APPM 4570/5570

Unit #6b:

Confidence Intervals (one sample)

EXAMPLES

(Ch. 8.1, 8.2.1, 8.2.2, 8.4.1)

Table of CIs (One Sample)

Confidence Interval Equation	Assumptions/requirements (One Sample)
$\bar{x} \pm z_{\alpha/2}\sigma/\sqrt{n}$	$n \geq 30$ or $n < 30$ and population distribution is normal.
$\bar{x} \pm z_{\alpha/2}s/\sqrt{n}$	$n \geq 40$, any distribution.
$\bar{x} \pm t_{\alpha/2,n-1}s/\sqrt{n}$	$n < 40$ and population distribution is normal.
$\widehat{p} \pm z_{\alpha/2}\sqrt{\widehat{p}(1-\widehat{p})/n}$	$n\widehat{p}, n(1-\widehat{p}) > 10$.
$\dfrac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} < \sigma^2 < \dfrac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}$	population distribution is normal

Example 1

Suppose a student measuring the boiling temperature of a certain liquid observes the readings (in degrees Celsius):

102.5, 101.7, 103.1, 100.9, 100.5, and 102.2

on 6 different samples of the liquid.

Example 1

Suppose a student measuring the boiling temperature of a certain liquid observes the readings (in degrees Celsius):

102.5, 101.7, 103.1, 100.9, 100.5, and 102.2

on 6 different samples of the liquid.

(a) If they know that the measurements follows a normal distribution, what is the 95% confidence interval for the population mean?

Example 1

Suppose a student measuring the boiling temperature of a certain liquid observes the readings (in degrees Celsius):

102.5, 101.7, 103.1, 100.9, 100.5, and 102.2

on 6 different samples of the liquid.

(a) If they know that the measurements follows a normal distribution, what is the 95% confidence interval for the population mean?

(b) Suppose the student finds out that the boiling temperatures for this liquid are normally distributed with true standard deviation $\sigma = 1.0$. Now what is the 95% confidence interval for the population mean?

Example 1 (solution)

(a) Here since the sample size is small (i.e. $n \leq 30$)

Example 1 (solution)

(a) Here since the sample size is small (i.e. $n \leq 30$) and we do not know the (true) variance

Example 1 (solution)

(a) Here since the sample size is small (i.e. $n \leq 30$) and we do not know the (true) variance and the population is assumed normal

Example 1 (solution)

(a) Here since the sample size is small (i.e. $n \leq 30$) and we do not know the (true) variance and the population is assumed normal, we should use the *t*-distribution CI since there will be ***more variability*** (from using the sample variance) and so we want a ***wider interval***.

Example 1 (solution)

(a) Here since the sample size is small (i.e. $n \leq 30$) and we do not know the (true) variance and the population is assumed normal, we should use the *t*-distribution CI since there will be ***more variability*** (from using the sample variance) and so we want a ***wider interval***. Note, here the sample standard deviation is $\sigma \approx s = 0.985$, and $\alpha/2 = (1 - 0.95)/2 = 0.025$ and so

$$t_{\alpha/2,n-1} = t_{.025,5} = \texttt{qt(0.975,5) = 2.570582} \approx 2.57$$

Example 1 (solution)

(a) Here since the sample size is small (i.e. $n \leq 30$) and we do not know the (true) variance and the population is assumed normal, we should use the *t*-distribution CI since there will be ***more variability*** (from using the sample variance) and so we want a ***wider interval***. Note, here the sample standard deviation is $\sigma \approx s = 0.985$, and $\alpha/2 = (1 - 0.95)/2 = 0.025$ and so

$t_{\alpha/2,n-1} = t_{.025,5} =$ `qt(0.975,5) = 2.570582` ≈ 2.57, so the 95% CI is:

$$\overline{x} \pm t_{\alpha/2,n-1} \cdot s/\sqrt{n} = 101.82 \pm (2.57)\frac{0.985}{\sqrt{6}} = 101.82 \pm 1.03 \text{ or } (100.79, 102.85)$$

Example 1 (solution)

(a) Here since the sample size is small (i.e. $n \leq 30$) and we do not know the (true) variance and the population is assumed normal, we should use the *t*-distribution CI since there will be ***more variability*** (from using the sample variance) and so we want a ***wider interval***. Note, here the sample standard deviation is $\sigma \approx s = 0.985$, and $\alpha/2 = (1 - 0.95)/2 = 0.025$ and so

