Moments

Ye Wang
University of North Carolina at Chapel Hill

Mathematics and Statistics For Political Research
POLI783

Expectation

- In many applications, we care less about the full distribution of a random variable and more about certain "summary statistics."
- E.g., What is the average income of Americans? Does a gasoline tax improve the air quality on average?
- The average is often the first number reported in an empirical paper—the mean of each variable, the average treatment effect (ATE), etc.
- For a random variable, this "average" is captured by its expectation.
- If X is a discrete random variable, then its expectation can be defined as

$$\mathbb{E}[X] = \sum_{x \in \{x_1, x_2, \dots\}} x p_X(x).$$

- This is the weighted average of all possible values of X, where the weights are given by the p.m.f.

Expectation

- X's expectation can be seen as a mapping from its distribution to a real number.
- It is a function of the distribution function (p.m.f.) and known as a "functional."
- Therefore, if two variables have the same distribution, they also have the same expectation.
- But the converse is not true (expectation is not an injective).
- If $X \sim Bern(p)$, then $\mathbb{E}[X] = p * 1 + (1 - p) * 0 = p$.

Example

- We randomly assign 3 units into treatment with the probability of 0.5.
- What is the expected size of the treatment group?
- Remember that

$$X = \begin{cases} 0, & \text{if } \{T, T, T\}, p = \frac{1}{8}, \\ 1, & \text{if } \{T, T, H\}, \{H, T, T\}, \{T, H, T\}, p = \frac{3}{8}, \\ 2, & \text{if } \{T, H, H\}, \{H, T, H\}, \{H, H, T\}, p = \frac{3}{8} \\ 3, & \text{if } \{H, H, H\}, p = \frac{1}{8}. \end{cases}$$

- Then, $\mathbb{E}[X] = 0 * \frac{1}{8} + 1 * \frac{3}{8} + 2 * \frac{3}{8} + 3 * \frac{1}{8} = \frac{3}{2}$.

Properties of expectation

- Linearity:
$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y].$$
- This property holds even when X and Y are dependent!
- Jensen's inequality: If $g(\cdot)$ is concave,
$$g(\mathbb{E}[X]) \geq \mathbb{E}[g(X)],$$
with equality only when $g(\cdot)$ is linear.
- Happiness is often concave in income.
- $g(\mathbb{E}[X])$: happiness of an individual with the average income; $\mathbb{E}[g(X)]$: average happiness in the society.
- Jensen's inequality: With diminishing marginal happiness from income, the average-income person is happier than the average member of the society.
- $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ if X and Y are independent.

Expectation of a binomial distribution

- Remember that if $X_i \sim Bern(p)$, then

$$Y = \sum_{i=1}^{n} X_i \sim Bin(n, p).$$

- Using the linearity of expectation, we know that if $Y \sim Bin(n, p)$,

$$\mathbb{E}[Y] = \sum_{i=1}^{n} \mathbb{E}[X_i] = np.$$

Sample average

- Consider a collection of i.i.d. r.v.s, $(X_1, X_2, \ldots, X_n)$, with $\mathbb{E}[X_i] = \mu$.
- The sample average is defined as

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

- $\bar{X}$ is also a r.v. with the expectation

$$\mathbb{E}\left[\bar{X}\right] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[X_i] = \frac{1}{n} \sum_{i=1}^{n} \mu = \mu.$$

- This property of $\bar{X}$ is known as the unbiasedness.
- What if $\mathbb{E}[X_i] = \mu_i$?
- This is known as independent but not identically distributed (i.n.i.d.).

Expectation of continuous random variables

- The expectation of a continuous r.v. X equals

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx = \int_{-\infty}^{\infty} x dF_X(x).$$

- Again, the definite integral is a generalization of summation.
- Assume that $U \sim Unif(a, b)$, then

$$\mathbb{E}[U] = \int_a^b x \frac{1}{b-a} dx = \frac{1}{2(b-a)}(b^2 - a^2) = \frac{b+a}{2}.$$

- The same properties hold for the expectation of continuous r.v.s.

