One Dimensional Distributions

Guy Lebanon

January 26, 2011

In this note we give an overview of some important one dimensional random variables (RV) and show how to compute and plot them with R. Recall the following three functions associated with a random variable X:

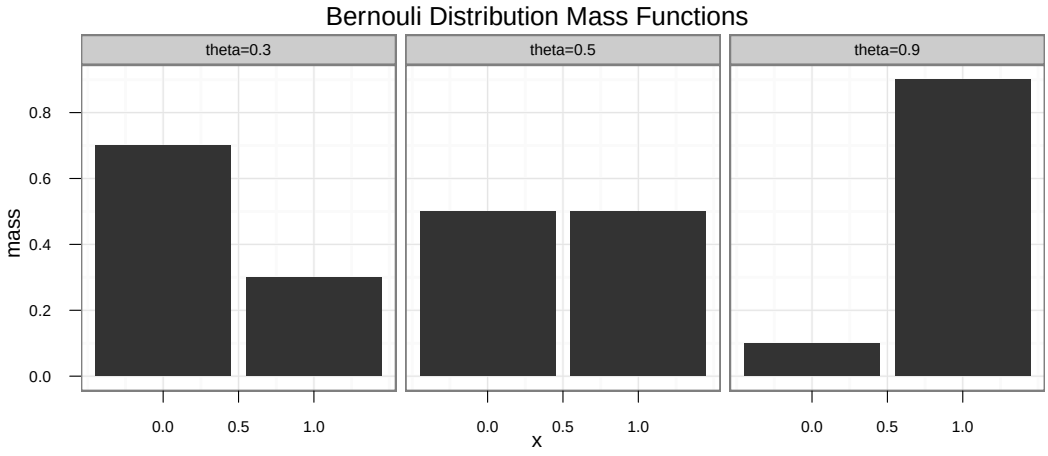
- For a discrete RV the mass function is $p_X(x) = P(X = x)$. The continuous analog is the probability density function (pdf) that may be interpreted as $f_X(x) \approx P(x < X < x + \Delta)/\Delta$ as $0 < \Delta \to 0$.
- The cumulative distribution function (cdf) $F_X(x) = P(X \leq x)$ is the probability that X is less than or equal to a specific value. In the continuous case we have $f_X(x) = dF_X(x)/dx$ and $F_X(x) = \int_{-\infty}^{x} f_X(r)dr$.
- The quantile function (qf) $Q_X(r) = F_X^{-1}(r)$ is the inverse of the cdf. $Q_X(r)$ is the x value at which $P(X \leq x) = r$ (or the minimal such value if F is not one to one).

Discrete Random Variables

The Bernoulli Trial RV

The simplest discrete RV is perhaps the Bernoulli trial RV. It has the following mass function $p_X(0) = 1 - \theta$, $p_X(1) = \theta$, $0 \leq \theta \leq 1$ and $p_X(x) = 0$ for all other values of x. θ is said to be the parameter of the distribution. The Bernoulli trial RV may be used to characterize the probability that a possibly biased coin falls on heads ($X = 1$) or tails ($X = 0$). Sometime X is interpreted as an experiment or trial that may either fail $X = 0$ or succeed $X = 1$ with probabilities $1-\theta, \theta$ respectively. The expectation of X is θ and the variance is $\theta(1 - \theta)$. We show below three different mass functions.

```
> x=c(0,1); # possible values of X are 0 and 1
> # Compute three different Bernoulli mass functions with theta 0.3, 0.5, 0.9
> # The function dbinom(x,1,theta) is used since the Bernouli is the same as a binomial
> # with n=1 (see next page)
> y1=dbinom(x,1,0.3);y2=dbinom(x,1,0.5);y3=dbinom(x,1,0.9);
> # construct a data-frame with columns: x, mass, and theta
> D=data.frame(mass=c(y1,y2,y3));D$x=x; theme_set(theme_bw(base_size=8));
> D$parameter[1:2]='theta=0.3';D$parameter[3:4]='theta=0.5';D$parameter[5:6]='theta=0.9';
> # display mass function using bar plot for each parameter
> print(qplot(x,mass,data=D,geom='bar',stat='identity',facets=.~parameter,
+             main='Bernouli Distribution Mass Functions'));
```



