Random Variables

Let Ω be any set, $\mathcal{F}$ any Sigma Field on Ω, and P any probability measure defined for each element of $\mathcal{F}$; such a triple $(\Omega, \mathcal{F}, \mathsf{P})$ is called a *probability space*. Let $\mathbb{R}$ denote the real numbers $(-\infty, \infty)$ and $\mathcal{B}$ the Borel sets on $\mathbb{R}$ generated by (for example) the half-open sets $(a, b]$.

Definition. *A real-valued Random Variable is a function $X : \Omega \to \mathbb{R}$ that is "$\mathcal{F}\backslash\mathcal{B}$-measurable," i.e., that satisfies $X^{-1}(B) = \{\omega : X(\omega) \in B\} \in \mathcal{F}$ for each Borel set $B \in \mathcal{B}$ (or, equivalently, simply for each set B of the form $(-\infty, b]$ for some rational $-\infty < b < \infty$).*

This is sometimes denoted simply "$X^{-1}(\mathcal{B}) \subset \mathcal{F}$." Since the probability measure P is only defined on sets $F \in \mathcal{F}$, a random variable *must* satisfy this condition if we are to be able to find the probability $\Pr[X \in B]$ for each Borel set B, or even if we want to find the distribution function (DF) $F_X(b) \equiv \Pr[X \le b]$ for each rational number b. Note that set-inverses are rather well-behaved functions from one class of sets to another; specifically, for any collection $\{A_\alpha\} \subset \mathcal{B}$,

$$[X^{-1}(A_\alpha)]^c = X^{-1}\big((A_\alpha)^c\big) \qquad \bigcap_\alpha X^{-1}(A_\alpha) = X^{-1}\Big(\bigcap_\alpha A_\alpha\Big) \qquad \bigcup_\alpha X^{-1}(A_\alpha) = X^{-1}\Big(\bigcup_\alpha A_\alpha\Big)$$

and thus, measurable or not, $X^{-1}(\mathcal{B})$ *is* a Sigma Field if $\mathcal{B}$ is; it is denoted $\mathcal{F}_X$ (or $\sigma(X)$), called the "sigma field generated by X," and is the smallest sigma field $\mathcal{G}$ such that X is $(\mathcal{G}\backslash\mathcal{B})$- measurable. In particular, X is $(\mathcal{F}\backslash\mathcal{B})$- measurable if and only if $\sigma(X) \subset \mathcal{F}$.

In probability and statistics, sigma field's represent *information*: a random variable Y is measurable over $\mathcal{F}_X$ if and only if the value of Y can be found from that of X, *i.e.*, if there exists some function φ such that $Y = \varphi(X)$. Note the difference in perpspective between real analysis, on the one hand, and probability/statistics, on the other; in analysis it is only *Lebesgue* measurability that mathematicians worry about, and only to avoid paradoxes and pathologies. In probability and statistics we study measurability for a variety of sigma field's, and the (technical) concept of measurability corresponds to the (empirical) notion of *observability.*

DISTRIBUTIONS.

A random variable X on a probability space $(\Omega, \mathcal{F}, \mathsf{P})$ *induces* a measure μ_X on $(\mathbb{R}, \mathcal{B})$, called the *distribution measure* (or simply the *distribution*), via the relation

$$\mu(B) = \mathsf{P}[X \in B],$$

sometimes written more succinctly as $\mu_X = \mathsf{P} \circ X^{-1}$ or even $\mathsf{P}X^{-1}$.

Functions of Random Variables

Let $(\Omega, \mathcal{F}, \mathsf{P})$ be a probability space, X a (real-valued) random variable, and $f : \mathbb{R} \to \mathbb{R}$ a (real-valued $\mathcal{B}\backslash\mathcal{B}$) measurable function. Then $Y = f(X)$ is a random variable, *i.e.*, $Y^{-1}(B) = X^{-1}(f^{-1}(B)) \in \mathcal{F}$ for any $B \in \mathcal{B}$. Also every continuous or piecewise-continuous real-valued function on $\mathbb{R}$ is $\mathcal{B}\backslash\mathcal{B}$-measurable.

Random Vectors

Denote by $\mathbb{R}^2$ the set of points (x, y) in the plane, and by $\mathcal{B}^2$ the sigma field generated by rectangles of the form $\{(x, y) : a < x \le b, c < y \le d\} = (a, b] \times (c, d]$. Note that finite unions of those rectangles form a field $\mathcal{F}_0^2$, so the minimal sigma field and minimal λ system containing $\mathcal{F}_0^2$ coincide, and the assignment $\lambda_0^2\big((a, b] \times (c, d]\big) = (b - a) \times (d - c)$ has a unique extension to a measure on all of $\mathcal{B}^2$, called two-dimensional Lebesgue measure (and denoted λ^2). Of course, it's just the area of sets in the plane.

