AUA CS108, Statistics, Fall 2020

Lecture 13

Michael Poghosyan

25 Sep 2020

Contents

Theoretical Quantiles, again

Now, if q_α is the α-quantile of some Distribution, and X is a r.v. from that Distribution, then

$$\mathbb{P}(X \leq q_\alpha) \geq \alpha \qquad \text{and} \qquad \mathbb{P}(X \geq q_\alpha) \geq 1 - \alpha.$$

Theoretical Quantiles, again

Now, if q_α is the α-quantile of some Distribution, and X is a r.v. from that Distribution, then

$$\mathbb{P}(X \leq q_\alpha) \geq \alpha \qquad \text{and} \qquad \mathbb{P}(X \geq q_\alpha) \geq 1 - \alpha.$$

Note: Here we are taking inequalities, and not, say, $\mathbb{P}(X \leq q_\alpha) = \alpha$, since, in the Discrete r.v. case, we can have no q_α with exact equality. Say, if $X \sim \mathit{Bernoulli}(0.2)$, and $\alpha = 0.4$, then no q_α exists with $\mathbb{P}(X \leq q_\alpha) = \alpha$.

Theoretical Quantiles, again

Now, if q_α is the α-quantile of some Distribution, and X is a r.v. from that Distribution, then

$$\mathbb{P}(X \leq q_\alpha) \geq \alpha \qquad \text{and} \qquad \mathbb{P}(X \geq q_\alpha) \geq 1 - \alpha.$$

Note: Here we are taking inequalities, and not, say, $\mathbb{P}(X \leq q_\alpha) = \alpha$, since, in the Discrete r.v. case, we can have no q_α with exact equality. Say, if $X \sim \textit{Bernoulli}(0.2)$, and $\alpha = 0.4$, then no q_α exists with $\mathbb{P}(X \leq q_\alpha) = \alpha$.

Note: If $\alpha = 0.5$, we call $q_\alpha = q_{0.5}$ to be the **Median of the Distribution**.

Theoretical Quantiles, again

Now, if q_α is the α-quantile of some Distribution, and X is a r.v. from that Distribution, then

$$\mathbb{P}(X \leq q_\alpha) \geq \alpha \qquad \text{and} \qquad \mathbb{P}(X \geq q_\alpha) \geq 1 - \alpha.$$

Note: Here we are taking inequalities, and not, say, $\mathbb{P}(X \leq q_\alpha) = \alpha$, since, in the Discrete r.v. case, we can have no q_α with exact equality. Say, if $X \sim \mathit{Bernoulli}(0.2)$, and $\alpha = 0.4$, then no q_α exists with $\mathbb{P}(X \leq q_\alpha) = \alpha$.

Note: If $\alpha = 0.5$, we call $q_\alpha = q_{0.5}$ to be the **Median of the Distribution**. So if we consider a Continuous r.v. and draw the PDF of that r.v., then the Median is the (leftmost) point dividing the area under the PDF curve into 50%-50% portions.

Q-Q Plots

Next, we consider three important statistical problems: Check visually if

Q-Q Plots

Next, we consider three important statistical problems: Check visually if

- two given Datasets (possibly, of different sizes) are from the same Distribution;

Q-Q Plots

Next, we consider three important statistical problems: Check visually if

- two given Datasets (possibly, of different sizes) are from the same Distribution;
- a given Dataset comes from a given Distribution;

Q-Q Plots

Next, we consider three important statistical problems: Check visually if

- two given Datasets (possibly, of different sizes) are from the same Distribution;
- a given Dataset comes from a given Distribution;
- given two theoretical Distributions, check if one of them is a shifted-scaled version of the other one, or check if one has *fatter tails* than the other one

Q-Q Plots, Data vs Data

Now, assume we have two Datasets, not necessarily of the same size:

$$x: \; x_1, x_2, ..., x_n \qquad \text{and} \qquad y: \; y_1, y_2, ..., y_m$$

Q-Q Plots, Data vs Data

Now, assume we have two Datasets, not necessarily of the same size:

$$x: \; x_1, x_2, ..., x_n \qquad \text{and} \qquad y: \; y_1, y_2, ..., y_m$$

Question: Are x and y coming from the same Distribution?

