Modern Statistics **Lecture 6 - 3/9/2025**

Lecture 6 Expectation & convergance of R.V.

Lecturer:Xiangyu Chang *Scribe: Zipeng Zheng,Xiangdong Meng*

Edited by: Zhihong Liu

1 Recall

$$\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}(X))^2\right]. \tag{1}$$

2 Covariance and Correlation

Definition 1 (Covariance)**.** *Covariance measures joint variability — the extent of variation between two random variables:*

$$\mathrm{Cov}(X,Y) \triangleq \mathbb{E}\left[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\right]. \tag{2}$$

Definition 2 (Correlation)**.** *Correlation is a measure of the degree of linear relationship between two variables:*

$$\rho(X,Y) \triangleq \frac{\mathrm{Cov}(X,Y)}{\sigma_X \cdot \sigma_Y} \quad \text{where } \sigma_X^2 = \mathrm{Var}(X),\ \sigma_Y^2 = \mathrm{Var}(Y). \tag{3}$$

2.1 Properties of Covariance

1. $\mathrm{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$.

   Proof.

   $$\begin{aligned}\mathrm{Cov}(X,Y) &= \mathbb{E}\left[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\right], \\ &= \mathbb{E}[XY - X\mathbb{E}(Y) - Y\mathbb{E}(X) + \mathbb{E}(X)\mathbb{E}(Y)], \\ &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).\end{aligned}$$

   Special case $X = Y$: $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$.

   $$\begin{aligned}\mathrm{Cov}(X,X) &= \mathbb{E}\left[(X - \mathbb{E}(X))^2\right], \\ &= \mathbb{E}(X^2) - [\mathbb{E}(X)]^2, \\ &= \mathrm{Var}(X).\end{aligned}$$

   □

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

   Proof.

   $$\begin{aligned}\mathrm{Cov}(X,Y+Z) &= \mathbb{E}\left[(X - \mathbb{E}(X))(Y + Z - \mathbb{E}(Y+Z))\right], \\ &= \mathrm{Cov}(X,Y) + \mathrm{Cov}(X,Z).\end{aligned}$$

   □

3. $\mathrm{Cov}(aX, bY) = ab \cdot \mathrm{Cov}(X, Y)$.

   Proof.

$$\begin{aligned}\mathrm{Cov}(aX, bY) &= ab \cdot \mathbb{E}(XY) - ab \cdot \mathbb{E}(X)\mathbb{E}(Y), \\ &= ab \cdot \mathrm{Cov}(X, Y).\end{aligned}$$

   □

4. Symmetry: $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$.
5. If $Y = aX + b$, then $\mathrm{Cov}(X, Y) = a \cdot \mathrm{Var}(X)$.

   Proof.

$$\begin{aligned}\mathrm{Cov}(X, Y) &= \mathrm{Cov}(X, aX) + \mathrm{Cov}(X, b), \\ &= a \cdot \mathrm{Var}(X).\end{aligned}$$

   □

6. $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X, Y)$.

   Proof.

$$\begin{aligned}\mathrm{Var}(X + Y) &= \mathrm{Cov}(X + Y, X + Y), \\ &= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X, Y).\end{aligned}$$

   □

7. $\mathrm{Var}\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2\,\mathrm{Var}(X_i) + 2\sum_{1 \le i < j \le n} a_i a_j\,\mathrm{Cov}(X_i, X_j)$.

   Proof.

$$\begin{aligned}\mathrm{Var}\left(\sum_{i=1}^n a_i X_i\right) &= \sum_{i=1}^n \sum_{j=1}^n a_i a_j\,\mathrm{Cov}(X_i, X_j), \\ &= \sum_{i=1}^n a_i^2\,\mathrm{Var}(X_i) + 2 \sum_{1 \le i < j \le n} a_i a_j\,\mathrm{Cov}(X_i, X_j).\end{aligned}$$

   □

2.2 Properties of Correlation

1. If $Y = aX + b$, then $\rho(X, Y) = \pm 1$.

   Proof.

$$\begin{aligned}\rho(X, Y) &= \frac{a \cdot \mathrm{Var}(X)}{\sqrt{\mathrm{Var}(X)} \cdot |a| \sqrt{\mathrm{Var}(X)}}, \\ &= \frac{a}{|a|} = \pm 1.\end{aligned}$$

   □

2. Boundedness: $|\rho_{X,Y}| \le 1$ (equivalent to $|\mathrm{Cov}(X, Y)| \le \sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}$).

