Домашнее задание 2

Дородный Дмитрий СКБ172

6 декабря 2019 г.

Содержание

1 Выборочные параметры гамма распределения

1.1 Среднее

$n = 5$
1.534
2.349
2.100
1.475
2.169
$n = 10$
1.865
1.717
1.580
2.052
1.561

$n = 100$
1.864
1.873
1.857
1.747
1.819
$n = 1000$
1.828
1.828
1.827
1.841
1.859
$n = 100000$
1.831
1.833
1.830
1.835
1.830

1.2 Дисперсия

$n = 5$
0.230
0.235
0.582
0.256
0.767
$n = 10$
0.581
0.845
0.579
0.347
0.230
$n = 100$
0.663
0.669
0.647
0.605
0.497
$n = 1000$
0.564
0.648
0.572
0.619
0.591

$n = 100000$
0.613
0.608
0.615
0.617
0.605

1.3 Оценка параметров гамма распределения

1.3.1 Выводы формул и свойств оценок

$M = \frac{\alpha}{\lambda}, D = \frac{\alpha}{\lambda^2}$, решая систему и по методу моментов получаем: $\lambda^* = \frac{\bar{X^*}}{S_x^2}, \alpha^* = \frac{\bar{X^*}^2}{S_x^2}$, где $S_x^2, \bar{X}^*$ - выборочная дисперсия и выборочное среднее соответственно. Вообще говоря, гамма распределение принадлежит экспоненциальному семейству, так как можно представить в форме
$f(x,\theta) = exp(A(\theta)B(x) + C(\theta) + D(x)$
$f(x|\alpha,\lambda) = \frac{x^{\lambda-1}e^{\frac{-x}{\alpha}}}{\Gamma(\lambda)\alpha^{\lambda}}$
$\Gamma(\theta,\lambda) = \frac{x^{\lambda-1}e^{\frac{-x}{\theta}}}{\Gamma(\lambda)\theta^{\lambda}} = exp(\frac{-x}{\theta} - \lambda ln(\theta) + (\lambda-1)ln(x)) + \Gamma(\lambda)$.
$A(\theta) = \frac{1}{\theta}$
$B(x) = -x$
$C(\theta) = \lambda ln(\theta)$
$D(x) = (\lambda-1)ln(x)$
Получаем распредление из эксп. семейства, тогда оцениваемый параметр:
$\tau(\theta) = -\frac{C'(\theta)}{A'(\theta)} = \frac{\lambda}{\theta}\frac{\theta^2}{-1} = \lambda\theta$
Лишняя λ, домножим и разделим первый член на λ:
$A(\theta) = \frac{\lambda}{\theta}, B(x) = \frac{-x}{\lambda}$
Теперь $\tau(\theta) = -\frac{C'(\theta)}{A'(\theta)} = \frac{\lambda}{\theta}\frac{\theta^2}{-\lambda} = \theta$
Оценка параметра:
$T^* = T^*(x) = \frac{1}{n}\sum_{i=1}^{n} B(X_i) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\lambda}X_i = \frac{1}{\lambda}\bar{X}$, которая является эффективной, как оценка экспоненциальной модели. Проверим несмещенность:
$ET^* = E\frac{\bar{X}}{lambda} = \frac{1}{\lambda}E\bar{X} = \frac{1}{\lambda}\frac{1}{n}E\sum_{i=1}^{n} X_i = \frac{n\lambda\theta}{n\lambda} = \theta = \tau(\theta)$
Мат. ожидание оценки параметра равно самому параметру, следовательно, оценка несмещенная.
Т.к. оценка эффективная, то можно определить ее дисперсию как:
$D_\theta T = \frac{\tau'(\theta)}{nA'(\theta)} = \frac{\theta^2}{\lambda n}$
Очевидно, что дисперсия стремится к нулю с ростом n, следовательно, оценка состоятельная.
Полученная оценка эффективная, значит она так же оптимальна (т.к. не будет лучше, чем нижняя граница в неравенстве Рао-Крамера)

1.3.2 Проверка на совпадение с реальными параметрами

$n = 5$
$\lambda^* = 6.657$ $\alpha^* = 10.217$

$\lambda^* = 9.963$ $\alpha^* = 23.406$
$\lambda^* = 3.605$ $\alpha^* = 7.574$
$\lambda^* = 5.753$ $\alpha^* = 8.488$
$\lambda^* = 2.827$ $\alpha^* = 6.135$

$n = 10$
$\lambda^* = 3.207$ $\alpha^* = 5.984$
$\lambda^* = 2.030$ $\alpha^* = 3.487$
$\lambda^* = 2.727$ $\alpha^* = 4.309$
$\lambda^* = 5.911$ $\alpha^* = 12.135$
$\lambda^* = 6.775$ $\alpha^* = 10.582$

