

STATS 700 Solution: HW0

Yuanzhi Li

September 18, 2017

1 Coinflip Trivia

Denote X_1, X_2, \dots, X_n as the results for the n coin flips, each of which takes value in $\{H, T\}$. From the problem setup, we can see that they are independent and identically distributed random variables. Each X_k follows a Bernoulli distribution: $\Pr(X_k = H) = 0.4$ and $\Pr(X_k = T) = 0.6$.

Question 1

Since the sequence of $\{X_k : k = 1, \dots, n\}$ are mutually independent, we have:

$$\begin{aligned}\Pr(X_1 = H, X_2 = H, X_3 = T) &= \Pr(X_1 = H) \cdot \Pr(X_2 = H) \cdot \Pr(X_3 = T) \\ &= 0.4 \times 0.4 \times 0.6 \\ &= 0.096.\end{aligned}$$

Question 2

Note that by taking intersection with the event $\{X_1 = T\}$, we fix the result for the first flip. Rewriting the conditional probability as:

$$\begin{aligned}\Pr(X_1 = T \mid \text{two of } X_1 \dots X_5 \text{ are } H) &= \frac{\Pr(X_1 = T, \text{two of } X_2 \dots X_5 \text{ are } H)}{\Pr(\text{Two of } X_1 \dots X_5 \text{ are } H)} \\ &= \frac{\Pr(X_1 = T) \cdot \Pr(\text{Two of } X_2 \dots X_5 \text{ are } H)}{\Pr(\text{Two of } X_1, \dots, X_5 \text{ are } H)} \\ &= \frac{0.6 \times \binom{4}{2} \times 0.4^2 \times 0.6^2}{\binom{5}{2} \times 0.4^2 \times 0.6^3} \\ &= 0.6.\end{aligned}$$

The code for numerical simulation is attached below:

```
set.seed(133)
Loop = 1e5
count.P = 0 ## The empirical frequency of B
count.Q = 0 ## The empirical frequency of AB
for(i in 1:Loop){
  rnd.X = rbinom(5,1,0.4)
  if(sum(rnd.X) == 2) count.P = count.P + 1
  if(rnd.X[1]==0 && sum(rnd.X[2:5]) == 2) count.Q = count.Q + 1
}
Prob = count.Q/count.P ## The conditional probability
Prob
```

```
## [1] 0.6026477
```

The results for $N = 10^5$ sets of experiments give an estimated probability of 0.6026. The true probability is 0.6.

Question 3

Let Y_n denote the number of ‘heads’ for the first n tosses. Then it is clear that Y_n follows a Binomial distribution $B(n, 0.4)$.

Assume that we win \$6 for each toss regardless of the result, then we win an additional \$4 if it is a ‘head’. This experiment would be equivalent to the original one. Then the final reward we would win from the first n trials equals to $Z_n := 6n + 4Y_n$. From the property of Binomial distribution, we can see:

$$\begin{aligned} E(Z_n) &= 6n + 4E(Y_n) = 6n + 1.6n = 7.6n, \\ \text{var}(Z_n) &= 16\text{var}(Y_n) = 16 \times 0.4 \times 0.6n = 3.84n. \end{aligned}$$

```
set.seed(137)
Loop = 1e5
n = 10
Reward = rep(NA, Loop)
for(i in 1:Loop){
  rnd.X = rbinom(n, 1, 0.4)
  Reward[i] = 6 * n + 4 * sum( rnd.X == 1 )
}
Emp.mean = mean(Reward)
Emp.mean
```

```
## [1] 75.99856
```

```
Emp.var = var(Reward)
Emp.var
```

```
## [1] 38.49286
```

The results for $N = 10^5$ sets of experiments give an estimated mean of 75.9986, an estimated variance of 38.4929. The true values are $E(Z_{10}) = 76$, $\text{var}(Z_{10}) = 38.4$.

2 Exchangeability

Question 1

First, we calculate the marginal probability conditioning on P . Since $\{X_k : k = 1, \dots, n\}$ are mutually independent when conditioned on P , we have:

$$\begin{aligned}\Pr(X_1 = x_1, \dots, X_k = x_k \mid P) &= \prod_{j=1}^k \Pr(X_j = x_j \mid P) \\ &= \prod_{j=1}^k P^{x_j} (1 - P)^{1-x_j} \\ &= P^t (1 - P)^{k-t},\end{aligned}$$

where $t = \sum_{j=1}^k x_j$.

Hence, the marginal probability of X_1, \dots, X_k is

$$\begin{aligned}\Pr(X_1 = x_1, \dots, X_k = x_k) &= \int \Pr(X_1 = x_1, \dots, X_k = x_k \mid P) \cdot \Pr(P) \, dP \\ &= \int_0^1 p^t (1 - p)^{k-t} \, dp \\ &= B(t + 1, k - t + 1).\end{aligned}$$

For any finite permutation on any subset of X_1, \dots, X_n , say X_{s_1}, \dots, X_{s_k} . The marginal distribution would be:

$$\Pr(X_{s_1}, \dots, X_{s_k}) = B\left(\sum_{j=1}^k X_{s_j} + 1, k - \sum_{j=1}^k X_{s_j} + 1\right),$$

regardless of the order of s_1, \dots, s_k . This shows that X_1, \dots, X_n are exchangeable.

Question 2

$$\begin{aligned}\text{RHS} &= \prod_{i=1}^n \Pr(X_i = x_i) \\ &= \prod_{i=1}^n \int_0^1 p^{x_i} (1 - p)^{1-x_i} \, dp \\ &= \left(\frac{1}{2}\right)^n.\end{aligned}$$

Recall that LHS is $B(\sum_{i=1}^n x_i + 1, n + 1 - \sum_{i=1}^n x_i)$, which apparently doesn't agree to RHS. This implies that X_1, \dots, X_n are not independent and identically distributed.

