MATH 6070: HW1

Jie Ma

February 2, 2017

Exercise 1.2 from text: Show how the CLT for iid random variables may be used to construct an approximate $100(1-\alpha)\%$ confidence interval for p.

Solution: We consider the statistics $\bar{p} = \sum x_i/n$, where $x_i \sim BIN(1,p)$. Based on CLT, we have that

$$\frac{\bar{p}-p}{\sigma/\sqrt{n}} \to N(0,1), \tag{1}$$

in distribution. Hence, if the sample size n is large, we may just use the standard confidence interval for normal distribution so that the confidence interval is

$$[-z_{1-\alpha/2} < \frac{\bar{p}-p}{\sqrt{p(1-p)/n}} < z_{1-\alpha/2}]. \tag{2}$$

From this equation, we directly observe that p is the solution of the quadratic equation

$$(\bar{p}-p)^2 = \frac{Z^2_{1-\alpha/2}}{n} p(1-p), \tag{3}$$

from which p can be solved to give

$$p = \frac{A + 2\bar{p} \pm \sqrt{A^2 + 4A\bar{p} - 4A\bar{p}^2}}{2(A+1)}, \tag{4}$$

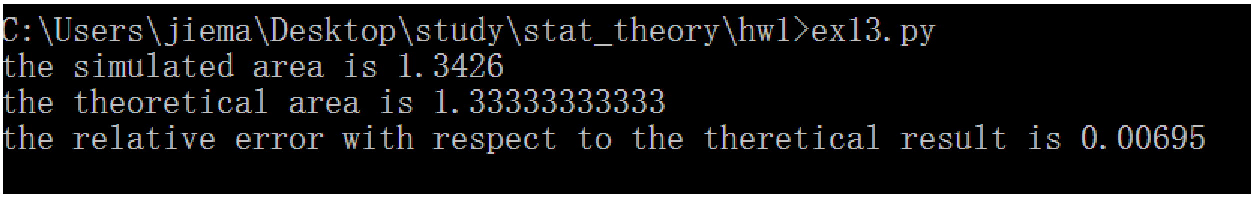
where $A := \frac{Z^2_{1-\alpha/2}}{n}$.

3. Exercise 1.3 from text: Consider the purely mathematical problem of finding a definite integral $\int_a^b f(x)dx$. Show that the usual laws of large numbers provide a method for approximately finding the value of the integral by using appropriate averages.

Solutions: We proceed by generating a sequence of points in the plane (x_i, y_i), where x_i and y_i are uniformly distributed random variables, whose support are $[a,b]$ and $[y_{min}, y_{max}]$. Then if $y_i < f(x_i)$, we count it as a success; if $y_i > f(x_i)$, we treat it as a failure. In the end, by LLN, we have the following expression

$$\frac{\text{Number of success}}{n} \approx \frac{\int_a^b f dx}{\text{area of the rectangle}}, \tag{5}$$

for n large. For example, if we choose the function to be $f(x) = -x^2 + 2x$ on $[0, 2]$, then its theoretical area is exactly 4/3. By running a Monte Carlo simulation with Python, with the total number of points $N = 10000$ generated, we find that the simulated area is close to 4/3, and the relative error is less than 0.7%. See figure (1) and (2) for reference.



```
C:\Users\jiema\Desktop\study\stat_theory\hw1>ex13.py
the simulated area is 1.3426
the theoretical area is 1.33333333333
the relative error with respect to the theretical result is 0.00695
```

Figure 1: Simulated Results.

