**2018**

1. (a) Let $\hat{\theta}$ be an estimator of an unknown parameter $\theta$.

i. Define the bias of $\hat{\theta}$, $\mathrm{bias}(\hat{\theta})$; [1 mark]

ii. Define the mean squared error of $\hat{\theta}$, $\mathrm{MSE}(\hat{\theta})$; [1 mark]

iii. Show that
$$\mathrm{MSE}(\hat{\theta}) = \mathrm{var}(\hat{\theta}) + (\mathrm{bias}(\hat{\theta}))^2.$$

[5 marks]

See lecture notes

(b) A continuous random variable $X$ has density function

$$f_X(x) = \begin{cases} 2\lambda^2 x e^{-(\lambda x)^2}, & x \geq 0, \\ 0, & x < 0, \end{cases}$$

where $\lambda > 0$. Suppose $x_1, \ldots, x_n$ are observations of independent variables $X_1, \ldots, X_n$, respectively, all with the same distribution as $X$.

i. Find the log-likelihood function $l(\mu)$ for this sample.
[Note, $\mu$ is a variable from which the estimator $\hat{\lambda}$ of $\lambda$ is selected.] [5 marks]

ii. Hence, show that the maximum likelihood estimate $\hat{\lambda}$ of $\lambda$ is

$$\hat{\lambda} = \left( \frac{n}{\sum_{i=1}^{n} x_i^2} \right)^{1/2}.$$

[Hint: be careful to check that this is actually a maximum of $l(\mu)$.] [8 marks]

iii. Now consider the random variable

$$L = \frac{\sum_{i=1}^{n} X_i^2}{n}.$$

Assuming $\lambda^2 X^2$ has mean and variance $1$, show that $L$ is an unbiased estimator for $\lambda^{-2}$ and hence, find $\mathrm{MSE}(L)$.
Does it necessarily follow that $\hat{\lambda}$ is an unbiased estimator for $\lambda$? [5 marks]

Solution:

i) The likelihood function

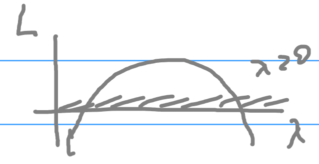
$$L(\lambda) = \prod_{i=1}^{n} 2\lambda^2 x_i e^{-\lambda^2 x_i^2} = 2^n \lambda^{2n} \prod_{i=1}^{n} x_i \, e^{-\lambda^2 \sum_{i=1}^{n} x_i^2}$$

The log likelihood function

$$\ell(\lambda) = \ln L(\lambda) = \ln 2^n + 2n \ln \lambda + \sum_{i=1}^{n} \ln x_i - \lambda^2 \sum_{i=1}^{n} x_i^2$$

$\ln(\lambda)$ is defined only for $\lambda > 0$

$\lambda \to 0 \Rightarrow \ell(\lambda) \to -\infty$ , $\lambda \to \infty \Rightarrow \ell(\lambda) \to -\infty$



$\Rightarrow$ at least 1 maximum

ii) $$\frac{d\ell(\lambda)}{d\lambda} = \frac{2n}{\lambda} - 2\lambda \cdot \sum_{i=1}^{n} x_i^2 = 0$$

$\lambda > 0 \Rightarrow 2n = 2\lambda^2 \sum_{i=1}^{n} x_i^2$ $\qquad \hat{\lambda} = +\left(\frac{n}{\sum_{i=1}^{n} x_i^2}\right)^{1/2}$

$$\frac{d^2\ell(\lambda)}{d\lambda^2} = -\frac{2n}{\lambda^2} - 2\sum_{i=1}^{n} x_i^2 < 0 \qquad \forall\, x_i \in [0,1],\ \lambda > 0,\ n > 0$$

$$\Rightarrow \hat{\lambda} = \left(\frac{n}{\sum_{i=1}^{n} x_i^2}\right)^{1/2} \text{ is the MLE}$$

