PROBABILITY SUMMARIES — O.Knill, MA2c, Caltech, April 1995

BOOLEAN ALGEBRA
Ω finite set. A set $\mathcal{A}$ of subsets of Ω is called a finite **Boolean algebra**, if

$$\boxed{\begin{array}{l} \Omega \in \mathcal{A}, \\ A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}, \\ A, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}. \end{array}}$$

PROPERTIES: A finite Boolean algebra $(\Omega, \mathcal{A})$ is closed under all set theoretical operations: if $A, B \in \mathcal{A}$, then

$$\boxed{\begin{array}{l} \emptyset \in \mathcal{A},\ A \cap B \in \mathcal{A} \\ A \setminus B \in \mathcal{A},\ A \Delta B \in \mathcal{A} \end{array}}$$

PROBABILITY MEASURE: $P : \mathcal{A} \to \mathbf{R}$

$$\boxed{\begin{array}{l} \mathrm{P}[A] \geq 0, \text{ (non-negativity)} \\ \mathrm{P}[\Omega] = 1, \text{ (normalization)} \\ \mathrm{P}[\bigcup_{i=1}^n A_i] = \sum_{i=1}^n \mathrm{P}[A_i] \\ A_i \cap A_j = \emptyset, \text{ all } i, j, \text{ (additivity)} \end{array}}$$

PROBABILITY SPACE.

$(\Omega, \mathcal{A}, P)$ is a **finite probability space** if $\mathcal{A}$ is a Boolean algebra on the finite set Ω and P is a probability measure on $(\Omega, \mathcal{A})$.

PROPERTIES:
$A \subset B \Rightarrow \mathrm{P}[A] \leq \mathrm{P}[B]$.
$\mathrm{P}[A^c] = 1 - \mathrm{P}[A]$,
$\mathrm{P}[\emptyset] = 0$
$\sum_j P[B \cap A_j] = P[B]$
if $\bigcup_j A_j = B, A_j \cap A_i = \emptyset$.

BAYES RULE: $\boxed{\mathrm{P}[A_i|B] = \mathrm{P}[B|A_i]P[A_i]/\sum_j P[B|A_j]P[A_j]}$ if $\bigcup_j A_j = \Omega, A_j \cap A_i = \emptyset$ and $\mathrm{P}[A|B] = \mathrm{P}[A \cap B]/\mathrm{P}[B]$.

SWITCH ON, SWITCH OFF:
$\mathrm{P}[\bigcup_{i=1}^n A_i] = \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < i_2 < \ldots < i_k \leq n} \mathrm{P}[A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}]$

CONSTRUCTING NEW PROBABILITY SPACES:

Change of algebra: $(\Omega, \mathcal{B}, Q)$, where $\mathcal{B} \subset \mathcal{A}$ is a Boolean algebra and Q is restriction of P to $\mathcal{B}$.
Product space:
$(\Omega, \mathcal{A}, P) = (\Omega_1, \mathcal{A}_1, P_1) \times (\Omega_2, \mathcal{A}_2, P_2)$, where $\Omega = \Omega_1 \times \Omega_2$, $\mathcal{A}$ is the smallest algebra containing $\mathcal{A}_1 \times \mathcal{A}_2 = \{A_1 \times A_2\ | A_i \in \mathcal{A}_i\}$ and $\mathrm{P}[(A_1 \times A_2)] = P_1[A_1] \cdot P_2[A_2]$.

Conditional probability space:
$(B, \mathcal{A} \cap B, \mathrm{P}[\cdot|B])$, if $\mathrm{Pr}[B] > 0$ and $\mathcal{A} \cap A = \{A \cap B \mid A \in \mathcal{A}\}$ and $\mathrm{P}[A|B] = \frac{\mathrm{P}[A \cap B]}{\mathrm{P}[B]}$

INDEPENDENT EVENTS: $(\Omega, \mathcal{A}, P)$ finite probability space. $A, B \in \mathcal{A}$ are called **independent** if

$$\mathrm{P}[A \cap B] = \mathrm{P}[A] \cdot \mathrm{P}[B] .$$

A finite set $\{A_i\}_{i \in I}$ of events is called **independent** if for all $J \subset I$

$$\mathrm{P}[\bigcap_{i \in J} A_i] = \prod_{i \in J} \mathrm{P}[A_i] .$$

PROPERTIES:
$A, B \in \mathcal{A}$ are independent, if and only if either $\mathrm{P}[B] = 0$ or $\mathrm{P}[A|B] = \mathrm{P}[A]$.