$t_{\alpha/2,n-1} = t_{.025,5} =$ `qt(0.975,5) = 2.570582` ≈ 2.57, so the 95% CI is:

$$\overline{x} \pm t_{\alpha/2,n-1} \cdot s/\sqrt{n} = 101.82 \pm (2.57)\frac{0.985}{\sqrt{6}} = 101.82 \pm 1.03 \text{ or } (100.79, 102.85)$$

(b) Although n is small

Example 1 (solution)

(a) Here since the <u>sample size is small</u> (i.e. $n \leq 30$) and we <u>do not know the (true) variance</u> and the <u>population is assumed normal</u>, we should use the <u>*t*-distribution</u> CI since there will be ***more variability*** (from using the sample variance) and so we want a ***wider interval***. Note, here the <u>sample standard deviation</u> is $\sigma \approx s = 0.985$, and $\alpha/2 = (1 - 0.95)/2 = 0.025$ and so

$t_{\alpha/2,n-1} = t_{.025,5} =$ `qt(0.975,5) = 2.570582` ≈ 2.57, so the 95% CI is:

$$\overline{x} \pm t_{\alpha/2,n-1} \cdot s/\sqrt{n} = 101.82 \pm (2.57)\frac{0.985}{\sqrt{6}} = 101.82 \pm 1.03 \text{ or } (100.79, 102.85)$$

(b) Although n is small, since we <u>*assume a normal population*</u>

Example 1 (solution)

(a) Here since the sample size is small (i.e. $n \leq 30$) and we do not know the (true) variance and the population is assumed normal, we should use the *t*-distribution CI since there will be ***more variability*** (from using the sample variance) and so we want a ***wider interval***. Note, here the sample standard deviation is $\sigma \approx s = 0.985$, and $\alpha/2 = (1 - 0.95)/2 = 0.025$ and so

$t_{\alpha/2,n-1} = t_{.025,5} =$ `qt(0.975,5) = 2.570582` ≈ 2.57, so the 95% CI is:

$$\overline{x} \pm t_{\alpha/2,n-1} \cdot s/\sqrt{n} = 101.82 \pm (2.57)\frac{0.985}{\sqrt{6}} = 101.82 \pm 1.03 \text{ or } (100.79, 102.85)$$

(b) Although n is small, since we *assume a normal population* and we know the true standard deviation

Example 1 (solution)

(a) Here since the sample size is small (i.e. $n \leq 30$) and we do not know the (true) variance and the population is assumed normal, we should use the *t*-distribution CI since there will be ***more variability*** (from using the sample variance) and so we want a ***wider interval***. Note, here the sample standard deviation is $\sigma \approx s = 0.985$, and $\alpha/2 = (1 - 0.95)/2 = 0.025$ and so

$t_{\alpha/2,n-1} = t_{.025,5} =$ `qt(0.975,5) = 2.570582` ≈ 2.57, so the 95% CI is:

$$\overline{x} \pm t_{\alpha/2,n-1} \cdot s/\sqrt{n} = 101.82 \pm (2.57)\frac{0.985}{\sqrt{6}} = 101.82 \pm 1.03 \text{ or } (100.79, 102.85)$$

(b) Although n is small, since we *assume a normal population* and we know the true standard deviation, we will use a CI based on z-scores (*otherwise we would use t-scores*).

Example 1 (solution)

(a) Here since the sample size is small (i.e. $n \leq 30$) and we do not know the (true) variance and the population is assumed normal, we should use the *t*-distribution CI since there will be ***more variability*** (from using the sample variance) and so we want a ***wider interval***. Note, here the sample standard deviation is $\sigma \approx s = 0.985$, and $\alpha/2 = (1 - 0.95)/2 = 0.025$ and so

$t_{\alpha/2,n-1} = t_{.025,5} =$ `qt(0.975,5) = 2.570582` ≈ 2.57, so the 95% CI is:

$$\overline{x} \pm t_{\alpha/2,n-1} \cdot s/\sqrt{n} = 101.82 \pm (2.57)\frac{0.985}{\sqrt{6}} = 101.82 \pm 1.03 \text{ or } (100.79, 102.85)$$

