Fundação Getúlio Vargas

Inferência Estatística

Trabalho III: Análise Bayesiana

Rener de Souza Oliveira
rener.oliveira@fgv.edu.br

Rio de Janeiro, RJ
20 de Outubro de 2020

Conteúdo

1 Introdução **2**

2 A distribuição Normal-Gama **2**

2.1 Distribuições à posteriori 3

2.2 Análise dos hiperparâmetros 5

3 A distribuição marginal de μ **6**

4 O Dilema da Palmirinha **7**

4.1 Posterioris 7

4.2 Intervalos de Credibilidade 7

1 Introdução

"Neste trabalho, vamos derivar os principais resultados de uma análise bayesiana conjugada de dados normalmente distribuídos. Para tal, começamos com uma reparametrização. Em particular, fazemos $\tau = 1/\sigma^2$, de modo que os parâmetros de interesse $\theta = (\mu, \sigma^2)$ se tornem $\phi = (\mu, \tau)$. O parâmetro τ é chamado de *precisão*. Suponha que observamos uma amostra aleatória $X_1, \dots, X_n$ com distribuição normal com parâmetros μ e τ, ambos desconhecidos"[1]

2 A distribuição Normal-Gama

Iremos derivar uma família de distribuições que serão conjugadas que funcionará como priori conjugada para ϕ. Ou seja, dada uma priori para ϕ de distribuição normal-gama, teremos que a distribuição da posteriori também será normal-gama.
Primeiramente, vamos determinar a distribuição conjunta condicional dos dados sob a nova parametrização. Para um dado $X_i \sim \text{Normal}(\mu, \sigma^2)$ temos a distribuição de densidade de probabilidade como:

$$f(x_i|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}\right]$$

Fazendo $\sigma^2 = 1/\tau$, temos:

$$f(x_i|\mu, \tau) = \sqrt{\frac{\tau}{2\pi}} \exp\left[-\frac{\tau}{2}(x_i-\mu)^2\right]$$

Usaremos a notação $X_i \sim \text{Normal}_2(\mu, \tau)$, para se referir à esta nova parametrização que utiliza precisão. Assumindo que os dados são independentes e identicamente distribuídos, temos que a função de verossimilhança, ou distribuição conjunta consicional dos dados é:

$$\begin{aligned} f(\vec{X}|\phi) &= \prod_{i=1}^{n} f(x_i|\phi) \\ &= \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left[-\frac{\tau}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right] \end{aligned} \tag{1}$$

Vamos assumir que a distribuição à priori de τ é $\text{Gama}(\alpha_0, \beta_0)$ e a distribuição à priori condicional de $\mu|\tau$ é Normal de média μ_0 e precisão $\lambda_0\tau$. Essa é uma premissa razoável por dois motivos: o primeiro é o que o espaço de definição dos parâmetros coincide com o espaço onde as densidades das distrivuições citadas estão definidas, μ é definino na reta assim como a normal, e τ é definido para os positivos assim como a gama; o segundo motivo é que pela expressão da verossimilhança (1), se multiplicarmos com a densidade de uma normal, a forma funcional parece que também irá se comportar como uma normal, o mesmo ocorre para gama. Mostraremos isso de forma mais clara a seguir, provando que as distribuições à posteriori são da mesma família. Mas antes, calcularemos a priori conjunta de $\phi = (\mu, \tau)$:
Dados $\xi_1(\mu|\tau) = \sqrt{\frac{\lambda_0\tau}{2\pi}} \exp\left[-\frac{\lambda_0\tau}{2}(\mu-\mu_0)^2\right]$ e $\xi_2(\tau) = \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)}\tau^{\alpha_0-1}e^{-\beta_0\tau}$, temos que a distribuição conjunta $\xi(\mu, \tau)$ é:

$$\begin{aligned} \xi(\mu, \tau) &= \xi_1(\mu|\tau)\xi_2(\tau) \\ &= \frac{\beta_0^{\alpha_0}\sqrt{\lambda_0}}{\Gamma(\alpha_0)\sqrt{2\pi}}\tau^{\alpha_0-\frac{1}{2}}e^{-\beta_0\tau}e^{-\frac{\lambda_0\tau(\mu-\mu_0)^2}{2}} \end{aligned} \tag{2}$$

$$\text{para } \mu \in \mathbb{R} \text{ e } \tau \in \mathbb{R}_{>0}.$$

Definição 1 (Distribuição Normal-Gama:) *A distribuição acima é conhecida como* Normal-Gama *de parâmetros* $\mu_0, \lambda_0, \alpha_0, \beta_0$. *Pode ser fatorizada facilmente como um produto das densidades de uma* $\text{Gama}(\alpha_0, \beta_0)$ *e uma* $\text{Normal}_2(\mu_0, \lambda_0\tau)$, *pela própria construção. Além disso a região de definição dos parâmetros é:* $\mu_0 \in \mathbb{R}$, $\lambda_0 \in \mathbb{R}_{>0}$, $\alpha_0 \in \mathbb{R}_{>0}$ *e* $\beta_0 \in \mathbb{R}_{>0}$.

