ECE 30200 - Probabilistic Methods in Electrical and Computer Engineering

Zeke Ulrich

July 4, 2025

Contents

- *Background* 2
  - *Series* 2
  - *Combinatorics* 2
  - *Approximations* 2
  - *Calculus* 3
  - *Linear Algebra* 3
  - *Set Theory* 4
  - *Probability Laws* 4
  - *Probability Properties* 5
- *Formal Definitions* 6
  - *Outcomes* 6
  - *Events* 6
  - *Probability Laws* 6
  - *Probability Space* 6
  - *Probability Properties* 7
- *Discrete Random Variables* 8
- *Continuous Random Variables* 13
- *Reference* 17
  - *Series* 17
  - *Combinatorics* 17
  - *Approximations* 17
  - *Calculus* 18
  - *Linear Algebra* 18

Background

The following formulas will be instrumental and may be familar.

Series

$$\sum_{k=0}^{n} r^k = \frac{1-r^{n+1}}{1-r} \tag{1}$$

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6} \tag{2}$$

$$\sum_{k=1}^{\infty} k r^{k-1} = \frac{1}{(1-r)^2} \tag{3}$$

Combinatorics

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \tag{4}$$

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k \tag{5}$$

$$\binom{n}{k} + \binom{n}{k-1} = \binom{n+1}{k} \tag{6}$$

$$P(n,k) = \frac{n!}{(n-k)!} \tag{7}$$

where $P(n,k)$ is the number of ways to arrange k objects out of n (permutations).

$$C(n,k) = \binom{n}{k} = \frac{n!}{k!(n-k)!} \tag{8}$$

where $C(n,k)$ is the number of ways to choose k objects out of n (combinations).

Approximations

$$\begin{aligned} f(x) &= f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \dots && (9) \\ &= \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n && (10) \end{aligned}$$

$$\begin{aligned} 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots &= \sum_{k=0}^{\infty} \frac{x^k}{k!} && (11) \\ &= e^x && (12) \end{aligned}$$

$$\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots \tag{13}$$
$$= \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!} \tag{14}$$

$$\cos(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \dots \tag{15}$$
$$= \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!} \tag{16}$$

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots \tag{17}$$
$$= \sum_{n=1}^{\infty} (-1)^{n+1} \frac{x^n}{n} \tag{18}$$

Calculus

$$\frac{d}{dx} \int_a^x f(t)\, dt = f(x) \tag{19}$$
$$\int_a^b f'(x)\, dx = f(b) - f(a) \tag{20}$$
$$\int f(g(x)) g'(x)\, dx = \int f(u)\, du \tag{21}$$
$$\int u\, dv = uv - \int v\, du \tag{22}$$
$$\int \frac{1}{(x-a)(x-b)}\, dx = \frac{1}{b-a} \ln \left| \frac{x-a}{x-b} \right| + C \tag{23}$$

Linear Algebra

$$\vec{y} = \beta_1 \vec{x}_1 + \beta_2 \vec{x}_2 + \dots + \beta_N \vec{x}_N \tag{24}$$

$$\langle \vec{a}, \vec{b} \rangle = \vec{a}\vec{b}^T \tag{25}$$
$$= \sum_{i=1}^{n} a_i b_i \tag{26}$$

where $\langle \vec{a}, \vec{b} \rangle$ denotes the inner product of vectors $\vec{a}$ and $\vec{b}$.

$$\|\vec{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p} \tag{27}$$

where $\|\vec{x}\|_p$ is the p-norm (or ℓ_p-norm) of vector $\vec{x}$.

$$\cos(\theta) = \frac{\langle \vec{a}, \vec{b} \rangle}{\|\vec{a}\|_2 \|\vec{b}\|_2} \tag{28}$$

where θ is the angle between vectors $\vec{a}$ and $\vec{b}$.

