MATH 320: Probability

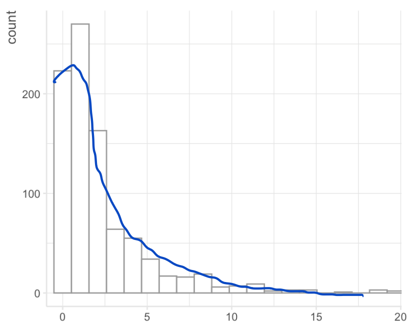
Lecture 11: Continuous Distributions

Chapters 3: Distributions (3.1, 3.2, 3.3)

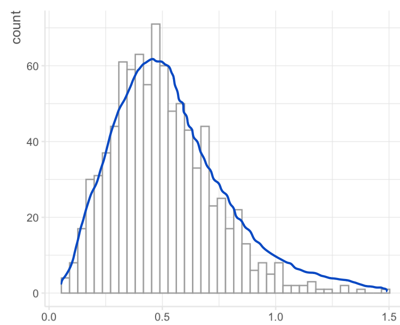
Introduction

- Recall that statistical distributions are used to model populations; this is the goal!
- Suppose an insurance company has the following data from two samples of $n = 900$ policyholder losses for two different policies.

  We can think of these as observations from loss random variables X and Y.



X



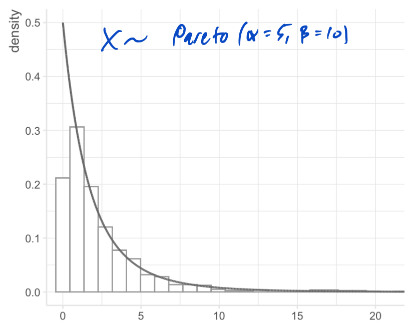
Y

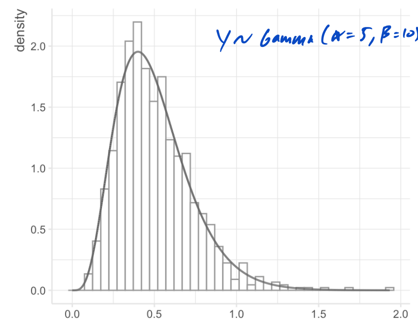
- The question is: How can me model the population that these losses are coming from?

  Said another way: In theory there is some distribution that these losses follow, and we want to figure out which family it is and what the parameter values are $\Longrightarrow$ Find equation for smooth curve.
- Once the researcher knows this, they can figure out lots of things that are useful in analyzing how losses will occur in the future. For example (suppose X is in thousands of dollars):
  - Can find the expected claim amount $E(X)$. Then use this to set a premium amount to ensure at least breakeven.
  - Can figure out the probability of "major" losses, say $P(X > 15 = \$15,000)$.
  - Or if there is a deductible of say \$2,000, can find the percentage of claims that the insurance company is not responsible to pay for: $P(X < 2 = \$2,000)$.

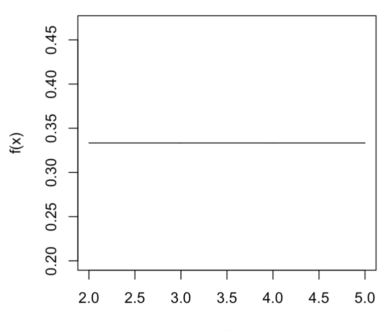
- If we already have data, we could simply approximate these values with empirical formulas:
  - $E(X) \approx \frac{1}{n}\sum_{i=1}^{n} x_i$ $= \frac{1}{900} \sum_{i=1}^{900} x_i$
  - $P(X > 15 = \$15,000) \approx$ $\frac{n(X>15)}{900}$
  - $P(X < 2 = \$2,000) \approx$ $\frac{n(x<2)}{900}$
- But these are just approximations and are based on a single sample of data. Having population information is much better and much more generalizable!

  Observation: When we overlay these density curves, they match the histograms excellently.



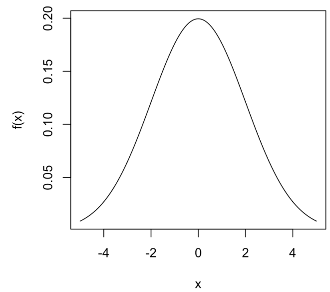


- Here is an overview of different population shapes and the distributions families that can be used to model them.
- If a population is distributed evenly between two numbers like



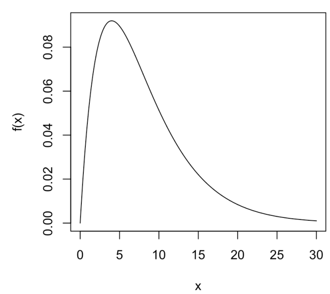
<u>uniform</u> distribution must be used.

- If a population is bell-shaped and symmetric like



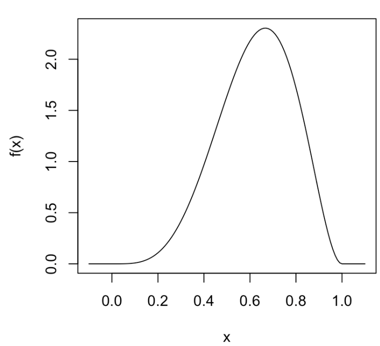
Normal or t distributions can be used, among others.

- If a population is skewed like



Gamma, lognormal or χ^2 distributions can be used, among others.

- If a population has bounded range (support) between two points and is not evenly distributed



Beta distribution can be used. This is useful when modeling probabilities $0 \le p \le 1$.

Very cool tangent!

- We saw in the previous example that we could model the data with a gamma distribution, specifically:
$$Y \sim \text{Gamma}\,(\alpha = 5, \beta = 10)$$

- In practice, once the researcher selects the family (gamma), they would then have to estimate the parameter values using techniques such as maximum likelihood estimation (will learn this next semester!).
  - The goal would be to have: $\hat{\alpha} \approx 5$ and $\hat{\beta} \approx 10$.
  - This strategy would be in the context of Classical statistics, where parameter values are fixed constants; there is another branch of statistics called Bayesian statistics.

- In Bayes, parameters are considered random variables and can have their own probability distributions. For example: $\alpha \sim \text{Exponential}\,(3)$ and $\beta \sim \text{Uniform}\,(1, 5)$.

Uniform distribution

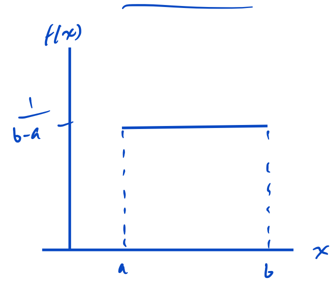
Definition

- The continuous uniform distribution is defined by spreading probability uniformly over an interval $[a, b]$ (X can also be thought of as the outcome when a point is randomly selected from the interval $[a, b]$).

- If $X \sim \text{Uniform}\,(a, b)$

  Min max

$$f(x \mid a, b) = \begin{cases} \dfrac{1}{b-a} & a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$



Area = bh

$1 = (b-a)h$

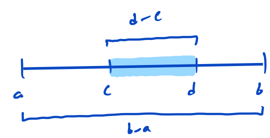
$\Rightarrow h = \frac{1}{b-a}$

  where the parameters a and b are real numbers.

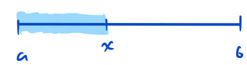
- Characteristics of a uniform distribution.
  - Constant probability over the entire interval.
  - Bounded support. $a, b \neq \infty$
  - Symmetric. $\Rightarrow$ mean = median

Probabilities

- The probability of any subinterval in the range (support) is proportional only to the length of the subinterval.



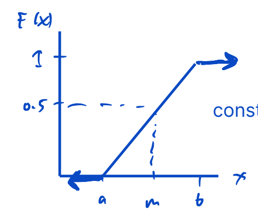
- For $a \leq c \leq d \leq b$
$$P(c \leq X \leq d) = \frac{d-c}{b-a}$$



- We can generalize this to find $P(X \leq x)$ for values of x in the interval $[a, b]$.
$P(X \leq x) =$ $P(a \leq X \leq x) = \frac{x-a}{b-a}$

- Then we can define the cdf $F(x)$ for a uniform random variable X on $[a, b]$ by:

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



constant Slope $= \frac{1}{b-a}$

Lifetime random variables and survival functions

- In many applied problems, the random variable interest is a time variable T. This time could represent:
  - Time until death of a person (a standard insurance application).
  - Time until the machine part fails.
  - Time until a disease ends.
  - Time it takes to serve a customer in a store.
- The uniform distribution doesn't give a very realistic model of human lifetimes, but is often used as an illustration of a lifetime model because of its simplicity.
- Example: Let T be the time from birth until death of a randomly selected member of a population. Assume that T has a uniform distribution on $[0, 100]$. Then

$$f(t) = \begin{cases} 1/100 & 0 \leq t \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad F(t) = \begin{cases} 0 & t < 0 \\ t/100 & 0 \leq t \leq 1 \\ 1 & t > 100 \end{cases}$$

  - The function $F(t)$ gives us the probability that the person dies by age t.
  - For example: The probability of death by age 57 is

$P(T \leq 57) = F(57) = \frac{57}{100} = 0.57$

- Most of us are interested in the probability that we will survive past a certain age. In this example, we might wish to find the probability that we survive beyond age 57. This is simply the probability that we do *not* die by age 57.

