

# Some example Mathematical lecture notes

Dr. A. Lecturer

March 4, 2024



# Contents

<b>Preface</b>	<b>i</b>
<b>1 Determinants</b>	<b>1</b>
1.1 Introduction . . . . .	1
1.2 Two by two matrices . . . . .	2
1.2.1 First properties . . . . .	2
1.3 Geometric meaning . . . . .	3
<b>2 Variance</b>	<b>5</b>
2.1 Introduction . . . . .	5
2.1.1 Definition . . . . .	6
2.1.2 Properties . . . . .	7



# Preface

These are some example lecture notes, designed to test out various systems for converting LaTeX notes to more accessible formats. The content is modified from the Wikipedia articles on the determinant [1] and variance [2].



# Chapter 1

## Determinants

### 1.1 Introduction

In mathematics, the **determinant** is a scalar value that is a function of the entries of a square matrix. The determinant of a matrix is commonly denoted  $\det(A)$ ,  $\det A$ , or  $|A|$ . Its value characterizes some properties of the matrix and the linear map represented by the matrix. In particular, the determinant is nonzero if and only if the matrix is invertible and the linear map represented by the matrix is an isomorphism. The determinant of a product of matrices is the product of their determinants.

The determinant of a  $2 \times 2$  matrix is

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc,$$

and the determinant of a  $3 \times 3$  matrix is

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = aei + bfg + cdh - ceg - bdi - afh.$$

The determinant of an matrix can be defined in several equivalent ways, the most common being Leibniz formula, which expresses the determinant as a sum of  $n!$  (the factorial of  $n$ ) signed products of matrix entries. It can be computed by the Laplace expansion, which expresses the determinant as a linear combination of determinants of submatrices, or with Gaussian elimination, which expresses the determinant as the product of the diagonal entries of a diagonal matrix that is obtained by a succession of elementary row operations.

Determinants can also be defined by some of their properties. Namely, the determinant is the unique function defined on the matrices that has the four following properties:

1. The determinant of the identity matrix is 1.
2. The exchange of two rows multiplies the determinant by  $-1$ .
3. Multiplying a row by a number multiplies the determinant by this number.

4. Adding to a row a multiple of another row does not change the determinant.

The above properties relating to rows (properties 2-4) may be replaced by the corresponding statements with respect to columns.

Determinants occur throughout mathematics. For example, a matrix is often used to represent the coefficients in a system of linear equations, and determinants can be used to solve these equations (Cramer's rule), although other methods of solution are computationally much more efficient. Determinants are used for defining the characteristic polynomial of a matrix, whose roots are the eigenvalues. In geometry, the signed  $n$ -dimensional volume of a  $n$ -dimensional parallelepiped is expressed by a determinant, and the determinant of (the matrix of) a linear transformation determines how the orientation and the  $n$ -dimensional volume are transformed. This is used in calculus with exterior differential forms and the Jacobian determinant, in particular for changes of variables in multiple integrals.

## 1.2 Two by two matrices

The determinant of a matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad (1.1)$$

is denoted either by “det” or by vertical bars around the matrix, and is defined as

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$$

For example,

$$\det \begin{pmatrix} 3 & 7 \\ 1 & -4 \end{pmatrix} = \begin{vmatrix} 3 & 7 \\ 1 & -4 \end{vmatrix} = (3 \cdot (-4)) - (7 \cdot 1) = -19.$$

### 1.2.1 First properties

The determinant has several key properties that can be proved by direct evaluation of the definition for  $2 \times 2$ -matrices, and that continue to hold for determinants of larger matrices. They are as follows: first, the determinant of the identity matrix  $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$  is 1. Second, the determinant is zero if two rows are the same:

$$\begin{vmatrix} a & b \\ a & b \end{vmatrix} = ab - ba = 0.$$

This holds similarly if the two columns are the same. Moreover,

$$\begin{vmatrix} a & b+b' \\ c & d+d' \end{vmatrix} = a(d+d') - (b+b')c = \begin{vmatrix} a & b \\ c & d \end{vmatrix} + \begin{vmatrix} a & b' \\ c & d' \end{vmatrix}.$$

Finally, if any column is multiplied by some number  $r$  (i.e., all entries in that column are multiplied by that number), the determinant is also multiplied by that number:

$$\begin{vmatrix} r \cdot a & b \\ r \cdot c & d \end{vmatrix} = rad - brc = r(ad - bc) = r \cdot \begin{vmatrix} a & b \\ c & d \end{vmatrix}.$$



