Preliminaries

This section revises some parts of Core A Probability, which are essential for this course, and lists some other mathematical facts to be used (without proof) in the following.

Probability space

We recall that a *sample space* Ω is a collection of all possible outcomes of a probabilistic experiment; an *event* is a collection of possible outcomes, ie., a subset of the sample space. We introduce the *impossible* event $\varnothing$ and the *certain* event Ω; also, if $A \subset \Omega$ and $B \subset \Omega$ are events, it is natural to consider other events such that $A \cup B$ (**A or B**), $A \cap B$ (**A and B**), $A^{\mathsf{c}} \equiv \Omega \setminus A$ (**not A**), and $A \setminus B$ (**A but not B**).

Definition 0.1. Let $\mathcal{A}$ be a collection of subsets of Ω. We shall call $\mathcal{A}$ a field if it has the following properties:

1. $\varnothing \in \mathcal{A}$;
2. if A_1, $A_2 \in \mathcal{A}$, then $A_1 \cup A_2 \in \mathcal{A}$;
3. if $A \in \mathcal{A}$, then $A^{\mathsf{c}} \in \mathcal{A}$.

Remark 0.1.1. *Obviously, every field is closed w.r.t. taking finite unions or intersections.*

Definition 0.2. Let $\mathcal{F}$ be a collection of subsets of Ω. We shall call $\mathcal{F}$ a σ-field if it has the following properties:

1. $\varnothing \in \mathcal{F}$;
2. if A_1, $A_2, \dots \in \mathcal{F}$, then $\bigcup_{k=1}^{\infty} A_k \in \mathcal{F}$;
3. if $A \in \mathcal{F}$, then $A^{\mathsf{c}} \in \mathcal{F}$.

Remark 0.2.1. *Obviously, property 2 above can be replaced by the equivalent condition* $\bigcap_{k=1}^{\infty} A_k \in \mathcal{F}$.

Clearly, if Ω is fixed, the smallest σ-field in Ω is just $\{\varnothing, \Omega\}$ and the biggest σ-field consists of all subsets of Ω. We observe the following simple fact:

Exercise 0.3. *Show that if $\mathcal{F}_1$ and $\mathcal{F}_2$ are σ-fields, then*[1] *$\mathcal{F}_1 \cap \mathcal{F}_2$ is a σ-field, but, in general, $\mathcal{F}_1 \cup \mathcal{F}_2$ is not a σ-field.*

If A and B are events, we say that A and B are incompatible (or disjoint), if $A \cap B = \varnothing$.

Definition 0.4. Let Ω be a sample space, and $\mathcal{F}$ be a σ-field of events in Ω. A probability distribution P on $(\Omega, \mathcal{F})$ is a collection of numbers $\mathsf{P}(A)$, $A \in \mathcal{F}$, possessing the following properties:

[1] and, in fact, an intersection of arbitrary (even **uncountable**!) collection of σ-fields;

A1 for every event $A \in \mathcal{F}$, $\mathsf{P}(A) \geq 0$;

A2 $\mathsf{P}(\Omega) = 1$;

A3 for any pair of incompatible events A and B, $\mathsf{P}(A \cup B) = \mathsf{P}(A) + \mathsf{P}(B)$;

A4 for any *countable* collection A_1, A_2, ... of mutually incompatible[2] events,

$$\mathsf{P}\Big(\bigcup_{k=1}^{\infty} A_k\Big) = \sum_{k=1}^{\infty} \mathsf{P}(A_k).$$

Remark 0.4.1. *Notice that the additivity axiom* **A4** *above does not extend to* uncountable *collections of incompatible events.*

Remark 0.4.2. *Obviously, property* **A4** *above and Definition 0.2 are non-trivial only in examples with infinitely many different events, ie., when the collection $\mathcal{F}$ of all events (and, therefore, the sample space Ω) is infinite.*

The following properties are immediate from the above axioms:

P1 for any pair of events A, B in Ω we have

$$\mathsf{P}(B \setminus A) = \mathsf{P}(B) - \mathsf{P}(A \cap B)\,, \qquad \mathsf{P}(A \cup B) = \mathsf{P}(A) + \mathsf{P}(B \setminus A);$$

in particular, $\mathsf{P}(A^{\mathsf{c}}) = 1 - \mathsf{P}(A)$;

P2 if events A, B in Ω are such that $\varnothing \subseteq A \subseteq B \subseteq \Omega$, then

$$0 = \mathsf{P}(\varnothing) \leq \mathsf{P}(A) \leq \mathsf{P}(B) \leq \mathsf{P}(\Omega) = 1\,.$$

P3 if $A_1, A_2, \ldots, A_n$ are events in Ω, then $\mathsf{P}\big(\cup_{k=1}^{n} A_k\big) \leq \sum_{k=1}^{n} \mathsf{P}(A_k)$ with the inequality becoming an equality if these events are mutually incompatible;

Definition 0.5. A *probability space* is a triple $(\Omega, \mathcal{F}, \mathsf{P})$, where Ω is a sample space, $\mathcal{F}$ is a σ-field of events in Ω, and $\mathsf{P}(\,\cdot\,)$ is a probability measure on $(\Omega, \mathcal{F})$.

In what follows we shall always assume that some probability space $(\Omega, \mathcal{F}, \mathsf{P})$ is fixed.

Conditional probability, independence

Definition 0.6. The *conditional probability* of event A given event B such that $\mathsf{P}(B) > 0$, is

$$\mathsf{P}(A \,|\, B) \stackrel{\mathsf{def}}{=} \frac{\mathsf{P}(A \cap B)}{\mathsf{P}(B)}\,.$$

It is easy to see that if $E \in \mathcal{F}$ is any event with $\mathsf{P}(E) > 0$, then $\mathsf{P}(\,\cdot\,|E)$ is a probability measure on $(\Omega, \mathcal{F})$, ie., axioms **A1**–**A4** and properties **P1**–**P3** hold (just $\mathsf{P}(\,\cdot\,)$ replace with $\mathsf{P}(\,\cdot\,|E)$). We list some additional useful properties of conditional probabilities:

[2] ie., $A_k \cap A_j = \varnothing$ for all $k \neq j$;

P4 *multiplication rule* for probabilities: if A and B are events, then

$$\mathsf{P}(A\cap B)=\mathsf{P}(A)\,\mathsf{P}(B\,|\,A)=\mathsf{P}(B)\,\mathsf{P}(A\,|\,B)\,;$$

more generally, if A_1, ..., A_n are arbitrary events in $\mathcal{F}$, then

$$\mathsf{P}\Bigl(\bigcap_{k=1}^{n}A_k\Bigr)=\mathsf{P}(A_1)\cdot\prod_{k=2}^{n}\mathsf{P}\Bigl(A_k\,|\,\bigcap_{j=1}^{k-1}A_j\Bigr)\,;\tag{0.1}$$

for example, $\mathsf{P}(A\cap B\cap C)=\mathsf{P}(A)\,\mathsf{P}(B\,|\,A)\,\mathsf{P}(C\,|\,A\cap B)$.

P5 *partition theorem* or *formula of total probability*: we say that events B_1, ..., B_n form a partition of Ω if they are mutually incompatible (disjoint) and their union $\cup_{k=1}^{n}B_k$ is the entire space Ω. The partition theorem says that if B_1, ..., B_n form a partition of Ω, then for any event A we have

$$\mathsf{P}(A)=\sum_{k=1}^{n}\mathsf{P}(B_k)\cdot\mathsf{P}(A\,|\,B_k)\,.\tag{0.2}$$

P6 *Bayes' theorem*: for any events A, B, we have

$$\mathsf{P}(A\,|\,B)=\frac{\mathsf{P}(A)\,\mathsf{P}(B\,|\,A)}{\mathsf{P}(B)}\,;$$

in particular, if D is an event and C_1, ..., C_n form a partition of Ω, then

$$\mathsf{P}(C_k\,|\,D)=\frac{\mathsf{P}(C_k)\,\mathsf{P}(D\,|\,C_k)}{\sum_{k=1}^{n}\mathsf{P}(C_k)\,\mathsf{P}(D\,|\,C_k)}\,;\tag{0.3}$$

Exercise 0.7. *Check carefully (ie., by induction) property* **P4** *above.*

Then next definition is one of the most important in probability theory.

