

# Inferência Estatística

## 3º Trabalho

Igor Patrício Michels

21/10/2020

### Introdução

Trabalho elaborado pelo aluno Igor Patrício Michels referente a disciplina de Inferência Estatística, do quarto período da Graduação em Matemática Aplicada da FGV-EMAp. Nele faremos uma análise bayesiana da distribuição normal.

O enunciado e eventuais funções utilizadas para resolução deste ou de outros trabalhos podem ser encontrados **nesse repositório do GitHub**.

### Uma Análise Bayesiana no Caso Normal

#### Precisão

Antes de começar nossa análise vamos definir o termo precisão, bem como dar uma pequena intuição de seu significado.

**Definição 1.** *Seja  $\sigma^2$  a variância de uma distribuição. O parâmetro  $\tau = (\sigma^2)^{-1}$  é chamado de precisão.*

Note que a precisão é definida como o recíproco da variância, uma medida de dispersão dos dados. Assim, se possuímos uma variância baixa afirmar ver que os valores tenderam a estarem próximos da média da distribuição. Já se a variância é alta os valores poderão estar mais espalhados ao redor da média, como ilustrado na figura 1.

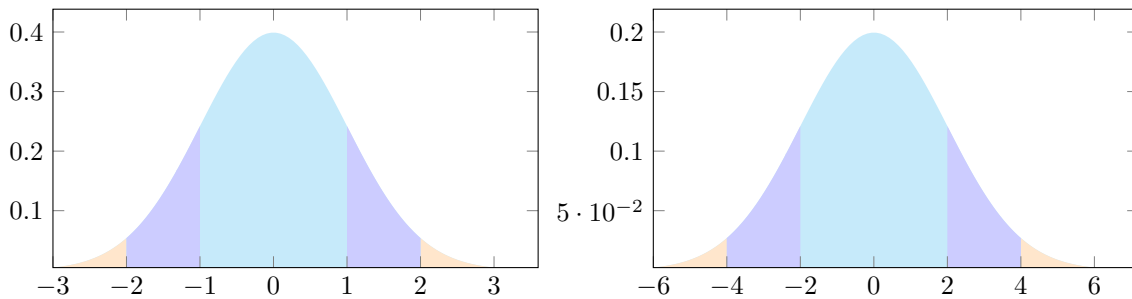


Figura 1: à esquerda temos a plotagem de uma  $\mathcal{N}(0, 1^2)$ , já à direita a plotagem de uma  $\mathcal{N}(0, 2^2)$ .

Note que na figura 1 ambas as p.d.f.'s são iguais, com exceção a escala, a qual foi modificada em virtude da mudança na variância. Esse detalhe nos possibilita perceber a regra 68-95-99.7<sup>1</sup> da normal, além de reforçar a ideia de que, quanto maior a variância, maior a dispersão dos valores em torno da média.<sup>2</sup>

<sup>1</sup>A regra 68-95-99.7 é a propriedade de que aproximadamente 68% dos valores estão a menos de um desvio padrão da média da normal, bem como 95% e 99.7% estão a uma distância inferior a dois e três desvios padrão, respectivamente.

<sup>2</sup>Aqui poderíamos apenas citar a regra 96-95-99.7, entretanto a plotagem da distribuição da normal facilita o entendimento, uma vez que a forma da curva permanece a mesma, mas a escala do eixo  $x$  aumenta.

Dessa forma, a precisão pode ser considerada uma medida de proximidade dos dados, pois a precisão é definida como o recíproco da variância. Assim, quando temos uma variância baixa os valores se encontram perto da média e a precisão é alta, já quando a variância é grande os valores estão mais dispersos e a precisão é menor.

Definida a ideia de precisão podemos nos perguntar qual a utilidade da mesma além de ser uma métrica como citado anteriormente. Podemos responder a essa dúvida mostrando qual a função de densidade de probabilidade de uma variável com distribuição normal em cada uma das parametrizações:

$$X \sim \mathcal{N}(\mu, \sigma^2) \longrightarrow f_X(x \mid \mu, \tau) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{1}{2} \left( \frac{x - \mu}{\sigma} \right)^2 \right]$$

$$X \sim \mathcal{N}_2(\mu, \tau) \longrightarrow f_X(x \mid \mu, \tau) = \sqrt{\frac{\tau}{2\pi}} \exp \left[ -\frac{1}{2} \tau (x - \mu)^2 \right],$$

onde  $\mathcal{N}_2$  representa a normal parametrizada com a média e precisão. Já para uma distribuição conjunta temos as seguintes parametrizações:

- se  $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ :

$$\begin{aligned} f_X(x_1, \dots, x_n \mid \mu, \tau) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{1}{2} \left( \frac{x_i - \mu}{\sigma} \right)^2 \right] \\ &= \left( \frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \prod_{i=1}^n \exp \left[ -\frac{1}{2} \left( \frac{x_i - \mu}{\sigma} \right)^2 \right] \\ &= \left( \sqrt{2\pi\sigma^2} \right)^{-\frac{n}{2}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right] \end{aligned}$$

- já se  $X_1, \dots, X_n \sim \mathcal{N}_2(\mu, \tau)$ :

$$\begin{aligned} f_X(x_1, \dots, x_n \mid \mu, \tau) &= \prod_{i=1}^n \sqrt{\frac{\tau}{2\pi}} \exp \left[ -\frac{1}{2} \tau (x_i - \mu)^2 \right] \\ &= \left( \sqrt{\frac{\tau}{2\pi}} \right)^n \prod_{i=1}^n \exp \left[ -\frac{1}{2} \tau (x_i - \mu)^2 \right] \\ &= \left( \sqrt{\frac{\tau}{2\pi}} \right)^n \exp \left[ -\frac{1}{2} \tau \sum_{i=1}^n (x_i - \mu)^2 \right] \end{aligned} \tag{1}$$

Perceba que, nos dois casos, a segunda representação é mais amigável, uma vez que é mais fácil de manipular os parâmetros manualmente pois não ocorrem divisões por um parâmetro, o que pode ser complicado. Além disso, com essa segunda representação podemos perceber facilmente pequenas variações [4].

