Lecture 6

Moments, Skewness, Kurtosis, Median, Quantiles, Mode

Moments

The expected value and the variance of a random variable are particular cases of the quantities known as the *moments* of this variable.

In mathematics, a moment is a specific quantitative measure of the shape of a function.

The mathematical concept is closely related to the concept of moment in physics.

Moments

If a certain function represents physical density of a body, then

- the zeroth moment is the total mass,
- the first moment divided by the total mass is the center of mass,
- the second moment is the rotational inertia.

The kth moment <u>about a point</u> a of a <u>discrete</u> r.v. X with *pmf* $p_X(x)$ defined on a domain $D_X = \{x_i \in \mathbb{R} : p_X(x_i) > 0\}$ is the number

$$\mu_k = \sum_{D_X} (x_i - a)^k \cdot p_X(x_i)\,.$$

The kth moment <u>about a point</u> a of a <u>continuous</u> r.v. X with *pdf* $f_X(x)$ defined on a domain $D_X = \{x \in \mathbb{R} : f_X(x) > 0\}$ is the number

$$\mu_k = \int_{D_X} (x - a)^k \cdot f_X(x) dx\,.$$

If $a = 0$, we can rewrite previous formulae as

$$\mu'_k = \begin{cases} \sum_{D_X} x_i^k \cdot p_X(x_i) \;\; for\ discrete\ r.v.; \\ \int_{D_X} x^k \cdot f_X(x)dx \;\; for\ continuous\ r.v. \end{cases}$$

We will refer to μ'_k as to kth *raw moment* of a random variable.

Moments

$$\mu_k' = \begin{cases} \sum_{D_X} x_i^k \cdot p_X(x_i) \ \ for\ discrete\ r.v.; \\ \int_{D_X} x^k \cdot f_X(x)dx \ \ for\ continuous\ r.v. \end{cases}$$

Note that if $k = 0$, $\mu_k' = 1$ for any random variable.

If $k = 1$, the formulae above transform into the expressions for the expected value.

$$\mu'_k = \begin{cases} \sum_{D_X} x_i^k \cdot p_X(x_i) \;\; for\; discrete\; r.v.; \\ \int_{D_X} x^k \cdot f_X(x)dx \;\; for\; continuous\; r.v. \end{cases}$$

So, we may redefine the expected value of a random variable X as its first raw moment:

$$E[X] = \mu'_1$$

Moreover,

$$E[X^2] = \mu'_2, \qquad E[X^3] = \mu'_3, \qquad \ldots$$

Moments

If $a = E[X]$, we can rewrite general formulae for moments as

$$\mu_k = \begin{cases} \sum_{D_X} (x_i - E[X])^k \cdot p_X(x_i) \ \ for\ discrete\ r.v.; \\ \int_{D_X} (x - E[X])^k \cdot f_X(x)dx \ \ for\ continuous\ r.v. \end{cases}$$

We will refer to μ_k as to kth *central moment* of a random variable.

Moments

$$\mu_k = \begin{cases} \sum_{D_X} (x_i - E[X])^k \cdot p_X(x_i) \;\; for\ discrete\ r.v.; \\ \int_{D_X} (x - E[X])^k \cdot f_X(x)dx \;\; for\ continuous\ r.v. \end{cases}$$

As you can see, the second central moment corresponds to the variance of a random variable X:

$$Var(X) = \mu_2$$

Also,

$$\mu_k = E\left[X - E[X]\right]^k$$

Moments

$$\mu_k = \begin{cases} \sum_{D_X} (x_i - E[X])^k \cdot p_X(x_i) \;\; for\ discrete\ r.v.; \\ \int_{D_X} (x - E[X])^k \cdot f_X(x) dx \;\; for\ continuous\ r.v. \end{cases}$$

Obviously, zeroth central moment equals one, and the first central moment equals 0.

Sometimes it is convenient to express central moments in terms of raw moments. The general equation is

$$\mu_n = E[(X - E[X])^n] = \sum_{j=0}^{n} \binom{n}{j} (-1)^{n-j} \mu'_j \mu^{n-j},$$

where $\mu = \mu'_1$ - the expected value.

We also can reverse this and express raw moments in terms of central moments:

$$\mu'_m = E[X^m] = \sum_{j=0}^{m} \binom{m}{j} \mu_j \mu^{m-j},$$

where $\mu = \mu'_1$ - the expected value.

In probability theory and statistics, a *standardized moment* of a probability distribution is a moment (normally a higher degree central moment) that is *normalized*.

The normalization is typically a division by an expression of the *standard deviation* which renders the moment scale invariant.