3 Gamma-Poisson

From the problem set up, we have:

$$Y \mid N \sim \chi^2(2N), \quad N \sim \text{Poisson}(\theta).$$

Further, the moments are:

$$\begin{aligned} E(Y \mid N) &= 2N, & \text{var}(Y \mid N) &= 4N, \\ E(N) &= \theta, & \text{var}(N) &= \theta. \end{aligned}$$

Question 1

From the law of iterated expectation and total variance, we have:

$$\begin{aligned} E(Y) &= E\{E(Y \mid N)\} = E(2N) = 2\theta, \\ \text{var}(Y) &= E\{\text{var}(Y \mid N)\} + \text{var}\{E(Y \mid N)\} \\ &= E(4N) + \text{var}(2N) = 8\theta \end{aligned}$$

We conduct simulation with increasing sample sizes at $m = 10, 100, 1000$ and 10000 . For each set of simulation, we generate 5000 samples of size m and then calculate the averages of 5000 estimates along with confidence intervals. The code for simulation is attached below:

```
set.seed(133)
## Number of samples for generating confidence interval
Repeat.num = 5000
m = c(1e1, 1e2, 1e3, 1e4)
Emp.mean = rep(NA, Repeat.num)
Emp.var = rep(NA, Repeat.num)

## Average of point estimations
Est.mean = rep(NA, length(m))
Est.var = rep(NA, length(m))

## Confidence intervals
CI.mean = matrix(NA, length(m), 2)
CI.var = matrix(NA, length(m), 2)
for(i in 1:length(m)){
  Emp.mean = 0
  Emp.var = 0
  for (j in 1:Repeat.num){
    rnd.N = rpois(m[i], 1)
    rnd.Y = rchisq(m[i], 2*rnd.N)
    Emp.mean[j] = mean(rnd.Y)
    Emp.var[j] = var(rnd.Y)
  }
  Est.mean[i] = mean( Emp.mean )
  CI.mean[i,] = quantile(Emp.mean, c(0.025, 0.975))
  Est.var[i] = mean(Emp.var)
  CI.var[i,] = quantile(Emp.var, c(0.025, 0.975))
}
```

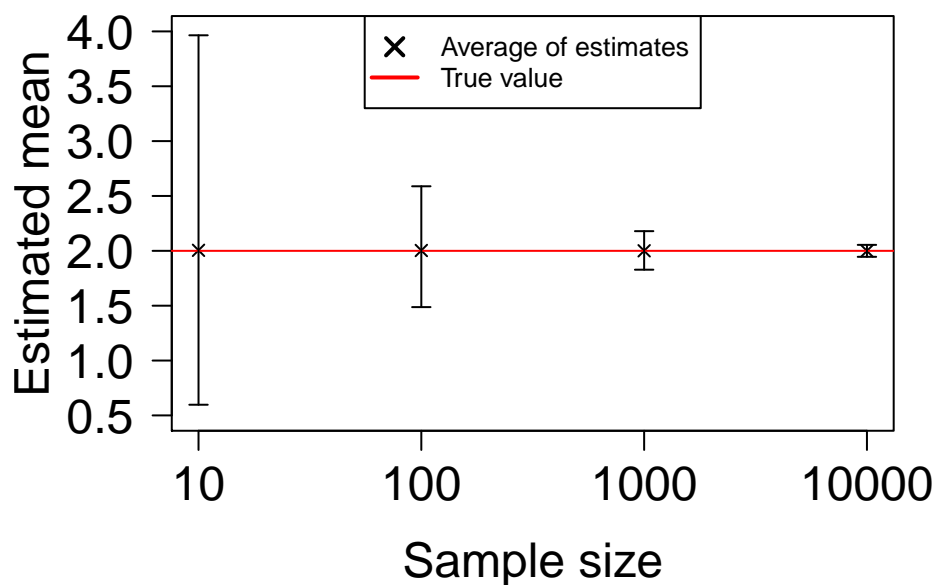
```

}

## Visualizing the results
## Plot the estimated mean and confidence interval
opar = par()
par(cex.axis=1.5,cex.lab=1.5,las = 1)
plot(Est.mean~m,xlab = 'Sample size', ylab = 'Estimated mean',
     ylim = c(0.5,4), pch=4, cex=0.8 , log = c('x','y'),
     main = 'The 95% confidence intervals \n of the estimated mean.',xaxt='n')
axis(1,at = m)
abline(h=2,col='red')
segments(x0 = m,y0=CI.mean[,1],y1=CI.mean[,2])
segments(x0 = 1.1*m,y0 = CI.mean[,1], x1=m/1.1)
segments(x0 = 1.1*m,y0 = CI.mean[,2], x1=m/1.1)
legend('top',legend = c('Average of estimates','True value'),lwd=2,pt.cex=1.2,cex=0.8,
      lty = c(-1,1), pch=c(4,-1),col=c('black','red'))

```

**The 95% confidence intervals
of the estimated mean.**

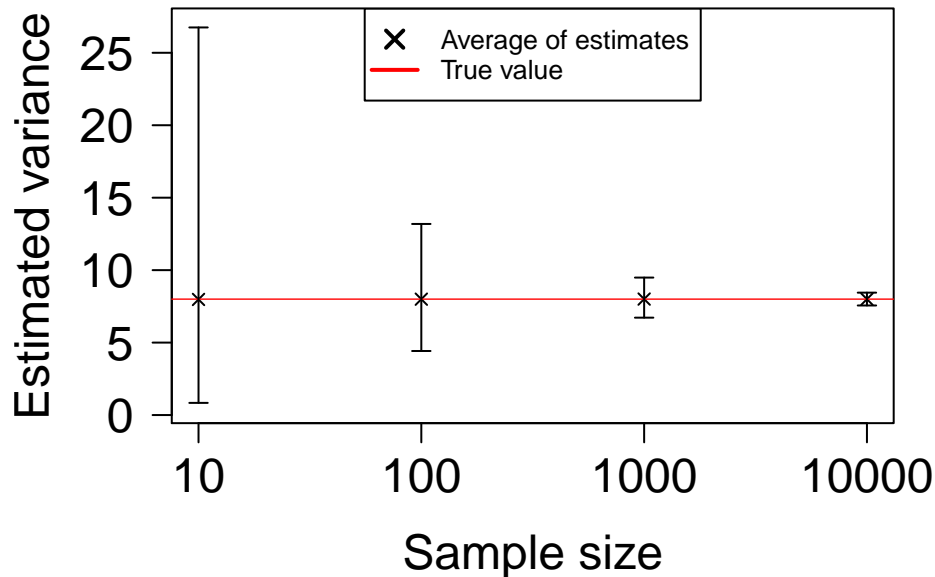


```

## Plot the estimated variance and confidence interval
plot(Est.var~m,xlab = 'Sample size', ylab = 'Estimated variance',log=c('x','y'),
     ylim = c(0.5,27), pch=4, cex=0.8 ,
     main = 'The 95% confidence intervals \n of the estimated variance.',xaxt='n')
axis(1,at = m)
abline(h=8, col='red',lwd=0.7)
segments(x0 = m,y0=CI.var[,1],y1=CI.var[,2])
segments(x0 = 1.1*m,y0 = CI.var[,1], x1=m/1.1)
segments(x0 = 1.1*m,y0 = CI.var[,2], x1=m/1.1)
legend('top',legend = c('Average of estimates','True value'),lwd=2,pt.cex=1.2,cex=0.8,
      lty = c(-1,1), pch=c(4,-1),col=c('black','red'))

```

The 95% confidence intervals of the estimated variance.

