Probability and Statistics (MA6.101)

Monsoon 2021, IIIT Hyderabad
23 November, Tuesday (Lecture 24)

Taught by Prof. Pawan Kumar

Probability

Covariance

The covariance between two RVs X and Y is

$$\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y].$$

The properties of covariance are as follows:

- $\text{Cov}(X, X) = \text{Var}(X)$
- If X, Y are independent, $\text{Cov}(X, Y) = 0$ (but the converse is not necessarily true).
- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
- $\text{Cov}(aX, Y) = a\text{Cov}(X, Y)$
- $\text{Cov}(X + c, Y) = \text{Cov}(X, Y)$
- $\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$

More generally, we have

$$\text{Cov}\left(\sum_i a_i X_i, \sum_j b_j Y_j\right) = \sum_i \sum_j a_i b_j \text{Cov}(X_i, Y_j).$$

We can use the covariance to find the variance of a sum. If $Z = X + Y$, then

$$\text{Var}(Z) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y).$$

More generally, if $Z = aX + bY$ for $a, b \in \mathbb{R}$,

$$\text{Var}(Z) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X, Y).$$

Correlation Coefficient

The correlation coefficient ρ_{XY} or $\rho(X,Y)$ of two RVs X, Y is defined as

$$\rho_{XY} = \rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} = \frac{\mathrm{Cov}(X,Y)}{\rho_X \rho_Y}.$$

We can define the standardised version U of a random variable X as

$$U = \frac{X - E[X]}{\rho_X},$$

in terms of which we can say

$$\rho_{XY} = \mathrm{Cov}(U,V),$$

where V is the standardised version of Y.

Some properties of the correlation coefficient are:

- $-1 \leq \rho_{XY} \leq 1$
- If $\rho(X,Y) = 1$, then $Y = ax + b$, where $a > 0$.

If $\rho_{XY} = 0$, then X, Y are *uncorrelated*; if it is > 0, they are *positively correlated*; if it is < 0, it is *negatively correlated*.

If X and Y are uncorrelated, $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$. This holds for the sum of set of pairwise uncorrelated variables also.

Bivariate Normal Distribution

Two RVs X, Y are called bivariate or jointly normal if $aX + bY$ has a normal ditribution for all $a, b \in \mathbb{R}$. Clearly this implies that X and Y are independently normal.

If they are independent and normal, then they are jointly normal. Further, if $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$, then

$$(X+Y) \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2 + 2\rho(X,Y)\sigma_X\sigma_Y).$$

The definition of a standard bivariate normal distribution can be defined analogously as

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

where $\rho \in (-1, 1)$.

If X, Y are bivariate normal *and* uncorrelated, then they are independent.

Union Bound and Extension

The union bound states tha the probability of the union of events is smaller than the first term in the expansion of the inclusion-exclusion principle. Thus, for $n = 2$, we have

$$P(A \cup B) \leq P(A) + P(B),$$

and more generally,

$$P\left(\cup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i).$$

Every time we add one more term to the expansion, we reverse the inequality. Thus we have

$$P\left(\cup_{i=1}^{n} A_i\right) \geq \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j),$$

and

$$P\left(\cup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cup A_j \cup A_k).$$

Markov and Chebyshev Inequalities

If X is a nonnegative RV, then

$$P(X \geq a) \leq \frac{E[X]}{a}.$$

This is the Markov inequality.

Moreover, if $b > 0$, then

$$P(|X - E[X]| \geq b) \leq \frac{\mathrm{Var}(X)}{b^2}.$$

This is the Chebyshev inequality.

Chernoff Bound

If X is an RV and $a \in \mathbb{R}$, let $M_X(s) = E[e^{sX}]$ be the MGF. Then

$$P(X \geq a) \leq e^{-sa} M_X(s), \forall s > 0,$$

$$P(X \leq a) \leq e^{-sa} M_X(s), \forall s < 0.$$

Since this holds for any s, we have

$$P(X \geq a) = \min_{s>0} e^{-sa} M_X(s),$$

$$P(X \leq a) = \min_{s<0} e^{-sa} M_X(s).$$

Cauchy-Schwarz Inequality

For any two RVs X, Y,

$$E[XY] \leq \sqrt{E[X^2]E[Y^2]}.$$

It can be used to prove that $|\rho_{XY}| \leq 1$.

