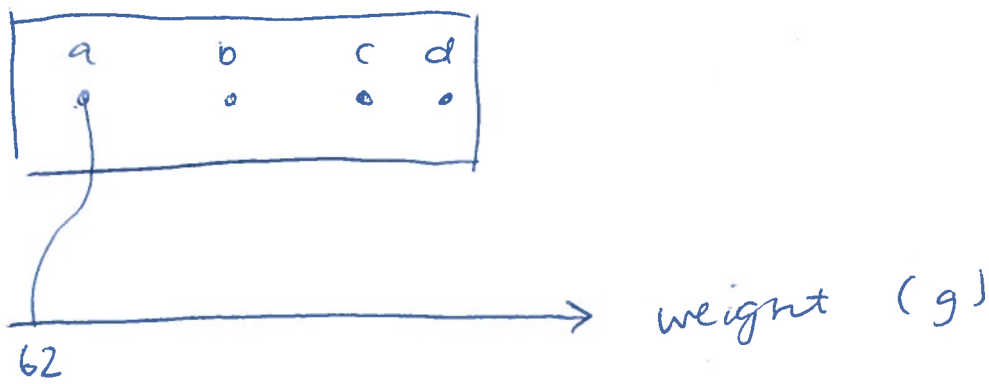
Expt: Pick item at random and record weight.



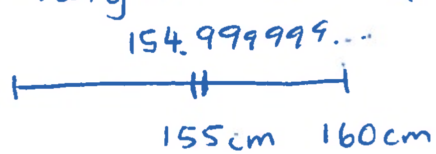
Ch 4 Random Variables and Distributions

Random variable: a function (rule) that assigns a number with each outcome in the sample space, denoted with capital letters, $X, Y, Z...$
The value that a random variable can assume will be denoted with corresponding lower-case letters x, y, z or $x_1, x_2, x_3...$

There are two types of random variables:

- **Discrete** random variables can take on a finite or countable set of values. ex. # courses student takes each semester.
- **Continuous** random variables are defined on a continuous range and can take on an uncountable set of values

exact height of a student.



ex. toss fair coin twice

$X = \#$ heads. $S = \{TT, \dots, HH\}$

T T $\quad$ H T $\quad$ T H $\quad$ H H

$1/4 \quad 1/4 \quad 1/4 \quad 1/4$

$X = 0$

$f(0) = P(X=0) = \frac{1}{4}$

$X = 1$

$P(X=1) = 1/2$

general addn rule.

$X = 2$

$P(X=2) = \frac{1}{4}$

$X_i =$ weight student i.

Discrete Random Variables

The **probability mass function** (pmf) of a discrete random variable X is defined as

$$\boxed{f(x) = P(X = x)}$$

It is a function that gives the probability of occurrence for each possible value x of X. It has the following properties:

1. $$f(x) \geq 0 \text{ for all } x \in X$$

2. $$\sum_{all\ x} f(x) = 1$$

The **cumulative distribution function** of X (cdf) or **distribution function** of X is defined as

$$F(x) = P(X \leq x) = \sum_{k \leq x} f(k) \text{ for all real } x$$

eg. pmf:

X	0	1	2
$P(X=x)$	1/4	1/2	1/4

$F(x) =$
$P(X \leq x)$

$P(X \leq 0) = \frac{1}{4}$

$P(X \leq 1) = \frac{1}{4} + \frac{1}{2} = 3/4$

$P(X \leq 2) = \frac{1}{4} + \frac{1}{2} + \frac{1}{4} = 1$

Example 1

A fair coin that is tossed twice, where X = number of heads.

(a) What is the probability of getting 2 heads? $P(X=2) = 1/4$

(b) What is the probability of getting at least 1 head?

$P(X \geq 1) = P(X=1) + P(X=2) = 3/4$

P(at most 5)

Note: The following table gives a list of some key words you may need to know. Suppose a discrete random variable, X, had possible values of 0 to 5.

Key Word	**Symbol**	**Value for X**
more than 2	$X > 2$	3, 4, 5
no more than 2	$X \leq 2$	$0, 1, 2$
fewer than 2	$X < 2$	$0, 1$
no less than 2	$X \geq 2$	$2, 3, 4, 5$
at least 2	$X \geq 2$	2, 3, 4, 5
at most 2	$X \leq 2$	0, 1, 2
exactly 2	X = 2	2

eg. Random exp.
toss 2 times.

$S = \{HH, HT, TH, TT\}$

X = # heads from tossing 2 times.

Possible values x 0, 1, 2

Run: toss 2 times.	X value
1	1
2	0
3	2
4	2
5	0
6	⋮
⋮	

long run average of values.

mean of X.

Mean and Variance of a Discrete Random Variable

- To find the mean, expectation, expected value.

$$E(X) = \sum_{x \in D} x f(x)$$ where D is the set of possible values

In general,

$$E(g(x)) = \sum_{D} g(x) f(x)$$

- To find the variance,

population variance

$$\sigma^2 = Var(X) = E[(X-\mu)^2] = \sum_{x} (x-\mu)^2 f(x)$$

population mean

true mean

$E(X) = \mu$

Not same $\bar{x}$ ← sample.

S^2 sample variance

$$Var(X) = E(X^2) - [E(X)]^2$$ proof for this later

$E(g(x)) = \sum g(x) f(x)$

$g(x) = (X - \mu)^2$

$E[(X-\mu)^2] = \sum (\cancel{X} - \mu)^2 f(x)$

X = # defective in hour.

a) $E(X)$

Example 2

Suppose a certain production line produces a variable number of defective parts in an hour, with probabilities shown in this table:

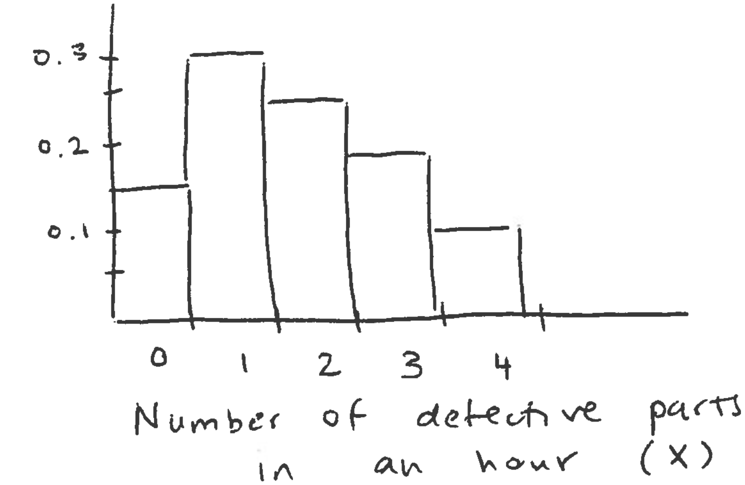
pmf.

X = number of defective parts in an hour	0	1	2	3	4
$P(X = x)$	0.15	0.30	0.25	0.20	0.10

(a) How many defective parts are typically produced in an hour on the production line?

(b) Find the variance, $Var(X)$ and standard deviation, $SD(X)$.

a)



$$\mu = E(X) = \sum x f(x)$$

$$= 0 \times 0.15 + 1 \times 0.30 + \ldots + 4 \times 0.10$$

$$= \underline{\underline{1.8}}$$

long run average.

Hour	# defective
1	0
2	1
3	4
4	2
	⋮

b) $Var(X) = \underline{\underline{E(X^2)}} - \underline{\boxed{E(X)}}^2$

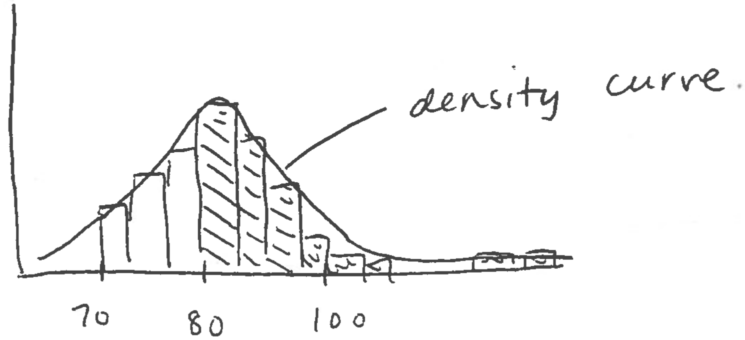
$$E(X^2) = 0^2 \times 0.15 + 1^2 \times 0.3 + 2^2 \times 0.25 + \ldots + 4^2 \times 0.1$$

$$= 4.7$$

$$\sigma^2 = Var(X) = 4.7 - 1.8^2 = 1.46$$

$$\sigma = SD(X) = \sqrt{1.46} = 1.21$$

speed limit...



