

## Stat 4620: Day 18

(3/28)

### Chap. 5 Revisited: Central Limit Theorem

**Theorem 1.** Let  $X_1, \dots, X_n$  be a random sample from a distribution that has density  $f(\cdot)$  with mean  $\mu$  and variance  $\sigma^2$ . Then the random variable

$$\sqrt{n}(\bar{X}_n - \mu) = \frac{\sum X_i - n\mu}{\sqrt{n}\sigma}$$

converges in distribution to a  $N(0, 1)$  random variable.

*Proof.* The moment generating function of a random variable  $X$  is  $M(t) = E[e^{tX}] = \int e^{tx} f(x) dx$ . Note that

$$\begin{aligned} M(0) &= \int e^{0x} f(x) dx = \int f(x) dx = 1 \\ M'(0) &= \left. \frac{\partial}{\partial t} \int e^{tx} f(x) dx \right|_{t=0} = \int x f(x) dx = E(X) \\ M''(0) &= \left. \frac{\partial^2}{\partial t^2} \int e^{tx} f(x) dx \right|_{t=0} = \int x^2 f(x) dx = E(X^2) \\ &\vdots \\ M^{(k)}(0) &= \left. \frac{\partial^k}{\partial t^k} \int e^{tx} f(x) dx \right|_{t=0} = \int x^k f(x) dx = E(X^k) \end{aligned}$$

Let  $m(t)$  be the mgf of  $X - \mu$ , i.e.

$$m(t) = E[e^{t(X-\mu)}]$$

then  $m(0) = 1$ ,  $m'(0) = 0$  and  $m''(0) = \sigma^2$ . By Taylor's expansion, for some  $\xi$  between 0 and  $t$ ,

$$\begin{aligned} m(t) &= m(0) + m'(0)t + \frac{m''(0)t^2}{2} + \frac{m'''(\xi)t^3}{3!} \\ &= 1 + \frac{\sigma^2 t^2}{2} + \frac{m'''(\xi)t^3}{3!} \end{aligned}$$

Now consider the mgf of  $\frac{\sum X_i - n\mu}{\sqrt{n}\sigma}$

$$\begin{aligned}
 M(t; n) &= E \left[ \exp \left( t \frac{\sum X_i - n\mu}{\sigma\sqrt{n}} \right) \right] \\
 &= E \left[ \exp \left( t \frac{X_1 - \mu}{\sigma\sqrt{n}} \right) \exp \left( t \frac{X_2 - \mu}{\sigma\sqrt{n}} \right) \cdots \exp \left( t \frac{X_n - \mu}{\sigma\sqrt{n}} \right) \right] \\
 &= E \left[ \exp \left( t \frac{X_1 - \mu}{\sigma\sqrt{n}} \right) \right] E \left[ \exp \left( t \frac{X_2 - \mu}{\sigma\sqrt{n}} \right) \right] \cdots E \left[ \exp \left( t \frac{X_n - \mu}{\sigma\sqrt{n}} \right) \right] \\
 &= \left[ m \left( \frac{t}{\sigma\sqrt{n}} \right) \right]^n \\
 &= \left[ 1 + \frac{\sigma^2 \left( \frac{t}{\sigma\sqrt{n}} \right)^2}{2} + \frac{m'''(\xi) \left( \frac{t}{\sigma\sqrt{n}} \right)^3}{3!} \right]^n \\
 &= \left[ 1 + \frac{t^2}{2n} + \frac{m'''(\xi)t^3}{3!\sigma^3 n^{3/2}} \right]^n \\
 &\longrightarrow \lim_{n \rightarrow \infty} \left[ 1 + \frac{t^2}{2n} \right]^n = e^{\frac{t^2}{2}}
 \end{aligned}$$

