

Lsn24

Clark

Admin

Recall that at the end of class last lesson we mentioned that *Consistent* estimator is a statistic that, if we collect enough samples of, we are guaranteed that it has the same value as the parameter we are trying to estimate.

Note for our discussions we can think of our estimators as a function of both our data and of the number of data points that we collect, for example, if we have Y_1, \dots, Y_n are IID from some distribution we can write

$$\hat{\theta}_n = \frac{\sum_{i=1}^n Y_i}{n}$$

as an estimator of $\theta = \mu = E[Y]$. Note here that as n increases our estimator changes. What consistence is saying is that if n gets really really big, eventually we will converge to μ .

Formally, a consistent estimator of θ is one in which:

Our book uses a Theorem called Tchebysheff's inequality to prove that if we have an unbiased estimator, that is $E[\hat{\theta}_n] = \theta$, then all we have to show is that the Variance of our estimator *asymptotically* goes to 0.

For example, \bar{Y} is a consistent estimator in our problem above for μ because

In fact, what we've just shown above is what is called the *Weak Law of Large Numbers* which states that for IID data the sample mean converges to the population mean as long as the Variance of $Y_i < \infty$.

Although not explicitly shown in our text, we can actually go one step further. In Statistics there's another inequality called Markov's inequality that states, for any random variable

$$P(|X| > t) \leq \frac{E|X|^r}{t^r}$$

So, using this, we can state that for any estimator, biased or otherwise, we can write

Recalling that $E[(\hat{\theta}_n - \theta)^2]$ is the same thing as $MSE(\hat{\theta}_n) = \text{Bias } \hat{\theta}_n + \text{Var}(\hat{\theta}_n)^2$ it follows that for ANY estimator all we need to show is that both the bias and the variance asymptotically go to zero.

Recall that for $Y_1, \dots, Y_n \sim \text{Unif}(0, \theta)$ one estimator of θ is $\hat{\theta} = Y_{(n)}$, let's show that $Y_{(n)}$ is consistent.

Once we have a consistent estimator, $\hat{\theta}_n$, for θ and say we have another consistent estimator, $\hat{\phi}_n$ for a parameter ϕ , then we can form a consistent estimator for $\theta + \phi$ or $\theta\phi$ or $\frac{\theta}{\phi}$ by using $\hat{\theta}_n$ and $\hat{\phi}_n$.

This allows us to really quickly solve problems like 9.17 in our text. Suppose $X_1, X_2, \dots, X_n \sim f(\mu_1, \sigma_1)$ and $Y_1, Y_2, \dots, Y_n \sim g(\mu_2, \sigma_2)$ where $f(\cdot)$ and $g(\cdot)$ are some unknown distributions but have finite variance and mean. Find a consistent estimator for $\mu_1 - \mu_2$.

Let's work 9.19 on the boards.

In general, most logical estimators of parameters are consistent, but not in some special cases. Let's consider $Y_1, Y_2, \dots, Y_n \sim N(\mu, \sigma)$ but let's let our data not be independent, let's assume $Cor(Y_i, Y_j) = \rho$ for all pairs. Let's see what happens to $Var(\bar{Y})$ as $n \rightarrow \infty$. To answer this we need the fact that:

So now let's begin.

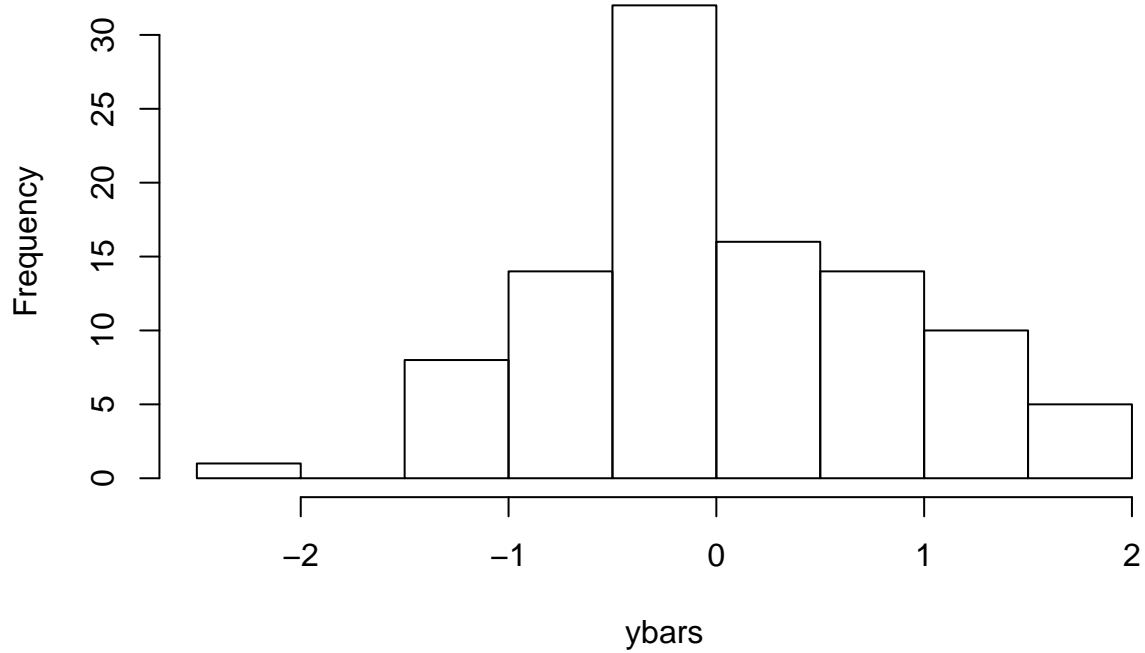
So what? What's the point? *MA476 doesn't have any application....* If your Data Are Correlated, you can't just go about your business and pretend that they are not and getting more data that is also correlated won't help!

```
library(MASS)
ybars<-c()
for(l in 1:100){
mu <- rep(0,400)
Sigma <- matrix(.7, nrow=400, ncol=400) + diag(400)*.3

rawvars <- mvrnorm(n=1, mu=mu, Sigma=Sigma)
ybars[l]<-mean(rawvars)}

hist(ybars)
```

Histogram of ybars



```
mean(ybars)
```

```
## [1] 0.08767961
```

```
var(ybars)
```

```
## [1] 0.7172589
```

So looking at the plot above, clearly $P(|\bar{Y} - \mu| > \epsilon) \neq 0$ for all ϵ for $n = 400$. But what we should have above is even if we have an infinite number of data points, we have no guarantee that we will accurately be estimating μ .