ПРИМЕРНЫЙ ВАРИАНТ КОНТРОЛЬНОЙ РАБОТЫ ПО МАТЕМАТИЧЕСКОЙ СТАТИСТИКЕ

1. Выборка $X_1, \dots, X_n$ соответствует распределению Релея, плотность которого имеет вид $f(x) = \frac{2x}{\theta} \exp\left(\frac{-x^2}{\theta}\right)$ *при* $x > 0$. Найдите оценку максимального правдоподобия параметра θ. Докажите несмещённость и состоятельность этой оценки.
2. Выборка $X_1, \dots, X_n$соответствует распределению Пуассона $П(\theta)$. Пользуясь критерием эффективности, постройте эффективную по Рао-Крамеру оценку параметра θ.
3. В избирательном округе А было опрошено 100 избирателей возраста 60+ и 200 избирателей возраста 30-40 лет. Среди опрошенных избирателей старшего возраста 62 человека объявили, что будут голосовать за кандидата NN; среди опрошенных избирателей среднего возраста таких оказалось 105 человек. Постройте асимптотический доверительный интервал уровня надёжности 0.95 для разности вероятностей поддержки кандидата NN среди избирателей среднего и старшего возраста.
4. Среднее время сборки изделия составляет 90 минут. Инженер предложил новый метод сборки изделий. В результате сборки 6 изделий новым способом время сборки составило 79, 74, 102, 95, 70, 90 минут. Можно ли считать, что время сборки в среднем сократилось? Предполагается, что время сборки имеет нормальное распределение. Уровень значимости считать равным 0.05. Опишите процедуру проверки соответствующей гипотезы.
5. На экзамене по английскому языку студент должен ответить на 100 вопросов. На каждый вопрос предлагается 4 варианта ответа, один из которых правильный. Студент NN указал верные ответы на 30 вопросов. Можно ли считать (на уровне значимости 0.05), что студент NN не знает английский язык (т.е., выбор ответа на вопросы является случайным)? Опишите процедуру проверки соответствующей гипотезы.

Комментарии.
В данном варианте не представлены, а в других вариантах могут быть задачи на следующие темы:

1) Доказательство эффективности по Рао-Крамеру предъявленной оценки параметра;
2) Метод моментов;
3) Построение доверительных интервалов для среднего и дисперсии в одновыборочных гауссовских моделях и доверительных интервалов для разности средних в двухвыборочных гауссовских моделях с известными дисперсиями;
4) Проверка гипотезы о значении дисперсии гауссовской СВ (в одновыборочной гауссовской модели) и о равенстве вероятностей «успеха» двух биномиальных СВ
5) Исследование свойств оценок и мощности критериев.

1. Выборка $X_1, \dots, X_n$ соответствует распределению Релея, плотность которого имеет вид $f(x) = \frac{2x}{\theta}\exp\left(\frac{-x^2}{\theta}\right)$ *при* $x > 0$. Найдите оценку максимального правдоподобия параметра θ. Докажите несмещённость и состоятельность этой оценки.

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

1) строим функцию правдоподобия

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

$$\ln(L(\theta)) = \sum_{i=1}^{n} \left(\ln 2 + \ln x_i - \ln\theta - \frac{x_i^2}{\theta}\right) = n\ln 2 + \sum_{i=1}^{n} \ln x_i - n\ln\theta - \frac{1}{\theta}\sum_{i=1}^{n} x_i^2 - n\ln\theta - \frac{1}{\theta}\sum_{i=1}^{n} x_i^2$$

$$\ln(L(\theta)) = n\ln 2 + \sum_{i=1}$$

2) производная по θ

$$\frac{\partial}{\partial\theta}\ln(L(\theta)) = -\frac{n}{\theta} + \frac{1}{\theta}\sum_{i=1}^{n} x_i^2$$

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

$$\frac{1}{\theta^2}\sum_{i=1}^{n} x_i^2 = \frac{n}{\theta} \quad \Big| \cdot \theta^2$$

$$\sum_{i=1}^{n} x_i^2 = n\theta$$

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

3) несмещенность оценки

$$E[\hat{\theta}] = E\left[\frac{1}{n}\sum_{i=1}^{n} x_i^2\right] = \frac{1}{n}\sum_{i=1}^{n} E[x_i^2] = E[x_i^2] = \theta \Rightarrow \text{несмещенная}$$

