W12 Homework B

Due date: Tuesday 4/7, 11:59pm

Summations

01 ★

Counting flip flops

A bag contains 50 marbles, 30 blue and 20 red. A sequence of zeros and ones is created by pulling the marbles out one at a time (without replacement) and writing a 1 if the marble drawn is blue and a zero if it is red.

How many pairs of adjacent digits in the sequence are expected to differ from each other?

Hint: Use a sum of 49 indicators.

Solution >

(1) Let $X_1, \dots, X_{49}$ be a sequence of indicators where X_i is the event that the i-th entry differs from the $i+1$-th entry. By a similar argument to above, the X_i are identically distributed for each i, and are, in fact, independent.

---

(2) Let S be the number of pairs of entries that differ from each other. Then $S = \sum_{i=1}^{49} X_i$. By the above,

$$E[S] = 49E[X_1] = 49\Big(0 \cdot P[X_1 = 0] + 1 \cdot P[X_1 = 1]\Big)$$

---

(3) Let Y_i denote the i-th entry. Now,

$$P[X_1 = 1] = P[Y_1 = 1, Y_2 = 0] + P[Y_1 = 0, Y_2 = 1]$$

$$= \frac{30}{50} \cdot \frac{20}{49} + \frac{20}{50} \cdot \frac{30}{49} = \frac{24}{49}$$

Thus, $E[S] = 49 \cdot \frac{24}{49} = 24$.

Central Limit Theorem

02 ★

Community college ages

At a community college, the mean age of the students is 22.3 years, and the standard deviation is 4 years. A random sample of 64 students is drawn.

What is the probability that the average age of the students in the random sample is less than 23 years?

Solution >

Sample mean RV:

$\bar{X} = M_{64}(X)$ has $\mu = 22.3$ and $\sigma = 4/\sqrt{(64)} = 0.5$.

Standardize this and approximate with a normal distribution:

$$\begin{aligned} &P[\bar{X} < 23] \\ &= P[Z < 0.7/0.5] \\ &= P[Z < 1.4] \\ &= 0.9192 \end{aligned}$$

03

Normal approximation - Ventilator filters

A mechanical ventilator model uses air filters that last 100 hours on average with a standard deviation of 30 hours.

How many filters should be stocked so that the supply lasts 2,000 hours with probability at least 95%? Use a normal approximation to estimate the answer.

State the reason that the normal approximation is applicable.

Solution >

In this case the normal approximation is applicable since we have a large sample size (need a large number of filters to last 2000 hours) and they follow follow independent but identical distributions.

Let X_i be how long the i-th filter lasts. Let $S = \sum_{i=1}^{n} X_i$ where we want to find n such that $P[S \geq 2000] \geq 0.95$. By normal approximation and the Central Limit Theorem, we have

$$\Phi\left(\frac{2000 - n \cdot 100}{30 \cdot \sqrt{n}}\right) \geq 0.95$$

$$\Phi\left(-\frac{2000 - n \cdot 100}{30\sqrt{n}}\right) \leq 0.05 \qquad \text{(by symmetry)}$$

$$\frac{2000 - 100n}{30\sqrt{n}} \leq \Phi^{-1}(0.05)$$

$$n \geq 22.3$$

and thus $n = 23$ filters are required.

04

Winning the lottery

Suppose a lottery game requires that you purchase a $10 game card and advertises a 10% probability of winning a prize.

Use the Central Limit Theorem to approximate the probability of winning at least 20 times when you purchase 100 of these game cards.

Solution >

$$E[X] = 100(.1) = 10$$

$$\mathrm{Var}[X] = 100(.1)(.9) = 9$$

$$P[X \geq 20] = P\left[\frac{X-10}{3} > \frac{20-10}{3}\right]$$

$$\ggg \quad \approx P\left[Z > \frac{(20-0.5)-10}{3}\right] = 0.000762$$

05 ★

Normal approximation - Heads v. tails

Flip a coin 10,000 times. Let H measure the number of heads, and T measure the number of tails. Estimate the probability that H and T are within 100 of each other.

Hint: Write an inequality for the condition, then sub a formula for T in terms of H.