$(\Omega, \mathcal{A}, P) = (\Delta, \mathcal{B}, Q)^n$ (product space). Given $B_i \in \mathcal{B}$ then

$$A_i = \{\omega = (\omega_1, \ldots, \omega_n) \mid \omega_i \in B_i\}$$

are all independent.

RANDOM VARIABLE: A **random variable** on a finite probability space $(\Omega, \mathcal{A}, P)$ is a map $X : \Omega \to \mathbf{R}$ such that for all $a \in \mathbf{R}$, we have $\{X = a\} \in \mathcal{A}$.

EXPECTATION: The **expectation** of a random variable X is defined as

$$\mathrm{E}[X] = \sum_{a \in X(\Omega)} a \cdot \mathrm{P}[X = a] = \sum_{A \in \mathcal{A}, A \text{ atom}} X(A) \cdot \mathrm{P}[A] ,$$

where an **atom** is a set in $\mathcal{A}$ so that $B \subset A, B \in \mathcal{A} \Rightarrow B = A$ or $B = \emptyset$. If $\mathcal{A} = \{A \subset \Omega\}$, then the atoms are all of the form $\{\omega\}$ and

$$\mathrm{E}[X] = \sum_{\omega \in \Omega} X(\omega) \mathrm{P}[\{\omega\}] .$$

By the definition of a random variable, X must be constant on each atom A and $X(A)$ is defined as the common value, X takes on A. The two expressions for $\mathrm{E}[X]$ in the box to the left are seen to be the same using $a = X(A)$ and $\mathrm{P}[X = a] = \sum_{A \text{ atom } X(A) = a} \mathrm{P}[A]$.

PROPERTIES OF EXPECTATION: For random variables X, Y and $\lambda \in \mathbf{R}$

$$\mathrm{E}[X+Y] = \mathrm{E}[X] + \mathrm{E}[Y] \qquad \mathrm{E}[\lambda X] = \lambda \mathrm{E}[X]$$
$$X \leq Y \Rightarrow \mathrm{E}[X] \leq \mathrm{E}[Y] \qquad \mathrm{E}[X^2] = 0 \Leftrightarrow X = 0$$
$$\mathrm{E}[X] = c \text{ if } X(\omega) = c \text{ is constant} \qquad \mathrm{E}[X - \mathrm{E}[X]] = 0\ .$$

PROOF OF THE ABOVE PROPERTIES:

$\mathrm{E}[X+Y] = \sum_{A \text{ atom}} (X+Y)(A) \cdot \mathrm{P}[A] = \sum_{A \text{ atom}} (X(A) + Y(A)) \cdot \mathrm{P}[A] = \mathrm{E}[X] + \mathrm{E}[Y]$
$\mathrm{E}[\lambda X] = \sum_{A \text{ atom}} (\lambda X)(A) \mathrm{P}[A] = \lambda \sum_{A \text{ atom}} X(A) \mathrm{P}[A] = \lambda \mathrm{E}[X]$
$X \leq Y \Rightarrow X(A) \leq Y(A)$, for all atoms A and $\mathrm{E}[X] \leq \mathrm{E}[Y]$
$\mathrm{E}[X^2] = 0 \Leftrightarrow X^2(A) = 0$ for all atoms A $\Leftrightarrow X = 0$
$X(\omega) = c$ is constant $\Rightarrow \mathrm{E}[X] = c \cdot \mathrm{P}[X = c] = c \cdot 1 = c$
$\mathrm{E}[X - \mathrm{E}[X]] = \mathrm{E}[X] - \mathrm{E}[\mathrm{E}[X]] = \mathrm{E}[X] - \mathrm{E}[X] = 0$