## Inferência utilizando a Precisão

Dada a distribuição conjunta condicional dos dados em 1, podemos buscar uma distribuição a priori para  $\mu$  e  $\tau$ . Assim como em [1] e [2], vamos reescrever  $P(\mu, \tau) = P(\mu \mid \tau) \cdot P(\tau)$ , dessa forma, vamos encontrar uma priori em duas etapas. Na primeira etapa, note que, dado  $\tau$ , a distribuição conjunta tem a forma de uma normal, assim escrevemos

$$\mu \mid \tau \sim \mathcal{N}_2(\mu_0, \lambda_0 \tau).$$

Já para a segunda etapa, quando precisamos encontrar a priori de  $\tau$ , podemos notar que, do ponto de vista de  $\tau$ , o núcleo da distribuição em 1 tem a mesma forma que o núcleo de uma distribuição Gama. Dessa forma, temos que

$$\tau \sim \text{Gama}(\alpha_0, \beta_0).$$

Já para o cálculo da posteriori, podemos seguir uma ideia similar a que DeGroot fez em [3]. Sejam  $\xi_1$  e  $\xi_2$  as prioris de  $\mu \mid \tau$  e  $\tau$ , respectivamente. Assim, temos que

$$\xi_1(\mu \mid \tau) \propto \sqrt{\tau} \exp \left[ -\frac{1}{2} \lambda_0 \tau (\mu - \mu_0)^2 \right] \text{ e que } \xi_2(\tau) \propto \tau^{\alpha_0-1} \exp [-\beta_0 \tau].$$

Dessa forma, a posteriori da distribuição conjunta deve satisfazer

$$\begin{aligned} \xi(\mu, \tau \mid x_1, x_2, \dots, x_n) &\propto f_X(x_1, \dots, x_n \mid \mu, \tau) \xi_1(\mu \mid \tau) \xi_2(\tau) \\ &\propto \tau^{\alpha_0-1+(n+1)/2} \exp \left[ -\frac{\tau}{2} \left( \lambda_0 (\mu - \mu_0)^2 + \sum_{i=1}^n (x_i - \mu)^2 \right) - \beta_0 \tau \right]. \end{aligned} \quad (2)$$

Agora, perceba que

$$\begin{aligned} \sum_{i=1}^n (x_i - \mu)^2 &= \sum_{i=1}^n (x_i - \bar{x}_n + \bar{x}_n - \mu)^2 \\ &= \sum_{i=1}^n \left[ (x_i - \bar{x}_n)^2 + 2(x_i - \bar{x}_n)(\bar{x}_n - \mu) + (\bar{x}_n - \mu)^2 \right] \\ &= \sum_{i=1}^n (x_i - \bar{x}_n)^2 + \sum_{i=1}^n 2(x_i - \bar{x}_n)(\bar{x}_n - \mu) + \sum_{i=1}^n (\bar{x}_n - \mu)^2 \\ &= \bar{s}_n^2 + 2(\bar{x}_n - \mu) \sum_{i=1}^n (x_i - \bar{x}_n) + n(\bar{x}_n - \mu)^2 \\ &= \bar{s}_n^2 + 2(\bar{x}_n - \mu) \cdot 0 + n(\bar{x}_n - \mu)^2 \\ &= \bar{s}_n^2 + n(\bar{x}_n - \mu)^2. \end{aligned}$$

Assim, podemos reescrever 2 como

$$\xi(\mu, \tau \mid x_1, x_2, \dots, x_n) \propto \tau^{\alpha_0-1+(n+1)/2} \exp \left[ -\frac{\tau}{2} \left( \lambda_0 (\mu - \mu_0)^2 + \bar{s}_n^2 + n(\bar{x}_n - \mu)^2 \right) - \beta_0 \tau \right]. \quad (3)$$

Note também que

$$\begin{aligned} n(\bar{x}_n - \mu)^2 + \lambda_0 (\mu - \mu_0)^2 &= n\bar{x}_n^2 - 2n\bar{x}_n\mu + n\mu^2 + \lambda_0\mu^2 - 2\lambda_0\mu\mu_0 + \lambda_0\mu_0^2 \\ &= (n + \lambda_0) \left( \mu^2 - \frac{2n\bar{x}_n\mu + 2\lambda_0\mu\mu_0}{n + \lambda_0} + \frac{(n + \lambda_0)(n\bar{x}_n^2 + \lambda_0\mu_0^2)}{(n + \lambda_0)^2} \right), \end{aligned} \quad (4)$$

desenvolvendo a última fração de 4 separadamente, temos

$$\begin{aligned} \frac{(n + \lambda_0)(n\bar{x}_n^2 + \lambda_0\mu_0^2)}{(n + \lambda_0)^2} &= \frac{n^2\bar{x}_n^2 + n\lambda_0\mu_0^2 + n\lambda_0\bar{x}_n^2 + \lambda_0^2\mu_0^2}{(n + \lambda_0)^2} \\ &= \frac{\lambda_0^2\mu_0^2 + 2\mu_0\lambda_0n\bar{x}_n + n^2\bar{x}_n^2 + n\lambda_0\bar{x}_n^2 - 2\mu_0\lambda_0n\bar{x}_n + n\lambda_0\mu_0^2}{(n + \lambda_0)^2} \\ &= \frac{(\lambda_0^2\mu_0^2 + 2\mu_0\lambda_0n\bar{x}_n + n^2\bar{x}_n^2) + n\lambda_0(\bar{x}_n^2 - 2\mu_0\bar{x}_n + \mu_0^2)}{(n + \lambda_0)^2} \\ &= \frac{(\lambda_0\mu_0 + n\bar{x}_n)^2 + n\lambda_0(\bar{x}_n - \mu_0)^2}{(n + \lambda_0)^2}. \end{aligned}$$

