DS203: Programming in Data Science
IE605: Engineering Statistics

Introduction to Probability and Statistics
Lecture 04

Manjesh K. Hanawal

28th August 2020

Previous Lecture:

- Distribution of functions of random variable
- Generate RVs with a given distribution

This Lecture:

- Joint distributed Random Variable
- Marginal PMF and PDF
- Independence of Random Variables
- Correlation of Random Variables

Jointly Distributed Random Variables

Let RVs $X = (X_1, X_2, X_3, \ldots, X_m)$ are defined on the same Ω.

Joint CDF of X is a map $F_X : \mathbb{R}^m \to [0, 1]$ given by

$$F_X(x_1, x_2, \ldots, x_m) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_m \leq x_m).$$

Jointly Distributed Random Variables

Let RVs $X = (X_1, X_2, X_3, \ldots, X_m)$ are defined on the same Ω.

Joint CDF of X is a map $F_X : \mathbb{R}^m \to [0, 1]$ given by

$$F_X(x_1, x_2, \ldots, x_m) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_m \leq x_m).$$

Example 1: n coins tossed $X = (X_1, X_2, \ldots, X_n)$, where X_i is outcome of ith coin. We may be interested in finding $P(X_1 = 1, X_2 = 0, X_3 = 0, \ldots, X_n = 1)$

Jointly Distributed Random Variables

Let RVs $X = (X_1, X_2, X_3, \ldots, X_m)$ are defined on the same Ω.

Joint CDF of X is a map $F_X : \mathbb{R}^m \to [0, 1]$ given by

$$F_X(x_1, x_2, \ldots, x_m) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_m \leq x_m).$$

Example 1: *n* coins tossed $X = (X_1, X_2, \ldots, X_n)$, where X_i is outcome of ith coin. We may be interested in finding $P(X_1 = 1, X_2 = 0, X_3 = 0, \ldots, X_n = 1)$

Example : Portfolio Management

$X = (X_1, X_2, \ldots, X_n)$, where X_i is the amount invested in ith share/stock. C is the amount available. $\sum_{i=1}^{n} X_i = C$.

Marginal Densities

- For two variables: $F_X(x_1, x_2) = P(X_1 \leq x_1, X_2 \leq x_2)$.
  $F_{X_1}(x_1) = \lim_{x_2 \to \infty} F_X(x_1, x_2)$ and $F_{X_2}(x_2) = \lim_{x_1 \to \infty} F_X(x_1, x_2)$
- $F_{X_1}(x_1)$ and $F_{X_2}(x_2)$ are **marginal CDF** of X_1 and X_2

Marginal Densities

- For two variables: $F_X(x_1, x_2) = P(X_1 \leq x_1, X_2 \leq x_2)$.
  $F_{X_1}(x_1) = \lim_{x_2 \to \infty} F_X(x_1, x_2)$ and $F_{X_2}(x_2) = \lim_{x_1 \to \infty} F_X(x_1, x_2)$
- $F_{X_1}(x_1)$ and $F_{X_2}(x_2)$ are **marginal CDF** of X_1 and X_2

Discrete RVs:

- If X_1 and X_2 are both discrete, we can define joint PMF as
  $P_X(x_1, x_2) = P(X_1 = x_1, X_2 = x_2)$ and $\sum_{x_1, x_2} P_X(x_1, x_2) = 1$.
  $P_{X_1}(x_1) = \sum_{x_2} P(X_1 = x_1, X_2 = x_2)$, similarly for $P_{X_2}(x_2)$
- $P_{X_1}(x_1)$ and $P_{X_2}(x_2)$ are **marginal PMF** of X_1 and X_2

Marginal Densities

- For two variables: $F_X(x_1, x_2) = P(X_1 \leq x_1, X_2 \leq x_2)$.
  $F_{X_1}(x_1) = \lim_{x_2 \to \infty} F_X(x_1, x_2)$ and $F_{X_2}(x_2) = \lim_{x_1 \to \infty} F_X(x_1, x_2)$
- $F_{X_1}(x_1)$ and $F_{X_2}(x_2)$ are **marginal CDF** of X_1 and X_2

Discrete RVs:

- If X_1 and X_2 are both discrete, we can define joint PMF as
  $P_X(x_1, x_2) = P(X_1 = x_1, X_2 = x_2)$ and $\sum_{x_1, x_2} P_X(x_1, x_2) = 1$.
  $P_{X_1}(x_1) = \sum_{x_2} P(X_1 = x_1, X_2 = x_2)$, similarly for $P_{X_2}(x_2)$
- $P_{X_1}(x_1)$ and $P_{X_2}(x_2)$ are **marginal PMF** of X_1 and X_2

Example: $X = (X_1, X_2)$ where $X_1 \in \{1, 2, 3\}$ and $X_2 \in \{2, 4, 5\}$ with joint PMF given by

