3 Random Variables & Distributions

Let \( \Omega \) be any set, \( \mathcal{F} \) any \( \sigma \)-field on \( \Omega \), and \( P \) any probability measure defined for each element of \( \mathcal{F} \); such a triple \((\Omega, \mathcal{F}, P)\) 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 1** A real-valued Random Variable is a function \( X : \Omega \to \mathbb{R} \) that is \( \mathcal{F}\mathcal{B}\)-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 \( P[X \in B] \) for each Borel set \( B \), or even if we want to find the distribution function (DF) \( F_X(b) \equiv P[X \leq b] \) 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 \( \{ A_\alpha \} \subset \mathcal{B} \),

\[
[X^{-1}(A_\alpha)]^c = X^{-1}((A_\alpha)^c) \quad \text{and} \quad \bigcup_\alpha X^{-1}(A_\alpha) = X^{-1}\left(\bigcup_\alpha A_\alpha\right)
\]

from which it follows that \( \bigcap_\alpha X^{-1}(A_\alpha) = X^{-1}\left(\bigcap_\alpha A_\alpha\right) \) and thus, measurable or not, \( X^{-1}(\mathcal{B}) \) is a \( \sigma \)-field if \( \mathcal{B} \) is. It is denoted \( \mathcal{G}_X \) (or \( \sigma(X) \)), called the “sigma field generated by \( X \),” and is the smallest sigma field \( \mathcal{G} \) such that \( X \) is \( (\mathcal{G}\setminus\mathcal{B}) \)-measurable. In particular, \( X \) is \( (\mathcal{G}\setminus\mathcal{B}) \)-measurable if and only if \( \sigma(X) \subset \mathcal{F} \).

Warning: The backslash character \( \setminus \) in this notation is entirely unrelated to the backslash character that appears in the common notation for set exclusion, \( A \setminus B \equiv A \cap B^c \).

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 \( \varphi \) such that \( Y = \varphi(X) \). Note the difference in perspective 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.

3.1 Distributions

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

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

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

3.1.1 Functions of Random Variables

Let $(\Omega, \mathcal{F}, P)$ be a probability space, $X$ a (real-valued) random variable, and $g: \mathbb{R} \to \mathbb{R}$ a (real-valued $\mathcal{B}\setminus\mathcal{B}$) measurable function. Then $Y = g(X)$ is a random variable, i.e.,

$$Y^{-1}(B) = X^{-1}(g^{-1}(B)) \in \mathcal{F}$$

for any $B \in \mathcal{B}$. Also every continuous or piecewise-continuous real-valued function on $\mathbb{R}$ is $\mathcal{B}\setminus\mathcal{B}$-measurable. How are $\sigma(X)$ and $\sigma(Y)$ related?

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

An $\mathcal{F}\setminus\mathcal{B}^2$-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 $(X_1, X_2) : \Omega \to \mathbb{R}^2$ is $\mathcal{F}\setminus\mathcal{B}^2$-measurable, i.e., is a $\mathbb{R}^2$-RV.
Also, any Borel 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 random variables if \( X \) and \( Y \) are.

The same ideas work in any finite number of dimensions, so without any special notice we will regard \( n \)-tuples \((X_1, \ldots, X_n)\) as \( \mathbb{R}^n \)-valued RV’s, or \( \mathcal{F} \setminus \mathcal{B}^n \)-measurable functions, and will use Lebesgue \( n \)-dimensional measure \( \lambda^n \) on \( \mathbb{R}^n \). Again \( \sum_i X_i \), \( \prod_i X_i \), \( \min_i X_i \), and \( \max_i X_i \) are all random variables. For any metric space \((E, \delta)\) with Borel sets \( \mathcal{E} \), a \( \mathcal{F}\setminus \mathcal{E} \)-measurable function \( X : \Omega \to E \) will be called an “\( E \)-valued random variable” (although some authors prefer the term “random element of \( E \)” unless \( E \) is \( \mathbb{R} \) or perhaps \( \mathbb{R}^n \)).

