Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное
учреждение высшего образования
«Московский государственный технический
университет имени Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

Математическая статистика

Домашнее задание №1

Вариант №15

Куприй Александр, ИУ7–63Б

2020 г.

Задача №1
Предельные теоремы теории вероятностей

Условие

В Москве рождается в год около 122500 детей. Считая вероятность рождения мальчика равной 0.51, найти вероятность того, что число мальчиков, которые родятся в Москве в текущем году, превысит число родившихся девочек не менее, чем на 1500.

Решение

Пусть:

- X - случайная величина, принимающая значения равные числу мальчиков, родившихся за год в Москве.
- Y - случайная величина, принимающая значения равные числу девочек, родившихся за год в Москве.

Из условия имеем следующую систему уравнений:

$$\begin{cases} X + Y = 122500 \\ X - Y \geqslant 1500 \end{cases}$$

Следовательно:

$$X \geqslant 62000$$

Условие данной задачи описывает схему испытаний Бернулли. При этом $n = 122500$ — число испытаний. Тогда по интегральной теореме Муавра-Лапласа имеем:

$$P\Big\{k_1 \leqslant k \leqslant k_2\Big\} = P\Big\{62000 \leqslant k \leqslant 122500\Big\} \approx \Phi(x_2) - \Phi(x_1)$$

где, k — число успехов; $x_i = \frac{k_i - np}{\sqrt{npq}}, i = \overline{1,2}$; $p = 0.51$ — вероятность успеха, $q = 1 - p = 0.49$ — вероятность неудачи.

$$P\Big\{x \in [62000, 122500]\Big\} = \Phi\left(\frac{122500 - 62475}{\sqrt{30612.75}}\right) - \Phi\left(\frac{6200 - 62475}{\sqrt{30612.75}}\right) \approx$$
$$\approx \Phi(343) + \Phi(2.7) \approx \underline{\underline{0.9964}}$$

Задача №2
Метод моментов

Условие

С использованием метода моментов для случайной выборки $\vec{X} = (X_1, \ldots, X_n)$ из генеральной совокупности X найти точечные оценки указанных параметров заданного закона распределения:

$$f_X(x) = \frac{1}{2^{\frac{\theta}{2}}\Gamma(\frac{\theta}{2})} x^{\frac{\theta}{2}-1} e^{\frac{-x}{2}}, \quad x > 0$$

Решение

$$f_X(x) = \frac{1}{2^{\frac{\theta}{2}}\Gamma(\frac{\theta}{2})} x^{\frac{\theta}{2}-1} e^{\frac{-x}{2}} = \frac{\frac{1}{2}^{\frac{\theta}{2}}}{\Gamma(\frac{\theta}{2})} x^{\frac{\theta}{2}-1} e^{-\frac{1}{2}x} \tag{1}$$

Гамма распределение.

Данный закон распределения зависит от единственного параметра θ, следовательно система метода моментов будет содержать только одно уравнение с одной неизвестной. Это уравнение имеет вид:

$$MX = \hat{\mu}_1(\overrightarrow{X}) \tag{2}$$

При этом:

$$\hat{\mu}_1(\overrightarrow{X}) = \overline{X} \tag{3}$$

Теперь необходимо вычислить математическое ожидание случайной величины, подчиняющейся данному закону распределения:

$$MX = \int_{-\infty}^{+\infty} x \cdot f_X(x) dx$$

$$\begin{aligned} MX &= \int_0^{+\infty} x \frac{\frac{1}{2}^{\frac{\theta}{2}}}{\Gamma(\frac{\theta}{2})} x^{\frac{\theta}{2}-1} e^{-\frac{1}{2}x} dx = \int_0^{+\infty} \frac{\frac{1}{2}^{\frac{\theta}{2}}}{\Gamma(\frac{\theta}{2})} x^{\frac{\theta}{2}-1} e^{-\frac{1}{2}x} dx = \begin{bmatrix} \frac{x}{2} = t \\ x = 2t \end{bmatrix} = \\ &= \int_0^{+\infty} \frac{(2t)^{\frac{\theta}{2}} \frac{1}{2}^{\frac{\theta}{2}} e^{-t}}{\Gamma(\frac{\theta}{2})} 2dt = \frac{2}{\Gamma(\frac{\theta}{2})} \int_0^{+\infty} t^{\frac{\theta}{2}} e^{-t} dt = \frac{2\Gamma(\frac{\theta}{2}+1)}{\Gamma(\frac{\theta}{2})} = \frac{2\Gamma(\frac{\theta}{2})\frac{\theta}{2}}{\Gamma(\frac{\theta}{2})} = \theta \end{aligned} \tag{4}$$

Приравняем теоретические моменты к их выбранным аналогам, найдём неизвестный параметр:

$$MX = \overline{x} = \theta \tag{5}$$

Ответ: $\theta = \overline{x}$

Задача №3
Метод максимального правдоподобия

Условие

С использованием метода максимального правдоподобия для случайной выборки $\vec{X} = (X_1, \ldots, X_n)$ из генеральной совокупности X найти точечные оценки параметров заданного закона распределения:

$$f_X(x) = \frac{4\theta^3}{\sqrt{\pi}} x^2 e^{-\theta^2 x^2}$$

А так же вычислить выборочные значения найденных оценок для выборки:

$$\vec{x}_5 = (1, 4, 7, 2, 3)$$

Решение

Рассматриваемая случайная величина является непрерывной. Следовательно, функция правдоподобия принимает вид:

$$\mathcal{L}(x_1, ..., x_n, \theta) = f_X(x_1, \theta) \cdot ... \cdot f_X(x_n, \theta) = \left(\frac{4\theta^3}{\sqrt{\pi}}\right)^n \cdot e^{-\theta^2 \sum\limits_{i=1}^{n} x_i^2} \cdot (x_1^2 \cdot ... \cdot x_n^2)$$

Для упрощения вычислений, проинтегрируем полученную функцию:

$$\ln \mathcal{L} = n \ln \frac{4\theta^3}{\sqrt{\pi}} - \theta^2 \sum_{i=1}^{n} x_i^2 + \ln(x_1^2 \cdot ... \cdot x_n^2) =$$

$$n \ln 4 + 3n \ln \theta - \frac{n}{2} \ln \pi - \theta^2 \sum_{i=1}^{n} x_i^2 + \ln(x_1^2 \cdot ... \cdot x_n^2)$$

Зная необходимое условие экстремума $\frac{\partial \ln L}{\partial \theta} = 0$, найдем искомый параметр:

$$\frac{\partial \ln L}{\partial \theta} = \frac{3n}{\theta} - 2\theta \sum_{i=1}^{n} x_i^2 = 0 \Rightarrow \frac{3n - 2\theta^2 \sum\limits_{i=1}^{n} x_i^2}{\theta} = 0 \Rightarrow \hat{\theta} = \theta = \pm \sqrt{\frac{3n}{2 \sum\limits_{i=1}^{n} x_i^2}}$$

Зная достаточное условие экстремума $\frac{\partial^2 \ln \mathcal{L}}{\partial \theta^2} \neq 0$, проверим полученное значение:

$$\frac{\partial^2 \ln \mathcal{L}}{\partial \theta^2} = -\frac{3n}{\theta^2} - 2 \sum_{i=1}^{n} x_i^2 \Bigg|_{\theta = \hat{\theta}} = -4 \sum_{i=1}^{n} x_i^2 = \left|\overline{x}^2 \geq 0\right| \Rightarrow -4\overline{x}^2 \leq 0$$

Вычислим выборочные значения найденных оценок для выборки $\vec{x_5} = (1, 4, 7, 2, 3)$

$$\hat{\theta} = \pm\sqrt{\frac{3 \cdot 5}{2 \cdot \frac{1^2+4^2+7^2+2^2+3^2}{5}}} = \pm 0.69$$

