Seoul National University
Dept. of Civil and Environmental Engineering

Instructor: Junho Song
junhosong@snu.ac.kr

457.212 Statistics for Civil & Environmental Engineers

In-Class Material: Class 23

Testing Validity of Distributions: (1) Probability Papers (A&T: 7.2)

Given: Sample data set $\{x_1, x_2, ..., x_n\}$

Question: Does it follow a certain type of distribution or not? (e.g. Normal, Lognormal…)

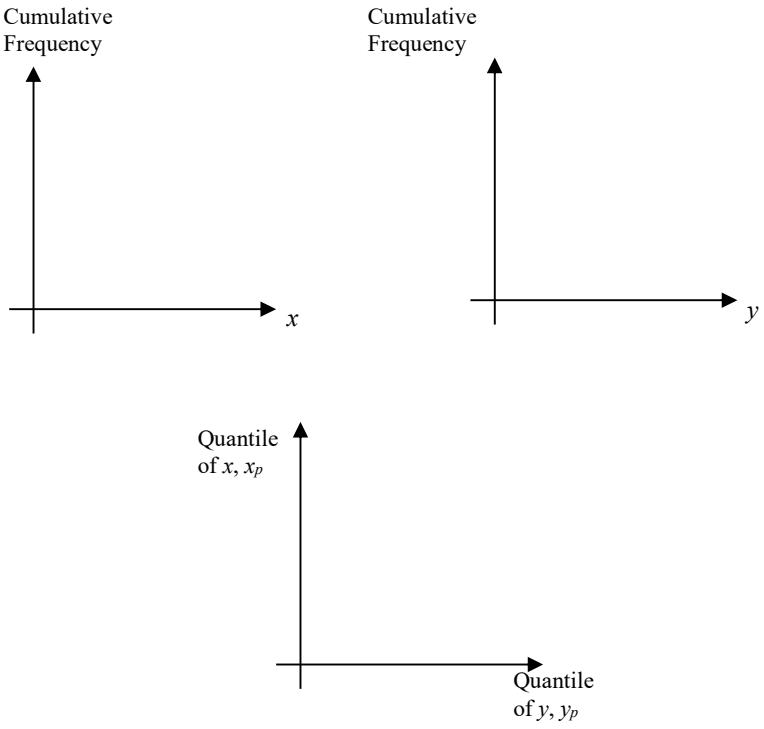
"Goodness-of-Fit" Test

(1) Visual: by probability paper (plot)

(2) Numerical: Chi-square test, K-S test

1. **Q-Q (Quantile-Quantile) plot:**

Two samples $\{x_1, x_2, ..., x_n\}$ and $\{y_1, y_2, ..., y_m\}$: Following the same distribution?

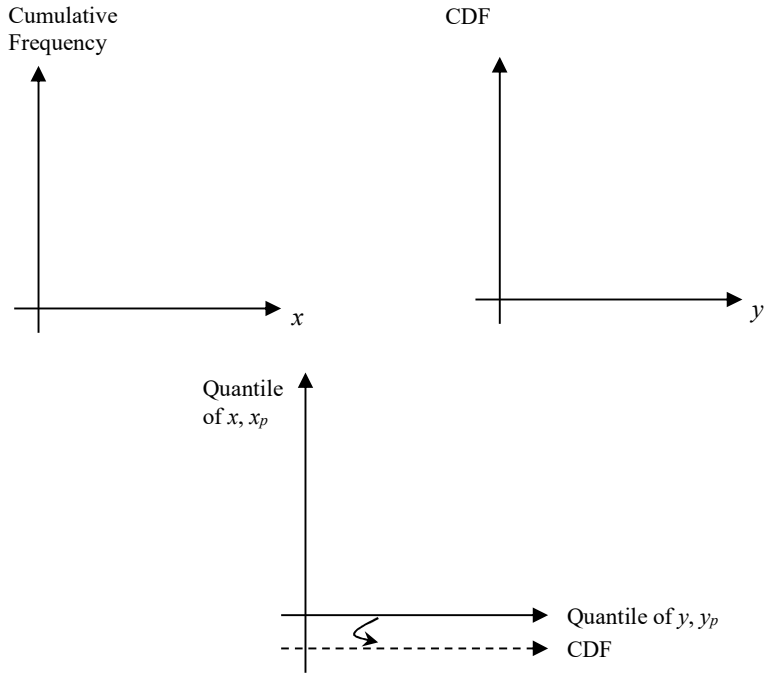


Linear pattern (i.e. scaling and shifting) indicates that the two sets follow the same type of distribution.