VARIANCE, STANDARD DEVIATION

Variance

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

Standard deviation

$$\sigma[X] = \sqrt{\mathrm{Var}[X]}\ .$$

Covariance

$$\mathrm{Cov}[X,Y] = \mathrm{E}[(X - \mathrm{E}[X]) \cdot (Y - \mathrm{E}[Y])]\ .$$

Correlation of $\mathrm{Var}[X] \neq 0, \mathrm{Var}[Y] \neq 0$

$$\mathrm{Corr}[X,Y] = \frac{\mathrm{Cov}[X,Y]}{\sigma[X]\sigma[Y]}\ .$$

$\mathrm{Corr}[X,Y] = 0$: **uncorrelated** X and Y.

PROPERTIES of VAR, COV, and CORR:

$\mathrm{Var}[X] \geq 0$.
$\mathrm{Var}[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2$.
$\mathrm{Var}[\lambda X] = \lambda^2 \mathrm{Var}[X]$.

$\mathrm{Cov}[X,Y] = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y]$.
$\mathrm{Cov}[X,Y] \leq \sigma[X]\sigma[Y]$ (Schwarz inequality).

$-1 \leq \mathrm{Corr}[X,Y] \leq 1$.
$\mathrm{Corr}[X,Y] = 1$ if $X - \mathrm{E}[X] = Y - \mathrm{E}[Y]$
$\mathrm{Corr}[X,Y] = -1$ if $X - \mathrm{E}[X] = -(Y - \mathrm{E}[Y])$.

EXAMPLE: BERNOULLI DISTRIBUTION $(\Omega = \{0,1\}^n, \mathcal{A}, P = Q^n)$, where $Q[\{1\}] = p, Q[\{0\}] = q = 1 - p$.

$$X(\omega) = \sum_{i=1}^{n} \omega_i$$

$$P[X = k] = \binom{n}{k} p^k q^{n-k}$$

$$E[X] = \sum_{k=1}^{n} k \binom{n}{k} p^k q^{n-k} = pn$$

$$\mathrm{Var}[X] = \sum_{k=1}^{n} k^2 \binom{n}{k} p^k q^{n-k} - E[X]^2 = npq$$

INDEPENDENT RANDOM VARIABLES: X, Y are **independent** if for all $a, b \in \mathbf{R}$

$$P[X = a; Y = b] = P[X = a] \cdot P[Y = b]\ .$$

A finite collection $\{X_i\}_{i \in I}$ of random variables are **independent**, if for all $J \subset I$ and $a_i \in \mathbf{R}$

$$P[X_i = a_i, i \in J] = \prod_{i \in J} P[X_i = a_i]\ .$$

PROPERTIES:

- If X and Y are independent, then $\mathrm{E}[X \cdot Y] = \mathrm{E}[X] \cdot \mathrm{E}[Y]$.
- If X_i is a set of independent random variables, then $\mathrm{E}[\prod_{i=1}^n X_i] = \prod_{i=1}^n \mathrm{E}[X_i]$.
- If X, Y are independent then $\mathrm{Cov}[X,Y] = 0$.
- A constant random variable is independent to any other random variable.

REGRESSION LINE: The **regression line** of two random variables X, Y is defined as $y = ax+b$, where

$$a = \frac{\mathrm{Cov}[X,Y]}{\mathrm{Var}[X]}, b = \mathrm{E}[Y] - a\mathrm{E}[X] \;.$$

PROPERTY: Given $X, \mathrm{Cov}[X,Y], \mathrm{E}[Y]$, and the regression line $y = ax+b$ of X, Y. The random variable $\tilde{Y} = aX + b$ minimizes $\mathrm{Var}[Y - \tilde{Y}]$ under the constraint $\mathrm{E}[Y] = \mathrm{E}[\tilde{Y}]$ and is the best guess for Y, when knowing only $\mathrm{E}[Y]$ and $\mathrm{Cov}[X,Y]$. We check $\mathrm{Cov}[X,Y] = \mathrm{Cov}[X,\tilde{Y}]$.