Q-Q Plots, Data vs Data

Now, assume we have two Datasets, not necessarily of the same size:

$$x: \; x_1, x_2, ..., x_n \qquad \text{and} \qquad y: \; y_1, y_2, ..., y_m$$

Question: Are x and y coming from the same Distribution?

Q-Q Plot helps to answer to this question visually.

Q-Q Plots, Data vs Data

Now, assume we have two Datasets, not necessarily of the same size:

$$x: \; x_1, x_2, ..., x_n \qquad \text{and} \qquad y: \; y_1, y_2, ..., y_m$$

Question: Are x and y coming from the same Distribution?

Q-Q Plot helps to answer to this question visually. To draw the Q-Q Plot for Datasets, we take some levels of quantiles, say, for some k,

$$\alpha = \frac{1}{k}, \frac{2}{k}, ..., \frac{k-1}{k}$$

and then draw the points (q_α^x, q_α^y).

Q-Q Plots, Data vs Data

Now, assume we have two Datasets, not necessarily of the same size:

$$x: \ x_1, x_2, ..., x_n \qquad \text{and} \qquad y: \ y_1, y_2, ..., y_m$$

Question: Are x and y coming from the same Distribution?

Q-Q Plot helps to answer to this question visually. To draw the Q-Q Plot for Datasets, we take some levels of quantiles, say, for some k,

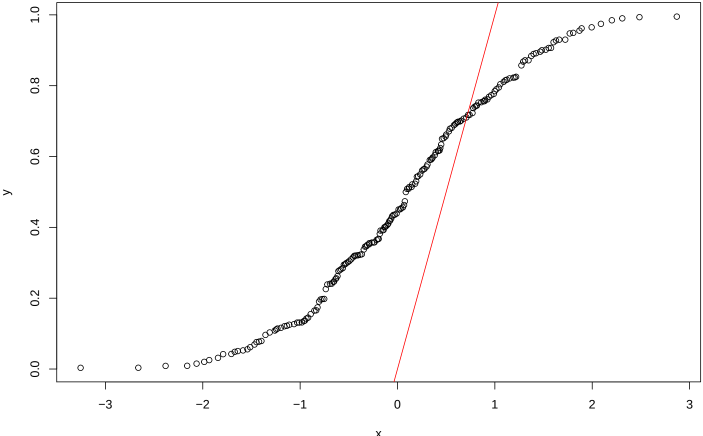
$$\alpha = \frac{1}{k}, \frac{2}{k}, ..., \frac{k-1}{k}$$

and then draw the points (q_α^x, q_α^y).

Idea: If x and y are coming from the same Distribution, then the Quantiles of x and y need to be approximately the same, $q_\alpha^x \approx q_\alpha^y$, so geometrically, the points (q_α^x, q_α^y) need to be close to the bisector line.

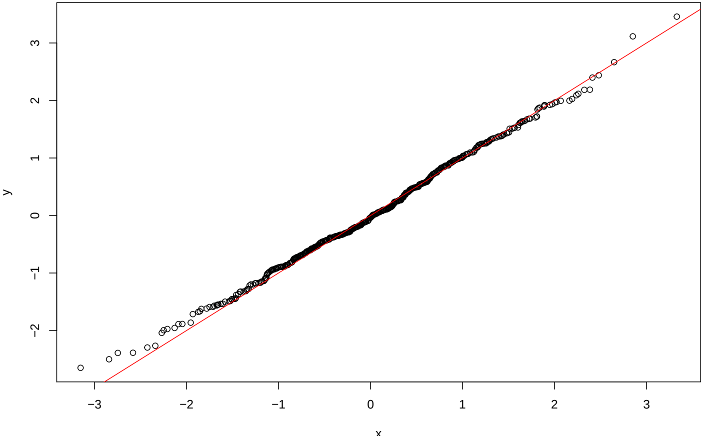
Example, Q-Q Plots, Data vs Data

```r
x <- rnorm(1000)
y <- runif(200)
qqplot(x,y)
abline(0,1, col="red")
```



Example, Q-Q Plots, Data vs Data

```
x <- rnorm(1000)
y <- rnorm(500)
qqplot(x,y)
abline(0,1, col="red")
```



Example, Q-Q Plot by Hands, Data vs Data

Example: Assume

$$x: \ -1, 2, 1, 2, 3, 2, 1 \qquad y: \ 0, 3, 4, 1, 1, 1, 1, 2$$

Draw the Q-Q Plot for x and y.