Distribuição Multinomial: Propriedades, Inferência e Exemplos

André F. B. Menezes

Department of Statistics, Universidade Estadual de Maringá, PR, Brasil

3 de Abril de 2018

Organização

Referências

- Agresti, A., 1990. **Categorical Data Analysis.** Wiley: New York.
- Agresti, A., 2007. **An Introduction to Categorical Data Analysis.** John Wiley & Sons, Inc: New Jersey.
- Bishop, Y. M. M., Fienberg, S. E., Holland, P.W., 1975. **Discrete Multivariate Analysis: Theory and Practice.** Cambridge, MA: The MIT Press.
- Davis, C. S., Jones, M. P., 1992. **Maximum likelihood estimation: for the Multinomial distribution.** Teaching Statistics 14 (3), 9–11.
- Gelman, A., Carlin, J. B., Stern, H. S., Rubin, D. B., 2004. **Bayesian Data Analysis.** Chapman & Hall/CRC: New York.
- Johnson, N. L., Kotz, S., Balakrishnan, N., 1996. **Discrete multivariate distributions.** Wiley: New York.

Definição

Considere uma série de n ensaios independentes, em que cada um dos quais apenas um dos k eventos mutuamente exclusivos $E_1, E_2, \ldots, E_k$ pode ser observado, e na qual a probabilidade de ocorrência do evento E_k é igual a p_k (com $p_1 + p_2 + \ldots + p_k = 1$). Sejam $N_1, N_2, \ldots, N_k$ variáveis aleatórias denotando o número de ocorrências dos eventos $E_1, E_2, \ldots, E_k$, respectivamente, nos n ensaios, com $\sum_{i=1}^{k} N_i = n$. Então a distribuição conjunta de $N_1, N_2, \ldots, N_k$ é dada por

$$\Pr\left[N_1 = n_1, \ldots, N_k = n_k\right] = n! \prod_{i=1}^{k} \left(\frac{p_i^{n_i}}{n_i!}\right), \qquad (1)$$

em que $n_1, \ldots, n_k$ são inteiros não negativos satisfazendo $\sum_{i=1}^{k} n_i = n$.

Definição

- A função massa de probabilidade conjunta apresentada em (1) é denominada na literatura de distribuição Multinomial, com parâmetros $(n; p_1, p_2, \ldots, p_k)$;
- Os parâmetros p_i, $i = 1, \ldots, k$, são chamados de probabilidades das células e n é chamado de índice;
- A distribuição Binomial é um caso espacial quando existem $k = 2$ categorias/eventos. Além disso, pode-se notar que a distribuição marginal dos N_i, $i = 1, \ldots, k$, é Binomial com parâmetros (n, p_i);
- Portanto, a distribuição Multinomial pode ser considerada uma generalização multivariada da distribuição Binomial.

Definição

- Johnson et al. (1996) apontam que a distribuição Multinomial pode surgir em diversos contextos.
- Se $X_1, X_2, \ldots, X_k$ são variáveis aleatórias com distribuição de Poisson com média $\lambda_1, \lambda_2, \ldots, \lambda_k$, respectivamente, então a distribuição condicional conjunta de $X_1, X_2, \ldots, X_k$ dado $\sum_{i=1}^{k} X_i = n$ é Multinomial com parâmetros $(n; \pi_1, \pi_2, \ldots, \pi_k)$ em que

$$\pi_i = \frac{\lambda_i}{\lambda_1 + \lambda_2 + \cdots + \lambda_k}, \qquad i = 1, \ldots, k.$$

Propriedades

Os principais momentos da distribuição Multinomial são dados por

$$
\begin{aligned}
\mathbb{E}\left(N_i\right) &= n\,p_i, \qquad\qquad \operatorname{Var}\left(N_i\right) = n\,p_i\left(1-p_i\right), \\
\operatorname{Cov}\left(N_i, N_j\right) &= -n\,p_i\,p_j, \qquad \operatorname{Corr}\left(N_i, N_j\right) = -\sqrt{\frac{p_i\,p_j}{\left(1-p_i\right)\left(1-p_j\right)}}, \quad i \neq
\end{aligned}
$$

A matriz de variância-covariância da distribuição Multinomial pode ser escrita como

$$
V = n\left(\text{diag}(\boldsymbol{p}) - \boldsymbol{p}\,\boldsymbol{p}^\top\right), \tag{2}
$$

em que $\boldsymbol{p} = (p_1, \ldots, p_k)$ e $\top$ denota o operador transposto.