Ответ:

1. $\hat{\theta} = \pm\sqrt{\frac{3n}{2\sum\limits_{i=1}^{n} x_i^2}}$

2. $\hat{\theta} = \pm 0.69$

Задача №4
Доверительные интервалы

Условие

По результатам n = 10 измерений прибором, не имеющим систематической ошибки, получены следующие отклонения емкости конденсатора от номинального значения (пФ):

$$5.4, -13.9, -11.0, 7.2, -15.6, 29.2, 1.4, -0.3, 6.6, -9.9.$$

Найти 90%-ые доверительные интервалы для среднего значения отклонения емкости от номинального значения и ее среднего квадратичного отклонения.

Решение

1. Пусть X – случайная величина, принимающая значения отклонения ёмкости конденсатора от номинального значения.

   Необходимо построить доверительные интервалы для математического ожидания и среднего квадратичного отклонения.

2. Построим доверительный интервал для математического ожидания θ.

   Так как среднеквадратичное отклонение неизвестно используем центральную статистику.

   $$T(\vec{X}, \theta) = \frac{\theta - \overline{X}}{S(\vec{X})}\sqrt{n} \sim St(n-1)$$

   Следуя общей идее построения доверительных интервалов, выберем $\alpha_1, \alpha_2 > 0$ такие, что $\alpha_1 + \alpha_2 = 1 - \gamma$. В соответствии со свойствами непрерывных случайных величин можно записать

   $$\gamma = P\{t_{\alpha_1} < T(\vec{X}, \theta) < t_{1-\alpha_2}\}$$

   где t_{α_1}, t_{α_2} – квантили соответствующих уровней распределения Стьюдента с $n - 1 = 9$ степенями свободы.

   Поскольку размах доверительного интервала обычно стараются минимизировать, то выбирают $\alpha_1 = \alpha_2 = \frac{1-\gamma}{2}$, поэтому $t_{1-\alpha_2} = t_{1-\frac{1-\gamma}{2}} = t_{\frac{1+\gamma}{2}}$. В силу симметричности функции плотности стандартного нормального распределения заключаем, что $t_{\alpha_1} = -t_{\alpha_2} = -t_{\frac{1+\gamma}{2}}$

Тогда

$$\gamma = P\Big\{ -t_{\frac{1+\gamma}{2}} < \frac{\theta - \overline{X}}{S(\vec{X})}\sqrt{n} < t_{\frac{1+\gamma}{2}} \Big\}$$

Или

$$\gamma = P\Big\{ \overline{X} - \frac{t_{\frac{1+\gamma}{2}} S(\vec{X})}{\sqrt{n}} < \theta < \overline{X} + \frac{t_{\frac{1+\gamma}{2}} S(\vec{X})}{\sqrt{n}} \Big\}$$

Тогда в качестве верхней и нижней границ γ-доверительного интервала для параметра θ могут быть использованы статистики:

$$\underline{\theta}(\vec{X}) = \overline{X} - \frac{t_{\frac{1+\gamma}{2}} S(\vec{X})}{\sqrt{n}},$$

$$\overline{\theta}(\vec{X}) = \overline{X} + \frac{t_{\frac{1+\gamma}{2}} S(\vec{X})}{\sqrt{n}}$$

Вычислим:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = -0.09$$

$$\frac{1+\gamma}{2} = \frac{1+0.9}{2} = 0.95$$

$$t_{\frac{1+\gamma}{2}} = t_{0.95} = 1.8331$$

$$S(\vec{X}) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = 166.78$$

$$\underline{\theta}(\vec{X}) = \overline{X} - \frac{t_{\frac{1+\gamma}{2}} S(\vec{X})}{\sqrt{n}} = -0.09 - \frac{1.8331 \cdot 166.78}{\sqrt{10}} = -96.77$$

$$\overline{\theta}(\vec{X}) = \overline{X} + \frac{t_{\frac{1+\gamma}{2}} S(\vec{X})}{\sqrt{n}} = -0.09 + \frac{1.8331 \cdot 166.78}{\sqrt{10}} = 96.59$$

3. Построим доверительный интервал для среднеквадратичного отклонения σ.

