R Functions for Probability Distributions

Young W. Lim

2018-03-22 Thr

Outline

Based on

"Probability with R: An Introduction with Computer Science Applications"
Jane Horgan

R Functions

	cdf	i cdf	pdf	rv
	probability	quantile	density	random
Beta	pbeta	qbeta	dbeta	rbeta
Binomial	pbinom	qbinom	dbinom	rbinom
Cauchy	pcauchy	qcauchy	dcauchy	rcauchy
Chi-Square	pchisq	qchisq	dchisq	rchisq
Exponential	pexp	qexp	dexp	rexp
Gamma	pgamma	qgamma	dgamma	rgamma
Geometric	pgeom	qgeom	dgeom	rgeom
Hypergeometric	phyper	qhyper	dhyper	rhyper
Log Normal	plnorm	qlnorm	dlnorm	rlnorm
Negative Binomial	pnbinom	qnbinom	dnbinom	rnbinom
Normal	pnorm	qnorm	dnorm	rnorm
Poisson	ppois	qpois	dpois	rpois
Uniform	punif	qunif	dunif	runif

http://www.stat.umn.edu/geyer/old/5101/rlook.html#dist

Finite Populations

```
sample(1:100)
        # a random permutation of the numbers 1 to 100

sample(x,10,replace=F)
        # a random sample of size N=10
        # drawn from the elements of x without replacement

sample(1:6,10,replace=T,prob=c(1/7,1/7,1/7,1/7,1/7,2/7))
        # a random sample of size 10 from the digits 1 to 6
        # with unequal probabilities of selection
        # eg. a loaded die
```

http://people.reed.edu/~jones/141/sim1.html

Bernoulli(p) Trials

- Bernoulli trials are sequences of independent dichotomous trials
- each with probability p of success.
- The sample space consists of the two possible outcomes.
- For example, the results of 10 successive (independent) tosses of a fair coin would constitute 10 Bernoulli(1/2) trials. In R:
- `rbinom(N,1,p)`
  - generates a sequence of N trials, each with probability p of success.

`http://people.reed.edu/~jones/141/sim1.html`

Binomial(n,p) Trials

- this arises from a sequence of n independent Bernoulli trials
- each with probability p of success.
- Since the number of successes in n trials can range from 0 to n
- the sample space is just the integers 0,1,2, ... n.
- the number of heads that occur in 20 independent tosses of a fair coin has a Binomial(20,.5) distribution.
- `rbinom(N,n,p)`
  - generates N independent samples from a binomial(n,p) distribution.

http://people.reed.edu/~jones/141/sim1.html

Geometric(p) Trials (1)

- The geometric distribution models the number of failures
- until the first success in a sequence of independent Bernoulli trials,
- each with probability p of success.
- the geometric is sometimes defined to be the number of trials including the first success.
- R uses the number of failures.

http://people.reed.edu/~jones/141/sim1.html

Geometric(p) Trials (2)

- Since we might have to repeat the experiment arbitrarily many times before the first success,
- the sample space is the non-negative integers: 0,1,2,...
- if we roll a fair die, and count the number of rolls before the first "6"appears, we have a geometric distribution with $p = 1/6$.
- In other words, if the sequence of rolls yields $\{4,2,4,5,6,\dots\}$, then we had 4 rolls before the first 6.
- `rgeom(N,p)`
  - generates N geometric(p) counts.

http://people.reed.edu/~jones/141/sim1.html

Poisson(m) Trials (1)

- The Poisson distibution arises as an approximation to the binomial(n,p) for large n and small p,
- hence the common reference to modelling "rare events".
- the sample space for the Poisson distribution is the non-negative integers
- the number of phone calls arriving at a switchboard in a given time interval is likely to be approximately Poisson.
- The parameter m is the mean count.
- `rpois(N,m)`
  - generates N values from a poisson(m) distribution.

http://people.reed.edu/~jones/141/sim1.html

Poisson(m) Trials (2)

- To compare the poisson distribution to a similar binomial distribution, set m=n*p.
- You can either directly compare the probabilities for each possible outcome,
- or you could plot the cumulative distribution functions against each other.
- The sample space for a binomial(n,p) distribution is 0:n;
- let's try it for a binomial(20,.1) and the poisson(2) distributions:

```
k <- 0:20
bin <- dbinom(k,20,.1)
pois <- dpois(k,2)
plot(k,bin)
points(k,pois,pch="+")
```

http://people.reed.edu/~jones/141/sim1.html

Poisson(m) Trials (3)

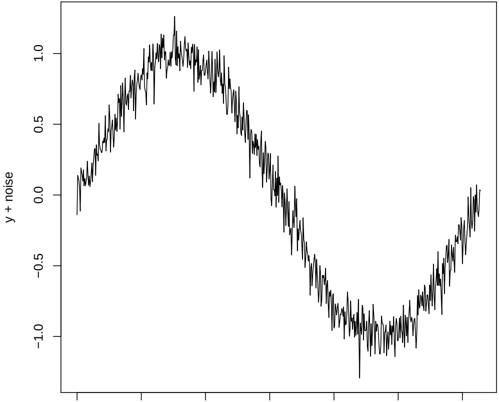
- Analogous to a quantile-quantile plot, we can plot the cumulative relative frequencies of the two distributions against each other:
- If the two distributions agreed exactly, the plotted points would lie on a straight line.
- The abline function adds a line with the given intercept and slope, in this case 0 and 1 respectively.

```
k <- 0:20
bin <- pbinom(k,20,.1)
pois <- ppois(k,2)
plot(bin,pois)
abline(0,1)
```

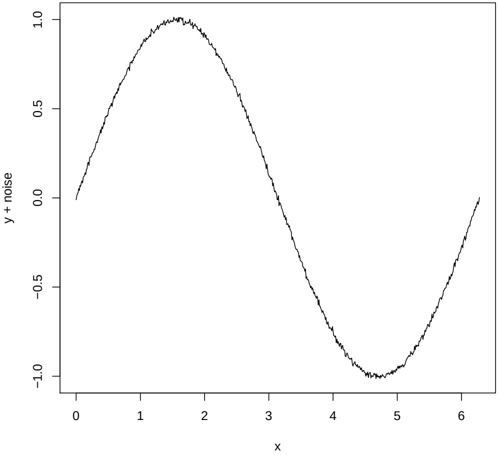
http://people.reed.edu/~jones/141/sim1.html

Normal distribution noise

```
x <- seq(0, 2*pi, 0.01)
y <- sin(x)
noise <- rnorm(length(x), 0.0, 0.1)
plot(x, y+noise, "l")
```



Normal distribution noise




```
x <- seq(0, 2*pi, 0.01)

y <- sin(x)

num <- 100
noise <- rep(0, length(x))
for (i in 1:num) {
  set.seed(i)
  noise <- noise +
  rnorm(length(x), 0.0, 0.1)
}
noise <- noise / num

plot(x, y+noise, "l")
```

```
x <- seq(0, 2*pi, 0.01)
y <- sin(x)
noise <- runif(length(x), -0.1, +0.1)
plot(x, y+noise, "l")
```