PROBABILITY 2019: PART 2
PROBABILITY ON THE UNIT INTERVAL

GIOVANNI PISTONE

Contents

The problem of finding a rigorous mathematical model for infinite sequences of binary independent repeated trials (0-1 outcomes) was solved by Emile Borel in the first years of XX century, by using the Lebesgue integration theory, at that time new. That solution was further generalised by A. Kolmogorov in the thirties with the use of abstract measure theory but actually the Borel solution is fully usable: an important author such as Norbert Wiener still used it in the fifties. Nowadays, it is quite common to think to all distribution as a result of a simulation, which consists precisely of transformation of the uniform distribution on the unit interval to some other distribution of interest.

Measure theory is presented in the lecture notes by Bertand Lods. A compact treatment is offered in W. Rudin [1, Ch. 11]. See also the slides of these lectures.

1. Bernoulli trials: the Borel construction

Consider the infinite sample space $S^{\mathbb{N}} = \{0,1\}^{\mathbb{N}}$. Each sample point is an infinite sequence of 0 and 1, $x = (x_1, x_2, \dots)$. The coordinate projections are denoted by X_t, $t \in \mathbb{N}$ namely, $X_t(x) = x_t$. There is a natural projection $X^t \colon S^{\mathbb{N}} \to S^t = \{0,1\}^t$, $t \in \mathbb{N}$, given by $X^t(x_1, x_2, \dots) = (x_1, x_2, \dots, x_t)$.

1.1. The Borel σ-algebra of S.

If $\mathcal{S}_t$ is the set of all sub-sets of S^t, then $\mathcal{G}_t = (X^t)^{-1}\mathcal{S}_t$ is an algebra of sub-sets of $S^{\mathbb{N}}$. It holds $\mathcal{G}_s \subset \mathcal{G}_t$ if $s \leqslant t$. The union of all $\mathcal{G}_t$ is a field contained in the set of all sub-sets of $S^{\mathbb{N}}$, which is a σ-field. The intersection of all σ-fields of $S^{\mathbb{N}}$ that contain all $\mathcal{G}_t$ is a σ-field denoted $\mathcal{G}_\infty$, the Borel σ-field of $S^{\mathbb{N}}$. The Bernoulli measurable space is $(S^{\mathbb{N}}, \mathcal{G}_\infty)$.

Exercise 1. This exercise shows why we look for a σ-algebra. For each given $x \in S^{\mathbb{N}}$, define the sequence of frequencies $f_n(x) = \frac{1}{t}\sum_{t=1}^{n}(x_t = 1)$. The real sequence of frequencies $(f_n(x))_{n\in\mathbb{N}}$ either converges to a real number in $[0,1]$ or is oscillating. If it is oscillating, there exist natural numbers $a, b \in \mathbb{N}$, $a < b$, such that the sequence is infinitely often

Date: DRAFT March 20, 2019.

above $1/a$ and infinitely often below $1/b$. The set $\left\{x \in S^{\mathbb{N}} \middle| f_m(x) \geqslant 1/a\right\}$ belong to the field $\mathcal{S}_n$. The set where the inequality holds infinitely often is the set

$$\bigcap_{n\in\mathbb{N}} \bigcup_{m\geqslant n} \left\{x \in S^{\mathbb{N}} \middle| f_m(x) \geqslant 1/a\right\}$$

which belongs to $\mathcal{G}_\infty$. Same for the other bound. In conclusion, the set of non-convergence is

$$\bigcup_{a,b\in\mathbb{N},a<b} \left(\left(\bigcap_{n\in\mathbb{N}} \bigcup_{m\geqslant n} \left\{x \in S^{\mathbb{N}} \middle| f_m(x) \geqslant 1/a\right\}\right) \bigcup \left(\bigcap_{n\in\mathbb{N}} \bigcup_{m\geqslant n} \left\{x \in S^{\mathbb{N}} \middle| f_m(x) \leqslant 1/b\right\}\right)\right)$$

which belongs to $\mathcal{G}_\infty$. The set of convergence is the complementary set.

Exercise 2 (The Law of Large Numbers (LLN)). Assume there exists a probability measure $\mathbb{P}$ on $(S^{\mathbb{N}}, \mathcal{G}_\infty)$. The sequence of random variables $(f_n)_{n\in\mathbb{N}}$ satisfies the LLN if there exists a random variable f_∞ such that $\mathbb{P}\left(\left\{x \in S^{\mathbb{N}} \middle| \lim_{n\to\infty} f_n(x) = f(x)\right\}\right) = 1$. The problem is well posed if the set $\left\{x \in S^{\mathbb{N}} \middle| \lim_{n\to\infty} f_n(x) = f(x)\right\}$ is measurable.