which is the mgf of  $N(0, 1)$ . □

```

> # SIMULATION OF CENTRAL LIMIT THEOREM
> sample(1:6,10, replace=T)                # 10 rolls of a die
[1] 6 4 4 2 5 1 6 1 5 5
> A<-matrix(sample(1:6, 10*1000, replace=T), ncol=10)  # 1000 ROWS (SAMPLES) OF n=10 ROLLS
> head(A)                                           # PRINT FIRST 6 SAMPLES
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
[1,]    3    2    6    3    2    5    1    6    2    2
[2,]    3    1    4    4    3    1    5    2    2    6
[3,]    1    1    5    5    1    3    5    3    2    6
[4,]    4    2    4    5    5    1    4    6    4    2
[5,]    1    3    3    4    1    1    2    4    5    5
[6,]    4    1    3    1    6    3    6    5    3    6
> means<-apply(A,1,mean)                          # MEAN OF B=1000 SAMPLES
> head(cbind(A,means))
      means
[1,] 3 2 6 3 2 5 1 6 2 2 3.2
[2,] 3 1 4 4 3 1 5 2 2 6 3.1
[3,] 1 1 5 5 1 3 5 3 2 6 3.2
[4,] 4 2 4 5 5 1 4 6 4 2 3.7
[5,] 1 3 3 4 1 1 2 4 5 5 2.9
[6,] 4 1 3 1 6 3 6 5 3 6 3.8
> hist(A[,1])                                     # HISTOGRAM OF COLUMN 1
> hist(A[,2])                                     # HISTOGRAM OF COLUMN 2
> hist(A[,3])                                     # HISTOGRAM OF COLUMN 3
> hist(A[,4])                                     # HISTOGRAM OF COLUMN 4
> hist(means)                                     # HISTOGRAM OF MEANS OF COL 1-10

```

## Sec. 6.2: Likelihood-based confidence interval

Recall theorem (Asymptotic normality of mle):

$$\hat{\theta}_n \sim N\left(\theta_0, \frac{1}{n\mathbf{I}(\theta_0)}\right)$$

Then

$$\begin{aligned} .95 &= P\left[-1.96 \leq \frac{\hat{\theta}_n - \theta_0}{\sqrt{\frac{1}{n\mathbf{I}(\theta_0)}}} \leq 1.96\right] \\ &\vdots \\ &= P\left[\hat{\theta}_n - 1.96\sqrt{\frac{1}{n\mathbf{I}(\theta_0)}} \leq \theta_0 \leq \hat{\theta}_n + 1.96\sqrt{\frac{1}{n\mathbf{I}(\theta_0)}}\right] \end{aligned}$$

So a 95% confidence interval for  $\theta_0$  is

$$\hat{\theta}_n \pm 1.96\sqrt{\frac{1}{n\mathbf{I}(\theta_0)}}$$

**Theorem 9** (Delta Method). *Let  $x_1, \dots, x_n$  be a random sample from  $f(x; \theta)$ ,  $\theta \in \Omega$ . Let  $\theta_0$  denote the true value. Assume that (R0)-(R5) hold, and suppose that  $0 < \mathbf{I}(\theta) < \infty$ . Let  $g(\theta)$  be differentiable. Then*

$$\sqrt{n}\left(g(\hat{\theta}_n) - g(\theta_0)\right) \xrightarrow{D} N\left(0, \frac{[g'(\theta_0)]^2}{\mathbf{I}(\theta_0)}\right)$$

*Proof.*

$$g(\hat{\theta}_n) = g(\theta_0) + g'(\theta_0)(\hat{\theta}_n - \theta_0) + \frac{1}{2}g''(\xi_n)(\hat{\theta}_n - \theta_0)^2$$

where  $\xi_n$  is between  $\hat{\theta}_n$  and  $\theta_0$ . Rearranging and multiplying by  $\sqrt{n}$

$$\sqrt{n}\left(g(\hat{\theta}_n) - g(\theta_0)\right) = g'(\theta_0)\sqrt{n}(\hat{\theta}_n - \theta_0) + \frac{1}{2}g''(\xi_n)\left[\sqrt{n}(\hat{\theta}_n - \theta_0)\right](\hat{\theta}_n - \theta_0)$$

Since  $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N\left(0, \frac{1}{\mathbf{I}(\theta_0)}\right)$  and  $(\hat{\theta}_n - \theta_0) \xrightarrow{P} 0$ , the result follows.  $\square$

### Corollary

It follows that a 95% confidence interval for  $g(\theta_0)$  is

$$g(\hat{\theta}_n) \pm 1.96\sqrt{\frac{[g'(\theta_0)]^2}{n\mathbf{I}(\theta_0)}}$$

**Example** (Poisson( $\theta$ )) Let  $x_1, \dots, x_n$  be a random sample from  $p(x; \theta) = \frac{e^{-\theta}\theta^x}{x!}$ ,  $x = 0, 1, \dots$

1. Construct a 95% confidence interval for the mean  $\theta$
2. Construct a 95% confidence interval for the standard deviation  $\sqrt{\theta}$