Moments

Let X be a random variable for which we have defined the kth central moment μ_k and variance $Var(X) = \mu_2 \equiv \sigma_X^2$, where σ_X is the standard deviation of X.

Then the standardized moment of order k is given by

$$\nu_k = \frac{\mu_k}{\sigma_X^k}$$

Such normalization leads to a fact that standardized moments are dimensionless quantities.

Third and fourth standardized moments are widely used in probability and statistics.

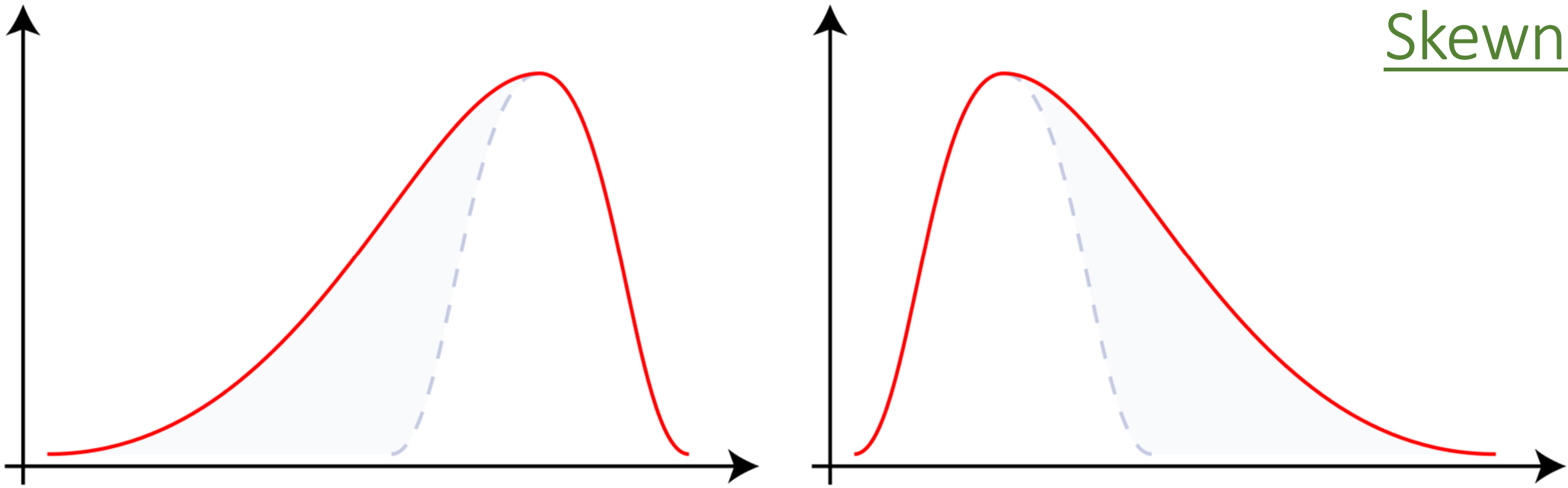
The moments of higher order have little practical use.

Third standardized moment ν_3 of a random variable X is often referred to as the *skewness*.

Skewness is a measure of the asymmetry of the probability distribution of a real-valued random variable about its mean.

The skewness value, $skew(X)$, can be positive, zero, negative, or undefined.

