IE605: Engineering Statistics

Lecture 05

Manjesh K. Hanawal

Previous Lecture:

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

This Lecture:

- Joint distribution of function of RVs
- Moment Generating Functions (MGFs)
- Conditional PMF and PDF
- Markov's and Chebyshev's inequalities
- Limit theorems: Law of Large Numbers (LLN)
- Limit theorems: Central Limit Theorem (CLT)

Joint Distribution of Function of Random Variables

Let X_1 and X_2 are RVs. Define $Y_1 = g_1(X_1, X_2)$, $Y_2 = g_2(X_1, X_2)$. What is the joint distribution of (Y_1, Y_2).

- Example 1: Sum and Difference of Coordinates
  $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$
- Example 2: (Cartesian to Polar Coordinates)
  $Y_1 = \sqrt{X_1^2 + X_2^2}$ and $Y_2 = \tan^{-1}\left(\frac{X_2}{X_1}\right)$

Joint Distribution of Function of Random Variables

Let X_1 and X_2 are RVs. Define $Y_1 = g_1(X_1, X_2)$, $Y_2 = g_2(X_1, X_2)$. What is the joint distribution of (Y_1, Y_2).

- Example 1: Sum and Difference of Coordinates
  $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$
- Example 2: (Cartesian to Polar Coordinates)
  $Y_1 = \sqrt{X_1^2 + X_2^2}$ and $Y_2 = \tan^{-1}\left(\frac{X_2}{X_1}\right)$

$$F(y_1, y_2) = \iint\limits_{\substack{(x_1, x_2) \\ g_1(x_1, x_2) \leq y_1 \\ g_2(x_1, x_2) \leq y_2}} f(x_1, x_2)\, dx_1\, dx_2$$

Joint Distribution Contd...

Assume:

- For given (y_1, y_2), $y_1 = g_1(x_1, x_2)$ and $y_2 = g_2(x_1, x_2)$ are uniquely solvable for x_1 and x_2,
- g_1 and g_2 have continuous partial derivatives such that the Jacobian matrix is non-singular

$$J(x_1, x_2) = \begin{vmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} \end{vmatrix} = \frac{\partial g_1}{\partial x_1}\frac{\partial g_2}{\partial x_2} - \frac{\partial g_2}{\partial x_1}\frac{\partial g_1}{\partial x_2} \neq 0$$

$$f_{Y_1 Y_2}(y_1, y_2) = \frac{f_{X_1 X_2}(x_1, x_2)}{|J(x_1, x_2)|}.$$

Joint Distribution Contd...

Example: $Y_1 = g_1(X_1, X_2) = X_1 + X_2$, $Y_2 = g_2(X_1, X_2) = X_1 - X_2$.
where $X_1 \sim Exp(\lambda_1)$, $X_2 \sim Exp(\lambda_2)$ and are independent.
Given (y_1, y_2), $x_1 = \frac{y_1+y_2}{2}$ and $x_2 = \frac{y_1-y_2}{2}$,

$$J(x_1, x_2) = \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} = -2$$

$$f_Y(y_1, y_2) = \frac{f_X(x_1, x_2)}{2} = \frac{f_{X_1}(x_1) f_{X_2}(x_2)}{2} = \lambda_1 \lambda_2 e^{-\lambda_1 \frac{(y_1+y_2)}{2} - \lambda_2 \frac{(y_1-y_2)}{2}}$$

Joint Distribution Contd...

Example: $Y_1 = g_1(X_1, X_2) = X_1 + X_2,\ Y_2 = g_2(X_1, X_2) = X_1 - X_2.$
where $X_1 \sim Exp(\lambda_1),\ X_2 \sim Exp(\lambda_2)$ and are independent.
Given (y_1, y_2), $x_1 = \frac{y_1+y_2}{2}$ and $x_2 = \frac{y_1-y_2}{2}$,

$$J(x_1, x_2) = \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} = -2$$

$$f_Y(y_1, y_2) = \frac{f_X(x_1, x_2)}{2} = \frac{f_{X_1}(x_1) f_{X_2}(x_2)}{2} = \lambda_1 \lambda_2 e^{-\lambda_1 \frac{(y_1+y_2)}{2} - \lambda_2 \frac{(y_1-y_2)}{2}}$$

Example 2: (Cartesian to Polar Coordinates)
$Y_1 = \sqrt{X_1^2 + X_2^2}$, $Y_2 = \tan^{-1}\left(\frac{X_2}{X_1}\right)$, where $X_1 \sim \mathcal{N}(0, 1)$ and $X_2 \sim \mathcal{N}(0, 1)$ are independent.
Joint distribution of polar coordinates (Y_1, Y_2)?

