UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

MARLEY APOLINARIO SARAIVA

CLASSE DE DISTRIBUIÇÕES MISTURA DE ESCALA T-STUDENT E ENSAIOS EM ANÁLISE DE DADOS CIRCULARES: distribuição wrapped Birnbaum-Saunders

Recife

2019

MARLEY APOLINARIO SARAIVA

CLASSE DE DISTRIBUIÇÕES MISTURA DE ESCALA T-STUDENT E ENSAIOS EM ANÁLISE DE DADOS CIRCULARES: distribuição wrapped Birnbaum-Saunders

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.

Área de Concentração: Matemática / Probabilidade e Estatística.

Orientador: Prof. Dr. Francisco José de Azevêdo Cysneiros
Coorientador: Prof. Dr. Aldo William Medina Garay

Recife
2019

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

S243c Saraiva, Marley Apolinario
Classe de distribuições mistura de escala *t-student* e ensaios em análise de dados circulares: distribuição *wrapped* Birnbaum-Saunders / Marley Apolinario Saraiva. – 2019.
131 f.: il., fig., tab.

Orientador: Francisco José de Azêvedo Cysneiros.
Tese (Doutorado) – Universidade Federal de Pernambuco. CCEN, Estatística, Recife, 2019.
Inclui referências e apêndices.

1. Estatística. 2. Dados circulares. I. Cysneiros, Francisco José de Azêvedo (orientador). II. Título.

310 CDD (23. ed.) UFPE- MEI 2019-027

MARLEY APOLINARIO SARAIVA

CLASSE DE DISTRIBUIÇÕES MISTURA DE ESCALA T-STUDENT E ENSAIOS EM ANÁLISE DE DADOS CIRCULARES: DISTRIBUIÇÃO WRAPPED BIRNBAUM-SAUNDERS

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: 26 de fevereiro de 2019.

BANCA EXAMINADORA

Prof.(º) Francisco José de Azevêdo Cysneiros
UFPE

Prof.(º) Audrey Helen Mariz de Aquino Cysneiros
UFPE

Prof.(º) Roberto Ferreira Manghi
UFPE

Prof.(º) Clécio da Silva Ferreira
UFJF

Prof.(º) Jalmar Manuel Farfan Carrasco
UFBA

Dedico este trabalho à Patrícia, Heitor e Ivan.

AGRADECIMENTOS

Agradeço à Deus.

Agradeço à minha esposa, Patrícia. Eu não consigo traduzir em palavras o quanto lhe sou grato, por tudo. Por seu cuidado com nossos filhos enquanto estive ausente, por sua infinita paciência e colaboração comigo, por seu amor e carinho, por sua preocupação com minha saúde e meu bem estar, por sua compreensão por não termos tanto tempo juntos quanto gostaríamos. Sem ela eu não teria conseguido. TE AMO.

Agradeço aos meus filhos, Heitor e Ivan, que são minha principal motivação, por brincarem comigo pra me animar e por me acordarem pontualmente às 6h todos os dias. Papai Ama.

Agradeço aos meus pais, Manoel e Deni, por me ensinarem a nunca desistir.

Agradeço ao meu amigo David pelas longas conversas sobre o doutorado.

Agradeço a todos os colegas da pós-graduação em estatística, aos colegas de turma, Thiago, Renata e Fernanda e, em especial a Wenia e Luana. Obrigado por me suportarem.

Agradeço à Fran pelos muitos cafés com conselhos.

Agradeço ao Prof. Francisco Cysneiros pela orientação e ao Prof. Aldo Garay pela coorientação.

Agradeço aos professores com os quais tive a oportunidade de cursar disciplinas.

Agradeço à Valéria Bittencourt, secretária da pós-graduação, pelo apoio, disponibilidade e paciência.

Agraço à FACEPE pelo apoio financeiro.

RESUMO

Nesta tese são apresentados resultados de duas linhas de pesquisa distintas, a saber, análise de dados circulares e modelos mistura de escala. Na primeira delas, na análise de dados circulares, foi proposta uma nova distribuição circular denominada de *wrapped* Birnbaum-Saunders. Para esta distribuição foram encontradas expressões para sua função densidade de probabilidade, função de distribuição acumulada e momentos trigonométricos, além disso, realizamos um estudo de simulação de Monte Carlo com o objetivo de avaliar o desempenho dos estimadores de máxima verossimilhança dos parâmetros do modelo e também fizemos uma aplicação a um conjunto de dados reais. A respeito da segunda linha de pesquisa abordada neste trabalho, foi proposta uma nova classe de distribuições denominada de mistura de escala t-Student (SMt). Calculamos alguns de seus momentos e demonstramos que uma distribuição desta classe converge para a distribuição da classe mistura de escala Normal (SMN) com mesmo fator de escala quando os graus de liberdade vão para o infinito. Também mostrou-se que esta classe possui curtose maior que da classe SMN sendo potencialmente mais apropriada para explicar dados com observações aberrantes. Caracterizamos algumas distribuições pertencentes a esta classe (t-Student, t-Student contaminada, Weibull-t e Slash-t) e utilizamos o método dos momentos e o método da máxima verossimilhança para obtenção dos estimadores dos parâmetros dos modelos dessa nova classe proposta. Obtivemos expressões para o erro padrão assintótico dos estimadores por meio da matriz de informação observada de Fisher e foi realizado um estudo de simulação de Monte Carlo para avaliar o desempenho dos estimadores. Além disso, foi feita uma aplicação a dados reais.

Palavras-chave: Algoritmo EM. Dados circulares. Distribuição Birnbaum-Saunders. Distribuição t-Student. Distribuição *wrapped*. Mistura de escala de distribuições.

ABSTRACT

In this thesis we presented the results of two distinct lines of research, namely circular data analysis and scale mixture models. In the first one, in the circular data analysis, a new circular distribution called *wrapped* Birnbaum-Saunders was proposed. For this distribution we have found expressions for their probability density function, distribution function and trigonometric moments, in addition, we conducted a Monte Carlo simulation study to evaluate the performance of the of the maximum likelihood estimators of the parameters and also we made an application to a real dataset. Regarding the second line of research addressed in this thesis, a new class of distributions called Scale Mixture of Student-t (SMt) was proposed. We compute some of its moments and show that a distribution of this class converges to the distribution of Scale Mixture of Normal (SMN) class with the same scale factor when the degrees of freedom goes to infinity. It has also been shown that this class has greater kurtosis than SMN class being potentially more appropriate to explain data with aberrant observations. We characterized some distributions belonging to this class (Student-t, contaminated Student-t, Weibull-t and Slash-t) and we used the method of moments and the maximum likelihood method to obtain the estimators of the parameters of the proposed new class. We obtained expressions for the standard asymptotic error of the estimates through Fisher's observed information matrix and we also conducted a simulation study to evaluate the performance of the parameter estimates. In addition, an application was made to actual data.

Keywords: Circular data. Birnbaum-Saunders distributions. EM algorithm. Scale mixture distributions. Student-t distribution. *Wrapped* Distribution.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

a.a.	Amostra aleatória
EQM	Erro quadrático médio
EM	Algoritmo EM
EMM	Estimador pelo método dos momentos
EMV	Estimador de máxima verossimilhança
fc	Função característica
fda	Função de distribuição acumulada
fdp	Função densidade de probabilidade
MD	Maximização direta
MM	Método dos momentos
MV	Máxima verossimilhança
Normal-C	Normal contaminada
SMN	Mistura de Escala Normal
SMt	Mistura de Escala t-Student
t-C	t-Contaminada
v.a.	variável aleatória
VR	Viés Relativo
WBS	Wrapped Birnbaum-Saunders
W-t	Weibull-t

SUMÁRIO

1 INTRODUÇÃO

Nesta tese apresentamos resultados de duas linhas de pesquisa distintas, a saber, análise de dados circulares e modelos mistura de escala. Dados circulares surgem naturalmente em muitos campos da ciência, tal como medicina, biologia, física, meteorologia, entre outros (MARDIA; JUPP, 2000) e, em geral, as medidas circulares associadas a estes dados são registradas em fenômenos direcionais ou periódicos, como, por exemplo, a direção do movimento de um animal após um estímulo (direcional) ou o horário da chegada de um paciente a uma sala de emergência de um hospital (periódico). Outros exemplos de dados circulares podem ser obtidos em Jammalamadaka e SenGupta (2001), dos quais é possível destacar os dados de migrações de pássaros em que biólogos registram a direção do voo de pássaros recém liberados à medida em que desaparecem no horizonte ou também os dados de uma aplicação à medicina em que o ângulo de flexão do joelho foi medido para avaliar a recuperação de pacientes.

A motivação para trabalhar com análise de dados circulares surgiu da necessidade de se modelar dados circulares assimétricos, como por exemplo, os dados das direções do movimento de formigas ao serem submetidas a um estímulo (FISHER, 1995). Ainda são raros os modelos probabilísticos circulares assimétricos com expressões dos momentos trigonométricos tratáveis e, por isso, propomos um novo modelo probabilístico obtido pelo método *wrapped* a partir da distribuição Birnbaum-Saunder reparametrizada (SANTOS-NETO et al., 2012).

Com relação à segunda linha de pesquisa, a motivação surgiu da necessidade de se obter modelos probabilísticos com caudas ainda mais pesadas do que os que já são conhecidos na literatura quando os dados são unimodais e simétricos. Esta necessidade têm se mostrado recorrente historicamente e podemos citar como exemplo o surgimento da distribuição t-Student como uma alternativa à distribuição normal, justamente pela necessidade de um modelo probabilístico com caudas mais pesadas que o modelo normal. Posteriormente, surgiu a classe de distribuições mistura de escala normal (ANDREWS; MALLOWS, 1974), da qual a distribuição t-Student pertence, para suprir novamente a necessidade de modelos com caudas mais pesadas. A classe SMN, cuja definição dada em Garay et al. (2017b),

está a definida a seguir:

Definição 1.0.1. Diz-se que uma variável aleatória X tem distribuição Mistura de Escala Normal com média μ, parâmetro de escala σ^2 e fator de escala U se possui a seguinte representação estocástica:

$$X = \mu + U^{-\frac{1}{2}} Z, \quad Z \perp U \tag{1.1}$$

em que $Z \perp U$ representa que as v.a. Z e U são independentes, $Z \sim N(0, \sigma^2)$, isto é, Z segue uma distribuição normal com média zero e variância σ^2, $\mu \in \mathbb{R}$, $\sigma^2 > 0$ e U é uma variável aleatória positiva com função de distribuição acumulada $H(\cdot|\boldsymbol{\lambda})$ indexada pelo parâmetro $\boldsymbol{\lambda}$, que pode ser escalar ou vetor. Será utilizada a notação $X \sim \text{SMN}(\mu, \sigma^2; \boldsymbol{\lambda})$. Quando $\mu = 0$ e $\sigma^2 = 1$ tem-se que X segue uma distribuição SMN padrão e denota-se $X \sim \text{SMN}(0, 1; \boldsymbol{\lambda})$.

Nota-se de (1.1) que $X|U \sim N(\mu, u^{-1}\sigma^2)$. Portanto, integrando em U a densidade conjunta de X e U obtém-se a densidade marginal de X:

$$f_{SMN}\left(x|\mu, \sigma^2; \boldsymbol{\lambda}\right) = \frac{1}{\sqrt{2\pi}\sigma} \int_0^\infty \sqrt{u} \exp\left[-\frac{u}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] dH\left(u|\boldsymbol{\lambda}\right).$$

em que U é chamada de fator de escala e $H(\cdot|\boldsymbol{\lambda})$ é a distribuição de mistura. Dependendo da escolha do fator de escala U temos uma distribuição em particular da classe SMN. Alguns exemplos de distribuições que pertencem à classe SMN são a própria distribuição normal, a distribuição t-Student, a distribuição normal contaminada, a distribuição Slash, entre outras.

Assim, propomos uma nova classe de distribuições com caudas pesadas chamada de mistura de escala t-Student, que é definida, como veremos mais adiante, substituindo a variável aleatória Z em (1.0.1) por uma T com distribuição t-Student.

As duas linhas de pesquisa mencionadas são independentes uma da outra e por isso estão apresentadas na tese em capítulos isolados. No Capítulo 2 apresentamos a distribuição Wrapped Birnbaum-Saunders , sua função densidade de probabilidade, função de distribuição acumulada e momentos trigonométricos, além disso, realizamos um estudo de simulação de Monte Carlo com o objetivo de avaliar o desempenho dos estimadores de máxima verossimilhança dos parâmetros do modelo e também fizemos uma aplicação a um conjunto de dados reais. No Capítulo 3 apresentamos a classe mistura de escala

t-Student (SMt), calculamos alguns de seus momentos e demonstramos que esta classe converge para a classe mistura de escala Normal (SMN) quando os graus de liberdade vão para o infinito. Mostramos também que esta classe possui curtose maior que da classe SMN sendo potencialmente mais apropriada para explicar dados com observações aberrantes. Nós utilizamos o método dos momentos e o método da máxima verossimilhança para obtenção dos estimadores dos parâmetros dos modelos e obtivemos expressões para o erro padrão assintótico dos estimadores por meio da matriz de informação observada de Fisher. Finalmente, realizamos um estudo de simulação de Monte Carlo para avaliar o desempenho dos estimadores e aplicamos o modelo proposto a um conjuntos de dados reais.

2 ENSAIOS SOBRE DADOS CIRCULARES

As distribuições circulares são utilizadas para modelar dados circulares e estas podem ser caracterizadas por serem simétricas ou assimétricas, amodal, unimodal ou multimodal, entre outras características de interesse. Mesmo que grande parte das distribuições circulares sejam simétricas, existem várias situações práticas em que distribuições assimétricas são necessárias e, em geral, é difícil lidar com modelos assimétricos e essa dificuldade, em parte, é devida à falta de algumas propriedades matemáticas que os modelos simétricos possuem. Por exemplo, os modelos assimétricos geralmente possuem constantes normalizadoras complexas (número complexo) e também momentos trigonométricos complexos, o que pode causar problemas nas análises.

Há várias maneiras de se obter uma distribuição circular, uma delas é por meio do enrolamento (*wrapping*, em inglês) em torno do círculo unitário de uma distribuição definida no conjunto dos números reais. Levy (1939) propôs um método para introduzir uma variável *wrapped* em uma distribuição simétrica ou assimétrica, que é conhecido na literatura como *wrapped distribution* e desde então foram publicados vários trabalhos nesta área. Distribuições enroladas, ou *wrapped distributions*, constituem uma classe de modelos probabilísticos muito rica e útil para dados circulares. O termo *wrapped distribution* foi traduzido do inglês para "distribuição enrolada" em português pelo glossário de termos do ISI-*International Statistical Institute*. No entanto, neste trabalho será usado o termo distribuição *wrapped* para se referir a este tipo de distribuição circular.

Jammalamadaka e Kozubowski (2004) discutiram distribuições circulares obtidas por meio do método *wrapped* a partir das distribuições exponencial e Laplace. Coelho (2007) obteve uma expressão para a função densidade de probabilidade da distribuição *wrapped* gama, obtida a partir da distribuição gama tradicional, e mostrou como ela pode ser obtida como uma mistura de distribuições gama truncadas. Rao, Sarma e Girija (2007) derivaram novos modelos circulares a partir das já conhecidas distribuições de tempo de vida, como lognormal, logística, Weibull e a distribuição de valores extremos. Roy e Adnan (2012a) exploraram a distribuição *wrapped* Gompertz e discutiram sua aplicação em ornitologia. Em outro trabalho, Roy e Adnan (2012b) desenvolveram uma nova classe

de distribuições circulares chamada de *wrapped* exponencial ponderada. Jacob e Jayakumar (2013) propuseram uma nova família de distribuições circulares por meio da *wrapped* geométrica e estudaram suas propriedades. Rao, Girija e Devaraaj (2013) discutiram as características da distribuição *wrapped* gama. Adnan e Roy (2014) derivaram a distribuição *wrapped variance* gama e mostraram sua aplicabilidade a dados de direção do vento. Souza (2014) propôs uma classe de distribuições circulares chamada de *wrapped* elíptica, da qual fazem parte a *wrapped* cauchy, a *wrapped* Laplace, *wrapped* normal, *wrapped* t-Student, entre outras. Joshi e K (2017) propuseram uma versão *wrapped* da distribuição Lindley e derivaram expressões para sua função característica, momentos trigonométricos, coeficientes de assimetria e curtose.

Existem dificuldades em se obter modelos circulares com assimetria e momentos trigonométricos tratáveis, por isso, para tentar superar essas dificuldades nosso trabalho veio preencher essa lacuna existente na literatura, no qual é proposta neste trabalho uma nova distribuição *wrapped* obtida a partir da distribuição Birnbaum-Saunders reparametrizada proposta por Santos-Neto et al. (2012) em termos dos parâmetros da distribuição Birnbaum-Saunders clássica (BIRNBAUM; SAUNDERS, 1969).

Este capítulo está organizado da seguinte forma: na Seção 2.1 é apresentada a versão *wrapped* da distribuição Birnbaum-Saunders em sua forma reparametrizada e são mostradas as expressões para seus momentos trigonométricos. A Seção 2.2 é dedicada aos estimadores de máxima verossimilhança dos parâmetros desta distribuição. Na Seção 2.3 é apresentado um estudo de simulação para avaliar o desempenho destes estimadores e na Seção 2.4 é apresentada uma aplicação a dados reais.

2.1 Distribuição *Wrapped* Birnbaum Saunder

Seja Y uma variável aleatória (v.a.) real com função densidade probabilidade (fdp) f, função de distribuição acumulada (fda) F e função característica (fc) φ. Então, a v.a. *wrapped* correspondente é dada por $\theta \equiv Y \,(\text{mod } 2\pi)$ (em que "$\equiv$" significa que θ é congruente a Y) e possui fdp, fda e fc dadas, respectivamente, por: (MARDIA; JUPP, 2000; FISHER, 1995; JAMMALAMADAKA; SENGUPTA, 2001)

$$f_\theta(\theta) = \sum_{k=-\infty}^{+\infty} f(\theta + 2k\pi),\ \theta \in [0, 2\pi),$$
$$F_\theta(\theta) = \sum_{k=-\infty}^{+\infty} \left\{F(\theta + 2k\pi) - F(2k\pi)\right\},\ \theta \in [0, 2\pi),$$
$$\varphi_p = \varphi(p) = \mathbb{E}\left[\mathrm{e}^{ip\theta}\right] = \mathbb{E}[\cos(p\theta)] + i\mathbb{E}[\sin(p\theta)],\ p \in \mathbb{Z} \text{ e } i = \sqrt{-1}.$$

Birnbaum e Saunders (1969) introduziram a distribuição Birnbaum-Saunders indexada por dois parâmetros $(\alpha, \beta)^\top$, em que α é o parâmetro de forma e β é o parâmetro de escala. Santos-Neto et al. (2014) propuseram uma nova parametrização para a distribuição Birnbaum-Saunders indexada por outros dois parâmetros μ e δ, em que $0 < \mu = \beta\,(1 + \alpha^2/2)$ é o parâmetro de escala e a média da distribuição enquanto que $0 < \delta = 2/\alpha^2$ é o parâmetro de forma e precisão. Seja Y uma v.a. com fdp dada por

$$f_Y(y; \mu, \delta) = \frac{\mathrm{e}^{\delta/2}\sqrt{\delta+1}}{4\sqrt{\mu\pi}y^{3/2}}\left[y + \frac{\delta\mu}{\delta+1}\right]\exp\left\{-\frac{\delta}{4}\left[\frac{y(\delta+1)}{\delta\mu} + \frac{\delta\mu}{y(\delta+1)}\right]\right\},$$

em que $y > 0$, $\mu > 0$ e $\delta > 0$, então, Y segue uma distribuição Birnbaum-Saunders indexada por dois parâmetros μ e δ, cuja notação é $Y \sim \mathcal{BS}(\mu, \delta)$ com $\mathbb{E}[Y] = \mu$ e $\mathrm{Var}[Y] = \frac{\mu^2(2\delta+5)}{(\delta+1)^2}$. Detalhes teóricos podem ser encontrados em Santos-Neto et al. (2014).

Uma importante propriedade da distribuição Birnbaum-Saunders reparametrizada em μ e δ é a sua relação com a distribuição normal, isto é, se $Y \sim \mathcal{BS}(\mu, \delta)$ e $Z \sim \mathrm{N}(0, 1)$ então

$$Y = \frac{\delta\mu}{\delta+1}\left[\frac{Z}{\sqrt{2\delta}} + \sqrt{\left\{\frac{Z}{\sqrt{2\delta}}\right\}^2 + 1}\right]^2 \sim \mathcal{BS}(\mu, \delta)$$

e

$$Z = \sqrt{\frac{\delta}{2}}\left[\sqrt{\frac{(\delta+1)Y}{\delta\mu}} - \sqrt{\frac{\delta\mu}{(\delta+1)Y}}\right] \sim \mathrm{N}(0, 1). \tag{2.1}$$

Esta relação será utilizada para gerar amostras aleatórias da distribuição Birnbaum-Saunders .

A v.a. *wrapped* correspondente a Y é obtida como segue. Seja $Y \sim \mathcal{BS}(\mu, \delta)$, então, $\theta \equiv Y (\mathrm{mod}\ 2\pi)$ segue uma distribuição Wrapped Birnbaum-Saunders indexada por dois parâmetros μ e δ com fdp dada por

$$f_\theta(\theta) = \sum_{k=0}^{\infty} f_Y(\theta + 2k\pi) = \tag{2.2}$$
$$\frac{\mathrm{e}^{\frac{\delta}{2}}\sqrt{\delta+1}}{4\sqrt{\pi\mu}}\sum_{k=0}^{\infty}\frac{\theta + 2k\pi + \frac{\delta\mu}{\delta+1}}{(\theta+2k\pi)^{3/2}}\exp\left\{-\frac{\delta}{4}\left[\frac{(\theta+2k\pi)(\delta+1)}{\delta\mu} + \frac{\delta\mu}{(\theta+2k\pi)(\delta+1)}\right]\right\},$$

em que $\theta \in (0, 2\pi]$, $\mu \in (0, 2\pi]$ é a media circular da distribuição e $\delta > 0$ é o parâmetro de forma. A notação $\theta \sim \mathcal{WBS}(\mu, \delta)$ será usada.

Proposição 2.1.1. *A série*

$$\sum_{k=0}^{\infty} \frac{\theta + 2k\pi + \frac{\delta\mu}{\delta+1}}{(\theta + 2k\pi)^{3/2}} \exp\left\{-\frac{\delta}{4}\left[\frac{(\theta + 2k\pi)(\delta + 1)}{\delta\mu} + \frac{\delta\mu}{(\theta + 2k\pi)(\delta + 1)}\right]\right\}$$

(contida em 2.2) é uma série absolutamente convergente.

Demonstração. Pelo teste da razão de d'Alembert. O termo geral da série é

$$a_n = \frac{\theta + 2n\pi + \frac{\delta\mu}{\delta+1}}{(\theta + 2n\pi)^{3/2}} \exp\left\{-\frac{\delta}{4}\left[\frac{(\theta + 2n\pi)(\delta + 1)}{\delta\mu} + \frac{\delta\mu}{(\theta + 2n\pi)(\delta + 1)}\right]\right\}.$$

Tem-se que $a_n > 0$, pois, $\theta, \mu, \delta > 0$. Daí,

$$\frac{a_{n+1}}{a_n} = \frac{\frac{\theta+2(n+1)\pi+\frac{\delta\mu}{\delta+1}}{(\theta+2(n+1)\pi)^{3/2}} \exp\left\{-\frac{\delta}{4}\left[\frac{(\theta+2(n+1)\pi)(\delta+1)}{\delta\mu} + \frac{\delta\mu}{(\theta+2(n+1)\pi)(\delta+1)}\right]\right\}}{\frac{\theta+2n\pi+\frac{\delta\mu}{\delta+1}}{(\theta+2n\pi)^{3/2}} \exp\left\{-\frac{\delta}{4}\left[\frac{(\theta+2n\pi)(\delta+1)}{\delta\mu} + \frac{\delta\mu}{(\theta+2n\pi)(\delta+1)}\right]\right\}}$$

$$\frac{a_{n+1}}{a_n} = \left[1 + \frac{2\pi}{\theta + 2n\pi + \frac{\delta\mu}{\delta+1}}\right]\left[1 + \frac{2\pi}{\theta + 2n\pi}\right]^{-\frac{3}{2}}$$
$$\times \exp\left\{-\frac{\delta}{4}\left[\frac{2\pi(\delta + 1)}{\delta\mu} + \frac{2\pi\delta\mu}{(\theta + 2n\pi)(\theta + 2(n + 1)\pi)(\delta + 1)^2}\right]\right\}.$$

Agora, note que

$$\lim_{n\to\infty} \exp\left\{-\frac{\delta}{4}\left[\frac{2\pi(\delta+1)}{\delta\mu} + \frac{2\pi\delta\mu}{(\theta+2n\pi)(\theta+2(n+1)\pi)(\delta+1)^2}\right]\right\} = \exp\left\{-\frac{1}{2}\left[\frac{\pi(\delta+1)}{\mu}\right]\right\},$$
$$\lim_{n\to\infty}\left\{1 + \frac{2\pi}{\theta + 2n\pi + \frac{\delta\mu}{\delta+1}}\right\} = 1 \text{ e } \lim_{n\to\infty}\left\{1 + \frac{2\pi}{\theta + 2n\pi}\right\} = 1.$$

Então $\lim_{n\to\infty} \frac{a_{n+1}}{a_n} = \exp\left\{-\frac{1}{2}\left[\frac{\pi(\delta+1)}{\mu}\right]\right\} < 1$ e, desta forma, a série é absolutamente convergente. □

A proposição a seguir apresentará um resultado importante que será utilizado para obter-se os momentos trigonométricos.

Proposição 2.1.2. *Seja $Y \sim \mathcal{BS}(\mu, \delta)$ e seja θ sua v.a. wrapped correspondente, isto é, $\theta \sim \mathcal{WBS}(\mu, \delta)$. Então*

$$\mathbb{E}[\cos(p\theta)] \cong \cos(\mu)\left\{1 - \frac{\mu^2}{2}\left[\frac{p^2(2\delta + 5)}{(\delta + 1)^2} + (p - 1)^2\right]\right\} + \mu(p - 1)\sin(\mu) \quad e \tag{2.3}$$

$$\mathbb{E}[\sin(p\theta)] \cong \sin(\mu)\left\{1 - \frac{\mu^2}{2}\left[\frac{p^2(2\delta + 5)}{(\delta + 1)^2} + (p - 1)^2\right]\right\} + \mu(p - 1)\cos(\mu), \tag{2.4}$$

em que $p \in \mathbb{Z}$, $p \neq 0$ e " $\cong$ " significa "aproximadamente igual a". Essa aproximação é obtida fazendo a expansão em série de Taylor até a segunda ordem do seno e do cosseno de $p\theta$.

Demonstração. Sabe-se que $\theta \equiv Y (\text{mod } 2\pi)$ se, e somente se, $\exists \kappa \in \mathbb{Z}$ tal que $\theta = Y + 2\kappa\pi$. Então $\cos(p\theta) = \cos(pY + 2p\kappa\pi) = \cos(pY)\cos(2p\kappa\pi) - \sin(pY)\sin(2p\kappa\pi) = \cos(pY)$. Similarmente, $\sin(p\theta) = \sin(pY + 2p\kappa\pi) = \sin(pY)\cos(2p\kappa\pi) + \cos(pY)\sin(2p\kappa\pi) = \sin(pY)$. Então tem-se que $\cos(p\theta) = \cos(pY)$ e $\sin(p\theta) = \sin(pY)$, daí, segue que $\mathbb{E}[\cos(p\theta)] = \mathbb{E}[\cos(pY)]$ e $\mathbb{E}[\sin(p\theta)] = \mathbb{E}[\sin(pY)]$. Pela expansão de Taylor de segunda ordem de $\cos(pY)$ e de $\sin(pY)$ em torno de μ (média circular), obtém-se o seguinte:

$$\cos(pY) = \cos(\mu) + \sum_{r=1}^{\infty} \left\{ \cos^{(r)}(\mu) \frac{(pY-\mu)^r}{r!} \right\},$$

$$\cos(pY) = \cos(\mu) + \sum_{r=1}^{2} \left\{ \cos^{(r)}(\mu) \frac{(pY-\mu)^r}{r!} \right\} + o_p([pY]^2),$$

$$\sin(pY) = \sin(\mu) + \sum_{r=1}^{\infty} \left\{ \sin^{(r)}(\mu) \frac{(pY-\mu)^r}{r!} \right\} \text{ e}$$

$$\sin(pY) = \sin(\mu) + \sum_{r=1}^{2} \left\{ \sin^{(r)}(\mu) \frac{(pY-\mu)^r}{r!} \right\} + o_p([pY]^2)$$

e, ignorando os termos $o_p([pY]^2)$ e calculando o valor esperado, tem-se,

$$\begin{aligned}
\mathbb{E}[\cos(pY)] &\cong \cos(\mu) + \cos'(\mu)\mathbb{E}[pY-\mu] + \frac{1}{2}\cos''(\mu)\mathbb{E}(pY-\mu)^2 \\
&= \cos(\mu) - \sin(\mu)[p\mathbb{E}(Y) - \mu] - \frac{1}{2}\cos(\mu)\mathbb{E}[pY-\mu]^2 \\
&= \cos(\mu)\left[1 - \frac{\mathbb{E}[pY-\mu]^2}{2}\right] - \sin(\mu)[\mu(p-1)] \\
&= \cos(\mu)\left\{1 - \frac{\mu^2}{2}\left[\frac{p^2(2\delta+5)}{(\delta+1)^2} + (p-1)^2\right]\right\} - \mu(p-1)\sin(\mu) \\
\mathbb{E}[\sin(pY)] &\cong \sin(\mu) + \sin'(\mu)\mathbb{E}[pY-\mu] + \frac{1}{2}\sin''(\mu)\mathbb{E}(pY-\mu)^2 \\
&= \sin(\mu) + \cos(\mu)[p\mathbb{E}(Y) - \mu] - \frac{1}{2}\sin(\mu)\mathbb{E}[pY-\mu]^2 \\
&= \sin(\mu)\left[1 - \frac{\mathbb{E}[pY-\mu]^2}{2}\right] + \cos(\mu)[\mu(p-1)] \\
&= \sin(\mu)\left\{1 - \frac{\mu^2}{2}\left[\frac{p^2(2\delta+5)}{(\delta+1)^2} + (p-1)^2\right]\right\} + \mu(p-1)\cos(\mu),
\end{aligned}$$

em que $Y \sim \mathcal{BS}(\mu, \delta)$ e os símbolos $'$ e $''$ denotam a primeira e a segunda derivadas,

respectivamente. Portanto,

$$\mathbb{E}[\cos(p\theta)] = \mathbb{E}[\cos(pY)] \cong \cos(\mu)\left\{1 - \frac{\mu^2}{2}\left[\frac{p^2(2\delta+5)}{(\delta+1)^2} + (p-1)^2\right]\right\} - \mu(p-1)\sin(\mu),$$
$$\mathbb{E}[\sin(p\theta)] = \mathbb{E}[\sin(pY)] \cong \sin(\mu)\left\{1 - \frac{\mu^2}{2}\left[\frac{p^2(2\delta+5)}{(\delta+1)^2} + (p-1)^2\right]\right\} + \mu(p-1)\cos(\mu).$$

□

Agora será mostrada a expressão da fda de θ. Seja $\theta \sim \mathcal{WBS}(\mu, \delta)$, então a fda de θ é dada por

$$\begin{aligned} F_\theta(\theta) &= \sum_{k=0}^{\infty} [F(\theta + 2k\pi) - F(2k\pi)] \\ &= \sum_{k=0}^{\infty} \left\{ \Phi\left[\sqrt{\frac{\delta}{2}}\left(\sqrt{\frac{(\delta+1)(\theta+2k\pi)}{\delta\mu}} - \sqrt{\frac{\delta\mu}{(\delta+1)(\theta+2k\pi)}}\right)\right] \right. \\ &\left. -\Phi\left[\sqrt{\frac{\delta}{2}}\left(\sqrt{\frac{(\delta+1)2k\pi}{\delta\mu}} - \sqrt{\frac{\delta\mu}{(\delta+1)2k\pi}}\right)\right]\right\}, \end{aligned}$$

em que F é a fda de $Y \sim \mathcal{BS}(\mu, \delta)$ e Φ é a fda da distribuição normal padrão, considerando a relação dada na equação 2.1.

2.1.1 Momentos Trigonométricos

Na análise de dados circulares os dois principais parâmetros são a média circular μ e o comprimento resultante da média ρ, que também é chamado de concentração circular. Estes dois parâmetros são tão importantes para a análise de dados circulares quanto a média e a variância são para análise de dados reais e podem ser definidos a partir do primeiro momento trigonométrico, cuja definição será dada a seguir.

Momentos Trigonométricos Amostrais

Seja $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_n)^\top$ uma amostra aleatória (a.a.) de $\theta \sim \mathcal{WBS}(\mu, \delta)$, então os momentos amostrais do seno e do cosseno de θ são

$$\bar{C} = \frac{1}{n}\sum_{j=1}^{n}\cos(\theta_j) \quad \text{e} \quad \bar{S} = \frac{1}{n}\sum_{j=1}^{n}\sin(\theta_j).$$

Assim, o primeiro momento trigonométrico amostral é dado por:

$$m_1' = \bar{C} + i\bar{S} = \bar{R}\mathrm{e}^{i\bar{\theta}},$$

em que $\bar{R} = (\bar{C}^2 + \bar{S}^2)^{\frac{1}{2}}$ é o comprimento resultante da direção média amostral (parâmetro de concentração circular amostral) e $\bar{\theta}$ é a direção média amostral, que é dada por $\bar{\theta} = \arctan^\star\left(\bar{S}/\bar{C}\right)$ e

$$\arctan^\star\left(\bar{S}/\bar{C}\right) = \begin{cases} \arctan\left(\bar{S}/\bar{C}\right) & \text{se } \bar{C} > 0 \text{ e } \bar{S} > 0, \\ \frac{\pi}{2} & \text{se } \bar{C} = 0 \text{ e } \bar{S} > 0, \\ \arctan\left(\bar{S}/\bar{C}\right) + 2\pi & \text{se } \bar{C} \geq 0 \text{ e } \bar{S} < 0, \\ \arctan\left(\bar{S}/\bar{C}\right) + \pi & \text{se } \bar{C} < 0. \end{cases}$$

Estendendo esta noção, define-se o p-ésimo momento trigonométrico amostral como

$$\begin{aligned} m'_p &= a_p + ib_p, \text{ em que} \\ a_p = \frac{1}{n}\sum_{i=1}^{n}\cos(p\theta_i) \;\; &\text{e} \;\; b_p = \frac{1}{n}\sum_{i=1}^{n}\sin(p\theta_i). \end{aligned} \tag{2.5}$$

Então,

$$m'_p = a_p + ib_p = \bar{R}_p \mathrm{e}^{i\bar{\theta}_p},$$

em que $\bar{\theta}_p$ e $\bar{R}_p$ denotam a direção média e o comprimento resultante da média amostral de $p\theta_1, \cdots, p\theta_n$.

Momentos Trigonométricos Populacionais

Seja $Y \sim \mathcal{BS}(\mu, \delta)$, então sua fc é $\varphi : \mathbb{R} \to \mathbb{C}$, dada por (SANTOS-NETO et al., 2014)

$$\varphi(t) = \mathbb{E}(\mathrm{e}^{itY}) = \frac{1}{2}\left\{\left[1 + \frac{\sqrt{\delta+1}}{\sqrt{1+\delta-4ti\mu}}\right]\exp\left(\frac{\delta\left[\sqrt{\delta+1} - \sqrt{1+\delta-4ti\mu}\right]}{2\sqrt{\delta+1}}\right)\right\},$$

em que $t \in \mathbb{R}$, $i = \sqrt{-1}$, $\mathbb{E}[\cos(p\theta)]$ é dado em (2.3) e $\mathbb{E}[\sin(p\theta)]$ é dado em (2.4). Agora seja $\theta \sim \mathcal{WBS}(\mu, \delta)$, então sua fc é $\varphi(p)$ com $p \in \mathbb{Z}$, em que φ é a fc de $Y \sim \mathcal{BS}(\mu, \delta)$, ou seja, $\varphi(p)$ é a sequência $\{\varphi(0), \varphi(\pm1), \varphi(\pm2), \ldots\}$ ou simplesmente $\{\varphi_0, \varphi_{\pm1}, \varphi_{\pm2}, \ldots\}$. O valor da fc de θ avaliada no número inteiro p é chamado de $p-$ésimo momento trigonométrico de θ.

Para $p = 0$ tem-se $\varphi_0 = \mathbb{E}(\mathrm{e}^0) = 1$, para $p = 1$ tem-se o primeiro momento trigonométrico de θ que é dado por

$$\varphi_1 = \frac{1}{2}\left\{\left[1 + \frac{\sqrt{\delta+1}}{\sqrt{1+\delta-4i\mu}}\right]\exp\left(\frac{\delta\left[\sqrt{\delta+1} - \sqrt{1+\delta-4i\mu}\right]}{2\sqrt{\delta+1}}\right)\right\}$$

e para $p = 2$ tem-se o segundo momento trigonométrico de θ que é dado por

$$\varphi_2 = \frac{1}{2}\left\{\left[1 + \frac{\sqrt{\delta+1}}{\sqrt{1+\delta-8i\mu}}\right]\exp\left(\frac{\delta\left[\sqrt{\delta+1} - \sqrt{1+\delta-8i\mu}\right]}{2\sqrt{\delta+1}}\right)\right\}.$$

Por outro lado, os momentos trigonométricos populacionais também são definidos de forma similar ao caso amostral dado em (2.5), assim tem-se que (MARDIA; JUPP, 2000, págs. 26–27)

$$\begin{aligned}\varphi_p &= \mathbb{E}(\mathrm{e}^{ip\theta}) = \mathbb{E}[\cos(p\theta)] + i\mathbb{E}[\sin(p\theta)] \\ &\cong \cos(\mu)\left\{1 - \frac{\mu^2}{2}\left[\frac{p^2(2\delta+5)}{(\delta+1)^2} + (p-1)^2\right]\right\} - \mu(p-1)\sin(\mu) \\ &+ i\left\{\sin(\mu)\left\{1 - \frac{\mu^2}{2}\left[\frac{p^2(2\delta+5)}{(\delta+1)^2} + (p-1)^2\right]\right\} + \mu(p-1)\cos(\mu)\right\},\end{aligned}$$

em que a aproximação é obtida através dos resultados da Proposição 2.1.2. Assim, para $p = 1$ tem-se a seguinte aproximação para o primeiro momento trigonométrico

$$\varphi_1 \cong [\cos(\mu) + i\sin(\mu)]\left\{1 - \frac{\mu^2}{2}\left[\frac{(2\delta+5)}{(\delta+1)^2}\right]\right\}.$$

Uma vez definidos os momentos trigonométricos, é possível obter uma aproximação para ρ, que é dado por $\rho = \sqrt{\mathbb{E}^2[\cos(\theta)] + \mathbb{E}^2[\sin(\theta)]}$. Assim, tem-se que o parâmetro de concentração circular é aproximado por

$$\rho \cong 1 - \frac{\mu^2(2\delta+5)}{2(\delta+1)^2}. \tag{2.6}$$

Uma vez que $0 < \rho < 1$, segue que $0 < 1 - \frac{\mu^2(2\delta+5)}{2(\delta+1)^2} < 1$ e, portanto, para que seja possível aproximar ρ pela expressão dada em (2.6) a seguinte desigualdade deve ser satisfeita:

$$0 < \mu < \frac{\sqrt{2}(\delta+1)}{\sqrt{2\delta+5}}. \tag{2.7}$$

2.2 Estimação de Máxima Verossimilhança

Seja $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)^\top$ uma amostra aleatória de $T \sim \mathcal{WBS}(\mu, \delta)$. Então, o logaritmo da função de verossimilhança (função log-verossimilhança) é dado por

$$\begin{aligned}&\ell(\mu, \delta; \boldsymbol{\theta}) = \frac{n}{2}(\delta + \log(\delta+1) - \log(16\pi\mu)) + \\ &\sum_{j=1}^{n}\log\left\{\sum_{k=0}^{\infty}\frac{\left(\frac{\delta\mu}{\delta+1} + \theta_j + 2k\pi\right)\exp\left\{-\frac{\delta}{4}\left(\frac{\delta\mu}{(\delta+1)(\theta_j+2k\pi)} + \frac{(\delta+1)(\theta_j+2k\pi)}{\delta\mu}\right)\right\}}{(\theta_j + 2k\pi)^{\,3/2}}\right\}.\end{aligned}$$

A função escore é dada por $\mathbf{U}(\boldsymbol{\beta}) = (\mathbf{U}(\mu), \mathbf{U}(\delta))^\top$ em que $\boldsymbol{\beta} = (\mu, \delta)^\top$, enquanto que

$$\mathbf{U}(\mu) = \frac{\partial \ell(\mu, \delta; \boldsymbol{\theta})}{\partial \mu} = -\frac{n}{2\mu} + \sum_{j=1}^{n} \left\{ \left[\sum_{k=0}^{\infty} \frac{\delta \exp\left(-\frac{\delta}{4}\left(\frac{\delta\mu}{(\delta+1)(\theta_j+2k\pi)} + \frac{(\delta+1)(\theta_j+2k\pi)}{\delta\mu}\right)\right)}{(\delta+1)\left(\theta_j+2k\pi\right)^{3/2}} \right.\right.$$
$$\left.\left. - \frac{\delta\left(\frac{\delta\mu}{\delta+1} + \theta_j + 2k\pi\right)\left(\frac{\delta}{(\delta+1)(\theta_j+2k\pi)} - \frac{(\delta+1)(\theta_j+2k\pi)}{\delta\mu^2}\right) \mathrm{e}^{-\frac{\delta}{4}\left(\frac{\delta\mu}{(\delta+1)\left(\theta_j+2k\pi\right)} + \frac{(\delta+1)\left(\theta_j+2k\pi\right)}{\delta\mu}\right)}}{4\left(\theta_j+2k\pi\right)^{3/2}} \right] \right.$$
$$\left. \times \left[\sum_{k=0}^{\infty} \frac{\left(\frac{\delta\mu}{\delta+1} + \theta_j + 2k\pi\right) \exp\left(-\frac{1}{4}\delta\left(\frac{\delta\mu}{(\delta+1)(\theta_j+2k\pi)} + \frac{(\delta+1)(\theta_j+2k\pi)}{\delta\mu}\right)\right)}{\left(\theta_j+2k\pi\right)^{3/2}} \right]^{-1} \right\}$$

e

$$\mathbf{U}(\delta) = \frac{\partial \ell(\mu, \delta; \boldsymbol{\theta})}{\partial \delta} = \frac{n}{2}\left(\frac{1}{\delta+1} + 1\right) + \sum_{j=1}^{n} \left\{ \sum_{k=0}^{\infty} \left\{ \left(\frac{\delta\mu}{\delta+1} + \theta_j + 2k\pi\right) \right.\right.$$
$$\left(\frac{1}{4}\left(-\frac{\delta\mu}{(\delta+1)\left(\theta_j+2k\pi\right)} - \frac{(\delta+1)\left(\theta_j+2k\pi\right)}{\delta\mu}\right) - \right.$$
$$\left.\frac{\delta}{4}\left(-\frac{(\delta+1)\left(\theta_j+2k\pi\right)}{\delta^2\mu} + \frac{\mu}{(\delta+1)\left(\theta_j+2k\pi\right)} - \frac{\delta\mu}{(\delta+1)^2\left(\theta_j+2k\pi\right)} + \frac{\theta_j+2k\pi}{\delta\mu}\right)\right)$$
$$\exp\left(-\frac{\delta}{4}\left(\frac{\delta\mu}{(\delta+1)\left(\theta_j+2k\pi\right)} + \frac{(\delta+1)\left(\theta_j+2k\pi\right)}{\delta\mu}\right)\right)\left(\theta_j+2k\pi\right)^{-3/2} +$$
$$\left.\left.\frac{\left(\frac{\mu}{\delta+1} - \frac{\delta\mu}{(\delta+1)^2}\right)\exp\left(-\frac{1}{4}\delta\left(\frac{\delta\mu}{(\delta+1)(\theta_j+2k\pi)} + \frac{(\delta+1)(\theta_j+2k\pi)}{\delta\mu}\right)\right)}{\left(\theta_j+2k\pi\right)^{3/2}} \right\}\right\} \times$$
$$\left(\sum_{k=0}^{\infty} \frac{\left(\frac{\delta\mu}{\delta+1} + \theta_j + 2k\pi\right)\exp\left(-\frac{\delta}{4}\left(\frac{\delta\mu}{(\delta+1)(\theta_j+2k\pi)} + \frac{(\delta+1)(\theta_j+2k\pi)}{\delta\mu}\right)\right)}{\left(\theta_j+2k\pi\right)^{3/2}}\right)^{-1}.$$

O estimador de máxima verossimilhança (EMV) não possui expressão analítica em forma explícita para μ e δ. Desta forma, faz-se necessário o uso de métodos numéricos de otimização, mais especificamente, utilizou-se o método quasi-Newton BFGS (NOCEDAL; WRIGHT, 1999, Cap. 8) para obter o EMV que, segundo Mittelhammer, Judge e Miller (2000), é o mais confiável dentre os métodos de otimização não-linear.

Seja $\boldsymbol{\beta} = (\mu, \delta)^\top$ o vetor de parâmetros e seja $\widehat{\boldsymbol{\beta}} = (\hat{\mu}, \hat{\delta})^\top$ o vetor dos EMV. Sob certas condições de regularidade apropriadas o EMV é consistente e assintoticamente normal com matriz de covariâncias igual à inversa da matriz de informação esperada de Fisher. Então, segue que

$$\widehat{\boldsymbol{\beta}} \overset{A}{\sim} \mathrm{N}_2(\boldsymbol{\beta}, \mathcal{I}^{-1}(\boldsymbol{\beta})), \tag{2.8}$$

quando n é suficientemente grande, $\overset{A}{\sim}$ denota aproximadamente distribuído, em que $\mathcal{I}(\boldsymbol{\beta})$ é a matriz de informação esperada de Fisher e $\mathcal{I}^{-1}(\boldsymbol{\beta})$ é a sua inversa. Obter a matriz

de informação esperada de Fisher, neste caso, é de grande complexidade analítica. Deste modo, será usada a matriz de informação observada de Fisher $\mathbf{I}_n(\boldsymbol{\beta})$, enquanto que a raiz quadrada dos elementos da diagonal da inversa desta matriz, $\mathbf{I}_n^{-1}(\boldsymbol{\beta})$, podem ser usados para aproximar os correspondentes erros padrão (veja Efron e Hinkley (1978) para detalhes a respeito do uso da matriz de informação observada de Fisher), em que

$$\mathbf{I}_n(\boldsymbol{\beta}) = \begin{bmatrix} \mathrm{I}(\boldsymbol{\beta})_{\mu\mu} & \mathrm{I}(\boldsymbol{\beta})_{\mu\delta} \\ \mathrm{I}(\boldsymbol{\beta})_{\delta\mu} & \mathrm{I}(\boldsymbol{\beta})_{\delta\delta} \end{bmatrix}, \quad \mathbf{I}_n^{-1}(\boldsymbol{\beta}) = \begin{bmatrix} \mathrm{I}(\boldsymbol{\beta})^{\mu\mu} & \mathrm{I}(\boldsymbol{\beta})^{\mu\delta} \\ \mathrm{I}(\boldsymbol{\beta})^{\delta\mu} & \mathrm{I}(\boldsymbol{\beta})^{\delta\delta} \end{bmatrix},$$

$\mathrm{I}(\boldsymbol{\beta})_{\mu\mu} = -\frac{\partial^2 \ell(\mu,\delta;\boldsymbol{\theta})}{\partial \mu^2}$, $\mathrm{I}(\boldsymbol{\beta})_{\mu\delta} = \mathrm{I}(\boldsymbol{\beta})_{\delta\mu} = -\frac{\partial^2 \ell(\mu,\delta;\boldsymbol{\theta})}{\partial \mu \partial \delta}$ e $\mathrm{I}(\boldsymbol{\beta})_{\delta\delta} = -\frac{\partial^2 \ell(\mu,\delta;\boldsymbol{\theta})}{\partial \delta^2}$. No Apêndice A são mostradas as expressões para estas segundas derivadas parciais.

Neste trabalho, $\mathbf{I}_n^{-1}(\widehat{\boldsymbol{\beta}})$ é a inversa de matriz de informação observada de Fisher avaliada em $\widehat{\boldsymbol{\beta}}$ e $\mathrm{I}(\widehat{\boldsymbol{\beta}})^{\mu\mu}$ e $\mathrm{I}(\widehat{\boldsymbol{\beta}})^{\delta\delta}$ são os elementos $(1,1)$ e $(2,2)$ da sua diagonal principal, respectivamente. Desta forma, a partir de (2.8), é possível construir intervalos de confiança assintóticos (ICA) para as componentes de $\boldsymbol{\beta} = (\mu, \delta)^\top$, que são dados por $\left(\hat{\mu} \pm z_{1-\alpha/2}\sqrt{\mathrm{I}(\widehat{\boldsymbol{\beta}})^{\mu\mu}}\right)$ e $\left(\hat{\delta} \pm z_{1-\alpha/2}\sqrt{\mathrm{I}(\widehat{\boldsymbol{\beta}})^{\delta\delta}}\right)$, em que $z_{1-\alpha/2}$ é o quantil $(1-\alpha/2)$ da distribuição normal padrão.

Para o parâmetro de concentração circular ρ tem-se que o ICA com nível de confiança $1-\alpha/2$ é dado por $\left(\hat{\rho} \pm z_{1-\alpha/2}\sqrt{\mathrm{Var}[d(\widehat{\boldsymbol{\beta}})]}\right)$, $d(\widehat{\boldsymbol{\beta}}) = \hat{\rho} = 1 - \dfrac{\hat{\mu}^2(2\hat{\delta}+5)}{2(\hat{\delta}+1)^2}$ e $\mathrm{Var}(\widehat{\boldsymbol{\beta}}) = \mathbf{I}_n^{-1}(\widehat{\boldsymbol{\beta}})$, tal que $\mathrm{Var}[d(\widehat{\boldsymbol{\beta}})] = D^\top(\widehat{\boldsymbol{\beta}})\mathrm{Var}(\widehat{\boldsymbol{\beta}})D(\widehat{\boldsymbol{\beta}})$, em que $D(\widehat{\boldsymbol{\beta}}) = \frac{\partial d(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\Big|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} = \left(\frac{-\hat{\mu}(2\hat{\delta}+5)}{(\hat{\delta}+1)^2} \quad \frac{\hat{\mu}^2(\hat{\delta}+4)}{(\hat{\delta}+1)^3}\right)^\top$.

2.3 Simulação

Foi conduzido um estudo de simulação de Monte Carlo para avaliar a performance do EMV de $\boldsymbol{\beta} = (\beta_1, \beta_2)^\top = (\mu, \delta)^\top$. O número de réplicas de Monte Carlo foi fixado em $m = 5000$ e foram geradas amostras de tamanho 15, 30, 50, 100, 500 e 1000. Para cada amostra gerada foram obtidas as estimativas de máxima verossimilhança (EMV) de μ, δ e ρ e foi calculado um ICA para μ e ρ. Note que o valor real de δ foi de 42 para satisfazer a desigualdade (2.7) em todos os casos e as estimativas obtidas para este parâmetro foram omitidas uma vez que o interesse está na estimação da concentração circular (ρ).

Diz-se que a concentração dos dados direcionais em torno da direção média μ é acentuada, média ou moderada, se ρ estiver próximo de 1, de $0,5$ ou de zero, respectivamente.

Assim, criamos cenários de simulação cuja concentração é acentuada (cenários 1 e 2), média (cenário 3) e moderada (cenário 4) e observamos que o valor da concentração afeta negativamente a qualidade das estimativas, como veremos mais adiante.

Para avaliar a performance dos estimadores foram calculados a média das estimativas obtidas em cada réplica de Monte Carlo $\overline{\hat{\beta}}_{MC}$ e $\widehat{EP}_j$ é a estimativa do erro-padrão correspondente, o erro quadrático médio (EQM), o viés relativo (VR), a assimetria, a curtose, a proporção de vezes em que o ICA conteve o verdadeiro valor dos parâmetros, que é a taxa de cobertura (TX), além da taxa à esquerda (TX ESQ), que é a estimativa da probabilidade do verdadeiro valor do parâmetro ser menor do que o limite inferior do ICA, e da taxa à direita (TX DIR), que é a estimativa da probabilidade do verdadeiro valor do parâmetro ser maior do que o limite superior do ICA. Considerando o nível nominal adotado para o ICA de 95%, a TX ESQ e TX DIR são ambas iguais a $2,5\%$. Então, tem-se o seguinte:

$$\overline{\hat{\beta}}_{MC} = \frac{1}{m}\sum_{j=1}^{m}\hat{\beta}_j, \quad \text{EP} = \frac{1}{m}\sum_{j=1}^{m}\widehat{EP}_j,$$
$$\text{VR} = \frac{1}{m}\sum_{j=1}^{m}\left[\frac{\hat{\beta}_j - \beta}{\beta}\right] \quad \text{e} \quad \text{EQM} = \frac{1}{m}\sum_{j=1}^{m}\left(\hat{\beta}_j - \beta\right)^2,$$

em que $\hat{\beta}_j$ é a estimativa de β obtida a partir da amostra gerada na réplica j do processo de simulação de MC. Os valores iniciais do processo iterativo para obtenção do EMV foram obtidos através das estatísticas amostrais $\bar{\text{R}}$ e $\bar{\theta}$. As simulações foram feitas usando a linguagem de programação matricial Ox, que é distribuída gratuitamente para fins acadêmicos e está disponível em http://www.doornik.com. Para gerar números aleatórios da distribuição Wrapped Birnbaum-Saunders utilizado o gerador descrito no Algoritmo 1, que é baseado no gerador de números aleatórios da distribuição Birnbaum-Saunders dado no Algoritmo 1 dado em (SANTOS-NETO et al., 2014).

Algoritmo 1: Gerador de números aleatórios Wrapped Birnbaum-Saunders

1. Gere um número aleatório z da v.a. $Z \sim \text{N}(0,1)$;
2. Estabeleça valores para μ e δ de $Y \sim \mathcal{BS}(\mu, \delta)$;
3. Calcule o número aleatório y de $Y \sim \mathcal{BS}(\mu, \delta)$ usando a fórmula dada em (2.1);
4. Calcule o número aleatório θ de $\theta \sim \mathcal{WBS}(\mu, \delta)$ usando a fórmula $\theta \equiv Y (\text{mod } 2\pi)$;
5. Repita os passos de 1 a 4 até obter o montante desejado de números aleatórios.

