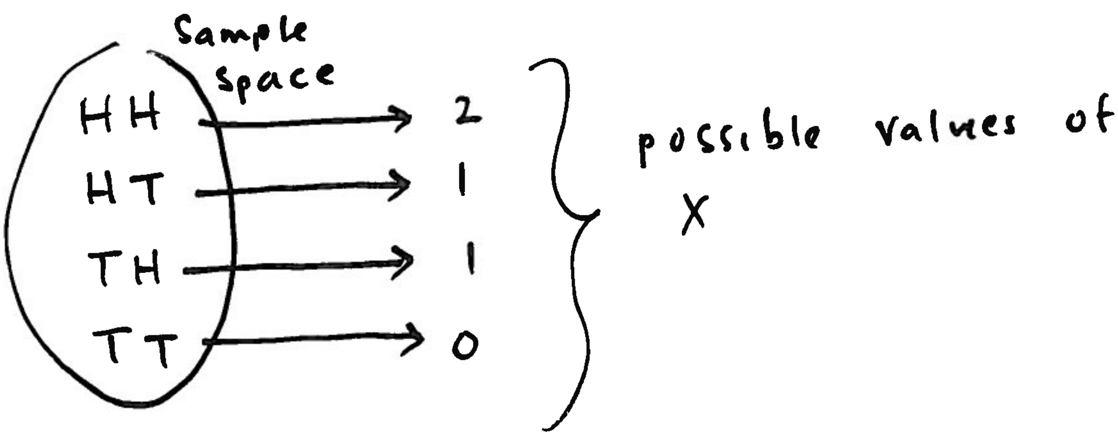
Ch. 4 - Random Variables and Distributions

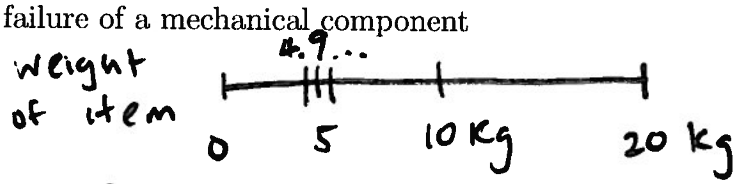
Random variable: a function (rule) that assigns a number with each outcome in the sample space. Usually denoted with capital letters, (X, Y, Z) and its possible realized values are denoted by the same lowercase letters (x, y, z). i.e. X is a random quantity before the experiment is performed and x is the random quantity after the experiment has been performed.
E.g. Toss a coin twice and X = number of heads



There are two types of random variables:

- **Discrete** random variables can take on a finite or countable set of values.
  e.g. number of defective items, number of sales for a store, whether it will rain tomorrow or not

- **Continuous** random variables are defined on a continuous range. Can take on an uncountable set of values
  e.g. weight of an item, time until failure of a mechanical component



infinite # values it could take on.

1 Discrete Random Variables

Probability mass function (denoted $f(x)$) is a function that gives the probability of occurrence for each possible realized value x of the random variable X.

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

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

Properties of $f(x)$

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

Distribution function of X (denoted $F(x)$) is defined as

$F(1) = P(X \leq 1)$

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

Let's consider a simple example of tossing a coin 2 times, where X = number of heads.

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

x # heads	$f(x) = P(X = x)$	$F(x) = P(X \leq x)$
0	$1/4$ $\quad P(X=0) = \frac{1}{4}$	$P(X \leq 0) = 1/4$
1	$1/2$ $\quad P(X=1) = \frac{1}{2}$	$P(X \leq 1) = P(X=0) + P(X=1) = 0.75$
2	$1/4$	$P(X \leq 2) = 1$

$\sum f(x) = 1$

We will learn more about discrete random variables in ch. 6.

Cont.

$$P(X=x) = P(x \le X \le x) = \int_x^x f(t)\,dt = 0$$

2 Continuous Random Variables

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. It differs from the discrete density function in that the probability that a continuous random variable will equal a particular value is 0. (i.e. $P(X = a) = 0$ for a continuous random variable). It cannot be expressed in tabular form and instead an equation is used to describe a continuous probability distribution.

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

$$P(a \le X \le b) = P(a < X \le b) = P(a \le X < b) = P(a < X < b)$$

Properties of $f(x)$

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

Cumulative distribution function (cdf) (denoted as $F(x)$) 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

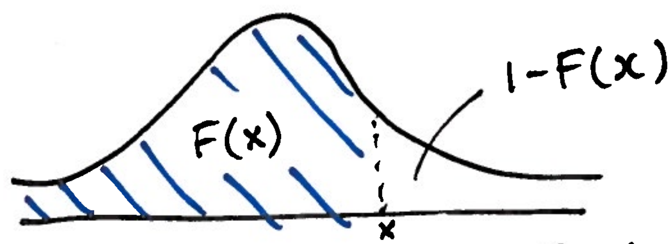
$$\begin{aligned} F'(x) &= f(x) \\ f(x) \to F(x) &\quad \text{by integration and} \\ F(x) \to f(x) &\quad \text{by differentiation} \end{aligned}$$

Why do we learn about $F(x)$?
$F(x)$ is useful for computing probabilities, such that

