Problem Set 2;
due Monday September 14

Exercise 1 Let X have a normal distribution with mean μ and variance σ^2. Find the probability density function of $Y = e^X$ (This is known as a lognormal distribution). Confirm your answer for $\mu = 0$ and $\sigma^2 = 1$ by simulations.

Solution

$f_Y(y) = \dfrac{dg^{-1}(y)}{dy} f_X\left(g^{-1}(y)\right)$

$g(x) = e^{-x}$, $g^{-1}(y) = ln(y)$

Hence,

$$f_y = \frac{1}{\sqrt{2\pi}} e^{-\dfrac{(ln(y)-\mu)^2}{2\sigma^2}}$$

Exercise 2 Suppose the random variable X_n is equal to the total obtained in the n tosses of a symmetric die.

(a) Applying Chebyshev's inequality, estimate $P(|X_n - 3.5| > \epsilon)$. For what values of n one has $P(|X_n - 3.5| > 0.1) \leq 0.1$?

(b) Applying the central limit theorem, choose n such that the inequality $P(|X_n - 3.5| > 0.1) \leq 0.1$ in $2a$ holds.

Solution

(a) $\sigma_X^2 = E(X^2) - (E(X))^2 = \dfrac{1}{6}(1+4+9+16+25+36) - \dfrac{1}{6}(1+2+3+4+5+6) = \dfrac{91}{6} - 3.5^2 = \dfrac{35}{12}$,

$Var\left(\dfrac{X_n}{n}\right) = \dfrac{\sigma_X^2}{n} = \dfrac{35}{12n}$,

According to the Chebyshev's inequality:

$P\left(\left|\dfrac{X_n}{n} - 3.5\right| > \epsilon\right) \leq \dfrac{35}{12n\epsilon^2}$

For $P(|X_n - 3.5| > 0.1) \leq 0.1$,

$\dfrac{35}{12n(0.1)^2} \leq 0.1$

$n \geq \dfrac{3500}{12}$

$n \geq 2917$

(b) $P\left(\left|\frac{X_n}{n}-3.5\right|>0.1\right)=P\left(\frac{\sqrt{n}}{\sigma}\left|\frac{X_n}{n}-3.5\right|>\frac{0.1\sqrt{n}}{\sigma}\right)\approx P\left(|N(0,1)|>\frac{0.1\sqrt{n}}{\sigma}\right)$

$P\left(|N(0,1)|>\frac{0.1\sqrt{n}}{\sigma}\right)=2\left(1-\Phi\left(\frac{0.1\sqrt{n}}{\sigma}\right)\right)=2\left(1-\Phi\left(\frac{0.1\sqrt{n}}{1.71}\right)\right)=$
$2\left(1-\Phi\left(0.058\sqrt{n}\right)\right)$

Exercise 3

Suppose that X_i *i.i.d* $N(0,1)$ and let $s_n^2=\frac{\sum_{i=1}^n X_i^2}{n}$.

(a) Show that $s_n^2\xrightarrow{p}1$.

(b) Show that $\sqrt{n}\left(s_n^2-1\right)\Rightarrow N\left(0,\omega^2\right)$, and derive an expression for ω^2.

(c) Suppose $n=15$, find $P\left(s_n^2\le 0.73\right)$ using an exact calculation.

(d) Redo part (c) using the normal approximation that you developed in (b).

(e) Let $A_n^2=\frac{\sum_{i=1}^n\left(X_i-\overline{X_n}\right)^2}{n-1}$ Show that $A_n^2\to 1$.

The next problem illustrates the how the difference in the rate at which the tail of the probability density function goes to zero affects the rate of convergence in central limit theorems.

Solution

(a) $s_n^2=\frac{\sum_{i=1}^n X_i^2}{n}=\frac{\sum_{i=1}^n\left(X_i^2\right)}{n}$

According to $N(0,1)$ the $E(X_i)=0$. Hence,

$s_n^2=E\left(X_i^2\right)-\left(E\left(X_i\right)\right)^2=Var\left(X_i\right)=1$

(b) $Var\left(X_i^2\right)=E\left(X_i^4\right)-\left(E\left(X_i^2\right)\right)^2=\sigma^4\left(p-1\right)!!-1=3-1=2$

$\frac{\sqrt{n}\left(s_n^2-1\right)}{\sqrt{Var\left(X_i^2\right)}}=\frac{\sqrt{n}\left(s_n^2-1\right)}{\sqrt{Var\left(X_i^2\right)}}=\sqrt{\frac{n}{Var\left(X_i^2\right)}}\left(s_n^2-1\right)=\sqrt{\frac{n}{Var\left(X_i^2\right)}}\left(\frac{\sum_{i=1}^n X_i^2}{n}-\left(E\left(X_i\right)\right)^2\right)$

$Var\left(X_i^2\right)=\omega^2$

$$\sqrt{\frac{n}{\omega^2}}\left(\frac{\sum_{i=1}^n X_i^2}{n}-\left(E\left(X_i\right)\right)^2\right)\to N(0,1)$$

$$\sqrt{n}\left(\frac{\sum_{i=1}^n X_i^2}{n}-\left(E\left(X_i\right)\right)^2\right)\to\omega N(0,1)$$

$$\sqrt{n}\left(\frac{\sum_{i=1}^n X_i^2}{n}-\left(E\left(X_i\right)\right)^2\right)\to N\left(0,\omega^2\right),\text{ where }\omega^2=2$$