Skewness

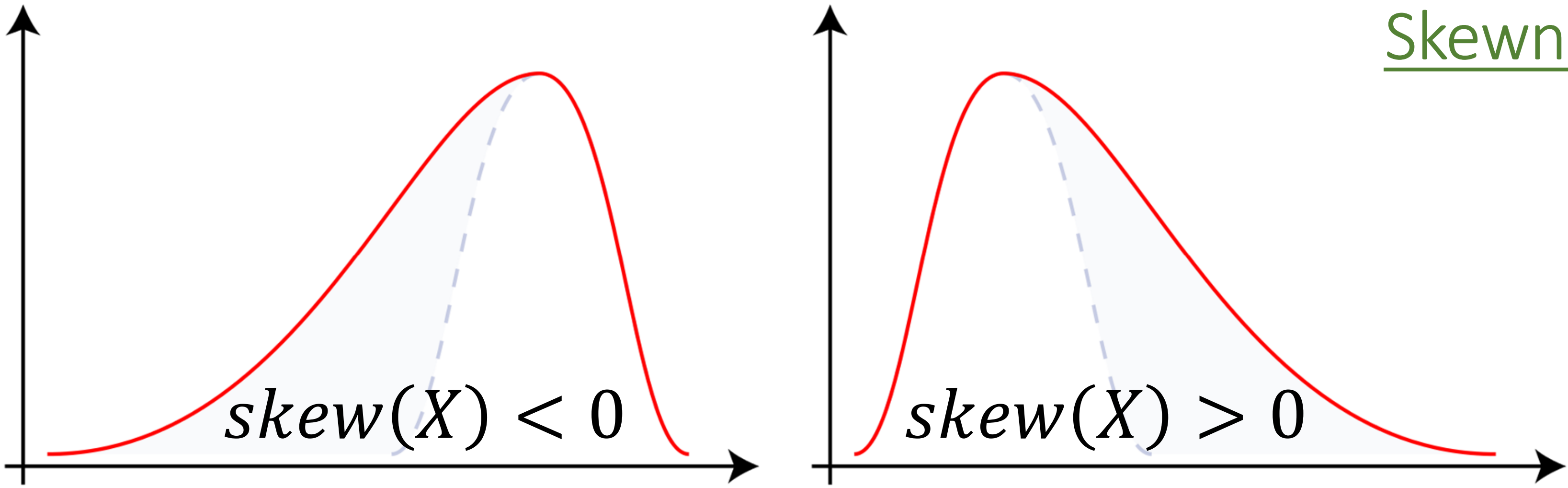


Consider the two distributions in the figure above.

Within each graph, the values on the right side of the distribution taper differently from the values on the left side.

These tapering sides are called *tails*, and they provide a visual means to determine which of the two kinds of skewness a distribution has

Skewness



negative skew: The left tail is longer; the mass of the distribution is concentrated on the right of the figure. The distribution is said to be left-skewed, left-tailed, or skewed to the left.

positive skew: The right tail is longer; the mass of the distribution is concentrated on the left of the figure. The distribution is said to be right-skewed, right-tailed, or skewed to the right.

If a distribution is symmetric about its expected value, its skewness equals zero.

However, the *converse is not true* in general. An asymmetric distribution might have the skewness value of zero, if one of its tails is long and thin, and the other is short but fat.

Skewness

If a distribution has finite expected value $E[X]$ and standard deviation σ_X, its skewness can be expressed in terms of the third raw moment:

$$skew(X) \equiv \nu_3 = \frac{E[X^3] - 3E[X]\sigma_X^2 - E[X]^3}{\sigma_X^3}$$

Skewness

Distribution	Parameters	Skewness	Remarks
Bernoulli	$0 \leq p \leq 1$	$\frac{1-2p}{\sqrt{p(1-p)}}$	
Binomial	$n \geq 1, 0 \leq p \leq 1$	$\frac{1-2p}{\sqrt{np(1-p)}}$	
Geometric	$0 \leq p \leq 1$	$\frac{2-p}{\sqrt{1-p}}$	
Disc. Uniform	$a, b \in \mathbb{Z}$	0	
Cont. Uniform	$a, b \in \mathbb{R}$	0	
Exponential	$\lambda > 0$	2	
Weibull	$\eta, \beta > 0$	$\frac{E[X^3] - 3E[X]\sigma_X^2 - E[X]^3}{\sigma_X^3}$	
Normal	$\mu, \sigma \in \mathbb{R}$	0	

Fourth standardized moment ν_4 of a random variable X is called *kurtosis* (from Greek: kurtos, meaning “curved”).

Kurtosis is a *measure of the "tailedness"* of the probability distribution of a real-valued random variable, i.e. its value describes the thickness of the distribution’s tails.

Fourth standardized moment ν_4 of a random variable X is called *kurtosis* (from Greek: kurtos, meaning “curved”).

Kurtosis is a *measure of the "tailedness"* of the probability distribution of a real-valued random variable, i.e. its value describes the thickness of the distribution’s tails.

If we would calculate the kurtosis of a normal distribution, we'd found that its value equals three, no matter what parameters the distribution has.

A lot of mathematicians prefer to compare kurtosis of any distribution with that of normal distribution. For that they subtract three from the value of kurtosis.

$$kurt(X) = \nu_4 - 3 = \frac{\mu_4}{\sigma_X^4}$$

The resulting quantity, $kurt(X)$, is called *excess kurtosis* of the random variable X.

Thus, the excess kurtosis of a normal distribution equals zero.

Distributions with zero excess kurtosis are called ***mesokurtic***.

The most prominent example of a mesokurtic distribution is the normal distribution.

A few other well-known distributions can be mesokurtic, depending on parameter values; e.g., the binomial distribution is mesokurtic for $p = \frac{1}{2} \pm \sqrt{\frac{1}{12}}$.

Kurtosis

A distribution with positive excess kurtosis is called ***leptokurtic*** ("lepto-" means "slender").

In terms of shape, a leptokurtic distribution has fatter tails than the normal distribution.

