Lecture 8

Ciprian Crainiceanu

Department of Biostatistics
Johns Hopkins Bloomberg School of Public Health
Johns Hopkins University

September 8, 2020

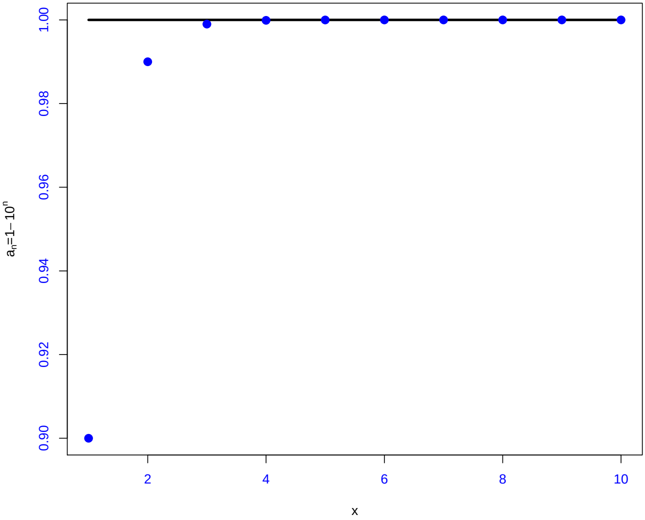
Table of contents

Outline

1. Define convergent series
2. Define the Law of Large Numbers
3. Define the Central Limit Theorem
4. Create Wald confidence intervals using the CLT

Numerical limits

- Imagine a sequence
    - $a_1 = .9$,
    - $a_2 = .99$,
    - $a_3 = .999, \ldots$
- Clearly this sequence converges to 1
- Definition of a limit: For any fixed distance we can find a point in the sequence so that the sequence is closer to the limit than that distance from that point on
- $|a_n - 1| = 10^{-n}$

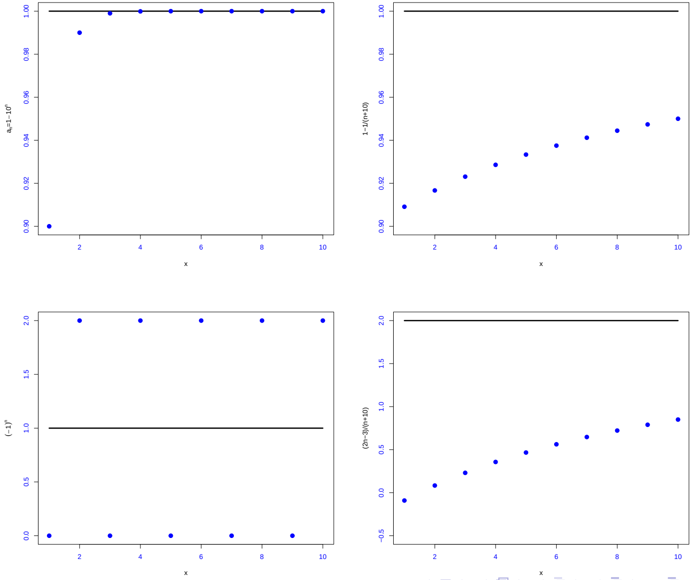


$a_n = 1 - 10^n$
1.00
0.98
0.96
0.94
0.92
0.90
2
4
6
8
10
x

More examples

- Sequence 1: $a_n = 1 - 10^n$
- Sequence 2: $a_n = 1 - \frac{1}{n+10}$
- Sequence 4: $a_n = 1 + (-1)^n$
- Sequence 3: $a_n = \frac{2n-3}{n+10}$

Examples



Limits of random variables

- The problem is harder for random variables
- Consider $\bar{X}_n$ the sample average of the first n of a collection of iid observations
    - Example $\bar{X}_n$ could be the average of the result of n coin flips (i.e. the sample proportion of heads)
- We say that $\bar{X}_n$ **converges in probability** to a limit if for any fixed distance the *probability* of $\bar{X}_n$ being closer (further away) than that distance from the limit converges to one (zero)
- $P(|\bar{X}_n - \text{limit}| < \epsilon) \to 1$

Why RVs are different

- Each experiment is different
- RVs are functions, not numbers
- Only after the experiment outcome is observed do we have a random variable realization
- Two scientists, same experiment
  - obtain different data
  - the summary of their experiments (e.g. the mean) converges to the same limit

