EXERCISE SET 3,
MS-A0503, FIRST COURSE IN PROBABILITY AND STATISTICS

Explorative exercises

I will expect that you study the explorative problems BEFORE the first lecture of the week. It is VERY STRONGLY RECOMMENDED that you work on them in groups. I LIKE CAPITAL LETTERS, apparently.

Problem 1. Assume we want to estimate the average height of a university student, knowing *nothing* about the distribution of heights. (In other words, if X is the random variable denoting the height of a randomly selected student, then we want to estimate $E(X)$.) Select N individuals at random, and denote their heights by $x_1, x_2, \ldots, x_N$. Which of the following three ways to estimate $E(X)$ would you say is best? Which is worst? Does your answer depend on the number of samples?

(1)
$$E(X) \approx \frac{x_1 + x_2 + \cdots + x_N}{N}$$

(2)
$$E(X) \approx \frac{x_1 + x_2 + \cdots + x_N}{N-1}$$

(3)
$$E(X) \approx x_1$$

Problem 2. Assume we want to estimate the *variance* $\mathrm{Var}X$ of the height X of a random university students, still knowing nothing about the distribution of heights. Select N individuals at random, and denote their heights by $x_1, x_2, \ldots, x_N$, and let $\bar{x} = \frac{x_1+x_2+\cdots+x_N}{N}$

(1) Argue that
$$s^2 = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N-1}$$
is a random variable.

(2) Find a formula for the expected value $E(s^2)$.

(3) Does this mean that s^2 is a good approximation for $\mathrm{Var}X$? Can you think of any other approximations of $\mathrm{Var}X$

Problem 3. Let X be the height of a random university student, as in Problems 1 and 2. Collect a group of 4-10 students and write down how tall each of you are. Using what you did in Problems 1 and 2, estimate $E(X)$ and $\mathrm{Var}X$. Keep your estimates - we will compare them to the estimates of other groups later.

Problem 4. Assume that we know that a given electronic device has a lifespan that is exponentially distributed, but we do not know the parameter λ. In order to estimate λ, we buy four identical copies of the device, and run them independently. We observe that they fail after time 1.2, 2.0, 1.2 and 4.1 respectively. Estimate the parameter λ in three different ways:

(1) Using that $E(X) = 1/\lambda$.
(2) Using that $\mathrm{Var}X = 1/\lambda^2$
(3) Challenging: Computing the joint probability density function in the point $(1.2, 2.0, 1.2, 4.1)$, as a function of λ.

Is any of these estimates better than the other in your opinion? Discuss!

Problem 5. How would you approximate the *standard deviation* of X from the N observations $x_1, \ldots x_N$? If you found a formula $\hat{\sigma}$ for this approximation, can you compute $E(\hat{\sigma})$?

Homework problems

The homework problems are reported during the second exercise session of the week. You are allowed and encouraged to work in groups, but every student should be prepared to present the solutions individually. During the last exercise session of the week, the teacher will ask you to mark what problems you have solved, and you get points according to how many problems you marked as solved. If you mark a problem as solved, however, you should also be prepared to present your solution in front of the class.

Homework 1. A roulette wheel has 38 slots. If you bet 1€ on a specified number, you either win 35€ if the roulette ball lands on that number or lose 1 if it does not. If you continually make such bets, approximate the probability that

- You are winning after 1 bet.
- You are winning after 34 bets.
- You are winning after 1,000 bets.

Homework 2. The Scholastic Aptitude Test mathematics test scores across the population of high school seniors follow a normal distribution with mean 500 and standard deviation 100. If five seniors are randomly chosen, find the probability that

(1) All scored below 600.
(2) Exactly three of them scored above 640.

Homework 3. Twelve percent of the population is left-handed. Find the probability that there are between 10 and 14 left-handers in a random sample of 100 members of this population.

Homework 3 (from last week). An IQ test produces scores that are normally distributed with mean value 100 and standard deviation 14.2. The top 1 percent of all scores are in what range? Hint: You probably need to look up some values of the normal distribution. I suggest you do so in Mellin's statistical tables (on the course homepage), as this is the resource you will have available on the exam.

Week 4, Homework 1

Let X_n = # times the ball lands on your number in the first n rounds.

Then $X_n \sim Bin(n, \frac{1}{38})$.