(c) $\sum_{i=1}^{n} X_i^2$ - it's χ^2 distribution with n dimension of freedom. Hence,

$P\left(s_n^2 \leq 0.73\right) \approx P\left(\frac{\chi^2}{n} \leq 0.73\right) = P\left(\chi^2 \leq 0.73n\right)$

For n = 15

$$P\left(\chi^2 \leq 10.95\right) = \frac{\gamma\left(\frac{15}{2}, \frac{10.95}{2}\right)}{\Gamma\left(\frac{15}{2}\right)} = 0.24$$

(d) $P\left(s_n^2 \leq 0.73\right) = P\left(s_n^2 - 1 \leq 0.73 - 1\right) = P\left(\frac{\sqrt{n}\left(s_n^2 - 1\right)}{\omega} \leq \frac{\sqrt{n}\left(0.73 - 1\right)}{\omega}\right)$

$$P\left(\frac{\sqrt{n}\left(s_n^2 - 1\right)}{\omega} \leq \frac{\sqrt{n}\left(0.73 - 1\right)}{\omega}\right) \approx P\left(N\left(0,1\right) \leq -\frac{\sqrt{15} \cdot 0.27}{\sqrt{2}}\right)$$

$$P\left(N\left(0,1\right) \leq -\frac{\sqrt{15} \cdot 0.27}{\sqrt{2}}\right) = P\left(N\left(0,1\right) \geq \frac{\sqrt{15} \cdot 0.27}{\sqrt{2}}\right) = 1 - P\left(N\left(0,1\right) \leq \frac{\sqrt{15} \cdot 0.27}{\sqrt{2}}\right) =$$

$$1 - \Phi\left(\frac{\sqrt{15} \cdot 0.27}{\sqrt{2}}\right) = 1 - \Phi\left(0.74\right) = 1 - 0.77 = 0.23$$

(e) $A_n^2 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X_n}\right)^2}{n - 1} = var(\hat{X})$

$\lim_{x \to \infty} var(\hat{X}) = var(X) = 1$

So $A_n^2 = 1$, QED

Exercise 4

Let $X_1, X_2, \cdots$ be a sequence of independent uniform random variables, and define $Y_n = \sqrt{\frac{12}{n}} \sum_{i=1}^{n}\left(X_i - \frac{1}{2}\right)$, and let $Z_1, Z_2, \cdots$ be a sequence exponential random variables, and define $W_n = \frac{\sum_{i=1}^{n}\left(Z_i - 1\right)}{\sqrt{n}}$

(a) Calculate the mean and variance and third moment of Y_n and W_n.

(b) Calculate for $n = 1$ the probability that Y_n and W_n are larger than 1.96 and the probability they are smaller than -1.96.

(c) Estimate the probability that Y_n and W_n are larger than 1.96 and the probability they are smaller than -1.96 by drawing 10000 random variables from their respective distributions (for $n = 1$).

(d) Repeat the previous exercise for $n = 2$, $n = 5$ and $n = 100$. Which of the two sequences converges to a normal distribution faster?

Solution

(a) Lets find $E(X_i)$ and $E(X_i^2)$

$E(X_i) = \int_{-\infty}^{\infty} x f(x)\, dx = \int_a^b x dx = \frac{x^2}{2}|_a^b = \frac{b^2}{2} - \frac{a^2}{2}$

For $x \in [0, 1]$ $E(X_i) = \frac{1}{2}$

$E(X_i^2) = \int_{-\infty}^{\infty} x^2 f(x)\, dx = \int_a^b x^2 dx = \frac{x^3}{3}|_a^b = \frac{b^3}{3} - \frac{a^3}{3}$

For $x \in [0, 1]$ $E(X_i^2) = \frac{1}{3}$

Lets find expectation of Y_n:

$E(Y_n) = E\left(\frac{12}{n}\sum_{i=1}^n\left(X_i - \frac{1}{2}\right)\right) = \frac{12}{n}E\left(\sum_{i=1}^n\left(X_i - \frac{1}{2}\right)\right) = \frac{12}{n}\sum_{i=1}^n E\left(X_i - \frac{1}{2}\right) =$
$\sqrt{\frac{12}{n}}\sum_{i=1}^n\left(E(X_i) - E\left(\frac{1}{2}\right)\right) = \sqrt{\frac{12}{n}}\sum_{i=1}^n\left(\frac{1}{2} - \frac{1}{2}\right) = 0$

Lets find variance of Y_n:

$Var(X_i) = E(X_i^2) - (E(X_i))^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$

$Var(Y_n) = Var\left(\sqrt{\frac{12}{n}}\sum_{i=1}^n\left(X_i - \frac{1}{2}\right)\right) = \frac{12}{n}Var\left(\sum_{i=1}^n\left(X_i - \frac{1}{2}\right)\right) = \frac{12}{n}\sum_{i=1}^n Var(X_i) =$
$\frac{12}{n}\cdot n \cdot Var(X_i) = \frac{12}{n}\cdot\frac{1}{12}\cdot n = 1$

Lets find third moment of Y_n:

$E(Y_n^3) = E\left(\left(\sqrt{\frac{12}{n}}\sum_{i=1}^n\left(X_i - \frac{1}{2}\right)\right)^3\right) = \left(\frac{12}{n}\right)^{1.5} E\left(\left(\sum_{i=1}^n\left(X_i - \frac{1}{2}\right)\right)^3\right) =$
$\left(\frac{12}{n}\right)^{1.5}\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n E\left(X_i - \frac{1}{2}\right)\left(X_j - \frac{1}{2}\right)\left(X_k - \frac{1}{2}\right) = \left(\frac{12}{n}\right)^{1.5}\left(E\left(X_i - \frac{1}{2}\right)\right)^3 =$
$\left(\frac{12}{n}\right)^{1.5}\left(E(X_i) - E\left(\frac{1}{2}\right)\right) = \left(\frac{12}{n}\right)^{1.5}\cdot 0 = 0$

Lets calculate expectation and variance of Z_i

$E(Z_i) = \int_0^{\infty} x \cdot e^{-x} dx = e^{-x}(-x-1)|_0^{\infty} = 1$

$E\left(Z_i^2\right) = \int_0^\infty x^2 \cdot e^{-x} dx = 2$

$Var\left(Z_i\right) = E\left(X_i^2\right) - \left(E\left(Z_i\right)\right)^2 = 2 - 1 = 1$

Lets find expectation of W_n:

$E\left(W_n\right) = E\left(\frac{\sum_{i=1}^n \left(Z_i - 1\right)}{\sqrt{n}}\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(E\left(Z_i\right) - E\left(1\right)\right) = \frac{n}{\sqrt{n}} \cdot (1-1) = 0$

Lets find variance of W_n:

$Var\left(W_n\right) = Var\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \left(Z_i - 1\right)\right) = \frac{1}{n} Var\left(\sum_{i=1}^n \left(Z_i - 1\right)\right) = \frac{1}{n} \sum_{i=1}^n Var\left(Z_i\right) = \frac{1}{n} \cdot n \cdot 1 = 1$

Lets find third moment of W_n:

$E\left(W_n^3\right) = E\left(\left(\frac{Z_i - 1}{\sqrt{n}}\right)^3\right) = \left(\frac{1}{\sqrt{n}}\right)^3 E\left(\left(Z_i - 1\right)^3\right) = \left(\frac{1}{\sqrt{n}}\right)^3 \left(E\left(Z_i - 1\right)\right)^3 = \left(\frac{1}{\sqrt{n}}\right)^3 \left(E\left(Z_i\right) - E\left(1\right)\right)^3 = \left(\frac{1}{\sqrt{n}}\right)^3 (1-1) = 0$

(b) 1) $P\left(\sqrt{12}\left(X_i - \frac{1}{2}\right) \leq -1.96\right) = P\left(X_i \leq \frac{-1.96}{\sqrt{12}} + \frac{1}{2}\right) = P\left(X_i \leq 0.0658\right)$
$P\left(X_i \leq 0.0658\right) = 0$

2) $P\left(\sqrt{12}\left(X_i - \frac{1}{2}\right) \geq 1.96\right) = 1 - P\left(\sqrt{12}\left(X_i - \frac{1}{2}\right) \leq 1.96\right) = 1 - P\left(X_i \leq \frac{1.96}{\sqrt{12}} + \frac{1}{2}\right) = 1 - 1 = 0$

3) $P\left(Z_i - 1 \leq -1.96\right) = p\left(Z_i \leq -0.96\right) = 0$

4) $P\left(Z_i - 1 \geq 1.96\right) = P\left(Z_i \geq 2.96\right) = 1 - P\left(Z_i \leq 2.96\right) = 1 - 1 + e^{-x} = 0.0518$

(c) $Y_n = \sqrt{12} * (X - 0.5)$ $W_n = Z - 1$

It was generated 10000 uniform random variables in vector X and the same sizy exponential distributed vector Z.

