Common Continuous random variables

CE 311S

Earlier, we saw a number of distribution families

- Binomial
- Negative binomial
- Hypergeometric
- Poisson

These were useful because they represented common situations: if we recognized a binomial random variable, we had formulas to give the pmf, mean, variance, and so forth.

The same is true for continuous distributions as well

Some of the distributions we will see are:

- Uniform (last class)
- Exponential
- Gamma
- Normal
- Lognormal
- Chi-squared

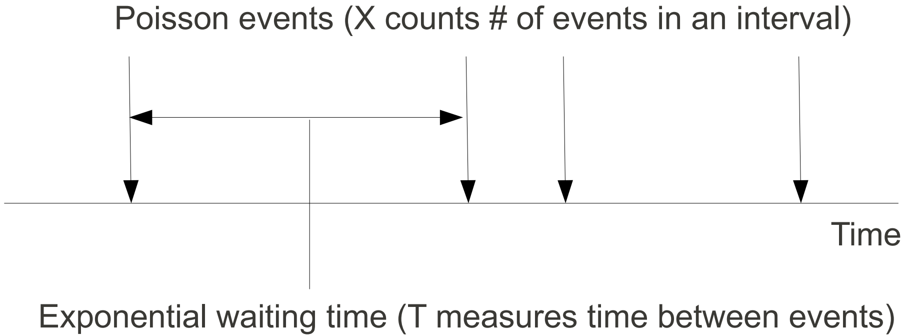
However, unlike the discrete distributions it isn't always easy to derive the pdf from first principles. In these situations, I will give the pdf or cdf to you.

EXPONENTIAL DISTRIBUTION

One exception to this rule is the **exponential distribution**.

The exponential distribution can represent the waiting time between Poisson events.

The Poisson distribution counts the number of times an event happens.



The exponential distribution describes the time **between** successive events.

We can use this fact to derive the exponential distribution. Assume we "start the clock" ($x = 0$) right after an event happens, and let X be the (random) time the *next* event happens, assuming a Poisson process which occurs at a rate λ.

Let's derive the cumulative distribution function $F(x) = P(X \leq x)$ (for any $x > 0$; $F(x) = 0$ if $x \leq 0$).

This is Method I for finding pdfs: the easiest way to find the pdf is to start with the **cdf**, which is of a form similar to what we've used previously to calculate probabilities.

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

However, the event $X > x$ simply means that there were no events in the time $[0, x]$ (the true time for the next event X is greater than the current time x)

The expected number of events during this time interval is λx.

From the Poisson distribution, the probability of seeing 0 events is $e^{-\lambda x}(\lambda x)^0/0!$ or simply $e^{-\lambda x}$.

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

So, the cdf for the exponential distribution is

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Therefore, the pdf is obtained by

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

So, the cdf for the exponential distribution is

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Therefore, the pdf is obtained by derivatizing

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

So, the cdf for the exponential distribution is

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Therefore, the pdf is obtained by ~~derivatizing~~ differentiating the cdf.

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

So, the cdf for the exponential distribution is

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Therefore, the pdf is obtained by ~~derivatizing~~ differentiating the cdf.

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

$f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & \text{otherwise} \end{cases}$ is a valid pdf because it is nonnegative and integrates to one:

$$\int_0^\infty f(x)\ dx = \lim_{x \to \infty} F(x) = \lim_{x \to \infty} (1 - e^{-\lambda x}) = 1.$$

We can calculate the mean and variance of the exponential distribution by integration:

$$E[X] = \int_0^\infty x\lambda e^{-\lambda x} \; dx = \frac{1}{\lambda}$$

$$V[X] = \int_0^\infty x^2\lambda e^{-\lambda x} \; dx - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

The equation for $E[X]$ should be intuitive: if events occur at a rate λ, the average time *between* events is $1/\lambda$.

Example

Taxis arrive at an average rate of one every five minutes. What is the probability that I have to wait between 5 and 10 minutes for one?

Let T be the time until the next taxi arrives; T has an exponential distribution, but what is λ?