We win after n rounds iff $35X_n - 1\cdot(n - X_n) > 0$

$$\iff$$

$$36X_n > n$$

1) $\mathbb{P}[\text{winning after 1 round}] = \mathbb{P}[X_1 = 1] = \frac{1}{38} \approx 0.03$

2) $\mathbb{P}[\text{winning after 34 rounds}] = \mathbb{P}[X_{34} \geq 1]$

$$= 1 - \left(1 - \frac{1}{38}\right)^{34}$$

$$= 1 - \left(\frac{38}{37}\right)^4 \cdot \left(1 - \frac{1}{38}\right)^{38}$$

$$= 1 - \left(\frac{38}{37}\right)^4 \cdot \frac{1}{e} \approx 1 - 0.41 = 0.59.$$

3) $\mathbb{P}[\text{winning after 1000 rounds}] = \mathbb{P}[X_{1000} > \frac{1000}{36}]$.

But $\dfrac{X_{1000} - \frac{1000}{38}}{\sqrt{1000 \cdot \frac{1}{38} \cdot \frac{37}{38}}} \approx Y \sim N(0,1)$

$$\mathbb{P}\left[X_{1000} \geq \frac{1000}{36}\right] \approx \mathbb{P}\left[Y \cdot \sqrt{1000 \cdot \frac{37}{38^2}} + \frac{1000}{38} > \frac{1000}{36}\right]$$

$$= \mathbb{P}\left[Y > \frac{1000\left(\frac{1}{36} - \frac{1}{38}\right)}{\sqrt{1000}\ \frac{\sqrt{37}}{38}} = \frac{\sqrt{1000}\left(\frac{38}{36} - 1\right)}{\sqrt{37}} = \frac{1}{18}\sqrt{\frac{1000}{37}}\right]$$

$$= 1 - \Phi\left(\frac{1}{18}\sqrt{\frac{1000}{37}}\right) \approx 1 - \Phi(0.289) \approx 0.387$$

~~Homew~~ Week 4, Homework 2

$X_1 \dots X_5$ (score of the five seniors) are indep Random variables with distribution $\sim N(500, 100)$

1) $\mathbb{P}[X_i < 600 \quad i=1\dots5] = \mathbb{P}[X<600]^5 = \mathbb{P}[Y<1]$

where $Y = \frac{X-500}{100} \sim N(0,1)$

$$\mathbb{P}[X_i < 600 \quad i=1..5] = \Phi(1)^5 \approx 0.8413^5 \approx 0.42$$

2) $\mathbb{P}[\text{exactly three scores over } 640]$

$$= \binom{5}{3} \mathbb{P}[X>640]^3 \, \mathbb{P}[X \le 640]^2$$

$$= \binom{5}{3} \mathbb{P}[Y>1.4]^3 \, \mathbb{P}[Y \le 1.4]^2$$

$$= \binom{5}{3} \Phi(1.4)^3 \left[1-\Phi(1.4)\right]^2 = 10 \cdot 0.9192^3 \cdot 0.0808^2$$

$$\approx 0.051$$

Week 4, Homework 3

$X = \#\text{ left-handers} \sim Bin(100, 0.12)$

$$Y = \frac{X - E[X]}{\sqrt{Var X}} = \frac{X - 12}{\sqrt{100 \cdot 0.12 \cdot 0.88}} \approx Z \sim N(0,1)$$

$$P[10 \leq X \leq 14] = P\left[\frac{-2}{\sqrt{1.2 \cdot 8.8}} \leq Y \leq \frac{2}{\sqrt{1.2 \cdot 8.8}}\right]$$

$$\approx P\left[\frac{-2.5}{\sqrt{1.2 \cdot 8.8}} \leq Z \leq \frac{2.5}{\sqrt{1.2 \cdot 8.8}}\right]$$

(Z continuous, Y discrete, so $\frac{2}{\sqrt{1.2 \cdot 8.8}} \leq Y \iff \frac{3}{\sqrt{1.2 \cdot 8.8}} < Y$)

$$\approx P[-0.77 \leq Z \leq 0.77]$$

$$= \Phi(0.77) - \Phi(-0.77)$$

$$= 2 \cdot \Phi(0.77) - 1$$

$$\approx 2 \cdot 0.7794 - 1$$

$$= 0.5588$$

Week 3, Problem 3

$IQ \sim N(100, 14.2)$

So $\quad X = \dfrac{IQ - 100}{14.2} \sim N(0.1)$

Now $\quad 0.01 \approx \mathbb{P}[X > 2.33]$

$$= \mathbb{P}[IQ > 2.33 \cdot 14.2 + 100]$$

$$\approx \mathbb{P}[IQ > 133]$$

So the 1% of the population with highest IQ have IQ 133 or higher.