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

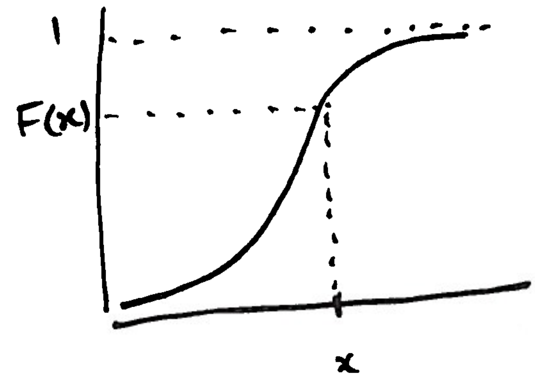
Also, $$P(X > a) = 1 - F(a)$$

density

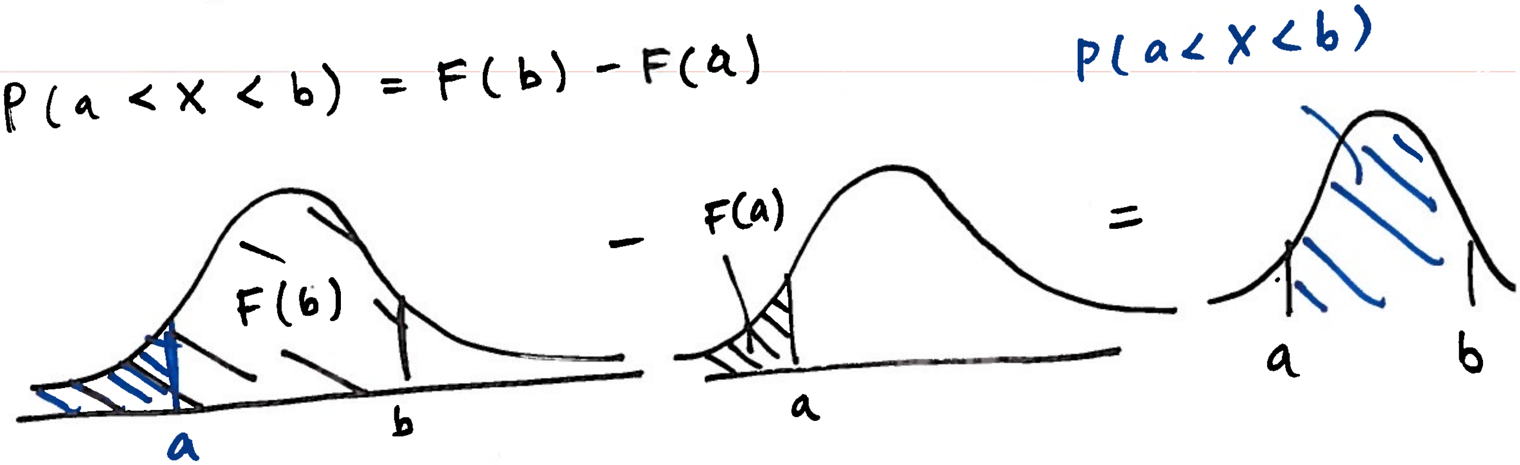


area under curve = 1

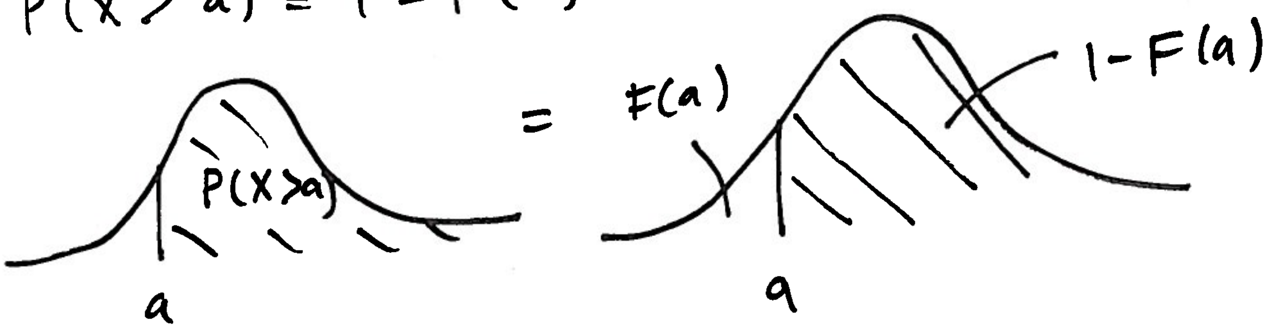
distribution:



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



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



Example 1.

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

c constant

(a) Find the value of c.

$y = \frac{1}{360}$

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

Solution: (we will work through this in class)

a) property 2

$$\int_{-\infty}^{\infty} f(x)\,dx = 1 \quad \Rightarrow \int_0^{360} \frac{1}{c}\,dx = 1$$

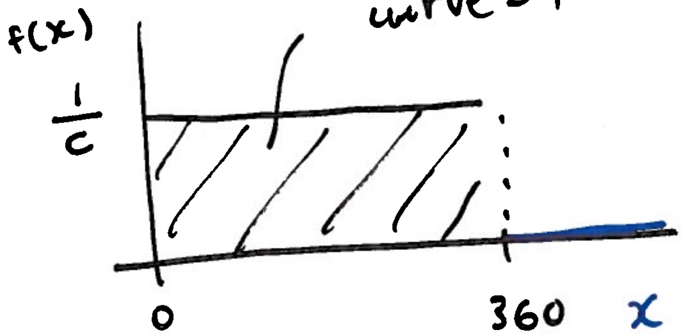
$$\frac{1}{c}x \Big|_0^{360} = 1$$

height at curve values of x

$$\frac{1}{c}[360 - 0] = 1$$

$c = 360$

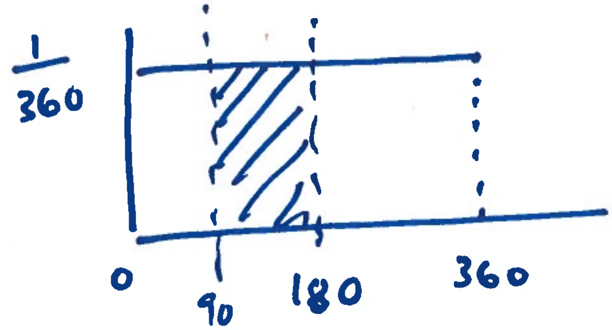
area under curve = 1



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

$c = 360$

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



$$= \int_{90}^{180} \frac{1}{360}\, dx$$

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

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

Method 2.