The Binomial RV

The Binomial RV X counts the number of successes in n independent Bernoulli experiments with parameter θ (without ordering). If we kept the ordering, then the required probability for x successes would be $\theta^x(1-\theta)^{n-x}$. Since we disregard the ordering we need to count all possible outcomes leading to x successes (there are n choose x such outcomes):

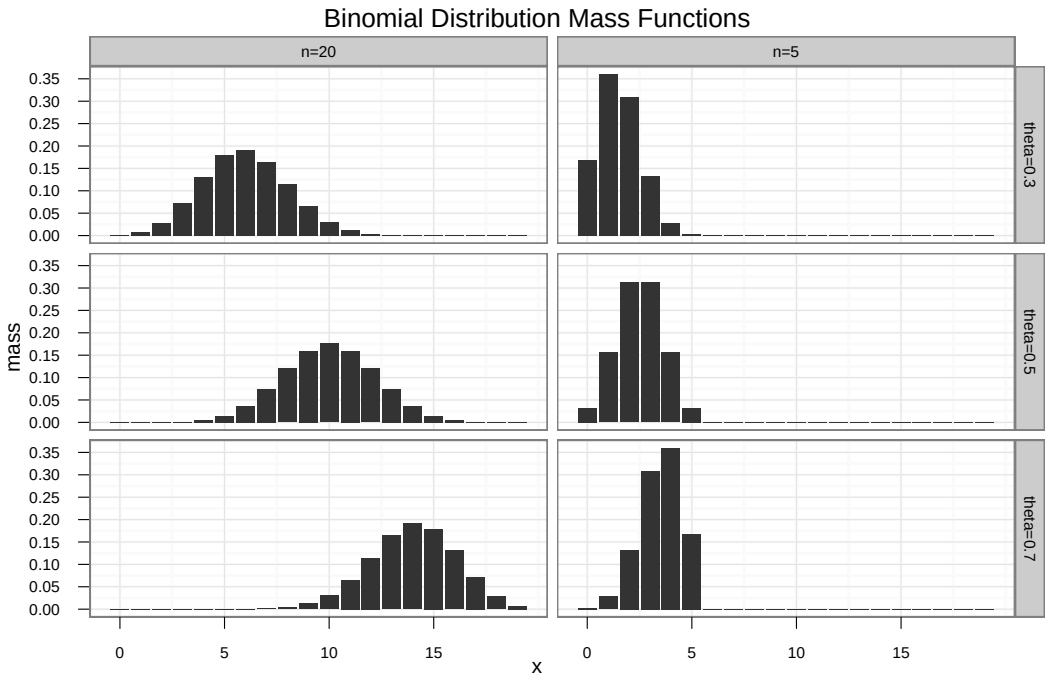
$$p_X(x) = \left(\begin{array}{c} n \\ x \end{array} \right) \theta^x (1-\theta)^{n-x}$$

for $x = 0, 1, \ldots, n$ and $p_X(x) = 0$ otherwise. The fact that $\sum_{x=0}^{n} p_X(x) = 1$ may be ascertained by the binomial theorem

$$1 = 1^n = (\theta + (1-\theta))^n = \sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array} \right) \theta^k (1-\theta)^{n-k}$$

Notice that a Bernouli distribution is a binomial distribution with $n = 1$ and also that a binomial RV is a sum of n independent Bernouli RVs with the corresponding θ. The expectation and variance of a binomial RV are $n\theta$ and $n\theta(1-\theta)$.

```
> x=0:19;
> y1=dbinom(x,5,0.3); y2=dbinom(x,5,0.5); y3=dbinom(x,5,0.7); #fix n=5, vary theta
> y4=dbinom(x,20,0.3);y5=dbinom(x,20,0.5);y6=dbinom(x,20,0.7); # fix n=20, vary theta
> D=data.frame(mass=c(y1,y2,y3,y4,y5,y6)); D$x=x;
> D$n[1:20] ='n=5';D$n[21:40] ='n=5';D$n[41:60] ='n=5';
> D$n[61:80] ='n=20';D$n[81:100] ='n=20';D$n[101:120] ='n=20';
> D$theta[1:20] ='theta=0.3';D$theta[21:40] ='theta=0.5';D$theta[41:60] ='theta=0.7';
> D$theta[61:80] ='theta=0.3';D$theta[81:100] ='theta=0.5';D$theta[101:120] ='theta=0.7';
> print(qplot(x,mass,data=D,geom='bar',stat='identity',facets=theta~n,
+              main='Binomial␣Distribution␣Mass␣Functions'));
```



Note that the trends in the above figure are in agreement with the expectation and variance formulas above: as θ increases the distribution moves to the right and as n increases the spread or variance of the distribution increases.

