Probability Distributions

ORIGIN := 1

Statistics is based upon comparisons of measurements collected from one or more limited *samples,* with *expected values* characterizing the underlying *population* from which the samples have been drawn. In fact, these expected values are sometimes/always not easily determined. Important assumptions are always involved linking samples with populations and these assumptions underlie the usefulness of descriptive statistics, such as mean and variance.

The logic of statistics is typically based on a pair of quantities:

X < observed sample values
P(X) < probability of the sampled values given a specified model of probability.

Models of probability differ depending on what's being analyzed, and are generally of two types:

Continuous **< Here an infinite (or nearly so) number of observations are possible as in measuring temperature, length, weight, etc. of some animal.**

Discrete **< Here only a limited number of values are expected such as "heads" versus "tails" in a coin toss, or "1", "2", "3", "4", "5", or "6" in a roll of a single die.**

In either case, specific observations (X) are associated with probability P(X) using *Probability functions* where the area under the curve gives the probabilty for each value of x. In working with statistical tests, the classical way to estimate P(X) from X or the inverse was to consult obligatory tables. With the advent of the microcomputers, this is now much more efficiently handled by standard functions of four different types: d, p, q, and r.

Prototype in R:

```
? dnorm()
? dt()
? dchisq()
? df()
? dbinom()
```

Returns information about how to call standard d,p,q,r functions, e.g. for Normal Distribution:

The Normal Distribution
Description
Density, distribution function, quantile function and random generation for the normal distribution with mean equal to mean and standard deviation equal to sd.
Usage

```
dnorm(x, mean=0, sd=1, log = FALSE)
pnorm(q, mean=0, sd=1, lower.tail = TRUE, log.p = FALSE)
qnorm(p, mean=0, sd=1, lower.tail = TRUE, log.p = FALSE)
rnorm(n, mean=0, sd=1)
```

Example Discrete Probability functions:

coin toss:

There are two possible observations: H < "heads" = 1
T < "tails" = 2

$x := \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ **< here each class is given an arbitrary number**

For a fair coin: P(H) = 1/2
P(T) = 1/2

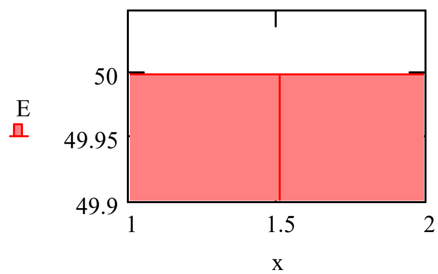
$P := \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$

For 100 coin tosses, expected number of H = 100(1/2) = 50
expected number of T = 100(1/2) = 50

$E := 100 \cdot \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$ $E = \begin{pmatrix} 50 \\ 50 \end{pmatrix}$

Expected probabilities:

Expected for each class for ^ n = 100 coin tosses



<- Two classes in x have equal numbers

single die:

There are six possible observations: "1" = 1
"2" = 2
"3" = 3
"4" = 4
"5" = 5
"6" = 6

$i := 1..6$

$x_i := i$

$x = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{pmatrix}$

For a fair die, all probabilities are 1/6 for obtaining one of the numbers on any throw:

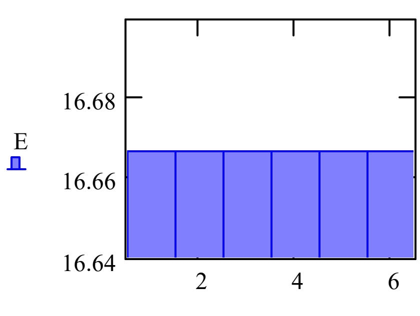
$P := \frac{1}{6}$ $P = 0.167$

For 100 die tosses, expected number for each:

$E_i := P \cdot 100$ $E = \begin{pmatrix} 16.667 \\ 16.667 \\ 16.667 \\ 16.667 \\ 16.667 \\ 16.667 \end{pmatrix}$

^ expected number

Expected probabilities:



<- Six classes in x with equal values that need not be whole numbers

ORIGIN := 0

Binomial distribution:

If one conducts multiple trials with two possible outcomes, such as tossing a coin resulting in either a "heads" or "tails", the expected number of "heads" in a set of trials follows the binomial distribution.