A $\mathcal{F}\backslash\mathbb{R}^2$-measurable mapping $X : \Omega \to \mathbb{R}^2$ is called a (two-dimensional) *random vector*, or simply an $\mathbb{R}^2$-valued random variable, or (a bit ambiguously) an $\mathbb{R}^2$-RV. It's easy to show that the components X_1, X_2 of a $\mathbb{R}^2$-RV X are each RV's, and conversely that for any two random variables X_1 and X_2 the two-dimensional RV $(X, Y) : \Omega \to \mathbb{R}^2$ is $\mathcal{F}\backslash\mathbb{R}^2$-measurable, *i.e.*, is a $\mathbb{R}^2$-RV.

Also, any measurable (and in particular, any piecewise-continuous) function $f : \mathbb{R}^2 \to \mathbb{R}$ induces a random variable $f(X, Y)$: this shows that such combinations as $X + Y$, X/Y, $X \wedge Y$, $X \vee Y$, *etc.* are all measurable random variables.

The same ideas work in any finite number of dimensions, so without any special notice we will regard n-tuples $(X_1, ..., X_n)$ as $\mathbb{R}^n$-valued RV's, or $\mathcal{F}\backslash\mathcal{B}^n$-measurable functions, and will use Lebesgue n-dimensional measure λ^n on $\mathcal{B}^n$. Again $\sum_i X_i$, $\prod_i X_i$, $\min_i X_i$, and $\max_i X_i$ are all random variables.

Even if we have *infinitely many* random variables we can verify the measurability of $\sum_i X_i$, $\inf_i X_i$, and $\sup_i X_i$, and of $\liminf_i X_i$, and $\limsup_i X_i$ as well: for example,

$$[\omega : \sup_i X_i(\omega) \le r] = \bigcap_{i=1}^{\infty} [\omega : X_i(\omega) \le r]$$
$$[\omega : \limsup_i X_i(\omega) \le r] = \bigcup_{i=1}^{\infty} \bigcap_{j=i}^{\infty} [\omega : X_i(\omega) \le r].$$

The event "X_i converges" is the same as

$$[\omega : \limsup_i X_i(\omega) - \liminf_i X_i(\omega) = 0],$$

and so is $\mathcal{F}$- measurable and has a well defined probability $\mathsf{P}[\limsup_i X_i = \liminf_i X_i]$. This is one point where countible additivity (and not just finite additivity) of P is crucial, and where $\mathcal{F}$ needs to be a sigma field (and not just a F).

Example: Discrete RV's

If an RV X can take on only a finite or countable set of values, say b_i, then each set $\Lambda_i = [\omega : X(\omega) = b_i]$ must be in $\mathcal{F}$, the Λ_i are disjoint, and X can be represented in the form

$$X(\omega) = \sum_i b_i 1_{\Lambda_i}(\omega), \qquad \text{where} \tag{$*$}$$
$$1_\Lambda(\omega) = \begin{cases} 1 & \text{if } \omega \in \Lambda \\ 0 & \text{if } \omega \notin \Lambda \end{cases}$$

is the so-called *indicator function* of Λ. By including a term with $b_i = 0$, if necessary, we can assume that $\Omega = \cup \Lambda_i$ so the $\{\Lambda_i\}$ form a "countable partition" of Ω. Any RV can be approximated as well as we like by a simple RV of the form $(*)$.

EXPLICIT CONSTRUCTION OF SIGMA FIELDS (OMIT ON FIRST READING)

Ordinals and Transfinite Induction

Every finite set S (say, with $n < \infty$ elements) can be *totally ordered* $a_1 \prec a_2 \prec a_3 \prec \ldots$ in $n!$ ways, but in some sense every one of these is the same— if $\prec_1$ and $\prec_2$ are two orderings, there exists a 1–1 order-preserving isomorphism $\varphi : (S, \prec_1) \longleftrightarrow (S, \prec_2)$. Thus *up to isomorphism* there is only one ordering for any finite set.

For countably infinite sets there are many different orderings. The obvious one is $a_1 \prec a_2 \prec a_3 \prec \ldots$, ordered just like the positive integers $\mathbb{N}$; this ordering is called ω, the first *limit ordinal.* But we could pick any element (say, $b_1 \in S$) and order the remainder of S in the usual way, but declare $a_n \prec b_1$ for every $n \in \mathbb{N}$; one element is "bigger" (in the ordering) than all the others. This is *not* isomorphic to ω, and it is called $\omega+1$, the *successor* to ω. If we set aside two elements (say, $b_1 \prec b_2$) to follow all the others we have $\omega + 2$, and similarly we have $\omega + n$ for each $n \in \mathbb{N}$. The limit of all these is $\omega + \omega$, or 2ω... it is the ordering we would get if we lexicographically ordered the set $\{(i,j) : i = 1, 2 \; j \in \mathbb{N}\}$ of the first two rows of integers in the first quadrant, declaring $(1,i) \prec (2,j)$ for every i,j and otherwise $(i,j) \prec (i,k)$ if $j < k$.