(b) Although n is small, since we *assume a normal population* and we know the true standard deviation, we will use a CI based on *z*-scores (*otherwise we would use t-scores*). Note `qnorm(0.975) = 1.959964`

Example 1 (solution)

(a) Here since the sample size is small (i.e. $n \leq 30$) and we do not know the (true) variance and the population is assumed normal, we should use the *t*-distribution CI since there will be ***more variability*** (from using the sample variance) and so we want a ***wider interval***. Note, here the sample standard deviation is $\sigma \approx s = 0.985$, and $\alpha/2 = (1 - 0.95)/2 = 0.025$ and so

$t_{\alpha/2,n-1} = t_{.025,5} =$ `qt(0.975,5) = 2.570582` ≈ 2.57, so the 95% CI is:

$$\overline{x} \pm t_{\alpha/2,n-1} \cdot s/\sqrt{n} = 101.82 \pm (2.57)\frac{0.985}{\sqrt{6}} = 101.82 \pm 1.03 \text{ or } (100.79, 102.85)$$

(b) Although n is small, since we *assume a normal population* and we know the true standard deviation, we will use a CI based on z-scores (*otherwise we would use t-scores*). Note `qnorm(0.975) = 1.959964` thus the 95% CI for the mean is:

$$\overline{x} \pm (z_{\alpha/2})\frac{\sigma}{\sqrt{n}} = 101.82 \pm (1.96)\frac{1.0}{\sqrt{6}} = 101.82 \pm 0.800 \text{ or } (101.02, 102.62)$$

Example 2

The following data represent a sample of the assets (in millions of dollars) of 45 credit unions in southwestern Pennsylvania. Find the <u>90% confidence interval of the mean</u>.

12.4, 15.1, 11.8, 13.3, 16.7, 17.1, 15.2, 22.3, 22.4,
14.0, 11.0, 13.4, 17.9, 13.8, 20.2, 13.2, 11.8, 15.8,
19.3, 16.1, 12.9, 15.2, 18.7, 17.7, 18.6, 15.6, 15.9,
17.6, 13.1, 17.6, 13.2, 18.5,15.0, 14.2, 17.9, 17.0,
13.6, 7.9, 19.2, 17.4, 15.2, 21.8, 11.5, 14.0, 12.9

Sample Mean = 15.67
Sample SD = 3.13

Example 2 (solution)

Note that since $n = 45$

Example 2(solution)

Note that since $n = 45$, we can justify the use of the CLT and assume the population has a normal distribution.

Example 2 (solution)

Note that since $n = 45$, we can justify the use of the CLT and assume the population has a normal distribution. Furthermore, since n is large we will use a CI based on z-scores.

Example 2(solution)

Note that since $n = 45$, we can justify the use of the CLT and assume the population has a normal distribution. Furthermore, since n is large we will use a CI based on z-scores. Since we do not know the true variance we approximate it with the sample variance.

Example 2 (solution)

Note that since $n = 45$, we can justify the use of the CLT and assume the population has a normal distribution. Furthermore, since n is large we will use a CI based on z-scores. Since we do not know the true variance we approximate it with the sample variance.

Here note that $\alpha/2 = (1 - 0.9)/2 = 0.05$ and so we have that

$$z_{\alpha/2} = z_{.05} = \texttt{qnorm(0.95) = 1.644854} \approx 1.64$$

Example 2 (solution)

Note that since $n = 45$, we can justify the use of the CLT and assume the population has a normal distribution. Furthermore, since n is large we will use a CI based on z-scores. Since we do not know the true variance we approximate it with the sample variance.

Here note that $\alpha/2 = (1 - 0.9)/2 = 0.05$ and so we have that

$$z_{\alpha/2} = z_{.05} = \texttt{qnorm(0.95) = 1.644854} \approx 1.64$$

(Note that $\phi(1.64) = 0.95$)

Example 2(solution)

Note that since $n = 45$, we can justify the use of the CLT and assume the population has a normal distribution. Furthermore, since n is large we will use a CI based on z-scores. Since we do not know the true variance we approximate it with the sample variance.