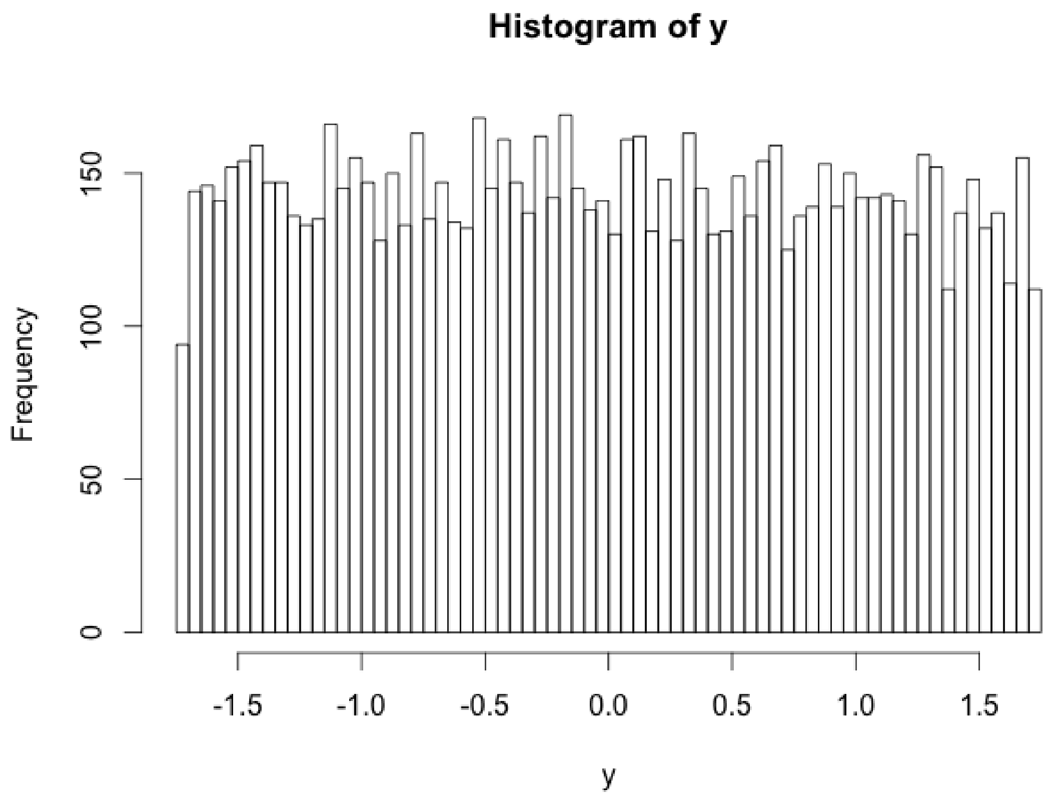
We can estimate the probability by processing the real data.

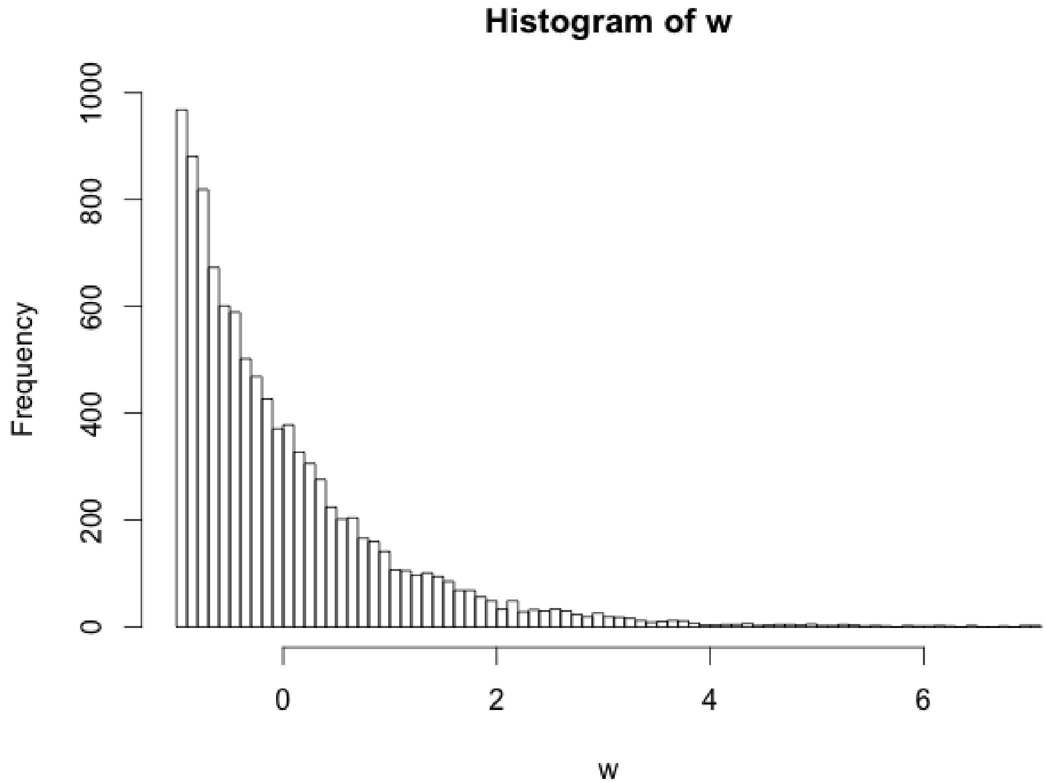
$P(Y_n > 1.96) = 0$

$P(W_n > 1.96) = 0.0513$

$P(Y_n < -1.96) = 0$

$P(W_n < -1.96) = 0$


Histogram of y
Frequency
0
50
100
150
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
y



