ECE 302: Lecture 4.4 Median, Mode, and Mean

Prof Stanley Chan

School of Electrical and Computer Engineering
Purdue University

Median

Given a sequence of numbers

n	1	2	3	4	5	6	7	8	9	...	100
x_n	1.5	2.5	3.1	1.1	-0.4	-4.1	0.5	2.2	-3.4	...	-1.4

Find the **median**.

- Step 1: You sort the sequence
- Step 2: You pick the one in the middle

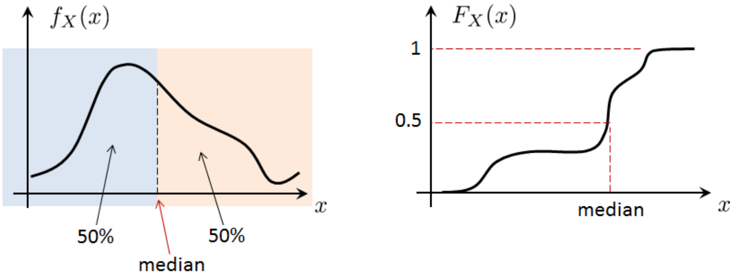
If we have a random variable, what is the **ideal** median?

Median from PMF

Definition

Let X be a continuous random variable with PDF f_X. The median of X is a point $c \in \mathbb{R}$ such that

$$\int_{-\infty}^{c} f_X(x)dx = \int_{c}^{\infty} f_X(x)dx. \tag{1}$$



Median from CDF

Theorem

The median of a random variable X is the point c such that

$$F_X(c) = \frac{1}{2}. \tag{2}$$

Proof.

Since $F_X(x) = \int_{-\infty}^{x} f_X(x')dx'$, we have

$$F_X(c) = \int_{-\infty}^{c} f_X(x)dx = \int_{c}^{\infty} f_X(x)dx = 1 - F_X(c).$$

Rearranging the terms shows that $F_X(c) = \frac{1}{2}$. □

Example

Example 1. (Uniform random variable) Let X be a continuous random variable with

- PDF: $f_X(x) = \frac{1}{b-a}$ for $a \leq x \leq b$, and is 0 otherwise.
- CDF: $F_X(x) = \frac{x-a}{b-a}$ for $a \leq x \leq b$.

Find median.

Solution: Want $F_X(c) = 1/2$.

$$c = \frac{a+b}{2}$$

Example

Example 2. (Exponential random variable) Let X be a continuous random variable with

- PDF: $f_X(x) = \lambda e^{-\lambda x}$ for $x \geq 0$
- CDF: $F_X(x) = 1 - e^{-\lambda x}$ for $x \geq 0$

Find median.

Solution: Want $F_X(c) = 1/2$.

$$c = (\log 2)/\lambda$$

Mode

Given a sequence of numbers

n	1	2	3	4	5	6	7	8	9	...	100
x_n	50	50	30	10	-40	-10	50	20	-30	...	-1

How to find the **mode**?

- Step 1: Sort the sequence
- Step 2: Pick the number that occurs most often?

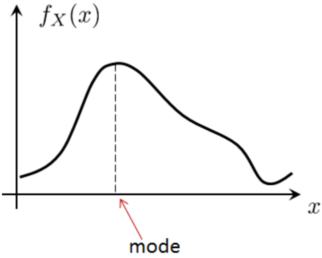
What is the **ideal** mode?

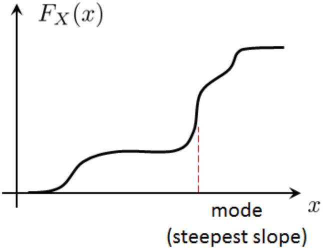
Mode from PDF and CDF

Definition

Let X be a continuous random variable. The mode is the point c such that $f_X(x)$ attains the maximum:

$$c = \underset{x \in \Omega}{\text{argmax}}\; f_X(x) = \underset{x \in \Omega}{\text{argmax}}\; \frac{d}{dx} F_X(x). \tag{3}$$





Example

Example 1. Let X be a continuous random variable with

- PDF $f_X(x) = 6x(1-x)$ for $0 \leq x \leq 1$.

Find mode.

