# Inferência Bayesiana para os parâmetros da distribuição Weibull Estendida de Marshall–Olkin

André F. B. Menezes

Universidade Estadual de Maringá

Departamento de Estatística

7 de Novembro de 2017

# Organização

# Introdução

## Método de Marshall-Olkin (1997)

$$
\begin{aligned}
S(x \mid \theta, \alpha) &= \frac{\alpha\, S_1(x \mid \theta)}{1 - \bar{\alpha}\, S_1(x \mid \theta)}, \\
f(x \mid \theta, \alpha) &= \frac{\alpha\, f_1(x \mid \theta)}{[1 - \bar{\alpha}\, S_1(x \mid \theta)]^2}, \\
h(x \mid \theta, \alpha) &= \frac{h_1(x \mid \theta)}{1 - \bar{\alpha}\, S_1(x \mid \theta)}
\end{aligned}
$$

em que $\alpha > 0$ é o novo parâmetro denominado de *tilt*, $S_1(x \mid \theta)$, $f_1(x \mid \theta)$ e $h_1(x \mid \theta)$ são as funções sobrevivência, densidade e risco da distribuição base, respectivamente.

# Introdução

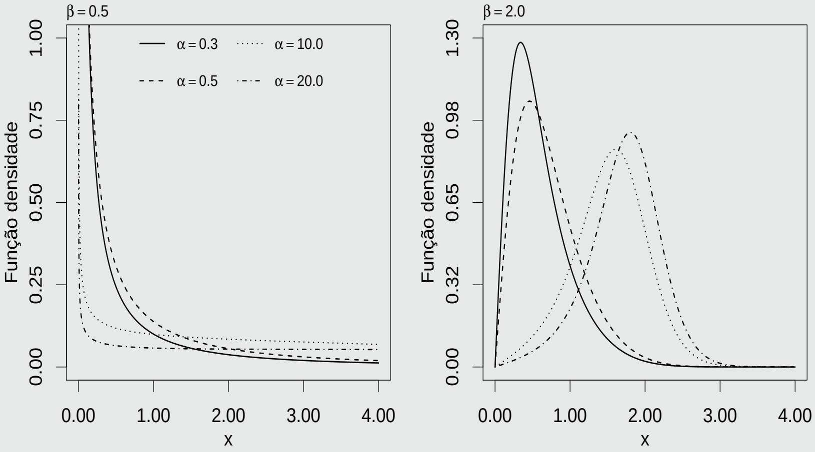
## Distribuição Weibull Estendida de Marshall–Olkin (MOEW)

$$\begin{aligned} S(x \mid \lambda, \alpha, \beta) &= \frac{\alpha\, \mathrm{e}^{-(\lambda x)^{\beta}}}{1 - \bar{\alpha}\, \mathrm{e}^{-(\lambda x)^{\beta}}}, \\ f(x \mid \lambda, \alpha, \beta) &= \frac{\alpha\, \beta\, \lambda^{\beta}\, x^{\beta - 1}\, \mathrm{e}^{-(\lambda x)^{\beta}}}{\left[1 - \bar{\alpha}\, \mathrm{e}^{-(\lambda x)^{\beta}}\right]^{2}}, \\ h(x \mid \lambda, \alpha, \beta) &= \frac{\beta\, \lambda^{\beta}\, x^{\beta - 1}}{1 - \bar{\alpha}\, \mathrm{e}^{-(\lambda x)^{\beta}}} \end{aligned}$$

$\lambda > 0$ é o parâmetro de taxa $(1/\mu)$, $\beta > 0$ e $\alpha > 0$ governam a forma das funções de densidade e risco.

