# Modern Statistics

Xiangyu Chang

March 21, 2026

**Abstract**

To be undated.

# 1 Lecture 6: Exceptation and Variance

## 1.1 Recall

### 1.1.1 Definition of Expectation

**Definition 1.1** (Expectation)**.** The expected value, or mean, or first moment, of $X$ is defined to be

$$\mathbb{E}(X) = \int x dF(x) = \begin{cases} \sum_x xf(x) & \text{if } X \text{ is discrete} \\ \int xf(x)dx & \text{if } X \text{ is continuous} \end{cases}$$

- **The sufficient conditions for the existence of** $\mathbb{E}(X)$.

$$\begin{cases} |\mathbb{E}(X)| < +\infty, & \text{if } X \text{ is discrete} \\ \int |x| f_X(x)dx < +\infty, & \text{if } X \text{ is continuous} \end{cases}$$

## 1.2 The $k^{th}$ moment of $X$

**Definition 1.2** ($k^{th}$ moment)**.** The $k^{th}$ moment of $X$ is defined to be $\mathbb{E}(X^k)$ ,assuming that $\mathbb{E}(|X|^k) < \infty$.

**Theorem 1.3.** *If* $\mathbb{E}(X^k) < \infty$ *(exist),* $k \geqslant 1$*,* $i \leqslant k$*, then* $\mathbb{E}(X^i) < \infty$ *(exist)*

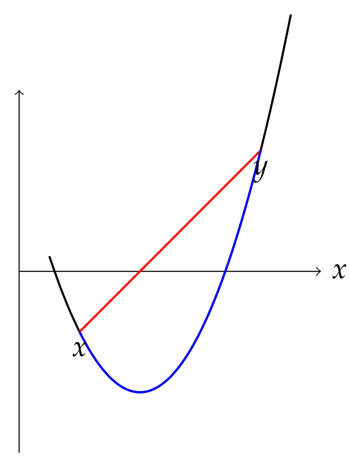
Proof:

$$\begin{aligned}
\mathbb{E}(X^i) &= \int_{\mathbb{R}} |x|^i dF_X(x) \\
&= \int_{|x|>1} |x|^i dF_X(x) + \int_{|x|\leqslant 1} |x|^i dF_X(x) \\
&\leqslant 1 + \int_{|x|>1} |x|^i dF_X(x) \\
&\leqslant 1 + \int_{|x|>1} |x|^k dF_X(x) \\
&\leqslant 1 + \int_{\mathbb{R}} |x|^k dF_X(x) \\
&= 1 + \mathbb{E}(|X|^k) < \infty
\end{aligned}$$

## 1.3 Properties of Expectation

1. $\mathbb{E}(\sum_{i=1}^n a_i X_i) = \sum_{i=1}^n a_i \mathbb{E}(X_i)$
2. If $X_1, X_2, \cdots, X_k$ are mutually independent, then $\mathbb{E}(\prod_{i=1}^k X_i) = \prod_{i=1}^k \mathbb{E}(X_i)$
3. Convex Function: For all $0 \leqslant \lambda \leqslant 1$, $g(\lambda x + (1-\lambda)y) \leqslant \lambda g(x) + (1-\lambda)g(y)$

$$\Rightarrow \mathbb{E}[g(X)] \geqslant g(\mathbb{E}[X]) \text{ (Jensen's Inequality)}$$




Red: $\lambda g(x) + (1-\lambda)g(y)$
Blue: $g(\lambda x + (1-\lambda)y)$
The red - line is always above the blue - line.

**Example 1.4.** $X \sim Binomial(n,p)$, $\mathbb{P}(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$

Method 1:

$$
\begin{aligned}
\mathbb{E}(X) &= \sum_{k=0}^{n} k \cdot \mathbb{P}(X = k) \\
&= \sum_{k=0}^{n} k \cdot C_n^k \cdot p^k \cdot (1-p)^{n-k} \\
&= \sum_{k=1}^{n} n \cdot \binom{n-1}{k-1} \cdot p^k \cdot (1-p)^{n-k} \\
&= np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k} \\
&= np \sum_{k=0}^{n-1} \binom{n-1}{k} p^{k} (1-p)^{n-k-1} \\
&= np
\end{aligned}
$$