Inferência Clássica

Observando uma amostra aleatória $\boldsymbol{n} = (n_1, \ldots, n_k)$ da distribuição Multinomial com $\sum\limits_{i=1}^{n} n_i = n$, então os estimadores de máxima verossimilhança, $\widehat{\boldsymbol{p}} = (\widehat{p}_1, \widehat{p}_2, \ldots, \widehat{p}_k)$ de $\boldsymbol{p} = (p_1, p_2, \ldots, p_k)$, são obtidos maximizando a função de log-verossimilhança, cuja expressão é dada por

$$\ell = \ell(\boldsymbol{p} \mid \boldsymbol{n}) = \log n! - \sum_{i=1}^{k} \log n_i! + \sum_{i=1}^{k} n_i \, \log p_i. \qquad (3)$$

- A restrição $\sum_{i=1}^{k} p_i = 1$ pode ser incorporada utilizando os multiplicadores de Lagrange (ver, por exemplo, Bishop et al., 1975, p. 446).
- Alternativamente, a redundância nos parâmetros pode ser eliminada tratando ℓ como uma função de $p_1, p_2, \ldots, p_{k-1}$ como em Agresti (1990) e Davis e Jones (1992).

Diferenciando (3) com respeito aos parâmetros têm-se

$$\frac{\partial \ell}{\partial p_i} = \frac{n_i}{p_i} - \frac{n_k}{1 - \sum_{i=1}^{k-1} p_i} = \frac{n_i}{p_i} - \frac{n_k}{p_k}. \tag{4}$$

Igualando $\dfrac{\partial \ell}{\partial p_i}$ a zero obtemos

$$n_i = \frac{n_k}{p_k}\, p_i. \tag{5}$$

Somando ambos os lados de (5) em $i = 1, \ldots, k$, obtemos que $\widehat{p}_k = \dfrac{n_k}{n}$. De (5) temos então que

$$\widehat{p}_i = \frac{n_i}{n_k}\, \widehat{p}_k = \frac{n_i}{n_k}\, \frac{n_k}{n} = \frac{n_i}{n}, \tag{6}$$

para $i = 1, \ldots, k - 1$.

Estatística de Pearson

- Considere a hipótese nula $\mathcal{H}_0 : p_j = p_{j0}$, $j = 1, \ldots, k$ em que $\sum_{j=1}^{k} p_{j0} = 1$.
- Quando $\mathcal{H}_0$ é verdadeira, os valores esperados de $\{n_j\}$, denominadas frequências esperadas, são $\mu_j = n\, p_{j0}$, $j = 1, \ldots, k$. Pearson em 1900 propôs a seguinte estatística de teste

$$\chi^2 = \sum_{j=1}^{k} \frac{(n_j - \mu_j)^2}{\mu_j}. \tag{7}$$

Estatística da razão de verossimilhanças

- Sob H_0 a verossimilhança é maximizada quando $\widehat{p}_j = p_{j0}$. No caso geral, ela é maximizada quando $\widehat{p}_j = n_j/n$. A razão de verossimilhanças é igual a

$$\Lambda = \frac{\prod_{j=1}^{k} (p_{j0})^{n_j}}{\prod_{i=1}^{k} (n_j/n)^{n_j}}. \tag{8}$$

- Assim, a estatística da razão de verossimilhança, denotada por S_{LR}, é

$$S_{LR} = -2\,\log\Lambda = 2\sum_{j=1}^{k} n_j \log\left(\frac{n_j}{n\,p_{j0}}\right). \tag{9}$$

- Alternativamente, a inferência sob os parâmetros da distribuição Multinomial pode ser realizada sob o enfoque Bayesiano.
- A distribuição a priori conjugada para $\boldsymbol{p}$ é a versão multivariada da distribuição Beta denominada de distribuição Dirichlet, cuja densidade é dada por

$$\pi(\boldsymbol{p}) = \frac{\Gamma\left(\sum\limits_{i=1}^{k} \alpha_i\right)}{\prod\limits_{i=1}^{k} \Gamma(\alpha_i)} \prod_{i=1}^{k} p_i^{\alpha_i - 1}, \tag{10}$$

em que $\alpha_i > 0$, $p_i > 0$ com $\sum\limits_{i=1}^{k} p_i = 1$.

