Chapter 5 - Limit Theorems

Why do we care about Limit Theorems? By limit theorems we mean what is the behavior as some quantity grows. Often we look at when the sample size (n) grows.
Ideally as our sample size grows the quantities we are using will 'converge' to their true values in some sense.
We also can derive many great results about *approximating* distributions. That is, distributions that can be used in 'large' samples.

Random Sample - $Y_1, ..., Y_n$ are a random sample (RS) of size n if they are independent and identically distributed (iid).

What does this allow us to know about the joint distribution of $Y_1, ..., Y_n$?

Statistic - Functions of $Y_1, ..., Y_n$ from a RS that don't involve unknown parameters.

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2$$

Considering a particular value of $\mu_0, T = \dfrac{\bar{Y} - \mu_0}{S/\sqrt{n}}$

Sampling Distribution - Distribution of a statistic

Example: If $Y_i \overset{iid}{\sim} N(\mu, \sigma^2)$ then the exact distribution of $\bar{Y}$ is $\bar{Y} \sim N(\mu, \sigma^2/n)$. How to prove? Recall $m_Y(t) = e^{\mu t + t^2\sigma^2/2}$.

Theory exists that says if $Y_i \overset{iid}{\sim} f_Y(y)$ where $Var(Y) < \infty$ then $\bar{Y} \overset{\bullet}{\sim} N(\mu, \sigma^2/n)$.

- If we use a exact (or true for any n) distribution for inference, we call it <u>exact inference</u>.
- If we use a 'large-sample' (or asymptotic) distribution for inference, we call it <u>approximate</u> (or <u>asympotic</u>) <u>inference</u>.

Definitions and Ideas of Convergence in Probability and Distribution

Since RVs are not defined deterministically (we may observe different observations each time we see them!), we can't consider the usual calculus convergence ideas. Instead we have different types of convergence.

3 major types of convergence of RVs (others exist)

$$\text{Almost Sure (with probability 1)} \implies \text{Probability} \implies \text{Distribution}$$

Convergence in Probability Definition - A sequence of RVs $Y_1, ..., Y_n, ...$ converges in probability to a RV Y if for every $\epsilon > 0$

$$lim_{n\to\infty} P(|Y_n - Y| \geq \epsilon) = 0 \iff lim_{n\to\infty} P(|Y_n - Y| < \epsilon) = 1$$

Denoted by $Y_n \xrightarrow{p} Y$.

We'll mostly care about convergence in probability to a constant, call it c.

$$lim_{n\to\infty} P(|Y_n - c| < \epsilon) = \lim_{n\to\infty} P(-\epsilon < Y_n - c < \epsilon) = \lim_{n\to\infty} P(c - \epsilon < Y_n < c + \epsilon) = 1$$

Convergence in probability idea -We will investigate the process of $X_n \xrightarrow{p} 0$, where $X_1, X_2, X_3, ..., X_n$ are a sequence of iid random variables that $X_n = \frac{\sum_{i=1}^{n} Y_i}{n^2}$ and $Y_i \sim N(0,1)$.

- n is the number of points in each sample path
- M is the number of sample paths
- $\epsilon = 0.05$
- Probability=$\frac{number|X_n - 0| \geq \epsilon}{M}$ for each column