iii) $$L = \frac{\sum_{i=1}^{n} x_i^2}{n}$$

$E(\lambda^2 x^2) = 1 \qquad Var(\lambda^2 x^2) = 1$

$$E(L) = E\left(\frac{\sum_{i=1}^{n} x_i^2}{n}\right) = E\left(\frac{\lambda^2 \sum_{i=1}^{n} x_i^2}{\lambda^2 n}\right) = \frac{E\left(\sum_{i=1}^{n} \lambda^2 x_i^2\right)}{\lambda^2 n} =$$

$$= \frac{\sum_{i=1}^{n} E(\lambda^2 x_i^2)}{\lambda^2 n} = \frac{n}{\lambda^2 n} = \frac{1}{\lambda^2}$$

$\Rightarrow$ unbiased

$$MSE(L) = Var(L) + Bias^2(L) = Var(L)$$

It can be shown that

$$Var(L) = \frac{1}{\lambda^4 n} \qquad Var\left(\frac{\sum x_i^2}{n}\right) = Var\left(\frac{\lambda^2 \sum x_i^2}{\lambda^2 n}\right) = \frac{\sum Var(\lambda^2 x_i^2)}{\lambda^4 n}$$

Does it necessarily follow that $\hat{\lambda}$ is an unbiased estimator for $\lambda$? ? No

$\hat{\lambda} = \left(\frac{n}{\sum_{i=1}^{n} x_i^2}\right)^{1/2}$ $\qquad E(\hat{\lambda}) = E\left(\frac{\sqrt{n}}{(\sum x_i^2)^{1/2}}\right) \neq \frac{\sqrt{n}}{(\sum E\, x^2)^{1/2}}$

2. Let $X$ be a continuous random variable with density function

$$f_X(x) = \begin{cases} \frac{2}{(x+1)^2}, & 0 \leq x \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find the cumulative distribution $F_X(x)$ and hence, show that, if $a = \frac{1}{39}$ and $b = \frac{19}{21}$,

$$\mathrm{P}(a < X < b) = 0.9.$$

[8 marks]

(b) Let $Y$ be a random variable such that, for unknown $\theta$, the pivot

$$\frac{\theta - Y^2}{\theta}$$

has the same distribution as random variable $X$. Using the result from (a), find a $90\%$ confidence interval for $\theta$ based on the single observation of $Y$, $y = -2$. [4 marks]

a) The cumulative function

$$F_X(x) = \int_{-\infty}^{x} \frac{2}{(x+1)^2}\, dx = \int_{0}^{x} \frac{2}{(x+1)^2}\, dx = -\frac{2}{(x+1)}\Big|_0^x =$$

$$= -\frac{2}{x+1} + 2 \qquad \text{for } x \in [0, 1]$$

$$\Rightarrow \quad F_X(x) = \begin{cases} 0 & , \ x < 0 \\ 2 - \frac{2}{x+1} & , \ 0 \leq x \leq 1 \\ 1 & , \ x > 1 \end{cases}$$

$$P(a<X<b) = P(X<b) - P(X<a)$$

$$P(X<a) = F_X(a)$$

$$F_X(a) = F_X\left(\frac{1}{39}\right) = 2 - \frac{2\cdot 39}{40} = \frac{2}{40} = \frac{1}{20}$$

$$P(X<b) = F(b)$$

$$F_X(b) = F_X\left(\frac{19}{21}\right) = 2 - \frac{2\cdot 21}{19+21} = \frac{2\cdot 19}{40} = \frac{19}{20}$$

$$P(a<X<b) = \frac{19}{20} - \frac{1}{20} = \frac{18}{20} = \frac{9}{10} = 0{,}9$$

b) $T = \dfrac{\theta - y^2}{\theta} \sim f_X(x)$

$$P(L < \theta < U) = 0{,}9 \quad - ?$$

$P(a < T < b) = 0{,}9$ (explain!)