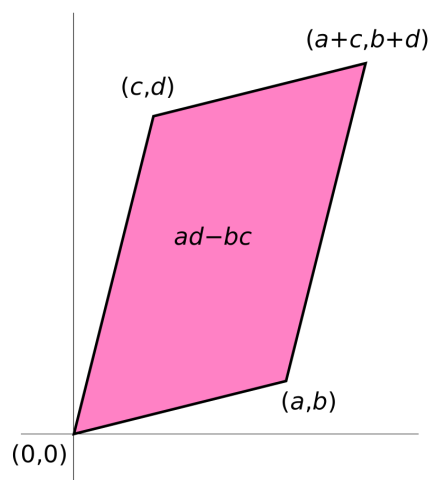


Figure 1.1: The area of the parallelogram is the absolute value of the determinant of the matrix formed by the vectors representing the parallelogram's sides.

### 1.3 Geometric meaning

If the matrix entries are real numbers, the matrix  $A$  can be used to represent two linear maps: one that maps the standard basis vectors to the rows of  $A$ , and one that maps them to the columns of  $A$ . In either case, the images of the basis vectors form a parallelogram that represents the image of the unit square under the mapping. The parallelogram defined by the rows of the matrix (1.1) is the one with vertices at  $(0,0)$ ,  $(a,b)$ ,  $(a+c,b+d)$ , and  $(c,d)$ , as shown in Figure 1.1.

The absolute value of  $ad-bc$  is the area of the parallelogram, and thus represents the scale factor by which areas are transformed by  $A$ . (The parallelogram formed by the columns of  $A$  is in general a different parallelogram, but since the determinant is symmetric with respect to rows and columns, the area will be the same.)

The absolute value of the determinant together with the sign becomes the *oriented area* of the parallelogram. The oriented area is the same as the usual area, except that it is negative when the angle from the first to the second vector defining the parallelogram turns in a clockwise direction (which is opposite to the direction one would get for the identity matrix).

[Original Wikipedia article continues...]



## Chapter 2

# Variance

### 2.1 Introduction

In probability theory and statistics, **variance** is the expected value of the squared deviation from the mean of a random variable. The standard deviation (SD) is obtained as the square root of the variance. Variance is a measure of dispersion, meaning it is a measure of how far a set of numbers is spread out from their average value. It is the second central moment of a distribution, and the covariance of the random variable with itself, and it is often represented by  $\sigma^2$ ,  $s^2$ ,  $\text{Var}(X)$ ,  $V(X)$ , or  $\mathbb{V}(X)$ .

An advantage of variance as a measure of dispersion is that it is more amenable to algebraic manipulation than other measures of dispersion such as the expected absolute deviation; for example, the variance of a sum of uncorrelated random variables is equal to the sum of their variances. A disadvantage of the variance for practical applications is that, unlike the standard deviation, its units differ from the random variable, which is why the standard deviation is more commonly reported as a measure of dispersion once the calculation is finished.

There are two distinct concepts that are both called “variance”. One, as discussed above, is part of a theoretical probability distribution and is defined by an equation. The other variance is a characteristic of a set of observations. When variance is calculated from observations, those observations are typically measured from a real-world system. If all possible observations of the system are present, then the calculated variance is called the population variance. Normally, however, only a subset is available, and the variance calculated from this is called the sample variance. The variance calculated from a sample is considered an estimate of the full population variance. There are multiple ways to calculate an estimate of the population variance, as discussed in the section below.

The two kinds of variance are closely related. To see how, consider that a theoretical probability distribution can be used as a generator of hypothetical observations. If an infinite number of observations are generated using a distribution, then the sample variance calculated from that infinite set will match the value calculated using the distribution’s equation for variance. Variance has a central role in statistics, where some ideas that use it include descriptive