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 \( \lim \inf_i X_i \), and \( \lim \sup_i X_i \) as well: for example,

\[
[\omega : \sup_i X_i(\omega) \leq r] = \bigcap_{i=1}^{\infty} [\omega : X_i(\omega) \leq r]
\]

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

The event “\( X_i \) converges” is the same as

\[
[\omega : \lim \sup_i X_i(\omega) - \lim \inf_i X_i(\omega) = 0] = \bigcap_{k=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{i,j=n}^{\infty} [\omega : |X_i(\omega) - X_j(\omega)| < \varepsilon_k]
\]

for any positive sequence \( \varepsilon_k \to 0 \), and so is \( \mathcal{F} \)-measurable and has a well defined probability \( P[\lim \sup_i X_i = \lim \inf_i X_i] \). This is one point where countable additivity (and not just finite additivity) of \( P \) is crucial, and where \( \mathcal{F} \) must be a sigma field (and not just a field).

### 3.1.3 Example: Discrete RV’s

If a RV \( X \) can take on only a finite or countable set of distinct values, say \( \{b_i\} \), then each set \( \Lambda_i = \{\omega : X(\omega) = b_i\} \) must be in \( \mathcal{F} \); the events \( \Lambda_i \) are
disjoint; and \( X \) can be represented in the form
\[
X(\omega) = \sum_i b_i 1_{\lambda_i}(\omega), \quad \text{where} \quad (*)
\]
\[
1_{\lambda}(\omega) = \begin{cases} 1 & \text{if } \omega \in \lambda \\ 0 & \text{if } \omega \notin \lambda \end{cases}
\]
is the so-called indicator function of \( \lambda \). 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 \( \Omega \). Any RV can be approximated uniformly as well as we like by a RV of the form \((*)\). Note that the indicator function \( 1_{\lambda} \) of the limit supremum \( A \equiv \limsup \lambda_i \lambda_i \) of a sequence of events is equal pointwise to the indicator \( 1_{\lambda}(\omega) = \limsup \lambda_i 1_{\lambda}(\omega) \) of their limit supremum.

### 3.2 Explicit Construction of Sigma Fields [optional]

#### Ordinals and Transfinite Induction

Every finite set \( S \) (say, with \( n < \infty \) elements) can be totally ordered \( a_1 < a_2 < a_3 < \cdots < a_n \) 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 \( \varphi : (S, \prec_1) \leftrightarrow (S, \prec_2) \). 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 \( a_1 < a_2 < a_3 < \cdots \), ordered just like the positive integers \( \mathbb{N} \); this ordering is called \( \omega \), 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 \( a_n < b_1 \) for every \( n \in \mathbb{N} \); one element is “bigger” (in the ordering) than all the others. This is not isomorphic to \( \omega \), and it is called \( \omega + 1 \), the successor to \( \omega \). If we set aside two elements (say, \( b_1 < b_2 \)) 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 \( 2\omega \)...

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 \( (1,i) < (2,j) \) 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 \( 3\omega, 4\omega, \ldots, \) and the successor ordinals \( m\omega + n \). The limit of all these is \( \omega^2 \) or \( \omega^2 \), 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 \(\omega^3\) and \(\omega^\omega\) (which turns out to be the same as \(2^\omega\)).

In general an ordinal is a successor ordinal if it has a maximal element, and otherwise is a limit ordinal. Every ordinal \(\alpha\) has a successor \(\alpha + 1\), and every set of ordinals \(\{\alpha_n\}\) has a limit (least upper bound) \(\lambda\). Let \(\Omega\) be the first uncountable ordinal.

Proofs and constructions by transfinite induction typically 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 \(\mathcal{F}_0\) be the class of open subsets of some topological space \(X\) (perhaps the real numbers \(X = \mathbb{R}\), for example).