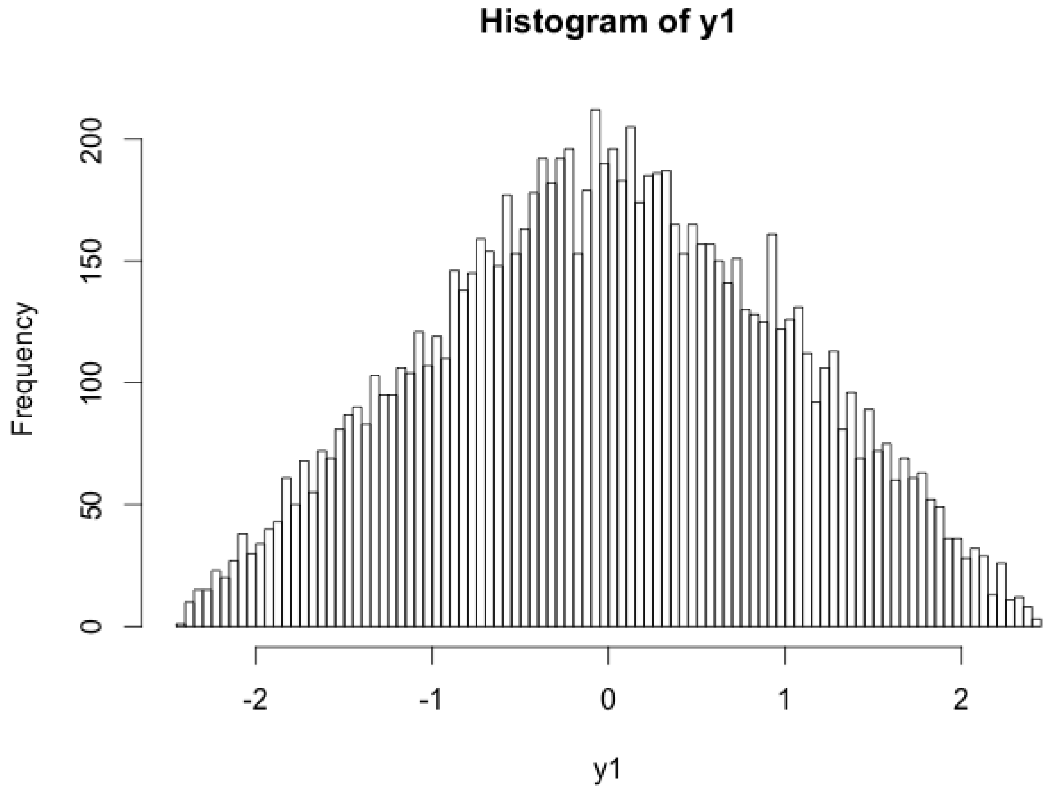
(d) n = 2

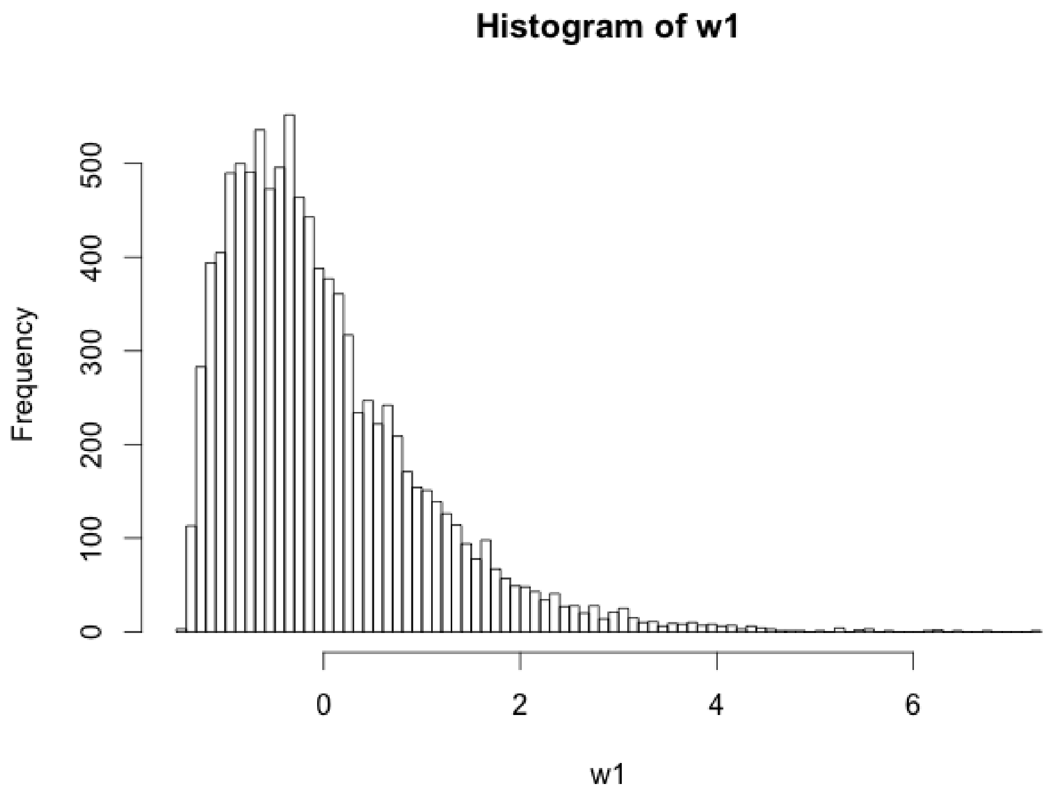
$P(Y_n > 1.96) = 0.0187$

$P(W_n > 1.96) = 0.0477$

$P(Y_n < -1.96) = 0.0205$

$P(W_n < -1.96) = 0$


Histogram of y1
Frequency
0
50
100
150
200
-2
-1
0
1
2
y1



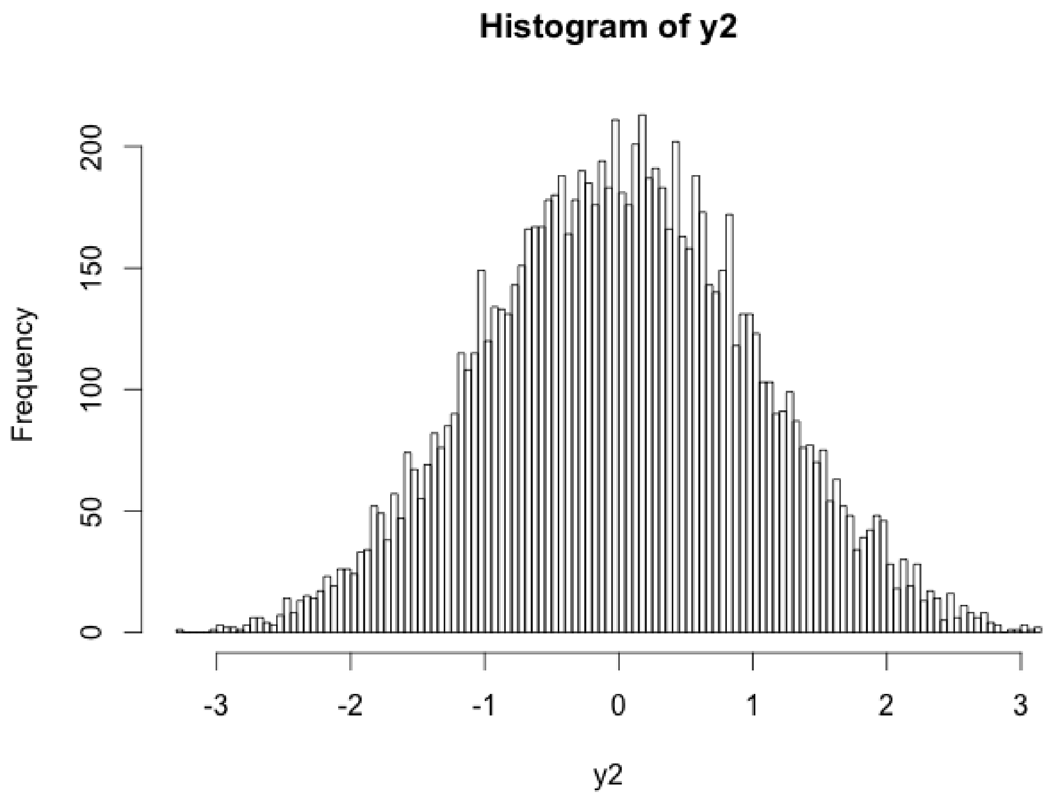
n = 5

$P(Y_n > 1.96) = 0.0278$

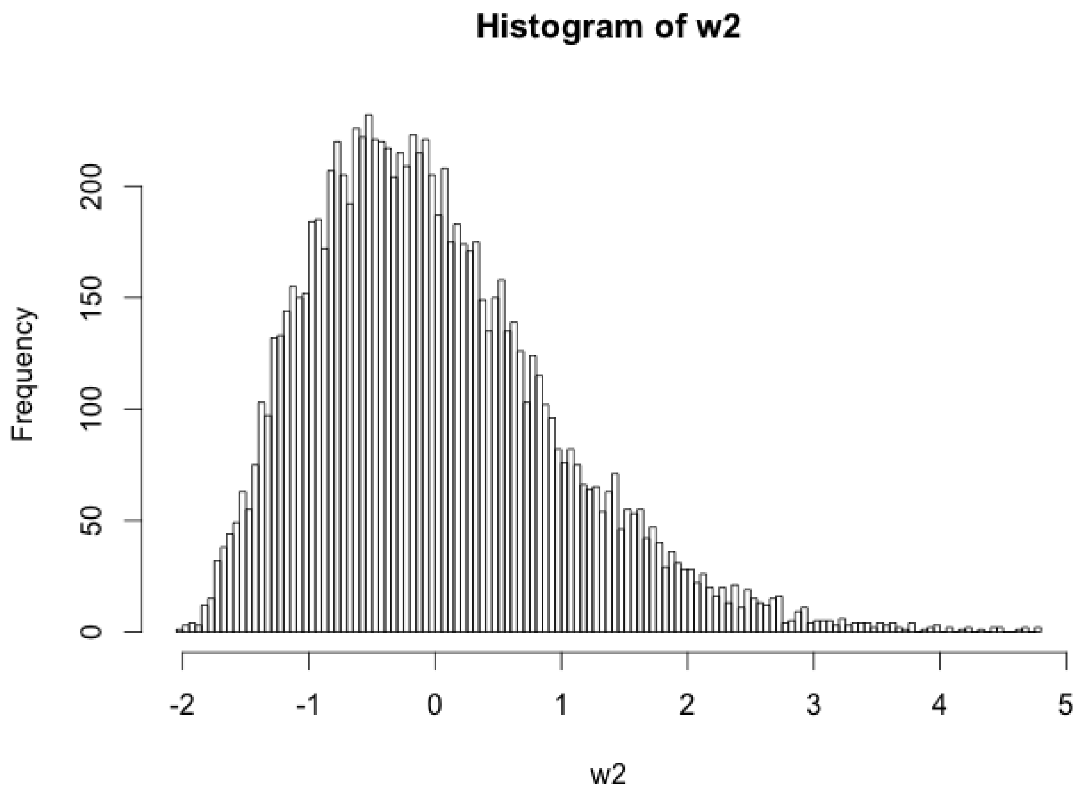
$P(W_n > 1.96) = 0.0402$

$P(Y_n < -1.96) = 0.0231$

$P(W_n < -1.96) = 4e - 04$



Histogram of y2
Frequency
0
50
100
150
200
-3
-2
-1
0
1
2
3
y2



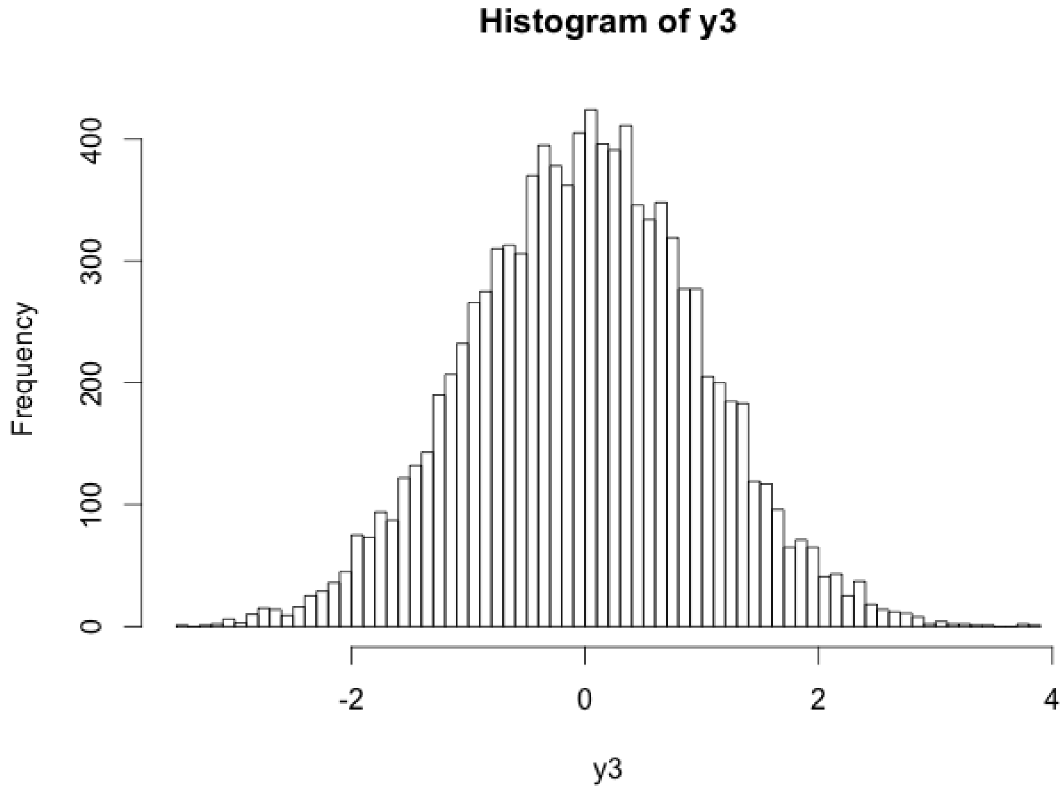
n = 100

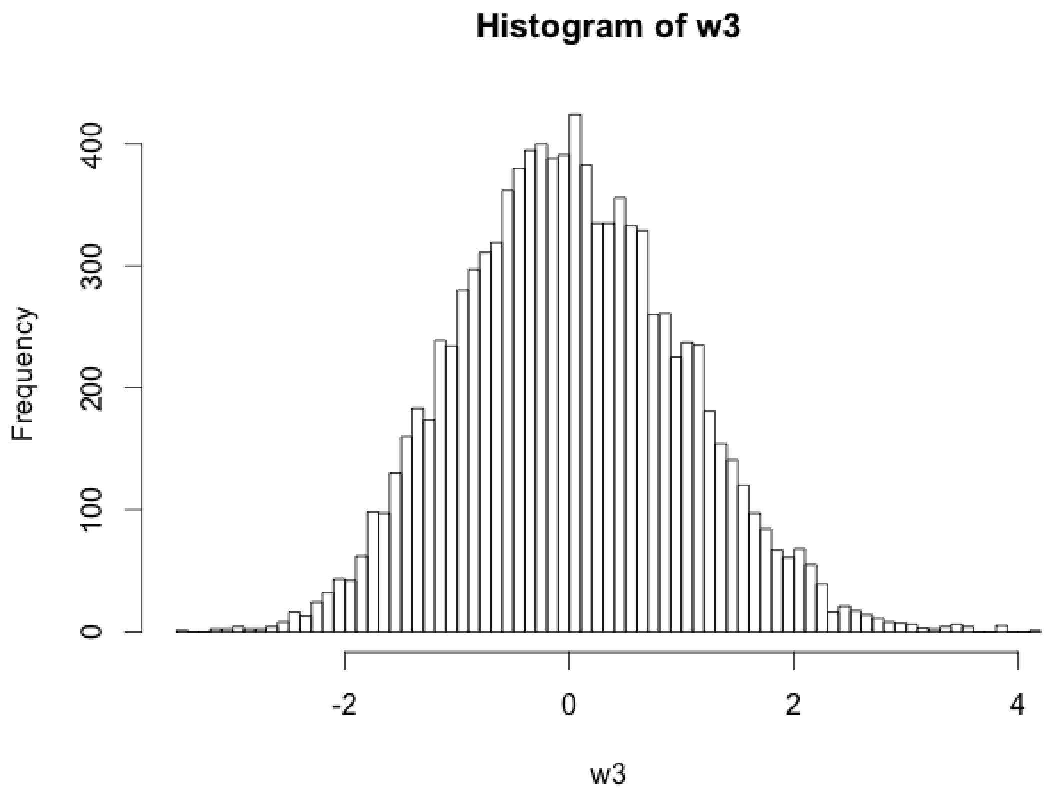
$P(Y_n > 1.96) = 0.0254$

$P(W_n > 1.96) = 0.0311$

$P(Y_n < -1.96) = 0.0239$

$P(W_n < -1.96) = 0.017$


Histogram of y3
Frequency
0
100
200
300
400
-2
0
2
4
y3



It is obvious that Y is closer to Normal distribution. It is well known formula to generate normal distributed values