Souza (2014), no contexto dos modelos *wrapped* elípticos, fez um estudo de simulação para determinar o número de termos no somatório que indicou uma boa aproximação para a função de log-verossimilhança e concluiu que as estimativas dos parâmetros já são aproximadamente não viesadas para grandes amostras com $k \geq 5$ fixo. No nosso trabalho, avaliamos o incremento na função de log-verossimilhança com a soma de um termo adicional e concluímos que k=8 garantiu, nas nossas simulações, uma aproximação para a função de log-verossimilhança com erros menores ou iguais a 10^{-5}.

As Tabelas 1–4 mostram que o EQM do EMV decresce à medida em que o tamanho da amostra cresce, como esperado. Para valores mais baixos de ρ e n (Tabela 4) o VR de $\hat{\rho}$ apresentou valores mais altos que nos demais casos, entretanto, quando n aumenta esse viés diminui drasticamente, indicando que $\hat{\rho}$ pode ser não-viesado assintoticamente. Os valores próximos de zero do viés de $\hat{\mu}$ é um indício de que este é um estimador não-viesado para μ. Os valores de assimetria e curtose estão próximos dos valores de assimetria e curtose da distribuição normal, como esperado. O nível de cobertura do ICA também aumenta com o tamanho da amostra se aproximando do nível de confiança que é de 95%.

Ainda considerando as Tabelas 1–4, é mostrado que o EQM de $\hat{\mu}$ e de $\hat{\rho}$ é maior quando ρ diminui, ou seja, quando os dados estão menos concentrados em torno da média circular o desempenho das estimativas é inferior ao caso em que os dados estão mais concentrados. A TX se aproxima do nível nominal de 95% à medida em que o tamanho de amostra aumenta e, para valores menores do tamanho de amostra n, a TX DIR se apresentou maior do que a TX ESQ no caso de $\hat{\mu}$, e ao contrário, a TX ESQ se apresentou maior do que a TX DIR no caso de $\hat{\rho}$.

De modo geral, o desempenho do EMV pode ser considerado satisfatório levando-se em conta tamanhos amostrais moderados a partir de $n = 50$, sendo que, foi observado nas simulações que a qualidade das estimativas é afetada negativamente com a diminuição do ρ, independentemente do valor de μ.

Tabela 1 – Desempenho dos EMV dos parâmetros do modelo Wrapped Birnbaum-Saunders quando $\mu = \frac{\pi}{4} = 0,78539$, $\rho = 0,98515$ e diversos tamanho de amostra

n	$\hat{\mu}$	EP	VR	EQM	ASSIM	CURT	TX ESQ	TX	TX DIR
15	0,78696	0,04222	0,00199	0,00195	0,11938	3,06505	2,46%	92,36%	5,18%
30	0,78556	0,03066	0,00021	0,00100	0,11467	3,06098	2,24%	93,24%	4,52%
50	0,78503	0,02402	−0,00047	0,00059	0,01753	2,95497	2,00%	93,86%	4,14%
100	0,78484	0,01712	−0,00071	0,00030	0,10419	2,97448	2,18%	94,06%	3,76%
500	0,78536	0,00770	−0,00005	0,00006	0,01580	3,01447	2,26%	95,00%	2,74%
1000	0,78521	0,00544	−0,00025	0,00003	0,02379	3,03059	2,04%	95,22%	2,74%
n	$\hat{\rho}$	EP	VR	EQM	ASSIM	CURT	TX ESQ	TX	TX DIR
15	0,98607	0,00562	0,00093	0,00003	−0,85704	3,93686	15,72%	84,28%	0,00%
30	0,98561	0,00409	0,00047	0,00002	−0,66154	3,82152	10,76%	89,22%	0,02%
50	0,98539	0,00322	0,00024	0,00001	−0,44349	3,19008	8,60%	91,34%	0,06%
100	0,98526	0,00230	0,00010	0,00001	−0,34464	3,25081	6,76%	92,86%	0,38%
500	0,98516	0,00103	0,00001	0,00000	−0,11967	3,03301	4,20%	94,64%	1,16%
1000	0,98517	0,00073	0,00001	0,00000	−0,12227	3,02988	3,82%	94,34%	1,84%

Tabela 2 – Desempenho dos EMV dos parâmetros do modelo Wrapped Birnbaum-Saunders quando $\mu = \frac{3\pi}{4} = 2,35619$, $\rho = 0,86639$ e diversos tamanho de amostra

n	$\hat{\mu}$	EP	VR	EQM	ASSIM	CURT	TX ESQ	TX	TX DIR
15	2,36088	0,12666	0,00199	0,01751	0,11958	3,06523	2,46%	92,36%	5,18%
30	2,35667	0,09197	0,00020	0,00903	0,11470	3,06174	2,24%	93,24%	4,52%
50	2,35508	0,07206	−0,00047	0,00529	0,01759	2,95510	2,00%	93,86%	4,14%
100	2,35453	0,05135	−0,00071	0,00271	0,10453	2,97449	2,18%	94,06%	3,76%
500	2,35607	0,02310	−0,00005	0,00053	0,01584	3,01430	2,26%	95,00%	2,74%
1000	2,35564	0,01634	−0,00024	0,00027	0,02374	3,03020	2,04%	95,26%	2,70%
n	$\hat{\rho}$	EP	VR	EQM	ASSIM	CURT	TX ESQ	TX	TX DIR
15	0,87465	0,05057	0,00954	0,00269	−0,86095	3,95798	15,70%	84,30%	0,00%
30	0,87056	0,03685	0,00481	0,00137	−0,66506	3,83242	10,80%	89,18%	0,02%
50	0,86855	0,02897	0,00249	0,00086	−0,44456	3,19146	8,62%	91,34%	0,04%
100	0,86730	0,02066	0,00106	0,00045	−0,34367	3,25164	6,84%	92,78%	0,38%
500	0,86644	0,00929	0,00006	0,00009	−0,12193	3,03280	4,22%	94,62%	1,16%
1000	0,86646	0,00657	0,00008	0,00004	−0,12327	3,02455	3,68%	94,44%	1,88%

Tabela 3 – Desempenho dos EMV dos parâmetros do modelo Wrapped Birnbaum-Saunders quando $\mu = \frac{5\pi}{4} = 3,92699$, $\rho = 0,62886$ e diversos tamanho de amostra

n	$\hat{\mu}$	EP	VR	EQM	ASSIM	CURT	TX ESQ	TX	TX DIR
15	3,93559	0,21472	0,00219	0,05004	0,16307	3,16908	2,46%	92,32%	5,22%
30	3,92908	0,15538	0,00053	0,02544	0,15448	3,01983	2,28%	93,52%	4,20%
50	3,92585	0,12154	−0,00029	0,01515	0,05786	2,96598	1,68%	94,08%	4,24%
100	3,92460	0,08645	−0,00061	0,00786	0,05046	3,03574	1,98%	93,98%	4,04%
500	3,92675	0,03884	−0,00006	0,00146	0,04290	3,03116	2,18%	95,22%	2,60%
1000	3,92608	0,02746	−0,00023	0,00075	0,03895	2,96594	2,14%	95,06%	2,80%
n	$\hat{\rho}$	EP	VR	EQM	ASSIM	CURT	TX ESQ	TX	TX DIR
15	0,65069	0,14408	0,03472	0,02132	−0,88997	3,95850	15,78%	84,22%	0,00%
30	0,63922	0,10419	0,01649	0,01103	−0,72874	4,07690	10,98%	89,02%	0,00%
50	0,63411	0,08155	0,00836	0,00693	−0,51083	3,31647	9,12%	90,86%	0,02%
100	0,63126	0,05792	0,00382	0,00347	−0,35889	3,22666	6,54%	93,16%	0,30%
500	0,62894	0,02600	0,00013	0,00065	−0,13540	2,98917	3,70%	94,98%	1,32%
1000	0,62904	0,01837	0,00029	0,00035	−0,15737	3,04016	3,52%	94,54%	1,94%

Tabela 4 – Desempenho dos EMV dos parâmetros do modelo Wrapped Birnbaum-Saunders quando $\mu = \frac{7\pi}{4} = 5,49779$, $\rho = 0,27256$ e diversos tamanho de amostra

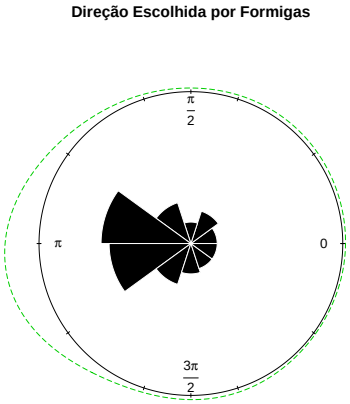
n	$\hat{\mu}$	EP	VR	EQM	ASSIM	CURT	TX ESQ	TX	TX DIR
15	5,45976	0,29920	−0,00692	0,10194	0,02573	2,92279	1,98%	90,98%	7,04%
30	5,47578	0,22819	−0,00400	0,05516	0,17023	3,10061	1,84%	92,66%	5,50%
50	5,48112	0,18324	−0,00303	0,03277	0,05491	2,91487	1,38%	93,92%	4,70%
100	5,49235	0,13270	−0,00099	0,01781	0,14127	3,10126	1,64%	94,28%	4,08%
500	5,49850	0,06004	0,00013	0,00359	0,02243	2,94236	2,12%	94,68%	3,20%
1000	5,49685	0,04239	−0,00017	0,00181	0,07571	3,06765	2,26%	94,74%	3,00%
n	$\hat{\rho}$	EP	VR	EQM	ASSIM	CURT	TX ESQ	TX	TX DIR
15	0,40648	0,25350	0,49136	0,05420	−0,08455	2,29095	18,60%	81,40%	0,00%
30	0,34266	0,20103	0,25720	0,03056	−0,05732	2,33322	12,68%	87,32%	0,00%
50	0,30884	0,16507	0,13311	0,02124	−0,05273	2,40797	10,06%	89,94%	0,00%
100	0,28474	0,12113	0,04468	0,01272	−0,15604	2,60861	7,32%	92,68%	0,00%
500	0,27189	0,05494	−0,00244	0,00297	−0,25161	3,20920	3,84%	95,14%	1,02%
1000	0,27268	0,03874	0,00045	0,00151	−0,15283	2,99835	3,76%	94,86%	1,38%

2.4 Aplicação

O experimento de Jander (1957) sobre a direção escolhida por formigas em resposta a um estímulo já motivou estudos interessantes em modelagem de dados circulares, como, por exemplo, em Umbach e Jammalamadaka (2009). Neste experimento, as formigas são submetidas a um estímulo olfativo e a direção do movimento escolhido por cada uma delas é medido de forma apropriada. As formigas tendem a caminhar em direção ao alvo, entretanto, para realizar alguma inferência de forma assertiva, é interessante verificar se os dados podem ser modelados como uma amostra aleatória de alguma distribuição circular, em particular, o interesse é verificar se os dados utilizados por Fisher (1995, Exemplo 4.4, pág. 60), que contém observações da direção escolhida por 100 formigas, podem ser modelados como uma amostra aleatória da distribuição Wrapped Birnbaum-Saunders .

Foram calculados a média e a concentração circular amostrais baseados no primeiro momento trigonométrico amostral, cujos valores obtidos foram $\bar{\boldsymbol{\theta}} = 3,1962$ e $\bar{R} = 0,6101$, respectivamente. A Figura 1 mostra a distribuição dos dados por meio de um diagrama de rosas, que é um gráfico para dados circulares similar ao histograma para dados reais. Na Figura 1:(a) a linha pontilhada representa a estimativa não paramétrica da fdp dos dados obtida pela método kernel para dados circulares com kernel *wrapped normal* e parâmetro de suavização igual a $\hat{\rho}$; já na Figura 1:(b) o parâmetro de suavização considerado foi $\bar{R}$. Com estes gráficos é possível verificar que os dados apresentam indícios de assimetria.

(a)



(b)

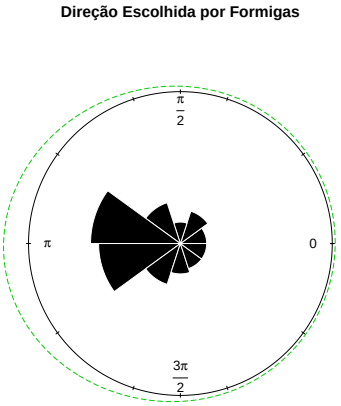


Figura 1 – Diagrama de rosas e estimativa não-paramétrica da distribuição dos dados, em que: (a) $\hat{\rho}$ é utilizado como parâmetro de suavização e (b) $\bar{R}$ é utilizado como parâmetro de suavização

Além da análise visual do gráfico, foi calculado o coeficiente de assimetria, que é dado por $s = \beta_2/(1-\rho)^{1.5}$, em que β_2 é o valor esperado de $\sin[2(\theta-\mu)]$. O valor de s foi obtido de duas formas: (i) considerando $\hat{\mu}$ e $\hat{\rho}$ obtidos pelo método da MV no modelo Wrapped Birnbaum-Saunders , tem-se que $\hat{s}_1 = \hat{\beta}_2/(1-\hat{\rho})^{1.5}$, em que $\hat{\beta}_2 = \frac{1}{n}\sum_{i=1}^{n}\sin[2(\theta_i - \hat{\mu})]$; (ii) considerando os valores amostrais $\bar{\theta}$ e $\bar{R}$, tem-se que $\hat{s}_2 = \hat{\beta}_2/(1-\bar{R})^{1.5}$, em que $\hat{\beta}_2 = \frac{1}{n}\sum_{i=1}^{n}\sin[2(\theta_i - \bar{\theta})]$. Os valores da assimetria circular obtidos foram, respectivamente, $\hat{s}_1 = -0,3654$ e $\hat{s}_2 = -0,3337$, sendo que ambos indicam assimetria. A Figura 1 também mostra que não há evidências de bimodalidade.

Umbach e Jammalamadaka (2009) desenvolveram uma versão assimétrica da distribuição Von Mises para modelar estes dados e compararam este novo modelo com o modelo Von Mises por meio das estatísticas de Kuiper e de Watson (JAMMALAMADAKA; SENGUPTA, 2001). Para verificar a bondade do ajuste do modelo *wrapped* Birnbaum-Saunders (W$\mathcal{BS}$) a estes mesmos dados também foram calculadas as estatísticas de Kuiper e de Watson, obtidas em Jammalamadaka e SenGupta (2001), e os valores foram comparados com os valores obtidos por Umbach e Jammalamadaka (2009). A Tabela 5 mostra as EMV para μ e ρ nos modelos W$\mathcal{BS}$, Von Mises e Von Mises assimétrico e as estatísticas de Kuiper e de Watson para os três modelos. O menor valor das estatísticas de Kuiper e de Watson indica um melhor ajuste e, neste caso, o modelo W$\mathcal{BS}$ apresentou um ajuste melhor aos dados do que os outros dois modelos.

Tabela 5 – Estimativas de MV dos parâmetros μ e ρ em diversos modelos e estatística de Kuiper e de Watson

Parâmetro	W$\mathcal{BS}$ (EP)	EMV VM (EP)	VMa (EP)
μ	$3,4754$ $(0,1365)$	$3,1963$ $(0,1029)$	$2,8808$ (0.1564)
ρ	$0,3095$ $(0,1292)$	$0,6084$ (0.0487)	$0,6173$ $(0,0488)$
Estatística de Kuiper	$2,6595$	$12,5958$	$11,2918$
Estatística de Watson	$0,2574$	$0,3261$	$0,2966$

2.5 Conclusões

Neste capítulo apresentamos a proposta de uma nova distribuição circular chamada de *wrapped* Birnbaum-Saunders que é uma alternativa assimétrica para modelagem de dados circulares. Derivou-se expressões para a sua função densidade de probabilidade, função de distribuição acumulada e para os seus momentos trigonométricos. Foi usado o método da máxima verossimilhança para estimar seus parâmetros e realizou-se um estudo de simulação de Monte Carlo para avaliar o desempenho dos estimadores. Finalmente, usou-se um conjunto de dados real para mostrar a aplicabilidade do modelo proposto e comparou-se sua performance com a dos modelos Von Mises simétrico e assimétrico por meio das estatísticas de Kuiper e de Watson para dados circulares. Concluiu-se que a distribuição *wrapped* Birnbaum-Saunders apresentou desempenho satisfatório nas simulações e se ajustou de forma mais apropriada aos dados da aplicação do que os outros dois modelos.

3 MISTURA DE ESCALA T-STUDENT

Neste capítulo será apresentada a classe de distribuições *Mistura de Escala t-Student* (SMt) que é uma alternativa à classe mistura de escala Normal (SMN), proposta por Andrews e Mallows (1974), e que já foi utilizada em diversos trabalhos como, por exemplo, West (1987) que estabeleceu uma relação entre a família exponencial potência e a classe SMN, Garay et al. (2017b) que desenvolveram um modelo de regressão linear para dados censurados com erros pertencentes à classe SMN, entre outros. As distribuições pertencentes à classe SMt possuem caudas mais pesadas que as distribuições da classe SMN com mesmo fator de escala, como por exemplo, o caso da t-Contaminada e da Slash-t, da classe SMt, que possuem caudas mais pesadas que a normal contaminada e Slash, da classe SMN, respectivamente (veja Figura 2). Além disso, as distribuições da classe SMt possuem curtose maior que da classe SMN com mesmo fator de escala, podendo também ser utilizadas em situações em que os dados apresentem observações aberrantes.

Também são apresentados os métodos para se obter os estimadores dos parâmetros da classe SMt, como por exemplo, o Método dos Momentos (MM) e o Método da Máxima Verossimilhança (MV), sendo que para o MV desenvolvemos o algoritmo EM e a maximização direta (MD) por meio do método quasi Newton BFGS, sendo obtidos também intervalos de confiança assintóticos (ICA) para os parâmetros. Realizamos estudos de simulação de Monte Carlo (MC) para avaliar o desempenho dos estimadores e também aplicamos os resultados obtidos a um conjunto de dados reais.

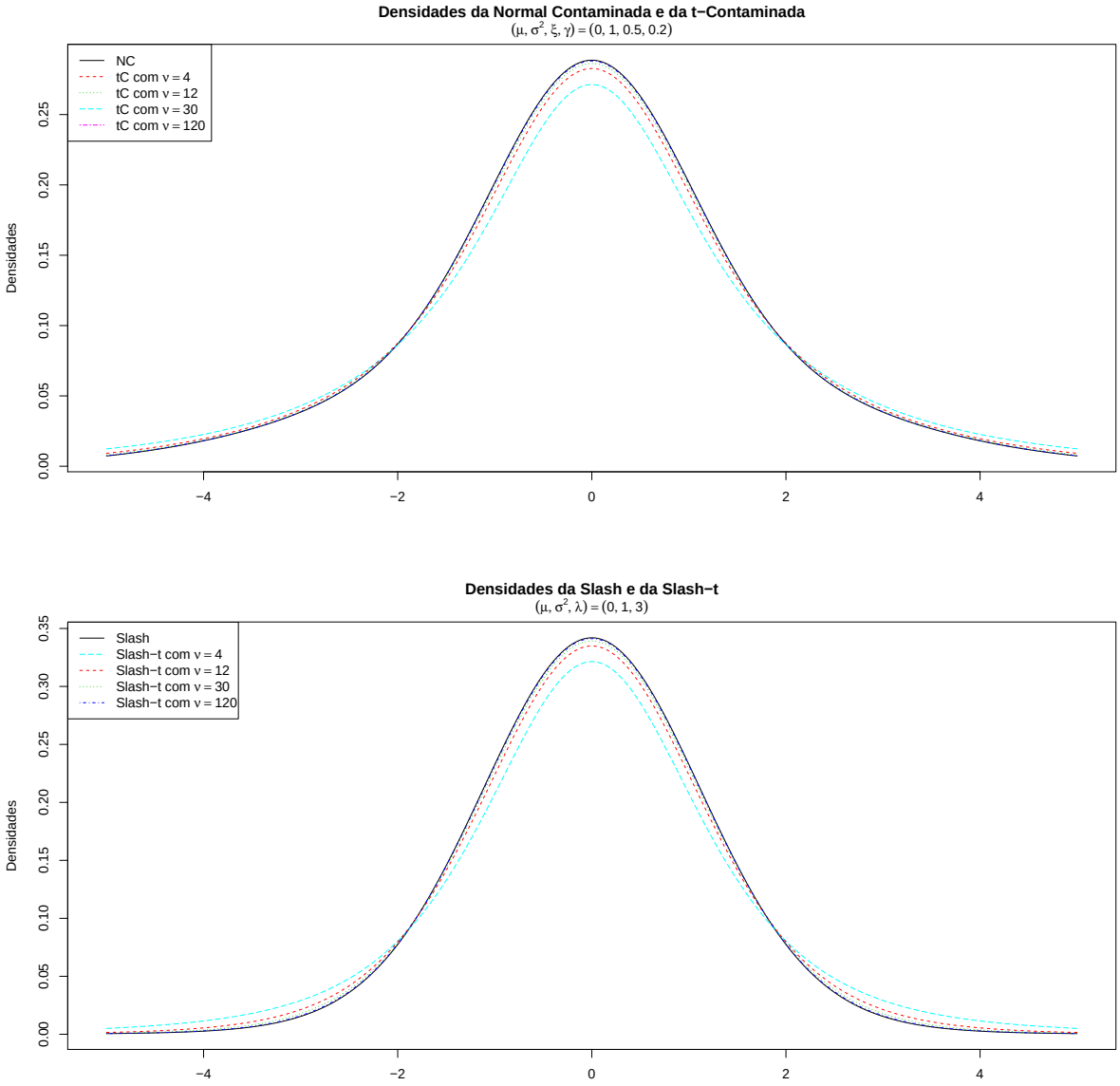


Figura 2 – Densidades concorrentes das Classes mistura de escala normal (SMN) e mistura de escala t-Student (SMt)

3.1 Modelos Mistura de Escala t-Student

Nesta seção será apresentada a classe de distribuições mistura de escala t-Student (SMt) e algumas de suas propriedades. Substituindo a variável aleatória Z em (1.1) por uma variável aleatória $T \sim t_\nu(0, \sigma^2)$ obtém-se a classe de distribuições mistura de escala t-Student, cuja definição é apresentada a seguir:

Definição 3.1.1. Uma variável aleatória X tem distribuição na classe Mistura de Escala t-Student com parâmetro de locação μ, parâmetro de escala σ^2, ν graus de liberdade e

fator de escala U se possui a seguinte representação estocástica:

$$X = \mu + U^{-\frac{1}{2}}T, \quad T \perp U \tag{3.1}$$

em que $T \sim t_\nu(0, \sigma^2)$, isto é, T segue uma distribuição t-Student com ν graus de liberdade, média zero e parâmetro de escala σ^2, $\mu \in \mathbb{R}$, $\sigma^2 > 0$ e U é uma variável aleatória positiva com função de distribuição acumulada $H(\cdot|\boldsymbol{\lambda})$ indexada pelo parâmetro $\boldsymbol{\lambda}$, que pode ser escalar ou vetor. Será utilizada a notação $X \sim \text{SMt}_\nu(\mu, \sigma^2; \boldsymbol{\lambda})$. Quando $\mu = 0$ e $\sigma^2 = 1$ tem-se que X segue uma distribuição SMt padrão e denota-se por $X \sim \text{SMt}_\nu(0, 1; \boldsymbol{\lambda})$.

Nota-se de (3.1) que $X|U \sim t_\nu(\mu, u^{-1}\sigma^2)$. Portanto, integrando em U a densidade conjunta de X e U obtém-se a densidade marginal de X:

$$f_{SMt}(x|\mu, \sigma^2, \nu, \boldsymbol{\lambda}) = \frac{c}{\sigma} \int_0^\infty \sqrt{u} \left[\nu + u\left(\frac{x-\mu}{\sigma}\right)^2\right]^{-\left(\frac{\nu+1}{2}\right)} dH(u|\boldsymbol{\lambda}), \tag{3.2}$$

em que $c = \frac{\Gamma\left(\frac{\nu+1}{2}\right)\nu^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi}}$ e $H(\cdot|\boldsymbol{\lambda})$ é a distribuição de mistura. Dependendo da escolha do fator de escala U temos uma distribuição em particular da classe SMt. A seguir são apresentados alguns exemplos de distribuições da classe SMt assumindo algumas distribuições usuais para a variável aleatória U.

É importante notar que, assim como a SMN (detalhes em Garay et al. (2017b)), existe uma relação entre as distribuições da classe SMt e as distribuições elípticas. Diz-se que a variável aleatória X possui distribuição elíptica univariada, com parâmetro de locação μ e parâmetro de dispersão σ^2, se a sua função de densidade de probabilidade é dada por

$$f(x) = \sigma^{-1} g(z), \tag{3.3}$$

em que $z = (x-\mu)^2/\sigma^2$ e $g : \mathbb{R} \to [0, \infty)$ satisfaz a desigualdade $\int_0^\infty g(z)dz < \infty$. Evidentemente que (3.2) tem a forma (3.3) e portanto, as distribuições da classe SMt também são distribuições elípticas.

• **Distribuição t-Student:** Assumindo que em (3.1) U é uma variável aleatória degenerada em 1 cuja função de probabilidade dada por

$$\mathbb{P}(U = 1) = 1 \text{ e } \mathbb{P}(U \neq 1) = 0. \tag{3.4}$$

Tem-se que $X \sim t_\nu(\mu, \sigma^2)$, ou seja, X segue uma distribuição t-Student com parâmetros μ, σ^2 e com ν graus de liberdade, cuja função de densidade de probabilidade é dada por

$$f_\nu(x|\mu,\sigma^2) = \frac{c}{\sigma}\left[\nu + \left(\frac{x-\mu}{\sigma}\right)^2\right]^{-\left(\frac{\nu+1}{2}\right)}, \tag{3.5}$$

em que c foi dado em (3.2).

• **Distribuição t-Student Contaminada:** Assumindo que em (3.1) U é uma variável aleatória dicotômica tal que $\mathbb{P}(U=\gamma)=\xi$ e $\mathbb{P}(U=1)=1-\xi$, tem-se que X terá distribuição t-Student contaminada. Neste caso, a função de probabilidade de U é dada por

$$h(u|\xi,\gamma) = \xi\,\mathbb{I}_{\{\gamma\}}(u) + (1-\xi)\,\mathbb{I}_{\{1\}}(u);\ \gamma,\xi \in (0,1), \tag{3.6}$$

em que $\mathbb{I}_B(u)$ é a função indicadora de $u \in B$.

Portanto, sendo $\boldsymbol{\lambda} = (\xi,\gamma)^\top$, tem-se que

$$f_{SMt}(x|\mu,\sigma^2;\nu,\boldsymbol{\lambda}) = \xi\, f_\nu(x|\mu,\gamma^{-1}\sigma^2) + (1-\xi)\, f_\nu(x|\mu,\sigma^2), \tag{3.7}$$

em que $f_\nu(x|\mu,\sigma^2)$ representa a função densidade de probabilidade da distribuição t-Student com parâmetros μ, σ^2 e com ν graus de liberdade.

• **Distribuição Slash-t:** Assumindo que em (3.1) $U \sim \text{Beta}(\lambda,1)$ com fdp dada por $h(u|\lambda) = \lambda u^{\lambda-1}$, em que $\lambda > 0$, segue que

$$\begin{aligned} f_{SMt}(x|\mu,\sigma^2;\nu,\lambda) &= \lambda \int_0^1 u^{\lambda-1} f_\nu(x|\mu,\lambda^{-1}\sigma^2)du \\ &= \frac{c\lambda}{\sigma}\int_0^1 u^{\lambda-\frac{1}{2}}\left[\nu + u\left(\frac{x-\mu}{\sigma}\right)^2\right]^{-\left(\frac{\nu+1}{2}\right)} du. \end{aligned} \tag{3.8}$$

• **Distribuição Weibull-t:** Assumindo que em (3.1) $U \sim \text{Weibull}(\lambda,1)$ com fdp dada por $h(u|\lambda) = \lambda u^{\lambda-1}\exp(-u^\lambda)$, em que $\lambda > 0$, segue que

$$\begin{aligned} f_{SMt}(x|\mu,\sigma^2;\nu,\lambda) &= \lambda \int_0^\infty u^{\lambda-1}\exp(-u^\lambda) f_\nu(x|\mu,\lambda^{-1}\sigma^2)du \\ &= \frac{c\lambda}{\sigma}\int_0^\infty u^{\lambda-\frac{1}{2}}\exp(-u^\lambda)\left[\nu + u\left(\frac{x-\mu}{\sigma}\right)^2\right]^{-\left(\frac{\nu+1}{2}\right)} du. \end{aligned} \tag{3.9}$$

Tanto em (3.8) como em (3.9) a integral que define a densidade de probabilidade pode ser obtida por métodos numéricos, uma vez que não há solução em forma fechada similarmente à distribuição Slash. Neste trabalho as integrais foram calculadas pelo método de

integração de Gauss-Legendre por meio do pacote *distEX* desenvolvido por Ruckdeschel et al. (2006) na linguagem de programação R (R Core Team, 2018). Detalhes teóricos sobre o método de integração de Gauss-Legendre podem ser encontrados em Piessens et al. (1983).

A seguir é apresentado um resultado importante em que demonstra-se algumas propriedades relativas às distribuições da classe SMt.

Proposição 3.1.2. *Seja* $X \sim \mathrm{SMt}_\nu(\mu, \sigma^2; \boldsymbol{\lambda})$*, logo, segue que:*

i. *Se* $\mathbb{E}[U^{-\frac{1}{2}}] < \infty$ *então* $\mathbb{E}[X] = \mu$*;*

ii. *Se* $\mathbb{E}[U^{-1}] < \infty$ *e* $\nu > 2$ *então* $\mathbb{E}[X^2] = \mu^2 + \frac{\nu\sigma^2}{\nu-2}\mathbb{E}[U^{-1}]$ *e* $\mathbb{E}[X^3] = \mu^3 + \frac{3\mu\nu\sigma^2}{\nu-2}\mathbb{E}[U^{-1}]$*;*

iii. *Se* $\mathbb{E}[U^{-1}] < \infty$*,* $\mathbb{E}[U^{-2}] < \infty$ *e* $\nu > 4$ *então*

$$\mathbb{E}[X^4] = \mu^4 + \frac{\nu\sigma^2}{\nu-2}\left\{6\mu^2\mathbb{E}[U^{-1}] + \frac{3\nu\sigma^2}{\nu-4}\mathbb{E}[U^{-2}]\right\} \text{ e}$$

$$\mathbb{E}[X^5] = \mu^5 + \frac{\nu\sigma^2}{\nu-2}\left\{10\mu^3\mathbb{E}[U^{-1}] + \frac{15\mu\nu\sigma^2}{\nu-4}\mathbb{E}[U^{-2}]\right\};$$

iv. *Se* $\mathbb{E}[U^{-1}] < \infty$*,* $\mathbb{E}[U^{-2}] < \infty$*,* $\mathbb{E}[U^{-3}] < \infty$ *e* $\nu > 6$ *então*

$$\mathbb{E}[X^6] = \mu^6 + \frac{\nu\sigma^2}{\nu-2}\left\{15\mu^4\mathbb{E}[U^{-1}] + \frac{45\mu^2\nu\sigma^2}{\nu-4}\mathbb{E}[U^{-2}] + \frac{15\nu^2\sigma^4}{(\nu-6)(\nu-4)}\mathbb{E}[U^{-3}]\right\}.$$

Demonstração. Supondo que $\mathbb{E}[U^{-r}] < \infty$, para $r \in \left\{-3, -2, -1, -\frac{1}{2}\right\}$, tem-se que

i. $\mathbb{E}[X] = \mu + \mathbb{E}[U^{-\frac{1}{2}}T] = \mu + \mathbb{E}[U^{-\frac{1}{2}}]\mathbb{E}[T] = \mu$;

ii. $\mathbb{E}[X^2] = \mathrm{Var}[X] + \mathbb{E}^2[X] = \mu^2 + \frac{\nu\sigma^2}{\nu-2}\mathbb{E}[U^{-1}]$ para $\nu > 2$;

$$\begin{aligned}
\mathbb{E}[X^3] &= \mathbb{E}[\mu^3 + 3\mu^2U^{-1/2}T + 3\mu U^{-1}T^2 + U^{-3/2}T^3] \\
&= \mu^3 + 3\mu^2\mathbb{E}[U^{-1/2}]\mathbb{E}[T] + 3\mu\mathbb{E}[U^{-1}]\mathbb{E}[T^2] + \mathbb{E}[U^{-3/2}]\mathbb{E}[T^3] \\
&= \mu^3 + 3\mu\mathbb{E}[U^{-1}]\mathbb{E}[T^2] = \mu^3 + \frac{3\mu\nu\sigma^2}{\nu-2}\mathbb{E}[U^{-1}] \quad \text{para} \quad \nu > 2;
\end{aligned}$$

iii.

$$\begin{aligned}
\mathbb{E}[X^4] &= \mathbb{E}[(\mu + U^{-1/2}T)^4] = \mathbb{E}[(\mu + U^{-1/2}T)^2(\mu + U^{-1/2}T)^2] \\
&= \mathbb{E}[\mu^4 + 4\mu^3 U^{-1/2}T + 6\mu^2 U^{-1}T^2 + 4\mu U^{-3/2}T^3 + U^{-2}T^4] \\
&= \mu^4 + 4\mu^3\mathbb{E}[U^{-1/2}]\mathbb{E}[T] + 6\mu^2\mathbb{E}[U^{-1}]\mathbb{E}[T^2] \\
&+ 4\mu\mathbb{E}[U^{-3/2}]\mathbb{E}[T^3] + \mathbb{E}[U^{-2}]\mathbb{E}[T^4] \\
&= \mu^4 + 6\mu^2\mathbb{E}[U^{-1}]\mathbb{E}[T^2] + \mathbb{E}[U^{-2}]\mathbb{E}[T^4] \\
&= \mu^4 + \frac{6\mu^2\nu\sigma^2}{\nu-2}[U^{-1}] + \frac{3\nu^2\sigma^4}{(\nu-4)(\nu-2)}\mathbb{E}[U^{-2}] \\
&= \mu^4 + \frac{\nu\sigma^2}{\nu-2}\left\{6\mu^2\mathbb{E}[U^{-1}] + \frac{3\nu\sigma^2}{\nu-4}\mathbb{E}[U^{-2}]\right\} \quad \text{para} \quad \nu > 4;.
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[X^5] &= \mathbb{E}[(\mu + U^{-1/2}T)^5] = \mathbb{E}[(\mu + U^{-1/2}T)^4(\mu + U^{-1/2}T)] \\
&= \mathbb{E}[(\mu^4 + 4\mu^3 U^{-1/2}T + 6\mu^2 U^{-1}T^2 + 4\mu U^{-3/2}T^3 + U^{-2}T^4)(\mu + U^{-1/2}T)] \\
&= \mathbb{E}[\mu^5 + 5\mu^4 U^{-1/2}T + 10\mu^3 U^{-1}T^2 + 10\mu^2 U^{-3/2}T^3 + 5\mu U^{-2}T^4 + U^{-5/2}T^5] \\
&= \mu^5 + 5\mu^4\mathbb{E}[U^{-1/2}]\mathbb{E}[T] + 10\mu^3\mathbb{E}[U^{-1}]\mathbb{E}[T^2] + 10\mu^2\mathbb{E}[U^{-3/2}]\mathbb{E}[T^3] \\
&+ 5\mu\mathbb{E}[U^{-2}]\mathbb{E}[T^4] + \mathbb{E}[U^{-5/2}]\mathbb{E}[T^5] \\
&= \mu^5 + 10\mu^3\mathbb{E}[U^{-1}]\mathbb{E}[T^2] + 5\mu\mathbb{E}[U^{-2}]\mathbb{E}[T^4] \\
&= \mu^5 + \frac{10\mu^3\nu\sigma^2}{\nu-2}\mathbb{E}[U^{-1}] + \frac{15\mu\nu^2\sigma^4}{(\nu-4)(\nu-2)}\mathbb{E}[U^{-2}] \\
&= \mu^5 + \frac{\nu\sigma^2}{\nu-2}\left\{10\mu^3\mathbb{E}[U^{-1}] + \frac{15\mu\nu\sigma^2}{\nu-4}\mathbb{E}[U^{-2}]\right\} \quad \text{para} \quad \nu > 4;.
\end{aligned}$$

iv.

$$\begin{aligned}
\mathbb{E}[X^6] &= \mathbb{E}[(\mu + U^{-1/2}T)^6] = \mathbb{E}[(\mu + U^{-1/2}T)^5(\mu + U^{-1/2}T)] \\
&= \mathbb{E}[\mu^6 + 6\mu^5 U^{-1/2}T + 15\mu^4 U^{-1}T^2 + 20\mu^3 U^{-3/2}T^3 \\
&+ 15\mu^2 U^{-2}T^4 + 6\mu U^{-5/2}T^5 + U^{-3}T^6] \\
&= \mu^6 + 6\mu^5\mathbb{E}[U^{-1/2}]\mathbb{E}[T] + 15\mu^4\mathbb{E}[U^{-1}]\mathbb{E}[T^2] + 20\mu^3\mathbb{E}[U^{-3/2}]\mathbb{E}[T^3] \\
&+ 15\mu^2\mathbb{E}[U^{-2}]\mathbb{E}[T^4] + 6\mu\mathbb{E}[U^{-5/2}]\mathbb{E}[T^5] + \mathbb{E}[U^{-3}]\mathbb{E}[T^6] \\
&= \mu^6 + 15\mu^4\mathbb{E}[U^{-1}]\mathbb{E}[T^2] + 15\mu^2\mathbb{E}[U^{-2}]\mathbb{E}[T^4] + \mathbb{E}[U^{-3}]\mathbb{E}[T^6] \\
&= \mu^6 + \frac{15\mu^4\nu\sigma^2}{\nu-2}\mathbb{E}[U^{-1}] + \frac{45\mu^2\nu^2\sigma^4}{(\nu-4)(\nu-2)}\mathbb{E}[U^{-2}]
\end{aligned}$$

$$
\begin{aligned}
&+ \frac{15\nu^3\sigma^6}{(\nu-6)(\nu-4)(\nu-2)}\mathbb{E}[U^{-3}] \\
&= \mu^6 + \frac{\nu\sigma^2}{\nu-2}\left\{15\mu^4\mathbb{E}[U^{-1}] + \frac{45\mu^2\nu\sigma^2}{\nu-4}\mathbb{E}[U^{-2}] + \frac{15\nu^2\sigma^4}{(\nu-6)(\nu-4)}\mathbb{E}[U^{-3}]\right\} \\
&\text{para} \quad \nu > 6.
\end{aligned}
$$

□

Para mais detalhes a respeito dos momentos da v.a. $T \sim t_\nu(0, \sigma^2)$, veja Johnson, Kotz e Balakrishnan (1995, pág. 365), Wolfram Research, Inc. (2018) e Wolfram|Alpha (2019).

Como aplicação direta da Proposição 3.1.2 é possível obter alguns dos momentos centrais de $X \sim \text{SMt}_\nu(\mu, \sigma^2; \boldsymbol{\lambda})$, que são dados a seguir:

i. Se $\mathbb{E}[U^{-1}] < \infty$ e $\nu > 2$ então $\text{Var}[X] = \frac{\nu\sigma^2}{\nu-2}\mathbb{E}[U^{-1}]$ e $\mathbb{E}[X-\mu]^3 = 0$;

ii. Se $\mathbb{E}[U^{-1}] < \infty$, $\mathbb{E}[U^{-2}] < \infty$ e $\nu > 4$ então $\mathbb{E}[X-\mu]^4 = \frac{3\nu^2\sigma^4}{(\nu-4)(\nu-2)}\mathbb{E}[U^{-2}]$ e $\mathbb{E}[X-\mu]^5 = 0$;

iii. Se $\mathbb{E}[U^{-1}] < \infty$, $\mathbb{E}[U^{-2}] < \infty$, $\mathbb{E}[U^{-3}] < \infty$ e $\nu > 6$ então $\mathbb{E}[X-\mu]^6 = \frac{15\nu^3\sigma^6}{(\nu-6)(\nu-4)(\nu-2)}\mathbb{E}[U^{-3}]$.

Uma notação usual e compacta para os momentos centrais é

$$\mathbb{E}[X-\mu]^r = \mu_r,$$

de modo que μ_r representa o r-ésimo momento central de X. Uma vez que se conheça os quatro primeiros momentos centrais de X se torna possível obter sua curtose, que é dada por

$$\kappa_4^{\text{SMt}} = \frac{\mu_4 - 4\mu_1\mu_3 + 6\mu_2\mu_1^2 - 3\mu_1^4}{(\mu_2 - \mu_1^2)^2} = \frac{\mu_4}{\mu_2^2} = \frac{3(\nu-2)}{\nu-4}\frac{\mathbb{E}[U^{-2}]}{\mathbb{E}^2[U^{-1}]},$$

para $\nu > 4$, uma vez que $\mu_1 = \mu_3 = 0$. Semelhantemente, é possível obter curtose da classe SMN, que é dada por $\kappa_4^{\text{SMN}} = \frac{3\mathbb{E}[U^{-2}]}{\mathbb{E}^2[U^{-1}]}$. Assim, é possível observar que $\frac{\kappa_4^{\text{SMt}}}{\kappa_4^{\text{SMN}}} = \frac{\nu-2}{\nu-4} > 1$, para $\nu > 4$, ou seja, a curtose na classe SMt é maior do que na classe SMN considerando um mesmo fator de escala U. Além disso, o limite desse quociente quando ν tende ao infinito é 1, ou seja, $\lim_{\nu\to\infty}\frac{\nu-2}{\nu-4} = 1$. Portanto, um dos atrativos da classe SMt em relação à classe SMN é a sua maior capacidade em acomodar dados aberrantes devido a ter uma maior curtose. Por exemplo, quando $U \sim \text{Beta}(\lambda, 1)$ tem-se que, se X pertence à classe

SMt então terá distribuição Slash-t cuja fdp está dada na equação (3.8) e, se X pertence à classe SMN terá distribuição Slash e a curtose em ambos os casos é dada, respectivamente, por $\kappa_4^{\text{SMt}} = \frac{3(\lambda-1)^2(\nu-2)}{\lambda(\lambda-4)(\nu-4)}$ e $\kappa_4^{\text{SMN}} = \frac{3(\lambda-1)^2}{\lambda(\lambda-4)}$. Com $\nu = 6$, por exemplo, $\kappa_4^{\text{SMt}} = 2\kappa_4^{\text{SMN}}$, isto é, com seis graus de liberdade a curtose da classe SMt é o dobro da curtose da classe SMN quando o fator de escala é $U \sim \text{Beta}(\lambda, 1)$.

No gráfico da Figura 3, com $\nu = 4$ fixado, é possível ver que as distribuições da classe SMt possuem caudas mais pesadas do que da classe SMN, isto significa que a área abaixo da curva para valores extremos é maior nas distribuições da classe SMt, indicando que esta classe SMt é resistente a valores aberrantes. A distribuição com caudas mais pesadas neste exemplo é a Slash-t, seguida pela Slash, t-Student contaminada, t-Student, normal contaminada e normal.

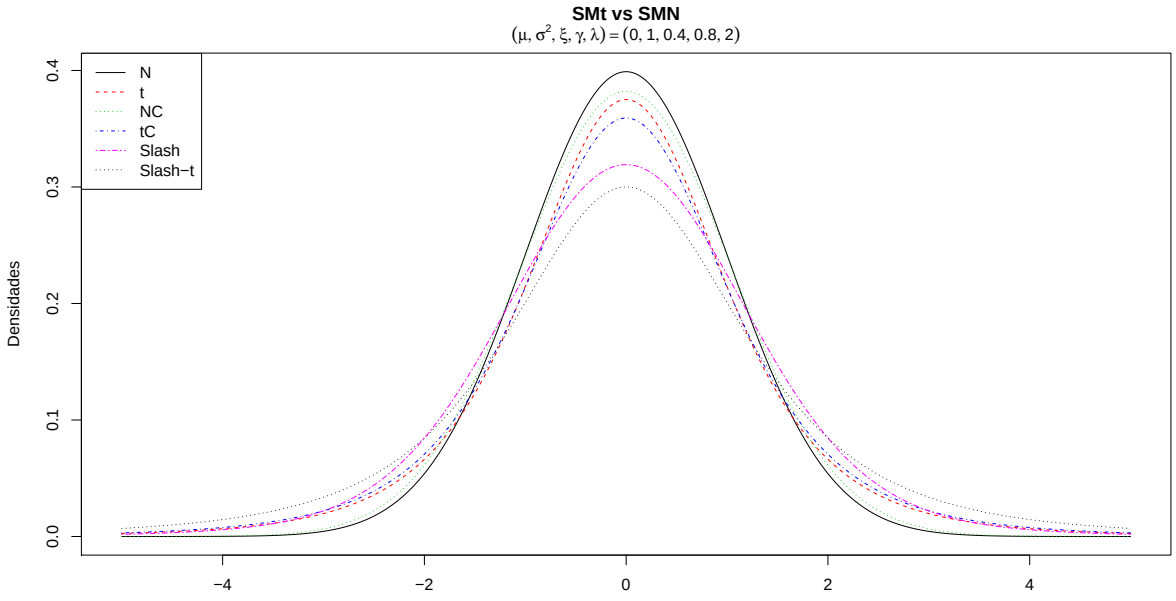


Figura 3 – Densidades da classe SMt e da classe SMN. Normal (N), t-Student (t), Normal Contaminada (NC), t-Student Contaminada (tC), Slash e Slash-t.

Um outro resultado importante da classe SMt é que esta classe converge para a classe SMN quando os graus de liberdade tendem ao infinito. Este resultado é mostrado na seguinte proposição.

Proposição 3.1.3. *Se* $X \sim \text{SMt}_\nu(\mu, \sigma^2; \boldsymbol{\lambda})$ *e* $Y \sim \text{SMN}(\mu, \sigma^2; \boldsymbol{\lambda})$ *com o mesmo fator de escala* U *e a função de distribuição de* $U^{-\frac{1}{2}}$ *existe, então,* $X \xrightarrow{D} Y$ *quando* $\nu \to \infty$, *em*

que $\xrightarrow{D}$ *representa convergência em distribuição.*

Demonstração. Sabe-se que se $T \sim t_\nu(0, \sigma^2)$ então $T \xrightarrow{D} Z$ quando $\nu \to \infty$, em que $Z \sim \mathrm{N}(0, \sigma^2)$, ou seja, a fda F_T de T converge para a fda Φ_Z de Z para todo t ponto de continuidade de F_T, quando $\nu \to \infty$. Agora, seja $X \sim \mathrm{SMt}_\nu(\mu, \sigma^2; \boldsymbol{\lambda})$ então $X = \mu + U^{-1/2}T$, $U \perp T$, e seja $Y \sim \mathrm{SMN}(\mu, \sigma^2; \boldsymbol{\lambda})$ então $Y = \mu + U^{-1/2}Z$, $U \perp Z$. Além disso, a fda de $U^{-1/2}$ não depende de ν. Dessa forma, sendo $H_{U^{-1/2}}$ a fda de $U^{-1/2}$, tem-se que $\mu + H_{U^{-1/2}}F_T$ converge para $\mu + H_{U^{-1/2}}\Phi_Z$ para todo t ponto de continuidade de F_T e, portanto, $X \xrightarrow{D} Y$, quando $\nu \to \infty$. □

3.2 Estimação dos Parâmetros nos Modelos SMt

Neste trabalho o valor de ν (os graus de liberdade) será considerado fixado e uma justificativa para isto pode ser encontrada em Berkane, Kano e Bentler (1994), Lange, Little e Taylor (1989) e Massuia et al. (2015), que apontam dificuldades em estimar ν devido a problemas com relação à máximos locais e verossimilhaça ilimitada próximo à fronteira do espaço paramétrico. Além disso, em Garay et al. (2017a) também se considerou ν fixado, uma vez que foi mostrado em Lucas (1997) que a proteção contra valores aberrantes é preservada apenas se os graus de liberdade são fixados.

Com o objetivo de se estimar os parâmetros do modelo SMt foram utilizados os métodos dos momentos (MM) e de máxima verossimilhança (MV). Além de se estimar os parâmetros, também há interesse em se estimar a variância de X, cuja expressão depende do valor esperado de U^{-1}. Logo, na Tabela 6 são apresentadas as expressões para a variância de X considerando diferentes distribuições para o fator de escala U.

Tabela 6 – Valor esperado de U^{-s}, $s \in \{1, 2, 3\}$ e variância de X para algumas distribuições da classe SMt

U	Degenerada em 1	Dicotômica(ξ, γ)	Weibull$(\lambda, 1)$	Beta$(\lambda, 1)$
$\mathbb{E}[U^{-s}]$	1	$\left[1 - \xi + \frac{\xi}{\gamma^s}\right]$	$\Gamma\left(\frac{\lambda - s}{\lambda}\right)$	$\frac{\lambda}{\lambda - s}$
X	t-Student	t-Contaminada	Weibull-t	Slash-t
$\mathrm{Var}(X)$	$\left[\frac{\nu\sigma^2}{\nu-2}\right]$	$\left[\frac{\nu\sigma^2}{\nu-2}\right]\left[1 - \xi + \frac{\xi}{\gamma}\right]$	$\left[\frac{\nu\sigma^2}{\nu-2}\right]\Gamma\left(\frac{\lambda-1}{\lambda}\right)$	$\left[\frac{\nu\sigma^2}{\nu-2}\right]\left(\frac{\lambda}{\lambda-1}\right)$

Com o objetivo de se obter os EMV nos modelos sob a classe SMt utilizou-se a maximização direta pelo método quasi-Newton BFGS de otimização não-linear com primeiras derivadas numéricas e também foram utilizados o algoritmo EM e tipo EM, dependendo do caso.

3.2.1 Método dos Momentos

Seja $\boldsymbol{x} = (x_1, \ldots, x_n)^\top$ uma amostra aleatória da v.a. $X \sim \text{SMt}_\nu(\mu, \sigma^2; \boldsymbol{\lambda})$ em que $\boldsymbol{\theta} = (\mu, \sigma^2, \boldsymbol{\lambda})^\top$ é o vetor de parâmetros. Para estimar $\boldsymbol{\theta}$ pelo MM nos baseamos no primeiro momento não-central e nos momentos centrais pares, uma vez que os momentos não-centrais de ordem maior ou igual a dois possuem expressões mais extensas que as expressões dos momentos centrais e cuja tratativa algébrica seria mais tediosa. Utilizaremos a definição usual de estimador de método dos momentos (EMM), que consiste na solução do sistema de equações obtido quando se iguala os momentos populacionais aos respectivos momentos amostrais, com tantas equações quanto parâmetros que se deseja estimar. Detalhes sobre o EMM podem ser obtidos em Matyas (1999, pág. 7).

Uma restrição deste método é a imposição de que $\nu > 6$, o que pode ser considerado um valor alto para os graus de liberdade. Além disso, os sistemas de equações obtidos são não lineares cuja solução, mesmo numericamente, pode ser bastante complicada. O objetivo é utilizar o EMM como valor inicial no processo iterativo para obtenção do EMV e, considerando ν e λ fixados, obtém-se um sistema de equações com duas incógnitas e duas equações com solução analítica explícita, sendo que a restrição a respeito dos graus de liberdade passa a ser $\nu > 2$. Assim, obtém-se o seguinte sistema de equações:

$$\begin{cases} \mu = \bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n} \\ \mu_2 = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} \\ \mu_4 = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})^4}{n} \\ \mu_6 = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})^6}{n} \end{cases} \implies \begin{cases} \mu = \bar{x} \\ \dfrac{\nu\sigma^2}{\nu-2}\mathbb{E}[U^{-1}] = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} \\ \dfrac{3\nu^2\sigma^4}{(\nu-4)(\nu-2)}\mathbb{E}[U^{-2}] = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})^4}{n} \\ \dfrac{15\nu^3\sigma^6}{(\nu-6)(\nu-4)(\nu-2)}\mathbb{E}[U^{-3}] = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})^6}{n} \end{cases} \tag{3.10}$$

Para cada fator de escala U o sistema de equações em (3.10) assumirá uma forma

distinta, uma vez que varia com os valores esperados $\mathbb{E}[U^{-s}]$, $s \in \{1,2,3\}$. Portanto, a partir de (3.10) e da Tabela 6 o EMM, se existir, será uma solução do sistema de equações em cada caso, como se pode ver a seguir.

• **Distribuição t-Student (U é degenerada):** Os estimadores pelo método dos momentos de $\boldsymbol{\theta} = (\mu, \sigma^2)^\top$ são dados por:

$$\begin{cases} \mu = \bar{x} \\ \dfrac{\nu\sigma^2}{\nu-2} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})^2 \end{cases} \implies \begin{cases} \hat{\mu} = \bar{x} \\ \hat{\sigma}^2 = \left(\dfrac{\nu-2}{\nu}\right)\left(\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})^2\right) \end{cases}$$

Neste caso em que se tem apenas os parâmetros μ e σ^2 as expressões analíticas do EMM são dadas de forma explícita.