$\lambda = 1/E[T] = 1/5$

$P(5 \leq T \leq 10) = F(10) - F(5) = (1 - e^{-10/5}) - (1 - e^{-5/5}) = 0.23$

Because Poisson events are independent, the exponential distribution should be "memoryless"

This means that *no matter how long I've been waiting*, nothing should change:

- The expected waiting time is still the same.
- The probability of waiting any period of time is still the same.
- ... and so forth.

In some ways, this is a good thing. We can start watching at any time without having to worry about when the last event happened.

Is it memoryless?

Memorylessness can be tested by checking if
$P(X \geq x + x_0 \mid X \geq x_0) = P(X \geq x)$ for any value of x_0.

$$P(X \geq x + x_0 \mid X \geq x_0) = \frac{P((X \geq x + x_0) \cap (X \geq x_0))}{P(X \geq x_0)}$$

$$= \frac{P(X \geq x + x_0)}{P(X \geq x_0)}$$

$$= \frac{1 - F(t + x_0)}{F(x_0)}$$

$$= \frac{e^{-\lambda(x+x_0)}}{e^{-\lambda x_0}}$$

$$= e^{-\lambda x} = P(X \geq x)$$

THE GAMMA FUNCTION

The exponential distribution is a special case of a more general continuous distribution, the **gamma distribution**.

Before discussing the gamma distribution, we need to define the **gamma function.**

Remember the factorial function $n!$?

The gamma function is a *continuous* function which shows this same behavior.

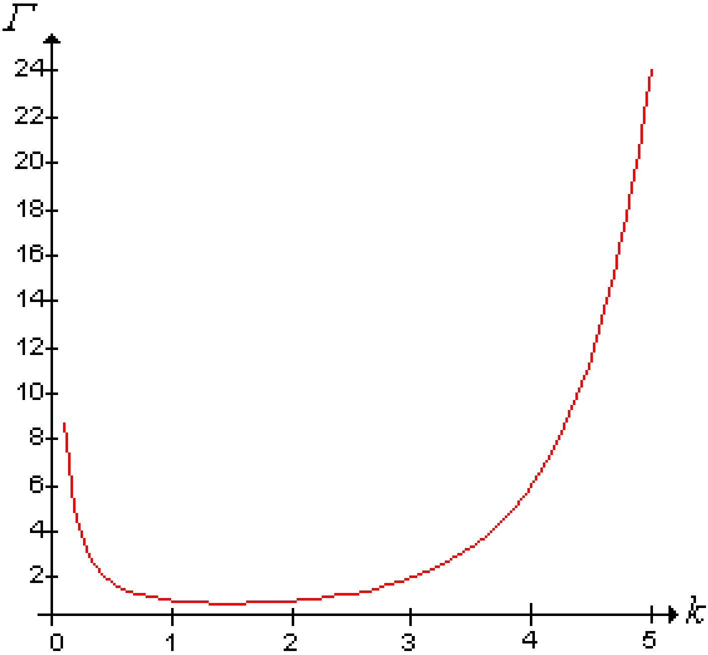
Specifically, define

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx$$

for any $\alpha > 0$.

The gamma function has the following properties:

1. $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$ for any $\alpha > 1$
2. If n is a positive integer, $\Gamma(n) = (n-1)!$
3. $\Gamma(1/2) = \sqrt{\pi}$


