Universidade Federal de Pernambuco
Centro de Ciencias Exatas e da Natureza
Programa de Pós-Graduação em Estatística

Thiago VedoVatto

INFERENCE, INFORMATION THEORY AND SEGMENTATION BASED ON AN EXTENDED CAUCHY-RAYLEIGH DISTRIBUTION:

Applications to Heavy Tailed Data

Recife
2019

Thiago VedoVatto

INFERENCE, INFORMATION THEORY AND SEGMENTATION BASED ON AN EXTENDED CAUCHY-RAYLEIGH DISTRIBUTION:

Applications to Heavy Tailed Data

Tese apresentada ao Programa de Pós-Graduação em Estatística do Centro de Ciências Exatas e da Natureza na Universidade de Federal de Pernambuco, para a obtenção de Título de Doutor em Ciências, na Área de Estatística.

Área de concentração: Inferência Paramétrica

Orientador: Abraão David Costa do Nascimento

Recife
2019

Catalogação na fonte
Bibliotecária Monick Raquel Silvestre da S. Portes, CRB4-1217

V416i Vedovatto, Thiago
Inference, information theory, and segmentation based on the extended cauchy-rayleigh model: applications to heavy-tailed data / Thiago Vedovatto. – 2019.
161 f.: il., fig., tab.

Orientador: Abraão David Costa do Nascimento.
Tese (Doutorado) – Universidade Federal de Pernambuco. CCEN, Estatística, Recife, 2019.
Inclui referências e apêndices.

1. Estatística. 2. Inferência. 3. Probabilidade. I. Nascimento, Abraão David Costa do (orientador). II. Título.

310 CDD (23. ed.) UFPE- MEI 2019-019

THIAGO VEDOVATTO

INFERENCE, INFORMATION THEORY, AND SEGMENTATION BASED ON THE EXTENDED CAUCHY-RAYLEIGH MODEL: APPLICATIONS TO HEAVY-TAILED DATA

Tese apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Doutor em Estatística.

Aprovada em: 25 de fevereiro de 2019.

BANCA EXAMINADORA

Prof.(º) Abraão David Costa do Nascimento
UFPE

Prof.(º) Francisco José de Azevêdo Cysneiros
UFPE

Prof.(º) Maria do Carmo Soares Lima
UFPE

Prof.(º) Hemílio Fernandes Campos Coelho
UFPB

Prof.(º) Renata Maria Cardoso Rodrigues de Souza
UFPE

Dedico essa tese à minha mãe, Clarice Ana Vedovatto, meu pai, Nilson Vedovatto, e meu irmão, Alexandre Vedovatto.

AGRADECIMENTOS

Agradeço aos professores do Instituto Federal de Goiás (IFG) em especial aos professores Iran Martins do Carmo e Maxwell Gonçalves Araújo por todo o apoio que me deram durante meu afastamento. Aos meus professores da Universidade Federal de São Carlos (UFSCar) pela forte inspiração para seguir em frente nessa jornada de qualificação dentro da estatística e igualmente aos professores da Universidade de Brasília (UnB).

A minha querida amiga Enai Taveira da Cunha, hoje professora da Universidade Federal Rural do Semi-Árido (UFERSA), por sua amizade sincera durante as horas tão difíceis que passei por aqui em Recife.

Aos meus amigos do Restaurante Universitário (RU) Andréa Leite, Larissa Lopes, Laryssa Kyonara, Juliana Yassunari, Raysa Albuquerque e Julio César por terem transfomado minhas refeições no RU em momentos de profunda alegria. A amizade de todos vocês foi uma das maiores conquistas de minha jornada em Recife. Será sempre com muitas saudades que lembrarei desses momentos.

A minha amiga Naara Marinheiro pela companhia nas caminhadas na pista da COMPESA. Sua amizade foi muito importante naquele momento.

Aos meus colegas de apartamento/república: Afonso, Alane, Benoit, Fernandes, Gislaine, Luiz Paulo, Márcio, Norah Patrícia, Ricardo Santiago, Robson (Quitin), Rodrigo Torres Cardona, Viriato e Yarianne.

A Alexandra Elbakyan fundadora do Sci-Hub (sci-hub.tw) e aos mantenedores do Library Genesis (gen.lib.rus.ec) por desafiarem a retrógrada e parasitária estrutura da publicação científica vigente, baseada na exploração de verbas públicas de pesquisa.

A secretária Valéria por ser tão eficiente e hábil em lidar de modo sério mesmo diante de tantos problemas que lhe são impostos. Aos funcionários do RU, pelo empenho em preparar refeições que me permitiram ter forças para realizar essa pesquisa. Ao pessoal da limpeza e a todos os funcionários terceirizados pelas funções tão importantes e as vezes invisíveis aos olhos de tantos.

Obrigado à professora Maria Helena por me ensinar a ler em 1992. Sua contribuição à esse trabalho é, de longe, a mais importante.

Finalmente, obrigado ao meu orientador Abraão David Costa do Nascimento por toda a paciencia e sabedoria ao longo desses quatro anos de trabalho.

ABSTRACT

The *Cauchy-Rayleigh* (CR) distribution has been successfully used to describe asymmetrical and heavy-tail events from radar imagery. Employing such model to describe lifetime data may then seem attractive, but drawbacks arise: its probability density function does not cover non-modal behavior as well as its *hazard rate function* (hrf) assumes only one form. To outperform this difficulty, it is investigated the *exponentiated Cauchy-Rayleigh* (ECR) distribution. This byparameteric model is flexible enough to accommodate hrf with decreasing, decreasing-increasing-decreasing and upside-down bathtub forms. Several closed-form mathematical expressions for the ECR model are obtained: median, mode, some moments, (h,ϕ)-entropies and Fisher information matrix. Their non-existence respective cases are also determined. It is proposed two estimators for the ECR parameters: *maximum likelihood* (ML) and percentile-based methods. Both of this methods may be biased for small and moderate sample sizes. To overcome it we furnish a expression for its second-order bias according to Cox and Snell (1968) and propose a third bias-corrected ML estimator. Further discussions about hypotheses-based inference and estimation formulas on censored-data are furnished as well. A simulation study is done to assess the estimators performance. The estimates existence and uniqueness are not guaranteed, thus procedures to constrained estimation are developed to overcome this trouble. Notes about hypothesis tests are given under censored and uncensored schemes. An application in a survival dataset illustrates the proposed model usefulness. Results point out that the ECR distribution may outperform classical lifetime biparametric models, such as the gamma, Birnbaum-Saunders, Weibull and log-normal laws, before heavy-tail data. The likelihood ratio test are compared against entropy-based tests as urban texture detector using a San Francisco synthetic aperture radar imagery. This same image is also the final application target which consist of compare segmentation algorithms based on CR and ECR entropies densities finite mixture. It is recurred to a result due Pardo *et al.* (1997) which provides the (h,ϕ)-entropies asymptotic distribution. The finite mixture log-likelihood function is maximized using the Expectation-Maximization algorithm.

Keywords: Cauchy-Rayleigh. Heavy-tail Distributions. Hypotheses Test. (h,ϕ)-Entropies. *Synthetic aperture radar* (SAR). Segmentation.

RESUMO

A distribuição de *Cauchy-Rayleigh* (CR) tem sido usada com sucesso para descrever dados assimétricos e eventos com caudas pesadas de imagens de radar. Empregar tal modelo para descrever dados de sobrevivência poder ser atrativo, mas inconvenientes surgem: sua função de densidade de probabilidade não abriga comportamento amodal bem como sua *função de taxa de falha* (hrf) assume apenas uma forma. Para superar essa dificuldade, é investigada a distribuição *Cauchy-Rayleigh exponencializada* (ECR). Este modelo biparamétrico é flexivel o bastante para acomodar hrf com formas decrescente, decrescente-crescente-decrescente e banheira invertida. Várias expressões matemáticas em forma fechada para o modelo ECR são obtidas: mediana, moda, alguns momentos, (h,ϕ)-entropias e matriz de informação de Fisher. Seus respectivos casos de não existência também são determinados. São propostos dois estimadores para os parâmetros da ECR: métodos de *máxima verossimilhança* (ML) e estimação quantílica. Ambos os métodos podem ser viesados para tamanhos de amostra pequenos e moderados. Para superar isto, fornecemos uma expressão para o viés de segunda ordem de acordo com Cox e Snell (1968) e propusemos um estimador de máxima verossimilhança com viés de terceira ordem corrigido. Discussões adicionais sobre inferência baseada em hipóteses e fórmulas de estimação em dados censurados são fornecidas. Um estudo de simulação é feito para aferir a performance dos estimadores. A existência e unicidade das estimativas não é garantida, assim procedimentos para estimação restrita são desenvolvidos para superar esse problema. Notas sobre teste de hipótese são dadas considerando esquemas censurados e não-censurados. Uma aplicação em dados de sobrevida ilustra a utilidade do modelo proposto. Os resultados apontam que a distribuição ECR pode superar modelos biparamétricos de sobrevivência clássicos como gama, Birnbaum-Saunders, Weibull e log-normal. O teste da razão de verossimilhança é comparado com testes baseados em entropias como detector de texturas urbanas em uma imagem synthetic aperture radar da cidade de São Francisco. Esta mesma imagem é também alvo da aplicação final que consiste em comparar algorítmos de segmentação baseados em misturas finitas das densidades das entropias dos modelos CR e ECR. Recorremos a um resultado de Pardo *et al.* (1997) o qual nos dá distribuição assintótica das (h,ϕ)-entropias. A função de log-verossimilhança das misturas finitas é maximizada usando o algorítmo EM.

Palavras-chave: Cauchy-Rayleigh. Distribuições de Cauda Pesada. Radar de Abertura Sintética. Segmentação. Teste de Hipótese. (h,ϕ)-Entropias.

LIST OF ABBREVIATIONS AND ACRONYMS

F^{α}R	F^{α} record
αS	α-stable
a.k.a.	also known as
AD	Anderson-Darling
AET	Arimoto entropy test
AIC	Akaike information criterion
ARI	adjusted rand index
BBXII	beta Burr type-XII
BIC	bayesian information criterion
BS	Birnbaum-Saunders
Bs	Bowley skewness
BX	Burr type-X
BXII	Burr type-XII
CAIC	consistent Akaike information criterion
cdf	cumulative distribution function
cf	characteristic function
cgf	cumulant-generating function
cpsf	critical percentile-based score function
CR	Cauchy-Rayleigh
CS	Cox-Snell corrected
CvM	Cramér-von Mises
CW	compound Weibull
EBT	entropy-based test
EBXII	exponentiated Burr type-XII
ecdf	empiriral cumulative density function
ECR	exponentiated Cauchy-Rayleigh
EE	exponentiated exponential
ELi	exponentiated Lindley
ELL	exponentiated log-logistic
EM	Expectation-Maximization
Exp	exponential
Exp-G	exponentiated
FIM	Fisher information matrix
FM	finite mixture
FP	Feller-Pareto

FW	flexible Weibull
GCE	global consistency error
GFP	generalized Feller-Pareto
GG	generalized gamma
GHN	generalized half-normal
GoF	goodness-of-fit
HQIC	Hannan-Quinn information criterion
hrf	hazard rate function
HTR	heavy-tailed Rayleigh
IαS	isotropic α-stable
IC	interval censored
ICA	independent component analysis
ICL	integrated completed likelihood
IG	inverse gamma
iid	independent and identically distributed
ISP	image segmentation performance
ITSM	inverse transform sampling method
K	K-Bessel
KBXII	Kumaraswamy Burr type-XII
KL	Kullback-Leibler
KLL	Kumaraswamy log-logistic
KS	Kolmogorov-Smirnov
LC	left censored
LCa	log-Cauchy
LCE	local consistency error
LE	Lindley exponential
lf	likelihood function
LL	log-logistic
lle	log-likelihood equation
llf	log-likelihood function
LN	log-normal
LRT	likelihood ratio test
LSA	latent semantic analysis
LSE	least squares estimator
McBXII	McDonald Burr type-XII

McLL	McDonald log-logistic
MCR	miss-classification rate
MDR	multifactor dimensionality reduction
Mk	Moors kurtosis
ML	maximum likelihood
MLE	maximum likelihood estimator
NH	Nadarajah-Haghighi's exponential
NPC	normal parsimonious clustering
PB	percentile-based
PBE	percentile-based estimator
PCA	principal component analysis
pdf	probability density function
pef	percentile-based estimation function
pllf	profile log-likelihood function
PLS	partial least squares
PolSAR	polarimetric SAR
ppef	profile percentile-based estimation function
ppsf	profile percentile-based score function
psf	profile score function
pwm	probability weighted moment
q-We	q-Weibull
qf	quantile function
RC1	type I right censored
RC2	type II right censored
RET	Rényi entropy test
RST	Rao score test
RVI	regularly varying at infinity
SαS	symmetric α-stable
SAR	synthetic aperture radar
SET	Shannon entropy test
sf	survival function
SM	Singh-Maddala
SMED	segmentation based on mixture of stochastic entropy distributions
SSD	sample standard deviation

SVD	singular value decomposition
TB	transformed beta
TET	Tsallis entropy test
TTT	total time on test
VOI	variation of information
WT	Wald test

LIST OF SYMBOLS

Notation	Description
$\mathrm{B}(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}\,\mathrm{d}t$	beta[1] function
$\mathrm{Bias}(\hat{\theta}) = \mathbb{E}(\hat{\theta}) - \hat{\theta}$	bias of the of θ estimate.
$F_X(x) = \Pr(X \leq x)$	cumulative distribution function of a random variable X
$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{t^2}{2}\right)\,\mathrm{d}t$	cdf of the standard normal distribution.
$\varphi_X(t) = \mathbb{E}(\mathrm{e}^{itX})$	characteristic function of a random variable X
$K_X(t) = \log\left[\mathbb{E}(\mathrm{e}^{tX})\right]$	cumulant-generating function of a random variable X
$X_n \xrightarrow[n\to\infty]{d} X$	convergence in distribution
Δ	discriminant of a quadratic equation
$\mathcal{H}_A(q) = \frac{q}{1-q}\left[\left(\int_{-\infty}^{\infty} f(x)^q\,\mathrm{d}x\right)^{\frac{1}{q}} - 1\right]$	Arimoto entropy of a continuous random variable X with additional parameter $q > 0$ and $q \neq 1$.
$\mathcal{H}$	generic entropy
$\mathcal{H}_R(q) = \frac{1}{1-q}\log\left(\int_{-\infty}^{\infty} f(x)^q\,\mathrm{d}x\right)$	Rényi entropy of a continuous random variable X with additional parameter $q > 0$ and $q \neq 1$.
$\mathcal{H}_S = \mathbb{E}\left\{-\log[f(X)]\right\}$	Shannon entropy[2] of a continuous random variable X.
$\mathcal{H}_S^c(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = -\int_{-\infty}^{\infty} f_1(x)\log f_2(x)\,\mathrm{d}x$	cross Shannon entropy of a continuous random variable X_1 relative to a random variable X_2.
$\mathcal{H}_T(q) = \frac{1}{q-1}\left(1 - \int_{-\infty}^{\infty} f(x)^q\,\mathrm{d}x\right)$	Tsallis entropy[3] of a continuous random variable X with additional parameter $q \in \mathbb{R}$ and $q \neq 1$.
$\kappa_1 = K'(0)$	first cumulant
$\kappa_{\theta_i\theta_j}^{(\theta_k)} = \frac{\partial}{\partial\theta_k}\kappa_{\theta_i\theta_j}$	first derivative of the cumulants with respect to the parameters.

[1] a.k.a. Euler integral of the first kind
[2] a.k.a. Boltzmann-Gibbs entropy
[3] a.k.a. Havrda-Charvât entropy

Notation	Description
$\kappa_{\theta_1 \cdots \theta_n} = \mathbb{E}\left(U_{\theta_1 \cdots \theta_n}\right)$	cumulant of the n^{th} log-likelihood derivative.
$\boldsymbol{\gamma} = \lim_{n \to \infty} \left(-\log n + \sum_{k=1}^{n} \frac{1}{k} \right)$	Euler-Mascheroni constant.
$\mathbf{K}_{i,j} = [\mathbf{K}(\boldsymbol{\theta})]_{i,j} = -\kappa_{\theta_i \theta_j}$	$(i,j)^{th}$ element of the Fisher information matrix for $\boldsymbol{\theta}$
$\mathbf{K}^{\text{-}1}{}_{i,j} = [\mathbf{K}^{\text{-}1}(\boldsymbol{\theta})]_{i,j} = -\kappa^{\theta_i, \theta_j}$	$(i,j)^{th}$ element of the inverse Fisher information matrix for $\boldsymbol{\theta}$
$\boldsymbol{\Psi} = (\pi_1 \ldots, \pi_{g-1}, \boldsymbol{\xi}^\top)^\top$	finite mixture parameter vector
$\mathbf{L}_\theta = \frac{\partial \mathfrak{p}(\boldsymbol{\theta})}{\partial \theta}$	profile percentile-based score function
$\Gamma(a) = \int_0^\infty t^{a-1} \, \mathrm{e}^{-t} \, \mathrm{d}t$	gamma[4] function
$\Psi(x) = \frac{\mathrm{d}}{\mathrm{d}x} \log \Gamma(x)$	digamma function.
${}_2\mathrm{F}_1(a,b;c;x) = \sum_{k=0}^{\infty} \frac{(a)_k (b)_k}{(c)_k} \frac{x^k}{k!}$	Gauss's hypergeometric function
$\mathrm{F}_1(a,b_1,b_2;c;x,y) = \sum_{i,j=0}^{\infty} \frac{(a)_{i+j} (b_1)_i (b_2)_j}{(c)_{i+j}} \frac{x^i}{i!} \frac{y^j}{j!}$	first Appell's hypergeometric function
$(a)_k = \frac{\Gamma(a+k)}{\Gamma(a)}$	Pochhammer symbol.
$h_X(t) = \frac{f_X(x)}{1 - F_X(x)}$	hazard rate function of a random variable X.
$\mathcal{K}_M = \frac{\sum_{k=1}^{4} (-1)^k Q\left(\frac{2k-1}{8}\right)}{\sum_{k=1}^{2} (-1)^k Q\left(\frac{2k-1}{4}\right)}$	Moors kurtosis of implicit random variable.
$\mathcal{L}(\boldsymbol{\theta}) = \mathcal{L}(\boldsymbol{\theta}\|\mathbf{x}) = \prod_{i=1}^{n} f_X(x_i\|\boldsymbol{\theta})$	likelihood function of $\boldsymbol{\theta}$ given the data $\mathbf{x}$
$\ell(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta}\|\mathbf{x}) = \sum_{i=1}^{n} \log[f_X(x_i\|\boldsymbol{\theta})]$	log-likelihood function of $\boldsymbol{\theta}$ given the data $\mathbf{x}$
$\boldsymbol{\Phi}(x; s, a) = \sum_{n=0}^{\infty} \frac{x^n}{(n+a)^s}$	Lerch transcendental Phi function.
$\mathcal{E}(\mathcal{C}, \mathcal{C}', p_i) = \frac{\lvert R(\mathcal{C}, p_i) \backslash R(\mathcal{C}', p_i) \rvert}{\lvert R(\mathcal{C}, p_i) \rvert}$	Local error of the refinement $\mathcal{C}$ for the cluster $\mathcal{C}'$ in pixel p_i.
$R(\mathcal{C}, p_i)$	Set of pixels corresponding to the region in cluster $\mathcal{C}$ that contains the pixel p_i.

[4] a.k.a. complete gamma

Notation	Description
$\boldsymbol{\Lambda}(\boldsymbol{x}) = \dfrac{\sup_{\boldsymbol{\theta}_0} \mathcal{L}(\boldsymbol{\theta}\|\boldsymbol{x})}{\sup_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}\|\boldsymbol{x})}$	likelihood ratio test statistic for testing $H_0 : \boldsymbol{\theta} \in \Theta_0$ versus $H_1 : \boldsymbol{\theta} \in \Theta_0^c$
$\widetilde{M} = Q_X(1/2)$	median of a random variable X
$\mathrm{Mo}_X = \arg\max_x f_X(x)$	mode of random variable X
$\mathfrak{M} = \{\mathrm{CR}\,, \mathrm{ECR}\}$	A set containing the ECR and CR models
$\mu = \mathbb{E}(X)$	expected value of the random variable X.
$\mu'_{r,s,t} = \mathbb{E}\left(X^r F(X)^s [1 - F(X)]^t\right)$	$(r,s,t)^{th}$ probability weighted moments of random variable X.
$\overline{X^r_{s,t}} = \frac{1}{n}\sum_{i=1}^n \left(x_i^r F(x_i)^s [1 - F(x_i)]^t\right)$	$(r,s,t)^{th}$ sample probability weighted moments of random variable X.
$m_r(x_0) = \int_{-\infty}^{x_0} x^r f_X(x)\,\mathrm{d}x$	r^{th} incomplete moment of a continuous random variable X.
$X \sim \mathcal{N}(\mu,\sigma^2)$	X following normal distribution with mean μ and variance σ^2.
Θ	parameter space
$\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$	parameter vector
$f_X(x) = \dfrac{\mathrm{d}}{\mathrm{d}x} F_X(x)$	probability density function of a continuous random variable X
$f_{(i)}(x) = \dfrac{f(x)F(x)^{i-1}S(x)^{n-i}}{\mathrm{B}(i,n-i+1)}$	probability density function of the i^{th} order statistic $X_{(i)}$ for $i = 1, \ldots, n$
$X_{(1)} \leq \ldots \leq X_{(n)}$	i^{th} order statistic $X_{(i)}$ for $i = 1, \ldots, n$
$\phi\left(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}\right) = \dfrac{\exp\left[-\frac{1}{2}\left(\boldsymbol{x} - \boldsymbol{\mu}\right)^\top \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{x} - \boldsymbol{\mu}\right)\right]}{\sqrt{(2\pi)^p \lvert\boldsymbol{\Sigma}\rvert}}$	pdf of the multivariate normal distribution.
$\mathfrak{p}(\boldsymbol{\theta}) = \sum_{i=1}^n \left[Q(p_i) - x_{(i)}\right]^2$	percentile-based estimation function of $\boldsymbol{\theta}$ given the data $\mathbf{x}$
$\mathfrak{C}_{(H,V)} = \{HH, HV, VV\}$	Set of SAR polarization channels under a linear polarization basis (H,V)
$\Pr(E)$	probability of a random variable E
$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}x$	expectation, or the expected value of a continuous random variable X
$\mathrm{Var}(X) = \sigma^2 = \mathbb{E}[(X - \mu)^2]$	variance of a random variable X

Notation	Description
$Q_X(p) = \inf\{x \in \mathbb{R}; p \leq F_X(x)\}$	quantile function of a random variable X.
$\mathcal{R}(\boldsymbol{x}) = n^{-1}\mathbf{U}(\boldsymbol{\theta}_0)^\top \mathbf{K}^{-1}(\boldsymbol{\theta}_0)\mathbf{U}(\boldsymbol{\theta}_0)$	Rao score test statistic for testing $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ versus $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$
$\overline{\mathcal{Q}}_i$	$100i\%$ percentile of a sample.
$\mathbf{U} = (U_{\theta_1}, \ldots, U_{\theta_n})$	score vector relative to parameter vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$.
$U_\theta = \dfrac{\partial \ell(\boldsymbol{\theta})}{\partial \theta}$	score vector component relative to parameter θ.
$U_{\theta_1 \cdots \theta_n} = \dfrac{\partial \ell(\boldsymbol{\theta})}{\partial \theta_1 \cdots \partial \theta_n}$	partial log-likelihood function derivative relative to parameters $\theta_1, \ldots, \theta_n$.
$\mathbb{I}$	set of imaginary numbers
$\mathbb{R}$	set of real numbers
$\mathbb{Z}$	set of integers numbers
$S_X(t) = 1 - F_X(t)$	survival function of a random variable X.
$\mathcal{B}_\Omega$	σ-algebra of Ω subsets
$\mathcal{S}_B = \dfrac{\sum_{k=1}^{3}(-1)^{k-1}\binom{2}{k-1}Q\left(\frac{k}{4}\right)}{\sum_{k=1}^{2}(-1)^{k}Q\left(\frac{2k-1}{4}\right)}$	Bowley skewness of implicit random variable.
$\mathfrak{W}_{5\times 5}$	Set of all 5×5 windows of the San Francisco image
$\mathcal{W}(\boldsymbol{x}) = n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})^\top \mathbf{K}(\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})$	Wald test statistic for testing $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ versus $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$

CONTENTS

1 INTRODUCTION

Synthetic aperture radar (SAR) systems have been indicated as important tools to solve remote sensing issues. Among others, this fact may be justified by their capability in operating on all weather conditions and in providing high resolution images. However, SAR images are strongly contaminated by an interference pattern called *speckle noise*. Thus, working with SAR imagery requires a suitable modeling. The *Cauchy-Rayleigh* (CR)[1] distribution has received great attention as descriptor of SAR features. In this work, we propose a new distribution that extends the CR law as modeling for both lifetime data and SAR features.

The reason of applying the CR law for describing lifetimes is firstly to accommodate heavy-tail distributions. However, its *probability density function* (pdf) does not cover non-modal behavior (as the exponential distribution does) as well as the CR *hazard rate function* (hrf) assumes only one form, limiting its employment in practice. To that end, we extend the CR distribution by using the *exponentiated* (Exp-G) class. We denote the new model as *exponentiated Cauchy-Rayleigh* (ECR) distribution.

As a second goal we aim to combining the CR and ECR models with information theory measures to construct both extractors of SAR features and segmenters of SAR clusters. Kuruoglu and Zerubia (2004) presented evidence that the CR model may outperform the Weibull, *log-normal* (LN) and *K-Bessel* (K) distributions in SAR imagery analysis, which is expected because urban areas show impulsive characteristics that correspond to underlying heavy-tailed distributions. Li and Ekman (2010), Li and Ekman (2011) introduced the CR law as a model of scattering clusters in state space based on simulation model for single and multiple polarization channels. Hill *et al.* (2014) produced a novel bivariate shrinkage technique to provide a quantitative improvement in image denoising using the CR model. Recently, Bibalan and Amindavar (2015) and Bibalan and Amindavar (2016) furnished a mathematical treatment to the *heavy-tailed Rayleigh* (HTR) distribution by means of mixtures of CR and Rayleigh models. Other papers that used the CR in SAR modelling are Peng *et al.* (2017) and Pappas *et al.* (2017). Recently, texture surfaces motivated many segmentation studies and algorithms like Jain and Farrokhnia (1990), Oakley and Hancock (1994), Randen and Husoy (1999), Hsin (2000), Huan and Hou (2008), Zhang *et al.* (2008), Karoui *et al.* (2010), Jyothirmayi *et al.* (2015), Kumar *et al.* (2016) and Akbulut *et al.* (2018). Li and Ekman (2010), Li and Ekman (2011) developed this approach for modeling amplitude of ultrasound images through the HTR distribution.

Some recent works that use *entropy-based tests* (EBTs) in SAR and *polarimetric SAR* (PolSAR) imagery are Frery *et al.* (2013) and Nascimento *et al.* (2014b). Frery *et*

[1] a.k.a. generalized Cauchy (BIBALAN; AMINDAVAR, 2016).

al. (2013) and Nascimento *et al.* (2014b) have shown evidences that employing entropy measures in order to understand SAR partner may be an efficient strategy, comparatively to others such like those based on *likelihood ratio test* (LRT) statistics. An important EBTs advantage is the smaller demand of computational effort than LRT alternatives. The employment of some entropy measures on texture segmentation has been made with success, some of the last studies are Nobre *et al.* (2016), Naidu *et al.* (2017), Abdel-Khalek *et al.* (2017), Yin *et al.* (2017), Nguyen *et al.* (2018), Pham *et al.* (2018), Wang *et al.* (2018), Wu *et al.* (2018) and Wang *et al.* (2018).

In this thesis, we have the following main objectives:

First goal: Extend the CR distribution by using the Exp-G class exploring flexibility, structural properties, parameter estimators, asymptotic theory, hypothesis test and confidence intervals of the new model.

Second goal: For urban SAR imagery understanding develop CR- and ECR- based procedures to identify homogeneity EBTs and LRTs.

Third goal: Propose segmentation methods for SAR imagery by means of the k-means and *finite mixture* (FM)-based algorithms using stochastic entropies.

1.1 SPECIFIC OBJECTIVES

This work performs contributions in various areas: inference, SAR imagery, information theory, survival analysis and asymptotic theory. Figure 1.1 presents a diagram illustrating this relationship. The following technical objectives are addressed in this work:

Obj. 1 Defining the ECR distribution from the CR model, employing the Exp-G class.

Obj. 2 Deriving closed-form expressions for some of ECR structural properties.

Obj. 3 Obtaining the ECR *maximum likelihood* (ML) and *percentile-based* (PB) estimates under unconstrained and constrained estimation schemes.

Obj. 4 Getting the ML estimates for the ECR parameters under censored and uncensored estimation schemes.

Obj. 5 Comparing numerically the ECR ML and PB estimators under uncensored and unconstrained schemes focusing on convergence, bias, *sample standard deviation* (SSD) and existence rate of their corresponding estimates.

Obj. 6 Using the Cox and Snell (1968) methodology to obtain expressions for the second order bias of *maximum likelihood estimators* (MLEs) and employ them to obtain the *Cox-Snell corrected* (CS)-MLEs.

Obj. 7 Assessing the flexibility of the ECR distribution by comparing its *goodness-of-fit* (GoF) measures to similar models in a real survival dataset.

Obj. 8 Developing the LRTs and EBTs supported by the ECR class.

Obj. 9 Evaluating the power and nominal level of the proposed tests by means of simulation studies.

Obj. 10 Applying the proposed LRTs and EBTs as models for SAR urban imagery.

Obj. 11 Comparing the LRT and EBTs GoF measures (supported by the ECR class) in urban SAR imagery.

Obj. 12 Measuring how able the LRT and EBTs (supported by the ECR class) are to detect equivalence of SAR imagery textures.

Obj. 13 Quantifying how able the LRT and EBTs (supported by the ECR class) are to detect redundancy of information in different SAR imagery channels.

Obj. 14 Employing the deduced ECR stochastic entropies to support feature vectors of SAR imagery.

Obj. 15 Using these feature vectors to support k-means and FM-based clustering algorithms.

Obj. 16 Presenting empirical evidences that the ECR new shape parameter improves these clustering algorithms.

Obj. 17 Minimizing the *miss-classification rate* (MCR) of the FM-based methods optimizing the order parameter of the q-entropies.

Obj. 18 Showing the main characteristics of the segmented images obtained by means of each reasoned clustering algorithms.

Obj. 19 Determining what of the proposed FM-based clustering algorithms optimizes the *image segmentation performance* (ISP) and GoF measures in segmenting the San Francisco SAR image.

1.2 COMPUTATIONAL PLATFORM

The computer we used in this work was an Intel® Core™ i7-4500U CPU at 1.80 GHz processor with base x64, Ubuntu 16.04.5 operating system. The computational platforms required to develop this work are

Plat. 1. `R` version 3.4.4 (R, 2018);

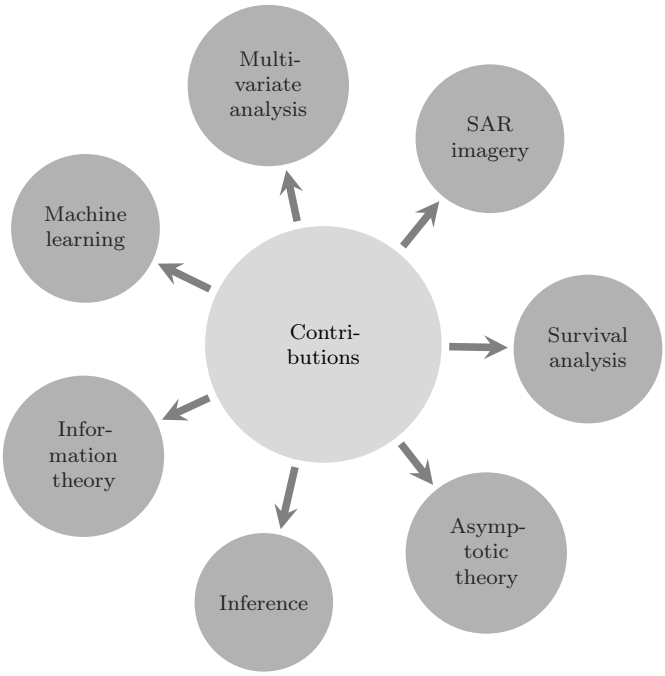


Figure 1.1: Diagram of the thesis contributions.

Plat. 2. `Wolfram Mathematica` version 11.1.1 (WOLFRAM RESEARCH, 2018);

Plat. 3. `Maple` version 2015.0 (WATERLOO MAPLE, 2015).

Plat. 1. was used to fit the studied models in real datasets, to perform numerical studies and to produce the graphics. All mathematical expressions deduced in this work were obtained manually and, subsequently, confirmed by Plat. 2. and Plat. 3..

1.3 SCOPE AND ORGANIZATION OF THIS THESIS

The present chapter was a brief introduction of the main concepts and ideas that will be developed in Chapters 3 to 5. This work is organized as follows. A useful review of the main concepts, methods and techniques used in this work is presented in Chapter 2. Chapter 3 presents the main structural properties of the ECR model, investigates the ML, CS-ML and PB estimators under censored/constrained and uncensored/unconstrained schemes and gives a comparison of the ECR model to similar and competitor models achieving Obj. 1 to Obj. 7. A study of the CR and ECR (h,ϕ)-entropies is made in Chapter 4. These entropies give support to develop the EBTs. These tests are compared to the classical LRT. This chapter attends Obj. 1 and Obj. 8 to Obj. 13. Chapter 5 has some segmentation methods based on FM models and ECR (h,ϕ)-entropies answering Obj. 14 to Obj. 19. The main conclusions of this text are summarized in Chapter 6.

The last pages are destined to some brief appendices. The first one, Appendix A, is reserved to figures relative of the simulation studies carried out in Chapter 3. Appendix B

bring tables corresponding to applications and simulations made in Chapters 4 and 5. Some proofs of the main results presented in Chapters 3 and 4 can be found in Appendix C.

2 LITERATURE REVIEW AND USEFUL REMARKS

Here we introduce the main concepts, methods and techniques which will be explored in Chapters 3 to 5. This review unfolds as follow. Section 2.1 presents a brief discussion about the origins and applications of the CR model obtained from the α-*stable* (αS) distribution. The Exp-G class of distributions is the theme of Section 2.2 followed by two possible applications of this class. Section 2.3 is dedicated to CS-MLEs. The main information theory measures is introduced in Section 2.4. Section 2.5 uses these measures to support two-sample LRTs and EBTs. Some SAR imagery and FM concepts are explored in Sections 2.6 and 2.7. Section 2.8 introduces the *Expectation-Maximization* (EM) algorithm and its particular configuration assuming multivariate normal components. Finally, Section 2.9 defines the ISP measures adopted in the Chapter 5 applications.

2.1 THE α-STABLE AND CR MODELS

Lévy (1925) pioneered the αS distribution, which has been widely used in financial time series (MANDELBROT, 1960; FAMA; SCHWERT, 1977; VOIT, 2005) and, recently, applied in SAR image processing (PIERCE, 1996; WANG *et al.*, 2008; PENG *et al.*, 2011). This law has not tractable pdf expression and is often represented by its *characteristic function* (cf) given by: For $-\infty < t < \infty$ and $\boldsymbol{\jmath} = \sqrt{-1}$,

$$\varphi_{\alpha S}(t) = \begin{cases} \exp\left\{\boldsymbol{\jmath}\zeta t - \lambda|t|^{\alpha}\left[1 + \boldsymbol{\jmath}\rho\,\mathrm{sign}(t)\tan\left(\frac{\alpha\pi}{2}\right)\right]\right\}, & \text{if } \alpha \neq 1, \\ \exp\left\{\boldsymbol{\jmath}\zeta t - \lambda|t|\left[1 + \boldsymbol{\jmath}\rho\,\mathrm{sign}(t)\frac{2}{\pi}\log|t|\right]\right\}, & \text{if } \alpha = 1, \end{cases}$$

where $-\infty < \zeta < \infty$, $\lambda > 0$, $0 < \alpha \leq 2$ and $-1 \leq \rho \leq 1$ are the location, scale, characteristic and symmetry parameters, respectively.

Now consider the (zero-mean) *symmetric* α-*stable* (SαS) distribution having cf

$$\varphi_{S\alpha S}(t) = \exp(-\lambda|t|^{\alpha}). \tag{2.1}$$

According to Kuruoglu and Zerubia (2004), we can define the bivariate *isotropic* α-*stable* (IαS) distribution, which has cf as

$$\varphi_{I\alpha S}(t_{\mathbb{R}},t_{\mathbb{I}}) = \exp(-\lambda|\boldsymbol{t}|^{\alpha}), \tag{2.2}$$

where $t_{\mathbb{R}}$ and $t_{\mathbb{I}}$ can be understood as outcomes of real and imaginary parts of a complex random variable, say $\boldsymbol{t}$, respectively, and $|\boldsymbol{t}| = \sqrt{t_{\mathbb{R}}^2 + t_{\mathbb{I}}^2}$ represents the amplitude of $\boldsymbol{t}$.

From (2.1), the bivariate IαS pdf can be obtained by taking the Fourier transform of (2.2):

$$f_{\alpha,\lambda}(x_{\mathbb{R}},x_{\mathbb{I}}) = \frac{1}{(2\pi)^2}\int_{t_{\mathbb{R}}}\int_{t_{\mathbb{I}}}\exp(-\lambda|\boldsymbol{t}|^{\alpha})\exp[-j2\pi(x_{\mathbb{R}}t_{\mathbb{R}}+x_{\mathbb{I}}t_{\mathbb{I}})]\,\mathrm{d}t_{\mathbb{I}}\,\mathrm{d}t_{\mathbb{R}}.$$

Representing this integral into its polar form in terms of $s=|\boldsymbol{t}|$ and $\omega=\arctan({}^{t_{\mathbb{R}}}/_{t_{\mathbb{I}}})$, one has

$$f_{\alpha,\lambda}(x_{\mathbb{R}},x_{\mathbb{I}}) = \frac{1}{(2\pi)^2}\int_0^{2\pi}\int_0^{\infty} s\exp(-\lambda s^{\alpha})J_0(s|\boldsymbol{x}|)\,\mathrm{d}s\,\mathrm{d}\omega, \tag{2.3}$$

where J_0 is the zero order Bessel function of the first kind (ABRAMOWITZ; STEGUN, 1972) and $|\boldsymbol{x}|=\sqrt{x_{\mathbb{R}}^2+x_{\mathbb{I}}^2}$. Since ω does not appear in the integrand of the (2.3), it collapses in

$$f_{\alpha,\lambda}(x_{\mathbb{R}},x_{\mathbb{I}}) = \frac{1}{2\pi}\int_0^{\infty} s\exp(-\lambda s^{\alpha})J_0(s|\boldsymbol{x}|)\,\mathrm{d}s. \tag{2.4}$$

Now consider to determine the pdf of the corresponding amplitude. Using polar coordinates transformation (where r and ϕ indicate amplitude and phase of $x_{\mathbb{R}}+\boldsymbol{j}x_{\mathbb{I}}$ respectively), the following joint pdf can be obtained, for $r>0$ and $0\leq\phi\leq 2\pi$,

$$f(r,\phi) = r f_{\alpha,\lambda}(r\cos(\phi),r\sin(\phi)). \tag{2.5}$$

Under the common assumption that ϕ is uniformly distributed on $[0,2\pi]$ (KURUOGLU; ZERUBIA, 2004), it is possible to determine an expression for the amplitude distribution from replacing (2.4) in (2.5) and integrating over ϕ:

$$g_{\alpha,\lambda}(r) = r\int_0^{\infty} s\exp(-\lambda s^{\alpha})J_0(rs)\,\mathrm{d}s. \tag{2.6}$$

This expression is know as the HTR pdf (NIKIAS; SHAO, 1995).

When $\alpha=1$ in (2.6) and using the identity 6.623.2 given by Gradshteyn and Ryzhik (2007), the CR pdf is obtained, for $r>0$, as

$$g_{\mathrm{CR}}(r) = \frac{\lambda r}{(\lambda^2+r^2)^{3/2}},$$

where $\lambda>0$ is the scale parameter. This pdf is associated with the amplitude of a coefficient on which the components are jointly Cauchy distributed (KURUOGLU; ZERUBIA, 2004). The CR *cumulative distribution function* (cdf) is

$$G_{\mathrm{CR}}(r) = 1-\frac{\lambda}{\sqrt{\lambda^2+r^2}}. \tag{2.7}$$

The integral in (2.6) has closed-form only for $\alpha=1$ and $\alpha=2$. When $\alpha=2$, and using

the identity 6.631.1 obtained by Gradshteyn and Ryzhik (2007), the integral results

$$g_{(\alpha=2)}(r) = \frac{r}{2\lambda}\exp\left(-\frac{r^2}{4\lambda}\right),$$

which correspond to the classical Rayleigh distribution.

Under a non-physical perspective, some remarks are important. The CR distribution is a nested case in the *compound Weibull* (CW)[1] distribution with cdf, for $x > 0$ is given by

$$F_{\mathrm{CW}}(x) = 1 - \exp(-ax^c)\left[1 + \left(\frac{x}{\lambda}\right)^c\right]^{-k},$$

where $c \geq 0$ and $k \geq 0$ are shape parameters and $a \geq 0$ and $\lambda > 0$ are scale parameters. The CR distribution is obtained at $c = 2$, $k = 1/2$ and $a = 0$. At $a = 0$, the CW distribution coincides with the *Singh-Maddala* (SM)[2] model, which is another notable CR generalization. The CR distribution is also a particular case of other models, like the *Feller-Pareto* (FP) (FELLER, 1971; ARNOLD; LAGUNA, 1977) and *transformed beta* (TB)[3] (KLUGMAN *et al.*, 2012) laws. However, although this uniparametric model has been successfully employed as reviewed above, it does not seem to exist works about mathematical properties of its biparametric extension the ECR law. In what follows, this gap is covered, presenting several new closed-form properties of the ECR model. Some ECR asymptotic theory results are also provided. It is hoped that the proposed analytical developments are

- used to improve processing of images whose features of pixels are distributed ECR data such like in (LI; EKMAN, 2011; HILL *et al.*, 2014; PENG; ZHAO, 2014; BIBALAN; AMINDAVAR, 2016; PENG *et al.*, 2017);
- employed in the lifetime data context.

2.2 THE EXPONENTIATED CLASS OF DISTRIBUTIONS

The exponentiation method may be described briefly as: Let X be a random variable with cdf $G(x)$, its Exp-G version is a random variable with cdf and pdf given, respectively, by

$$F(x) = G(x)^\beta$$

[1] a.k.a. Weibull-gamma distribution (RINNE, 2009, p. 157).

[2] a.k.a. three-parameter *Burr type-XII* (BXII) distribution (MCDONALD, 1984), which is the Pareto VI model with null location parameter. Dubey (1968) deduce the SM model as a compound with Weibull and gamma distributions and also named it as compound Weibull. An important reparameterization of this distribution is the *q-Weibull* (q-We) model (TSALLIS, 1988).

[3] a.k.a. beta prime, inverted beta, beta distribution of the second kind (JOHNSON *et al.*, 1995) or beta log-logistic (LEMONTE, 2014).

and

$$f(x) = \beta g(x) G(x)^{\beta-1},$$

where $\beta > 0$ is an additional shape parameter and $g(x)$ is the pdf of the baseline $G(x)$. The hrf and the reversed hrf can be expressed, respectively, as

$$h(x) = \beta g(x) \frac{G(x)^{\beta-1}}{1 - G(x)^{\beta}}$$

and

$$h^*(x) = \beta \frac{g(x)}{G(x)}. \tag{2.8}$$

An interesting property of the Exp-G hrfs was deduced by Gupta *et al.* (1998):

- if $\beta > 1$ and G admits increasing hrf, then F admits increasing hrf;
- if $\beta < 1$ and G admits decreasing hrf, then F admits decreasing hrf.

The Exp-G class also presents β proportional reversed hrf, as evinced by Eq. (2.8). Clearly the *survival function* (sf) of the Exp-G class can be written in the form

$$S(x) = 1 - G(x)^{\beta}. \tag{2.9}$$

Al-Hussaini and Ahsanullah (2015, p. 17) presented several interesting properties of the Exp-G class. Following their notes note that any (absolutely continuous) baseline sf S_G can be written in the form

$$S_G(x) = \exp[-u(x)], \quad 0 \leq a < x < b \leq \infty, \tag{2.10}$$

where

$$u(x) = -\ln S_G(x). \tag{2.11}$$

Thus $u(x)$ is a continuous, monotone increasing, differentiable function of x such that $\lim_{x \to a^+} u(x) = 0$ and $\lim_{x \to b^-} u(x) = \infty$. Applying (2.10) in (2.9) we obtain that the Exp-G hrfs can be written in the form

$$S(x) = 1 - \{1 - \exp[-u(x)]\}^{\beta}. \tag{2.12}$$

Al-Hussaini and Ahsanullah (2015, p. 18) showed that r^{th} moment of a random variable X following the sf (2.12) is given by

$$\mathbb{E}(X^r) = r \sum_{i=1}^{n} (-1)^{i-1} \binom{\beta}{i} \int_0^{\infty} x^{r-1} \exp[-iu(x)] \, \mathrm{d}x. \tag{2.13}$$

Table 2.1: Recent Exp-G distributions.

Baseline model	Works
Logistic	Verhulst (1838), Ahuja and Nash (1967) and Ali *et al.* (2007)
Weibull	Mudholkar and Srivastava (1993) and Nadarajah and Kotz (2006)
Gamma	Gupta *et al.* (1998) and Nadarajah and Kotz (2006)
Exponential	Gupta and Kundu (1999)
Generalized Pareto	Adeyemi and Adebanji (2004)
Rayleigh[1]	Kundu and Raqab (2005) and Surles and Padgett (2005)
Pareto	Nadarajah (2005) and Ali *et al.* (2007)
Fréchet[2]	Nadarajah and Kotz (2006) and Ali *et al.* (2007)
Gumbel	Nadarajah and Kotz (2006)
Generalized uniform	Ali *et al.* (2007)
Double Exponential	Ali *et al.* (2007)
Double Weibull	Ali *et al.* (2007)
Double Fréchet	Ali *et al.* (2007)
Generalized Gamma	Cordeiro *et al.* (2011)
Generalized Wald[3]	Lemonte and Cordeiro (2011)
Lindley	Nadarajah *et al.* (2011)
Kumaraswamy	Lemonte *et al.* (2013)
Nadarajah-Haghighi's exponential (NH)	Lemonte (2013)
Perks	Singh and Choudhary (2016)

[1] The exponentiated Rayleigh distribution matches the two-parameter *Burr type-X* (BX) model;
[2] The Fréchet distribution matches the inverse Weibull model;
[3] The Wald distribution matches the inverse normal model.

where

$$n = \begin{cases} \beta, & \text{if } \beta \in \mathbb{Z}_+ \\ \infty & \text{if } \beta \in \mathbb{R}_+ - \mathbb{Z}_+ \end{cases}.$$

The expansion (2.13) can be used to obtain the expected value, variance, kurtosis and asymmetry of the random variable X, but these quantities may not have closed-form expressions or even do not exist for some baselines models.

The *quantile function* (qf) of a random variable X following the sf (2.12) is given by

$$Q(p) = u^{-1}(p)\left[-\ln(1-p^{1/\beta})\right],$$

where $u^{-1}(p)$ is the inverse function of (2.11).

A useful review of the Exp-G distributions was done by Al-Hussaini and Ahsanullah (2015). This technique has triggered several models and some examples are the given in Table 2.1. Gompertz (1825) and Verhulst (1838) pioneered this method. Ahuja and Nash (1967) deduced the cumulants and moment generating functions of the Gompertz and Verhulst distributions, some asymptotic properties were also studied.

2.2.1 Some possible applications of the ECR model

The theoretical mechanism of extension that we used can be understood as arisen from one between two contexts. First, assuming $\beta_0 \in \mathbb{Z}_+$, consider a device works based on β_0 independent components in a parallel system. Assume that the device fails if all β_0

components fail. Let $X_1,\ldots,X_{\beta_0}$ denote the $X_i \sim \text{CR}(\lambda)$ for $i = 1,\ldots,\beta_0$ and let X be the device lifetime. Setting the event *to fail* as equivalent to $[X \leq x]$, the probability of device fails is

$$\Pr(X \leq x) = \Pr\left(\bigcap_{i=1}^{\beta_0} X_i \leq x\right) = [\Pr(X_1 \leq x)]^{\beta_0}.$$

So the device lifetime obeys the ECR(β_0,λ) distribution. We can rewrite this motivation considering X as the largest order statistic in a CR(λ) random sample of size β_0. Note that

$$\Pr(X_{(\beta_0)} = \max(X_1,\ldots,X_{\beta_0}) \leq x) = \Pr\left(\bigcap_{i=1}^{\beta_0} X_i \leq x\right) = [\Pr(X_1 \leq x)]^{\beta_0}.$$

Then we have $X_{(\beta_0)} \sim$ ECR(β_0,λ). Thus, from this approach, X may represent the maximum lifetimes among β_0 lifetime components which are structured in parallel system and represent a random sample from $Y \sim$ CR(λ).

In SAR imagery the observed intensity is the sum of the contributions of the different scatterers at an under-study surface (DELIGNON; PIECZYNSKI, 2002). In the classical approach, the number of scatterers in each elementary cell is assumed to be large enough and approximately constant. The reflect electrical field is then normal; the observed intensity admits an exponential distribution, and the amplitude admits a Rayleigh model. However, when the number of scatterers varies, the resulting field distribution may be non-normal, and so the observed intensity may no longer be exponential. Let $X_1,\ldots,X_{\beta_0}$ be a sample of scatterer intensities in an elementary cell, where β_0 is a Poisson random variable. The intensity of the back-scattered field is the sum of the scatterer intensities in an elementary cell $X = \sum_{i=1}^{\beta_0} X_i$. The distribution of X is hard to deduce when $X_i \sim$ CR(λ), but as seen before $X_{(\beta_0)} \sim$ ECR(β_0,λ). Then we can study the intensity of the back-scattered field by means of the maximum intensity of its elementary cells.

A second motivation is based on F^α *record* (F^αR) models. Arnold *et al.* (1998, p. 187) introduced the F^αR model as an infinite sequence $\{Y_n,n \geq 1\}$ of independent random variables, where the cdf of Y_n has the form $F_n(y) = [F(y)]^{\alpha_n}$, $\alpha_n > 0$, and the associated record statistics, the cdf is assumed to be continuous and the α_n's are not necessarily integer. The F^αR models have been studied by Yang (1975), Nevzorov (1986), Ballerini and Resnick (1987), Nevzorov (1990), Nagaraja (1994) and Hofmann and Nagaraja (2000). If we assume $\alpha_n = 1$ for all n, we obtain the classical record model (ARNOLD *et al.*, 1998, p. 7). Yang (1975) assumed α_n's to be geometrically increasing, $\alpha_n = \alpha_1\gamma^{n-1}$ and $\gamma > 1$, and proposed a model in which the increasing frequency of record breaks in the Olympic games are attributed to the population size increase. Verhulst (1847), as mentioned by Al-Hussaini and Ahsanullah (2015, p. 2), used exponentiated distributions to represent population growth. Recently, the record models were used to

support CS-MLEs (TEIMOURI; NADARAJAH, 2013).

2.3 COX-SNELL CORRECTED MAXIMUM LIKELIHOOD ESTIMATORS

Let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$ and $\hat{\theta}_i$, for $i = 1, \ldots, p$, be the MLE for θ_i obtained by maximizing the *log-likelihood function* (llf):

$$\ell(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta}|\mathbf{x}) = \sum_{i=1}^{n} \log[f_X(x_i|\boldsymbol{\theta})].$$

The elements *Fisher information matrix* (FIM) is denoted by $\mathbf{K}_{i,j} = [\mathbf{K}(\boldsymbol{\theta})]_{i,j} = -\kappa_{\theta_i\theta_j}$ and the associated cumulants obey the notation:

$$U_{\theta_1\cdots\theta_n} = \frac{\partial \ell(\boldsymbol{\theta})}{\partial\theta_1 \cdots \partial\theta_n} \text{ and } \kappa_{\theta_1\cdots\theta_n} = \mathbb{E}\left(U_{\theta_1\cdots\theta_n}\right).$$

We will assume the notation $\kappa_{\theta_i\theta_j}^{(\theta_k)} = \frac{\partial}{\partial\theta_k}\kappa_{\theta_i\theta_j}$ for the derivatives of FIM components and $\mathbf{K}^{-1}{}_{i,j} = [\mathbf{K}^{-1}(\boldsymbol{\theta})]_{i,j} = -\kappa^{\theta_i,\theta_j}$ for the inverse FIM components.

Cox and Snell (1968) proposed a general formula for the second order bias of the MLE of the vector $\boldsymbol{\theta}$ in terms of the cumulants. This procedure was revisited by Cordeiro and Klein (1994), which established that the bias of $\hat{\theta}_i$ assumes the following formulation:

$$\text{Bias}(\hat{\theta}_i) = \mathbb{E}(\hat{\theta}_i) - \theta_i = \sum_{r,s,t} \kappa^{\theta_i,\theta_r}\kappa^{\theta_s,\theta_t}\left(\kappa_{\theta_r\theta_s}^{(\theta_t)} - \frac{1}{2}\kappa_{\theta_r\theta_s\theta_t}\right) + \sum_{r=2}^{\infty} O(n^{-r}). \tag{2.14}$$

Thus, the CS-MLE, $\widetilde{\theta}_i$, is defined by

$$\widetilde{\theta}_i = \hat{\theta}_i - \widehat{\text{Bias}}(\hat{\theta}_i), \tag{2.15}$$

where $\widehat{\text{Bias}}(\hat{\theta}_i)$ represents the second order bias of $\hat{\theta}_i$ evaluated in $\hat{\theta}_i$. Cox and Snell (1968) showed that

1. $\mathbb{E}\left[\widehat{\text{Bias}}(\hat{\theta}_i)\right] = O(n^{-2})$;
2. $\mathbb{E}(\hat{\theta}_i) = O(n^{-1})$;
3. $\mathbb{E}(\widetilde{\theta}_i) = \theta_i + O(n^{-2})$.

Therefore, the bias of $\widetilde{\theta}_i$ has order of n^{-2}. Thus, the CS-MLEs are expected to have better asymptotic features than MLEs. Some recent works using this method are seen in Table 2.2.

Table 2.2: Recent CS-MLE works.

Distribution	Works
Beta	Cribari-Neto and Vasconcellos (2002)
Negative Binomial	Saha and Paul (2005)
Birnbaum-Saunders	Lemonte *et al.* (2007)
Gamma	Giles and Feng (2009) and Singh *et al.* (2015)
Generalized Logistic	Lagos-Álvarez *et al.* (2011)
Kumaraswamy	Lemonte (2011)
Half-logistic	Giles (2012)
Lomax	Aguilar (2012) and Giles *et al.* (2013)
Nakagami	Schwartz *et al.* (2013)
Weibull	Teimouri and Nadarajah (2013)
Wishart complex	Nascimento *et al.* (2014a)
Generalized Rayleigh	Ling and Giles (2014)
Doubly-truncated Poisson	Godwin (2016)
Zero-inflated Poisson	Schwartz and Giles (2016)
Log-logistic	Reath (2016)
Weighted Lindley	Wang and Wang (2017)
Skew normal	Zhang and Liu (2017)
Fréchet	Mazucheli *et al.* (2018a)
Unit-gamma	Mazucheli *et al.* (2018b)

2.4 INFORMATION THEORY MEASURES

Entropy can be defined as the uncertainty of a single random variable (COVER; THOMAS, 2006, p. 6). From a historical perspective, the first entropy measure was proposed by Shannon (1948). Subsequently, many entropy measures have been introduced (PARDO, 2006, p. 6). In order to study the different entropy measures at a systematic way, Burbea and Rao (1982a), Burbea and Rao (1982b), Burbea and Rao (1982c) and Salicrú *et al.* (1993) considered the (h,ϕ)-entropies class by means of Definition 2.1.

Definition 2.1 (ϕ- and (h,ϕ)-entropies)**.** *An (h,ϕ)-entropy is any entropy measure in the form*

$$\mathcal{H}_{\phi}^{h}(X) \equiv \mathcal{H}_{\phi}^{h}(\Pr) \equiv \mathcal{H}_{\phi}^{h}(\boldsymbol{\theta}) = h\left(\int_{\Omega} \phi\left(f_{\boldsymbol{\theta}}(x)\right) d\mu(x)\right),$$

where $h : \mathbb{R} \to \mathbb{R}$, $\phi : (0,\infty) \to \mathbb{R}$, h is differentiable and either ϕ is concave and h is increasing and ϕ is convex and h is decreasing. If ϕ is a continuous concave function and $h(x) = x$ then this entropy is an ϕ-entropy.

Table 2.3 shows some elements of the class introduced by Definition 2.1. Note that Shannon and Tsallis entropies assumes h as the identify function, then they can be classified as ϕ-entropies while Rényi and Arimoto entropies have more elaborated h and ϕ functions.

Table 2.3: ϕ- and (h,ϕ)-entropies ($q > 0, q \neq 1$).