Substituindo isso em 4, temos

$$\begin{aligned}
n(\bar{x}_n - \mu)^2 + \lambda_0(\mu - \mu_0)^2 &= (n + \lambda_0) \left( \mu^2 - \frac{2n\bar{x}_n\mu + 2\lambda_0\mu\mu_0}{n + \lambda_0} + \frac{(\lambda_0\mu_0 + n\bar{x}_n)^2 + n\lambda_0(\bar{x}_n - \mu_0)^2}{(n + \lambda_0)^2} \right) \\
&= (n + \lambda_0) \left( \mu^2 - \frac{2\mu(n\bar{x}_n + \lambda_0\mu_0)}{n + \lambda_0} + \frac{(\lambda_0\mu_0 + n\bar{x}_n)^2}{(n + \lambda_0)^2} \right) + \frac{n\lambda_0(\bar{x}_n - \mu_0)^2}{n + \lambda_0} \\
&= (n + \lambda_0) \left( \mu - \frac{\lambda_0\mu_0 + n\bar{x}_n}{n + \lambda_0} \right)^2 + \frac{n\lambda_0(\bar{x}_n - \mu_0)^2}{n + \lambda_0}.
\end{aligned}$$

Dessa forma, substituindo o resultado acima em 3, obtemos

$$\xi(\mu, \tau \mid x_1, x_2, \dots, x_n) \propto \tau^{\alpha_0 - 1 + (n+1)/2} \exp \left[ -\frac{\tau}{2} \left( \bar{s}_n^2 + (n + \lambda_0) \left( \mu - \frac{\lambda_0\mu_0 + n\bar{x}_n}{n + \lambda_0} \right)^2 + \frac{n\lambda_0(\bar{x}_n - \mu_0)^2}{n + \lambda_0} \right) - \beta_0\tau \right].$$

Definindo

$$\begin{aligned}
\lambda_1 &= \lambda_0 + n \\
\mu_1 &= \frac{\lambda_0\mu_0 + n\bar{x}_n}{n + \lambda_0} \\
\alpha_1 &= \alpha_0 + \frac{n}{2} \\
\beta_1 &= \beta_0 + \frac{\bar{s}_n^2}{2} + \frac{n\lambda_0(\bar{x}_n - \mu_0)^2}{2(\lambda_0 + n)},
\end{aligned}$$

podemos escrever o termo dentro da exponencial como<sup>3</sup>

$$\begin{aligned}
&-\frac{\tau}{2} \left( \bar{s}_n^2 + (n + \lambda_0) \left( \mu - \frac{\lambda_0\mu_0 + n\bar{x}_n}{n + \lambda_0} \right)^2 + \frac{n\lambda_0(\bar{x}_n - \mu_0)^2}{n + \lambda_0} \right) - \beta_0\tau \\
&-\frac{1}{2}(\lambda_0 + n)\tau \left( \mu - \left( \frac{\lambda_0\mu_0 + n\bar{x}_n}{n + \lambda_0} \right) \right)^2 - \tau \left( \beta_0 + \frac{\bar{s}_n^2}{2} + \frac{n\lambda_0(\bar{x}_n - \mu_0)^2}{2(n + \lambda_0)} \right) \\
&-\frac{1}{2}\lambda_1\tau(\mu - \mu_1)^2 - \tau\beta_1.
\end{aligned}$$

Já o expoente de  $\tau$  pode ser escrito como

$$\begin{aligned}
\alpha_0 - 1 + \frac{n+1}{2} &= \alpha_0 + \frac{n}{2} - 1 + \frac{1}{2} \\
&= \alpha_1 - 1 + \frac{1}{2}.
\end{aligned}$$

Dessa forma, temos

$$\begin{aligned}
\xi(\mu, \tau \mid x_1, x_2, \dots, x_n) &\propto \tau^{\alpha_1 - 1 + 1/2} \exp \left[ -\frac{1}{2}\lambda_1\tau(\mu - \mu_1)^2 - \beta_1\tau \right] \\
&= \tau^{1/2} \exp \left[ -\frac{1}{2}\lambda_1\tau(\mu - \mu_1)^2 \right] \cdot \tau^{\alpha_1 - 1} \exp[-\beta_1\tau].
\end{aligned}$$

Assim, obtemos a distribuição conjunta a posteriori de  $\mu$  e  $\tau$ . Além disso, o termo

$$\tau^{1/2} \exp \left[ -\frac{1}{2}\lambda_1\tau(\mu - \mu_1)^2 \right],$$

---

<sup>3</sup>Como as expressões são muito grandes, estou manipulando as mesmas de uma linha para outra.

é, exceto por um fator que independe de  $\mu$  e  $\tau$ , a distribuição a posteriori de  $\mu \mid \tau$ , sendo da forma  $\text{Normal}_2$  com média  $\mu_1$  e precisão  $\lambda_1\tau$ . Já o termo

$$\tau^{\alpha_1-1} \exp[-\beta_1\tau]$$

é, também exceto por um fator multiplicativo que independe de  $\mu$  e  $\tau$ , a distribuição a posteriori de  $\tau$ . Note que essa distribuição é uma Gama de parâmetros  $\alpha_1$  e  $\beta_1$ .