Definition 0.8. We say that events A and B are *independent* if

$$\mathsf{P}\bigl(A\cap B\bigr)=\mathsf{P}(A)\,\mathsf{P}(B)\,;\tag{0.4}$$

under (0.4), we have $\mathsf{P}(A\,|\,B)=\mathsf{P}(A)$, ie., event A is *independent of* B; similarly, $\mathsf{P}(B\,|\,A)=\mathsf{P}(B)$, ie., event B is *independent of* A.

More generally,

Definition 0.9. A collection of events A_1, ..., A_n is called (*mutually*) independent, if

$$\mathsf{P}\Bigl(\bigcap_{k=1}^{n}A_k\Bigr)=\prod_{k=1}^{n}\mathsf{P}(A_k)\,.\tag{0.5}$$

It is immediate from (0.5) that every sub-collection of $\bigl\{A_1,\dots,A_n\bigr\}$ is also mutually independent.

Random variables

It is very common for the sample space Ω of possible outcomes to be a set of *real numbers*. Then the outcome to the "probabilistic experiment" is often called a *random variable* and denoted by a capital letter such as X. In this case the events are subsets $A \subseteq \mathbb{R}$ and it is usual to write $\mathsf{P}(X \in A)$ instead of $\mathsf{P}(A)$ and similarly $\mathsf{P}(X = 1)$ for $\mathsf{P}(\{1\})$, $\mathsf{P}(1 < X < 5)$ for $\mathsf{P}(A)$ where $A = (1, 5)$ and so on. The *probability distribution* of a r.v. X is the collection of probabilities $\mathsf{P}(X \in A)$ for all intervals $A \subseteq \mathbb{R}$ (and other events that can be obtained from intervals via axioms **A1**–**A4**).

Let X be a random variable (so the sample space Ω is a subset of $\mathbb{R}$). We say that X is a *discrete r.v.* if in addition Ω is countable, i.e., if the possible values for X can be enumerated in a (possibly infinite) list. In this case the function $p(x) \stackrel{\mathsf{def}}{=} \mathsf{P}(X = x)$ (defined for all real x) is called the *probability mass function* of X and the corresponding probability distribution of X is defined via

$$\mathsf{P}(X \in A) = \sum_{x \in A} \mathsf{P}(X = x) = \sum_{x \in A} p(x)\,.$$

If X takes possible values x_1, x_2, ..., then, by axiom **A3**, $\sum_{k \geq 1} p(x_k) = 1$ and if x is NOT one of the possible values of X then $p(x) = 0$.

Similarly, a random variable X has a *continuous probability distribution* if there exists a non-negative function $f(x)$ on $\mathbb{R}$ such that for any interval $(a, b) \subseteq \mathbb{R}$

$$\mathsf{P}\big(a < X < b\big) = \int_a^b f(x)\,dx\,;$$

in particular, by axiom **A3**, we must have $\int_{-\infty}^{\infty} f(x)\,dx = 1$. The function $f(\,\cdot\,)$ is then called the *probability density function* (or *pdf*) of X.

In Core A Probability you saw a number of random variables with discrete (Bernoulli, binomial, geometric, Poisson) or continuous (uniform, exponential, normal) distribution.

