Common Continuous Distributions

Statistics 104

Autumn 2004

Taken from Statistics 110 Lecture Notes

Common Continuous Distributions

- Uniform
- Exponential
- Normal
- Gamma
- Cauchy

Uniform Distributions

This distribution describes events that are equally likely in a range (a, b). As mentioned before, it is what people often consider as a random number.

The PDF for the for the uniform distribution ($U(a, b)$) is

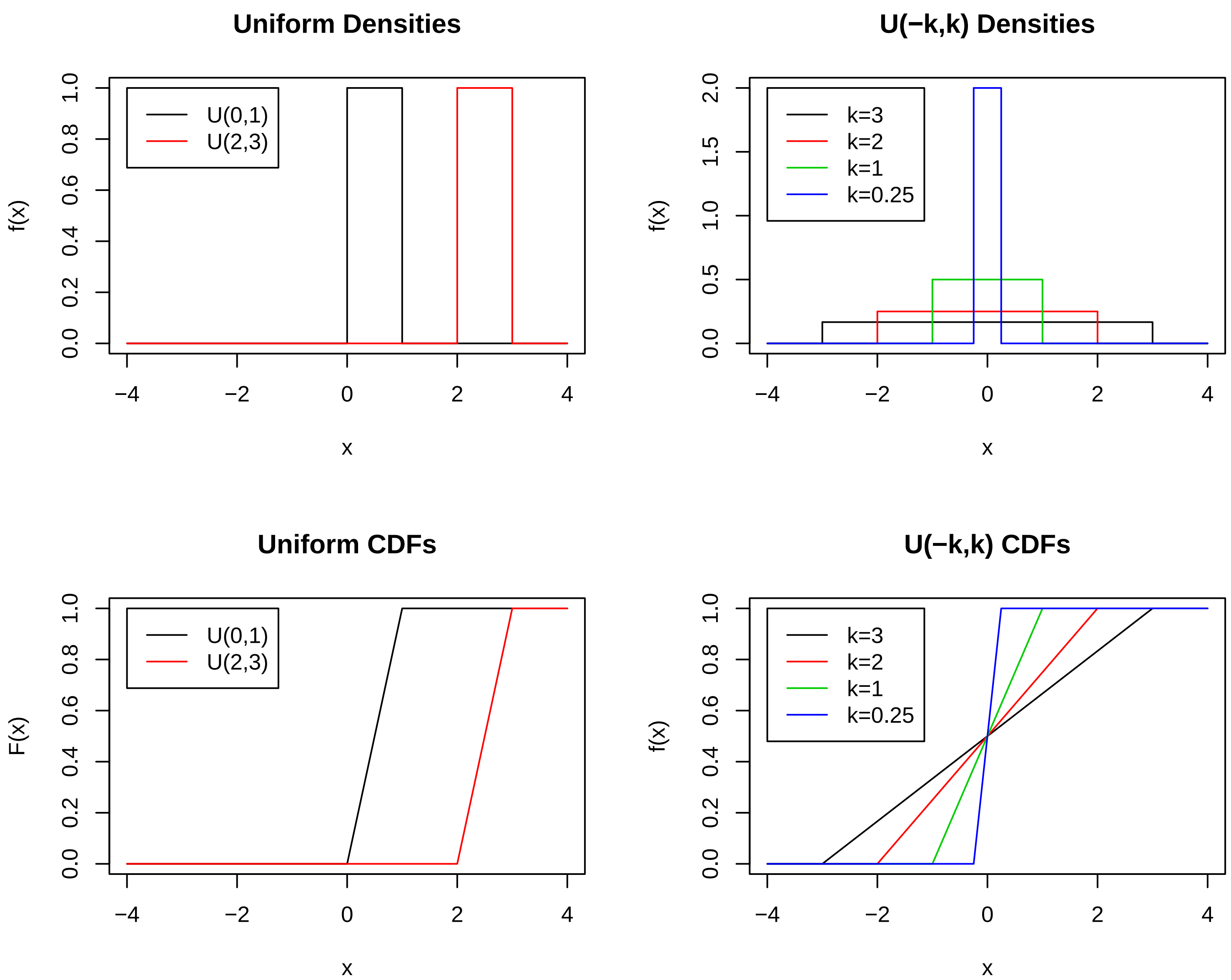
$$f(x) = \frac{1}{b-a}; \quad a \leq x \leq b$$

The CDF is

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

The first two moments of the uniform are

$$E[X] = \frac{a+b}{2}; \quad \mathrm{Var}(X) = \frac{(b-a)^2}{12}; \quad SD(X) = \sqrt{\frac{(b-a)^2}{12}}$$


Uniform Densities
U(0,1)
U(2,3)
f(x)
x
U(−k,k) Densities
k=3
k=2
k=1
k=0.25
f(x)
x
Uniform CDFs
U(0,1)
U(2,3)
F(x)
x
U(−k,k) CDFs
k=3
k=2
k=1
k=0.25
f(x)
x

Exponential Distribution

The exponential distribution is often used to describe the time to an event.