1D RANDOM WALK

$(\Omega = \{-1,1\}^N = \{(\omega = (\omega_1 < \dots, \omega_N) \mid \omega_i \in \{-1,1\}\}, \mathcal{A} = \{A \subset \Omega\}, \mathrm{P}[A] = |A|/|\Omega|)$.
The random variables $X_k(\omega) = \omega_k$ define the k'th step. The random variables $S_n = \sum_{k=1}^n X_k(\omega)$ describe the location of the walker at time n. If X_k is interpreted as the win or loss in a game at time k, then S_n is the **total win or loss** up to time n. Ω is the set of all possible trajectories up to time N.

RANDOM WALK PROPERTIES

a) $I \subset \{1, \dots, N\}$, $x_i \in \{-1,1\}$

$$\mathrm{P}[X_i = x_i, i \in I] = 2^{-|I|} \;.$$

b) $\mathrm{E}[X_k] = 0$,
c) $\mathrm{E}[S_k] = 0$.
d) $n+x$ even: $\mathrm{P}[X_n = x] = 2^{-n} \begin{pmatrix} n \\ \frac{n+x}{2} \end{pmatrix}$.
$n+x$ odd: $\mathrm{P}[X_n = x] = 0$.

GAMBLING SYSTEM. A **gambling system** attached to the random walk is sequence of random variables V_k such that every event $\{V_n = c\}$ is a union of sets of the form $\{\omega_1 = x_1, \dots, \omega_{n-1} = x_{n-1}\}$.
Let V_k be a gambling system, then

$$S_n^V = \sum_{i=1}^n V_k X_k$$

is the **total winnings** with this system. General rule: **You can't beat the system**: $\mathrm{E}[S_N^V] = 0$.

CARDINALITY

$f : A \to B$ is 1 : 1 or injective: $x \neq y \Rightarrow f(x) \neq f(y) \forall x, y \in A$.
$f : A \to B$ is **onto** or surjective): $f(A) = B$.
f is **bijective** $\Leftrightarrow$ f is 1:1 and onto.
A, B are **equivalent**: exists bijection $f : A \to B$.

A equivalent to $\mathbf{N}$: **countably infinite**.
A equivalent to finite set: **finite**.
A not finite nor countably infinite: **uncountable**.

σ ALGEBRA

A σ**-algebra** on Ω is a set $\mathcal{A}$ of subsets of Ω satisfying

$\Omega \in \mathcal{A}$,
$A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$,
$\{A_1, A_2, \dots\} \subset \mathcal{A}$ countable $\Rightarrow \bigcup_{i=1}^\infty A_i \in \mathcal{A}$.

$P : \mathcal{A} \to \mathbf{R}$ is a **probability measure** if

$\mathrm{P}[A] \geq 0$, (non-negativity)
$\mathrm{P}[\Omega] = 1$, (normalization)
$\{A_1, A_2, \dots\}$ countable set of disjoint sets $\Rightarrow$
$\mathrm{P}[\bigcup_{i=1}^\infty A_i] = \sum_{i=1}^\infty \mathrm{P}[A_i]$ **(σ-additivity)**

A **probability space** $(\Omega, \mathcal{A}, P)$ consists of a set Ω, a σ-algebra $\mathcal{A}$ on Ω and a probability measure P on $\mathcal{A}$. If Ω is **finite**, the probability space is called a finite probability space. If Ω is countable, it is called **discrete**.

METRIC. $d : \Omega \times \Omega \to \mathbf{R}$ is called a **metric** if

(i) $d(x,y) \geq 0$,
(ii) $d(x,z) \leq d(x,y) + d(y,z)$
(iii) $d(x,y) = 0 \Leftrightarrow x = y$

The pair (Ω, d) where Ω is a set and d is a metric is called a **metric space**. Examples. $(\mathbf{R^n}, \mathbf{d(x-y)} = ||\mathbf{x}-\mathbf{y}||$, $(\{0,1\}^{\mathbf{N}}, d(x,y) = \sum_{i=1}^\infty |x_i - y_i| 2^{-i})$.