Here note that $\alpha/2 = (1 - 0.9)/2 = 0.05$ and so we have that

$z_{\alpha/2} = z_{.05} =$ `qnorm(0.95) = 1.644854` ≈ 1.64

(Note that $\phi(1.64) = 0.95$)

Thus the 90% CI is

$$\overline{x} \pm (z_{\alpha/2}) \frac{s}{\sqrt{n}} = 15.67 \pm (1.64) \frac{3.13}{\sqrt{45}} = 15.67 \pm 0.77 \text{ or } (14.9, 16.44)$$

Example 3

You are studying the lifetime of a piece of machinery, and you know the failure times for this type of machinery is exponentially distributed with parameter $\lambda = 1/3$.

Example 3

You are studying the lifetime of a piece of machinery, and you know the failure times for this type of machinery is exponentially distributed with parameter $\lambda = 1/3$.

(a) Because your data is not normally distributed, you will collect a sample containing at least 40 data points. If you get exactly 40 measurements, what will the range of your 99% confidence interval for the true mean be?

Example 3

You are studying the lifetime of a piece of machinery, and you know the failure times for this type of machinery is exponentially distributed with parameter $\lambda = 1/3$.

(a) Because your data is not normally distributed, you will collect a sample containing at least 40 data points. If you get exactly 40 measurements, what will the range of your 99% confidence interval for the true mean be?

Solution: Note that $E(X_i) = 1/\lambda = 3$ and $V(X_i) = 1/\lambda^2 = 9$ for $i = 1, 2, \ldots, 40$

Example 3

You are studying the lifetime of a piece of machinery, and you know the failure times for this type of machinery is exponentially distributed with parameter $\lambda = 1/3$.

(a) Because your data is not normally distributed, you will collect a sample containing at least 40 data points. If you get exactly 40 measurements, what will the range of your 99% confidence interval for the true mean be?

Solution: Note that $E(X_i) = 1/\lambda = 3$ and $V(X_i) = 1/\lambda^2 = 9$ for $i = 1, 2, \ldots, 40$ and note that by the CLT we have the the arithmetic average is normally distributed with $\mu = 3$ and $\sigma = 3/\sqrt{40}$

Example 3

You are studying the lifetime of a piece of machinery, and you know the failure times for this type of machinery is <u>exponentially distributed</u> with parameter $\lambda = 1/3$.

(a) Because your data is not normally distributed, you will collect a sample containing at least 40 data points. If you get exactly 40 measurements, what will the range of your 99% confidence interval for the true mean be?

Solution: Note that $E(X_i) = 1/\lambda = 3$ and $V(X_i) = 1/\lambda^2 = 9$ for $i = 1, 2, \ldots, 40$ and note that by the CLT we have the the <u>arithmetic average is normally distributed</u> with $\mu = 3$ and $\sigma = 3/\sqrt{40}$ and thus the 99% CI is

$$\overline{x} \pm (z_{.005})\frac{\sigma}{\sqrt{n}} = 3 \pm (2.58)\frac{3}{\sqrt{40}} = 3 \pm 1.22 \text{ or } (1.78, 4.22)$$

(Note that `qnorm(0.995) = 2.575829`)

Example 3 (cont'd)

(b) How many data points do you need to ensure the width of your 99% confidence interval is 1/2 the value of the mean?

Example 3 (cont'd)

(b) How many data points do you need to ensure the width of your 99% confidence interval is 1/2 the value of the mean?

Solution: Here we need to set $\mu/2$ equal to the width of the confidence interval and solve for n

Example 3(cont'd)

(b) How many data points do you need to ensure the width of your 99% confidence interval is 1/2 the value of the mean?

Solution: Here we need to set *μ/2* equal to the width of the confidence interval and solve for *n*, thus we get

$$\frac{3}{2} = 2 \cdot z_{.005} \cdot s/\sqrt{n} \Rightarrow n = \left(\frac{4 \cdot z_{.005} \cdot s}{3}\right)^2$$

$$= \left(\frac{4 \cdot (2.58) \cdot 3}{3}\right)^2 = 106.5$$

Thus we need *n* = *107*.

Example 4

The article "Repeatability and Reproducibility for Pass/Fail Data" (*J. of Testing and Eval.,* 1997: 151–153) reported that in n = 48 trials in a particular laboratory, 16 resulted in ignition of a particular type of substrate by a lighted cigarette.

Example 4

The article "Repeatability and Reproducibility for Pass/Fail Data" (*J. of Testing and Eval.,* 1997: 151–153) reported that in n = 48 trials in a particular laboratory, 16 resulted in ignition of a particular type of substrate by a lighted cigarette.