• **Distribuição t-Student Contaminada (U é dicotômica):** Os estimadores dos parâmetros $\boldsymbol{\theta} = (\mu, \sigma^2, \xi, \gamma)^\top$ obtidos pelo MM são solução do sistema de equações dado por:

$$\begin{cases} \mu = \bar{x} \\ \dfrac{\nu\sigma^2}{\nu-2}\left(1 - \xi + \dfrac{\xi}{\gamma}\right) = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})^2 \\ \dfrac{3\nu^2\sigma^4}{(\nu-4)(\nu-2)}\left(1 - \xi + \dfrac{\xi}{\gamma^2}\right) = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})^4 \\ \dfrac{15\nu^3\sigma^6}{(\nu-6)(\nu-4)(\nu-2)}\left(1 - \xi + \dfrac{\xi}{\gamma^3}\right) = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})^6 \end{cases} \implies$$

$$\begin{cases} \hat{\mu} = \bar{x} \\ \hat{\sigma}^2 = \left[\dfrac{\nu-2}{\nu\left(1 - \hat{\xi} + \frac{\hat{\xi}}{\hat{\gamma}}\right)}\right]\left(\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})^2\right) \\ 1 - \hat{\xi} + \dfrac{\hat{\xi}}{\hat{\gamma}^2} = \left(\dfrac{(\nu-4)(\nu-2)}{3\nu^2\hat{\sigma^4}}\right)\left(\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})^4\right) \\ 1 - \hat{\xi} + \dfrac{\hat{\xi}}{\hat{\gamma}^3} = \left(\dfrac{(\nu-6)(\nu-4)(\nu-2)}{15\nu^3\hat{\sigma^6}}\right)\left(\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})^6\right) \end{cases} \tag{3.11}$$

• **Distribuição Weibull-t:** Os estimadores dos parâmetros $\boldsymbol{\theta} = (\mu, \sigma^2, \lambda)^\top$ obtidos

pelo MM são solução do sistema de equações dado por:

$$\begin{cases} \mu = \bar{x} \\ \dfrac{\nu\sigma^2}{\nu-2}\Gamma\left(\dfrac{\lambda-1}{\lambda}\right) = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i-\bar{x})^2 \\ \dfrac{3\nu^2\sigma^4}{(\nu-4)(\nu-2)}\Gamma\left(\dfrac{\lambda-2}{\lambda}\right) = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i-\bar{x})^4 \end{cases} \Longrightarrow$$

$$\begin{cases} \hat{\mu} = \bar{x} \\ \hat{\sigma}^2 = \left(\dfrac{\nu-2}{\nu\Gamma\left(\frac{\hat{\lambda}-1}{\hat{\lambda}}\right)}\right)\left(\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i-\bar{x})^2\right) \\ \Gamma\left(\dfrac{\hat{\lambda}-2}{\hat{\lambda}}\right) = \left(\dfrac{(\nu-4)(\nu-2)}{3\nu^2\hat{\sigma^4}}\right)\left(\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i-\bar{x})^4\right) \end{cases} \tag{3.12}$$

• **Distribuição Slash-t:** Os estimadores dos parâmetros $\boldsymbol{\theta} = (\mu, \sigma^2, \lambda)^\top$ obtidos pelo MM são solução do sistema de equações dado por:

$$\begin{cases} \mu = \bar{x} \\ \dfrac{\nu\sigma^2}{\nu-2}\left(\dfrac{\lambda}{\lambda-1}\right) = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i-\bar{x})^2 \\ \dfrac{3\nu^2\sigma^4}{(\nu-4)(\nu-2)}\left(\dfrac{\lambda}{\lambda-2}\right) = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i-\bar{x})^4 \end{cases} \Longrightarrow$$

$$\begin{cases} \hat{\mu} = \bar{x} \\ \hat{\sigma}^2 = \left(\dfrac{\nu-2}{\nu\left(\frac{\hat{\lambda}}{\hat{\lambda}-1}\right)}\right)\left(\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i-\bar{x})^2\right) \\ \left(\dfrac{\hat{\lambda}}{\hat{\lambda}-2}\right) = \left(\dfrac{(\nu-4)(\nu-2)}{3\nu^2\hat{\sigma^4}}\right)\left(\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i-\bar{x})^4\right) \end{cases} \Longrightarrow$$

$$\begin{cases} \hat{\mu} = \bar{x} \\ \hat{\sigma}^2 = \left(\dfrac{(\nu-2)(\hat{\lambda}-1)}{\nu\hat{\lambda}}\right)\left(\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i-\bar{x})^2\right) \\ \hat{\lambda} = \dfrac{2(\nu-4)(\nu-2)\frac{1}{n}\sum\limits_{i=1}^{n}(x_i-\bar{x})^4}{(\nu-4)(\nu-2)\frac{1}{n}\sum\limits_{i=1}^{n}(x_i-\bar{x})^4 - 3\nu^2\hat{\sigma^4}} \end{cases} \tag{3.13}$$

Os sistemas de equações (3.11), (3.12) a (3.13) não são sistemas de equações lineares e não podem ser resolvidos analiticamente, sendo necessário, em cada caso, buscar soluções por métodos numéricos iterativos. Se ν e $\boldsymbol{\lambda}$ são conhecidos, obtém-se solução analítica explícita para $\hat{\mu}$ e $\hat{\sigma}^2$.

3.2.2 Maximização Direta

Seja $\boldsymbol{x} = (x_1, \ldots, x_n)^\top$ uma a.a. da v.a. $X \sim \text{SMt}_\nu(\mu, \sigma^2; \boldsymbol{\lambda})$. O logaritmo da função de verossimilhança (função de log-verossimilhança) de $\boldsymbol{\theta} = (\mu, \sigma^2, \boldsymbol{\lambda})^\top$, dado $\boldsymbol{x}$, é dada por

$$\ell(\boldsymbol{\theta}; \boldsymbol{x}) = \sum_{i=1}^{n} \log\left\{f_{SMt}(x_i|\mu, \sigma^2; \nu, \boldsymbol{\lambda})\right\}$$

$$\ell(\boldsymbol{\theta}; \boldsymbol{x}) = \sum_{i=1}^{n} \log\left\{\frac{c}{\sigma}\int_0^\infty \sqrt{u}\left[\nu + u\left(\frac{x_i - \mu}{\sigma}\right)^2\right]^{-\left(\frac{\nu+1}{2}\right)} dH(u|\boldsymbol{\lambda})\right\},$$

em que $c = \frac{\Gamma\left(\frac{\nu+1}{2}\right)\nu^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi}}$.

Neste caso, os estimadores de máxima verossimilhança não possuem expressões analíticas explicitamente e, por isso, recorreu-se ao método quasi-Newton de otimização não linear BFGS (NOCEDAL; WRIGHT, 1999) para obter-se as EMV de $\boldsymbol{\theta}$.

3.2.3 Algoritmo EM

O algoritmo EM (*Expectation-Maximization*) foi introduzido por Dempster, Laird e Rubin (1977) com o intuito de obter-se as EMV dos parâmetros ($\boldsymbol{\theta}$) do modelo, de forma iterativa. Este algoritmo é aplicado ao problema de estimação a partir de dados incompletos, aumentando o vetor de dados observados ($\boldsymbol{x}_{obs}$) com a inclusão de variáveis latentes ($\boldsymbol{x}_{nobs}$) que não são diretamente observadas. Deste modo, obtém-se o vetor de dados completos $\boldsymbol{x}_{comp} = (\boldsymbol{x}_{obs}, \boldsymbol{x}_{nobs})$, de tal forma que a função de log-verossimilhança completa é representada por $\ell_c(\boldsymbol{\theta}; \boldsymbol{x}_{comp}) = \log(f(\boldsymbol{x}_{comp}; \boldsymbol{\theta}))$, ou simplesmente, $\ell_c(\boldsymbol{\theta})$.

Na iteração $k+1$ do algoritmo EM consideram-se dois passos:

• **Passo E (Expectation):** Neste passo calcula-se a esperança da log-verossimilhança completa, denotada por $\text{Q}\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}\right)$, condicionada ao vetor de dados observados $\boldsymbol{x}_{obs}$, ou seja,

$$\text{Q}\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}\right) = \mathbb{E}\left\{\ell_c(\boldsymbol{\theta})\Big|\boldsymbol{x}_{obs}, \hat{\boldsymbol{\theta}}^{(k)}\right\},$$

em que $\hat{\boldsymbol{\theta}}^{(k)}$ é a estimativa de $\boldsymbol{\theta}$ obtida na iteração k.

• **Passo M (Maximization):** Neste passo maximiza-se a esperança condicional da log-verossimilhança completa em relação aos parâmetros do modelo, substituindo os dados não observados por seus respectivos valores esperados condicionais obtidos no **Passo E**. Na iteração $k+1$ obtém-se $\hat{\boldsymbol{\theta}}^{(k+1)}$ que maximiza $\mathrm{Q}\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}\right)$.

Estes passos alternam-se até convergir e geralmente os critérios de convergência considerados são $\left\|\hat{\boldsymbol{\theta}}^{(k)} - \hat{\boldsymbol{\theta}}^{(k-1)}\right\| < \epsilon$ ou $\left\|\ell\left(\hat{\boldsymbol{\theta}}^{(k+1)}\right) - \ell\left(\hat{\boldsymbol{\theta}}^{(k)}\right)\right\| < \epsilon$, para um $\epsilon > 0$ suficientemente pequeno. Neste trabalho utilizamos $\epsilon = 10^{-4}$.

A maximização simultânea em termos de todas as componentes do vetor $\boldsymbol{\theta}$ pode ser, em alguns casos, um problema extremamente difícil, mesmo do ponto de vista computacional. Além disso, obter expressões analíticas explicitamente para os valores esperados condicionais do **Passo E** também pode ser uma tarefa complicada. Por essas razões é que existem variações do algoritmo EM às quais serão nomeadas de algoritmos tipo EM, como por exemplo, o algoritmo MCEM, proposto por Wei e Tanner (1990a) em que se obtém os valores esperados condicionais do **Passo E** por meio de integração de Monte Carlo, o algoritmo ECM, proposto por Meng e Rubin (1993), que maximiza as coordenadas de $\boldsymbol{\theta}$ marginalmente e o algoritmo ECME proposto por Liu e Rubin (1994) que é uma extensão do algoritmo ECM e que converge monotonicamente mais rapidamente que este último.

Como a estratégia do algoritmo EM é introduzir uma ou mais variáveis latentes no modelo de modo a "facilitar" de alguma forma a obtenção do EMV, tem-se como consequência ter que calcular uma ou mais esperanças condicionais no Passo E do algoritmo. No caso da classe SMt, devido à sua representação estocástica, o fator de escala U surge quase que naturalmente como uma variável latente no algoritmo EM. Utilizando-se do fato de que a distribuição t-Student possui uma representação estocástica (MASSUIA et al., 2015) dada a seguir, a variável aleatória W aparece como uma outra variável latente.

$$T = W^{-\frac{1}{2}}Z, \quad W \perp Z,$$

em que $T \sim t_\nu\left(0, \sigma^2\right)$, $W \sim Gama\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$, isto é, W possui distribuição Gama com parâmetros $(\frac{\nu}{2}, \frac{\nu}{2})$ e $\mathbb{E}[W] = 1$, $Z \sim \mathrm{N}\left(0, \sigma^2\right)$, isto é, Z possui distribuição normal com média zero e variância σ^2.

Então, considerando que $X \sim \mathrm{SMt}_\nu(\mu, \sigma^2; \boldsymbol{\lambda})$ pode-se escrever o seguinte:

$$X = \mu + U^{-\frac{1}{2}} W^{-\frac{1}{2}} Z, \quad U \perp W \perp Z. \tag{3.14}$$

Como consequência de (3.14) tem-se que $X|U, W \sim \mathrm{N}(\mu, u^{-1}w^{-1}\sigma^2)$ e já é sabido de (3.1) que $X|U \sim t_\nu(\mu, u^{-1}\sigma^2)$. Assim, considerando as variáveis U e W como variáveis latentes e $\boldsymbol{\theta} = (\mu, \sigma^2, \boldsymbol{\lambda})^\top$ como o vetor de parâmetros, obtém-se a função de verossimilhança completa que é dada por

$$\begin{aligned} f_c(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{w}|\boldsymbol{\theta}) &= \prod_{i=1}^{n} f(x_i, u_i, w_i|\boldsymbol{\theta}) \\ &= \prod_{i=1}^{n} f(x_i|u_i, w_i, \mu, \sigma^2) f(w_i|u_i) h(u_i|\boldsymbol{\lambda}) \\ &= \prod_{i=1}^{n} f(x_i|u_i, w_i, \mu, \sigma^2) f(w_i) h(u_i|\boldsymbol{\lambda}) \end{aligned}$$

em que $f(x|u, w, \mu, \sigma^2)$ é a função de densidade de $X|U, W \sim \mathrm{N}(\mu, u^{-1}w^{-1}\sigma^2)$, $f(w)$ é a densidade de $W \sim Gama\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$ e $h(u|\boldsymbol{\lambda})$ é a função densidade de U. Consequentemente, a função de log-verossimilhança completa é

$$\begin{aligned} \ell_c(\boldsymbol{\theta}) &= \log\left[f_c(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{w}|\boldsymbol{\theta})\right] \\ &= \sum_{i=1}^{n} \log[f(x_i|u_i, w_i, \mu, \sigma^2)] + \sum_{i=1}^{n} \log[f(w_i)] + \sum_{i=1}^{n} \log[h(u_i|\boldsymbol{\lambda})] \\ &= \sum_{i=1}^{n} \log\left\{ \frac{u_i^{\frac{1}{2}} w_i^{\frac{1}{2}}}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-u_i w_i}{2\sigma^2}(x_i - \mu)^2\right] \right\} \\ &+ \sum_{i=1}^{n} \log\left[\frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)} w_i^{\frac{\nu}{2}-1} \exp\left(-\frac{\nu}{2} w_i\right) \right] + \sum_{i=1}^{n} \log\left[h(u_i|\boldsymbol{\lambda})\right] \\ &= \sum_{i=1}^{n} \left[\frac{\log u_i}{2} + \frac{\log w_i}{2} - \frac{\log \sigma^2}{2} - \frac{\log 2\pi}{2} - \frac{u_i w_i}{2\sigma^2}(x_i - \mu)^2 \right] \\ &+ \sum_{i=1}^{n} \left[\frac{\nu}{2}\log\left(\frac{\nu}{2}\right) - \log\Gamma\left(\frac{\nu}{2}\right) + \frac{\nu}{2}\log w_i - \log w_i - \frac{\nu}{2} w_i \right] + \sum_{i=1}^{n} \log\left[h(u_i|\boldsymbol{\lambda})\right], \end{aligned}$$

em que $\boldsymbol{x} = (x_1, \cdots, x_n)^\top$ são os dados observados e $\boldsymbol{u} = (u_1, \cdots, u_n)^\top$ e $\boldsymbol{w} = (w_1, \cdots, w_n)^\top$ são os dados não observados.

- **Passo E:**

No passo E do algoritmo EM calcula-se o valor esperado da log-verossimilhança completa condicionada aos dados observados e escreve-se $\mathrm{Q}\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}\right)$, ou seja, na iteração

$(k+1)$ temos que

$$\mathrm{Q}\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}\right) = \mathbb{E}\left\{\ell_c(\boldsymbol{\theta})\middle|\boldsymbol{x}, \hat{\boldsymbol{\theta}}^{(k)}\right\},$$

em que $\hat{\boldsymbol{\theta}}^{(k)}$ é a estimativa de $\boldsymbol{\theta}$ obtida na (k)-ésima iteração.

A menos de uma constante de proporcionalidade que não depende de $\boldsymbol{\theta}$, a função $\ell_c(\boldsymbol{\theta})$ pode ser simplificada de modo que o cálculo da função $\mathrm{Q}\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}\right)$ fica dependente dos valores esperados condicionais $\widehat{\mathbb{E}}^{(k)}_{U_iW_i} = \mathbb{E}\left[U_iW_i|\boldsymbol{x}, \boldsymbol{\theta}^{(k)}\right]$ e $\widehat{\mathbb{E}}^{(k)}_{\log h(U_i)} = \mathbb{E}\left[\log\left\{h(U_i)\right\}|\boldsymbol{x}, \boldsymbol{\theta}^{(k)}\right]$, ou seja,

$$\begin{aligned}\mathrm{Q}\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}\right) &= -\frac{n\log\sigma^2}{2} - \sum_{i=0}^{n}\frac{(x_i-\mu)^2}{2\sigma^2}\mathbb{E}\left[U_iW_i|\boldsymbol{x}, \boldsymbol{\theta}^{(k)}\right] + \sum_{i=1}^{n}\mathbb{E}\left[\log\left\{h(U_i)\right\}|\boldsymbol{x}, \boldsymbol{\theta}^{(k)}\right] \\ &= -\frac{n\log\sigma^2}{2} - \sum_{i=0}^{n}\frac{(x_i-\mu)^2}{2\sigma^2}\widehat{\mathbb{E}}^{(k)}_{U_iW_i} + \sum_{i=1}^{n}\widehat{\mathbb{E}}^{(k)}_{\log h(U_i)}.\end{aligned}$$

• **Passo M:**

No Passo M é obtido o valor $\hat{\boldsymbol{\theta}}^{(k+1)}$ que maximiza a função $\mathrm{Q}\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}\right)$ obtida no **Passo E**, isto é, $\hat{\boldsymbol{\theta}}^{(k+1)} = \mathrm{argmax}_{\boldsymbol{\theta}}\left\{\mathrm{Q}\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}\right)\right\}$. Para isto, derivando $\mathrm{Q}\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}\right)$ em relação a μ e a σ^2 e igualando a zero obtém-se

$$\hat{\mu}^{(k+1)} = \frac{\sum_{i=1}^{n} x_i\widehat{\mathbb{E}}^{(k)}_{U_iW_i}}{\sum_{i=1}^{n}\widehat{\mathbb{E}}^{(k)}_{U_iW_i}} \quad \text{e} \quad \hat{\sigma}^{2^{(k+1)}} = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \hat{\mu}^{(k+1)}\right)^2\widehat{\mathbb{E}}^{(k)}_{U_iW_i},$$

respectivamente. É interessante observar que $\hat{\mu}^{(k+1)}$ e $\hat{\sigma}^{2^{(k+1)}}$ não dependem do valor de $\widehat{\mathbb{E}}^{(k)}_{\log h(U_i)}$. A obtenção do $\hat{\boldsymbol{\lambda}}^{(k+1)}$ vai depender, em cada caso, do fator de escala escolhido.

Estes passos alternam-se até convergir segundo algum critério que, em geral, pode ser $\left\|\hat{\boldsymbol{\theta}}^{(k+1)} - \hat{\boldsymbol{\theta}}^{(k)}\right\| < \epsilon$ ou $\left\|\ell\left(\hat{\boldsymbol{\theta}}^{(k+1)}\right) - \ell\left(\hat{\boldsymbol{\theta}}^{(k)}\right)\right\| < \epsilon$ com $\epsilon > 0$ suficientemente pequeno.

A seguir é mostrado o algoritmo EM para obter o EMV dos parâmetros do modelo nas distribuições t-Student, t-Student contaminada, Slash-t e Weibull-t, todas pertencentes à classe SMt.

t-Student

Quando U é uma variável aleatória degenerada como em (3.4) segue que X tem distribuição t-Student com fdp dada em (3.5), então, a função de log-verossimilhança completa, a menos de uma constante de proporcionalidade que não depende de $\boldsymbol{\theta}$, é dada por

$$\ell_c(\boldsymbol{\theta}) \propto \sum_{i=1}^{n}\log\left\{\sqrt{\left(\frac{w_i}{2\pi\sigma^2}\right)}\exp\left[-\frac{w_i}{\sigma^2}(x_i-\mu)^2\right]\right\}$$

e após algumas manipulações algébricas obtém-se

$$\ell_c(\boldsymbol{\theta}) \propto -\frac{n\log\sigma^2}{2} - \frac{1}{2\sigma^2}\sum_{i=1}^{n} w_i\left(x_i-\mu\right)^2.$$

• **Passo E:** Na iteração $k+1$ do algoritmo EM deve-se calcular a função $\mathrm{Q}\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}\right)$ que é dada por

$$\mathrm{Q}\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}\right) \propto -\frac{n\log\sigma^2}{2} - \frac{1}{2\sigma^2}\sum_{i=1}^{n} \widehat{\mathbb{E}}_{W_i}^{(k)}\left(x_i-\mu\right)^2,$$

em que

$$\widehat{\mathbb{E}}_{W_i}^{(k)} = \left[\frac{\nu+1}{\nu+\frac{\left(x_i-\hat{\mu}^{(k)}\right)^2}{\hat{\sigma}^{2(k)}}}\right].$$

• **Passo M:** Maximiza-se a função Q, obtida no **Passo E**, em relação a $\boldsymbol{\theta}$ e obtém-se

$$\hat{\mu}^{(k+1)} = \frac{\sum\limits_{i=1}^{n} x_i\widehat{\mathbb{E}}_{W_i}^{(k)}}{\sum\limits_{i=1}^{n}\widehat{\mathbb{E}}_{W_i}^{(k)}} \quad \text{e} \quad \hat{\sigma}^{2(k+1)} = \frac{1}{n}\sum_{i=1}^{n}\left(x_i-\hat{\mu}^{(k+1)}\right)^2\widehat{\mathbb{E}}_{W_i}^{(k)}.$$

Observa-se que, neste caso, como U é degenerada, o valor esperado $\widehat{\mathbb{E}}_{U_iW_i}^{(k)}$ reduz-se a $\widehat{\mathbb{E}}_{W_i}^{(k)}$. O processo é iterado até que $\left\|\hat{\boldsymbol{\theta}}^{(k+1)}-\hat{\boldsymbol{\theta}}^{(k)}\right\| < \varepsilon$, para ε suficientemente pequeno.

t-Student Contaminada

Quando U é uma variável aleatória dicotômica como em (3.6) segue que X tem distribuição t-Student contaminada com fdp dada em (3.7) e a função de log-verossimilhança completa, a menos de uma constante de proporcionalidade que não depende de $\boldsymbol{\theta}$, é dada por

$$\begin{aligned}\ell_c(\boldsymbol{\theta}) \propto \frac{1}{1-\gamma}\Bigg\{&\sum_{i=1}^{n}(u_i-\gamma)\left[\log(1-\xi) - \frac{\log\sigma^2}{2} - \frac{w_i}{2\sigma^2}(x_i-\mu)^2\right]\\ &+\sum_{i=1}^{n}(1-u_i)\left[\log\xi + \frac{\log\gamma}{2} - \frac{\log\sigma^2}{2} - \frac{\gamma w_i}{2\sigma^2}(x_i-\mu)^2\right]\Bigg\}\end{aligned}$$

e após algumas manipulações algébricas obtém-se

$$\begin{aligned}\ell_c(\boldsymbol{\theta}) \propto &-\frac{n\log\sigma^2}{2} - \frac{1}{2\sigma^2}\sum_{i=1}^{n}u_iw_i\left(x_i-\mu\right)^2 + \left[\frac{\log\xi}{1-\gamma} + \frac{\log\gamma}{2(1-\gamma)}\right]\left[n-\sum_{i=1}^{n}u_i\right]\\ &+\frac{\log(1-\xi)}{1-\gamma}\left[\sum_{i=1}^{n}u_i - n\gamma\right].\end{aligned}$$

• **Passo E:** Na iteração $k+1$ deve-se calcular a função $\mathrm{Q}\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}\right)$ que é dada por

$$\mathrm{Q}\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}\right) \propto -\frac{n\log\sigma^2}{2} - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\widehat{\mathbb{E}}_{U_iW_i}^{(k)}\left(x_i-\mu\right)^2 + \left[\frac{\log\xi}{1-\gamma} + \frac{\log\gamma}{2(1-\gamma)}\right]\left[n - \sum_{i=1}^{n}\widehat{\mathbb{E}}_{U_i}^{(k)}\right] + \frac{\log(1-\xi)}{1-\gamma}\left[\sum_{i=1}^{n}\widehat{\mathbb{E}}_{U_i}^{(k)} - n\gamma\right],$$

em que

$$\widehat{\mathbb{E}}_{U_i}^{(k)} = \frac{1}{f_{SMt}\left(x_i\middle|\hat{\boldsymbol{\theta}}^{(k)}\right)}\left[\left(1-\hat{\xi}^{(k)}\right)f_\nu\left(x_i\middle|\hat{\mu}^{(k)},\hat{\sigma}^{2^{(k)}}\right) + \hat{\gamma}^{(k)}\hat{\xi}^{(k)}f_\nu\left(x_i\middle|\hat{\mu}^{(k)},\frac{\hat{\sigma}^{2^{(k)}}}{\hat{\gamma}^{(k)}}\right)\right]$$

e

$$\widehat{\mathbb{E}}_{U_iW_i}^{(k)} = \frac{\left(1-\hat{\xi}^{(k)}\right)f_\nu\left(x_i\middle|\hat{\mu}^{(k)},\hat{\sigma}^{2^{(k)}}\right)\left[\frac{\nu+1}{\nu+\frac{\left(x_i-\hat{\mu}^{(k)}\right)^2}{\hat{\sigma}^{2^{(k)}}}}\right] + \hat{\gamma}^{(k)}\hat{\xi}^{(k)}f_\nu\left(x_i\middle|\hat{\mu}^{(k)},\frac{\hat{\sigma}^{2^{(k)}}}{\hat{\gamma}^{(k)}}\right)\left[\frac{\nu+1}{\nu+\hat{\gamma}^{(k)}\frac{\left(x_i-\hat{\mu}^{(k)}\right)^2}{\hat{\sigma}^{2^{(k)}}}}\right]}{f_{SMt}\left(x_i\middle|\hat{\boldsymbol{\theta}}^{(k)}\right)}.$$

Como será visto mais adiante, não há expressões analíticas em forma fechada para obter-se os parâmetros ξ e γ, não sendo necessário o cálculo de $\widehat{\mathbb{E}}_{U_i}^{(k)}$ para este fim, entretanto, este valor será necessário para o cálculo da matriz de informação empírica.

• **Passo M:** Maximiza-se a função Q em relação a μ e σ^2 e obtém-se

$$\hat{\mu}^{(k+1)} = \frac{\sum_{i=1}^{n} x_i\widehat{\mathbb{E}}_{U_iW_i}^{(k)}}{\sum_{i=1}^{n}\widehat{\mathbb{E}}_{U_iW_i}^{(k)}} \quad \text{e} \quad \hat{\sigma}^{2^{(k+1)}} = \frac{1}{n}\sum_{i=1}^{n}\widehat{\mathbb{E}}_{U_iW_i}^{(k)}\left(x_i-\hat{\mu}^{(k+1)}\right)^2.$$

Note que não há expressões analíticas em forma fechada para os parâmetros ξ e γ, por isso, será usado aqui o algoritmo tipo EM, ECME (LIU; RUBIN, 1994), considerando agora que $\boldsymbol{\theta}_1 = (\mu,\sigma^2)^\top$ e $\boldsymbol{\theta}_2 = (\xi,\gamma)^\top$, obtém-se $\hat{\boldsymbol{\theta}}_2^{(k+1)} = \mathrm{argmax}_{\boldsymbol{\theta}_2}\left\{\ell\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}_1^{(k+1)}\right)\right\}$. O processo é iterado até que $\left\|\hat{\boldsymbol{\theta}}^{(k+1)} - \hat{\boldsymbol{\theta}}^{(k)}\right\| < \varepsilon$, para ε suficientemente pequeno.

Weibull-t

Da definição (3.1.1) tem-se que quando $U \sim \text{Weibull}(\lambda,1)$ a v.a. X tem fdp dada em (3.9). Então, a função log-verossimilhança completa, a menos de uma constante de proporcionalidade que não depende de $\boldsymbol{\theta} = (\mu,\sigma^2,\lambda)^\top$, é dada por

$$\ell_c(\boldsymbol{\theta}) \propto -\frac{n\log\sigma^2}{2} - \frac{1}{2\sigma^2}\sum_{i=1}^{n}u_iw_i(x_i-\mu)^2 + n\log\lambda + (\lambda-1)\sum_{i=1}^{n}\log u_i - \sum_{i=1}^{n}u_i^\lambda.$$

• **Passo E:** Na iteração $(k+1)$ deve-se calcular a função $\mathrm{Q}\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}\right)$ que é dada por

$$\mathrm{Q}\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k-1)}\right) = n\left(\log\lambda - \frac{\log\sigma^2}{2}\right) - \sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma^2}\widehat{\mathbb{E}}^{(k)}_{U_iW_i} + (\lambda-1)\sum_{i=1}^{n}\widehat{\mathbb{E}}^{(k)}_{\log U_i} - \sum_{i=1}^{n}\widehat{\mathbb{E}}^{(k)}_{U_i^{\lambda}},$$

em que $\widehat{\mathbb{E}}^{(k)}_{U_iW_i}$, $\widehat{\mathbb{E}}^{(k)}_{\log U_i}$ e $\widehat{\mathbb{E}}^{(k)}_{U_i^{\lambda}}$ não possuem expressão analítica em forma fechada e, por isso, será usado o algoritmo do tipo EM, SAEM, que consiste em obter estes valores esperados condicionais do **Passo E** por meio de uma aproximação estocástica (detalhes sobre o SAEM podem ser obtido em Delyon, Lavielle e Moulines (1999) e Jank (2006) e um breve resumo está no Apêndice B). São introduzidos dois passos intermediários para substituir o Passo E por sua aproximação estocástica usando dados simulados. Portanto, na iteração $k+1$, temos o seguinte:

• **Passo S (*Sampling*):** Seja $\mathbf{U}^{(k+1)} = (U_1^{(k+1)}, \ldots, U_n^{(k+1)})$ o vetor de n observações, em que $U_i^{(k+1)}$ é gerado da distribuição condicional de U dado X, que é dada por

$$f(u|x) = \frac{f_{\nu}\left(x|\mu, \frac{\sigma^2}{u}\right)h(u|\lambda,1)}{f_{SMt}(x|\mu,\sigma^2,\lambda)}, \tag{3.15}$$

em que $f_{SMt}(x|\mu,\sigma^2,\lambda)$ é a fdp da distribuição da classe SMt quando o fator de escala U segue uma distribuição Weibull$(\lambda,1)$. Esta geração foi feita pelo método SIR, *Sampling Importance Resampling* (RUBIN, 1988), cujos detalhes estão no Apêndice C.

• **Passo AE (*Approximation Expectation*):** Uma vez gerada a amostra, é possível obter as aproximações estocásticas para $\widehat{\mathbb{E}}^{(k)}_{U_iW_i}$, $\widehat{\mathbb{E}}^{(k)}_{\log U_i}$ e $\widehat{\mathbb{E}}^{(k)}_{U_i^{\lambda}}$ fazendo

$$\widehat{\mathbb{E}}^{(k+1)}_{U_iW_i} = \widehat{\mathbb{E}}^{(k)}_{U_iW_i} + \kappa_{k+1}\left[\frac{1}{m}\sum_{i=1}^{m}\mathbb{E}[U_i^{k+1}W_i|\boldsymbol{x};\boldsymbol{\theta}^{(k)}] - \widehat{\mathbb{E}}^{(k)}_{U_iW_i}\right]$$

e, de forma similar, obtém-se as aproximações estocásticas de $\widehat{\mathbb{E}}^{(k)}_{\log U_i}$ e $\widehat{\mathbb{E}}^{(k)}_{U_i^{\lambda}}$.

• **Passo M:** Neste passo são obtidos os estimadores de μ, σ^2 e λ que são dadas respectivamente por: $\hat{\mu}^{(k+1)} = \frac{\sum_{i=1}^{n} x_i\widehat{\mathbb{E}}^{(k)}_{U_iW_i}}{\sum_{i=1}^{n}\widehat{\mathbb{E}}^{(k)}_{U_iW_i}}$, $\hat{\sigma}^{2^{(k+1)}} = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \hat{\mu}^{(k+1)}\right)^2\widehat{\mathbb{E}}^{(k)}_{U_iW_i}$ e, uma vez que não há expressão em forma fechada para $\hat{\lambda}^{(k+1)}$, segue que $\hat{\lambda}^{(k+1)} = \mathrm{argmax}_{\lambda}\ell\left(\boldsymbol{\theta}|\hat{\mu}^{(k+1)}, \hat{\sigma}^{2^{(k+1)}}\right)$. O processo é iterado até que $\left\|\hat{\boldsymbol{\theta}}^{(k+1)} - \hat{\boldsymbol{\theta}}^{(k)}\right\| < \varepsilon$, para ε suficientemente pequeno.

Slash-t

Da definição (3.1.1) tem-se que quando $U \sim \mathrm{Beta}(\lambda,1)$ a v.a. X tem distribuição slash-t com fdp dada em (3.8). Então, a função log-verossimilhança completa, a menos

de uma constante de proporcionalidade que não depende de $\boldsymbol{\theta} = (\mu, \sigma^2, \lambda)^\top$, é dada por

$$\ell_c(\boldsymbol{\theta}) \propto -\frac{n \log \sigma^2}{2} - \frac{1}{2\sigma^2} \sum_{i=1}^{n} u_i w_i (x_i - \mu)^2 + n \log \lambda + \lambda \sum_{i=1}^{n} \log u_i.$$

• **Passo E:** Na iteração $k+1$ deve-se calcular a função $\mathrm{Q}\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}\right)$ que é dada por

$$\mathrm{Q}\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k-1)}\right) \propto -\frac{n \log \sigma^2}{2} - \frac{1}{2\sigma^2} \sum_{i=1}^{n} \widehat{\mathbb{E}}^{(k)}_{U_i W_i} (x_i - \mu)^2 + n \log \lambda + \lambda \sum_{i=1}^{n} \widehat{\mathbb{E}}^{(k)}_{\log U_i},$$

em que $\widehat{\mathbb{E}}^{(k)}_{U_i W_i}$ e $\widehat{\mathbb{E}}^{(k)}_{\log U_i}$ também não possuem expressão analítica em forma fechada e, semelhantemente ao caso da Weibull-t, será utilizado o algoritmo SAEM. Em cada iteração é necessário gerar observações da distribuição condicional de U dado X, que é dada por

$$f(u|x) = \frac{f_\nu\left(x|\mu, \frac{\sigma^2}{u}\right) h(u|\lambda, 1)}{f_{SMt}(x|\mu, \sigma^2, \lambda)}, \tag{3.16}$$

em que $f_{SMt}(x|\mu, \sigma^2, \lambda)$ é a fdp da distribuição da classe SMt quando o fator de escala U segue uma distribuição Beta$(\lambda, 1)$. Esta geração também foi feita pelo método SIR, cujos detalhes estão no Apêndice C. Uma vez gerada a amostra, é possível obter as aproximações estocásticas para $\widehat{\mathbb{E}}^{(k)}_{U_i W_i}$ e $\widehat{\mathbb{E}}^{(k)}_{\log U_i}$.

• **Passo M:** Neste passo são obtidas as estimadores de μ, σ^2 e λ que são dadas respectivamente por: $\hat{\mu}^{(k+1)} = \frac{\sum_{i=1}^{n} x_i \widehat{\mathbb{E}}^{(k)}_{U_i W_i}}{\sum_{i=1}^{n} \widehat{\mathbb{E}}^{(k)}_{U_i W_i}}$, $\hat{\sigma}^{2^{(k+1)}} = \frac{1}{n} \sum_{i=1}^{n} \left(x_i - \hat{\mu}^{(k+1)}\right)^2 \widehat{\mathbb{E}}^{(k)}_{U_i W_i}$ e, uma vez que não há expressão em forma fechada para $\hat{\lambda}^{(k+1)}$, segue que $\hat{\lambda}^{(k+1)} = \mathrm{argmax}_\lambda \left\{\ell\left(\boldsymbol{\theta}|\hat{\mu}^{(k+1)}, \hat{\sigma}^{2^{(k+1)}}\right)\right\}$. O processo é iterado até que $\left\|\hat{\boldsymbol{\theta}}^{(k+1)} - \hat{\boldsymbol{\theta}}^{(k)}\right\| < \varepsilon$, para ε suficientemente pequeno.

3.3 Erro Padrão Assintótico

O erro padrão assintótico do EMV foi obtido no caso da maximização direta e do algoritmo EM, uma vez que, satisfeitas as condições de regularidade apresentadas em Casella e Berger (2002, Seção 10.6.2), o EMV é consistente e assintoticamente normal com matriz de covariâncias igual à inversa da matriz de informação esperada de Fisher. Obter a matriz de informação de Fisher esperada, neste caso, é de grande complexidade analítica, deste modo, será usada a matriz de informação observada de Fisher para aproximar os erros padrão (veja Efron e Hinkley (1978) para detalhes a respeito do uso da matriz de informação de Fisher observada).

No caso da MD foi utilizada a matriz de informação observada de Fisher obtida por meio de derivadas numéricas e no caso do algoritmo EM (ou tipo EM) foi utilizada a matriz de informação empírica para se estimar a matriz de informação de Fisher.

3.3.1 Matriz de informação Observada de Fisher

Obter a matriz de informação de esperada Fisher, neste caso, é de grande complexidade analítica. Deste modo, será usada a matriz de informação observada de Fisher $\mathbf{I}(\boldsymbol{\theta})$ com derivadas obtidas numericamente, enquanto que a raiz quadrada dos elementos da diagonal da inversa desta matriz, $\mathbf{I}^{-1}(\boldsymbol{\theta})$, podem ser usados para aproximar os correspondentes erros padrão (veja Efron e Hinkley (1978) para detalhes a respeito do uso da matriz de informação observada de Fisher), em que se $\boldsymbol{\theta} = (\mu, \sigma^2, \boldsymbol{\lambda})^\top$, por exemplo, tem-se

$$\mathbf{I}(\boldsymbol{\theta}) = \begin{bmatrix} \mathrm{I}(\boldsymbol{\theta})_{\mu\mu} & \mathrm{I}(\boldsymbol{\theta})_{\mu\sigma^2} & \mathrm{I}(\boldsymbol{\theta})_{\mu\boldsymbol{\lambda}} \\ \mathrm{I}(\boldsymbol{\theta})_{\sigma^2\mu} & \mathrm{I}(\boldsymbol{\theta})_{\sigma^2\sigma^2} & \mathrm{I}(\boldsymbol{\theta})_{\sigma^2\boldsymbol{\lambda}} \\ \mathrm{I}(\boldsymbol{\theta})_{\boldsymbol{\lambda}\mu} & \mathrm{I}(\boldsymbol{\theta})_{\boldsymbol{\lambda}\sigma^2} & \mathrm{I}(\boldsymbol{\theta})_{\lambda\lambda} \end{bmatrix}, \quad \mathbf{I}^{-1}(\boldsymbol{\theta}) = \begin{bmatrix} \mathrm{I}(\boldsymbol{\theta})^{\mu\mu} & \mathrm{I}(\boldsymbol{\theta})^{\mu\sigma^2} & \mathrm{I}(\boldsymbol{\theta})^{\mu\boldsymbol{\lambda}} \\ \mathrm{I}(\boldsymbol{\theta})^{\sigma^2\mu} & \mathrm{I}(\boldsymbol{\theta})^{\sigma^2\sigma^2} & \mathrm{I}(\boldsymbol{\theta})^{\sigma^2\boldsymbol{\lambda}} \\ \mathrm{I}(\boldsymbol{\theta})^{\boldsymbol{\lambda}\mu} & \mathrm{I}(\boldsymbol{\theta})^{\boldsymbol{\lambda}\sigma^2} & \mathrm{I}(\boldsymbol{\theta})^{\lambda\lambda} \end{bmatrix},$$

$\mathrm{I}(\boldsymbol{\theta})_{\mu\mu} = -\frac{\partial^2\ell(\boldsymbol{\theta};y)}{\partial\mu^2}$, $\mathrm{I}(\boldsymbol{\theta})_{\mu\sigma^2} = \mathrm{I}(\boldsymbol{\theta})_{\sigma^2\mu} = -\frac{\partial^2\ell(\boldsymbol{\theta};y)}{\partial\mu\partial\sigma^2}$, $\mathrm{I}(\boldsymbol{\theta})_{\mu\boldsymbol{\lambda}} = \mathrm{I}(\boldsymbol{\theta})_{\boldsymbol{\lambda}\mu} = -\frac{\partial^2\ell(\boldsymbol{\theta};y)}{\partial\mu\partial\boldsymbol{\lambda}}$, $\mathrm{I}(\boldsymbol{\theta})_{\sigma^2\boldsymbol{\lambda}} = \mathrm{I}(\boldsymbol{\theta})_{\boldsymbol{\lambda}\sigma^2} = -\frac{\partial^2\ell(\boldsymbol{\theta};y)}{\partial\sigma^2\partial\boldsymbol{\lambda}}$, $\mathrm{I}(\boldsymbol{\theta})_{\sigma^2\sigma^2} = -\frac{\partial^2\ell(\boldsymbol{\theta};y)}{\partial{\sigma^2}^2}$ e $\mathrm{I}(\boldsymbol{\theta})_{\boldsymbol{\lambda}\boldsymbol{\lambda}} = -\frac{\partial^2\ell(\boldsymbol{\theta};y)}{\partial\boldsymbol{\lambda}^2}$. Evidentemente que a dimensão de $\mathbf{I}(\boldsymbol{\theta})$ irá depender da dimensão de $\boldsymbol{\lambda}$. No caso da Weibull-t e Slash-t, por exemplo, $\mathbf{I}(\boldsymbol{\theta})$ tem dimensão 3×3, já no caso da t-Contaminada, sua dimensão é 4×4.

Neste trabalho, $\mathbf{I}^{-1}(\widehat{\boldsymbol{\theta}})$ é a inversa de matriz de informação de observada Fisher avaliada em $\widehat{\boldsymbol{\theta}}$, enquanto que $\mathrm{I}(\widehat{\boldsymbol{\theta}})^{\mu\mu}$, $\mathrm{I}(\widehat{\boldsymbol{\theta}})^{\sigma^2\sigma^2}$ e $\mathrm{I}(\widehat{\boldsymbol{\theta}})^{\boldsymbol{\lambda}\boldsymbol{\lambda}}$ são os elementos da sua diagonal principal. Assim, é possível construir intervalos de confiança assintóticos (ICA) para as componentes de $\boldsymbol{\theta} = (\mu, \sigma^2, \boldsymbol{\lambda})^\top$, que são dados por $\left(\hat{\mu} \pm z_{1-\alpha/2}\sqrt{\mathrm{I}(\widehat{\boldsymbol{\theta}})^{\mu\mu}}\right)$, $\left(\hat{\sigma}^2 \pm z_{1-\alpha/2}\sqrt{\mathrm{I}(\widehat{\boldsymbol{\theta}})^{\sigma^2\sigma^2}}\right)$ e $\left(\hat{\boldsymbol{\lambda}} \pm z_{1-\alpha/2}\sqrt{\mathrm{I}(\widehat{\boldsymbol{\theta}})^{\boldsymbol{\lambda}\boldsymbol{\lambda}}}\right)$, em que $z_{1-\alpha/2}$ é o quantil $(1-\alpha/2)$ da distribuição normal padrão. Aqui se comete um abuso de notação, uma vez que $\boldsymbol{\lambda}$ pode ser um escalar ou um vetorial, caso em que a expressão para o seu ICA se desdobrará para cada uma de suas componentes.

Para se obter um ICA para uma função contínua de $\boldsymbol{\theta}$, $d(\boldsymbol{\theta})$, é possível utilizar o método delta e obter a variância de $d(\widehat{\boldsymbol{\theta}})$, ou seja, $\mathrm{Var}[d(\widehat{\boldsymbol{\theta}})] = D^\top(\widehat{\boldsymbol{\theta}})\mathrm{Var}(\widehat{\boldsymbol{\theta}})D(\widehat{\boldsymbol{\theta}})$, em que $\mathrm{Var}(\widehat{\boldsymbol{\theta}}) = \mathbf{I}_n^{-1}(\widehat{\boldsymbol{\theta}})$, $D(\boldsymbol{\theta}) = \frac{\partial d(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = \left(\frac{\partial d(\boldsymbol{\theta})}{\partial\mu} \quad \frac{\partial d(\boldsymbol{\theta})}{\partial\sigma^2} \quad \frac{\partial d(\boldsymbol{\theta})}{\partial\boldsymbol{\lambda}}\right)^\top$ e $D(\widehat{\boldsymbol{\theta}})$ é $D(\boldsymbol{\theta})$ avaliada em $\hat{\boldsymbol{\theta}}$. Portanto, tem-se que o ICA com nível de confiança $1-\alpha/2$ para $d(\boldsymbol{\theta})$ é dado

por $\left(d(\widehat{\boldsymbol{\theta}}) \pm z_{1-\alpha/2}\sqrt{\text{Var}[d(\widehat{\boldsymbol{\theta}})]}\right)$. Desta maneira é possível obter um ICA para $\text{Var}(X)$, $\log[\text{Var}(X)]$, $\log(\sigma^2)$ e $\log(\lambda)$.

3.3.2 Matriz de informação Empírica de Fisher

A partir da log-verossimilhança completa é possível estimar a matriz de informação de Esperada Fisher por meio da matriz de informação empírica de Fisher que é definida por

$$\mathbf{I}_e(\boldsymbol{\theta}|\boldsymbol{x}) = \sum_{i=1}^{n} \mathbf{v}(\boldsymbol{x}_{obs_i}|\boldsymbol{\theta})\mathbf{v}^\top(\boldsymbol{x}_{obs_i}|\boldsymbol{\theta}) - \frac{1}{n}\mathbf{V}(\boldsymbol{x}|\boldsymbol{\theta})\mathbf{V}^\top(\boldsymbol{x}|\boldsymbol{\theta}), \tag{3.17}$$

em que $\boldsymbol{\theta} = (\mu, \sigma^2, \boldsymbol{\lambda})^\top$, $\mathbf{V} = \sum_{i=1}^{n} \mathbf{v}(\boldsymbol{x}_{obs_i}|\boldsymbol{\theta})$ e, como observado por Louis (1982),

$$\mathbf{v}(\boldsymbol{x}_{obs_i}|\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta}; \boldsymbol{x}_{obs_i})}{\partial \boldsymbol{\theta}} = \mathbb{E}\left[\frac{\partial \ell_c(\boldsymbol{\theta}|\boldsymbol{x})}{\partial \boldsymbol{\theta}}\middle| \boldsymbol{x}_{obs_i}, \boldsymbol{\theta}\right].$$

Portanto, substituindo $\boldsymbol{\theta}$ por seu EMV em (3.17), a matriz de informação empírica se reduz a

$$\mathbf{I}_e(\widehat{\boldsymbol{\theta}}|\boldsymbol{x}) = \sum_{i=1}^{n} \widehat{\mathbf{v}}_i\widehat{\mathbf{v}}_i^\top,$$

em que $\widehat{\mathbf{v}}_i = (\widehat{\mathbf{v}}_{\mu_i}, \widehat{\mathbf{v}}_{\sigma_i^2}, \widehat{\mathbf{v}}_{\boldsymbol{\lambda}_i})^\top$ é um vetor de escores individuais e

$$\begin{aligned}
\widehat{\mathbf{v}}_{\mu_i} &= \left(\frac{x_i - \hat{\mu}}{\hat{\sigma}^2}\right)\mathbb{E}_{U_iW_i}, \\
\widehat{\mathbf{v}}_{\sigma_i^2} &= \frac{1}{2}\left[\left(\frac{x_i - \hat{\mu}}{\hat{\sigma}^2}\right)^2 \mathbb{E}_{U_iW_i} - \frac{1}{\hat{\sigma}^2}\right] \text{ e} \\
\widehat{\mathbf{v}}_{\boldsymbol{\lambda}_i} &= \frac{1}{h(u_i|\boldsymbol{\lambda})}\frac{\partial h(u_i|\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}}.
\end{aligned}$$

Na prática, essa matriz de informação empírica é muito utilizada como uma aproximação à matriz de informação observada. Uma discussão mais detalhada a respeito da matriz de informação empírica pode ser encontrada em Louis (1982), McLachlan e Basford (1997, Seção 4.3) e Scott (2002).

A expressão para $\widehat{\mathbf{v}}_{\boldsymbol{\lambda}_i}$ depende do fator de escala U escolhido, então tem-se:

- $U \sim \text{Beta}(\lambda, 1)$ **(Slash-t):** $\widehat{\mathbf{v}}_{\boldsymbol{\lambda}_i} = \frac{1}{\hat{\lambda}} + \mathbb{E}_{\log(U_i)}$;
- $U \sim \text{Weibull}(\lambda, 1)$**(Weibull-t):** $\widehat{\mathbf{v}}_{\boldsymbol{\lambda}_i} = \frac{1}{\hat{\lambda}} + \mathbb{E}_{\log(U_i)} - \mathbb{E}_{U_i^{\hat{\lambda}}\log(U_i)}$;
- U **é dicotômica como em** (3.6)**(t-Contaminada):**

$$\widehat{\mathbf{v}}_{\xi_i} = \frac{1}{1-\hat{\gamma}}\left(\frac{1 - \mathbb{E}_{U_i}}{\hat{\xi}} - \frac{\hat{\gamma} - \mathbb{E}_{U_i}}{1-\hat{\xi}}\right)$$

$$\widehat{\mathbf{v}}_{\gamma_i} = \frac{1-\mathbb{E}_{U_i}}{2(1-\hat{\gamma})^2}\left\{2\left[\log\hat{\xi}+\log\left(1-\hat{\xi}\right)\right]+1+\hat{\gamma}\left(\log\hat{\gamma}-1\right)\right\}.$$

Uma vez obtida uma estimativa da matriz de informação de Fisher, os ICA são obtidos de forma semelhante ao que foi apresentado no caso da maximização direta na Seção (3.3.1).

3.4 Simulação

Nesta seção são apresentados estudos de simulação de Monte Carlo (MC) que foram realizados para avaliar o desempenho dos estimadores dos parâmetros de alguns modelos da classe SMt. As simulações e os gráficos foram feitos no software R (R Core Team, 2018).

Os principais objetivos das simulações foram avaliar as propriedades assintóticas dos estimadores e estudar a consistência dos EMV via MD e algoritmo EM. As estimativas obtidas pelo MM foram utilizadas como valores iniciais no processo iterativo de obtenção do EMV.

Nos cenários simulados nesta seção, o parâmetro $\boldsymbol{\lambda}$ e os graus de liberdade ν foram considerados conhecidos, e foram estimados os parâmetros μ, σ^2 e também a variância de X, $\mathrm{Var}(X)$.

Foram analisados também o logaritmo da variância e de σ^2, com o intuito de que tanto os histogramas quanto os ICA fossem definidos no conjunto dos números reais, dessa forma, foram construídos ICA para μ, $\log(\sigma^2)$ e $\log\left[\mathrm{Var}(X)\right]$.

Seja m o número de réplicas de MC, a estimativa de MC é a média das estimativas obtidas em cada réplica e, além disso, foram calculados o viés relativo (VR), o erro quadrático médio (EQM) e o erro padrão do EMV em cada amostra, sendo que

$$\bar{\hat{\theta}}_{MC} = \frac{1}{m}\sum_{j=1}^{m}\hat{\theta}_j, \quad \mathrm{EP} = \frac{1}{m}\sum_{j=1}^{m}\widehat{EP}_j,$$
$$\mathrm{VR} = \frac{1}{m}\sum_{j=1}^{m}\left[\frac{\hat{\theta}_j-\theta}{\theta}\right] \quad \text{e} \quad \mathrm{EQM} = \frac{1}{m}\sum_{j=1}^{m}\left(\hat{\theta}_j-\theta\right)^2,$$

em que $\hat{\theta}_j$ é a estimativa de θ obtida a partir da amostra gerada na réplica j do processo de simulação de MC e $\widehat{EP}_j$ é a estimativa do erro-padrão correspondente. Para simplificar a notação será utilizado apenas $\hat{\theta}$ para representar tanto as estimativas de θ quanto a média

delas $\overline{\hat{\theta}}_{MC}$. Também foi calculada a taxa de cobertura (TX) do ICA, a taxa à esquerda (TX ESQ) e a taxa à direira (TX DIR).

Para cada uma das distribuições da classe SMt definidas neste trabalho, foram considerados dois cenários de simulação que estão detalhados na Tabela 7, na qual são apresentados os valores dos parâmetros que foram considerados. Os valores dos parâmetros foram escolhidos de modo a se ter o valor da média positivo e negativo e o valor da variância menor do que 3. Foram feitas 5000 réplicas de Monte Carlo em cada caso.

Tabela 7 – Cenários Simulados da Classe SMt

	Cenário 1			
	t-Student	t-Contaminada	Weibull-t	Slash-t
μ	4	4	4	4
σ^2 $\left(\log\sigma^2\right)$	$0,7$ $(-0,3566)$	$0,8$ $(-0,2231)$	$0,7$ $(-0,3566)$	$0,7$ $(-0,3566)$
$\boldsymbol{\lambda}$ (fixado)	–	$(0,2; 0,5)$	$2,2$	3
$\mathrm{Var}(X)$ $(\log[\mathrm{Var}(X)])$	$1,4$ $(0,3365)$	$2,56$ $(0,94)$	$1,90$ (0.6418)	$1,75$ (0.5596)
ν (fixado)	4	4	5	5
	Cenário 2			
	t-Student	t-Contaminada	Weibull-t	Slash-t
μ	-4	-4	-4	-4
σ^2 $\left(\log\sigma^2\right)$	$1,4$ $(0,3365)$	1 (0)	$1,4$ $(0,3365)$	$1,4$ $(0,3365)$
$\boldsymbol{\lambda}$ (fixado)	–	$(0,5; 0,5)$	10	6
$\mathrm{Var}(X)$ $(\log[\mathrm{Var}(X)])$	$2,8$ (1.0296)	3 $(1,0986)$	$2,49$ (0.9136)	$2,8$ (1.0296)
ν (fixado)	4	4	5	5

3.4.1 Gráficos e Tabelas das Simulações

Nesta seção são apresentados os gráficos e tabelas com os resultados das simulações de MC para avaliar o desempenho dos estimadores dos parâmetros de alguns modelos da classe SMt.

A Tabela 8 apresenta os resultados obtidos com as simulações do modelo t-Student. A Figura 4 apresenta o VR e a Figura 5 apresenta o EQM, ambos em função do tamanho da amostra n e, é possível observar um padrão de convergência para zero dessas duas quantidades quando o tamanho da amostra aumenta, indicando que os estimado-

res possuem boas propriedades assintóticas, em particular, o EMV apresentou melhores propriedades assintóticas que o EMM em todos os parâmetros de todos os cenários.

De modo geral, as propriedades assintóticas do EMV não diferem significativamente com o método utilizado para sua obtenção, MD e EM. Nos histogramas apresentados desde a Figura 19 até a Figura 36 e na Tabela 8 é possível observar que para valores pequenos de tamanho de amostra, $n = 15$ e às vezes $n = 30$, o EMM possui VR e EQM relativamente altos quando comparados com os valores do VR e EQM do EMV. Também nos histogramas nota-se que a distribuição dos estimadores se assemelha a uma distribuição normal em todos os casos.

Um padrão de convergência similar a este que foi observado para os estimadores dos parâmetros modelo t-Student também foi observado nos outros modelos simulados, como pode ser visto na Tabela 9, na Figura 6 e na Figura 7 para o modelo t-Contaminada, na Tabela 10, na Figura 8 e na Figura 9 para o modelo Weibull-t e na Tabela 11, na Figura 10 e na Figura 11 para o modelo Slash-t. Os histogramas com a distribuição dos estimadores dos parâmetros destes modelos estão mostrados desde a Figura 37 até a Figura 90.

Observando os valores obtidos para a taxa de cobertura (TX), a taxa à direita (TX DIR) e a taxa à esquerda (TX ESQ), pode-se observar que os ICA para μ são aproximadamente simétricos em todos os casos, com a TX se aproximando da taxa nominal de 95% quando se aumenta o tamanho da amostra e, mesmo com amostras de tamanho $n = 15$, a TX não está distante da taxa nominal. Para o logaritmo do parâmetro σ^2 e da $\mathrm{Var}(X)$ a TX também se aproxima da taxa nominal quando se aumenta o tamanho da amostra, entretanto, para amostras menores, a TX DIR se mostrou muito maior que a TX ESQ, e este padrão pode ser observado também nos histogramas, como por exemplo, na Figura 88. Os valores do EP do EMV obtido pela MD e pelo EM são muito próximos em todos os casos.

De modo geral, considera-se que o desempenho dos estimadores foi satisfatório, tanto com relação à consistência quanto com relação às propriedades assintóticas.

