EC 587 HW1

Erik Andersen

October 2023

Question 1

a)

We have two normal random variables $X_1 \sim N(\mu_1, \sigma_1^2)$, and $X_2 \sim N(\mu_2, \sigma_2^2)$ The moment generating function for a normal random variable is as follows:

$$M_X = exp(\mu t + \frac{1}{2}\sigma^2 t^2)$$

Let $X = X_1 + X_2$. Its moment generating function is as follows:

$$\begin{aligned} M_X &= exp(X_1 + X_2) \\ &= exp(X_1)exp(X_2) \\ &= M_{X_1} M_{X_2} \\ &= exp(\mu_1 t + \frac{1}{2}\sigma_1^2 t^2) exp(\mu_2 t + \frac{1}{2}\sigma_2^2 t^2) \\ &= exp((\mu_1 + \mu_2) + \frac{1}{2}(\sigma_1^2 \sigma_2^2) t^2) \end{aligned}$$

Since the sum has the moment generating function of a normal random variable, the sum itself is a normal random variable. So, $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2$

b)

Using the same logic as part a), we know that if $\frac{(X_i - \mu_i)}{\sqrt{\sigma_i^2}} \forall i \in \{1, 2\}$, the sum of $X = X_1 + X_2$ will be normal with parameters that are the sum of the parameters of X_1 and X_2.

$$\frac{X_1 - \mu_1}{\sqrt{\sigma_1^2}} + \frac{X_2 - \mu_2}{\sqrt{\sigma_2^2}} \sim N(0, 2)$$

c)

We have two random variables $X_1, X_2 \sim Bernoulli(p)$. Let $X = X_1 + X_2$. Consider the distribution of outcomes of X.

$$\begin{aligned} &P(X = 0) = (1 - p)^2 \\ &P(X = 1) = p(1 - p) + (1 - p)p = 2p(1 - p) \\ &P(X = 2) = p^2 \\ \Rightarrow &P(X = x) = \binom{2}{x} p^x (1 - p)^{2-x} \end{aligned}$$

This is the probability mass function of a binomial random variable with n=2. This means that $X_1 + X_2 \sim Bernoulli(n, p)$.

d)

We now have two Bernoulli random variables with differing probabilities, p_1 and p_2. Similarly to the last part, we can consider the distribution of outcomes of $X = X_1 + X_2$.

$$\begin{aligned} P(X = 0) &= (1 - p_1)(1 - p_2) \\ P(X = 1) &= p_1(1 - p_2) + (1 - p_1)p_2 \\ P(X = 2) &= p_1 p_2 \end{aligned}$$

This time, there is no nice simple solution for the distribution. But it turns out this is called a Poisson binomial random variable. So the distribution is either characterized by the above, or $X \sim PoissonBinomial(p_1, p_2)$.

Question 2

Suppose that we have a sequence of i.i.d. random variables X_i for i = 1, ..., N , each of which has finite mean μ and variance σ

a)

What is the mean and variance of $\frac{1}{N}\Sigma X_i$ for finite N?
First we will find the expected value.

$$\begin{aligned} E[\frac{1}{N}\Sigma X_i] &= \frac{1}{N}\Sigma E[X_i] \\ &= \frac{1}{N}\Sigma\mu \\ &= \frac{1}{N}n\mu \\ &= \mu \end{aligned}$$

The expected value of the sample mean is the population mean. Now we will find the variance.

$$\begin{aligned} Var(\frac{1}{N}\Sigma X_i) &= \frac{1}{N^2}\Sigma Var(X_i) \\ &= \frac{1}{N}\Sigma\sigma^2 \\ &= \frac{1}{N^2}n\sigma^2 \\ &= \frac{\sigma^2}{N} \end{aligned}$$

The variance of the sample mean is the population variance divided by N.

b)

To find the probability limit, we will use Chebyshev's inequality.

$$P(|X - E[X]| \geq \delta) \leq \frac{V(X)}{\delta^2}$$

Letting $\delta = k\sigma$, and plugging in definition of X give us:

$$P(|\frac{1}{N}\Sigma X_i - \mu| \geq \sigma k) \leq \frac{V(X)}{N\sigma^2 k^2} = \frac{1}{N\sigma k}$$

Taking the limit as $N \to \infty$ gives us

$$P(|\frac{1}{N}\Sigma X_i - \mu| \geq \sigma k) \leq 0$$

This says that the probability that the absolute difference between the sample mean and the population mean is greater than any number is 0. This implies that the sample mean must equal the population mean, and the variance must be 0. This is the same result as we found in part a) if we let N $\to \infty$.