total number of trials (n): $n := 20$

probability of obtaining a heads (p) (more genererally termed "success") $p := \frac{1}{2}$

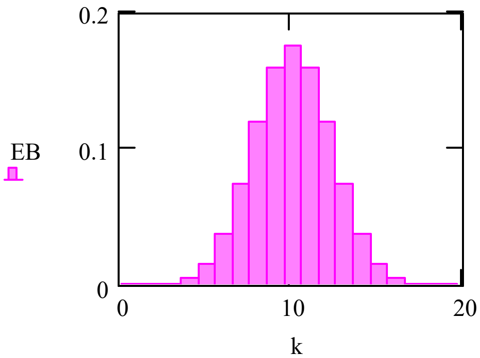
number of times one obtains a "heads" Note that this is a range of discrete possibilities (ranging from 0 to n) $i := 0 .. n - 1$

$k_i := i$

Expected probability for each k: (E_k): $EB := \text{dbinom}(k, n, p)$

EB =

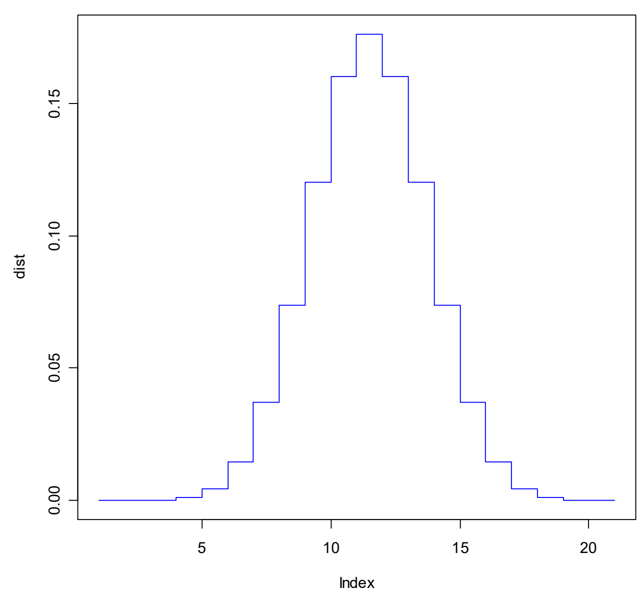
	0
0	$9.537 \cdot 10^{-7}$
1	$1.907 \cdot 10^{-5}$
2	$1.812 \cdot 10^{-4}$
3	$1.087 \cdot 10^{-3}$
4	$4.621 \cdot 10^{-3}$
5	0.015
6	0.037
7	0.074
8	0.12
9	0.16
10	0.176
11	0.16
12	0.12
13	0.074
14	0.037
15	0.015
16	$4.621 \cdot 10^{-3}$
17	$1.087 \cdot 10^{-3}$
18	$1.812 \cdot 10^{-4}$
19	$1.907 \cdot 10^{-5}$



< **It is unlikely to find 0 or 20 heads as outcome of 20 coin tosses. An intermediate number is much more likely.**

Prototype in R:

```
# BINOMIAL DISTRIBUTION:
# SETS UP VARIABLE X AS A RANGE
x=seq(0,20,1)
#NUMBER OF TRIALS:
n=20
#PROBABILITY
p=0.5
dist=dbinom(x,n,p)
plot(dist,type="s")
```



Example Continuous Probability Density functions:

Normal Distribution:

Many forms of data are continuous, so the probability function is continuous and the area under the curve represents probability (often called "probability density"). Normal distributions are common, and underlie many statistical methods.