Exercise 3 (Bernoulli shift). The mapping $T\colon S^{\mathbb{N}} \to S^{\mathbb{N}}$ defined by $T(x_1, x_2, x_3, \dots) = (x_2, x_3, \dots)$ is measurable. In fact, $T^{-1}(y_1, y_2, \dots) = \{(0, y_1, y_2, \dots), (1, y_1, y_2, \dots)\}$ hence $T^-1(\mathcal{G}_t) \subset \mathcal{G}_{t+1}$. The set where the LLN holds is invariant for the Bernoulli shift.

1.2. **Bernoulli measure on the Bernoulli measurable space.** Given any

$$x = (x_1, x_2, \dots) \in \{0, 1\}^\infty = S^{\mathbb{N}} ,$$

the series $\omega(x) = \sum_{t=1}^\infty x_t(1/2)^t$ is absolutely convergent to the real number ω in the interval $[0, 1]$ whose expression in base 2 is $(\omega)_2 = 0.x_1x_2\dots$ The mapping $\omega\colon S^{\mathbb{N}} \to [0, 1]$ is not injective for example, $1/2 = \sum_{k=2}^\infty (1/2)^k$. In binary notation, $.100\cdots = .011\cdots$. The partial sum $\sum_{t=1}^n x_t(1/2)^t = \frac{\sum_{t=1}^k x_t 2^k}{2^n}$ is the left approximation of $\omega(x)$ by a binary rational. It follows that the mapping is surjective because for each $\omega \in [0, 1]$ we can always construct a sequence x such that $\omega(x)$ gives the required value. The mapping ω is a random variable from $(S^{\mathbb{N}}, \mathcal{G}_\infty)$ to $([0, 1], \mathcal{B})$, where $\mathcal{B}$ is a σ-algebra containing all binary intervals. In particular, every probability measure on $(S^{\mathbb{N}}, \mathcal{G}_\infty)$ has an image in $([0, 1], \mathcal{B})$.

Exercise 4. Consider the effect of the Bernoulli shift T. As $(T(x))_t = x_{t+1}$,

$$\omega(T(x)) = \sum_{t=1}^\infty x_{t+1} \left(\frac{1}{2}\right)^t = 2 \sum_{s\geqslant 2} x_s \left(\frac{1}{2}\right)^s = 2\left(\omega(x) - \frac{x_1}{2}\right) = 2\omega(x) - X_1(x) .$$

It follows that

$$X_1(x) = 2\omega(x) - \omega(T(x)) .$$

In a similar way, one can prove that

$$X_2(x) = 2\omega(T^2(x)) - \omega(T^2(x)) ,$$

and so on.

Exercise 5 (Simulation). Let us provide the a measurable mapping $Y\colon [0,1] \to S^{\mathbb{N}}$ such that $\omega(Y(\theta)) = \theta$, $\theta \in [0,1]$. Let $H\colon \mathbb{R} \to \mathbb{R}$ be defined by $H(\theta) = 0$ if $\theta \in]-\infty, 1[$ and $H(\theta) = 1$ if $\theta \in [1/2, +\infty[$. One can check that

$$X_t(\theta) = H\left(2^t\theta - \sum_{j=1}^{t-1} X_j(\theta)2^{t-j}\right) \quad , \ t \in \mathbb{N} \ ,$$

is such that $\omega(X(\theta)) = \sum_{t=1}^{\infty} X_t(\theta)(1/2)^t = \theta$, where $X(\theta)$ is the sequence $(X_1(\theta), X_2(\theta), \dots) \in S^{\mathbb{N}}$. The set $\{X_1 = 1\}$ is the set of all $\omega \in [0,1]$ such that $2\omega \geqslant 1$ that is, $[1/2, 1]$. The set $\{X_1 = 1, X_2 = 1\}$ is the set where $2\omega \geqslant 1$ and $4\omega - 2 \geqslant 1$ that is, $[3/4, 1]$. The set $X_1 = 0, X_2 = 1$ is the set where $2\omega < 1$ and $4\omega \geqslant 1$ that is $[1/4, 1/2[$. The set $\{X_2 = 1\}$ is $[1/4, 1/2[\cup [3/4, 1]$.