Given (y_1, y_2), $x_1 = y_1 \cos(y_2)$ and $x_2 = y_1 \sin(y_2)$ (complete!)

Moment Generating Functions

Moment Generating Function (MGF) of a random variable X, denoted $\phi_X : \mathcal{R} \to \mathcal{R}_+$, is defined as

$$\phi_X(t) = \mathbb{E}\left(e^{tX}\right) = \begin{cases} \sum_{i=1}^{\infty} P_X(x_i) e^{tx_i} & \text{if } X \text{ is discrete} \\ \int_x f_X(x) e^{tx} dx & \text{if } X \text{ is continuous.} \end{cases}$$

Moment Generating Functions

Moment Generating Function (MGF) of a random variable X, denoted $\phi_X : \mathcal{R} \to \mathcal{R}_+$, is defined as

$$\phi_X(t) = \mathbb{E}\left(e^{tX}\right) = \begin{cases} \sum_{i=1}^{\infty} P_X(x_i) e^{tx_i} & \text{if } X \text{ is discrete} \\ \int_x f_X(x) e^{tx} dx & \text{if } X \text{ is continuous.} \end{cases}$$

- $\phi'_X(t) = \mathbb{E}\left(Xe^{tX}\right) \to \phi'_X(0) = \mathbb{E}(X)$ (first moment)
- $\phi''_X(t) = \mathbb{E}\left(X^2 e^{tX}\right) \to \phi''_X(0) = \mathbb{E}(X^2)$ (second moment)
- $\phi^{(n)}_X(t) = \mathbb{E}\left(X^n e^{tX}\right) \to \phi^{(n)}_X(0) = \mathbb{E}(X^n)$ (nth moment)

From MGF of a RV all its moment can be generated, specifically, mean and variance.

$$Var(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$$

Properties of MGF

- MGF of sum of independent RVs: $(X_1, X_2, \ldots, X_n)$ are independent, then

$$\phi_{X_1+X_2\ldots+X_n}(t) = \prod_{i=1}^{n} \phi_{X_i}(t)$$

- MGF uniquely determines the distributions
  one-to-one correspondence between the MGF and distribution.

Properties of MGF

- MGF of sum of independent RVs: $(X_1, X_2, \ldots, X_n)$ are independent, then

$$\phi_{X_1+X_2\ldots+X_n}(t) = \prod_{i=1}^{n} \phi_{X_i}(t)$$

- MGF uniquely determines the distributions
  one-to-one correspondence between the MGF and distribution.

Distribution	MGF $\phi(t)$
$Ber(p)$	$pe^t + (1-p)$
$Bin(n,p)$	$(pe^t + (1-p))^n$
$Geo(n,p)$	$\frac{pe^t}{1-(1-p)e^t}$
$Poi(\lambda)$	$e^{\lambda(e^t-1)}$

Distribution	MGF $\phi(t)$
$Uni(a,b)$	$\frac{e^{bt}-e^{at}}{t(b-a)}$
$Exp(\lambda)$	$\frac{\lambda}{\lambda-t}$
$\mathcal{N}(\mu,\sigma^2)$	$\exp\{\mu t + \frac{t^2\sigma^2}{2}\}$
$Gamma(n,\lambda)$	$\left(\frac{\lambda}{\lambda-t}\right)^n$

Characteristic function: $\Phi_X(t) = \mathbb{E}(e^{jtX})$, where $j = \sqrt{-1}$ (always exists).