Method 2: $X = \sum_{i=1}^{n} X_i$, $X_i \sim B(p)$ (The binomial distribution is the sum of $n$ Bernoulli trials)

(Using property 1) $\mathbb{E}(X_i) = p$

$$
\mathbb{E}(X) = \mathbb{E}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \mathbb{E}(X_i) = np
$$

### 1.4 Variance

**Definition 1.5** (Variance)**.** The variance of a random variable $X$ is defined as

$$
\mathrm{Var}(X) \stackrel{\mathrm{Def}}{=} \mathbb{E}[(X - \mathbb{E}(X))^2]
$$

If $\mathbb{E}(X) = 0$, then $\mathrm{Var}(X) = \mathbb{E}[X^2]$.

The variance is a measure of how much a random variable deviates from its mean.

**Definition 1.6** (Standard Deviation)**.** The standard deviation is defined as $\mathrm{sd}(X) = \sqrt{\mathrm{Var}(X)}$.

### 1.5 Properties of Variance

1. $\mathrm{Var}(X) = \mathbb{E}[X^2] - [\mathbb{E}(X)]^2$
   Proof:

$$
\begin{aligned}
\mathrm{Var}(X) &= \mathbb{E}[X^2 - 2X\mathbb{E}(X) + (\mathbb{E}(X))^2] \\
&= \mathbb{E}[X^2] - 2\mathbb{E}[X \cdot \mathbb{E}(X)] + \mathbb{E}[(\mathbb{E}(X))^2] \\
&= \mathbb{E}[X^2] - 2(\mathbb{E}(X))^2 + (\mathbb{E}(X))^2 \\
&= \mathbb{E}[X^2] - (\mathbb{E}(X))^2
\end{aligned}
$$

2. $\text{Var}(X) \geqslant 0 \Rightarrow \mathbb{E}[X^2] \geqslant (\mathbb{E}(X))^2$
   Proof:
   Using Jensen's Inequality: Let $g(x) = x^2$.
   Then $\mathbb{E}[g(x)] \geqslant g(\mathbb{E}(X))$, so $\mathbb{E}[X^2] \geqslant (\mathbb{E}(X))^2$.
3. $\text{Var}(aX + b) = a^2\text{Var}(X)$
   Proof:

$$\begin{aligned}\text{Var}(aX+b) &= \mathbb{E}[(aX+b) - \mathbb{E}(aX+b)]^2 \\ &= \mathbb{E}[aX + b - a\mathbb{E}(X) - b]^2 \\ &= \mathbb{E}[aX - a\mathbb{E}(X)]^2 \\ &= a^2\mathbb{E}[X - \mathbb{E}(X)]^2 \\ &= a^2\text{Var}(X)\end{aligned}$$

4. Let $X_1, X_2, \cdots, X_n$ be independent random variables.
   Then $\text{Var}(\sum_{i=1}^n a_i X_i) = \sum_{i=1}^n a_i^2 \text{Var}(X_i)$
   Proof:

$$\begin{aligned}\text{Var}(\sum_{i=1}^n a_i X_i) &= \mathbb{E}[\sum_{i=1}^n a_i X_i - \mathbb{E}[\sum_{i=1}^n a_i X_i]]^2 \\ &= \mathbb{E}[\sum_{i=1}^n a_i (X_i - \mathbb{E}(X_i))]^2 \\ &= \mathbb{E}[\sum_{i=1}^n a_i^2 (X_i - \mathbb{E}(X_i))^2 + 2\sum_{1\leqslant i<j\leqslant n} a_i a_j (X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))] \\ &= \sum_{i=1}^n a_i^2 \mathbb{E}(X_i - \mathbb{E}(X_i))^2 + 2\sum_{1\leqslant i<j\leqslant n} a_i a_j \mathbb{E}[(X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))]\end{aligned}$$

   Since $\sum_{i=1}^n a_i^2 \mathbb{E}(X_i - \mathbb{E}(X_i))^2 = \sum_{i=1}^n a_i^2 \text{Var}(X_i)$.
   Next, we prove that $2\sum_{1\leqslant i<j\leqslant n} a_i a_j \mathbb{E}[(X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))] = 0$:
   Since any $X_i$ and $X_j$ are independent of each other, we have $\mathbb{E}[(X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))] = \mathbb{E}[X_i - \mathbb{E}(X_i)] \cdot \mathbb{E}[X_j - \mathbb{E}(X_j)] = 0 \times 0 = 0$.

