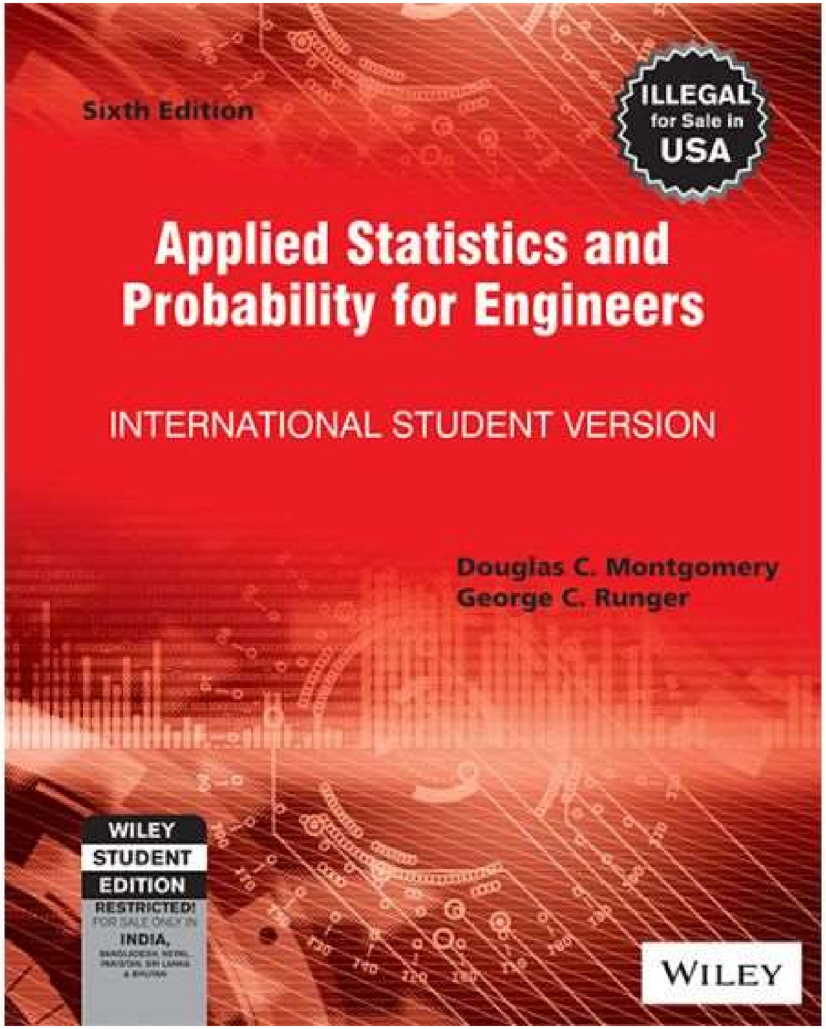
Probability and Statistics

Textbook

Chapters 1 - 10



Engineering Method and Statistical Thinking

- Develop a clear and concise description of problem
- Identify the important factors that affect the problem
- Propose a model – include assumptions, limitations; requires engineering knowledge
- Conduct experiments / collect data – test and validate the model
- Refine the model
- Use (manipulate!) the model
- Conduct experiments to confirm proposed solution
- Draw conclusions or make recommendations

Measure fuel efficiency

- Description of the problem?

  Scope.

- Factors affecting? Speed, distance

  age, maintenance traffic, weight, aerodynamics, engine char.

  fuel char.

- Propose a model

  $F.E = f_1(\text{speed, traffic, engine} \dots)$

  $f_2(\text{how angry a driver is?})$

- Conduct experiments

  Collect data

- Refine the model

Collecting Data

Retrospective Study	Observational Study	Designed Experiments
Already available! May not have all the parameters and ranges?	Can collect additional parameters for a short period No complete control	Establish cause and effect Test interaction effects

Mechanistic and Empirical Models

$$V = IR$$

built from knowledge of physical mechanism

$$V = IR + \varepsilon$$

ε captures effects of unmodelled sources of variability

Random variables: fn. that assigns a value (real number) to each outcome in the S. S. of a random expt.

- Discrete RV - finite set of values
- Continuous RV - intervals (infinitely many)

Prob. while rolling a N-face die with equally likely outcomes. $P = \dfrac{1}{N}$ $SS = \{1,2, \dots N\}$ $x = \{1, 2, \dots N\}$

Discrete RV : X : outcomes $x_1, x_2 \dots x_n$

$$\sum_{i=1}^{n} f(x_i) = 1$$

$$f(x_i) \geq 0$$

Prob. Mass fn. (PMF)

$$P(X = x_i) = f(x_i)$$

Cum. Dist. fn. (CDF)

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x_i)$$

$$0 \leq F(x) \leq 1$$

$$\text{if } x \leq y \Leftrightarrow F(x) \leq F(y)$$

Mean : $\mu = E[X] = \sum_{\forall x} x \cdot f(x)$ (Location)

Variance : $\sigma^2 = V[X] = E[(X-\mu)^2] = \sum_{\forall x} (x-\mu)^2 \cdot f(x)$

σ – standard deviation (spread)

second moment about the mean

$= E[x^2] - (E[x])^2$

$= \sum x^2 \cdot f(x) - \mu^2$

Eg. N faced die, $X = [1, 2, \ldots N]$ - equally likely. mean, variance?

mean $= \sum x \cdot f(x) = \sum_{i=1}^{N} i \cdot \frac{1}{N} = \frac{1}{\cancel{N}} \cdot \frac{\cancel{N}(N+1)}{2}$

Var. $= \sum x^2 \cdot f(x) - \mu^2 =$

$\frac{N^2-1}{12}$

Median, Mode.

