Inferência para a média de uma distribuição Normal
com variância desconhecida: O modelo Normal-NormalGama

Definição do Problema

Considere como antes $X_1, \ldots, X_n \,|\, \mu, \sigma^2 \overset{iid}{\sim} \mathrm{N}(\mu,\sigma^2)$, mas agora tanto a média μ quanto a variância σ^2 são desconhecidas, i.é. o parâmetro $\theta = (\mu, \sigma^2)$ tem dimensão 2. Usualmente o foco da inferência é o parâmetro μ e σ^2 é chamado um parâmetro *nuissance* ou *de estorvo*.

Como vimos na Seção anterior, a notação fica mais simples quando o problema é parametrizado em termos da precisão $\phi = \sigma^{-2}$, de forma que pensamos numa amostra $X_1, \ldots, X_n \,|\, \mu, \phi \overset{iid}{\sim} \mathrm{N}(\mu,\phi^{-1})$.

É natural que especificar uma distribuição a priori bivariada $p(\mu, \phi)$ é mais difícil que especificar uma univariada, como foi o caso de $p(\theta)$ na Seção **??** ou de $p(\mu)$ na Seção 1.1. Uma opção conveniente é partir o problema em "*pedazos*" univariados, especificando sucessivamente $p(\mu \,|\, \phi)$ e $p(\phi)$ (essa ideia de "*dividir para conquistar*" é muito útil em modelagem e, de fato, vamos usar ela extensivamente para abordar o problema de *modelos hierárquicos* nas próximas unidades).

Calculando a condicional \mu | \phi

Começamos assim pela distribuição a priori condicional de $\mu \,|\, \phi$. Quando condicionamos no valor de ϕ (ou, equivalentemente, no valor de $\sigma^2 = \phi^{-1}$), o problema passa a ser o que estudamos na Seção anterior. Dessa forma, sabemos que a distribuição conjugada $p(\mu, \phi)$ deve satisfazer que a condicional de $\mu \,|\, \phi$ deve ser Normal e também que a distribuição a posteriori condicional de $\mu \,|\, x_1, \ldots, x_n; \phi$ é Normal com média e variância dadas pelas equações (2) e (3).

Suponha que parametrizamos a distribuição a priori condicional de forma que $\mu \,|\, \phi \sim \text{Normal}(\mu_0,(\lambda_0\, \phi)^{-1})$. Substituindo σ^2 por ϕ^{-1} e τ^2 por $(\lambda_0\, \phi)^{-1}$ nas equações (2) e (3), o comentário anterior implica que, a posteriori, $\mu \,|\, \phi, x_1, \ldots, x_n \sim \text{Normal}(\mu^*,(\lambda^*\, \phi)^{-1})$, onde

$$\mu^* = \frac{\lambda_0\, \mu_0 + n\, \bar{x}}{\lambda_0 + n} \quad \lambda^* = \lambda_0 + n$$

Dessa forma, resolvemos o problema da distribuição a priori condicional de $\mu \,|\, \phi$ e também o de calcular a posteriori condicional de $\mu \,|\, \phi, x_1, \ldots, x_n$.

Calculando a marginal de \phi

Resta especificar a distribuição a priori marginal de ϕ e calcular a distribuição a posteriori também marginal de $\phi \,|\, x_1, \ldots, x_n$. Por enquanto, considere o caso com uma priori $p(\phi)$ geral.

Lembre primeiro que $[\bar{X} = \sum_{i=1}^n X_i, S^2 = \sum_{i=1}^n (X_i - \bar{X})^2]$ é um estatístico suficiente e que, dado o par (μ, ϕ), $\bar{X}$ e S^2 são independentes com $\bar{X} \,|\, \mu, \phi \sim$ Normal$[\mu,(n\,\phi)^{-1}]$ e $\phi\, S^2 \,|\, \mu, \phi \sim \chi^2_{n-1} \equiv$ Gama$(\frac{n-1}{2}, \frac{1}{2})$. Logo. a verossimilhança é

$$
\begin{aligned}
p(x_1, \ldots, x_n \,|\, \mu, \phi) &\propto p(\bar{x}, S^2 \,|\, \mu, \phi) \propto p(\bar{x} \,|\, \mu, \phi)\, p(S^2 \,|\, \mu, \phi) \\
&\propto \phi^{1/2} \exp\left\{-\frac{n\,\phi}{2}\,(\bar{x} - \mu)^2\right\} \times \phi^{(n-1)/2} \exp\left\{-\frac{\phi}{2}\, s^2\right\} \\
&\propto \phi^{n/2} \exp\left\{-\frac{\phi}{2}\,[s^2 + (\mu - \bar{x})^2]\right\} .
\end{aligned}
$$