*Solution to (1)*

$$\ln p(x; \theta) = -\theta + x \ln \theta - \ln x!$$

$$\frac{\partial}{\partial \theta} \ln p(x; \theta) = -1 + \frac{x}{\theta}$$

$$\mathbf{I}(\theta) = V \left[ -1 + \frac{x}{\theta} \right] = \left( \frac{1}{\theta} \right)^2 V(x) = \left( \frac{1}{\theta} \right)^2 \theta = \frac{1}{\theta}$$

The mle is solution to

$$0 = \sum \frac{\partial}{\partial \theta} \ln p(x_i; \theta) = -n + \frac{\sum x_i}{\hat{\theta}} \quad \text{hence } \hat{\theta} = \bar{x}$$

A 95% confidence interval for  $\theta$  is

$$\hat{\theta} \pm \sqrt{\frac{1}{n\mathbf{I}(\theta_0)}}$$

or

$$\hat{\theta} \pm \sqrt{\frac{\theta_0}{n}}$$

which is estimated by

$$\bar{x} \pm 1.96 \sqrt{\frac{\bar{x}}{n}}$$

*Solution to (2):* 95% confidence interval for  $g(\theta) = \sqrt{\theta}$

$$g'(\theta) = \frac{1}{2\sqrt{\theta}}$$

95% confidence interval is

$$g(\hat{\theta}) \pm 1.96 \sqrt{\frac{[g'(\theta_0)]^2}{n\mathbf{I}(\theta_0)}}$$

or

$$\sqrt{\bar{x}} \pm 1.96 \sqrt{\frac{\left[ \frac{1}{2\sqrt{\bar{x}}} \right]^2}{n \frac{1}{\bar{x}}}}$$

$$\sqrt{\bar{x}} \pm 1.96 \sqrt{\frac{1}{4n}}$$

## Simulation in R

```
> # CONFIDENCE INTERVAL FOR THETA
> W<-matrix(rpois(30*1000, 9), ncol=30)      # 1000 SAMPLES OF SIZE n=30 FROM POIS(9)
> W[1,]                                         # PRINT FIRST SAMPLE (n=30)
 [1]  9 15  8 13  8  8 10  9 13 12  9  7  8 14 10  3 10  4  7  9  5  4 10  8 12 15 15  6
[29]  5 12
> mean(W[1,])                                 # MEAN OF FIRST SAMPLE
[1] 9.266667
> rowmeans<-apply(W,1,mean)                   # MEAN OF B=1000 SAMPLES
> head(rowmeans)                              # PRINT FIRST 6 MEANS
[1] 9.266667 9.066667 9.300000 9.200000 9.566667 9.433333
```

```

> lower1<-rowmeans-1.96*sqrt(rowmeans/30)      # LOWER ENDPOINT
> upper1<-rowmeans+1.96*sqrt(rowmeans/30)      # UPPER ENDPOINT
> head(cbind(lower1,upper1))                   # PRINT FIRST 6 INTERVALS
      lower1  upper1
[1,]  8.177342 10.35599
[2,]  7.989162 10.14417
[3,]  8.208718 10.39128
[4,]  8.114601 10.28540
[5,]  8.459850 10.67348
[6,]  8.334257 10.53241
> sum(9<lower1)                                # HOW MANY INTERVALS TOO HIGH
[1] 25
> sum(upper1<9)                                # HOW MANY INTERVALS TOO LOW
[1] 30
> 1-(25+30)/1000                               # PERCENTAGE OF CORRECT INTERVALS
[1] 0.945
>
> # CONFIDENCE INTERVAL FOR SQUAREROOT THETA
> lower2<-sqrt(rowmeans)-1.96*sqrt(1/(4*30))   # LOWER ENDPOINT
> upper2<-sqrt(rowmeans)+1.96*sqrt(1/(4*30))   # UPPER ENDPOINT
> head(cbind(lower2,upper2))                   # PRINT FIRST 6 INTERVALS
      lower2  upper2
[1,]  2.865197 3.223043
[2,]  2.832168 3.190013
[3,]  2.870667 3.228513
[4,]  2.854227 3.212073
[5,]  2.914080 3.271926
[6,]  2.892450 3.250296
> sum(3<lower2)                                # HOW MANY INTERVALS TOO HIGH
[1] 27
> sum(upper2<3)                                # HOW MANY INTERVALS TOO LOW
[1] 25
> 1-(27+25)/1000                               # PERCENTAGE OF CORRECT INTERVALS
[1] 0.948

```