Lecture 18: Convergence in Distribution and Asymptotic Normality

MATH 667-01
Statistical Inference
University of Louisville

November 21, 2017

Introduction

- We discuss convergence in distribution and the Central Limit Theorem as described in Section 5.5 of Casella and Berger (2002)[1].
- The proof of this version of the Central Limit Theorem is based on a result on the convergence of mgfs in Section 2.3.
- We also define and discuss a few results on asymptotic efficiency given in Section 10.1.
- Finally, we give asymptotic results regarding the MLE in Section 10.1 and the likelihood ratio test in Section 10.3 under regularity assumptions from Section 10.6.

[1] Casella, G. and Berger, R. (2002). *Statistical Inference, Second Edition*. Duxbury Press, Belmont, CA.

- *Definition L18.1* (Def 5.5.10 on p.235): A sequence of random variables, $X_1, X_2, \ldots$, *converges in distribution* to a random variable X if $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$, at all points x where $F_X(x)$ is continuous.
- *Example L18.1*: For each positive integer n, suppose X_n is a random variable which is $1 + \frac{1}{n}$ with probability 1. Find a random variable X such that $X_n \to X$ in distribution. What is the cdf of X?
- *Answer to Example L18.1*: The cdf of X_n is $F_{X_n}(x) = I_{[1+1/n,\infty)}(x)$. As $n \to \infty$, $F_{X_n}(x) \to I_{(1,\infty)}(x)$. This is not a cdf, but $I_{[1,\infty)}(x)$ is the cdf of a random variable X which is 1 with probability 1.
  Since *Definition L18.1* only requires that $F_{X_n}(x) \to F_X(x)$ at all points where X is continuous, we have $X_n \to X$ in distribution.

- *Example L18.2*: Suppose that $X_n \to X$ in distribution and $Y_n \to Y$ in distribution. Is it necessarily true that $X_n + Y_n \to X + Y$ in distribution?
- *Answer to Example L18.2*: No! Here is a counterexample. Let $X_1, X_2, \ldots$ be iid Normal$(0, 1)$ random variables and let $Y_n = -X_n$ for $n = 1, 2, \ldots$. Letting $Z \sim$ Normal$(0, 1)$, $X_n \to Z$ in distribution and $Y_n \to Z$ in distribution. If the statement is true, then $X_n + Y_n \to 2Z$ which would be Normal$(0, 4)$, but $X_n + Y_n = 0$ with probability 1.

- *Theorem L18.1* (Thm 2.3.12 on p.66): Suppose $\{X_n, n = 1, 2, \ldots\}$ is a sequence of random variables, each with mgf $M_{X_n}(t)$. Furthermore, suppose that $\lim_{n\to\infty} M_{X_n}(t) = M_X(t)$, for all t in a neighborhood of 0, and $M_X(t)$ is an mgf. Then there is a unique cdf F_X whose moments are determined by $M_X(t)$ and, for all X where $F_X(x)$ is continuous, we have $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$.

- *Theorem L18.2* (Thm 5.5.14 on p.236): Let $X_1, X_2, \ldots$ be a sequence of random variables whose mgfs exist in a neighborhood of 0 (that is, $M_{X_i}(t)$ exists for $|t| < h$, for some positive h). Let $\mathsf{E}[X_i] = \mu$ and $\mathsf{Var}[X_i] = \sigma^2 > 0$. Define $\bar{X}_n = (1/n)\sum_{i=1}^{n} X_i$. Let $G_n(x)$ denote the cdf of $\sqrt{n}(\bar{X}_n - \mu)/\sigma$. Then, for any x, $-\infty < x < \infty$,

$$\lim_{n\to\infty} G_n(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy,$$

that is, $\sqrt{n}\dfrac{(\bar{X}_n - \mu)}{\sigma}$ has a limiting standard normal distribution.

