Priemerované diskrétne zmiešané logaritmické rozdelenia
Robust 2018

Samuel Hudec

Univerzita Mateja Bela
Fakulta prírodných vied
Katedra matematiky

25. 1. 2018

Ciele prezentacie

- Priemerované diskrétne zmiešané rozdelenia
- Parametrický priestor
- Optimalizačná úloha
- Niektoré príklady najdené v literatúre
- Ďalšie rozdelenia
- Záver

Priemerované diskrétne zmiešané rozdelenia

Nech X_1, X_2 sú nezávislé diskrétne náhodné premenné nadobúdajúce hodnoty $x_1, x_2, \ldots$ zo známymi (štandardnými) rozdeleniami pravdepodobnosti a pmf tvaru $\{\frac{f_j(x_i)}{C_j}\}_{i=1}^{\infty}$, $j = 1, 2$. Potom diskrétna priemerovaná zmiešaná distribúcia má pravdepodobnostnú funkciu

$$P(x_i) = \frac{af_1(x_i) + bf_2(x_i)}{aC_1 + bC_2}, \quad i = 1, 2, \ldots.$$

kde $f_i(x)$, $i = 1, 2$ sú ,,nenormalizované časti" známych pravdepodobnostných distribúcii s váhami a, b a aC_1, bC_2 sú ich vážené normalizačné konštanty.

Priemerované diskrétne zmiešané rozdelenia (Typ 1a)

$$P(x) = \frac{a^x - rb^x}{x[r\ln(1-b) - \ln(1-a)]}, \quad \text{pre} \quad x = 1, 2, 3 \ldots.$$

kde $-1 \leq a < 1$ a $-1 \leq b < 1$ aby naš rad konvergoval a musí platiť

$$a^x - rb^x \geq 0, \quad \text{pre} \quad x = 1, 2, 3...$$

alebo

$$a^x - rb^x \leq 0. \quad \text{pre} \quad x = 1, 2, 3...$$

aby rad bol pmf.

Priemerované diskrétne zmiešané rozdelenia (Typ 2)

$$P(x) = \frac{a^x + (-1)^{x+1} r b^x}{x[r \ln(1+b) - \ln(1-a)]}, \quad \text{pre} \quad x = 1, 2, 3 \ldots .$$

kde $-1 \leq a < 1$ a $-1 < b \leq 1$. Aby funkcia bola pmf musí platiť

$$a^x + (-1)^{x+1} r b^x \geq 0, \quad \text{pre} \quad x = 1, 2, 3...$$

alebo

$$a^x + (-1)^{x+1} r b^x \leq 0. \quad \text{pre} \quad x = 1, 2, 3...$$

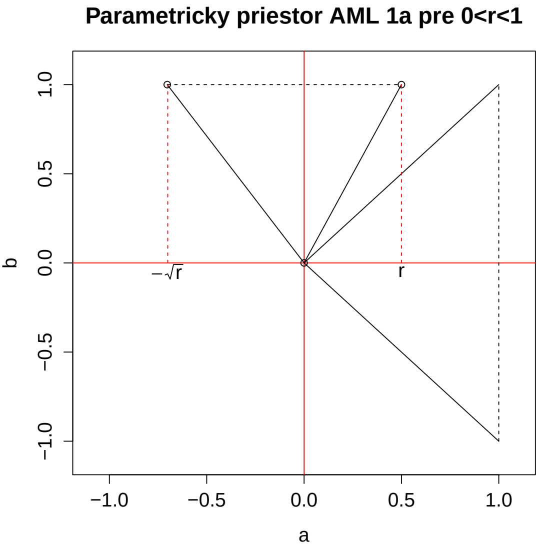
a tiež

$$r \ln(1+b) - \ln(1-a) \neq 0, \quad \text{teda} \quad (1+b)^r \neq 1 - a$$

Parametrický priestor (Typ 1a)

1. $r > 1$,
$\{0 < a < 1, 0 < b \leq \frac{a}{r} \wedge -\frac{a}{\sqrt{r}} \leq b\} \cup \{0 < b < 1, -a \leq b \wedge 0 < a \leq b\}$
2. $r = 1$,
$\{-1 \leq a < 0, -1 \leq b < 1\} \cup \{0 < a < 1, b \in \langle -a, a) \cup (a, 1)\}$
3. $0 < r < 1$,
$\{0 < b \leq a < 1 \wedge -a \leq b < 0\} \cup \{-1 < b < 0, -a \leq \sqrt{r}b < 1 \wedge 0 < a \leq rb < 1\}$
4. $-1 < r < 0$, $\{-1 \leq a < 1, a \leq rb\} \cup \{0 < a < 1, -a \leq b\}$
5. $r = -1$, $-a \leq b < 1$
6. $r < -1$, $\{-1 \leq a < 1, -a \leq b\} \cup \{0 < a < 1, a \leq rb\}$

