BIOS 7731 HW 7

Tim Vigers

1. a) Chebychev's inequality : $P[|X| \geq a] \leq \frac{E[X^2]}{a^2}$

can be used to derive the Bernoulli Law:

$$P[|\bar{X} - p| \geq \varepsilon] \leq \frac{E[(\bar{X}-p)^2]}{\varepsilon^2} = \frac{Var(\sum_{i=1}^{n} X_i)}{n^2 \varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2}.$$

Taking this formula, we set $\varepsilon = 0.1$. Because we want the probability that $\bar{X} < 0.4$ or $\bar{X} > 0.6$ to be 10%, we set the right hand side of the Bernoulli Law equal to 0.1 and solve for n:

$$\frac{p(1-p)}{n\varepsilon^2} = 0.1 = \frac{(0.5)^2}{n(0.1)^2}$$

So $\frac{1}{n} = \frac{(0.1)^3}{(0.5)^2} = \frac{0.001}{0.25}$, so n must be

≥ 250.

b) For the normal approximation, we start with the central limit theorem which tells us that

$$\bar{X} \longrightarrow N(p, p(1-p)/n) = N(0.5, 0.25/n)$$

And

$$\frac{\sqrt{n}(\bar{X} - 0.5)}{\sqrt{0.25}} \sim N(0,1)$$

Then we set:

$$P\left(\frac{\sqrt{n}\,(\bar{x}-0.5)}{\sqrt{0.25}} \leq \frac{\sqrt{n}\,(0.4-0.5)}{\sqrt{0.25}}\right) = 0.9$$

Then solve for n:

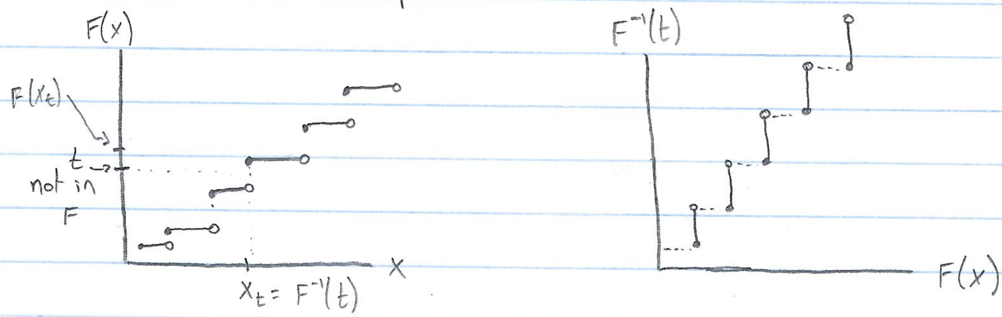
$$\frac{\sqrt{n}\,(0.1)}{\sqrt{0.25}} = 1.645$$

So with this approximation $n \approx 68$, which is much lower than the estimate based on Chebychev.

2\. a) If we let $Z(s) = s$, then it's clear that Z_n converges in probability to Z. As $n \to \infty$, the interval of the indicator approaches 0, so the sequence converges to $s = Z$, because $\lim_{n\to\infty} P(|Z_n - Z| \geq \varepsilon) = 0$.

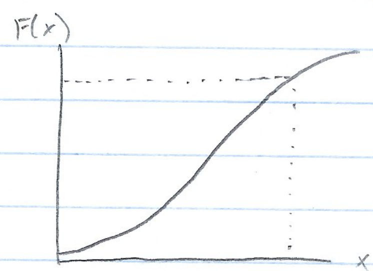
b) Although the sequence converges in probability, it does not converge almost surely. There is no value of s such that $Z_n(s) \to Z$. For every possible value of s, the value of Z_n alternates between s and s + 1. In other words, $P(\lim_{n\to\infty} |Z_n - Z| < \varepsilon) \neq 1$. However, because the sequence converges in probability, this implies that a subsequence can be found that converges almost surely.

3. a) Let $x_t = F^{-1}(t)$. By the right continuity of F, $F(x_t) \geq t$ (proved in class). As a visual demonstration plot $F(x)$ and its inverse:

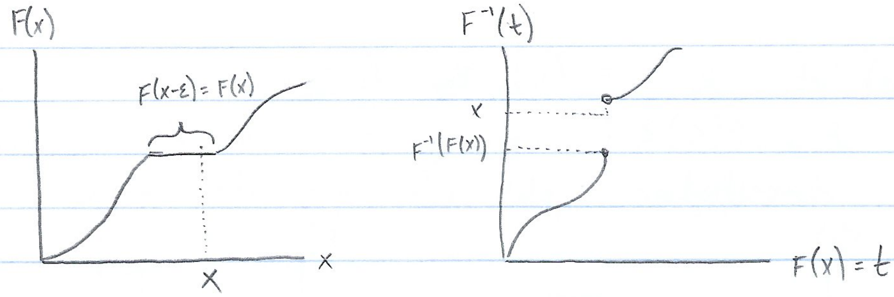


I realize the plot of $F^{-1}(t)$ is bizarre notation, but in that plot solid lines indicate areas where the function is undefined. I did this to make the connection between the plots more obvious. If we pick a value of t such that F is discontinuous at that point, then mark the corresponding point $x_t = F^{-1}(t)$ on the x-axis, it's clear that $F(x_t) > t$ because the CDF is nondecreasing and right continuous. If t is in the range of F, we simply have $F(F^{-1}(t)) = F(x) = t$ by defintion of the inverse CDF.

b) This is similar to above, but the idea is that the inequality only applies when there are discontinuities. If $F(x)$ is defined for all x in $(-\infty, \infty)$, and strictly increasing, it naturally follows that $F^{-1}(F(x)) = x$ for all x by the definition of the inverse CDF:



However, if there are any discontinuities or flat sections in the CDF, then the inverse CDF will be undefined there:



Then, because the inverse CDF is left-continuous, the inequality "kicks in" and we have $F^{-1}(F(x)) < x$ (see plots above).

BD 5.3.13. a) From the Delta method we know that if $\sqrt{n}\,(g(Y_n) - g(\theta)) \to N(0, \sigma^2 g'(\theta)^2)$ if $\sqrt{n}\,(Y_n - \theta)$ converges to $N(0, \sigma^2)$ in distribution.
So, we can rewrite $\sqrt{S_n} - \sqrt{n}$ as:

$$\sqrt{S_n} - \sqrt{n} = \sqrt{n}\left(\frac{\sqrt{S_n}}{\sqrt{n}} - 1\right)$$

In this case $g(x) = \sqrt{x}$, so by the Delta method:

$$\sqrt{n}\left(\frac{\sqrt{S}}{\sqrt{n}} - 1\right) \to N\left(0, \sigma^2 \tfrac{1}{2\sqrt{\theta}}\right) = N(0, \tfrac{1}{2}).$$

b) From a) we know that $\sqrt{n}\,(S_n/n - 1) \xrightarrow{P} N(0, \sigma^2)$.
So we can set $g(x) = \sqrt{2x}$. So,

$$\sqrt{n}\left(\sqrt{\frac{2S_n}{n}} - \sqrt{2\theta}\right) \to N(0, \sigma^2 g'(\theta)^2)$$

$$\to N\left(0, \frac{\sigma^2}{\sqrt{2\theta}}\right) = N(0, \tfrac{1}{\sqrt{2}})$$

Thus, we know that

$$\sqrt{n}\left(\sqrt{\frac{2S_n}{n}} - \sqrt{2}\right) \longrightarrow N(0,1)$$

and $\sqrt{2S_n} - \sqrt{2n} \longrightarrow N(0,1)$. Therefore,

$$P(S_n \le x) \approx \Phi\left(\sqrt{2x} - \sqrt{2n}\right).$$

BIOS 7731 HW 7

Tim Vigers

28 October 2020

BD 5.3.13

c)

For each n (degrees of freedom), find the critical value. Plug the critical value and n into the two approximations from above and compare.

```r
# Vector of n
n = c(5,10,25)
# Get critical values
q9 = qchisq(0.9,df=n)
q99 = qchisq(0.99,df=n)
# CLT approximations
clt9 = pnorm((q9-n)/sqrt(2*n))
clt99 = pnorm((q99-n)/sqrt(2*n))
# Part b
b9 = pnorm(sqrt(2*q9)-sqrt(2*n))
b99 = pnorm(sqrt(2*q99)-sqrt(2*n))
# Make tables
t9 = cbind(n,b9,clt9)
kable(t9,caption = "90th %ile",col.names = c("n","Part b)","CLT"),
      digits = 3)
```

Table 1: 90th %ile

n	Part b)	CLT
5	0.872	0.910
10	0.881	0.910
25	0.889	0.908

```r
t99 = cbind(n,b99,clt99)
kable(t99,caption = "99th %ile",col.names = c("n","Part b)","CLT"),
      digits = 3)
```

Table 2: 99th %ile

n	Part b)	CLT
5	0.99	0.999
10	0.99	0.998
25	0.99	0.997

For the $x_{0.90}$ case, the approximation from part b) slightly underestimates the probability while the CLT approach slightly overestimates. Both seem to perform well, however, and get very close to the correct value as n increases. In the $x_{0.99}$ case, the approximation from part b) is correct for every value of n, while the CLT approximation is too large. The CLT approximation again improves as n increases, but I think the approximation from part b) is better overall.