$$P\left(\frac{1}{39} < T < \frac{19}{21}\right) = 0{,}9$$

$a < T$ : $\quad a < \dfrac{\theta - y^2}{\theta} \Rightarrow y^2 < \theta(1-a)$

$$\frac{y^2}{1-a} < \theta \Rightarrow \frac{2^2}{1-\frac{1}{39}} = \frac{78}{19} = L$$

$T < b$ : $\quad \dfrac{\theta - y^2}{\theta} < b \Rightarrow \theta(1-b) < y^2$

$$\theta < \frac{y^2}{1-b} \Rightarrow \frac{2^2}{1-\frac{19}{21}} = 42 = U$$

$$\Rightarrow P(4{,}11 < \theta < 42) = 0{,}9$$

c) For the random variable $X$, calculate the probability of $X$ lying in each of the following intervals.

i. $[0,0.2)$;
ii. $[0.2,0.4)$;
iii. $[0.4,0.6)$;
iv. $[0.6,0.8)$;
v. $[0.8,1]$.

[5 marks]

| interval | I | II | III | IV | V |
|---|---|---|---|---|---|
| Frequency, $E_i$ | $1/3$ | $5/21$ | $5/28$ | $5/36$ | $1/9$ |

$$E_I = F(0.2) - F(0) = 2 - \frac{2}{1.2} - 2 + \frac{2}{1} = \frac{2}{6} = \frac{1}{3}$$

$$E_{II} = F(0.4) - F(0.2) = 5/21$$

. . .

(d) A statistical experiment is performed in which an independent random sample of size 100 from a certain distribution is taken. The following table records how many observations lie in each of the same intervals from part (c):

| Interval | $[0,0.2)$ | $[0.2,0.4)$ | $[0.4,0.6)$ | $[0.6,0.8)$ | $[0.8,1]$ |
|---|---|---|---|---|---|
| Frequency | 40 | 31 | 12 | 10 | 7 |

Using the expected values from part (c), perform a $\chi^2$ goodness of fit test of the hypothesis "the data is drawn from the $X$ distribution", at the 0.1 significance level. What is the conclusion?

[8 marks]

| interval | I | II | III | IV | V |
|---|---|---|---|---|---|
| observed $O_i$ | 40 | 31 | 12 | 10 | 7 |
| estimated $n \cdot E_i$ | $100/3$ | $500/21$ | $500/28$ | $500/36$ | $100/9$ |

$\Rightarrow$ $\chi^2$ statistic $\approx 8.036$

goodness of fit

$$\chi^2_{stat} = \sum_{i=1}^{5} \frac{(O_i - nE_i)^2}{nE_i}$$

We have 5 intervals $\Rightarrow$ sampling distribution is appoximately $\chi^2_4$

$\Rightarrow$ critical value is $\chi^2_{0,1,4} \approx 7,779$

What is the conclusion?

$\chi^2_{obs} > \chi^2_{0,1,4} \Rightarrow$ reject $H_0$

3. (a) Let $X_1$ and $X_2$ be two random variables with distributions $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, respectively. Assuming that $X_1$ and $X_2$ are independent and $X_1 - X_2 \sim N(\mu, \sigma^2)$, show that $\mu = \mu_1 - \mu_2$ and $\sigma^2 = \sigma_1^2 + \sigma_2^2$. [5 marks]

The derivation is based on properties of expectation and the definition of variance

(b) Suppose we are carrying out a hypothesis test at the $\alpha$-significance level. Let $H_0$ be the null hypothesis and $H_1$ be the alternative hypothesis. Define

i. the type I error;
ii. the type II error;
iii. the power of the test.