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

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

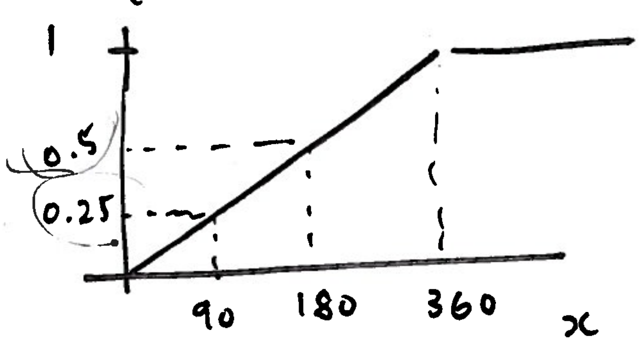
$$F(x) = \begin{cases} 0 & x \le 0 \\ \frac{x}{360} & 0 \le x \le 360 \\ 1 & x > 360 \end{cases}$$

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

$F(180) - F(90)$

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

cdf.



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

Example 2.

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

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

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

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

Solution: (in class)

a) $F(x) = \int_{-\infty}^{x} f(t)\,dt$] def. of cdf.

$= \int_0^x \frac{1}{8} + \frac{3}{8} t \, dt = \frac{1}{8} t + \frac{3}{8} \frac{t^2}{2} \Big|_0^x$

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

$\int_{-\infty}^{0} \overset{=0}{f(t)\,dt} + \int_0^x$

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

Notice:

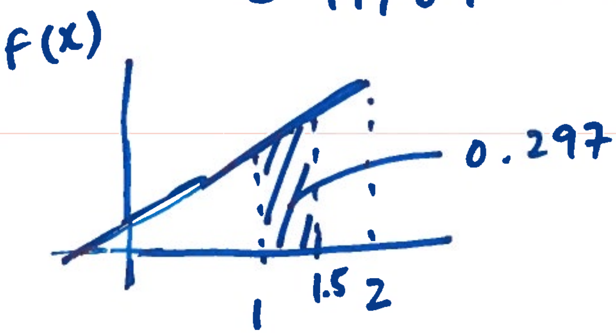
$f(x) = F'(x)$

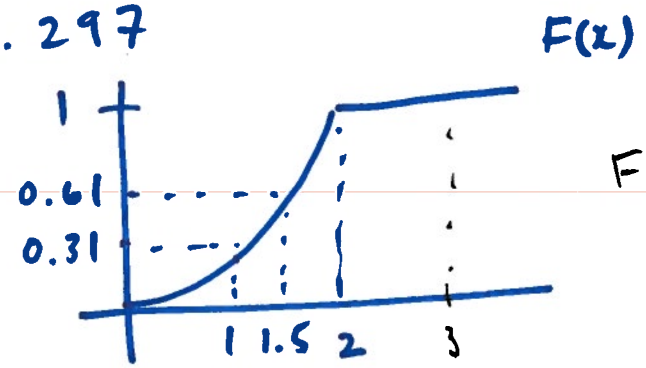
$= \frac{d}{dx}\left(\frac{x}{8} + \frac{3}{16} x^2\right) = \frac{1}{8} + \frac{3}{8} x$

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

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

$$= 19/64 = 0.297$$





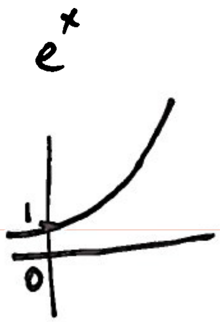
$F(3) =$

$0 \le x \le 2$

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

$$= 1 - F(1)$$

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



3 Summarizing the main features of $f(x)$

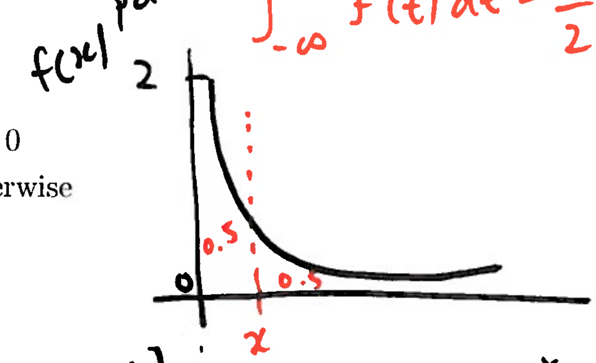
How to find median, Q_1, Q_3, IQR from $f(x)$

- Steps to find the median:
  - (a) Find $F(x)$
  - (b) then solve for x such that $F(x) = 0.5$
    x is the median
- 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 $IQR = Q_3 - Q_1$

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

Example 3.

