Statistics 110 Autumn 2006–2007
B. Srinivasan Out: Sept. 29, 2006
Handout #7 **Due: Oct. 6, 2006**

Homework #2

Reading: Rice 2.1-4 (skip 2.2.2 and 2.2.4), 4.1-2 (skip 4.2.1),10.1-2

Submit in the inbox outside Sequioa 229 before 5:00 PM.

This second problem set should give you experience with basic `R` programming, sets and counting, and introductory probability. Note: we will do problem 3 from HW1 (concerning discrete random variables) next week.

1. *He had more than a histogram of code on him, officer...*

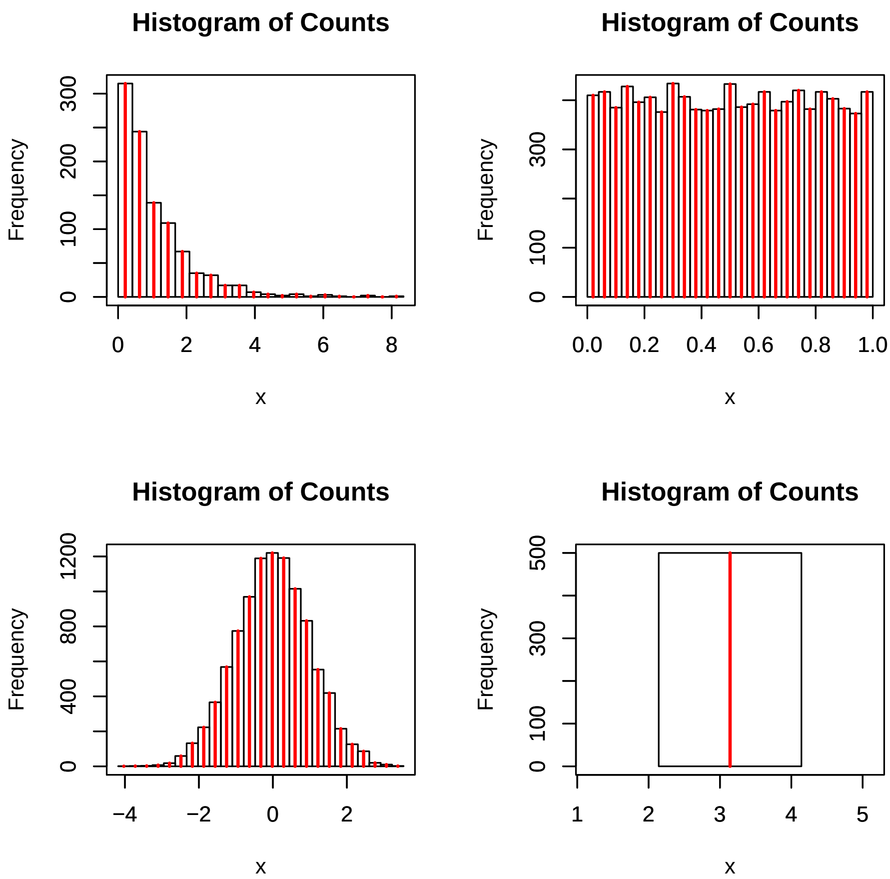
   The simple histogram code in `week1_lecture2.R` does not generalize well to continuous data because it relies upon tabulation (and hence implicitly upon tests for equality). So for this problem, you will write a short `R` script to build your own histogram.

   Begin by making an `R` function titled `myhist` in an `R` file of the same name. Your function should accept two arguments, `x` (the vector of numerical data) and `N` (the number of histogram bins). Use the command `range` to get the min and max values of `x` and divide this range into `N` bins. Calculate the number of values of `x` that land up in each bin. Try not to use a `for` loop, because such loops are slow in `R` (as they are in all interpreted languages). You may find the following functions to be helpful: `match,seq,sort` and `diff`. Once you have calculated the number of counts which land up in each bin, obtain the midpoints of each bin. Plot the bin midpoints vs. the number of counts per bin with `plot` using the `type=h` invocation, also specifying `col="red"` and `lwd=2`. Finally, overplot the histogram generated with `hist` using `par(new=TRUE)`. You will need to explicitly specify `xlim` and `ylim` in both the invocation of `plot` and `hist` to get things to line up properly.

   Your code should check that there are enough elements of `x` to build a histogram and more than `N=2` bins. It should also deal intelligently with the special case where all the elements of `x` are constant. To put your function through its paces, use `source` to load your function into the workspace and check that your output looks like the following for the given continuous input. (See the web pdf for color output). **Include all `R` code and plots with your HW. Color printing would be nice but is not mandatory.**

```
> rm(list = ls())
> source("myhist.R")
> layout(matrix(1:4, 2, 2))
> x <- rexp(1000, rate = 1)
> N <- 20
> myhist(x, N)
> x <- rnorm(10000)
> N <- 25
> myhist(x, N)
```

```
> x <- runif(10000)
> n <- 3
> myhist(x, N)
> x <- rep(3.14159, 500)
> myhist(x, N)
```



.

