Anna Janicka

# Probability Calculus 2021/2022
# Lecture 7

## 1. Variance

The expected value is not the only parameter useful in describing random variables. A second commonly used characteristic is the variance (and its square root, the standard deviation).

**Definition 1.** *Let $X$ be a random variable such that $\mathbb{E}|X| < \infty$ and $\mathbb{E}(X - \mathbb{E}X)^2 < \infty$. The* **variance** *of $X$ is defined as*

$$D^2X = VarX = \mathbb{E}(X - \mathbb{E}X)^2.$$

*The* **standard deviation** *of variable $X$ is the square root of the variance: $\sigma_X = \sqrt{D^2X}$.*

Note that the variance depends only on the distribution of the random variable, therefore in many cases we will refer to the variance of a distribution. The two conditions for the existence of the variance may be simplified slightly: it may be shown that it suffices that $\mathbb{E}X^2 < \infty$. If a random variable is limited, its variance always exists.

In many cases, it is easier to calculate the variance using a simplified formula rather than straight from the definition:

$$\begin{aligned} D^2X =& \mathbb{E}(X - \mathbb{E}X)^2 = \mathbb{E}(X^2 - 2X\mathbb{E}X + (\mathbb{E}X)^2) = \mathbb{E}(X^2) - 2\mathbb{E}X\mathbb{E}X + (\mathbb{E}X)^2 = \\ & \mathbb{E}X^2 - (\mathbb{E}X)^2. \end{aligned}$$

The variance of a random variable is the mean of the square of the deviance of the random variable from its mean. Therefore, we expect that variables with small variances will take on values (relatively) close to their means, while variables with large variances – far from their means. In many cases we wish to describe the deviation of a variable from its mean with the use of a parameter which is of the same measure as the values of the variable itself – in such cases, we will revert to the standard deviation.

Examples:

(1) Let $X$ and $Y$ be two variables with zero mean: $\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = \frac{1}{2}$ and $\mathbb{P}(Y = 100) = \mathbb{P}(Y = -100) = \frac{1}{2}$. The variance of $X$ is equal to

$$D^2X = \frac{1}{2}(1^2 + (-1)^2) - 0^2 = 1, \text{ with } \sigma_X = 1,$$

while the variance of $Y$ is equal to

$$D^2Y = \frac{1}{2}(100^2 + (-100)^2) - 0^2 = 100^2 \text{ with } \sigma_Y = 100.$$

We see that the variance (and the standard deviation) of $Y$ is much larger than that of $X$, as the values of $Y$ are further from the mean than the values of $X$.

(2) Let $X$ be the number of points obtained on a die in a single roll. We already know that $\mathbb{E}X = \frac{7}{2}$ and $\mathbb{E}X^2 = \frac{91}{6}$, so that

$$D^2X = \frac{91}{6} - \frac{49}{4} = \frac{35}{12}$$

.

(3) Let $X$ have a uniform distribution over $[a, b]$. We already know that $\mathbb{E}X = \frac{a+b}{2}$. We have

$$\mathbb{E}X^2 = \int_a^b x^2 \frac{1}{b-a} dx = \frac{1}{b-a} \cdot \frac{b^3 - a^3}{3} = \frac{b^2 + ab + a^2}{3},$$

and therefore

$$D^2X = \frac{b^2 + ab + a^2}{3} - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12}.$$

The variance, unlike the expected value, is not a linear operator. It has, however, some properties which allow to simplify calculations of linear combinations of random variables:

**Theorem 1.** *Let $X$ be a random variable with a variance.*
*(i) $D^2X \geqslant 0$, and the equality holds if and only if there exists a value $a \in \mathbb{R}$ such that $\mathbb{P}(X = a) = 1$ (i.e., $X \sim \delta_a$).*
*(ii) $D^2(bX) = b^2 D^2X$ for any $b \in \mathbb{R}$.*
*(iii) $D^2(X + c) = D^2X$ for any $c \in \mathbb{R}$.*