Examples of leptokurtic distributions include the Rayleigh distribution, exponential distribution, Poisson distribution.

Kurtosis

A distribution with negative excess kurtosis is called ***platykurtic*** (“platy-" means “broad“).

In terms of shape, a platykurtic distribution has thinner tails than the normal distribution.

An example of platykurtic distributions is the uniform distribution. The most platykurtic distribution of all is the Bernoulli distribution with $p = 1/2$.

Kurtosis

If a distribution has finite expected value $E[X]$ and standard deviation σ_X, its kurtosis can be expressed in terms of the fourth raw moment:

$$\nu_4 = \frac{E[X^4] - 4E[X]E[X^3] + 6E[X]^2E[X^2] - 3E[X]^4}{\sigma_X^4}$$

For excess kurtosis you should subtract 3.

Kurtosis

Distribution	Parameters	Excess Kurtosis	Remarks
Bernoulli	$0 \leq p \leq 1$	$\frac{1-6p(1-p)}{p(1-p)}$	
Binomial	$n \geq 1, 0 \leq p \leq 1$	$\frac{1-6p(1-p)}{np(1-p)}$	
Geometric	$0 \leq p \leq 1$	$6+\frac{p^2}{1-p}$	
Disc. Uniform	$a, b \in \mathbb{Z}$	$-\frac{6(n^2+1)}{5(n^2-1)}$	$n = b - a + 1$ – number of values
Cont. Uniform	$a, b \in \mathbb{R}$	-1.2	
Exponential	$\lambda > 0$	6	
Weibull	$\eta, \beta > 0$	See previous slide	
Normal	$\mu, \sigma \in \mathbb{R}$	0	

Median

The expected value of a random variable was introduced to provide a numerical value for the center of its distribution.

For some random variables, however, it is preferable to use another quantity for this purpose, either because $E[X]$ does not exist or because the distribution of X is very skewed and $E[X]$ does not represent the center very well.

Median

The latter case occurs, for instance, when X stands for the income of a randomly selected person from a set of ten people, with nine earning 20 thousand dollars and one of them earning 20 million dollars. Saying that the average income is

$$E[X] = \frac{1}{10}(9 \cdot 20000 + 1 \cdot 2000000) = 2018000 \text{ dollars}$$

is worthless and misleading.

Median

In such cases we use the *median* to represent the center. Also, for some random variables, $E[X]$ does not exist, but a median always does.

We want to define the median so that half of the probability is below it and half above it.

Median

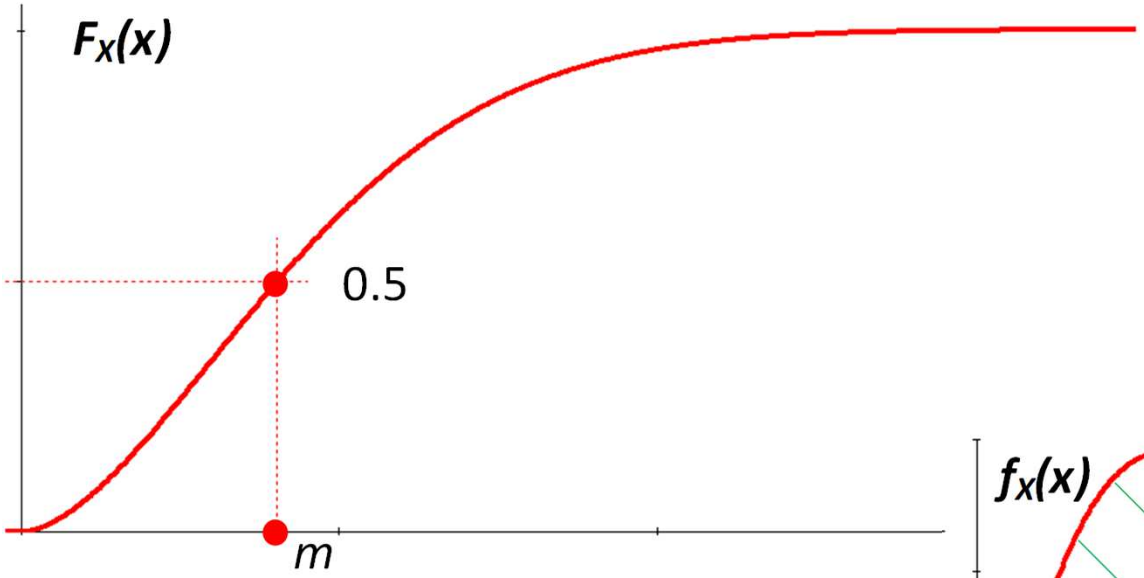
For any real-valued probability distribution, a *median* is defined as any real number m that satisfies the inequalities:

