Computational statistics

Random variable.

$(\Omega, \mathcal{F}, P)$

$X : \Omega \to \mathbb{R}^d$ is $\mathcal{F}$. measurable.

$$P(X \in A) = P\{\omega \in \Omega \mid X(\omega) \in A\} = P(X^{-1}(A)) \quad \text{—} (*)$$

Exs: Tossing a coin 'n' number of times.

(Assumption: Tosses are independent) $P(H) = p$

$$\Omega = \{x_1 x_2 \dots x_n \mid x_i \in \{H, T\} \text{ for } i = 1, 2, \dots, n\}$$

$p(x_1 x_2 \dots x_n) = p^k (1-p)^{n-k}$ if there are k-heads

Define X = no. of heads in n tosses.

$X : \Omega \to \mathbb{R}$; $\text{im}(X) = \{0, 1, 2, \dots, n\}$

$$\text{Prob}(X = 0) = P\{\omega \in \Omega \mid X(\omega) = 0\} = P\{TT\dots T\} = (1-p)^n$$

$\text{Prob}\{X = k\} = \text{Prob}\{\omega \in \Omega \mid X(\omega) = k\}$

$= \binom{n}{k} p^k (1-p)^{n-k}$ for $k \in \{0, 1, \cdots, n\}$

p.m.f. here is

$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$ for $x = 0, 1, 2, \cdots, n$

$= 0$ o.w.

In general, any function $f: \mathbb{R} \to \mathbb{R}$ is a pmf

if $f(x) \geq 0$

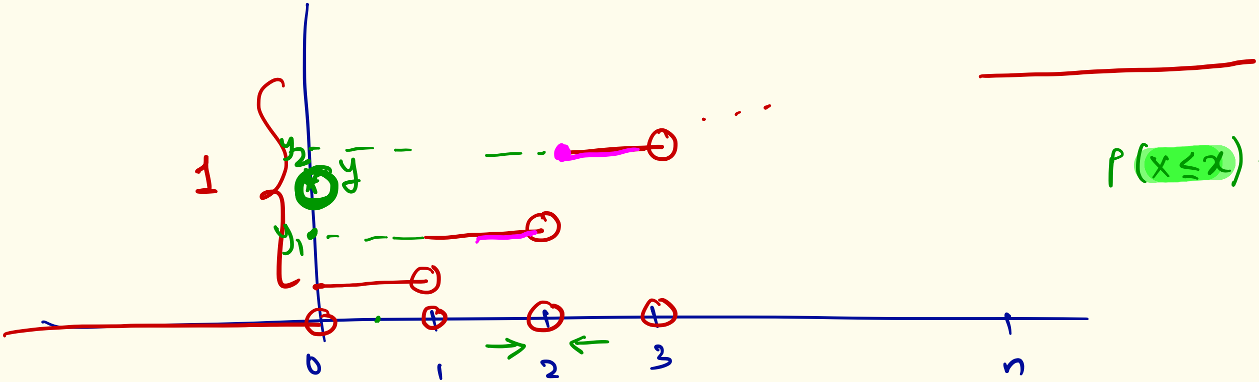
and $\sum f(x) = 1$ where $f(x)$ is non-zero on a countable set.

(K. L. Chung)

CDF (F)