- *Sketch of proof of Theorem L18.2*: Let $Y_i = \frac{X_i - \mu}{\sigma}$ so that $\mathsf{E}[Y_i] = 0$ and $\mathsf{Var}[Y_i] = \mathsf{E}[Y_i^2] = 1$.
- Then $\sum_{i=1}^n Y_i = \frac{\sum_{i=1}^n X_i - n\mu}{\sigma} = \frac{n(\bar{X}_n - \mu)}{\sigma}$ so that $\frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$ is the random variable that we are interested in.
- We will show that the mgf of $\frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i$ converges to the mgf of a Normal$(0,1)$ random variable. First, we know

$$M_{\sum_{i=1}^n Y_i/\sqrt{n}}(t) = M_{\sum_{i=1}^n Y_i}\left(\frac{t}{\sqrt{n}}\right) = \left(M_{Y_i}\left(\frac{t}{\sqrt{n}}\right)\right)^n.$$

- *Sketch of proof of Theorem L18.2 continued*: The Taylor series expansion of $M_{Y_i}(t)$ is

$$
\begin{aligned}
M_{Y_i}(t) &= M_{Y_i}(0) + M'_{Y_i}(0)t + \frac{1}{2}M''_{Y_i}(0)t^2 + \sum_{k=3}^{\infty} \frac{1}{k!} M_{Y_i}^{(k)}(0)t^k \\
&= 1 + (\mathsf{E}[Y_i])t + \frac{1}{2}\mathsf{E}[Y_i^2]t^2 + \sum_{k=3}^{\infty} \frac{1}{k!} M_{Y_i}^{(k)}(0)t^k \\
&= 1 + \frac{1}{2}t^2 + \sum_{k=3}^{\infty} \frac{1}{k!} M_{Y_i}^{(k)}(0)t^k
\end{aligned}
$$

where $\lim\limits_{t \to 0} \dfrac{\sum_{k=3}^{\infty} \frac{1}{k!} M_{Y_i}^{(k)}(0)t^k}{t^2} = 0$.

- *Sketch of proof of Theorem L18.2 continued*: So, it follows that

$$\begin{aligned} M_{\sum_{i=1}^n Y_i/\sqrt{n}}(t) &= \left(1 + \frac{1}{2}\left(\frac{t}{\sqrt{n}}\right)^2 + \sum_{k=3}^{\infty} \frac{1}{k!} M_{Y_i}^{(k)}(0) \left(\frac{t}{\sqrt{n}}\right)^k\right)^n \\ &= \left(1 + \frac{t^2}{n}\left[\frac{1}{2} + \frac{\sum_{k=3}^{\infty} \frac{1}{k!} M_{Y_i}^{(k)}(0) \left(\frac{t}{\sqrt{n}}\right)^k}{(\frac{t}{\sqrt{n}})^2}\right]\right)^n . \end{aligned}$$

- For $t \neq 0$, Taylor's Theorem implies that $\frac{\sum_{k=3}^{\infty} \frac{1}{k!} M_{Y_i}^{(k)}(0) \left(\frac{t}{\sqrt{n}}\right)^k}{(\frac{t}{\sqrt{n}})^2} \to 0$ as $n \to \infty$ so that $M_{\sum_{i=1}^n Y_i/\sqrt{n}}(t) \to e^{t^2/2}$ as $n \to \infty$.
- Since this is the mgf of a standard normal distribution, we have shown that $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ converges in distribution to Normal$(0, 1)$ by *Theorem L18.1*.

