Probability and Statistics
Autumn 2024

MATH-235
Mats J. Stensrud

---

**Problem Sheet 12** [1]

---

Based on Chapters 10.2 and 10.3 in the course book. Introduction to Statistics.

# Optional Revision Problems

**Exercise 1.** Let $X$ and $Y$ be positive random variables, not necessarily independent. Assume that the various expressions below exist. Write the most appropriate of $\leq$, $\geq$, $=$, or ? in the blank for each part (where "?" means that no relation holds in general).

1. $P(X+Y>2)$ _____ $\frac{E(X)+E(Y)}{2}$
2. $P(X+Y>3)$ _____ $P(X>3)$
3. $E(\cos(X))$ _____ $\cos(E(X))$
4. $E(X^{1/3})$ _____ $(E(X))^{1/3}$
5. $E(X^c)$ _____ $(E(X))^c$ for some constant $c \in \mathbb{R}$
6. $E(E(X|Y)+E(Y|X))$ _____ $E(X)+E(Y)$

**Solution 1.**

1. $P(X+Y>2) \leq \frac{E(X)+E(Y)}{2}$
   by Markov inequality and linearity
2. $P(X+Y>3) \geq P(X>3)$
   since $X>3$ implies $X+Y>3$ since $Y>0$
3. $E(\cos(X)) ? \cos(E(X))$
   e.g., let $W \sim \text{Bern}(1/2)$ and try for $X=(\pi/2)W$ and $X=\pi/2+(\pi/2)W$. Alternatively, argue that the convexity of $cos(x)$ depends on the value of $x$, so Jensen inequality can hold both ways ($\leq$ or $\geq$), depending on the support of $X$.
4. $E(X^{1/3}) \leq (E(X))^{1/3}$
   by Jensen inequality
5. $E(X^c) ? (E(X))^c$
   by Jensen inequality, convexity depends on the value of $c$, e.g. try $c=2$ and $c=1/2$.
6. $E(E(X|Y)+E(Y|X)) = E(X)+E(Y)$
   by linearity and Adam's law

[1] Exercises are based on the coursebook Statistics 110: Probability by Joe Blitzstein

**Exercise 2.** Let $X$ and $Y$ be i.i.d. positive random variables. Assume that the various expressions below exist. Write the most appropriate of $\leq$, $\geq$, $=$, or ? in the blank for each part (where "?" means that no relation holds in general).

1. $E(e^{X+Y})$ _____ $e^{2E(X)}$
2. $E(X^2 e^X)$ _____ $\sqrt{E(X^4)E(e^{2X})}$
3. $E(X|3X)$ _____ $E(X|2X)$
4. $E(X^7 Y)$ _____ $E(X^7)E(Y|X)$
5. $E\left(\frac{X}{Y}+\frac{Y}{X}\right)$ _____ $2$
6. $P(|X-Y|>2)$ _____ $\frac{\text{Var}(X)}{2}$

**Solution 2.**

1. $E(e^{X+Y}) \geq e^{2E(X)}$
   write $E(e^{X+Y}) = E(e^X e^Y) = E(e^X)E(e^Y) = E(e^X)E(e^X)$ using the fact that $X, Y$ are i.i.d., and then apply Jensen inequality for $e^X$.
2. $E(X^2 e^X) \leq \sqrt{E(X^4)E(e^{2X})}$
   by Cauchy-Schwarz inequality
3. $E(X|3X) = E(X|2X)$
   knowing $2X$ is equivalent to knowing $3X$
4. $E(X^7 Y) = E(X^7)E(Y|X)$ by Adam's law and taking out what's known, or using independence directly.
5. $E\left(\frac{X}{Y}+\frac{Y}{X}\right) \geq 2$
   since $E\left(\frac{X}{Y}\right) = E(X)E\left(\frac{1}{Y}\right) \geq \frac{E(X)}{E(Y)} = 1$, and similarly $E\left(\frac{Y}{X}\right) \geq 1$
6. $P(|X-Y|>2) \leq \frac{\text{Var}(X)}{2}$
   by Chebyshev, applied to the random variable $W = X - Y$, which has variance $2\text{Var}(X)$: $P(|W - E(W)| > 2) \leq \text{Var}(W)/4 = \text{Var}(X)/2$

