PROBABILITY THEORY

GIOVANNI PANTI

This is a *personal* geodesic along probability theory. It is not a textbook, nor an attempt at it. It does *not* constitute a syllabus for my course. Definitions, lemmas, and examples are often merged in the text, and it is up to the reader to discern which is which. Corrections, suggestions, observations, ..., are most welcome. Version of December 19, 2024.

1. Basics

Definition 1.1. Let $\emptyset \in \mathcal{C} \subseteq \mathscr{P}(\Omega)$, for some nonempty set Ω. Then $\mathcal{C}$ is:

(1) a *semialgebra* $\mathcal{S}$ if it is closed under finite intersections (including the empty one, so $\Omega \in \mathcal{S}$) and, for every $A \in \mathcal{S}$, A^c is a finite disjoint union of elements of $\mathcal{S}$;
(2) an *algebra* $\mathcal{A}$ if it is closed under all boolean operations;
(3) a *σ-algebra* $\mathcal{F}$ if it is closed under all boolean operations and countable unions.

Example 1.2. $\mathscr{P}(\Omega)$. Finite/cofinite subsets of an infinite set. Countable/cocountable subsets of $\mathbb{R}$. $\emptyset$, $\mathbb{R}$, plus all intervals $(-\infty, b]$, $(a, +\infty)$, $(a, b]$, for a, b in $\mathbb{R}$ or in $\mathbb{Q}$.

If A, B belong to the semialgebra $\mathcal{S}$, then $A \setminus B$ is a finite disjoint union of elements of $\mathcal{S}$.

Note that an algebra is an algebra $(\mathcal{A}, \triangle, \cap, \emptyset, \Omega)$ in the algebraic sense of the word, over the field $F_2 = \{0, 1\} = \{\emptyset, \Omega\}$.

The pullback of any semialgebra/algebra/σ-algebra by any function is a semialgebra/algebra/σ-algebra. In particular, this holds for $\Omega' \subseteq \Omega$.

Lemma 1.3. *Given any nonempty $\mathcal{C} \subseteq \mathscr{P}(\Omega)$, the set of all finite intersections of elements of $\mathcal{C}$ and complements of elements of $\mathcal{C}$ is a semialgebra* (*in general larger than $\mathcal{C}$, even though $\mathcal{C}$ might already be a semialgebra*).

Proof. Indeed, it contains $\emptyset$ and Ω, and is closed under finite intersections. Also, for $n \geq 2$, $(A_1 \cap \cdots \cap A_n)^c$ is the disjoint union of the $2^n - 1$ sets $A_1^{f(1)} \cap \cdots \cap A_n^{f(n)}$, for $f : \{1, \ldots, n\} \to \{\mathtt{nothing}, c\}$ not identically `nothing`. $\square$

Lemma 1.4. *Let $(\Omega_i, \mathcal{S}_i)_{i \in I}$ be a family of semialgebras. Then the set of all finite intersections of $\pi_i^{-1} S_i$, for $i \in I$ and $S_i \in \mathcal{S}_i$, is a semialgebra in $\prod_{i \in I} \Omega_i$, called the semialgebra of* cylinders.

Proof. Look at $\left(\pi_1^{-1}[S_1] \cap \cdots \cap \pi_t^{-1}[S_t]\right)^c$. By the proof of Lemma 1.3, it is a finite disjoint union of intersections of stuff of the form $\pi_i^{-1}[S_i]$ or $\left(\pi_j^{-1}[S_j]\right)^c$. We now observe that $\cdots \cap \left(\pi_j^{-1}[S_j]\right)^c \cap \cdots = \cdots \cap \pi_j^{-1}[\dot\bigcup_m S_{j,m}] \cap \cdots = \dot\bigcup_m \cdots \cap \pi_j^{-1}[S_{j,m}] \cap \cdots$. $\square$

Example 1.5. The following are semialgebras.

(1) *Cylinders* in $\mathbb{R}^d$: finite intersections of π_i^{-1} of $\emptyset$, $\mathbb{R}$, $(-\infty, b]$, $(a, +\infty)$, $(a, b]$ in $\mathbb{R}$.
(2) *Cylinders* in $m^{\mathbb{Z}_{\geq 0}}$: finite intersections of π_i^{-1} of subsets of m.

(3) *Blocks* in $m^{\mathbb{Z}_{\geq 0}}$: the emptyset, plus all $[a_0, \ldots, a_{t-1}]$'s, for $a_0, \ldots, a_{t-1} \in m$.

Definition 1.6. The intersection of any nonempty family of algebras/σ-algebras is an algebra/σ-algebra; hence we may speak about the algebra/σ-algebra $\mathcal{A}(\mathcal{C})/\mathcal{F}(\mathcal{C})$ *generated* by any $\mathcal{C}$. The Borel σ-algebra $\mathcal{B}$.

Example 1.7. The intersection of the semialgebras $\{\emptyset, \{a, b, c\}, \{a\}, \{b\}, \{c\}\}$ and $\{\emptyset, \{a, b, c\}, \{a\}, \{b, c\}\}$ is not a semialgebra.

Definition 1.8. of *measurable map* $F : (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}')$.

Lemma 1.9. *Let $F : (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}')$ be a map, and let $\mathcal{C}'$ be a subset of $\mathcal{F}'$ that generates $\mathcal{F}'$ as a σ-algebra. Assume that the F-counterimage of every element of $\mathcal{C}'$ is in $\mathcal{F}$. Then F is measurable. In particular, a continuous map between topological spaces is a measurable map w.r.t. the relative Borel σ-algebras.*

Proof. Just note that $\{A \subseteq \Omega' : F^{-1}A \in \mathcal{F}\}$ is a σ-algebra. $\square$

Theorem 1.10. *The Borel σ-algebra of $\mathbb{R}$ is generated both by the family of all open intervals with rational endpoints, and by the family of all intervals $(-\infty, a]$, with rational a's. Hence, we can check measurability of functions just by checking the counterimages of $(-\infty, a]$'s.*

Proof. Key point: $\mathcal{C} \subseteq \mathcal{F}(\mathcal{D})$ implies $\mathcal{F}(\mathcal{C}) \subseteq \mathcal{F}(\mathcal{D})$. We have $\{\text{open sets}\} \subseteq \mathcal{F}(\{\text{open intervals}\})$, and hence the first statement. For the second one, we need $\{(-\infty, a]\} \subseteq \mathcal{F}(\{\text{open intervals}\})$ and $\{\text{open intervals}\} \subseteq \mathcal{F}(\{(-\infty, a]\})$, which is easy. $\square$

Example 1.11. Both cylinders and blocks generate the Borel σ-algebra of $m^{\mathbb{Z}_{\geq 0}}$.

Lemma 1.12. *The algebra $\mathcal{A}(\mathcal{S})$ generated by the semialgebra $\mathcal{S}$ is the set $\mathcal{Q}$ of all finite disjoint unions of elements of $\mathcal{S}$.*

Proof. One direction is clear. For the other, it is enough to check that $\mathcal{Q}$ is an algebra. Closure by finite intersections is clear. For complements, we have

$$(S_1 \,\dot\cup \cdots \dot\cup\, S_t)^c = S_1^c \cap \cdots \cap S_t^c = (T_{1,1} \,\dot\cup \cdots \dot\cup\, T_{1,r_1}) \cap \cdots \cap (T_{1,t} \,\dot\cup \cdots \dot\cup\, T_{t,r_t}),$$

which is in $\mathcal{Q}$, since it is a finite intersection of elements of $\mathcal{Q}$. $\square$

Example 1.13. The algebra $\mathcal{A}$ generated by blocks in $m^{\mathbb{Z}_{\geq 0}}$ equals the algebra generated by cylinders. We have $A \in \mathcal{A}$ iff A is clopen in $m^{\mathbb{Z}_{\geq 0}}$ iff A is expressible as a finite boolean combination of sets of the form $(\omega_i = a)$. For $m = 2$, $\mathcal{A}$ is the free boolean algebra on countably many generators.

Definition 1.14. Let $\mathcal{C}$ be either a semialgebra, or an algebra, or a σ-algebra on Ω. A *[positive] measure* on $(\Omega, \mathcal{C})$ is a map $\mu : \mathcal{C} \to [0, +\infty]$ satisfying $\mu(\emptyset) = 0$ and conditional σ-additivity. If $\mu(\Omega) = 0$ then μ is *trivial*; if not otherwise specified, "measure" will always mean "nontrivial measure". Relaxing σ-additivity to finite additivity we get the definition of a *finitely additive measure*. A measure is *finite* if $\mu(\Omega) < +\infty$, and is a *probability* if $\mu(\Omega) = 1$; in this case we use P for μ. A measure is *σ-finite* if Ω can be written as a countable union of μ-finite sets.

Example 1.15. The counting measure $(\Omega, \mathscr{P}(\Omega), \sharp)$ on a finite, countable, and uncountable set. Dirac measures. Measures are closed under finite or countable nonnegative combinations, and probabilities under finite or countable affine combinations.

Non-principal ultrafilters provide examples of $\{0,1\}$-valued finitely additive probabilities. Also, setting μ on a countable set to be $0/\infty$ on finite/infinite subsets we obtain a finitely additive measure which is not a measure.

Let μ be finitely additive on an algebra. Then $A \subseteq B$ implies $\mu(A) \leq \mu(B)$; if $\mu(B) < \infty$, then $\mu(B \setminus A) = \mu(B) - \mu(A)$. Moreover, $\mu(A \cup B) \leq \mu(A) + \mu(B)$ and $\mu(A \triangle B) = 0$ implies $\mu(A) = \mu(B) = \mu(A \cap B)$.