The PDF for the for the exponential distribution $(Exp(\lambda))$ is

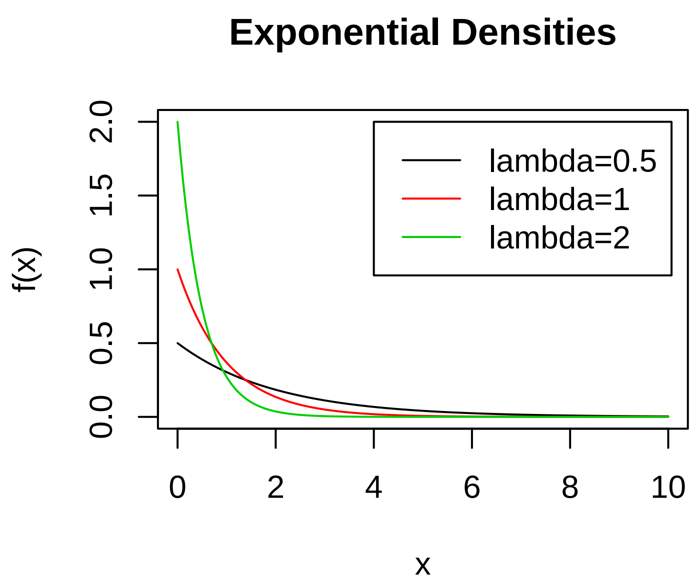
$$f(x) = \lambda e^{-\lambda x}; \quad x \geq 0$$

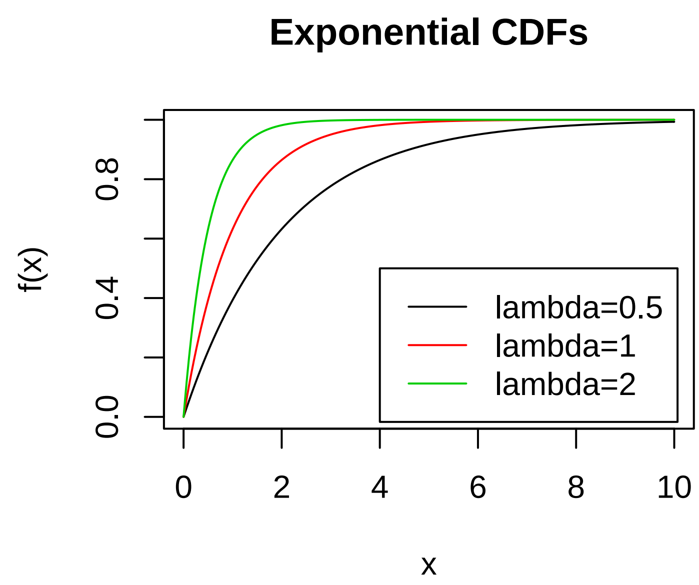
The CDF is

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

The first two moments of the exponential are

$$E[X] = \frac{1}{\lambda}; \quad \mathrm{Var}(X) = \frac{1}{\lambda^2}; \quad SD(X) = \frac{1}{\lambda}$$





The exponential distribution has an interesting property. It is said to be memoryless. That is, it satisfies

$$
\begin{aligned}
P[T > t + s | T > s] &= \frac{P[T > t + s \text{ and } T > s]}{P[T > s]} \\
&= \frac{P[T > t + s]}{P[T > s]} = \frac{e^{-\lambda t + s}}{e^{-\lambda s}} = e^{-\lambda t}
\end{aligned}
$$

So the chance that what we are observing survives another t units of time doesn't depend on how long we have observed it so far.

This property limits where the exponential distribution can be used. For example, we wouldn't want to use it to model human lifetimes.

Another potential drawback is that the parameter λ describes the distribution completely. Once you know this, you also know the standard deviation, skewness, kurtosis, etc.

Even with these drawbacks, the exponential distribution is widely used. Examples where it may be appropriate are

- Queuing theory - times between customer arrivals
- Times to relapse in leukemia patients
- Times to equipment failures
- Distances between crossovers during meiosis

The parameter λ can often be thought of as a rate or a speed parameter. The exponential distribution can be parameterized in terms of the mean time to the event, $\mu = \frac{1}{\lambda}$. With this parameterization the PDF and CDF are

$$f(x) = \frac{e^{-x/\mu}}{\mu}; \quad x \geq 0 \qquad F(x) = \begin{cases} 1 - e^{-x/\mu} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Normal Distributions

The normal distribution is almost surely the most common distribution used in probability and statistics. It is also referred to as the Gaussian distribution, as Gauss was an early promoter of its use (though not the first, which was probably De Moivre). It is also what most people mean when they talk about bell curve. It is used to describe observed data, measurement errors, an approximation distribution (Central Limit Theorem).

