# Theory of statistics 2

Department of Statistics and Operations Research

September 25, 2019

# Pivotal Quantity PQ- Confidence Interval (C.I) by PQ

Let $X$ be a random variable and $f(x;\theta)$ its pdf. Our interest is to find the $100(1-\alpha)\%$ C.I. of a function $\tau(\theta)$. In other word, our aim is to get from the basis $\underline{X} = (X_1, \ldots, X_n)$ the interval $(T_1(\underline{X}), T_2(\underline{X}))$ which satisfies:

$$\mathbb{P}\left(T_1(\underline{X}) < \tau(\theta) < T_2(\underline{X})\right) = 1 - \alpha.$$

Indeed, there are so many solutions of $(T_1(\underline{X}), T_2(\underline{X}))$ depending on the length $L = T_2(\underline{X}) - T_1(\underline{X})$, but $L$ is minimized by the way of Pivotal Quantity (PQ).

### Definition

A random variable $Q(\underline{X};\theta)$ with distribution $g(q)$ is said PQ, if $g(q)$ is free of the parameter $\theta$.

### Lemma

Let $X_1, \ldots, X_n$ be $n$ random variables iid with distribution $f(x;\theta)$. The statistic $Q(\underline{X};\theta) \sim -2\sum \log(F_{x_i,\theta}) \sim \mathcal{X}^2_{2n}$ is a PQ.

The purpose of the PQ method is to get the two values $q_1$ and $q_2$ using the two following steps:
1. $\mathbb{P}(q_1 < Q(\underline{X};\theta) < q_2) = \mathbb{P}(T_1(\underline{X}) < \tau(\theta) < T_2(\underline{X})) = 1 - \alpha$.
2- The length $L = T_2(\underline{X}) - T_1(\underline{X})$ is minimum.

## Example 1: The normal distribution $N(\mu, \sigma^2)$ with $\sigma$ known

Let $X$ be a random variable with normal distribution $N(\mu, \sigma^2)$. Let $X_1, \ldots, X_n$ be $n$ copies of $X$. Our aim is to find $100(1-\alpha)\%$ of $\mu$.

$Q(\underline{X}; \mu) = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim g(q) = N(0,1)$ represents the PQ of $\mu$.

Then, we look for getting $q_1$ and $q_2$.

Step 1:

$$\begin{aligned}\mathbb{P}\left(q_1 < \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < q_2\right) &= \mathbb{P}\left(\overline{X} - q_1\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} - q_2\frac{\sigma}{\sqrt{n}}\right) \\ &= \int_{q_1}^{q_2} g(q)dq = 1-\alpha.\end{aligned}$$

## Example 1: The normal distribution $N(\mu, \sigma^2)$ with $\sigma$ known

Step 2:

$L = \overline{X} - q_1 \frac{\sigma}{\sqrt{n}} - \left( \overline{X} - q_2 \frac{\sigma}{\sqrt{n}} \right) = \frac{\sigma}{\sqrt{n}} (q_2 - q_1)$ must be minimum.

The last equality of step 1 indicate that $q_2$ is a function of $q_1$. Differentiate this equality with respect $q_1$, we get

$$g(q_2)\frac{dq_2}{dq_1} - g(q_1) = 0 \Rightarrow \frac{dq_2}{dq_1} = \frac{g(q_1)}{g(q_2)}.$$

Now, let us differentiate $L$ with respect to $q_1$, we get

$$\frac{dL}{dq_1} = \frac{\sigma}{\sqrt{n}} \left( \frac{dq_2}{dq_1} - 1 \right) = \frac{\sigma}{\sqrt{n}} \left( \frac{g(q_1)}{g(q_2)} - 1 \right).$$

It follows that

$$\frac{dL}{dq_1} = \frac{\sigma}{\sqrt{n}} (e^{-(q_1^2 - q_2^2)} - 1).$$

## Example 1: The normal distribution $N(\mu, \sigma^2)$ with $\sigma$ known

Thus $\dfrac{dL}{dq_1} = 0$ if, and only if $q_1 = q_2$ or $q_1 = -q_2$. This implies that

$$\begin{cases} L \searrow, & \text{if } q_1 < -q_2; \\ L \nearrow, & \text{if } -q_2 < q_1 < q_2; \\ L \searrow, & \text{if } q_1 > q_2. \end{cases}$$

Since $q_1 < q_2$ (from the first step), then the minimum of the function $L$ is obtained on $q_1 = -q_2$. It follows that $q_2 = z_{1-\frac{\alpha}{2}}$.

