ECE 131A
Probability and Statistics
Instructor: Lara Dolecek

Homework 7 Solutions
Wednesday, March 3, 2021
Due: Wednesday, March 10, 2021
before class begins

TAs: Lev Tauz
Debarnab Mitra

levtauz@ucla.edu
debarnabucla@g.ucla.edu

Please upload your homework to Gradescope by March 10, 3:59 pm.
Please submit a single PDF directly on Gradescope
You may type your homework or scan your handwritten version. Make sure all the work is discernible.

Chapters 6.4 and 7.1-7.3 *Probability, Statistics, and Random Processes* by A. Leon-Garcia

1. Consider two random variables X and Y. Prove that the correlation coefficient $\rho_{X,Y}$ satisfies $-1 \leq \rho_{X,Y} \leq 1$. **Hint:** Consider the function $\mathbb{E}[(\frac{X-\mathbb{E}[X]}{\sigma_X} \pm \frac{Y-\mathbb{E}[Y]}{\sigma_Y})^2]$.

   Solution:

$$\begin{aligned}
0 &\leq \mathbb{E}[(\frac{X-\mathbb{E}[X]}{\sigma_X} \pm \frac{Y-\mathbb{E}[Y]}{\sigma_Y})^2] \\
&= \mathbb{E}[(\frac{X-\mathbb{E}[X]}{\sigma_X})^2] \pm 2\mathbb{E}[\frac{(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])}{\sigma_X\sigma_Y}] + \mathbb{E}[(\frac{Y-\mathbb{E}[Y]}{\sigma_Y})^2] \\
&= 1 \pm 2\rho_{X,Y} + 1 = 2(1 \pm \rho_{X,Y}) \\
&\implies -1 \leq \rho_{X,Y} \leq 1
\end{aligned}$$

2. Two points are picked uniformly at random in the interval $[0, L]$. What is the expected distance between these points?

   Solution: Let the selected points be X_1 and X_2 such that $0 \leq X_1 \leq L$, $0 \leq X_2 \leq L$. Since we select the points X_1 and X_2 uniformly at random, the joint pdf of X_1 and X_2 becomes:

$$f_{X_1X_2}(x_1, x_2) = \begin{cases} \frac{1}{L^2} & \text{for } 0 \leq x_1 \leq L, 0 \leq x_2 \leq L \\ 0 & \text{otherwise} \end{cases}$$

   where the constant $\frac{1}{L^2}$ is chosen such that $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X_1X_2}(x_1, x_2)dx_1dx_2 = 1$. The expected distance is

$$\begin{aligned}
E(|X_1 - X_2|) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |x_1 - x_2| f_{X_1X_2}(x_1, x_2)dx_1dx_2 \\
&= 2\int_{x_1=0}^{L}\int_{x_2=x_1}^{L} (x_2 - x_1)\frac{1}{L^2}dx_1dx_2 \\
&= \frac{L}{3}
\end{aligned}$$

   Symmetry Argument: We can also solve this problem using symmetry: Suppose X_1 is the smaller point and X_2 is the larger point. The three segments on the line $(X_1, X_2 -$

$X_1, L - X_2)$ are identically distributed due to symmetry. Thus $E[X_1] = E[X_2 - X_1] = E[L - X_2]$ Since $X_1 + X_2 - X_1 + L - X_2 = L$, by linearity of expectation we get the following:

$$\begin{aligned} X_1 + X_2 - X_1 + L - X_2 &= L \\ \implies E[X_1] + E[X_2 - X_1] + E[L - X_2] &= L \\ \implies E[X_2 - X_1] + E[X_2 - X_1] + E[X_2 - X_1] &= L \\ \implies E[X_2 - X_1] &= \frac{L}{3} \end{aligned}$$

3. Consider the jointly Gaussian random variables X and Y that have the following joint PDF:

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{x^2}{\sigma_X^2} + \frac{y^2}{\sigma_Y^2} - \frac{2\rho xy}{\sigma_X\sigma_Y}\right)\right].$$