$P(X_1, X_2)$	$X_2 = 2$	$X_2 = 4$	$X_2 = 5$
$X_1 = 1$	.1	.05	.2
$X_1 = 2$	.1	.1	.15
$X_1 = 3$	.15	.1	0.05

$P_{X_1}(1) =$ $\quad$ $P_{X_2}(2) =$

$P_{X_1}(2) =$ $\quad$ $P_{X_2}(4) =$

$P_{X_1}(3) =$ $\quad$ $P_{X_2}(5) =$

Continuous Case

We say $X = (X_1, X_2, X_3, \ldots, X_m)$ are **jointly continuous** if $\exists \;\; f_X : R^m \to R$ such that for any $(x_1, x_2, \ldots, x_m) \in \mathbb{R}^m$

$$F_X(x_1, \ldots, x_m) = \int_{\infty}^{x_1} \cdots \int_{\infty}^{x_m} f_X(y_1, y_2, \ldots, y_m) dy_1 dy_2 \ldots dy_m.$$

f_X is called the **joint PDF** of X

Continuous Case

We say $X = (X_1, X_2, X_3, \ldots, X_m)$ are **jointly continuous** if $\exists \ f_X : R^m \to R$ such that for any $(x_1, x_2, \ldots, x_m) \in \mathbb{R}^m$

$$F_X(x_1, \ldots, x_m) = \int_{\infty}^{x_1} \cdots \int_{\infty}^{x_m} f_X(y_1, y_2, \ldots, y_m) dy_1 dy_2 \ldots dy_m.$$

f_X is called the **joint PDF** of X

Example 1: Weather Report
$X = (X_1, X_2)$, where X_1 denote the humidity level and X_2 is the temperature.

Example 2: Healthcare
$X = (X_1, X_2)$, where X_1 denote blood sugar level and X_2 could be BMI.

Continuous case contd.

- If X_1 and X_2 are jointly continous with PDF f_X $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_X(x_1, x_2)dx_1 dx_2 = 1$.
- Define $f_{X_1}(x_1) = \int_{-\infty}^{\infty} f_X(x_1, x_2)dx_1$, similarly for $f_{X_2}(x_2)$
- $f_{X_1}(x_1)$ and $f_{X_2}(x_2)$ are **marginal PDF** of X_1 and X_2

Continuous case contd.

- If X_1 and X_2 are jointly continous with PDF f_X $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_X(x_1, x_2)dx_1 dx_2 = 1$.
- Define $f_{X_1}(x_1) = \int_{-\infty}^{\infty} f_X(x_1, x_2)dx_1$, similarly for $f_{X_2}(x_2)$
- $f_{X_1}(x_1)$ and $f_{X_2}(x_2)$ are **marginal PDF** of X_1 and X_2

Example: $X = (X_1, X_2)$ is jointly continuous with PDF given by

$$f_X(x_1, x_2) = \begin{cases} c(1 + x_1 x_2) & \text{if } 2 \leq x_1 \leq 3, 1 \leq x_2 \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

What is $f_{X_1}(x_1)$?

Independence of RVs

$X := (X_1, X_2, \ldots, X_m)$ are independent if its joint CDF is such that for all $x_i \in \mathbb{R}, i = 1, 2 \ldots, m$,

$$F_X(x_1, x_2, \ldots x_m) = F_{X_1}(x_1)F_{X_2}(x_2) \ldots F_{X_m}(x_m)$$

This simplifies to for the case of two RVs as

- Discrete case: $P_X(x_1, x_2) = P_{X_1}(x_1)P_{X_2}(x_2)$
- Continuous case: $f_X(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2)$
- For independent RVs it is enough to specify their marginal PMF/PDF.

Independence of RVs contd..

Example: n coins tossed: $X = (X_1, X_2, \ldots, X_n)$, where
$X_i \sim Ber(p_i)$ and X_is are independent.
$P(X_1 = x_1, X_2 = x_2..X_n = x_n) = P_{X_1}(x_1) \times P_{X_1}(x_1) \times .. \times P_{X_n}(x_n)$.

Special Case: If $p_i = p$, $\sum_{i=1}^{n} X_i \sim Bin(n, p)$.

Property of Independent RVs $(X_1, X_2, \ldots, X_n)$ are independent

$$\implies E(X_1 X_2, \times \ldots, X_n) = E(X_1)E(X_2)\ldots E(X_n)$$

Let $X = (X_1, X_2, \ldots, X_n)$ are independent and each random variable has the same distribution, then $(X_1, X_2, \ldots, X_n)$ are said to be **independent and identically distributed (i.i.d.)**.

For i.i.d distributed random variables, we just need to specify one common distribution!

Covariance of RVs

Covariance of random variable X_1 and X_2 is defined as
$$Cov(X_1, X_2) = E\big((X_1 - E(X_1))(X_2 - E(X_2))\big)$$

- $Cov(X_1, X_2) = E(X_1X_2) - E(X_1)E(X_2)$
- If X_1 and X_2 are independent $Cov(X_1, X_2) = 0$
- What does $|Cov(X_1, X_2)| > 0$ indicates?