Chegando à conjunta de \mu e \phi

Por outra parte, como especificamos que $\mu \,|\, \phi \sim \text{Normal}(\mu_0,(\lambda_0\,\phi)^{-1})$, segue que a densidade a priori conjunta é

$$p(\mu,\phi) = p(\phi)\,p(\mu\,|\,\phi) \propto p(\phi)\,\phi^{1/2}\exp\left\{-\frac{\phi}{2}\,\lambda_0\,(\mu-\mu_0)^2\right\}\,. \tag{9}$$

$$\begin{aligned} p(\mu,\phi\,|\,x_1,\ldots,x_n) &\propto p(\mu,\phi)\,p(x_1,\ldots,x_n\,|\,\mu,\phi) \\ &\propto p(\phi)\,\phi^{(n+1)/2}\,\exp\left\{-\frac{\phi}{2}\left[s^2+(\mu-\bar{x})^2+\lambda_0\,(\mu-\mu_0)^2\right]\right\}\,. \end{aligned} \tag{10}$$

Finalmente, usando o fato que $\mu\,|\,\phi, x_1,\ldots,x_n \sim \text{Normal}(\mu^*,(\lambda^*\,\phi)^{-1})$,

A posteriori marginal de \phi

$$p(\phi \,|\, x_1, \ldots, x_n) = \frac{p(\mu, \phi \,|\, x_1, \ldots, x_n)}{p(\mu \,|\, \phi, x_1, \ldots, x_n)}$$

$$\propto p(\phi)\, \frac{\phi^{(n+1)/2} \, \exp\left\{-\frac{\phi}{2}\left[s^2 + (\mu - \bar{x})^2 + \lambda_0 \,(\mu - \mu_0)^2\right]\right\}}{\phi^{1/2} \, \exp\left\{-\frac{\phi}{2}\, \lambda^* \,(\mu - \mu^*)^2\right\}}$$

$$\propto p(\phi)\, \phi^{n/2} \, \exp\left\{-\frac{\phi}{2}\left[s^2 + (\mu - \bar{x})^2 + \lambda_0 \,(\mu - \mu_0)^2 - \lambda^* \,(\mu - \mu^*)^2\right]\right\}$$

Veja o núcleo da Gama ->

$$\propto p(\phi)\, \left[\phi^{n/2} \, \exp\left\{-\frac{\phi}{2}\left[s^2 + \frac{n\,\lambda_0}{\lambda_0 + n}\,(\bar{x} - \mu_0)^2\right]\right\}\right] \quad (11)$$

(para o cálculo acima, da equação é útil observar que o termo da esquerda depende somente de ϕ, enquanto o da direita depende também de μ. A equação 11 é portanto uma identidade que vale para todo μ. Em termos práticos, na direita é possível substituir μ por qualquer valor e o resultado final deveria ser o mesmo. Fazer a substituição $\mu = \bar{x}$ ou $\mu = \mu_0$ facilita o cálculo).

Quais as regras de atualização?

Proposição 1. *Suponha que* $X_1, \ldots, X_n \,|\, \mu, \phi \overset{iid}{\sim} \mathrm{Normal}(\mu, \phi^{-1})$ *e, a priori,* $\mu \,|\, \phi \sim Normal(\mu_0, (\lambda_0\, \phi)^{-1})$ *e* $\phi \sim \mathrm{Gama}\ (\alpha_0, \beta_0)$. *Então, a posteriori,* $\mu \,|\, \phi, x_1, \ldots, x_n \sim$

$\mathrm{Normal}(\mu^*, (\lambda^*\, \phi)^{-1})$ *e* $\phi \,|\, x_1, \ldots, x_n \sim \mathrm{Gama}(\alpha^*, \beta^*)$, *onde* μ^* *e* λ^* *são dados nas equações* (6) *e* (7) *e*

$$\alpha^* = \alpha_0 + \frac{n}{2}\,, \tag{12}$$

$$\beta^* = \beta_0 + \frac{1}{2}\, s^2 + \frac{n\, \lambda_0}{2\,(\lambda_0 + n)}\, (\bar{x} - \mu_0)^2\,. \tag{13}$$

