Estimação bayesiana dos parâmetros da distribuição multinomial: uma aplicação para atuação profissional de alunos do Profmat da UFSJ

Isabela da Silva Lima isabela_lima30@hotmail.com
Universidade Federal de Lavras, Lavras, UFLA, MG, Brazil

Carla Regina Guimarães Brighenti carlabrighenti@ufsj.edu.br
Universidade Federal de São João del-Rei, São João del-Rei, UFSJ, MG, Brazil

Resumo

Um dos problemas básicos da inferência estatística é a estimação de parâmetros, e uma forma de realizá-la é através da abordagem bayesiana, que através da incorporação de *prioris* pode-se melhorar a estimação de parâmetros. Com o objetivo de estimar a proporção da atuação profissional de alunos do Mestrado Profissional em Matemática em Rede Nacional-Profmat, da Universidade Federal de São João del-Rei (UFSJ), esse trabalho utilizou a abordagem bayesiana, onde gerou-se três *prioris* e estimou-se os parâmetros da distribuição multinomial para comparação, utilizando dados da secretaria do programa da UFSJ. Além disso, a partir do *software* R, plotou-se o *simplex* de cada distribuição *a priori* e *a posteriori* para a visualização dos parâmetros. Concluiu-se que as estimativas foram precisas utilizando a abordagem bayesiana, e também foi possível comparar a influência de cada *priori* na estimação dos parâmetros.

Palavras-chave

Dirichlet, *Simplex*, Docentes.

1 Introdução

A inferência é uma área de grande relevância dentro da Estatística, que segundo Bussab e Morettin em [11] tem o objetivo de fazer generalizações sobre uma população, com base nos dados contidos em uma amostra. Um dos problemas básicos da inferência é a estimação de parâmetros, e uma forma de realizá-la é através da abordagem bayesiana.

A inferência bayesiana é uma alternativa importante em relação aos procedimentos clássicos de estimação, e esta abordagem vem sendo utilizada com sucesso em várias áreas da ciência. Ela permite a incorporação de uma informação *a priori* sobre o parâmetro que se quer estimar e toda sua teoria está baseada no teorema de Bayes, veja [12].

Dentre as estimativas de parâmetros, tem-se a do parâmetro p sobre proporções, que é útil na identificação e mensuração de elementos portadores de uma certa característica, como por exemplo, na mensuração de proporções de peças defeituosas e não defeituosa, de indivíduos pertencerem a uma determinada classe social, entre outros. Desse modo, pode-se ainda observar que a estimação de proporções ambienta-se em questões de dois tipos: dicotômicas e politômicas, como por ser visto em [4].

Problemas de caráter dicotômico, são aqueles que possuem duas opções de resposta, geralmente representadas pelo binômio sim/não, como por exemplo, ao mensurar a quantidade de peças defeituosas tem-se como resposta sim que corresponde à peça defeituosa ou não correspondente a não defeituosa. Para esse tipo de problema é suficiente utilizar a distribuição binomial para a análise dos dados, e esta possui teoria bem desenvolvida na literatura.

Já as questões politômicas são compostas de mais de duas categorias ou classes de resposta, como por exemplo, ao mensurar a raça de alunos, tem-se brancos, pardos, negros, etc. Assim, deve-se considerar a distribuição multinomial para a análise. No entanto, esta não conta com literatura tão acessível para melhor compreensão, pois é menos utilizada.

Desse modo, um exemplo onde a distribuição multinomial é adequada é na análise de dados da atuação profissional de alunos ingressantes no Mestrado Profissional em Matemática em Rede Nacional-Profmat, pois pode-se ter ingressantes que são professores da rede pública de ensino, e essas se dividem em escolas estaduais, municipais e estatais, professores da rede particular, ou ainda, que possuem outras profissões. Nesse caso, quando o interesse é estimar as proporções de cada uma dessas categorias utiliza-se a distribuição multinomial e pode-se fazê-la através da inferência bayesiana.

O Mestrado Profissional em Matemática em Rede Nacional-Profmat é um programa de mestrado semipresencial na área de Matemática, formado por uma rede de instituições de Ensino Superior, no contexto da Universidade Aberta do Brasil/Coordenação e Aperfeiçoamento Pessoal de Nível Superior (CAPES), e coordenado pela Sociedade Brasileira de Matemática (SBM), com apoio do Instituto Nacional de Matemática Pura e Aplicada (IMPA). Surgiu em 2011, como um programa integrado por 48 instituições, com oferta de 1192 vagas em 54 *campi* , de acordo com [2]. Atualmente, conta com 1400 vagas em 97 *campi*.