Esse conjunto de priors utilizadas acima para esse caso normal em que desconhecemos tanto a média quanto a precisão é conhecida como Distribuição Normal-Gama. Note que essa distribuição é fechada no cálculo da posteriori, ou seja, a posteriori tem a mesma distribuição da priori. Já a distribuição marginal de  $\mu$  a posteriori pode ser encontrada integrando o produto de  $P(\tau)$  com  $P(\mu \mid \tau)$  no domínio de  $\tau$ , isto é  $\mathbb{R}_{>0}$ . Fazendo isto, temos

$$\begin{aligned} P(\mu) &= \int_0^\infty P(\mu \mid \tau) P(\tau) d\tau \\ &\propto \int_0^\infty \tau^{1/2} \exp\left[-\frac{1}{2}\lambda_1\tau(\mu - \mu_1)^2\right] \tau^{\alpha_1-1} \exp[-\beta_1\tau] d\tau \\ &\propto \int_0^\infty \tau^{\alpha_1-1+1/2} \exp\left[-\frac{1}{2}\lambda_1\tau(\mu - \mu_1)^2 - \beta_1\tau\right] d\tau \\ &\propto \left(\frac{2\beta_1 + \lambda_1(\mu - \mu_1)^2}{2}\right)^{-(\alpha_1+\frac{1}{2})} \\ &\propto \left(1 + \frac{(\mu - \mu_1)^2}{\frac{2\beta_1}{\lambda_1}}\right)^{-(\alpha_1+\frac{1}{2})}, \end{aligned}$$

onde podemos ver que  $\mu$  tem distribuição  $t$  deslocada, isto é,

$$\mu \sim X = \mu_1 + \left(\frac{\beta_1}{\alpha_1\lambda_1}\right)^{\frac{1}{2}} T,$$

com  $T$  tendo distribuição  $t$  de Student com  $2\alpha_1$  graus de liberdade.

### Paralelo com o Frequentismo

Fazendo um paralelo com a ideia frequentista, note que a média a posteriori ( $\mu_1$ ) é função da média amostral ( $\bar{x}_n$ ), que é o EMV, e dos hiperparâmetros iniciais:

$$\mu_1 = \frac{\lambda_0\mu_0 + n\bar{x}_n}{n + \lambda_0}.$$

Note que o lado direito da expressão acima pode ser interpretada como uma média ponderada entre  $\mu_0$  e  $\bar{x}_n$ . Dessa forma,  $\mu_0$  tem peso  $\lambda_0$  e  $\bar{x}_n$  tem peso  $n$  e, com isso, nota-se que quando o número de amostras cresce ( $n \rightarrow \infty$ ), temos que  $\mu_1 \rightarrow \bar{x}_n$ . Dessa forma, podemos interpretar  $\lambda_0$  como se fosse uma espécie de “métrica de certeza” ou então uma “métrica de aversão” em relação as ideias, a priori, sobre  $\mu_0$ , como se você tivesse feito o experimento anteriormente  $\lambda_0$  vezes, obtendo uma média  $\mu_0$ .

Exemplificando essa ideia, tomando  $\mu_0$  e  $\lambda_0 > n$  teremos, após a atualização dos hiperparâmetros, um valor de  $\mu_1$  mais próximo de  $\mu_0$  que de  $\bar{x}_n$ . Dessa forma, um  $\lambda_0$  maior que  $n$  implica que, de certo modo, estamos confiando mais nos conhecimentos prévios sobre o experimento que em nossas observações. Podemos notar também que, se  $\lambda_0 \gg n$ , a diferença entre  $\mu_1$  e  $\mu_0$  é muito pequena, o que pode nos levar a ideia do parágrafo anterior, ou seja, a de considerar  $\lambda_0$  como uma “métrica de certeza” ou uma “métrica de aversão”, assim, quanto maior  $\lambda_0$ , podemos dizer que estamos mais certos que sobre  $\mu$  a partir de nossa ideia  $\mu_0$  ou então que estamos mais aversos a alterações na atualização do parâmetro  $\mu_0$ . Note que, nos dois casos, necessitamos uma grande quantidade de amostras para uma alteração significativa de  $\mu_0$  para  $\mu_1$ , quando a

mesma for possível.<sup>4</sup> Uma outra visualização para  $\lambda_0$  seria como quantidade de pseudo-observações, como se você já tivesse realizado esse experimento em algum momento anteriormente, obtendo média  $\mu_0$ .

Já os hiperparâmetros  $\alpha_i$  e  $\beta_i$  acabam, conforme resultado acima, dando forma a distribuição de  $\mu$ . Além disso, note que  $\alpha_0$  e  $\beta_0$  são proporcionais, assim, podemos escolher  $\alpha_0$  e, feito isso, temos  $\beta_0$ . Dessa forma, podemos interpretar  $\alpha_0$  como se fosse a quantidade de pseudo-observações já realizadas.

Após termos visto o desenvolvimento das distribuições a posteriori vamos ilustrar essas ideias com um exemplo.

## Um pequeno exemplo

**Exemplo 1.** *Palmirinha anda preocupada com a concentração de amido em sua pamonha. Ela pede para Valciclei, seu assistente, amostrar  $n = 10$  pamonhas e medir sua concentração de amido.*

*Ele, muito prestativo, rapidamente faz o experimento, mas, porque comeu todas as amostras depois que foram medidas, precisou fazer uma visita de emergência ao banheiro. Desta feita, apenas teve tempo de anotar em um papel a média e variância amostrais,  $\bar{x}_n = 8.307849$  e  $\bar{s}_n^2 = 7.930452$ .*

*Palmirinha tem uma reunião com investidores em pouco tempo, então decide voltar aos seus tempos de bayesiana old school e analisar os dados utilizando prioris conjugadas. Ela supõe que a concentração de amido segue uma distribuição normal com parâmetros  $\mu$  e  $\tau$  e que as observações feitas por Valciclei são independentes entre si. Ela suspeita que a concentração de amido na pamonha fique em torno de 10 mg/L, com desvio padrão de 2 mg/L. Com sua larga experiência na confecção de pamonhas, ela suspeita ainda que o coeficiente de variação da concentração de amido seja em torno de 1/2. Palmirinha tem um quadro em seu escritório, que diz*

$$cv = \frac{\sigma}{\mu}.$$

*Agora,*

- 1. Com os dados anotados por Valciclei, é possível computar a distribuição a posteriori de  $\mu$  e  $\tau$ ? Justifique.*
- 2. Em caso afirmativo, ajude Palmirinha a encontrar  $a, b \in \mathbb{R}$ ,  $a < b$  de modo que  $\Pr(\mu \in (a, b) \mid \mathbf{x}) = 0.95$ .*

Nesse exemplo, apesar de parecer que não é possível encontrar a distribuição a posteriori de  $\mu$  e  $\tau$ , é possível sim, conforme visto anteriormente, encontrar tal distribuição. Abaixo temos duas soluções.

