Statistics - Problem Set 3

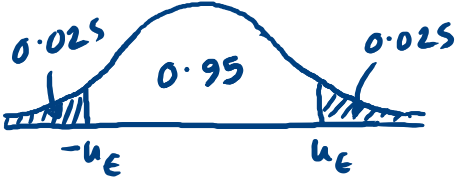
4. Find 95% confidence interval for the mean of a normal distribution with $\sigma = 3$, given the sample $(2{\cdot}3, -0{\cdot}2, -0{\cdot}4, -0{\cdot}9)$.
(Given $P(U > 1{\cdot}96) = 0{\cdot}25$ where U is a $N(0,1)$ variate)

Sol. ✓ population r.v. $X \sim N(m, \sigma)$ with $\sigma = 3$ (known)

✓ Choose statistic $u = \dfrac{\bar{x} - m}{\sigma/\sqrt{n}}$ whose sampling distribution is $N(0,1)$.

✓ 95% confidence interval : $\left(\bar{x} - \dfrac{\sigma u_\epsilon}{\sqrt{n}},\ \bar{x} + \dfrac{\sigma u_\epsilon}{\sqrt{n}}\right)$

✓ where u_ϵ is chosen s.t : $P(|U| < u_\epsilon) = 0.95$



$\Rightarrow 1 - P(|U| > u_\epsilon) = 0.95$

$\Rightarrow P(|U| > u_\epsilon) = 0.05$

$\Rightarrow P(U > u_\epsilon) = 0.025$

$\Rightarrow u_\epsilon = 1.960$

$\bar{x} = \dfrac{0.8}{4} = 0.2$, $n = 4$

Thus 95% C.I is $\left(0.2 - \frac{3 \times 1.96}{2},\ 0.2 + \frac{3 \times 1.96}{2}\right)$

$$\equiv (-2.74,\ 3.14)$$

(Ans).

(5) In a random sample of $\boxed{400}$ articles $\boxed{40}$ are found to be defective. Obtain 95% confidence interval for the true proportion of defectives in the population of such articles. Given $\frac{1}{\sqrt{2\pi}}\int_0^{1.96} e^{-\frac{x^2}{2}}\,dx = 0.4750$.

x:

$\frac{x_1 - Np}{\sqrt{Npq}}$

$X \sim \text{binomial}(400, p)$ → true proportion

$n \to X$

$\{0, 1, 2, \ldots, n\}$

random sample of size 1 : 40 x_1

$U = \frac{X_1 - Np}{\sqrt{Npq}}$ is approximately normal (0,1) when :

(asymptotically) $N = 400$

Sol. Let p be the true proportion of defectives in the population.

Here population r.v. $X \sim \text{Binomial}(400, p)$

We have a sample unit : 40 (from the population of X)

Appropriate statistic : $\boxed{u = \frac{x - Np}{\sqrt{Npq}}}$ whose sampling distribution is approximately normal (0,1).

Prob ③

Here x : observed sample, $N = 400$

95% Confindence Interval of p : $\left(\frac{x}{N} - u_\epsilon \sqrt{\frac{x(N-x)}{N^3}} , \frac{x}{N} + u_\epsilon \sqrt{\frac{x(N-x)}{N^3}} \right)$

where u_ϵ is chosen s.t. : $P(|U| < u_\epsilon) = 1.96$

$$\Rightarrow \int_{-u_\epsilon}^{u_\epsilon} \varphi_U(u)\,du = 0.95$$

$$= \frac{1}{\sqrt{2\pi}} \int_0^{u_\epsilon} e^{-\frac{x^2}{2}}\,dx = 0.4750$$

$$\Rightarrow \boxed{u_\epsilon = 1.96}$$ (from given value)

