STA122 Lab Session # 2: Random Variables in R

Course Instructor: Prof. Merlise Clyde
Teaching Assistant: Debdeep Pati (dp55@stat.duke.edu)

January 20, 2009

1 Random variables

- Functions are provided to evaluate the cumulative distribution function $P(X \leq x)$, the probability density function and the quantile function (given q, the smallest x such that $P(X \leq x) > q$), and to simulate from the distribution.

```
Distribution   R name   additional arguments
beta           beta     shape1,shape2, ncp
binomial       binom    size, prob
Cauchy         cauchy   location, scale
chi-squared    chisq    df, ncp
exponential    exp      rate
F              f        df1, df2, ncp
gamma          gamma    shape, scale
geometric      geom     prob
hypergeometric hyper    m,n,k
log-normal     lnorm    meanlog, sdlog
logistic       logis    location, scale
neg. binomial  nbinom   size, prob
normal         norm     mean, sd
Poisson        pois     lambda
Students t    t        df, ncp
uniform        unif     min, max
Weibull        weibull  shape, scale
Wilcoxon       wilcox   m, n
```

- Prefix the name given here by d for the density, p for the CDF, q for the quantile function and `r`for simulation (random deviates).

- **Some Examples:**

```
> ##100 N(0,1)random variables
> x <- rnorm(100,0,1)
> ## 2-tailed p-value for t distribution
> 2*pt(-2.43, df = 13)
> ## upper 1% point for an F(2, 7) distribution
> qf(0.01, 2, 7, lower.tail = FALSE)
```

- To draw the histogram and add the density function the command is

  ```
  > x <- rnorm(100)
  > hist(x,freq=F)
  > curve(dnorm(x),add=T)
  ```

  The freq=F argument to hist ensures that the histogram is in terms of densities rather than absolute counts. The curve function graphs an expression (in terms of x) and its add=T allows it to overplot an existing plot.

- So things are generally set up correctly, but sometimes the top of the density function gets chopped off. The reason is of course that the height of the normal density played no role in the setting of the y-axis for the histogram. It will not help to reverse the order and draw the curve first and add the histogram, because then the highest bars might get clipped. The solution is first to get hold of the magnitude of the y values for both plot elements and make the plot big enough to hold both

  ```
  > h <- hist(x, plot=F)
  > ylim <- range(0, h$density, dnorm(0))
  > hist(x, freq=F, ylim=ylim)
  > curve(dnorm(x), add=T)
  ```

  When called with plot=F, hist will not plot anything, but it will return a structure containing the bar heights on the density scale. This and the fact that the maximum of dnorm(x) is dnorm(0) allows us to calculate a range covering both the bars and the normal density. The zero in the range call ensures that the bottom of the bars will be in range too. The range of y values is then passed to the hist function via the ylim argument.

2 Prior, posterior, HPD credible set

```
# comments
> x = c(seq(39.8 - 4*2, 39.8, length=100), seq(39.8, 39.8 +4*2, length=100))
#generate a sequence to plot the density #note
that the mode is explicitly included twice in the sequence


# plot the prior distribution

plot(x, dnorm(x, mean=38, sd=sqrt(9)),type="l", main="Prior
Density", ylab="density")
 # dnorm gives the normal density # type
=l connects points by lines

# plot the standardized likelihood
 plot(x,dnorm(x, mean=39.8,sd=sqrt(4/10000)), type="l", ylab="",main="Likelihood")
 lines(x,dnorm(x, mean=38, sd=sqrt(9)), lty=2, col=2)
 # add the prior density
lines(x, dnorm(x, mean=39.8, sd=sqrt(1/2500.111)), col=3) # overlay
```

```
the posterior density x = seq(39.7, 39.9, length=100)

#Plot the posterior

plot(x, dnorm(x,mean=39.8, sd=sqrt(1/2500.111)), type="l",
ylab="density",main="Posterior Density", xlab=expression(mu))

# Calculate an HPD Region
hl = qnorm(.025, 39.8, sqrt(1/2500.111))
hu = qnorm(.975, 39.8, sqrt(1/2500.111))
#> hu #[1] 39.8392 #> hl#[1] 39.7608

hd = dnorm(hu, 39.8, sqrt(1/2500.111))
 hd = dnorm(hl, 39.8,
sqrt(1/2500.111))
 #add a horzontal line at the height of highest
densities
abline(h= hd)
# add verticle line segments at the points
of highest densities
segments(c(hl, hu), c(0, 0), c(hl, hu), c(hd,
hd))
```

3 Another Bayesian Analysis showing washing away of prior as sample size increases

```
ybar  <- 6
 mu0   <- 2
 tau0  <- 1
 sigma <- 1
  theta <- seq(-2, 8,0.05)
  mu1 <-4
  mu10 <-5.6363
  tau1 <-0.70710
  tau10 <-0.301513
plot(theta,dnorm(theta,mu0,tau0),type='l',ylim=c(0,1.4))
lines(theta,dnorm(theta,mu1,tau1),lty=2)
lines(theta,dnorm(theta,mu2,tau10),lty=3)

title(main=paste("Prior and Posterior dfs for a Normal Mean, ybar=6,
mu0=2, tau0=1"),cex=.8)

 legend(-2,1.38,legend= c("Prior","Posterior
with n = 1", "Posterior with n = 10"),lty=c(1,2,3))
```