Lecture - 7

Instructor: Dr. Arabin Kumar Dey

2 Quantitle/Percentile:

If the CDF of X is continuous and strictly increasing then it has an inverse function F^{-1}. For each q between 0 and 1, $F^{-1}(q)$ is called the **q-quantile or 100qth percentile**. The probability that a continuous X is below its q-quantile is precisely q, but we will show this is not exactly true for discrete random variables.

- The median is 50% percentile or 0.5-quantile. The 25% and 75% percentiles (0.25 and 0.75-quantiles) are called the first and third quartiles and the median is second quartile. The three quartiles divide the range of the range of the random variable into four groups of equal probability.
- Similarly the 20%, 40%, 60% and 80% percentiles are called quintiles and 10%, 20%, $\cdots$, 90% percentiles are called deciles.

For discrete distributon we may see $P(X < F^{-1}(q)) < q < P(X \leq F^{-1}(q))$. Also if the function is increasing but not strictly increasing, there exists an interval of q-quantiles. Therefore we modify the definition of quantile : *The set q-quantile is the closed interval* $[x_q^-, x_q^+]$ *where:* $x_q^- = \inf\{x : F(x) \geq q\}$ *and* $x_q^+ = \inf\{x : F(x) > q\}$.

We get a uniquely defined quantile if we have $x_q^- = x_q^+$.

Example 2.1. *For exponential distribution qth qunatile π_q can be written as*

$$\pi_q = \frac{1}{\lambda}\ln(\frac{1}{1-q})$$

whereas for cauchy it can be written as

$$\pi_q = m + \lambda\tan(q\pi - \frac{\pi}{2})$$

3 Understanding the meaning of QQ plot

We are going to illustrate the intrepretation of QQ-plot with the experiment below as they relate to thickness of its tails. We calculate the quantiles for normal and cauchy distribution. Theoretical QQ-plot for these two distributions can be produced through by plotting quantiles of normal distribution against quantiles of cauchy distribution.

	$X \sim N(0,1)$	$X \sim C(0,1)$
$\pi_{0.8} = F_X^{-1}(0.8)$	0.842	1.376
$\pi_{0.85} = F_X^{-1}(0.85)$	1.036	1.963
$\pi_{0.9} = F_X^{-1}(0.9)$	1.282	3.078
$\pi_{0.95} = F_X^{-1}(0.95)$	1.645	6.314
$\pi_{0.975} = F_X^{-1}(0.975)$	1.960	12.706
$\pi_{0.99} = F_X^{-1}(0.975)$	2.326	31.821

In particular, we would have to plot the points

(0.842, 1.376), (1.036, 1.963), (1.282, 3.078), $\cdots$, (2.326, 31.821).

Note that all these points are above diagonal $y = x$, and in fact they drift further and further away above the diagonal. This fact is at the core of the interpretation of a QQ-plot comparing two distributions: points above the diagonal in the rightmost part of the plot indicate that the upper tail of the first distribution (whose quantiles are on the horizontal axis) is thinner than the tail of the distribution whose quantiles are on the verticle axis. Similarly, points below the diagonal on the left part of the plot indicate that the second distribution has a heavier lower tail.

4 Order Statistics and Sample Quantile:

Suppose that $X_1, X_2, \cdots, X_n$ is the random sample from a probability distribution F. We define the empirical cdf or estimate of cdf from the sample as

$$F_n(x) = \frac{\sum_{i=1}^{n} I(X_i \leq x)}{n}.$$

where $I(X_i \leq x)$ is 1 if $X_i \leq x$ and 0 otherwise.

We define order statistics $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$ are $X_1, X_2, \cdots, X_n$ ordered from smallest to largest. $X_{(1)}$ is the smallest observation and $X_{(n)}$ is the largest observation.

Suppose the sample is 6,4,8,2,3,4. Then the order statistics are 2,3,4,4,6,8. We write

$$F_n(x) = \begin{cases} 0 & \text{if } x < 2 \\ \frac{1}{6} & \text{if } 2 \leq x < 3 \\ \frac{2}{6} & \text{if } 3 \leq x < 4 \\ \frac{4}{6} & \text{if } 4 \leq x < 6 \\ \frac{5}{6} & \text{if } 6 \leq x < 8 \\ 1 & \text{if } x > 8 \end{cases}$$