Lecture 8:

Moments

Summaries of a distribution

Summaries of a distribution

The mean is one of the **measures of central tendency**. Other important ones are:

Definition 6.1.1 (Median). c is a **median** of a r.v. X if $P(X \leq c) \geq 1/2$ and $P(X \geq c) \geq 1/2$. ($\geq$, because CDF may jump)

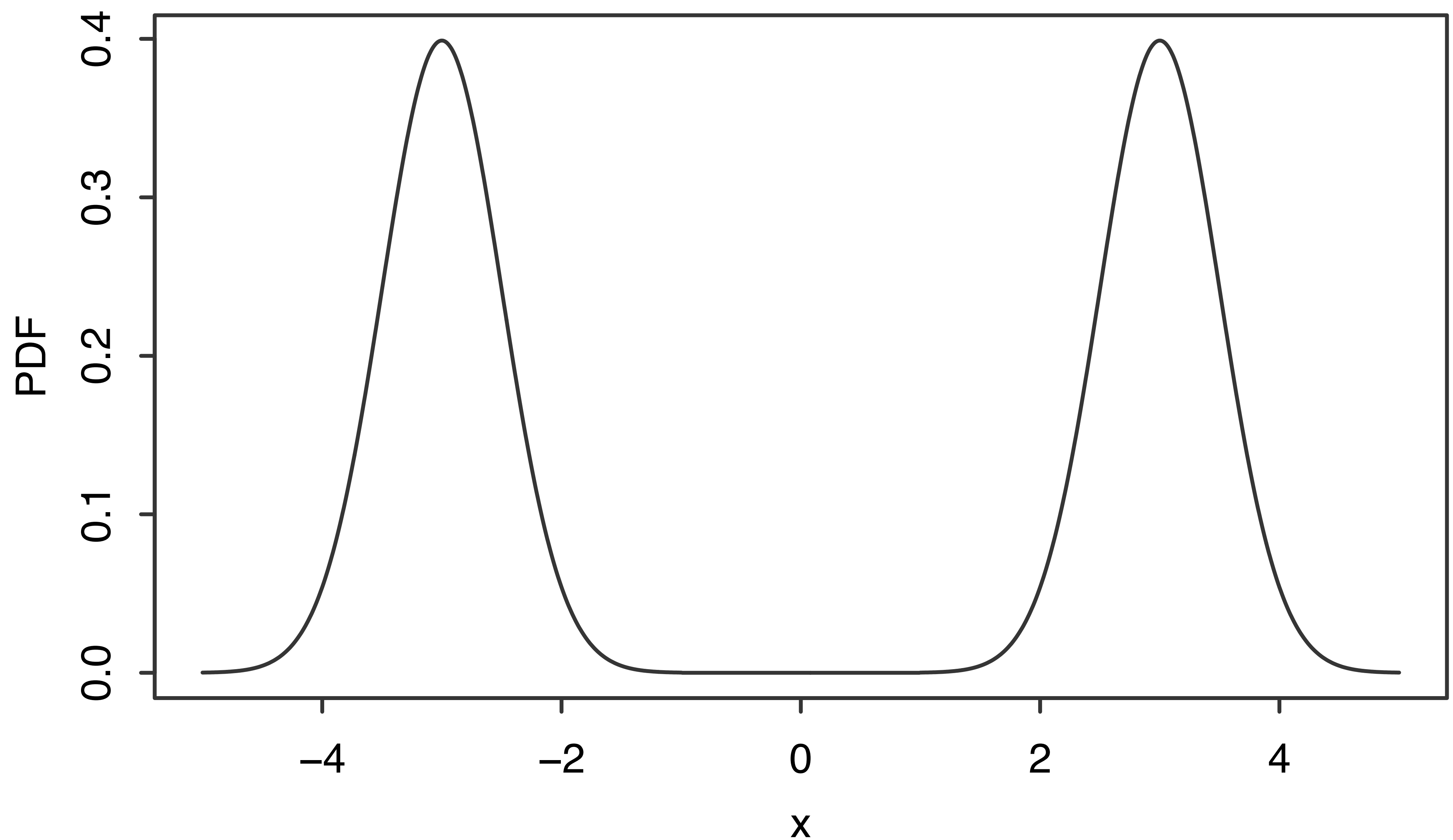
Definition 6.1.2 (Mode). c is a **mode** of a r.v. X if it maximises the PMF: $\forall x: \;\; P(X = c) \geq P(X = x)$. For a continuous r.v. X with PDF f, c is a mode if it maximises the PDF: $\forall x: \;\; f(c) \geq f(x)$

All r.v.-s that have the same distribution – have the same median and mode (just like the mean). So we talk about means, medians and modes of distributions (rather than of r.v.-s). For example, for $Z \sim \mathcal{N}(0,1)$ – median is 0 (from symmetry, $\Phi(0) = 1/2$).