$P(T > 57) = 1 - F(57) = 1 - \frac{57}{100} = 0.43$

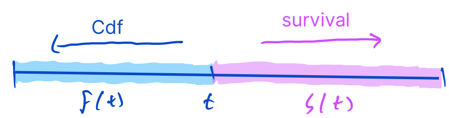
- The probability of surviving from birth past a given age t is called a **survival probability** and denoted by $S(t)$.

- Definition: The **survival function** is

$$\boxed{S(t) = P(T > t) = 1 - F(t)}$$

≈ Complement of cdf

Right probability =1- left probability



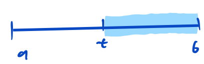
- The survival function for a uniform random variable is

$S(t) = 1 - F(t)$

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

$= \frac{b-a-(t-a)}{b-a}$

$\boxed{= \frac{b-t}{b-a}, \quad a \le t \le b}$



Mean and variance

- If $X \sim \text{Uniform}\,(a, b)$

$$E(X) = \frac{a+b}{2} \qquad V(X) = \frac{(b-a)^2}{12} \implies SD(X) = \sqrt{\frac{(b-a)^2}{12}}$$

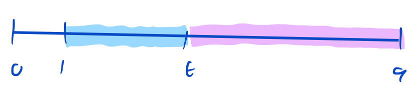
$=$ median m (midpoint of interval)

Summarizing example

- Let T be the time in months from initial use of a machine part until failure, where $T \sim \text{Uniform}\,(1, 9)$.

  a) Find the pdf, cdf and survival function of T.

$f(t) = \begin{cases} \frac{1}{8} & 1 \le t \le 9 \\ 0 & \text{otherwise} \end{cases}$

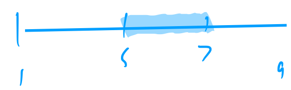


$F(t) = \frac{t-1}{8}, \quad 1 \le t \le 9$

$S(t) = \frac{9-t}{8}, \quad 1 \le t \le 9$

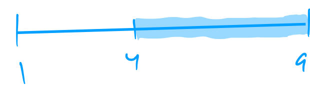
b) Find the probability the machine part fails between 5 and 7 months.

$P(5 \le T \le 7) = \frac{7-5}{8} = \frac{2}{8}$



c) Find the probability the machine part lasts longer than 4 months.

$P(T \ge 4) = S(4) = \frac{9-4}{8} = \frac{5}{8}$



d) Find the expected value and standard deviation for the lifetime of the machine part.

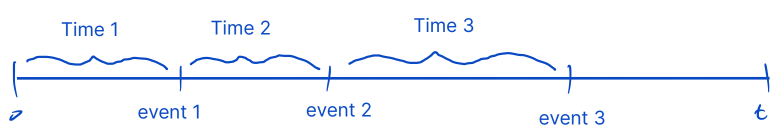
$E(T) = \frac{1+9}{2} = 5$

$V(T) = \frac{(9-1)^2}{12} = \frac{64}{12} \implies SD(T) = \sqrt{\frac{64}{12}}$

Exponential distribution

(Brief) Motivation

- Recall when observing a Poisson process, we counted the number of occurrences in a given interval. This number was a discrete random variable with a Poisson distribution.



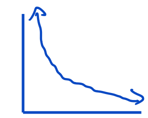
Time between events ~ Exponential

number of events ~ Poisson

- But not only is the number of occurrences a random variable, the waiting times between successive occurrences are also random variables. These are continuous and follow an exponential distribution.
- Formal statement: It can be shown that the exponential distribution gives the probability for the waiting time between successive Poisson events.

Derivation of density

$e^{-\lambda x}$



- The main part of the density function is an exponential decay function with parameter λ, which represents average number of events occurring (i.e. average rate of events) per unit of time in a Poisson process.
- To get this function to be a valid pdf, when integrated over the appropriate range $[0, \infty)$ it needs to equal 1.

Happens immediately   Never occurs

$$\int_0^\infty e^{-\lambda t}\, dt = -\frac{1}{\lambda} e^{-\lambda t} \Big|_0^\infty \quad f(\infty) - f(0)$$

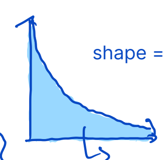
$$= 0 - \left(-\frac{1}{\lambda}\right)$$

$$\lambda \times [\ ] \quad = \left[\frac{1}{\lambda}\right] \times \lambda \quad \text{need to multiply through by } \lambda$$

$$\overset{\checkmark}{=} 1 \implies \int_0^\infty \lambda e^{-\lambda t}\, dt \overset{\checkmark}{=} 1$$

This process is the same as finding c such that: $\int_0^\infty c\, e^{-\lambda t}\, dt = 1$, $c = \lambda$

- The value c is called a **normalizing constant**.
  - All of the distributions we will study from now on have a normalizing constant, some of which are quite complex.
  - The purpose is the same though, functions of x determine the shape of a function (e.g. $e^{-\lambda x}$), then c converts it to a valid pdf.



Definition

- If $T \sim \text{Exponential}\,(\lambda)$

$$f(t \mid \lambda) = \lambda e^{-\lambda t}, \qquad t \geq 0 \quad \text{and} \quad \lambda > 0$$

- Characteristics of an exponential distribution.
  - Gives the probability for the waiting time between successive Poisson events.
  - Right-skewed density function.
  - Unbounded support. $0 \leq t < \infty$
  - The exponential random variable is a continuous version of the geometric random variable, which waits for the first Success in a discrete sample space. F F F S
  - Also has the memoryless property: $P(T > a + b \mid T > a) = P(T > b)$.

    Exponential and geometric are the only distributions with the memoryless property.

Parameter

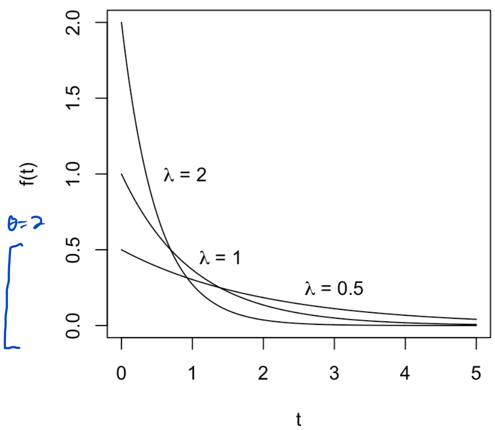
- As stated, λ is the rate at which events occur in a Poisson process. Here is how it affects the exponential density function.
- As λ increases, events happen more often in a time interval, which consequently means we are waiting less and less for the next event to occur.

Note different "scales"

$\theta = 0.5$

$\theta = 1$

$\theta = 2$



We see that for larger values of λ, there is more probability close to zero and less in the tail.

- Just like with some of the discrete distributions, there are different versions of the exponential distribution.
  - All versions have a random variable that is giving probabilities for waiting times. So they are not different in the way that we had two different versions of the geometric random variable (counting number of trials vs number of failures).
  - Rather they differ in what the parameters represent.

    1. The definition given above uses the **rate parameterization** of the exponential distribution. rate $= \lambda$
    2. There is also a **scale parameterization**, where the scale parameter θ is defined by

    $$\text{scale } \theta = \frac{1}{\text{rate}} = \frac{1}{\lambda} \quad \Longrightarrow \quad T \sim Exp(\theta = 1/\lambda) \quad \& \quad f(t) = \frac{1}{\lambda} e^{-t/\lambda}$$

  - This obviously will impact the formulas for the cdf, mean, variance, etc.

Probabilities

- Probabilities for the exponential distribution are easy to solve by hand (unlike the next distributions).
- Example: Accidents at a busy intersection occur at an average rate of $\lambda = 2$ per month according to a Poisson process. Let T be the random variable for the time between accidents.

  a) Find $f(t)$. $f(t) = 2e^{-2t}, \quad t \geq 0$

  b) Cdf $P(T \leq t)$: Find the probability that the waiting time for the next accident is less than 2.

$$P(T<2) = \int_0^2 2e^{-2t}\,dt = -e^{-2t}\Big|_0^2 \quad F(2) - F(0) = -e^{-2(2)} - (\underbrace{\cancel{-e^{-2(0)}}}_{=1}) = 1 - e^{-2(\underset{t}{\underset{\downarrow}{2}})}$$

  c) Survival $P(T > t)$: Find the probability that the waiting time for the next accident is longer than 1 month.

$$P(T>1) = \int_1^\infty 2e^{-2t}\,dt = -e^{-2t}\Big|_1^\infty \quad F(\infty) - F(1) = \underbrace{\cancel{\frac{-1}{e^{2(\infty)}}}}_{\to 0} - (-e^{-2(1)}) = e^{-2(\underset{t}{\underset{\downarrow}{1}})}$$

  d) Interval $P(a \leq X \leq b)$: Find the probability that the waiting time for the next accident is between 0.5 and 1.5 months

$$P(0.5 < T < 1.5) = \int_{0.5}^{1.5} 2e^{-2t}\,dt = -e^{-2t}\Big|_{0.5}^{1.5} \quad F(1.5) - F(0.5) = -e^{-2(1.5)} - (-e^{-2(0.5)}) = e^{-2(\underset{a}{\underset{\downarrow}{0.5}})} - e^{-2(\underset{b}{\underset{\downarrow}{1.5}})}$$