We would get the successor to this, $2\omega + 1$, by extending the lexicographical ordering as we add $(3,1)$ to S; in an obvious way we get $2\omega + n$ and eventually the limit ordinals 3ω, 4ω, *etc.*, and the successor ordinals $m\omega+n$. The limit of all these is $\omega\omega$ or ω^2, the lexicographical ordering of the entire first quadrant of integers (i,j). It too has successors $\omega^2 + n$ (graphically you can think about integer triplets (i,j,k)), and limits like $\omega^2 + \omega$ and ω^3 and ω^ω (which turns out to be the same as 2^ω).

In general an ordinal is a *successor* ordinal if it has a maximal element, and otherwise is a *limit* ordinal. Every ordinal α has a successor $\alpha + 1$, and every set of ordinals $\{\alpha_n\}$ has a limit (least upper bound) λ. Let Ω be the first *un*countable ordinal.

Proofs and constructions by transfinite induction usually have one step at each successor ordinal, and another at each limit ordinal. The *Borel sets* can be defined by transfinite construction as follows. Let $\mathcal{F}_1$ be any class of subsets of some probability space $\mathcal{X}$ (perhaps $\mathcal{F}_1$ is the open sets in $\mathcal{X} = \mathbb{R}$, for example).

Succ: For any ordinal α, let $\mathcal{F}_{\alpha+1}$ be the class of countable unions of sets $E_n \in \mathcal{F}_\alpha$ and their complements $E_m : E_m^c \in \mathcal{F}_\alpha$.

Lim: For any limit ordinal λ, let $\mathcal{F}_\lambda = \cup_{\alpha \prec \lambda} \mathcal{F}_\alpha$.

Together these define $\mathcal{F}_\alpha$ for all ordinals, limit and successor; the sigma field *generated by* $\mathcal{F}_1$ is just $\mathcal{F}_\Omega$. It remains to prove that:

1. $\mathcal{F}_1 \subset \mathcal{F}_\Omega$, *i.e.*, $\mathcal{F}_\Omega$ contains the open sets;
2. $E \in \mathcal{F}_\Omega \Longrightarrow E^c \in \mathcal{F}_\Omega$, *i.e.*, $\mathcal{F}_\Omega$ is closed under complements;
3. $E_n \in \mathcal{F}_\Omega \Longrightarrow \cup_{n=1}^\infty E_n \in \mathcal{F}_\Omega$, *i.e.*, $\mathcal{F}_\Omega$ is closed under countable unions;
4. $\mathcal{F}_\Omega \subset \mathcal{G}$ for any sigma field $\mathcal{G}$ containing $\mathcal{F}_1$.

Item 1. is trivial since $\mathcal{F}_\Omega = \cup_{\alpha \prec \Omega} \mathcal{F}_\alpha$, and in particular contains $\mathcal{F}_1$. Item 2. follows by transfinite induction upon noting that $E \in \mathcal{F}_\alpha \Longrightarrow E^c \in \mathcal{F}_{\alpha+1}$. Item 3 follows by noting that $E_n \in \mathcal{F}_\Omega \Longrightarrow E_n \in \mathcal{F}_{\alpha_n}$ for some $\alpha_n \prec \Omega$, and $\beta = \sup_{n<\infty} \alpha_n$ is an ordinal satisfying $\alpha_n \preceq \beta \prec \Omega$ and hence $E_n \in \mathcal{F}_\beta$ for all n and $\cup_{n=1}^\infty E_n \in \mathcal{F}_{\beta+1}$. Verifying the minimality condition Item 4 is left as an exercise.

It isn't immediately obvious from the construction that we couldn't have stopped earlier— for example, that $\mathcal{F}_2$ or $\mathcal{F}_\omega$ isn't already the Borel sets, unchanging as we allow successively more intersections and unions. In fact that happens if the original space $\mathcal{X}$ is countable or finite; in the case of $\mathbb{R}$, however, one can show that $\mathcal{F}_\alpha \neq \mathcal{F}_{\alpha+1}$ for every $\alpha \prec \Omega$.

Do you think this explicit construction is clearer or more complicated than the completion argument used in Billingsly's book?

