Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Estatística

Dissertação de Mestrado

O Modelo Hierárquico Poisson-Gama sob Distribuições *a priori* Não-Informativas

por

Fernando Oscar Schmitt

Orientador: Prof. Gustavo Leonel Gilardoni Avalle

Julho de 2013

Fernando Oscar Schmitt

O Modelo Hierárquico Poisson-Gama sob Distribuições *a priori* Não-Informativas

Dissertação apresentada ao Departamento de Estatística do Instituto de Ciências Exatas da Universidadede de Brasília como requisito parcial à obtenção do título de Mestre em Estatística.

Universidade de Brasília
Brasília, julho de 2013

À Alice Maria.

Resumo

Na presente dissertação são apresentados, como contribuições originais, uma família de densidades *a priori* para o modelo hierárquico Poisson-Gama, bem como resultados acerca de condições necessárias e suficientes para a propriedade das correspondentes densidades *a posteriori* e para a existência de momentos das variáveis latentes.

A família introduzida inclui como casos especiais as densidades de Jeffreys e outras utilizadas na literatura. A questão da convergência ou divergência das distribuições *a posteriori* associadas é respondida completamente.

As densidades *a priori* de Jeffreys são formuladas em termos de funções hipergeométricas generalizadas, e é apresentado um código em "R" para simulação de amostras da distribuição *a posteriori* por meio do algoritmo Metropolis Adaptivo.

Palavras Chave: *abordagem bayesiana, densidade a priori de Jeffreys, propriedade da distribuição a posteriori, existência de momentos a posteriori, distribuição binomial negativa.*

Abstract

The present study proposes a new family of prior distributions for the Poisson-Gamma hierachical model, as well as new results regarding necessary and sufficient conditions for propriety of the corresponding posteriors and finiteness of moments of latent variables.

The proposed family includes Jeffreys' densities and others commonly used in the literature. The conditions for propriety or impropriety of the corresponding posteriors are treated in full detail.

Jeffreys' priors are expressed in terms of generalized hypergeometric functions, and code in "R" is presented for drawing samples from the posterior distribution using the Adaptive Metropolis algorithm.

Keywords: *bayesian methods, Jeffreys' prior, posterior propriety, finiteness of posterior moments, negative binomial distribution.*

Conteúdo

1 Introdução

Na presente dissertação, estudamos o modelo hierárquico onde cada variável observável possui uma distribuição de Poisson cujas médias seguem uma distribuição Gama. Esse modelo é conhecido na literatura como Poisson-Gama. Para maior generalidade, admite-se a presença de covariáveis (fixas) como fatores que permitem reescalonar cada observação de forma diferenciada.

Como exemplo de aplicação desse modelo, podemos citar inicialmente um conjunto de observações X_j indexadas por $j \in \{1, .., n\}$ que representam o número de ocorrências de um evento em uma unidade de tempo, de acordo com um processo de Poisson homogêneo. O tempo total em que houve observação da contagem de ocorrências pode ser diferente para cada j, e é denotado por t_j. Alem disso, o valor esperado do número de ocorrências em cada unidade de tempo também pode variar em j, e é denotado por λ_j. Por fim, os valores λ_j são modelados como realizações independentes e identicamente distribuídas de uma variável aleatória Gama.

A utilização de diferentes valores esperados para cada observação tem a vantagem de permitir a inclusão de *sobredispersão* no modelo. De acordo com a definição acima, a distribuição de X_j é Poisson com parâmetro $\lambda_j t_j$. Assim, é possível tratar dados onde, por exemplo, o número de ocorrências por unidade de tempo é grande mas com pouca variabilidade entre as diferentes observações. O mesmo não seria possivel se λ_j fosse considerado fixo.

Como o modelo é visto sob o ponto de vista bayesiano, estudaremos distribuições *a priori* para os parâmetros da distribuição Gama, denotados por α e β (este utilizado como parâmetro de escala). Além disso, analisaremos o modelo tanto sob a ótica em que os parâmetros λ_j são de interesse para as inferências, como sob a ótica em que são apenas intermediários na especificação do modelo.

O interesse recai principalmente sobre densidades *a priori* impróprias, uma vez que (i) há uma grande disponibilidade de pacotes que implementam algoritmos de amostragem do tipo MCMC (*Markov Chain Monte Carlo*) e (ii) já foi mostrado na literatura [12] que esses algoritmos podem resultar em amostras aparentemente corretas mesmo quando utilizados com formas funcionais que não definem uma densidade *a posteriori* própria.

A contribuição original do presente trabalho consiste na apresentação de uma família de densidades *a priori* para o modelo em questão (18), que inclui as densidades de Jeffreys [11] e outras formas funcionais utilizadas na literatura. Para essa família, é possivel estabelecer condições necessárias e suficientes de convergência da densidade *a posteriori* correspondente (Teorema 14).

Como corolário, mostramos que a distribuição *a posteriori* associada à densidade *a priori* de Jeffreys é própria sempre que existir pelo menos uma observação não nula (Corolário 15), e que a densidade *a posteriori* correspondente a distribuição de Jeffreys de independência é *sempre* imprópria, quaisquer que sejam os dados observados (Corolário 16).

Ainda em decorrência do Teorema obtido, apresentamos condições de existência dos momentos das variáveis latentes ($\mathrm{E}[\lambda_j^w]$ e $\mathrm{E}[\tilde{\lambda}^w]$), para $w > 0$, em função dos parâmetros da densidade *a priori* e dos dados observados (Teoremas 21 e 22).

Por fim, destacamos que a formulação da densidade *a priori* de Jeffreys em termos de funções hipergeométricas generalizadas permite obter amostras aproximadamente independentes da correspondente densidade *a posteriori* por meio de um algoritmo MCMC. Uma implementação dessa amostragem é apresentada em anexo, em linguagem "R", com base no algoritmo Metropolis Adaptivo (Vihola [13]).

No restante desta seção, apresentaremos a notação utilizada e os conceitos básicos necessários para o estudo do modelo em questão, nas duas acepções mencionadas. Na seção seguinte será tratado o conceito de distribuição *a priori não-informativa*, com ênfase para as metodologias de Jeffreys. Na terceira seção a densidades *a priori* de Jeffreys para o modelo Poisson-Gama serão obtidas de forma explícita. Na quarta seção apresentaremos uma forma funcional de uma família de densidades *a priori* para esse modelo, que engloba as metodologias de Jeffreys e outras densidades utilizadas na literatura. Na quinta seção são apresentadas situações em que é possível estabelecer condições necessárias e suficientes para a integrabilidade da distribuição *a posteriori* do modelo Poisson-Gama quando a densidade *a priori* pertence à família citada. Na sexta seção é tratada a questão da existência de momentos *a posteriori* de ordem arbitrária para os parâmetros latentes e para uma variável preditiva, identicamente distribuida e condicionalmente independente.

1.1 Notação

Será utilizada, por abuso de notação, sempre a letra p para indicar densidades (ou funções) de probabilidade, e a variável a qual se aplica será indicada pelo argumento entre parênteses. Por exemplo, $p(\theta)$ indica a densidade (marginal) de θ (se este for contínuo), e $p(x|\theta)$ indica a densidade (ou função) de probabilidade da variável $X|\theta$. Sempre que for necessário maior clareza quanto às diferentes densidades, o abuso de notação não será utilizado.

O foco da inferência bayesiana consiste na obtenção da distribuição de θ condicionada aos valores observados de X, denominada distribuição *a posteriori*, que pode ser expressa (a menos de uma constante de normalização dependente apenas de X) nos termos da regra de Bayes:

$$p(\theta|x) \propto p(x|\theta)p(\theta).$$

A distribuição *a posteriori* condensa toda a informação que se pode atribuir ao parâmetro não-observável pela combinação do conhecimento *a priori* sobre θ e da verossimilhança destes em função dos dados observados. A partir dela é possível obter, por exemplo, no caso em que $\theta \in \mathbb{R}$, números LI e LS que satisfazem desigualdades da forma

$$Prob\,[LI \leq \theta \leq LS|X = x] = \int_{LI}^{LS} p(\theta|x)\,d\theta \geq 1 - \alpha,$$

similares aos *intervalos de confiança* do paradigma frequentista, mas com interpretação consideravelmente distinta.

Os modelos bayesianos são denotados na seguinte forma:

$$\begin{array}{ccc} X|\theta & \sim & p(x|\theta) \\ \theta & \sim & p(\theta) \end{array} \tag{1}$$

O modelo hierárquico bayesiano consiste na relação entre quantidades aleatórias em que a dependência da parte observável X em relação aos parâmetros se dá por meio de uma variável ou vetor latente λ, e é denotado da seguinte forma:

$$\begin{array}{ccc} X|\lambda & \sim & p(x|\lambda) \\ \lambda|\phi & \sim & p(\lambda|\phi) \\ \phi & \sim & p(\phi). \end{array} \tag{2}$$

O modelo hierárquico pode ser reduzido ao caso simples (1) de duas formas. A primeira consiste em integrar a variável latente na distribuição conjunta:

$$X|\phi \sim p(x|\phi) = \int_\Lambda p(x|\lambda)p(\lambda|\phi)\, d\lambda,$$
$$\phi \sim p(\phi).$$

Nesse caso, a distribuição *a posteriori* é dada por

$$p(\phi|x) \propto p(x|\phi)p(\phi) \propto p(\phi) \int_\Lambda p(x|\lambda)p(\lambda|\phi)\, d\lambda,$$

e o modelo recai no caso simples tomando $\theta = \phi$. A segunda forma consiste em considerar a distribuição de $\lambda|\phi$ como parte da ditribuição *a priori*, e tomar como parâmetro o par $\theta = (\lambda, \phi)$:

$$X|\lambda, \phi \sim p(x|\lambda, \phi) = p(x|\lambda),$$

$$(\lambda, \phi) \sim p(\lambda, \phi) = p(\lambda|\phi)p(\phi).$$

No segundo caso, a distribuição *a posteriori* conjunta é dada por $p(\lambda, \phi|x) \propto p(x|\lambda)p(\lambda, \phi) = p(x|\lambda)p(\lambda|\phi)p(\phi)$, que pode ser marginalizada, para se obter a distribuição *a posteriori* de ϕ:

$$p(\phi|x) \propto p(\phi) \int_\Lambda p(x|\lambda)p(\lambda|\phi)\, d\lambda,$$

ou do parâmetro latente:

$$p(\lambda|x) \propto p(x|\lambda) \int_\Phi p(\lambda|\phi)p(\phi)\, d\phi = p(x|\lambda)p(\lambda).$$

1.2 Modelos com densidade *a priori* imprópria

Como uma extensão do modelo (1), consideraremos situações em que $p(\theta)$ é *imprópria*, isto é, tal que

$$\int_{\Theta} p(\theta)\, d\theta = \infty,$$

e buscaremos determinar se (ou sob quais condições) a densidade *a posteriori* é própria.

A integrabilidade da densidade *a posteriori* é condição fundamental para a aplicabilidade do modelo a dados reais, com vistas a extrair inferências sobre os parâmetros. Se a densidade *a posteriori* é imprópria, não é possivel fazer inferências sobre o parâmetro θ *e nem mesmo sobre as variáveis latentes* λ_j. Por outro lado, se a densidade *a posteriori* for integrável, então também será própria a densidade *a posteriori* das variáveis latentes. Isso é demostrado no seguinte resultado:

Teorema 1. *Para o modelo hierárquico* (2), *associado a uma densidade* $p(\phi)$ *imprópria, podemos afirmar que* $p(\lambda|X)$ *é própria* se e somente se $p(\phi|X)$ *o for.*

Demonstração. Observe que $p(\lambda|X, \phi)$ é própria por definição do modelo, uma vez que pode ser obtida de

$$p(\lambda|X, \phi) = \frac{p(X|\lambda)p(\lambda|\phi)}{p(X|\phi)},$$

que definem variáveis aleatórias com distribuições próprias. A integrabilidade de $p(\lambda|X)$ é dada pela expressão

$$\int_{\Lambda} p(\lambda|X)\, d\lambda = \int_{\Lambda}\int_{\Phi} p(\lambda|X, \phi)p(\phi|X)\, d\phi d\lambda = \int_{\Phi} p(\phi|X)\, d\phi.$$

Da igualdade acima observa-se que a primeira integral é finita se e somente se a última também for. ■

A determinação da integrabilidade de $p(\phi|X)$, portanto, é condição primordial para a utilização de densidades *a priori* impróprias.

1.3 O modelo Poisson-Gama

Estudaremos o modelo bayesiano hierárquico Poisson-Gama na formulação de Hadjicostas e Berry [12]:

$$\begin{array}{rcl} X_j|\lambda_j & \sim & p(x_j|\lambda_j), \\ \lambda_j|\alpha,\beta & \sim & p(\lambda_j|\alpha,\beta), \end{array} \tag{3}$$

em que

$$\begin{aligned} p(x_j|\lambda_j) &= (\lambda_j t_j)^{x_j} e^{-\lambda_j t_j}/x_j!, \\ p(\lambda_j|\alpha,\beta) &= \frac{1}{\beta^\alpha \Gamma(\alpha)} \lambda_j^{\alpha-1} e^{-\lambda_j/\beta}, \end{aligned}$$

onde $t_1, ..., t_n$ são covariáveis fixas e α e β são hiperparâmetros de forma e escala da distribuição gama.

Integrando o vetor de parâmetros latentes obtemos a expressão:

$$p(\boldsymbol{x}|\alpha,\beta) = \prod_{j=1}^{n} \left(\frac{t_j^{x_j}}{\beta^\alpha \Gamma(\alpha) x_j!} \right) \int_0^\infty \cdots \int_0^\infty \prod_{j=1}^{n} \lambda_j^{\alpha+x_j-1} e^{-(1/\beta+t_j)\lambda_j}\, d\boldsymbol{\lambda},$$

que é fatorável em:

$$p(\boldsymbol{x}|\alpha,\beta) = \prod_{j=1}^{n} \frac{t_j^{x_j}}{\beta^\alpha \Gamma(\alpha) x_j!} \frac{\Gamma(\alpha+x_j)}{(1/\beta+t_j)^{\alpha+x_j}}.$$

Segue, portanto, que os X_j são condicionalmente independentes, dados α e β, com função de probabilidade dada por:

$$p(x_j|\alpha,\beta) = \frac{\Gamma(\alpha+x_j)}{\Gamma(\alpha) x_j!} \frac{(\beta t_j)^{x_j}}{(1+\beta t_j)^{\alpha+x_j}}.$$

Reparametrizando a função de probabilidade acima em termos de α e $r = \beta/(1+\beta)$ e definindo:

$$r_j = \frac{\beta t_j}{1+\beta t_j} = \frac{t_j r}{1+(t_j-1)r}, \tag{4}$$

a função de probabilidade assume a forma

$$p(x_j|\alpha, r) = \binom{\alpha + x_j - 1}{x_j} r_j^{x_j}(1 - r_j)^{\alpha}, \tag{5}$$

e identifica-se com a distribuição binomial negativa (generalizada para o caso de ambos os parâmetros poderem assumir valores reais - também chamada de distribuição de Pólya). Utilizaremos a parametrização acima definida em todo o restante dessa dissertação.