$F : \mathbb{R} \longrightarrow \mathbb{R}$

$F(x) = \text{Prob}(X \le x) \quad \forall x \in \mathbb{R}$

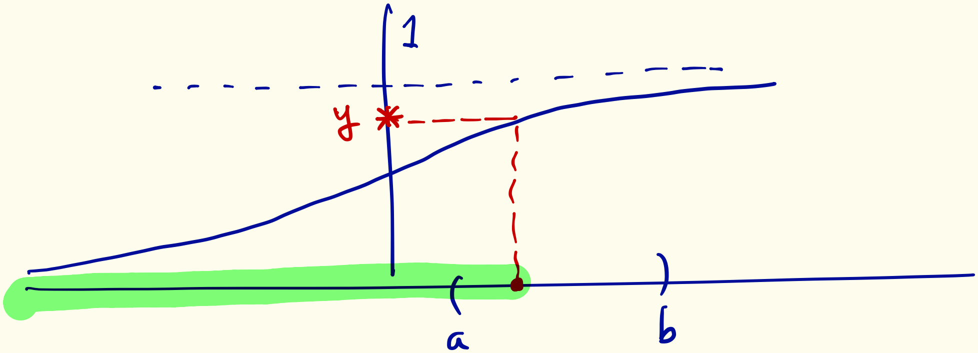


$P(X \le x) = y$

$F_X(\tfrac{1}{2}) = \text{Prob}(X \le \tfrac{1}{2}) = \text{Prob}(X = 0)$

(i) $0 \le F_X(x) \le 1$ (ii) $\lim\limits_{x \to \infty} F_X(x) = 1$ (iii) $\lim\limits_{x \to \infty^-} F_X(x) = 0$

(iv) $F_X(x)$ is non-decreasing (v) $F_X(x)$ is right continuous.



$$F_X(x) = \frac{1}{1+e^{-x}}$$

$$y \in [0,1]$$

$$F^{-1}(y) = x$$

$$\boxed{P(X \le x) = y}$$

$$P(a \le X \le b) = F(b) - F(a)$$

$$=$$

$f_X(x)$ is the pdf (density);

$$f_X(x) = \frac{d}{dx} F_X(x)$$

Some important densities:

$\mu = \frac{\alpha+\beta}{2}$

$var = \frac{(\alpha-\beta)^2}{12}$

Parameters

i) $X \sim U[\alpha, \beta]$

$$f_X(x) = \frac{1}{\beta - \alpha} \qquad x \in [\alpha, \beta]$$

α & β

$\alpha < \beta$

ii) $X \sim N(\mu, \sigma^2)$

mean $= \mu$

var $= \sigma^2$

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad -\infty < x < \infty$$

$\mu \in \mathbb{R}$

$\sigma > 0$

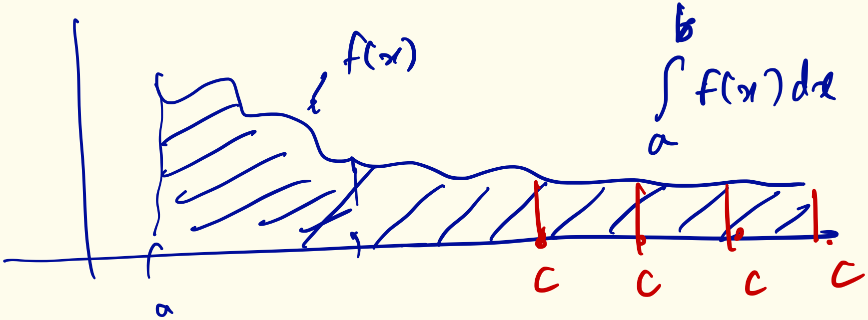
iii) $X \sim \text{gamma}(\alpha, \lambda)$

$$f_X(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \qquad x > 0 \; ; \; \alpha, \lambda > 0$$

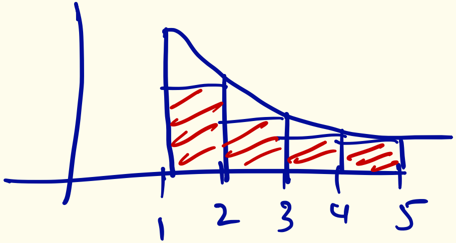
$\mu = \frac{\alpha}{\lambda}$; var $= \frac{\alpha}{\lambda^2}$

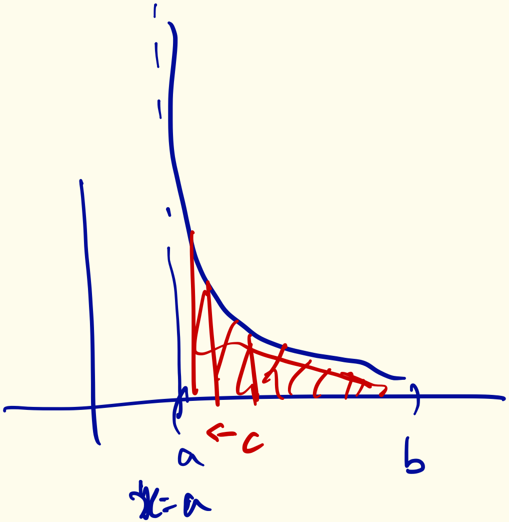
Gamma function

$$\boxed{\frac{\Gamma(\alpha)}{\lambda^\alpha} = \int_0^\infty e^{-\lambda x} x^{\alpha-1} \, dx}$$



