

Central Limit Theorem

Lecturer: Jie Fu

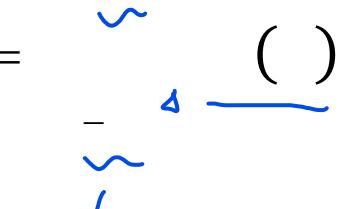
Outlines

- Functions of random variables.
- Expectation, moment, variance of a random variable:
 - Can be defined for both continuous and discrete RV.
- Important property of variance.
- Central Limit Theorem
- Applications of C.L.T.

Expectation

- The expectation of a continuous random variable is defined by

$$[] = \int_{-\infty}^{\infty} x f(x) dx$$



- The expectation of a discrete random variable is defined by

$$[] = \sum_k k p(k)$$



- As for discrete random variables, the expectation can be interpreted as
 - "center of gravity" of the PDF
 - anticipated average value of in a large number of independent repetitions of the experiment.

Example

$$\lim_{x \rightarrow +\infty} x e^{-\lambda x} = \lim_{x \rightarrow +\infty} \frac{x}{e^{\lambda x}} = \lim_{x \rightarrow +\infty} \frac{1}{e^{\lambda x}} = 0$$

1. Let X be the outcome of rolling a **fair 6-sided die**. The probability mass function (PMF) is:

$$P(X=k) = \frac{1}{6} \quad \text{for } k=1, 2, 3, 4, 5, 6$$

The expectation is calculated as:

$$E(X) = \sum_{k=1}^6 k \cdot P(X=k) = \frac{1}{6} (1+2+3+4+5+6) = \frac{3 \times 7}{6} = 3.5$$

2. Let X be an exponential distribution with pdf.

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0, \lambda > 0$$

The expectation is calculated as:

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} x \cdot \lambda e^{-\lambda x} dx = xe^{-\lambda x} \Big|_0^{+\infty} + \int_0^{+\infty} e^{-\lambda x} dx \\ &= 0 + \left(-\frac{1}{\lambda}\right) e^{-\lambda x} \Big|_0^{+\infty} = \frac{1}{\lambda} \end{aligned}$$

$$\boxed{\begin{aligned} \int u dv &= uv - \int v du \\ u &= x \\ dv &= \lambda e^{-\lambda x} dx \end{aligned}}$$

"Average wait time"

Function of random variable

$$g(x) = x^2 + 2x + 3$$

- For any real-valued function (\cdot) , (\cdot) is also a random variable.

$$X \sim N(0, 1)$$

$$Y = ax + b \sim N(a \cdot 0 + b, a^2)$$

- The expectation of (\cdot) is

$$E(\cdot) = \int_{-\infty}^{\infty} (\cdot) \underbrace{f(x)}_{\sim} dx$$

$$E(g(X)) = \int_{-\infty}^{\infty} (x^2 + 2x + 3) \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

Example

life span of a bulb

$$E(X) = \frac{1}{\lambda} = 5 \Rightarrow \lambda = \frac{1}{5}$$

- A company manufactures LED lightbulbs, and the lifespan (in years) of each bulb follows an Exponential distribution with average life space of 5 years. The company offers a warranty where if a lightbulb fails within 3 years, it is replaced for free. The replacement cost per bulb is \$10.
- What is the expected cost the company need to pay per light bulb under the warranty replacement?

Y : cost

X : lifespan.

$$Y = \begin{cases} 0 & \text{if } X \leq 3 \\ 10 & \text{o.w.} \end{cases}$$

$$E(Y) = \int_{-\infty}^{+\infty} 10 \cdot 1(X \leq 3) \cdot \underbrace{\left(\frac{1}{5}\right)}_{g(x)} e^{-\frac{1}{5}x} dx$$

PDF

$$\begin{aligned} &= 10 \times (1 - SF(3)) \\ &= 10 \times (1 - e^{-\lambda \cdot 3}) \\ &= 10 \cdot (1 - e^{-\frac{3}{5}}) \\ &\approx 9.45 \end{aligned}$$

$$\begin{aligned} &dx = \int_0^3 10 \cdot \left(\frac{1}{5}\right) e^{-\frac{1}{5}x} dx \\ &= 10 \cdot \left[\int_0^3 \frac{1}{5} e^{-\frac{1}{5}x} dx \right] = 10 \cdot F_X(3) \end{aligned}$$

Moments and variance

integer ≥ 1

- The n th moment of X is defined by $[]$.

- The Δ variance of X is defined by $\text{Constant} = []$.