- Cdf and survival function:

  If $T \sim \text{Exponential}(\lambda)$

  $F(t) = 1 - e^{-\lambda t}$

  $S(t) = 1 - F(t)$

  $= 1 - (1 - e^{-\lambda t})$

  $= e^{-\lambda t}$

  Note that we could derive these formally:

$$F(t) = \int_0^t f(u)\,du \quad \& \quad S(t) = \int_t^\infty f(u)\,du$$

- Using the cdf: Recall once the cdf $F(x)$ is known for a random variable X, it can be used to find the probability that X lies in any interval since

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

For the exponential distribution, we have:

$$P(a \le T \le b) = F(b) - F(a) = (1 - e^{-\lambda b}) - (1 - e^{-\lambda a}) = e^{-\lambda a} - e^{-\lambda b}$$

Mean and variance

scale parameterization

$E(T) = \theta$

$V(T) = \theta^2$

- If $T \sim \text{Exponential}\,(\lambda)$

$$E(T) = \frac{1}{\lambda} \qquad V(T) = \frac{1}{\lambda^2} \implies SD(T) = \sqrt{\frac{1}{\lambda^2}} = \frac{1}{\lambda} = E(T)$$

- Example: Find the mean and variance for the previous accident example:

$T \sim Exp(\lambda = 2)$

$E(T) = \frac{1}{2}$ months

$V(T) = \frac{1}{2^2} = \frac{1}{4}$ months^2

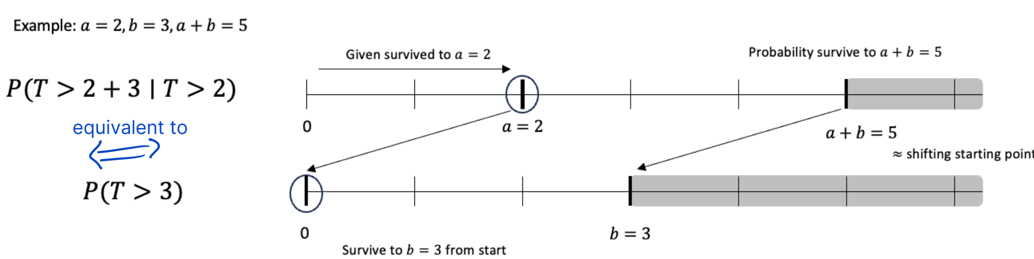
Memoryless property

- Just like the geometric distribution, the exponential distribution has the **memoryless property**.
- Theorem: If $T \sim \text{Exponential}\,(\lambda)$

$$P(T > a + b \mid T > a) = P(T > b)$$

- (Using a concrete example to help conceptualize) The probability that a bulb will operate for more than time $a + b$ time units, given that has already operated for at least a time units, is the same as the probability that a new bulb will operate for at least b time units.

  In practical terms, this means that the length of time a light bulb has already operated does not affect its chances of operating for additional time units.



- Proof:

$$P(T > a+b \mid T > a) = \frac{P(T > a+b)}{P(T > a)} = \frac{S(a+b)}{S(a)} = \frac{e^{-\lambda(\cancel{a}+b)}}{\cancel{e^{-\lambda a}}} = e^{-\lambda b} = S(b) \overset{\checkmark}{=} P(T > b)$$

- Example: Let T be the time to failure of a machine part, where $T \sim \text{Exponential}\,(\lambda = 0.001)$.
  (a) Given that the part has operated 100 hours, find the probability it will operate for 150 hours.

  $$P(T > 150 \mid T > 100) = P(T > 50) = e^{-0.001(50)}$$

  (b) Given that the part has operated 100 hours, find the probability it will operate for $100 + x$ hours $(0 \le x < \infty)$.

  $$P(T > 100+x \mid T > 100) = P(T > x) = e^{-0.001x} = S(x)$$

- Note that survival functions are unique, just like cdfs and pdfs.

  And the final expression in part (b) is the survival function $S(x)$ for a random variable which is exponentially distributed on $[0, \infty)$ with $\lambda = 0.001$; so it must have this distribution. This has a nice intuitive interpretation:
  - Lifetime of a new part $\sim$ $\text{Exp}(\lambda = 0.001)$ on $[0, \infty)$
  - Remaining lifetime of a 100 hr old part $\sim$ $\text{Exp}(\lambda = 0.001)$ on $[0, \infty)$

Waiting time for a Poisson process

- In the intro, we stated that the exponential distribution gives the waiting time between Poisson events. Here's why.
- If the number of events in a time period of length 1 is a Poisson random variable with parameter λ, then the number of events in a time period of length t is a Poisson random variable with parameter $\underline{\ \lambda t\ }$. Let this be the random variable X.

  This is a reasonable assumption (think about the accidents in 1 month vs 3 months example).
- We can find the probability of no accidents in an interval of length t easily:

  $$P(X=0) = \frac{e^{-\lambda t}\cancel{(\lambda t)^0}}{\cancel{0!}} = e^{-\lambda t}$$

  However, no accidents in an interval of length t is equivalent to saying the waiting time T for the next accident is greater than t. Thus

  $$P(X=0) = P(T > t) = S(t) = e^{-\lambda t} \implies T \sim \text{Exp}(\lambda)$$

- Example: Customers arrive in a certain shop according to an approximate Poisson process at a mean rate of 40 per 2 hours. What is the probability that the shopkeeper will have to wait more than 5 minutes for the arrival of the first customer?

$\to \lambda = \frac{40 \text{ customers}}{2 \text{ hours}} = \frac{20 \text{ customers}}{1 \text{ hour}} * \frac{1 \text{ hour}}{60 \text{ min}} = 1/3 \implies T \sim Exp(\lambda = 1/3)$

$\to P(T > 5) = S(5) = e^{-\frac{1}{3}(5)} \approx 0.1889$

Last example

- The lifetime of a machine has an exponential distribution with a mean of 3 years. The manufacturer is considering offering a warranty and considers two types of warranties.
  - W_1 – Warranty 1 pays 3 if the machine fails in the first year, 2 if the machine fails in the second year, and 1 if the machine fails in the third year, with no payment if the machine fails after 3 years.
  - W_2 – Warranty 2 pays $3e^{-T}$ at time T years.

  Find the expected warranty payment under each of the two warranties.

$\to E(T) = \frac{1}{\lambda} = 3 \to T \sim Exp(\lambda = 1/3)$

$\implies \lambda = 1/3$

$$\to W_1 = \begin{cases} P(T \le 1) & w_1 = 3 \\ P(1 \le T \le 2) & w_1 = 2 \\ P(2 \le T \le 3) & w_1 = 1 \end{cases} \implies E(W_1) = \sum w_1 f(w_1)$$

$= 3\,P(T \le 1) + 2\,P(1 \le T \le 2) + 1\,P(2 \le T \le 3)$

$= \quad + 2[P(T \le 2) - P(T \le 1)] + [P(T \le 3) - P(T \le 1)]$

$= P(T \le 1) + P(T \le 2) + P(T \le 3)$

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

≈ 1.4022

$\to W_2 = 3e^{-T} \implies E(W_2) = \int_0^\infty \underbrace{g(t)}_{= W_2} f(t)\, dt$

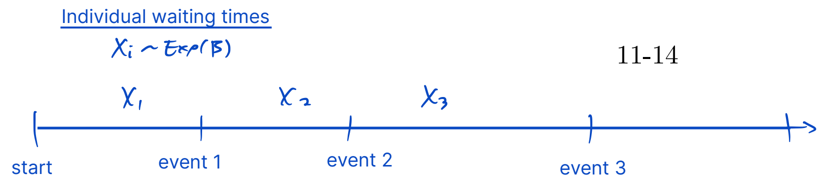
$= \int_0^\infty 3e^{-t} \frac{1}{3} e^{-\frac{1}{3}t}\, dt$

$= \int_0^\infty e^{-\frac{4}{3}t}\, dt$

$= -\frac{3}{4} e^{-\frac{4}{3}t} \Big|_0^\infty \quad F(\infty) - F(0)$

$= -\frac{3}{4}(0 - 1)$

$= 3/4$



Gamma distribution

(Brief) Motivation

$X = X_1 + X_2 + X_3$

$\downarrow \sim Gamma\ (\alpha = 3, \beta)$

- Relationship between the exponential distribution and the gamma distribution:

Recall

Geometric = number of trials for 1st success

negative Binomial = number of trials for r successes

$\sum_{i=1}^{r} Geo(p) = NB(r, p)$

analogous $\Longleftrightarrow$ same relationship

Exponential = waiting time for first event

Gamma = waiting time for α^{th} event

for a poisson process

$\perp\!\!\!\perp$ events

$\sum_{i=1}^{\alpha} Exp(\beta) = Gamma\ (\alpha, \beta)$

- The gamma distribution can also applied in other problems where the exponential distribution is useful like analysis of failure time of a machine part or survival time for a disease.