Tabela 8 – Desempenho dos estimadores dos parâmetros μ, σ^2 e da variância $\mathrm{Var}(X)$ do modelo t-Student

CENÁRIO 1 ($\mu = 4$, $\sigma^2 = 0,7$ **e** $\mathrm{Var}(X) = 1,4$)

n	$\hat{\mu}$			VR			EQM			EP		TX ESQ		TX DIR		TX	
	MM	MD	EM	MM	MD	EM	MM	MD	EM	MD	EM	MD	EM	MD	EM	MD	EM
15	3,99916	3,99724	4,00020	−0,00021	−0,00069	0,00005	0,09389	0,06585	0,06780	0,24822	0,25747	3,54%	3,48%	3,82%	3,48%	92,64%	93,04%
30	4,00325	4,00043	4,00403	0,00081	0,00011	0,00101	0,04580	0,03242	0,03250	0,17905	0,18119	2,72%	3,06%	3,00%	2,88%	94,28%	94,06%
50	3,99793	4,00025	3,99818	−0,00052	0,00006	−0,00045	0,02765	0,01990	0,01903	0,13929	0,14042	2,88%	2,34%	2,74%	2,48%	94,38%	95,18%
100	4,00062	4,00129	4,00053	0,00016	0,00032	0,00013	0,01440	0,00998	0,01017	0,09859	0,09890	2,92%	2,96%	2,62%	2,66%	94,46%	94,38%
500	4,00033	4,00018	4,00070	0,00008	0,00005	0,00017	0,00276	0,00199	0,00193	0,04422	0,04427	2,70%	2,36%	2,66%	2,44%	94,64%	95,20%
1000	4,00027	3,99956	4,00023	0,00007	−0,00011	0,00006	0,00139	0,00095	0,00097	0,03132	0,03128	2,20%	2,60%	1,96%	2,52%	95,84%	94,88%
n	$\hat{\sigma}^2$			VR			EQM			EP		TX ESQ		TX DIR		TX	
	MM	MD	EM	MM	MD	EM	MM	MD	EM	MD	EM	MD	EM	MD	EM	MD	EM
15	0,66628	0,68960	0,70531	−0,04817	−0,01486	0,00758	0,45870	0,12737	0,13161	0,33466	0,46614	1,18%	1,32%	5,52%	5,04%	93,30%	93,64%
30	0,67850	0,70193	0,70518	−0,03072	0,00276	0,00740	0,19899	0,06041	0,06239	0,24028	0,32649	1,42%	1,74%	4,30%	3,92%	94,28%	94,34%
50	0,69431	0,70169	0,70352	−0,00813	0,00241	0,00503	0,23455	0,03562	0,03579	0,18578	0,25299	1,76%	1,76%	4,12%	3,32%	94,12%	94,92%
100	0,69094	0,69901	0,69894	−0,01295	−0,00141	−0,00152	0,08615	0,01744	0,01705	0,13091	0,18053	1,68%	1,68%	3,54%	3,34%	94,78%	94,98%
500	0,69782	0,69933	0,70026	−0,00311	−0,00095	0,00037	0,02102	0,00349	0,00337	0,05853	0,08215	2,08%	2,36%	3,18%	2,58%	94,74%	95,06%
1000	0,70021	0,70101	0,69915	0,00030	0,00144	−0,00121	0,03405	0,00182	0,00177	0,04148	0,05833	2,78%	2,08%	3,04%	3,16%	94,18%	94,76%
n	$\widehat{\mathrm{Var}(X)}$			VR			EQM			EP		TX ESQ		TX DIR		TX	
	MM	MD	EM	MM	MD	EM	MM	MD	EM	MD	EM	MD	EM	MD	EM	MD	EM
15	1,33257	1,37920	1,41062	−0,04817	−0,01486	0,00758	1,83478	0,50949	0,52643	0,66931	0,46614	1,18%	1,32%	5,52%	5,04%	93,30%	93,64%
30	1,35700	1,40387	1,41035	−0,03072	0,00276	0,00740	0,79595	0,24163	0,24958	0,48056	0,32649	1,42%	1,74%	4,30%	3,92%	94,28%	94,34%
50	1,38862	1,40338	1,40704	−0,00813	0,00241	0,00503	0,93820	0,14246	0,14315	0,37157	0,25299	1,76%	1,76%	4,12%	3,32%	94,12%	94,92%
100	1,38187	1,39803	1,39788	−0,01295	−0,00141	−0,00152	0,34458	0,06978	0,06819	0,26181	0,18053	1,68%	1,68%	3,54%	3,34%	94,78%	94,98%
500	1,39565	1,39867	1,40052	−0,00311	−0,00095	0,00037	0,08407	0,01395	0,01349	0,11706	0,08215	2,08%	2,36%	3,18%	2,58%	94,74%	95,06%
1000	1,40042	1,40202	1,39831	0,00030	0,00144	−0,00121	0,13622	0,00726	0,00707	0,08295	0,05833	2,78%	2,08%	3,04%	3,16%	94,18%	94,76%

CENÁRIO 2 ($\mu = -4$, $\sigma^2 = 1,4$ **e** $\mathrm{Var}(X) = 2,8$)

n	$\hat{\mu}$			VR			EQM			EP		TX ESQ		TX DIR		TX	
	MM	MD	EM	MM	MD	EM	MM	MD	EM	MD	EM	MD	EM	MD	EM	MD	EM
15	−4,00119	−3,99972	−4,00350	0,00030	−0,00007	0,00088	0,18777	0,13560	0,13181	0,35513	0,36064	3,80%	3,32%	3,82%	3,68%	92,38%	93,00%
30	−3,99540	−3,99513	−3,99430	−0,00115	−0,00122	−0,00142	0,09160	0,06445	0,06500	0,25408	0,25624	2,88%	3,06%	2,52%	2,88%	94,60%	94,06%
50	−4,00292	−4,00228	−4,00257	0,00073	0,00057	0,00064	0,05530	0,03963	0,03806	0,19678	0,19858	2,72%	2,34%	2,92%	2,48%	94,36%	95,18%
100	−3,99912	−3,99933	−3,99925	−0,00022	−0,00017	−0,00019	0,02880	0,02024	0,02034	0,13966	0,13986	2,88%	2,96%	3,12%	2,66%	94,00%	94,38%
500	−3,99953	−3,99919	−3,99901	−0,00012	−0,00020	−0,00025	0,00552	0,00389	0,00386	0,06258	0,06261	2,66%	2,36%	2,42%	2,44%	94,92%	95,20%
1000	−3,99962	−3,99962	−3,99968	−0,00009	−0,00009	−0,00008	0,00277	0,00194	0,00194	0,04424	0,04424	2,76%	2,60%	2,58%	2,52%	94,66%	94,88%
n	$\hat{\sigma}^2$			VR			EQM			EP		TX ESQ		TX DIR		TX	
	MM	MD	EM	MM	MD	EM	MM	MD	EM	MD	EM	MD	EM	MD	EM	MD	EM
15	1,33257	1,41062	1,37896	−0,04817	0,00758	−0,01503	1,83478	0,52642	0,50922	0,68298	0,46584	1,32%	1,30%	5,12%	4,98%	93,56%	93,72%
30	1,35700	1,41285	1,41035	−0,03072	0,00918	0,00740	0,79595	0,25270	0,24958	0,48307	0,32649	1,54%	1,74%	4,24%	3,92%	94,22%	94,34%
50	1,38862	1,40098	1,40703	−0,00813	0,00070	0,00502	0,93820	0,14426	0,14315	0,37100	0,25299	1,56%	1,76%	3,88%	3,32%	94,56%	94,92%
100	1,38187	1,40159	1,39787	−0,01295	0,00113	−0,00152	0,34458	0,06852	0,06819	0,26227	0,18053	1,88%	1,68%	3,12%	3,34%	95,00%	94,98%
500	1,39565	1,40012	1,40052	−0,00311	0,00009	0,00037	0,08407	0,01367	0,01349	0,11713	0,08215	2,28%	2,36%	2,66%	2,58%	95,06%	95,06%
1000	1,40042	1,39845	1,39830	0,00030	−0,00111	−0,00121	0,13622	0,00700	0,00707	0,08273	0,05833	2,16%	2,08%	3,24%	3,16%	94,60%	94,76%
n	$\widehat{\mathrm{Var}(X)}$			VR			EQM			EP		TX ESQ		TX DIR		TX	
	MM	MD	EM	MM	MD	EM	MM	MD	EM	MD	EM	MD	EM	MD	EM	MD	EM
15	2,66514	2,82123	2,75793	−0,04817	0,00758	−0,01503	7,33913	2,10569	2,03687	1,36597	0,46584	1,32%	1,30%	5,12%	4,98%	93,56%	93,72%
30	2,71399	2,82570	2,82071	−0,03072	0,00918	0,00740	3,18380	1,01080	0,99831	0,96614	0,32649	1,54%	1,74%	4,24%	3,92%	94,22%	94,34%
50	2,77725	2,80196	2,81407	−0,00813	0,00070	0,00502	3,75282	0,57706	0,57259	0,74199	0,25299	1,56%	1,76%	3,88%	3,32%	94,56%	94,92%
100	2,76375	2,80317	2,79575	−0,01295	0,00113	−0,00152	1,37834	0,27408	0,27274	0,52453	0,18053	1,88%	1,68%	3,12%	3,34%	95,00%	94,98%
500	2,79130	2,80025	2,80104	−0,00311	0,00009	0,00037	0,33628	0,05468	0,05396	0,23425	0,08215	2,28%	2,36%	2,66%	2,58%	95,06%	95,06%
1000	2,80085	2,79689	2,79661	0,00030	−0,00111	−0,00121	0,54487	0,02799	0,02827	0,16545	0,05833	2,16%	2,08%	3,24%	3,16%	94,60%	94,76%

Tabela 9 – Desempenho dos estimadores dos parâmetros μ, σ^2 e da variância $\mathrm{Var}(X)$ do modelo t-Contaminada

CENÁRIO 1 ($\mu = 4,0$, $\sigma^2 = 0,8$ e $\mathrm{Var}(X) = 2,54$)

n	$\hat{\mu}$			VR			EQM			EP		TX ESQ		TX DIR		TX	
	MM	MD	EM	MM	MD	EM	MM	MD	EM	MD	EM	MD	EM	MD	EM	MD	EM
15	3,99788	3,99891	3,99911	−0,00053	−0,00027	−0,00022	0,16962	0,10081	0,10073	0,30368	0,30855	4,00%	4,04%	3,62%	3,42%	92,38%	92,54%
30	4,00040	4,00086	4,00078	0,00010	0,00022	0,00019	0,08267	0,04786	0,04788	0,21675	0,21822	3,22%	3,16%	2,94%	3,08%	93,84%	93,76%
50	4,00040	4,00008	3,99973	0,00010	0,00002	−0,00007	0,05114	0,02842	0,02857	0,16844	0,16931	2,90%	2,92%	2,66%	2,58%	94,44%	94,50%
100	4,00413	4,00128	4,00126	0,00103	0,00032	0,00031	0,02504	0,01440	0,01396	0,11935	0,11965	2,76%	2,42%	2,90%	2,96%	94,34%	94,62%
500	3,99837	3,99960	3,99960	−0,00041	−0,00010	−0,00010	0,00510	0,00294	0,00293	0,04779	0,05349	2,54%	2,52%	3,08%	2,92%	94,38%	94,56%
1000	4,00103	4,00033	4,00034	0,00026	0,00008	0,00008	0,00254	0,00140	0,00141	0,03786	0,03786	2,30%	2,44%	2,04%	2,36%	95,66%	95,20%
n	$\hat{\sigma}^2$			VR			EQM			EP		TX ESQ		TX DIR		TX	
	MM	MD	EM	MM	MD	EM	MM	MD	EM	MD	EM	MD	EM	MD	EM	MD	EM
15	0,74160	0,80558	0,80537	−0,07300	0,00698	0,00672	0,88263	0,19934	0,19947	0,52248	0,56129	1,20%	1,18%	5,52%	5,34%	93,28%	93,48%
30	0,76717	0,80392	0,80388	−0,04104	0,00490	0,00485	0,47534	0,09663	0,09663	0,36813	0,38043	1,60%	1,60%	4,04%	4,02%	94,36%	94,38%
50	0,79021	0,80227	0,80253	−0,01224	0,00284	0,00316	0,26357	0,05523	0,05650	0,28498	0,29014	1,68%	2,02%	3,98%	3,82%	94,34%	94,16%
100	0,78929	0,79948	0,79971	−0,01338	−0,00065	−0,00036	0,32154	0,02608	0,02588	0,20130	0,20315	1,72%	2,28%	3,36%	3,44%	94,92%	94,28%
500	0,80103	0,79845	0,79843	0,00128	−0,00194	−0,00197	0,16087	0,00524	0,00524	0,08905	0,09015	2,12%	2,06%	3,50%	2,88%	94,38%	95,06%
1000	0,79951	0,79969	0,79964	−0,00061	−0,00038	−0,00045	0,01720	0,00264	0,00265	0,06359	0,06365	2,38%	2,56%	2,68%	2,90%	94,94%	94,54%
n	$\widehat{\mathrm{Var}(X)}$			VR			EQM			EP		TX ESQ		TX DIR		TX	
	MM	MD	EM	MM	MD	EM	MM	MD	EM	MD	EM	MD	EM	MD	EM	MD	EM
15	2,37313	2,57787	2,57719	−0,07300	0,00698	0,00672	9,03814	2,04128	2,04255	0,52248	0,56129	1,20%	1,18%	5,52%	5,34%	93,28%	93,48%
30	2,45494	2,57255	2,57243	−0,04104	0,00490	0,00485	4,86750	0,98954	0,98948	0,36813	0,38043	1,60%	1,60%	4,04%	4,02%	94,36%	94,38%
50	2,52866	2,56726	2,56808	−0,01224	0,00284	0,00316	2,69892	0,56558	0,57855	0,28498	0,29014	1,68%	2,02%	3,98%	3,82%	94,34%	94,16%
100	2,52574	2,55834	2,55907	−0,01338	−0,00065	−0,00036	3,29255	0,26704	0,26502	0,20130	0,20315	1,72%	2,28%	3,36%	3,44%	94,92%	94,28%
500	2,56329	2,55504	2,55496	0,00128	−0,00194	−0,00197	1,64730	0,05371	0,05364	0,08905	0,09015	2,12%	2,06%	3,50%	2,88%	94,38%	95,06%
1000	2,55843	2,55902	2,55885	−0,00061	−0,00038	−0,00045	0,17613	0,02708	0,02718	0,06359	0,06365	2,38%	2,56%	2,68%	2,90%	94,94%	94,54%

CENÁRIO 2 ($\mu = -4,0$, $\sigma^2 = 1,0$ e $\mathrm{Var}(X) = 3,00$)

n	$\hat{\mu}$			VR			EQM			EP		TX ESQ		TX DIR		TX	
	MM	MD	EM	MM	MD	EM	MM	MD	EM	MD	EM	MD	EM	MD	EM	MD	EM
15	−4,00134	−4,00186	−4,00166	0,00034	0,00046	0,00041	0,19997	0,13822	0,13819	0,35474	0,36261	3,94%	3,80%	3,54%	3,20%	92,52%	93,00%
30	−3,99972	−4,00029	−3,99984	−0,00007	0,00007	−0,00004	0,09777	0,06713	0,06535	0,25346	0,25595	3,14%	2,98%	3,00%	2,86%	93,86%	94,16%
50	−3,99889	−4,00027	−4,00040	−0,00028	0,00007	0,00010	0,05989	0,03893	0,03891	0,19706	0,19834	2,70%	2,86%	2,68%	2,82%	94,62%	94,32%
100	−3,99721	−3,99845	−3,99861	−0,00070	−0,00039	−0,00035	0,02913	0,01985	0,01929	0,13978	0,14019	2,62%	2,34%	2,96%	2,74%	94,42%	94,92%
500	−4,00170	−4,00058	−4,00063	0,00042	0,00015	0,00016	0,00601	0,00395	0,00401	0,06270	0,06272	2,40%	2,58%	2,84%	2,90%	94,76%	94,52%
1000	−3,99892	−3,99958	−3,99956	−0,00027	−0,00011	−0,00011	0,00300	0,00194	0,00194	0,04438	0,04439	2,36%	2,54%	2,42%	2,30%	95,22%	95,16%
n	$\hat{\sigma}^2$			VR			EQM			EP		TX ESQ		TX DIR		TX	
	MM	MD	EM	MM	MD	EM	MM	MD	EM	MD	EM	MD	EM	MD	EM	MD	EM
15	0,92443	0,99813	0,99805	−0,07557	−0,00187	−0,00195	0,83095	0,27266	0,27279	0,49762	0,54234	1,08%	1,00%	5,56%	5,28%	93,36%	93,72%
30	0,95865	0,99812	0,99934	−0,04135	−0,00188	−0,00066	0,52018	0,13307	0,13301	0,35119	0,36549	1,76%	1,54%	4,36%	4,52%	93,88%	93,94%
50	0,98774	0,99833	0,99814	−0,01226	−0,00167	−0,00186	0,37155	0,08117	0,07895	0,27199	0,27796	2,02%	2,06%	3,52%	3,64%	94,46%	94,30%
100	0,98490	0,99712	0,99751	−0,01510	−0,00288	−0,00249	0,25022	0,03727	0,03710	0,19217	0,19438	1,86%	1,92%	3,28%	3,34%	94,86%	94,74%
500	0,99917	0,99794	0,99793	−0,00083	−0,00206	−0,00207	0,25914	0,00746	0,00742	0,08589	0,08613	2,12%	2,08%	3,10%	2,90%	94,78%	95,02%
1000	0,99908	0,99931	0,99936	−0,00092	−0,00069	−0,00064	0,01938	0,00374	0,00375	0,06074	0,06081	2,24%	2,48%	2,76%	3,00%	95,00%	94,52%
n	$\widehat{\mathrm{Var}(X)}$			VR			EQM			EP		TX ESQ		TX DIR		TX	
	MM	MD	EM	MM	MD	EM	MM	MD	EM	MD	EM	MD	EM	MD	EM	MD	EM
15	2,77328	2,99439	2,99415	−0,07557	−0,00187	−0,00195	7,47855	2,45396	2,45512	0,49762	0,54234	1,08%	1,00%	5,56%	5,28%	93,36%	93,72%
30	2,87595	2,99436	2,99803	−0,04135	−0,00188	−0,00066	4,68162	1,19761	1,19706	0,35119	0,36549	1,76%	1,54%	4,36%	4,52%	93,88%	93,94%
50	2,96321	2,99500	2,99441	−0,01226	−0,00167	−0,00186	3,34393	0,73056	0,71056	0,27199	0,27796	2,02%	2,06%	3,52%	3,64%	94,46%	94,30%
100	2,95470	2,99135	2,99253	−0,01510	−0,00288	−0,00249	2,25197	0,33543	0,33392	0,19217	0,19438	1,86%	1,92%	3,28%	3,34%	94,86%	94,74%
500	2,99750	2,99381	2,99378	−0,00083	−0,00206	−0,00207	2,33223	0,06712	0,06679	0,08589	0,08613	2,12%	2,08%	3,10%	2,90%	94,78%	95,02%
1000	2,99725	2,99792	2,99808	−0,00092	−0,00069	−0,00064	0,17440	0,03364	0,03371	0,06074	0,06081	2,24%	2,48%	2,76%	3,00%	95,00%	94,52%

Tabela 10 – Desempenho dos estimadores dos parâmetros μ e σ^2 e da variância $\mathrm{Var}(X)$ do modelo Weibull-t

CENÁRIO 1 ($\mu = 4$, $\sigma^2 = 0,7$ e $\mathrm{Var}(X) = 1,9005$)

n	$\hat{\mu}$			VR			EQM			EP		TX ESQ		TX DIR		TX	
	MM	MD	EM	MM	MD	EM	MM	MD	EM	MD	EM	MD	EM	MD	EM	MD	EM
15	3,99769	4,00315	3,99856	−0,00058	0,00079	−0,00036	0,12469	0,08426	0,08419	0,27589	0,28342	4,16%	3,54%	4,01%	4,02%	91,83%	92,44%
30	4,00216	3,99836	3,99853	0,00054	−0,00041	−0,00037	0,06095	0,04146	0,04139	0,19687	0,19958	3,24%	3,01%	3,00%	3,24%	93,76%	93,75%
50	3,99808	3,99969	4,00108	−0,00048	−0,00008	0,00027	0,03862	0,02504	0,02365	0,15377	0,15444	3,10%	2,55%	2,92%	3,16%	93,98%	94,29%
100	3,99968	4,00098	3,99759	−0,00008	0,00025	−0,00060	0,01927	0,01193	0,01224	0,10902	0,10899	3,04%	2,71%	2,22%	3,02%	94,74%	94,27%
500	3,99944	4,00023	4,00281	−0,00014	0,00006	0,00070	0,00382	0,00238	0,00252	0,04877	0,04881	2,48%	3,11%	2,68%	3,11%	94,85%	93,78%
1000	3,99958	4,00032	3,99987	−0,00011	0,00008	−0,00003	0,00185	0,00117	0,00114	0,03450	0,03453	2,50%	2,10%	2,56%	2,36%	94,94%	95,54%
n	$\hat{\sigma}^2$			VR			EQM			EP		TX ESQ		TX DIR		TX	
	MM	MD	EM	MM	MD	EM	MM	MD	EM	MD	EM	MD	EM	MD	EM	MD	EM
15	0,65533	0,70195	0,70721	−0,06381	0,00278	0,01030	0,50993	0,14593	0,14589	0,50612	0,54793	1,25%	1,52%	5,25%	5,00%	93,50%	93,48%
30	0,67484	0,69912	0,70481	−0,03594	−0,00125	0,00687	0,25579	0,06730	0,06787	0,35708	0,37049	1,50%	1,61%	4,60%	4,20%	93,90%	94,20%
50	0,69999	0,70366	0,70233	−0,00001	0,00524	0,00334	0,52573	0,03884	0,03763	0,27617	0,28246	1,40%	1,62%	3,76%	3,36%	94,84%	95,01%
100	0,68985	0,70343	0,69916	−0,01451	0,00491	−0,00120	0,06580	0,01988	0,01888	0,19514	0,19714	2,04%	1,76%	3,10%	3,42%	94,86%	94,82%
500	0,69753	0,70008	0,70079	−0,00352	0,00011	0,00112	0,01999	0,00386	0,00379	0,08708	0,08747	2,96%	2,42%	3,02%	2,94%	94,03%	94,65%
1000	0,70046	0,70017	0,70011	0,00066	0,00024	0,00016	0,00866	0,00193	0,00194	0,06182	0,06178	2,58%	2,36%	2,56%	2,96%	94,86%	94,68%
n	$\widehat{\mathrm{Var}(X)}$			VR			EQM			EP		TX ESQ		TX DIR		TX	
	MM	MD	EM	MM	MD	EM	MM	MD	EM	MD	EM	MD	EM	MD	EM	MD	EM
15	1,77923	1,90578	1,92007	−0,06381	0,00278	0,01030	3,75882	1,07571	1,07537	0,50612	0,54793	1,25%	1,52%	5,25%	5,00%	93,50%	93,48%
30	1,83220	1,89812	1,91356	−0,03594	−0,00125	0,00687	1,88546	0,49610	0,50030	0,35708	0,37049	1,50%	1,61%	4,60%	4,20%	93,90%	94,20%
50	1,90048	1,91045	1,90684	−0,00001	0,00524	0,00334	3,87528	0,28632	0,27737	0,27617	0,28246	1,40%	1,62%	3,76%	3,36%	94,84%	95,01%
100	1,87293	1,90982	1,89822	−0,01451	0,00491	−0,00120	0,48499	0,14655	0,13914	0,19514	0,19714	2,04%	1,76%	3,10%	3,42%	94,86%	94,82%
500	1,89380	1,90070	1,90264	−0,00352	0,00011	0,00112	0,14737	0,02845	0,02792	0,08708	0,08747	2,96%	2,42%	3,02%	2,94%	94,03%	94,65%
1000	1,90174	1,90095	1,90080	0,00066	0,00024	0,00016	0,06385	0,01423	0,01431	0,06182	0,06178	2,58%	2,36%	2,56%	2,96%	94,86%	94,68%

CENÁRIO 2 ($\mu = -4$, $\sigma^2 = 1,4$ e $\mathrm{Var}(X) = 2,4935$)

n	$\hat{\mu}$			VR			EQM			EP		TX ESQ		TX DIR		TX	
	MM	MD	EM	MM	MD	EM	MM	MD	EM	MD	EM	MD	EM	MD	EM	MD	EM
15	−4,00249	−3,99161	−4,00142	0,00062	−0,00210	0,00036	0,16576	0,12927	0,13687	0,35285	0,36536	3,80%	3,28%	3,36%	3,28%	92,84%	93,45%
30	−3,99949	−4,00400	−4,00516	−0,00013	0,00100	0,00129	0,07878	0,06446	0,06565	0,25375	0,25645	2,56%	2,92%	3,44%	2,73%	94,00%	94,35%
50	−4,00104	−3,99686	−3,99724	0,00026	−0,00079	−0,00069	0,05109	0,03969	0,03996	0,19797	0,19856	3,36%	2,98%	2,66%	2,53%	93,98%	94,49%
100	−4,00004	−4,00391	−4,00318	0,00001	0,00098	0,00080	0,02543	0,02026	0,01993	0,14007	0,14078	2,58%	2,83%	3,14%	2,77%	94,28%	94,40%
500	−4,00043	−3,99860	−3,99859	0,00011	−0,00035	−0,00035	0,00503	0,00397	0,00391	0,06283	0,06291	2,22%	2,51%	2,44%	2,28%	95,34%	95,20%
1000	−4,00057	−4,00045	−4,00035	0,00014	0,00011	0,00009	0,00241	0,00199	0,00198	0,04446	0,04445	2,46%	2,20%	2,62%	2,47%	94,92%	95,33%
n	$\hat{\sigma}^2$			VR			EQM			EP		TX ESQ		TX DIR		TX	
	MM	MD	EM	MM	MD	EM	MM	MD	EM	MD	EM	MD	EM	MD	EM	MD	EM
15	1,30832	1,37834	1,38934	−0,06548	−0,01547	−0,00762	1,19821	0,46038	0,47571	0,46494	0,51637	1,24%	1,19%	5,68%	5,53%	93,08%	93,29%
30	1,35053	1,39465	1,38634	−0,03534	−0,00382	−0,00976	0,39972	0,21717	0,21754	0,32807	0,34487	1,38%	1,30%	4,38%	4,48%	94,24%	94,22%
50	1,38801	1,40378	1,39300	−0,00857	0,00270	−0,00500	0,34083	0,13129	0,13089	0,25415	0,26202	1,84%	1,70%	3,60%	3,56%	94,56%	94,74%
100	1,38364	1,39706	1,40327	−0,01169	−0,00210	0,00234	0,12764	0,06640	0,06564	0,17964	0,18286	2,34%	2,20%	3,32%	3,22%	94,34%	94,58%
500	1,40009	1,39862	1,40047	0,00006	−0,00099	0,00034	0,04111	0,01308	0,01318	0,08034	0,08062	2,58%	2,23%	2,96%	3,31%	94,46%	94,46%
1000	1,40190	1,39977	1,39842	0,00135	−0,00016	−0,00113	0,01527	0,00651	0,00657	0,05682	0,05694	2,48%	2,30%	2,92%	2,93%	94,60%	94,77%
n	$\widehat{\mathrm{Var}(X)}$			VR			EQM			EP		TX ESQ		TX DIR		TX	
	MM	MD	EM	MM	MD	EM	MM	MD	EM	MD	EM	MD	EM	MD	EM	MD	EM
15	2,33019	2,45489	2,47447	−0,06548	−0,01547	−0,00762	3,80089	1,46039	1,50902	0,46494	0,51637	1,24%	1,19%	5,68%	5,53%	93,08%	93,29%
30	2,40535	2,48394	2,46913	−0,03534	−0,00382	−0,00976	1,26796	0,68889	0,69008	0,32807	0,34487	1,38%	1,30%	4,38%	4,48%	94,24%	94,22%
50	2,47211	2,50019	2,48100	−0,00857	0,00270	−0,00500	1,08116	0,41647	0,41519	0,25415	0,26202	1,84%	1,70%	3,60%	3,56%	94,56%	94,74%
100	2,46433	2,48824	2,49929	−0,01169	−0,00210	0,00234	0,40490	0,21063	0,20823	0,17964	0,18286	2,34%	2,20%	3,32%	3,22%	94,34%	94,58%
500	2,49362	2,49101	2,49430	0,00006	−0,00099	0,00034	0,13041	0,04148	0,04182	0,08034	0,08062	2,58%	2,23%	2,96%	3,31%	94,46%	94,46%
1000	2,49684	2,49306	2,49065	0,00135	−0,00016	−0,00113	0,04844	0,02064	0,02084	0,05682	0,05694	2,48%	2,30%	2,92%	2,93%	94,60%	94,77%

Tabela 11 – Desempenho dos estimadores dos parâmetros μ e σ^2 e da variância $\mathrm{Var}(X)$ do modelo Slash-t

CENÁRIO 1 ($\mu = 4$, $\sigma^2 = 0,7$ **e** $\mathrm{Var}(X) = 1,75$)																	
n	$\hat{\mu}$			**VR**			**EQM**			**EP**		**TX ESQ**		**TX DIR**		**TX**	
	MM	**MD**	**EM**	**MM**	**MD**	**EM**	**MM**	**MD**	**EM**	**MD**	**EM**	**MD**	**EM**	**MD**	**EM**	**MD**	**EM**
15	3,99760	3,99664	3,99400	−0,00060	−0,00084	−0,00150	0,11423	0,08555	0,09018	0,28493	0,29763	3,56%	3,36%	3,80%	3,64%	92,64%	93,00%
30	4,00243	4,00038	4,00241	0,00061	0,00010	0,00060	0,05737	0,04311	0,04291	0,20664	0,20878	3,12%	2,90%	3,00%	2,54%	93,88%	94,56%
50	3,99976	3,99882	4,00057	−0,00006	−0,00030	0,00014	0,03570	0,02649	0,02658	0,16023	0,16170	2,72%	3,08%	3,06%	2,84%	94,22%	94,08%
100	4,00123	4,00146	3,99966	0,00031	0,00037	−0,00008	0,01774	0,01328	0,01315	0,11335	0,11419	2,92%	2,48%	2,82%	2,66%	94,26%	94,86%
500	3,99941	3,99960	4,00000	−0,00015	−0,00010	0,00000	0,00347	0,00255	0,00261	0,05097	0,05106	2,14%	2,68%	2,64%	2,54%	95,22%	94,78%
1000	3,99962	3,99930	4,00025	−0,00010	−0,00017	0,00006	0,00177	0,00129	0,00132	0,03600	0,03608	2,54%	2,65%	2,48%	2,00%	94,98%	95,36%
n	$\hat{\sigma}^2$			**VR**			**EQM**			**EP**		**TX ESQ**		**TX DIR**		**TX**	
	MM	**MD**	**EM**	**MM**	**MD**	**EM**	**MM**	**MD**	**EM**	**MD**	**EM**	**MD**	**EM**	**MD**	**EM**	**MD**	**EM**
15	0,64046	0,68367	0,69823	−0,08506	−0,02332	−0,00252	0,24435	0,11536	0,12551	0,47779	0,52843	1,06%	1,30%	5,56%	4,54%	93,38%	94,16%
30	0,68197	0,70517	0,69889	−0,02576	0,00738	−0,00158	0,15739	0,06004	0,05836	0,33734	0,35657	1,72%	1,38%	4,44%	3,88%	93,84%	94,74%
50	0,68597	0,70065	0,69841	−0,02004	0,00093	−0,00228	0,08609	0,03507	0,03480	0,26122	0,27006	1,66%	1,90%	3,48%	4,08%	94,86%	94,02%
100	0,69269	0,69759	0,69747	−0,01044	−0,00345	−0,00361	0,04274	0,01709	0,01703	0,18482	0,18785	1,88%	1,66%	3,92%	3,58%	94,20%	94,76%
500	0,70319	0,70149	0,69993	0,00456	0,00213	−0,00010	0,02035	0,00337	0,00328	0,08260	0,08308	2,44%	2,10%	2,96%	2,78%	94,60%	95,12%
1000	0,69846	0,69919	0,69956	−0,00220	−0,00116	−0,00062	0,00462	0,00164	0,00165	0,05841	0,05864	2,00%	2,18%	2,74%	2,58%	95,26%	95,23%
n	$\widehat{\mathrm{Var}(X)}$			**VR**			**EQM**			**EP**		**TX ESQ**		**TX DIR**		**TX**	
	MM	**MD**	**EM**	**MM**	**MD**	**EM**	**MM**	**MD**	**EM**	**MD**	**EM**	**MD**	**EM**	**MD**	**EM**	**MD**	**EM**
15	1,60115	1,70918	1,74559	−0,08506	−0,02332	−0,00252	1,52717	0,72099	0,78446	0,47779	0,52843	1,06%	1,30%	5,56%	4,54%	93,38%	94,16%
30	1,70492	1,76292	1,74724	−0,02576	0,00738	−0,00158	0,98369	0,37523	0,36474	0,33734	0,35657	1,72%	1,38%	4,44%	3,88%	93,84%	94,74%
50	1,71493	1,75162	1,74602	−0,02004	0,00093	−0,00228	0,53807	0,21919	0,21748	0,26122	0,27006	1,66%	1,90%	3,48%	4,08%	94,86%	94,02%
100	1,73172	1,74397	1,74368	−0,01044	−0,00345	−0,00361	0,26715	0,10680	0,10646	0,18482	0,18785	1,88%	1,66%	3,92%	3,58%	94,20%	94,76%
500	1,75797	1,75374	1,74982	0,00456	0,00213	−0,00010	0,12717	0,02109	0,02051	0,08260	0,08308	2,44%	2,10%	2,96%	2,78%	94,60%	95,12%
1000	1,74615	1,74797	1,74891	−0,00220	−0,00116	−0,00062	0,02890	0,01026	0,01031	0,05841	0,05864	2,00%	2,18%	2,74%	2,58%	95,26%	95,23%
CENÁRIO 2 ($\mu = -4$, $\sigma^2 = 1,4$ **e** $\mathrm{Var}(X) = 2,8$)																	
n	$\hat{\mu}$			**VR**			**EQM**			**EP**		**TX ESQ**		**TX DIR**		**TX**	
	MM	**MD**	**EM**	**MM**	**MD**	**EM**	**MM**	**MD**	**EM**	**MD**	**EM**	**MD**	**EM**	**MD**	**EM**	**MD**	**EM**
15	−4,01269	−4,01354	−4,00222	0,00317	0,00338	0,00056	0,18811	0,15234	0,14975	0,37227	0,38683	3,44%	3,46%	4,20%	3,22%	92,36%	93,32%
30	−3,99884	−4,00156	−3,99685	−0,00029	0,00039	−0,00079	0,09195	0,07352	0,07343	0,26961	0,27168	2,74%	2,98%	3,24%	3,14%	94,02%	93,88%
50	−3,99795	−3,99943	−4,00367	−0,00051	−0,00014	0,00092	0,05692	0,04528	0,04521	0,20853	0,21030	2,78%	2,72%	3,12%	3,06%	94,10%	94,22%
100	−4,00061	−4,00049	−4,00187	0,00015	0,00012	0,00047	0,02901	0,02272	0,02242	0,14775	0,14867	2,82%	2,71%	3,00%	2,26%	94,18%	95,04%
500	−4,00010	−3,99985	−3,99998	0,00003	−0,00004	−0,00001	0,00562	0,00435	0,00442	0,06643	0,06646	2,46%	2,57%	2,36%	2,75%	95,18%	94,68%
1000	−3,99991	−4,00035	−4,00123	−0,00002	0,00009	0,00031	0,00281	0,00213	0,00218	0,04693	0,04701	2,48%	2,53%	2,34%	2,34%	95,18%	95,14%
n	$\hat{\sigma}^2$			**VR**			**EQM**			**EP**		**TX ESQ**		**TX DIR**		**TX**	
	MM	**MD**	**EM**	**MM**	**MD**	**EM**	**MM**	**MD**	**EM**	**MD**	**EM**	**MD**	**EM**	**MD**	**EM**	**MD**	**EM**
15	1,30266	1,37390	1,38672	−0,06953	−0,01865	−0,00948	1,12935	0,45450	0,47575	0,46602	0,51807	1,10%	1,20%	5,58%	5,48%	93,32%	93,32%
30	1,36362	1,41354	1,38996	−0,02598	0,00967	−0,00717	0,52559	0,23677	0,22715	0,32908	0,34668	1,82%	1,44%	4,36%	4,02%	93,82%	94,54%
50	1,37265	1,39602	1,39198	−0,01953	−0,00285	−0,00573	0,34050	0,12558	0,13303	0,25502	0,26280	1,66%	1,70%	3,24%	4,02%	95,10%	94,28%
100	1,38912	1,39531	1,39618	−0,00777	−0,00335	−0,00273	0,17149	0,06621	0,06633	0,18042	0,18319	1,72%	1,94%	3,60%	3,75%	94,68%	94,31%
500	1,40202	1,40249	1,38328	0,00145	0,00178	−0,01194	0,03874	0,01293	0,01435	0,08064	0,08095	2,38%	2,05%	2,86%	3,19%	94,76%	94,76%
1000	1,40016	1,39943	1,40089	0,00011	−0,00040	0,00063	0,01635	0,00637	0,00671	0,05702	0,05722	2,14%	2,00%	2,86%	3,05%	95,00%	94,95%
n	$\widehat{\mathrm{Var}(X)}$			**VR**			**EQM**			**EP**		**TX ESQ**		**TX DIR**		**TX**	
	MM	**MD**	**EM**	**MM**	**MD**	**EM**	**MM**	**MD**	**EM**	**MD**	**EM**	**MD**	**EM**	**MD**	**EM**	**MD**	**EM**
15	2,60532	2,74779	2,77345	−0,06953	−0,01865	−0,00948	4,51739	1,81800	1,90298	0,46602	0,51807	1,10%	1,20%	5,58%	5,48%	93,32%	93,32%
30	2,72724	2,82708	2,77993	−0,02598	0,00967	−0,00717	2,10236	0,94707	0,90860	0,32908	0,34668	1,82%	1,44%	4,36%	4,02%	93,82%	94,54%
50	2,74530	2,79203	2,78396	−0,01953	−0,00285	−0,00573	1,36199	0,50232	0,53212	0,25502	0,26280	1,66%	1,70%	3,24%	4,02%	95,10%	94,28%
100	2,77823	2,79063	2,79237	−0,00777	−0,00335	−0,00273	0,68596	0,26485	0,26532	0,18042	0,18319	1,72%	1,94%	3,60%	3,75%	94,68%	94,31%
500	2,80405	2,80499	2,76656	0,00145	0,00178	−0,01194	0,15496	0,05171	0,05741	0,08064	0,08095	2,38%	2,05%	2,86%	3,19%	94,76%	94,76%
1000	2,80031	2,79887	2,80178	0,00011	−0,00040	0,00063	0,06541	0,02549	0,02683	0,05702	0,05722	2,14%	2,00%	2,86%	3,05%	95,00%	94,95%

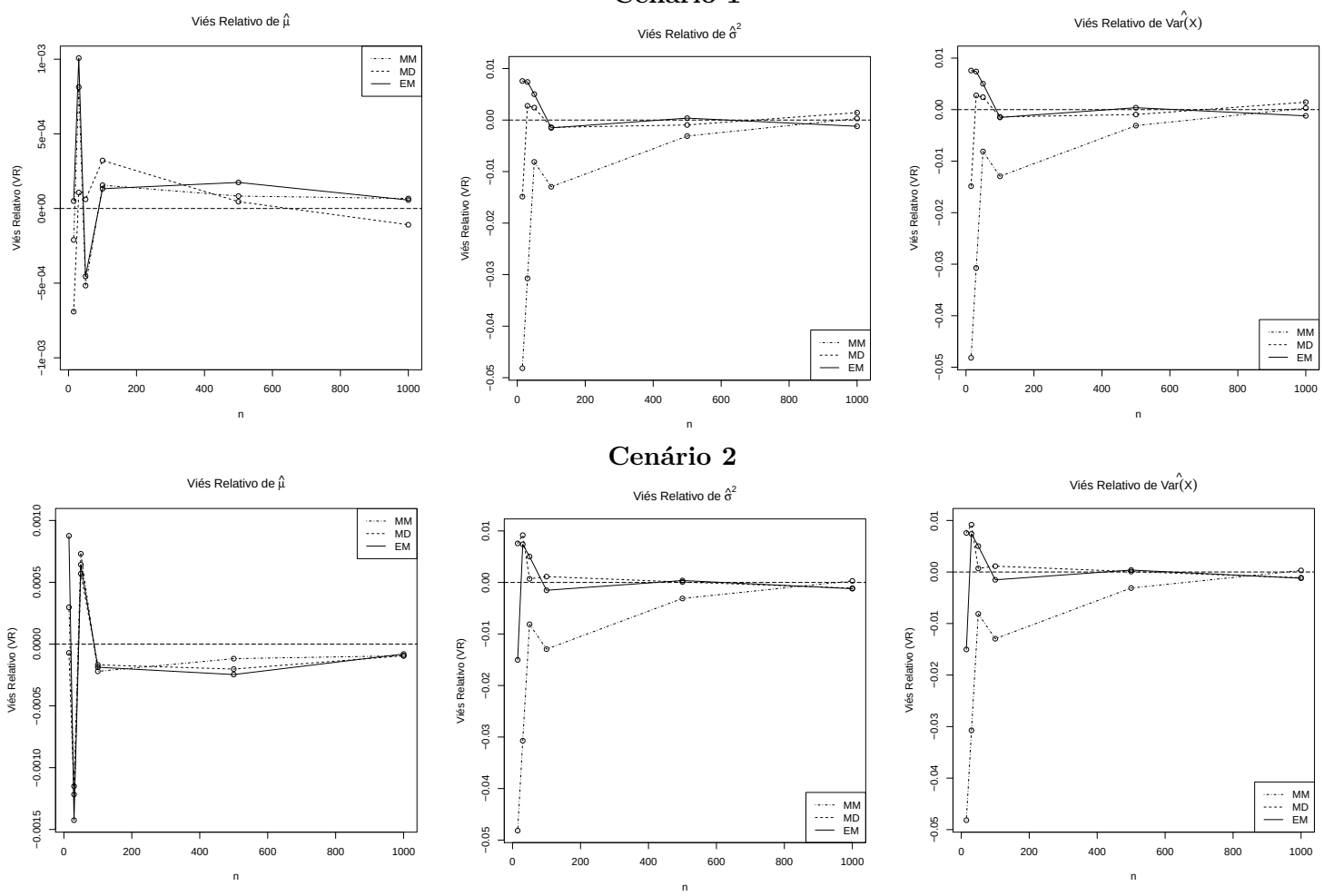


Figura 4 – Viés Relativo (VR) das estimativas dos parâmetros do modelo t-Student

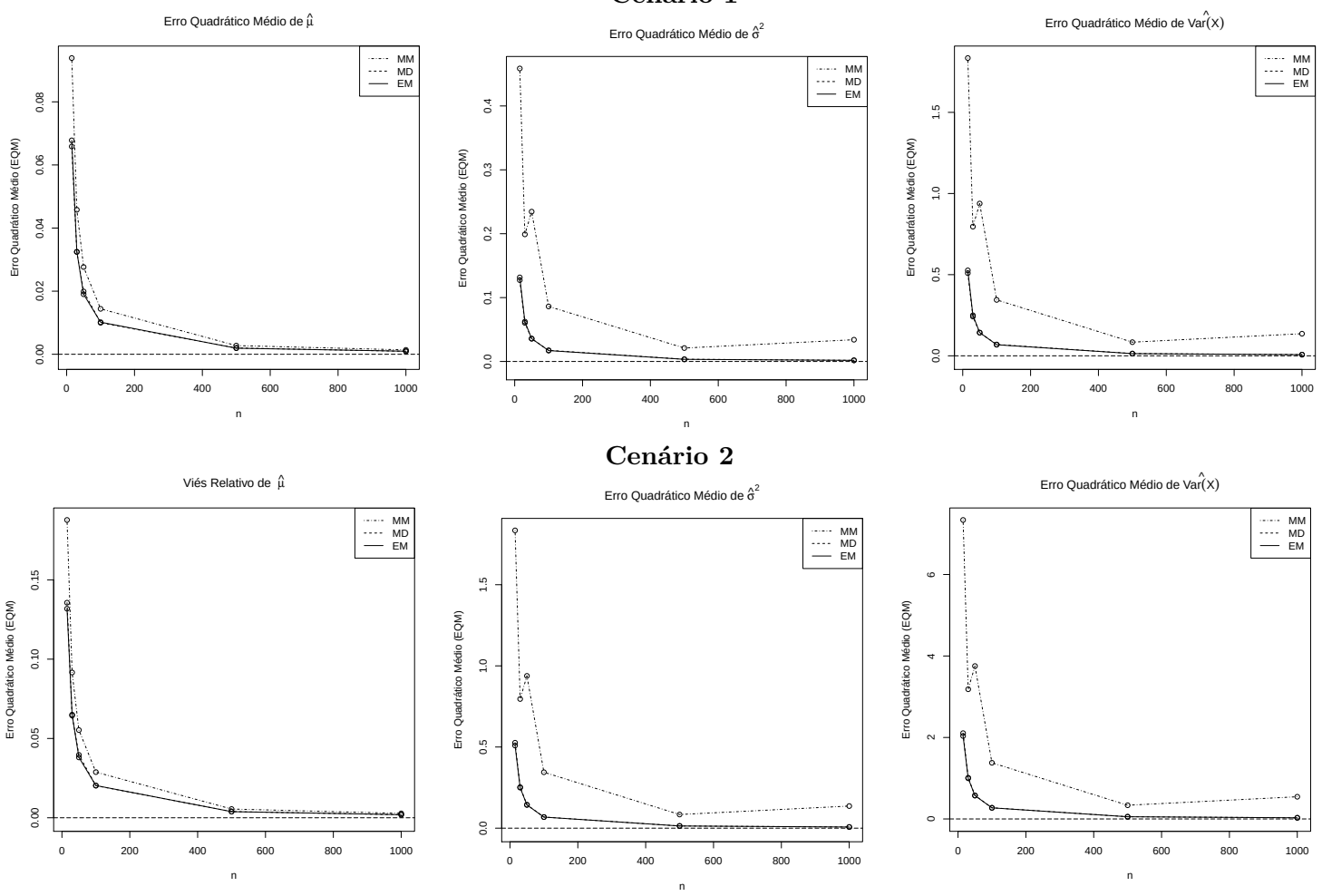


Figura 5 – Erro Quadrático Médio (EQM) das estimativas dos parâmetros do modelo t-Student

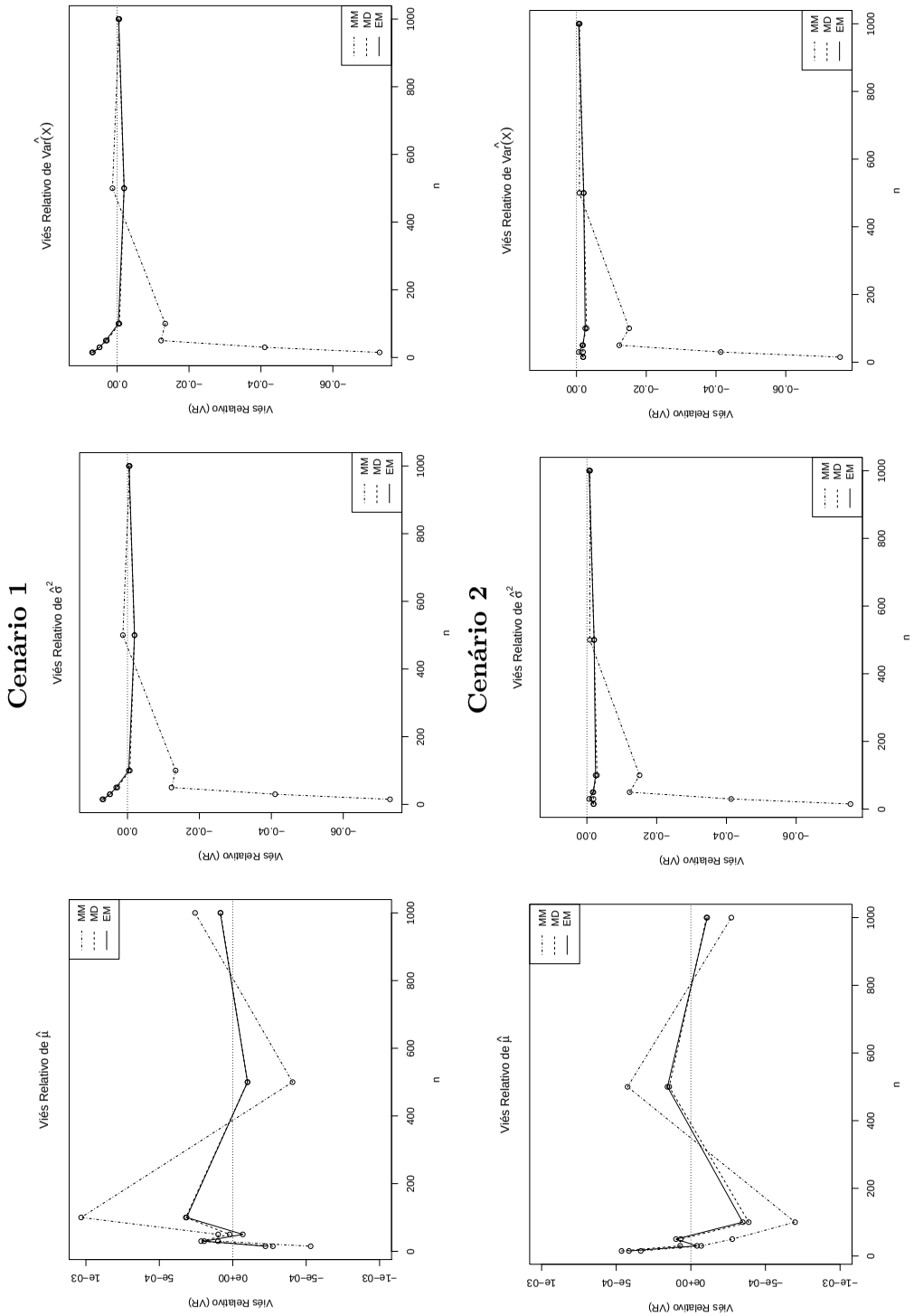


Figura 6 – Viés Relativo (VR) das estimativas dos parâmetros do modelo t-Contaminada

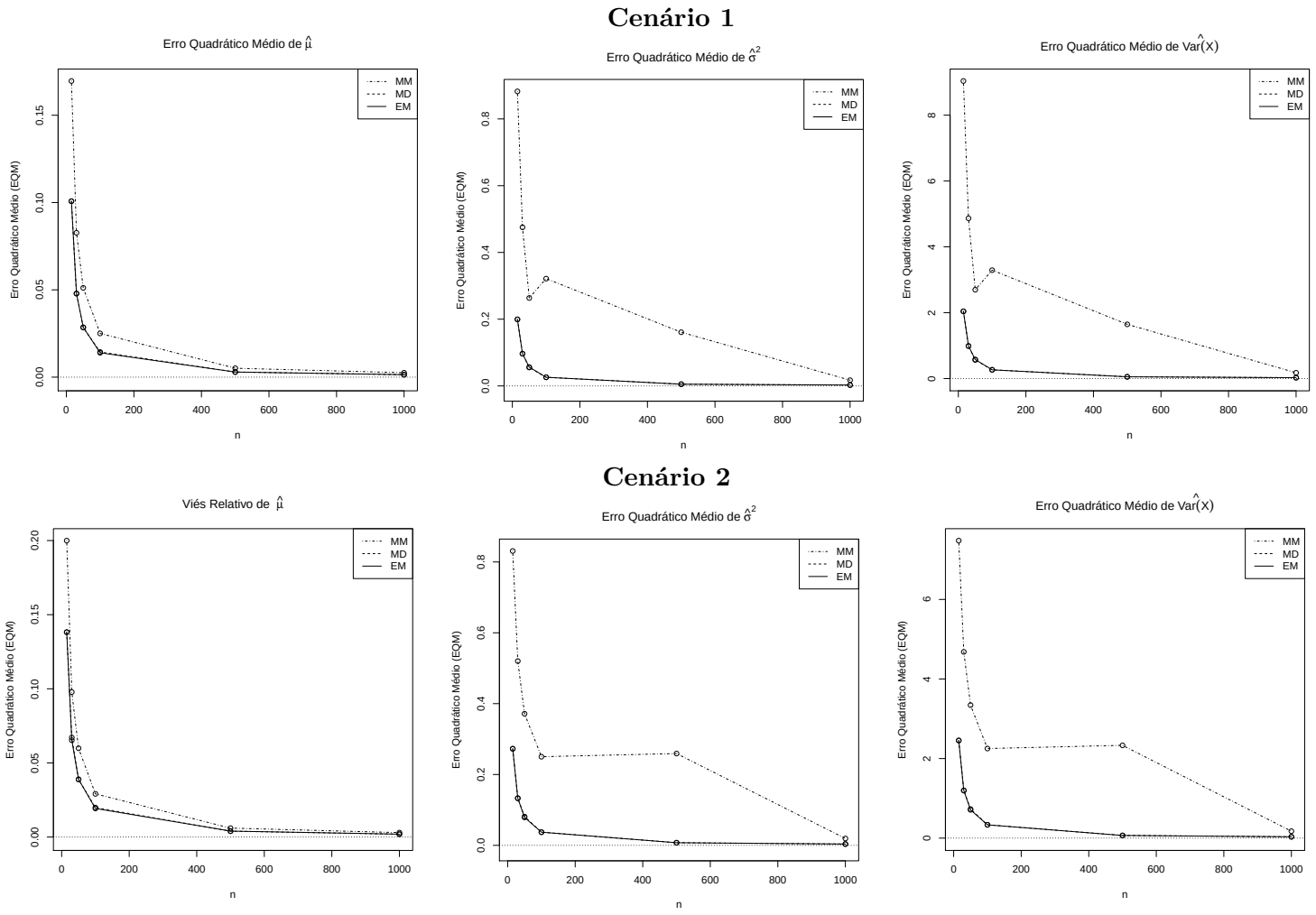


Figura 7 – Erro Quadrático Médio (EQM) das estimativas dos parâmetros do modelo t-Contaminada

