# Distribuições à priori

Prof. Caio Azevedo

# Motivação

- As distribuições à priori são de fundamental importância na inferência bayesiana.
- De fato, diferentes escolhas podem levar, algumas vezes, à resultados "significativamente" diferentes (para uma mesma verossimilhança).
- Contudo, existem alguns métodos para "construção", escolha e comparação de prioris.

# Propriedades desejadas

- Fundamentalmente, uma priori deve apresentar as seguintes características:
  - Respeitar o espaço paramétrico.
  - Conduzir à uma posteriori própria (integrável).
  - Refletir, apropriadamente, o conhecimento de especialistas.
  - Conduzir à um processo de inferência com “boas propriedades”.
  - Não “dominar” a verossimilhança (à menos que exista uma contundente razão para isto).
- OBS: uma priori não, necessariamente, precisa ser uma fdp nem mesmo ser integrável.

# Tipos de priori

- Quanto à propriedade (integrabilidade):
  - Própria: quando ela é integrável (ainda que não seja uma fdp).

    Exemplos:

$$\theta \sim gama(a,b); p(\theta) = I_{(0,1)}(\theta); p(\theta) = \frac{1}{\theta} I_{(a,b)}(\theta)$$

  - Imprópria: quando ela não é integrável. Exemplos:

$$p(\theta) \propto I_{(0,\infty)}(\theta); p(\theta) \propto I_{(a,\infty)}(\theta)$$

# Tipos de priori (cont.)

- Quanto à depender ou não da amostra (verossimilhança) que se está analisando
  - "Subjetiva": depende, no máximo, do modelo (verossimilhança) adotado.

$$\theta \sim N(a, b)$$

  - "Objetiva": depende somente do modelo (verossimilhança) adotado e/ou da amostra onservada. Exemplos: priori de Jeffreys, priori de máxima entropia, priori empírica, priori de referência (Berger e Bernardo).
  - Observação: desde que os hiperparâmetros sejam escolhidos sem considerar a amostra em questão, as prioris conjugadas são consideradas como subjetivas.

# Tipos de priori (cont.)

- Quanto ao nível de informação.
  - Não informativa: assumem ignorância total em relação ao parâmetro, ou seja, é proporcional à uma constante.

$$p(\theta) = I_{(0,1)}(\theta); p(\theta) = I_{(0,\infty)}(\theta)$$

  - Informativas: assumem algum grau de conhecimento acerca do parâmetro.
    - Pouco informativa ou vaga: $\theta \sim N(0, 10000)$.
    - Moderadamente informativa: $\theta \sim N(0, 100)$.
    - Muito informativa: $\theta \sim N(0, 1)$.

# Observações importantes

- Prioris muito informativas podem conflitar com a verossimilhança.
- Prioris pouco informativas, ainda que levem à posterioris próprias, podem gerar uma certa instabilidade (quando se obtem a posteriori numericamente).
- Prioris impróprias, podem levar à posterioris impróprias. Neste caso, não se pode usar tais prioris. Ainda que a posteriori seja própria, a utilização de prioris impróprias pode comprometer o uso de certas metodologias bayesianas como o fator de Bayes.

## Observações importantes

- Prioris próprias, em geral, conduzem à posterioris próprias.
- Se a verossimilhança for integrável ($\int_\Theta p(\boldsymbol{x}|\theta)d\theta < \infty$, ou uniformemente limitada) a posteriori será própria, ainda que a priori seja imprópria.

# Metodologias mais utilizadas

- Escolher prioris que sejam apropriadas (sendo densidades ou não).
- Família conjugada (conjugada condicional).
- Priori de Jeffreys (e de Jeffreys sob independência).
- Máxima entropia.
- Não informativas (proporcionais à constante).
- Priori de referência.

