Разбор нулевика к экзамену по теории вероятностей

2019-2020 учебный год

№ 1

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

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

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:

Найдем функцию плотности ξ :

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

$$f'_x = -\lambda e^{-(\lambda x + \mu y)} + \lambda e^{-\lambda x}$$
$$f''_{xy} = \lambda \mu e^{-(\lambda x + \mu y)}$$

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

Чтобы посчитать вектор математического ожидания, найдем еще функции плотности для каждого элемента случайного вектора:

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

Абсолютно аналогично считается и вторая функция плотности. $f_{\xi_2}(y) = \mu e^{-\mu y}$ при $y \geq 0$.

Теперь можно посчитать вектор математического ожидания, то есть посчитать математическое ожидание для ξ_1 и ξ_2:

$$E\xi_1 = \int_0^{+\infty} x \lambda e^{-\lambda x} dx = \int_0^{+\infty} x \lambda \left(-\frac{1}{\lambda}\right) e^{-\lambda x} d(-\lambda x) = -\int_0^{+\infty} x d e^{-\lambda x} = -\left(x e^{-\lambda x} \Big|_0^{+\infty} - \int_0^{+\infty} e^{-\lambda x} dx \right) =$$
$$= \int_0^{+\infty} \frac{e^{-\lambda x}}{-\lambda} d(-\lambda x) = -\frac{1}{\lambda} e^{-\lambda x} \Big|_0^{+\infty} = \frac{1}{\lambda}$$

Аналогично для ξ_2 : $E\xi_2 = \dfrac{1}{\mu}$

Прежде чем считать ковариационную матрицу, есть смысл посмотреть на независимость ξ_1 и ξ_2, так как если они независимы, то вне главной диагонали ковариационной матрицы будут стоять нули. Можно проверить это через функцию распределения:

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

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

Отсюда также следует, что ξ_1 и ξ_2 некоррелируемы.

Посчитаем дисперсии ξ_1 и ξ_2:
$D\xi_1 = E\xi_1^2 - (E\xi_1)^2$.
$E\xi_1$ уже посчитано, осталось посчитать $E\xi_1^2$.

$$E\xi_1^2 = \int\limits_0^{+\infty} x^2\lambda e^{-\lambda x}dx = \int\limits_0^{+\infty} x^2\lambda\left(-\frac{1}{\lambda}\right)e^{-\lambda x}d(-\lambda x) = -\int\limits_0^{+\infty} x^2 de^{-\lambda x} = -\left(x^2e^{-\lambda x}\Big|_0^{+\infty} - \int\limits_0^{+\infty} 2xe^{-\lambda x}dx\right) =$$

$$= \int\limits_0^{+\infty}\frac{2xe^{-\lambda x}}{-\lambda}d(-\lambda x) = -\frac{2}{\lambda}\int\limits_0^{+\infty} xde^{-\lambda x} = -\frac{2}{\lambda}\left(xe^{-\lambda x}\Big|_0^{+\infty} - \int\limits_0^{+\infty} e^{-\lambda x}dx\right) = \frac{2}{\lambda}\int\limits_0^{+\infty}\frac{e^{-\lambda x}}{-\lambda}d(-\lambda x) =$$

$$= -\frac{2}{\lambda^2}\int\limits_0^{+\infty} e^{-\lambda x}\Big|_0^{+\infty} = \frac{2}{\lambda^2}$$

Таким образом, $D\xi_1 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$

Аналогично $D\xi_2 = \frac{1}{\mu^2}$

Значит, ковариационная матрица будет иметь вид:

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

Осталось только посчиатть вероятность попадания в область А. Надо взять двойной интеграл:

$$\int\limits_0^1\int\limits_0^1 f(x,y)dxdy$$

Так как величины независимы, то этот интеграл принимает следующую форму:

$$\int\limits_0^1 f_{\xi_1}(x)dx\cdot\int\limits_0^1 f_{\xi_2}(y)dy = \int\limits_0^1 \lambda e^{-\lambda x}dx\cdot\int\limits_0^1 \mu e^{-\mu x}dy = \int\limits_0^1 \lambda\left(-\frac{1}{\lambda}\right)e^{-\lambda x}d(-\lambda x)\cdot\int\limits_0^1 \mu\left(-\frac{1}{\mu}\right)e^{-\mu x}d(-\mu y) =$$

$$= -e^{-\lambda x}\Big|_0^1\cdot\left(-e^{-\mu x}\Big|_0^1\right) = (1-e^{-\lambda})(1-e^{-\mu})$$