The Law of Large Numbers

- Establishing that a random sequence converges to a limit is hard
- Fortunately, we have a theorem that does all the work for us, called the **Law of Large Numbers**
- The law of large numbers states that if $X_1, \ldots X_n$ are iid from a population with mean μ and variance σ^2 then $\bar{X}_n$ converges in probability to μ
- (There are many variations on the LLN; we are using a particularly lazy one)

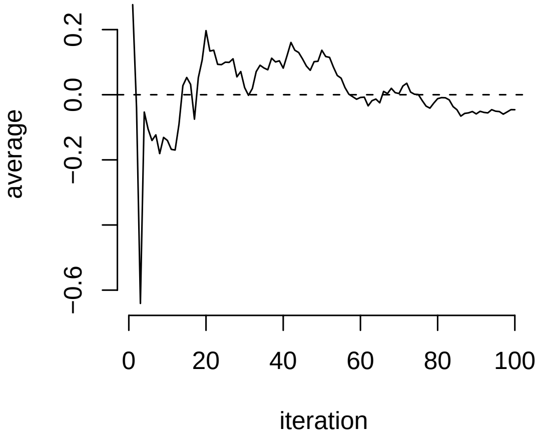
Proof using Chebyshev's inequality

- Recall Chebyshev's inequality states that the probability that a random variable variable is more than k standard deviations from its mean is less than $1/k^2$
- Therefore for the sample mean

$$P\left\{|\bar{X}_n - \mu| \geq k \text{ sd}(\bar{X}_n)\right\} \leq 1/k^2$$

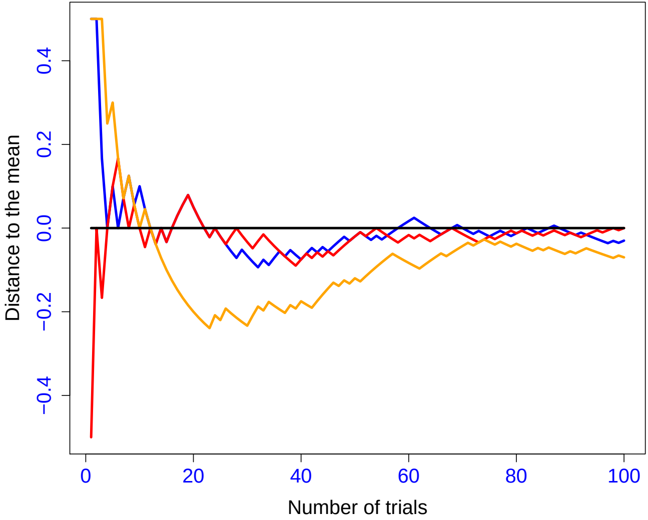
- Pick a distance ϵ and let $k = \epsilon/\text{sd}(\bar{X}_n)$

$$P(|\bar{X}_n - \mu| \geq \epsilon) \leq \frac{\text{sd}(\bar{X}_n)^2}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$$


average
0.2
0.0
−0.2
−0.6
0
20
40
60
80
100
iteration

Generating sequences of means

```
x1=rbinom(100,1,0.5)
x2=rbinom(100,1,0.5)
x3=rbinom(100,1,0.5)
xbar1=rep(0,length(x1))
xbar2=xbar1
xbar3=xbar1

for (i in 1:length(x1))
   {xbar1[i]=mean(x1[1:i])
   xbar2[i]=mean(x2[1:i])
   xbar3[i]=mean(x3[1:i])}
```


Distance to the mean
0.4
0.2
0.0
−0.2
−0.4
0
20
40
60
80
100
Number of trials

The strength of the weak LLN

- Widely used in sampling/polling
- A main reason why Nate Silver (and other Statisticians) was right in the 2012 presidential election when all the "pundits" were wrong
- He might have been a bit "too right"

```
sum(rbinom(50,1,0.8)<1)
```

- A main reason why big data is over-hyped
- Data and scientific complexity $>>>$ Data size

Convergence of transformed data

- If $X_1, \ldots, X_n$ are iid random variables then

$$\frac{1}{n}\sum_{i=1}^{n} f(X_i) \to E[f(X)]$$

- $E[f(X)]$ needs to exists; otherwise, no go
- $\frac{1}{n}\sum_{i=1}^{n} X_i^2 \underset{n}{\longrightarrow} E[X^2]$
- $\frac{1}{n}\sum_{i=1}^{n} X_i^3 \underset{n}{\longrightarrow} E[X^3]$
- $\frac{1}{n}\sum_{i=1}^{n} \exp(X_i) \underset{n}{\longrightarrow} E[\exp(X)]$
- $\frac{1}{n}\sum_{i=1}^{n} \sin(X_i) \underset{n}{\longrightarrow} E[\sin(X)]$