- O hiperparâmetro α_i pode ser interpretado como uma contagem virtual para o valor i, antes de observar o n_i. Grandes valores para α_i corresponde a um forte conhecimento a priori sobre a distribuição, ao passo que pequenos valores de α_i corresponde a ignorância.
- A distribuição Dirichlet tem média e variância expressas, respectivamente, por

$$\mathbb{E}\left(p_i\right) = \frac{\alpha_i}{\alpha_0} \qquad \text{e} \qquad \mathrm{Var}\left(p_i\right) = \frac{\alpha_i\left(\alpha_0 - \alpha_i\right)}{\alpha_0^2\left(\alpha_0 + 1\right)}, \tag{11}$$

em que $\alpha_0 = \sum_{i=1}^{k} \alpha_i$.

- A distribuição a posteriori é proporcional a

$$\pi(\boldsymbol{p} \mid \boldsymbol{n}) \propto \pi(\boldsymbol{p})\, L(\boldsymbol{p} \mid \boldsymbol{n}) \tag{12}$$

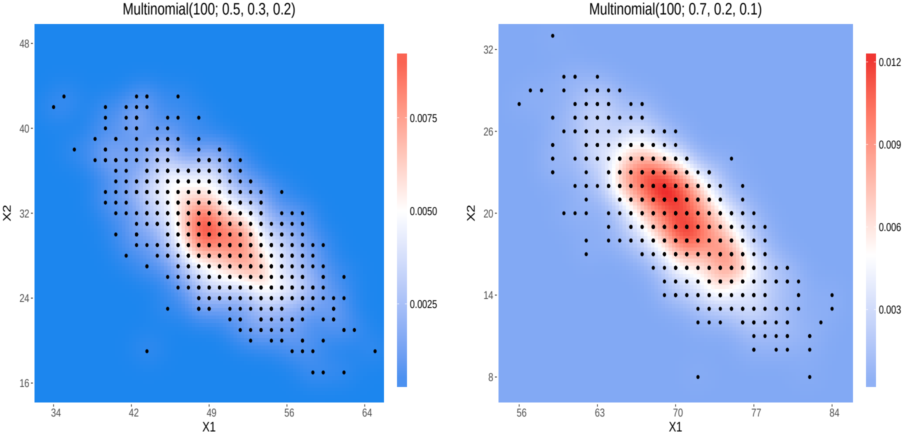
Substituindo as expressões pode-se mostrar que

$$\pi(\boldsymbol{p} \mid \boldsymbol{n}) = \frac{\Gamma\left(\sum\limits_{i=1}^{k} n_i + \alpha_i\right)}{\prod\limits_{i=1}^{k} \Gamma(n_i + \alpha_i)} \prod_{i=1}^{k} p_i^{n_i + \alpha_i - 1} \sim \mathcal{D}(n_i + \alpha_i), \tag{13}$$

isto é, a densidade a posteriori tem distribuição Dirichlet com parâmetros $n_i + \alpha_i$.

Dados simulados

- Utilizando o software R simulamos 1000 amostras pseudos aleatórias da distribuição Multinomial supondo três categorias (X_1, X_2 e X_3) com parâmetros $n = 100$, $\boldsymbol{p} = (0.5, 0.3, 0.2)$ e $\boldsymbol{p} = (0.7, 0.2, 0.1)$.



Votação eleitoral

- Consideramos as respostas de uma questão de uma pesquisa amostral com três repostas possíveis;
- Um total de $n = 1447$ adultos foram entrevistados para indicar a sua preferência, sendo que $n_1 = 727$ apoiaram George Bush, $n_2 = 583$ apoiaram Michael Dukakis e $n_3 = 137$ apoiaram outros candidatos ou não expressaram opinião;
- O objetivo desta análise é comparar a proporção de votos populacional entre Bush e Dukakis. Assim a estatística mais usual para realizar a comparação é $\theta = p_1 - p_2$, isto é, a diferença da proporção populacional entre os dois principais candidatos.