1.4 Conceitos básicos

Para estabelecer a convergência (ou a divergência) das integrais relativas às densidades estudadas, utilizaremos o conceito de ordem (também chamada de equivalência assintótica) e de “O-grande” (*big-O*), que é menos restritiva:

Definição 1. *Sejam f e g funções definidas em uma vizinhança de x_0. Diremos que $f(x)$ é de ordem $g(x)$ (ou $f(x) \sim g(x)$) em $x \to x_0$ se*

$$\lim_{x \to x_0} \frac{f(x)}{g(x)} = K,$$

para alguma constante positiva K.

Definição 2. *Sejam f e g funções definidas em uma vizinhança de x_0. Diremos que $f(x) \in \mathcal{O}(g(x))$, $x \to x_0$ se $|f(x)/g(x)|$ é limitada em alguma vizinhança de x_0.*

Ambas as definições acima podem ser estendidas para o caso $x_0 = \infty$, tomando como vizinhança de x_0, nesse caso, qualquer intervalo da forma (a, ∞). Uma terceira definição que será utilizada é a de expansão assintótica, conforme a seguir:

Definição 3. *Seja f uma função definida em uma vizinhança de ∞. Dizemos que f tem expansão assintótica $f(x) \simeq \sum_{k=0}^{\infty} a_k x^{-k}$ em $x \to \infty$ se, para todo $n \in \mathbb{N}$, vale a seguinte afirmativa:*

$$f(x) - \sum_{k=0}^{n-1} a_k x^{-k} \in \mathcal{O}(x^{-n}),\ x \to \infty$$

Nota: não é necessário que a expansão assintótica defina uma série convergente. Apenas suas somas parciais são consideradas na definição. Por isso, o simbolo $\simeq$ é utilizado ao invés da igualdade.

Seguem alguns resultados básicos que serão utilizados ao longo do estudo:

Teorema 2. *Dada uma função f contínua em $(0,\infty)$ tal que*

$$f(x) \sim \begin{cases} x^a & \text{em} \quad x \to 0 \quad \text{e} \\ x^b & \text{em} \quad x \to \infty, \end{cases}$$

então a integral $\int_0^\infty f(x)\,dx$ converge se $a > -1$ e $b < -1$, e diverge se $a \leq -1$ ou $b \geq -1$.

Corolário 3. *A integral $\int_0^1 x^a(1-x)^b\,dx$ converge se e somente se $a, b > -1$.*

A demonstração dos resultados acima pode ser obtida facilmente na literatura de análise matemática introdutória.

Lema 4. *Seja f com expansão assintótica $f(x) \simeq \sum_{k=0}^\infty a_k x^{-k}$ em $x \to \infty$. Então*

$$\lim_{x\to\infty} f(x) = a_0.$$

Demonstração. Tomando $n = 1$ na Definição 3, temos $f(x) - a_0 \in \mathcal{O}(x^{-1}),\ x \to \infty$. Da Definição 2, segue que

$$\left|\frac{f(x) - a_0}{x^{-1}}\right| < K,$$

para algum $K > 0$ e para todo x suficientemente grande. Segue então

$$|f(x) - a_0| < \frac{K}{|x|} \to 0.$$

Logo $f(x) \to a_0$. ∎

Lema 5. *Seja f definida como no Lema 4. Então*

$$g(x) = x(f(x) - a_0) \simeq \sum_{k=0}^\infty b_k x^{-k} \text{ em } x \to \infty,$$

onde $b_k = a_{k+1}$.

Demonstração. Seja $n \in \mathbb{N}$. Podemos escrever

$$\left|\frac{g(x) - \sum_{k=0}^{n-1} b_k x^{-k}}{x^{-n}}\right| = \left|\frac{xf(x) - xa_0 - \sum_{k=0}^{n-1} b_k x^{-k}}{x^{-n}}\right| =$$

$$= \left|\frac{f(x) - a_0 - \sum_{k=0}^{n-1} b_k x^{-k-1}}{x^{-n-1}}\right| = \left|\frac{f(x) - \sum_{k=0}^{n} a_k x^{-k}}{x^{-(n+1)}}\right|.$$

Como $f(x) - \sum_{k=0}^{n} a_k x^{-k} \in \mathcal{O}(x^{-(n+1)})$ em $x \to \infty$, podemos tomar x_0 tal que o valor absoluto acima é limitado, $\forall x > x_0$. Isso implica que $g(x) - \sum_{k=0}^{n-1} b_k x^{-k} \in \mathcal{O}(x^{-n})$ em $x \to \infty$. ■

2 Distribuições *a priori* não-informativas

A crítica tradicional ao paradigma bayesiano consiste na dificuldade de se definir uma distribuição *a priori* que possa ser considerada objetiva, ou que represente um estado de nenhum conhecimento sobre θ anterior à observação dos dados. Esse problema já aparece em 1774, quando Laplace (segundo Fienberg [5]) apresenta o *princípio da indiferença*, ou a hoje chamada regra de sucessão de Laplace. De acordo com esse princípio, a distribuição que representa desconhecimento total sobre θ é aquela que associa iguais probabilidades (ou densidades) a quaisquer valores possíveis, ou seja, a distribuição uniforme. No entanto, se o conjunto de valores possíveis para θ tiver medida infinita, essa regra não é capaz de fornecer uma densidade normalizável para θ, pois

$$\int_\Theta p(\theta)d\theta = \int_\Theta c d\theta = \infty.$$

Mesmo assim, dependendo da forma da função de verossimilhança, é possível que a relação $p(\theta|x) \propto p(x|\theta)$ defina uma distribuição *a posteriori própria* (isto é, normalizável) para θ. Portanto, é válido analisar as chamadas distribuições *a priori impróprias* (não-normalizáveis), desde que a distribuição *a posteriori* seja própria. Pode ocorrer ainda que esta seja integrável apenas para alguns dos valores possíveis de X, embora a situação ideal seja aquela em que sua aplicabilidade não dependa dos dados observados.

Retornando à regra de sucessão de Laplace, considere uma variável $X \sim Bin(n, \theta)$ com distribuição binomial em que n é fixo e conhecido e θ segue uma distribuição *a priori* uniforme no intervalo $[0, 1]$:

$$p(x|\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x},$$

$$p(\theta) = \begin{cases} 1 & \text{se } 0 \leq \theta \leq 1, \\ 0 & \text{caso contrário.} \end{cases}$$

Nesse modelo, a distribuição de θ dado X é $p(\theta|x) \propto \theta^x(1-\theta)^{n-x}$, ou seja, $\theta|X = x$ tem distribuição Beta com parâmetros $\alpha = x + 1$ e $\beta = n - x + 1$. Tomando a média dessa distribuição $\left(\frac{\alpha}{\alpha+\beta} = \frac{x+1}{n+2}\right)$, obtém-se a regra de sucessão de Laplace, que consiste informalmente em "adicionar um sucesso e uma falha à amostra", e é uma decorrência de seu princípio de indiferença.

Revendo esse modelo, Haldane [8] afirma que a suposição de uniformidade introduz viés (no sentido frequentista) nas inferências, e que um estimador bayesiano objetivo de θ deveria ser $\frac{x}{n}$. Para chegar a este resultado, ele propõe utilizar como densidade *a priori* a expressão:

$$p(\theta) \propto \frac{1}{\theta(1-\theta)}.$$

A densidade *a priori* de Haldane, como pode ser chamada, não é integrável no intervalo $[0, 1]$, de sorte que se trata de uma densidade imprópria. Por outro lado, a distribuição *a posteriori* é dada por $p(\theta|x) \propto \theta^{x-1}(1-\theta)^{n-x-1}$, que somente é própria se $x \neq 0$ e $x \neq n$. Nesse caso, $\theta|x$ segue uma distribuição Beta com parâmetros $\alpha = x$ e $\beta = n - x$. Essa densidade, no entanto, não se aplica a todos os possíveis valores que X pode assumir.

2.1 ***Probability matching*** no modelo normal

Um dos critérios para avaliar a objetividade de uma distribuição *a priori* consiste na comparação entre as inferências que são obtidas a partir da correspondente distribuição *a posteriori* e as que seriam calculadas sob o paradigma frequentista. Quando a probabilidade *a posteriori* de certas regiões do espaço paramétrico (intervalos por exemplo) coincide com a sua probabilidade de cobertura, do ponto de vista clássico (Datta e Sweeting [3]), dizemos que a distribuição *a priori* associada ao modelo é *Probability Matching Prior*.

Para exemplificar essa propriedade, considere uma classe de densidades *a priori* definida por $p(\mu, \sigma) = 1/\sigma^\lambda$, $\lambda > 0$, associada a uma amostra *iid* de tamanho n do modelo normal, parametrizado pela média e desvio-padrão.

A função de verossimilhança é dada por:

$$p(\mathbf{x}|\mu, \sigma) = (2\pi)^{-n/2}\sigma^{-n} \exp\left[\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right].$$

A densidade *a posteriori* pode, portanto, ser escrita como

$$p(\mu, \sigma|\mathbf{x}) \propto \sigma^{-n-\lambda} \exp\left\{\frac{-A}{2\sigma^2}\right\},$$

onde $A = \sum_{i=1}^{n}(x_i - \mu)^2 = (n-1)s^2 + n(\bar{x} - \mu)^2$, e $\bar{x}$ e s^2 são definidos por:

$$\bar{x} = \sum_{i=1}^{n} x_i \quad \text{e} \quad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

A marginal $p(\mu|\mathbf{x})$ pode ser obtida pela transformação $z = 1/2\sigma^2$:

$$p(\mu|\mathbf{x}) \propto \int_0^\infty \sigma^{-n-\lambda} \exp\left\{\frac{-A}{2\sigma^2}\right\} d\sigma \propto \int_0^\infty z^{\frac{n+\lambda}{2}} e^{-Az} z^{-3/2}\, dz.$$

Definindo $\alpha = (n + \lambda - 1)/2$ e notando que o integrando é uma densidade Gama, tem-se

$$p(\mu|\mathbf{x}) \propto A^{-\alpha} \propto \left[1 + \frac{n(\bar{x} - \mu)^2}{(n-1)s^2}\right]^{-\frac{n+\lambda-1}{2}} = \left[1 + \frac{1}{\nu}\left(\sqrt{\frac{\nu}{n-1}}\frac{\mu - \bar{x}}{s/\sqrt{n}}\right)^2\right]^{-\frac{\nu+1}{2}},$$

onde $\nu = n + \lambda - 2$.

Por fim, considere a variável $t = \frac{\mu - \bar{x}}{s/\sqrt{n}}$. Decorre do exposto que $\sqrt{\frac{\nu}{n-1}}t \Big| \boldsymbol{x}$ tem distribuição *t de Student* com ν graus de liberdade, onde $\nu = n + \lambda - 2$. Assim, no caso $\lambda = 1$ (e somente nesse caso), o resultado se reduz a $t|\boldsymbol{x} \sim t_{n-1}$, o que coincide com a distribuição da quantidade pivotal t, quando o modelo é analisado sob o paradigma frequentista.

2.2 A densidade *a priori* de Jeffreys

Jeffreys [11] apresenta outra questão acerca da escolha da distribuição *a priori*, que chamaremos de *coerência*, ou *invariância*: considere um modelo associado a uma reparametrização dada por $\varphi : \Theta \to E$. Se a distribuição *a priori* representa desconhecimento sobre θ, deve também representá-lo sobre $\eta = \varphi(\theta)$. Por isso, é desejável que a forma de obter a densidade seja equivalente, qualquer que seja a parametrização do modelo.

Para apresentar essa propriedade mais formalmente, deixaremos de utilizar o abuso de notação introduzido na seção inicial, passando a adotar uma letra diferente para cada densidade ou função de probabilidade. Após a definição e um exemplo, o abuso de notação será retomado.

Considere um modelo $X|\theta \sim f(x|\theta)$ e uma bijeção $\eta = \varphi(\theta)$. Sejam $\pi_1(\theta)$ e $\pi_2(\eta)$ respectivamente as distribuições *a priori* dos modelos original e reparametrizado, sendo este definido por $X|\eta \sim g(x|\eta) = f(x|\varphi^{-1}(\eta))$. Diremos que existe coerência se a distribuição *a posteriori* de θ for a mesma quando calculada por qualquer uma das seguintes alternativas:

1. diretamente por meio da fórmula de Bayes: $p_1(\theta|x) \propto f(x|\theta)\pi_1(\theta)$
2. a partir da distribuição *a posteriori* do modelo transformado $p_2(\eta|x) \propto g(x|\eta)\pi_2(\eta)$, considerando θ como uma função de η ($\theta = \varphi^{-1}(\eta)$).

Na segunda alternativa, a densidade de $\theta|X$ é dada por

$$q(\theta|x) = p_2(\varphi(\theta)|x)|\varphi'(\theta)|$$

(vide James [10]).

A distribuição uniforme para o modelo binomial sugerida pelo princípio da indiferença *não* satisfaz essa propriedade. Por exemplo, tomando a bijeção $\varphi : [0,1] \to [0,1]$, $\eta = \varphi(\theta) = \theta^a$, para algum valor $a > 0$ fixado, a verossimilhança pode ser reescrita como:

$$g(x|\eta) = f(x|\varphi^{-1}(\eta)) = \binom{n}{x}\eta^{x/a}(1-\eta^{1/a})^{n-x},$$

que, associada a uma distribuição uniforme em η, leva a densidade *a posteriori* dada por $p_2(\eta|x) \propto g(x|\eta)$. Ao realizar a transformação de volta para a variável θ, a densidade obtida é dada por:

$$q(\theta|x) = p_2(\varphi(\theta)|x)|\varphi'(\theta)| \propto \theta^x(1-\theta)^{n-x}\theta^{a-1},$$

que indica uma distribuição Beta com parâmetros $\alpha = x + a$ e $\beta = n - x + 1$. Verifica-se então que o princípio da indiferença, quando aplicado a diferentes potências do parâmetro θ da binomial, leva a diferentes "regras de sucessão", resultado de diferentes distribuições *a posteriori* para θ.

Retornando a Jeffreys (1946), o autor propõe uma metodologia para definição de uma densidade *a priori* que satisfaz a propriedade de coerência, embora nem sempre resulte em uma distribuição própria. A densidade *a priori* de Jeffreys é dada por

$$p(\theta) \propto \sqrt{\det \mathcal{I}(\theta)},$$

onde $\mathcal{I}(\theta)$ é a matriz de informação de Fisher, definida por

$$\mathcal{I}(\theta)_{ij} = \mathrm{E}\left[\left(\frac{\partial}{\partial\theta_i}\log p(x|\theta)\right)\left(\frac{\partial}{\partial\theta_j}\log p(x|\theta)\right)\right].$$

Ressalte-se que a metodologia de Jeffreys requer algumas condições de regularidade para validade da propriedade de coerência, e que não é a única que satisfaz essa propriedade (vide, por exemplo, Hartigan [9]). No Anexo I, trazemos a demonstração dessa propriedade para um caso razoavelmente regular.

No caso multiparamétrico, uma variação da proposta de Jeffreys consiste em tomar o produto das densidades *a priori* de cada parâmetro, calculada como se os demais fossem fixos. A densidade assim obtida será chamada de distribuição *a priori de independência*, uma vez que coincide com a metodologia original no caso em que a matriz de informação de Fisher tem forma diagonal.

Na seção seguinte veremos exemplos dessas duas alternativas. Convém ressaltar, no entanto, que as distribuições de independência não necessariamente satisfazem a propriedade de invariância definida anteriormente.

2.2.1 Densidade *a priori* de Jeffreys para modelos de locação e escala

Nesta seção serão considerados modelos uniparamétricos de locação e de escala, bem como o modelo biparamétrico onde ambas são desconhecidas, para fins de obtenção da densidade *a priori* de Jeffreys.

