Causal Contrasts from Experimental Study Data

In a randomized experimental study, inference concerning the causal effect of binary variable Z on Y can be made by direct comparison of sample averages. Suppose that $Z \sim Bernoulli(p)$ for $0 < p < 1$. Two estimators of the causal contrast

$$\delta = \mathbb{E}[Y(1) - Y(0)]$$

can be proposed:

$$\widetilde{\delta} = \frac{1}{np}\sum_{i=1}^{n} Z_i Y_i - \frac{1}{n(1-p)}\sum_{i=1}^{n}(1-Z_i)Y_i = \frac{1}{np(1-p)}\sum_{i=1}^{n}(Z_i - p)Y_i \tag{1}$$

and

$$\widehat{\delta} = \frac{1}{n\widehat{p}}\sum_{i=1}^{n} Z_i Y_i - \frac{1}{n(1-\widehat{p})}\sum_{i=1}^{n}(1-Z_i)Y_i = \frac{1}{n\widehat{p}(1-\widehat{p})}\sum_{i=1}^{n}(Z_i - \widehat{p})Y_i \tag{2}$$

where

$$\widehat{p} = \frac{1}{n}\sum_{i=1}^{n} Z_i.$$

It follows that (2) has lower variance than (1).

(1) The variance of $\widetilde{\delta}$ is computed as follows: we have that

$$\widetilde{\delta} = \frac{1}{np(1-p)}\sum_{i=1}^{n}(Z_i - p)Y_i$$

which has variance

$$\frac{1}{np^2(1-p)^2}\text{Var}[(Z-p)Y].$$

In the experimental study, $Z \perp\!\!\!\perp X$.

$$\text{Var}[(Z-p)Y] = \mathbb{E}_Z[(Z-p)^2 \text{Var}_{Y|Z}[Y|Z]] + \text{Var}_Z[(Z-p)\mathbb{E}_{Y|Z}[Y|Z]].$$

We have that

$$\mathbb{E}_{Y|Z}[Y|Z] = \mathbb{E}_X[\mathbb{E}_{Y|X,Z}[Y|X,Z]] = \mathbb{E}_X[\mu(X,Z)] = \mu(Z)$$

say. Note that

$$\mathbb{E}_Z[(Z-p)\mu(Z)] = -p(1-p)\mu(0) + p(1-p)\mu(1) = p(1-p)(\mu(1)-\mu(0)) = p(1-p)\delta.$$

If $\text{Var}_{Y|X,Z}[Y|X,Z] = \sigma^2(X,Z)$, we also have that

$$\begin{aligned}\text{Var}_{Y|Z}[Y|Z] &= \mathbb{E}_X[\text{Var}_{Y|X,Z}[Y|X,Z]] + \text{Var}_X[\mathbb{E}_{Y|X,Z}[Y|X,Z]] \\ &= \mathbb{E}_X[\sigma^2(X,Z)] + \text{Var}_X[\mu(X,Z)] = \sigma^2(Z)\end{aligned}$$

say. Therefore

$$\begin{aligned}\text{Var}[(Z-p)Y] &= \mathbb{E}_Z[(Z-p)^2\sigma^2(Z)] + \text{Var}_Z[(Z-p)\mu(Z)] \\ &= \left\{p^2(1-p)\sigma^2(0) + p(1-p)^2\sigma^2(1)\right\} + \left\{\mathbb{E}_Z[(Z-p)^2\mu(Z)^2] - \{\mathbb{E}_Z[(Z-p)\mu(Z)]\}^2\right\} \\ &= p(1-p)\left\{p\sigma^2(0) + (1-p)\sigma^2(1)\right\} \\ &\qquad + p^2(1-p)\{\mu(0)\}^2 + p(1-p)^2\{\mu(0)\}^2 - p^2(1-p)^2(\mu(1)-\mu(0))^2 \\ &= p(1-p)\left\{p\sigma^2(0) + (1-p)\sigma^2(1)\right\} + p(1-p)(p\mu(0) + (1-p)\mu(1))^2\end{aligned}$$

Therefore

$$\text{Var}[\widetilde{\delta}] = \frac{1}{np(1-p)}\left(p\sigma^2(0) + (1-p)\sigma^2(1) + (p\mu(0) + (1-p)\mu(1))^2\right).$$

(2) The variance of $\widehat{\delta}$

$$\widehat{\delta} = \frac{1}{n\widehat{p}(1-\widehat{p})}\sum_{i=1}^{n}(Z_i - \widehat{p})Y_i$$

can be computed by noting that the estimator arises as the WLS estimator of δ from the model

$$Y_i = \beta_0 + Z_i\delta + \varepsilon_i$$

where

$$\text{Var}_{\varepsilon|Z}[\varepsilon|Z = z] = \sigma^2(z).$$

Under homoscedasticity, $\sigma^2(z) \equiv \sigma^2$.

For large n, the variance covariance matrix of the WLS estimators is known to be the inverse of the expectation of

$$\frac{n}{\sigma^2(Z)}\begin{bmatrix} 1 & Z \\ Z & Z^2 \end{bmatrix}.$$

Taking expectations (elementwise) we have

$$n\begin{bmatrix} \dfrac{(1-p)}{\sigma^2(0)} + \dfrac{p}{\sigma^2(1)} & \dfrac{p}{\sigma^2(1)} \\ \dfrac{p}{\sigma^2(1)} & \dfrac{p}{\sigma^2(1)} \end{bmatrix}$$

which has inverse

$$\frac{1}{n}\frac{\sigma^2(0)\sigma^2(1)}{p(1-p)}\begin{bmatrix} \dfrac{p}{\sigma^2(1)} & -\dfrac{p}{\sigma^2(1)} \\ -\dfrac{p}{\sigma^2(1)} & \dfrac{(1-p)}{\sigma^2(0)} + \dfrac{p}{\sigma^2(1)} \end{bmatrix}$$

Picking off the element (2,2) element of this matrix, we find that for large n

$$\text{Var}[\widehat{\delta}] \doteq \frac{1}{np(1-p)}(p\sigma^2(0) + (1-p)\sigma^2(1)).$$

We therefore have from (1) that

$$\text{Var}[\widetilde{\delta}] = \frac{1}{np(1-p)}\left(p\sigma^2(0) + (1-p)\sigma^2(1) + (p\mu(0) + (1-p)\mu(1))^2\right) \geq \text{Var}[\widehat{\delta}].$$

Most commonly we will assume homoscedasticity where $\sigma^2(z) \equiv \sigma^2$, in which case

$$\text{Var}[\widetilde{\delta}] = \frac{1}{np(1-p)}\left(\sigma^2 + (p\mu(0) + (1-p)\mu(1))^2\right)$$

$$\text{Var}[\widehat{\delta}] = \frac{1}{np(1-p)}\sigma^2$$

Simulation 1: In the simulation below, we assume that

$$X \sim Normal(\mu_X, \sigma_X^2)$$

with $\mu_X = 1$ and $\sigma_X = 1$, and

$$Y|X = x, Z = z \sim Normal(\alpha x + \delta z, \sigma^2)$$

with $\sigma^2(x, z) \equiv \sigma^2 = 1$, and

$$\mu(x, z) = \mathbb{E}_{Y|X,Z}[Y|X = x, Z = z] = \alpha x + \delta z$$

with $\alpha = 0.5$ and $\delta = 2$, and hence

$$\mathbb{E}[Y(1) - Y(0)] = \mathbb{E}_{Y|X,Z}[Y|X = x, Z = 1] - \mathbb{E}_{Y|X,Z}[Y|X = x, Z = 0] = \delta.$$

We also have

$$\mu(z) = \mathbb{E}_X[\mu(X, z)] = \mathbb{E}_X[\alpha X + \delta z] = \alpha\mathbb{E}_X[X] + \delta z = \alpha\mu_X + \delta z$$