Prova da regra de atualização

Demonstração. Como foi explicado acima das equações (6) e (7), a parte referente à $p(\mu \,|\, \phi, x_1, \ldots, x_n)$ é consequência da Seção 1.1. Para achar a distribuição a posteriori marginal de ϕ, substitua $p(\phi) \propto \phi^{\alpha-1} e^{-\beta_0 \phi}$ na equação (11) para obter que

$$p(\phi \,|\, x_1, \ldots, x_n) \propto \phi^{\alpha_0 + n/2 - 1} \exp\left\{ -\phi \left[\beta_0 + \frac{1}{2} s^2 + \frac{n \lambda_0 (\bar{x} - \mu_0)^2}{2(\lambda_0 + n)} \right] \right\}, \qquad (14)$$

o que implica que $\phi \,|\, x_1, \ldots, x_n \sim \text{Gamma}(\alpha^*, \beta^*)$ $\square$

Posteriori Marginal de \mu

Como mencionamos no inicio desta Seção, usualmente o interesse é no parâmetro μ. Nesse caso, precisamos da distribuição a posteriori marginal $p(\mu \,|\, x_1, \ldots, x_n)$.

Segue da proposição 1 que, dado $x_1, \ldots, x_n$, $Z = (\lambda^* \phi)^{1/2} (\mu - \mu^*)$ segue uma distribuição Normal Padrão. Considere então a distribuição a posteriori conjunta de (Z, ϕ). O Jacobiano da transformação de $(Z, \phi) \mapsto (\mu = \mu^* + (\lambda^* \phi)^{-1/2} Z; \phi)$ é $(\lambda^* \phi)^{-1/2} \propto \phi^{-1/2}$. Logo, a densidade a posteriori de (Z, ϕ) é

$$
\begin{aligned}
p(Z, \phi \,|\, x_1, \ldots, x_n) &\propto p(\mu, \phi \,|\, x_1, \ldots, x_n) \left| \frac{\partial\,(\mu, \phi)}{\partial\,(Z, \phi)} \right| \\
&\propto p(\mu \,|\, \phi, x_1, \ldots, x_n)\, p(\phi, x_1, \ldots, x_n)\, \phi^{-1/2} \\
&\propto \phi^{1/2} \exp\left\{ -\frac{\lambda^* \phi}{2} (\mu - \mu^*)^2 \right\} \times \phi^{\alpha^* - 1} \exp\{-\beta^* \phi\} \times \phi^{-1/2} \\
&\propto \exp\left\{ -\frac{1}{2} z^2 \right\} \times \phi^{\alpha^* - 1} \exp\{-\beta^* \phi\} \,.
\end{aligned}
$$

Posteriori Marginal de \mu

Logo, a posteriori, Z e ϕ são independentes com $Z \,|\, x_1, \ldots, x_n \sim \mathrm{N}(0{,}1)$ e, como já sabíamos, $\phi \,|\, x_1, \ldots, x_n \sim \text{Gama}(\alpha^*{,}\beta^*)$. Das propriedades da distribuição Gama sabemos que isso implica que $2\,\beta^*\,\phi \,|\, x_1, \ldots, x_n \sim \text{Gama}(\alpha^*{,}1/2)$, também denominada de distribuição χ^2 com $(2\,\alpha^*)$ graus de liberdade. Da definição da distribuição t de Student como a razão entre duas variáveis aleatórias independentes, uma com distribuição Normal(0,1) e a outra sendo a raiz quadrada de uma χ^2_m dividida pelos graus de liberdade m, segue que

$$\sqrt{\frac{\lambda^*\,\alpha^*}{\beta^*}}\,(\mu - \mu^*) = \frac{Z}{\sqrt{2\,\beta^*\,\phi/(2\,\alpha^*)}} \,\Big|\, x_1, \ldots, x_n \sim \; t_{2\,\alpha^*}\,.$$

Note que, como a distribuição a priori a a posteriori pertencem a mesma família Normal-Gama, um resultado semelhante deve valer também para a distribuição a priori marginal de μ. É conveniente escrever o resultado para uso futuro (veja, por exemplo, a subseção 1.2.3).

Resumindo num lema

Lema 2. *Suponha que* $\mu \,|\, \phi \sim$ *Normal*$(\mu_0,(\lambda_0\, \phi)^{-1})$ *e* $\phi \sim$ Gama (α_0,β_0), *onde* α, β, λ_0 *e* ϕ *são parâmetros positivos. Então*

$$\sqrt{\frac{\lambda_0\, \alpha}{\beta}}\, (\mu - \mu_0) \sim \; t_{2\,\alpha}\,. \tag{16}$$

Tornando a priori não informativa

O intervalo de intervalo de confiança clássico para μ é

$$\bar{x} \pm t_{n-1;a/2} \frac{s}{\sqrt{n\,(n-1)}}\,, \tag{17}$$

S é só a soma, sem
Dividir por n-1.