## Example 1: The normal distribution $N(\mu, \sigma^2)$ with $\sigma$ known

The confidence interval is equal to

$$\begin{aligned}(T_1(\underline{X}), T_2(\underline{X})) &= \left(\overline{X} - q_1 \frac{\sigma}{\sqrt{n}}, \overline{X} - q_2 \frac{\sigma}{\sqrt{n}}\right) \\ &= \left(\overline{X} \pm z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right).\end{aligned}$$

## Example 2: The normal distribution $N(\mu, \sigma^2)$ with $\sigma$ unknown

Let $X$ be a random variable with normal distribution $N(\mu, \sigma^2)$. Let $X_1, \ldots, X_n$ be $n$ copies of $X$ with $n \leq 30$. Our aim is to find $100(1-\alpha)\%$ of $\mu$. $Q(\underline{X}; \mu) = \dfrac{\overline{X} - \mu}{S/\sqrt{n}} \sim g(q) = t_{n-1}$ represents the PQ of $\mu$. Then, we look for getting $q_1$ and $q_2$.

Step 1:

$$\begin{aligned}
\mathbb{P}\left(q_1 < \frac{\overline{X} - \mu}{S/\sqrt{n}} < q_2\right) &= \mathbb{P}\left(\overline{X} - q_1 \frac{S}{\sqrt{n}} < \mu < \overline{X} - q_2 \frac{S}{\sqrt{n}}\right) \\
&= \int_{q_1}^{q_2} g(q)dq = 1 - \alpha.
\end{aligned}$$

## Example 2: The normal distribution $N(\mu, \sigma^2)$ with $\sigma$ unknown

Step 2:

$L = \overline{X} - q_1 \frac{S}{\sqrt{n}} - \left(\overline{X} - q_2 \frac{S}{\sqrt{n}}\right) = \frac{S}{\sqrt{n}}(q_2 - q_1)$ must be minimum.

The last equality of step 1 indicate that $q_2$ is a function of $q_1$. Differentiate this equality with respect $q_1$, we get

$$g(q_2)\frac{dq_2}{dq_1} - g(q_1) = 0 \Rightarrow \frac{dq_2}{dq_1} = \frac{g(q_1)}{g(q_2)}.$$

Now, let us differentiate $L$ with respect to $q_1$, we get

$$\frac{dL}{dq_1} = \frac{S}{\sqrt{n}}\left(\frac{dq_2}{dq_1} - 1\right) = \frac{S}{\sqrt{n}}\left(\frac{g(q_1)}{g(q_2)} - 1\right).$$

Recall that the t-distribution with degree of freedom $\nu$ is given by

$$g(q) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1 + \frac{q^2}{\nu}\right)^{-\frac{\nu+1}{2}}.$$

## Example 2: The normal distribution $N(\mu, \sigma^2)$ with $\sigma$ unknown

It follows that

$$\frac{dL}{dq_1} = \frac{S}{\sqrt{n}} \left( \left( \frac{\nu + q_1^2}{\nu + q_2^2} \right)^{-\frac{\nu+1}{2}} - 1 \right).$$

Thus $\frac{dL}{dq_1} = 0$ if, and only if $q_1 = q_2$ or $q_1 = -q_2$. This implies that

$$\begin{cases} L \searrow, & \text{if } q_1 < -q_2; \\ L \nearrow, & \text{if } -q_2 < q_1 < q_2; \\ L \searrow, & \text{if } q_1 > q_2. \end{cases}$$

Since $q_1 < q_2$ (from the first step), then the minimum of the function $L$ is obtained on $q_1 = -q_2$. It follows that $q_2 = t_{1-\frac{\alpha}{2}}$.