$\phi(x)$	$h(x)$	Entropy
$-x \log x$	x	Shannon (1948)
$\frac{x^q - x}{1-q}$	x	Tsallis (1988)
x^q	$\frac{\log x}{1-q}$	Rényi (1961)
$x^{1/q}$	$\frac{1-x^q}{1-q}$	Arimoto (1971)

A large value of the entropy indicates the greater uncertainty in the data. In terms of interpretability, a large value of the entropy indicates the greater uncertainty in the data. Deriving closed-form entropy measures is sought in various applications. Mahmoud and El-Ghafour (2013) deduced the Shannon entropy for the *generalized Feller-Pareto* (GFP) distribution and Gomes *et al.* (2015) obtained the Rényi and Shannon entropies for *McDonald Burr type-XII* (McBXII) model, but these results do not have closed-form expressions. In the following lines we show some results involving the ECR-based entropies due to Shannon (1948), Rényi (1961), Tsallis (1988) and Arimoto (1971).

The Shannon entropy provides an absolute limit on the best possible average length of lossless encoding of an information source and is closely related to the Boltzmann constant and Gini entropy (DOUGHERTY, 1964). The Rényi and Shannon entropies are important in ecology and statistics as indexes of diversity (HILL, 1973; JOST, 2006). As pointed by Rathie and Silva (2008), the Tsallis entropy exactly matches the Havrda and Charvât entropy (HAVRDA; CHARVÂT, 1967) introduced aiming to form a quantificatory theory of classificatory processes. Tsallis (1988) used this entropy as a basis for generalizing the standard statistical mechanics. Finally, the Arimoto entropy, as defined in Pardo (2006, p. 20), is another generalization of the Shannon entropy. Arimoto (1971) proposed it to investigate the finite-parameter estimation problem treating the error-probability of decision and the other equivocation measures.

The class of q-entropies associated to a probability model with pdf $f(x)$ can be defined by means of the following integral:

$$\boldsymbol{I}(q,\boldsymbol{\theta}) = \int_0^\infty [f(x)]^q \, \mathrm{d}x, \quad q \neq 1. \tag{2.16}$$

We named it q-integral. Once deduced the q-integral, Proposition 2.1 can be used to obtain its associated q-entropies under any model. This result can also be used to define Rényi, Tsallis and Arimoto q-entropies.

Proposition 2.1 (q-entropies)**.** *For a given q-integral $\boldsymbol{I}(q,\boldsymbol{\theta})$ the correspondent Rényi, Tsallis and Arimoto q-entropies can be expressed, respectively, by*

$$\mathcal{H}_R(q,\boldsymbol{\theta}) = \frac{1}{1-q} \log \boldsymbol{I}(q,\boldsymbol{\theta}),$$

$$\mathcal{H}_T(q,\boldsymbol{\theta}) = \frac{1}{1-q}\left[\boldsymbol{I}(q,\boldsymbol{\theta}) - 1\right]$$

and

$$\mathcal{H}_A(q,\boldsymbol{\theta}) = \frac{q}{1-q}\left\{\left[\boldsymbol{I}(q,\boldsymbol{\theta})\right]^{1/q} - 1\right\}.$$

The q-entropies gradients relative to distributions parameters are fundamental to implement EBTs. These gradients can be easily obtained by means of the q-integral (2.16) and Proposition 2.2.

Proposition 2.2. *The ECR Rényi, Tsallis and Arimoto q-entropies gradients relative to the parametric vector* $\boldsymbol{\theta}$ *are*

$$\nabla\mathcal{H}_R(q,\boldsymbol{\theta}) = \frac{1}{(1-q)\boldsymbol{I}(q,\boldsymbol{\theta})}\nabla\boldsymbol{I}(q,\boldsymbol{\theta}),$$
$$\nabla\mathcal{H}_T(q,\boldsymbol{\theta}) = \frac{1}{1-q}\nabla\boldsymbol{I}(q,\boldsymbol{\theta})$$

and

$$\nabla\mathcal{H}_A(q,\boldsymbol{\theta}) = \frac{\left[\boldsymbol{I}(q,\boldsymbol{\theta})\right]^{(1-q)/q}}{1-q}\nabla\boldsymbol{I}(q,\boldsymbol{\theta}).$$

A notable relation between all (h,ϕ-entropies are given by

$$\lim_{q\to 1}\mathcal{H}_R(q,\boldsymbol{\theta}) = \lim_{q\to 1}\mathcal{H}_T(q,\boldsymbol{\theta}) = \lim_{q\to 1}\mathcal{H}_A(q,\boldsymbol{\theta}) = \mathcal{H}_S(\boldsymbol{\theta}). \tag{2.17}$$

Another useful concept is the cross entropy, which can be used to deduce the *Kullback-Leibler* (KL) divergence and several others divergence measures. The cross Shannon entropy is a measure of the average number of bits needed to identify an event from a set of possibilities between two probability distributions. The cross Shannon entropy of a continuous random variable X_1 relative to a continuous random variable X_2, with pdfs given, respectively, by $f_1(x)$ and $f_2(x)$, is given by

$$\mathcal{H}_S^c(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = -\int_{-\infty}^{\infty} f_1(x)\log f_2(x)\,\mathrm{d}x. \tag{2.18}$$

2.5 SOME TWO-SAMPLE HYPOTHESIS TESTS

Let $(\Omega,\mathcal{B},\Pr)_{\boldsymbol{\theta}\in\Theta}$ be the statistical space associated with the random variable X, where $\mathcal{B}$ is the σ-field of the Borel subsets $A\subset\Omega$ and $\{\Pr\}_{\boldsymbol{\theta}\in\Theta}$ is a family of probability distributions on the measurable space $(\Omega,\mathcal{B})$ with Θ an open subset of $\mathbb{R}^n$, $n\geq 1$. We assume that the probability distributions Pr are absolutely continuous with respect to a σ-finite measure μ on $(\Omega,\mathcal{B})$. For simplicity μ is either the Lebesque measure or a counting measure. On the basis of two independent random samples $\boldsymbol{X}_1 = X_{11},\ldots,X_{1n_1}$ and $\boldsymbol{X}_2 = X_{21},\ldots,X_{2n_2}$ of sizes n_1 and n_2, respectively, from two populations having

distribution belonging to

$$\mathcal{F} = \{f(x;\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p\} \tag{2.19}$$

a set of parametric model densities, with parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, we can solve problems of image pre- and post-processing by testing hypotheses of kind

$$H_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 \text{ versus } H_1 : \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2. \tag{2.20}$$

To carry out (2.20), we recur to the following test classes: LRTs and EBTs. A LRT consists the ratio of the likelihood at the hypothesized parameter values to the likelihood of the data at the maximum. A short review of the LRT theory can be seen in Bickel and Doksum (2000, p. 255). On the other hand, an EBT is settled on weighted difference of entropy measures at the hypothesized parameter values. A complete reference of this class of test can be obtained in the work of Pardo *et al.* (1997).

2.5.1 Likelihood ratio test

This test statistic is the ratio of the likelihood at the hypothesized parameter values to the likelihood of the data at the maximum. The joint *likelihood function* (lf) of the pooled sample $\boldsymbol{X} = (\boldsymbol{X}_1,\boldsymbol{X}_2)$ can be written as

$$\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{x}) = \mathcal{L}(\boldsymbol{\theta}_1|\boldsymbol{x}_1)\mathcal{L}(\boldsymbol{\theta}_2|\boldsymbol{x}_2),$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) \in \Theta \times \Theta$. To test $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$ versus $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$, the LRT statistic is

$$\begin{aligned} \boldsymbol{\Lambda}(\boldsymbol{x}) &= \frac{\sup\{\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{x}) : \boldsymbol{\theta} \in \Theta_1\}}{\sup\{\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{x}) : \boldsymbol{\theta} \in \Theta_0\}} \\ &= \frac{\mathcal{L}(\hat{\boldsymbol{\theta}}_1|\boldsymbol{x}_1)\mathcal{L}(\hat{\boldsymbol{\theta}}_2|\boldsymbol{x}_2)}{\mathcal{L}(\hat{\boldsymbol{\theta}}_0|\boldsymbol{x})}, \end{aligned} \tag{2.21}$$

where $\Theta_0 = \{\boldsymbol{\theta} : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2\}$, $\Theta_1 = \{\boldsymbol{\theta} : \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2\}$, $\hat{\boldsymbol{\theta}}_0$ is the constrained MLE, based on the pooled sample and $\hat{\boldsymbol{\theta}}_1$ and $\hat{\boldsymbol{\theta}}_2$ are the unconstrained ones, based on each given sample. Wilks (1938) proved that

$$2\log\boldsymbol{\Lambda}(X) \xrightarrow[n\to\infty]{d} \chi^2_k,$$

where k is the difference between the number of free parameters specified by $\boldsymbol{\theta} \in \Theta_1$ and the number of free parameters specified by $\boldsymbol{\theta} \in \Theta_0$.

2.5.2 Entropy-based tests

Pardo *et al.* (1997) used (h,ϕ)-entropy statistics to construct confidence intervals and statistical hypotheses tests. Let $\hat{\boldsymbol{\theta}}$ be the MLE of $\boldsymbol{\theta}$. Suppose that the standard

regularity conditions assumptions as described in Pardo (2006, p. 58) hold. In these conditions Pardo *et al.* (1997) deduced an (h,ϕ)-EBT for the equality of r populational entropies. As given in Pardo *et al.* (1997, p. 498) to test the hypotheses that the (h,ϕ)-entropy for r populations with pdfs $f(x,\boldsymbol{\theta}_1),\ldots,f(x,\boldsymbol{\theta}_r)$ and sizes $n_1,\ldots,n_r$ is the same, that is,

$$H_0 : \mathcal{H}_h^\phi(\boldsymbol{\theta}_1) = \cdots = \mathcal{H}_h^\phi(\boldsymbol{\theta}_r)$$

we can use the statistic

$$\mathcal{Z}_r(\boldsymbol{x}) = \sum_{j=1}^{r} \frac{n_j \left[\mathcal{H}_h^\phi(\hat{\boldsymbol{\theta}}_j) - \overline{\mathcal{H}}\right]^2}{\hat{\sigma}_j^2},$$

which has approximately a chi-square distribution with $r-1$ degrees of freedom under H_0 where

$$\overline{\mathcal{H}} = \frac{\displaystyle\sum_{j=1}^{r} \frac{n_j \mathcal{H}_h^\phi(\hat{\boldsymbol{\theta}}_j)}{\hat{\sigma}_j^2}}{\displaystyle\sum_{j=1}^{r} \frac{n_j}{\hat{\sigma}_j^2}},$$

$$\hat{\sigma}_i^2 = \mathbf{T}^\top \mathbf{K}^{-1} \mathbf{T}|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_i}, \quad (i = 1,\ldots,r), \tag{2.22}$$

$$\mathbf{T} = \nabla \mathcal{H}_h^\phi(\boldsymbol{\theta}) = \left(\frac{\partial \mathcal{H}_h^\phi(\boldsymbol{\theta})}{\partial \theta_1}, \ldots, \frac{\partial \mathcal{H}_h^\phi(\boldsymbol{\theta})}{\partial \theta_n} \right)^\top, \tag{2.23}$$

$$\boldsymbol{\theta} = (\theta_1,\ldots,\theta_n).$$

and $\mathbf{K}^{-1}$ is the inverse FIM. It is possible to reduce with some algebra when $r = 2$ getting

$$\mathcal{Z}(\boldsymbol{x}) = \frac{\left[\mathcal{H}_h^\phi(\hat{\boldsymbol{\theta}}_1) - \mathcal{H}_h^\phi(\hat{\boldsymbol{\theta}}_2)\right]^2}{\dfrac{\hat{\sigma}_1^2}{n_1} + \dfrac{\hat{\sigma}_2^2}{n_2}}. \tag{2.24}$$

A nominal level α EBT rejects H_0 in (2.20) if

$$\mathcal{Z}(\boldsymbol{x}) > c_{1,1-\alpha},$$

where $c_{1,1-\alpha}$ is the $1-\alpha$ quantile of χ_1^2. Pardo (2006, p. 70) noticed that in situations where the (h,ϕ)-entropy is one-to-one function of $\boldsymbol{\theta}$ the test (2.20) is equivalent to the test

$$H_0 : \mathcal{H}_h^\phi(\boldsymbol{\theta}_1) = \mathcal{H}_h^\phi(\boldsymbol{\theta}_2) \text{ versus } H_1 : \mathcal{H}_h^\phi(\boldsymbol{\theta}_1) \neq \mathcal{H}_h^\phi(\boldsymbol{\theta}_2). \tag{2.25}$$

2.6 SAR IMAGERY

Following the notes of Oliver and Quegan (2004) the operation of SAR consists of sending electromagnetic pulses towards a target and analyzing the returning echo, whose intensity depends on the physical properties of the target surface. Noise is inherent to image acquisition. An important source of noise when coherent illumination is used is due to the interference of the signal backscattering by the elements of the target surface. The resulting effect is called speckle noise. Kuruoglu and Zerubia (2004) argues that the presence of speckle noise degrade SAR image significantly and may hide some important details of the image, leading to the loss of crucial information. Thus, working with SAR imagery requires specialized modeling and post-processing techniques.

Several assumptions have been made for developing statistical models for speckled data. Kuruoglu and Zerubia (2004) listed the following procedures: 1. the scatters are statistically independent, 2. the number of scatterers is large, 3. the scattering amplitude and the instantaneous phase are independent random variables, 4. the phase is uniformly distributed over the range $[0,2\pi]$, 5. the individual scatterer dominates the whole scene, 6. the reflection surface is large when compared to the size of individual reflectors.

Shapiro and Stockman (2001, p. 237) present two approaches for texture analysis: structural and statistical. This work is focused on statistical approach which consider texture as quantitative measure of the arrangement of intensities in a region. We opted to use the (h,ϕ)-entropies and likelihoods under the ECR class as texture measures.

The SAR system illuminates the earth landscape and receives the scattered wave reflected by the earth surface represented by the scattering matrix (LEE; POTTIER, 2009)

$$\begin{bmatrix} S_{HH} & S_{HV} \\ S_{VH} & S_{VV} \end{bmatrix}.$$

The scattering matrix gives the backscattered response of earth landscape to an emitted electromagnetic wave. This matrix is usually given in a linear polarization basis (H: horizontal and V: vertical), defining the set of polarization channels $\mathfrak{C}_{(H,V)}$:

$$\begin{aligned} \mathfrak{C}_{(H,V)} &= \{H,V\} \times \{H,V\} \\ &= \{HH, HV, VH, VV\}. \end{aligned}$$

Due to the reciprocity theorem (ULABY; ELACHI, 1990), the two cross-polarization scattering coefficients are equal, thus it is possible to redefine $\mathfrak{C}_{(H,V)}$ as

$$\mathfrak{C}_{(H,V)} = \{HH, HV, VV\}. \tag{2.26}$$

2.7 THE FINITE MIXTURE MODELS

The importance of FM models is supported by the increasing rate at which articles on mixture applications appear in the literature. McLachlan and Peel (2000, p. xix) points that, because of their flexibility, FM models have been increasing exploited as a convenient, semiparametric way in which to model unknown distributional shapes. Some important references for FM models are Everitt and Hand (1981), Titterington *et al.* (1985), Redner and Walker (1984), McLachlan and Basford (1987), McLachlan and Peel (2000) and McLachlan and Krishnan (2008). The book of Cornell (2002) brings a large number of applications, experiments and techniques based on this class of models.

According to McLachlan and Peel (2000, ch. 1), let $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n$ denote a random sample of size n, where $\boldsymbol{X}_j$ is a p-dimensional continuous random vector with pdf $f(\boldsymbol{x}_j)$ on $\mathbb{R}^p$. A realization of a random vector $\boldsymbol{X} = \left(\boldsymbol{X}_1^\top,\ldots,\boldsymbol{X}_n^\top\right)^\top$ is denoted by

$$\boldsymbol{x} = \left(\boldsymbol{x}_1^\top,\ldots,\boldsymbol{x}_n^\top\right)^\top, \tag{2.27}$$

where $\boldsymbol{x}_j$ is the observed value of the random vector $\boldsymbol{X}_j$ and is called the *feature vector*.

The *mixture density* $f(\boldsymbol{x}_j)$ is given by

$$f(\boldsymbol{x}_j;\boldsymbol{\Psi}) = \sum_{i=1}^{g} \pi_i f(\boldsymbol{x}_j;\boldsymbol{\theta}_i), \tag{2.28}$$

where the vector $\boldsymbol{\Psi}$ containing all the unknown parameters in the FM can be written as

$$\boldsymbol{\Psi} = (\pi_1\ldots,\pi_{g-1},\boldsymbol{\xi}^\top)^\top, \tag{2.29}$$

where $\boldsymbol{\xi}$ is the vector containing all the parameters in $\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_g$ known *a priori* to be distinct, the $f(\boldsymbol{x}_j;\boldsymbol{\theta}_i)$ denotes a generic member of the parametric family

$$\{f(\boldsymbol{x}_j;\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\},$$

the *mixing proportions* π_i are non-negative quantities such that

$$\sum_{i=1}^{g} \pi_i = 1. \tag{2.30}$$

From (2.30), π_g is redundant ($\pi_g = 1 - \sum_{i=1}^{g-1} \pi_i$), then we have arbitrarily omitted the g^{th} mixing proportion π_g in defining $\boldsymbol{\Psi}$ in (2.29). The pdfs $f(\boldsymbol{x}_j;\boldsymbol{\theta}_1),\ldots,f(\boldsymbol{x}_j;\boldsymbol{\theta}_g)$ are called the *component mixtures*. The expression (2.28) defines the pdf of a FM with g components.

The llf for $\boldsymbol{\Psi}$ associated to the FM (2.28) is given by

$$\ell(\boldsymbol{\Psi}) = \sum_{i=1}^{g}\sum_{j=1}^{n}\{\log \pi_i + \log f(\boldsymbol{x}_j;\boldsymbol{\theta}_i)\}\,.$$

The FMs can be tackled from parametric or non-parametric statistical points of view. Jordan and Xu (1995) regard (2.28) as non-parametric by allowing the number of components g to grow and as parametric when g is fixed.

For FM models, the likelihood equations are almost certain to be nonlinear and beyond hope of solution by analytic means and, as a consequence, one must resort to seeking an approximate solution via some iterative procedure (REDNER; WALKER, 1984, p. 203). The first publication relating to the FMs estimation problem appears to be that of Pearson (1894). The standard procedure to solve this class of problems is known as EM algorithm whose terminology is due to Dempster *et al.* (1977).

2.8 THE EM ALGORITHM AND INCOMPLETE-DATA STRUCTURE

The first papers which investigated algorithms to obtain the FM ML estimates were Wolfe (1965), Wolfe (1967), Day (1969) and Wolfe (1970). The problem of maximize directly the FM lf is typically a hard task. The EM algorithm is a more efficiency technique to obtain the ML estimates. To fit the normal FMs using MLEs, we can use the recursive formulas due to Dempster *et al.* (1977), which assumes the data can be viewed as being incomplete. This approach was named EM algorithm and simplifies considerably the estimation process introducing a component-label vector, but a number of issues like multiple local maximum remain.

Rewriting McLachlan and Peel (2000, ch. 1-2), let $\boldsymbol{Z}_j$ be a g-dimensional *component-label vector*, where its i^{th} element, $Z_{ij} = (\boldsymbol{Z}_j)_i$, is defined to be one or zero, according to whether the component of origin of $\boldsymbol{X}_j$ in the FM (2.28) is equal to i or not ($i = 1,\ldots,g$). Thus $\boldsymbol{Z}_j$ follows a multinomial distribution consisting of one draw on g categories with probabilities $\pi_1,\ldots,\pi_g$; that is,

$$\Pr(\boldsymbol{Z}_j = \boldsymbol{z}_j) = \pi_1^{z_{1j}}\pi_2^{z_{2j}}\cdots\pi_g^{z_{gj}}.$$

We write $\boldsymbol{Z}_j \sim \text{Mult}_g(1,\boldsymbol{\pi})$ and assume that they are distributed unconditionally as

$$\boldsymbol{Z}_1,\ldots,\boldsymbol{Z}_n \overset{iid}{\sim} \text{Mult}_g(1,\boldsymbol{\pi}), \tag{2.31}$$

where $\boldsymbol{\pi} = (\pi_1,\ldots,\pi_g)^\top$ satisfying (2.30).

In the EM framework, the feature data $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n$ are viewed as being incomplete since their associated component-label vectors, $\boldsymbol{z}_1,\ldots,\boldsymbol{z}_n$, are not available. The complete-

data vector is expressed by

$$\boldsymbol{x}_c = (\boldsymbol{x}^\top, \boldsymbol{z}^\top)^\top,$$

where $\boldsymbol{x}$ is the observed-data (incomplete-data vector) given by (2.27) and where

$$\boldsymbol{z} = \left(\boldsymbol{z}_1^\top, \ldots, \boldsymbol{z}_n^\top\right)^\top$$

is the unobservable vector of component-indicator variables taken to be the realized values of the random vectors $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$ distributed according (2.31).

The complete-data vector $\boldsymbol{x}_c$ turns the FM estimation into a more straightforward routine than on the basis of observed data $\boldsymbol{x}$, assuming each component density $f_i(\boldsymbol{x})$ be estimated directly from those feature data $\boldsymbol{x}_j$ with $z_{ij} = (\boldsymbol{z}_j)_i = 1$. This assumption means that the distribution of the complete-data vector $\boldsymbol{X}_c$ implies the appropriate distribution for the incomplete-data vector $\boldsymbol{Y}$. The complete-data llf for $\boldsymbol{\Psi}$, $\ell_c(\boldsymbol{\Psi})$, is given by

$$\ell_c(\boldsymbol{\Psi}) = \sum_{i=1}^{g} \sum_{j=1}^{n} z_{ij} \left\{\log \pi_i + \log f_i(\boldsymbol{x}_j; \boldsymbol{\theta}_i)\right\}.$$

Still perusing McLachlan and Peel (2000, ch. 2-3), the EM algorithm treats $\boldsymbol{z}_{ij}$ as missing data proceeding iteratively in two steps named E-step and M-step. These two steps are alternated iteratively until

$$\ell(\boldsymbol{\Psi}^{(k+1)}) - \ell(\boldsymbol{\Psi}^k) < \epsilon,$$

for a given $\epsilon > 0$. Dempster *et al.* (1977) showed that the (incomplete-data) llf value is not decreased after an EM iteration; that is, for $k \in \mathbb{Z}^+$

$$\ell(\boldsymbol{\Psi}^{(k+1)}) \geq \ell(\boldsymbol{\Psi}^k).$$

2.8.1 E-step

Let $\boldsymbol{\Psi}^{(0)}$ be the value specified initially for $\boldsymbol{\Psi}$. Then on the first iteration of the EM algorithm, the E-step requires the computation of the conditional expectation of $\ell_c(\boldsymbol{\Psi})$ given $\boldsymbol{x}$, using $\boldsymbol{\Psi}^{(0)}$ for $\boldsymbol{\Psi}$, which can be written as

$$Q\left(\boldsymbol{\Psi}, \boldsymbol{\Psi}^{(0)}\right) = \mathbb{E}_{\boldsymbol{\Psi}^{(0)}}\left[\ell_c(\boldsymbol{\Psi}) | \boldsymbol{x}\right].$$

It follows that on the $(k+1)^{th}$ iteration in the E-step requires the calculation $Q\left(\boldsymbol{\Psi}, \boldsymbol{\Psi}^{(k)}\right)$, where $\boldsymbol{\Psi}^{(k)}$, is the value of $\boldsymbol{\Psi}$ the k^{th} EM iteration. Note that, for $i = 1, \ldots, g$

and $j = 1, \ldots, n$,

$$\begin{aligned} \mathbb{E}_{\boldsymbol{\Psi}^{(k)}}(Z_{ij}|\boldsymbol{x}) &= \Pr_{\boldsymbol{\Psi}^{(k)}}(Z_{ij} = 1|\boldsymbol{x}) \\ &= \tau_i\left(\boldsymbol{x}_j; \boldsymbol{\Psi}^{(k)}\right), \end{aligned} \tag{2.32}$$

where (2.32) corresponds to the posterior probability that the j^{th} member of the sample with observed value $\boldsymbol{y}_i$ belongs to the i^{th} component of the mixture. It is possible to express

$$\begin{aligned} \tau_i\left(\boldsymbol{x}_j; \boldsymbol{\Psi}^{(k)}\right) &= \frac{\pi_i^{(k)} f\left(\boldsymbol{y}_j; \boldsymbol{\theta}_i^{(k)}\right)}{f\left(\boldsymbol{y}_j; \boldsymbol{\Psi}^{(k)}\right)} \\ &= \frac{\pi_i^{(k)} f\left(\boldsymbol{y}_j; \boldsymbol{\theta}_i^{(k)}\right)}{\sum_{h=1}^{g} \pi_h^{(k)} f\left(\boldsymbol{y}_j; \boldsymbol{\theta}_h^{(k)}\right)}. \end{aligned} \tag{2.33}$$

By (2.32), it is possible to express

$$Q\left(\boldsymbol{\Psi}, \boldsymbol{\Psi}^{(k)}\right) = \sum_{i=1}^{g}\sum_{j=1}^{n} \tau_i(\boldsymbol{x}_j; \boldsymbol{\Psi}^{(k)}) \left\{\log \pi_i + \log f\left(\boldsymbol{x}_j; \boldsymbol{\theta}_i\right)\right\}.$$

2.8.2 M-step

If the z_{ij} were observable, then the complete-data MLE of π_i would be given simply by

$$\hat{\pi}_i = \frac{1}{n}\sum_{j=1}^{n} z_{ij} \quad (i = 1, \ldots, g). \tag{2.34}$$

In the complete-data llf, the updated estimate of π_i is given by replacing each z_{ij} in (2.34) by

$$\pi_i^{(k+1)} = \frac{1}{n}\sum_{j=1}^{n} \tau_i\left(\boldsymbol{x}_j; \boldsymbol{\Psi}^{(k)}\right) \quad (i = 1, \ldots, g). \tag{2.35}$$

The M-step on the $(k+1)^{th}$ iteration updates the estimate $\boldsymbol{\xi}^{(k+1)}$ maximizing globally $Q\left(\boldsymbol{\Psi}, \boldsymbol{\Psi}^{(k)}\right)$ with respect to $\boldsymbol{\Psi}$ over the entire parameter space. By (2.34), the $\boldsymbol{\xi}^{(k+1)}$ coincides as one root of

$$\sum_{i=1}^{g}\sum_{j=1}^{n} \tau_i(\boldsymbol{x}_j; \boldsymbol{\Psi}^{(k)}) \frac{\partial \log f\left(\boldsymbol{x}_j; \boldsymbol{\theta}_i\right)}{\partial \boldsymbol{\xi}} = \mathbf{0}.$$

Typically, it is not available explicit formulas for FM estimates.

2.8.3 Multivariate normal finite mixture

We now specialize these results to the case of a FM with normal components. If the mixture components in (2.28) follow multivariate normal distributions with pdf

$$f(\boldsymbol{x}_j;\boldsymbol{\theta}_i) = \phi\left(\boldsymbol{x}_j;\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i\right),$$

then a g-component multivariate normal FM distribution can be written as

$$f(\boldsymbol{x}_j;\boldsymbol{\Psi}) = \sum_{i=1}^{g} \pi_i \phi\left(\boldsymbol{x}_j;\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i\right), \tag{2.36}$$

where $\phi\left(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma}\right)$ is the multivariate normal random variable pdf with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ given by

$$\phi\left(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma}\right) = \frac{\exp\left[-\frac{1}{2}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)^{\top}\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)\right]}{\sqrt{(2\pi)^p|\boldsymbol{\Sigma}|}},$$

where $\boldsymbol{\xi}$ in (2.29) contains the elements of the component means $\boldsymbol{\mu}_i$ and covariance matrices $\boldsymbol{\Sigma}_i$ $(i = 1,\ldots,g)$.

For multivariate normal FMs the posterior probability given in (2.33) can be rewritten as

$$\tau_i\left(\boldsymbol{x}_j;\boldsymbol{\Psi}^{(k)}\right) = \frac{\pi_i^{(k)}\phi\left(\boldsymbol{x}_j;\boldsymbol{\mu}_i^{(k)},\boldsymbol{\Sigma}_i^{(k)}\right)}{\sum_{h=1}^{g}\pi_h^{(k)}\phi\left(\boldsymbol{x}_j;\boldsymbol{\mu}_h^{(k)},\boldsymbol{\Sigma}_h^{(k)}\right)}$$

for $i = 1,\ldots,g$ and $j = 1,\ldots,n$.

The M-step for normal components exists in closed-form and is given by

$$\boldsymbol{\mu}_i^{(k+1)} = \frac{\sum_{j=1}^{n}\tau_{ij}^{(k)}\boldsymbol{x}_j}{\sum_{j=1}^{n}\tau_{ij}^{(k)}} \tag{2.37}$$

and

$$\boldsymbol{\Sigma}_i^{(k+1)} = \frac{\sum_{j=1}^{n}\tau_{ij}^{(k)}\left(\boldsymbol{x}_j-\boldsymbol{\mu}_i^{(k+1)}\right)\left(\boldsymbol{x}_j-\boldsymbol{\mu}_i^{(k+1)}\right)^{\top}}{\sum_{j=1}^{n}\tau_{ij}^{(k)}} \tag{2.38}$$

for $i = 1,\ldots,g$ and $j = 1,\ldots,n$, where

$$\tau_{ij}^{(k+1)} = \tau_i(\boldsymbol{x}_j;\boldsymbol{\Psi}^{(k)}).$$

The updated estimate of the i^{th} mixing proportion π_i is as given in (2.35). Several alternative EM algorithms are explored by Gupta and Chen (2010).

2.9 IMAGE SEGMENTATION PERFORMANCE MEASURES

Clustering has become a increasingly popular part of data analysis, but is a relatively young domain of research. Several distance measures were proposed to empirically assess the performance of a clustering algorithm by comparing its output to a given reference clustering. In this work we opted by the following measures MCR, *variation of information* (VOI) (MEILĂ, 2007), *global consistency error* (GCE) (MARTIN *et al.*, 2001), *local consistency error* (LCE) (MARTIN *et al.*, 2001) and *adjusted rand index* (ARI) (HUBERT; ARABIE, 1985).

Following the notes of Meilă (2007), let $\mathcal{C}$ a partition of a set of points D into a mutually disjoint subsets $C_1,\ldots,C_K$ called clusters, i.e.,

$$\mathcal{C} = \{C_1,\ldots,C_K\} \text{ such that } C_k \cap C_l = \emptyset \text{ and } \bigcup_{k=1}^{K} C_k = D,$$

where K is the number of non-empty clusters. Let n and n_k be the number of data points respectively in D and in cluster C_k, thus

$$n = \sum_{k=1}^{K} n_k, \quad n_k > 0 \text{ for } i = 1,\ldots,K.$$

Let $\mathcal{C}' = \{C'_1,\ldots,C'_{K'}\}$ be another clustering of the same dataset D, with cluster sizes $n'_{k'}$.

The *confusion matrix* of the pair $(\mathcal{C},\mathcal{C}')$ can be defined as the $K \times K'$ matrix, whose kk'^{th} element is the number of points in the intersection of clusters C_k of $\mathcal{C}$ and $C'_{k'}$ of $\mathcal{C}'$, thus

$$n_{kk'} = |C_k \cap C'_{k'}|.$$

The MCR computes the proportion of pixels which is miss-classified. This is one of the most simple measures and can be defined from the confusion matrix as

$$\text{MCR}(\mathcal{C},\mathcal{C}') = \frac{1}{n} \sum_{k \neq k'} n_{kk'}.$$

The ARI is a measure of agreement between the partition estimated independent of the groups labeling and the true classification. Its expected value is zero for random partitions, and it is bounded above by one, with higher values representing better partition

accuracy. We can also define it from the confusion matrix as

$$\mathrm{ARI}(\mathcal{C},\mathcal{C}') = \frac{\sum_{k=1}^{K}\sum_{k'=1}^{K'}\binom{n_{kk'}}{2} - \frac{1}{\binom{n}{2}}\left[\sum_{k=1}^{K}\binom{n_k}{2}\right]\left[\sum_{k'=1}^{K'}\binom{n'_{k'}}{2}\right]}{\frac{1}{2}\left[\sum_{k=1}^{K}\binom{n_k}{2} + \sum_{k'=1}^{K'}\binom{n'_{k'}}{2}\right] - \frac{1}{\binom{n}{2}}\left[\sum_{k=1}^{K}\binom{n_k}{2}\right]\left[\sum_{k'=1}^{K'}\binom{n'_{k'}}{2}\right]}.$$

The VOI measures the amount of information lost and gained in changing from clustering $\mathcal{C}$ to clustering $\mathcal{C}'$. This index can be defined as

$$\mathrm{VOI}(\mathcal{C},\mathcal{C}') = \mathcal{H}(\mathcal{C}) + \mathcal{H}(\mathcal{C}') - 2\mathcal{I}(\mathcal{C},\mathcal{C}'),$$

where

$$\mathcal{H}(\mathcal{C}) = -\sum_{k=1}^{K}\Pr(k)\log\Pr(k),$$

$$\Pr(k) = \frac{n_k}{n},$$

$$\mathcal{I}(\mathcal{C},\mathcal{C}') = -\sum_{k=1}^{K}\sum_{k'=1}^{K'}\Pr(k,k')\log\frac{\Pr(k,k')}{\Pr(k)\Pr(k')}$$

and

$$\Pr(k,k') = \frac{n_{kk'}}{n}.$$

Martin *et al.* (2001) combine the local refines error by means of the GCE, which forces all local refinements to be in the same direction, and LCE, which allows refinements in different directions and in different parts of the image. This indexes are defined by

$$\mathrm{GCE}(\mathcal{C},\mathcal{C}') = \frac{1}{n}\min\left\{\sum_i \mathcal{E}(\mathcal{C},\mathcal{C}',p_i), \sum_i \mathcal{E}(\mathcal{C}',\mathcal{C},p_i)\right\}$$

and

$$\mathrm{LCE}(\mathcal{C},\mathcal{C}') = \frac{1}{n}\sum_i \min\left\{\mathcal{E}(\mathcal{C},\mathcal{C}',p_i),\mathcal{E}(\mathcal{C}',\mathcal{C},p_i)\right\},$$

where

$$\mathcal{E}(\mathcal{C},\mathcal{C}',p_i) = \frac{|R(\mathcal{C},p_i)\backslash R(\mathcal{C}',p_i)|}{|R(\mathcal{C},p_i)|},$$

$R(\mathcal{C},p_i)$ is the set of pixels corresponding to the region in cluster $\mathcal{C}$ that contains the pixel p_i and $\backslash$ denotes the difference set. The measures proposed by Martin *et al.* are such that $0 \leq \mathrm{LCE} \leq \mathrm{GCE} \leq 1$, where zero signifies no error.

3 THE EXPONENTIATED CAUCHY-RAYLEIGH MODEL

3.1 INTRODUCTION

Here it is presented the main structural properties of the ECR distribution. We have twofold goal. First using the original and a new two-parameter extended CR distributions in the survival analysis context as alternatives to the gamma, *Birnbaum-Saunders* (BS), Weibull, LN models and other lifetime ones. Several works have used FMs models to describe heavy-tail lifetimes (HARRISON; MILLARD, 1991; MCCLEAN; MILLARD, 1993; GARDINER *et al.*, 2014). But this modeling scheme becomes the associated estimation process hard.

The reason of applying the CR law for describing lifetimes is firstly to accommodate heavy-tail distributions. However, its pdf does not cover non-modal behavior (as the exponential does) as well as the CR hrf assumes only one form, limiting its employment in practice. First, we extend the CR distribution by using the Exp-G class reviewed in Section 2.2. This model has two parameters and is flexible enough to accommodates hrfs with decreasing, decreasing-increasing-decreasing and upside-down bathtub forms. As second goal, several closed-form mathematical expressions for the ECR model are proposed: median, mode, probability weighted, log-, incomplete and order statistic moments and FIM. There are several concepts for the term *closed-form*; as instance, one has the work of Stover and Weisstein (2017). Here, we understand *closed-form* as

> "all expressions in terms of a finite number of known (special) functions, which have well-defined analytic properties."

Some of these quantities do not exist for specific ECR parametric regions, as it will be discussed in our presentation. Further, we propose two estimation procedures for the ECR parameters: MLEs and *percentile-based estimators* (PBEs). It is known MLEs present a bias of order $O(n^{-1})$ for small and moderate sample sizes (n), where $O(\cdot)$ is the Landau notation to represent order. To overcome it we furnish a expression for the CS-ECR, proposing a second bias-corrected MLE. A simple scheme to obtain constrained estimates are developed for the abML and PB methods. A Monte Carlo simulation study is made to quantify the performance of proposed estimators, adopting bias and SSD as figures of merit. An application to active repair times measured from airborne communication transceiver is performed to illustrate the usefulness of the ECR model. Results point out that the ECR distribution may outperform well-defined biparametric models; such as-beyond other nineteen ones-the gamma, BS, Weibull and LN laws, before heavy-tail data.

This chapter is outlined as follows. The proposed model is presented in Section 3.2

and some of its descriptive properties and shape analysis are listed in Section 3.3. Section 3.4 discusses some ECR mathematical expressions. Section 3.5 addresses procedures to obtain the MLEs, PBEs and CS-MLE. A simulation study tasting the convergence, consistency and existence of these estimators is presented in Section 3.6. Two real data applications are yielded in Section 3.7. Finally, Section 3.8 provides concluding remarks.

3.2 THE BROACHED MODEL

In this section, we aim to introduce a CR extension as alternative to classical biparametric models (as the gamma and Weibull distributions), which we denote as the ECR model. This distribution has cdf, for $x > 0$, given by

$$F(x) = \left(1 - \frac{\lambda}{\sqrt{\lambda^2 + x^2}}\right)^{\beta}, \tag{3.1}$$

where $\lambda > 0$ is a scale parameter and $\beta > 0$ is an additional shape parameter and, as a consequence, its pdf, for $x > 0$, is

$$f(x) = \beta\lambda \frac{x}{(\lambda^2 + x^2)^{3/2}} \left(1 - \frac{\lambda}{\sqrt{\lambda^2 + x^2}}\right)^{\beta-1}. \tag{3.2}$$

From (3.1) and (3.2), the ECR sf[1] and hrf is expressed, respectively, as

$$S(x) = 1 - \left(1 - \frac{\lambda}{\sqrt{\lambda^2 + x^2}}\right)^{\beta}$$

and

$$h(x) = \beta\lambda \frac{x}{(\lambda^2 + x^2)^{3/2}} \left(1 - \frac{\lambda}{\sqrt{\lambda^2 + x^2}}\right)^{\beta-1} \left[1 - \left(1 - \frac{\lambda}{\sqrt{\lambda^2 + x^2}}\right)^{\beta}\right]^{-1}. \tag{3.3}$$

In this chapter, an ECR random variable X with parameters β (shape) and λ (scale) will be denoted by $X \sim \text{ECR}(\beta,\lambda)$.

The ECR law is a model nested in the *exponentiated Burr type-XII* (EBXII) (AL-HUSSAINI; AHSANULLAH, 2015)[2], *Kumaraswamy log-logistic* (KLL) (SANTANA, 2010), GFP[3] (ZANDONATTI, 2001), *Kumaraswamy Burr type-XII* (KBXII) (PARANAÍBA, 2012) and McBXII (MEAD, 2014; GOMES *et al.*, 2015) distributions. In general, the properties obtained for these models do not have closed-form expressions, which may imply in computational difficulties and an awkward mathematical treatment (where divergence cases may happen). As foregoing discussed, the ECR model can be very useful, but some

[1] a.k.a. tail function.

[2] a.k.a. *exponentiated log-logistic* (ELL) (LIMA; CORDEIRO, 2017).

[3] a.k.a. *beta Burr type-XII* (BBXII) (PARANAÍBA, 2012) and *McDonald log-logistic* (McLL) (TAHIR *et al.*, 2014).

of its properties do not has closed-form expressions and, therefore, a specialized treatment is required. We do it in the next sections.

The ECR qf is figured out by inverting (3.1), for $0 < p < 1$,

$$Q(p) = \frac{\lambda}{1 - p^{1/\beta}} \sqrt{2p^{1/\beta} - p^{2/\beta}}. \tag{3.4}$$

The ECR median is then determined from using $p = 1/2$ in (3.4):

$$\widetilde{M} = \frac{\lambda}{2^{1/\beta} - 1} \sqrt{2^{\frac{\beta+1}{\beta}} - 1}.$$

We can also use (3.4) for generating outcomes of an ECR random variable by the *inverse transform sampling method* (ITSM) as follows:

- Generate u as an outcome of $U \sim U(0,1)$;
- Obtain $x = Q(u)$ as an outcome from the ECR distribution.

The identifiability property is crucial in statistics (inference, regression, time series, etc.) The Proposition 3.1 ensures the ECR identifiability. A parametric family of pdfs is said to be *identifiable* if distinct parameter values determine distinct members of the family.

Proposition 3.1. *The ECR laws with parameters* $\boldsymbol{\theta}_1 = (\beta_1,\lambda_1)$ *and* $\boldsymbol{\theta}_2 = (\beta_2,\lambda_2)$ *are distinct if* $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$.

3.3 SOME ELEMENTARY PROPERTIES

In this section, we derive some ECR descriptive properties. Proposition 3.2 assures that the ECR pdf and hrf asymptotes are more flexible than the CR law. The ECR pdf and hrf limits assume three distinct values when x approaches zero (in terms of β), while it is always null for the CR case.

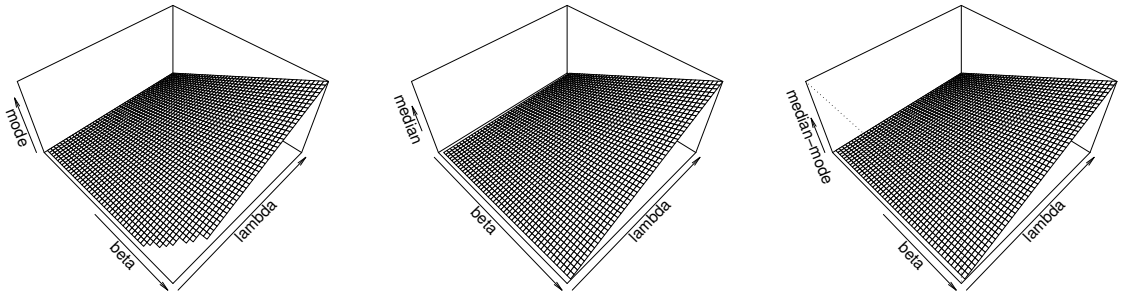
Proposition 3.2. *The ECR pdf and hrf have the following limiting behavior:*

$$\lim_{x\to 0^+} f(x) = \lim_{x\to 0^+} h(x) = \begin{cases} \infty, & \beta < 1/2, \\ \frac{\sqrt{2}}{2\lambda}, & \beta = 1/2, \\ 0, & \beta > 1/2, \end{cases} \tag{3.5}$$

and

$$\lim_{x\to\infty} h(x) = 0.$$

A random variable X is called *regularly varying at infinity* (RVI) with tail index $a > 0$ if $S_X(cx) \sim c^{-a} S_X(x)$ as $x \to \infty$, where $S_X(x) = 1 - F_X(x)$. Cooke *et al.* (2014,



(a) mode (b) median (c) median - mode

Figure 3.1: Surfaces of the ECR mode and median and its differences.

p. 55), citing Chistyakov (1964), defined the *subexponential* distributions containing the RVI class. Foss *et al.* (2011, Lemma 3.2), crediting Chistyakov (1964) again, pointed out that any subexponential distribution with positive support is *long-tailed* and therefore it is *heavy-tailed* too. Thus, the ECR model belongs to all these classes. Proposition 3.3 assigns the ECR distribution to the RVI class.

Proposition 3.3. *The ECR distribution is RVI with tail index* $a = 1$.

Proposition 3.4 shows a closed-form expression for the ECR mode. The ECR pdf is non-modal for $\beta \leq 1/2$ as induced from Proposition 3.2 and unimodal for $\beta > 1/2$.

Proposition 3.4. *The ECR mode is*

$$\mathrm{Mo} = \frac{\lambda}{2\sqrt{2}}\sqrt{(\beta+1)^2 + (\beta-1)\sqrt{\beta^2+6\beta+17}}, \quad \beta > 1/2. \tag{3.6}$$

Corollary 3.4.1. *The ECR mode has the following limiting behavior* $\lim_{\beta \to 1/2^+} \mathrm{Mo} = 0$.

Corollary 3.4.2. *The CR mode is* $\mathrm{Mo}_{CR} = \lambda/\sqrt{2}$.

Figure 3.1 exhibits plots for the ECR mode and median and $\widetilde{M} - \mathrm{Mo}$. The difference $\widetilde{M} - \mathrm{Mo}$ in Figure 3.1 indicates that our model has always positive skewness.

Proposition 3.5 shows the ECR distribution belongs to the scale family. Two immediate consequences of this proposition are $\mathbb{E}(X^k) = \lambda^k \mathbb{E}(Z^k)$ and

$$\mathbb{E}(\log X) = \log \lambda + \mathbb{E}(\log Z). \tag{3.7}$$

In the rest of this chapter, we denote $Z \sim \mathrm{ECR}(\beta,1)$ as the standard ECR random variable.

Proposition 3.5. *If* Z *is the standard ECR random variable, then* $X = \lambda Z \sim \mathrm{ECR}(\beta,\lambda)$.

The following shape analysis is based on Propositions 3.2 and 3.4. Figure 3.2 summarizes the possible shapes of (3.2) and (3.3) and Figures 3.3(a) and 3.3(b) depict some of them. The ECR pdf can assume three shapes:

1. unimodal if $\beta > 1/2$;
2. non-modal (starting at $\sqrt{2}/2\lambda$) if $\beta = 1/2$;
3. non-modal (starting at the origin) if $\beta < 1/2$.

It is also noticeable that the ECR distribution may be more flexible than the CR model. The ECR hrf has four shapes:

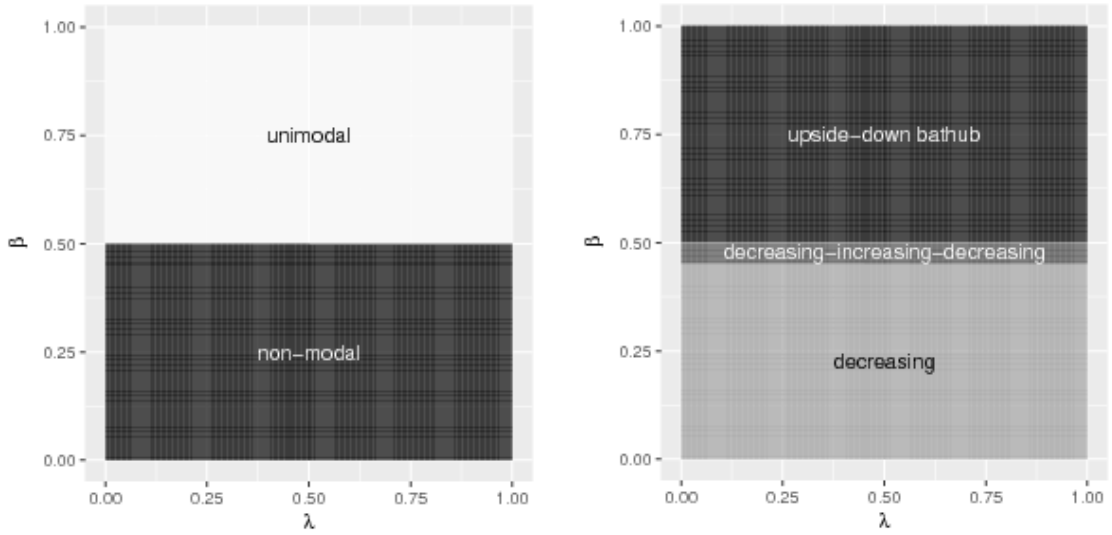
1. upside-down bathtub (starting at the origin) if $\beta > 1/2$;
2. upside-down bathtub (starting at $\sqrt{2}/2\lambda$) if $\beta = 1/2$;
3. decreasing-increasing-decreasing if $0.45259 < \beta < 1/2$;
4. decreasing if $\beta \leq 0.45259$.

The decreasing-increasing-decreasing and decreasing hrf shapes threshold are obtained numerically based on the change of signal of the ECR hrf first derivative. For any value $0.45259 < \beta < 1/2$ the ECR hrf first derivative change its signal when $x \to \infty$. In contrast, the CR hrf only assumes upside-down bathtub shape (starting at the origin). The parameter λ only has scale effects in pdf and hrf shapes. It is notable that in Figure 3.3(b) three different features of hrf may be obtained with the additional parameter assuming values minors than 1. Figure 3.3(a) also sketches the histograms of simulated observations using the ITSM described above. The pdfs came near to the corresponding histograms midpoints. In other words, the additional shape parameter β allow for a degree of flexibility. So, the new model can be very useful in many practical situations for modeling positive real data.

As will be presented in Corollary 3.7.1 the ECR model does not admit integer moments, then it is not possible to define the classical skewness and kurtosis measures, but we can employ alternative (quantile based) measures like the Bowley skewness (KENNEY; KEEPING, 1962), and the Moors kurtosis (MOORS, 1988), defined, respectively, by

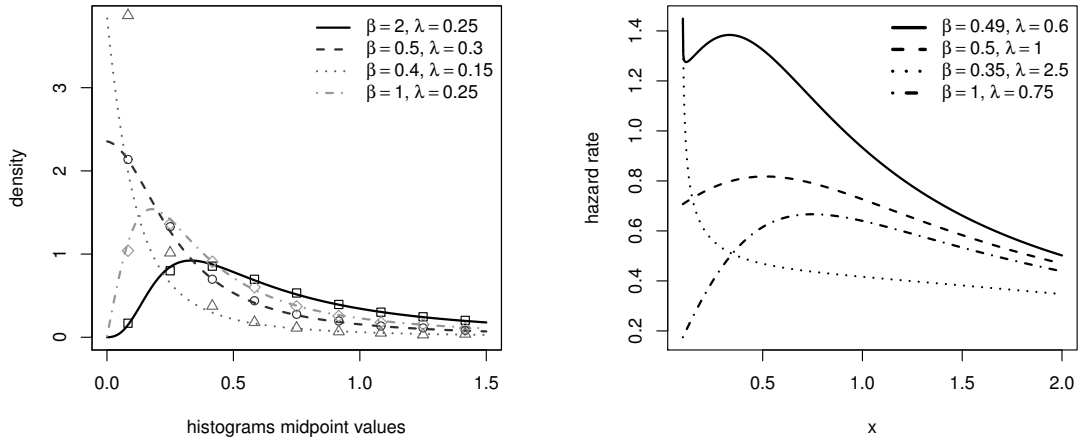
$$\mathcal{S}_B = \frac{\sum_{k=1}^{3}(-1)^{k-1}\binom{2}{k-1}Q\left(\frac{k}{4}\right)}{\sum_{k=1}^{2}(-1)^{k}Q\left(\frac{2k-1}{4}\right)} \text{ and } \mathcal{K}_M = \frac{\sum_{k=1}^{4}(-1)^{k}Q\left(\frac{2k-1}{8}\right)}{\sum_{k=1}^{2}(-1)^{k}Q\left(\frac{2k-1}{4}\right)}.$$

These measures are less sensitive to outliers and exist even for distributions without moments. For symmetric unimodal distributions, positive kurtosis indicates heavy tails and peakedness relative to the normal distribution, whereas negative kurtosis indicates light tails and flatness. For the normal distribution, $\mathcal{S}_B = \mathcal{K}_M = 0$. In the ECR case these measures only depends of the β parameter. In Figure 3.4, we plot the measures Bowley



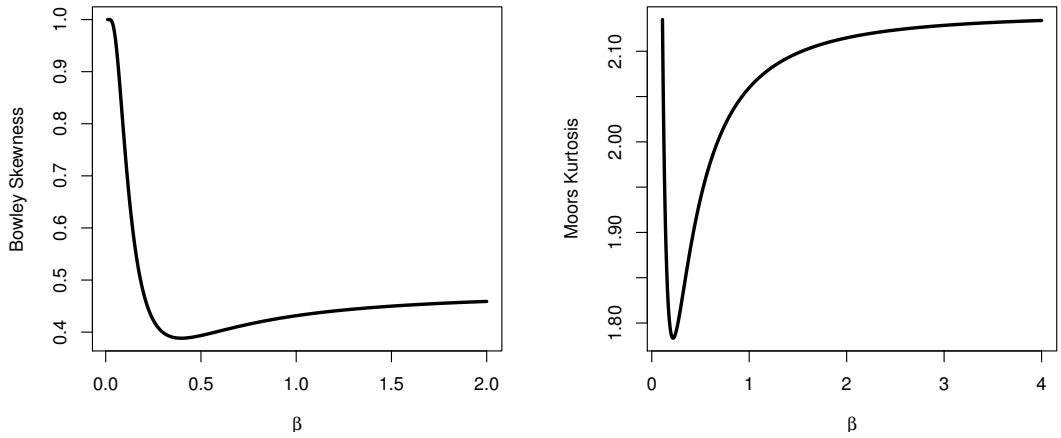
(a) pdf. (b) hrf.

Figure 3.2: Summaries of the ECR shapes.



(a) pdf and sample histogram midpoints (b) hrf

Figure 3.3: ECR hrf and pdf curves for some parameter values.



(a) Bowley skewness. (b) Moors kurtosis.

Figure 3.4: Curves of the ECR quantile measures.

skewness and Moors kurtosis curves for some β values. These plots indicate that both measures are very sensitive on this shape parameter. It is possible obtain numerically the minimum values of these measures. The minimum Bowley skewness is 0.38838 and is obtained when $\beta \approx 0.396897$ while the mininum Moors kurtosis is 1.78325 and is achieved when $\beta \approx 0.218776$. The Proposition 3.6 shows that the Moors kurtosis increases indefinitely when β goes to zero and stabilizes around 2.14174 when β goes to infinity. The behavior of the Bowley skewness is similar raising to 1 as β approximates of 0 and stabilizing around 0.476281 as β increases.

Proposition 3.6. *The Bowley skewness $\mathcal{S}_B$ and the Moors kurtosis $\mathcal{K}_M$ have the following limiting behaviors:*

1. $\lim_{\beta\to 0^+} \mathcal{S}_B = 1$;
2. $\lim_{\beta\to 0^+} \mathcal{K}_M = \infty$;
3. $\lim_{\beta\to\infty} \mathcal{S}_B = \dfrac{\log(2)\log(4) - \log\left(\frac{4}{3}\right)\log(8)}{\log(2)\log(3)} \approx 0.476281\ldots$;
4. $\lim_{\beta\to\infty} \mathcal{K}_M = \log\left(\frac{4}{3}\right)\log(4)\left[\frac{1}{\log\left(\frac{8}{7}\right)\log(3)} + \frac{1}{\log\left(\frac{8}{3}\right)\log(8)} - \frac{1}{\log\left(\frac{8}{5}\right)\log(3)}\right] \approx 2.14174\ldots$.

3.4 MOMENTS

Expressions for the r-moments, incomplete and order statistics moments of some ECR extensions were explored by Paranaíba (2012), Mead (2014), Al-Hussaini and Ahsanullah (2015), Silva *et al.* (2015), Gomes *et al.* (2015) and Cordeiro *et al.* (2016). Unfortunately, these extended results do not have closed-form expressions. In general, they were deduced from power series expansions. In this section, it is showed that the ECR moments often do not exist and, therefore, it is required a detailed discussion about it.

3.4.1 Probability weighted moments

The *probability weighted moments* (pwms) with order indexes $r, s, t \in \mathbb{R}$, $\mu'_{r,s,t}$, for a random variable $X \sim \text{ECR}(\beta,\lambda)$ are defined by $\mu'_{r,s,t} = \mathbb{E}\left(X^r F(X)^s [1-F(X)]^t\right)$. Proposition 3.7 presents a closed-form expression for the ECR pwms with $t \in \mathbb{Z}$.

Proposition 3.7. *Let $X \sim \text{ECR}(\beta,\lambda)$, for $-2(s+1)\beta < r < 1$ and $t \in \mathbb{Z}$,*

$$\mu'_{r,s,t} = \beta(\lambda\sqrt{2})^r \sum_{i=0}^{t} (-1)^t \binom{t}{i} \text{B}\left(1-r, \frac{r}{2} + (s+i+1)\beta\right) \times {}_2\text{F}_1\left(-\frac{r}{2}, \frac{r}{2} + (s+i+1)\beta; 1 - \frac{r}{2} + (s+i+1)\beta; \frac{1}{2}\right),$$

where ${}_2\text{F}_1$ is the Gauss's hypergeometric function defined as

$${}_2\text{F}_1(a,b;c;x) = \sum_{k=0}^{\infty} \frac{(a)_k (b)_k}{(c)_k} \frac{x^k}{k!}, \quad (a)_k = \frac{\Gamma(a+k)}{\Gamma(a)},$$

$\text{B}(\cdot,\cdot)$ *is the beta function and* $\Gamma(\cdot)$ *is the gamma function. The convergence is fulfilled when $c > b > 0$, $|x| < 1$ and the branch of $(1-tx)^{-a}$ is chosen so that $(1-tx)^{-a} \to 1$ as $t \to 0$.*

Foss *et al.* (2011, p. 33) showed that in the RVI class with index $a > 0$ all moments of order $0 < r < a$ are finite, while all moments with order $r > a$ are infinite. This fact occurs in the ECR distribution, and they will be presented in Corollary 3.7.1. Closed-form expressions for the ECR moments consist in a simple product involving the beta and the Gauss's hypergeometric functions. Figure 3.5(a) depicts some ECR pwm curves. The pwm values increase as s and t values decrease.

