

Problem Set I - Solutions

Content: Sampling Distributions, CLT and Order Statistics

Question 1

A particular brand of dishwasher soap is sold in three sizes: 25 oz, 40 oz, and 65 oz. Twenty percent of all purchasers select a 25-oz box, 50% select a 40-oz box, and the remaining 30% choose a 65-oz box. Let X_1 and X_2 denote the package sizes selected by two independent purchasers.

- Determine the sampling distribution of $\bar{X}$, calculate $E(\bar{X})$, and compare to μ .
- Determine the sampling distribution of the sample variance S^2 , calculate $E(S^2)$, and compare to σ^2 .

Solution

The distribution of X is

$$P(X = 25) = 0.20, \quad P(X = 40) = 0.50, \quad P(X = 65) = 0.30,$$

and X_1, X_2 are independent. Hence the joint probabilities are $P(X_1 = x_1, X_2 = x_2) = P(X_1 = x_1)P(X_2 = x_2)$.

				$P(X_2 = x_2)$	$x_2 = 25$	$x_2 = 40$	$x_2 = 65$
	$x_1 = 25$	$x_1 = 40$	$x_1 = 65$	0.20	0.04	0.10	0.06
$P(X_1 = x_1)$	0.20	0.50	0.30	0.50	0.10	0.25	0.15
				0.30	0.06	0.15	0.09

- Sampling distribution of $\bar{X}$.**

$$\bar{X} = \frac{X_1 + X_2}{2}.$$

Possible values and probabilities:

$\bar{x}$	25	32.5	40	45	52.5	65
$P(\bar{X} = \bar{x})$	0.04	0.20	0.25	0.12	0.30	0.09

Then

$$E(\bar{X}) = 25(0.04) + 32.5(0.20) + 40(0.25) + 45(0.12) + 52.5(0.30) + 65(0.09) = 44.5 = \mu.$$

b. **Sampling distribution of S^2 and $E(S^2)$.**

For $n = 2$,

$$S^2 = \frac{1}{n-1} \sum_{i=1}^2 (X_i - \bar{X})^2 = \frac{(X_1 - X_2)^2}{2}.$$

Thus the possible values are:

s^2	0	112.5	312.5	800
$P(S^2 = s^2)$	0.38	0.20	0.30	0.12

(where $0.38 = 0.04 + 0.25 + 0.09$).

Hence

$$E(S^2) = 0(0.38) + 112.5(0.20) + 312.5(0.30) + 800(0.12) = 212.25 = \sigma^2.$$

Question 2

Let $Y_1, Y_2, \dots, Y_n$ be a random sample of size n from a normal distribution with mean μ and variance σ^2 .

a. Then, for a positive constant k , consider:

$$\tilde{Y} = \frac{1}{2kn} \sum_{i=1}^n Y_i$$

Show that $\tilde{Y}$ is normally distributed with mean $\mu_{\tilde{Y}} = \frac{\mu}{2k}$ and variance $\sigma_{\tilde{Y}}^2 = \frac{\sigma^2}{4k^2n}$.

b. Consider $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$. Show that $\bar{Y}$ is normally distributed with mean $\mu_{\bar{Y}} = \mu$ and variance $\sigma_{\bar{Y}}^2 = \frac{\sigma^2}{n}$.

Solution:

$$\tilde{Y} = \frac{1}{2kn} (Y_1 + Y_2 + \dots + Y_n) = a_1 Y_1 + a_2 Y_2 + \dots + a_n Y_n, \quad a_i = \frac{1}{2kn}$$

Thus, $\tilde{Y}$ is a linear combination of normally distributed random variables. Since a linear combination of normal variables is also normal, $\tilde{Y}$ is normally distributed.

$$\text{Then, } E(\tilde{Y}) = \frac{1}{2kn} \sum_{i=1}^n E(Y_i) = \frac{1}{2kn} (n\mu) = \frac{\mu}{2k},$$

$$\text{Var}(\tilde{Y}) = \sum_{i=1}^n a_i^2 \text{Var}(Y_i) = n \left(\frac{1}{2kn} \right)^2 \sigma^2 = \frac{\sigma^2}{4k^2n},$$

$$E(\bar{Y}) = \frac{1}{n} \sum_{i=1}^n E(Y_i) = \mu,$$

$$\text{Var}(\bar{Y}) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n Y_i\right) = \frac{1}{n^2} (n\sigma^2) = \frac{\sigma^2}{n}.$$

Question 3

Let $X_1, \dots, X_n$ be a random sample from $N(\mu, \sigma^2)$. Find $E(s^2)$, $V(s^2)$.