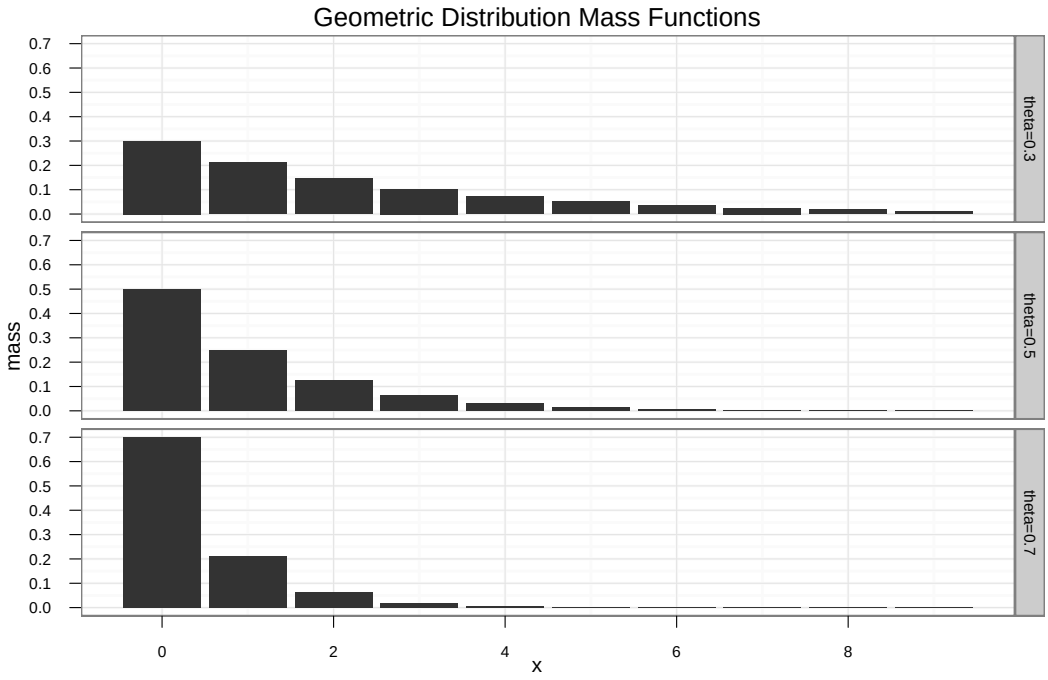
The Geometric RV

Imagine now that we have a sequence of independent Bernoulli trials $\{Z_i\}_{i=0}^{\infty}$ with parameter θ. The geometric RV X, counts the number of failures ($Z_i = 0$) before we encounter a success ($Z_j = 1$). The pmf is $p_X(x) = \theta(1-\theta)^x$ for $x \in \mathbb{N} = \{0, 1, \ldots\}$ and 0 otherwise. Using the power series formula we ascertain

$$\sum_{m=0}^{\infty} p_X(m) = \theta(1 + (1-\theta) + (1-\theta)^2 + \cdots) = \theta \frac{1}{1-(1-\theta)} = 1.$$

The expectation of a geometric RV is $(1-\theta)\theta$ and the variance is $(1-\theta)/\theta^2$. The plot below confirms that as θ increases it is more unlikely to get high values.

```
> x=0:9;
> y1=dgeom(x,0.3); y2=dgeom(x,0.5); y3=dgeom(x,0.7);
> D=data.frame(mass=c(y1,y2,y3)); D$x=x;
> D$theta[1:10] ='theta=0.3';D$theta[11:20] ='theta=0.5';D$theta[21:30] ='theta=0.7';
> print(qplot(x,mass,data=D,geom='bar',stat='identity',facets=theta~.,
+               main='Geometric Distribution Mass Functions'));
```