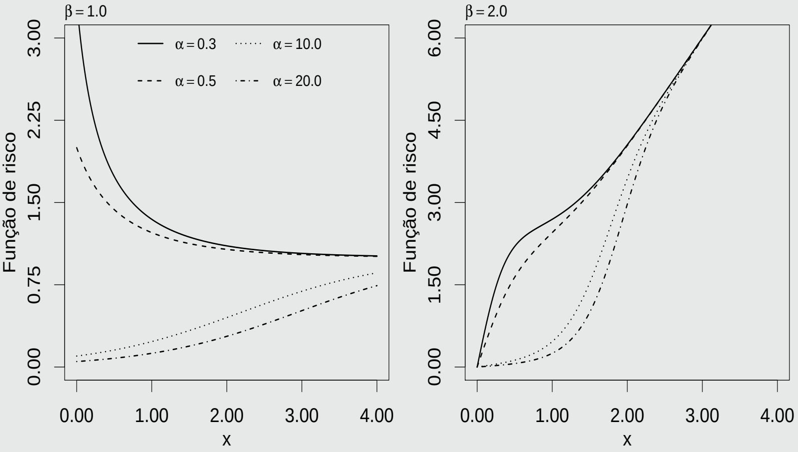
# Introdução

## Função densidade de probabilidade



# Introdução

## Função de risco



# Introdução

## Casos particulares

- $0 < \alpha < 1 \rightarrow$ distribuição Weibull-Geométrica.
- $\alpha = 1 \rightarrow$ distribuição Weibull.
- $\beta = 1 \rightarrow$ distribuição Exponencial Estendida de Marshall–Olkin.
- $\alpha = \beta = 1 \rightarrow$ distribuição Exponencial.

# Função de Verossimilhança

## Dados completos

Seja $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^\top$ uma amostra aleatória e da distribuição MOEW.

- Função de verossimilhança:

$$L(\boldsymbol{\theta} \mid \boldsymbol{x}) = \prod_{i=1}^{n} f(x_i \mid \boldsymbol{\theta}) = \prod_{i=1}^{n} \frac{\alpha\, \beta\, \lambda^\beta\, x_i^{\beta-1}\, \mathrm{e}^{-(\lambda x_i)^\beta}}{\left[1 - \bar{\alpha}\, \mathrm{e}^{-(\lambda x_i)^\beta}\right]^2}.$$

# Função de Verossimilhança

## Dados com censura intervalar

Considere que $\delta_i = 1$ se a observação não foi censurada e $\delta_i = 0$ se a observação é censurada por intervalo.

- Função de verossimilhança:

$$L(\boldsymbol{\theta} \mid \boldsymbol{t}, \{\boldsymbol{l}, \boldsymbol{r}\}, \boldsymbol{\delta}) = \prod_{i=1}^{n} \left[f(t_i \mid \boldsymbol{\theta})\right]^{\delta_i} \left[S(l_i \mid \boldsymbol{\theta}) - S(r_i \mid \boldsymbol{\theta})\right]^{1-\delta_i}.$$

Sendo:

- $\boldsymbol{t} = (t_1, \ldots, t_{n_1})$ as $n_1$ observações exatamente observadas;
- $\{\boldsymbol{l}, \boldsymbol{r}\} = (\{l_{n_1+1}, r_{n_1+1}\}, \ldots, \{l_{n_2}, r_{n_2}\})$ as $n_2$ observações com censura intervalar;

# Inferência Bayesiana

## Distribuição a posteriori

Toda inferência Bayesiana se concentra na distribuição a posteriori. A distribuição a posteriori é proporcional a

$$\pi(\boldsymbol{\theta} \mid \text{Dados}) \propto \pi(\boldsymbol{\theta})\, L(\boldsymbol{\theta} \mid \text{Dados}),$$

em que:

- $\boldsymbol{\theta} = (\lambda, \beta, \alpha)^\top$ é o vetor de parâmetros da distribuição MOEW;
- $\pi(\boldsymbol{\theta})$ é a distribuição a priori conjunta de $\boldsymbol{\theta}$;
- $L(\boldsymbol{\theta} \mid \text{Dados})$ representa a função de verossimilhança da amostra observada.