2. *The Gambler and the Omega (Adapted from Rice 1.8.2)*

   A gambler throws two six sided dice sequentially and records the face values in order.

   (a) List the sample space, Ω.

   (b) Let X_i = the face value of the i^{th} die ($i = 1, 2$). List the elements that make up the following events.

      i. $A = \{(X_1, X_2)|\ X_1 + X_2 \geq 5\}$

      ii. $B = \{(X_1, X_2)|\ X_1 > X_2\}$

      iii. $C = \{(X_1, X_2)|\ X_1 = 4\}$

      Can you assign a geometric interpretation to these sets?

   (c) Now list the elements of the following derived events

i. $A \cap C$
ii. $B \cup C$
iii. $A \cap (B \cup C)$

3. *Set Tripping*

(a) Let $\Omega = \{1, 2, 3, ..., 15\}$, $A = \{1, 3, 5, 7, 9\}$ and $B = \{2, 5, 7, 11, 13\}$. Calculate the following quantities:

i. A^c
ii. $A \cap \Omega$
iii. $A \cup \Omega$
iv. $A \cap \emptyset$
v. $A \cup \emptyset$
vi. $A \cap B$
vii. $A^c \cap B$
viii. $A \cup B$
ix. $B - A$ (Recall that $B - A = B \cap A^c$)
x. $(B - A) \cup (A - B)$ (this is known as the symmetric set difference, $A \bowtie B$)

(b) For the same sets as in the previous question, which of the following set comparisons are true?

i. $A \subset B$
ii. $A \subset \Omega$
iii. $(B - A) \subset B$
iv. $(A \cap B) \cup (A \cap B^c) \subseteq A$

(c) Prove that

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$$

via first order logic. Specifically, begin by noting that

$$(A \cup B) \cap C = \{x | (x \in A \text{ or } x \in B) \text{ and } x \in C\}$$

Define the logical variable $p = 1$ if $x \in A$, $p = 0$ if $x \notin A$. Define similar logical variables q, r for the other two sets. By means of an exhaustive truth table for (p, q, r) and related logical variables like $p \wedge q$ and $q \vee r$, prove the equality. The point of this problem is to show how one would quantify your intuition from Venn Diagrams. By enumerating all the possible situations regarding set membership and checking each one, we can prove the equality.

(d) Suppose P is a probability measure. Prove that

$$P(A \cap B) \geq P(A) + P(B) - 1$$

(e) If P is a probability measure and A, B are sets, prove that

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

from the axioms of probability. Hint: consider manipulating $(B \cap A) \cup (B \cap A^c)$. Does this expression hold if A or B are the null set? Apply this formula iteratively to prove:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

(f) An alternative, equivalent expression which may be easier to remember is $P(A \cup B) = P(A) + P(B - A)$. Prove that $P(B) - P(A \cap B) = P(B - A)$ using the axioms of probability.

4. *Counting your Blessings*

   (a) Let $S = \{A, B, C, D, E\}$. How many unique ways are there to select 3 elements from S...

      i. with replacement, where order matters?
      ii. with replacement, where order does not matter?
      iii. without replacement, where order matters?
      iv. without replacement, where order does not matter?

   Note that "with replacement" indicates that we put a letter back after we select it, so that it can possibly be selected again. If "order matters", then a sequence such as ADC is considered different from ACD and hence these would represent two unique elements rather than one in your enumeration. You may find it helpful to explicitly enumerate the selections to check your combinatoric calculations.

   (b) Now let's specify a set function P, such that $P(x) = 1/5$ for $\{x \in S\}$. Is this a valid probability measure? If so, what is the probability of selecting ABC under each selection paradigm? Please calculate Ω for each selection paradigm (you can use a compact notation rather than writing out each element).

   (c) Repeat the previous part with P which takes the values $P(A) = 1/3, P(B) = 1/3, P(C) = 1/3, P(D) = 0, P(E) = 0$.

5. *Basic Probability*

   (a) (Rice 1.12) In a game of poker, five players are each dealt 5 cards from a 52-card deck. How many ways are there to deal the cards?

   (b) (Rice 1.18) A lot of n items contains k defectives, and m are selected randomly and inspected. How should the value of m be chosen so that the probability that at least one defective item turns up is .90? Calculate your answer explicitly for:

      i. $n = 1000, k = 10$
      ii. $n = 10000, k = 100$

   (c) (Rice 1.22) A standard deck of 52 cards is shuffled thoroughly and n cards are turned up. What is the probability that a face card turns up? For what value of n is the probability about .5?

   (d) (Rice 1.24) If n balls are distributed randomly into k urns, what is the probability that the last urn contains j balls?

(e) (Rice 1.32) A wine taster claims that she can distinguish four vintages of a particular Cabernet. What is the probability that she can do this by merely guessing? (She is confronted with four unlabeled glasses).

(f) (Rice 1.39) A monkey at a typewriter types each of the 26 letters of the alphabet exactly once, the order being random.

   i. What is the probability that the word *Hamlet* appears somewhere in the string of letters?

   ii. How many independent monkey typists would you need in order that the probability that the word appears is at least .90?