$x = 40$, $N = 400$

Therefore, 95% C.I. of p:

$$\left(\frac{40}{400} - 1.96\sqrt{\frac{40(400-40)}{400^3}},\ \frac{40}{400} + 1.96\sqrt{\frac{40(400-40)}{400^3}} \right)$$

$$\equiv (0.0706,\ 0.13)$$ (Ans) ✓

6. The probability density function of a population random variable X is given by

$$f(x;\alpha) = \begin{cases} \frac{2}{\alpha^2}(\alpha - x), & 0 < x < \alpha, \\ 0, & \text{elsewhere}. \end{cases}$$

(i) $f(x,\alpha) \geq 0$

(ii) $\int_0^\alpha f(x;\alpha)\,dx = 1$

Obtain 95% confidence limits for the parameter α on the basis of a random sample $\underline{x}$ of unit size from the population of X by using the sampling distribution of the statistic $\frac{\alpha - x}{\alpha}$.

$y = \frac{\alpha - x}{\alpha} \rightarrow Y = \frac{\alpha - X}{\alpha}$

$f_Y(y) = ?$ $\quad f_X(x)$.

To find the sampling distribution of y.

$$\frac{dy}{dx} = -\frac{1}{\alpha} < 0$$

$\hookrightarrow$ strictly dec.

$$y = \frac{\alpha - x}{\alpha}$$

if x varies from 0 to α

then y varies from 0 to 1

$$f_Y(y) = f_X(x) \left| \frac{dx}{dy} \right|$$

$$= \frac{2}{\alpha^2}(\alpha - x) \cdot \alpha$$

$$f_Y(y) = \begin{cases} 2y, & 0 < y < 1 \\ 0, & \text{elsewhere} \end{cases}$$

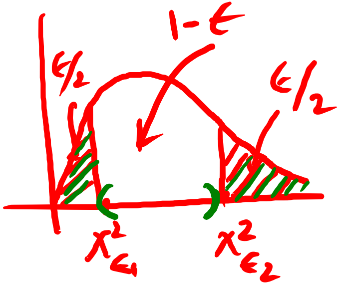
$\longrightarrow$ independent of α.

choose statistic

$$Y = \frac{\alpha - x}{\alpha}.$$

For 95% C.I. we choose two values y_1, y_2 s.t.

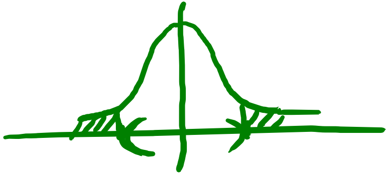
$$P(y_1 < Y < y_2) = 0.95$$



$$\epsilon = 0.05$$

We choose s.t.

$$P(Y \le y_1) = P(Y \ge y_2) = 0.025$$



$$\int_0^{y_1} 2y\,dy = 0.025 \qquad 0 < y_1 < 1$$

$$\Rightarrow y_1^2 = 0.025 \Rightarrow y_1 = \sqrt{0.025}$$

$$\int_{y_2}^{1} 2y\,dy = 0.025 \Rightarrow \int_0^{y_2} 2y\,dy = 0.975 \Rightarrow y_2 = \sqrt{0.975}.$$

$$P\left(y_1 < \frac{\alpha - X}{\alpha} < y_2\right) = 0.95$$

find the interval for α.

Sol: Population r.v. X

To find the sampling distribution of $Y = \frac{\alpha - X}{\alpha}$

$$\frac{dy}{dx} = -\frac{1}{\alpha} < 0 \quad , \quad 0 < x < \alpha \Rightarrow 0 < y < 1$$

$$f_Y(y) = \text{p.d.f. of } Y = f_X(x)\left|\frac{dx}{dy}\right| = \frac{2}{\alpha^2}(\alpha - x)\alpha$$

$$= \begin{cases} 2y , & 0 < y < 1 \\ 0 , & \text{elsewhere} \end{cases}$$

Note: Distribution of Y independent of the unknown parameter α; and it Y contains only unknown paramete α. So we choose this as statistic.