# Inferência Bayesiana

## Métodos MCMC

- A distribuição a posteriori não possui forma fechada.
- Os métodos de Monte Carlo em Cadeias de Markov (MCMC) foram utilizados para gerar amostras das distribuições marginais posteriori.
- A cadeia foi gerada considerando $1,000,000$ simulações para cada parâmetro.
- Com um *burn-in* de tamanho $1000$.
- Os valores simulados foram selecionados de $100$ em $100$.
- Resultando em amostras finais de tamanho $20,000$.
- A procedure `MCMC` do SAS foi utilizada.

# Aplicações

## Aplicação I

- Os dados foram simulados utilizando o método da transformação inversa, isto é

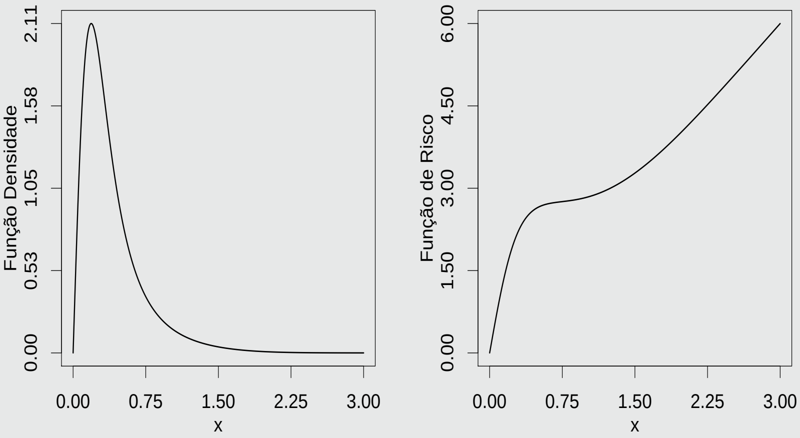
$$x_i = \frac{1}{\lambda}\left[-\log\left(\frac{u_i}{\alpha+\bar{\alpha}\,u_i}\right)\right]^{1/\beta}, \qquad i=1,\ldots,n.$$

em que $u_i$ são números aleatórios da distribuição uniforme padrão.

- O seguinte cenário foi estabelecido: $n=(100,300,500)$, $\lambda=1.0$, $\beta=2.0$ e $\alpha=0.1$

# Aplicações

## Aplicação I



# Aplicação I

## Definição das distribuições a priori

- $\pi(\lambda) \sim$ Gamma$(0.001; 0.001)$.
- $\pi(\beta) \sim$ Gamma$(0.001; 0.001)$.
- $\pi(\alpha) \sim$ Uniforme$(0; 1)$.

# Aplicação I

## Estimativas Clássicas e Bayesianas.

| $n$ | Parâmetro | EMV (E.P.) | Média (D.P.) | IC $95\%$ | HPD $95\%$ |
|---|---|---|---|---|---|
| 100 | $\lambda$ | 1.8894 (0.5828) | 1.8310 (0.3381) | (0.7471, 3.0317) | (1.1490, 2.4374) |
| | $\beta$ | 1.4222 (0.2273) | 1.4388 (0.1549) | (0.9767, 1.8678) | (1.1568, 1.7507) |
| | $\alpha$ | 0.6446 (0.4796) | 0.6217 (0.2260) | (-0.2954, 1.5845) | (0.2376, 0.9993) |
| 300 | $\lambda$ | 1.3913 (0.2275) | 1.5025 (0.2544) | (0.9455, 1.8372) | (1.0365, 2.0378) |
| | $\beta$ | 1.7209 (0.1402) | 1.6623 (0.1398) | (1.4460, 1.9957) | (1.4025, 1.9411) |
| | $\alpha$ | 0.3082 (0.1405) | 0.4062 (0.1802) | (0.0329, 0.5835) | (0.1198, 0.7917) |
| 500 | $\lambda$ | 0.7853 (0.1460) | 0.8304 (0.1571) | (0.4990, 1.0715) | (0.5135, 1.1434) |
| | $\beta$ | 1.9647 (0.1024) | 1.9266 (0.1044) | (1.7639, 2.1655) | (1.7205, 2.1288) |
| | $\alpha$ | 0.0697 (0.0349) | 0.0898 (0.0458) | (0.0012, 0.1381) | (0.0178, 0.1823) |

- Verdadeiros valores dos parâmetros: $\lambda = 1, \beta = 2.0$ e $\alpha = 0.1$.