Fundamental bridge

- We can show the following relationship between probability and expectation:

$$\mathbb{P}(A) = \mathbb{E}[\mathbf{1}\{A\}].$$

- The probability of a random event is the expectation for a r.v. indicating its occurrence.
- This is known as the fundamental bridge between probability and expectation.
- We can derive the result from the definition of expectation:

$$\begin{aligned}\mathbb{E}[\mathbf{1}\{A\}] &= \int_{-\infty}^{\infty} \mathbf{1}\{x \in A\} f_X(x) dx \\ &= \int_{\mathbf{1}\{x \in A\}} f_X(x) dx = \mathbb{P}(A).\end{aligned}$$

Expectation of transformed r.v.s

- We can also compute the expectation for any transformation of a r.v.
- Let $Y = g(X)$, then by its definition

$$\mathbb{E}[Y] = \mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx.$$

- This is sometimes known as the "law of the unconscious statistician" (LOTUS).
- What an unconscious statistician would do if they forget to derive Y's distribution first.
- For a discrete r.v., it becomes

$$\mathbb{E}[Y] = \mathbb{E}[g(X)] = \sum_{x \in \mathcal{X}} g(x) p_X(x).$$

Exercises

- Consider the discrete r.v. X we have been using, how do we compute $\mathbb{E}\left[X^2\right]$?
- Applying the LOTUS, we know that

$$\begin{aligned}\mathbb{E}\left[X^2\right] =& 0 * \frac{1}{8} + 1^2 * \frac{3}{8} + 2^2 * \frac{3}{8} + 3^2 * \frac{1}{8} \\ =& \frac{3}{8} + \frac{12}{8} + \frac{9}{8} = 3.\end{aligned}$$

Exercises

- Suppose $R \sim Unif(0,1)$ and $A = \pi R^2$, compute $\mathbb{E}[A]$ and $\mathbb{E}\left[A^2\right]$:

$$\begin{aligned}\mathbb{E}[A] =& \mathbb{E}[\pi R^2] = \int_0^1 \pi r^2 f_R(r) dr = \pi \int_0^1 r^2 dr \\ =& \frac{\pi}{3} r^3 |_0^1 = \frac{\pi}{3}.\end{aligned}$$

- Similarly,

$$\begin{aligned}\mathbb{E}\left[A^2\right] =& \mathbb{E}\left[\pi^2 R^4\right] = \int_0^1 \pi^2 r^4 f_R(r) dr = \pi^2 \int_0^1 r^4 dr \\ =& \frac{\pi^2}{5} r^5 |_0^1 = \frac{\pi^2}{5}.\end{aligned}$$

Moments

- The expectation is one of the "moments" of a r.v.
- A moment is a real number generated from the r.v.'s distribution (i.e., a functional).
- The kth moment and the kth central moment of a r.v. X are defined as

$$\mu_k = \mathbb{E}\left[X^k\right], \text{ and } \tilde{\mu}_k = \mathbb{E}\left[(X - \mathbb{E}[X])^k\right],$$

  respectively.
- The first moment of X is $\mu = \mathbb{E}[X]$, X's expectation.
- It captures the concentration of the variable.
- The second centralized moment is known as the r.v.'s variance.
- It describes the variation of the r.v. around its expectation.

Variance

- From its definition, we can see that

$$
\begin{aligned}
\mathrm{Var}[X] =& \int_{-\infty}^{\infty} (x - \mathbb{E}[X])^2 f_X(x)dx \\
=& \int_{-\infty}^{\infty} \left(x^2 - 2x\mathbb{E}[X] + \mathbb{E}^2[X]\right) f_X(x)dx \\
=& \int_{-\infty}^{\infty} x^2 f_X(x)dx - 2\mathbb{E}^2[X] + \mathbb{E}^2[X] \\
=& \mathbb{E}\left[X^2\right] - \mathbb{E}^2[X] \leq \mathbb{E}\left[X^2\right].
\end{aligned}
$$

- To compute $\mathrm{Var}[X]$, we need to know $\mathbb{E}[X]$ and $\mathbb{E}\left[X^2\right]$.
- E.g., $\mathbb{E}\left[X^2\right] = \mathbb{E}[X] = p$ if $X \sim Bern(p)$, thus $\mathrm{Var}[X] = p - p^2 = p(1-p)$.
- $\sqrt{\mathrm{Var}[X]}$ is known as the standard deviation of X.
- Sometimes, we denote the standard deviation as σ and the variance as σ^2.