(a) Prove that Y is a Gaussian random variable by deriving its marginal PDF, $f_Y(y)$. Find the mean and variance of Y.
Solution:
The marginal PDF of Y, $f_Y(y)$ is derived as follows:

$$\begin{aligned} f_Y(y) &= \int_{x=-\infty}^{\infty} f_{X,Y}(x,y)dx \\ &= \int_{x=-\infty}^{\infty} \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{x^2}{\sigma_X^2} + \frac{y^2}{\sigma_Y^2} - \frac{2\rho xy}{\sigma_X\sigma_Y}\right)\right] dx. \end{aligned}$$

To perform this integral, we need to complete a square inside the argument of the exponential.

$$\begin{aligned} &-\frac{1}{2(1-\rho^2)}\left(\frac{x^2}{\sigma_X^2} + \frac{y^2}{\sigma_Y^2} - \frac{2\rho xy}{\sigma_X\sigma_Y}\right) \\ &= -\frac{1}{2(1-\rho^2)}\left(\left[\frac{x}{\sigma_X} - \frac{\rho y}{\sigma_Y}\right]^2 - \frac{\rho^2 y^2}{\sigma_Y^2} + \frac{y^2}{\sigma_Y^2}\right) \\ &= -\frac{1}{2(1-\rho^2)}\left[\frac{x}{\sigma_X} - \frac{\rho y}{\sigma_Y}\right]^2 - \frac{1}{2(1-\rho^2)}\frac{(1-\rho^2)y^2}{\sigma_Y^2} \\ &= -\frac{1}{2(1-\rho^2)\sigma_X^2}\left[x - \frac{\rho\sigma_X y}{\sigma_Y}\right]^2 - \frac{y^2}{2\sigma_Y^2}. \end{aligned}$$

Substituting this exponential argument in the integral of $f_Y(y)$ gives us:

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_Y} \exp\left[-\frac{y^2}{2\sigma_Y^2}\right] \int_{x=-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_X\sqrt{1-\rho^2}} \exp\left[-\frac{\left[x - \frac{\rho\sigma_X y}{\sigma_Y}\right]^2}{2\sigma_X^2(1-\rho^2)}\right] dx$$

The value of this integral is 1 because it is the pdf of a Gaussian with mean $\frac{\rho\sigma_X y}{\sigma_Y}$ and variance $2\sigma_X^2(1-\rho^2)$. Thus,

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_Y} \exp\left[-\frac{y^2}{2\sigma_Y^2}\right], \quad -\infty < y < -\infty,$$

which proves that Y is a Gaussian random variable with mean 0 and variance σ_Y^2.

(b) Prove that $f_{X|Y}(x|y)$ corresponds to another Gaussian random variable by determining its closed form equation, then find its mean and variance. **Solution:** The conditional PDF $f_{X|Y}(x|y)$ is derived as follows:

$$\begin{aligned} f_{X|Y}(x|y) &= f_{X,Y}(x,y)/f_Y(y) \\ &= \frac{1}{\sqrt{2\pi}\sigma_X\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{x^2}{\sigma_X^2}+\frac{y^2}{\sigma_Y^2}-\frac{2\rho xy}{\sigma_X\sigma_Y}\right)+\frac{y^2}{2\sigma_Y^2}\right]. \end{aligned}$$

One more time, we operate on the exponential argument:

$$\begin{aligned} &-\frac{1}{2(1-\rho^2)}\left(\frac{x^2}{\sigma_X^2}+\frac{y^2}{\sigma_Y^2}-\frac{2\rho xy}{\sigma_X\sigma_Y}\right)+\frac{y^2}{2\sigma_Y^2} \\ &= -\frac{1}{2(1-\rho^2)}\left[\frac{x}{\sigma_X}-\frac{\rho y}{\sigma_Y}\right]^2 + \frac{1}{2(1-\rho^2)}\left[\frac{\rho^2 y^2}{\sigma_Y^2}-\frac{y^2}{\sigma_Y^2}\right]+\frac{y^2}{2\sigma_Y^2} \\ &= -\frac{1}{2(1-\rho^2)}\left[\frac{x}{\sigma_X}-\frac{\rho y}{\sigma_Y}\right]^2 - \frac{y^2}{2\sigma_Y^2}+\frac{y^2}{2\sigma_Y^2} \\ &= -\frac{1}{2\sigma_X^2(1-\rho^2)}\left[x-\frac{\rho\sigma_X y}{\sigma_Y}\right]^2. \end{aligned}$$