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

**Lim:** For any limit ordinal \(\lambda\), let \(\mathcal{F}_\lambda = \cup_{\alpha < \lambda} \mathcal{F}_\alpha\).

Together these define \(\mathcal{F}_\alpha\) for all ordinals, limit and successor; the sigma field generated by \(\mathcal{F}_0\) is just \(\mathcal{F}_\Omega\), where \(\Omega\) is the first uncountable ordinal. It remains to prove that:

1. \(\mathcal{F}_0 \subset \mathcal{F}_\Omega\), i.e., \(\mathcal{F}_\Omega\) contains the open sets (including \(X\) itself);
2. \(E \in \mathcal{F}_\Omega \implies E^c \in \mathcal{F}_\Omega\), i.e., \(\mathcal{F}_\Omega\) is closed under complements;
3. \(E_n \in \mathcal{F}_\Omega \implies \cup_{n=1}^\infty E_n \in \mathcal{F}_\Omega\), i.e., \(\mathcal{F}_\Omega\) is closed under countable unions;
4. \(\mathcal{F}_\Omega \subset \mathcal{G}\) for any sigma field \(\mathcal{G}\) containing \(\mathcal{F}_0\).

Item 1. is trivial since \(\mathcal{F}_\Omega = \cup_{\alpha < \Omega} \mathcal{F}_\alpha\), and in particular contains \(\mathcal{F}_0\). Item 2. follows by transfinite induction upon noting that \(E \in \mathcal{F}_\alpha \implies E^c \in \mathcal{F}_{\alpha+1}\). Item 3 follows by noting that \(E_n \in \mathcal{F}_\Omega \implies E_n \in \mathcal{F}_{\alpha_n}\) for some \(\alpha_n < \Omega\), and \(\beta = \sup_{n<\infty} \alpha_n\) is an ordinal satisfying \(\alpha_n \leq \beta < \Omega\) 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 \(X\) 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 \(\alpha < \Omega\).

Do you think this explicit construction is clearer or more complicated than the completion argument used in the text?
3.3 Infinite Coin Toss

For each $\omega \in \Omega = (0, 1]$ and integer $n \in \mathbb{N}$ let $\delta_n(\omega)$ be the $n^{th}$ bit in the nonterminating binary expansion of $\omega$. 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 $0.01111111b$. If we had 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

$$\delta_n(\omega) = (\lfloor 2^n \omega \rfloor + 1) \pmod{2}. \quad (2)$$

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 $\delta_1$ equal to zero on $[0, \frac{1}{2})$ and one on $[\frac{1}{2}, 1]$; $\delta_2$ equal to zero on $[0, \frac{1}{4}] \cup \left[\frac{1}{2}, \frac{3}{4}\right]$ and one on $\left[\frac{1}{4}, \frac{1}{2}\right] \cup \left[\frac{3}{4}, 1\right]$; and, in general, $\delta_n$ equal to one on a union of $2^{n-1}$ intervals, each of length $2^{-n}$ (for a total length of $\frac{1}{2}$), and equal to zero on the complementary set, also of length $\frac{1}{2}$. For the Lebesgue probability measure $P$ on $\Omega$ that just assigns to each event $E \in \mathcal{B}^1$ its length $P(E)$, we have $P[X_n = 0] = P[X_n = 1] = \frac{1}{2}$.

**Q 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? How about $E_2 \equiv \{\omega : \delta_2(\omega) = 1\}$?

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 $X_\alpha$ is measurable, i.e., the smallest one $\sigma(A)$ containing

$$A = \{X_\alpha^{-1}(B) : B \in \mathcal{B}(\mathbb{R})\}.$$ 

For each fixed $n$ the $\sigma$-algebra $\mathcal{F}_n$ generated by $\delta_1, \ldots, \delta_n$ is just the field $\mathcal{F}_n = \{\cup_i (a_i, b_i / 2^n)\}$ 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 $(\frac{i}{2^n}, \frac{i+1}{2^n})$ ($0 \leq i < 2^n$) it contains, so there are $2^{2n}$ 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 it still isn’t a sigma field, since it isn’t closed under taking countable unions and intersections; for example, it contains the set $E_n = \{\omega : \delta_n(\omega) = 1\}$ for each $n \in \mathbb{N}$ and finite intersections like $E_1 \cap \ldots \cap E_n = (1 - 2^{-n}, 1]$, but not their countable intersection $\bigcap_{n=1}^{\infty} E_n = (1 \}$. By definition the “join” $\mathcal{F} = \bigvee_n \mathcal{F}_n \equiv \sigma(\cup_n \mathcal{F}_n)$ 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 $\mathbb{P}$, which assigns to any interval $(a, b]$ its length, is determined on each $\mathcal{F}_n$ by the rule $\mathbb{P} \left[ \bigcup_i (a_i/2^n, b_i/2^n] \right] = \sum (b_i - a_i)2^{-n}$ or, equivalently, by the joint distribution of the random variables $\delta_1, \ldots, \delta_n$: independent Bernoulli’s, each with $\mathbb{P} [\delta_i = 1] = \frac{1}{2}$. For any number $0 < p < 1$ we can make a similar measure $\mathbb{P}_p$ on $(\Omega, \mathcal{F}_n)$ by requiring $\mathbb{P}_p [\delta_n = 1] = p$ and, more generally,

