Modern Statistics **Lecture 5 - 03/03/2025**

Lecture 5 Expectation (Moment)

Lecturer:Xiangyu Chang *Scribe: Chuan Liu,Zishuo Wang*

Edited by: Zhihong Liu

1 Recall

1.1 Definition of Expectation

Definition 1 (Expectation)**.** *The expected value, or mean, or first moment, of X is defined to be*

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

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

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

Example 1. *The probability density function is given as $f_X(x) = \frac{1}{\pi(1+x^2)}$.*

$$\begin{aligned} \int_{-\infty}^{+\infty} f(x)dx &= \int_{-\infty}^{+\infty} \frac{1}{\pi(1+x^2)} dx \\ &= \frac{1}{\pi} \arctan x \Big|_{-\infty}^{+\infty} \\ &= \frac{1}{\pi} \cdot \pi = 1 \end{aligned}$$

The first possible method to solve for $\mathbb{E}(X)$ of this distribution:

$$\begin{aligned} \mathbb{E}(X) &= \int_{-\infty}^{+\infty} \frac{x}{\pi(1+x^2)} dx \\ &= \lim_{a \to +\infty} \int_{-a}^{a} \frac{x}{\pi(1+x^2)} dx \\ &= \lim_{a \to +\infty} \frac{1}{2\pi} \log(1+x^2) \Big|_{-a}^{a} \\ &= 0 \end{aligned}$$

Similarly,

$$\begin{aligned}
\mathbb{E}(X) &= \int_{-\infty}^{+\infty} \frac{x}{\pi(1+x^2)} dx \\
&= \lim_{a\to+\infty} \int_{-a}^{2a} \frac{x}{\pi(1+x^2)} dx \\
&= \lim_{a\to+\infty} \frac{1}{2\pi} \log(1+x^2)\Big|_{-a}^{a} \\
&= \frac{1}{2\pi} \log 4
\end{aligned}$$

Since the two results are different, obviously the method is incorrect.

The second possible method to solve for $\mathbb{E}(X)$ *of this distribution:*

$$\begin{aligned}
\mathbb{E}(X) &= \int_{-\infty}^{+\infty} \frac{x}{\pi(1+x^2)} dx \\
&= \int_{-\infty}^{0} \frac{x}{\pi(1+x^2)} dx + \int_{0}^{+\infty} \frac{x}{\pi(1+x^2)} dx \\
&= \lim_{a\to+\infty} \int_{-a}^{a} \frac{x}{\pi(1+x^2)} dx
\end{aligned}$$

$$\int_{-\infty}^{0} \frac{x}{\pi(1+x^2)} dx + \int_{0}^{+\infty} \frac{x}{\pi(1+x^2)} dx = \lim_{a\to+\infty} \int_{-a}^{0} \frac{x}{\pi(1+x^2)} dx + \lim_{b\to+\infty} \int_{0}^{b} \frac{x}{\pi(1+x^2)} dx$$

But $\lim\limits_{a\to+\infty} \int_{-a}^{0} \frac{x}{\pi(1+x^2)} dx + \lim\limits_{b\to+\infty} \int_{0}^{b} \frac{x}{\pi(1+x^2)} dx \neq \lim\limits_{a\to+\infty} \int_{-a}^{a} \frac{x}{\pi(1+x^2)} dx.$

Obviously, this method is incorrect. From this, the sufficient conditions for the existence of $\mathbb{E}(X)$ *can be further understanded.*

If $\mathbb{E}(X)$ *exists, then* $\int |x| f(x) dx < +\infty$.

$$\int |x| dF(x) = \frac{2}{\pi} \int_{0}^{\infty} \frac{x\ dx}{1+x^2} = \left[x \tan^{-1}(x)\right]_{0}^{\infty} - \int_{0}^{\infty} \tan^{-1} x\ dx = \infty$$

So the mean does not exist. If you simulate a Cauchy distribution many times and take the average, you will see that the average never settles down. This is because the Cauchy has thick tails and hence extreme observations are common.

2 The k^{th} moment of X

Definition 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. *If* $\mathbb{E}(X^k) < \infty$ *(exist),* $k \geq 1$, $i \leq k$, *then* $\mathbb{E}(X^i) < \infty$ *(exist)*

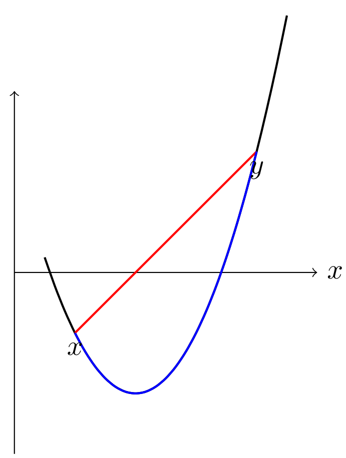
Proof:

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

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 $0 \le \lambda \le 1$, $g(\lambda x + (1-\lambda)y) \le \lambda g(x) + (1-\lambda) g(y)$

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



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

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) $\mathbb{E}(X_i) = p$

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

4 Variance

Definition 3 (Variance)**.** *The variance of a random variable X is defined as*

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

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

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. $\mathrm{Var}(X) \geq 0 \Rightarrow \mathbb{E}[X^2] \geq (\mathbb{E}(X))^2$

   Proof:

   Using Jensen's Inequality: Let $g(x) = x^2$.

   Then $\mathbb{E}[g(x)] \geq g(\mathbb{E}(X))$, so $\mathbb{E}[X^2] \geq (\mathbb{E}(X))^2$.

3. $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$

   Proof:

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

4. Let $X_1, X_2, \cdots, X_n$ be independent random variables.

   Then $\mathrm{Var}(\sum_{i=1}^n a_i X_i) = \sum_{i=1}^n a_i^2 \mathrm{Var}(X_i)$

   Proof:

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

   Since $\sum_{i=1}^n a_i^2 \mathbb{E}(X_i - \mathbb{E}(X_i))^2 = \sum_{i=1}^n a_i^2 \mathrm{Var}(X_i)$.

   Next, we prove that $2\sum_{1 \leq i < j \leq n} a_i a_j \mathbb{E}[(X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))] = 0$:

   Since any X_i and X_j are independent of each other, we have $\mathbb{E}[(X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))] = \mathbb{E}[X_i - \mathbb{E}(X_i)] \cdot \mathbb{E}[X_j - \mathbb{E}(X_j)] = 0 \times 0 = 0$.

6 Sample Mean and Sample Variance

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

Definition 5 (Sample Mean)**.** *The sample mean is defined as* $\bar{x}_n = \frac{1}{n}\sum_{i=1}^n x_i$.

Definition 6 (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 2. *Suppose that $\{x_i\}_{i=1}^n$ are independent and identically - distributed random variables with $\mathbb{E}(x_i) = \mu$ and $\mathrm{Var}(x_i) = \sigma^2$. 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. $\mathrm{Var}(\bar{x}_n) = \frac{\sigma^2}{n}$

   Proof:

   $\mathrm{Var}(\bar{x}_n) = \mathrm{Var}(\frac{1}{n}\sum_{i=1}^n x_i) = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}(x_i) = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}$.

   As $n \to \infty$, $\mathrm{Var}(\bar{x}_n) \to 0$.

   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:

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

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