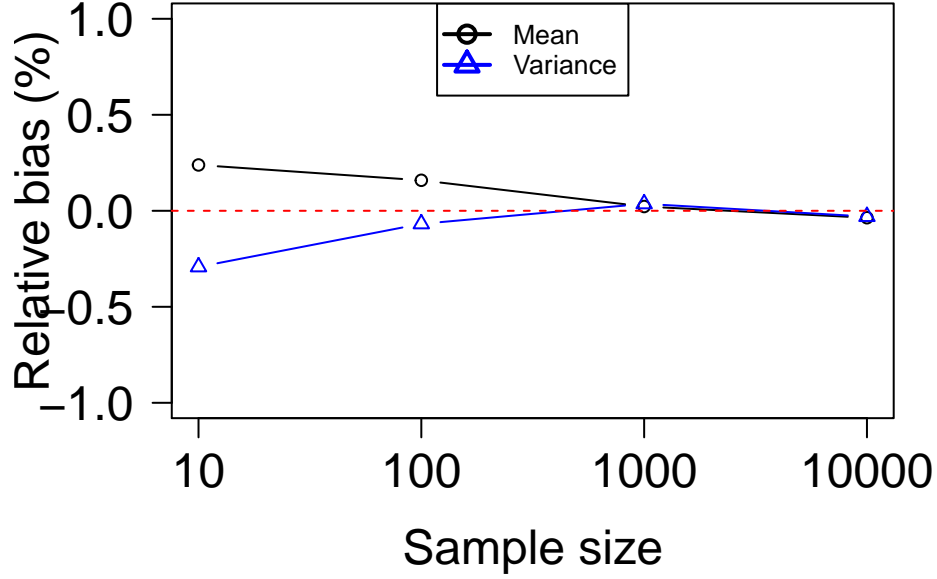


The figures above show the change of 95% confidence intervals with different sample sizes. Note that both the x -axis (sample size) and y -axis (estimates) are rescaled by taking logarithm. We can see that the averaged sample mean and averaged sample variance among the 5000 simulation centered toward the true values, regardless of the sample size m . However, the confidence intervals shrink rapidly as sample size m increases. When $m = 10$, the confidence interval is relatively wide; While when m grows larger, the confidence intervals become condensed at the true values, which indicated that nearly all the estimates are close to the true values. Intuitively, this result verifies that the estimators converges to the true values with probability one.

```
## Plot the relative bias
opar = par()
par(cex.axis=1.5,cex.lab=1.5,las = 1)

plot((Est.mean-2)/2*1e2~m,xlab = 'Sample size', ylab = 'Relative bias (%)',log='x',
     ylim = c(-0.01,0.01)*1e2, cex = 0.8, type = 'b',
     main = 'Relative bias of the estimators.',xaxt='n')
axis(1,at = m)
points((Est.var-8)/8*1e2~m,xlab = 'log sample size', ylab = 'log mean', pch=2 ,
      cex = 0.8, type = 'b', col = 'blue')
abline(h=0,col='red', lty = 2)
legend('top',legend = c('Mean','Variance'),lwd=2,pt.cex=1.2,cex=0.8, lty = c(1,1),
      pch=c(1,2),col=c('black','blue'))
```

Relative bias of the estimators.



This figure above shows the averaged relative bias of the estimators with the log sample size $\log(m)$. The relative bias is not that large as we would imagine even if $m = 10$. Also, we can see that the relative bias is steadily decreasing as m grows large. When $m \geq 1000$, the relative bias for estimating the mean and variance look identical to each other.

Question 2

We show this proposition by utilizing characteristic function. Denote $Z_\theta = (Y - 2\theta)/\sqrt{8\theta}$. First, we calculate the characteristic function of the LHS variable Z_θ . Note that the characteristic function of a $\chi^2(m)$ distributed random variable is

$$f_m(\xi) = \frac{1}{(1 - 2i\xi)^{m/2}}.$$

With a given N , Y follows chi-squared distribution with dof $2N$. Hence the conditional expectation of $\exp(i\xi Y)$ is

$$\mathbb{E}(e^{i\xi Y} | N) = f_{2N}(\xi) = \left(\frac{1}{1 - 2i\xi} \right)^N.$$

Then we can calculate the (unconditional) characteristic function of Y , $g_Y(\xi)$, by the law of total expectation:

$$\begin{aligned} g_Y(\xi) &= \mathbb{E}(e^{i\xi Y}) = \mathbb{E}\{\mathbb{E}(e^{i\xi Y} | N)\} \\ &= \mathbb{E}\left(\frac{1}{1 - 2i\xi} \right)^N \\ &= \sum_{j=1}^{\infty} \left(\frac{1}{1 - 2i\xi} \right)^j e^{-\theta} \frac{\theta^j}{j!} \\ &= \exp\left(\frac{2i\xi\theta}{1 - 2i\xi} \right), \end{aligned}$$

With $g_Y(\xi)$, it follows that the characteristic function of Z_θ is:

$$\begin{aligned}
h_\theta(\xi) &= E\left(e^{i\xi(Y-2\theta)/\sqrt{8\theta}}\right) = E\left(e^{i(\xi/\sqrt{8\theta})Y} \cdot e^{-\sqrt{\theta/2}}\right) \\
&= E\left(e^{i(\xi/\sqrt{8\theta})Y}\right) \cdot e^{-\sqrt{\theta/2}} \\
&= g_Y(\xi/\sqrt{8\theta}) \cdot e^{-\sqrt{\theta/2}} \\
&= \exp\left(\frac{2i\xi\theta}{\sqrt{8\theta} - 2i\xi} - \frac{\sqrt{\theta}}{\sqrt{2}}\right) \\
&= \exp\left(\frac{-\xi^2\sqrt{2\theta}}{\sqrt{8\theta} - 2i\xi}\right).
\end{aligned}$$

Since $\theta \in \mathbb{R}$, we can see that $\sqrt{8\theta} - 2i\xi \neq 0$ for all $\xi \in \mathbb{R}$, which implies $h_\theta(\xi)$ is continuous with respect to ξ . Then, we show that for every $\xi \in \mathbb{R}$, $h_\theta(\xi)$ converges as θ goes to $+\infty$. For any fixed ξ , it follows that

$$\lim_{\theta \rightarrow +\infty} \frac{-\xi^2\sqrt{2\theta}}{\sqrt{8\theta} - 2i\xi} = \frac{-\xi^2}{2}.$$

From the continuity of the function e^z on the complex plane, we can see that:

$$\lim_{\theta \rightarrow +\infty} h_\theta(\xi) = e^{-\xi^2/2}, \quad \forall \xi \in \mathbb{R}.$$

Note that RHS of the above equation is the characteristic function of standard normal distribution. This completes the proof.