## Example 2: The normal distribution $N(\mu, \sigma^2)$ with $\sigma$ unknown

The confidence interval is equal to

$$
\begin{aligned}
(T_1(\underline{X}), T_2(\underline{X})) &= \left(\overline{X} - q_1 \frac{S}{\sqrt{n}}, \overline{X} - q_2 \frac{S}{\sqrt{n}}\right) \\
&= \left(\overline{X} \pm t_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}\right).
\end{aligned}
$$

If $n > 30$, The confidence interval is equal to

$$
\begin{aligned}
(T_1(\underline{X}), T_2(\underline{X})) &= \left(\overline{X} - q_1 \frac{S}{\sqrt{n}}, \overline{X} - q_2 \frac{S}{\sqrt{n}}\right) \\
&= \left(\overline{X} \pm z_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}\right).
\end{aligned}
$$

## Example 3: The normal distribution $N(\mu, \sigma^2)$ with $\mu$ known

Let $X$ be a random variable with normal distribution $N(\mu, \sigma^2)$. Let $X_1, \ldots, X_n$ be $n$ copies of $X$. Our aim is to find $100(1-\alpha)\%$ of $\sigma^2$. $Q(\underline{X}; \sigma^2) = \sum_{i=1}^{n} \left( \frac{X_i - \mu}{\sigma} \right)^2 \sim g(q) = \mathcal{X}_n^2$ represents the PQ of $\sigma^2$. Then, we look for getting $q_1$ and $q_2$.

Step 1:

$$\begin{aligned}
&\mathbb{P}\left( q_1 < \sum_{i=1}^{n} \left( \frac{X_i - \mu}{\sigma} \right)^2 < q_2 \right) \\
&= \mathbb{P}\left( \frac{\sum_{i=1}^{n} (X_i - \mu)^2}{q_2} < \sigma^2 < \frac{\sum_{i=1}^{n} (X_i - \mu)^2}{q_1} \right) \\
&= \int_{q_1}^{q_2} g(q) dq = 1 - \alpha.
\end{aligned}$$

## Example 3: The normal distribution $N(\mu, \sigma^2)$ with $\mu$ known

Step 2:

$L = \left(\frac{1}{q_1} - \frac{1}{q_2}\right) \sum_{i=1}^{n} (X_i - \mu)^2$ must be minimum.

The last equality of step 1 indicate that $q_2$ is a function of $q_1$. Differentiate this equality with respect $q_1$, we get

$$g(q_2)\frac{dq_2}{dq_1} - g(q_1) = 0 \Rightarrow \frac{dq_2}{dq_1} = \frac{g(q_1)}{g(q_2)}.$$

Now, let us differentiate $L$ with respect to $q_1$, we get

$$\frac{dL}{dq_1} = (\frac{dq_2}{dq_1}\frac{1}{q_2^2} - \frac{1}{q_1^2})\sum_{i=1}^{n}(X_i - \mu)^2 = 0 \Rightarrow \frac{1}{q_2^2}\frac{g(q_1)}{g(q_2)} - \frac{1}{q_1^2} = 0.$$

## Example 3: The normal distribution $N(\mu, \sigma^2)$ with $\mu$ known

Since $\frac{dq_2}{dq_1} = \frac{g(q_1)}{g(q_2)} = \frac{q_1^{\frac{n}{2}-1} e^{-\frac{q_1}{2}}}{q_2^{\frac{n}{2}-1} e^{-\frac{q_2}{2}}}$, then $q_1$ and $q_2$ must satisfy

$$q_1^{\frac{n}{2}+1} e^{-\frac{q_1}{2}} = q_2^{\frac{n}{2}+1} e^{-\frac{q_2}{2}}. \tag{1}$$

## Remark

The minimum of the function $L$ is given on the point that satisfies the equation (1) and it is obtained by a numerical way.