Discrete Distributions:

$b =$

Uniform dist: $f(x_i) = \frac{1}{N}$ $\quad X: a, a+1, \ldots, (a+N)$

$$\mu = \sum x \cdot f(x) = a \cdot \frac{1}{N} + (a+1) \cdot \frac{1}{N} + \ldots = \frac{(a+b)}{2}$$

$$\sigma^2 = E[(x-\mu)^2] = \frac{(b-a+1)^2 - 1}{12}$$

Binomial ~~dist~~ dist:

n indep. Bernoulli trials $<$ p — success — 1 ; $(1-p)$ — 0. e.g. flip a coin

x successes in n trials

$${}^nC_x = \frac{n!}{(n-x)!\,x!}$$

PMF $\quad f(x) = {}^nC_x (p)^x (1-p)^{n-x}$

$E[x] = 1 \cdot p + 0 \cdot (1-p) = p$

$V[x] = (0-p)^2 (1-p) + (1-p)^2 p = p(1-p)$

$\mu = E[x] = \underline{np}$

$\sigma^2 = np(1-p)$

Geometric dist. X: # of trials to 1st success.

PMF $f(x) = (1-p)^{x-1} \cdot p$

$$\boxed{\sum_{i=1}^{\infty} a^i = \frac{a}{1-a}}$$

$0 \le a \le 1$

$$\mu = \sum_{k=1}^{\infty} k p (1-p)^{k-1} = p \sum_{k=1}^{\infty} k \cdot q^{k-1}$$

$$= \frac{d}{dq}\left[p \sum q^k\right] = \frac{d}{dq}\left[\frac{pq}{1-q}\right] = \frac{p}{(1-q)^2} = \frac{1}{p}$$

$$\sigma^2 = E[X^2] - (E[X])^2 = \frac{(1-p)}{p^2}$$

Lack of memory property!

Negative Binomial: r successes, how many trials give r.

PMF : $f(x) = {}^{x-1}C_{r-1} \cdot (1-p)^{x-r}\, p^{r-1} \cdot p = {}^{x-1}C_{r-1}(1-p)^{x-r} p^{r}$

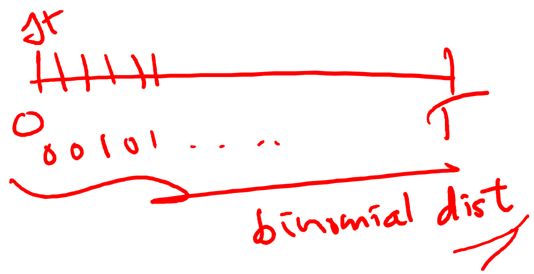
$$\mu = E[X] = \frac{r}{p} \qquad V[X] = \frac{r(1-p)}{p^2}$$

Poisson Dist:

time $[0, T]$ Divide into $\Delta t = \frac{T}{n}$ p : prob. of veh. arriving in Δt

λ – mean for unit time

$$E[X] = \lambda T = np \Rightarrow p = \frac{\lambda T}{n}$$



Poisson Dist. :

PMF $P(X=x) = {}^nC_x \, p^x (1-p)^{n-x} = \cancel{\frac{n!}{(n-x)!\,x!}} \left(\frac{\lambda T}{\cancel{n}}\right)^x \underbrace{\left(1-\frac{\lambda T}{n}\right)^{n-x}}_{}$

$$= \frac{(\lambda T)^x}{x!} \left[\left(1-\frac{\lambda T}{n}\right)^{\frac{n-x}{T}}\right]^T$$

$a = -x$

$$= \frac{(\lambda T)^x}{x!} e^{-\lambda T}$$

$$\boxed{\begin{gathered}\lim_{x \to 0} (1+ax)^{1/x} = e^a \\ \Delta t \to 0,\ n \to \infty\end{gathered}}$$

Poisson dist.

$$E[x] = V[x] = \lambda T$$

$$E[x] = \sum_{x=0}^{\infty} x \cdot \frac{e^{-\lambda T} (\lambda T)^{x}}{x!} = \lambda T \underbrace{\sum_{x=1}^{\infty} \frac{e^{-\lambda T} (\lambda T)^{x-1}}{(x-1)!}}_{=1} = \lambda T$$

$$V[x] = E[x^2] - (E[x])^2$$

$$E[x^2] = \sum_{x=0}^{\infty} x^2 \frac{e^{-\lambda T} (\lambda T)^{x}}{x!} = \lambda T \sum_{x=1}^{\infty} x \cdot \frac{e^{-\lambda T} (\lambda T)^{x-1}}{(x-1)!}$$

$$= \lambda T \underbrace{\sum_{x=1}^{\infty} (x-1) \frac{e^{-\lambda T} (\lambda T)^{x-1}}{(x-1)!}}_{\lambda T} + \lambda T$$

$$V[x] = \cancel{(\lambda T)^2} + \lambda T - \cancel{(\lambda T)^2} = \lambda T$$

Continuous Dist. pdf
prob. density fn.