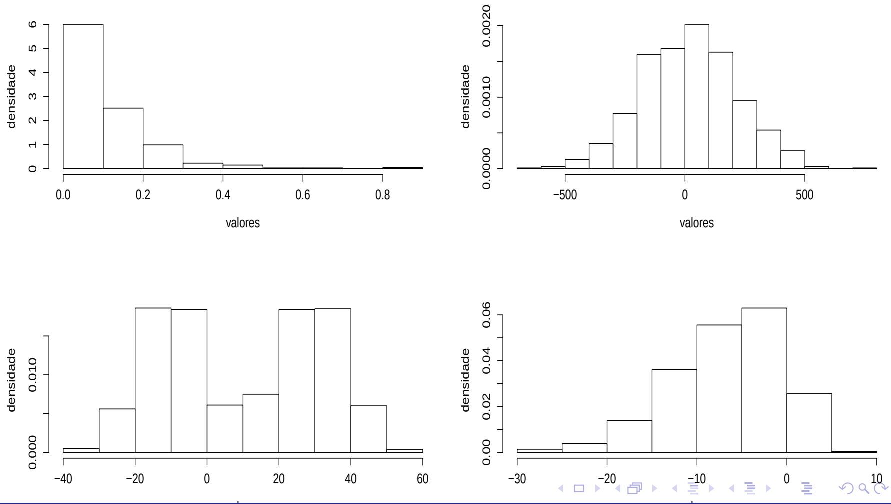
# Método do histograma

- Consiste em construir um histograma usando informações sobre o problema como: os valores mais prováveis do parâmetros fornecidos por especialistas, ou fornecidos por um mesmo especialista (com as respectivas probabilidades) ou estimativas oriundas de estudos anteriores.
- Nesse caso, tal histograma fornece uma aproximação para distribuição à priori e ele pode ser usado como um indicador de uma possível distribuição.

# Método do histograma

- Pode-se então, por exemplo, considerar alguma densidade com base nesses histogramas.
- No slide a seguir, no sentido horário, começando com o histograma da primeira linha e coluna: exponencial, normal, normal assimétrica e mistura de duas normais

# Exemplos hipotéticos relativos ao método do histograma



# Densidades apropriadas

- Sejam $X_i|\theta \overset{iid}{\sim} U[0,\theta], \theta > 0$
- Podemos assumir que: $p(\theta) = ae^{-\frac{a}{\theta}}\mathbb{1}_{(0,\infty}(\theta)$ ($IG(1,a) \equiv$ inverso de uma distribuição gama$(1, a^{-1}) \equiv$ inverso de uma distribuição $exp(a^{-1})$).
- Portanto:

$$p(\theta|\boldsymbol{x}) \quad \propto \quad \theta^{-n}e^{-\frac{a}{\theta}}\mathbb{1}_{[y_n,\infty)}(\theta)\mathbb{1}_{(0,\infty)}(\theta) = \theta^{-n}e^{-\frac{a}{\theta}}\mathbb{1}_{[y_n,\infty)}(\theta)$$

- Exercício: encontrar a constante de normalização da densidade acima.

# Família conjugada

- Família conjugada: escolhe-se uma priori que leva à uma posteriori na mesma família. Exemplo: famílias conjugadas naturais.
- Em geral, as prioris dessas famílias dependem de hiperparâmetros.
- A escolha deles (hiperparâmetros) requer conhecimento prévio sobre o problema ou pode ser feita através da amostra.

# Família conjugada (cont.)

- Para discutir a escolha dos hiperparâmetros, vamos considerar o Exemplo 4: verossimilhança de Poisson($\lambda$) com priori conjugada ($gama(a, b^{-1})$).
- Relembrando

$$p(\boldsymbol{x}|\lambda) \propto e^{-n\lambda}\lambda^{n\overline{x}}$$

que corresponde ao núcleo de uma gama($n\overline{x} + 1, n^{-1}$). Assim gama($a, b^{-1}$) é a família de distribuições a priori conjugadas para o modelo Poisson.

# Família conjugada (cont.)

- Assim,

$$\begin{aligned} p(\lambda|\boldsymbol{x}) &\propto p(\boldsymbol{x}|\lambda)p(\lambda) \propto e^{-n\lambda}\lambda^{n\overline{x}}e^{-b\lambda}\lambda^{a-1} \\ &= \lambda^{n\overline{x}+a-1}e^{(n+b)\lambda} \end{aligned}$$

  que corresponde ao núcleo de uma gama$(n\overline{x}+a,(n+b)^{-1})$. Assim $\lambda|\boldsymbol{x} \sim$ gama$(n\overline{x}+a,(n+b)^{-1})$.

