

Теория вероятностей и математическая статистика.

Нулевой вариант экзамена

Автор: Сурова София, БПИ191

декабрь 2022

Задача 1.

Случайный вектор $\xi = (\xi_1, \xi_2)^T$ имеет функцию распределения

$$F(x, y) = \begin{cases} 1 + e^{-(\lambda x + \mu y)} - e^{-\lambda x} - e^{-\mu y}, & x \geq 0, y \geq 0 \\ 0, & \text{иначе} \end{cases}$$

Найти математическое ожидание и ковариационную матрицу вектора ξ , исследовать случайные величины ξ_1 и ξ_2 на независимость и некоррелированность. Найти вероятность попадания ξ в область $A = \{|x| \leq 1, |y| \leq 1\}$

Решение

Совместная плотность распределения - вторая смешанная производная совместной функции распределения

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y} = \frac{\partial^2 F(x, y)}{\partial y \partial x}$$

Найдём совместную плотность распределения

$$t(x, y) = \frac{\partial F(x, y)}{\partial y} = \begin{cases} -\mu e^{-\lambda x - \mu y} + \mu e^{-\mu y}, & x \geq 0, y \geq 0 \\ 0, & \text{иначе} \end{cases}$$

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial y \partial x} = \frac{\partial t(x, y)}{\partial x} = \begin{cases} \lambda \mu e^{-\lambda x - \mu y}, & x \geq 0, y \geq 0 \\ 0, & \text{иначе} \end{cases}$$

Найдём одномерные плотности распределения f_{ξ_1}, f_{ξ_2}

$$f_{\xi_1}(x) = \int_{-\infty}^{+\infty} f(x, y) dy = \int_0^{+\infty} \lambda \mu e^{-\lambda x - \mu y} dy = \lambda \mu \int_0^{+\infty} \frac{e^{-\lambda x - \mu y}}{-\mu} d(-\lambda x - \mu y) = -\lambda e^{-\lambda x - \mu y} \Big|_0^{+\infty} = \lambda e^{-\lambda x}, \quad x \geq 0$$

$$f_{\xi_1}(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

$$f_{\xi_2}(y) = \int_{-\infty}^{+\infty} f(x, y) dx = \int_0^{+\infty} \lambda \mu e^{-\lambda x - \mu y} dx = \lambda \mu \int_0^{+\infty} \frac{e^{-\lambda x - \mu y}}{-\lambda} d(-\lambda x - \mu y) = -\mu e^{-\lambda x - \mu y} \Big|_0^{+\infty} = \mu e^{-\mu y}, \quad y \geq 0$$

$$f_{\xi_2}(y) = \begin{cases} \mu e^{-\mu y}, & y \geq 0 \\ 0, & y < 0 \end{cases}$$

Математическое ожидание непрерывного случайного вектора

Математическое ожидание случайного вектора - это вектор математических ожиданий элементов случайного вектора

$$E\xi = (E\xi_1, E\xi_2)$$

Найдём математическое ожидание непрерывной случайной величины ξ_1

$$E\xi_1 = \int_{-\infty}^{+\infty} x f_{\xi_1}(x) dx = \int_0^{+\infty} \lambda x e^{-\lambda x} dx = \int_0^{+\infty} \frac{\lambda x}{-\lambda} d(e^{-\lambda x}) = - \int_0^{+\infty} x d(e^{-\lambda x}) = -(x e^{-\lambda x} \Big|_0^{+\infty} - \int_0^{+\infty} e^{-\lambda x} dx) =$$

$$= \int_0^{+\infty} e^{-\lambda x} dx = \int_0^{+\infty} \frac{e^{-\lambda x}}{-\lambda} d(-\lambda x) = -\frac{e^{-\lambda x}}{\lambda} \Big|_0^{+\infty} = \frac{1}{\lambda}$$

Аналогично найдём математическое ожидание непрерывной случайной величины ξ_2

$$\begin{aligned} E\xi_2 &= \int_{-\infty}^{+\infty} y f_{\xi_2}(y) dy = \int_0^{+\infty} \mu y e^{-\mu y} dy = \int_0^{+\infty} \frac{\mu y}{-\mu} d(e^{-\mu y}) = - \int_0^{+\infty} y d(e^{-\mu y}) = -(y e^{-\mu y} \Big|_0^{+\infty} - \int_0^{+\infty} e^{-\mu y} dy) = \\ &= \int_0^{+\infty} e^{-\mu y} dy = \int_0^{+\infty} \frac{e^{-\mu y}}{-\mu} d(-\mu y) = -\frac{e^{-\mu y}}{\mu} \Big|_0^{+\infty} = \frac{1}{\mu} \end{aligned}$$

