Week 3. Statistics I

August 23, 2019

1 Statistics I

This notebook is the first in a two part **introduction/refresher** to statistics for econometrics. This notebook focuses on the following distributions key distributions:

- The Bernoulli Distribution
- The Discrete Uniform Distribution
- The Continuous Uniform Distribution
- The Binomial Distribution
- The Normal Distribution
- The Chi-squared Distribution
- The t-distribution
- The F-distribution
- The Poisson Distribution
- The Logarithmic Distribution
- The Exponential Distribution

There is lots of information about these distribtions on Wikipedia and in various textbooks written by far better educated people. We don't go into proofs or cover each distribution exhaustively. Instead we provide **easy-to-follow** definitions of statistical concepts and mathematical objects and operators. What is attempted here is to give an intuitive feel for what these distributions look like and to decode some of the jargon for a non-expert (including formerly-expert!) audience. We include loads of **examples** to keep things flowing too. R is an extremely useful tool for this task because it lets us simulate what we know must be true from statistical theory, often without needing to import any extra packages (unlike in Python) and always free of charge (unlike Stata/SAS/SPSS etc).

1.0.1 The Bernoulli Distribution

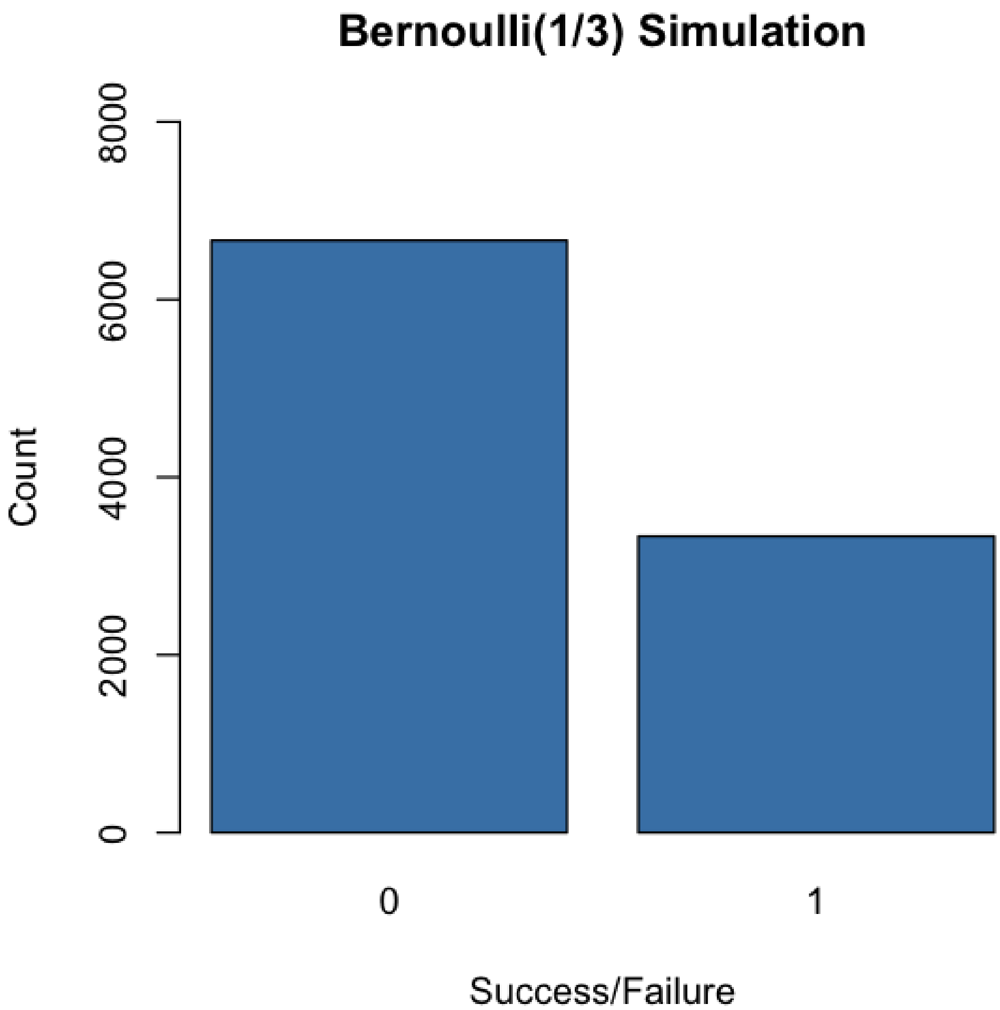
The Bernoulli distribution describes the probability an independent binary event occurs. A brief word of warning: ***it is very simple***. Most of us have an extremely intuitive understanding of this distribution because of how simple it is. For this reason having a strong understanding of it is key. Later we will see it is also the foundation of key distributions like the The Binomial Distribution and the The Logarithmic Distribution.

We can describe any binary event Bernoulli distributed by calling one outcome a *'success'* and denoting it with 1 and the other outcome a 'failure' and denoting it with a 0. The simplest example of a Bernoulli trial (or experiment) is a coin toss, in which we can call heads our success - occuring with probability $p = 0.5$. The Bernoulli distribution is called a **discrete distribution** as variables take a finite number (2) of potential outcomes. This is in contrast to a **continuous distribution** (e.g. The Normal Distribution) where there are an infinite number of outcomes.

A more interesting example of a Bernoulli experiment we can consider is rolling a number less than 3 on a fair dice. In this case rolling the numbers 1 or 2 represent a *'success'* and are denoted with 1, and rolling 3-6 is a *'failure'*, denoted with 0. We use X to denote our **random variable**, or the outcome of our dice roll. X is called a *random* variable because we don't know in advance what its outcome will be. So our dice roll X is distributed around $Bernoulli(1/3)$, occuring with probability: $P(X) = 1 = 1/3$

To get a better sense of how a Bernoulli random variable behaves we simulate 10,000 'dice rolls' and graph the number of 'successes' (1s and 2s) and the number of 'failures' (3s to 6s). Before we start our first simulation it is crucial to know that simulation is inherently *random*. I won't go into pseudorandom numbers here as we aren't looking for preceision or reproducibility but don't be shocked when these values change as you reload the notebook.

[1]:
```
options(repr.plot.width=5, repr.plot.height=5)
a<-data.frame(rbinom(10000,1,(1/3)))
b<-table(a[,1])
bp<-barplot(b,col='steelblue',ylim=c(0,8000),xlab='Success/
 ↪Failure',ylab='Count',main='Bernoulli(1/3) Simulation')
```



As expected our simulation recorded roughly 2/3rds failures and 1/3rd succeses. Why? To explain this introduce the **probability mass function** (pmf). The pmf of a random variable (for example our X from before) is simply a function which states the probability with which each potential outcome occurs. As we only have two potential outcomes, 0 and 1, for the Bernoulli distribution there are only two potential probabilities we need to consider. Pmf's are important because they give a complete description of how likely each outcome is to occur. Because the probablity of any event which is possible (so a 0 *or* a 1 for the Bernoulli distribution) must be 1, pmf's must sum to 1. In the case of our example our random variable X is distributed with the (pmf):

$ P(X) =

$$\begin{cases} 1/3 & X = 1 \\ 2/3 & X = 0 \\ 0 & \text{otherwise} \end{cases}$$

$

This is just an example case of the Bernoulli distribution's pmf, which is:

$ P(X) =

$$\begin{cases} p & X = 1 \\ 1 - p & X = 0 \\ 0 & \text{o/w} \end{cases}$$

$

We say the **expected value** of a Bernoulli distributed random variable is just p. The expected value of a random variable is the same as its **population mean** or the share of the total outcomes which would be our outcome if we could conduct an extremely large number of trials. This is simple enough to see when we introduce the mathematical definition of expected value for discrete random variables as: $E(X) = \sum_{i=1}^{n} p_i x_i$.