c)

Now we will find the variance of $\frac{1}{\sqrt{N}}\Sigma X_i$.

$$Var(\frac{1}{\sqrt{N}}\Sigma X_i) = \frac{1}{N}\Sigma Var(X_i)$$
$$= \frac{1}{N}N\sigma^2$$
$$= \sigma^2$$

As $N \to \infty$, the variance doesn't change because it doesn't depend on N.

d)

By the central limit theorem since $X_i \sim iid$, in large samples, the sample mean is approximately normally distributed. So the distribution of the sample mean is $\frac{1}{N}\Sigma X_i \sim N(\mu, \frac{\sigma^2}{N})$. As $N \to \infty$, the variance goes to zero. So the sample mean collapses to the population mean as the sample size goes to infinity.

e)

Now we will find the distribution of $\sqrt{N}(\frac{1}{N}\Sigma(X_i - \mu)$. Notice that we are applying the continuous transformation $g(X) = \sqrt{N}(X - \frac{1}{N}\Sigma\mu)$ to the distribution from part d). We can use the delta method to find the distribution of the new function. The derivative of $g(X)$ wrt X is $\sqrt{N}$. The delta method states that the new variance is $g'(X)^2V$ where V is the variance of the untransformed function. This gives us:

$$\sqrt{N}(\frac{1}{N}\Sigma(X_i - \mu) \sim N(0, g'(X)^2V) = N(0, \sqrt{N}^2\frac{\sigma^2}{N}) = N(0, \sigma^2)$$

f)

Now suppose $X \sim Bernoulli(p)$. We will again find the distribution of the sample mean of a finite number of X's. We showed in 1.c that the sum of n Bernoulli variables with the same probability is a binomial random variables with n trials, and probability p. So to characterize the distribution we just need to find the variance.

$$Var(\frac{1}{N}\Sigma X_i) = \frac{1}{N^2}\Sigma Var(X_i)$$
$$= \frac{1}{N^2}\Sigma p(1-p)$$
$$= \frac{p(1-p)}{N}$$

So $\Sigma X_i \sim Binom(n, \frac{p(1-p}{N})$. We can also find the expected value.

$$E[\frac{1}{N}\Sigma X_i] = \frac{1}{N}\Sigma E[X_i]$$
$$= \frac{1}{N}\Sigma p$$
$$= p$$

g)

We can use the delta method again similarly to part e using the same transformation $g(X) = \sqrt{N}(X - \frac{1}{N}\Sigma\mu)$. The delta method tells us that the transformed variable will have a mean zero normal distribution. So we need to find the variance to characterize the distribution.

$$\sqrt{N}(\frac{1}{N}\Sigma(X_i - \mu) \sim N(0, g'(X)^2V) = N(0, \sqrt{N}^2 np(1-p)) = N(0, p(1-p))$$

Question 3

a)

Remember that the theoretical mean for each N is 0, and the variance is $\frac{2}{N}$.

N	Means	Variances
10	0.0124117547	0.299925683
100	-0.0005648683	0.020274050
1000	-0.0002311633	0.002033869

We can see all the means are basically 0, and the variances line up very closely to the predicted $\frac{2}{N}$.

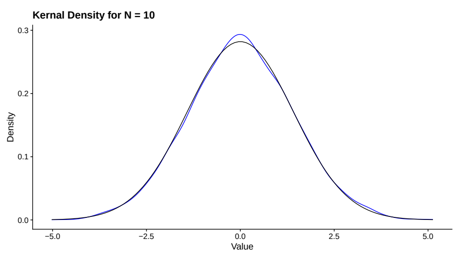
b)

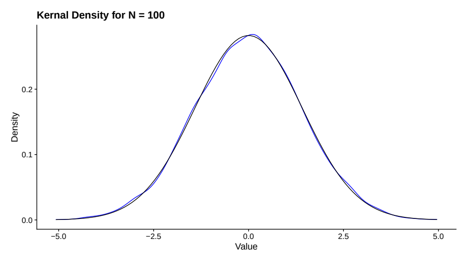
N	Probability > 0.1
10	0.8719
100	0.6094
1000	0.1116

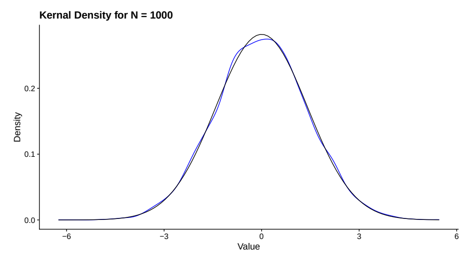
As we know from Chebyshev's inequality, as N increases, the probability of the random variable being larger than some value in this case 0.1 decreases. In the limit, this value will be zero.

c)

Below are the graphs for N=10, N=100, and N=1000. We can see that as we found in 2.c, the variance does not depend







on N. We can also see that the simulated distribution is basically identical to the normal distribution. The KS test fails to reject the null of the distributions being separate as shown below.