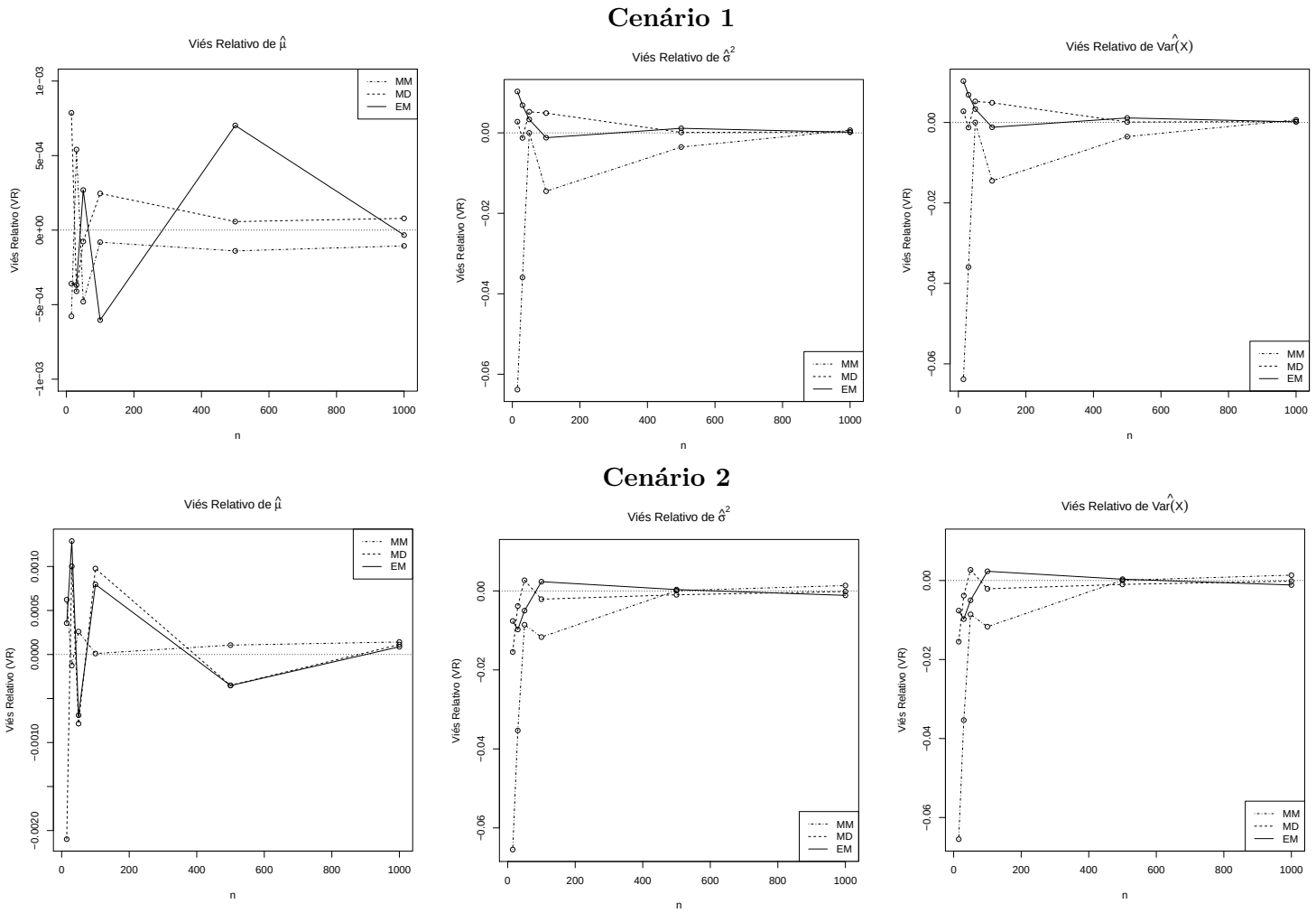


Figura 8 – Viés Relativo (VR) das estimativas dos parâmetros do modelo Weibull-t

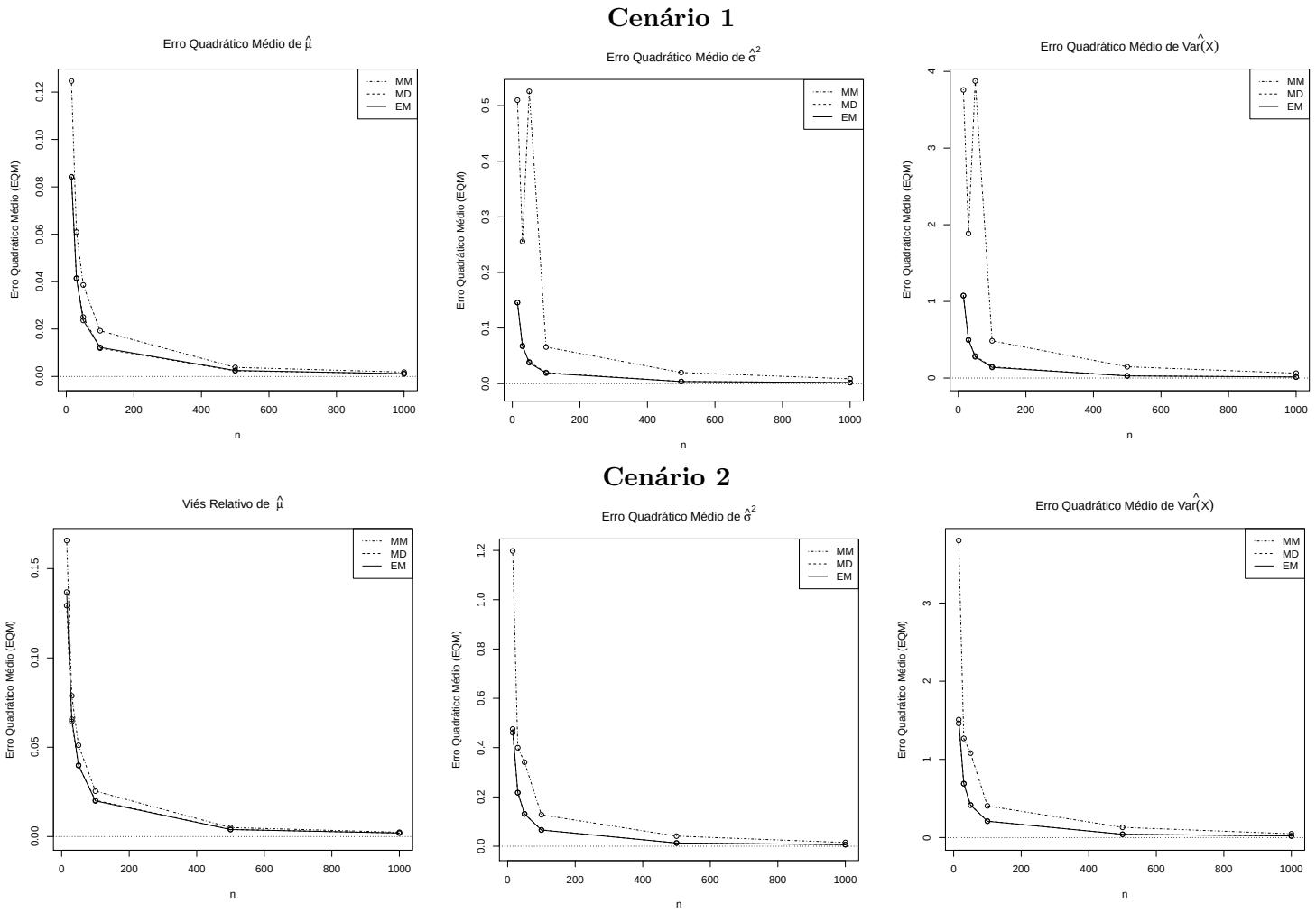


Figura 9 – Erro Quadrático Médio (EQM) das estimativas dos parâmetros do modelo Weibull-t

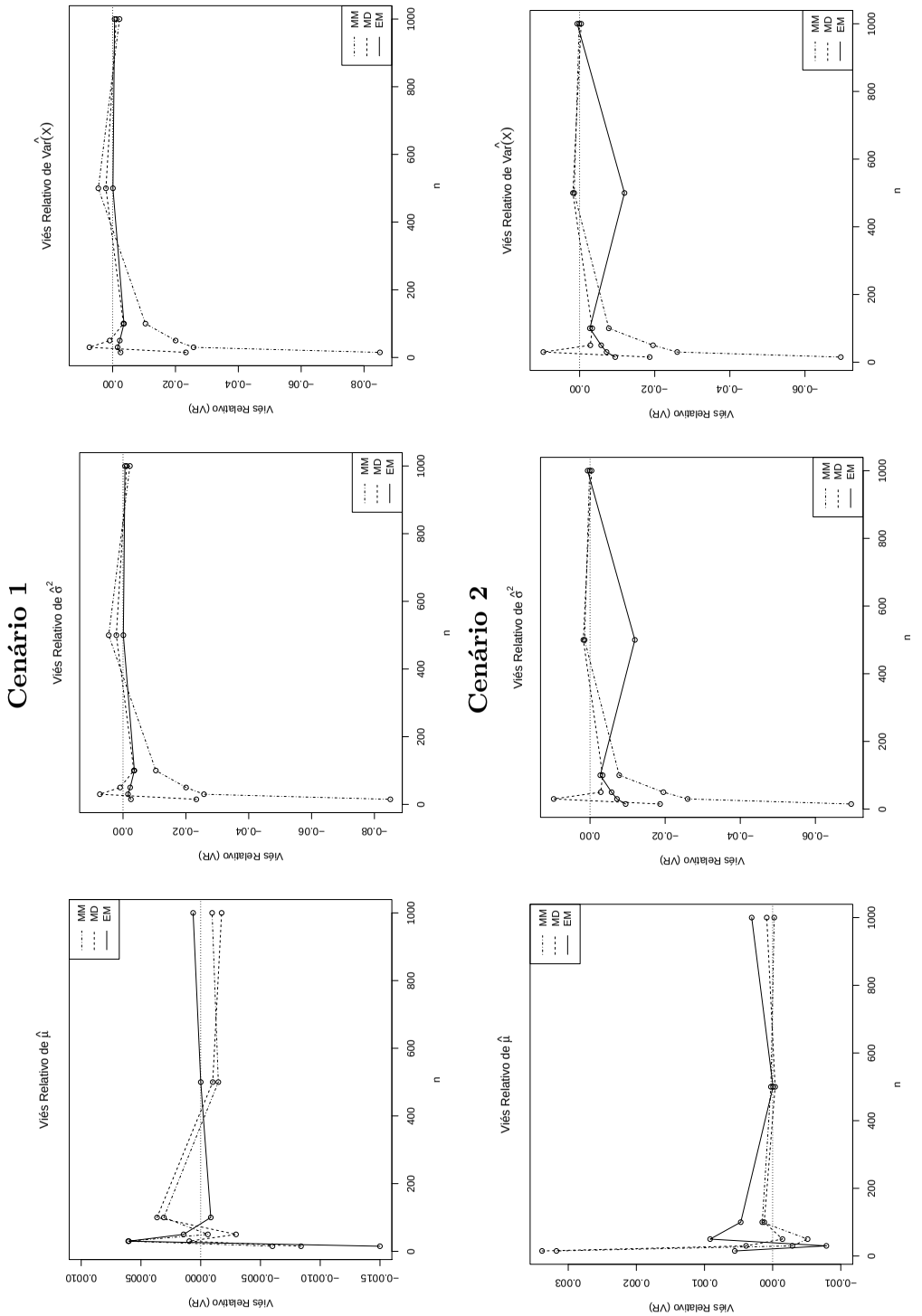


Figura 10 – Viés Relativo (VR) das estimativas dos parâmetros do modelo Slash-t

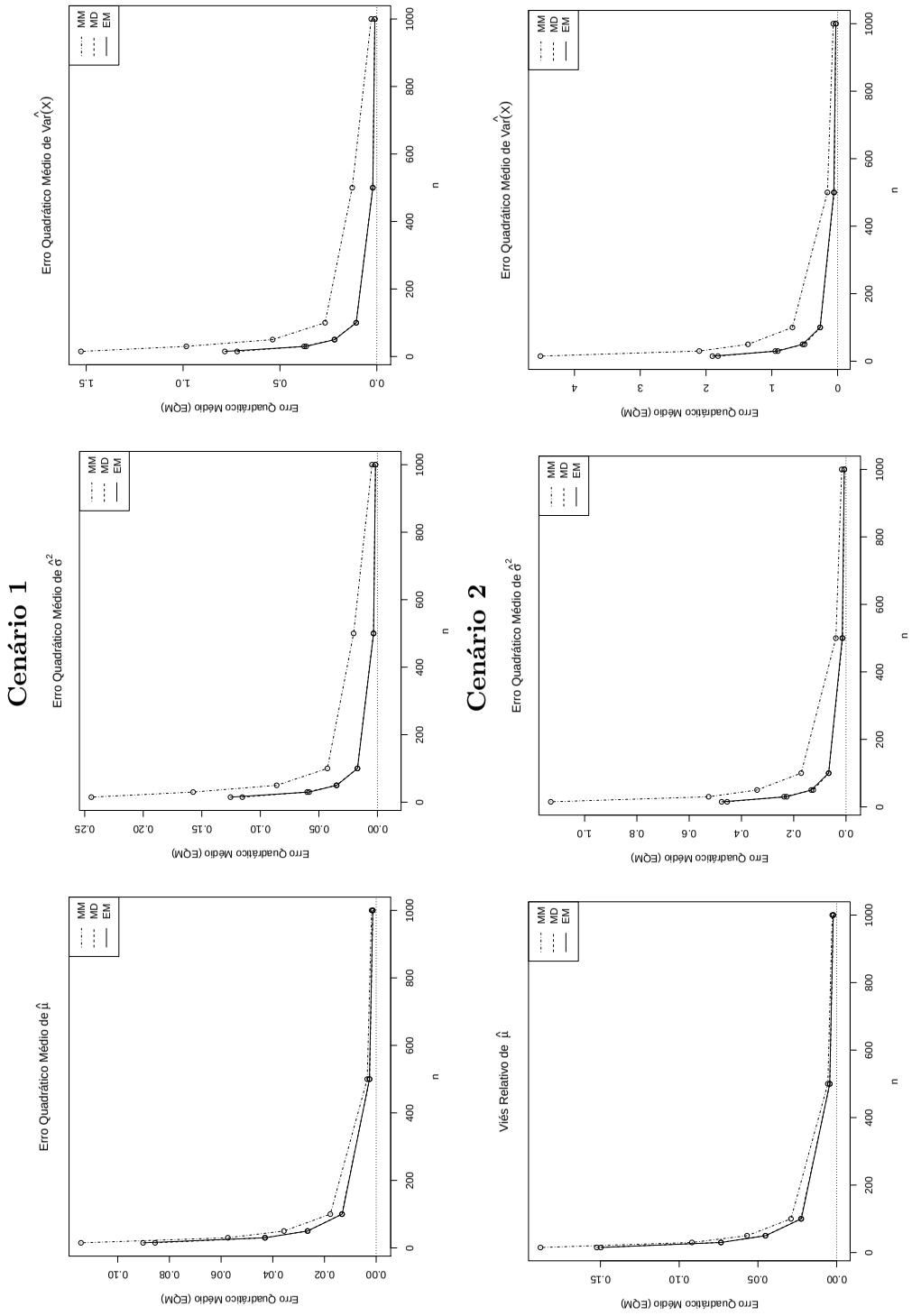


Figura 11 – Erro Quadrático Médio (EQM) das estimativas dos parâmetros do modelo Slash-t

3.5 Aplicação

Nesta seção é apresentada uma aplicação dos resultados obtidos nas seções anteriores deste capítulo utilizando o conjunto de dados da PNAD (IBGE, 2015) relativos à renda dos indivíduos com idade a partir de 15 anos, residentes no estado de Roraima no ano de 2015 divulgados pelo Instituto Brasileiro de Geografia e Estatística IBGE (2019). A Figura 12 mostra o histograma destes dados em que é possível observar indícios de simetria.

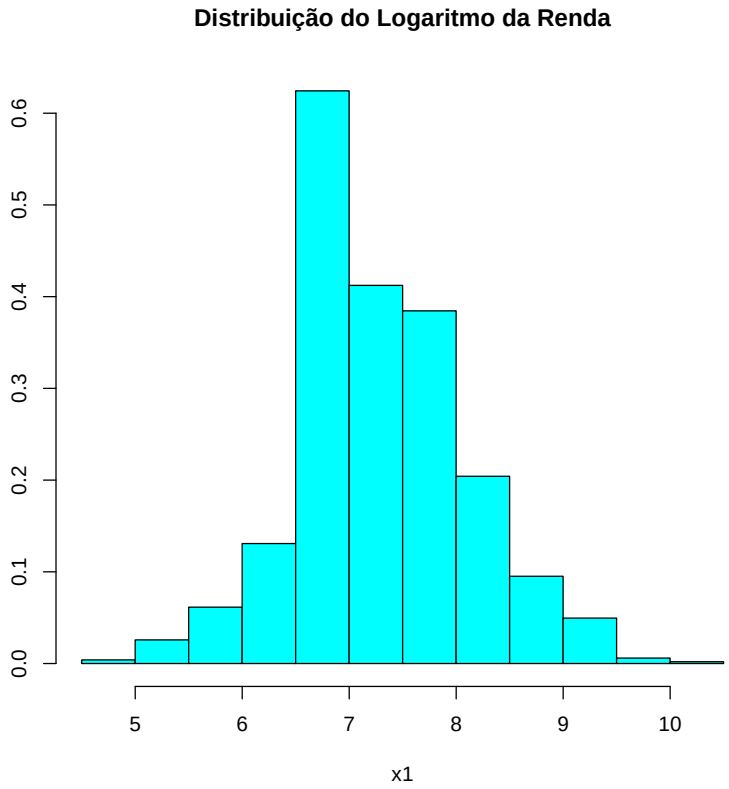


Figura 12 – Distribuição do logaritmo da renda da população do estado de Roraima

Para o ajuste dos modelos da classe SMt fixaremos os valores de ν e $\boldsymbol{\lambda}$, como foi feito no estudo de simulação, os quais serão escolhidos entre um conjunto de possíveis valores utilizando o menor AIC do modelo ajustado como critério de escolha. Assim, ajustamos os modelos t-Student, t-contaminada, Weibull-t e Slash-t, da classe SMt. Com o objetivo de comparar os nossos resultados com a classe SMN, ajustamos ao mesmo conjunto de dados, os modelos normal, t-Student, normal-contaminada e o modelo Slash, todos da classe SMN, utilizando o pacote *SMNCensReg* (GARAY; MASSUIA; LACHOS, 2015) disponível no software R. É importante salientar que o modelo t-Student pertence a

ambas as classes e, neste caso, as estimativas obtidas para μ e σ^2 coincidem, como mostra a Tabela 12, e o valor fixado para ν na classe SMt ficou próximo ao valor de ν obtido na classe SMN. Os modelos foram comparados utilizando o critério *Akaike Information Criteria* (AIC) (AKAIKE, 1974).

A Figura 13 mostra a variação do AIC com sinal trocado (-AIC), para visualizar o ponto de máximo ao invés do ponto de mínimo, no modelo t-Student e é possível observar que o menor valor do AIC está associado a algum valor próximo a $\nu = 7,5$.

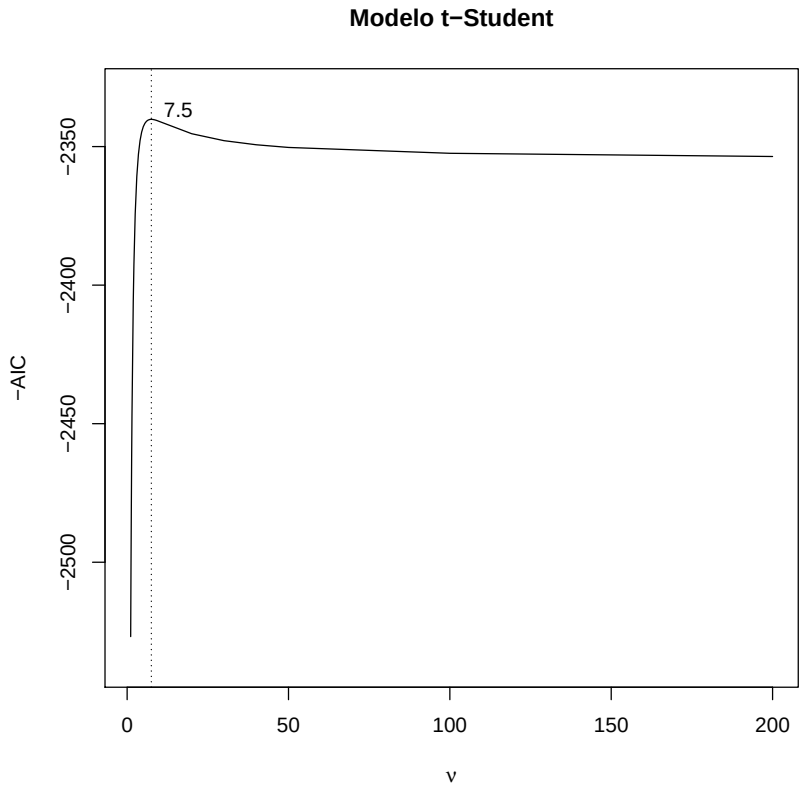


Figura 13 – Variação do AIC em função dos graus de liberdade na distribuição t-Student

A Figura 14 mostra um gráfico tri-dimensional com a variação do -AIC em função de $\boldsymbol{\lambda} = (\xi, \gamma)^\top$ no modelo t-Contaminada, com $\nu = 7,5$ que é o valor associado ao menor AIC. É possível observar que o menor valor do AIC está associado a algum par de valores próximo a $\boldsymbol{\lambda} = (0,7; 0,25)$.

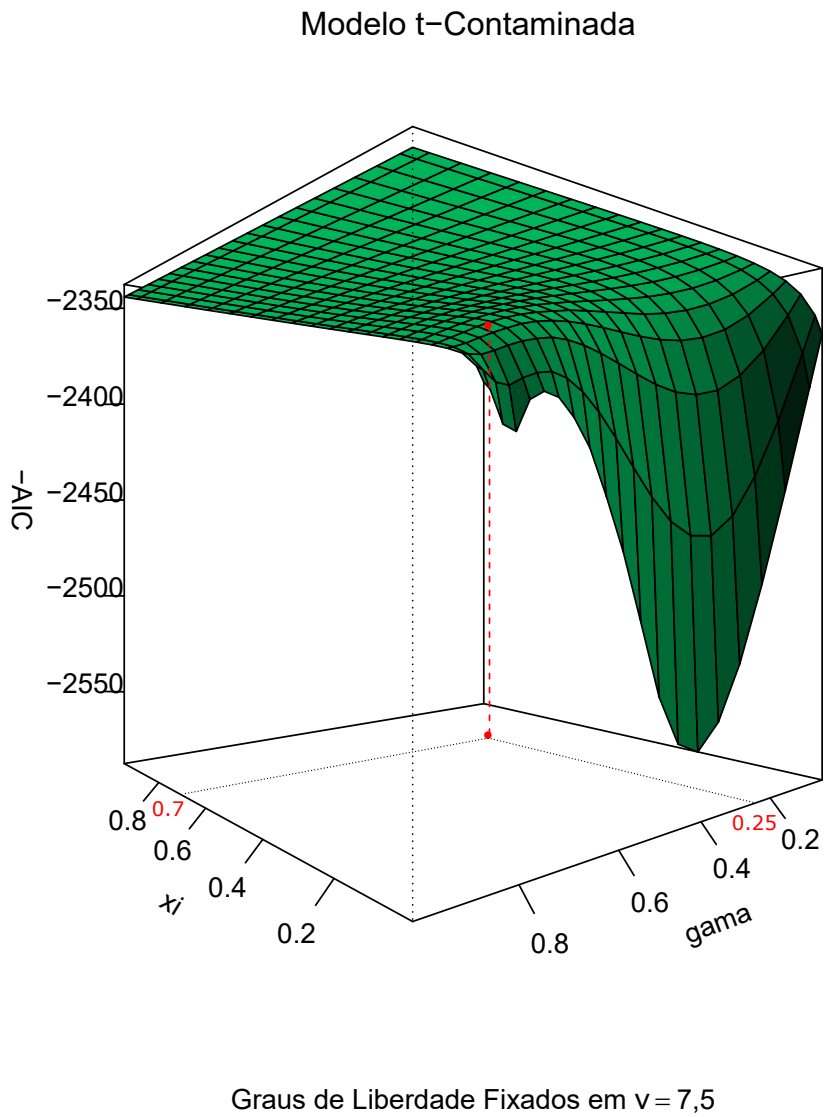


Figura 14 – Variação do AIC em função de $\boldsymbol{\lambda} = (\xi, \gamma)^\top$ na distribuição t-Contaminada, com $\nu = 7,5$

A Figura 15 mostra um gráfico tri-dimensional com a variação do -AIC em função de λ e ν no modelo Weibull-t e é possível observar que o menor valor do AIC está associado a algum par de valores próximo a $\lambda = 9$ e $\nu = 7,5$.

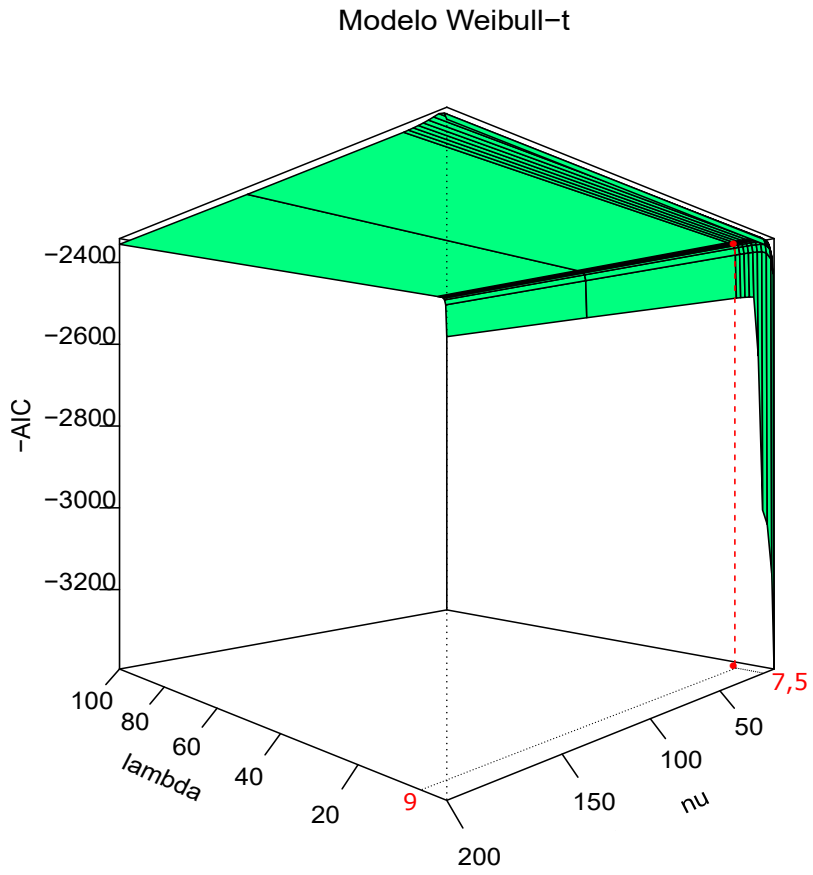


Figura 15 – Variação do AIC em função dos parâmetros λ e ν na distribuição Weibull-t

A Figura 16 mostra um gráfico tri-dimensional com a variação do -AIC em função de λ e ν no modelo Slash-t e a Figura 17 mostra o gráfico de -AIC em função apenas de λ, com $\nu = 7,5$ fixado, e é possível observar que o menor valor do AIC está associado a algum par de valores próximo a $\lambda = 120$ e $\nu = 7,5$. Observamos que o valor do AIC do modelo Slash-t diminuiu à medida em que o valor de λ aumenta até em torno de $\lambda = 120$, sendo que a partir desse valor o AIC começou a decrescer, de modo que este foi o valor escolhido.

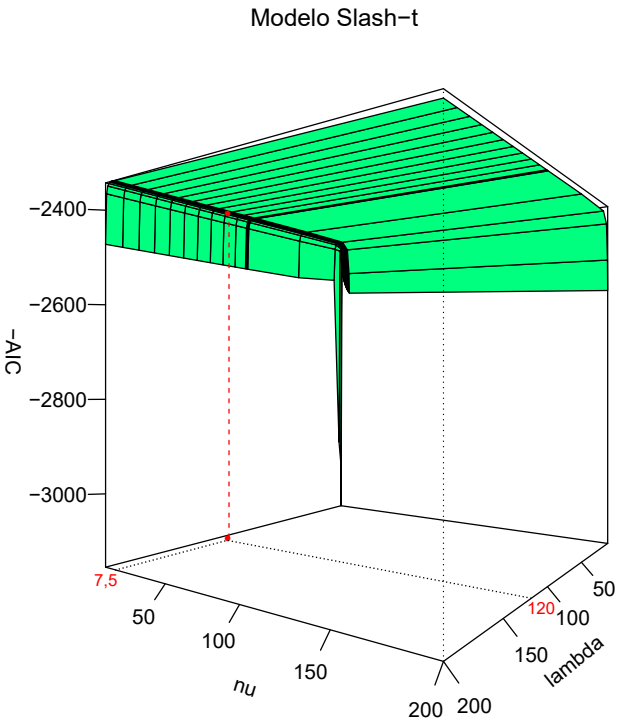


Figura 16 – Variação do AIC em função dos parâmetros λ e ν na distribuição Slash-t

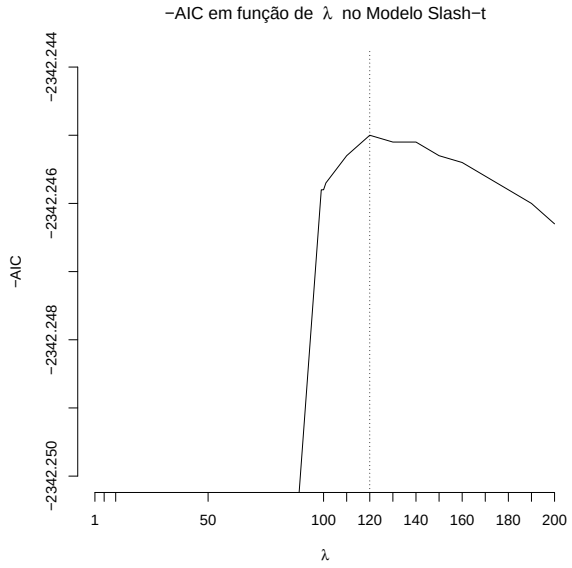


Figura 17 – Variação do AIC em função de λ com $\nu = 7,5$ na distribuição Slash-t

Na Tabela 12 é possível observar que o EMV da média μ em todos os modelos apresentaram valores próximos uns dos outros, sendo que o modelo t-Contaminada foi o

que apresentou o menor EP. O modelo que apresentou o maior valor da log-verossimilhança foi o t-contaminada e, além disso, também apresentou o menor valor do AIC. Desta forma, considerando o critério AIC e o valor da log-verossimilhança, o modelo t-contaminada é o mais apropriado para estes dados. Assim, conclui-se que o modelo t-contaminada se ajustou melhor aos dados do logaritmo da renda dos indivíduos com mais de 15 anos de idade residentes no estado de Roraima.

Tabela 12 – Resultados da estimação dos parâmetros de vários modelos para o logaritmo da renda dos indivíduos de Roraima

Classe SMN	**Normal**		**t-Student**		**Normal Contaminada**		**Slash**	
Parâmetro	**EMV**	**EP**	**EMV**	**EP**	**EMV**	**EP**	**EMV**	**EP**
μ	$7,2953$	$0,0251$	$7,2596$	$0,0243$	$7,2797$	$0,0249$	$7,2792$	$0,0248$
σ^2	$0,6005$	$0,0246$	$0,4496$	$0,0238$	$0,4214$	$0,0196$	$0,3832$	$0,0182$
$\boldsymbol{\lambda}$	–	–	–	–	(0,4;0,5)	–	$2,6821$	–
ν	–	–	$7,4931$	–	–	–	–	–
$\ell(\hat{\boldsymbol{\theta}})$	$-1174,4049$	–	$-1167,0875$	–	$-1168,9725$	–	$-1171,2477$	–
AIC	$2352,8099$	–	$2340,1751$	–	$2345,9450$	–	$2348,4954$	–
Classe SMt	**Weibull-t**		**t-Student**		**t-Contaminada**		**Slash-t**	
Parâmetro	**EMV**	**EP**	**EMV**	**EP**	**EMV**	**EP**	**EMV**	**EP**
μ	$7,2584$	$0,0243$	$7,2597$	$0,0243$	$7,2184$	$\mathbf{0,0231}$	$7,2607$	$0,0244$
σ^2	$0,4087$	$0,0217$	$0,4497$	$0,0238$	$0,1553$	$\mathbf{0,0096}$	$0,4491$	$0,0238$
$\boldsymbol{\lambda}$	5	–	–	–	(0,7;0,25)	–	120	–
ν	9	–	$7,5$	–	$7,5$	–	$7,5$	–
$\ell(\hat{\boldsymbol{\theta}})$	$-1166,7171$	–	$-1167,0875$	–	$\mathbf{-1163,8289}$	–	$-1167,0960$	–
AIC	$2341,4343$	–	$2340,1751$	–	$\mathbf{2337,6578}$	–	$2342,1920$	–

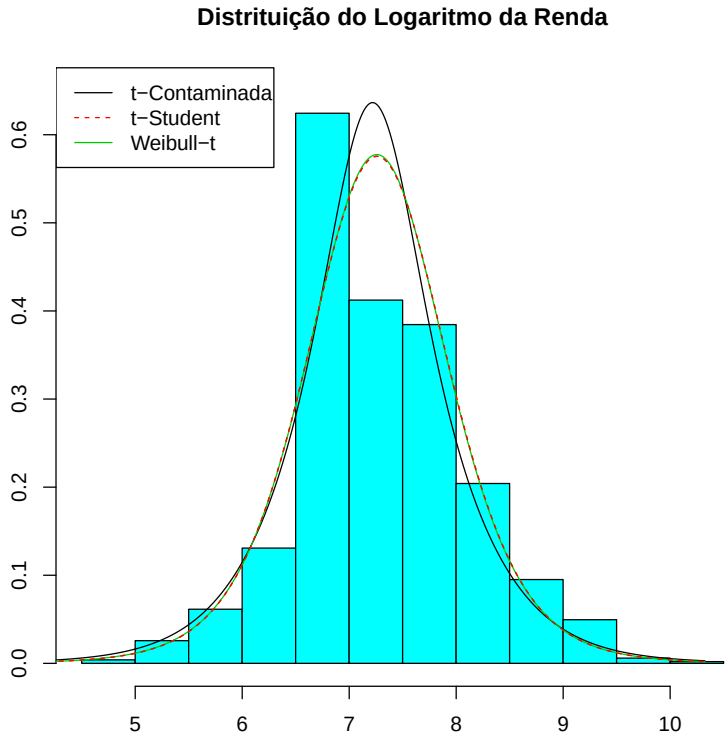


Figura 18 – Distribuição do logaritmo da renda da população do estado de Roraima e modelos ajustados

O gráfico da Figura 18 mostra o histograma dos dados juntamente a fdp dos modelos ajustados t-contaminada, t-Student e Weibull-t, que foram os três que apresentaram maiores valores de $\ell(\hat{\boldsymbol{\theta}})$.

3.6 Conclusões

Neste capítulo apresentamos a classe de distribuições mistura de escala t-Student, que é uma alternativa à classe mistura de escala normal, com o objetivo de modelar dados que seguem distribuições com caudas pesadas. Comparamos a curtose dessas duas classes, com mesmo fator de escala, e concluímos que as distribuições da classe SMt possuem curtose maior. Mostramos que a classe SMt converge para a classe SMN quando os graus de liberdade vão para o infinito, fixado o fator de escala, sendo esta uma propriedade atrativa desta nova classe.

Utilizamos o MM e o MV para estimar os parâmetros desta nova classe e realizamos estudos de simulação de MC para avaliar o desempenho dos estimadores obtidos considerando as quatro distribuições apresentadas que pertencem à classe SMt, a saber: t-Student, t-Contaminada, Weibull-t e Slash-t; e verificamos que o desempenho dos estimadores foi satisfatório. Além disso, fizemos uma aplicação a um conjunto de dados reais em que o modelo t-Contaminada foi o mais apropriado.

4 CONSIDERAÇÕES FINAIS

Nesta tese discutimos alguns aspectos importantes nas duas linhas de pesquisa abordadas. Em análise de dados circulares, um dos aspectos discutidos foi uma nova distribuição circular assimétrica chamada de *wrapped* Birnbaum-Saunders em que foi mostrado via simulação que a estimação de seus parâmetros pelo método da máxima verossimilhança sofre grande influência do parâmetro de concentração circular, sendo que, quanto menor a concentração dos dados, se torna mais difícil obter estimativas robustas dos parâmetros. Foi mostrado que os momentos trigonométricos desta distribuição podem ser aproximados via expansão em série de Taylor de segunda ordem para contornar as dificuldades matemáticas em lidar com suas expressões complexas. Acreditamos que esta nova distribuição possui grande potencial na modelagem de dados circulares assimétricos.

Em modelos mistura de escala, discutimos a classe de distribuições SMt e provamos alguns resultados a respeito dos seus momentos. Mostramos que esta classe converge para a classe SMN e possui grande potencial na modelagem de dados com presença de observações aberrantes, principalmente porque foi provado que esta classe possui curtose maior que da classe SMN. Mostramos via simulação que os EMV dos parâmetros do modelo são eficientes, quando ν e $\boldsymbol{\lambda}$ são fixados. Apresentamos uma aplicação em que uma distribuição da classe SMt se ajustou melhor aos dados do que outras distribuições da classe SMN, mostrando a aplicabilidade da classe proposta.

Concluindo, esta tese representa um esforço inicial para apresentar alguns tópicos nestas duas áreas de pesquisa e divulgar sua utilidade.

REFERÊNCIAS

ADNAN, M. A. S.; ROY, S. Wrapped variance gamma distribution with an application to wind direction. *Journal of Environmentl Statistics*, v. 6, n. 2, p. 1–10, 9 2014. ISSN 1945-1296.

AKAIKE, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, v. 19, n. 6, p. 716–723, December 1974. ISSN 0018-9286.

ANDREWS, D. F.; MALLOWS, C. L. Scale mixtures of normal distributions. *Journal of the Royal Statistical Society. Series B (Methodological)*, [Royal Statistical Society, Wiley], v. 36, n. 1, p. 99–102, 1974. ISSN 00359246.

BERKANE, M.; KANO, Y.; BENTLER, P. M. Pseudo maximum likelihood estimation in elliptical theory: Effects of misspecification. *Computational Statistics and Data Analysis*, v. 18, n. 2, p. 255 – 267, 1994. ISSN 0167-9473.

BIRNBAUM, Z. W.; SAUNDERS, S. C. A new family of life distributions. *Journal of Applied Probability*, Applied Probability Trust, v. 6, n. 2, p. 319–327, 1969. ISSN 00219002.

BOYLES, R. A. On the convergence of the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, [Royal Statistical Society, Wiley], v. 45, n. 1, p. 47–50, 1983. ISSN 00359246. Disponível em: <http://www.jstor.org/stable/2345622>.

CASELLA, G.; BERGER, R. L. *Statistical Inference*. 2. ed. Thomson Learning, 2002. (Duxbury advanced series in statistics and decision sciences). ISBN 9780534243128. Disponível em: <https://books.google.com.br/books?id=0x_vAAAAMAAJ>.

COELHO, C. A. The wrapped gamma distribution and wrapped sums and linear combinations of independent gamma and laplace distributions. *Journal of Statistical Theory and Practice*, v. 1, n. 1, p. 1–29, 2007.

DELYON, B.; LAVIELLE, M.; MOULINES, E. Convergence of a stochastic approximation version of the em algorithm. *The Annals of Statistics*, Institute of Mathematical Statistics, v. 27, n. 1 (Feb. 1999), p. 94–128, 1999.

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. ISSN 00359246.

EFRON, B.; HINKLEY, D. V. Assessing the accuracy of the maximum likelihood estimator: Observed versus expected fisher information. *Biometrika*, [Oxford University Press, Biometrika Trust], v. 65, n. 3, p. 457–482, 1978. ISSN 00063444.

FISHER, N. I. *Statistical Analysis of Circular Data*. Cambridge: Cambridge University Press, 1995. ISBN 0-521-56890-0.

GALARZA, C. E.; LACHOS, V. H.; BANDYOPADHYAY, D. Quantile regression in linear mixed models: a stochastic approximation em approach. *Statistics and its interface*, v. 10, n. 3, p. 471–482, 2017.

GARAY, A. M. et al. Censored linear regression models for irregularly observed longitudinal data using the multivariate-t distribution. *Statistical Methods in Medical Research*, v. 26, n. 2, p. 542–566, 2017. PMID: 25296865.

GARAY, A. M. et al. Linear censored regression models with scale mixtures of normal distributions. *Statistical Papers*, Springer, v. 58, n. 1, p. 395–402, 2017. ISSN 0932-5026.

GARAY, A. M.; MASSUIA, M. B.; LACHOS, V. *SMNCensReg: Fitting Univariate Censored Regression Model Under the Family of Scale Mixture of Normal Distributions*. [S.l.], 2015. R package version 3.0. Disponível em: <https://CRAN.R-project.org/package=SMNCensReg>.

IBGE. *Pesquisa Nacional por Amostra de Domicílios - PNAD*. [S.l.], 2015. Acessado em 16-03-2019. Disponível em: <https://www.ibge.gov.br/estatisticas-novoportal/sociais/populacao/9127-pesquisa-nacional-por-amostra-de-domicilios.html?=&t=downloads>.

IBGE. *Pesquisa Nacional por Amostra de Domicílios - PNAD*. [S.l.], 2019. Acessado em 16-03-2019. Disponível em: <https://www.ibge.gov.br/estatisticas-novoportal/sociais/populacao/9127-pesquisa-nacional-por-amostra-de-domicilios.html?=&t=o-que-e>.

JACOB, S.; JAYAKUMAR, K. Wrapped geometric distribution: A new probability model for circular data. *Journal of Statistical Theory and Applications*, Atlantis Press, v. 12, n. 4, p. 348–355, 2013.

JAMMALAMADAKA, R.; KOZUBOWSKI, T. New families of wrapped distributions for modeling skew circular data. *Communications in Statistics - Theory and Methods*, v. 33, n. 9, p. 2059–2074, 2004.

JAMMALAMADAKA, R.; SENGUPTA, A. *Topics in Circular Statistics*. [S.l.]: World Scientific Publishing Co. Pte. Ltd., 2001. v. 5. (Series on multivariate analysis, v. 5). ISBN 9810237782.

JANDER, R. Die optische richtungsorientierung der roten waldameise (formica ruea l.). *Zeitschrift für vergleichende Physiologie*, v. 40, n. 2, p. 162–238, 1957. ISSN 1432-1351.

JANK, W. Implementing and diagnosing the stochastic approximation em algorithm. *Journal of Computational and Graphical Statistics*, [American Statistical Association, Taylor & Francis, Ltd., Institute of Mathematical Statistics, Interface Foundation of America], v. 15, n. 4, p. 803–829, 2006. ISSN 10618600.

JOHNSON, N. L.; KOTZ, S.; BALAKRISHNAN, N. *Continuous Univariate Distributions*. 2. ed. [S.l.]: John Willey and Sons, Inc., 1995. v. 2. 365 p. ISBN 0-471-58494-0.

JOSHI, S.; K, J. K. Wrapped lindley distribution. *Communications in Statistics - Theory and Methods*, 2017.

LANGE, K. L.; LITTLE, R. J. A.; TAYLOR, J. M. G. Robust statistical modeling using the t distribution. *Journal of the American Statistical Association*, Taylor and Francis, v. 84, n. 408, p. 881–896, 1989.

LEVY, P. L'addition des variables aléatoires définies sur une circonférence. *Bulletin de la Société Mathématique de France*, v. 67, p. 1–41, 1939.

LIU, C.; RUBIN, D. B. The ecme algorithm: A simple extension of em and ecm with faster monotone convergence. *Biometrika*, [Oxford University Press, Biometrika Trust], v. 81, n. 4, p. 633–648, 1994. ISSN 00063444.

LOUIS, T. A. Finding the observed information matrix when using the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, [Royal Statistical Society, Wiley], v. 44, n. 2, p. 226–233, 1982. ISSN 00359246.

LUCAS, A. Robustness of the student t based m-estimator. *Communications in Statistics - Theory and Methods*, Taylor & Francis, v. 26, n. 5, p. 1165–1182, 1997.

MARDIA, K.; JUPP, P. *Directional statistics*. 2 sub. ed. [S.l.]: John Wiley and Sons, LTD, 2000. (Wiley series in probability and statistics). ISBN 0-471-95333-4.

MASSUIA, M. B. et al. Influence diagnostics for student-t censored linear regression models. *Statistics*, Taylor & Francis, v. 49, n. 5, p. 1074–1094, 2015.

MATTOS, T. do B.; GARAY, A. M.; LACHOS, V. H. Likelihood-based inference for censored linear regression models with scale mixtures of skew-normal distributions. *Journal of Applied Statistics*, Taylor & Francis, v. 45, n. 11, p. 2039–2066, 2018. Disponível em: <https://doi.org/10.1080/02664763.2017.1408788>.

MATYAS, L. (Ed.). *Generalized method of moments estimation*. [S.l.]: Cambrige University Press, 1999. ISBN 978-0-521-66967-2.

MCLACHLAN, G. J.; BASFORD, K. E. djvu. *Mixture Models: Inference and Applications to Clustering*. [S.l.]: Marcel Dekker, INC., 1997. v. 84. (STATISTICS: Textbooks and Monographs, v. 84). ISBN 0-8247-7691-7.

MENG, X.-L.; RUBIN, D. B. Maximum likelihood estimation via the ecm algorithm: A general framework. *Biometrika*, [Oxford University Press, Biometrika Trust], v. 80, n. 2, p. 267–278, 1993. ISSN 00063444.

MITTELHAMMER, R. C.; JUDGE, G. G.; MILLER, D. J. *Econometric Foundations*. New York, NY, USA: Cambridge University Press, 2000. ISBN 0-521-62394-4.

NOCEDAL, J.; WRIGHT, S. J. *Numerical Optimization*. New York: Springer-Verlag New York, Inc., 1999. (Springer Series in Operations Research). ISBN 0-387-98793-2.

PIESSENS, R. et al. *Quadpack*: A subroutine package for automatic integration. [S.l.]: Springer-Verlag Berlin Heidelberg, 1983. (Springer Series in Computational Mathematics). ISBN 978-3-642-61786-7.

POLYAK, B.; JUDITSKY, A. Acceleration of stochastic approximation by averaging. *SIAM Journal on Control and Optimization*, v. 30, n. 4, p. 838–855, 1992.

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2018. Disponível em: <https://www.R-project.org/>.

RAO, A. D.; GIRIJA, S.; DEVARAAJ, V. On characteristics of wrapped gamma distribution. *IRACST – Engineering Science and Technology: An International Journal (ESTIJ)*, v. 3, n. 2, april 2013. ISSN 2250-3498.

RAO, A. V. D.; SARMA, I. R.; GIRIJA, S. V. S. On wrapped version of some life testing models. *Communications in Statistics - Theory and Methods*, v. 36, n. 11, p. 2027–2035, 2007.

ROY, S.; ADNAN, M. A. S. Wrapped generalized gompertz distribution: an application to ornithology. *Journal of Biometrics & Biostatistics*, 2012.

ROY, S.; ADNAN, M. A. S. Wrapped weighted exponential distributions. *Statistics & Probability Letters*, v. 82, n. 1, p. 77 – 83, 2012. ISSN 0167-7152.

RUBIN, D. B. Comment. *Journal of the American Statistical Association*, Taylor & Francis, v. 82, n. 398, p. 543–546, 1987.

RUBIN, D. B. Using the sir algorithm to simulate posterior distributions. *Bayesian Statistics 3*, p. 395–402, 1988.

RUCKDESCHEL, P. et al. S4 classes for distributions. *R News*, v. 6, n. 2, p. 2–6, May 2006.

SANTOS-NETO, M. et al. On new parameterizations of the birnbaum-saunders distribution. *Pakistan Journal of Statistics*, v. 28, n. 1, p. 1–26, 2012. ISSN 1012-9367.

SANTOS-NETO, M. et al. A reparameterized birnbaum–saunders distribution and its moments, estimation and applications. *REVSTAT–Statistical Journal*, v. 12, n. 3, p. 247–272, 2014.

SCOTT, W. A. Maximum likelihood estimation using the empirical fisher information matrix. *Journal of Statistical Computation and Simulation*, Taylor & Francis, v. 72, n. 8, p. 599–611, 2002.

SOUZA, I. C. A. de. *Modelagem de dados circulares e na reta baseada em distribuições derivadas da família de contornos elípticos*. 137 f. Doutorado em Estatística — Universidade Federal de Pernambuco, Recife-PE, 2014.

UMBACH, D.; JAMMALAMADAKA, R. Building asymmetry into circular distributions. *Statistics & Probability Letters*, v. 79, n. 5, p. 659 – 663, 2009. ISSN 0167-7152.

WEI, G. C. G.; TANNER, M. A. A monte carlo implementation of the em algorithm and the poor mans data augmentation algorithms. *Journal of the American Statistical Association*, Taylor and Francis Ltd., v. 85, n. 411, p. 699–704, 1990. ISSN 0162-1459.

WEI, G. C. G.; TANNER, M. A. Posterior computations for censored regression data. *Journal of the American Statistical Association*, [American Statistical Association, Taylor & Francis, Ltd.], v. 85, n. 411, p. 829–839, 1990. ISSN 01621459.

WEST, M. On scale mixtures of normal distributions. *Biometrika*, v. 74, n. 3, p. 646–648, 1987.

Wolfram Research, Inc. Mathematica, Version 11.1. Champaign, IL, 2018. 2018.

WOLFRAM|ALPHA. [S.l.], 2019. Acessado em 10-01-2019. Disponível em: <https://www.wolframalpha.com/input/?i=moments+of+student+t+distribution>.

WU, C. F. J. On the convergence properties of the em algorithm. *The Annals of Statistics*, Institute of Mathematical Statistics, v. 11, n. 1, p. 95–103, 1983. ISSN 00905364. Disponível em: <http://www.jstor.org/stable/2240463>.