Para um modelo de locação, isto é, $p(x|\mu) = g(x-\mu)$, $\mu \in \mathbb{R}$, pode-se mostrar que a informação de Fisher para μ é constante, de forma que a densidade *a priori* de Jeffreys é imprópria e dada por $p(\mu) \propto 1$. Mesmo assim, a densidade *a posteriori* é sempre própria, e dada por $p(\mu|x) = g(x-\mu)$, pois

$$\int_{-\infty}^{+\infty} g(x-\mu)\, d\mu = -\int_{+\infty}^{-\infty} g(z)\, dz = 1.$$

No modelo de escala, $p(x|\sigma) = (1/\sigma)g(x/\sigma)$, $0 < \sigma < \infty$, a informação de Fisher tem a forma $\mathcal{I}(\sigma) = c/\sigma^2$, de forma que a densidade *a priori* de Jeffreys é dada por $p(\sigma) \propto 1/\sigma$, também imprópria. No entanto, a densidade *a posteriori* é normalizável *se* $x \neq 0$, e é dada por $p(\sigma|x) \propto g(x/\sigma)(1/\sigma^2)$. Para verificar essa afirmação, considere a transformação $z = x/\sigma$ na integral abaixo:

$$\int_0^\infty \frac{1}{\sigma} g\left(\frac{x}{\sigma}\right) \frac{1}{\sigma}\, d\sigma = \begin{cases} \frac{1}{x}\int_0^\infty g(z)\, dz & \text{se } x > 0, \\ \frac{-1}{x}\int_{-\infty}^0 g(z)\, dz & \text{se } x < 0. \end{cases}$$

De onde se conclui que $p(\sigma|x) = g(x/\sigma)(h(x)/\sigma^2)$, em que

$$h(x) = \begin{cases} |x|/\operatorname{Prob}[Z > 0] & \text{se } x > 0, \\ |x|/\operatorname{Prob}[Z < 0] & \text{se } x < 0, \end{cases}$$

e $Z \sim g(z)$.

No modelo completo de locação-escala, isto é, $p(x|\mu,\sigma) = \frac{1}{\sigma} g\left(\frac{x-\mu}{\sigma}\right)$, tem-se

$$\log p(x|\mu,\sigma) = -\log\sigma + \log g\left(\frac{x-\mu}{\sigma}\right).$$

Definindo $h = (\log \circ g)' = g'/g$, segue, pela regra da cadeia:

$$\frac{\partial \log p}{\partial \mu} = h\left(\frac{x-\mu}{\sigma}\right)\left(\frac{-1}{\sigma}\right),$$

$$\frac{\partial \log p}{\partial \sigma} = \frac{-1}{\sigma} + h\left(\frac{x-\mu}{\sigma}\right)\left(-\frac{x-\mu}{\sigma^2}\right).$$

Assim, os elementos da matriz de informação de Fisher,

$$\mathcal{I}(\mu, \sigma) = \begin{pmatrix} i_{\mu\mu} & i_{\mu\sigma} \\ i_{\mu\sigma} & i_{\sigma\sigma} \end{pmatrix},$$

são dados por:

$$i_{\mu\mu} = \mathrm{E}\left[\frac{1}{\sigma^2}\left(h\left(\frac{x-\mu}{\sigma}\right)\right)^2\right],$$

$$i_{\mu\sigma} = \mathrm{E}\left[h\left(\frac{x-\mu}{\sigma}\right)\frac{1}{\sigma}\left(\frac{1}{\sigma} + h\left(\frac{x-\mu}{\sigma}\right)\frac{x-\mu}{\sigma^2}\right)\right],$$

$$i_{\sigma\sigma} = \mathrm{E}\left[\left(\frac{1}{\sigma} + h\left(\frac{x-\mu}{\sigma}\right)\frac{x-\mu}{\sigma^2}\right)^2\right].$$

As variáveis $z = \frac{x-\mu}{\sigma}$ e suas funções são quantidades pivotais. Portanto, suas distribuições não dependem dos parâmetros. Segue que

$$i_{\mu\mu} = \frac{1}{\sigma^2}\,\mathrm{E}\left[(h(z))^2\right],$$

$$i_{\mu\sigma} = \mathrm{E}\left[h(z)\frac{1}{\sigma}\left(\frac{1}{\sigma} + h(z)z\frac{1}{\sigma}\right)\right] = \frac{1}{\sigma^2}\,\mathrm{E}\left[h(z)\left(1 + h(z)z\right),\right]$$

$$i_{\sigma\sigma} = \mathrm{E}\left[\left(\frac{1}{\sigma} + h(z)z\frac{1}{\sigma}\right)^2\right] = \frac{1}{\sigma^2}\,\mathrm{E}\left[(1 + h(z)z)^2\right].$$

Dessa forma, a matriz de informação é dada por:

$$\mathcal{I}(\theta) = \frac{1}{\sigma^2}\begin{pmatrix} c_{11} & c_{12} \\ c_{12} & c_{22} \end{pmatrix}.$$

onde c_{ij} não depende de θ, de forma que densidade *a priori* de Jeffreys nesse caso é $p(\theta) \propto 1/\sigma^2$.

Cumpre observar que o exemplo da seção 2.1 pode ser visto sob a ótica de um modelo de locação-escala, de forma que a densidade *a priori* de Jeffreys na versão completa é dada por $p(\mu, \sigma) \propto 1/\sigma^2$ e na versão de independência é dada por $p(\mu, \sigma) \propto 1/\sigma$. Conforme exposto naquela seção, a versão de independência é uma distribuição *probability matching*, mas a versão completa da densidade *a priori* de Jeffreys para o modelo normal *não* satisfaz essa propriedade.

2.2.2 O Princípio da Verossimilhança e a distribuição *a priori* de Jeffreys

No paradigma bayesiano, o *princípio da verossimilhança* consiste em tomar a função de verossimilhança como a especificação completa do modelo, sendo equivalentes duas funções que difiram apenas por uma constante em relação aos parâmetros.

Como um exemplo, considere os experimentos definidos a seguir:

1. Lançar uma moeda n vezes, e observar o número de vezes x em que o resultado foi 'cara'.
2. Lançar uma moeda até que sejam observadas α vezes o resultado 'coroa', e registrar o número total de lançamentos m.

No primeiro caso, x segue uma distribuição binomial de parâmetros n e θ, onde $\theta \in (0, 1)$ é desconhecido e representa a probabilidade de obter 'cara' em cada lançamento (desconsideramos o caso $\theta \in \{0, 1\}$ porque resultaria em probabilidade nula de obter uma das faces). No segundo caso, a variável $y = m - \alpha$ também representa o número de 'caras' observado mas segue uma distribuição binomial negativa de parâmetros α e θ.

As funções de verossimilhança nos dois casos acima são:

$$p(x|\theta) = \binom{n}{x} \theta^x (1 - \theta)^{n-x},$$

$$p(y|\theta) = \binom{m-1}{y} \theta^y (1 - \theta)^{m-y}.$$

Se os dois experimentos resultarem nas mesmas observações, isto é, se $x = y$ e $n = m$, então as funções são equivalentes, na ótica do princípio da verossimilhança. Veremos a seguir que nesse caso a densidade *a priori* de Jeffreys *não* respeita esse princípio, por levar a expressões diferentes para cada uma das situações acima.

No caso do modelo binomial, a densidade *a priori* de Jeffreys pode ser obtida por

$$\mathcal{I}(\theta) = \mathrm{E}\left[\left(\frac{d}{d\theta}\log p(x|\theta)\right)^2\right] = \mathrm{E}\left[\left(\frac{x}{\theta} - \frac{n-x}{1-\theta}\right)^2\right] = \mathrm{E}\left[\left(\frac{x-n\theta}{\theta(1-\theta)}\right)^2\right].$$

Como $\mathrm{E}[X] = n\theta$ e $\mathrm{Var}[X] = n\theta(1-\theta)$,

$$\mathcal{I}(\theta) = \left(\frac{1}{\theta(1-\theta)}\right)^2 \mathrm{Var}\left[X\right] = \frac{n}{\theta(1-\theta)},$$

que significa que a densidade *a priori* de Jeffreys é dada por $p(\theta) \propto \theta^{-1/2}(1-\theta)^{-1/2}$, ou seja, a distribuição Beta$(1/2, 1/2)$. A densidade *a posteriori* é sempre própria e dada por Beta$(x + 1/2, n - x + 1/2)$.

No caso da distribuição Binomial Negativa, a informação de Fisher é dada por:

$$\mathcal{I}(\theta) = \mathrm{E}\left[-\frac{d^2 \log p(y|\theta)}{d^2\theta}\right] = \mathrm{E}\left[\frac{y}{\theta^2} + \frac{m-y}{(1-\theta)^2}\right] = \frac{1}{\theta(1-\theta)^2}$$

pois $\mathrm{E}[y] = \alpha\theta/(1-\theta)$. Segue que a densidade *a priori* de Jeffreys tem a forma $p(\theta) \propto \theta^{-1/2}(1-\theta)^{-1}$. A distribuição *a posteriori* tem a forma Beta$(y + 1/2, m - y)$.

Observa-se que as inferências serão diferentes de acordo com o *design* do experimento que foi adotado, mesmo que os dados observados sejam idênticos. Isso ocorre porque a definição da densidade de Jeffreys faz uso do conceito de informação de Fisher, que por sua vez consiste em um valor esperado sobre o espaço amostral, isto é, depende não apenas dos dados observados, mas também dos que poderiam ter sido observados.

2.2.3 Invariância da distribuição *a priori* de Jeffreys

Nessa seção, não será utilizado nenhum abuso de notação, para evitar qualquer confusão entre os conceitos.

Para verificar que a distribuição *a priori* de Jeffreys satisfaz a propriedade de coerência, considere uma família de densidades (ou de funções de probabilidade) $f_1(x|\theta)$, com suporte independente de θ e derivadas parciais contínuas em relação a cada componente θ_i, e um difeomorfismo $\varphi : \Theta \to E$, entre os conjuntos abertos $\Theta, E \subset \mathbb{R}^d$. Valem, então, as seguintes expressões:

$$
\begin{array}{r|l}
f_1(x|\theta), & f_2(x|\eta) = f_1(x|\varphi^{-1}(\eta)), \\
\pi_1(\theta) \propto \sqrt{\det \mathcal{I}_1(\theta)}, & \pi_2(\eta) \propto \sqrt{\det \mathcal{I}_2(\eta)}, \\
p_1(\theta|x) \propto f_1(x|\theta)\pi_1(\theta). & p_2(\eta|x) \propto f_2(x|\eta)\pi_2(\eta).
\end{array}
$$

Teorema 6. *A distribuição a posteriori de θ é a mesma, tomando diretamente $p_1(\theta|x)$ ou partindo da densidade a posteriori para η e fazendo a mudança de variável $\theta = \varphi^{-1}(\eta)$.*

Demonstração. Inicialmente, provaremos a seguinte propriedade da matriz de informação de Fisher:

$$\mathcal{I}_2(\varphi(\theta)) = J_{\varphi^{-1}}(\varphi(\theta))^T \mathcal{I}_1(\theta)\, J_{\varphi^{-1}}(\varphi(\theta)). \tag{6}$$

Para tanto, observe que

$$\mathcal{I}_2(\eta)_{ij} = \mathrm{E}\left[\frac{\partial L_2(\eta; X)}{\partial \eta_i}\frac{\partial L_2(\eta; X)}{\partial \eta_j}\right], \qquad L_2(\eta; x) = \log f_2(x|\eta).$$

Pela regra da cadeia multidimensional,

$$\frac{\partial L_2}{\partial \eta_i}(\eta; x) = \frac{1}{f_2(x|\eta)}\frac{\partial f_2(x|\eta)}{\partial \eta_i} = \frac{1}{f_1(x|\varphi^{-1}(\eta))}\sum_{k=1}^{d}\left.\frac{\partial f_1(x|\theta)}{\partial \theta_k}\right|_{\theta=\varphi^{-1}(\eta)}\frac{\partial \theta_k}{\partial \eta_i}.$$

Segue, portanto, que:

$$\frac{\partial L_2}{\partial \eta_i}(\varphi(\theta); x) = \sum_{k=1}^{d}\frac{\partial L_1(\theta; x)}{\partial \theta_k}\left.\frac{\partial \theta_k}{\partial \eta_i}\right|_{\eta=\varphi(\theta)}, \qquad L_1(\theta; x) = \log f_1(x|\theta).$$

Substituindo esses resultados na expressão para os elementos da matriz de informação, tem-se

$$\mathcal{I}_2(\varphi(\theta))_{ij} = \mathrm{E}\left[\sum_{k=1}^{d}\sum_{l=1}^{d} C_{ki}A(X)_{kl}C_{lj}\right],$$

onde

$$A(x)_{kl} = \frac{\partial L_1(\theta;x)}{\partial \theta_k}\frac{\partial L_1(\theta;x)}{\partial \theta_l} \quad \text{e} \quad C_{ij} = \left.\frac{\partial \theta_i}{\partial \eta_j}\right|_{\eta=\varphi(\theta)}.$$

Para completar a prova, basta observar que

$$\sum_{k=1}^{d}\sum_{l=1}^{d} C_{ki}A(x)_{kl}C_{lj} = (C^T A(x)\, C)_{ij},$$

$$C = J_{\varphi^{-1}}(\varphi(\theta)) \quad \text{e} \quad \mathrm{E}\,[A(X)] = \mathcal{I}_1(\theta),$$

de forma que a propriedade (6) segue pela linearidade do valor esperado:

$$\mathcal{I}_2(\varphi(\theta)) = \mathrm{E}\left[C^T A(X)\, C\right] = C^T\,\mathrm{E}\,[A(X)]\, C.$$

Retomando o enunciado, a distribuição de uma função de uma variável aleatória é dada pela fórmula

$$q(\theta|x) = p_2(\varphi(\theta)|x)|\det J_\varphi(\theta)| \propto f_2(x|\varphi(\theta))\pi_2(\varphi(\theta))|\det J_\varphi(\theta)|,$$

segundo James [10], onde $J_\varphi(\theta)$ é a matriz jacobiana da transformação:

$$J_\varphi(\theta)_{ij} = \frac{\partial \eta_i}{\partial \theta_j}(\theta), \qquad i,j = 1,...,d.$$

Como $f_2(x|\varphi(\theta)) = f_1(x|\theta)$, basta provar que $\pi_1(\theta) = \pi_2(\varphi(\theta))|\det J_\varphi(\theta)|$, ou, equivalentemente,

$$\det \mathcal{I}_1(\theta) = \det \mathcal{I}_2(\varphi(\theta))\,(\det J_\varphi(\theta))^2\,.$$

Usando propriedades do determinante, o lado direito pode ser escrito como $\det\left(J_\varphi^T(\theta)\mathcal{I}_2(\varphi(\theta))J_\varphi(\theta)\right)$, de forma que a proposição decorre diretamente de (6), pois a matriz jacobiana da função inversa satisfaz $J_{\varphi^{-1}}(\varphi(\theta)) = (J_\varphi(\theta))^{-1}$. ■

3 A densidade *a priori* de Jeffreys no modelo Poisson-Gama

Nessa seção será exibida a expressão analítica da densidade *a priori* de Jeffreys na parametrização dada por α e $r = \beta/(1+\beta)$. Nesses termos, o modelo integrado se escreve como

$$X_j \sim p(x_j|\alpha, r) = \frac{\Gamma(\alpha + x_j)}{\Gamma(\alpha) x_j!} r_j^{x_j} (1 - r_j)^{\alpha},$$

onde

$$r_j = \frac{t_j \beta}{1 + t_j \beta} = \frac{t_j r}{1 + (t_j - 1) r},$$

e os valores t_j são considerados fixos e positivos.