- The following result on convergence in distribution is referred to as Slutsky's Theorem.
- *Theorem L18.3* (Thm 5.5.17 on p.239): If $X_n \to X$ in distribution and $Y_n \to a$, a constant, in probability, then
  a. $Y_n X_n \to aX$ in distribution.
  b. $X_n + Y_n \to X + a$ in distribution.
- Slutsky's Theorem can be used to obtain the following generalization of the Central Limit Theorem called the Delta Method.
- *Theorem L18.4* (Thm 5.5.24 on p.243): Let $Y_1, Y_2, \ldots$ be a sequence of random variables such that $\sqrt{n}(Y_n - \theta) \to \text{Normal}(0, \sigma^2)$ in distribution. For a given function g and a specific value of θ, suppose that $g'(\theta)$ exists and is not 0. Then

$$\sqrt{n}\,[g(Y_n) - g(\theta)] \to \text{Normal}(0, \sigma^2[g'(\theta)]^2) \text{ in distribution.}$$

- *Sketch of proof of Theorem L18.4 continued*: The Taylor expansion of $g(Y_n)$ around $Y_n = \theta$ is

$$g(Y_n) = g(\theta) + g'(\theta)(Y_n - \theta) + \text{Remainder}$$

where Remainder $\to 0$ as $Y_n \to 0$. Since $Y_n \to \theta$ in probability, it follows that Remainder $\to 0$ in probability. Applying *Theorem L18.3* to

$$\sqrt{n}\,\{g(Y_n) - g(\theta)\} = g'(\theta)\sqrt{n}(Y_n - \theta),$$

we see that

$$\sqrt{n}\,\{g(Y_n) - g(\theta)\} \to \text{Normal}(0, [g'(\theta)]^2\sigma^2) \text{ in distribution}$$

since $\sqrt{n}(Y_n - \theta) \to \text{Normal}(0, \sigma^2)$ in distribution.

- *Example L18.3*: Suppose that $X_1, X_2, \ldots$ is a sequence of iid Bernoulli(p) random variables with $p \in (0,1)$. Letting $\hat{p}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, show that

  (a) $\dfrac{\sqrt{n}\,(\hat{p}_n - p)}{\sqrt{p(1-p)}} \to$ Normal$(0,1)$ in distribution.

  (b) $\dfrac{\sqrt{n}\,(\hat{p}_n - p)}{\sqrt{\hat{p}_n(1-\hat{p}_n)}} \to$ Normal$(0,1)$ in distribution.

  (c) $\dfrac{\sqrt{n\hat{p}_n}\,(\ln \hat{p}_n - \ln p)}{\sqrt{1-\hat{p}_n}} \to$ Normal$(0,1)$ in distribution.

- *Answer to Example L18.3*: Since $\mathsf{E}[X_i] = p$ and $\mathsf{Var}[X_i] = p(1-p)$, *Theorem L18.2* implies that the statement (a) holds.
- Since $h(\hat{p}_n) = \sqrt{\frac{p(1-p)}{\hat{p}_n(1-\hat{p}_n)}}$ is a continuous function of $\hat{p}_n$,

$$\frac{\sqrt{n}\,(\hat{p}_n - p)}{\sqrt{\hat{p}_n(1-\hat{p}_n)}} = \sqrt{\frac{p(1-p)}{\hat{p}_n(1-\hat{p}_n)}}\frac{\sqrt{n}\,(\hat{p}_n - p)}{\sqrt{p(1-p)}} \to \text{Normal}(0,1) \text{ in distribution}$$

since $\hat{p}_n \to p$ by *Example L17.1* so that *Theorem L17.3* implies that $\sqrt{\frac{p(1-p)}{\hat{p}_n(1-\hat{p}_n)}} \to 1$ in probability. Then part (a) and *Theorem L18.3* imply statement (b).

- *Answer to Example L18.3 continued*: Letting $g(p) = \ln p$, note that $g'(p) = \frac{1}{p}$. Since $\sqrt{n}(\hat{p}_n - p) \to$ Normal$(0, p(1-p))$ in distribution, *Theorem L18.4* implies that $\sqrt{n}\,(\ln \hat{p}_n - \ln p) \to$ Normal$(0, p(1-p)\left(\frac{1}{p}\right)^2)$ in distribution. This is equivalent to $\sqrt{\frac{pn}{1-p}}\,(\ln \hat{p}_n - \ln p) \to$ Normal$(0, 1)$ in distribution. Since *Theorem L17.3* implies $\sqrt{\frac{\hat{p}_n n}{1-\hat{p}_n}} \to \sqrt{\frac{pn}{1-p}}$ in probability, statement (c) follows from *Theorem L18.3*.