Summaries of a distribution

Note that a distribution can have multiple medians and modes!

This distribution has 2 modes, and ∞-many medians (b.c. the PDF is 0 from -1 to 1):



Summaries of a distribution

Theorem 6.1.4. Let X – r.v. with mean = μ and median = m.

1) The value that minimises mean squared error $E(X-c)^2$ is $c = \mu$

2) The value that minimises mean absolute error $E|X-c|$ is $c = m$

Summaries of a distribution

Theorem 6.1.4. Let X – r.v. with mean = μ and median = m.

1) The value that minimises mean squared error $E(X-c)^2$ is $c=\mu$

2) The value that minimises mean absolute error $E|X-c|$ is

Proof: 1) First prove a useful fact: $E(X-c)^2 = \text{Var}(X) + (\mu - c)^2$:

$\text{Var}(X) = \text{Var}(X-c) = E(X-c)^2 - \big(E(X-c)\big)^2 = E(X-c)^2 - (\mu - c)^2$

$\text{Var}(X)$ doesn't depend on c, $E(X-c)^2$ is minimised by $c=\mu$.

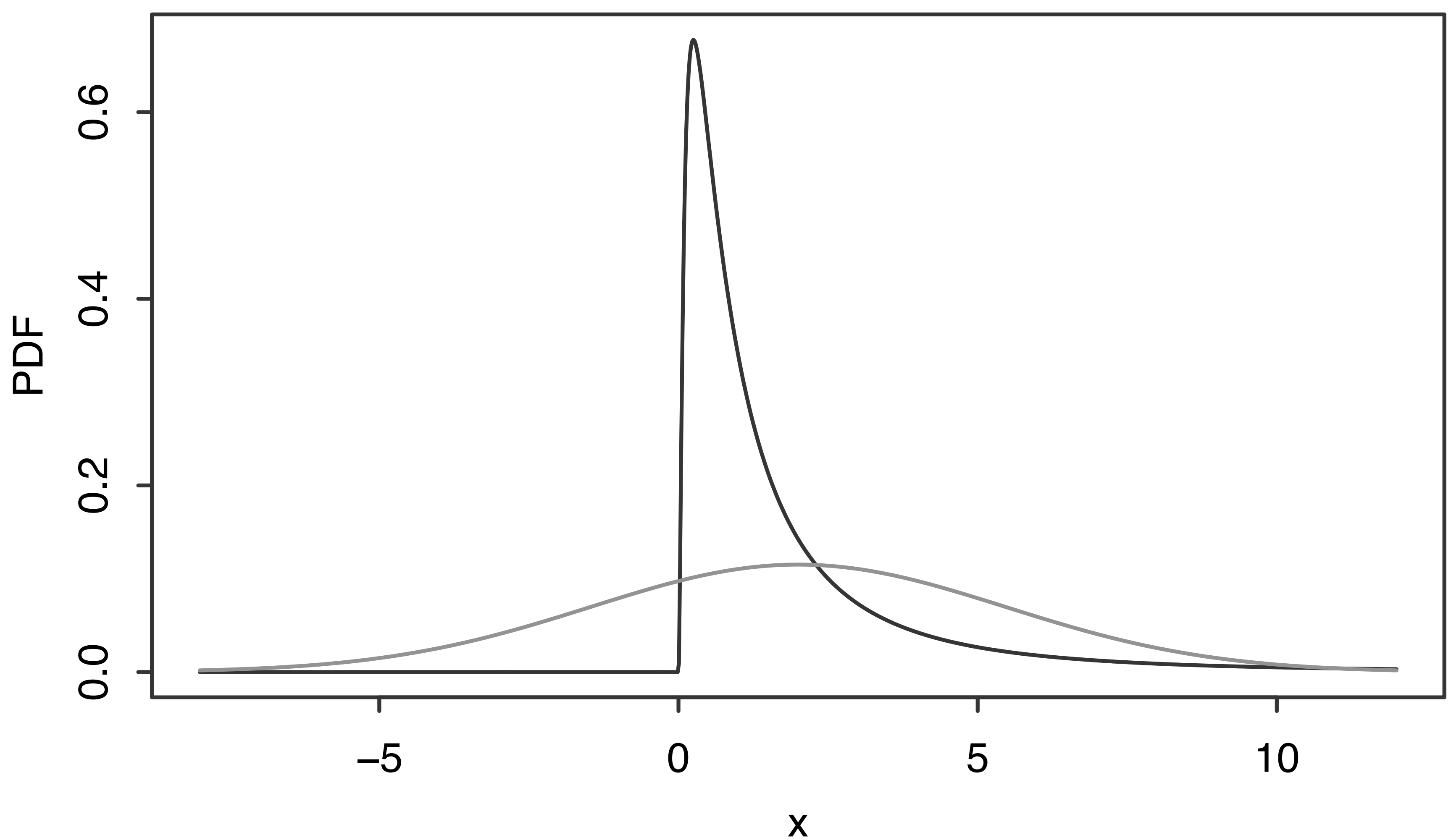
2) For $a \neq m$, show $E|X-m| \leq E|X-a|$ (that is, $E(Y) \geq 0$ for $Y = |X-a| - |X-m|$). Assume $m < a$ – then if $X \leq m$, $Y = a - X - (m - X) = a - m$, and if $X > m$, $Y = m - a$.

