Ch 4: Continuous RV

Contents

1 Preliminaries **2**
1.1 Continuous Random Variable 3
1.2 Uniform Distribution 12

2 Normal Distribution **16**
2.1 Normal Distribution 17
2.2 Empirical Rule 20
2.3 Standard Normal Distribution 22
2.4 Standardization of Normal: 25
2.5 Binomial Approximation 41

3 Exponential Distribution **42**
3.1 Exponential Distribution 43
3.2 Poisson Process 49
3.3 Memoryless property: 50

---

1 Preliminaries

[ToC]

1.1 Continuous Random Variable

[ToC]

- r.v. whose range is a interval on a real line or a disjoint union of such intervals.

- This leads to major over-haul in pmf $P(X = a)$.

- Suppose X is a r.v. which takes any value within the interval $[0, 1]$ with equal probability.

(Called Uniform(0,1) r.v.)

What value can we assign to $P(X = .5)$?

It also must satisfy that for any constant c, $P(X = c) = 0$.

Probability density function (pdf) of continuous r.v. X is a functon $f(x)$ such that for any two numbers a and b with $a \leq b$,

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

Pdf must satisfy:

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

■

Cumulative Distribution Function (CDF) of r.v. X is a functon $F(x)$ defined for every number x by

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(x)dx$$

If X is a continuous r.v. with pdf $f(x)$ and cdf $F(x)$ then at every x at which the derivative $F'(x)$ exists,

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

Cdf must satisfy:

1. $F(-\infty) = 0$ and $F(\infty) = 1$.
2. non-decreasing.
3. right continuous.

For any number a and b with $a < b$,

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

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

Percentiles

Let p a number between 0 and 1. The $(100 \times p)$th percentile of the distribution of a continuous r.v. X, denoted η_p, is a number such that

$$F(\eta_p) = p$$

Example Let rv X have pdf

$$f(x) = \begin{cases} Kx^2 & \text{if } 1 \leq x \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

1. What is K?
2. What is $P(1.5 \leq X \leq 2)$
3. What is $F(x)$
4. What is 70th percentile of X?

Expected Values

Expected or mean value of a continuous r.v. X with pdf f(x) is

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

If $h(\cdot)$ is any function, then

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

Therefore,

$$E(h(X)) \quad = \quad h\Big(E(x)\Big)$$

if $h(\cdot)$ is a linear function. In other words, $E(aX + b) = aE(X) + b$. $\square$

Variance

Variance of a continuous r.v. X with pdf f(x) is

$$V(X) = \sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx = E\left[(x-\mu)^2\right]$$

and standard deviation (SD) of X is

$$\sigma = \sqrt{\sigma^2}.$$

Example

Let rv X have pdf

$$f(x) = \begin{cases} Kx^2 & \text{if } 1 \leq x \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

1. What is $E(X)$?
2. What is $V(X)$?

1.2 Uniform Distribution

[ToC]

- pdf

$$f(x) = \frac{1}{B - A} \qquad \text{for } A \le x \le B$$

  and 0 otherwise.

- CDF

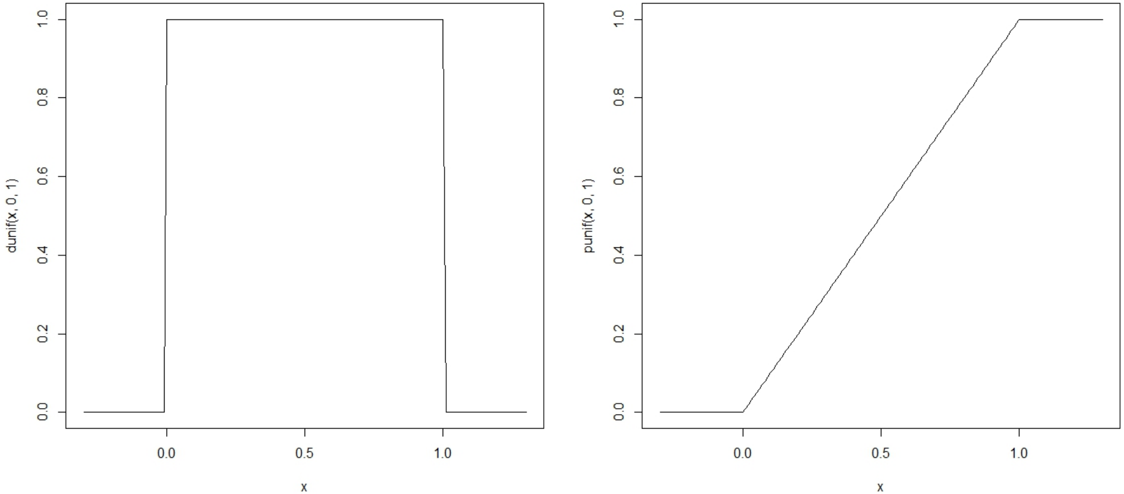
$$F(x) = P(X \le x) = \frac{x - A}{B - A}.$$