Definition

Scale parameterization

$f(x) = \dfrac{x^{\alpha-1} e^{-x/\beta}}{\beta^{\alpha}\, \Gamma(\alpha)}$

- If $X \sim \text{Gamma}\,(\alpha, \beta)$

Rate parameterization

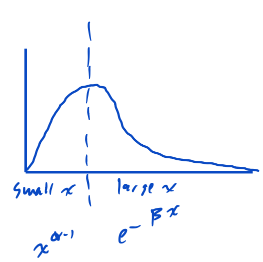
$$f(x \mid \alpha, \beta) = \underbrace{\frac{\beta^{\alpha}}{\Gamma(\alpha)}}_{\text{Normalizing constant}} \underbrace{x^{\alpha-1}\, e^{-\beta x}}_{\text{main function}}, \qquad x \geq 0 \quad \text{and} \quad \alpha, \beta > 0$$

- Intuitive idea behind pdf: $\to$ setting shape
  - The main part of $f(x)$ is $x^{\alpha-1}\, e^{-\beta x}$.

    As $x \to \infty$, $x^{\alpha-1} \to \infty$ for $\alpha > 1$.

    As $x \to \infty$, $e^{-\beta x} \to 0$ for $\beta > 0$.

    When x is small, $x^{\alpha-1}\, e^{-\beta x}$ is dominated by $x^{\alpha-1}$; when x is large, it is dominated by $e^{-\beta x}$. Thus as $x \to \infty$, $x^{\alpha-1} e^{-\beta x} \to 0$.



- Once we model the shape of the density function, we need to find the constant c such that

$$\int_0^\infty c\, x^{\alpha-1}\, e^{-\beta x}\, \mathrm{d}x = 1$$

By some calculus, we can show that $c = \frac{\Gamma(\alpha)}{\beta^\alpha}$, where

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}\, \mathrm{e}^{-x}\, \mathrm{d}x, \qquad \alpha > 0$$

- The quantity $\Gamma(\alpha)$ is known as the **gamma function**. Here is a summary of the $\Gamma(\alpha)$:
  - Gives a value for any $\alpha > 0$.
  - We can think of the gamma function as an extension of the factorial definition from the positive integers to all positive real numbers.

    Integration by parts can show

    ex) $\Gamma(2.5) = 1.5\,\Gamma(1.5) = 1.5\,(0.5)\,\Gamma(0.5) \approx 1.329$

$$\boxed{\Gamma(t) = (t-1)\,\Gamma(t-1)}$$

  - Whenever $\alpha = n$, where n is a positive integer, repeated integration by parts shows that

$$\Gamma(n) = (n-1)\,\Gamma(n-1) = (n-1)(n-2)\cdots(2)(1)\,\Gamma(1)$$

$$\int_0^\infty \underbrace{e^{-x}}_{\sim \mathrm{Exp}(\lambda=1)} dx = 1$$

Thus, when n is a positive integer, we have

$$\Gamma(n) = (n-1)!, \qquad n = 1, 2, \ldots$$

  - The pdf for the gamma distribution is essentially the gamma function as defined above, except it introduces another parameter in the exponential term, β.

- Characteristics of a gamma distribution:
  - Gives the probability for the waiting time until the α^{th} occurrence in a Poisson process.
  - Right-skewed density function.
  - Unbounded support. $0 \le x < \infty$

– Two important special cases of the gamma distribution:

1. When $\alpha = 1 \to X \sim \text{Exponential}\,(\beta)$. Thus gamma can be considered a generalized version of the exponential random variable.

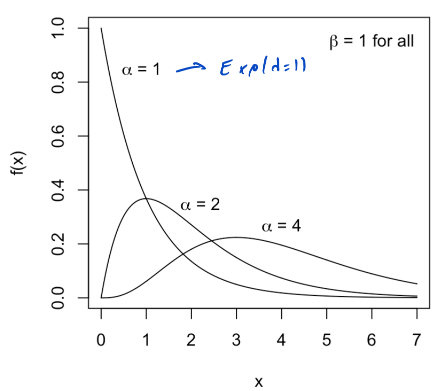
$$f(x) = \frac{\beta^1}{\Gamma(1)} x^{1-1} e^{-\beta x} = \beta e^{-\beta x} \sim Exp(\lambda = \beta)$$

rate

2. When $\alpha = r/2, \;\; \beta = 2 \to X \sim \chi^2(r)$ (read "Chi-squared"). This distribution is important in statistical inference, especially in analysis of variance (aka ANOVA), which is the basis of experimental design methods.

Parameters

- The parameters of the distributions we have studied so far have a direct interpretation, but here α and β do not. But we can still give some meaning to them. Here is how they affect the gamma density function.
- α = **shape** parameter.

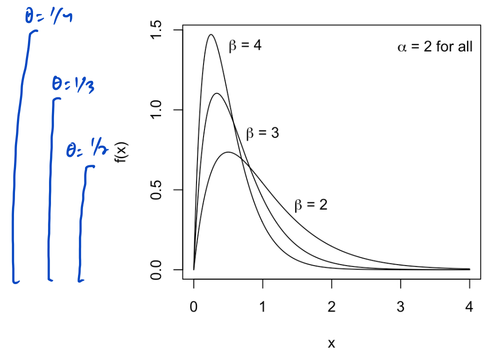


We see that for different values of α (and constant β), there are different <u>Shapes</u> of the gamma densities. Hence "shape" parameter.

Additionally, for larger values of α, the density increases longer out from zero because the polynomial part of the density function is more powerful for small x.

Notice that when <u>$\alpha = 1$</u> we have the familiar negative exponential curve. All exponential densities have this shape because of this fixed α.

- β = **rate** parameter (or $\theta = 1/\beta$ = **scale** parameter).



We see that for different values of β and constant α, the general Shapes of the gamma densities look similar, but the Scales are different. Hence "scale" parameter, which is equal to the inverse rate.

- Remember that the rate vs scale parameterizations will cause slight differences in the formulas for the pdf, mean, variance, etc. in other resources.

Probabilities

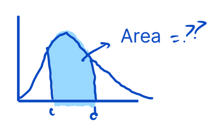
- Probabilities for the gamma distribution are difficult. If $X \sim \text{Gamma}(\alpha, \beta)$:
  - Can only be solved by hand if α is an integer.

    $\alpha = 1 \to$ Exponential. ✓

    $\alpha = 2 \to$ Requires integration by parts once. ✓ reasonable

    $\alpha > 2 \to$ Requires repeated integration by parts $\alpha - 1$ times. Possible, but not practical
  - If α is not an integer and $0 < c < d < \infty$, cannot calculate $P(c < X < d)$ by integration because it is impossible to give a closed-form expression for

    $$\int_c^d \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \mathrm{e}^{-\beta x} \, \mathrm{d}x$$

    Thus we must use statistical software to compute the probability that a gamma random variable falls in an interval.
  - Software do implement a general gamma cdf, but not anything one needs to know.

Mean and variance

Scale parameterization
$E(X) = \alpha\theta$
$V(X) = \alpha\theta^2$

- If $X \sim \text{Gamma}(\alpha, \beta) \qquad \alpha, \beta > 0$

$$E(X) = \frac{\alpha}{\beta} \qquad\qquad V(X) = \frac{\alpha}{\beta^2} \implies SD(X) = \sqrt{\alpha/\beta^2}$$

- If $\alpha = 1$, $E(X)$ and $V(X)$ for exponential follow.

  ✓ $= \frac{1}{\beta}$ ✓ $= \frac{1}{\beta^2}$

Summarizing examples:

1. Let X be random variable for the waiting time (in months) from the start of observation until the second accident at an intersection. Assume $X \sim \text{Gamma}\,(\alpha = 2, \beta = 3)$.

   (a) Find the pdf of X.

   $$f(x) = \frac{3^2}{\underbrace{\Gamma(2)}_{=1!}} x^{2-1} e^{-3x} = 9x e^{-3x}, \quad 0 \le x < \infty$$

   (b) Find $E(X)$ and $V(X)$.

   $$E(X) = \frac{\alpha}{\beta} = \frac{2}{3} \text{ months}$$

   $$V(X) = \frac{\alpha}{\beta^2} = \frac{2}{3^2} = \frac{2}{9} \text{ months}^2$$

   (c) Find the probability the total waiting time for the second accident is between 1 and 2 months.

   $$P(1 \le X \le 2) = \int_1^2 f(x)\,dx$$

   Integration by parts

   $u = 9x \quad dv = e^{-3x}$

   $du = 9 \quad v = -\frac{1}{3}e^{-3x}$

   $\left[uv - \int v\,du\right]_a^b$

   $$= \int_1^2 9x e^{-3x}\,dx$$

   $$= \left[9x\left(-\tfrac{1}{3}e^{-3x}\right) - \int -\tfrac{1}{3}e^{-3x}(9)\,dx\right]_1^2$$

   $$= \left[-3x e^{-3x} - e^{-3x}\right]_1^2 = F(2) - F(1)$$

   $$= \left(\underbrace{-3(2)e^{-3(2)} - e^{-3(2)}}_{-7e^{-6}}\right) - \left(\underbrace{-3(1)e^{-3(1)} - e^{-3(1)}}_{-4e^{-3}}\right)$$

   $$= 4e^{-3} - 7e^{-6}$$

   $$\approx 0.1818$$

2. Biologists investigating a stretch of desert find certain fossils according to a Poisson process at a mean rate of 500 per kilometer.
   1000 m

   What is the probability that the biologists will have to investigate more than 5 meters in order to find the first four fossils?