## Example 3: The normal distribution $N(\mu, \sigma^2)$ with $\mu$ known

Indeed, this is an implicit equation which is solved by the true and false in the following steps:
Step 1: Start by a possible value for $q_1$.
Step 2: Find $q_2$ from the integration $\int_{q_1}^{q_2} g(q)dq = 1 - \alpha$.
Step 3: Check if $q_1$ and $q_2$ satisfy $q_1^{\frac{n}{2}+1} e^{-\frac{q_1}{2}} = q_2^{\frac{n}{2}+1} e^{-\frac{q_2}{2}}$.
Step 4: Stop if Step 3 is true, and if not go to Step 1.

## Example 3: The normal distribution $N(\mu, \sigma^2)$ with $\mu$ known

In this case $q_1 = \mathcal{X}^2_{n,1-\frac{\alpha}{2}}$ and $q_2 = \mathcal{X}^2_{n,\frac{\alpha}{2}}$. The confidence interval of $\sigma^2$ is equal to:

$$(T_1(\underline{X}), T_2(\underline{X})) = \left( \frac{\sum_{i=1}^{n} (X_i - \mu)^2}{\mathcal{X}^2_{n,\frac{\alpha}{2}}}, \frac{\sum_{i=1}^{n} (X_i - \mu)^2}{\mathcal{X}^2_{n,1-\frac{\alpha}{2}}} \right).$$

In the case where $\mu$ is unknown

$$Q(\underline{X}; \sigma^2) = \sum_{i=1}^{n} \left( \frac{X_i - \overline{X}}{\sigma} \right)^2 = \frac{(n-1)S^2}{\sigma^2} \sim g(q) = \mathcal{X}^2_{n-1}.$$

$$(T_1(\underline{X}), T_2(\underline{X})) = \left( \frac{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2}{\mathcal{X}^2_{n-1,\frac{\alpha}{2}}}, \frac{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2}{\mathcal{X}^2_{n-1,1-\frac{\alpha}{2}}} \right).$$

## Example 4: The exponential distribution $\exp(\theta)$

Let $X$ be a random variable with exponential distribution $\exp(\theta)$. Let $X_1, \ldots, X_n$ be $n$ copies of $X$. Our aim is to find $100(1-\alpha)\%$ of $\theta$. $Q(\underline{X};\theta) = 2\theta \sum_{i=1}^{n} X_i = 2\theta S \sim g(q) = \mathcal{X}^2_{2n}$ represents the PQ of $\theta$. Then, we look for getting $q_1$ and $q_2$.

Step 1:

$$\mathbb{P}\left(q_1 < 2\theta S < q_2\right) = \mathbb{P}\left(\frac{q_1}{2S} < \theta < \frac{q_2}{2S}\right) = \int_{q_1}^{q_2} g(q)dq = 1-\alpha.$$

## Example 4: The exponential distribution $\exp(\theta)$

Step 2:

$L = \frac{1}{2S}(q_2 - q_1)$ must be minimum.

The last equality of step 1 indicate that $q_2$ is a function of $q_1$.

Differentiate this equality with respect $q_1$, we get

$$g(q_2)\frac{dq_2}{dq_1} - g(q_1) = 0 \Rightarrow \frac{dq_2}{dq_1} = \frac{g(q_1)}{g(q_2)}.$$

Now, let us differentiate $L$ with respect to $q_1$, we get

$$\frac{dL}{dq_1} = \frac{1}{2S}\left(\frac{dq_2}{dq_1} - 1\right) = 0 \Rightarrow \frac{g(q_1)}{g(q_2)} - 1 = 0.$$

**Example 4: The exponential distribution $\exp(\theta)$**

Then $g(q_1) = g(q_2)$ and consequently $q_1$ and $q_2$ must satisfy

$$q_1^{n-1} e^{-\frac{q_1}{2}} = q_2^{n-1} e^{-\frac{q_2}{2}}.$$

This is an implicit equation which is solved by the true and false way.