Corollary 3.7.1. *Let $X \sim \text{ECR}(\beta,\lambda)$, for $r \in (-2\beta, 1)$,*

$$\mathbb{E}(X^r) = \beta(\lambda\sqrt{2})^r \, \text{B}\left(1-r, \frac{r}{2} + \beta\right) \, {}_2\text{F}_1\left(-\frac{r}{2}, \frac{r}{2} + \beta; 1 - \frac{r}{2} + \beta; \frac{1}{2}\right). \tag{3.8}$$

Corollary 3.7.2 presents the CR moments, derived from the Corollary 3.7.1. The obtained result agrees with the known results for the SM distribution as presented by Kleiber and Kotz (2003, eq. 6.46) (Assuming $a = 2$, $b = \lambda$ and $q = 1/2$ in the BXII parameterization adopted by the referred authors).

Corollary 3.7.2. *Let $Y \sim \text{CR}(\lambda)$, for $r \in (-2, 1)$,*

$$\mathbb{E}(Y^r) = \frac{\lambda^r}{\sqrt{\pi}} \Gamma\left(\frac{1-r}{2}\right) \Gamma\left(1 + \frac{r}{2}\right) = \frac{\lambda^r}{2} \text{B}\left(\frac{1-r}{2}, 1 + \frac{r}{2}\right). \tag{3.9}$$

In Table 3.1 we present a comparison between the sample and expected probability weighted moments. We can note that the corresponding values difference increases as

Table 3.1: Some expected values of the standard ECR pwms and its corresponding sample moments.

			$\beta = 0.5$		$\beta = 1$		$\beta = 2$	
r	s	t	$\mu'_{r,s,t}$	$\overline{X^r_{s,t}}$	$\mu'_{r,s,t}$	$\overline{X^r_{s,t}}$	$\mu'_{r,s,t}$	$\overline{X^r_{s,t}}$
0.75	0	0	2.249	1.999	3.778	3.368	6.282	5.545
0.50	0	0	1.297	1.282	1.854	1.832	2.597	2.540
0.25	0	0	1.041	1.041	1.259	1.258	1.493	1.484
0.50	1	0	0.927	0.915	1.299	1.276	1.814	1.752
0.50	0	1	0.370	0.368	0.555	0.555	0.784	0.788
0.50	1	1	0.173	0.172	0.246	0.246	0.342	0.341

the order index r nearing 1, this is normal because the expected probability weighted moments go to infinity as the order index r approaches 1 while the sample probability weighted moments are always finite. The corresponding sample moments are calculated using random samples of size 10,000.

3.4.2 First log-moment and an induced regression model

Proposition 3.8 contains an useful result to define a log-linear regression model or a location quantity (since the mean is infinite) for the ECR distribution.

Proposition 3.8. *Let* $X \sim \mathrm{ECR}(\beta,\lambda)$*, then*

$$\mathbb{E}(\log X) = \log\lambda + \frac{1}{2}\mathbf{\Phi}\left(\frac{1}{2}; 1,\beta\right) + \Psi(1+\beta) + \boldsymbol{\gamma} - \frac{1}{\beta},$$

where $\mathbf{\Phi}(\cdot;\cdot,\cdot)$ *is the Lerch transcendental Phi function,* $\Psi(\cdot)$ *is the digamma function and* $\boldsymbol{\gamma} = 0.57721\ldots$ *is the Euler-Mascheroni constant defined by, respectively:*

$$\mathbf{\Phi}(x; s,a) = \sum_{n=0}^{\infty} \frac{x^n}{(n+a)^s}, \quad \Psi(x) = \frac{\mathrm{d}}{\mathrm{d}x}\log\Gamma(x) \text{ and } \boldsymbol{\gamma} = \lim_{n\to\infty}\left(-\log n + \sum_{k=1}^{n}\frac{1}{k}\right).$$

An outline of a possible ECR regression is given as follows. Let $\epsilon_i \sim \mathrm{ECR}(\alpha,1)$ and $\boldsymbol{z}_i = (1, z_{i1}, \ldots, z_{ip})^\top$ be a vector of predict variables. Assuming that there is interest in describing linearly the expected value of logarithm of a RVI positive variable, the following result holds:

$$Y_i = \mathrm{e}^{\boldsymbol{z}_i^\top\boldsymbol{\beta}}\epsilon_i \sim \mathrm{ECR}(\alpha, \exp(\boldsymbol{z}_i^\top\boldsymbol{\beta})) \quad \Longleftrightarrow \quad \log(Y_i) = \boldsymbol{z}_i^\top\boldsymbol{\beta} + \log(\epsilon_i),$$

where $\boldsymbol{\beta}_0^* := \beta_0 + \mathbb{E}[-\log(\epsilon_i)]$ and $\boldsymbol{\beta} = (\beta_0^*, \beta_1, \ldots, \beta_p)^\top$ is a vector of regression coefficients. In this case, notice that $\mathbb{E}(\log Y_i) = \beta_0 + \sum_{k=1}^{p}\beta_k z_{ik}$ and it is possible to proof that $\mathrm{Var}(\log Y_i) = \mathrm{Var}(\log\epsilon_i) < \infty$. We omit the expression of $\mathrm{Var}(\log Y_i)$ by simplicity, but it may be obtained from author's contact.

3.4.3 Incomplete moments

Let $X \sim \text{ECR}(\beta,\lambda)$, its incomplete moments are presented by Proposition 3.9. The incomplete moments are defined by $m_r(x_0) = \int_0^{x_0} x^r f(x)\,\mathrm{d}x$. It is known incomplete moments consist in the main part of important inequality tools in income applications (BUTLER; MCDONALD, 1989) such as Lorenz curves (LORENZ, 1905) and Gini measure (GINI, 1921). For future applications, users of the ECR law may employ the Proposition 3.9 to obtain these measures.

Proposition 3.9. *Let $X \sim \text{ECR}(\beta,\lambda)$, for $r > -2\beta$*

$$m_r(x_0) = \frac{\beta 2^{r/2+1}\lambda^r u_0^{\beta+r/2}}{2\beta + r}\text{F}_1\left(\frac{r}{2}+\beta,r,-\frac{r}{2};\frac{r}{2}+\beta+1;u_0,\frac{u_0}{2}\right),$$

where $u_0 = 1 - \lambda/\sqrt{x_0^2+\lambda^2}$ and F_1 is the first Appell's hypergeometric function defined by

$$\text{F}_1(a,b_1,b_2;c;x,y) = \sum_{i,j=0}^{\infty}\frac{(a)_{i+j}(b_1)_i(b_2)_j}{(c)_{i+j}}\frac{x^i}{i!}\frac{y^j}{j!}.$$

3.4.4 Order statistics

Moments of the order statistics play an important role in quality control and reliability issues (AHSANULLAH *et al.*, 2013), where researchers wish to predict the future failure items based on recorded failure times. Let $X_1,\ldots,X_n$ be a random sample from the ECR distribution and $X_{(1)} \leq \ldots \leq X_{(n)}$ denote the corresponding order statistics. Let $f_{(i)}$ be the pdf of the i^{th} order statistic $X_{(i)}$:

$$f_{(i)}(x) = \frac{f(x)F(x)^{i-1}S(x)^{n-i}}{\text{B}(i,n-i+1)},$$

where $F(x)$ and $f(x)$ are given in (3.1) and (3.2), respectively. Then we can express the ECR order statistics pdf as

$$f_{(i)}(x) = \frac{\beta\lambda}{\text{B}(i,n-i+1)}\frac{x}{(\lambda^2+x^2)^{3/2}}\left(1-\frac{\lambda}{\sqrt{\lambda^2+x^2}}\right)^{i\beta-1}\left[1-\left(1-\frac{\lambda}{\sqrt{\lambda^2+x^2}}\right)^{\beta}\right]^{n-i}.$$

In Proposition 3.10, we derive a closed-form expression for the ECR order statistics moments.

Proposition 3.10. *Let $X_{(i)}$ be the i^{th} order statistic of a n-points random sample from*

Table 3.2: Some values of the corresponding sample and expected r-moments order statistics of a random variable $X \sim \text{ECR}(1/2,1/2)$ of size $n = 10$.

i	1	2	3	4	5	6	7	8	9	10
$\mathbb{E}(X_{(i)}^{-2})$	-	-	87.050	27.086	12.135	6.200	3.285	1.682	0.757	0.237
$\overline{X_{(i)}^{-2}}$	-	-	81.931	28.488	12.794	6.183	3.281	1.679	0.765	0.239
$\mathbb{E}(X_{(i)}^{-1})$	-	13.946	6.773	4.311	3.022	2.198	1.601	1.128	0.723	0.355
$\overline{X_{(i)}^{-1}}$	-	14.056	6.824	4.334	3.031	2.201	1.604	1.129	0.726	0.356
$\mathbb{E}(X_{(i)}^{1/4})$	0.462	0.582	0.661	0.727	0.789	0.853	0.926	1.021	1.173	1.562
$\overline{X_{(i)}^{1/4}}$	0.463	0.583	0.660	0.726	0.790	0.852	0.925	1.019	1.168	1.570
$\mathbb{E}(X_{(i)}^{1/2})$	0.230	0.350	0.447	0.538	0.632	0.739	0.873	1.068	1.435	2.857
$\overline{X_{(i)}^{1/2}}$	0.232	0.351	0.448	0.540	0.633	0.740	0.874	1.068	1.444	2.836
$\mathbb{E}(X_{(i)}^{3/4})$	0.121	0.216	0.308	0.405	0.514	0.650	0.839	1.148	1.850	7.319
$\overline{X_{(i)}^{3/4}}$	0.120	0.216	0.310	0.406	0.517	0.654	0.845	1.146	1.835	7.487

$X \sim \text{ECR}(\beta,\lambda)$, *for* $r \in (-2i\beta,1)$,

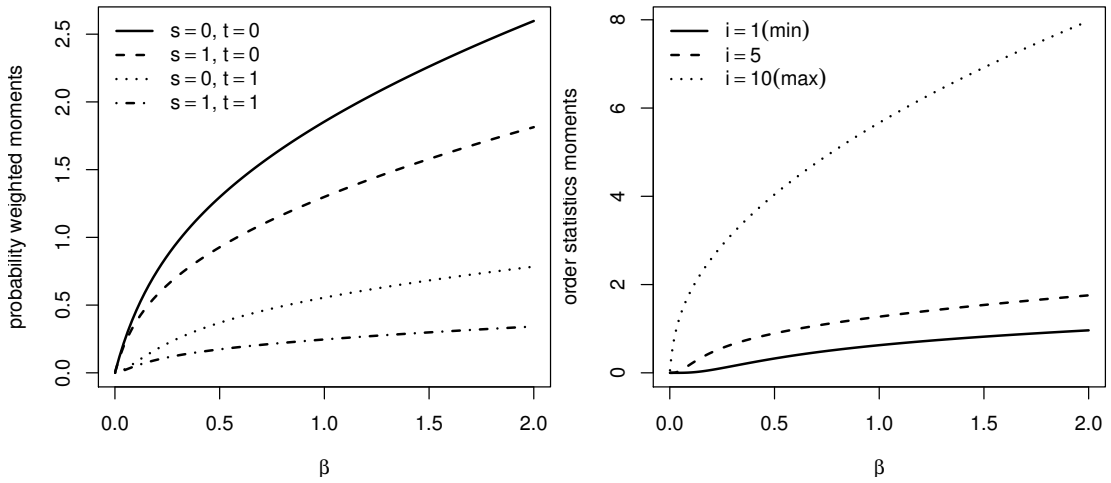
$$\mathbb{E}(X_{(i)}^r) = \frac{\beta(\lambda\sqrt{2})^r}{\text{B}(i,n-i+1)} \sum_{j=0}^{n-i} (-1)^j \binom{n-i}{j} \text{B}\left(1-r,\frac{r}{2}+(i+j)\beta\right) \times {}_2\text{F}_1\left(\frac{-r}{2},\frac{r}{2}+(i+j)\beta;1-\frac{r}{2}+(i+j)\beta;\frac{1}{2}\right).$$

In Table 3.2 we present a comparison between sample $\overline{X_{(i)}^{-n}}$ and corresponding expected r-moments order statistics $\mathbb{E}(X_{(i)}^{-n})$. The r-moments order statistics $\mathbb{E}(X_1^{-2})$, $\mathbb{E}(X_2^{-2})$ and $\mathbb{E}(X_1^{-1})$ are omitted because the corresponding order statistics are not defined by Proposition 3.10. The sample r-moments order statistics are calculated using 10,000 random samples of size $n = 10$.

Figure 3.5(b) depicts some ECR order statistics moments curves. As expected the order statistics moments build up as the values of i increases.

3.5 INFERENCE AND SECOND-ORDER ASYMPTOTIC THEORY

In this section, we develop inference procedures under censored and uncensored samples. Uncensored schemes are explored using ML and PB methods, while the censored schemes are focused only on ML method. A simple scheme to obtain constrained estimates are developed for the abML and PB methods. The MLEs are the background for developed hypotheses tests. The first and second order bias of the MLE are also studied.



(a) probability weighted moments. (b) order statistics moments.

Figure 3.5: Curves of the standard ECR moments for some parameter values and $r = 0.5$.

3.5.1 Estimators for uncensored samples

We develop two procedures to estimate the ECR parameters under the uncensored samples. We consider the ML and PB methods. Let $x_1, \ldots, x_n$ be an observed uncensored sample obtained from the ECR distribution with vector parameter $\boldsymbol{\theta} = (\beta,\lambda)^\top \in \Theta = \mathbb{R}^+ \times \mathbb{R}^+$.

Maximum likelihood estimators

Firstly, it is provided a procedure to find MLEs for the ECR parameters. The lf evaluated at $\boldsymbol{\theta}$ is

$$\mathcal{L}(\boldsymbol{\theta}) = \prod_{i=1}^{n} \beta\lambda \frac{x_i}{(\lambda^2 + x_i^2)^{3/2}} \left[1 - \frac{\lambda}{\sqrt{\lambda^2 + x_i^2}}\right]^{\beta-1}$$

and the corresponding llf is

$$\ell(\boldsymbol{\theta}) = n\log(\beta\lambda) + T_1(\boldsymbol{x}) + 3\,T_2(\lambda,\boldsymbol{x}) + (\beta - 1)\,T_3(\lambda,\boldsymbol{x}), \tag{3.10}$$

where $\boldsymbol{\theta}$ is the parametric space and

$$T_1(\boldsymbol{x}) = \sum_{i=1}^{n} \log x_i,$$

$$T_2(\lambda,\boldsymbol{x}) = \sum_{i=1}^{n} \log\left(\frac{1}{\sqrt{\lambda^2 + x_i^2}}\right) \tag{3.11}$$

and

$$T_3(\lambda,\boldsymbol{x}) = \sum_{i=1}^{n} \log\left(1 - \frac{\lambda}{\sqrt{\lambda^2 + x_i^2}}\right). \tag{3.12}$$

The ML estimate for $\boldsymbol{\theta}$ is the pair that maximizes the llf (or, equivalently, the lf); that is, $\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}\in\Theta}[\ell(\boldsymbol{\theta})]$, where Θ is the parametric space. MLEs have not closed-form and, therefore, it is required using iterative numerical methods such as Newton-Raphson and BFGS, implemented in several computational platforms like that in Section 1.2.

Other fashion to find ML estimates is by llf derivatives, known as score functions. The components of the score vector $\mathbf{U} = (U_\beta(\boldsymbol{\theta}),U_\lambda(\boldsymbol{\theta}))^\top$ are

$$U_\beta(\boldsymbol{\theta}) = \frac{n}{\beta} + T_3(\lambda,\boldsymbol{x}),$$

and

$$U_\lambda(\boldsymbol{\theta}) = \frac{n}{\lambda} + (1-\beta)T_4(\lambda,\boldsymbol{x}) - (\beta+2)T_5(\lambda,\boldsymbol{x}),$$

where

$$T_4(\lambda,\boldsymbol{x}) = \sum_{i=1}^{n} \frac{1}{\sqrt{\lambda^2 + x_i^2}},$$

and

$$T_5(\lambda,\boldsymbol{x}) = \lambda \sum_{i=1}^{n} \frac{1}{\lambda^2 + x_i^2}.$$

The ML estimates, $\hat{\boldsymbol{\theta}} = (\hat{\beta},\hat{\lambda})^\top$, can also be obtained numerically by solving the *log-likelihood equations* (lles) system

$$\begin{cases} U_\beta(\boldsymbol{\theta}) & = 0 \\ U_\lambda(\boldsymbol{\theta}) & = 0. \end{cases} \tag{3.13}$$

From (3.13) it is possible to obtain a semi-closed MLE for β.

$$\hat{\beta}(\lambda) = -\frac{n}{T_3(\lambda,\boldsymbol{x})}. \tag{3.14}$$

It is clear by (3.14) that when λ is known the ML estimate, $\hat{\beta}$, exist and is unique. Proposition 3.11 is useful for constrained ML estimation processes. Suppose we need an estimate for β constrained to (0,1), the following result shows how to find the corresponding constrain for the λ parameter induced by (3.14).

Proposition 3.11. *Let* $\Theta_\beta \subset \mathbb{R}^+$ *and* $\Theta_\lambda^\beta = \left\{\lambda \in \mathbb{R}^+ | \hat{\beta}(\lambda) \in \Theta_\beta\right\}$ *then for a given sample*

$$\inf \Theta_\lambda^\beta = \begin{cases} \arg\min\limits_{\lambda} \left|\hat{\beta}(\lambda) - \sup \Theta_\beta\right| & \text{if } \exists \sup \Theta_\beta \\ 0 & \text{if } \nexists \sup \Theta_\beta \end{cases}$$

and

$$\sup \Theta_\lambda^\beta = \begin{cases} \arg\min\limits_{\lambda} \left|\hat{\beta}(\lambda) - \inf \Theta_\beta\right| & \text{if } \inf \Theta_\beta > 0 \\ \infty & \text{if } \inf \Theta_\beta = 0. \end{cases}$$

Table 3.3: Simulated datasets.

DS1	0.0773	0.0844	0.1661	0.2780	0.4245	0.7309	0.7442	0.9227	0.9689	1.0
DS2	0.0053	0.0096	0.0115	0.0253	0.0266	0.0319	0.0462	0.1055	0.3072	1.0
DS3	0.0130	0.0314	0.0349	0.0354	0.0489	0.0912	0.1117	0.1162	0.1735	1.0
DS4	0.0177	0.0400	0.0484	0.0540	0.0709	0.0965	0.1279	0.1419	0.5060	1.0
DS5	0.0317	0.0446	0.0644	0.1165	0.1167	0.1654	0.1923	0.1958	0.4448	1.0
DS6	0.0421	0.0483	0.0991	0.1072	0.1224	0.1845	0.1879	0.2662	0.6115	1.0
DS7	0.0576	0.0581	0.0622	0.0815	0.1491	0.1947	0.2169	0.2637	0.5897	1.0

By replacing β by $\hat{\beta}(\lambda)$ in (3.10), the following *profile log-likelihood function* (pllf) for λ is obtained:

$$\ell_\beta(\lambda) = n\left[\log\left(\frac{-n\lambda}{T_3(\lambda,\boldsymbol{x})}\right) - 1\right] + T_1(\boldsymbol{x}) + 3T_2(\lambda,\boldsymbol{x}) - T_3(\lambda,\boldsymbol{x}).$$

We can also obtain the ML estimates for λ by maximizing the pllf $\ell_\beta(\lambda)$ with respect to λ. The *profile score function* (psf) is

$$U_{\lambda|\beta=\hat{\beta}(\lambda)}(\boldsymbol{\theta}) = \frac{n}{\lambda} + \left(1 + \frac{n}{T_3(\lambda,\boldsymbol{x})}\right) T_4(\lambda,\boldsymbol{x}) - \left(2 - \frac{n}{T_3(\lambda,\boldsymbol{x})}\right) T_5(\lambda,\boldsymbol{x}).$$

The ML estimate for λ is always a root of the psf.

Corollary 3.11.1 shows how to find constrained ML estimates in a rectangular region of the parametric space. The existence of constrained estimates is assured only when $\mathfrak{L}$ is a compact set (IZMAILOV; SOLODOV, 2009, p. 8).

Corollary 3.11.1. *Let $\Theta_\beta, \Theta_\lambda \subset \mathbb{R}^+$, Θ_λ^β as given in Proposition 3.11, $\mathfrak{L} = \Theta_\lambda \cap \Theta_\lambda^\beta$ and $\mathfrak{C}$ be the set of all psf roots. For a given sample $(\hat{\beta}(\lambda^*),\lambda^*)$ are ML estimates, constrained to $\Theta_\beta \times \Theta_\lambda$, if*

1. $\lambda^* \in \mathfrak{C}$, $\lambda^* = \inf \mathfrak{L}$ *or* $\lambda^* = \sup \mathfrak{L}$;
2. $\lambda^* \in \mathfrak{L}$;
3. $\ell_\beta(\lambda^*) \geq \ell_\beta(\lambda)\ \forall \lambda \in \mathfrak{C}$;
4. $\ell_\beta(\lambda^*) \geq \lim\limits_{\lambda \to (\inf \mathfrak{L})^+} \ell_\beta(\lambda)$;
5. $\ell_\beta(\lambda^*) \geq \lim\limits_{\lambda \to (\sup \mathfrak{L})^-} \ell_\beta(\lambda)$.

Table 3.3 contains seven simulated datasets for which the ML and PB estimation process present problems like multiple local maximums of the estimation functions and non-existence of the estimative.

Figure 3.6 depicts the pllfs and psfs of the simulated datasets DS1-DS4 in Table 3.3. Figure 3.6(a) and Figure 3.6(b) present the corresponding plots of the DS1 dataset and

is clear the existence of a local maximum in the pllf, psf has two critical points, but it is not possible to obtain a global maximum. Figure 3.6(c) and Figure 3.6(d) display the associated plots of the DS2 dataset. The psf does not present any critical points and then the pllf does not have a maximum again. The plots of the DS3 dataset are presented in Figure 3.6(e) and Figure 3.6(f). This psf has three critical points and the pllf presents a local and a global maximum. Figures 3.6(g) and 3.6(h) represent the DS4 dataset, the pllf present two "numerically equal" maximums.

Percentile-based estimators

For data coming from a distribution having closed-form cdf, a natural way to obtain unknown parameters is to determine the argument which minimizes the square distance between theoretical and sample percentiles. The PBEs are similar to *least squares estimators* (LSEs) both techniques are based on minimizing square distances between expected values of order-statistics functions and their corresponding sample statistics. While the PBE is based on the qf, the LSE method is based on the cdf. Murthy *et al.* (2004) discussed this method for the Weibull distribution, while Gupta and Kundu (2001a) studied the *exponentiated exponential* (EE) case.

Let $x_{(1)},\ldots,x_{(n)}$ be the order statistics obtained from a set of n *independent and identically distributed* (iid) random variables following the ECR distribution with vector parameter $\boldsymbol{\theta} = (\beta,\lambda)^\top$, the *percentile-based estimation function* (pef) at $\boldsymbol{\theta}$ is

$$\mathfrak{p}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left[\frac{\lambda}{1-p_i^{1/\beta}} \sqrt{\left(2-p_i^{1/\beta}\right) p_i^{1/\beta}} - x_{(i)} \right]^2 . \tag{3.15}$$

In this case, p_i is the cdf value at the i^{th} order statistic, $F(x_{(i)})$. Murthy *et al.* (2004, p. 63) describe several ways to define sample p_i. We employ the *mean rank* given by $p_i = i/(n+1)$, which is an unbiased estimator of p_i.

The pef partial derivatives[4] are

$$\mathbf{L}_\beta(\boldsymbol{\theta}) = \frac{2\lambda}{\beta^2}\left[\lambda T_6(\beta,\boldsymbol{x}) - T_7(\beta,\boldsymbol{x})\right]$$

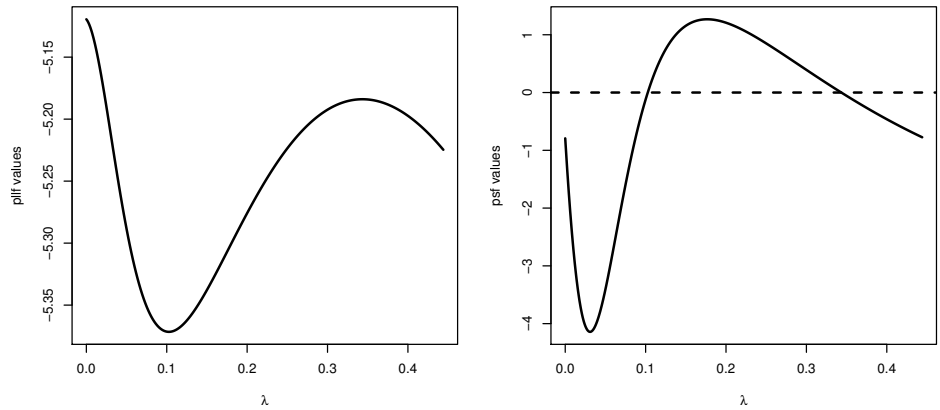
and

$$\mathbf{L}_\lambda(\boldsymbol{\theta}) = 2\left[T_8(\beta,\boldsymbol{x}) - \lambda T_9(\beta,\boldsymbol{x})\right],$$

where

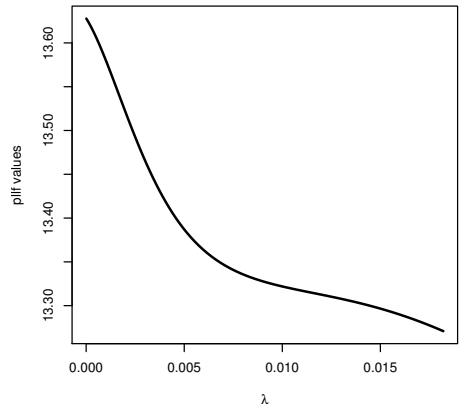
$$T_6(\beta,\boldsymbol{x}) = \sum_{i=1}^{n} \frac{p_i^{1/\beta}\log p_i}{\left(1-p_i^{1/\beta}\right)^3},$$

[4]These derivatives are analogous to the score functions of the ML estimation

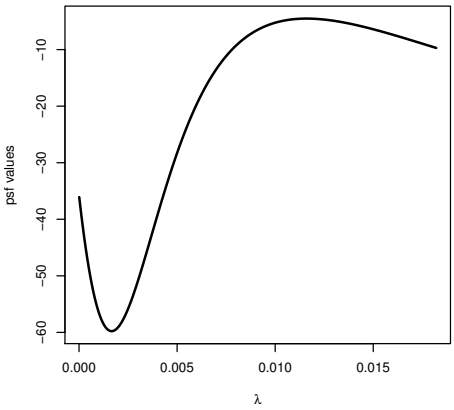


(a) DS1 pllf.

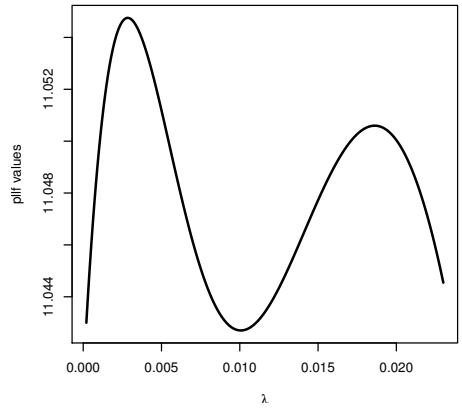
(b) DS1 psf.



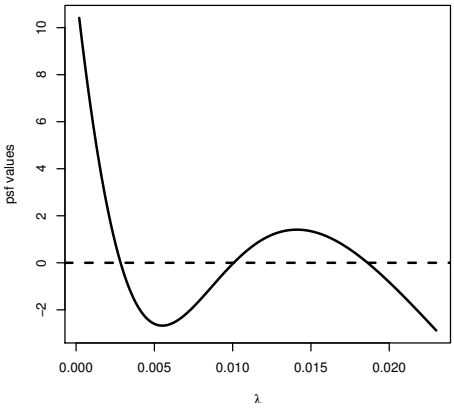
(c) DS2 pllf.



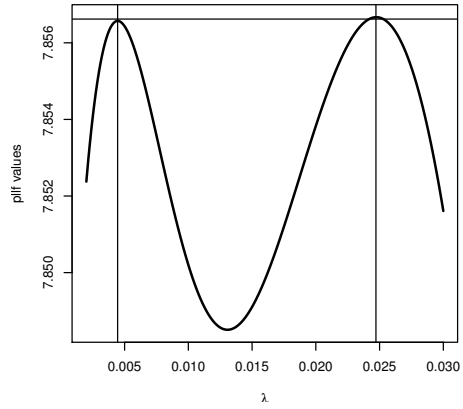
(d) DS2 psf.



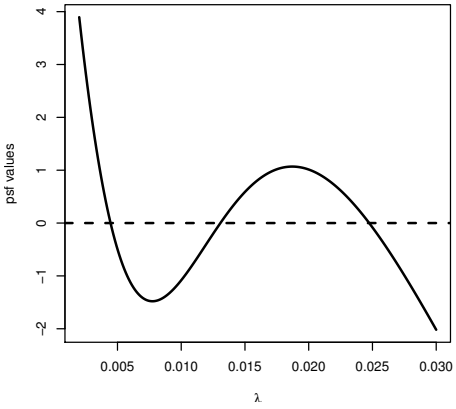
(e) DS3 pllf.



(f) DS3 psf.



(g) DS4 pllf.



(h) DS4 psf.

Figure 3.6: Profile ML estimation plots for simulated datasets in Table 3.3.

$$T_7(\beta,\boldsymbol{x}) = \sum_{i=1}^{n} \frac{x_{(i)} \log p_i}{\left(1-p_i^{1/\beta}\right)^2} \sqrt{\frac{p_i^{1/\beta}}{2-p_i^{1/\beta}}},$$

$$T_8(\beta,\boldsymbol{x}) = -\sum_{i=1}^{n} \frac{x_{(i)} \sqrt{\left(2-p_i^{1/\beta}\right) p_i^{1/\beta}}}{1-p_i^{1/\beta}}$$

and

$$T_9(\beta,\boldsymbol{x}) = n - \sum_{i=1}^{n} \frac{1}{\left(1-p_i^{1/\beta}\right)^2}.$$

The PB estimates, $\breve{\boldsymbol{\theta}} = (\breve{\beta},\breve{\lambda})^\top$, can be obtained numerically by solving the nonlinear equations system

$$\begin{cases} \mathbf{L}_\beta(\boldsymbol{\theta}) & = 0 \\ \mathbf{L}_\lambda(\boldsymbol{\theta}) & = 0. \end{cases} \tag{3.16}$$

In general, PBEs must be obtained numerically, similarly to MLEs. Note that, from (3.16), it is possible to obtain two semi-closed PBEs for λ given by

$$\breve{\lambda}(\beta) = \frac{T_8(\beta,\boldsymbol{x})}{T_9(\beta,\boldsymbol{x})} \tag{3.17}$$

and

$$\breve{\lambda}^*(\beta) = \frac{T_7(\beta,\boldsymbol{x})}{T_6(\beta,\boldsymbol{x})}.$$

By (3.17), when β is known the PB estimate, $\hat{\beta}$, exist and is given by $\breve{\lambda}(\beta)$. Proposition 3.12 is useful for constrained PB estimation processes, it is analogous to Proposition 3.11.

Proposition 3.12. *Let* $\Theta_\lambda \subset \mathbb{R}^+$ *and* $\Theta_\beta^\lambda = \left\{\beta \in \mathbb{R}^+ | \hat{\lambda}(\beta) \in \Theta_\lambda\right\}$ *then for a given sample*

$$\inf \Theta_\beta^\lambda = \begin{cases} \arg\min_\beta \left|\breve{\lambda}(\beta) - \sup \Theta_\lambda\right| & \text{if } \exists \sup \Theta_\lambda \\ 0 & \text{if } \nexists \sup \Theta_\lambda \end{cases}$$

and

$$\sup \Theta_\beta^\lambda = \begin{cases} \arg\min_\beta \left|\breve{\lambda}(\beta) - \inf \Theta_\lambda\right| & \text{if } \inf \Theta_\lambda > 0 \\ \infty & \text{if } \inf \Theta_\lambda = 0. \end{cases}$$

If we consider the ECR nested model where $\beta = \beta_0$ the PB estimate for the scale parameter λ can only be obtained using the first estimator and, consequently, the estimative always exists and is unique. Thus, we can set the PB estimate for β as solution the equation $\breve{\lambda}(\beta) = \breve{\lambda}^*(\beta)$ or, equivalently,

$$T_6(\beta,\boldsymbol{x})T_8(\beta,\boldsymbol{x}) = T_7(\beta,\boldsymbol{x})T_9(\beta,\boldsymbol{x}).$$

By replacing λ by $\breve{\lambda}(\beta)$ in (3.15), we obtain the *profile percentile-based estimation function* (ppef):

$$\mathfrak{p}_{\lambda}(\beta) = \sum_{i=1}^{n} \left[\frac{T_8(\beta,\boldsymbol{x})}{T_9(\beta,\boldsymbol{x}) \left(1 - p_i^{1/\beta}\right)} \sqrt{\left(2 - p_i^{1/\beta}\right) p_i^{1/\beta}} - x_{(i)} \right]^2 .$$

We can also obtain the PB estimates for β by minimizing one of the ppef with respect to β.

The *profile percentile-based score function* (ppsf), psf analog, is defined by

$$\mathbf{L}_{\beta|\lambda=\breve{\lambda}(\beta)}(\beta) = \frac{2}{\beta^2} \left(\frac{T_8(\beta,\boldsymbol{x})}{T_9(\beta,\boldsymbol{x})}\right) \left[\left(\frac{T_8(\beta,\boldsymbol{x})}{T_9(\beta,\boldsymbol{x})}\right) T_6(\beta,\boldsymbol{x}) - T_7(\beta,\boldsymbol{x}) \right].$$

The ppsf frequently presents numerical errors when $\beta \to 0$. To overcome this trouble consider the *critical percentile-based score function* (cpsf) given by

$$\mathbf{L}^*_{\beta|\lambda=\breve{\lambda}(\beta)}(\beta) = T_6(\beta,\boldsymbol{x})T_8(\beta,\boldsymbol{x}) - T_7(\beta,\boldsymbol{x})T_9(\beta,\boldsymbol{x}).$$

The Proposition 3.13 shows that cpsf can be used as a ppsf alternative to find the ppef critical points with the advantage that it is less sensitive to numerical errors. Obviously the PB estimate for β is a root of the cpsf. The cpsf limiting behavior is presented in Proposition 3.14.

Proposition 3.13. *β_0 is a ppsf root if and only if it is a cpsf root and*

$$\operatorname{sign}\left(\mathbf{L}_{\beta|\lambda=\breve{\lambda}(\beta)}(\beta)\right) = -\operatorname{sign}\left(\mathbf{L}^*_{\beta|\lambda=\breve{\lambda}(\beta)}(\beta)\right) \quad \forall \beta > 0.$$

Proposition 3.14. *The cpsf has the following limit behavior*

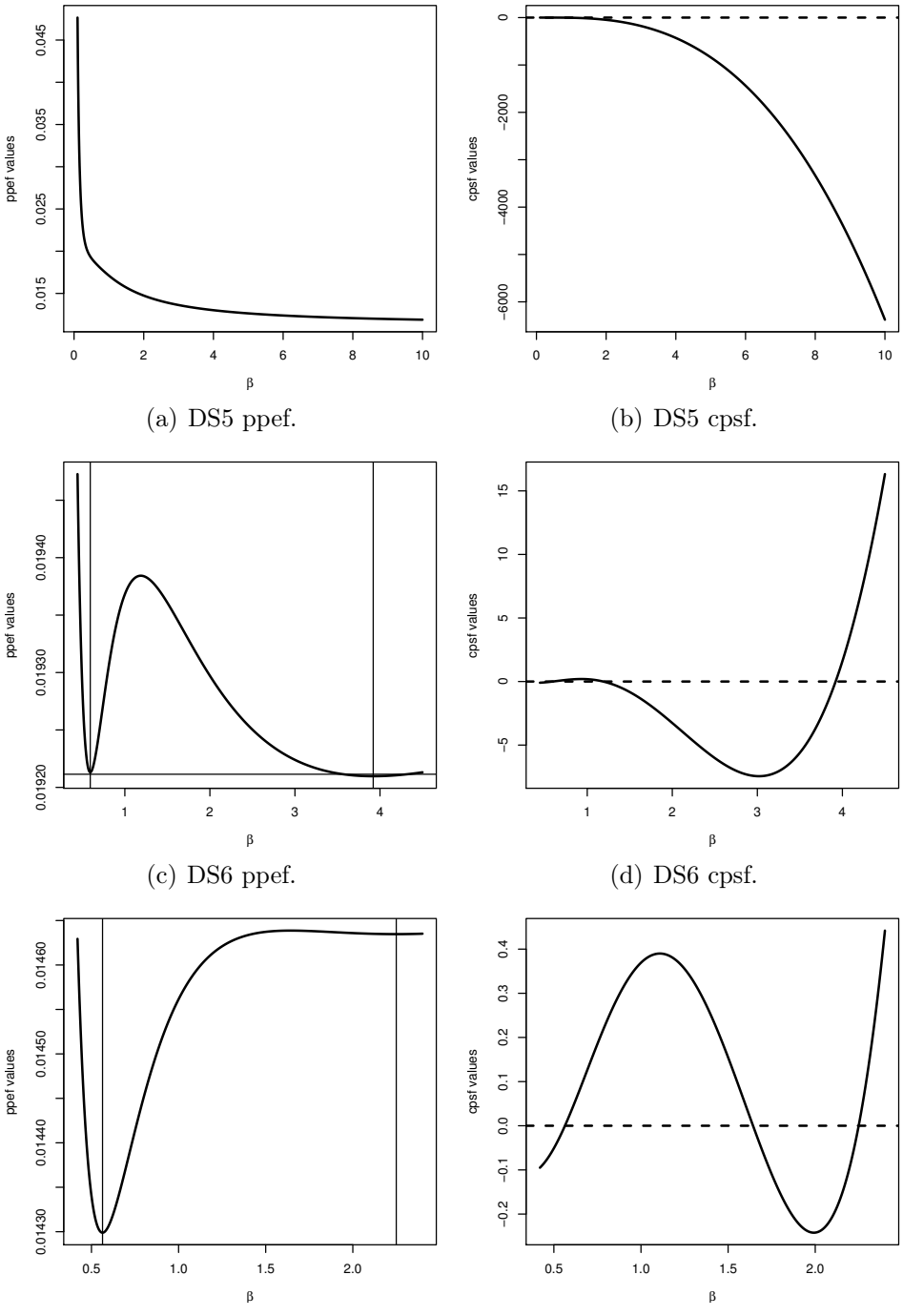
$$\lim_{\beta\to 0} \mathbf{L}^*_{\beta|\lambda=\breve{\lambda}(\beta)}(\beta) = 0$$

and

$$\lim_{\beta\to \infty} \left|\mathbf{L}^*_{\beta|\lambda=\breve{\lambda}(\beta)}(\beta)\right| = \infty.$$

Figure 3.7 depicts the ppef and cpsf of the DS5, DS6 and DS7 datasets as given Table 3.3 and shows that PBE has the same issues of the MLE. Figure 3.7(b) does not have any root and consequently Figure 3.7(a) illustrate a ppef with no global minimum. Figures 3.7(c) and 3.7(d) show that the DS6 dataset has ppef with three critical points, two of them correspond to "numerically equal" minimums. CS7 dataset present local minimums in its ppef as presented in Figures 3.7(e) and 3.7(f).

Corollary 3.14.1 shows how to find constrained PB estimates in a rectangular region of the parametric space analogous to Corollary 3.11.1. The existence of PB estimates is again assured only when $\mathfrak{L}$ is a compact set (IZMAILOV; SOLODOV, 2009, p. 8).



(a) DS5 ppef. (b) DS5 cpsf.

(c) DS6 ppef. (d) DS6 cpsf.

(e) DS7 ppef. (f) DS7 cpsf.

Figure 3.7: Profile PB estimation plots for simulated datasets in Table 3.3.

Corollary 3.14.1. *Let $\Theta_\beta, \Theta_\lambda \subset \mathbb{R}^+$, Θ_β^λ as given in Proposition 3.12, $\mathfrak{L} = \Theta_\beta \cap \Theta_\beta^\lambda$ and $\mathfrak{C}$ be the set of all cpsf roots. For a given sample $(\beta^*, \breve{\lambda}(\beta^*))$ are PB estimates, constrained to $\Theta_\beta \times \Theta_\lambda$, if*

1. $\beta^* \in \mathfrak{C}$, $\beta^* = \inf \mathfrak{L}$ *or* $\beta^* = \sup \mathfrak{L}$;
2. $\beta^* \in \mathfrak{L}$;
3. $\mathfrak{p}_\lambda(\beta^*) \leq \mathfrak{p}_\lambda(\beta)\ \forall \beta \in \mathfrak{C}$;
4. $\mathfrak{p}_\lambda(\beta^*) \leq \lim\limits_{\beta \to (\inf \mathfrak{L})^+} \mathfrak{p}_\lambda(\beta)$;
5. $\mathfrak{p}_\lambda(\beta^*) \leq \lim\limits_{\beta \to (\sup \mathfrak{L})^-} \mathfrak{p}_\lambda(\beta)$.

3.5.2 Maximum likelihood estimators for censored samples

Here we briefly discuss the ECR MLEs under left, right or interval censored datasets. Let $\mathcal{D}$ and $\mathcal{D}^*$ denote the sets of the individuals whose lifetimes are uncensored and censored respectively. Estimators under censored samples for exponentiated distributions were also studied by Gupta and Kundu (2001b) and Al-Hussaini (2010).

Type I right censoring

Type I right censoring occurs when the event of interest is observed only if it occurs prior to some prespecified time. Let T_i and L_i represent, respectively, the i^{th} individual lifetime and the fixed censoring time in a random sample of n individuals. One observes only $x_i = \min\{T_i, L_i\}$ and whether it is the lifetime or censored time. The *type I right censored* (RC1) llf is

$$\ell^{RC1}(\boldsymbol{\theta}) = m \log(\beta\lambda) + D_1(\boldsymbol{x}) + 3\, D_2(\lambda, \boldsymbol{x}) + (\beta - 1)\, D_3(\lambda, \boldsymbol{x}) + R_1(\boldsymbol{\theta}, \boldsymbol{x}),$$

where $\boldsymbol{\theta}$ is the parametric space and m denotes the number of uncensored individuals and

$$\begin{aligned} D_1(\boldsymbol{x}) &= \sum_{i \in \mathcal{D}} T_1(x_i), \\ D_2(\lambda, \boldsymbol{x}) &= \sum_{i \in \mathcal{D}} T_2(\lambda, x_i), \\ D_3(\lambda, \boldsymbol{x}) &= \sum_{i \in \mathcal{D}} T_3(\lambda, x_i) \end{aligned}$$

and

$$R_1(\boldsymbol{\theta}, \boldsymbol{x}) = \sum_{i \in \mathcal{D}^*} \log \left[1 - \left(1 - \frac{\lambda}{\sqrt{\lambda^2 + x_i^2}} \right)^\beta \right].$$

It is possible to obtain the RC1 ML estimates for $\boldsymbol{\theta}$ maximizing the RC1 llf. The RC1 lles system is

$$\begin{cases} \dfrac{m}{\beta} + D_3(\lambda,\boldsymbol{x}) - R_2(\boldsymbol{\theta},\boldsymbol{x}) & = 0, \\ \dfrac{m}{\lambda} + (1-\beta)D_4(\lambda,\boldsymbol{x}) - (\beta+2)D_5(\lambda,\boldsymbol{x}) - \beta R_3(\boldsymbol{\theta},\boldsymbol{x}) & = 0. \end{cases}$$

where

$$D_4(\lambda,\boldsymbol{x}) = \sum_{i\in\mathcal{D}} T_4(\lambda,x_i),$$
$$D_5(\lambda,\boldsymbol{x}) = \sum_{i\in\mathcal{D}} T_5(\lambda,x_i),$$
$$R_2(\boldsymbol{\theta},\boldsymbol{x}) = \sum_{i\in\mathcal{D}^*} \frac{\left(1-\frac{\lambda}{\sqrt{\lambda^2+x_i^2}}\right)^{\beta} \log\left(1-\frac{\lambda}{\sqrt{\lambda^2+x_i^2}}\right)}{1-\left(1-\frac{\lambda}{\sqrt{\lambda^2+x_i^2}}\right)^{\beta}}$$

and

$$R_3(\boldsymbol{\theta},\boldsymbol{x}) = \sum_{i\in\mathcal{D}^*} \frac{\left[\frac{\lambda^2}{\left(\lambda^2+x_i^2\right)^{3/2}} - \frac{1}{\sqrt{\lambda^2+x_i^2}}\right]\left(1-\frac{\lambda}{\sqrt{\lambda^2+x_i^2}}\right)^{\beta-1}}{1-\left(1-\frac{\lambda}{\sqrt{\lambda^2+x_i^2}}\right)^{\beta}}.$$

We can also obtain the RC1 ML estimates solving numerically the RC1 lles. This system appears does not have a (semi-)closed solution.

Type II right censoring

Type II right censoring occurs when the study continues until the event of interest is observed m times, where m is some predetermined integer ($m < n$). The *type II right censored* (RC2) llf is

$$\ell^{RC2}(\boldsymbol{\theta}) = \log\frac{n!}{(n-m)!} + m\log(\beta\lambda) + D_{(1)}(\boldsymbol{x}) + 3\,D_{(2)}(\lambda,\boldsymbol{x}) + (\beta-1)\,D_{(3)}(\lambda,\boldsymbol{x}) + R_{(1)}(\boldsymbol{\theta},\boldsymbol{x}),$$

where m denote the number of uncensored individuals and

$$D_{(1)}(\boldsymbol{x}) = T_1(x_{(1)},\ldots,x_{(m)}),$$
$$D_{(k)}(\lambda,\boldsymbol{x}) = T_k(\lambda,x_{(1)},\ldots,x_{(m)}), \quad k = 2,\ldots,5,$$

and

$$R_{(k)}(\boldsymbol{\theta},\boldsymbol{x}) = (n-m)R_k(\boldsymbol{\theta},x_{(m)}), \quad k = 1,\ldots,3.$$

The RC2 ML estimates for $\boldsymbol{\theta}$ can be obtained maximizing the RC2 llf. The RC2 lles is

$$\begin{cases} \dfrac{m}{\beta} + D_{(3)}(\lambda,\boldsymbol{x}) - R_{(2)}(\boldsymbol{\theta},\boldsymbol{x}) & = 0, \\ \dfrac{m}{\lambda} + (1-\beta)D_{(4)}(\lambda,\boldsymbol{x}) - (\beta+2)D_{(5)}(\lambda,\boldsymbol{x}) - \beta R_{(3)}(\boldsymbol{\theta},\boldsymbol{x}) & = 0. \end{cases}$$

We can also obtain the RC2 ML estimates solving numerically the RC2 lles. This system appears does not have a (semi-)closed solution.

Left censoring

Left censoring occurs when the event of interest is observed prior to a certain time t, but the exact time of occurrence is unknown. Let $x_1,\ldots,x_n$ be an observed sample obtained from the ECR distribution with m uncensored observations and $n-m$ *left censored* (LC) ones. Note that for left censoring $m > 0$. The LC llf is

$$\ell^{LC}(\boldsymbol{\theta}) = m\log(\beta\lambda) + D_1(\boldsymbol{x}) + 3\,D_2(\lambda,\boldsymbol{x}) - \ D_3(\lambda,\boldsymbol{x}) + \beta\,T_3(\lambda,\boldsymbol{x}). \tag{3.18}$$

The LC score functions is

$$U_\beta^{LC} = \frac{m}{\beta} + T_3(\lambda,\boldsymbol{x}), \tag{3.19}$$

and

$$U_\lambda^{LC} = \frac{m}{\lambda} + D_4(\lambda,\boldsymbol{x}) - 2\,D_5(\lambda,\boldsymbol{x}) - \beta\left[T_4(\lambda,\boldsymbol{x}) + T_5(\lambda,\boldsymbol{x})\right].$$

The LC ML estimates, $\hat{\boldsymbol{\theta}} = (\hat{\beta},\hat{\lambda})^\top$, can also be obtained numerically by solving the nonlinear equations system

$$U_\beta^{LC} = U_\lambda^{LC} = 0.$$

From Eq. (3.19) it is possible to obtain a semi-closed LC MLE for β. For $U_\beta^{LC} = 0$ the LC MLE for β, for a given λ, is

$$\hat{\beta}^{LC}(\lambda) = -\frac{m}{T_3(\lambda,\boldsymbol{x})}.$$

By replacing β by $\hat{\beta}^{LC}(\lambda)$ in (3.18), the following left censored pllf for λ is obtained:

$$\ell_\beta^{LC}(\lambda) = m\left[\log\left(\frac{-m\lambda}{T_3(\lambda,\boldsymbol{x})}\right) - 1\right] + D_1(\boldsymbol{x}) + 3\,D_2(\lambda,\boldsymbol{x}) - D_3(\lambda,\boldsymbol{x}).$$

We can also obtain the LC ML estimates for λ by maximizing the LC pllf $\ell_\beta^{LC}(\lambda)$ with respect to λ. The LC psf is

$$U^{LC}_{\lambda|\beta=\hat{\beta}^{LC}(\lambda)} = m\left[\frac{1}{\lambda} + \frac{T_4(\lambda,\boldsymbol{x}) + T_5(\lambda,\boldsymbol{x})}{T_3(\lambda,\boldsymbol{x})}\right] + D_4(\lambda,\boldsymbol{x}) - 2\,D_5(\lambda,\boldsymbol{x}).$$

The LC ML estimate for λ can also be defined as a root of the LC psf.

Interval censoring

Interval censoring occurs when it is only known that the event of interest occurred between times a and b. Let $x_1,\ldots,x_n$ be an observed sample obtained from the ECR distribution. Assume that we only known that $x_i \in (a_i,b_i), \forall i \in \{1,\ldots,n\}$. Then the *interval censored* (IC) llf is

$$\ell^{IC}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log\left[T_4(\lambda,b_i)^{\beta} - T_4(\lambda,a_i)^{\beta}\right].$$

We must employ some two dimensional maximization algorithm to obtain the IC ML estimates using the IC llf. The IC lle system is

$$\begin{cases} \displaystyle\sum_{i=1}^{n} \frac{\left(1-\frac{\lambda}{\sqrt{a_i^2+\lambda^2}}\right)^{\beta} \log\left(1-\frac{\lambda}{\sqrt{a_i^2+\lambda^2}}\right) - \left(1-\frac{\lambda}{\sqrt{b_i^2+\lambda^2}}\right)^{\beta} \log\left(1-\frac{\lambda}{\sqrt{b_i^2+\lambda^2}}\right)}{\left(1-\frac{\lambda}{\sqrt{a_i^2+\lambda^2}}\right)^{\beta} - \left(1-\frac{\lambda}{\sqrt{b_i^2+\lambda^2}}\right)^{\beta}} = 0, \\ \beta \displaystyle\sum_{i=1}^{n} \frac{\frac{a_i^2}{\left(a_i^2+\lambda^2\right)^{3/2}}\left(1-\frac{\lambda}{\sqrt{a_i^2+\lambda^2}}\right)^{\beta-1} - \frac{b_i^2}{\left(b_i^2+\lambda^2\right)^{3/2}}\left(1-\frac{\lambda}{\sqrt{b_i^2+\lambda^2}}\right)^{\beta-1}}{\left(1-\frac{\lambda}{\sqrt{b_i^2+\lambda^2}}\right)^{\beta} - \left(1-\frac{\lambda}{\sqrt{a_i^2+\lambda^2}}\right)^{\beta}} = 0. \end{cases}$$

This system also appears does not have a (semi-)closed solution, but we can also obtain the IC MLEs solving numerically this system.

3.5.3 The Fisher information matrix

The FIM is defined by

$$\mathbf{K} = \mathbf{K}(\boldsymbol{\theta}) = -\frac{1}{n}\begin{bmatrix} \kappa_{\beta\beta} & \kappa_{\beta\lambda} \\ \kappa_{\lambda\beta} & \kappa_{\lambda\lambda} \end{bmatrix}, \tag{3.20}$$

where the entries of $\mathbf{K}$ obey the notation defined in Section 2.3. The FIM entries are obtained in Proposition 3.15. In general, determining this matrix requires hard integrations, but we provide solutions for all in the ECR case. Mahmoud and El-Ghafour (2015) obtained the FIM for the GFP distribution and Gomes *et al.* (2015) for McBXII law it is also possible to obtain the ECR FIM from their expressions.

Proposition 3.15. *The ECR FIM elements in* (3.20) *are*

$$\kappa_{\beta\beta} = -\frac{n}{\beta^2}, \tag{3.21a}$$

$$\kappa_{\beta\lambda} = \kappa_{\lambda\beta} = \frac{n}{\lambda}\left(\frac{2}{\beta+2} - \frac{3}{\beta+1}\right) \tag{3.21b}$$

and

$$\kappa_{\lambda\lambda} = \frac{n}{\lambda^2}\left(\frac{18}{\beta+2} - \frac{36}{\beta+3} + \frac{16}{\beta+4} - 1\right). \tag{3.21c}$$

Corollary 3.15.1 assures that for any non-zero vector $v \in \mathbb{R}^2$ we have $v^\top \mathbf{K} v > 0$.

Corollary 3.15.1. *The ECR FIM is a positive definite matrix.*

Corollary 3.15.2 presents the inverse FIM, which is also useful for determining analytic expressions for the asymptotic standard errors of the MLEs and construct confidence intervals.

Corollary 3.15.2. *The inverse FIM is*

$$\mathbf{K}^{-1} = \mathbf{K}^{-1}(\boldsymbol{\theta}) = -n\begin{bmatrix} \kappa^{\beta,\beta} & \kappa^{\beta,\lambda} \\ \kappa^{\lambda,\beta} & \kappa^{\lambda,\lambda} \end{bmatrix},$$

where

$$\kappa^{\beta,\beta} = -\left(\frac{\varpi}{n}\right)\beta^3(\beta+1)\left(\beta^2+11\beta+36\right),$$
$$\kappa^{\beta,\lambda} = \kappa^{\lambda,\beta} = \left(\frac{\lambda\varpi}{n}\right)\beta^2(\beta+3)(\beta+4)^2,$$
$$\kappa^{\lambda,\lambda} = -\left(\frac{\lambda^2\varpi}{n}\right)(\beta+1)(\beta+2)(\beta+3)(\beta+4)$$

and

$$\varpi = \frac{(\beta+1)(\beta+2)}{\beta(\beta^3-7\beta^2+10\beta+72)}. \tag{3.22}$$

3.5.4 Hypothesis tests and confidence intervals

The likelihood ratio, Wald and Rao score tests are classical procedures to verify if a given sample provides evidence to reject or not a null hypothesis.

Likelihood ratio test

Suppose we need testing $H_0 : \boldsymbol{\theta} \in \Theta_0$ versus $H_1 : \boldsymbol{\theta} \in \Theta - \Theta_0$, where Θ_0 is the parametric space under the null hypothesis. The LRT statistic is

$$\boldsymbol{\Lambda}(\boldsymbol{x}) = \frac{\mathcal{L}(\hat{\boldsymbol{\theta}}_0|\boldsymbol{x})}{\mathcal{L}(\hat{\boldsymbol{\theta}}|\boldsymbol{x})},$$

where $\hat{\boldsymbol{\theta}}$ is the unconstrained MLE and $\hat{\boldsymbol{\theta}}_0$ is the constrained one. (WILKS, 1938) proved that

$$-2\log\boldsymbol{\Lambda}(\boldsymbol{x}) \xrightarrow[n\to\infty]{d} \chi^2_k,$$

where k is the difference between the number of free parameters specified by $\boldsymbol{\theta} \in \Theta$ and the number of free parameters specified by $\boldsymbol{\theta} \in \Theta_0$. In the ECR case we can express

$$\log \boldsymbol{\Lambda}(\boldsymbol{x}) = n \log \left(\frac{\hat{\beta}_0 \hat{\lambda}_0}{\hat{\beta}\hat{\lambda}} \right) + 3[T_2(\hat{\lambda}_0,\boldsymbol{x}) - T_2(\hat{\lambda},\boldsymbol{x})] + (\hat{\beta}_0 - 1)T_3(\hat{\lambda}_0,\boldsymbol{x}) - (\hat{\beta} - 1)T_3(\hat{\lambda},\boldsymbol{x}).$$

For testing $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ versus $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ the nominal level α LRT rejects if

$$-2 \log \boldsymbol{\Lambda}(\boldsymbol{x}) > c_{k,1-\alpha},$$

where $c_{k,1-\alpha}$ is the $1-\alpha$ quantile of χ^2_k.

