cse 107 4-29-25

Independence ....

given 3 r.v.s X, Y, Z,

We say X, Y, Z are independent iff

$$p_{X,Y,Z}(x,y,z) = p_X(x) \cdot p_Y(y) \cdot p_Z(z)$$

- if X, Y, Z are indep., so are $f(X), g(Y), h(Z)$.

- also

$$E[XYZ] = E[X]E[Y]E[Z]$$

- and

$$Var(X+Y+Z) = Var(X) + Var(Y) + Var(Z)$$

Variance of the Binomial (n, p)

Recall X is Binomial(n, p) iff

$$X = X_1 + X_2 + \cdots + X_n$$

where X_i are indep. Bernoulli(p).

Also $Var(X_i) = p(1-p)$. Then since X_i are indep.

$$Var(X) = Var(X_1 + \cdots + X_n)$$
$$= Var(X_1) + \cdots + Var(X_n)$$
$$= p(1-p) + \cdots + p(1-p)$$
$$= np(1-p)$$

Ex.

Suppose we have a weighted coin, but we don't know $P(\text{head}) = p$. How can we estimate p? Flip coin n times, let X_i $(0,1)$ be outcome of i^{th} flip. let

$$S_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

we have $E[X_i] = p$, so

$$E[S_n] = E\left[\tfrac{1}{n}X_1 + \cdots + \tfrac{1}{n}X_n\right]$$

$$= \sum_{i=1}^{n} E\left[\tfrac{1}{n}X_i\right] = \frac{1}{n}\cdot\sum_{i=1}^{n} p = \frac{1}{\cancel{n}}\cdot \cancel{n}p = \boxed{p}$$

S_n is called the <u>Sample Mean</u>

How good an estimate of p is S_n?

<u>note</u>: X_i are indep, hence so are $\frac{1}{n}X_i$. Thus

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

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

$$= \sum_{i=1}^{n} \frac{1}{n^2} \cdot p \cdot (1-p) = n \cdot \frac{1}{n^2} p(1-p)$$

$$= \frac{p(1-p)}{n}$$

observe that

$$\mathrm{Var}(S_n) \to 0$$

as $n \to \infty$.

<u>Exercise</u>

Prove that if $\mathrm{Var}(Y) = 0$, then $Y = \text{constant} = E[Y]$.

<u>Note</u>

if X_i are not Bernoulli, but still independent, same calculation gives

$$Var(S_n) = \frac{Var(X_i)}{n} \to 0.$$

since $E[S_n] = E[X_i] = \underbrace{E[X]}_{\text{common mean}}$

we have S_n is a good approx. to mean $E[X]$.

Ex. (#38 on p. 132)

Alice passes through 4 traffic lights on way to work, each is equally likely to be red or green, independently.

(a) find PMF, mean, variance of # of red lights.

Soln

let X = # red lights. X is Binomial$(4, \frac{1}{2})$. so

$$P_X(k) = \binom{4}{k}\left(\frac{1}{2}\right)^k\left(\frac{1}{2}\right)^{4-k}$$

$$= \binom{4}{k}\left(\frac{1}{2}\right)^4$$

$$= \frac{4!}{k!\,(4-k)!}\cdot\frac{1}{16}$$

$$= \frac{3}{2k!\,(4-k)!} \qquad (k=0,1,2,3,4)$$

$$E[X] = np = 4\cdot\frac{1}{2} = \boxed{2}$$

$$\mathrm{Var}(X) = np(1-p) = 4\cdot\frac{1}{2}\cdot\frac{1}{2} = \boxed{1}$$

(b) Suppose Alice is delayed by 2 mins. at each red light. Find Variance of her commute time.