INFINITE COIN TOSS

For each $\omega \in \Omega = (0,1]$ and $n \in \mathbb{N}$ let $\delta_n(\omega)$ be the n^{th} *bit* in the binary expansion of ω. There's some ambiguity in the dyadic expansion of rationals... for example, one-half can be written either as $0.10b$ or as the infinitely repeating $0.01111111...b$. If we used the convention that the dyadic rationals have only finitely many 1's in their expansion (so $1/2 = 0.10b$) then $\delta_n(\omega) = \lfloor 2^n\omega \rfloor \pmod 2$; with our convention that all expansions must have infinitely many ones, we have

$$\delta_n(\omega) = (\lceil 2^n\omega \rceil + 1) \pmod 2.$$

We can think of $\{\delta_n\}$ as an infinite sequence of *random variables*, all defined on the same measurable space $(\Omega, \mathcal{B}^1)$, with the random variable δ_1 equal to zero on $(0, 1/2]$ and one on $(1/2, 1]$; δ_2 equal to zero on $(0, 1/4] \cup (1/2, 3/4]$ and one on $(1/4, 1/2] \cup (3/4, 1]$; and, in general, δ_n equal to one on a union of 2^{n-1} intervals, each of length 2^{-n} (for a total length of $1/2$), and equal to zero on the complementary set, also of length $1/2$. For the Lebesgue probability measure P on Ω that just assigns to each event $E \in \mathcal{B}^1$ its length $\mathsf{P}(E)$, we have $\mathsf{P}[X_n = 0] = \mathsf{P}[X_n = 1] = 1/2$.

Question 1: If we had used the other convention that every binary expansion must have infinitely many zero's (instead of one's), so e.g. $1/2 = 0.10b$, *then* what would the event $E_1 \equiv \{\omega : \delta_1(\omega) = 1\}$ have been?

The sigma field "generated by" any family of random variables $\{X_\alpha\}$ (whether countable or not) is defined to be the smallest sigma field for which each X_α is measurable, *i.e.*, the smallest one containing each $X_\alpha{}^{-1}(B)$ for every Borel set $B \subset \mathbb{R}$. For each fixed n the $\sigma-$algebra $\mathcal{F}_n$ generated by $\delta_1, ..., \delta_n$ is just the field $\mathcal{F}_n = \{\cup_i (a_i/2^n, b_i/2^n]\}$ consisting of all (finite) unions of left-open intervals with both endpoints an integer over 2^n. Each set in $\mathcal{F}_n$ can be specified by listing which of the 2^n intervals $(\frac{i}{2^n}, \frac{i+1}{2^n}]$ $(0 \le i < 2^n)$ it contains, so there are 2^{2^n} sets in $\mathcal{F}_n$ altogether. The union $\cup\mathcal{F}_n$ consists of all finite unions of left-open intervals with dyadic rational endpoints; it is closed under taking complements but still isn't a sigma field, since it isn't closed under taking *countable* unions and intersections; for example, it contains the set $E_n = \{\omega : \delta_n{=}1\}$ for each $n \in \mathbb{N}$ and finite intersections like $E_1 \cap ... \cap E_n = (1 - 2^{-n}, 1]$, but not the countable intersection $\cap_{n=1}^\infty E_n = \{1\}$. By definition the "join" $\mathcal{F} = \sigma(\cup_n \mathcal{F}_n) = \bigvee_n \mathcal{F}_n$ is just the smallest sigma field that contains each $\mathcal{F}_n$ (and so contains their union); this is just the familiar Borel sets in $(0, 1]$.

Lebesgue measure P, which assigns to any interval $(a, b]$ its length, is determined on each $\mathcal{F}_n$ by the rule $\mathsf{P}\big[\cup_i (a_i/2^n, b_i/2^n]\big] = \sum(b_i - a_i)2^{-n}$ or, equivalently, by the joint distribution of the random variables $\delta_1, ..., \delta_n$: independent Bernoulli's, each with $\mathsf{P}[\delta_i = 1] = 1/2$. For any number $0 < p < 1$ we can make a similar measure P_p on $(\Omega, \mathcal{F}_n)$ by requiring $\mathsf{P}_p[\delta_n = 1] = p$ and, more generally,

$$\mathsf{P}[\delta_i = d_i,\ 1 \le i \le n] = p^{\Sigma d_i}(1{-}p)^{n-\Sigma d_i};$$

the four intervals in $\mathcal{F}_2$ would have probabilities $[(1{-}p)^2,\ p(1{-}p),\ p(1{-}p),$ and $p^2]$, for example, instead of $[1/4,\ 1/4,\ 1/4,\ 1/4]$. This determines a measure on each $\mathcal{F}_n$, which extends uniquely to a measure P_p on $\mathcal{F} = \bigvee_n \mathcal{F}_n$. For $p = 1/2$ this is Lebesgue Measure, characterized by the property that $P\big[(a, b]\big] = b - a$ for each $0 \le a \le b \le 1$, but the other P_p's are new. This example (the family δ_n of random variables on the spaces $(\Omega, \mathcal{F}, \mathsf{P}_p)$) is an important one, and lets us build other important examples.