$n := 50 \qquad i := 0..n$ **< Out of all possible values, we will arbitrarily look at a set of n points. At the scale we plot things here, this might as well be continuous...**

$b := -\left(\frac{1}{2} \cdot n\right) \qquad c := 0.1$ **< Arbitrary linear transformation so we can see things in the plot.**

$x_i := c \cdot (i + b)$ **< Individual transformed values we plot on our x axis below.**

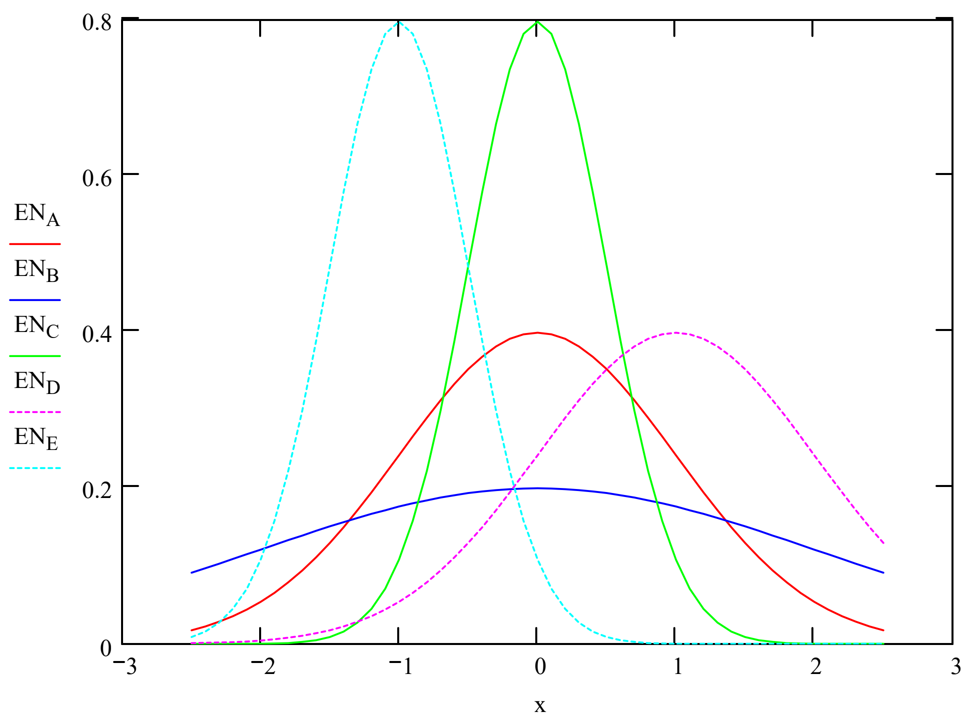
$\mu := 0 \qquad \sigma := 1$ **< parameters of the standard normal curve where μ is the mean of the distribution and σ is the standard deviation**

$EN_A := dnorm(x, \mu, \sigma)$

$EN_B := dnorm[x, \mu, (2\sigma)] \qquad EN_D := dnorm[x, (\mu + 1), \sigma]$

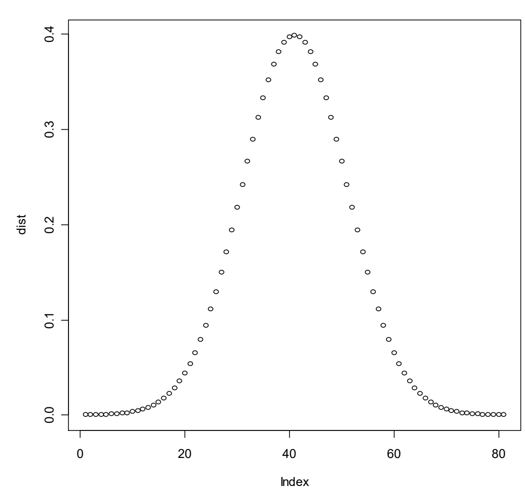
$EN_C := dnorm[x, \mu, (0.5\sigma)] \qquad EN_E := dnorm[x, (\mu - 1), (0.5 \cdot \sigma)]$

< The Normal distribution defines a family of curves with different values of μ and σ.



Prototype in R:

```
# NORMAL DISTRIBUTION:
# SETS UP VARIABLE X AS A RANGE
x=seq(-4,4,0.1)
# SPECIFY MEAN (mu) &
# STANDARD DEVIATION (sigma)
mu=0
sigma=1
dist=dnorm(x,mu,sigma)
plot(dist, type="p")
```



Student's t Distribution:

Student's t distribution is similar to the Normal Distribution, but *leptokurtic* - with greater concentrations of point in the tails and around the mean.