The PDF for the for the normal distribution ($N(\mu, \sigma^2)$) is

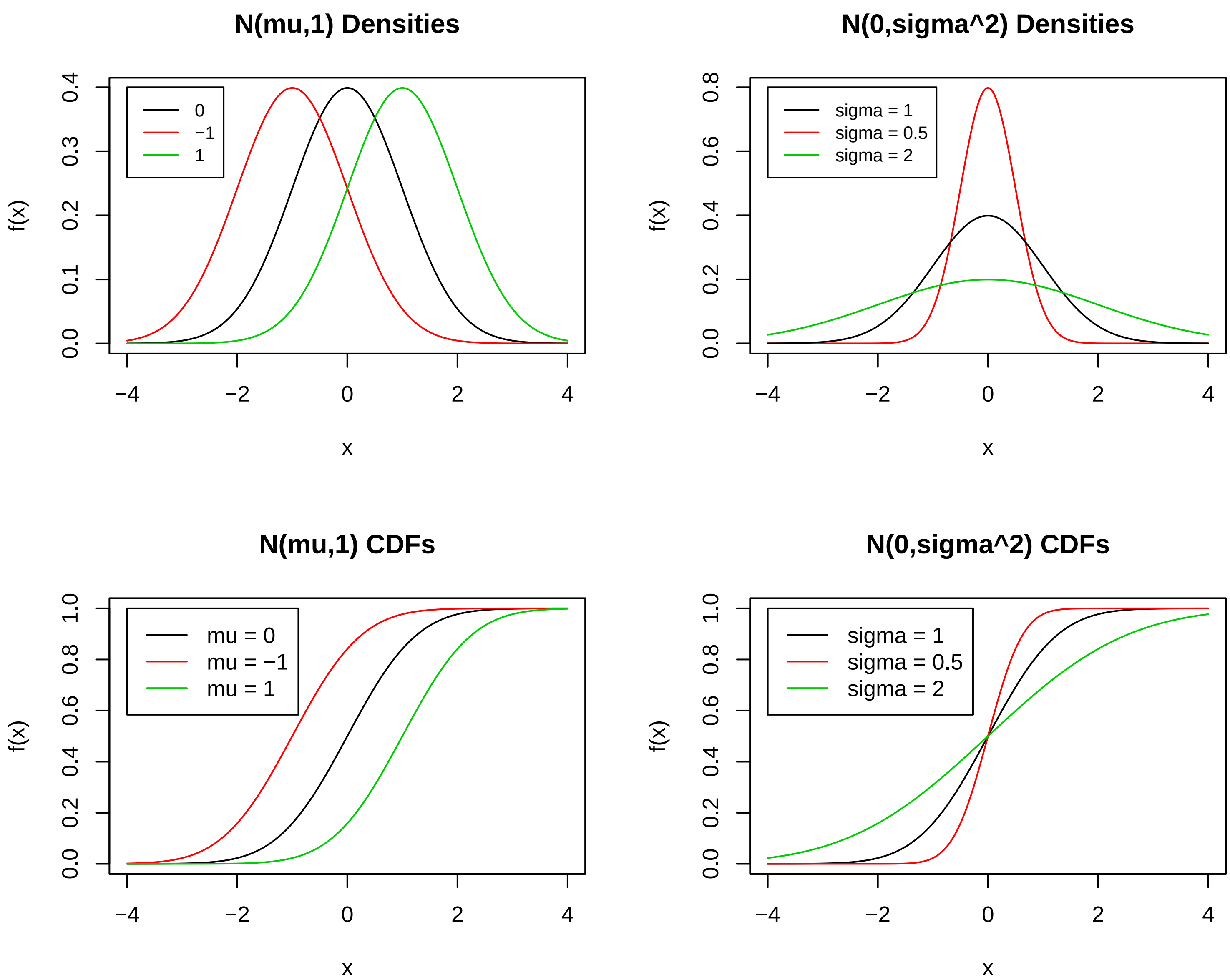
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

The two parameters of the distribution are the mean (μ) and the variance (σ^2). A special case is the **standard normal density** which has $\mu = 0$ and $\sigma^2 = 1$ and its PDF is often denoted by $\phi(x)$. As we shall see, once we understand the standard normal ($N(0,1)$), we understand all normal distributions.

The CDF for the normal distribution doesn't have a nice form. The CDF for the standard normal is often denoted by $\Phi(x)$ which is of the form

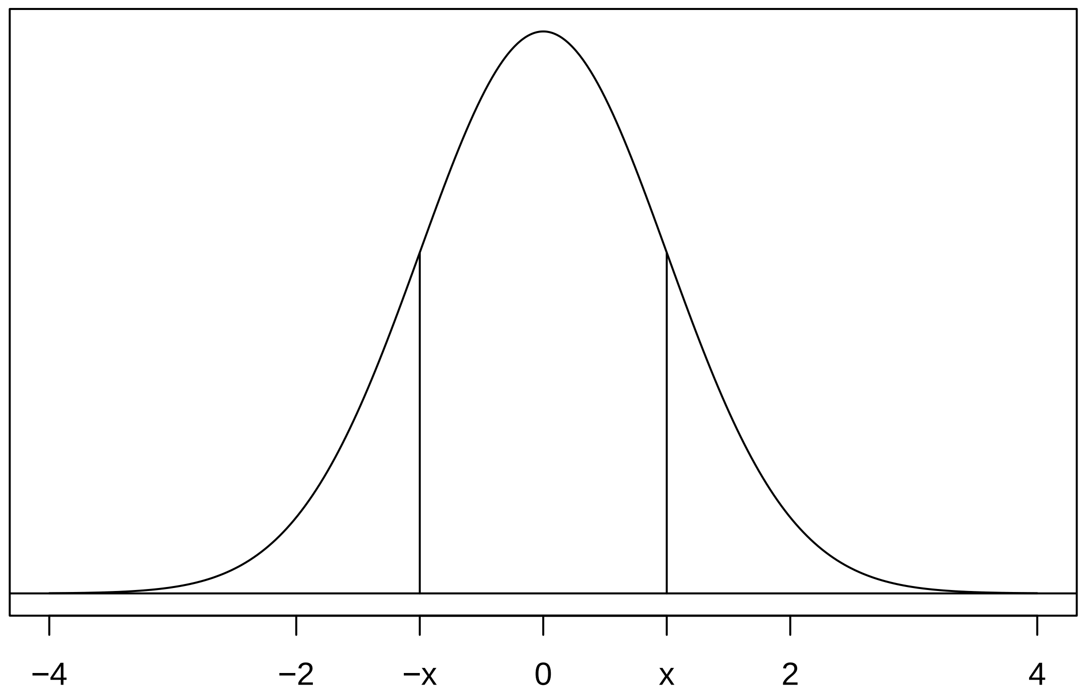
$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$$

The CDF for any other normal distribution is based on $\Phi(x)$.


N(mu,1) Densities
0
−1
1
f(x)
x
N(0,sigma^2) Densities
sigma = 1
sigma = 0.5
sigma = 2
f(x)
x
N(mu,1) CDFs
mu = 0
mu = −1
mu = 1
f(x)
x
N(0,sigma^2) CDFs
sigma = 1
sigma = 0.5
sigma = 2
f(x)
x

The normal distribution is an example of a symmetric distribution, with the point of symmetry being the mean μ. For a symmetric distribution with mean 0

$$P[X \leq -x] = P[X \geq x]$$



Theorem. *Let* $X \sim N(\mu, \sigma^2)$ *and let* $Y = aX + b$*. Then* $Y \sim N(a\mu + b, (a\sigma)^2)$

Proof.

$$
\begin{aligned}
F_Y(y) &= P[Y \leq y] \\
&= P[aX + b \leq y] \\
&= P\left[X \leq \frac{y-b}{a}\right] \\
&= F_X\left(\frac{y-b}{a}\right)
\end{aligned}
$$

Therefore

$$
\begin{aligned}
f_Y(y) &= \frac{d}{dy} F_X\left(\frac{y-b}{a}\right) \\
&= \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)
\end{aligned}
$$

Note that to this point, we haven't made any assumptions about the distribution of X. Any linear transformation of a RV gives this relationship.