   $\rightarrow X \sim \text{Gamma}\left(\alpha = 4, \beta = \frac{500 \text{ fossils}}{1000 \text{ meters}}\right)$

   $\beta = 0.5$

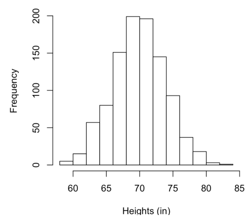
   $\rightarrow P(X > 5) = \int_5^\infty f(x)\,dx = 1 - \text{gammacdf}(\alpha = 4, \beta = 0.5, x = 5) \approx 0.7576$

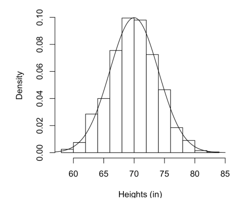
   * If there was a calculator function

Normal distribution

Applications

- The **normal distribution** is the most widely-used of all the distributions we have and will discuss. It can be used to model a wide range of natural phenomena that follow the "bell-shaped" pattern in their relative frequency distribution (histogram).





- Examples include variables such as test scores, physical measurements (height, weight, length) of organisms, and repeated measurements of the same quantity on different occasions or by different observers (measurement error), stock portfolio returns, insurance portfolio losses, etc.
- Every normal density curve has this shape, and the normal density model is used to find probabilities for all of the natural phenomena who histograms display this pattern.
- Random variables whose histograms are well-approximated by a normal density curve are called **approximately normal**.

Definition

mean   variance   normalizing constant   main function

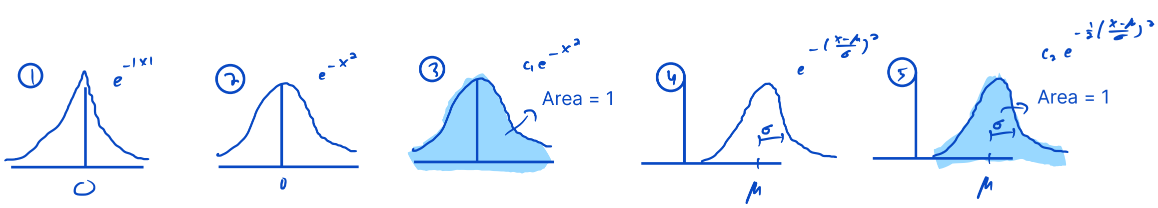
Notation
$\exp[a] = e^a$

- If $X \sim \text{Normal}\,(\mu, \sigma^2)$

$$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \qquad -\infty < x < \infty \quad \text{and} \quad -\infty < \mu < \infty, \quad \sigma > 0$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

- Intuitive idea behind pdf:
  1. Set the shape $\rightarrow$ Bell-shaped and symmetric (exponential decay from center in both directions)
  2. Change to smooth curve
  3. Find constant to make valid pdf.
  4. Add location and scale parameters to shift and scale the density, respectively.
  5. Adjust constant to take into account new parameters.



- Characteristics of a normal distribution.
  - Density function is bell-shaped and symmetric.
  - Unbounded support (range). $-\infty < x < \infty$

    But the density decreases exponentially as x runs away from the center. Thus, the normal distribution is also used for bounded data in practice (e.g. test scores). $\rightarrow$ ACT = [0, 36]

- The density function for the normal distribution is difficult to integrate as we will see. For example, to show that the normal distribution is a valid pdf, we need to use polar coordinates.

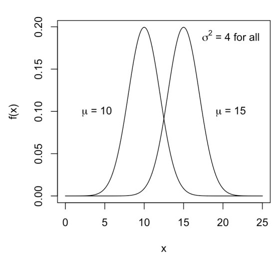
$$\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] dx \overset{\checkmark}{=} 1$$

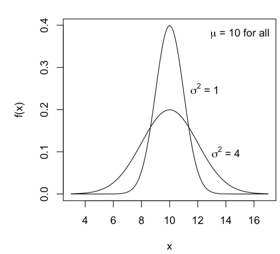
Parameters, expected value and variance

- The normal distribution is somewhat special in the sense that it's two parameters provide us with complete information about the exact shape (the spread) and location (where it's centered) of the distribution.
- If $X \sim \text{Normal}\,(\mu, \sigma^2)$

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

- These are really easy to derive using [redacted]. Moment generating functions (will learn later).
- Here is how they affect the normal density function.





- In practice, often the standard deviation σ is used rather than the variance σ^2. So it is common to see $X \sim \text{Normal}\,(\mu, \sigma)$.

Probabilities

- Suppose we are looking at a national exam whose scores X are approximately normal with $\mu = 500$ and $\sigma = 100$. To find the probability a score is between 600 and 750, we have to evaluate the integral:

$\rightarrow X \sim \text{Normal}\,(\mu = 500,\ \sigma = 100)$

$\rightarrow P(600 < X < 750) = \int_{600}^{750} \frac{1}{100\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-500}{100}\right)^2\right] dx = ??? \ldots$

- This cannot be done in closed-form (just like the gamma distribution), but it can be approximated using numerical methods using software, such as TI-84s as shown next.

Continuing ex) $\to X \sim Normal(\mu = 500, \sigma = 100)$

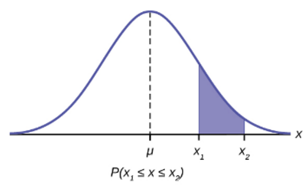
$\to P(600 < X < 750) =$ NormalCdf $\begin{pmatrix} lower = 600 \\ upper = 750 \\ \mu = 500 \\ \sigma = 100 \end{pmatrix} \approx 0.1524$

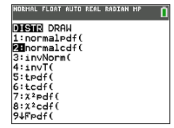
$\boxed{\text{* show work using calculator}}$

Finding Probabilities

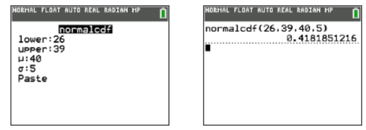
1. Choose correct Dist: 2ND → VARS
   - ALWAYS want **normalcdf()**
2. Enter in information (endpoints and parameters)
   - lower = lower boundary
   - upper = upper boundary
   - μ = Mean
   - **σ** = SD

- If you have TI-83, you would type normalcdf(lower, upper, mean, sd)





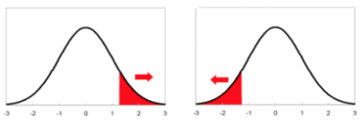
Ex) If $X \sim Normal(\mu = 40, \sigma = 5)$,
$P(26 < X < 39)$



*** Special Cases

If finding a left tailed probability, enter lower = -10000 (some really big negative number)

If finding a right tailed probability, enter upper = 10000 or 10^6 (really big positive number)



- In the olden days before such software was readily available, another way of finding normal probabilities involving tables of areas for a standard normal distribution was developed.

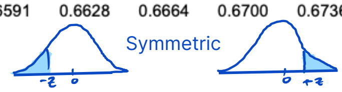
  NOTE: Exam P formula sheet only gives LEFT probabilities for POSITIVE Z values $\Longrightarrow$ We need to use properties of the normal curve to find probabilities for $-Z$. So you have to know this way.

Entries represent the area under the standardized normal distribution from $-\infty$ to z, Pr(Z<z)

The value of z to the first decimal is given in the left column. The second decimal place is given in the top row.

ex) $P(Z < 0.12) = 0.5478$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879



Left of $-Z$ = Right of $+Z$

$$
\begin{aligned}
P(Z < -0.12) &= P(Z > 0.12) \\
&= 1 - P(Z < 0.12) \\
&= 1 - 0.5478 \\
&= 0.4522
\end{aligned}
$$

- We will now cover this method, and the basic properties of the normal distribution which are behind it, in a series of steps.

- **Step 1: Linear transformation of normal random variables**
  - Theorem: If $X \sim \text{Normal}\,(\mu, \sigma^2)$ and $Y = aX + b$. Then
  $$\boxed{Y \sim \text{Normal}\,(a\mu + b, a^2\sigma^2)}$$
  $E(Y) = E(aX+b) = aE(X) + b$
  $V(Y) = V(aX+b) = a^2 V(X)$
  - It is easy to show that the mean and variance of Y using properties we have learned before.
  - But the crucial statement is that Y is also normally distributed.

    This is actually really easy to show with moment generating functions ([redacted]) and can also be shown using the pdf and a transformation of X.
  - Example: If $X \sim \text{Normal}\,(\mu = 10, \sigma^2 = 4)$, find the distribution of $Y = 0.5X - 3$.

    $E(Y) = E(0.5X - 3) = 0.5\underbrace{E(X)}_{=10} - 3 = 2$

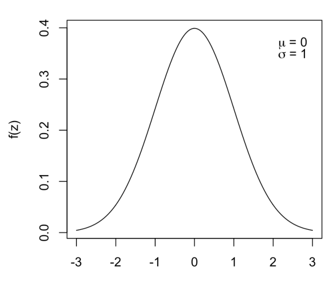
    $V(Y) = V(0.5X - 3) = 0.5^2 \underbrace{V(X)}_{=4} = 1$

    $\Rightarrow Y \sim \text{Normal}(\mu = 2, \sigma^2 = 1)$

- **Step 2: Transformation to a standard normal**
  - Using the linear transformation property of normal random variables, we can transform any normal random variable X with μ and standard deviation σ into a **standard normal** random variable with mean 0 and standard deviation 1.
  - The transformation used to do this is:
  $$\boxed{Z = \frac{X - \mu}{\sigma} = \frac{1}{\sigma}X - \frac{\mu}{\sigma}}$$
  - Using the previous theorem, we know $Z \sim \text{Normal}$ and can easily confirm the mean and standard deviation.