$f(x) \geq 0$ $\quad \int_{-\infty}^{\infty} f(x)\,dx = 1$

$P(a \leq x \leq b) = \int_a^b f(x)\,dx$

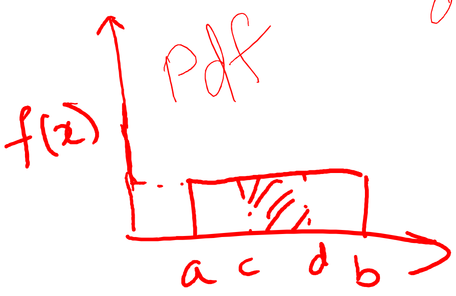
CDF: $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)\,dx$

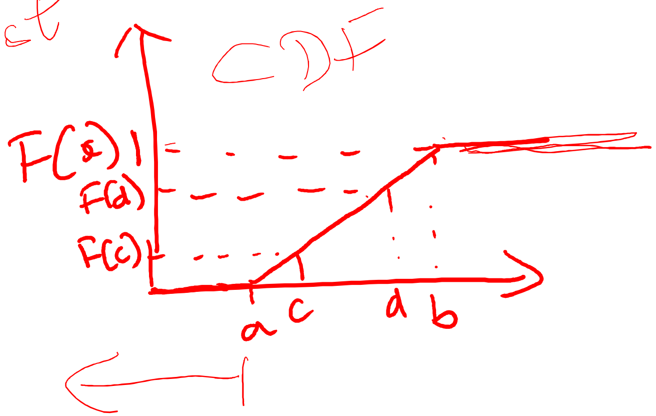
$f(x) = \dfrac{dF(x)}{dx}$, $F(-\infty) = 0$ & $F(+\infty) = 1$

$P(c \leq x \leq d) =$ area under the curve.

$= F(d) - F(c)$

Uniform dist





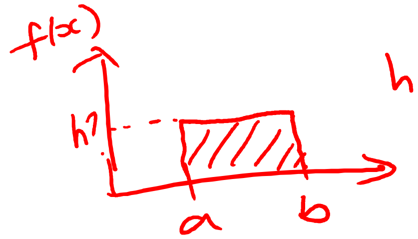
Expected value : $E[X] = \mu = \int_{-\infty}^{\infty} x \cdot f(x)\,dx$

Expected value of a fn. of a R.V. $h(x)$

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

Variance $\sigma^2 = V[X] = \int_{-\infty}^{\infty} (x-\mu)^2 \cdot f(x)\,dx = \int_{-\infty}^{\infty} x^2 \cdot f(x)\,dx - \mu^2$

Continuous uniform dist.