$$\mathbb{P} [\delta_i = d_i, \ 1 \leq i \leq n] = p^{\sum d_i} (1-p)^{n-\sum d_i}.$$ 

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

Under each of these probability distributions all the $\delta_n$ are both identically distributed and independent, i.e.,

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

Any probability assignment to intervals $(a, b] \subset \Omega$ determines some joint probability distribution for all the $\{\delta_n\}$, but typically the $\delta_n$ 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 $\mathbb{P}_F \{(a, b]\} \equiv F(b) - F(a)$ determines a probability distribution on every $\mathcal{F}_n$ that extends uniquely to $\mathcal{F}$, determining the joint distribution of all the $\{\delta_n\}$.

**Q 2:** For $F(x) = x^2$, are $\delta_1$ and $\delta_2$ identically distributed? Independent? Find the marginal probability distribution for each $\delta_n$ under $\mathbb{P}_F$.

### 3.4 Measurability and Observability

Fix any measure $\mathbb{P}_p$ on $(\Omega, \mathcal{F})$ (say, Lebesgue measure $\mathbb{P} = \mathbb{P}_{0.5}$), and define a new sequence of random variables $Y_n$ on $(\Omega, \mathcal{F}, \mathbb{P})$ by

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

Page 7
the sum of \( n \) independent terms, each \( \pm 1 \) with probability \( 1/2 \) each. This is the “symmetric random walk” (it would be assymmetric with \( p \neq 0.5 \)), 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 \( \text{Bi}(n, 0.5) \) random variable \( S_n = \sum_{i=1}^{n} \delta_i \), the partial sums of the \( \delta_i \)'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 \( n \) \( \delta_i \)'s are all the same, the finite field \( \mathcal{F}_n \) of all unions of half-open intervals with endpoints of the form \( j2^{-n} \). A random variable \( Z \) on \( (\Omega, \mathcal{F}, P) \) 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 \( n \) \( \delta \)'s. Thus “measurability” means something for us—\( Z \) is **measurable** over \( \mathcal{F}_n \) if and only if you can tell its value by **observing** the first \( n \) values of \( \delta_i \) (or, equivalently, of \( Y_i \) or \( S_i \)). We’ll see that a function \( Z \) on \( \Omega \) 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 \( n \) \( \delta_i \)'s, as \( n \to \infty \).

### 3.5 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 \( (\Omega, \mathcal{F}, P) \), the unit interval with Lebesgue measure (length). For example, set:

\[
U_1(\omega) = \sum_{i=1}^{\infty} 2^{-i} \delta_{2i}(\omega) \quad U_3(\omega) = \sum_{i=1}^{\infty} 2^{-i} \delta_{5i}(\omega)
\]

\[
U_2(\omega) = \sum_{i=1}^{\infty} 2^{-i} \delta_{3i}(\omega) \quad U_4(\omega) = \sum_{i=1}^{\infty} 2^{-i} \delta_{7i}(\omega)
\]

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 \( \{2, 3, 5, 7, 11, 13, \ldots\} \).

**Q 3:** Why did I use primes in \( \delta_{2i}, \delta_{3i}, \delta_{5i}, \delta_{7i} \)? Give another choice that would have worked.