Votação eleitoral

- Assumindo que as contagens n_1, n_2 e n_3 são observações da distribuição Multinomial com tamanho amostral n e respectivas probabilidades p_1, p_2 e p_3. Então, a função de verossimilhança pode ser escrita na forma

$$L(\boldsymbol{p} \mid \boldsymbol{n}) \propto p_1^{n_1}\, p_2^{n_2}\, p_3^{n_3}. \tag{14}$$

Votação eleitoral

- Inferências a respeito de θ sera realizada sob as abordagens Clássica e Bayesiana.
- Os intervalos assintóticos para os parâmetros e para θ foram obtidos utilizando o erro padrão assintótico dado pela inversa da matriz de informação de Fisher observada.
- Sob a abordagem Bayesiana atribuímos distribuição a priori não informativa Dirichlet, isto é, $\pi(\boldsymbol{p}) \sim \mathcal{D}(1, 1, 1)$. Portanto como temos que $\pi(\boldsymbol{p} \mid \boldsymbol{n}) \sim \mathcal{D}(y_1 + 1, y_2 + 1, y_3 + 1)$.
- Valores pseudoaleatórios da distribuição Dirichlet foram gerados baseado no fato de que se W_1, W_2 e W_3 tem distribuição $\mathrm{Gamma}(\alpha_i, 1), i = 1, 2, 3$ e $T = W_1 + W_2 + W_3$, então a distribuição das proporções $(W_1/T, W_2/T, W_3/T)$ tem distribuição $\mathcal{D}(\alpha_1, \alpha_2, \alpha_3)$.

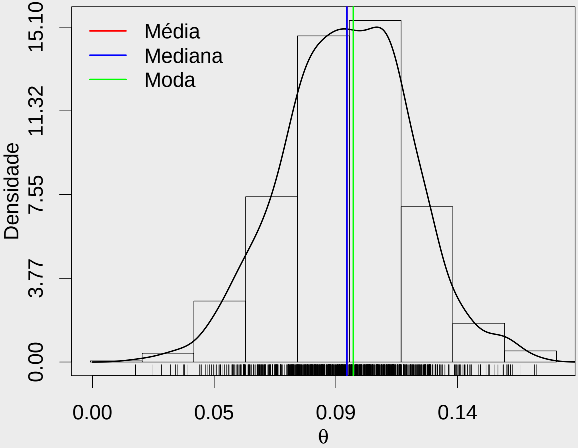
Votação eleitoral

Resumos inferências sob o enfoque Clássico e Bayesiano.

Parâmetro	EMV (E.P.)	Média (D.P.)	I.C. 95%	H.P.D. 95%
p_1	0.5024 (0.0131)	0.5020 (0.0130)	(0.4767, 0.5282)	(0.4763, 0.5264)
p_2	0.4029 (0.0129)	0.4029 (0.0127)	(0.3776, 0.4282)	(0.3795, 0.4305)
p_3	0.0947 (0.0079)	0.0951 (0.0077)	(0.0796, 0.1098)	(0.0801, 0.1100)
θ	0.0995 (0.0248)	0.0991 (0.0245)	(0.0508, 0.1483)	(0.0497, 0.1456)

Votação eleitoral

Densidade da distribuição posteriori simulada de $\theta = p_1 - p_2$.



Tópicos adicionais

- Estudo sobre os intervalos de confiança simultâneos para os parâmetros da distribuição Multinomial:
  - Goodman, L. A., 1965. **On Simultaneous Confidence Intervals for Multinomial Proportions.** Technometrics 7, 247–254.
  - May, W. K., Johnson, W. D., 1997. **A SAS macro for constructing simultaneous confidence intervals for multinomial proportions.** Computer Methods and Programs in Biomedicine 53, 153–162.
- Inclusão de covariáveis:
  - High, R., 2013. **Models for Ordinal Response Data.** SAS Global Forum, Statistics and Data Analysis.