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\vec{y} \tag{29}$$

where $\hat{\beta}$ is the vector of least squares coefficients, $\mathbf{X}$ is the data matrix, and $\vec{y}$ is the target vector

Set Theory

The *set difference* $A \setminus B$ is the set of elements that are in A but not in B:

$$A \setminus B = \{x \mid x \in A \text{ and } x \notin B\} \tag{30}$$

Some important properties of set operations are:

- **Commutativity:**

$$A \cup B = B \cup A \tag{31}$$
$$A \cap B = B \cap A \tag{32}$$

- **Associativity:**

$$(A \cup B) \cup C = A \cup (B \cup C) \tag{33}$$
$$(A \cap B) \cap C = A \cap (B \cap C) \tag{34}$$

- **Distributivity:**

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \tag{35}$$
$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \tag{36}$$

- **Identity:**

$$A \cup \emptyset = A \tag{37}$$
$$A \cap \Omega = A \tag{38}$$

- **Complement:**

$$A \cup A^c = \Omega \tag{39}$$
$$A \cap A^c = \emptyset \tag{40}$$

Probability Laws

A probability law must satisfy three axioms:

1. Non-negativity: $P(A) \geq 0 \forall A \in F$
2. Normalization: $P(\Omega) = 1$
3. Additivity: For any disjoint subsets $\{A_1, A_2, \dots\}$, it holds that

$$P\left[\bigcup_{n=1}^{\infty} A_n\right] = \sum_{n=1}^{\infty} P\left[A_n\right]$$

Probability Properties

$$P[A \cup B] = P[A] + P[B] - P[A \cap B] \tag{41}$$

$$P[A \cup B] \leq P[A] + P[B] \tag{42}$$

$$A \subseteq B \implies P[A] \leq P[B] \tag{43}$$

Formal Definitions

Outcomes

An *outcome* is the result of some *experiment*. If that experiment is flipping a coin, the outcome is either heads or tails. We could express the outcome of heads as H, and the outcome of tails as T. The set of all possible outcomes for an experiment is known as a sample space and is denoted by Ω. In this case $\Omega = \{H, T\}$.

Events

An *event* F is a subset of the sample space Ω. The formal definitions of probability are expressed with set notation. So the event where we have neither heads nor tails is written as $\{\}$. The event of heads could be expressed as $\{H\}$, and the event of tails could be expressed as $\{T\}$. The event of either heads or tails is $\{H, T\}$.

Probability Laws

A *probability law* is a function P that maps an event A to a real number in $[0, 1]$. For the coin example, the probability law might be $P(\{\}) = 0$, $P(\{H\}) = 0.5$, $P(\{T\}) = 0.5$, and $P(\{\Omega\}) = 1$. A probability law must satisfy three axioms:

1. Non-negativity: $P(A) \geq 0 \forall A \in F$
2. Normalization: $P(\Omega) = 1$
3. Additivity: For any disjoint subsets $\{A_1, A_2, \dots\}$, it holds that

$$P\left[\bigcup_{n=1}^{\infty} A_n\right] = \sum_{n=1}^{\infty} P\left[A_n\right]$$

Probability Space

A probability space is a triplet Ω, F, P.

Probability Properties

$$P\left[A \cup B\right] = P\left[A\right] + P\left[B\right] - P\left[A \cap B\right] \tag{44}$$

$$P\left[A \cup B\right] \leq P\left[A\right] + P\left[B\right] \tag{45}$$

$$A \subseteq B \implies P\left[A\right] \leq P\left[B\right] \tag{46}$$

$$P\left[A|B\right] = \frac{P[A \cap B]}{P[B]} \tag{47}$$

Outcomes are statistically *independent* if $P(A|B) = P(A)$ (assuming P(B) > 0), or equivalently $P(A \cap B) = P(A)P(B)$.