The Poisson RV

The pmf of the Poisson RV X is

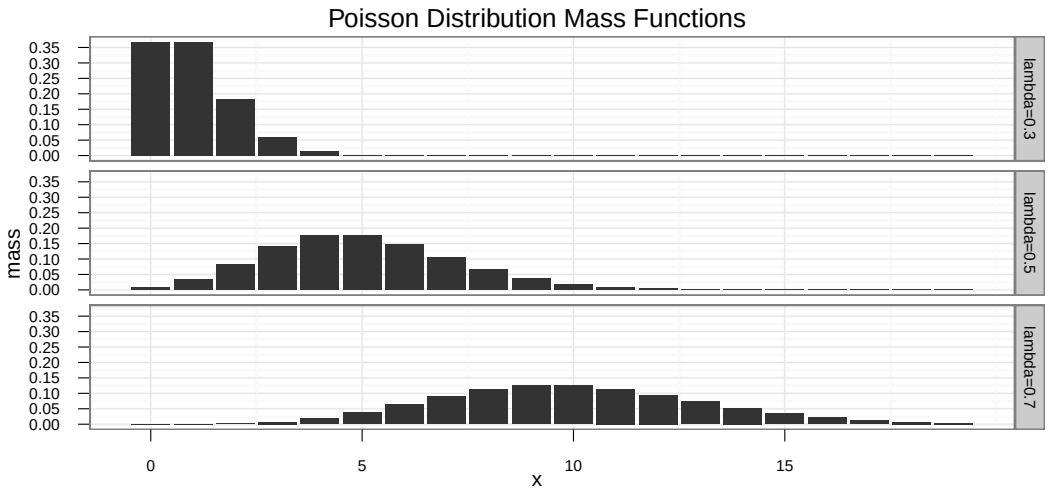
$$p_X(x) = \frac{\lambda^x e^{-\lambda}}{x!}.$$

for $x \in \mathbb{N} = \{0, 1, \ldots\}$ and $p_X(x) = 0$ otherwise. Here as well

$$\sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} = e^{-\lambda} \sum_k \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1.$$

The Poisson RV is often used to count the number of occurrences of an event at a particular region (e.g. cars arriving at an intersection or phone calls arriving at a switchboard). In the Poisson case both expectation and variance equals λ which implies that θ expresses the rate of occurrences. The binomial RV pmf approaches the Poisson pmf when $n \to \infty, \theta \to 0$ but $n\theta$ stays constant: $n\theta = \lambda$.

```
> x=0:19;
> y1=dpois(x,1); y2=dpois(x,5); y3=dpois(x,10);
> D=data.frame(mass=c(y1,y2,y3)); D$x=x;
> D$lambda[1:20] ='lambda=0.3';D$lambda[21:40] ='lambda=0.5';D$lambda[41:60] ='lambda=0.7';
> print(qplot(x,mass,data=D,geom='bar',stat='identity',facets=lambda~.,
+               main='Poisson Distribution Mass Functions'));
```