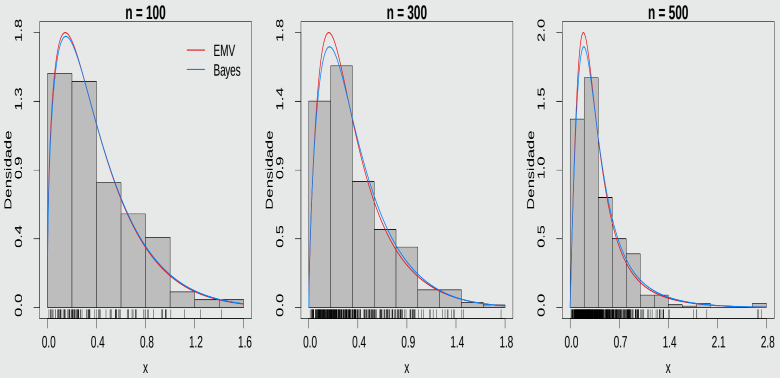
# Aplicação I

## Critérios para avaliar a convergência das cadeias ($n = 100$).

| Parâmetro | Heidelberger-Welch | | Geweke | |
|---|---|---|---|---|
| | Estatística | valor-*p* | Estatística | valor-*p* |
| $\lambda$ | 0.0639 | 0.7892 | -0.8358 | 0.4033 |
| $\beta$ | 0.1540 | 0.3775 | -0.3016 | 0.7629 |
| $\alpha$ | 0.0665 | 0.7733 | -0.2364 | 0.8131 |

