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

$$[\;\;] = \;{}_{-\infty}^{\infty} \qquad (\;\;)$$

- The expectation of a discrete random variable   is defined by

$$[\;\;] = \sigma_k \qquad (\;\; = \;\;)$$

- 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

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

1. Let X be the outcome of rolling a **fair 6-sided die**. The probability mass function (PMF) is: $P(X=k) = \frac{1}{6}$ for $k = 1, 2, 3, 4, 5, 6$

The expectation is calculated as:

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

2. Let X be an exponential distribution with pdf.

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

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

The expectation is calculated as:

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x \cdot \underbrace{\lambda e^{-\lambda x}} dx = x e^{-\lambda x}\Big|_0^{+\infty} + \int_0^{+\infty} e^{-\lambda x} dx$$

$$= 0 + \left(-\frac{1}{\lambda}\right) e^{-\lambda x}\Big|_0^{+\infty} = \frac{1}{\lambda}$$

"Average wait time"

Function of random variable

$g(x) = x^2 + 2x + 5$

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

$$\boxed{X \sim N(0, 1)}$$

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

- The expectation of $(\ )$ is

$$[\ (\ )] = \int_{-\infty}^{\infty} (\ ) \underbrace{\ (\ )\ }\ .$$

$$E(g(X)) = \int_{-\infty}^{\infty} (x^2 + 2x + 5) \cdot \frac{1}{2\pi} e^{-\frac{x^2}{1}} dx$$

Example

$\mathbb{E}(X) = \frac{1}{\lambda} = 5 \Rightarrow \lambda = \frac{1}{5}$ ← life span of a bulb

- 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 $\qquad X$: lifespan.

$$Y = \begin{cases} 10 & \text{if } X \le 3 \\ 0 & \text{o.w.} \end{cases} \qquad Y = \underbrace{10 \cdot \mathbb{1}(X \le 3)}$$

$$\mathbb{E}(Y) = \int_{-\infty}^{+\infty} \underbrace{10 \cdot \mathbb{1}(X \le 3)}_{g(x)} \cdot \underbrace{\left(\tfrac{1}{5}\right) e^{-\frac{1}{5}x}}_{\text{PDF}} dx = \int_0^3 10 \cdot \left(\tfrac{1}{5}\right) e^{-\frac{1}{5}x} dx$$

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

$$= 10 \times (1 - SF(3))$$

$$= 10 \times (1 - e^{-\lambda \cdot 3})$$

$$= 10 \cdot (1 - e^{-\frac{3}{5}})$$

$$\doteq 9.45$$

Moments and variance

integer ≥ 1

1st moment $\mathbb{E}(X)$

- The th moment of is defined by [ ].

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

- The variance of is defined by

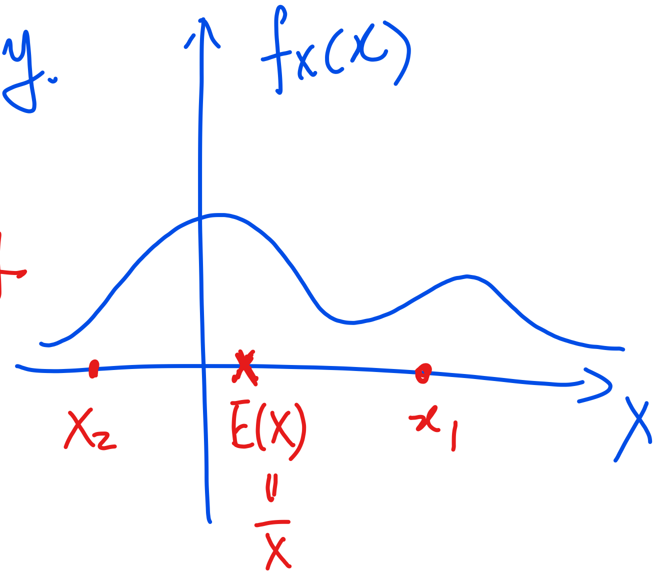
constant

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

"center of gravity."

$$= \int_{-\infty}^{\infty} (\ - \ [\ ])^2 \ (\ )$$

"Average" squared distance to the center of gravity



- Please verify the equality.

discrete

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

Property of variance

property: $\mathbb{E}(X_1 + X_2) = \mathbb{E}(X_1) + \mathbb{E}(X_2)$, $\mathbb{E}(aX) = a\,\mathbb{E}(X)$ exercise

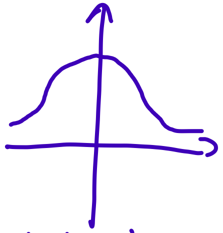
- 0 ≤ [ ] = [²] − ([ ])²

$$Var(X) = \mathbb{E}\left([X - \mathbb{E}(X)]^2\right) = \mathbb{E}\left(X^2 - 2X\cdot\mathbb{E}(X) + (\mathbb{E}(X))^2\right) = \mathbb{E}(X^2) - 2\,\mathbb{E}(X\cdot \mathbb{E}(X)) + \mathbb{E}\left(\mathbb{E}(X)^2\right)$$

$$= \mathbb{E}(X^2) - 2\,\mathbb{E}(X)\cdot\mathbb{E}(X) + (\mathbb{E}(X))^2 = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$$

- If = , then [ ] = [ ], [ ] = ² [ ].