Soln

let T = commute time. Then

$$T = T_0 + 2X$$

where T_0 is her commute time with no red lights. Then

$$\begin{aligned} Var(T) &= Var(2X) \\ &= 2^2 \cdot Var(X) \\ &= 4 \cdot 1 = \boxed{4} \end{aligned}$$

3.1 Continuous Random Variables, PDFs

Defn

A r.v. $X : \Omega \to \mathbb{R}$ is called continuous iff there exists a function

$$f_X : \mathbb{R} \to \mathbb{R}$$

called the probability density function (PDF) of X, such that

$$P(X \in I) = \int_I f_X(x)\,dx$$

for 'any' subset $I \subseteq \mathbb{R}$.

In particular, if $I = [a,b]$

$$P(a \le X \le b) = \int_a^b f_X(x)\,dx .$$

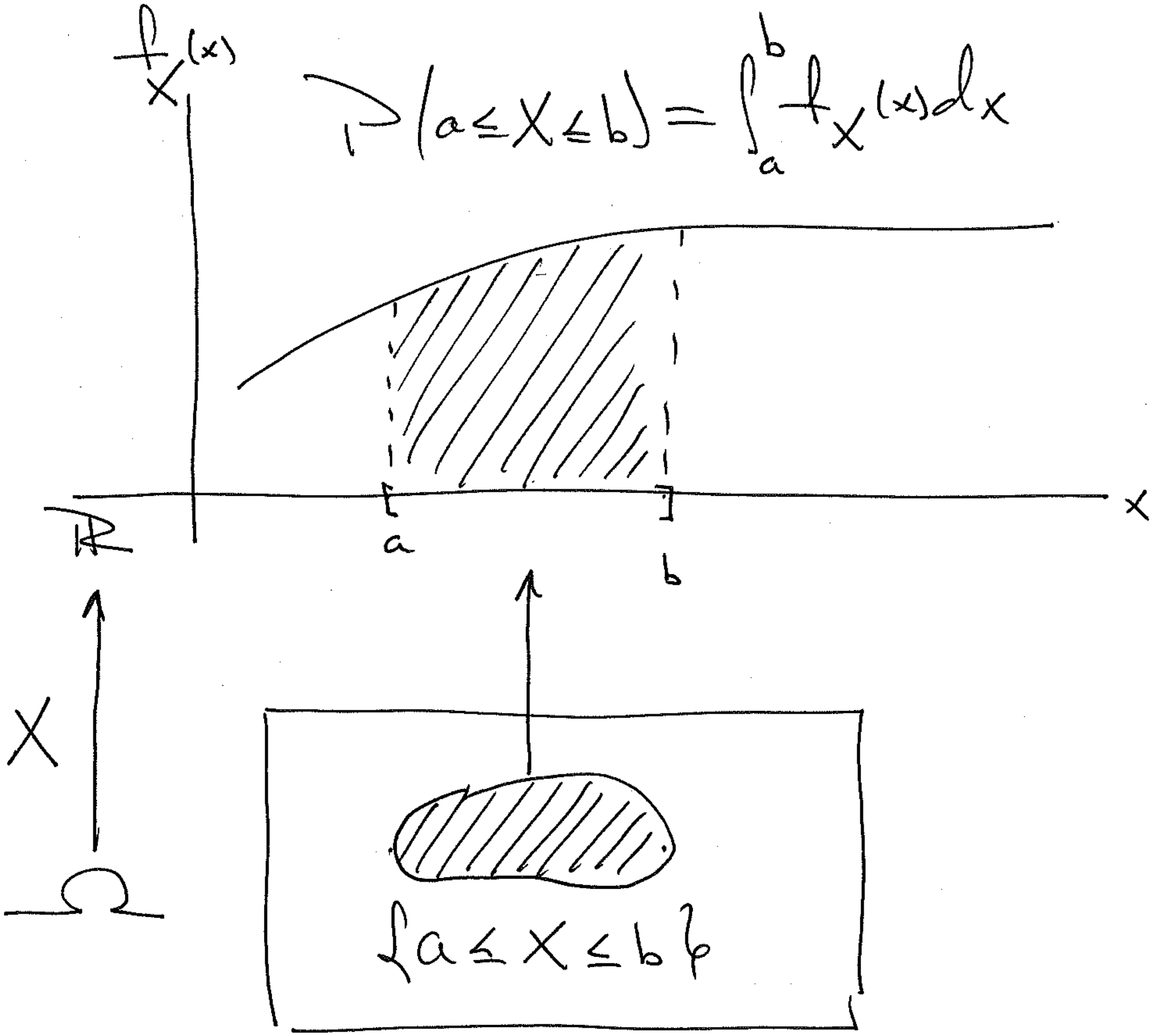
<u>note</u>: $P(X=a) = \int_a^a f_X(x)\,dx = 0.$