# Week 12 Exercises

**Exercise 3.** Let $U_1, U_2, \ldots, U_{60}$ be i.i.d. Unif$(0,1)$ and $X = U_1 + U_2 + \cdots + U_{60}$.

1. Which **important distribution** is the distribution of $X$ very close to? Specify what the parameters are, and state which theorem justifies your choice.
2. Give a simple but accurate approximation for $P(X > 17)$. Justify briefly.

**Solution 3.**

1. By the **Central Limit Theorem**, the distribution of $X = U_1 + U_2 + \cdots + U_{60}$ is approximately normal ($\mathcal{N}(\mu, \sigma^2)$) when the number of terms is large. The parameters are:

$$\mathbb{E}[X] = 60 \cdot \mathbb{E}[U] = 60 \cdot \frac{1}{2} = 30,$$

$$\text{Var}(X) = 60 \cdot \text{Var}(U) = 60 \cdot \frac{1}{12} = 5.$$

Thus, $X \sim \mathcal{N}(30, 5)$.

2. To approximate $P(X > 17)$, we standardize $X$ to the standard normal distribution $Z = \frac{X-30}{\sqrt{5}}$. Then:

$$P(X > 17) = 1-P(X \leq 17) = 1-P\left(\frac{X-30}{\sqrt{5}} \leq \frac{17-30}{\sqrt{5}}\right) = 1-P\left(Z \leq \frac{-13}{\sqrt{5}}\right) = 1-\Phi\left(\frac{-13}{\sqrt{5}}\right).$$

Using symmetry of the standard normal distribution:

$$P(X > 17) = \Phi\left(\frac{13}{\sqrt{5}}\right).$$

Since $\frac{13}{\sqrt{5}} > 5$, and we already have $\Phi(3) \approx 0.9985$ by the 68-95-99.7% rule, the value is extremely close to 1.

**Exercise 4.** 1. Let $Y = e^X$, with $X \sim \text{Expo}(3)$. Find the mean and variance of $Y$.

2. For $Y_1, \ldots, Y_n$ i.i.d. with the same distribution as $Y$ from part 1., what is the approximate distribution of the sample mean $\bar{Y}_n = \frac{1}{n}\sum_{j=1}^{n} Y_j$ when $n$ is large?

**Solution 4.** 1. By LOTUS,

$$\mathbb{E}(Y) = \int_0^\infty e^x(3e^{-3x})dx = \left[-\frac{3}{2}e^{-2x}\right]_{x=0}^{\infty} = \frac{3}{2},$$

$$\mathbb{E}(Y^2) = \int_0^\infty e^{2x}(3e^{-3x})dx = \left[-3e^{-x}\right]_{x=0}^{\infty} = 3.$$

So $\mathbb{E}(Y) = 3/2, \text{Var}(Y) = 3 - 9/4 = 3/4$.

2. By the CLT, $\bar{Y}_n$ is approximately $\mathcal{N}\left(\frac{3}{2}, \frac{3}{4n}\right)$ for large $n$, as $E(\bar{Y}_n) = E(Y)$ and $Var(\bar{Y}_n) = \frac{1}{n}Var(Y)$.

**Exercise 5.** Let $X_1, X_2, \ldots$ be i.i.d. positive r.v.s. with mean $\mu$, and let $W_n = \frac{X_1}{X_1+\cdots+X_n}$.

1. Find $\mathbb{E}(W_n)$.
   **Hint:** Consider

$$\frac{X_1}{X_1+\cdots+X_n} + \frac{X_2}{X_1+\cdots+X_n} + \cdots + \frac{X_n}{X_1+\cdots+X_n}.$$

2. What random variable does $nW_n$ converge to (with probability 1) as $n \to \infty$?

**Solution 5.** 1. The expression in the hint equals 1, and by linearity and symmetry its expected value is $n\mathbb{E}(W_n)$. So $\mathbb{E}(W_n) = 1/n$.
**Sanity check:** in the case that the $X_j$ are actually constants, $\frac{X_1}{X_1+\cdots+X_n}$ reduces to $\frac{1}{n}$.