Bayes Theorem states that for any two events A and B such that $P[A] > 0$ and $P[B] > 0$,

$$P[A|B] = \frac{P[B|A]P[A]}{P[B]} \tag{48}$$

The *Law of Total Probability* states that if $\{A_1, A_2, \ldots, A_n\}$ is a partition of Ω, then for any $B \subseteq \Omega$,

$$P[B] = \sum_{i=1}^{n} P[B|A_i]P[A_i] \tag{49}$$

Discrete Random Variables

A *random variable* X is a function $X : \Omega \implies \Re$ that maps an outcome $\epsilon \in \Omega$ to a number $X(\epsilon)$ on the real line. We call it a variable because it has multiple states.

The *expectation* of a random variable X is

$$E[X] = \sum_{x \in X(\Omega)} x p_X(x) \tag{50}$$

The difference between $E[X]$ and the mean is that $E[X]$ is computed from the ideal histogram, while mean is computed from the empirical histogram. In general for any functions g and h,

$$E[g(X)] = \sum_x g(x) p_X(x) \tag{51}$$

$$E[g(X) + h(X)] = E[g(X)] + E[h(X)] \tag{52}$$

$$E[cX] = cE[X] \tag{53}$$

$$E[X + c] = E[X] + c \tag{54}$$

The *variance* of a random variable X is

$$Var[X] = E\left[(X - \mu)^2\right] \tag{55}$$

or alternatively, the second moment minus the first moment squared.

$$E[X^2] - E[X]^2 \tag{56}$$

The *probability mass function* (PMF) $p_X(a)$ of a random variable X specifies the probability of obtaining a number $X(\epsilon) = a$. We denote a PMF as

$$p_X(a) = P[X = a] \tag{57}$$

PMFs are represented with histograms. A PMF should satisfy

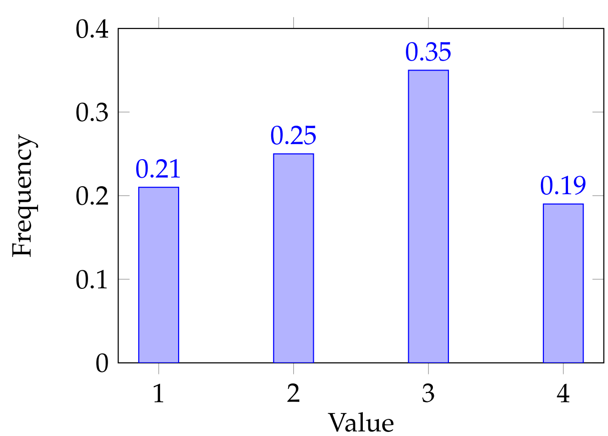

Figure 1: PMF

$$\sum_{x \in X(\Omega)} p_X(x) = 1 \tag{58}$$

The *cumulative distribution function* is given by

$$F_X(x) = P\left[X \leq x\right] \tag{59}$$
$$= \sum_{u \leq x} p_X(u) \tag{60}$$

and represents the sum of every impulse of the PMF up to x.

A *Bernoulli random variable* has a state of either 0 or 1. The probability of getting 1 is p and the probability of getting 0 is $1 - p$. We write

$$X \sim Bernoulli(p) \tag{61}$$

or

$$X \sim B(p) \tag{62}$$

to say that X is drawn from a Bernoulli distribution with a parameter p. For a Bernoulli distribution,

$$E[X] = p \tag{63}$$
$$E[X^2] = p \tag{64}$$
$$Var[X] = p(1-p) \tag{65}$$

Say $S \sim B(1-p)$. Let

$$P(R = 0|S = 0) = 1 - \epsilon_0 \tag{66}$$
$$P(R = 1|S = 0) = \epsilon_0 \tag{67}$$

then $R|S = 0 \sim B(\epsilon_0)$. Let

$$P(R = 0|S = 1) = \epsilon_1 \tag{68}$$
$$P(R = 1|S = 0) = 1 - \epsilon_1 \tag{69}$$

then $R|S = 0 \sim B(1 - \epsilon_1)$. Overall,

$$R|S \sim B(\epsilon_0^{1-S}(1-\epsilon_1)^S) \tag{70}$$

A *Rademacher random variable* has two states, -1 and 1. The probability of getting each is 0.5.