- Expectation and Variance:

$$E(X) = \frac{B + A}{2} \qquad V(X) = \frac{(B - A)^2}{12}$$

R code for Uniform(a, b)

```
dunif(.5, 0, 1)         #- F(3):  CDF
punif(.5,0, 1)          #- p(3):  pmf


layout( matrix(1:2, 1, 2) )

x <- seq(-.3,1.3,.01)
plot(x, dunif(x, 0, 1), type="l", ylim=c(0,1))   #- PMF plot -
plot(x, punif(x, 0, 1), type="l", ylim=c(0,1))   #- CDF plot -
```



dunif(x, 0, 1)
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.5
1.0
x
punif(x, 0, 1)
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.5
1.0
x

Example:

1. If $X \sim$Unif(0,1), what is $P(X > .7)$?
2. If $X \sim$Unif(2,7), what is P(X=5)?
3. If $X \sim$Unif(2,7), what is the 80th percentile of X?

2

Normal Distribution

[ToC]

2.1 Normal Distribution

[ToC]

- pdf for $N(\mu, \sigma^2)$

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$$

- CDF

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

  but this is analytically non-tractable, and must be evaluated numerically. We have a table for the case $(\mu, \sigma^2) = (0, 1)$.

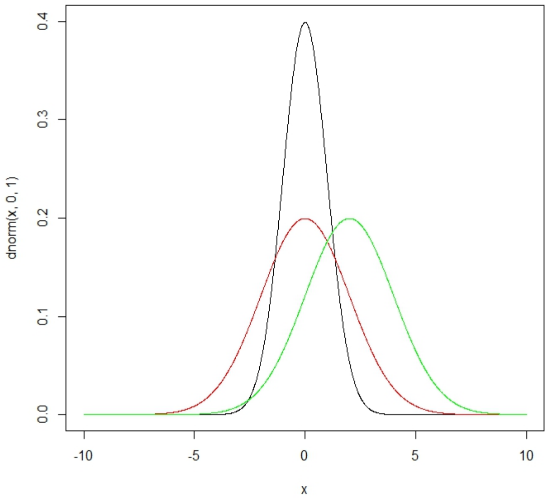
- Mean and Variance

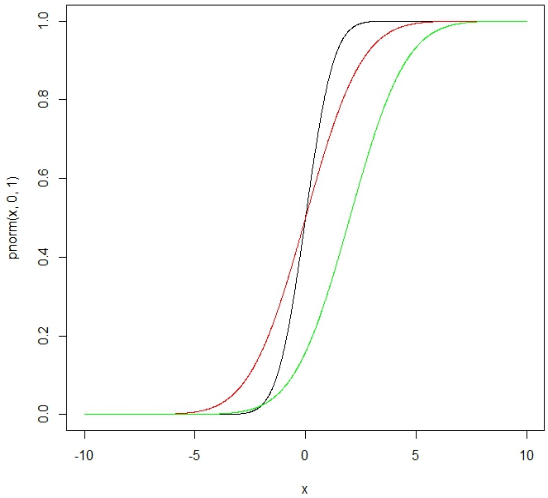
$$E(X) = \mu \qquad V(X) = \sigma^2$$

R code for Normal(μ, σ^2)

```
dnorm(.5, 0, 1)          #- F(3):  CDF
pnorm(.5, 0, 1)          #- p(3):  pmf


layout( matrix(1:2, 1, 2) )

x <- seq(-10,10,.01)
plot(x, dnorm(x, 0, 1), type="l", ylim=c(0,.4))   #- PMF plot -
lines(x,dnorm(x, 0, 2), col="red")
lines(x,dnorm(x, 2, 2), col="green")

plot(x, pnorm(x, 0, 1), type="l", ylim=c(0,1))   #- CDF plot -
lines(x, pnorm(x, 0, 2), col="red")
lines(x, pnorm(x, 2, 2), col="green")
```