[1] Trecho retirado da descrição do trabalho.

De fato, trata-se de uma densidade válida pois fatorizando $\tau^{\alpha_0-1/2} = \tau^{\alpha_0-1}\tau^{1/2}$ a integral dupla da função sobre o espaço de definição de ϕ é facilmente calculável via integração iterada de Fubini[2] e resulta em 1:

$$\begin{aligned}\int_0^{+\infty}\int_{-\infty}^{+\infty}\xi(\mu,\tau)d\mu d\tau &= \int_0^{+\infty}\left(\frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)}\tau^{\alpha_0-1}e^{-\beta_0\tau}\int_{-\infty}^{+\infty}\sqrt{\frac{\lambda_0\tau}{2\pi}}e^{-\frac{\lambda_0\tau(\mu-\mu_0)^2}{2}}d\mu\right)d\tau\\ &= \int_0^{+\infty}\frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)}\tau^{\alpha_0-1}e^{-\beta_0\tau}d\tau = 1\end{aligned}$$

2.1 Distribuições à posteriori

Com base nos resultados (1) e (2), podemos derivar a distribuição à posteriori conjunta de $\phi = (\mu, \tau)$ e a partir dela, a distribuição condicional de $\mu|\tau$ e as marginais à posteriori de τ e μ.
Notações:

$$\begin{aligned}\bar{x}_n &= \frac{1}{n}\sum_{i=1}^{n}x_i\\ s_n^2 &= \sum_{i=1}^{n}(x_i-\bar{x}_n)^2\end{aligned}$$

Teorema 1 *[1] Suponha que $X_1, ..., X_n$ formam uma amostra aleatória de um distribuição Normal de média μ desconhecida e precisão τ também desconhecida ($\mu \in \mathbb{R}$ e $\tau \in \mathbb{R}_{>0}$). Se a distribuição conjunta à priori de $\phi = (\mu,\tau)$ é* Normal-Gama$(\mu_0, \lambda_0, \alpha_0, \beta_0)$*, então a distribuição à posteiori é* Normal-Gama$(\mu_1, \lambda_1, \alpha_1, \beta_1)$*, onde:*

$$\mu_1 = \frac{\lambda_0\mu_0 + n\bar{x}_n}{\lambda_0+n} \qquad \lambda_1 = \lambda_0 + n$$

$$\alpha_1 = \alpha_0 + \frac{n}{2} \qquad \beta_1 = \beta_0 + \frac{1}{2}s_n^2 + \frac{n\lambda_0(\bar{x}_n-\mu_0)^2}{2(\lambda_0+n)}$$

Consequentemente, a distribuição condicional de $\mu|\tau$ é Normal(μ_1, λ_1) *e a marginal à posteriori de τ é* Gama(α_1, β_1).

Prova:
Assumindo que a distribuição conjunta à posteriori $f(\phi|\vec{X})$ seja realmente Normal-Gama$(\mu_1, \lambda_1, \alpha_1, \beta_1)$, teremos da Definição (1):

$$\begin{aligned}f(\phi|\vec{X}) &\propto \tau^{\alpha_1-\frac{1}{2}}e^{-\beta_1\tau}e^{-\frac{\lambda_1\tau(\mu-\mu_1)^2}{2}}\\ &\propto \tau^{\alpha_1+\frac{1}{2}-1}e^{-\beta_1\tau}\exp\left[-\frac{\lambda_1\tau(\mu-\mu_1)^2}{2}\right]\\ &\propto \tau^{\alpha_1-1}e^{-\beta_1\tau}\tau^{\frac{1}{2}}\exp\left[-\frac{\lambda_1\tau(\mu-\mu_1)^2}{2}\right]\end{aligned}$$