For every at most countable family $\{a_i\}_{i \in I}$ in $[0, +\infty]$, the sum $\sum_i a_i$ is well defined and does not depend on the order.

If Ω is at most countable, we always endow it with the σ-algebra $\mathscr{P}(\Omega)$; it is then easy to describe all measures and all probabilities on Ω.

Definition 1.16. Let μ be a f.a. measure on a semialgebra. If for every event and countable family of events $A, \{A_i\}_{i<\omega}$ we have

$$A = \bigcup_{i<\omega} A_i \text{ implies } \mu(A) \leq \sum_{i<\omega} \mu(A_i),$$

then we say that μ is *σ-subadditive*.

Theorem 1.17. *Let μ be a f.a. measure on the algebra $\mathcal{A}$. T.f.a.e.*:

(1) *μ is a measure*;
(2) *for every $A, A_0, A_1, \ldots \in \mathcal{A}$, if A_n converges monotonically increasing to A, then $\mu(A_n)$ converges to $\mu(A)$*;
(3) *μ is σ-subadditive.*

If the above conditions hold then:

(4) *for every $A, A_0, A_1, \ldots \in \mathcal{A}$, if at least one of the A_n is μ-finite and A_n converges monotonically decreasing to A, then $\mu(A_n)$ converges to $\mu(A)$.*

Proof. (1) implies (2): consider $B_i = A_i \setminus A_{i-1}$. (2) implies (3): consider $B_i = A_0 \cup \cdots \cup A_i$ and assume $\sum_{i<\omega} \mu(A_i) = \alpha < \mu(A)$. Then all partial sums are $\leq \alpha$ and therefore $\mu(B_i)$ cannot converge to $\mu(A)$. (3) implies (1): let $A = \dot{\bigcup} A_i$. We need $\mu(A) \geq \sum \mu(A_i)$, which is true since $\mu(A)$ is $\geq$ than the measure of any finite union of the A_i's.
(2) implies (4) by passing to the complements. $\square$

Given a sequence of events $(A_n)_{n<\omega}$, the functions $\operatorname{limsup}_n \mathbb{1}_{A_n}$ and $\operatorname{liminf}_n \mathbb{1}_{A_n}$ (defined componentwise) exist and are $\{0,1\}$-valued; hence they define sets

$$B = \operatorname*{limsup}_n A_n = \bigcap_n \bigcup_{k \geq n} A_k,$$

$$C = \operatorname*{liminf}_n A_n = \bigcup_n \bigcap_{k \geq n} A_k.$$

The function $\lim_n \mathbb{1}_{A_n}$ may or may not exist; if it exists it is $\{0,1\}$-valued and hence defines a set $D = \lim_n A_n$.

Lemma 1.18. *B and C are events with $B \supseteq C$. $B = C$ iff D exists. If D exists, then it is an event. In that case, for every measure μ such that $\mu(\bigcup_{k \geq n} A_k)$ is finite for some n, we have $\mu(D) = \lim_n \mu(A_n)$.*

Proof. For the last claim we have, for every n,

$$\bigcap_{k\geq n} A_k \subseteq D \subseteq \bigcup_{k\geq n} A_k,$$

as well as

$$\bigcap_{k\geq n} A_k \subseteq A_n \subseteq \bigcup_{k\geq n} A_k.$$

□

2. Independence and conditional probability

Definition 2.1. A *μ-measurable partition* of $(X, \mathcal{X}, \mu)$ is a finite or countable family $\{E_i\}_{i\in I}$ of elements of $\mathcal{X}$ such that:

(1) $E_i \cap E_j = \emptyset$ for $i \neq j$;
(2) $\mu(X \setminus \bigcup E_i) = 0$.

We identify two μ-measurable partitions $\{E_i\}$ and $\{E_i'\}$ on the same set of indices if $\mu(E_i \triangle E_i') = 0$ for every i.

For the rest of this section we deal with probability spaces $(\Omega, \mathcal{F}, P)$ only.

Definition 2.2. A family $\{A_i\}_{i\in I}$ of events is P-independent if for every finite $J \subseteq I$ we have

$$P(\bigcap_{j\in J} A_j) = \prod_{j\in J} P(A_j).$$

A family of pairwise P-independent elements is not necessarily P-independent. If A, B are P-independent, so are A, B^c.

Definition 2.3. Let $P(E) > 0$; we then have the probability space $(\Omega, \mathcal{F}, P(-|E))$, and $P(-|E)$ is the *conditional probability given* E.

The events A such that P and $P(-|E)$ agree on A are precisely those events which are P-independent with E.

Lemma 2.4. *Fix $A_1, \ldots, A_n$ and assume $P(A_1 \cap \cdots \cap A_{n-1}) > 0$. Then*

$$\begin{aligned} P(A_1 \cap \cdots \cap A_n) &= P(A_1 \cap \cdots \cap A_{n-1}) \cdot P(A_n | A_1 \cap \cdots \cap A_{n-1}) \\ &= P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap \cdots A_{n-1}). \end{aligned}$$

We agree once for all that $\infty \cdot 0 = 0$. In particular, if $P(E) = 0$, then we have that $P(A|E)$ is undefined, but the product $P(A|E)P(E)$ is defined and equals 0.

Theorem 2.5 (The Bayes Theorem)**.** *Let $\{E_i\}_{i\in I}$ be a measurable partition.*

(1)
$$P = \sum_i P(E_i) P(-|E_i).$$

(2) *If $P(A) > 0$ then, for every i, we have*
$$P(E_i|A) = P(A|E_i)\frac{P(E_i)}{P(A)}.$$

3. Random variables

Lemma 3.1. *Let $F : (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}')$ be a measurable map, and let μ be a measure on $(\Omega, \mathcal{F})$. Then $F_*\mu$ is a measure on $(\Omega', \mathcal{F}')$.*

Definition 3.2. A *random variable* is a measurable map $X : (\Omega, \mathcal{F}, P) \to (R, \mathcal{B})$, where R is any of the topological spaces $\mathbb{R}, \mathbb{R}^d, \mathbb{C}$. The probability X_*P on R is said to be the *distribution*, or the *law*, of X. The variable X is *discrete* if $X[\Omega]$ is discrete (and hence at most countable, since R has a countable basis) in R. It is *continuous* if $P(X = x) = (X_*P)(\{x\}) = 0$ for every $x \in R$.

Definition 3.3. Let μ be a measure on $(R, \mathcal{B})$, for R one of $\mathbb{R}, \mathbb{R}^d, \mathbb{C}$. The *support* of μ is the complement of the set of all x such that x belongs to an open set of μ-measure 0. Of course $\operatorname{supp}(\mu)$ is closed, and is a subset of the closure of $X[\Omega]$.

Definition 3.4. Given any Borel probability μ on $\mathbb{R}$, its *cumulative distribution function*, or *repartition function*, is the function $M : \mathbb{R} \to [0, 1]$ defined by $M(x) = \mu\big((-\infty, x]\big)$. Note that $\mu((-\infty, x)) = \lim_{x' \nearrow x} M(x') = M(x - 0)$ and $\mu(\{x\}) = M(x) - M(x - 0)$.

If μ is discrete, then it has strictly positive value only in the points of the finite or countable set $A = \operatorname{supp}(\mu)$. Its *discrete-density distribution function* is then the function $m : \mathbb{R} \to \mathbb{R}_{\geq 0}$ defined by

$$m(x) = \begin{cases} \mu\big(\{x\}\big), & \text{if } x \in A; \\ 0, & \text{otherwise.} \end{cases}$$

Example 3.5. The following are examples of discrete distributions on $\mathbb{R}$; plots done with SageMath, `http://www.sagemath.org/`.

(1) The *Binomial*, or *Bernoulli* distribution, $Bin(n, p)$, for $n \in \mathbb{Z}_{>0}$ and $p \in (0, 1)$:

$$m(k) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

(2) The *Hypergeometric* distribution, $Hyp(N, K, n)$, for $1 \leq K, n \leq N$, see Figure 1:

$$m(k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}.$$

(3) The *Poisson* distribution, $Poisson(\mu)$, for $\mu \in \mathbb{R}_{>0}$, see Figure 2:

$$m(k) = \exp(-\mu)\frac{\mu^k}{k!}.$$

(4) The *Geometric* distribution, $Geom(p)$, for $p \in (0, 1)$:

$$m(k) = p(1 - p)^k.$$

(5) The *Zeta* distribution, $Zeta(\alpha)$, for $\alpha \in \mathbb{R}_{>1}$:

$$m(k) = \zeta(\alpha)^{-1}\frac{1}{k^\alpha}.$$

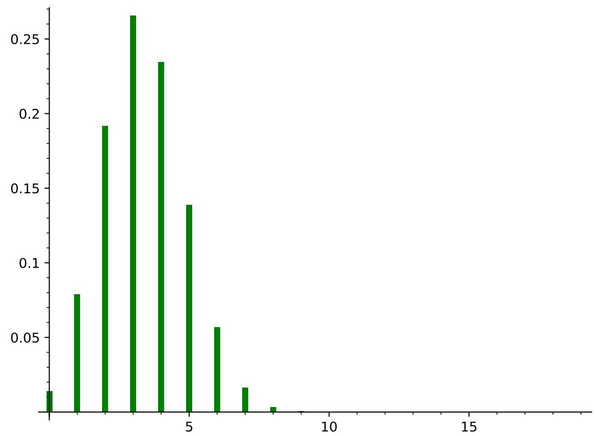

FIGURE 1

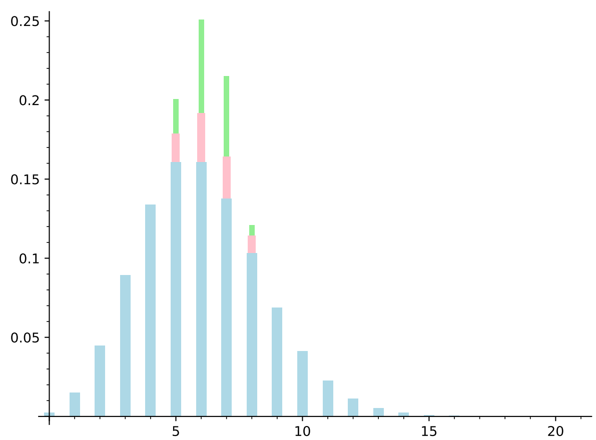

FIGURE 2

4. CONSTRUCTION OF MEASURES

Theorem 4.1 (The Monotone Class Theorem)**.** *Let $\mathcal{C} \subseteq \mathscr{P}(\Omega)$ be closed under finite intersections (in particular, $\Omega \in \mathcal{C}$). Let $\mathcal{M}$ be the smallest overclass of $\mathcal{C}$ which is closed under countable increasing unions and increasing differences. Then $\mathcal{M} = \mathcal{F}(\mathcal{C})$.*

Proof. [JP03, Theorem 6.2 p. 36] □

Lemma 4.2. *Let $\mathcal{A}$ be the algebra generated by the semialgebra $\mathcal{S}$, and let μ be a f.a. measure on $\mathcal{S}$. Then μ can be extended uniquely to a f.a. measure on $\mathcal{A}$. If μ is σ-subadditive on $\mathcal{S}$, then the extension remains σ-subadditive, and hence is a measure on $\mathcal{A}$.*

Theorem 4.3 (The Carathéodory-Hahn-Kolmogorov Extension Theorem)**.** *Let $\mathcal{F}$ be the σ-algebra generated by the algebra $\mathcal{A}$. Then every σ-finite measure on $\mathcal{A}$ can be uniquely extended to $\mathcal{F}$, and the extension is again σ-finite.*

Proof. If an extension exists, then it is clearly σ-finite. Uniqueness is easy: let μ, ν be measures on $\mathcal{F}$ that agree on $\mathcal{A}$. For every $A \in \mathcal{A}$ such that $\mu(A) = \nu(A)$ is finite, let $\mathcal{M}_A = \{B \in \mathcal{F} : \mu(B \cap A) = \nu(B \cap A)\}$. Then $\mathcal{M}_A$ is a monotone overclass of $\mathcal{A}$, and hence equals $\mathcal{F}$. Using σ-finiteness we can write $\Omega = \dot{\bigcup}_{i<\omega} A_i$ with $A_i \in \mathcal{A}$ of finite $(\mu = \nu)$-measure. Then, for every $B \in \mathcal{F}$, $\mu(B) = \sum_i \mu(B \cap A_i) = \sum_i \nu(B \cap A_i) = \nu(B)$. Existence is difficult: see [Bil95, Theorems 11.2 and 11.3]. □

Lemma 4.4. *Let μ, ν be measures on $(X, \mathcal{X})$, and let $\mathcal{S}$ be a semialgebra that generates $\mathcal{X}$. Assume that $\mu = \nu$ on $\mathcal{S}$ and that X is the union of at most countably many elements of $\mathcal{S}$ of finite $\mu(= \nu)$-measure. Then $\mu = \nu$.*

Corollary 4.5. *There exists a unique σ-finite measure on $(\mathbb{R}, \mathcal{B})$ that assigns to each interval its lenght. This measure is translation-invariant*; *it is named the* Lebesgue measure*, and denoted by λ.*

Proof. Let $\mathcal{S}$ be the semialgebra of Example 1.2. Define λ to have value $a - b$ on $(b, a]$, value 0 on $\emptyset$, and $+\infty$ on all unbounded intervals. Then λ is a f.a., σ-subadditive measure on $\mathcal{S}$ (additivity is clear, while σ-subadditivity requires a compactness argument). Hence it extends uniquely to a measure on the generated σ-algebra, namely $\mathcal{B}$. For translation-invariance, fix r and let $\mu = (T_r)_* \lambda$. Then $\mu = \lambda$ on $\mathcal{S}$. □

Corollary 4.6. *Given any probability vector $(p_0, \dots, p_{n-1})$, there exists a unique probability on $(n^\omega, \mathcal{B})$ that assigns to each block $[a_0, \dots, a_{t-1}]$ the number $p_{a_0} \cdots p_{a_{t-1}}$.*

5. Probability measures on $\mathbb{R}$

A function $M : \mathbb{R} \to \mathbb{R}$ is *right continuous* in c if for every $\varepsilon > 0$ there exists $\delta > 0$ such that for every $x \in [c, c + \delta)$ we have $|Mx - Mc| < \varepsilon$. If M is nondecreasing, then M is right continuous in c iff for every sequence $x_0 \geq x_1 \geq \cdots$ converging to c the sequence Mx_n converges to Mc iff there exists a sequence $x_0 \geq x_1 \geq \cdots$ converging to c such that Mx_n converges to Mc,

Probability measures on $\mathbb{R}$ are completely described by the following theorem.

Theorem 5.1. *A function $M : \mathbb{R} \to [0, 1]$ is the repartition function of a probability μ on $\mathbb{R}$ iff M is nondecreasing, right continuous, tending to 0 for $x \to -\infty$ and to 1 for $x \to +\infty$. If this happens, then μ and M determine each other.*

Remark 5.2. We have $\{b\} = \bigcap_{a \nearrow b} (a, b]$, and therefore $\mu(\{b\}) = \lim_{a \nearrow b}(Mb - Ma) = Mb - M(b - 0)$, with $M(b - 0) = \sup\{Mx : x < b\}$. Thus the random variable X is continuous according to Definition 3.2 iff the repartition function of $X_* P$ is continuous.

Example 5.3. The following are examples of continuous probability distributions on $\mathbb{R}$ which are induced by a Riemann-integrable density function.

1. The *Uniform* distribution with parameters $\alpha < \beta$, given by $m = (\beta - \alpha)^{-1} \mathbb{1}_{[\alpha, \beta]}$.
2. The *Exponential* distribution with parameter $\beta > 0$, given by $m(x) = \beta \exp(-\beta x)$ for $x > 0$, and $m(x) = 0$ otherwise.
3. More generally, the *Gamma* distribution with parameters $\alpha, \beta > 0$, given by

$$m(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha - 1} \exp(-\beta x)$$

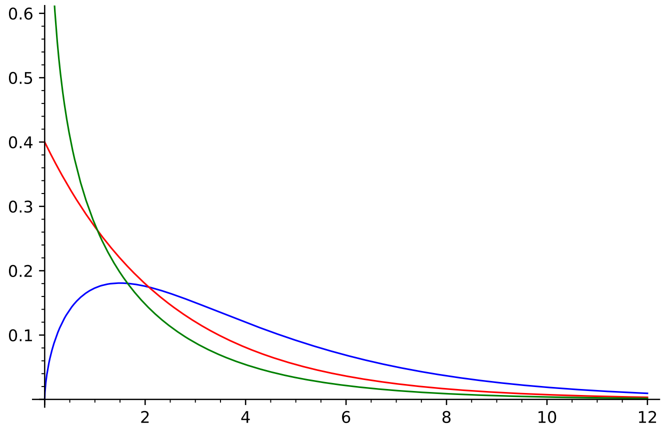

FIGURE 3

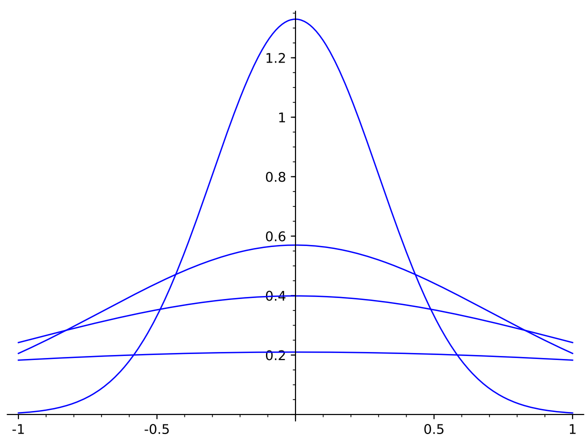

FIGURE 4

if $x > 0$, and $m(x) = 0$ otherwise; see Figure 3.

(4) The *Normal*, or *gaussian* distribution, $Normal(\mu, \sigma^2)$; see Figure 4.

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

Example 5.4. Repartition functions can be extremely slippery. Let P be the $(5/7, 2/7)$-probability on $\Omega = 2^{\mathbb{N}}$, and let $X : \Omega \to \mathbb{R}$ be the random variable induced by the binary expansion. The repartition function of $\mu = X_* P$ is a typical devil's staircase; see Figure 5.