Conditional PMF

X_1 and X_2 are discrete with joint PMF $P(X_1 = x_1, X_2 = x_2)$. We may want to know about $X_2 = x_2$ given that $X_1 = x_1$

- Conditional PMF is defined as

$$P_{X_2|X_1}(x_2|x_1) = P(X_2 = x_2|X_1 = x_1) = \frac{P(X_2 = x_2, X_1 = x_1)}{P(X_1 = x_1)}$$

- Conditional CDF is $F_{X_2|X_1}(x_2|x_1) = \sum_{x \leq x_2} P_{X_2|X_1}(x|x_1)$
- Conditional Expectation:

$$\mathbb{E}(X_2|X_1 = x_1) = \sum_x x P_{X_2|X_1}(x|x_1)$$

Conditional PMF

X_1 and X_2 are discrete with joint PMF $P(X_1 = x_1, X_2 = x_2)$. We may want to know about $X_2 = x_2$ given that $X_1 = x_1$

- Conditional PMF is defined as

$$P_{X_2|X_1}(x_2|x_1) = P(X_2 = x_2|X_1 = x_1) = \frac{P(X_2 = x_2, X_1 = x_1)}{P(X_1 = x_1)}$$

- Conditional CDF is $F_{X_2|X_1}(x_2|x_1) = \sum_{x \le x_2} P_{X_2|X_1}(x|x_1)$
- Conditional Expectation:

$$\mathbb{E}(X_2|X_1 = x_1) = \sum_x x P_{X_2|X_1}(x|x_1)$$

Example:

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

$P_{X_2|X_1}(4|1) =$
$\mathbb{E}(X_2|X_1 = 1) =$
$P_{X_2|X_1}(5|2) =$
$\mathbb{E}(X_2|X_1 = 2) =$

Conditional PDF

X_1 and X_2 are jointly continuous with PDF $f(X_1 = x_1, X_2 = x_2)$.
We may want to know PDF of $X_2 = x_2$ given that $X_1 = x_1$

- Conditional PDF is defined as

$$f_{X_2|X_1}(x_2|x_1) = \frac{f_{X_1 X_2}(x_1, x_2)}{f_{X_1}(x_1)}$$

- Conditional Expectation:

$$\mathbb{E}(X_2|X_1 = x_1) = \int_x x f_{X_2|X_1}(x|x_1)dx$$

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

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

Find $f_{X_2|X_1}(x_2|2.5)$ and $\mathbb{E}(X_2|X_1 = 2.5)$.

Markov's Inequality

Let X is a non-negative RV. For any $a > 0$

$$P(X \geq a) \leq \frac{\mathbb{E}(X)}{a}$$

Useful when $a \geq \mathbb{E}(X)$.

$$\begin{aligned} X &= X\mathbb{1}_{\{X<a\}} + X\mathbb{1}_{\{X\geq a\}} \\ \implies \mathbb{E}(X) &= \mathbb{E}\left(X\mathbb{1}_{\{X<a\}}\right) + \mathbb{E}\left(X\mathbb{1}_{\{X\geq a\}}\right) \\ &\geq 0 + a\mathbb{E}\left(\mathbb{1}_{\{X\geq a\}}\right) \\ &= aP(X \geq a) \end{aligned}$$

Chebyshev's Inequality

For any RV X and $d > 0$

$$P(|X - \mathbb{E}(X)| \geq d) \leq \frac{Var(X)}{d^2}$$

$|X - \mathbb{E}(X)|$ is non-negative. We apply Markov inequality on it.

$$\begin{aligned} P(|X - \mathbb{E}(X)| \geq d) &= P(|X - \mathbb{E}(X)|^2 \geq d^2) \\ &\leq \frac{\mathbb{E}\left(|X - \mathbb{E}(X)|^2\right)}{d^2} \\ &= \frac{Var(X)}{d^2} \end{aligned}$$

Chebyshev's inequlity bounds 'deviation' of a RV around its mean.

Application of Markov and Chebyshev's inequality

Factory output: Suppose a factory produce a certain number of items each week. The number of items produced is random due to uncertainty in availability raw material. Suppose that the factory produce on an average 500 items every week.