A *binomial random variable* has a PMF of

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, k = 0, 1, \ldots n \tag{71}$$

where $0 < p < 1$ is the binomial parameter, and n is the total number of states. We write

$$X \sim Binomial(n, p) \tag{72}$$

to say that X is drawn from a binomial distribution with a parameter p of size n. If $X \sim Binomial(n, p)$, then

$$E[X] = np \tag{73}$$
$$E[X^2] = np(np + (1-p)) \tag{74}$$
$$Var[X] = np(1-p) \tag{75}$$

Let X be a *geometric random variable*. Then the PMF of X is

$$p_X(k) = (1-p)^{k-1}p, k = 1, 2, \ldots \tag{76}$$

We write

$$X \sim Geometric(p) \tag{77}$$

to say that X was drawn from a geometric distribution with a parameter p. If $X \sim Geometric(p)$ then

$$E[X] = \frac{1}{p} \tag{78}$$
$$E[X^2] = \frac{2}{p^2} - \frac{1}{p} \tag{79}$$
$$Var[X] = \frac{1-p}{p^2} \tag{80}$$

Let X be a *Poisson random variable*. Then the PMF of X is

$$p_X(k) = \frac{\lambda^k}{k!}e^{-\lambda}, k = 0, 1, 2, \ldots \tag{81}$$

where $\lambda > 0$ is the Poisson rate. We write $X \sim Poisson(\lambda)$ to say that X was drawn from a Poisson distribution with a parameter λ. If $X \sim Poisson(\lambda)$ then

$$E[X] = \lambda \tag{82}$$
$$E[X^2] = \lambda + \lambda^2 \tag{83}$$
$$Var[X] = \lambda \tag{84}$$

For small p and large n,

$$\binom{n}{k} p^k (1-p)^{n-k} \approx \frac{\lambda^k}{k!}e^{-\lambda} \tag{85}$$

Joint Distributions are higher-dimensional PDFs, PMFs, or CDFs. We write

$$f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) \equiv f_{\vec{X}}(\vec{x}) \tag{86}$$

The *joint PMF* of two random variables X and Y is notated by

$$p_{X,Y}(x, y) = P[X = x \text{ and } Y = y] \tag{87}$$

and represents the probability of both.

A *marginal PMF* is defined as

$$p_X(x) = \sum_{y \in \Omega_Y} p_X(x, y) \tag{88}$$

or w.l.o.g.

$$p_Y(y) = \sum_{x \in \Omega_X} p_X(x, y) \tag{89}$$

That is, it is the joint PMF summed over one of the variables.

The *conditional PMF* is given by

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x, y)}{p_Y(y)} \tag{90}$$

If two random variables X and Y are independent, then

$$p_{X,Y} = p_X(x)p_Y(y) \tag{91}$$

$$f_{X,Y} = f_X(x)f_Y(y) \tag{92}$$

If a sequence of random variables $X_1, X_2, \ldots, X_N$ are independent, then their joint PDF (or joint PMF) can be factorized as

$$f_{X_1,X_2,\ldots,X_N}(x_1, x_2, \ldots, x_N) = \prod_{n=1}^{N} f_{X_n}(x_n) \tag{93}$$

The *joint CDF* of two random variables X and Y is the function $F_{X,Y}(x, y)$ such that

$$F_{X,Y}(x, y) = P[X \leq x \cap Y \leq y] \tag{94}$$

If X and Y are discrete, then

$$F_{X,Y}(x, y) = \sum_{y' \leq y} \sum_{x' \leq x} p_{X,Y}(x', y') \tag{95}$$

For two random variables X and Y, the *marginal CDF* is

$$F_X(x) = F_{X,Y}(x, \infty) \tag{96}$$

$$F_Y(y) = F_{X,Y}(\infty, y) \tag{97}$$

Let X and Y be two random variables. The *joint expectation* is

$$E[XY] = \sum_{y \in \Omega_Y} \sum_{x \in \Omega_X} xy \times p_{X,Y}(x, y) \tag{98}$$

