Basic Probability Distributions in R

M. George Akritas

- The Binomial p.m.f.: dbinom(x,n,p), $x = 0, 1, \ldots, n$; x can also be a vector.
  - dbinom(4, 10, 0.5) gives the $P(X = 4)$ for $X \sim Bin(10, 0.5)$, same as factorial(10)/(factorial(4)*factorial(10-4))*0.5**10.
  - dbinom(0:10, 10, 0.5) gives all probabilities $P(k)$, $k = 0, \ldots 10$.
  - sum(dbinom(46:54, 100, 0.5)) gives $P(46 \leq X \leq 54)$ for $X \sim Bin(100, 0.5)$
- The Binomial c.d.f..: pbinom(x,n,p), $x = 0, 1, \ldots, n$; x can also be a vector.
- Sampling: rbinom(m, n, p) gives a random sample of size m
  - set.seed(111); rbinom(50, 10, 0.4)
  - x=rbinom(10000,10,0.5); table(x)/10000. Compare with dbinom(0:10, 10, 0.5).

- Basic plot command works by specifying the points on the x- and y- axes: plot(0:10,dbinom(0:10, 10, 0.5)).
  - Add color/plotting character: plot(0:10,dbinom(0:10, 10, 0.5), pch=4,col=2)
  - Customize labels and connecting the dots: plot(0:10,dbinom(0:10, 10, 0.5), pch=1,col=2, xlab="Sample Space", ylab="Binomial Probabilities"); lines(0:10,dbinom(0:10, 10, 0.5), col=3)
- Bar graph: p=dbinom(0:10, 10, 0.5); barplot(p)
  - Customize the axes: barplot(p,xlim=c(0,12),ylim=c(0,0.25))
  - Add color: barplot(p,xlim=c(0,12),ylim=c(0,0.25),col="green")

- Histogram of empirical probabilities:
  - Try "hist(x,seq(-0.5,10.5,1))"
  - Add color "hist(x,seq(-0.5,10.5,1),col=5)"
- Bar graph for the empirical probabilities: barplot(table(x)/10000,col="3")

- The Negative Binomial
  - p.m.f.: dnbinom(x,r,p), $x = 0, 1, 2 \ldots$, so $x =$ the number of failures which occur before the r-th success; x can also be a vector.
  - c.d.f.: pnbinom(x,r,p)
  - sampling: rnbinom(m,r,p) gives a random sample of size m.
- The Hypergeometric
  - p.m.f.: dhyper(x, M, N-M, n)
  - c.d.f.: phyper(x, M, N-M, n)
  - sampling: rhyper(m, M, N-M, n) gives a random sample of size m.
- The Poisson
  - p.m.f.: dpois(x, lambda), $x = 0, 1, 2 \ldots$; x can also be a vector.
  - c.d.f.: ppois(x, lambda)
  - sampling: dpois(m, lambda) gives a random sample of size m.

- p.d.f.: dnorm(x,mu,sigma), $-\infty < x < \infty$; x can also be a vector.
  - default: dnorm(x) is the same as dnorm(x,0,1)
- c.d.f.: pnorm(x,mu,sigma)
  - default: pnorm(x) is the same as pnorm(x,0,1)
- quantiles: qnorm(p,mu,sigma), gives the $100p$th percentile, $0 < p < 1$. For example, qnorm(.9,mu,sigma) gives the 90th percentile for the specified values of mu and sigma.
  - default: qnorm(p) is the same as qnorm(p,0,1)
- sampling: rnorm(m,mu,sigma) gives a random sample of size m.
  - default: rnorm(m) is the same as rnorm(m,0,1)

1. Standard use of plot:
   plot(seq(-3,3,0.01), dnorm(seq(-3,3,0.01))) plots the pdf of the standard normal distribution from -3 to 3.
   - Try also: plot(seq(-3,3,0.01), dnorm(seq(-3,3,0.01)),type="l")
   - Superimpose two PDFs:
     plot(x,dnorm(x,0,0.5),type="l",col="blue");
     lines(x,dnorm(x),type="l", col="red")
2. Use of curve: curve(dnorm(x,0,0.3),from=-3, to=3)
   - Superimpose with curve: curve(dnorm(x,0,0.3),from=-3, to=3,col="blue"); curve(dnorm(x,0,1),from=-3, to=3, col="red", add=T)
   - The "from" and "to" can be omitted: curve(sin(2*x),-pi,pi)

Marking and shading

1. Plot the standard normal PDF and mark the 90th percentile: curve(dnorm,-3,3); lines(qnorm(0.9),dnorm(qnorm(0.9)), type="h", col="red")
2. Shade the area under the N(0,1) pdf to the right of the 90th percentile: x1=seq(qnorm(0.9),3,0.01); y1=dnorm(x1) curve(dnorm,-3,3); lines(x1,y1,type="h",col="red")
3. curve(dnorm,-3,3)
   polygon(c(rep(1,201),rev(seq(1,3,.01))),c(dnorm(seq(1,3,.01)), dnorm(rev(seq(1,3,.01)))),col="orange", lty=2, lwd=2, border="red")

1. Exponential(λ), with density $f(x) = \lambda e^{-\lambda x}$:
   - dexp(x,lambda), pexp(x,lambda), qexp(x,lambda), rexp(m,lambda).
2. Uniform from A to B:
   - dunif(x,A,B), punif(x,A,B), qunif(x,A,B), runif(x,A,B).
3. Even more:
   - Beta distribution with parameters alpha>0, beta>0. It generalizes the Uniform(0,1) (alpha=1, beta=1): dbeta(x, alpha ,beta) etc
   - Gamma distribution with parameters alpha>0, beta>0. No probability mass on negative values – generalizes the exponential distribution (alpha=1, lambda=1/beta), and the Erlang (alpha integer): dgamma(x, shape=alpha, scale = beta) etc.
   - Weibull distribution with parameters alpha>0, beta>0. No probability mass on negative values – generalizes the the exponential distribution (alpha=1, lambda=1/beta): dweibull(x, alpha, beta) etc