Parametrický priestor (Typ 1a)



Parametrický priestor (Typ 2)

1. $r > 1$, $\{0 < a < 1, 0 < b \leq \frac{a}{\sqrt{r}} \wedge -\frac{a}{r} \leq b < 0\} \cup \{-1 < b < 0, b \leq a < 0 \wedge 0 < a \leq -b\}$
2. $r = 1$, $\{-1 < a < 0, -1 < b \leq a\} \cup \{0 < a < 1, b \in (-1, -a) \cup (-a, a)\}$
3. $0 < r < 1$, $\{0 < a < 1, 0 < b \leq a \wedge -a \leq b < 0\} \cup \{-1 < b < 0, \sqrt{r}b \leq a < 0 \wedge 0 < a \leq -rb\}$
4. $-1 < r < 0$, $\{-1 < a < 0, br \leq a\} \cup \{0 < a < 1, -1 < b \leq a\}$
5. $r = -1$, $-1 < b \leq a < 1$
6. $r < -1$, $\{-1 < a < 0, b \leq a\} \cup \{0 < a < 1, -1 < b < \frac{a}{r}\}$

Charakteristiky rozdelenia (Typ 1a)

$$G(t) = \frac{r\ln(1-bt) - \ln(1-at)}{r\ln(1-b) - \ln(1-a)},$$

$$\begin{aligned}
E(X) &= \frac{1}{r\ln(1-b) - \ln(1-a)}\left(\frac{a}{(1-a)} - \frac{rb}{(1-b)}\right)\\
E(X^2) &= \frac{1}{r\ln(1-b) - \ln(1-a)}\left(\frac{a}{(1-a)^2} - \frac{rb}{(1-b)^2}\right)\\
E(X^3) &= \frac{1}{r\ln(1-b) - \ln(1-a)}\left(-\frac{(a^2+a)}{(a-1)^3} + r\frac{(b^2+b)}{(b-1)^3}\right),\\
&\dots
\end{aligned}$$

Charakteristiky rozdelenia (Typ 1a)

$$\mu'_{[t]} = \frac{(t-1)!}{r\ln(1-b) - \ln(1-a)}\left(\frac{a^t}{(1-a)^t} - \frac{rb^t}{(1-b)^t}\right).$$

$$\begin{aligned} E(X - E(X)) =& 0 \\ E(X - E(X))^2 =& \frac{1}{C^2}\left(\frac{C-a}{(1-a)^2}a - \frac{C+rb}{(1-b)^2}rb + \frac{2rab}{(1-a)(1-b)}\right) \\ & \dots \end{aligned}$$

kde $C = r\ln(1-b) - \ln(1-a)$.

Optimalizačná úloha

Odhady prarametrov zoznam metód

- Metóda štyroch frekvencii
- Momentova medóda
- Metóda maximalnej vierohodnosti
- Metóda modifikovaným χ^2
- Kombinácie
- MCMC (Stan)

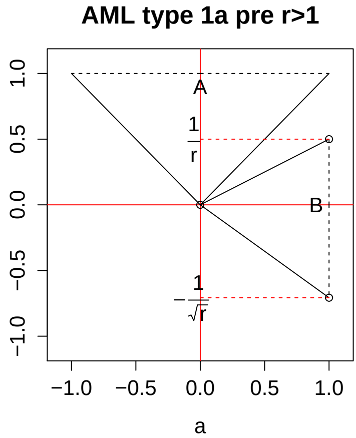
Optimalizačná úloha (likelihood (Typ 1a))

