Introduction to Probability & Statistics

Assignment 5, 2024/25

Instructions

Submit your answers to all of the **Assignment Questions (AQ)**. You should upload your solutions to the VLE as a *pdf file*.
Marks will be awarded for clear, logical explanations, as well as for correctness of solutions.
You should also look at the other questions in preparation for your Week 11 seminar.

Practice Questions

PQ1. You perform 28 independent experiments measuring a random variable X which you know has mean 457 and variance 676. Use Chebychev's inequality to give a lower bound on the probability that the mean of your measurements is between 433 and 481.

Answer

First we note that

$$\mathbb{E}\left[\bar{X}_{28}\right] = \mathbb{E}\left[X\right] = 457$$

and

$$\mathrm{Var}\left(\bar{X}_{28}\right) = \frac{\mathrm{Var}\left(X\right)}{28} = \frac{676}{28}.$$

Also we rewrite

$$\mathbb{P}\left(433 < \bar{X}_{28} < 481\right) = \mathbb{P}\left(|\bar{X}_{28} - 457| < 24\right) = 1 - \mathbb{P}\left(|\bar{X}_{28} - \mathbb{E}\left[\bar{X}_{28}\right]| \geq 24\right).$$

Chebychev's inequality tells us that

$$\mathbb{P}\left(|\bar{X}_{28} - \mathbb{E}\left[\bar{X}_{28}\right]| \geq 24\right) \leq \frac{\mathrm{Var}\left(\bar{X}_{28}\right)}{(24)^2} = \frac{169}{4032}.$$

Combining these two equations gives

$$\mathbb{P}\left(433 < \bar{X}_{28} < 481\right) \geq 1 - \frac{169}{4032} = \frac{3863}{4032} \approx 0.96.$$

PQ2. Let $X_1, \dots, X_n$ be an i.i.d. sample from the $\mathrm{N}(\mu, \sigma^2)$ distribution.

a) Write down a $100(1-\alpha)\%$ confidence interval estimator for μ in the case where σ^2 is known. Hence or otherwise, find a 95% confidence interval for μ in the case where $\sigma^2 = 9$ and a random sample of size 16 has been taken with values $x_1, \dots, x_{16}$ and it has been found that

$$\sum_{i=1}^{16} x_i = 50, \quad \sum_{i=1}^{16} (x_i - \bar{x}_{16})^2 = 115.$$

b) It is proposed that, from a second independent random sample of size 16 a 99% confidence interval for μ be constructed and that, from a third independent random sample of size 32, a 98% confidence interval for μ be constructed. State the probability that *neither* of these two confidence intervals will contain μ.

Answer

a) The confidence interval estimator is

$$\left[\bar{X}_n - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X}_n + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right].$$

With $\sigma = 3$, $n = 16$, $\bar{x}_n = 50/16$ and $z_{0.025} = 1.96$ substituted into the above expression we get the confidence interval $[1.655, 4.595]$.

b) The event that the second sample leads to a confidence interval that does not contain μ and the event that the third sample leads to a confidence interval that does not contain μ are independent because the samples are independent. Thus the probability that both samples lead to a confidence interval that does not contain μ is equal to the product of the individual probabilities, $0.01 \cdot 0.02 = 0.0002$.

PQ3. Lengths of baguettes are assumed to follow a $\mathrm{N}(\mu, \sigma^2)$ distribution. Six baguettes were measured, giving the following lengths in cms: 66, 69, 62, 62, 64, 67.

a) Calculate unbiased estimates for μ and σ^2.

b) Calculate a 90% confidence interval for μ.

Answer

a) An unbiased estimate for μ is $\hat{\mu} = \bar{x} = 65cm$. An unbiased estimate for σ^2 is $\hat{\sigma^2} = s_n^2 = 8$.

b) A 90% confidence interval is given by

$$\begin{aligned}\left[\bar{x}_n - t_{n-1,\alpha/2}\frac{s_n}{\sqrt{n}}, \bar{x}_n + t_{n-1,\alpha/2}\frac{s_n}{\sqrt{n}}\right] &= \left[65 - t_{5,0.05}\sqrt{8/6}, 65 + t_{5,0.05}\sqrt{8/6}\right] \\ &= \left[65 - 2.02\sqrt{8/6}, 65 + 2.02\sqrt{8/6}\right] \\ &= [62.67, 67.33].\end{aligned}$$

PQ4. Let $X_1, \ldots, X_n$ be an i.i.d. sample from a distribution with mean μ and variance σ^2. Let $\bar{X}_n$ denote the sample mean, and define

$$Z_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}.$$

Check that $\mathbb{E}[Z_n] = 0$ and $\mathrm{Var}(Z_n) = 1$.