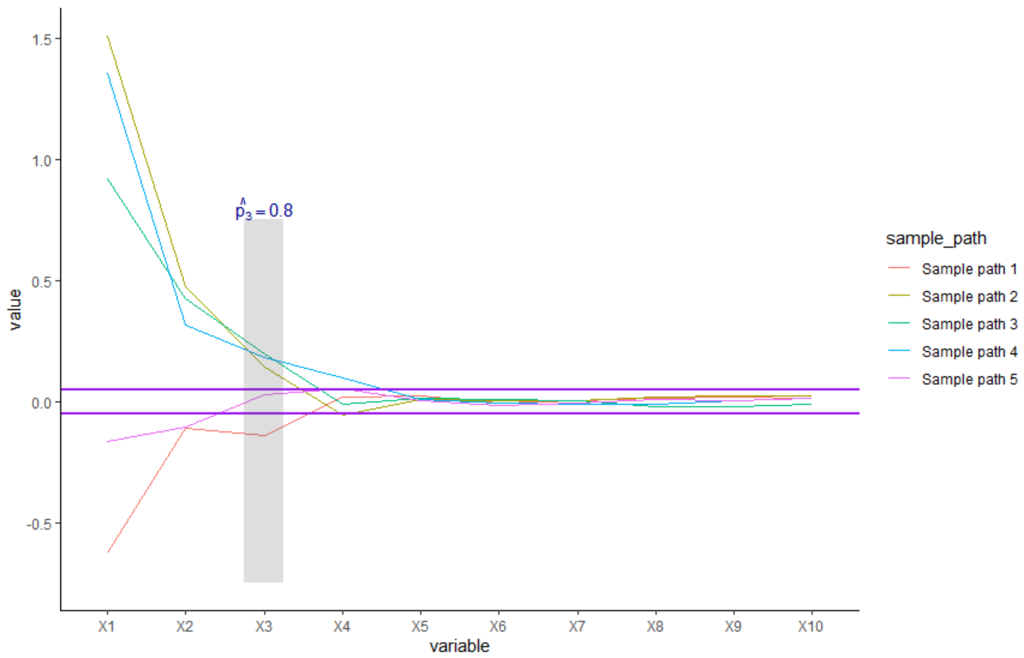
```
#Load library and set seed

Packages <- c("ConvergenceConcepts","formattable","kableExtra","dplyr", "knitr", "magrittr","reshape2","ggplot2

lapply(Packages, library, character.only = TRUE)
set.seed(1)

#Generate this many columns
n=10
#Simulate the sequence this many times
M=5
#Set up epsilon value
epsilon=0.05
#Define the function to generate the sequence
f=function(n){cumsum(rnorm(n))/((1:n)**2)}
#Generate dataset with n columns and M rows
data <- abs(round(generate(nmax=n,M=M,f)$data,digits=2))
#Extract the probability for each column that the distance between X and 0 is greater than epsilon
Pn_critr <- criterion(data,epsilon=epsilon,mode="p")$crit

#Rename the row name and column names of dateframe
data=data.frame(data)
rownames(data)=c("Sample path 1", "Sample path 2","Sample path 3","Sample path 4","Sample path 5")
colnames(data)=c("X1","X2","X3","X4","X5","X6","X7","X8","X9","X10")
data %>%
  kable(caption = "data") %>%
  kable_styling()
```

data

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
Sample path 1	-0.626	-0.111	-0.142	0.020	0.026	-0.005	0.006	0.016	0.020	0.013
Sample path 2	1.512	0.475	0.142	-0.058	0.008	0.004	0.003	0.017	0.023	0.025
Sample path 3	0.919	0.425	0.197	-0.013	0.016	0.010	0.004	-0.020	-0.022	-0.013
Sample path 4	1.359	0.314	0.183	0.099	0.009	-0.006	-0.012	-0.010	0.005	0.012
Sample path 5	-0.165	-0.104	0.031	0.052	0.006	-0.016	-0.004	0.009	0.006	0.013

Convergence in probability is looking for the probability that the $|X_n - 0| \geq \epsilon$ in each column.

```
#Add rownames to a column in order to melt
sample_path <- rownames(data)
datanew <- cbind(sample_path,datanew)
data_melted = melt(datanew, id.vars = 'sample_path')
ggplot(data_melted, aes(x = variable, y = value)) + geom_line(aes(color = sample_path, group =sample_path))+
  geom_hline(yintercept = epsilon,size=1,col="purple")+geom_hline(yintercept = -epsilon,size=1,col="purple")+
  annotate("rect", xmin = 2.75, xmax = 3.25, ymin = -0.75, ymax = 0.75,
  alpha = .2)+
  annotate("text",x = 3, y = 0.8, label = "hat(p[3])==0.8",parse = TRUE,col="darkblue")+
  theme_classic()
```



Let's find the probability of each column and summarize it in a table.