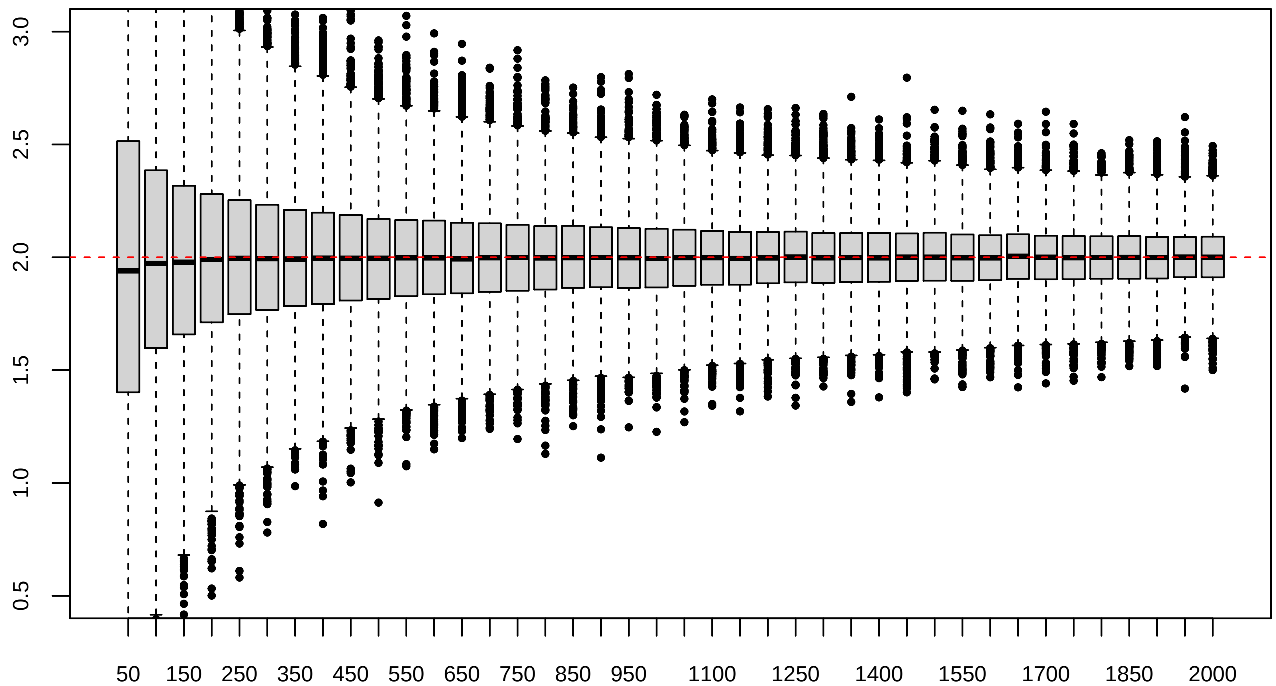
and that

$$\begin{aligned}\sigma^2(z) &= \mathbb{E}_X[\sigma^2(X, z)] + \text{Var}_X[\mu(X, z)]\\ &= \sigma^2 + \text{Var}_X[\alpha X + \delta z]\\ &= \sigma^2 + \alpha^2\text{Var}_X[X]\\ &= \sigma^2 + \alpha^2\sigma_X^2\end{aligned}$$

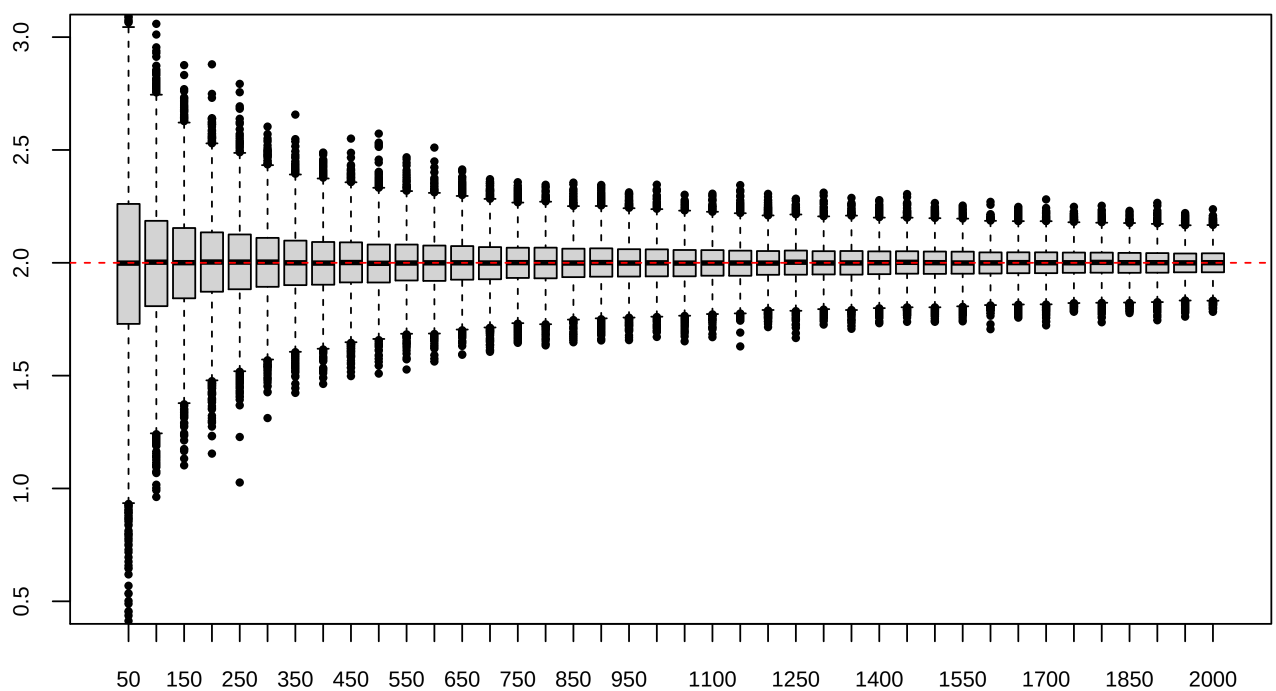
We study how the variances change as n increases. We choose $p = 0.2$.

```
set.seed(23)
nreps<-10000;nvec<-seq(50,2000,by=50)
p<-0.2;delta<-2;al<-0.5;sig<-1
muX<-1;sigX<-1
mu0<-al*muX; mu1<-mu0+delta
ests.mat<-array(0,c(length(nvec),nreps,2))
for(i in 1:length(nvec)){
    n<-nvec[i]
    for(irep in 1:nreps){
        X<-rnorm(n,muX,sigX)
        Z<-rbinom(n,1,p)
        Y<-rnorm(n,delta*Z+al*X,sig)
        p.hat<-mean(Z)
        ests.mat[i,irep,1]<-sum(Z*Y)/(n*p)-sum((1-Z)*Y)/(n*(1-p))
        ests.mat[i,irep,2]<-sum(Z*Y)/(n*p.hat)-sum((1-Z)*Y)/(n*(1-p.hat))
    }
}
#Variance of delta tilde
true.var.tilde<-((sig^2+al^2*sigX^2+(p*mu0+(1-p)*mu1)^2)/(p*(1-p)))
true.var.tilde

+ [1] 35.375

#Variance of delta hat
true.var.hat<-(sig^2+al^2*sigX^2)/(p*(1-p))
true.var.hat

+ [1] 7.8125

#Efficiency
true.var.tilde/true.var.hat

+ [1] 4.528
```