Daí podemos extrair:

$$\begin{aligned}f(\tau|\vec{X}) &= \int_0^{\infty}f(\phi|\vec{X})d\mu\\ &\propto \tau^{\alpha_1-1}e^{-\beta_1\tau}\end{aligned} \tag{3}$$

e também:

$$\begin{aligned}f(\mu|\tau,\vec{X}) &= \frac{f(\mu,\tau|\vec{X})}{f(\tau|\vec{X})}\\ &\propto \frac{\tau^{\alpha_1-1}e^{-\beta_1\tau}\tau^{\frac{1}{2}}\exp\left[-\frac{\lambda_1\tau(\mu-\mu_1)^2}{2}\right]}{\tau^{\alpha_1-1}e^{-\beta_1\tau}}\\ &\propto \tau^{\frac{1}{2}}\exp\left[-\frac{\lambda_1\tau(\mu-\mu_1)^2}{2}\right]\end{aligned} \tag{4}$$

De (3) segue que a distribuição à posteriori de $\tau|\vec{X} \sim \text{Gama}(\alpha_1, \beta_1)$ e de (4) segue que $\mu|\tau, \vec{X} \sim \text{Normal}_2(\mu_1, \lambda_1\tau)$ Provamos então a segunda parte do Teorema, mas ainda não provamos que $f(\phi|\vec{X}) \sim \text{Normal-Gama}(\mu_1, \lambda_1, \alpha_1, \beta_1)$. Multiplicando à densidade da distribuição à priori (2) com a função de verossimilhança dos dados (1), teremos:

$$\xi(\mu,\tau)f(\vec{X}|\mu,\tau) \propto \tau^{\alpha_0+\frac{n}{2}-\frac{1}{2}}e^{-\beta_0\tau}\exp\left[-\frac{\tau}{2}\left(\lambda_0(\mu-\mu_0)+\sum_{i=1}^{n}(x_i-\mu)^2\right)\right] \tag{5}$$

Identidade 1

$$\sum_{i=1}^{n}(x_i-\mu)^2 = s_n^2 + n(\bar{x}_n-\mu)^2$$

Prova:

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

Aplicando a Identidade acima na expressão (5), temos:

$$\xi(\mu,\tau)f(\vec{X}|\mu,\tau) \propto \tau^{\alpha_0+\frac{n}{2}-\frac{1}{2}}e^{-\beta_0\tau}\exp\left[-\frac{\tau}{2}\left(\lambda_0(\mu-\mu_0)+s_n^2+n(\bar{x}_n-\mu)^2\right)\right]$$

Vamos agora para mais manipulações algébricas,

Identidade 2

$$\lambda_0(\mu-\mu_0)+n(\bar{x}_n-\mu)^2 = (\lambda_0+n)(\mu-\mu_1)+\frac{n\lambda_0(\bar{x}_n-\mu_0)^2}{\lambda_0+n}$$

Prova:

$$\begin{aligned}
\lambda_0(\mu-\mu_0)+n(\bar{x}_n-\mu)^2 &= \lambda_0(\mu^2+2\mu\mu_0+\mu_0^2)+n(\mu^2-2\mu\bar{x}_n+\bar{x}_n^2) \\
&= \mu^2(\lambda_0+n)+\mu(-2\lambda_0\mu_0-2n\bar{x}_n)+\lambda_0\mu_0^2+n\bar{x}_n^2 \\
&= (\lambda_0+n)\left[\mu^2-\frac{2\mu(\lambda_0\mu_0+n\bar{x}_n)}{\lambda_0+n}\right]+\lambda_0\mu_0^2+n\bar{x}_n^2 \\
&= (\lambda_0+n)(\mu-\mu_1)^2-\mu_1^2(\lambda_0+n)+\lambda_0\mu_0^2+n\bar{x}_n^2 \\
&= (\lambda_0+n)(\mu-\mu_1)^2+\frac{-(\lambda_0\mu_0+n\bar{x}_n)^2+(\lambda_0+n)(\lambda_0\mu_0^2+n\bar{x}_n^2)}{\lambda_0+n} \\
&= (\lambda_0+n)(\mu-\mu_1)^2+\frac{-\cancel{\lambda_0^2\mu_0^2}-2\lambda_0\mu_0n\bar{x}_n-\cancel{n^2\bar{x}_n^2}+\cancel{\lambda_0^2\mu_0^2}+\lambda_0n\bar{x}_n^2+n\lambda_0\mu_0^2+\cancel{n^2\bar{x}_n^2}}{\lambda_0+n} \\
&= (\lambda_0+n)(\mu-\mu_1)^2+\frac{n\lambda_0(\bar{x}_n^2-2\mu_0\bar{x}_n+\mu_0^2)}{\lambda_0+n} \\
&= (\lambda_0+n)(\mu-\mu_1)^2+\frac{n\lambda_0(\bar{x}_n-\mu_0)^2}{\lambda_0+n}
\end{aligned}$$