$P(X > 80)$

What % cars speed?

Continuous Random Variables

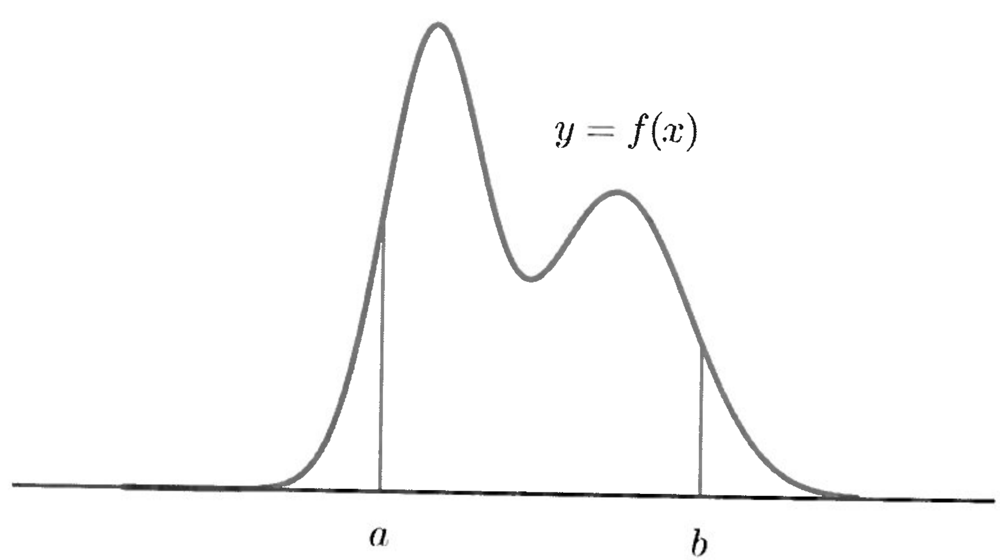
The **probability density function (pdf)** (denoted $f(x)$) is a function that allows us to work out the probability of occurrence over a range of x-values.

$$\boxed{P(a \leq X \leq b) = \int_a^b f(x)dx}$$

$f(x)$ must have the following two properties:

1. $f(x) \geq 0$ for all x
2. $\int_{-\infty}^{+\infty} f(x)dx = 1$

$P(a < X < b) =$ area of shaded region



discrete.

$f(x) = P(X = x)$

continuous:

$P(X = x) = 0$

infinite # of values.

9.99999...

0 10

$$P(X=a) = P(a \le X \le a) = \int_a^a f(x)\,dx = 0$$

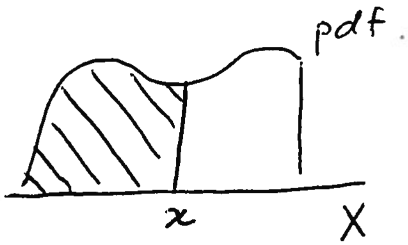
$$P(a \le X \le b) = P(\cancelto{0}{X=a}) + P(\cancelto{0}{X=b}) + P(a < X < b)$$

$$= \underline{\underline{P(a < X < b)}}$$

discrete:

$F(x) = \Sigma f(t)$

The **cumulative distribution function (cdf)** gives the probability of being less than or equal to a particular value



$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt \text{ for all } x$$

Note that: The derivative of the distribution gives the density

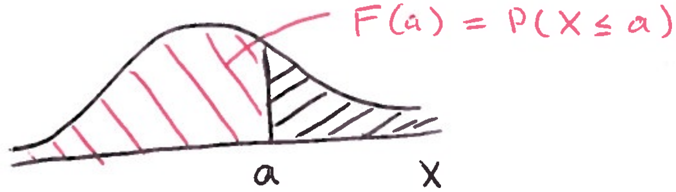
$$F'(x) = f(x)$$

$f(x) \rightarrow F(x)$ by integration and

$F(x) \rightarrow f(x)$ by differentiation

Notice that, in particular,

- $P(X > a) = 1 - F(a)$
- $P(a < X < b) = F(b) - F(a)$



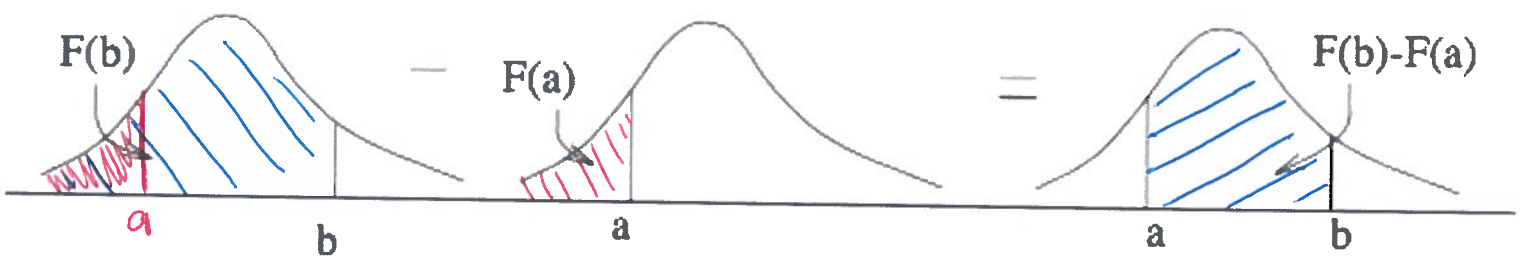


Figure 3.1: Probability on (a, b) under density function $f(x)$

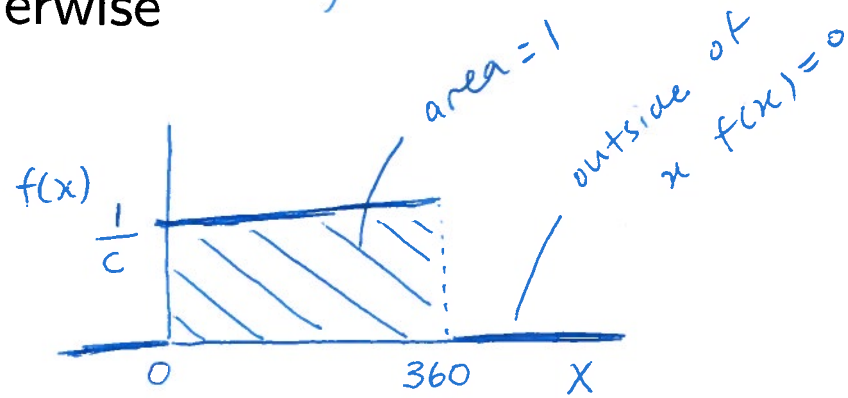
Example 3

$$f(x) = \begin{cases} \frac{1}{c} & \text{if } 0 \le x < 360 \\ 0 & \text{otherwise} \end{cases}$$

support of x

(a) Find the value of c.