Тогда математическое ожидание случайного вектора

$$E\xi = \left(\frac{1}{\lambda}, \frac{1}{\mu} \right)$$

Примечание. Вы огромный молодец, если учили основные распределения и сразу увидели, что это плотность распределения похожа на экспоненциальное распределение. Вы дважды огромный молодец, если учили и математические ожидания и дисперсии для основных распределений, так как знаете, что $E\xi_1 = \frac{1}{\lambda}$, $D\xi_1 = \frac{1}{\lambda^2}$, $E\xi_2 = \frac{1}{\mu}$, $D\xi_2 = \frac{1}{\mu^2}$. Но в предположении, что не все учили основные распределения, разберём, как найти математическое ожидание и дисперсию у неизвестной непрерывной случайной величины.

Ковариационная матрица случайного вектора

$$K = \begin{pmatrix} cov(\xi_1, \xi_1) & cov(\xi_1, \xi_2) \\ cov(\xi_2, \xi_1) & cov(\xi_2, \xi_2) \end{pmatrix} = \begin{pmatrix} D\xi_1 & cov(\xi_1, \xi_2) \\ cov(\xi_1, \xi_2) & D\xi_2 \end{pmatrix}$$

Найдём дисперсию непрерывной случайной величины ξ_1

$$\begin{aligned} E\xi_1^2 &= \int_{-\infty}^{+\infty} x^2 f_{\xi_1}(x) dx = \int_0^{+\infty} \lambda x^2 e^{-\lambda x} dx = \int_0^{+\infty} \frac{\lambda x^2}{-\lambda} d(e^{-\lambda x}) = - \int_0^{+\infty} x^2 d(e^{-\lambda x}) = -(x^2 e^{-\lambda x} \Big|_0^{+\infty} - \int_0^{+\infty} e^{-\lambda x} d(x^2)) = \\ &= \int_0^{+\infty} 2x e^{-\lambda x} dx = -\frac{2}{\lambda} \int_0^{+\infty} x d(e^{-\lambda x}) = -\frac{2}{\lambda} \int_0^{+\infty} x d(e^{-\lambda x}) = -\frac{2}{\lambda} (x e^{-\lambda x} \Big|_0^{+\infty} - \int_0^{+\infty} e^{-\lambda x} dx) = \frac{2}{\lambda} \cdot \frac{1}{\lambda} = \frac{2}{\lambda^2} \\ D\xi_1 &= E\xi_1^2 - (E\xi_1)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2} \end{aligned}$$

Найдём дисперсию непрерывной случайной величины ξ_2

$$\begin{aligned} E\xi_2^2 &= \int_{-\infty}^{+\infty} y^2 f_{\xi_2}(y) dy = \int_0^{+\infty} \mu y^2 e^{-\mu y} dy = \int_0^{+\infty} \frac{\mu y^2}{-\mu} d(e^{-\mu y}) = - \int_0^{+\infty} y^2 d(e^{-\mu y}) = -(y^2 e^{-\mu y} \Big|_0^{+\infty} - \int_0^{+\infty} e^{-\mu y} d(y^2)) = \\ &= \int_0^{+\infty} 2y e^{-\mu y} dy = -\frac{2}{\mu} \int_0^{+\infty} y d(e^{-\mu y}) = -\frac{2}{\mu} \int_0^{+\infty} y d(e^{-\mu y}) = -\frac{2}{\mu} (y e^{-\mu y} \Big|_0^{+\infty} - \int_0^{+\infty} e^{-\mu y} dy) = \frac{2}{\mu} \frac{1}{\mu} = \frac{2}{\mu^2} \\ D\xi_2 &= E\xi_2^2 - (E\xi_2)^2 = \frac{2}{\mu^2} - \frac{1}{\mu^2} = \frac{1}{\mu^2} \end{aligned}$$

Чтобы не считать ковариацию, попробуем посмотреть на независимость случайных величин. Если величины независимы, то ковариация равна 0. Заодно и проверим величины на независимость.