## Uma solução

Primeiramente, note que, dados  $\lambda_0$ ,  $\mu_0$ ,  $\alpha_0$  e  $\beta_0$ , precisamos apenas de  $n$ ,  $\bar{x}_n$  e  $\bar{s}_n^2$  para computar os parâmetros da distribuição a posteriori, a qual já vimos ser Normal-Gama. Agora, perceba que Valciclei fez suas estatísticas baseado em dados de dez pamonhas, ou seja,  $n = 10$ . Além disso, as estatísticas que Valciclei conseguiu anotar antes de fazer sua visita de emergência ao banheiro foram, justamente, a média amostral ( $\bar{x}_n$ ) e a variância amostral ( $\bar{s}_n^2$ ).

Agora precisamos apenas verificar quais são os hiperparâmetros para nossa priori. Note que Palmirinha suspeita que a concentração de amido está em torno de 10 mg/L, ou seja, tomamos  $\mu_0 = 10$ .

Agora, sabemos que Palmirinha suspeita que o desvio padrão da concentração seja igual a 2, ou seja, uma variância igual a 4, além de sua suspeita de que o coeficiente de variação esteja próximo de  $\frac{1}{2}$ . Equacionando essas informações, já usando que  $\mu_0 = 10$ , temos:

$$\sigma^2 = 4 \implies \tau = \frac{1}{4} \implies \frac{\beta_0}{\alpha_0} = \frac{1}{4} \text{ e } cv = \frac{\sigma}{\mu} = \frac{\sqrt{\tau^{-1}}}{\mu} \implies E[cv \mid \mu_0 = 10] = \frac{\sqrt{\alpha_0}}{10 \cdot \sqrt{\beta_0}} \implies \frac{1}{2} = \frac{\sqrt{\alpha_0}}{10 \cdot \sqrt{\beta_0}}.$$

Note que os dois resultados acima são excludentes, uma vez que devemos ter

$$\frac{1}{4} = \frac{\beta_0}{\alpha_0} = \frac{1}{25}.$$

Sendo assim, irei encontrar o parâmetro com 4 prioris distintas:

---

<sup>4</sup>Note que podemos ter  $\mu_0$  muito próximo do valor esperado de  $\mu$ , então ter um  $\lambda_0$  alto, indicando certeza ou aversão, acaba não sendo relevante, pois o esperado é que, com  $n$  grande,  $\bar{x}_n$  também esteja próximo ao valor esperado de  $\mu$ , logo, as diferença entre  $\mu_0$  e  $\mu_1$  acaba se tornando pequena, mesmo com  $\lambda_0$  pequeno.

- usando que  $\frac{\beta_0}{\alpha_0} = \frac{1}{4}$ ;
- usando que  $\frac{\beta_0}{\alpha_0} = \frac{1}{25}$ ;
- usando que  $\frac{\beta_0}{\alpha_0}$  é média aritmética entre  $\frac{1}{4}$  e  $\frac{1}{25}$ , ou seja,  $\frac{29}{200}$  e;
- usando que  $\frac{\beta_0}{\alpha_0}$  é média geométrica entre  $\frac{1}{4}$  e  $\frac{1}{25}$ , ou seja,  $\frac{1}{10}$ .

Note que, dessa forma, escolhendo  $\alpha_0$ , já temos o valor de  $\beta_0$ . Agora, precisamos encontrar o valor de  $\lambda_0$ . Para tanto, podemos utilizar o teorema 8.6.3 do DeGroot [3], o qual afirma que, com  $\alpha_0 > 1$ ,

$$Var(\mu) = \frac{\beta_0}{\lambda_0(\alpha_0 - 1)} \implies \lambda_0 = \frac{\beta_0}{Var(\mu)(\alpha_0 - 1)}.$$

Note que agora nosso problema passa a ser determinar um valor para  $Var(\mu)$ , o que podemos fazer de forma similar ao que DeGroot fez em seu exemplo 8.6.3, dividindo a variância amostral entre  $Var(\mu)$  e  $E[\tau]$ . Fazendo

$$Var(\mu) = \frac{\bar{s}_n^2}{2} = 3.965226 \text{ e } E[\tau] = \frac{1}{\frac{\bar{s}_n^2}{2}} = \frac{1}{3.965226},$$

já temos o valor de  $Var(\mu)$  e, conseqüentemente, podemos encontrar  $\lambda_0$  ao tomar um  $\alpha_0 > 1$  e, além disso, obtemos que o valor esperado de  $\tau$  está próximo a  $\frac{1}{4}$ , que é o que Palmirinha estava supondo. Dessa forma, podemos fazer as contas nos quatro casos dados acima. Para isso, necessitamos escolher um valor para  $\alpha_0$ , assim, optarei por sempre escolher o valor que se encontra no denominador da fração que representa  $\frac{\beta_0}{\alpha_0}$ . Conseqüentemente,  $\beta_0$  sempre o numerador da fração. Os cálculos de  $\lambda_0$ , bem como dos hiperparâmetros a posteriori e do intervalo de credibilidade foram realizados em um Notebook com Python 3.