(b) Find $P(90 \le X \le 180)$



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

$$= \underbrace{\int_{-\infty}^{0} f(x)\, dx}_{0} + \int_{0}^{360} \frac{1}{c}\, dx + \underbrace{\int_{360}^{\infty} f(x)\, dx}_{0}$$

$$= \int_{0}^{360} \frac{1}{c}\, dx = \frac{1}{c} x \Big|_{0}^{360} = \frac{1}{c}\left[360 - 0\right] = \frac{360}{c}$$

Set $\frac{360}{c} = 1 \rightarrow c = 360$

$$f(x) = \begin{cases} \frac{1}{360} & 0 \leq x \leq 360 \\ 0 & \text{o.w.} \end{cases}$$

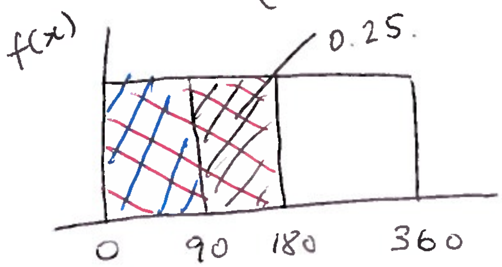
Method 2:

$l \times w = 1$

$360 \times \frac{1}{c} = 1$

$c = 360$

b)



Method 1:

$$P(90 \leq X \leq 180) = \int_{90}^{180} \frac{1}{360}\, dx$$

$$= \frac{1}{360} x \Big|_{90}^{180}$$

$$= \frac{1}{360}[180 - 90] = \frac{1}{4}$$

Method 2:

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

$$F(x) = \int_{-\infty}^{x} f(t)\, dt. \quad \text{(definition cdf continuous)}$$

$$F(x) = \int_{0}^{x} \frac{1}{360}\, dt = \frac{1}{360} t \Big|_{0}^{x} = \frac{x}{360}$$

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{360} & 0 \leq x < 360 \\ 1 & x \geq 360 \end{cases}$$

$P(X \leq x)$

$P(X \leq 361) = 1$

$$P(90 \le X \le 180) = F(180) - F(90)$$

$$= \frac{180}{360} - \frac{90}{360} = \frac{1}{4}$$

//

Last Class:

Discrete random variables | **Continuous r.v.**

- countable set values

eg. # days in month

pmf:
$f(x) = P(X = x)$

Properties:

1) $f(x) \geq 0$ for all x
2) $\sum_x f(x) = 1$

cdf:
$$F(x) = P(X \leq x) = \sum_{k \leq x} f(k)$$

$P(X < 2)$

$P(X \leq 2)$

- uncountable set of values

eg. time required to run 1km.

pdf:
$$P(a < X < b) = \int_a^b f(x)\,dx$$

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

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt$$

$P(X = a) = 0$

$P(X \leq a) = P(X < a)$

Example 4

pdf

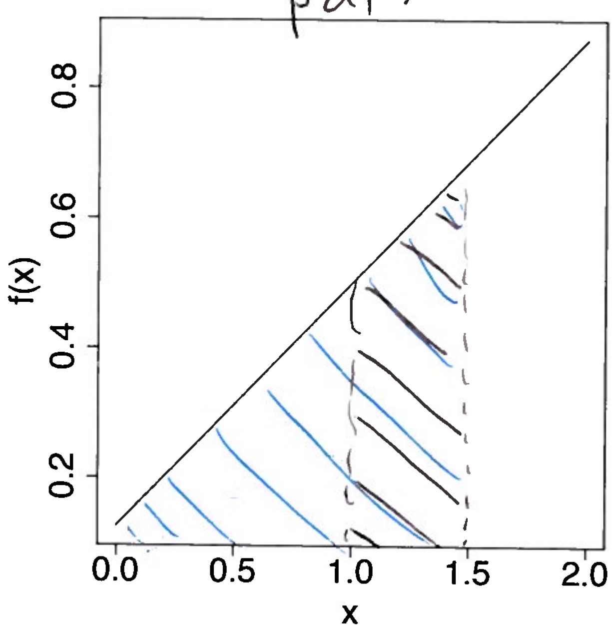
$$f(x) = \begin{cases} \frac{1}{8} + \frac{3}{8}x & 0 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find the cdf, $F(x)$

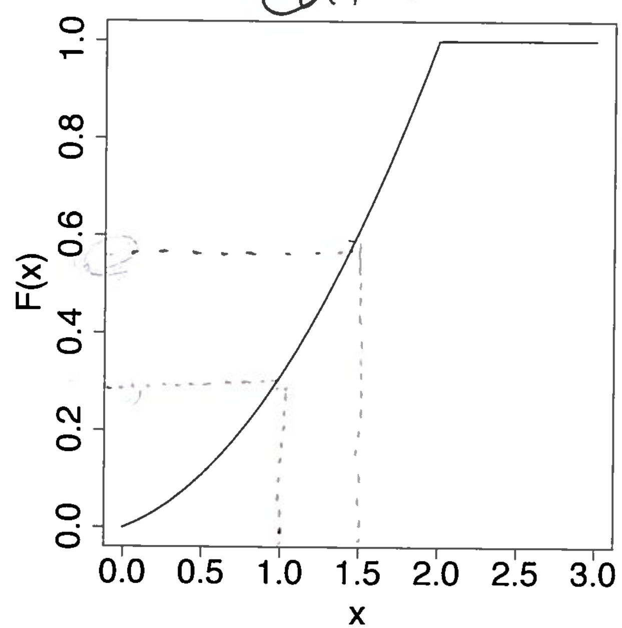
(b) Use $F(x)$ to find $P(1 \le X \le 1.5)$

(c) Find $P(X > 1)$

pdf.



cdf.



$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\, dt.$$

$$= \int_0^x \frac{1}{8} + \frac{3}{8} t \quad dt$$

$$= \frac{1}{8} t + \frac{3}{8} \frac{t^2}{2} \Big|_0^x$$

$$= \frac{x}{8} + \frac{3}{16} x^2$$

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{8} + \frac{3}{16} x^2 & 0 \le x < 2 \\ 1 & x \geq 2 \end{cases}$$

$P(X \le x)$

$f(x) \to F(x)$

$F(x) \to f(x)$

Notice:

$$f(x) = F'(x)$$

$$= \frac{d}{dx}\left(\frac{x}{8} + \frac{3}{16} x^2 \right) = \frac{1}{8} + \frac{3}{8} x$$

$$f(x) = \begin{cases} \frac{1}{8} + \frac{3}{8} x & 0 \le x \le 2 \\ 0 & \text{otherwise.} \end{cases}$$

$$P(1 \le X \le 1.5) = F(1.5) - F(1)$$

$$= \left(\frac{1.5}{8} + \frac{3}{16} 1.5^2\right) - \left(\frac{1}{8} + \frac{3}{16} 1^2\right)$$

$$= 0.2968$$

c) $P(X > 1) = 1 - P(X \leq 1)$

$$= 1 - F(1)$$

$$= 1 - \left(\frac{1}{8} + \frac{3}{16} 1^2 \right) = \frac{11}{16} = 0.6875.$$

How to find median, Q_1, Q_3, IQR from $f(x)$
Steps to find the median:

(a) Find $F(x)$

(b) then solve for x such that $F(x) = 0.5$
x is the median

Q_1

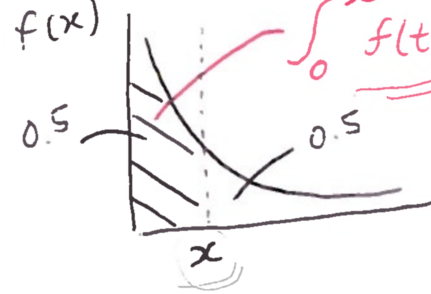
$F(x) = 0.25$

$F(x) = 0.75$

- To find Q_1 and Q_3, do the same as above but instead of 0.5, use 0.25 and 0.75 for Q_1 and Q_3 respectively.
- To find IQR, use $IQR = Q_3 - Q_1$

Example 5

$$f(x) = \begin{cases} 2e^{-2x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$



$\int_0^x f(t)\,dt = \frac{1}{2}$

Find the median, Q_1, Q_3 and IQR.

