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. $E(E(X|Y)+E(Y|X))$ _____ $E(X)+E(Y)$

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

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

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.

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?

Exercise 5. Let $X_1, X_2, \ldots$ be i.i.d. positive r.v.s. with mean μ, 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 $n \to \infty$?

Exercise 6. Suppose that the random variables X_1 and X_2 have means μ_1 and μ_2 and variances σ_1^2 and σ_2^2, with $\text{corr}(X_1, X_2) = \rho$.

1. If a_1, a_2, b_1, b_2 are constants, prove that
   $$\text{Cov}(a_1X_1 + a_2X_2, b_1X_1 + b_2X_2) = \sum_{i=1}^{2}\sum_{j=1}^{2} a_i b_j \text{Cov}(X_i, X_j).$$
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