A matriz de informação de Fisher $\begin{pmatrix} i_{\alpha\alpha} & i_{\alpha r} \\ i_{\alpha r} & i_{rr} \end{pmatrix}$ (para uma observação) tem como componentes:

$$i_{\alpha\alpha} = -\,\mathrm{E}\left[\frac{\partial^2}{\partial \alpha^2} \log p(X|\alpha, r)\right] = \mathrm{E}[\psi'(\alpha) - \psi'(\alpha + X)]$$

$$i_{\alpha r} = -\,\mathrm{E}\left[\frac{\partial^2}{\partial r^2} \log p(X|\alpha, r)\right] =$$

$$= \mathrm{E}\left[\left(\frac{X}{r} - \frac{\alpha t_j}{1 - r}\right) \frac{t_j - 1}{(1 + (t_j - 1) r)^2} + \left(\frac{X}{r^2} + \frac{\alpha t_j}{(1 - r)^2}\right) \frac{1}{1 + (t_j - 1) r}\right]$$

$$i_{rr} = -\,\mathrm{E}\left[\frac{\partial^2}{\partial r \partial \alpha} \log p(X|\alpha, r)\right] = \mathrm{E}\left[\frac{t_j}{(1 - r)(1 + (t_j - 1) r)}\right],$$

onde

$$\psi(x) = \frac{d}{dx} \log \Gamma(x)$$

é a função *digama*, cuja derivada ψ' é chamada *função trigama*. A primeira componente é dada por:

$$i_{\alpha\alpha} = \sum_{x=1}^{\infty} \frac{\Gamma(\alpha + x)}{\Gamma(\alpha) x!} r_j^x (1 - r_j)^{\alpha} (\psi'(\alpha) - \psi'(\alpha + x)).$$

Usando o fato de que $\psi'(\alpha) - \psi'(\alpha + x)$ pode ser escrito como $\sum_{n=0}^{x-1} (\alpha + n)^{-2}$, que segue de uma representação em série da função trigama [4], Fisher aponta em artigo de 1941 [6] que a série acima "curiosamente" se reduz a

$$i_{\alpha\alpha} = \sum_{x=1}^{\infty} \mathrm{B}(\alpha, x) \frac{r_j^x}{x}, \tag{7}$$

onde $\mathrm{B}(\alpha, x) = \Gamma(\alpha)\Gamma(x)/\Gamma(\alpha + x)$.

Assim como em [2], também não foi possível obter uma demonstração para o resultado acima. Tomando-o como válido, é possivel escrevê-lo em termos da função hipergeométrica generalizada:

$$i_{\alpha\alpha} = \frac{r_j}{\alpha}\ {}_3F_2\left(\begin{matrix}1,\, 1,\, 1\\ 2,\, \alpha+1\end{matrix}\, ; r_j\right), \tag{8}$$

onde

$${}_3F_2\left(\begin{matrix}a_1,\, a_2,\, a_3\\ b_1,\, b_2\end{matrix}\, ; z\right) = \sum_{k=0}^{\infty} \frac{(a_1)_k (a_2)_k (a_3)_k}{(b_1)_k (b_2)_k} \frac{z^k}{k!}. \tag{9}$$

e $(a)_k = \Gamma(a + k)/\Gamma(a)$ é o símbolo de Pochhammer.

A componente $i_{\alpha r}$ pode ser simplificada utilizando a linearidade do valor esperado e o fato de que

$$\mathrm{E}[X_j] = \alpha \frac{r_j}{1 - r_j} = \alpha t_j \frac{r}{1 - r},$$

resultando em

$$i_{\alpha r} = \alpha \frac{r_j}{r^2 (1 - r)^2}.$$

Por fim a componente i_{rr} pode ser escrita como:

$$i_{rr} = \frac{r_j}{r(1 - r)}.$$

Dessa forma, o determinante da matriz de informação para n observações, é dado por:

$$\left(\sum_{j=1}^{n} i_{\alpha\alpha}\right)\left(\sum_{j=1}^{n} i_{\alpha r}\right) - \left(\sum_{j=1}^{n} i_{rr}\right)^2,$$

e a densidade *a priori* de Jeffreys pode ser escrita como:

$$p_J(\alpha, r) \propto \frac{1}{r(1 - r)} \sqrt{R \sum_{j=1}^{n} r_j \left({}_3F_2\left(\begin{matrix}1,\, 1,\, 1\\ 2,\, \alpha+1\end{matrix}\, ; r_j\right) - 1\right)}, \tag{10}$$

onde $R = \sum_{j=1}^{n} r_j$.

Para a densidade de independência, tomamos

$$p_{JI}(\alpha, r) \propto \sqrt{\left(\sum_{j=1}^{n} i_{\alpha\alpha}\right)\left(\sum_{j=1}^{n} i_{\alpha r}\right)}$$

que resulta em

$$p_{JI}(\alpha, r) \propto \frac{1}{r(1-r)}\sqrt{R\sum_{j=1}^{n} r_j \; {}_3F_2\left(\begin{matrix}1,\,1,\,1\\2,\,\alpha+1\end{matrix};r_j\right)} \tag{11}$$

Na seção seguinte, veremos as densidades aqui expostas no contexto de uma família mais geral de densidades *a priori* para o modelo Poisson-Gama. No restante dessa seção veremos formulações equivalentes de (8) e propriedades assintóticas da função hipergeométrica generalizada que resulta de sua aplicação.

3.1 Comportamento assintótico da função ${}_3F_2$

Nesta seção adotaremos a notação simplificada

$$F(\alpha, r) = {}_3F_2\left(\begin{matrix}1,\,1,\,1\\2,\,\alpha+1\end{matrix};r\right).$$

Teorema 7. *A expressão $F(\alpha, r)$ satisfaz as seguintes desigualdades:*

$$1 \leq F(\alpha, r) \leq \alpha\psi'(\alpha), \tag{12}$$

para $r \in (0,1)$. Além disso, a expressão à direita apresenta o seguinte comportamento assintótico:

$$\alpha\psi'(\alpha) \sim \begin{cases}\alpha^{-1} & \text{em} \quad \alpha \to 0 \quad \text{e}\\ 1 & \text{em} \quad \alpha \to \infty.\end{cases}$$

Demonstração. Utilizando (7) e (8),

$$F(\alpha, r) = \frac{\alpha}{r} \sum_{x=1}^{\infty} \mathrm{B}(\alpha, x) \frac{r^x}{x} = 1 + \sum_{x=2}^{\infty} \alpha \, \mathrm{B}(\alpha, x) \frac{r^{x-1}}{x}, \tag{13}$$

uma vez que $\mathrm{B}(\alpha, 1) = 1/\alpha$.

Como os coeficientes da série acima são todos positivos, fica demonstada a desigualdade à esquerda de (12). Para demonstrar a desigualdade à direita, observe, de (13), que a função $F(\alpha, r)$ é estritamente crescente em $r \in (0, 1)$, de forma que basta mostrar

$$\lim_{r \to 1} F(\alpha, r) = \alpha \psi'(\alpha).$$

A série hipergeométrica generalizada (9) de $F(\alpha, r)$ converge absolutamente para todo $r \in [0, 1]$ (vide [4, seção 16.2(iii)]). Portanto, o limite acima é simplesmente o valor de $F(\alpha, 1)$. Aplicando as transformações em [4, 15.4.1 e 16.5.2], obtemos

$$F(\alpha, 1) = \frac{\Gamma(\alpha + 1)}{\Gamma(\alpha)} \int_0^1 (1 - t)^{\alpha - 1} \frac{-\log(1 - t)}{t} \, dt. \tag{14}$$

Por outro lado, tomando a representação integral da função trigama [1, 6.4.1],

$$\psi'(z) = \int_0^{\infty} \frac{u e^{-zu}}{1 - e^{-u}} \, du,$$

e fazendo a mudança de variável para [$t = 1 - e^{-u}$, obtemos

$$\psi'(z) = \int_0^1 \frac{-\log(1 - t)(1 - t)^{z-1}}{t} \, dt.$$

Substituindo esta expressão em (14), completamos a prova das desigualdades (12) $F(\alpha, 1) = \alpha \psi'(\alpha)$.

Para obter o comportamento assintótico do termo limitante, aplicamos inicialmente a série dada em [4, 5.15.1]:

$$\psi'(\alpha) = \sum_{n=0}^{\infty} \left(\frac{1}{\alpha + n} \right)^2 = \frac{1}{\alpha^2} + \sum_{n=1}^{\infty} \left(\frac{1}{\alpha + n} \right)^2 \leq \frac{1}{\alpha^2} + \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{1}{\alpha^2} + \frac{\pi^2}{6}. \tag{15}$$

Segue, portanto, que $\alpha^2\psi'(\alpha) \to 1$, de forma que $\alpha\psi'(\alpha) \sim \alpha^{-1}$ em $\alpha \to 0$.

Alem disso, a função trigama satisfaz a expansão assintótica

$$\psi'(\alpha) \simeq \frac{1}{\alpha} + \frac{1}{2\alpha^2} + \sum_{k=1}^{\infty} \frac{B_{2k}}{\alpha^{2k+1}}, \tag{16}$$

onde B_{2k} são os números de Bernoulli [4, 5.15.18]. Seguindo a notação da Definição 3, verifica-se que $a_0 = 0$. Aplicando o Lema 4, podemos escrever

$$\alpha\psi'(\alpha) \simeq 1 + \frac{1}{2\alpha} + \sum_{k=1}^{\infty} \frac{B_{2k}}{\alpha^{2k}}. \tag{17}$$

Finalmente, pelo Lema 5, concluímos $\lim_{\alpha\to\infty} \alpha\psi'(\alpha) = 1$, pois $a_1 = 1$. ■

O resultado acima permitirá tratar a distribuição *a posteriori* associada à densidade *a priori* de Jeffreys de independência (11). Para tratar o caso da densidade de Jeffreys (10) na formulação original, é necessário estudar a expressão $F(\alpha, r) - 1$, que aparece no seu radicando. Ocorre que essa expressão não admite cota inferior *positiva* em $(0, 1)$, pois $F(\alpha, r) \to 1$ quando $r \to 0$. Estudaremos, portanto a expressão

$$G(\alpha, r) = \frac{F(\alpha, r) - 1}{r}.$$

Teorema 8. *A expressão $G(\alpha, r)$ definida acima satisfaz as seguintes desigualdades:*

$$\frac{1}{2 + 2\alpha} \leq G(\alpha, r) \leq \alpha\psi'(\alpha) - 1,$$

em $r \in (0, 1)$. Além disso, vale $\alpha\psi'(\alpha) - 1 \sim \alpha^{-1}$ tanto em $\alpha \to 0$ quanto em $\alpha \to \infty$.

Demonstração. Aplicando a representação (13), podemos escrever

$$G(\alpha, r) = \frac{1}{r}\left(\frac{\alpha}{r}\sum_{x=1}^{\infty} \mathrm{B}(\alpha, x)\frac{r^x}{x} - 1\right) = \frac{1}{r}\left(\frac{\alpha}{r}\sum_{x=2}^{\infty} \mathrm{B}(\alpha, x)\frac{r^x}{x}\right).$$

Segue que

$$G(\alpha, r) = \alpha\left(\frac{\mathrm{B}(\alpha, 2)}{2} + \sum_{x=3}^{\infty} \mathrm{B}(\alpha, x)\frac{r^{x-2}}{x}\right).$$

Como os coeficientes da série são positivos, podemos escrever

$$G(\alpha, r) \geq \alpha\left(\frac{\mathrm{B}(\alpha, 2)}{2}\right) = \frac{1}{2 + 2\alpha}.$$

Para verificar o limite em $r \to 1$, aplica-se o resultado do Teorema 7:

$$\lim_{r\to 1} G(\alpha, r) = F(\alpha, 1) - 1 = \alpha\psi'(\alpha) - 1.$$

Para obter o comportamento assintótico em $\alpha \to 0$, observamos que

$$\alpha(\alpha\psi'(\alpha) - 1) = \alpha^2\psi'(\alpha) - \alpha \to 1,$$

uma vez que $\lim_{\alpha\to 0} \alpha^2\psi'(\alpha) = 1$.

Por outro lado, para analisar o comportamento em $\alpha \to \infty$ aplicamos o Lema 5 à expressão (17), obtendo

$$\alpha(\alpha\psi'(\alpha) - 1) \simeq \frac{1}{2} + \sum_{k=1}^{\infty} \frac{B_{2k}}{\alpha^{2k-1}}.$$

Aplicando o Lema 4 à expansão obtida acima, segue

$$\lim_{\alpha\to\infty} \alpha(\alpha\psi'(\alpha) - 1) = \frac{1}{2}.$$

Conclui-se, portanto, que $\alpha\psi'(\alpha) - 1 \sim \alpha^{-1}$ tanto em $\alpha \to 0$ quanto em $\alpha \to \infty$. ■

4 Uma família de densidades *a priori*

Na presente dissertação, trataremos de uma família de densidades *a priori* que engloba algumas formas funcionais comumente utilizadas bem como as densidades obtidas pela técnica proposta por Jeffreys e pela sua variação que resulta na densidade *a priori* de independência.

Retomando a parametrização definida em (5), vamos considerar densidades *a priori* para um modelo

$$X_j|\alpha, r \sim p(x_j|\alpha, r),$$

nos parâmetros $\alpha \in (0, \infty)$ e $r \in (0, 1)$. A família de densidades que será estudada possui a seguinte forma funcional:

$$p(\alpha, r) \propto r^a (1-r)^b f(\alpha, r), \tag{18}$$

onde a e b são hiperparâmetros reais e f é uma função contínua, positiva no domínio $(0, \infty) \times (0, 1)$, que satisfaz as seguintes condições:

1. f admite cotas inferior e superior positivas, $f_I(\alpha)$ e $f_S(\alpha)$, que dependem somente de α;

2. O comportamento assintótico das cotas pode ser expresso como

$$f_I(\alpha) \sim \begin{cases} \alpha^{u+\varepsilon} & \text{em} \quad \alpha \to 0 \quad \text{e} \\ \alpha^{v-\delta} & \text{em} \quad \alpha \to \infty, \end{cases} \tag{19}$$

   e

$$f_S(\alpha) \sim \begin{cases} \alpha^{u} & \text{em} \quad \alpha \to 0 \quad \text{e} \\ \alpha^{v} & \text{em} \quad \alpha \to \infty, \end{cases} \tag{20}$$

onde $u, v \in \mathbb{R}$ e $\varepsilon, \delta \geq 0$ são hiperparâmetros fixos.

Informalmente, trata-se da família de densidades cuja dependência em r pode ser minorada e majorada por uma expressão proporcional à densidade Beta de parâmetros a e b. Alem disso, é necessário que as expressões majorante e minorante tenham comportamento polinomial em α nos dois extremos de integração (0 e ∞). Os parâmetros ε e δ permitem flexibilidade para que as cotas inferior e superior tenham comportamentos diferentes entre si.