Convex Functions and Jensen's Inequality

A function $g : I \to \mathbb{R}$ is convex if for any two $x, y \in I$ and any $\alpha \in [0, 1]$, we have

$$g(\alpha x + (1 - \alpha)y) \leq \alpha g(x) + (1 - \alpha)g(y).$$

If the above inequality is $\geq$, then g is concave.

Jensen's inequality states that if $g(x)$ is convex on R_X and $E[g(X)]$ and $g(E[X])$ are finite, then

$$E[g(X)] \geq g(E[X]).$$

Sample Mean

Let $X_1, \ldots, X_n$ be n independent and identically distributed (same mean) RVs. Their sample mean is then defined as

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}.$$

Weak Law of Large Numbers

Let X_i be n i.i.d RVs with mean $E[X_i] = \mu < \infty$. Then for any $\epsilon > 0$, \$$\lim_{n\to\infty} P(|X - \mu| \geq \infty) = 0$.

Central Limit Theorem

Let X_i be n i.i.d RVs with mean $E[X_i] = \mu < \infty$ and variance $0 < \mathrm{Var}(X_i) = \sigma^2 < \infty$. Then,

$$Z_n = \frac{X - \mu}{\sigma/\sqrt{n}} = \frac{\sum_{i=1}^{n} - n\mu}{\sqrt{n}\sigma}$$

converges in distribution to the standard normal distribution as n goes to infinity, *i.e.*

$$\lim_{n\to\infty} P(X_n \leq z) = \Phi(z), \forall x \in \mathbb{R},$$

where Φ is the standard normal CDF. Note that is does not matter what the X_i's distributions are.

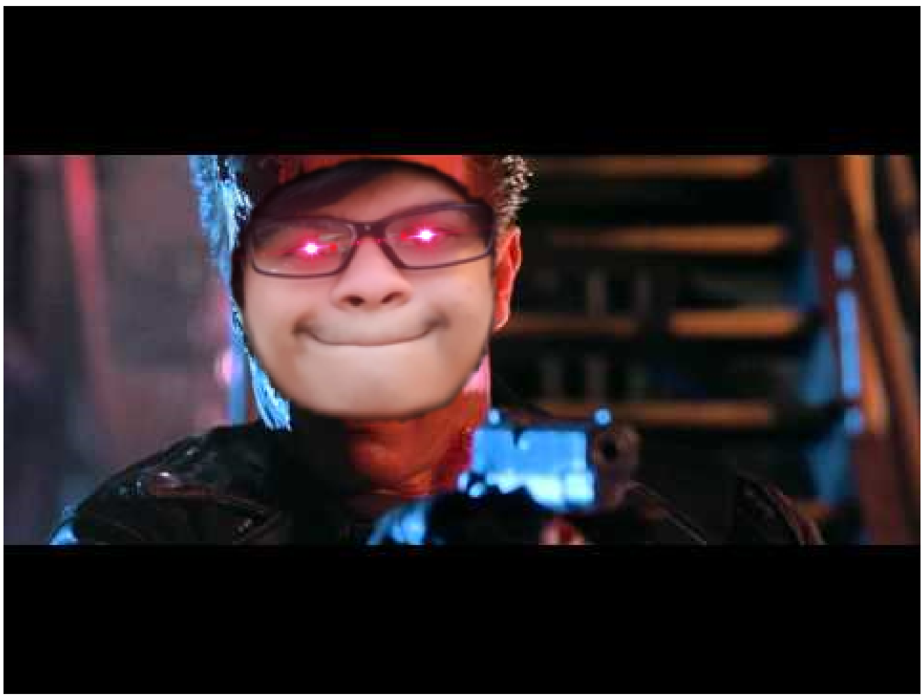

Figure 1: George's Art