Final Probability and Statistics AESB1212
25 January 2019, 9.00 – 11.00

Last name, initials: | | Grade: | |

Student number: | | | | | | | |

Open questions

ALWAYS MOTIVATE YOUR ANSWERS AND WRITE NEATLY.

1. Scientists measure the temperature in a lava flow. The temperature measurements have a normal distribution with unknown expectation μ_A (in Celsius) and known variance $\sigma_A^2 = 4$. To improve the measurements, 6 repeated measurements are performed giving an average value of $\bar{x}_6 = 967.6$.

   (a) (3 points) Compute the two-sided 95% confidence interval for μ_A.

The 95% confidence interval is given by

$$\left(\bar{x}_n - z_{0.025}\frac{\sigma_A}{\sqrt{n}},\ \bar{x}_n + z_{0.025}\frac{\sigma_A}{\sqrt{n}}\right).$$

We have $n=6$, $\bar{x}_6 = 967.6$, $\sigma_A = 2$ and $z_{0.025} = 1.96$.

That yields $967.6 \pm 1.96 \cdot \frac{2}{\sqrt{6}} = 967.6 \pm 1.60$.

The two-sided 95% confidence interval for μ_A is given by

$(966.00, 969.20)$.

The scientists acquired a new thermometer. The measurements of the new thermometer follow a normal distribution with unknown expectation μ_B (in Celsius) and unknown variance σ_B. The temperature in a lava lake is measured 8 times, giving an average value $\bar{x}_8 = 1043.8$ and a sample standard deviation $s_8 = 1.52$.

   (b) (3 points) Compute the two-sided 99% confidence interval for μ_B.

The variance is unknown, so the confidence interval is of the form $\left(\bar{x}_n - t_{n-1,0.005}\frac{s_8}{\sqrt{n}},\ \bar{x}_n + t_{n-1,0.005}\frac{s_8}{\sqrt{n}}\right)$.

The value are $n=8$, $\bar{x}_8 = 1043.8$, $s_8 = 1.52$ and $t_{7,0.005} = 3.499$.

This gives $1043.8 \pm 3.499\,\frac{1.52}{\sqrt{8}} = 1043.8 \pm 1.88$.

The two-sided confidence interval for μ_B is given by

$(1041.92, 1045.68)$.

2. A t-test for the null hypothesis $H_0 : \mu = 15$ is performed by means of a dataset consisting of $n = 25$ measurements. The sample mean is 16.5 and the sample variance is 9. The significance level of the test is 0.01.

(a) (3 points) Decide whether the null hypothesis is rejected in favour of $H_1 : \mu \neq 15$.

$H_0 : \mu = 15$ $H_1 : \mu \neq 15$, so we perform a two-sided test.

$T = \frac{\bar{X}_n - \mu}{S_n / \sqrt{n}} = \frac{16.5 - 15}{3 / \sqrt{25}} = 2.5$ since $\bar{X}_n = 16.5$,

$\mu = 15$, $S_{25}^2 = 9$, $n = 25$.

The critical value of the two-sided test is

$t_{n-1, \alpha/2} = t_{24, 0.005} = 2.797$.

$T = 2.5 < 2.797 = t_{n-1, \alpha/2}$, so we do not reject.

There is not enough evidence to reject $H_0 : \mu = 15$ in favour of $H_1 : \mu \neq 15$.

(b) (3 points) The analysis is repeated with the alternative $H_1 : \mu > 15$. Decide whether the null hypothesis is rejected when using this alternative.

$H_0 : \mu = 15$ $H_1 : \mu > 15$, so we perform a one-sided test.

$T = \frac{\bar{X}_n - \mu}{S_n / \sqrt{n}} = 2.5$ with the same values as above.

The critical value of the one-sided test is

$t_{n-1, \alpha} = t_{24, 0.01} = 2.492$.

$T = 2.5 > 2.492 = t_{n-1, \alpha}$, so we reject.

There is enough evidence to reject $H_0 : \mu = 15$ in favour of $H_1 : \mu > 15$.

3. Assume two measurement devices can measure the speed of a tennis serve. Assume the measurements from the two devices are realizations of two random variables X_1 and X_2. Suppose X_1 and X_2 are independent with $E[X_1] = E[X_2] = \mu$, where μ is the true speed, and that the variances are given by $\text{Var}(X_1) = 9$ and $\text{Var}(X_2) = 16$.

