Probability Distributions in S-PLUS

Continuous

Distributions	root
beta	`beta`
Cauchy	`cauchy`
chi-square	`chisq`
exponential	`exp`
F	`f`
gamma	`gamma`
normal	`norm`
student's t	`t`
uniform	`unif`
Weibull	`weibull`

In the continuous case, `d`*root* returns the density, `p`*root* a probability, `q`*root* a quantile, `r`*root* a random number.

Example Probability

If X follows $N(0,1)$, then to find $P(X \leq 1.25)$, that is, the amount of area under the density curve to the left of $x = 1.25$,

```
> pnorm(1.25)
```

By default, the `norm` function assumes $\mu = 0, \sigma = 1$ (that is, you are working with the standard normal distribution). For other means and standard deviations, specify them in the argument. For example, if X follows $N(\mu = 2, \sigma = 3)$, then to find $P(X \leq 2.8)$,

```
> pnorm(2.8, 2, 3)
```

If T follows a t-distribution with 7 degrees of freedom, then to find the probability that $T \leq 3.9$, type

```
> pt(3.9, 7)  # pt(t-value, d.f)
```

If F follows an F-distribution with 4 and 8 degrees of freedom, then to find $P(F \leq 5.3)$:

```
> pf(5.3, 4, 8)
```

Quantiles

To find the 25^{th} percentile, that is, the value q such that $P(X \leq q) = .25$ for X from $N(0,1)$,

```
> qnorm(.25)
[1] -0.6744898
```

In other words, the amount of area under the density function to the left of $x = -0.6744$ is 0.25.

The .75 quantile for $N(2,3)$ can be found by

```
> qnorm(.75, 2,3)
[1] 4.023469
```

In other words, the amount of area under the density curve and to the left of $x = 4.023469$ is .75.

For T from a t-distribution with 13 degrees of freedom, to find value t such that $P(T > t) = .025$, which is equivalent to $P(T \leq t) = .975$, type

```
> qt(.975, 13)
```

Random numbers

To generate 100 random numbers from the normal distribution $N(0,1)$, type

```
> rnorm(100)
> x<- rnorm(100)
> hist(x)
```

Plotting the density curve

To plot the density curve for $N(0,1)$ for $-3 \leq x \leq 3$, first create a vector of x values:

```
> x <- seq(-3,3, by=.1)
> x  #vector of values from -3 to 3
> y <- dnorm(x)
> plot(x, y, type="l")
```

The `type="l"` (that's the letter "el") option to `plot` tells S-PLUS to create a line plot. See what happens if you omit this.

To plot the density curve for the F-distribution with 4 and 10 degrees of freedom,

```
> x <- seq(0,4,length=100)
> y <- df(x, 4, 10)
> plot(x,y,type="l")
```

Discrete

Distribution	root
binomial	`binom`
geometric	`geom`
hypergeometric	`hyper`
negative binomial	`nbinom`
Poisson	`pois`

Preface each of the above roots with either `d`, `p`, `q` or `r`.

`d`*root* returns a probability, `p`*root* returns a cumulative probability, `q`*root* returns a quantile, and `r`*root* returns a random number.

The quantile function is the inverse of the CDF, $F(t) = P(X \leq t) = \sum_{k \leq t} P(X = k)$.

Example Binomial

Suppose you have a biased coin that has a probability of 0.8 of coming up heads.

The probability of getting 5 heads in 16 tosses of this coin is

```
> dbinom(5,16,.8)
```

Check this answer by calculating directly $\binom{16}{5} .8^5 \cdot .2^{11}$,

```
> choose(16,5)*.8^5*.2^11
```

The probability of getting at most 5 heads in 16 tosses is

```
> pbinom(5,16,.8)
```

In other words, `pbinom(5, 16, .8)` is computing:

```
dbinom(0,16,.8)+dbinom(1,16,.8)
+dbinom(2,16,.8)+dbinom(3,16,.8)
+dbinom(4,16,.8)+dbinom(5,16,.8)
```

The .25 quantile is

```
> qbinom(.25,16,.8)
[1] 12
```

This is the smallest number of successes such that the probability of at most this many successes is greater than or equal to .25.

Check this:

```
> pbinom(11,16,.8)
> pbinom(12,16,.8)
```

Example (cont.) Geometric

Find the probability of getting the first head on the fourth toss. This is the geometric distribution. The arguments to `geom` are `geom(failures, p)`.

```
> dgeom(3,.8)
```

The probability that the first head occurs on one of the first four tosses (that is, on the first, second, third or fourth toss) is

```
> pgeom(3,.8)
```

Example Poisson

Suppose a certain region of California experiences about 5 earthquakes a year. Assume occurrences follow a Poisson distribution. What is the probability of 3 earthquakes in a given year?

Here $\lambda = 5$

```
> dpois(3,5)
```

Check the answer:

```
> 5^3*exp(-5)/(3*2)
```

Random numbers

To generate random numbers from a particular distribution, preface the root name with an `r`.

For example, we continue our previous example of a biased coin with $p = .8$ of coming up heads. Toss this coin 25 times. The command `rbinom(1,25,.8)` will return a random number of successes.

```
> rbinom(1,25,.8)
```

Now, lets run this experiment 10 times (that is, we do 10 sets of tossing a coin 25 times) and record the number of successes.

```
> set.seed(0)
```

This sets the seed for the random number generator so that we all get the same results.

```
> heads <- rbinom(10,25,.8)
> heads
[1] 17 19 21 18 20 18 22 18 22 17
```

In the first experiment of tossing the coin 25 times, 17 heads occurred. In the second experiment of tossing the coin 25 times, 19 heads occurred, etc.

```
> table(heads)
> barplot(table(heads))
```

Repeat the above, except now run the experiment 100 times.

```
> heads2 <- rbinom(100,25,.8)
```

Repeat above, running experiments 250, 500 and 1000 times.

What do you notice about the shape of the plots as the number of experiments increase?