$\mathbf{N}(\mu = 0, \sigma = 1)$, $\mathbf{N}(\mu = 0, \sigma = 2)$ and $\mathbf{N}(\mu = 2, \sigma = 2)$





2.2 Empirical Rule

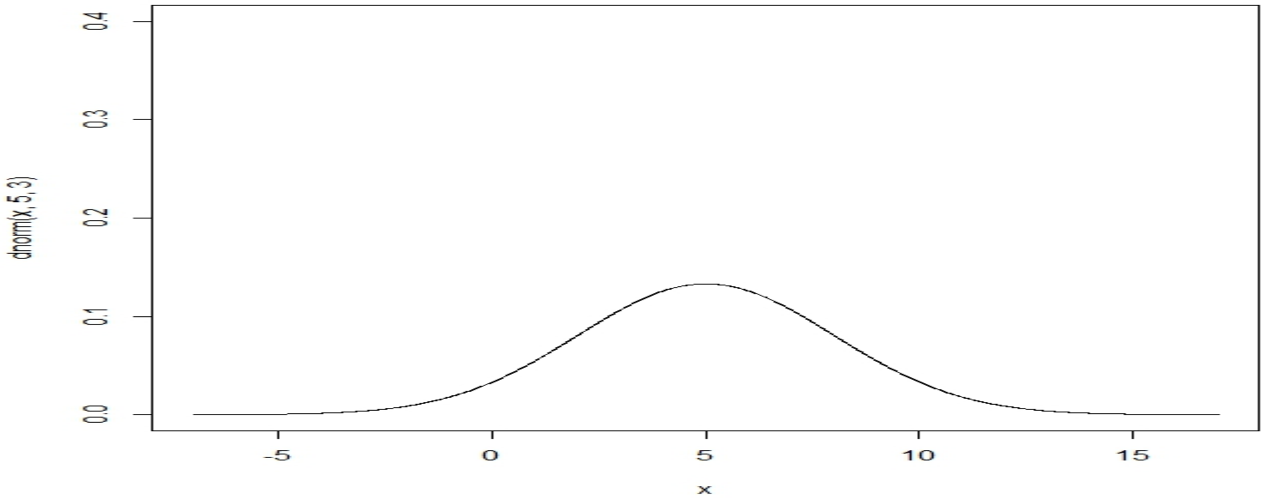
[ToC]

In $X \sim \mathrm{N}(\mu, \sigma^2)$, then

1. with probability .68, X is within 1 SD away from μ.
2. with probability .95, X is within 2 SD away from μ.
3. with probability .99.7, X is within 3 SD away from μ.

```
x <- seq(-7,17,.01)
plot(x, dnorm(x, 5, 3), type="l", ylim=c(0,.4))
```

$\mathbf{N}(\mu = 5, \sigma^2 = 3^2)$



2.3 Standard Normal Distribution

[ToC]

- N(0,1) is called Standard Normal Distribution.
- Pdf of standard normal distribution is

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

- CDF

$$F(t) = P(Z \le t) = \Phi(t).$$

- Table A.3 in the textbook lists values of $\Phi(t)$.

z_α **Notation**

- Z is used to denote Standard Normal random variable.

- z_α denotes $(1-\alpha)100$ th percentile of Z.

- i.e. $z_{.05} = [95\text{th percentile of } Z]$

Using Normal Table

- Find $P(Z \leq 1.4)$

- Find $P(Z > .53)$

- Find 90th percentile of Z

- Find $Z_{.05}$

2.4 Standardization of Normal:

[ToC]