GENERATING ALGEBRAS. Let $\{\mathcal{A}_i\}_{i \in I}$ be a collection of σ-algebras in Ω. Then $\bigcap_{i \in I} \mathcal{A}_i$ is a σ-algebra. It follows that if $\mathcal{S}$ is a set of subsets of Ω, then there exists a **smallest σ-algebra**, which contains $\mathcal{S}$.

BOREL ALGEBRA. Let (Ω, d) be a metric space and let $\mathcal{S}$ be the set of open balls $B_r(x) = \{y \in \Omega \mid d(x,y) < r\}$. The smallest σ-algebra which contains $\mathcal{S}$ is called the **Borel $\sigma-$ algebra** on Ω.

BOREL CANTELLI LEMMA:

a) If $\sum_n \mathrm{P}[A_n] < \infty$, then $\mathrm{P}[A_\infty] = 0$.
b) If A_i are pairwie independent and $\sum_n \mathrm{P}[A_n] = \infty$, then $\mathrm{P}[A_\infty] = 1$.

NOTATION:
$A_\infty := \limsup_{n \to \infty} A_n := \bigcap_{m=1}^\infty \bigcup_{n \geq m} A_n$.
We have $A_\infty = \{\omega \mid \omega$ is in infinitely many $A_i\}$.

EXPECTATION OF DISCRETE RANDOM VARIABLE. A random variable X on a probability space $(\Omega, \mathcal{A}, P)$ is called **discrete**, if $\Omega(X)$ is countable or finite. In this case, the expectation of X is defined as

$$\mathrm{E}[X] = \sum_{a \in X(\Omega)} a \cdot \mathrm{P}[X = a]$$

if the sum converges. $\mathcal{L}^1$ is the set of random variables, for which $\mathrm{E}[|X|] < \infty$.
Variance, Covariance etc. are defined as in the finite case noting that if $f(X) \in \mathcal{L}^1$, then

$$\mathrm{E}[f(X)] = \sum_{a \in X(\Omega)} f(a) \cdot \mathrm{P}[X = a] \; .$$

For example if $X^2 \in \mathcal{L}^1$, then

$$\mathrm{Var}[X] = \mathrm{E}[(X - \mathrm{E}[X])^2] = \sum_{a \in X(\Omega)} (a - m)^2 \mathrm{P}[X = a], \; m = \mathrm{E}[X] \; .$$

EXAMPLES:

$\mathrm{P}[X = k] = \frac{\lambda^k}{k!} e^{-\lambda}$	Poisson	$\mathrm{E}[X] = \lambda$	Electrons from cathode
$\mathrm{P}[X = k] = (1-p)^{k-1} p$	Geometric	$\mathrm{E}[X] = 1/p$	Waiting time for success
$\mathrm{P}[X = k] = \zeta(s)^{-1} k^{-s}$	Zeta	$\mathrm{E}[X] = \zeta(s+1)/\zeta(s)$	$\zeta(s) = \sum_{n=1}^{\infty} n^{-s}$

INTEGRATION=EXPECTATION. $\mathcal{S}$: set of random variables taking finitely many values: Define for $X \in \mathcal{S}$

$$\mathrm{E}[X] := \sum_{a \in X(\Omega)} a \cdot \mathrm{P}[X = a] \; .$$

$\mathcal{L}^1$: set of random variables X for which $\sup_{Y \in \mathcal{S}, Y \leq |X|} \mathrm{E}[Y] < \infty$. For $X \in \mathcal{L}^1$ and $X \geq 0$, the **integral** or **expectation** is defined as

$$\mathrm{E}[X] := \sup_{Y \in \mathcal{S}, Y \leq X} \mathrm{E}[Y] \; .$$

In general, we decompose X into $X = X^+ - X^-$ with $X^{\pm} \geq 0$ and put $\mathrm{E}[X] = \mathrm{E}[X^+] - \mathrm{E}[X^-]$. We write also $\int_\Omega X \; dP$ for $\mathrm{E}[X]$ since **expectation** is **integration**. Variance, Covariance etc. are defined as in the finite case: $\mathrm{Var}[X] = \mathrm{E}[(X - E[X])^2], \mathrm{Cov}[X, Y] = \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])]$.