Let p denote the long-run **proportion** of all such trials that would result in ignition.

Example 4

The article "Repeatability and Reproducibility for Pass/Fail Data" (*J. of Testing and Eval.,* 1997: 151–153) reported that in n = 48 trials in a particular laboratory, 16 resulted in ignition of a particular type of substrate by a lighted cigarette.

Let p denote the long-run **proportion** of all such trials that would result in ignition. A point estimate for p is

$$\widehat{p} = \frac{16}{48} = 0.333$$

Example 4

The article "Repeatability and Reproducibility for Pass/Fail Data" (*J. of Testing and Eval.,* 1997: 151–153) reported that in n = 48 trials in a particular laboratory, 16 resulted in ignition of a particular type of substrate by a lighted cigarette.

Let p denote the long-run **proportion** of all such trials that would result in ignition. A point estimate for p is

$$\widehat{p} = \frac{16}{48} = 0.333$$

Thus a confidence interval for p with a confidence level of approximately 95% (by the CLT since np > 10 and n(1-p)>10) is

Example 4

The article "Repeatability and Reproducibility for Pass/Fail Data" (*J. of Testing and Eval.,* 1997: 151–153) reported that in n = 48 trials in a particular laboratory, 16 resulted in ignition of a particular type of substrate by a lighted cigarette.

Let p denote the long-run **proportion** of all such trials that would result in ignition. A point estimate for p is

$$\widehat{p} = \frac{16}{48} = 0.333$$

Thus a confidence interval for p with a confidence level of approximately 95% (by the CLT since np > 10 and n(1-p)>10) is

$$\widehat{p} \pm z_{\alpha/2}\sqrt{\widehat{p}(1-\widehat{p})/n} = .333 \pm (1.96)\sqrt{\frac{(.333)(.667)}{48}} = .333 \pm .133 = (.200, .466)$$

(Note `qnorm(0.975) = 1.959964`*)*

Example 5(non-normal confidence interval)

A professor measures how late (in minutes) students come to class on Monday mornings. There are 17 students in the class, and the arrival time of each one is measured.

Example 5 (non-normal confidence interval)

A professor measures how late (in minutes) students come to class on Monday mornings. There are 17 students in the class, and the arrival time of each one is measured.

The distribution of these arrival times is unknown (assume it's not normal), but the professor wants to *make a 90% confidence interval for the latest arrival time*.

Example 5 (non-normal confidence interval)

A professor measures how late (in minutes) students come to class on Monday mornings. There are 17 students in the class, and the arrival time of each one is measured.

The distribution of these arrival times is unknown (assume it's not normal), but the professor wants to *make a 90% confidence interval for the latest arrival time*.

What is the correct approach to building a CI?

Example 5 (non-normal confidence interval)

A professor measures how late (in minutes) students come to class on Monday mornings. There are 17 students in the class, and the arrival time of each one is measured.

The distribution of these arrival times is unknown (assume it's not normal), but the professor wants to *make a 90% confidence interval for the latest arrival time*.

What is the correct approach to building a CI?

When $n < 40$ and the distribution is unknown, we have to

Example 5 (non-normal confidence interval)

A professor measures how late (in minutes) students come to class on Monday mornings. There are 17 students in the class, and the arrival time of each one is measured.

The distribution of these arrival times is unknown (assume it's not normal), but the professor wants to *make a 90% confidence interval for the latest arrival time*.

What is the correct approach to building a CI?

When $n < 40$ and the distribution is unknown, we have to

- make a specific *assumption* about the form of the *population distribution* and then

Example 5 (non-normal confidence interval)

A professor measures how late (in minutes) students come to class on Monday mornings. There are 17 students in the class, and the arrival time of each one is measured.

The distribution of these arrival times is unknown (assume it's not normal), but the professor wants to *make a 90% confidence interval for the latest arrival time*.

What is the correct approach to building a CI?

When $n < 40$ and the distribution is unknown, we have to

- make a specific *assumption* about the form of the *population distribution* and then
- *derive a CI* based on that assumption.