Solution. The mode of X happens at $\underset{x}{\text{argmax}}\ f_X(x)$.

$$x = \frac{1}{2}$$

Mean

Given a sequence of numbers

n	1	2	3	4	5	6	7	8	9	...	100
x_n	1.5	2.5	3.1	1.1	-0.4	-4.1	0.5	2.2	-3.4	...	-1.4

Find the **mean**.

$$\overline{X} = \frac{1}{N}\sum_{n=1}^{N} X_n.$$

How to find mean from CDF?

Mean from CDF: Positive Case

Lemma

Let $X > 0$. Then, $\mathbb{E}[X]$ can be computed from F_X as

$$\mathbb{E}[X] = \int_0^\infty (1 - F_X(t))\, dt. \tag{4}$$

Proof:

$$\begin{aligned}
\int_0^\infty (1 - F_X(t))\, dt &= \int_0^\infty [1 - \mathbb{P}[X \leq t]]\, dt = \int_0^\infty \mathbb{P}[X > t] dt \\
&= \int_0^\infty \int_t^\infty f_X(x) dx dt \overset{(a)}{=} \int_0^\infty \int_0^x f_X(x) dt dx \\
&= \int_0^\infty \int_0^x dt f_X(x) dx = \int_0^\infty x f_X(x) dx = \mathbb{E}[X].
\end{aligned}$$

Interchange the integrations

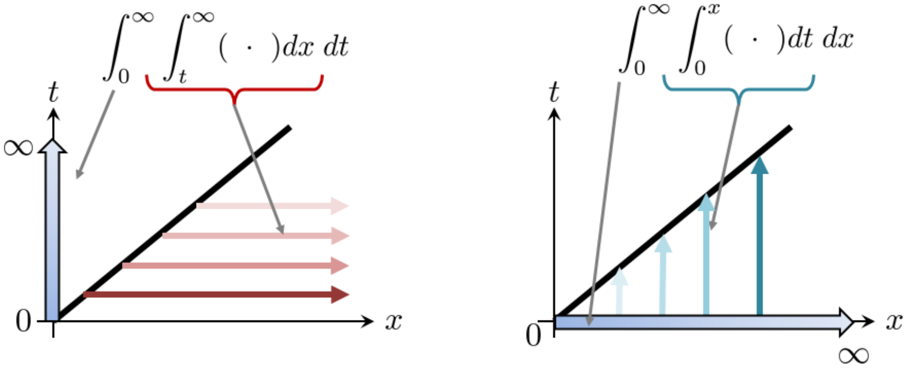


Figure: A double integration can be evaluated in two ways: x then t, or t then x.

Negative Case

Lemma

Let $X < 0$. Then, $\mathbb{E}[X]$ can be computed from F_X as

$$\mathbb{E}[X] = \int_{-\infty}^{0} F_X(t)dt. \tag{5}$$

Proof.

$$\begin{aligned}
\int_{-\infty}^{0} F_X(t)dt &= \int_{-\infty}^{0} \mathbb{P}[X \leq t]dt \\
&= \int_{-\infty}^{0}\int_{-\infty}^{t} f_X(x)dxdt \\
&= \int_{-\infty}^{0}\int_{x}^{0} f_X(x)dtdx = \int_{-\infty}^{0} xf_X(x)dx = \mathbb{E}[X]. \qquad \square
\end{aligned}$$

The overall result

Theorem

The mean of a random variable X can be computed from the CDF as

$$\mathbb{E}[X] = \int_0^\infty (1 - F_X(t))\,dt - \int_{-\infty}^0 F_X(t)dt. \tag{6}$$

Proof. Let $X = X^+ - X^-$. Then,

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{E}[X^+ - X^-] \\ &= \mathbb{E}[X^+] - \mathbb{E}[X^-] \\ &= \int_0^\infty (1 - F_X(t))\,dt - \int_{-\infty}^0 F_X(t)dt. \qquad \square\end{aligned}$$

Summary

We have learned three things:

Median

50% of the area, from left and from right.

Mode

Peak of PDF, steepest slope of CDF

Mean

- PDF: $\int_0^\infty t f_X(t) dt$, for $X > 0$
- CDF: $\int_0^\infty (1 - F_X(t))\, dt$, for $X > 0$

Questions?