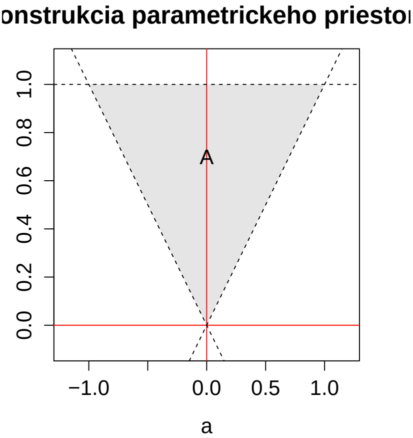
log-likelihood (logaritmická vierohodnostná) funkcia

$$\ln \mathcal{L}(x_1, x_2, ..., x_n; \boldsymbol{\theta}) = \sum_{i=1}^{n} \ln \frac{a^{x_i} - rb^{x_i}}{x_i[r \ln(1-b) - \ln(1-a)]}.$$

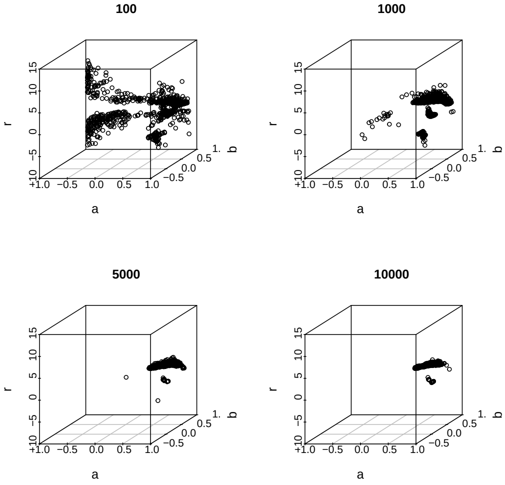
Funkciu použijeme na numerické optimalizovanie za nelinearnych reštrikcii (parametrický priestor). Z mnoho funkcii ktoré nám R ponúka sme si zvolili funkciu `constrOptim.nl` z knižnice `alabama`, ktorá využíva `Augmented Lagrangian Adaptive Barrier Minimization Algorithm`.

Optimalizačná úloha (Typ 1a)





Optimalizačná úloha (Typ 1a)



Optimalizačná úloha

Asymptotické vlastnosti odhadu maximalnej vierohodnotsi

$$\hat{\boldsymbol{\theta}} \approx N_3\left(\boldsymbol{\theta}; \frac{1}{n}\mathbf{J}^{-1}(\boldsymbol{\theta})\right),$$

Napiseme 95% intervaly spoľahlivosti pre parametre r, a, b

$$\left(\hat{a} - u_{0,025}\sqrt{(1/N)\mathbf{J}_{11}^{-1}(\hat{r}, \hat{a}, \hat{b})}, \quad \hat{a} + u_{0,025}\sqrt{(1/N)\mathbf{J}_{11}^{-1}(\hat{r}, \hat{a}, \hat{b})}\right),$$

$$\left(\hat{b} - u_{0,025}\sqrt{(1/N)\mathbf{J}_{22}^{-1}(\hat{r}, \hat{a}, \hat{b})}, \quad \hat{b} + u_{0,025}\sqrt{(1/N)\mathbf{J}_{22}^{-1}(\hat{r}, \hat{a}, \hat{b})}\right),$$

$$\left(\hat{r} - u_{0,025}\sqrt{(1/N)\mathbf{J}_{33}^{-1}(\hat{r}, \hat{a}, \hat{b})}, \quad \hat{r} + u_{0,025}\sqrt{(1/N)\mathbf{J}_{33}^{-1}(\hat{r}, \hat{a}, \hat{b})}\right),$$

Testy dobrej zhody

Nech $\xi_i^{(n)}$ sú empirické početnosti hodnôt z realizácie náhodneho výberu $X_1, X_2, \ldots, X_n$, ktoré patria do i-teho intervalu. Odhad metodou minimalneho χ^2 vektora prametrov budeme označovať $\tilde{\boldsymbol{\theta}} = (\tilde{r}, \tilde{a}, \tilde{b})$. Ak realizácia

$$\sum_{i=1}^{k} \frac{(\xi_i^{(n)(real)} - np_i(\tilde{\boldsymbol{\theta}})^{(real)})^2}{np_i(\tilde{\boldsymbol{\theta}})^{(real)}} \geqq \chi^2_{k-1-u}(1-\alpha), \tag{1}$$

kde $\chi^2_{k-1-u}(1-\alpha)$ je $(1-\alpha)$ kvantil rozdelenia χ^2, tak zamietame hypotézu, že náhodný výber pochádza z ,,našej" rozdelenia na hladine významnosti α.

