PART A: INDEPENDENCE

Contents

1 Long independent sequences
  1a Independent events ................................... 1
  1b Measure-theoretic foundations .......................... 10
  1c Independent random variables .......................... 18

1 Long independent sequences

1a Independent events

1a1 Reminder. Fair coin: finite probability space \( \Omega = \{0,1\}^n \) with 
\[ p(\omega) = p_n(\omega) = 2^{-n} \] for all \( \omega \in \Omega \); the number of “heads” — random 
variable \( H = H_n : \Omega_n \to \mathbb{R}, H(\omega) = a_1 + \cdots + a_n \) for \( \omega = (a_1, \ldots, a_n) \in \Omega \); 
its distribution
\[
\mathbb{P}(H = k) = P_n(\{\omega : H(\omega) = k\}) = \sum_{\omega : H(\omega) = k} p(\omega).
\]

1a2 Reminder. Binomial distribution (the fair case):
\[
\mathbb{P}(H = k) = 2^{-n} \binom{n}{k}, \quad \binom{n}{k} = \frac{n!}{k!(n-k)!}; \quad H_n \sim \text{Binom}(n, 0.5).
\]

1a3 Reminder. Random signs: random variables \( X_1, \ldots, X_n : \Omega_n \to \mathbb{R}, \)
\( X_k(\omega) = 2a_k - 1 \) for \( \omega = (a_1, \ldots, a_n) \in \Omega \).

Simple random walk: random variables \( S_0, \ldots, S_n : \Omega \to \mathbb{R}, S_k = X_1 + \cdots + X_k. \)

1a4 Remark.
\[
S_n = 2H_n - n;
\]
\[
\mathbb{P}(S_n = k) = \mathbb{P}\left(H_n = \frac{n+k}{2}\right) = \frac{n!}{2^n \binom{n-k}{k+k}^n} \quad \text{for } k = -n, -n + 2, \ldots, n.
\]
1a5 Proposition. For every $\varepsilon > 0$,

$$P\left(\frac{1}{n}|S_n| \leq \varepsilon\right) \to 1 \quad \text{as } n \to \infty.$$ 

Thus, $P\left(\left|\frac{1}{n}H_n - \frac{1}{2}\right| \leq \varepsilon\right) \to 1$ as $n \to \infty$. The frequency is close to the probability...

This is a special case of the Weak Law of Large Numbers, see [1a1]. See also [1a24].

How to prove 1a5? Cumbersome sums of binomial coefficients? No, this is the old way. The newer way: via Pythagorean theorem in the ($2^n$-dimensional) Euclidean space of random variables!

The Euclidean space $L_2(\Omega) = L_2(\Omega_n, P_n)$ consists of all functions $X : \Omega \to \mathbb{R}$ and is endowed with the norm and scalar product

$$\|X\| = \sqrt{\sum_{\omega \in \Omega} |X(\omega)|^2 p(\omega)} = \sqrt{\langle X, X \rangle},$$

$$(1a6) \quad \langle X, Y \rangle = \sum_{\omega \in \Omega} X(\omega)Y(\omega)p(\omega).$$

Its dimension is equal to the number of points in $\Omega$ (think, why; any restriction on $p(\cdot)$?).

Recall the expectation $\mathbb{E}X$ of a random variable $X$:

$$(1a7) \quad \mathbb{E}X = \sum_{\omega \in \Omega} X(\omega)p(\omega) = \langle X, 1 \rangle,$$

an important linear functional on $L_2(\Omega)$. Note that

$$(1a8) \quad \|X\|^2 = \mathbb{E}X^2, \quad \langle X, Y \rangle = \mathbb{E}XY. \quad \text{(that is, } \mathbb{E}(X^2))$$

1a9 Exercise. The random signs $X_1, \ldots, X_n$ are orthonormal, that is,

$$\mathbb{E}X_iX_j = \begin{cases} 1 & \text{for } i = j, \\ 0 & \text{otherwise.} \end{cases}$$

Prove it. Are they a basis?

1a10 Exercise. $\|S_k\| = \sqrt{k}$ for $k = 0, 1, \ldots, n$.

Prove it.

$^{1}$[KS, Sect. 2.1, Th. 2.5].
1a11 Exercise. $\mathbb{P}( |X| \geq \varepsilon ) \leq \left( \frac{1}{\varepsilon} \|X\| \right)^2$ for all $\varepsilon > 0$ and $X \in L_2(\Omega)$.
Prove it.

1a12 Exercise. Prove Proposition 1a5.

Here is the normal approximation to the binomial distribution.

1a13 Proposition. ¹

$$\mathbb{P}( S_n = k ) = \frac{2}{\sqrt{2\pi n}} \exp \left( -\frac{k^2}{2n} \right) \cdot \left( 1 + \alpha_n \left( \frac{k}{\sqrt{n}} \right) \right) \quad \text{for } k + n \text{ even},$$

where $\alpha_n(\cdot) \to 0$ uniformly on bounded intervals.

(Clearly, $\mathbb{P}( S_n = k )$ vanishes if $n + k$ is odd.)

1a14 Remark. The convergence $\alpha_n(\cdot) \to 0$ cannot be uniform on $\mathbb{R}$, since $\alpha_n(\sqrt{n} + \frac{1}{\sqrt{n}}) = -1$ (think, why). What about $\alpha_n(\sqrt{n})$? Well, it is $-1 + \frac{1}{2} \sqrt{2\pi n} e^{n/2} 2^{-n} \to -1$ (think, why).

How to prove Prop. 1a13? Some calculations with binomial coefficients (but not their sums...) are needed.

1a15 Reminder.

$$n! = n^n e^{-n} \sqrt{2\pi n} \beta(n), \quad \beta(n) \to 1; \quad \beta(n) = 1 + O(1/n). \quad \text{(Stirling)}$$

Thus,

$$\ln n! = n \ln n - n + \frac{1}{2} \ln n + \frac{1}{2} \ln(2\pi) + \ln \beta(n) + O(1/n). \quad \text{(1a16)}$$

1a17 Exercise.

$$\ln \mathbb{P}( S_n = cn ) = - \frac{1}{2} \ln n - \frac{n}{2} \left( (1 - c) \ln (1 - c) + (1 + c) \ln (1 + c) \right) - \frac{1}{2} \ln (1 - c^2) + \ln 2 - \frac{1}{2} \ln (2\pi) + \ln \beta(n) - \ln \beta \left( \frac{n(1 - c)}{2} \right) - \ln \beta \left( \frac{n(1 + c)}{2} \right)$$

whenever $cn \in \{-n, -n + 2, \ldots, n\}$.

Prove it. (Combine 1a14 with 1a15 and enjoy many cancellations!)

¹[KS, Sect. 2.2, Th. 2.10]; [D, Sect. 2.1, Th. (14)].
We introduce a function $\gamma : [-1, 1] \to \mathbb{R}$ by

\begin{equation}
\gamma(c) = \frac{1}{2}(1 + c) \ln(1 + c) + \frac{1}{2}(1 - c) \ln(1 - c) \quad \text{for } c \in (-1, 1),
\end{equation}

$\gamma(-1) = \gamma(+1) = \ln 2$.

Now 1a17 becomes

$$\ln \mathbb{P}(S_n = cn) = -n\gamma(c) - \frac{1}{2} \ln n - \frac{1}{2} \ln(1 - c^2) + \ln 2 - \frac{1}{2} \ln(2\pi) + o(1),$$

if $n(1 \pm c) \gg 1$; moreover,

\begin{equation}
\mathbb{P}(S_n = cn) = \frac{2}{\sqrt{n}} \frac{1}{\sqrt{1 - c^2}} \frac{1}{\sqrt{2\pi}} e^{-n\gamma(c)} \cdot \left(1 + O\left(\frac{1}{n(1 - |c|)}\right)\right).
\end{equation}

A numerical example: $n = 200, c = -0.9, \mathbb{P}(S_{200} = -180) = ?$ Really $1.397 \cdot 10^{-44}$; approximately (as above) $1.409 \cdot 10^{-44}$; by normal approximation: $3.7 \cdot 10^{-37}$ (oops...).

About the function $\gamma$:

- $\gamma(-c) = \gamma(c)$;
- $\gamma'(0) = 0$;
- $\gamma''(0) = 1$ since $(x \ln x)' = 1 + \ln x, \quad (x \ln x)'' = \frac{1}{x}$;
- $\gamma'''(0) = 0$; thus $\gamma(c) = \frac{1}{2}c^2 + O(c^4)$ as $c \to 0$.

Proof of 1a13

$$\left| \ln \mathbb{P}(S_n = k) - \ln \left(\frac{2}{\sqrt{2\pi n}} \exp\left(-\frac{k^2}{2n}\right)\right) \right| \leq \left( k = cn \right)$$

$$\leq \left| \ln \mathbb{P}(S_n = cn) - \ln \left(\frac{2}{\sqrt{n}} \frac{1}{\sqrt{1 - c^2}} \frac{1}{\sqrt{2\pi}} e^{-n\gamma(c)}\right) \right| +$$

$$+ \left| \ln \left(\frac{2}{\sqrt{n}} \frac{1}{\sqrt{1 - c^2}} \frac{1}{\sqrt{2\pi}} e^{-n\gamma(c)}\right) - \ln \left(\frac{2}{\sqrt{2\pi n}} e^{-nc^2/2}\right) \right| \leq$$

$$\leq O\left(\frac{1}{n(1 - |c|)}\right) + \ln \frac{1}{\sqrt{1 - c^2}} + n\gamma(c) - nc^2/2 + O(\gamma(c) - nc^2/2).$$
Claim:

\[ O \left( \frac{1}{n(1-|c|)} \right) + O(c^2) + O(nc^4) = O \left( nc^4 + \frac{1}{n} \right) . \]

Proof of the claim. If \( nc^4 + \frac{1}{n} \leq \delta \leq \frac{1}{4} \) then: \( nc^4 \leq \delta; \frac{1}{n} \leq \delta; c^4 \leq \frac{1}{n}\delta \leq \delta^2; \)

Thus, \( \Pr(S_n = k) = \frac{2}{\sqrt{2\pi n}} e^{-\frac{k^2}{2n}} \left( 1 + O \left( \frac{k^4}{n^3} + \frac{1}{n} \right) \right) . \)

If \( |k| = O(\sqrt{n}) \) then \( \frac{k^4}{n^3} = O \left( \frac{n^2}{n^3} \right) = O \left( \frac{1}{n} \right) . \) Thus, \( \sup_{[a,b]} |\alpha(\cdot)| = O(1/n) \) for all \( a, b. \)

End of proof of 1a13

Proposition 1a13 is a special case of the Local Limit Theorem. In contrast, the next result is global.

**1a20 Theorem.** 1

\[ \Pr \left( a\sqrt{n} < S_n < b \sqrt{n} \right) \to \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} \, dx \quad \text{as } n \to \infty \]

whenever \( -\infty \leq a \leq b \leq \infty. \)

This is the De Moivre-Laplace theorem, a special case of the Central Limit Theorem.

Proof of 1a20

First, assume that \( -\infty < a < b < \infty. \) Then:

\[ \Pr \left( a\sqrt{n} < S_n < b \sqrt{n} \right) = \sum_{k \in (a\sqrt{n}, b\sqrt{n}), k+n \text{ even}} \Pr(S_n = k) ; \]

let \( \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} ; \)

\[ \Pr \left( S_n = k \right) = \frac{2}{\sqrt{n}} \varphi \left( \frac{k}{\sqrt{n}} \right) \left( 1 + \alpha_n \left( \frac{k}{\sqrt{n}} \right) \right) ; \]

\[ ^1[\text{KS, Sect. 2.2}]; [\text{D, Sect. 2.1, (1.5)}]. \]
\[
\frac{2}{\sqrt{n}} \varphi \left( \frac{k}{\sqrt{n}} \right) \left( 1 + \inf_{[a,b]} \alpha_n \right) \leq \mathbb{P} \left( S_n = k \right) \leq \frac{2}{\sqrt{n}} \varphi \left( \frac{k}{\sqrt{n}} \right) \left( 1 + \sup_{[a,b]} \alpha_n \right);
\]
\[
\left( 1 + \inf_{[a,b]} \alpha_n \right) \sum_{k \in (a \sqrt{n}, b \sqrt{n}), k + n \text{ even}} \frac{2}{\sqrt{n}} \varphi \left( \frac{k}{\sqrt{n}} \right) \leq \mathbb{P} \left( a \sqrt{n} < S_n < b \sqrt{n} \right) \leq \left( 1 + \sup_{[a,b]} \alpha_n \right) \sum_{k \in (a \sqrt{n}, b \sqrt{n}), k + n \text{ even}} \frac{2}{\sqrt{n}} \varphi \left( \frac{k}{\sqrt{n}} \right).
\]

It remains to prove that the sum converges to the integral. We divide \([a, b]\) into intervals of length \(2/\sqrt{n}\) and get an integral sum; only the first and last terms differ, but contribute only \(O(1/\sqrt{n})\) anyway.

The case \(-\infty < a < b < \infty\) is done. The case \(a = b\) is trivial. It is sufficient (think, why) to consider the case \(-\infty < a < b = \infty\). We could do it via the equality
\[
\int_{-\infty}^{\infty} e^{-x^2/2} dx = \sqrt{2\pi} \quad \text{well-known in analysis, but it is instructive to do it differently, and get the integral equality as a by-product.}\]

The argument introduced below will be reused in the proof of \([\text{a21}]\).

We note that
\[
\lim \inf_{n \to \infty} \mathbb{P} \left( a \sqrt{n} < S_n \right) \geq \frac{1}{\sqrt{2\pi}} \int_a^{\infty} e^{-x^2/2} dx,
\]
since it exceeds \(\lim \inf_{n \to \infty} \mathbb{P} \left( a \sqrt{n} < S_n < b \sqrt{n} \right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx\) for every \(b \in (a, \infty)\). It remains to prove that
\[
\lim \sup_{n \to \infty} \mathbb{P} \left( a \sqrt{n} < S_n \right) \leq \frac{1}{\sqrt{2\pi}} \int_a^{\infty} e^{-x^2/2} dx.
\]
We have
\[
\mathbb{P} \left( a \sqrt{n} < S_n \right) = \mathbb{P} \left( S_n = k_n \right) + \mathbb{P} \left( S_n = k_n + 2 \right) + \ldots
\]
where \(k_n = \min(\mathbb{Z} \cap (a \sqrt{n}, \infty))\); and
\[
\mathbb{P} \left( S_n = k + 2 \right) = 2^{-n} \frac{n!}{\left( \frac{n-k-2}{2} \right)! \left( \frac{n+k+2}{2} \right)!} \leq\]
\[
\leq \frac{1 - \frac{k}{n}}{1 + \frac{k}{n}} \mathbb{P} \left( S_n = k \right) \leq \frac{1 - a/\sqrt{n}}{1 + a/\sqrt{n}} \mathbb{P} \left( S_n = k \right)
\]

The doubt is: maybe a part \(1 - \frac{1}{\sqrt{2\pi}} \int_a^{\infty} e^{-x^2/2} dx > 0\) of the distribution of \(S_n\) escapes to infinity when \(n \to \infty\)?
for \(k = k_n, k_n + 2, \ldots\); therefore

\[
\frac{\mathbb{P}(S_n > a\sqrt{n})}{\mathbb{P}(S_n = k_n)} \leq 1 + \frac{1 - a/\sqrt{n}}{1 + a/\sqrt{n}} + \left(\frac{1 - a/\sqrt{n}}{1 + a/\sqrt{n}}\right)^2 + \cdots = \frac{1 + a/\sqrt{n}}{2a/\sqrt{n}} = \frac{\sqrt{n}}{2a} + \frac{1}{2},
\]

and we get

\[
\limsup_{n \to \infty} \mathbb{P}(a\sqrt{n} < S_n) \leq \limsup_{n \to \infty} \frac{\sqrt{n}}{2a} \mathbb{P}(S_n = k_n) = \frac{1}{a} \sqrt{\frac{2\pi}{e}} e^{-a^2/2}
\]

by \text{[1a13]} since \(k_n/\sqrt{n} \to a\). It follows that

\[
\limsup_{n \to \infty} \mathbb{P}(a\sqrt{n} < S_n) \leq \limsup_{n \to \infty} \mathbb{P}(a\sqrt{n} < S_n < b\sqrt{n}) + \limsup_{n \to \infty} \mathbb{P}(b\sqrt{n} \leq S_n) \leq \frac{1}{\sqrt{2\pi}} \int_a^b \sqrt{\frac{x}{n}} e^{-x^2/2} \, dx + \frac{1}{b} \sqrt{\frac{2\pi}{e}} e^{-b^2/2}
\]

for every \(b \in (a, \infty)\); we take \(b \to \infty\).

End of proof of \text{[1a20]}

\textbf{1a21 Proposition.} \(^1\) For every \(c \in (0, 1)\),

\[
\frac{1}{n} \ln \mathbb{P}(S_n > cn) \to -\gamma(c) \quad \text{as } n \to \infty,
\]

where \(\gamma(c) = \frac{1}{2}(1 + c) \ln(1 + c) + \frac{1}{2}(1 - c) \ln(1 - c)\).

This is a special case of the Large Deviations Principle.\(^2\)

Proposition \text{[1a21]} suggests the approximation (for large \(c\) and \(n\))

\[
\mathbb{P}(S_n > c) \approx e^{-n\gamma(c/n)} = \frac{n^n}{\sqrt{(n-c)^{n-c}(n+c)^{n+c}}}.
\]

However, Theorem \text{[1a20]} suggests another approximation,

\[
\mathbb{P}(S_n > c) \approx \frac{1}{\sqrt{2\pi}} \int_{c\sqrt{n}}^\infty e^{-x^2/2} \, dx \approx \exp \left( -\frac{c^2}{2n} \right).
\]

A paradox! What do you think? A clue: for \(n = 200\),

\[
\begin{array}{cccccccc}
c & 0 & 30 & 60 & 90 & 120 & 150 & 180 \\
2^{-\frac{n}{(n+c)/2}} & 6 \cdot 10^{-2} & 6 \cdot 10^{-3} & 6 \cdot 10^{-6} & 5 \cdot 10^{-11} & 1 \cdot 10^{-18} & 3 \cdot 10^{-29} & 1 \cdot 10^{-44} \\
\frac{2}{\sqrt{2\pi n}} \exp(-\frac{c^2}{2n}) & 6 \cdot 10^{-2} & 6 \cdot 10^{-3} & 7 \cdot 10^{-6} & 9 \cdot 10^{-11} & 1 \cdot 10^{-17} & 2 \cdot 10^{-26} & 4 \cdot 10^{-37} \\
\exp(-n\gamma(c/n)) & 1 & 1 \cdot 10^{-1} & 1 \cdot 10^{-4} & 8 \cdot 10^{-10} & 2 \cdot 10^{-17} & 3 \cdot 10^{-28} & 1 \cdot 10^{-43} \\
\end{array}
\]

\(^1\)[D, Sect. 2.1, Exercise 1.3].

\(^2\)[KS, Sect. 10.4]; [D, Sect. 1.9].
Proof of 1a21

We reuse the argument of the last part of the proof of 1a20: \( P(S_n > cn) = P(S_n = k_n) + P(S_n = k_n + 2) + \ldots \) where \( k_n = \min(\mathbb{Z} \cap (cn, \infty)) \);

\[
P(S_n = k + 2) \leq \frac{1-c}{1+c} P(S_n = k)
\]

for \( k = k_n, k_n + 2, \ldots \); therefore

\[
1 \leq \frac{P(S_n > cn)}{P(S_n = k_n)} \leq 1 + \frac{1-c}{1+c} + \left( \frac{1-c}{1+c} \right)^2 + \ldots = \frac{1+c}{2c},
\]

and we get \( \frac{1}{n} \ln P(S_n > cn) - \frac{1}{n} \ln P(S_n = k_n) \to 0 \). By (1a19),

\[
\frac{1}{n} \ln P(S_n = k_n) = \frac{1}{n} \ln \left( \frac{2}{\sqrt{n}} \frac{1}{\sqrt{1-c^2}} \frac{1}{\sqrt{2\pi}} \right) = \gamma \left( \frac{k_n}{n} \right) + o(1) \to \gamma(c),
\]

since \( \frac{k_n}{n} \to c \).

End of proof of 1a21

1a22 Reminder. Unfair coin: The same \( \Omega \) and \( H \) as in 1a1, but different probabilities

\[
p_n(a_1, \ldots, a_n) = p^{a_1+\ldots+a_n}(1-p)^{n-(a_1+\ldots+a_n)} = \prod_{k=1}^{n} p(a_k) \quad \text{for } a_1, \ldots, a_n \in \{0, 1\}.
\]

It is convenient to write \( H = H_{n,p} \); this function on \( \Omega \) does not depend on \( p \), but its distribution depends on \( p \).

1a23 Reminder. Binomial distribution:

\[
P(H_{n,p} = k) = \binom{n}{k} p^k (1-p)^{n-k}; \quad H_n \sim \text{Binom}(n,p).
\]

1a24 Proposition. \(^1\) For every \( p \in [0,1] \) and \( \varepsilon > 0 \),

\[
P\left( \left| \frac{1}{n} H_{n,p} - p \right| \leq \varepsilon \right) \to 1 \quad \text{as } n \to \infty.
\]

\(^1\) [KS, Sect. 2.1, Th. 2.5].
This is more general than (the corollary of) 1a5, but still, a special case of the Weak Law of Large Numbers, see 1c1.

It is easy to prove 1a24 similarly to 1a5, but anyway, 1a24 will follow from 1c1.

Rather unexpectedly, 1a24 can be used for proving Weierstrass’s approximation theorem: polynomials are dense in $C[0,1]$.\footnote{[KS, Sect. 2.1, Th. 2.7]; [D, Sect. 1.5, Example 5.1].}

Here is the idea of the probabilistic proof of Weierstrass’s approximation theorem. Consider the distribution $\mu_{n,p}$ of $\frac{1}{n}H_{n,p}$,

$$\mu_{n,p} = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \delta_{k/n};$$

it belongs to the $(n+1)$-dimensional linear space of signed measures on $\{0, \frac{1}{n}, \frac{2}{n}, \ldots, 1\}$, and the vector-function $p \mapsto \mu_{n,p}$ is polynomial (of degree $n$). By 1a24, $\mu_{n,p}$ is close to $\delta_p$ (the unit mass at $p$). Thus, the map $p \mapsto \delta_p$ is approximately polynomial! Now, given a continuous function $f : [0, 1] \to \mathbb{R}$ we have $f(p) = \int f \; d\delta_p \approx \int f \; d\mu_{n,p} = P_n(p)$, $P_n$ being a polynomial. Namely,

$$|f(p) - \mathbb{E} f\left(\frac{1}{n}H_{n,p}\right)| \leq \mathbb{E} \left| f\left(\frac{1}{n}H_{n,p}\right) - f(p) \right| \leq \max_{[p-\varepsilon, p+\varepsilon]} |f(\cdot) - f(p)| + \mathbb{P}\left(\left|\frac{1}{n}H_{n,p} - p\right| > \varepsilon\right) \cdot 2 \max_{[0,1]} |f(\cdot)|;$$

the former summand is made small using uniform continuity of $f$, the latter summand — using $\mathbb{E}\left|\frac{1}{n}H_{n,p} - p\right|^2 = \frac{p(1-p)}{n} \leq \frac{1}{4\varepsilon^2}$.

1a25 Proposition.\footnote{[KS, Sect. 2.3]; [D, Sect. 2.6(a)].} For every $\lambda \in (0, \infty)$ and $k = 0, 1, 2, \ldots$

$$\mathbb{P}\left(H_{n,\lambda/n} = k\right) \rightarrow \frac{\lambda^k}{k!} e^{-\lambda} \quad \text{as } n \rightarrow \infty.$$

This is the Poisson Limit Theorem.

Proof.

$$\mathbb{P}\left(H_{n,\lambda/n} = k\right) = \frac{n!}{k!(n-k)!} \binom{\lambda}{n}^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = \binom{n}{n-k} \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^{n-k} \rightarrow \lambda^k \frac{\lambda^k}{k!} e^{-\lambda}.$$

$\square$
1b Measure-theoretic foundations

Some measure theory

1b1 Reminder. An algebra of sets (on $X$): a set $E$ of subsets of $X$ such that

$$\emptyset, X \in E; \quad \forall E \in E \quad X \setminus E \in E; \quad \forall E, F \in E \quad E \cap F, E \cup F \in E.$$ 

A σ-algebra of sets (on $X$): an algebra $A$ such that

$$\forall A_1, A_2, \cdots \in A \quad \bigcap_n A_n, \bigcup_n A_n \in A.$$ 

A measurable space: $(X, A)$. 

A probability measure (on $A$, or on $(X, A)$): a map $\mu : A \to [0,1]$ such that $\mu(X) = 1$ and

$$\mu\left(\bigcup_n A_n\right) = \sum_n \mu(A_n)$$

whenever $A_1, A_2, \cdots \in A$ are pairwise disjoint. (In such case we may write $\uplus_n A_n$.) 

A probability space: $(X, A, \mu)$.

1b2 Reminder. A box (in $\mathbb{R}^d$): a set of the form $I_1 \times \cdots \times I_d$ where $I_1, \ldots, I_d \subset \mathbb{R}$ are bounded intervals (open, closed, or neither). 

An elementary set (in $\mathbb{R}^d$): a finite union of boxes. 

The elementary algebra (on $\mathbb{R}^d$): the algebra generated by all boxes; consists of all elementary sets and their complements (“co-elementary sets”). 

The Borel σ-algebra $\mathcal{B}(\mathbb{R}^d)$: the σ-algebra generated by all boxes, or equivalently, by all open sets.

1b3 Theorem. 1 (Hahn-Kolmogorov) Let $E$ be an algebra on $X$, $A = \sigma(E)$ the σ-algebra generated by $E$, and $\mu_0 : E \to [0,1]$ a map. Then the following are equivalent:

(a) there exists one and only one probability measure $\mu$ on $A$ such that $\mu|_{E} = \mu_0$;

(b) there exists at least one such $\mu$;

(c) $\mu_0(X) = 1$, and $\mu_0\left(\bigcup_n E_n\right) = \sum_n \mu_0(E_n)$ whenever $E_1, E_2, \cdots \in E$ are pairwise disjoint and $\bigcup_n E_n \in E$.

---

1[Tao, Th. 1.7.8].
Clearly, \((a)\Rightarrow (b)\Rightarrow (c)\). In order to prove\(^1\) \((c)\Rightarrow (a)\) we assume \((c)\); define the outer measure
\[
\mu^*(Z) = \inf \left\{ \sum_n \mu_0(E_n) : E_1, E_2, \cdots \in \mathcal{E}, \cup_n E_n \supset Z \right\}
\]
for arbitrary \(Z \subset X\); and call a set \(A \subset X\) \(\mu\)-measurable if
\[
\inf_{E \in \mathcal{E}} \mu^*(A \triangle E) = 0.
\]

\[1\text{b}4\] Exercise. \(\mu^*(\cup_n Z_n) \leq \sum_n \mu^*(Z_n)\) for arbitrary \(Z_1, Z_2, \cdots \subset X\).
Prove it. (Do you need \((c)\)?)

\[1\text{b}5\] Exercise. \(\mu^*(E) = \mu_0(E)\) for all \(E \in \mathcal{E}\).
Prove it. (Do you need \((c)\)?)

\[1\text{b}6\] Exercise. \(\mu\)-measurable sets are a \(\sigma\)-algebra.
Prove it.

Taking into account that all sets of \(\mathcal{E}\) are \(\mu\)-measurable we conclude that all sets of \(\mathcal{A} = \sigma(\mathcal{E})\) are \(\mu\)-measurable.

We define \(\mu\) as the restriction of \(\mu^*\) to \(\mathcal{A}\).

\[1\text{b}7\] Exercise. \(\mu^*(A \cup B) + \mu^*(A \cap B) = \mu^*(A) + \mu^*(B)\) for all \(\mu\)-measurable \(A, B\).
Prove it.
Hint: \(\mu_0(E \cup F) + \mu_0(E \cap F) = \mu_0(E) + \mu_0(F)\) for all \(E, F \in \mathcal{E}\).

Thus, \(\mu(A \cup B) = \mu(A) + \mu(B)\).

\[1\text{b}8\] Exercise. \(|\mu^*(Z) - \mu^*(W)| \leq \mu^*(Z \triangle W)\) for arbitrary \(Z, W \subset X\).
Prove it.

\[1\text{b}9\] Exercise. \(\mu^*(\psi_n A) = \sum_n \mu^*(A_n)\) for \(\mu\)-measurable \(A_n\).
Prove it.
Hint: \(\mu^*(\psi^\infty_{n=1} A_n) - \mu^*(\psi^N_{n=1} A_n) \leq \mu^*(\psi^\infty_{n=N+1} A_n)\).

Thus, \(\mu\) is a probability measure, which completes the proof of existence.
Here is uniqueness. Let \(\mu_1\) be another such measure. Then \(\mu_1(A) \leq \mu^*(A) = \mu(A)\) for all \(A \in \mathcal{A}\) (since \(A \subset \cup_n E_n\) implies \(\mu_1(A) \leq \sum_n \mu_1(E_n) = \sum_n \mu(E_n)\)). The same holds for \(X \setminus A\), thus, \(\mu_1(A) = 1 - \mu_1(X \setminus A) \geq 1 - \mu(X \setminus A) = \mu(A)\) and finally \(\mu_1(A) = \mu(A)\).

Theorem \[1\text{b}3\] is proved.

1b10 Corollary. If two probability measures are equal on an algebra $\mathcal{E}$ then they are equal on the generated $\sigma$-algebra $\sigma(\mathcal{E})$.

1b11 Corollary. A probability measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ is uniquely determined by its values on boxes. The same holds for closed boxes, and for open boxes.

1b12 Exercise. A probability measure $\mu$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is uniquely determined by its cumulative distribution function (CDF)

$$F_\mu(x) = \mu\left((\infty, x]\right) \text{ for } x \in \mathbb{R}.$$  

Prove it.

1b13 Exercise. (Lebesgue-Stieltjes measure) Let $F : \mathbb{R} \to [0, 1]$ be an increasing function, $F(-\infty) = 0$, $F(+\infty) = 1$. Then

(a) there exists one and only one additive function $\mu_0$ on the elementary algebra $\mathcal{E}$ (of $\mathbb{R}$) such that

$$\mu_0((a, b)) = F(b-) - F(a+) \text{ for } -\infty < a < b < \infty;$$
$$\mu_0(\{a\}) = F(a+) - F(a-) \text{ for } -\infty < a < \infty;$$

(b) for every $E \in \mathcal{E}$ and $\varepsilon > 0$ there exists a compact elementary set $K \subset E$ such that $\mu_0(K) \geq \mu_0(E) - \varepsilon$;

(c) $\mu_0$ satisfies Condition 1b3(c).

Prove it.

1b14 Exercise. The correspondence $\mu \mapsto F_\mu$ is a bijective correspondence between all probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and all increasing functions $F : \mathbb{R} \to [0, 1]$ such that $F(-\infty) = 0$, $F(+\infty) = 1$ and $\forall x \in \mathbb{R} \ F(x+) = F(x)$.

Prove it.

In fact, 1b12 generalizes readily to $\mathbb{R}^d$;

$$F_\mu(x_1, \ldots, x_d) = \mu\left((\infty, x_1] \times \cdots \times (\infty, x_d]\right);$$

unfortunately, 1b14 does not.

Discrete measures:

$$\mu = \sum_k p_k \delta_{x_k}; \quad \mu(B) = \sum_{k: x_k \in B} p_k = \sum_{x \in B} p(x); \quad \int f \, d\mu = \sum f(x_k)p_k.$$
Absolutely continuous measures:

\[\mu = \int p(x)\delta_x \, dx; \quad \mu(B) = \int_B p(x) \, dx; \quad \int f \, d\mu = \int f(x)p(x) \, dx.\]

Singular measures: nonatomic but concentrated on a set of zero Lebesgue measure.

The product of two measurable spaces \((X_1, \mathcal{A}_1)\) and \((X_2, \mathcal{A}_2)\) is defined to be \((X_1 \times X_2, \mathcal{A}_1 \times \mathcal{A}_2)\) where \(\mathcal{A}_1 \times \mathcal{A}_2\) is the \(\sigma\)-algebra generated by \(A_1 \times A_2\) for all \(A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2\).

1b15 Exercise. (a) A probability measure \(\mu\) on \((X_1 \times X_2, \mathcal{A}_1 \times \mathcal{A}_2)\) is uniquely determined by \(\mu(A_1 \times A_2)\) for all \(A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2\).

(b) The same holds for \(\mu(E_1 \times E_2)\) for all \(E_1 \in \mathcal{E}_1, E_2 \in \mathcal{E}_2\) provided that an algebra \(\mathcal{E}_1\) generates \(A_1\), and an algebra \(\mathcal{E}_2\) generates \(A_2\).

Prove it.

Hint: (b) the \(\sigma\)-algebra generated by \(E_1 \times E_2\) contains \(E_1 \times A_2\).

In particular, \(\mathcal{B}(\mathbb{R}^{d_1}) \times \mathcal{B}(\mathbb{R}^{d_2}) = \mathcal{B}(\mathbb{R}^{d_1+d_2})\).

1b16 Reminder. Let \((X_1, \mathcal{A}_1, \mu_1), (X_2, \mathcal{A}_2, \mu_2)\) be two probability spaces. The formula

\[\mu(A) = \int_{X_1} \left( \int_{X_2} 1_A(x_1, x_2) \mu_2(dx_2) \right) \mu_1(dx_1)\]

defines correctly a measure \(\mu\) on \((X_1 \times X_2, \mathcal{A}_1 \times \mathcal{A}_2)\). Clearly, \(\mu\) satisfies

\[(1b17) \quad \mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2) \quad \text{for} \quad A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2.\]

By 1b15, \(\mu\) is the only measure satisfying (1b17). It follows that \(\int_{X_1} \int_{X_2} \cdots = \int_{X_2} \int_{X_1} \cdots\). We write \(\mu = \mu_1 \times \mu_2\), say that \(\mu\) is the product measure, and \((X_1 \times X_2, \mathcal{A}_1 \times \mathcal{A}_2, \mu_1 \times \mu_2) = (X_1, \mathcal{A}_1, \mu_1) \times (X_2, \mathcal{A}_2, \mu_2)\) is the product of probability spaces.

1b18 Reminder. Let \((X, \mathcal{A}, \mu) = (X_1, \mathcal{A}_1, \mu_1) \times (X_2, \mathcal{A}_2, \mu_2)\).

(a) (Tonelli)

\[\int_X f \, d\mu = \int_{X_1} \mu_1(dx_1) \int_{X_2} \mu_2(dx_2) f(x_1, x_2) \in [0, +\infty]\]

for every measurable \(f : X \to [0, \infty]\).

(b) (Fubini)

\[\int_X f \, d\mu = \int_{X_1} \mu_1(dx_1) \int_{X_2} \mu_2(dx_2) f(x_1, x_2) \in \mathbb{R}\]

for every integrable \(f : X \to \mathbb{R}\).
In particular,

\[(1b19) \quad \int_X f_1(x_1)f_2(x_2) \, \mu(dx_1dx_2) = \left( \int_{X_1} f_1 \, d\mu_1 \right) \left( \int_{X_2} f_2 \, d\mu_2 \right).\]

The same holds for the product of three, four, \ldots probability spaces.

**1b20 Remark.** 
Associativity of the multiplication: the space \((X_1, \mathcal{A}_1, \mu_1) \times (X_2, \mathcal{A}_2, \mu_2) \times (X_3, \mathcal{A}_3, \mu_3) \times (X_4, \mathcal{A}_4, \mu_4)\) is the same as \(((X_1, \mathcal{A}_1, \mu_1) \times (X_2, \mathcal{A}_2, \mu_2)) \times ((X_3, \mathcal{A}_3, \mu_3) \times (X_4, \mathcal{A}_4, \mu_4))\). That is, \((\mu_1 \times \mu_2 \times \mu_3 \times \mu_4) = \mu_1 \times \mu_2 \times \mu_3 \times \mu_4\), which follows from the uniqueness; both measures satisfy \(\mu(A_1 \times A_2 \times A_3 \times A_4) = \mu_1(A_1)\mu_2(A_2)\mu_3(A_3)\mu_4(A_4)\) (think, why). The same holds for more than two factors in a group, and more than two groups.

**1b21 Reminder.** 
A measurable map from a measurable space \((X, \mathcal{A})\) to a measurable space \((Y, \mathcal{B})\) is \(f : X \to Y\) such that \(\forall B \in \mathcal{B} \ f^{-1}(B) \in \mathcal{A}\).

The composition of measurable maps, \((X, \mathcal{A}) \to (Y, \mathcal{B}) \to (Z, \mathcal{C})\), is again a measurable map.

It is sufficient to check the condition \(f^{-1}(B) \in \mathcal{A}\) for all \(B\) of a set that generates \(\mathcal{B}\).

When \(Y = \mathbb{R}^d\), the Borel \(\sigma\)-algebra \(\mathcal{B} = \mathcal{B}(\mathbb{R}^d)\) is meant by default. Thus, a map \(f : X \to \mathbb{R}^d\) is measurable iff \(f^{-1}(B) \in \mathcal{A}\) for every box \(B \subset \mathbb{R}^d\), or equivalently, for every open set \(B \subset \mathbb{R}^d\). A real-valued function \(f : X \to \mathbb{R}\) is measurable iff \(\{x : f(x) \leq a\} \in \mathcal{A}\) for all \(a \in \mathbb{R}\).

If \((Y, \mathcal{B}) = (Y_1, \mathcal{B}_1) \times (Y_2, \mathcal{B}_2)\) then a map \(f : X \to Y\) boils down to \(f_1 : X \to Y_1\) and \(f_2 : X \to Y_2\): \(f(x) = (f_1(x), f_2(x)) \in Y_1 \times Y_2\). In this case \(f\) is measurable iff \(f_1, f_2\) are measurable (think, why). In particular: \(Y_1 = Y_2 = \mathbb{R}, Y = \mathbb{R}^2\).

Every continuous map \(\mathbb{R}^{d_1} \to \mathbb{R}^{d_2}\) is measurable. (The Borel \(\sigma\)-algebras are meant!)

If \(f, g : (X, \mathcal{A}) \to \mathbb{R}\) are measurable then \(f + g\) is measurable. Here is a short and general proof:

\[
\begin{align*}
X \quad &\xrightarrow{x \mapsto (f(x), g(x))} \mathbb{R}^2 \quad \text{measurable} \\
&\xrightarrow{(a, b) \mapsto a+b} \mathbb{R} \quad \text{continuous} \\
&\xrightarrow{\quad \text{measurable}} \mathbb{R}
\end{align*}
\]

The same holds for \(f, g : (X, \mathcal{A}) \to \mathbb{R}^d\).

If \(f, f_1, f_2, \ldots : \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}\), \(f_n(\cdot) \to f(\cdot)\) pointwise, and \(f_n\) are measurable then \(f\) is measurable. Also \(\sup_n f_n(\cdot), \limsup_n f_n(\cdot)\) etc.
1b22 Reminder. Given a measurable map $\varphi : (X, \mathcal{A}) \to (Y, \mathcal{B})$ and a probability measure $\mu$ on $(X, \mathcal{A})$, the formula

$$\nu(B) = \mu(\varphi^{-1}(B)) \quad \text{for } B \in \mathcal{B}$$

defines a probability measure $\nu$ on $(Y, \mathcal{B})$, — the induced measure. We have

$$\int_X f \circ \varphi \, d\mu = \int_Y f \, d\nu \in [0, \infty]$$

for every measurable $f : (Y, \mathcal{B}) \to [0, \infty]$. For $f : (Y, \mathcal{B}) \to \mathbb{R}$ consider $f = f^+ - f^-$. The $\sigma$-algebra $\sigma(\varphi) = \{\varphi^{-1}(B) : B \in \mathcal{B}\}$ generated by $\varphi : X \to (Y, \mathcal{B})$ is the least $\sigma$-algebra (on $X$) that makes $\varphi$ measurable. The $\sigma$-algebra $\sigma(\varphi_1, \varphi_2)$ generated by $\varphi_1 : X \to (Y_1, \mathcal{B}_1)$ and $\varphi_2 : X \to (Y_2, \mathcal{B}_2)$ is, by definition, the least $\sigma$-algebra that makes $\varphi_1, \varphi_2$ measurable. It is the same as $\sigma(\sigma(\varphi_1) \cup \sigma(\varphi_2))$.

About convergence theorems (monotone, dominated) I give no reminder; I just assume that you never forget them!

Random variables

By a random variable we mean a measurable function on a given probability space $(\Omega, \mathcal{F}, P)$. (By default all random variables — on a single probability space.) Usually it maps $\Omega$ to $\mathbb{R}$, but can also map $\Omega$ to a given measurable space; then it may be called a random element of that space. A random element of $\mathbb{R}^d$ (endowed with the Borel $\sigma$-algebra) may be called a $d$-dimensional random vector, or just a $d$-dimensional random variable, — basically the same as $d$ one-dimensional random variables (the coordinates).

Subsets of $\Omega$ belonging to $\mathcal{F}$ are called events; $P(A)$ is called the probability of an event $A$.

Random variables generate $\sigma$-algebras (of events): $\sigma(X)$, $\sigma(X, Y)$ etc.

By the distribution of a random variable $X : \Omega \to \mathbb{R}$ we mean the induced measure $P_X$ on $\mathbb{R}$,

$$P_X(B) = P(X^{-1}(B)) = P\{\omega : X(\omega) \in B\} = \mathbb{P}(X \in B)$$

for Borel sets $B \subset \mathbb{R}$.

Random variables $X, Y : \Omega \to \mathbb{R}$ are called identically distributed if $P_X = P_Y$; that is, $\mathbb{P}(X \in B) = \mathbb{P}(Y \in B)$ for every Borel $B \subset \mathbb{R}$, or equivalently, every interval $B \subset \mathbb{R}$; still equivalently, if $F_X = F_Y$, where $F_X$ is the CDF,

$$F_X(x) = \mathbb{P}(X \leq x).$$
For random elements, $P_X = P_Y$ still applies (in contrast to intervals and CDF). For random vectors CDF applies, and boxes may be used instead of the intervals.

A 2-dim random vector $\omega \mapsto (X(\omega), Y(\omega))$ has a 2-dim distribution $P_{X,Y}$, called also the *joint distribution* of $X$ and $Y$. Usually $P_{X,Y}$ is far from being uniquely determined by $P_X$, $P_Y$. Two such vectors $(X, Y)$ and $(U, V)$ are identically distributed iff $P_{X,Y} = P_{U,V}$. Then $P_X = P_U$ and $P_Y = P_V$ (but the converse fails).

If $\varphi : \mathbb{R} \to \mathbb{R}$ is a Borel function and $X : \Omega \to \mathbb{R}$ a random variable then their composition $\varphi(X) : \Omega \to \mathbb{R}$ is another random variable. Likewise, if $\varphi : \mathbb{R}^2 \to \mathbb{R}$ is a Borel function and $X, Y : \Omega \to \mathbb{R}$ random variables then $\varphi(X, Y) : \Omega \to \mathbb{R}$ is a random variable.

The *expectation* $\mathbb{E} X$ of a random variable $X : \Omega \to \mathbb{R}$ is, by definition, the Lebesgue integral

$$\mathbb{E} X = \int \Omega X \, dP$$

provided that $X$ is integrable, that is, $\mathbb{E} |X| < \infty$. (Otherwise $\mathbb{E} X = \mathbb{E} X^+ - \mathbb{E} X^-; \text{the four cases...}$)

By 11b22,

$$\mathbb{E} X = \int \Omega X \, dP = \int \mathbb{E} X$$

$$\mathbb{E} \varphi(X) = \int \varphi(X) \, dP = \int \varphi \, dP_X = \int z \, P_{\varphi(X)}(dz);$$

$$\mathbb{E} \varphi(X, Y) = \int \varphi(X, Y) \, dP = \int \varphi \, dP_{X,Y} = \int z \, P_{\varphi(X,Y)}(dz)$$

etc.

If $X, Y$ are identically distributed then $\mathbb{E} X = \mathbb{E} Y$.

Using Tonelli’s theorem on $\Omega \times \mathbb{R}$ (or alternatively, approximation) we get

$$\mathbb{E} X = \int_0^\infty \mathbb{P}(X > a) \, da - \int_0^\infty \mathbb{P}(X < -a) \, da$$

(if integrable... four cases...)

Some examples of random variables with distributions of different kind ($\Omega$ is $(0, 1)$ with Lebesgue measure):

$$X\left(\sum_{1}^{\infty} 2^{-k} \beta_k\right) = \sum_{1}^{10} 2^{-k} \beta_k \quad \text{— discrete;}$$

$$X\left(\sum_{1}^{\infty} 2^{-k} \beta_k\right) = \sum_{1}^{\infty} 2^{-k} \beta_{2k} \quad \text{— absolutely continuous;}$$

$$X\left(\sum_{1}^{\infty} 2^{-k} \beta_k\right) = \sum_{1}^{\infty} 2^{-2k} \beta_k \quad \text{— singular.}$$
Independence

Random variables $X, Y : \Omega \rightarrow \mathbb{R}$ are called independent if

$$P_{X,Y} = P_X \times P_Y;$$

that is, $\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A) \mathbb{P}(Y \in B)$ for all Borel sets $A, B \subset \mathbb{R}$, or equivalently, all intervals $A, B \subset \mathbb{R}$; still equivalently, if $F_{X,Y}(x,y) = F_X(x)F_Y(y)$ for all $X, Y \in \mathbb{R}$. For random elements, $P_{X,Y} = P_X \times P_Y$ still applies (in contrast to intervals and CDF). For random vectors CDF applies, that is, $P_{X,Y} = P_X \times P_Y$ for random vectors $X, Y$.

For random vectors CDF applies, and boxes may be used instead of the intervals. If $A$ applies (in contrast to intervals and CDF). For random vectors CDF applies, that is, $P_{X,Y} = P_X \times P_Y$ for random vectors $X, Y$.

Two 2-dim random vectors $(X, Y)$ and $(U, V)$ are independent iff $P_{X,Y,U,V} = P_{X,Y} \times P_{U,V}$, that is, $\mathbb{P}((X, Y) \in A, (U, V) \in B) = \mathbb{P}((X, Y) \in A) \mathbb{P}((U, V) \in B)$ for $A, B \subset \mathbb{R}^2$ (Borel sets, or only boxes). Then, $f(X,Y)$ and $g(U,V)$ are independent for all Borel $f, g : \mathbb{R}^2 \rightarrow \mathbb{R}$. In particular, $X$ and $U$ are independent; also $X$ and $V$; $Y$ and $U$; $Y$ and $V$. (But the converse fails.)

Random variables $X_1, \ldots, X_n$ are called independent, if

$$P_{X_1, \ldots, X_n} = P_{X_1} \times \cdots \times P_{X_n};$$

that is, $\mathbb{P}(X_1 \in B_1, \ldots, X_n \in B_n) = \mathbb{P}(X_1 \in B_1) \cdots \mathbb{P}(X_n \in B_n)$. (This is much stronger than the pairwise independence. A counterexample: random signs conditioned by $X_1 \ldots X_n = +1$.)

Events $A_1, \ldots, A_n$ are called independent, if their indicators $1_{A_1}, \ldots, 1_{A_n}$ are independent random variables. For $n = 2$ this boils down to $\mathbb{P}(A_1, A_2) = \mathbb{P}(A_1) \mathbb{P}(A_2)$, but for $n > 2$ it does not.

One says that $\sigma$-algebras $A_1, \ldots, A_n \subset \mathcal{F}$ are independent if

$$\forall A_1 \in A_1 \ldots \forall A_n \in A_n \quad (A_1, \ldots, A_n \text{ are independent events}).$$

Random variables $X_1, \ldots, X_n$ are independent iff their $\sigma$-algebras $\sigma(X_1), \ldots, \sigma(X_n)$ are independent. The same holds for events (the $\sigma$-algebra generated by an event $A$ being just $\{\emptyset, A, \Omega \setminus A, \Omega\}$).

If random variables $X_1, X_2, X_3, X_4$ are independent then random vectors $(X_1, X_2)$ and $(X_3, X_4)$ are independent, since

$$P_{(X_1,X_2), (X_3,X_4)} = P_{X_1,X_2,X_3,X_4} = P_{X_1} \times P_{X_2} \times P_{X_3} \times P_{X_4} = (P_{X_1} \times P_{X_2}) \times (P_{X_3} \times P_{X_4}) = P_{X_1,X_2} \times P_{X_3,X_4}. $$

Thus, $f(X_1, X_2)$ and $g(X_3, X_4)$ are independent for all Borel $f, g : \mathbb{R}^2 \rightarrow \mathbb{R}$. (Pairwise independence of $X_1, X_2, X_3, X_4$ is not sufficient! The same
counterexample: \((X_1X_2)(X_3, X_4) = +1\). The same holds for more than two factors in a group, and more than two groups.

If \(X, Y\) are independent then

\[
\mathbb{E} f(X, Y) = \int_{\mathbb{R}^2} f(x, y) dP_X \times dP_Y = \int_{\mathbb{R}^2} f(x, y) dP_X(dx) dP_Y(dy);
\]

\[
\mathbb{E} f(X)g(Y) = \left(\int f(x) dP_X\right) \left(\int g(y) dP_Y\right) = \left(\mathbb{E} f(X)\right) \left(\mathbb{E} g(Y)\right);
\]

\[
\mathbb{E} (XY) = \left(\mathbb{E} X\right)\left(\mathbb{E} Y\right).
\]

The same holds for more than two random variables.

Independence for discrete:

\[
p_{X_1, \ldots, X_n}(x_1, \ldots, x_n) = p_{X_1}(x_1) \ldots p_{X_n}(x_n);
\]

\[
\mathbb{E} f(X_1, \ldots, X_n) = \sum_{x_1, \ldots, x_n} f(x_1, \ldots, x_n) p_{X_1}(x_1) \ldots p_{X_n}(x_n).
\]

Independence for absolutely continuous:

\[
p_{X_1, \ldots, X_n}(x_1, \ldots, x_n) = p_{X_1}(x_1) \ldots p_{X_n}(x_n);
\]

\[
\mathbb{E} f(X_1, \ldots, X_n) = \int \cdots \int f(x_1, \ldots, x_n) p_{X_1}(x_1) \ldots p_{X_n}(x_n) dx_1 \ldots dx_n.
\]

1c  **Independent random variables**

Let \(X_1, X_2, \ldots\) be independent identically distributed random variables. Their sums \(S_k = X_1 + \cdots + X_k\) are a (one-dimensional) random walk.

1c1  **Theorem.**  

If \(|X_1| < \infty\) then

\[
P\left(\left|\frac{1}{n}S_n - \mathbb{E} X_1\right| \leq \varepsilon\right) \to 1 \quad \text{as } n \to \infty.
\]

This is the Weak Law of Large Numbers.

Clearly, 1c1 implies 1a5 and 1a24.

Interestingly, 1c1 helps to integrate numerically functions of many (say, 20 or 200) variables (“Monte-Carlo method”).

---

\(^1\)KS, Sect. 7.1, Th. 7.2; [D, Sect. 1.5, Corollary (5.8)].

\(^2\)KS, Sect. 3.8; [D, Sect. 1.5, Exercise 5.3].
Proof of 1c1

Two main ideas: orthogonality (as for la5) and approximation by $L_2$; $L_2$ is dense in $L_1$, but we need also independence...

Let Borel functions $\varphi_1, \varphi_2, \ldots : \mathbb{R} \rightarrow \mathbb{R}$ satisfy $\forall x \varphi_i(x) \rightarrow x$ as $i \rightarrow \infty$, and $\forall i \forall \lambda |\varphi_i(\lambda)| \leq |\lambda|$. Let $X \in L_1 = L_1(\Omega)$ and $Y_i = \varphi_i(X)$. Then $Y_i \in L_1$ and $Y_i \rightarrow X$ in $L_1$ (that is, $\|Y_i - X\|_1 = E|Y_i - X| \rightarrow 0$) by the dominated convergence theorem. For example, we may take $Y_i = \mathbf{1}_{(-i,i)}(X)$ or $Y_i = \text{mid}(-i, X, i)$.

Let $X_k$ be as in 1c1; however, only pairwise independence will be used. Given $\varepsilon > 0$ we define $Y_k = \varphi(X_k)$ with a bounded $\varphi$ such that $\|Y_k - X_k\|_1 \leq \varepsilon$. Then

$$\frac{Y_1 + \cdots + Y_n}{n} - EY_1 \parallel_2 = \frac{1}{n} \|(Y_1 - EY_1) + \cdots + (Y_n - EY_n)\parallel_2 = \frac{1}{n} \sqrt{n} \|Y_1 - EY_1\|_2 \rightarrow 0 \text{ as } n \rightarrow \infty$$

by orthogonality ensured by the pairwise independence: $\langle Y_k - EY_k, Y_l - EY_l \rangle = E((Y_k - EY_k)(Y_l - EY_l)) = (E(Y_k - EY_k))(E(Y_l - EY_l)) = 0 \cdot 0 = 0$ for $k \neq l$.

We have

$$\left\| \frac{X_1 + \cdots + X_n}{n} - EX_1 \right\|_1 \leq \left\| \frac{X_1 + \cdots + X_n}{n} - \frac{Y_1 + \cdots + Y_n}{n} \right\|_1 + \left\| \frac{Y_1 + \cdots + Y_n}{n} - EY_1 \right\|_1 + |EY_1 - EX_1| \leq \left\| X_1 - Y_1 \right\|_1 + \left\| \frac{Y_1 + \cdots + Y_n}{n} - EY_1 \right\|_2 + \|X_1 - Y_1\|_1 \leq 2\varepsilon + o(1);$$

$$\limsup_n \left( \ldots \right) \leq 2\varepsilon \text{ for every } \varepsilon. \text{ Convergence in } L_1 \text{ is proved; convergence in probability follows (by the Markov inequality).}$$

End of proof of 1c1

Recall the cumulative distribution function $F$ defined by

$$F(t) = \mathbb{P}(X_1 \leq t) \text{ for } t \in \mathbb{R}.$$ 

The empirical distribution function is the random function $F_n$ defined by

$$F_n(t) = \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{(-\infty,t]}(X_k).$$
1c2 Theorem. ¹ For every \( \varepsilon > 0 \),
\[
\mathbb{P} \left( \sup_{t} |F_n(t) - F(t)| \leq \varepsilon \right) \to 1 \quad \text{as } n \to \infty.
\]
This is the weak form of the Glivenko-Cantelli theorem.

1c3 Lemma. For every probability measure² \( \mu \) on \( \mathbb{R} \) and every \( \varepsilon > 0 \) there exist \( m \) and \( t_1 < \cdots < t_m \) such that
\[
\mu \left( -\infty, t_1 \right) \leq \varepsilon, \quad \mu \left( t_1, t_2 \right) \leq \varepsilon, \quad \ldots, \quad \mu \left( t_{m-1}, t_m \right) \leq \varepsilon, \quad \mu \left( t_m, +\infty \right) \leq \varepsilon.
\]
Proof. We take \( t_1 = \sup \{ t : \mu((-\infty, t)) \leq \varepsilon \} \) (the set is not empty!), then
\[
\mu((-\infty, t_1)) = \lim_{k} \mu((-\infty, t_1 - \frac{1}{k})) \leq \varepsilon \quad \text{but} \quad \mu((-\infty, t_1)) \leq \lim_{k} \mu((-\infty, t_1 + \frac{1}{k})) \geq \varepsilon. \quad \text{If } \varepsilon \geq 0.5 \text{ then we are done. Otherwise,}
\]
\[
t_2 = \sup \{ t : \mu((t_1, t)) \leq \varepsilon \}, \quad \text{then} \quad \mu((t_1, t_2)) \leq \varepsilon \quad \text{but} \quad \mu((t_1, t_2)) \geq \varepsilon, \quad \text{thus} \quad \mu((-\infty, t_2)) \geq 2\varepsilon. \quad \text{And so on...}
\]

Proof of 1c2
Let \( F(t) = \mu((-\infty, t)) \), random functions \( F_n(t) = \mu_n((-\infty, t)) \), and \( \varepsilon \) be as in 1b4. Lemma 1c3 gives us \( t_1, \ldots, t_m \). By 1a24,
\[
\mathbb{P} \left( |\mu_n((-\infty, t_k)) - \mu((-\infty, t_k))| > \varepsilon \right) \to 0 \quad \text{as } n \to \infty
\]
for each \( k = 1, \ldots, m \). Sum it over \( k \):
\[
\mathbb{P} \left( \max_{k} |\mu_n((-\infty, t_k)) - \mu((-\infty, t_k))| > \varepsilon \right) \to 0 \quad \text{as } n \to \infty.
\]
The same holds for open intervals \((-\infty, t_k)\). Assuming that \( |\mu_n((-\infty, t_k)) - \mu((-\infty, t_k))| \leq \varepsilon \) and \( |\mu_n((-\infty, t_k)) - \mu((-\infty, t_k))| \leq \varepsilon \) we have for every \( k \) and³ every \( t \in (t_{k-1}, t_k) \)
\[
\mu_n((-\infty, t)) \in [\mu_n((-\infty, t_{k-1})), \mu_n((-\infty, t_k))] \subset
\subset [\mu((-\infty, t_{k-1})), \mu((-\infty, t_k))] + \varepsilon \subset \mu((-\infty, t)) + \varepsilon.
\]
and therefore⁴ \( |\mu_n((-\infty, t)) - \mu((-\infty, t))| \leq 3\varepsilon. \) Thus,
\[
\mathbb{P} \left( \sup_{t} |F_n(t) - F(t)| \leq 3\varepsilon \right) = \mathbb{P} \left( \sup_{t} |\mu_n((-\infty, t)) - \mu((-\infty, t))| \leq 3\varepsilon \right) \to 0 \quad \text{as } n \to \infty.
\]
End of proof of 1c2

¹[KS, Sect. 2.1, Th. 2.9].
²The measure may have both atoms and a continuous part, of course.
³The two unbounded intervals are treated similarly.
⁴You can easily improve 3\( \varepsilon \) to 2\( \varepsilon \).