4.1 Propriedade e impropriedade *a priori*

Trata-se, nesta seção, de estabelecer condições de convergência da expressão

$$\int_0^\infty f_S(\alpha) \int_0^1 r^a (1-r)^b \, dr \, d\alpha, \tag{21}$$

ou de divergência da expressão

$$\int_0^\infty f_I(\alpha) \int_0^1 r^a (1-r)^b \, dr \, d\alpha. \tag{22}$$

A integral em relação a r nas expressões acima converge se e somente se ambos os parâmetros a e b forem maiores do que -1 (conforme Corolário 3). Disso se conclui imediatamente que a densidade é imprópria sempre que $a \leq -1 \quad \text{ou} \quad b \leq -1$.

No caso contrário, a propriedade depende da convergência da integral em relação a α. Aplicando o Teorema 2, vemos que a integral da cota superior

$$\int_0^\infty f_S(\alpha) \, d\alpha$$

converge se e somente se $u > -1$ e $v < -1$.

Se a integral acima não for convergente, a propriedade da densidade *a priori* não pode ser estabelecida. No entanto, podemos avaliar a integral da cota inferior

$$\int_0^\infty f_I(\alpha) \, d\alpha$$

que diverge se e somente se $u + \varepsilon \leq -1$ ou $v - \delta \geq -1$.

Em resumo, temos as seguintes condições *suficientes* de

1. *propriedade*: $a > -1$ e $b > -1$ e $u > -1$ e $v < -1$
2. *impropriedade*: $a \leq -1$ ou $b \leq -1$ ou $u + \varepsilon \leq -1$ ou $v - \delta \geq -1$

Nota: se $\varepsilon = \delta = 0$, as condições acima são também *necessárias*.

Por outro lado, se $\varepsilon > 0$ ou $\delta > 0$ (e se $a, b > -1$), existe uma região de valores de u, v para os quais não se pode estabelcer nem a propriedade nem a impropriedade. Essa região é dada por

$$(u \leq -1 \quad \text{ou} \quad v \geq -1) \quad \text{e} \quad u > -1 - \varepsilon \quad \text{e} \quad v < -1 + \delta.$$

4.2 Os parâmetros que correspondem às densidades de Jeffreys

Para explicitar os parâmetros que definem a densidade *a priori* de Jeffreys (10), vamos escrevê-la na forma

$$p_J(\alpha, r) \propto \frac{r^{1/2}}{1-r} \sqrt{\frac{R}{r} \sum_{j=1}^{n} \frac{r_j^2}{r^2} \frac{F(\alpha, r_j) - 1}{r_j}}.$$

As quantidades R/r e $(r_j/r)^2$, consideradas como funções de $r \in (0, 1)$, são limitadas superior e inferiormente por valores positivos. Para verificar isto, basta notar que:

$$\lim_{r \to 0} \frac{R}{r} = \sum_{j=1}^{n} t_j, \qquad \lim_{r \to 1} \frac{R}{r} = n,$$

$$\lim_{r \to 0} \frac{r_j^2}{r^2} = t_j^2, \qquad \lim_{r \to 1} \frac{r_j^2}{r^2} = 1.$$

Assim, as expressões definem funções contínuas no intervalo compacto $[0, 1]$, sendo portanto limitadas. Além disso, são estritamente positivas, uma vez que $r_j \in (0, 1)$ e $t_j > 0$.

Para aplicar o Teorema 8, podemos escrever p_J como $p_J(\alpha, r) \propto r^{1/2}(1 - r)^{-1} f(\alpha, r)$. Segue que $f(\alpha, r)$ possui cotas inferior e superior nas formas

$$f_I(\alpha) = \sqrt{c \frac{1}{2 + 2\alpha}} \text{ e}$$

$$f_S(\alpha) = \sqrt{C(\alpha \psi'(\alpha) - 1)},$$

e c e C são constantes que não dependem de r nem de α. O comportamento assintótico de f_I é dado por

$$f_I(\alpha) \sim \begin{cases} \alpha^0 & \text{em} \quad \alpha \to 0 \quad \text{e} \\ \alpha^{-1/2} & \text{em} \quad \alpha \to \infty, \end{cases}$$

e o de f_S por

$$f_S(\alpha) \sim \begin{cases} \alpha^{-1/2} & \text{em} \quad \alpha \to 0 \quad \text{e} \\ \alpha^{-1/2} & \text{em} \quad \alpha \to \infty. \end{cases}$$

Concluímos assim que a densidade *a priori* de Jeffreys pertence à família (18) com parâmetros

$$a = 1/2,\ b = -1,\ u = v = -1/2,\ \varepsilon = 1/2,\ \delta = 0. \tag{23}$$

Segue que essa densidade é *imprópria*.

A densidade de Jeffreys de independência (11) pode ser escrita como

$$p_{JI}(\alpha, r) \propto \frac{1}{1-r}\sqrt{\frac{R}{r}\sum_{j=1}^{n}\frac{r_j}{r}F(\alpha, r_j)}.$$

Pelo Teorema 7, podemos escrever a densidade acima como

$$p_{JI}(\alpha, r) \propto (1-r)^{-1} f(\alpha, r),$$

onde $f(\alpha, r)$ possui cotas inferior e superior nas formas $f_I(\alpha) = c$ e $f_S(\alpha) = \sqrt{C\alpha\psi'(\alpha)}$, para certas constantes c e C. O comportamento assintótico de f_S é dado por

$$f_S(\alpha) \sim \begin{cases} \alpha^{-1/2} & \text{em} \quad \alpha \to 0 \quad \text{e} \\ 1 & \text{em} \quad \alpha \to \infty. \end{cases}$$

Concluímos assim que a densidade *a priori* de Jeffreys de independência pertence à família (18) com parâmetros

$$a = 0,\ b = -1,\ u = -1/2,\ v = 0,\ \varepsilon = 1/2,\ \delta = 0. \tag{24}$$

Segue que essa densidade também é *imprópria*.

4.3 Outras densidades especiais

Nessa seção consideraremos densidades definidas por expressões da forma

$$p_{\alpha\beta}(\alpha,\beta) = \alpha^{k_1}\beta^{k_2} \text{ e}$$

$$p_{HB}(\alpha,\beta) = \alpha^{k_1}(\alpha+s_1)^{k_2}\beta^{k_3}(\beta+s_2)^{k_4},$$

onde β é o parâmetro de escala em (3), k_1, k_2, k_3 e k_4 são hiperparâmetros reais e s_1 e s_2 são hiperparâmetros reais positivos. A forma funcional de p_{HB} é proposta por Hadjicostas e Berry [12].

Como as expressões estão em termo de α e β, vamos supor a inveriância para transformá-las em termos de α e r. O determinante jacobiano da transformação

$$\varphi(\alpha,r) = \left(\alpha, \frac{r}{1-r}\right)$$

é dado por

$$\det J_\varphi = (1-r)^{-2}.$$

Dessa forma, as densidades se escrevem como

$$p_{\alpha\beta}(\alpha,r) = \alpha^{k_1}r^{k_2}(1-r)^{-k_2-2} \text{ e}$$

$$p_{HB}(\alpha,r) = \alpha^{k_1}(\alpha+s_1)^{k_2}r^{k_3}(1-r)^{-k_3-k_4-2}(r+(1-r)s_2)^{k_4}.$$

A primeira forma funcional se enquadra na família (18) com $f_I(\alpha) = f_S(\alpha) = \alpha^{k_1}$, com parâmetros

$$a = k_2,\ b = -k_2-2,\ u = v = k_1,\ \varepsilon = \delta = 0. \tag{25}$$

Para a outra densidade (p_{HB}), observamos que $(r+(1-r)s_2)$ é limitado por constantes positivas para todo $r \in (0,1)$. Dessa forma, $f(\alpha,r) = \alpha^{k_1}(\alpha+s_1)^{k_2}(r+(1-r)s_2)^{k_4}$ admite cotas

$$f_I(\alpha) \sim f_S(\alpha) \sim \begin{cases} \alpha^{k_1} & \text{em} \quad \alpha \to 0 \quad \text{e} \\ \alpha^{k_1+k_2} & \text{em} \quad \alpha \to \infty. \end{cases}$$

Conclui-se, portanto, que a densidade proposta por Hadjicostas e Berry se enquadra na família (18) com os seguintes parâmetros:

$$a = k_3,\ b = -k_3 - k_4 - 2,\ u = k_1,\ v = k_1 + k_2,\ \varepsilon = \delta = 0. \tag{26}$$

5 A densidade *a posteriori* para o modelo Poisson-Gama

A partir da família de densidades *a priori* considerada na seção anterior, vamos tratar a densidade *a posteriori* dada por

$$p(\alpha, r|\boldsymbol{x}) \propto p(\boldsymbol{x}|\alpha, r) r^a (1-r)^b f(\alpha, r) \propto$$

$$\propto \prod_{j=1}^{n} \left(\frac{\Gamma(\alpha + x_j)}{\Gamma(\alpha)} r_j^{x_j} (1 - r_j)^\alpha \right) r^a (1-r)^b f(\alpha, r). \tag{27}$$

Definindo $x_{n+1} = a$, $y_j = \alpha$ (para $j \in \{1, ..., n\}$), $y_{n+1} = b$ e $t_{n+1} = 1$, podemos escrever a densidade *a posteriori* como

$$p(\alpha, r|\boldsymbol{x}) \propto Q(\alpha) R(\alpha, r) f(\alpha, r), \tag{28}$$

onde

$$Q(\alpha) = \prod_{j=1}^{n} \frac{\Gamma(\alpha + x_j)}{\Gamma(\alpha)} \quad \text{e}$$

$$R(\alpha, r) = \prod_{j=1}^{n+1} r_j^{x_j} (1 - r_j)^{y_j}.$$

Essa fatoração da densidade *a posteriori* permite tratar a integração com respeito a r primeiramente, e depois com respeito a α.

Além disso, outras duas observações são relevantes para o análise da integrabilidade da expressão acima:

Lema 9. *As transformações* $g_t : (0,1) \longrightarrow (0,1)$*, definidas por*

$$\left\{ g_t(r) = \frac{tr}{1+(t-1)r} \right\}_{t>0},$$

formam um grupo em relação à operação de composição tal que

$$g_t \circ g_s = g_{ts} \quad \text{e} \quad g_t^{-1} = g_{1/t}.$$

Demonstração. Basta verificar que

$$g_t(g_s(r)) = \frac{t}{1/g_s(r) + (t-1)} = \frac{t}{\frac{1/r+(s-1)}{s} + (t-1)} =$$

$$= \frac{ts}{1/r + (s-1) + (ts - s)} = \frac{ts}{1/r + (ts-1)},$$

e que $g_1(r) = r$. ■

Lema 10. *A expressão* $\Gamma(x+p)/\Gamma(x+q)$*, com* $p, q \geq 0$*, apresenta o seguinte comportamento assintótico:*

$$\frac{\Gamma(x+p)}{\Gamma(x+q)} \sim \begin{cases} \left\{ \begin{array}{lll} x, & \text{se} & p \neq 0, q = 0 \\ 1, & \text{se} & p, q > 0 \\ 1/x, & \text{se} & p = 0, q \neq 0 \end{array} \right\} & \text{em } x \to 0 \quad \text{e} \\ \qquad x^{p-q} & \text{em } x \to \infty. \end{cases}$$

Demonstração. Quando $x \to 0$, há três possibilidades: se $p, q > 0$, então

$$\frac{\Gamma(x+p)}{\Gamma(x+q)} \to \frac{\Gamma(p)}{\Gamma(q)}.$$

Se $p = 0$ e $q > 0$,

$$x\frac{\Gamma(x)}{\Gamma(x+q)} = \frac{\Gamma(x+1)}{\Gamma(x+q)} \to \frac{\Gamma(1)}{\Gamma(q)}.$$

Por fim, se $p > 0$ e $q = 0$,

$$\frac{1}{x}\frac{\Gamma(x+p)}{\Gamma(x)} = \frac{\Gamma(x+p)}{\Gamma(x+1)} \to \frac{\Gamma(p)}{\Gamma(1)}.$$

Em $x \to \infty$, vale a aproximação de Stirling $\Gamma(z) \sim \sqrt{2\pi} z^{z-1/2} e^{-z}$:

$$(x+p)^{q-p} \frac{\Gamma(x+p)}{\Gamma(x+q)} \sim (x+p)^{q-p} \frac{(x+p)^{x+p-1/2} e^{-x-p}}{(x+q)^{x+q-1/2} e^{-x-q}} =$$

$$= \left(\frac{x+p}{x+q} \right)^{x+q-1/2} e^{q-p} = \left(1 + \frac{p-q}{x+q} \right)^{x+q} \left(\frac{x+p}{x+q} \right)^{-1/2} e^{q-p}.$$

Tomando o limite e usando o fato que

$$\lim_{n \to \infty} \left(1 + \frac{y}{n} \right)^n = e^y,$$

e que $\sqrt{(x+q)/(x+p)} \to 1$, segue

$$(x+p)^{q-p} \frac{\Gamma(x+p)}{\Gamma(x+q)} \to 1, \; x \to \infty.$$

Para concluir pelo enunciado, basta observar que $(x+p)^{q-p} \sim x^{q-p}$, pois $(x+p)/x \to 1$ quando $x \to \infty$. ■

5.1 Resultados intermediários para o estudo de integrabilidade da densidade *a posteriori*

Nesta seção apresentaremos dois resultados que serão utilizados para estabelecer as condições de integrabilidade da distribuição *a posteriori*. Serão utilizadas ao longo do texto as seguintes definições, para simplificar a notação:

$$s = \sum_{j=1}^{n} x_j \quad \text{e} \quad n' = |\{j : \, x_j \neq 0\}|,$$

onde $|\{...\}|$ indica o número de elementos de um conjunto.

Além disso, definimos b^* como

$$b^* = \begin{cases} 0 & \text{se } b > -1 \\ 1 & \text{se } b = -1. \end{cases} \tag{29}$$

Lema 11. *O termo $Q(\alpha)$ satisfaz o seguinte comportamento assintótico:*

$$Q(\alpha) \sim \begin{cases} \alpha^{n'} & \text{em } \alpha \to 0, \text{ e} \\ \alpha^{s} & \text{em } \alpha \to \infty. \end{cases}$$

Demonstração. Para cada $j \in \{1, ..., n\}$, o fator $\Gamma(\alpha + x_j)/\Gamma(\alpha)$ é igual a 1, se $x_j = 0$, ou então se reduz a

$$\prod_{i=1}^{x_j} (\alpha + x_j - i)\,,$$

aplicando sucessivamente a relação $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$. Em ambos os casos, trata-se de polinômios de grau x_j. Definindo $\prod_{i=1}^{0} c_i = 1$, $Q(\alpha)$ pode ser escrito como

$$Q(\alpha) = \prod_{j=1}^{n} \prod_{i=1}^{x_j} (\alpha + x_j - i)\,,$$

que resulta em um polinômio de grau s em α.

Para cada observação não nula, $Q(\alpha)$ possui um fator da forma α, correspondente ao termo em que $i = x_j$ no produtório acima. Os demais fatores convergem para uma constante positiva quando α tende a 0. Segue, portanto, que $Q(\alpha) \sim \alpha^{n'}$ em $\alpha \to 0$.