Solution:

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} \& \Rightarrow \frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$$

$$\Rightarrow E\left[\frac{(n-1)s^2}{\sigma^2}\right] = \frac{(n-1)E(s^2)}{\sigma^2} = n-1$$

$$\Rightarrow E(s^2) = \sigma^2$$

$$V\left(\frac{(n-1)s^2}{\sigma^2}\right) = \frac{(n-1)^2 V(s^2)}{\sigma^4} = 2(n-1)$$

$$\Rightarrow V(s^2) = \frac{2\sigma^4}{n-1}$$

Question 4

Let Z_1, Z_2 be a random sample of size 2 from $N(0,1)$ and X_1, X_2 a random sample of size 2 from $N(1,1)$.

Suppose the Z_i 's are independent of X_j 's. Find the sampling distribution of

a. $\bar{X} + \bar{Z}$

b. $(Z_1 + Z_2) / \sqrt{[(X_2 - X_1)^2 + (Z_2 - Z_1)^2]} / 2$

c. $[(X_1 - X_2)^2 + (Z_1 - Z_2)^2 + (Z_1 + Z_2)^2] / 2$

d. $(X_2 + X_1 - 2)^2 / (X_2 - X_1)^2$

Solution:

$$\text{a) } \bar{X} = \frac{X_1+X_2}{2} \sim N(1, 1/2) \text{ and } \bar{Z} = \frac{Z_1+Z_2}{2} \sim N(0, 1/2) \Rightarrow \bar{X} + \bar{Z} \sim N(1, 1)$$

$$\text{b) } Z_1 + Z_2 \sim N(0, 2) \Rightarrow \frac{Z_1+Z_2}{\sqrt{2}} \sim N(0, 1) \quad X_2 - X_1 \sim N(0, 2) \Rightarrow \frac{X_2-X_1}{\sqrt{2}} \sim N(0, 1)$$

$$Z_2 - Z_1 \sim N(0, 2) \Rightarrow \frac{Z_2-Z_1}{\sqrt{2}} \sim N(0, 1)$$

$$\frac{(X_2-X_1)^2}{2} \sim \chi_{(1)}^2 \text{ and } \frac{(Z_2-Z_1)^2}{2} \sim \chi_{(1)}^2 \Rightarrow \frac{(X_2-X_1)^2}{2} + \frac{(Z_2-Z_1)^2}{2} \sim \chi_{(2)}^2$$

$$\frac{(Z_1+Z_2)/\sqrt{2}}{\sqrt{\frac{(X_2-X_1)^2}{2} + \frac{(Z_2-Z_1)^2}{2}}/\sqrt{2}} \sim t_{(2)}$$

$$\text{c) } X_1 - X_2 \sim N(0, 2) \Rightarrow \frac{X_1-X_2}{\sqrt{2}} \sim N(0, 1) \Rightarrow \frac{(X_1-X_2)^2}{2} \sim \chi_{(1)}^2$$

$$Z_1 - Z_2 \sim N(0, 2) \Rightarrow \frac{Z_1-Z_2}{\sqrt{2}} \sim N(0, 1) \Rightarrow \frac{(Z_1-Z_2)^2}{2} \sim \chi_{(1)}^2$$

$$Z_1 + Z_2 \sim N(0, 2) \Rightarrow \frac{Z_1+Z_2}{\sqrt{2}} \sim N(0, 1) \Rightarrow \frac{(Z_1+Z_2)^2}{2} \sim \chi_{(1)}^2$$

$$\frac{(X_1-X_2)^2}{2} + \frac{(Z_1-Z_2)^2}{2} + \frac{(Z_1+Z_2)^2}{2} \sim \chi_{(3)}^2$$

$$\text{d) } X_2 + X_1 - 2 \sim N(0, 2) \Rightarrow \frac{X_2+X_1-2}{\sqrt{2}} \sim N(0, 1) \Rightarrow \frac{(X_2+X_1-2)^2}{2} \sim \chi_{(1)}^2$$

$$X_2 - X_1 \sim N(0, 2) \Rightarrow \frac{X_2-X_1}{\sqrt{2}} \sim N(0, 1) \Rightarrow \frac{(X_2-X_1)^2}{2} \sim \chi_{(1)}^2$$

$$\frac{\left(\frac{(X_2+X_1-2)^2}{2}\right)_{/1}}{\left(\frac{(X_2-X_1)^2}{2}\right)_{/1}} \sim F(1, 1)$$

