Random sampling

Sampling with replacement

When sampling with replacement, you put back what you just drew.

- Imagine you have a bag with 5 red, 3 blue and 2 orange chips in it. What is the probability that the first chip you draw is blue?

5 ● , 3 ● , 2 ●

$$Prob(1^{st}\text{ chip }B) = \frac{3}{5+3+2} = \frac{3}{10} = 0.3$$

- Suppose you did indeed pull a blue chip in the first draw. If drawing with replacement, what is the probability of drawing a blue chip in the second draw?

1^{st} draw: 5 ●, 3 ●, 2 ●

2^{nd} draw: 5 ●, 3 ●, 2 ●

$$Prob(2^{nd} \text{ chip } B | 1^{st} \text{ chip } B) = \frac{3}{10} = 0.3$$

- Suppose you actually pulled an orange chip in the first draw. If drawing with replacement, what is the probability of drawing a blue chip in the second draw?

1^{st} draw: 5 ●, 3 ●, 2 ●

2^{nd} draw: 5 ●, 3 ●, 2 ●

$$Prob(2^{nd} \text{ chip } B|1^{st} \text{ chip } O) = \frac{3}{10} = 0.3$$

- When drawing with replacement, probability of the second chip being blue does not depend on the color of the first chip.

$$Prob(B \mid B) = Prob(B \mid O)$$

- In addition, this probability is equal to the probability of drawing a blue chip in the first draw.

$$Prob(B \mid B) = Prob(B)$$

$\Leftrightarrow$ 1^{st} draw $\perp$ 2^{nd} draw

independent

- When drawing with replacement, draws are independent.

Independent and identically distributed (random sample)

- A sequence of random trials is independent and identically distributed (i.i.d.) if each trial has the same probability distribution as the others and all are mutually independent.

- Drawing by sampling with ~~replaces~~ replacement is i.i.d.

- Preferred by most of the statistical and machine learning methods due to its simplicity.

Sampling without replacement

When drawing without replacement you do not put back what you just drew.

- Suppose you pulled a blue chip in the first draw. If drawing without replacement, what is the probability of drawing a blue chip in the second draw?

1^{st} draw: 5 🔴 , 3 🔵 , 2 🟠

2^{nd} draw: 5 🔴 , 2 🔵 , 2 🟠

$$Prob(2^{nd} \text{ chip } B | 1^{st} \text{ chip } B) = \frac{2}{9} = 0.22$$

- If drawing without replacement, what is the probability of drawing two blue chips in a row?

1^{st} draw: 5 ● , 3 ● , 2 ●
2^{nd} draw: 5 ● , 2 ● , 2 ●

$$\begin{aligned} Prob(1^{st}\text{ chip } B) \cdot Prob(2^{nd}\text{ chip } B \,|\, 1^{st}\text{ chip } B) &= 0.3 \times 0.22 \\ &= 0.066 \end{aligned}$$

- When drawing without replacement, the probability of the second chip being blue given the first was blue is not equal to the probability of drawing a blue chip in the first draw.

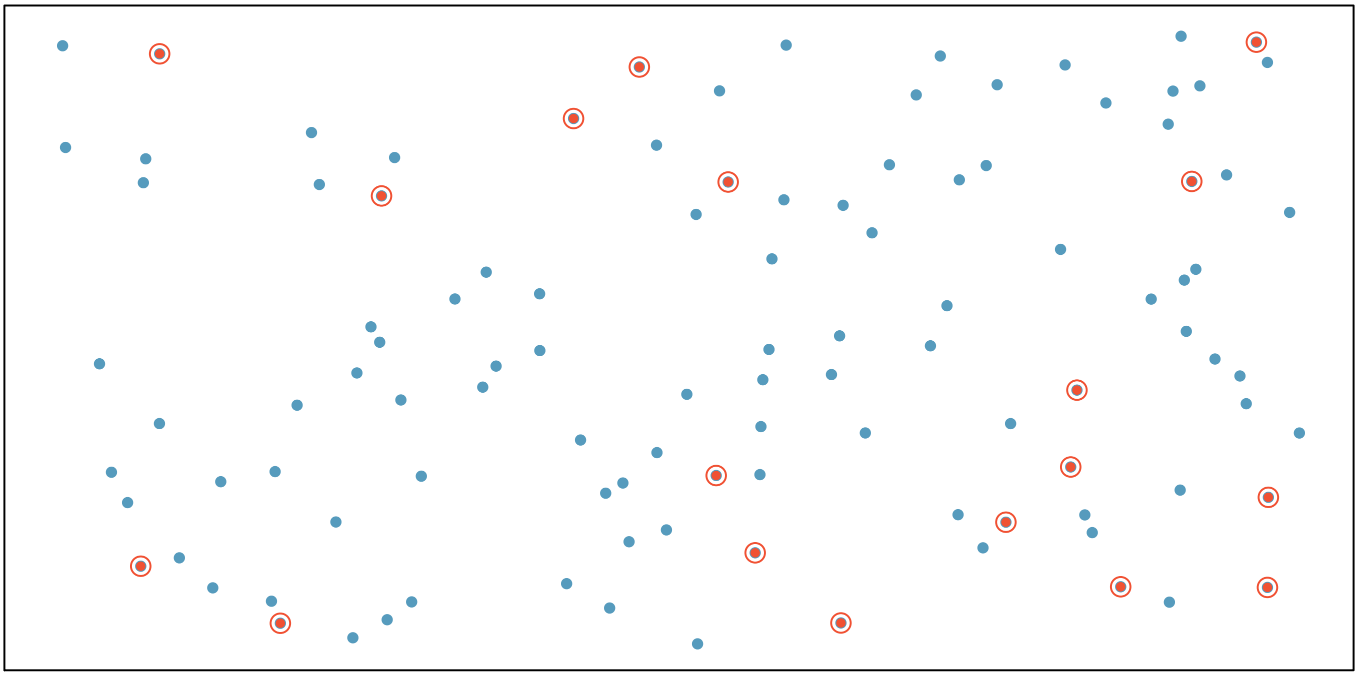
$$Prob(B \,|\, B) \neq Prob(B)$$

- When drawing without replacement, draws are not independent.