The code for simulation is attached below:

```

set.seed(133)
m = c(1e2, 1e3, 1e4, 1e5)
theta = c(1, 1e1, 1e2, 1e3)
Color.name = c('black', 'blue', 'purple', 'green3')

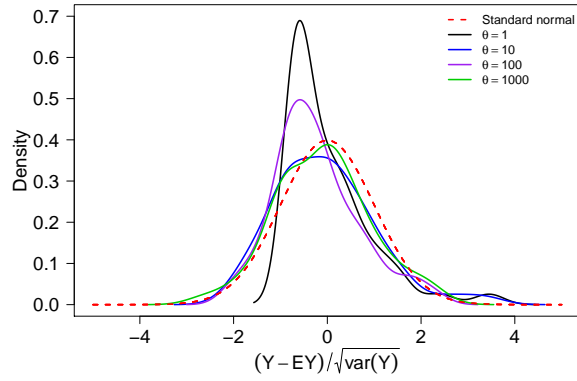
for(i in 1:4){
  pdf(paste0('GP', i, '.pdf'), 8, 6)
  par(cex.axis=1.5, cex.lab=1.5, las=1)
  plot(0, 0, col='white', main = NULL, ylab = 'Density',
        xlab = expression((Y-EY)/sqrt(var(Y))),
        xlim = c(-5, 5), ylim = c(0, 0.7))
  legend('topright', legend = c('Standard normal', expression(theta==1),
                                expression(theta==10), expression(theta==100), expression(theta==1000)),
        lwd=2, pt.cex=1.2, cex=1, lty = c(2, 1, 1, 1), col=c('red', Color.name), y.intersp = 1, bty = 'n')

  for (j in 1:4){
    rnd.N = rpois(m[i], theta[j])
    rnd.Y = rchisq(m[i], 2*rnd.N)
    Z = (rnd.Y - 2*theta[j])/sqrt(8*theta[j])
    lines(density(Z), col = Color.name[j], lwd=2)
    lines(seq(-5, 5, 0.1), dnorm(seq(-5, 5, 0.1)), col='red', lty = 2, lwd=2)
  }

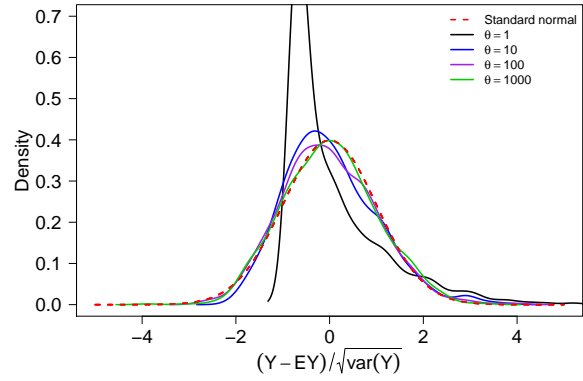
  dev.off()
}

```

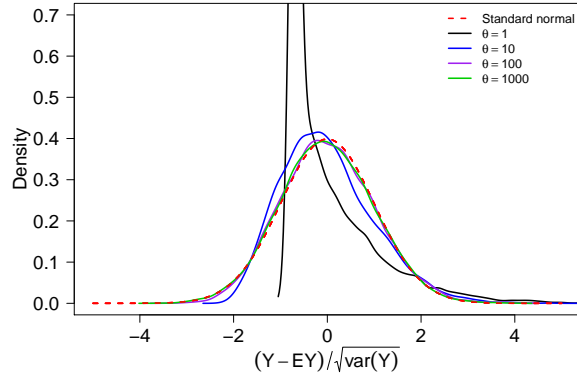

Figure 1: Asymptotic Normality with different Monte Carlo sample sizes m and parameter θ . Each panel show with a fixed m with increasing θ .



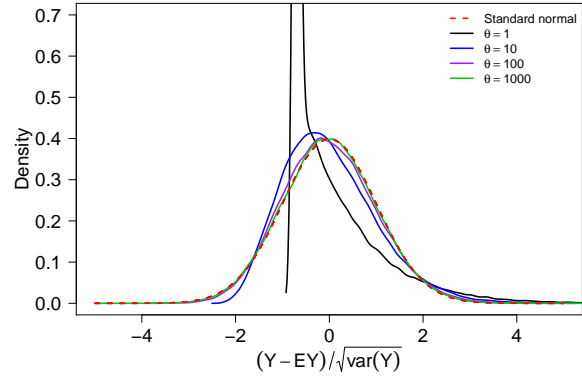
(a) Sample size $m = 100$



(b) Sample size $m = 1000$



(c) Sample size $m = 10000$



(d) Sample size $m = 100000$

Figure 1 below shows how the empirical distributions trend toward a standard normal distribution as θ goes to infinity. For a fixed m , larger θ yields a smoother density function that is closer to the limiting standard normal. With $\theta = 1$, the empirical distribution is very different from a standard normal; Starting from $\theta = 10$ the empirical distribution is pretty close to a standard normal. From the four panels, we can also see that with a larger sample m , the empirical distribution approximates standard normal distribution better as θ grows larger. When using $m = 100$ Monte Carlo samples, the fitted density is still bumpy, and it looks far different from a standard normal. With larger m , the fitted density approximates the limiting distribution better.