$$\xi_1, \xi_2 \text{ независимы} \Leftrightarrow f(x, y) = f_{\xi_1}(x) f_{\xi_2}(y)$$

$$f_{\xi_1}(x) f_{\xi_2}(y) = \begin{cases} \lambda e^{-\lambda x} \mu e^{-\mu y}, & x \geq 0, y \geq 0 \\ 0, & \text{иначе} \end{cases} = \begin{cases} \lambda \mu e^{-\lambda x - \mu y}, & x \geq 0, y \geq 0 \\ 0, & \text{иначе} \end{cases} = f(x, y) \Rightarrow \xi_1, \xi_2 \text{ независимы}$$

$$\xi_1, \xi_2 \text{ независимы} \Rightarrow cov(\xi_1, \xi_2) = 0 \quad (\text{обратное неверно})$$

$$K = \begin{pmatrix} \frac{1}{\lambda^2} & 0 \\ 0 & \frac{1}{\mu^2} \end{pmatrix}$$

Исследование на коррелированность и независимость

Как уже определили ранее

$$f(x, y) = f_{\xi_1}(x)f_{\xi_2}(y) \Rightarrow \xi_1, \xi_2 \text{ независимы}$$

$$\xi_1, \xi_2 \text{ независимы} \Rightarrow \text{cov}(\xi_1, \xi_2) = 0 \Rightarrow \xi_1, \xi_2 \text{ некоррелированные}$$

Примечание. Разберёмся, что делать в иных случаях? Пользоваться следующими свойствами.

$$f(x, y) = f_X(x)f_Y(y) \Rightarrow X, Y \text{ независимы}$$

$$f(x, y) \neq f_X(x)f_Y(y) \Rightarrow X, Y \text{ зависимы}$$

$$X, Y \text{ независимы} \Rightarrow \text{cov}(X, Y) = 0$$

$$\text{cov}(X, Y) \neq 0 \Rightarrow X, Y \text{ зависимы}$$

$$\text{cov}(X, Y) > 0 \Rightarrow X, Y \text{ коррелированные, положительно коррелированные}$$

$$\text{cov}(X, Y) = 0 \Rightarrow X, Y \text{ некоррелированные}$$

$$\text{cov}(X, Y) < 0 \Rightarrow X, Y \text{ коррелированные, отрицательно коррелированные}$$

Вероятность попадания случайного вектора в область

$$\begin{aligned} P(\xi \in A) &= \int_{-1}^1 dx \int_{-1}^1 f(x, y) dy = \{ \text{т.к. СВ независимы} \} = \int_0^1 f_{\xi_1}(x) dx \int_0^1 f_{\xi_2}(y) dy = \int_0^1 \lambda e^{-\lambda x} dx \int_0^1 \mu e^{-\mu y} dy = \\ &= \left[-\frac{1}{\lambda} \int_0^1 \lambda e^{-\lambda x} d(-\lambda x) \right] \left[-\frac{1}{\mu} \int_0^1 \mu e^{-\mu y} d(-\mu y) \right] = -e^{-\lambda x} \Big|_0^1 - e^{-\mu y} \Big|_0^1 = (1 - e^{-\lambda})(1 - e^{-\mu}) \end{aligned}$$

Ответ:

$$E\xi = (1/\lambda, 1/\mu), K = \begin{pmatrix} \frac{1}{\lambda^2} & 0 \\ 0 & \frac{1}{\mu^2} \end{pmatrix}, \xi_1, \xi_2 \text{ независимые и некоррелированные, } P(\xi \in A) = (1 - e^{-\lambda})(1 - e^{-\mu})$$

Задача 2

В продукции цеха детали отличного качества составляют 80%. В каких пределах с вероятностью 0.99 будет находиться количество деталей отличного качества, если взять 10000 деталей? Построить оценку с помощью неравенства Чебышёва и по теореме Муавра-Лапласа.

Решение

Пусть ξ - количество деталей отличного качества, тогда $\xi \sim \mathbb{B}\square(10000, 0.8)$.

Примечание. Нельзя рассматривать распределение Пуассона, так как хоть число испытаний и большое, но вероятность успеха не мала.

Первая форма неравенства Чебышёва

Случайная величина ξ принимает неотрицательные значения, и мат. ожидание конечно $E\xi = np$, тогда

$$P(\xi < \epsilon) \geq 1 - \frac{E\xi}{\epsilon}$$

$$1 - \frac{E\xi}{\epsilon} = 0.99 \Rightarrow \epsilon = \frac{E\xi}{0.01} = 100E\xi = 100np = 800000$$

То есть первая форма неравенства Чебышёва даёт следующую оценку: с вероятностью не ниже 0.99 количество деталей отличного качества находится в пределах от 0 до 800000.