Exercise 5 Suppose that $X_n \approx N\left(1, \frac{1}{n}\right))$ for all n. Use the delta method to find an approximate normal distribution for $\sqrt{n}(e^{X_n} - e^1)$.

Solution

We have $X_n \sim N\left(1, \frac{1}{n}\right)$

$X_n - 1 \sim N\left(0, \frac{1}{n}\right)$

$X_n - 1 \sim \frac{1}{\sqrt{n}} N(0, 1)$

$\sqrt{n}\,(X_n - 1) \sim N(0, 1)$

According to the delta-method $\sqrt{n}\,(g(x) - g(\theta))\;\; N\left(0, \sigma^2 \cdot g'(\theta)^2\right)$

Hence,

$g(x) = e^x, g'(x) = e^x$

$\sqrt{n}(e^{X_n} - e^1) \sim N(0, e)$

Exercise 6

In a survey of 400 likely voters, 215 responded that they would vote for the incumbent and 185 responded that they would vote for the challenger. Let p denote the fraction of all likely voters that preferred the incumbent at the time of the survey, and let $\hat{p}$ be the fraction of survey respondents that preferred the incumbent.

(a) Use the survey results to estimate $\hat{p}$.

(b) Use the estimator of the variance of $\hat{p}$, $\frac{\hat{p}(1-\hat{p})}{n}$, to calculate the standard error of your estimator.