If X and Y are discrete, then joint expectation is also called *correlation*. This can be written in matrix form as

$$\begin{bmatrix} p_{X,Y}(x_1, y_1) & p_{X,Y}(x_1, y_2) & \ldots & p_{X,Y}(x_1, y_N) \\ p_{X,Y}(x_2, y_1) & p_{X,Y}(x_2, y_2) & \ldots & p_{X,Y}(x_2, y_N) \\ \vdots & \vdots & \ddots & \vdots \\ p_{X,Y}(x_N, y_1) & p_{X,Y}(x_N, y_2) & \ldots & p_{X,Y}(x_N, y_N) \end{bmatrix} \tag{99}$$

then the joint expectation is

$$E[XY] = \sum_{i=1}^{N} \sum_{j=1}^{N} x_i y_j \times p_{X,Y}(x_i, y_j) \tag{100}$$

Let the matrix in Equation 99 be $\mathbf{P}$. Let

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \tag{101}$$

$$\vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \tag{102}$$

then

$$E[XY] = \begin{bmatrix} x_1 & x_2 & \dots & x_N \end{bmatrix} \begin{bmatrix} p_{X,Y}(x_1, y_1) & p_{X,Y}(x_1, y_2) & \dots & p_{X,Y}(x_1, y_N) \\ p_{X,Y}(x_2, y_1) & p_{X,Y}(x_2, y_2) & \dots & p_{X,Y}(x_2, y_N) \\ \vdots & \vdots & \ddots & \vdots \\ p_{X,Y}(x_N, y_1) & p_{X,Y}(x_N, y_2) & \dots & p_{X,Y}(x_N, y_N) \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \tag{103}$$

$$= \vec{x}^T \mathbf{P} \vec{y} \tag{104}$$

$E[XY]$ is a weighted inner product between the states. $\vec{x}$ and $\vec{y}$ are the states of the random variables X and Y. Recalling that the magnitude of the inner product of $\vec{a}$ and $\vec{b}$ is $|a||b|\cos(\theta)$ and that cosine is bounded, we have

$$-1 \leq \frac{E[XY]}{\sqrt{E[X^2]}\sqrt{E[Y^2]}} \leq 1 \tag{105}$$

Notice that the correlation of X, Y is proportional to the covariance.

Continuous Random Variables

A *continuous random variable* is analogous to the discrete case. Recall that a probability is just a size of a set. It's easy to find the size of a discrete set because you can just count elements, but for an uncountable set new methods are needed. Luckily the intution for continuous random variables is intuitive, it's still just the size of a set A relative to Ω. Formally, if each event in A is equally likely, then

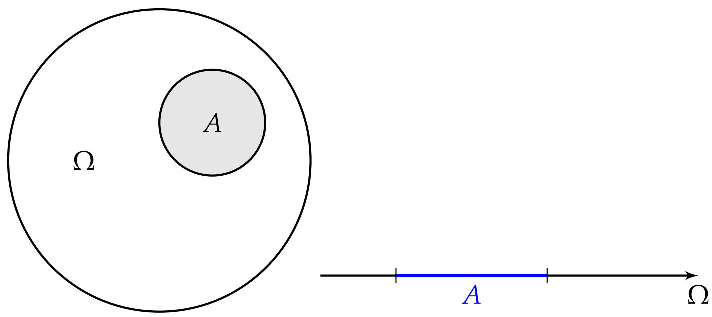

Figure 2: Continuous random variables

$$P[\{x \in A\}] = \frac{\int_A dx}{|\Omega|} \tag{106}$$

If we relax the assumption of equiprobability, then more generally

$$P[\{x \in A\}] = \int_A f_X(x)dx \tag{107}$$

$f_X(x)$ is called the *probability density function* (PDF). It is analogous to the probability mass function.

