Math 472: Worksheet 1

February 20, 2026

Instructions: *Please work on the problems in groups of 2 or 3. You don't need to finish this today, but I will ask each group to submit solutions on February 27.*

Problem 1. This problem aims to help visualize an answer to the following question: *"Suppose that our population of interest does not have a normal distribution. What does the sampling distribution of $\overline{Y}$ look like, and what is the effect of the sample size on the sampling distribution of $\overline{Y}$?"*

For this problem we will assume the population is a chi-squared distribution with 3 degrees of freedom. You can visualize the shape of this distribution using the `R` code

```
# Draw a random sample of size n=1000000 from a chi-squared distribution with 3 degrees
# of freedom
y <- rchisq(1000000, df=3)
hist(y, probability = TRUE, breaks = 100)
```

Use `R` to complete the following.

(a) By drawing samples from the population and computing the sample mean and sample variance, estimate the mean μ, variance σ^2, and standard deviation σ of the population. (When n is large, your answers will be close to the values from Proposition 60 in the lecture notes).

(b) We can visualize the sampling distribution of $\overline{Y} = (Y_1 + \ldots + Y_n)/n$ by generating a large number (say, $k = 100,000$) of sample means for a specified value of n, and then plotting this histogram (use the setting `probability=TRUE` and specify a reasonable number of breaks).

Do this for $n = 4, 16, 32$ and 64.

To be extra spicy, you can overlay the normal curve approximation given by the central limit theorem by running the following code after you plot your histogram:

```
curve(dnorm(x, mean=3, sd=sqrt(6/n)), add=TRUE)
```

(c) What do you notice about the adequacy of the normal approximation of the sampling distribution of $\overline{Y}$ as the sample size n increases?

(d) What happens to the standard error of $\overline{Y}$ as n goes from 4 to 16, from 16 to 64, and from 64 to 256? Formulate a general conjecture which relates the sample size to the standard error of $\overline{Y}$? Can you prove your conjecture?

Problem 2 (1-dimensional Brownian motion)**.** In this problem, we will model the price of a stock over the course of a day. Let $x(t)$ be the price of the stock at time t, where t is measured in seconds. Assume that at the beginning of the day ($t = 0$), the stock trades at \$5, and that trading continues for 9 hours.

We will model the volatility of the stock price in the following way. Assume that each second, the change in the stock prices is normally distributed with mean $\mu = 0$ and standard deviation $\sigma = 1/180$. In other words, for every $n = 0, 1, 2, \ldots$, we have

$$x(n+1) = x(n) + V_n \tag{1}$$

where V_n is an independent normal random variable with mean $\mu = 0$ and standard deviation $\sigma = 1/294$. We can think of $V_1, V_2, \ldots$ as representing the random *volatility* of the stock price from moment to moment.

(a) Plot the stock price over the course of the day. (You can use the command `plot()` to do this, and I recommend using the optional argument `type="l"` to tell `R` to draw connected lines.)

(b) Let Y be the change in the price of the stock (i.e., the final price minus the initial price). Generate a sample of $n = 5000$ values of Y, which you should save as the vector `y_values`.

(Hint: one way to do this is to first write a function which returns Y, and then use the `replicate()` command.)

(c) Compute the sample mean and sample standard deviation of `y_values`.

(d) Plot a histogram of your sample from (b) using

```
hist(y_values, probability=TRUE, breaks=100)
```

After doing this, overlay a normal curve on top of your histogram with the command

```
curve(dnorm(x,mean=??,sd=??), add=true)
```

(where you have to substitute sensible values for the mean and standard deviation).

(e) Prove that Y has a standard normal distribution.

(Hint: use Eq. (1)).

(f) Suppose you purchase the stock at the beginning of the day and sell at the end. What is the probability that you make at least \$1?

Problem 3 (2-dimensional Brownian motion)**.** Suppose we track the motion of a pollen grain suspended in a petri dish under microscope. The pollen grain is constantly being bombarded by water molecules from every direction, with each collision transferring a small amount of momentum to the pollen grain. This causes the pollen grain to slowly move around in an random, irregular way.

Let $(x(t), y(t))$ be the position of the pollen grain after $t \in \mathbb{N}$ units of time. Assume that at time $t = 0$, the pollen grain starts at position $(x(0), y(0)) = (0, 0)$, and that at each subsequent time step, its position is given by

$$(x(t+1), y(t+1)) = (x(t) + X_t, y(t) + Y_t),$$

where $X_1, X_2, \ldots,$ and $Y_1, Y_2 \ldots$ are independent normal random variables with mean 0 and variance $\sigma^2 > 0$.

(a) Assume that $\sigma = 1/10$. Plot the motion of the pollen grain from time $t = 0$ to $t = 100$.

Hint: One way to do this is generate a vector of 100 x-coordinates and a vector of 100 y-coordinates independently, and then plot them using the following command:

```
plot(x,y, pch=16, type="l")
```

(b) Let D be the distance of the pollen grain from the origin after 100 times steps (we are still assuming $\sigma = 1/10$). Generate a large number of independent samples of D. Use them to estimate the mean and variance of D, and plot a histogram of your D values.

(c) Now, assume that $\sigma = 1/100$. Plot the motion of the pollen grain from time $t = 0$ to $t = 10000$.

(d) With $\sigma = 1/100$, let D be the distance of the pollen grain from the origin after 1000. Generate a large number of independent samples of D. Use them to estimate the mean and variance of D, and plot a histogram of your D values.

(e) Compare the distributions of (b) and (d). Does the distribution obtained in change substantially?

(f) What is the distribution of D^2 in both cases?