$h \cdot (b-a) = 1$

$h = \frac{1}{b-a}$

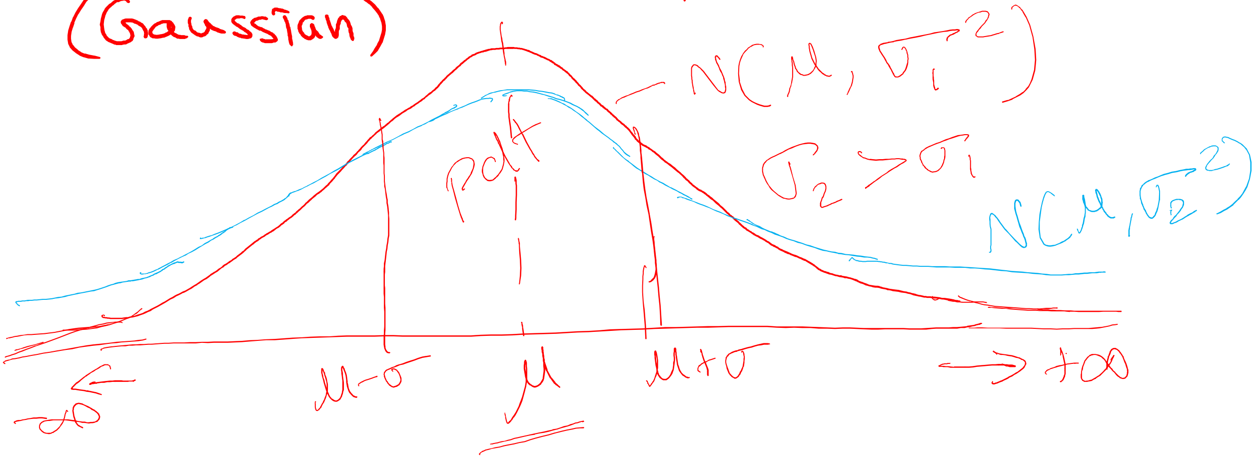
$$f(x) = \frac{1}{b-a} \quad //$$

$$E[X] = \frac{a+b}{2}, \quad V[X] = \frac{(b-a)^2}{12}$$

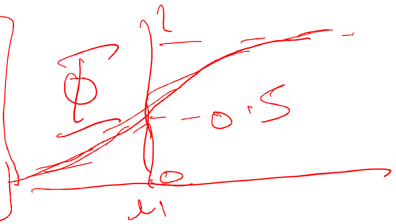
CDF : $F(x) = \int_{a}^{x} \frac{1}{b-a} \cdot dx = x - a$

Normal dist. (Gaussian)

pdf. $f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ $\quad -\infty < x < +\infty$



$$\boxed{\underline{\underline{N}}(\underset{\uparrow\,\text{mean}}{\mu}, \underset{\uparrow\,\text{var.}}{\sigma^2})}$$



$P(\mu-\sigma \le X \le \mu+\sigma) = 0.6827$

$P(\mu-2\sigma \le X \le \mu+2\sigma) = 0.9565$

$(\quad 3\sigma \qquad 3\sigma) = 0.9973$

CDF – closed form doesn't exist

Standard normal dist.

N-RV

std. normal variate $\rightarrow$ $z = \frac{X-\mu}{\sigma}$

std. normal variate

$Z \sim N(0, 1)$.

$$P(X \le x) = P\left(\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right) = P(Z \le z) = \Phi(z)$$

$(\mu + 2\sigma)$

$\uparrow$ CDF Tables std. normal var.

E.g. The dia of a shaft in a storage device $N(0.2508, (0.0005)^2)$

4.16 Specification: 0.2500 ± 0.0015 inches

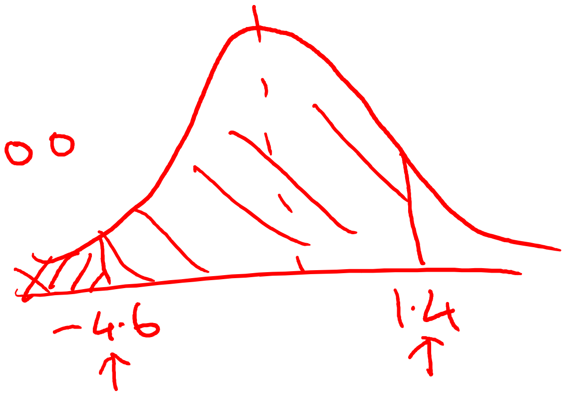
What proportion of shafts conform to the specs.

$$P(0.2485 < X < 0.2515) = P\left(\frac{0.2485 - 0.2508}{0.0005} < Z < \frac{0.2515 - 0.2508}{0.0005}\right)$$

$$= P(-4.6 < Z < 1.4)$$

$$= \Phi(1.4) - \Phi(-4.6)$$

$$\Phi(1.4) - \Phi(-4.6) = \underline{\underline{0.9192}} - 0.0000$$



Exponential Distribution

Time gap between successive events $- X$

$$P(X > x) \equiv P(N = 0 \text{ upto } x) = \frac{e^{-\lambda x}(\lambda x)^0}{0!} = \underline{e^{-\lambda x}}$$

CDF $\quad F(x) = P(X \leq x) = 1 - P(X > x)$

$$= 1 - e^{-\lambda x}$$

pdf $\quad f(x) = \frac{dF(x)}{dx} = \lambda e^{-\lambda x} \quad //$ exp. dist

$$\mu = \lambda^{-1}, \quad \sigma^2 = \lambda^{-2}$$

E.g. $\lambda^{-1} = 0.5$ min

$P(X < 15\,s) = P(X < 0.25m) = F(0.25) = 1 - e^{\frac{-0.25}{0.5}} = \underline{0.4}$

Say $\emptyset$ vehicles arrive $[0, 2m]$. Prob of next veh arriving in next 15s?

$P(B/A) = P(A \cap B)/P(A)$

$P(X < 2.25 / X > 2)$

$= \dfrac{P(2 < X < 2.25)}{P(X > 2)} = \dfrac{F(2.25) - F(2)}{1 - F(2)}$

$= \dfrac{\left(1 - e^{-2.25/0.5}\right) - \left(1 - e^{-2.0/0.5}\right)}{e^{-2.0/0.5}} = 0.4$

Lack of memory property.

$P(X < t_1 + t_2 / X > t_1) = P(X < t_2)$

Erlang and Gamma dist.