Useful facts

- Functions of convergent random sequences converge to the function evaluated at the limit
- This includes sums, products, differences, ...
- Example: $\bar{X}_n^2$ converges to μ^2
- Notice that this is different than $(\sum X_i^2)/n$ which converges to $E[X_i^2] = \sigma^2 + \mu^2$
- We can use this to prove that the sample variance converges to σ^2

Continued

$$
\begin{aligned}
\sum (X_i - \bar{X}_n)^2/(n-1) &= \frac{\sum X_i^2}{n-1} - \frac{n(\bar{X}_n)^2}{n-1} \\
&= \frac{n}{n-1} \times \frac{\sum X_i^2}{n} - \frac{n}{n-1} \times (\bar{X}_n)^2 \\
&\xrightarrow{p} 1 \times (\sigma^2 + \mu^2) - 1 \times \mu^2 \\
&= \sigma^2
\end{aligned}
$$

Hence we also know that the sample standard deviation converges to σ

Quizz!!!

- Example of a sequence of unbiased estimators that is not convergent?
- Example of a convergent sequence of estimators that are not unbiased?

Discussion

- An estimator is **consistent** if it converges to what you want to estimate
- The LLN basically states that the sample mean is consistent
- We just showed that the sample variance and the sample standard deviation are consistent as well
- Recall also that the sample mean and the sample variance are unbiased as well
- (The sample standard deviation is not unbiased, by the way)

The Central Limit Theorem

- The **Central Limit Theorem** (CLT) is one of the most important theorems in statistics
- For our purposes, the CLT states that the distribution of averages of iid variables, properly normalized, becomes that of a standard normal as the sample size increases
- The CLT applies in an endless variety of settings

Convergence in distribution

- Consider a sequence of rvs $X_n, n \geq 1$. We say that X_n converges in distribution to X if

$$P(X_n \leq x) = F_n(x) \underset{n}{\longrightarrow} F(x) = P(X \leq x)$$

  for every x
- This is sometimes referred to as the *weak convergence of random variables*

The CLT

- Let $X_1, \ldots, X_n$ be a collection of iid random variables with mean μ and variance σ^2
- Let $\bar{X}_n$ be their sample average
- Then
$$P\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq z\right) \to \Phi(z)$$
- Notice the form of the normalized quantity
$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{\text{Estimate} - \text{Mean of estimate}}{\text{Std. Err. of estimate}}.$$
- We say that $Z_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$ converges in distribution to $Z \sim N(0,1)$

Example

- Simulate a standard normal random variable by rolling n six sided dice
- Let X_i be the outcome for die i
- Then note that $\mu = E[X_i] = 3.5$
- $\mathrm{Var}(X_i) = 2.92$
- SE $\sqrt{2.92/n} = 1.71/\sqrt{n}$
- Standardized mean

$$\frac{\bar{X}_n - 3.5}{1.71/\sqrt{n}}$$


1 die rolls
2 die rolls
6 die rolls
0.0
0.2
0.4
−3
−1
1
2
3

R simulations: exponential

Assume that $X_1, \ldots, X_n$ are iid with an exp(1) distribution

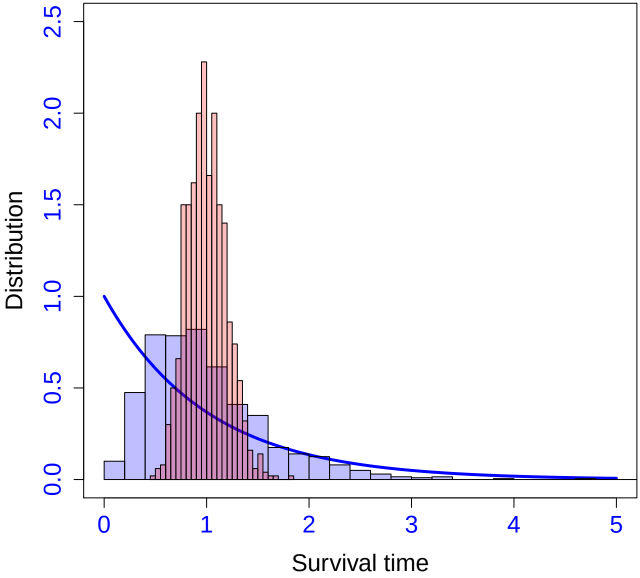
$$f(x) = \exp(-x) \text{ for } x > 0$$

- $E[X_i] = 1$, $\mathrm{Var}(X) = 1$
- Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$
- Simulate $\bar{X}_n$ for $n = 3$, $n = 30$ and plot
- Show histograms of $\bar{X}_n$ and