- Por outro lado, note que

$$\begin{aligned} p(\boldsymbol{x}|\lambda)p(\lambda) &= \frac{e^{-n\lambda}\lambda^{n\overline{x}}}{\prod_{i=1}^{n} x_i!}\frac{b^a}{\Gamma(a)}e^{-b\lambda}\lambda^{a-1}\mathbb{1}_{(0,\infty)}(\lambda) \\ &= \frac{e^{-(n+b)\lambda}\lambda^{(n\overline{x}+a)-1}b^a}{\prod_{i=1}^{n} x_i!\Gamma(a)}\mathbb{1}_{(0,\infty)}(\lambda) \end{aligned}$$

# Família conjugada (cont.)

- Podemos também encontrar a constante de normalização, ou seja (exercício):

$$
\begin{aligned}
p(\boldsymbol{x}) &= \int_0^\infty p(\boldsymbol{x}|\lambda)p(\lambda)d\lambda = \frac{b^a}{\prod_{i=1}^n x_i!\Gamma(a)}\int_0^\infty e^{-(n+b)\lambda}\lambda^{(n\overline{x}+a)-1}d\lambda \\
&= \frac{\Gamma(n\overline{x}+a)b^a}{(n+b)^{n\overline{x}+a}\Gamma(a)\prod_{i=1}^n x_i!}
\end{aligned}
$$

- Logo

$$
\begin{aligned}
p(\lambda|\boldsymbol{x}) &= \frac{\frac{e^{-(n+b)\lambda}\lambda^{(n\overline{x}+a)-1}b^a}{\prod_{i=1}^n x_i!\Gamma(a)}\mathbb{1}_{(0,\infty)}(\lambda)}{\frac{\Gamma(n\overline{x}+a)b^a}{(n+b)^{n\overline{x}+a}\Gamma(a)\prod_{i=1}^n x_i!}} \\
&= \frac{(n+b)^{n\overline{x}+a}}{\Gamma(n\overline{x}+a)}e^{-(n+b)\lambda}\lambda^{(n\overline{x}+a)-1}\mathbb{1}_{(0,\infty)}(\lambda)
\end{aligned}
$$

# Família conjugada (cont.)

- Vamos apresentar algumas ideias sobre como escolher (obter) os valores dos hiperparâmetros.
- Apesar de considerarmos o exemplo em questão, essas ideias podem ser usadas para quaisquer outras situações, em princípio.
- Se $\lambda \sim gama(a, b^{-1})$, então $\mu = \mathcal{E}(\theta) = a/b$ e $\sigma^2 = \mathcal{V}(\theta) = a/b^2$. Podemos pedir ao pesquisador que fixe $\mu$ e $\sigma^2$ e, então, calculamos $a$ e $b$ a partir desses valores, ou seja:

$$a = \frac{\mu^2}{\sigma^2}; b = \frac{\mu}{\sigma^2}$$

# Família conjugada (cont.)

- Por outro lado, podemos utilizar os dados (inferência bayesiana empírica) para obter os hiperparâmetros.
- Voltando à constantes de normalização, temos que

$$p(\boldsymbol{x}|a,b) \quad = \quad \frac{\Gamma(n\overline{x}+a)b^{a}}{(n+b)^{n\overline{x}+a}\Gamma(a)\prod_{i=1}^{n}x_i!}\mathbb{1}_{\{0,1,2,\ldots\}^n}(\boldsymbol{x})$$

$\mathbb{1}_{\{0,1,2,\ldots\}^n}(\boldsymbol{x}) = \prod_{i=1}^{n}\mathbb{1}_{\{0,1,2,\ldots\}}(x_i)$ que corresponde à distribuição preditiva à priori da amostra

- Assim, eliminou-se $\lambda$ da função acima, originando uma espécie de verossimilhança para os hiperparâmetros. Portanto, podemos obter as estimativas de MV para $(a,b)$ e usá-las na priori.

# Família conjugada (cont.)

- Podemos, ainda, tratar os hiperparâmetros (a,b) como parâmetros de interesse, atribuir prioris para eles, e estimá-los através de suas posteriores.
- Essa estrutura é chamada de bayesiana hierárquica. Em nosso caso, podemos considerar:

$$
\begin{aligned}
X_i|\lambda &\overset{i.i.d}{\sim} Poisson(\lambda) \\
\lambda &\sim gama(a, b^{-1}) \\
a &\sim gama(c_1, d_1) \\
b &\sim gama(c_2, d_2)
\end{aligned}
$$

com $(c_1, d_1, c_2, d_2)'$ conhecidos.