Testy dobrej zhody

#	$r > 1$	$r < -1$
N=50	0.71	0.99
N=100	0.17	0.89
N=500	0.04	0.14
N=1000	0.036	0.1
N=5000	0.039	0.07
N=10000	0.046	0.067

Niektoré príklady najdené v literatúre

Janardan-logarithmic distribution je definovaná ako

$$P(x) = \frac{(\beta\theta)^x + \alpha\theta^x}{x[-\ln(1-\beta\theta) - \alpha\ln(1-\theta)]}, \quad \text{pre} \quad x = 1, 2, 3\ldots,$$

kde $r = \alpha$, $a = \beta\theta$ a $b = \theta$.

Quine logarithmic distribution najdené v literatúrach

$$P(x) = \frac{p^x + (-1)^{x+1}\frac{1}{3}p^x}{x[1/3\ln 1 + p - \ln(1-p)]}, \quad \text{pre} \quad x = 1, 2, 3\ldots$$

$a = b =$ a $r = 1/3$.

Niektoré príklady najdené v literatúre

Boen Sylwester distribution najdené v literatúrach

$$P(x) = \frac{\left(\frac{q}{1+q}\right)^x + (-1)^{x+1}\left(\frac{1-q}{q}\right)^x}{x[\ln(1+q) - \ln(q)]}, \quad \text{pre} \quad x = 1, 2, 3\ldots$$

$a = \frac{q}{1+q}$, $b = \frac{1-q}{q}$ a $r = 1$.

Feller Arley logarithmic distribution najdené v literatúrach

$$P(x) = \frac{p_2^x + (-1)^{x+1}\left(\frac{q_2-p_1}{p_1}\right)^x}{x[-\ln(1-q_1)]}, \quad \text{pre} \quad x = 1, 2, 3\ldots$$

$a = p_2$, $b = \frac{q_2-p_1}{p_1}$ a $r = 1$.

Niektoré príklady najdené v literatúre

Hunter logarithmic distribution 1 najdené v literatúrach

$$P(x) = \frac{\alpha^x + (-1)^{x+1}\left(\frac{1}{p}\right)^x}{x[\ln(1+1/q) - \ln(1-\alpha)]}, \quad \text{pre} \quad x = 1, 2, 3 \ldots$$

$a = \alpha$, $b = \frac{1}{p}$ a $r = 1$.

Hunter logarithmic distribution 2 najdené v literatúrach

$$P(x) = \frac{\alpha^x + (-1)^{x+1}\left(\frac{p}{p}\right)^x}{x[-\ln p - \ln(1-\alpha)]}, \quad \text{pre} \quad x = 1, 2, 3 \ldots$$

$a = \alpha$, $b = \frac{p}{p}$ a $r = 1$.

Ďalšie rozdelenia

Ako ďalšie sa venujeme zmiešavaniu rozdelení geometrickej a logaritmickej pod názvom Type 3. Všeobecná forma rozdelenia sa dá napísať ako

$$P(x) = \frac{r f_g(x) + f_{log}(x)}{rC_g + C_{log}}, \quad \text{pre} \quad x = 0, 1, 2, 3 \ldots.$$

kde $f_g(x)$ a $f_{log}(x)$ sú „nenormalizované" zložky geometrického a logaritmického rozdelenia.

Ďalšie rozdelenia

Ako ďalšie sa venujeme zmiešavaniu rozdelení geometrickej a logaritmickej pod názvom Type 3. Všeobecná forma rozdelenia sa dá napísať ako

$$P(x) = \frac{rf_g(x) + f_{log}(x)}{rC_g + C_{log}}, \quad \text{pre} \quad x = 0, 1, 2, 3 \ldots .$$

kde $f_g(x)$ a $f_{log}(x)$ sú „nenormalizované" zložky geometrického a logaritmického rozdelenia.

- Negativne binomické rozdelenia
- a iné

Ďalšie rozdelenia

$$P_a(x) = \begin{cases} \dfrac{r}{\frac{r}{(1-q)} - \ln(1-\theta)}, & \text{pre } x = 0, \\ \dfrac{rq^x + \frac{\theta^x}{x}}{\frac{r}{(1-q)} - \ln(1-\theta)}, & \text{pre } x = 1, 2, \ldots. \end{cases}$$

$$P_b(x) = \frac{rq^x + \frac{\theta^x}{x}}{rq(1-q)^{-1} - \ln(1-\theta)}, \quad \text{pre} \quad x = 1, 2, 3 \ldots$$