Consequently, we conclude that:

$$f_{X|Y}(x|y) = \frac{1}{\sqrt{2\pi}\sigma_X\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2\sigma_X^2(1-\rho^2)}\left[x-\frac{\rho\sigma_X y}{\sigma_Y}\right]^2\right],$$

where $-\infty < x < \infty$. This proves that $f_{X|Y}(x|y)$ corresponds to another Gaussian random variable with mean $\rho\sigma_X y/\sigma_Y$, and variance $\sigma_X^2(1-\rho^2)$.

4. Let X and Y be jointly Gaussian random variables with $\mathbb{E}[Y] = 0$, $\sigma_X = 1$, $\sigma_Y = 2$ and $\mathbb{E}[X|Y] = \frac{Y}{4} + 1$.

   (a) Find the joint pdf of X and Y.

   Solution:
   Let m_x and m_y be the means of X and Y. By the definition of marginal distributions for jointly Gaussian random variables, we know that

$$\mathbb{E}[X|Y] = \rho_{X,Y}\frac{\sigma_X}{\sigma_Y}(Y-m_y)+m_x.$$

As such,

$$\mathbb{E}[X|Y] = \frac{Y}{4} + 1 \implies m_x = 1, \rho_{X,Y} = \frac{1}{2}.$$

Now we have all the parameters to find the joint pdf of X and Y which is

$$\begin{aligned} f_{X,Y}(x,y) &= \frac{\exp[\frac{1}{2(1-\rho_{X,Y}^2)}((\frac{x-m_x}{\sigma_X})^2 - 2\rho_{X,Y}(\frac{x-m_x}{\sigma_X})(\frac{y-m_y}{\sigma_Y}) + (\frac{y-m_y}{\sigma_Y})^2)]}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{X,Y}^2}} \\ &= \frac{\exp[\frac{2}{3}((x-1)^2 - \frac{(x-1)y}{2} + (\frac{y}{2})^2)]}{2\pi\sqrt{3}} \end{aligned}$$

(b) Find the conditional pdf of $f_{X|Y}(x|y)$.
Solution:

Since X conditioned on Y is still a Gaussian, we only need to find its mean and variance to calculate its pdf. The mean is already given in the problem statement, i.e. $\mathbb{E}[X|Y] = \frac{Y}{4} + 1$. From class, the variance is $Var(X|Y) = \sigma_X^2(1 - \rho_{X,Y}^2) = \frac{3}{4}$. Thus, the conditional pdf is

$$f(x|y) = \frac{1}{\sqrt{\frac{3}{2}\pi}} e^{\frac{-2(x-(\frac{y}{4}+1))^2}{3}}$$

5. Assume that $X_1, X_2, ..., X_n$ are independent random variables with possibly different distributions and let S_n be their sum. Let $m_k = E(X_k)$, $\sigma_k^2 = VAR(X_k)$, and $M_n = m_1 + m_2 + \cdots + m_n$. Assume that $\sigma_k^2 < R$ and $m_k < T$ for all k. Prove that, for any $\epsilon > 0$,

$$P(|\frac{S_n}{n} - \frac{M_n}{n}| < \epsilon) \to 1$$

as $n \to \infty$.
Solution:

$$\frac{S_n}{n} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

$$E[\frac{S_n}{n}] = \frac{m_1 + m_2 + \cdots + m_n}{n} = \frac{M_n}{n}$$

$$VAR[\frac{S_n}{n}] = \frac{VAR(X_1 + X_2 + \cdots + X_n)}{n^2} = \frac{\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2}{n^2} \le \frac{nR}{n^2} = \frac{R}{n},$$

using the Chebyshev's inequality we get:

$$\begin{aligned} P[|\frac{S_n}{n} - \frac{M_n}{n}| \ge \epsilon] &\le \frac{Var(\frac{S_n}{n})}{\epsilon^2} \\ &\le \frac{R}{n\epsilon^2} \to 0 \text{ as } n \to \infty \end{aligned}$$