Step 1:

a) $F(x) = \int_0^x 2e^{-2t}\,dt = -\frac{2}{2} e^{-2t} \Big|_0^x$

$= -e^{-2x} + 1$

$= 1 - e^{-2x}$

Recall:

$\int e^{ax}\,dx = \frac{1}{a} e^{ax}$

Step 2: Set $F(x) = 0.5$

$1 - e^{-2x} = 0.5$

$e^{-2x} = 0.5$

$\ln(e^{-2x}) = \ln(0.5)$

$-2x = \ln(0.5)$

$x = 0.347.$

Check: Exercise:

$Q_1 = 0.144$ $IQR = Q_3 - Q_1 = 0.549$

$Q_3 = 0.693$

Mean and Variance of a Continuous Random Variable

- To find the mean,

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

In general,

$$E(g(x)) = \int_{-\infty}^{\infty} g(x) f(x) dx$$

- To find the variance,

$g(x) = (X - \mu)^2$

$$Var(X) = E[(X - \mu)^2] \qquad E(g(x))$$
$$= \int (x - \mu)^2 f(x) dx$$
$$Var(X) = E(X^2) - [E(X)]^2$$

Example 6

Find the mean and standard deviation of

$$f(x) = \begin{cases} 2x & 0 \leq x \leq 1 \\ 0 & otherwise \end{cases}$$

$$E(X) = \int_0^1 x \cdot 2x \, dx = \int_0^1 2x^2 \, dx = 2\frac{x^3}{3}\Big|_0^1 = \underline{\underline{2/3}}$$

$$Var(X) = \underline{E(X^2) - E(X)^2}$$

$$E(X^2) = \int_0^1 x^2 \cdot 2x = \int_0^1 2x^3 \, dx = \frac{2x^4}{4}\Big|_0^1 = \frac{1}{2}$$

$$Var(X) = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{18}$$

$$SD(X) = \sqrt{1/18} = 0.236$$

Properties of the Mean and Variance

Stretch or shrinking by "a" units

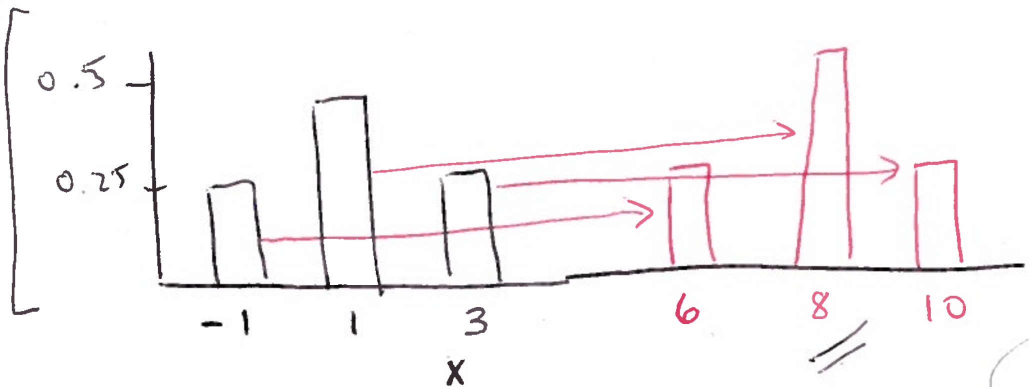
1. $E(aX \pm b) = aE(X) \pm b$ — shifting distribution "b" units. where a and b are constants
2. $Var(aX \pm b) = a^2Var(X)$ where a and b are constants
3. $E(X \pm Y) = E(X) \pm E(Y)$ where X and Y are random variables
4. $E(XY) = E(X)E(Y)$ where X and Y are *independent* random variables
5. If X and Y are independent random variables,
   $Var(X + Y) = Var(X) + Var(Y)$
   $Var(X - Y) = Var(X) + Var(Y)$

$E(X) = 1$

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

$= 2$

$Y = X + 7$



$E(Y) = 6 \times 0.25 + 8 \times 0.5 + 10 \times 0.25 = 8$

$E(X+7) = E(X) + 7 = 8$

$Var(Y) = Var(X)$

$E(2X) = E(X) \times 2 = 2$

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

$E(2X) = -2 \times 0.25 + 2 \times 0.5 + 6 \times 0.25 = 2$

1. $E(aX \pm b) = aE(X) \pm b$

Recall: discrete r.v.

$$E[g(X)] = \sum_x g(x) P(X=x)$$

$$E(\underbrace{aX+b}_{g(X)}) = \sum_x (aX+b) P(X=x)$$

$$= a \underbrace{\sum_x x P(X=x)} + b \underbrace{\sum_x P(X=x)}_{=1}$$

$$= a E(X) + b$$

2. $Var(aX+b) = a^2 Var(X)$

Recall: discrete.

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

$$= \sum_x (x-\mu)^2 P(X=x)$$

$$Var(aX+b) = E[((aX+b) - \underbrace{E(aX+b)}_{aE(X)+b \text{ (1)}})^2]$$

$$= E[(aX + \cancel{b} - a\underbrace{E(X)}_{\mu} - \cancel{b})^2]$$

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

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

$$= a^2 Var(X).$$

5 $\underline{Var(X-Y)} = Var(X) + Var(Y)$

want to show ↑

$\mu_X = E(X)$

$\mu_Y = E(Y)$

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

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

$$= E[((X-Y) - (\mu_X - \mu_Y))^2]$$

$$= E[((\underbrace{X-\mu_X}_{a}) - (\underbrace{Y-\mu_Y}_{b}))^2]$$

$$(a-b)^2 = a^2 - 2ab + b^2$$

$$(*) = E[(X-\mu_X)^2 \underbrace{- 2(X-\mu_X)(Y-\mu_Y)}_{0} + (Y-\mu_Y)^2]$$

$$-2E[(X-\mu_X)(Y-\mu_Y)] = -2E[XY - X\mu_Y - Y\mu_X + \mu_X\mu_Y]$$

$$= -2[E(XY) - E(X\mu_Y) - E(Y\mu_X) + E(\mu_X\mu_Y)]$$

$$= -2[\mu_X\mu_Y - \mu_Y E(X) - \mu_X E(Y) + \mu_X\mu_Y]$$

X, Y independent ↑

$$= -2[\cancel{\mu_X}\mu_Y - \mu_Y\cancel{\mu_X} - \mu_X\cancel{\mu_Y} + \mu_X\cancel{\mu_Y}]$$

$$= 0$$

(*) $= E[(X-\mu_X)^2 + (Y-\mu_Y)^2]$

$= E[(X-\mu_X)^2] + E[(Y-\mu_Y)^2] = \underline{\underline{Var(X) + Var(Y)}}$ //

$Var(X) = E(X^2) - \mu^2$ want to show.

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

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

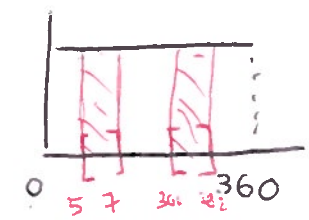
$= E[X^2] - 2\mu \underbrace{E(X)}_{\mu} + \mu^2 = E[X^2] - 2\mu^2 + \mu^2$

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

Some Continuous Models

Uniform Random Variables

If X is a uniform random variable, $X \sim U(a, b)$, X is evenly distributed over the interval $[a, b]$.

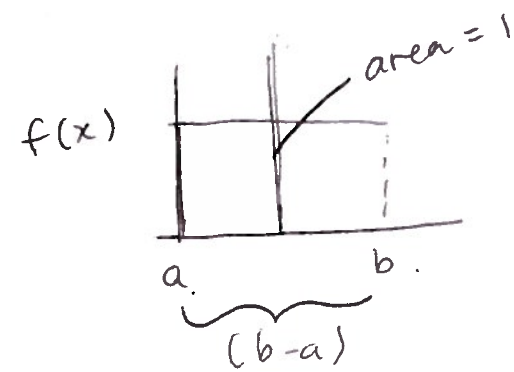


pdf:

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

$$E(X) = \frac{a+b}{2}$$

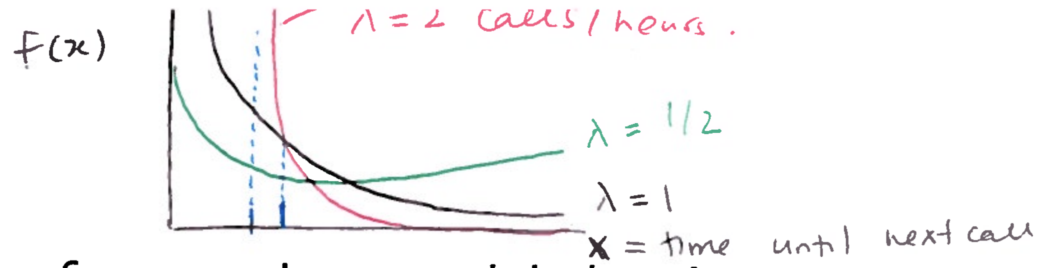
$$Var(X) = \frac{(b-a)^2}{12}$$



$$F(x) = \frac{x-a}{b-a}$$

$$(b-a) f(x) = 1$$

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



Exponential Random Variables

Exponential random variables are often used to model the time until an event occurs.