- What is the probability that production this week is at least 1000? Let number of items produced is X. We want $P(X \geq 1000)$. From Markov Inequality

$$P(X \geq 1000) \leq \frac{500}{1000} = 0.5$$

- If $Var(X) = 100$, what is the probability that production this week is between 400 and 600? We want $P(400 < X < 600) = P(|X - \mathbb{E}(X)| < 100)$. From Chebyshev's inequality

$$P(|X - \mathbb{E}(X)| \geq 100) \leq \frac{100}{100^2} = \frac{1}{100}$$

Hence

$$P(|X - \mathbb{E}(X)| < 100) = 1 - P(|X - \mathbb{E}(X)| \geq 100) \geq \frac{99}{100}$$

Limit Theorems: Law of Large Numbers (LLN)

Let $X_1, X_2, X_3, \ldots$ be a sequence of i.i.d. RVs with common mean $\mu = E(X_1)$. Define $S_n = \sum_{i=1}^{n} X_i$ for all $n \geq 1$.

$$\text{LLN: } \lim_{n\to\infty} \frac{S_n}{n} = E(X_1) \text{ with probability 1.}$$

Consider an event E in an experiment. The experiment is repeated infinitely. For each trial i define RV X_i

$$X_i = \begin{cases} 1 & \text{if event E occurs} \\ 0 & \textit{otherwise} \end{cases}$$

- $S_n = \sum_{i=1}^{n} X_i$, counts the number of time E occurs
- S_n/n gives fraction of time E occurs
- From LLN $\lim_{n\to\infty} \frac{S_n}{n} = \mathbb{E}(X_1) = P(E)$
- Fraction of time event E occurs is its probability

LLN Contd..

Examples

1. X_i's are i.i.d with $X_i \sim Exp(\lambda)$. Then $\lim_{n\to\infty} \dfrac{S_n}{n} = \dfrac{1}{\lambda}$
2. X_i's are i.i.d with $X_i \sim Poi(\lambda)$. Then $\lim_{n\to\infty} \dfrac{S_n}{n} = \lambda$
3. X_i's are i.i.d with some unknown mean. $\lim_{n\to\infty} \dfrac{S_n}{n}$ gives the mean.

LLN for parameter estimation.

- In real life we will have only finite samples. .
- We can use $\hat{\mu}_n = \frac{S_n}{n}$ as an (estimate) of μ.
- For finite n, $|\hat{\mu}_n - \mu| \neq 0$. $\lim_{n\to\infty} |\hat{\mu}_n - \mu| = 0$

Limit Theorem: Central Limit Theorem (CLT)

Let $X_1, X_2, X_3, \ldots$ be a sequence of i.i.d. RVs with common mean $\mu = E(X_1)$ and $\sigma^2 = Var(X)$. Define $S_n = \sum_{i=1}^{n} X_i$ for all $n \geq 1$.

CLT: $\lim_{n\to\infty} \frac{S_n - n\mu}{\sqrt{n\sigma^2}} \equiv \mathcal{N}(0,1)$ in distributions.

For any $a \in \mathcal{R}$.

$$P\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq a\right) \approx \int_{-\infty}^{a} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = \Phi(a).$$

- $\Phi(\cdot)$ is the CDF of $\mathcal{N}(0,1)$.
- $\Phi(a) + \Phi(-a) = 1$ for all a (symmetry of $\mathcal{N}(0,1)$)
- $\Phi(\cdot)$ tables are used.

CLT Contd..

$$P\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq a\right) = P\left(\frac{S_n}{n} - \mu \leq a\sqrt{\frac{\sigma^2}{n}}\right)$$

$$= P\left(\hat{\mu}_n - \mu \leq a\sqrt{\frac{\sigma^2}{n}}\right) \approx \Phi(a).$$

Example: 100 i.i.d. samples are available of an experiments with variance 5 and unknown mean. What is the probability that error in estimate mean $(\hat{\mu}_n)$ is no more that 0.1.

- Unknown mean is μ. We want $P\left(-0.1 \leq \hat{\mu}_{100} - \mu \leq 0.1\right)$.

$$P\left(-0.1 \leq \hat{\mu}_{100} - \mu \leq 0.1\right)$$
$$= P\left(\hat{\mu}_{100} - \mu \leq 0.1\right) - P\left(\hat{\mu}_{100} - \mu \leq -0.1\right)$$
$$\approx \Phi\left(0.1\sqrt{20}\right) - \Phi\left(-0.1\sqrt{20}\right) = 2\Phi\left(0.1\sqrt{20}\right) - 1$$

Number of Samples Required

Suppose we want the estimation error to be smaller than $\epsilon > 0$

$$P\left(|\hat{\mu}_n - \mu| \leq \epsilon\right) \approx 2\Phi\left(\epsilon\sqrt{n/\sigma^2}\right) - 1.$$

add we want this probability to be smaller than δ. Then we set

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