Voltando ao trabalho, e aplicando a nova identidade acima temos:

$$\begin{aligned}
\xi(\mu,\tau)f(\vec{X}|\mu,\tau) &\propto \tau^{\alpha_0+\frac{n}{2}-\frac{1}{2}}e^{-\beta_0\tau}\exp\left[-\frac{\tau}{2}\left(s_n^2+(\lambda_0+n)(\mu-\mu_1)^2+\frac{n\lambda_0(\bar{x}_n-\mu_0)^2}{\lambda_0+n}\right)\right] \\
&\propto \tau^{\alpha_0+\frac{n}{2}-\frac{1}{2}}\exp\left[-\beta_0\tau-\frac{\tau s_n^2}{2}-\frac{\tau n\lambda_0(\bar{x}_n-\mu_0)^2}{2(\lambda_0+n)}\right]\exp\left[-\frac{\tau}{2}(\lambda_0+n)(\mu-\mu_1)^2\right] \\
&\propto \tau^{\alpha_1-\frac{1}{2}}e^{-\beta_1\tau}\exp\left[-\frac{\lambda_1\tau(\mu-\mu_1)^2}{2}\right]
\end{aligned}$$

Esse é o numerador da dossa distribuição à posteriori conjunta. O denominador será a integral do numerador sobre o espaço de definição de $\phi=(\mu,\tau)$, mas o numerador é o núcleo de uma distribuição Normal-Gama$(\mu_1,\lambda_1,\alpha_1,\beta_1)$, logo, à menos que uma constante que não depende de ϕ, a integral é igual a 1, portanto:

$$\xi(\mu,\tau|\vec{X}) \propto \tau^{\alpha_1-\frac{1}{2}}e^{-\beta_1\tau}\exp\left[-\frac{\lambda_1\tau(\mu-\mu_1)^2}{2}\right], \tag{6}$$

que é núcleo da distribuição almejada, o que finaliza a prova do teorema.

2.2 Análise dos hiperparâmetros

Vejamos novamente as espressões dos parâmetros à posteriori:

$$\mu_1=\frac{\lambda_0\mu_0+n\bar{x}_n}{\lambda_0+n} \qquad \lambda_1=\lambda_0+n$$

$$\alpha_1=\alpha_0+\frac{n}{2} \qquad \beta_1=\beta_0+\frac{1}{2}s_n^2+\frac{n\lambda_0(\bar{x}_n-\mu_0)^2}{2(\lambda_0+n)}$$

As expressões de μ_1 e λ_1 são bastante sugestivas, a primeira é uma média ponderada entre μ_0 e $\bar{x}_n$ onde o peso de μ_0 é o hiperparâmetro λ_0 e o peso de $\bar{x}_n$ é o número de observações n. Ou seja, a média a posteriori vai juntar o conhecimento à priori da média, com o conhecimento obtido a partir dos dados por $\bar{x}_n$ que é estimador de máxima verossimilhança pra média.
Além disso $\lim_{n\to\infty}\mu_1=\lim_{n\to\infty}\left(\frac{\lambda_0}{\lambda_0+n}\mu_0+\frac{n}{\lambda_0+n}\bar{x}_n\right)=\bar{x}_n$,
mostrando que a média à posteriori converge para o estimador de máxima verossimilhança com n grande.
O hiperparâmetro λ_0 também é interpretado como número de pseudo-observações[3] representando de certa forma, observações conhecidas à priori, ou simplesmente, o conhecimento do estatístico que está fazendo a elicitação das prioris.
A expressão $\lambda_1=\lambda_0+n$ incrementa esse número de pseudo-obervações com as novas obervações, vindas dos dados.
Também podemos interpretar os parâmetros da Gama nesse contexto de pseudo-observações[3]. Podemos enxergar como se tivéssesmos à priori $2\alpha_0$ pseudo-observações[2] do experimento Normal$_2(\mu,\tau)$, a variância amostral desse experimento é o valor esperado de τ^{-1}, ou seja $\frac{\beta_0}{\alpha_0}$.
Interpretando α_0 como metade da quantidade de pseudo-observações. O incremento ao observar os dados é de $n/2$ como sugere a expresão à posteriori.
A interpretação de β_0 neste contexto será metade da soma dos erros quadráticos à priori; isso pois, como a variância amostral do pseudo-experimento é β_0/α_0 ao multiplicar pela quantidade de pseudo-observações teremos a soma dos erros quadráricos amostrais que é $2\beta_0$.
A expresssão de β_1 deve então simbolizar a soma dos erros quadráticos do experimento à posteriori; Assim, somando à β_0, $s_n^2/2$ que é metade do erro quadrático dos dados em relacão a média amostral, já temos parte da expressão.