Time until 'r' arrivals/events occur in a Poisson process

$$P(X > x) = \sum_{k=0}^{r-1} \frac{e^{-\lambda x} (\lambda x)^k}{k!} = 1 - F(x)$$

diff. & re-arrange

$$f(x) = \frac{\lambda^r x^{r-1} e^{-\lambda x}}{(r-1)!} \qquad r = 1, 2, 3, \ldots \quad x > 0.$$

Gamma function:

$$\Gamma(r) = \int_0^\infty x^{r-1} e^{-x} \, dx, \quad r > 0$$

$$\Gamma(r) = (r-1)\Gamma(r-1)$$

For integer values, $\Gamma(r) = (r-1)!$

Gamma dist: pdf: $f(x) = \dfrac{\lambda^{x} x^{r-1} e^{-\lambda x}}{\Gamma(r)}$ $\qquad \mu = \dfrac{r}{\lambda}$ $\qquad \sigma^2 = \dfrac{r}{\lambda^2}$

Lognormal dist.

If $W \sim N(\theta, \omega^2)$, then $\boxed{X = e^{W}}$ is dist. lognormal RV

pdf: $f(x) = \dfrac{1}{x\omega\sqrt{2\pi}}\, e^{\left[-\frac{(\ln(x)-\theta)^2}{2\omega^2}\right]}$, $0 < x < \infty$

$\mu = e^{\theta + \frac{\omega^2}{2}}$, $V(X) = \sigma^2 = e^{2\theta+\omega^2}\left(e^{\omega^2} - 1\right)$

E.g. (4.26): A laser has a lifetime in hrs that follows a lognormal dist.

$\theta = 10h$, $\omega = 1.5h$. $P(X > 10{,}000h)$

$P(X > 10000h) = 1 - P(\exp(W) \le 10000)$

$= 1 - P(W \le \ln(10000))$

$= 1 - \Phi\left[\dfrac{\ln(10000) - 10}{1.5}\right] = 1 - \Phi(-0.52)$

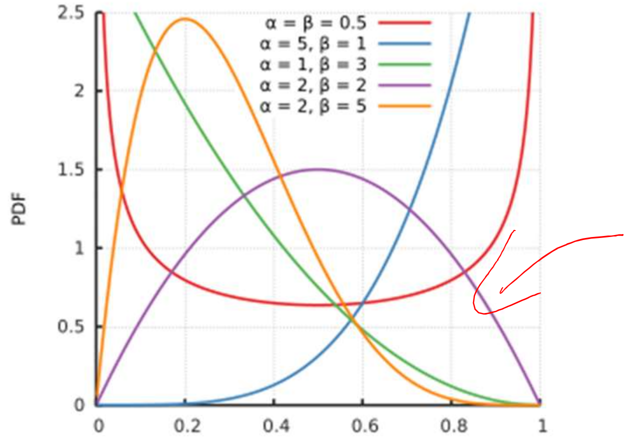
$= ?\ \underline{0.6985} \sim 0.7$

Beta distribution

$\alpha > 0, \ \beta > 0, \ x \in [0,1]$

pdf : $f(x) = \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$

$\mu = \dfrac{\alpha}{\alpha+\beta}, \quad \sigma^2 = \dfrac{\alpha\beta}{(\alpha+\beta)^2(1+\alpha+\beta)}$



Joint Prob. Dist. X & Y are discrete RV

Joint PMF $\quad f_{XY}(x, y) \geq 0$ & $\sum_x \sum_y f_{XY}(x,y) = 1$

$$\begin{array}{c} f_{XY}(x,y) \\ \parallel \\ P(X = x, Y = y) \end{array}$$

X & Y are continuous!

Joint pdf $\quad \underline{\underline{f_{XY}(x,y) \geq 0}}$ & $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\,dx\,dy = \underline{\underline{1}}$

$$P((X,Y) \in \underbrace{\underline{\underline{R}}}_{\text{region}}) = \iint_{\underline{\underline{R}}} f_{XY}(x,y)\,dx\,dy$$

Marginal prob. distn.

$$f_X(x) = \int_Y f_{XY}(x,y)\,dy$$

Conditional prob. dist. $\quad f_{Y/x}(y) = \dfrac{f_{XY}(x,y)}{f_X(x)}$

Conditional mean & variance

$$E[Y/x] = \int_y y \cdot f_{Y/x}(y)\, dy$$

$$V[Y/x] = \int_y y^2 f_{Y/x}(y)\, dy - \mu^2_{Y/x}$$

<u>E.g. (5-2)</u>

X : time until a computer server connects to a machine (millisec.)

Y : time until server authorizes (millisec) $X < Y$

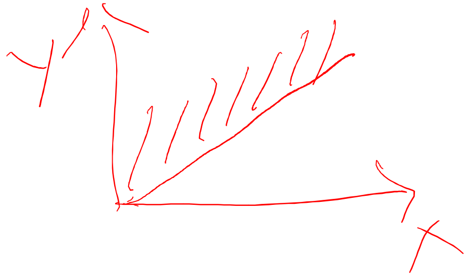
Assume : $f_{XY}(x,y) = \boxed{6\times10^{-6} \exp(-0.001x - 0.002y)}$ $x<y$.