Assim, no primeiro caso, onde  $\frac{\beta_0}{\alpha_0} = \frac{1}{4}$ , temos:

$$\left\{ \begin{array}{ll} \alpha_0 &= 4 \\ \beta_0 &= 1 \\ \mu_0 &= 10 \\ Var(\mu) &= 3.965226 \implies \lambda_0 \approx 0.08406414497769694, \\ n &= 10 \\ \bar{x}_n &= 8.307849 \\ \bar{s}_n^2 &= 7.930452 \end{array} \right.$$

o que nos retorna os seguintes hiperparâmetros a posteriori:

$$\left\{ \begin{array}{ll} \lambda_1 &= 10.084064144977697 \\ \mu_1 &= 8.321955338966417 \\ \alpha_1 &= 9 \\ \beta_1 &= 5.084576277941813 \end{array} \right.$$

e o intervalo de credibilidade (7.824678461062495, 8.819232216870338).

Já no segundo caso, quando  $\frac{\beta_0}{\alpha_0} = \frac{1}{25}$ , temos:

$$\left\{ \begin{array}{ll} \alpha_0 &= 25 \\ \beta_0 &= 1 \\ \mu_0 &= 10 \\ Var(\mu) &= 3.965226 \implies \lambda_0 \approx 0.010508018122212118, \\ n &= 10 \\ \bar{x}_n &= 8.307849 \\ \bar{s}_n^2 &= 7.930452 \end{array} \right.$$

nos retornando os seguintes hiperparâmetros a posteriori:

$$\left\{ \begin{array}{ll} \lambda_1 &= 10.010508018122213 \\ \mu_1 &= 8.309625248851837 \\ \alpha_1 &= 30 \\ \beta_1 &= 4.980254406354445 \end{array} \right.$$

e o intervalo de credibilidade (8.05203362327982, 8.567216874423854).

No terceiro dos casos, com  $\frac{\beta_0}{\alpha_0} = \frac{29}{200}$ , temos:

$$\left\{ \begin{array}{lcl} \alpha_0 & = & 200 \\ \beta_0 & = & 29 \\ \mu_0 & = & 10 \\ Var(\mu) & = & 3.965226 \\ n & = & 10 \\ \bar{x}_n & = & 8.307849 \\ \bar{s}_n^2 & = & 7.930452 \end{array} \right. \implies \lambda_0 \approx 0.03675166137215896,$$

nos retornando os seguintes hiperparâmetros a posteriori:

$$\left\{ \begin{array}{lcl} \lambda_1 & = & 10.03675166137216 \\ \mu_1 & = & 8.314045164121694 \\ \alpha_1 & = & 205 \\ \beta_1 & = & 33.01765022657346 \end{array} \right.$$

e o intervalo de credibilidade sendo o intervalo (8.065026690836966, 8.563063637406422).

Por fim, no caso em que  $\frac{\beta_0}{\alpha_0} = \frac{1}{10}$ , temos:

$$\left\{ \begin{array}{lcl} \alpha_0 & = & 10 \\ \beta_0 & = & 1 \\ \mu_0 & = & 10 \\ Var(\mu) & = & 3.965226 \\ n & = & 10 \\ \bar{x}_n & = & 8.307849 \\ \bar{s}_n^2 & = & 7.930452 \end{array} \right. \implies \lambda_0 \approx 0.028021381659232316,$$

nos retornando os seguintes hiperparâmetros a posteriori:

$$\left\{ \begin{array}{lcl} \lambda_1 & = & 10.028021381659233 \\ \mu_1 & = & 8.312577391293896 \\ \alpha_1 & = & 15 \\ \beta_1 & = & 5.005231760281796 \end{array} \right.$$

e o intervalo de credibilidade resultante é o intervalo (7.940037725332982, 8.68511705725481).

Podemos ver essas distribuições de  $\mu$  graficamente. Em ciano temos as distribuições a priori, já em azul é a posteriori e, em vermelho, o intervalo de credibilidade. Assim, temos as seguintes distribuições:

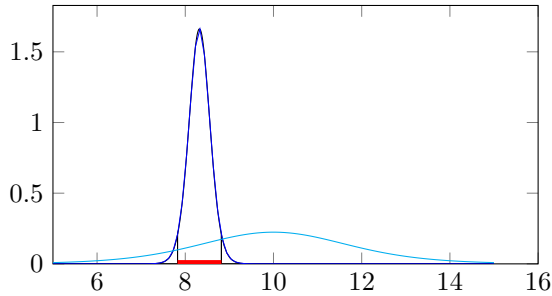


Figura 2: Caso em que  $\frac{\beta_0}{\alpha_0} = \frac{1}{4}$ .

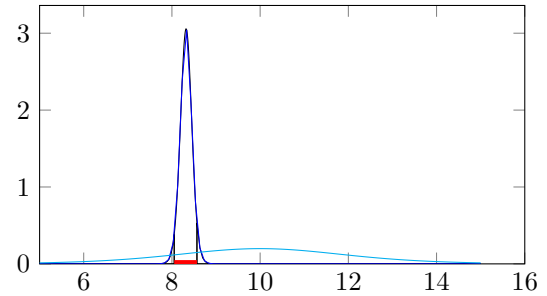


Figura 3: Caso em que  $\frac{\beta_0}{\alpha_0} = \frac{1}{25}$ .



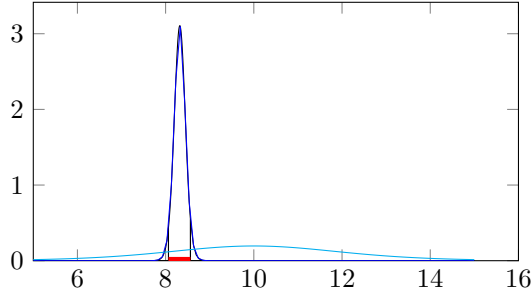


Figura 4: Caso em que  $\frac{\beta_0}{\alpha_0} = \frac{29}{200}$ .