Instructor: Junho Song
junhosong@snu.ac.kr

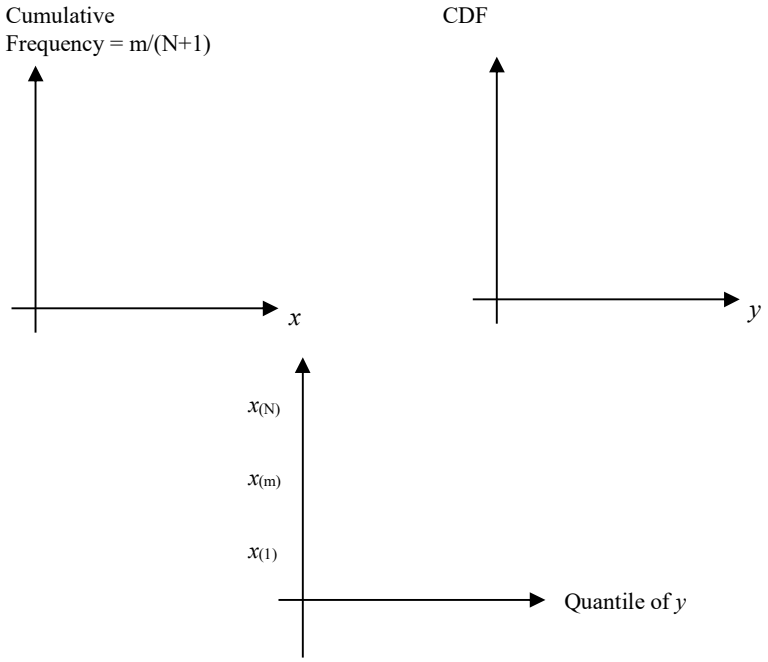
2. Probability Paper (Plot)

A sample data set $\{x_1, x_2, ..., x_n\}$ and a probabilistic distribution $F_Y(y)$



Compute quantile $y_p = F_Y^{-1}(p)$ → "Probability Plot"
Use a special graph paper (that has pre-calculated CDF) → "Probability Paper"

(a) Probability paper for Normal distribution: Use **<u>standard</u>** Normal random variable



Instructor: Junho Song
junhosong@snu.ac.kr

(b) Probability paper for Lognormal distribution

Take the ( ) of the same values and use the normal probability plot

Why?

(c) Probability paper for a general distribution

Find the standardized distribution for the distribution to test through a linear transformation

Example 1: Shifted exponential distribution

$$f_Y(y) = \lambda \exp[-\lambda(x-a)] \quad x \geq a$$
$$0 \quad \text{otherwise}$$

Example 2: Gumbel distribution

$$F_X(x) = \exp[-e^{-\alpha(x-u)}]$$

Instructor: Junho Song
junhosong@snu.ac.kr

Example 3: Construct a Rayleigh distribution probability paper for the sample, {1.413, 5.451, 1.805, 6.606, 3.464, 3.589, 4.793, 2.107, 1.023, 2.551}. The PDF of the Rayleigh distribution is

$$f_X(x) = \frac{x}{b^2}\exp\left(-\frac{x^2}{2b^2}\right), \quad 0 < x$$

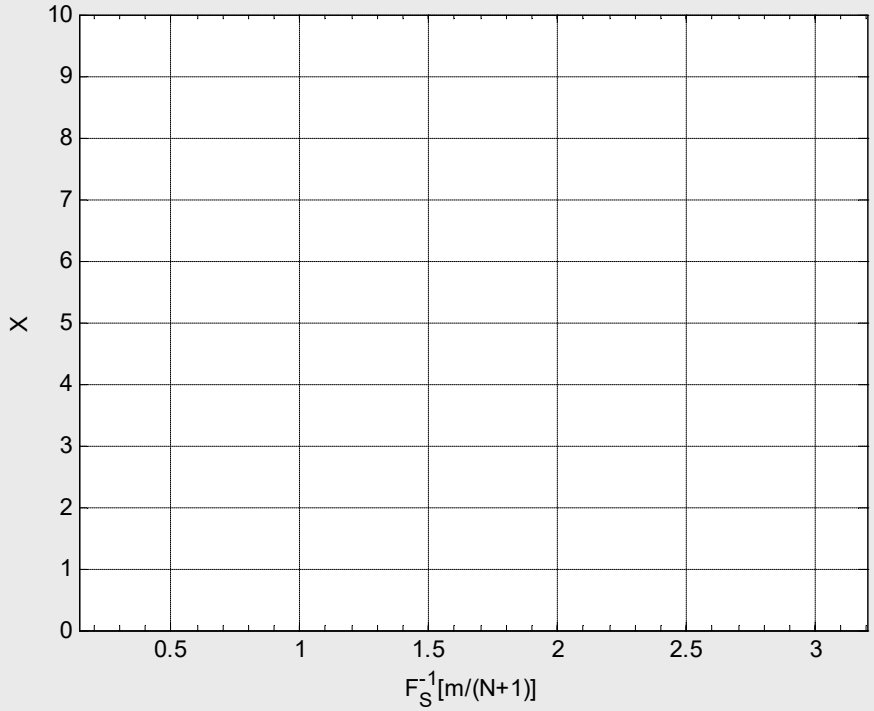
(1) Find the standard variate and its CDF, quantile for the m-th data.

(2) Order the sample in the increasing order and compute the cumulative frequency of each data and its corresponding quantile.

m	$X_{(m)}$	$m/(n+1)$	$s_p = s_{m/(n+1)}$
1	1.023	1/11 = 0.09	0.437
2	1.413	2/11 = 0.18	0.634
3	1.805	0.27	0.798
4	2.107	0.36	0.951
5	2.551	0.45	1.101
6	3.464	0.55	1.256
7	3.589	0.64	1.422
8	4.793		
9	5.451		
10	6.606		

Instructor: Junho Song
junhosong@snu.ac.kr

(3) Plot the pair of the standard variate and the corresponding sample value with arithmetic scales on both axes.



(4) Plot the pair of the standard variate and the corresponding sample value on the given probability paper of the Rayleigh distribution.