$n := 50 \quad i := 0..n$ — **< Out of all possible values, we will arbitrarily look at a set of n points. At the scale we plot things here, this might as well be continuous...**

$b := -\left(\frac{1}{2} \cdot n\right) \quad c := 0.1$ — **< Arbitrary linear transformation so we can see things in the plot.**

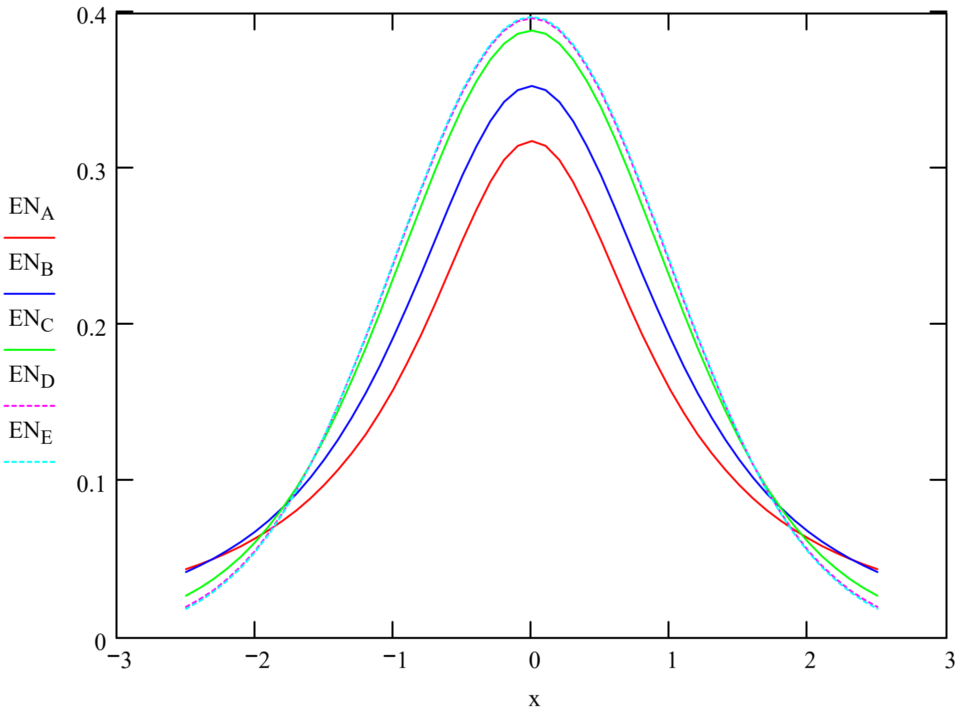
$x_i := c \cdot (i + b)$ — **< Individual transformed values we plot on our x axis below.**

$df_t := 1$ — **< "degrees of freedom" for the t distribution**

$EN_A := dt(x, df_t)$

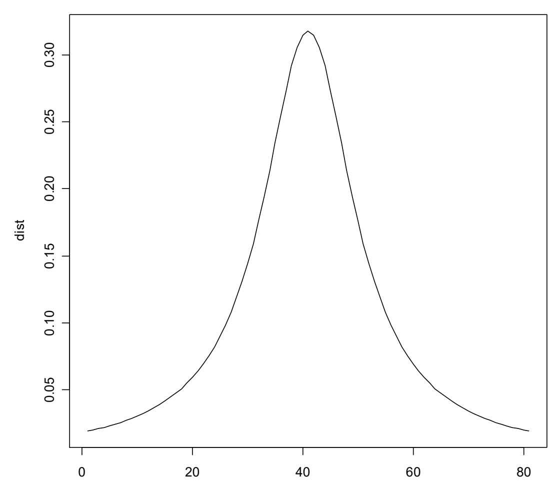
$EN_B := dt(x, 2 \cdot df_t) \qquad EN_D := dt(x, 50 \cdot df_t)$

$EN_C := dt(x, 10 \cdot df_t) \qquad EN_E := dt(x, 100 \cdot df_t)$



Prototype in R:

```
#STUDENT'S t DISTRIBUTION:
# SETS UP VARIABLE X AS A RANGE
x=seq(-4,4,0.1)
# SPECIFY DEGREES OF FREEDOM
dft=1
#PLOTTING WITH LINE TYPE "1" (the letter l)
dist=dt(x,dft)
plot(dist,type="l")
```



Chi-Square (χ^2) Distribution:

This distribution is commonly encountered in statistics, especially in what is known as "Goodness of Fit" tests.