To be a valid PDF, $f_X(x)$ must satisfy

$$\int_{-\infty}^{\infty} f_X(x)\,dx = P(X \in \mathbb{R}) = P(\Omega) = 1 .$$

<u>note</u>: also $f_X(x) \ge 0$ for all $x \in \mathbb{R}$.

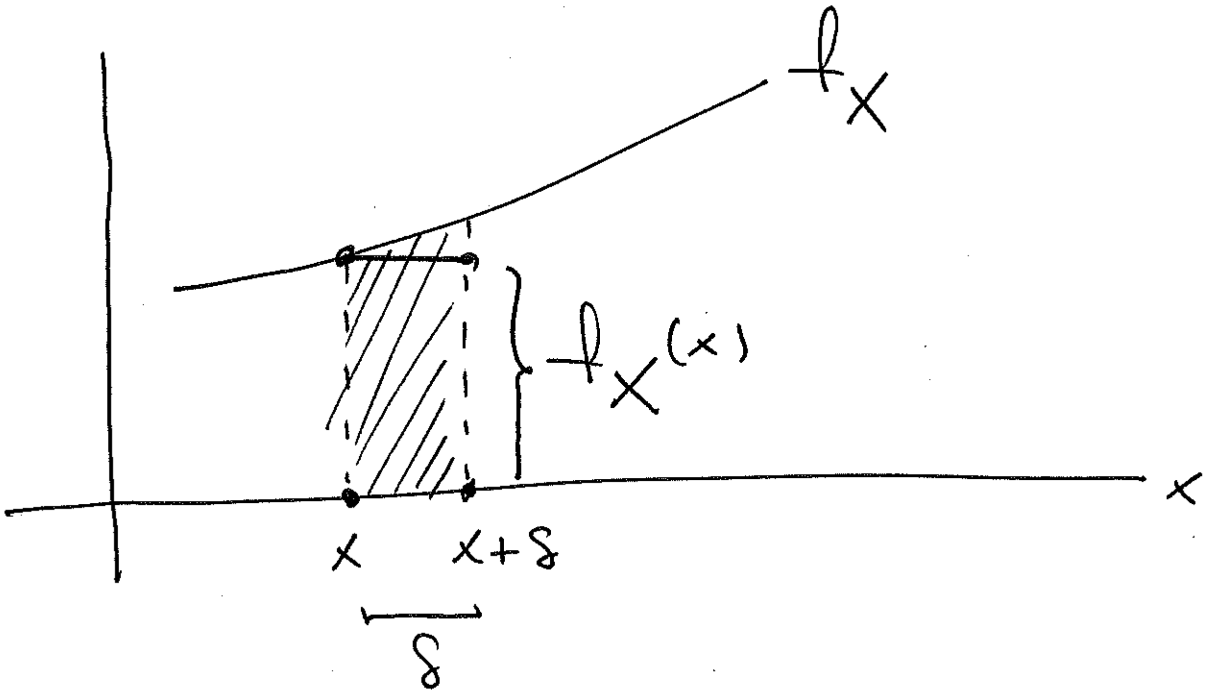
Picture



<u>note</u>: unlike PMF, the PDF does not give 'mass' at a point $x \in \mathbb{R}$.