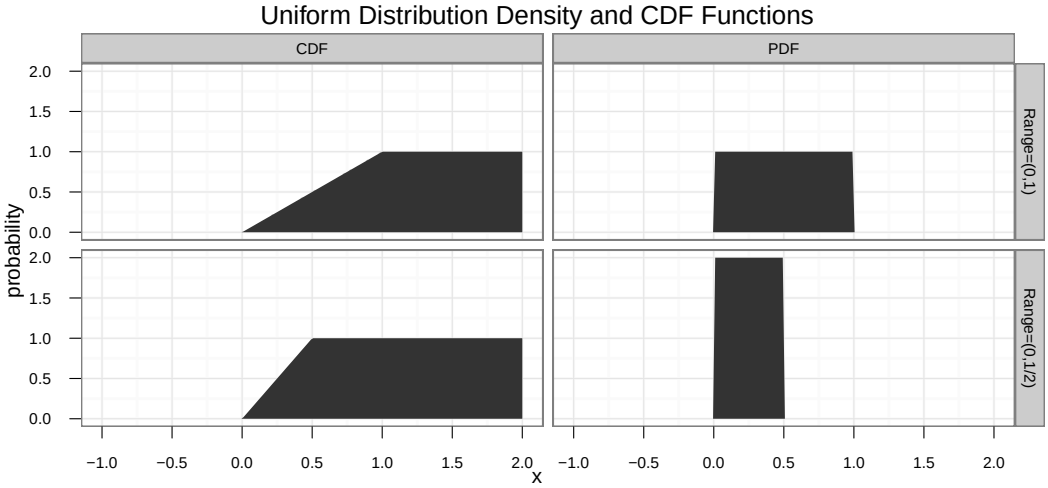


Continuous Random Variables

The Uniform distribution

The simplest continuous RV is perhaps the uniform RV on the interval $[a, b]$. It has the pdf $f_X(x) = 1/(b-a)$ for $a \le x \le b$ and 0 otherwise. Clearly $\int_{-\infty}^{\infty} f_X(x)dx = \frac{b-a}{b-a} = 1$. Note that the pdf is positive and constant between a and b and therefore any result $\{X = c\}, a \le c \le b$ is equally likely. The expectation is $(a+b)/2$ and the variance is $(b-a)^2/12$. In the plot below we show both the pdf and the cdf for the uniform distribution over two ranges (the y axis indicates probability or density).

```
> x=seq(-1,2,length=200);
> y1=dunif(x,0,1/2); y2=dunif(x,0,1); y3=punif(x,0,1/2); y4=punif(x,0,1);
> D=data.frame(probability=c(y1,y2,y3,y4));
> D$parameter[1:200]='Range=(0,1/2)';D$parameter[201:400]='Range=(0,1)'
> D$parameter[401:600]='Range=(0,1/2)';D$parameter[601:800]='Range=(0,1)'
> D$type[1:400]='PDF'; D$type[401:800]='CDF';
> print(qplot(x,probability,data=D,geom='area',facets=parameter~type,
+              main='Uniform Distribution Density and CDF Functions'));
```



Note that the density f_X may be higher than 1: its interpretation is $f_X(x) \approx P(x < X < x + \Delta)/\Delta$ as $0 < \Delta \to 0$.

The Exponential RV

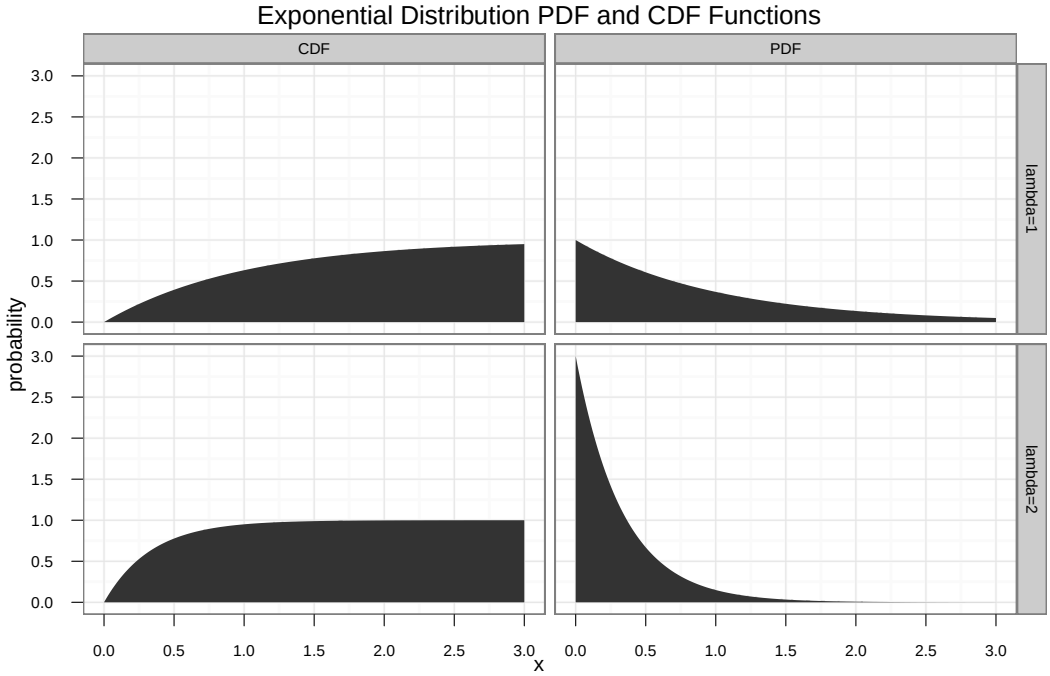
The exponential RV with parameter $\lambda > 0$ has the pdf $f_X(x) = \lambda e^{-\lambda x}$ for $x > 0$ and 0 otherwise. The cdf is