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

\[
X_1(\omega) = \inf \{x \in \mathbb{R} : F(x) \geq U_1(\omega)\} \quad X_3(\omega) = \inf \{x \in \mathbb{R} : F(x) \geq U_3(\omega)\}
\]

\[
X_2(\omega) = \inf \{x \in \mathbb{R} : F(x) \geq U_2(\omega)\} \quad X_4(\omega) = \inf \{x \in \mathbb{R} : F(x) \geq U_4(\omega)\}
\]
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 \( F(x) = P[X_n \leq x] \). For example, we could
take \( X_n = \Phi^{-1}(U_n) \) to get independent random variables with the standard
normal distribution or \( X_n = -\log(1 - U_n) \) for the exponential distribution.
Independent normal random variables can be constructed even more efficiently via:

\[
Z_1(\omega) = \cos(2\pi U_1) \sqrt{-2 \ln U_2} \quad Z_2(\omega) = \sin(2\pi U_1) \sqrt{-2 \ln U_2}
\]

\[
Z_3(\omega) = \cos(2\pi U_3) \sqrt{-2 \ln U_4} \quad Z_4(\omega) = \sin(2\pi U_3) \sqrt{-2 \ln U_4}.
\]

We’ve seen that from ordinary length measure on the unit interval (or, equivalently,
from a single uniformly-distributed random variable \( \omega \)) we can con-
struct first an infinite sequence of independent 0–1 bits \( \delta_n \); then an in-
finite 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(s) we choose.

3.5.1 The Cantor Distribution

Set \( Y \equiv \sum_{n=1}^{\infty} 2\delta_n 3^{-n} \) for the random variables \( \delta_n(\omega) \) of Equation (2).
Then the ternary expansion of \( y = Y(\omega) \) includes only zero’s (where \( \delta_n = 0 \)) and
two’s (where \( \delta_n = 1 \)), never ones, and so \( y \) lies in the Cantor set. Since
\( Y \) takes on uncountably many different values, it cannot have a discrete
distribution. Its CDF can be given analytically by the expression

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

in terms of the ternary expansion \( t_n \equiv [3^n y] \) (mod 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, i.e., outside the Cantor set. This distribution is an example of a singular distribution, one that is neither absolutely continuous nor discrete.

**Theorem 1** Let $F(x)$ be any distribution function. Then there exist unique numbers $p_d \geq 0$, $p_{ac} \geq 0$, $p_{sc} \geq 0$ with $p_d + p_{ac} + p_{sc} = 1$ and distribution functions $F_d(x)$, $F_{ac}(x)$, $F_{sc}(x)$ with the properties that $F_d$ is discrete with some probability mass function $f_d(x)$, $F_{ac}$ is absolutely continuous with some probability density function $f_{ac}(x)$, and $F_{sc}$ is singular continuous, satisfying $F(x) = p_d F_d(x) + p_{ac} F_{ac}(x) + p_{sc} F_{sc}(x)$ and

$$F_d(x) = \sum_{t \leq x} f_d(t), \quad F_{ac}(x) = \int_{t \leq x} f_{ac}(t) \, dt, \quad F'_{sc}(x) = 0.$$

### 3.6 Expectation and Integral Inequalities

**Discrete RV’s**

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

$$Y(\omega) = \sum_i b_i \mathbf{1}_{\Lambda_i}(\omega), \quad \text{where } \mathbf{1}_{\Lambda_i}(\omega) = \begin{cases} 1 & \text{if } \omega \in \Lambda_i \\ 0 & \text{if } \omega \notin \Lambda_i \end{cases} \quad \text{(**)}$$
is the so-called *indicator function* of \( \Lambda_i \). By adding a term with \( b_i = 0 \), if necessary, we can assume that \( \Omega = \bigcup \Lambda_i \) so the \( \{ \Lambda_i \} \) form a “countable partition” of \( \Omega \). Any RV \( X \) can be approximated as well as we like by a simple RV of the form \((\ast)\) by choosing \( \epsilon > 0 \), setting \( b_i = i\epsilon \), and

\[
\Lambda_i \equiv \{ \omega : b_i \leq X(\omega) < b_i + \epsilon \} \quad X_\epsilon(\omega) = \sum_{\lambda_i} b_i 1_{\Lambda_i}(\omega) = \epsilon \left[ X(\omega)/\epsilon \right]
\]