So for a Bernoulli random variable X this is just $E(X) = p$, or $E(X) = 1/3$ in the previous example. Clearly this fits with what our simulation showed us, as seen below:

[2]:
```
mean(a[,1])
```

0.3335

If you're still not sure how we got here just remember that Bernoulli random variables can only be 1s or 0s. If there is a p (0p1) chance of them being 1s then we should expect their average to be p. So if p=0.2 our sample could look like:

[3]:
```
rbinom(100,1,0.2)
```

1. 0 2. 0 3. 0 4. 0 5. 1 6. 0 7. 0 8. 1 9. 0 10. 0 11. 0 12. 0 13. 0 14. 0 15. 1 16. 0 17. 0 18. 1 19. 0 20. 0 21. 1 22. 0 23. 0 24. 0 25. 1 26. 0 27. 0 28. 1 29. 1 30. 0 31. 0 32. 0 33. 1 34. 0 35. 1 36. 0 37. 0 38. 1 39. 1 40. 0 41. 0 42. 0 43. 0 44. 0 45. 0 46. 0 47. 0 48. 1 49. 0 50. 1 51. 0 52. 1 53. 0 54. 0 55. 0 56. 0 57. 0 58. 0 59. 0 60. 1 61. 0 62. 1 63. 0 64. 0 65. 0 66. 1 67. 0 68. 0 69. 0 70. 0 71. 0 72. 0 73. 1 74. 0 75. 0 76. 0 77. 0 78. 0 79. 1 80. 1 81. 1 82. 0 83. 0 84. 1 85. 0 86. 0 87. 0 88. 0 89. 0 90. 0 91. 1 92. 0 93. 0 94. 0 95. 0 96. 0 97. 0 98. 0 99. 0 100. 0

In which the breakdown of 1s and 0s would be roughly

[4]:
```
table(rbinom(100,1,0.2))
```

```
 0  1
88 12
```

Hopefully you get the picture a little better by now

1.0.2 The Discrete Uniform Distribution

This is another relatively simple distribution which makes pretty good intuitive sense. The Discrete Uniform distribution describes a random variable which takes one of an finite number of values in a set, each with the exact same probability (hence the 'uniform'). When we say set we simply mean a collection of objects. We write them using braces (curly brackets) and by convention call them S (for set!). For example $S =${1,2,3,4,5} is a set. A good intuitive example of a Discrete Uniform distributed random variable is the roll of a dice. While our Bernoulli dice example was contrained to either a 1 or 0, we can write the pmf of rolling a dice as:

$ P(X) =

$$\begin{cases} 1/6 & X = 1,2,3,4,5,6 \\ 0 & \text{o/w} \end{cases}$$

$

To decode, this means our dice roll takes any value in {1,2,3,4,5,6} with the same probability, 1/6. This generalises to:

$ P(X) =

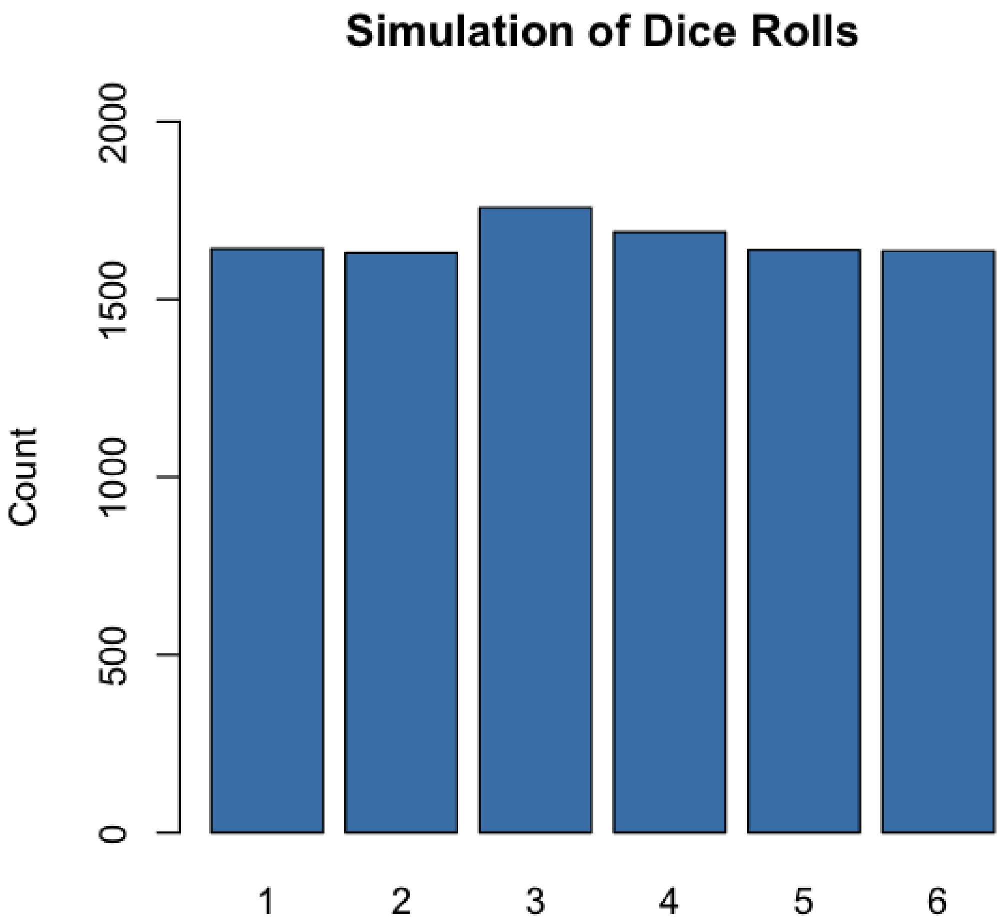
$$\begin{cases} 1/n & \forall X \in S \text{ where |S|=n} \\ 0 & \text{o/w} \end{cases}$$

$

Which is mathspeak for X has a probability $1/n$ for all X in the set S, where S has size n.

Let's look at another simulation to get a sense of the Discrete Uniform distribution. Once we understand this it will be much easier to take on the Continuous Uniform distriubtion. I really wish I could resist but I have to say it: Alea Iacta Est (loosely translated to: Let's Roll Some Dice - Julius Caesar)

[5]:
```
a<-data.frame(sample(1:6,10000,replace=T))
b<-table(a[,1])
bp<-barplot(b,ylab='Count',ylim=c(0,2000),col='steelblue',main='Simulation of␣
 ↪Dice Rolls')
```



Hopefully this seems intuitive. We rolled our fair (simulated) die 10,000 times and each number came up roughly 1/6th of the time. We should expect this. We know from before that the expected value of X is $E(X) = \sum_{i=1}^{n} p_i x_i$.

Using the distribution we introduced above ($ P(X) =

$$\begin{cases} 1/6 & X = 1,2,3,4,5,6 \\ 0 & \text{o/w} \end{cases}$$

$) we can sub in the values of our example to find that: $E(X) = \sum_{i=1}^{6} p_i x_i = \sum_{i=1}^{6} \frac{1}{6} x_i = \frac{1}{6} \sum_{i=1}^{6} x_i = \frac{1}{6}(1+2+3+4+5+6) = 3.5$

Which fits pretty well with our observed value of $\bar{X}$, as seen below:

[6]:
```
mean(a[,1])
```

3.4964

1.0.3 The Continuous Uniform Distribution

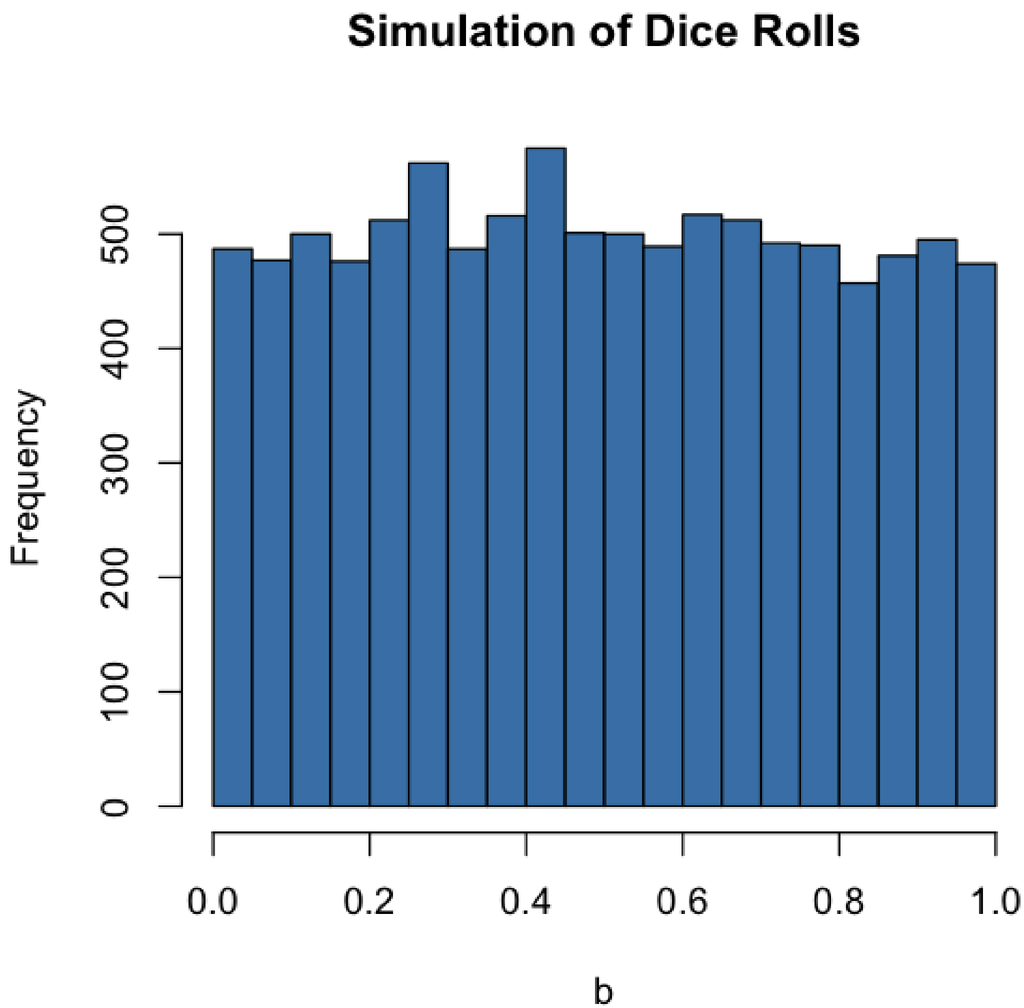
This our first 'hard' distribution. By hard we simply mean unlike its discrete cousin or the Bernoulli distribution we don't have an easy intuitive understanding for Continuous Uniform distributed random variables. They don't exist in our daily lives in as obvious ways! The Continuous Uniform distribution describes a random variable which takes one of an infinite number of values between two real numbers (a and b). The name **continous** just comes from the fact a random variable with this distribution can take an infinite number of values. Don't be put off by infinity though, there are an infinte number of real numbers between 0 and 1!

Like we said, there is no good natural examples of this distribution. The closest we can come to a good (artificial) example of a Continuous Uniform distributed random variable is by imagining we want to pick a number between 0 and 1 absolutely at random. We could pick 0.01, 0.5, 0.93234324324 and any other number between 0 and 1 we wanted. To do this in a completely random way, we'd need the Continuous Uniform distribution.

The obvious next step for us is to simulate a *Continuous* Uniform distribution, but before we do we need to think about one issue. In the Discrete Uniform distribution we had a finite set of n values, or in the dice case, 6. Each of these has probability $1/n$. So the problem when we have infinite potential is that the probability of any one value occuring is 0, as $\lim_{n\to\infty} \frac{1}{n} = 0$. Obviously every possible outcome having an exactly 0 probability of occuring is not great news when we're trying to do statistics, so we need a new approach…

But first: a simulation

[7]:
```
a<-data.frame(runif(10000))
b<-a[,1]
bp<-hist(b,ylab='Frequency',col='steelblue',main='Simulation of Dice Rolls')
```



We simulated 10,000 draws from the $U(0,1)$ distribution and found that each bin of length 0.05 contained around 500 observations. This makes sense as the interval [0,1] has 20 of these bins (take [0.15,0.2] or [0.8,0.85] for example), so 10,000 observations from $U(0,1)$ should lead to about 10000/20=500 observations per bin! This is just like in our Discrete Uniform case except instead of the observations being integers they are any real number. For example a random sample of 10 points from our discrete distribution could look like:

[8]:
```
sample(1:6,10,replace=T)
```

1. 3 2. 2 3. 2 4. 5 5. 6 6. 3 7. 1 8. 1 9. 5 10. 4

But from our Continuous Uniform example it could look like:

[9]:
```
runif(10)
```

1. 0.994200188899413 2. 0.970767290331423 3. 0.935012394096702 4. 0.133364760084078 5. 0.0606731963343918 6. 0.183409092249349 7. 0.700597072020173 8. 0.35861148010008 9. 0.907729021739215 10. 0.106625801883638

The last thing for us to cover before moving on is the expected value of a $U(a,b)$ distributed random variable X, or $E(X)$. In the discrete case we calculated expected value using $E(X) = \sum_{i=1}^{n} p_i x_i$, but as mentioned $p_i = 0$ for every i in our distribution and there are infinite i's! This points us in a calculus direction. For continuous random variables we calculate expected value using $E(X) = \int_{-\infty}^{\infty} x f(x) \mathrm{d}x$, where $f(x)$ is the pdf of X.

We already know the pdf of the $U(a,b)$ distribution is $ f(x) =

$$\begin{cases} \frac{1}{b-a} & \text{for } x \in [a,b] \\ 0 & \text{o/w} \end{cases}$$

$, so our expected value is simply $E(X) = \int_{-\infty}^{a} 0 \mathrm{d}x + \int_{a}^{b} x \frac{1}{b-a} \mathrm{d}x + \int_{b}^{\infty} 0 \mathrm{d}x = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}$

So for our random number between 0 and 1 example, we would expect to choose 0.5 because that's exactly in the centre of our distribution! But of course we wouldn't be complete without a simulation:

[10]:
```
mean(b)
```

0.496373007915542

So over 10,000 samples we get a mean pretty close to 0.5.

1.0.4 The Binomial Distribution

The Binomial distribution is fundamental in statistics as it describes the probability of a sum of Bernoulli events (like a coin toss) occuring a given number of times. It is a discrete distribution, because obviously there are a finite number of possible sums of Bernoulli random variables. In fact in a sense we can say that we have already studied the Binomial distribution, because the Bernoulli distribution is just the Binomial distribution when n=1. A good example of when we can use the Binomial distribution is to find the probability of a fair coin showing heads 7 times in 10 flips. In R this is simply:

[11]:
```
dbinom(7,10,0.5)
```

0.1171875

Which is the equivalent of: $P(X = 7) = \binom{10}{7} 0.5^7 (1 - 0.5)^3$

Likewise we could have found the probability of our fair coin showing heads **at least** 7 times out of 10 with:

[12]:
```
pbinom(7,10,0.5)
```

0.9453125

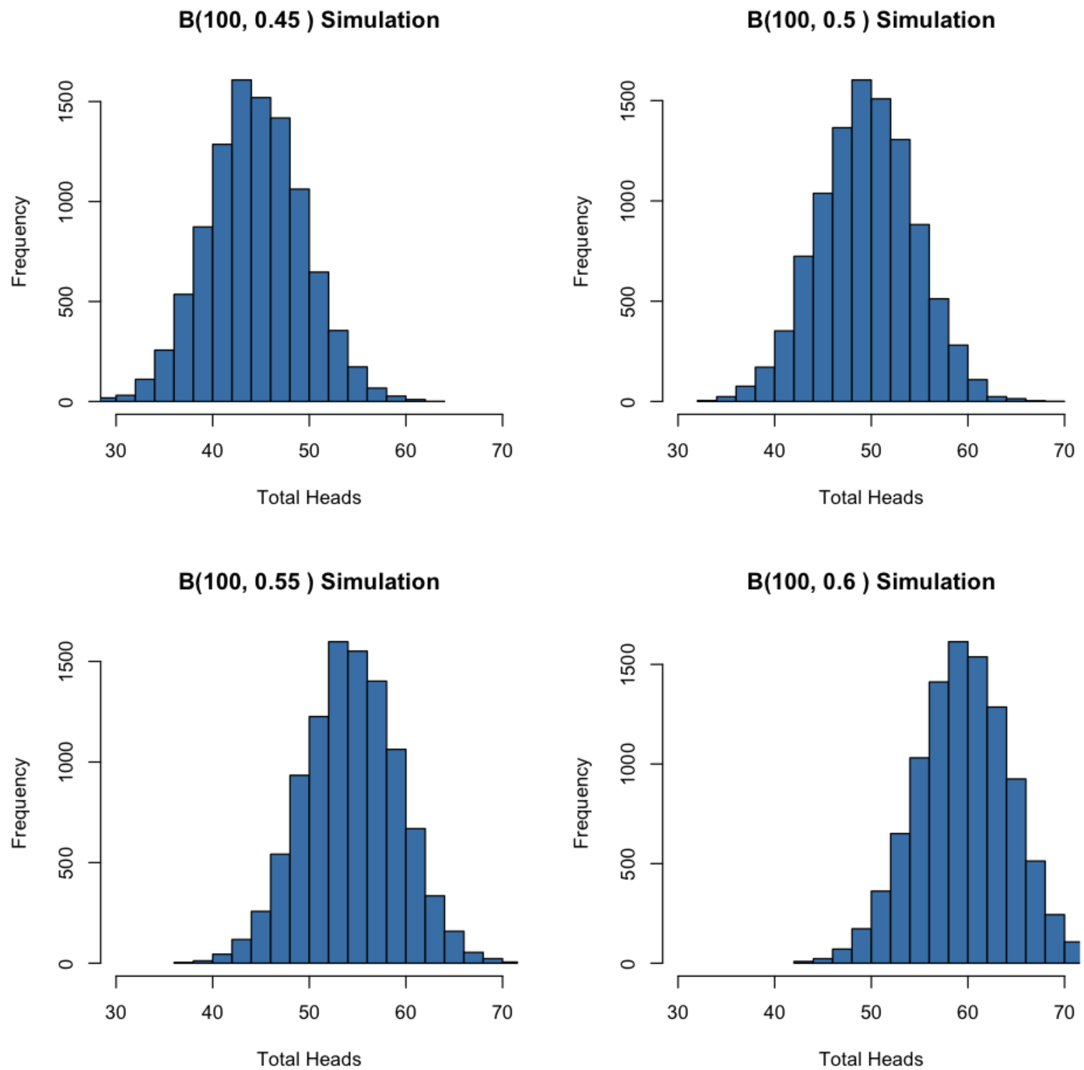
Which is the equivalent of: $P(X7) = \binom{10}{7}0.5^7(1-0.5)^3$

The Binomial distribution $B(n,k)$ observes the probability distribution function: $P(X=k) = \begin{cases} \binom{n}{k}p^k(1-p)^{n-k} & k \in [0,n] \\ 0 & \text{o/w} \end{cases}$

Where \$ n($_{k=\frac{n!}{k!(n-k)!}\$}$)

We see this more clearly by plotting the distribution for several different parameters:

[13]:
```
library(repr)
options(repr.plot.width=8, repr.plot.height=8)
par(mfrow=c(2,2))
for (i in 1:4) {p<-0.4+i/20
    c<-paste('B(100,',p,')')
                hist(rbinom(10000,100,p),col='steelblue',xlab='Total␣
 ↪Heads',main=paste(c, 'Simulation'),xlim=c(30,70))}
```



What we are seeing graphically is 10,000 simulations of a Binomial trial, using 4 different probabilities. Evidently the **mean** of the distributions increase as we increase the probability of successs (0.45, 0.5, 0.55, 0.6) but the shape of the distributions remains (almost exactly) the same. We'll get a better sense of this 'shape' question when we introduce the Normal distribution in the next section.

Another key question we need to answer for the Binomial distribution is what is its expected value. The simple answer is just for $X \sim B(n, p)$, $E(X) = np$. How did we get this result? Well we can think about integrating over our probability distribution like before, so: $E(X) = \sum_{k=0}^{n} k\binom{n}{k} p^k (1-p)^{n-k}$

For our coin toss example this is $E(X) = \sum_{k=0}^{10} k\binom{10}{k}(0.5)^k(1-0.5)^{10-k} = 10(0.5) = 5$. Let's check this with a simulation:

[14]:
```
mean(rbinom(10000,10,0.5))
```

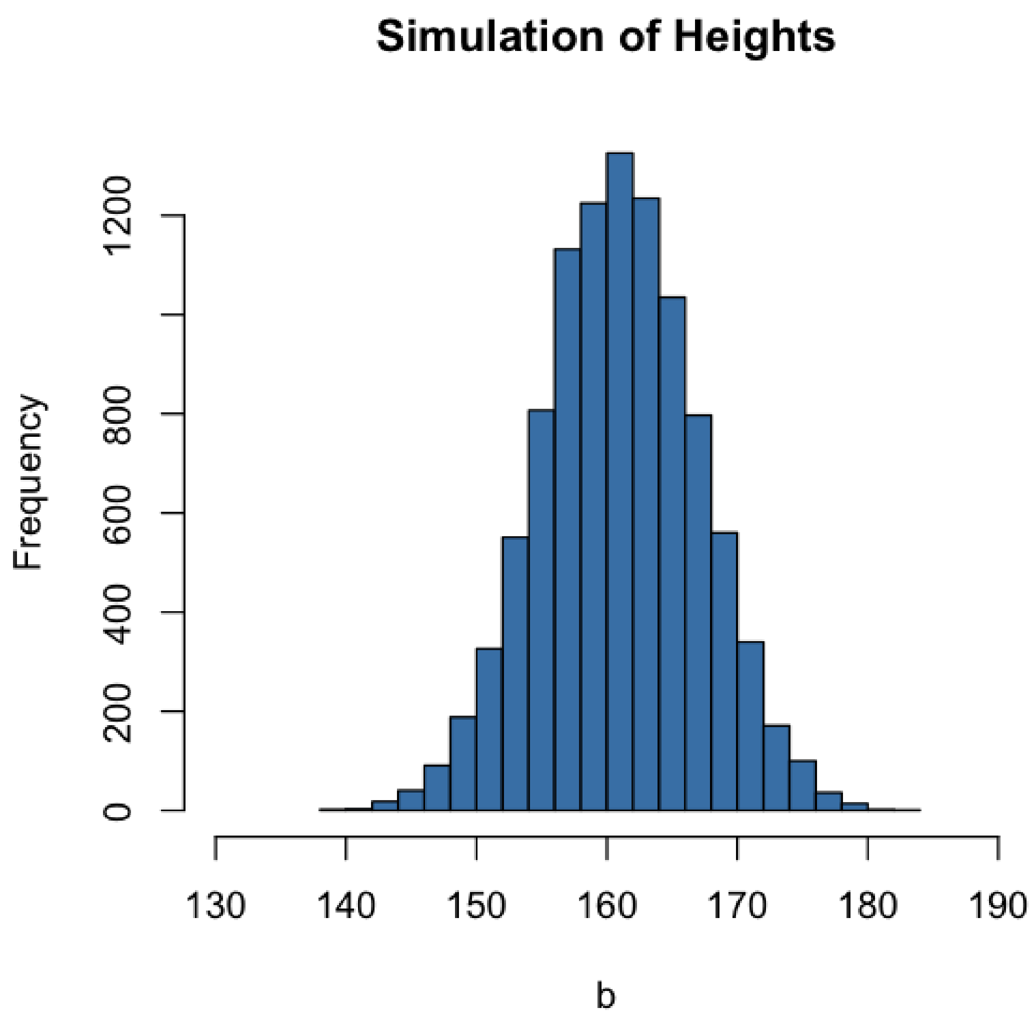
5.0122

So in 10,000 simulations of 10 coin tosses the average number of heads was almost exactly 5.

1.0.5 The Normal Distribution

At last we have reached the Normal distribution! Before we get into the Normal distribution just a brief heads up that this is *by far* the most important distribution we will cover. Not to worry though, we see the normal distribution around us every day. The Normal Distribution is continuouss distribution which is symmetric and bell-shaped, meaning that the bulk of the observations occur around the mean and become less likely at the same rate as they get further from it in both directions (smaller and larger than average). An extremely intuitive example is height. I'm not going to get too scientific but the average height for a British (adult) woman is around 161cm (5'3“). The Normal Distribution means that the proportion of British women who are at least 20cm taller than this, so taller than 181cm (5'11”), is around the same proportion who are at least 20cm shorter than this, or under 141cm (4'7"). We conduct a simulation to make more sense of this

[15]:
```
options(repr.plot.width=5, repr.plot.height=5)
a<-data.frame(rnorm(10000,161,6))
b<-a[,1]
hist(b,ylab='Frequency',col='steelblue',main='Simulation of␣
↪Heights',breaks=20,xlim=c(130,190))
```



In our simulation of the heights of 10,000 British women we see that only a tiny amount of the population is taller than 181cm or shorter than 141cm. The vast majority fall within the range of around 150-170cm.