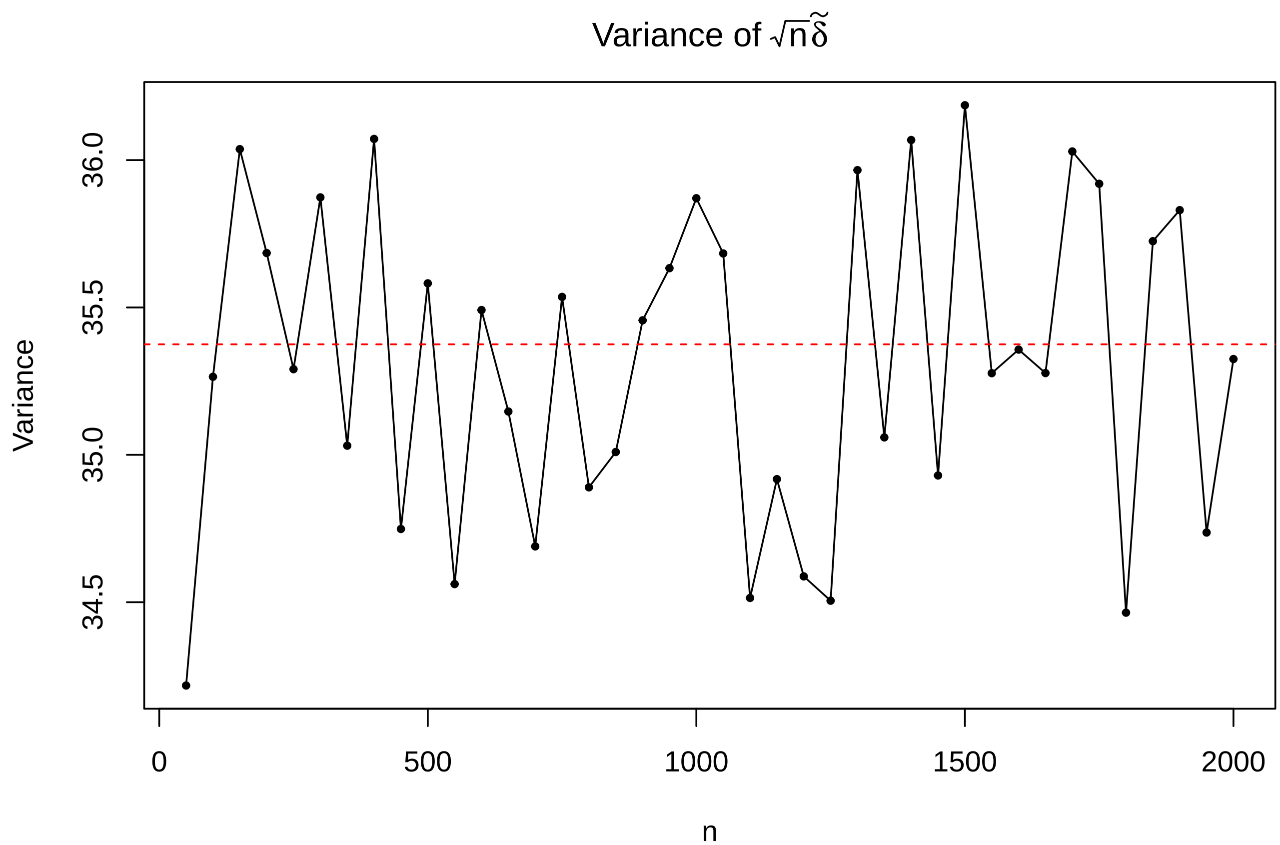
Variability of $\tilde{\hat{\delta}}$



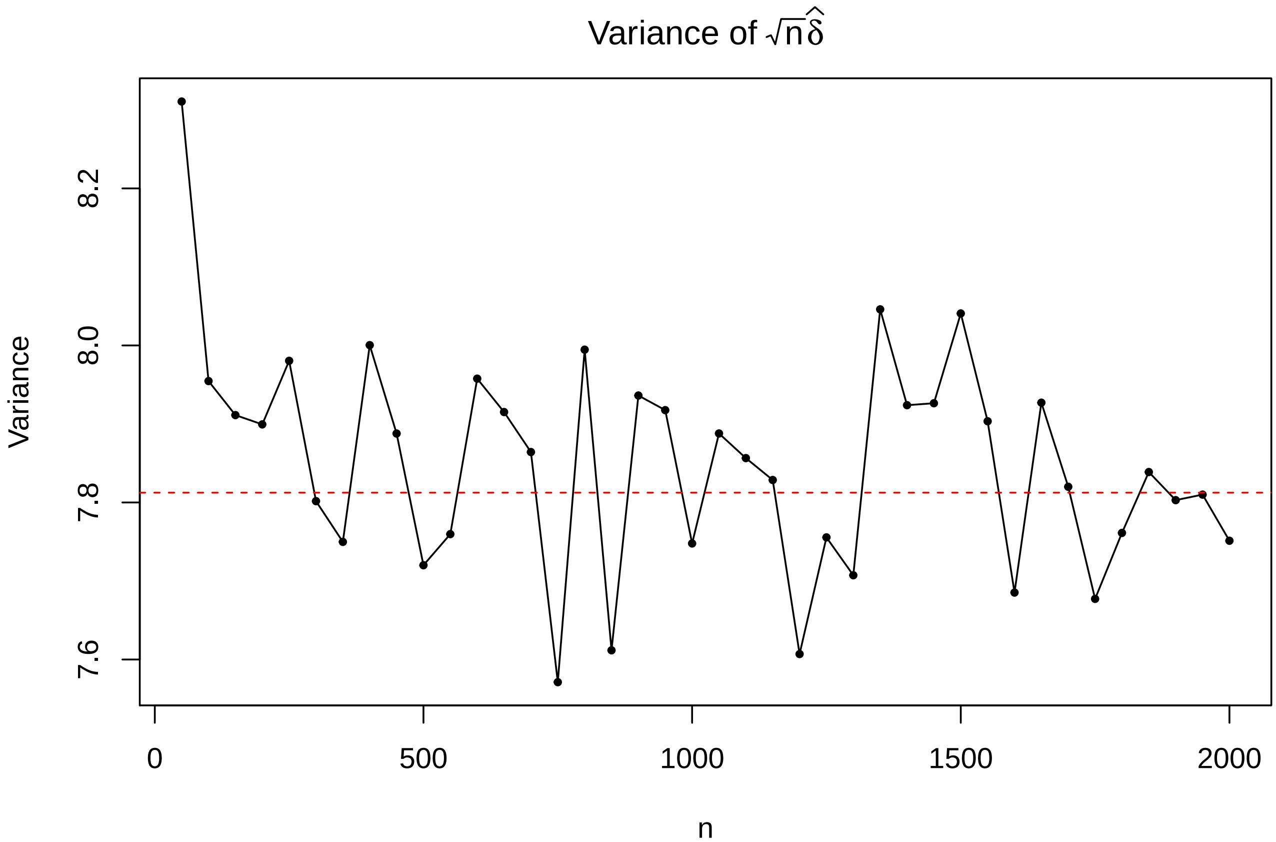
Variability of $\hat{\hat{\delta}}$

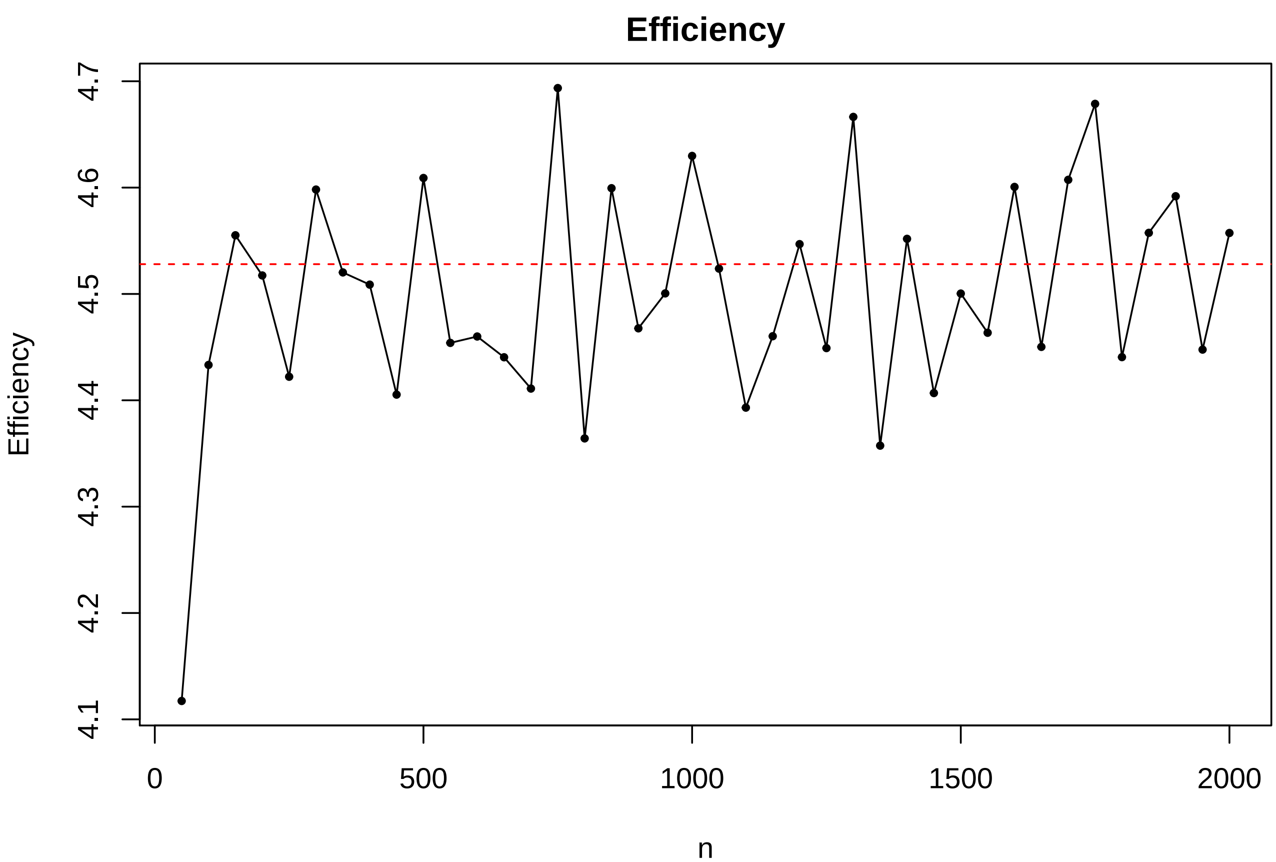


```
v1<-apply(ests.mat[,,1],1,var)
#Variance for tilde delta
nvec*v1

+  [1] 34.21735 35.26480 36.03715 35.68467 35.29061 35.87350 35.03104 36.07176
+  [9] 34.74845 35.58190 34.56152 35.49123 35.14699 34.68936 35.53579 34.88977
+ [17] 35.00965 35.45635 35.63328 35.87057 35.68328 34.51448 34.91756 34.58762
+ [25] 34.50509 35.96589 35.05918 36.06805 34.92996 36.18612 35.27704 35.35706
+ [33] 35.27736 36.02925 35.91959 34.46452 35.72490 35.83024 34.73638 35.32501

par(mar=c(4,4,3,1))
plot(nvec,nvec*v1,type='l',xlab='n',ylab='Variance')
points(nvec,nvec*v1,pch=19,cex=0.5)
abline(h=true.var.tilde,col='red',lty=2)
title(expression(paste('Variance of ',sqrt(n)*widetilde(delta))))
```




```
v2<-apply(ests.mat[,,2],1,var)
#Variance for delta hat
nvec*v2

+  [1] 8.310651 7.954632 7.911290 7.899423 7.980439 7.801714 7.749813 8.000389
+  [9] 7.887804 7.720068 7.759712 7.957697 7.915180 7.864222 7.571164 7.994609
+ [17] 7.611749 7.936225 7.917605 7.747808 7.887879 7.856500 7.828615 7.606973
+ [25] 7.755549 7.707327 8.045958 7.923902 7.926390 8.040731 7.903443 7.685256
+ [33] 7.927103 7.820016 7.677205 7.761254 7.838776 7.802961 7.810025 7.751256

par(mar=c(4,4,3,1))
plot(nvec,nvec*v2,type='l',xlab='n',ylab='Variance')
points(nvec,nvec*v2,pch=19,cex=0.5)
abline(h=true.var.hat,col='red',lty=2)
title(expression(paste('Variance of ',sqrt(n)*widehat(delta))))
```





The efficiency is the ratio of variances of the two estimators: $\text{Var}[\tilde{\delta}]/\text{Var}[\hat{\delta}]$.