Variance

- From its definition, we know that $\mathrm{Var}[X] \geq 0$.
- $\mathrm{Var}[X] = 0$ if X is a constant.
- $\mathrm{Var}[aX + b] = a^2 \,\mathrm{Var}[X]$.
- If X and Y are independent, $\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$.
- For $Y \sim Bin(n, p)$, we know that

$$\mathrm{Var}[Y] = \sum_{i=1}^{n} \mathrm{Var}[X_i] = \sum_{i=1}^{n} p(1-p) = np(1-p).$$

- For the sample average $\bar{X}$ with $\mathrm{Var}[X_i] = \sigma^2$, we can see that

$$\mathrm{Var}[\bar{X}] = \frac{1}{n^2} \sum_{i=1}^{n} \mathrm{Var}[X_i] = \frac{1}{n^2} \sum_{i=1}^{n} \sigma^2 = \frac{\sigma^2}{n}.$$

General normal distribution

- If $Z \sim \mathcal{N}(0,1)$, then $X = \mu + \sigma Z$ follows normal distribution with mean μ and variance σ^2.
- The p.d.f. of X is

$$\phi_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

- Conversely, $\frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$ (standardization).

General normal distribution

- If $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$, and $X_1 \perp\!\!\!\perp X_2$, then
$$X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$
- For the sample average where $X_i \sim \mathcal{N}(\mu, \sigma^2)$, we know that $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$.
- Therefore, $\frac{\bar{X}-\mu}{\sqrt{\sigma^2/n}} \sim \mathcal{N}(0, 1)$.
- Cramer's theorem: if $X_1 \perp\!\!\!\perp X_2$ and $X_1 + X_2$ is normal, then both X_1 and X_2 are normal.
- Not true for other distributions!

Chi-squared distribution

- Let $V = Z_1^2 + Z_2^2 + \cdots + Z_n^2$, where $Z_1, Z_2, \ldots, Z_n$ are i.i.d. $\mathcal{N}(0,1)$.
- V obeys the Chi-square distribution with n degrees of freedom:

$$V \sim \chi_n^2.$$

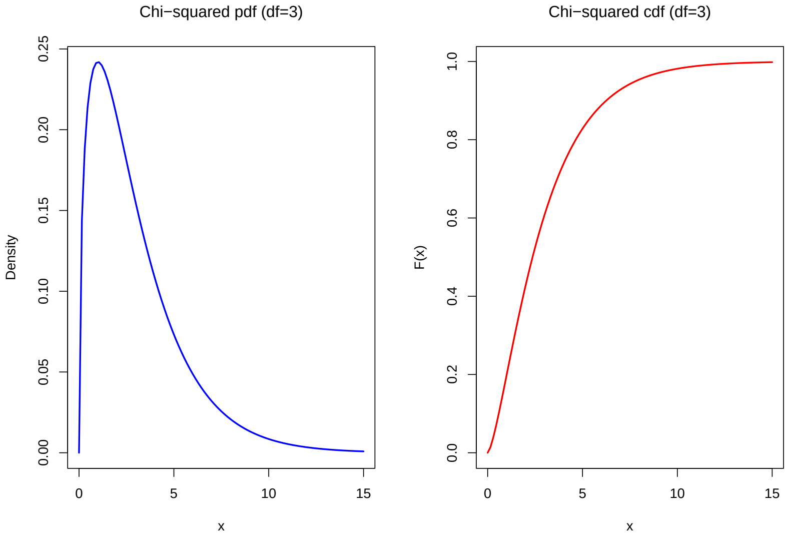
- $\mathbb{E}[V] = n$ and $\mathrm{Var}[V] = 2n$.
- For an i.i.d. sample, $\{X_i\}_{i=1}^n$, the sample variance is defined as

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X} \right)^2 .$$

- If $X_i \sim \mathcal{N}(0,1)$, then

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}^2.$$

Chi-squared distribution



Student-t distribution

- If $Z \sim \mathcal{N}(0,1)$, $V \sim \chi^2_n$, and $Z \perp\!\!\!\perp V$, then

$$T = \frac{Z}{\sqrt{V/n}}$$

follows the student-t distribution with n degrees of freedom:

$$T \sim t_n.$$

- Consider the studentized sample average:

$$T = \frac{\bar{X} - \mu}{\sqrt{s^2/n}} = \frac{(\bar{X} - \mu)/\sqrt{\sigma^2/n}}{\sqrt{s^2/n}/\sqrt{\sigma^2/n}} = \frac{(\bar{X} - \mu)/\sqrt{\sigma^2/n}}{\sqrt{(n-1)s^2/(\sigma^2(n-1))}}.$$

- The numerator obeys $\mathcal{N}(0,1)$ and the denominator obeys χ^2_{n-1}, hence $T \sim t_{n-1}$.

Student-t distribution