Let I - ind.($X \leq m$). Then $E(Y) = E(YI) + E(Y(1-I)) \geq$

$\geq (a-m)E(I) + (m-a)E(1-I) = \ldots = (a-m)(2P(X \leq m) - 1) \geq 0$

Summaries of a distribution

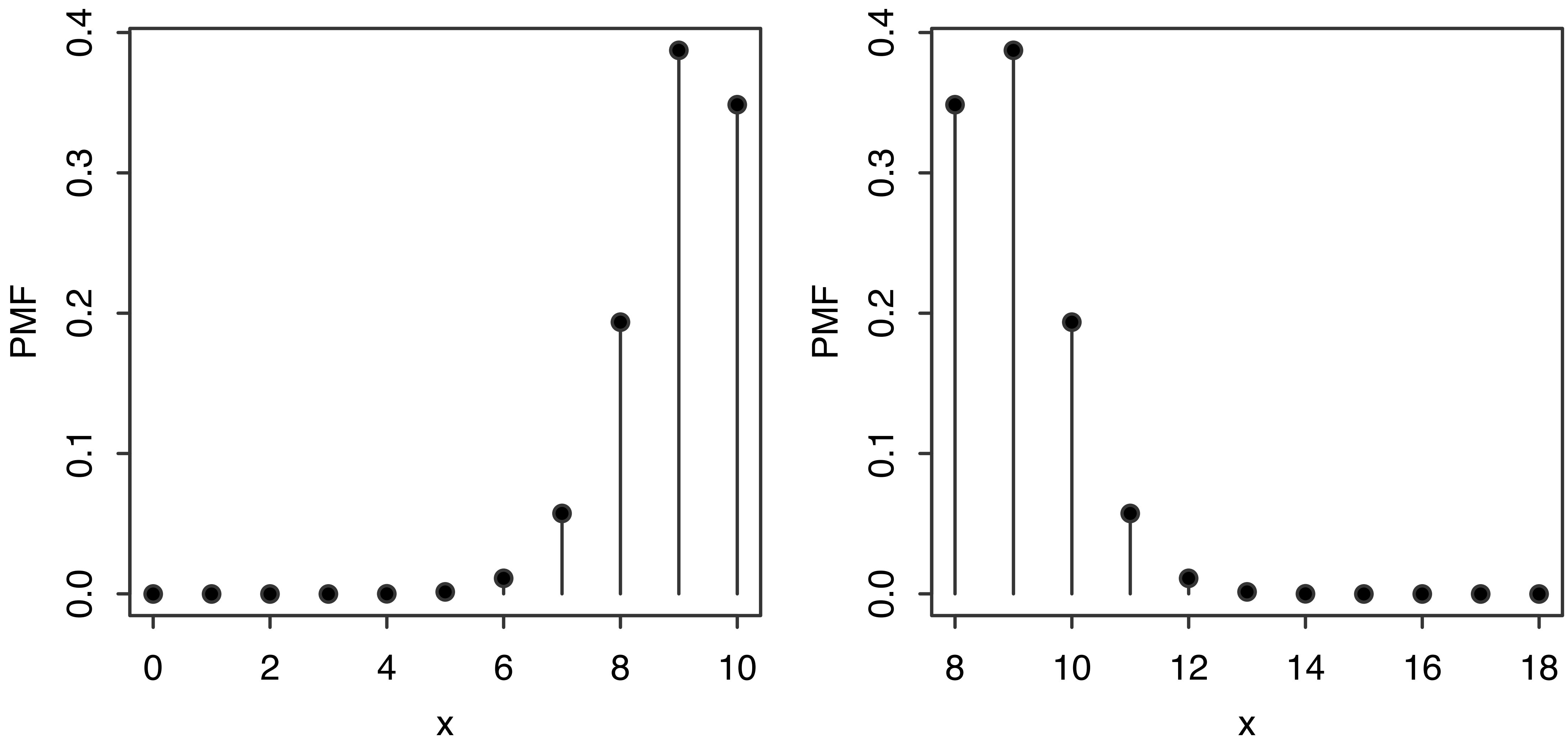
Variance is one of the measures of the **spread**. There are, however, some other major features that are not captured by the variance:



Here are two distributions (normal and **log-normal**) with same mean (=2) and variance (=12). Log-normal is **skewed** to the right

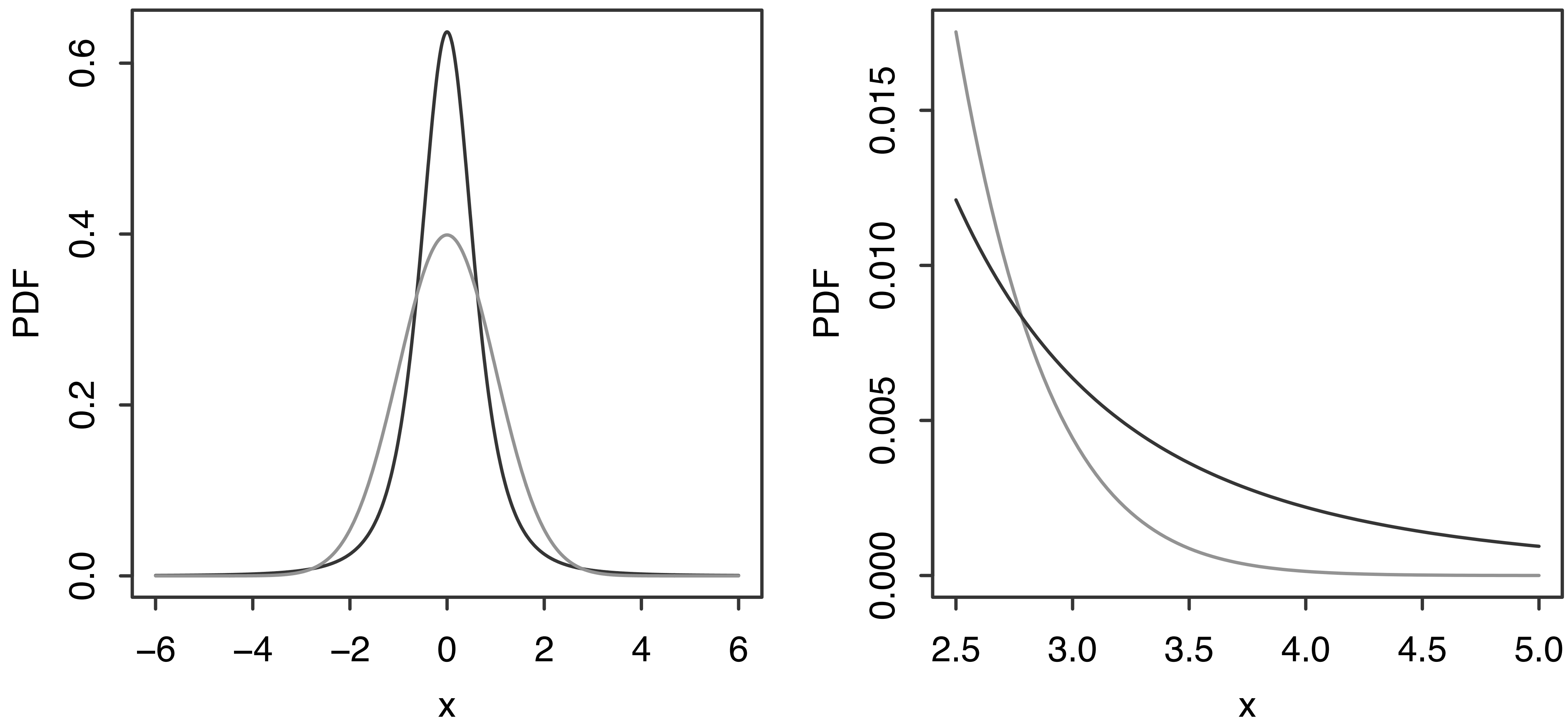
Summaries of a distribution

Two cases of Binomial distribution – Bin$(10, 0.9)$ and Bin$(10, 0.1)$. They have same mean and variance. First is left-skewed, second is right-skewed:



Summaries of a distribution

Those were a-symmetric distributions. Bus symmetric distributions can also be quite different:



Here are the Normal (light grey) and Student's t-distribution (dark). The latter has a higher peak and **heavier tails** (magnified on the right image). This is measured with **kurtosis**.