$X \sim \mathrm{N}(\mu, \sigma^2)$ $\qquad\qquad$ $Z \sim \mathrm{N}(0,\ 1)$

$$Z = \tfrac{X-\mu}{\sigma} \quad \Longrightarrow$$

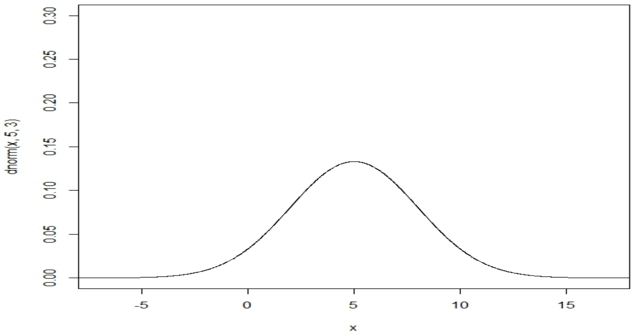
$$\Longleftarrow \quad X = \mu + Z\sigma$$

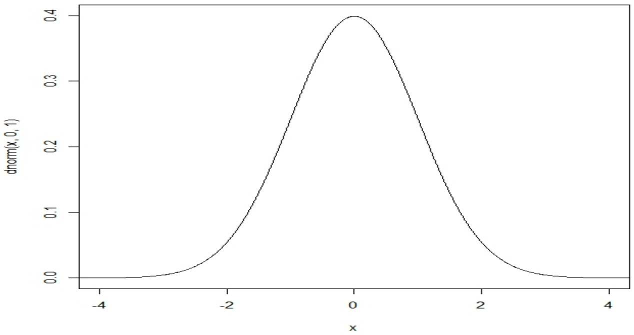
Use Standardization to find $F(x)$

Using standardization, you can use $\Phi(\cdot)$ to figure out the cdf of X.

$$P(X \le a) \quad = \quad P\Big(\frac{X-\mu}{\sigma} \le \frac{a-\mu}{\sigma}\Big) \quad = \quad \Phi\Big(\frac{a-\mu}{\sigma}\Big)$$

Find $P(X \le 8)$ in $\mathrm{N}(5, 3^2)$.





- $$P(a \leq X \leq b) \;=\; P(X \leq b) - P(X \leq a) \;=\; \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

- $$P(X > a) \;=\; 1 - P(X \leq a) = 1 - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

- $P(X \geq a)$

- $P(X = a)$

Example: Tree Height

Diameter at breast height (in.) of trees of certain type is normally distributed with $\mu = 8.8$ and $\sigma = 2.8$.

1. What is probability that randomly chosen tree has diameter less than 10in?
2. What is probability that randomly chosen tree has diameter greater than 20in?
3. What is probability that randomly chosen tree has diameter between 5 and 15?
4. What is range of diameter represents the middle 68% of the trees?

$X \sim N(8.8, 2.8^2)$

What is probability that randomly chosen tree has diameter greater than 20in?

$X \sim N(8.8, 2.8^2)$

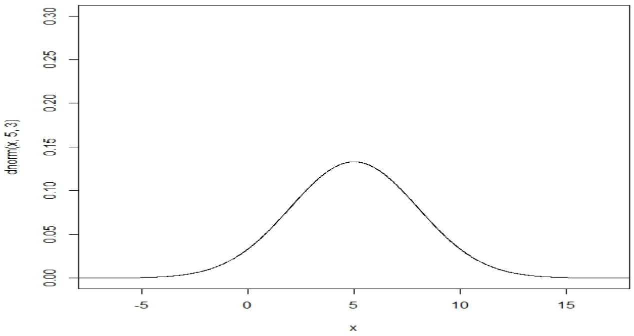
What is probability that randomly chosen tree has diameter between 5 and 15?

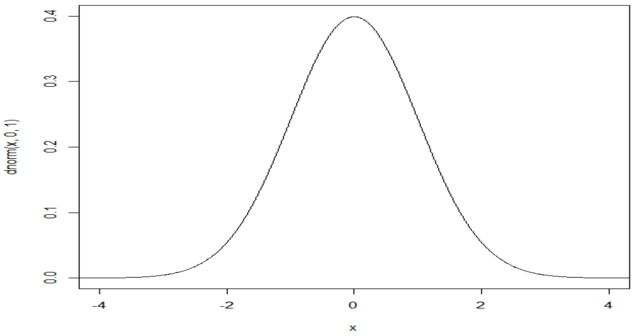
$X \sim N(8.8, 2.8^2)$

What is range of diameter represents the middle 68% of the trees?

Finding percentile of $\mathbf{N}(\mu, \sigma^2)$

Find 90th percentile of $\mathrm{N}(5, 3^2)$.