$X \sim N(0,1)$

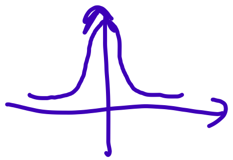


$$Var(Y) = var[aX] = \mathbb{E}\left((aX - \mathbb{E}(aX))^2\right] = \mathbb{E}\left[a^2X^2 - 2aX\cdot\mathbb{E}(aX) + (\mathbb{E}(aX))^2\right]$$

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

$$= \mathbb{E}(a^2X^2) - 2a\,\mathbb{E}(X)\cdot\mathbb{E}(aX) + (\mathbb{E}(aX))^2$$

$$= a^2\,\mathbb{E}(X^2) - 2a^2(\mathbb{E}(X))^2 + (a\,\mathbb{E}(X))^2 = a^2\,\mathbb{E}(X^2) - a^2(\mathbb{E}(X))^2 = a^2 Var(X)$$



- If = + , then

[ ] = [ ] + , [ ] = [ ].

$$Var(Y) = \mathbb{E}\left(\left[(X+b) - \mathbb{E}(X+b)\right]^2\right)$$

$$= \mathbb{E}\left[(X + b - \mathbb{E}(X) - b)^2\right]$$

$$= \mathbb{E}(X - \mathbb{E}(X))^2 = Var(X)$$

- if = + , then

$$Var(Y) = var(aX + b)$$

$$= Var(aX) = a^2 Var(X)$$

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

[ ] = $a^2 Var(X)$

- If $\quad = \quad {}_1 + \quad {}_2$ for two independent RV, then

$[\ ] =$ $\mathbb{E}(X_1) + \mathbb{E}(X_2)$

$[\ ] =$ $Var(X_1) + Var(X_2)$

for all $B_1 \subseteq \mathcal{X}_1$, $B_2 \subseteq \mathcal{X}_2$

$P(X_1 \in B_1$ 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:

$\mathbb{E}(X_1+X_2) = \mathbb{E}(X_1) + \mathbb{E}(X_2)$

- A machine process 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?

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

time taken to process one part:

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

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

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

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

X_i: time it takes to process ith part, $i = 1, \dots, 100$

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

$Var(S_{100}) = Var\left(\sum_{i=1}^{100} X_i\right) = 300 \times \frac{4}{3} = 400$

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

- Let $_1$, , be a sequence of independent identically distributed random variable with mean and variance ^{2}
- Let = $X_1 + X_2 +$ $+ X_n$, what is the mean of ? What is the variance of ?

$n=100$ $n=200$. 10000 ,

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

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

Background

The distribution of spreads out as increases.

$E(S_n) = n\mu$

$Var(S_n) = n\sigma^2$

But the situation is different if we consider the sample mean

$\star \quad = \frac{1 + \quad +}{\quad} = \frac{}{}.$

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

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

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

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

$E(aX) = aE(X)$

$Var(aX) = a^2 Var(X)$

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

Background

- Given our calculation:

$$[\quad] = \quad, \ \mathrm{var}(\quad) = \frac{\ ^2}{}.$$

- 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 and .
- is defined as follows. recall $\mathbb{E}(S_n) = n\mu$

1. subtract from , to obtain the zero-mean random variable −
2. then divide by √ , to form the random variable.

$$= \frac{-}{\sqrt{}}$$

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

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

Formally

- Let ${}_1$, , be a sequence of independent identically distributed random variable with mean and variance
  2

- Define

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

- Let’s use jupyter notebook

The Central Limit Theorem

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

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

in the sense that

$$\lim_{\to\infty} \mathrm{P}(\quad \le \quad) = \Phi(\quad)$$

CDF of N(0,1)

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

$P(\underline{S_{100} < 320})$

$S_{100} \sim N(100 \cdot \mu, 100 \cdot \sigma^2)$

$Z_{100} \sim N(0, 1)$

$$\boxed{Z_{100} = \frac{S_{100} - n\mu}{\sigma\sqrt{n}}}$$

$$P(S_{100} < 320) = P\left(\frac{S_{100} - n\mu}{\sigma\sqrt{n}} < \underbrace{\frac{320 - n\mu}{\sigma\sqrt{n}}}_{\text{constant.}}\right)$$

CDF

Example 2

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

Average time wait: $\mu = \frac{1}{\lambda}$

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?

1. $X_i \sim \exp(\lambda = \frac{1}{4})$

$P(X_i \geq 5)$

2. X_i : ith customer wait time,

$X_1, X_2, \ldots, X_{40}$

$S_n = S_{40} = \sum_{i=1}^{40} X_i$

$\frac{S_n}{n} = M_n = \frac{1}{40}\left(\sum_{i=1}^{40} X_i\right)$

$$\boxed{P(\underbrace{M_{40}}_{RV} \geq 5) = P(\qquad\qquad) \qquad M_n \xrightarrow{n\to+\infty} N\left(\mu, \frac{\sigma^2}{n}\right)}$$

$$P(M_{40} \geq 5) = P\left(\frac{M_{40} - \mu}{\underbrace{\sqrt{\sigma^2/n}}} \geq \frac{5-\mu}{\sigma/\sqrt{n}}\right)$$

Standardization. $\left(N(\mu, \sigma^2) \right.$

$\downarrow$

$N(0,1)$

Advanced thinking: Polling

- p: fraction of population that will vote "yes" in a referendum
- i-th random people polled: 1 : yes, 0: no

X_i : vote for ith person

$\sim$ Bernoulli (p)

- Let X_i be the random variable.
- $M_n = \frac{X_1 + X_2 + \ldots + X_n}{n}$ the fraction of "yes" in our sample.
- We would like small error:

R.V. "Average"

$$P\left(|M_n - p| \leq 0.01 \right) \geq 0.95 \quad \text{for all } p.$$

Constant

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

$M_n \sim N\left(\mu, \frac{\sigma^2}{n} \right)$ n ?

Bernoulli · X_i , mean : p

variance p(1-p)