Asymptotic Efficiency

- *Definition L18.2* (Def 10.1.7 on p.470): For a sequence of estimators $\{T_n\}$ and a sequence of normalizing constants $\{k_n\}$, τ^2 is called the *limiting variance* if $\lim_{n\to\infty} k_n \mathsf{Var}[T_n] = \tau^2 < \infty$.
- *Definition L18.3* (Def 10.1.9 on p.471): Suppose that $k_n(T_n - \tau(\theta)) \to \mathsf{Normal}(0, \sigma^2)$ in distribution for a sequence of estimators $\{T_n\}$ and a sequence of normalizing constants $\{k_n\}$. Then σ^2 is called the *asymptotic variance* of T_n.
- *Definition L18.4* (Def 10.1.11 on p.471): A sequence of estimators $\{W_n\}$ is called *asymptotically efficient* for a parameter $\tau(\theta)$ if $\sqrt{n}(W_n - \tau(\theta)) \to \mathsf{Normal}(0, v(\theta))$ in distribution where

$$v(\theta) = \frac{[\tau'(\theta)]^2}{\mathsf{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\ln f(X|\theta)\right)^2\right]}.$$

- *Example L18.4*: Suppose that $X_1, X_2, \ldots$ is a sequence of iid Bernoulli(p) random variables with $p \in (0,1)$. Letting $\hat{p}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, show that $\ln \hat{p}_n$ is asymptotically efficient for $\ln p$.
- *Answer to Example L18.4*: In *Example L18.3*(c) we showed that $\sqrt{n}\,(\ln \hat{p}_n - \ln p) \to$ Normal$(0, \frac{1-p}{p})$ in distribution. Since

$$\begin{aligned} \frac{d}{dp}\left[\ln p^x (1-p)^{1-x}\right] &= \frac{d}{dp}\left[x \ln p + (1-x)\ln(1-p)\right] \\ &= \frac{x}{p} - \frac{1-x}{1-p} \\ &= \frac{x-p}{p(1-p)}, \end{aligned}$$

- *Answer to Example L18.4 continued*: we have

$$
\begin{aligned}
\mathsf{E}\left[\left(\frac{d}{dp}\ln f(X|p)\right)^2\right] &= \mathsf{E}\left[\left(\frac{X-p}{p(1-p)}\right)^2\right] \\
&= \frac{\mathsf{E}[(X-p)^2]}{p^2(1-p)^2} \\
&= \frac{\mathsf{Var}[X]}{p^2(1-p)^2} \\
&= \frac{p(1-p)}{p^2(1-p)^2} \\
&= \frac{1}{p(1-p)}.
\end{aligned}
$$

So, $v(p) = \dfrac{(\frac{d}{dp}\ln p)^2}{\frac{1}{p(1-p)}} = \dfrac{\frac{1}{p^2}}{\frac{1}{p(1-p)}} = \dfrac{1-p}{p}.$

Asymptotic Efficiency of the MLE

- *Theorem L18.5* (Thm 10.1.12 on p.472): Let $X_1, X_2, \ldots$ be iid $f(x|\theta)$, and let $L_n(\theta|\boldsymbol{x}) = \prod_{i=1}^n f(x_i|\theta)$ be the likelihood function based on $X_1, \ldots, X_n$ and $\hat{\theta}_n$ denote the maximum likelihood estimator (MLE) of θ. Let $\tau(\theta)$ be a continuous function of θ. If assumptions (A1)–(A6) on slide 17.12 hold, then

$$\sqrt{n}\left(\tau(\hat{\theta}_n) - \tau(\theta)\right) \to \text{Normal}(0, v(\theta)) \text{ in distribution}$$

where $v(\theta) = \dfrac{[\tau'(\theta)]^2}{\mathsf{E}_\theta\left(\left(\frac{\partial}{\partial\theta}\ln f(X|\theta)\right)^2\right)}$ is the Crámer-Rao Lower Bound.