$n = 100$
$\lambda^* = 2.812$ $\alpha^* = 5.245$
$\lambda^* = 2.800$ $\alpha^* = 5.246$
$\lambda^* = 2.869$ $\alpha^* = 5.331$
$\lambda^* = 2.883$ $\alpha^* = 5.037$
$\lambda^* = 3.656$ $\alpha^* = 6.653$

$n = 1000$
$\lambda^* = 3.239$ $\alpha^* = 5.924$
$\lambda^* = 2.817$ $\alpha^* = 5.15$
$\lambda^* = 3.191$ $\alpha^* = 5.831$
$\lambda^* = 2.973$ $\alpha^* = 5.477$
$\lambda^* = 3.144$ $\alpha^* = 5.845$

$n = 100000$
$\lambda^* = 2.984$ $\alpha^* = 5.465$
$\lambda^* = 3.011$ $\alpha^* = 5.521$
$\lambda^* = 2.976$ $\alpha^* = 5.449$
$\lambda^* = 2.972$ $\alpha^* = 5.456$
$\lambda^* = 3.023$ $\alpha^* = 5.536$

Выборки были сгенерированы с параметрами $\lambda = 3, \alpha = 5.5$
Можно заметить, что с ростом выборки оценка весьма точная.

2 Выборочные параметры распределения Бореля-Таннера

2.1 Среднее

$n = 5$
7.2
5.8
5.6
6.0

8.4

$n = 10$

6.7

6.7

7.3

7.2

5.8

$n = 100$

6.75

6.58

6.78

6.76

6.62

$n = 1000$

6.622

6.727

6.634

6.733

6.761

$n = 100000$

6.66641

6.66666

6.67107

6.67369

6.66809

2.2 Дисперсия

$n = 5$

12.2

4.199

3.3

1.5

6.3

$n = 10$

5.788

3.344

11.566

12.400

2.622

$n = 100$

11.058

6.205

7.971

9.456

8.076

$n = 1000$
7.542
6.895
7.095
7.943
8.001
$n = 100000$
7.438
7.380
7.366
7.486
7.444

2.3 Оценки для распределения Бореля-Таннера

2.3.1 Выводы формул и свойств оценок

Функция правдоподобия для данного распределния содержит несокращающиеся произведения факториалов и переменных, и это так же не является распределением экспоненциальног осемейства, т.к. в функции распределения сумма составных функций в экспоненциальной форме - следовательно не получится представить в виде, подходящеи для распределения экспоненциального семейства. Воспользуемся методом моментов: возьмем 1й момент и 2й центральный момент
$m_1 = \bar{x} = \frac{r}{1-\alpha}$
$m_2 = S_x^2 = \frac{r}{(1-\alpha)^3}$
Вначале, найдем оценку для параметра r (решение системы получено в Wolfram):
$r = -\frac{m_1^2 - m_1^{\frac{3}{2}}\sqrt{m_1+4m_2}}{2m_2}$
$r = -\frac{m_1^2 - m_1^2\sqrt{1+4\frac{m_2}{m_1}}}{2m_2}$
$r = -\frac{1-\sqrt{1+4m_1\frac{m_2}{m_1^2}}}{2\frac{m_2}{m_1^2}}$
$\frac{m_2}{m_1^2} = \frac{S_x^2}{\bar{x}^2}$ - коэффициент вариации v
$r = \frac{\sqrt{1+4\bar{x}v}-1}{2v}$
Т.к. r - принимает целые значения (кол-во человек в начале очереди), так что надо прибавить 0.5 и взять целую часть:
$r = [\frac{\sqrt{1+4\bar{x}v}-1}{2v} + 0.5]$
Выразим второй параметр (метод моментов):
$\alpha^* = 1 - \frac{r}{\bar{x}}$