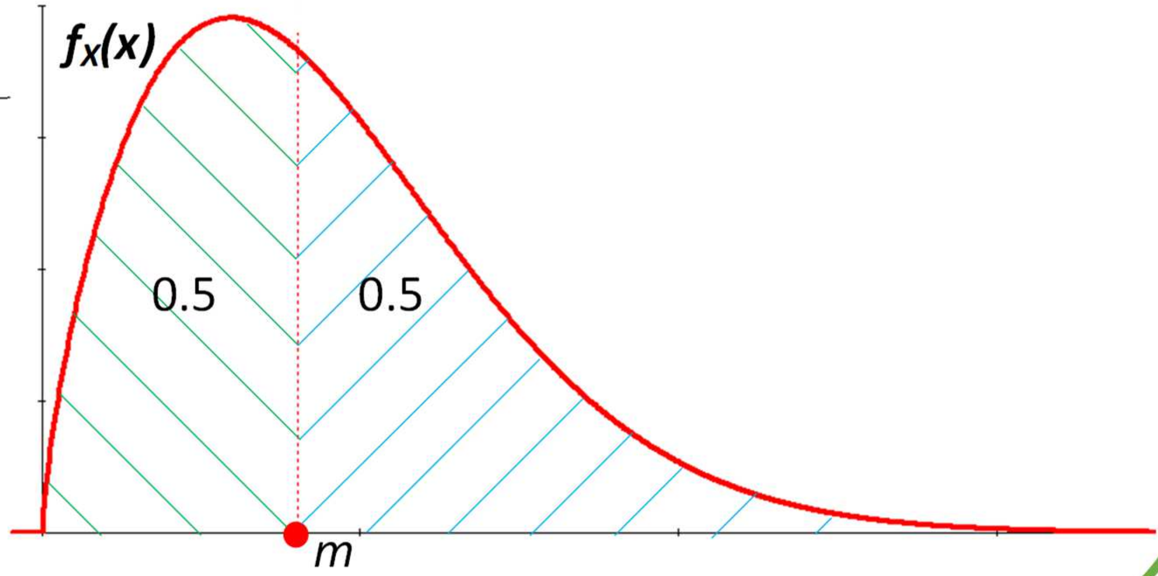
$$Pr\{X \leq m\} \geq \frac{1}{2} \qquad\qquad Pr\{X \geq m\} \geq \frac{1}{2}$$

For an absolutely continuous distribution with *pdf* $f_X(x)$ and *cdf* $F_X(x)$, the median is any real number m such that

$$\int_{-\infty}^{m} f_X(x)dx = \int_{m}^{\infty} f_X(x)dx = \frac{1}{2} \quad \text{or} \quad F_X(m) = \frac{1}{2}$$