Simulation 2: In the simulation below, we assume that

$$X \sim Normal(\mu_X, \sigma_X^2)$$

with $\mu_X = 1$ and $\sigma_X = 1$, and

$$Y|X = x, Z = z \sim Normal(\alpha x + \delta z, \sigma^2(x,z))$$

with $\sigma^2(x,z) = \sigma^2(1+z)$, and

$$\mu(x,z) = \mathbb{E}_{Y|X,Z}[Y|X = x, Z = z] = \alpha x + \delta z$$

with $\alpha = 0.5$ and $\delta = 2$, and hence

$$\mathbb{E}[Y(1) - Y(0)] = \mathbb{E}_{Y|X,Z}[Y|X = x, Z = 1] - \mathbb{E}_{Y|X,Z}[Y|X = x, Z = 0] = \delta.$$

We also have

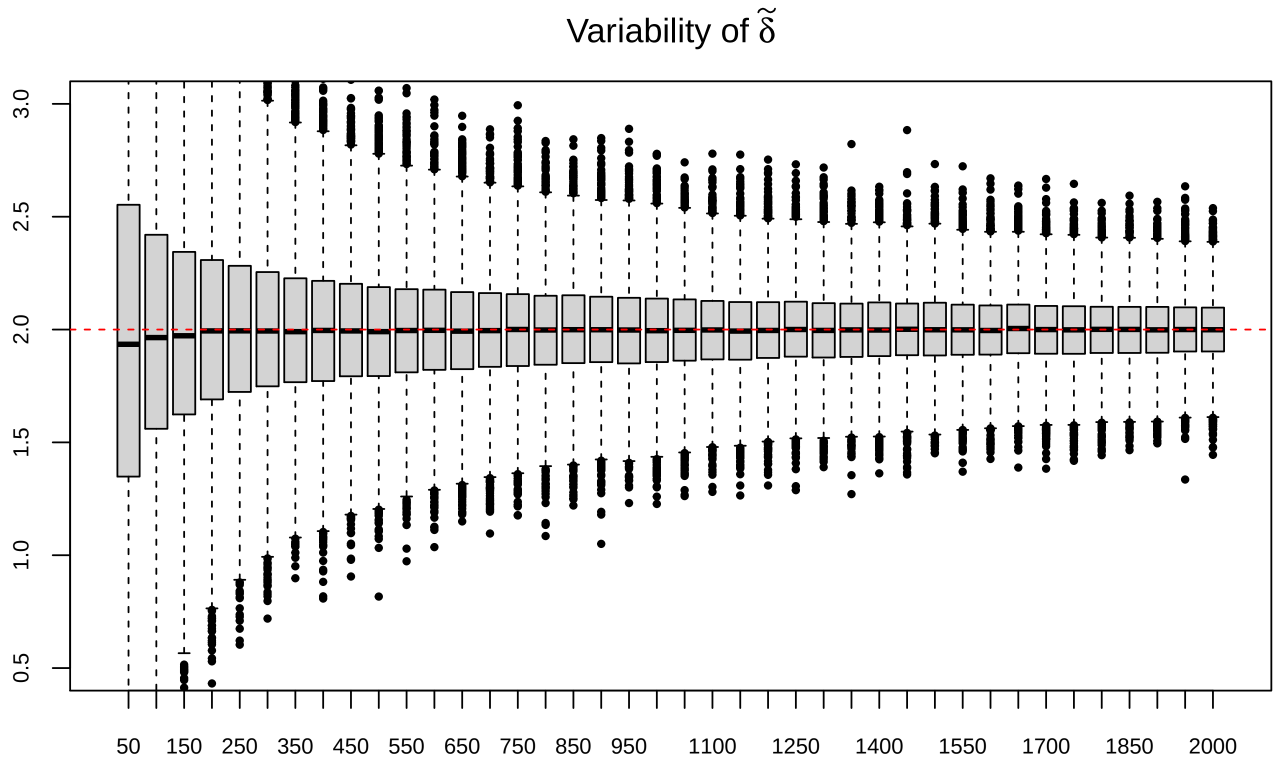
$$\mu(z) = \mathbb{E}_X[\mu(X,z)] = \mathbb{E}_X[\alpha X + \delta z] = \alpha \mathbb{E}_X[X] + \delta z = \alpha \mu_X + \delta z$$

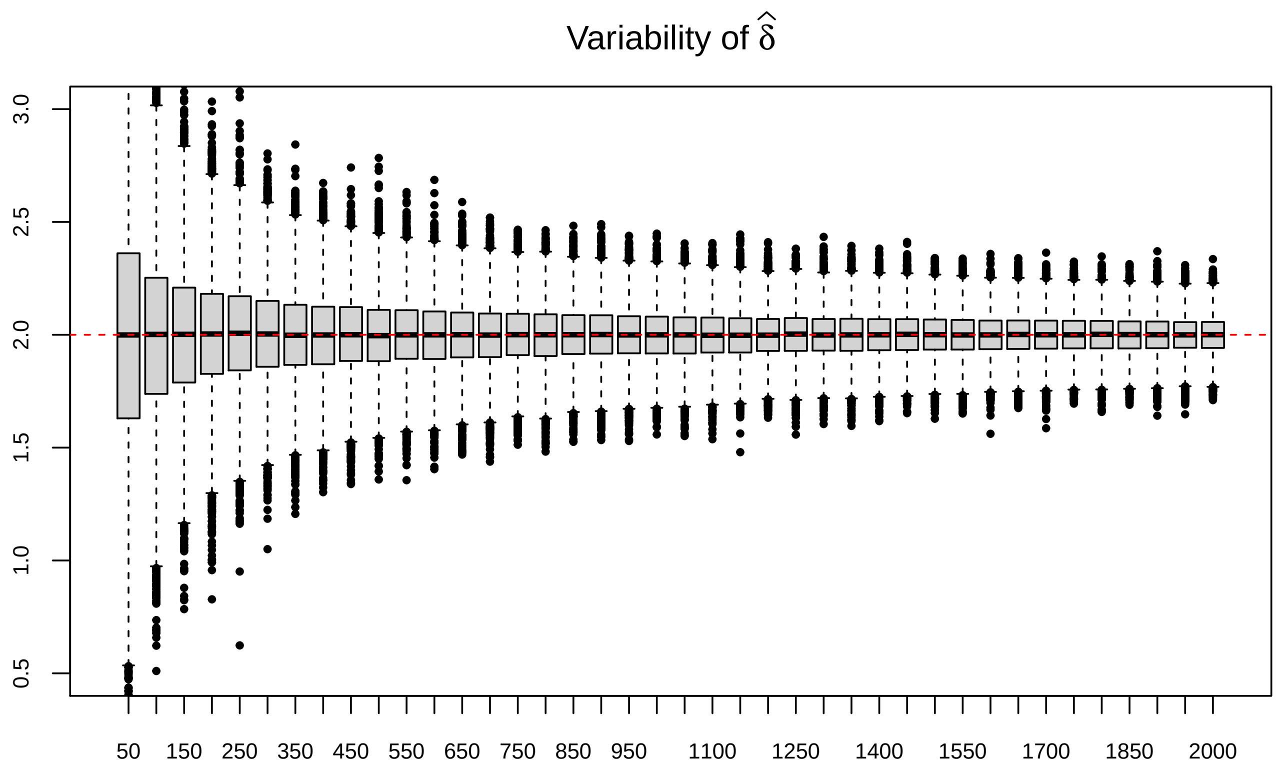
and that

$$\begin{aligned}\sigma^2(z) &= \mathbb{E}_X[\sigma^2(X,z)] + \text{Var}_X[\mu(X,z)] \\ &= \sigma^2(1+z) + \text{Var}_X[\alpha X + \delta z] \\ &= \sigma^2(1+z) + \alpha^2 \text{Var}_X[X] \\ &= \sigma^2(1+z) + \alpha^2 \sigma_X^2\end{aligned}$$