Wald test

Among its potentialities, FIM allows to determine the asymptotic confidence interval based on the following result (LEHMANN; ROMANO, 2005, p. 510)

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow[n\to\infty]{d} \mathcal{N}(\mathbf{0},\mathbf{K}^{-1}(\boldsymbol{\theta})),$$

which implies that

$$n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^\top \mathbf{K}(\boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow[n\to\infty]{d} \chi^2_k,$$

where k is the dimension of the parameter vector $\boldsymbol{\theta}$. Thus, the $1-\alpha$ Wald confidence ellipsoid for $\boldsymbol{\theta}$ is

$$\{\boldsymbol{\theta} : n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^\top \mathbf{K}(\boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \leq c_{k,1-\alpha}\},$$

where $c_{k,1-\alpha}$ is the $1-\alpha$ quantile of χ^2_k. For testing $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ versus $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ the nominal level α Wald test rejects H_0 if

$$\mathcal{W}(\boldsymbol{x}) = n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^\top \mathbf{K}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) > c_{k,1-\alpha}.$$

In the ECR case we can express

$$\mathcal{W}(\boldsymbol{x}) = \left(\frac{\hat{\beta} - \beta_0}{\hat{\beta}} \right)^2 - \frac{2(\hat{\lambda} - \lambda_0)(\hat{\beta} - \beta_0)}{\hat{\lambda}} \left(\frac{2}{\hat{\beta} + 2} - \frac{3}{\hat{\beta} + 1} \right) - \left(\frac{\hat{\lambda} - \lambda_0}{\hat{\lambda}} \right)^2 \left(\frac{18}{\hat{\beta} + 2} - \frac{36}{\hat{\beta} + 3} + \frac{16}{\hat{\beta} + 4} - 1 \right).$$

Rao score test

As described by (BICKEL; DOKSUM, 2000, p. 399), for the simple hypothesis $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ versus $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, *Rao score test* (RST) is based on the observation that, by

the central limit theorem,

$$n^{-1/2}\mathbf{U}(\boldsymbol{\theta}_0)\xrightarrow[n\to\infty]{d}\mathcal{N}(\mathbf{0},\mathbf{K}(\boldsymbol{\theta}_0)),$$

where $\mathbf{U}(\boldsymbol{\theta}_0)$ is the likelihood score vector. The RST statistic is

$$\mathcal{R}(\boldsymbol{x}|\boldsymbol{\theta}_0) = n^{-1}\mathbf{U}(\boldsymbol{\theta}_0)^\top\mathbf{K}^{-1}(\boldsymbol{\theta}_0)\mathbf{U}(\boldsymbol{\theta}_0)\xrightarrow[n\to\infty]{d}\chi^2_k.$$

The test rejects H_0 if

$$\mathcal{R}(\boldsymbol{x}|\boldsymbol{\theta}_0) > c_{k,1-\alpha}.$$

In the ECR case we can express

$$\mathcal{R}(\boldsymbol{x}|\boldsymbol{\theta}) = \varpi\left[\beta^3(\beta+1)\left(\beta^2+11\beta+36\right)U_\beta^2 + 2\lambda\beta^2(\beta+3)(\beta+4)U_\beta U_\lambda + \lambda^2(\beta+1)(\beta+2)(\beta+3)(\beta+4)U_\lambda^2\right],$$

where ϖ is given by (3.22).

3.5.5 Cox-Snell corrected maximum likelihood estimators

Corollaries 3.15.2 and 3.15.3 and Proposition 3.16 are the fundamental results to construct the CS-MLEs described in Section 2.3. The notation adopted is the same of that section. Corollary 3.15.3 presents the first derivatives of the FIM components.

Corollary 3.15.3. *The first derivatives of the FIM components are*

$$\kappa_{\beta\beta}^{(\beta)} = \frac{2n}{\beta^3},$$

$$\kappa_{\beta\beta}^{(\lambda)} = 0,$$

$$\kappa_{\beta\lambda}^{(\beta)} = \frac{n}{\lambda}\left[\frac{3}{(\beta+1)^2} - \frac{2}{(\beta+2)^2}\right],$$

$$\kappa_{\beta\lambda}^{(\lambda)} = \frac{n}{\lambda^2}\left(\frac{3}{\beta+1} - \frac{2}{\beta+2}\right),$$

$$\kappa_{\lambda\lambda}^{(\beta)} = \frac{2n}{\lambda^2}\left[\frac{18}{(\beta+3)^2} - \frac{8}{(\beta+4)^2} - \frac{9}{(\beta+2)^2}\right]$$

and

$$\kappa_{\lambda\lambda}^{(\lambda)} = \frac{2n}{\lambda^3}\left(1 - \frac{18}{\beta+2} + \frac{36}{\beta+3} - \frac{16}{\beta+4}\right).$$

Proposition 3.16 presents the expected value of the third derivatives of the llf, which can be obtained using essentially the same procedures employed in Proposition 3.15.

Proposition 3.16. *The expected value of the third derivatives of the llf (third order*

cumulants of the llf) in (3.10) *with respect to its parameters are*

$$\kappa_{\beta\beta\beta} = \frac{2n}{\beta^3}, \tag{3.24a}$$

$$\kappa_{\beta\beta\lambda} = 0, \tag{3.24b}$$

$$\kappa_{\beta\lambda\lambda} = \frac{n}{\lambda^2}\left(\frac{9}{\beta+1} - \frac{28}{\beta+2} + \frac{27}{\beta+3} - \frac{8}{\beta+4}\right) \tag{3.24c}$$

and

$$\kappa_{\lambda\lambda\lambda} = \frac{2n}{\lambda^3}\left(1 - \frac{81}{\beta+2} + \frac{378}{\beta+3} - \frac{606}{\beta+4} + \frac{405}{\beta+5} - \frac{96}{\beta+6}\right). \tag{3.24d}$$

The second order biases of the MLEs are listed in Proposition 3.17. As expressed in (2.15), they are the key to obtain the CS-MLEs.

Proposition 3.17. *Let X be an ECR random variable with vector parameter $\boldsymbol{\theta} = (\beta,\lambda)^\top$. The second order biases of the ECR MLEs evaluated in $\hat{\boldsymbol{\theta}} = (\hat{\beta},\hat{\lambda})^\top$ are*

$$\widehat{\mathrm{Bias}}(\hat{\beta}) = \frac{1}{n}\left[\hat{\beta}^3 + 13\hat{\beta}^2 + 122\hat{\beta} + 380 - \frac{699840}{19321(\hat{\beta}+5)} + \frac{96000}{361(\hat{\beta}+6)} + \frac{432\left(4085783\hat{\beta}^2 - 8192586\hat{\beta} - 40352456\right)}{2641\left(\hat{\beta}^3 - 7\hat{\beta}^2 + 10\hat{\beta} + 72\right)^2} - \frac{12\left(70740551\hat{\beta}^2 + 3809213278\hat{\beta} - 35831044156\right)}{6974881\left(\hat{\beta}^3 - 7\hat{\beta}^2 + 10\hat{\beta} + 72\right)}\right]$$

and

$$\widehat{\mathrm{Bias}}(\hat{\lambda}) = \frac{\hat{\lambda}}{n}\left[8\hat{\beta} + 86 + \frac{49}{270\hat{\beta}} - \frac{1679616}{96605(\hat{\beta}+5)} + \frac{80000}{1083(\hat{\beta}+6)} - \frac{8\left(84037561\hat{\beta}^2 + 21509105\hat{\beta} - 393761162\right)}{7923\left(\hat{\beta}^3 - 7\hat{\beta}^2 + 10\hat{\beta} + 72\right)^2} + \frac{356431397749\hat{\beta}^2 - 158970444943\hat{\beta} - 4636191041858}{376643574\left(\hat{\beta}^3 - 7\hat{\beta}^2 + 10\hat{\beta} + 72\right)}\right].$$

Unfortunately, for certain sample sizes and $\hat{\beta}$ values, the ECR CS-MLE outcomes can be outside of the parametric space. This problem also occurred in some models like inverse Weibull (MAZUCHELI *et al.*, 2018a). Thus, it is interesting to outline a map that identifies correctable regions, on which an constrained optimization would be acceptable. Figure 3.8 displays portions of the ECR CS-correctable region in terms of sample size and $\hat{\beta}$. It is noticeable that this region does not depend of $\hat{\lambda}$, as indicated by the $\widehat{\mathrm{Bias}}(\hat{\beta})$.

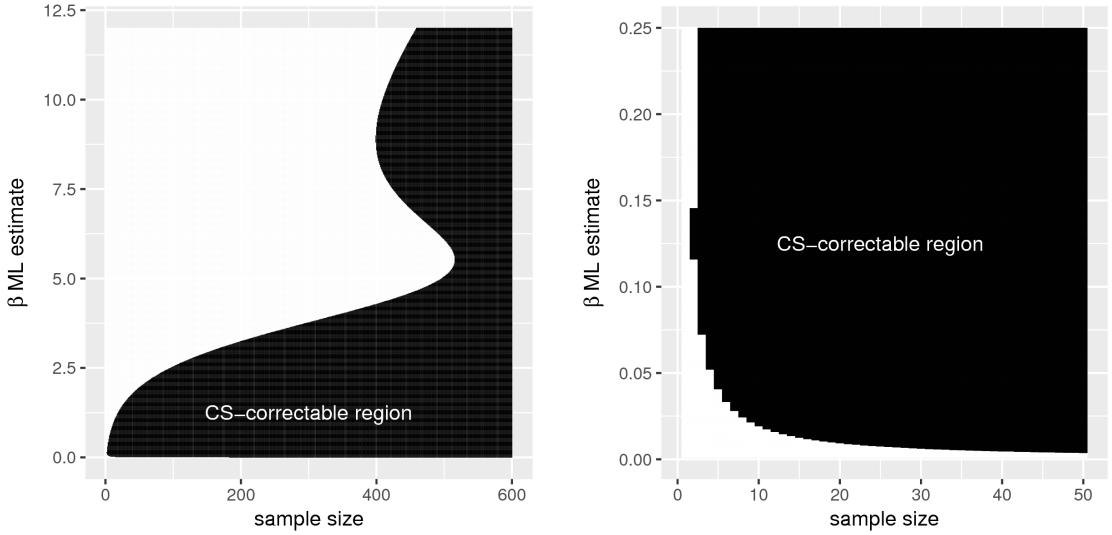


Figure 3.8: CS-correctable region for ECR ML estimates.

The ECR CS-MLEs are obtained using the MLEs second order biases expressed in (2.15). The ECR CS-MLEs when β or λ are known can be obtained by Proposition 3.18:

Proposition 3.18. *Let an ECR random variable with parameters*

1. *β and $\lambda = \lambda_0$. The second order bias of the β MLE evaluated in $\hat{\beta}$ is*

$$\widehat{\text{Bias}}(\hat{\beta}|\lambda = \lambda_0) = \frac{\hat{\beta}}{n}.$$

2. *$\beta = \beta_0$ and λ. The second order bias of the λ MLE evaluated in $\hat{\lambda}$ is*

$$\widehat{\text{Bias}}(\hat{\lambda}|\beta = \beta_0) = \frac{\hat{\lambda}}{n}\frac{(\beta_0+2)(\beta_0+3)(\beta_0+4)}{\beta_0(\beta_0+5)(\beta_0+6)}\left[\frac{\beta_0^4+24\beta_0^3+216\beta_0^2+761\beta_0+294}{(\beta_0^2+11\beta_0+36)^2}\right].$$

The CR CS-MLE for λ is indicated in Corollary 3.18.1.

Corollary 3.18.1. *Let a CR random variable with parameter λ. The second order bias of the λ MLE evaluated in $\hat{\lambda}$ is $\widehat{\text{Bias}}_{CR}(\hat{\lambda}) = {}^{45\hat{\lambda}}/_{56n}$.*

3.6 SIMULATION STUDIES

In this section, we assess the performance of the proposed estimators by means of three simulation studies, we investigate the convergence of the bias-corrected MLEs. In particular, the asymptotic behavior of MLE ($\hat{\beta}$,$\hat{\lambda}$), CS-MLE ($\tilde{\beta}$,$\tilde{\lambda}$) and PBE ($\breve{\beta}$,$\breve{\lambda}$) is quantified adopting relative biases and SSDs as comparison criteria. The final study is focused on the issue of existence and non-existence of the ML and PB estimates for given samples. To that end, we used the platform described in Section 1.2.

3.6.1 Convergence study

The first study analyzed the convergence of the estimates as the sample size increases. Define the following sets

$$\mathfrak{N} = \{5, 10, \ldots, 250\}$$

and

$$\mathfrak{P} = \{(0.5,0.3),(0.4,0.15),(1,0.25),(0.35,2.5),(0.47,0.6),(1,0.5),(5,5),(8,8),(5,0.1),(10,0.1)\}.$$

For each element $(\boldsymbol{\theta},n) \in \mathfrak{N} \times \mathfrak{P}$, it was generated 1,000 ECR distributed samples with parameters $\boldsymbol{\theta} = (\beta,\lambda)$ and size n using the ITSM. For each Monte Carlo replica, it was obtained both bias and SSD and, subsequently, the Monte Carlo averages are calculated for each $(\boldsymbol{\theta},n)$.

Figures A.1–A.11 exhibits bias and SSD values for several sample sizes. It is noticeable the $\hat{\beta}$ bias outperforms that due to $\breve{\beta}$ for smaller size samples. Under CS-correctable region, $\tilde{\beta}$ presents the smallest bias for the majority of cases. With respect to SSD, $\breve{\beta}$ tends to assume the smallest value at small samples; while, $\tilde{\beta}$ has the best asymptotic behavior.

In general, the $\hat{\lambda}$ bias is the smallest, while the $\breve{\lambda}$ one is the highest. In small values of β, the $\hat{\lambda}$ and $\tilde{\lambda}$ SSDs present similar values, while $\breve{\lambda}$ is the highest one.

3.6.2 Relative bias and sample standard deviation studies

This study investigated the convergence of the relative biases and SSDs of the parameter estimates in different regions of the parametric space. Let these sets

$$\mathfrak{N} = \{10,100\}$$

and

$$\mathfrak{P} = \{0.1, 0.2, \ldots, 6\}.$$

For each element $(\beta,\lambda,n) \in \mathfrak{P}^2 \times \mathfrak{N}$ it was generated 1,000 ECR distributed samples with parameters (β,λ) and size n using the ITSM. For each random sample, we calculated both bias and SSD. Finally, we compute the average of biases and SSDs for each element (β,λ,n). Results are depicted in Figures A.12–A.15.

In Figures A.12 and A.13 biases and SSDs do not appear to suffer influence of λ value, that is, biases and SSDs seem to be only function of β values. Figures also reveal that CS-MLEs are desirable for only small values of β.

In Figures A.14 and A.15, we can observe that the $\hat{\beta}$ relative bias and SSD decrease for when β increases, while the relative bias and SSD of $\tilde{\beta}$ and $\breve{\beta}$ do not have this property. Small values of β, the smallest relative biases and SSDs for β estimates are obtained by $\tilde{\beta}$

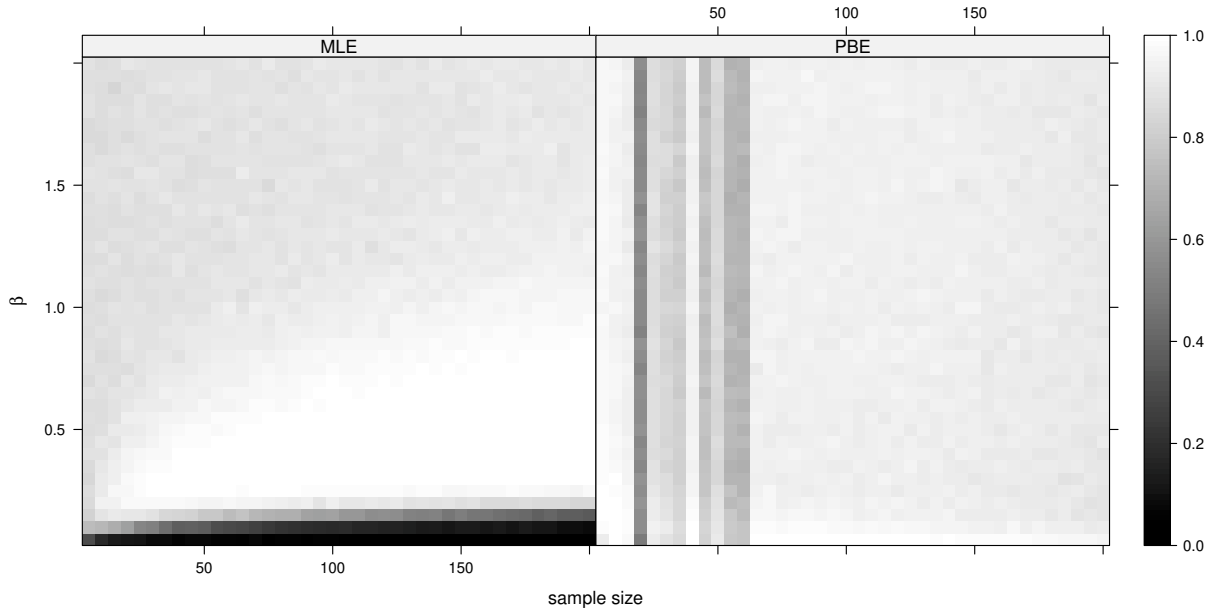


Figure 3.9: Existence rates of the estimate of the standard ECR shape parameter.

and $\breve{\beta}$. The smallest biases and SSDs for λ estimates are, typically, attained by $\hat{\lambda}$ and $\tilde{\lambda}$. In particular, for smallest values of β, the $\tilde{\lambda}$ has the smallest biases and SSDs.

3.6.3 Existence of estimates study

The last simulation study evaluated the proportion of non-existence of the ML and PB estimates in different points of the parametric space. Consider the sets

$$\mathfrak{N} = \{5,10,\ldots,200\}$$

and

$$\mathfrak{P} = \{0.05,0.10,\ldots,2\}.$$

For each element $(\beta,n) \in \mathfrak{P} \times \mathfrak{N}$, we generated 10,000 standard ECR distributed samples and the proportion of them that had ML or PB estimates was registered.

Figure 3.9 summarizes the evaluated existence rates of estimates. Figure 3.9 reveals that MLE presents its smallest existence rates when $\beta < 0.2$. When $\beta > 0.2$ the existence rates build up as n increases. The region where $0.3 < \beta < 1$ appears to be the highest existence rates. Figure 3.9 unveils a different behavior for the PBE, the existence rates look to be influenced only by the sample size. The lowest existence rates of the PB estimates are found when $25 < n < 60$.

3.7 APPLICATION

An application to real data is done to illustrate the ECR potentiality. To that end, we use the uncensored real datasets previously discussed by Crowley and Hu (1977) and

Table 3.4: Crowley and Hu's dataset.

1	1	2	2	2	4	4	5	5	7	8
11	15	15	15	16	17	20	20	27	29	31
34	35	36	38	39	42	44	49	50	52	57
60	65	67	67	68	71	71	76	77	79	80
84	89	95	99	101	109	148	152	187	206	218
262	284	284	307	333	339	674	732	851	1031	1386

Table 3.5: Descriptive statistics.

Dataset	Mean	$\overline{\mathcal{Q}}_{1/4}$	$\overline{\mathcal{Q}}_{1/2}$	$\overline{\mathcal{Q}}_{3/4}$	Variance	Skew.	Kurt.	Min.	Max.
Crowley and Hu	143.70	16.92	58.50	112.25	64506.42	3.104648	13.00242	1	1386

which consist of 66 patient survival days with respect to a Stanford heart transplant list. We consider only patients that died in the followup time and were not submitted to prior bypass surgery. The data are presented in Table 3.4.

Let $\overline{\mathcal{Q}}_i$ be the $i\%$ percentile of a sample obtained by means of the methodology recommended by Hyndman and Fan (1996). Table 3.5 gives a descriptive summary, which anticipates right-skewed data.

In many applications empirical hrf shape can indicate a particular model. The plot of *total time on test* (TTT) (AARSET, 1987) can be useful in this sense. The TTT plot is obtained by measuring $G(r/n) = (\sum_{i=1}^{r} y_{(i)} + (n-r)y_{r:n}) / \sum_{i=1}^{n} y_{(i)}$ versus r/n for $r = 1, \ldots, n$. The under study TTT plots are presented in Figure 3.10. Results indicate a upside-down bathtub-shaped hrf, pattern which is covered by the ECR hrf.

Beyond, Figure 3.11 presents the pllf and ppef curves in order to anticipate the ECR optimization framework at under study data. The concavity of the pllf curve in Figure 3.11(a) indicates a global maximum around $\lambda_0 = 80.5$ and the convexity of the ppef curve in Figure 3.11(b) indicates a global minimum around $\beta_0 = 0.175$. We employed these values as start points to obtain the ML and PB estimates using the Newton-Raphson method.

We compare the fits of the ECR distribution defined in (2.7) with the CR pdf and some other models namely *generalized gamma* (GG) (PRENTICE, 1974), LN, BS (BIRNBAUM; SAUNDERS, 1969), *log-logistic* (LL)[5], Weibull, Fréchet[6], EE, *inverse gamma* (IG)[7], *generalized half-normal* (GHN) (COORAY; ANANDA, 2008), *exponentiated Lindley* (ELi), *Lindley exponential* (LE), Chen (CHEN, 2000), Wald[8], *flexible Weibull* (FW) (BEBBINGTON *et al.*, 2007), BX (SURLES; PADGETT, 2005) and *log-Cauchy* (LCa). These models have sample space supported on $\mathbb{R}^+$.

[5] a.k.a. Fisk model (FISK, 1961).
[6] a.k.a. inverse Weibull (BURY, 1999).
[7] a.k.a. Vinci model (VINCI, 1921).
[8] a.k.a. inverse normal model (MICHAEL *et al.*, 1976).

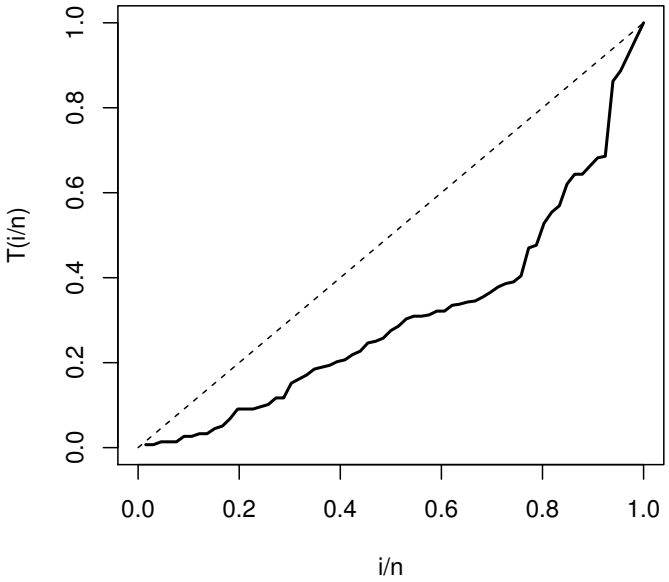
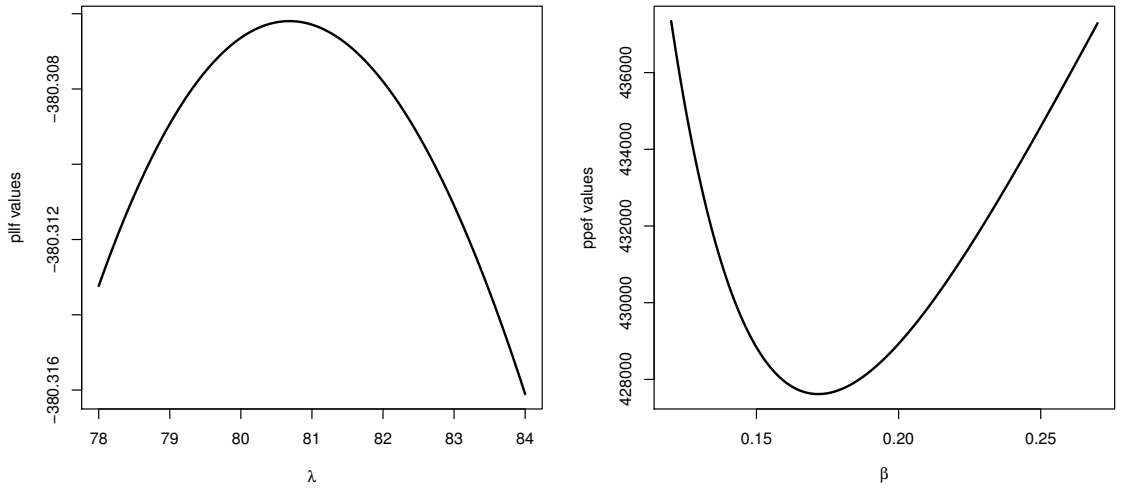


Figure 3.10: TTT-plot of the Crowley and Hu's dataset.



(a) pllf. (b) ppef.

Figure 3.11: ECR estimation functions.

- The GG model with shape parameters μ,$q \in \mathbb{R}$ and scale parameter $\sigma > 0$, can be written $y = \mu + \sigma x$ where the error pdf is written

$$f_{GG}(y) = \begin{cases} |q| \frac{(q^{-2})^{q^{-2}}}{\Gamma(q^{-2})} \exp\left\{q^{-2}\left[qy - \exp(qy)\right]\right\} & \text{se } q \neq 0, \\ \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y^2\right) & \text{se } q = 0. \end{cases}$$

  Some special cases for $t = \exp(y)$ are Weibull($q = 1$), *exponential* (Exp)($q = \sigma = 1$), log-normal($q = 0$) and gamma($\sigma = 1$,$q > 0$).

- The LN model with shape parameters $\mu > 0$ and $\sigma > 0$ has pdf (for $x > 0$) given by

$$f_{LN}(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\frac{\left[\log(x) - \mu\right]^2}{\sigma^2}\right\}.$$

- The BS model with shape parameter $\alpha > 0$ and scale parameter $\beta > 0$ has cdf (for $x > 0$) given by

$$F_{BS}(x) = \Phi\left(\frac{1}{\alpha}\left[\left(\frac{x}{\beta}\right)^{0.5} - \left(\frac{\beta}{x}\right)^{0.5}\right]\right),$$

  where Φ(x) is the cdf of the standard normal distribution defined as

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{t^2}{2}\right) \mathrm{d}t.$$

- The LL model with shape parameter $\eta > 0$ and scale parameter $\varphi > 0$ has pdf (for $x > 0$) given by

$$f_{LL}(x) = \frac{\eta}{\varphi}\left(\frac{x}{\varphi}\right)^{\eta-1}\left[1 + \left(\frac{x}{\varphi}\right)^{\eta}\right]^{-2}.$$

- The Weibull model with shape parameter $p > 0$ and scale parameter $b > 0$ has pdf (for $x > 0$) given by

$$f_{Weibull}(x) = \frac{a}{\beta}\left(\frac{x}{\beta}\right)^{a-1} \exp\left[-\left(\frac{x}{\beta}\right)^{a}\right].$$

- The Fréchet model with shape parameter $\lambda > 0$ and scale parameter $\sigma > 0$ has pdf (for $x > 0$) given by

$$f_{Frechet}(x) = \frac{\lambda}{\sigma}\left(\frac{\sigma}{x}\right)^{\lambda+1} \exp\left[-\left(\frac{\sigma}{x}\right)^{\lambda}\right].$$

- The EE model with shape parameter $\alpha > 0$ and scale parameter $\lambda > 0$ has pdf (for

$x > 0$) given by

$$f_{EE}(x) = \alpha\lambda \, \mathrm{e}^{-\lambda x} \left(1 - \mathrm{e}^{-\lambda x}\right)^{1-\alpha}.$$

- The gamma model with shape parameter $p > 0$ and scale parameter $b > 0$ has pdf (for $x > 0$) given by

$$f_{gamma}(x) = \frac{1}{b\Gamma(p)} \left(\frac{x}{b}\right)^{p-1} \exp\left(-\frac{x}{b}\right).$$

- The IG model with shape parameter $\alpha > 0$ and scale parameter $\beta > 0$ has pdf (for $x > 0$) given by

$$f_{IG}(x) = \frac{\beta}{\Gamma(\alpha)} (\beta x)^{-\alpha-1} \exp\left(-\frac{1}{\beta x}\right).$$

- The GHN model with shape parameter $\alpha > 0$ and scale parameter $\theta > 0$ has pdf (for $x > 0$) given by

$$f_{GHN}(x) = \sqrt{\frac{2}{\pi}} \left(\frac{\alpha}{x}\right) \left(\frac{x}{\theta}\right)^{\alpha} \exp\left[-\frac{1}{2}\left(\frac{x}{\theta}\right)^{2\alpha}\right].$$

- The ELi model with shape parameters $\alpha > 0$ and scale parameter $\lambda > 0$ has pdf (for $x > 0$) given by

$$f_{ELi}(x) = \frac{\alpha\lambda^2}{1+\lambda}(1+x) \, \mathrm{e}^{-\lambda x} \left[1 - \frac{1+\lambda+\lambda x}{1+\lambda} \, \mathrm{e}^{-\lambda x}\right]^{\alpha-1}.$$

- The LE model with shape parameters $\theta > 0$ and scale parameter $\lambda > 0$ has pdf (for $x > 0$) given by

$$f_{LE}(x) = \frac{\theta^2\lambda}{1+\theta} \left[1 - \mathrm{e}^{-\lambda x}\right]^{\theta-1} \left[1 - \log\left(1 - \mathrm{e}^{-\lambda x}\right)\right].$$

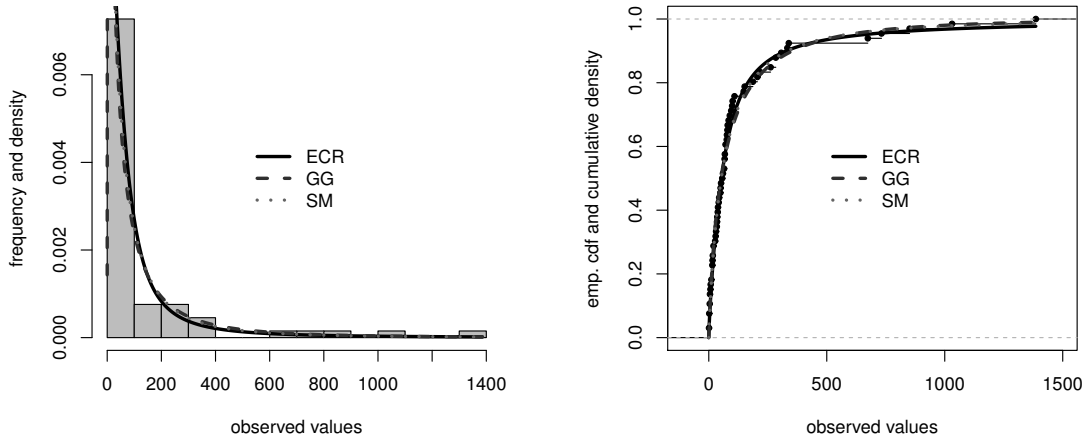
- The Chen model with shape parameters $\kappa \geq 0$ and $\theta > 0$ has pdf (for $x > 0$) given by

$$f_{Chen}(x) = \theta\kappa x^{\kappa-1} \exp\left[\theta\left(1 - \mathrm{e}^{x^{\kappa}}\right) + x^{\kappa}\right].$$

- The Wald model with shape parameter $\mu \geq 0$ and scale parameter $\lambda > 0$ has pdf (for $x > 0$) given by

$$f_{Wald}(x) = \sqrt{\frac{\lambda}{2\pi x^3}} \exp\left[\frac{-\lambda(x-\mu)^2}{2\mu^2 x}\right].$$

- The FW model with shape parameter $\alpha > 0$ and scale parameter $\beta > 0$ has cdf (for



(a) pdfs vs. histograms.

(b) cdfs vs. ecdfs.

Figure 3.12: Some of the best fitted models.

$x \geq 0$) given by

$$f_{FW}(x) = \left(\alpha + \frac{\beta}{x^2}\right) \exp\left(\alpha x - \frac{\beta}{x}\right) \exp\left[-\exp\left(\alpha x - \frac{\beta}{x}\right)\right].$$

- The BX model with shape parameter $\alpha \geq 0$ and scale parameter $\lambda > 0$ has pdf (for $x > 0$) given by

$$f_{BX}(x) = 2\alpha\lambda^2 x \exp[-(\lambda x)^2]\{1 - \exp[-(\lambda x)^2)]^{\alpha-1}\}.$$

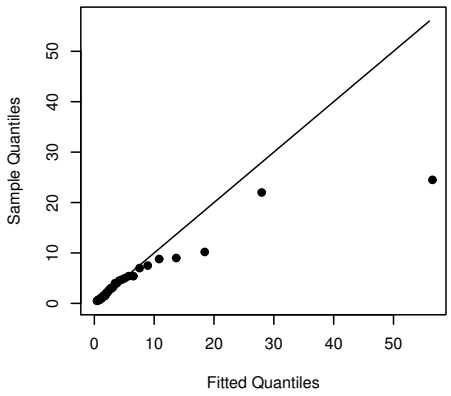
- The LCa model with shape parameter $\mu \in \mathbb{R}$ and scale parameter $\sigma > 0$ has pdf (for $x > 0$) given by

$$f_{LCa}(x) = \frac{1}{x\pi} \frac{\sigma}{(\log x - \mu)^2 + \sigma^2}.$$

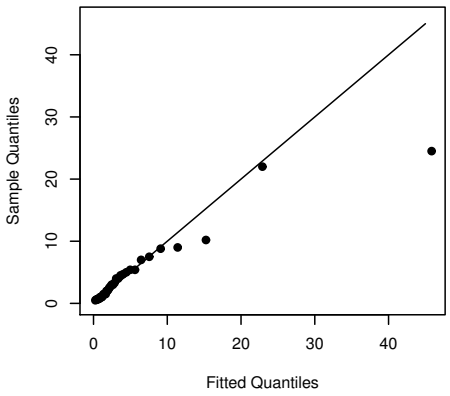
Figure 3.12 displays fitted and empirical densities for the studied dataset. For better exhibition, we regard only models which are to provide closer fits to the histogram. The empirical and fitted cdfs of these models are displayed in Figure 3.12. The qq-plots for some fitted models can be seen in Figure 3.13. From these plots, we note that the ECR model provides a good fit.

Table 3.6 presents the ML estimates and their std. errors and asymptotic confidence intervals for fitted models, indicating there is not non-significant fits from their respective asymptotic confidence intervals. The PB and CS-MLE are also presented. Table 3.6 present the ML estimates and their std. errors and asymptotic confidence intervals for fitted models, indicating there is not non-significant fits from their respective asymptotic confidence intervals.

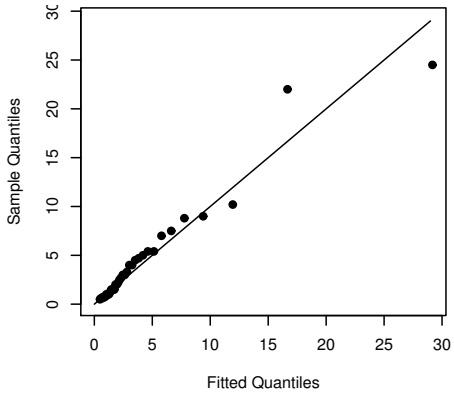
To compare quantitatively discussed models, Table 3.7 elects the GoF statistics of the fitted models. We use the following GoF statistics:



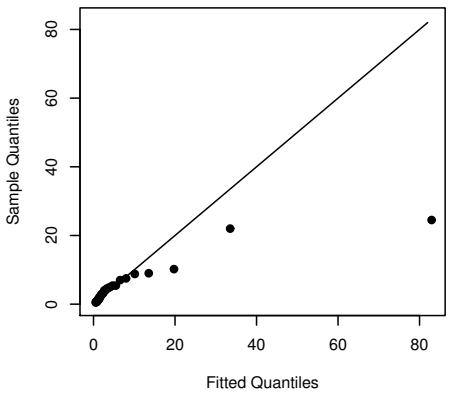
(a) ECR.



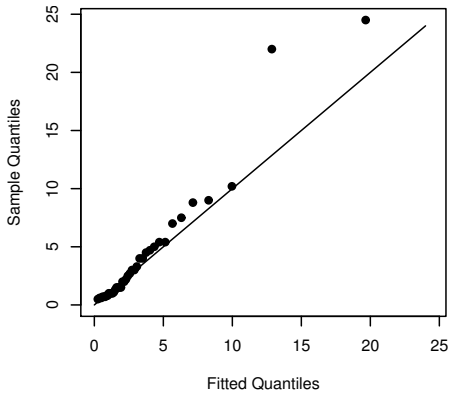
(b) CR.



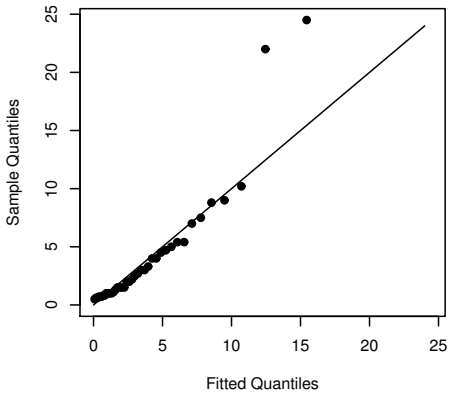
(c) GG.



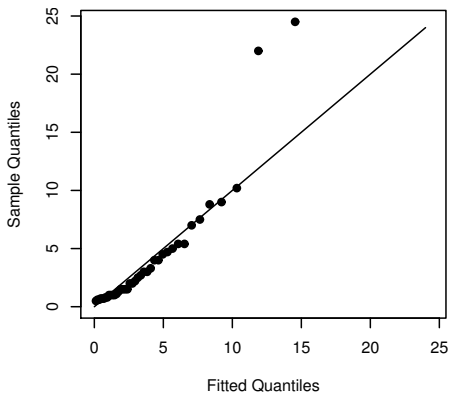
(d) SM.



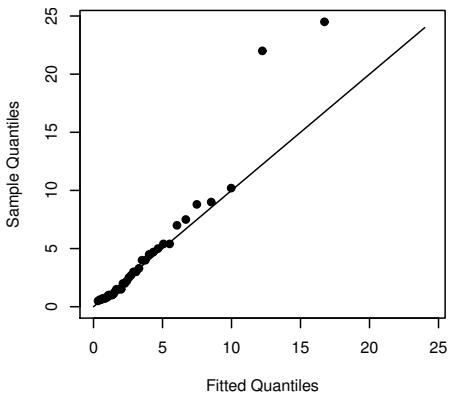
(e) LL.



(f) Weibull.



(g) gamma.



(h) LN.

Figure 3.13: qq-plots of Crowley and Hu's dataset for some models.

Table 3.6: Estimates for Crowley and Hu's dataset.

model	estimates (std. err.)		
GG(μ,σ,q)	4.1496 (0.3233)	1.5892 (0.1481)	0.3690 (0.1536)
	(3.5160,4.7832)	(1.2988,1.8795)	(0.0687,0.6707)
SM(c,k,λ)	1.8953 (0.0565)	0.9110 (0.2660)	129.9019 (10.5690)
	(1.7845,2.0060)	(0.3879,1.4323)	(109.1868,150.6171)
ECR(β,λ)^{ML}	0.3867 (0.0400)	80.6840 (11.7096)	
	(0.3082,0.4651)	(68.9744,92.3936)	
ECR(β,λ)^{CS}	0.3652 (0.0389)	80.2282 (10.8532)	
	(-,-)	(-,-)	
LL(a,b)	1.0806 (0.1122)	50.3379 (4.1945)	
	(0.8607,1.3006)	(42.1168,58.5590)	
LN(μ,σ)	3.8491 (0.2021)	1.6429 (0.1435)	
	(3.4529,4.2453)	(1.3615,1.9242)	
Weibull(a,β)	0.6692 (0.0585)	104.1081 (4.2048)	
	(0.5546,0.7839)	(95.8666,112.3496)	
LE(θ,λ)	0.7915 (0.0972)	0.0037 (0.0015)	
	(0.6010,0.9820)	(0.0007,0.0067)	
GHN(α,θ)	0.4964 (0.0421)	142.7885 (4.2051)	
	(0.4138,0.5790)	(134.5464,151.0300)	
Fréchet(λ,σ)	0.5846 (0.0511)	20.1116 (4.4516)	
	(0.4844,0.6847)	(11.3865,28.8367)	
Chen(β,λ)	0.2104 (0.0122)	0.0687 (0.0151)	
	(0.1864,0.2344)	(0.03903,0.0983)	
BS(α,β)	2.2664 (0.1982)	37.6227 (4.2131)	
	(1.8779,2.6548)	(29.3649,45.8804)	
gamma(p,b)	0.5583 (0.0410)	257.4058 (40.4295)	
	(0.4779,0.6387)	(178.1640,336.6470)	
EE(α,λ)	0.5503 (0.0822)	0.0045 (0.0011)	
	(0.3891,0.7114)	(0.0023,0.0067)	
ELi(α,λ)	0.2740 (0.0421)	0.0058 (0.0011)	
	(0.1914,0.3566)	(0.0037,0.0008)	
IG(α,β)	0.4539 (0.0661)	5.1499 (1.2391)	
	(0.3243,0.5836)	(2.7212,7.5786)	
ECR(β,λ)^{PB}	0.1721 (0.0545)	147.4861 (10.2357)	
	(-,-)	(-,-)	
BX(β,λ)	0.1960 (0.0261)	0.0017 (0.0002)	
	(0.1448,0.2473)	(0.0013,0.0022)	
Wald(μ,λ)	143.7060 (4.8431)	12.3159 (2.4221)	
	(134.2134,153.1986)	(7.5685,17.0633)	
FW(α,β)	0.0015 (0.0006)	8.3432 (1.4112)	
	(0.0003,0.0025)	(5.5773,11.1093)	
BXII(c,k)	0.0153 (0.0012)	16.9950 (2.5626)	
	(0.0128,0.0177)	(11.9723,22.0178)	
CR(λ)	24.4910 (0.4479)		
	(23.6131,25.3689)		

Table 3.7: Goodness-of-fit tests for Crowley and Hu's dataset.

model	CvM	AD	KS	AIC	CAIC	BIC	HQIC
GG(μ,σ,q)	0.069	0.399	0.079	765.532	765.919	772.101	768.127
SM(c,k,λ)	0.052	0.327	0.072	766.012	766.399	772.581	768.608
ECR(β,λ)^{ML}	0.039	0.286	**0.057**	**764.612**	**764.803**	**768.992**	**766.343**
ECR(β,λ)^{CS}	**0.038**	**0.285**	0.063	764.612	764.803	768.992	766.343
LL(a,b)	0.074	0.454	0.063	765.481	765.672	769.861	767.212
LN(μ,σ)	0.102	0.579	0.089	764.915	765.105	769.294	766.645
Weibull(a,β)	0.114	0.689	0.118	767.444	767.635	771.824	769.175
LE(θ,λ)	0.140	0.838	0.139	768.735	768.926	773.115	770.466
GHN(α,θ)	0.191	1.150	0.142	773.212	773.403	777.592	774.943
Fréchet(λ,σ)	0.356	2.036	0.146	780.374	780.564	784.753	782.104
Chen(β,λ)	0.295	1.766	0.151	782.206	782.397	786.585	783.937
BS(α,β)	0.229	1.229	0.155	770.982	771.173	775.362	772.713
gamma(p,b)	0.195	1.172	0.159	772.658	772.849	777.037	774.389
EE(α,λ)	0.210	1.261	0.168	773.709	773.900	778.088	775.440
ELi(α,λ)	0.258	1.562	0.185	778.603	778.793	782.982	780.333
IG(α,β)	0.563	3.183	0.210	792.664	792.854	797.043	794.394
ECR(β,λ)^{PB}	0.069	0.445	0.223	783.887	784.077	788.266	785.617
BX(β,λ)	0.386	2.274	0.238	788.552	788.742	792.931	790.282
Wald(μ,λ)	0.425	2.383	0.261	787.709	787.899	792.088	789.439
FW(α,β)	0.931	4.691	0.262	816.638	816.829	821.018	818.369
BXII(c,k)	0.676	3.828	0.323	824.816	825.007	829.196	826.547
CR(λ)	0.213	1.254	0.132	785.023	785.085	787.212	785.888

- Criteria under cdfs:
  - *Cramér-von Mises* (CvM) (CRAMÉR, 1928);
  - *Anderson-Darling* (AD) (ANDERSON; DARLING, 1952);
  - *Kolmogorov-Smirnov* (KS) (SMIRNOV, 1948).
- Criteria under pdfs:
  - *Akaike information criterion* (AIC) (AKAIKE, 1974);
  - *bayesian information criterion* (BIC) (SCHWARZ, 1978);
  - *consistent Akaike information criterion* (CAIC) (HURVICH; TSAI, 1989);
  - *Hannan-Quinn information criterion* (HQIC) (HANNAN; QUINN, 1979).

Smaller GoF values are associated with better fits. The GoFs under pdf class may indicate superiority relations in nested models while the criteria under cdf can be used to compare nested and non-nested models.

Table 3.7 reveals that the new model presents smallest cdf based GoF values. To test $\mathcal{H}_0 : \beta = 1$ (or $F_{\text{CR}} = F_{\text{ECR}}$), we employ the likelihood ratio statistic, which yields a

Table 3.8: Adjusted cdfs of the studied models evaluated in the sample quartiles.

model	$F\left(\overline{\mathcal{Q}}_{1/4}\right)$	$F\left(\overline{\mathcal{Q}}_{1/2}\right)$	$F\left(\overline{\mathcal{Q}}_{3/4}\right)$
GG(μ,σ,q)	0.253	0.529	0.688
SM(c,k,λ)	0.240	0.526	0.696
ECR(β,λ)^{ML}	0.226	0.527	**0.713**
ECR(β,λ)^{CS}	0.246	0.547	**0.728**
LL(a,b)	0.235	0.541	0.704
LN(μ,σ)	0.267	0.553	0.702
Weibull(a,β)	0.257	**0.493**	0.651
LE(θ,λ)	0.244	0.472	0.629
GHN(α,θ)	0.271	0.479	0.625
Fréchet(λ,σ)	0.331	0.585	0.694
Chen(β,λ)	0.297	0.480	0.615
BS(α,β)	0.359	0.578	0.694
gamma(p,b)	0.240	0.454	0.610
EE(α,λ)	0.237	0.447	0.601
ELi(α,λ)	0.234	0.434	0.585
IG(α,β)	0.399	0.635	0.705
ECR(β,λ)^{PB}	0.420	0.633	**0.761**
BX(β,λ)	**0.249**	0.404	0.521
Wald(μ,λ)	0.428	0.701	0.801
FW(α,β)	0.465	0.612	0.667
BXII(c,k)	0.521	0.653	0.707
CR(λ)	0.177	0.614	**0.787**

p-value $< 10^{-5}$, indicating there is statistical difference between the ECR and CR models. In summary, the ECR model may be a good alternative for describing heavy-tailed lifetime (or positive real) data.

Table 3.8 shows the adjusted cdfs of the models previously listed evaluated in the sample quartiles. In general, the CR and ECR distributions, under ML estimates, present cdf values for the third quartile closer to its expected value of 3/4 in comparison to the other models. The ECR PB estimate present the best approximation to the third quartile, but this corresponding approximations to the first and second quartiles are poor. The ECR CS-ML estimates obtained better approximations for the first and third quartiles than ECR ML estimates. Thus, beside all mathematical and inferential we have presented and discussed, these results reveal the ECR distribution may have better capacity to fit tail probabilities than the remainder ones.

3.8 CONCLUDING REMARKS

We deepened a discussion about an extended Cauchy-Rayleigh distribution, particular case of some known models like heavy-tailed Rayleigh (NIKIAS; SHAO, 1995), gener-

alized Feller-Pareto (ZANDONATTI, 2001), exponentiated Burr type-XII (AL-HUSSAINI; AHSANULLAH, 2015), Kumaraswamy Burr type-XII (PARANAÍBA, 2012) and McDonald Burr type-XII (GOMES *et al.*, 2015). We referred to it as *exponentiated Cauchy-Rayleigh* (ECR) distribution. Some of its mathematical properties were derived and discussed: Closed-form expressions for mode and probability weighted, log-, incomplete and order statistic moments. The ECR model obeyed the property of regularly varying at infinity and can take decreasing, decreasing-increasing-decreasing and upside-down bathtub-shape hazard rate functions, confirming its usefulness to describe lifetime data. We provided procedures to estimate the ECR parameters through original and bias-corrected maximum likelihood estimators and a percentile-based method. An important characteristic of this new model was that the mode, median, moments, Fisher information components, its inverse components and the second order bias can be decomposed in two factors where the first factor is function of the scale parameter λ and the second one is function of the new shape parameter β. A simulation study to assess proposed estimators was performed. Each procedure revealed advantages over specific parameter points and sample sizes. The ECR employment was illustrated through an application to real data. Results indicated that our proposal can furnish better performance than classical lifetime models like gamma, Birnbaum-Saunders, Weibull and log-normal. The formulae related with the new model were manageable and may turned into adequate tools comprising the arsenal of applied statistics. We hope that the broached model may attract wider applications for modeling positive real data.

4 STATISTICS-BASED IDENTIFIERS FOR ECR DISTRIBUTED URBAN PATTERNS

4.1 INTRODUCTION

Nowadays, it is required up-to-date information about the urban areas because to the fast growth of urbanization and population of the cities. Aghababaee *et al.* (2013) noted that urban managers and decision makers use this information from any possible data sources. One of the best source of information for urban area is the remote sensing.

Urban areas usually present a *double-bounce scattering* response. The *double-bounce scattering* for the orthogonal illuminated building, according to the radar line of sight is caused by right angle structure, consisting of building block walls and roads surfaces. However there is a problem in the oblique urban areas where the buildings are not aligned orthogonal to the radar line of sight (AZMEDROUB *et al.*, 2016).

SAR and PolSAR remotely sensed data are widely used in urban analysis due to their vast coverage, frequent observation and fair prices (XIANG *et al.*, 2015; SIDDIQUE *et al.*, 2016; XU *et al.*, 2017; JI *et al.*, 2018; QUAN *et al.*, 2018). SAR imagery has long been used as an appropriate and effective data source for many applications (REN *et al.*, 2011; AUSHERMAN *et al.*, 1984; WILEY, 1985; YANG *et al.*, 2013; ZHOU *et al.*, 2018).

The usage of optical remote sensing data is usually limited by lots of factors such as cloud contamination, atmospheric effect, and presence of different materials in complex urban areas. Consequently, they cannot have an acceptable performance. On the other hand, radar sensors because of being independent of atmospheric and weather conditions and not being affected by cloud have gained more attentions. SAR imaging is a well-developed, coherent microwave remote sensing technique for providing large scale two-dimensional high spatial resolution images of the Earth surface reflexivity. Lee and Pottier (2009) shows that operating in microwave spectral region avoids effects of clouds, fog, rain, and smoke.

These factors in addition to the high resolution SAR sensors make these sensors the best choice for the urban analysis. Azmedroub *et al.* (2016) argued that urban areas are very important both socially and environmentally. Their delineation and characterization using remote sensing has attracted continuously researchers by use of different approaches and methods.

Kuruoglu and Zerubia (2004) presented evidence that the CR model may outperform the Weibull, LN and K distributions in SAR imagery analysis, which is expected because urban areas show impulsive characteristics that correspond to underlying heavy-tailed distributions. Li and Ekman (2010), Li and Ekman (2011) introduced the CR law as a model of scattering clusters in state space based on simulation model for single and

multiple polarization channels. Hill *et al.* (2014) produced a novel bivariate shrinkage technique to provide a quantitative improvement in image denoising using the CR model. Recently, Bibalan and Amindavar (2015), Bibalan and Amindavar (2016) furnished a mathematical treatment to the HTR distribution by means of mixtures of CR and Rayleigh models. Other papers that used the CR in SAR modelling are Peng *et al.* (2017) and Pappas *et al.* (2017). This approach was developed for modeling amplitude of ultrasound images through the HTR distribution.

It is known it is not possible to determine the existence of uniformly most powerful tests for such a configuration hypotheses. Thus, this study depends of a reasonable suggestions of hypothesis tests. In this chapter, we opt to use two: likelihood ratio and entropy-based methodology.

Some recent works that use EBTs in SAR and PolSAR imagery are Frery *et al.* (2013) and Nascimento *et al.* (2014b). An important advantage of the EBTs is that they require a smaller computational effort that the LRT alternative. While the two-sample LRT computes the lf three times, which folds several internal summations, the EBTs require only the normalized squared difference between two stochastic entropies. Then the EBTs can reduce the response time of SAR systems in comparison of the LRT option.

The aim of this chapter is to advance the techniques of SAR imagery analysis in order to develop CR and ECR based procedures to test (2.20) from SAR point of view. To accomplish this goal, some hypothesis tests based on likelihood ratio and entropy difference principles are proposed and assessed. All hypothesis tests are based on the CR and ECR models defined by (2.7) and (3.1) respectively. These combinations are not applied in SAR imagery literature yet. We point the main characteristics of each ECR based tests focusing on urban areas extraction. Simulation studies are done to access the nominal level and power of the ECR based tests. In general, the nominal level is underestimated for all tests and their powers are very close each other. An application was made in a San Francisco SAR image and the ECR EBTs alternatives present best properties. identifying urban patterns

This chapter unfolds as follows. Section 4.2 reveals some closed-form expressions for the Shannon, Rényi, Tsallis and Arimoto entropy measures under the CR and ECR models and discusses their relationships. The applied LRTs and EBTs are scrutinized in Section 4.3. These tests performances are analyzed in Section 4.4 focusing on nominal level and power. Section 4.5 contains an application of these methodologies in San Francisco SAR image. Conclusions and future research fields are presented in Section 4.6.

4.2 ECR ENTROPY MEASURES

As pointed in Section 2.4 the q-integral is a key to define the q-entropies class. Lemma 4.1 defined the ECR q-entropy in the parametric space $\boldsymbol{\theta} = (\beta,\lambda)^\top$. Proposition 4.1

shows closed-form expressions for the ECR q-entropies settled in Proposition 2.1 obtained simply applying Lemma 4.1.

Lemma 4.1 (The ECR q-integral)**.** *The ECR q-integral, as defined in* (2.16)*, is given by*

$$\boldsymbol{I}(q,\boldsymbol{\theta}) = \beta^q \left(\frac{\lambda}{\sqrt{2}}\right)^{1-q} \mathrm{B}\left(\left(\beta-\frac{1}{2}\right)q+\frac{1}{2},2q-1\right) \times {}_2\mathrm{F}_1\left(\frac{1}{2}(1-q),q\left(\beta-\frac{1}{2}\right)+\frac{1}{2};\left(\frac{3}{2}+\beta\right)q-\frac{1}{2};\frac{1}{2}\right).$$

where $q > 1/2$, $q \neq 1$ and if $\beta < 1/2$ then $q < 1/(1-2\beta)$.

Proposition 4.1. *The ECR Rényi, Tsallis and Arimoto q-entropies are*

$$\mathcal{H}_R(q,\boldsymbol{\theta}) = \log\lambda + \frac{q}{1-q}\log\beta - \frac{1}{2}\log 2 + \frac{1}{1-q}\log\left[\mathrm{B}\left(\left(\beta-\frac{1}{2}\right)q+\frac{1}{2},2q-1\right)\right] + \frac{1}{1-q}\log\left[{}_2\mathrm{F}_1\left(\frac{1}{2}(1-q),q\left(\beta-\frac{1}{2}\right)+\frac{1}{2};\left(\frac{3}{2}+\beta\right)q-\frac{1}{2};\frac{1}{2}\right)\right],$$

$$\mathcal{H}_T(q,\boldsymbol{\theta}) = \frac{1}{1-q}\left[\frac{\beta^q\lambda^{1-q}}{2^{(1-q)/2}}\mathrm{B}\left(\left(\beta-\frac{1}{2}\right)q+\frac{1}{2},2q-1\right) \times {}_2\mathrm{F}_1\left(\frac{1-q}{2},q\left(\beta-\frac{1}{2}\right)+\frac{1}{2};\left(\frac{3}{2}+\beta\right)q-\frac{1}{2};\frac{1}{2}\right)-1\right]$$

and

$$\mathcal{H}_A(q,\boldsymbol{\theta}) = \frac{q}{1-q}\left\{1-\beta\left(\frac{\lambda}{\sqrt{2}}\right)^{\frac{1-q}{q}}\left[\mathrm{B}\left(q\left(\beta-\frac{1}{2}\right)+\frac{1}{2},2q-1\right) \times {}_2\mathrm{F}_1\left(\frac{1-q}{2},q\left(\beta-\frac{1}{2}\right)+\frac{1}{2};q\left(\frac{3}{2}+\beta\right)-\frac{1}{2};\frac{1}{2}\right)\right]^{\frac{1}{q}}\right\}.$$

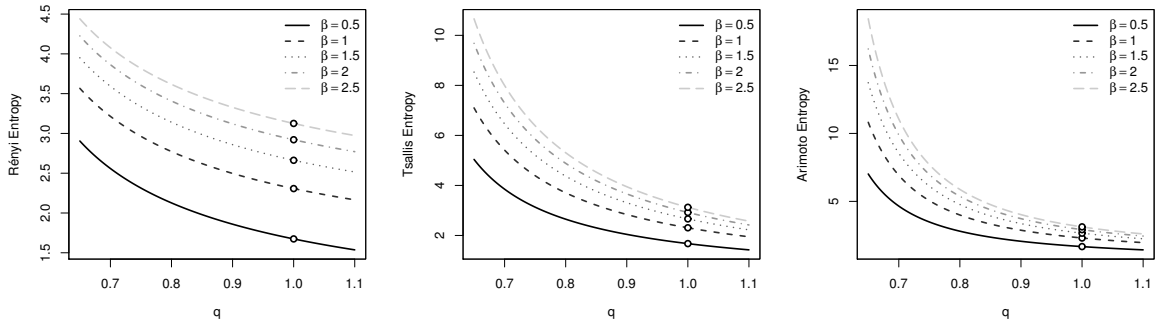
where $q > 1/2$, $q \neq 1$ and if $\beta < 1/2$ then $q < 1/(1-2\beta)$.