Proof. Case 1: Discrete variables.

$$\text{By Cauchy-Schwarz: } \left(\sum_{i=1}^{n} X_i Y_i\right)^2 \leq \left(\sum_{i=1}^{n} X_i^2\right)\left(\sum_{i=1}^{n} Y_i^2\right).$$
$$\Rightarrow |\mathbf{X}^T \mathbf{Y}| \leq \|\mathbf{X}\|_2 \cdot \|\mathbf{Y}\|_2.$$

Case 2: Continuous variables.

$$\text{By Cauchy-Schwarz: } \left(\int XY \; dF_{X,Y}\right)^2 \leq \left(\int X^2 dF_X\right)\left(\int Y^2 dF_Y\right).$$
$$\Rightarrow (\mathbb{E}(XY))^2 \leq \mathbb{E}(X^2)\mathbb{E}(Y^2).$$

General case:

$$\text{Let } X' = X - \mathbb{E}X,\; Y' = Y - \mathbb{E}Y.$$
$$(\mathbb{E}(X'Y'))^2 \leq \mathbb{E}(X'^2)\mathbb{E}(Y'^2).$$
$$\Rightarrow \operatorname{Cov}(X,Y)^2 \leq \operatorname{Var}(X)\operatorname{Var}(Y).$$

□

3 Expectation of Random Vector

$$\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_k \end{pmatrix} \in \mathbb{R}^k, \quad \mathbb{E}[\mathbf{X}] \triangleq \begin{pmatrix} \mathbb{E}X_1 \\ \vdots \\ \mathbb{E}X_k \end{pmatrix}.$$

$$\operatorname{Var}(\mathbf{X}) \triangleq \mathbb{E}\left[(\mathbf{X} - \mathbb{E}\mathbf{X})(\mathbf{X} - \mathbb{E}\mathbf{X})^\top\right] \in \mathbb{R}^{k\times k}.$$

$$\operatorname{Cov}(X_i, X_j) = \mathbb{E}[(X_i - \mathbb{E}X_i)(X_j - \mathbb{E}X_j)].$$

3.1 Properties

1. Covariance matrix representation:

$$\overline{Z} = \begin{pmatrix} \operatorname{Var}(X_1) & \operatorname{Cov}(X_1,X_2) & \cdots & \operatorname{Cov}(X_1,X_k) \\ \operatorname{Cov}(X_2,X_1) & \operatorname{Var}(X_2) & \cdots & \operatorname{Cov}(X_2,X_k) \\ \vdots & \vdots & \ddots & \vdots \\ \operatorname{Cov}(X_k,X_1) & \operatorname{Cov}(X_k,X_2) & \cdots & \operatorname{Var}(X_k) \end{pmatrix}.$$

The covariance matrix $\overline{Z}$ is positive semi-definite.

Proof. For any non-zero vector $\mathbf{a} = (a_1, \ldots, a_k)^\top \in \mathbb{R}^k$:

$$\mathbf{a}^\top \overline{Z} \mathbf{a} = \sum_{i=1}^{k} \sum_{j=1}^{k} a_i a_j \operatorname{Cov}(X_i, X_j),$$
$$= \operatorname{Var}\left(\sum_{i=1}^{k} a_i X_i\right) \geq 0.$$

Alternative proof using matrix notation:

$$\begin{aligned}\mathbf{a}^\top \overline{Z}\mathbf{a} &= \mathbf{a}^\top \mathbb{E}[(\mathbf{X}-\mathbb{E}\mathbf{X})(\mathbf{X}-\mathbb{E}\mathbf{X})^\top]\mathbf{a}, \\ &= \mathbb{E}[\mathbf{a}^\top(\mathbf{X}-\mathbb{E}\mathbf{X})(\mathbf{X}-\mathbb{E}\mathbf{X})^\top \mathbf{a}], \\ &= \mathbb{E}\left[\left(\mathbf{a}^\top(\mathbf{X}-\mathbb{E}\mathbf{X})\right)^2\right] \geq 0.\end{aligned}$$

Dimension compatibility:

Component	**Dimension**
$\mathbf{a}^\top \in \mathbb{R}^{1\times k}$	$(\mathbf{X}-\mathbb{E}\mathbf{X}) \in \mathbb{R}^{k\times 1}$
$\mathbf{a} \in \mathbb{R}^{k\times 1}$	$(\mathbf{X}-\mathbb{E}\mathbf{X})^\top \in \mathbb{R}^{1\times k}$

.