Вторая форма неравенства Чебышёва

Мы знаем дисперсию ξ , она конечна $D\xi = npq = 10000 * 0.8 * 0.2 = 1600$, тогда

$$P(|\xi - E\xi| < \epsilon) \geq 1 - \frac{D\xi}{\epsilon^2}$$

$$1 - \frac{D\xi}{\epsilon^2} = 0.99 \Rightarrow \epsilon = \sqrt{\frac{D\xi}{0.01}} = \sqrt{100D\xi} = 10\sqrt{1600} = 10 * 40 = 400$$

$$P(|\xi - E\xi| < \epsilon) = P(-400 < \xi - 8000 < 400) = P(7600 < \xi < 8400)$$

То есть вторая форма неравенства Чебышёва даёт более точную оценку: с вероятностью не ниже 0.99 количество деталей отличного качества находится в пределах от 7600 до 8400.

Интегральная теорема Муавра-Лапласа

Интегральная теорема Муавра-Лапласа даёт аппроксимацию биномиального распределения к нормальному закону при большом числе испытаний

$$P(a < \xi < b) = \Phi_0\left(\frac{b - np}{\sqrt{npq}}\right) - \Phi_0\left(\frac{a - np}{\sqrt{npq}}\right)$$

$$P(|\xi - E\xi| < \epsilon) = \Phi_0\left(\frac{\epsilon}{\sqrt{npq}}\right) - \Phi_0\left(\frac{-\epsilon}{\sqrt{npq}}\right) = 2\Phi_0\left(\frac{\epsilon}{\sqrt{npq}}\right)$$

$$2\Phi_0\left(\frac{\epsilon}{\sqrt{npq}}\right) = 0.99 \Rightarrow \frac{\epsilon}{\sqrt{npq}} = 2.58 \text{ по таблице функции Лапласа} \Rightarrow \epsilon = 2.58\sqrt{D\xi} = 2.58 * 40 = 103.2$$

$$P(|\xi - E\xi| < \epsilon) = P(-103.2 < \xi - 8000 < 103.2) = P(7896.8 < \xi < 8103.2)$$

То есть интегральная теорема Муавра-Лапласа даёт ещё более точную оценку: с вероятностью не ниже 0.99 количество деталей отличного качества находится в пределах от 7896.8 до 8103.2

Задача 3

Выборка X_1, \dots, X_n соответствует распределению Релея, плотность которого имеет вид

$$f(x) = \frac{2x}{\theta} \exp\left(-\frac{x^2}{\theta}\right), \text{ при } x > 0$$

Найдите оценку максимального правдоподобия параметра θ .

Решение

Начинаем с функции правдоподобия

$$L(X_1, \dots, X_n, \theta) = \prod_{i=1}^n f(x_i, \theta) = \prod_{i=1}^n \frac{2x_i}{\theta} \exp\left(-\frac{x_i^2}{\theta}\right) = \prod_{i=1}^n \frac{2}{\theta} \cdot \prod_{i=1}^n \exp\left(-\frac{x_i^2}{\theta}\right) \cdot \prod_{i=1}^n x_i = \frac{2^n}{\theta^n} \exp\left(-\frac{\sum_{i=1}^n x_i^2}{\theta}\right) \prod_{i=1}^n x_i$$

Найти оценку максимального правдоподобия можно, решив следующее уравнение

$$\frac{\partial \ln L(X_1, \dots, X_n, \theta)}{\partial \theta} = 0$$

Чтобы не ошибиться, сначала прологарифмируем, потом будем дифференцировать.



$$\ln L(X_1, \dots, X_n, \theta) = \ln\left(\frac{2^n}{\theta^n}\right) + \ln\left(\exp\left(-\frac{\sum_{i=1}^n x_i^2}{\theta}\right)\right) + \ln\left(\prod_{i=1}^n x_i\right) = n \ln 2 - n \ln \theta - \frac{\sum_{i=1}^n x_i^2}{\theta} + \sum_{i=1}^n \ln(x_i)$$
$$\frac{\partial(n \ln 2 - n \ln \theta - \frac{1}{\theta} \sum_{i=1}^n x_i^2 + \sum_{i=1}^n \ln(x_i))}{\partial \theta} = -\frac{n}{\theta} + \frac{1}{\theta^2} \sum_{i=1}^n x_i^2 = 0$$

Выразим параметр $\hat{\theta}$ из полученного уравнения и получим оценку максимального правдоподобия.