    $E(Z) = E\left(\frac{1}{\sigma}X - \frac{\mu}{\sigma}\right) = \frac{1}{\sigma}\underbrace{E(X)}_{=\mu} - \frac{\mu}{\sigma} = 0$ ✓

    $V(Z) = V\left(\frac{1}{\sigma}X - \frac{\mu}{\sigma}\right) = \frac{1}{\sigma^2}\underbrace{V(X)}_{=\sigma^2} = 1$ ✓
  - Thus, if we have the standard normal random variable $Z \sim \text{Normal}\,(0, 1)$
  $$f(z) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}z^2\right], \qquad -\infty < z < \infty$$



– This density function equation looks simpler, but still requires numerical integration.

– Forming $(X-\mu)/\sigma$ is known as **standardizing** X. Note that we can standardize any random variable, not just normals.

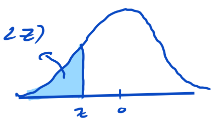
– Example: Suppose $X \sim \text{Exponential}(\lambda)$. $\Longrightarrow E(X) = 1/\lambda \quad V(X) = 1/\lambda^2$

$$Z = \frac{X-\mu}{\sigma} = \frac{X - 1/\lambda}{1/\lambda} = \lambda X - 1 \Rightarrow E(Z) = 0, \; V(Z) = 1$$

However Z doesn't necessarily have the same distribution of X. $Z \sim ??$

- **Step 3: Using z-tables**

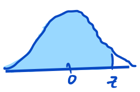
$P(Z < z)$



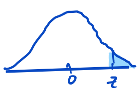
– Tables of areas under the density curve for the distribution of Z have been constructed for use in probability calculations.

– The table gives values for the cdf of Z, $F_Z(z) = P(Z \le z)$. LEFT probabilities

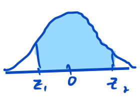
– Can find any probability we would like using the cdf:



Cdf (left) $P(Z \le z) =$ Table



Survival (right) $P(Z > z) = 1 - P(Z \le z)$



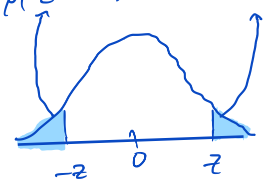
Interval $P(z_1 \le Z \le z_2) = F_Z(z_2) - F_Z(z_1)$

$= P(Z \le z_2) - P(Z \le z_1)$

Symmetry

$P(Z < -z) = P(Z > +z)$

$= 1 - P(Z < +z)$

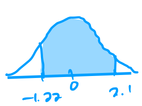


– Examples: $P(-1.22 \le Z \le 2.1) = P(Z \le 2.1) - P(Z \le -1.22)$

$= 0.9821 - (1 - P(Z < 1.22))$ $\quad$ ↳ $= 0.8888$

$= 0.9821 - 0.1112$

$= 0.8709$



$P(Z > -0.95)$

$= P(Z \le 0.95)$

$= 0.8289$

- **Step 4: Finding probabilities for any normal X**

– Once we know how to find probabilities for Z, we can use the transformation in step 1 to find probabilities for any normal random variable X with mean μ and standard deviation σ using the identity:

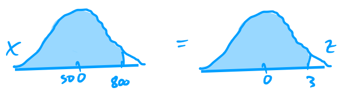
$$P(x_1 \le X \le x_2) = P\left(\frac{x_1-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{x_2-\mu}{\sigma}\right) = P(z_1 \le Z \le z_2)$$

where $z_1 = \dfrac{x_1-\mu}{\sigma}$ and $z_2 = \dfrac{x_2-\mu}{\sigma}$.

Z - scores = distance from mean in standard deviation units

- National exam examples: If $X \sim \text{Normal}\,(\mu = 500, \sigma = 100)$. Find the following probabilities:

  $P(X \leq 800) =$ $P\left(\frac{X-500}{100} \leq \frac{800-500}{100}\right) = P(Z \leq 3) = 0.9987$



  $P(600 \leq X \leq 750) =$ $P(1 \leq Z \leq 2.5) = P(Z \leq 2.5) - P(Z \leq 1)$

  $z_1 = \frac{600-500}{100} = 1$ $\qquad = 0.9938 - 0.8413$

  $z_2 = \frac{750-500}{100} = 2.5$ $\qquad = 0.1525$

  (calc answer $\approx$ 0.1524)

Percentiles

- We can also find percentiles of the standard normal distribution from the table.
- Recall for $0 \leq p \leq 1$ the **100p^{th} percentile** of X is the number x_p defined by

$$F(x_p) = p \longrightarrow F_Z(z_p) = p$$

- If X is a normal random variable with mean μ and standard deviation σ, then we can easily find x_p, using z_p and the basic relationship of X and Z.

$$z_p = \frac{x_p - \mu}{\sigma} \implies x_p = z_p\sigma + \mu$$

- Examples: If you scored in the 70^{th} percentile of test scores, what was your score?

$\rightarrow X \sim \text{Normal}(\mu = 500, \sigma = 100)$

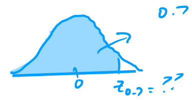
$x_{0.7} = 100\, z_{0.7} + 500 = \boxed{553}$

$z_{0.7} \to 0.53$

$\rightarrow F(z_{0.7}) = 0.7$

$\to$ 0.53 (find 0.7 probability on z - table (middle values)

$0.7 = P(Z \leq z_{0.7})$



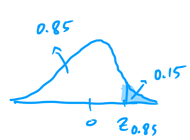
$\implies z_{0.7} = 0.53$

$\rightarrow$ Prob $= 0.7019$

Closest Z to get probability at least 0.7

$P(Z \leq z_{0.7}) \geq 0.7$

Find the test score corresponding to the top 15% of scores.



$\rightarrow$ Top 15% $= z_{1-0.15} = z_{0.85}$

$\rightarrow F(z_{0.85}) = 0.85 \longrightarrow x_{0.85} = 100\, z_{0.85} + 500 = 604$

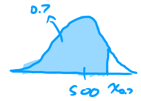
$z_{0.85} \to 1.04$

$\to 1.04$

$\implies$ gives prob $= 0.8508 \geq 0.8$ ✓

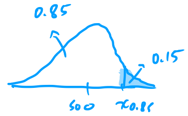
ex) → $P(X \le x_{0.7}) = 0.7$

→ $x_{0.7} = \text{invNorm}\begin{pmatrix} \text{area} = 0.7 \\ \mu = 500 \\ \sigma = 100 \end{pmatrix} = 552.44$



→ $P(X \le x_{0.85}) = 0.85$

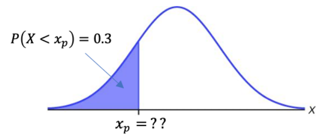
$\Rightarrow x_{0.85} = 603.64$



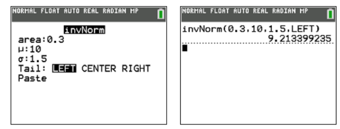
- This can also be done using software such as TI-84s.

<u>Finding Percentiles</u>

1. Choose invNorm(): 2ND → VARS
2. Enter in information (area and parameters)
   - area = Probability (LEFT, percentile!)
   - μ = Mean
   - **σ** = SD



Ex) If $X \sim Normal(\mu = 10, \sigma = 1.5)$, $x_{0.30} \approx 9.21$



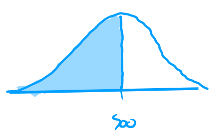
If you have TI-83, you would type invNorm(area,mean,sd)

*** If you have a right tail (upper probability), you need to rewrite as probability to the left!

Harder examples

1. Let $X \sim \mathrm{N}(\mu, \sigma^2)$, $P(X < 500) = 0.5$[redacted] and $P(X > 650) = 0.0228$[redacted]

   Find μ and σ^2.



