Selecting r out of n objects (symbols):

Order matters $\rightarrow$ Duplication allowed $\downarrow$	Yes	No
No	P_r^n	C_r^n
Yes	n^r	C_r^{n+r-1}

Binomial formula

$$(x+y)^n = \sum_{i=0}^{n} \binom{n}{i} x^{n-i} y^i$$

Multinomial formula

$$(x+y+z)^n \sum_{\substack{i,j,k \geq 0 \\ i+j+k=n}} \binom{n}{i,j,k} x^i y^j z^k$$

Partitioning n objects into groups of size $n_1, n_2, ...n_k$:

$$\binom{n}{n_1, n_2, ...n_k}$$

(when groups of the same size are considered *distinct*). Also, the number of permutations of n_1 a's, n_2 b's, ...

For a specific **RANDOM EXPERIMENT**, we should be able to construct the corresponding `samples space` of `simple events`.

There are *three* important types of random experiments with `equally likely` simple events:

1. Randomly permuting (arranging) n objects, ether in a row ($n!$ simple events) of in a circle, with $(n-1)!$ simple events.
2. Rolling a regular 'die' with k sides, n times (k^n equally likely simple events - ordered n-tuples).
3. Dealing k 'cards' out of n (sample space consists of C_k^n *unordered* 'hands').

For any of these, $\Pr(A) = \frac{\#(A)}{\#(\Omega)}$, where A is any `event`.

To deal with `events` in general, we use **rules of Boolean algebra.**

Two distributive laws:

$$(A \cap B) \cup (C \cap D) =$$
$$(A \cup C) \cap (A \cup D \cap (B \cup C) \cap (B \cup D)$$

and

$$(A \cup B) \cap (C \cup D) =$$
$$(A \cap C) \cup (A \cap D) \cup (B \cap C) \cup (B \cap D)$$

and two De'Morgan's laws:

$$\begin{aligned}\overline{A \cap B \cap C} &= \bar{A} \cup \bar{B} \cup \bar{C} \\ \overline{A \cup B \cup C} &= \bar{A} \cap \bar{B} \cap \bar{C}\end{aligned}$$

Rules of Probability

$$\begin{aligned}\Pr(\bar{A}) &= 1 - \Pr(A) \\ \Pr(A \cap \bar{B}) &= \Pr(A) - \Pr(A \cap B) \\ \Pr(A \cup B) &= \Pr(A) + \Pr(B) - \Pr(A \cap B)\end{aligned}$$

Potential mistakes ('notationally' speaking)

$$A \cap B \cup C$$

$$\Pr(A) \cap \Pr(B)$$

$$\Pr(A + B)$$

Maxim: Probability of *any* Boolean expression can be always reduced to a linear combination of probabilities of the individual events and of their 'simple' (no bars, no unions, no duplicates) intersections.

This is because we know how to simplify the probability of:

- a *union* of any number of events:
  $\Pr(G \cup H) = \Pr(G) + \Pr(H) - \Pr(G \cap H)$,
- a *complement*: $\Pr(\bar{G}) = 1 - \Pr(G)$,
- an *intersection* of *several complements* (De Morgan makes them into a single complement),
- an *intersection* with a *complement*:
  $\Pr(G \cap \bar{H}) = \Pr(G) - \Pr(G \cap H)$,
- an *intersection* of *unions* is changed into *unions* of *intersections* by distributive law,
- an *intersection* with *duplicates* - remove the duplicates: $P(A \cap B \cap A) = \Pr(A \cap B)$.

A systematic application of these rules is *guaranteed* to work. The remaining formulas, such as $A \cap \bar{A} = \varnothing$, will speed up the process.

Know how to :

- Draw a **probability tree** of a simple 2 or three-stage experiment, use **product rule** to compute the probability of each path.
- Compute the conditional probability of B given that A has happened, for *any* A and B (*some* of these probabilities are very simple - when B follows A, just let A happen), otherwise

$$\Pr(B|A) = \frac{\Pr(A \cap B)}{\Pr(A)}$$

- A special case: in a multi-stage experiment, compute the conditional probability of something which happened at an early stage(s), given what happened later (**Bayes' rule**).
- Apply the formula of **total probability**, using a convenient **partition**

$$\Pr(B) = \Pr(A_1)\Pr(B|A_1) + \Pr(A_2)\Pr(B|A_2) + ...$$

Mutual independence of k events (which we should be able to establish, *on our own*, based on the description of the experiment and the definition of the events), implies:

$$\Pr(A_1 \cap \bar{A}_3 \cap A_7 \cap \bar{A}_{12}) = \Pr(A_1) \cdot \Pr(\bar{A}_3) \cdot \Pr(A_7) \cdot \Pr(\bar{A}_{12})$$

(all together, $3^k - 1 - 2k$ of these, i.e. 4 for two events, 20 for three, 72 for four, etc.).