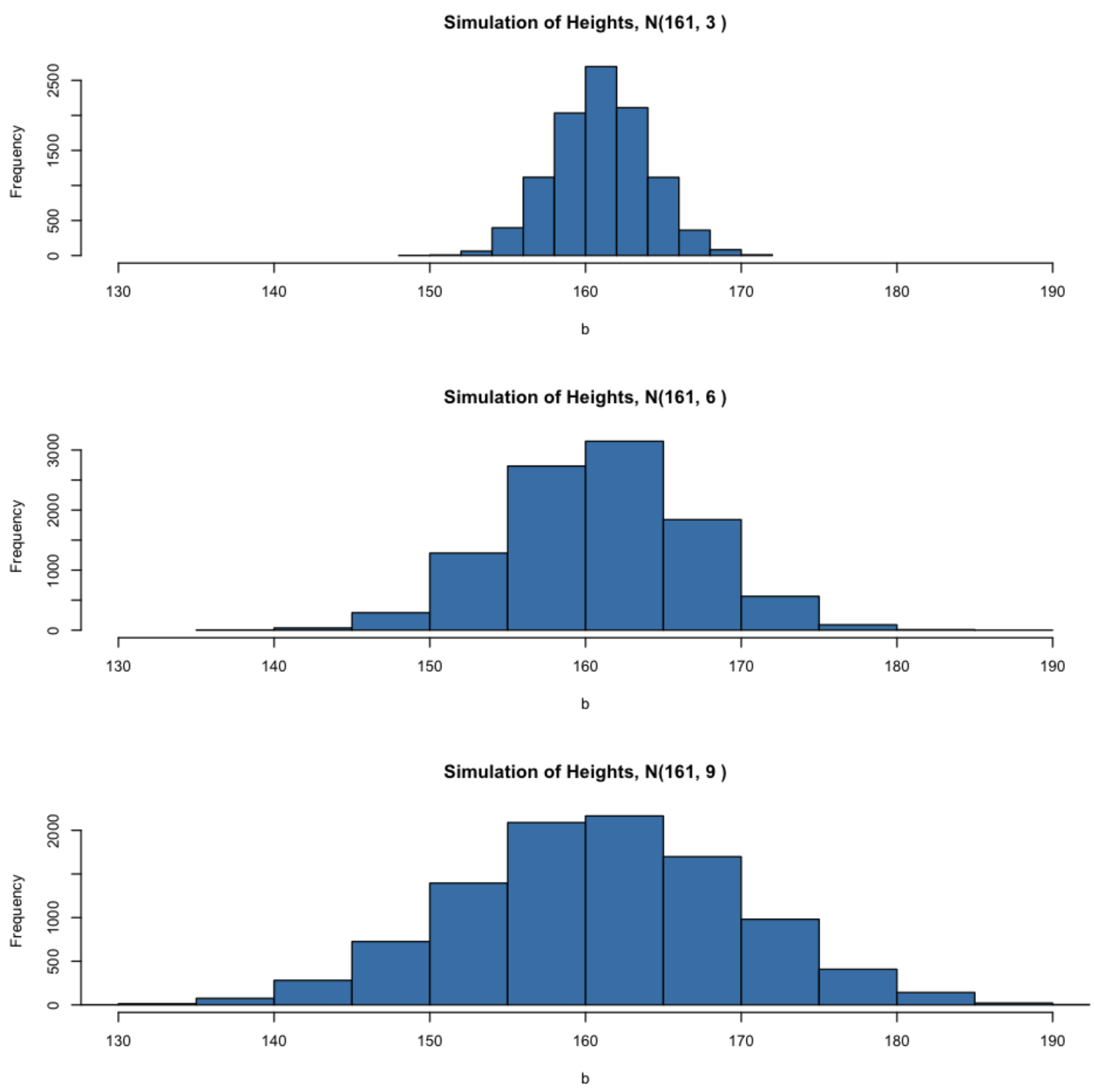
Before we can say more about the normal distribution we need to introduce a concept which has been notably absent so far, **standard deviation**. A data's standard deviation is a measure of the average difference between observations and their mean. Mathematically we calculate standard as expected but with a little twist. Instead of simply taking the average of the difference between each observation and the mean we find the average of the *squares* of these distances, then find the square root of that, so: $sd(X) = \sqrt{\sum_{i=1}^{n}(X_i - \mu)^2}$

What's the point of the square and the square root? Well that requires us to think about **variance** which is just standard deviation squared! The reason we square each value for our variance is so that our negative values and positive ones don't cancel out and leave us with 0. It's easy to see that $\sum_{i=1}^{n}(X_i - \mu) = \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu = \sum_{i=1}^{n} X_i - n\mu = \sum_{i=1}^{n} X_i - n\frac{1}{n}\sum_{i=1}^{n} X_i = 0$

Squaring each difference solves this problem and the average of these squares gives us variance. In a really simple sense, it's how much values vary! Standard deviation is used because the squaring also squares the units, so we'd meassure the variance of a company's net income in ^{2}, which isn't very helpful. So standard deviation is the square root of variance to get us back to £. Let's see how variance/standard deviation affect our normal distributions by extending our height example.

[16]:

```
options(repr.plot.width=8, repr.plot.height=8)
par(mfrow=(c(3,1)))
for (i in 1:3) {a<-data.frame(rnorm(10000,161,3*i))
b<-a[,1]
    hist(b,ylab='Frequency',col='steelblue',main=paste('Simulation of Heights, N(161,',3*i,')') ,xlim=c(130,190))}
```



As we can see our distributions have spread out a lot as standard deviation increased from 3 to 6 and then to 9. A reasonable proportion of women are between 140-145cm in our high variance (final) histogram, a tiny fraction are in our middle (true population) histogram and exactly 0 are in our top one.

Because variance is the average of squared deviations from the mean it can also be written as $Var(X) = E((X - E(X))^2) = E(X^2) - E(X)^2$, which uses the **second population moment**. We have already seen the **first population moment**, $E(X) = \int\limits_{-\infty}^{\infty} xf(x)\mathrm{d}x$, quite a lot, so it shouldn't be too surprising to see $E(X^2) = \int\limits_{-\infty}^{\infty} x^2 f(x)\mathrm{d}x$.

The standard deviation and variance are vital concepts because they form one of the two key variables in a Normal distribution; along with the mean. When we say a variable X is normally distributed, we describe the particular normal distribution it follows as $X \sim N(\mu, \sigma^2)$, where μ is the population mean (161cm in our height example) and σ is our standard deviation (which we graphed as 3,6 and 9cm above). Including the standard deviation is crucial because, as we have seen above, random variables which are normally distributed with the same mean *don't* have to follow the same distribution.

In a sense this is quite a strange way to think about our distribution, because we have already been given the population mean μ! In all our other examples we have had to derive it using $E(X)$. What may be even stranger is that we actually need the population mean and the population variance (σ^2) to construct the pdf, which is: $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$. This formula probably looks pretty scary right now but don't worry we aren't going to use it in this notebook. It's just included for illustration.

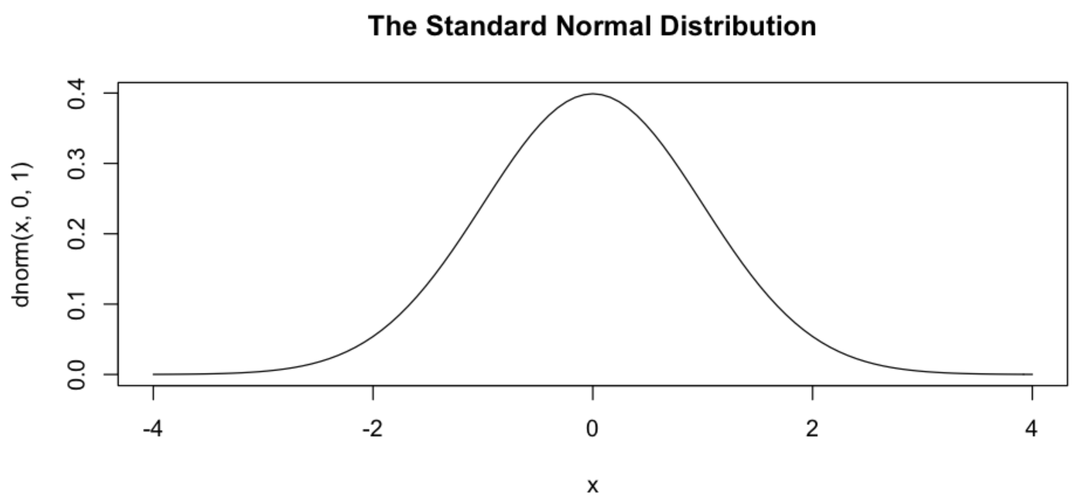
There is a LOT more we could say about the Normal distribution, importantly how it relates to the other distributions we have already covered, but we leave that for Important Distributions II.

