PROBLEM CORNER

Md S. Warasi
Department of Mathematics and Statistics
Radford University, VA 24142
email: msarker@radford.edu

Problem 1

Consider the experiment of rolling a six-sided fair die. The aim of this problem is to illustrate the law of large numbers in identifying the true mean, μ, of the distribution when a die is rolled once. To accomplish this, do the following.

(a) Roll the die 5 times and calculate the sample mean of the observations. For example, the sample mean for the observations $\{5, 3, 4, 6, 1\}$ is $\overline{x} = 3.8$. Repeat this with 10, 30, 50, 100, and 200 trials. Plot the sample mean $\overline{x}$ (vertical axis) against the number of trials (horizontal axis). What does $\overline{x}$ converge to? By the law of large numbers, the sample mean should gradually approach the true mean as the number of trials increases.

(b) Repeat the entire process in part-(a) using a software with 1, 2, 3, ..., 1000 trials. This should provide a better illustration of the law of large numbers. Find an approximate value of the true mean.

(c) Calculate the exact value of the true mean. Use an intuitive approach or use the knowledge taught in elementary statistics courses. Provide a rationale for your answer.

Solution: The statistical software R has been used to simulate the experiment of rolling a die. The plots for parts (a) and (b) are shown in Figures 1 and 2, respectively. It is evident especially from Figure 2 which involves larger numbers of trials that the sample mean $\overline{x}$ converges to a value that is very close to 3.5.

The population mean, μ, is the center (balancing point) of a distribution. Because the distribution of the observations for rolling a fair die is symmetric (Figure 3), the mean must be the middle point of the possible observed data. Because the possible values are $\{1, 2, ..., 6\}$, the true mean is the average of the minimum and maximum; i.e., $\mu = \frac{1+6}{2} = 3.5$.

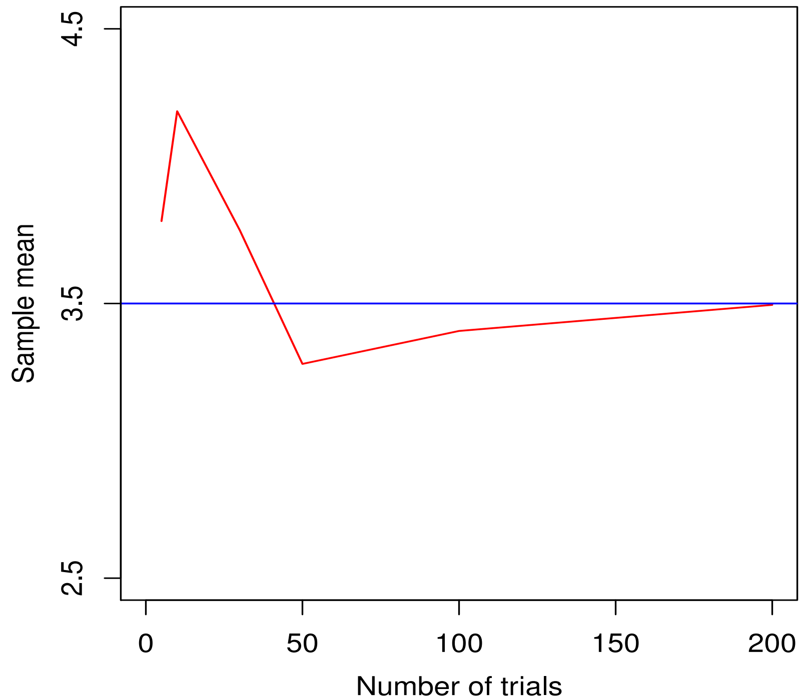

Figure 1: Convergence of the sample mean with the number of trials 10, 30, 50, 100, and 200.