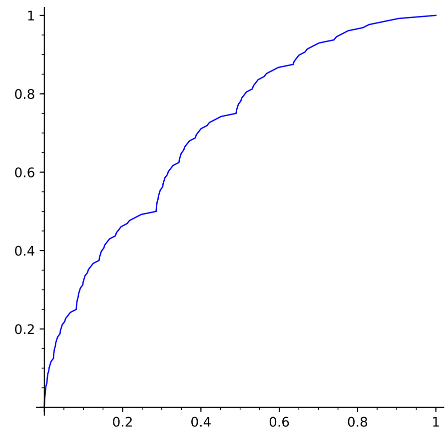

FIGURE 5

6. INTEGRATION THEORY

We see $[0,+\infty]$ and $[-\infty,+\infty]$ as topological spaces, homeomorphic to $[0,1]$ and $[-1,1]$ via, e.g., the cotangent function. They are then endowed with the Borel σ-algebra $\mathcal{B}$. We agree that $0\cdot(+\infty)=0$.

Lemma 6.1. *Let $f_0, f_1, \ldots : (X,\mathcal{X}) \to ([-\infty,+\infty],\mathcal{B})$ be measurable.*

(1) $\sup f_n$, $\inf f_n$, $\operatorname{limsup} f_n$, $\operatorname{liminf} f_n$ *(all of them pointwise defined) are all measurable.*
(2) *If the pointwise limit* $\lim f_n$ *exists, then it is measurable.*
(3) *If $f_1,\ldots,f_d : X \to \mathbb{R}$ are measurable, then $(f_1,\ldots,f_d) : X \to \mathbb{R}^d$ is measurable.*
(4) *If, moreover, $g : \mathbb{R}^d \to \mathbb{R}$ is measurable, then $g(f_1,\ldots,f_d) : X \to \mathbb{R}$ is measurable; in particular, the set of measurable functions from X to $\mathbb{R}$ is an $\mathbb{R}$-algebra.*

Proof. $\{x : (\sup f_n)(x) \le a\} = \bigcap_n \{x : f_n(x) \le a\}$ and $\sup f_n$ is measurable. Analogously $\inf f_n$ is measurable. We have $\operatorname{limsup}_n f_n = \inf_n(\sup_{m\ge n} f_m)$, which is thus measurable. One proves the last two items by noting that the set of all products of open intervals with rational endpoints is a countable basis for the topology of $\mathbb{R}^d$, and therefore generates the Borel σ-algebra. □

Definition 6.2. Let $(X,\mathcal{X},\mu)$ be a σ-finite measure space. A *step function* is a measurable function $s : X \to \mathbb{R}_{\ge 0}$ whose range is finite; equivalently, it is a function that can be written (nonuniquely) as a finite sum $s = \sum_{i<n} a_i \mathbb{1}_{A_i}$.

Lemma 6.3. *Every step function can be written in disjoint form $s = \dot{\sum}_{j<m} b_j \mathbb{1}_{B_j}$ in such a way that*

$$\sum_{i<n} a_i\mu(A_i) = \sum_{j<m} b_j\mu(B_j), \text{ (it may be } +\infty).$$

If $\dot{\sum}_{j<m} b_j \mathbb{1}_{B_j} = \dot{\sum}_{k<t} c_k \mathbb{1}_{C_k}$, then $\sum_{j<m} b_j\mu(B_j) = \sum_{k<t} c_k\mu(C_k)$.

Definition 6.4. Let $s = \sum_{i<n} a_i \mathbb{1}_{A_i} : X \to \mathbb{R}_{\geq 0}$ be a step function. We define

$$\mu(s) = \int_X s \, \mathrm{d}\mu = \sum_{i<n} a_i \mu(A_i) \in [0, +\infty],$$

and note that $\mu(-)$ is positively linear.

Let $f : X \to [0, +\infty]$ be measurable. We define

$$\mu(f) = \int_X f \, \mathrm{d}\mu = \sup\{\mu(s) : 0 \leq s \leq f\} \in [0, +\infty],$$

and note that $f \leq g$ implies $\mu(f) \leq \mu(g)$.

Lemma 6.5. *Let $f : X \to [0, +\infty]$ be measurable.*

(1) *There exists an increasing sequence $s_0, s_1, \dots : X \to \mathbb{R}_{\geq 0}$ of step functions that converges pointwise to f.*
(2) *For every such sequence, $\mu(s_n)$ converges to $\mu(f)$ (this is a preliminary version of the Monotone Convergence Theorem).*
(3) *$\mu(-)$ is positively linear (i.e., if g is another such function and $r \geq 0$, then $\mu(f + g) = \mu(f) + \mu(g)$ and $\mu(rf) = r\mu(f)$).*

For $f : X \to [-\infty, +\infty]$ we set $f^+ = f \vee 0$, $f^- = (-f) \vee 0$, $|f| = f^+ + f^- = f^+ \vee f^- = f \vee (-f)$.

If at least one of $\mu(f^+)$, $\mu(f^-)$ is finite, then we say that f is *integrable* w.r.t. μ, and set

$$\int f \, \mathrm{d}\mu = \int f(x) \, \mathrm{d}\mu(x) = \int f(x)\mu(\, \mathrm{d}x) = \int f^+ \, \mathrm{d}\mu - \int f^- \, \mathrm{d}\mu \in [-\infty, +\infty].$$

Since $f^+, f^- \leq |f| = f^+ + f^-$, we have that both of $\mu(f^+)$ and $\mu(f^-)$ are finite iff so is $\mu(|f|)$.

Definition 6.6. If $f = f_1 + i f_2 : X \to \mathbb{C}$ is measurable, we say that $f \in \mathcal{L}_1(\mu) = \mathcal{L}_1(\mu, \mathbb{C})$ if $\mu(|f|) < +\infty$. Since $|f_1|, |f_2| \leq |f| \leq |f_1| + |f_2|$, this is equivalent to $f_1, f_2 \in \mathcal{L}_1(\mu, \mathbb{R})$; we then set $\mu(f) = \mu(f_1) + i\mu(f_2)$.

Theorem 6.7.
(1) *$\mathcal{L}_1$ is a complex vector space, and $\mu : \mathcal{L}_1 \to \mathbb{C}$ is a positive $\mathbb{C}$-linear functional.*
(2) *If $f \in \mathcal{L}_1$, then $|\mu(f)| \leq \mu(|f|)$.*

Theorem 6.8. *Let $f : X \to \mathbb{C}$ be a measurable function with finite or countable range. Then $f \in \mathcal{L}_1$ iff the series*

$$\sum \{a\mu(f^{-1}a) : a \in \mathrm{range}(f)\}$$

is absolutely convergent (this clearly requires that $\mu(f^{-1}a) < +\infty$ for every $a \neq 0$). If this happens, then $\mu(f)$ equals the sum of that series.

Theorem 6.9. *Let $R : (X, \mathcal{X}, \mu) \to (Y, \mathcal{Y})$ and $f : (Y, \mathcal{Y}) \to (\mathbb{C}, \mathcal{B})$ be measurable. Then $f \in \mathcal{L}_1(R_*\mu)$ iff $f \circ R \in \mathcal{L}_1(\mu)$. If this happens (or if f is $[0, +\infty]$-valued), then*

$$\int_X f \circ R \, \mathrm{d}\mu = \int_Y f \, \mathrm{d}(R_*\mu).$$

Proof. Case 1: f is a step function. Case 2: f is $[0,+\infty]$-valued. Use the MCT for step functions. Case 3: $\mathbb{R}$-valued. Case 4: $\mathbb{C}$-valued. $\square$

Theorem 6.10 (The Monotone Convergence Theorem)**.** *Let $f_0 \le f_1 \le \cdots$ be a monotone sequence of $\mathbb{R}_{\ge 0}$-valued measurable functions, and let $f = \lim f_n$. Then $\mu(f) = \lim \mu(f_n)$.*

Theorem 6.11 (The Fatou Lemma)**.** *Let $f_0, f_1, \dots$ be a sequence of $\mathbb{R}_{\ge 0}$-valued measurable functions. Then $\mu(\liminf f_n) \le \liminf \mu(f_n)$.*

Theorem 6.12 (The Dominated Convergence Theorem)**.** *Let $f_0, f_1, \dots$ be a sequence of $\mathbb{C}$-valued measurable functions. Assume that $\lim f_n = f$ (μ-a.e.) exists, and that there exists an $\mathbb{R}_{\ge 0}$-valued $g \in \mathcal{L}_1$ such that $|f_n| \le g$ for every n. Then all f_ns and f are in $\mathcal{L}_1$, and $\lim \mu(f_n) = \mu(f)$.*

Theorem 6.13 (The Markov and Chebishev Inequalities)**.** *Let $f : (X,\mu) \to \mathbb{R}_{\ge 0}$ be measurable, $a \ge 0$. Then*:

(1) $a\mu(f \ge a) \le \mu(f)$;
(2) *for every $p \ge 1$, we have $a^p\mu(f \ge a) \le \mu(f^p)$.*

Proof. Observe that $a\mathbb{1}_{(f \ge a)} \le f$ and integrate. We obtain (2) by applying (1) to f^p and a^p. $\square$

Theorem 6.14. *Let $f_0, f_1, \dots : X \to \mathbb{C}$ be measurable.*

(1) *If they are $\mathbb{R}_{\ge 0}$-valued, then $\sum_n f_n$ is a $[0,+\infty]$-valued measurable function, and*

$$\int_X \Big(\int_{\mathbb{Z}_{\ge 0}} f_n(x)\,\mathrm{d}\sharp(n) \Big)\,\mathrm{d}\mu(x) = \int_X \big(\sum_n f_n\big)\,\mathrm{d}\mu = \sum_n \int_X f_n\,\mathrm{d}\mu = \int_{\mathbb{Z}_{\ge 0}} \Big(\int_X f_n(x)\,\mathrm{d}\mu(x) \Big)\,\mathrm{d}\sharp(n). \tag{6.1}$$

(2) *Assume that $\mu(\sum_n |f_n|) = \sum_n \mu(|f_n|) < +\infty$. Then*:
   (i) *$\sum_n f_n$ converges absolutely a.e., and determines a function in $\mathcal{L}_1(\mu)$.*
   (ii) (6.1) *holds* (*this is a preliminary form of the Fubini theorem*), *and the series to the right converges absolutely*;
   (iii) $\lim_n f_n = 0$ *a.e.*.