From lecture notes

3a) $y = X_1 + X_2 \sim N(\mu, \sigma^2)$

$\mu = \mu_1 + \mu_2$ , $\sigma^2 = \sigma_1^2 + \sigma^2$

$E(X_1 + X_2) = E(X_1) + E(X_2)$

$E(X_1 - X_2) = E(X_1) + E(-X_2) = E(X_1) - E(X_2)$

$M = X_1 - X_2 \sim N(\mu, \sigma^2)$

$\mu = E(M) = \mu_1 - \mu_2$

$\sigma^2 = Var(M) = Var(X_1) + (-1)^2 Var(X_2)$

(c) A statistics module has been running for many years and, in the past, it has been found that each year the number of students passing the exam has distribution $\mathrm{Bi}(n, 0.75)$, where $n$ are the number of students taking the module that year.

A lecturer is teaching the module for the first time and 105 out of 150 students pass the exam. Perform a hypothesis test at the 0.05-significance level, where the null hypothesis is "The probability of a student passing the module is 0.75" and the alternative hypothesis is "The probability of a student passing the module is less than 0.75". What is the conclusion?

[Hint: Clearly state any assumptions made and recall the conditions under which a binomial distribution can be approximated by a normal distribution.] [10 marks]

1 $H_0: p = 0{,}75$

$H_A: p < 0{,}75$

2 $n > 30$, $n = 150$ $\quad \frac{X}{n} = \frac{105}{150} = 0{,}7$

$\frac{X}{n} \cdot n > 5$, $\left(1 - \frac{X}{n}\right) \cdot n > 5 \Rightarrow$

we can use the following test statistic

$$T = \frac{\frac{X}{n} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim N(0,1)$$

- Observed $T = \frac{0{,}7 - 0{,}75}{\sqrt{\frac{0{,}75 \cdot 0{,}25}{150}}} = -\frac{0{,}05}{0{,}0354} = -1{,}414$

- For one-sided test $T_{crit} = -1{,}645$, i.e.

$P(T < -1{,}645) = 0{,}05$

- $T_{observed} > T_{crit} \Rightarrow$ What is the conclusion?

(d) At another university, 300 students are taking a statistics module. Two lecturers $A$ and $B$ each teach $150$ students. After the exam has been taken, $98$ of lecturer $A$'s students have passed, while $92$ of lecturer $B$'s students have passed.

Assuming that, for each lecturer, the number of students passing the exam has a binomial distribution, perform a hypothesis test at the $0.05$-significance level to test the null hypothesis "Students taught by lecturer $A$ or lecturer $B$ have the same probability of passing" against the alternative hypothesis "Students taught by lecturer $A$ have a *different* probability of passing than those taught by lecturer $B$". What is the conclusion?

[Hint: Recall the result from part (a) and again state clearly state any assumptions made]. [7 marks]

for group A $\quad n = 150$

$X_A \sim Bin(n, p_A)$, $E(X_A) = np_A$
$Var(X_A) = np_A(1-p_A)$

for group B

$X_B \sim Bin(n, p_B)$, $E(X_B) = np_B$
$Var(X_B) = np_B(1-p_B)$

$H_0: p_A - p_B = 0$
$H_A: p_A - p_B \neq 0$

Under $H_0$ we assume that $p_A = p_B \Rightarrow$
$Var(X_A) = Var(X_B) \Rightarrow$ we can use the following pivot

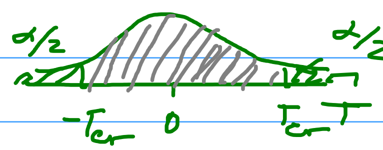
$$T = \frac{\frac{x_A - x_B}{n} - (p_A - p_B)}{\sqrt{\frac{\frac{x_A}{n}(1-\frac{x_A}{n})}{n} + \frac{\frac{x_B}{n}(1-\frac{x_B}{n})}{n}}} \sim N(0,1)$$

$$T = \frac{0,04 - 0}{\sqrt{0,00151 + 0,00158}} = \frac{0,04}{\sqrt{0,00309}} = \frac{0,04}{0,0556} =$$

$$= \underline{\underline{0,72}}$$

$T_{crit}$ : $P(|T| > T_{crit}) = 0,05$

$T_{crit} = 1,96$




$T < T_{crit} \Rightarrow$ What is the conclusion?

do not reject $H_0$