Final Probability and Statistics AESB1212
Practice Exam

Only the use of a **non-graphical** calculator and a clean copy of the formula sheet is allowed.
This exam consists of nine multiple choice questions and three open questions. You should answer the open questions on the exam sheet.

Grade: Every correct multiple choice question counts for 2 points; for the open questions, points are denoted per part. Then:

$$\text{Grade} = \frac{MC + OQ}{4} + 1,$$

where MC is the amount of points scored for the multiple choice part and OQ for the open questions part.

Explanation MC sheet: Colour the boxes with a pencil. Fill in the version, course code, your name and student number. The latter should be ticked as well. Finally, sign the sheet with your signature.

Version 1

1. $X_1, X_2, \ldots, X_{100}$ are i.i.d. random variables, having a $U(0,6)$ distribution.
The sum $\sum_{i=1}^{100} X_i$ has approximately the following distribution:

A. $N(300, 300)$ **B.** $N(200, 600)$ **C.** $U(300, 600)$
D. $N(300, 200)$ **E.** $N(200, 300)$ **F.** $U(0, 600)$

2. Let X have a $Exp(2)$ distribution. Using Chebyshev's inequality it holds that:

A. $\mathrm{P}(|X - \frac{1}{2}| < 2) \leq \frac{1}{4}$ **B.** $\mathrm{P}(|X - 2| < 2) \geq \frac{15}{16}$
C. $\mathrm{P}(|X - \frac{1}{2}| < 2) \leq \frac{1}{16}$ **D.** $\mathrm{P}(|X - 2| < 2) \geq \frac{3}{4}$
E. $\mathrm{P}(|X - \frac{1}{2}| < 2) \geq \frac{15}{16}$ **F.** $\mathrm{P}(|X - 2| < 2) \leq \frac{1}{16}$

3. Consider the following two statements:
(I) If a random sample becomes twice as large and all other factors stay equal, then a confidence interval becomes twice as small.
(II) A Type II error occurs if the null hypothesis is falsely not rejected.
Which of these statements is/are true?

A. Both **B.** None **C.** Only (I) **D.** Only (II)

4. The following values of the empirical distribution function of a certain dataset are given:

t	0	1	2	3	4	5	6
$F_n(t)$	0	0.20	0.30	0.55	0.70	0.90	1.0

If you would draw an histogram of this dataset, what would be the height of bin $(2, 4]$?

A. 0.20 **B.** 0.30 **C.** 0.50 **D.** 0.40 **E.** 0.35 **F.** 0.25

5. Let T_1 and T_2 be two independent unbiased estimators for θ. It is given that $\mathrm{Var}(T_1) = 1$ and $\mathrm{Var}(T_2) = 2$. Consider the following new estimator for θ: $T_3 = 2T_1 - T_2$.
What is the mean squared error of T_3?

A. 2 **B.** 0 **C.** $4 + \theta^2$ **D.** 6 **E.** 4 **F.** θ^2

6. Let $X_1, \ldots, X_n$ be a random sample from a $Pois(\mu)$ distribution. Consider the following two estimators for $e^{-\mu} = P(X_i = 0)$:

$$S = \frac{\#\{X_i = 0\}}{n} \quad \text{and} \quad T = e^{-\bar{X}_n},$$

where we use $\#A$ to denote the number of elements in A. Define the bias of an estimator T as $\text{Bias}[T] = \text{E}[T] - \theta$.

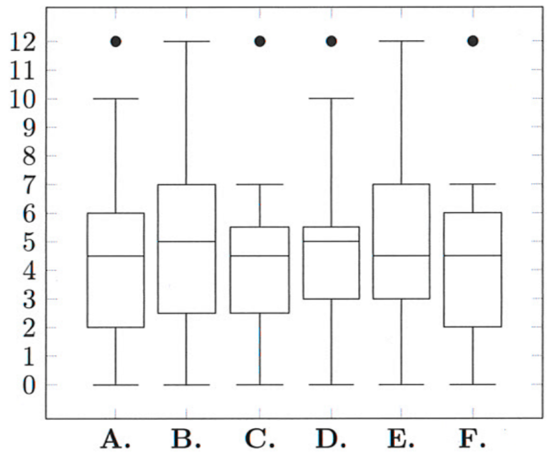
Then it is the case that

A. S and T are both unbiased estimators
B. S is unbiased and T has a negative bias
C. T is unbiased and S has a negative bias
D. S is unbiased and T has a positive bias
E. T is unbiased and S has a positive bias
F. S and T are both biased

7. Consider the following ordered dataset:

$$0 \quad 1 \quad 3 \quad 3 \quad 4 \quad 5 \quad 5 \quad 5 \quad 7 \quad 12$$

Which boxplot below corresponds to the given dataset?



8. We model the measurement X of the speed of a car at the highway as a normally distributed random variable with parameters $\mu = v$ (km/h) and $\sigma^2 = 25$, where v is the exact speed of the car. To check whether the driver is driving faster than 130 km/h, we test $H_0 : v = 130$ against $H_1 : v > 130$. The test statistic is the measurement X.