Is it a valid jpdf

$$\int_0^\infty \left(\int_x^\infty 6\times10^{-6} e^{-0.001x-0.002y} dy \right) dx = 6\times10^{-6} \int_0^\infty \left(\int_x^\infty e^{-0.002y} dy \right) e^{-0.001x} dx$$

$$= 6\times10^{-6} \int_0^\infty \left(\frac{e^{-0.002x}}{0.002} \right) e^{-0.001x} dx.$$

$$= \frac{\cancel{6\times10^{-6}}}{\cancel{0.002}} \left(\frac{1}{\cancel{0.003}} \right) = 1$$



$$P(X<1000 \text{ and } Y<2000) = 6\times10^{-6}\int_0^{1000}\left(\int_x^{2000} e^{-0.002y}\,dy\right)e^{-0.001x}\,dx$$

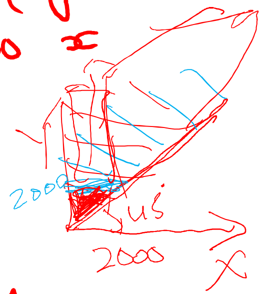
$$= 6\times10^{-6}\int_0^{1000}\left(\frac{e^{-0.002x}-e^{-4}}{0.002}\right)e^{-0.001x}\,dx = 0.915.$$

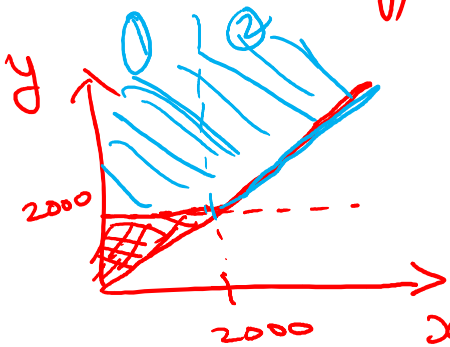
E.g. 5-4

$$P(Y>2000 \text{ milliseconds}) = \int_0^{2000}\left(\int_{2000}^{\infty} 6\times10^{-6}e^{-0.001x-0.002y}\,dy\right)dx + \int_{2000}^{\infty}\left(\int_x^{\infty} 6\times10^{-6}e^{-0.001x-0.002y}\,dy\right)dx$$

$$= 0.0475+0.0025$$

$$= 0.05$$





Marginal pdf

$$f_Y(y) = \int_x f_{XY}(x,y)\,dx\,dy$$

$$= \int_0^y 6\times10^{-6}e^{-0.001x-0.002y}\,dx = 6\times10^{-6}e^{-0.002y}(1-e^{-0.001y})$$

$$P(Y>2000) = 6\times10^{-6}\int_{2000}^{\infty} e^{-0.002y}(1-e^{-0.001y})\,dy = 0.05$$

Conditional prob. density: $f_{Y/x}(y) = \dfrac{f_{XY}(x,y)}{f_X(x)}$

$E(Y/x) = \int y\, f_{Y/x}(y)\,dy$

$V(Y/x) = \int y^2 f_{Y/x}(y)\,dy - \mu^2_{Y/x}$

E.g. Mobile response times

5-1, 5-7 & 5-9

$f_{XY}(x,y)$ x : # of bars of signal strength

y: response time (in s)

y	1	2	3	$f_Y(y)$
4	0.15	0.1	0.05	0.3
3	0.02	0.1	0.05	0.17
2	0.02	0.03	0.2	0.25
1	0.01	0.02	0.25	0.28
$f_X(x)$	0.2	0.25	0.55	

$$\frac{f_{Y/x}(y)}{?} = \frac{f_{XY}(x,y)}{f_X(x)}$$

	1	2	3
4	$0.15/0.2 = 0.75$	$\frac{0.1}{0.25} = 0.4$	0.091
3	0.1	0.4	0.091
2	0.1	0.12	0.364
1	0.05	0.08	0.454
	1	1	

$$E(Y/1) = \sum y \cdot f_{Y/x}(y) = 1 * 0.05 + 2 * 0.1 + 3 * 0.1 + 4 * 0.75$$

$$= 3.55 \text{ s}$$

$E(Y/2)$ & $E(Y/3)$

$$V(Y/1) = \sum (y - \mu_{Y/x})^2 \cdot f_{Y/x}(y)$$

$$= (1 - 3.55)^2 \cdot 0.05 + (2 - 3.55)^2 \cdot 0.1 + (3 - 3.55)^2 \cdot 0.1$$

$$+ (4 - 3.55)^2 \cdot 0.75 = 0.748 \text{ s}^2$$

More than 2 R.Vs: $X_1, X_2, \ldots X_p$ are p RVs

$$f_{X_1 X_2 \ldots X_p}(x_1, x_2, \ldots, x_p) \geq 0.$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1 X_2 \ldots X_p}(x_1, x_2 \ldots x_p) = 1.$$ ← property of pdf