Quando α tende a ∞, todos os fatores são de ordem α, de forma que $Q(\alpha) \sim \alpha^{s}$ em $\alpha \to \infty$. ■

Lema 12. *A integral $\int_0^1 R(\alpha, r)\, dr$ é finita (para todo $\alpha > 0$) se e somente se os parâmetros a e b da distribuição a priori satisfazem as seguintes condições:*

$$a > -1 - s \quad \text{e} \quad b \geq -1. \tag{30}$$

Além disso, quando convergente, a integral satisfaz as seguintes desigualdades:

$$c\,\mathrm{B}(s + a + 1, n\alpha + b + 1) \leq \int_0^1 R(\alpha, r)\, dr \leq C\,\mathrm{B}(s + a + 1, n\alpha + b + 1),$$

onde c e C são constantes positivas que só dependem de t_j e x_j, e B é a função Beta.

Demonstração. Utilizando a notação do Lema 9, podemos escrever o fator R como

$$R(\alpha, r) = \prod_{j=1}^{n+1} g_{t_j}(r)^{x_j}(1 - g_{t_j}(r))^{y_j}.$$

Definindo $\mu = \max(t_1, ..., t_n)$, considere a transformação de variáveis $r \mapsto r_\mu = g_\mu(r)$. A tranformação inversa é dada por $r = g_{1/\mu}(r_\mu)$. Logo, $r_j = g_{t_j}(r) = g_{t_j}(g_{1/\mu}(r_\mu)) = g_{t_j/\mu}(r_\mu)$. A expressão $R(\alpha, r)$ se escreve nesses termos como:

$$R(\alpha, g_{1/\mu}(r_\mu)) = \prod_{j=1}^{n+1} g_{t_j/\mu}(r_\mu)^{x_j}(1 - g_{t_j/\mu}(r_\mu))^{y_j} =$$

$$= \prod_{j=1}^{n+1} \left(\frac{u_j r_\mu}{1 + (u_j - 1)r_\mu}\right)^{x_j} \left(\frac{1 - r_\mu}{1 + (u_j - 1)r_\mu}\right)^{y_j} =$$

$$= \frac{\prod_{j=1}^{n+1} \left(u_j^{x_j}\right) r_\mu^{s+a}(1 - r_\mu)^{n\alpha+b}}{\prod_{j=1}^{n+1} (1 + (u_j - 1)r_\mu)^{x_j+y_j}}, \tag{31}$$

onde $u_j = t_j/\mu$ e $s = \sum_{j=1}^{n} x_j$.

O denominador em (31) pode ser reescrito como:

$$\left[\prod_{j=1}^{n} (1 + (u_j - 1)r_\mu)^{x_j+y_j}\right] (1 + (1/\mu - 1)r_\mu)^{a+b}. \tag{32}$$

O termo $(1 + (1/\mu - 1)r_\mu)^{a+b}$ pode ser majorado por uma constante positiva, pois o termo entre parênteses é positivo para todo $r_\mu \in [0, 1]$.

No produtório de (32), as expressões $(1 + (u_j - 1)r_\mu)$ são positivas e menores ou iguais a 1, pois $u_j \leq 1$ para todo $j \leq n$. Como os respectivos expoentes $x_j + y_j = x_j + \alpha$ são positivos, o fator

$$\prod_{j=1}^{n} (1 + (u_j - 1)s)^{x_j+y_j}$$

é menor ou igual a 1 para todo $r_\mu \in [0, 1]$ (e para todo $\alpha > 0$).

Nota: embora esse fator também admita uma cota inferior positiva (em relação à variável r_μ), esta não será utilizada, pois apresentaria dependência em α, devido à presença dos expoentes y_j.

Conclui-se que a expressão (32) pode ser majorada por uma constante. Disso decorre que a expressão de R em (31) admite cota inferior na forma:

$$c_1\, r_\mu^{s+a}(1-r_\mu)^{n\alpha+b} \leq R(\alpha, g_{1/\mu}(r_\mu)), \tag{33}$$

para alguma constante $c_1 > 0$.

Por outro lado, definindo $\nu = \min(t_1, ..., t_{n+1})$, $r_\nu = g_\nu(r)$ e $v_j = t_j/\nu$, podemos escrever

$$R(\alpha, g_{1/\nu}(r_\nu)) = \frac{\prod_{j=1}^{n+1}\left(v_j^{x_j}\right) r_\nu^{s+a}(1-r_\nu)^{n\alpha+b}}{\prod_{j=1}^{n+1}\left(1+(v_j-1)r_\nu\right)^{x_j+y_j}}.$$

O denominador da expressão acima é dado por

$$\prod_{j=1}^{n}\left(1+(v_j-1)r_\nu\right)^{x_j+y_j}\left(1+(1/\nu-1)r_\nu\right)^{a+b}.$$

As expressões $(1+(v_j-1)r_\mu)$ são positivas e maiores ou iguais a 1, pois $v_j \geq 1$ para todo $j \leq n$. Segue, em analogia ao caso anterior, que o denominador acima pode ser minorado por uma constante positiva. Disso obtemos uma cota superior para a expressão de R como função de r_ν:

$$R(\alpha, g_{1/\nu}(r_\nu)) \leq C_1\, r_\nu^{s+a}(1-r_\nu)^{n\alpha+b}, \tag{34}$$

para uma certa constante C_1.

O jacobiano das transformações da forma $r = g_{1/\tau}(r_\tau)$, onde $\tau \in \{\mu, \nu\}$, é dado por

$$J = \frac{dr}{dr_\tau} = g'_{1/\tau}(r_\tau) = \frac{1/\tau}{\left(1+(1/\tau-1)r_\tau\right)^2}.$$

Em ambos os casos, a expressão de J é positiva para todo $r_\tau \in [0,1]$ e pode, portanto, ser minorada e majorada por constantes.

Para provar a suficiência das condições do enunciado para a convergência da integral de $R(\alpha, r)$, suponha que a e b satisfazem as restrições dadas em (30). Sejam ξ, ζ tais que $0 < \xi < \zeta < 1$. Pelo teorema de mudança de variável, podemos escrever:

$$\int_{\xi}^{\zeta} R(\alpha, r)\, dr = \int_{g_\nu(\xi)}^{g_\nu(\zeta)} R(\alpha, g_{1/\nu}(r_\nu))|J|\, dr_\nu \quad \text{e}$$

$$\int_{\xi}^{\zeta} R(\alpha, r)\, dr = \int_{g_\mu(\xi)}^{g_\mu(\zeta)} R(\alpha, g_{1/\mu}(r_\mu))|J|\, dr_\mu.$$

As integrais no lado direito admitem, respectivamente, cotas superior e inferior na forma:

$$\int_{g_\nu(\xi)}^{g_\nu(\zeta)} R(\alpha, g_{1/\nu}(r_\nu))|J|\, dr_\nu < C_2 \int_{g_\nu(\xi)}^{g_\nu(\zeta)} r_\nu^{s+a}(1 - r_\nu)^{n\alpha+b}\, dr_\nu \quad \text{e}$$

$$c_2 \int_{g_\mu(\xi)}^{g_\mu(\zeta)} r_\mu^{s+a}(1 - r_\mu)^{n\alpha+b}\, dr_\mu < \int_{g_\mu(\xi)}^{g_\mu(\zeta)} R(\alpha, g_{1/\mu}(r_\mu))|J|\, dr_\mu,$$

para certas constantes c_2 e C_2. Tomando o limite em que $\xi \to 0^+, \zeta \to 1^-$, os integrandos da forma $r_\tau^{s+a}(1 - r_\tau)^{n\alpha+b}$, $\tau \in \{\mu, \nu\}$, resultam em $\mathrm{B}(s + a + 1, n\alpha + b + 1)$, pois tanto $s + a + 1$ quanto $n\alpha + b + 1$ são positivos por hipótese.

Disso se conclui pelo enunciado:

$$c\,\mathrm{B}(s + a + 1, n\alpha + b + 1) \leq \int_0^1 R(\alpha, r)\, dr \leq C\,\mathrm{B}(s + a + 1, n\alpha + b + 1),$$

para certas constantes c e C.

Para provar a necessidade das restrições em a e b, suponha por absurdo que $s + a \leq -1$ ou $b < -1$. Nesse caso, a cota inferior

$$c_2 \int_{g_\mu(\xi)}^{g_\mu(\zeta)} r_\mu^{s+a}(1 - r_\mu)^{n\alpha+b}\, dr_\mu$$

diverge no limite em que $\xi \to 0^+, \zeta \to 1^-$. Portanto, a integral de $R(\alpha, r)$ em $r \in (0, 1)$ não pode ser convergente. ■

Corolário 13. *Sob as mesmas condições de integrabilidade de $R(\alpha, r)$, a função*

$$S(\alpha) = \int_0^1 R(\alpha, r)\, dr$$

admite cotas inferior $S_I(\alpha)$ e superior $S_S(\alpha)$ cujo comportamento assintótico é dado por

$$S_I(\alpha) \sim S_S(\alpha) \sim \begin{cases} \alpha^{-b^*} & \text{em } \alpha \to 0 \quad \text{e} \\ \alpha^{-s-a-1} & \text{em } \alpha \to \infty. \end{cases}$$

Demonstração. Pelo Lema 12, verifica-se que as cotas obtidas para $S(\alpha)$ possuem comportamento assintótico determinado pela expressão $\mathrm{B}(s + a + 1, n\alpha + b + 1)$, tanto em $\alpha \to 0$ quanto em $\alpha \to \infty$.

Pela definição da função Beta, podemos escrever:

$$\mathrm{B}(s + a + 1, n\alpha + b + 1) = \Gamma(s + a + 1)\frac{\Gamma(n\alpha + b + 1)}{\Gamma(s + n\alpha + a + b + 2)}.$$

O fator $\Gamma(s + a + 1)$ é constante. Aplicando o Lema 10 no outro fator, com $x = n\alpha$, $p = b + 1$ e $q = s + a + b + 2$, concluímos que

$$\mathrm{B}(s + a + 1, n\alpha + b + 1) \sim \begin{cases} \left\{ \begin{array}{cc} 1, & \text{se} \quad b > -1 \\ \alpha^{-1}, & \text{se} \quad b = -1 \end{array} \right\} & \text{em } \alpha \to 0 \quad \text{e} \\ \alpha^{-s-a-1} & \text{em } \alpha \to \infty. \end{cases}$$

Utilizando a definição de b^*, esse resultado pode ser expresso na forma compacta do enunciado do Corolário. ■

5.2 Condições de integrabilidade da densidade *a posteriori*

Teorema 14. *Para que a densidade $p(\alpha, r|\boldsymbol{x})$ seja própria, são suficientes as seguintes condições:*

$$a > -1 - s \quad \text{e} \quad b \geq -1 \quad \text{e} \quad u > b^* - 1 - n' \quad \text{e} \quad v < a. \tag{35}$$

Conversamente, para que a densidade $p(\alpha, r|\boldsymbol{x})$ seja imprópria, são suficientes as condições:

$$a \leq -1 - s \quad \text{ou} \quad b < -1 \quad \text{ou} \quad u + \varepsilon \leq b^* - 1 - n' \quad \text{ou} \quad v - \delta \geq a. \quad (36)$$

Nota: no caso $\varepsilon = \delta = 0$, as condições acima são também necessárias.

Demonstração. Suponhamos válidas as condições em (35). Aplicando a decomposição em (28) e as propriedades de f, podemos obter uma cota superior para a densidade *a posteriori* na forma:

$$p(\alpha, r|\mathbf{x}) < C\, Q(\alpha) R(\alpha, r) f_S(\alpha),$$

para alguma constante C.

As condições $a > -1 - s$ e $b \geq -1$ garantem a aplicabilidade do Lema 12. Portanto, integrando o lado direito da expressão acima (ignorando a constante) com respeito a r obtemos uma majoração da forma

$$\int_0^1 Q(\alpha) R(\alpha, r) f_S(\alpha)\, dr \leq Q(\alpha) S_S(\alpha) f_S(\alpha) = T(\alpha).$$

O comportamento assintótico de $T(\alpha)$ pode ser obtido combinando os Lemas 11 e 12 e a expressão (20):

$$T(\alpha) \sim \begin{cases} \alpha^{n'-b^*+u} & \text{em } \alpha \to 0 \quad \text{e} \\ \alpha^{-a-1+v} & \text{em } \alpha \to \infty. \end{cases}$$

Das condições $u > b^* - 1 - n'$ e $v < a$, segue que os expoentes de α acima satisfazem

$$n' - b^* + u > -1 \quad \text{e} \quad -a - 1 + v < -1.$$

Pelo Teorema 2, conclui-se que $T(\alpha)$ é integrável, e portanto $p(\alpha, r|\mathbf{x})$ é própria.

Para provar que as condições (36) são suficientes para que a integral da densidade *a posteriori* divirja, vamos considerar uma cota inferior da forma

$$p(\alpha, r|\mathbf{x}) \geq c\, Q(\alpha) R(\alpha, r) f_I(\alpha),$$

onde c é uma constante positiva.

Suponha, inicialmente, que vale $a \leq -1-s$ ou $b < -1$. Nesse caso, pelo Lema 12, temos

$$\int_0^1 R(\alpha, r)\, dr = \infty,$$

para todo $\alpha \in (0, \infty)$ (caso $a \leq -1-s$) ou somente num intervalo da forma $(0, \alpha_0 = -(b+1)/n]$ (caso somente $b < -1$). Em ambos os casos, é possível definir um intervalo fechado I onde a integral de $R(\alpha, r)$ diverge para todo $\alpha \in I$. Nesse intervalo I, as funções Q e f_I admitem cotas constantes positivas, visto que são funções contínuas e positivas restritas a um domínio compacto. Segue, portanto, que

$$\int_0^\infty \int_0^1 p(\alpha, r|\mathbf{x})\, dr\, d\alpha \geq \int_I \int_0^1 p(\alpha, r|\mathbf{x})\, dr\, d\alpha \geq$$

$$\geq \int_I K \int_0^1 R(\alpha, r)\, dr\, d\alpha = \infty.$$

Para completar a prova, suponha que $a > -1-s$ e $b \geq -1$, de forma que a integral de $R(\alpha, r)$ seja convergente, mas que $u + \varepsilon \leq b^* - 1 - n'$ ou $v - \delta \geq a$. Nesse caso, a cota inferior para $p(\alpha, r|\mathbf{x})$ pode ser integrada com respeito a r (ignorando a constante c), resultando em

$$\int_0^1 Q(\alpha) R(\alpha, r) f_I(\alpha)\, dr \geq Q(\alpha) S_I(\alpha) f_I(\alpha) = t(\alpha).$$

O comportamento assintótico de $t(\alpha)$ é obtido considerando as propriedades de f_I em (19), analogamente à prova de (35), e é dado por:

$$t(\alpha) \sim \begin{cases} \alpha^{n'-b^*+u+\varepsilon} & \text{em } \alpha \to 0 \quad \text{e} \\ \alpha^{-a-1+v-\delta} & \text{em } \alpha \to \infty. \end{cases}$$

Por hipótese, $u + \varepsilon \leq b^* - 1 - n'$ ou $v - \delta \geq a$, de forma que os expoentes acima satisfazem

$$n' - b^* + u + \varepsilon \leq -1 \quad \text{ou} \quad -a - 1 + v - \delta \geq -1.$$

Aplicando o Teorema 2, conclui-se que a integral de $t(\alpha)$ em $(0, \infty)$ é divergente.

