A Modicum of Measure Theory
K.L. Chung, *A Course in Probability Theory, pp.* . 15—48

CONSTRUCTION & EXTENSION OF MEASURES

For any finite set $\Omega = \{\omega_1, ..., \omega_n\}$, the "power set" $P = \mathfrak{P}(\Omega)$ of *all* subsets of Ω has $|\mathfrak{P}| = 2^n$ elements; set theorists denote it $\mathfrak{P}(\Omega) = 2^\Omega$, even for infinite sets Ω. Let's consider a number of properties that classes of events (*i.e.*, sets $\mathcal{A} \subset 2^\Omega$) might have. A class $\mathcal{A}$ is called a:

FIELD if $E_1^c \in \mathcal{A}$ and $E_1 \cup E_2 \in \mathcal{A}$ whenever $E_1, E_2 \in \mathcal{A}$;

σ-ALGEBRA (or Borel field) if $E_1^c \in \mathcal{A}$ and $\cup_{i=1}^\infty E_i \in \mathcal{A}$, whenever $E_i \in \mathcal{A}$, $i \in \mathbb{N}$;

MONOTONE CLASS if $\cup_{i=1}^\infty E_i \in \mathcal{A}$, whenever $E_i \subset E_{i+1} \in \mathcal{A}$ and $\cap_{i=1}^\infty E_i \in \mathcal{A}$ whenever $E_i \supset E_{i+1} \in \mathcal{A}$, $i = 1, 2, ...$

Note that if $\mathcal{A}_\alpha$ is a (F, σA, resp. MC) for each α in any index set, then $\cap_\alpha \mathcal{A}_\alpha$ is also a (F, σA, resp. MC) (not true for unions; note we don't require that the index set be finite or even countable). Since also 2^Ω is a (F, σA, resp. MC), it follows that for any collection $\mathcal{A}_0 \subset 2^\Omega$ there exists a *smallest* (F, σA, resp. MC): namely, the intersection of all (F, σA, resp. MC)'s containing $\mathcal{A}_0$. We denote the smallest (F, σA, resp. MC) containing $\mathcal{A}_0$ by $\mathcal{F}(\mathcal{A}_0)$, $\sigma(\mathcal{A}_0)$, and $\mathcal{M}(\mathcal{A}_0)$, respectively.

For example, if Ω is arbitrary and $\mathcal{A}_0 = \{\,\{\omega\}\,\}$, the singletons, then $\mathcal{F}(\mathcal{A}_0) = \sigma(\mathcal{A}_0) = 2^\Omega$ if Ω is finite, but $\mathcal{F}(\mathcal{A}_0)$ is the finite and co-finite sets, $\sigma(\mathcal{A}_0)$ the countable and co-countable sets if Ω is infinite. $\mathcal{M}(\mathcal{A}_0)$ is just $\mathcal{A}_0$ itself!

For probability and measure theory we need probabilities to be defined for all sets in a σ-algebra $\mathcal{F}$; we'd like the luxury of defining the measure on a much smaller collection, either a field $\mathcal{F}_0$ or a collection $\mathcal{A}$ that generates a field $\mathcal{F}_0 = \mathcal{F}(\mathcal{A})$. To do this we need to know that, subject to some obvious consistency conditions, we can always *extend* a pre-measure μ_0 defined only on a field $\mathcal{F}_0$ to *some* measure μ on the σ-algebra $\mathcal{F} = \sigma(\mathcal{F}_0)$, and we need to prove that this μ is unique— *i.e.*, that, if μ_1 and μ_2 are two measures on $\mathcal{F}$ such that $\mu_1(F) = \mu_2(F)$ for $F \in \mathcal{F}_0$, then also $\mu_1(F) = \mu_2(F)$ for $F \in \mathcal{F}$, *i.e.*, μ_1 and μ_2 agree on the entire σ-algebra.

It turns out to be easier to show that μ_0 extends uniquely to the monotone class $\mathcal{M}(\mathcal{A}_0)$ than it is to show unique extension to the σ-algebra $\sigma(\mathcal{A}_0)$; luckily, when $\mathcal{A}_0$ is a *field*, these are the same:

Theorem. *Let $\mathcal{F}_0$ be a field; then $\mathcal{M}(\mathcal{F}_0) = \sigma(\mathcal{F}_0)$. (Sketch proof; set $\mathcal{C} = \{E \in \mathcal{M} : E^c \in \mathcal{M}\}$ and show $\mathcal{C}$ an MC, $\mathcal{A} \subset \mathcal{C}$, and conclude $\mathcal{M} \subset \mathcal{C} \subset \mathcal{M}$).*

How can we specify μ_0 on $\mathcal{F}_0$? Two examples:

1. $\mathcal{A} = \{\,\{\omega\}\,\}$: Given any $\{\omega_i\}$ and $\{p_i \geq 0\}$ with $\sum_i p_i = 1$, set $\mu_0(A) = \sum[p_i : \omega_i \in A]$. In fact, this is also μ; it's the only kind of discrete measure there is, and the only kind on a finite or countable set Ω.
2. $\Omega = (-\infty, \infty)$, and $\mathcal{A} = \{(a, b]\}$ for $a \leq b$, $a, b \in \mathbb{Q}$. Now $\mathcal{F}_0 = \mathcal{F}(\mathcal{A})$ consists of finite disjoint unions of left-open rational intervals, possibly augmented by sets of the form $(-\infty, b]$, (a, ∞), or $(-\infty, \infty)$. The σ-field $\sigma(\mathcal{A})$ is *not* just countable unions of such sets; it is a larger collection, called the "Borel sets" in the real line, that includes all open and closed sets, the Cantor set, and many others. It can be constructed explicitly by transfinite induction (!), but is not easily described. It includes every set of real numbers we'll need in this course.

Given any DF $F(x)$, we can define a pre-pm μ_0 on $\mathcal{A}$ by setting $\mu_0((a, b]) = F(b) - F(a)$. If $F = F_d$ is purely discontinuous this just assigns probability $p_i = F(a_i) - F(a_i-)$ to each a_i where $F(x)$ jumps; if $F(x) = F_{ac} = \int_{-\infty}^x f(t)\,dt$ is absolutely continuous this just assigns probability $\mu(A) = \int_A f(t)\,dt$ to A (and in fact this is the usual *definition* of that integral!)

How does the extension idea work? Suppose μ_0 is defined on a field $\mathcal{F}_0$, and $\mathcal{F} = \sigma(\mathcal{F}_0)$. Define two new set functions μ^* and μ_* on 2^Ω by:

$$\mu^*(E) = \inf\Big[\sum_{i=0}^{\infty} \mu_0(F_i) : \; E \subset \bigcup_{i=0}^{\infty} F_i, \; F_i \in \mathcal{F}_0\Big] \qquad\qquad \mu_*(E) = 1 - \mu^*(E^c)$$

On reflection it's clear that $\mu_*(E) \le \mu^*(E)$ for each set $E \in 2^\Omega$, and $\mu_*(E) = \mu_0(E) = \mu^*(E)$ for each set $E \in \mathcal{F}_0$; thus there is an obvious well-defined extension of μ_0 to a set function on the *μ-completion*, $\overline{\mathcal{F}}^\mu = \{E \in 2^\Omega : \mu_*(E) = \mu^*(E)\} = \{E \in 2^\Omega : \mu^*(E) + \mu^*(E^c) = 1\}$. It remains to show that: (1) The extension μ is nonnegative and countably additive on $\overline{\mathcal{F}}^\mu$ (an $\epsilon/2^n$ argument); and (2) The σ-algebra $\mathcal{F} = \sigma(\mathcal{F}_0)$ is contained in $\overline{\mathcal{F}}^\mu$ (just show that $\overline{\mathcal{F}}^\mu$ is a σ-algebra containing $\mathcal{F}_0$); and (3) The extension to $\mathcal{F}$ is unique (show $\{E \in \mathcal{F} : \mu(E) = \nu(E)\}$ is an MC containing $\mathcal{F}_0$).

RANDOM VARIABLES

Let Ω be any set, $\mathcal{F}$ any σ-algebra 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 probability $\Pr[X \le b]$ 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 $\{A_\alpha\} \subset \mathcal{B}$,

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

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

Theorem. *Let X be a real-valued random variable on some probability space $(\Omega, \mathcal{F}, \mathsf{P})$, and let $\mathcal{F}_X = \sigma(X) = X^{-1}(\mathcal{B})$. Then a random variable Y on $(\Omega, \mathcal{F}, \mathsf{P})$ is $\mathcal{F}_X$-measurable (i.e., $\mathcal{F}_Y \subset \mathcal{F}_X$) if and only if there exists a Borel-measurable real-valued function $\phi : \mathbb{R} \to \mathbb{R}$ such that $Y = \phi(X)$.*

Thus in probability and statistics, σ-algebra'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.*, $Y = \phi(X)$ for some function ϕ. 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 people are concerned about, and only to avoid paradoxes and pathologies— the whole object is to show that *every* function and set of interest are measurable, to justify the usual operations (interchanging limits, integration formulas, *etc.*). In probability and statistics, on the other hand, measurability is meaningful, and directly connected with the *observability* or *estimability* of uncertain quantities.

Filtrations

Often in Stochastic Processes we have not one σ-algebra F_X but a nested family of them, $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_3 \subset ... \subset \mathcal{F}$, like

$$\mathcal{F}_n = \sigma\{X_1, ..., X_n\},$$

the smallest σA $\subset \mathcal{F}$ for which the first n of some family of RV's is measurable or, for continuous-time processes,

$$\mathcal{F}_t = \sigma\{X_s : s \le t\}.$$

In this case a random variable Y is F_t-measurable if and only if its value can be determined by observing X_s up to time t; we will soon be concerned with the problem of making optimal predictions at time t of some future event, *i.e.*, of finding $\mathcal{F}_t$-measurable approximations to non-$\mathcal{F}_t$-measurable random variables $Y \in \mathcal{F}$.