Under each of these probability distributions all the δ_n are both identically distributed and independent, *i.e.*,

$$\mathsf{P}[\delta_1 \in A_1, \ldots, \delta_n \in A_n] = \prod_{i=1}^n \mathsf{P}[\delta_1 \in A_i].$$

Any probability assignment to intervals $(a, b] \subset \Omega$ determines *some* joint probability distribution for all the $\{\delta_n\}$ (but, for most of them, the δ_n will be neither independent nor identically distributed). For any DF (*i.e.*, non-decreasing right-continuous function $F(x)$ satisfying $F(0) = 0$ and $F(1) = 1$), the prescription $\mathsf{P}_F\big((a, b]\big) \equiv F(b) - F(a)$ determines a probability distribution on every $\mathcal{F}_n$ that extends uniquely to $\mathcal{F}$, determining the joint distribution of all the $\{\delta_n\}$.

Question 2: For $F(x) = x^2$, are δ_1 and δ_2 identically distributed? Independent? Find the marginal probability distribution for each δ_n under P_F.

MEASURABILITY AND OBSERVABILITY

Fix any measure P_p on $(\Omega, \mathcal{F})$ (say, Lebesgue measure $\mathsf{P} = \mathsf{P}_{.5}$), and define a new sequence of random variables Y_n on $(\Omega, \mathcal{F}, \mathsf{P})$ by

$$Y_n(\omega) = \sum_{i=1}^{n} (-1)^{\delta_n(\omega)} = \sum_{i=1}^{n} \big(2\delta_n(\omega) - 1\big),$$

the sum of n independent terms, each ± 1 with probability 1/2 each. This is the "symmetric random walk" (it would be assymetric with P_p for $p \neq .5$), starting at the origin and moving left or right with equal probability at each step; each Y_n is $2S_n - n$ for the binomial $B(n, .5)$ random variable $S_n = \sum_{i=1}^{n} \delta_i$, the partial sums of the δ_n's.

The sigma field generated by the first n Y_i's, that generated by the first n S_i's, and that generated by the first n δ_i's are all the same, the finite field $\mathcal{F}_n$ of all unions of half-open intervals with endpoints of the form $j2^{-n}$, and a random variable Z on $(\Omega, \mathcal{F}, \mathsf{P})$ is $\mathcal{F}_n$-measurable if *and only if* Z can be written as a function $Z = \varphi_n(\delta_1, \ldots, \delta_n)$ of the first n δ's. Thus "measurability" *means* something for us— Z is **measureable** over $\mathcal{F}_n$ if and only if you can tell its value by **observing** the first n values of δ_i (or, equivalently, of Y_i or S_i). We'll see that a function Z on Ω is $\mathcal{F}$-measurable (*i.e.*, is a random variable) if and only if you can approximate it arbitrarily well by a function of the first n δ_i's, as $n \to \infty$.

UNIFORMS, NORMALS, AND MORE

From the infinite sequence of independent random bits $\{\delta_n\}$ we can construct as many random variables as we like of *any* distribution, all on the same space $(\Omega, \mathcal{F}, \mathsf{P})$, the unit interval with Lebesgue measure (length). For example, set:

$$\begin{aligned} U_1(\omega) &= \sum_{i=1}^{\infty} 2^{-i}\delta_{2^i}(\omega) \qquad & U_3(\omega) &= \sum_{i=1}^{\infty} 2^{-i}\delta_{5^i}(\omega) \\ U_2(\omega) &= \sum_{i=1}^{\infty} 2^{-i}\delta_{3^i}(\omega) \qquad & U_4(\omega) &= \sum_{i=1}^{\infty} 2^{-i}\delta_{7^i}(\omega), \end{aligned}$$

each the sum of *different* (and therefore independent) random bits; it is easy to see that $\{U_n\}$ will be independent, uniformly distributed random variables for $n = 1, 2, 3, 4$, and that we could construct as many of them as we like using successive primes $\{2, 3, 5, 7, 11, 13, \ldots\}$.

Question 3: Why did I use δ_{2^i}, δ_{3^i}, δ_{5^i}, δ_{7^i}? Give another choice that would have worked.