Furthermore, the null hypothesis is rejected (and the driver receives a speeding ticket) if the measured speed is 136 km/h or higher.

What is then the value of P_{II}, the probability of an error of type II, for a driver whose exact speed is 132 km/h?

A. $P_{II} = 0.8888$	**B.** $P_{II} = 0.7881$	**C.** $P_{II} = 0.4364$
D. $P_{II} = 0.1112$	**E.** $P_{II} = 0.5636$	**F.** $P_{II} = 0.2119$

9. Consider 100 batteries. The lifetimes of the batteries are i.i.d. random variables with expectation $\mu = 100$ hours and $\sigma^2 = 225$.

   Approximate the probability that the total lifetime of the 100 batteries is less than 10150 hours.

   A. 0.8413 **B.** 0.1587 **C.** 0.6103 **D.** 0.3897 **E.** 0.5398 **F.** 0.4602

Final Probability and Statistics AESB1212
Practice Exam

Last name, initials: | | Grade: | |

Student number: | | | | | | | |

Open questions

ALWAYS MOTIVATE YOUR ANSWERS AND WRITE NEATLY.

1. The time instants of incoming request at a data server can be modelled with a Poisson process. Let S_n be the number of requests in n minutes and let λ be the intensity (requests per minute) of the Poisson process.

   (a) (3 points) Use the Central Limit Theorem to deduce that if n is large, then S_n approximately has a normal distribution. Also specify its parameters.

Let X_i be the number of requests in the i-th minute. $S_n = X_1 + X_2 + \dots + X_n$

$X_1, X_2, \dots, X_n$ are i.i.d. $Pois(\lambda)$ random variables.

By the CLT, S_n is approximately normal for n large.

$\mathbb{E}[S_n] = \mathbb{E}[X_1 + X_2 + \dots + X_n] = n\,\mathbb{E}[X_1] = n\lambda$

$Var(S_n) = n\, Var(X_1) = n\lambda$

For large n, S_n is well approximated by $N(n\lambda, n\lambda)$.

Recall that λ is the intensity (requests per minute) of the Poisson process. You want to test $H_0 : \lambda = 1$ against $H_1 : \lambda > 1$.

   (b) (3 points) Suppose that $S_{60} = 72$ requests were counted in 60 minutes. Use part (a) to compute the corresponding p-value.

   If you did not solve part (a) you may assume that $S_{60} \overset{d}{\approx} N(50, 50)$.

Large values are in favour of $H_1: \lambda > 1$.

The p-value is given by $P(S_{60} > 72)$.

Under H_0 we have that S_{60} is close to $Y \sim N(60, 60)$

Let $Z \sim N(0,1)$.

$P(Y > 72) = P\left(\frac{Y-60}{\sqrt{60}} > \frac{72-60}{\sqrt{60}}\right) = P\left(Z > \frac{12}{\sqrt{60}}\right)$

$\approx P(Z > 1.55) = 0.0606$. The p-value is close to 0.0606.

2. In a factory bags of crisps are produced. The amount of crisps machine Apollo puts in a bag has a normal distribution with unknown expectation μ_A (gram) and known variance $\sigma_A^2 = 16$. To ensure the quality, a random sample of 25 bags filled by Apollo is taken weekly.

   The sample mean of the data set of last week is $\bar{x}_{25} = 148.22$ gram and the sample standard deviation is $s_{25} = 4.36$ gram.

   (a) (3 points) Compute the two-sided 90% confidence interval for μ_A.

σ is known, so $100(1-\alpha)\%$ confidence interval for μ_A is given by $\bar{X}_n \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

Here $n = 25$, $\alpha = 0.10$ and $\sigma_A = 4$

So we get $\bar{x}_{25} \pm z_{0.05} \frac{4}{\sqrt{25}} = 148.22 \pm 1.645 \frac{4}{5}$

$= 148.22 \pm 1.316$

This gives $(146.90, 149.54)$ as 90% confidence interval for μ_A.

The factory has recently acquired a new machine, called Beethoven. The amount of crisps filled by Beethoven has a normal distribution with expectation μ_B gram and variance $\sigma_B^2 = 9$. The producer claims to fill 150 gram of crisps in the bags.

To check Beethoven a random sample of size 25 was taken last week. The sample mean of this dataset is $\bar{x}_{25} = 148.47$ gram and the sample standard deviation is $s_{25} = 3.25$ gram.

   (b) (4 points) Formulate the relevant hypotheses and test statistic (with distribution!) and use a suitable statistical test to investigate whether Beethoven is putting too few crisps in the bags at a significance level of $\alpha = 0.05$.

$H_0: \mu_B = 150$, $H_1: \mu_B < 150$. The variance is known, so the test statistic is $T = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$ under H_0.

We have a left-sided test and $\alpha = 0.05$, so the critical value is $z_{0.95} = -z_{0.05} = -1.645$

Observed value of test statistic

$$t = \frac{148.47 - 150}{3/5} = -2.55$$

Since $t = -2.55 < -1.645 = z_{0.05}$ we reject H_0.