DISTRIBUTION FUNCTION. A random variable X has the **distribution function** $F(t) = \mathrm{P}[X \leq t]$. **Absolutely continuous random variable**: the **probability density function** if $F' = f$ exists. **Discrete random variable**: F is piecewise constant with countably many jump discontinuities. The **expectation**, **variance** and $\mathrm{E}[g(X)]$ for $g(X) \in \mathcal{L}^1$ is in the continuous case

$$m = \mathrm{E}[X] = \int_{-\infty}^{\infty} x f(x) \; dx, \; \mathrm{Var}[X] = \int_{-\infty}^{\infty} (x - m)^2 f(x) \; dx \; , \mathrm{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) \; dx$$

Comparison: for discrete random variables this was

$$m = \mathrm{E}[X] = \sum_{a \in X(\Omega)} a \mathrm{P}[X = a], \; \mathrm{Var}[X] = \sum_{a \in X(\Omega)} (a - m)^2 \mathrm{P}[X = a], \mathrm{E}[g(X)] = \sum_{a \in X(\Omega)} g(a) \mathrm{P}[X = a]$$

Sometimes, one does not know the distribution of the random variable, then $\mathrm{E}[X], \mathrm{Var}[X]$ and $\mathrm{E}[g(X)]$ have to be computed by integrating (rsp. summing) over Ω.

DISCRETE DISTRIBUTIONS:

Distribution	$\mathrm{P}[x = k]$	Parameters	Domain	Mean	Variance
Binomial	$\binom{n}{k} p^k (1-p)^{n-k}$	$n \in \mathbf{N}$, $p \in [0, 1]$	$\{0, \dots, n\}$	np	$np(1-p)$
Poisson	$\frac{\lambda^k}{k!} e^{-\lambda}$	$\lambda > 0$	$\{0, 1, \dots\}$	λ	λ
Geometric	$(1-p)^{k-1} p$	$p \in (0, 1)$	$\{1,2, \dots\}$	$1/p$	$1/p^2$

ABSOLUTELY CONTINUOUS DISTRIBUTIONS:

Distribution	Density $f(x) =$	Parameters	Domain	Mean	Variance
Uniform	$1_{[a,b]} \cdot (b-a)^{-1}$	$a < b$	$[a, b]$	$(a+b)/2$	$(b-a)^2/12$
Exponential	$\lambda e^{-\lambda x}$	$\lambda > 0$	$\mathbf{R}^+$	$1/\lambda$	$1/\lambda^2$
Normal	$(2\pi\sigma^2)^{-1/2}\, e^{-\frac{(x-m)^2}{2\sigma^2}}$	$m \in \mathbf{R}, \boldsymbol{\sigma^2 > 0}$	$\mathbf{R}$	m	σ^2
Erlang	$\frac{\lambda^k x^{k-1}}{(k-1)!} e^{-\lambda x}$	$\mathbf{R}^+$	$\lambda > 0, k \in \mathbf{N}$	k/λ	k/λ^2

PROPERTIES OF DISTRIBUTION FUNCTIONS:

$F(t) \in [0,1]$	$\mathrm{P}[a < X \leq b] = F(b) - F(a)$
$a \leq b \Rightarrow F(a) \leq F(b)$	$\lim_{t\to-\infty} F(t) = 0, \lim_{t\to\infty} F(t) = 1$
$\lim_{\epsilon \searrow 0} F(a+\epsilon) = F(a)$	$\lim_{\epsilon \searrow 0} F(a-\epsilon) = F(a) - \mathrm{P}[X = a]$

Every function $F : \mathbf{R} \to \mathbf{R}$ satisfying the above properties belongs to a random variable X: define the probability space $(\Omega, \mathcal{A}, P)$, where $\mathcal{A}$ is the Borel σ-algebra on $\Omega = \mathbf{R}$ and P is defined by $\mathrm{P}[[a,b)] = F(b) - F(a) = \mathrm{P}[X \in [a,b)]$.