Question 5

When a production process is operating correctly, the resistance in ohms of components produced has a normal distribution with standard deviation 3.6. A random sample of four components was taken. What is the probability that the sample variance is bigger than 30?

Solution:

We have $n = 4$, $\sigma = 3.6$, $\sigma^2 = (3.6)^2 = 12.96$

The required probability is

$$\begin{aligned} P(s^2 > 30) &= P\left(\frac{(n-1)s^2}{\sigma^2} > \frac{30(n-1)}{\sigma^2}\right) \\ &= P\left(\chi_3^2 > \frac{30(3)}{12.96}\right) = P(\chi_3^2 > 6.94) \end{aligned}$$

From the table, we find $P(\chi_3^2 > 6.25) = 0.10$ and $P(\chi_3^2 > 7.81) = 0.05$. Then, since 6.94 is between 6.25 and 7.81, the probability we require must be between 0.05 and 0.10 ; that is $0.05 < P(s^2 > 30) < 0.10$. Notice also that the sample variance is bigger than 30 if and only if the sample standard deviation is bigger than $\sqrt{30} = 5.48$. The probabilities of these two events are therefore the same. It follows that $0.05 < P(s > 5.48) < 0.10$

Question 6

A student in Turkey is studying for University Entrance Exam. She/he has created a contingent plan of solving a certain number of questions every day. For him/her to keep up with the plan, she/he needs to solve 100 questions in less than 2 hours every night. The times it takes for him/her solve a question are independent random variables with mean 1.5 minutes and variance 1.0. Claim with statistical reference that it is highly unlikely for the student to keep up with the plan.

Solution:

If we let Y_i denote the time for the student solving the i th question, then we want to have the following:

$$P\left(\sum_{i=1}^{100} Y_i \leq 120\right) = P\left(\bar{Y} \leq \frac{120}{100}\right) = P(\bar{Y} \leq 1.20)$$

Because the sample size is large, the central limit theorem tells us that $\bar{Y}$ is approximately normally distributed with mean $\mu_{\bar{Y}} = \mu = 1.5$ and variance $\sigma^2_{\bar{Y}} = \sigma^2/n = 1/100$.

$$P(\bar{Y} \leq 1.20) = P\left(\frac{\bar{Y} - 1.5}{1/\sqrt{100}} \leq \frac{1.2 - 1.5}{1/\sqrt{100}}\right) \approx P(Z \leq -3) = 0.0013$$

Question 7

A pollster believes that 20% of the voters in a certain area favor a bond issue. If 64 voters are randomly sampled from the large number of voters in this area, approximate the probability that the sampled fraction of voters favoring the bond issue will not differ from the true fraction by more than 0.06.

Solution:

If Y is a Binomial random variable with parameters n and p and if n is large enough, then Y has approximately the same distribution with W , where W is normally distributed with mean $\mu_W = np$ and variance

$\sigma^2 w = np(1-p)$. Therefore, $\frac{Y-np}{\sqrt{np(1-p)}}$ follows a standard normal distribution.

OR

If $\bar{Y}$ is a Binomial random variable with parameters n and p and if n is large enough, then $\bar{Y}$ has approximately the same distribution with U , where U is normally distributed with mean $\mu_U = p$ and variance $\sigma_u^2 = p(1-p)$. Therefore, $\frac{\bar{Y}-p}{\sqrt{p(1-p)/n}}$ follows a standard normal distribution