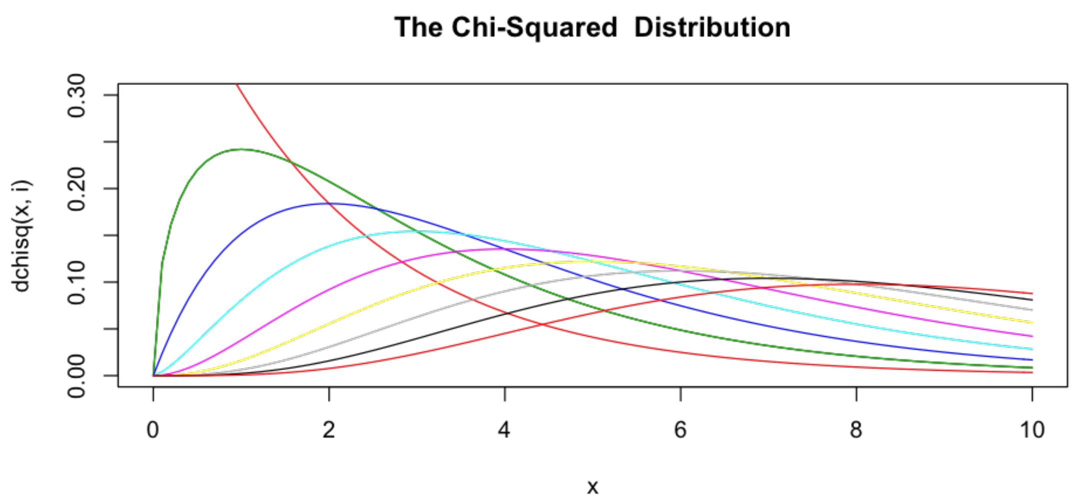
1.0.6 The Chi-Squared Distribution

The Chi-squared distribution is a direct product of the Normal distribution, just like the Binomial distribution is a direct product of the Bernoulli distribution. An extremely important special case of the Normal distribution is the Standard Normal distribution, or $N(0, 1)$, which we graph below. When a random variable, say X, is distributed around $N(0, 1)$, then X^2 is distributed around the Chi-squared distribution χ^2_1. The subscript 1 denotes 1 **degree of freedom**, or the fact that there is 1 (squared) random variable. When two random variables, say X and Y are distributed around $N(0, 1)$, then $X^2 + Y^2$ is distributed around the Chi-squared distribution χ^2_2, and for $X_i \sim N(0, 1)$ $\forall i$, $\sum\limits_{i=1}^{n} X_i^2 \sim \chi^2_n$.

[17]:
```
par(mfrow=c(2,1))
curve(dnorm(x,0,1),xlim=c(-4,4),main='The Standard Normal Distribution')
curve(dchisq(x,i),main='The Chi-Squared ␣
↪Distribution',col=1,xlim=c(0,10),ylim=c(0,0.3))
```

```
for (i in 2:10){curve(dchisq(x,i),main='The Chi-Squared ␣
↪Distribution',col=i,add=T)}
```





We don't usually use the Chi-squared distribution much on its own, but it is an essential component of the following two distributions: the t-distribution and the F-distribution. So some colourful lines will suffice for our purposes.

1.0.7 The t-distribution

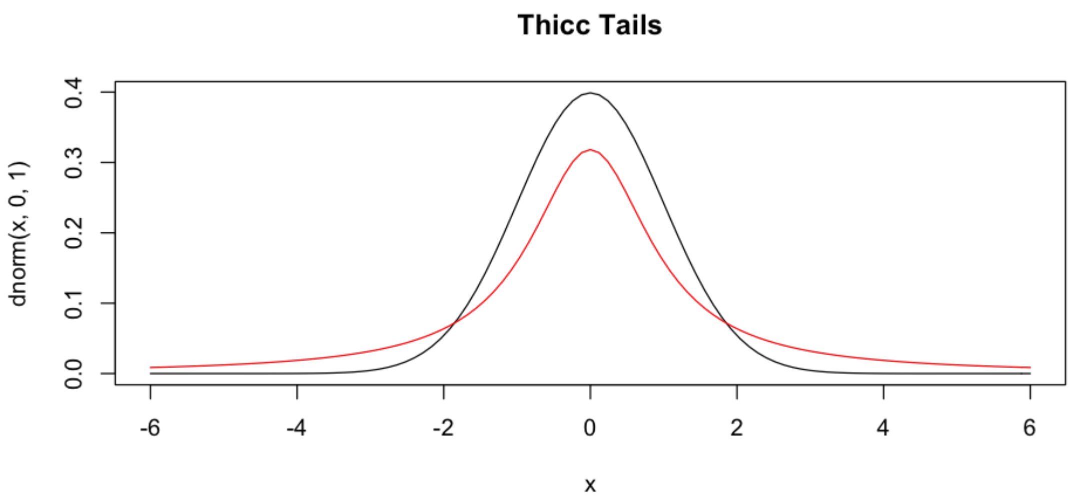
The t-distribution is essential in econometrics. It is used for conducting **hypothesis testing**, like whether our regression coefficients are actually meaningful or whether they're likely to show up through random chance. The t-distribution is very very similar to the Standard Normal distribution. It is symmetric, bell-shaped and continous, but has one key difference: **fatter tails**. Fat tails simply mean that outliers are more common than in the Standard Normal distribution, or that there is more area in the 'tails' of the distribution on either side of the mean. This is useful when we are dealing with more uncertainty (real-world data is not that friendly) and we want to be a little more hesitant before saying "There's no way that could have happened through chance alone".

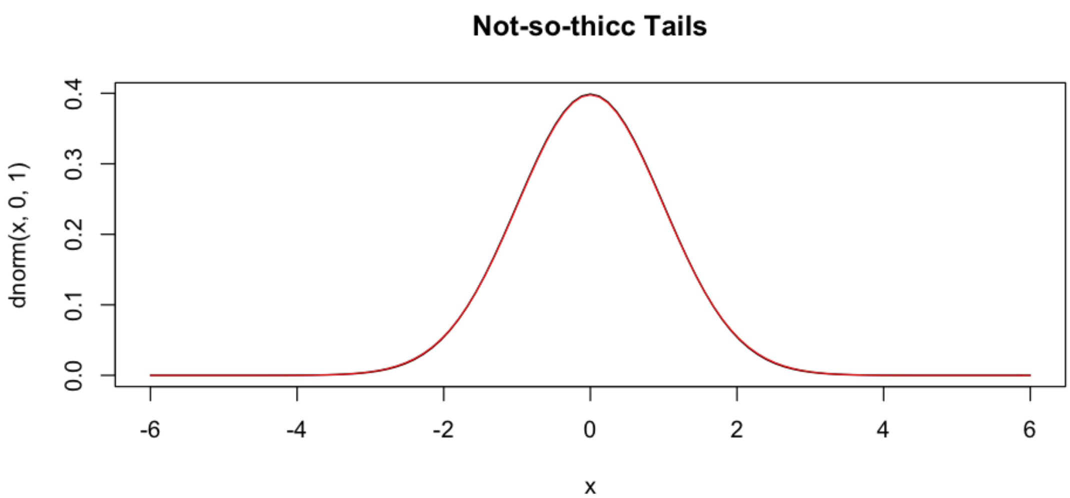
A random variable Z follows the t-distribution t_n if it can be written as $Z = \frac{X}{\sqrt{\frac{Y}{n}}}$ where $X \sim N(0,1)$ and $Y \sim \chi^2_n$.

This probably all sounds really technical and irrelevant, especially considering all the useful examples we had before, but it turns out when we do regressions in econometrics that a lot of our coefficients will follow the t-distribution. If you want to learn more about this then head to the econometrics section and check out how we use t-distributions for hypothesis testing.