2. By LLN, with probability 1 we have

$$nW_n = \frac{X_1}{(X_1+\cdots+X_n)/n} \to \frac{X_1}{\mu} \quad \text{as } n \to \infty.$$

**Sanity check:** the answer should be a random variable since it's asked what random variable $nW_n$ converges to. It should not depend on $n$ since we let $n \to \infty$.

**Exercise 6.** Suppose that the random variables $X_1$ and $X_2$ have means $\mu_1$ and $\mu_2$ and variances $\sigma_1^2$ and $\sigma_2^2$, with $\text{corr}(X_1, X_2) = \rho$.

1. If $a_1, a_2, b_1, b_2$ are constants, prove that

$$\text{Cov}(a_1X_1 + a_2X_2, b_1X_1 + b_2X_2) = \sum_{i=1}^{2}\sum_{j=1}^{2} a_ib_j\text{Cov}(X_i, X_j).$$

2. Prove the statement below, with or without using part 1.:

$$\text{Var}(a_1X_1 + a_2X_2) = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + 2a_1a_2\rho\sigma_1\sigma_2.$$

3. What is the distribution of $\overline{X}_1 - \overline{X}_2$, for two independent averages $\overline{X}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i$ and $\overline{X}_2 = \frac{1}{n_2}\sum_{j=1}^{n_2} X_j$ for iid $X_i,\ X_j$; satisfying

$$\overline{X}_1 \sim \mathcal{N}\left(\mu_1, \frac{\sigma_1^2}{n_1}\right), \quad \overline{X}_2 \sim \mathcal{N}\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)?$$

   **Hint:** Remember that the sum of two normally distributed random variables is still normally distributed.

   *For the rest of the exercise suppose that $n_1 = n_2 = n$, $\mu_1 = \mu_2 = \mu$, and $\sigma_1 = \sigma_2 = \sigma$.*

4. Using the Chebyshev's inequality give a bound $B$ as a function of $n$, that ensures that the probability of the sample difference $(\overline{X}_1 - \overline{X}_2)$ being further than $B$ away from the true mean of the difference $(\mu - \mu = 0)$ is less than 0.05. That is, find $B$ such that

$$P(|(\overline{X}_1 - \overline{X}_2)| > B) \leq 0.05$$

5. Find the same $B_N$ but instead using the Chebyshev's inequality, use the fact that $\overline{X}_1 - \overline{X}_2$ is normally distributed.

   **Hint:** $\Phi(-1.96) \approx 0.025$

6. Which theorem/result would imply that the $\overline{X}_1 - \overline{X}_2$ is indeed normally distributed, without knowing anything about the distribution of the $X_i,\ X_j$-s

**Solution 6.** 1. Remember that $\text{Cov}(X, aY + bZ) = a\text{Cov}(X, Y) + b\text{Cov}(X, Z)$. Applying this rule sequentially we get

$$\begin{aligned}
\text{Cov}(a_1X_1 + a_2X_2, b_1X_1 + b_2X_2) &= a_1\text{Cov}(X_1, b_1X_1 + b_2X_2) + a_2\text{Cov}(X_2, b_1X_1 + b_2X_2) \\
&= a_1b_1\text{Cov}(X_1, X_1) + a_1b_2\text{Cov}(X_1, X_2) \\
&+ a_2b_1\text{Cov}(X_2, X_1) + a_2b_2\text{Cov}(X_2, X_2) \\
&= \sum_{i=1}^{2}\sum_{j=1}^{2} a_ib_j\text{Cov}(X_i, X_j).
\end{aligned}$$