Marginal PDF: $f_{X_i}(x_i) = \underbrace{\int\int \cdots \int}_{(p-1) \text{ integrals}} f_{X_1 \ldots X_p}(x_1 \ldots x_p)\, dx_1 \ldots dx_{i-1} dx_{i+1} \ldots dx_p$

$$E[X_i] = \int_{-\infty}^{\infty} x_i f_{X_i}(x_i)\, dx_i \qquad V[X_i] = \int_{-\infty}^{\infty} (x_i - \mu_{X_i})^2 f_{X_i}(x_i)\, dx_i$$

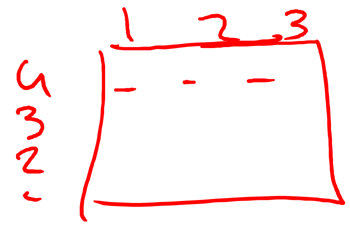
Independence: $X_1, X_2, \ldots X_p$ are independent, if and only if

$$f_{X_1 X_2 \ldots X_p}(x_1, x_2, \ldots x_p) = f(x_1) f(x_2) \ldots f(x_p)$$

$$\forall\, x_1, x_2 \ldots x_p$$

Covariance: $\sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)] = E[XY] - \mu_X \mu_Y$

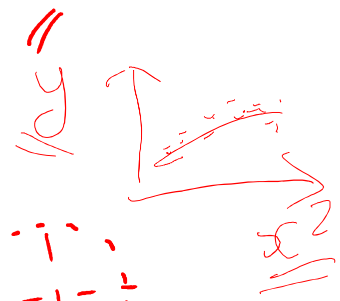
E.g. Mobile signal strength vs response time



$\sigma_{XY} = ?$ $\quad \mu_X = 2.35 \quad \mu_Y = 2.49$

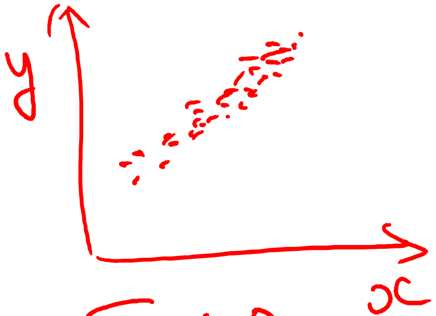
$(x-\mu_X)(y-\mu_Y) \cdot f_{XY}(x,y)$

$$\sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)] = (1-2.35)(1-2.49)\cdot 0.01 + (2-2.35)(1-2.49)\,0.02 + \dots = -0.5815$$



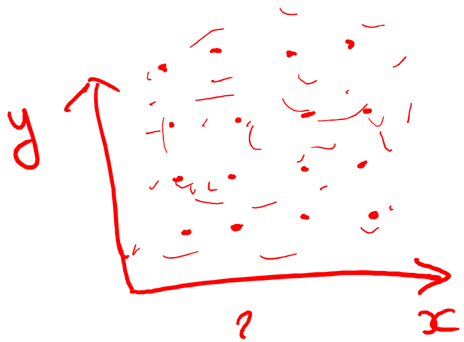
Cov: linear relationship

Scatter plot

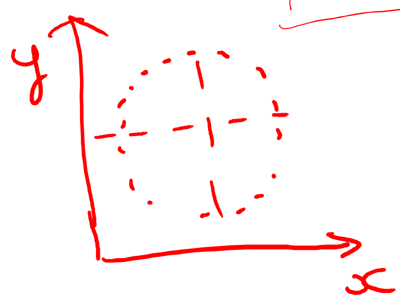


$\sigma_{XY} > 0$

$\sigma_{XY} < 0$



$\sigma_{XY} \overset{?}{\simeq} 0$



$\sigma_{XY} \simeq 0$

No linear rel.

Correlation: $\rho_{XY} = \dfrac{cov(X,Y)}{\sqrt{V(X)V(Y)}} = \dfrac{\sigma_{XY}}{\sigma_X \sigma_Y}$ $\quad -1 \le \rho_{XY} \le 1$

Exercise: Find cor in prev. e.g.

Independence (linear): $\sigma_{XY} = \rho_{XY} = 0$.

Statistics
- Descriptive (subj.) — statistical measures, graphs (visualization
- Inferential (obj.) — statistical & probability theory to test specific hypothesis.

Inferential Statistics: Estimating a statistic - 'point' estimate of a population parameter
& use it for inference

Sample average - 'point' estimate of expected value of the population
it is a random variable & it has a sampling distribution

X is a RV with PPF $f(x)$ with unknown parameter θ

$$\hat{\theta} = h(\underbrace{X_1, X_2, \ldots X_n}_{\text{random sample of size 'n'}})$$ is a point estimator of 'θ'

$\underbrace{X_1, X_2 \ldots X_n}_{\text{Random sample}}$: X_i's are independent, every X_i has same prob. distn. (iid)

Statistic: Any function of the observations in a random sample

Sampling dist. prob distn. of a statistic e.g. $\bar{x}$

CLT - Central Limit Theorem

$X_1, X_2, \ldots X_n$ is a random sample from a population with mean μ and var. σ^2. [no assumption on distribution].