1st moment $E(X)$

$$\int_{-\infty}^{\infty} x^n \cdot f_X(x) dx.$$

$$[] = \Delta [(- [])^2]$$

$$= \Delta \underline{[(- [])^2]} ()$$

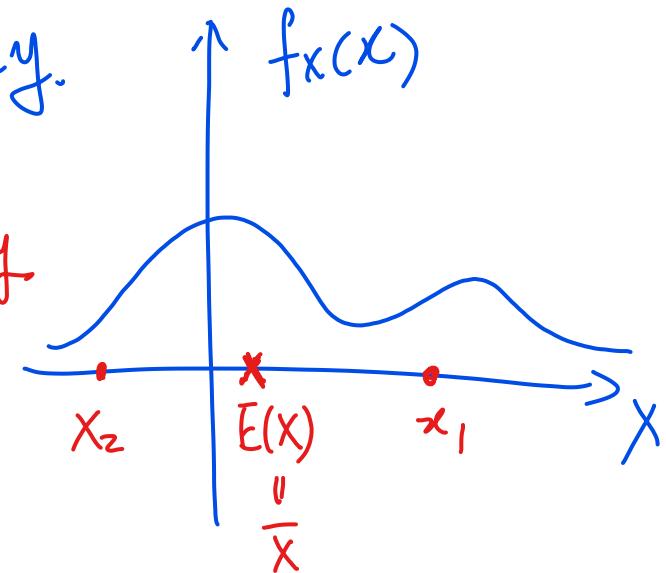
"Center of gravity."

"Average" squared distance to the center of gravity

- Please verify the equality.

discrete

$$\text{Var}(X) = \sum_k (x_k - E(X))^2 P(X=x_k)$$



Property of variance

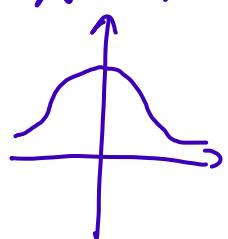
property: $E(X_1 + X_2) = E(X_1) + E(X_2)$
 $E(aX) = aE(X)$

- 0 $[] = []^2 - ([])^2$

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

- If $=$, then

$$X \sim N(0, 1)$$

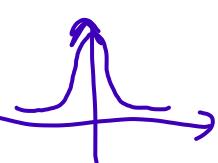


$$\text{Var}(Y) = \text{Var}[ax] = E((ax - E(ax))^2) = E(a^2X^2 - 2ax \cdot E(ax) + (E(ax))^2)$$

$$[] = [].$$

$$[] = \underbrace{a^2}_{\text{underbrace}} [].$$

$$Y = \frac{1}{3}X$$



$$Y = -\frac{1}{3}X$$

$$\begin{aligned} &= E(a^2X^2) - 2aE(X) \cdot E(ax) + (E(ax))^2 \\ &= a^2E(X^2) - 2a^2(E(X))^2 + (aE(X))^2 = a^2E(X^2) - a^2(E(X))^2 \end{aligned}$$

$\Rightarrow a^2 \text{Var}(X)$

- If $Y = aX + b$, then

$$[Y] = \underline{[a]} [X] + \underline{b}, \quad [Y^2] = \underline{\underline{[a^2 X^2]}}.$$

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

- if $Y = aX + b$, then

$$\begin{aligned} \text{Var}(Y) &= \text{Var}(\underbrace{aX}_\sim + b) \\ &= \text{Var}(aX) = a^2 \text{Var}(X) \end{aligned}$$

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

- If $= X_1 + X_2$ for two independent RV, then

$$[] = E(X_1) + E(X_2)$$

$$[] = \text{Var}(X_1) + \text{Var}(X_2)$$

for all $B_1 \subseteq \mathcal{X}_1$, $B_2 \subseteq \mathcal{X}_2$
 $P(X_1 \in B_1 \text{ and } X_2 \in B_2) = P(X_1 \in B_1) \cdot P(X_2 \in B_2)$

outline

- Expectation, moment, variance of a random variable:
 - Can be defined for both continuous and discrete RV.
- Important property of variance.
- Next: Central limit theorem.

Motivating problem:

$$E(X_1 + X_2) = E(X_1) + E(X_2)$$

- A machine processes parts, one at a time, in a time independently and uniformly distributed in [1,5].
- What is the probability the machine processes at least 100 parts in 320 time units?

time taken to process one part:

$$X \sim \text{uniform}(1, 5)$$

$$f_X(x) = \begin{cases} \frac{1}{4} & 1 \leq x \leq 5 \\ 0 & x < 1 \text{ or } x > 5 \end{cases}$$

X_i : time it takes to process i th part,

$$\text{Total time: } \sum_{i=1}^{100} X_i = S_{100}$$

$$\text{"Average time"} \cdot E(S_{100}) = E\left(\sum_{i=1}^{100} X_i\right) = \sum_{i=1}^{100} E(X_i) = 300.$$

$$P(S_{100} \leq 320)$$

$$E(X_i) = \frac{1+5}{2} = 3$$

$$\text{Var}(X_i) = \frac{(b-a)^2}{12} = \frac{4^2}{12} = \frac{4}{3}$$

$$i = 1, \dots, 100$$

$$\text{Var}(S_{100}) = \text{Var}\left(\sum_{i=1}^{100} X_i\right)$$

$$= 300 \times \frac{4}{3} = 400$$

- Let X_1, X_2, \dots be a sequence of independent identically distributed random variables with mean μ and variance σ^2
- Let $S_n = X_1 + X_2 + \dots + X_n$, what is the mean of S_n ? What is the variance of S_n ?