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



Find the median, Q_1, Q_3 and IQR.
Solution: (in class)

Step a)

$$F(x) = \int_0^x 2e^{-2t}\,dt = 2\left[-\frac{1}{2}e^{-2t}\right]_0^x$$

$$= -(e^{-2x} - \underbrace{e^0}_{1}) = \underline{1 - e^{-2x}}$$

Step b)

$$F(x) = 0.5$$

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

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

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

$$-2x = -\ln(2)$$

$$x = \frac{\ln(2)}{2} = 0.347$$

$$\ln\left(\frac{1}{2}\right) = \underbrace{\ln(1)}_{0} - \underbrace{\ln(2)}$$

Exercise:
$Q_1 = 0.144$
$Q_3 = 0.693$
$IQR = 0.549.$

$E(X-\mu)^2 = \Sigma (x-\mu)^2 f(x)$

<u>Mean and Variance of a *Discrete* Random Variable</u>

- To find the mean, (expectation)

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

In general,

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

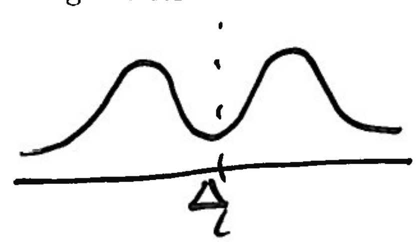
- To find the variance,

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

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

Example 4. Find the mean, variance and standard deviation using the following probability model:

x	2	4	6
$f(x)$	0.5	0.3	0.2



<u>Solution:</u>

$$\begin{aligned} E(X) &= (2 \times 0.5) + (4 \times 0.3) + (6 \times 0.2) \\ &= 3.4 \\ Var(X) &= E(X^2) - [E(X)]^2 \end{aligned}$$

So we need to find $E(X^2)$

$$\begin{aligned} E(X^2) &= (2^2 \times 0.5) + (4^2 \times 0.3) + (6^2 \times 0.2) \\ &= 14 \end{aligned}$$

Thus

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

$$\begin{aligned} Var(X) &= 14 - 3.4^2 = 2.44 \\ SD(X) &= \sqrt{2.44} = 1.56 \end{aligned}$$

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,

$$\int_{-\infty}^{\infty} x^2 f(x)\, dx$$

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

Example 5. Find the mean and standard deviation of

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

Solution

$$\int_0^1 x \cdot 2x \; dx$$

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx$$
$$= \int_0^1 2x^2 dx$$
$$= \frac{2}{3} x^3 \Big|_0^1 = \frac{2}{3}$$

$$2\int_0^1 x^2 \cdot x^1 \, dx$$

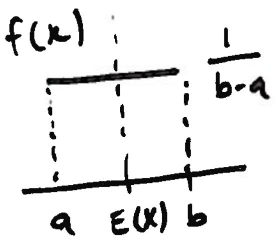
$$E(X^2) = \int_0^1 2x^3 dx = \frac{1}{2}$$
$$Var(X) = E(X^2) - E(X)^2 = \frac{1}{2} - (\frac{2}{3})^2 = \frac{1}{18}$$
$$SD(X) = \sqrt{\frac{1}{18}} = 0.236$$

4 Some Continuous Models

We will now introduce two continuous random variables: the uniform and exponential random variables. Ch. 5 we will discuss the normal random variable, which is the most common continuous distributions in statistics.

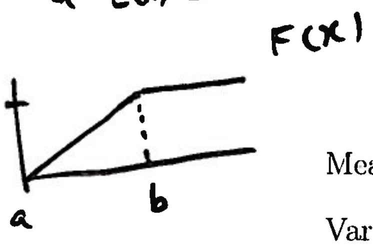
Uniform Random Variables ex. 1 was a uniform r.v.

If X is a uniform random variable, we write $X \sim U(a, b)$. This indicates that X is uniformly (evenly) distributed over the interval $[a, b]$.



Density Function: $f(x) = \frac{1}{b-a}, \qquad a \leq x \leq b$

Distribution Function:



$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}$$

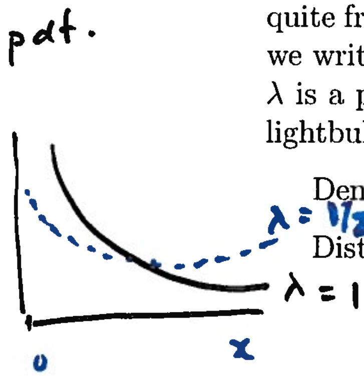
Mean: $E(X) = \frac{a+b}{2}$

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

Exponential Random Variables

Exponential random variables are often used to model the time until an event occurs. For science and engineering you encounter exponential distributions quite frequently. If X is an exponential random variable with rate of λ, then we write, $X \sim \exp(\lambda)$

λ is a positive constant and is the reciprocal of the mean lifetime. i.e. If a lightbulb has a mean lifetime of 5 years then $\lambda = \frac{1}{5}$



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

Distribution Function:

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases}$$

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

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

Suppose you have the pdf of X ($f(x)$) and you want to find the pdf of $Y = X^2$. How?
Steps:

1. Find the cdf of X; $F(x) = \int_{-\infty}^{x} f(t)dt$
2. Find the cdf of Y; $F_Y(Y) = P(Y \le y)$
3. Differentiate the cdf of Y to get $f_Y(y)$; $f_Y(y) = F_Y'(y)$

Remember: you need to consider the support of y.

Problem 4.13 in course notes is important. We will solve it in class.

Example 6. 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;

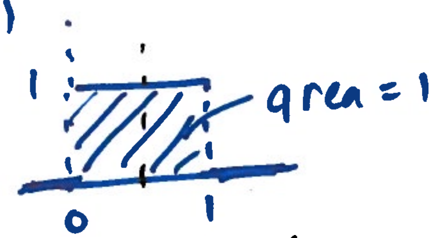
(b) Calculate the mean (expectation) $E(X)$ and variance $Var(X)$;

(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.