Interpreting moments

Interpreting moments

Definition 6.2.1 (Kinds of moments). Let X – r.v. with mean μ and variance σ^2. For any integer $n > 0$, n-th **moment** of X is $E(X^n)$. The n-th **central moment** is $E\big((X-\mu)^n\big)$, and n-th **standardised moment** is $E\left(\left(\frac{X-\mu}{\sigma}\right)^n\right)$. (If they exist).

We saw that the **mean** is the 1-st moment, and for a discrete r.v. it has a meaning of the **“center of mass”**.

Variance is the 2-nd central moment, and it has a meaning of **“moment of inertia”**.

We’ll now define **skewness** and **kurtosis**.

Interpreting moments

Definition 6.2.2 (Skewness). **Skewness** of a r.v. X with mean μ and variance σ^2 is its 3rd standardised moment: Skew$(X) = E\left(\frac{X-\mu}{\sigma}\right)^3$

(By standardising, we make skewness independent of location and scale of X. Also, the units in which we measure X won't affect).

Definition 6.2.3 (Symmetry of a r.v.) A r.v. X has a distribution **symmetric about** μ, if $(X - \mu)$ has the same distr. as $(\mu - X)$.

Proposition 6.2.5 (Symmetry in terms of PDF) A continuous r.v. X is symmetric iff its PDF $f(x)$ satisfies $f(x) = f(2\mu - x)$.

Interpreting moments

Proposition 6.2.6 (Odd central moments of symm. distr.-s). Let X - r.v. symmetric about its mean μ. Then **any** of its odd central moments equals zero, $E(X-\mu)^n$ (n-odd) (if it exists!).

Proof follows from the definition of symmetry of a distribution.
So, since 1-st central moment is always 0 (linearity), we can use 3-rd central moment to measure asymmetry = **skewness.**

Why not use, say, 5-th central moment? At least two reasons:
1) Higher order moments are harder to compute analytically
2) If we measure a **sample moment** and the sample has an **outlier** (improbably large value) – 5th power of $(x_i - \mu)$ will be an even larger number, suppressing the rest of the sum, so 5th moment is less stable (**robust**) than 3-rd.

Interpreting moments

Concerning outliers – rare (extreme) events (values) in a sample – the probability of these is given by the **tails** of the distribution: “fat tails” = outliers are more probable. It is believed that rare events that happen in our life (catastrophes, stock market crashes, etc.) are a consequence of fat tails.
The “fatness” of tails is measured by:

Definition 6.2.7 (Kurtosis). The **kurtosis** of a r.v. X with mean μ and variance σ^2 is a (shifted) version of the 4-th standardised moment:
Kurt$(X) = E\left(\frac{X-\mu}{\sigma}\right)^4 - 3$

(Subtracting 3, so that normal distribution has zero kurtosis!)

Sample moments

Sample moments

Having a sample from a certain distribution, one can be interested in estimates of parameters of that distribution from the sample – that is the subject of **statistical inference**.

Definition 6.3.1 (Sample moments). Let $X_1, \ldots, X_n$ – i.i.d. random variables. k-th **sample moment** is: $M_k = \frac{1}{n} \sum_{j=1}^{n} X_j^k$

The **sample mean** $\overline{X}_n$ is the 1-st sample moment. In contrast, the **population mean** or **true mean** is $E(X_j)$, the mean of the distribution from which X_j were drawn from.

The **law of large numbers** says that k-th sample moment of i.i.d. random variables ***converges*** to the k-th true moment of the distribution, as $n \to \infty$.

Sample moments

Theorem 6.3.2 (Mean and variance of sample mean). Let $X_1, \ldots, X_n$ – i.i.d. r.v.-s with mean μ and variance σ^2. Then the sample mean $\overline{X}_n$ is **unbiased** for estimating μ – that is, $E(\overline{X}_n) = \mu$.
The variance is given by Var$(\overline{X}_n) = \sigma^2/n$.
Proof: From linearity for the mean. For variance – Var(sum ind. r.v.-s) = sum(vars).

Definition 6.3.3 (Sample variance and sample std). Let $X_1, \ldots, X_n$ – i.i.d. r.v.-s. The **sample variance** is the following r.v.:
$S_n^2 = \frac{1}{n-1} \sum_{j=1}^{n} (X_j - \overline{X}_n)^2$. The **sample std** is the sqrt of that.

$1/(n-1)$ instead of $1/n$ makes this an **unbiased** estimator!