Γ
24
22
20
18
16
14
12
10
8
6
4
2
0
1
2
3
4
5
k

GAMMA DISTRIBUTION

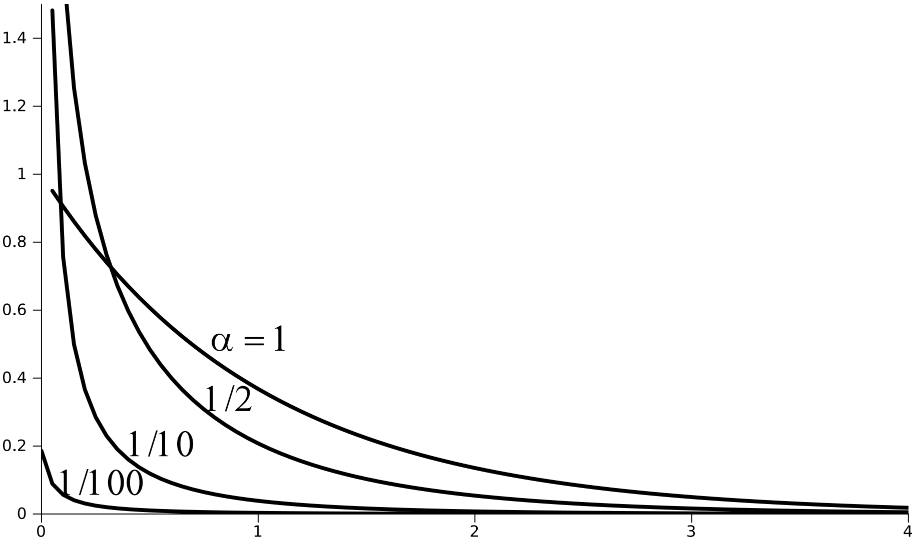
The gamma distribution is described by the following pdf:

$$f(x) = \begin{cases} \frac{1}{\Gamma(\alpha)}\lambda^{\alpha} x^{\alpha-1} e^{-\lambda x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

where α and λ are any positive real numbers.

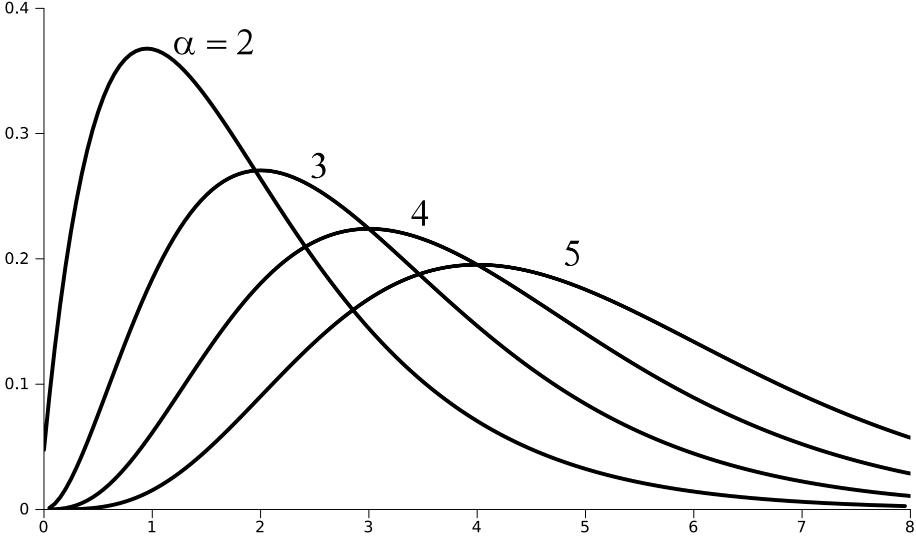
By integration we can show that $E[X] = \alpha/\lambda$ and $V[X] = \alpha/\lambda^2$

α is called a *shape* parameter.