Neste sentido, é importante analisar o perfil de ingressantes nesse programa, dessa forma, este trabalho teve como objetivo estimar a proporção da atuação profissional de alunos ingressantes no Profmat da Universidade Federal de São João del-Rei (UFSJ), sob três categorias: professores de escolas estaduais, professores de outras escolas e atuantes em outras profissões, utilizando a distribuição multinomial, a partir da inferência bayesiana, com dados obtidos da secretaria do programa Profmat da UFSJ.

2 Referencial Teórico

2.1 Distribuição Multinomial

A distribuição multinomial, é uma distribuição discreta multivariada, sendo uma generalização da distribuição binomial, utilizada em problemas que possuem mais de duas categorias de resposta. De acordo com [3], a distribuição multinomial é definida supondo um experimento cujo resultado seja um dos eventos $E_1, E_2, ..., En$, com probabilidade $P[E_i] = p_i$. Para $i = 1, 2, ..., n$, $0 \leq p_i \leq 1$ e $\sum_{i=1}^{n} p_i = 1$, e seja X_i uma variável aleatória que conta o número de ocorrências de E_i em m repetições independentes desse experimento. Então, o vetor aleatório $(X_1, X_2, ..., X_n)$ tem distribuição chamada **multinomial**, com parâmetros $p_1, p_2, ..., p_{n-1}$ e m, dada por:

$$f_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n) = P[X_1 = x_1, X_2 = x_2, ..., X_n = x_n] =$$

$$= \frac{m!}{x_1!x_2! \cdot ... \cdot x_n!} p_1^{x_1} p_2^{x_2} \cdot ... \cdot p_n^{x_n} = m! \prod_{i=1}^{n} \frac{p_i^{x_i}}{x_i!}. \tag{1}$$

em que cada X_i é um inteiro positivo, $p_1, p_2, ..., p_n$ são proporções populacionais, $\sum_{i=1}^{n} x_i = m$ e $\sum_{i=1}^{n} p_i = 1$. Assim, escreve-se que:

$$(\boldsymbol{X}|m, \boldsymbol{p}) \sim \text{Multinomial}(m, \boldsymbol{p}).$$

Um resultado importante é que a distribuição marginal de qualquer componente X_i da distribuição multinomial corresponde a uma binomial (m, p_i), em que m é o número de tentativas e p_i é a probabilidade de obtenção de X_i em cada tentativa. Desse modo, tem-se que os equacionamentos dos parâmetros de amostragem (esperança e variância) utilizados para amostragem de variáveis dicotômicas são válidos para a amostragem de variáveis politômicas. Ou seja, a esperança de X_i é mp_i e a sua variância é $mp_i(1-p_i)$, que são equivalentes ao caso binomial, veja Teorema Multinomial de [4].

Ainda, tem-se que dadas duas variáveis aleatórias multinomiais X_i e X_vj, com $i \neq j$. A covariância entre X_i e X_j é dada por, veja [5]:

$$Cov(X_i, X_j) = -mp_ip_j. \tag{2}$$

2.2 Estimação Bayesiana dos parâmetros da distribuição Multinomial

Segundo Reis et al. (2009), em [14], a estatística bayesiana ganhou força ao longo do tempo com a evolução dos computadores em termos de capacidade de processamento e despertou grande interesse principalmente em áreas aplicadas, como uma metodologia alternativa em relação aos procedimentos clássicos de estimação.

Depois da análise dos dados, o propósito de qualquer estatístico é fazer inferências ou predições, com certo grau de confiança, sobre o fenômeno que se está estudando, a partir dos dados que representam a variabilidade ou a incerteza na observação da característica ou fenômeno, veja [7].

Diferentemente da estatística frequentista, em que somente se admite probabilidade através de medidas de frequências relativas, na análise bayesiana entende-se que a probabilidade é uma medida racional e condicional de incerteza. Nesta abordagem, o valor do parâmetro de interesse é desconhecido, o qual é considerado uma variável aleatória, e o intuito é tentar reduzir esse desconhecimento sobre o parâmetro, veja [8].