Figure 4.1 displays curves of the standard ECR q-entropies. One can observe the next relation holds:

$$\mathcal{H}_R < \mathcal{H}_T < \mathcal{H}_A.$$

These plots also reveal that the deduced entropies respect, as expected, the property (2.17). In later discussions, these quantities will be used like contrast measures and it is expected those in terms of $\mathcal{H}_A$ are more sensible to variations of the additional shape parameter.

The CR q-entropies can be deduced from Proposition 4.1 as shown in Corollary 4.1.1. As expected these expressions are relatively simpler than corresponding ECR q-entropies.



(a) Rényi. (b) Tsallis. (c) Arimoto.

Figure 4.1: Standard ECR q-entropy curves.

Corollary 4.1.1. *The CR Rényi, Tsallis and Arimoto q-entropies are*

$$\mathcal{H}_R(q,\lambda|\beta=1) = \log\lambda + \frac{1}{1-q}\log\left[\frac{\pi 2^{2-3q}\Gamma(2q-1)}{\Gamma(q/2)\Gamma(3q/2)}\right],$$

$$\mathcal{H}_T(q,\lambda|\beta=1) = \frac{1}{q-1} - \frac{\pi(8\lambda)^{1-q}\Gamma(2q-2)}{\Gamma\left(\frac{q}{2}\right)\Gamma\left(\frac{3q}{2}\right)}$$

and

$$\mathcal{H}_A(q,\lambda|\beta=1) = \frac{q}{1-q}\left\{\left[\frac{\lambda^{1-q}\pi 2^{2-3q}\Gamma(2q-1)}{\Gamma(q/2)\Gamma(3q/2)}\right]^{1/q} - 1\right\},$$

where $q > 1/2$ *and* $q \neq 1$.

Lemma 4.2 presents a key result to obtain the q-entropies gradients relative to the parametric vector $\boldsymbol{\theta}$.

Lemma 4.2. *The ECR q-integral gradients relative to the parametric vector* $\boldsymbol{\theta}$ *is*

$$\nabla \boldsymbol{I}(q,\boldsymbol{\theta}) = \left(\frac{\partial \boldsymbol{I}(q,\boldsymbol{\theta})}{\partial\beta}, \frac{\partial \boldsymbol{I}(q,\boldsymbol{\theta})}{\partial\lambda}\right)^\top = \left(q\left(\frac{\boldsymbol{I}(q,\boldsymbol{\theta})}{\beta} + \frac{\Upsilon(q,\beta)}{\lambda^{q-1}}\right), \left(\frac{1-q}{\lambda}\right)\boldsymbol{I}(q,\boldsymbol{\theta})\right)^\top,$$

where $q > 1/2$, $q \neq 1$ *and if* $\beta < 1/2$ *then* $q < 1/(1-2\beta)$ *and*

$$\Upsilon(q,\beta) = \beta^q \int_0^1 \frac{u^{q(\beta-1/2)-1/2}\ln u}{(2-u)^{(1-q)/2}(1-u)^{2(1-q)}}\,\mathrm{d}u. \tag{4.1}$$

Figure 4.2 presents the surface of the special function defined by (4.1). The gradients of the ECR q-entropies can be obtained by means of the Proposition 2.2.

A particular case of the cross Shannon entropy (2.18) for the ECR model can be seen in Proposition 4.2. Unfortunately, it was not so simple to obtain the most general result assuming $\beta_1 \neq \beta_2$ and $\lambda_1 \neq \lambda_2$.

Proposition 4.2. *Let* f_i *be the pdf of the random variable* $X_i \sim \mathrm{ECR}(\boldsymbol{\theta}_i)$, *where* $\boldsymbol{\theta}_i =$

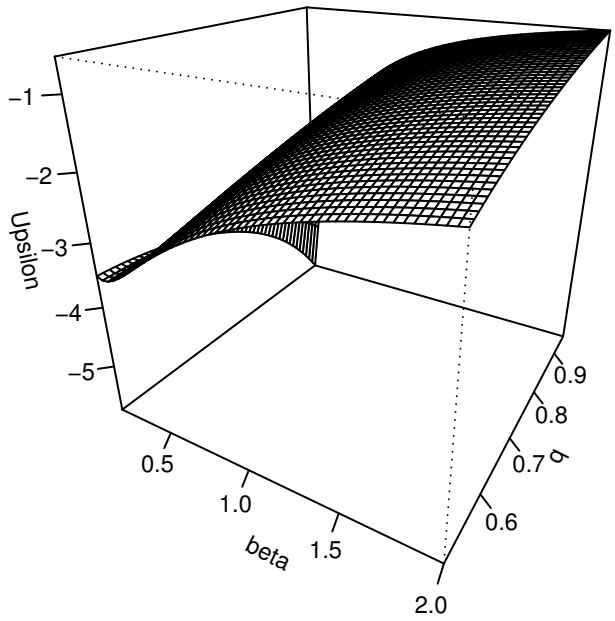


Figure 4.2: Surface of the special function Υ.

$(\beta_i,\lambda)^\top$ *for* $i = 1,2$. *The ECR cross Shannon entropy for* X_1 *relative to* X_2 *is*

$$\mathcal{H}_S^c(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = 2\left[\Psi(1+\beta_1)+\boldsymbol{\gamma}\right] - \frac{1}{2}\boldsymbol{\Phi}\left(\frac{1}{2};1,\beta_1\right) + \log\left(\frac{\lambda}{\beta_2}\right) + \frac{\beta_2}{\beta_1}.$$

The Shannon entropy in Corollary 4.2.1 can be obtained as a particular case of the cross Shannon entropy in Proposition 4.2. It can also be obtained as a particular case of the Rényi, Tsallis and Arimoto entropies as provided in Proposition 4.1 or using Definition 2.1 supported by Table 2.3.

Corollary 4.2.1. *The ECR Shannon entropy is*

$$\mathcal{H}_S(\boldsymbol{\theta}) = 2\left[\Psi(1+\beta)+\boldsymbol{\gamma}\right] - \frac{1}{2}\boldsymbol{\Phi}\left(\frac{1}{2};1,\beta\right) + \log\left(\frac{\lambda}{\beta}\right) + 1.$$

Figure 4.3 shows when $q = 3/4$ the ECR entropies values increase as their parameters increase. The CR Shannon entropy can be reckoned from Corollary 4.2.1 as shown in Corollary 4.2.2 and the ECR Shannon entropy gradient are presented in Corollary 4.2.3.

Corollary 4.2.2. *The CR Shannon entropy is* $\mathcal{H}_S(\lambda|\beta=1) = \log\left(\frac{\lambda}{2}\right) + 3$.

Corollary 4.2.3. *The ECR Shannon entropy gradient are*

$$\nabla\mathcal{H}_S(\boldsymbol{\theta}) = \left(\frac{\partial\mathcal{H}_S(\boldsymbol{\theta})}{\partial\beta}, \frac{\partial\mathcal{H}_S(\boldsymbol{\theta})}{\partial\lambda}\right)^\top = \left(2\Psi^{(1)}(1+\beta) + \frac{1}{2}\boldsymbol{\Phi}\left(\frac{1}{2};2,\beta\right) - \frac{1}{\beta}, \frac{1}{\lambda}\right)^\top.$$

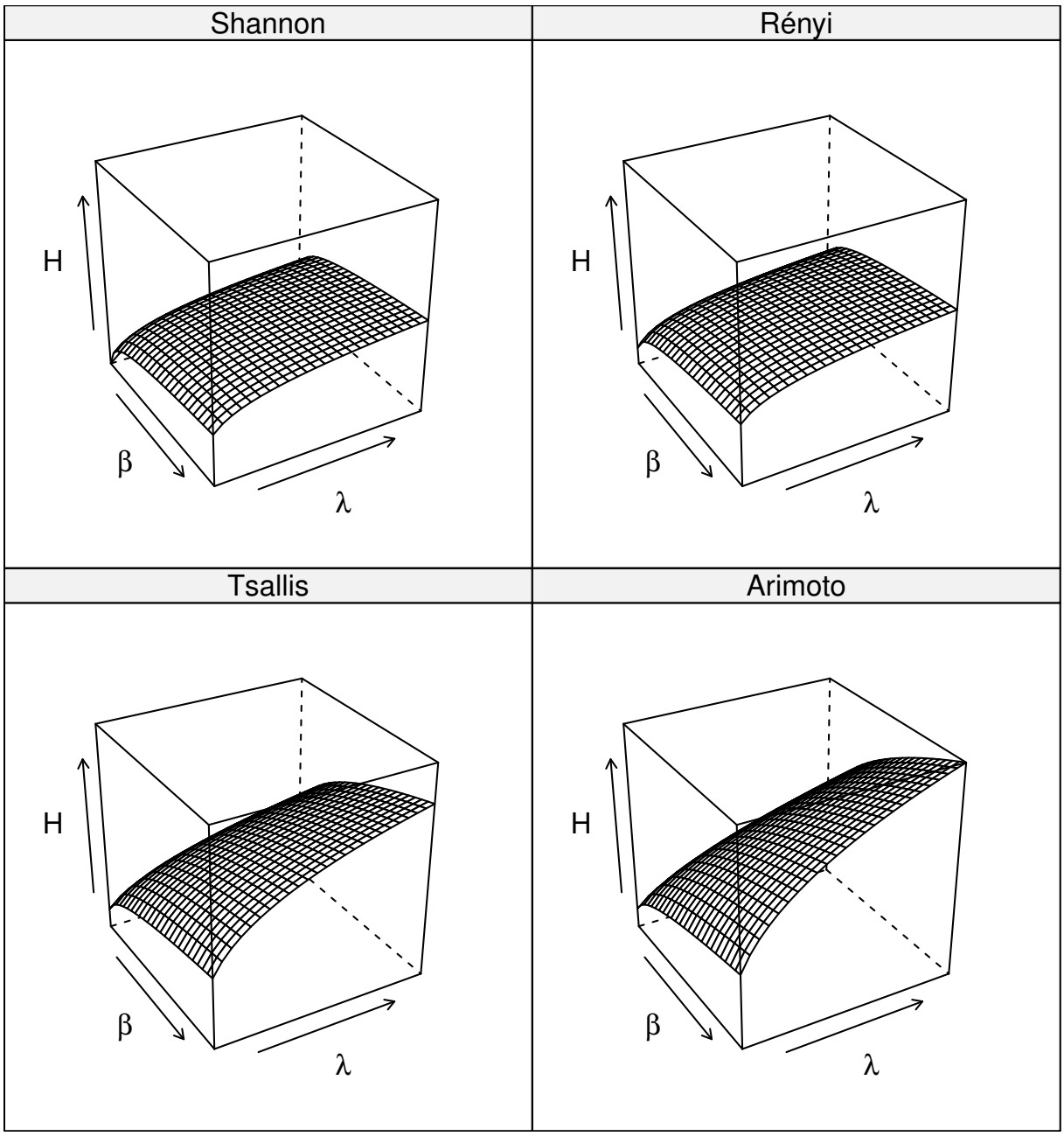


Figure 4.3: ECR entropy surfaces considering $q = 3/4$.

4.3 SOME TWO-SAMPLE HYPOTHESIS TESTS IN THE ECR MODEL

4.3.1 Likelihood ratio tests

Considering $\mathcal{F}$ in (2.19) as the class of ECR models, the LRT statistic (2.21) for test (2.20) becomes

$$\log \boldsymbol{\Lambda}(\boldsymbol{x}) = \sum_{i=1}^{2} \left[n_i \log \left(\frac{\hat{\beta}_i \hat{\lambda}_i}{\hat{\beta}_0 \hat{\lambda}_0} \right) + 3T_2(\hat{\lambda}_i, \boldsymbol{x}_i) + (\hat{\beta}_i - 1) T_3(\hat{\lambda}_i, \boldsymbol{x}_i) \right] - 3T_2(\hat{\lambda}_0, \boldsymbol{x}) - 3T_2(\hat{\lambda}_0, \boldsymbol{x}) - (\hat{\beta}_0 - 1) T_3(\hat{\lambda}_0, \boldsymbol{x}), \quad (4.2)$$

where $\hat{\boldsymbol{\theta}}_i = (\hat{\beta}_i, \hat{\lambda}_i)$ for $i = 0, 1, 2$. T_2 and T_3 are given, respectively, by Eqs. (3.11) and (3.12). Given a nominal level α, the ECR LRT rejects H_0 if

$$2 \log \boldsymbol{\Lambda}(\boldsymbol{x}) > c_{2,1-\alpha},$$

where $c_{2,1-\alpha}$ is the $1-\alpha$ quantile of χ^2_2. In the CR LRT case the expression (4.2) reduces to

$$\log \boldsymbol{\Lambda}_{\mathrm{CR}}(\boldsymbol{x}) = \sum_{i=1}^{2} \left[n_i \log \left(\frac{\hat{\lambda}_i}{\hat{\lambda}_0} \right) + 3T_2(\hat{\lambda}_i, \boldsymbol{x}_i) \right] - 3T_2(\hat{\lambda}_0, \boldsymbol{x}),$$

and the nominal level α CR LRT rejects H_0 if

$$2 \log \boldsymbol{\Lambda}_{\mathrm{CR}}(\boldsymbol{x}) > c_{1,1-\alpha},$$

where $c_{2,1-\alpha}$ is the $1-\alpha$ quantile of χ^2_2.

4.3.2 Entropy-based tests

Considering the notes in Section 2.5 the CR or ECR models the *Shannon entropy test* (SET) can be implemented applying the Corollaries 4.2.1 and 4.2.3 in (2.24) while the *Rényi entropy test* (RET), *Tsallis entropy test* (TET) and *Arimoto entropy test* (AET) can be obtained using the Propositions 2.2 and 4.1. The ECR FIM are obtained from Corollary 3.15.2.

The CR EBTs statistics are relatively simple to be deduced. Proposition 4.3 shows these statistics. The CR SET and CR RET statistics are equivalent, then the order parameter q has influence only in CR TET and CR AET. Another interesting property of the CR EBTs statistics is that they not contain the gamma function instead of theirs corresponding entropies in Corollary 4.1.1. Explicit expressions for the ECR EBTs statistics are relatively complex and can be obtained by author contact.

Proposition 4.3. *The CR EBTs statistics are*

$$\mathcal{Z}_{\mathrm{CR}}^{\mathrm{SET}}(\boldsymbol{x}) = \mathcal{Z}_{\mathrm{CR}}^{\mathrm{RET}}(\boldsymbol{x}) = \frac{4}{5}\left(\frac{n_1 n_2}{n_1+n_2}\right)\left[\log\left(\frac{\hat{\lambda}_1}{\hat{\lambda}_2}\right)\right]^2,$$

$$\mathcal{Z}_{\mathrm{CR}}^{\mathrm{TET}}(\boldsymbol{x}) = \frac{4}{5}\left(\frac{n_1 n_2}{n_2\hat{\lambda}_1^{2(1-q)} + n_1\hat{\lambda}_2^{2(1-q)}}\right)\left(\frac{\hat{\lambda}_1^{1-q} - \hat{\lambda}_2^{1-q}}{1-q}\right)^2$$

and

$$\mathcal{Z}_{\mathrm{CR}}^{\mathrm{AET}}(\boldsymbol{x}) = \frac{4}{5}\left(\frac{n_1 n_2}{n_2\hat{\lambda}_1^{2(1-q)/q} + n_1\hat{\lambda}_2^{2(1-q)/q}}\right)\left[\frac{q}{1-q}\left(\hat{\lambda}_1^{(1-q)/q} - \hat{\lambda}_2^{(1-q)/q}\right)\right]^2.$$

In what follows, we present some essays on which the performance of tests is both computed and compared. Before, we anticipate some comparisons made in theoretical terms:

1. LRT is more simple of being implemented because do not depend of integrations, often hard. But, when analytically tractable, EBTs depend on data only through MLEs, while LRT also consider the productions of densities;

2. In practical sense, entropies often bring behaviors associated with phenomenon, while parameters reveal the same by a more confuse path.

4.4 NUMERICAL RESULTS

In this section, we assess the proposed tests performances by means of two simulation studies under finite samples. The computations were done using the platform described in Section 1.2. In the following studies consider the sets

$$\mathfrak{N} = \{25{,}121{,}289\},$$
$$\mathfrak{L} = \{0.01{,}0.05{,}0.10\}$$

and

$$\mathfrak{T} = \{\mathrm{LRT}, \mathrm{SET}, \mathrm{RET}, \mathrm{TET}, \mathrm{AET}\}.$$

When required the entropy order parameter q was arbitrated as $3/4$ and all the CR and ECR estimates were obtained by means of the MLEs as described in Section 3.5.1.

4.4.1 Nominal level study

The first study was focused on obtaining numerically the test sizes or the probability of type I error. Let the set

$$\mathfrak{P}_1 = \{0.15{,}0.3, \ldots, 3\}$$

and

$$\mathfrak{S}_1 = \mathfrak{N} \times \mathfrak{L} \times \mathfrak{T} \times \mathfrak{P}_1.$$

For each element $(n,\alpha,t,\beta) \in \mathfrak{S}_1$ it was used the ITSM, as presented in Section 3.2, to generate 10,000 pairs of standard ECR random samples with size n and shape parameter β. It was applied the two-samples ECR based test t with nominal level α to compare these samples and finally was associated to each element of $\mathfrak{S}_1$ its correspondent rejection rate computed under all 10,000 tests made.

Figure 4.4 shows a lattice plot of the computed values, we can see that the size of the tests converge, as expected, to the nominal level as the sample size increases. The tests tend to underestimate their sizes for smaller samples and grater values of β.

Tables B.8 and B.9 summarize the minimum and maximum distance between the computed sizes and their corresponding nominal levels. Table B.2 shows the number of times each test are ranked in these tables. Jointly LRT and SET appears 133 times in Table B.8 while LRT and AET, jointly, appears 140 times in Table B.9. RET and TET have the smallest number of times the extreme values in Figure 4.4.

4.4.2 Power study

The second study was formulated to compute the power of the tests. Define

$$\mathfrak{P}_2 = \{0.1,0.2,\ldots,2\}$$

and

$$\mathfrak{S}_2 = \mathfrak{N} \times \mathfrak{L} \times \mathfrak{T} \times \mathfrak{P}_2.$$

For each element $(n,\alpha,t,\theta) \in \mathfrak{S}_2$ it was used the ITSM to generate a ordered set $\mathfrak{A}_1$ of 10,000 ECR random samples with size n and both parameters equal to θ and a second ordered set $\mathfrak{A}_2$ of 10,000 ECR random samples with size n and unit parameters. It was applied the two-samples ECR based test t with nominal level α to compare the corresponding samples in $\mathfrak{A}_1$ and $\mathfrak{A}_2$ and finally was associated to each element of $\mathfrak{S}_2$ its correspondent rejection rate computed under all 10,000 tests made.

Note that the all the samples are generated under the line $\beta = \lambda$ inside the parametric space then, as depicted in Figure 4.3, under this path all entropies are increasing functions of θ, the tests (2.20) and (2.25) are equivalent and therefore LRT and the EBTs are comparable.

Figure 4.5 shows the estimated power curves of the tests. As expected the power of the tests increase as the sample size increases.

Tables B.10 and B.11 summarize the maximum and minimum powers of the tests in function of the θ. Table B.2 shows the number of times each test are ranked in these

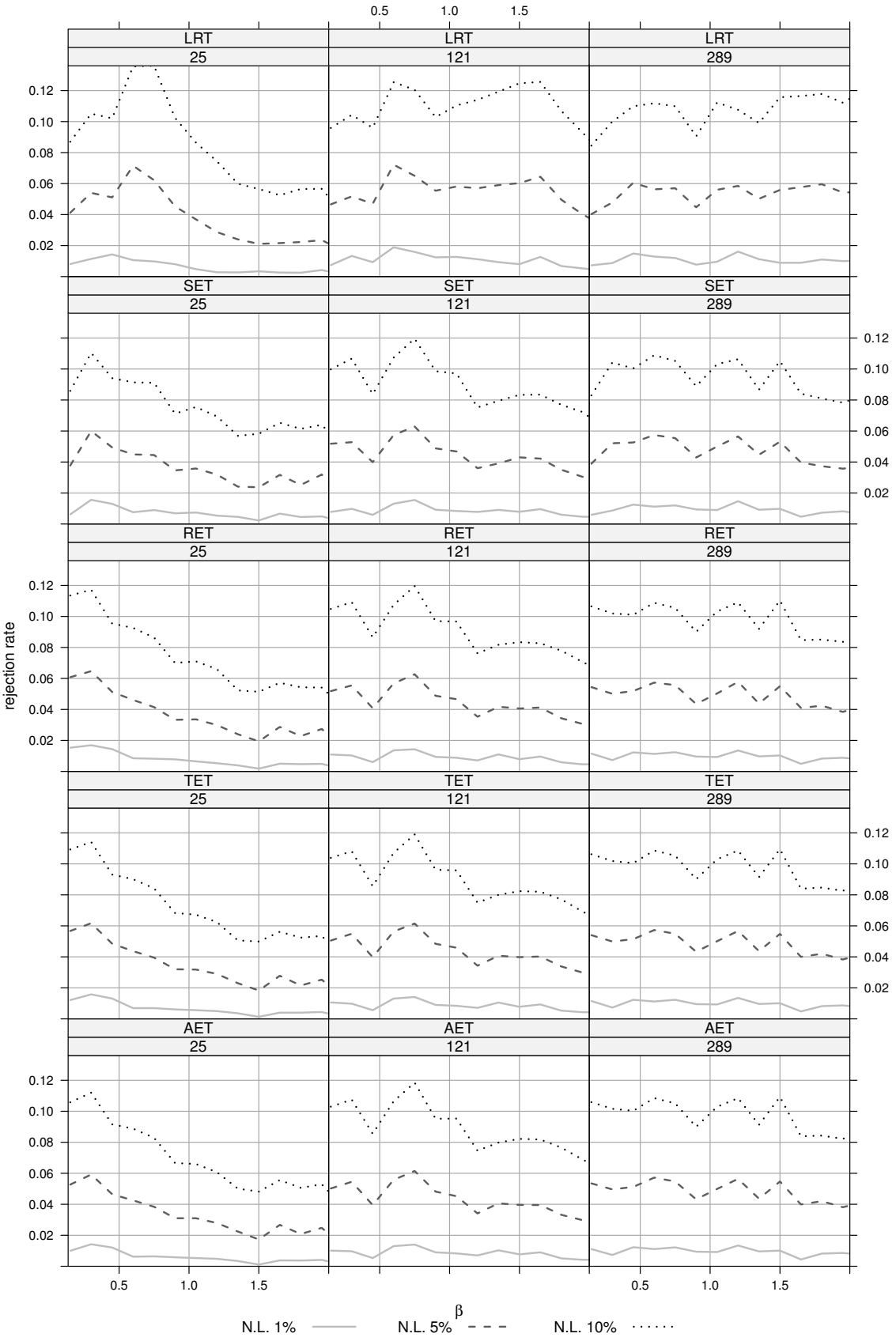


Figure 4.4: Sizes of the tests in $\mathfrak{T}$ with nominal levels in $\mathfrak{L}$ for sample sizes in $\mathfrak{N}$ based on the standard ECR regarding $\beta \in \mathfrak{P}_1$.

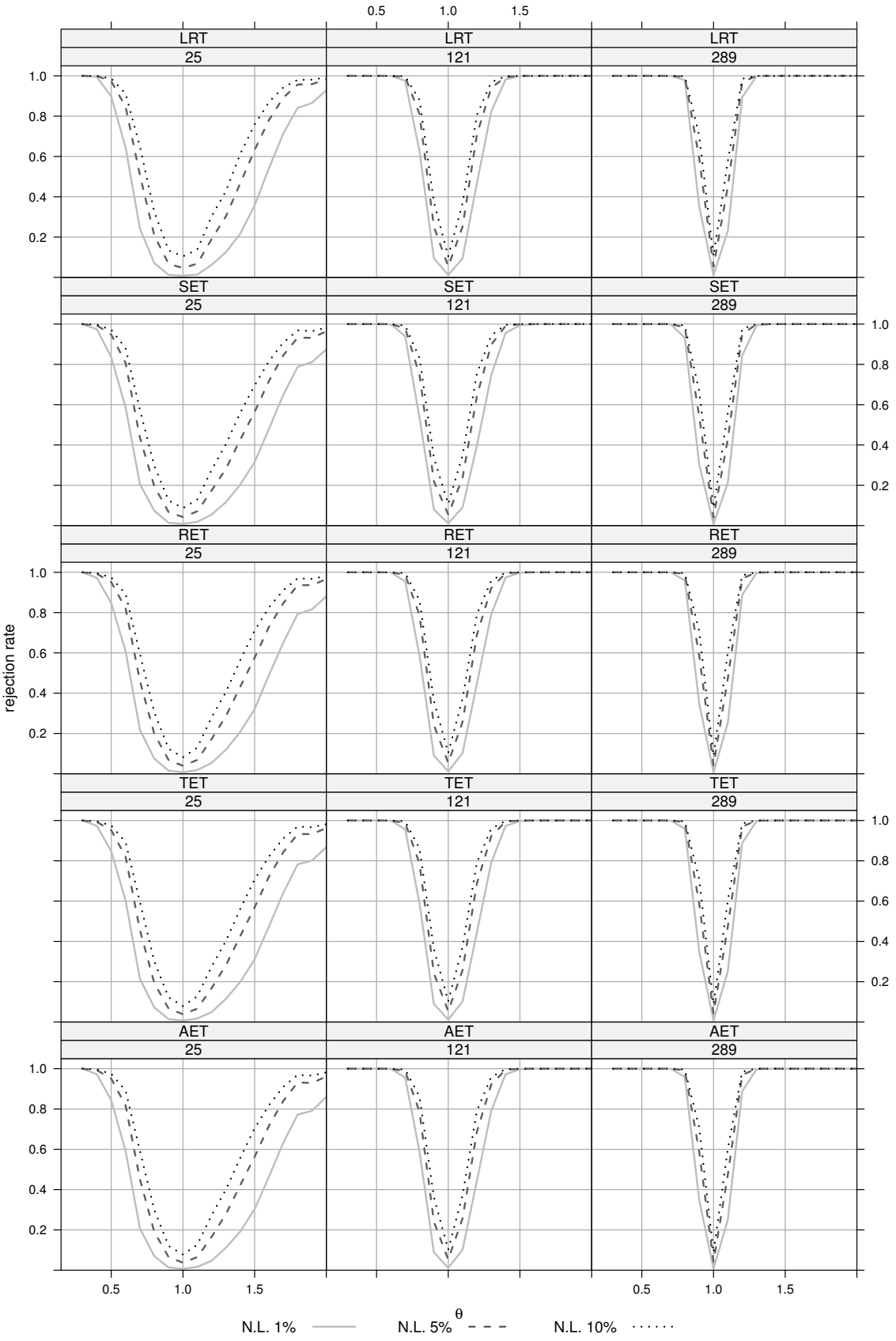


Figure 4.5: Sizes of the tests in $\mathfrak{T}$ with nominal levels in $\mathfrak{L}$ for sample sizes in $\mathfrak{N}$ based on the ECR assuming $(\beta,\lambda) \in \mathfrak{P}_2$ and $\theta = \beta = \lambda$.

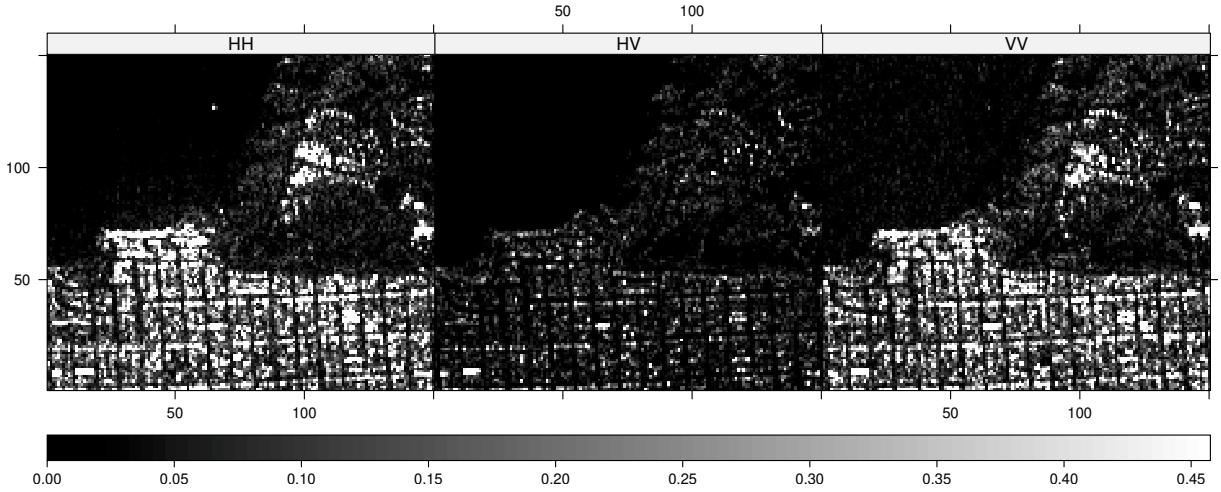


Figure 4.6: SAR intensities maps of the San Francisco image clipped at the 95^{th} percentile grouped by channels in $\mathfrak{C}_{(H,V)}$.

tables. The LRT typically reaches the highest powers and the SET and AET present the slowest ones.

4.5 EXPERIMENTAL RESULTS

In this section, we investigate the performance of the proposed scattering mechanism identification scheme on real SAR data. Three studies are made to analyze data GoF, urban texture identification and redundancy information between polarization channels. As in Section 4.4, the entropy order parameter q was arbitrated as $3/4$ when required. In all of the following studies the CR and ECR estimates are obtained by means of the MLE as described in Section 3.5.1.

4.5.1 The San Francisco image

We used a PolSAR image of San Francisco bay acquired in April 9, 2008 by RADARSAT-2 system at C-band with dimensions of 150×150 pixels. This image was studied by Yang *et al.* (2013) and Azmedroub *et al.* (2016) using PolSAR techniques.

Figures 4.6 and 4.7 show, respectively, the collected SAR signal intensities maps and histograms for its three polarization channels. The intensity histograms suggest the employ of a heavy-tailed model to describe the San Francisco textures. In order to get a better view the maps in Figure 4.6 were clipped at the 95^{th} intensity percentile. Figure 4.8 shows its corresponding optical image.

There are three distinct regions in this image: sea, urban and forest. We did not have access to the real nature of all pixels in this image, but it is possible to sampling sea, urban or forest zones using Figure 4.8(a) as an optical reference. Figure 4.8(b) presents a manual segmentation of the San Francisco image using this approach, which

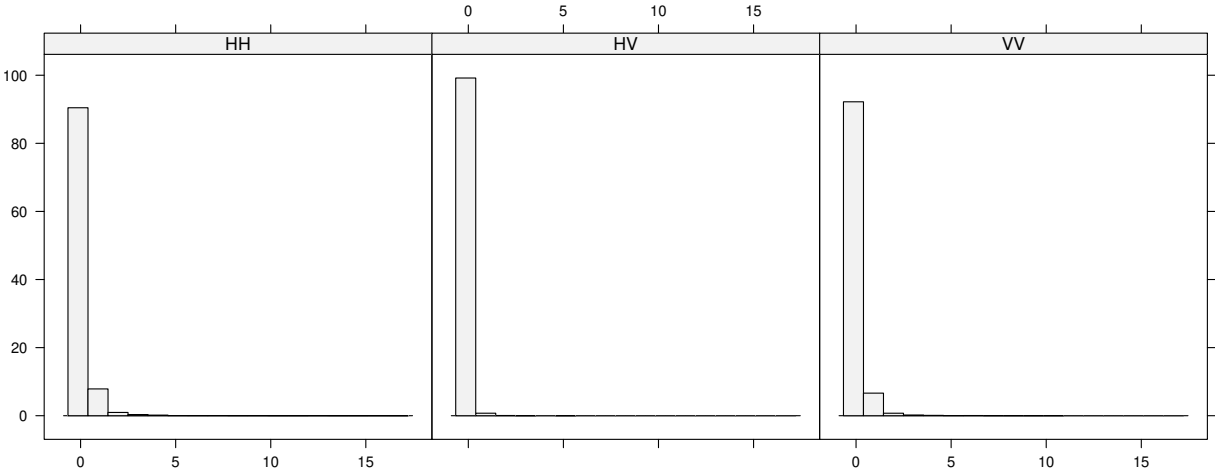
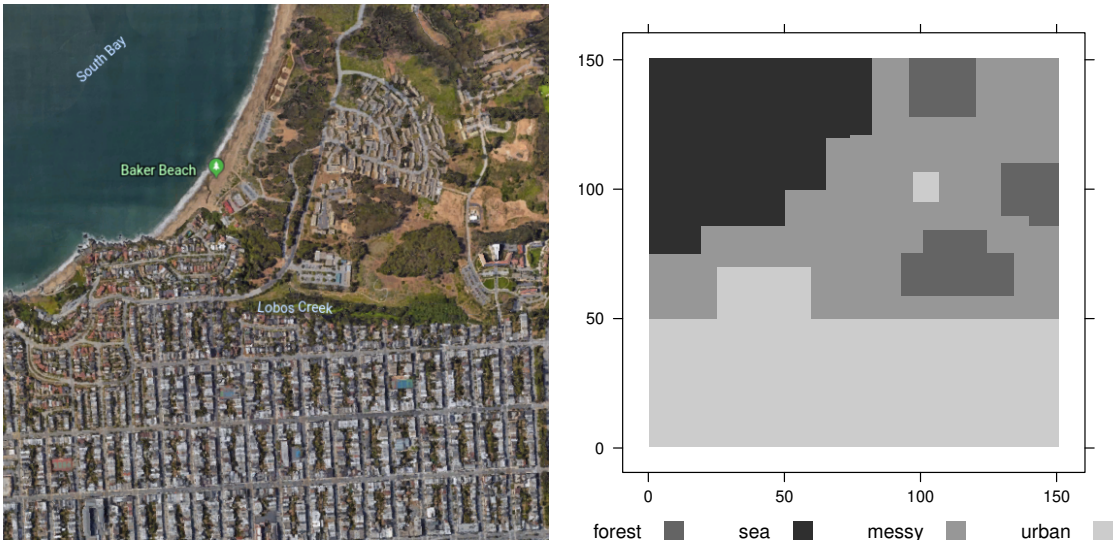


Figure 4.7: SAR intensities histograms of the San Francisco image grouped by channels in $\mathfrak{C}_{(H,V)}$.



(a) Optical image obtained from Google earth©. (b) Samples of the main textures.

Figure 4.8: Additional San Francisco images corresponding to SAR maps in Figure 4.6.

adds an additional messy texture including all pixels not classifiable by comparison with Figure 4.8(a). The number of pixels in each samples can be verified in Table 4.1. Clearly the urban texture compose the largest region in this figure presenting, in general, higher intensities as disposed on Figure 4.6.

4.5.2 Goodness-of-fit study

Let the following set of windows associated to the San Francisco image in Section 4.5.1:

$$\mathfrak{W}_{5\times5} = \{\mathfrak{W}_{(i,j)} : (i,j) \in \mathfrak{I}^2\}, \tag{4.3}$$

Table 4.1: Number of pixels by main textures for the samples depicted in Figure 4.8(b).

texture	forest	sea	messy	urban
number of pixels	1990	4743	7638	8129

where $\mathfrak{W}_{(i,j)}$, defined by

$$\mathfrak{W}_{(i,j)} = \left\{(i^*,j^*) \in \mathfrak{I}^2, \max\left(|i-i^*|,|j-j^*|\right) \leq 2\right\},$$

is the 5×5 window centered at the pixel of coordinates (i,j) and $\mathfrak{I} = \{1,\ldots,150\}$. Following the recommendations given by Frery *et al.* (2004) and Rodrigues *et al.* (2016) we consider the set $\mathfrak{W}_{5\times5}$ which slices the whole San Francisco image in 150×150 overlapping windows. Frery *et al.* (2004) argues that larger the windows size, the smoother the analysis, which implies that most sites corresponded to heterogeneous or extremely heterogeneous spots. Smaller windows size have more heterogeneous areas.

Regard the sets

$$\mathfrak{M} = \{\mathrm{CR}\,,\mathrm{ECR}\} \tag{4.4}$$

and

$$\mathfrak{S}_3 = \mathfrak{W}_{5\times5} \times \mathfrak{C}_{(H,V)} \times \mathfrak{M}.$$

For each element $(\mathfrak{W}_{(i,j)},c,P) \in \mathfrak{S}_3$ it was found the probability model P ML estimates for intensities in the window $\mathfrak{W}_{(i,j)}$ and SAR polarization channel c at the San Francisco image. The AD test was applied to obtained estimates in order to check the probability model P GoF.

The AD test places more weight on observations in the tails of the distribution when compared to its KS and CvM alternatives (YAP; SIM, 2011). Then it is adequate compare GoF of heavy-tailed distributions like the CR and ECR models as proved in Proposition 3.3.

Figure 4.9 illustrates the obtained AD tests p-values and evinces that the greatest ones, for both models, are concentrated inside urban area, which indicates the use of these models to describe textures of urban SAR data.

The ECR model presents a relatively small number of windows where the ML estimates do not exist, but this is expected due to what was observed in Section 3.6.3. This problem does not occurs in the CR model.

Figure 4.10 depicts the p-values histograms of the applied GoF test considering now only windows inside the urban region. The histograms modes are closed to the maximum capitally in HH and VV polarization channels and for greater p-values the ECR histograms present higher frequencies than the CR ones. Greater AD p-values correspond to smaller values of the AD statistic and better GoF measures.

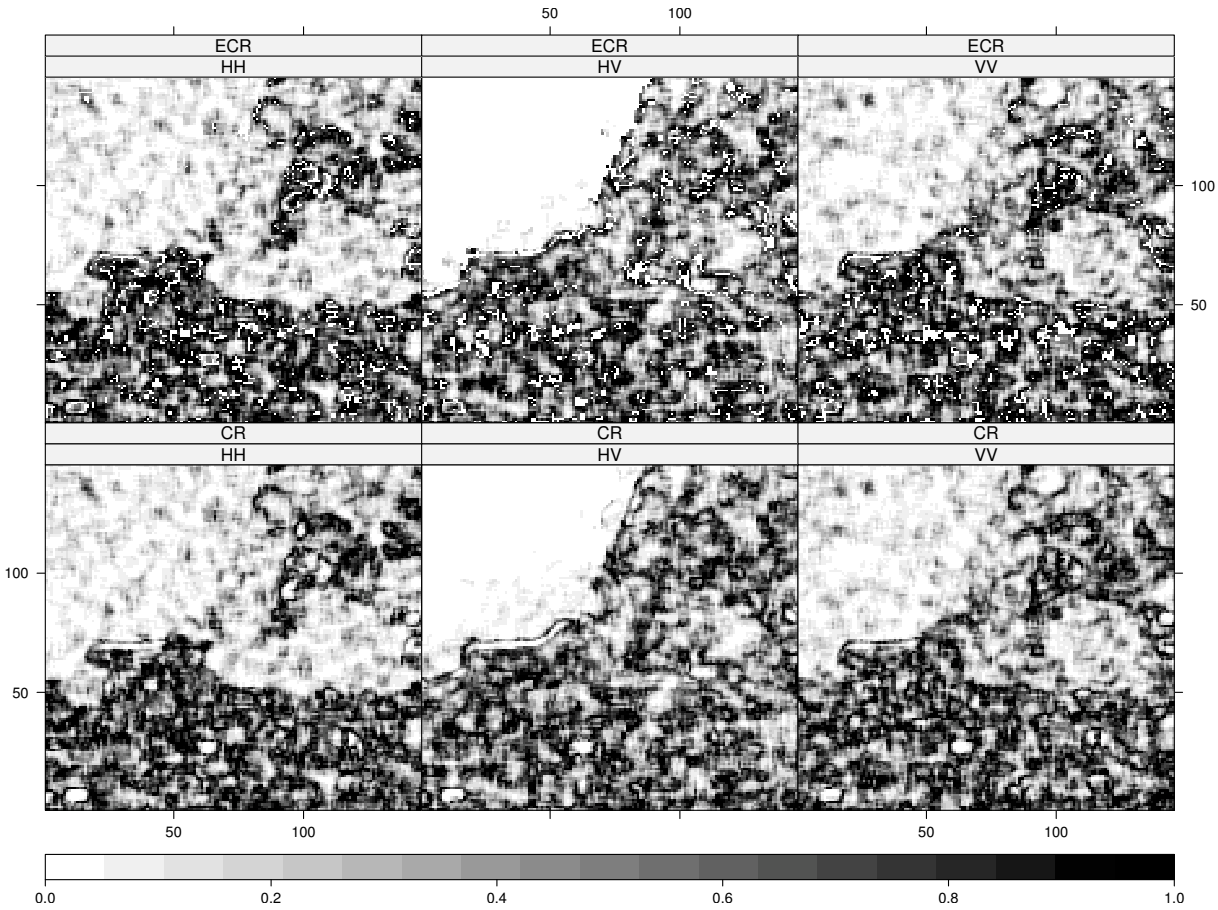


Figure 4.9: AD p-value maps based on the set of models $\mathfrak{M}$ of all SAR intensity samples corresponding to the windows in $\mathfrak{W}_{5\times5}$ grouped by the set of channels $\mathfrak{C}_{(H,V)}$ and the same set of models.

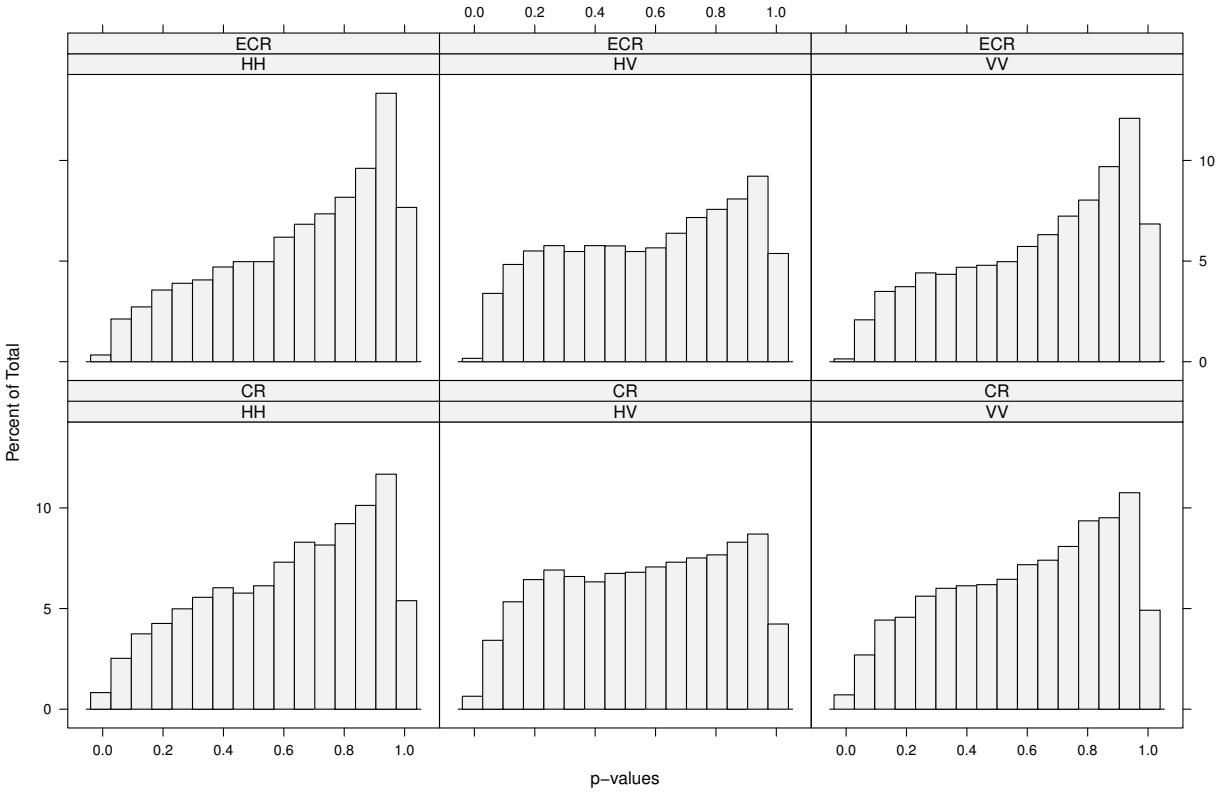


Figure 4.10: AD p-value histograms based on the set of models $\mathfrak{M}$ of all urban SAR intensity samples corresponding to the windows in $\mathfrak{W}_{5\times5}$ grouped by the set of channels $\mathfrak{C}_{(H,V)}$ and the same set of models.

Table 4.2: Percentages of urban SAR intensity samples corresponding to the windows in $\mathfrak{W}_{5\times5}$ rejected by AD GoF test as drawn of the models in $\mathfrak{M}$ grouped by the set of significance levels $\mathfrak{L}$ and the same set of models.

level	CR	ECR
1%	0.39%	0.03%
5%	1.39%	0.91%
10%	3.85%	3.30%

Finally Table 4.2 presents the proportions of urban windows that are rejected by AD GoF test as drawn from the ECR or CR pdf considering the main nominal levels. These rejection rates are smaller in the ECR law, therefore this CR extension presents advantage over its baseline.

4.5.3 Texture identification study

In this second study we verified how able the CR and ECR based LRT and EBTs to test if two SAR samples, obtained of the same polarization channel, provide or not of urban areas.

Consider the set

$$\mathfrak{S}_4 = \mathfrak{C}_{(H,V)} \times \mathfrak{N} \times \mathfrak{M} \times \mathfrak{L} \times \mathfrak{T}.$$

For each element $(c,n,P,\alpha,t) \in \mathfrak{S}_4$ it was collected 10,000 pairs of intensity samples with size n in the channel c inside the urban region of the San Francisco SAR image, as detached in Figure 4.8. For each element of $\mathfrak{S}_4$ the seed to random collect the pairs of samples was set to n. After this initial step it was applied the two-sample test t based on the probability model P with nominal level α to each of these sample pairs. Finally the rejection rate associate to each element of $\mathfrak{S}_4$ are computed.

Tables B.1, B.3 and B.4 tabulate the obtained mean rejection rates. The CR LRTs present higher rejection rates than their corresponding ECR LRTs in all polarization channels. The EBTs assume opposite behavior achieving superior rejection rates in ECR based tests. These facts provides empirical evidence that the additional ECR shape parameter β improves only the EBTs.

The rejection rates are below their corresponding nominal levels and appears does not reach them as window size increases, but it is important to note that larger windows sizes are not recommended in signal processing literature (FRERY *et al.*, 2004), furthermore this behavior is expected due to Figure 4.4.

The VV polarization channel presents the rejection rates closer to the corresponding nominal levels. Finally the ratio between the rejection rates corresponding to the nominal level 10% and the expected value of 10% are the lowest when compared to other nominal levels.

4.5.4 Information redundancy study

In this second study we verified the capacity of the LRT and the EBTs to detect if two SAR textures provides or not of urban areas from the points of view of two different channels.

Regard the set

$$\mathfrak{S}_5 = \mathcal{P}_2\left(\mathfrak{C}_{(H,V)}\right) \times \mathfrak{N} \times \mathfrak{M} \times \mathfrak{L} \times \mathfrak{T},$$

where $\mathcal{P}_2(A)$ is the set of all subsets of A with cardinality 2.

For each element $((c_1,c_2),n,P,\alpha,t) \in \mathfrak{S}_5$ it was generated ordered sets $\mathfrak{A}_1$ and $\mathfrak{A}_2$ of 10,000 collected intensity samples of size n in the channels c_1 and c_2, respectively, inside the urban zone of the San Francisco SAR image. The seed used to generate each of these ordered sets was specified as n. After this initial phase it was applied the two-sample test t based on the probability model P with nominal level α to the corresponding samples in the ordered sets $\mathfrak{A}_1$ and $\mathfrak{A}_2$. Finally the rejection rate associate to each element of $\mathfrak{S}_5$ are computed.

Tables B.5 to B.7 tabulate the obtained mean rejection rates. When comparing channels HH and VV its notable that the rejection rates are null for all sample sizes and the CR based tests and practically null for the ECR ones. This indicates that HH and VV

polarization channels presents the same texture information.

The highest rejection rates are achieved when comparing HH or VV polarization channels with HV one, particularly when sample sizes are 121 and 289 the rejection rates reaches 100% for these comparisons, therefore the HV polarization channel presents some kind of texture information which is not present in HH or VV ones.

4.6 CONCLUDING REMARKS

This paper has proposed new strategies to quantify differences in *synthetic aperture radar* (SAR) urban patterns by means of its intensities. The differences was detected by means of the hypotheses tests. It was considered two classes of hypotheses tests: *likelihood ratio tests* (LRTs) and *entropy-based tests* (EBTs). To support the main objective it was deduced closed-form expressions for the *exponentiated Cauchy-Rayleigh* (ECR) Shannon, Rényi, Tsallis and Arimoto entropies. The corresponding *Cauchy-Rayleigh* (CR) entropies can be obtained from these previous expressions. The regarded ECR based tests, generally, underestimated the corresponding nominal level for small sample sizes, but as long as the sample size increases the corresponding nominal level are achieved as expected. The power of all test has behaved similarly in the simulated samples, but this behavior is not expected in all scenarios because the LRTs are focused on the distributions homogeneity while EBTs on the entropies equality. The ECR model presented slightly smaller goodness-of-fit measures than CR one in all polarization channels of urban areas of the studied San Francisco SAR image. The texture identification study for the San Francisco urban patterns showed that the CR LRTs and the ECR EBTs have rejection rates closer to the expected nominal level. The final study attested that all tests are able to detect the known information redundancy between the channels HH and VV, but the CR LRTs and the ECR EBTs present the highest rejection rates when applied to detect information redundancy between the channels pairs HH-HV and HV-VV. Many further analyses are needed in future works to discriminate between the different types of urban areas using SAR imagery. A natural sequence of this work is the use of ECR constrained *maximum likelihood estimators* (MLEs) which can overcome the non-existence of maximum likelihood estimates as reported in Section 4.5.2. The constrained MLEs also can make feasible the use of the CS-MLEs as defined in Section 3.5.5.

5 IMAGE SEGMENTATION USING CR AND ECR ENTROPY-BASED MODELS

5.1 INTRODUCTION

Recently, texture surfaces motivated many segmentation studies and algorithms like Jain and Farrokhnia (1990), Oakley and Hancock (1994), Randen and Husoy (1999), Hsin (2000), Huan and Hou (2008), Zhang *et al.* (2008), Karoui *et al.* (2010), Jyothirmayi *et al.* (2015), Kumar *et al.* (2016) and Akbulut *et al.* (2018). The employment of some entropy measures on texture segmentation has been made with success, some of the last studies are Nobre *et al.* (2016), Naidu *et al.* (2017), Abdel-Khalek *et al.* (2017), Yin *et al.* (2017), Nguyen *et al.* (2018), Pham *et al.* (2018), Wang *et al.* (2018), Wu *et al.* (2018) and Wang *et al.* (2018).

Understanding texture as quantitative measure of the arrangement of intensities (not merely its distribution) in a image (SHAPIRO; STOCKMAN, 2001, p. 237), we can define feature vectors to summary the texture of each image cropping. Thus, a crucial step in a segmentation methodology is the extraction of these feature vectors. Some techniques commonly used in feature vector extraction are *principal component analysis* (PCA) (HOTELLING, 1936), *singular value decomposition* (SVD) (GOLUB; KAHAN, 1965), *latent semantic analysis* (LSA) (DEERWESTER *et al.*, 1990), *partial least squares* (PLS) regression (LINDGREN *et al.*, 1993), *independent component analysis* (ICA) (COMON, 1994), Kernel PCA (SCHÖLKOPF *et al.*, 1998), Isomap (TENENBAUM *et al.*, 2000), multi-linear PCA (LATHAUWER *et al.*, 2000) and *multifactor dimensionality reduction* (MDR) (RITCHIE *et al.*, 2001). In this chapter, it is suggested that to use the classes of the CR and ECR information theory measures may consist in good strategy to segment SAR imagery. The employed methods adopt the CR and ECR entropies, as defined in Chapter 4, as feature vectors to model SAR imagery textures. We can not find any research using CR or ECR entropies as feature vectors to describe SAR imagery textures.

In this chapter, it is compared the CR and ECR entropies as basis to obtaining feature vectors for SAR image segmentation algorithms. First, the k-means algorithm is used as a benchmark comparing to the proposed FM based algorithms. It is proposed the CR and ECR versions for FM-based algorithms, assuming CR and ECR distributed intensities inside considered windows. It is associated to each SAR image pixel the parametric intensities entropy of the windows centered in the corresponding pixel regarding the desired probability model and (h,ϕ)-entropy. In terms of methodology each pixel of SAR image is associated with the parametric entropy. Let the following set

$$\mathfrak{H}_q = \{\mathcal{H}_R,\mathcal{H}_T,\mathcal{H}_A\} \times \{0.55,\ldots,0.95\}.$$

Each element $(\mathcal{H}_q,q) \in \mathfrak{H}_q$ personifies the $\mathcal{H}_q$ entropy as given in Proposition 2.1 with order parameter arbitrated to q. Finally, one defines the set $\mathfrak{H}$ by adding the Shannon entropy to the set $\mathfrak{H}_q$

$$\mathfrak{H} = \{\mathcal{H}_S\} \cup \mathfrak{H}_q. \tag{5.1}$$

Regarding a model $P \in \mathfrak{M}$, as defined in (4.4), and an (h,ϕ)-entropy $\mathcal{H} \in \mathfrak{H}$ the studied techniques will be named $(P,\mathcal{H})$-*segmentation based on mixture of stochastic entropy distributions* (SMED).

This chapter is organized as follows. Section 5.2 details the procedure to obtain the class of feature vectors associated to each pixel of the studied image, while Section 5.3 presents the theoretical structure of the used mixture model. The initialization of model parameters and the used algorithm is also show. The GoF and ISP measures are computed and analyzed in Section 5.4, displaying useful figures and tables. Conclusions and forthcoming investigations are presented in Section 5.5.

5.2 FEATURE VECTORS EXTRACTION

In the following, notes, it is considered the same San Francisco image displayed in Section 4.5.1. To obtain feature vector, consider the sets of entropy channels $\mathfrak{C}_{(H,V)}$, windows $\mathfrak{W}_{5\times5}$ and probability models $\mathfrak{M}$ as defined, respectively, in (2.26), (4.3) and (4.4).

For element $(\mathfrak{W}_j,P,\mathcal{H}) \in \mathfrak{W}_{5\times5} \times \mathfrak{M} \times \mathfrak{H}$, it was found the ML estimates, regarding the probability model P, for the $\mathfrak{W}_j$ associated intensities and each SAR polarization channel in $\mathfrak{C}_{(H,V)}$. The values of the entropy $\mathcal{H}$, for each polarization channel in $\mathfrak{C}_{(H,V)}$, are computed using these ML estimates. An assumption which is inherent in this last step is that the window intensities follow the current probability model P. These entropy values define the vector

$$\boldsymbol{x}_j^{(P,\mathcal{H})} = (x_{j1},x_{j2},x_{j3})^\top \tag{5.2}$$

$$= \Big(\mathcal{H}(\hat{\boldsymbol{\theta}}_{HH}), \mathcal{H}(\hat{\boldsymbol{\theta}}_{HV}), \mathcal{H}(\hat{\boldsymbol{\theta}}_{VV})\Big)^\top \tag{5.3}$$

which is the feature vector associated to the j^{th} pixel regarding the entropy $\mathcal{H}$ under the probability model P. Clearly, by Definition 2.1 and Table 2.3, $\boldsymbol{x}_j^{(P,\mathcal{H})} \in \mathbb{R}^3$. From now, the entropy channels in $\mathfrak{C}_{(H,V)}$ will be indexed as elicited by (5.2) and (5.3). Due to the actual extraction of the feature vectors, it is expected a presence of spacial correlation structure among the values of $\boldsymbol{x}_j^{(P,\mathcal{H})}$.

Unfortunately, as pointed in Section 3.6.3, some intensity samples were not ML estimable under CR and ECR models. To overcome this trouble it is employed constrained estimation. The ML estimates were obtained using Corollary 3.11.1 assuming $\Theta_\beta = \mathbb{R}^+$ and Θ_λ as the range of λ estimates.

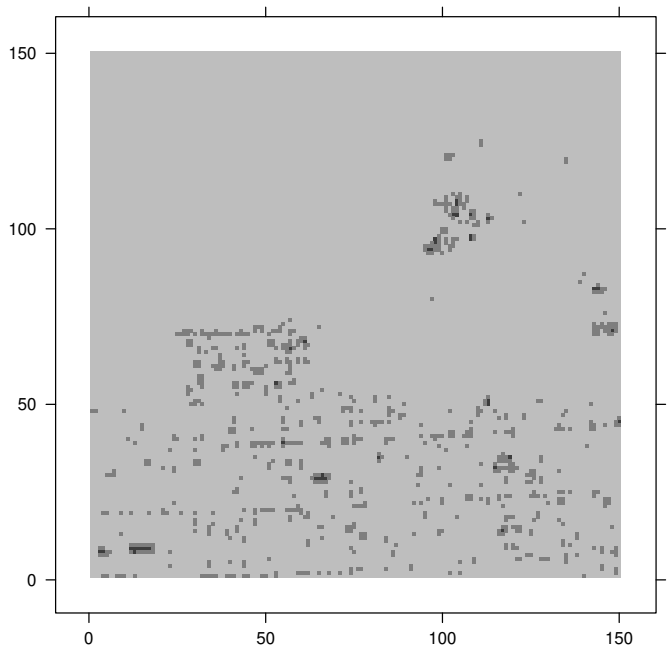


Figure 5.1: Segmentation obtained applying k-means directly on channel intensities.