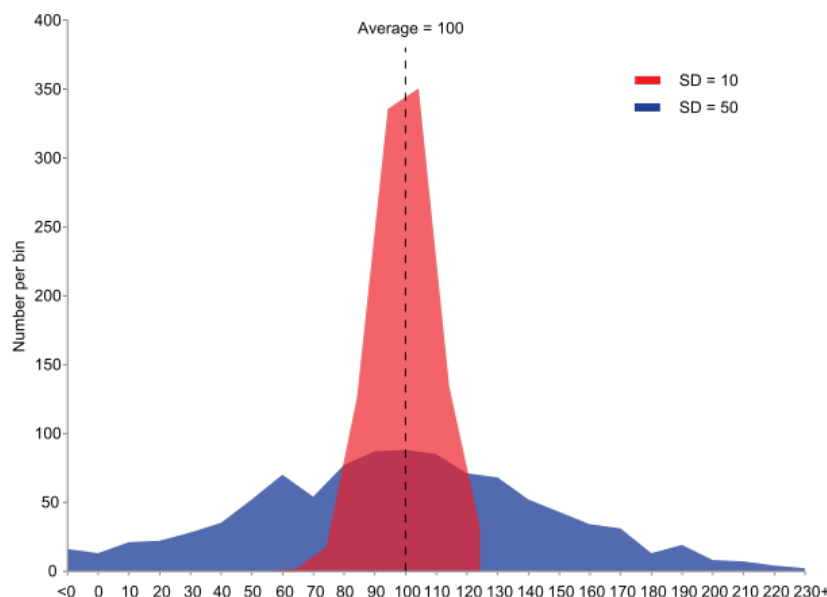


Figure 2.1: Example of samples from two populations with the same mean but different variances. The red population has mean 100 and variance 100 (SD=10) while the blue population has mean 100 and variance 2500 (SD=50) where SD stands for Standard Deviation.

statistics, statistical inference, hypothesis testing, goodness of fit, and Monte Carlo sampling.

### 2.1.1 Definition

The variance of a random variable  $X$  is the expected value of the squared deviation from the mean of  $X$ ,  $\mu = E[X]$ :

$$\text{Var}(X) = E[(X - \mu)^2].$$

This definition encompasses random variables that are generated by processes that are discrete, continuous, neither, or mixed. The variance can also be thought of as the covariance of a random variable with itself:

$$\text{Var}(X) = \text{Cov}(X, X).$$

The variance is also equivalent to the second cumulant of a probability distribution that generates  $X$ . The variance is typically designated as  $\text{Var}(X)$ , or sometimes as  $V(X)$  or  $\mathbb{V}(X)$ , or symbolically as  $\sigma_X^2$  or simply  $\sigma^2$  (pronounced “sigma squared”). The expression for the variance can be expanded as follows:

$$\begin{aligned} \text{Var}(X) &= E[(X - E[X])^2] \\ &= E[X^2 - 2X E[X] + E[X]^2] \\ &= E[X^2] - 2E[X]E[X] + E[X]^2 \\ &= E[X^2] - E[X]^2 \end{aligned}$$

In other words, the variance of  $X$  is equal to the mean of the square of  $X$  minus the square of the mean of  $X$ . This equation should not be used for computations using floating point arithmetic, because it suffers from catastrophic cancellation if the two components of the equation are similar in magnitude.

[...Original Wikipedia article continues ...]

### Commonly used probability distributions

Table 2.1 lists the variance for some commonly used probability distributions.

Name	Probability distribution function	Mean	Variance
Binomial distribution	$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$	$np$	$np(1-p)$
Geometric distribution	$\Pr(X = k) = (1-p)^{k-1} p$	$\frac{1}{p}$	$\frac{(1-p)}{p^2}$
Normal distribution	$f(x   \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$\mu$	$\sigma^2$
Uniform distribution	$f(x   a, b) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b, \\ 0 & \text{for } x < a \text{ or } x > b \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponential distribution	$f(x   \lambda) = \lambda e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Poisson distribution	$f(k   \lambda) = \frac{e^{-\lambda} \lambda^k}{k!}$	$\lambda$	$\lambda$

Table 2.1: Basic properties of some commonly used probability distributions.

### 2.1.2 Properties

#### Basic properties

[This section is altered from to the original Wikipedia article to try out the AMS theorem environments.]

**Theorem 1.** *The variance of any random variable  $X$  is non-negative*

$$\text{Var}(X) \geq 0.$$

*Proof.* Since  $(X - \mu)^2 \geq 0$ ,

$$\text{Var}(X) = \text{E}[(X - \mu)^2] \geq 0.$$

□

[Original Wikipedia article continues ...]



# Bibliography

- [1] Wikipedia contributors. Determinant — Wikipedia, the free encyclopedia, 2024. [Online; accessed 21-February-2024].
- [2] Wikipedia contributors. Variance — Wikipedia, the free encyclopedia, 2024. [Online; accessed 21-February-2024].