4 Gaussian Example

4.1 When ρ is known

Theory

When ρ and $\sigma_X = \sigma_Y = \sigma$ are known, there are only three parameters (μ_X, μ_Y, σ) . Then the likelihood is (since ρ is known we can drop ρ in deriving the likelihood):

$$\begin{aligned}\mathcal{L}(\mu_X, \mu_Y, \sigma \mid \mathbf{X}, \mathbf{Y}) &= \prod_{i=1}^n \left[\frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \exp \left(-\frac{1}{2\sigma^2(1-\rho^2)} [(X_i - \mu_X)^2 + (Y_i - \mu_Y)^2 \right. \right. \\ &\quad \left. \left. - 2\rho(X_i - \mu_X)(Y_i - \mu_Y)] \right) \right] \\ &\propto \sigma^{-2n} \exp \left(-\frac{1}{2\sigma^2(1-\rho^2)} \sum_{i=1}^n [(X_i - \mu_X)^2 + (Y_i - \mu_Y)^2 \right. \\ &\quad \left. - 2\rho(X_i - \mu_X)(Y_i - \mu_Y)] \right).\end{aligned}$$

Taking derivative of $\log(\mathcal{L})$ with respect to μ_X , μ_Y and σ , we have:

$$\begin{aligned}\frac{\partial \log(\mathcal{L})}{\partial \mu_X} &\propto -\frac{1}{2\sigma^2(1-\rho^2)} \sum_{i=1}^n [2(\mu_X - X_i) - 2\rho(\mu_Y - Y_i)], \\ \frac{\partial \log(\mathcal{L})}{\partial \mu_Y} &\propto -\frac{1}{2\sigma^2(1-\rho^2)} \sum_{i=1}^n [2(\mu_Y - Y_i) - 2\rho(\mu_X - X_i)], \\ \frac{\partial \log(\mathcal{L})}{\partial \sigma} &\propto \frac{-2n}{\sigma} + \frac{1}{\sigma^3(1-\rho^2)} \sum_{i=1}^n [(X_i - \mu_X)^2 + (Y_i - \mu_Y)^2 - 2\rho(X_i - \mu_X)(Y_i - \mu_Y)].\end{aligned}$$

Denote \bar{X} , \bar{Y} as the sample mean of X_i and Y_i . Let all the derivatives above equal to 0, the MLE estimate of (μ_X, μ_Y, σ) satisfy:

$$\begin{aligned}\hat{\mu}_X - \rho\hat{\mu}_Y &= \bar{X} - \rho\bar{Y}, \\ \hat{\mu}_Y - \rho\hat{\mu}_X &= \bar{Y} - \rho\bar{X}, \\ \frac{1}{\hat{\sigma}^2(1-\rho^2)} \sum_{i=1}^n [(X_i - \hat{\mu}_X)^2 + (Y_i - \hat{\mu}_Y)^2 - 2\rho(X_i - \hat{\mu}_X)(Y_i - \hat{\mu}_Y)] &= 2n.\end{aligned}$$

Denote $S_{WV} = \sum_{i=1}^n (W_i - \bar{W})(V_i - \bar{V})$ as the sample covariance between W and V . Solving the above system gives the MLE estimators:

$$\begin{aligned}\hat{\mu}_X &= \bar{X} \\ \hat{\mu}_Y &= \bar{Y} \\ \hat{\sigma}^2 &= \frac{[S_{XX} + S_{YY} - 2\rho S_{XY}]}{2n(1-\rho^2)}\end{aligned}$$

Simulation

The R code is attached:

```
Inv.quadform <- function(X, Sigma){
  Y = t(X) %*% solve(Sigma, X)
  return(Y)
}

set.seed(133)
library(MASS)

N = c(20,100,500,2000)
M = 5000
sigma.X = 1
sigma.Y = 1
rho = 0.3
Sigma = matrix( c(sigma.X^2, sigma.X*sigma.Y*rho,
                  sigma.X*sigma.Y*rho, sigma.Y^2), 2, 2 )

mu.X = 10
mu.Y = 10
Fin.Width.Fieller = rep(NA,length(N))
Fin.Width.Delta = Fin.Width.Fieller
Fin.Cover.Delta = Fin.Width.Fieller
Fin.Cover.Fieller = Fin.Width.Fieller

for(j in 1:length(N)){ ##Vary the sample size n
  Cover.Fieller = 0
  Cover.Delta = 0
  Width.Fieller = 0
  Width.Delta = 0
  n = N[j]
  for(i in 1:M){## Simulation starts
    #####
    ##Generating Data
    Data = mvrnorm(n, mu = c(mu.X, mu.Y), Sigma)

    #####
    ## Estimating Delta method
    mu.hat = apply(Data, 2, mean)
    Data.centered = apply(Data, 1, '-', mu.hat )
    sigma2.hat = sum( apply(Data.centered, 2, Inv.quadform, Sigma) )/(2*n)
    Sigma.hat = matrix( c(sigma2.hat/n, rho*sigma2.hat/n, rho*sigma2.hat/n,
                        sigma2.hat/n), 2, 2 )
    temp.vec = c(-mu.hat[2]/mu.hat[1]^2, 1/mu.hat[1])
    Delta.var.est = t(temp.vec) %*% Sigma.hat %*% temp.vec
    Delta.CI = mu.hat[2]/mu.hat[1] + c( -qnorm(0.975) * sqrt(Delta.var.est),
                                     qnorm(0.975) * sqrt(Delta.var.est) )

    ## Estimating Fieller's method
    a = mu.hat[2]
    b = mu.hat[1]
    t.quant = qt(0.975, n-2)
    f1 = a*b - t.quant^2 * rho * sigma2.hat / n
```

```

f2 = b^2 - t.quant^2 * sigma2.hat / n
f0 = a^2 - t.quant^2 * sigma2.hat / n
D = f1^2 - f0*f2
Fieller.CI = c((f1 - sqrt(D)) / f2, (f1 + sqrt(D)) / f2)

## Evaluate the width and coverage rate
Width.Fieller = Width.Fieller + diff(Fieller.CI)
Width.Delta = Width.Delta + diff(Delta.CI)

if(Fieller.CI[1]<=1 && Fieller.CI[2]>=1) Cover.Fieller = Cover.Fieller + 1
if(Delta.CI[1]<=1 && Fieller.CI[2]>=1) Cover.Delta = Cover.Delta + 1
}

Fin.Width.Fieller[j] = Width.Fieller / M
Fin.Width.Delta[j] = Width.Delta / M

Fin.Cover.Fieller[j] = Cover.Fieller / M
Fin.Cover.Delta[j] = Cover.Delta / M
}
pdf('Gaussian_rates.pdf',8,6)
opar= par()
par(cex.axis=1.5,cex.lab=1.5,las=1,mar=c(5.1,6.1,4.1,1.1))
plot(Fin.Cover.Delta*100~N, type= 'b', ylim=c(94,96), pch = 4,
      xlab = 'Sample size',ylab='',log='x')
title( ylab = 'Coverage rates (%)',line=4.5)
points(Fin.Cover.Fieller*100~N, type='b', pch = 5,col='blue')
abline(h=95, col = 'red',lty=2)
legend('topright',legend = c('Delta','Fieller','Nominal level'),lwd=2,pt.cex=1.2,cex=1,
      lty = c(1,1,2), pch=c(4,5,-1),col=c('black','blue','red'),y.intersp =1)
dev.off()

## pdf
## 2