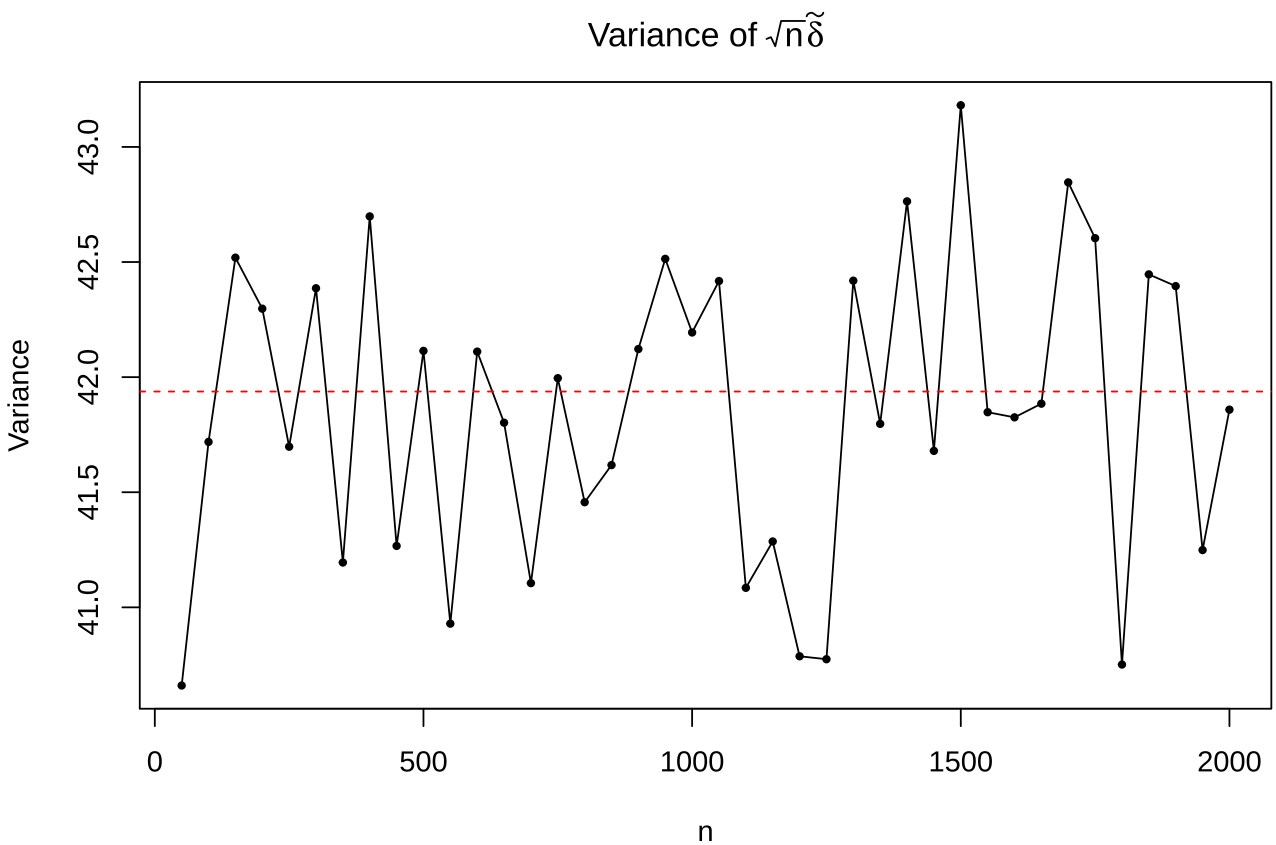
We study how the variances change as n increases. We choose $p = 0.2$.

```
set.seed(23)
nreps<-10000;nvec<-seq(50,2000,by=50)
p<-0.2;delta<-2;al<-0.5;sig<-1
muX<-1;sigX<-1
mu0<-al*muX; mu1<-mu0+delta
sigs<-c(0,0)
sigs[1]<-sqrt(sig^2*(1)+al^2*sigX^2)
sigs[2]<-sqrt(sig^2*(1+1)+al^2*sigX^2)
ests.mat<-array(0,c(length(nvec),nreps,2))
for(i in 1:length(nvec)){
    n<-nvec[i]
    for(irep in 1:nreps){
        X<-rnorm(n,muX,sigX)
        Z<-rbinom(n,1,p)
        Y<-rnorm(n,delta*Z+al*X,sigs[Z+1])
        p.hat<-mean(Z)
        ests.mat[i,irep,1]<-sum(Z*Y)/(n*p)-sum((1-Z)*Y)/(n*(1-p))
        ests.mat[i,irep,2]<-sum(Z*Y)/(n*p.hat)-sum((1-Z)*Y)/(n*(1-p.hat))
    }
}
#Variance of delta tilde
true.var.tilde<-(((1-p)*sigs[2]^2+p*sigs[1]^2)+al^2*sigX^2+(p*mu0+(1-p)*mu1)^2)/(p*(1-p))
true.var.tilde

+ [1] 41.9375

#Variance of delta hat
true.var.hat<-(((1-p)*sigs[2]^2+p*sigs[1]^2)+al^2*sigX^2)/(p*(1-p))
true.var.hat

+ [1] 14.375

#Efficiency
true.var.tilde/true.var.hat

+ [1] 2.917391
```

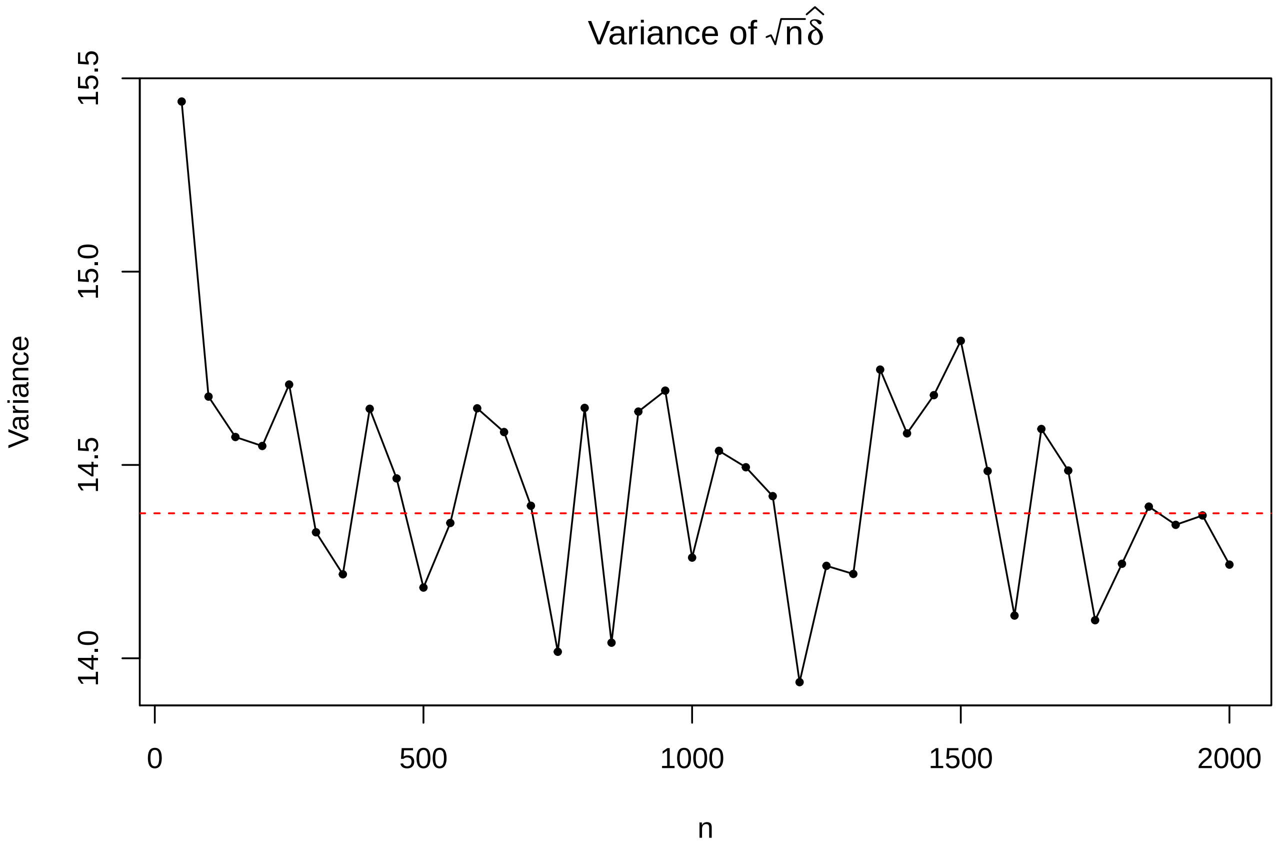

Variability of $\tilde{\delta}$
0.5
1.0
1.5
2.0
2.5
3.0
50
150
250
350
450
550
650
750
850
950
1100
1250
1400
1550
1700
1850
2000


Variability of $\hat{\delta}$
0.5
1.0
1.5
2.0
2.5
3.0
50
150
250
350
450
550
650
750
850
950
1100
1250
1400
1550
1700
1850
2000

