5A RandomSample

Contents

5A Random Sampling

[ToC]

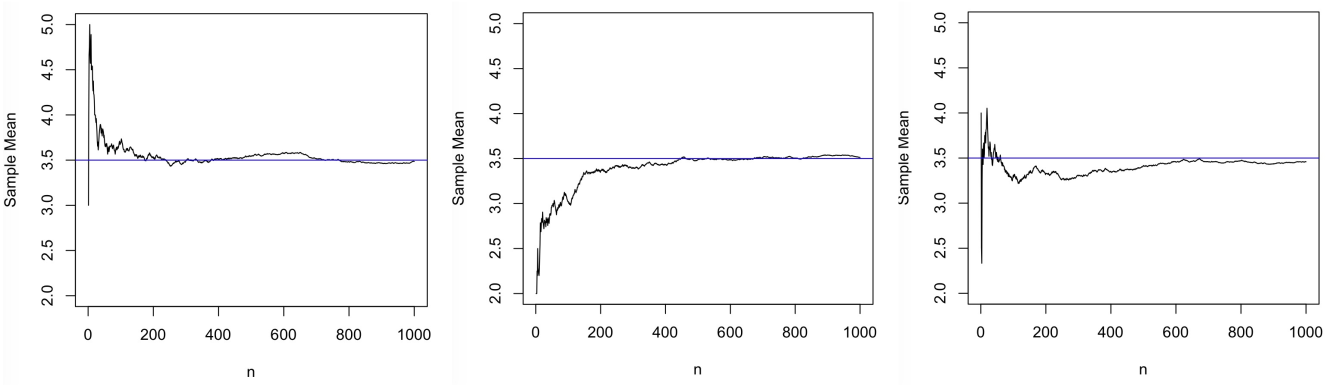
A.1 Random Sample from $F(x)$

Random variables $X_1, X_2, \ldots, X_n$ are said to be a **random sample** of size n from distribution F if

1. The X_i's are independent
2. Each X_i has distribution(CDF) $F(x)$ (or pdf $f(x)$).
   1st run of the Experiment $\rightarrow$ realization of X_1
   2nd run of the Experiment $\rightarrow$ realization of X_2
   3rd run of the Experiment $\rightarrow$ realization of X_3
   $\vdots$
3. $\{X_1\, X_2, \ldots, X_3\}$ is the dataset.
4. F is called the population distribution.

A.2 Back to Sample Mean

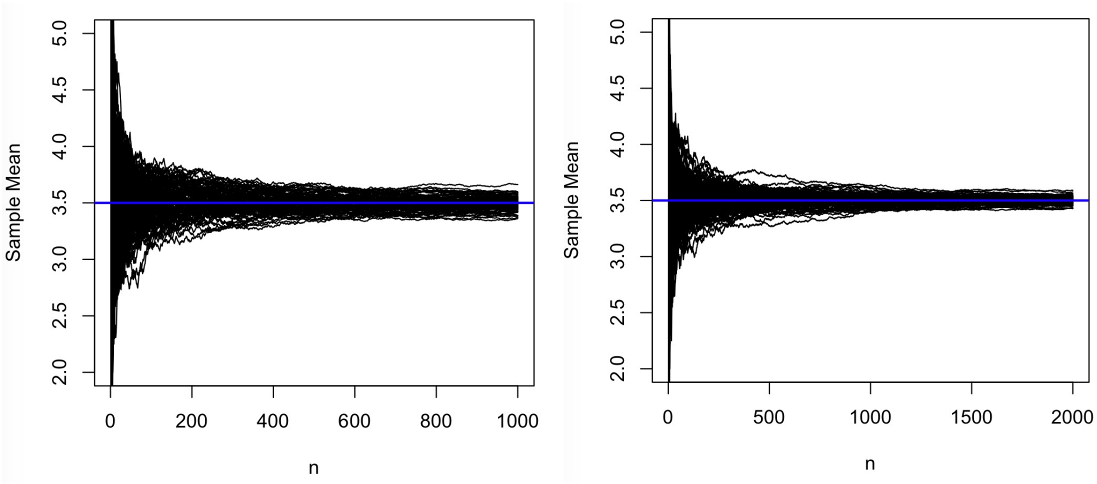
Roll a die 1000 times and follow $\bar{X}$.