→ symmetric ⇒ mean = median = 500 = μ

→ $P(X > 650) = 0.0228$

⇒ $P(X < 650) = 1 - 0.0228$

$P\left(\frac{X-500}{\sigma} < \frac{650-500}{\sigma}\right) = 0.9772$

$P(\underset{\hookrightarrow\, = 2}{Z < z}) = 0.9772$

$\Rightarrow \frac{650-500}{\sigma} = 2 \Rightarrow \sigma = 75$ & $\sigma^2 = 5625$

2. Let $X \sim \mathrm{N}(\mu = 25, \sigma^2 = 36)$. Find c such that $P(|X -$ 25 [redacted]$| \le c) = 0.9544$

$P(|X-25| \le c) = P\left(\frac{-c}{6} \le \frac{X-25}{6} \le \frac{c}{6}\right) = P\left(\frac{-c}{6} \le Z \le \frac{c}{6}\right)$

$= P(Z \le \frac{c}{6}) - P(Z \le -\frac{c}{6})$

$= \quad \downarrow \quad - (1 - P(Z \le \frac{c}{6}))$

$= 2P(Z \le \frac{c}{6}) - 1 = 0.9544$

$\Rightarrow P(Z \le \frac{c}{6}) = \frac{1.9544}{2}$

$P(\underset{\hookrightarrow\, = 2}{Z \le z}) = 0.9772$

$\Rightarrow \frac{c}{6} = 2 \Rightarrow c = 12$

3. If $X \sim \mathrm{N}(\mu = 1, \sigma^2 = 16)$, find $P(X^2 - 4X < 21)$.

$P(X^2 - 4X < 21) = P(X^2 - 4X - 21 < 0)$

$= P[\underbrace{(X-7)(X+3)}_{} < 0]$

true when

$(+)(-)$ or $(-)(+)$

$x > 7 \Rightarrow$ ✗

$x > -3 \Rightarrow$ ✓

$x < -3 \Rightarrow$ ✗

$= P(-3 < X < 7)$

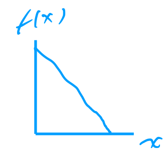
$= P\left(\frac{-3-1}{4} \le \frac{X-1}{4} \le \frac{7-1}{4}\right)$

$= P(-1 \le Z \le 1.5)$

NormalCdf $\begin{pmatrix} & X & Z \\ \text{lower} = & -3 & -1 \\ \text{upper} = & 7 & 1.5 \\ \mu = & 1 & 0 \\ \sigma = & 4 & 1 \end{pmatrix} \approx 0.7754$

Central Limit Theorem

Sums of independent, identically distribution random variables



- We will now demonstrate one reason why the normal distribution is so useful in applications.
- Motivating example: Recall the previous straight-line density example, where the random variable X represented the loss on a single warranty insurance policy. It was not normally distributed.

  We found that $E(X) = \frac{100}{3}$ and $V(X) = \frac{5{,}000}{9}$ and were able to find probabilities for X. However, this information applies only to a single policy.

  The company selling insurance has more than one policy and must look at its total book of business by adding up all of the losses on all policies.

- Generalizing this scenario: Suppose the loss on a single insurance policy follows a non-normal density X. Companies will have say 1,000 policies and be interested in the total loss S, which is the sum of the losses on all individual policies X_i:

$$S = X_1 + X_2 + \cdots + X_{1000}$$

- If we assume that all of the policies are *iid* (independent and follow the same distribution), then the **Central Limit Theorem (CLT)** shows the sum is approximately normal (even though the individual policies X_i are not) $\Longrightarrow$ which means we can use normal probability methods to find probabilities for the total loss.

  Normalcdf() and z-tables

- **Central Limit Theorem** Let $X_1, \ldots, X_n$ be independent random variables, all of which have the same probability distribution and thus the same mean μ and variance σ^2. If n is large, the sum

$$S = X_1 + X_2 + \cdots + X_n$$

  will be approximately normal with mean $n\mu$ and variance $n\sigma^2$.

  Written succinctly: If $X_i \overset{iid}{\sim} f(x)$ with mean μ + variance σ^2

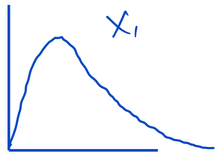
  then $S = \sum_{i=1}^{n} X_i$

  $\overset{\cdot}{\sim}$ Normal $(n\mu, n\sigma^2)$ if n is large

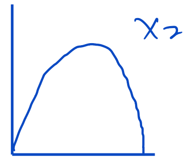
  approx

- Notes about CLT:
  - This is a super powerful theorem!
  - X_i do not have to be normally distributed. When this is the case, the CLT results in an approximately normal distribution.
  - How large must n be? This depends on how close the original distribution is to the normal.

ex) $n_1 >> n_2$ for accurate CLT approximations



skewed



Symmetric

Some elementary statistics books define $n > 30$ as "large". This will not always be the case. For example, skewed distributions will require larger values of n compared to symmetric distributions in order for the results of the CLT to be decent.

– In general, as n increases, approximations based on the CLT get better and better.

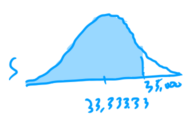
– If $X_i \overset{iid}{\sim} \text{Normal}\,(\mu, \sigma^2)$, then the CLT results in an exactly normal distribution for all n.

In theory
As $n \to \infty$
$S \sim$ Normal
exactly

- Return to motivating example: Find the distribution of the sum of losses S. $E(X) = \frac{100}{3}$ $V(X) = \frac{5000}{9}$

$$S = \sum_{i=1}^{1000} X_i \longrightarrow S \sim \text{Normal}\left(\mu = 1000\left(\tfrac{100}{3}\right) \simeq 33{,}333.33,\ \sigma = \sqrt{1000\left(\tfrac{5000}{9}\right)} = 745.35\right)$$

Find the probability total losses are less than \$35,000.



$$P(S < 35{,}000) = \text{Normalcdf}\begin{pmatrix} \text{lower} = -10{,}000 \\ \text{upper} = 35{,}000 \\ \mu = 33{,}333.33 \\ \sigma = 745.35 \end{pmatrix} \simeq 0.9873$$

This shows the company that it is not likely to need more than \$35,000 to pay claims, which is helpful in planning.

More examples

1. Another example: Suppose the number of claims filed on for a particular policy follow a Poisson distribution with a mean of 2 claims per year and the company has a portfolio of 500 active policies this year, which are assumed to be independent.

   (a) Find the distribution of the total number of filed claims for the entire portfolio.

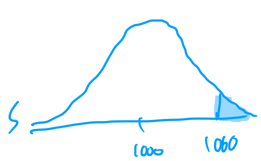
$\to X_i \overset{iid}{\sim} \text{Poisson}(\lambda = 2)$

$\Rightarrow E(X) = V(X) = \lambda = 2$

$$\Rightarrow S = \sum_{i=1}^{500} X_i \sim \text{Normal}\left(\mu = 500(2) = 1000,\ \sigma^2 = 500(2) = 1000\right)$$

☆ By CLT ☆

   (b) Find the probability there will be more 1060 claims this year.



$$P(S > 1060) = \text{Normalcdf}\begin{pmatrix} \text{lower} = 1060 \\ \text{upper} = 10000 \\ \mu = 1000 \\ \sigma = \sqrt{1000} \end{pmatrix} \simeq 0.0289$$

2. Two instruments are used to measure the height, h, of a tower. The error made by the less accurate instrument is normally distributed with mean 0 and standard deviation $0.0056h$. The error made by the more accurate instrument is normally distributed with mean 0 and standard deviation $0.0044h$. Assuming the two measurements are independent random variables, what is the probability that their average value is within $.005h$ of 0?

$\rightarrow X_1 \sim Normal(\mu = 0, \sigma = 0.0056h)$
$X_2 \sim Normal(\mu = 0, \sigma = 0.0044h)$ $\Big\}$ $X_1 \perp\!\!\!\perp X_2$

$\rightarrow P(-0.005h \le \frac{X_1+X_2}{2} \le 0.005h) = P(-0.01h \le X_1+X_2 \le 0.01h)$

$\hookrightarrow X_1 + X_2 \sim Normal(\mu = \mu_1 + \mu_2, \sigma^2 = \sigma_1^2 + \sigma_2^2)$

$= 0 + 0 = 0$, $\sigma^2 = (0.0056h)^2 + (0.0044h)^2$

$= Normal\,cdf\left(\begin{array}{l} lower = -0.01\cancel{h} \\ upper = 0.01\cancel{h} \\ \mu = 0 \\ \sigma = \sqrt{(0.0056\cancel{h})^2 + (0.0044\cancel{h})^2} \\ h\text{'s cancel} \end{array}\right)$

$= 0.8347$

Summary

- In general, normal distribution is quite valuable because it applies in so many situations were independent and identical components are being added.

  Even though normal is continuous, the CLT works with discrete distributions as we saw above.

- And the CLT showed us why so many random variables are approximately normally distributed.

  This occurs because many useful random variables are themselves sums of other independent random variables.

→ Contextually → Total claims = Sum of individual claims

Total losses = Sum of individual losses

→ Distributions → $\sum_{i=1}^{n}$ Bernoulli $\sim$ Binomial (n, p)

$\sum_{i=1}^{r}$ Geometric $\sim$ Negative Binomial (r, p)

$\sum_{i=1}^{\lambda}$ Poisson $(\lambda = 1)$ $\sim$ Poisson (λ)

$\sum_{i=1}^{\alpha}$ Exponential $\sim$ Gamma (α, β)

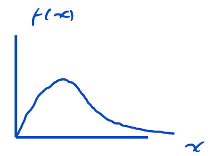
for large summations, we can approximate the ending distributions with normal

Lognormal distribution

(Brief) Motivation

- An alternative to the gamma distribution for asymmetric data is the lognormal distribution.

  This distribution is more widely used than gamma to model skewed populations because we can take advantage of the properties of the normal distribution.



- Examples include insurance claim severity and investment returns.

Definition

- A random variable is called **lognormal** if its natural logarithm is normally distributed.

  $\boxed{\text{If } Y \sim \text{Lognormal} \Longleftrightarrow \ln(Y) \sim \text{Normal}\,(\mu, \sigma^2).}$

  $\ln(Y) = \ln(e^{\text{normal}})$

  $\downarrow = \text{normal}$

- Stated another way: A random variable Y is lognormal if $Y = \mathrm{e}^X$ for some normal random variable X with mean μ and variance σ^2.