Also: any Boolean combination of a *group* of these is independent of a Boolean combination of the *others* (none of which may come from the first group).

Final rule: probability of *intersection* of *independent* events equals the *product* of individual probabilities. Thus, we can compute the probability of *any* Boolean combination of *independent* events, based on the *individual* probabilities.

RANDOM VARIABLE (RV for short) is defined by assigning, to each simple event, a **single number** (in this chapter, almost always an integer). Recall, for comparison, that an **event** is defined by assigning, to each simple event, a Boolean value ('yes', I am an element of A, 'no', I am outside of A). Note that, if X is a RV, $X = 3$ defines an *event*.

Distribution of a RV is a *table* with two rows: the RV's possible values, and the corresponding probabilities (we should be able to build it).

Probability function is an expression, say $f(i)$, which enables us to *compute* the probability of $X = i$ for each possible i (the **range** of possible values of i *must* be properly specified).

Probability generating function $P(z) = \sum_{\text{All } i} f(i) \times z^i$. Differentiating, and substituting $z = 1$, yields the *factorial moments*.

Once we have the distribution of X, we can answer *any* question concerning X (and forget the experiment it came from).

In addition to *probability function* $f(i)$, we can also define (less important) **distribution function** $F(k) = \sum_{i=L}^{k} f(i)$ (same range, equivalent information - be able to build a *table* of values).

Bivariate distribution: is a 2D table of all possible values of both X and Y, and their *joint* probabilities, i.e. $\Pr(X = i \cap Y = j)$.

Marginal (regular) **distribution** of X is obtained from a *bivariate distribution* by adding the probabilities in each 'row' (column). Note that this summation is over the *conditional range* of Y, when it's done 'symbolically'.

Independence of 2 or more RVs (a natural concept) can be established based on two conditions: (i) $f(i, j, ..)$ is 'separable', and (ii) all ranges are 'marginal'. The joint distribution thus becomes *redundant* (can be constructed from the individual 'marginals').

Conditional distribution of Y given $X = \mathbf{i}$ is constructed by 're-normalizing' the probabilities of the $\mathbf{i}^{th}$ row (column). It has all properties of a regular distribution. When X and Y are *independent*, conditional distribution is the same as the corresponding marginal (i.e. *ignore,* or remove, the condition). This is the same as what we had for *events*: $\Pr(A \cup B \mid C \cap \bar{D}) = \Pr(A \cup B)$ when A, B, C and D are (mutually) independent.

Transforming a RV, e.g. $U \equiv (X-1)^2$ defines a new RV U. To find its distribution, add a new row of U values to the X distribution (a table), then list all *unique* values of U (and their total probabilities) in a separate table (from the smallest to the largest).

Transforming X and Y into $U = g(X, Y)$: Using the bivariate table of X-Y probabilities, compute and insert (under a slash) the values of U. Then collect the *unique* values and their *total* probabilities in a usual 'neat' table.

EXPECTED VALUE (mean):

$$\begin{aligned} \mathbb{E}(X) &= \mu_x = \sum_{\text{All } i} i \times \Pr(X = i) \\ \mathbb{E}[g(X)] &= \sum_{\text{All } i} g(i) \times \Pr(X = i) \neq g(\mu_x) \end{aligned}$$

We can bypass the distribution of $Y \equiv g(X)$.

Exception (**linear transformation**):

$$\mathbb{E}[aX + b] = a\mu_x + b$$

Similarly (**bivariate case**)

$$\mathbb{E}[h(X, Y)] = \sum_{\text{All } i \text{ and } j} h(i, j) \times \Pr(X = i \cap Y = j) \neq h(\mu_x, \mu_y)$$

Exception (linear case):

$$\mathbb{E}[aX + bY + c)] = a\mu_x + b\mu_y + c$$

even when X and Y are *not* independent (commonly, we are simply *adding* several RVs).

This can be extended to a linear combination of *any* number of RVs.