Covariance of RVs

Covariance of random variable X_1 and X_2 is defined as
$$Cov(X_1, X_2) = E\big((X_1 - E(X_1))(X_2 - E(X_2))\big)$$

- $Cov(X_1, X_2) = E(X_1X_2) - E(X_1)E(X_2)$
- If X_1 and X_2 are independent $Cov(X_1, X_2) = 0$
- What does $|Cov(X_1, X_2)| > 0$ indicates?

X_1 and X_2 are defined as indicators of two events A and B

$$X_1 = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{otherwise} \end{cases} \qquad X_2 = \begin{cases} 1 & \text{if } B \text{ occurs} \\ 0 & \text{otherwise} \end{cases}$$

Covariance of RVs

Covariance of random variable X_1 and X_2 is defined as
$$Cov(X_1, X_2) = E\big((X_1 - E(X_1))(X_2 - E(X_2))\big)$$

- $Cov(X_1, X_2) = E(X_1X_2) - E(X_1)E(X_2)$
- If X_1 and X_2 are independent $Cov(X_1, X_2) = 0$
- What does $|Cov(X_1, X_2)| > 0$ indicates?

X_1 and X_2 are defined as indicators of two events A and B

$$X_1 = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{otherwise} \end{cases} \qquad X_2 = \begin{cases} 1 & \text{if } B \text{ occurs} \\ 0 & \text{otherwise} \end{cases}$$

$$Cov(X_1, X_2) = P(X_1 = 1, X_2 = 1) - P(X_1 = 1)P(X_2)$$

$$\begin{aligned} Cov(X_1, X_2) > 0 &\iff P(X_1 = 1, X_2 = 1) > P(X_1 = 1)P(X_2 = 1) \\ &\iff \frac{P(X_1 = 1, X_2 = 1)}{P(X_2 = 1)} > P(X_1 = 1) \\ &\iff P(X_1 = 1 | X_2 = 1) > P(X_1 = 1) \end{aligned}$$

Properties of Covariance

- $|Cov(X_1, X_2)| > 0$ indicates that occurrence or nonoccurence of X_2 improves knowledge of X_1 and they are correlated.
- $Cov(X_1, X_2) > 0$ is an indication that when X_1 increases X_2 also increases and vice versa.
- $Cov(X_1, X_2) < 0$ is an indication that when X_1 decreases X_2 also decreases and vice versa.

Properties of Covariance

- $|Cov(X_1, X_2)| > 0$ indicates that occurrence or nonoccurence of X_2 improves knowledge of X_1 and they are correlated.
- $Cov(X_1, X_2) > 0$ is an indication that when X_1 increases X_2 also increases and vice versa.
- $Cov(X_1, X_2) < 0$ is an indication that when X_1 decreases X_2 also decreases and vice versa.

- $Cov(X_1, X_1) = Var(X_1)$
- $Cov(X_1, X_2) = Cov(X_2, X_1)$
- $Cov(aX_1, X_2) = aCov(X_1, X_2)$
- $Cov(X_1 + X_2, X_3) = Cov(X_1, X_2) + Cov(X_1, X_3)$

(Verify!)

Fundamental Theorems of Probability

let $X_1, X_2, X_3, \ldots$ be a sequence of RVs all defined on the same Ω. Assume they are i.i.d with mean $E(X_1)$ and $= Var(X_1)$. Define $S_n = \sum_{i=1}^{n} X_i$ for all $n \geq 1$.

Law of Large Numbers: $\lim_{n\to\infty} \frac{S_n}{n} = E(X_1)$

Central Limit Theorem: $\lim_{n\to\infty} \frac{S_n - nE(X_1)}{\sqrt{nVar(X_1)}} \equiv \mathcal{N}(0,1)$

Example 1: X_i's are i.i.d with $X_i \sim Exp(\lambda)$. Then $\lim_{n\to\infty} \frac{S_n}{n} = \lambda$

Example 1: X_i's are i.i.d with $X_i \sim Poi(\lambda)$. Then $\lim_{n\to\infty} \frac{S_n}{n} = \lambda$

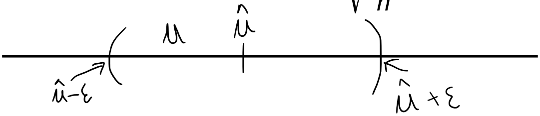
Confidence Interval

- In real life we will have only finite samples. .
- Let $\mu = E(X_1)$ and $\hat{\mu} = \frac{S_n}{n}$ (estimate). $|\hat{\mu} - \mu| \neq 0$
- We would like to know $\left|\hat{\mu} - \mu\right| > \epsilon$ for some $\epsilon > 0$

$$P\left(|\hat{\mu} - \mu| > \epsilon\right) \leq 2\exp(-n\epsilon^2)$$

$$2\exp(-n\epsilon^2) = \delta \implies n = \frac{1}{\epsilon^2}\log(\delta/2)$$

$$2\exp(-n\epsilon^2) = \delta \implies \epsilon = \sqrt{\frac{1}{n}\log(\delta/2)}$$



End!