It is possible to apply the k-means directly on polarization channel intensities regarding the feature vector

$$\boldsymbol{x}_j = (x_{j1},x_{j2},x_{j3})^\top,$$

where x_{jc} is the intensity of the j^{th} pixel in c^{th} channel, according indexes induced by (5.2) and (5.3). This method results in a problematic segmentation as illustrated in Figure 5.1. The high variability of the SAR intensities may be understood as one of the reasons for this poor performance.

5.3 THE APPLIED SMED MODELS

Pardo *et al.* (1997) obtained the asymptotic distribution of $\mathcal{H}_h^\phi(\hat{\boldsymbol{\theta}})$ where $\hat{\boldsymbol{\theta}}$ is a consistent asymptotically normal estimator of the parameter $\boldsymbol{\theta}$ obtained from a random sample of size n. This result can be expressed as

$$\sqrt{n}\left[\mathcal{H}_h^\phi(\hat{\boldsymbol{\theta}}) - \mathcal{H}_h^\phi(\boldsymbol{\theta})\right] \xrightarrow[n\to\infty]{d} \mathcal{N}\left(0,\sigma^2(\boldsymbol{\theta})\right), \tag{5.4}$$

provided $\sigma^2(\boldsymbol{\theta}) = \mathbf{T}^\top\mathbf{K}^{-1}\mathbf{T} > 0$ with $\mathbf{T}$ is the entropy gradient and $\mathbf{K}^{-1}$ is the inverse FIM given, respectively, by (2.23) and Corollary 3.15.2.

Let the three-dimensional random vector $\boldsymbol{X}_j^{(P,\mathcal{H})}$, which realizes a feature vector like (5.3). The applied model assumes that

- Supported by (5.4), the random vector $\boldsymbol{X}_j^{(P,\mathcal{H})}$ follows a distribution in the three-

variate normal class;

- The joint pdf of the feature vector associated with each individual texture, model P and entropy $\mathcal{H}$ is given by (2.36) following a g-component three-variate normal mixture distribution model.

These assumptions associate each pair $(P,\mathcal{H}) \in \mathfrak{M} \times \mathfrak{H}$ with its corresponding g-component mixture of stochastic entropy densities model, whose associated entropy $\mathcal{H}$ was defined using the model P ML estimates. The estimation process can be supported by the classical g-component normal mixtures as described by Hasselblad (1966), Day (1969), Behboodian (1970), Johns (1970) and Redner and Walker (1984).

A serious estimation problem arises from these assumptions: there are $p(p+1)/2$ free parameters for each component covariance matrix $\mathbf{\Sigma}_i$ in (2.36). Celeux and Govaert (1995) introduced parsimony into the normal FM covariance matrices by means of decompositions in the form $\mathbf{\Sigma}_i = \alpha_i \mathbf{D}_i \mathbf{A}_i \mathbf{D}_i^\top$, where $\mathbf{D}_i$ is the matrix of eigenvectors, $\mathbf{A}_i$ is a diagonal matrix with entries proportional to the eigenvalues, and α_i is the associated constant of proportionality (BANFIELD; RAFTERY, 1993). The resulting models are known as the *normal parsimonious clustering* (NPC) models; this family members have between one and $gp(p+1)/2$ free covariance parameters (BROWNE; MCNICHOLAS, 2014). As commented formerly, the entire NPC family was fitted and the models which optimize the ISP measures assumes a diagonal covariance matrix.

In the modeling of feature vectors obtained in Section 5.2 it is assumed the diagonal structure for the covariance-matrix motivated by the performance evaluation done in Section 5.4.

$$\mathbf{\Sigma}_i = \begin{bmatrix} \sigma_{i1}^2 & 0 & 0 \\ 0 & \sigma_{i2}^2 & 0 \\ 0 & 0 & \sigma_{i3}^2 \end{bmatrix}. \tag{5.5}$$

where σ_{ic}^2 is the variance associated to the intensity entropies in i^{th} segment for c^{th} polarization entropy channel according indexes induced by (5.2) and (5.3). This diagonal structure optimizes the ISP measures for all $(P,\mathcal{H})$-SMED models when compared to other NPC class elements. Evidences will be present at final.

Let $(\mathfrak{W}_j,P,\mathcal{H}) \in \mathfrak{W}_{5\times5} \times \mathfrak{M} \times \mathfrak{H}$, thus, the studied g-component $(P,\mathcal{H})$-SMED model for the i^{th} segment $(i = 1,\ldots,g)$ has pdf

$$\phi\left(\boldsymbol{x}_j^{(P,\mathcal{H})};\boldsymbol{\mu}_i,\mathbf{\Sigma}_i\right) \approx \frac{1}{(2\pi)^{3/2}} \left(\prod_{c=1}^{3} \sigma_{ic}^2\right)^{-1/2} \sum_{c=1}^{3} \left(\frac{x_{jc}-\mu_{ic}}{\sigma_{ic}}\right)^2,$$

where

$$\boldsymbol{\mu}_i = (\mu_{i1},\mu_{i2},\mu_{i3})$$

and

$$\boldsymbol{x}_j^{(P,\mathcal{H})} = (x_{j1},x_{j2},x_{j3}).$$

As pointed in (5.4), for $c = 1, 2, 3$

$$\sqrt{n}\,(X_{jc} - \mu_{jc}) \xrightarrow[n\to\infty]{d} \mathcal{N}\left(0,\sigma_{jc}^2\right) \Leftrightarrow \sqrt{n}\left(\begin{bmatrix} x_{j1} \\ x_{j2} \\ x_{j3} \end{bmatrix} - \begin{bmatrix} \mu_{i1} \\ \mu_{i2} \\ \mu_{i3} \end{bmatrix}\right) \xrightarrow[n\to\infty]{d} \mathcal{N}_3\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{i1}^2 & 0 & 0 \\ 0 & \sigma_{i2}^2 & 0 \\ 0 & 0 & \sigma_{i3}^2 \end{bmatrix}\right),$$

$$\sigma_{jc}^2 = \mathbf{T}^\top \mathbf{K}^{-1} \mathbf{T},$$

$$\mathbf{T} = \nabla \mathcal{H}_h^\phi(\boldsymbol{\theta}) = \left(\frac{\partial \mathcal{H}_h^\phi(\boldsymbol{\theta})}{\partial \beta}, \frac{\partial \mathcal{H}_h^\phi(\boldsymbol{\theta})}{\partial \lambda}\right)^\top,$$

where μ_{jc} is the expected value of the intensity stochastic entropy $\mathcal{H}$ regarding the probability model P in the c^{th} polarization channel according indexes induced by (5.2) and (5.3) on window $\mathfrak{W}_j$. The inverse FIM $\mathbf{K}^{-1}$ can be obtained from Corollary 3.15.2.

5.3.1 Initialization of model parameters

The efficiency of the EM algorithm in estimating the parameters is strongly influenced by the initial estimates of the model parameters. It was used the Hartigan and Wong (1979) version of the k-means algorithm as an initialization technique for the EM algorithm as implemented in R (2018). This specification of an initial parameter value is suggested by McLachlan and Peel (2000, p. 54). The employed implementation of the EM algorithm is due to Grün and Leisch (2008). The k-means is also used as a benchmark for the SMED methods.

5.3.2 SMED algorithm

The entries for the $(P,\mathcal{H})$-SMED procedures are:

Entry 1 A SAR intensities map for each polarization channel in the set $\mathfrak{C}_{(H,V)}$ as defined in (2.26);

Entry 2 The number of segments to be found on the image in the previous entry;

Entry 3 An (h,ϕ)-entropy in the set $\mathfrak{H}$ as defined in (5.1);

Entry 4 A probability model in the set $\mathfrak{M}$ as defined in (4.4).

The $(P,\mathcal{H})$-SMED algorithm has the following steps:

Step 1 Obtain the feature vectors by means of the procedure described in Section 5.2 considering the previous entries;

Step 2 Obtain the initial estimates of the model parameters using the classical k-means algorithm as described in Section 5.3.1;

Step 3 Obtain the refined ML estimates $\hat{\pi}_i$, $\hat{\boldsymbol{\mu}}_i$ and $\hat{\boldsymbol{\sigma}}_i^2$ of the corresponding model parameters π_i, $\boldsymbol{\mu}_i$ and $\boldsymbol{\sigma}_i^2$ for $i = 1, \ldots, g$ by means of the EM algorithm as described in Section 2.8 and the recursive formulas (2.35), (2.37) and (2.38);

Step 4 Assign the j^{th} pixel into the i^{th} region (segment) which corresponding (P,$\mathcal{H}$)-SMED component lf, $\mathcal{L}_j(\hat{\boldsymbol{\theta}}_i)$, is maximized

$$\mathcal{L}_j(\hat{\boldsymbol{\theta}}_i) = \phi\left(\boldsymbol{x}_j^{(P,\mathcal{H})}; \hat{\boldsymbol{\mu}}_i, \hat{\boldsymbol{\Sigma}}_i\right).$$

Considering the probability models in $\mathfrak{M}$ and the entropies in the set $\mathfrak{H}$ this algorithm incorporates eight segmentation methods named: 1. (CR ,$\mathcal{H}_S$)-SMED; 2. (CR ,$\mathcal{H}_R$)-SMED; 3. (CR ,$\mathcal{H}_T$)-SMED; 4. (CR ,$\mathcal{H}_A$)-SMED; 5. (ECR ,$\mathcal{H}_S$)-SMED; 6. (ECR ,$\mathcal{H}_R$)-SMED; 7. (ECR ,$\mathcal{H}_T$)-SMED; 8. (ECR ,$\mathcal{H}_A$)-SMED. This nomenclature make explicit the probability model P and entropy $\mathcal{H}$ which support the corresponding configuration of the (P,$\mathcal{H}$)-SMED algorithm. We will refer as (CR)-SMED and (ECR)-SMED to the SMEDs algorithms based on the CR and ECR model, respectively.

5.4 PERFORMANCE EVALUATION AND COMPARATIVE STUDY

To demonstrate the ability of the suggested models, texture segmentation is to be performed by using the same San Francisco SAR image described in Section 4.5. Some entropy vectors (feature vectors) obtained using the previously described procedure in Section 5.2 was depicted in Figures 5.2 and 5.3 considering disconnectedly the SAR polarization channels. These figures consider the Rényi entropy with $q = 0.55$, the Tsallis and Arimoto entropies with $q = 0.95$ and the Shannon entropy. The computations were done using the platform described in Section 1.2.

The visual aspect of the entropy intensities is depicted in Figure 5.2 and corroborates the existence of three textures as pointed in Figure 4.8 on Section 4.5.1. In general, the smallest entropies occur on sea zones while the greatest appear in urban regions. Note that this numerical aspect is aligned with the physical formation of SAR image. The energy returned by the under study region (when it is illuminated) is smaller in homogeneous targets (as ocean region) and larger in heterogeneous target (as urban and forest regions).

Figure 5.3 presents the corresponding intensity histograms of the entropy maps in Figure 5.2. The three modes in channel HH of the feature data suggest the possibility that the data have been drawn from a FM distribution with at least three components. HV and VV channels present two modes each and then, considering all polarization channels, the number of components can vary between three and twelve. However, the multi-modality

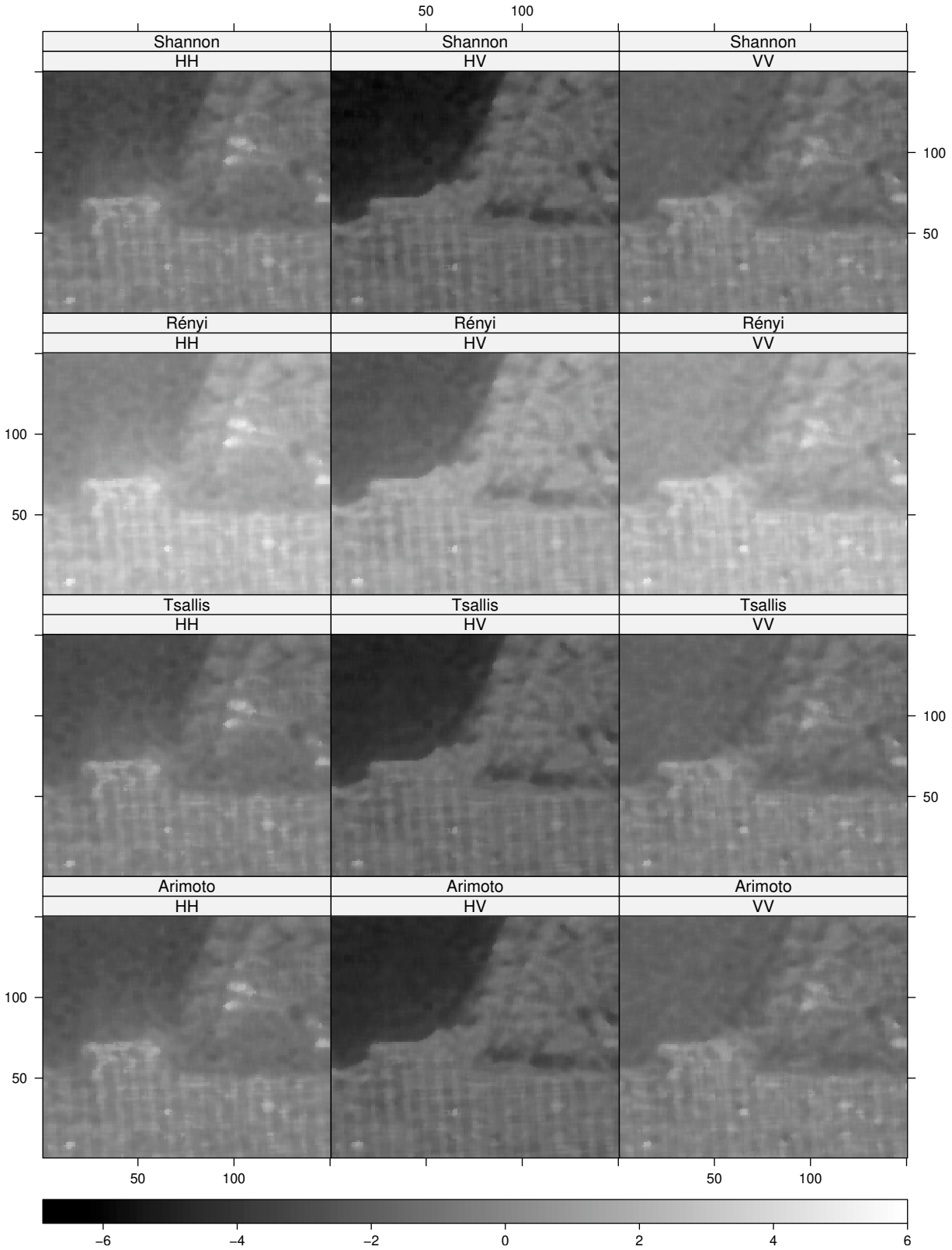


Figure 5.2: ECR entropy maps assuming $q = 0.55$ for Rényi and $q = 0.95$ for Tsallis and Arimoto.

Table 5.1: k-means based on the feature vectors centroids for q-optimized in $\mathfrak{H}$ ($k = 3$).

	model	Shannon	Rényi ($q = 0.55$)	Tsallis ($q = 0.95$)	Arimoto ($q = 0.95$)
$\boldsymbol{\mu}_1^\top$	CR	$-$(1.230,2.110,1.308)	(1.687,0.804,1.610)	$-$(1.078,1.880,1.150)	$-$(1.076,1.870,1.147)
$\boldsymbol{\mu}_2^\top$	CR	$-$(0.083,1.360,0.193)	(2.840,1.551,2.719)	$-$(0.036,1.220,0.078)	$-$(0.037,1.220,0.078)
$\boldsymbol{\mu}_3^\top$	CR	$-$(2.790,5.010,1.999)	(0.136, $-$ 2.081,0.927)	$-$(2.507,4.340,1.817)	$-$(2.497,4.310,1.813)
$\boldsymbol{\mu}_1^\top$	ECR	$-$(1.227,2.090,1.296)	(1.631,0.770,1.561)	$-$(1.083,1.870,1.148)	$-$(1.079,1.860,1.140)
$\boldsymbol{\mu}_2^\top$	ECR	$-$(0.102,1.370,0.217)	(2.756,1.480,2.641)	$-$(0.006,1.240,0.111)	$-$(0.008,1.240,0.110)
$\boldsymbol{\mu}_3^\top$	ECR	$-$(2.782,4.720,2.018)	(0.076, $-$ 1.870,0.839)	$-$(2.505,4.120,1.839)	$-$(2.495,4.100,1.830)

labels: 1=forest, 2=urban and 3=sea

Table 5.2: k-means based on feature vectors ISP measures for q optimized by corresponding (ECR,$\mathcal{H}$)-SMED models under CR and ECR laws ($k = 3$).

entropy	q	model	MCR	GCE	LCE	VOI	ARI
Shannon	-	CR	0.0690	0.0565	0.0225	0.4127	0.8287
Rényi	0.55	CR	0.0690	0.0565	0.0225	0.4127	0.8287
Tsallis	0.95	CR	0.0744	0.0590	0.0205	0.4291	0.8173
Arimoto	0.95	CR	0.0753	0.0597	0.0207	0.4328	0.8154
Shannon	-	ECR	0.0694	0.0553	0.0204	0.4057	0.8261
Rényi	0.55	ECR	0.0662	0.0533	0.0200	0.3940	0.8331
Tsallis	0.95	ECR	0.0771	0.0590	0.0187	0.4315	0.8097
Arimoto	0.95	ECR	0.0772	0.0591	0.0186	0.4319	0.8095

in histograms of the data does not always imply that the data are drawn from a FM distribution.

Considering the Figures 4.8, 5.2 and 5.3 and previous notes it was assumed $g = 3$ in the following segmentations. Table 5.1 presents the centroids for the k-means algorithm assuming $k = 3$ regarding the CR and ECR models. These centroids are used as EM initial values for the corresponding FM models as declared previously.

The measures discussed in Section 2.9 were employed to evaluate the clustering solutions. Table 5.2 summarizes the k-means ISP measures for (P,$\mathcal{H}$)-SMED models, which minimizes the MCR assuming three clusters. In general, a (P,$\mathcal{H}$)-SMED model which minimizes the MCR also presents optimum values for all ISP measures. The lowest MCR value was achieved by the ECR Rényi entropy with $q = 0.55$, which also achieves the minors GCE and VOI values and greater ARI value. It is interesting to note that the higher GCE values are associated to lower LCE ones.

The estimates of the q-optimized (P,$\mathcal{H}$)-SMED models can be checked in Table 5.3. The mean vectors of the SMED models based on the Shannon, Tsallis and Arimoto entropies are closer of each other than the correspondent Rényi entropy. This is expected because Tsallis and Arimoto entropies converges to Shannon when q goes to one as pointed in (2.17). The variance vectors of the Rényi-based mixture are, surprisingly, closer to Shannon-based mixture than Tsallis and Arimoto ones.

Figure 5.4 shows the segmented textures for the San Francisco image using k-means

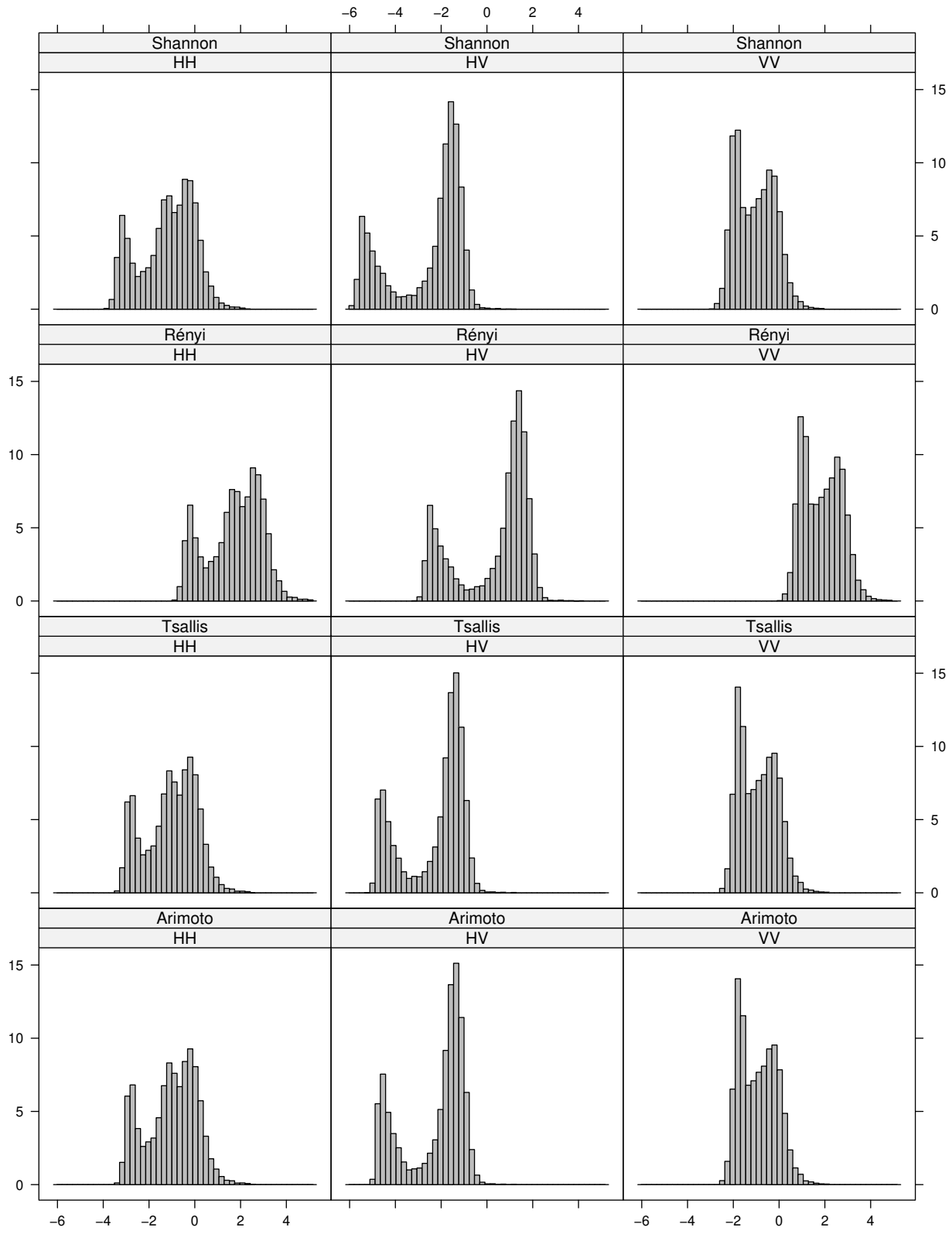


Figure 5.3: ECR entropy histograms assuming $q = 0.55$ for Rényi and $q = 0.95$ for Tsallis and Arimoto.

Table 5.3: q-optimized parameter estimates of the (ECR,$\mathcal{H}$)-SMED models ($g = 3$).

$\boldsymbol{\Psi}$	Shannon	Rényi ($q = 0.55$)	Tsallis ($q = 0.95$)	Arimoto ($q = 0.95$)
π_1	0.273	0.276	0.273	0.273
π_2	0.469	0.465	0.476	0.476
π_3	0.258	0.259	0.251	0.251
$\boldsymbol{\mu}_1^\top$	−(1.343,2.288,1.433)	(1.584,0.651,1.498)	−(1.231,2.116,1.315)	−(1.230,2.112,1.313)
$\boldsymbol{\mu}_2^\top$	−(0.188,1.409,0.288)	(2.744,1.505,2.632)	−(0.099,1.276,0.199)	−(0.099,1.274,0.199)
$\boldsymbol{\mu}_3^\top$	−(2.845,5.064,2.007)	(0.091, − 2.126,0.920)	−(2.588,4.417,1.827)	−(2.579,4.390,1.822)
$\boldsymbol{\sigma}_1^\top$	(0.124,0.458,0.128)	(0.127,0.429,0.129)	(0.110,0.402,0.112)	(0.110,0.399,0.111)
$\boldsymbol{\sigma}_2^\top$	(0.276,0.137,0.213)	(0.264,0.136,0.202)	(0.284,0.122,0.216)	(0.284,0.121,0.216)
$\boldsymbol{\sigma}_3^\top$	(0.225,0.206,0.053)	(0.227,0.210,0.054)	(0.158,0.112,0.043)	(0.155,0.109,0.043)

labels: 1=forest, 2=urban and 3=sea

and (P,$\mathcal{H}$)-SMED models. The implementations of the EM algorithm due to Grün and Leisch (2008) and Scrucca *et al.* (2016) obtained only two segments for all CR entropy measures, even after it was tried several initialization methods, thus, it was omitted in Figures 5.2 and 5.4 the plots relative to the CR model. This fact is an empirical evidence that the additional shape parameter β of the ECR model is imperative to make the entropy-based segmentation works using the EM algorithm. The k-means clusters present a large number of misclassified sea and forest zones, respectively, inside forest and urban regions than the corresponding FM-based segmentation models. On the other hand models of FM-based segmentations misclassify more forest zones inside sea regions than corresponding k-means clusters.

We used the AIC (AKAIKE, 1974), BIC (SCHWARZ, 1978; FRALEY; RAFTERY, 1998) and *integrated completed likelihood* (ICL) (BIERNACKI *et al.*, 2000) as model selection methods. Biernacki *et al.* (2000) showed that ICL appears to be more robust than AIC and BIC to violation of some of the mixture model assumptions and it can select a number of clusters leading to a sensible partitioning of the data. The ICL interpretation is analogous of AIC and BIC one as described in Section 3.7.

Table 5.4 condenses GoF and ISP measures for the (ECR,$\mathcal{H}$)-SMED. The smaller GoF measure values for Rényi-based FM segmentations are obtained when $q = 0.55$. Tables 5.2 and 5.4 reveal that this same Rényi-based segmentation model also achieves the smallest MCR, GCE and VOI and greatest ARI values among all regarded (ECR,$\mathcal{H}$)-SMED algorithms. For ECR-SMED under Tsallis and Arimoto entropies, the smallest GoF measures are obtained when $q = 0.75$, while the optimum ISP measures are obtained in higher q values. On comparing the ISP measures of k-means and (ECR,$\mathcal{H}$)-SMED models in Tables 5.2 and 5.4, it becomes clear that (ECR,$\mathcal{H}$)-SMED performance beats k-means one.

The assumption of the diagonal structure for the covariance-matrix $\boldsymbol{\Sigma}_i$ in (5.5) was motivated by the fact that the ISP measures for the San Francisco image achieve values closer of the optimum using this structure for all (ECR,$\mathcal{H}$)-SMED methods than with any

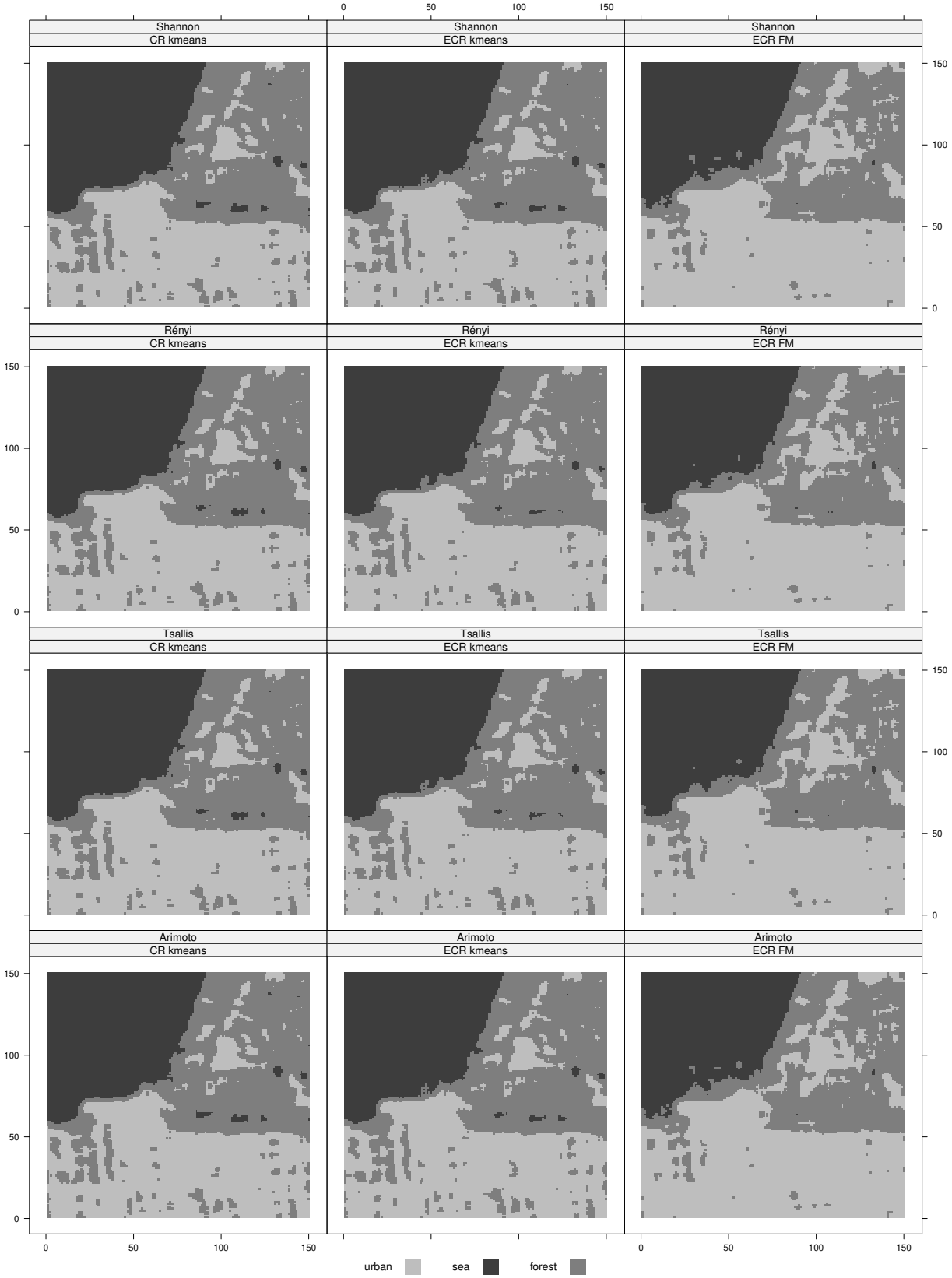


Figure 5.4: Segmentation maps based on entropies assuming $q = 0.55$ for Rényi and $q = 0.95$ for Tsallis and Arimoto.

Table 5.4: GoF and ISP measures of the textured image based on (ECR,$\mathcal{H}$)-SMED models regarding ECR entropies for various q values ($g = 3$).

entropy	q	AIC	BIC	ICL	MCR	GCE	LCE	VOI	ARI
Shannon	-	119412	119572	121472	0.0388	0.0382	0.0341	0.2845	0.8981
	0.55	**118568**	**118729**	**120596**	**0.0378**	**0.0372**	**0.0326**	0.2805	**0.9006**
	0.60	118721	118882	120753	0.0379	0.0373	0.0329	0.2808	0.9004
	0.65	118861	119022	120897	0.0379	0.0373	0.0329	**0.2801**	0.9005
	0.70	118986	119147	121026	0.0379	0.0374	0.0336	0.2805	0.9004
Rényi	0.75	119095	119256	121138	0.0380	0.0375	0.0336	0.2812	0.9000
	0.80	119188	119349	121235	0.0383	0.0378	0.0337	0.2826	0.8993
	0.85	119266	119427	121316	0.0383	0.0378	0.0338	0.2826	0.8994
	0.90	119329	119489	121383	0.0384	0.0379	0.0342	0.2825	0.8991
	0.95	119377	119538	121434	0.0386	0.0381	0.0343	0.2839	0.8985
	0.55	206236	206396	208310	0.0421	0.0401	**0.0308**	0.3194	0.8933
	0.60	138895	139055	140986	0.0413	0.0398	0.0323	0.3148	0.8949
	0.65	109876	110036	111980	0.0410	0.0398	0.0340	0.3125	0.8957
	0.70	96350	96510	98465	0.0408	0.0399	0.0357	0.3101	0.8959
Tsallis	0.75	**91211**	**91372**	**93329**	0.0394	0.0389	0.0353	0.3002	0.8989
	0.80	91322	91483	93439	0.0393	0.0390	0.0356	0.2987	0.8989
	0.85	95022	95183	97144	0.0391	0.0389	0.0357	0.2965	0.8992
	0.90	101329	101489	103457	0.0388	**0.0382**	0.0360	0.2894	**0.9003**
	0.95	109606	109766	111720	**0.0384**	0.0385	0.0365	**0.2888**	0.8986
	0.55	285565	285726	287503	0.0684	0.0518	**0.0158**	0.4146	0.8274
	0.60	156778	156938	158799	0.0521	0.0449	0.0220	0.3627	0.8675
	0.65	107505	107666	109570	0.0453	0.0418	0.0276	0.3351	0.8850
	0.70	87943	88104	90034	0.0421	0.0402	0.0310	0.3190	0.8928
Arimoto	0.75	**82542**	**82702**	**84659**	0.0407	0.0396	0.0338	0.3105	0.8962
	0.80	84621	84781	86742	0.0398	0.0392	0.0357	0.3024	0.8978
	0.85	90835	90995	92955	0.0392	0.0390	0.0356	0.2980	0.8990
	0.90	99350	99510	101477	0.0387	0.0385	0.0361	0.2913	**0.8997**
	0.95	109093	109254	111209	**0.0386**	**0.0383**	0.0365	**0.2877**	0.8991

other structure in the NPC class.

5.5 CONCLUDING REMARKS

This final chapter has been dedicated to the *Cauchy-Rayleigh* (CR) and *exponentiated Cauchy-Rayleigh* (ECR) entropy-based segmentation techniques called (P,$\mathcal{H}$)-*segmentation based on mixture of stochastic entropy distributions* (SMED) methods. Generally, evidences have indicated that San Francisco *synthetic aperture radar* (SAR) imagery segmentation by using channel intensities as input did not present satisfactory results, thus, more sophisticated methodologies have been required. These reasons make plausible to model the entropies associated to each SAR imagery pixel using the FM of multivariate normal densities structures. The feature vectors required by the (P,$\mathcal{H}$)-SMED method can be also used by the standard k-means algorithm, thus, it was used as an *image segmentation performance* (ISP) measures reference. The (ECR ,$\mathcal{H}$)-SMEDs ISP measures indicated best segmentations when compared to its corresponding k-means ones. It was not possible to fit any (CR ,$\mathcal{H}$)-SMEDs assuming three segments, which provides empirical evidence that the ECR additional shape parameter β is imperative to make the (ECR ,$\mathcal{H}$)-SMEDs method works. The order parameter of q-entropies influenced in the segmentation performances, the optimum ISP measure values were, in general, obtained assuming the Rényi with $q = 0.55$ in the (ECR ,$\mathcal{H}$)-SMED model. The order parameter of q-entropies presented optimum GoF measures when $q = 0.95$ for Tsallis and Arimoto entropies and $q = 0.55$ for the Rényi entropy.

6 CONCLUSION

We deepened a discussion about an extended *Cauchy-Rayleigh* (CR) distribution, particular case of some known models like heavy-tailed Rayleigh (NIKIAS; SHAO, 1995), generalized Feller-Pareto (ZANDONATTI, 2001), exponentiated Burr type-XII (AL-HUSSAINI; AHSANULLAH, 2015), Kumaraswamy Burr type-XII (PARANAÍBA, 2012) and McDonald Burr type-XII (GOMES *et al.*, 2015). We refereed to it as *exponentiated Cauchy-Rayleigh* (ECR) distribution. Some of its mathematical properties was derived and discussed: Closed-form expressions for its moments and Shannon, Rényi, Tsallis and Arimoto entropies was found. The ECR model obeyed the property of regularly varying at infinity and can take decreasing, decreasing-increasing-decreasing and upside-down bathtub-shape hazard rate functions, confirming its usefulness to describe lifetime data. We provided procedures to estimate the ECR parameters through original and bias-corrected *maximum likelihood estimators* (MLEs) and a percentile-based method. An important characteristic of this new model was that the mode, median, moments, Fisher information components, its inverse components and the second order bias can be decomposed in two factors where the first factor was function of the scale parameter λ and the second one was function of the new shape parameter β. A simulation study to assess proposed estimators was performed. Each procedure revealed advantages over specific parameter points and sample sizes. The ECR usefulness was illustrated through an application to real data. Results indicated that our proposal can furnish better performance than classical lifetime models like gamma, Birnbaum-Saunders, Weibull and log-normal. The formulae related with the new model was manageable and turned into adequate tools comprising the arsenal of applied statistics.

It was proposed strategies to quantify differences in *synthetic aperture radar* (SAR) urban patterns by means of its intensities using the CR and ECR models. These differences were detected using the classes of hypotheses tests: *likelihood ratio tests* (LRTs) and *entropy-based tests* (EBTs). In general, all regarded test underestimates the corresponding nominal level for small sample sizes, but as long as the sample size increases these corresponding nominal level are achieved as expected. The ECR model presented slightly smaller *goodness-of-fit* (GoF) measures than CR one in all polarization channels of urban areas of the studied San Francisco SAR image. The texture identification study for the San Francisco urban patterns showed that the CR LRTs and the ECR EBTs had rejection rates closer to the expected nominal level. The final study attested that all tests are able to detect the known information redundancy between the channels HH and VV, but the CR LRTs and the ECR EBTs presented the highest rejection rates when applied to detect information redundancy between the channels pairs HH-HV and HV-VV.

The final study defined the *segmentation based on mixture of stochastic entropy*

distributions (SMED) methods which consist in CR and ECR entropies-based segmentation techniques. Generally, evidences indicated that the studied San Francisco SAR imagery segmentation supported directly by channel intensities did not present satisfactory results, thus more sophisticated methodologies were required. The studied solution was based on the classical *finite mixture* (FM) of multivariate normal densities (MCLACHLAN; PEEL, 2000, ch. 3) and in the result of Pardo *et al.* (1997), which assured that (h,ϕ)-entropies are asymptotically distributed. These motivations made plausible to model the entropies associated to each SAR imagery pixel using the FM of multivariate normal densities structures. The feature vectors required by the $(P,\mathcal{H})$-SMED method can be also used by the standard k-means algorithm, thus it was used as an *image segmentation performance* (ISP) measures reference. The $(\mathrm{ECR},\mathcal{H})$-SMEDs ISP measures indicated best segmentations when compared with its corresponding k-means ones. It was not possible to fit any $(\mathrm{CR},\mathcal{H})$-SMEDs assuming three segments, which provided empirical evidence that the ECR additional shape parameter β was imperative to make the $(\mathrm{ECR},\mathcal{H})$-SMEDs method works. The q-entropies order parameter had influenced in the segmentation performances and the optimum ISP measure values were, in general, obtained assuming the ECR Rényi with $q = 0.55$.

6.1 FUTURE WORKS

We hope that the broached model may attract wider applications for modeling positive real data. A natural sequence of this work is the use of ECR constrained MLEs which can overcome the non-existence of maximum likelihood estimates as reported in Section 4.5.2. The constrained MLEs also can make feasible the use of the Cox-Snell corrected-MLEs as defined in Section 3.5.5. This thesis has also raised several points, which require deeper studies. Some of them are:

1. Develop the log-ECR regression model;
2. Improve the notes of censored estimators using the ECR;
3. Explore other methods to obtain estimators;
4. Extend the CR/ECR aiming to describe non-urban SAR imagery;
5. Develop the ECR-FM models.

REFERENCES

AARSET, M. V. How to identify a bathtub hazard rate. **IEEE Transactions on Reliability**, R-36, n. 1, p. 106–108, 1987.

ABDEL-KHALEK, S. *et al.* A two-dimensional image segmentation method based on genetic algorithm and entropy. **Optik**, v. 131, p. 414–422, 2017.

ABRAMOWITZ, M.; STEGUN, I. A. (Ed.). **Handbook of Mathematical Functions**. 10. ed. Washington, D.C.: NBS, 1972. 1037 p. (National Bureau of Standards Applied Mathematics, 55).

ADEYEMI, S.; ADEBANJI, T. The exponentiated generalized Pareto distribution. **Ife Journal of Science**, v. 6, n. 2, 2004.

AGHABABAEE, H.; NIAZMARDI, S.; AMINI, J. Urban area extraction in SAR data. **International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences**, v. 15, n. 1, p. 5–8, 2013.

AGUILAR, D. L. F. d. **Estimação Pontual e Intervalar dos Parâmetros da Distribuição Lomax**. Master Thesis — Universidade Federal de Pernambuco, Recife, Fev 2012.

AHSANULLAH, M.; NEVZOROV, V. B.; SHAKIL, M. **An Introduction to Order Statistics**. Paris: Atlantis Press, 2013. (Atlantis Studies in Probability and Statistics, v. 3).

AHUJA, J. C.; NASH Stanley W. The generalized Gompertz-Verhulst family of distributions. **Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)**, Springer, v. 29, n. 2, p. 141–156, 1967.

AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, v. 19, n. 6, p. 716–723, 1974.

AKBULUT, Y. *et al.* An effective color texture image segmentation algorithm based on hermite transform. **Applied Soft Computing**, v. 67, p. 494–504, 2018.

AL-HUSSAINI, E. K. Inference based on censored samples from exponentiated populations. **TEST**, v. 19, n. 3, p. 487–513, 2010.

AL-HUSSAINI, E. K.; AHSANULLAH, M. **Exponentiated Distributions**. Paris: Atlantis Press, 2015. (Atlantis Studies in Probability and Statistics, v. 5).

ALI, M. M.; PAL, M.; WOO, J. Some exponentiated distributions. **The Korean Communications in Statistics**, v. 14, n. 1, p. 93–109, 2007.

ANDERSON, T. W.; DARLING, D. A. Asymptotic theory of certain goodness-of-fit criteria based on stochastic processes. **The Annals of Mathematical Statistics**, The Institute of Mathematical Statistics, v. 23, n. 2, p. 193–212, 1952.

ARIMOTO, S. Information-theoretical considerations on estimation problems. **Information and Control**, v. 19, n. 3, p. 181–194, 1971.

ARNOLD, B. C.; BALAKRISHNAN, N.; NAGARAJA, H. N. **Records**. New York: Wiley, 1998.

ARNOLD, B. C.; LAGUNA, L. **On Generalized Pareto Distributions with Applications to Income Data**. Iowa: Department of Economics, Iowa State University, 1977. (International Studies in Economics).

AUSHERMAN, D. A. *et al.* Developments in radar imaging. **IEEE Transactions on Aerospace and Electronic Systems**, v. 20, n. 4, p. 363–400, 1984.

AZMEDROUB, B.; OUARZEDDINE, M.; SOUISSI, B. Extraction of urban areas from polarimetric SAR imagery. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 9, n. 6, p. 2583–2591, 2016.

BAILEY, W. N. **Generalized Hypergeometric Series**. New York: Hafner Publishing Co Ltd, 1973. 108 p. p. (Cambridge Tracts in Mathematics).

BALLERINI, R.; RESNICK, S. I. Records in the presence of a linear trend. **Advances in Applied Probability**, Applied Probability Trust, v. 19, n. 4, p. 801–828, 1987.

BANFIELD, J. D. D.; RAFTERY, A. E. Model-based Gaussian and non-Gaussian clustering. **Biometrics**, v. 49, n. 3, p. 803–821, 1993.

BEBBINGTON, M.; LAI, C.; ZITIKIS, R. A flexible Weibull extension. **Reliability Engineering & System Safety**, v. 92, n. 6, p. 719–726, 2007.

BEHBOODIAN, J. On a mixture of normal distributions. **Biometrika**, Oxford University Press, Biometrika Trust, v. 57, n. 1, p. 215–217, 1970.

BIBALAN, M. H.; AMINDAVAR, H. A new alpha and gamma based mixture approximation for heavy-tailed Rayleigh distribution. In: **IEEE International Conference on Acoustics, Speech and Signal Processing**. Brisbane, Australia: IEEE, 2015. p. 3711–3715.

_____. Non-Gaussian amplitude pdf modeling of ultrasound images based on a novel generalized Cauchy-Rayleigh mixture. **EURASIP Journal on Image and Video Processing**, n. 1, p. 48–50, 2016.

BICKEL, P. J.; DOKSUM, K. A. **Mathematical Statistics**: Basic ideas and selected topics. 2. ed. Prentice Hall, 2000. 556 p.

BIERNACKI, C.; CELEUX, G.; GOVAERT, G. Assessing a mixture model for clustering with the integrated completed likelihood. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, IEEE Computer Society, Washington, DC, USA, v. 22, n. 7, p. 719–725, 2000.

BIRNBAUM, Z. W.; SAUNDERS, S. C. A new family of life distributions. **Journal of Applied Probability**, Cambridge University Press, v. 6, n. 2, p. 319–327, 1969.

BROWNE, R. P.; MCNICHOLAS, P. D. Estimating common principal components in high dimensions. **Advances in Data Analysis and Classification**, v. 8, n. 2, p. 217–226, 2014.

BURBEA, J.; RAO, C. R. Entropy differential metric, distance and divergence measures in probability spaces: A unified approach. **Journal of Multivariate Analysis**, v. 12, n. 4, p. 575–596, 1982.

_____. On the convexity of higher order Jensen differences based on entropy functions (corresp.). **IEEE Transactions on Information Theory**, v. 28, n. 6, p. 961–963, 1982.

_____. On the convexity of some divergence measures based on entropy functions. **IEEE Transactions on Information Theory**, v. 28, n. 3, p. 489–495, 1982.

BURY, K. **Statistical Distributions in Engineering**. Cambridge: Cambridge University Press, 1999.

BUTLER, R. J.; MCDONALD, J. B. Using incomplete moments to measure inequality. **Journal of Econometrics**, v. 42, n. 1, p. 109–119, 1989.

CELEUX, G.; GOVAERT, G. Gaussian parsimonious clustering models. **Pattern Recognition**, v. 28, n. 5, p. 781–793, 1995.

CHEN, Z. A new two-parameter lifetime distribution with bathtub shape or increasing failure rate function. **Statistics & Probability Letters**, v. 49, n. 2, p. 155–161, 2000.

CHISTYAKOV, V. P. A theorem on sums of independent positive random variables and its applications to branching random processes. **Theory of Probability & its Applications**, v. 9, n. 4, p. 640–648, 1964.

COMON, P. Independent component analysis, a new concept? **Signal Processing**, v. 36, n. 3, p. 287–314, 1994.

COOKE, R. M.; NIEBOER, D.; MISIEWICZ, J. Regularly varying and subexponential distributions. In: **Fat-Tailed Distributions**. John Wiley & Sons, Inc., 2014. v. 1, cap. 4, p. 49–63.

COORAY, K.; ANANDA, M. M. A. A generalization of the half-normal distribution with applications to lifetime data. **Communications in Statistics - Theory and Methods**, v. 37, n. 9, p. 1323–1337, 2008.

CORDEIRO, G. M.; KLEIN, R. Bias correction in ARMA models. **Statistics & Probability Letters**, v. 19, n. 3, p. 169–176, 1994.

CORDEIRO, G. M. *et al.* The McBurr XII and log-McBurr XII models with applications to lifetime data. **International Journal of Statistics and Probability**, v. 5, n. 1, p. 1, 2016.

CORDEIRO, G. M.; ORTEGA, E. M. M.; SILVA, G. O. The exponentiated generalized gamma distribution with application to lifetime data. **Journal of Statistical Computation and Simulation**, v. 81, n. 7, p. 827–842, 2011.

CORNELL, J. **Experiments with Mixtures**: Designs, models, and the analysis of mixture data. 3. ed. New York: John Wiley & Sons, Inc, 2002. (Wiley Series in Probability and Statistics).

COVER, T. M.; THOMAS, J. A. **Elements of Information Theory**. 2nd ed. ed. New York, NY: Wiley-Interscience, 2006. 748 p. (Wiley Series in Telecommunications and Signal Processing).

COX, D. R.; SNELL, E. J. A general definition of residuals. **Journal of the Royal Statistical Society. Series B (Methodological)**, Royal Statistical Society, Wiley, v. 30, n. 2, p. 248–275, 1968.

CRAMÉR, H. On the composition of elementary errors. **Scandinavian Actuarial Journal**, Taylor & Francis, v. 1928, n. 1, p. 13–74, 1928.

CRIBARI-NETO, F.; VASCONCELLOS, K. L. P. Nearly unbiased maximum likelihood estimation for the Beta distribution. **Journal of Statistical Computation and Simulation**, Taylor & Francis, v. 72, n. 2, p. 107–118, 2002.

CROWLEY, J.; HU, M. Covariance analysis of heart transplant survival data. **Journal of the American Statistical Association**, American Statistical Association, Taylor & Francis, Ltd., v. 72, n. 357, p. 27–36, 1977.

DAY, N. E. Estimating the components of a mixture of normal distributions. **Biometrika**, Oxford University Press, Biometrika Trust, v. 56, n. 3, p. 463–474, 1969.

DEERWESTER, S. *et al.* Indexing by latent semantic analysis. **Journal of the American society for information science**, Wiley Online Library, v. 41, n. 6, p. 391–407, 1990.

DELIGNON, Y.; PIECZYNSKI, W. Modeling non-Rayleigh speckle distribution in SAR images. **IEEE Transaction on Geoscience and Remote Sensing**, v. 40, n. 6, p. 1430–1435, 2002.

DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the EM algorithm. **Journal of the Royal Statistical Society. Series B (Methodological)**, Royal Statistical Society, Wiley, v. 39, n. 1, p. 1–38, 1977.

DOUGHERTY, J. P. Lectures on gas theory. by ludwig boltzmann. translated by stephen g. brush. university of california press, 1964. 490 pp. **Journal of Fluid Mechanics**, Cambridge University Press, v. 20, n. 4, p. 698–699, 1964.

DUBEY, S. D. A compound Weibull distribution. **Naval Research Logistics Quarterly**, Wiley Subscription Services, Inc., A Wiley Company, v. 15, n. 2, p. 179–188, 1968.

EVERITT, B. S.; HAND, D. J. **Finite Mixture Distributions**. London: Chapman and Hall, 1981. 143 p. (Monographs on Applied Probability and Statistics).

FAMA, E. F.; SCHWERT, G. W. Human capital and capital market equilibrium. **Journal of Financial Economics**, v. 4, n. 1, p. 95–125, 1977.

FELLER, W. **An Introduction to Probability and its Applications**. 2. ed. New York: John Wiley & Sons, Inc., 1971. 525 p.

FISK, P. R. The graduation of income distributions. **Econometrica**, Wiley, Econometric Society, v. 29, n. 2, p. 171–185, 1961.

FOSS, S.; KORSHUNOV, D.; ZACHARY, S. **An Introduction to Heavy-Tailed and Subexponential Distributions**. New York: Springer, 2011. (Springer Series in Operations Research and Financial Engineering 38).

FRALEY, C.; RAFTERY, A. E. How many clusters? which clustering method? answers via model-based cluster analysis. **The Computer Journal**, v. 41, n. 8, p. 578–588, 1998.

FRERY, A. C.; CINTRA, R. J.; NASCIMENTO, A. D. C. Entropy-based statistical analysis of PolSAR data. **IEEE Transaction on Geoscience and Remote Sensing**, v. 51, n. 6, p. 3733–3743, 2013.

FRERY, A. C.; CRIBARI-NETO, F.; SOUZA, M. O. de. Analysis of minute features in speckled imagery with maximum likelihood estimation. **EURASIP Journal on Advances in Signal Processing**, v. 2004, n. 16, p. 375370, 2004.

GARDINER, J. C. *et al.* Fitting heavy-tailed distributions to health care data by parametric and bayesian methods. **Journal of Statistical Theory and Practice**, v. 8, n. 4, p. 619–652, 2014.

GILES, D. E. Bias reduction for the maximum likelihood estimators of the parameters in the half-Logistic distribution. **Communications in Statistics - Theory and Methods**, Taylor & Francis, v. 41, n. 2, p. 212–222, 2012.

GILES, D. E.; FENG, H. **Bias of the maximum likelihood estimators of the two-parameter gamma distribution revisited**. Victoria, CA, 2009.

GILES, D. E.; FENG, H.; GODWIN, R. T. On the bias of the maximum likelihood estimator for the two-parameter Lomax distribution. **Communications in Statistics - Theory and Methods**, Taylor & Francis, v. 42, n. 11, p. 1934–1950, 2013.

GINI, C. Measurement of inequality of incomes. **The Economic Journal**, Royal Economic Society, Wiley, v. 31, n. 121, p. 124–126, 1921.

GODWIN, R. T. Bias reduction for the maximum likelihood estimator of the doubly-truncated Poisson distribution. **Communications in Statistics - Theory and Methods**, v. 45, n. 7, p. 1887–1901, 2016.

GOLUB, G.; KAHAN, W. Calculating the singular values and pseudo-inverse of a matrix. **Journal of the Society for Industrial and Applied Mathematics, Series B: Numerical Analysis**, SIAM, v. 2, n. 2, p. 205–224, 1965.

GOMES, A. E.; SILVA, C. Q.; CORDEIRO, G. M. Two extended Burr models: Theory and practice. **Communications in Statistics - Theory and Methods**, v. 44, n. 8, p. 1706–1734, 2015.

GOMPERTZ, B. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. **Philosophical Transactions of the Royal Society of London**, The Royal Society, v. 115, p. 513–583, 1825.

GRADSHTEYN, I. S.; RYZHIK, I. M. **Table of Integrals, Series and Products**. 7. ed. Burlington: Academic Press, 2007. 1221 p.

GRÜN, B.; LEISCH, F. FlexMix version 2: Finite mixtures with concomitant variables and varying and constant parameters. **Journal of Statistical Software, Articles**, v. 28, n. 4, p. 1–35, 2008.

GUPTA, M. R.; CHEN, Y. Theory and use of the EM algorithm. **Foundations and Trends in Signal Processing**, v. 4, n. 3, p. 223–296, 2010.

GUPTA, R. C.; GUPTA, P. L.; GUPTA, R. D. Modeling failure time data by Lehman alternatives. **Communications in Statistics - Theory and Methods**, v. 27, n. 4, p. 887–904, 1998.

GUPTA, R. D.; KUNDU, D. Generalized exponential distributions. **Australian & New Zealand Journal of Statistics**, Blackwell Publishers Ltd, v. 41, n. 2, p. 173–188, 1999.

_____. Generalized exponential distribution: Different method of estimations. **Journal of Statistical Computation and Simulation**, v. 69, n. 4, p. 315–337, 2001.

_____. Generalized exponential distribution: Statistical inferences. Unpublished. 2001.

HANNAN, E. J.; QUINN, B. G. The determination of the order of an autoregression. **Journal of the Royal Statistical Society. Series B (Methodological)**, Royal Statistical Society, Wiley, v. 41, n. 2, p. 190–195, 1979.

HARRISON, G. W.; MILLARD, P. H. Balancing acute and long-term care: The mathematics of throughput in departments of geriatric medicine. **Methods of Information in Medicine**, v. 30, n. 3, p. 221–228, 1991.

HARTIGAN, J. A.; WONG, M. A. Algorithm AS 136: A k-means clustering algorithm. **Journal of the Royal Statistical Society. Series C (Applied Statistics)**, Wiley, Royal Statistical Society, v. 28, n. 1, p. 100–108, 1979.

HASSELBLAD, V. Estimation of parameters for a mixture of normal distributions. **Technometrics**, Taylor & Francis, Ltd., American Statistical Association, American Society for Quality, v. 8, n. 3, p. 431–444, 1966.

HAVRDA, J.; CHARVÂT, F. Quantification method of classification processes: Processes concept of structural α-entropy. **Kybernetica**, v. 1, n. 3, p. 30–35, 1967.

HILL, M. O. Diversity and evenness: A unifying notation and its consequences. **Ecology**, v. 54, n. 2, p. 427–432, 1973.

HILL, P. R. *et al.* Dual-tree complex wavelet coefficient magnitude modelling using the bivariate Cauchy–Rayleigh distribution for image denoising. **Signal Processing**, v. 105, p. 464–472, 2014.

HOFMANN, G.; NAGARAJA, H. N. Random and Poisson paced record models in the F^{α} setup. **Journal of Applied Probability**, Applied Probability Trust, v. 37, n. 2, p. 374–388, 2000.

HOTELLING, H. Relations between two sets of variates. **Biometrika**, Oxford University Press, Biometrika Trust, v. 28, n. 3/4, p. 321–377, 1936.

HSIN, H. Texture segmentation using modulated wavelet transform. **IEEE Transactions on Image Processing**, v. 9, n. 7, p. 1299–1302, 2000.

HUAN, Z.; HOU, Y. An segmentation algorithm of texture image based on DWT. In: **Fourth International Conference on Natural Computation**. Jinan, China: IEEE, 2008. v. 5, p. 433–436.

HUBERT, L.; ARABIE, P. Comparing partitions. **Journal of Classification**, v. 2, n. 1, p. 193–218, 1985.

HURVICH, C. M.; TSAI, C. Regression and time series model selection in small samples. **Biometrika**, v. 76, n. 2, p. 297–307, 1989.