```

I programmed the Delta method from scratch, while the Fieller's method follows from: Fieller, E. C. (1954). Some problems in interval estimation. Figure 2 and 3 show the results for the simulations with $n = 20, 100, 500, 2000$. Each set of simulation consists 5000 Monte Carlo samples. We can see that both methods approximate a valid confidence with coverage rates of 95%. As the sample size n goes large, the results of these two method agree. Also, Fieller's method gives a slightly wider confidence interval than Delta method. This shows that the two methods might be asymptotically equivalent when the Data come from a normal distribution.

4.2 When ρ is unknown

Theory

Denote $\mathbf{X} = (X, Y)^T$, $\boldsymbol{\mu} = (\mu_X, \mu_Y)^T$ and $\boldsymbol{\Sigma}$ as the variance-covariance matrix of \mathbf{X} . We can formulate the linear transformation in matrix form. Since (X, Y) follows a bivariate Gaussian distribution, the linear

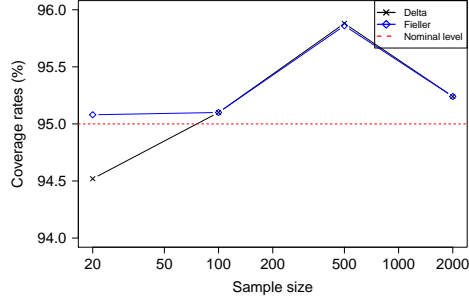


Figure 2: Empirical coverage rates of the 95% C.I. The black dot shows Fiellers method, while the blue cross shows Delta method. The red dashed line shows the nominal level.

Figure 3: Average width of the C.I.

Samples	Delta method	Fieller's method
20	1.0038×10^{-1}	1.0775×10^{-1}
100	4.6102×10^{-2}	4.6694×10^{-2}
500	2.0730×10^{-2}	2.0782×10^{-2}
2000	1.0366×10^{-2}	1.0373×10^{-2}

transformation (W, Z) also follows a Gaussian distribution.

$$\begin{aligned}
\begin{pmatrix} W \\ Z \end{pmatrix} &= \begin{pmatrix} \frac{1}{\sigma_X} & -\frac{\rho}{\sigma_Y} \\ 0 & \frac{1}{\sigma_Y} \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} + \begin{pmatrix} \frac{\rho\mu_Y}{\sigma_Y} - \frac{\mu_X}{\sigma_X} \\ -\frac{\mu_Y}{\sigma_Y} \end{pmatrix} \\
&:= \mathbf{A} \begin{pmatrix} X \\ Y \end{pmatrix} + \boldsymbol{\eta} \\
&\sim \mathcal{N}(\mathbf{A}\boldsymbol{\mu} + \boldsymbol{\eta}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T) \\
&\sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} 1 - \rho^2 & 0 \\ 0 & 1 \end{bmatrix}\right).
\end{aligned}$$

First, we can derive the MLE for (μ_X, μ_Y) . This time, the likelihood is more complicated:

$$\begin{aligned}
\mathcal{L}(\mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho \mid \mathbf{X}, \mathbf{Y}) &\propto \sigma_X^{-n} \sigma_Y^{-n} \exp \left(-\frac{1}{2(1-\rho^2)} \sum_{i=1}^n \left[\frac{(X_i - \mu_X)^2}{\sigma_X^2} + \frac{(Y_i - \mu_Y)^2}{\sigma_Y^2} \right. \right. \\
&\quad \left. \left. - \frac{2\rho}{\sigma_X \sigma_Y} (X_i - \mu_X)(Y_i - \mu_Y) \right] \right).
\end{aligned}$$

It is sufficient to take derivatives of the log-likelihood function $\log(\mathcal{L})$ only with respect to μ_X and μ_Y . We have:

$$\begin{aligned}
\frac{\partial \log(\mathcal{L})}{\partial \mu_X} &\propto -\frac{1}{2(1-\rho^2)} \sum_{i=1}^n \left[\frac{2(\mu_X - X_i)}{\sigma_X^2} - \frac{2\rho}{\sigma_X \sigma_Y} (\mu_Y - Y_i) \right], \\
\frac{\partial \log(\mathcal{L})}{\partial \mu_Y} &\propto -\frac{1}{2(1-\rho^2)} \sum_{i=1}^n \left[\frac{2(\mu_Y - Y_i)}{\sigma_Y^2} - \frac{2\rho}{\sigma_X \sigma_Y} (\mu_X - X_i) \right].
\end{aligned}$$

Letting all these derivatives to zero, we can obtain the MLE for (μ_X, μ_Y) as:

$$\hat{\mu}_X = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \hat{\mu}_Y = \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i.$$

Second, we derive the MLE for (σ_X, σ_Y) . Taking derivatives of the likelihood with respect to (σ_X, σ_Y) , and substituting $(\hat{\mu}_X, \hat{\mu}_Y)$ as their MLE, we have:

$$\begin{aligned}\frac{\partial \log \mathcal{L}}{\partial \sigma_X} &\propto -n\sigma_X^{-1} - \frac{1}{2(1-\rho^2)} [-2\sigma_X^{-3}S_{XX} + 2\rho\sigma_X^{-2}\sigma_Y^{-1}S_{XY}], \\ \frac{\partial \log \mathcal{L}}{\partial \sigma_Y} &\propto -n\sigma_Y^{-1} - \frac{1}{2(1-\rho^2)} [-2\sigma_Y^{-3}S_{YY} + 2\rho\sigma_Y^{-2}\sigma_X^{-1}S_{XY}].\end{aligned}$$