Ответ: $E\xi = (\frac{1}{\lambda}, \frac{1}{\mu})^T$; $K_\xi = \begin{pmatrix} \frac{1}{\lambda^2} & 0 \\ 0 & \frac{1}{\mu^2} \end{pmatrix}$; ξ_1 и ξ_2 независимы и некоррелируемы;
$P(\xi \in A) = (1-e^{-\lambda})(1-e^{-\mu})$

№ 2

Задание: 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:
Определим СВ $\xi = \{$количество деталей отличного качества среди 10000$\}$. Эта СВ описывается биномиальным законом: $\xi \sim Bi(10000;\ 0{,}8)$.
Используя неравенство Чебышева, оценим количество деталей отличного качества:

$$P(|\xi - E\xi| \geq \varepsilon) \leq \frac{D\xi}{\varepsilon^2} \Leftrightarrow$$

$$\Leftrightarrow P(|\xi - E\xi| \leq \varepsilon) \geq 1 - \frac{D\xi}{\varepsilon^2}$$

Подставим числа из условия:

$$P(|\xi - 8000| \leq \varepsilon) \geq 1 - \frac{1600}{\varepsilon^2} = 0{,}99$$

Отсюда $\varepsilon = 400$. Тогда СВ лежит в границах:
$|\xi - 8000| \le 400 \Leftrightarrow 7600 \le \xi \le 8400$

Оценим количество деталей отличного качества при помощи теоремы Муавра-Лапласа. Так как общее количество деталей достаточно большое, можно сказать, что СВ ξ имеет нормальное распределение: $\xi \sim N(8000{,}1600)$. Посчиатем вероятность:
$P(|\xi - m| < \delta) = 2\Phi_0(\dfrac{\delta}{\sigma})$
Подставим числа:
$P(|\xi - 8000| < \delta) = 2\Phi_0(\dfrac{\delta}{40}) = 0{,}99$
$\Phi_0(\dfrac{\delta}{40}) = 0{,}495$
$\dfrac{\delta}{40} = 2{,}58$

$\delta = 103{,}2$
Тогда СВ лежит в границах:
$|\xi - 8000| < 103{,}2 \Leftrightarrow 7896{,}8 < \xi < 8103{,}2$

Ответ: По Чебышеву: $7600 \le \xi \le 8400$
По Муавру-Лапласу: $7896{,}8 < \xi < 8103{,}2$

№ 3

Задание: Выборка $X_1, \ldots, X_n$ соответствует распределению Релея, плотность которого имеет вид $f(x) = \dfrac{2x}{\Theta} \exp\left(\dfrac{-x^2}{\Theta}\right)$ при $x > 0$. Найдите оценку максимального правдоподобия параметра Θ. Докажите несмещённость и состоятельность этой оценки.
Решение:
Посчитаем функцию правдоподобия:

$$L(X_1, \ldots, X_n, \Theta) = \prod_{i=1}^{n} \frac{2}{\Theta} X_i \exp\left(\frac{-X_i^2}{\Theta}\right) = \frac{2^n}{\Theta^n} \exp\left(-\frac{\sum\limits_{i=1}^{n} X_i^2}{\Theta}\right) \prod_{i=1}^{n} X_i$$

Теперь прологарифмируем ее и продифференцируем по параметру Θ:

$$\ln L(X_1, \ldots, X_n, \Theta) = n \ln 2 - n \ln \Theta - \frac{\sum\limits_{i=1}^{n} X_i^2}{\Theta} + \sum_{i=1}^{n} \ln X_i$$

$$\frac{\partial \ln L(X_1, \ldots, X_n, \Theta)}{\partial \Theta} = -\frac{n}{\hat{\Theta}} + \sum_{i=1}^{n} X_i^2 \cdot \frac{1}{\hat{\Theta}^2} = 0$$

$$-n\hat{\Theta} + \sum_{i=1}^{n} X_i^2 = 0$$

$$\hat{\Theta} = \frac{\sum\limits_{i=1}^{n} X_i^2}{n}$$

Проверим несмещенность этой оценки. Оценка является несмещенной, если $E\hat{\Theta} = \Theta :$

$$E\hat{\Theta} = E\frac{\sum\limits_{i=1}^{n} X_i^2}{n} = \frac{1}{n} E \sum_{i=1}^{n} X_i^2 = (\text{так как они все независимы}) = \frac{1}{n} \sum_{i=1}^{n} EX_i^2 =$$

$$= (\text{так как они все одинаково распределены}) = EX_1^2 = \int\limits_0^{+\infty} X_1^2 \frac{2X_1}{\Theta} \exp\left(-\frac{X_1^2}{\Theta}\right) dX_1 =$$