Example: Find Percentile

Suppose X is Normal random variable with $\mu = 5$ and $\sigma = 2$. What is the 70th percentile of X?

Example: Find Percentile

Suppose X is Normal random variable with $\mu = 5$ and $\sigma = 2$. What is the 70th percentile of X?

- Z-table says

$$\Phi(0.52) = .6985$$

- That means for N(0,1), .52 is the 70th percentile.
- De-standardize .52 to $\mathrm{N}(5, 2^2)$ by

$$X = \mu + Z\sigma \;=\; 5 + (.52)2 = 6.04.$$

- 6.04 is the 70th percentile of X.

Example: Find Percentile 2

Suppose X is a Normal random variable with μ and $\sigma = 2$. For what value of μ, the 70th percentile of X equal to 3.5?

Example: Find Percentile 2

Suppose X is a Normal random variable with μ and $\sigma = 2$. For what value of μ, the 70th percentile of X equal to 3.5?

- Z-table says 70th percentile is at .52.

$$\Phi(0.52) = .6985$$

- De-standardize .52 to $\mathrm{N}(\mu, 2^2)$ by

$$X = \mu + Z\sigma \ = \ \mu + (.52)2.$$

- We need this to equal 3.5. Set up equation as

$$\mu + (.52)2 = 3.5 \qquad \Rightarrow \qquad \mu = 2.46.$$

Example: Tree Height 2

Diameter at breast height (in.) of trees of certain type is normally distributed with $\mu = 8.8$ and $\sigma = 2.8$.

1. To protect younger tree from being cut, we want to ban cutting of smallest 70% of the trees. For what diameters should we ban the cutting?

2. For what value of c does interval $(8.8 \pm c)$ contain 95% of diameters?

Example: Cereal Box

Cereal box is being filled at a factory. Box says it contains 32oz. Let the machine to have $\sigma^2 = 2$ and define [underfilled] as Box< 30, [overfilled] as Box> 33.

1. Determine μ if we want $P(\text{underfilled}) = .03$?
2. For that μ, what is $P(\text{overfilled})$?
3. For the same μ, what σ is needed so that $P(\text{overfilled}) = .05$?

2 For that μ, what is $P(\text{overfilled})$?

3 For the same μ, what σ is needed so that $P(\text{overfilled}) = .05$?

2.5 Binomial Approximation

[ToC]

- If n is sufficiently large ($np \geq 10$ and $n(1-p) \geq 10$),

$$\text{Binomal}(n, p) \quad \approx \quad \text{Normal}(np, np(1-p))$$

- Continuity correction of binomial approximation is done by the formula

$$P(X \leq x) \quad = \quad \Phi\left(\frac{x + .5 - np}{\sqrt{np(1-p)}}\right).$$

3

Exponential Distribution

[ToC]

3.1 Exponential Distribution

[ToC]

- pdf of Exp(λ)

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

- CDF

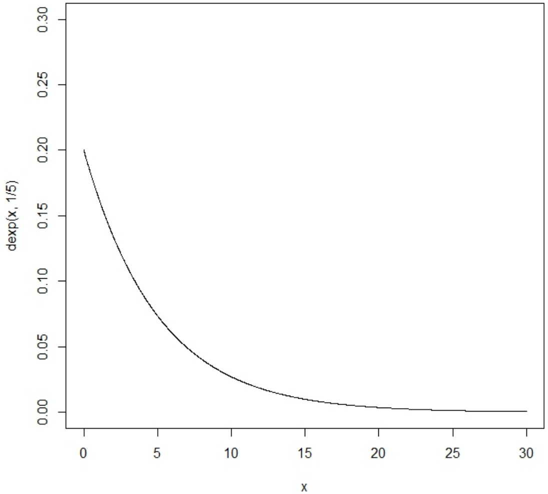
$$F(x) = P(X \leq x) = 1 - \lambda e^{-\lambda x}$$