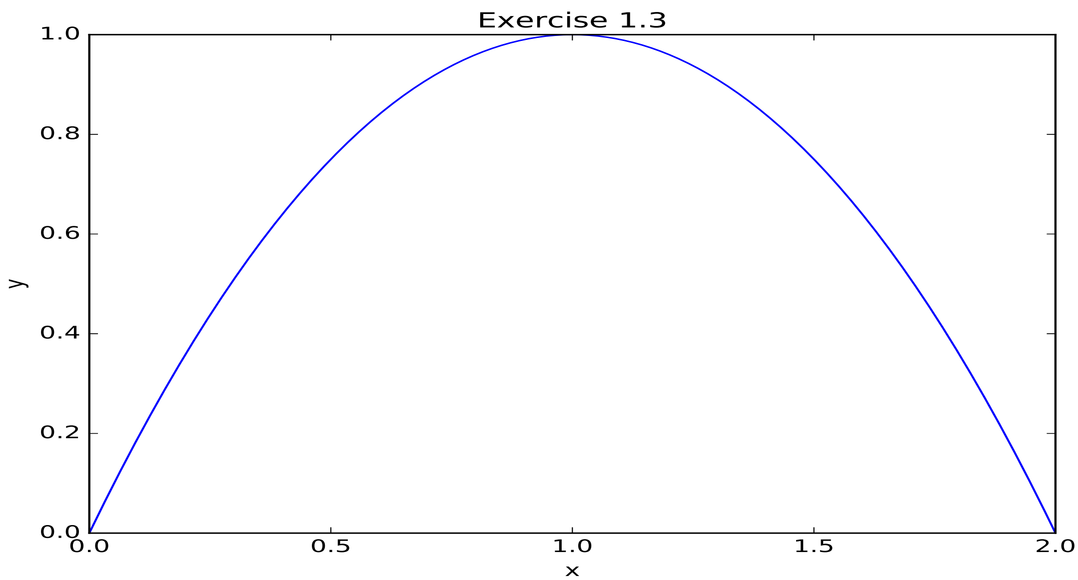


Figure 2: the function $y = -x^2 + 2x$

2. Let X_n be the discrete uniform on $\{1/2^n, 2/2^n \dots 1\}$. Prove that X_n converges to $U(0,1)$ in laws.

Proof: Suppose we are considering a point $x \in [0,1]$ such that $m/2^n \leq x < (m+1)/2^n$. Then for the random variable X_n,

$$P(X_n \leq x) = m/2^n, \tag{6}$$

and for the uniformly distributed random variable $X \sim U(0,1)$, we have:

$$P(X \leq x) = x. \tag{7}$$

Since $|m/2^n - x| \leq |m/2^n - (m+1)/2^n| = 1/2^n \to 0$, as $n \to +\infty$, it follows that the statement is true. Based on this result, we say that in order to generate a uniform distribution, we flip the coin for n times then we get a sequence of numbers, which consist of 0's and 1's and the length of the sequence is n. Now, since any sequence of 0's and 1's represents a binary representation of a unique number, ranging from 0 to $2^n - 1$, what we do is we simply map that binary representation to that number in $\mathbb{Z}$, and then we basically recover the uniform distribution, for n that is large enough.

3. Fix $\lambda > 0$. For $n > \lambda$, let X_n be $Binomial(n, \lambda/n)$. Prove that X_n converges in distribution to a $Poisson(\lambda)$.

Proof: Since the pdf for the Possion is $pdf = \frac{\lambda^k e^{-\lambda}}{k!}$, it follows that the MGF of the Possion is

$$\mathbb{E}(e^{tx}) = \sum_{k=0}^{+\infty} \frac{e^{tk}\lambda^k e^{-\lambda}}{k!}. \tag{8}$$

After simplification, we obtain

$$\mathbb{E} = e^{-\lambda} \sum_{k=0}^{+\infty} \frac{(\lambda e^t)^k}{k!} = e^{\lambda(e^t - 1)}. \tag{9}$$

In the mean time, since $X_n \sim B(n, \lambda/n)$, we learn that

$$P(X_n = k) = C_n^k (\lambda/n)^k (1 - \lambda/n)^{n-k}. \tag{10}$$

Now, it follows that

$$\begin{aligned} \mathbb{E}(e^{tX_n}) &= \sum_{k=0}^{n} e^{tk} C_n^k (\lambda/n)^k (1-\lambda/n)^{n-k} \\ &= \left(\frac{e^t \lambda}{n} + 1 - \frac{\lambda}{n} \right)^n. \end{aligned}$$

By letting $n \to +\infty$, we obtain that

$$\lim_{n \to +\infty} \left[\left(1 + \frac{1}{n/(\lambda(e^t - 1))} \right)^{n/(\lambda(e^t-1))} \right]^{\lambda(e^t-1)} = e^{(e^{\lambda t} - 1)},$$

where we have used the limit definition of e. Therefore, we conclude that the MGF is the same as that of the Possion's. Hence, the proof is finished.

4. Fix $\lambda > 0$. For $n > \lambda$ let X_n be $Geometric(\lambda/n)$. Show that X_n/n converges in distribution to an $Exponential(\lambda)$.