(4) We will use the above theorem to simplify calculations of the parameters of a normal distribution $\mathcal{N}(m, \sigma^2)$. Let $m \in \mathbb{R}$, $\sigma > 0$ and let $X \sim \mathcal{N}(0, 1)$ and $Y = \sigma X + m$. We have:

$$F_Y(t) = \mathbb{P}(\sigma X + m \leqslant t) = \mathbb{P}\left(X \leqslant \frac{t-m}{\sigma}\right) = \int_{-\infty}^{\frac{t-m}{\sigma}} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right) dx.$$

Substituting $x = \frac{y-m}{\sigma}$, we have $dx = \frac{dy}{\sigma}$ and

$$F_Y(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y-m)^2}{2\sigma^2}\right) dy,$$

which shows that $Y = \sigma X + m$ is distributed normally: $Y \sim \mathcal{N}(m, \sigma^2)$.

Making use of the properties of the expected value and variance, we may write:

$$\mathbb{E}Y = \mathbb{E}(\sigma X + m) = \sigma \mathbb{E}X + m = m,$$

and

$$D^2Y = D^2(\sigma X + m) = D^2(\sigma X) = \sigma^2 D^2 X.$$

We will now calculate the variance of the standard normal distribution:

$$\begin{aligned} D^2X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \mathbb{E}X^2 &= \int_{-\infty}^{\infty} x^2 \cdot \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) dx \\ &= \int_{-\infty}^{\infty} x \cdot \left(-\frac{1}{\sqrt{2\pi}} \exp(-x^2/2)\right)' dx \\ &= \left(-x\frac{1}{\sqrt{2\pi}} \exp(-x^2/2)\right)\Bigg|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} 1 \cdot \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) dx = 0 + 1 = 1. \end{aligned}$$

This means that

$$D^2Y = \sigma^2 \cdot 1 = \sigma^2,$$

i.e. the $m$ and $\sigma$ parameters of the normal distribution describe the mean and the standard deviation, respectively.

## 2. Moments

Both the mean and the variance of a random variable are special cases of parameters referred to as *moments*:

**Definition 2.** *For $p \in (0, \infty)$, we define:*
*(i) the* **absolute moment** *of rank $p$ for random variable $X$ as $\mathbb{E}|X|^p$ (if this value is finite);*
*For $p \in \mathbb{N}$, we define:*
*(ii) the* **moment** *of rank $p$ for random variable $X$ as $\mathbb{E}X^p$ (provided that the $p$-th absolute moment exists);*
*(iii) the* **central moment** *of rank $p$ for random variable $X$ as $\mathbb{E}(X - \mathbb{E}X)^p$ (provided that the $p$-th absolute moment exists).*

The mean is the first moment; the variance is the second central moment.

Moments play a big role in statistics. The most popular distribution parameters used for describing random variables, apart from the mean and the variance, are based on the third and fourth central moments:

**Definition 3.** *Let $X$ be a random variable such that $\mathbb{E}|X|^3 < \infty$. The* **skewness** *of $X$ is*

$$\alpha_3 = \frac{\mathbb{E}(X - \mathbb{E}X)^3}{(D^2X)^{3/2}} = \frac{\mathbb{E}(X - \mathbb{E}X)^3}{\sigma_X^3}.$$

**Definition 4.** *Let $X$ be a random variable such that $\mathbb{E}|X|^4 < \infty$. The* **kurtosis** *of $X$ is*

$$\alpha_4 = \frac{\mathbb{E}(X - \mathbb{E}X)^4}{(D^2X)^2} - 3 = \frac{\mathbb{E}(X - \mathbb{E}X)^4}{\sigma_X^4} - 3.$$

Skewness describes the shape of the distribution. It may be:

- positive (the random variable is then said to be right-skewed, right-tailed, or skewed to the right), when the "right tail" of a unimodal distribution is longer or fatter (the mass of the distribution is on the left hand side, i.e. with relatively few high values); or
- negative (the random variable is then said to be left-skewed, left-tailed, or skewed to the left), when the "left tail" of a unimodal distribution is longer or fatter (the mass of the distribution is on the right hand side, i.e. with relatively few low values); or
- zero – if the distribution is symmetric around the mean, but not only then (the situation becomes complicated especially for multi-modal distributions).