N	KS P-Value
10	0.08628215
100	0.2190385
1000	0.6441907

d)

Remember that the theoretical values for the mean and variance are p and $\frac{p(1-p)}{N}$ respectively. For the simulated p=0.8, that gives us a predicted mean of 0.8 and predicted variance of $\frac{0.16}{N}$.

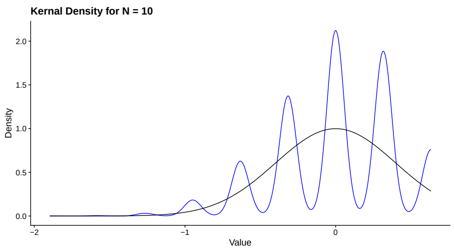
N	Means	Variances
10	0.8031200	0.0159818638
100	0.8003310	0.0015872992
1000	0.7999907	0.0001588263

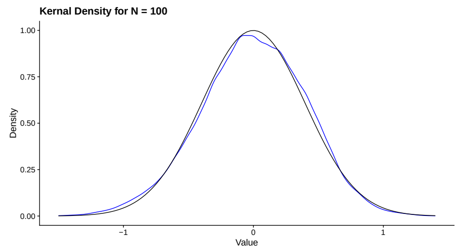
The empirical values are almost identical to the predicted values.

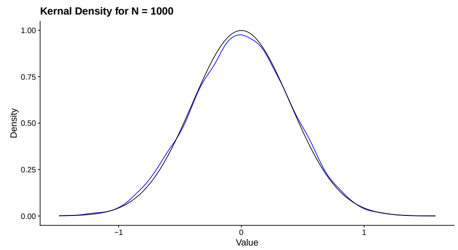
e)

From 2.g, we know that the distribution should theoretically approach a normal distribution. Below are the graphs for N=10, N=100, and N=1000. We see they do indeed approach a normal distribution. The graph for N=10 looks so weird because the sample means are choppy because of the small sample sizes. The KS tests below show the distributions indeed approach the normal distribution. As expected, it takes longer to approach a normal distribution when the underlying data generating process is not normal.

N	KS P-Value
10	0.4098308
100	0.82953591
1000	0.8677999







Question 4

Suppose the true values of β_0 and β_1 in the following model are 0 and 2 respectively. The model is as follows.

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

a)

Suppose $x_i \sim iidN(0,4)$ and $\epsilon_i \sim iidN(0,1)$. We will find the unconditional mean and variance of the vector $\hat{\beta}$.

From the definition of $\hat{\beta}_1$ we will show it is unbiased.

$$\begin{aligned}
\hat{\beta}_1 &= \frac{cov(x,y)}{var(x)} \\
&= \frac{E[(x-E[x])(y-E[y])]}{E[(x-E(x)]^2} \\
&= \frac{E[x(\beta_0+\beta_1+\epsilon-\beta_0]}{E[x^2]} \\
&= \frac{\beta_1 E[x^2]}{E[x^2]} \\
&= \beta_1
\end{aligned}$$

Where the second equality uses the definition of covariance and variance, the third plugs in the definition of y and uses the fact that the expectations of x and of ϵ are both 0. Now we'll find the expectation of $\hat{\beta}_0$.

$$\begin{aligned}
y_i &= \beta_0 + \beta_1 x_i + \epsilon_i \\
\bar{y} &= \hat{\beta}_0 + \hat{\beta}_1 \bar{x} \\
\frac{1}{n}\Sigma(\beta_0 + \beta_1 x_i + \epsilon) &= \hat{\beta}_0 + \hat{\beta}_1 \bar{x} \\
\beta_0 + \beta_1 \bar{x} + \epsilon &= \hat{\beta}_0 + \hat{\beta}_1 \bar{x} \\
\hat{\beta}_0 &= \beta_0 + \bar{x}(\beta_1 - \hat{\beta}_1) + \bar{\epsilon} \\
E[\beta_0] &= \beta_0 + \bar{\epsilon}
\end{aligned}$$