Let $Y = \#$ the favor the bond issue and $p = 0.2$ where $1-p = 0.8$. Then, the probability of interest is:

$$\begin{aligned} P\left(\left|\frac{Y}{n} - p\right| \leq 0.06\right) &= P\left(-0.06 \leq \frac{Y}{n} - p \leq 0.06\right) \\ P\left(\frac{-0.06}{\sqrt{\frac{p(1-p)}{n}}} \leq \frac{\frac{Y}{n} - p}{\sqrt{\frac{p(1-p)}{n}}} \leq \frac{0.06}{\sqrt{\frac{p(1-p)}{n}}}\right) &\approx P\left(\frac{-0.06}{\sqrt{\frac{0.2(0.8)}{64}}} \leq Z \leq \frac{0.06}{\sqrt{\frac{0.2(0.8)}{64}}}\right) \\ P(-1.2 \leq Z \leq 1.2) &= 0.7698 \end{aligned}$$

Question 8

The fracture strength of tempered glass averages 14 (measured in thousands of pounds per square inch) and has standard deviation 2.

- What is the probability that the average fracture strength of 100 randomly selected pieces of this glass exceeds 14.5?
- Find an interval that includes, with probability 0.95, the average fracture strength of 100 randomly selected pieces of this glass.

Solution:

Let $\bar{Y}$ denote the sample mean strength of 100 random selected pieces of glass. Thus, the quantity $(\bar{Y}-14)/0.2$ has an approximate standard normal distribution.

- $P(\bar{Y} > 14.5) = P\left(\frac{\bar{Y}-14}{0.2} > \frac{14.5-14}{0.2}\right) \approx P(Z > 2.5) = 0.0062$.
- We have that $P(-1.96 < Z < 1.96) = 0.95$. So, denoting the required interval as (a, b) such that $P(a < Y < b) = 0.95$, we have that $-1.96 = (a - 14)/0.2$ and $1.96 = (b - 14)/0.2$.

Question 9

Show that if T has a t distribution with v degrees of freedom, then $U = T^2$ has an F distribution with 1 numerator degree of freedom and v denominator degrees of freedom. ($T = Z/\sqrt{W/v}$, $F = \frac{W_1/v_1}{W_2/v_2}$. Let Z be a standard normal random variable and let W , W_1 and W_2 a chi-square distributed random variables with v , v_1 and v_2 degrees of freedom respectively. Z and W , W_1 and W_2 are independent)

Solution:

Define $T = Z/\sqrt{W/v}$. Then $T^2 = Z^2/(W/v)$. Since Z^2 has χ^2 distribution with 1 degree of freedom and Z and W independent, T^2 has an F distribution with 1 numerator and v denominator of degrees of freedom.

Question 10

A manufacturer of canned peas is concerned that the mean weight of product be close to the advertised weight. In addition, he does not want too much variability in the weights of the cans of peas; otherwise, a large proportion will differ markedly from the advertised weight. Assume that the population distribution of weights is normal. If a random sample of twenty cans is checked, find the numbers K_1 and K_2 such that

$$P\left(\frac{s_x^2}{\sigma_x^2} < K_1\right) = 0.05 \text{ and } P\left(\frac{s_x^2}{\sigma_x^2} > K_2\right) = 0.05$$

Solution:

$$0.05 = P\left(\frac{s_X^2}{\sigma_X^2} < K_1\right) = P\left[\frac{(n-1)s_X^2}{\sigma_X^2} < (n-1)K_1\right] = P\left[\chi_{(n-1)}^2 < (n-1)K_1\right]$$

Degrees of freedom: 19, (n = 20)

$$0.05 = P(\chi_{19}^2 < 19K_1)$$

From the table, $19K_1 = 10.12$,

$K_1 = 0.533$ and,

$$0.05 = P\left(\frac{s_X^2}{\sigma_X^2} > K_2\right)$$

$$0.05 = P\left[\frac{(n-1)s_X^2}{\sigma_X^2} > (n-1)K_2\right] = P\left[\chi_{(n-1)}^2 > (n-1)K_2\right]$$

From the table, $19K_2 = 30.14$

$K_2 = 1.586$

Question 11

Suppose that $n=20$ observations are to be taken on toxicology test measurement which has a normal distribution and that $\sigma^2 = 1, 4$. Let S^2 denote the sample variance of the 20 measurements.

- Find a number a such that $P(a \leq S^2) = 0.975$
- Find a number b such that $P(S^2 \leq b) = 0.975$
- If a and b are as in parts (a) and (b), what is $P(a \leq S^2 \leq b)$?