$$F_X(x) = P(X \leq x) = \int_0^x \lambda e^{-\lambda x} = -e^{-\lambda x}\Big|_0^x = 1 - e^{-\lambda x}$$

for $x > 0$ and 0 otherwise. Note that the pdf decreases exponentially fast as x grows (for positive x) - therefore it is more probable that X will receive a small value. This RV is often used to model time between successive arrivals of customers, cars, or phone calls. Its expectation is $1/\lambda$ and its variance is $1/\lambda^2$. It is the unique continuous RV with the memoryless property:

$$P(X > t + h | X > t) = \frac{P(\{X > t + h\} \cap \{X > t\})}{P(X > t)} = \frac{P(\{X > t + h\})}{P(X > t)} = \frac{e^{-\lambda(t+h)}}{e^{-\lambda t}} = e^{-\lambda h} = P(X > h).$$

```
> x=seq(0,3,length=200);
> y1=dexp(x,1); y2=dexp(x,3);
> y3=pexp(x,1); y4=pexp(x,3);
> D=data.frame(probability=c(y1,y2,y3,y4));
> D$parameter[1:200]='lambda=1';D$parameter[201:400]='lambda=2'
> D$parameter[401:600]='lambda=1';D$parameter[601:800]='lambda=2';
> D$type[1:400]='PDF';
> D$type[401:800]='CDF';
> print(qplot(x,probability,data=D,geom='area',facets=parameter~type,
+              main='Exponential Distribution PDF and CDF Functions'));
```



The Normal or Gaussian RV

The normal or Gaussian RV X with parameters $\mu \in \mathbb{R}$ and $\sigma > 0$ has the following pdf

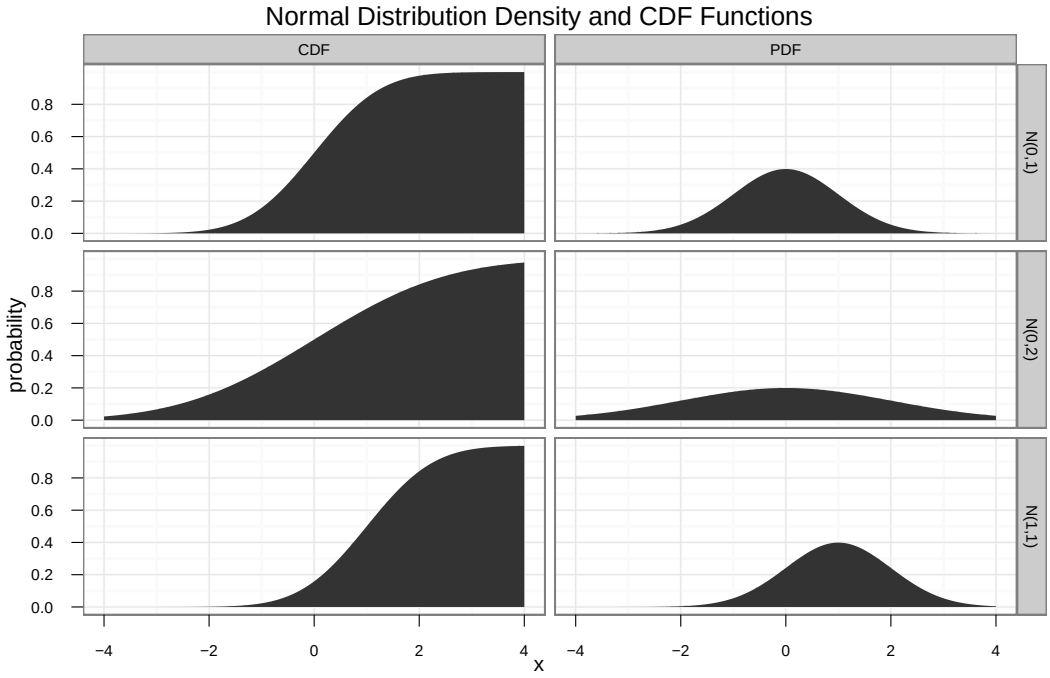
$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