Asymptotic Efficiency of the MLE

- *Sketch of proof of Theorem L18.5*: Here we prove the result for $\tau(\theta) = \theta$. Consider the Taylor series expansion of the derivative of $\ell(\theta|\boldsymbol{x}) = \sum_{i=1}^{n} \ell(\theta|x_i) = \sum_{i=1}^{n} \ln f(x_i|\theta)$:

$$\ell'(\theta|\boldsymbol{x}) = \ell'(\theta_0|\boldsymbol{x}) + (\theta - \theta_0)\ell''(\theta_0|x) + \dots$$

Plugging in $\theta = \hat{\theta}_n$ and ignoring higher order terms, we obtain

$$\begin{aligned} \ell'(\hat{\theta}_n|\boldsymbol{x}) &\approx \ell'(\theta_0|\boldsymbol{x}) + (\hat{\theta}_n - \theta_0)\ell''(\theta_0|x) \\ 0 &\approx \ell'(\theta_0|\boldsymbol{x}) + (\hat{\theta}_n - \theta_0)\ell''(\theta_0|\boldsymbol{x}) \\ \sqrt{n}(\hat{\theta}_n - \theta_0) &\approx \frac{\sqrt{n}\frac{1}{n}\ell'(\theta_0|\boldsymbol{x})}{-\frac{1}{n}\ell''(\theta_0|\boldsymbol{x})}. \end{aligned}$$

- *Sketch of proof of Theorem L18.5 continued*: Since $\mathsf{E}[\ell'(\theta|x_i)] \overset{9.8}{=} 0$ and $\mathsf{Var}[\ell'(\theta|x_i)] \overset{9.9}{=} \mathsf{E}[(\ell'(\theta|x_i))^2] \overset{9.12}{=} I(\theta)$, $\sqrt{n}(\frac{1}{n}\sum_{i=1}^n \ell'(\theta_0|x_i) - 0) \to \mathsf{Normal}(0, I(\theta_0))$ in distribution by *Theorem L18.6*.
  Since $\mathsf{E}[\ell''(\theta|x_i)] = -I(\theta)$ by *Theorem L9.4*, $\frac{1}{n}\sum_{i=1}^n \ell''(\theta_0|x_i) \to -I(\theta_0)$ in probability by *Theorem L17.4*.
  Letting $W \sim \mathsf{Normal}(0, I(\theta_0))$, we have by *Theorem L18.3*,

  $$\sqrt{n}(\hat{\theta}_n - \theta_0) \to \frac{1}{I(\theta_0)}W$$

  in distribution. Finally, $\frac{1}{I(\theta_0)}W \sim \mathsf{Normal}(0, \frac{1}{I(\theta_0)})$.

- *Example L18.5*: Suppose that $X_1, X_2, \ldots, X_n$ is a sequence of iid Bernoulli(p) random variables with $p \in (0,1)$. Letting $\hat{p}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, construct an approximate 95% confidence interval for $\ln p$.
- *Answer to Example L18.5*: From *Example L18.3*(c), we know that $\dfrac{\sqrt{n\hat{p}_n}\,(\ln \hat{p}_n - \ln p)}{\sqrt{1-\hat{p}_n}} \to$ Normal$(0,1)$ in distribution. This implies that

$$P\left(-z_{.025} \leq \frac{\sqrt{n\hat{p}_n}\,(\ln \hat{p}_n - \ln p)}{\sqrt{1-\hat{p}_n}} \leq z_{.025}\right) \to .95$$

$$\Rightarrow P\left(\ln \hat{p}_n - z_{.025}\sqrt{\frac{1-\hat{p}_n}{n\hat{p}_n}} \leq \ln p \leq \ln \hat{p}_n + z_{.025}\sqrt{\frac{1-\hat{p}_n}{n\hat{p}_n}}\right) \to .95$$

as $n \to \infty$ where $z_{.025} \approx 1.960$ is the value such that $P(Z \leq z_{.025}) = .975$ where $Z \sim$ Normal$(0,1)$.