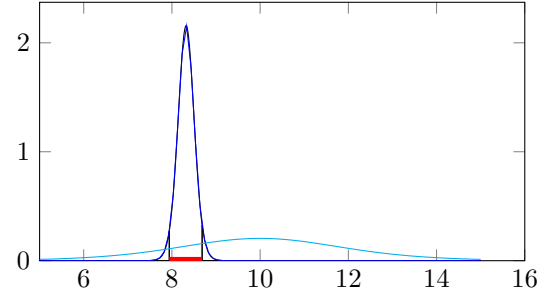


Figura 5: Caso em que  $\frac{\beta_0}{\alpha_0} = \frac{1}{10}$ .

Dessa forma, podemos ver que o intervalo de credibilidade varia conforme nossa escolha de priori. E, como estamos sob o paradigma bayesiano, podemos interpretar o intervalo de credibilidade de 95% como um intervalo em que temos 95% de probabilidade de  $\mu$  estar ali. Note que essa interpretação pode ser contraditória pois, como visto acima, esse intervalo varia conforme nossa escolha de priori, entretanto, quando  $n$  cresce, o intervalo de credibilidade diminui, ou seja, a variância a posteriori de  $\mu$  diminui e, conseqüentemente, temos mais certeza sobre a distribuição de  $\mu$ . Para ilustrar isso fiz algumas contas e simulações no Notebook Python que pode ser encontrado no GitHub.

### Uma outra solução

De forma análoga a anterior, precisamos apenas encontrar os valores de  $\mu_0$ ,  $\alpha_0$ ,  $\beta_0$  e  $\mu_0$ . Feito isso, podemos fazer o cálculo da posteriori, isto é, a atualização desses parâmetros.

De acordo com o enunciado, Palmirinha acha que  $\mu$  tem média 10 e desvio padrão igual a 2, ou seja, tomando  $\alpha_0 > \frac{1}{2}$ , podemos dizer que, pelo Teorema 8.6.3 do DeGroot [3],  $E[\mu] = \mu_0$ . Dessa forma, podemos definir  $\mu_0 = 10$  e, por enquanto,  $\alpha_0 > \frac{1}{2}$ . Além disso, o mesmo teorema afirma que, se  $\alpha_0 > 1$ , então

$$Var(\mu) = \frac{\beta_0}{\lambda_0(\alpha_0 - 1)} \implies \lambda_0 = \frac{\beta_0}{Var(\mu)(\alpha_0 - 1)},$$

ou seja, escolhendo  $Var(\mu) = 4$ , conforme suspeita a priori de Palmirinha, temos  $\lambda_0$  a partir das escolhas e  $\beta_0$  e  $\alpha_0$ , com este último maior que 1.

Agora, manipulando o coeficiente de variação, temos que

$$\begin{aligned} cv &= \frac{\sigma}{\mu} \\ &= \frac{\sqrt{\sigma^2}}{\mu} \\ &= \frac{\sqrt{\tau^{-1}}}{\mu} \\ &= \frac{1}{\mu\sqrt{\tau}}. \end{aligned}$$

Assim, se temos que o coeficiente de variação é de  $\frac{1}{2}$ , conforme suspeita da Palmirinha, então vale que

$$\frac{1}{2} = \frac{1}{\mu\sqrt{\tau}} \implies \tau = \frac{4}{\mu^2}.$$

Agora, podemos supor um  $\mu$  fixo, dessa forma, temos que

$$E[\tau] = \frac{4}{\mu^2}.$$

Escolhendo para fixar  $\mu$  podemos pensar em tomar  $\mu_0$  como valor para  $\mu$ , o que é razoável dado que  $\mu_0$  é a média e a moda a priori para  $\mu$ . Dessa forma, temos que

$$\frac{\alpha_0}{\beta_0} = \frac{4}{\mu_0^2} \implies \frac{\alpha_0}{\beta_0} = \frac{1}{25}.$$

Como temos que  $\alpha_0 > 1$ , podemos tomar  $\alpha_0 = 2$  e, consequentemente,  $\beta_0 = 50$ . Dessa forma,

$$\lambda_0 = \frac{50}{4(2-1)} = \frac{25}{2}$$

e, com isso, podemos atualizar os parâmetros. Temos

$$\left\{ \begin{array}{ll} \alpha_0 &= 2 \\ \beta_0 &= 50 \\ \mu_0 &= 10 \\ Var(\mu) &= 4 \\ n &= 10 \\ \bar{x}_n &= 8.307849 \\ \bar{s}_n^2 &= 7.930452 \\ \lambda_0 &= 12.5, \end{array} \right.$$

o que nos gera os seguintes parâmetros a posteriori:

$$\left\{ \begin{array}{ll} \lambda_1 &= 22.5 \\ \mu_1 &= 9.247932888888888 \\ \alpha_1 &= 7 \\ \beta_1 &= 61.91904546333612 \end{array} \right.$$

e o intervalo de credibilidade igual a  $(7.903138300680498, 10.592727477097279)$ .

