Session 2 Exercises

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 (remember, elements of the **p** vector include both 0 and 1) is equally likely. Use R to enter the values of **p** and the associated probabilities into the vectors **p** and **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 **plot()** function with argument **type = “h”).** What family of random variable does this prior distribution look like?

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 output by **pdisc()**. Plot the resulting **post** distribution with the values of **p** on the horizontal axis and the values of the posterior distribution on the vertical axis (**hint:** again use **plot()** function with argument **type = “h”).** What family of random variable does the posterior distribution look like?

Highest Probability

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?