  If $X \sim \text{Normal}\,(\mu, \sigma^2)$ and $Y = \mathrm{e}^X \Longleftrightarrow Y \sim \text{Lognormal}$

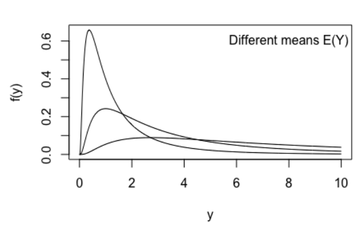
$$f(y \mid \mu, \sigma^2) = \frac{1}{y\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\ln(y) - \mu)^2}{2\sigma^2}\right], \qquad y \geq 0 \quad \text{and} \quad -\infty < \mu < \infty, \quad \sigma > 0$$

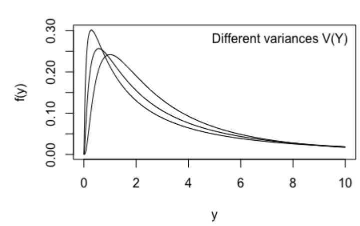
very similar to normal pdf except

- Characteristics of a lognormal distribution.
  - Right-skewed density function.
  - Unbounded support. $0 \leq x < \infty$

Parameters, expected value and variance

- Note that the parameters μ and σ^2 represent the mean and variance of the normal random variable X which appears in the exponent.
- Here is how they affect the lognormal density function.





- The mean and variance of the actual lognormal distribution Y are:

  If $X \sim \text{Normal}\,(\mu, \sigma^2)$ and $Y = \mathrm{e}^X$

  $$E(Y) = \mathrm{e}^{\mu + \frac{\sigma^2}{2}} \qquad V(Y) = \mathrm{e}^{2\mu+\sigma^2}(\mathrm{e}^{\sigma^2} \text{[redacted]}) \Rightarrow SD(X) = \sqrt{V(Y)}$$

Probabilities

- Just like with the normal, we cannot integrate the pdf, but we do not need to. The cdf can be found directly from the cdf for the normally distributed exponent.

- $F_Y(y) =$ $P(Y \le y) = P(e^X \le y) = P[\ln(e^X) \le \ln(y)] = P(X \le \ln(y))$ ↳ Normal

- This means we just need to do algebra in the probability statement, then use normal-cdf() or z-tables like normal on the resulting number.

Example

- Let the claim severity $X \sim \text{Normal}\,(\mu = 7, \sigma^2 = \text{[redacted]})$ and $Y = \mathrm{e}^X$.

  0.25

  (a) Find $E(Y)$ and $V(Y)$.

  $E(Y) = e^{7 + 0.25/2} = 1{,}242.65$ $\qquad V(Y) = e^{2(7)+0.25}(e^{0.25} - 1) = 438{,}584.80$

  (b) Find the probability a claim is less than or equal to 1300.

  $P(Y \le 1300) = P(e^X \le 1300) = P(X \le \ln(1300)) = P\left(Z \le \frac{\ln(1300) - 7}{0.5}\right)$

  $= \text{Normalcdf}\begin{pmatrix} \text{lower} = -10000 \\ \text{upper} = \ln(1300) \\ \mu = 7 \\ \sigma = 0.5 \end{pmatrix} \approx 0.6337$

  (c) Find the probability a claim is between 900 and 1200.

  $P(900 \le Y \le 1200) = P(900 \le e^X \le 1200)$

  $= P(\ln(900) \le X \le \ln(1200))$

  $= \text{Normal cdf}\begin{pmatrix} \text{lower} = \ln(900) \\ \text{upper} = \ln(1200) \\ \mu = 7 \\ \sigma = 0.5 \end{pmatrix}$

  ≈ 0.225

Beta distribution

(Brief) Motivation

- The beta distribution is defined on the interval $[0, 1]$.
- Thus it is often used as a model for probabilities such as the proportion of impurities in a chemical product or the proportion of time that a machine is under repair

Definition

- If $X \sim \text{Beta}\,(\alpha, \beta)$

normalizing constant

main function

$$f(x \mid \alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1}, \qquad 0 \leq x \leq 1 \quad \text{and} \quad \alpha, \beta > 0,$$

where

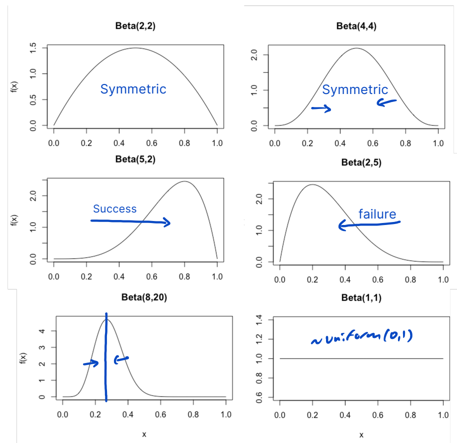
$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,\mathrm{d}x = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

- Intuitive idea behind pdf: Like the gamma distribution, we construct the formula as the function of x and then find the constant out from to make the density valid.
- Characteristics of a beta distribution.
  - In general, asymmetric density function.
  - Bounded support. $0 \leq x \leq 1$

Parameters, expected value and variance

- [redacted] The beta random variable X is often used as a model for probability and proportion. Thus, x represents the probability of success and $1 - x$ represents the probability of failure.

  So, α and β sort of represent the chances (or weight) of success and failure, respectively.
- For example, when $\alpha = 5$ and $\beta = 2$, the chance of success is stronger than the chance of failure. So the mode and mean lean towards 1. In addition, as the parameters increase, the distributions tighten around the expectations.



- If $X \sim \text{Beta}\,(\alpha, \beta)$

$$E(X) = \frac{\alpha}{\alpha+\beta} \qquad V(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \implies SD(X) = \sqrt{V(X)}$$

Probabilities

- When α and β are integers greater than 1, the cdf can be found by integrating a polynomial. Else we need to use software.

Example

- A management firm handles investment accounts for a large number of clients. The percent of clients who contact the firm for information or services in a given month is a beta random variable with $\alpha = 4$ and $\beta = 3$.

  (a) Find the pdf $f(x)$ and the cdf $F(x)$.

$$\to f(x) = \frac{1}{B(4,3)} x^{4-1}(1-x)^{3-1} = 60x^3(1-x)^2 = 60x^3(1-2x+x^2) = 60x^3 - 120x^4 + 60x^5, \quad 0 \le x \le 1$$

$$B(4,3) = \frac{\Gamma(4)\Gamma(3)}{\Gamma(4+3)} = \frac{3!\,2!}{6!}$$

$$\to F(x) = \int_0^x f(t)\,dt = \int_0^x [60t^3 - 120t^4 + 60t^5]\,dt = 15t^4 - 24t^5 + 10t^6 \Big|_0^x \quad F(x) - F(0) = 15x^4 - 24x^5 + 10x^6, \quad 0 \le x \le 1$$

  (b) Find the probability the percent of clients contacting the firm in a month is less than 40%.

$$F(0.4) = 15(0.4)^4 - 24(0.4)^5 + 10(0.4)^6 = 0.1792$$

  (c) Find the expected value and variance for the percent of clients contacting the firm.

$$\to E(X) = \frac{\alpha}{\alpha+\beta} = \frac{4}{4+3} = \frac{4}{7}$$

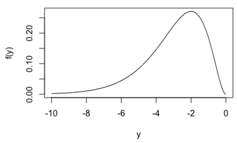
$$\to V(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \frac{3(4)}{(3+4)^2(3+4+1)} \approx 0.0306$$

Transformations

- We have covered the standard shapes that can be modeled by these distributions. To be more flexible, we can use functions of these random variables to model variations of the standard shapes.
- Examples:

1. If a population is left-skewed and the range is $-\infty < Y \leq 0$. How can we model the population?

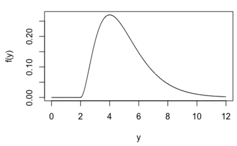
$\rightarrow X \sim$ Gamma OR $X \sim$ Lognormal



$\rightarrow Y = -X$

2. If a population is right-skewed, but the range is $a \leq Y$ $<$ [redacted] ∞, $a \neq 0$. How can we model the population?

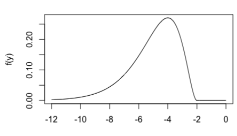
$\rightarrow X \sim$ Gamma OR $X \sim$ Lognormal



$\Rightarrow Y = X + a$

3. If a population is left-skewed and the range is $-\infty < Y \leq a$, $a \neq 0$. How can we model the population?

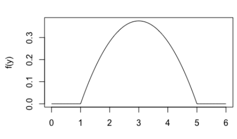
$\rightarrow X \sim$ Gamma OR $X \sim$ Lognormal



$\rightarrow Y = -(X + a)$

4. If a population is bounded between $a \leq Y \leq b$. How can we model the population?

$\rightarrow X \sim$ Beta



$\rightarrow \underset{[a,b]}{Y} \longrightarrow \underset{[0,1]}{X}$:

$a \leq Y \leq b$

$0 \leq Y - a \leq b - a$

$0 \leq \frac{Y-a}{b-a} \leq 1$

$0 \leq X \leq 1$

$\rightarrow X = \frac{Y-a}{b-a}$

$\Rightarrow Y = (b-a)X + a$