Problem Set 2 - Random Variables, Distributions and Expectations

ECO 104 - Statistics for Business and Economics - I, Summer-2025

Faculty: Sk. Tanvir Hossain
Teaching Assistant: Habiba Afroz

`Due Date - 3rd August (Sec 9) and 4th August (Sec 7),`
`Group Submision (Max 3 Members) -` **`Please Submit Hard Copy`** `in Class`

1. Summation formulas
   (a) Show that the following formula for the sample variance

   $$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

   is equal to

   $$s^2 = \frac{1}{n}\left(\sum_{i=1}^{n} x_i^2\right) - \bar{x}^2$$

   Notice in this case we don't have $n-1$ in the denominator.

   (b) Also you can write a similar sample covariance formula,

   $$s_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

   Show that this can be rewritten as (this one is easier to use in some cases)

   $$s_{xy} = \frac{1}{n}\left(\sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y}\right)$$

2. A discrete random variable X has the following PMF:

x	1	2	3	4
$f(x)$	0.1	0.3	0.4	0.2

   (a) Verify that this is a valid PMF. Plot the PMF.
   (b) Claulate $\mathbb{P}(X \leq 2)$ and $\mathbb{P}(X > 2)$.
   (c) Calculate $\mathbb{P}(X = 3)$ and $\mathbb{P}(X \neq 3)$.
   (d) Calculate $\mathbb{P}(1 \leq X \leq 3)$, $\mathbb{P}(1 \leq X < 3)$ and $\mathbb{P}(X \geq 3)$.
   (e) Calculate $\mathbb{E}(X)$ and $\mathbb{V}(X)$.
   (f) Create a new random variable $Y = X^2$ and calculate $\mathbb{E}(Y)$ and $\mathbb{V}(Y)$.
   (g) Create a new random variable $Y = 2X + 3$ and calculate $\mathbb{E}(Y)$ and $\mathbb{V}(Y)$.

(h) Create a new random variable $Y = 2X^2 + 3$ and calculate $\mathbb{E}(Y)$ and $\mathbb{V}(Y)$.

(i) Let $\mu = \mathbb{E}(X)$ and $\sigma^2 = \mathbb{V}(X)$. Now create a new random variable $Z = \frac{X-\mu}{\sigma}$ what are the values of Z, calculate $\mathbb{E}(Z)$ and $\mathbb{V}(Z)$ where (Note: In this case we say Z is a standardized version of the random variable X).

(j) Calculate Cumulative Distribution Function (CDF) $F(x)$ for $x = 1, 2, 3, 4$. and plot the CDF.

3. Answer the following conceptual questions

(a) Using LOTUS show that $\mathbb{E}(aX^2 + bX + c) = a\mathbb{E}(X^2) + b\mathbb{E}(X) + c$ for any random variable X and constants a, b and c (But it's a linearity of expectation, so when you apply, you can apply directly without using LOTUS).

(b) Suppose we have a random variable X with mean $\mu = \mathbb{E}(X)$ and variance $\sigma^2 = \mathbb{E}((X-\mu)^2)$. Now define another random variable $Z = \frac{X-\mu}{\sigma}$, show that $\mathbb{E}(Z) = 0$ and variance of $\mathbb{V}(Z) = 1$ (Note: this holds for any random variable X regardless of its distribution, and it's a very important property).

(c) A student claims that for any random variable X, $\mathbb{E}(X^2) = [\mathbb{E}(X)]^2$. Provide a counterexample and explain why this is generally false.

(d) If two random variables X and Y have the same expected value, do they necessarily have the same distribution? Provide an example to support your answer.

(e) Explain the difference between empirical PMF and theoretical PMF. Construct an example to illustrate this difference.

4. We already know the Mean and Variance of a distribution, recall population mean gives the central location of the distribution and variance gives the spread of the distribution. Now we will calculate some other summary measures of a distribution, which are called **skewness and kurtosis**.

Let X and Y be two rando, barwith PMF given by:

x	1	2	3	4	5
$f(x)$	0.1	0.2	0.3	0.2	0.2

and

y	1	2	3	4	5
$f(y)$	0.4	0.3	0.2	0.1	0.0

(a) Plot the PMF of X and Y.

(b) Calculate the mean and variance of X and Y.

(c) Calculate $\mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]$, where $\mu = \mathbb{E}(X)$ and $\sigma^2 = \mathbb{V}(X)$. **This number measures the skewness of the distribution of X. which shows how symmetric the distribution is around its mean**.

(d) Calculate the skewness of Y

(e) Which distribution is more skewed, X or Y?

(f) Calculate $\mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^4\right]$. This number measures the kurtosis of the distribution of X.

(g) Calculate the kurtosis of Y, which distribution is more peaked, X or Y?

(h) Calculate skewness and kurtosis of the random variable X.