Setting the above equations to zero. We can see that $(\hat{\sigma}_X, \hat{\sigma}_Y)$ satisfy (Since $\rho \in (-1, 1)$):

$$\hat{\sigma}_X^2 = \frac{S_{XX} - \hat{\rho}S_{XY}\sqrt{S_{XX}/S_{YY}}}{n(1-\hat{\rho}^2)}, \quad \hat{\sigma}_Y^2 = \frac{S_{YY} - \hat{\rho}S_{XY}\sqrt{S_{YY}/S_{XX}}}{n(1-\hat{\rho}^2)}. \quad (1)$$

Then, instead of deriving the MLE from the likelihood above, we take a look at the likelihood function of (W_i, Z_i) . We shall see that the MLE for ρ also maximizes the likelihood for (W_i, Z_i) . From the joint distribution of the Gaussian random vector (W, Z) , we have:

$$\begin{aligned}1 - \hat{\rho}^2 &= \frac{1}{n} \sum_{i=1}^n (W_i - \bar{W})^2 \\ &= \frac{1}{n} \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\hat{\sigma}_X} - \hat{\rho} \frac{Y_i - \bar{Y}}{\hat{\sigma}_Y} \right)^2 \\ &= \frac{1}{n} \left[\frac{S_{XX}}{\hat{\sigma}_X^2} + \frac{\hat{\rho}^2 S_{YY}}{\hat{\sigma}_Y^2} - \frac{2\hat{\rho}S_{XY}}{\hat{\sigma}_X \hat{\sigma}_Y} \right] \\ &= \frac{1 - \hat{\rho}^2}{1 - \hat{\rho}S_{XY}/\sqrt{S_{XX}S_{YY}}} \left[1 + \hat{\rho}^2 - \frac{2\hat{\rho}S_{XY}}{\sqrt{S_{XX}S_{YY}}} \right].\end{aligned}$$

Thus we can obtain the MLE for ρ :

$$\hat{\rho} = \frac{S_{XY}}{\sqrt{S_{XX}S_{YY}}} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}}.$$

Since

$$\begin{aligned}S_{XX} &= \sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n [(X_i - \mu_X)^2 + (\bar{X} - \mu_X)^2 + 2(X_i - \mu_X)(\mu_X - \bar{X})] \\ &= \sum_{i=1}^n (X_i - \mu_X)^2 + 2n\mu_X \bar{X} - n\mu_X^2 - n(\bar{X})^2.\end{aligned}$$

By the Law of Large Numbers, we can see that $S_{XX}/n \xrightarrow{a.s.} \sigma_X^2$ as n goes to infinity. Similarly we know that $S_{YY}/n \xrightarrow{a.s.} \sigma_Y^2$ and $S_{XY}/n \xrightarrow{a.s.} \rho\sigma_X\sigma_Y$. Thus, we know immediately that $\hat{\rho} \xrightarrow{a.s.} \rho$ as n goes to infinity.

Simulation

The code for this part is attached below:

```
set.seed(133)
library(MASS)

N = c(20, 100, 1000, 1e4)
M = 5000
sigma.X = 1
```

```

sigma.Y = 2
val.rho = c(0.15,0.95)
mu.X = 10
mu.Y = 5

Est.rho = matrix(NA,length(N),2)
CI.rho = matrix(NA,length(N),4)

for(k in 1:2){
  rho = val.rho[k]
  Sigma = matrix( c(sigma.X^2, sigma.X*sigma.Y*rho,
                    sigma.X*sigma.Y*rho, sigma.Y^2), 2, 2 )
  for(j in 1:length(N)){ ##Vary the sample size n
    n = N[j]
    rho.hat = 0
    for(i in 1:M){## Simulation starts
      #####
      ##Generating Data
      Data = mvrnorm(n, mu = c(mu.X, mu.Y), Sigma)

      #####
      ## Estimating
      Est.Sigma = cov(Data)
      S.XX = Est.Sigma[1,1]*(n-1)/n
      S.YY = Est.Sigma[2,2]*(n-1)/n
      S.XY = Est.Sigma[1,2]*(n-1)/n

      rho.hat[i] = S.XY/sqrt(S.XX*S.YY)
    }
    Est.rho[j,k] = mean(rho.hat)
    CI.rho[j,c(2*k-1,k*2)] = quantile(rho.hat, c(0.025,0.975) )
  }
}

#### Visualizing the result
## Plot the confidence intervals
opar = par()
par(cex.axis=1.5,cex.lab=1.5,las=1)
plot(Est.rho[,1]~N,
     main='Estimated values \n and 95% confidence intervals of the estimated mean.',
     xlab = 'Sample size', ylab = 'Estimated coefficient',pch=4, cex=0.8,ylim=c(-0.36,1),log='x')
abline(h=val.rho[1],lty = 2)
segments(x0 = N,y0=(CI.rho[,1]),y1=CI.rho[,2])
segments(x0 = N*1.1,y0 = CI.rho[,1], x1=N/1.1)
segments(x0 = N*1.1,y0 = CI.rho[,2], x1=N/1.1)

points(Est.rho[,2]~N,pch=4, cex=0.8,col='purple')
abline(h=val.rho[2],col='purple',lty = 2)
segments(x0 = N,y0=(CI.rho[,3]),y1=CI.rho[,4],col='purple')
segments(x0 = N*1.1,y0 = CI.rho[,3], x1=N/1.1,col='purple')
segments(x0 = N*1.1,y0 = CI.rho[,4], x1=N/1.1,col='purple')