Let $F(x)$ be any DF (right-continuous, non-decreasing function on $\mathbb{R}$ with limits 0 and 1 $x \to -\infty$ and $x \to +\infty$, respectively) and define:

$$\begin{aligned} X_1(\omega) &= \inf[x \in \mathbb{R} : F(x) \geq U_1(\omega)] \qquad & X_3(\omega) &= \inf[x \in \mathbb{R} : F(x) \geq U_3(\omega)] \\ X_2(\omega) &= \inf[x \in \mathbb{R} : F(x) \geq U_2(\omega)] \qquad & X_4(\omega) &= \inf[x \in \mathbb{R} : F(x) \geq U_4(\omega)]; \end{aligned}$$

it's not hard to see or show (we'll do it in a week or so) that the $\{X_n\}$ are independent, each with DF $F(x) = P[X_n \le x]$.

Independent normal random variables can be constructed even more easily:

$$\begin{aligned} Z_1(\omega) &= \cos(U_1)\sqrt{-2\ln U_2} \qquad X_3(\omega) = \cos(U_3)\sqrt{-2\ln U_4} \\ X_2(\omega) &= \sin(U_1)\sqrt{-2\ln U_2} \qquad X_4(\omega) = \sin(U_3)\sqrt{-2\ln U_4}; \end{aligned}$$

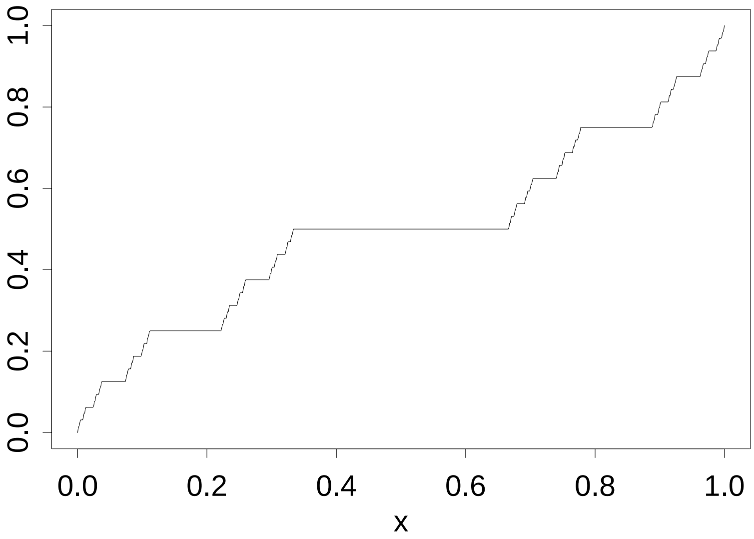
We've seen that from ordinary length measure on the unit interval (or, equivalently, from a single uniformly-distributed random variable ω) we can construct first an infinite sequence of independent $0-1$ bits δ_n; then an infinite sequence of independent uniform random variables U_n; then an infinite sequence of independent normal random variables Z_n or, more generally, random variables X_n with any distribution we choose.

The Cantor Distribution

Set $Y \equiv \sum_{n=1}^{\infty} 2\delta_n 3^{-n}$; then the ternery expansion of $y = Y(\omega)$ includes only zero's (where $\delta_n = 0$) and two's (where $\delta_n = 1$), and so lies in the Cantor set; thus Y takes on uncountably many different values, and cannot be a discrete random variable. Its CDF can be given analytically by the expression

$$F(y) = \sum_{n=1}^{\infty} \{2^{-n} : t_n > 0,\ t_m \neq 1,\ 1 <= m < n\},$$

in terms of the ternary expansion $t_n \equiv \lfloor 3^n y \rfloor \pmod 3$ of $y = \sum_{n=1}^{\infty} t_n 3^{-n}$ or graphically as



Evidently $F(x)$ has derivative $F' = 0$ wherever it is differentiable; this distribution is an example of a *singular* distribution, one that is neither absolutely continuous nor discrete.

Theorem. *Let $F(x)$ be any distribution function. Then there exist unique numbers $p_d \ge 0$, $p_c \ge 0$, $p_s \ge 0$ with $p_d + p_c + p_s = 1$ and distribution functions $F_d(x)$, $F_c(x)$, $F_s(x)$ with the properties that F_d is discrete with some probability mass function $f_d(x)$, F_c is absolutely continuous with some probability density function $f_c(x)$, and F_s is singular, satisfying $F(x) = p_d F_d(x) + p_c F_c(x) + p_s F_s(x)$ and*

$$F_d(x) = \sum_{t \le x} f_d(t), \qquad F_c(x) = \int_{t \le x} f_c(t)\, dt, \qquad F_s'(x) = 0.$$