Formally, a probability density function is a mapping $f_X : \Omega \implies \Re$, with the following properties:

- Non-negativity: $f_X(x) \geq 0 \forall x \in \Omega$
- Unity: $\int_\Omega f_X(x)dx = 1$
- Measure of a set: $P[\{x \in A\}] = \int_A f_X(x)dx$

We can express a PDF in terms of a PMF with a train of delta functions like so:

$$f_X(x) = \sum_{x_k \in \Omega} p_X(x_k)\delta(x - x_k) \tag{108}$$

We can also define the probability density function as the derivative of the CDF, like so:

$$f_X(x) = \frac{d}{dx} p(X \leq x) \tag{109}$$

The expectation of a continuous random variable is

$$E[X] = \int_\Omega x f_X(x) dx \tag{110}$$

Properties of the expectation for continuous random variables:

- $E[aX] = aE[X]$
- $E[X + a] = E[X] + a$
- $E[aX + b] = aE[X] + b$

A random variable X has an expectation if it is absolutely integrable,

$$E[|X|] = \int_\Omega |x| f_X(x) dx < \infty \tag{111}$$

The variance of a continuous random variable X is

$$\begin{aligned} Var[X] &= E[(X - \mu)^2] &(112)\\ &= \int_\Omega (x - \mu)^2 f_X(x) dx &(113)\\ &= E[X^2] - \mu^2 &(114) \end{aligned}$$

A continuous *uniform random variable* has a PDF of

$$f_X(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{else} \end{cases} \tag{115}$$

We write

$$X \sim Uniform(a, b) \tag{116}$$

to mean that X is drawn from a uniform distribution on an interval $[a, b]$. It has a CDF given by

$$F_X(x) = \begin{cases} 0 & a < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases} \tag{117}$$

If $X \sim Uniform(a, b)$ then

$$\begin{aligned} E[X] &= \frac{a+b}{2} &(118)\\ Var[X] &= \frac{(b-a)^2}{12} &(119) \end{aligned}$$

A continuous *exponential random variable* has a PDF of

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{else} \end{cases} \tag{120}$$

An exponential random variable is the interarrival time between two consecutive Poisson events

We write

$$X \sim Exponential(\lambda) \tag{121}$$

to mean that X is drawn from an exponential distribution of parameter λ. It has a CDF given by

$$F_X(x) = 1 - e^{-\lambda x} \tag{122}$$

If $X \sim Exponential(\lambda)$, then

$$E[X] = \frac{1}{\lambda} \tag{123}$$

$$Var[X] = \frac{1}{\lambda^2} \tag{124}$$

A *Gaussian random variable* is a random variable X such that its PDF is

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{125}$$

We write

$$X \sim Gaussian(\mu, \sigma^2) \tag{126}$$

or

$$X \sim \mathcal{N}\left(\mu, \sigma^2\right) \tag{127}$$

to mean that X is drawn from a Gaussian of parameter (μ, σ^2). If $X \sim \mathcal{N}(\mu, \sigma^2)$, then

$$E[X] = \mu \tag{128}$$

$$Var[X] = \sigma^2 \tag{129}$$

The *standard Gaussian* random variable has a PDF given by

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \tag{130}$$

The CDF of the standard Gaussian is defined as the Φ function.