Exercise 6. If $([0,1], \mathcal{F}, m)$ is the unit interval endowed with the Lebesgue measure, then

$$\mathbb{X}\colon [0,1] \ni \omega \mapsto (X_j(\omega)\colon j \in \mathbb{N})$$

is a *measurable function.* In fact, for all $t \in \mathbb{N}$ and all $x_1, \dots, x_t = 0, 1$, the counter-image of the set

$$\left\{x \in S^{\mathbb{N}} \middle| X_1(x) = x_1, \dots, X_t(x) = x_t\right\}$$

is a union of intervals. Here, measurable means that for each $B\mathcal{S}_\infty$ the set $\mathbb{X}^{-1}(B)$ is measurable in $[0,1]$. As a consequence, $B \mapsto \mathbb{P}(B) = m(\mathbb{X}^{-1}(B))$ is a probability measure on the Bernoulli measurable space. This follows from the following representation of the Bernoulli trials X_t. Let us consider the function on $[0,1]$ defined by $T(\omega) = 2\omega$ on $[0, 1/2[$, $2\omega - 1$ on $[1/2, 1[$, and 0 at 1. Multiplying by 2 the equality $\omega = \sum_{t=1}^{\infty} X_t(\omega)\frac{1}{2^t}$ we get

$$2\omega = \sum_{t=1}^{\infty} X_t(\omega)\frac{1}{2^{t-1}} = \sum_{s=0}^{\infty} X_{s+1}(\omega)\frac{1}{2^s} = X_1(\omega) + \sum_{s=1}^{\infty} X_{s+1}(\omega)\frac{1}{2^s}$$

$(S^{\mathbb{N}}, \mathcal{S}_\infty, \mathbb{P})$ is the Bernoulli probability space with probability of success $1/2$.

Give any coherent family of probability measures on the sequence $S_t = \{0,1\}^t$, there exists a probability measure on $[0,1]$ whose images are the given measures.

Exercise 7 (Bernoulli trials). Given the Bernoulli probabilities on S^t, compute the first images in $[0,1]$ and their distribution functions. Provide an intuitive argument to show that the sequence of distribution functions is convergent to a distribution function.

Exercise 8 (Independence of Bernoulli trials). Show that the random variables $X_1, \dots, X_n$ are independent i.e., given functions $\phi_1, \dots, \phi_n\colon S \to \mathbb{R}$, it holds

$$\mathbb{E}_p\left[\phi_1(X_1)\cdots\phi_n(X_n)\right] = \mathbb{E}_p\left[\phi_1(X_1)\right]\cdots\mathbb{E}_p\left[\phi_n(X_n)\right] \ .$$

1.3. **Weak LLN for Bernoulli trials.** Let X_t be a sequence of Bernoulli trials with parameter p. It follows that $\mathbb{P}\left(X_t = 1\right) = p$ and $\mathbb{E}_p\left[X_t\right] = p$. Define the frequences $F_n = \frac{1}{n}\sum_{t=1}^{n} X_t$. Then $\mathbb{E}_p\left[F_n\right] = p$ and $\mathbb{E}_p\left[(F_n - p)^2\right] = \frac{p(1-p)}{n}$. We have

$$1 - \mathbb{P}_p\left(p - \epsilon \leqslant F_n \leqslant p + \epsilon\right) \leqslant \frac{1}{n\epsilon^2}p(1-p) \ ,$$

hence the limit as $n \to \infty$ of the RHS is 0.

1.4. **Strong LLN for Bernoulli trials.** Let us compute a better estimate of the probability of deviation from the mean value. For each $\beta \in \mathbb{R}$ we have

$$\mathbb{E}_p\left[\mathrm{e}^{\beta(F_n-a)}\right] = \mathrm{e}^{-\beta a}\,\mathbb{E}_p\left[\prod_{t=1}^{n} \mathrm{e}^{\beta X_t/n}\right] = \mathrm{e}^{-\beta a}\prod_{t=1}^{n}\mathbb{E}_p\left[\mathrm{e}^{\beta X_t/n}\right] = \mathrm{e}^{-\beta a}\left((1-p)+\mathrm{e}^{\beta/n}p\right)^n \ .$$

The log applied on both sides gives

$$\log \mathbb{E}_p\left[\mathrm{e}^{\beta(F_n-a)}\right] = -n\left(a\beta/n - \log\left((1-p)+\mathrm{e}^{\beta/n}p\right)\right) \ .$$

If we define

$$\kappa(a) = \sup\left\{ax - \log\left((1-p)+\mathrm{e}^{x}p\right) | x \in D\right\} \ ,$$

we have the inequality

$$\mathbb{E}_p\left[\mathrm{e}^{\beta(F_n-p)}\right] \leqslant \mathrm{e}^{-n\kappa(a)}$$

for all $\beta/n \in D$. Notice that the RHS is summable: $\sum_n \mathrm{e}^{-nh(p)} < +\infty$.