APÊNDICE A – SEGUNDAS DERIVADAS

Neste apêndice são apresentadas as segundas derivadas parciais da função de log-verossimilhança da distribuição *wrapped* Birnbaum-Saunders.

$$\frac{\partial^2 \ell(\mu,\delta;\boldsymbol{\theta})}{\partial \mu^2} = \frac{n}{2\mu^2} + \sum_{j=1}^{n}\left\{\left[\sum_{k=0}^{\infty}\left(\frac{\delta^2\left(\frac{\delta}{(\delta+1)(\theta_j+2k\pi)} - \frac{(\delta+1)(\theta_j+2k\pi)}{\delta\mu^2}\right)^2\left(\frac{\delta\mu}{\delta+1}+\theta_j+2k\pi\right)\exp\left(-\frac{\delta}{4}\left[\frac{\delta\mu}{(\delta+1)(\theta_j+2k\pi)}+\frac{(\delta+1)(\theta_j+2k\pi)}{\delta\mu}\right]\right)}{16\left(\theta_j+2k\pi\right)^{1.5}}\right.\right.$$

$$-\frac{\delta^2\left(\frac{\delta}{(\delta+1)(\theta_j+2k\pi)} - \frac{(\delta+1)(\theta_j+2k\pi)}{\delta\mu^2}\right)\exp\left(-\frac{\delta}{4}\left[\frac{\delta\mu}{(\delta+1)(\theta_j+2k\pi)}+\frac{(\delta+1)(\theta_j+2k\pi)}{\delta\mu}\right]\right)}{2(\delta+1)\left(\theta_j+2k\pi\right)^{1.5}}$$

$$\left.\left.-\frac{(\delta+1)\left(\frac{\delta\mu}{\delta+1}+\theta_j+2k\pi\right)\exp\left(-\frac{\delta}{4}\left[\frac{\delta\mu}{(\delta+1)(\theta_j+2k\pi)}+\frac{(\delta+1)(\theta_j+2k\pi)}{\delta\mu}\right]\right)}{2\mu^3\left(\theta_j+2k\pi\right)^{0.5}}\right)\right]\times\left\{\sum_{k=0}^{\infty}\frac{\left(\frac{\delta\mu}{\delta+1}+\theta_j+2k\pi\right)\exp\left(-\frac{\delta}{4}\left(\frac{\delta\mu}{(\delta+1)(\theta_j+2k\pi)}+\frac{(\delta+1)(\theta_j+2k\pi)}{\delta\mu}\right)\right)}{\left(\theta_j+2k\pi\right)^{1.5}}\right\}^{-1}$$

$$-\frac{\left[\sum_{k=0}^{\infty}\left(\frac{\delta\exp\left(-\frac{\delta}{4}\left[\frac{\delta\mu}{(\delta+1)\left(\theta_j+2k\pi\right)}+\frac{(\delta+1)\left(\theta_j+2k\pi\right)}{\delta\mu}\right]\right)}{(\delta+1)(\theta_j+2k\pi)^{1.5}} - \frac{\delta\left(\frac{\delta\mu}{\delta+1}+\theta_j+2k\pi\right)\left(\frac{\delta}{(\delta+1)\left(\theta_j+2k\pi\right)}-\frac{(\delta+1)\left(\theta_j+2k\pi\right)}{\delta\mu^2}\right)\exp\left(-\frac{\delta}{4}\left[\frac{\delta\mu}{(\delta+1)\left(\theta_j+2k\pi\right)}+\frac{(\delta+1)\left(\theta_j+2k\pi\right)}{\delta\mu}\right]\right)}{4(\theta_j+2k\pi)^{1.5}}\right)\right]^2}{\left(\sum_{k=0}^{\infty}\frac{\left(\frac{\delta\mu}{\delta+1}+\theta_j+2k\pi\right)\exp\left(-\frac{\delta}{4}\left[\frac{\delta\mu}{(\delta+1)\left(\theta_j+2k\pi\right)}+\frac{(\delta+1)\left(\theta_j+2k\pi\right)}{\delta\mu}\right]\right)}{(\theta_j+2k\pi)^{1.5}}\right)^2}\Bigg\}$$

$$\frac{\partial^2 \ell(\mu,\delta;\boldsymbol{\theta})}{\partial \mu\partial\delta} = \sum_{j=1}^{n}\left\{\left[\sum_{k=0}^{b}\left(-\frac{\delta\exp\left(-\frac{\delta}{4}\left[\frac{\delta\mu}{(\delta+1)\left(2k\pi+\theta_j\right)}+\frac{(\delta+1)\left(2k\pi+\theta_j\right)}{\delta\mu}\right]\right)\left(\frac{\mu}{\delta+1}-\frac{\delta\mu}{(\delta+1)^2}\right)\left(\frac{\delta}{(\delta+1)\left(2k\pi+\theta_j\right)}-\frac{(\delta+1)\left(2k\pi+\theta_j\right)}{\delta\mu^2}\right)}{4\left(2k\pi+\theta_j\right)^{1.5}}\right.\right.$$

$$\left.\left.-\frac{\delta\exp\left(-\frac{\delta}{4}\left[\frac{\delta\mu}{(\delta+1)\left(2k\pi+\theta_j\right)}+\frac{(\delta+1)\left(2k\pi+\theta_j\right)}{\delta\mu}\right]\right)}{(\delta+1)^2\left(2k\pi+\theta_j\right)^{1.5}} - \frac{\delta\exp\left(-\frac{\delta}{4}\left[\frac{\delta\mu}{(\delta+1)\left(2k\pi+\theta_j\right)}+\frac{(\delta+1)\left(2k\pi+\theta_j\right)}{\delta\mu}\right]\right)\left(2k\pi+\frac{\delta\mu}{\delta+1}+\theta_j\right)\left(-\frac{\delta}{(\delta+1)^2\left(2k\pi+\theta_j\right)}+\frac{(\delta+1)\left(2k\pi+\theta_j\right)}{\delta^2\mu^2}-\frac{2k\pi+\theta_j}{\delta\mu^2}+\frac{1}{(\delta+1)\left(2k\pi+\theta_j\right)}\right)}{4\left(2k\pi+\theta_j\right)^{1.5}}\right]\Bigg\}$$

$$
\begin{aligned}
&- \left[\delta \exp\left(-\frac{\delta}{4}\left[\frac{\delta\mu}{(\delta+1)(2k\pi+\theta_j)}+\frac{(\delta+1)(2k\pi+\theta_j)}{\delta\mu}\right]\right)\left(2k\pi+\frac{\delta\mu}{\delta+1}+\theta_j\right)\left(\frac{\delta}{(\delta+1)(2k\pi+\theta_j)}-\frac{(\delta+1)(2k\pi+\theta_j)}{\delta\mu^2}\right)\left(\frac{1}{4}\left(-\frac{\delta\mu}{(\delta+1)(2k\pi+\theta_j)}-\frac{(\delta+1)(2k\pi+\theta_j)}{\delta\mu}\right)\right.\right.\\
&\left.\left.-\frac{\delta}{4}\left(\frac{\mu}{(\delta+1)(2k\pi+\theta_j)}-\frac{\delta\mu}{(\delta+1)^2(2k\pi+\theta_j)}+\frac{2k\pi+\theta_j}{\delta\mu}-\frac{(\delta+1)(2k\pi+\theta_j)}{\delta^2\mu}\right)\right)\right]\left[4(2k\pi+\theta_j)^{1.5}\right]^{-1}\\
&+\frac{\delta\exp\left(-\frac{\delta}{4}\left[\frac{\delta\mu}{(\delta+1)\left(2k\pi+\theta_j\right)}+\frac{(\delta+1)\left(2k\pi+\theta_j\right)}{\delta\mu}\right]\right)\left(\frac{1}{4}\left(-\frac{\delta\mu}{(\delta+1)\left(2k\pi+\theta_j\right)}-\frac{(\delta+1)\left(2k\pi+\theta_j\right)}{\delta\mu}\right)-\frac{\delta}{4}\left(\frac{\mu}{(\delta+1)\left(2k\pi+\theta_j\right)}-\frac{\delta\mu}{(\delta+1)^2\left(2k\pi+\theta_j\right)}+\frac{2k\pi+\theta_j}{\delta\mu}-\frac{(\delta+1)\left(2k\pi+\theta_j\right)}{\delta^2\mu}\right)\right)}{(\delta+1)(2k\pi+\theta_j)^{1.5}}\\
&-\frac{\exp\left(-\frac{\delta}{4}\left[\frac{\delta\mu}{(\delta+1)\left(2k\pi+\theta_j\right)}+\frac{(\delta+1)\left(2k\pi+\theta_j\right)}{\delta\mu}\right]\right)\left(2k\pi+\frac{\delta\mu}{\delta+1}+\theta_j\right)\left(\frac{\delta}{(\delta+1)\left(2k\pi+\theta_j\right)}-\frac{(\delta+1)\left(2k\pi+\theta_j\right)}{\delta\mu^2}\right)}{4(2k\pi+\theta_j)^{1.5}}+\frac{\exp\left(-\frac{\delta}{4}\left[\frac{\delta\mu}{(\delta+1)\left(2k\pi+\theta_j\right)}+\frac{(\delta+1)\left(2k\pi+\theta_j\right)}{\delta\mu}\right]\right)}{(\delta+1)(2k\pi+\theta_j)^{1.5}}\Bigg)\Bigg]\\
&\times\left[\sum_{k=0}^{b}\frac{\exp\left(-\frac{\delta}{4}\left[\frac{\delta\mu}{(\delta+1)\left(2k\pi+\theta_j\right)}+\frac{(\delta+1)\left(2k\pi+\theta_j\right)}{\delta\mu}\right]\right)\left(2k\pi+\frac{\delta\mu}{\delta+1}+\theta_j\right)}{(2k\pi+\theta_j)^{1.5}}\right]^{-1}-\left(\sum_{k=0}^{b}\left(\frac{\delta\ \exp\left(-\frac{\delta}{4}\left[\frac{\delta\mu}{(\delta+1)\left(2k\pi+\theta_j\right)}+\frac{(\delta+1)\left(2k\pi+\theta_j\right)}{\delta\mu}\right]\right)}{(\delta+1)(2k\pi+\theta_j)^{1.5}}\right.\right.\\
&\left.\left.-\frac{\exp\left(-\frac{\delta}{4}\left[\frac{\delta\mu}{(\delta+1)\left(2k\pi+\theta_j\right)}+\frac{(\delta+1)\left(2k\pi+\theta_j\right)}{\delta\mu}\right]\right)\delta\left(2k\pi+\frac{\delta\mu}{\delta+1}+\theta_j\right)\left(\frac{\delta}{(\delta+1)\left(2k\pi+\theta_j\right)}-\frac{(\delta+1)\left(2k\pi+\theta_j\right)}{\delta\mu^2}\right)}{4(2k\pi+\theta_j)^{1.5}}\right)\right)\sum_{k=0}^{b}\left(\frac{\exp\left(-\frac{\delta}{4}\left[\frac{\delta\mu}{(\delta+1)\left(2k\pi+\theta_j\right)}+\frac{(\delta+1)\left(2k\pi+\theta_j\right)}{\delta\mu}\right]\right)\left(\frac{\mu}{\delta+1}-\frac{\delta\mu}{(\delta+1)^2}\right)}{(2k\pi+\theta_j)^{1.5}}\right.\\
&\left.+\frac{\exp\left(-\frac{\delta}{4}\left[\frac{\delta\mu}{(\delta+1)\left(2k\pi+\theta_j\right)}+\frac{(\delta+1)\left(2k\pi+\theta_j\right)}{\delta\mu}\right]\right)\left(2k\pi+\frac{\delta\mu}{\delta+1}+\theta_j\right)\left(\frac{1}{4}\left(-\frac{\delta\mu}{(\delta+1)\left(2k\pi+\theta_j\right)}-\frac{(\delta+1)\left(2k\pi+\theta_j\right)}{\delta\mu}\right)-\frac{\delta}{4}\left(\frac{\mu}{(\delta+1)\left(2k\pi+\theta_j\right)}-\frac{\delta\mu}{(\delta+1)^2\left(2k\pi+\theta_j\right)}+\frac{2k\pi+\theta_j}{\delta\mu}-\frac{(\delta+1)\left(2k\pi+\theta_j\right)}{\delta^2\mu}\right)\right)}{(2k\pi+\theta_j)^{1.5}}\right)\\
&\times\left(\sum_{k=0}^{b}\frac{\exp\left(-\frac{\delta}{4}\left[\frac{\delta\mu}{(\delta+1)\left(2k\pi+\theta_j\right)}+\frac{(\delta+1)\left(2k\pi+\theta_j\right)}{\delta\mu}\right]\right)\left(2k\pi+\frac{\delta\mu}{\delta+1}+\theta_j\right)}{(2k\pi+\theta_j)^{1.5}}\right)^{-2}\Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2\ell(\mu,\delta;\boldsymbol{\theta})}{\partial\delta^2}=\sum_{j=1}^{n}\Bigg\{\sum_{k=0}^{b}\Bigg\{&\left[2\exp\left(-\frac{\delta}{4}\left[\frac{\delta\mu}{(\delta+1)(2k\pi+\theta_j)}+\frac{(\delta+1)(2k\pi+\theta_j)}{\delta\mu}\right]\right)\left(\frac{\mu}{\delta+1}-\frac{\delta\mu}{(\delta+1)^2}\right)\right.\\
&\left.\left(\frac{1}{4}\left(-\frac{\delta\mu}{(\delta+1)(2k\pi+\theta_j)}-\frac{(\delta+1)(2k\pi+\theta_j)}{\delta\mu}\right)-\frac{\delta}{4}\left(\frac{\mu}{(\delta+1)(2k\pi+\theta_j)}-\frac{\delta\mu}{(\delta+1)^2(2k\pi+\theta_j)}+\frac{2k\pi+\theta_j}{\delta\mu}-\frac{(\delta+1)(2k\pi+\theta_j)}{\delta^2\mu}\right)\right)\right]
\end{aligned}
$$

$$\times \left(2k\pi+\theta_j\right)^{-1.5}+\left[\exp\left(-\frac{\delta}{4}\left[\frac{\delta\mu}{(\delta+1)\left(2k\pi+\theta_j\right)}+\frac{(\delta+1)\left(2k\pi+\theta_j\right)}{\delta\mu}\right]\right)\left(2k\pi+\frac{\delta\mu}{\delta+1}+\theta_j\right)\left[\frac{1}{4}\left(-\frac{\delta\mu}{(\delta+1)\left(2k\pi+\theta_j\right)}-\frac{(\delta+1)\left(2k\pi+\theta_j\right)}{\delta\mu}\right)\right.$$
$$\left.-\frac{\delta}{4}\left(\frac{\mu}{(\delta+1)\left(2k\pi+\theta_j\right)}-\frac{\delta\mu}{(\delta+1)^2\left(2k\pi+\theta_j\right)}+\frac{2k\pi+\theta_j}{\delta\mu}-\frac{(\delta+1)\left(2k\pi+\theta_j\right)}{\delta^2\mu}\right)\right]^2\Bigg]\times\left(2k\pi+\theta_j\right)^{-1.5}$$
$$+\left[\exp\left(-\frac{\delta}{4}\left[\frac{\delta\mu}{(\delta+1)\left(2k\pi+\theta_j\right)}+\frac{(\delta+1)\left(2k\pi+\theta_j\right)}{\delta\mu}\right]\right)\left(2k\pi+\frac{\delta\mu}{\delta+1}+\theta_j\right)\right.$$
$$\left(\frac{1}{2}\left(-\frac{\mu}{(\delta+1)\left(2k\pi+\theta_j\right)}+\frac{\delta\mu}{(\delta+1)^2\left(2k\pi+\theta_j\right)}+\frac{(\delta+1)\left(2k\pi+\theta_j\right)}{\delta^2\mu}-\frac{2k\pi+\theta_j}{\delta\mu}\right)\right.$$
$$\left.\left.-\frac{\delta}{4}\left(-\frac{2\mu}{(\delta+1)^2\left(2k\pi+\theta_j\right)}+\frac{2\delta\mu}{(\delta+1)^3\left(2k\pi+\theta_j\right)}+\frac{2(\delta+1)\left(2k\pi+\theta_j\right)}{\delta^3\mu}-\frac{2\left(2k\pi+\theta_j\right)}{\delta^2\mu}\right)\right)\right]\times\left(2k\pi+\theta_j\right)^{-1.5}$$
$$+\frac{\exp\left(-\frac{\delta}{4}\left[\frac{\delta\mu}{(\delta+1)(2k\pi+\theta_j)}+\frac{(\delta+1)(2k\pi+\theta_j)}{\delta\mu}\right]\right)\left(\frac{2\delta\mu}{(\delta+1)^3}-\frac{2\mu}{(\delta+1)^2}\right)}{\left(2k\pi+\theta_j\right)^{1.5}}\Bigg\}\times\left\{\sum_{k=0}^{b}\frac{\exp\left(-\frac{\delta}{4}\left[\frac{\delta\mu}{(\delta+1)(2k\pi+\theta_j)}+\frac{(\delta+1)(2k\pi+\theta_j)}{\delta\mu}\right]\right)\left(2k\pi+\frac{\delta\mu}{\delta+1}+\theta_j\right)}{\left(2k\pi+\theta_j\right)^{1.5}}\right\}^{-1}$$
$$-\left[\sum_{k=0}^{b}\left(\frac{\exp\left(-\frac{\delta}{4}\left[\frac{\delta\mu}{(\delta+1)(2k\pi+\theta_j)}+\frac{(\delta+1)(2k\pi+\theta_j)}{\delta\mu}\right]\right)\left(\frac{\mu}{\delta+1}-\frac{\delta\mu}{(\delta+1)^2}\right)}{\left(2k\pi+\theta_j\right)^{1.5}}\right.\right.$$
$$+\left[\exp\left(-\frac{\delta}{4}\left[\frac{\delta\mu}{(\delta+1)\left(2k\pi+\theta_j\right)}+\frac{(\delta+1)\left(2k\pi+\theta_j\right)}{\delta\mu}\right]\right)\left(2k\pi+\frac{\delta\mu}{\delta+1}+\theta_j\right)\left(\frac{1}{4}\left(-\frac{\delta\mu}{(\delta+1)\left(2k\pi+\theta_j\right)}-\frac{(\delta+1)\left(2k\pi+\theta_j\right)}{\delta\mu}\right)\right.\right.$$
$$\left.\left.\left.-\frac{\delta}{4}\left(\frac{\mu}{(\delta+1)\left(2k\pi+\theta_j\right)}-\frac{\delta\mu}{(\delta+1)^2\left(2k\pi+\theta_j\right)}+\frac{2k\pi+\theta_j}{\delta\mu}-\frac{(\delta+1)\left(2k\pi+\theta_j\right)}{\delta^2\mu}\right)\right)\right]\times\left(2k\pi+\theta_j\right)^{-1.5}\right)\Bigg]^2$$
$$\left.\times\left(\sum_{k=0}^{b}\frac{\exp\left(-\frac{\delta}{4}\left[\frac{\delta\mu}{(\delta+1)(2k\pi+\theta_j)}+\frac{(\delta+1)(2k\pi+\theta_j)}{\delta\mu}\right]\right)\left(2k\pi+\frac{\delta\mu}{\delta+1}+\theta_j\right)}{\left(2k\pi+\theta_j\right)^{1.5}}\right)^{-2}\right\}-\frac{n}{2(\delta+1)^2}$$

APÊNDICE B – *STOCHASTIC APPROXIMATION* EM - SAEM

O algoritmo de aproximação estocástica EM (SAEM) é uma alternativa baseada em simulação para o algoritmo EM em situações em que o passo E é difícil ou impossível. Um dos atrativos do SAEM é que, ao contrário de outras versões Monte Carlo do EM, este converge com o número de réplicas fixado (e geralmente pequeno). A ideia básica é que em vez de aumentar continuamente o tamanho da simulação, o SAEM alcança uma aproximação cada vez melhor ao passo E calculando uma média ponderada da aproximação na iteração atual com todas as iterações anteriores. Usando uma sequência decrescente de pesos (ou tamanho do passo), essa abordagem assegura que as informações das iterações anteriores sejam descartadas gradualmente e que seja dada mais ênfase às iterações mais recentes, que geralmente contêm informações mais precisas. Delyon, Lavielle e Moulines (1999) mostraram que este método converge com um tamanho de simulação fixado (e pequeno).

Seja $\boldsymbol{x}$ o vetor de dados observados, seja $\boldsymbol{u}$ o vetor de dados não observados e seja $\boldsymbol{\theta}$ o vetor de parâmetros. Seja $f(\boldsymbol{x}, \boldsymbol{u}; \boldsymbol{\theta})$ a densidade conjunta dos dados completos, $(\boldsymbol{x}, \boldsymbol{u})$, seja $L(\boldsymbol{\theta}; \boldsymbol{x}) = \int f(\boldsymbol{x}, \boldsymbol{u}; \boldsymbol{\theta}) d\boldsymbol{u}$ a função de verossimilhança e seja $\hat{\boldsymbol{\theta}}$ o ponto de máximo de $L(\cdot; \boldsymbol{x})$.

Em cada iteração, o algoritmo EM executa um passo de Esperança (Passo E) e um passo de Maximização (Passo M). Seja $\boldsymbol{\theta}^{(k)}$ o valor corrente do parâmetro. Então, na $k+1$-ésima iteração do algoritmo, o passo E calcula a esperança condicional da log-verossimilhança completa, condicional nos dados observados e no valor corrente do parâmetro,

$$\mathrm{Q}\left(\boldsymbol{\theta}\middle|\boldsymbol{\theta}^{(k)}\right) = \mathbb{E}\left\{\log\left[f(\boldsymbol{x}, \boldsymbol{u}; \boldsymbol{\theta})\right] | \boldsymbol{x}; \boldsymbol{\theta}^{(k)}\right\} = \int \log\left[f(\boldsymbol{x}, \boldsymbol{u}; \boldsymbol{\theta})\right] f(\boldsymbol{u}|\boldsymbol{x}; \boldsymbol{\theta}^{(k)}) d\boldsymbol{u}. \tag{B.1}$$

A $k+1$-ésima atualização do EM, $\boldsymbol{\theta}^{(k+1)}$, maximiza (B.1), isto é, $\boldsymbol{\theta}^{(k+1)}$ satisfaz a seguinte relação

$$\mathrm{Q}\left(\boldsymbol{\theta}^{(k+1)}\middle|\boldsymbol{\theta}^{(k)}\right) \geq \mathrm{Q}\left(\boldsymbol{\theta}\middle|\boldsymbol{\theta}^{(k)}\right)$$

para todo $\boldsymbol{\theta}$ no espaço paramétrico. Este é conhecido como o passo M. Dado um valor inicial $\boldsymbol{\theta}^{(0)}$, o algoritmo EM produz uma sequência $\left\{\boldsymbol{\theta}^{(0)}, \boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots\right\}$ que, sob certas condições de regularidade converge para $\hat{\boldsymbol{\theta}}$ (veja Boyles (1983) e Wu (1983)).

A implementação do EM às vezes é complicada pelo fato de que, ou o passo E ou o passo M (ou ambos) são difíceis (ou impossíveis) teoricamente ou computacionalmente. Enquanto soluções para o passo M têm sido propostas, por exemplo, por Meng e Rubin (1993), que propuseram o algoritmo tipo EM chamado de ECM, o foco aqui é em situações relacionadas ao passo E em que não há soluções em forma fechada para o valor esperado condicional. Wei e Tanner (1990b) propuseram o algoritmo Monte Carlo EM (MCEM) que aproxima (B.1) por uma média de Monte Carlo dada por

$$\tilde{\mathrm{Q}}\left(\boldsymbol{\theta}\middle|\boldsymbol{\theta}^{(k)}\right) = \frac{1}{m_k}\sum_{j=1}^{m_k}\log f(\boldsymbol{x}, \boldsymbol{u}_j; \boldsymbol{\theta}), \tag{B.2}$$

em que $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{m_k}$ são simulados a partir da distribuição condicional $f(\boldsymbol{u}|\boldsymbol{x}; \boldsymbol{\theta}^{(k)})$. Então, pela lei dos grandes números, $\tilde{\mathrm{Q}}\left(\boldsymbol{\theta}\middle|\boldsymbol{\theta}^{(k)}\right)$ será uma aproximação razoável para $\mathrm{Q}\left(\boldsymbol{\theta}\middle|\boldsymbol{\theta}^{(k)}\right)$ se m_k for suficientemente grande. Aproximar (B.1) por (B.2) introduz o erro de Monte Carlo

$$E^{(t)}(m_k) = \left| \int \log f(\boldsymbol{x}, \boldsymbol{u}; \boldsymbol{\theta}) f(\boldsymbol{u}|\boldsymbol{x}; \boldsymbol{\theta}^{(k)}) d\boldsymbol{u} - \frac{1}{m_k}\sum_{j=1}^{m_k}\log f(\boldsymbol{x}, \boldsymbol{u}_j; \boldsymbol{\theta}) \right|,$$

cuja magnitude depende do tamanho da amostra de Monte Carlo m_k. Ao invés de constantemente aumentar o tamanho da amostra de Monte Carlo em todo o algoritmo, existem versões que convergem com um valor constante (e tipicamente pequeno) de m_k.

Seja κ_k uma sequência de tamanhos de passo (*step size*) positivos tal que $\sum \kappa_k = \infty$ e $\sum \kappa_k^2 < \infty$ e defina

$$\begin{aligned} \tilde{P}^{(k+1)}(\boldsymbol{\theta}) &= (1-\kappa_k)\tilde{P}^{(k)}(\boldsymbol{\theta}) + \kappa_k \tilde{\mathrm{Q}}\left(\boldsymbol{\theta}\middle|\boldsymbol{\theta}^{(k)}\right) \\ &= \tilde{P}^{(k)}(\boldsymbol{\theta}) + \kappa_k\left(\frac{1}{m_k}\sum_{j=1}^{m_k}\log f(\boldsymbol{x}, \boldsymbol{u}_k; \boldsymbol{\theta}) - \tilde{P}^{(k)}(\boldsymbol{\theta})\right). \end{aligned} \tag{B.3}$$

$\tilde{P}^{(k+1)}$ é uma combinação convexa da informação de todas as iterações prévias (contida em $\tilde{P}^{(k)}$) e a informação da iteração corrente. Quanto maior o κ_k mais rápido se "esquece" o passado. Seja $\tilde{\boldsymbol{\theta}}^{(k+1)}$ o argmax$\tilde{P}^{(k+1)}(\boldsymbol{\theta})$. O algoritmo definido em (B.3) é o algoritmo de aproximação estocástica EM (SAEM). Uma característica notável é que ele converge com

m_k constante (DELYON; LAVIELLE; MOULINES, 1999). O SAEM também é muito atraente do ponto de vista conceitual pois faz uso de todos os dados: a atualização atual do parâmetro $\tilde{\boldsymbol{\theta}}^{(k+1)}$ utiliza os dados da iteração atual $k+1$ e os dados de todas as iterações anteriores. Outro destaque do método é que, ao menos a princípio, a única decisão a ser tomada é a escolha do tamanho do passo κ_k que é feita uma única vez, usualmente antes do início do algoritmo.

Polyak e Juditsky (1992) mostraram que para tamanhos de passo $\kappa_k \propto (1/k)^{\alpha}$, em que $0.5 < \alpha \leq 1$, o método converge a uma razão ótima (se usado conjuntamente com *offline averaging*). Portanto, a uma primeira vista, o SAEM se mostra mais fácil de implementar que o MCEM que requer uma decisão a respeito do valor de m_k em cada iteração. Enquanto tamanho de passo grande ($\alpha \approx 0.5$) rapidamente faz com que o algoritmo alcance a vizinhança da solução, ao mesmo tempo infla o erro de Monte Carlo. Por outro lado, tamanho de passo pequeno ($\alpha \approx 1$) resulta em uma rápida redução do erro de Monte Carlo ao mesmo tempo que diminui a taxa de convergência do método.

A partir da (B.3) é possível notar que κ_k tem grande influência na velocidade de convergência do algoritmo. Se $\kappa_k = 1 \; \forall \; k$, o SAEM não terá memória e será equivalente ao MCEM, o que faz com que o algoritmo convirja rapidamente para a vizinhança da solução, mas, a convergência para a solução em si é lenta. Galarza, Lachos e Bandyopadhyay (2017) sugeriram o uso da seguinte escolha para κ_k:

$$\kappa_k = \begin{cases} 1 & \text{se } 1 \leq k \leq cS, \\ \\ \frac{1}{k-cS} & \text{se } cS+1 \leq k \leq S, \end{cases}$$

em que S é o número máximo de iterações e c $(0 \leq c \leq 1)$ é um valor limítrofe que determina a porcentagem de iterações iniciais sem memória. Por exemplo, se $c = 0$ o algoritmo terá memória em todas as iterações e portanto convergirá lentamente para a solução. Por outro lado, se $c = 1$, o algoritmo não terá memória e irá convergir rapidamente para a vizinhança da solução mas não tão rápido para a solução em si, assim como o MCEM. No primeiro caso ($c = 0$) o valor de S deve ser grande para garantir que o algoritmo atinja a solução. No segundo caso ($c = 1$) o algoritmo produzirá uma Cadeia de Markov em que a média das observações da cadeia será uma estimativa razoável. Assim, um valor de c entre 0 e 1 $(0 < c < 1)$ irá garantir a convergência para a vizinhança

da solução nas primeiras cS iterações e uma convergência quase-certa para a solução nas demais iterações. Portanto, esta combinação conduzirá o a um algoritmo rápido e que produz estimativas razoavelmente boas (MATTOS; GARAY; LACHOS, 2018). Nas aplicações práticas, os valores de c e S irão variar de acordo com o modelo e com os dados. Para monitorar a convergência das estimativas Galarza, Lachos e Bandyopadhyay (2017) sugeriram um monitoramento gráfico entre avaliações sucessivas da log-verossimilhança e, no caso da classe SMt, os valores escolhidos para c e S foram $0,4$ e 10, respectivamente.

APÊNDICE C – *SAMPLING IMPORTANCE RESAMPLING* - SIR

Foi usado o método SIR, proposto por Rubin (1987) e Rubin (1988), para gerar amostras da distribuição condicional de de U dado X. Este método é útil para gerar uma amostra de tamanho m aproximadamente independente e identicamente distribuída da densidade alvo $f(y)$, em que $y \in \mathcal{S}_Y \subseteq \mathbb{R}$. Portanto, seja $g(y)$ a densidade proposta com mesmo suporte $\mathcal{S}_Y$. O método consiste em dois passos:

• **Passo 1(*Sampling*):** Gere uma amostra aleatória $Y_1, Y_2, \ldots, Y_j$ de $g(y)$ e obtenha os pesos

$$W(Y_j) = \frac{f(Y_j)}{g(Y_j)}, \quad j = 1, \ldots, J$$

e as probabilidades

$$\pi_j = \frac{W(Y_j)}{\sum_{j=1}^{J} W(Y_j)} \quad j = 1, \ldots, J.$$

• **Passo 2(*Importance resampling*):** Sorteie m valores (m muito menor do que J) $Y_1^\star, Y_2^\star, \ldots, Y_m^\star$ dos J valores $Y_1, Y_2, \ldots, Y_J$ com respectivas probabilidades $\pi_1, \pi_2, \ldots, \pi_J$. Rubin (1987) sugere $J/m = 20$.

Na iteração $k+1$ do algoritmo EM, no caso da Weibull-t, a função de densidade objetivo $f(y)$ está dada em (3.15) e a função de densidade proposta $g(y)$ é a função de densidade de uma distribuição Gamma com parâmetros fixados em 1 e $|\hat{\mu}^{k+1}|$. Já no caso da Slash-t, a função de densidade objetivo $f(y)$ está dada em (3.16) e a função de densidade proposta $g(y)$ é a função de densidade de uma distribuição beta com parâmetros $\left|\frac{\hat{\mu}^{k+1}}{1-\hat{\mu}^{k+1}}\right|$ e 1.

APÊNDICE D – DISTRIBUIÇÃO DOS ESTIMADORES DA CLASSE SMT

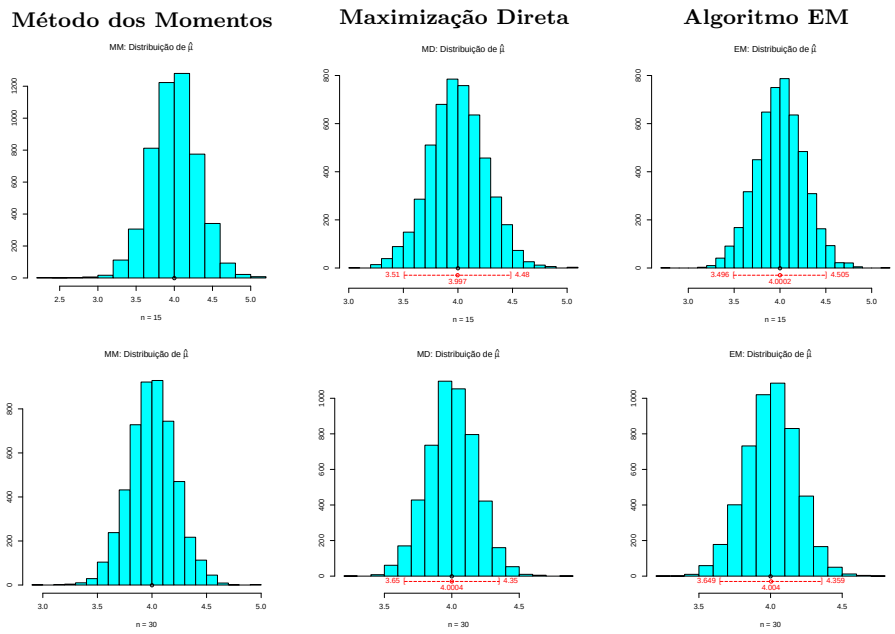


Figura 19 – Distribuição dos estimadores do parâmetro μ no modelo t-Student no Cenário 1 com $n = (15, 30)$

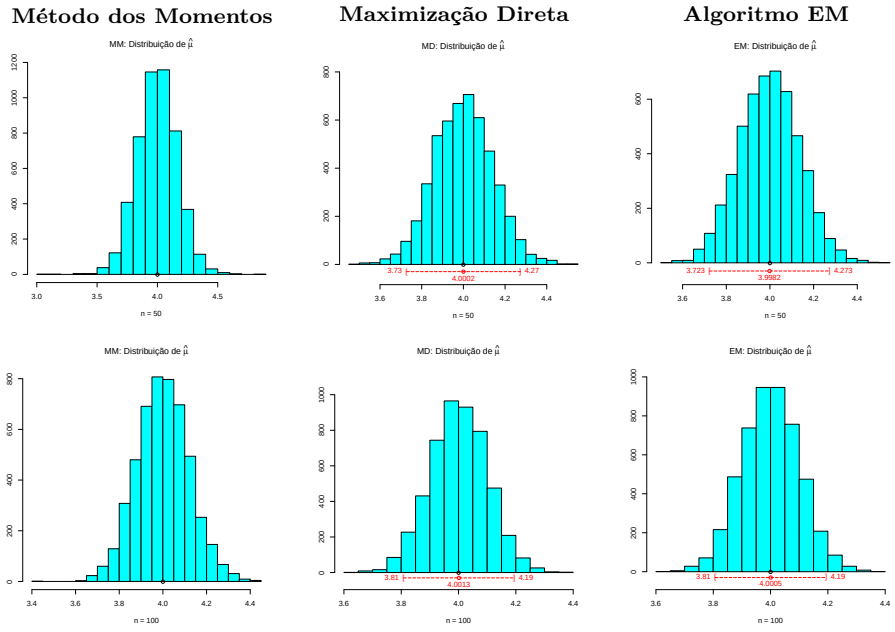


Figura 20 – Distribuição dos estimadores do parâmetro μ no modelo t-Student no Cenário 1 com $n = (50, 100)$

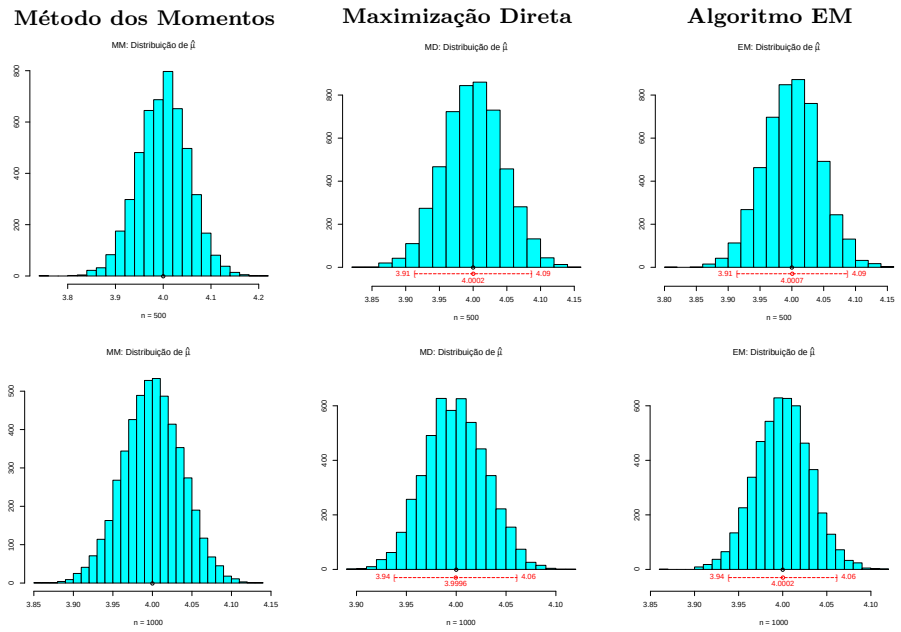


Figura 21 – Distribuição dos estimadores do parâmetro μ no modelo t-Student no Cenário 1 com $n = (500, 1000)$

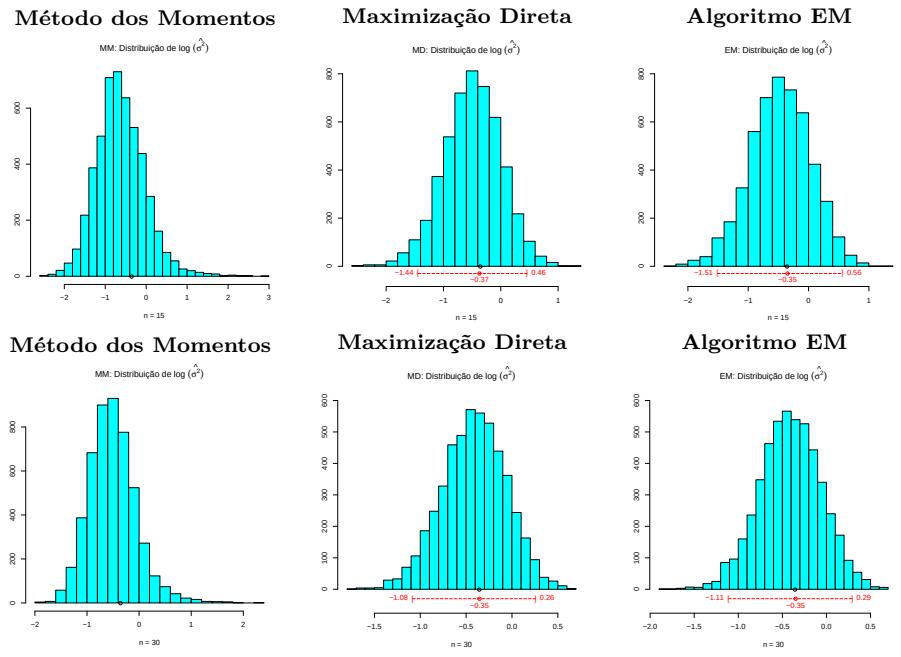


Figura 22 – Distribuição dos estimadores do parâmetro σ^2 no modelo t-Student no Cenário 1 com $n = (15, 30)$

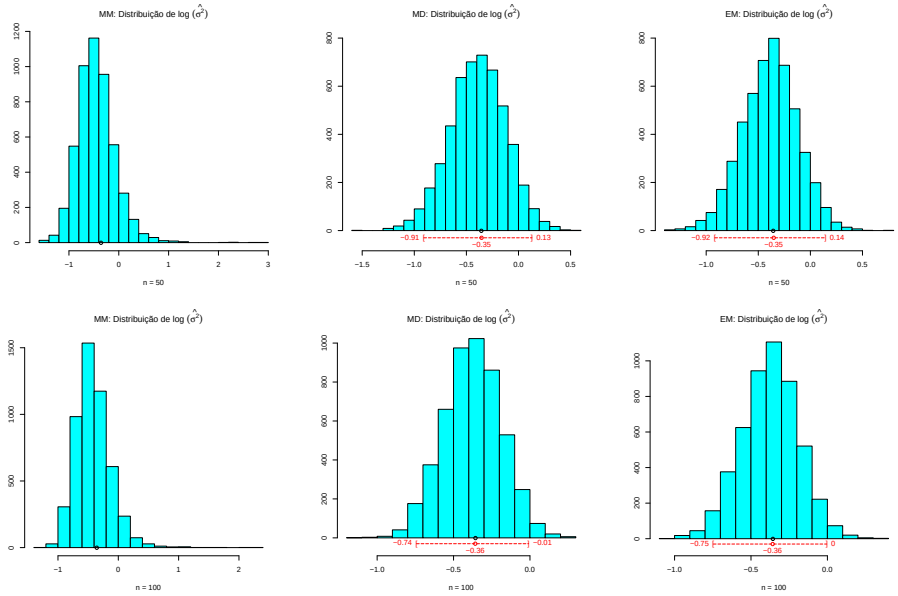


Figura 23 – Distribuição dos estimadores do parâmetro σ^2 no modelo t-Student no Cenário 1 com $n = (50, 100)$

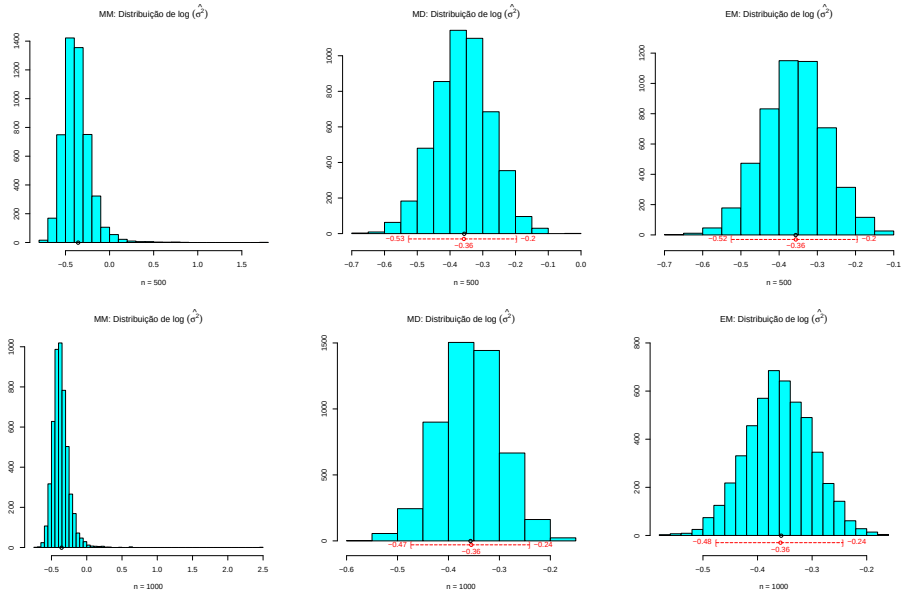


Figura 24 – Distribuição dos estimadores do parâmetro σ^2 no modelo t-Student no Cenário 1 com $n = (500, 1000)$

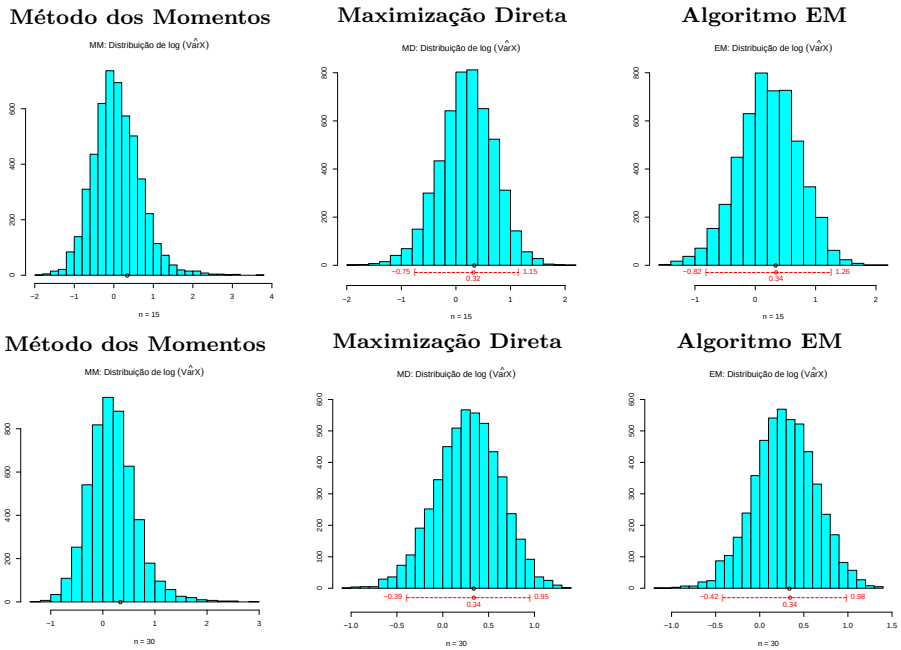


Figura 25 – Distribuição dos estimadores de Var(X) no modelo t-Student no Cenário 1 com $n = (15, 30)$

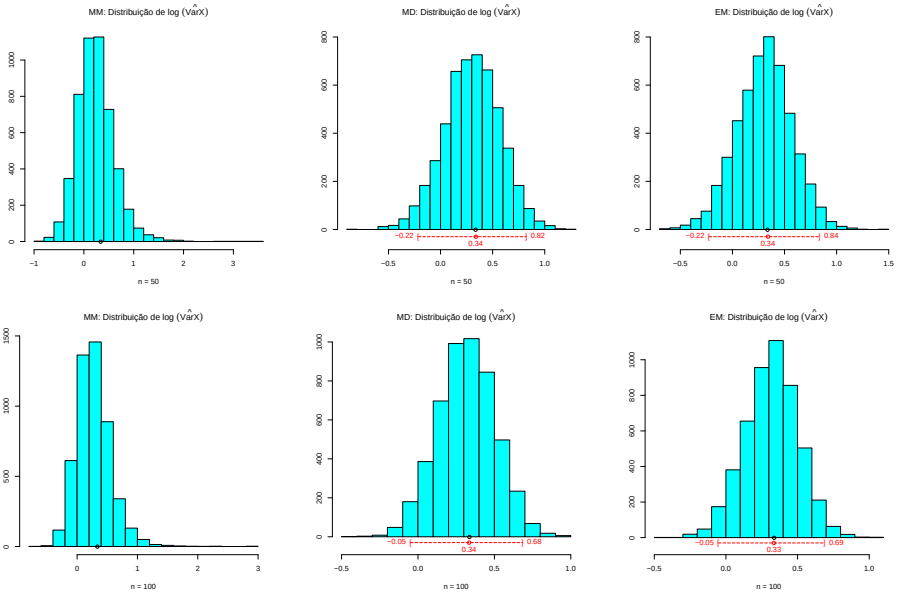


Figura 26 – Distribuição dos estimadores de Var(X) no modelo t-Student no Cenário 1 com $n = (50, 100)$

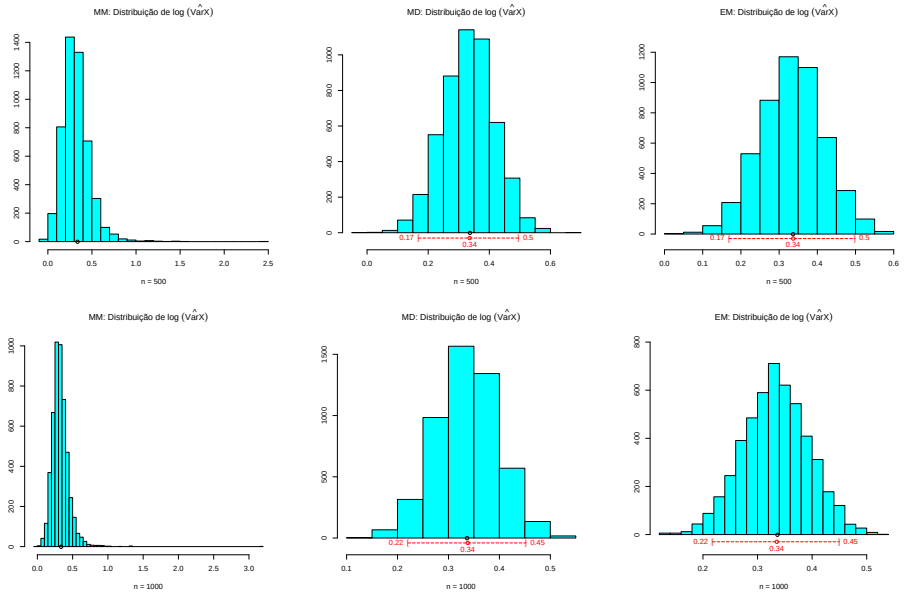


Figura 27 – Distribuição dos estimadores de Var(X) no modelo t-Student no Cenário 1 com $n = (500, 1000)$

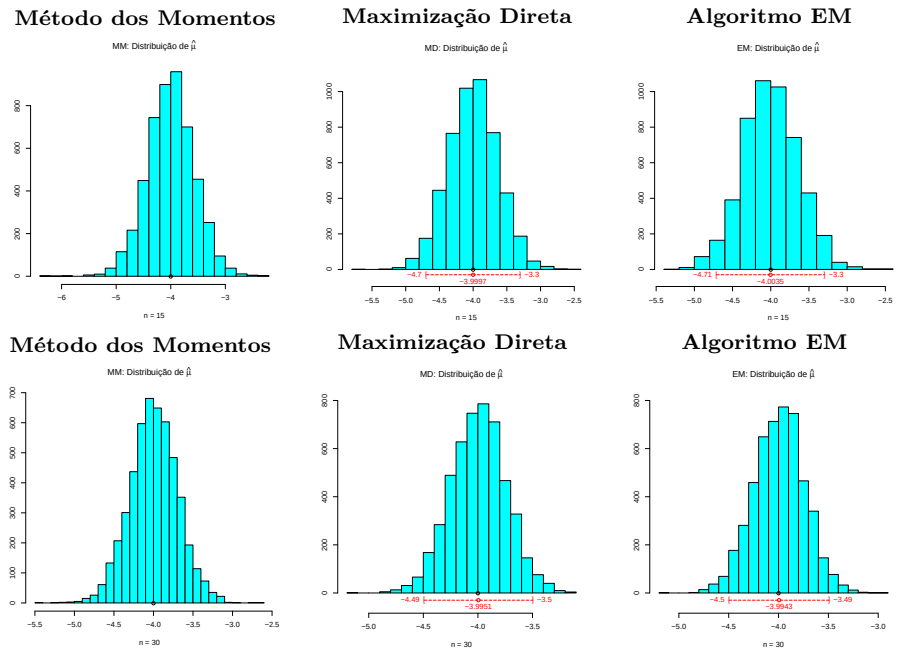


Figura 28 – Distribuição dos estimadores de μ no modelo t-Student no Cenário 2 com $n = (15, 30)$

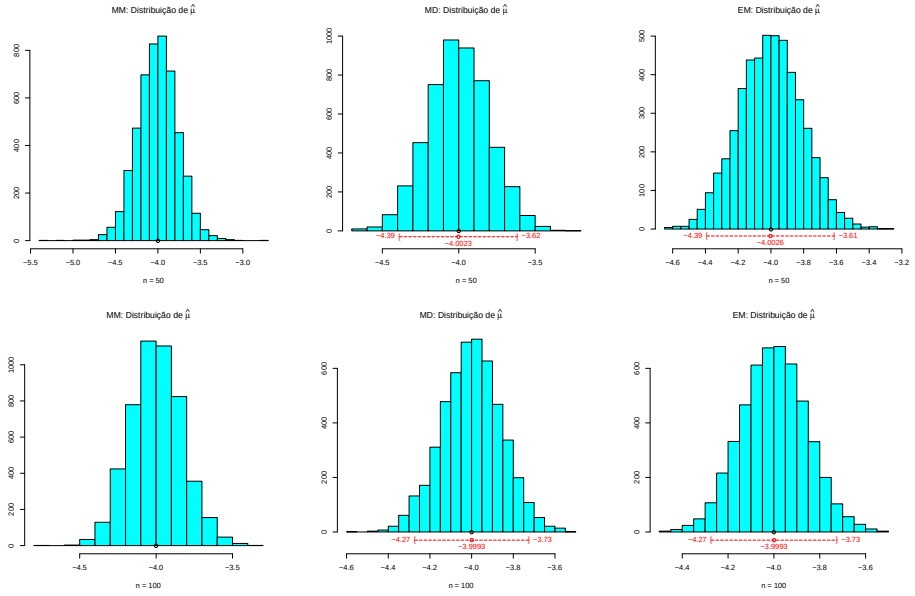


Figura 29 – Distribuição dos estimadores de μ no modelo t-Student no Cenário 2 com $n = (100, 500)$

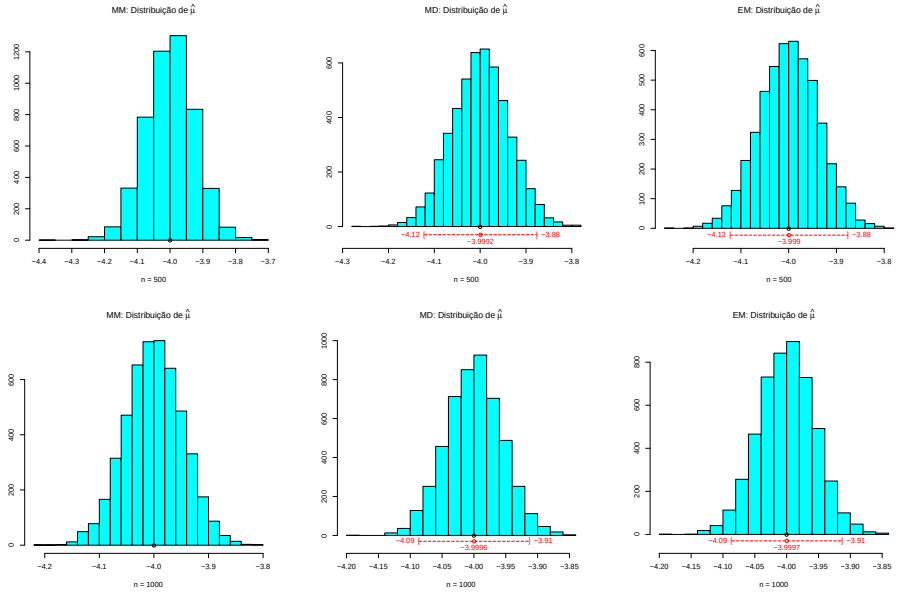


Figura 30 – Distribuição dos estimadores de μ no modelo t-Student no Cenário 2 com $n = (500, 1000)$

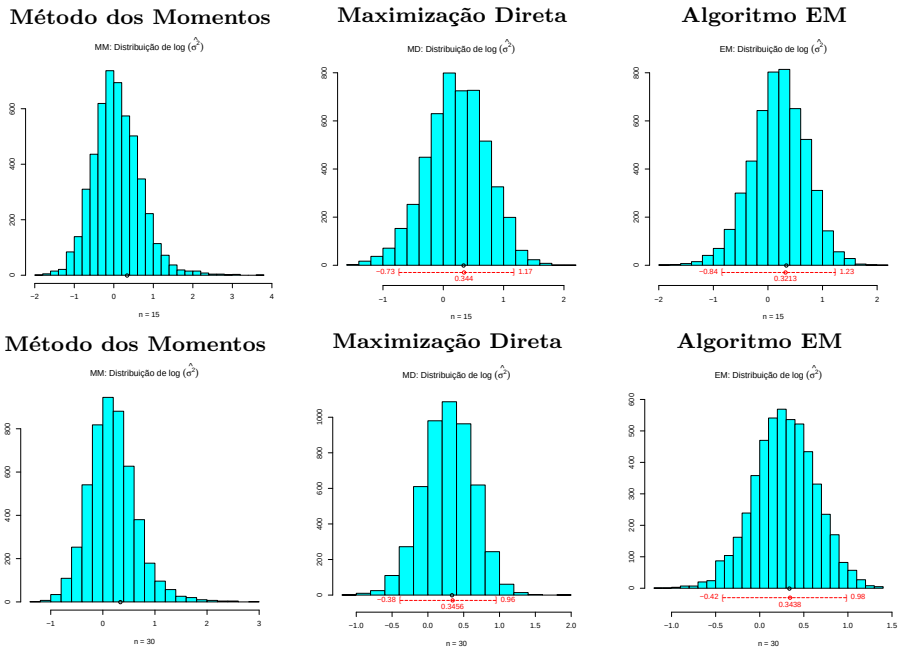


Figura 31 – Distribuição dos estimadores de σ^2 no modelo t-Student no Cenário 2 com $n = (15, 30)$

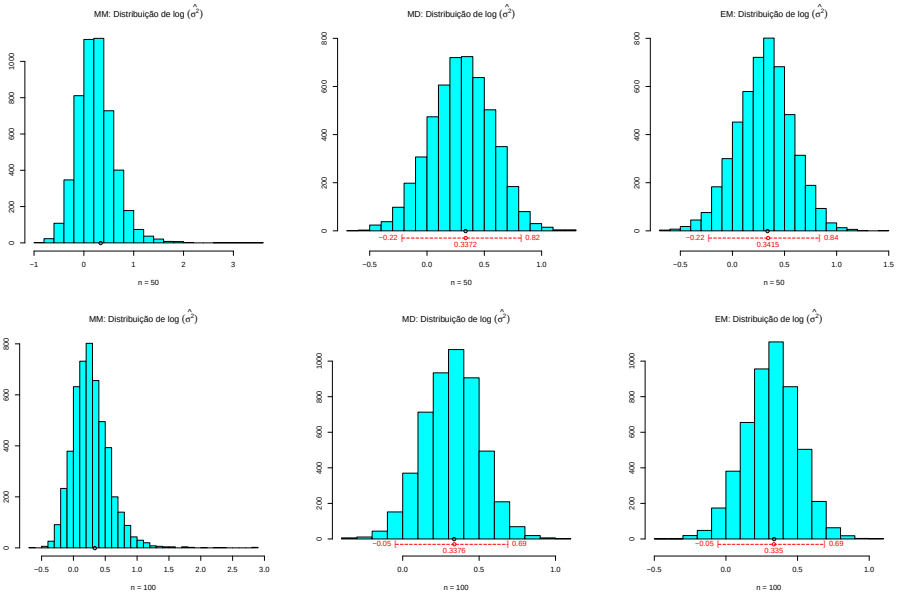


Figura 32 – Distribuição dos estimadores de σ^2 no modelo t-Student no Cenário 2 com $n = (100, 500)$

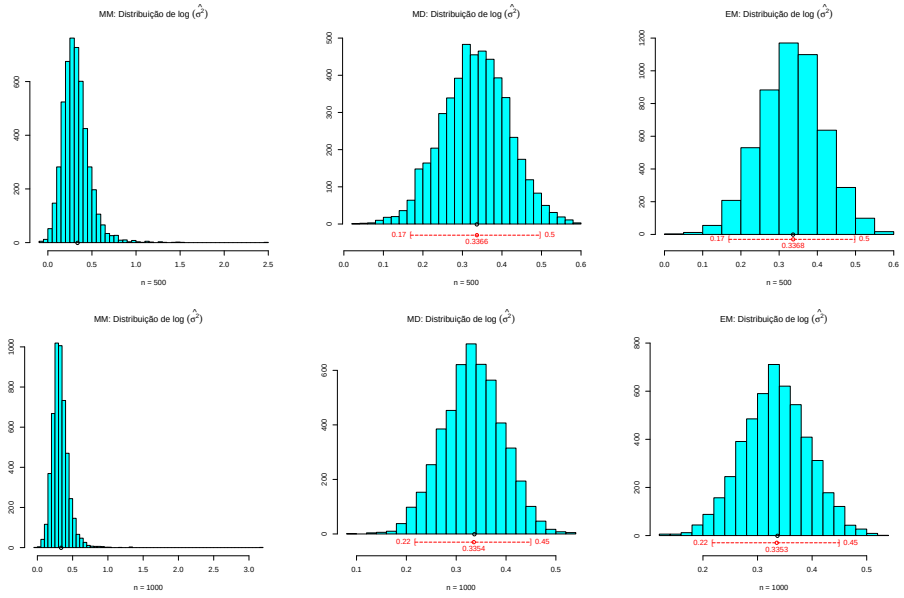


Figura 33 – Distribuição dos estimadores de σ^2 no modelo t-Student no Cenário 2 com $n = (500, 1000)$