## 1.6 Sample Mean and Sample Variance

Let $x_1, x_2, \cdots, x_n$ be the observed values.

**Definition 1.7** (Sample Mean)**.** The sample mean is defined as $\bar{x}_n = \frac{1}{n}\sum_{i=1}^n x_i$.

**Definition 1.8** (Sample Variance)**.** The sample variance is defined as $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x}_n)^2$.

**Theorem 1.9.** *Suppose that $\{x_i\}_{i=1}^n$ are independent and identically - distributed random variables with $\mathbb{E}(x_i) = \mu$ and $\text{Var}(x_i) = \sigma^2$. Let $\bar{x}_n$ be the sample mean and $S_n^2$ be the sample variance.*

1. $\mathbb{E}(\bar{x}_n) = \mu$
   Proof:
   $\mathbb{E}(\bar{x}_n) = \mathbb{E}(\frac{1}{n}\sum_{i=1}^{n} x_i) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(x_i) = \frac{1}{n} \cdot n\mu = \mu.$
   Note that this result does not require the random variables to be independent and identically - distributed. It only requires that $\mathbb{E}(x_i)$ are equal for all $i$.
2. $\text{Var}(\bar{x}_n) = \frac{\sigma^2}{n}$
   Proof:
   $\text{Var}(\bar{x}_n) = \text{Var}(\frac{1}{n}\sum_{i=1}^{n} x_i) = \frac{1}{n^2}\sum_{i=1}^{n} \text{Var}(x_i) = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}.$
   As $n \to \infty$, $\text{Var}(\bar{x}_n) \to 0$.
   This implies that when the sample size is large enough, the sample mean is close to the true mean.
3. $\mathbb{E}(S_n^2) = \sigma^2$
   Proof:

$$
\begin{aligned}
\mathbb{E}(S_n^2) \cdot (n-1) &= \mathbb{E}[\sum_{i=1}^{n}(x_i - \bar{x}_n)^2] \\
&= \mathbb{E}[\sum_{i=1}^{n}(x_i^2 - 2\bar{x}_n x_i) + n \cdot \bar{x}_n^2] \\
&= \mathbb{E}[\sum_{i=1}^{n} x_i^2 - 2\bar{x}_n \cdot n\bar{x}_n + n \cdot \bar{x}_n^2] \\
&= \mathbb{E}[\sum_{i=1}^{n} x_i^2 - n \cdot \bar{x}_n^2] \\
&= \sum_{i=1}^{n} \mathbb{E}(x_i^2) - n \cdot \mathbb{E}(\bar{x}_n^2) \\
&= \sum_{i=1}^{n}[(\mathbb{E}(x_i))^2 + \text{Var}(x_i)] - n \cdot [(\mathbb{E}(\bar{x}_n))^2 + \text{Var}(\bar{x}_n)] \\
&= n(\mu^2 + \sigma^2) - n(\mu^2 + \frac{\sigma^2}{n}) \\
&= (n-1)\sigma^2
\end{aligned}
$$

   So, $\mathbb{E}(S_n^2) = \sigma^2$.

### 1.7 Covariance and Correlation

**Definition 1.10** (Covariance)**.** Covariance measures joint variability — the extent of variation between two random variables:

$$\text{Cov}(X, Y) \triangleq \mathbb{E}\left[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\right]. \tag{1}$$

**Definition 1.11** (Correlation)**.** Correlation is a measure of the degree of linear relationship between two variables:

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

### 1.7.1 Properties of Covariance

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

   *Proof.*

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

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

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

   ■

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

   *Proof.*

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

   ■

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

   *Proof.*

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

   ■

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

   *Proof.*

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

   ■

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

   *Proof.*

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

   ■

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

   *Proof.*

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

   ■