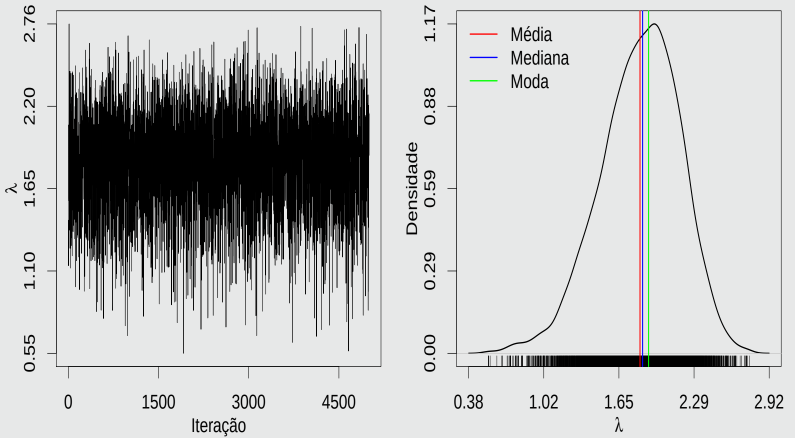
# Aplicação I

## Histograma e curva teórica ajustada.



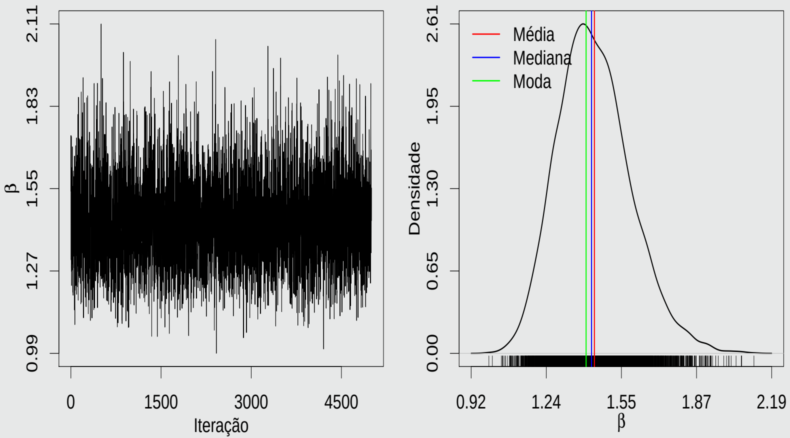
# Aplicação I

## Convergência e densidade a posteriori de $\lambda$ ($n = 100$).



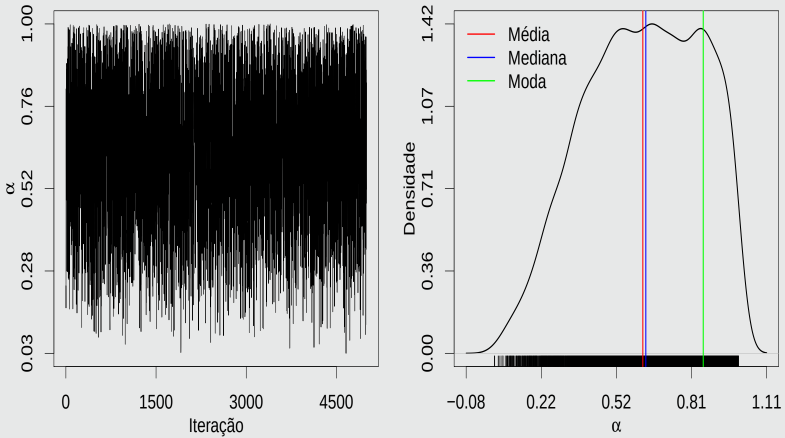
# Aplicação I

## Convergência e densidade a posteriori de $\beta$ $(n = 100)$.



# Aplicação I

## Convergência e densidade a posteriori de $\alpha$ $(n = 100)$.



# Aplicação II

- Os dados consistem de um estudo realizado na Dinamarca entre 1993 e 1972 envolvendo 731 pacientes diagnosticados como diabéticos do Tipo I entre 1933 e 1972, e seguidos até a morte, a emigração ou 31 de dezembro de 1984.
- O tempo de sobrevivência de interesse é o tempo (meses) desde o início da diabetes até o início da nefropatia diabética (DN).
- Todos os paciente desenvolveram a nefropatia diabética no momento de admissão ou até o final do estudo, logo não há observação censurada a direita.
- 596 tempos foram exatamente observados e 136 observações apresentaram censura intervalar.
- Entre os 731 pacientes, têm-se 277 mulheres e 454 homens.

# Aplicação II

## Considerações

- A distribuição MOEW foi reparametrizada em termos de $\mu$, em que $\lambda = 1/\mu$.
- $\pi(\lambda) \sim$ Gamma$(0.001; 0.001)$.
- $\pi(\beta) \sim$ Gamma$(0.001; 0.001)$.
- $\pi(\alpha) \sim$ Uniforme$(0; 1)$.