$$= \int\limits_0^{+\infty} \frac{X_1^2}{\Theta} \exp\left(-\frac{X_1^2}{\Theta}\right) dX_1^2 = \int\limits_0^{+\infty} (-\Theta) \frac{X_1^2}{\Theta} \exp\left(-\frac{X_1^2}{\Theta}\right) d\left(-\frac{X_1^2}{\Theta}\right) = -\int\limits_0^{+\infty} X_1^2 d\exp\left(-\frac{X_1^2}{\Theta}\right) =$$

$$= -\left(X_1^2 \exp\left(-\frac{X_1^2}{\Theta}\right)\Bigg|_0^{+\infty} - \int\limits_0^{+\infty} \exp\left(-\frac{X_1^2}{\Theta}\right) dX_1^2\right) = \int\limits_0^{+\infty} (-\Theta) \exp\left(-\frac{X_1^2}{\Theta}\right) d\left(-\frac{X_1^2}{\Theta}\right) =$$

$$= -\Theta \exp\left(-\frac{X_1^2}{\Theta}\right)\Bigg|_0^{+\infty} = \Theta \Rightarrow \hat{\Theta} \text{ несмещенная}$$

Проверим состоятельность оценки. Оценка является состоятельной, если $\hat{\Theta} \xrightarrow[n\to\infty]{P} \Theta$

Другими словами, $P(|\hat{\Theta} - \Theta| \le \varepsilon) \to 1$

$P(|\hat{\Theta} - E\hat{\Theta}| \le \varepsilon) \to 1$

По неравенству Чебышева

$$P(|\hat{\Theta} - E\hat{\Theta}| \le \varepsilon) \ge 1 - \frac{D\hat{\Theta}}{\varepsilon^2}$$

$$D\hat{\Theta} = D\frac{\sum\limits_{i=1}^n X_i^2}{n} = \frac{1}{n^2} D\sum\limits_{i=1}^n X_i^2 = \frac{1}{n^2}\sum\limits_{i=1}^n DX_i^2 = \frac{1}{n} DX_1^2 = \frac{1}{n}(EX_1^4 - (EX_1^2)^2)$$

EX_1^2 посчитано выше, осталось посчитать только EX_1^4.

$$EX_1^4 = \int\limits_0^{+\infty} X_1^4 \frac{2X_1}{\Theta} \exp\left(-\frac{X_1^2}{\Theta}\right) dX_1 = \int\limits_0^{+\infty} \frac{X_1^4}{\Theta} \exp\left(-\frac{X_1^2}{\Theta}\right) dX_1^2 = \int\limits_0^{+\infty} (-\Theta) \frac{X_1^4}{\Theta} \exp\left(-\frac{X_1^2}{\Theta}\right) d\left(-\frac{X_1^2}{\Theta}\right) =$$

$$= -\int\limits_0^{+\infty} X_1^4 d\exp\left(-\frac{X_1^2}{\Theta}\right) = -\left(X_1^4 \exp\left(-\frac{X_1^2}{\Theta}\right)\Bigg|_0^{+\infty} - \int\limits_0^{+\infty} \exp\left(-\frac{X_1^2}{\Theta}\right) dX_1^4\right) =$$

$$= 2\int\limits_0^{+\infty} X_1^2 \exp\left(-\frac{X_1^2}{\Theta}\right) dX_1^2 = -2\Theta \int\limits_0^{+\infty} X_1^2 d\exp\left(-\frac{X_1^2}{\Theta}\right) =$$

$$= -2\Theta\left(X_1^2 \exp\left(-\frac{X_1^2}{\Theta}\right)\Bigg|_0^{+\infty} - \int\limits_0^{+\infty} \exp\left(-\frac{X_1^2}{\Theta}\right) dX_1^2\right) = 2\Theta \int\limits_0^{+\infty} (-\Theta) \exp\left(-\frac{X_1^2}{\Theta}\right) d\left(-\frac{X_1^2}{\Theta}\right) =$$

$$= -2\Theta^2 \exp\left(-\frac{X_1^2}{\Theta}\right)\Bigg|_0^{+\infty} = 2\Theta^2$$

$$D\hat{\Theta} = \frac{1}{n}(EX_1^4 - (EX_1^2)^2) = \frac{1}{n}(2\Theta^2 - \Theta^2) = \frac{\Theta^2}{n} \xrightarrow{n\to\infty} 0$$

$$\Rightarrow P(|\hat{\Theta} - E\hat{\Theta}| \le \varepsilon) \to 1 \Rightarrow \hat{\Theta} \text{ состоятельная}$$

Ответ: $\hat{\Theta} = \dfrac{\sum\limits_{i=1}^n X_i^2}{n}$

№ 4

Задание: В результате проведенного исследования было установлено, что у 309 светлоглазых мужчин жены также имеют светлые глаза, а у 214 светлоглазых мужчин жены темноглазые. У 119 темноглазых мужчин жены также темноглазые, а у 132 темноглазых мужчин жены светлоглазые. Имеется ли зависимость между цветом глаз мужей и их жен?