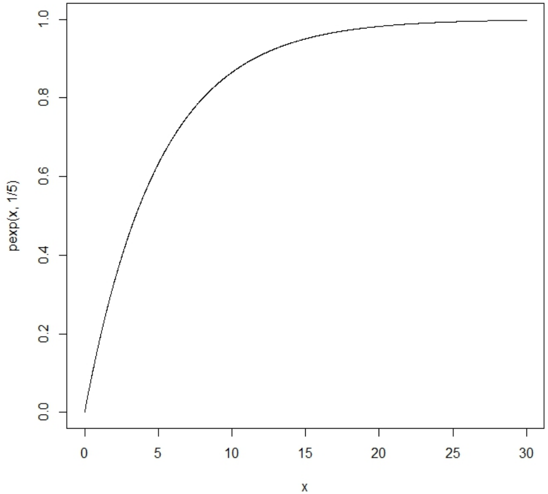
- Mean and Variance

$$E(X) = 1/\lambda \qquad V(X) = 1/\lambda^2$$

R code for Exponential(λ)

```
dexp(.5, 1/5)         #- F(3):  CDF
pexp(.5, 1/5)         #- p(3):  pmf


layout( matrix(1:2, 1, 2) )

x <- seq(0,30,.01)
plot(x, dexp(x, 1/5), type="l", ylim=c(0,.3))   #- PMF plot -
plot(x, pexp(x, 1/5), type="l", ylim=c(0,1))   #- CDF plot -
```


dexp(x, 1/5)
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0
5
10
15
20
25
30
x


pexp(x, 1/5)
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
25
30
x

CDF of exponential

$$
\begin{aligned}
F(x;\lambda) &= P(X \leq x) \\
&= \int_0^x \lambda e^{-\lambda y} dy \\
&= -\frac{\lambda}{\lambda} e^{-\lambda y} \Big|_0^x \\
&= 1 - e^{-\lambda x}.
\end{aligned}
$$

Mean

If $X \sim \text{Exp}(\lambda)$, then

$$\begin{aligned}
E(X) &= \int_0^\infty x \cdot \lambda e^{-\lambda x} dx \\
&\text{Integraing by parts,} \\
&= -x\lambda e^{-\lambda x}\Big|_0^\infty + \int_0^\infty e^{-\lambda x} dx \\
&= 0 - \frac{1}{\lambda} e^{-\lambda x}\Big|_0^\infty \\
&= \frac{1}{\lambda}
\end{aligned}$$

Variance

If $X \sim \text{Exp}(\lambda)$, then

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

$$\begin{aligned}
E(X^2) &= \int_0^\infty x^2 \cdot \lambda e^{-\lambda x} dx \\
&\text{Integraing by parts,} \\
&= -x^2 \lambda e^{-\lambda x}\Big|_0^\infty + \int_0^\infty 2x e^{-\lambda x} dx \\
&= 0 + \frac{2}{\lambda}\int_0^\infty x\lambda e^{-\lambda x} dx \\
&= 0 - \frac{2}{\lambda}E(X) \quad = \quad \frac{2}{\lambda^2}
\end{aligned}$$

Therefore,

$$V(X) \quad = \quad E(X^2) - \Big(E(X)\Big)^2 \quad = \quad \frac{1}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 \quad = \quad \frac{1}{\lambda^2}.$$

3.2 Poisson Process

- Exponential distribution is often used as a model of distribution of times between the occurrence of successive events, such as time between two calls received in customer service desk, or time between two accidents on highway.

- Poisson process is a model for occurrence of events over time such that number of events in time interval of length t is Poisson distributed with parameter λt, and number of events in any non-overlapping time intervals are independent. Then time between two successive events is exponentially distributed with parameter λ.

3.3 Memoryless property:

- Let $X \sim \text{Exp}(\lambda)$. Then probability of X being more than $t + t_0$ given that it already is more than t_0 is the same as probability of X being more than t.

$$
\begin{aligned}
P(X \geq t + t_0 | X \geq t_0) &= \frac{P\Big((X \geq t + t_0) \cap (X \geq t_0)\Big)}{P(X \geq t_o)} \\
&= \frac{P(X \geq t + t_0)}{P(X \geq t_o)} \\
&= \frac{1 - P(X < t + t_0)}{1 - P(X < t_o)} \\
&= \frac{1 - P(X \leq t + t_0)}{1 - P(X \leq t_o)} \\
&= \frac{e^{-\lambda(t+t_0)}}{e^{-\lambda t_0}} \\
&= e^{-\lambda t} \\
&= 1 - P(X \leq t) \\
&= P(X \geq t).
\end{aligned}
$$

Example: Half Life of C14

about 5700

which means lambda is about .0001216

Min and Max of Exponential RV