- *Independence* implies that

  $$\mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$$

  and also

  $$\mathbb{E}[g(X) \cdot h(Y)] = \mathbb{E}[g(X)] \cdot \mathbb{E}[h(Y)]$$

- Univariate **moments** are `simple` $\mathbb{E}(X^k)$, `central` $\mathbb{E}[(X-\mu)^k]$ and `factorial` $\mathbb{E}\left[X(X-1)(X-2)..(X-k+1)\right]$

- The most important of these are the **mean** $\mu \equiv \mathbb{E}(X)$ and the **variance** $\text{Var}(X) \equiv \mathbb{E}[(X-\mu)^2] = \mathbb{E}[X^2] - \mu^2$.

- **Standard deviation** is $\sigma \equiv \sqrt{\text{Var}(X)}$ represents a *typical* (but not quite the average) distance between random values of X and the theoretical mean μ.

- Important formula: $\text{Var}(aX + b) = a^2 \text{Var}(X)$.

- *Bivariate* moments: The only important one is the $(1^{st}, 1^{st})$ central moment, called

covariance between X and Y:

$$\text{Cov}(X,Y) \equiv \mathbb{E}[(X-\mu_x)\cdot(Y-\mu_y)] = \mathbb{E}(X\cdot Y) - \mu_x \cdot \mu_y$$

- Basic properties: $\text{Cov}(X,Y) =$Cov(Y,X), $\text{Cov}(X,X) =$Var(X), and when X and Y are *independent*, $\text{Cov}(X,Y) = 0$ (not necessarily reverse).
- **Correlation coefficient:**
$$\rho_{xy} \equiv \frac{\text{Cov}(X,Y)}{\sigma_x \cdot \sigma_y}$$
(dimensionless).
- **Important formula:**
$$\begin{aligned}&\text{Var}(a_1X_1 + a_2X_2 + ...a_nX_n + c) = a_1^2\text{Var}(X_1) + a_2^2\text{Var}(X_2) + ...\\&... + a_k^2\text{Var}(X_k) + 2a_1a_2\text{Cov}(X_1,X_2) + 2a_1a_3\text{Cov}(X_1,X_3) + ...\\&... + 2a_{n-1}a_n\text{Cov}(X_{n-1},X_n)\end{aligned}$$
Special case:
$$\begin{aligned}&\text{Var}(X_1 + X_2 + ...X_n) = \text{Var}(X_1) + \text{Var}(X_2) + ... + \text{Var}(X_n) +\\&+2\text{Cov}(X_1,X_2) + 2\text{Cov}(X_1,X_3) + ... + 2\text{Cov}(X_{n-1},X_n)\end{aligned}$$
And, more **special** yet: When all X's are *independent*, and drawn from the *same distribution* (IID RVs), such as playing n rounds of the same game:
$$\text{Var}(X_1 + X_2 + ...X_n) = n \cdot Var(X)$$
- **Distributive law of covariance**
- **Moment generating function** (can get from $P(z)$ by $z \to e^t$):
$$M_x(t) = \mathbb{E}(e^{tX})$$
$$\mathbb{E}(X^k) = \left.\tfrac{d^k M_x(t)}{dt^k}\right|_{t=0}$$
$$M_{X+Y+Z}(t) = M_x(t)\cdot M_y(t)\cdot M_z(t) \quad \textit{need independence}$$
$$M_{X_1+X_2+...+X_n}(t) = M_x(t)^n \qquad \text{IID case}$$
$$M_{aX+b}(t) = e^{bt}\cdot M_x(at)$$
- **Conditional expected value**, variance, moment generating function, etc. uses the corresponding conditional distribution (everything else is the same).
- **Probability generating function** (only for *integer-valued* RVs):
$$P_x(z) \equiv \mathbb{E}(z^X) = p_0 + p_1z + p_2z^2 + p_3z^3 + ...$$
Its main property (shared with *moment* generating function) is:
$$P_{X_1+X_2+X_3}(z) = P_1(z)\cdot P_2(z)\cdot P_3(z)$$
when X_1, X_2, X_3 are mutually *independent.* This time, the result can be (Taylor) *expanded*, where the coefficient of z^i is equal to $\Pr(X_1 + X_2 + X_3 = i)$, which would

be very difficult (if not practically impossible) to find otherwise.