# Família conjugada (cont.)

- Neste caso, $(c_1, d_1, c_2, d_2)'$ passam a ser os hiperparâmetros e nosso objetivo é encontrar as posteriores marginais, ou seja:

$$p(\lambda|\boldsymbol{x}), p(a|\boldsymbol{x}), p(b|\boldsymbol{x})$$

a partir da posteriori conjunta:

$$p(\lambda, a, b|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|\lambda)p(\lambda|a,b)p(a)p(b)}{\int_0^\infty \int_0^\infty \int_0^\infty p(\boldsymbol{x}|\lambda)p(\lambda|a,b)p(a)p(b)d\lambda da db}$$

# Família conjugada (cont.)

- Por exemplo, $p(\lambda|\boldsymbol{x}) = \int_0^\infty \int_0^\infty p(\lambda, a, b|\boldsymbol{x}) dadb$
- Na grande maioria dos caros, tais posteriores não apresentam forma analítica conhecida, e precisamos usar métodos numéricos para obter aproximações delas.
- Mesmo em casos em que se utiliza prioris que não correspondem à família conjugada, podemos utilizar as idéias apresentadas anteriormente para obter/escolher os hiperparâmetros.

# Prioris não-informativas

- Se não dispomos de informações que possam nos levar à escolha de uma priori adequada, podemos utilizar prioris não informativas (proporcionais à uma constante) $p(\theta) \propto \mathbb{1}_A(\theta)$, em que $A$ pode ou não ser o espaço paramétrico original.
- Exemplo: Seja $X_i|\theta \overset{iid}{\sim} X|\theta$, $X|\theta \sim \exp(\theta)$ e $p(\theta) \propto \mathbb{1}_{(0,\infty)}(\theta)$, assim

$$p(\theta|\boldsymbol{x}) \propto e^{-\frac{n\overline{x}}{\theta}}\theta^{-n}\mathbb{1}_{(0,\infty)}(\theta) = \frac{(n\overline{x})^{n+1}}{\Gamma(n+1)}e^{-\frac{n\overline{x}}{\theta}}\theta^{-n}\mathbb{1}_{(0,\infty)}(\theta)$$

que corresponde à uma distribuição $IG(n+1, n\overline{x})$, que é própria.

# Máxima entropia

- Seja $\theta \in \Theta^*$, $\Theta^* = \{\theta_1, \theta_2, ..., \theta_k\}$. Então a entropia de uma dada função $h(\theta)$ é definida por

$$\mathcal{E}(h(\theta)) = -\sum_{\theta \in \Theta^*} \ln(h(\theta))h(\theta)$$

- A definição acima também vale se $\Theta^* = \{\theta_1, \theta_2, ...\}$.
- Vamos assumir que $h(\theta) = p(\theta)$, tal que

$$P(\Theta = \theta_i) = p(\theta_i) = p_i, i = 1, 2, ..., n$$

.

# Máxima entropia

- Assim, a priori de máxima entropia para $\theta$, sob a restrição de que $\sum_{i=1}^{k} p_i = 1$ é obtida maximizando-se

$$g(\boldsymbol{p}, \lambda) = -\sum_{i=1}^{k} \ln(p_i) p_i + \lambda \left( \sum_{i=1}^{k} p_i - 1 \right)$$

que resulta em $p(\theta_i) = \dfrac{1}{k} \mathbb{1}_{\{\theta_1, \theta_2, ..., \theta_k\}}(\theta)$, $\boldsymbol{p} = (p_1, ..., p_k)'$.

- Exercício: repetir o processo acima, considerando $\Theta^* = \{\theta_1, \theta_2, ...\}$ e adicionando a restrição de que $E(\Theta) = a$.

# Priori de Jeffreys

- Uma outra forma, é obter as prioris de Jeffreys (PJ) ou de Jeffreys sob independência (PJI). Em geral, estes dois tipos de prioris são vagas ou pouco informativas.
- A PJ e a PJI são obtidas através da informação de Fisher ($I(\theta)$).
- Lembrando: quanto maior for a valor daInformação de Fisher menor a variância do estimador de MV.