Instead it gives 'mass density', 'mass per unit length'.

Let $\delta > 0$ be small, so if $f_X(x)$ is continuous it, is approximately constant on $[x, x+\delta]$



Thus

$$P(x \le X \le x+\delta) = \int_x^{x+\delta} f_X(t)\,dt \approx f_X(x) \cdot \delta$$

so $$f_X(x) \approx \frac{P(x \le X \le x+\delta)}{\delta}$$

$$= \frac{\text{Probability}}{\text{length}}$$

$\therefore$ $f_X(x)$ has units Probability per unit length, i.e. density.

note: f_X can have jump discontinuities.

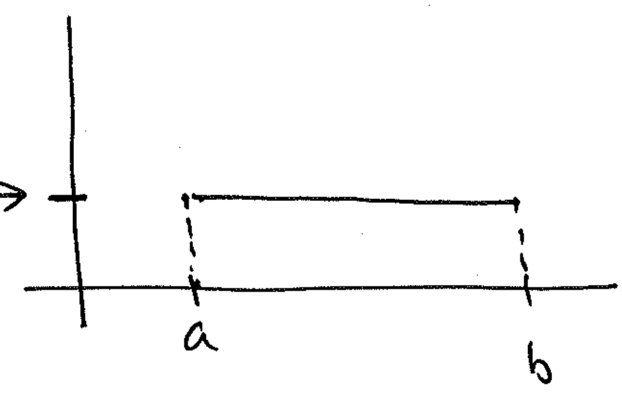
Ex. Continuous uniform on $[a, b]$

$$f_X(x) = \begin{cases} c & \text{if } a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

to find c, use $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$.

$$1 = \int_{-\infty}^{\infty} f_X(x)\,dx = \int_a^b c\,dx = c \cdot x \Big|_a^b = c \cdot (b-a)$$

$$\therefore \quad c = \frac{1}{b-a}$$



$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

Ex. Piecewise Constant PDF

Alice's time to work is 40-44 minutes in good weather, and 44-50 minutes in bad, uniformly distributed in each case. Suppose

$$P(\text{good}) = \frac{2}{3}, \quad \text{so } P(\text{bad}) = \frac{1}{3}.$$

find PDF of her driving time.

Solution

$$f_X(x) = \begin{cases} c_1 & \text{if } 40 \le x \le 44 \ (\text{good}) \\ c_2 & \text{if } 44 < x \le 50 \ (\text{bad}) \\ 0 & \text{otherwise} \end{cases}$$

Also

$$\frac{2}{3} = P(\text{good}) = \int_{40}^{44} c_1\,dx = c_1 \cdot (44-40) = 4c_1$$

$$\therefore \quad c_1 = \frac{2}{3} \cdot \frac{1}{4} = \frac{2}{12} = \boxed{\frac{1}{6}}$$