[2] $2\alpha_0$ não necessariamente precisa ser igual a λ_0. Essa independência é introduzida na interpretação de forma que os pseudo-experimentos Normal e Gamma sejam diferentes.

O terceiro termo $\dfrac{n\lambda_0(\bar{x}_n-\mu_0)^2}{2(\lambda_0+n)}$ pode ser entendido como um "termo de interação", e sua existência se justifica pelo fato dos erros anteriores terem sido calculados em relação a médias diferentes: $2\beta_0$ é erro quadrático das pseudo-observações, apesar de não termos explicitado, é óbvio que isso é calculado em ralação à μ_0 que pode ser interpretado como média amostral do pseudo-experimento neste contexto. Assim, adicionamos uma penalidade sobre a distância dessa média μ_0 e a observada pelos dados $\bar{x}_n$, com o termo $\dfrac{n\lambda_0}{\lambda_0+n}$ funcionando como uma constante balizadora entre pseudo-obervações e observações dos dados de fato.
Essa interpretação do β_0 e β_1 como erros quadráticos faz sentido, pois numa situação onde esses erros são altos, significa que há muita incerteza sobre a distribuição dos dados, e isso é refletido como uma redução do valor esperado de τ, já que β_0 β_1 ficam no denominador de $E(\tau|\vec{X})=\alpha_1/\beta_1$. Na expressão de $\alpha_1=\alpha_0+n/2$ essa noção é ainda mais clara, pois, fixando β_1, quando maior for o n (quanto mais dados observamos), nossa precisão aumenta pelo fato de termos mais informações.

3 A distribuição marginal de μ

Por definição, a função de densidade da distribuição marginal à posteriori de μ é calculada por:

$$f_\mu(\mu|\vec{X})=\int_0^\infty \xi(\mu,\tau|\vec{X})d\tau$$

De (6), temos:

$$\begin{aligned}
f_\mu(\mu|\vec{X}) &= c\cdot\int_0^\infty \tau^{\alpha_1-\frac{1}{2}}e^{-\beta_1\tau}\exp\left[-\frac{\lambda_1\tau(\mu-\mu_1)^2}{2}\right]d\tau\\
&= c\cdot\int_0^\infty \tau^{\alpha_1+\frac{1}{2}-1}\exp\left[-\tau\left(\beta_1+\frac{\lambda_1(\mu-\mu_1)^2}{2}\right)\right]d\tau,
\end{aligned}$$

Onde c é a constante da densidade da Normal-Gamma: $c=\dfrac{\beta_1^{\alpha_1}\lambda_1^{\frac{1}{2}}}{\Gamma(\alpha_1)(2\pi)^{\frac{1}{2}}}$. Note que o integrando acima é o núcleo de uma distribuição Gama $\left(\alpha_1+\frac{1}{2},\beta_1+\dfrac{\lambda_1(\mu-\mu_1)^2}{2}\right)$. Assim:

$$\begin{aligned}
f_\mu(\mu|\vec{X}) &= c\cdot\frac{\Gamma(\alpha_1+\frac{1}{2})}{\left(\beta_1+\frac{\lambda_1(\mu-\mu_1)^2}{2}\right)^{\alpha_1+\frac{1}{2}}}\\
&= \frac{\beta_1^{\alpha_1}\lambda_1^{\frac{1}{2}}\Gamma(\alpha_1+\frac{1}{2})}{\Gamma(\alpha_1)(2\pi)^{\frac{1}{2}}}\cdot\left(\beta_1+\frac{\lambda_1(\mu-\mu_1)^2}{2}\right)^{-\alpha_1-\frac{1}{2}}\\
&= \frac{\beta_1^{-\frac{1}{2}}\lambda_1^{\frac{1}{2}}\Gamma(\alpha_1+\frac{1}{2})}{\Gamma(\alpha_1)(2\pi)^{\frac{1}{2}}}\left(1+\frac{\lambda_1(\mu-\mu_1)^2}{2\beta_1}\right)^{-\alpha_1-\frac{1}{2}}\\
&= \frac{\alpha_1^{\frac{1}{2}}\beta_1^{-\frac{1}{2}}\lambda_1^{\frac{1}{2}}\Gamma(\frac{2\alpha_1+1}{2})}{\Gamma(\frac{2\alpha_1}{2})(2\alpha_1\pi)^{\frac{1}{2}}}\left(1+\frac{\lambda_1\alpha_1(\mu-\mu_1)^2}{2\alpha_1\beta_1}\right)^{-\frac{2\alpha_1-1}{2}}
\end{aligned}$$