Now if $X \sim N(\mu, \sigma^2)$, then

$$f_Y(y) = \frac{1}{|a|\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{y-b-a\mu}{a\sigma}\right)\right\}$$

which is the density for a $N(a\mu + b, (a\sigma)^2)$. $\square$

From this all normal density curves must have the same basic shape and if $X \sim N(\mu, \sigma^2)$

$$f(x) = \frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)$$

and

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

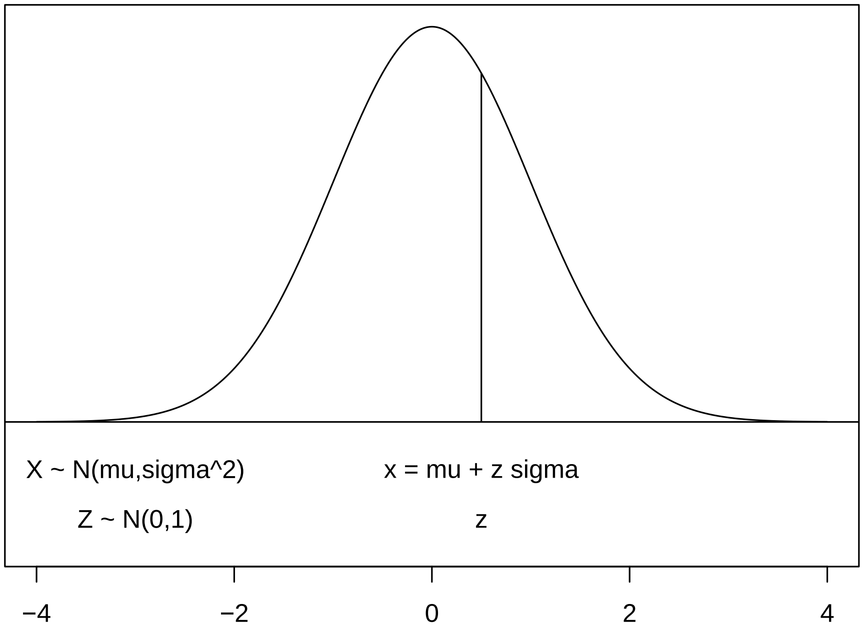
One consequence of this, is that to get probabilities involving normal distributions we only need a single function, or table.

If $X \sim N(\mu, \sigma^2)$, then

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

The values Z are sometimes referred to as the standard scores.

Suppose for example that blood pressure (X) can be modelled (approximately) by a normal distribution with $\mu = 120$ and $\sigma = 20$.



If we are interested in the $P[X \leq 140]$, this is the same as the $P[Z \leq 1]$ since

$$z = \frac{140 - 120}{20} = 1$$

Table 2 of Rice gives the CDF of the standard normal.

Note that this table is actually taken from Moore and McCabe's Introduction to the Practice of Statistics. However it is the same a Table 2 in Rice.

z is the standard normal variable. The value of P for $-z$ equals 1 minus the value of P for $+z$; for example P for -1.62 equals $1 - 0.9474 = 0.0526$.

Table entry for z is the area under the standard normal curve to the left of z.

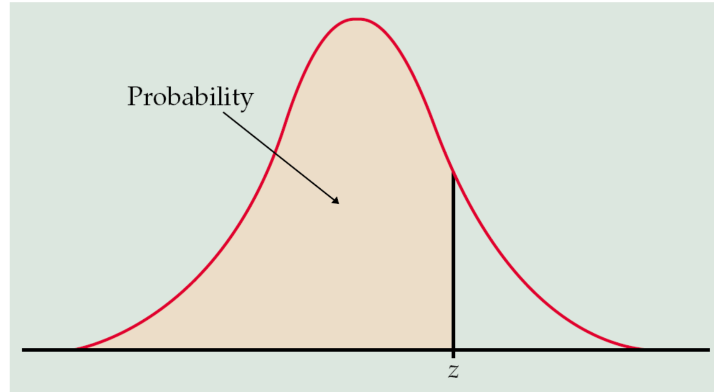
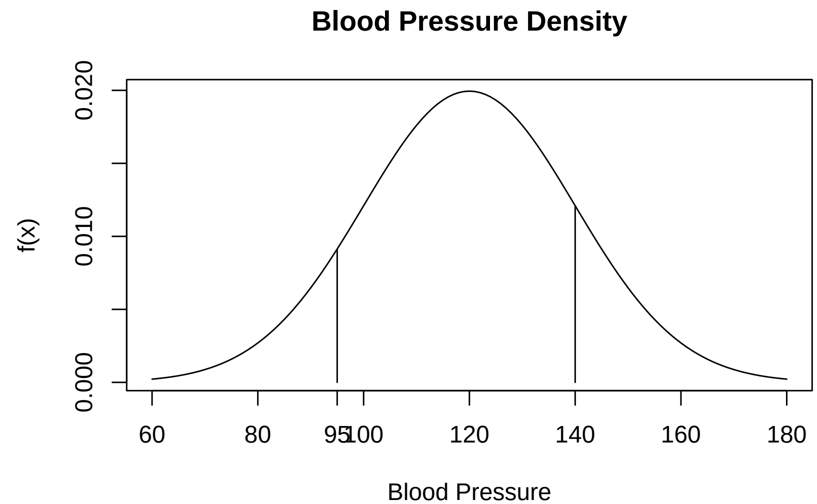