2. Use the relation $Var(X) = Cov(X, X)$. Then by part 1.

$$\begin{aligned}\mathrm{Var}(a_1X_1 + a_2X_2) &= \mathrm{Cov}(a_1X_1 + a_2X_2, a_1X_1 + a_2X_2)\\ &= a_1a_1\mathrm{Cov}(X_1, X_1) + a_1a_2\mathrm{Cov}(X_1, X_2)\\ &+ a_2a_1\mathrm{Cov}(X_2, X_1) + a_2a_2\mathrm{Cov}(X_2, X_2)\\ &= a_1^2\mathrm{Var}(X_1) + a_2^2\mathrm{Var}(X_2) + 2a_1a_2\mathrm{Cov}(X_1, X_2).\end{aligned}$$

Finally, by definition $Corr(X, Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$, thus
$\mathrm{Cov}(X_1, X_2) = \mathrm{Corr}(X_1, X_2)\sqrt{\mathrm{Var}(X_1)\mathrm{Var}(X_2)} = \rho\sigma_1\sigma_2$.

3. Using the Hint we know that $\overline{X}_1 - \overline{X}_2$ is normally distributed. By the linearity of the expectation

$$E(\overline{X}_1 - \overline{X}_2) = E(\overline{X}_1) - E(\overline{X}_2) = \mu_1 - \mu_2.$$

For the variance we can either use part 2. and that independence implies that the correlation is 0, or that if two random variables are independent, then the variance of the sum is the sum of the variances, i.e.

$$Var(\overline{X}_1 + (-\overline{X}_2)) = Var(\overline{X}_1) + Var(-\overline{X}_2) = Var(\overline{X}_1) + Var(\overline{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.$$

4. By Chebyshev's inequality

$$P(|(\overline{X}_1 - \overline{X}_2)| > B) \leq \frac{Var(\overline{X}_1 - \overline{X}_2)}{B^2}.$$

From part 3., the variance is $\frac{2\sigma^2}{n}$, thus solving the inequality

$$\frac{\frac{2\sigma^2}{n}}{B^2} \leq 0.05,$$

gives $\sqrt{20 \cdot 2\frac{\sigma^2}{n}} \leq B$.

5. We are interested in the probability

$$P(|(\overline{X}_1 - \overline{X}_2)| > B_N) = P((\overline{X}_1 - \overline{X}_2) > B_N) + P((\overline{X}_1 - \overline{X}_2) < -B_N),$$

where the equality follows from the axioms of probability. By standardization

$$P\left(\frac{\overline{X}_1 - \overline{X}_2}{\sqrt{2\sigma^2/n}} > \frac{B_N}{\sqrt{2\sigma^2/n}}\right) + P\left(\frac{\overline{X}_1 - \overline{X}_2}{\sqrt{2\sigma^2/n}} < -\frac{B_N}{\sqrt{2\sigma^2/n}}\right) = 1 - \Phi\left(\frac{B_N}{\sqrt{2\sigma^2/n}}\right) + \Phi\left(-\frac{B_N}{\sqrt{2\sigma^2/n}}\right).$$

By symmetry

$$1 - \Phi\left(\frac{B_N}{\sqrt{2\sigma^2/n}}\right) + \Phi\left(-\frac{B_N}{\sqrt{2\sigma^2/n}}\right) = 2\Phi\left(-\frac{B_N}{\sqrt{2\sigma^2/n}}\right)$$

Using the hint if $\frac{B_N}{\sqrt{2\sigma^2/n}} > 1.96$, then $2\Phi\left(-\frac{B_N}{\sqrt{2\sigma^2/n}}\right) \leq 0.05$. Thus the desired inequality holds as long as $B_N \geq 1.96 \cdot \sqrt{\frac{2\sigma^2}{n}}$.

6. Informally, the central limit theorem says that for reasonably large $n$, the sample means approximate two normal distributions, and consequently their difference is approximately normal as well.

   By using the central limit theorem instead of the Chebyshev's inequality we can reduce the probabilistic bounds by quite a big margin:

$$\frac{1.96 \cdot \sqrt{\frac{2\sigma^2}{n}}}{\sqrt{20 \cdot 2\frac{\sigma^2}{n}}} = \frac{1.96}{\sqrt{20}} \approx 0.44.$$