```

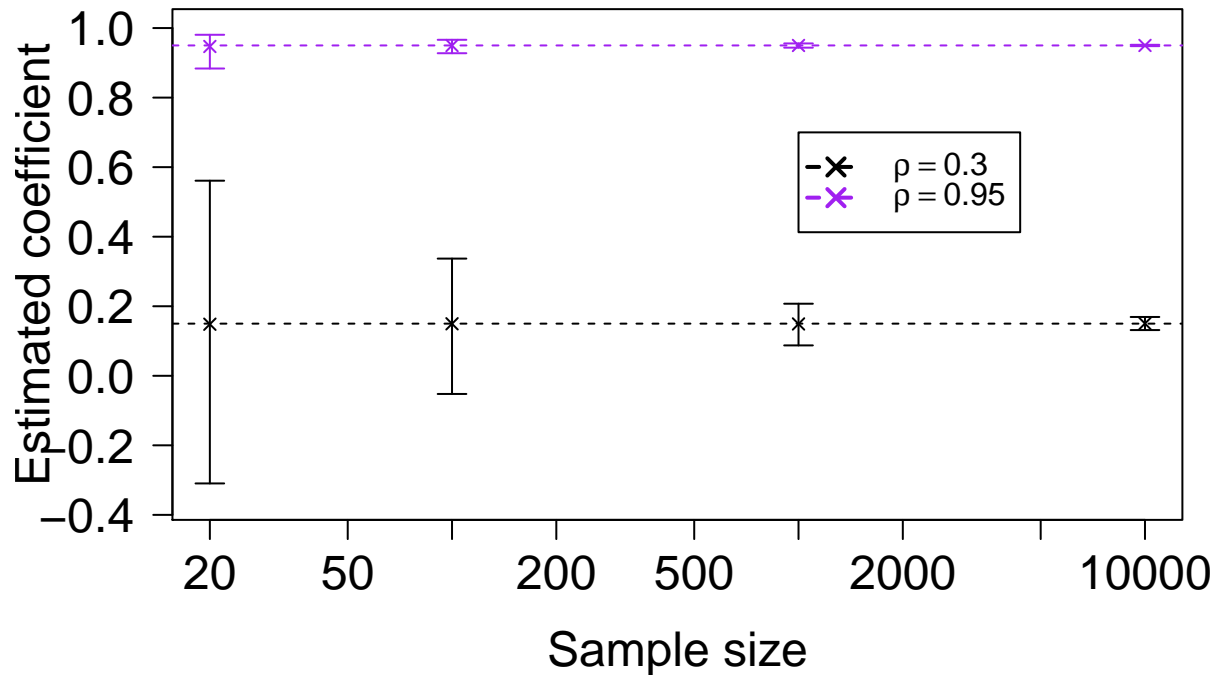


```

legend(x=xy.coords(1000,0.7),legend = c(expression(rho==0.3),expression(rho==0.95)),
      lwd=2,pt.cex=1.2,cex=, lty = c(2,2), pch=c(4,4),col=c('black','purple'),y.intersp =0.8)

```

Estimated values and 95% confidence intervals of the estimated mean.



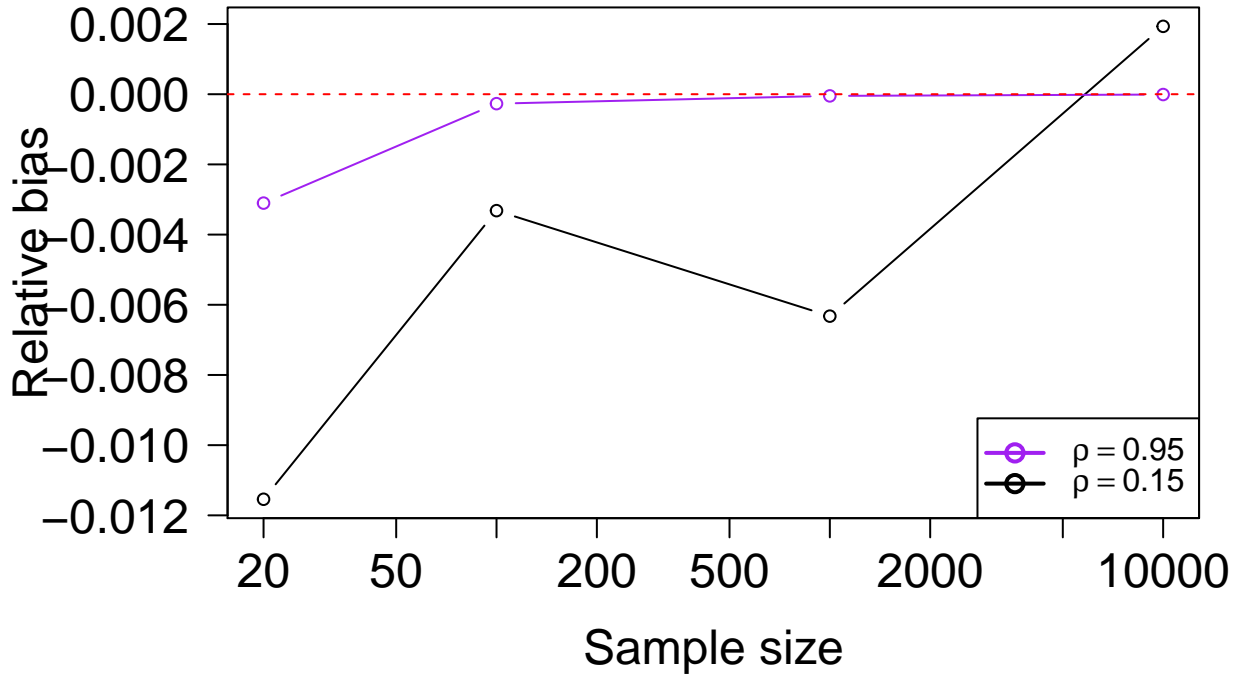
```

## Plot the relative bias
par(mar=c(5.1,6.1,4.1,1.1))
plot((Est.rho[,1]-val.rho[1])/val.rho[1]~N,log='x',main = 'The relative bias of the estimator.' ,xlab =
title(ylab='Relative bias',line = 4.5)
lines((Est.rho[,2]-val.rho[2])/val.rho[2]~N,xlab = 'log sample size', ylab = 'Relative bias', cex = 0.8)
abline(h=0,col='red', lty = 2)

legend('bottomright',legend = c(expression(rho==0.95), expression(rho==0.15)),lwd=2,pt.cex=1.2,cex=1, l

```

The relative bias of the estimator.



We set the means $\mu_X = 10$, $\mu_Y = 5$; The variances $\sigma_X^2 = 1$, $\sigma_Y^2 = 4$. We conduct simulations in two scenarios, corresponding to different values of $\rho = 0.3$ and $\rho = 0.95$. The figures above show the results for our simulations. We conduct four set of simulations with sample sizes 20, 100, 1000, 10000. For each set of simulation, 5000 Monte Carlo samples are generated to calculate the confidence interval for ρ .

We can see that as n grows larger, the MLE $\hat{\rho}$ converges toward the true correlation coefficient ρ in both cases. When $\rho = 0.95$, the convergence occurs much more rapid. We can see this from both the width of the confidence intervals and the relative bias. This show that the MLE performs better when there is a strong correlation between X and Y .