Solution:

These values can be found by using percentiles from the chi-square distribution. With $\sigma^2 = 1.4$ and $n = 20$, $\frac{19}{1.4}S^2$ has a chi-square distribution with 19 degrees of freedom.

a. Similarly, $P(S^2 \geq a) = P(\frac{n-1}{\sigma^2}S^2 \geq \frac{n-1}{\sigma^2}a) = .974$. Thus, $\frac{19}{1.4}a = 8.96055$, the 2.5% -tile of this chi-square distribution, and so $a = .656$.

b. $P(S^2 \leq b) = P(\frac{n-1}{\sigma^2}S^2 \leq \frac{n-1}{\sigma^2}b) = P(\frac{19}{1.4}S^2 \leq \frac{19}{1.4}b) = .975$. It must be true that $\frac{19}{1.4}b = 32.8523$, the 97.5% -tile of this chi-square distribution, and so $b = 2.42$.

c. $P(a \leq S^2 \leq b) = .95$.

Question 12

Briefly state the Central Limit Theorem and show that

$$\frac{1}{2n} \int_0^n e^{-x/n} dx = \frac{1}{4}$$

by using the Central Limit Theorem.

Solution:

If $X_1, X_2, X_3, \dots, X_n$ constitute a random sample from an infinite population with mean μ , variance σ^2 , and moment generating function $M_X(t)$, then the limiting distribution of

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

as $n \rightarrow \infty$ is the standard normal distribution.

In practice, this approximation is used when $n \geq 30$ regardless of the actual shape of the population sampled.

To show that

$$\frac{1}{2n} \int_0^n e^{-x/n} dx = \frac{1}{4},$$

first realize that

$$X \sim \text{Exp}(n, n^2).$$

Then

$$P(X \leq n) = \int_0^n \frac{1}{n} e^{-x/n} dx.$$

Hence,

$$P(X \leq n) = \frac{1}{n} \int_0^n e^{-x/n} dx.$$

By using the Central Limit Theorem, as $n \rightarrow \infty$ the distribution behaves like the normal distribution:

$$P\left(\frac{X - n}{n/\sqrt{n}} \leq \frac{n - n}{n/\sqrt{n}}\right) = P(Z \leq 0) = \frac{1}{2}.$$

Therefore,

$$\frac{1}{2} = \lim_{n \rightarrow \infty} \frac{1}{n} \int_0^n e^{-x/n} dx.$$

So,

$$\lim_{n \rightarrow \infty} \frac{1}{2n} \int_0^n e^{-x/n} dx = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}.$$

Question 13

Suppose that the length of time Y it takes a worker to complete a certain task has the probability density function given by

$$f(y) = \begin{cases} e^{-(y-\theta)}, & y > \theta, \\ 0, & \text{elsewhere,} \end{cases}$$

where θ is a positive constant that represents the minimum time until task completion. Let $Y_1, Y_2, \dots, Y_n$ denote a random sample of size n from this distribution. Find:

a. The density function of

$$Y_{(1)} = \min(Y_1, Y_2, \dots, Y_n).$$

Solution. For $y > \theta$,

$$F_Y(y) = P(Y \leq y) = \int_{\theta}^y e^{-(t-\theta)} dt = 1 - e^{-(y-\theta)}, \quad \Rightarrow \quad P(Y > y) = e^{-(y-\theta)}.$$

Then

$$P(Y_{(1)} > y) = P(Y_1 > y, \dots, Y_n > y) = [P(Y > y)]^n = \left(e^{-(y-\theta)}\right)^n = e^{-n(y-\theta)}, \quad y > \theta.$$

So

$$F_{Y_{(1)}}(y) = 1 - e^{-n(y-\theta)}, \quad y > \theta,$$

and differentiating gives the density

$$f_{Y_{(1)}}(y) = \frac{d}{dy} F_{Y_{(1)}}(y) = ne^{-n(y-\theta)}, \quad y > \theta,$$

and $f_{Y_{(1)}}(y) = 0$ for $y \leq \theta$.

b. $E(Y_{(1)})$.

Solution. From part (a), $Y_{(1)} - \theta$ has an exponential distribution with rate n . Hence

$$E(Y_{(1)} - \theta) = \frac{1}{n} \quad \Rightarrow \quad E(Y_{(1)}) = \theta + \frac{1}{n}.$$