$\bar{X}$ is the mean of the sample is Normally distributed for large enough N, with mean μ and var. $\frac{\sigma^2}{N}$

$\bar{X} \sim N(\mu, \frac{\sigma^2}{N})$

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{N}} \sim N(0, 1)$$

Aside: $X_1, X_2 \ldots X_n \sim N(\mu, \sigma^2)$

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{N}$$

$$\bar{X} \simeq N(\mu, \frac{\sigma^2}{n})$$

$N(\bar{X}, \frac{S^2}{n})$

Eg. (7-2) X has a continuous uniform distribution $f(x) = \begin{cases} 1/2 & 4 \le x \le 6 \\ 0 & \text{o.w} \end{cases}$

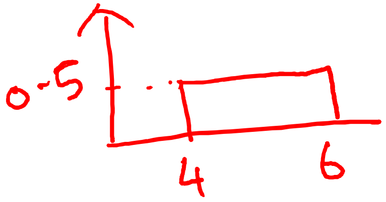
Dist of sample mean from random sample of size 40.

$\mu = 5, \quad \sigma^2 = 1/3$

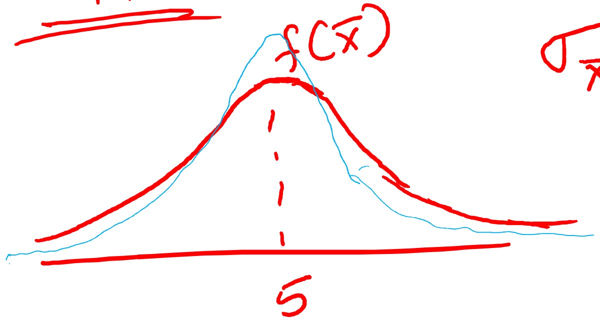
Apply CLT : $\bar{X}$ is approx. normal

$\mu_{\bar{X}} = \mu = 5$

$\sigma_{\bar{X}}^2 = \frac{\sigma^2}{N} = \frac{1/3}{40} = \frac{1}{120}$



40. $\Rightarrow$



$n = 100$

$\mu_{\bar{X}} = 5, \quad \sigma_{\bar{X}}^2 = \frac{1/3}{100} = \frac{1}{300}$

Difference in sample mean : 2 independent populations with means μ_1, μ_2 and variances σ_1^2, σ_2^2

$$\mu_{\bar{X}_1 - \bar{X}_2} = \mu_{\bar{X}_1} - \mu_{\bar{X}_2} = \mu_1 - \mu_2$$

$$\sigma^2_{\bar{X}_1 - \bar{X}_2} = \sigma^2_{\bar{X}_1} + \sigma^2_{\bar{X}_2} = \frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}$$

Extended CLT :
$$z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/N_1 + \sigma_2^2/N_2}} \sim N(0,1) \quad \text{as } \underline{\underline{N}} \to \infty$$

E.g. (7-3)

Old process : effective life $\sim N(5000, 40^2)$ $N_1 = 16$

New process : eff. life $\sim N(5050, 30^2)$ $N_2 = 25$

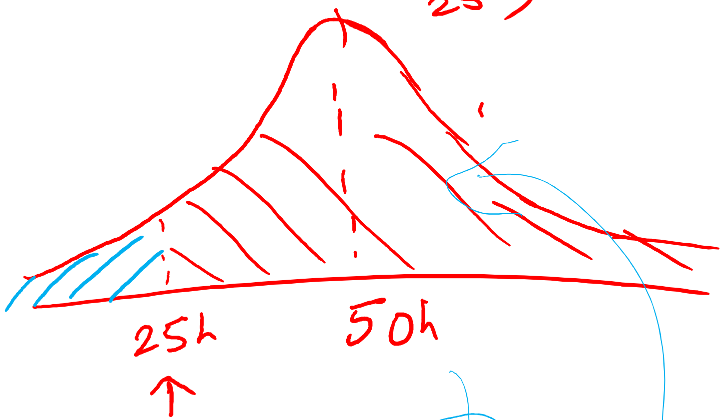
What is the probability that $\bar{X}_2 - \bar{X}_1$ is at least 25 hours.

$$P(\bar{X}_2 - \bar{X}_1 \geq 25\text{h})$$

$$\bar{X}_1 \simeq N\left(5000, \frac{40^2}{16}\right) \simeq N(5000, 10^2)$$

$$\bar{X}_2 \simeq N\left(5050, \frac{30^2}{25}\right) \simeq N(5050, 36)$$

$$\bar{X}_2 - \bar{X}_1 \simeq N(50, 136)$$



$$z = \frac{25-50}{\sqrt{136}} = -2.14$$

$$P(z \geq -2.14) = 1 - \boxed{P(z < -2.14)}$$

$$= 0.984$$

$$1 - \boxed{0.116}$$