De acordo com Ehlers (2011), em [6], a principal característica da inferência bayesiana é utilizar probabilidade para quantificar a intensidade da incerteza a respeito do parâmetro de interesse, e esta é uma medida subjetiva, que pode variar de acordo com o pesquisador, pois a experiência e a fonte dessa informação que cada um possui são diferenciadas.

Assim, assume-se uma distribuição de probabilidade associada, conhecida como distribuição *a priori*, a qual é especificada antes da observação dos dados e descreve o conhecimento, ou ainda, o grau da crença do pesquisador sobre o parâmetro desconhecido, e esse conhecimento pode ser formalmente incorporado na análise. Ela pode ser extraída de fundamentos subjetivos, da experiência do pesquisador na área em questão por análises realizadas anteriormente, por considerações particulares ou ainda por informações disponíveis na literatura sobre o assunto estudado, veja [7].

A teoria da inferência Bayesiana está fundamentada no teorema de Bayes, o qual é basicamente um resultado de uma probabilidade condicional, dado por [12]:

$$\pi(\theta|X) = \frac{L(X|\theta)\pi(\theta)}{\int L(X|\theta)\pi(\theta)d\theta}. \tag{3}$$

onde: $\pi(\theta|X)$ é distribuição a *posteriori* do parâmetro θ; $\pi(\theta)$ é a distribuição a *priori* de θ; $L(X|\theta)$ é a função de verossimilhança de θ, que é uma distribuição conjunta para os dados amostrais, a qual representa a informação sobre θ que foi obtida dos dados.

Assim, a partir do Teorema de Bayes, a distribuição *a priori* é combinada com a informação contida nos dados amostrais, obtendo assim a distribuição *a posteriori*. Portanto, toda a inferência relativa a um determinado parâmetro é realizada utilizando-se a distribuição *a posteriori*, pois ela contém toda a informação probabilística a respeito do parâmetro, e esta pode ser resumida por meio da média, da moda, da mediana e do intervalo de credibilidade, com base em Filho et al. (2008), em [10].

No entanto, segundo Paulino et al. (2018), em [12], especificar distribuições a *priori* quando a sua obtenção e quantificação são de natureza essencialmente subjetiva, e o intuito é transformá-la em uma informação que possa ser utilizada, muitas vezes é uma tarefa difícil. Essas dificuldades costumam ser contornadas através da adoção de uma forma distribucional conveniente denominada família das distribuições conjugadas.

Segundo Ehlers em [6], uma *priori* é dita conjugada se as distribuições a *priori* e a *posteriori* pertencem a mesma classe de distribuições. Tem-se que *prioris* conjugadas são os casos mais importantes ao utilizar uma abordagem com hiperparâmetros, que em geral, facilita a análise e consiste em definir uma família paramétrica de densidades.

Nesta abordagem, a distribuição a *priori* é representada por uma forma funcional, cujos parâmetros são especificados de acordo com o conhecimento que se tem sobre o parâmetro de interesse. Desse modo, os parâmetros indexadores da família de distribuições a *priori* recebem o nome de hiperparâmetros para distingui-los dos parâmetros de interesse θ, e a ideia é que a atualização do conhecimento que se tem de θ envolva apenas uma mudança nos hiperparâmetros, veja [6].

No caso da distribuição multinomial sua distribuição *a priori* conjugada é a distribuição de Dirichlet, que é a generalização multivariada da distribuição beta, com um parâmetro vetorial não-negativo e real $\boldsymbol{\alpha}$, dada por: [12].

$$f_{\boldsymbol{X}}(x_1, ..., x_n) = \frac{\Gamma(\alpha_0)}{\prod_{i=1}^{n} \Gamma(\alpha_i)} \prod_{i=1}^{n} p_i^{\alpha_i - 1},\ 0 < p_i < 1, \tag{4}$$

sendo $\alpha_0 = \displaystyle\sum_{i=1}^{n} \alpha_i$, $\Gamma(t) = \displaystyle\int_0^{\infty} x^{t-1} e^{-x} dx$ a função gama e $\displaystyle\sum_{i=1}^{n} p_i = 1$. A distribuição marginal é uma beta com parâmetros α_i e $(\alpha_0 - \alpha_i)$, para cada i, de onde tem-se:

$$E(X_i) = \frac{\alpha_i}{\alpha_0}, \qquad Var(X_i) = \frac{\alpha_i(\alpha_0 - \alpha_i)}{\alpha_0^2(\alpha_0 + 1)}, \qquad Cov(X_i, X_j) = \frac{-\alpha_i \alpha_j}{\alpha_0^2(\alpha_0 + 1)}. \tag{5}$$

Em (4) tem-se um vetor n-dimensional, em que os elementos são interpretados como as probabilidades de ter as proporções $(p_1, ..., p_n)$.