Com base na distribuição (15), o equivalente bayesiano desse intervalo de confiança seria

$$\mu^* \pm t_{2\,\alpha^*;a/2} \sqrt{\frac{\beta^*}{\lambda^*\,\alpha^*}}\,, \tag{18}$$

Quais valores dos hiperparâmetros?

Tem valores de $(\mu_0, \lambda_0, \alpha_0, \beta_0)$ para quais (17) e (18) são iguais?

$2\,\alpha^* = 2\,\alpha_0 + n$ tería que ser igual a $(n-1)$, isto é $\alpha_0 = -(1/2)$

$\mu^* = \bar{x}$, e da equação (2) obtemos então $\lambda_0 = 0$ Assim o valor de \mu_0 não importa

$$\frac{s}{\sqrt{n\,(n-1)}} = \sqrt{\frac{\beta^*}{\lambda^*\,\alpha^*}} = \sqrt{\frac{\beta_0 + \frac{1}{2}\,s^2 + \frac{n\,\lambda_0}{2\,(\lambda_0+n)}\,(\bar{x}-\mu_0)^2}{(\lambda_0+n)\,(\alpha_0+n/2)}}$$

$\alpha_0 = -(1/2)$ e $\lambda_0 = 0$ dá $\beta_0 = 0$ É chamada de priori matching, nesse caso não informativa.

Um problema, porém, é que nem a distribuição Normal está bem definida para variância infinita (i.é. precisão $\lambda_0\,\phi = 0$), nem a distribuição Gama está definida quando $\alpha_0 = -(1/2) \le 0$ e/ou $\beta_0 = 0 \le 0$. Nesse sentido, veja que da equação (9) temos que

Substituindo os valores

$$p(\mu,\phi) \propto \phi^{\alpha_0-1}\,\exp\{-\beta_0\,\phi\} \times \phi^{1/2}\exp\left\{-\frac{\phi}{2}\,\lambda_0\,(\mu-\mu_0)^2\right\}$$
$$\propto \phi^{\alpha_0-1/2}\,\exp\left\{-\frac{\phi}{2}\,\lambda_0\,(\mu-\mu_0)^2-\beta_0\,\phi\right\}$$

Leva a $$p(\mu,\phi) \propto \frac{1}{\phi}\,,$$ Que é imprópria.

Veja que pode ser pensada como $p(\mu,\phi) \propto p(\mu) \times p(\phi) \propto 1 \times \phi^{-1}$

isto é, a priori μ e ϕ seriam independentes com $p(\mu) \propto 1$ e $p(\phi) \propto \phi^{-1}$.

Distribuição Preditiva

O problema do cálculo de distribuições preditivas no caso desta Seção pode tomar a seguinte forma, semelhante à que foi discutida na Seção 1.1. Considere $X_1, \ldots, X_n, X_{n+1}, \ldots, X_{n+m} \mid \mu, \phi \overset{iid}{\sim} \text{Normal}(\mu,(\cancel{\lambda_0}\, \phi)^{-1})$ com a distribuição a priori $\mu \mid \phi \sim \text{Normal}(\mu_0,(\lambda_0\, \phi)^{-1})$ e $\phi \sim$ Gama (α_0,β_0). Com base na observação de $x_1, \ldots, x_n$ desejamos prever $\bar{X}_{(n+1):(n+m)} = m^{-1} \sum_{i=n+1}^{m+n} X_i$ ou, em outras palavras, procuramos a distribuição condicional de $\bar{X}_{(n+1):(n+m)}$ dado $x_1, \ldots, x_n$.

Da subseção 1.1.2 sabemos que $\bar{X}_{(n+1):(n+m)} \mid x_1, \ldots, x_n; \phi \sim \text{Normal}[\mu^*,(\lambda^*\, \phi)^{-1}]$. Por outro lado, da subseção **??**, sabemos também que $\phi \mid x_1, \ldots, x_n \sim \text{Gama}(\alpha^*,\beta^*)$.

Portanto, se olharmos o par $(\bar{X}_{(n+1):(n+m)}, \phi)$ condicionado a $x_1, \ldots, x_n$, temos uma distribuição Normal-Gama e o Lema 2 implica que

$$\sqrt{\frac{\left(\frac{1}{m} + \frac{1}{\lambda^*}\right)\ \alpha^*}{\beta^*}}\,(\bar{X}_{(n+1):(n+m)} - \mu^*) \mid x_1, \ldots, x_n \sim\ t_{2\,\alpha^*}\,.$$

Previsão para a velocidade da luz

Exemplo 2 (Continuação). Como a distribuição a posteriori de μ [equação (15)] é simétrica com respeito a μ^*, segue que a estimativa bayesiana com respeito a PQ, PA ou perda zero-um é $\mu^* = 27.75$.