- How quickly does $\bar{X}$ converge to $E(X)$?

A.3 How quickly does $\bar{X}$ converge to $E(X)$?

Overlay of 100 $\bar{X}$.



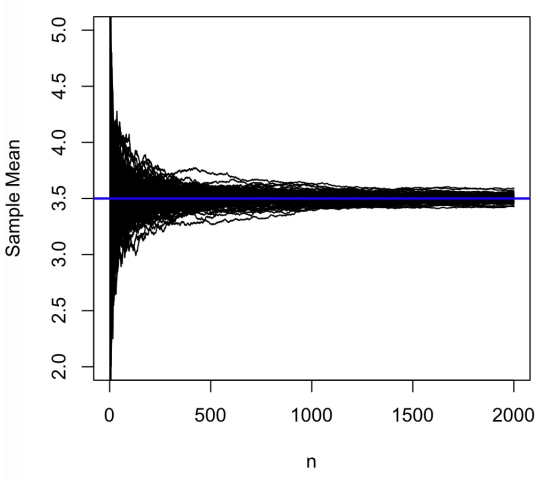
A.4 Central Limit Theorem

- When $X_1, \ldots, X_n$ are random sample from an experiment with mean μ and SD σ,

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

- We can use this information and talk about how accurately $\bar{X}$ estimates μ.
- Java Applet

http://onlinestatbook.com/stat_sim/sampling_dist/

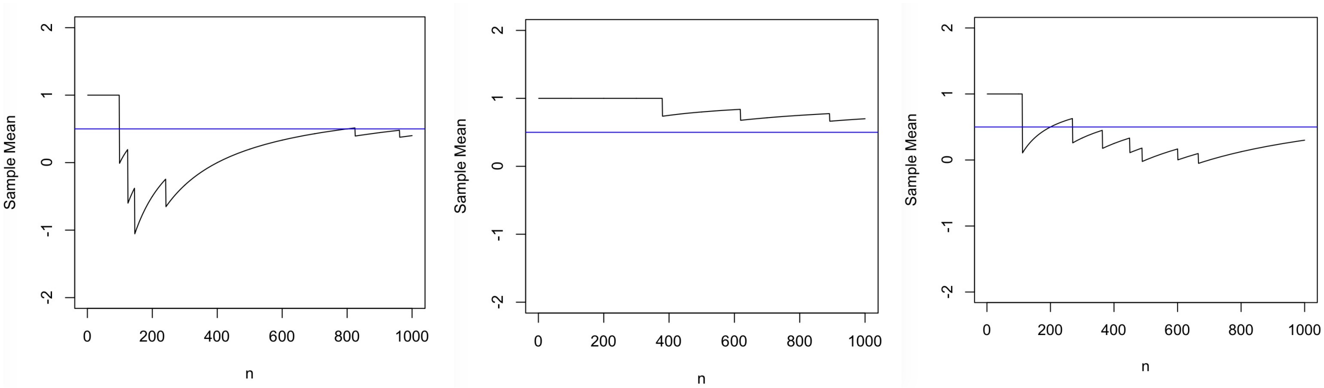


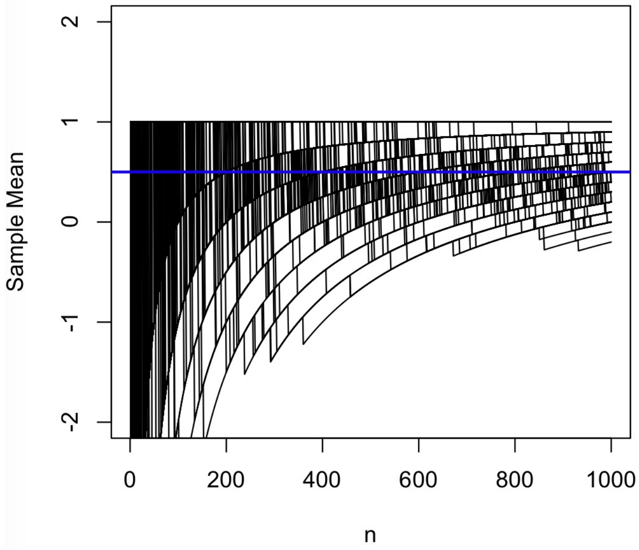
Sample Mean
5.0
4.5
4.0
3.5
3.0
2.5
2.0
0
500
1000
1500
2000
n