EXPECTATION AND INTEGRAL INEQUALITIES

Discrete RV's

If a random variable Y can take on only a finite or countably infinite set of values, say b_i, then each set $\Lambda_i = [\omega : Y(\omega) = b_i]$ must be in $\mathcal{F}$; the Λ_i are disjoint, and Y can be represented in the form

$$Y(\omega) = \sum_i b_i 1_{\Lambda_i}(\omega), \qquad \text{where } 1_{\Lambda_i}(\omega) = \begin{cases} 1 & \text{if } \omega \in \Lambda_i \\ 0 & \text{if } \omega \notin \Lambda_i \end{cases} \tag{$\star$}$$

is the so-called *indicator function* of Λ_i. By adding a term with $b_i = 0$, if necessary, we can assume that $\Omega = \cup\Lambda_i$ so the $\{\Lambda_i\}$ form a "countable partition" of Ω. Any RV X can be approximated as well as we like by a simple RV of the form ($\star$) by choosing $\epsilon > 0$, setting $b_i = i\epsilon$, and

$$\Lambda_i = \{\omega : b_i \le X(\omega) < b_i + \epsilon\} \qquad X_\epsilon(\omega) = \sum_{-\infty}^{\infty} b_i 1_{\Lambda_i}(\omega)$$

It is easy to define the *expectation* of such a simple RV, or (equivalently) the *integral* of X_ϵ over $(\Omega, \mathcal{F}, \mathsf{P})$, if X is bounded below or above (to avoid indeterminate sums):

$$\mathsf{E}X_\epsilon = \int_\Omega X_\epsilon(\omega)\,\mathsf{P}(d\omega) = \int_\Omega X_\epsilon(\omega)\,d\mathsf{P}(\omega) = \int_\Omega X_\epsilon\,d\mathsf{P} = \sum_i b_i \mathsf{P}(\Lambda_i)$$

Since $X_\epsilon(\omega) \le X(\omega) \le X_\epsilon(\omega) + \epsilon$, we have $\mathsf{E}X_\epsilon \le \mathsf{E}X < \mathsf{E}X_\epsilon + \epsilon$, *i.e.*,

$$\sum_i i\epsilon \mathsf{P}[i\epsilon \le X < (i{+}1)\epsilon] \le \mathsf{E}X < \sum_i i\epsilon \mathsf{P}[i\epsilon \le X < (i{+}1)\epsilon] + \epsilon. \tag{$\star\star$}$$

This determines the value of $\mathsf{E}\,X = \int_\Omega X\,d\mathsf{P}$ for each random variable X. If we take $\epsilon = 2^{-n}$ above, and simplify the notation by writing X_n for $X_{2^{-n}} = 2^{-n}\lfloor 2^n X\rfloor$, the sequence X_n increases monotonically to X and we can define $\mathsf{E}X = \lim_n \mathsf{E}X_n$.

Note that even for $\Omega = (0,1]$, $\mathsf{P} = \lambda(dx)$ (Lebesgue measure), and X continuous, the passage to the limit suggested in ($\star\star$) is *not* the same as the limit of Riemann sums that is used to introduce integration in undergraduate calculus courses; for the Riemann sum it is the x-axis that is broken up into integral multiples of some ϵ, determining the integral of *continuous* functions, while here it is the y axis that is broken up, determining the integral of all *measurable* functions. The two definitions of integral agree for continuous functions where they are both defined, of course, but the present one is much more general.

If X is *not* bounded below or above, we can set $X^+ \equiv 0 \vee X$ and $X^- \equiv 0 \vee -X$, so that $X = X^+ - X^-$ with both X^+ and X^- bounded below (by zero), so their expectations are well-defined; if either $\mathsf{E}X^+ < \infty$ or $\mathsf{E}X^- < \infty$, we can unambiguously define $\mathsf{E}X \equiv \mathsf{E}X^+ - \mathsf{E}X^-$, while if $\mathsf{E}X^+ = \mathsf{E}X^- = \infty$ we regard $\mathsf{E}X$ as undefined.

For any measurable set $\Lambda \in \mathcal{F}$ we write $\int_\Lambda X\,d\mathsf{P}$ for $\mathsf{E}X1_\Lambda$. For $\Omega \subset \mathbb{R}$, if P gives positive probability to either $\{a\}$ or $\{b\}$ then the integrals over the sets (a,b), $(a,b]$, $[a,b)$, and $[a,b]$ may all be different; the notation $\int_a^b X\,d\mathsf{P}$ isn't expressive enough to distinguish them.