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 succinctly as $\mu_X = \mathsf{P} \circ X^{-1}$.

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 also 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 σ-algebra 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 σ-algebra and minimal MF 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 and Y 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 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 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,

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

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 σ-algebra (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

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

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 σ-ALGEBRAS

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 $\phi : (S, \prec_1) \longleftrightarrow (S, \prec_2)$. 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 σ-algebra *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 σ-algebra $\mathcal{G}$ 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 Chung's text?

INFINITE COIN TOSS

For each $\omega \in \Omega = (0,1]$ and $n \in \mathbb{N}$ let $b_n(\omega)$ be the nth *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 dyadic rationals have finitely many 1's in their expansion (so $1/2 = 0.10b$) then $b_n(\omega) = \lfloor 2^n \omega \rfloor \pmod 2$; with Chung's convention that all expansions must have infinitely many ones, we have

$$b_n(\omega) = (\lceil 2^n \omega \rceil - 1) \pmod 2.$$

We can think of $\{b_n\}$ as an infinite sequence of *random variables*, all defined on the same measurable space $(\Omega, \mathcal{B}^1)$. For each n the σ-algebra generated by $b_1, ..., b_n$ is the same $\mathcal{F}_n$ we studied before, generated by sets of the form $(0, i/2^n]$ for integers $0 \le i \le 2^n$, and $\mathcal{B}^1 = \sigma(\cup_{n=1}^\infty \mathcal{F}_n)$.

For each $0 < p < 1$ we can define a probability measure P_p on $(\Omega, \mathcal{B}^1)$ such that $\mathsf{P}[b_n = 1] = p$ for all n with the $\{b_n\}$ all independent, *i.e.*, such that

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

For $p = 1/2$ this is Lebesgue Measure, characterized by the property that $P\big[(a,b]\big] = b - a$ for each $0 \le a \le b \le 1$. This example (the family b_n of random variables on the spaces $(\Omega, \mathcal{F}, \mathsf{P}_p)$) is an important one, and lets us build other important examples.

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 y_i, then each set $\Lambda_i = [\omega : Y(\omega) = y_i]$ must be in $\mathcal{F}$; the Λ_i are disjoint, and Y can be represented in the form

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

is the so-called *indicator function* of Λ_i. By adding a term with $y_i = 0$, 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 $x_i = i\epsilon$, and

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

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):

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

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 $X_{2^{-n}}$, 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 $X^+ = 0 \vee X$ and $X^- = 0 \vee -X$, 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 $\mathsf{E}X = \mathsf{E}X^+ - \mathsf{E}X^-$, 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:

1. $\int_\Lambda X\,d\mathsf{P}$ is well-defined if and only if $\int_\Lambda |X|\,d\mathsf{P} < \infty$, and $\left|\int_\Lambda X\,d\mathsf{P}\right| \le \int_\Lambda |X|\,d\mathsf{P}$.

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 $\int_\Lambda (\liminf X_n)\,d\mathsf{P} \le \liminf\big(\int_\Lambda X_n\,d\mathsf{P}\big)$. The two sides may be unequal (example?), and the result is false for lim sup.

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. If $S = \sum_{n=1}^\infty \mathsf{P}[|X| \ge n] < \infty$, then $S \le \mathsf{E}X \le S+1$. If X takes on only positive integer values, $\mathsf{E}X = S$. For $p > 0$, $\mathsf{E}|X|^p < \infty \Leftrightarrow \sum_{n=1}^\infty n^{p-1}\mathsf{P}[|X| \ge n] < \infty$.

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 $q = \frac{p}{p-1}$ (so $1/p + 1/q = 1$; *e.g.*, $p = q = 2$ or $p = 1.01$, $q = 101$): then $\mathsf{E}\,XY \le \mathsf{E}\,|XY| \le \big[\mathsf{E}|X|^p\big]^{\frac{1}{p}}\big[\mathsf{E}|y|^q\big]^{\frac{1}{q}}$. In particular,

   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 $\phi(x)$ be a convex function on $\mathbb{R}$, X an integrable RV. Then $\phi(\mathsf{E}[X]) \le \mathsf{E}[\phi(X)]$. Examples: $\phi(x) = |x|^p$, $p \ge 1$; $\phi(x) = e^x$; $\phi(x) = [0 \vee x]$.

10. **Chebychev's Inequality:** If ϕ is positive and increasing, then $\mathsf{P}[|X| \ge u] \le \frac{\mathsf{E}[\phi(|X|)]}{\phi(u)}$. In particular $\mathsf{P}[|X-\mu| > u] \le \frac{\sigma^2}{u^2}$ and $\mathsf{P}[|X| > u] \le \frac{\sigma^2+\mu^2}{u^2}$.