Por fim, observamos que, no caso $\varepsilon = \delta = 0$, as condições (35) e (36) são opostas, de forma que não há região de indecisão acerca da propriedade ou impropriedade de $p(\alpha, r|\mathbf{x})$ em função dos parâmetros a, b, u, v, e dos observáveis n' e s. ■

5.2.1 Integrabilidade de algumas densidades ***a posteriori*** especiais

Nessa seção, veremos que o Teorema 14 permite estabelecer condições simples para a convergência ou divergência das densidades especiais tratadas na seção 4.

Corolário 15. *Para que a densidade a posteriori de Jeffreys seja própria é necessário e suficiente que haja pelo menos uma observação não-nula.*

Demonstração. Suponha inicialmente $n' \geq 1$. Aplicando as condições (35) aos parâmetros (23), que todas as condições são satisfeitas:

1. $a = 1/2 > -1 - s$ (sempre vale uma vez que $s \geq 0$).
2. $b = -1 \geq -1$.
3. $u = -1/2 > b^* - 1 - n' = -n' \leq -1$.
4. $v = -1/2 < a = 1/2$.

Conversamente, se $n' = 0$, a terceira condição em (36) implica a divergência: $u + \varepsilon = 0 \leq b^* - 1 - n' = 0$. ■

Corolário 16. *A densidade a posteriori de Jeffreys de independência é sempre imprópria.*

Demonstração. Pela quarta condição em (36), aplicada aos parâmetros de p_{JI}, temos $v - \delta = 0 \geq a = 0$, o que implica em divergência *a posteriori*, independentemente dos dados observados. ■

Corolário 17. *Para que a densidade a posteriori associada a $p_{\alpha\beta}$ seja convergente, é necessário e suficiente um dos dois conjuntos de condições a seguir:*

1. $k_2 = -1$ *e* $1 < -k_1 < n'$ *e* $n' \geq 2$*;*
2. $1 < -k_2 < 1 + s$ *e* $-k_2 < -k_1 < 1 + n'$ *e* $n' \geq 1$*.*

Demonstração. Suponha inicialmente a primeira alternativa de condições. Aplicando (25), temos como condições equivalentes:

1. $a = k_2 = -1 > -1 - s$, pois $n' \geq 1$ implica $s \geq 1$.
2. $b = -k_2 - 2 = -1 \geq -1$.
3. $u = k_1 > -n'$. $(b^* = 1)$
4. $v = k_1 < -1 = k_2 = a$.

Supondo o segundo conjunto de condições, obtemos as seguinte formulação equivalente:

1. $a = k_2 > -1 - s$.
2. $b = -k_2 - 2 > -1$.
3. $u = k_1 > -1 - n'$. $(b^* = 0)$
4. $v = k_1 < k_2 = a$.

Como $\varepsilon = \delta = 0$, tratam-se de condições também necessárias para a integrabilidade *a posteriori* ■

Corolário 18. *Para que a densidade a posteriori associada a* p_{HB} *seja convergente, é necessário e suficiente um dos dois conjuntos de condições a seguir:*

1. $k_1 + n' + 1 > 0$ *e* $k_3 + s + 1 > 0$ *e* $k_3 + k_4 + 1 < 0$ *e* $k_3 > k_1 + k_2$*;*
2. $k_1 + n' > 0$ *e* $k_3 + s + 1 > 0$ *e* $k_3 + k_4 + 1 = 0$ *e* $k_3 > k_1 + k_2$*.*

Demonstração. Aplicando (26), obtemos, na primeira alternativa:

1. $a = k_3 > -1 - s$.
2. $b = -k_3 - k_4 - 2 > -1$.
3. $u = k_1 > -1 - n'$. $(b^* = 0)$
4. $v = k_1 + k_2 < k_3 = a$.

Na segunda alternativa:

1. $a = k_3 > -1 - s$.
2. $b = -k_3 - k_4 - 2 = -1$.
3. $u = k_1 > -n'$. $(b^* = 1)$
4. $v = k_1 + k_2 < k_3 = a$.

■

Nota: as condições do Corolário 18 são equivalentes às de [12, Teorema 3.1].

5.2.2 Observação sobre a integrabilidade da densidade *a posteriori*

A propriedade da distribuição *a posteriori* dadas n observações independentes também pode ser analisada sob a ótica da coleta sequencial de dados. É uma consequencia do paradigma bayesiano que a expressão da densidade *a posteriori* pode ser escrita como

$$p(\theta|x_1,...,x_n) \propto p(x_1,...,x_n|\theta)p(\theta) =$$

$$= \prod_{i=1}^{n} (p(x_i|\theta))\, p(\theta) =$$

$$= \prod_{i=2}^{n} (p(x_i|\theta))\, p(x_1|\theta)p(\theta) \propto$$

$$\propto \prod_{i=2}^{n} (p(x_i|\theta))\, p(\theta|x_1).$$

Ou seja, dada a primeira observação, a densidade *a posteriori* obtida pode ser usada como densidade *a priori* para as demais observações, e assim se obtém o mesmo resultado que se obteria partindo da densidade *a priori* original. Mais geralmente, vale

$$p(\theta|x_1,...,x_n) \propto \prod_{i=k+1}^{n} (p(x_i|\theta))\, p(\theta|x_1,...,x_k),$$

para todo $k \in \{1,...,n-1\}$.

As condições de convergência do Teorema 14 apresentam explicitamente essa propriedade: se as condições (35) são válidas para certos valores de s e n', serão também válidas para quaisquer valores iguais ou superiores dessas estatísticas. Como s e n' não decrescem quando novas observações são acrescentadas à amostra, se a densidade *a posteriori* associada a uma amostra for integrável, então a densidade associada a uma amostra maior também será.

6 Existência de momentos para as variáveis latentes

O modelo hierárquico também pode ser visto sob a ótica em que as variáveis latentes, ao invés de serem integradas para dar origem a um modelo simples, são considerada parte do espaço paramétrico. Sob esse ponto de vista alternativo, cabe questionar a propriedade da distribuição *a posteriori* desse vetor.

Inicialmente, vamos obter a distribuição de $\lambda_j|\boldsymbol{x},\alpha,r$. Essa densidade será utilizada também para a inferência sobre os momentos de $\lambda_j|\boldsymbol{x}$. e em seguida vamos mostrar uma condição necessária e suficiente para a propriedade de $\lambda_j|\boldsymbol{x}$.

Lema 19. *A distribuição de $\lambda_j|\boldsymbol{x},\alpha,r$ é Gama, na forma*

$$\lambda_j|\boldsymbol{x},\alpha,r \sim \text{Gama}\left(\alpha + x_j, t_j/g_{t_j}(r)\right). \tag{37}$$

Demonstração. A partir das densidades de $X_j|\lambda_j \sim \text{Poisson}(\lambda_j t_j)$, $\lambda_j|\alpha,r \sim \text{Gama}(\alpha,(1-r)/r)$ e $p(\boldsymbol{x}|\alpha,r)$, e observando que

$$p(\boldsymbol{\lambda}|\boldsymbol{x},\alpha,r) = \frac{p(\boldsymbol{\lambda},\boldsymbol{x}|\alpha,r)}{p(\boldsymbol{x}|\alpha,r)} = \frac{p(\boldsymbol{x}|\boldsymbol{\lambda},\alpha,r)p(\boldsymbol{\lambda}|\alpha,r)}{p(\boldsymbol{x}|\alpha,r)} = \frac{p(\boldsymbol{x}|\boldsymbol{\lambda})p(\boldsymbol{\lambda}|\alpha,r)}{p(\boldsymbol{x}|\alpha,r)},$$

obtemos

$$p(\boldsymbol{\lambda}|\boldsymbol{x},\alpha,r) = \prod_{j=1}^{n} \frac{(\lambda_j t_j)^{x_j} e^{-\lambda_j t_j}}{x_j!} \frac{\lambda_j^{\alpha-1} e^{-\lambda_j/\beta}}{\beta^\alpha \Gamma(\alpha)} \frac{\Gamma(\alpha) x_j! (1+\beta t_j)^{\alpha+x_j}}{\Gamma(\alpha+x_j)(\beta t_j)^{x_j}},$$

onde $\beta = r/(1-r)$. Essa densidade pode ser fatorada e simplificada, resultando em

$$p(\lambda_j|\boldsymbol{x},\alpha,r) = \frac{(1+\beta t_j)^{\alpha+x_j} \lambda_j^{\alpha+x_j-1} e^{-\lambda_j(1/\beta+t_j)}}{\beta^{\alpha+x_j}\Gamma(\alpha+x_j)}.$$

Observando ainda as identidades

$$\frac{1+\beta t_j}{\beta} = \frac{1}{\beta} + t_j = \frac{1-r}{r} + t_j = \frac{t_j}{g_{t_j}(r)},$$

conclui-se que a distribuição de $\lambda_j|\boldsymbol{x},\alpha,r$ é Gama, com parâmetros de forma $\alpha + x_j$ e escala $t_j/g_{t_j}(r)$. ■

Corolário 20.

$$\mathrm{E}[\lambda_j^w|\boldsymbol{x},\alpha,r] = \left(\frac{g_{t_j}(r)}{t_j}\right)^w \frac{\Gamma(\alpha+x_j+w)}{\Gamma(\alpha+x_j)}.$$

6.1 Existência de momentos *a posteriori* de λ_j

O interesse dessa seção recai sobre a existência dos momentos *a posteriori* de λ_j, isto é, sobre

$$\mathrm{E}[\lambda_j^w|\boldsymbol{x}],$$

onde $w > 0$ é considerado fixo. Com efeito, deseja-se saber sob que condições é convergente a integral

$$\int_0^\infty \lambda_j^w p(\lambda_j|\boldsymbol{x})\, d\lambda_j. \tag{38}$$

Essa integral pode ser escrita como

$$\int_0^\infty \lambda_j^w p(\lambda_j|\boldsymbol{x})\, d\lambda_j = \int_0^\infty \lambda_j^w \int_0^\infty \int_0^1 p(\lambda_j,\alpha,r|\boldsymbol{x})\, dr\, d\alpha\, d\lambda_j =$$

$$= \int_0^\infty \int_0^1 \int_0^\infty \lambda_j^w p(\lambda_j,\alpha,r|\boldsymbol{x})\, d\lambda_j\, dr\, d\alpha.$$

A integral interna, por sua vez, pode ser expressa como

$$\int_0^\infty \lambda_j^w p(\lambda_j,\alpha,r|\boldsymbol{x})\, d\lambda_j = p(\alpha,r|\boldsymbol{x}) \int_0^\infty \lambda_j^w p(\lambda_j|\boldsymbol{x},\alpha,r)\, d\lambda_j.$$

Aplicando o Corolário 20, conclui-se que a convergência dos momentos pode ser analisada segundo a integrabilidade da expressão

$$I(\alpha,r) = \left(\frac{g_{t_j}(r)}{t_j}\right)^w \frac{\Gamma(\alpha+x_j+w)}{\Gamma(\alpha+x_j)} p(\alpha,r|\boldsymbol{x}),$$

em relação a α e r.

Teorema 21. *Se a densidade a priori pertence à família* (18)*, é suficiente que a densidade a posteriori seja própria para que os momentos* $\mathrm{E}[\lambda_j^w|\boldsymbol{x}]$ *sejam todos finitos.*

Demonstração. A expressão $I(\alpha, r)$ pode ser escrita como

$$I(\alpha, r) \propto p(\boldsymbol{x}|\alpha, r)p(\alpha, r)r^w \left(\frac{g_{t_j}(r)}{rt_j}\right)^w W(\alpha),$$

onde

$$W(\alpha) = \frac{\Gamma(\alpha + x_j + w)}{\Gamma(\alpha + x_j)}.$$

Definindo

$$q(\alpha, r) = p(\alpha, r)r^w \left(\frac{g_{t_j}(r)}{rt_j}\right)^w W(\alpha),$$

e aplicando o Lema 10, segue que a expressão $W(\alpha)$ apresenta o seguinte comportamento assintótico:

$$W(\alpha) \sim \begin{cases} \alpha^{1-\min(x_j,1)} & \text{em } \alpha \to 0 \quad \text{e} \\ \alpha^w & \text{em } \alpha \to \infty. \end{cases}$$

Como o termo $g_{t_j}(r)/(rt_j)$ é limitado por constantes positivas, concluímos que $q(\alpha, r)$ também pertence à família (18), pois tem a forma

$$q(\alpha, r) = r^{a+w}(1-r)^b f_w(\alpha, r),$$

onde

$$f_w(\alpha, r) = f(\alpha, r) \left(\frac{g_{t_j}(r)}{rt_j}\right)^w W(\alpha).$$

Assim, a integrabilidade de $I(\alpha, r)$ pode ser avaliada aplicando diretamente o Teorema 14, com as seguintes alterações nos parâmetros:

1. $a \longrightarrow a + w$,
2. $u \longrightarrow u + 1 - \min(x_j, 1)$,
3. $v \longrightarrow v + w$,

mantendo os demais inalterados. Como a densidade $p(\alpha, r|\boldsymbol{x})$ é própria por suposição, valem as seguintes condições (35): $a > -1 - s$, $u > b^* - 1 - n'$ e

$v < a$. Como $w > 0$, segue da primeira condição que $a + w > a > -1 - s$. Segue também da última que $v + w < a + w$. Como $1 - \min(x_j, 1) \geq 0$, segue ainda que $u + 1 - \min(x_j, 1) \geq u > b^* - 1 - n'$. Conclui-se que a integral de $I(\alpha, r)$ é convergente. ■

Nota: em alguns casos pode ocorrer que a integral de $I(\alpha, r)$ seja convergente mesmo quando $p(\alpha, r|\boldsymbol{x})$ é imprória. Isso acontece, por exemplo, com a densidade *a priori* de Jeffreys quando $n' = 0$. Nesses casos, embora convergente, a integral de $I(\alpha, r)$ não pode ser interpretada estatísticamente como um momento, visto que $\lambda_j|\boldsymbol{x}$ não possui distribuição própria.

6.2 Existência de momentos preditivos das variáveis latentes

Considere uma variável $\tilde{\lambda} \sim \text{Gama}(\alpha, (1-r)/r)$, condicionalmente independente de λ_j $(j = 1, ..., n)$ dados α e r. Essa variável representa o parâmetro latente de uma observação futura, caso venha a realizar-se. Nessa seção obteremos condições suficientes para a existência dos momentos *a posteriori* de $\tilde{\lambda}$, dados por

$$\text{E}[\tilde{\lambda}^w|\boldsymbol{x}] = \int_0^\infty \tilde{\lambda}^w p(\tilde{\lambda}|\boldsymbol{x})\, d\tilde{\lambda},$$

fixado $w > 0$.

Teorema 22. *Sob uma densidade a priori pertencente à família* (18) *tal que a correspondente densidade a posteriori é própria, os momentos* $\text{E}[\tilde{\lambda}^w|\boldsymbol{x}]$ *existem se e somente se* $w \leq b + 1$.