X = time until next phone call occurs.
λ = rate of incoming phone calls.

If X is an exponential random variable with rate of λ, $X \sim \exp(\lambda)$, the pdf is

rate parameter.

pdf: $$f(x) = \lambda e^{-\lambda x}, \qquad x \geq 0$$

where λ is a positive constant and is the reciprocal of the mean lifetime.

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

$$Var(X) = \frac{1}{\lambda^2}$$

Last Class:

Properties of Mean + Variance:

1. $E(aX \pm b) = aE(X) \pm b$
2. $Var(aX \pm b) = a^2 Var(X)$
3. $E(X \pm Y) = E(X \pm Y)$
4. $E(XY) = E(X)E(Y)$ if X, Y independent.
5. $Var(X \pm Y) = Var(X) + Var(Y)$ X, Y independent.

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

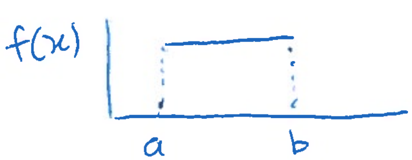
Uniform r.v.

$X \sim U(a, b)$ parameters

$$f(x) = \frac{1}{b-a} \qquad a \le x \le b$$

$$E(X) = \frac{a+b}{2}$$

$$Var(X) = \frac{(b-a)^2}{12}$$



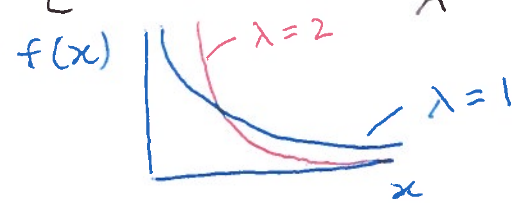
Exponential r.v.

$X \sim exp(\lambda)$ parameter

$$f(x) = \lambda e^{-\lambda x} \qquad x \geq 0.$$

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

$$Var(X) = \frac{1}{\lambda^2}$$



$$\boxed{E(X) = \frac{1}{\lambda}}$$ prove.

Integration by parts!

$u = x$

$du = dx$

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

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

$uv - \int v\,du$

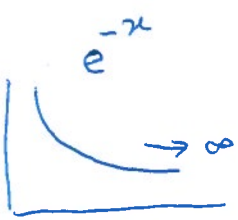
$$E(X) = \int_{-\infty}^{\infty} x f(x)\,dx$$

$$= \int_0^{\infty} x \cdot \lambda e^{-\lambda x}\,dx$$

$$= \lambda \int_0^{\infty} x \cdot e^{-\lambda x}\,dx$$

$$= \cancel{\lambda}\left[\underbrace{x}_{u} \cdot \underbrace{\left(-\frac{1}{\cancel{\lambda}}e^{-\lambda x}\right)}_{v}\Bigg|_0^{\infty} + \int \underbrace{\frac{1}{\cancel{\lambda}}e^{-\lambda x}}_{v}\,\underbrace{dx}_{du}\right]$$

$$= -\overset{0}{\cancel{x e^{-\lambda x}}}\Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda x}\,dx$$



$$= -\frac{e^{-\lambda x}}{\lambda}\Bigg|_0^{\infty} = +\frac{e^{0}}{\lambda} = \frac{1}{\lambda} \quad //$$

X continuous r.v. density $f_X(x)$

cdf: $F_X(x)$

eg. $Y = g(X)$

eg. $Y = X^2$

$Y = e^X$

input $\quad$ output

$$X \longrightarrow \boxed{g(x)} \longrightarrow Y$$

want to find density $f_Y(y)$ and cdf $F_Y(y)$

Problem 4.13 from the course notes

Example 7

Consider a random variable X which follows the uniform distribution on the interval $(0, 1)$.

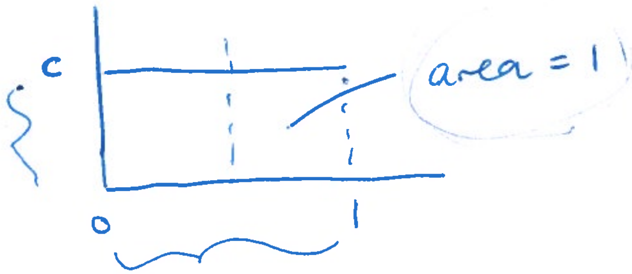
(a) Give the density function $f(x)$ and obtain the cumulative distribution function $F(x)$ of X;

(b) Calculate the mean (expectation) $E(X)$ and variance $Var(X)$;

(c) Let $Y = \sqrt{X}$. Find the E(Y) and Var(Y);

(d) Obtain the distribution function $G(y)$ and furthermore the density function $g(y)$ of random variable Y.

a) $X \sim U(0,1)$



$l \times w = \text{area}.$

$c \times 1 = 1$

$c = 1$

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

$= \frac{1}{1-0}$

$= 1 \qquad 0 \le x \le 1$

$F_X(x) = \int_0^x 1 \, dt = t \Big|_0^x = x$

$P(X \le x)$

$$F_X(x) = \begin{cases} 0 & x < 0 \\ x & 0 \le x < 1 \\ 1 & x \ge 1 \end{cases}$$

b) $E(X) = \frac{a+b}{2} = \frac{0+1}{2} = \frac{1}{2}$

$\mathrm{Var}(X) = \frac{(b-a)^2}{12} = \frac{(1-0)^2}{12} = \frac{1}{12}$

c) $\mathrm{Var}(Y)$, $E(Y)$ $\qquad Y = \sqrt{X}$

$E(g(X)) = \int g(x) f(x) \, dx$

$E(Y) = E(\sqrt{X}) = \int_0^1 \sqrt{x} \cdot 1 = \int_0^1 x^{1/2} = \frac{2}{3} x^{3/2} \Big|_0^1 = 2/3$

$\mathrm{Var}(Y) = E(Y^2) - E(Y)^2$

$$E(Y^2) = \int_0^1 x \cdot 1 \, dx = \frac{x^2}{2}\Big|_0^1 = \frac{1}{2}$$

$Y = \sqrt{X}$

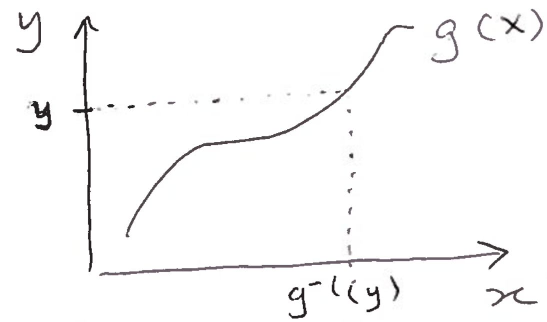
$Y^2 = X$

$$\mathrm{Var}(Y) = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{18}$$

---

want: pdf and cdf of Y

have: pdf + cdf of X.