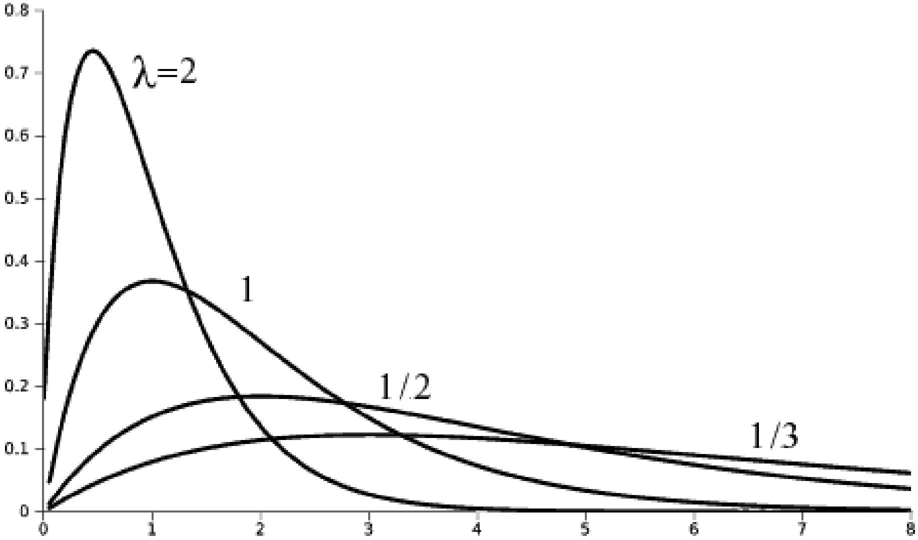
(Here λ is fixed at 1 and α varies.)

α is called a *shape* parameter.



(Here λ is fixed at 1 and α varies.)

λ is called a *scale* parameter.



(Here α is fixed at 2 and λ varies.)

$$f(x) = \begin{cases} \frac{1}{\Gamma(\alpha)}\lambda^{\alpha}x^{\alpha-1}e^{-\lambda x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Notice that if $\alpha = 1$, this simplifies to the exponential distribution with parameter λ.

This means that the gamma distribution family includes the exponential distribution as a special case.

If you add up n independent exponential random variables with parameter λ, the result is a gamma-distributed random variable with parameters $\alpha = n$ and λ.

$$f(x) = \begin{cases} \frac{1}{\Gamma(\alpha)} \lambda^\alpha x^{\alpha-1} e^{-\lambda x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

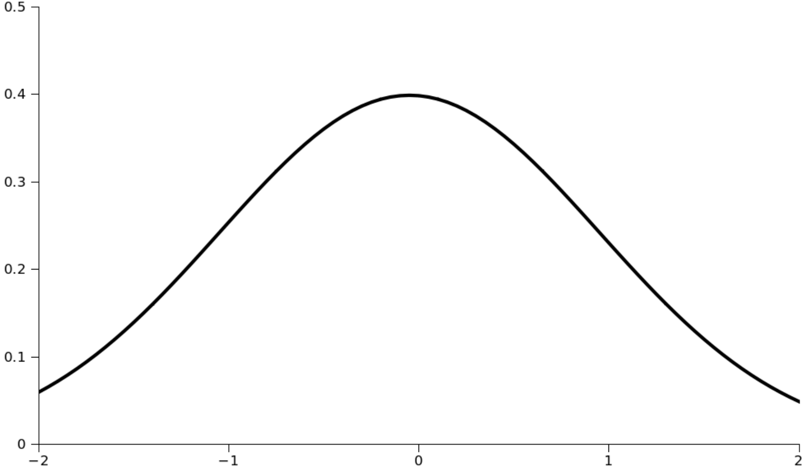
One other important special case is the *chi-squared distribution* with ν "degrees of freedom." It is simply the gamma distribution with $\alpha = \nu/2$ and $\lambda = 1/2$.

$$f(x) = \begin{cases} \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{(\nu/2)-1} e^{-x/2} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

We'll come back to this distribution later in the course.

NORMAL DISTRIBUTION

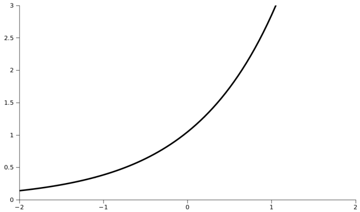
In many cases, we want a “bell curve” shaped distribution

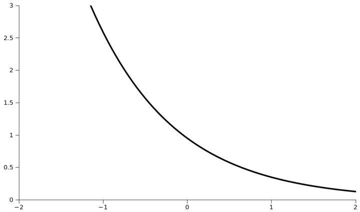


What type of mathematical function has this shape?

Think about one tail at a time...