RANDOM VECTOR. $X = (X_1, X_2, \dots, X_d)$ is called a **random vector** if X_i are random variables. The **distribution function** of X (also called **joint distribution** of $X_1, X_2, \dots, X_d$) is defined as

$$F(t_1, \dots, t_d) = \mathrm{P}[X_1 \leq t_1, X_2 \leq t_2, \dots, X_d \leq t_d]\ .$$

The distribution is **continuous** if there exists a function $f : \mathbf{R^d} \to \mathbf{R}^+$ such that

$$F(t) = \int_{-\infty}^{t_1} \int_{-\infty}^{t_2} \dots \int_{-\infty}^{t_d} f(y_1, y_2, \dots, y_d)\ dy_d \dots dy_2 dy_1\ .$$

TRANSFORMATION OF RANDOM VARIABLES:

• Let F be a continuous invertible distribution function. Let X be a random variable which is uniformly distributed in $[0,1]$. Then $Y = F^{-1}(X)$ gives random numbers with distribution F.

• Given a continuous random variable X with density f and a differentiable invertible function $\phi : \mathbf{R} \to \mathbf{R}$ with inverse ψ. The random variable $Y = \phi(X)$ has the density

$$g(t) = f(\psi(t))|\psi'(t)|\ .$$

• Given a continuous random vector X with density f and a differentiable invertible function $\phi : \mathbf{R^d} \to \mathbf{R^d}$ with inverse ψ. The random variable $Y = \phi(X)$ has the density

$$g(u) = f(\psi(u))|\mathrm{Det}(D\psi(u))|\ .$$

CHARACTERISTIC FUNCTION The **characteristic function** of X is defined as $\phi_X(t) = \mathrm{E}[e^{itX}]$.
Discrete case: $\phi_X(t) = \sum_{a \in X(\Omega)} e^{ita} \mathrm{P}[X = a]$. Continuous case: $\phi_X(t) = \int_{-\infty}^{\infty} e^{itx} f(x)\ dx$.

Distribution	Parameter	Charact. function
Normal	$m \in \mathbf{R}$, $\sigma^2 > 0$	$e^{mit - \sigma^2 t^2/2}$
Standard normal		$e^{-t^2/2}$
Uniform	$[-a, a]$	$\sin(at)/(at)$
Exponential	$\lambda > 0$	$\lambda/(\lambda - it)$
Binomial	$n \in \mathbf{N}$, $p \in [0,1]$	$(p + (1-p)e^{it})^n$
Poisson	$\lambda > 0$	$e^{\lambda(e^{it}-1)}$
Geometric	$p \in (0,1)$	$\frac{pe^{it}}{(1-(1-p)e^{it})}$

CALCULATION OF MOMENTS: $\mathrm{E}[X^k] = (-i)^k \phi_X^{(k)}(0)$. Especially, $\mathrm{E}[X] = -i\phi'_X(0)$.

SUMS OF INDEPENDENT RANDOM VARIABLES: X_i independent with distribution $\phi_i, S = \sum_{i=1}^n X_i$, then $\phi_S(t) = \phi_1(t) \cdot \phi_2(t) \cdots \phi_n(t)$.

THE GAMMA FUNCTION. Some distributions use the Gamma function: $\Gamma(n) = \int_0^\infty z^{n-1} e^{-z}\, dz$. For $n \in \mathbf{N}$, we have $(n-1)!$. Proof. $\Gamma(1) = 1$, $\Gamma(n) = (n-1)\Gamma(n-1)$ by partial integration. Computations like $\Gamma(1/2) = \sqrt{\pi}, \Gamma(3/2) = \sqrt{\pi}/2$ use $\int_{\mathbf{R}} e^{-x^2/2} = \sqrt{2\pi}$.

CHEBYCHEV-MARKOV INEQUALITY. Let $h : \mathbf{R}^+ \to \mathbf{R}^+$ be a monotone function and $X \geq 0$ a random variable with $h(X) \in \mathcal{L}^1$. Then for all $c > 0$

$$h(c) \cdot \mathrm{P}[X \geq c] \leq \mathrm{E}[h(X)] \ .$$

Proof. Take the expectation of $h(c) 1_{X \geq c}(\omega) \leq h(X)(\omega)$. Use the monotonicity and linearity of the expectation.