Portanto, o modelo Multinomial-Dirichlet é dado pela combinação do modelo probabilístico multinomial com uma distribuição *a priori* de Dirichlet, generalizando assim os resultados obtidos com o modelo binomial e a distribuição *a priori* beta [9]. E possui a seguinte estrutura hierárquica:

$$(X_1, ..., X_n | m, \boldsymbol{p}) \sim \text{Multinomial}(m, p_1, ..., p_n).$$

$$(p_1, ..., p_n|\boldsymbol{\alpha}) \sim \text{Dirichlet}(\alpha_1, ..., \alpha_n).$$

com função de probabilidade e densidade conjunta definidas em (1) e (4), respectivamente. Ou seja, os vetores de probabilidade $\boldsymbol{p}$, parâmetros da distribuição Multinomial, seguem uma distribuição de Dirichlet com parâmetros $\boldsymbol{\alpha}$.

Sendo assim, se a distribuição *a priori* é uma Dirichlet e a variável observada segue uma distribuição Multinomial, então a distribuição *a posteriori* será uma distribuição de Dirichlet, com outro parâmetro. A distribuição *a posteriori* é encontrada utilizando o teorema de Bayes (3), segundo [1], como segue:

$$\pi(\boldsymbol{p}|X) = \frac{\left(\dfrac{m!}{x_1! \cdot ... \cdot x_n!}\right) \prod_{i=1}^n p_i^{x_i} \dfrac{\Gamma\left(\sum_{i=1}^n \alpha_i\right)}{\prod_{i=1}^n \Gamma(\alpha_i)} \prod_{i=1}^n p_i^{\alpha_i - 1}}{\int_{\boldsymbol{p}} \left(\dfrac{m!}{x_1! \cdot ... \cdot x_n!}\right) \prod_{i=1}^n p_i^{x_i} \dfrac{\Gamma\left(\sum_{i=1}^n \alpha_i\right)}{\prod_{i=1}^n \Gamma(\alpha_i)} \prod_{i=1}^n p_i^{\alpha_i - 1} d\boldsymbol{p}}$$

$$\pi(\boldsymbol{p}|X) \propto \left(\frac{m!}{x_1! \cdot ... \cdot x_n!}\right) \prod_{i=1}^n p_i^{x_i} \frac{\Gamma\left(\sum_{i=1}^n \alpha_i\right)}{\prod_{i=1}^n \Gamma(\alpha_i)} \prod_{i=1}^n p_i^{\alpha_i - 1}$$

$$\pi(\boldsymbol{p}|X) \propto \prod_{i=1}^n p_i^{x_i + \alpha_i - 1}.$$

Portanto, *a posteriori* segue uma distribuição de Dirichlet com parâmetros:

$$\boldsymbol{p}|X \sim \text{Dirichlet}(\alpha_1^* = x_1 + \alpha_1, ..., \alpha_n^* = x_n + \alpha_n).$$

sendo $X = (x_1, ..., x_n, \alpha_1, ..., \alpha_n)^T$.