The lower end could be described by e^x

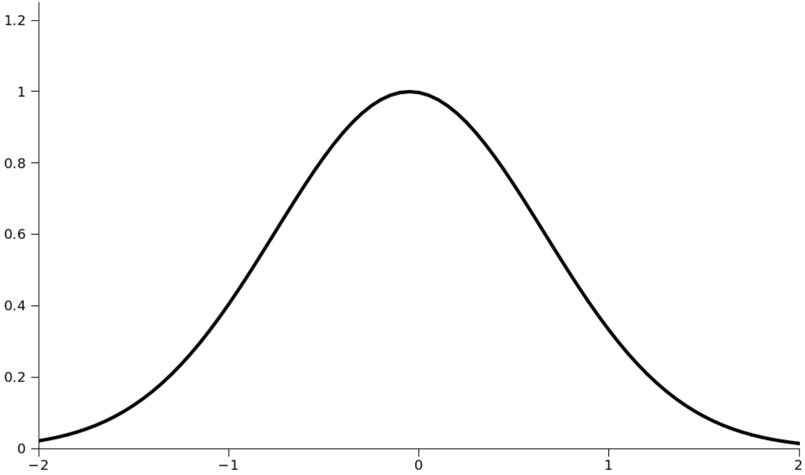




The upper end could be described by e^{-x}

In other words, we want to take the exponential of something which goes to 0 as x is either very positive or very negative

One such function is e^{-x^2}



In other words, we want to take the exponential of something which goes to 0 as x is either very positive or very negative

But is e^{-x^2} a valid PDF?

- **Is it nonnegative?** Yes, for all x.
- **Does $\int_{-\infty}^{\infty} e^{-x^2} dx = 1$?** No, but finding out what this integral is actually equal to is surprisingly hard.

It turns out that it does not integrate to 1 (oddly, it actually integrates to $\sqrt{\pi}$), but adding a few constants does the trick:

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 1$$

This is called the **standard normal distribution**, and shows up very frequently. The notation Z is usually used to mean a continuous random variable with the standard normal distribution.

Its PDF is given as:

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

for any $z \in \mathbb{R}$

The standard normal distribution has mean 0, and its variance and standard deviation are both equal to 1.

Calculating probabilities with the PDF is trickier than usual, because there is no closed-form expression for $\int_a^b f(z)dz$.

To solve this problem, tables have been calculated with values of the CDF, which is denoted

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$$

One of these tables has been posted on Canvas. (Hard copies of this will be provided with future exams.)

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
$\vdots$										
-1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
$\vdots$										

So, for instance, $\Phi(-1.63) = 0.0516$

Many software programs can calculate $\Phi(x)$ for you (`normcdf` in Matlab, `NORMDIST` in Excel, `pnorm` in R)

Example

The table tells you that $\Phi(1.25) = 0.8944$ What are the following probabilities?

- $P(Z \leq 1.25)$?
- $P(Z \geq 1.25)$?
- $P(Z \leq -1.25)$?
- $P(-1.25 \leq Z \leq 1.25)$?

We can also use the tables to find percentiles of the standard normal distribution.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
$\vdots$										
-1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
$\vdots$										

The 5th-percentile is the value of z such that $\Phi(z) = 0.05$; interpolating the table, this happens if $z = -1.645$

NONSTANDARD NORMAL DISTRIBUTIONS

There are many cases where we want the general shape of the standard normal distribution, but don't want a mean of 0 or a variance of 1.

Let's consider the easy approach: we want X to be a normal random variable, but with a mean μ and standard deviation σ.

Let $X = \mu + \sigma Z$ where Z is a standard normal distribution.

We can check that $E[X] = \mu$ and $V[X] = \sigma^2$.

In this case, we say X has a normal distribution with mean μ and standard deviation σ.

However, it would be really annoying to have tables with values of the CDF for all possible values of μ and σ.

The good news is that we can use the equations $X = \mu + \sigma Z$ and $Z = (X - \mu)/\sigma$ to directly apply everything we know about the standard normal distribution.