Example 5(cont'd)

For example, if we assume the data is uniformly distributed and parameter Θ = latest arrival time, then we would have to find:

Example 5 (cont'd)

For example, if we assume the data is <u>uniformly distributed</u> and parameter Θ = latest arrival time, then we would have to find:

1. A r.v. $Z = f(X_1, X_2, \ldots, X_n)$ which contains the parameter Θ and whose <u>distribution is known</u>.

Example 5 (cont'd)

For example, if we assume the data is <u>uniformly distributed</u> and parameter Θ = latest arrival time, then we would have to find:

1. A r.v. $Z = f(X_1, X_2, \ldots, X_n)$ which contains the parameter Θ and whose <u>distribution is known</u>.
2. Then determine suitable points $a < b$ such that the r.v. Z lies in the interval (a,b) with <u>probability</u> $1 - \alpha$

Example 5 (cont'd)

For example, if we assume the data is uniformly distributed and parameter Θ = latest arrival time, then we would have to find:

1. A r.v. $Z = f(X_1, X_2, \ldots, X_n)$ which contains the parameter Θ and whose distribution is known.
2. Then determine suitable points $a < b$ such that the r.v. Z lies in the interval (a,b) with probability $1 - \alpha$
3. In the probability expression from step (2), isolate Θ.

Example 5 (cont'd)

For example, if we assume the data is uniformly distributed and parameter Θ = latest arrival time, then we would have to find:

1. A r.v. $Z = f(X_1, X_2, \ldots, X_n)$ which contains the parameter Θ and whose distribution is known.
2. Then determine suitable points $a < b$ such that the r.v. Z lies in the interval (a,b) with probability $1 - \alpha$
3. In the probability expression from step (2), isolate Θ.
4. The interval in step (3) is now a $100(1 - \alpha)\%$ confidence interval of parameter Θ based on the distribution of Z.

Example 5(cont'd)

For example, if we assume the data is uniformly distributed and parameter Θ = latest arrival time, then we would have to find:

1. A r.v. $Z = f(X_1, X_2, \ldots, X_n)$ which contains the parameter Θ and whose distribution is known.
2. Then determine suitable points $a < b$ such that the r.v. Z lies in the interval (a,b) with probability $1 - \alpha$
3. In the probability expression from step (2), isolate Θ.
4. The interval in step (3) is now a $100(1 - \alpha)\%$ confidence interval of parameter Θ based on the distribution of Z.

(See **Example 6** for a technical example.)

Example 6

Let $X_1, X_2, \ldots, X_n$ be a *random sample* from a uniform distribution on the interval $(0,\theta)$

Example 6

Let $X_1, X_2, \ldots, X_n$ be a *random sample* from a uniform distribution on the interval $(0,\theta)$, now given

$$\text{if} \quad Z = \frac{\max\limits_{1 \le i \le n} \{X_i\}}{\Theta} \quad \text{then the pdf of } Z \text{ is} \quad f_Z(z) = \begin{cases} nz^{n-1}, & 0 \le z \le 1 \\ 0, & \text{else} \end{cases}$$

Example 6

Let $X_1, X_2, \ldots, X_n$ be a *random sample* from a uniform distribution on the interval $(0,\theta)$, now given

$$\text{if} \quad Z = \frac{\max\limits_{1\le i\le n}\{X_i\}}{\Theta} \quad \text{then the pdf of } Z \text{ is} \quad f_Z(z) = \begin{cases} nz^{n-1}, & 0 \le z \le 1 \\ 0, & \text{else} \end{cases}$$

then, using this pdf, it can be shown that

$$P\left((\alpha/2)^{1/n} < Z < (1-\alpha/2)^{1/n}\right) = 1 - \alpha$$

Example 6

Let X_1, X_2, …, X_n be a *random sample* from a uniform distribution on the interval *(0,θ)*, now given

$$\text{if} \quad Z = \frac{\max\limits_{1 \le i \le n} \{X_i\}}{\Theta} \quad \text{then the pdf of } Z \text{ is} \quad f_Z(z) = \begin{cases} nz^{n-1}, & 0 \le z \le 1 \\ 0, & \text{else} \end{cases}$$

then, using this pdf, it can be shown that

$$P\left((\alpha/2)^{1/n} < Z < (1-\alpha/2)^{1/n}\right) = 1 - \alpha$$

Now use this probability statement to derive a 90% CI for *Θ*.