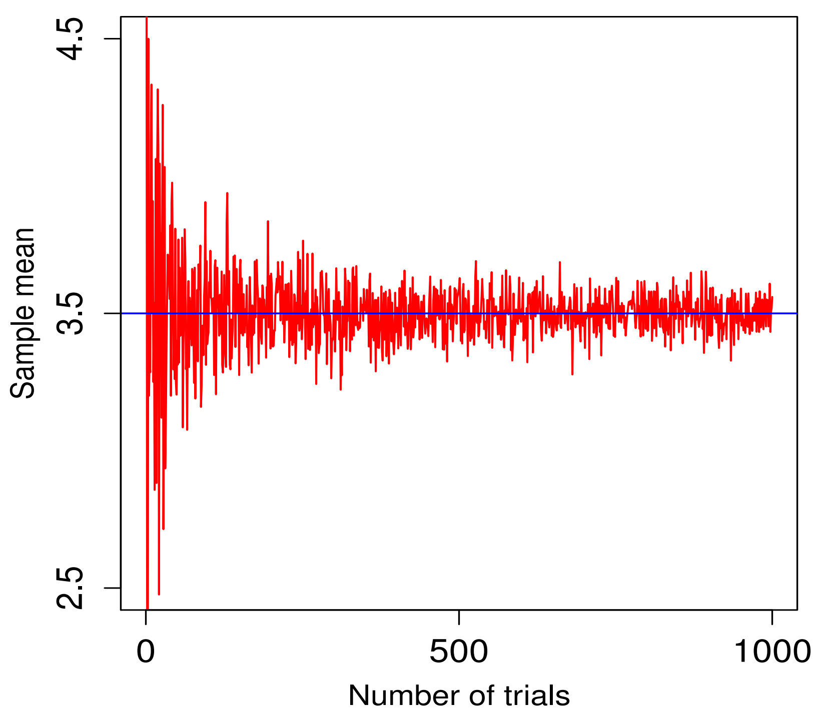

Figure 2: Convergence of the sample mean with the number of trials 1, 2, ..., 1000.

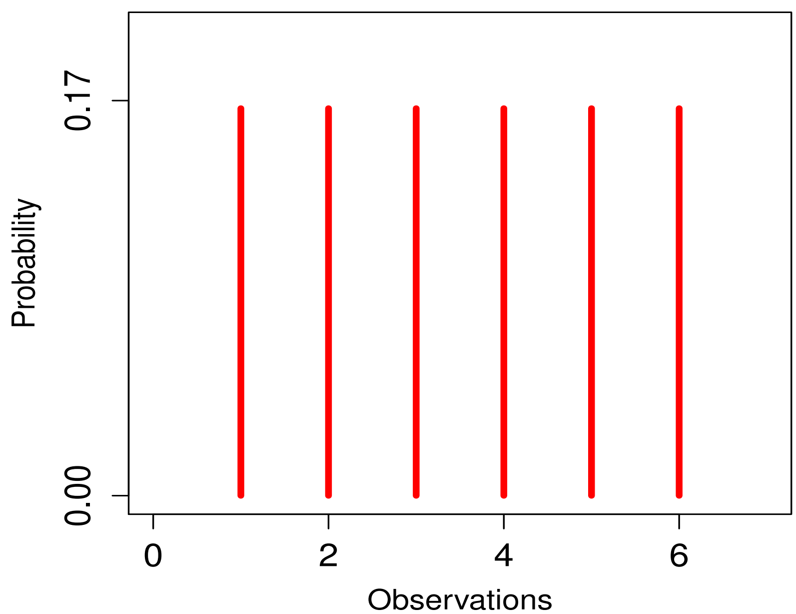

Figure 3: Probability histogram for the observations of a six-sided fair die.

Problem 2

Consider the experiment of rolling an N-sided fair die, where the number of sides N is unknown. When the die is rolled, the minimum possible value is 1 and the maximum possible value is N. Suppose, one observes the following data when the die is rolled 10 times. Find a reasonable estimate for N.

Observed data									
10	14	4	3	17	15	6	19	2	9

A widely used approach to solve this type of problem is the maximum likelihood estimation that involves forming the likelihood function $L(N)$ and maximizing it with respect to N. To do so, one easily identifies that the trials are independent and each of the 10 observations has the identical probability of $1/N$ to be observed. Thus, the probability that the observations are obtained jointly is

$$\begin{aligned} L(N) &= \frac{1}{N} \times \frac{1}{N} \times \ldots \times \frac{1}{N} \\ &= \left(\frac{1}{N}\right)^{10}. \end{aligned}$$

The N that maximizes $L(N)$ is called the maximum likelihood estimate for N. Therefore, finding a reasonable estimate for N involves completing the following steps: (i) find the N that maximizes $L(N)$, which is the estimate in general, (ii) apply this to the given observed data for calculating the estimate for N.

Solution: Let $X_1, X_2, ..., X_{10}$ be a sequence of random variables for the observations when the N-sided die is rolled 10 times. It is straightforward to observe that $L(N)$ is decreasing in N (Figure 4); i.e., the maximum value of $L(N)$ occurs at the minimum possible value of N. Note that each of the observations $X_1, X_2, ..., X_{10}$ must be between 1 and N, inclusive. This suggests that N cannot be smaller than any of $X_1, X_2, ..., X_{10}$. The condition would be satisfied when the minimum possible value of N is the maximum of $X_1, X_2, ..., X_{10}$. Hence, the maximum likelihood estimate of N, denoted by $\widehat{N}$, is

$$\widehat{N} = \max\{X_1, X_2, ..., X_{10}\}.$$

Because the maximum value of the observed data $\{10, 14, 4, 3, 17, 15, 6, 19, 2, 9\}$ is 19, the maximum likelihood estimate for N is $\widehat{N} = 19$.