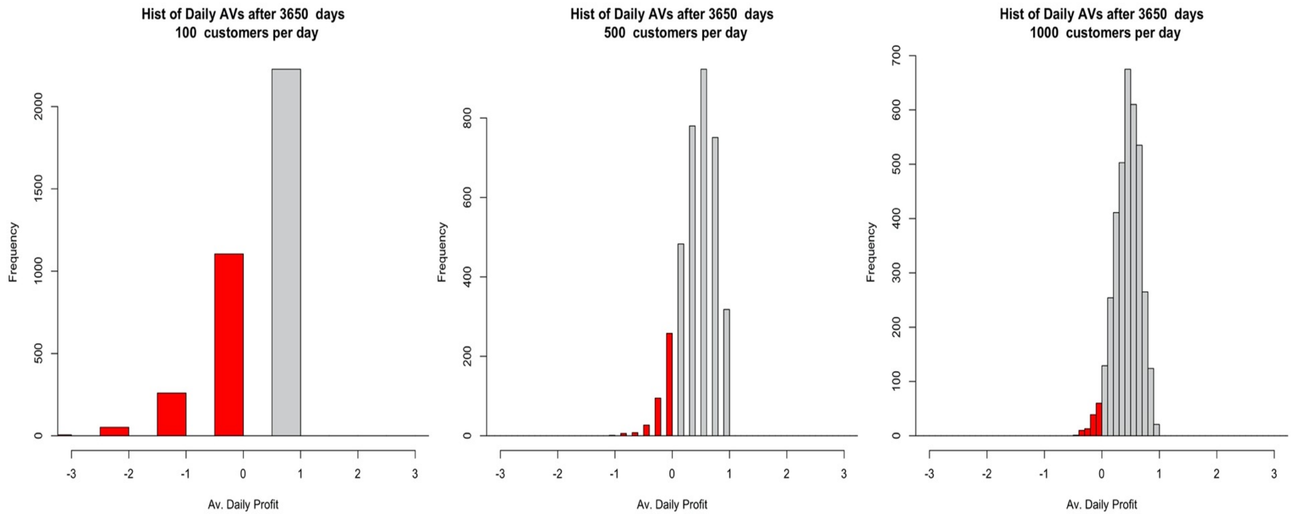
A.5 Ex: Spinning Wheel Game

Recall the game of spinning wheel, which customers pay \$1 to play a game, with .005 chance of winning \$100. Let n be a number of customers that you will get per day.

- $\bar{X}$ will represent your potential profit (loss) per customer for the day.
- $E(X) = .5$ and $V(X) = 49.75$.
- Let's study the behavior of $\bar{X}$ by simulation study.


Sample Mean
n
2
1
0
-1
-2
0
200
400
600
800
1000


Sample Mean
2
1
0
-1
-2
0
200
400
600
800
1000
n


Hist of Daily AVs after 3650 days
100 customers per day
Frequency
Av. Daily Profit
Hist of Daily AVs after 3650 days
500 customers per day
Frequency
Av. Daily Profit
Hist of Daily AVs after 3650 days
1000 customers per day
Frequency
Av. Daily Profit

A.6 Proportion of days (profit ≤ 0)

(.390)	(.116)	(.031)

A.7 Central Limit Theorem

- Law of Large Numbers

$$\bar{X} \to E(X)$$

- When $X_1, \ldots, X_n$ are random sample from any distribution, if n is large enough $(n > 30)$, approximately,

$$\bar{X} \sim N\Big(\mu, \frac{\sigma}{\sqrt{n}}\Big)$$

  μ and σ are mean and SD of one X.

- Use 1

  We can use this to see how quickly (slowly) $\bar{X} \to E(X)$ is happening.(probability)

- Use 2

  We can use this to talk about how accurately $\bar{X}$ estimates μ.(statistics)

A.8 Ex:

Suppose $X_1, \ldots, X_{30}$ are random sample from mean 5 and sd 4.

- What is $P(4.5 \leq \bar{X} \leq 5.5)$
- For what value of α, $P(\mu - \alpha \leq \bar{X} \leq \mu + \alpha) = .95$?