The mapping $x \mapsto h(x) = ax - \log\left((1-p)+\mathrm{e}^{x}p\right)$ has $h(0) = 0$ and has derivative

$$h'(x) = a - \frac{p\mathrm{e}^{x}}{(1-p)+\mathrm{e}^{x}p} \ ,$$

in particular $h'(0) = a - p$. The second derivative is negative i.e., the function is concave.

Let us apply the computations to the probabilities of deviations F_n from p. One case is deviation at left. Write $a = p - \epsilon$ and $\beta/n \in]-\infty, 0[= D$ to get

$$\mathbb{P}_p\left(F_n < p - \epsilon\right) = \mathbb{P}_p\left(F_n - a < 0\right) = \\ \mathbb{P}_p\left(\beta(F_n - a) > 0\right) = \mathbb{P}_p\left(\mathrm{e}^{\beta(F_n-a)} > 1\right) \leqslant \mathbb{E}_p\left[\mathrm{e}^{\beta(F_n-a)}\right] \leqslant \mathrm{e}^{-n\kappa(a)} \ .$$

Exercise 9. Conclude the argument above to prove the Strong LLN for Bernoulli trials.

2. Distribution function and quantile function

On the real measurable space $(\mathbb{R}, \mathcal{B})$ we define the *distribution function of the probability measure* μ to be the real function $\mathbb{R} \ni x \mapsto F_\mu(x) = \mu(]-\infty, x])$. The distribution function of the real random variable X is the distribution function of the induced probability measure, $F_X(x) = \mathbb{P}\left(X \leqslant x\right)$. The class of intervals $\{]-\infty, x] | x \in \mathbb{R}\}$ is closed under intersection, $]-\infty, x] \cap]-\infty, y] =]-\infty, x \wedge y]$, hence $\mu = \nu$ if $F_\mu = F_\nu$.

The distribution function of μ has the following properties: 1) $\lim_{x\to-\infty} F_\mu(x) = 0$*; 2)* $\lim_{x\to+\infty} F_\mu(x) = 1$*; 3)* $(F_\mu(y) - F_\mu(x))(y - x) \geqslant 0$*; 4)* $\lim_{y\downarrow x} F_\mu(y) = F_\mu(x)$. Notice that $\mu\{x\} = F_\mu(x) - F_\mu(x-)$.

We will show below that, conversely, *any function $F \colon \mathbb{R} \to [0, 1]$ with the properties 1) to 4) is a distribution function of a unique probability measure.*

Given any distribution function F and any real t the set $\{F \geqslant t\} = \{x \in \mathbb{R} | F(x) \geqslant t\}$ is a left-closed interval $[Q(t), +\infty[$. In fact, $F(y) \geqslant t$ implies $F(z) \geqslant t$ for all $z \geqslant y$ and the minimum of $\{x \in \mathbb{R} | F \geqslant t\}$ obtains at some $Q(t) \in \mathbb{R} \cup \{-\infty, +\infty\}$. Notice that for all $x, t \in \mathbb{R}$, $Q(t) \leqslant x$ is equivalent to $F(x) \geqslant t$ and $Q(t) < x$ is equivalent to $F(x) < t$. If $t \leqslant 0$, then $Q(t) = -\infty$; if $t > 1$ then $Q(t) = +\infty$; if $t \in]0, 1[$ then $Q(t)$ is finite; if $t = 1$ then the relevant condition is $F(x) = 1$ hence $Q(1)$ can be either $+\infty$ of finite. The restriction of Q to $]0, 1[$ is called the *quantile function of F*.

The quantile function is non-decreasing and left-continuous. In fact: 1) $s < t$ implies $\{x \in \mathbb{R} | F(x) \geqslant t\} \subseteq \{x \in \mathbb{R}\} F(x) \geqslant s$ hence $Q(s) \leqslant Q(t)$; 2) $\{t \in]0, 1[| Q(t) \leqslant x\}$ is a (relatively) closed interval.

A non-decreasing function has at most a numerable many jump points. It can be seen by evaluating the number of jumps larger than a given $\epsilon > 0$. A non decreasing function is continuous in all points except a nuberable set.

Exercise 10. As $Q(t) = \inf\{x \in \mathbb{R} | F(x) \geqslant t\}$, we have $Q(F(x)) = \inf\{y \in \mathbb{R} | F(y) \geqslant F(x)\} = x^- \leqslant x$. If F is not invertible then $x^- < x$ for some x and $F(Q(F(x))) = F(x)$. If F is invertible, then $x^- = x$ and $Q = F^{-1}$.