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

\[
EX_\epsilon = \int_\Omega X_\epsilon(\omega) P(d\omega) = \int_\Omega X_\epsilon(\omega) dP(\omega) = \int_\Omega X_\epsilon dP = \sum_i b_i P(\Lambda_i)
\]

Since \( X_\epsilon(\omega) \leq X(\omega) < X_\epsilon(\omega) + \epsilon \), we have \( EX_\epsilon \leq EX < EX + \epsilon \), i.e.,

\[
\sum_i i\epsilon P[i\epsilon \leq X < (i+1)\epsilon] \leq EX < \sum_i i\epsilon P[i\epsilon \leq X < (i+1)\epsilon] + \epsilon. \quad (\ast\ast)
\]

This determines the value of \( EX = \int_\Omega X dP \) for each random variable \( X \) bounded above or below. If we take \( \epsilon = 2^{-n} \) above, and simplify the notation by writing \( X_n \) for \( X_{2^{-n}} = 2^{-n}[2^n X] \), the sequence \( X_n \) increases monotonically to \( X \) and we can define \( EX = \lim EX_n \).

Note that even for \( \Omega = (0,1], P = \lambda(dx) \) (Lebesgue measure), and \( X \) continuous, the passage to the limit suggested in \((\ast\ast)\) 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 \( \epsilon \), 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 Lebesgue integral is much more general.

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

For any measurable set \( \Lambda \in \mathcal{F} \) we write \( \int_\Lambda X dP \) for \( EX 1_\Lambda \). For \( \Omega \subset \mathbb{R} \), if \( P \) gives positive probability to either \( \{a\} \) or \( \{b\} \) then the integrals over the sets

Page 11
(a, b), (a, b), [a, b], and [a, b] may all be different, so the notation $\int_a^b X \, dP$

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 $X$. We’ll prove each of these

in class.

1. $\int_{\Lambda} X \, dP$ is well-defined and finite if and only if $\int_{\Lambda} |X| \, dP < \infty$, and

   \[
   \int_{\Lambda} X \, dP = \int_{\Lambda} |X| \, dP \quad \text{if } \int_{\Lambda} X \, dP < \infty \text{ and } \int_{\Lambda} |X| \, dP < \infty
   \]

   We can also define $\int_{\Lambda} X \, dP \leq \infty$ for any $X$ bounded below by some $b > -\infty$.

2. **Lebesgue’s Monotone Convergence Thm**: If $0 \leq X_n \nearrow X$, then

   \[
   \int_{\Lambda} X_n \, dP \nearrow \int_{\Lambda} X \, dP < \infty. \quad \text{In particular, the sequence of integrals converges (possibly to } +\infty).
   \]

3. **Lebesgue’s Dominated Convergence Thm**: If $X_n \to X$, and if $|X_n| \leq Y$ for some RV $Y \geq 0$ with $EY < \infty$, then $\int_{\Lambda} X_n \, dP \to \int_{\Lambda} X \, dP$ and $\int_{\Lambda} |X| \, dP \leq \int_{\Lambda} Y \, dP < \infty$. In particular, the sequence of integrals converges to a finite limit.

4. **Fatou’s Lemma**: If $X_n \geq 0$ on $\Lambda$, then

   \[
   \int_{\Lambda} \left( \lim \inf X_n \right) \, dP \leq \lim \inf \left( \int_{\Lambda} X_n \, dP \right).
   \]

   The two sides may be unequal (example?), and the result is false for $\lim sup$. Is “$X_n \geq 0$” necessary?

5. **Fubini’s Thm**: If either each $X_n \geq 0$, or $\sum_n \int_{\Lambda} |X_n| \, dP < \infty$, then the order of integration and summation can be exchanged: $\sum_n \int_{\Lambda} X_n \, dP = \int_{\Lambda} \sum_n X_n \, dP$. If both these conditions fail, the orders may not be exchangeable (example?)

6. For any $p > 0$, $E|X|^p = \int_0^\infty px^{p-1}P[|X| > x] \, dx$ and $E|X|^p < \infty \iff \sum_{n=1}^{\infty} n^{p-1}P[|X| \geq n] < \infty$. The case $p = 1$ is easiest and most important: if $S \equiv \sum_{n=1}^{\infty} P[|X| \geq n] < \infty$, then $S \leq E|X| < S+1$. If $X$ takes on only nonnegative integer values, $EX = S$. 