TABLE A Standard normal probabilities (*continued*)

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936

We can use the table to get the following probabilities



- $P[X \leq 140]$

$$P[X \leq 140] = P[Z \leq 1] = 0.8413$$

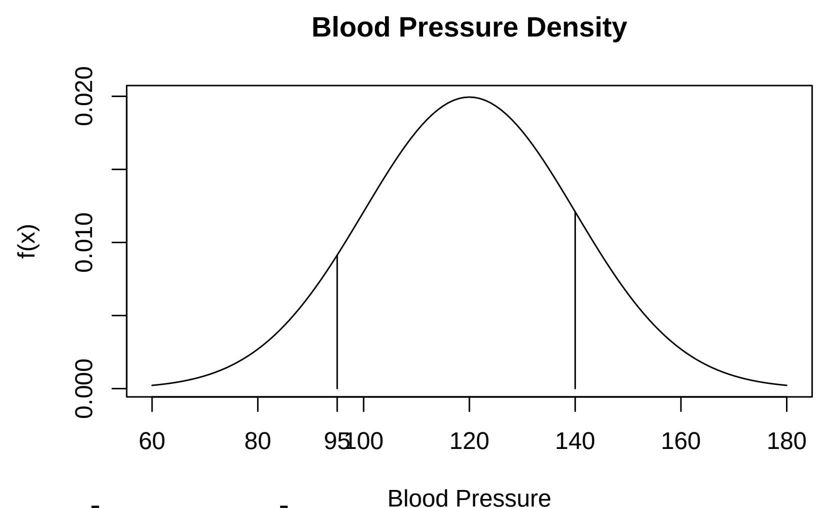
- $P[X \geq 95]$

$$\begin{aligned} P[X \geq 95] &= P\left[Z \geq \frac{95 - 120}{20}\right] \\ &= P[Z \geq -1.25] \\ &= 1 - P[Z \leq -1.25] \end{aligned}$$

Using the fact that $P[Z \leq -1.25] = 1 - P[Z \leq 1.25]$ (table flip), $P[Z \leq -1.25] = 1 - 0.8944 = 0.1056$. Therefore

$$P[X \geq 95] = 1 - 0.1056 = 0.8944$$

- $P[95 \leq X \leq 140]$



$$\begin{aligned} P[95 \leq X \leq 140] &= P[X \leq 140] - P[X \leq 95] \\ &= P[Z \leq 1] - P[Z \leq -1.25] \\ &= 0.8413 - 0.1056 = 0.7357 \end{aligned}$$

Gamma Distributions

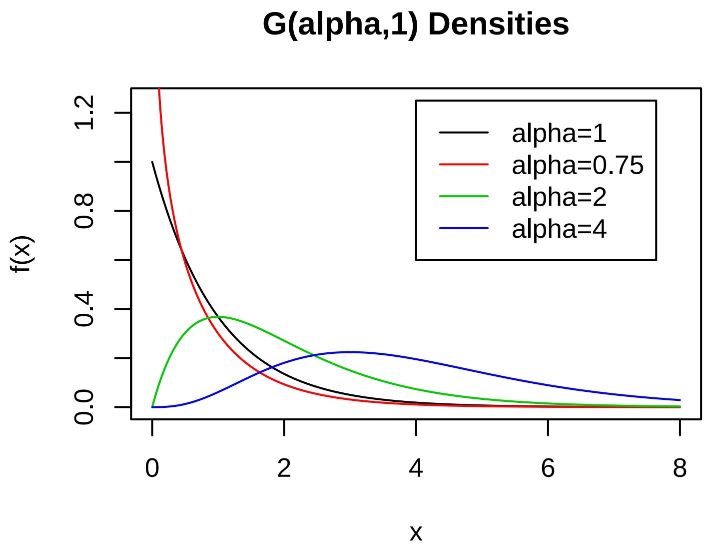
The gamma distribution can be used model a wide range of non-negative RVs. It has been used to model times between earthquakes, the size of automobile insurance claims, rainfall amounts, plant yields.

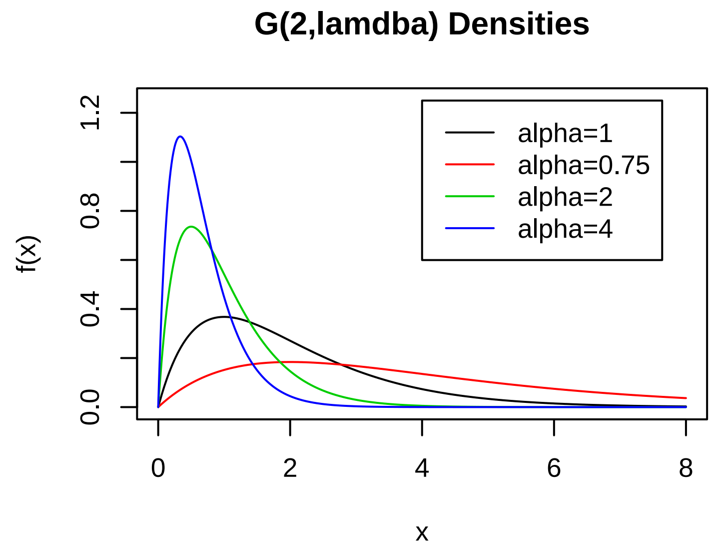
The PDF for the for the gamma distribution ($G(\alpha, \lambda)$) is