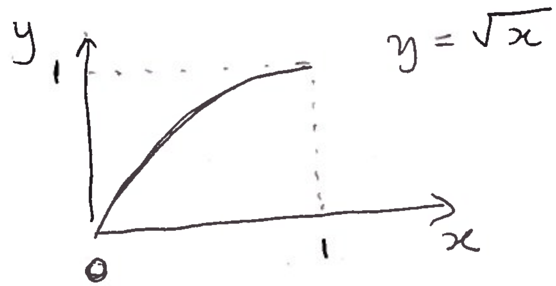
want $\rightarrow$ $F_Y(y) = P(Y \le y)$

$= P(X \le g^{-1}(y))$

$= F_X(g^{-1}(y))$

d) $G_Y(y) = P(Y \le y)$

$Y = \sqrt{X}$



support of x: $0 \le x \le 1$

support of y: $0 \le x \le 1$

d) $X \sim U(0,1)$

$$f_X(x) = \begin{cases} 1 & 0 \le x \le 1 \\ 0 & \text{o.w.} \end{cases}$$

$$F_X(x) = \begin{cases} 0 & x < 0 \\ x & 0 \le x < 1 \\ 1 & x \ge 1 \end{cases}$$

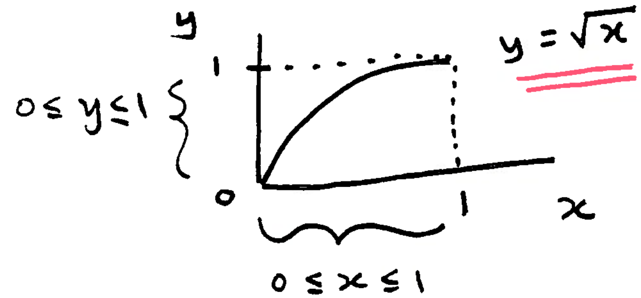
$Y = \sqrt{X}$ want to find

$G_Y(y)$ cdf

$g_Y(y)$ pdf

Idea: Find cdf

$$G_Y(y) = P(Y \le y) \xrightarrow{\text{derive}} g_Y(y)$$



support x: $0 \le x < 1$

support y: $0 \le y < 1$

$$G_Y(y) = P(Y \le y) = P(\sqrt{X} \le y) = P(X \le y^2)$$
$$= F_X(y^2)$$

recall:
$P(X \le x) = F_X(x)$

$$G_Y(y) = F_X(y^2) = \begin{cases} 0 & y < 0 \\ y^2 & 0 \le y < 1 \\ 1 & y \ge 1 \end{cases}$$

$$g_Y(y) = \frac{d}{dy} G_Y(y) = \frac{d}{dy}(y^2) = 2y$$

$$g_Y(y) = \begin{cases} 2y & 0 \le y \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

$$E(Y) = \int_{-\infty}^{\infty} y \cdot g(y)\, dy$$

$E(Y) = 2/3$, $Var(Y) = 1/18$ } solved part. (b)

check:

$$E(Y) = \int_0^1 y \cdot 2y \, dy = \frac{2}{3} y^3 \Big|_0^1 = \frac{2}{3} \checkmark$$

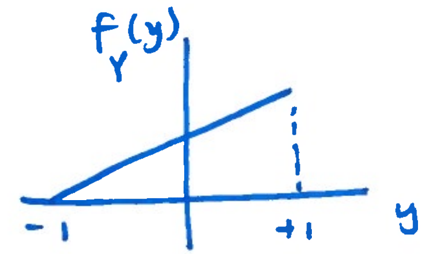
$$E(Y^2) = \int_0^1 y^2 \cdot 2y \, dy = \frac{2y^4}{4} \Big|_0^1 = \frac{1}{2}$$

$$Var(Y) = E(Y^2) - E(Y)^2 = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{18} \checkmark$$

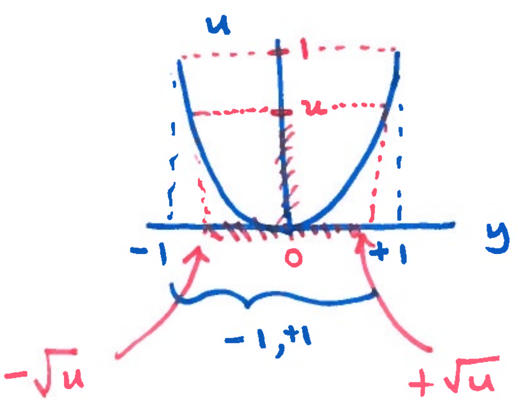
match part (b)

Example 8

$$f_Y(y) = \begin{cases} \frac{y+1}{2} & -1 \le y \le 1 \\ 0 & otherwise \end{cases}$$



Find the pdf of $U = Y^2$.



$$\left.\begin{array}{l} -1 \le y \le 1 \\ 0 \le y^2 \le 1 \\ 0 \le u \le 1 \end{array}\right\} \text{support.}$$

$Y_1 = -\sqrt{u} \qquad -1 \le y < 0$

$Y_2 = +\sqrt{u} \qquad 0 < y < 1$

$$\begin{aligned} \Big[F_U(u) &= P(U \le u) \\ &= P(Y^2 \le u) \\ &= P(-\sqrt{u} \le Y \le \sqrt{u}) \end{aligned}$$

method A :

cdf method

method B.

integration method.

method A:

$$F_Y(y) = P(Y \le y)$$

$$= \int_{-1}^{y} \frac{t+1}{2}\, dt = \frac{t^2}{4} + \frac{1}{2}t \Big|_{-1}^{y} = \left(\frac{y^2}{4} + \frac{y}{2}\right) - \left(\frac{1}{4} - \frac{1}{2}\right)$$

$$= \frac{y^2 + 2y + 1}{4}$$

$$F_U(u) = P(-\sqrt{u} \le Y \le +\sqrt{u})$$

$$= F_Y(\sqrt{u}) - F_Y(-\sqrt{u})$$

$$= \frac{\cancel{u} + 2\sqrt{u} + \cancel{1}}{4} - \left[\frac{\cancel{u} - 2\sqrt{u} + \cancel{1}}{4}\right]$$

$$= \frac{4\sqrt{u}}{4} = \sqrt{u}$$

method B:

$$F_U(u) = \int_{-\sqrt{u}}^{\sqrt{u}} f(y)\, dy = \int_{-\sqrt{u}}^{\sqrt{u}} \frac{y+1}{2}\, dy = \frac{1}{2}\left[\frac{y^2}{2} + y\right]_{-\sqrt{u}}^{\sqrt{u}}$$

$$= \frac{1}{2}\left[\frac{u}{\cancel{2}} + \sqrt{u} - \left(\frac{u}{\cancel{2}} - \sqrt{u}\right)\right] = \frac{1}{2}\left[\sqrt{u} + \sqrt{u}\right] = \sqrt{u}.$$

method A and B produce same result.

$$F_U(u) = \begin{cases} 0 & u < 0 \\ \sqrt{u} & 0 \le u < 1 \\ 1 & u \geq 1 \end{cases} \xrightarrow{\text{derivative.}} f_U(u) = \begin{cases} \frac{1}{2} u^{-\frac{1}{2}} & 0 \le u \le 1 \\ 0 & \text{o.w.} \end{cases}$$

$$Var(aX) = a^2 Var(X)$$

Sum of Independent Random Variables

Random experiments are often independently repeated creating a sequence of n independent random variables (e.g. roll a die repeatedly, measure the lifetime of a component repeatedly).

If $X_1, X_2, X_3, ..., X_n$ are n independent random variables and $Y = a_1X_1 + a_2X_2 + ... + a_nX_n$ where $a_1, a_2, ..., a_n$ are constants,

$$E(Y) = a_1E(X_1) + a_2E(X_2) + ... + a_nE(X_n)$$
$$Var(Y) = a_1^2Var(X_1) + a_2^2Var(X_2) + ... + a_n^2Var(X_n)$$

If the n random variables X_i have a common mean μ and common variance σ^2. We call $\{X_1, ..., X_n\}$ a random sample and we get:

$$E(Y) = (a_1 + a_2 + ... + a_n)\mu$$
$$Var(Y) = (a_1^2 + a_2^2 + ... + a_n^2)\sigma^2$$

Average of Independent Random Variables

If $X_1, X_2, ..., X_n$ are n independent random variables and $\bar{X} = \frac{X_1+X_2+...+X_n}{n}$

$$E(\bar{X}) = \frac{1}{n}[E(X_1) + E(X_2) + ... + E(X_n)]$$
$$Var(\bar{X}) = \frac{1}{n^2}[Var(X_1) + Var(X_2) + ...Var(X_n)]$$

If the n random variables X_i have a common mean μ and common variance σ^2.

$$E(\bar{X}) = \mu$$
$$Var(\bar{X}) = \frac{\sigma^2}{n}$$

$$E(\bar{X}) = E\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{1}{n} E(X_1 + \dots + X_n) = \frac{1}{n}[E(X_1) + \dots + E(X_n)]$$

$$Var(\bar{X}) = Var\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{1}{n^2}(Var(X_1) + \dots + Var(X_n)) \circledast$$

common mean μ. $E(X_i) = \mu$ $i = 1, \dots, n$

" " variance σ^2 $Var(X_i) = \sigma^2$

$$E(\bar{X}) = \frac{1}{n}[E(X_1) + \dots + E(X_n)] = \frac{1}{n}[\underbrace{\mu + \mu + \dots + \mu}_{n \text{ times}}] = \frac{n\mu}{n} = \mu$$

$$\circledast\ Var(\bar{X}) = \frac{1}{n^2}(Var(X_1) + \dots + Var(X_n))$$

$$= \frac{1}{n^2}(\sigma^2 + \dots + \sigma^2) = \frac{\cancel{n}\sigma^2}{n^{\cancel{2}}} = \frac{\sigma^2}{n}$$

$S = X_1 + \dots + X_5$

$E(S) = E(X_1) + \dots + E(X_5) = 5\,E(X_i) = 5 \times 7/2 = 35/2$

1	2	...	6.
$\frac{1}{6}$			1/6.