CHEBYCHEV INEQUALITY. If $X \in \mathcal{L}^2$, then for all $c > 0$

$$\mathrm{P}[|X - \mathrm{E}[X]| \geq c] \leq \frac{\mathrm{Var}[X]}{c^2} \ .$$

Proof. Apply Chebychev-Markov's inequality to $Y = |X - \mathrm{E}[X]|$ and $h(x) = x^2$.

CONVERGENCE IN PROBABILITY. A sequence of random variables X_n **converges in probability** to a random variable X, if for all $\epsilon > 0$,

$$\lim_{n \to \infty} \mathrm{P}[|X_n - X| \geq \epsilon] = 0 \ .$$

WEAK LAW OF LARGE NUMBERS.
Assume X_i have common expectation $E[X_i] = m$ and satisfy $\sup_n \frac{1}{n} \sum_{i=1}^n \mathrm{Var}[X_i] < \infty$. If X_n are pairwise uncorrelated, then $\sum_{i=1}^n X_i/n \to m$ in probability.
Proof. Since in general $\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2 \cdot \mathrm{Cov}[X,Y]$ and X_n are pairwise uncorrelated, $\mathrm{Var}[X_n + X_m] = \mathrm{Var}[X_n] + \mathrm{Var}[X_m]$ for $n \neq m$ and by induction $\mathrm{Var}[S_n] = \sum_{i=1}^n \mathrm{Var}[X_i]$. Using linearity, we obtain $E[S_n/n] = m$ and

$$\mathrm{Var}[S_n/n] = E[(S_n)^2/n^2] - E[S_n]^2/n^2 = \mathrm{Var}[S_n]/n^2 = \frac{1}{n^2} \sum_{i=1}^n \mathrm{Var}[X_n] \to 0 \ .$$

With Chebychev's we obtain

$$P[|S_n/n - m| \geq \epsilon] \leq \frac{\mathrm{Var}[S_n/n]}{\epsilon^2} \to 0 \ (n \to \infty) \ .$$

SPECIAL CASE. If X_i are independent random variables with the same distribution for which the mean m and variance exists, then $\sum_{i=1}^n X_i/n \to m$ in probability.

ALMOST EVERYWHERE CONVERGENCE: X_n converges **almost everywhere** to a random variable X, if $\mathrm{P}[X_n \to X, n \to \infty] = 1$.

STRONGER WEAK LAW OF LARGE NUMBERS: Assume X_i have common expectation $E[X_i] = m$ and satisfy $M = \sup_n \mathrm{E}[X^4] < \infty, \sup_n \mathrm{E}[X^2]^2 < \infty$. If X_i are independent, then $\sum_n \mathrm{P}[|S_n/n - m| \geq \epsilon]$ converges for all $\epsilon > 0$.
Proof. Estimation of $\mathrm{E}[X_n^4]$ with Chebychev-Markov's gives $\mathrm{P}[|S_n/n - m| \geq \epsilon] \leq C/n^2$ for some constant C.

STRONG LAW OF LARGE NUMBERS: If X_n are independent random variables with $M = \sup_n \mathrm{E}[X_n^4] < \infty$, $\sup_n \mathrm{E}[X_n^2] < \infty$ with common expectation $\mathrm{E}[X_n] = m$. Then $S_n/n \to m$ almost everywhere.
Proof. Follows from the stronger weak law because complete convergence implies almost everywhere convergence.

CONVERGENCE IN DISTRIBUTION. X_n converges **in distribution** to a random variable X, if for all $t \in \mathbf{R}$, $\mathrm{P}[X_n \leq t] \to \mathrm{P}[X \leq t]$ for $n \to \infty$.

CENTRAL LIMIT THEOREM:
Given X_n which are independent with mean m and variance σ^2. Let X be a random variable with standard normal distribution. Then

$$\frac{S_n - nm}{\sigma\sqrt{n}} \to X$$

in distribution, where $S_n = X_1 + X_2 + \ldots + X_n$.
Proof. A calculation shows that the characteristic functions of $S_n^* = (S_n - E[S_n])/\sigma[S_n]$ converge to the characteristic function of X.