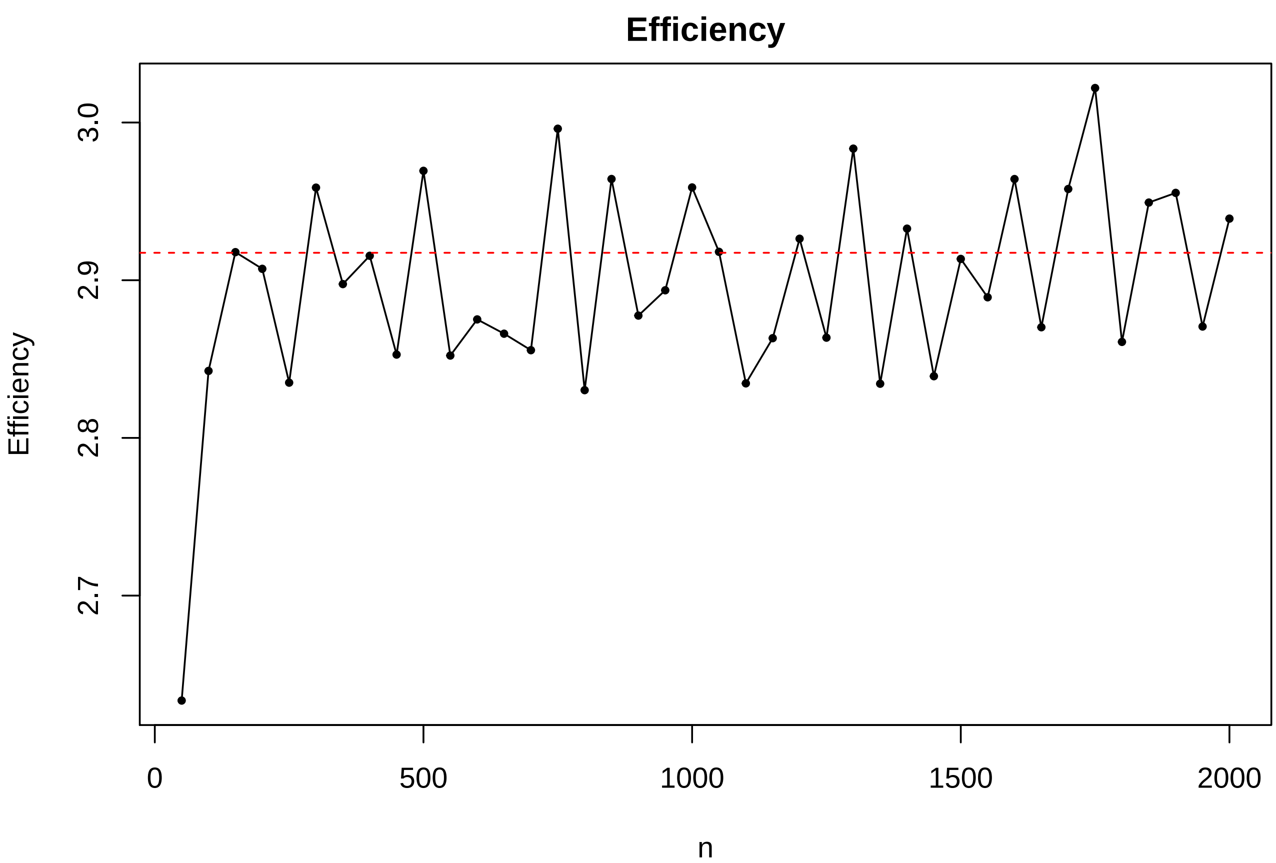
```
v1<-apply(ests.mat[,,1],1,var)
#Variance for tilde delta
nvec*v1

+  [1] 40.66042 41.71864 42.51910 42.29740 41.69781 42.38620 41.19487 42.69828
+  [9] 41.26688 42.11413 40.92922 42.11086 41.80208 41.10523 41.99550 41.45670
+ [17] 41.61794 42.12201 42.51375 42.19395 42.41750 41.08477 41.28620 40.78751
+ [25] 40.77464 42.41893 41.79722 42.76359 41.67977 43.18117 41.84774 41.82535
+ [33] 41.88475 42.84599 42.60364 40.75158 42.44638 42.39542 41.24886 41.85874

par(mar=c(4,4,3,1))
plot(nvec,nvec*v1,type='l',xlab='n',ylab='Variance')
points(nvec,nvec*v1,pch=19,cex=0.5)
abline(h=true.var.tilde,col='red',lty=2)
title(expression(paste('Variance of ',sqrt(n)*widetilde(delta))))
```




```
v2<-apply(ests.mat[,,2],1,var)
#Variance for delta hat
nvec*v2

+  [1] 15.44000 14.67682 14.57232 14.54907 14.70803 14.32590 14.21720 14.64532
+  [9] 14.46525 14.18287 14.34995 14.64649 14.58510 14.39444 14.01686 14.64758
+ [17] 14.04031 14.63807 14.69225 14.26036 14.53649 14.49409 14.41936 13.93823
+ [25] 14.23908 14.21814 14.74661 14.58165 14.68051 14.82108 14.48435 14.11041
+ [33] 14.59308 14.48554 14.09843 14.24440 14.39225 14.34518 14.36933 14.24224

par(mar=c(4,4,3,1))
plot(nvec,nvec*v2,type='l',xlab='n',ylab='Variance')
points(nvec,nvec*v2,pch=19,cex=0.5)
abline(h=true.var.hat,col='red',lty=2)
title(expression(paste('Variance of ',sqrt(n)*widehat(delta))))
```



Variance of $\sqrt{n}\hat{\delta}$
Variance
15.5
15.0
14.5
14.0
0
500
1000
1500
2000
n



Efficiency
Efficiency
3.0
2.9
2.8
2.7
0
500
1000
1500
2000
n

Simulation 3: In the simulation below, we assume that

$$X \sim Normal(\mu_X, \sigma_X^2)$$

with $\mu_X = 1$ and $\sigma_X = 1$, and

$$Y|X = x, Z = z \sim Normal(\alpha x + \delta z, \sigma^2(x, z))$$

with $\sigma^2(x, z) = \sigma^2(x^2 + z)$, and

$$\mu(x, z) = \mathbb{E}_{Y|X,Z}[Y|X = x, Z = z] = \alpha x + \delta z$$

with $\alpha = 0.5$ and $\delta = 2$, and hence

$$\mathbb{E}[Y(1) - Y(0)] = \mathbb{E}_{Y|X,Z}[Y|X = x, Z = 1] - \mathbb{E}_{Y|X,Z}[Y|X = x, Z = 0] = \delta.$$

We also have

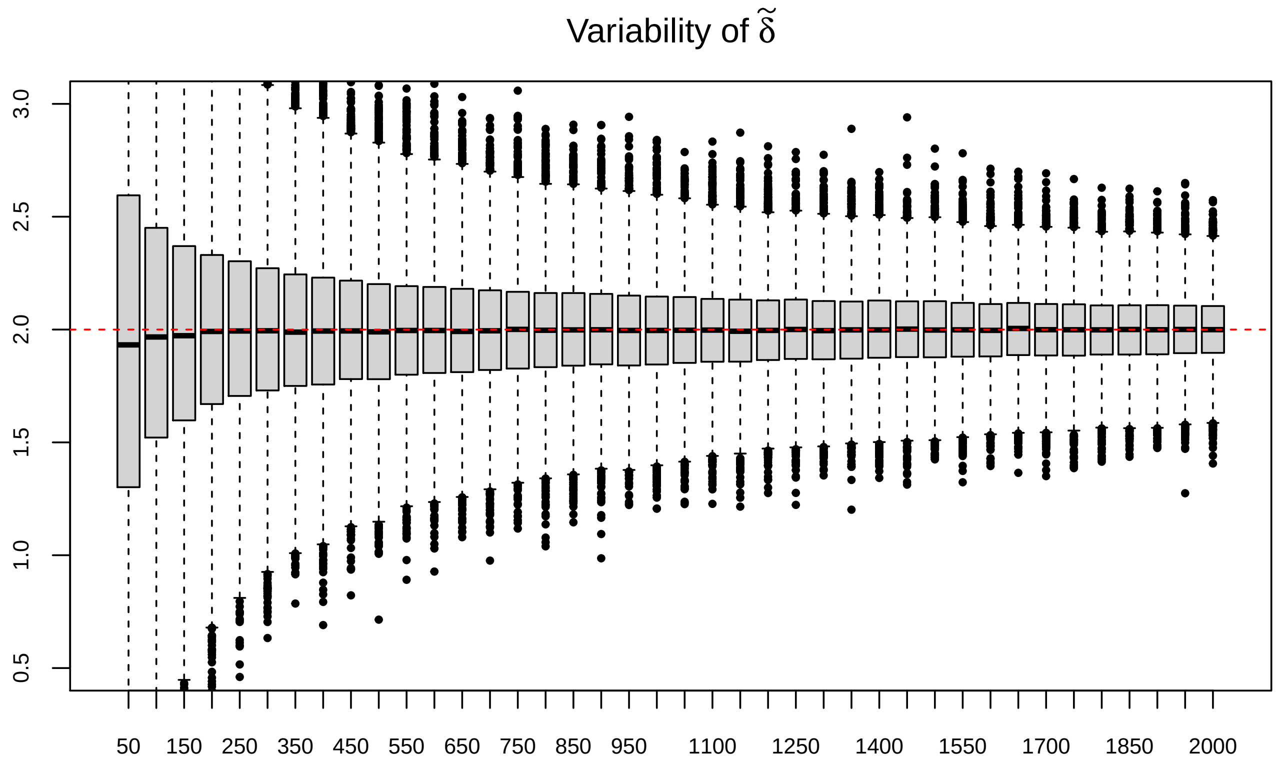
$$\mu(z) = \mathbb{E}_X[\mu(X, z)] = \mathbb{E}_X[\alpha X + \delta z] = \alpha\mathbb{E}_X[X] + \delta z = \alpha\mu_X + \delta z$$