95% confidence interval of α is (y_1, y_2) where y_1, y_2 are chosen as

$$P(y_1 < Y < y_2) = 0.95$$

choose $P(Y \leq y_1) = P(Y \geq y_2) = 0.025$

$P(Y \leq y_1) = 0.025 \Rightarrow \int_0^{y_1} 2y\,dy = 0.025 \quad (\because 0 < Y_1 < 1)$

$$\Rightarrow y_1^2 = 0.025$$

$$\Rightarrow y = \sqrt{0.025} \quad \because (y_1 > 0)$$

$$P(Y \geq y_2) = 0.025$$

$$\Rightarrow \int_{y_2}^{1} 2y\,dy = 0.025 \qquad (\because 0 < y_2 < 1)$$

$$\Rightarrow 1 - y_2^2 = 0.025$$

$$\Rightarrow y_2 = \sqrt{0.975}$$

Thus $$P\left(\sqrt{0.025} < Y < \sqrt{0.975}\right) = 0.95$$

$$\Rightarrow P\left(\sqrt{0.025} < \frac{\alpha - X}{\alpha} < \sqrt{0.975}\right) = 0.95$$

$$\Rightarrow P\left(\frac{X}{1-\sqrt{0.025}} < \alpha < \frac{X}{1-\sqrt{0.975}}\right) = 0.95$$

$\therefore$ 95% C.I. $\left(\frac{X}{1-\sqrt{0.025}}, \frac{X}{1-\sqrt{0.975}}\right)$. (Ans).

⑦ The population of scores of 10-year children in a test is known to have a standard deviation 5·2. If a random sample of size 20 shows a mean 16·9, find 95% confidence limits for the mean score of the population assuming that the population is normal. Given $\frac{1}{\sqrt{2\pi}}\int_{1.96}^{\infty} e^{-\frac{x^2}{2}}\,dx = 0.25$.

$u = \dfrac{\bar{x} - m}{\sigma/\sqrt{n}}$

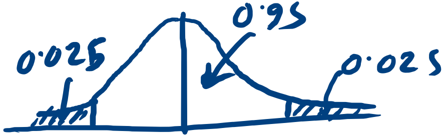
Sol. Population r.v. $X \sim N(m, \sigma)$, $\sigma = 5{\cdot}2$ (known)

To find C.I. of m choose statistic $u = \dfrac{\bar{x} - m}{\sigma/\sqrt{n}}$

95% Confidence Interval is $\left(\bar{x} - \dfrac{\sigma u_\epsilon}{\sqrt{n}},\ \bar{x} + \dfrac{\sigma u_\epsilon}{\sqrt{n}}\right)$

where $\bar{x} = 16{\cdot}9$, $n = 20$ and u_ϵ is chosen s.t.

$$P(|U| < u_\epsilon) = 0{\cdot}95$$



$\Rightarrow \quad P(U > u_\epsilon) = 0{\cdot}025$

$\Rightarrow \quad u_\epsilon = 1{\cdot}96$ (from the given value)

$\therefore$ 95% C.I. $\left(16{\cdot}9 - \dfrac{5{\cdot}2 \times 1{\cdot}96}{\sqrt{20}},\ 16{\cdot}9 + \dfrac{5{\cdot}2 \times 1{\cdot}96}{\sqrt{20}}\right) \simeq (14.62,\ 19{\cdot}18)$ Ans