<u>Median</u>

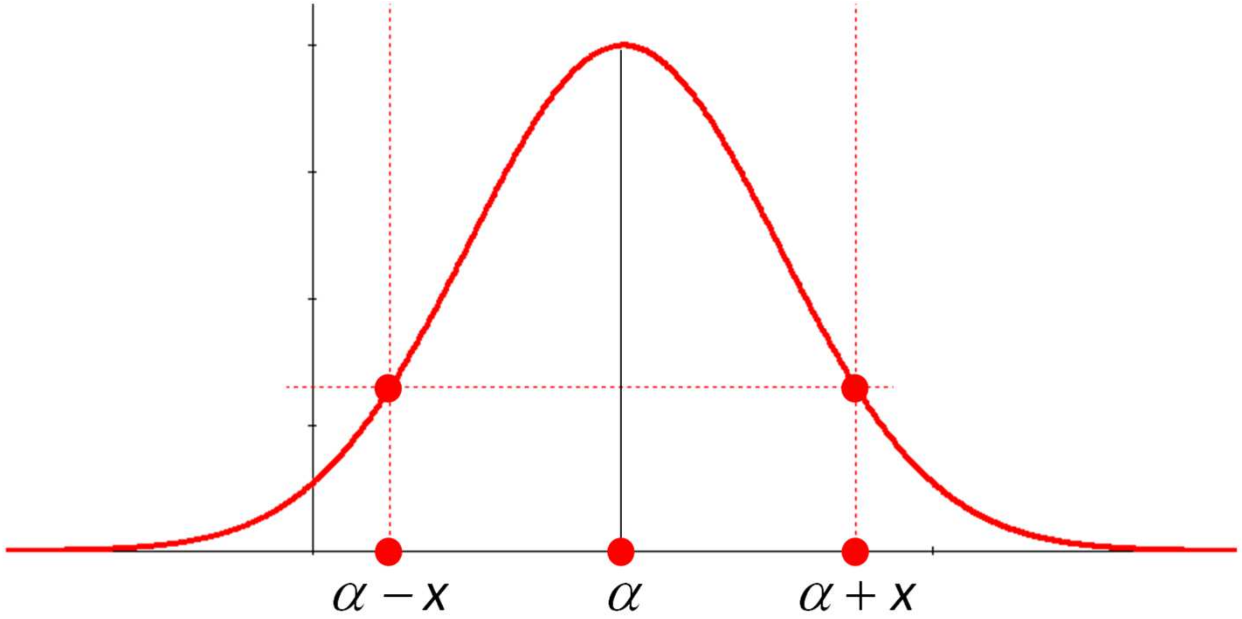




$$med(X) \equiv m_X$$

Median

If the distribution of a random variable X is symmetric about a point α, that is, its *pdf* satisfies



$f_X(\alpha - x) = f_X(\alpha + x)$ for all x, then α is a median of X.

Median

Distribution	Parameters	Median	Remarks
Bernoulli	$0 \leq p \leq 1$	$\begin{cases} 0, & if\ p < 0.5 \\ [0,1], & if\ p = 0.5 \\ 1, & if\ p > 0.5 \end{cases}$	
Binomial	$n \geq 1, 0 \leq p \leq 1$	$\lfloor np \rfloor$ or $\lceil np \rceil$	there is no single formula
Geometric	$0 \leq p \leq 1$	$\left\lceil \frac{-1}{\log_2(1-p)} \right\rceil$	not unique if $\frac{-1}{\log_2(1-p)}$ is an integer
Uniform	$a, b \in \mathbb{R}$	$\frac{a+b}{2}$	
Exponential	$\lambda > 0$	$\frac{\ln 2}{\lambda}$	
Weibull	$\eta, \beta > 0$	$\eta(\ln 2)^{1/\beta}$	
Normal	$\mu, \sigma \in \mathbb{R}$	μ	

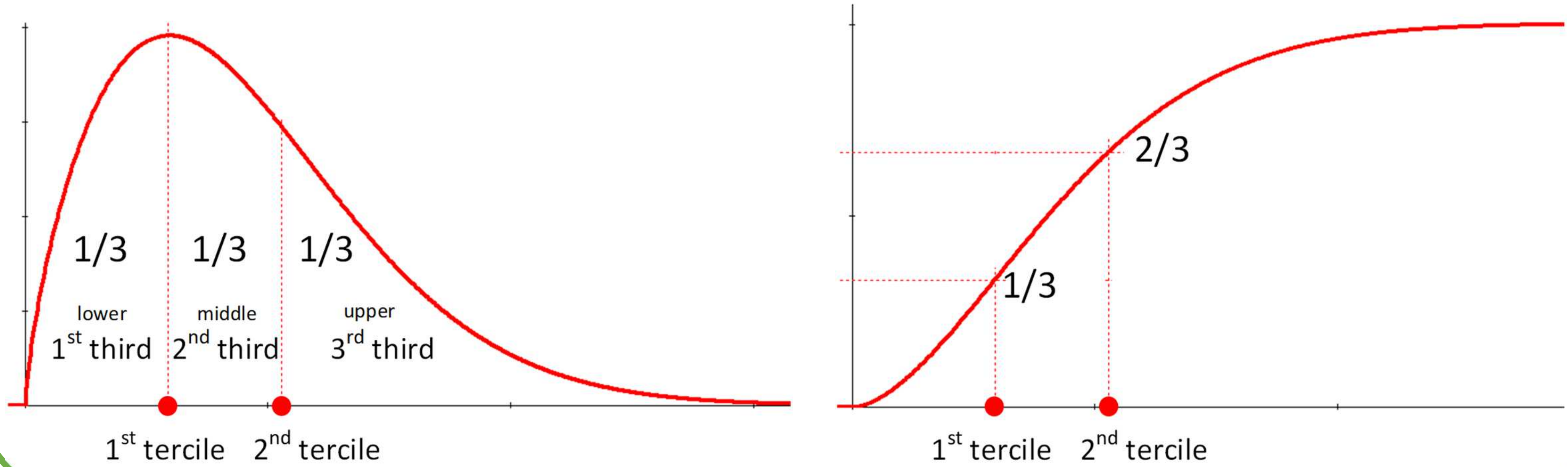
Quantiles

In probability theory *quantiles* are cut points dividing the range of a probability distribution into continuous intervals with equal probabilities.

Common quantiles have special names: for instance *tercile*, *quartile*, *decile*, etc.