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt \tag{131}$$

The CDF of the standard Gaussian is related to the *error function*, which is defined as

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \tag{132}$$

by the relation

$$\Phi(x) = \frac{1}{2}\left(1 + \text{erf}\left(\frac{x}{\sqrt{2}}\right)\right) \tag{133}$$

The CDF of an arbitrary Gaussian is related via the transformation

$$F_X(x) = \phi\left(\frac{x-\mu}{\sigma}\right) \tag{134}$$

In addition to mean and variance, we introduce two more useful quantities, *skewness* and *kurtosis*.

$$E[X] = \mu \tag{135}$$

$$E\left[(X-\mu)^2\right] = \sigma^2 \tag{136}$$

$$E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \gamma \tag{137}$$

$$E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] = \kappa \tag{138}$$

Excess kurtosis is defined as $\kappa - 3$

Skewness measures the asymmetry of a distribution. A Gaussian distribution has skewness 0. Kurtosis measures how heavy-tailed the distribution is. If the kurtosis is positive, then the tails decay faster than a Gaussian. If the kurtosis is negative, then the distribution has a tail that decays more slowly than a Gaussian.

Reference

Series

$$\sum_{k=0}^{n} r^k = \frac{1-r^{n+1}}{1-r} \tag{1}$$

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6} \tag{2}$$

$$\sum_{k=1}^{\infty} k r^{k-1} = \frac{1}{(1-r)^2} \tag{3}$$

Combinatorics

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \tag{4}$$

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k \tag{5}$$

$$\binom{n}{k} + \binom{n}{k-1} = \binom{n+1}{k} \tag{6}$$

$$P(n,k) = \frac{n!}{(n-k)!} \tag{7}$$

where $P(n,k)$ is the number of ways to arrange k objects out of n (permutations).

$$C(n,k) = \binom{n}{k} = \frac{n!}{k!(n-k)!} \tag{8}$$

where $C(n,k)$ is the number of ways to choose k objects out of n (combinations).

Approximations

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \ldots \tag{9}$$

$$= \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n \tag{10}$$

$$1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots = \sum_{k=0}^{\infty} \frac{x^k}{k!} \tag{11}$$

$$= e^x \tag{12}$$

$$\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \ldots \tag{13}$$

$$= \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!} \tag{14}$$

$$\cos(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \dots \tag{15}$$
$$= \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!} \tag{16}$$

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots \tag{17}$$
$$= \sum_{n=1}^{\infty} (-1)^{n+1} \frac{x^n}{n} \tag{18}$$

Calculus

$$\frac{d}{dx} \int_a^x f(t)\, dt = f(x) \tag{19}$$

$$\int_a^b f'(x)\, dx = f(b) - f(a) \tag{20}$$

$$\int f(g(x))g'(x)\, dx = \int f(u)\, du \tag{21}$$

$$\int u\, dv = uv - \int v\, du \tag{22}$$

$$\int \frac{1}{(x-a)(x-b)}\, dx = \frac{1}{b-a} \ln \left| \frac{x-a}{x-b} \right| + C \tag{23}$$

Linear Algebra

$$\vec{y} = \beta_1 \vec{x}_1 + \beta_2 \vec{x}_2 + \dots + \beta_N \vec{x}_N \tag{24}$$

$$\langle \vec{a}, \vec{b} \rangle = \vec{a}\vec{b}^T \tag{25}$$
$$= \sum_{i=1}^{n} a_i b_i \tag{26}$$

where $\langle \vec{a}, \vec{b} \rangle$ denotes the inner product of vectors $\vec{a}$ and $\vec{b}$.

$$\|\vec{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p} \tag{27}$$

where $\|\vec{x}\|_p$ is the p-norm (or ℓ_p-norm) of vector $\vec{x}$.

$$\cos(\theta) = \frac{\langle \vec{a}, \vec{b} \rangle}{\|\vec{a}\|_2 \|\vec{b}\|_2} \tag{28}$$

where θ is the angle between vectors $\vec{a}$ and $\vec{b}$.

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\vec{y} \tag{29}$$

where $\hat{\beta}$ is the vector of least squares coefficients, $\mathbf{X}$ is the data matrix, and $\vec{y}$ is the target vector