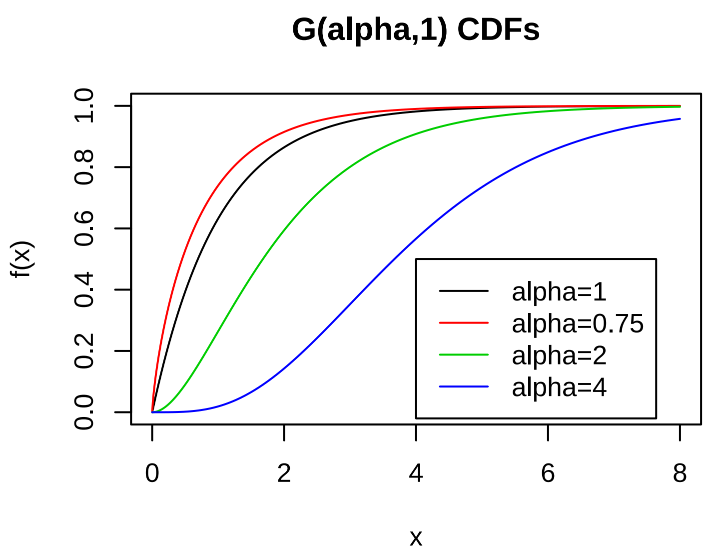
$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}; \quad x \geq 0$$

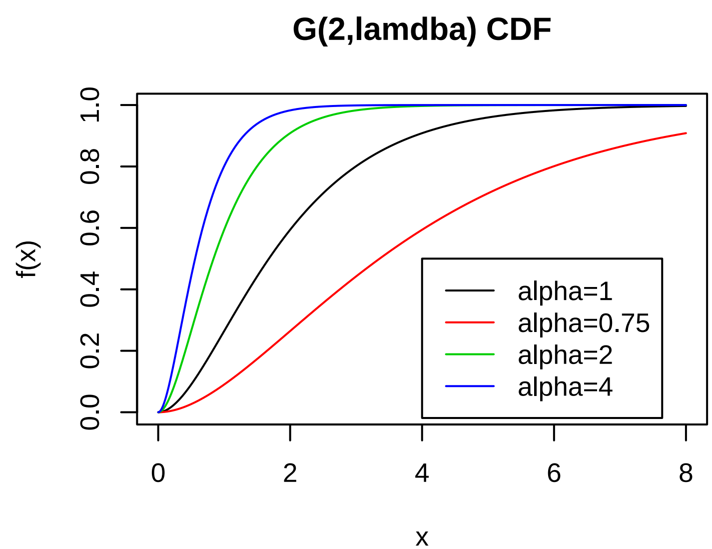
The parameter α is the shape parameter of the gamma distribution and $\frac{1}{\lambda}$ is the scale parameter.

The gamma distribution is a generalization of exponential distribution as $Exp(\lambda) = G(1, \lambda)$.


G(alpha,1) Densities
f(x)
0.0
0.4
0.8
1.2
0
2
4
6
8
x
alpha=1
alpha=0.75
alpha=2
alpha=4


G(2,lamdba) Densities
f(x)
0.0
0.4
0.8
1.2
0
2
4
6
8
x
alpha=1
alpha=0.75
alpha=2
alpha=4


G(alpha,1) CDFs
f(x)
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
x
alpha=1
alpha=0.75
alpha=2
alpha=4


G(2,lamdba) CDF
f(x)
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
x
alpha=1
alpha=0.75
alpha=2
alpha=4

The “normalization constant”

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

is the Gamma function evaluated at α.

Some useful properties of the Gamma function are

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

2. $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$ (Prove by integration by parts).

3. If n is a positive integer, then $\Gamma(n) = (n - 1)!$ (Direct consequence of the first two facts).

4. $\Gamma(0.5) = \sqrt{\pi}$ (Useful in showing that the variance for a normal is σ^2).

The CDF of the gamma doesn’t have a nice closed form so you need tables or a computer to find probabilities involving the gamma.

The first two moments of the gamma are

$$E[X] = \frac{\alpha}{\lambda}; \quad \mathrm{Var}(X) = \frac{\alpha}{\lambda^2}; \quad SD(X) = \frac{\sqrt{\alpha}}{\lambda}$$

Proof. Let $X \sim G(\alpha, 1)$. Then

$$\begin{aligned} E[X^n] &= \int_0^\infty x^n \frac{x^{\alpha-1}e^{-x}}{\Gamma(\alpha)} dx \\ &= \int_0^\infty \frac{x^{\alpha+n-1}e^{-x}}{\Gamma(\alpha)} dx \\ &= \frac{\Gamma(\alpha+n)}{\Gamma(\alpha)} \end{aligned}$$

So

$$E[X] = \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} = \frac{\alpha\Gamma(\alpha)}{\Gamma(\alpha)} = \alpha$$

and

$$E[X] = \frac{\Gamma(\alpha+2)}{\Gamma(\alpha)} = \frac{(\alpha+1)\alpha\Gamma(\alpha)}{\Gamma(\alpha)} = (\alpha+1)\alpha$$

This implies that

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2 = \alpha$$

Let $Y = \frac{X}{\lambda}$. Then $Y \ \sim G(\alpha, \lambda)$ as

$$f_Y(y) = \frac{\lambda(\lambda y)^{\alpha-1} e^{-\lambda y}}{\Gamma(\alpha)}$$

Thus

$$E[Y] = \frac{E[X]}{\lambda} = \frac{\alpha}{\lambda}; \quad \mathrm{Var}(Y) = \frac{\mathrm{Var}(X)}{\lambda^2} = \frac{\alpha}{\lambda^2};$$