7. L_p SPACES

In this section $f, g, \dots : (X, \mathcal{X}, \mu) \to (\mathbb{C}, \mathcal{B})$ are complex-valued measurable functions. Let $1 \le p < +\infty$. We set $\mathcal{L}_p(\mu) = \{f : f^p \in \mathcal{L}_1(\mu)\}$. Let $\|f\|_p = \big(\mu(|f|^p)\big)^{1/p}$. By Minkowski's inequality $\|f+g\|_p \le \|f\|_p + \|g\|_p$; in particular, $\mathcal{L}_p(\mu)$ is a $\mathbb{C}$-vector space. Moreover, $\|-\|_p$ is a seminorm on $\mathcal{L}_p(\mu)$, and a norm on the set $L_p(\mu)$ of equivalence classes.

Lemma 7.1. *Let $f, g \in \mathcal{L}_p$. Then $\|f-g\|_p = 0$ iff $\mu(f \ne g) = 0$.*

Proof. $\mu(f \ne g) = 0$ implies $|f-g|^p = 0$ (a.e.) and $\int |f-g|^p = 0$. Conversely, for every $n \ge 1$, by Chebishev $(1/n)^p\mu(|f-g| \ge 1/n) \le \int |f-g|^p$. Therefore $\mu(|f-g| \ge 1/n) = 0$ for every n, and $\mu(f \ne g) = 0$. $\square$

Let $\|f\|_\infty$ be the infimum of all $M \in [0, +\infty]$ such that $\mu(M < |f|) = 0$; sometimes $\|f\|_\infty$ is called the *essential supremum* of f. Let $\mathcal{L}_\infty(\mu) = \{f : \|f\|_\infty < +\infty\}$. Again $\|-\|_\infty$ is a seminorm, and again $\|f - g\|_\infty = 0$ amounts to $\mu(f \neq g) = 0$. Thus the set $L_\infty(\mu)$ of equivalence classes inherits from $\mathcal{L}_\infty(\mu)$ a complex vector space structure.

All $\mathcal{L}_p$'s and L_p's are closed under complex conjugation.

Theorem 7.2. *Let $f, g \in L_2(\mu)$. Then $fg \in L_1(\mu)$, and the Cauchy-Schwarz-Bunyakovsky-Hölder inequality*

$$|\langle f, g\rangle| := \left| \int_X \bar{f} g \, \mathrm{d}\mu \right| \leq \|f\|_2 \|g\|_2$$

holds.

Theorem 7.3. *Let P be a probability*; *then*

$$L_\infty(P) \subseteq \cdots \subseteq L_3(P) \subseteq L_2(P) \subseteq L_1(P).$$

8. $E(X)$, $\mathrm{Var}(X)$, $G_X(z)$ FOR COMMON R.V.S

Suppose $X \in L_p(P)$, with $p < \infty$. Then $E(X^p)$ is the *pth moment* of X, and $E\big((X - E(X))^p\big)$ its *pth central moment*.

If $X \in L_2(P)$ and $E(X) = \mu$ (= *expectation* = *mean*), define

$$\mathrm{Var}(X) = E(|X - \mu|^2) = \sigma^2 \in \mathbb{R}_{\geq 0}.$$

We have:

(1) $\mathrm{Var}(X) = E(|X|^2) - |\mu|^2$.
(2) For $a \geq 0$, $a^2 P(|X - \mu| \geq a) \leq \sigma^2$.
(3) $\mathrm{Var}(X + a) = \mathrm{Var}(X)$ and $\mathrm{Var}(aX) = |a|^2 \mathrm{Var}(X)$.
(4) If $\mathrm{Var}(X) = 0$, then $X = \mu \mathbb{1}$ (a.e.).

Definition 8.1. Let X have values in $\mathbb{Z}_{\geq 0}$, with $X_* P = \mu$ and discrete-density function m. Then, for every z in the complex closed unit disk, the series

$$G_X(z) = \sum_{k=0}^{\infty} m(k) z^k = \int_{\mathbb{R}} z^x \, \mathrm{d}\mu(x) = \int_\Omega z^{X(\omega)} \, \mathrm{d}P(\omega) = E(z^X)$$

(written also G_μ or G_m) converges absolutely, and determines the *generating function* of X.

Since G can be differentiated termwise inside its disk of convergence, it determines μ via

$$m(k) = \frac{G^{(k)}(0)}{k!}.$$

We have $G(1) = 1$ and, if the radius of convergence is > 1, we also have $G'(1) = E(X)$, $G''(1) = E(X^2) - E(X)$, $G'''(1) = E(X^3) - 3E(X^2) + 2E(X), \ldots$.

X	$E(X)$	$\mathrm{Var}(X)$	$G_X(z)$ or $\varphi_X(u)$
$\mathit{Bin}(n,p)$	np	$np(1-p)$	$(1-p+pz)^n$
$\mathit{Hyp}(N,K,n)$	Kn/N		
$\mathit{Poisson}(\mu)$	μ	μ	$\exp\bigl(\mu(z-1)\bigr)$
$\mathit{Geom}(p)$	$(1-p)/p$	$(1-p)/p^2$	$p/\bigl(1-(1-p)z\bigr)$
$\mathit{Zeta}(\alpha)$	$\zeta(\alpha-1)/\zeta(\alpha)$ if $\alpha>2$ ∞ otherwise		
$\mathit{Uniform}(a,b)$	$(a+b)/2$	$(b-a)^2/12$	
$\mathit{Gamma}(\alpha,\beta)$	α/β	α/β^2	
$\mathit{Normal}(\mu,\sigma^2)$	μ	σ^2	$\exp(i\mu u-\sigma^2u^2/2)$

9. The Borel-Cantelli Lemma

Theorem 9.1. *Let $(A_n)_{n<\omega}$ be a sequence of measurable sets in $(X,\mathcal{X},\mu)$.*

(1) *If $\sum_{n<\omega}\mu(A_n)<\infty$, then $\mu(\limsup_n A_n)=0$.*
(2) *If $\mu=P$ is a probability, $\sum_{n<\omega}P(A_n)=\infty$, and the A_n's are independent, then $P(\limsup_n A_n)=1$.*

10. Stochastic processes

A family (even more than countable) of sub-σ-algebras $\{\mathcal{E}_i\}_{i\in I}$ of $\mathcal{F}$ is *P-independent* if for every finite subset J of I and for every choice of $A_j\in\mathcal{E}_j$ we have

$$P(\bigcap_{j\in J}A_j)=\prod_{j\in J}P(A_j).$$

A family $\{X_i\}_{i\in I}$ of random variables is *P-independent* if the family $\{X_i^{-1}\mathcal{B}\}_{i\in I}$ is P-independent.

Definition 10.1. A *stochastic process* is a sequence of r.v.s $X_0,X_1,\dots:(\Omega,\mathcal{F},P)\to(\mathbb{R},\mathcal{B})$ ($\mathbb{R}$ might be replaced by $\mathbb{C}$ or $\mathbb{R}^d$). It is:

- *independent* if the X_n's are independent;
- *identically distributed* if $(X_n)_*P=(X_m)_*P$ for every n and m;
- *stationary* if the push-forward probability $\overline{X}_*P$ on $\mathbb{R}^{\mathbb{Z}_{\geq 0}}$ is shift-invariant, that is, for every n and every choice of events $A_0,\dots,A_{n-1}$, we have
$$P(X_0\in A_0\cap\dots\cap X_{n-1}\in A_{n-1})=P(X_1\in A_0\cap\dots\cap X_n\in A_{n-1}).$$

I.i.d. implies stationary, which implies i.d.. Let $\Omega=\{0,1\}$ with uniform P. Let $X_n=\mathrm{id}$ if $n\equiv 0\pmod 3$ and $X_n=1-\mathrm{id}$ otherwise. Then $\overline{X}_*P\bigl([011]\bigr)=1/2$ and $\overline{X}_*P\bigl([*011]\bigr)=0$; thus $\overline{X}$ is i.d. and non-stationary.

Lemma 10.2. *Let $g,g_0,g_1,\dots$ be measurable function.*

(1) *If $X_0,X_1,\dots$ is independent, then $g_0\circ X_0,g_1\circ X_1,\dots$ is independent.*
(2) *If $X_0,X_1,\dots$ is identically distributed, then $g\circ X_0,g\circ X_1,\dots$ is identically distributed.*

Definition 10.3. For every n, let

$$\mathcal{C}_n = \big(\sigma\text{-algebra generated by } \bigcup_{m \geq n} X_m^{-1}\mathcal{B}\big) = \bigvee_{m \geq n} X_m^{-1}\mathcal{B}.$$

Then $\mathcal{C}_\infty = \bigcap_n \mathcal{C}_n$ is the *tail σ-algebra* of the process.