$E(X_i) = 7/2$

Last Class:

Sum independent r.v.

$Y = a_1X_1 + \dots + a_nX_n$

$E(Y) = E(a_1X_1 + \dots + a_nX_n)$

$= a_1E(X_1) + a_2E(X_2) + \dots + a_nE(X_n)$

$Var(Y) =$

$Var(a_1X_1 + \dots + a_nX_n)$

$= a_1^2 Var(X_1) + \dots + a_n^2 Var(X_n)$

if X_i have common mean and variance

$E(Y) = (a_1 + \dots + a_n)\mu$

$Var(Y) = (a_1^2 + \dots + a_n^2)\sigma^2$

Average of Independent r.v.

$$\bar{X} = \frac{X_1 + \dots + X_n}{n}$$

$$E(\bar{X}) = \frac{1}{n}\left(E(X_1) + \dots + E(X_n)\right)$$

$$Var(\bar{X}) = \frac{1}{n^2}\left(Var(X_1) + \dots + Var(X_n)\right)$$

if X_i have common mean and variance

$E(\bar{X}) = \mu$

$$Var(\bar{X}) = \frac{\sigma^2}{n}$$

Maximum of Independent Random Variables

There are cases when we are interested in the maximum or minimum of a random sample.

For instance, the maximum can be used to model:

- The lifetime of a system of n independent components connected in parallel,
- The completion time of a project of n independent subprojects, which can be completed simultaneously.

We will work on ex 4.8 and 4.9 from the course text.

Example 9

A system consists of five components connected in parallel. The lifetime (in thousands of hours) of each component is an exponential random variable with mean $\mu = 3$.

(a) Calculate the median and standard deviation for each component

(b) Calculate the probability that a component fails before 3500 hours.

(c) Calculate the probability that the system will fail before 3500 hours. Compare this with the probability that a component fails before 3500 hours.

(d) Calculate the median life for the system.

$Y = \max(X_1, \dots, X_n)$ where $X_1, \dots, X_n$ independent.

↳ pdf

9) Let X_i be lifetime of each component.

$X_i \sim \exp(\lambda = \frac{1}{3})$

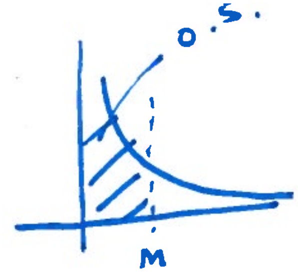
$E(X) = \frac{1}{\lambda} = 3 \Rightarrow \lambda = \frac{1}{3}$. $\quad f(x) = \frac{1}{3} e^{-\frac{1}{3}x} \quad x \geq 0$

lifetime system continues until last component fails.

$Y = \max(X_1, \dots, X_5)$ lifetime system.

a) $X \sim \exp(\lambda = \frac{1}{3})$

$Var(X) = \frac{1}{\lambda^2} = 9$ $\qquad SD(X) = 3$



median

$F_X(x) = \int_0^x \frac{1}{3} e^{-\frac{1}{3}t}\,dt = -\frac{1}{3}\,3\,e^{-1/3\,t}\Big|_0^x$

$= -e^{-1/3\,t}\Big|_0^x = 1 - e^{-x/3}$

$F_X(x) = 0.5$

$1 - e^{-x/3} = 0.5$

$e^{-x/3} = 0.5$

$-\frac{x}{3} = \ln(0.5)$

$x = 2.08.$

b) $P(X \leq 3.5) = F_X(3.5) = 1 - e^{-\frac{1}{3}(3.5)} = 0.6886$

c) $Y = \max(X_1, X_2, \ldots, X_5)$

To find pdf, find cdf.

$$F_Y(y) = P(Y \leq y)$$
$$= P(\{X_1 \leq y\} \cap \{X_2 \leq y\} \cap \{\ldots \cap \{X_5 \leq y\})$$
$$= P(X_1 \leq y)\, P(X_2 \leq y) \ldots P(X_5 \leq y) \qquad X_i \text{ independent}$$
$$= [P(X \leq y)]^5 \qquad X_i \text{ identically distributed.}$$
$$= [F_X(y)]^5$$

$$F_Y(\overset{y}{\cancel{3.5}}) = \left[1 - e^{-1/3\, y}\right]^5$$

$$F_Y(3.5) = \left[1 - e^{-1/3 \cdot 3.5}\right]^5 = 0.1548$$

d) $F_Y(y) = 0.5$

$$\left(1 - e^{-1/3 \cdot \overset{y}{\cancel{3.5}}}\right)^5 = 0.5$$
$$1 - e^{-y/3} = 0.5^{1/5}$$
$$1 - 0.5^{1/5} = e^{-y/3}$$
$$\ln(1 - 0.5^{1/5}) = -y/3.$$
$$y = 6.133.$$

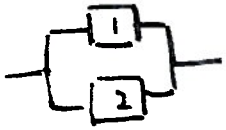
Minimum of Independent Random Variable

The minimum can be used to model:

- The lifetime of a system of n independent components connected in series,
- The completion time of a project pursued by n independent competing teams

Last Class:

Maximum of Independent r.v.



eg. independent components in parallel.

Setup: Given pdf of n independent r.v. $X_1, \dots, X_n$

Question: Find the pdf of the maximum

Steps:

$Y = \max(X_1, \dots, X_n)$

Thought: If y is max. then each of $X_1, \dots, X_n$ are $\le y$.

$$F_Y(y) = P(Y \le y)$$

and

$$= P(X_1 \le y, X_2 \le y, \dots, X_n \le y)$$

$$= P(X_1 \le y) \cdots P(X_n \le y)$$

X_is independent.

$$= F_{X_1}(y)\, F_{X_2}(y) \cdots F_{X_n}(y)$$

If X_is have same pdf

$$= [F_X(y)]^n$$

pdf: $f_Y(y) = F_Y'(y)$

$$= n\, F_X(y)^{n-1} f_X(y)$$

Minimum of Indep. r.v.

eg.

indep. components in series

Q: Find pdf of min.

Steps:

$Y = \min(X_1, \dots, X_n)$

$$F_Y(y) = P(Y \le y)$$

$$= 1 - P(Y > y)$$

$$= 1 - P(X_1 > y, \dots, X_n > y)$$

$$= 1 - P(X_1 > y) \cdots P(X_n > y)$$

X_is indep.

$$= 1 - [1 - F_{X_1}(y)] \cdots [1 - F_{X_n}(y)]$$

If X_is have same pdf.

$$= 1 - [1 - F_X(y)]^n$$

pdf: $f_Y(y) = F_Y'(y)$

$$= -n[1 - F_X(y)]^{n-1}(-f_X(y))$$

Example 10

A system consists of five components connected in series. The lifetime (in thousands of hours) of each component is an exponential random variable with mean $\mu = 3$.

(a) Calculate the probability that the system fails before 3500 hours. Compare this with the probability that a component fails before 3500 hours.

(b) Calculate the median life, mean life and standard deviation for the system.