Quantiles

The groups created are termed *halves*, *thirds*, *quarters*, etc., though sometimes the terms for the quantile are used for the groups created, rather than for the cut points.



Quantiles

Let X be a continuous r.v. with $F_X(x)$ continuous and strictly increasing from 0 to 1 on some finite or infinite interval.

Then, for any $p \in (0, 1)$, the solution x_p of $F_X(x_p) = p$ or, in other words, $x_p = F_X^{-1}(p)$ is called the *p*-quantile or the 100*p* percentile and the function F_X^{-1} - the *quantile function* of X or of the distribution of X.

Quantiles

For general X the *p*-quantile is defined as

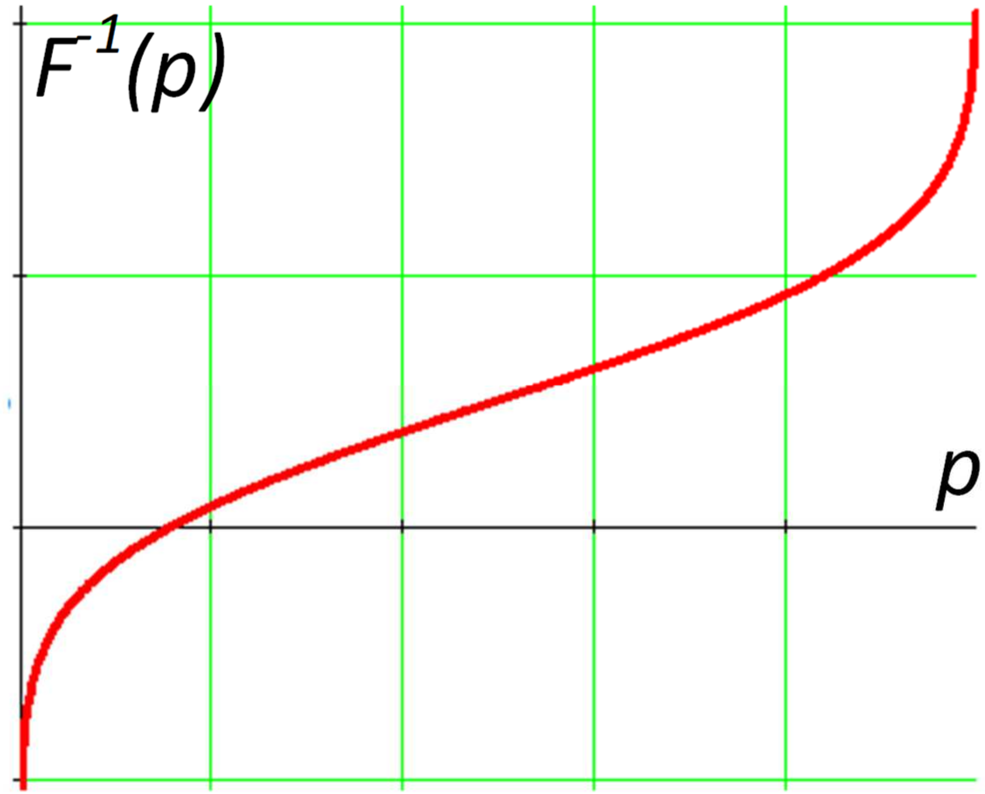
$$x_p = min\{x\colon\ F_X(x) \geq p\},$$

and we define the quantile function F_X^{-1} by

$$F_X^{-1}(p) = x_p,$$

for all $p \in (0, 1)$.

The quantile function F_X^{-1} is also called inverse cumulative distribution function (*inverse cdf*).



Quantiles

Quantiles or percentiles are often used to describe statistical data such as exam scores, home prices, incomes, etc. For example, a student's score of 650 on the math SAT is much better understood if it is also stated that this number is at the 78th percentile, meaning that 78% of the students who took the test scored 650 or less.

Clearly, the 50th percentile is also a median.

Furthermore, the 25th percentile is also called the *first quartile*, the 50th percentile the *second quartile*, and the 75th percentile *the third quartile*.

Mode

Another measure of central tendency is the *mode* of the distribution.