Lemma 10.4. *Let $\{\mathcal{F}_i\}_{i \in I}$ be a family of independent σ-algebras. Let $\{I_j\}_{j \in J}$ be a partition of I and, for every $j \in J$, let $\mathcal{M}_j = \bigvee_{i \in I_j} \mathcal{F}_i$. Then the family $\{\mathcal{M}_j\}_{j \in J}$ is independent.*

Proof. [Bil95, Theorem 4.2 p. 50] □

Theorem 10.5 (The Kolmogorov 0-1 Law)**.** *The tail σ-algebra of an independent process is trivial.*

Example 10.6. Fix $p \in [0, 1]$, and let N_p be the set of all $x \in [0, 1]$ such that 5 appears in the decimal expansion of x with frequency p. Then either $\lambda(N_p) = 0$ or $\lambda(N_p) = 1$.

11. Product measures

Definition 11.1. Let $(X, \mathcal{X})$, $(Y, \mathcal{Y})$ be measurable spaces. By Lemma 1.4, the class

$$\mathcal{S} = \{\pi_1^{-1}A \cap \pi_2^{-1}B : A \in \mathcal{X}, B \in \mathcal{Y}\} = \{A \times B : A \in \mathcal{X}, B \in \mathcal{Y}\}$$

is a semialgebra. We define $\mathcal{X} \times \mathcal{Y} = \mathcal{F}(\mathcal{S})$.

Lemma 11.2. *Let X, Y be topological spaces with a countable basis, $\mathcal{X}, \mathcal{Y}$ their Borel σ-algebras, $\mathcal{B}$ the Borel σ-algebra of $X \times Y$. Then $\mathcal{B} = \mathcal{X} \times \mathcal{Y}$.*

If $X, Y : \Omega \to \mathbb{R}$ are r.v.s, then $(X, Y) : \Omega \to \mathbb{R}^2$ is a r.v.. Since r.v.s are closed under postcomposition with continuous functions, the set of all r.v.s from Ω to $\mathbb{R}$ is an $\mathbb{R}$-algebra.

Lemma 11.3. *Let $f : X \times Y \to \mathbb{C}$ be $\mathcal{X} \times \mathcal{Y}$-measurable. Then every $f(a, -)$ is $\mathcal{Y}$-measurable and every $f(-, b)$ is $\mathcal{X}$-measurable.*

Let $C \in \mathcal{X} \times \mathcal{Y}$, and define

$$\varphi_C(x) = \int_Y \mathbb{1}_C(x, -)\, \mathrm{d}\nu \in [0, \infty],$$
$$\psi_C(y) = \int_X \mathbb{1}_C(-, y)\, \mathrm{d}\mu \in [0, \infty].$$

Theorem 11.4. *φ_C is $\mathcal{X}$-measurable, ψ_C is $\mathcal{Y}$-measurable, and*

$$\int_X \varphi_C \, \mathrm{d}\mu = \int_Y \psi_C \, \mathrm{d}\nu.$$

The function ρ that associates that number to C is a σ-finite measure, satisfies $\rho(A \times B) = \mu(A)\nu(B)$, and is the only measure on $(X \times Y, \mathcal{X} \times \mathcal{Y})$ that satisfies such an identity. We denote it by $\rho = \mu \times \nu$, and call it the product measure *of μ and ν.*

Theorem 11.5 (The Tonelli-Fubini Theorem)**.** *Let $(X, \mathcal{X}, \mu)$, $(Y, \mathcal{Y}, \nu)$, $(X \times Y, \mathcal{X} \times \mathcal{Y}, \mu \times \nu)$ be as above, and assume that $f : X \times Y \to \mathbb{C}$ is $\mathcal{X} \times \mathcal{Y}$-measurable.*

(1) *If* $f \geq 0$*, then*

$$\varphi_f(x) = \int_Y f(x,-)\,\mathrm{d}\nu \quad \textit{is } \mathcal{X}\textit{-measurable},$$
$$\psi_f(y) = \int_X f(-,y)\,\mathrm{d}\mu \quad \textit{is } \mathcal{Y}\textit{-measurable},$$

and the identity

$$\int_{X\times Y} f\,\mathrm{d}\mu\times\nu = \int_X \varphi_f\,\mathrm{d}\mu = \int_Y \psi_f\,\mathrm{d}\nu \tag{11.1}$$

holds.

(2) *If any of the integrals in* (11.1)*, with* $|f|$ *for* f*, is finite, then*:
   (a) $f \in L_1(\mu\times\nu)$;
   (b) $f(x,-) \in L_1(\nu)$ *for* μ*-every* x*, and* $\varphi_f \in L_1(\mu)$;
   (c) $f(-,y) \in L_1(\mu)$ *for* ν*-every* y*, and* $\psi_f \in L_1(\nu)$;
   (d) *the identity in* (11.1) *holds.*

12. The Multiplication Theorem

Theorem 12.1. *Let* X, Y *be r.v.'s and define* $Z(\omega) = (X(\omega), Y(\omega))$*. Then* Z *is a r.v., and* X*,* Y *are independent iff* $Z_*P = X_*P \times Y_*P$*.*

Corollary 12.2. *Let* $X, Y \in L_1(P)$ *be independent. Then* $XY \in L_1(P)$ *and* $E(XY) = E(X)E(Y)$*.*

Corollary 12.3. *Let* X, Y *be* $\mathbb{Z}_{\geq 0}$*-valued and independent, and assume that* G_X, G_Y *have radius of convergence* $\geq r > 1$*. Then* $G_{X+Y}(z) = G_X(z)G_Y(z)$ *for* $|z| < r$*.*

Taking into account the fact that a generating function determines the variable, this implies that the sum of two independent Poisson variables, of parameters μ and ν, is Poisson of parameter $\mu + \nu$.

Proposition 12.4. *Let* $X : (\Omega, P) \to \mathbb{R}$ *and* $X' : (\Omega', P') \to \mathbb{R}$ *be r.v.s. Let* $Y = X \circ \pi_1, Y' = X' \circ \pi_2 : (\Omega\times\Omega', \mathcal{F}\times\mathcal{F}') \to \mathbb{R}$*. Then* X *and* Y *have the same law,* X' *and* Y' *have the same law, and* Y, Y' *are independent.*

13. Covariance

Let $X, Y \in L_2(P)$, $\mu = E(X)$, $\nu = E(Y)$. The *covariance* of X and Y is

$$\operatorname{Cov}(X,Y) = E\big(\overline{(X-\mu)}(Y-\nu)\big) \in \mathbb{C}.$$

Lemma 13.1. (1) $\operatorname{Var}(X) = \operatorname{Cov}(X,X)$;
(2) Cov *is hermitian sesquilinear*;
(3) $\operatorname{Cov}(X,Y) = E(\overline{X}Y) - \overline{\mu}\nu$;
(4) *if* X, Y *are independent, their covariance is* 0;
(5) *the variance of a finite sum of independent variables is the sum of the variances.*

The *correlation coefficient* of X, Y is

$$\rho = \frac{\operatorname{Cov}(X,Y)}{\big(\operatorname{Var}(X)\ \operatorname{Var}(Y)\big)^{1/2}}.$$

Lemma 13.2. *ρ belongs to the closed unit disc (it is in $[-1,1]$ if X, Y are real-valued), and equals 0 if X, Y are independent.*

Proof.
$$\left|\operatorname{Cov}(X,Y)\right| = \left|E\big(\overline{(X-\mu)}(Y-\nu)\big)\right| \leq \left[E(|X-\mu|^2)\,E(|Y-\nu|^2)\right]^{1/2}$$
by Cauchy-Schwarz. □

Definition 13.3. Let $X = (X_1, \ldots, X_d) : \Omega \to \mathbb{C}^d$ be a r.v., with all components in $L_2(P)$. Its *covariance matrix* is the hermitian-symmetrix matrix Σ^2 whose ijth entry is $\operatorname{Cov}(X_i, X_j)$.

Lemma 13.4. *Let $S : \mathbb{C}^d \to \mathbb{C}^m$ be a linear map. Then the covariance matrix of $S \circ X$ is $\overline{S}\Sigma^2 S^\top$. In particular, Σ^2 is positive semidefinite.*

14. Densities

Lemma 14.1. *Let $m : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ be in $L_1(\lambda)$, with $\int m \,\mathrm{d}\lambda = 1$. Then the function $\mu_m : \mathcal{B} \to [0,1]$ defined by*
$$\mu_m(A) = \int_{\mathbb{R}^d} \mathbb{1}_A \, m \,\mathrm{d}\lambda$$
is a probability. The map $m \mapsto \mu_m$ is injective.

If μ is in the range of the above map, then we say that μ *has a density* (which is unique).

Lemma 14.2. *Let $f : \mathbb{R}^d \to \mathbb{C}$ be measurable. Then $f \in L_1(\mu_m)$ iff $fm \in L_1(\lambda)$. If this happens, then*
$$\int_{\mathbb{R}^d} f \,\mathrm{d}\mu_m = \int_{\mathbb{R}^d} fm \,\mathrm{d}\lambda.$$

Let $A : \mathbb{R} \to \mathbb{R}$ be a homeomorphism, μ a probability on $\mathbb{R}$ with repartition function M. Then $A_*\mu$ has repartition function $M \circ A^{-1}$ if A is increasing, and $1 - M(A^{-1}x - 0)$ if A is decreasing.

Let $T : I \to \mathbb{R}$, with I an interval in $\mathbb{R}$ and T piecewise C^1 and strictly monotone. Let μ be the probability on I induced by the continuous density $m : I \to \mathbb{R}_{\geq 0}$.