Тут нам нужно посчитать $E[X^2]$ Релея, хотя почему вообще Релея, он Рэлей

$$E[X^2] = \int_0^\infty x^2 \cdot \frac{2x}{\theta}\exp\left(-\frac{x^2}{\theta}\right)dx = \frac{2}{\theta}\int_0^\infty x^3 \exp\left(-\frac{x^2}{\theta}\right)dx = \left|\begin{array}{l}\text{много}\\ \text{вычислений}\end{array}\right| = \theta \cdot \frac{3}{4} \cdot \frac{2}{\sqrt{\pi}} \cdot \frac{\sqrt{\pi}}{2} = \theta$$

4) состоятельность оценки

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

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \theta\right| > \varepsilon\right) \xrightarrow{n\to\infty} 0, \forall \varepsilon > 0$$

$$P(|X - EX| < \varepsilon) \le \frac{\mathcal{D}X}{\varepsilon^2} \text{ — неравенство Чебышева}$$

$$\mathcal{D}X = \mathcal{D}(\hat{\theta}) = \frac{1}{n^2}\cdot n \cdot \mathcal{D}(x_i^2) = \frac{1}{n^2}\cdot n \cdot \underbrace{2\theta^2}_{\text{ну типа знаем}} = \frac{2\theta^2}{n}$$

$$\frac{2\theta^2}{\varepsilon^2 n} \xrightarrow[n\to\infty]{} 0 \Rightarrow \text{состоятельная}$$

Ответ: $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} x_i^2$, выборка несмещенная и состоятельная

2. Выборка $X_1, \dots, X_n$ соответствует распределению Пуассона $\Pi(\theta)$. Пользуясь критерием эффективности, постройте эффективную по Рао-Крамеру оценку параметра θ.

$$X_1, \dots, X_n \sim Pois(\theta) \qquad P(X=k) = \frac{\theta^k e^{-\theta}}{k!}, \; k = 0, 1, 2, \dots$$

$$L(\theta) = \prod_{i=1}^{n} \frac{\theta^{x_i} e^{-\theta}}{x_i!}$$

$$\ln(L(\theta)) = \sum_{i=1}^{n} (x_i \ln\theta - \theta - \ln x_i!)$$

$$\frac{\partial}{\partial\theta}\ln(L(\theta)) = \sum_{i=1}^{n}\left(\frac{x_i}{\theta} - 1\right) = \frac{1}{\theta}\sum_{i=1}^{n} x_i - n = 0$$

$$\frac{1}{\theta}\sum_{i=1}^{n} x_i = n \Rightarrow \hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$I(\theta) = -E\left[\frac{\partial^2}{\partial\theta^2}\ln(L(\theta))\right] = \frac{n}{\theta}$$

$$\frac{\partial^2}{\partial \theta^2} \ln(L(\theta)) = -\sum_{i=1}^{n} \frac{x_i}{\theta^2}$$

$$E\left[-\sum_{i=1}^{n} \frac{x_i}{\theta^2}\right] = -\sum_{i=1}^{n} \frac{E[x_i]}{\theta^2} = -\sum_{i=1}^{n} \frac{\theta}{\theta^2} = -\frac{n}{\theta}$$

$$D(\hat{\theta}) = D\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} \underbrace{D(x_i)}_{= \theta \text{ по распределению Пуассона}} = \frac{1}{n^2} \cdot n \cdot \theta = \frac{\theta}{n} = I(n) \Rightarrow \text{оценка эффективная}$$

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

3. В избирательном округе А было опрошено 100 избирателей возраста 60+ и 200 избирателей возраста 30-40 лет. Среди опрошенных избирателей старшего возраста 62 человека объявили, что будут голосовать за кандидата NN; среди опрошенных избирателей среднего возраста таких оказалось 105 человек. Постройте асимптотический доверительный интервал уровня надёжности 0.95 для разности вероятностей поддержки кандидата NN среди избирателей среднего и старшего возраста.