Let F be a distribution function with quantile function Q. Let m be the Lebesgue probability measure on $]0,1[$ and let $\mu = Q_{\#}m$ be the image of m under q. The distribution function of μ is F, $F_{Q_{\#}m} = F$. In fact, $Q^{-1}(]-\infty, x]) = \{t \in]0,1[| Q(t) \leqslant x\} = \{t \in]0,1[| F(x) \geqslant t\} =]0, F(x)]$ so that $m(Q^{-1}(]-\infty, x])) = F(x)$.

If μ is a probability measure on $(\mathbb{R}, \mathcal{B})$ and ϕ is any integrable function then

$$\int \phi(x)\ \mu(dx) = \int_0^1 \phi(Q(t))\ dt\ .$$

If moreover $F_\mu \in C^1(\mathbb{R})$ then the Change of Variable Theorem gives

$$\int \phi(x)\ \mu(dx) = \int_0^1 \phi(F_\mu^{-1}(t))\ dt = \int_{F_\mu(-\infty)}^{F_\mu(+\infty)} \phi(x) F'_\mu(x)\ dx = \int_{F_\mu(-\infty)}^{F_\mu(+\infty)} \phi(x) f_\mu(x)\ dx\ ,$$

where $f_\mu = F'_\mu$ is the *density* of μ. This is a special case of the general notion of density.

3. Weak convergence

Let μ and ν_n, $n \in \mathbb{N}$ be real probability measures with distribution functions F and F_n and quantile functions Q and Q_n, respectively. The following conditions are equivalent.

(1) $\lim_{n\to\infty} Q_n = Q$ *almost surely;*
(2) *For all continuous and bounded ϕ, $\phi \in C_b$, it holds* $\lim_{n\to\infty} \int \phi\ d\mu_n = \int \phi\ d\mu$. *This convergence is called* weak convergence, $\mu_n \to \mu$.
(3) *F_n converges to F at all continuity points of F.*

Exercise 11 (Proof). (1) $\Rightarrow$ (2): If $\phi \in C_{\mathrm{b}}$,

$$\lim_{n\to\infty} \int \phi\ d\mu_n = \lim_{n\to\infty} \int_0^1 \phi(Q_n(t))\ dt = \int_0^1 \phi(Q_n(t))\ dt = \int \phi\ d\mu$$

by bounded convergence. (2) $\Rightarrow$ (3): Given $x \in \mathbb{R}$ and $\epsilon > 0$ consider the functions $f, g \in C_{\mathrm{b}}(\mathbb{R})$ defined by

$$f(y) = \begin{cases} 1 & \text{if } y \leqslant x - \epsilon \\ -\frac{1}{\epsilon}(y - x) & \text{if } x - \epsilon < y < x \\ 0 & \text{if } y \geqslant x \end{cases} \qquad g(x) = \begin{cases} 1 & \text{if } y \leqslant x \\ -\frac{1}{\epsilon}(y - x - \epsilon) & \text{if } x < y < x + \epsilon \\ 0 & \text{if } y \geqslant x + \epsilon \end{cases}$$

As $f \leqslant \mathbf{1}_{]-\infty,x]} \leqslant g$,

$$\int f\ d\mu \leqslant F(x) \leqslant \int g\ d\mu \quad \text{and} \quad \int f\ d\mu_n \leqslant F_n(x) \leqslant \int g\ d\mu_n$$

which in turn gives $\lim_{n\to\infty} |F(x) - F_n(x)| = 0$ if $\lim_{\epsilon\to 0} \int (g - f)\ d\mu = 0$. (3) $\Rightarrow$ (1): Let s be a continuity point of Q. For each $\epsilon > 0$ choose two continuity points of F, say x, y such that $x < Q(s) < y$ and $y - x \leqslant \epsilon$. As Q is continuous at s there exists $t > s$ such that $Q(t) \leqslant y$. The first inequality is equivalent to $F(x) < s$ and the second one implies $s < F(y)$. It follows that $F_n(x) < s \leqslant F_n(y)$ that is $x < Q_n(s) \leqslant y$. It is important to remark that in the proof of (2) $\Rightarrow$ (3) we do not need the totality of function in C_{b}. We need only bounded continuous functions that separate points in the sense explained in the proof.

References

[1] Walter Rudin, *Principles of mathematical analysis*, third ed., McGraw-Hill Book Co., New York-Auckland-Düsseldorf, 1976, International Series in Pure and Applied Mathematics.

Collegio Carlo Alberto Room 203a
E-mail address: giovanni.pistone@carloalberto.org
URL: https://www.giannidiorestino.it/