$$Z_n = \frac{\bar{X}_n - 1}{1/\sqrt{n}} = \sqrt{n}(\bar{X}_n - 1)$$

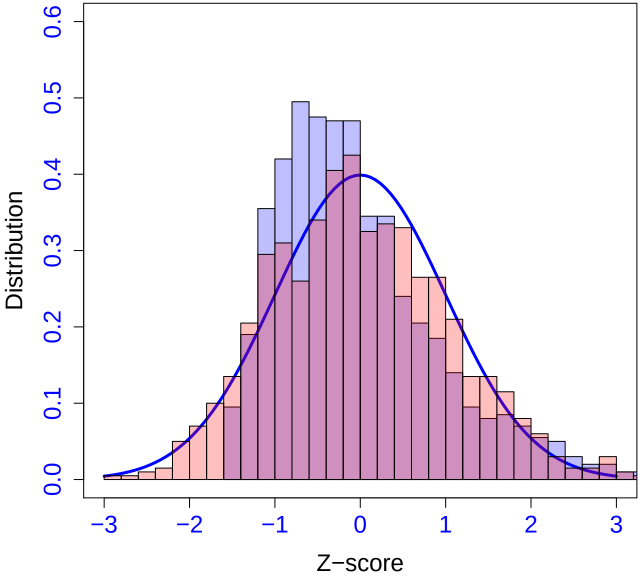
R simulations: exponential

```
xh=seq(0,5,length=101)
he=dexp(xh,rate=1)
n=c(3,30)
mx=matrix(rep(0,2000),ncol=2)

for (i in 1:1000)
   {mx[i,1]=mean(rexp(n[1], rate = 1))
   mx[i,2]=mean(rexp(n[2], rate = 1))}

plot(xh,he,type="l",col="blue",lwd=3,
   ylim=c(0,2.5))
hist(mx[,1],prob=T,add=T,col=rgb(0,0,1,1/4),
   breaks=25)
hist(mx[,2],prob=T,add=T,col=rgb(1,0,0,1/4),
   breaks=25)
```


Distribution
0.0
0.5
1.0
1.5
2.0
2.5
0
1
2
3
4
5
Survival time

R simulations: exponential Z-score

```
zx=mx for (j in 1:2)
   {zx[,j]<-sqrt(n[j])*(mx[,j]-1)}
xx=seq(-3,3,length=101)
yx=dnorm(xx)

plot(xx,yx,type="l",col="blue",lwd=3)
hist(zx[,1],prob=T,add=T,col=rgb(0,0,1,1/4),
   breaks=50)
hist(zx[,2],prob=T,add=T,col=rgb(1,0,0,1/4),
   breaks=50)
```


Distribution
0.0
0.1
0.2
0.3
0.4
0.5
0.6
−3
−2
−1
0
1
2
3
Z−score

Coin CLT

- Let X_i be the 0 or 1 result of the i^{th} flip of a possibly unfair coin
- The sample proportion, say $\hat{p}_n$, is the average of the coin flips
- $E[X_i] = p$ and $\mathrm{Var}(X_i) = p(1-p)$
- Standard error of the mean is $\sqrt{p(1-p)/n}$
- Then

$$z_n = \frac{\hat{p}_n - p}{\sqrt{p(1-p)/n}}$$

will be approximately normally distributed

Coin CLT: z-score

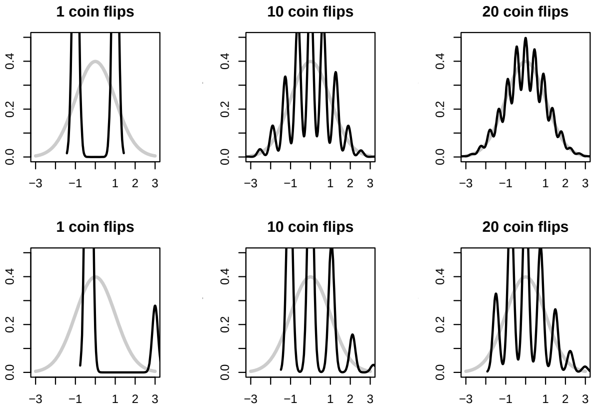
Recall that with n Bernoulli trials $\hat{p}_n$ can take the values $0/n, 1/n, \ldots, n/n$ and

$$P(\hat{p}_n = k/n) = P(\sum_{i=1}^{n} X_i = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