Fazendo $\nu=2\alpha_1$, $c_2=\left(\dfrac{\lambda_1\alpha_1}{\beta_1}\right)^{\frac{1}{2}}$ e $y=c_2(\mu-\mu_1)$, temos:

$$f_\mu(\mu|\vec{X})=c_2\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\nu\pi}}\left(1+\frac{y^2}{2}\right)^{-\frac{\nu+1}{2}},$$

que é c_2 vezes a densidade da distribuição t de Student[4] com $\nu=2\alpha_1$ graus de liberdade.
A distribuição marginal de μ nada mais é do que uma t_ν escalonada por c_2^{-1} e deslocada por μ_1. Isso vem por manipulação direta de y. Com isso, sabendo que $E(y)=0$ para $\nu>1$ e $Var(y)=\dfrac{\nu}{\nu-2}$ para $\nu>2$, temos:

$$E(\mu) = c_2^{-1}E(y) + \mu_1 = \mu_1, \ para \ \alpha_1 > \frac{1}{2}$$
$$Var(\mu) = c_2^{-2}Var(y) = \frac{\beta_1}{\lambda_1\alpha_1}\frac{\alpha_1}{\alpha_1 - 1} = \frac{\beta_1}{\lambda_1(\alpha_1 - 1)} \ para \ \alpha_1 > 1$$

Todos esses resultados podem ser derivados de forma análoga para a distribuição marginal à priori de μ, bastando trocar $\mu_1, \lambda_1, \alpha_1, \beta_1$ respectivamente por $\mu_0, \lambda_0, \alpha_0, \beta_0$.
Verbalizando melhor, temos:

Teorema 2 (Marginal à priori de μ) *A distribuição de* $\left(\frac{\lambda_0\alpha_0}{\beta_0}\right)^{\frac{1}{2}}(\mu - \mu_0)$ *é* t *de Student com* $2\alpha_0$ *graus de liberdade e:*

$$E(\mu) = \mu_0, \ para \ \alpha_1 > \frac{1}{2}$$
$$Var(\mu) = \frac{\beta_0}{\lambda_0(\alpha_0 - 1)} \ para \ \alpha_1 > 1$$

4 O Dilema da Palmirinha

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 μ e τ 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

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

4.1 Posterioris

Sabendo que $\mu|\vec{X} \sim \left(\frac{\lambda_1\alpha_1}{\beta_1}\right)^{-\frac{1}{2}} \text{t}_{2\alpha_1} + \mu_1$ e que, pelo Teorema (1), temos

$$\mu_1 = \frac{\lambda_0\mu_0 + n\bar{x}_n}{\lambda_0 + n} \qquad \lambda_1 = \lambda_0 + n$$
$$\alpha_1 = \alpha_0 + \frac{n}{2} \qquad \beta_1 = \beta_0 + \frac{1}{2}s_n^2 + \frac{n\lambda_0(\bar{x}_n - \mu_0)^2}{2(\lambda_0 + n)},$$

, dada uma atribuição de valores para os hiperparâmetros $\mu_0, \lambda_0, \alpha_0, \beta_0$, a distribuição à posteriori de *mu* fica definida, pois tais valores mais as estatísticas suficientes calculadas por Valciclei ($\bar{x}_n$ e s_n^2) definem de forma fechada $\mu_1, \lambda_1, \alpha_1, \beta_1$ e consequentemente, a distribuição $\mu|\vec{X}$.
Analogamente, com tais informações, poderíamos gerar a distribuição de $\tau|\vec{X}$, pois esta é Gama(α_1, β_1) que é está definida com as informações de Valciclei e os hiperparâmetros.