The last thing to note is that as the degrees of freedom of the t-distribution, which is just the degrees of freedom of the Chi-squared it is made up of, increases, the t-distribution approaches the Standard Normal distribution. So a t_{100} distribution is essentially the same as a Standard Normal distribution - it loses its fat tails! We see this below:

[18]:
```
par(mfrow=c(2,1))
curve(dnorm(x,0,1),xlim=c(-6,6),main='Thicc Tails')
curve(dt(x,1),main='The Chi-Squared  Distribution',col=2,add=T)
curve(dnorm(x,0,1),xlim=c(-6,6),main='Not-so-thicc Tails')
curve(dt(x,100),main='The Chi-Squared  Distribution',col=2,add=T)
```



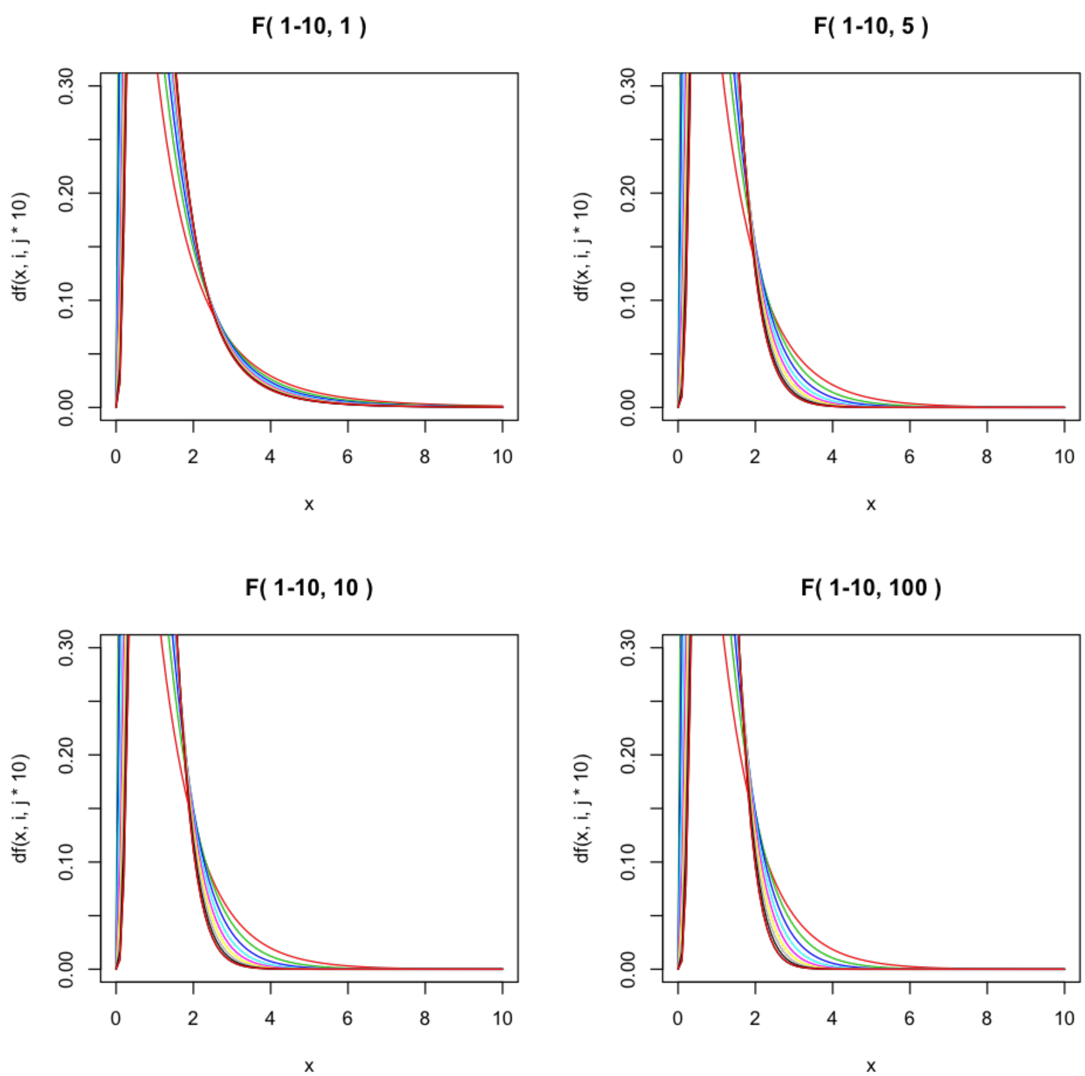


1.0.8 The F-distribution

The F-distribution is other main distribution based on the Chi-squared distribution. A random variable Z is distributed around the F-distribution $F_{n,m}$ if $Z = \frac{X/n}{Y/m}$ where $X \sim \chi^2_n$ and $Y \sim \chi^2_m$. It's not a symmetric distribution as we can see by plotting some more colourful lines:

[19]:
```
options(repr.plot.width=8, repr.plot.height=8)
par(mfrow=c(2,2))
for (j in c(1,5,10,100))
{curve(df(x,i,j*10),main=paste('F(','1-10,',j,')'),col=1,xlim=c(0,10),ylim=c(0,0.
↪3))
```

```
for (i in 2:10){curve(df(x,i,j*10),main='The Chi-Squared
  Distribution',col=i,add=T)}}
```



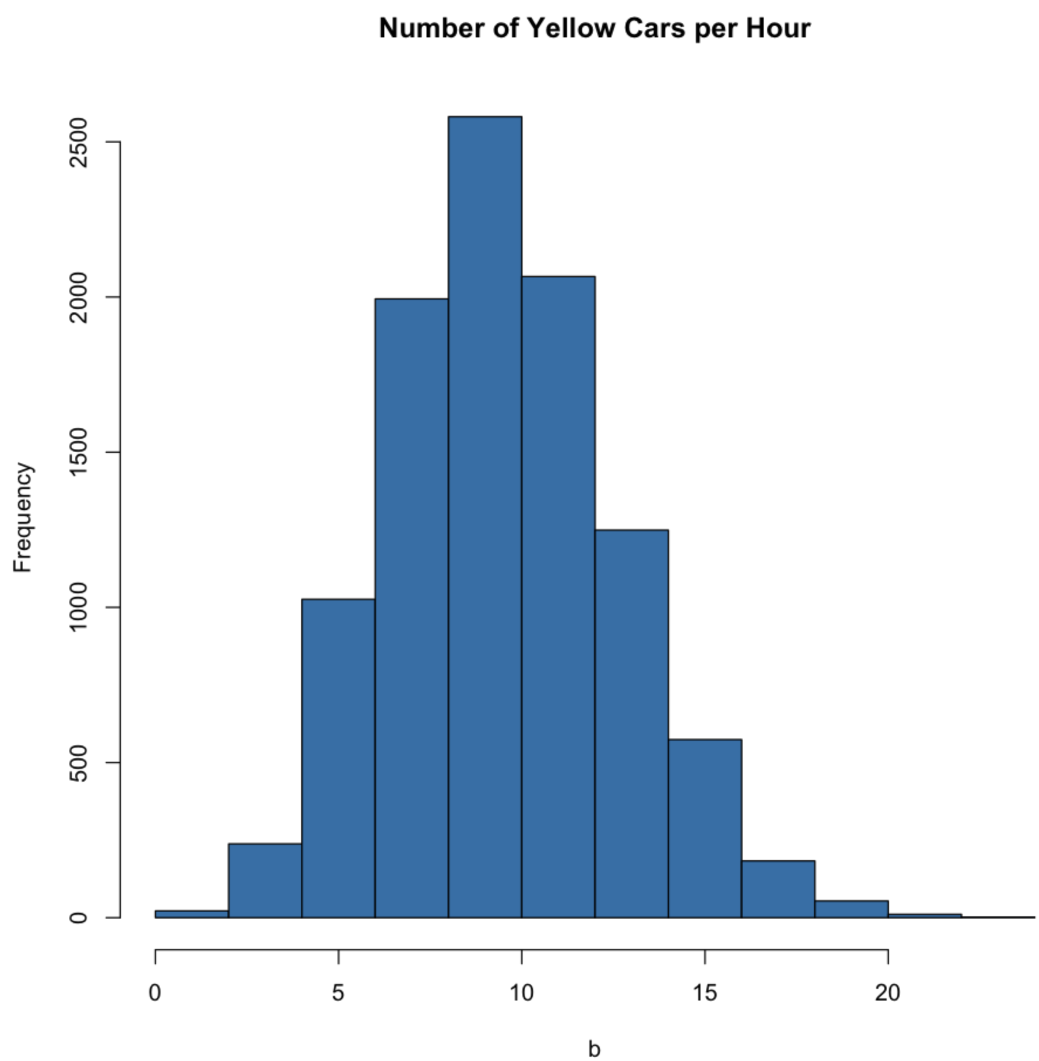
Once again the F-distribution seems extremely tedious and useless but some very key estimators in econometrics are distributed around it, so it's essential to know for anyone looking to skill up their metrics.