Примечание. Заметьте, что после дифференцирования и приравнивания к нулю параметр принарядился шляпой, теперь он не просто параметр, а оценка параметра. Оценку параметра от истинного значения параметра отличает то, что она носит шляпку.

$$-\frac{n}{\hat{\theta}} + \frac{1}{\hat{\theta}^2} \sum_{i=1}^n x_i^2 = 0 \Rightarrow -n\hat{\theta} + \sum_{i=1}^n x_i^2 = 0 \Rightarrow \hat{\theta} = \frac{1}{n} \sum_{i=1}^n x_i^2$$

Ответ: $\hat{\theta} = \frac{1}{n} \sum_{i=1}^n x_i^2$

Задача 4

Изучается зависимость между показателями X (оценка студента за контрольную работу по теории вероятностей) и Y (оценка студента за экзамен по программированию). Используя данные о показателях X и Y для группы 182, в которой обучается 30 человек, Кристиан Бенуа вычислил выборочный коэффициент корреляции и получил значение 0.45. Можно ли, опираясь на эти данные, считать, что показатели X и Y являются положительно коррелированными? Уровень значимости принять равным 0.05. Опишите процедуру проверки соответствующей гипотезы.

Решение

Нам необходимо проверить гипотезу о коррелированности двух величин.

1. Формулируем основную и альтернативную гипотезы

$$H_0 : \rho_{X,Y} = 0$$

$$H_A : \rho_{X,Y} > 0$$

2. Выбираем уровень значимости, в данном случае из условия

$$\alpha = 0.05$$

3. Выбираем статистику

$$T((X, Y)) = \frac{\sqrt{n-2} \hat{\rho}_{X,Y}}{\sqrt{1 - \hat{\rho}_{X,Y}^2}}$$

4. Определяем, какому распределению соответствует статистика при верной основной гипотезе

$$T((X, Y)) \Big|_{H_0} \sim t(n-2)$$

5. Строим доверительную и критическую области

$$\text{Д.О.} : (-\infty, t_{n-2, 1-\alpha}], \quad \text{К.О.} : (t_{n-2, 1-\alpha}; +\infty)$$

$$t_{n-2, 1-\alpha} = 1.7001$$

6. Вычисляем реализацию статистики на нашей выборке

$$t = \frac{\sqrt{30-2} * 0.45}{\sqrt{1 - 0.45^2}} \approx 2.6664$$

7. Определяем, куда попало значение статистики и делаем вывод

$$t = 2.6664 \in (1.7001, +\infty)$$

Значение статистики попало в критическую область, а значит мы отвергаем основную гипотезу в пользу альтернативной, то есть на уровне значимости 0.05 можно считать, что величины положительно коррелированы.

Задача 5

Из произведённой партии шоколадных батончиков случайным образом были выбраны шесть штук, вес которых (в граммах) составил 49.1; 50.3; 49.6; 51.2; 48.4; 49.8. Постройте центральный доверительный интервал уровня надёжности 0.95 для среднего веса батончика. Предполагается, что наблюдения имеют гауссовское распределение.

Решение

У нас есть выборка $X_1, \dots, X_6 \sim N(\theta_1, \theta_2^2)$, θ_1 и θ_2^2 неизвестны, но мы хотим оценить θ_1 с помощью доверительного интервала с уровнем надёжности $1 - \alpha$. Воспользуемся теоремой Фишера, и в качестве центральной статистики возьмём

$$G = \frac{(\bar{x} - \theta_1)\sqrt{n-1}}{\sqrt{s^2}} \sim t(n-1)$$

Тогда доверительный интервал

$$P(t_{n-1, \alpha/2} < G < t_{n-1, 1-\alpha/2}) = 1 - \alpha$$

$$P(t_{n-1, \alpha/2} < \frac{(\bar{x} - \theta_1)\sqrt{n-1}}{\sqrt{s^2}} < t_{n-1, 1-\alpha/2}) = 1 - \alpha$$

$$P\left(\bar{x} - \frac{\sqrt{s^2} \cdot t_{n-1, 1-\alpha/2}}{\sqrt{n-1}} < \theta_1 < \bar{x} + \frac{\sqrt{s^2} \cdot t_{n-1, 1-\alpha/2}}{\sqrt{n-1}}\right) = 1 - \alpha$$

Найдём выборочное математическое ожидание, выборочную смещённую дисперсию и квантили

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{49.1 + 50.3 + 49.6 + 51.2 + 48.4 + 49.8}{6} = \frac{298.4}{6} \approx 49.73$$

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{4.6734}{6} \approx 0.7789$$

$$t_{n-1, 1-\alpha/2} = 2.5706$$

Подставим найденные значения в доверительный интервал

$$\left(49.73 - \frac{\sqrt{0.7789} \cdot 2.5706}{\sqrt{5}}; 49.73 + \frac{\sqrt{0.7789} \cdot 2.5706}{\sqrt{5}}\right) \\ (48.7154; 50.7446)$$

Ответ: (48.7154; 50.7446)