Session 2 Exercises with Solutions

Estimating a Proportion with a Discrete Prior: Consider a population of "successes" and "failures" where the proportion of successes is p. Suppose p takes on the discrete set of values 0, .01, ..., .99, 1 and one assigns a uniform prior on these 101 values, meaning that the probability of any one of the 101 possible outcomes is equally likely. Use R to enter these values into a vector **`p`** and the associated probabilities into a vector **`prior`**, respectively. Plot the resulting **`prior`** distribution with the values of **`p`** on the horizontal axis and the values of **`prior`** on the vertical axis (**hint:** use the **plot()** function with argument **`type="h")`.** What family or type of random variable does this prior distribution look like?

```
# Discrete Prior
p <- seq(0, 1, by = 0.01)
p
prior <- 1 / 101 + 0 * p
prior
plot(p, prior,
     type="h",
     main="Prior Distribution")
# The prior distribution looks uniform
```

Posterior Distribution: Suppose one takes a random sample from the population without replacement and observes 20 successes and 12 failures. The function **`pdisc()`** in the **`LearnBayes`** package computes the associated posterior probabilities for p. After loading the **`LearnBayes`** package, call up help for the **`pdisc()`** function with the R command **`?pdisc`**. The inputs to **`pdisc()`** are the prior (vector of values of **`p`** and vector of **`prior`** probabilities) and a vector containing the number of successes and failures. Create a variable named **`post`** which captures the posterior distribution which is the output from **`pdisc()`**. Plot the resulting **`post`** distribution with the values of **`p`** on the horizontal axis and the values of the posterior distribution (**`post`**) on the vertical axis (**hint:** again use the **plot()** function with argument **`type="h")`.** What family of random variable does the posterior distribution look like?

```
# Posterior Distribution
library(LearnBayes)
post <- pdisc(p, prior, c(20, 12))
post
plot(p, post,
     type="h",
     main="Posterior Distribution")
# The posterior distribution looks normal
```

<u>Highest Probability</u>: A highest probability interval for a discrete distribution is obtained using the **`discint()`** function. Call up the help screen for the **`discint()`** function using the R command **`?discint`**. This function has two inputs: the probability distribution matrix where the first column contains the values and the second column contains the probabilities, and the desired probability content. Use **`discint()`** to separately compute 90, 95 and 99 percent probability intervals for p from the posterior distribution. What are the upper and lower bounds of the interval that contains **`p`** with a 90% probability? With a 95% probability? With a 99% probability interval?

```
# Highest Probabilities
discint(cbind(p, post), 0.90)
# The probability that p falls in the interval
# {0.49, 0.75} is approximately 0.90.
discint(cbind(p, post), 0.95)
# The probability that p falls in the interval
# {0.46, 0.77} is approximately 0.95.
discint(cbind(p, post), 0.99)
# The probability that p falls in the interval
# {0.40, 0.81} is approximately 0.99.
```