The expectation is μ and the variance is σ^2. The cdf of the normal distribution does not have a closed form. It may be expressed through the function Φ - the cdf of a normal distribution with parameters $\mu = 0, \sigma = 1$

$$F_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} \exp\left(-\frac{(x'-\mu)^2}{2\sigma^2}\right) dx' = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(x-\mu)/\sigma} e^{-t^2/2} dt = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

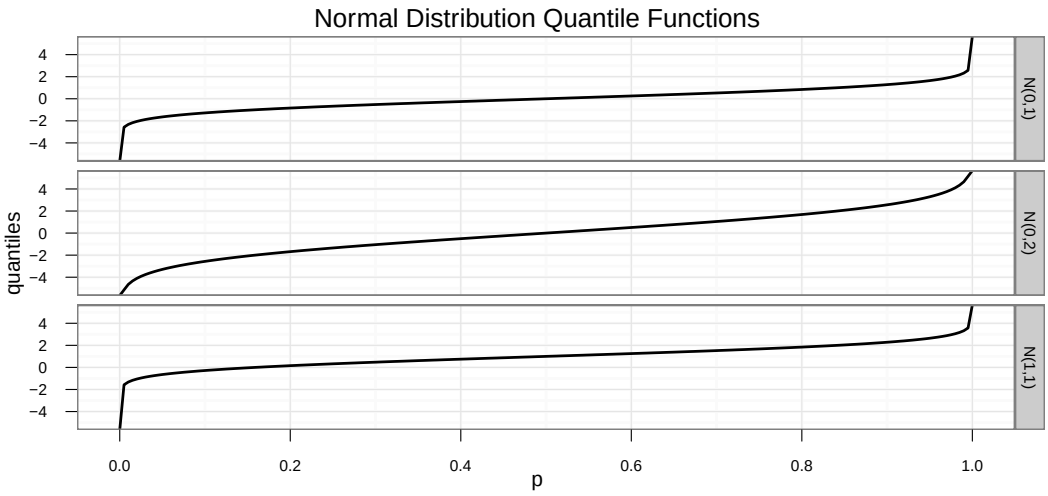
where we used the change of variables $t = (x' - \mu)/\sigma$. As a byproduct we get the useful result that if X is a Gaussian RV with some μ, σ, then $Y = g(X) = (X - \mu)/\sigma$ is a Gaussian RV with $\mu = 0, \sigma = 1$. Y is called a standard normal RV.

```
> x=seq(-4,4,length=200);
> y1=dnorm(x,0,1);y2=dnorm(x,1,1);y3=dnorm(x,0,2);
> y4=pnorm(x,0,1);y5=pnorm(x,1,1);y6=pnorm(x,0,2);
> D=data.frame(probability=c(y1,y2,y3,y4,y5,y6));D$x=x;
> D$parameter[1:200]='N(0,1)';D$parameter[601:800]='N(0,1)';
> D$parameter[201:400]='N(1,1)';D$parameter[801:1000]='N(1,1)';
> D$parameter[401:600]='N(0,2)';D$parameter[1001:1200]='N(0,2)';
> D$type[1:600]='PDF';D$type[601:1200]='CDF';
> print(qplot(x,probability,data=D,geom='area',facets=parameter~type,
+              main='Normal Distribution Density and CDF Functions'));
```



We also plot below the quantile function of the normal distribution.

```
> p=seq(0,1,length=200);
> y1=qnorm(p,1,1); y2=qnorm(p,0,1); y3=qnorm(p,0,2);
> D=data.frame(quantiles=c(y1,y2,y3));
> D$parameter[1:200]='N(1,1)';D$parameter[201:400]='N(0,1)';D$parameter[401:600]='N(0,2)';
> print(qplot(p,quantiles,data=D,geom='line',facets=parameter~.,
+              main='Normal Distribution Quantile Functions'));
```



Notice that keeping the variance fixed and changing the mean shifts the quantile function up or down. Keeping the mean fixed but chaning the variance makes the distribution move heavy tailed.