For example, a random variable which takes on four, equally probable, values: 1, 2, 3, 1000 would be skewed to the right ($\alpha_3 \approx 1.15$), while a random variable which takes on values 1, 1000, 1001, 1002 with probabilities equal to $\frac{1}{4}$ would be skewed to the left ($\alpha_3 \approx -1.15$). A random variable with equally probable values of $-1000, -1001, 1000, 1001$ has zero skewness.

Kurtosis describes the concentration ("peakedness") of the distribution. Kurtosis (nowadays) is defined by comparison with the standard normal distribution (for which the fourth central moment and thus the fraction from the definition is equal to 3), and is sometimes referred to as *excess kurtosis*. Kurtosis may be positive (a leptokurtic distribution), when the distribution is more "peaked" than the standard normal distribution; or negative (a platykurtic distribution), when the distribution is more flat than the standard normal distribution; or zero.

Example: Let $X$ be a random variable from a standard normal distribution. Then, $\mathbb{E}X = 0$, and the central moments reduce to ordinary moments. We have:

$$\alpha_3 = \frac{\mathbb{E}(X - \mathbb{E}X)^3}{\sigma^3} = \frac{\mathbb{E}X^3}{1} = \int_{-\infty}^{\infty} x^3 \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 0,$$

because the integrand is an odd function – the distribution is not skewed in either direction (that's because it is symmetric). We also have

$$\begin{aligned}\mathbb{E}X^4 &= \int_{-\infty}^{\infty} x^4 \cdot \frac{1}{2\pi} e^{-x^2/2} dx = \int_{-\infty}^{\infty} x^3 \cdot \left(-\frac{1}{\sqrt{2\pi}} \exp(-x^2/2)\right)' dx \\ &= \left(-x^3 \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)\right)\Bigg|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} 3x^2 \cdot \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) dx = 0 + 3\mathbb{E}X^2 = 3,\end{aligned}$$

which means that

$$\alpha_4 = \frac{\mathbb{E}(X - \mathbb{E}X)^4}{\sigma_X^4} - 3 = \frac{\mathbb{E}X^4}{1} - 3 = 3 - 3 = 0.$$

Therefore, the standard normal distribution has zero (excess) kurtosis.

## 3. Empirical distributions

In most real-life situations, we are faced with random variables for which we can not specify distributions nor, in some cases, even the range of values; this may be the case for example if we want to study the distribution of earnings in a population. Furthermore, in most cases we are also faced not with whole populations, but rather with their subsets (called samples). In such cases, if we wish to formulate statements or test hypotheses based on the available information, we will need to revert to statistical methods. At this point we will just signal that the latter are in many cases based on the so-called empirical distributions.

**Definition 5.** *Let $X_1, X_2, \ldots, X_n$ be random variables with unknown distributions. An* **empirical distribution (measure)** *for this sample is $\mu_n(A) = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}(A) = \frac{|\{i \leqslant n \colon X_i \in A\}|}{n}$*

**Definition 6.** *An* **empirical distribution function** *of the sample $X_1, X_2, \ldots, X_n$ is the function $F \colon \mathbb{R} \to [0,1]$, such that $F_n(t) = \mu_n((-\infty, t]) = \frac{|\{i \leqslant n \colon X_i \leqslant t\}|}{n}$.*

This is the CDF of the empirical distribution.

**Definition 7.** *A* **Quantile** *of rank $p$ of the sample $X_1, \ldots, X_n$ is any number $x_p$, such that $\mu_n((-\infty, x_p]) \geqslant p$, and $\mu_n([x_p, \infty)) \geqslant 1 - p$.*

These are the quantiles of the empirical distribution.

**Definition 8.** *A* **sample mean** *for $X_1, X_2, \ldots, X_n$ is equal to $m = \frac{X_1 + X_2 + \ldots + X_n}{n}$, i.e. the arithmetic mean of $X_1, X_2, \ldots, X_n$.*

**Definition 9.** *A* **sample variance** *for $X_1, X_2, \ldots, X_n$ is equal to $s^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2$, where $m$ is the sample mean.*

The sample mean and the sample variance are the mean and the variance of the empirical distribution, respectively.