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 X is defined by

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

- The expectation of a discrete random variable X is defined by

$$\mathbf{E}[X] = \sum_{\mathrm{k}} x_k \, P(X = x_k)$$

- As for discrete random variables, the expectation can be interpreted as
  - "center of gravity" of the PDF
  - anticipated average value of X 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:

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

The expectation is calculated as:

$$\mathbb{E}(X) = \sum_{k=1}^{6} k \cdot P(X=k) = \frac{1}{6}(1+2+3+4+5+6) = \frac{\cancel{3}\times 7}{\cancel{6}_2} = 3.5$$

2. Let X be an exponential distribution with pdf.

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

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

The expectation is calculated as:

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

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

"Average wait time"

Function of random variable

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

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

$X \sim N(0,1)$ $\quad Y = aX + b \sim N(a\cdot 0 + b, a^2)$

- The expectation of $g(X)$ is

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

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

Example

life span of a bulb

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

- 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 $\qquad X$: 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

- The nth moment of X is defined by $\mathbf{E}[X^n]$.

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

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

- The variance of X is defined by

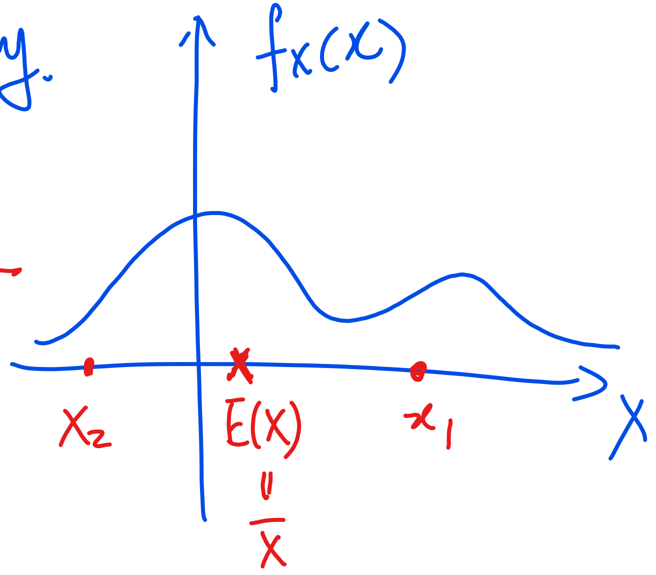
constant

$$\mathbf{Var}[X] = \mathbf{E}[(X - \mathbf{E}[X])^2]$$

"center of gravity."

$$= \int_{-\infty}^{\infty} (x - \mathbf{E}[X])^2 f_X(x)\,dx$$

"Averge" squared distance to the center of gravity



  - Please verify the equality.

discrete

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

Property of variance

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

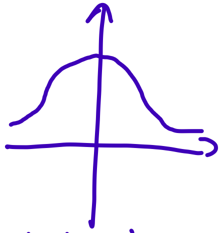
- $0 \le \mathbf{Var}[X] = \mathbf{E}[X^2] - (\mathbf{E}[X])^2$

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

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

- If $Y = aX$, then

$X \sim N(0,1)$



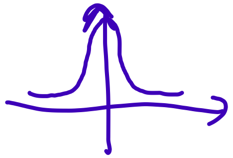
$$\mathbf{E}[Y] = a\mathbf{E}[X], \qquad \mathbf{Var}[Y] = a^2\mathbf{Var}[X].$$

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

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

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

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



- If $Y = X + b$, then

$$\mathbf{E}[Y] = \mathbf{E}[X] + b, \qquad \mathbf{Var}[Y] = \mathbf{Var}[X].$$

- if $Y = aX + b$, then

$$\mathbf{E}[Y] =$$
$$\mathbf{Var}[Y] =$$

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

$$\mathbf{E}[Y] =$$

$$\mathbf{Var}[Y] =$$

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:

- 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?

- Let $X_1, \cdots, X_n$ be a sequence of independent identically distributed random variable with mean μ and variance σ^2
- Let $S_n = \mathrm{X}_1 + \mathrm{X}_2 + \cdots + \mathrm{X}_\mathrm{n}$, what is the mean of S_n? What is the variance of S_n?

Background

The distribution of S_n spreads out as n increases.

But the situation is different if we consider the sample mean

$$M_n = \frac{X_1 + \cdots + X_n}{n} = \frac{S_n}{n}.$$

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

Background

- Given our calculation:

$$\mathbf{E}[M_n] = \mu, \ \ \mathrm{var}(M_n) = \frac{\sigma^2}{n}.$$

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

Background

- We will also consider a quantity which is intermediate between S_n and M_n.
- Z_n is defined as follows.

1. subtract $\boldsymbol{n\mu}$ from S_n, to obtain the zero-mean random variable $S_n - n\mu$
2. then divide by $\boldsymbol{\sigma\sqrt{n}}$, to form the random variable.

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

Formally

- Let $X_1, \cdots, X_n$ be a sequence of independent identically distributed random variable with mean μ and variance σ^2

- Define

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

- Let’s use jupyter notebook

The Central Limit Theorem

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

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

in the sense that

$$\lim_{n\to\infty} \mathrm{P}(Z_n \leq z) = \Phi(z)$$

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 X_i,
  - 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 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:

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

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