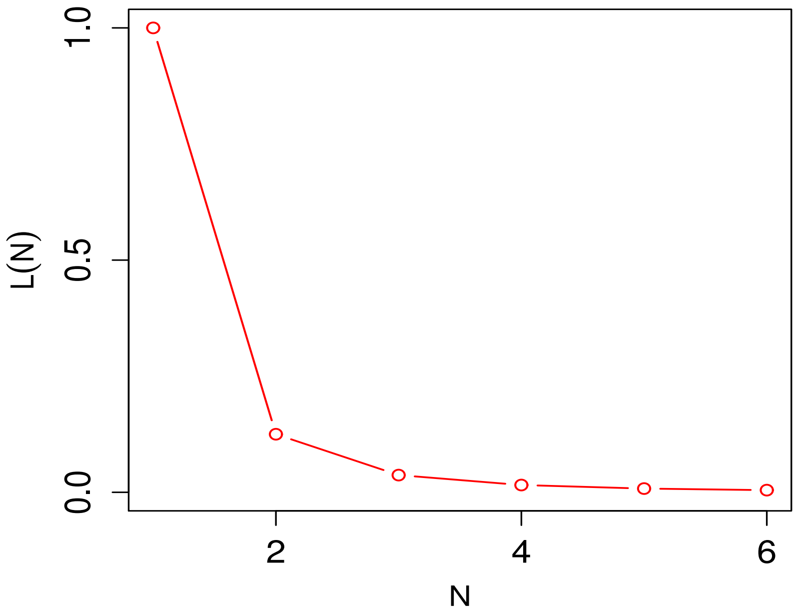

Figure 4: The likelihood function $L(N)$ when an N-sided die is rolled 3 times.

R codes:

```
#**************************************************#
#            R codes for Problem 1                 #
#**************************************************#

### 1(a)
n <- c(5, 10, 30, 50, 100, 200)
set.seed(123456)   # use the seed value to reproduce the results
xbar <- rep(-9, length(n))
for(i in 1:length(n)){
   xbar[i] <- mean( sample(1:6, n[i], replace=TRUE) )
}
par(mar=c(4.0,4.3,2,1.5))
plot(n, xbar, xlim=c(0, 200), ylim=c(2.5,4.5), xaxt="n", yaxt="n",
       type="n", main=" ", xlab=" ", ylab=" ")
lines(n, xbar, col="red", lty="solid", lwd=1.2)
abline(a=3.5, b=0, col="blue", lwd=1.2)
axis(1, at=c(0, 50, 100, 150, 200), cex.axis=1.1)
axis(2, at=c(2.5, 3.5, 4.5), cex.axis=1.1)
mtext("Number of trials", side=1, cex=1.1, line=2.5)
mtext("Sample mean", side=2, cex=1.1, line=2.5)
```

```
### 1(b)
n <- 1:1000
xbar <- rep(-9, length(n))
for(i in 1:length(n)){
   xbar[i] <- mean( sample(1:6, n[i], replace=TRUE) )
}
par(mar=c(4.0,4.3,2,1.5))
plot(n, xbar, xlim=c(0, 1000), ylim=c(2.5,4.5), xaxt="n", yaxt="n",
        type="n", main=" ", xlab=" ", ylab=" ")
lines(n, xbar, col="red", lty="solid", lwd=1.2)
abline(a=3.5, b=0, col="blue", lwd=1.2)
axis(1, at=c(0, 500, 1000), cex.axis=1.3)
axis(2, at=c(2.5, 3.5, 4.5), cex.axis=1.3)
mtext("Number of trials", side=1, cex=1.2, line=2.5)
mtext("Sample mean", side=2, cex=1.2, line=2.5)

### 1(c)
prob <- rep(1/6, 6)
plot(1:6, prob, xlim=c(0, 7), ylim=c(0, 1/5), xaxt="n", yaxt="n",
       type="n", main=" ", xlab=" ", ylab=" ")
lines(1:6, prob, col="red", type="h", lwd=4)
axis(1, at=c(0, 2, 4, 6), cex.axis=1.3)
axis(2, at=c(0, .17), cex.axis=1.3)
mtext("Observations", side=1, cex=1.2, line=2.5)
mtext(expression("Probability"), side=2, cex=1.2, line=2.5)

#************************************************#
#            R codes for Problem 2               #
#************************************************#

N <- 1:6
L <- (1/N)^3
plot(N, L, xlim=c(1, 6), ylim=c(0, 1), xaxt="n", yaxt="n", type="n",
         main=" ", xlab=" ", ylab=" ")
lines(N, L, col="red", type="b", lwd=1.2)
axis(1, at=c(2, 4, 6), cex.axis=1.3)
axis(2, at=c(0, .5, 1), cex.axis=1.3)
mtext(expression(N), side=1, cex=1.2, line=2.5)
mtext(expression(L(N)), side=2, cex=1.2, line=2.5)
```

Note: R is a free software widely used in research and applications. R can be downloaded from `https://www.r-project.org`.