# Prioris de Jeffreys

- Então, pensando na IF como uma distribuição de probabilidade, quanto maior for o valor da IF num determinado intervalo, maior a probabilidade (a priori) do parâmetro pertencer à este intervalo.
- Seja $\boldsymbol{\theta} = (\theta_1, .., \theta_k)'$ um vetor de parâmetros e $I(.)$ a informação de Fisher obtida à partir de $p(\boldsymbol{x}|\boldsymbol{\theta})$. A PJ é definida por

$$p^J(\boldsymbol{\theta}) \propto |I(\boldsymbol{\theta})|^{1/2}$$

# Prioris de Jeffreys (cont.)

- A priori de Jeffreys sob independência (PJI) consiste em se considerar os parâmetros como independentes (à priori) e obter a priori de Jeffreys para cada um deles. A PJI será o produtório destas prioris (de acordo com a suposição de independência desejada).
- Suponha que desejamos considerar os parâmetros independentes à priori. Neste caso, teremos:

$$p^I(\boldsymbol{\theta}) = \prod_{i=1}^{k} p^J(\theta_i)$$

# Prioris de Jeffreys (cont.)

- Consideremos o Exemplo 4 (Poisson ($\lambda$)). Nesse caso, temos que $I(\lambda) = \frac{n}{\lambda}$, logo

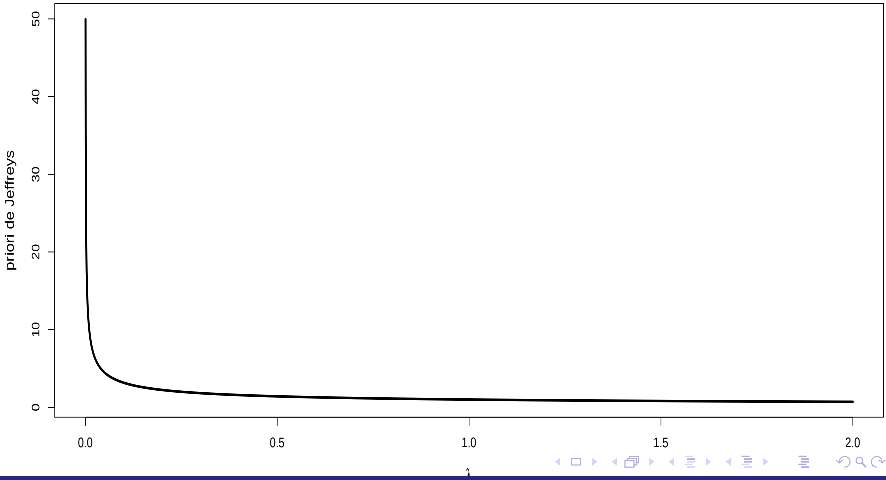
$$p^J(\lambda) \propto \lambda^{-1/2} I_{(0,\infty)}(\lambda)$$

a qual é imprópria. Por outro lado, a posteriori é dada por

$$p^J(\lambda|\boldsymbol{x}) \propto e^{-n\lambda} \lambda^{n\overline{x}-1/2-1} I_{(0,\infty)}(\lambda)$$

ou seja, $\lambda|\boldsymbol{x} \sim gama(n\overline{x} - 1/2, n^{-1})$, a qual é própria.

# Priori de Jeffreys para o modelo de Poisson



# Prioris de Jeffreys (cont.)

- Consideremos o Exemplo da ($N(\mu, \sigma^2)$) com ambos os parâmetros desconhecidos. Nesse caso, temos que ($\boldsymbol{\theta} = (\mu, \sigma^2)'$)

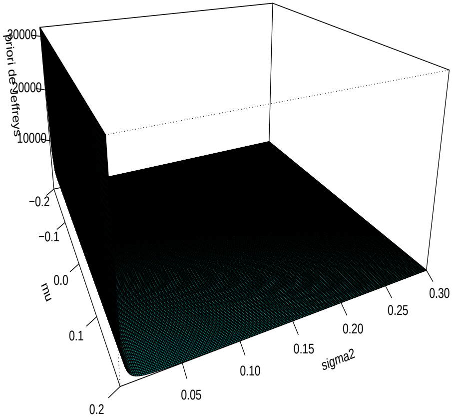
$$I(\boldsymbol{\theta}) = \begin{bmatrix} \dfrac{n}{\sigma^2} & 0 \\ 0 & \dfrac{n}{(\sigma^2)^2} \end{bmatrix}$$

- Assim,

$$p^J(\boldsymbol{\theta}) \propto \left(\sigma^2\right)^{-3/2} I_{(-\infty,\infty)}(\mu) I_{(0,\infty)}(\sigma^2)$$

a qual é imprópria.