Frequently in Probability and Statistics we need to calculate or estimate integrals and expectations; usually this is done through limiting arguments in which a sequence of integrals is shown to converge to the one whose value we need. Here are some important properties of integrals for any measurable set $\Lambda \in \mathcal{F}$ and random variables $\{X_n\}$, X, Y, useful for bounding or estimating the integral of a random variable X (they're only listed here for reference and so we can talk about them— don't worry, you won't have to remember them all or know how to prove them!):

1. $\int_\Lambda X\,d\mathsf{P}$ is well-defined and finite if and only if $\int_\Lambda |X|\,d\mathsf{P} < \infty$, and $\left|\int_\Lambda X\,d\mathsf{P}\right| \le \int_\Lambda |X|\,d\mathsf{P}$. We can also define $\int_\Lambda X\,d\mathsf{P} \le \infty$ for any X bounded below by some $b > -\infty$.
2. **Lebesgue's Monotone Convergence Thm:** If $0 \le X_n \nearrow X$, then $\int_\Lambda X_n\,d\mathsf{P} \nearrow \int_\Lambda X\,d\mathsf{P} \le \infty$. In particular, the sequence of integrals converges (possibly to $+\infty$).
3. **Lebesgue's Dominated Convergence Thm:** If $X_n \to X$, and if $|X_n| \le Y$ while $\mathsf{E}Y < \infty$ for some RV $Y \ge 0$, then $\int_\Lambda X_n\,d\mathsf{P} \to \int_\Lambda X\,d\mathsf{P}$ and $\int_\Lambda |X|\,d\mathsf{P} \le \int_\Lambda Y\,d\mathsf{P} < \infty$. In particular, the sequence of integrals converges to a finite limit.
4. **Fatou's Lemma:** If $X_n \ge 0$ on Λ, then $\int_\Lambda (\liminf X_n)\,d\mathsf{P} \le \liminf \left(\int_\Lambda X_n\,d\mathsf{P}\right)$. The two sides may be unequal (example?), and the result is false for $\limsup$.
5. **Fubini's Thm:** If *either* each $X_n \ge 0$, *or* $\sum_n \int_\Lambda |X_n|\,d\mathsf{P} < \infty$, then the order of integration and summation can be exchanged: $\sum_n \int_\Lambda X_n\,d\mathsf{P} = \int_\Lambda \sum_n X_n\,d\mathsf{P}$. If these conditions fail, the orders may not be exchangeable (example?)
6. For any $p > 0$, $\mathsf{E}|X|^p = \int_0^\infty p\,x^{p-1}\mathsf{P}[|X| > x]\,dx$ and $\mathsf{E}|X|^p < \infty \Leftrightarrow \sum_{n=1}^\infty n^{p-1}\mathsf{P}[|X| \ge n] < \infty$. The case $p = 1$ is easiest and most important: if $S \equiv \sum_{n=1}^\infty \mathsf{P}[|X| \ge n] < \infty$, then $S \le \mathsf{E}|X| \le S{+}1$. If X takes on only positive integer values, $\mathsf{E}X = S$.
7. If μ_X is the distribution of X, and if f is a measurable real-valued function on $\mathbb{R}$, then $\mathsf{E}f(X) = \int_\Omega f(X(\omega))\,d\mathsf{P} = \int_\mathbb{R} f(x)\,\mu_X(dx)$ if either side exists. In particular, $\mu_X = \mathsf{E}X = \int x\,\mu_X(dx)$ and $\sigma_X^2 = \mathsf{E}(X-\mu)^2 = \int (x-\mu)^2\,\mu_X(dx)$.
8. **Hölder's Inequality:** Let $p > 1$ and $q = \frac{p}{p-1}$ (*e.g.*, $p = q = 2$ or $p = 1.01$, $q = 100$). Then $\mathsf{E}\,XY \le \mathsf{E}\,|XY| \le \left[\mathsf{E}|X|^p\right]^{\frac{1}{p}}\left[\mathsf{E}|y|^q\right]^{\frac{1}{q}}$. In particular, for $p = q = 2$,

   Cauchy-Schwartz Inequality: $\mathsf{E}\,[XY] \le \mathsf{E}\,|XY| \le \sqrt{\mathsf{E}X^2\,\mathsf{E}Y^2}$.
9. **Jensen's Inequality:** Let $\varphi(x)$ be a convex function on $\mathbb{R}$, X an integrable RV. Then $\varphi(\mathsf{E}[X]) \le \mathsf{E}[\varphi(X)]$. Examples: $\varphi(x) = |x|^p$, $p \ge 1$; $\varphi(x) = e^x$; $\varphi(x) = [0 \vee x]$.
10. **Chebychev's Inequality:** If φ is positive and increasing, then $\mathsf{P}[|X| \ge u] \le \mathsf{E}[\varphi(|X|)]/\varphi(u)$. In particular $\mathsf{P}[|X-\mu| > u] \le \frac{\sigma^2}{u^2}$ and $\mathsf{P}[|X| > u] \le \frac{\sigma^2+\mu^2}{u^2}$.