Page 12
7. If \( \mu_X \) is the distribution of \( X \), and if \( f \) is a measurable real-valued function on \( \mathbb{R} \), then \( \mathbb{E} f(X) \equiv \int_{\Omega} f(X(\omega)) d\mathbb{P} = \int_{\mathbb{R}} f(x) \mu_X(dx) \) if either side exists. In particular, \( \mu \equiv \mathbb{E} X = \int x \mu_X(dx) \) and \( \sigma^2 \equiv \mathbb{E}(X - \mu)^2 = \int (x - \mu)^2 \mu_X(dx) = \int x^2 \mu_X(dx) - \mu^2 \).

8. Hölder’s Inequality: Let \( p > 1 \) and \( q = \frac{p}{p-1} \) (e.g., \( p = q = 2 \) or \( p = 1.01, q = 101 \)). Then \( \mathbb{E} |XY| \leq |\mathbb{E}|X|^p|^{\frac{1}{p}} |\mathbb{E}|Y|^q|^{\frac{1}{q}} \). In particular, for \( p = q = 2 \), Cauchy-Schwartz Inequality: \( \mathbb{E} XY \leq \mathbb{E} |XY| \leq \sqrt{\mathbb{E} X^2 \mathbb{E} Y^2} \).

9. Minkowski’s Inequality: Let \( 1 \leq p \leq \infty \) and let \( X, Y \in L_p(\Omega, \mathcal{F}, \mathbb{P}) \). Then
\[
(\mathbb{E}|X + Y|^p)^{\frac{1}{p}} \leq (\mathbb{E}|X|^p)^{\frac{1}{p}} + (\mathbb{E}|Y|^p)^{\frac{1}{p}}
\]
so the norm \( ||X||_p \equiv (\mathbb{E}|X|^p)^{\frac{1}{p}} \) obeys the triangle inequality on \( L_p(\Omega, \mathcal{F}, \mathbb{P}) \) (what if \( p < 1 \)?)

10. Jensen’s Inequality: Let \( \varphi(x) \) be a convex function on \( \mathbb{R} \), \( X \) an integrable RV. Then \( \varphi(\mathbb{E}|X|) \leq \mathbb{E}|\varphi(X)| \). Examples: \( \varphi(x) = |x|^p \), \( p \geq 1 \); \( \varphi(x) = e^x \); \( \varphi(x) = [0 \vee x] \). The equality is strict if \( \varphi(\cdot) \) is strictly convex and \( X \) has a non-degenerate distribution.

11. Markov’s & Chebychev’s Inequalities: If \( \varphi \) is positive and increasing, then \( \mathbb{P}(|X| \geq u) \leq \mathbb{E}(|\varphi(|X|)|)/\varphi(u) \). In particular \( \mathbb{P}(|X - \mu| > u) \leq \frac{\sigma^2}{u^2} \) and \( \mathbb{P}(|X| > u) \leq \frac{\sigma^2 + \mu^2}{u^2} \).

12. One-Sided Version: \( \mathbb{P}(X > u) \leq \frac{\sigma^2}{u^2 + (\mu + t)^2} \)
\( (p:\mathbb{P}((X - \mu + t) > (u - \mu + t)) \leq ?) \)

13. Hoeffding’s Inequality: If \( \{X_j\} \) are independent and \( (\exists \{a_j, b_j\}) \) s.t. \( \mathbb{P}(a_j \leq X_j \leq b_j) = 1 \), then (\( \forall \epsilon > 0 \)), \( S_n := \sum_{j=1}^{n} X_j \) satisfies \( \mathbb{P}(S_n - \mathbb{E}S_n \geq \epsilon) \leq \exp \left( -\frac{2\epsilon^2}{\sum_{j=1}^{n} (b_j - a_j)^2} \right) \). Hoeffding proved this improvement on Chebychev’s inequality (at UNC) in 1963. See also related Azuma’s inequality (1967), Bernstein’s inequality (1937), and Chernoff bounds (1952).