Proof: For the random variable X_n/n, its pdf is the following:

$$P(X_n/n = K/n) = (1-\lambda/n)^{K-1}\lambda/n. \tag{11}$$

Hence, it follows that the MGF for this random variable is

$$\begin{aligned} MGF &= \sum_{k=1}^{+\infty} e^{tk/n}(1-\lambda/n)^{k-1}\lambda/n \\ &= e^{t/n}\lambda/n \sum_{k=0}^{+\infty}(e^{t/n}(1-\lambda/n))^k \\ &= e^{t/n}\lambda/n \frac{1}{1-e^{t/n}(1-\lambda/n)} \end{aligned}$$

Now, by change of variable $x = 1/t$ and then using L'Hopital's rule, we have that

$$\lim_{x\to 0} \frac{\lambda x}{1-e^{tx}+\lambda x e^{tx}} = \lim_{x\to 0} \frac{\lambda}{\lambda - t} = \frac{\lambda}{\lambda - t}. \tag{12}$$

In the mean time, it is easy to check that the MGF for $Exponential(\lambda)$ is also $\lambda/(\lambda - t)$. Hence, the proof is finished.

Exercise 1.7b in the text: Suppose $a_n(X_n - \theta) \to N(0, \tau^2)$. What can be said about the limiting distribution of $|X_n|$ when $\theta \neq 0, 0$?

Solution:

Exercise 1.16a: X_i are standard Cauchy, iid. Show that $P(|X_n| > n \text{ infinitely often}) = 1$.

Proof:

$$P(|X_n| > n) = 2\int_n^{+\infty} \frac{1}{1+x^2}dx, \tag{13}$$

by symmetry. Denote $f(n) := \int_n^{\infty} \frac{1}{1+x^2}dx$. We want to show that

$$\sum_{n=0}^{+\infty} f(n) = +\infty. \tag{14}$$

Notice that for any n, $\frac{1}{1+x^2} > \frac{1}{2x^2}$ so that $\int_n^{\infty} \frac{1}{1+x^2}dx > \int_n^{+\infty} \frac{1}{2x^2}dx = \frac{1}{2n}$. Now it follows that $\sum_{n=0}^{+\infty} f(n) > \sum_{n=0}^{+\infty} \frac{1}{2n} = \infty$, since harmonic series diverges. Hence, by Borel-Cantelli, the proof is done.

Exercise 3.5 in the text: Suppose X_i are iid $Poi(\mu)$. Find the limit distribution of $\frac{1}{\bar{X}+\bar{X}^2+\bar{X}^3}$.

Solution: let $g(x) = \frac{1}{x+x^2+x^3}$. Then, since for Poisson random variables, mean is μ and var is μ as well, it follows that

$$\sqrt{n}(\bar{X}_n - \mu) \to N(0, \mu^2). \tag{15}$$

By applying DT, we have the following

$$\sqrt{n}\left(\frac{1}{\bar{x}+\bar{x}^2+\bar{x}^3} - g(\mu)\right) \to N(0, g'(\mu)^2\mu^2). \tag{16}$$

Since $g'(\mu) = -\frac{1+2\mu+3\mu^2}{(\mu+\mu^2+\mu^3)^2}$, it follows that

$$\sqrt{n}\left[\left(\frac{1}{\bar{X}+\bar{X}^2+\bar{X}^3}\right) - \frac{1}{\mu+\mu^2+\mu^3}\right] \to N\left(0, \frac{(1+2\mu+3\mu^3)^2}{(\mu+\mu^2+\mu^3)^4}\right). \tag{17}$$

Exercise 1.1 from text: By CLT, cince the variance is finite, we learn that $\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \to N(0,1)$. We may find that $S_n^2 \to \sigma^2$ by employing LLN twice. Thus, $\frac{\sqrt{n}(\bar{X}_n-\mu)}{S_n} \to N(0,1)$. Note that the left hand side statistics is just $t(n-1)$. Hence, the proof is done, in the case even when the sample is not drawn from normal distributions.

Exercise 3.11 in the text: X_i iid, mean 0, var 1. Show that $\frac{\sum_{i=1}^n X_i}{\sqrt{n}\log n} \to 0$ a.s.

Proof: If we can show that

$$\sum_{i=1}^{+\infty} \frac{2}{(\sqrt{n}\log n)^2} \tag{18}$$

converges, then by theorem 3.1 (v), the conclusion is apparent. This is true, since by integral test on positive monotonically decreasing sequence, we have that

$$\int_2^{+\infty} \frac{1}{n(\log n)^2} dn = \int_{\log 2}^{+\infty} \frac{1}{u^2} du < +\infty. \tag{19}$$

Hence, the proof is done.