Solution:

a)



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

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

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

cont. infinite

$P(X = a) = 0$

b) $E(X) = \frac{a+b}{2} = \frac{1}{2}$. $Var(X) = \frac{(b-a)^2}{12}$

$= \frac{1}{12}$

$E(Y) =$

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

$\underline{Y = \sqrt{X}}$ $= \int_0^1 \sqrt{x}\cdot 1\, dx = \int_0^1 (x)^{\frac{1}{2}}\, dx$

$= \frac{2}{3} x^{3/2} \Big|_0^1 = \frac{2}{3}$

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

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

$Var(Y) = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{2} - \frac{4}{9} = 0.056$

d) Step. 1. cdf of X.
$F(x) = x$ $0 < x < 1$ $\boxed{Y = \sqrt{X}}$

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

Support of x: $0 < x < 1$
y: $0 < y < 1$

$$G(y) = \begin{cases} 0 & y < 0 \\ y^2 & 0 < y < 1 \\ 1 & y > 1 \end{cases}$$ cdf of y.

$$g(y) = G'(y)$$ pdf of y.

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

$$g(y) = \begin{cases} 2y & 0 < y < 1 \\ 0 & \text{o.w} \end{cases}$$

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

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

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

$$Var(Y) = \frac{1}{18} = 0.056$$

match. part c.

Last class we saw:

$X \sim \exp(\lambda)$

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

$$E(X) = \frac{1}{\lambda} \qquad F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases}$$

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

$$E(X) = \int_0^\infty x \cdot \lambda e^{-\lambda x}\, dx = \underline{\underline{\lambda \int_0^\infty x e^{-\lambda x}\, dx}}$$

Integration by parts.

$u = x \qquad dv = e^{-\lambda x}\, dx \qquad uv - \int v\, du$

$du = dx \qquad v = -\frac{1}{\lambda} e^{-\lambda x}$

$$E(X) = \cancel{\lambda}\left[\underset{\uparrow \atop u}{x} \cdot \underbrace{\left(-\frac{1}{\cancel{\lambda}} e^{-\lambda x}\right)}_{v}\Bigg|_0^\infty + \frac{1}{\cancel{\lambda}}\int_0^\infty \underbrace{e^{-\lambda x}}_{v}\, \underbrace{dx}_{du}\right]$$

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

$$= 0 \qquad -\frac{1}{\lambda} e^{-\lambda x}\Big|_0^\infty = \frac{1}{\lambda}$$

$$Var(X) = \underline{\underline{E(X^2)}} - E(X)^2 \qquad E(g(X)) = \int g(x) f(x)\, dx$$

$$E(X^2) = \lambda \int x^2 \cdot e^{-\lambda x}\, dx$$

integration by parts.

$$u = x^2 \qquad dV = e^{-\lambda x}\, dx$$

$$du = 2x\, dx \qquad V = -\frac{1}{\lambda} e^{-\lambda x}$$

$$E(X^2) = \cancel{\lambda}\Big[\underbrace{\overset{u}{x^2}\overset{v}{\left(-\frac{1}{\lambda} e^{-\lambda x}\right)}}_{0}\Big|_0^\infty + \int \frac{1}{\cancel{\lambda}} e^{-\lambda x} \cdot 2x\, dx$$

$$= \int_0^\infty e^{-\lambda x} \cdot 2x\, dx \qquad \text{int. by parts.}$$

$$= \frac{2}{\lambda^2}$$

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

$$= \frac{1}{\lambda^2}$$

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

$$= \int_0^x \lambda e^{-\lambda t}\, dt = -\frac{\lambda}{\lambda} e^{-\lambda t}\Big|_0^x$$

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

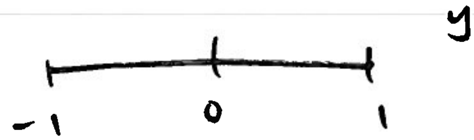
$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases}$$

Example 7.

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

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

<u>Solution:</u>



Since $-1 \le y \le 1$

$0 \le y^2 \le 1$

$0 \le u \le 1$

$$F_u(u) = P(U \le u)$$

$$= P(Y^2 \le u) = P(-\sqrt{u} \le Y \le \sqrt{u})$$

we can use <u>cdf</u> method (method A) or use integration directly (method B)

Method A:

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

$$= \int_{-1}^{y} \frac{t+1}{2}\, dy = \left. \frac{t^2}{4} + \frac{t}{2} \right|_{-1}^{y}$$

$$= \left(\frac{y^2}{4} + \frac{y}{2}\right) - \left(\frac{1}{4} - \frac{1}{2}\right)$$

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

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

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

$$= \frac{4\sqrt{u}}{4} = \sqrt{u} \quad \text{cdf}$$

Method B: $P(-\sqrt{u} \le Y \le \sqrt{u}) =$

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

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

Method A and B produce the same result:

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

$$\sqrt{u} = u^{\frac{1}{2}}$$

cdf

$$F_u(u) = \begin{cases} 0 & u < 0 \\ \sqrt{u} & 0 \le u \le 1 \\ 1 & u > 1 \end{cases}$$

pdf

$$f_u(u) = \begin{cases} 0 & \text{otherwise.} \\ \frac{1}{2} u^{-\frac{1}{2}} & 0 \le u \le 1 \end{cases}$$

5 Properties of the Mean and Variance

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

6 Sum and Average 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 $\{X_1, ..., X_n\}$ an independent random sample and we get:

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

eg. $E(X) = 3$, $Var(X) = 2$

$$E(2X+3) = 2E(X) + 3 = 9$$

$$Var(\tfrac{1}{2}X) = \tfrac{1}{4} Var(X) = \tfrac{2}{4} = \tfrac{1}{2}$$

eg. Given $E(X) = 3$ $Var(X) = 2$

$E(Y) = 5$ $Var(Y) = 1$

X, Y independent random variables.