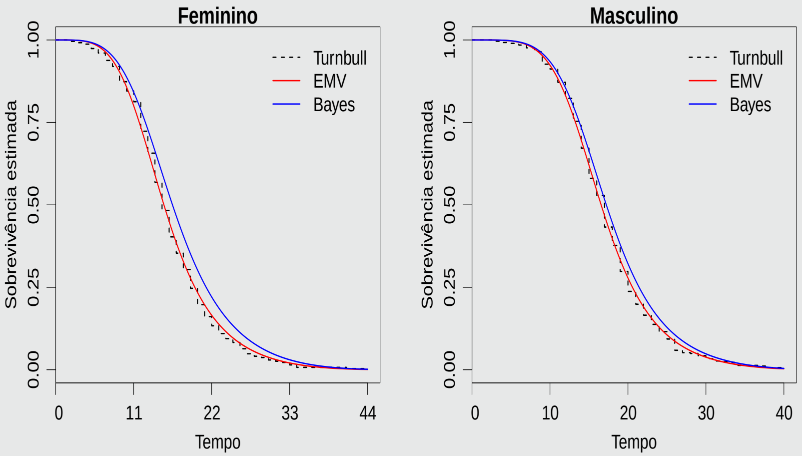
# Aplicação I

## Estimativas Clássicas e Bayesianas.

| Modelo | Gênero | Parâmetro | EMV (E.P.) | Média (D.P.) | IC $95\%$ | HPD $95\%$ |
|---|---|---|---|---|---|---|
| MOEW | Feminino | $\mu$ | 31.5621 (4.1393) | 31.3722 (5.1347) | (23.4492, 39.6751) | (23.1008, 40.9715) |
| | | $\beta$ | 4.1870 (0.2991) | 4.0692 (0.3100) | (3.6007, 4.7734) | (3.4488, 4.6691) |
| | | $\alpha$ | 0.0489 (0.0360) | 0.0754 (0.0565) | (-0.0216, 0.1194) | (0.0010, 0.1854) |
| | Masculino | $\mu$ | 33.0371 (3.0012) | 33.4082 (3.6710) | (27.1548, 38.9194) | (27.4911, 40.7049) |
| | | $\beta$ | 4.9068 (0.2547) | 4.8526 (0.2523) | (4.4076, 5.4059) | (4.3589, 5.3484) |
| | | $\alpha$ | 0.0343 (0.0197) | 0.0410 (0.0226) | (-0.0043, 0.0729) | (0.0032, 0.0852) |
| Weibull | Feminino | $\mu$ | 18.0958 (0.4457) | 18.1044 (0.4482) | (17.2183, 18.9732) | (17.2491, 19.0029) |
| | | $\beta$ | 2.6450 (0.1203) | 2.6368 (0.1201) | (2.4083, 2.8818) | (2.4070, 2.8771) |
| | Masculino | $\mu$ | 19.3296 (0.3357) | 19.3346 (0.3375) | (18.6698, 19.9894) | (18.6754, 19.9976) |
| | | $\beta$ | 2.9583 (0.1062) | 2.9533 (0.1065) | (2.7497, 3.1670) | (2.7454, 3.1628) |

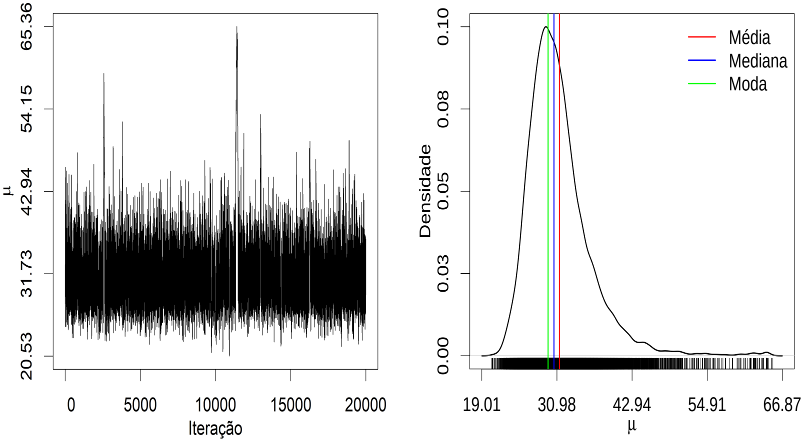
# Aplicação I

## Curvas de sobrevivência estimadas.



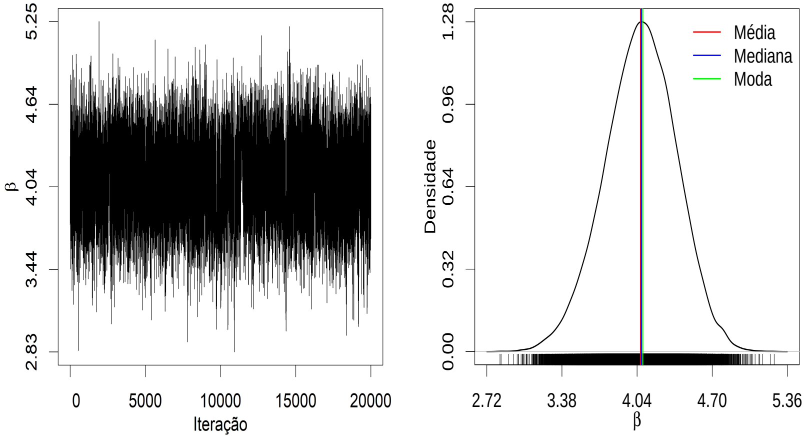
# Convergência e densidade a posteriori de $\mu$.