Группа 1 – опрошенные возраста 60+ лет

$n_1 = 100,\ x_1 = 62,\ \hat{p}_1 = \frac{x_1}{n_1} = \frac{62}{100} = 0{,}62$

Группа 2 – опрошенные возраста 30–40 лет

$n_2 = 200,\ x_2 = 105,\ \hat{p}_2 = \frac{x_2}{n_2} = \frac{105}{200} = 0{,}525$

$\hat{\theta} = \hat{p}_2 - \hat{p}_1 = 0{,}525 - 0{,}62 = -0{,}095$

$$\hat{\sigma}^2 = \frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2} = \frac{0{,}62(1-0{,}62)}{100} + \frac{0{,}525(1-0{,}525)}{200} \approx 0{,}0036$$

$\hat{\sigma} = \sqrt{0{,}0036} \approx 0{,}06$

$z_{1-\frac{\alpha}{2}} = z_{1-0{,}25} = z_{0{,}975} = 1{,}96$

$\hat{\theta} \pm z_{0{,}975} \cdot \hat{\sigma}$

$\hat{\theta} + z_{0{,}975} \cdot \hat{\sigma} = -0{,}095 + 1{,}96 \cdot 0{,}06 = 0{,}0226$

$\hat{\theta} - z_{0{,}975} \cdot \hat{\sigma} = -0{,}095 - 1{,}96 \cdot 0{,}06 = -0{,}2126$

$(-0{,}2126;\ 0{,}0226)$

4. Среднее время сборки изделия составляет 90 минут. Инженер предложил новый метод сборки изделий. В результате сборки 6 изделий новым способом время сборки составило 79, 74, 102, 95, 70, 90 минут. Можно ли считать, что время сборки в среднем сократилось? Предполагается, что время сборки имеет нормальное распределение. Уровень значимости считать равным 0.05. Опишите процедуру проверки соответствующей гипотезы.

H_0: Время сборки не изменилось или увеличилось $\mu \geq 90$

H_1: Время сборки снизилось $\mu < 90$.

$$\bar{x} = \frac{79+74+102+95+70+90}{6} = \frac{510}{6} = 85$$

$$s^2 = \frac{1}{n-1}\sum(x_i - \bar{x})^2 = \frac{1}{5}\left((79-85)^2 + (74-85)^2 + (102-85)^2 + (95-85)^2 + (70-85)^2 + (90-85)^2\right) =$$
$$= \frac{1}{5}(161 + 11^2 + 17^2 + 15^2) = 159{,}2$$

$s = \sqrt{159{,}2} \approx 12{,}62$

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{(85-90)\cdot\sqrt{6}}{12{,}62} \approx -0{,}97$$

$n-1 = 6-1 = 5$, $\alpha = 0{,}05$

$t_{критич} = 2{,}015$

$t = |-0{,}97| < t_{критич} = 2{,}015 \Rightarrow$ не отвергаем нулевую гипотезу

Ответ: нет, не можем

5. На экзамене по английскому языку студент должен ответить на 100 вопросов. На каждый вопрос предлагается 4 варианта ответа, один из которых правильный. Студент NN указал верные ответы на 30 вопросов. Можно ли считать (на уровне значимости 0.05), что студент NN не знает английский язык (т.е., выбор ответа на вопросы является случайным)? Опишите процедуру проверки соответствующей гипотезы.

H_0: Студент выбирает ответы случайно $p = \frac{1}{4}$

H_1: Студент знает английский $p \neq \frac{1}{4}$

$\hat{p} = \frac{30}{100} = 0{,}3$

$\sigma = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0{,}25 \cdot 0{,}75}{100}} \approx 0{,}0433$

$z = \frac{\hat{p} - p_0}{\sigma} = \frac{0{,}3 - 0{,}25}{0{,}0433} \approx 1{,}155$

$z_{критич} = z_{1-\frac{\alpha}{2}} = 1{,}96$

$|z| < 1{,}96 \Rightarrow$ НООН$_0$

Ответ: да.