$n := 50$ $i := 0..n$ **< Out of all possible values, we will arbitrarily look at a set of n point. At the scale we plot things here, this might as well be continuous...**

$b := 1.4$ $c := 0.3$ **< Arbitrary linear transformation so we can see things in the plot.**

$x_i := c \cdot (i + b)$ **< Individual transformed values we plot on our x axis below.**

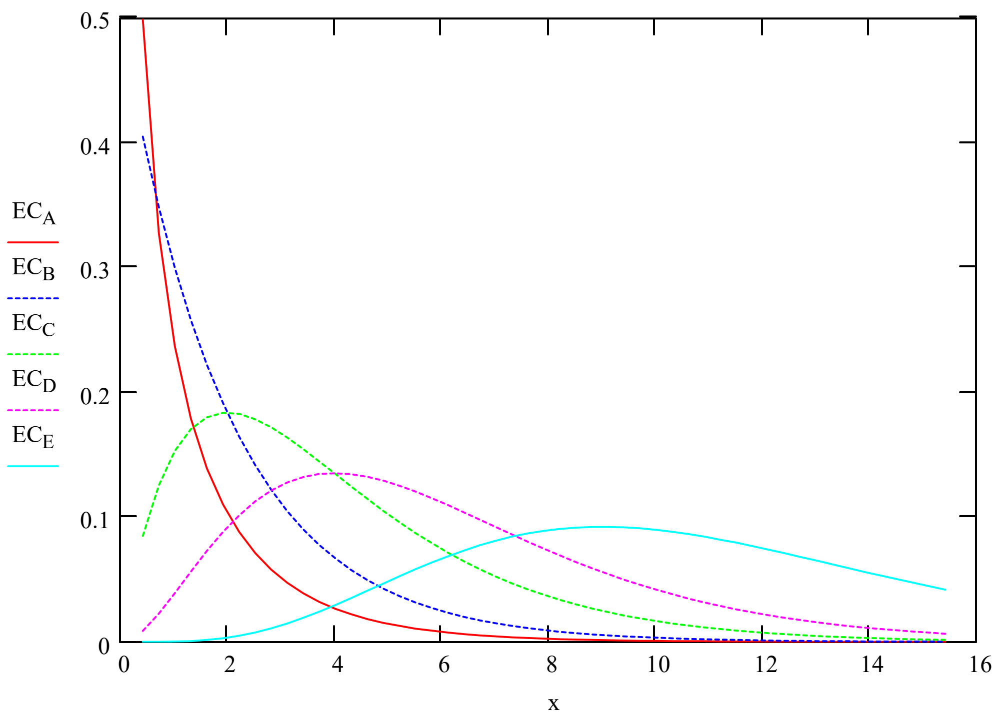
$df_{chisq} := 1$ **< df.chisq is a parameter for the χ^2 distribution called "degrees of freedom". Thus χ^2 defines a family of curves.**

$EC_A := \text{dchisq}(x, df_{chisq})$

$EC_B := \text{dchisq}[x, (df_{chisq} + 1)]$ $EC_D := \text{dchisq}[x, (df_{chisq} + 5)]$

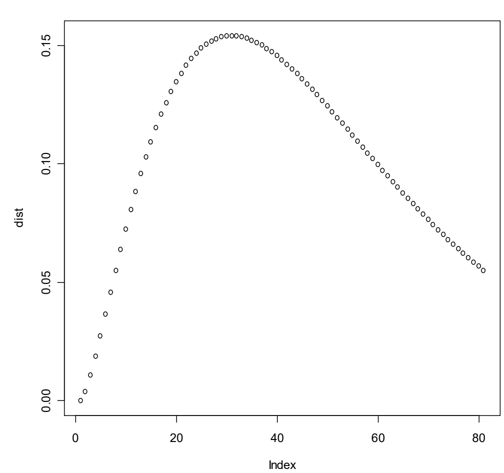
$EC_C := \text{dchisq}[x, (df_{chisq} + 3)]$ $EC_E := \text{dchisq}[x, (df_{chisq} + 10)]$

χ^2 family plotted below. As above, probability density is the area under each curve.