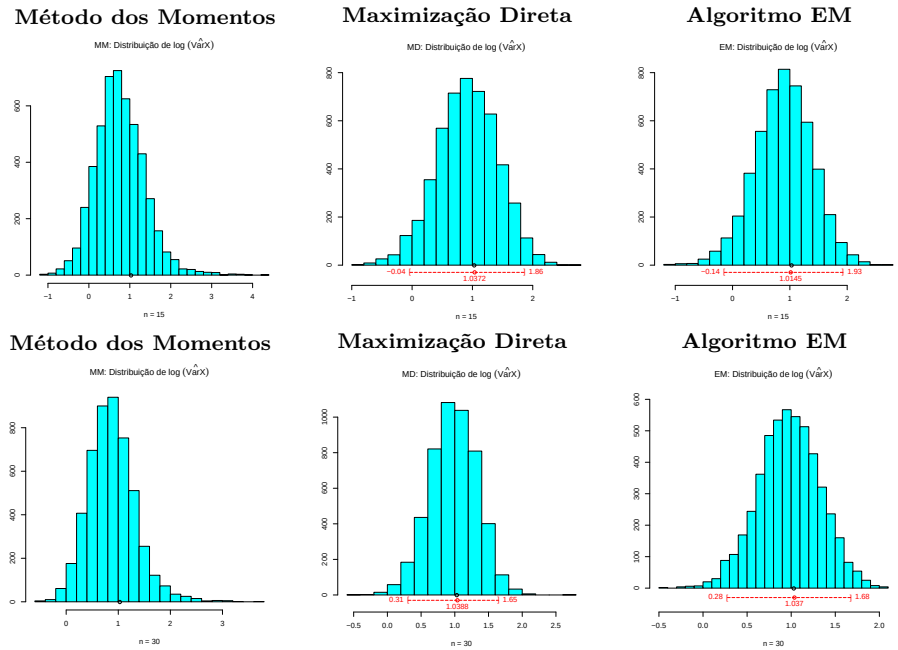


Figura 34 – Distribuição dos estimadores de Var(X) no modelo t-Student no Cenário 2 com $n = (15, 30)$

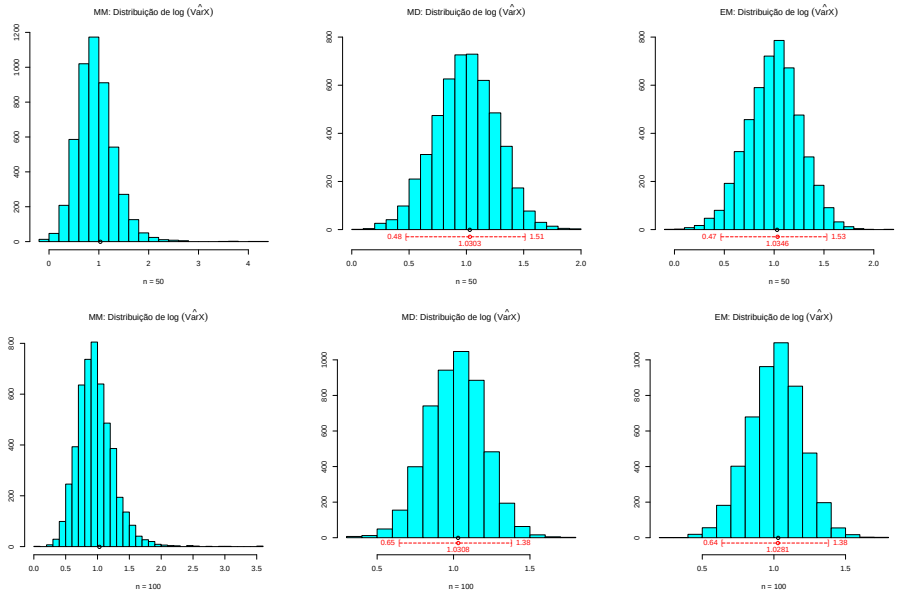


Figura 35 – Distribuição dos estimadores de Var(X) no modelo t-Student no Cenário 2 com $n = (100, 500)$

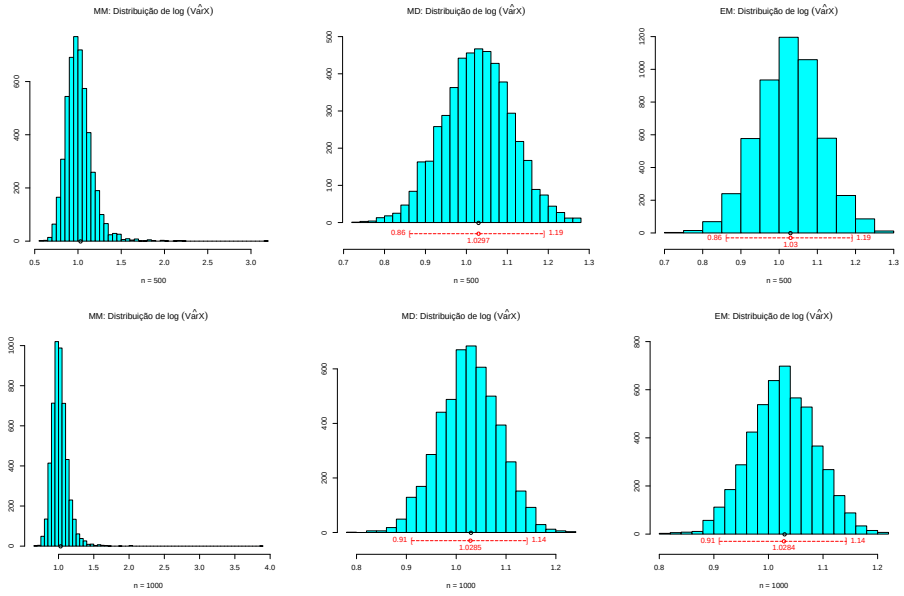


Figura 36 – Distribuição dos estimadores de Var(X) no modelo t-Student no Cenário 2 com $n = (500, 1000)$

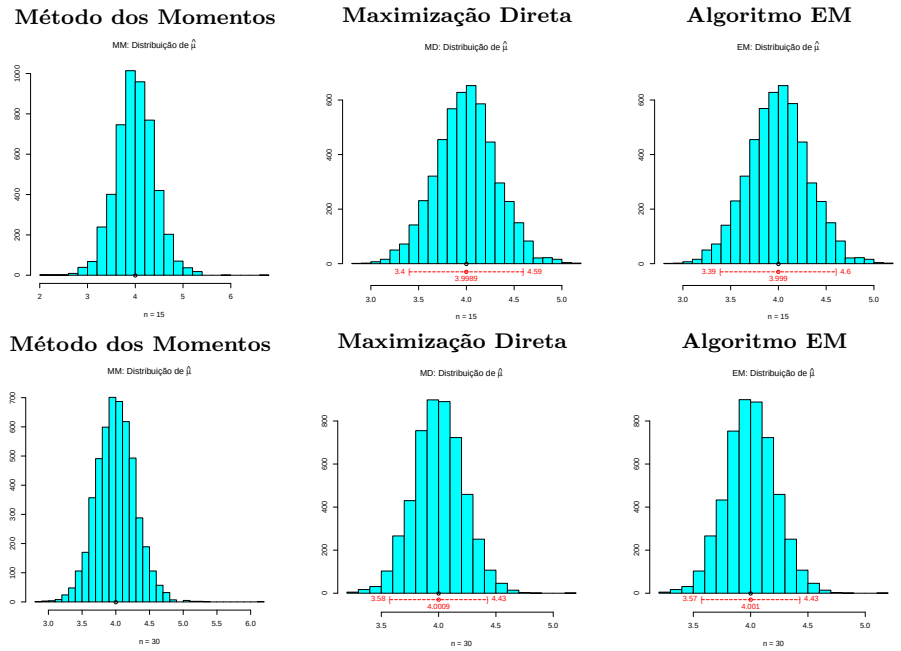


Figura 37 – Distribuição dos estimadores do parâmetro μ do modelo t-Contaminada no Cenário 1 com $n = (15, 30)$

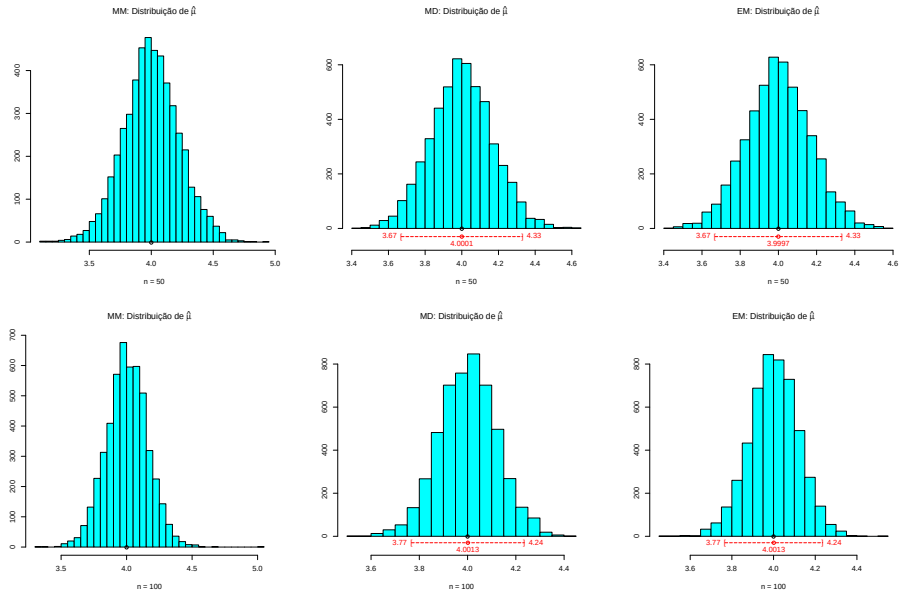


Figura 38 – Distribuição dos estimadores do parâmetro μ do modelo t-Contaminada no Cenário 1 com $n = (50, 100)$

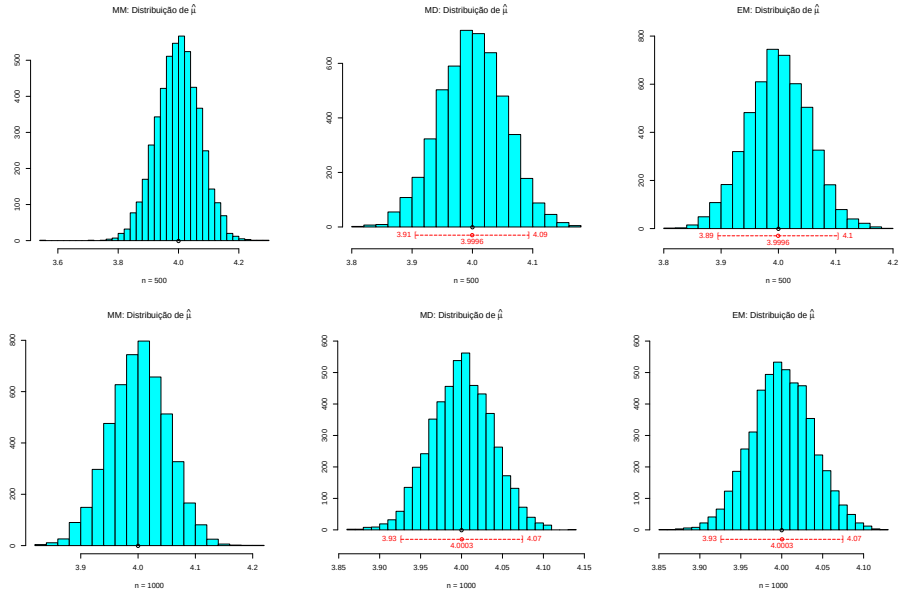


Figura 39 – Distribuição dos estimadores do parâmetro μ do modelo t-Contaminada no Cenário 1 com $n = (500, 1000)$

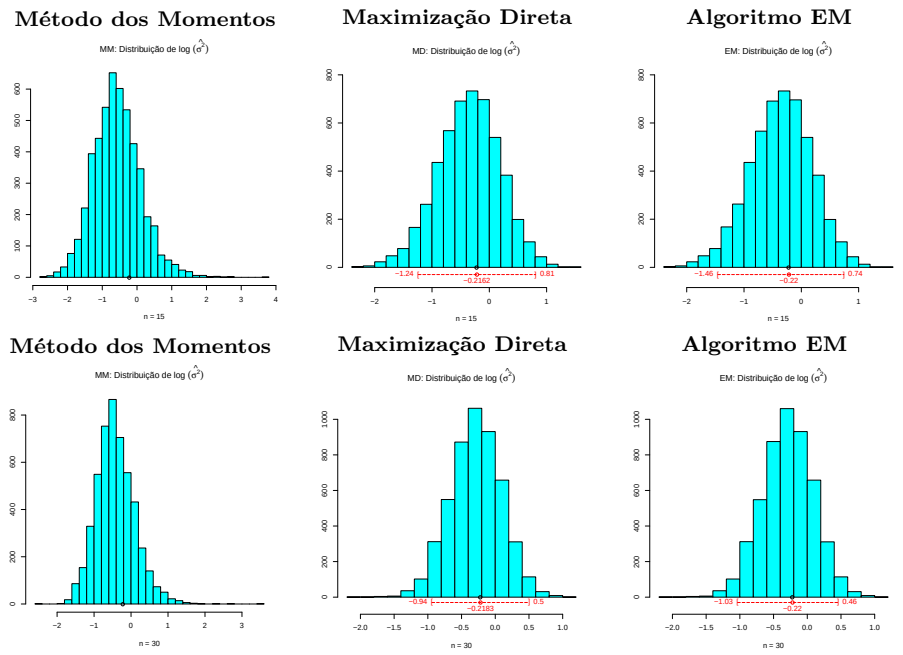


Figura 40 – Distribuição dos estimadores do parâmetro σ^2 do modelo t-Contaminada no Cenário 1 com $n = (15, 30)$

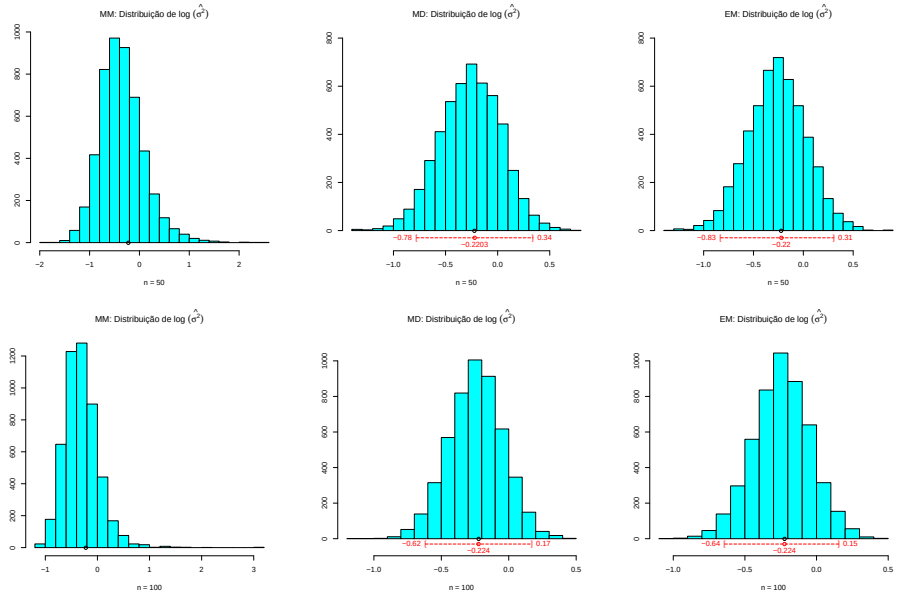


Figura 41 – Distribuição dos estimadores do parâmetro σ^2 do modelo t-Contaminada no Cenário 1 com $n = (50, 100)$

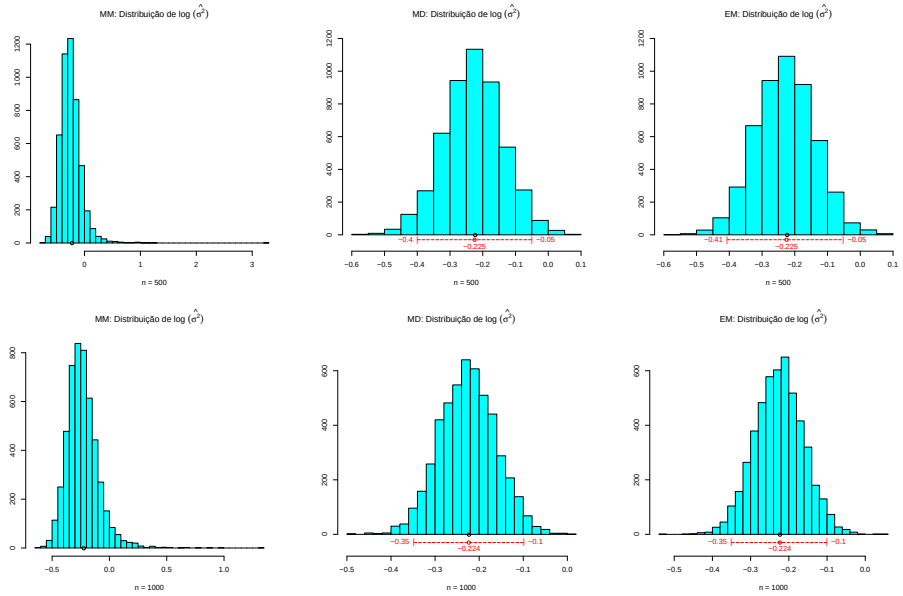


Figura 42 – Distribuição dos estimadores do parâmetro σ^2 do modelo t-Contaminada no Cenário 1 com $n = (500, 1000)$

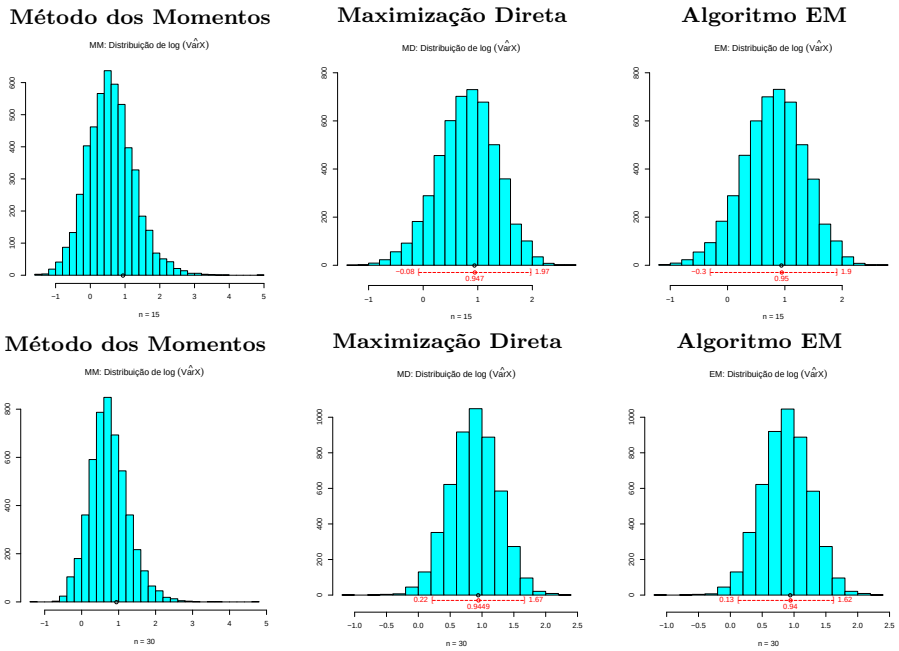


Figura 43 – Distribuição dos estimadores de Var(X) no modelo t-Contaminada no Cenário 1 com $n = (15, 30)$

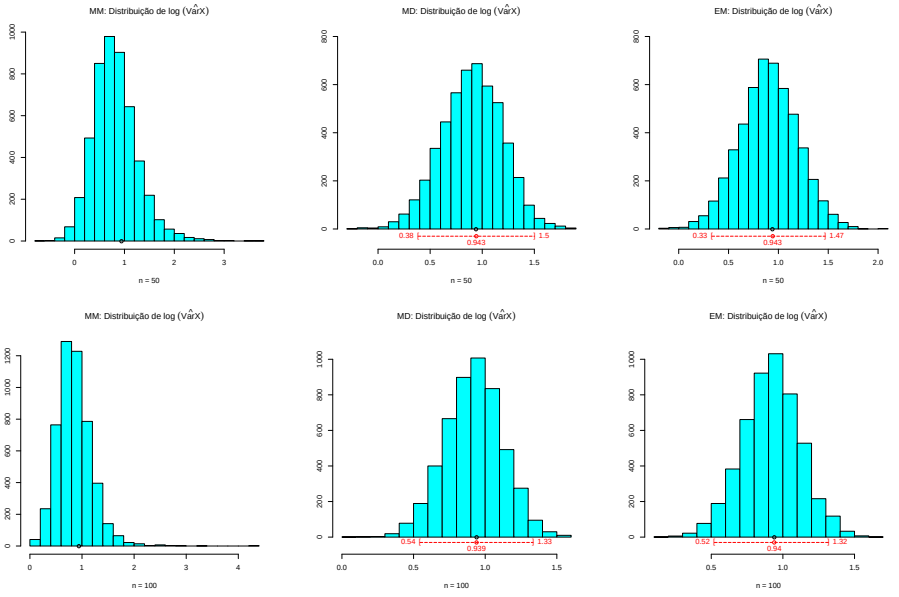


Figura 44 – Distribuição dos estimadores de Var(X) no modelo t-Contaminada no Cenário 1 com $n = (50, 100)$

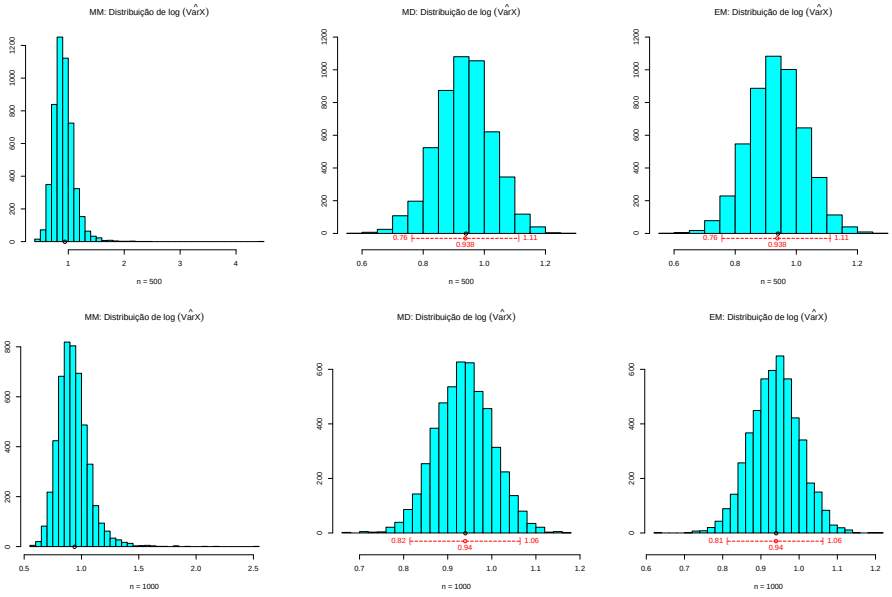


Figura 45 – Distribuição dos estimadores de Var(X) no modelo t-Contaminada no Cenário 1 com $n = (500, 1000)$

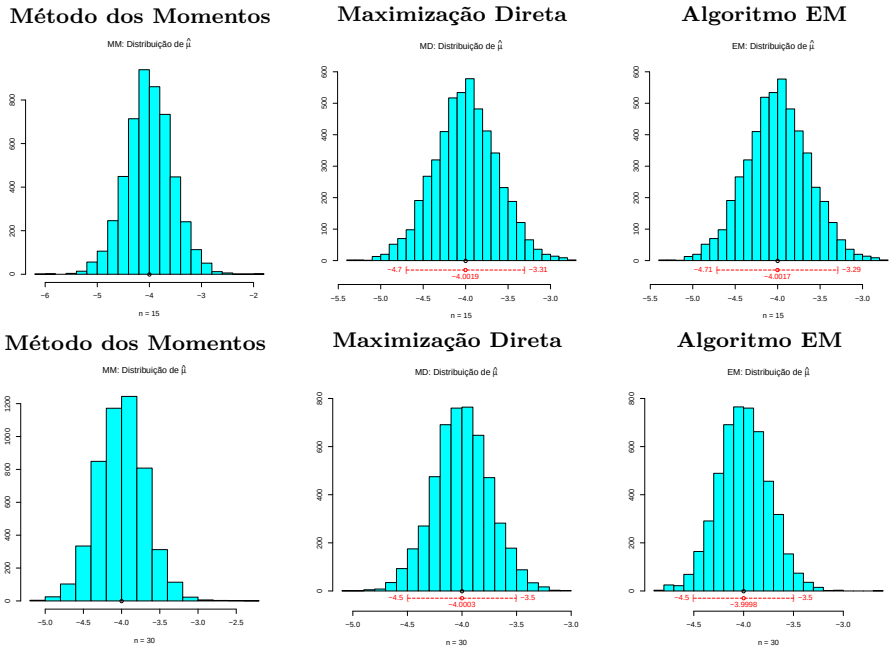


Figura 46 – Distribuição dos estimadores do parâmetro μ do modelo t-Contaminada no Cenário 2 com $n = (15, 30)$

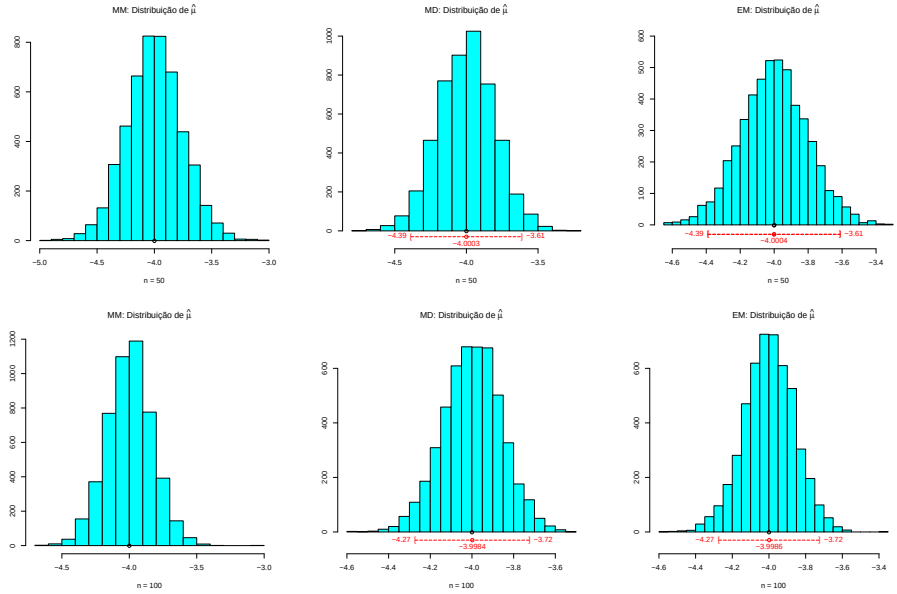


Figura 47 – Distribuição dos estimadores do parâmetro μ do modelo t-Contaminada no Cenário 2 com $n = (50, 100)$

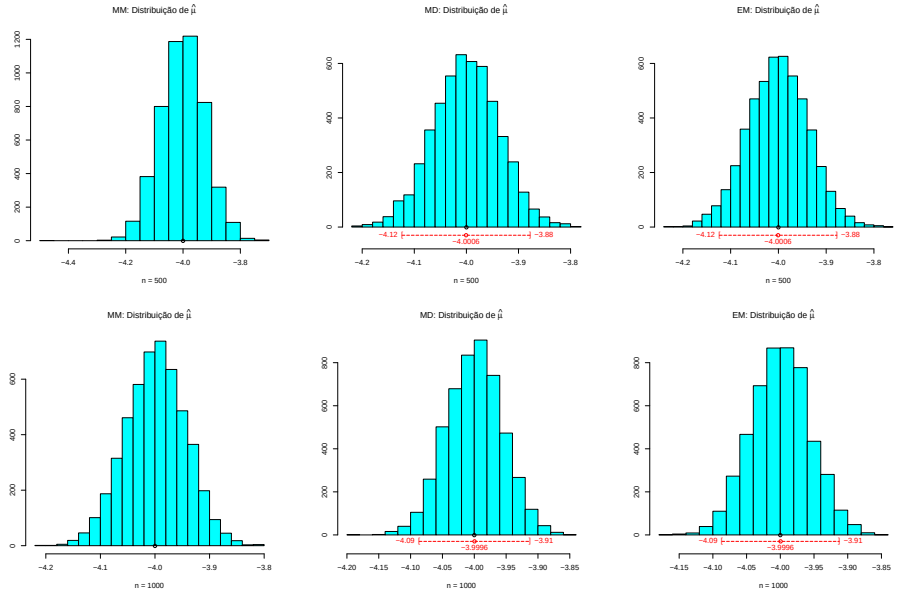


Figura 48 – Distribuição dos estimadores do parâmetro μ do modelo t-Contaminada no Cenário 2 com $n = (500, 1000)$

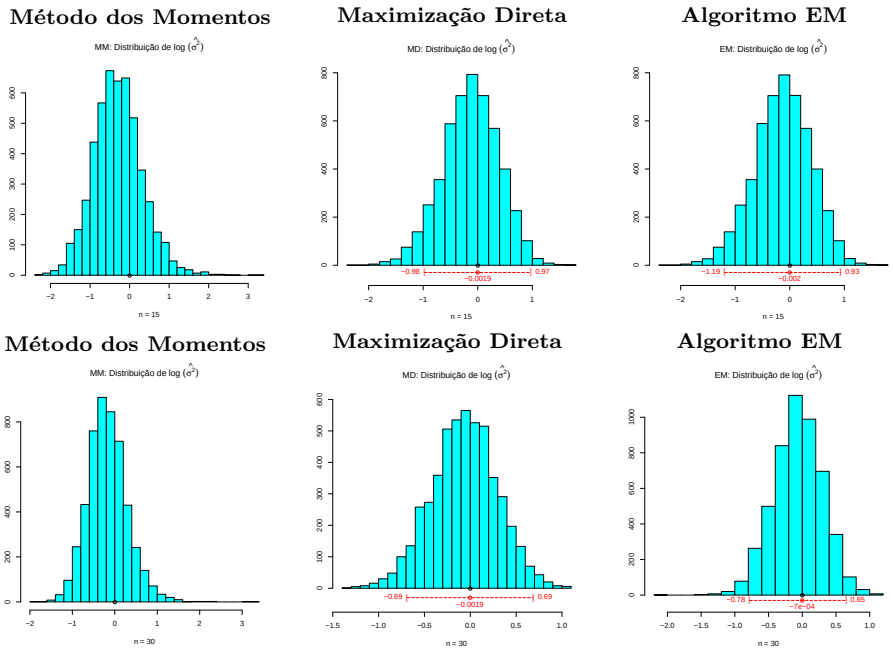


Figura 49 – Distribuição dos estimadores do parâmetro σ^2 do modelo t-Contaminada no Cenário 2 com $n = (15, 30)$

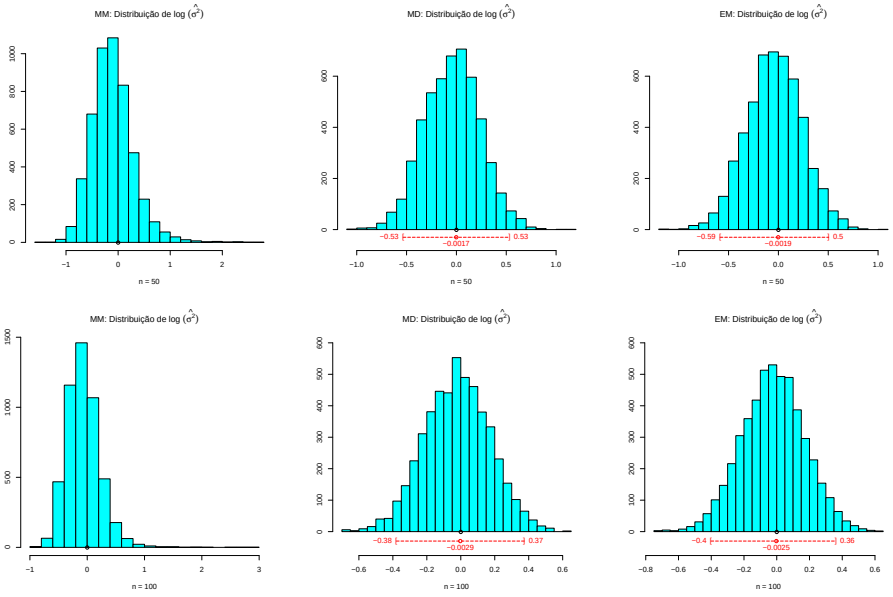


Figura 50 – Distribuição dos estimadores do parâmetro σ^2 do modelo t-Contaminada no Cenário 2 com $n = (50, 100)$

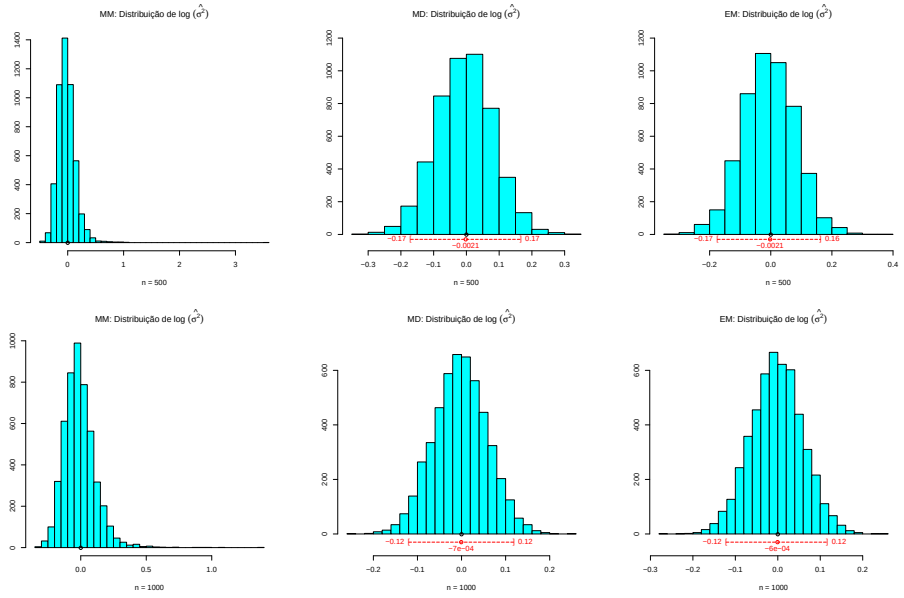


Figura 51 – Distribuição dos estimadores do parâmetro σ^2 do modelo t-Contaminada no Cenário 2 com $n = (500, 1000)$

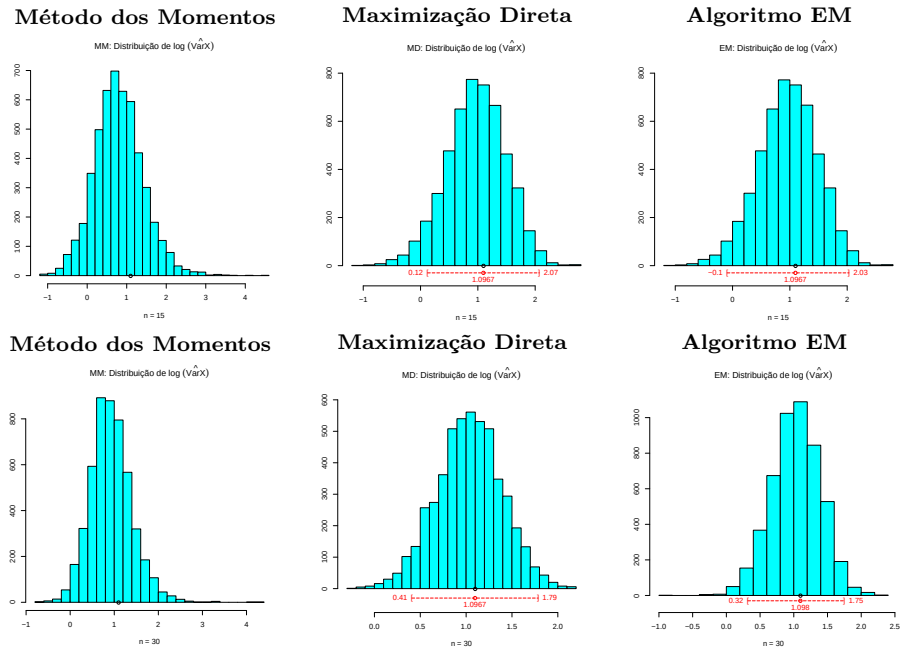


Figura 52 – Distribuição dos estimadores de Var(X) no modelo t-Contaminada no Cenário 2 com $n = (15, 30)$

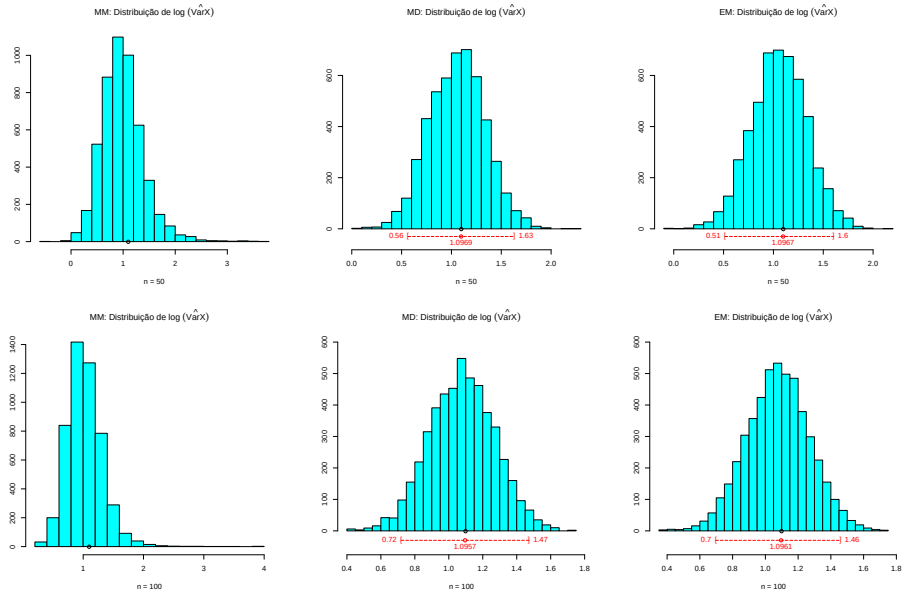


Figura 53 – Distribuição dos estimadores de Var(X) no modelo t-Contaminada no Cenário 2 com $n = (50, 100)$

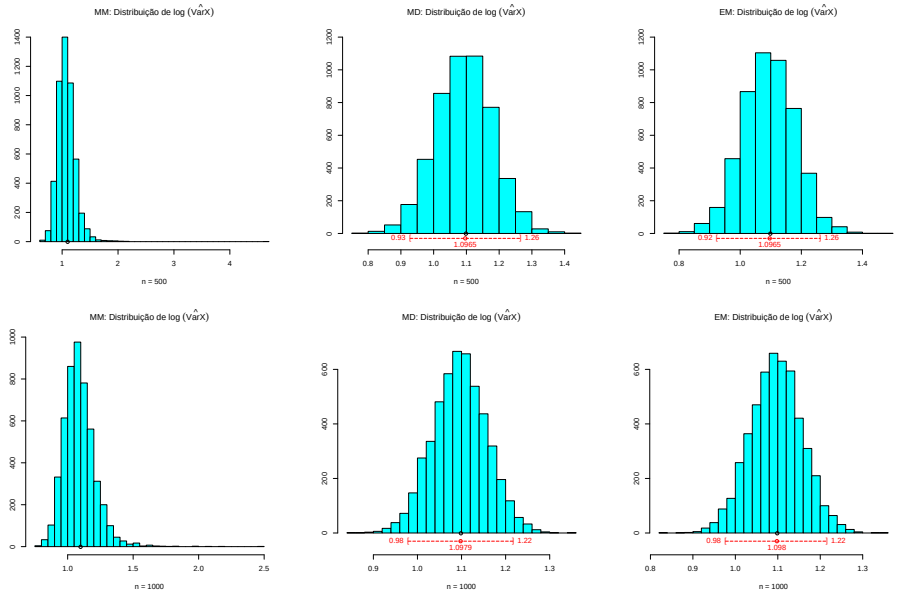


Figura 54 – Distribuição dos estimadores de Var(X) no modelo t-Contaminada no Cenário 2 com $n = (500, 1000)$

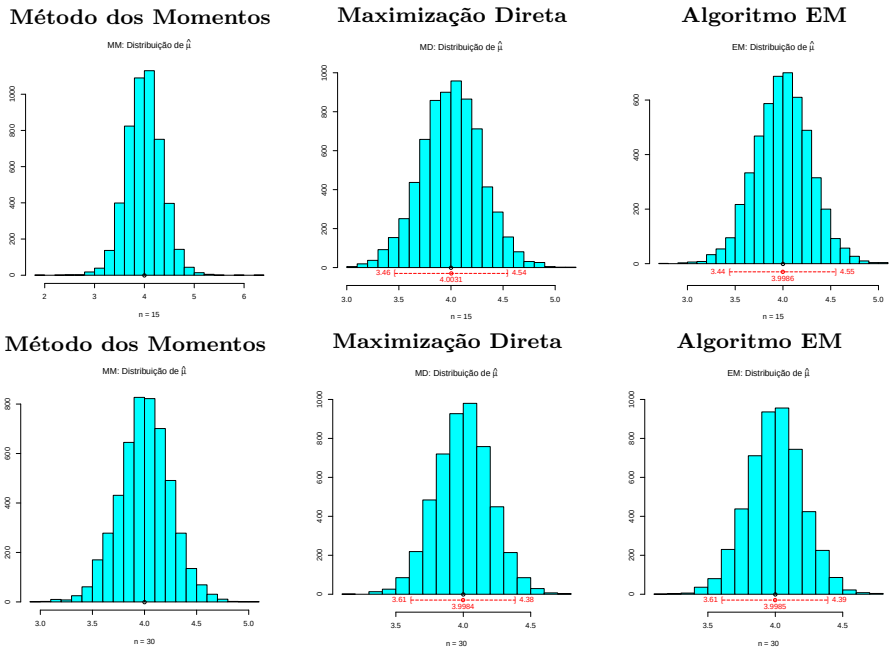


Figura 55 – Distribuição dos estimadores do parâmetro μ do modelo Weibull-t no Cenário 1 com $n = (15, 30)$

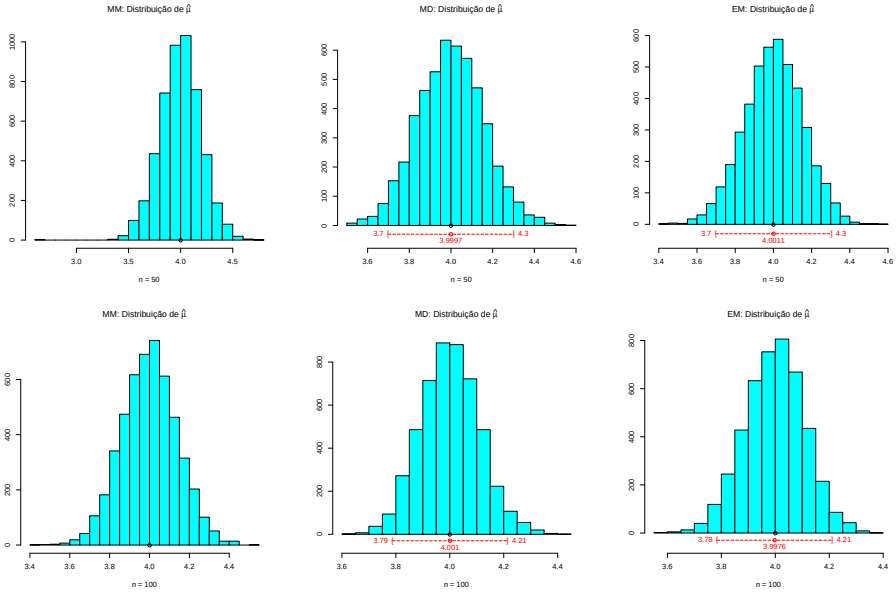


Figura 56 – Distribuição dos estimadores do parâmetro μ do modelo Weibull-t no Cenário 1 com $n = (50, 100)$

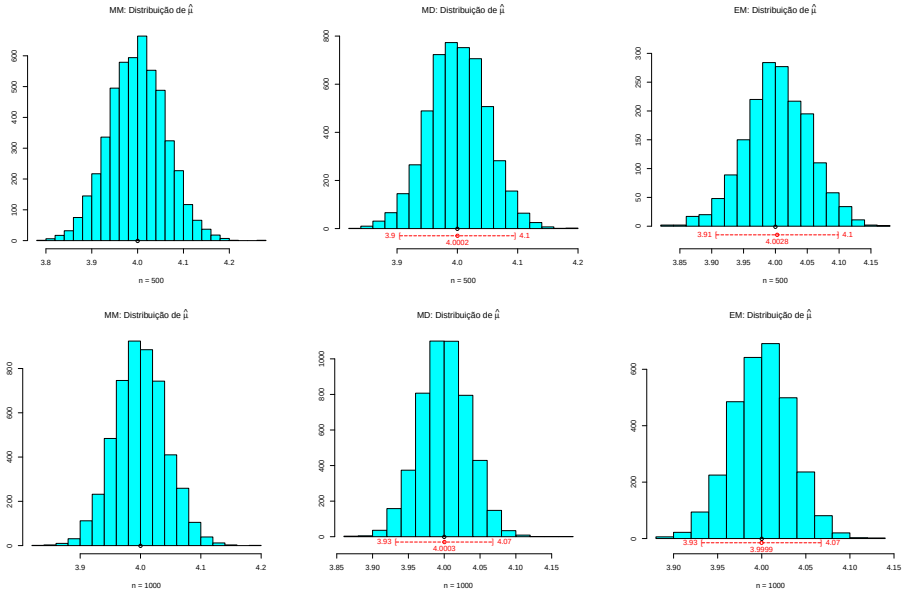


Figura 57 – Distribuição dos estimadores do parâmetro μ do modelo Weibull-t no Cenário 1 com $n = (500, 1000)$

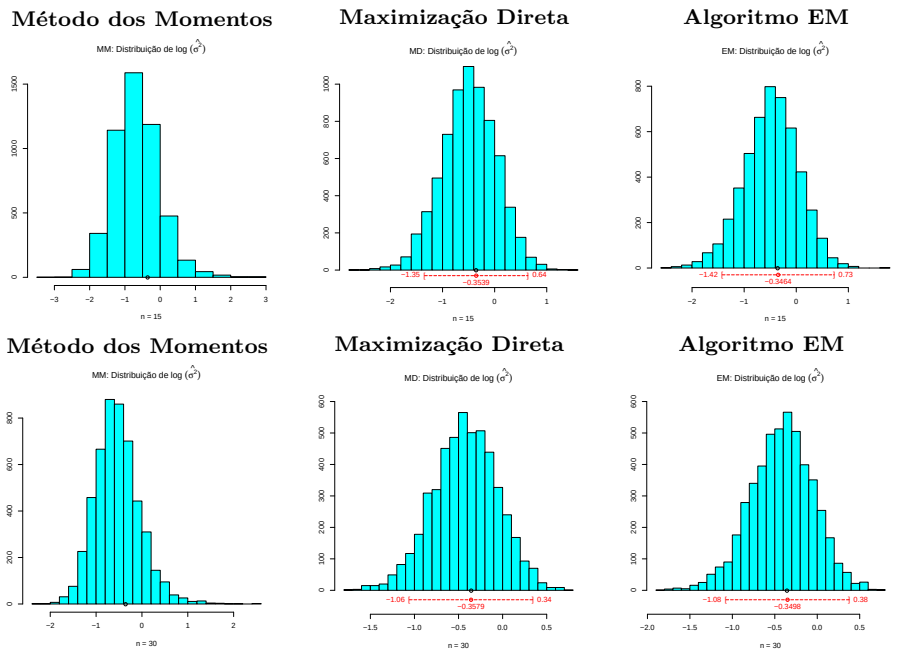


Figura 58 – Distribuição dos estimadores do parâmetro σ^2 do modelo Weibull-t no Cenário 1 com $n = (15, 30)$

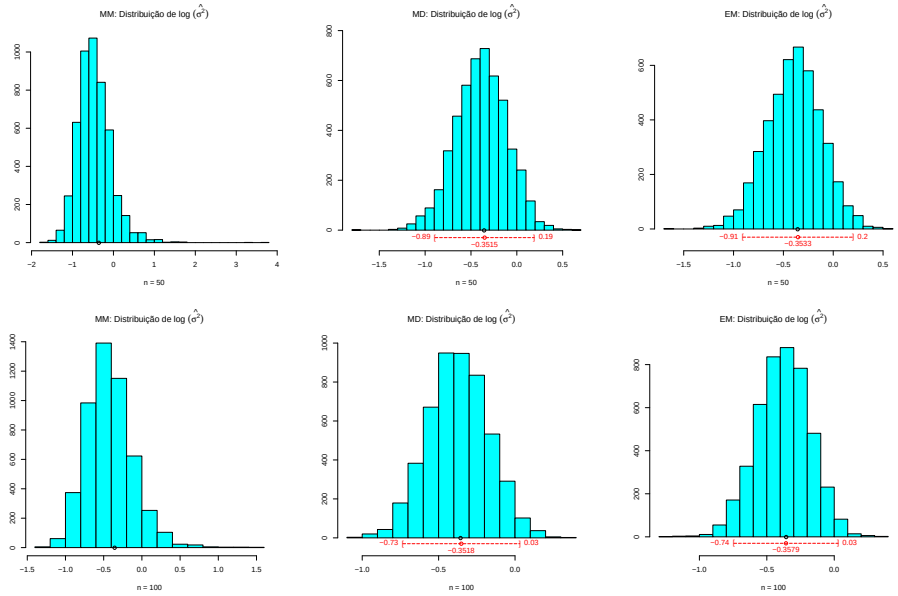


Figura 59 – Distribuição dos estimadores do parâmetro σ^2 do modelo Weibull-t no Cenário 1 com $n = (50, 100)$

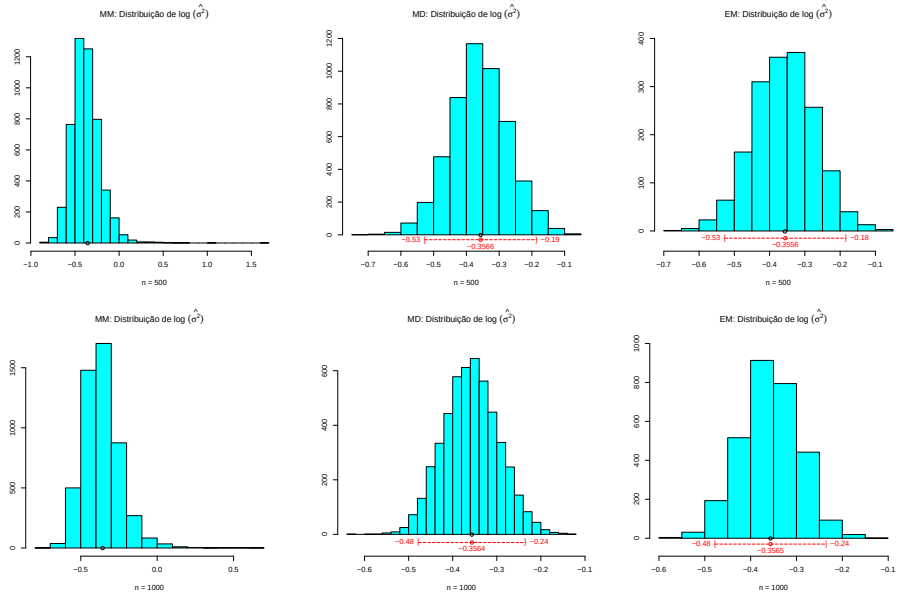


Figura 60 – Distribuição dos estimadores do parâmetro σ^2 do modelo Weibull-t no Cenário 1 com $n = (500, 1000)$

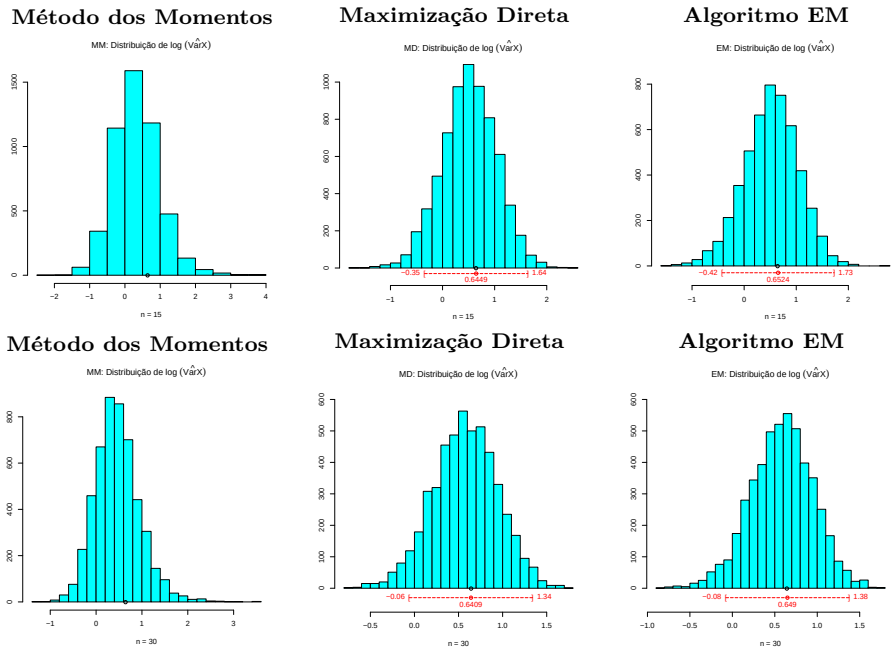


Figura 61 – Distribuição dos estimadores de Var(X) no modelo Weibull-t no Cenário 1 com $n = (15, 30)$