$\Delta \quad n=100 \quad m=200 \quad . \quad 10000$

$$E(S_n) = \sum_{i=1}^n E(X_i) = n \cdot \mu$$

$$\text{Var}(S_n) = \sum_{i=1}^n \text{Var}(X_i) = n \cdot \sigma^2$$

Background

The distribution of S_n spreads out as n increases.

But the situation is different if we consider the **sample mean**

$$\text{Sample Mean} = \bar{M}_n = \frac{1 + 2 + \dots + n}{n} = \frac{S_n}{n}.$$

The sample mean is itself a RV (why?) so we can compute its mean and variance

$$E(M_n) = E\left(\frac{S_n}{n}\right) = E\left(\frac{1}{n} \cdot S_n\right) = \frac{1}{n} E(S_n) = \frac{n\mu}{n} = \mu$$

$$\text{Var}(M_n) = \text{Var}\left(\frac{S_n}{n}\right) = \text{Var}\left(\frac{1}{n} \cdot S_n\right) = \left(\frac{1}{n}\right)^2 \cdot \text{Var}(S_n) = \frac{n\sigma^2}{n^2}$$

$$E(ax) = aE(x)$$

$$\text{Var}(ax) = a^2 \text{Var}(x)$$

$$\lim_{n \rightarrow \infty} \text{Var}(M_n) = 0.$$

$$= \frac{\sigma^2}{n}.$$

$$E(S_n) = n\mu$$

$$\text{Var}(S_n) = n\sigma^2$$

Background

- Given our calculation:

$$[] = , \text{ var}() = -^2.$$

- The variance of decreases to zero as increases.
- Thus, the bulk of the distribution of must be very close to the mean as increases.

Background

$$\mathbb{E}(a+X) = a + \mathbb{E}(X)$$

- We will also consider a quantity which is intermediate between μ and σ .
 - It is defined as follows. Recall $\mathbb{E}(S_n) = n\mu$
1. subtract μ from S_n , to obtain the zero-mean random variable $S_n - \mu$
 2. then divide by \sqrt{n} , to form the random variable.

$$= \frac{S_n - \mu}{\sqrt{n}}$$

$$\bullet \quad \mathbb{E}(Z_n) = \mathbb{E}\left(\frac{S_n - \mu}{\sqrt{n}}\right) = \frac{1}{\sqrt{n}} \mathbb{E}(S_n - \mu) = \frac{1}{\sqrt{n}} (\mathbb{E}(S_n) - \mu) = 0$$

$$\bullet \quad \text{Var}(Z_n) = \text{Var}\left(\frac{S_n - \mu}{\sqrt{n}}\right) = \frac{1}{n} \text{Var}(S_n - \mu) = \frac{\text{Var}(S_n)}{n} = \frac{n\sigma^2}{n} = \sigma^2$$

Formally

- Let X_1, X_2, \dots be a sequence of independent identically distributed random variable with mean μ and variance σ^2
- Define

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}}.$$

- Let's use jupyter notebook

The Central Limit Theorem

- *Theorem (The Central Limit Theorem)* The CDF of $\frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}}$ converges to standard normal CDF

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-x^2/2} dx$$

in the sense that

$$\lim P(\quad) = \underbrace{\Phi}_{\text{CDF of } N(0, 1)}(\quad)$$

Generality

- The central limit theorem is surprisingly general.
- Besides **independence**, and the implicit assumption that the **mean and variance are finite**, it places **no other requirement** on the distribution of the ,
 - which could be discrete, continuous, or mixed.

Going back to our example

Example 2

- A call center receives customer calls according to an exponential distribution with a mean wait time of 4 minutes.

Questions:

- 1. If a single customer calls, what is the probability that they wait more than 5 minutes?
- 2. If we take a random sample of 40 customers, what is the probability that their average wait time is more than 5 minutes?

Advanced thinking: Polling

- p : fraction of population that will vote "yes" in a referendum
- i-th random people polled: 1 : yes, 0: no
- Let M_n be the random variable.
- $M_n = \frac{1^+ + 2^+ + \dots +}{n}$ the fraction of "yes" in our sample.
- We would like small error:

$$|M_n - p| \leq 0.01$$

- How many samples to generate so that the probability of error greater than 0.01 is smaller than 0.05?