HYNDMAN, R. J.; FAN, Y. Sample quantiles in statistical packages. **The American Statistician**, American Statistical Association, Taylor & Francis, Ltd., v. 50, n. 4, p. 361–365, 1996.

IZMAILOV, A.; SOLODOV, M. **Otimização**: Condições de otimalidade, elementos de análise convexa e de dualidade. 2. ed. Rio de Janeiro: Instituto Nacional de Matemática Pura e Aplicada, 2009. 253 p.

JAIN, A. K.; FARROKHNIA, F. Unsupervised texture segmentation using gabor filters. In: **IEEE International Conference on Systems, Man, and Cybernetics Conference Proceedings**. Los Angeles, USA: IEEE, 1990. p. 14–19.

JI, Y. *et al.* Earthquake/tsunami damage level mapping of urban areas using full polarimetric SAR data. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 11, n. 7, p. 2296–2309, 2018.

JOHNS, S. On identifying the population of origin of each observation in a mixture of observations from two normal populations. **Technometrics**, Taylor & Francis, v. 12, n. 3, p. 553–563, 1970.

JOHNSON, N. L.; KOTZ, S.; BALAKRISHNAN, N. **Continuous Univariate Distributions**. 2. ed. New Jersey: John Wiley & Sons, Inc., 1995. 769 p. (Wiley Series in Probability and Mathematical Statistics, v. 2).

JORDAN, M. I.; XU, L. Convergence results for the EM approach to mixtures of experts architectures. **Neural Networks**, v. 8, n. 9, p. 1409–1431, 1995.

JOST, L. Entropy and diversity. **Oikos**, v. 113, n. 2, p. 363–375, 2006.

JYOTHIRMAYI, T. *et al.* Studies on image segmentation integrating generalized Laplace mixture model and hierarchical clustering. **International Journal of Computer Applications**, v. 128, n. 12, 2015.

KAROUI, I. *et al.* Variational region-based segmentation using multiple texture statistics. **IEEE Transactions on Image Processing**, v. 19, n. 12, p. 3146–3156, 2010.

KENNEY, J. F.; KEEPING, E. S. **Mathematics of Statistics**. 3. ed. Van Nostrand, New Jersey: Chapman & Hall Ltd., 1962.

KLEIBER, C.; KOTZ, S. **Statistical Size Distributions in Economics and Actuarial Sciences**. New Jersey: John Wiley & Sons, Inc., 2003. (Wiley Series in Probability and Statistics).

KLUGMAN, S. A.; PANJER, H. H.; WILLMOT, G. E. **Loss Models**: From data to decisions. 4. ed. New Jersey: John Wiley & Sons, Inc., 2012. (Wiley Series in Probability and Statistics).

KUMAR, K. N. *et al.* Studies on texture segmentation using D-dimensional generalized Gaussian distribution integrated with hierarchical clustering. **International Journal of Image, Graphics and Signal Processing**, v. 8, n. 3, p. 45–54, 2016.

KUNDU, D.; RAQAB, M. Z. Generalized Rayleigh distribution: Different methods of estimations. **Computational Statistics & Data Analysis**, v. 49, n. 1, p. 187–200, 2005.

KURUOGLU, E. E.; ZERUBIA, J. Modeling SAR images with a generalization of the Rayleigh distribution. **Image Processing, IEEE Transactions on**, v. 13, n. 4, p. 527–533, 2004.

LAGOS-ÁLVAREZ, B.; JIMÉNEZ-GAMERO, M. D.; ALBA-FERNÁNDEZ, V. Bias correction in the type I generalized Logistic distribution. **Communications in Statistics - Simulation and Computation**, Taylor & Francis, v. 40, n. 4, p. 511–531, 2011.

LATHAUWER, L.; MOOR, B.; VANDEWALLE, J. A multilinear singular value decomposition. **SIAM Journal on Matrix Analysis and Applications**, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, v. 21, n. 4, p. 1253–1278, 2000.

LEE, J.; POTTIER, E. **Polarimetric Radar Imaging**: From basics to applications. New York: CRC Press, 2009. (Optical Science and Engineering).

LEHMANN, E. L.; ROMANO, J. P. **Testing Statistical Hypotheses**. 3. ed. New York: Springer-Verlag, 2005. 784 p. (Springer Texts in Statistics).

LEMONTE, A. J. Improved point estimation for the Kumaraswamy distribution. **Journal of Statistical Computation and Simulation**, Taylor & Francis, v. 81, n. 12, p. 1971–1982, 2011.

_____. A new exponential-type distribution with constant, decreasing, increasing, upside-down bathtub and bathtub-shaped failure rate function. **Computational Statistics & Data Analysis**, v. 62, p. 149–170, 2013.

_____. The beta log-logistic distribution. **Brazilian Journal of Probability and Statistics**, Brazilian Statistical Association, v. 28, n. 3, p. 313–332, 2014.

LEMONTE, A. J.; BARRETO-SOUZA, W.; CORDEIRO, G. M. The exponentiated Kumaraswamy distribution and its log-transform. **Brazilian Journal of Probability and Statistics**, v. 27, n. 1, p. 31–53, 2013.

LEMONTE, A. J.; CORDEIRO, G. M. The exponentiated generalized inverse Gaussian distribution. **Statistics & Probability Letters**, v. 81, n. 4, p. 506–517, 2011.

LEMONTE, A. J.; CRIBARI-NETO, F.; VASCONCELLOS, K. L. P. Improved statistical inference for the two-parameter Birnbaum–Saunders distribution. **Computational Statistics & Data Analysis**, v. 51, n. 9, p. 4656 – 4681, 2007.

LÉVY, P. **Calcul des Probabilités**. Paris: Gauthier-Villars, 1925. (PCMI collection).

LI, X.; EKMAN, T. Cauchy power azimuth spectrum for clustered radio propagation MIMO channel model. In: **IEEE 72nd Vehicular Technology Conference - Fall**. Ottawa, Canada: IEEE, 2010. p. 1–5.

_____. Cauchy-Rayleigh scattering cluster based spatial-temporal-spectral correlation properties with MIMO-OFDM channel model. In: **International Conference on Wireless Communications and Signal Processing**. Nanjing, China: IEEE, 2011. p. 1–5.

LIMA, S. R. L.; CORDEIRO, G. M. The extended log-logistic distribution: Properties and application. **Anais da Academia Brasileira de Ciências**, v. 89, p. 3–17, 2017.

LINDGREN, F.; GELADI, P.; WOLD, S. The kernel algorithm for PLS. **Journal of Chemometrics**, v. 7, n. 1, p. 45–59, 1993.

LING, X.; GILES, D. E. Bias reduction for the maximum likelihood estimator of the parameters of the generalized Rayleigh family of distributions. **Communications in Statistics - Theory and Methods**, Taylor & Francis, v. 43, n. 8, p. 1778–1792, 2014.

LORENZ, M. O. Methods of measuring the concentration of wealth. **Publications of the American Statistical Association**, v. 9, n. 70, p. 209–219, 1905.

MAHMOUD, M. R.; EL-GHAFOUR, A. S. A. Shannon entropy for the generalized Feller-Pareto (GFP) family and order statistics of GFP subfamilies. **Applied Mathematical Sciences**, v. 7, n. 65, p. 3247–3253, 2013.

_____. Fisher information matrix for the generalized Feller-Pareto distribution. **Communications in Statistics - Theory and Methods**, v. 44, n. 20, p. 4396–4407, 2015.

MANDELBROT, B. B. The Pareto-Lévy law and the distribution of income. **International Economic Review**, Economics Department of the University of Pennsylvania, Wiley, Institute of Social and Economic Research, Osaka University, v. 1, n. 2, p. 79–106, 1960.

MARTIN, D. *et al.* A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: IEEE. **Proceedings Eighth IEEE International Conference on Computer Vision**. Vancouver, Canada, 2001. v. 2, p. 416–423.

MAZUCHELI, J.; MENEZES, A. F. B.; DEY, S. Bias-corrected maximum likelihood estimators of the parameters of the inverse Weibull distribution. **Communications in Statistics - Simulation and Computation**, Taylor & Francis, v. 0, n. 0, p. 1–10, 2018.

_____. Improved maximum-likelihood estimators for the parameters of the unit-gamma distribution. **Communications in Statistics - Theory and Methods**, Taylor & Francis, v. 47, n. 15, p. 3767–3778, 2018.

MCCLEAN, S.; MILLARD, P. Modelling in-patient bed usage behaviour in a department of geriatric medicine. **Methods of Information in Medicine**, v. 32, n. 1, p. 79–81, 1993.

MCDONALD, J. B. Some generalized functions for the size distribution of income. **Econometrica**, The Econometric Society, v. 52, n. 3, p. 647–663, 1984.

MCLACHLAN, G.; PEEL, D. **Finite Mixture Models**. New York: Wiley-Interscience, 2000. 438 p. (Wiley Series in Probability and Statistics).

MCLACHLAN, G. J.; BASFORD, K. E. **Mixture models**: Inference and applications to clustering. 1. ed. Dallas, TX: CRC Press, 1987. 273 p. (Statistics: A Series of Textbooks and Monographs).

MCLACHLAN, G. J.; KRISHNAN, T. **The EM Algorithm and Extensions**. 2. ed. New York: Wiley-Interscience, 2008. 400 p. (Wiley Series in Probability and Statistics).

MEAD, M. E. A new generalization of Burr XII distribution. **Journal of Statistics: Advances in Theory and Applications**, v. 12, n. 2, p. 53–73, 2014.

MEILĂ, M. Comparing clusterings: An information based distance. **Journal of Multivariate Analysis**, v. 98, n. 5, p. 873–895, 2007.

MICHAEL, J. R.; SCHUCANY, W. R.; HAAS, R. W. Generating random variates using transformations with multiple roots. **The American Statistician**, American Statistical Association, Taylor & Francis, Ltd., v. 30, n. 2, p. 88–90, 1976.

MOORS, J. J. A. A quantile alternative for kurtosis. **Journal of the Royal Statistical Society. Series D (The Statistician)**, Royal Statistical Society, Wiley, v. 37, n. 1, p. 25–32, 1988.

MUDHOLKAR, G. S.; SRIVASTAVA, D. K. Exponentiated Weibull family for analyzing bathtub failure-rate data. **IEEE Transactions on Reliability**, v. 42, n. 2, p. 299–302, 1993.

MURTHY, D. N. P.; XIE, M.; JIANG, R. **Weibull Models**. New Jersey: John Wiley & Sons, Inc., 2004. 396 p. (Wiley Series in Probability and Statistics).

NADARAJAH, S. Exponentiated Pareto distributions. **Statistics**, Taylor & Francis, v. 39, n. 3, p. 255–260, 2005.

NADARAJAH, S.; BAKOUCH, H. S.; TAHMASBI, R. A generalized Lindley distribution. **Sankhya B**, Springer, v. 73, n. 2, p. 331–359, 2011.

NADARAJAH, S.; KOTZ, S. The exponentiated Type distributions. **Acta Applicandae Mathematicae**, Springer Netherlands, v. 92, n. 2, p. 97–111, 2006.

NAGARAJA, H. N. **Record occurrence in the presence of a linear trend**. Departament of Statistics, September 1994.

NAIDU, M. S. R.; KUMAR, P. R.; CHIRANJEEVI, K. Shannon and fuzzy entropy based evolutionary image thresholding for image segmentation. **Alexandria Engineering Journal**, 2017.

NASCIMENTO, A. D. C.; FRERY, A. C.; CINTRA, R. J. Bias correction and modified profile likelihood under the Wishart complex distribution. **IEEE Transaction on Geoscience and Remote Sensing**, v. 52, n. 8, p. 4932–4941, 2014.

NASCIMENTO, A. D. C. *et al.* Comparing edge detection methods based on stochastic entropies and distances for PolSAR imagery. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 7, n. 2, p. 648–663, 2014.

NEVZOROV, V. B. Two caracterizations using records. In: KALASIKOV, V. V.; PENKOV, B.; ZOLOTAREV, V. M. (Ed.). Berlin: Springer-Verlag, 1986, (Statistics Problems for Stochastic Models, v. 2). cap. 1, p. 79–85.

_____. Records for nonidentically distributed random variables. In: GRIGELIONIS, B. *et al.* (Ed.). 5. ed. Mokslas: Procedings of the Vilnus Conference, 1990. v. 2, cap. 2, p. 25–31.

NGUYEN, D. C. T. *et al.* Superpixel and multi-atlas based fusion entropic model for the segmentation of X-ray images. **Medical Image Analysis**, v. 48, p. 58–74, 2018.

NIKIAS, C. L.; SHAO, M. **Signal Processing with α-stable Distributions and Applications**. New York: John Wiley & Sons, Inc., 1995.

NOBRE, R. H. *et al.* SAR image segmentation with renyi's entropy. **IEEE Signal Processing Letters**, v. 23, n. 11, p. 1551–1555, 2016.

OAKLEY, J.; HANCOCK, E. Texture segmentation using fuzzy clustering. In: **IEE Colloquium on Texture Classification**: Theory and applications. London, UK: IEEE, 1994. p. 5/1–5/4.

OLIVER, C.; QUEGAN, S. **Understanding Synthetic Aperture Radar Images**. Raleigh: SciTech Publishing, Inc., 2004.

PAPPAS, O.; BULL, D.; ACHIM, A. Cauchy-Rayleigh CFAR for ship detection in synthetic aperture radar. In: **XXVI'eme Colloque GRETSI**. Juan-Les-Pins, France: University of Bristol, 2017.

PARANAÍBA, P. F. **Caracterização e Extensões da Distribuição Burr XII: Propriedades e Aplicações**. Doctoral Dissertation — Universidade de São Paulo: Escola Superior de Agricultura Luiz de Queiroz, Piracicaba, 2012.

PARDO, L. **Statistical Inference Based on Divergence Measures**. New York: Chapman & Hall/CRC, 2006. (Statistics, Textbooks and Monographs, 185).

PARDO, L. *et al.* Large sample behavior of entropy measures when parameters are estimated. **Communications in Statistics - Theory and Methods**, Taylor & Francis, v. 26, n. 2, p. 483–501, 1997.

PEARSON, K. Contributions to the mathematical theory of evolution. **Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences**, The Royal Society, v. 185, p. 71–110, 1894.

PENG, Q. *et al.* **Modeling amplitude SAR image with the Cauchy-Rayleigh mixture**. 2017. 10463-10463-7 p.

PENG, Q.; ZHAO, L. SAR image filtering based on the Cauchy-Rayleigh mixture model. **Geoscience and Remote Sensing Letters, IEEE**, v. 11, n. 5, p. 960–964, 2014.

PENG, Y. *et al.* SAR image classification based on α-stable distribution. **Remote Sensing Letters**, v. 2, n. 1, p. 51–59, 2011.

PHAM, T. X.; SIARRY, P.; OULHADJ, H. Integrating fuzzy entropy clustering with an improved PSO for MRI brain image segmentation. **Applied Soft Computing**, v. 65, p. 230–242, 2018.

PIERCE, R. D. RCS characterization using the α-stable distribution. In: **Proceedings of the IEEE National Radar Conference**. Ann Arbor, USA: IEEE, 1996. p. 154–159.

PRENTICE, R. L. A log-gamma model and its maximum likelihood estimation. **Biometrika**, Oxford University Press, Biometrika Trust, v. 61, n. 3, p. 539–544, 1974.

QUAN, S. *et al.* Eigenvalue-based urban area extraction using polarimetric SAR data. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 11, n. 2, p. 458–471, 2018.

R, D. C. T. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2018. R-3.4.3.

RANDEN, T.; HUSOY, J. H. Texture segmentation using filters with optimized energy separation. **IEEE Transactions on Image Processing**, v. 8, n. 4, p. 571–582, 1999.

RATHIE, P. N.; SILVA, S. Shannon, Lévy, and Tsallis: A note. **Applied Mathematical Sciences**, v. 2, n. 28, p. 1359–1363, 2008.

REATH, J. **Improved Parameter Estimation Of The log-logistic Distribution With Applications**. Master Thesis — Michigan Technological University, Michigan, 2016.

REDNER, R. A.; WALKER, H. F. Mixture densities, maximum likelihood and the EM algorithm. **Society for Industrial and Applied Mathematics Review**, Society for Industrial and Applied Mathematics, v. 26, n. 2, p. 195–239, 1984.

REN, S. *et al.* Automated SAR reference image preparation for navigation. **Progress in Electromagnetics Research**, v. 121, p. 535–555, 2011.

RÉNYI, A. On measures of entropy and information. In: **Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability**. Berkeley, USA: University of California Press, 1961. (Contributions to the Theory of Statistics, v. 1), p. 547–561.

RINNE, H. **The Weibull Distribution**: A handbook. New York: Chapman & Hall/CRC, 2009. 782 p.

RITCHIE, M. D. *et al.* Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. **The American Journal of Human Genetics**, Elsevier, v. 69, n. 1, p. 138–147, 2001.

RODRIGUES, F. A. A. *et al.* SAR image segmentation using the roughness information. **IEEE Geoscience and Remote Sensing Letters**, v. 13, n. 2, p. 132–136, 2016.

SAHA, K.; PAUL, S. Bias-corrected maximum likelihood estimator of the negative binomial dispersion parameter. **Biometrics**, v. 61, n. 1, p. 179–185, 2005.

SALICRÚ, M. *et al.* Asymptotic distribution of (h,ϕ)-entropies. **Communications in Statistics - Theory and Methods**, Taylor & Francis, v. 22, n. 7, p. 2015–2031, 1993.

SANTANA, T. V. F. **As Distribuições Kumaraswamy-Log-Logística e Kumaraswamy-Logística**. Master Thesis — Universidade de São Paulo: Escola Superior de Agricultura Luiz de Queiroz, 2010.

SCHÖLKOPF, B.; SMOLA, A.; MÜLLER, K. Nonlinear component analysis as a kernel eigenvalue problem. **Neural Computation**, v. 10, n. 5, p. 1299–1319, 1998.

SCHWARTZ, J.; GILES, D. E. Bias-reduced maximum likelihood estimation of the zero-inflated Poisson distribution. **Communications in Statistics - Theory and Methods**, v. 45, n. 2, p. 465–478, 2016.

SCHWARTZ, J.; GODWIN, R. T.; GILES, D. E. Improved maximum-likelihood estimation of the shape parameter in the Nakagami distribution. **Journal of Statistical Computation and Simulation**, Taylor & Francis, v. 83, n. 3, p. 434–445, 2013.

SCHWARZ, G. Estimating the dimension of a model. **The Annals of Statistics**, The Institute of Mathematical Statistics, v. 6, n. 2, p. 461–464, 1978.

SCRUCCA, L. *et al.* mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models. **The R Journal**, v. 8, n. 1, p. 289–317, 2016.

SHANNON, C. E. A mathematical theory of communication. **Bell System Technical Journal**, Blackwell Publishing Ltd, v. 27, n. 3, p. 379–423, 1948.

SHAPIRO, L.; STOCKMAN, G. **Computer Vision**. Upper Saddle River: Prentice Hall, 2001.

SIDDIQUE, M. A. *et al.* Single-look SAR tomography as an add-on to PSI for improved deformation analysis in urban areas. **IEEE Transaction on Geoscience and Remote Sensing**, v. 54, n. 10, p. 6119–6137, 2016.

SILVA, R. V. *et al.* The exponentiated Burr XII Poisson distribution with application to lifetime data. **International Journal of Statistics and Probability**, v. 4, n. 4, p. 112–131, 2015.

SINGH, A. K.; SINGH, A.; MURPHY, D. J. On bias corrected estimators of the two parameter gamma distribution. In: **2015 12th International Conference on Information Technology - New Generations**. Las Vegas, NV: IEEE, 2015. p. 127–132.

SINGH, B.; CHOUDHARY, N. The exponentiated Perks distribution. **International Journal of System Assurance Engineering and Management**, p. 1–11, 2016.

SMIRNOV, N. Table for estimating the goodness of fit of empirical distributions. **The Annals of Mathematical Statistics**, The Institute of Mathematical Statistics, v. 19, n. 2, p. 279–281, 1948.

STOVER, C.; WEISSTEIN, E. W. **Closed-Form Solution**. 29-Mar 2017. MathWorld - A Wolfram Web Resource. Http://mathworld.wolfram.com/Closed-FormSolution.html.

SURLES, J. G.; PADGETT, W. J. Some properties of a scaled Burr type X distribution. **Journal of Statistical Planning and Inference**, v. 128, n. 1, p. 271–280, 2005.

TAHIR, M. H. *et al.* McDonald log-logistic distribution with an application to breast cancer data. **Journal of Statistical Theory and Applications**, v. 13, p. 65–82, 2014.

TEIMOURI, M.; NADARAJAH, S. Bias corrected MLEs for the Weibull distribution based on records. **Statistical Methodology**, v. 13, p. 12 – 24, 2013.

TENENBAUM, J. B.; SILVA, V. d.; LANGFORD, J. C. A global geometric framework for nonlinear dimensionality reduction. **Science**, American Association for the Advancement of Science, v. 290, n. 5500, p. 2319–2323, 2000.

TITTERINGTON, D. M.; SMITH, A. F. M.; MAKOV, U. E. **Statistical Analysis of Finite Mixture Distributions**. New York: John Wiley & Sons, 1985. 248 p. (Wiley Series in Probability and Mathematical Statistics (Applied section)).

TSALLIS, C. Possible generalization of Boltzmann-Gibbs statistics. **Journal of Statistical Physics**, v. 52, n. 1, p. 479–487, 1988.

ULABY, F.; ELACHI, C. **Radar Polarimetry for Geoscience Applications**. Artech House, 1990. (Artech House remote sensing library).

VERHULST, P. F. Notice sur la loi que la population poursuit dans son accroissement. **Correspondance Mathématique et Physique**, n. 10, p. 113–121, 1838.

_____. Mémoires de lacadémie impériale et royale des sciences et belles-lettres de bruxelles. A. Bruxeles; De Limprimerie Académique, 1847.

VINCI, F. Nuovi contributi allo studio della distribuzione dei redditi. **Giornale degli Economisti e Rivista di Statistica**, EGEA SpA, v. 61 (Anno 32), n. 11, p. 365–369, 1921.

VOIT, J. **The Statistical Mechanics of Financial Markets**. Heidelberg: Springer, 2005. 378 p.

WANG, B.; CHEN, L. L.; CHENG, J. New result on maximum entropy threshold image segmentation based on P system. **Optik**, v. 163, p. 81–85, 2018.

WANG, C.; LIAO, M.; LI, X. Ship detection in SAR image based on the α-stable distribution. **Sensors (Basel)**, v. 8, n. 8, p. 4948–4960, 2008.

WANG, L. *et al.* Active contours driven by edge entropy fitting energy for image segmentation. **Signal Processing**, v. 149, p. 27–35, 2018.

WANG, M.; WANG, W. Bias-corrected maximum likelihood estimation of the parameters of the weighted Lindley distribution. **Communications in Statistics - Simulation and Computation**, Taylor & Francis, v. 46, n. 1, p. 530–545, 2017.

WATERLOO MAPLE Inc. **Maple**. Waterloo, ON, 2015. Version 2015.0.

WILEY, C. A. Synthetic aperture radars. **IEEE Transactions on Aerospace and Eletronic Systems**, v. 21, n. 3, p. 440–443, 1985.

WILKS, S. S. Shortest average confidence intervals from large samples. **The Annals of Mathematical Statistics**, The Institute of Mathematical Statistics, v. 9, n. 3, p. 166–175, 1938.

WOLFE, J. H. **A computer program for the maximum likelihood analysis of types**. San Diego, CA, 1965.

_____. **NORMIX: Computational methods for estimating the parameters of multivariate normal mixtures of distributions**. San Diego, CA, 1967.

_____. Pattern Clustering by Multivariate Mixture Analysis. **Multivariate Behavioral Research**, Routledge, v. 5, n. 3, p. 329–350, 1970.

WOLFRAM RESEARCH Inc. **Wolfram Mathematica**. Champaign, IL, 2018. Version 11.1.1.

WU, H. *et al.* Joint entropy based learning model for image retrieval. **Journal of Visual Communication and Image Representation**, v. 55, p. 415–423, 2018.

XIANG, D.; BAN, Y.; SU, Y. Model-based decomposition with cross scattering for polarimetric SAR urban areas. **IEEE Geoscience and Remote Sensing Letters**, v. 12, n. 12, p. 2496–2500, 2015.

XU, R.; ZHANG, H.; LIN, H. Urban impervious surfaces estimation from optical and SAR imagery: A comprehensive comparison. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 10, n. 9, p. 4010–4021, 2017.

YANG, M. C. K. On the distribution of the inter-record times in an increasing population. **Journal of Applied Probability**, v. 12, p. 148–154, 1975.

YANG, W. *et al.* On the mixed scattering mechanism analysis of model-based decomposition for polarimetric SAR data. **Progress in Electromagnetics Research B**, v. 52, p. 327–345, 2013.

YAP, B. W.; SIM, C. H. Comparisons of various types of normality tests. **Journal of Statistical Computation and Simulation**, Taylor & Francis, v. 81, n. 12, p. 2141–2155, 2011.

YIN, S.; QIAN, Y.; GONG, M. Unsupervised hierarchical image segmentation through fuzzy entropy maximization. **Pattern Recognition**, v. 68, p. 245–259, 2017.

ZANDONATTI, A. **Distribuzioni de Pareto Generalizzate**. Doctoral Dissertation — University of Trento, Italy, 2001.

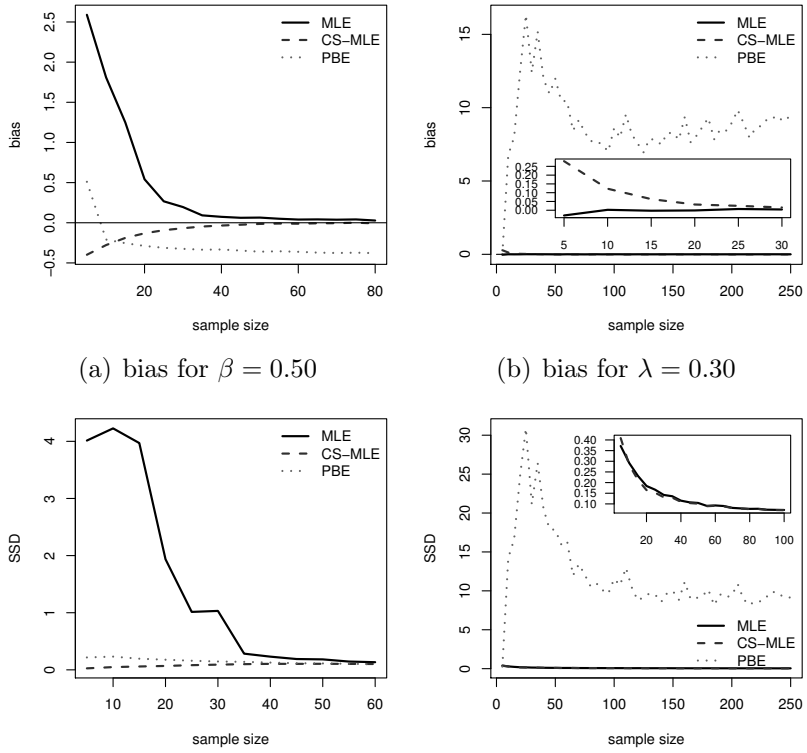
ZHANG, F.; YE, X.; LIU, W. Image decomposition and texture segmentation via sparse representation. **IEEE Signal Processing Letters**, v. 15, p. 641–644, 2008.

ZHANG, G.; LIU, R. Bias-corrected estimators of scalar skew normal. **Communications in Statistics - Simulation and Computation**, Taylor & Francis, v. 46, n. 2, p. 831–839, 2017.

ZHOU, F. *et al.* SAR ATR of ground vehicles based on LM-BN-CNN. **IEEE Transaction on Geoscience and Remote Sensing**, p. 1–12, 2018.

APPENDIX A - SIMULATION STUDIES

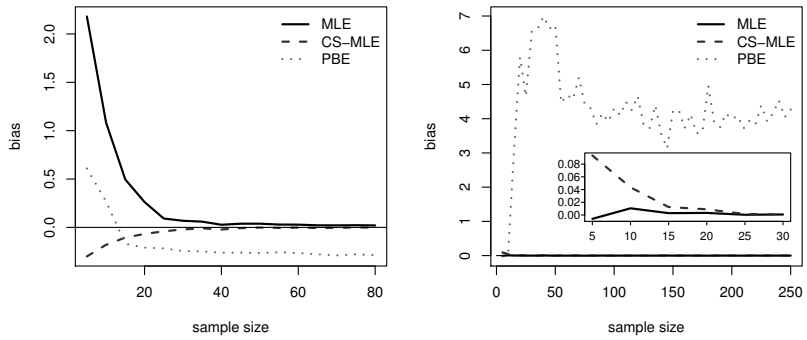
In this appendix we show the plots produced in the two simulation studies described in Section 3.6 of the Chapter 3 (The EXPONENTIATED CAUCHY-RAYLEIGH model). The plots in Figures A.1–A.11 are analyzed in Section 3.6.1 (Convergence study) while the plots in Figures A.12–A.15 are interpreted in Section 3.6.2 (Relative bias and sample standard deviation studies).



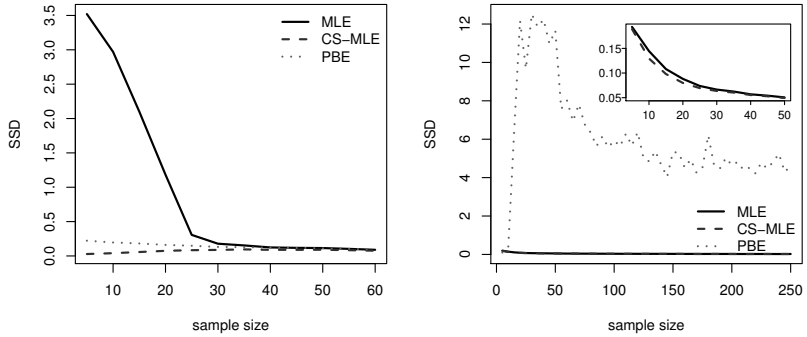
(a) bias for $\beta = 0.50$ (b) bias for $\lambda = 0.30$

(c) SSD for $\beta = 0.50$ (d) SSD for $\lambda = 0.30$

Figure A.1: ML, PB and ML-CS estimates mean biases and SSDs of a random sample following ECR(0.50,0.30).

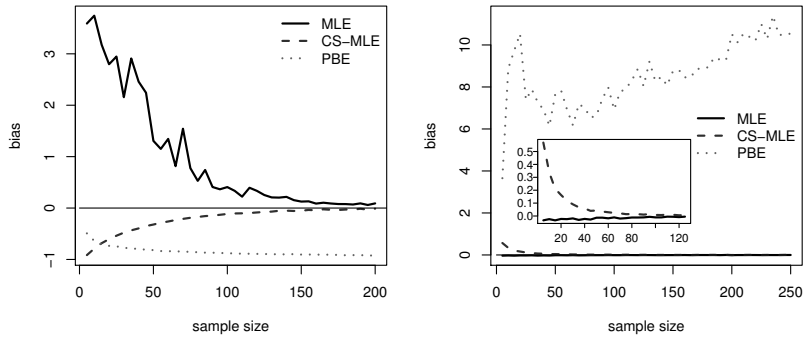


(a) bias for $\beta = 0.40$ (b) bias for $\lambda = 0.15$

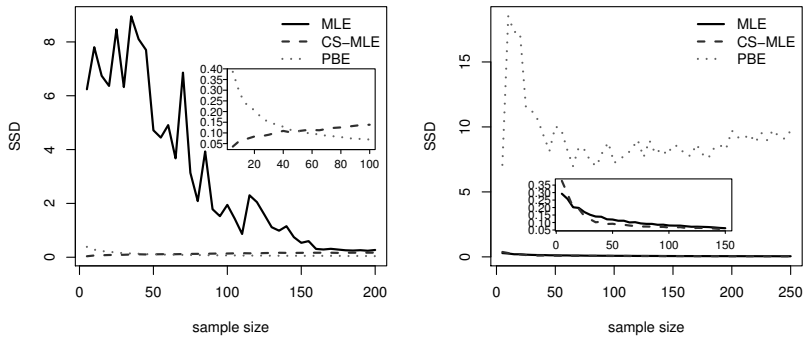


(c) SSD for $\beta = 0.40$ (d) SSD for $\lambda = 0.15$

Figure A.2: ML, PB and ML-CS estimates mean biases and SSDs of a random sample following ECR(0.40,0.15).

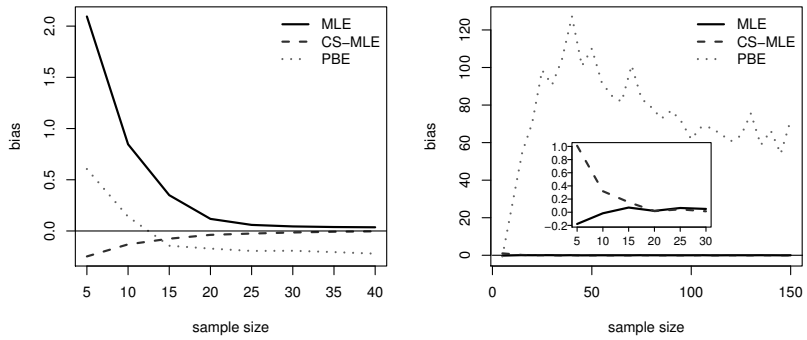


(a) bias for $\beta = 1.00$ (b) bias for $\lambda = 0.25$

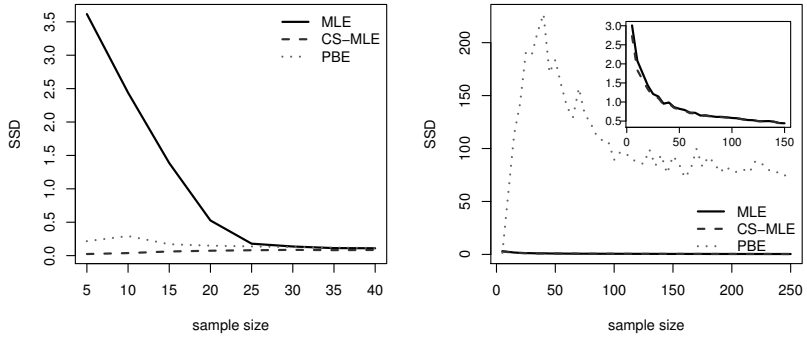


(c) SSD for $\beta = 1.00$ (d) SSD for $\lambda = 0.25$

Figure A.3: ML, PB and ML-CS estimates mean biases and SSDs of a random sample following ECR(1.00,0.25).

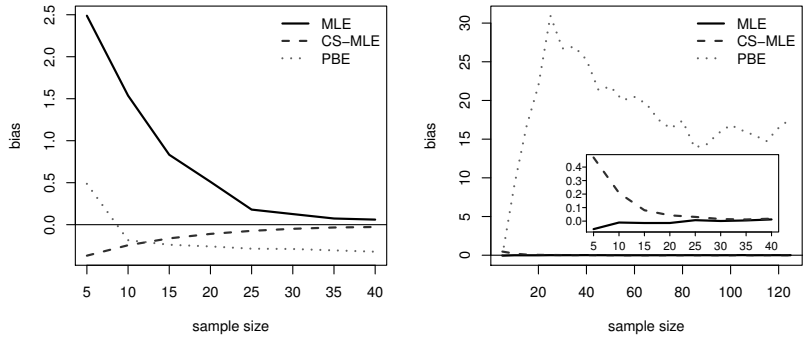


(a) bias for $\beta = 0.35$ (b) bias for $\lambda = 2.50$

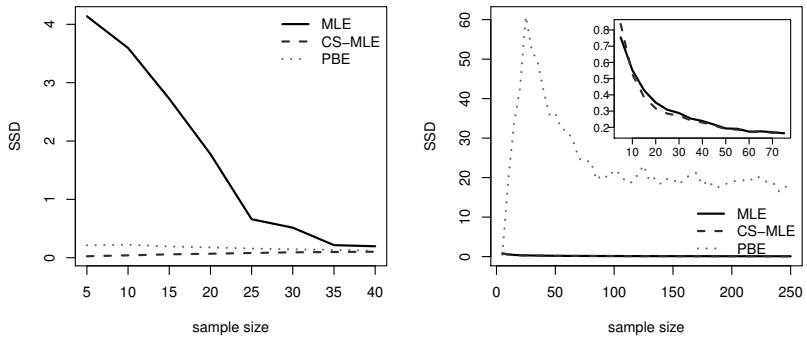


(c) SSD for $\beta = 0.35$ (d) SSD for $\lambda = 2.50$

Figure A.4: ML, PB and ML-CS estimates mean biases and SSDs of a random sample following ECR(0.35,2.50).

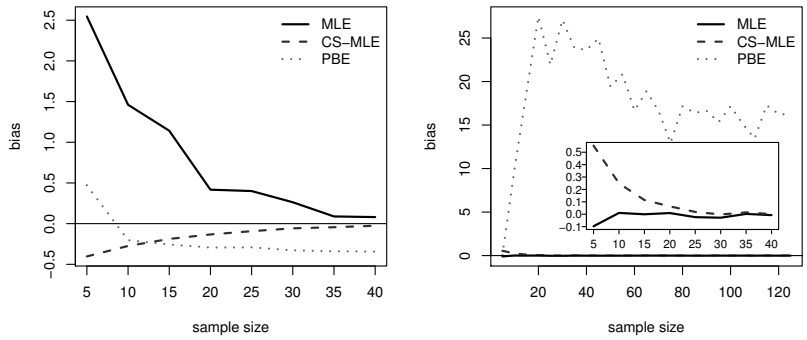


(a) bias for $\beta = 0.47$ (b) bias for $\lambda = 0.60$

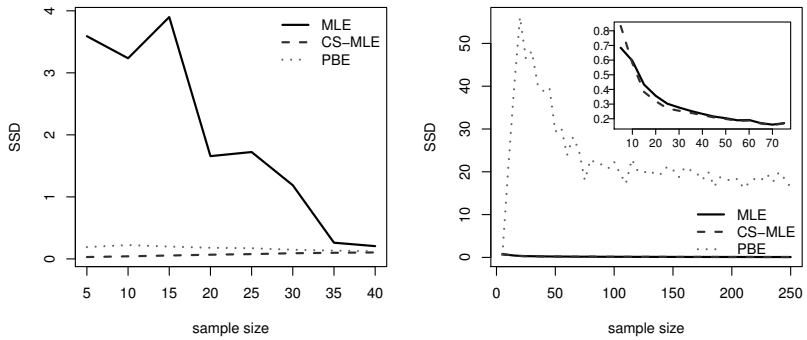


(c) SSD for $\beta = 0.47$ (d) SSD for $\lambda = 0.60$

Figure A.5: ML, PB and ML-CS estimates mean biases and SSDs of a random sample following ECR(0.47,0.60).

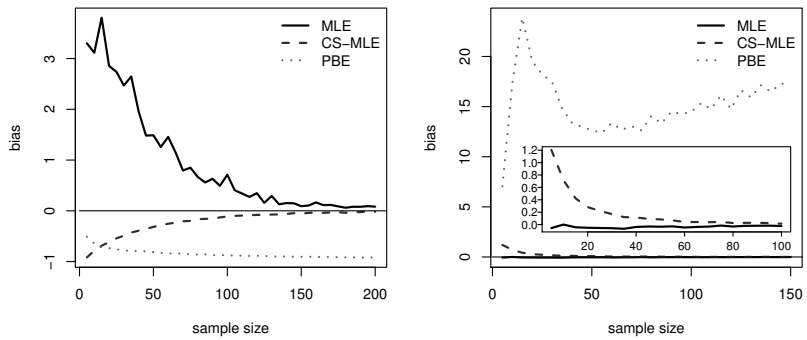


(a) bias for $\beta = 0.50$ (b) bias for $\lambda = 0.60$

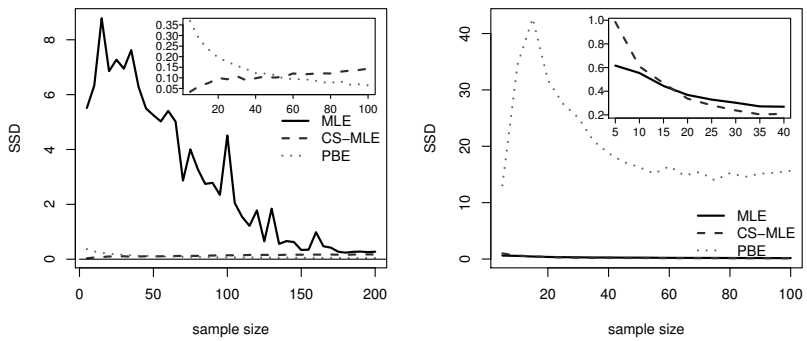


(c) SSD for $\beta = 0.50$ (d) SSD for $\lambda = 0.60$

Figure A.6: ML, PB and ML-CS estimates mean biases and SSDs of a random sample following ECR(0.50,0.60).

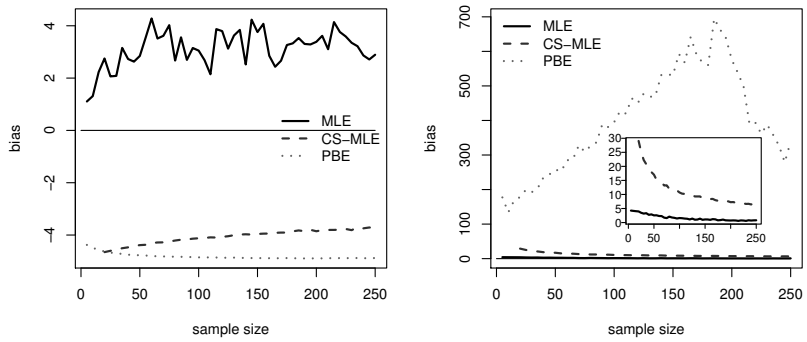


(a) bias for $\beta = 1.00$ (b) bias for $\lambda = 0.50$

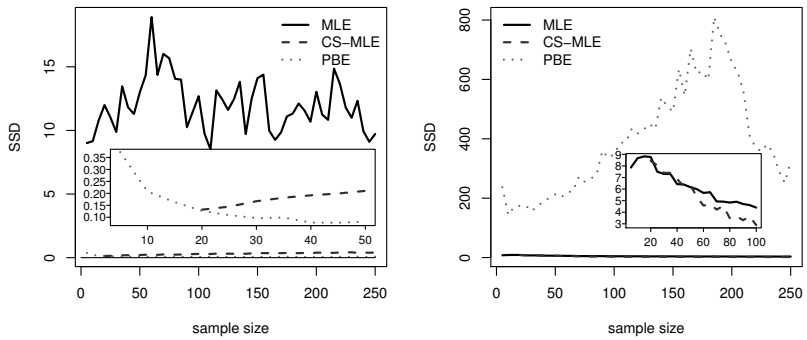


(c) SSD for $\beta = 1.00$ (d) SSD for $\lambda = 0.50$

Figure A.7: ML, PB and ML-CS estimates mean biases and SSDs of a random sample following ECR(1.00,0.50).

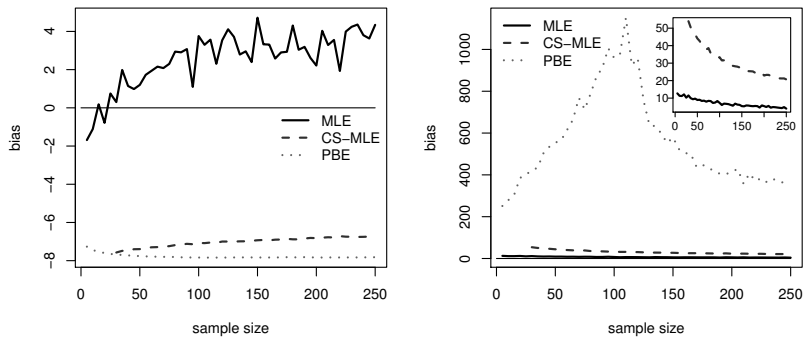


(a) bias for $\beta = 5.00$ (b) bias for $\lambda = 5.00$

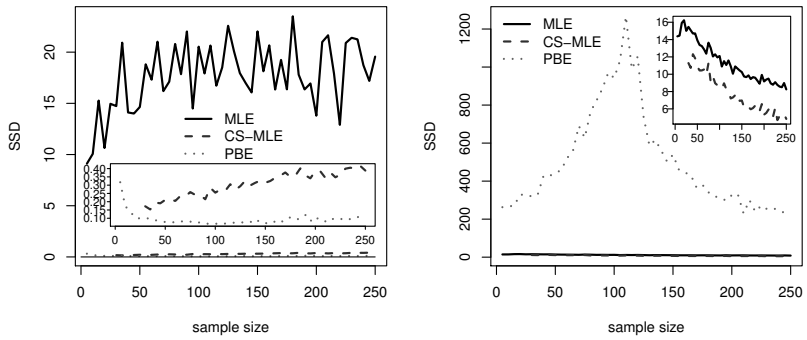


(c) SSD for $\beta = 5.00$ (d) SSD for $\lambda = 5.00$

Figure A.8: ML, PB and ML-CS estimates mean biases and SSDs of a random sample following ECR(5.00,5.00).

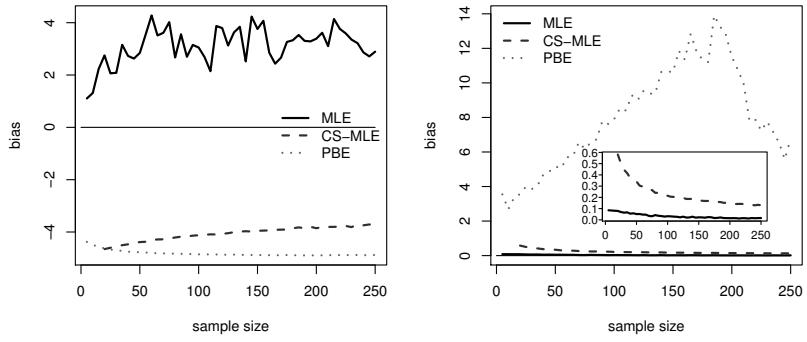


(a) bias for $\beta = 8.00$ (b) bias for $\lambda = 8.00$

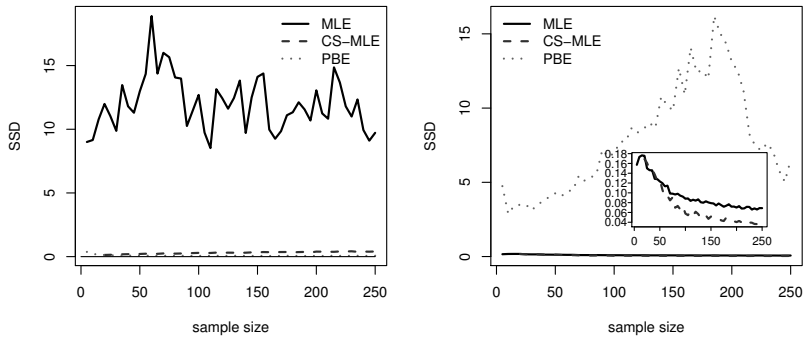


(c) SSD for $\beta = 8.00$ (d) SSD for $\lambda = 8.00$

Figure A.9: ML, PB and ML-CS estimates mean biases and SSDs of a random sample following ECR(8.00,8.00).

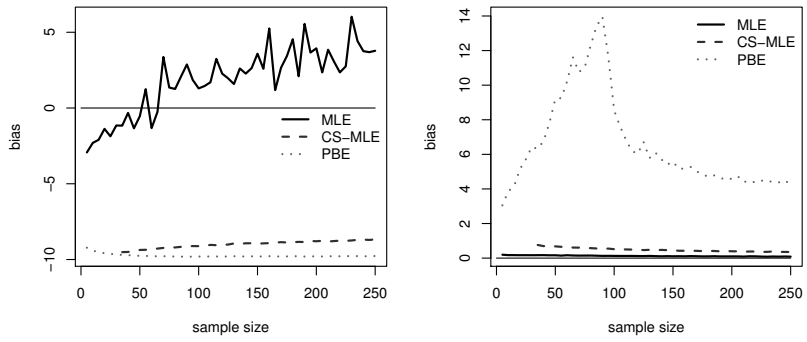


(a) bias for $\beta = 5.00$ (b) bias for $\lambda = 0.10$

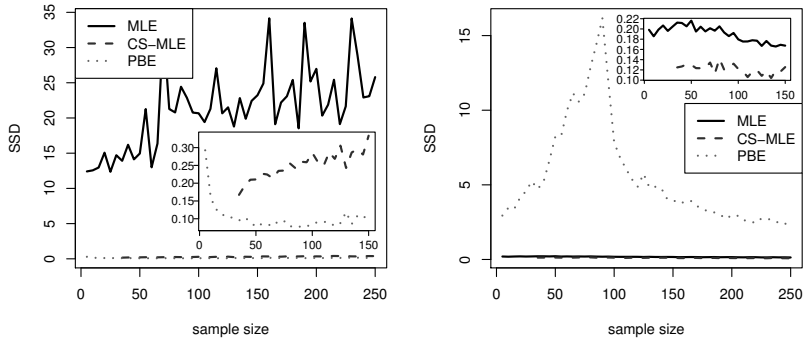


(c) SSD for $\beta = 5.00$ (d) SSD for $\lambda = 0.10$

Figure A.10: ECR(5.00,0.10) mean bias and SSD for some sample sizes.

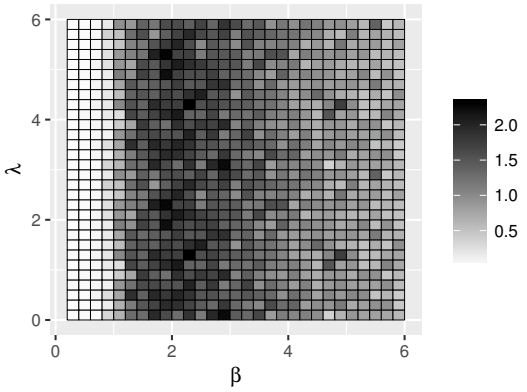


(a) bias for $\beta = 10.00$ (b) bias for $\lambda = 0.10$

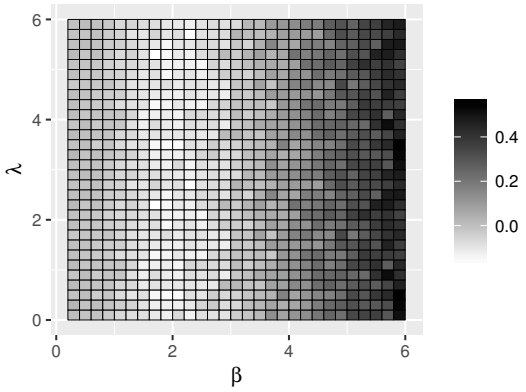


(c) SSD for $\beta = 10.00$ (d) SSD for $\lambda = 0.10$

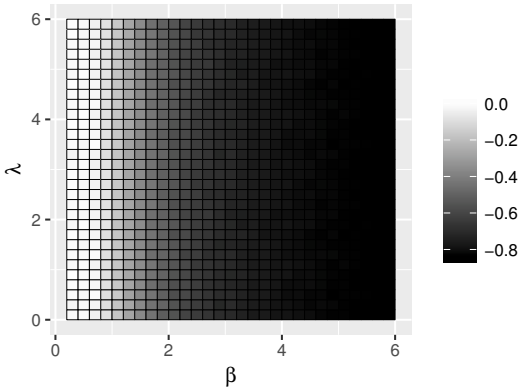
Figure A.11: ML, PB and ML-CS estimates mean biases and SSDs of a random sample following ECR(10.00,0.10).



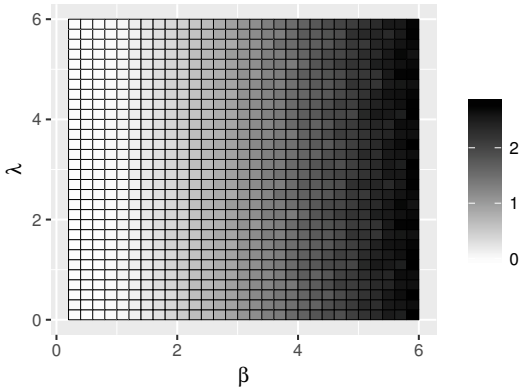
(a) MLEs of β



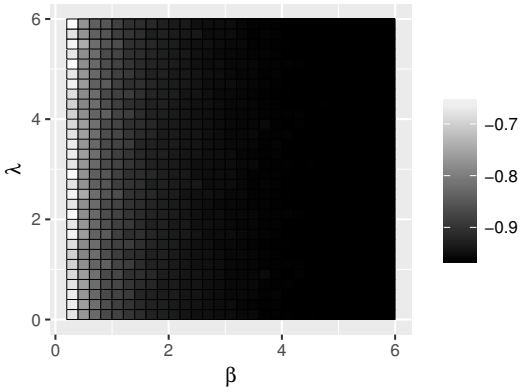
(b) MLEs of λ



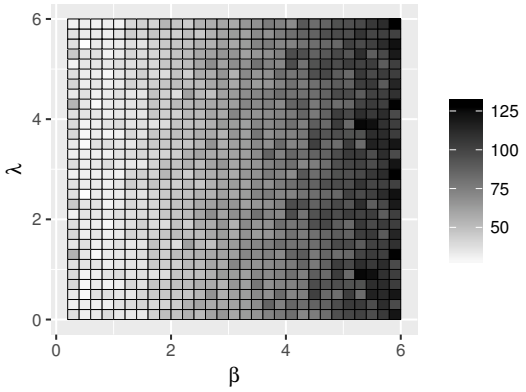
(c) CS-MLEs of β



(d) CS-MLEs of λ

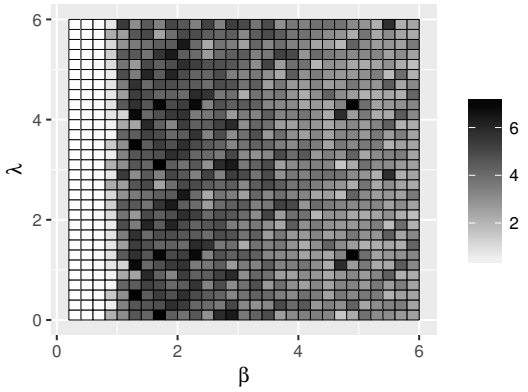


(e) PBEs of β

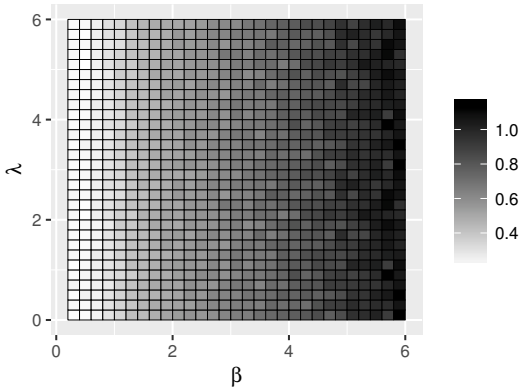


(f) PBEs of λ

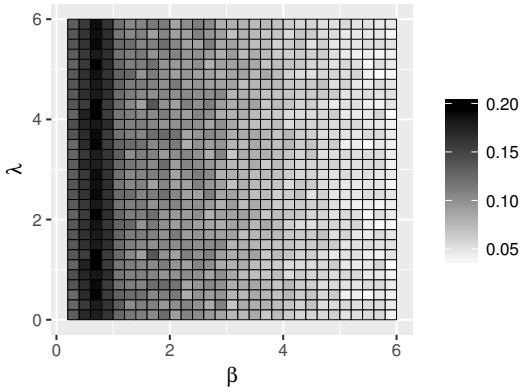
Figure A.12: ECR relative bias of the estimates for sample size 100.