□

Beta Distributions

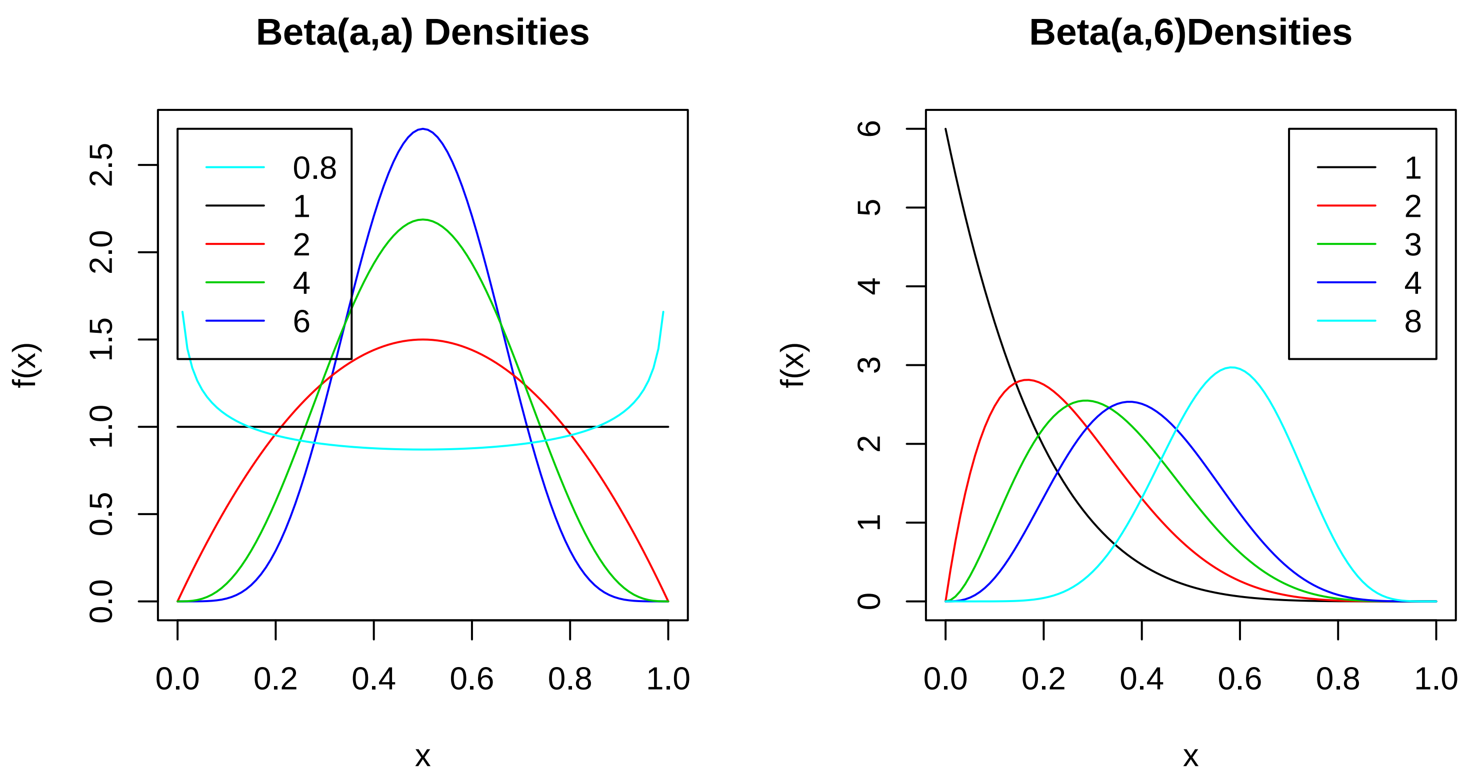
Beta distributions are useful for data that occur in fixed, finite intervals

The PDF for the for the beta distribution $(Beta(\alpha, \beta))$ is

$$f(x) = \frac{x^{a-1}(1-x)^{b-1}}{\beta(a,b)}; \quad 0 \leq x \leq 1$$

The function $\beta(a, b)$ is known as the Beta function and is

$$\begin{aligned} \beta(a,b) &= \int_0^1 x^{a-1}(1-x)^{b-1}dx \\ &= \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \end{aligned}$$



Like many continuous distributions, the CDF for the beta does not have a nice form and must be determined through tables or software.

The first two moments of the uniform are

$$E[X] = \frac{a}{a+b}; \quad \mathrm{Var}(X) = \frac{ab}{(a+b+1)(a+b)^2}$$

Proof.

$$\begin{aligned} E[X^n] &= \frac{1}{\beta(a,b)} \int_0^1 x^n x^{a-1} (1-x)^{b-1} \\ &= \frac{\beta(a+n,b)}{\beta(a,b)} \end{aligned}$$

So

$$E[X] = \frac{\beta(a+1,b)}{\beta(a,b)} = \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} = \frac{a}{a+b}$$

and

$$E[X^2] = \frac{\beta(a+1,b)}{\beta(a,b)} = \frac{(a+1)a}{(a+b+1)(a+b)}$$

which implies

$$\mathrm{Var}(X) = \frac{a}{a+b}\left(\frac{a+1}{a+b+1} - \frac{a}{a+b}\right) = \frac{a}{a+b}\frac{b}{(a+b+1)(a+b)}$$