Definition 0.10. For any random variable X, the *cumulative distribution function* (or cdf) of X is the function $F : \mathbb{R} \to [0, 1]$ that is given at all $x \in \mathbb{R}$ by

$$F(x) \stackrel{\mathsf{def}}{=} \mathsf{P}(X \leq x) = \begin{cases} \int_{-\infty}^{x} f(y)\,dy\,, & X \text{ a continuous r.v.;} \\ \sum_{x_k : x_k \leq x} p(x_k)\,, & X \text{ a discrete r.v.;} \end{cases} \tag{0.6}$$

If, in addition, $f(x)$ is continuous function on some interval (a, b) then by the fundamental theorem of calculus, for all $x \in (a, b)$, $F'(x) = f(x)$; ie., the cdf determines the pdf and vice versa. In fact, the cdf of a r.v. X always determines its probability distribution.

Remark 0.10.1. *Suppose X is a random variable and h is some real-valued function defined for all real numbers. Then $h(X)$ is also a random variable, namely, the outcome to a new "experiment" obtained by running the old "experiment" to produce the r.v. X and then evaluating $h(X)$.*

Joint distributions

It is essential for most useful applications of probability to have a theory which can handle many random variables simultaneously.

Definition 0.11. Let $(X_1, \dots, X_n)$ be a *multivariate* random variable (or *random vector*). Its cumulative distribution function is

$$F_{X_1,\dots,X_n}(x_1, \dots, x_n) \stackrel{\text{def}}{=} \mathsf{P}\big(X_1 \le x_1, \dots, X_n \le x_n\big)\,, \tag{0.7}$$

here and below we write $\big\{X_1 \le x_1, \dots, X_n \le x_n\big\} = \{X_1 \le x_1\} \cap \dots \cap \{X_n \le x_n\}$.

Bivariate variables: discrete case

Suppose (X, Y) is a bivariate r.v. and that X and Y are discrete r.v. taking possible values x_1, x_2, ... and y_1, y_2,... respectively. Then the collection of probabilities

$$p(x_j, y_k) \equiv \mathsf{P}(X = x_j, Y = y_k)\,, \qquad k \ge 1\,, j \ge 1\,,$$

determines the *joint probability distribution* of (X, Y). It is important to remember that given the joint distribution of X, Y we can recover the probability density function p_X (in this case it is called the *marginal probability distribution*) of X via

$$p_X(x_j) \equiv \mathsf{P}(X = x_j) = \sum_k \mathsf{P}(X = x_j, Y = y_k) = \sum_k p(x_j, y_k) \tag{0.8}$$

for any possible value x_j of X. Similarly, the marginal probability distribution of Y is given by

$$p_Y(y_k) = \sum_j \mathsf{P}(X = x_j, Y = y_k) = \sum_j p(x_j, y_k)\,.$$

Conditional distribution and independence

For any discrete bivariate rv (X, Y) the *conditional distribution* of X given Y has probability mass function

$$p(x \,|\, y) \equiv \mathsf{P}(X = x \,|\, Y = y) = \frac{p(x, y)}{p_Y(y)}$$

for all y with $p_Y(y) > 0$. There is also a r.v. version of the partition theorem (0.2); it is often called the *law of total probability*: for any X-event A,

$$\mathsf{P}(X \in A) = \sum_y \mathsf{P}\big(X \in A \,|\, Y = y\big)\, p_Y(y)\,. \tag{0.9}$$

We say that X and Y are *independent* if for all x, y

$$p(x, y) = p_X(x)\, p_Y(y)\,. \tag{0.10}$$

Alternatively, we have

Definition 0.12. Random variables X, Y are *independent* if for *every* X-event A and *every* Y-event B we have

$$\mathsf{P}\big(X \in A, Y \in B\big) \equiv \mathsf{P}\big((X,Y) \in A \times B\big) = \mathsf{P}(X \in A)\,\mathsf{P}(Y \in B)\,. \qquad (0.11)$$

The definitions (0.10), (0.11) can be easily extended to the case of any general multivariate distribution.