$E(X-Y) = ?$ $Var(X-Y) = ?$

$E(X) - E(Y) = 3 - 5$

$= -2$

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

$= 2 + 1 = 3$

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

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

If common mean and variance, then:

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

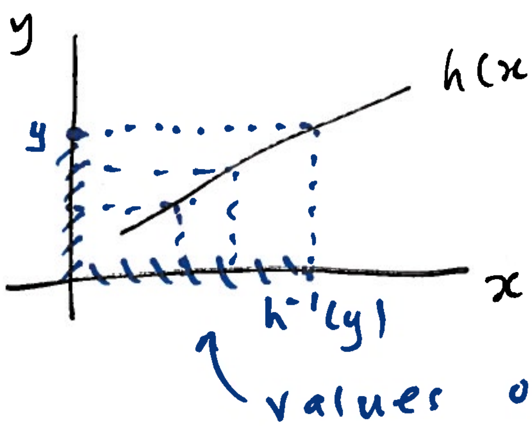
If common mean and variance, then:

$$= \frac{1}{n^2}[n\sigma^2] = \frac{\sigma^2}{n}$$

Last class

$X \sim f_X(x)$ and $F_X(x)$

$Y = \underbrace{h(x)}_{\text{some function of } x.}$ eg. $Y = X^2$ $\quad f_Y(y) = ?$

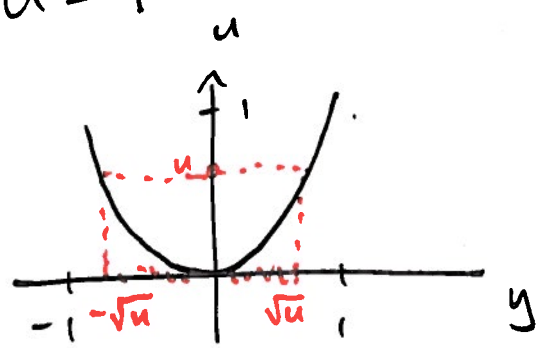


$F_Y(y) = P(Y \le y)$

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

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

$U = Y^2$



$F_U(u) = P(U \le u)$

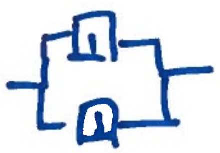
$= P(-\sqrt{u} \le y \le \sqrt{u})$

7 Maximum and Minimum of Independent Random Variables

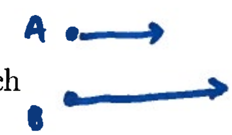
There are cases when we are interested in the maximum or minimum of a random sample.

Maximum:

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.

Setup of problem:

You are given a pdf of n independent random variables $X_1, X_2, ..., X_n$.

Question: Find the pdf of the maximum of $X_1, X_2, ..., X_n$

Steps:

Let $Y = \max(X_1, X_2, ..., X_n)$.

To find the pdf of Y we start by finding its cdf.

if y is the max of $X_1, \ldots, X_n$ then each of $X_1, \ldots, X_n$ must be less than or equal to y

$$\begin{aligned}
F_Y(y) &= P(Y \le y) \\
&= P(X_1 \le y, X_2 \le y, ..., X_n \le y) \\
&= P(X_1 \le y)P(X_2 \le y)...P(X_n \le y) \quad \text{(because } X_i\text{'s are independent)} \\
&= F_{X_1}(y)F_{X_2}(y)...F_{X_n}(y)
\end{aligned}$$

Furthermore, if all the X_is have the same pdf,

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

We found the cdf of Y, now how do we get the pdf of Y?

$$\begin{aligned}
f_Y(y) &= F_Y'(y) \\
&= n[F_X(y)]^{n-1} f_X(y)
\end{aligned}$$

Simple Ex:

Electronic components, length of life Y (hours)

with pdf given by:

$$f_Y(y) = \begin{cases} \frac{1}{100} e^{-y/100} & y > 0 \\ 0 & \text{otherwise} \end{cases}$$

Suppose 2 components, operate independently and connected in series in a certain system.

Find the pdf of X (the lifetime of the system).

Solution:

$X = \min(Y_1, Y_2)$

Why? The system fails when the 1st component fails.

Find cdf of X.

$$F_X(x) = P(X \le x)$$
$$= 1 - P(X > x)$$
$$= 1 - P(Y_1 > x, Y_2 > x)$$
$$= 1 - P(Y_1 > x)P(Y_2 > x)$$ b/c Y_1, Y_2 indep.

$$= 1 - [1 - \underbrace{(1 - e^{-x/100})}_{P(Y \le x) = F_Y(x)}]^2$$

should show work on test.

X

x

$$= \underbrace{\quad}_{[P(Y > x)]^2}$$

$$= 1 - (e^{-x/100})^2$$

We differentiate $F_X(x)$ to get pdf:

$$f_X(x) = -2(e^{-x/100})^1 \left(-\frac{1}{100} e^{-x/100}\right)$$ chain rule

$$= \frac{1}{50} e^{-x/50}$$

$$f_X(x) = \begin{cases} \frac{1}{50} e^{-x/50} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

the minimum of 2 exp. r.v. has an exponential distribution as well.

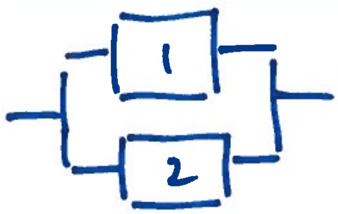
$E(Y) = 100$ hrs ← each component.

$E(X) = 50$

↑ system connected in series

Next example:

Consider 2 independent components in parallel



system does not fail until both components fail

Find the pdf of X, the lifetime of the system.

Solution:

$$X = \max(Y_1, Y_2)$$

↑ why?