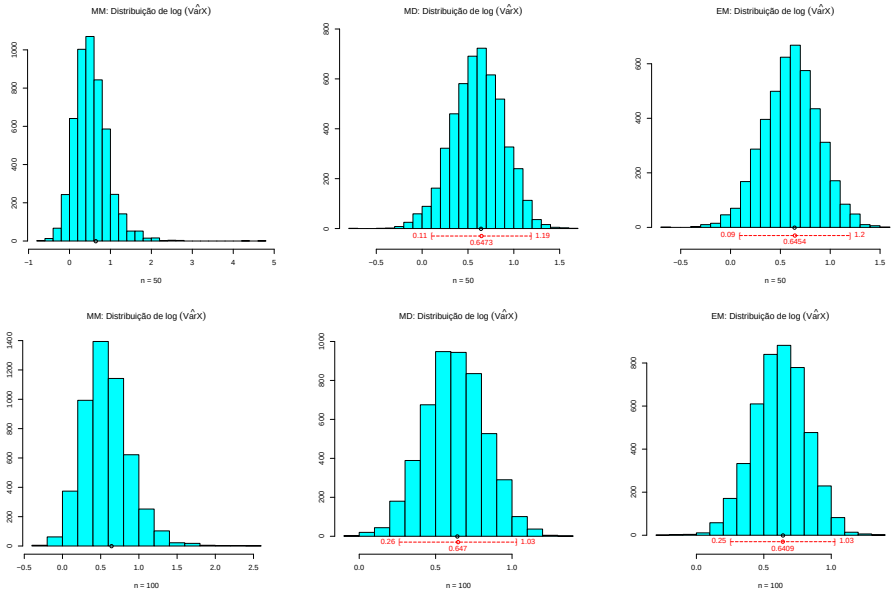


Figura 62 – Distribuição dos estimadores de Var(X) no modelo Weibull-t no Cenário 1 com $n = (50, 100)$

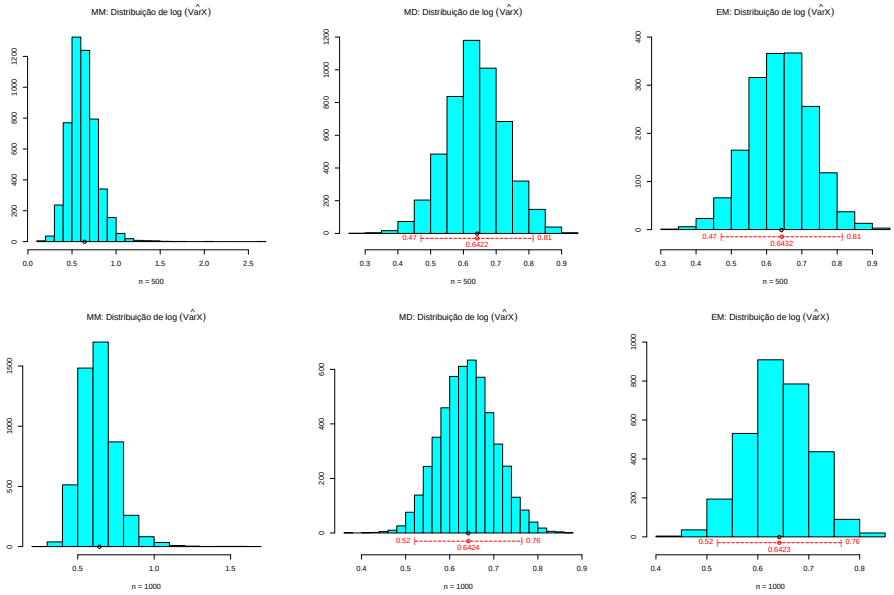


Figura 63 – Distribuição dos estimadores de Var(X) no modelo Weibull-t no Cenário 1 com $n = (500, 1000)$

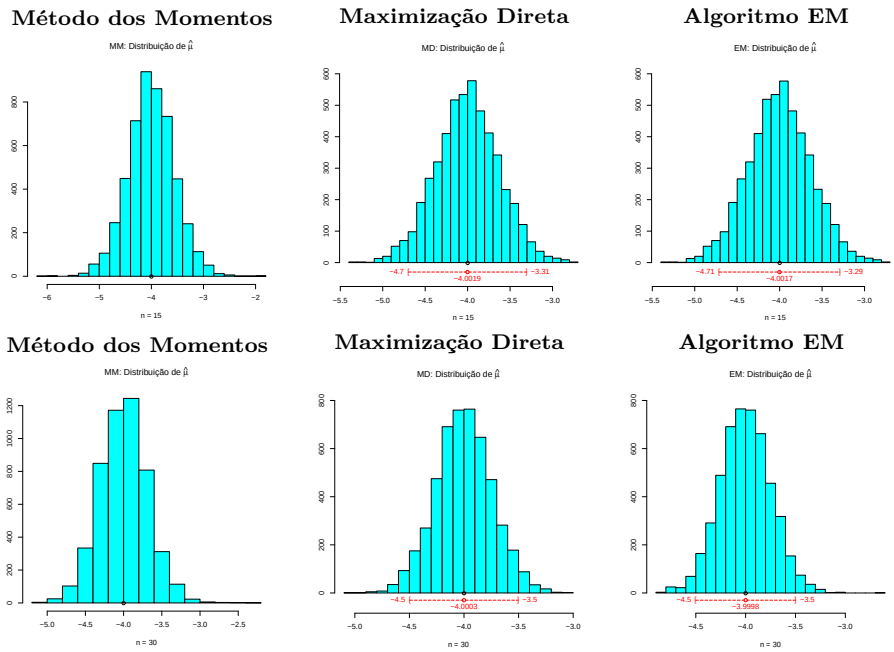


Figura 64 – Distribuição dos estimadores do parâmetro μ do modelo Weibull-t no Cenário 2 com $n = (15, 30)$

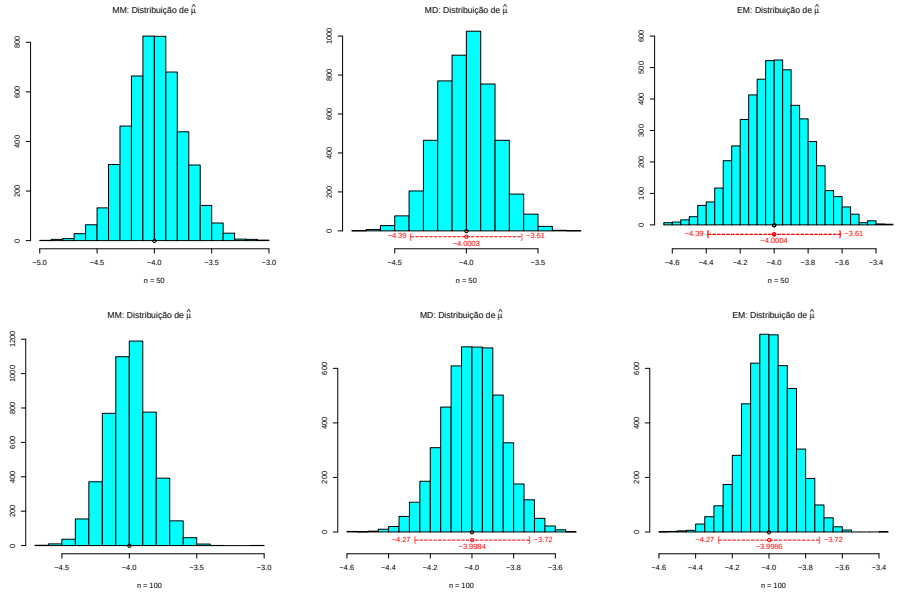


Figura 65 – Distribuição dos estimadores do parâmetro μ do modelo Weibull-t no Cenário 2 com $n = (50, 100)$

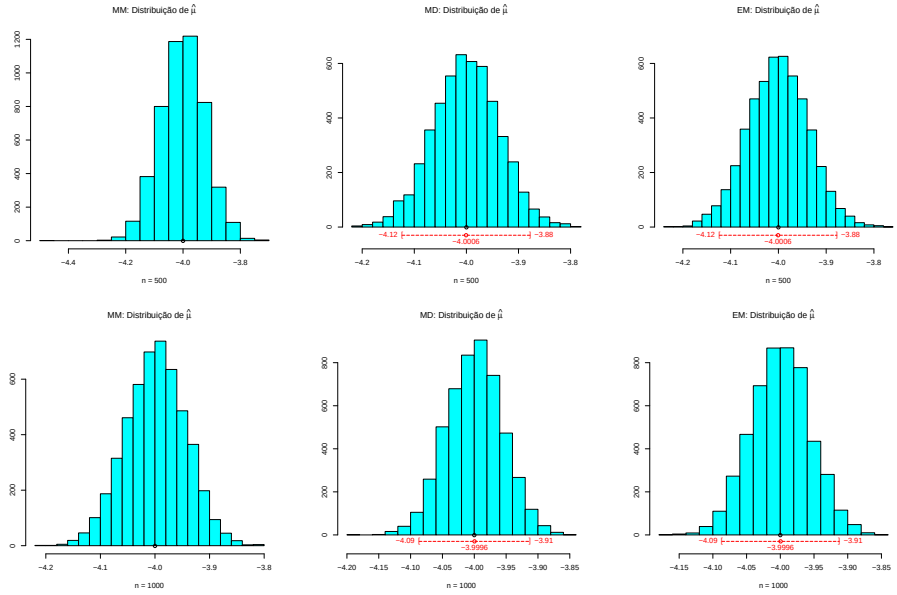


Figura 66 – Distribuição dos estimadores do parâmetro μ do modelo Weibull-t no Cenário 2 com $n = (500, 1000)$

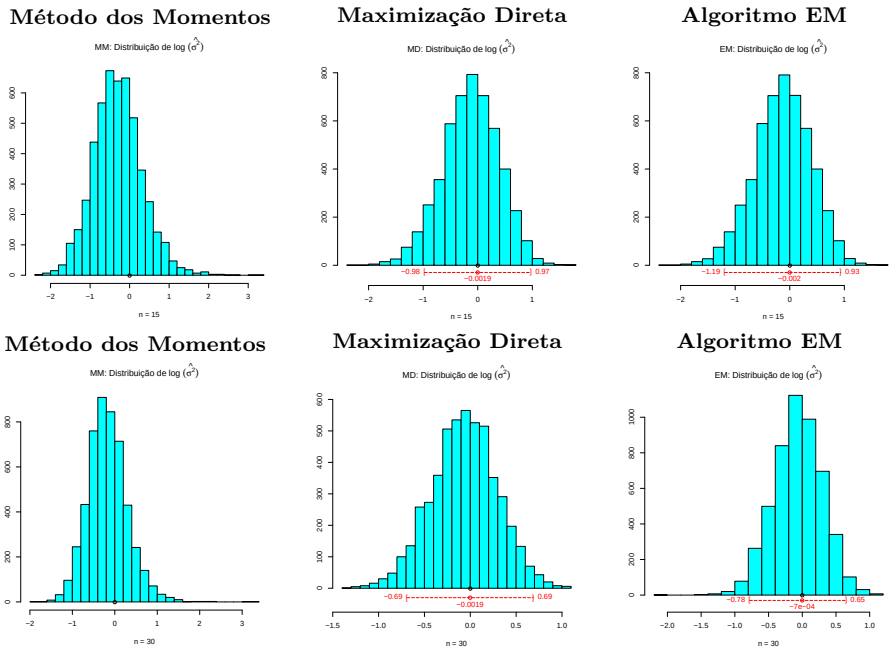


Figura 67 – Distribuição dos estimadores do parâmetro σ^2 do modelo Weibull-t no Cenário 2 com $n = (15, 30)$

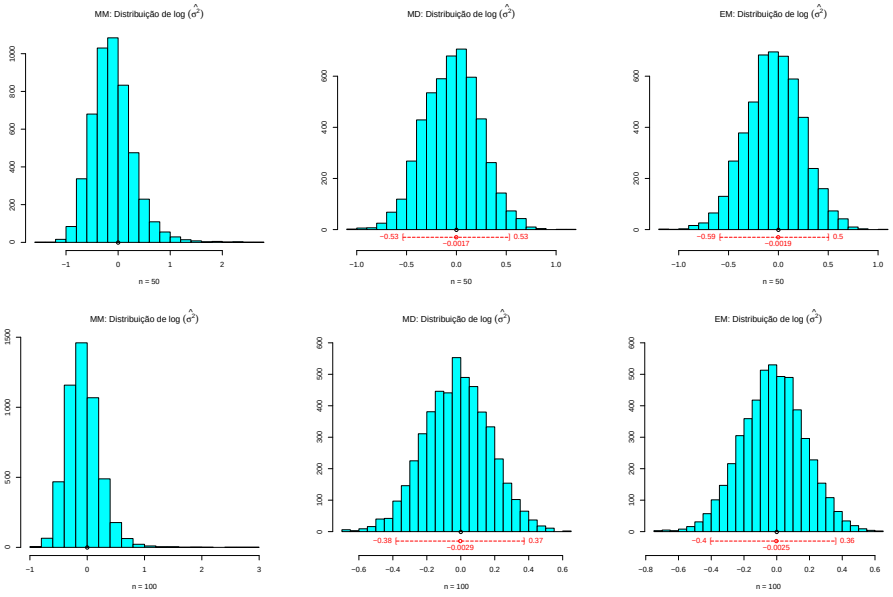


Figura 68 – Distribuição dos estimadores do parâmetro σ^2 do modelo Weibull-t no Cenário 2 com $n = (50, 100)$

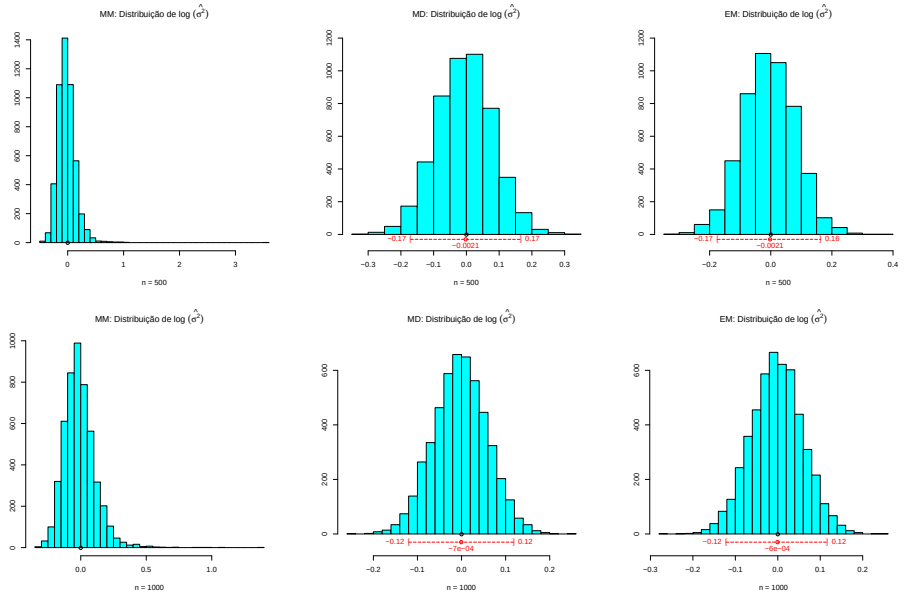


Figura 69 – Distribuição dos estimadores do parâmetro σ^2 do modelo Weibull-t no Cenário 2 com $n = (500, 1000)$

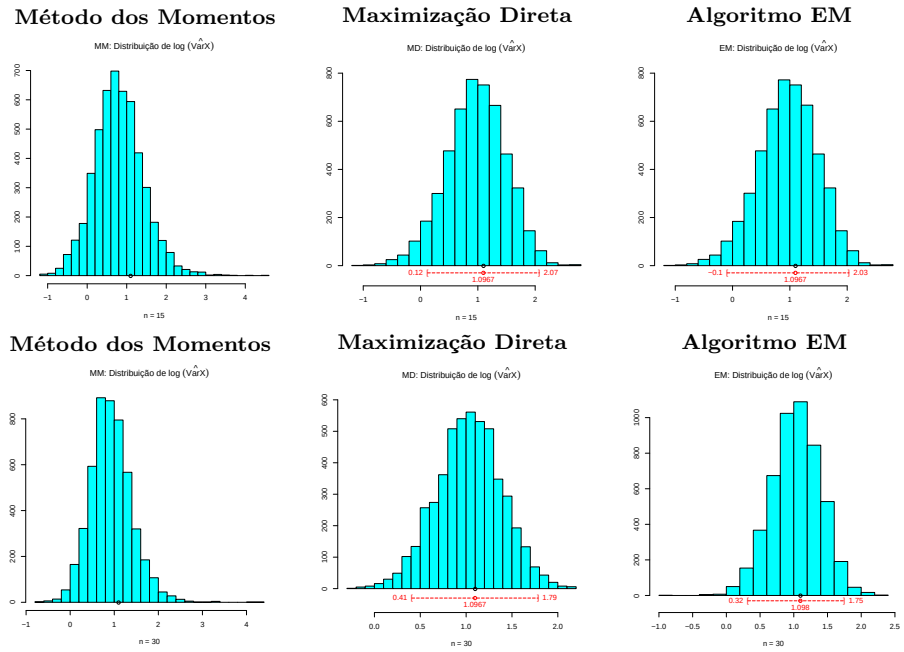


Figura 70 – Distribuição dos estimadores de Var(X) no modelo Weibull-t no Cenário 2 com $n = (15, 30)$

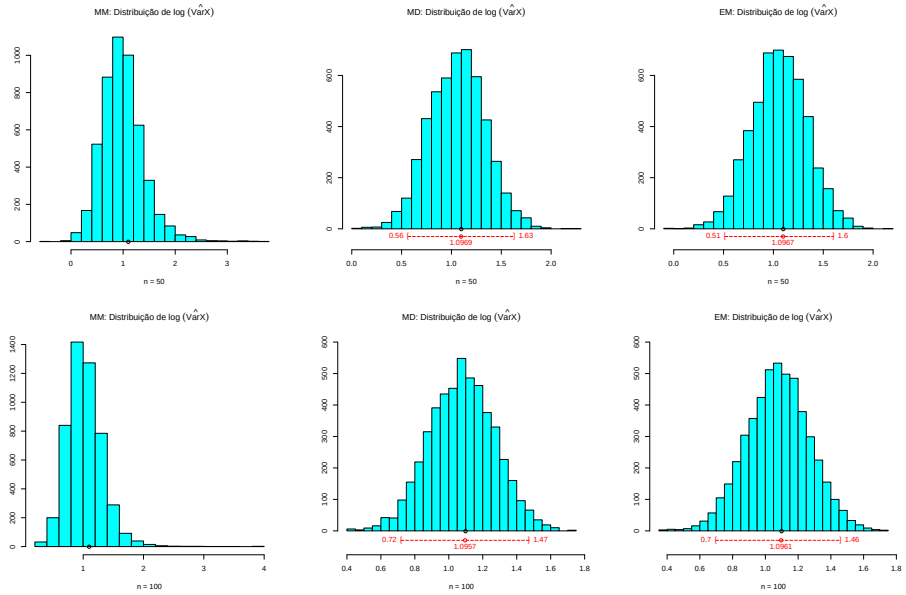


Figura 71 – Distribuição dos estimadores de Var(X) no modelo Weibull-t no Cenário 2 com $n = (50, 100)$

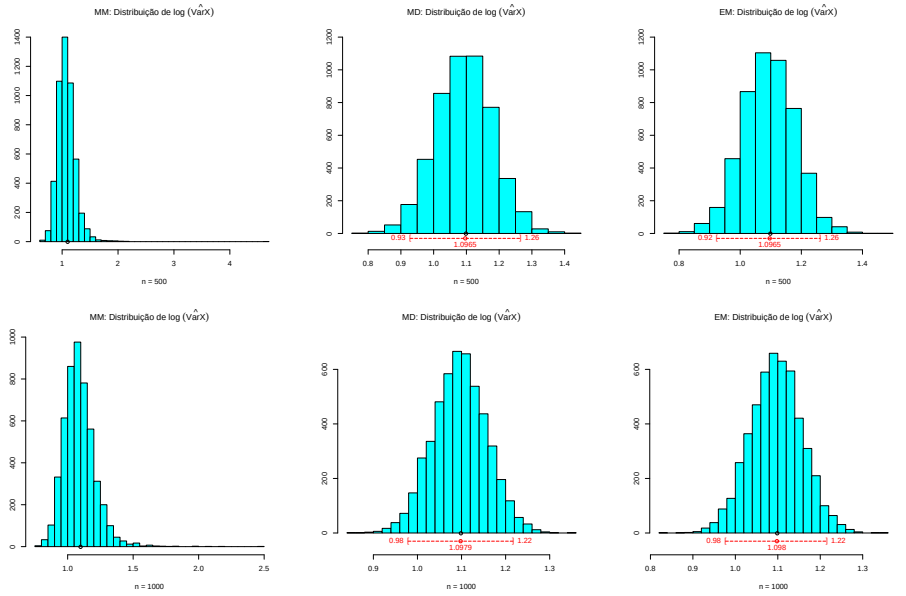


Figura 72 – Distribuição dos estimadores de Var(X) no modelo Weibull-t no Cenário 2 com $n = (500, 1000)$

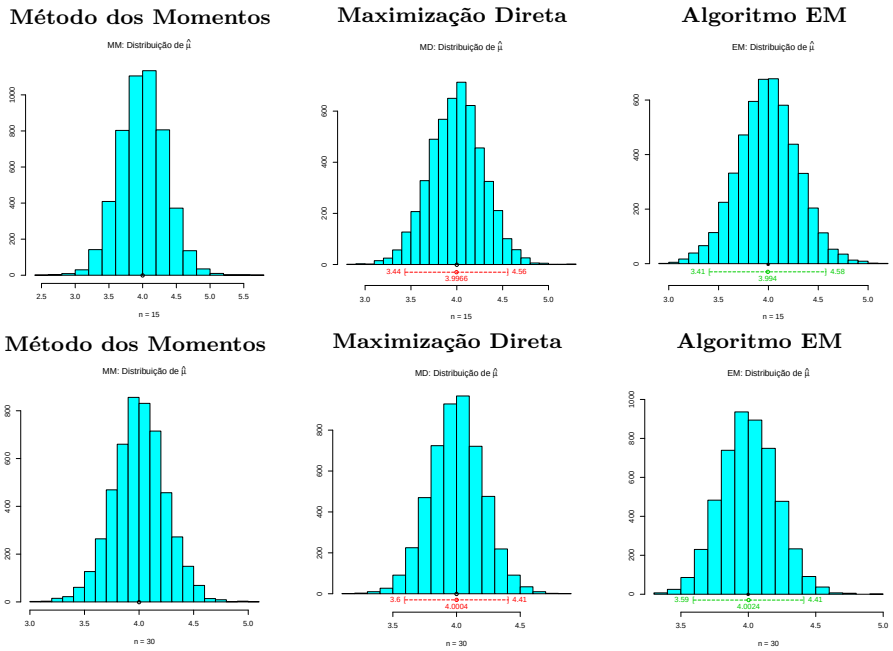


Figura 73 – Distribuição dos estimadores do parâmetro μ do modelo Slash-t no Cenário 1 com $n = (15, 30)$

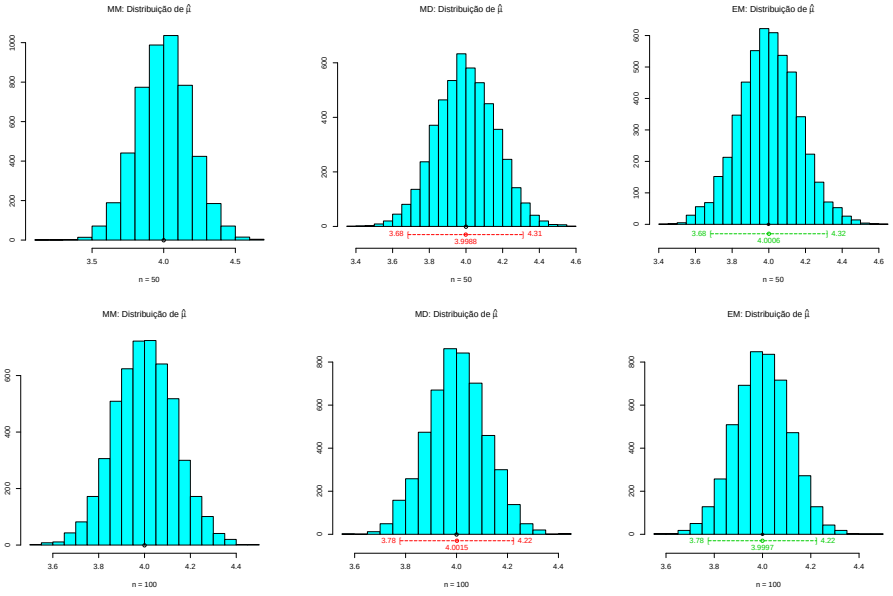


Figura 74 – Distribuição dos estimadores do parâmetro μ do modelo Slash-t no Cenário 1 com $n = (50, 100)$

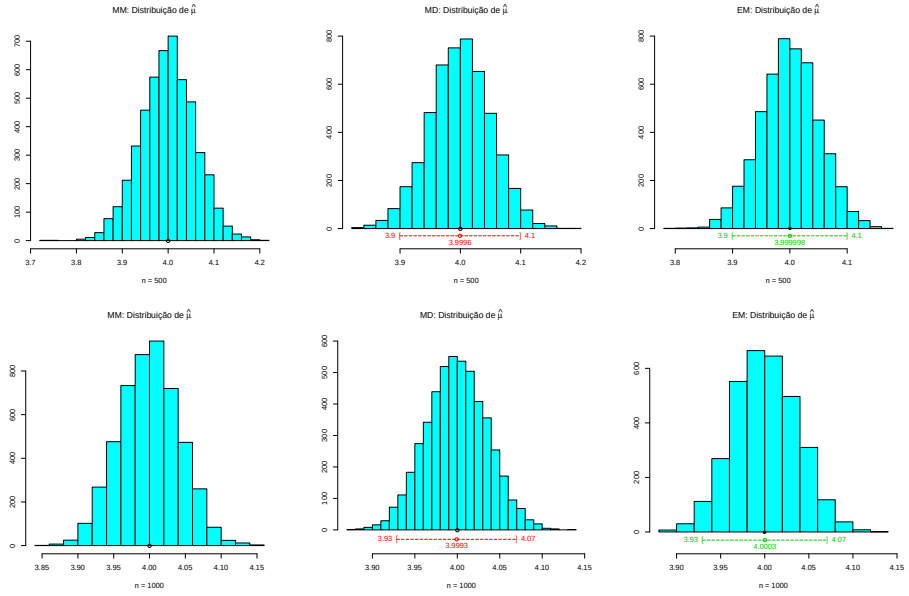


Figura 75 – Distribuição dos estimadores do parâmetro μ do modelo Slash-t no Cenário 1 com $n = (500, 1000)$

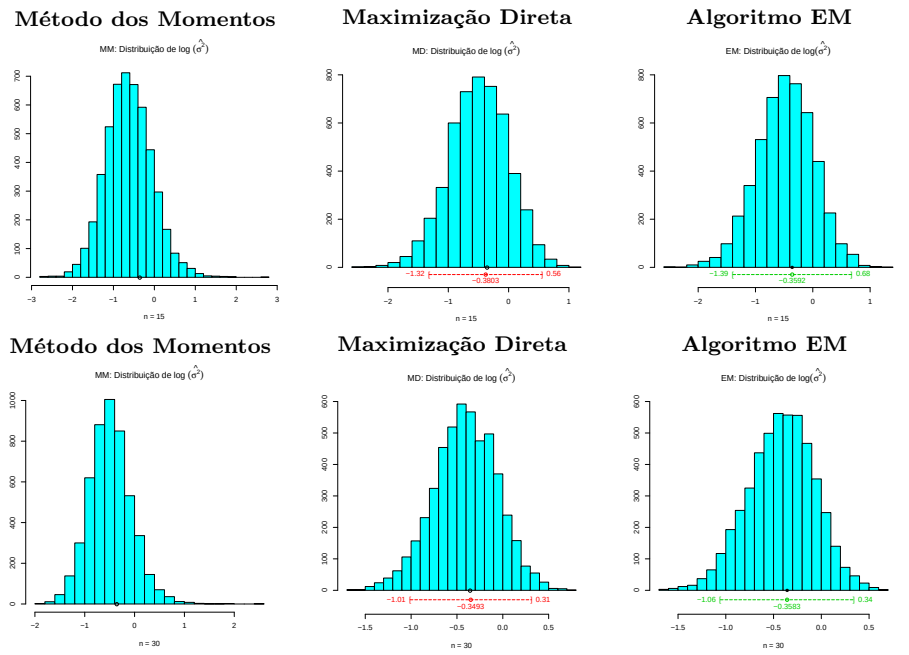


Figura 76 – Distribuição dos estimadores do parâmetro σ^2 do modelo Slash-t no Cenário 1 com $n = (15, 30)$

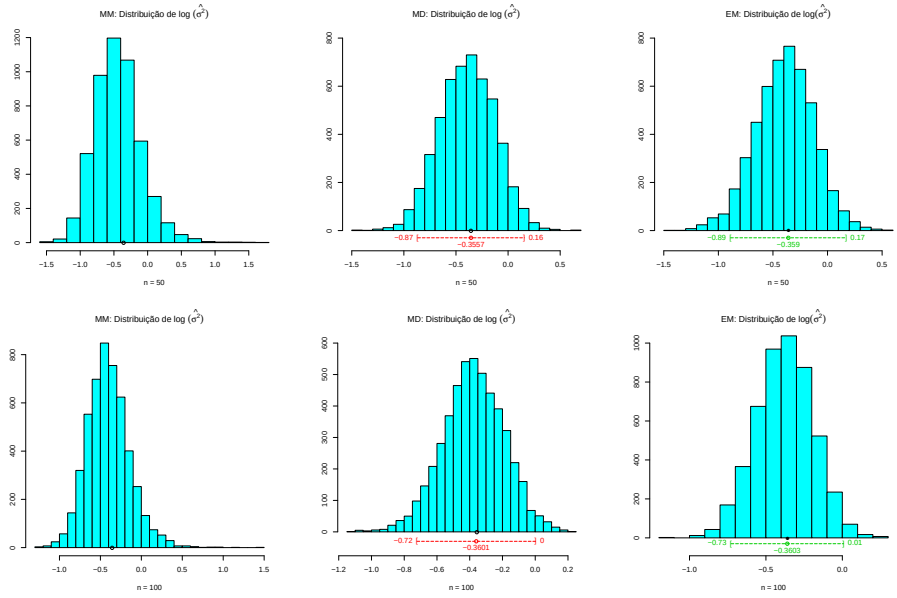


Figura 77 – Distribuição dos estimadores do parâmetro σ^2 do modelo Slash-t no Cenário 1 com $n = (50, 100)$

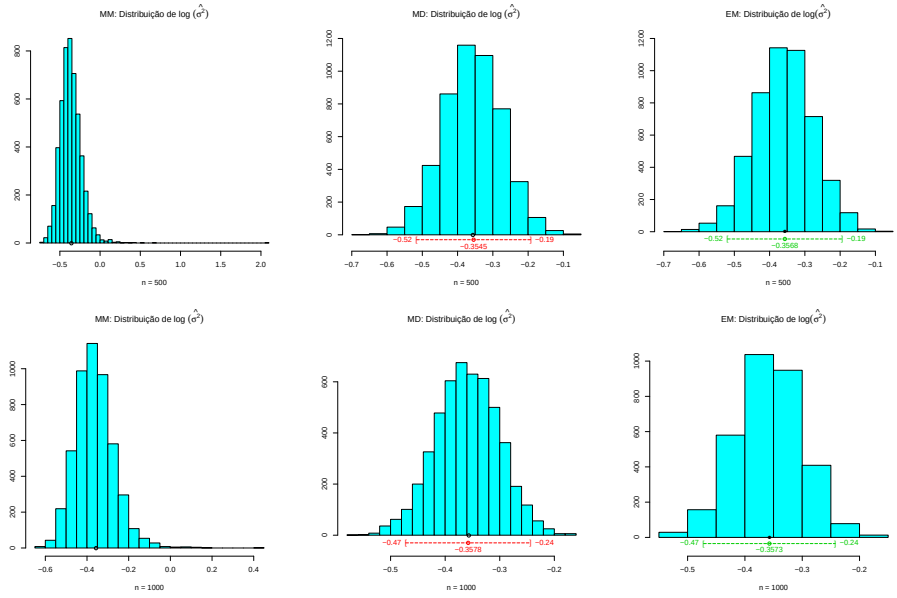


Figura 78 – Distribuição dos estimadores do parâmetro σ^2 do modelo Slash-t no Cenário 1 com $n = (500, 1000)$

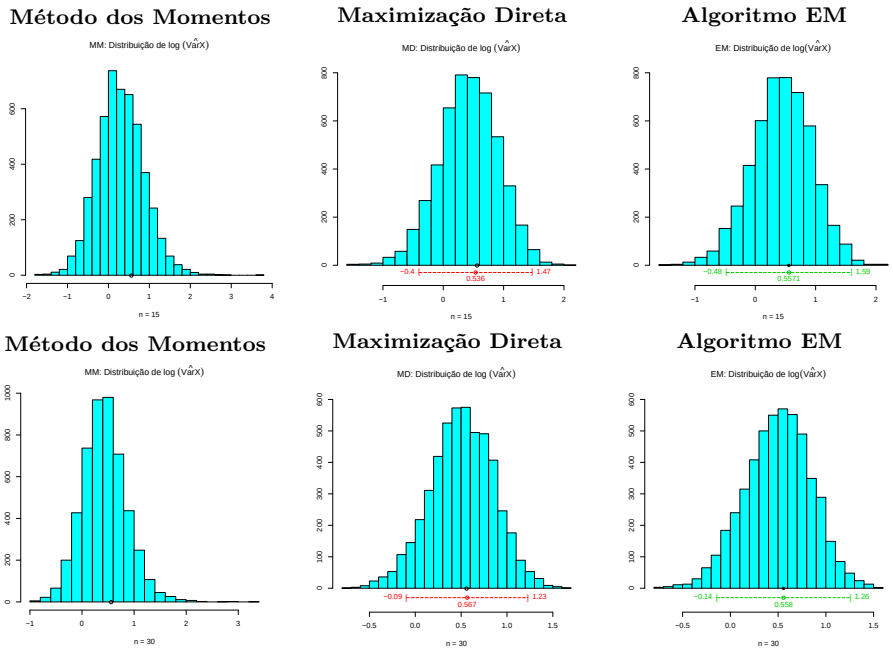


Figura 79 – Distribuição dos estimadores de Var(X) no modelo Slash-t no Cenário 1 com $n = (15, 30)$

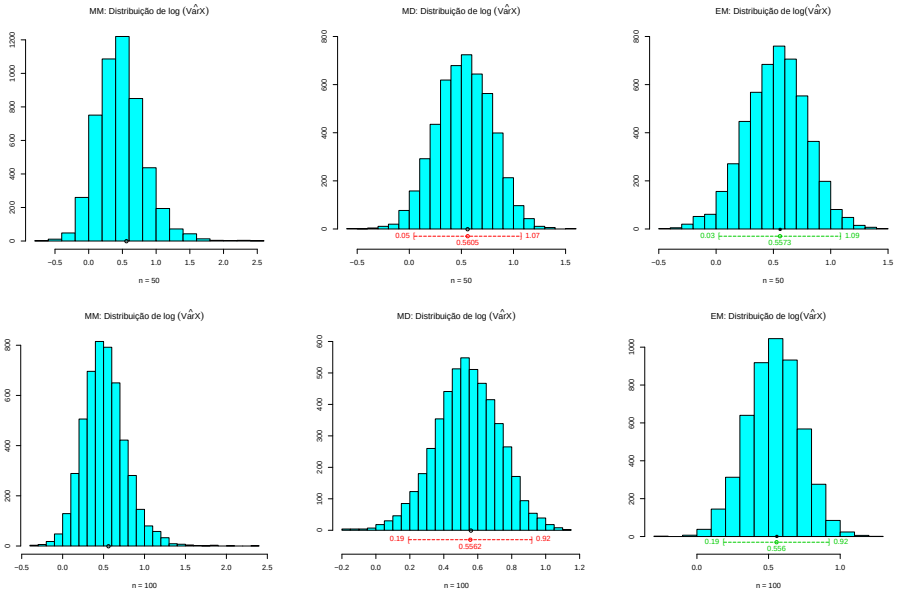


Figura 80 – Distribuição dos estimadores de Var(X) no modelo Slash-t no Cenário 1 com $n = (50, 100)$

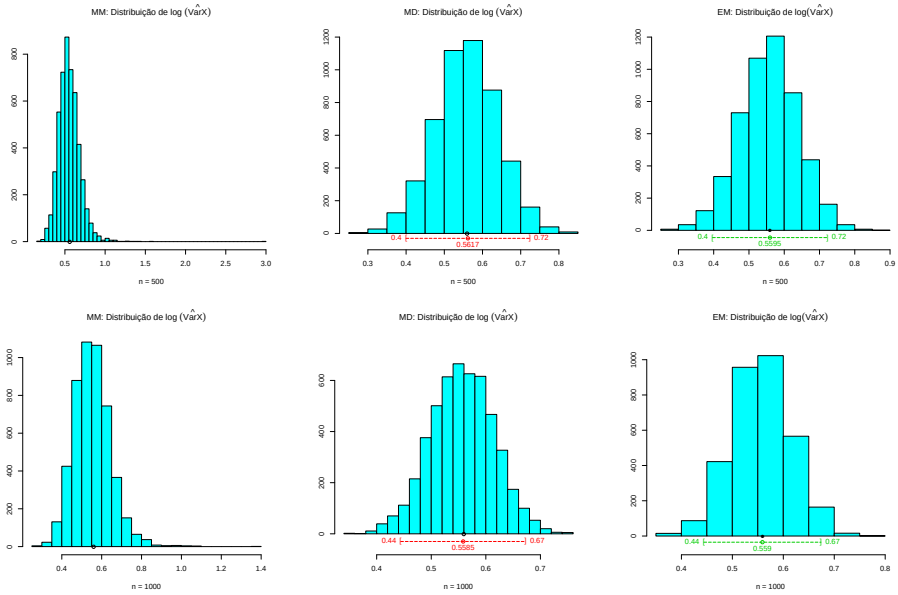


Figura 81 – Distribuição dos estimadores de Var(X) no modelo Slash-t no Cenário 1 com $n = (500, 1000)$

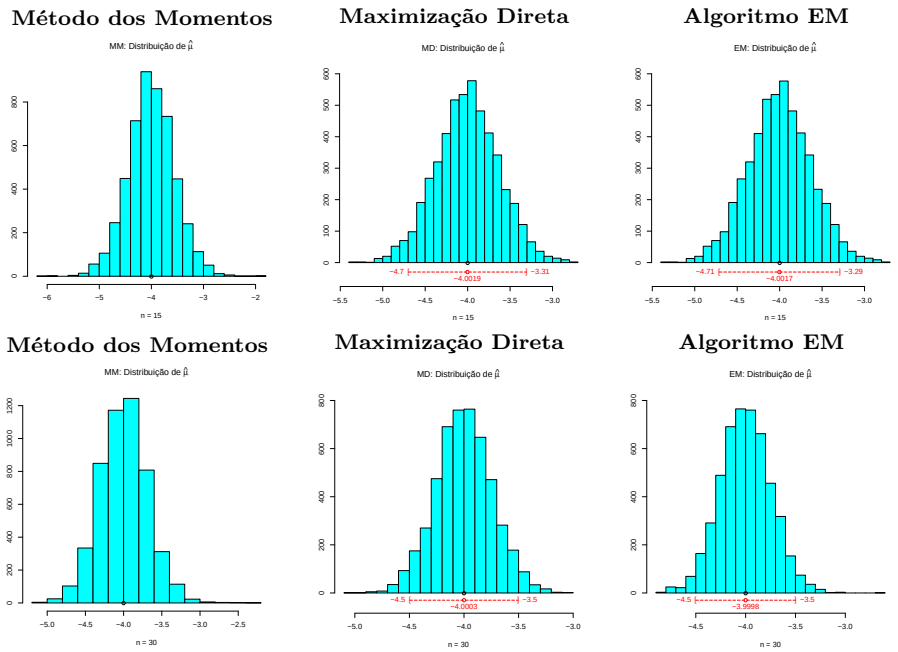


Figura 82 – Distribuição dos estimadores do parâmetro μ do modelo Slash-t no Cenário 2 com $n = (15, 30)$

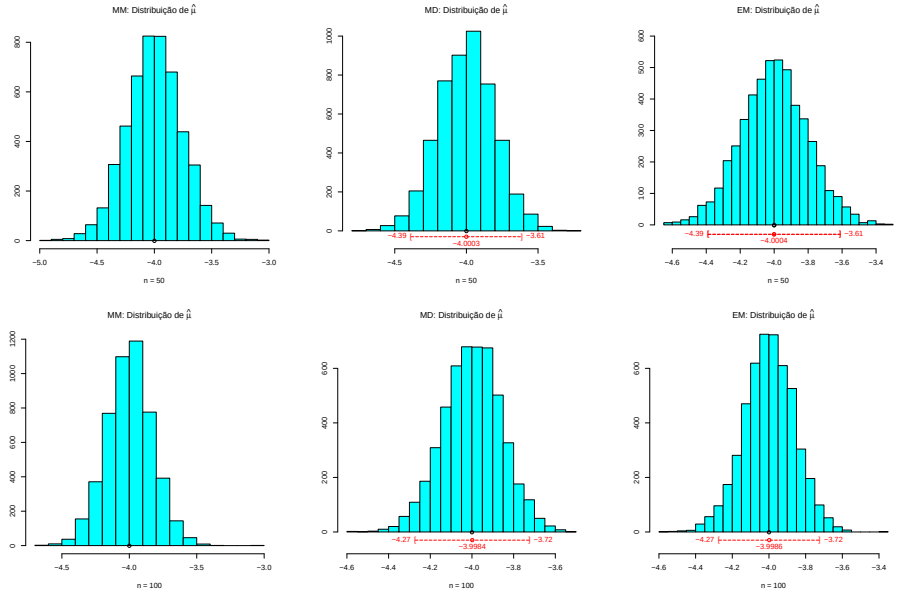


Figura 83 – Distribuição dos estimadores do parâmetro μ do modelo Slash-t no Cenário 2 com $n = (50, 100)$

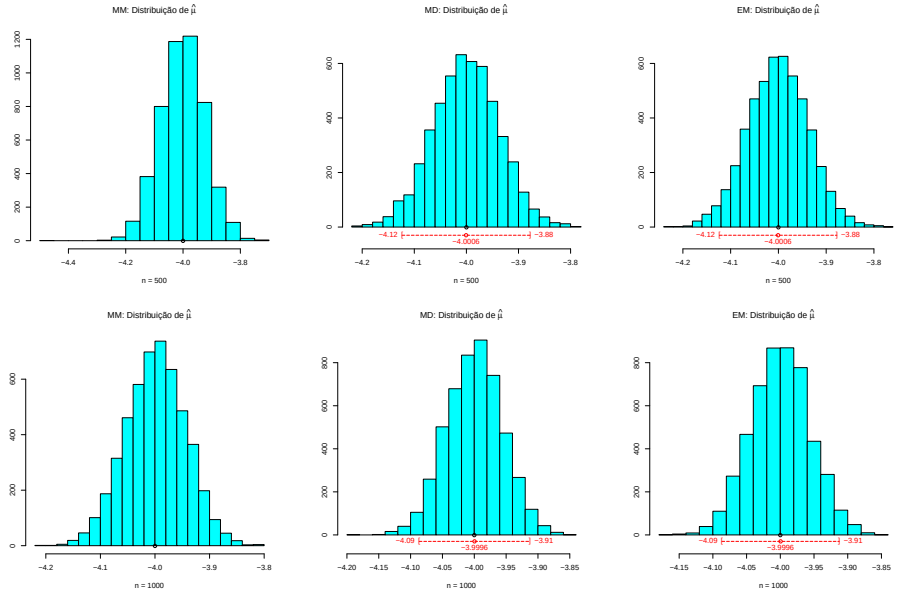


Figura 84 – Distribuição dos estimadores do parâmetro μ do modelo Slash-t no Cenário 2 com $n = (500, 1000)$

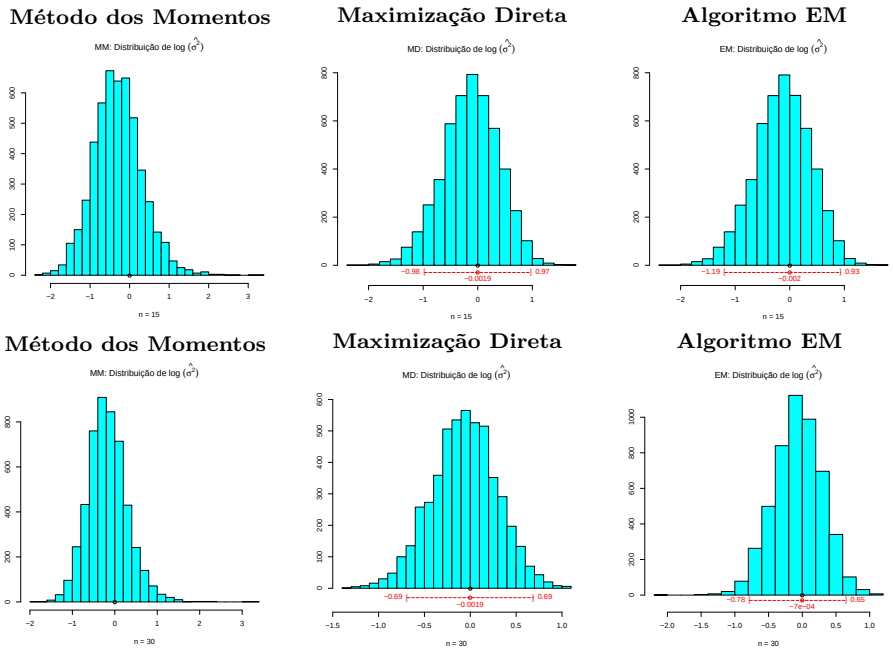


Figura 85 – Distribuição dos estimadores do parâmetro σ^2 do modelo Slash-t no Cenário 2 com $n = (15, 30)$

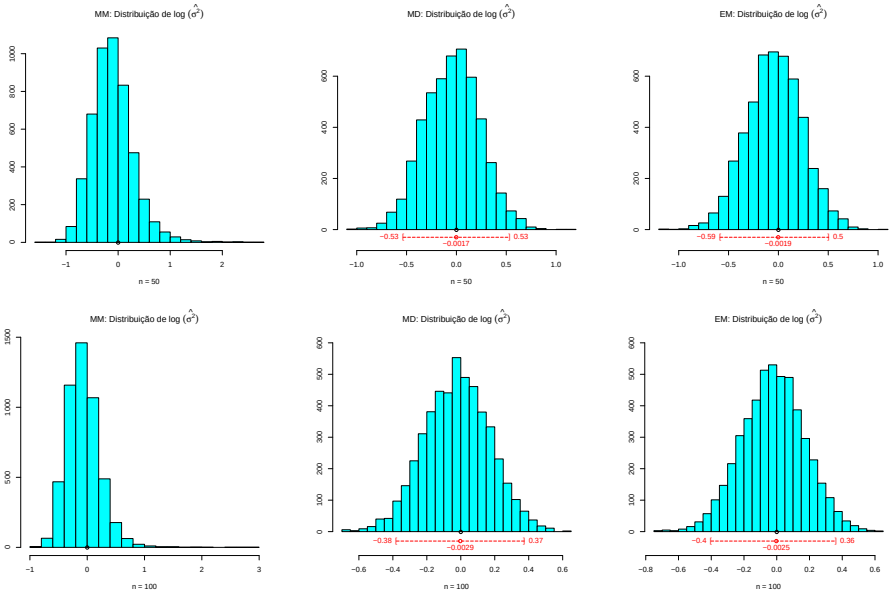


Figura 86 – Distribuição dos estimadores do parâmetro σ^2 do modelo Slash-t no Cenário 2 com $n = (50, 100)$

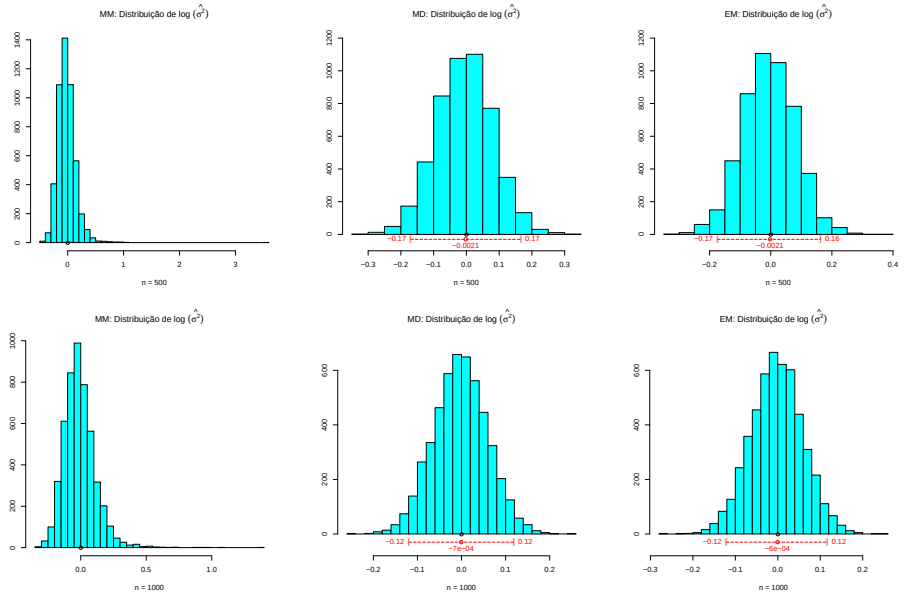


Figura 87 – Distribuição dos estimadores do parâmetro σ^2 do modelo Slash-t no Cenário 2 com $n = (500, 1000)$

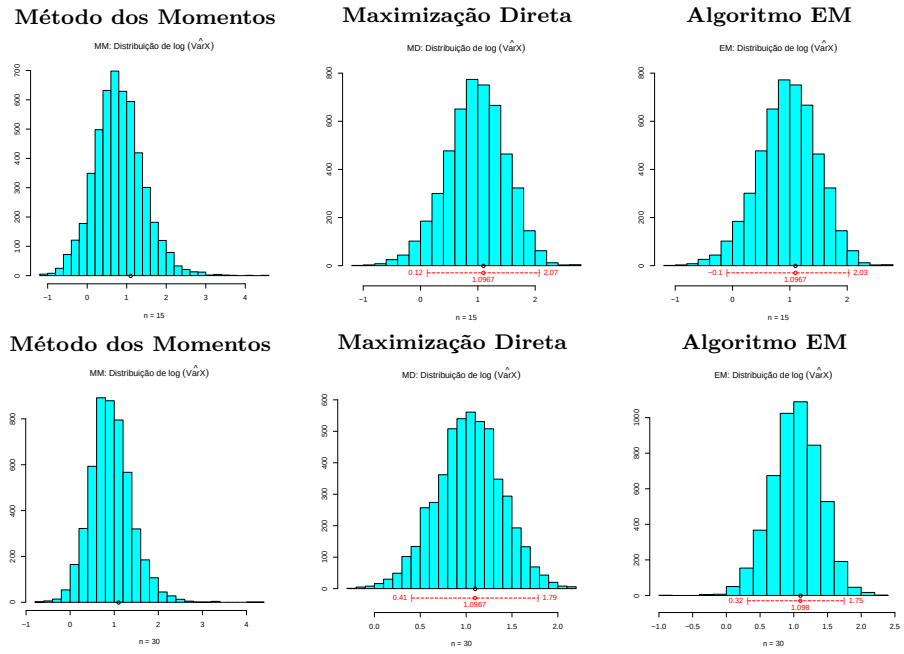


Figura 88 – Distribuição dos estimadores de Var(X) no modelo Slash-t no Cenário 2 com $n = (15, 30)$

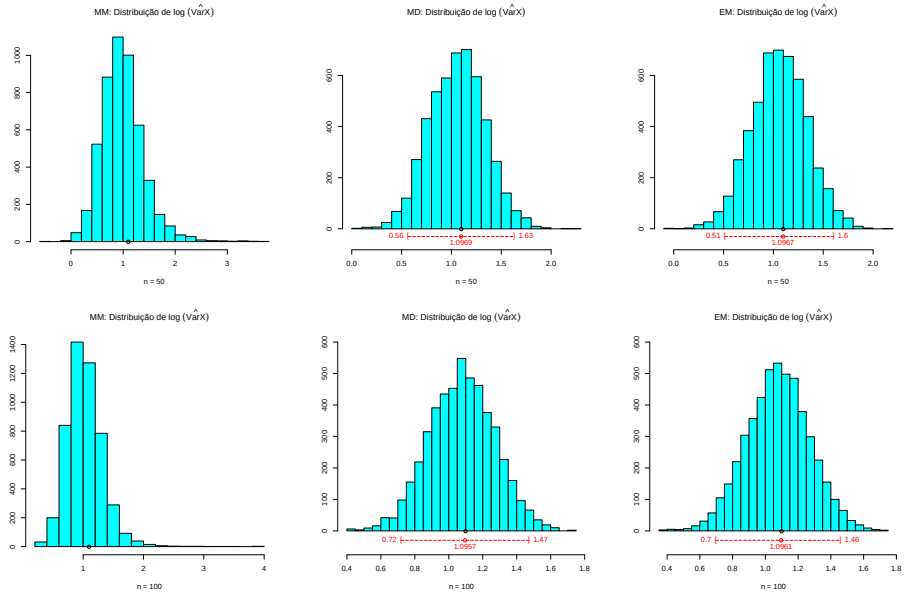


Figura 89 – Distribuição dos estimadores de Var(X) no modelo Slash-t no Cenário 2 com $n = (50, 100)$

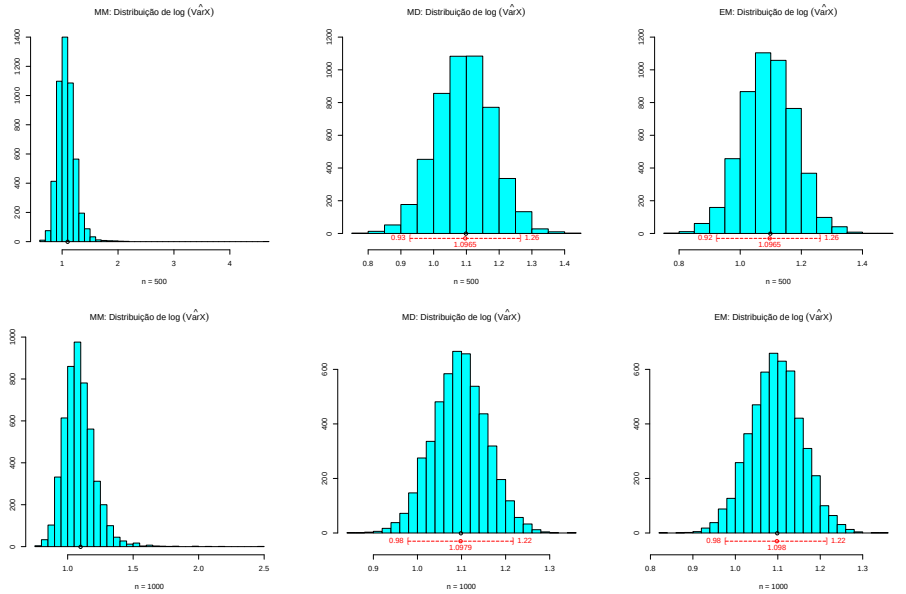


Figura 90 – Distribuição dos estimadores de Var(X) no modelo Slash-t no Cenário 2 com $n = (500, 1000)$