Simulation

```
> set.seed(348756)
> n=c(10,100,1000,10000,100000)
> p=.9
> R=1000000
> for (i in 1:length(n)){
+  count=0
+  for (j in 1:R){
+   p.hat=rbinom(1,size=n[i],prob=p)/n[i]
+   lower.limit=log(p.hat)-qnorm(.975)*sqrt((1-p.hat)/(n[i]*p.hat))
+   upper.limit=log(p.hat)+qnorm(.975)*sqrt((1-p.hat)/(n[i]*p.hat))
+   if ((lower.limit<log(p))&(log(p)<upper.limit))
+    count=count+1
+  }
+  cat("n=",n[i]," estimated conf=",count/R,"\n")
+ }
n= 10  estimated conf= 0.652169
n= 100  estimated conf= 0.932281
n= 1000  estimated conf= 0.944211
n= 10000  estimated conf= 0.950801
n= 1e+05  estimated conf= 0.950024
```

Asymptotic Distribution of the LRT

- *Theorem L18.6* (Thm 10.3.1 on p.489): Let $X_1, X_2, \dots$ be iid $f(x|\theta)$. Suppose $L_n(\theta|\boldsymbol{x}) = \prod_{i=1}^n f(x_i|\theta)$ is the likelihood function based on $X_1, \dots, X_n$ and let

$$\lambda_n(\boldsymbol{x}) = \frac{\sup\limits_{\Theta_0} L_n(\theta|\boldsymbol{x})}{\sup\limits_{\Theta} L_n(\theta|\boldsymbol{x})} = \frac{L_n(\theta_0|\boldsymbol{x})}{L_n(\hat{\theta}_n|\boldsymbol{x})}$$

  be the likelihood ratio test statistic for testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$. If assumptions (A1)–(A6) on slide 17.12 hold, then

$$-2 \ln \lambda_n(\boldsymbol{X}) \to \chi_1^2 \text{ in distribution.}$$

- *Sketch of proof of Theorem L18.6*: Consider the Taylor series expansion of $\ell(\theta_0|\boldsymbol{x})$ about θ:

$$\ell(\theta_0|\boldsymbol{x}) = \ell(\theta|\boldsymbol{x}) + \ell'(\theta|\boldsymbol{x})(\theta_0 - \theta) + \frac{1}{2}\ell''(\theta|\boldsymbol{x})(\theta_0 - \theta)^2 + \dots$$

Plugging in $\theta = \hat{\theta}_n$ so that $\ell'(\hat{\theta}_n|\boldsymbol{x}) = 0$ and then ignoring higher order terms, we have

$$\begin{aligned} \ell(\theta_0|\boldsymbol{x}) &\approx \ell(\hat{\theta}_n|\boldsymbol{x}) + \frac{1}{2}\ell''(\hat{\theta}_n|\boldsymbol{x})(\theta_0 - \hat{\theta}_n)^2 \\ \ell(\hat{\theta}_n|\boldsymbol{x}) - \ell(\theta_0|\boldsymbol{x}) &\approx -\frac{1}{2}\ell''(\hat{\theta}_n|\boldsymbol{x})(\theta_0 - \hat{\theta}_n)^2. \end{aligned}$$