The non-negativity follows from the expectation of squared terms. □

E.g. $(X_1, X_2) \to \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix}.$

If $X_1 = X_2$, then:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 \end{pmatrix}.$$

and $\mathrm{Rank}(\Sigma) = 1$.

E.g. Multinomial (n, p)

$$p = \begin{pmatrix} p_1 \\ \vdots \\ p_k \end{pmatrix} \quad \sum_k p_k = 1 \quad \sum_k X_k = n.$$

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_k \end{pmatrix} \sim \text{Multinomial } (n, p).$$

$$X_k \sim \text{Binomial}(n, \hat{p}_k).$$

$$\mathrm{Var}(X_k) = n\hat{p}_k(1-\hat{p}_k).$$

$$X_k = \sum_{i=1}^n X_{ki}, \quad X_{ki} \sim \mathrm{Ber}(\hat{p}_k).$$

$$\mathbb{E}(X_{k\cdot i}) = \hat{p}_k.$$

$$\mathrm{Var}(X_{k\cdot i}) = \hat{p}_k(1-\hat{p}_k).$$

$$\Sigma_{12} = \mathrm{Cov}(X_1, X_2) = \mathbb{E}X_1X_2 - \mathbb{E}X_1\mathbb{E}X_2,$$

$$= \mathbb{E}X_1X_2 - n^2\hat{p}_1\hat{p}_2.$$

$$\mathbb{E}X_1X_2 = \mathbb{E}\left[\left(\sum_{i=1}^{n} X_{1i}\right)\left(\sum_{j=1}^{n} X_{2j}\right)\right],$$

$$= \mathbb{E}\left[\sum_{i,j} X_{1i}X_{2j}\right] = \sum_{i,j} \mathbb{E}[X_{1i}X_{2j}],$$

$$= h(n-1)\hat{p}_1\hat{p}_2.$$

$$\Sigma_{12} = -n\hat{p}_1\hat{p}_2.$$

$$\Sigma_{ij} = -n\hat{p}_i\hat{p}_j \quad i \neq j.$$

$$\Sigma_{ii} = n\hat{p}_i(1-\hat{p}_i).$$

Conditional Expectation:

$$\mathbb{E}[Y \mid X] = \begin{cases} \sum_y y \cdot f_{Y|X}(y) = \sum_y y \cdot p(y \mid X), & \text{if } Y \text{ is discrete,} \\ \int y f_{Y|X}(y)\, dy, & \text{if } Y \text{ is continuous.} \end{cases}$$

Theorem 1 (the rule of iterative expectation)**.**

$$\mathbb{E}_X[\mathbb{E}_{Y|X}[Y \mid X]] = \mathbb{E}_Y[Y].$$

Proof. $\mathbb{E}[\mathbb{E}[Y \mid X = x]] = \int \mathbb{E}[Y \mid X = x] f_X(x)\, dx,$

$$= \iint y f_{Y|X}(y \mid x) f_X(x)\, dx\, dy,$$

$$= \iint y f_{Y|X}(x, y)\, dx\, dy.$$

$$= \iint y f_Y \cdot f_{X|Y}\, dy\, dx,$$

$$= \int y f_Y\, dy = \mathbb{E}[Y].$$

□

$\mu = \mathbb{E}[X].$

Sample Mean: $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \to \mathbb{E}\bar{X}_n = \mu.$

$$\lim_{n\to\infty} \bar{X}_n = \mu \quad ?$$

Sequence Limit:

$$\lim_{n\to\infty} X_n = X.$$

$$\forall\varepsilon > 0, \exists N(\varepsilon) \text{ such that if } n \geq N \text{ then } |X_n - X| \leq \varepsilon.$$

Convergence in Probability For any $\varepsilon > 0$, if

$$\lim_{n\to\infty} P(|X_n - X| \geq \varepsilon) = 0,$$

then $X_n \xrightarrow{P} X$.

Convergence in Quadratic Mean:

$$X_n \xrightarrow{L^2} X \text{ if } \lim_{n\to\infty} \mathbb{E}(X_n - X)^2 = 0.$$

Convergence in distribution:

$$X_n \xrightarrow{d} X \text{ if } \lim_{n\to\infty} F_n(t) = F(t) \text{ at continuous point.}$$