Prototype in R:

```
# CHI-SQUARE DISTRIBUTION:
# SETS UP VARIABLE X AS A RANGE
x=seq(0,8,0.1)
# SPECIFY DEGREES OF FREEDOM
dfchisq=5
#PLOTTING
dist=dchisq(x,dfchisq)
plot(dist)
```



F Distribution:

Typical distributions an a wide variety ANOVA and Regression tests.

$n := 50$ $\quad i := 0..n$

< Out of all possible values, we will arbitrarily look at a set of n point. At the scale we plot things here, this might as well be continuous...

$b := 1.4$ $\quad c := 0.3$

< Arbitrary linear transformation so we can see things in the plot.

$x_i := c \cdot (i + b)$

< Individual transformed values we plot on our x axis below.

$df_{F1} := 1$ $\quad df_{F2} := 1$

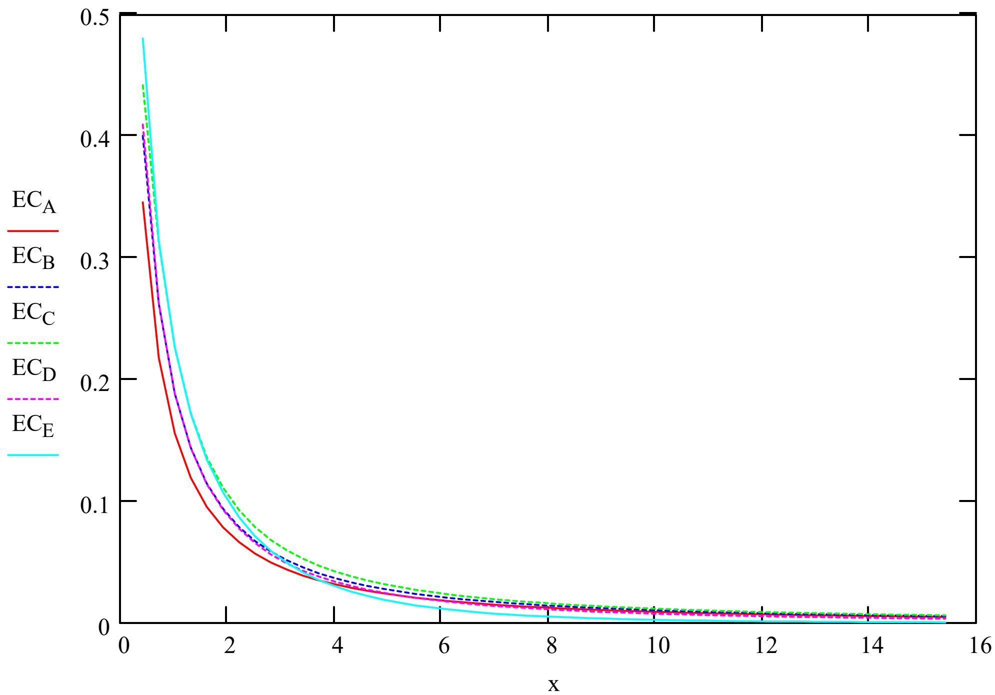
< df.F1 & df.F2 are numerator and denominator "degrees of freedom" respectively. Thus the F distribution defines a family of curves.

$EC_A := dF(x, df_{F1}, df_{F2})$

$EC_B := dF[x, (df_{F1} + 1), df_{F2}]$ $\quad EC_D := dF[x, df_{F1}, (df_{F2} + 1)]$

$EC_C := dF[x, (df_{F1} + 10), df_{F2}]$ $\quad EC_E := dF[x, df_{F1}, (df_{F2} + 10)]$

F family plotted below. Probability is density the area under each curve.



Prototype in R:

```
# F DISTRIBUTION:
# SETS UP VARIABLE X AS A RANGE
x=seq(0,8,0.1)
# SPECIFY DEGREES OF FREEDOM
dfF1=2
dfF2=7
#PLOTTING
dist=df(x,dfF1,dfF2)
plot(dist)
```