Demonstração. Analogamente à seção anterior, podemos escrever

$$\text{E}[\tilde{\lambda}^w|\boldsymbol{x}] = \int_0^\infty \int_0^1 \text{E}[\tilde{\lambda}^w|\boldsymbol{x}, \alpha, r] p(\alpha, r|\boldsymbol{x})\, dr\, d\alpha.$$

Como $p(\tilde{\lambda}|\boldsymbol{x}, \alpha, r) = p(\tilde{\lambda}|\alpha, r)$, e

$$\text{E}[\tilde{\lambda}^w|\alpha, r] = \left(\frac{r}{1-r}\right)^w \frac{\Gamma(\alpha + w)}{\Gamma(\alpha)},$$

a existência dos momentos preditivos pode ser avaliada pela integrabilidade da expressão

$$\tilde{I}(\alpha, r) = r^w(1-r)^{-w}\tilde{W}(\alpha)p(\alpha, r|\boldsymbol{x}),$$

onde

$$\tilde{W}(\alpha) = \frac{\Gamma(\alpha + w)}{\Gamma(\alpha)}.$$

O comportamento assintótico de $\tilde{W}(\alpha)$ é dado pelo Lema 10:

$$\tilde{W}(\alpha) \sim \begin{cases} \alpha & \text{em } \alpha \to 0 \quad \text{e} \\ \alpha^w & \text{em } \alpha \to \infty. \end{cases}$$

Seguindo o argumento do Teorema 21, a integrabilidade de $\tilde{I}(\alpha, r)$ pode ser avaliada aplicando diretamente o Teorema 14, com as seguintes alterações nos parâmetros:

1. $a \longrightarrow a + w$,
2. $b \longrightarrow b - w$,
3. $u \longrightarrow u + 1$,
4. $v \longrightarrow v + w$,

mantendo ε e δ inalterados.

As condições em (35) permanecem válidas aplicando as alterações 1, 3 e 4 acima, uma vez que $w > 0$. Além disso, se vale $w \leq b + 1$, então a segunda alteração acima também mantém as condições de integrabilidade válidas, e a integral de $\tilde{I}(\alpha, r)$ converge. Por outro lado, se $w > b + 1$, temos $b - w < -1$, e $\tilde{I}(\alpha, r)$ não é integrável. ■

Corolário 23. *Se $b = -1$, não existem momentos finitos de nenhuma ordem para $\tilde{\lambda}|\boldsymbol{x}$. Em particular, a densidade a posteriori de Jeffreys não possui momentos finitos para $\tilde{\lambda}|\boldsymbol{x}$.*

7 Conclusão

No presente trabalho, foi estudado o modelo hierárquico Poisson-Gama associado a uma família (18) de densidades *a priori* que inclui as densidades de Jeffreys e outras utilizadas na literatura.

A definição dessa família e os resultados que estabelecem condições suficientes e necessárias para a integrabilidade da correspondente distribuição *a posteriori* (Teorema 14) e para a existência de momentos das variáveis latentes (Teoremas 21 e 22) são contribuições originais.

Os principais corolários são as condições intuitivas para a integrabilidade da distribuição *a posteriori* associada à metodologia de Jeffreys e o resultado que afirma que a densidade *a priori* de independência de Jeffreys acarreta uma densidade *a posteriori* imprópria *sempre*.

Uma possível linha de pesquisa que pode se seguida a partir dos resultados aqui obtidos consiste em estudar os algoritmos MCMC quando (i) a densidade *a posteriori* é imprópria e (ii) a densidade *a posteriori* é própria mas os momentos de λ_j não existem - objetivos que não foram alcançados no presente trabalho devido a limitação de tempo.

Anexo - Código em “R”

O código apresentado a seguir permite extrair amostras das distribuições *a posteriori* associada às quatro densidades *a priori* estudadas como casos especiais. Além disso, são mostrados exemplos de uso das rotinas com base em dois conjuntos de dados encontrados na literatura e outro gerado aleatoriamente. O código pode ser facilmente adaptado para uma densidade *a priori* arbitrária.

```
require("compiler")
require("adaptMCMC")

# A função abaixo deve ser compilada para melhorar o desempenho.
# Calculo da função hipergeométrica generalizada com parametros
# a=(1,1,1) e b=(b1,b2)
# Essa função NÃO é vetorizada.
# Condições de validade: b1>1, b2>1, 0<r<1
PFQ <- cmpfun(function(b1, b2, r, eps=1E-6, ret_n=FALSE)
{
    if( length(b1)>1 || length(b2)>1 || length(r)>1 )
        stop("argumentos devem ter tamanho um.")
    dr <- (1/b1)*(1/b2)*r   #d_0 r
    s <- 1
    S <- 1
    n <- 1
    repeat
    {
        s <- s * dr   # s_n
        S <- S + s
        dr <- ((n+1)/(b1+n))*((n+1)/(b2+n))*r   # d_n r
        LB <- s/(1-dr)
        UB <- s/(1-r)
        if( (UB-LB) < 2*eps ) break
        n <- n + 1
    }
    val <- S + (0.5*LB + 0.5*UB)
    if( ret_n ) c( val=val, n=n ) else val
})

##############################################################
# Transformação em termos de covariáveis
# Função vetorizada.
```

```
g <- function(tj, r)
{
    (tj*r)/(1+(tj-1)*r)
}

# Definição das densidades a priori utilizadas. TODAS retornam
# o log. Estas funções NÃO checam os argumentos. 'a' e 'r'
# devem ter tamanho 1.
logpJeffreys <- function(a, r, tj, indep=FALSE)
{
    rj <- g(tj, r)
    Fj <- Vectorize(PFQ, "r")(2, 1 + a, rj)
    -log(r) -log(1-r) + log(sum(rj))/2 +
        log(sum(rj * (Fj - ifelse(indep,0,1))))/2
}

logpAlphaBeta <- function(a, r, k1, k2)
{
    k1*log(a) + k2*log(r) - (k2+2)*log(1-r)
}

logpHadjicostasBerry <- function(a, r, k1, k2, k3, k4, s1, s2)
{
    k1*log(a) + k2*log(a+s1) + k3*log(r) - (k3+2+k4)*log(1-r) +
        k4*log(r+(1-r)*s2)
}

allReal <- function(...)
{
    all(unlist(lapply(list(...), function(x) { is.numeric(x) &&
        length(x)==1 && is.finite(x) } )))
}

checkAlphaBeta <- function(k1, k2)
{
    if( ! allReal(k1,k2) )
        stop("Parâmetros inválidos para a priori AlphaBeta.")
}

checkHadjicostasBerry <- function(k1, k2, k3, k4, s1, s2)
{
    if( ! allReal(k1,k2,k3,k4,s1,s2) )
        stop(paste("Parâmetros inválidos para a priori de",
            "Hadjicostas-Berry."))
    if( (s1<=0) || (s2<=0) )
```

```
        stop(paste("Os parâmetros 's1' e 's2' para a priori",
            "de Hadjicostas-Berry devem ser positivos."))
}

# Calcula o log da densidade a priori com checagem dos
# argumentos e vetorizacao:
logprior <- function(a, r, tj, type=c("Jeffreys", "IndJeffreys",
    "AlphaBeta", "HadjicostasBerry"), ...)
{
    if( ! all(is.finite(c(a, r, tj))) )
        stop("Argumentos com valores inválidos.")
    if( any(tj <= 0) ) stop("As covariáveis devem ser positivas.")
    if( length(a) != length(r) )
        stop("Os argumentos 'a' e 'r' devem ter o mesmo tamanho.")
    ok <- (a > 0) & (0 < r) & (r < 1)
    result <- numeric(length(ok))
    result[!ok] <- -Inf
    type <- match.arg(type)
    if(type=="AlphaBeta") checkAlphaBeta(...)
    if(type=="HadjicostasBerry") checkHadjicostasBerry(...)
    f <- function(i) switch(type,
        Jeffreys = logpJeffreys(a[i], r[i], tj, indep=FALSE),
        IndJeffreys = logpJeffreys(a[i], r[i], tj, indep=TRUE),
        AlphaBeta = logpAlphaBeta(a[i], r[i], ...),
        HadjicostasBerry = logpHadjicostasBerry(a[i], r[i], ...)
    )
    result[ok] <- vapply(which(ok), f, numeric(1))
    result
}

# Log da verossimilhança:
# NÃO checa os argumentos. 'a' e 'r' devem ter tamanho 1.
# Versão que utiliza o pacote 'stats':
loglik <- function(a, r, tj, xj)
{
    sum(dnbinom(x=xj, size=a, prob=1-g(tj,r), log=TRUE))
}
# Versão "interna"
loglik2 <- function(a, r, tj, xj)
{
    rj <- g(tj, r)
    sum(lgamma(a+xj) - lgamma(a) - lgamma(xj+1) +
        xj*log(rj) + a*log(1-rj))
}
```

```
# Calcula a posteriori, com validação e vetorização.
# a = alpha
# r = r
# tj = vetor de covariáveis
# xj = vetor de observações
# logp = função que retorna o log da priori. São repassados os
# argumentos 'a', 'r', 'tj' e '...'.
# Assume-se que 'logp' é vetorizada em 'a' e 'r' e realiza a
# validação de todos os argumentos,
# gerando erro se algum for inválido e retornando '-Inf' para
# pontos fora do suporte da densidade.
logpost <- function(a, r, tj, xj, logp=logprior, ...)
{
    xj <- as.integer(xj)
    if( any(is.na(xj)) )
        stop("Observações inválidas.")
    if( any(xj < 0) )
        stop("As observações não podem ser negativas.")
    if( length(xj) != length(tj) )
        stop(paste("O número de observações deve ser igual",
            "ao de covariáveis."))
    result <- logp(a, r, tj, ...)
    ok <- is.finite(result)
    f <- function(i) loglik(a[i], r[i], tj, xj)
    result[ok] <- result[ok] + vapply(which(ok), f, numeric(1))
    result
}

# Reparametrizacao: u = log(a), v = log(r/(1-r))
# Gera a função que retorna a mudança de coordenadas
# e o determinante jacobiano (em valor absoluto).
changevar <- function(a, r, u="u", v="v")
{
    expr_a <- substitute(a)
    expr_r <- substitute(r)
    L <- list(a=expr_a, r=expr_r, dadu=D(expr_a, u),
        drdv=D(expr_r, v), drdu=D(expr_r, u), dadv=D(expr_a, v))
    f <- eval(substitute(function(u, v)list(a=a, r=r,
        J=abs(dadu*drdv - drdu*dadv)), L))
    attributes(f) <- NULL
    environment(f) <- parent.frame()
    f
}

# Simulador:
```

```
posteriorMCMC <- function(tj, xj, n=100, burnin=100, skip=10,
    plot=TRUE, FUN=colMeans, ...)
{
    phi <- changevar(a=exp(u), r=1/(exp(-v)+1))
    logdens <- function(pars)
    {
        temp <- phi(pars[1], pars[2])
        logpost(temp$a, temp$r, tj, xj, ...) + log(temp$J)
    }
    N <- n*skip+burnin
    getMCMC <- MCMC(logdens, n=N, init=c(u=0,v=0), acc.rate=0.5,
        list=FALSE)
    temp <- phi(getMCMC[,"u"], getMCMC[,"v"])
    result <- cbind(a=temp$a, r=temp$r)
    if(plot) plot(coda::mcmc(result))
    rows <- 1:N
    FUN(result[(rows > burnin) & (rows %% skip == 0),])
}

###############################################################

# Exemplos de uso:

# Bombas d'água:
pump_tj <- c(94.32,15.72,62.88,125.76,5.24,31.44,1.048,1.048,
    2.096,10.48)
pump_xj <- c(5,1,5,14,3,19,1,1,4,22)
# Dados aleatórios (amostra com n=1000, a=7, r=0.75):
random_tj <- runif(1000, 0, 10)
random_xj <- rpois(1000, rgamma(1000, shape=7, scale=3)*random_tj)
# Dados de McFarland et al. (1992) - usados em Hadjicostas-Berry:
HB_tj <- c(18,18,19,18,20,19,13,12,10,13)
HB_xj <- c(39,3,21,65,48,58,46,9,59,29)

# Dados aleatorios:

# Jeffreys completa:
posteriorMCMC(random_tj, random_xj, n=50, burnin=100, skip=5)
# Jeffreys de independencia:
posteriorMCMC(random_tj, random_xj, n=50, burnin=100, skip=5,
    type="IndJeffreys")
# Hadjicostas-Berry:
posteriorMCMC(random_tj, random_xj, n=200, burnin=200, skip=20,
    type="H", k1=-1, k2=-1, k3=-1, k4=-1, s1=1, s2=1)
# AlphaBeta
```

```
posteriorMCMC(random_tj, random_xj, n=200, burnin=200, skip=20,
    type="A", k1=-1, k2=-1)

# Dados das bombas:

# Jeffreys completa:
posteriorMCMC(pump_tj, pump_xj, n=100, burnin=200, skip=10)
# Jeffreys de independencia:
posteriorMCMC(pump_tj, pump_xj, n=100, burnin=200, skip=10,
    type="IndJeffreys")
# Hadjicostas-Berry:
posteriorMCMC(pump_tj, pump_xj, n=200, burnin=200, skip=20,
    type="H", k1=-1, k2=-1, k3=-1, k4=-1, s1=1, s2=1)
# AlphaBeta
posteriorMCMC(pump_tj, pump_xj, n=200, burnin=200, skip=20,
    type="A", k1=-1, k2=-1)

# Dados de McFarland et al:

# Jeffreys completa:
posteriorMCMC(HB_tj, HB_xj, n=100, burnin=200, skip=10)
# Jeffreys de independencia:
posteriorMCMC(HB_tj, HB_xj, n=100, burnin=200, skip=10,
    type="IndJeffreys")
# Hadjicostas-Berry:
posteriorMCMC(HB_tj, HB_xj, n=200, burnin=200, skip=20,
    type="H", k1=-1, k2=-1, k3=-1, k4=-1, s1=1, s2=1)
# AlphaBeta
posteriorMCMC(HB_tj, HB_xj, n=200, burnin=200, skip=20,
    type="A", k1=-1, k2=-1)
```

Referências

[1] Abramowitz, Milton; Stegun, Irene A. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, New York: Dover Publications, ISBN 978-0-486-61272-0.

[2] Bowman, K.O.; Shenton, L.R. (2007a). Skewness for maximum likelihood estimators of the negative binomial distribution. *Far East Journal of Theoretical Statistics*, 22(1),103-129.

[3] Datta; Sweeting (2005). Probability Matching Priors - Research Report 252, Department of Statistical Science, University College London.

[4] NIST (2012). *Digital Library of Mathematical Functions*. Versão 1.0.5. http://dlmf.nist.gov/.

[5] Fienberg, Stephen E. (2006). When did bayesian inference become "bayesian"?. *Bayesian Analysis 1* (1): 1–40.

[6] Fisher, R. A. (1941). The negative binomial distribution. *Annals of Eugenics*, Volume 11, Issue 1, pages 182–187.

[7] Gelman, A.; Carlin, J. B.; Stern, H.; Rubin, D. B. (1995). *Bayesian Data Analysis*. Chapman and Hall.

[8] Haldane, J.B.S. (1948). The precision of observed values of small frequencies. *Biometrika* Volume 35, Issue 3-4 Pp. 297-300.

[9] Hartigan, J. (1964). Invariant prior distributions. *Ann. Math. Statist.* Volume 35, Number 2, p. 836-845.

[10] James, B. (2004). *Probabilidade: Um curso em nível intermediário.* Publicação: IMPA, ISBN: 978-85-244-0101-5.

[11] Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proc. R. Soc. Lond. A* September 24, 186 1007 453-461.

[12] Hadjicostas, P.; Berry, S. M. (1999). Improper and proper posteriors with improper priors in a Poisson-gamma hierarchical model. *TEST* Volume 8, Number 1, 147-166, DOI: 10.1007/BF02595867.

[13] Vihola, M. (2012). Robust adaptive Metropolis algorithm with coerced acceptance rate. *Statistics and Computing*, September 2012, Volume 22, Issue 5, pp 997-1008.