4.2 Intervalos de Credibilidade

Queremos ajudar a Palmirinha a encontrar $a, b \in \mathbb{R}$, $a < b$ de modo que $\Pr(\mu \in (a,b) \mid \vec{X}) = 0.95$. De modo geral, sabendo que $\mu|\vec{X} \sim \left(\frac{\lambda_1\alpha_1}{\beta_1}\right)^{-\frac{1}{2}} \text{t}_{2\alpha_1} + \mu_1$, se quisermos $a < b$ na qual $\Pr(\mu \in (a,b) \mid \vec{X}) = \gamma$, com $0 < \gamma < 1$, é conveniente tratar o problema de forma simétrica já que estamos falando da distribuição t, sendo assim, tomemos c, de tal forma que $\Pr(-c < \mathrm{t}_{2\alpha_1} < c) = \gamma$; Se $T_{2\alpha_1}$ é a função de densidade acumulada da $\mathrm{t}_{2\alpha_1}$, então temos:

$$\begin{aligned}
\Pr(-c < \mathrm{t}_{2\alpha_1} < c) = \gamma &\Leftrightarrow T_{2\alpha_1}(c) - T_{2\alpha_1}(-c) = \gamma \\
&\Leftrightarrow T_{2\alpha_1}(c) - (1 - T_{2\alpha_1}(c)) = \gamma \\
&\Leftrightarrow 2T_{2\alpha_1}(c) - 1 = \gamma \\
&\Leftrightarrow T_{2\alpha_1}(c) = \frac{1+\gamma}{2} \\
&\Leftrightarrow c = T_{2\alpha_1}^{-1}\left(\frac{1+\gamma}{2}\right),
\end{aligned}$$

onde a última ezpressão é calculada computacionalmente ou por consulta em tabelas.
Sabendo que $\left(\left(\frac{\lambda_1\alpha_1}{\beta_1}\right)^{\frac{1}{2}}(\mu - \mu_1)\right) \sim \mathrm{t}_{2\alpha_1}$, podemos definir a e b via:

$$\begin{aligned}
&\Pr\left[-c < \left(\frac{\lambda_1\alpha_1}{\beta_1}\right)^{\frac{1}{2}}(\mu - \mu_1) < c\right] = \gamma \\
\Leftrightarrow &\Pr\left[-c\left(\frac{\lambda_1\alpha_1}{\beta_1}\right)^{-\frac{1}{2}} < \mu - \mu_1 < c\left(\frac{\lambda_1\alpha_1}{\beta_1}\right)^{-\frac{1}{2}}\right] = \gamma \\
\Leftrightarrow &\Pr\left[-c\left(\frac{\lambda_1\alpha_1}{\beta_1}\right)^{-\frac{1}{2}} + \mu_1 < \mu < c\left(\frac{\lambda_1\alpha_1}{\beta_1}\right)^{-\frac{1}{2}} + \mu_1\right] = \gamma
\end{aligned}$$

Assim, podemos fazer:

$$\begin{aligned}
a &= -c\left(\frac{\lambda_1\alpha_1}{\beta_1}\right)^{-\frac{1}{2}} + \mu_1 \\
b &= c\left(\frac{\lambda_1\alpha_1}{\beta_1}\right)^{-\frac{1}{2}} + \mu_1, \\
\text{onde } c &= T_{2\alpha_1}^{-1}\left(\frac{1+\gamma}{2}\right)
\end{aligned}$$

Por construção, $a < b$ e $P(a < \mu < b|\vec{X}) = \gamma$
Para elicitação dos hiperparâmetros à priori, comecemos com a distribuição marginal de μ, a qual palmirinha acredita ter média 10 e variância 4. Podemos então igualar os momentos de tal distribuição descritos no Teorema (2), e chegar a um sistema de equações que vai resolver nossa situação pra μ_0.

$$\begin{aligned}
E(\mu) &= \mu_0 = 10 \\
Var(\mu) &= \frac{\beta_0}{\lambda_0(\alpha_0 - 1)} = 4
\end{aligned}$$

Já definimos então $\mu_0 = 10$. Para os outros parâmetros vamos trabalhar com τ. A unica informação que temos sobre ele é o coeficiente de variação igual a 1/2. O cv é uma função de τ, e podemos invertê-la de forma a fazer τ como função do cv=1/2:

$$\begin{aligned}
\mathrm{cv} = \frac{\sigma}{\mu} &\Rightarrow \mathrm{cv} = \frac{1/\sqrt{\tau}}{\mu} \\
&\Rightarrow \sqrt{\tau} = \frac{1}{\mathrm{cv} \cdot \mu} \\
&\Rightarrow \tau = \left(\frac{2}{\mu}\right)^2
\end{aligned}$$

Uma coisa razoável a se fazer, para não nos preocuparmos com a distribuição induzida da expressão acima, é considerar μ como um valor fixo razoável. No caso, usaremos $E(\mu) = 10$, que é a informação da Palmirinha. Sendo assim, o valor esperado $E(\tau) = \frac{4}{100} = \frac{1}{25}$. Usaremos tal valor para igualar ao momento do parâmetro de precisão. Como τ tem distribuição Gamma(α_0, β_0), temos:

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

Vamos agora fixar um valor razoável para α_0, que apesar de ser arbitrário, não pode ser muito grande pois estpa ligado com a quantidade de informação que temos. na interpretação das pseudo-observações, α_0 está ligado à quantidade de pseudo-experimentos $(2\alpha_0)$, que no nosso caso está representando a experiência da Palmirinha, que apesar de ser vasta na área de confecção de pamonhas, vamos adotar α_0 baixo, para dar mais flexibilidade para a posteriori. Para que os momentos de μ estajam bem definidos é preciso que α_0 seja maior que 1, usaremos o próximo inteiro 2, para manter a integridade da interpretação de pseudo-experimentos.
Sendo assim, $\beta_0 = 25\alpha_0 = 50$
Na interpretação que demos sobre β_0 ele era metade da soma dos erros quadráticos à priori. Como temos $2 * \alpha_0/2 = 4$ pseudo-observação, a soma quadrática mencionada nada mais é do que a variância amostral[3] desses experimentos vezes a quantidade de experimentos, ou seja $25 \cdot 4 = 100$, que é consistente com o que foi dito.
Tendo α_0, β_0, encontramos λ_0 facilmente pela expressão que já tínhamos da variância de μ:

$$\begin{aligned} Var(\mu) = \frac{\beta_0}{\lambda_0(\alpha_0 - 1)} = 4 &\Rightarrow \frac{50}{\lambda_0} = 4 \\ &\Rightarrow \lambda_0 = 50/4 = 12.5 \end{aligned}$$

Tendo entao $(\mu_0, \lambda_0, \alpha_0, \beta_0) = (10, 12.5, 2, 50)$, usando os dados de Valciclei no Teorema 1 obtemos, $(\mu_1, \lambda_1, \alpha_1, \beta_1) \approx (9.25, 22.5, 7, 61.92)$, o cálculo foi feito com o código abaixo escrito em Python 3.7, que além de fazer isso, calcula o intervalo de credibilidade com $\gamma = 0.95$ como descrito anteriormente.
Segue o script:

```python
#Hiperparametros a priori
mu0 = 10
alpha0 = 2
beta0 = 25*alpha0
lambda0 = beta0/(4*(alpha0-1))
print(f"mu0 = {mu0}\nlambda0 = {lambda0}\nalpha0 = {alpha0}\nbeta0 = {beta0}\n\n")

#Estatisticas Suficientes
xn = 8.307849
sn = 7.930452
n = 10

#Hiperparametros a posteriori
mu1 = (lambda0*mu0+n*xn)/(lambda0+n)
lambda1 = lambda0+n
alpha1 = alpha0+n/2
beta1 = beta0 + sn/2 + (n*lambda0*(xn-mu0)**2)/(2*(lambda0+n))
print(f"mu1 = {mu1}\nlambda1={lambda1}\nalpha1 = {alpha1}\nbeta1 = {beta1}\n\n")

#Intervalo de credibilidade
gamma = 0.95
from scipy.stats import t
c = t.ppf((1+gamma)/2,df=2*alpha1)
c2 = (lambda1*alpha1/beta1)**(-1/2)
a = -c*c2+mu1
b = c*c2+mu1

print(f"P({a} < mu < {b}) = {gamma}")
```

A saída no console referente ao intervalo de credibilidade foi:

```
P(7.903138300680498 < mu < 10.592727477097279) = 0.95
```

Portanto, dadas as condicões sobre os hiperparâmetros à priori e as estatísticas de Valciclei, com $a \approx 7.9$ e $b \approx 10.59$, temos $P(a < \mu < b|\vec{X}) \approx 0.95$

[3]Usamos 25, que é o a variância induzida pelo cv.

Referências

[1] M.H. DeGroot and M.J. Schervish. *Probability and Statistics, 4th ed.*, pages 495–505. Addison-Wesley, 2012.

[2] Wikipedia. Fubini's theorem — Wikipedia, the free encyclopedia. http://en.wikipedia.org/w/index.php?title=Fubini's%20theorem&oldid=975839217, 2020.

[3] Wikipedia. Normal-gamma distribution, Interpretation of parameters — Wikipedia, the free encyclopedia. https://en.wikipedia.org/wiki/Normal-gamma_distribution?oldformat=true#Interpretation_of_parameters, 2020.

[4] Wikipedia. Student's t-distribution — Wikipedia, the free encyclopedia. http://en.wikipedia.org/w/index.php?title=Student's%20t-distribution&oldid=981752965, 2020.