3. Let the random variable X be the waiting time until the next student has to go to the toilet. Assume that X has an $Exp(\lambda)$ distribution with unknown λ. We test $H_0 : \lambda = 0.2$ against $H_1 : \lambda < 0.2$, where we use X as test statistic.

(a) (3 points) Determine the critical region for this test at significance level $\alpha = 0.05$.

Large values of X are in favour of $H_1: \lambda < 0.2$

So the critical region is of the form $[c_r, \infty)$.

Since $\alpha = 0.05$, we have to solve

$P(X \geq c_r) = 0.05$. Under $H_0: \lambda = 0.2$ we have

$X \sim Exp(0.2)$, so

$$P(X \geq c_r) = e^{-0.2 \cdot c_r} = 0.05$$

if $\quad -0.2 \cdot c_r = \ln 0.05$

which yields $\quad c_r \approx 14.98$

So the critical region is $[14.98, \infty)$

(b) (2 points) Suppose that the observed value for X is $x = 10$. Compute the p-value for this realisation.

Large values of X are again in favour of

$H_1: \lambda < 0.2$

So the p-value is $P(X \geq 10)$

Again $X \sim Exp(0.2)$ under H_0, so we get

$$\text{p-value} = P(X \geq 10) = e^{-0.2 \cdot 10} = e^{-2} = 0.135$$

Extra space Short solutions of MC questions

1. According to CLT: $\sum_{i=1}^{100} X_i \approx N(n\mu, n\sigma^2)$, where $\mu = \mathbb{E}[X_i]$ and $\sigma^2 = Var(X_i)$. Here $n=100$, $\mu = 3$ and $\sigma^2 = \frac{1}{12}(b-a)^2 = \frac{6^2}{12} = 3$. So $\sum_i X_i \sim N(300, 300)$

2. According to Chebyshev's inequality
$P(|X-\frac{1}{2}| \geq 2) \leq \frac{Var\, X}{2^2} = \frac{1/4}{4} = \frac{1}{16}$
So $P(|X - \frac{1}{2}| < 2) \geq 1 - \frac{1}{16} = \frac{15}{16}$.

3. When the random sample becomes twice as large, the length of a confidence interval is divided by $\sqrt{2}$. A Type II error means that we stay with H_0 although H_1 is true, i.e. we do not reject and this is false. Thus only (II).

4. $\text{height} = \frac{\#\{X_i : X_i \in (2,4]\}}{n \cdot |(2,4]|} = \frac{\#\{X_i : X_i \leq 4\} - \#\{X_i : X_i \leq 2\}}{n \cdot 2}$
$= \frac{F_n(4) - F_n(2)}{2} = \frac{0.7 - 0.3}{2} = 0.2$

5. $Bias(T_3) = \mathbb{E}[T_3] - \theta = 2\mathbb{E}[T_1] - \mathbb{E}[T_2] - \theta = 0$
$Var(T_3) = Var(2T_1 - T_2) \underset{\substack{\uparrow \\ T_1 \text{ and } T_2 \text{ independent}}}{=} 4\, Var(T_1) + Var(T_2) = 6$
$MSE(T_3) = Var(T_3) + (Bias(T_3))^2 = 6$

6. $\#\{X_i = 0\} \sim Bin(n, e^{-\mu})$ so $\mathbb{E}[S] = \frac{n \cdot e^{-\mu}}{n} = e^{-\mu}$
By Jensen's inequality $\mathbb{E}[T] \geq e^{-\mathbb{E}[\bar{X}_n]} = e^{-\mu}$

7. $n = 10$, so median is $\frac{x_{(5)} + x_{(6)}}{2} = \frac{4+5}{2} = 4.5$
First quartile is $q_{10}(0.25)$: $0.25 \cdot (10+1) = 2.75$
So $q_{10}(0.25) = x_{(2)} + 0.75\,(x_{(3)} - x_{(2)}) = 1 + 0.75 \cdot (3-1) = 2.5$

8. $P_{II} = P(X < 136 \mid V = 132) = P\left(\frac{X - 132}{5} < \frac{136 - 132}{5}\right)$
$= P(Z < 0.8) = 1 - P(Z > 0.8) = 0.7881$ with $Z \sim N(0,1)$

9. $\mathbb{E}[\sum_i X_i] = 10\,000$, $Var(\sum_i X_i) = 100 \cdot 225 = 22\,500$
$\sum_i X_i$ approximately $N(10\,000, 22\,500)$ by CLT
$P(\sum_i X_i < 10\,150) = P\left(\frac{\sum_i X_i - 10\,000}{\sqrt{22\,500}} < \frac{10\,150 - 10\,000}{\sqrt{22\,500}}\right)$
$\approx P(Z < \frac{150}{150}) = 1 - P(Z \geq 1) = 1 - 0.1587 = 0.8413$

Answers multiple choice:

1 A.

2 E.

3 D.

4 A.

5 D.

6 D.

7 C.

8 B.

9 A.