10) —[1]—[2]—[3]—[4]—[5]— If any one component fails, system fails.

$$X \sim \exp(\underset{\swarrow}{\lambda} = \tfrac{1}{3}) \quad \text{parameter.}$$

$$E(X) = \frac{1}{\lambda} = 3$$

$f(x) = \frac{1}{3} e^{-x/3}$ $x \geq 0$ pdf

$F(x) = 1 - e^{-x/3}$ cdf.

$$Y = \min(X_1, X_2, \dots, X_5)$$

a) $P(Y \leq 3.5) = F_Y(3.5)$

$$\begin{aligned}
F_Y(y) &= P(Y \leq y) \\
&= 1 - P(Y > y) \\
&= 1 - P(\{X_1 > y\} \cap \{X_2 > y\} \cap \dots \cap \{X_5 > y\}) \\
&= 1 - [P(X > y)]^5 \\
&= 1 - [1 - P(X \leq y)]^5 \\
&= 1 - [\cancel{1} - (\cancel{1} - e^{-y/3})]^5 \\
&= 1 - (e^{-y/3})^5 = 1 - e^{-y\,5/3}
\end{aligned}$$

$$F_Y(y) = \underbrace{1 - e^{-5/3 \cdot y}}_{\text{cdf of an exponential}}$$

$$f_Y(y) = -e^{-5/3\, y}\left(-\frac{5}{3}\right) = \frac{5}{3} e^{-5/3\, y}$$

min of exp. r.v. has an exponential distribution as well.

Notice: previous example. max does not.

$$Y \sim \exp\left(\lambda = \frac{5}{3}\right)$$

$$E(Y) = \frac{1}{\lambda} \quad \text{for } Y \sim \exp(\lambda)$$

$$E(Y) = \frac{3}{5}$$

$$\text{Var}(Y) = \frac{1}{\lambda^2} = \frac{1}{(5/3)^2} = \frac{9}{25} \qquad SD(Y) = \frac{3}{5}$$

median:

$$F_Y(y) = 0.5$$

$$1 - e^{-5/3\, y} = 0.5$$

$$0.5 = e^{-5/3\, y}$$

$$\ln(0.5) = -5/3\, y$$

$$y = -\frac{3}{5}\ln(0.5)$$

$$= 0.416$$

Problem 4.22 A system has two independent components A and B connected in parallel. If the operational life (in thousand of hours) of each component is a random variable with density

$$f(x) = \frac{1}{36}(x-4)(10-x) \qquad 4 < x < 10$$
$$= 0 \qquad \text{otherwise}$$

(a) Find the median and the mean life of each component. Find also the standard deviation and IQR.
(b) Calculate the distribution and density functions for the lifetime of the system. What is the expected lifetime of the system?

(c) Same as (b) but assuming that the components are connected "in series" instead of "in parallel".

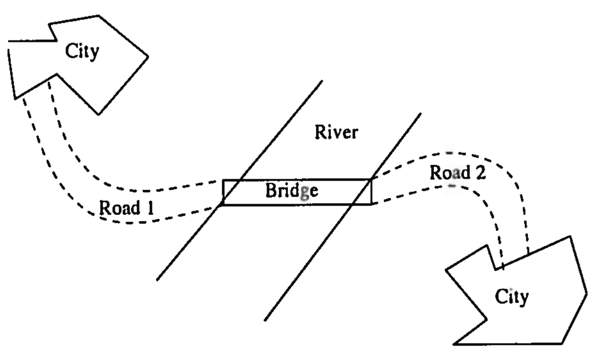
Problem 4.23 A large construction project consists of building a bridge and two roads linking it to two cities (see the picture below). The contractual time for the entire project is 18 months.

The construction of each road will require between 15 and 20 months and that of the bridge will require between 12 and 19 months. The three parts of the projects can be done simultaneously and independently. Let X_1, X_2 and Y represent the construction times for the two roads and the bridge, respectively and suppose that these random variables are uniformly distributed on their respective ranges.
(a) What is the expected time for completion of each part of the project? What are the corresponding standard deviations?

(b) What is the expected time for the completion of the entire project? What is the corresponding standard deviation?

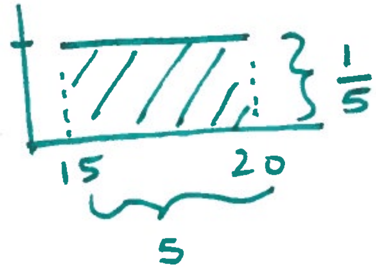
(c) What is the probability that the project will be completed within the contractual time?

Problem 4.24 Same as Problem 2.51, but assuming that the variables X_1, X_2 and Y have triangular distributions over their ranges.



Set up: Job 1:

$X_1 \sim U(15, 20)$



$$f_{X_1}(x) = \frac{1}{5}$$

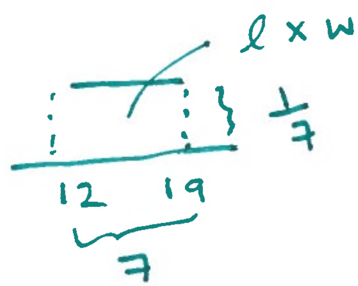
$$15 \le x \le 20$$

$$F_X(x) = \frac{x-a}{b-a} = \frac{x-15}{5} \qquad 15 \le x < 20$$

Job 2: Same as job 1.

Job 3:

$Y \sim U(12, 19)$



$$f_Y(y) = \frac{1}{7} \qquad 12 \le y \le 19$$

$$F_Y(y) = \frac{y-12}{7}$$

Completion time T entire project.

$T = \underline{\max}(X_1, X_2, Y)$

Is it $\max(X_1, X_2, Y)$ or

$\min\cancel{(X_1, X_2, Y)}$

Find cdf of 1.

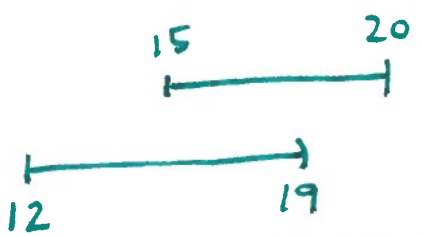
$$F_T(t) = P(T \le t)$$
$$= P(X_1 \le t,\ X_2 \le t,\ Y \le t)$$
$$= F_{X_1}(t)\, F_{X_2}(t)\, F_Y(t) \qquad \text{independent.}$$
$$= [F_X(t)]^2\, F_Y(t)$$
$$= \left(\frac{t-15}{5}\right)^2 \left(\frac{t-12}{7}\right)$$

↑ Support?

Is it $15 \cancel{\le t} \le 20$

No, $F_T(20) = 1^2\, \frac{8}{7} > 1$

Support in 3 parts:



roads

bridge.

$$F_T(t) = \left(\frac{t-15}{5}\right)^2 \left(\frac{t-12}{7}\right)$$

$15 \le t \le 19$

$(12, 15)\ (15, 19)\ (19, 20)$

↑

$F_T(t) = 0$ Why? Its not possible for $P(T \le t)$ to have value when $12 \le t < 15$

Fastest job can be complete is at 15 months

$$F_T(t) = \begin{cases} 0 & t < 15 \\ \left(\frac{t-15}{5}\right)^2\left(\frac{t-12}{7}\right) & 15 \le t \le 19 \\ \left(\frac{t-15}{5}\right)^2 & 19 < t \le 20 \\ 1 & t > 20 \end{cases}$$

at this point must have bridge been completed. so only roads relevant.

a) $E(X_1) = \frac{15+20}{2} = 17.5$

$Var(X_1) = \frac{(20-15)^2}{12}$

$SD(X_1) = \sqrt{\frac{25}{12}}$

$E(Y) = \frac{12+19}{2} = 15.5$

$Var(Y) = \frac{(19-12)^2}{12}$

$SD(Y) = \sqrt{\frac{49}{12}}$

b) Using $F_T(t) \rightarrow f_T(t)$

$E(T)$

$Var(T)$

$SD(T)$

tedious not hard.

c) $P(T \le 18) = F_T(18)$

$= \left(\frac{18-15}{5}\right)^2 \left(\frac{18-12}{7}\right) = 30.8\%$