- *Sketch of proof of Theorem L18.6 continued*: Since

$$\begin{aligned}
-2\ln\lambda(\boldsymbol{x}) &= -2\ln\frac{L(\theta_0|\boldsymbol{x})}{L(\hat{\theta}_n|\boldsymbol{x})} \\
&= -2\left(\ell(\theta_0|\boldsymbol{x}) - \ell(\hat{\theta}_n|\boldsymbol{x})\right) \\
&\approx \ell''(\hat{\theta}_n|\boldsymbol{x})(\theta_0 - \hat{\theta}_n)^2 \\
&= \frac{\ell''(\hat{\theta}_n|\boldsymbol{x})}{nI(\theta_0)}\left(\sqrt{n}\frac{\hat{\theta}_n - \theta_0}{\sqrt{1/I(\theta_0)}}\right)^2
\end{aligned}$$

Theorem L18.3 implies that $-2\ln\lambda(\boldsymbol{x}) \to \chi_1^2$ in distribution since $\frac{\ell''(\hat{\theta}_n|\boldsymbol{x})}{n} \to I(\theta_0)$ in probability by *Theorem L17.4* so that $\frac{\ell''(\hat{\theta}_n|\boldsymbol{x})}{nI(\theta_0)} \to 1$, and $\sqrt{n}\frac{\hat{\theta}_n-\theta_0}{\sqrt{1/I(\theta_0)}} \to$ Normal$(0,1)$ in distribution by *Theorem L18.5* so that $\left(\sqrt{n}\frac{\hat{\theta}_n-\theta_0}{\sqrt{1/I(\theta_0)}}\right)^2 \to \chi_1^2$.

- *Example L18.6*: Suppose we toss a coin one million times and observe 501000 heads. Does this provide sufficient evidence that the coin is not fair based on the asymptotic approximation to the likelihood ratio test at signficance level 0.05?
- *Answer to Example L18.6*: Here we want to perform the two-sided test $H_0 : p = 0.5$ versus $H_1 : p \neq 0.5$. Since $\ell(p) = \sum_{i=1}^{n} \{x_i \ln p + (1 - x_i)\ln(1 - p)\}$, we see that $\ell(.5) = n \ln .5$ and $\ell(\hat{p}) = n\left(\bar{x}\ln\bar{x} + (1 - \bar{x})\ln(1 - \bar{x})\right)$. So, we have

$$
\begin{aligned}
-2\ln\lambda(\boldsymbol{x}) &= -2\left(\ell(.5) - \ell(\bar{x})\right) \\
&= 2\left(\bar{x}\ln\bar{x} + (1 - \bar{x})\ln(1 - \bar{x}) - \ln .5\right).
\end{aligned}
$$

Since $n = 1000000$ and $\bar{x} = \frac{501000}{1000000} = .5001$, we have $-2\ln\lambda(\boldsymbol{x}) = 4.000003$. If H_0 is true, then $-2\ln\lambda(\boldsymbol{x})$ is approximately χ_1^2 so we reject H_0 since $\chi_{1,.05}^2 = 3.841$ and find that there is strong evidence that the coin is not fair.

- *Theorem L18.7* (Thm 10.3.3 on p.490): Let $X_1, X_2, \ldots$ be iid $f(x|\boldsymbol{\theta})$. Let $L_n(\boldsymbol{\theta}|\boldsymbol{x}) = \prod_{i=1}^n f(x_i|\boldsymbol{\theta})$ be the likelihood function based on $X_1, \ldots, X_n$ and $\lambda_n(\boldsymbol{x}) = \dfrac{\sup\limits_{\Theta_0} L_n(\boldsymbol{\theta}|\boldsymbol{x})}{\sup\limits_{\Theta} L_n(\boldsymbol{\theta}|\boldsymbol{x})}$ be the corresponding likelihood ratio test statistic. If $\boldsymbol{\theta} \in \Theta_0$ and assumptions (A1)–(A6) on slide 17.12 hold, then

$$-2\ln\lambda_n(\boldsymbol{X}) \to \chi^2_{df} \text{ in distribution}$$

where df is the difference between the number of free parameters specified by $\boldsymbol{\theta} \in \Theta_0$ and the number of free parameters specified by $\boldsymbol{\theta} \in \Theta$.