1.0.9 The Poisson Distribution

With the Poisson distribution we are out of the woods in the sense that we are back dealing with intuitive examples for a distribution which we have pretty nice formulas for. The Poisson distribution describes the number of occurences which happen at even intervals in a given unit of time. For example if you sat by the road and counted the number of yellow cars which passed you every hour (an important statistic for wherever 'spotto' is played), and you found on average 5 cars passed per hour, then the Poisson(10) distribution would be able to tell you the likelihood that 10 yellow cars pass you in the next hour, or 100, or 1 in the next minute. The crucial thing we need to keep in mind is that these events must be *independent*, so one happening doesn't change the chance of another happening, and occur at a constant rate. If yellow cars like to follow each other then independence wouldn't hold anymore and so our Poisson distribution wouldn't apply. And if there are times when yellow cars are more likely to be on the road then this could affect our constant rate assumption and mess up our Poisson distribution too.

Let's simulate our cars example.

[20]:
```
a<-data.frame(rpois(10000,10))
b<-a[,1]
hist(b,col='steelblue',main='Number of Yellow Cars per Hour')
```



In this example we sat by the highway for 416.66 days, or 10,000 hours (so we're officially experts!), and counted the numbers of yellow cars passing per hour. Clearly some hours there weren't any, and some hours there were more than 20. In fact we see that:

[21]:
```
max(a)
```

23

Which is clearly a lot of yellow cars considering our average is 10, but also not that many in a way; especially considering 100 yellow cars in an hour seemed plausible 5 minutes ago. But as usual we would have known this is we had an understanding of the pmf of the Poisson(λ) distribution:

$

$$\begin{cases} \frac{\lambda^k e^{-\lambda}}{k!} & \text{for } k = 0, 1, 2, ... \\ 0 & \text{o/w} \end{cases}$$

$

so for the Poisson(10):

$

$$\begin{cases} \frac{10^k e^{-10}}{k!} & \text{for } k = 0, 1, 2, ... \\ 0 & \text{o/w} \end{cases}$$

$

If we knew this beforehand then we could quickly see that P(X=100)= $\frac{10^{100} e^{-10}}{100!}$, and

[22]:
```
10^100*exp(-10)/factorial(100)
```

4.86464918206771e-63

Which is basically 0. So what would our expected value be for the Poisson(λ) distribution? It's just λ! And the variance? λ again! We don't prove these properties but we can easily see what they were in our simulation:

[23]:
```
mean(b)
var(b)
```

9.9979

9.78567415741574

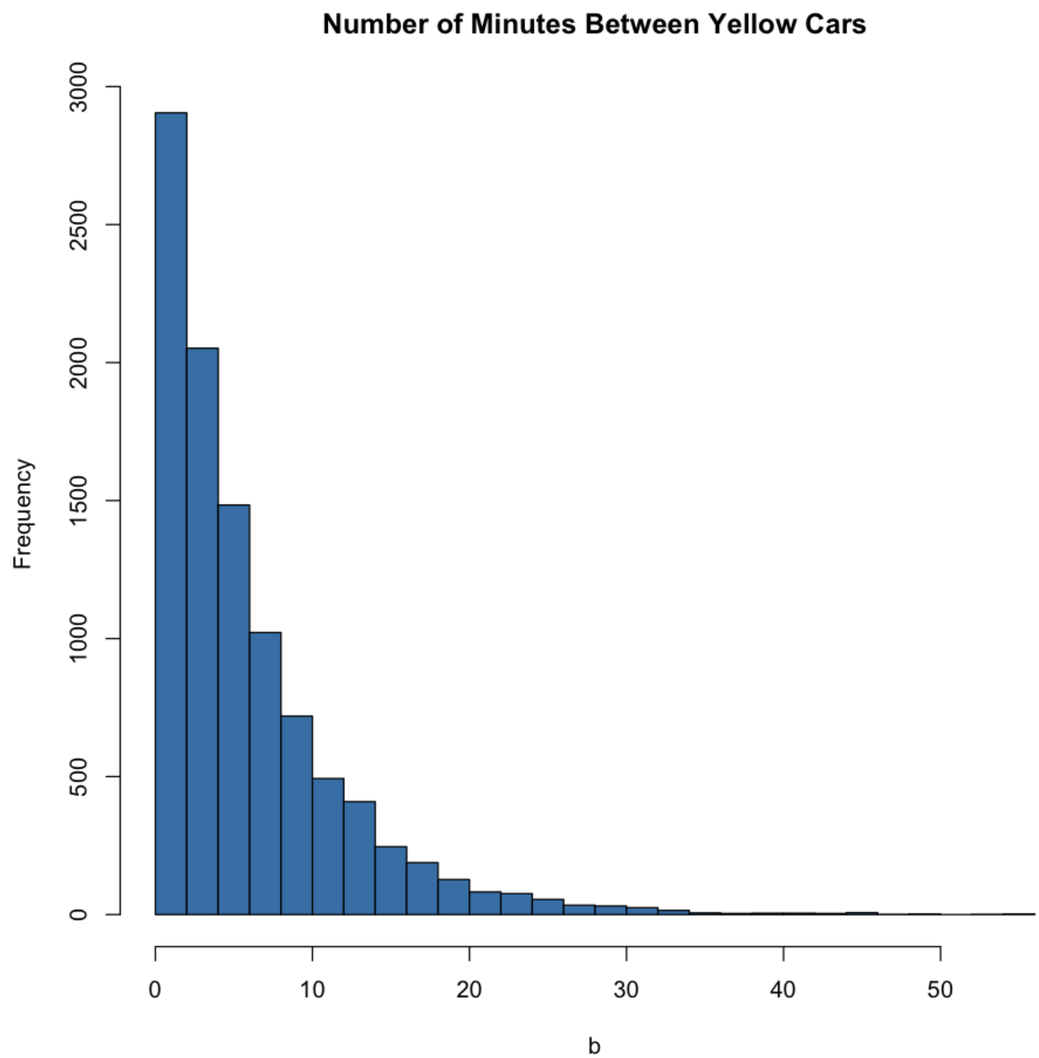
Esssentially spot on. Ok, only one more distribution to go now!

1.0.10 The Exponential Distribution

The Exponential distribution is the continuous cousin of the Poisson distribution. Where the Poisson distribution tells us the number of events per unit of time, the Exponential distribution tells us the amount of time between events. The same rules of independence and a constant rate still apply, but the good news is our previous example does too!

We simulate the amount of time it takes to wait for 10,000 yellow cars, each still showing up on average 10 times per hour.

[24]:
```
a<-data.frame(rexp(10000,10))
b<-a[,1]*60
hist(b,col='steelblue',main='Number of Minutes Between Yellow Cars',breaks=20)
```



I'm sure most of you who've gotten this far have a pretty good idea of what $E(X)$ should be when $X \sim exp(\lambda)$. 10 cars hour, so 6 minutes per car right?

[25]:
```
mean(b)
```

5.91346954045915

Right.

So $E(X) = \frac{1}{\lambda}$

This is also not too hard to see when we consider $f(x,\lambda) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x0 \\ 0 & \text{o/w} \end{cases}$ but we do need a bit of integration by parts which can be a hassle.

More integration by parts and we see $Var(X) = \frac{1}{\lambda^2}$

[26]:
```
var(b)
```

36.7203449070733

And there we have it! A brief and hopefully not too painful overview of some of the most useful distributions for econometrics. Please check out Important Distributions II for a more detailed look of what we've covered here, and once you've covered Linear Algebra, Statistics and Multivariate Calculus head over to the Econometrics session to get started on the Linear Regression model.