Answer

We know that $\mathbb{E}\left[\bar{X}_n\right] = \mu$ and $\mathrm{Var}\left(\bar{X}_n\right) = \sigma^2/n$. By linearity of expectation we have

$$\mathbb{E}[Z_n] = \frac{\sqrt{n}\mathbb{E}\left[\bar{X}_n - \mu\right]}{\sigma} = 0.$$

For the variance, we know that for any random variable Y, and constants $a, b \in \mathbb{R}$,

$\text{Var}(aY+b) = a^2\text{Var}(Y)$. Thus

$$\text{Var}(Z_n) = \text{Var}\left(\sqrt{n}\bar{X}_n/\sigma - \sqrt{n}\mu/\sigma\right) = (\sqrt{n}/\sigma)^2\text{Var}\left(\bar{X}_n\right) = 1\,.$$

PQ5. Recall that we say that T_m has a t-distribution with $m > 1$ degrees of freedom, and write $T_m \sim t(m)$, if it has density function given by

$$f(x) = k_m\left(1+\frac{x^2}{m}\right)^{-\frac{m+1}{2}}, \qquad x \in \mathbb{R}\,,$$

where k_m is a constant that ensures that the density integrates to 1. If $T_m \sim t(m)$, show that $\mathbb{E}[T_m] = 0$.
Hint: you shouldn't need to explicitly calculate any integrals here!

Answer

We note that $f(x)$ is an even function: $f(-x) = f(x)$. Thus

$$\begin{aligned}
\mathbb{E}[T_m] &= \int_{-\infty}^{\infty} xf(x)dx \\
&= \int_{-\infty}^{0} xf(x)dx + \int_{0}^{\infty} xf(x)dx \\
&= \int_{0}^{\infty} (-x)f(-x)dx + \int_{0}^{\infty} xf(x)dx \\
&= \int_{0}^{\infty} (-x)f(x)dx + \int_{0}^{\infty} xf(x)dx \\
&= \int_{0}^{\infty} ((-x)+x)f(x)dx \\
&= 0\,.
\end{aligned}$$

Assignment Questions – answers to be uploaded

AQ1. Let $X_1, \ldots, X_{16}$ be an i.i.d. sample from a $\text{N}(3,1)$ distribution, and let $S = X_1 + X_2 + \cdots + X_{16}$. Express $\mathbb{P}(S < 52)$ in terms of the distribution function Φ of the standard normal distribution.

AQ2. Let $Y_1, Y_2, \ldots$ be an i.i.d. sequence of random variables, each with a Uniform$(0,3)$ distribution. Define a new sequence of random variables $X_1, X_2, \ldots$ by

$$X_n = \frac{1}{n}\sum_{i=1}^{n} Y_i^2\,.$$

Using the Law of Large Numbers, determine the value of $a \in \mathbb{R}$ for which $\mathbb{P}(\lim_{n\to\infty} X_n = a) = 1$.

AQ3. Suppose the random variables X_1, X_2 and X_3 all have the same expectation μ. For what values of a and b is

$$M = -4(X_1 - 2) + 9(X_2 - 1) + aX_3 + b$$

an unbiased estimator for μ?

AQ4. From a dataset $x_1, \ldots, x_{10}$ it has been calculated that

$$\sum_{i=1}^{10} x_i = 491, \quad \sum_{i=1}^{10}(x_i - \bar{x}_{10})^2 = 41.$$

You model the dataset as a random sample from the normal distribution with mean μ and variance σ^2.

a) Assume that both μ and σ^2 are unknown. Determine a 95% confidence interval for the mean μ. You can use that $t_{9,0.025} \approx 2.26$.

b) Now assume that it is known that the variance is $\sigma^2 = 5$. Give a 95% confidence interval for the mean μ in this case. You can use that $z_{0.025} \approx 1.96$.