Solution >

Let H and T be the number of heads and tails respectively. Then we have the following two conditions:

$$H + T = 10000$$
$$|H - T| \leq 100$$

Thus, $T = 10000 - H \implies |H - (10000 - H) \leq 100 \implies$
$|2H - 10000| \leq 100 \implies -100 \leq 2H - 10000 \leq 100 \implies 4950 \leq H \leq 5050$.

Let H_i be the event that the i-th flip is a head. Then $H_i \sim \mathrm{Ber}(\frac{1}{2})$ for each i and $H = \sum_{i=1}^{10000} H_i$.

Thus, by CLT, $H \sim \mathcal{N}(10000 \cdot 0.5, 10000 \cdot 0.5 \cdot 0.5) = \mathcal{N}(5000, 2500)$.

By the normal approximation, using the continuity correction, we have:

$$
\begin{aligned}
P(4950 \leq H \leq 5050) &\approx P(4949.5 < H < 5050.5) \\
&= \Phi\left(\frac{5050.5 - 5000}{50}\right) - \Phi\left(\frac{4949.5 - 5000}{50}\right) \\
&= \Phi(1.01) - \Phi(-1.01) \\
&= 2\Phi(1.01) - 1 \\
&\approx 2(0.8438) - 1 = 0.687
\end{aligned}
$$

06

Indicator method, exchangeability, summation rules

A class has 40 students: 24 women and 16 men. Each period the teacher selects a random student to present an exercise on the board from among those who have not presented already.

Let X count the number of times a man was chosen after 15 class periods.

(a) Find $E[X]$.

(b) Find $\mathrm{Var}[X]$.

Hint: Is X_j independent of X_i? Do you know $E[X_j]$ anyway?

Solution

(a)

Let X_i be the indicator that a man was chosen in the i-th period. Then $X_i \sim \mathrm{Ber}(\frac{16}{40})$ for each i, and the X_i are independent for each i. Let X be the total number of times a man was chosen. We can use a similar argument to Problems 2 or 3, or we can simply use linearity of expectation:

$$
\begin{aligned}
E[X] &= E[X_1 + \ldots + X_{15}] \\
&= E[X_1] + \ldots + E[X_{15}] \\
&= 15 \cdot \frac{16}{40} \\
&= 6
\end{aligned}
$$

(b)

(1) Using the standard formula for the variance of a sum of random variables, we have:

$$
\begin{aligned}
\mathrm{Var}[X] &= \mathrm{Var}[X_1 + \ldots + X_{15}] \\
&= \mathrm{Var}[X_1] + \ldots + \mathrm{Var}[X_{15}] + 2 \sum_{1 \leq i < j \leq 15} \mathrm{Cov}(X_i, X_j)
\end{aligned}
$$

Since the variables are identically distributed, their variances are equal. Thus,

$$
\mathrm{Var}[X] = 15\mathrm{Var}[X_1] + 2 \sum_{1 \leq i < j \leq 15} \mathrm{Cov}(X_i, X_j)
$$

---

(2) Now, the sum has $\binom{15}{2}$ terms, and since each X_i is identically distributed, each term is identical.

Thus,

$$2\sum_{1\le i<j\le 15}\mathrm{Cov}(X_i,X_j)=\binom{15}{2}\mathrm{Cov}(X_i,X_j)$$

for some fixed $i \neq j$.

---

(3) We then have

$$\begin{aligned}\mathrm{Cov}(X_i,X_j) &= \mathrm{Cov}(X_1,X_2)\\ &= \mathbb{E}[X_iX_j]-\mathbb{E}[X_i]\mathbb{E}[X_j]\\ &= (1\cdot P(X_iX_j=1)+0\cdot P(X_iX_j=0))-\mathbb{E}[X_i]\mathbb{E}[X_j]\\ &= P(X_1=1,X_2=1)-\left(\frac{16}{40}\right)^2\frac{16}{40}\cdot\frac{15}{39}-\frac{4}{25}\\ &= -\frac{2}{325}\end{aligned}$$

Finally, $E[X_1^2]=1^2\cdot P[X_1=1]+0^2\cdot P[X_1=0]=\frac{2}{5}$, and thus $\mathrm{Var}[X_1]=15\cdot\frac{6}{25}$
Plugging these values in, we have $\mathrm{Var}[X]=15\cdot\frac{6}{25}-\frac{4}{325}\cdot 105\approx 2.308$.