$$\int_a^b f(x)\,dx$$





$f(x) = \dfrac{1}{x}$ $\qquad 1 \le x < \infty$

$$\int_1^\infty \frac{1}{x}\,dx$$

$$\lim_{c\to\infty} \int_1^c \frac{1}{x}\,dx = \lim_{c\to\infty} \log c$$

$$= \infty$$

iv) $X \sim \text{exponential}(\lambda) = \text{gamma}(1, \lambda)$ $\quad$ mean $= \frac{1}{\lambda}$, var $= \frac{1}{\lambda^2}$

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

v) $X \sim \text{Beta}(\alpha, \beta)$ $\quad$ mean $= \frac{\alpha}{\alpha+\beta}$

$f_X(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$ $\quad x \in [0,1]$

var $= \frac{\alpha\beta}{(\alpha+\beta)^2(1+\alpha+\beta)}$ $\quad \alpha, \beta > 0$

vi) $X \sim \text{Bernoulli}(p)$ $\quad$ mean $= p$; var $= p(1-p)$

$f_X(x) = p^x (1-p)^{1-x}$ $\quad x \in \{0, 1\}$ $\quad 0 \leq p \leq 1$

vii) $X \sim \text{Binomial}(n, p)$ $\quad$ mean $= np$; var $= np(1-p)$

$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$ $\quad x \in \{0, 1, 2, \ldots, n\}$ $\quad n \in \mathbb{N}$, $0 \leq p \leq 1$

viii) $X \sim \text{geometric}(p)$ $\quad$ mean $= \frac{1}{p}$; var $= \frac{1-p}{p^2}$

$f_X(x) = p(1-p)^{x-1}$ $\quad x \in \{1, 2, \ldots\}$ $\quad 0 \leq p \leq 1$

viii) $X \sim \text{poisson}(\lambda)$ mean = var = λ

$$f_X(x) = \frac{e^{-\lambda}\lambda^x}{x!} \qquad x \in \{0, 1, 2, \dots\} \qquad \lambda > 0$$

Moments of r.v.

mean :

$\mu = E(X) = \sum x f(x)$ X is discrete.

$= \int x f(x)$ X is continuous.

$E[h(X)] = \sum h(x) f(x)$ X is discrete

$= \int h(x) f_X(x)\, dx$ if X is cts.

median : $m \in \mathbb{R}$ is median if

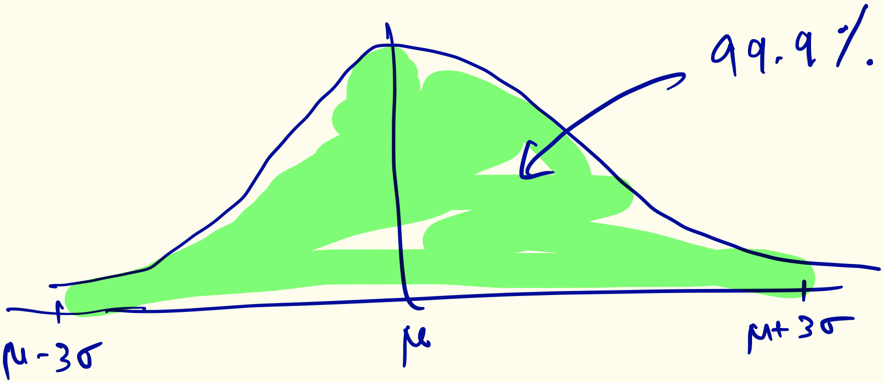
$$F_X(m) = \frac{1}{2}$$

Variance:

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

$$= E(X - \mu)^2$$

Ex: Normal $\quad X \sim N(\mu, \sigma^2)$



X : positive r.v. has pdf $f_X(x)$

For $x > 0$

$$E(X) = \int_0^\infty t f(t)\,dt = \int_0^x t f(t)\,dt + \int_x^\infty t f(t)\,dt$$

$$\geq \int_x^\infty x f(t)\,dt = x\, P(X \geq x)$$

$$E(X) \geq x\, P(X \geq x)$$

$$\Rightarrow \quad P(X \geq x) \leq \frac{E(X)}{x}$$

$$P(|X - \mu| \geq x) \leq \frac{\operatorname{var}(X)}{x^2}$$

$$P(|X - \mu| \geq x) \leq \frac{\sigma^2}{x^2}$$

$$|X - \mu| \geq x$$

$$\underbrace{(X - \mu)^2}_{} \geq \underbrace{x^2}_{}$$

Random vectors:

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}$$

For $x_1, x_2, \ldots, x_n \in \mathbb{R}$

$$F_X(x_1, \ldots, x_n) = \text{Prob}(X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n)$$

$$\text{Prob}(X \in B) = \iiint_B \cdots \int f_{X_1, \ldots, X_n}(x_1, \ldots, x_n)\, dx_1 \ldots dx_n$$

$\begin{pmatrix} X \\ Y \end{pmatrix} \leftarrow$ random vector. $\quad f_{X,Y}(x,y)$: joint pdf

$$f_{Y|X}(y \mid X = x) = \frac{f(x,y)}{f_X(x)} \qquad \text{conditional}$$

where $f_X(x) = \int f_{X,Y}(x,y)\, dy$: marginal density of X.

Independence of random variables.

The random variables $x_1, x_2, \ldots, x_n$ are called as independent r.v.s. if

$$f_{x_1,\ldots,x_n}(x_1,\ldots,x_n) = f_{x_1}(x_1)\, f_{x_2}(x_2) \cdots f_{x_n}(x_n)$$

(Joint density as product of marginals).

Expectation vector:

$$E(X) = \begin{pmatrix} E(X_1) \\ \vdots \\ E(X_n) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix} = \mu$$

$$\Sigma = E(x-\mu)(x-\mu)^T$$

Functions of random variables.

X is a cts r.v. with pdf $f_X(x)$.

Define $Z = aX + b$. (Let $z \in \mathbb{R}$, $a > 0$)

$$F_Z(z) = \text{Prob}(Z \le z)$$
$$= \text{Prob}(aX + b \le z)$$
$$= \text{Prob}\left(X \le \frac{z-b}{a}\right)$$
$$= F_X\left(\frac{z-b}{a}\right)$$

$$f_Z(z) = \frac{d}{dz} F_Z(z) = \frac{d}{dz} F_X\left(\frac{z-b}{a}\right)$$

$$= f_X\left(\frac{z-b}{a}\right) \cdot \frac{1}{a}$$

$X \sim N(0,1)$

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \qquad -\infty < x < \infty$$

$$\boxed{Z = \sigma X + \mu}$$

$$f_Z(z) = \frac{1}{\sigma} f_X\left(\frac{z-\mu}{\sigma}\right)$$

$Z \sim N(\mu, \sigma^2)$

Ex: X is cts. r.v. with pdf $f_X(x)$.

Define $Z = g(X)$ where g is monotonically decreasing.

$$F_Z(z) = \text{Prob}(Z \le z)$$
$$= \text{Prob}(g(X) \le z)$$
$$= \text{Prob}(X \ge g^{-1}(z))$$
$$= 1 - F_X(g^{-1}(z))$$

$$f_Z(z) = \frac{d}{dz} F_Z(z)$$

$$\boxed{f_Z(z) = - f_X(g^{-1}(z)) \frac{d}{dz} g^{-1}(z)}$$

Order statistics:

Let $X_1, X_2, \cdots, X_n$ be iid with pdf $f_X(x)$.

order statistics $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$

$$P(X_{(n)} \le x) = P(X_1 \le x, X_2 \le x, \cdots, X_n \le x)$$

$$= P(X_1 \le x)\, P(X_2 \le x) \cdots P(X_n \le x)$$

$$F_{X_{(n)}}(x) = (F_X(x))^n$$

$$\boxed{f_{X_{(n)}}(x) = n\,(F_X(x))^{n-1}\, f_X(x)}$$

Q: What is the density of $X_{(1)} = \min\{X_1, \cdots, X_n\}$

$$P(X_{(1)} \ge x) = 1 - F_{X_{(1)}}(x)$$

$$f_{X_{(1)}}(x) = n\,(1 - F_X(x))^{n-1}\, f_X(x)$$