# Priori de Jeffreys para o modelo $N(\mu, \sigma^2)$



# Prioris de Jeffreys (cont.)

- Por outro lado, podemos provar que (exercício):

$$p(\boldsymbol{x}|\boldsymbol{\theta}) \propto exp\left\{-\frac{n\left(\mu-\overline{x}\right)^2+(n-1)s^2}{2\sigma^2}\right\}\left(\sigma^2\right)^{-\left(\frac{n}{2}\right)}$$

em que $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ e $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2$.

# Distribuição norma-inversa-gama

- Obs: um vetor aleatório $(X, Y)'$ é dito seguir uma distribuição normal-inversa-gama ($\boldsymbol{Z} = (X, Y) \sim NIG(\alpha, \lambda, a, b)$) se:

$$
\begin{aligned}
f_{\boldsymbol{z}}(x, y|\mu, \lambda, a, b) &= \frac{\sqrt{\lambda}}{\sqrt{2\pi y}} \frac{b^a}{\Gamma(a)} \left(\frac{1}{y}\right)^{a+1} \\
&\times \exp\left\{-\frac{2b + \lambda(x-\alpha)^2}{2y}\right\} \mathbb{1}_{\mathcal{R}}(x) \mathbb{1}_{\mathcal{R}^+}(y)
\end{aligned}
$$

- A definição acima equivale à

$$X|Y = y \sim N(\mu, y/\lambda);\ Y \sim \text{IG}(a, b)$$

## Prioris de Jeffreys (cont.)

- Finalmente, temos que

$$
\begin{aligned}
p(\boldsymbol{\theta}|\boldsymbol{x}) &\propto exp\left\{-\frac{n\,(\mu-\overline{x})^2+(n-1)s^2}{2\sigma^2}\right\}\left(\sigma^2\right)^{-(\frac{n}{2}+1)-\frac{1}{2}} \\
&\times I_{(-\infty,\infty)}(\mu)I_{(0,\infty)}(\sigma^2)
\end{aligned}
$$

a qual correspondes ao núcleo de uma distribuição $NIG(\overline{x}, n, n/2, (n-1)s^2/2)$ (normal-inversa-gama), a qual é própria. Assim,

$$\mu|\boldsymbol{x} \sim t_{(n)}(\overline{x}, \sqrt{(n-1)s^2}/n); \sigma^2|\boldsymbol{x} \sim IG(n/2, (n-1)s^2/2)$$

# Priori de Jeffreys sob independência

- Voltemos ao exemplo da $N(\mu, \sigma^2)$.
- Temos que $I(\mu, \mu) \equiv I(\mu) = \frac{n}{\sigma^2}$ e $I(\sigma^2, \sigma^2) \equiv I(\sigma^2) = \frac{n}{(\sigma^2)^2}$.
- A PJI é obtida como o produtório da PJ para cada parâmetro, como se os outros fossem conhecidos. Com efeito $p^J(\mu) \propto \mathbb{1}_{\mathcal{R}}(\mu)$ e $p(\sigma^2) \propto (\sigma^2)^{-1}\mathbb{1}_{\mathcal{R}^+}(\sigma^2)$. Logo,

$$p^{JI}(\boldsymbol{\theta}) \propto (\sigma^2)^{-1}\mathbb{1}_{\mathcal{R}}(\mu)\mathbb{1}_{\mathcal{R}^+}(\sigma^2)$$

# Priori de Jeffreys sob independência

- Portanto, temos que

$$
\begin{aligned}
p(\boldsymbol{\theta}|\boldsymbol{x}) &\propto exp\left\{-\frac{n\,(\mu-\overline{x})^2+(n-1)s^2}{2\sigma^2}\right\}\left(\sigma^2\right)^{-\left(\frac{n}{2}+\frac{1}{2}\right)-\frac{1}{2}} \\
&\times I_{(-\infty,\infty)}(\mu) I_{(0,\infty)}(\sigma^2)
\end{aligned}
$$

a qual correspondes ao núcleo de uma distribuição $NIG(\overline{x}, n, (n+1)/2, (n-1)s^2/2)$ (normal-inversa-gama), a qual é própria. Assim,

$$
\mu|\boldsymbol{x} \sim t_{(n)}(\overline{x}, \sqrt{(n-1)s^2}/n); \sigma^2|\boldsymbol{x} \sim IG((n+1)/2, (n-1)s^2/2)
$$