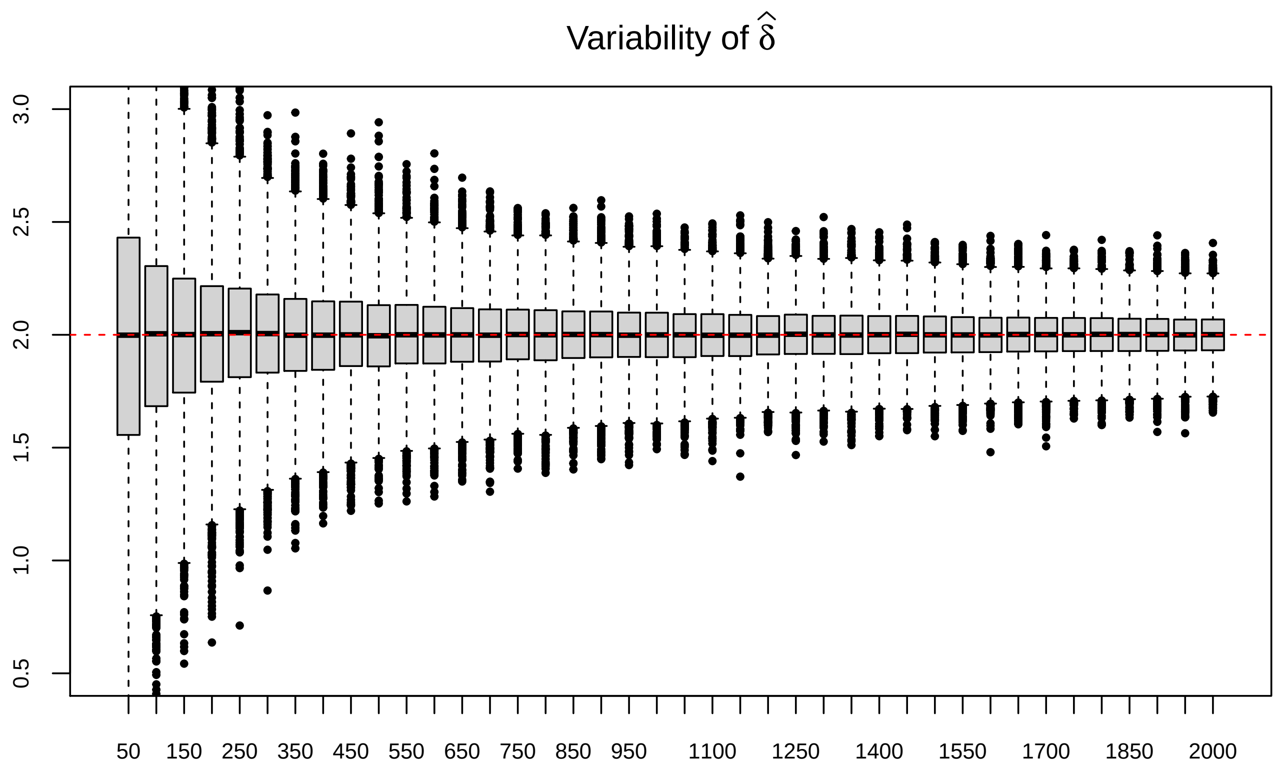
and that

$$\begin{aligned}
\sigma^2(z) &= \mathbb{E}_X[\sigma^2(X, z)] + \text{Var}_X[\mu(X, z)] \\
&= \sigma^2(\mathbb{E}_X[X^2] + z) + \text{Var}_X[\alpha X + \delta z] \\
&= \sigma^2(\mu_X^2 + \sigma_X^2 + z) + \alpha^2\text{Var}_X[X] \\
&= \sigma^2(\mu_X^2 + \sigma_X^2 + z) + \alpha^2\sigma_X^2
\end{aligned}$$

We study how the variances change as n increases. We choose $p = 0.2$.

```
set.seed(23)
nreps<-10000;nvec<-seq(50,2000,by=50)
p<-0.2;delta<-2;al<-0.5;sig<-1
muX<-1;sigX<-1
mu0<-al*muX; mu1<-mu0+delta
sigs<-c(0,0)
sigs[1]<-sqrt(sig^2*(muX^2+sigX^2)+al^2*sigX^2)
sigs[2]<-sqrt(sig^2*(muX^2+sigX^2+1)+al^2*sigX^2)
ests.mat<-array(0,c(length(nvec),nreps,2))
for(i in 1:length(nvec)){
    n<-nvec[i]
    for(irep in 1:nreps){
        X<-rnorm(n,muX,sigX)
        Z<-rbinom(n,1,p)
        Y<-rnorm(n,delta*Z+al*X,sigs[Z+1])
        p.hat<-mean(Z)
        ests.mat[i,irep,1]<-sum(Z*Y)/(n*p)-sum((1-Z)*Y)/(n*(1-p))
        ests.mat[i,irep,2]<-sum(Z*Y)/(n*p.hat)-sum((1-Z)*Y)/(n*(1-p.hat))
    }
}
#Variance of delta tilde
true.var.tilde<-(((1-p)*sigs[2]^2+p*sigs[1]^2)+al^2*sigX^2+(p*mu0+(1-p)*mu1)^2)/(p*(1-p))
true.var.tilde

+ [1] 48.1875

#Variance of delta hat
true.var.hat<-(((1-p)*sigs[2]^2+p*sigs[1]^2)+al^2*sigX^2)/(p*(1-p))
true.var.hat

+ [1] 20.625

#Efficiency
true.var.tilde/true.var.hat

+ [1] 2.336364
```

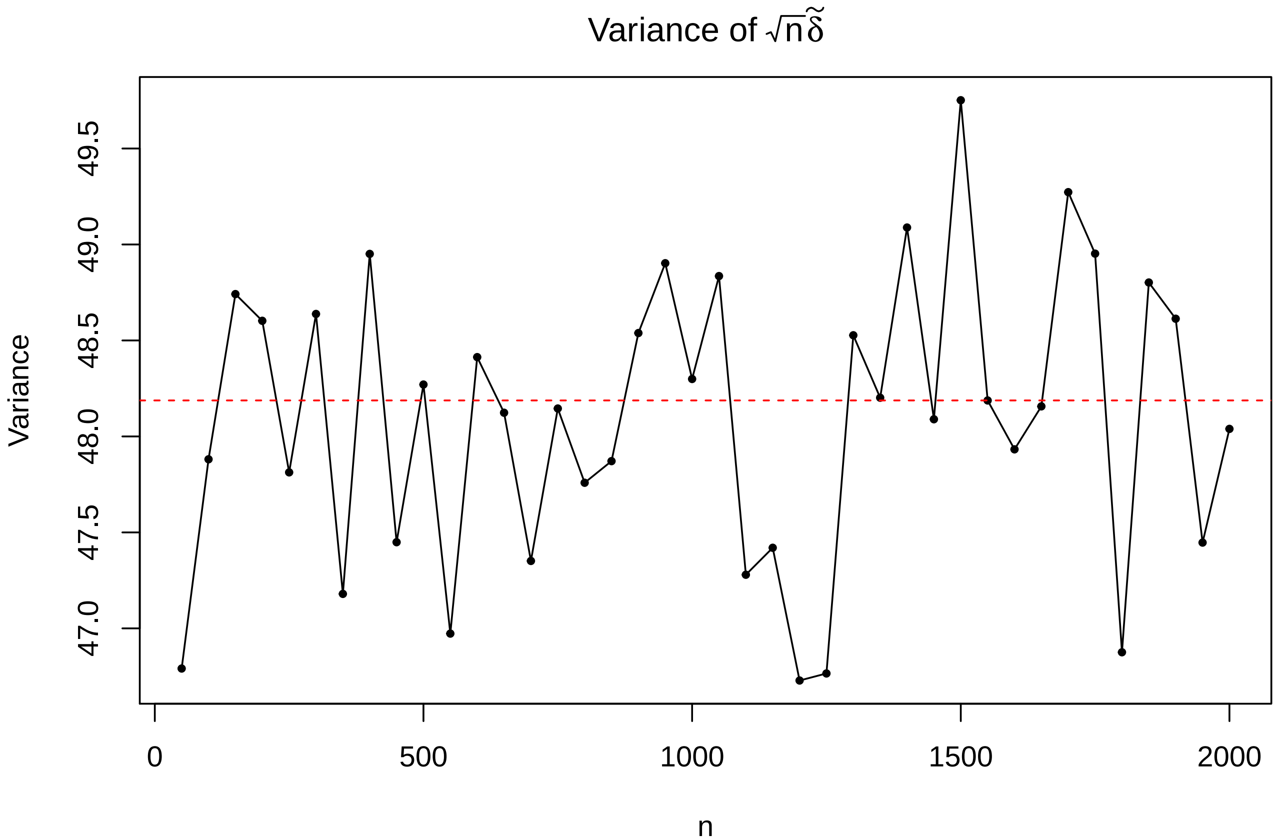

Variability of $\tilde{\delta}$
0.5 1.0 1.5 2.0 2.5 3.0
50 150 250 350 450 550 650 750 850 950 1100 1250 1400 1550 1700 1850 2000