The second equality aggregates, the third plugs in the definition of $\bar{y}$, and the final uses that the expectation of $\hat{\beta}_1$ is β. So β_1 is always unbiased if x is uncorrelated with the error, but β_0 needs the expectation of the error to be zero. Now we can find the variances.

$$\begin{aligned}
Var(\hat{\beta}_1) &= E[(\beta_1 - \beta_1)^2] \\
&= E[(\frac{\Sigma(x_i - E[x_i])\epsilon}{\Sigma(x_i - E[x])})^2] \\
&= E[(\frac{\Sigma(x_i - E[x_i])}{\Sigma(x_i - E[x])})^2]E[\epsilon^2] \\
&= E[(\frac{1}{\Sigma(x_i - E[x])})^2] \\
&= \frac{1}{nVar(x)} \\
&= \frac{1}{4n}
\end{aligned}$$

Similarly, we can get the variance of β_0.

$$\begin{aligned}
Var(\hat{\beta}_0) &= E[(\hat{\beta}_0 - \beta_0)^2] \\
&= E[(\bar{y} - \hat{\beta}_1\bar{x} - \beta_0)^2] \\
&= E[(\beta_0 + \beta_1\bar{x} - \hat{\beta}_1\bar{x} - \beta_0)^2] \\
&= E[(\bar{x}(\beta_1 - \hat{\beta}_1) + \bar{\epsilon})^2] \\
&= E[\bar{x}^2]E[(\beta_1 - \hat{\beta}_1)^2] + E[\bar{\epsilon}^2] \\
&= \frac{\bar{x}}{4n} + \frac{1}{n} \\
&= \frac{1}{n}
\end{aligned}$$

b)

$\hat{\beta}$ is normally distributed because the error terms are normally distributed, and weighted sums of normally distributed terms are also normally distributed. $\hat{\beta} \sim N(\beta, \frac{1}{4})$. Now we will find the distribution of $\sqrt{N}(\hat{\beta} - \beta)$.

$$\begin{aligned}
\sqrt{N}(\hat{\beta} - \beta) &= \sqrt{N}((X'X)^{-1}X'\epsilon) \\
&= N(X'X)^{-1}\frac{1}{\sqrt{N}}X'\epsilon \\
&= (N^{-1}X'X)^{-1}\frac{1}{\sqrt{N}}X'\epsilon \\
&\xrightarrow{d} N(\beta, \frac{1}{4})
\end{aligned}$$

Where the fourth line uses the central limit theorem, and the fifth line uses that the variance of X is 4 and the variance of ϵ is 1. So, $\sqrt{N}(\hat{\beta} - \beta) \sim N(0, 1)$.

c)

Now we will assume that the error terms have a uniform distribution $\epsilon \sim U(0, 2)$ and find the new expected value and variance. First, not the expected value of the errors is $\frac{0+2}{2} = 1$, and the variance is $\frac{(0-2)}{12}$. We can plug these values into the results from part a), and find the new values.

$$\begin{aligned}
E[\hat{\beta}_0] &= \beta_0 + \bar{\epsilon} = 1 \\
E[\hat{\beta}_1] &= \beta_1 = 2
\end{aligned}$$

Now we will find the variance.

$$Var(\hat{\beta}_0) = \frac{1}{3n} \tag{1}$$

$$Var(\hat{\beta}_1) = \frac{1}{4n} + \frac{1}{3n} \tag{2}$$

d)

By the central limit theorem, we know that the resulting distribution will still be normal, however it will now have the new moment conditions we calculated in part c. It will be a normal distribution with mean zero, and the new variances.

e)

Now suppose X and ϵ are correlated with correlation coefficient 0.4. We will find the new expected value of $\hat{\beta}$. We will use the following property of joint normal distributions to find the solution. If X and Y and two normal distributions with correlation coefficient ρ, then

$$E[X|Y] = E[X] + \rho\frac{Var(X)}{Var(Y)}(X - E[x]) \tag{3}$$

Now we can find the expected value.

$$E[\hat{\beta} = E[\beta] + E[(X'X)^{-1}X'E[\epsilon|X]]] \tag{4}$$

Using equation 4, we get:

$$\begin{aligned} E[\epsilon|X] &= E[\epsilon]|\rho\frac{Var(\epsilon)}{Var(x)}(X - E[X]) \\ &= 0 + 0.4\frac{1}{4}X \\ &= \frac{X}{10} \end{aligned}$$

Plugging $\frac{X}{10}$ into equation 5 gives us the following.

$$\begin{aligned} E[\hat{\beta} &= E[\beta] + E[(X'X)^{-1}X'\frac{X}{10}]] \\ &= E[\beta] + \frac{1}{10} \end{aligned}$$

So we find that when X and ϵ are correlated, $\hat{\beta}$ is biased.