Find cdf of X first,

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

$$= P(Y_1 \le x, Y_2 \le x) \qquad Y_1, Y_2 \text{ indep.}$$

$$= P(Y_1 \le x)\, P(Y_2 \le x)$$

$$= \Big(\underbrace{1 - e^{-x/100}}\Big)^2$$

show how you get this on a test/assign.

$$f_X(x) = F_X'(x) \quad \text{deriv.}$$

$$= 2(1-e^{-x/100})'\left(-e^{-x/100}\left(-\frac{1}{100}\right)\right)$$

$$= \frac{2}{100}\left(e^{-x/100} - e^{-x/150}\right)$$

$$f_X(x) = \begin{cases} \frac{1}{50}\left(e^{-x/100} - e^{-x/150}\right) & x > 0 \\ 0 & \text{o.w.} \end{cases}$$

We see that the maximum of 2 exp. r.v. is not an exp. r.v.

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

Example 8. 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 $\mu = 3$.

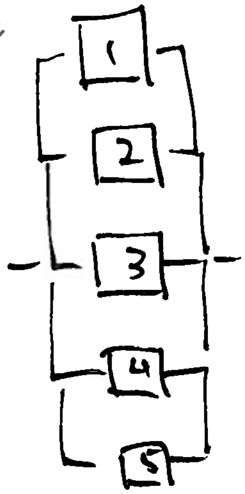
(a) Calculate the median and standard deviation for each component

(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, mean life and standard deviation for the system.

Ex. 8



Let X_i be the lifetime of each component. $i = \{1, 2, 3, 4, 5\}$

$X \sim \exp(\lambda = \frac{1}{3}) \Rightarrow f_X(x) = \frac{1}{3} e^{-\frac{1}{3}x}, \quad x \geq$

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

The max lifetime of $X_1, \dots, X_5$ is the lifetime of the system. (System continues working until last component fails.

a) $\boxed{F_X(x) = 1 - e^{-x/3}}$ $x \geq 0$ } Need to show on exam

set $F_X(x) = 0.5$

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

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

$-x/3 = \ln(\frac{1}{2})$

$-\frac{x}{3} = -\ln(2)$

$x = 3\ln(2)$

$= 2.08$

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

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

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

c) Let Y be the lifetime of the system.

$Y = \max(X_1, X_2, X_3, X_4, X_5)$

Find cdf:

$F_Y(y) = P(Y \le y) = P(X_1 \le y, X_2 \le y, \dots, X_5 \le y)$

$= P(X_1 \le y) P(X_2 \le y) \dots P(X_5 \le y)$ (independence)

$= [P(X \le y)]^5$

$= [F_X(y)]^5 = \left[1 - e^{-\frac{1}{3}y}\right]^5$ cdf of Y

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

d) median of system:

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

set $F_Y(y) = 0.5$

$\left[1 - e^{-\frac{1}{3}y}\right]^5 = 0.5$

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

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

$-\frac{y}{3} = \ln(1 - 0.5^{1/5})$

$y = -3 \ln(1 - 0.5^{1/5}) = 6.133$

To find mean and standard deviation, of the system; we are going to make use of

$E(Y) = \int_0^\infty y f(y)\, dy$ and $Var(Y) = E(Y^2) - E(Y)^2$

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

$$= \frac{d}{dy}\left[1 - e^{-y/3}\right]^5 = 5\left[1-e^{-y/3}\right]^4\left[-e^{-y/3}\left(-\tfrac{1}{3}\right)\right]$$

$$= \frac{5}{3}\underbrace{e^{-y/3}\left[1-e^{-y/3}\right]^4}$$

$$e^{-y/3}\left[1-e^{-y/3}\right]^4 = e^{-y/3}(1-e^{-y/3})(1-e^{-y/3})(1-e^{-y/3})(1-e^{-y/3})$$

$$= e^{-y/3}(1-2e^{-y/3}+e^{-2y/3})(1-2e^{-y/3}+e^{-2y/3})$$

$$= e^{-y/3}\left(1 - 2e^{-y/3} + e^{-2y/3} - 2e^{-y/3} + 4e^{-2y/3} - 2e^{-y} + e^{-2y/3} - 2e^{-y} + e^{-\frac{4}{3}y}\right]$$

$$= e^{-y/3} - 4e^{-2/3y} + 6e^{-y} - 4e^{-4/3y} + e^{-5/3y}$$

$$f_Y(y) = \frac{5}{3}\left[e^{-y/3} - 4e^{-2/3y} + 6e^{-y} - 4e^{-4/3y} + e^{-5/3y}\right]$$

For any $a>0$

$$\int_0^\infty y e^{-ay}\,dy = \left(\frac{1}{a}\right)^2$$

let $u = y$ $\quad dV = e^{-ay}\,dy$

$du = dy$ $\quad V = -\frac{1}{a}e^{-ay}$

$uv - \int v\,du$

$$\underbrace{-\frac{1}{a}y\,e^{-a/y}\Big|_0^\infty}_{0} + \frac{1}{a}\int e^{-ay}\,dy = \frac{1}{a}\,\frac{e^{-ay}}{-a}\Big|_0^\infty$$

$$= -\left(\frac{1}{a}\right)^2 e^{-ay}\Big|_0^\infty$$

$$= 0 - \left(-\frac{1}{a^2}\cdot 1\right)$$

$$= \left(\frac{1}{a}\right)^2$$

So

$$E(Y) = \int_0^\infty y f(y)\, dy$$

$$= \int_0^\infty y \frac{5}{3}\left[e^{-y/3} - 4e^{-2/3y} + 6e^{-y} - 4e^{-4y/3} + e^{-5/3y}\right]$$