Note que, na solução acima, precisamos escolher apenas um valor para  $\alpha_0$ . Assim, podemos nos perguntar o que aconteceria com um  $\alpha_0$  diferente. Chutando um pouco o balde e escolhendo  $\alpha_0 = 100$ , consequentemente  $\beta_0 = 2500$ , obtemos algumas alterações nos demais parâmetros a posteriori e, no fim, uma pequena modificação no intervalo de credibilidade, o qual passa a ser  $(6.576604381280694, 11.348808015004133)$ . Graficamente, temos os seguintes cenários:

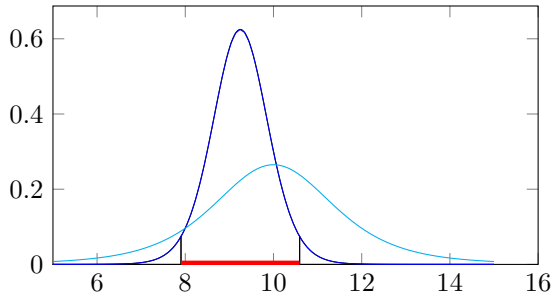


Figura 6: Distribuição de  $\mu$  quando  $\alpha_0 = 2$ .

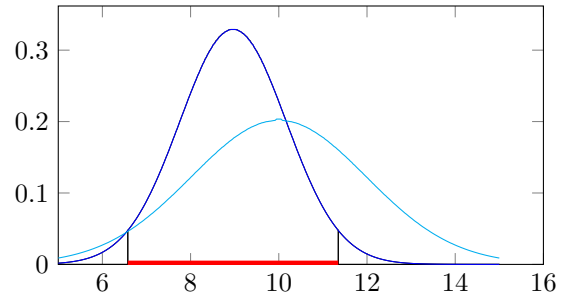


Figura 7: Distribuição de  $\mu$  quando  $\alpha_0 = 100$ .

Note que com  $\alpha_0$  maior temos um intervalo de credibilidade maior a posteriori, ou seja, nossa distribuição a posteriori de  $\mu$  tem maior variância. Agora, mantendo a média, variância amostral e parâmetros iniciais, mas mudando  $n$  para 100, como se Valciclei tivesse amostrado mais pamonhas. Fazendo isso, temos as seguintes distribuições para  $\mu$ :

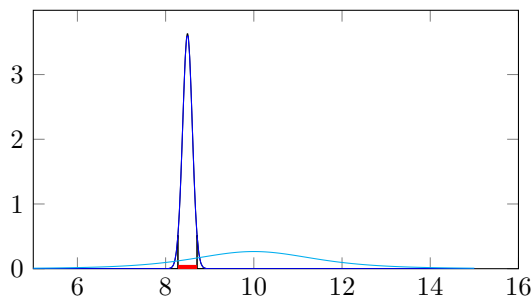


Figura 8: Distribuição de  $\mu$  quando  $\alpha_0 = 2$  e  $n = 100$ .

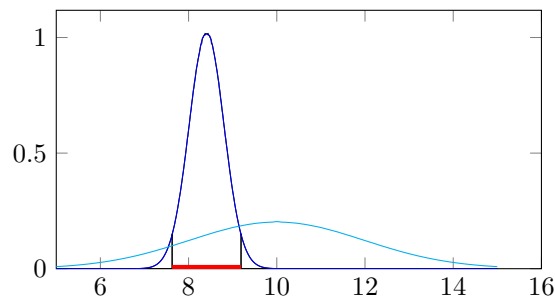


Figura 9: Distribuição de  $\mu$  quando  $\alpha_0 = 100$  e  $n = 100$ .

Note que, a medida com que  $n$  cresce, o intervalo de credibilidade diminui mostrando, assim como dito anteriormente, que a distribuição de  $\mu$  fica mais concentrada, com variância menor e, dessa forma, podemos ver as posteriores tendem a convergir para uma mesma distribuição, mesmo com prioris com  $\alpha_0$  distintos.

## Conclusão

Já no início pudemos ter uma boa ideia do significado da precisão, sendo visualizada como uma métrica de proximidade dos dados em relação a média. Após isso passamos a fazer uma análise bayesiana com a normal com média e precisão desconhecidas. Feita a análise, descobrimos mais uma família conjugada, agora com duas distribuições a priori. Além disso, a visualização de  $\lambda_0$  como um nível de certeza a priori acabou deixando o cálculo da média a posteriori,  $\mu_1$ , um pouco mais intuitivo. Além dele, vimos que  $\alpha_0$  e  $\beta_0$  são hiperparâmetros que irão contribuir na forma da posteriori de  $\mu$ , ditando a concentração em torno de  $\mu_1$ .

Chegando no exemplo vimos que a escolha da priori tem certa relevância no cálculo da posteriori, embora os intervalos de credibilidade que foram gerados ficaram próximos. Além disso, vimos que a medida em que  $n$  cresce essa relevância da priori diminui. Ou seja, a priori pode ser gerada de diversas formas, a depender da interpretação de quem vai elaborá-la, mas a posteriori nos dará resultados similares e, com uma grande amostragem, as posteriores tendem a uma mesma distribuição.

Assim, uma das maiores lições desse trabalho é que elicitar prioris pode ser muito complicado, podendo ser considerada uma arte na qual o ponto de vista e interpretação do artista pode ser crucial para obtermos um resultado razoável ou então um totalmente descartável.

Por fim, talvez seria interessante alguém falar com Palmirinha para ela buscar um assistente menos comilão, imagine ele fazendo uma amostragem de 100 pamonhas!

## Referências

- [1] Ricardo Ehlers. *Inferência Bayesiana (Notas de Aula)*. URL: <https://sites.icmc.usp.br/ehlers/bayes/>. (acessado: 02.10.2020).
- [2] Ricardo S. Ehlers. *Introdução a Inferência Bayesiana*. URL: <http://www.leg.ufpr.br/~paulojus/CE227/ce227/>. (acessado: 03.10.2020).
- [3] Mark J. Schervish Morris H. DeGroot. *Probability and Statistics*. 4ª ed. Pearson, 2011. ISBN: 9780321500465.
- [4] *What's in a name: Precision (inverse of variance)*. URL: <https://stats.stackexchange.com/questions/211419/whats-in-a-name-precision-inverse-of-variance>. (acessado: 30.09.2020).