Assim, a média, a variância e a covariância da distribuição *a posteriori* de Dirichlet são dadas respectivamente por:

$$E(\boldsymbol{p}|X_i) = \frac{x_i + \alpha_i}{\sum_{i=1}^n (x_i + \alpha_i)}, \tag{6}$$

$$Var(\boldsymbol{p}|X_i) = \frac{(x_i + \alpha_i)\{[\sum_{i=1}^n (x_i + \alpha_i)] - (x_i + \alpha_i)\}}{[\sum_{i=1}^n (x_i + \alpha_i)]^2 \{[\sum_{i=1}^n (x_i + \alpha_i)] + 1\}}, \tag{7}$$

$$Cov(\boldsymbol{p}|X_i, X_j) = \frac{-(x_i + \alpha_i)(x_j + \alpha_j)}{[\sum_{i=1}^n (x_i + \alpha_i)]^2 \{[\sum_{i=1}^n (x_i + \alpha_i)] + 1\}}. \tag{8}$$

3 Metodologia

Os dados das atuações profissionais de alunos do Profmat da Universidade Federal de São João del-Rei, do polo de São João del-Rei, foram obtidos a partir da secretaria do programa, onde havia o registro de 56 alunos, dos quais 26 são professores de escolas estaduais de ensino, 24 professores de outras escolas e 6 possuem outra profissão.

Assim, foi realizada a estimação da proporção de alunos sob essas três categorias através da distribuição multinomial utilizando a inferência bayesiana. Logo, ao utilizar a distribuição multinomial a distribuição a *priori* conjugada será a distribuição de Dirichlet.

Foram avaliadas três *prioris*: **1)** *a priori* uniforme, onde admite-se que todos os possíveis valores do parâmetro de interesse são igualmente prováveis. **2)** *Priori* de Dirichlet dada pelos parâmetros $(\alpha_1 = 0,32; \alpha_2 = 0,32; \alpha_3 = 0,16)$, considerando a informação de que a preferência de alunos para esse programa são para os atuantes em sala de aula, logo, essa *priori* incorpora essa informação no sentido que os valores dos parâmetros para as categorias de professores são maiores e iguais. E por fim, **3)** *Priori* de Dirichlet dada pelos parâmetros $(\alpha_1 = 0,0625; \alpha_2 = 0,20535; \alpha_3 = 0,23315)$, onde esta tem-se o maior valor do parâmetro para a categoria de outras profissões, sendo contrária a informação presente para avaliar a sua influência nas estimativas finais.

Logo, por intermédio do teorema de Bayes chegou-se na distribuição *a posteriori* para **p**, onde toda a inferência foi realizada a partir da média e variância da distribuição *a posteriori*, dadas por:

$$E(\boldsymbol{p}|X_i) = \frac{x_i + \alpha_i}{\sum_{i=1}^{n}(x_i + \alpha_i)}, \tag{9}$$

$$Var(\boldsymbol{p}|X_i) = \frac{(x_i + \alpha_i)\{[\sum_{i=1}^{n}(x_i + \alpha_i)] - (x_i + \alpha_i)\}}{[\sum_{i=1}^{n}(x_i + \alpha_i)]^2\{[\sum_{i=1}^{n}(x_i + \alpha_i)] + 1\}}, \tag{10}$$

Foram plotados *simplex* de dimensão 3 para todas as *prioris* e *posterioris*, utilizando o *software* R [13], para observar o comportamento que os parâmetros desempenham.

4 Resultados e Discussão

Dos dados da secretaria do Profmat da UFSJ, tem-se que dos 56 alunos do Profmat, 26 deles são professores de escolas estaduais, 24 são professores de outras escolas e 6 possuem outras profissões, tem-se que a proporção frequentista para cada uma das três categorias de atuação profissional estão representadas na Tabela 1:

Tabela 1: Proporção da atuação profissional

Professor escola estadual	Professor outras escolas	Outra profissão
0,4643	0,4286	0,1071

Para a estimação bayesiana, considerou-se três *prioris* para analisar a influência dessas. A primeira delas foi *a priori* uniforme, onde os parâmetros, denotados por hiperparâmetros, da distribuição de Dirichlet, são: $\alpha_1 = 1$, $\alpha_2 = 1$, $\alpha_3 = 1$, em que todos os possíveis valores do parâmetro de interesse são igualmente prováveis. Na Tabela 2 a seguir tem-se os valores dos hiperparâmetros de cada categoria com suas respectivas médias e variâncias *a priori*:

Tabela 2: Hiperparâmetros e valores da média e variância *a priori* de Dirichlet

Atuação profissional	α_i	Média (%)	Variância (%)
Professor escola estadual	1	33,33	5,55
Professor outras escolas	1	33,33	5,55
Outra profissão	1	33,33	5,55