Example

Let's say X has a normal distribution with mean 5 and standard deviation 2. What is the probability X is between 3 and 4?

$$P(3 \leq X \leq 4) = P(3 \leq 5 + 2Z \leq 4)$$

$$= P(-2 \leq 2Z \leq -1)$$

$$= P(-1 \leq Z \leq -1/2)$$

$$= \Phi(-1/2) - \Phi(-1) = 0.3085 - 0.1587 = 0.1498$$

We can use a similar technique to derive the PDF. For the standard normal distribution we have

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = 1$$

Substituting $z = (x - \mu)/\sigma$ and $dz = dx/\sigma$ we obtain

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

so the PDF for the normal distribution with mean μ and standard deviation σ is

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

The same idea works with percentiles. What is the 90th percentile of a normal distribution with mean 100 and standard deviation 10?

The 90th percentile of the standard normal distribution is approximately 1.28.

Using $X = 100 + 10Z$, the 90th-percentile of X is $100 + 12.8 = 112.8$

Some rules of thumb

For *any* normal distribution (regardless of μ or σ):

- The probability of being within 1 standard deviation of the mean is roughly 68%.
- The probability of being within 2 standard deviations of the mean is roughly 95%.
- The probability of being within 3 standard deviations of the mean is roughly 99.7%.

How can you find the exact values for these rules of thumb?

Why is this called the normal distribution?

Many common phenomena are exactly (or approximately) described by normal distributions.

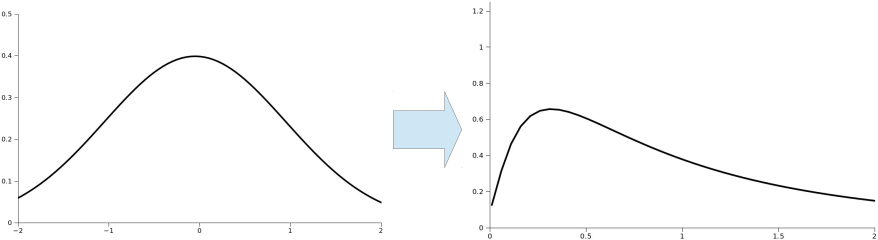
- Velocities of molecules in an ideal gas
- Binomial random variables where n is large
- Measurement errors in physical experiments
- Blood pressure of adult humans
- The change in the logarithm of daily stock prices

In general, phenomena which are the results of repeated **additive** properties will be approximately normal.

We will soon discuss the central limit theorem, which will give some justification for why it occurs so frequently.

LOGNORMAL DISTRIBUTION

The lognormal distribution is a small modification to the normal distribution.



What if we wanted to ensure that the random variable was positive?

One approach is to let Y have a normal distribution, and let $X = e^Y$.

In this case, X is said to have a lognormal distribution because $\log X$ has a normal distribution.

The pdf of X is given by $\begin{cases} \frac{1}{\sqrt{2\pi}\sigma x} \exp(-[\log x - \mu]^2/2\sigma^2) & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$
where μ and σ are the mean and standard deviation of $Y = \log X$, **not** X.

The mean and variance of X are given by these formulas:

$$E[X] = \exp(\mu + \sigma^2/2)$$
$$V[X] = \exp(2\mu + \sigma^2)(\exp(\sigma^2) - 1)$$

and problems using the lognormal distribution can be related back to the standard normal distribution by the equations $X = \exp(Y)$ and $Y = \mu + \sigma Z$, with Z the standard normal distribution.

Example

The daily traffic volume on I35 through downtown Austin can be represented by a lognormal distribution with $\mu = 11$ and $\sigma = 1$. What is the probability that the volume will be between 80,000 and 100,000 on any given day?

$$P(80000 \leq X \leq 100000) = P(80000 \leq \exp(Y) \leq 100000)$$

$$= P(11.29 \leq Y \leq 11.51)$$

$$= P(11.29 \leq 11 + Z \leq 11.51)$$

$$= P(0.29 \leq Z \leq 0.51)$$

$$= \Phi(0.51) - \Phi(0.29) = 0.6950 - 0.6141 = 0.0809$$