- This is especially important to take note of when the sample sizes are small. If we were dealing with, say, 10,000 chips in a (giant) bag, taking out one chip of any color would not have as big an impact on the probabilities in the second draw. $P(B|B) = \frac{2999}{9999} = 0.2999\ldots \simeq 0.3 \simeq P(B)$

- So, drawing with replacement isn't much different from drawing without replacement when the population size is big enough. 大致認.

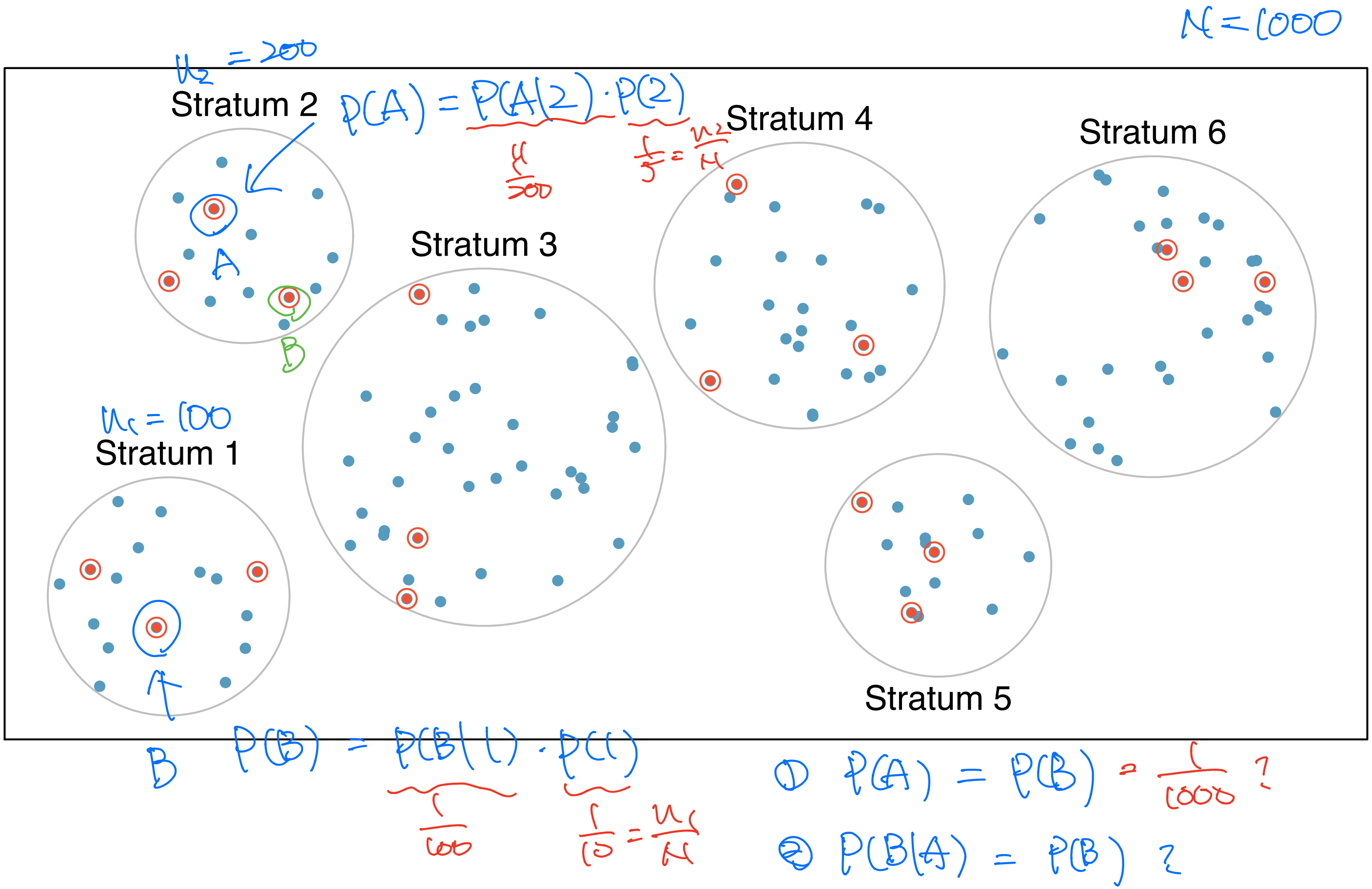
Simple random sample



sampling with/ without replacement

大致 i.i.d. (when population size is large enough)

Stratified sample



Readings

- Chapter 2.3 of OpenIntro Statistics

Homework #2: blackjack (simplified)

規則: 玩家與莊家各抽領張牌比大小

1. with v.s. without replacement

2. 抽牌順序

四種情況各自模擬100次計算並比較勝率(輸出表格)

加分題: 顯示玩家與莊家抽到的牌組 (hint)