8. Ten individuals are chosen at random from a normal (m, σ) population and their heights in inches are found to be: 63, 66, 63, 67, 68, 69, 70, 71, 72, 71. On the basis of the above data, obtain 95% confidence interval (i) for the parameter m when σ is unknown and (ii) for the parameter σ. Given

$$P(\chi^2(9) > 19.023) = 0.025$$

$$P(\chi^2(9) > 2.700) = 0.975$$

$$P(t(9) > 2.262) = 0.025$$

$$t = \frac{\bar{x} - m}{S/\sqrt{n}} \sim t(9)$$

$$\chi^2 = \frac{nS^2}{\sigma^2} \sim \chi^2(9)$$

Sol. $X \sim \text{normal}(m, \sigma)$

(i) To find C.I. of m when σ is unknown

Choose statistic $t = \dfrac{\bar{x} - m}{s/\sqrt{n}}$ whose sampling dist. is $t(9)$

since $n = 10$

95% C.I. : $\left(\bar{x} - \dfrac{s t_\epsilon}{\sqrt{n}},\ \bar{x} + \dfrac{s t_\epsilon}{\sqrt{n}}\right)$

where t_ϵ is chosen s.t. $P(|T| < t_\epsilon) = 0.95$

$\Rightarrow \quad P(T > t_\epsilon) = 0.025$

$\Rightarrow \quad t_\epsilon = 2.262$

To compute $\bar{x}$ and s:

x_i	$y_i = x_i - 68$	y_i^2
63	-5	25
66	-2	4
63	-5	25
67	-1	1
68	0	0
69	1	1
70	2	4
71	3	9
72	4	16
71	3	9
Total	0	94

$$\bar{x} = \frac{1}{10}\sum_{i=1}^{10} y_i + 68$$

$$= 0 + 68 = 68$$

$$S^2 = \frac{1}{10}\sum_{i=1}^{10} y_i^2 - (\bar{y})^2$$

$$= \frac{94}{10} - 0 = 9.4$$

$$s^2 = \frac{10}{9} \times 9.4 \Rightarrow s = \frac{\sqrt{94}}{3}.$$

<u>95% C.I.</u> :

$$\left(68 - \frac{\sqrt{94}}{3} \times \frac{2.262}{\sqrt{10}},\ 68 + \frac{\sqrt{94}}{3} \times \frac{2.262}{\sqrt{10}}\right)$$

$$\equiv (65.688,\ 70.312).$$

(Ans).

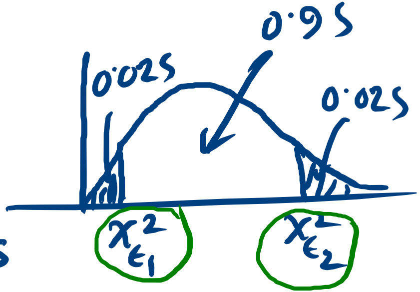
(11) To find C.I. of σ

Choose statistic $\chi^2 = \dfrac{nS^2}{\sigma^2}$ whose sampling distribution is $\chi^2(9)$.

95% C.I. : $\left(S\sqrt{\dfrac{n}{\chi^2_{\epsilon_2}}} \,,\; S\sqrt{\dfrac{n}{\chi^2_{\epsilon_1}}} \right)$

where $\chi^2_{\epsilon_1}$ and $\chi^2_{\epsilon_2}$ are chosen s.t.

$$P\left(\chi^2_{\epsilon_1} < \chi^2 < \chi^2_{\epsilon_2}\right) = 0.95$$



Choose $P\left(\chi^2 < \chi^2_{\epsilon_1}\right) = P\left(\chi^2 > \chi^2_{\epsilon_2}\right) = 0.025$

From given values, $\chi^2_{\epsilon_1} = 2.700$, $\chi^2_{\epsilon_2} = 0.975$

So, 95% C.I. $\left(\sqrt{\frac{10}{19.023}} \times \sqrt{9.4}, \sqrt{\frac{10}{2.700}} \times \sqrt{9.4}\right)$

$$= (2.223, 5.900).$$ (Ans).

9. The weights in gram of 12 items are 7, 13, 22, 15, 12, 14, 18, 8, 21, 23, 10, 17 taken at random from its population which is normal having standard deviation 5. Find 95% confidence interval for the mean of the population.

$$u = \frac{\bar{x} - m}{\sigma/\sqrt{n}}$$

Hints: Choose statistic $u = \dfrac{\bar{x} - m}{\sigma/\sqrt{n}}$

Ans $(12.171,\ 17.829)$

10. The marks obtained by 17 candidates in an examination have a mean $\bar{x}$ 57 and variance (s^2) 64. Find 99% confidence limits for the mean of the population of marks, assuming it to be normal.

σ: unknown

$$t = \frac{\bar{x} - m}{S/\sqrt{n}} \sim t(16)\text{-distribution}$$

$$\boxed{S^2 = \frac{n}{n-1} s^2}$$

Hints: Choose Statistic $t = \dfrac{\bar{x}-m}{s/\sqrt{n}}$

Ans $(51.158,\ 62.842)$.

* From table

$t_\epsilon = 2.921$

$\epsilon = 0.01$

- Basic Computational Topology

✓ - Topological Data Analysis

$$f(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}}, \quad x > 0, \quad \underbrace{\theta > 0}$$

$*$ $E(X) = \frac{1}{\theta}\int_0^\infty x e^{-\frac{x}{\theta}} dx = \theta$ (check!)

$\checkmark (x_1, x_2, \ldots, x_n)$

$$L = L(\underbrace{x_1, x_2, \ldots, x_n}; \theta) = \frac{1}{\theta^n} e^{-\frac{x_1+x_2+\cdots+x_n}{\theta}}$$

$$\log L = -n \log \theta - \frac{(x_1+x_2+\cdots+x_n)}{\theta} \qquad (x_1+x_2+\cdots+x_n) = n\bar{x}$$

$$\frac{d \log L}{d\theta} = -\frac{n}{\theta} + \frac{n\bar{x}}{\theta^2} = 0$$

$$\Rightarrow \quad \theta = \bar{x} \longrightarrow \text{local max}^{m}$$

$$\frac{d^2 \log L}{d\theta^2} = \frac{n}{\theta^2} - \frac{2n\bar{x}}{\theta^3} = \frac{n}{\theta^2}\left(1 - \frac{2\bar{x}}{\theta}\right) = \frac{n}{\theta^3}(\theta - 2\bar{x}) < 0$$

at $\theta = \bar{x}$

$\theta = \bar{x}$ is the only critical point.

$\Rightarrow \theta = \bar{x}$ is global max.

LLN

WLLN

$\hat{\theta} = \bar{x}$

$E(\hat{\theta}) = E\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) = \frac{n\theta}{\theta} = \theta$

$\hat{\theta} \xrightarrow[\text{in } P]{} \theta$ as $n \to \infty$

for any $\epsilon > 0$,

(Tchebycheff's inequality)

$P(|\hat{\theta} - \theta| > \epsilon) \leq \frac{Var(\hat{\theta})}{\epsilon^2}$ $\longrightarrow 0$ as $n \to \infty$

(*) $Var(\hat{\theta}) = E\{(\hat{\theta} - E(\hat{\theta}))^2\}$

$= E\{(\bar{X} - \theta)^2\}$

$= E\left\{\left(\frac{X_1 + X_2 + \cdots X_n}{n} - \theta\right)^2\right\} = \frac{1}{n^2} E\left\{\left(\sum_{i=1}^{n} (X_i - \theta)\right)^2\right\}$

$= \frac{n\theta^2}{n^2} = \frac{\theta^2}{n}$ (*)

(*) WLLN

$$0 \le P\left(|\hat{\theta} - \theta| > \epsilon\right) \le \frac{\theta^2}{n\epsilon^2}$$

R.H.S $\longrightarrow 0$ as $n \longrightarrow \infty$

$$\Rightarrow \lim_{n \to \infty} P\left(|\hat{\theta} - \theta| > \epsilon\right) = 0$$

$$\Rightarrow \hat{\theta} \xrightarrow[\text{in } P]{} \theta \text{ as } n \longrightarrow \infty.$$

(Alt: Use L.L.N to prove)