Other Questions (for seminars / extra practice)

OQ1. If X has expectation μ and standard deviation σ, the ratio $r = |\mu|/\sigma$ is called the *measurement signal-to-noise-ratio* of X. If we define $D = |(X-\mu)/\mu|$ as the *relative deviation* of X from its mean μ, show that, for $\alpha > 0$,

$$\mathbb{P}(D < \alpha) \geq 1 - \frac{1}{r^2\alpha^2}.$$

OQ2. Let X be the number of 1s and Y be the number of 2s that occur in n rolls of a fair die. Use indicator random variables to compute $\mathrm{Cov}(X, Y)$ and $\rho(X, Y)$.
Hint: this is just like the smarties example covered in lectures.

OQ3. In the lecture I proved Chebychev's inequality for the case of a continuous random variable. Provide a similar proof for the case of a discrete random variable.

OQ4. Assume that in Example 17.1 from the lectures (measuring a ball rolling down an inclined plane), we choose to stop the ball always after one time unit, so that

$$X_i = \frac{1}{2}a(1+U_i)^2 + V_i,$$

where the independent errors are normally distributed with $U_i \sim \mathrm{N}(0, \sigma_U^2)$, $V_i \sim \mathrm{N}(0, \sigma_V^2)$. Assume the variances of the errors are known. Calculate the bias of the estimator $A = 2\bar{X}_n$ for the acceleration parameter a. Propose an unbiased estimator for a.

OQ5. Consider the following dataset of lifetimes of ball bearings in hours:

6278	3113	5236	11584	12628	7725	8604	14266	6125	9350
3212	9003	3523	12888	9460	13431	17809	2812	11825	2398.

Suppose that we are interested in estimating the minimum lifetime of this type of ball bearing. The dataset is modelled as a realization of a random sample $X_1, \ldots, X_n$. Each random variable X_i is represented as $X_i = \delta + Y_i$, where Y_i has an $\mathrm{Exp}(\lambda)$ distribution and $\delta > 0$ is an unknown parameter that is supposed to model the minimum lifetime. The objective is to construct an unbiased estimator for δ. It is known that

$$\mathbb{E}[M_n] = \delta + \frac{1}{n\lambda} \quad \text{and} \quad \mathbb{E}\left[\bar{X}_n\right] = \delta + \frac{1}{\lambda},$$

where $M_n = \min(X_1, \ldots, X_n)$ and $\bar{X}_n = (X_1 + \cdots + X_n)/n$.

a) Check whether

$$T = \frac{n}{n-1}\left(\bar{X}_n - M_n\right)$$

is an unbiased estimator for $1/\lambda$.

b) Construct an unbiased estimator D for δ.

c) Use the dataset to compute an estimate for the minimum lifetime δ.

OQ6. Let M_n be the maximum of n independent Uniform$(0, 1)$ random variables. Show that for any fixed $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(|M_n - 1| > \varepsilon\right) = 0\,.$$

OQ7. **[Harder]** (A more general law of large numbers, see Exercise 13.12 in the textbook). Let $X_1, X_2, \ldots$ be a sequence of independent random variables with $\mathbb{E}\left[X_i\right] = \mu_i$ and $\mathrm{Var}\left(X_i\right) = \sigma_i^2$ for $i = 1, 2, \ldots$. Let $\bar{X}_n = (X_1 + \cdots + X_n)/n$. Suppose that there exists an $M \in \mathbb{R}$ such that $0 < \sigma_i^2 \le M$ for all i, and let a be an arbitrary positive number.

a) Apply Chebychev's inequality to show that

$$\mathbb{P}\left(\left|\bar{X}_n - \frac{1}{n}\sum_{i=1}^{n}\mu_i\right| > a\right) \le \frac{\mathrm{Var}\left(X_1\right) + \cdots + \mathrm{Var}\left(X_n\right)}{n^2a^2}.$$

b) Conclude from a) that

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\bar{X}_n - \frac{1}{n}\sum_{i=1}^{n}\mu_i\right| > a\right) = 0.$$

c) Check that the weak law of large numbers is a special case of this result.

! Challenge question

Suppose that $X_1, \ldots, X_n$ are mutually independent random variables, each distributed as Exp(λ). (That is, all events of the kind $\{X_1 \le x_1\}, \ldots, \{X_n \le x_n\}$ are mutually independent.) Let $Y_n = \min\{X_1, \ldots, X_n\}$, and $V_n = \max\{X_1, \ldots, X_n\}$.

a) Show that $Y_n \sim \mathrm{Exp}(\lambda n)$.
b) What is the distribution function of V_n?
c) Show that, for all $s > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(V_n - (\log n)/\lambda \le s\right) = \exp(-e^{-\lambda s})\,.$$