### Example 4: The exponential distribution $\exp(\theta)$

In this case $q_1 = \mathcal{X}^2_{2n,1-\frac{\alpha}{2}}$ and $q_2 = \mathcal{X}^2_{2n,\frac{\alpha}{2}}$. The confidence interval of $\theta$ is equal to:

$$(T_1(\underline{X}), T_2(\underline{X})) = \left( \frac{\mathcal{X}^2_{2n,1-\frac{\alpha}{2}}}{2S}, \frac{\mathcal{X}^2_{2n,\frac{\alpha}{2}}}{2S} \right).$$

## Example 5

Let $X$ be a random variable with distribution $f(x;\theta) = \theta x^{\theta-1}, \quad 0 < x < 1$. Let $X_1, \ldots, X_n$ be $n$ copies of $X$. Our aim is to find $100(1-\alpha)\%$ of $\theta$.

$Q(\underline{X};\theta) = -2\sum_{i=1}^{n} \log(F_{x_i,\theta}) \sim g(q) = \mathcal{X}^2_{2n}$ represents the PQ of $\theta$.

Since $F_{x_i,\theta} = X_i^{\theta}$, then

$$Q(\underline{X};\theta) = -2\theta \sum_{i=1}^{n} \log(x_i) = 2\theta K \sim g(q) = \mathcal{X}^2_{2n}$$

with $K = -\sum_{i=1}^{n} \log(x_i) > 0$. Then, we look for getting $q_1$ and $q_2$.

Step 1:

$$\mathbb{P}\left(q_1 < 2\theta K < q_2\right) = \mathbb{P}\left(\frac{q_1}{2K} < \theta < \frac{q_2}{2K}\right) = \int_{q_1}^{q_2} g(q)dq = 1-\alpha.$$

## Example 5

Step 2:

$L = \frac{1}{2K}(q_2 - q_1)$ must be minimum.

The last equality of step 1 indicate that $q_2$ is a function of $q_1$. Differentiate this equality with respect $q_1$, we get

$$g(q_2)\frac{dq_2}{dq_1} - g(q_1) = 0 \Rightarrow \frac{dq_2}{dq_1} = \frac{g(q_1)}{g(q_2)}.$$

Now, let us differentiate $L$ with respect to $q_1$, we get

$$\frac{dL}{dq_1} = \frac{1}{2K}\left(\frac{dq_2}{dq_1} - 1\right) = 0 \Rightarrow \frac{g(q_1)}{g(q_2)} - 1 = 0.$$

**Example 5**

Then $g(q_1) = g(q_2)$ and consequently $q_1$ and $q_2$ must satisfy

$$q_1^{n-1} e^{-\frac{q_1}{2}} = q_2^{n-1} e^{-\frac{q_2}{2}}.$$

This is an implicit equation which is solved by the true and false way.

## Example 5

In this case $q_1 = \mathcal{X}^2_{2n,1-\frac{\alpha}{2}}$ and $q_2 = \mathcal{X}^2_{2n,\frac{\alpha}{2}}$. The confidence interval of $\theta$ is equal to:

$$(T_1(\underline{X}), T_2(\underline{X})) = \left( \frac{\mathcal{X}^2_{2n,1-\frac{\alpha}{2}}}{2K}, \frac{\mathcal{X}^2_{2n,\frac{\alpha}{2}}}{2K} \right).$$

Remark

not any PQ is useful, for instance

$$Q(\underline{X};\theta) = -2\sum_{i=1}^{n}\log(F_{x_i,\theta}) = -2\sum_{i=1}^{n}\log(1-e^{-\theta X_i})$$

is a PQ but it cannot reach to

$$\mathbb{P}\left(T_1(\underline{X}) < \theta < T_2(\underline{X})\right) = 1-\alpha.$$

# Thank you