Exercise 5.1 in the text: Let $X_{ni} \sim Bin(1, \theta_{ni})$, $1 \le i \le n$. suppose $\sum_{i=1}^n \theta_{ni}(1-\theta_{ni}) \to +\infty$. Show that

$$\frac{\sum_{i=1}^n X_{ni} - \sum_{i=1}^n \theta_{ni}}{\sqrt{\sum_{i=1}^n \theta_{ni}(1-\theta_{ni})}} \to N(0,1). \tag{20}$$

Proof: WE will employ Thm 5.2 in the book, and set $\delta = 2$. It follows that

$$\frac{\sum_{i=1}^{n} E[(X_{ni} - \theta_{ni})^4]}{(\sum \theta_{ni}(1-\theta_{ni}))^2} = \frac{\sum_1^n \theta_{ni}(1-\theta_{ni})(3\theta_{ni}^2 - 3\theta_{ni} + 1)}{(\sum \theta_{ni}(1-\theta_{ni}))^2}.$$

Since for the function $3x^2 - 3x + 1$, the max it can take on $[0,1]$ is 1, it follows immediately the last equation is less than $\frac{\sum \theta_{ni}(1-\theta_{ni})}{(\sum \theta_{ni}(1-\theta_{ni}))^2} \to 0$.

6. Let $f(x)$ be the pdf of $t(2)$ distribution. Let $L(x) = \int_{-x}^{x} y^2 f(y)dy$. Prove that $L(x)$ is slowly varying.

Proof: When $\mu = 2$, the t distribution is

$$g = \frac{1}{2\sqrt{2}}(1 + t^2/2)^{-3/2}. \tag{21}$$

Thus, by ignoring the constant coefficient $\frac{1}{2\sqrt{2}}$ (this term will eventually disapper later when we do fractions and it does not affect the convergence of the integral), we have

$$\begin{aligned}
L(x) &= \int_{-x}^{x} y^2 \frac{1}{1+y^2/2}^{3/2} dy \\
&= \int_{-x}^{x} y \frac{ydy}{(1+y^2/2)^{3/2}} \\
&= -2\int_{-x}^{x} yd(1+y^2/2)^{3/2} \\
&= -2\left[(1+y^2)^{-1/2} y \Big|_{-x}^{x} - \int_{-x}^{x} \frac{1}{\sqrt{1+y^2/2}} dy \right].
\end{aligned}$$

The integral in the last equation can be evaluated by using change of variables $y = \sqrt{2}\tan t$, to find that it is equal to $4\sqrt{2}\sinh^{-1}(x/\sqrt{2})$. Thus, the last equation is found to be

$$L(x) = 4\sqrt{2}\sinh^{-1}\left(\frac{x}{\sqrt{2}}\right) - 4x\sqrt{\frac{2}{x^2+2}}. \tag{22}$$

Now, it is clear that both the denominator and numerator of $L(tx)/L(x)$ are going to infinity, as x goes to infinity. We may use L'Hopital's Rule. It follows that

$$\begin{aligned}
\lim_{x\to+\infty} \frac{L(tx)}{L(x)} &= \lim_{x\to+\infty} \frac{tL'(tx)}{L'(x)} \\
&= \lim_{x\to+\infty} \frac{t \cdot 2(tx)^2 f(tx)}{2x^2 f(x)} \\
&= t^3 \lim_{x\to+\infty} \frac{f(tx)}{f(x)} \\
&= t^3 \cdot t^{-3} = 1.
\end{aligned}$$

Hence, the proof is done.

Exercise 7.2 from the text: Find the limiting distribution for $X_{[3n/4]:n}/X_{[n/4]:n}$ for the exponential, Pareto, and uniform distribution.
Solution: Since

$$[(\sqrt{n}X_{k1:n}, \sqrt{n}X_{k2:n}) - (\epsilon_{1/4}, \epsilon_{3/4})] \to N(0, \Sigma), \tag{23}$$

where Σ is the covaraice matrix, it follows that $X_{[3n/4]:n}/X_{[n/4]:n}$ would have a Cauchy distribution

$$C(a = \rho\frac{\sigma_x}{\sigma_y}, b = \frac{\sigma_x}{\sigma_y}\sqrt{1-\rho^2}). \tag{24}$$

Hence, we would only need to compute ρ, σ_x and σ_y. For example, for Uniform distribution case, $\rho = \frac{1}{4}$, $\sigma_x = \sigma_y = \frac{\sqrt{3}}{4}$, it follows that the limiting distribution is $C(\frac{1}{4}, \frac{\sqrt{15}}{4})$ so its distribution is

$$\frac{\frac{\sqrt{15}}{4}}{\pi[(x-\frac{1}{4})^2 + \frac{15}{16}]}. \tag{25}$$

In the case of Exponential distribution, the we find that $\rho = \sigma_y = \frac{1}{\sqrt{3}\lambda}$, $\sigma_x = \frac{\sqrt{3}}{\lambda}$. It follows that the limiting distribution is $C(\frac{\sqrt{3}}{\lambda}, 3\sqrt{1-\frac{1}{3\lambda^2}})$.