(c) What is the p-value for the test $H_0 : p = 0.5$ vs. $H_a : p \neq 0.5$?

(d) What is the p-value for the test $H_0 : p = 0.5$ vs. $H_a : p > 0.5$?

(e) Did the survey contain statistically significant evidence that the incumbent was ahead of the challenger at the time of the survey?

Solution

(a) $\hat{p} = \frac{215}{400} = 0.5375$

(b) $SE(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.02492959$

(c) $t = \frac{\hat{p} - p}{SE(\hat{p})} = 1.504237$

p-value $= 2\Phi(-|t|) = 0.132$

(d) p-value $=$ 1-$\Phi(t) = 0.066$

(e) According to D and C we can not reject H_0 hypothesis because t statistic is less than 1.645. Moreover p-value is 0.066 that is bigger 0.05. So statisticaly the survey was wrong.

Exercise 7 Suppose X_i $i.i.d.$ $N(0, \sigma^2)$, and let $\hat{\sigma^2} = \frac{\sum_{i=1}^{n} X_i^2}{n}$.

(a) Show that $\hat{\sigma^2}$ is unbiased.

(b) Compute the variance of $\hat{\sigma^2}$.

(c) Compute the Rao-Cramer lower bound on the variance of any unbiased estimator of $\hat{\sigma^2}$.

Solution

(a) $E\left(\hat{\sigma^2}\right) = E\left(\dfrac{\sum_{i=1}^{n} X_i^2}{n}\right) = \dfrac{1}{n}\sum_{i=1}^{n} E\left(X_i^2\right) = E\left(X_i^2\right) = \sigma^2$

(b) $Var\left(\hat{\sigma^2}\right) = Var\left(\dfrac{\sum_{i=1}^{n} X_i^2}{n}\right) = \dfrac{1}{n^2}\sum_{i=1}^{n} Var\left(X_i^2\right) = \dfrac{1}{n^2}\sum_{i=1}^{n} \cdot 2\sigma^4 = \dfrac{2\sigma^4}{n}$

(c) $I\left(\sigma^2\right) = -E\left[\dfrac{\partial^2 L_n\left(\sigma^2\right)}{\partial\left(\sigma^2\right)^2}\right] = -E\left[\dfrac{n}{2\sigma^4} - \sum_{i=1}^{n} \dfrac{X_i^2}{\sigma^6}\right]\Bigg] = -\dfrac{n}{2\sigma^4} + \sum_{i=1}^{n} \dfrac{E\left[X_i^2\right]}{\sigma^6} =$
$-\dfrac{n}{2\sigma^4} + \dfrac{n\sigma^2}{\sigma^6} = \dfrac{n}{2\sigma^4}$

Therefore, the lower bound on the variance of any unbiased estimated σ^2 is $\dfrac{n}{2\sigma^4}$