22S:138
Lab session 2
More R/Splus functions for simple Bayesian analysis

Sept. 10, 2004

1 Built-in R/Splus functions for distributions

R and Splus have built-in functions for extracting characteristics of the following probability distributions:

```
Distribution  Name of Splus function for density (or pmf) evaluation
------------  ------------------------------------------------------
beta          dbeta
Cauchy        dcauchy
chisquare     dchisq
exponential   dexp
F             df
gamma         dgamma
logistic      dlogis
lognormal     dlnorm
normal        dnorm
t             dt
uniform       dunif
weibull       dweibull

poisson       dpois
binomial      dbinom
```

For each of the distributions 4 functions are available, which differ only by their first letter:

```
d  -- density or probability mass function (example: dbeta, dbinom)
p  -- cumulative density or cumulative probability mass function (example: pbeta, pbinom)
q --  quantiles (example:  qbeta, qbinom)
r --  generate a random sample (example: rbeta, rbinom)
```

You can get detailed documentation on all four of the functions for any of the distributions by entering "help(name of one of the functions)". For example

```
help(dgamma)
```

will produce documentation on all 4 of the functions for the gamma distribution.

To get used to using the documentation and using the functions, use the functions to get answers to the following questions. You will need to be able to do these kinds of things to solve real problems on an upcoming homework assignment.

1. What are the .25 and .75 quantiles of an exponential distribution with parameter lambda = 6?

2. Generate a random sample of 25 values from a normal distribution with mean 3 and standard deviation 4.

3. Find the probability that a random variable from a binomial distribution with $n = 12$ and $p = .4$ takes on a value less than or equal to 4.

Whenever you are using software to do things with probability distributions, it is important to know what parameterization the software package uses for the distribution of interest. For example, there are two different parameterizations of the gamma distribution in common use. Let's find out whether R or Splus use the same parameterization as that given in the Gill table of distributions.

1. Note the formulas for the mean and variance of a gamma random variable given in Gill table of distributions.

2. Now we will generate a random sample of 1000 values from a gamma(2, 10) distribution and will calculate their mean and variance.

```
gammasamp <- rgamma(1000, 2, 10)   # generate the random sample and
                                   # assign it to a vector named gammasamp
mean(gammasamp)                    # display the mean
var(gammasamp)                     # display the variance
rm(gammasamp)                      # remove the vector since it is no
                                   # longer needed
```

3. Does it appear that R uses the same parameterization? If you're not sure, repeat the exercise with a larger random sample.

2 A function to compute posterior predictive probabilities based on a Beta posterior distribution

This is not a built-in function. I wrote it and you must download it and "source" it to get access to it. Use Internet Explorer or Netscape to download the function `pbetap` into the directory `c:`
`temp`. Then in R

```
s138.pbetap <- source("c:\\temp\\pbetap")$value
```

You must provide the arguments to this function in parentheses when you call the function. The arguments are:

- the alpha parameter of the beta posterior distribution

- the beta parameter of the beta posterior distribution
- the sample size in the future sample for which you wish to compute posterior probabilities
- a vector of the numbers of successes for which you wish to compute posterior probabilities

For example, suppose I was trying to estimate a population proportion p. I had used a uniform prior and I observed 6 successes and 4 failures. Thus, my posterior distribution is:

$$p(p|y) = Beta(7, 5)$$

I plan to conduct a new study with sample size $n^* = 10$. If y^* denotes the number of successes in the hypothetical future study, I want to compute the posterior predictive probabilities of three possible values of y^* — 6, 7, and 8. The following call will do this:

```
s138.pbetap( 7, 5, 10, 6:8)
```

Note that the last argument ("6:8"), is Splus's notation for a sequence. It means the integers 6 through 8 inclusive.

1. Suppose you had carried out our survey of UI students. You had used a uniform prior and had gotten 7 yeses and 43 nos in your survey.
   (a) What is your posterior distribution for the population proportion p?
   (b) You are considering carrying out a follow-up survey of 25 more students. Use the function to compute the posterior predictive probabilities of each of the possible numbers of yeses in the follow-up study.
   (c) What is the most likely number of yeses in the new study?