Let $(X_1, \dots, X_n)$ be a random vector and $g : \mathbb{R}^n \to \mathbb{R}$ be a function. Then $g(X_1, \dots, X_n)$ is a random variable (obtained by the new "experiment" consisting of first carrying out the original experiment to determine the value of $(X_1, \dots, X_n)$ and then applying the function g to this ordered n-tuple to obtain a real number $g(X_1, \dots, X_n)$).

Exercise 0.13. *1). Let (X, Y, Z) be a random vector with independent components; show that for any function $h : \mathbb{R}^2 \to \mathbb{R}$ the variables $h(X, Y)$ and Z are independent.*

2). Let X_1, ..., X_k and Y_1, ..., Y_m be a collection of independent random variables. If the functions f and g are such that $f : \mathbb{R}^k \to \mathbb{R}$ and $g : \mathbb{R}^m \to \mathbb{R}$, show that the random variables $f(X_1, \dots, X_m)$ and $g(Y_1, \dots, Y_m)$ are independent.

Bivariate variables: continuous case

We will only consider the case where (X, Y) has a continuous joint pdf $f(x, y)$ defined for $(x, y) \in \mathbb{R}^2$. By analogy with the definition for discrete random variables,

$$\mathsf{P}\big((X,Y) \in A\big) = \iint_A f(x,y)\,dx\,dy$$

for any integrable set A. In this case X and Y have the *marginal* pdfs

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy\,, \qquad f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx$$

and for any interval (a, b) we have

$$\mathsf{P}(a < X < b) \equiv \int_a^b \int_{-\infty}^{\infty} f(x,y)\,dx\,dy = \int_a^b f_X(x)\,dx\,.$$

We define the continuous conditional density of X given Y by

$$f(x \,|\, y) = \begin{cases} f(x,y)/f_Y(y)\,, & \text{if } f_Y(y) > 0 \\ 0\,, & \text{if } f_Y(y) = 0\,. \end{cases}$$

Also, X and Y are *independent* if and only if $f(x, y) = f_X(x)\, f_Y(y)$ for every pair $(x, y) \in \mathbb{R}^2$.

Transformations $g(X, Y)$ in the continuous case are treated similarly to the discrete case.

Expectation

Definition 0.14. For any random variable X the *expected value* (or *mean*) of X is the number

$$\mathsf{E}(X) = \begin{cases} \sum_{x_k\in\Omega} x_k\, p(x_k)\,, & X \text{ discrete with pmf } p\,; \\ \int_{-\infty}^{\infty} x f(x)\, dx\,, & X \text{ continuous with pdf } f\,. \end{cases} \tag{0.12}$$

The following generalisation of this definition is of great importance to the whole theory.

If X is a discrete rv and takes values in $\Omega = \{x_1, x_2, \dots\}$ with probabilities $p(x_k)$ and the transformed rv $g(X)$ takes values y_1, y_2, ... with probabilities

$$q(y_m) \stackrel{\mathsf{def}}{=} \mathsf{P}(X \in G_m) = \sum_{x\in G_m} p(x)\,, \quad \text{where} \quad G_m \stackrel{\mathsf{def}}{=} \bigl\{x \in \Omega : g(x) = y_m\bigr\}\,,$$

then the sets G_m form a partition of Ω and it follows that

$$\mathsf{E}\bigl(g(X)\bigr) = \sum_m y_m q(y_m) = \sum_m \sum_{x\in G_m} g(x)\, p(x) = \sum_{k=1}^{\infty} g(x_k)\, p(x_k)\,.$$

Similarly, if X is continuous rv with pdf f, then

$$\mathsf{E}\bigl(g(X)\bigr) = \int_{-\infty}^{\infty} g(x)\, f(x)\, dx\,.$$

The most important properties of the expectation are:

E1 linearity: let f, g be real functions and let a, b be real numbers; then

$$\mathsf{E}\bigl(af(X) + bg(X)\bigr) = a\,\mathsf{E}\bigl(f(X)\bigr) + b\,\mathsf{E}\bigl(g(X)\bigr)\,, \tag{0.13}$$

provided the corresponding expectations exist.