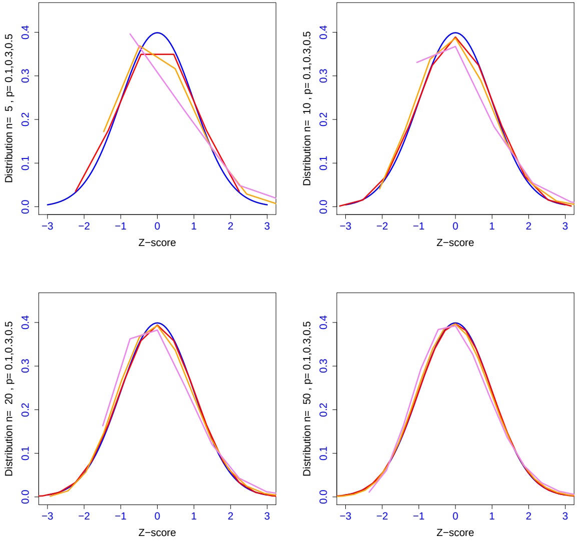
- for $n = 1$: z_n takes the values $\sqrt{(1-p)/p}$ with probability p and $-\sqrt{p/(1-p)}$ with probability $1-p$
- for a general n: z_n takes the values $\sqrt{n}\frac{k/n-p}{\sqrt{p(1-p)}}$ with probability $P(\hat{p}_n = k/n)$.

R simulations: coin flip Z-score

```
n=50
k=0:n
p=c(0.5,0.3,0.1)
values=matrix(rep(0,(n+1)*3),ncol=3)
pr=values
for (i in 1:length(p))
   {values[,i]=sqrt(n)*(k/n-p[i])/sqrt(p[i]*(1-p[i]))
   pr[,i]=dbinom(k,n,prob=p[i])/(values[2,i]-values[1,i])}

xx=seq(-3,3,length=101)
yx=dnorm(xx)
plot(xx,yx,type="l",col="blue",lwd=3)
lines(values[,1],pr[,1],lwd=3,col="red")
lines(values[,2],pr[,2],lwd=3,col="orange")
lines(values[,3],pr[,3],lwd=3,col="violet")
```


1 coin flips
10 coin flips
20 coin flips
1 coin flips
10 coin flips
20 coin flips
0.0
0.2
0.4
−3
−1
1
2
3

CLT for coin flips



CLT in practice

- In practice the CLT is mostly useful as an approximation

$$P\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq z\right) \approx \Phi(z).$$

- Recall 1.96 is a good approximation to the $.975^{th}$ quantile of the standard normal
- Consider

$$\begin{aligned} .95 &\approx P\left(-1.96 \leq \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq 1.96\right) \\ &= P\left(\bar{X}_n + 1.96\sigma/\sqrt{n} \geq \mu \geq \bar{X}_n - 1.96\sigma/\sqrt{n}\right), \end{aligned}$$

Confidence intervals

- Therefore, according to the CLT, the probability that the random interval

$$\bar{X}_n \pm z_{1-\alpha/2}\sigma/\sqrt{n}$$

  contains μ is approximately 95%, where $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal distribution
- This is called a 95% **confidence interval** for μ
- **Slutsky's theorem**, allows us to replace the unknown σ with s

Slutsky's theorem

If X_n and Y_n are random sequences, such that X_n converges in distribution to X and Y_n converges in probability to a constant c then

- $X_n + Y_n \underset{d}{\longrightarrow} X + c$
- $X_n Y_n \underset{d}{\longrightarrow} Xc$
- $X_n Y_n^{-1} \underset{d}{\longrightarrow} Xc^{-1}$

Sample proportions

- In the event that each X_i is 0 or 1 with common success probability p then $\sigma^2 = p(1-p)$
- The interval takes the form

$$\hat{p} \pm z_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

- Replacing p by $\hat{p}$ in the standard error results in what is called a Wald confidence interval for p
- Also note that $p(1-p) \leq 1/4$ for $0 \leq p \leq 1$
- Let $\alpha = .05$ so that $z_{1-\alpha/2} = 1.96 \approx 2$ then

$$2\sqrt{\frac{p(1-p)}{n}} \leq 2\sqrt{\frac{1}{4n}} = \frac{1}{\sqrt{n}}$$

- Therefore $\hat{p} \pm \frac{1}{\sqrt{n}}$ is a quick CI estimate for p