Variability of $\hat{\delta}$
0.5 1.0 1.5 2.0 2.5 3.0
50 150 250 350 450 550 650 750 850 950 1100 1250 1400 1550 1700 1850 2000

```r
v1<-apply(ests.mat[,,1],1,var)
#Variance for tilde delta
nvec*v1
```

```
+  [1] 46.79075 47.88094 48.74159 48.60232 47.81278 48.63795 47.17971 48.95088
+  [9] 47.44893 48.27030 46.97286 48.41309 48.12342 47.35135 48.14543 47.75861
+ [17] 47.87136 48.53856 48.90241 48.29923 48.83542 47.27945 47.41985 46.72831
+ [25] 46.76515 48.52697 48.20157 49.08828 48.08946 49.75154 48.18735 47.93267
+ [33] 48.15663 49.27255 48.95228 46.87559 48.80188 48.61322 47.44694 48.03926
```

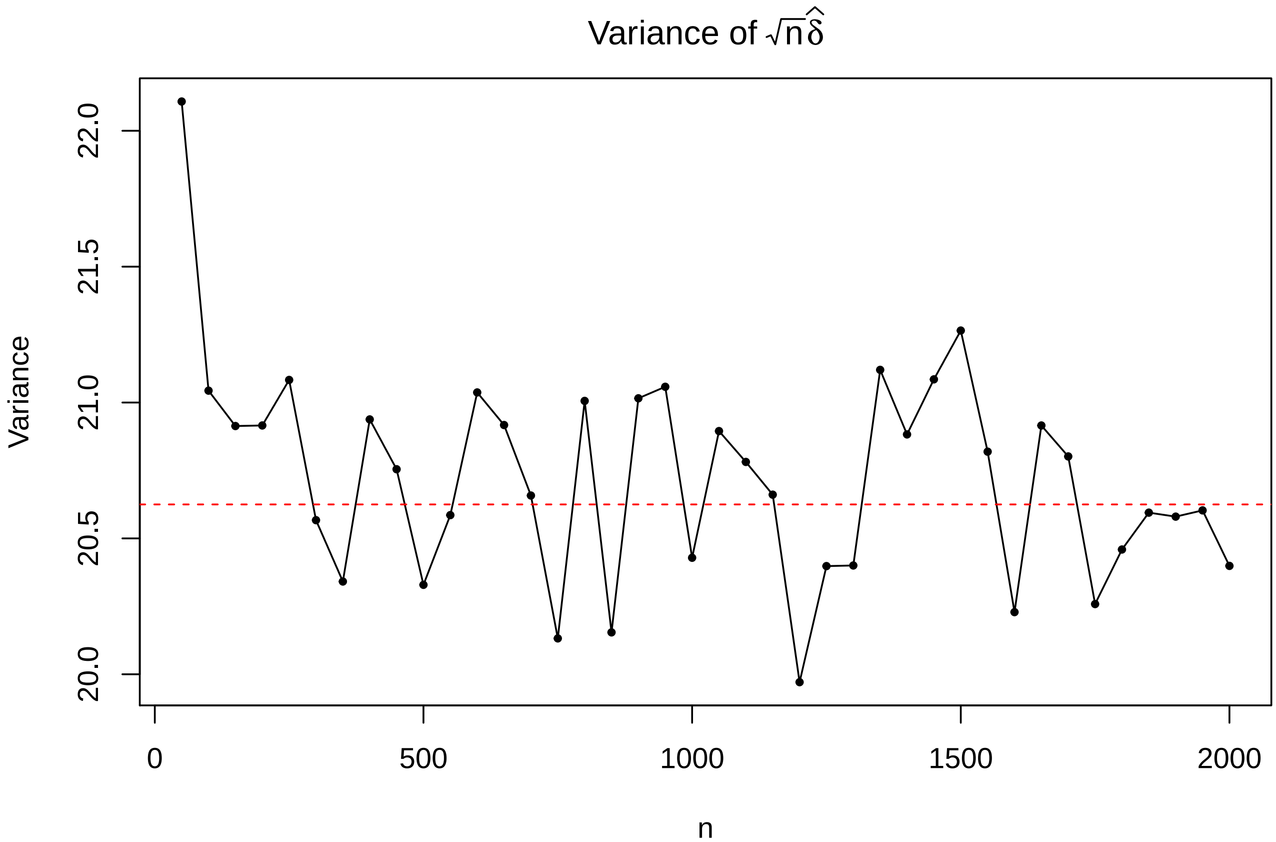
```r
par(mar=c(4,4,3,1))
plot(nvec,nvec*v1,type='l',xlab='n',ylab='Variance')
points(nvec,nvec*v1,pch=19,cex=0.5)
abline(h=true.var.tilde,col='red',lty=2)
title(expression(paste('Variance of ',sqrt(n)*widetilde(delta))))
```

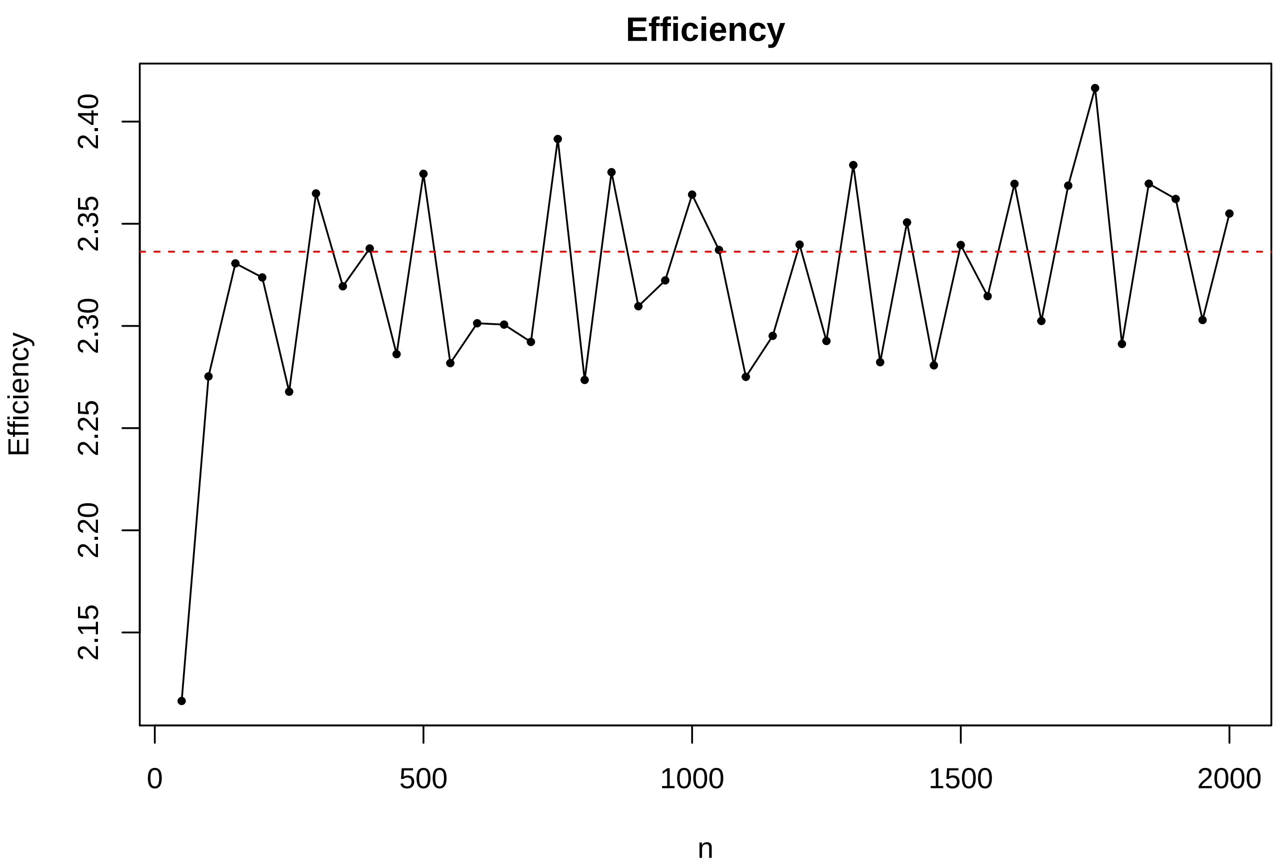



```r
v2<-apply(ests.mat[,,2],1,var)
#Variance for delta hat
nvec*v2
```

```
+  [1] 22.10768 21.04372 20.91352 20.91560 21.08312 20.56711 20.34131 20.93784
+  [9] 20.75448 20.32916 20.58573 21.03708 20.91718 20.65773 20.13202 21.00597
+ [17] 20.15421 21.01555 21.05790 20.42874 20.89494 20.78139 20.66087 19.97095
+ [25] 20.39806 20.40030 21.12015 20.88264 21.08527 21.26483 20.81918 20.22870
+ [33] 20.91550 20.80164 20.25836 20.45909 20.59479 20.58009 20.60306 20.39879
```

```r
par(mar=c(4,4,3,1))
plot(nvec,nvec*v2,type='l',xlab='n',ylab='Variance')
points(nvec,nvec*v2,pch=19,cex=0.5)
abline(h=true.var.hat,col='red',lty=2)
title(expression(paste('Variance of ',sqrt(n)*widehat(delta))))
```


Variance of $\sqrt{n}\hat{\delta}$
Variance
22.0
21.5
21.0
20.5
20.0
0
500
1000
1500
2000
n


Efficiency
Efficiency
2.40
2.35
2.30
2.25
2.20
2.15
0
500
1000
1500
2000
n