The **most COMMON DISTRIBUTIONS** of integer-valued RVs:

Name	$f(i)$	Range	Mean	Var	PGF
$\mathcal{B}$	$\binom{n}{i}p^i q^{n-i}$	$0..n$	np	npq	$(q+pz)^n$
$\mathcal{NB}$	$\binom{i-1}{k-1}p^k q^{i-k}$	$k..\infty$	$\frac{k}{p}$	$\frac{k}{p}\left(\frac{1}{p}-1\right)$	$\left(\frac{pz}{1-qz}\right)^k$
$\mathcal{HG}$	$\frac{\binom{K}{i}\binom{N-K}{n-i}}{\binom{N}{n}}$	$\times$	$n\frac{K}{N}$	$n\frac{K(N-K)}{N^2}\frac{N-n}{N-1}$	$\times$
$\mathcal{PS}$	$\frac{\Lambda^i}{i!}e^{-\Lambda}$	$0..\infty$	Λ	Λ	$e^{\Lambda(z-1)}$

Only for $\mathcal{NB}$ we were able to find

$$F(i) = 1 - \sum_{j=0}^{k-1} \binom{i}{j} p^j q^{i-j}$$

(all others, we must add individual probabilities).

Multinomial distribution is a simple extension of binomial. Instead of S and F, each trial can result in one of *three* (four, ...) possible outcomes, and X, Y and Z count how many of them happen in n independent trials. The individual `marginals` are all *binomial* (with obvious parameters). We need only two extra formulas:

$$f_{xyz}(i,j,k) = \Pr(X = i \cap Y = j \cap Z = k) = \binom{n}{i,j,k} p_x^i p_y^j p_z^k$$

for any $i, j, k \geq 0$ such that $i + j + k = n$. The tricky part now is to figure out which combinations of i, j and k contribute to the event whose probability we need (winning a series of games, etc.).

The PGF of $aX + bY + c$ is $(p_x z^a + p_y z^b + p_z)^n \times z^c$, where a, b and c must be integers.

Finally

$$\text{Cov}(X,Y) = -np_x p_y$$

For $U = U_0 + T$ and $V = V_0 + T$ where U_0. V_0 and T *are* multinomial (but U and V are not), we get the following extension:

$$\text{Cov}(U,V) = -n(p_u p_v - p_{uv})$$

To compute $\Pr(U = 3 \cap V = 5)$, we must first express it in terms of U_0, V_0 and T.

Multivariate Hypergeometric (box with red, blue and green marbles):

$$f_{xyz}(i,j,k) = \frac{\binom{K_1}{i}\binom{K_2}{j}\binom{K_3}{k}}{\binom{N}{n}}$$

$$\text{Cov}(X,Y) = -n \cdot \frac{K_1}{N} \cdot \frac{K_2}{N} \cdot \frac{N-n}{N-1}$$

$$\text{Cov}(U,V) = -n\left(\frac{K_1}{N} \cdot \frac{K_2}{N} - \frac{K_{12}}{N}\right) \cdot \frac{N-n}{N-1}$$

CONTINUOUS RVs

Key concept: `probability density function`

$$f(x) = \frac{dF(x)}{dx}$$

It must be non-negative, and integrate (over all values) to 1. Often, we are given $f(x)$ and have to find $F(x)$ by integrating $f(x)$.

Expected value: Everything is same as with discrete RVs, only summation changes to integration. *Old formulas still apply*:

$$\begin{aligned}\text{Var}(X) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 \\ \mathbb{E}(aX + bY + c) &= a\mathbb{E}(X) + b\mathbb{E}(Y) + c \\ \text{Var}(aX + bY + c) &= a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X, Y)\end{aligned}$$

Moment Generation function:

$$M_x(t) = \mathbb{E}(e^{tX})$$

Common continuous distributions:

Name	$f(x)$	Range	Mean	Var	MGF
Uniform	$\frac{1}{b-a}$	$a..b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{bt}-e^{at}}{t(b-a)}$
Exponential	$\frac{1}{\beta}e^{-x/\beta}$	$0..\infty$	β	β^2	$\frac{1}{1-\beta\cdot t}$
Gamma	$\frac{x^{k-1}}{(k-1)!\beta^k}e^{-x/\beta}$	$0..\infty$	$k\beta$	$k\beta^2$	$\left(\frac{1}{1-\beta\cdot t}\right)^k$

`Exponential distribution` describes time till the next arrival (from now). Also: time between consecutive arrivals, and time till the first arrival after 10:30 (any specific time).

n people fishing independently is the same as one person fishing at a rate equal to the sum of the individual rates.

`Gamma` distribution describes time till the k^{th} arrival (from 'now').

CENTRAL LIMIT THEOREM:

Let X have any distribution whatsoever, with a mean of μ and standard deviation of σ. Then

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}}$$

has, in the $n \to \infty$ limit, the **standardized Normal** distribution, which has the MGF of $\exp(\frac{t^2}{2})$ and PDF of

$$f(z) = \frac{\exp(-\frac{z^2}{2})}{\sqrt{2\pi}}$$

(not analytically integrable, but Maple can handle it).