   Используем центральную статистику.

$$T(\vec{X}, \sigma^2) = \frac{S^2(\vec{X})}{\sigma^2}(n-1) \sim \chi^2(n-1)$$

   Следуя общей идее построения доверительных интервалов, выберем $\alpha_1, \alpha_2 > 0$ такие, что $\alpha_1 + \alpha_2 = 1 - \gamma$. В соответствии со свойствами непрерывных случайных величин можно записать

$$\gamma = P\{h_{\alpha_1} < T(\vec{X}, \sigma^2) < h_{1-\alpha_2}\}$$

   где h_{α_1}, h_{α_2} – квантили соответствующих уровней распределения хи-квадрат с $n - 1 = 9$ степенями свободы.

   Поскольку размах доверительного интервала обычно стараются минимизировать, то выбирают $\alpha_1 = \alpha_2 = \frac{1-\gamma}{2}$, поэтому $h_{1-\alpha_2} = h_{1-\frac{1-\gamma}{2}} = h_{\frac{1+\gamma}{2}}$.
   Тогда $h_{\alpha_1} = h_{\frac{1-\gamma}{2}}$, $h_{\alpha_2} = h_{\frac{1+\gamma}{2}}$, так как график функции плотности не симметричен.

   Тогда

$$\gamma = P\left\{h_{\frac{1-\gamma}{2}} < \frac{S^2(\vec{X})}{\sigma^2}(n-1) < h_{\frac{1+\gamma}{2}}\right\}$$

   Или

$$\gamma = P\left\{\frac{1}{h_{\frac{1+\gamma}{2}}} < \frac{\sigma^2}{S^2(\vec{X})(n-1)} < \frac{1}{h_{\frac{1-\gamma}{2}}}\right\}$$

$$\gamma = P\left\{\frac{S^2(\vec{X})(n-1)}{h_{\frac{1+\gamma}{2}}} < \sigma^2 < \frac{S^2(\vec{X})(n-1)}{h_{\frac{1-\gamma}{2}}}\right\}$$

   Тогда в качестве верхней и нижней границ γ-доверительного интервала для параметра σ^2 могут быть использованы статистики:

$$\underline{\sigma}^2(\vec{X}) = \frac{S^2(\vec{X})(n-1)}{h_{\frac{1+\gamma}{2}}},$$

$$\overline{\sigma}^2(\vec{X}) = \frac{S^2(\vec{X})(n-1)}{h_{\frac{1-\gamma}{2}}}$$

Вычислим:

$$\frac{1+\gamma}{2} = \frac{1+0.9}{2} = 0.95$$

$$h_{\frac{1+\gamma}{2}} = h_{0.95} = 16.92$$

$$\frac{1-\gamma}{2} = \frac{1+0.9}{2} = 0.05$$

$$h_{\frac{1-\gamma}{2}} = h_{0.05} = 3.33$$

$$S(\vec{X}) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = 166.78$$

$$S^2(\vec{X}) = 27815.57$$

$$\underline{\sigma}^2(\vec{X}) = \frac{S^2(\vec{X})(n-1)}{h_{\frac{1+\gamma}{2}}} = \frac{27815.57 \cdot 9}{16.92} = 14795.52$$

$$\overline{\sigma}^2(\vec{X}) = \frac{S^2(\vec{X})(n-1)}{h_{\frac{1-\gamma}{2}}} = \frac{27815.57 \cdot 9}{3.33} = 75177.22$$

Ответ:

Доверительные интервалы уровня 0.9 для математического ожидания (-96.77, 96.59) и среднего квадратичного отклонения (14795.52, 75177.22).