For a discrete r.v. X with *pmf* $p_X(x)$ the mode, x_{mod} is the value x at which the probability mass function takes its maximum value:

$$x_{mod} = \arg\max_{x \in D} p_X(x)$$

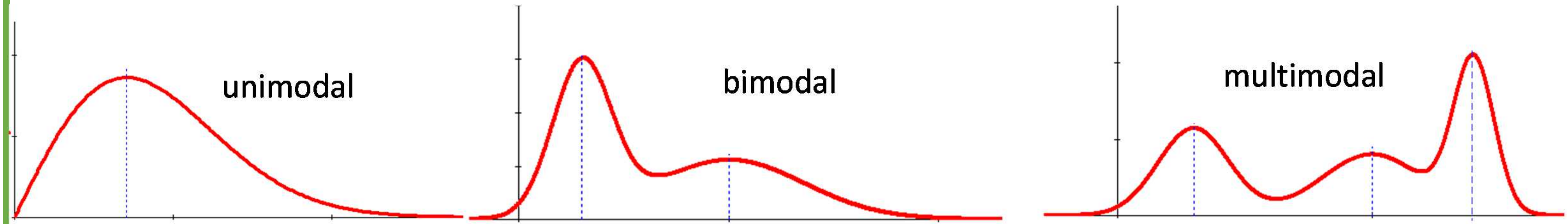
Mode

When the *pdf* of a continuous distribution has multiple local maxima, it is common to refer to all of the local maxima as *modes* of the distribution.

Such a continuous distribution is called *multimodal* (as opposed to *unimodal*).

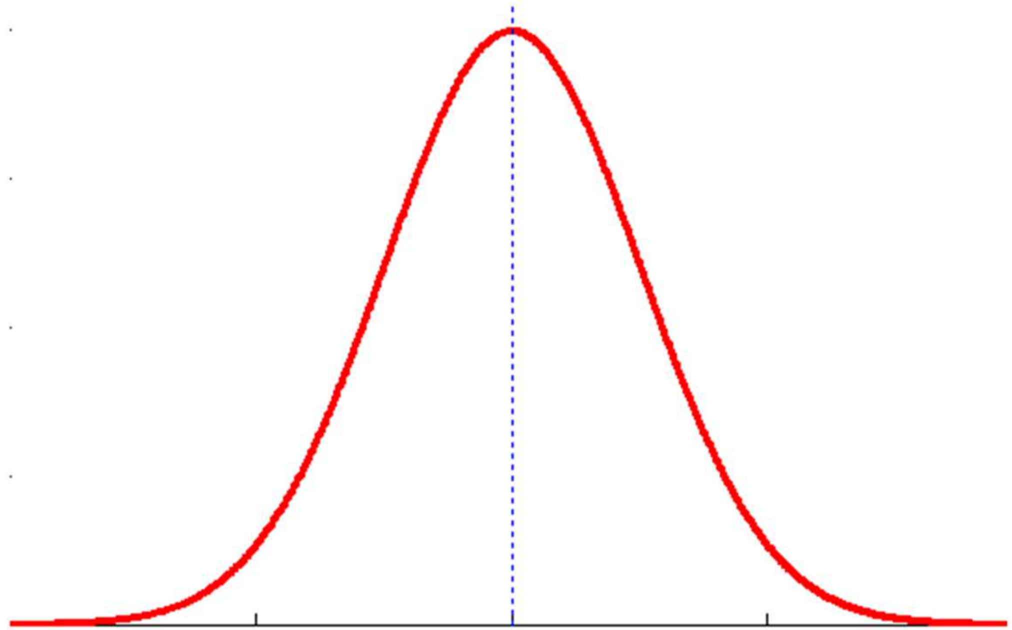
Mode

A mode of a continuous probability distribution is often considered to be any value x at which its *pdf* has a locally maximum value, so any peak is a mode.



Mode

In symmetric unimodal distributions, such as the normal distribution, the mean (if defined), median and mode all coincide.



Mode

Distribution	Parameters	Mode	Remarks
Bernoulli	$0 \le p \le 1$	$\begin{cases} 0, & if\ p < 0.5 \\ 0\ and\ 1, & if\ p = 0.5 \\ 1, & if\ p > 0.5 \end{cases}$	
Binomial	$n \ge 1, 0 \le p \le 1$	$\lfloor (n+1)p \rfloor$ or $\lceil (n+1)p \rceil - 1$	there is no single formula
Geometric	$0 \le p \le 1$	1	
Uniform	$a, b \in \mathbb{R}$	any value in the domain	
Exponential	$\lambda > 0$	0	
Weibull	$\eta, \beta > 0$	$\begin{cases} \eta \left(\frac{\beta-1}{\beta}\right)^{1/\beta}, & if\ \beta > 1 \\ 0, & if\ \beta \le 1 \end{cases}$	
Normal	$\mu, \sigma \in \mathbb{R}$	μ	

Lecture 6

Textbook Assignment

Géza Schay. *Introduction to Probability…*

❖Chapter 6.3 & 6.6. 198-205, 220-227 pp.

F.M. Dekking et al. *A Modern Introduction to…*

❖Chapter 5.6. 65-67 pp.