Question 5

a)

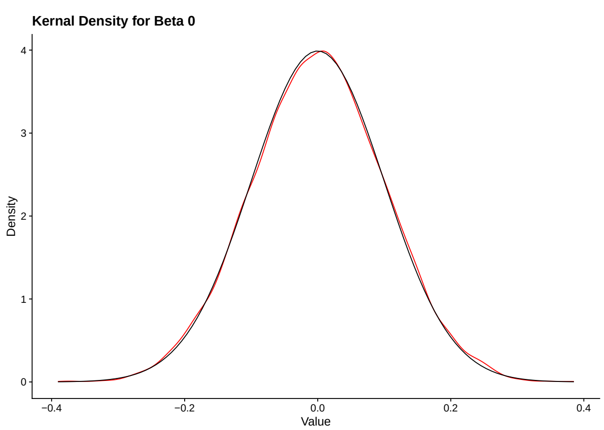
The following table shows the empirical means and variances for β_0 and β_1.

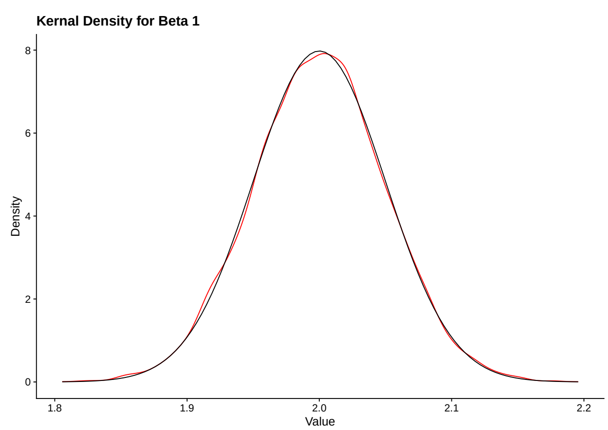
Beta	Means	Variances
β_0	-0.0009956245	0.009961758
β_1	2.0012230579	0.002573589

Recall that the theoretical values for the means are 0, and 2 which match very closely. The values for the variances are $\frac{1}{100} = 0.01$, and $\frac{1}{400} = 0.0025$ which also match very closely.

b)

Below are the graphs for the β's. The values for the KS test are below and they confirm that the distributions are indeed





the normal distribution predicted in the theoretical part.

Beta	KS P-Value
β_0	0.101
β_1	0.01968

WE fail to reject the null for both tests, so we have evidence the distributions are the same.

c)

Below is the table of KS test pvalues for the different n's and β_0.

N	KS P-Value
10	1.85e-06
100	0.1544
1000	0.6125

For β_0, N=10 is too small for it to be normal yet, but after that each distribution is normal. Below is the table for β_1.

N	KS P-Value
10	2.28e-10
100	0.007378
1000	0.7291

For β_1, it takes until N=1000 until the distribution is normal.

d)

Now we have normally distributed errors. Below are the means and variances for each β.

Beta	Means	Variances
β_0	1.000798	0.003322959
β_1	1.999941	0.000848996

Recall the means were predicted to be 1 and 2 respectively and the variances were predicted to be $\frac{1}{300} = 0.003$, and $\frac{1}{400} + \frac{1}{300} = 0.006$ which are close to the empirical value.

e)

Below are the values for the KS tests for β_0.

N	KS P-Value
10	0.0161
100	0.1676
1000	0.2736

It again takes until n=100 to fail to reject the null. Below is the table for β_1.

N	KS P-Value
10	0.6709
100	0.6583
1000	0.9558

This time, it converges very quickly to the predicted normal distribution.

f

We predicted that $\hat{\beta_0}$ should have mean $\beta_0 = 0$, while $\hat{\beta_1}$ should be biased by $\frac{1}{10}$, so $\hat{\beta_1} = 2.1$. Below are the means.

β	Value
β_0	0.0006266365
β_1	2.1007856573