2.3.2 Проверка свойств

Несмещенность
Оценка параметра несмеценная, если её математическое ожидание равно самому параметру:
$ET(\theta) = \theta$
$E\theta^* = E(1 - \frac{r}{\bar{x}})$
$E\theta^* = 1 - rE\frac{1}{\bar{x}}$
$E\theta^* = 1 - rE\frac{1-\theta}{r} = \theta$
Получаем что оценка несмещенная.
Состоятельность
Оценка параметра состоятеьная, если её дисперся стремится к 0 с ростом размера выборки
$DT(\theta) \overset{n \to \infty}{=} 0$
$D\theta^* = E(\theta^*)^2 - (E\theta)^2$
Вычислим $E(\theta^*)^2$
$E(\theta^*)^2 = E(1 - \frac{r}{\bar{x}})^2$
$E(\theta^*)^2 = E(1 - \frac{2r}{\bar{x}} + \frac{r^2}{\bar{x}^2})$
$E(\theta^*)^2 = 1 - 2rE\frac{1}{\bar{x}} + r^2E\frac{1}{\bar{x}^2}$
$E(\theta^*)^2 = 1 - 2rE\frac{1-\theta}{r} + r^2E\frac{(1-\theta)^2}{r^2}$
$E(\theta^*)^2 = 1 - 2 + 2\theta + 1 - 2\theta + \theta^2$
$E(\theta^*)^2 = \theta^2$
Тогда дисперсия будет:
$DT(\theta) = E(\theta^*)^2 - (E\theta)^2 = \theta^2 - \theta^2 = 0$
Получаем, что оценка состоятельная.
Для проверки эффективности и оптимальности используем неравенство Рао-Крамера, однако, оно применимо только к регулярным моделям.
Модель регулярна, если удовлетворяет следующим критериям:
1) Математическое ожидание квадрата вклада выборки сходится для любого значения параметра из множества
$0 < EV^2(X, \theta) < \infty \forall \theta \in \Theta$
2) можно менять местами интегрирование и дифференцирование в следующем интеграле:
$\int T(x)L(x, \theta)dx$, при интегрировании по всему выборочному пространству

Посчитаем вклад выборки для распределения Бореля-Таннера:
$V(X, \theta) = \sum_{i=1}^{n} \frac{\partial}{\partial \theta} \ln(P(x, \theta))$
$\ln(P) = \ln(r) - \sum_{i=1}^{x-r} + (x - r - 1)\ln(x) - \theta x + (x - r)ln(\theta)$
$\frac{\partial}{\partial \theta} \ln(P) = -x + \frac{x-r}{\theta}$
$V(X, \theta) = \sum_{i=1}^{n}(-x_i + \frac{x_i - r}{\theta})$
$V(X, \theta = -n\bar{x} + \frac{n\bar{x} - nr}{\theta}) = n(-\bar{x} + \frac{\bar{x} - r}{\theta})$
$EV^2(X, \theta) = n^2E(-\bar{x} + \frac{\bar{x} - r}{\theta})^2$
Что, очевидно, не сходится. Следовательно, модель нерегулярна. Тогда можно воспользоваться формулой для нерегулярных моделей, которая даст оптимальную (а следовательно, несмещенную и состоятельноую) оценку. Эф-

фективной оценки у нерегулярной модели быть не может.
$\tau^*(\theta) = H(X_{(n)}, \theta) = \tau(X_{(n)}) + \frac{\tau'(X_{(n)})}{nf(X_{(n)}, X_{(n)})}$
$\tau^*(\theta) = X_{(n)} + \frac{1}{n\frac{r}{(X_{(n)}-r)!}X_{(n)}^{2X_{(n)}-2r-1}e^{-X_{(n)}^2}}$

2.3.3 Проверка на совпадение с истиными параметрами

$r^* = 4.0$ $\alpha^* = 0.444$
$r^* = 4.0$ $\alpha^* = 0.31$
$r^* = 4.0$ $\alpha^* = 0.285$
$r^* = 5.0$ $\alpha^* = 0.166$
$r^* = 6.0$ $\alpha^* = 0.285$
$r^* = 4.0$ $\alpha^* = 0.402$
$r^* = 5.0$ $\alpha^* = 0.253$
$r^* = 4.0$ $\alpha^* = 0.452$
$r^* = 4.0$ $\alpha^* = 0.444$
$r^* = 4.0$ $\alpha^* = 0.31$
$r^* = 4.0$ $\alpha^* = 0.407$
$r^* = 4.0$ $\alpha^* = 0.392$
$r^* = 4.0$ $\alpha^* = 0.410$
$r^* = 4.0$ $\alpha^* = 0.408$
$r^* = 4.0$ $\alpha^* = 0.395$
$r^* = 4.0$ $\alpha^* = 0.395$
$r^* = 4.0$ $\alpha^* = 0.405$
$r^* = 4.0$ $\alpha^* = 0.397$
$r^* = 4.0$ $\alpha^* = 0.405$
$r^* = 4.0$ $\alpha^* = 0.408$
$r^* = 4.0$ $\alpha^* = 0.399$
$r^* = 4.0$ $\alpha^* = 0.399$
$r^* = 4.0$ $\alpha^* = 0.4$
$r^* = 4.0$ $\alpha^* = 0.4$
$r^* = 4.0$ $\alpha^* = 0.4$

Выборки были сгенерированы с параметрами $r = 4, \alpha = 0.4$
Видно, что оценка довольно точная.