$$= \frac{5}{3}\left[\int_0^\infty \underbrace{y e^{-y/3}\, dy}_{9} - 4\underbrace{\int_0^\infty y e^{-2/3y}\, dy}_{9/4} + \underbrace{6\int_0^\infty y e^{-y}\, dy}_{1} - 4\underbrace{\int y e^{-4/3y}\, dy}_{9/16} + \int_0^\infty \underbrace{y e^{-5/3y}\, dy}_{9/25}\right]$$

$$= \frac{5}{3}\left[9 - 4\left(\frac{9}{4}\right) + 6\cdot 1 - 4\cdot\frac{9}{16} + \frac{9}{25}\right]$$

$$= 6.85$$

$$Var(Y) = E(Y^2) - \underbrace{E(Y)}_{6.85}{}^2$$

$$E(Y^2) = \int_0^\infty y^2 f(y)\, dy$$

$$= \frac{5}{3}\left[\int_0^\infty y^2 e^{-y/3} dy - 4\int_0^\infty y^2 e^{-2y/3} dy + 6\int_0^\infty y^2 e^{-y} dy - 4\int_0^\infty y^2 e^{-y\frac{4}{3}} dy + \int_0^\infty y^2 e^{-5/3y} dy\right]$$

Notice:

$$\int_0^\infty y^2 e^{-ay}\, dy = \frac{2}{a^3}$$

$u = y^2 \qquad dV = e^{-ay}\, dy$

$du = 2y\, dy \qquad V = \frac{e^{-ay}}{-a}$

$\underline{uv} - \int v\, du$

$$= 0 - \int_0^\infty -\frac{1}{a} e^{-ay}\, 2y\, dy$$

$$= \frac{1}{a} 2 \underbrace{\int y e^{-ay}\, dy}_{\frac{1}{a^2} \text{ (from earlier)}} = \frac{2}{a^3}$$

$$E(Y^2) = \frac{5}{3}\left[\frac{2}{(\frac{1}{3})^3} - 4\left(\frac{2}{(\frac{2}{3})^3}\right) + 6\,\frac{2}{1^3} - 4 \cdot \frac{2}{(4/3)^3} + \frac{2}{(5/3)^3} \right] = 60.095$$

$Var(Y) = 60.095 - (6.85)^2 = 13.1725$

$SD(Y) = 3.63$

Minimum:

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

Setup of problem:

You are given the pdf of n independent random variables $X_1, X_2, ..., X_n$.

Question: Find the pdf of the minimum of $X_1, X_2, ..., X_n$

Steps:

Let $Y = \min(X_1, X_2, ..., X_n)$.

Again, to find the pdf of Y we start by finding its cdf.

if y is min of $X_1, \ldots, X_n$, then surely each $X_1, \ldots, X_n$ must be greater than or equal to y

$$
\begin{aligned}
F_Y(y) &= P(Y \leq y) \\
&= 1 - P(Y > y) \\
&= 1 - P(X_1 > y, X_2 > y, ..., X_n > y) \\
&= 1 - [P(X_1 > y)P(X_2 > y)...P(X_n > y)] \quad \text{(because } X_i\text{'s are independent)} \\
&= 1 - [1 - F_{X_1}(y)][1 - F_{X_2}(y)]...[1 - F_{X_n}(y)]
\end{aligned}
$$

Furthermore, if all the X_is have the same pdf,

$$F_Y(y) = 1 - [1 - F_X(y)]^n \quad .$$

How do we get the pdf of Y?

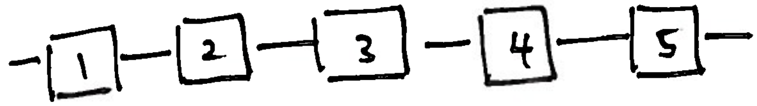
chain rule.

$$
\begin{aligned}
f_Y(y) &= F_Y'(y) \\
&= -n[1 - F_X(y)]^{n-1}(-f_X(y)) \\
&= n[1 - F_X(y)]^{n-1} f_X(y)
\end{aligned}
$$

Example 9. 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.



if any one component fails the system fails. Let X lifetime of each component

$X \sim \exp(\lambda = \frac{1}{3})$ $\quad f(x) = \frac{1}{3} e^{-\frac{1}{3}x}, \quad x \geq 0$

Y = lifetime of system $\quad E(X) = \frac{1}{\lambda} = 3$

$Y = \min(X_1, \dots, X_5)$ $\quad \lambda = \frac{1}{3}$

$$
\begin{aligned}
P(Y \leq 3.5) &= F_Y(3.5) \\
&= 1 - P(Y > 3.5) \\
&= 1 - [P(X > 3.5)]^5 \\
&= 1 - [1 - F_X(3.5)]^5 \\
&= 1 - [1 - (\underbrace{1 - e^{-\frac{1}{3}3.5}}_{\text{cdf of exp.}})]^5 \\
&= 1 - e^{-\frac{5}{3}3.5} = 0.9971
\end{aligned}
$$

Last class, $P(X \leq 3.5) = 0.6886$

b) Notice Y is exponentially dist.

$$F_Y(y) = \underbrace{1 - e^{-\frac{5}{3}y}}_{\text{cdf of an exponential!}}, \quad y \geq 0$$

$$f_Y(y) = -e^{-5/3\,y}\left(-\frac{5}{3}\right) = \frac{5}{3} e^{-5/3\,y}, \quad y \geq 0$$

$$E(Y) = \frac{1}{\lambda} = \frac{3}{5} = 0.6$$

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

median: $F_Y(y) \overset{set}{=} 0.5$

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

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

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

$$= 0.416$$