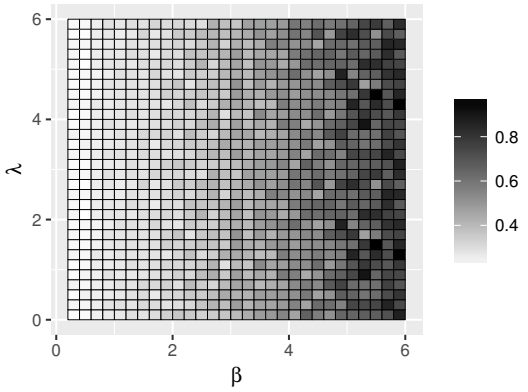
(a) MLEs of β



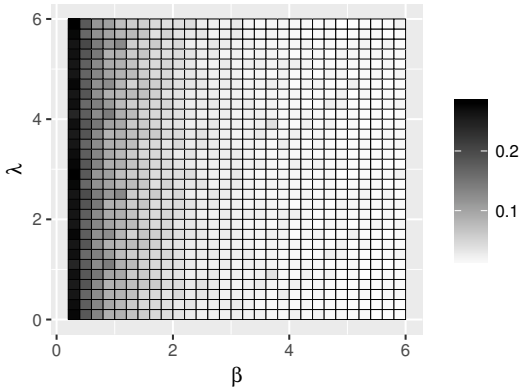
(b) MLEs of λ



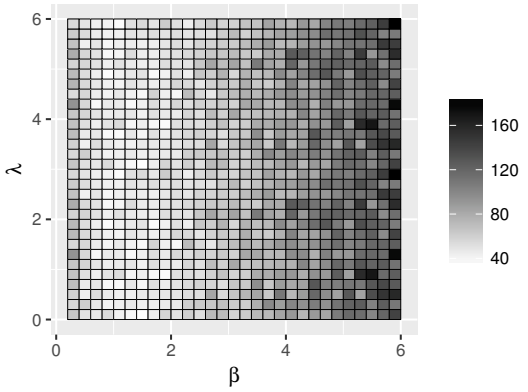
(c) CS-MLEs of β



(d) CS-MLEs of λ

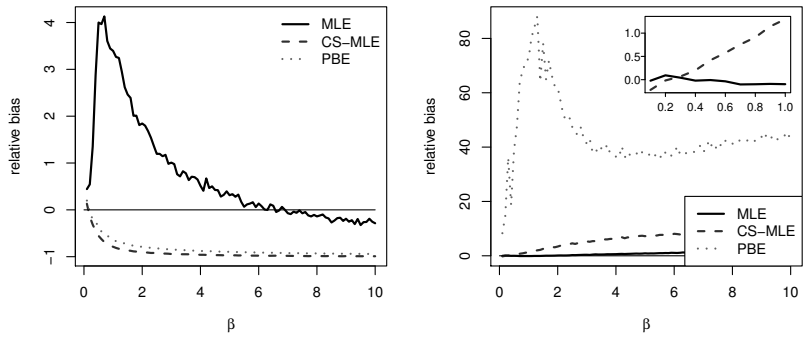


(e) PBEs of β

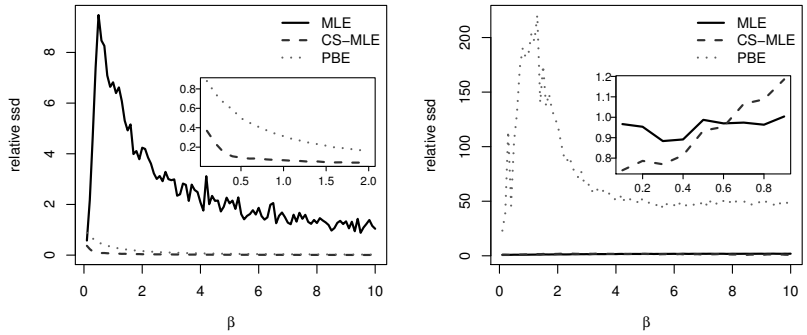


(f) PBEs of λ

Figure A.13: ECR relative SSDs of the estimates for sample size 100.

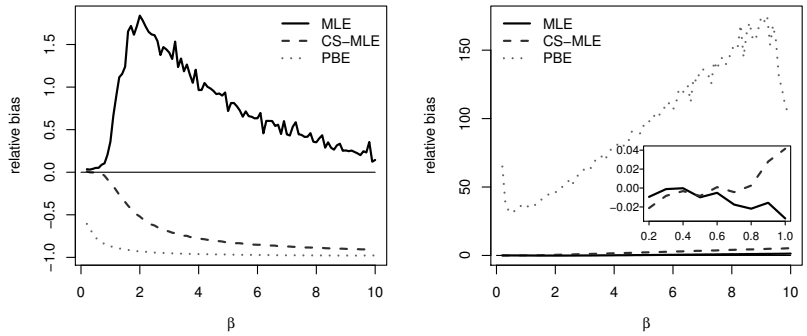


(a) β estimate relative biases (b) λ estimate relative biases

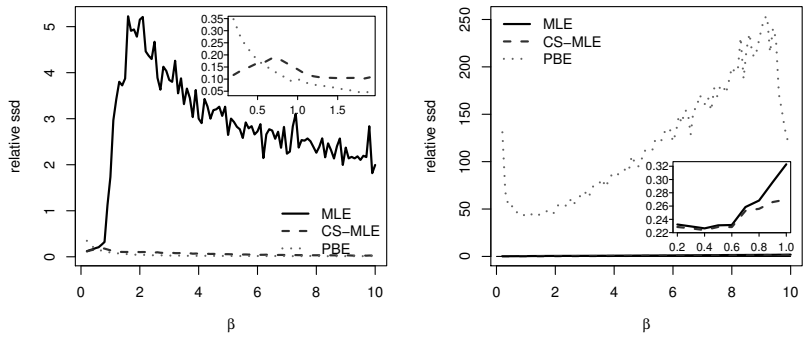


(c) β estimate relative SSDs (d) λ estimate relative SSDs

Figure A.14: Comparison between relatives bias and SSD of ECR estimates ($\lambda = 1$ and $n = 10$).



(a) β estimate relative biases (b) λ estimate relative biases



(c) β estimate relative SSDs (d) λ estimate relative SSDs

Figure A.15: Comparison between relatives bias and SSD of ECR estimates ($\lambda = 1$ and $n = 100$).

APPENDIX B - APPLICATIONS IN SAR DATA

Table B.1: Rejection rates (%) obtained in texture identification study of urban areas considering sample size 25.

		HH		HV		VV	
Test	α	CR	ECR	CR	ECR	CR	ECR
	1%	0.00	0.00	0.00	0.00	0.00	0.00
LRT	5%	0.90	0.00	0.00	0.00	1.05	0.00
	10%	5.10	0.45	3.90	0.60	5.40	0.60
	1%	0.00	0.00	0.00	0.00	0.00	0.00
SET	5%	0.00	0.30	0.00	0.45	0.00	0.90
	10%	0.90	3.60	0.45	3.00	1.05	4.20
	1%	0.00	0.00	0.00	0.00	0.00	0.00
RET	5%	0.00	0.30	0.00	0.60	0.00	1.05
	10%	0.90	4.20	0.45	4.35	1.05	4.80
	1%	0.00	0.00	0.00	0.00	0.00	0.00
TET	5%	0.00	0.30	0.00	0.60	0.00	1.05
	10%	0.90	4.20	0.45	3.75	1.05	4.80
	1%	0.00	0.00	0.00	0.00	0.00	0.00
AET	5%	0.00	0.30	0.00	0.60	0.00	1.05
	10%	0.90	4.05	0.45	3.75	1.05	4.80

Table B.2: Number of times the tests appear in Tables B.8 to B.11

	LRT	SET	RET	TET	AET
Table B.8	74	59	23	4	27
Table B.9	65	30	12	10	75
Table B.10	64	3	9	2	2
Table B.11	2	39	4	1	34

Table B.3: Rejection rates (%) obtained in texture identification study of urban areas considering sample size 121.

		HH		HV		VV	
Test	α	CR	ECR	CR	ECR	CR	ECR
	1%	0.00	0.00	0.00	0.00	0.00	0.00
LRT	5%	1.40	0.20	0.80	0.00	1.00	0.00
	10%	8.40	0.40	4.80	0.40	5.80	0.60
	1%	0.00	0.00	0.00	0.00	0.00	0.00
SET	5%	0.20	0.80	0.00	0.80	0.00	1.00
	10%	1.20	5.80	0.80	4.20	1.60	6.40
	1%	0.00	0.00	0.00	0.00	0.00	0.00
RET	5%	0.20	0.80	0.00	0.80	0.00	1.00
	10%	1.20	6.40	0.80	6.00	1.60	7.80
	1%	0.00	0.00	0.00	0.00	0.00	0.00
TET	5%	0.20	0.80	0.00	0.80	0.00	1.00
	10%	1.00	6.40	0.80	5.80	1.20	7.60
	1%	0.00	0.00	0.00	0.00	0.00	0.00
AET	5%	0.00	0.80	0.00	0.80	0.00	1.00
	10%	1.00	6.40	0.80	5.60	1.20	7.00

Table B.4: Rejection rates (%) obtained in texture identification study of urban areas considering sample size 289.

		HH		HV		VV	
Test	α	CR	ECR	CR	ECR	CR	ECR
	1%	0.00	0.00	0.00	0.00	0.00	0.00
LRT	5%	0.40	0.00	1.20	0.00	1.00	0.20
	10%	6.00	0.60	3.40	1.00	8.80	1.00
	1%	0.00	0.00	0.00	0.00	0.00	0.00
SET	5%	0.00	0.40	0.00	1.00	0.20	1.20
	10%	1.60	4.80	0.60	3.20	2.60	7.80
	1%	0.00	0.00	0.00	0.00	0.00	0.00
RET	5%	0.00	0.40	0.00	1.20	0.20	1.40
	10%	1.60	5.20	0.60	4.60	2.60	8.60
	1%	0.00	0.00	0.00	0.00	0.00	0.00
TET	5%	0.00	0.40	0.00	1.20	0.20	1.40
	10%	1.60	5.20	0.60	4.40	2.20	8.40
	1%	0.00	0.00	0.00	0.00	0.00	0.00
AET	5%	0.00	0.40	0.00	1.00	0.20	1.40
	10%	1.60	5.00	0.60	4.40	2.00	8.20

Table B.5: Rejection rates (%) obtained in information redundancy study considering sample size 25.

Test	α	HH-HV CR	HH-HV ECR	HH-VV CR	HH-VV ECR	HV-VV CR	HV-VV ECR
LRT	1%	40.63	12.07	0.00	0.00	17.86	3.48
	5%	95.34	68.79	0.00	0.00	78.49	38.71
	10%	99.55	90.31	0.00	0.00	95.05	68.67
SET	1%	5.15	33.75	0.00	0.00	1.35	13.13
	5%	74.17	91.51	0.00	0.00	43.90	69.60
	10%	96.35	98.76	0.00	0.02	80.77	91.25
RET	1%	5.15	40.15	0.00	0.00	1.35	16.62
	5%	74.17	93.37	0.00	0.00	43.90	73.75
	10%	96.35	98.99	0.00	0.02	80.77	92.81
TET	1%	4.41	38.63	0.00	0.00	1.13	15.63
	5%	72.99	93.14	0.00	0.00	42.58	73.15
	10%	96.18	98.95	0.00	0.02	80.35	92.66
AET	1%	3.92	37.00	0.00	0.00	0.94	14.91
	5%	72.05	92.98	0.00	0.00	41.58	72.66
	10%	96.00	98.92	0.00	0.02	79.96	92.51

Table B.6: Rejection rates (%) obtained in information redundancy study considering sample size 121.

Test	α	HH-HV CR	HH-HV ECR	HH-VV CR	HH-VV ECR	HV-VV CR	HV-VV ECR
LRT	1%	100.00	100.00	0.00	0.00	100.00	100.00
	5%	100.00	100.00	0.00	0.00	100.00	100.00
	10%	100.00	100.00	0.00	0.02	100.00	100.00
SET	1%	100.00	100.00	0.00	0.00	100.00	100.00
	5%	100.00	100.00	0.00	0.00	100.00	100.00
	10%	100.00	100.00	0.00	0.30	100.00	100.00
RET	1%	100.00	100.00	0.00	0.00	100.00	100.00
	5%	100.00	100.00	0.00	0.00	100.00	100.00
	10%	100.00	100.00	0.00	0.43	100.00	100.00
TET	1%	100.00	100.00	0.00	0.00	100.00	100.00
	5%	100.00	100.00	0.00	0.00	100.00	100.00
	10%	100.00	100.00	0.00	0.43	100.00	100.00
AET	1%	100.00	100.00	0.00	0.00	100.00	100.00
	5%	100.00	100.00	0.00	0.00	100.00	100.00
	10%	100.00	100.00	0.00	0.43	100.00	100.00

Table B.7: Rejection rates (%) obtained in information redundancy study considering sample size 289.

		HH-HV		HH-VV		HV-VV	
Test	α	CR	ECR	CR	ECR	CR	ECR
LRT	1%	100.00	100.00	0.00	0.00	100.00	100.00
	5%	100.00	100.00	0.00	0.00	100.00	100.00
	10%	100.00	100.00	0.00	0.36	100.00	100.00
SET	1%	100.00	100.00	0.00	0.00	100.00	100.00
	5%	100.00	100.00	0.00	0.27	100.00	100.00
	10%	100.00	100.00	0.00	2.92	100.00	100.00
RET	1%	100.00	100.00	0.00	0.00	100.00	100.00
	5%	100.00	100.00	0.00	0.38	100.00	100.00
	10%	100.00	100.00	0.00	3.72	100.00	100.00
TET	1%	100.00	100.00	0.00	0.00	100.00	100.00
	5%	100.00	100.00	0.00	0.38	100.00	100.00
	10%	100.00	100.00	0.00	3.72	100.00	100.00
AET	1%	100.00	100.00	0.00	0.00	100.00	100.00
	5%	100.00	100.00	0.00	0.38	100.00	100.00
	10%	100.00	100.00	0.00	3.72	100.00	100.00

Table B.8: Tests which present minors distance between the size and their corresponding nominal level in function of β

n	α	β 0.15	0.3	0.45	0.6	0.75	0.9	1.05	1.2	1.35	1.5	1.65	1.8	1.95	2.1	2.25	2.4	2.55	2.7	2.85	3
25	1%	AET	LRT	AET	LRT	LRT	LRT	SET	SET	SET	LRT	SET	SET[1]	SET	SET	SET	SET	SET	SET	SET	SET
	5%	AET	LRT	SET	RET	SET	LRT	LRT	SET	SET	SET	SET	SET	SET	SET	SET	SET	SET	SET	SET	SET
	10%	AET	LRT	LRT	RET	SET	LRT	LRT	LRT	LRT	SET	SET	SET	SET	SET	SET	SET	SET	SET	SET	SET
121	1%	AET	SET[2]	LRT	AET	AET	RET	RET	LRT	AET	LRT	SET[1]	LRT	LRT	RET	LRT	LRT	SET	LRT	SET	LRT
	5%	AET	LRT	LRT	AET	AET[2]	SET[1]	SET	LRT	RET	SET	SET	LRT	LRT	LRT	LRT	LRT	LRT	SET	SET	LRT
	10%	SET	LRT	LRT	AET	AET	SET	SET	LRT	RET	RET	SET	LRT	LRT	LRT	LRT	LRT	LRT	LRT	SET	LRT
289	1%	AET	SET[3]	AET[4]	AET	SET[3]	RET	LRT	AET	RET	AET[2]	LRT	LRT	LRT	LRT	LRT	SET	LRT	RET[3]	RET	LRT
	5%	AET	RET	AET	LRT	AET	LRT	SET	AET	LRT	SET	LRT	RET	LRT	LRT	LRT	LRT	RET	LRT	LRT	LRT
	10%	AET	LRT	AET	AET	AET	LRT	AET	SET	LRT	SET	RET	RET	LRT	RET	LRT	LRT	LRT	LRT	LRT	LRT

[1] Drew with RET.
[2] Drew with TET.
[3] Drew with LRT.
[4] Drew with RET and TET.

Table B.9: Tests which present majors distance between the size and their corresponding nominal level in function of β

		β																			
n	α	0.15	0.3	0.45	0.6	0.75	0.9	1.05	1.2	1.35	1.5	1.65	1.8	1.95	2.1	2.25	2.4	2.55	2.7	2.85	3
	1%	RET	RET	RET	AET	AET	AET	LRT	LRT	LRT	AET	LRT	LRT	AET	AET	AET	LRT	AET	AET	AET[1]	LRT
25	5%	SET	RET	AET	LRT	LRT	AET	AET	AET	AET	AET	LRT	AET	LRT	AET	LRT	LRT	LRT	LRT	LRT	AET[1]
	10%	SET	RET	AET	LRT	LRT	AET	AET	AET	AET	AET	LRT	AET	AET	AET	LRT	LRT	LRT	LRT	LRT	LRT
	1%	LRT	LRT	AET	LRT	LRT	LRT	LRT	AET	RET	AET	LRT	AET	AET	LRT	AET[1]	AET[2]	LRT	AET[1]	AET[1]	AET[1]
121	5%	LRT	RET	AET	LRT	LRT	LRT	LRT	AET	SET	AET	LRT	AET	AET	SET	AET	AET	AET	AET	AET	AET
	10%	RET	RET	SET	LRT	LRT	AET	LRT	AET	SET	LRT	LRT	AET	AET	AET	AET	AET[1]	AET	AET[1]	AET	AET
	1%	SET	AET[3]	LRT	LRT	RET	LRT	SET	LRT	LRT	LRT	AET	SET	SET	SET	SET	LRT	SET	SET	AET	AET[1]
289	5%	SET	LRT	LRT	SET	LRT	SET	LRT	LRT	AET	LRT	SET	SET	SET	SET	SET	AET	AET	AET	AET	AET
	10%	SET	SET	LRT	LRT	LRT	SET	LRT	RET	SET	LRT	LRT	SET	SET	LRT	AET	AET	SET	AET	AET	AET

[1] Drew with TET.
[2] Drew with SET.
[3] Drew with RET and TET.

Table B.10: Tests that present the greatest powers as a function θ.

n	α	θ 0.4	0.5	0.6	0.7	0.8	0.9	1	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9
25	1%	LRT	LRT	LRT	LRT	RET	RET	LRT	SET	LRT	LRT	LRT	LRT	LRT	LRT	LRT	LRT
	5%	LRT	LRT	LRT	LRT	LRT	RET	LRT	SET	LRT	LRT	LRT	LRT	LRT	LRT	LRT	LRT
	10%	LRT	LRT	LRT	LRT	LRT	LRT	LRT	LRT	LRT	LRT	LRT	LRT	LRT	LRT	LRT	LRT
121	1%			LRT	LRT	LRT	LRT	RET	RET	LRT	LRT	LRT	LRT				
	5%			LRT	LRT	LRT	LRT	SET	RET	LRT	LRT	RET	LRT				
	10%			LRT	LRT	LRT	LRT	SET	RET	LRT	LRT	LRT	LRT[1]				
289	1%					LRT	LRT	LRT	RET	LRT	LRT						
	5%					LRT	LRT	LRT	TET[2]	TET[2]	LRT						
	10%					LRT	TET[2]	LRT	TET	TET[2]	LRT						

[1] Drew with RET, TET and AET.
[2] Drew with AET.

Table B.11: Tests that present the lowest powers as a function θ.

n	α	θ 0.4	0.5	0.6	0.7	0.8	0.9	1	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9
25	1%	RET	SET	SET	SET	AET	LRT	AET	LRT	AET	AET	AET	AET	AET	AET	AET	AET
	5%	RET	SET	SET	SET	AET	AET	AET	AET	AET	AET	AET	AET	AET	AET	AET	AET
	10%	RET[1]	SET	SET	SET	SET	AET	AET	AET	AET	AET	AET	SET	SET	SET	SET	SET
121	1%			SET	SET	SET	SET	AET	SET	SET	SET	SET	SET				
	5%			SET[2]	SET	SET	SET	AET	SET	SET	SET	SET	SET				
	10%			SET[2]	SET	SET	SET	AET	SET	SET	SET	SET	SET				
289	1%					SET	SET	AET	SET	SET	SET						
	5%					SET	SET	SET	SET	SET	SET						
	10%					SET	SET	AET	SET	SET	SET						

[1] Drew with AET.
[2] Drew with RET, TET and AET.

APPENDIX C - PROPOSITIONS AND COROLLARIES PROOFS

In this appendix we set up the proofs of the main results obtained in Chapter 3 and Chapter 4.

Proof of Proposition 3.1. Let $\boldsymbol{\theta}_1 = (\beta_1,\lambda_1)$ and $\boldsymbol{\theta}_2 = (\beta_2,\lambda_2)$ ECR equivalent parametric points, i.e., $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta$ such that for all $p \in (0,1)$

$$\begin{aligned} & Q_{\boldsymbol{\theta}_1}(p) = Q_{\boldsymbol{\theta}_2}(p) \\ \iff & \frac{\lambda_1}{1-p^{1/\beta_1}}\sqrt{2p^{1/\beta_1}-p^{2/\beta_1}} = \frac{\lambda_2}{1-p^{1/\beta_2}}\sqrt{2p^{1/\beta_2}-p^{2/\beta_2}} \\ \iff & \frac{1}{1-p^{1/\beta_1}}\sqrt{2p^{1/\beta_1}-p^{2/\beta_1}} = \frac{\lambda}{1-p^{1/\beta_2}}\sqrt{2p^{1/\beta_2}-p^{2/\beta_2}} \quad \text{where } \lambda = \frac{\lambda_2}{\lambda_1} \\ \iff & \frac{1}{1-q}\sqrt{2q-q^2} = \frac{\lambda}{1-q^{1/\beta}}\sqrt{2q^{1/\beta}-q^{2/\beta}} \quad \text{where } \beta = \frac{\beta_2}{\beta_1} \text{ and } q = p^{1/\beta_1}. \end{aligned}$$

Then the existence of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ implies the existence of a parametric point $\boldsymbol{\theta} = (\beta,\lambda)$ which is ECR equivalent to $\boldsymbol{\theta}_0 = (1,1)$, i.e., for all $x > 0$

$$\begin{aligned} & F_{\boldsymbol{\theta}_0}(x) = F_{\boldsymbol{\theta}}(x) \\ \iff & 1 - \frac{1}{\sqrt{1+x^2}} = \left(1 - \frac{\lambda}{\sqrt{\lambda^2+x^2}}\right)^{\beta} \\ \iff & \log\left(1 - \frac{1}{\sqrt{1+x^2}}\right) = \beta \log\left(1 - \frac{\lambda}{\sqrt{\lambda^2+x^2}}\right). \end{aligned} \tag{C.1}$$

In particular consider $x = 1$ and $x = \lambda$ in Eq. (C.1). Then we obtain the following system:

$$\begin{cases} \log\left(1 - \frac{1}{\sqrt{2}}\right) & = \beta \log\left(1 - \frac{\lambda}{\sqrt{1+\lambda^2}}\right) \\ \beta \log\left(1 - \frac{1}{\sqrt{2}}\right) & = \log\left(1 - \frac{1}{\sqrt{1+\lambda^2}}\right). \end{cases} \tag{C.2}$$

Applying the first in the second Eq. (C.2) we can obtain:

$$\log\left(1 - \frac{1}{\sqrt{1+\lambda^2}}\right) \log\left(1 - \frac{\lambda}{\sqrt{1+\lambda^2}}\right) = \left[\log\left(1 - \frac{1}{\sqrt{2}}\right)\right]^2. \tag{C.3}$$

Consider the following change of variables:

$$u = 1 - \frac{1}{\sqrt{1+\lambda^2}} \iff \lambda = \frac{\sqrt{u(2-u)}}{1-u}, \quad u \in (0,1). \tag{C.4}$$

Applying (C.4) in Eq. (C.3) we obtain:

$$\log\left[1-\sqrt{u(2-u)}\right]\log u=\left[\log\left(1-\frac{1}{\sqrt{2}}\right)\right]^2.$$

This equation has only one solution $u=1-1/\sqrt{2}$ which implies that $\boldsymbol{\theta}=\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1=\boldsymbol{\theta}_2$. □

Proof of Proposition 3.2. Note that

$$\begin{aligned}\lim_{x\to 0^+} f(x) &= \beta\lambda \lim_{x\to 0^+} \frac{x}{(\lambda^2+x^2)^{\frac{3}{2}}}\left(1-\frac{\lambda}{\sqrt{\lambda^2+x^2}}\right)^{\beta-1}\\ &= \beta\lambda \lim_{x\to 0^+} \frac{x}{(\lambda^2+x^2)^{\frac{3}{2}}}\left(1-\frac{\lambda}{\sqrt{\lambda^2+x^2}}\right)^{\beta-1}\frac{\left(1+\frac{\lambda}{\sqrt{\lambda^2+x^2}}\right)^{\beta-1}}{\left(1+\frac{\lambda}{\sqrt{\lambda^2+x^2}}\right)^{\beta-1}}\\ &= \beta\lambda \lim_{x\to 0^+} \frac{x^{2\beta-1}}{(\lambda^2+x^2)^{\frac{1}{2}+\beta}}\left(1+\frac{\lambda}{\sqrt{\lambda^2+x^2}}\right)^{1-\beta}.\end{aligned}$$

Now the result in (3.5) is obtained. About the hrf limit note that

$$\lim_{x\to 0^+} h(x)=\lim_{x\to 0^+}\frac{f(x)}{1-F(x)}=\lim_{x\to 0^+} f(x).$$

Consider the following change of variables

$$u=1-\frac{\lambda}{\sqrt{\lambda^2+x^2}} \iff x=\lambda\frac{\sqrt{u(2-u)}}{1-u}, \tag{C.5}$$

Applying Eq. (C.5) we can set

$$\begin{aligned}\lim_{x\to\infty} h(x) &= \lim_{u\to 1^-} h(u)\\ &= \frac{\beta}{\lambda}\lim_{u\to 1^-}\frac{(2-u)^{1/2}(u-1)^2u^{\beta-1/2}}{1-u^\beta}.\end{aligned}$$

If $\beta\geq 1/2$ and applying the L'Hospital's rule we obtain

$$\begin{aligned}\lim_{x\to\infty} h(x) &= \frac{1}{\lambda}\lim_{u\to 1^-}(2-u)^{1/2}u^{1/2}(u-1)^2\left[\frac{1}{2(2-u)}+\frac{2}{u-1}+\frac{\beta-1/2}{u}\right]\\ &= 0.\end{aligned}$$

Otherwise if $\beta<1/2$ and applying the L'Hospital's rule we obtain

$$\begin{aligned}\lim_{x\to\infty} h(x) &= -\frac{\beta}{\lambda}\lim_{u\to 1^-}\frac{(u-1)(5u+9)(u^{1/2+\beta})}{(2-u)^{1/2}\left[(1-u^\beta)-2\beta\right]}\\ &= 0.\end{aligned}$$

□

Proof of Proposition 3.3. Let X be a ECR random variable and $c > 0$.

$$\begin{aligned}
\lim_{x\to\infty} \frac{S(cx)}{S(x)} &= \lim_{x\to\infty} \frac{1-\left(1-\frac{\lambda}{\sqrt{\lambda^2+(cx)^2}}\right)^{\beta}}{1-\left(1-\frac{\lambda}{\sqrt{\lambda^2+x^2}}\right)^{\beta}} \\
&= c^2 \lim_{x\to\infty} \left\{ \left[\frac{\lambda^2+x^2}{\lambda^2+(cx)^2}\right]^{3/2} \frac{\left(1-\frac{\lambda}{\sqrt{\lambda^2+(cx)^2}}\right)^{\beta-1}}{\left(1-\frac{\lambda}{\sqrt{\lambda^2+x^2}}\right)^{\beta-1}} \right\} \quad \text{(L'Hôpital's rule)} \\
&= c^2 \left(\frac{1}{c^2}\right)^{3/2} = \frac{1}{c}.
\end{aligned}$$

Then the tail index is $a = 1$. □

Proof of Proposition 3.4. Note that

$$\begin{aligned}
\frac{\mathrm{d}\log\left[f(x)\right]}{\mathrm{d}x} &= \frac{1}{x} - \frac{(\beta+2)x}{\lambda^2+x^2} + \frac{(\beta-1)x}{\left(\sqrt{\lambda^2+x^2}-\lambda\right)\sqrt{\lambda^2+x^2}} \\
&= \frac{\left(\lambda^2+x^2\right)\left(\sqrt{\lambda^2+x^2}-\lambda\right) - (\beta+2)x^2\left(\sqrt{\lambda^2+x^2}-\lambda\right) + (\beta-1)x^2\sqrt{\lambda^2+x^2}}{x\left(\lambda^2+x^2\right)\left(\sqrt{\lambda^2+x^2}-\lambda\right)} \\
&= \frac{x^2\left[(\beta+1)\lambda - 2\sqrt{\lambda^2+x^2}\right] + \lambda^2\left(\sqrt{\lambda^2+x^2}-\lambda\right)}{x\left(\lambda^2+x^2\right)\left(\sqrt{\lambda^2+x^2}-\lambda\right)} \\
&= \frac{\lambda^2\left[(\beta-1)\lambda + \beta\sqrt{\lambda^2+x^2}\right] - x^2\left[(1-\beta)\lambda + 2\sqrt{\lambda^2+x^2}\right]}{x\left(\lambda^2+x^2\right)^{3/2}}. \qquad \text{(C.6)}
\end{aligned}$$

The mode of the ECR pdf is the solution of

$$\frac{\mathrm{d}\log[f(x)]}{\mathrm{d}x} = 0. \qquad \text{(C.7)}$$

By the Eq. (C.6) the solution of Eq. (C.7) is obtained solving

$$\lambda^2\left[(\beta-1)\lambda + \beta\sqrt{\lambda^2+x^2}\right] - x^2\left[(1-\beta)\lambda + 2\sqrt{\lambda^2+x^2}\right] = 0.$$

After some manipulations we obtain

$$\frac{2x^2-\lambda^2\beta}{\sqrt{\lambda^2+x^2}} = (\beta-1)\lambda. \qquad \text{(C.8)}$$

For x_0 to be a solution of Eq. (C.8) it needs to satisfy the following conditions:

$$\beta > 1 \iff x_0 > \lambda\sqrt{\beta/2} \qquad \text{(C.9a)}$$

and

$$0 < \beta < 1 \iff 0 < x_0 < \lambda\sqrt{\beta/2}, \tag{C.9b}$$

the case $\beta = 1$ is trivial. Squaring both sides of Eq. (C.8) and simplifying we obtain a bi-quadratic equation

$$4x^4 - [(\beta+1)\lambda]^2 x^2 + \lambda^4(2\beta - 1) = 0. \tag{C.10}$$

Assuming $y = x^2$ the last expression became

$$4y^2 - [(\beta+1)\lambda]^2 y + \lambda^4(2\beta - 1) = 0. \tag{C.11}$$

The discriminant of this quadratic equation is

$$\Delta = (\beta-1)^2\left(\beta^2 + 6\beta + 17\right)\lambda^4,$$

which is always positive, then the two distinct real roots of Eq. (C.11) are

$$y_1 = \frac{\lambda^2}{8}\left[(\beta+1)^2 + \sqrt{(\beta-1)^2(\beta^2+6\beta+17)}\right]$$

and

$$y_2 = \frac{\lambda^2}{8}\left[(\beta+1)^2 - \sqrt{(\beta-1)^2(\beta^2+6\beta+17)}\right].$$

This solutions can be rewritten as

$$y_1^\star = \frac{\lambda^2}{8}\left[(\beta+1)^2 + (\beta-1)\sqrt{\beta^2+6\beta+17}\right]$$

and

$$y_2^\star = \frac{\lambda^2}{8}\left[(\beta+1)^2 - (\beta-1)\sqrt{\beta^2+6\beta+17}\right].$$

Note that for $\beta > 1$ we have $y_1 = y_1^\star$ and $y_2 = y_2^\star$, on the other hand if $\beta < 1$ we obtain $y_1 = y_2^\star$ and $y_2 = y_1^\star$ and finally, when $\beta = 1$ all solutions are equivalent. It is straightforward to prove that

$$y_1^\star \geq 0 \iff \beta \geq 1/2$$

and

$$y_2^\star \geq 0 \iff \beta \geq 0.$$

Then Eq. (C.10) has only two possible (real and non-negative) roots (for $\beta \geq 1/2$)

$$x_1 = \frac{\lambda}{2\sqrt{2}}\sqrt{(\beta+1)^2 + (\beta-1)\sqrt{\beta^2+6\beta+17}}, \quad \beta \geq 1/2 \tag{C.12a}$$

and

$$x_2 = \frac{\lambda}{2\sqrt{2}}\sqrt{(\beta+1)^2 - (\beta-1)\sqrt{\beta^2+6\beta+17}}, \quad \beta \geq 0. \tag{C.12b}$$

The Eq. (C.12b) do not satisfy any of the conditions in (C.9a) and (C.9b), then it is not a solution of Eq. (C.8). On the contrary the solution in (C.12a) satisfy both Eq. (C.9a) and

$$1/2 \leq \beta < 1 \iff 0 \leq x_1 < \lambda\sqrt{\beta/2},$$

which is (using Corollary 3.4.1) a restricted version of Eq. (C.9b). Then the unique solution of Eq. (C.8) is Eq. (C.12a). Note that the ECR pdf is not defined in zero and hence the ECR mode is given by Eq. (3.6). □

Proof of Proposition 3.5. Note that

$$\begin{aligned}
\Pr(X < x) &= \Pr(\lambda Z < x) \\
&= \Pr\left(Z < \frac{x}{\lambda}\right) \\
&= \left(1 - \frac{1}{\sqrt{1 - \left(\frac{x}{\lambda}\right)^2}}\right)^{\beta} \\
&= \left(1 - \frac{\lambda}{\sqrt{\lambda^2 + x^2}}\right)^{\beta}.
\end{aligned}$$

□

Proof of Proposition 3.7. Let $X \sim \text{ECR}(\beta,\lambda)$. Note that

$$\mu'_{r,s,t} = \beta\lambda \int_0^{\infty} x^r \frac{x}{(\lambda^2+x^2)^{3/2}} \left(1 - \frac{\lambda}{\sqrt{\lambda^2+x^2}}\right)^{(s+1)\beta-1} \left[1 - \left(1 - \frac{\lambda}{\sqrt{\lambda^2+x^2}}\right)^{\beta}\right]^t \mathrm{d}x.$$

We apply change of variable given in Eq. (C.5) and after some algebras we obtain

$$\mu'_{r,s,t} = \beta(\lambda\sqrt{2})^r \int_0^1 \frac{u^{r/2+(s+1)\beta-1}(1-u)^{-r}}{(1-u/2)^{-r/2}}(1-u^{\beta})^t \,\mathrm{d}u.$$

Using simple binomial expansion we can write

$$(1-u^{\beta})^t = \sum_{i=0}^{t}(-1)^i \binom{t}{i} i^{i\beta}.$$

Then we can express

$$\mu'_{r,s,t} = \beta(\lambda\sqrt{2})^r \sum_{i=0}^{t}(-1)^i \binom{t}{i} \int_0^1 \frac{u^{r/2+(s+i+1)\beta-1}(1-u)^{-r}}{(1-u/2)^{-r/2}} \,\mathrm{d}u. \tag{C.13}$$

Consider the following result

$$ {}_2\mathrm{F}_1(a,b;c;x) = \frac{1}{\mathrm{B}(b,c-b)} \int_0^1 \frac{t^{b-1}(1-t)^{c-b-1}}{(1-tx)^a}\,\mathrm{d}t, \tag{C.14} $$

showed in Bailey (1973, p. 4). The result is obtained applying Eq. (C.14) in (C.13). □

Proof of Corollary 3.7.2. Let $Z \sim \mathrm{CR}(1)$. Consider the following identity obtained by Bailey (1973, p. 11)

$$ {}_2\mathrm{F}_1(a,1-a;c;1/2) = \frac{\Gamma\left(\frac{c}{2}\right)\Gamma\left(\frac{1+c}{2}\right)}{\Gamma\left(\frac{a+c}{2}\right)\Gamma\left(\frac{1+c-a}{2}\right)}. \tag{C.15} $$

Assuming $\beta = 1$ in (3.8), applying Eq. (C.15) and after some algebras we can set

$$ \begin{aligned} \mathbb{E}(Z^r) &= \frac{2^{r/2+1}}{\sqrt{\pi}} \frac{\Gamma\left(\frac{1}{2}\left(2-\frac{r}{2}\right)\right)\Gamma\left(\frac{1}{2}\left(3-\frac{r}{2}\right)\right)}{\Gamma\left(\frac{2-r}{2}\right)} \mathrm{B}\left(1-r,\frac{r}{2}+1\right) \\ &= \frac{\Gamma\left(\frac{1-r}{2}\right)\,\Gamma\left(1+\frac{r}{2}\right)}{\sqrt{\pi}}. \end{aligned} $$

Then the results in (3.9) follows. □

Proof of Proposition 3.8. Note that

$$ \mathbb{E}(\log Z) = \beta \int_0^\infty \log z \frac{z}{(1+z^2)^{3/2}} \left\{1 - \frac{1}{\sqrt{1+z^2}}\right\}^{\beta-1} \mathrm{d}z. $$

We apply a simple change of variable assuming

$$ u = 1 - \frac{1}{\sqrt{1+z^2}} \iff z = \frac{\sqrt{u(2-u)}}{1-u}, $$

and after some algebras we can set

$$ \mathbb{E}(\log Z) = \frac{\beta}{2}\int_0^1 u^{\beta-1}\log u\,\mathrm{d}u + \frac{\beta}{2}\int_0^1 u^{\beta-1}\log(2-u)\,\mathrm{d}u - \beta\int_0^1 u^{\beta-1}\log(1-u)\,\mathrm{d}u. $$

The first integral is determined transforming $v = -\log u$. After this transform we obtain

$$ \int_0^1 u^{\beta-1}\log u\,\mathrm{d}u = -\int_0^\infty v\exp(-\beta v)\,\mathrm{d}v = -\frac{1}{\beta^2}. \tag{C.16} $$

In order to determine the second integral we need to look upon the following known expansion

$$ \log(2-u) = \log 2 - \sum_{n=1}^\infty \frac{1}{n}\left(\frac{u}{2}\right)^n. $$

Then we can express the second integral as

$$\begin{aligned}\int_0^1 u^{\beta-1}\log(2-u)\,\mathrm{d}u &= (\log 2)\int_0^1 u^{\beta-1}\,\mathrm{d}u - \sum_{n=1}^{\infty}\frac{1}{n2^n}\int_0^1 u^{\beta+n-1}\,\mathrm{d}u \\ &= \frac{\log 2}{\beta} - \sum_{n=1}^{\infty}\frac{1}{n(n+\beta)2^n} \\ &= \frac{\log 2}{\beta} - \frac{1}{\beta}\sum_{n=1}^{\infty}\left[\frac{1}{n2^n} - \frac{1}{(n+\beta)2^n}\right] \\ &= \frac{1}{\beta}\left[\sum_{n=0}^{\infty}\frac{1}{(n+\beta)2^n} - \frac{1}{\beta}\right] \\ &= \frac{1}{\beta}\left[\boldsymbol{\Phi}\left(\frac{1}{2};1,\beta\right) - \frac{1}{\beta}\right]. \end{aligned} \tag{C.17}$$

Consider another well known expansion

$$\log(1-u) = -\sum_{n=1}^{\infty}\frac{u^n}{n},$$

then the last integral is

$$\begin{aligned}\int_0^1 u^{\beta-1}\log(1-u)\,\mathrm{d}u &= -\sum_{n=1}^{\infty}\frac{1}{n}\int_0^1 u^{\beta+n-1}\,\mathrm{d}u \\ &= -\sum_{n=1}^{\infty}\frac{1}{n(\beta+n)} \\ &= -\frac{1}{\beta}\left[\Psi(1+\beta)+\right]. \end{aligned} \tag{C.18}$$

In the last step we use the identity obtained from Abramowitz and Stegun (1972, eq. 6.3.16)

$$\Psi(1+\beta) = -\boldsymbol{\gamma} + \sum_{n=1}^{\infty}\frac{\beta}{n(\beta+n)}.$$

Then we can set

$$\mathbb{E}(\log Z) = \frac{1}{2}\boldsymbol{\Phi}\left(\frac{1}{2};1,\beta\right) + \Psi(1+\beta) + \boldsymbol{\gamma} - \frac{1}{\beta},$$

and the result follows by applying Eq. (3.7). □

Proof of Proposition 3.9. The incomplete moments for a random variable $X \sim \mathrm{ECR}(\beta,\lambda)$ are defined by

$$m_r(x_0) = \int_{-\infty}^{x_0} x^r f_X(x)\,\mathrm{d}x.$$

Then we need to determine

$$m_r(x_0) = \beta\lambda \int_0^{x_0} x^r \frac{x}{(\lambda^2 + x^2)^{3/2}} \left(1 - \frac{\lambda}{\sqrt{\lambda^2 + x^2}}\right)^{\beta-1} \mathrm{d}x.$$

Assuming the change of variable in (C.5) and after some algebras we can express

$$m_r(x_0) = \beta(\lambda\sqrt{2})^r u_0 \int_0^{u_0} u^{r/2+\beta-1}(1-u)^{-r} \left(1 - \frac{u}{2}\right)^{r/2} \mathrm{d}u,$$

where $u_0 = 1 - \lambda/\sqrt{x_0^2+\lambda^2}$. Now assuming $t = u/u_0$ we obtain

$$m_r(x_0) = \beta(\lambda\sqrt{2})^r u_0^{r/2+\beta} \int_0^1 t^{r/2+\beta-1}(1-u_0 t)^{-r} \left[1 - \left(\frac{u_0}{2}\right) t\right]^{r/2} \mathrm{d}t. \tag{C.19}$$

Consider the following result

$$\mathrm{F}_1(a,b_1,b_2,c;x,y) = \frac{1}{\mathrm{B}(a,c-a)} \int_0^1 t^{a-1}(1-t)^{c-a-1}(1-xt)^{-b_1}(1-yt)^{-b_2} \,\mathrm{d}t, \tag{C.20}$$

due to (BAILEY, 1973, p. 77) where F_1 is the first Appell's hypergeometric function which is defined by

$$\mathrm{F}_1(a,b_1,b_2;c;x,y) = \sum_{i,j=0}^{\infty} \frac{(a)_{i+j}(b_1)_i(b_2)_j}{(c)_{i+j}} \frac{x^i}{i!}\frac{y^j}{j!}.$$

The convergence is obtained when $|x| < 1$, $|y| < 1$ and $c > a > 0$. The result is obtained by applying Eq. (C.20) in (C.19). □

Proof of Proposition 3.10. We need to determine the following integral

$$\mathbb{E}(X_{i:n}^r) = \frac{\beta\lambda}{\mathrm{B}(i,n-i+1)} \int_0^{\infty} x^r \frac{x}{(\lambda^2+x^2)^{3/2}} \left(1 - \frac{\lambda}{\sqrt{\lambda^2+x^2}}\right)^{i\beta-1} \times \\ \times \left[1 - \left(1 - \frac{\lambda}{\sqrt{\lambda^2+x^2}}\right)^{\beta}\right]^{n-i} \mathrm{d}x.$$

Using the change of variable in (C.5) and after some algebras we can set

$$\mathbb{E}(X_{i:n}^r) = \frac{\beta(\lambda\sqrt{2})^r}{\mathrm{B}(i,n-i+1)} \int_0^1 \frac{u^{r/2+i\beta-1}(1-u)^{-r}}{(1-u/2)^{-r/2}} (1-u^{\beta})^{n-i} \,\mathrm{d}u. \tag{C.21}$$

Note that we can express

$$(1-u^{\beta})^{n-i} = \sum_{j=0}^{n-i} (-1)^j \binom{n-i}{j} u^{j\beta}. \tag{C.22}$$

Now applying Eq. (C.22) in (C.21) we obtain

$$\mathbb{E}(X_{i:n}^r) = \frac{\beta(\lambda\sqrt{2})^r}{\mathrm{B}(i,n-i+1)} \sum_{j=0}^{n-i} (-1)^j \binom{n-i}{j} \int_0^1 \frac{u^{r/2+(i+j)\beta-1}(1-u)^{-r}}{(1-u/2)^{-r/2}}\,\mathrm{d}u.$$

Finally using the identity in (C.14) the result is obtained. □

Proof of Proposition 3.11. Here we present a resumed proof of this result. Note that $\hat{\beta}(\lambda)$ is a continuous decreasing function of λ and

$$\lim_{\lambda\to 0} \hat{\beta}(\lambda) = \infty$$

and

$$\lim_{\lambda\to\infty} \hat{\beta}(\lambda) = 0.$$

Then for a given interval $\Theta_\beta = [\beta_1,\beta_2]$ with $0 < \beta_1 < \beta_2 < \infty$, there exists an interval $[\lambda_2,\lambda_1]$ such that $\beta_1 = \hat{\beta}(\lambda_2)$ and $\beta_2 = \hat{\beta}(\lambda_1)$. The interval Θ_β is the image of the interval $[\lambda_2,\lambda_1]$ by the function $\hat{\beta}(\lambda)$. Finally

$$\lambda_1 = \arg\min_\lambda \left|\hat{\beta}(\lambda) - \sup\Theta_\beta\right|$$

and

$$\lambda_2 = \arg\min \lambda \left|\hat{\beta}(\lambda) - \inf\Theta_\beta\right|.$$

The cases where $\Theta_\beta = (\beta_1,\beta_2]$, $\Theta_\beta = (\beta_1,\beta_2)$, $\Theta_\beta = [\beta_1,\beta_2)$, $\Theta_\beta = (0,\beta_2]$, $\Theta_\beta = (\beta_1,\infty)$ and $\Theta_\beta = [\beta_1,\infty)$ are analogous. □

Proof of Proposition 3.12. Analogous to the proof of Proposition 3.11. □

Proof of Proposition 3.15. Let U be the ECR llf hessian matrix (observed information).

$$U = \begin{bmatrix} U_{\beta\beta} & U_{\beta\lambda} \\ U_{\lambda\beta} & U_{\lambda\lambda} \end{bmatrix}.$$

Its components are given by

$$U_{\beta\beta} = -\frac{n}{\beta^2}, \tag{C.23}$$

$$U_{\beta\lambda} = -\sum_{i=1}^{n} \frac{\sqrt{\lambda^2 + {x_i}^2} + \lambda}{\lambda^2 + {x_i}^2}$$

and

$$U_{\lambda\lambda} = -\sum_{i=1}^{n} \frac{{x_i}^4 + (\beta+4)\lambda^2{x_i}^2 - \lambda^3\left[(\beta+1)\lambda + (\beta-1)\sqrt{\lambda^2+{x_i}^2}\right]}{\lambda^2\left(\lambda^2+{x_i}^2\right)^2}.$$

The result in (3.21a) is derived taking the expectation of the constant obtained in (C.23). The second element in (3.21b) is obtained by the following integral.

$$\kappa_{\beta\lambda} = -n\beta\lambda \int_0^\infty \frac{x\left(\lambda + \sqrt{\lambda^2 + x^2}\right)}{\left(\lambda^2 + x^2\right)\left(\lambda^2 + x^2\right)^{3/2}} \left(1 - \frac{\lambda}{\sqrt{\lambda^2 + x^2}}\right)^{\beta-1} \mathrm{d}x.$$

Applying the transform in (C.5) and then after some algebras we obtain

$$\kappa_{\beta\lambda} = -\frac{n\beta}{\lambda} \int_0^1 (u-2)(u-1)u^{\beta-1}\,\mathrm{d}u,$$

which is a simple integral whose value is given by

$$\kappa_{\beta\lambda} = -\frac{n}{\lambda}\left[\frac{\beta+4}{(\beta+1)(\beta+2)}\right]$$

and after partial fractions decomposition we obtain (3.21b).

The third element in (3.21c) is provided by the following integral

$$\kappa_{\lambda\lambda} = -n\beta\lambda \int_0^\infty x \left\{ \frac{x^4 + (\beta+4)\lambda^2x^2 - \lambda^3\left[(\beta+1)\lambda + (\beta-1)\sqrt{\lambda^2+x^2}\right]}{\lambda^2\left(\lambda^2+x^2\right)^2\left(\lambda^2+x^2\right)^{3/2}} \right\} \times \\ \times \left(1 - \frac{\lambda}{\sqrt{\lambda^2+x^2}}\right)^{\beta-1} \mathrm{d}x.$$

Applying the change of variable in (C.5) and after some algebras we obtain

$$\kappa_{\lambda\lambda} = -\frac{n\beta}{\lambda^2}\left\{\int_0^1 \left[2(\beta+2)u^4 - 3(3\beta+5)u^3 + (14\beta+19)u^2 - 9(\beta+1)u + 2\beta\right] u^{\beta-1}\,\mathrm{d}u\right\},$$

which is a simple integral whose value is

$$\kappa_{\lambda\lambda} = -\frac{n\beta}{\lambda^2}\left[\frac{\beta^2+11\beta+36}{(\beta+2)(\beta+3)(\beta+4)}\right],$$

and after partial fractions decomposition we obtain (3.21c). □

Proof of Corollary 3.15.1. Note that in the FIM $\kappa_{\beta\beta} < 0$ and its determinant is given by

$$\boldsymbol{D} = \left(\frac{n}{\lambda}\right)^2 \frac{\beta^3 - 7\beta^2 + 10\beta + 72}{\beta(\beta+1)^2(\beta+2)^2(\beta+3)(\beta+4)}.$$

Consider the polynomial equation defined by the last factor numerator of $\boldsymbol{D}$:

$$\beta^3 - 7\beta^2 + 10\beta + 72 = 0.$$

This polynomial equation has only one real root given by

$$\beta_0 = \frac{1}{3}\left(-\sqrt[3]{944 - 27\sqrt{1213}} - \frac{19}{\sqrt[3]{944 - 27\sqrt{1213}}} + 7\right) \approx -2.2966\ldots,$$

and the associated polynomial do not assume negative values when $\beta > \beta_0$. Then $\boldsymbol{D} > 0$ and the result is proved. □

Proof of Proposition 3.16. Note that

$$U_{\beta\beta\beta} = \frac{2n}{\beta^3}, \tag{C.24a}$$

$$U_{\beta\beta\lambda} = 0, \tag{C.24b}$$

$$U_{\beta\lambda\lambda} = \sum_{i=1}^{n} \frac{\lambda\left(\lambda + \sqrt{\lambda^2 + x_i^2}\right) - x_i^2}{\left(\lambda^2 + x_i^2\right)^2}, \tag{C.24c}$$

and

$$U_{\lambda\lambda\lambda} = \sum_{i=1}^{n} \frac{2x_i^6 + 6\lambda^2 x_i^4 - 2\lambda^5\left[(\beta+1)\lambda + (\beta-1)\sqrt{\lambda^2 + x_i^2}\right]}{\lambda^3\left(\lambda^2 + x_i^2\right)^3} + \sum_{i=1}^{n} \frac{\lambda^3 x_i^2\left[6(\beta+3)\lambda + (\beta-1)\sqrt{\lambda^2 + x_i^2}\right]}{\lambda^3\left(\lambda^2 + x_i^2\right)^3} \tag{C.24d}$$

The results in (3.24a) and (3.24b) are derived taking the expected values of the constants obtained in Eqs. (C.24a) and (C.24b) respectively. The result in (3.24c) is achieved by taking the expected value of Eq. (C.24c) which is determined by the integral

$$\kappa_{\beta\lambda\lambda} = -n\beta\lambda \int_0^\infty x\left[\frac{\lambda\left(\lambda + \sqrt{\lambda^2 + x^2}\right) - x^2}{\left(\lambda^2 + x^2\right)^2\left(\lambda^2 + x^2\right)^{3/2}}\right]\left(1 - \frac{\lambda}{\sqrt{\lambda^2 + x^2}}\right)^{\beta-1} \mathrm{d}x.$$

Applying the change of variable in (C.5) it is possible to rewrite this integral in the form

$$\kappa_{\beta\lambda\lambda} = \frac{n\beta}{\lambda^2} \int_0^1 (u-2)(u-1)^2(2u-1)u^{\beta-1}\,\mathrm{d}u,$$

which integrand is polynomial and can be evaluated as

$$\kappa_{\beta\lambda\lambda} = -\frac{2n}{\lambda^2}\left[\frac{\beta^2 + 4\beta - 24}{(\beta+1)(\beta+2)(\beta+3)(\beta+4)}\right]$$

and the result in (3.24c) follows after partial fractions decomposition.

The result in (3.24d) is obtained by the following integral

$$\kappa_{\lambda\lambda\lambda} = n\beta\lambda \int_0^\infty x \left\{ \frac{2x^6 + 6\lambda^2 x^4 - 2\lambda^5 \left[(\beta+1)\lambda + (\beta-1)\sqrt{\lambda^2+x^2}\right]}{\lambda^3 \left(\lambda^2+x^2\right)^3 \left(\lambda^2+x^2\right)^{3/2}} \right\} \left(1 - \frac{\lambda}{\sqrt{\lambda^2+x^2}}\right)^{\beta-1} + x \left\{ \frac{\lambda^3 x^2 \left[6(\beta+3)\lambda + (\beta-1)\sqrt{\lambda^2+x^2}\right]}{\lambda^3 \left(\lambda^2+x^2\right)^3 \left(\lambda^2+x^2\right)^{3/2}} \right\} \left(1 - \frac{\lambda}{\sqrt{\lambda^2+x^2}}\right)^{\beta-1} \mathrm{d}x$$

Applying the change of variable in (C.5) and after some algebras it is possible to rewrite this integral in the form

$$\kappa_{\lambda\lambda\lambda} = \frac{n\beta}{\lambda^3} \int_0^1 \left[-8(\beta+2)u^6 + (51\beta+93)u^5 - 3(43\beta+71)u^4 + 3(55\beta+81)u^3 - 3(37\beta+47)u^2 + 36(\beta+1)u - 4\beta\right] u^{\beta-1}\,\mathrm{d}u,$$

which integrand is polynomial and can be evaluated as

$$\kappa_{\lambda\lambda\lambda} = \frac{2n}{\lambda^3} \left[\frac{\beta\left(\beta^4 + 20\beta^3 + 158\beta^2 + 691\beta + 1866\right)}{(\beta+2)(\beta+3)(\beta+4)(\beta+5)(\beta+6)}\right]$$

and the result in (3.24d) follows after partial fractions decomposition. □

Proof of Proposition 3.17. By Eq. (2.14) we can express the first order bias of the MLE as

$$^{(1/n)}\mathrm{Bias}(\hat{\theta}_i) = \sum_{r,s,t} \kappa^{\theta_i,\theta_r} \kappa^{\theta_s,\theta_t} \left(\kappa^{(\theta_t)}_{\theta_r\theta_s} - \frac{1}{2}\kappa_{\theta_r\theta_s\theta_t}\right). \tag{C.25}$$

Applying the Corollary 3.15.2, Corollary 3.15.3 and Proposition 3.16 in (C.25) the result is obtained after elementary algebras. □

Proof of Lemma 4.1. Note that

$$\boldsymbol{I}(q,\boldsymbol{\theta}) = (\beta\lambda)^q \int_0^\infty \left[\frac{x}{(\lambda^2+x^2)^{3/2}} \left(1 - \frac{\lambda}{\sqrt{\lambda^2+x^2}}\right)^{\beta-1}\right]^q \mathrm{d}x$$

Applying the change of variables in (C.5) and then after some algebras we can set

$$\boldsymbol{I}(q,\boldsymbol{\theta}) = \beta^q \left(\frac{\lambda}{\sqrt{2}}\right)^{1-q} \int_0^1 \frac{u^{q(\beta-1/2)-1/2}(1-u)^{2(q-1)}}{(1-u/2)^{(1-q)/2}}\,\mathrm{d}u$$

Now we use the identity in (C.14) and the result is obtained. □

Proof of Lemma 4.2. The first gradient component is

$$\begin{aligned}\frac{\partial \boldsymbol{I}(q,\boldsymbol{\theta})}{\partial \beta} &= \int_0^\infty \frac{\partial}{\partial \beta}\left[f(x)\right]^q \,\mathrm{d}x \\ &= q\int_0^\infty \left[f(x)\right]^{q-1}\frac{\partial f(x)}{\partial \beta}\,\mathrm{d}x.\end{aligned}$$

Note that

$$\frac{\partial f(x)}{\partial \beta} = \lambda\frac{x}{(\lambda^2+x^2)^{3/2}}\left(1-\frac{\lambda}{\sqrt{\lambda^2+x^2}}\right)^{\beta-1}\left[1+\beta\log\left(1-\frac{\lambda}{\sqrt{\lambda^2+x^2}}\right)\right].$$

Then we can express

$$\frac{\partial \boldsymbol{I}(q,\boldsymbol{\theta})}{\partial \beta} = q\left[\frac{\boldsymbol{I}(q,\boldsymbol{\theta})}{\beta}+\int_0^\infty \left[f(x)\right]^q\log\left(1-\frac{\lambda}{\sqrt{\lambda^2+x^2}}\right)\,\mathrm{d}x\right]$$

Applying the change of variables in (C.5) and then after some algebras we can set

$$\begin{aligned}\int_0^\infty \left[f(x)\right]^q\log\left(1-\frac{\lambda}{\sqrt{\lambda^2+x^2}}\right)\,\mathrm{d}x &= \frac{\beta^q}{\lambda^{q-1}}\int_0^1 \frac{u^{q(\beta-1/2)-1/2}\ln u}{(2-u)^{(1-q)/2}(1-u)^{2(1-q)}}\,\mathrm{d}u \\ &= \frac{\Upsilon(q,\beta)}{\lambda^{q-1}}.\end{aligned}$$

□

Proof of Proposition 4.2. By definition, the cross Shannon entropy is given by

$$\mathcal{H}_S^c(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = -\int_{-\infty}^\infty f_1(x)\log f_2(x)\,\mathrm{d}x,$$

then we can set

$$\mathcal{H}_S^c(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = -\beta_1\lambda\int_0^\infty \log\left[\beta_2\lambda\frac{x}{(\lambda^2+x^2)^{3/2}}\left(1-\frac{\lambda}{\sqrt{\lambda^2+x^2}}\right)^{\beta_2-1}\right]\frac{x}{(\lambda^2+x^2)^{3/2}}\left(1-\frac{\lambda}{\sqrt{\lambda^2+x^2}}\right)^{\beta_1-1}$$

Using again the change of variable in (C.5) we obtain

$$\begin{aligned}\mathcal{H}_S^c(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) =& -\beta_1\log\left(\frac{\beta_2}{\lambda}\right)\int_0^1 u^{\beta_1-1}\,\mathrm{d}u - 2\beta_1\int_0^1 u^{\beta_1-1}\log(1-u)\,\mathrm{d}u \\ &-\beta_1\left(\beta_2-\frac{1}{2}\right)\int_0^1 u^{\beta_1-1}\log u\,\mathrm{d}u - \frac{\beta_1}{2}\int_0^1 u^{\beta_1-1}\log(2-u)\,\mathrm{d}u.\end{aligned}$$

The first integral is straightforward and the last three were determined in Eqs. (C.16)–(C.18), then after some algebras the result is obtained. □