- Gênero Feminino



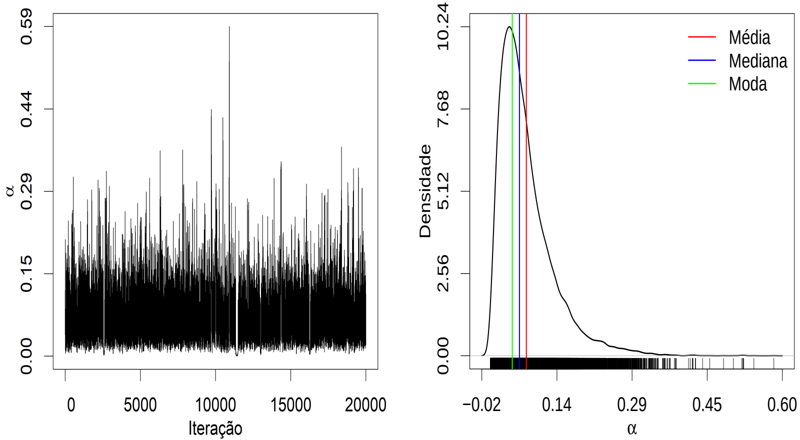
# Convergência e densidade a posteriori de $\beta$.

- Gênero Feminino



# Convergência e densidade a posteriori de $\alpha$.

- Gênero Feminino



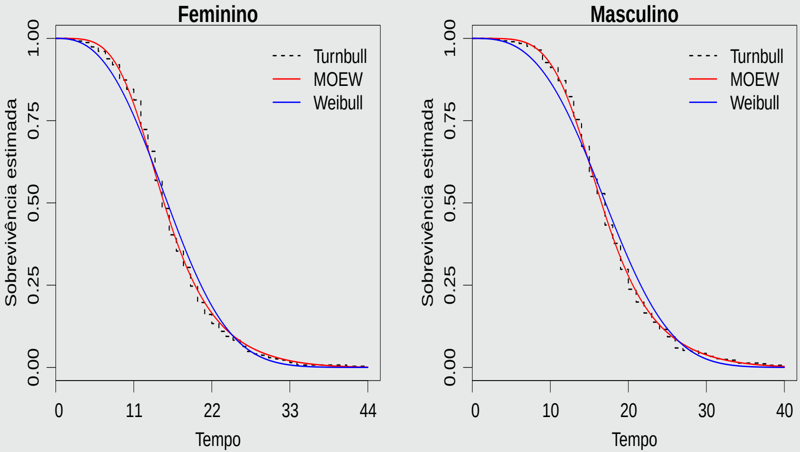
# Aplicação I

## Critérios de discriminação Clássicos e Bayesianos.

| Gênero | Modelo | AIC | DIC |
|---|---|---|---|
| Feminino | MOEW | 295.2428 | 285.9942 |
| | Weibull | 309.9924 | 309.9955 |
| Masculino | MOEW | 399.3692 | 393.8040 |
| | Weibull | 438.8576 | 438.8871 |

# Aplicação I

## Comparação entre as distribuição MOEW e Weibull.



# Referências

[1] Almalki, S. J., Nadarajah, S., 2014. Modifications of the Weibull distribution: A review. Reliability Engineering & System Safety 124 (Supplement C), 32–55.

[2] Marshall, A. W., Olkin, I., 1997. *A new method for adding a parameter to a family of distributions with application to the exponential and weibull families.* Biometrika 84 (3), 641–652.

[3] SAS, 2010a.*The MCMC Procedure*, SAS/STAT User's Guide, Version 9.4. Cary, NC: SAS Institute Inc.

[4] SAS, 2010b. *The NLMIXED Procedure*, SAS/STATR User's Guide, Version 9.4. Cary, NC: SAS Institute Inc.