Thus,

$$P\left(\left|\frac{S_n}{n} - \frac{M_n}{n}\right| < \epsilon\right) \to 1 \text{ as } n \to \infty.$$

6. *Application of CLT.*

   (a) A fair coin is tossed 100 times. Estimate the probability that the number of heads is between 40 and 60. Estimate the probability that the number is between 50 and 55.
   Solution:
   Let $n = 100$ and let $X_i = 1$ if i^{th} toss is a head, else 0. Then $S_n = X_1 + \ldots + X_n$. We have $\mu = nE[X_i] = 50$ and $\sigma^2 = nVAR[X_i] = np(1-p) = 25$. The central limit theorem gives:

$$\begin{aligned} P[40 \leq S_n \leq 60] &= P[\frac{40-50}{\sqrt{25}} \leq \frac{S_n - \mu}{\sigma} \leq \frac{60-50}{\sqrt{25}}] \\ &= Q(-2) - Q(2) = 0.9544 \end{aligned}$$

   Similarly,

$$\begin{aligned} P[50 \leq S_n \leq 55] &= P[\frac{50-50}{\sqrt{25}} \leq \frac{S_n - \mu}{\sigma} \leq \frac{55-50}{\sqrt{25}}] \\ &= Q(0) - Q(1) = 0.3413 \end{aligned}$$

   (b) Repeat part (a) for if we toss the coin 1000 times and for the intervals [400,600] and [500,550].
   Solution:
   We have $n = 1000$, $\mu = 500$, $\sigma^2 = 250$. The central limit theorem gives:

$$\begin{aligned} P[400 \leq S_n \leq 600] &= P[\frac{400-500}{\sqrt{250}} \leq \frac{S_n - \mu}{\sigma} \leq \frac{600-500}{\sqrt{250}}] \\ &= Q(-6.32) - Q(6.32) \approx 1 \end{aligned}$$

$$\begin{aligned} P[500 \leq S_n \leq 550] &= P[\frac{500-500}{\sqrt{250}} \leq \frac{S_n - \mu}{\sigma} \leq \frac{550-500}{\sqrt{250}}] \\ &= Q(0) - (3.16) = 0.4992 \end{aligned}$$

   (c) Suppose that 20% of voters are in favor of certain legislation. A large number n of voters are polled and a relative frequency estimate $f_A(n)$ for the above proportion is obtained. Use central limit theorem to estimate how many voters should be polled in order that the probability is at least .95 that $f_A(n)$ differs from 0.20 by less than 0.02.
   Solution:
   Let us define the relative frequency $f_A(n)$ as $\frac{X}{n}$, where X is the total number of

voters in favor of legislation. Since X is binomial random variable, then for $f_A(n)$ we have:

$$\mu = E[f_A(n)] = E(\frac{X}{n}) = \frac{E[X]}{n} = \frac{np}{n} = 0.2$$

$$\sigma^2 = VAR[f_A(n)] = VAR(\frac{X}{n}) = \frac{VAR(X)}{n^2} = \frac{np(1-p)}{n^2} = \frac{0.16}{n}$$

Now,

$$\begin{aligned} P[|f_A(n) - 0.2| < 0.02] &= P[0.18 < f_A(n) < 0.22] \\ &= P[\frac{0.18 - 0.2}{\sqrt{\frac{0.16}{n}}} \leq \frac{f_A(n) - \mu}{\sigma} \leq \frac{0.22 - 0.2}{\sqrt{\frac{0.16}{n}}}] \\ &= 1 - 2Q(\frac{0.02}{\sqrt{\frac{0.16}{n}}}) \geq 0.95 \end{aligned}$$

Therefore we have $Q(\frac{\sqrt{n}}{20}) \leq 0.025$, then the number of voters n should be at least $n \geq 1521$.