(a) (2 points) Consider the weighted average $X_\alpha = \alpha X_1 + (1-\alpha)X_2$, for $\alpha \in [0,1]$. Show that X_α is an unbiased estimator for μ.

We have

$$\mathbb{E}[X_\alpha] = \mathbb{E}[\alpha X_1 + (1-\alpha) X_2] = \alpha \mathbb{E}[X_1] + (1-\alpha) \mathbb{E}[X_2]$$
$$= \alpha\mu + (1-\alpha)\mu = \mu.$$

It follows that X_α is an unbiased estimator for μ.

(b) (2 points) Compute the variance of X_α.

The variance of X_α is given by

$$\text{Var}(X_\alpha) = \text{Var}(\alpha X_1 + (1-\alpha) X_2) = \alpha^2 \text{Var}(X_1) + (1-\alpha) \text{Var}(X_2)$$
$$= \alpha^2 \cdot 9 + (1 - 2\alpha + \alpha^2) \cdot 16 = 25\alpha^2 - 32\alpha + 16.$$

(c) (2 points) Find the value α that minimizes the variance of X_α.
If you did not solve part (b) you may assume that $\text{Var}(X_\alpha) = 75\alpha^2 - 96\alpha + 75$.

Let $f(\alpha) = \text{Var}(X_\alpha) = 25\alpha^2 - 32\alpha + 16.$

$$f'(\alpha) = 50\alpha - 32$$

$$f'(\alpha) = 0 \iff 50\alpha = 32 \iff \alpha = 0.64$$

$$f''(\alpha) = 50 > 0$$

At $\alpha = 0.64$ the first derivative is zero and the second derivative is positive.

$\alpha = 0.64$ is the global minimum of $f(\alpha) = \text{Var}(X_\alpha)$.

Extra space

1./6. $\frac{\#\{x_i : 4 < x_i \leq 6\}}{n \cdot |(4,6]|} = \frac{8}{20 \cdot 2} = 0.2 \to C$

2./9. Large observations are in favour of $H_1 : \lambda < 3$. Let $X \sim Exp(3)$. The p-value is $P(X \geq 2) = 1 - P(X \leq 2) = 1 - F(2) = e^{-3 \cdot 2} = 0.0025 \to F$

3./5. On average one call arrives per 4 minutes. The times between calls are $Exp(\frac{1}{4})$.
$P(X \geq 10) = e^{-\frac{10}{4}} = e^{-2.5} \to A$

4./7. $\mathbb{E}\left[\sum_{i=1}^{200} X_i\right] = 200 \cdot \frac{1}{2} = 100$
$Var\left(\sum_{i=1}^{200} X_i\right) = 200\, Var(X_1) = \frac{200}{2^2} = 50$ By the CLT the sum is well approximated by $N(100, 50) \to D$

5./8. $\mathbb{E}[T] = \mathbb{E}[X_1] = \frac{12-0}{2} = 6$, $Var(T) = \frac{1}{10} Var(X_1) = \frac{1}{10} \frac{12^2}{12} = \frac{12}{10}$
$\to C$

6./1. $0.25 \cdot (8+1) = 2.25$
$q_8(0.25) = x_{(2)} + 0.25(x_{(3)} - x_{(2)}) = 4 + 0.25 \cdot 2 = 4.5$
median $= 7$ $\quad 0.75 \cdot (8+1) = 6.25$
$q_8(0.75) = x_{(6)} + 0.25(x_{(7)} - x_{(6)}) = 8 + 0.25 \cdot 0 = 8 \to A$

7./2. Type I error means H_0 is true but is rejected. (I) is correct. A larger sample entails a smaller confidence interval, so (II) is not correct. $\to B$

8./3. Let $X \sim U(0,3)$. $P(\text{Type II error}) = P(\text{not reject})$
$= P(X < 1.8) = \frac{1.8}{3} = 0.6 \to F$

9./4. Let $Y \sim N(4,4)$. Then $\mathbb{E}[Y] = 4$, $Var(Y) = 4$.
$P(|Y - 4| \geq 8) \leq \frac{1}{8^2} \cdot 4 = \frac{1}{16} \to C$