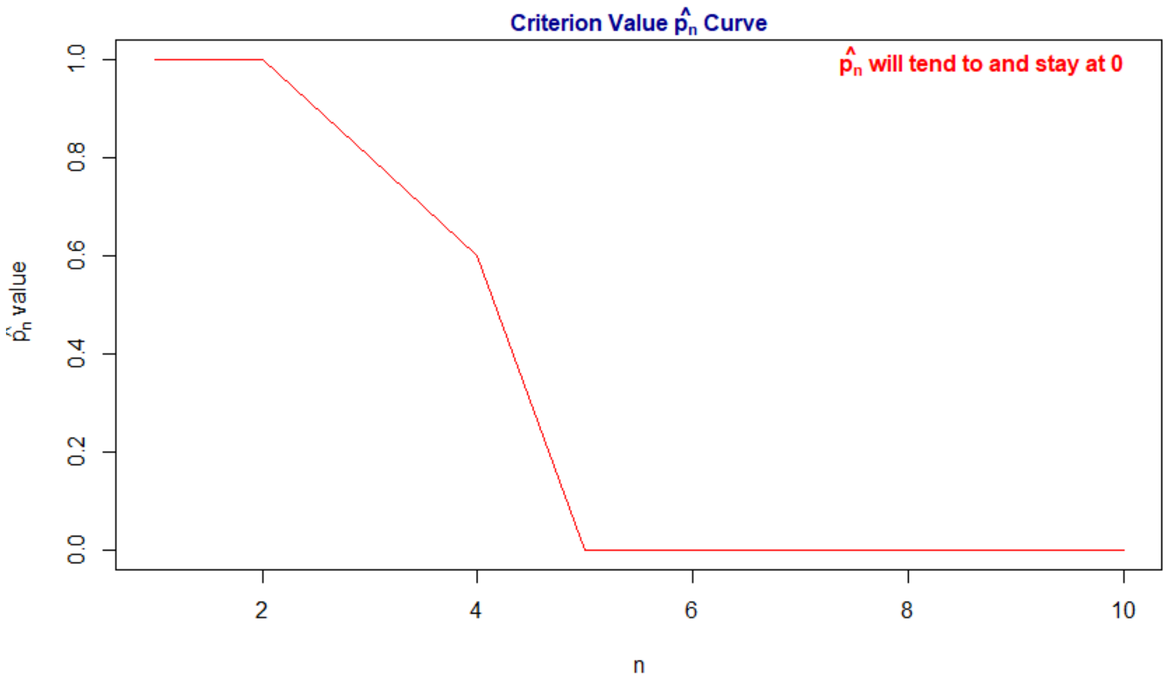
```
#Transpose column dataset probability and rename the column names
probability=t(Pn_critr)
rownames(probability)="Probability"
colnames(probability)=c("X1","X2","X3","X4","X5","X6","X7","X8","X9","X10")
#produce data table and combind probability dataset
rbind(data,probability)%>%rownames_to_column()%>%
mutate_if(is.numeric,function(x){
x=cell_spec(x,color = ifelse(x>=0.05,"red","black"))})%>%
column_to_rownames()%>%
kable(escape = F,row.names =T )%>%
kable_styling(bootstrap_options = c("striped", "hover"), full_width = T)%>%
row_spec(6, bold = T, color = "black", background = "lightblue")
```

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
Sample path 1	0.63	0.11	0.14	0.02	0.03	0	0.01	0.02	0.02	0.01
Sample path 2	1.51	0.48	0.14	0.06	0.01	0	0	0.02	0.02	0.02
Sample path 3	0.92	0.43	0.2	0.01	0.02	0.01	0	0.02	0.02	0.01
Sample path 4	1.36	0.31	0.18	0.1	0.01	0.01	0.01	0.01	0.01	0.01
Sample path 5	0.16	0.1	0.03	0.05	0.01	0.02	0	0.01	0.01	0.01
Probability	**1**	**1**	**0.8**	**0.4**	**0**	**0**	**0**	**0**	**0**	**0**

- $lim_{n\to\infty} P(|X_n - 0| \geq \epsilon) = 0$
- $X_n \xrightarrow{p} 0$

Let's graph the p_n curve.

```
#Plot the probability curve
plot(Pn_critr,xlab="n",ylab=bquote(hat(p[n])~"value"), main=mtext(bquote(bold("Criterion Value"~hat(p[n])~"Curv
text(x =n, y = max(Pn_critr),bquote(bold(hat(p[k])~"will tend to and stay at 0")),col="red",adj=1)
```



We can visualize this convergence concept from shinyapp.

```
#This is a Shiny app for visualizing convergence almost surely, in probability and in distribution. It is based

install.packages(c("shiny", "ConvergenceConcepts","shinydashboard"))

#After installing these, you can run the app from the R console using:
library("shinydashboard")
library("shiny")
library("ConvergenceConcepts")
runGitHub("XiaoxiaChampon/ConvergenceConcepts")
```

Let's prove it using the definition: