

## Quiz

Please find maximum likelihood estimation of  $\theta$  using sample  $X_1, \dots, X_n$ , generated from a normal distribution with parameters:

(a)  $\mu = 0, \sigma^2 = \theta^2$ ,

(b)  $\mu = \theta, \sigma^2 = 2\theta$ .

## Solution:

P.d.f. of a normal distribution with mean  $\mu$  and variance  $\sigma^2$  is:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

(a) For  $\mu = 0$  and  $\sigma^2 = \theta^2$  p.d.f. would be

$$f(x) = \frac{1}{\sqrt{2\pi}\theta} e^{-\frac{x^2}{2\theta^2}}.$$

Likelihood function:

$$\mathcal{L}(\theta) = \prod_{i=1}^n f(X_i; \theta) = \frac{1}{(2\pi)^{n/2} \cdot \theta^n} e^{-\frac{1}{2\theta^2} \sum_{i=1}^n X_i^2}.$$

Log-likelihood function:

$$l(\theta) = \log \mathcal{L}(\theta) = -\frac{n}{2} \log(2\pi) - n \log \theta - \frac{1}{2\theta^2} \sum_{i=1}^n X_i^2.$$

By necessary condition of extremum (in case of MLE – maximum):

$$\left. \frac{\partial l}{\partial \theta} \right|_{\theta=\hat{\theta}} = -\frac{n}{\hat{\theta}} + \frac{\sum_{i=1}^n X_i^2}{\hat{\theta}^3} = 0, \quad \Rightarrow \quad \boxed{\hat{\theta} = \sqrt{X^2}}.$$

(b) For  $\mu = \theta$  and  $\sigma^2 = 2\theta$  p.d.f. would be

$$f(x) = \frac{1}{2\sqrt{\pi}\theta} e^{-\frac{(x-\theta)^2}{4\theta}}.$$

Likelihood function:

$$\mathcal{L}(\theta) = \prod_{i=1}^n f(X_i; \theta) = \frac{1}{(2\sqrt{\pi})^n \cdot \theta^{n/2}} e^{-\frac{1}{4\theta} \sum_{i=1}^n (X_i - \theta)^2}.$$

Log-likelihood function:

$$l(\theta) = \log \mathcal{L}(\theta) = -n \log(2\sqrt{\pi}) - \frac{n}{2} \log \theta - \frac{1}{4\theta} \sum_{i=1}^n (X_i - \theta)^2.$$

By necessary condition of extremum (in case of MLE – maximum):

$$\left. \frac{\partial l}{\partial \theta} \right|_{\theta=\hat{\theta}} = -\frac{n}{2\hat{\theta}} + \frac{\sum_{i=1}^n (X_i - \theta)^2}{4\hat{\theta}^2} + \frac{\sum_{i=1}^n (X_i - \theta)}{2\hat{\theta}} = 0,$$

$$\hat{\theta}^2 + 2\hat{\theta} - \overline{X^2} = 0,$$

$$\hat{\theta} = -1 \pm \sqrt{1 + \overline{X^2}}.$$

Since  $\theta = \frac{\sigma^2}{2} \geq 0$ :

$$\boxed{\hat{\theta} = \sqrt{1 + \overline{X^2}} - 1}.$$

## Problem 1

Manager of a restaurant wants to estimate the mean amount  $\mu$  that a visitor spends for a lunch. A sample contains 36 visitors. Sample mean is  $\bar{x} = \$3.60$ . Manager knows that the standard deviation for one visitor is \$0.72. Find the confidence level corresponding to the interval (\$3.5; \$3.7).

### Solution:

$(1 - \alpha) \cdot 100\%$  confidence interval for  $\mu$ , when  $\sigma^2$  is known:

$$CI_{1-\alpha}(\mu) = \bar{X} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}.$$

From problem statement:

$$3.6 \pm z_{\alpha/2} \cdot \frac{0.72}{\sqrt{36}} = (3.5; 3.7),$$

Let's choose one of the boundaries (does not matter which one because of the symmetry around  $\bar{x}$ ):

$$\begin{aligned} 3.6 + z_{\alpha/2} \cdot \frac{0.72}{\sqrt{36}} &= 3.7, \\ z_{\alpha/2} &\approx 0.833. \end{aligned}$$

$z_{\alpha/2}$  is a critical value of  $\frac{\alpha}{2}$  in standard normal distribution:

$$P(Z > z_{\alpha/2}) = 1 - \Phi(z_{\alpha/2}) = \frac{\alpha}{2},$$

$$\frac{\alpha}{2} = 1 - \Phi(0.833) \approx 1 - 0.797 = 0.203,$$

$$\alpha = 2 \cdot 0.203 = 0.406.$$

Confidence level:

$$1 - \alpha = \boxed{0.594}.$$

## Problem 2

A college admission officer for an *MBA* program has determined that historically candidates have undergraduate grade point averages that are normally distributed with standard deviation 0.45. A random sample of twenty-five applications from the current year is taken, yielding a sample mean grade average of 2.90.

- (a) Find a 95% confidence interval for the population mean.
- (b) Based on these sample results, a statistician computes for the population mean a confidence interval running from 2.81 to 2.99. Find the probability content associated with this interval.

### Solution:

- (a)  $(1 - \alpha) \cdot 100\%$  confidence interval for  $\mu$ , when  $\sigma^2$  is known:

$$CI_{1-\alpha}(\mu) = \bar{X} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}.$$

From problem statement:

$$\begin{aligned} CI_{95\%}(\mu) &= 2.9 \pm z_{0.025} \cdot \frac{0.45}{\sqrt{25}} = \\ &= 2.9 \pm 1.96 \cdot \frac{0.45}{\sqrt{25}} = \\ &= 2.9 \pm 0.18 = \boxed{(2.72; 3.08)}. \end{aligned}$$

- (b) From problem statement:

$$\begin{aligned} 2.9 \pm z_{\alpha/2} \cdot \frac{0.45}{\sqrt{25}} &= (2.81; 2.99), \\ z_{\alpha/2} &= 1, \\ \frac{\alpha}{2} &= 1 - \Phi(z_{\alpha/2}) = 1 - \Phi(1) \approx 1 - 0.843 = 0.157, \\ \alpha &= 2 \cdot 0.157 = 0.314. \end{aligned}$$

Confidence level:

$$1 - \alpha = \boxed{0.686}.$$

## Problem 3

A random sample of 5 observations from a normal distribution with mean  $\mu$  and variance  $\sigma^2$  gives a sample mean 100. An independent random sample of size 10 from the same population has sample variance 9. Find a 90% confidence interval for the population mean.

### Solution:

$(1 - \alpha) \cdot 100\%$  confidence interval for  $\mu$ , when  $\sigma^2$  is unknown:

$$CI_{1-\alpha}(\mu) = \bar{X} \pm t_{n-1; \alpha/2} \cdot \frac{S}{\sqrt{n}}.$$

But the problem has two independent samples with different sizes:  $n_1 = 5$ ,  $n_2 = 10$ . So we need to adjust the formula above with specific sizes.

Original pivot function for the problem statement is:

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1},$$

where the number of degrees of freedom in  $t_{n-1}$  is derived from Fisher's lemma:

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2,$$

so the  $n$  for degrees of freedom should be taken from the sample, which is used to calculate  $S^2$  – in our case it's  $n_2$ .

$n$  in a square root from denominator is a part of E.S.E.  $(\bar{X}) = \frac{S}{\sqrt{n}}$ , so it should be taken from the sample, which is used to calculate  $\bar{X}$  – in our case it's  $n_1$ .

$(1 - \alpha) \cdot 100\%$  confidence interval is then:

$$CI_{1-\alpha}(\mu) = \bar{X} \pm t_{n_2-1; \alpha/2} \cdot \frac{S}{\sqrt{n_1}}.$$

From problem statement:

$$\begin{aligned} CI_{90\%}(\mu) &= 100 \pm t_{9; 0.05} \cdot \frac{3}{\sqrt{5}} = \\ &= 100 \pm 1.83 \cdot \frac{3}{\sqrt{5}} = \\ &= 100 \pm 2.46 = \boxed{(97.54; 102.46)}. \end{aligned}$$

## Problem 4

The reaction time of a patient to a certain stimulus is known to have a standard deviation of 0.05 seconds. How large a sample of measurements must a psychologist take in order to be 95% confident and 99% confident, respectively, that the error in the estimate of the mean reaction time will not exceed 0.01 seconds?

### Solution:

In a  $(1 - \alpha) \cdot 100\%$  confidence interval for  $\mu$ , when  $\sigma^2$  is known:

$$CI_{1-\alpha}(\mu) = \bar{X} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}},$$

accuracy of estimation  $e$  is its half-width:

$$e = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}.$$

A number observations required to get the accuracy  $e$ :

$$n \geq \frac{z_{\alpha/2}^2 \cdot \sigma^2}{e^2}.$$

From problem statement:

(a)  $1 - \alpha = 0.95$ :

$$n_{95\%} \geq \frac{z_{0.025}^2 \cdot 0.05^2}{0.01^2} = \frac{1.96^2 \cdot 0.05^2}{0.01^2} = 96.04,$$
$$n_{95\%} = \boxed{97}.$$

(b)  $1 - \alpha = 0.99$ :

$$n_{99\%} \geq \frac{z_{0.005}^2 \cdot 0.05^2}{0.01^2} = \frac{2.58^2 \cdot 0.05^2}{0.01^2} = 165.87,$$
$$n_{99\%} = \boxed{166}.$$

## Problem 5

During the Friday night shift,  $n = 28$  mints were selected at random from a production line and weighted. They had average weight of  $\bar{x} = 21.45$  grams and  $s = 0.31$  gram. Give the lower endpoint of a 90% one-sided confidence interval for  $\mu$ , the mean weight of all mints.

### Solution:

$(1 - \alpha) \cdot 100\%$  lower bound confidence interval for  $\mu$ , when  $\sigma^2$  is unknown:

$$CI_{1-\alpha}(\mu) = \left( \bar{X} - t_{n-1; \alpha} \cdot \frac{S}{\sqrt{n}}; +\infty \right).$$

From problem statement:

$$\begin{aligned} CI_{90\%}(\mu) &= \left( 21.45 - t_{27; 0.1} \cdot \frac{0.31}{\sqrt{28}}; +\infty \right) = \\ &= \left( 21.45 - 1.314 \cdot \frac{0.31}{\sqrt{28}}; +\infty \right) = \\ &= (21.37; +\infty). \end{aligned}$$

So the lower endpoint is 21.37.

## Problem 6

- (a) A student constructed two 95% confidence intervals for unknown parameter  $\theta : (-\infty; 4.2)$  and  $(0.5; \infty)$ . What could be the confidence of the interval  $(0.5; 4.2)$ ?
- (b) A student constructed two 95% confidence intervals for unknown parameter  $\theta : (-5; 4.2)$  and  $(0.5; 7)$ . What could be the confidence of the interval  $(0.5; 4.2)$ ?

**Solution:**

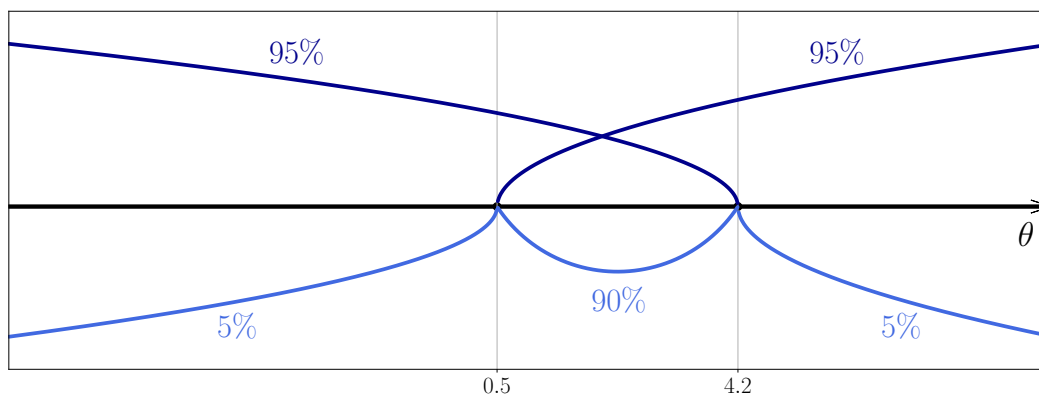


Figure 1: Intersection of infinite confidence intervals.

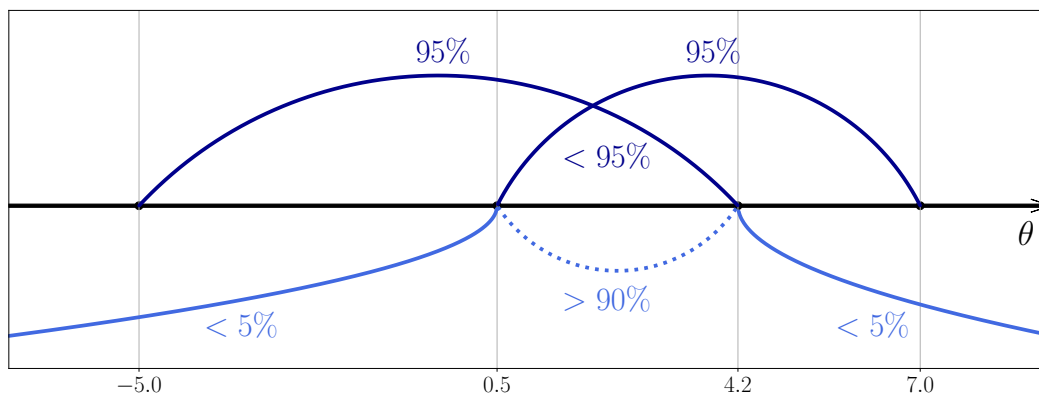


Figure 2: Intersection of finite confidence intervals.

- (a) Let's define  $A$  as an event of confidence interval  $(-\infty; 4.2)$  containing parameter  $\theta$ , and  $B$  as an event of confidence interval  $(0.5; \infty)$  containing parameter  $\theta$ .  
It's required to find  $P(A \cap B)$ .



We know that  $P(A) = P(B) = 0.95$ , and since intervals are exhaustive  $P(A \cup B) = 1$ .

Then

$$P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0.95 + 0.95 - 1 = \boxed{0.9}.$$

The illustration is in the fig. 1.

- (b) Let's define  $A$  as an event of confidence interval  $(-5; 4.2)$  containing parameter  $\theta$ , and  $B$  as an event of confidence interval  $(0.5; 7)$  containing parameter  $\theta$ .

It's required to find  $P(A \cap B)$ .

$P(A) = P(B) = 0.95$ , but intervals are not exhaustive, so  $0.95 < P(A \cup B) < 1$ .

Then

$$P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0.95 + 0.95 - (0.95; 1) = (0.9; 0.95),$$

$$\boxed{0.9 < P(A \cap B) < 0.95}.$$

The illustration is in the fig. 2.

## Problem 7

A manufacturer bonds a plastic coating to a metal surface. A random sample of nine observations on the thickness of this coating is taken from a week's output. The sample thickness (in millimeters) were as follows:

19.8   21.2   18.6   20.4   21.6   19.8   19.9   20.3   20.8

Assuming that the population distribution is normal, find a 90% confidence interval for the population variance.

### Solution:

$(1 - \alpha) \cdot 100\%$  confidence interval for  $\sigma^2$ , when  $\mu$  is unknown:

$$CI_{1-\alpha}(\sigma^2) = \left( \frac{(n-1)S^2}{\chi_{n-1; \alpha/2}^2}; \frac{(n-1)S^2}{\chi_{n-1; 1-\alpha/2}^2} \right).$$

A sample variance  $s^2$ :

$$s^2 = \frac{1}{8} \sum_{i=1}^9 (x_i - \bar{x})^2.$$

A sample mean  $\bar{x}$  required:

$$\bar{x} = \frac{1}{9} \sum_{i=1}^9 x_i = 20.27,$$

so  $s^2$  is

$$s^2 = \frac{1}{8} \sum_{i=1}^9 (x_i - 20.27)^2 \approx 0.79.$$

From problem statement:

$$\begin{aligned} CI_{90\%}(\sigma^2) &= \left( \frac{8 \cdot 0.79}{\chi_{8; 0.05}^2}; \frac{8 \cdot 0.79}{\chi_{8; 0.95}^2} \right) = \\ &= \left( \frac{6.3}{15.51}; \frac{6.3}{2.73} \right) = \boxed{(0.41; 2.31)}. \end{aligned}$$

## Problem 8

Let 10.1, 9.7 be a sample from the normal population  $X \sim \mathcal{N}(10, \sigma^2)$ . Let 20.1, 19.5, 20.4 be a sample from the normal population  $Y \sim \mathcal{N}(20, \sigma^2)$ . Find a two-sided 90% confidence interval for the population standard deviation  $\sigma$ .

### Solution:

We could use either variable  $X$  or  $Y$  individually in order to find an estimate for  $\sigma$ , but that would not be an optimal result, since we would not have used all given degrees of freedom.

Let's construct a random variable  $Q$  as follows:

$$Q = \sum_{i=1}^2 \left( \frac{X_i - 10}{\sigma} \right)^2 + \sum_{j=1}^3 \left( \frac{Y_j - 20}{\sigma} \right)^2.$$

Both terms are squares of standardized normal variables. Moreover, they are independent of each other. Thus,  $Q$  has a known  $\chi^2$ -distribution as a sum of independent  $\chi^2$ -distributions:

$$Q = \sum_{i=1}^2 Z_i^2 + \sum_{j=1}^3 Z_j^2 = \sum_{k=1}^5 Z_k^2 \sim \chi_5^2,$$

so  $Q$  may be a pivot function.

A 90% confidence interval with  $Q$ :

$$P(\chi_{5; 0.95}^2 \leq Q \leq \chi_{5; 0.05}^2) = 0.9.$$

From problem statement a value of  $Q$ :

$$q = \sum_{i=1}^2 \left( \frac{x_i - 10}{\sigma} \right)^2 + \sum_{j=1}^3 \left( \frac{y_j - 20}{\sigma} \right)^2 = \frac{0.52}{\sigma^2}.$$

Substituting into confidence interval:

$$P\left(\chi_{5; 0.95}^2 \leq \frac{0.52}{\sigma^2} \leq \chi_{5; 0.05}^2\right) = 0.9,$$
$$P\left(\frac{0.52}{\chi_{5; 0.05}^2} \leq \sigma^2 \leq \frac{0.52}{\chi_{5; 0.95}^2}\right) = 0.9,$$

Getting values from  $\chi^2$ -distribution table:

$$CI_{90\%}(\sigma^2) = \left( \frac{0.52}{11.07}; \frac{0.52}{1.145} \right) = (0.047; 0.454).$$

Confidence interval for standard deviation  $\sigma$  then:

$$CI_{90\%}(\sigma) = \left( \sqrt{0.047}; \sqrt{0.454} \right) = \boxed{(0.217; 0.674)}.$$