Example 6 (solution)

(Step 1 & 2) First note that here $\alpha = 1 - 0.9 = 0.1$

Example 6(solution)

(Step 1 & 2) First note that here *α = 1 – 0.9 = 0.1,* and so by the probability expression given, we get

$$P\left((0.1/2)^{1/n} < \frac{\max\limits_{1\le i\le n}\{X_i\}}{\Theta} < (1-0.1/2)^{1/n}\right)$$
$$= P\left((0.05)^{1/n} < \frac{\max\limits_{1\le i\le n}\{X_i\}}{\Theta} < (0.95)^{1/n}\right) = 0.9$$

Example 6 (solution)

(Step 1 & 2) First note that here *α = 1 – 0.9 = 0.1,* and so by the probability expression given, we get

$$P\left((0.1/2)^{1/n} < \frac{\max\limits_{1\le i\le n}\{X_i\}}{\Theta} < (1-0.1/2)^{1/n}\right)$$

$$= P\left((0.05)^{1/n} < \frac{\max\limits_{1\le i\le n}\{X_i\}}{\Theta} < (0.95)^{1/n}\right) = 0.9$$

(Step 3) Now isolating *Θ* yields

$$P\left(\frac{\max\limits_{1\le i\le n}\{X_i\}}{(0.05)^{1/n}} > \Theta > \frac{\max\limits_{1\le i\le n}\{X_i\}}{(0.95)^{1/n}}\right) = 0.9$$

Example 6(solution)

(Step 4) And so, finally

Example 6 (solution)

(Step 4) And so, finally, we see that the interval

$$\frac{\max\limits_{1\le i\le n}\{X_i\}}{(0.95)^{1/n}} < \Theta < \frac{\max\limits_{1\le i\le n}\{X_i\}}{(0.05)^{1/n}} \quad \text{or} \quad \left(\frac{\max\limits_{1\le i\le n}\{X_i\}}{(0.95)^{1/n}}, \frac{\max\limits_{1\le i\le n}\{X_i\}}{(0.05)^{1/n}} \right)$$

is a 90% confidence interval for Θ based on the distribution of Z (and the data $X_1, X_2, \ldots, X_n$).

Example 6(**comment about mean arrival time**)

What about a CI for the **average** (late) **arrival time**?

Example 6 (**comment about mean arrival time**)

What about a CI for the **average** (late) **arrival time**?
Recall that

Proposition

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with mean μ and standard deviation σ. Then

$$1. \quad E(\overline{X}) = \mu$$

$$2. \quad V(\overline{X}) = \sigma^2/n \quad \text{and} \quad \sigma_{\overline{X}} = \sigma/\sqrt{n}$$

Example 6(**comment about <u>mean arrival time</u>**)

What about a CI for the **average** (late) **arrival time**?
Recall that
<u>Proposition</u>

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with mean μ and standard deviation σ. Then

$$1.\quad E(\overline{X}) = \mu$$

$$2.\quad V(\overline{X}) = \sigma^2/n \quad \text{and} \quad \sigma_{\overline{X}} = \sigma/\sqrt{n}$$

And recall that for a uniformly distributed r.v. on $(0,\theta)$ we have that E(X) =$\theta/2$ and so the mean of the average is also $\theta/2$

Example 6(**comment about <u>mean arrival time</u>**)

What about a CI for the **average** (late) **arrival time**?
Recall that
<u>Proposition</u>

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with mean μ and standard deviation σ. Then

$$1. \quad E(\overline{X}) = \mu$$

$$2. \quad V(\overline{X}) = \sigma^2/n \quad \text{and} \quad \sigma_{\overline{X}} = \sigma/\sqrt{n}$$

And recall that for a uniformly distributed r.v. on *(0,θ)* we have that E(X) =*θ/2* and so the mean of the average is also *θ/2,* so if we find a CI for *θ/2* we will have a CI for the average arrival time …

Example 6(**comment about <u>mean arrival time</u>**)

Thus, for example, a 90% CI for the average (late) arrival time is

Example 6(**comment about <u>mean arrival time</u>**)

Thus, for example, a 90% CI for the average (late) arrival time is

$$\frac{\max\limits_{1\le i\le n}\{X_i\}}{2(0.95)^{1/n}} < \frac{\Theta}{2} < \frac{\max\limits_{1\le i\le n}\{X_i\}}{2(0.05)^{1/n}}$$