Theorem 14.3. *Under the above hypotheses $T_*\mu$ has a density $\mathcal{L}m$, which is explicitly given by*
$$\begin{aligned}(\mathcal{L}m)(y) &= \sum\left\{\frac{m(x)}{|T'(x)|} : x \in T^{-1}\{y\}\right\}\\ &= \sum\{|A'(y)|\, m(A(y)) : A \textit{ is an inverse branch of } T \textit{ and } y \in \operatorname{dom}(A)\}.\end{aligned}$$

The map $\mathcal{L}$ is the *Ruelle-Perron-Frobenius operator*, or *transfer operator*; more generally, Theorem 14.3 holds for $m \in L_1(\lambda)$.

Theorem 14.4. *Let $T : O \to \mathbb{R}^d$, with O open in $\mathbb{R}^d$ and T injective C^1 with never 0 jacobian determinant j_T. Assume that the probability μ on O is induced by the density $m \in L_1(\lambda)$, and let $A = T^{-1}$. Then $T_*\mu$ is induced by a density, which is explicitly given by*
$$\mathcal{L}m = \frac{m}{|j_T|} \circ T^{-1} = |j_A|(m \circ A)$$
on $T[O]$, and 0 otherwise. There's an analogous statement for the piecewise case.

15. Marginals

Definition 15.1. If P is a probability on $(\Omega \times \Omega', \mathcal{F} \times \mathcal{F}')$, then $\pi_{1*}P$, $\pi_{2*}P$ are the *marginals* of P. Conversely, any P that projects to a given pair P_1, P_2 is a *joining* of P_1 and P_2.

Theorem 15.2. *Let μ be a probability on $\mathbb{R}^2$ with marginals μ_1, μ_2, and assume that μ has a density m.*

(1) *μ_1 and μ_2 have densities, explicitly given by*

$$m_1(x) = \int_{\mathbb{R}} m(x,-)\,\mathrm{d}\lambda,$$
$$m_2(y) = \int_{\mathbb{R}} m(-,y)\,\mathrm{d}\lambda.$$

(2) *π_1 and π_2 are μ-independent iff $m(x,y) = m_1(x)m_2(y)$ in $L_1(\lambda^2)$.*

(3) *The set $A = \{a \in \mathbb{R} : m_1(a) \neq 0, +\infty\}$ has full μ_1-measure, and parametrizes a family of densities on $\mathbb{R}$, namely*

$$m(y|a) = \frac{m(a,y)}{m_1(a)}.$$

(4) *For every $a \in A$, let $(\mu|a)$ be the pushforward via ι_a of the probability on $\mathbb{R}$ of density $m(-|a)$; in other words,*

$$(\mu|a)(B) = \int_{\mathbb{R}} \mathbb{1}_B(a,y)m(y|a)\,\mathrm{d}\lambda(y).$$

We then have

$$\mu = \int_{\mathbb{R}} (\mu|x)\,\mathrm{d}\mu_1(x), .$$

which is the continuous version of the first Bayes identity.

16. The gaussian in $\mathbb{R}^d$

The standard normal distribution is the distribution of a random variable $Z = (Z_1, \dots, Z_d) : \Omega \to \mathbb{R}^d$ with $Z_1, \dots, Z_d$ independent standard normals. In other words, Z_*P is induced by the density

$$m(x_1, \dots, x_d) = \frac{1}{\sqrt{(2\pi)^d}} \exp\left(-\frac{1}{2}(x_1^2 + \cdots + x_d^2)\right) = \frac{1}{\sqrt{(2\pi)^d}} \exp\left(-\frac{1}{2}\langle x, x\rangle\right).$$

Applying $T(x) = Sx + \mu : \mathbb{R}^d \to \mathbb{R}^d$ we get the general case $T \circ Z \in Normal(\mu, \Sigma^2)$. It has density

$$m(x) = \frac{1}{\sqrt{(2\pi)^d \det(\Sigma^2)}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-2}(x-\mu)\right),$$

where $\Sigma^2 = SS^T$ is the covariance matrix, which is positive definite.

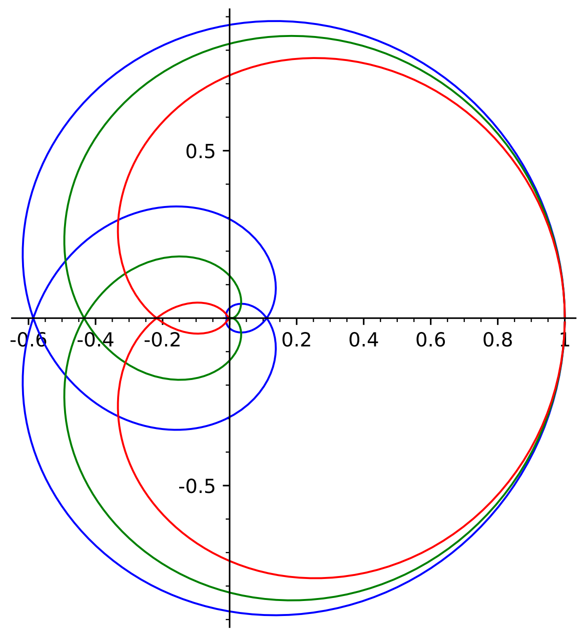

FIGURE 6

17. CHARACTERISTIC FUNCTIONS

Definition 17.1. Let μ be a probability on $\mathbb{R}^d$. Its *Fourier transform* is the function $\hat{\mu} : \mathbb{R}^d \to \mathbb{C}$ given by

$$\hat{\mu}(s) = \int_{\mathbb{R}^d} \exp\bigl(i\langle s, x\rangle\bigr)\, \mathrm{d}\mu(x).$$

If $\mu = X_* P$, then $\hat{\mu}$ is the *characteristic function* of X, written

$$\varphi_X(s) = E\bigl(\exp(i\langle s, X\rangle)\bigr).$$

If X is $\mathbb{Z}_{\geq 0}$-valued, then $\varphi_X(s) = G_X\bigl(\exp(is)\bigr)$, which is 2π-periodic.

Theorem 17.2. *$\hat{\mu}$ is uniformly continuous, bounded by $\mathbb{1}$, and $\hat{\mu}(0) = 1$.*

Theorem 17.3. (1) $\varphi_{aX} = \varphi_X(a-)$;
(2) $\varphi_{-X} = \overline{\varphi_X}$;
(3) *if $X = -X$ in law, then φ_X is $\mathbb{R}$-valued*;
(4) *if X and Y are independent, then $\varphi_{X+Y} = \varphi_X \varphi_Y$.*

It is not true that $\varphi_{X+Y} = \varphi_X \varphi_Y$ implies that X and Y are independent [JP03, p. 113].

Example 17.4. If X is $\mathit{Uniform}(-a, a)$, then $\varphi_X(s) = \sin(as)/(as)$.

Theorem 17.5 (The Theorem of Moments)**.** *Let $X : \Omega \to \mathbb{R}$ be in $L_m(P)$, for some $m \geq 1$. Then $\varphi_X \in C^m(\mathbb{R})$ and*

$$\varphi_X^{(m)}(s) = E\bigl((iX)^m \exp(isX)\bigr).$$

In particular, $E(X^m) = (-i)^m \varphi_X^{(m)}(0)$.

It has a d-dimensional version.

Theorem 17.6. *Let $X : \Omega \to \mathbb{R}^d$, and let $m \geq 1$ be such that, for every $1 \leq r \leq m$ and every $j_1, \dots, j_r$, the product $X_{j_1} \cdots X_{j_r}$ is in $L_1(P)$. Then $\varphi_X \in C^m(\mathbb{R}^d)$ and*

$$(\partial_{s_{j_1}} \cdots \partial_{s_{j_m}} \varphi_X)(s) = E\big((iX_{j_1}) \cdots (iX_{j_m}) \exp(i\langle s, X\rangle)\big).$$

Theorem 17.7. *The map $\mu \mapsto \hat{\mu}$ is injective.*

Proof. [JP03, Theorem 14.1] □

Corollary 17.8.
1. $\varphi_X = \varphi_Y$ *iff X and Y have the same law.*
2. $X = -X$ *in law iff φ_X is $\mathbb{R}$-valued*;

Lemma 17.9. *Let X be a standard normal. Then $\varphi_X(s) = \exp(-s^2/2)$.*

18. Convolutions

Definition 18.1. Let $\alpha : M \times X \to X$ be a left action of a monoid M on a set X. Let μ and ν be probabilities on M and X, respectively. Then $\mu * \nu = \alpha_*(\mu \times \nu)$ is the *convolution product* of μ and ν, which yields a left action of $\mathcal{P}(M)$ on $\mathcal{P}(X)$.

The main example is that of a group G acting on itself by left translations; we will deal with the case $G = (\mathbb{R}, +)$ only.

Theorem 18.2.
1. $\widehat{\mu * \nu} = \hat{\mu}\hat{\nu}$.
2. *If X, Y are independent then $(X_*P) * (Y_*P) = (X + Y)_*P$, and therefore $\varphi_{X+Y} = \varphi_X\varphi_Y$.*

Theorem 18.3. *Let μ, ν be probabilities on $(\mathbb{R}, \mathcal{B})$, and assume that μ has density m. Then $\mu * \nu$ has density r, which is given by*

$$r(z) = \int_{\mathbb{R}} m(z - y)\, \mathrm{d}\nu(y).$$

If ν has density n as well, then

$$r(z) = \int_{\mathbb{R}} m(z - y)n(y)\, \mathrm{d}\lambda(y),$$

which defines the convolution $r = m * n$ *of the densities m and n.*

19. Convergence of r.v.s

Definition 19.1. Let $f, f_0, f_1, f_2, \ldots : (X, \mathcal{X}, \mu) \to \mathbb{C}$ be measurable. If, for every $\varepsilon > 0$, $\mu(|f_n - f| > \varepsilon)$ converges to 0, then we say that f_n converges to f *in measure*.

Lemma 19.2. *Let $h : \mathbb{R}_{\geq 0} \to [0, M]$ be such that $h(0) = 0$, h is continuous nondecreasing, and strictly increasing in some right neighborhood of 0. Then $X_n \to X$ in probability iff $E\big(h \circ |X_n - X|\big) \to 0$.*

Theorem 19.3. *Fix $1 \leq p < \infty$.*
1. *If X_n converges to X either a.e. or in L_p, then it converges in probability.*
2. *If X_n converges to X in probability, then there exists a subsequence that converges a.e.*

(3) *If X_n converges to X in probability, and there exists $0 \leq Y \in L_p$ that dominates every X_n, then X_n converges to X in L_p.*

Given a stochastic process $X_0, X_1, \ldots$, we set

$$S_n = \sum_{k<n} X_k, \quad A_n = \frac{S_n}{n}.$$

Remark 19.4. The process $S_0, S_1, \ldots$ is the *random walk* induced by X on the group $\mathbb{R}$. More generally, if $X_0, X_1, \ldots$ take values in a not necessarily commutative semigroup, then $L_n = X_{n-1} \cdots X_0$ is the induced *left random walk* and $R_n = X_0 \cdots X_{n-1}$ the right one.

Theorem 19.5 (The Weak Law of Large Numbers)**.** *Assume the process is i.i.d., with all variables in $L_2(P)$; let $\mu = E(X_0)$. Then $A_n \to \mu \mathbb{1}$ in $L_2(P)$ and in probability.*

20. Weak convergence

Let $\mu, \mu_0, \mu_1, \ldots$ be probabilities on $\mathbb{R}^d$. Let $C_b(\mathbb{R}^d)$ be the set of all continuous bounded functions from $\mathbb{R}^d$ to $\mathbb{R}$ (or $\mathbb{C}$). If, for every $f \in C_b(\mathbb{R}^d)$,

$$\int f \, \mathrm{d}\mu_n \to \int f \, \mathrm{d}\mu,$$

then we say that μ_n converges to μ *weakly*. If $(X_n)_* P$ converges to $X_* P$ weakly, then we say that X_n converges to X *weakly* (or *in distribution*, or *in law*); this amounts to $E(f \circ X_n) \to E(f \circ X)$, for every $f \in C_b(\mathbb{R}^d)$.

Remark 20.1. One can replace $C_b(\mathbb{R}^d)$ with its subset of bounded Lipschitz functions [JP03, Theorem 18.7], or with any subset whose $\mathbb{R}$-span is uniformly dense.

Example 20.2. A sequence $r_0, r_1, \ldots$ in $[0,1]$ is *uniformly distributed* w.r.t. the Lebesgue measure λ if the sequence of Cesàro averages $n^{-1} \sum_{k=0}^{n-1} \delta_{r_k}$ converges weakly to λ. The basic example is $r_n = \alpha_1 n + \alpha_0 \pmod 1$, with α_1, α_0 real numbers and α_1 irrational. This is a first instance of the Weyl equidistribution theorem, and can be proved by using as test functions the family $\chi_k = \exp(2\pi i k -)$, for $k \in \mathbb{Z}$, whose $\mathbb{C}$-span is uniformly dense in $C(\mathbb{R}/\mathbb{Z}, \mathbb{C})$.

Theorem 20.3. (1) *Convergence in probability implies weak convergence.*
(2) *Weak convergence to a constant implies convergence in probability.*

Theorem 20.4. *Let $\mu, \mu_0, \mu_1, \ldots$ be probabilities on $\mathbb{R}$, and let $M, M_0, M_1, \ldots$ be their repartition functions. Then $\mu_n \to \mu$ weakly iff $M_n \to M$ at every point at which M is continuous.*

Theorem 20.5. *Let $\mu, \mu_0, \mu_1, \ldots$ be probabilities on a finite or countable space, and let $m, m_0, m_1, \ldots$ be their discrete-density functions. Then $\mu_n \to \mu$ weakly iff $m_n \to m$ pointwise.*

Example 20.6. Let p_n go to 0 as n goes to infinity, in such a way that np_n converges to some constant $\mu \in \mathbb{R}_{>0}$. Then the sequence $Bin(n, p_n)$ weakly converges to $Poisson(\mu)$. Thus, a binomial with large n and small p can be effectively (i.e., the discrete-density function is more manageable) approximated by a Poisson of parameter np.

Definition 20.7. A family $\{\mu_i\}_{i \in I}$ of probabilities on $\mathbb{R}^d$ is *tight* if for every $\varepsilon > 0$ there exists a compact $K \subset \mathbb{R}^d$ such that, for every i, $\mu_i(K^c) < \varepsilon$.

Theorem 20.8 (The Helly Selection Theorem)**.** *Every sequence extracted from a tight family contains a weakly converging subsequence.*

Theorem 20.9 (Slutski's Theorem)**.** *Let $X_0, X_1, \dots, Y_0, Y_1, \dots, Z : \Omega \to \mathbb{R}^d$. Let $\| \quad \|$ be any norm on $\mathbb{R}^d$. Assume that $X_n \to Z$ weakly and $\|X_n - Y_n\| \to 0$ in probability. Then $Y_n \to Z$ weakly.*

Theorem 20.10 (The Lévy Continuity Theorem)**.** *Let $(\mu_n)_{n<\omega}$ be a sequence of probabilities on $\mathbb{R}^d$. Then*:

(1) *if $\mu_n \to \mu$ weakly, then $\hat{\mu}_n \to \hat{\mu}$ everywhere* (*actually, uniformly on compacta*);
(2) *if $\hat{\mu}_n \to f$ everywhere and f is continuous at 0, then $f = \hat{\mu}$ for some μ and $\mu_n \to \mu$ weakly.*

21. The Strong Law of Large Numbers

Theorem 21.1. *Assume $X_0, X_1, \dots$ is i.i.d., with all variables in $L_2(P)$; let $\mu = E(X_0)$. Then $A_n \to \mu\mathbb{1}$ a.e.*

This yields, e.g., Monte Carlo integration and Borel's normal number theorem.

22. The Central Limit Theorem

Theorem 22.1. *Let $X_0, X_1, \dots : \Omega \to \mathbb{R}$ be i.i.d., with variables in $L_2(P)$ having mean μ and variance $\sigma^2 > 0$. Let*

$$Y_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{A_n - \mu}{\sigma/\sqrt{n}}.$$

Then Y_n converges weakly to a standard normal.

Example 22.2. We want to approximate $\pi/4$, namely the area of a disk of radius $1/2$ inscribed in the unit square, by firing bullets randomly. We need to compute the minimum n such that with probability 99% our shooting sequence will approximate the area up to $d \geq 1$ correct decimal places. We have $\mu = \pi/4$, $\sigma^2 = \mu - \mu^2$, and thus require

$$P(|A_n - \mu| < 10^{-d}) = P\left(\frac{|A_n - \mu|}{\sigma/\sqrt{n}} < \frac{10^{-d}}{\sigma/\sqrt{n}}\right) \approx P(|Z| < (\sqrt{n}/\sigma)10^{-d}) \geq 0.99.$$

The equality $\approx$ above is not mathematically rigorous, but works well in practice. By looking at tables, or using a calculator, this happens for $(\sqrt{n}/\sigma)10^{-d} > \sqrt{2}{\cdot}1.183$, i.e., $n > 2\sigma^2 1.83^2 100^d$; thus 11289 bullets for 2 digits.

Example 22.3. Let the functions $X_2, X_3, X_5, \dots : \mathbb{N} \to \mathbb{R}_{\geq 0}$ be defined by $X_p(n) = 1$ if $p \mid n$, and 0 otherwise. It is an astonishing fact that the X_ps behave as if they were an independent process. This provides heuristics for facts that can often be proved by non-probabilistic means; here's an easy example and a difficult one.

(1) Fix $d \geq 2$, let l be a large number and let $(a_1, \dots, a_d)$ vary in $\{1, \dots, l\}^d$. For $p \leq l$ we have $P(X_p(a_1) = \dots = X_p(a_d) = 1) = p^{-d}$, . The above heuristics gives then

$$P(a_1, \dots, a_d \text{ are relatively prime}) = \prod_{p \leq l} \bigl(1 - p^{-d}\bigr),$$

which tends to $\zeta(d)^{-1}$ for l tending to infinity.

(2) Let $\omega(n) = \sum_p X_p(n)$ be the number of prime divisors of n, and let m be the density function of a standard normal. Then the Erdös-Kac theorem says that, for every a,

$$\lim_{l\to\infty} \frac{1}{l} \sharp\left\{ n \le l : \frac{\omega(n) - \log\log l}{\sqrt{\log\log l}} \le a \right\} = \int_{-\infty}^{a} m(x)\,\mathrm{d}x.$$

Thus if we pick a number at random in $\{1, \dots, 10^{10000}\}$, the number of its prime factors will be gaussian distributed with average and variance 10.044. Therefore, since 90% of the mass of a standard normal lies between -1.644 and 1.644, with probability 90% the number of distinct prime factors of our number will be between 4.83 and 15.25.

References

[Bil95] P. Billingsley. *Probability and measure*. Wiley, third edition, 1995.
[JP03] J. Jacod and P. Protter. *Probability essentials*. Springer, second edition, 2003.

University of Udine
Email address: giovanni.panti@uniud.it