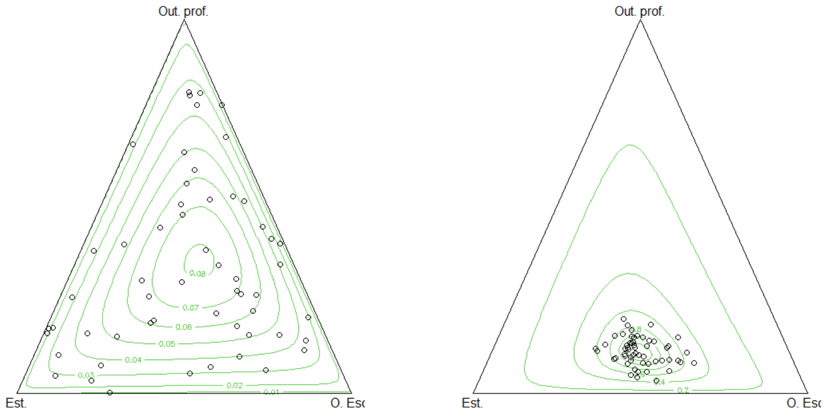
Os resultados da estimação bayesiana considerando *a priori* uniforme para a proporção dos três tipos de atuação profissional, que são dadas pelas médias *a posteriori*, pode ser observada na Tabela 3:

Tabela 3: Estimativas *a posteriori* para proporção da atuação profissional de alunos do Profmat

Atuação profissional	Média	Variância
Professor escola estadual	0,4576	0,0041
Professor outras escolas	0,4237	0,0041
Outra profissão	0,1186	0,0017

Pode-se observar que as estimativas ficaram bem próximas das frequentistas, mas para se ter melhor compreensão do impacto do parâmetro da *priori* uniforme $\alpha_1 = 1$, $\alpha_2 = 1$ e $\alpha_3 = 1$, e o da *posteriori* que foi $\alpha_1^* = 27$, $\alpha_2^* = 25$ e $\alpha_3^* = 7$ é apresentado o *simplex* de dimensão 3, que pode ser observado na Figura 1:

Figura 1: *Simplex* da *priori* uniforme com parâmetros (1; 1; 1) e da *posteriori* com os parâmetros (27; 25; 7)



Fonte: As autoras (2021)

Assim, pode-se observar que para *a priori* uniforme (*simplex* à esquerda) todos os pontos no *simplex* são igualmente prováveis, pois estão espalhados por todo o *simplex*. Já no *simplex* à direita, que representa o da *posteriori*, é possível perceber que os pontos estão entre as categorias de escolas estaduais e outras escolas, o que conclui que as estimativas das proporções entre essas duas categorias são muito próximas.

A segunda *priori* foi dada pelos parâmetros: $\alpha_1 = 0,32$; $\alpha_2 = 0,32$; $\alpha_3 = 0,16$. Na Tabela 4 a seguir tem-se os valores dos hiperparâmetros de cada categoria com suas respectivas médias e variâncias *a priori*:

Tabela 4: Hiperparâmetros e valores da média e variância *a priori* de Dirichlet

Atuação profissional	α_i	**Média (%)**	**Variância (%)**
Professor escola estadual	0,32	40,00	13,33
Professor outras escolas	0,32	40,00	13,33
Outra profissão	0,16	20,00	8,89

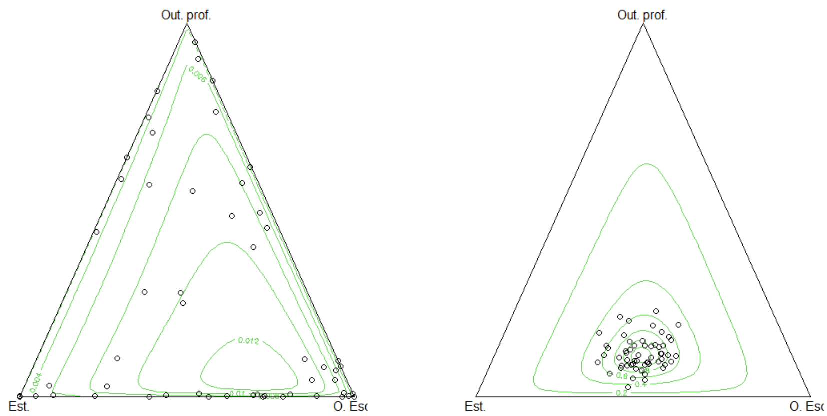
Os resultados da estimação bayesiana considerando a segunda *priori* para a proporção dos três tipos de atuação profissional, que são dadas pelas médias *a posteriori*, pode ser observada na Tabela 5:

Tabela 5: Estimativas *a posteriori* para proporção da atuação profissional de alunos do Profmat

Atuação profissional	**Média**	**Variância**
Professor escola estadual	0,4634	0,0043
Professor outras escolas	0,4282	0,0042
Outra profissão	0,1085	0,0017

Pode-se observar que a partir dessa *priori* as estimativas das proporções foram ainda mais próximas das frequentistas, sendo mais próximas do que utilizando *a priori* uniforme. Foi constuído o *simplex* também para se ter melhor compreensão do impacto do parâmetro da *priori* 2 $\alpha_1 = 0,32$, $\alpha_2 = 0,32$ e $\alpha_3 = 0,16$, e o da *posteriori* que foi $\alpha_1^* = 26,32$, $\alpha_2^* = 24,32$ e $\alpha_3^* = 6,16$, que é apresentado na Figura 2:

Figura 2: *Simplex* da *priori* uniforme com parâmetros $(0,32; 0,32; 0,16)$ e da *posteriori* com os parâmetros $(26,32; 24,32; 6,16)$



Fonte: As autoras (2021)

Assim, pode-se observar que para *a priori* 2 (*simplex* à esquerda) os pontos estão bem espalhados em torno do *simplex*. Já no *simplex* à direita, que representa o da *posteriori*, é possível perceber que os pontos estão entre as categorias também de escola estadual e outras escolas, mas com uma pouco mais de tendência à escola estadual.

A terceira *priori* foi dada pelos parâmetros: $\alpha_1 = 0,0625$; $\alpha_2 = 0,20535$; $\alpha_3 = 0,23215$. Na Tabela 6 a seguir tem-se os valores dos hiperparâmetros de cada categoria com suas respectivas médias e variâncias *a priori*:

Tabela 6: Hiperparâmetros e valores da média e variância *a priori* de Dirichlet

Atuação profissional	α_i	Média (%)	Variância (%)
Professor escola estadual	0,0625	12,50	7,29
Professor outras escolas	0,20535	41,07	16,13
Outra profissão	0,23215	46,43	16,58

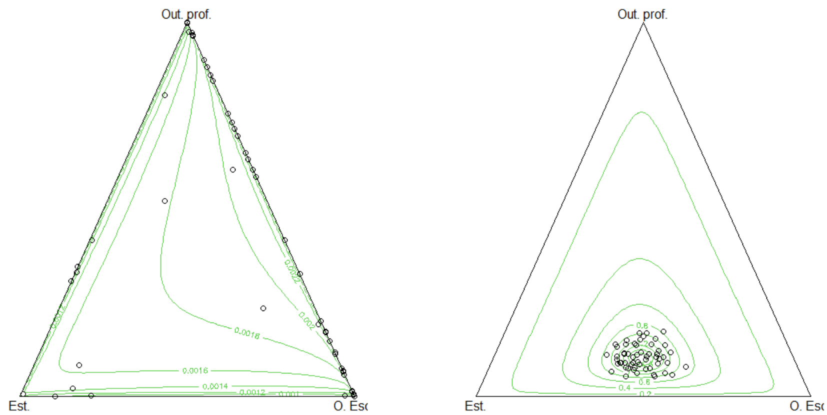
Os resultados da estimação bayesiana considerando a terceira *priori* para a proporção dos três tipos de atuação profissional, que são dadas pelas médias *a posteriori*, pode ser observada na Tabela 7:

Tabela 7: Estimativas *a posteriori* para proporção da atuação profissional de alunos do Profmat

Atuação profissional	Média	Variância
Professor escola estadual	0,4613	0,0043
Professor outras escolas	0,4284	0,0043
Outra profissão	0,1103	0,0017

Conclui-se que também tem-se estimativas muito próximas, apenas para a categoria de outras profissões ficou um pouco mais longe. Logo, para se ter melhor compreensão do impacto do parâmetro da *priori* 3 $\alpha_1 = 0,0625$, $\alpha_2 = 0,20535$ e $\alpha_3 = 0,23215$, e o da *posteriori* que foi $\alpha_1^* = 26,0625$, $\alpha_2^* = 24,20535$ e $\alpha_3^* = 6,23215$ é apresentado o *simplex* de dimensão 3, que pode ser observado na Figura 3:

Figura 3: *Simplex* da *priori* uniforme com parâmetros $(0,0625; 0,20535; 0,23215)$ e da *posteriori* com os parâmetros $(26,0625; 24,20535; 6,23215)$



Fonte: As autoras (2021)

Portanto, pode-se observar que para *a priori* 3 (*simplex* à esquerda) os pontos estão espalhadas em torno do *simplex*, mas mais aglomerados nas categorias outras escolas e outras profissões. Já no *simplex* à direita, que representa o da *posteriori*, mesmo *a priori* os pontos sendo mais concentrados para outras escolas e outras profissões é possível perceber novamente que os pontos se concentram mais para a categoria de professor de escola estadual e professor de outras escolas, assim como nos outros *simplex* plotados anteriormente para as outras *posterioris* obtidas.

5 Considerações finais

Foi possível concluir que *a priori* uniforme, como foi a menos informativa, gerou estimativas mais distantes, mas ainda não sendo ruins ao comparar com as frequentistas. A *priori* 2 foi a que desempenhou melhor influência nas estimativas, pois essa incorporou uma informação relevante de que a preferência são para os que atuam em sala de aula, e *a priori* 3, mesmo com maior tendência para a categoria outras profissões, ainda obteve-se estimativas precisas.

6 Agradecimentos

As autoras agradecem o apoio financeiro concedido pelo CNPq.

Referências

[1] Marianna Avetisyan and Jean-Paul Fox. The dirichlet-multinomial model for multivariate randomized response data and small samples. *Psicologica: International Journal of Methodology and Experimental Psychology*, 33(2):362–390, 2012.

[2] Marlova Estela Caldatto, Regina Maria Pavanello, and Dario Fiorentini. O PROFMAT e a Formação do Professor de Matemática: uma análise curricular a partir de uma perspectiva processual e descentralizadora. *Bolema: Boletim de Educação Matemática*, 30:906–925, 2016.

[3] George Casella and Roger L. Berger. *Statiscal Inference*. Duxbury Press, df, 2 edition, 2002.

[4] Angelo Henrique Lopes da Silva. Estimativa de proporções em questões politômicas. *Revista do TCU*, 1(125):18–27, 2012.

[5] Jacqueline Patrícia Duarte De Oliveira and Telles Timóteo Da Silva. Sobre as distribuições binomial e multinomial. *Revista de Matemática*, 5(1):1–28, 2018.

[6] Ricardo S Ehlers. Inferência bayesiana. *Departamento de Matemática Aplicada e Estatıstica, ICMC-USP*, 64, 2011.

[7] Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian data analysis*. Chapman & Hall, London, 3 edition, 2013.

[8] Paul Gerhard Kinas and Humber Agrelli Andrade. *Introdução à análise bayesiana (com R)*. Consultor Editorial, Porto Alegre, 2 edition, 2020.

[9] Dennis V Lindley. The Bayesian analysis of contingency tables. *The Annals of Mathematical Statistics*, pages 1622–1643, 1964.

[10] Sebastião Martins Filho, Fabyano Fonseca Silva, Antonio Policarpo Souza Carneiro, and Joel Augusto Muniz. Abordagem bayesiana das curvas de crescimento de duas cultivares de feijoeiro. *Ciência Rural*, 38(6):1516–1521, 2008.

[11] Pedro Alberto Morettin and Wilton Oliveira Bussab. *Estatística básica*. Editora Saraiva, São Paulo, 9 edition, 2017.

[12] Carlos Daniel Paulino, Maria Antônia Amaral Turkman, Bento Murteira, and Giovani L. Silva. *Estatística bayesiana*. Fundação Calouste Gulbenkian, Lisboa, 2 edition, 2018.

[13] R Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, 2020.

[14] Ricardo Luis dos Reis, Joel Augusto Muniz, Fabyano Fonseca Silva, Thelma Sáfadi, and Luiz Henrique de Aquino. Abordagem bayesiana da sensitividade de modelos para o coeficiente de endogamia. *Ciência Rural*, 39(6):1752–1759, 2009.