E2 monotonicity: if $h(x) \ge 0$ for all real x, then $\mathsf{E}\bigl(h(X)\bigr) \ge 0$; in other words, if the real functions f, g are such that $f(x) \le g(x)$ for all real x, then

$$\mathsf{E}\bigl(f(X)\bigr) \le \mathsf{E}\bigl(g(X)\bigr)\,, \tag{0.14}$$

provided the corresponding expectations exist.

Recall three important special cases: the *variance* $\mathsf{Var}(X)$ of a rv X, its r-th *moment* $\mathsf{E}(X^r)$, and its *moment generating function*, $M_X(t)$,

$$\mathsf{Var}(X) \stackrel{\mathsf{def}}{=} \mathsf{E}\bigl(X - \mathsf{E}(X)\bigr)^2\,, \qquad M_X(t) \stackrel{\mathsf{def}}{=} \mathsf{E}\bigl(e^{tX}\bigr)\,.$$

Exercise 0.15. *Let X be a rv, and let $g : \mathbb{R} \to [0, \infty]$ be an increasing function such that $\mathsf{E}\bigl(g(X)\bigr) < \infty$. Show that for any real a, one has*

$$P\bigl(X > a\bigr) \le \frac{\mathsf{E}\bigl(g(X)\bigr)}{g(a)}\,. \tag{0.15}$$

In particular, $\mathsf{P}\bigl(X > a\bigr) \le \mathsf{E}\bigl(\exp\{\lambda(X - a)\}\bigr)$ for any real a and any $\lambda > 0$.

Notice that the Markov inequality and the Chebyshev inequality are special cases of (0.15).

Multivariate case

In the multivariate case, the expectation is defined similarly and has properties analogous to the considered above. Additionally, we mention two other properties:

E3 multivariate linearity: let $(X_1, \dots, X_n)$ be a random vector, g_1, ..., g_n be real functions, and a_1, ..., a_n be real numbers. Then

$$\mathsf{E}\Big(\sum_{k=1}^n a_k\, g_k(X_k)\Big) = \sum_{k=1}^n a_k\, \mathsf{E}\big(g_k(X_k)\big)\,. \qquad (0.16)$$

E4 independence: if X_1, ..., X_n are independent rv's, so that their joint pmf/pdf factorises,

$$p_{X_1,\dots,X_n}(x_1,\dots,x_n) = \prod_{k=1}^n p_{X_k}(x_k)\,,$$

then for all real functions g_1, ..., g_n one has

$$\mathsf{E}\Big(\prod_{k=1}^n g_k(X_k)\Big) = \prod_{k=1}^n \mathsf{E}\big(g_k(X_k)\big)\,. \qquad (0.17)$$

We say that the variables X and Y are *uncorrelated* if their covariance,

$$\mathsf{Cov}(X,Y) \stackrel{\mathsf{def}}{=} \mathsf{E}\big((X-\mathsf{E}(X))(Y-\mathsf{E}(Y))\big) \equiv \mathsf{E}(XY) - \mathsf{E}(X)\,\mathsf{E}(Y)\,, \qquad (0.18)$$

vanishes, $\mathsf{Cov}(X,Y) = 0$. In particular, any pair of independent variables is uncorrelated.

By linearity property **E3**, the variance $\mathsf{Var}\big(\sum_{k=1}^n X_k\big)$ of the sum of rv's X_1, ..., X_n equals

$$\mathsf{Var}\Big(\sum_{k=1}^n X_k\Big) = \sum_{k=1}^n \mathsf{Var}\big(X_k\big) + 2\sum_{k<l} \mathsf{Cov}(X_k, X_l)\,.$$

Thus, if the variables X_1, ..., X_n are pairwise uncorrelated (in particular, independent), then

$$\mathsf{Var}\Big(\sum_{k=1}^n X_k\Big) = \sum_{k=1}^n \mathsf{Var}\big(X_k\big)\,. \qquad (0.19)$$

Conditional expectation

Let X be a discrete rv on a sample space Ω, and let $A \subseteq \Omega$ be an event. The *conditional expectation* of X given A is a number $\mathsf{E}(X \mid A)$ defined by

$$\mathsf{E}(X \mid A) = \sum_x x\, \mathsf{P}(X = x \mid A)\,, \qquad (0.20)$$

where the sum runs through all possible values of X.