Ďalšie rozdelenia

$$P_c(x) = \begin{cases} \dfrac{r}{r(1-b)^{-1} + \ln(\frac{1-a}{1-b})}, & \text{pre } x = 0, \\ \dfrac{rb^x + \frac{b^x - a^x}{x}}{r(1-b)^{-1} + \ln(\frac{1-a}{1-b})}, & \text{pre } x = 1, 2, \ldots. \end{cases}$$

$$P_d(x) = \begin{cases} \dfrac{r(1-q)^m}{r - \ln(1-q)}, & \text{pre } x = 0, \\ \dfrac{r\binom{m+x-1}{x}(1-q)^m q^x + \frac{q^x}{x}}{r - \ln(1-q)}, & \text{pre } x = 1, 2, \ldots. \end{cases}$$

Literatúra

- Anděl, J. *Matematická statistika* SNTL/ALFA, Praha 1978.
- Anděl, J., *Základy matematické statistiky* Univerzita Karlova, Praha 2002.
- Bloomfield, V. A., *Using R for Numerical Analysis in Science and Engineering* University of Minnesota, Minneapolis, USA 2014.
- Boen J. R., Sylwester D., *A quantitative discussion of the effectiveness of voiding as a defense against bladder infection* Biometrics Vol. 22, No. 1 (Mar., 1966), pp. 53-57. International Biometric Society 1966
- Cane, V., *Mathematical models for neutal networks* Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics an Probability 1967, 21-26.

Literatúra

- Gani J., *Problems in the Probability Theory of Storage Systems*,Journal of the Royal Statistical Society. Series B, Vol. 19, No. 2 (1957), pp. 181-206 JSTOR 1957
- Griffiths R. C., Pakes A. G., *An infinite-alleles version of the simple branching process* Advances in Applied Probablity Volume 20, Issue 3 September 1988, pp. 489-524 Cambridge University Press 1988
- Gross, D., Harris, C. M., *Fundamentals of Queueing Theory* Wiley, New York 1985.
- Hudec, S. *Zmiešané rozdelenia pravdepodobnosti typu 1* (diplomova práca), 2015
- Hunter, J. J., *Mathematical techniques of applied probability*, Volume 2: discrete time m odels, techniques and applications. 1983.

Literatúra

- Johnson, N. L., Kemp A. W., Knotz S. *Univariate Discrete Distrbutions* Wiley-Interscience, Hoboken (NJ), 2005.
- Kemp A. W., *Convolutions Involving Binomial Pseudo-Variables* The Indian Journal of Statistics, Vol. 41, No. 3/4 (Oct., 1979), pp. 232-243 JSTOR 1979
- Levikson B., *The age distribution of Markov processes* Advances in Applied Probablity Volume 14, Issue 3 September 1977, pp. 492-506 Cambridge University Press 1977
- Owen, J., Maillardet, R., Robinson, A., *Scientific Programming and Simulation using r* Taylor and Francis Group, LLC, 2009.
- Panaretos, S., *On Moran's property of the Poisson distribution* Biometrical Jurnal 25 1983, 69-76.

Literatúra

- Plunkett I. G. Jain G. C., *Three generalised negative binomial distributions* Biometrical Journal Wiley Online Library 1975
- Quine M. P., *Asymptotic Results for Estimators in a Subcritical Branching Process with Immigration* The Annals of Probability Volume 4, Number 2 (1976), 319-325. 1976
- Rubin, H., Vere-Jones, D., *Domains of attraction for the subcritical Galton-Watson branching process* Journal of Applied Probability 5 1968, 216-219.
- Schervish, M. J., DeGroot, M. H., *Probability and Statistic* Carnegie Mellon University: Pearson Education, Inc. 2002.

Literatúra

- Wimmer, G. *Diskrétne jednorzmerné rozdelnia pravdepodobnosti* Matfyz press, Praha 2000.
- Wimmer, G., Mačutek, J., Altmann, G., *Discrete averaged mixing applied to the logarithmic distributions* Mathematica Slovaca (2015).
- Wimmer, G., Altmann, G., *Thesaurus of univariate discrete probability distributions* Stamm, Essen,1999.
- Wimmer G. Altmann G., *The Multiple Poisson Distribution, Its Characteristics and a Variety of Forms* Biometrical Journal Wiley Online Library 1996
- Wimmer G. Altmann G., *Review Article: On Vocabulary Richness* Journal of Quantitative Linguistics Volume 6, Taylor and Francis 1999

Ďakujem za pozornosť.