$\square$

Cauchy Distributions

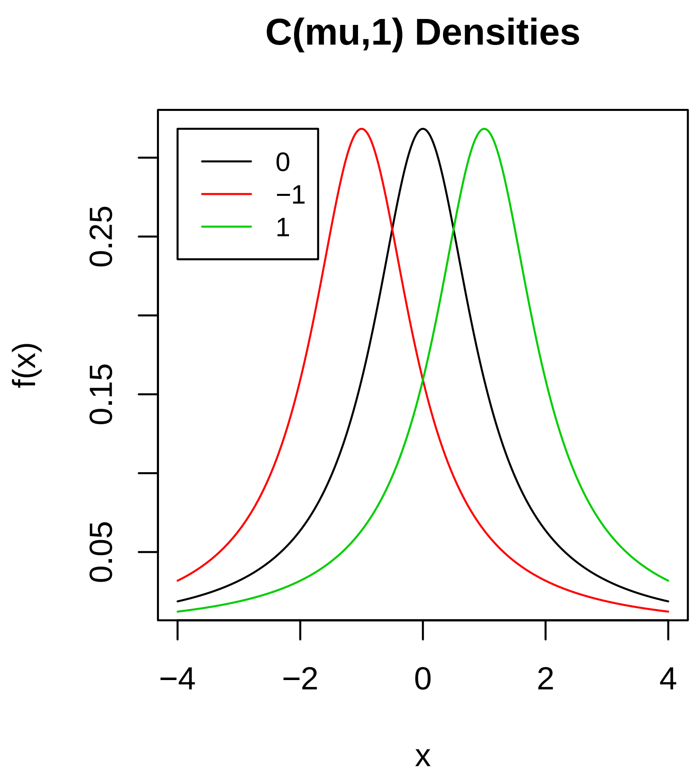
The Cauchy distribution (also known as the Lorentzian distribution), is often used for describing resonance behavior. It can also be used to describe outliers in data sets. However it is more commonly used in probability and statistics as a distribution that can be used for counter examples.

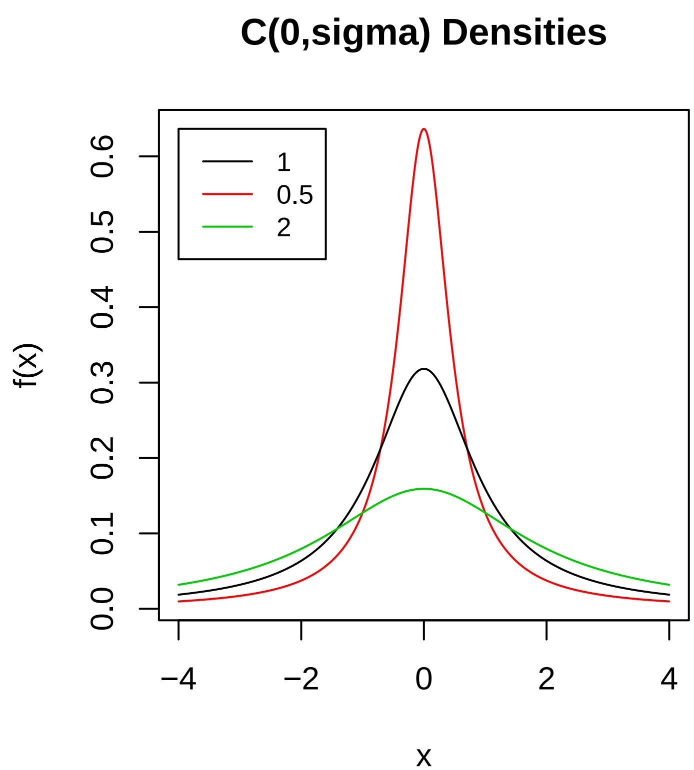
The PDF for the for the Cauchy distribution $\left(C(\mu, \sigma)\right)$ is

$$f(x) = \frac{1}{\pi\sigma\left(1 + (\frac{x-\mu}{\sigma})^2\right)}$$

The CDF is

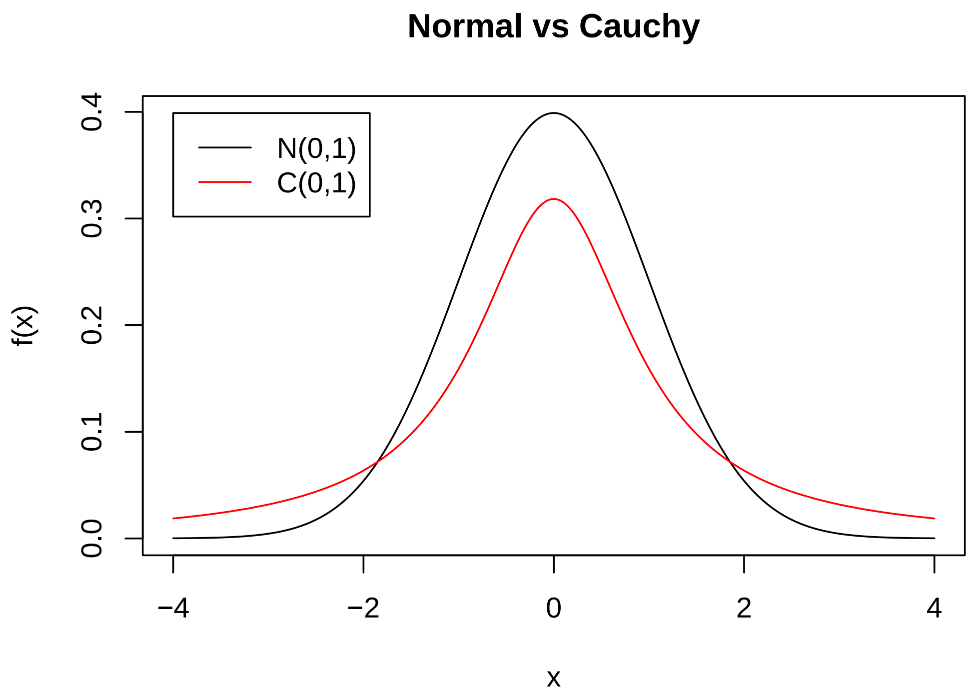
$$F(x) = \frac{1}{2} + \frac{1}{\pi}\mathrm{Arctan}\left(\frac{x-\mu}{\sigma}\right)$$





The Cauchy is another example of a location-scale distribution (the normal is the first we've discussed). If $X \sim C(0, 1)$, then $Y = \mu + \sigma X$ is $C(\mu, \sigma)$.

The Cauchy distribution is known as a heavy tailed distribution. Its tails decay to 0 very slowly.



So slowly in fact, that the Cauchy has no moments. For all n,

$$\int_{-\infty}^{\infty} |x^n| f(x) dx = \infty$$