In particular, we have the partition theorem for expectation: if events $B_1, \dots, B_n$ form a partition of the sample space Ω, then

$$\mathsf{E}(X) = \sum_{k=1}^{n} \mathsf{E}(X \mid B_k)\,\mathsf{P}(B_k)\,.$$

Using the definition (0.20), it is immediate to compute $\mathsf{E}(X \mid Y = y)$; we recall that then $\mathsf{E}(X \mid Y)$ is a random variable such that $\mathsf{E}\big(\mathsf{E}(X \mid Y)\big) = \mathsf{E}(X)$.

Limiting results

Theorem 0.16 (Law of Large Numbers)**.** *Let $X_1, \dots, X_n$ be iid (independent, identically distributed) rv's such that*

$$\mathsf{E}(X_k) \equiv \mu\,, \qquad \mathsf{Var}(X_k) = \sigma^2\,.$$

Denote $S_n \stackrel{\mathsf{def}}{=} \sum_{k=1}^{n} X_k$. Then for any fixed $a > 0$

$$\mathsf{P}\big(|n^{-1}S_n - \mu| > a\big) \to 0 \tag{0.21}$$

as $n \to \infty$.

Theorem 0.17 (Central Limit Theorem)**.** *Under the conditions of the previous theorem, denote*

$$S_n^* \stackrel{\mathsf{def}}{=} \frac{S_n - n\mu}{\sqrt{\mathsf{Var}(S_n)}} \equiv \frac{S_n - n\mu}{\sigma\sqrt{n}}\,.$$

Then, as $n \to \infty$, the distribution of S_n^* converges to that of the standard Gaussian random variable (ie., $\mathcal{N}(0,1)$): for every fixed $a \in \mathbb{R}$,

$$\mathsf{P}\big(S_n^* \le a\big) \to \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\,dy\,. \tag{0.22}$$

Moment generating functions

As mentioned before, the *moment generating function* (or *mgf*) of a rv X is defined via

$$M_X(t) \stackrel{\mathsf{def}}{=} \mathsf{E}\big(e^{tX}\big)\,. \tag{0.23}$$

We finish by listing several useful properties of mgf's.

M1 For each positive integer r

$$\mathsf{E}(X^r) = \frac{d^r M_X}{dt^r}(0)\,.$$

M2 [uniqueness] The mgf $M_X(t)$ of X uniquely determines the probability distribution of X, provided that $M_X(t)$ is finite in some neighbourhood of the origin.

M3 [linear transformation] If X has mgf $M_X(\mathrm{t})$, and $Y = aX + b$, then

$$M_Y(t) = e^{bt} M_X(at)\,.$$

M4 [independence] Suppose that $X_1, \ldots, X_n$ are independent rv's and let $Y = \sum_{k=1}^{n} X_k$. Then

$$M_Y(t) = \prod_{k=1}^{n} M_{X_k}(t)\,.$$

M5 [convergence] Suppose that $Y_1, Y_2, \ldots$ is an infinite sequence of rv's, and that Y is a further random variable. Suppose that $M_Y(\mathrm{t})$ is finite for $|t| < a$ for some positive a and that for all $t \in (-a, a)$

$$M_{Y_n}(t) \to M_Y(t) \quad \text{as } n \to \infty\,.$$

Then, as $n \to \infty$,

$$\mathsf{P}(Y_n \le c) \to \mathsf{P}(Y \le c)\,.$$

for all real c such that $\mathsf{P}(Y = c) = 0$.