Решение:

Пусть $\xi_1 = \{\text{количество светлоглазых мужей}\}, \xi_2 = \{\text{количество светлоглазых жен}\}$

ξ_2/ξ_1	0	1
0	309/774	214/774
1	132/774	119/774

ξ_1	0	1
P	441/774	333/774

ξ_2	0	1
P	523/774	251/774

Компоненты дискретного вектора независимы $\Leftrightarrow p_{ij} = p_{i\cdot}p_{\cdot j} \quad \forall i = \overline{1,2}, j = \overline{1,2}$

Рассмотрим $p_{11} = \dfrac{309}{774}$, при этом $p_{\cdot 1} = \dfrac{441}{774}, p_{1\cdot} = \dfrac{523}{774}$

$p_{11} \neq p_{\cdot 1} \cdot p_{1\cdot} \Rightarrow \xi_1$ и ξ_2 зависимые

Ответ: Имеется зависимость

№ 5

Задание: 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. Проверьте гипотезу о том, что средний вес батончика составляет 50 г. Уровень значимости принять равным 0.1. Предполагается, что наблюдения имеют гауссовское распределение.
Решение:
Вначале проведем вспомогательные вычисления, то есть посчитаем выборочное математическое ожидание и несмещенную выборочную дисперсию:
$\overline{X} = \dfrac{1}{6}\sum\limits_{i=1}^{6} X_i = \dfrac{1}{6}(49{,}1 + 50{,}3 + 49{,}6 + 51{,}2 + 48{,}4 + 49{,}8) = \dfrac{298{,}4}{6} = 49{,}73$
$\tilde{S}^2 = \dfrac{1}{5}\sum\limits_{i=1}^{6}(X_i - 49{,}73)^2 = \dfrac{1}{5}(0{,}3969 + 0{,}3249 + 0{,}0169 + 2{,}1609 + 1{,}7689 + 0{,}0049) = \dfrac{4{,}6734}{5} = 0{,}9347$
$\tilde{S} = \sqrt{0{,}9347} = 0{,}9668$

Так как неизвестно ни математическое ожидание рассматриваемого распределения, ни дисперсия, то доверительный интервал уровня надежности 0,95 будет иметь следующий вид:
$P(\overline{X} - \dfrac{\tilde{S}t_{5;0,975}}{\sqrt{6}} < \Theta_1 < \overline{X} + \dfrac{\tilde{S}t_{5;0,975}}{\sqrt{6}}) = 0{,}95$
$P(49{,}73 - \dfrac{0{,}9668 \cdot 2{,}5706}{2{,}4495} < \Theta_1 < 49{,}73 + \dfrac{0{,}9668 \cdot 2{,}5706}{2{,}4495}) = 0{,}95$

$P(48{,}7154 < \Theta_1 < 50{,}7446) = 0{,}95$
Доверительный интервал уровня надежности 0,95 равен $(48{,}7154; 50{,}7446)$

Проверим гипотезу о среднем весе батончика. Выборка распределена по гауссовскому закону, но ни математическое ожидание, ни дисперсия неизвестны.
$X_1, \ldots, X_6 \sim N(m, \sigma^2)$
Пройдем по всем пунктам проверки гипотезы.
1. Основная гипотеза: $H_0 : m = 50$, альтернативная гипотеза: $H_1 : m \neq 50$
2. $\alpha = 0{,}1$
3. $T(X) = \dfrac{(\overline{X} - 50)\sqrt{6}}{\tilde{S}} = \dfrac{(49{,}73 - 50)\sqrt{6}}{0{,}9668} = -\dfrac{0{,}6614}{0{,}9668} = -0{,}684$

4. $T(X)|_{H_0} \sim t(6)$

5. Доверительная область лежит в границах от $-t_{5;0,95}$ до $t_{5;0,95}$, то есть от $-2{,}015$ до $2{,}015$. Полученное в пункте 3 значение попадает в эту область, значит, гипотеза верна на уровне значимости 0,1.

Ответ: Доверительный интервал: $(48{,}7154; 50{,}7446)$; гипотеза верна.