and

$$\frac{1}{3} = P(\text{bad}) = \int_{44}^{50} c_2\,dx = 6 \cdot c_2$$

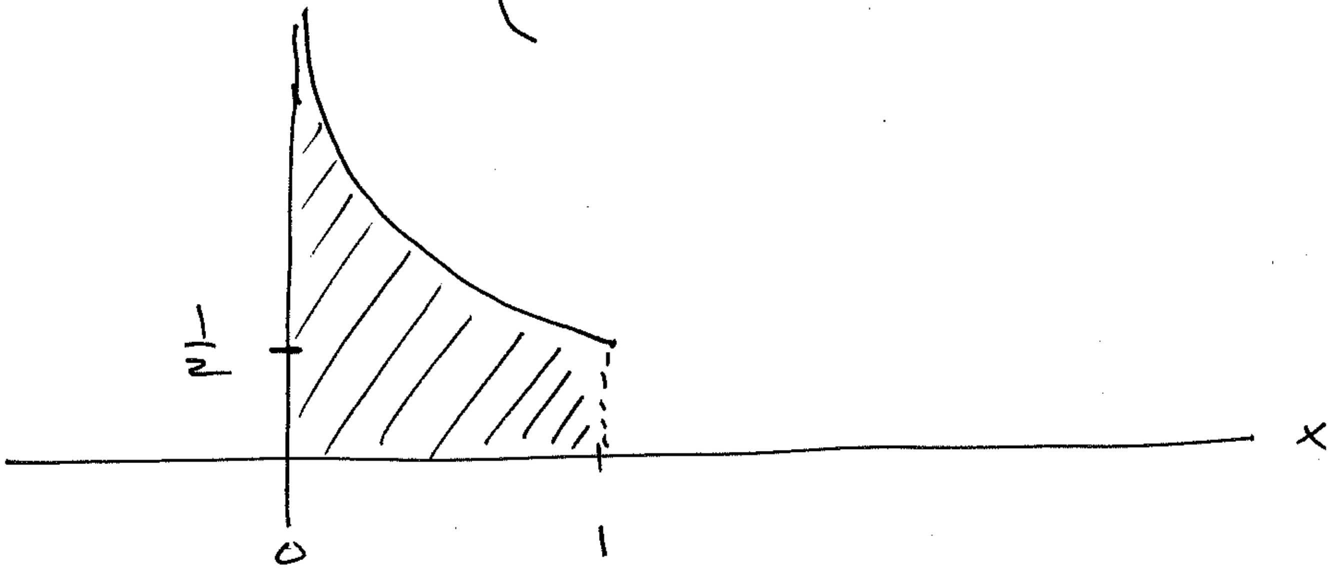
$$\therefore \quad c_2 = \frac{1}{3} \cdot \frac{1}{6} = \boxed{\frac{1}{18}}$$

Thus

$$f_X(x) = \begin{cases} \frac{1}{6} & \text{if } 40 \le x \le 44 \\ \frac{1}{18} & \text{if } 44 < x \le 50 \\ 0 & \text{otherwise} \end{cases}$$

Ex. let

$$f_X(x) = \begin{cases} \dfrac{1}{2\sqrt{x}} & \text{if } 0 < x \le 1 \\ 0 & \text{otherwise} \end{cases}$$



$$\int_0^1 \frac{1}{2\sqrt{x}}\,dx = \lim_{a\to 0^+} \int_a^1 \tfrac{1}{2} x^{-\frac{1}{2}}\,dx$$

$$= \lim_{a\to 0^+} \cancel{\left(\tfrac{1}{2}\right)} \cdot \frac{x^{\frac{1}{2}}}{\cancel{\left(\tfrac{1}{2}\right)}} \Bigg|_a^1 = \lim_{a\to 0^+} \left(\sqrt{1} - \sqrt{a}\right)$$

$$= 1$$

So, this is a valid PDF.

Defn

The expected value of a continuous r.v. X is

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

Expected value rule

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

see Problem #4 on P. 185, and solution, for the proof.

Defn

The n^{th} moment of X $(n \geq 0)$

is $E[X^n]$.

The Variance of X is

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

Theorem

$$Var(X) = E[X^2] - E[X]^2$$

Proof.

let $\mu = E[X]$. so

$$Var(X) = E[(X-\mu)^2]$$

$$= \int_{-\infty}^{\infty} (x-\mu)^2 \cdot f_X(x)\,dx$$

$$= \int_{-\infty}^{\infty} (x^2 - 2\mu x + \mu^2) \cdot f_X(x)\,dx$$

$$= \int_{-\infty}^{\infty} x^2 f_X(x)\,dx - 2\mu \int_{-\infty}^{\infty} x \cdot f_X(x)\,dx + \mu^2 \int_{-\infty}^{\infty} f_X(x)\,dx$$

$$= E[X^2] - 2\mu\cdot\mu + \mu^2$$

$$= E[X^2] - \mu^2 = E[X^2] - E[X]^2$$

$\blacksquare$