ENEE 620
FALL 2016
CHAPTER 1

Field

With Ω an arbitrary set, a non-empty collection of $\mathcal{F}$ of subsets of Ω is a *field* (also known as an *algebra*) on Ω if

(F1) $\emptyset \in \mathcal{F}$

(F2) Closed under complementarity: If $E \in \mathcal{F}$, then $E^c \in \mathcal{F}$

(F3) Closed under union: If $E \in \mathcal{F}$ and $F \in \mathcal{F}$, then $E \cup F \in \mathcal{F}$

By de Morgan's Laws, (F2) and (F3) automatically imply

(F3b) Closed under intersection: If $E \in \mathcal{F}$ and $F \in \mathcal{F}$, then $E \cap F \in \mathcal{F}$

Furthermore, (F3) implies the seemingly more general statement:

(F4) Closed under finite union: If $E_1, \ldots, E_n \in \mathcal{F}$, then $\cup_{i=1}^n E_i \in \mathcal{F}$

while (F3b) implies the seemingly more general statement:

(F4b) Closed under finite intersection: If $E_1, \ldots, E_n \in \mathcal{F}$, then $\cap_{i=1}^n E_i \in \mathcal{F}$

σ-Field

With Ω an arbitrary set, a non-empty collection of $\mathcal{F}$ of subsets of Ω is a σ-*field* (also known as an σ-*algebra*) on Ω if

(F1) $\emptyset \in \mathcal{F}$

(F2) Closed under complementarity: If $E \in \mathcal{F}$, then $E^c \in \mathcal{F}$

(F3) Closed under countable union: With I a *countable* index set, if $E_i \in \mathcal{F}$ for each $i \in I$, then $\cup_{i \in I} E_i \in \mathcal{F}$

Probability measures

Consider an arbitrary non-empty set Ω equipped with a σ-field $\mathcal{F}$. A *probability (measure)* $\mathbb{P}$ on $\mathcal{F}$ (or on $(\Omega, \mathcal{F})$) is a mapping $\mathbb{P} : \mathcal{F} \to [0,1]$ such that

(P1) $\mathbb{P}[\emptyset] = 0$ and $\mathbb{P}[\Omega] = 1$

(P2) σ-additivity: With I a countable index set, if $E_i \in \mathcal{F}$ for each $i \in I$, then

$$\mathbb{P}[\cup_{i \in I} E_i] = \sum_{i \in I} \mathbb{P}[E_i]$$

whenever the subsets $\{E_i, i \in I\}$ are *pairwise disjoint*, namely

$$E_i \cap E_j = \emptyset, \qquad \begin{array}{c} i \neq j \\ i, j \in I \end{array}$$

Probability models

A probability space (triple) is a triple $(\Omega, \mathcal{F}, \mathbb{P})$ where

- Ω is the *sample space*, i.e., the collection of all outcomes (samples) generated by the experiment $\mathcal{E}$.
- Events are collections of outcomes. The collection of events whose likelihood of occurrence can be defined is a *σ-field* $\mathcal{F}$ on Ω. In many cases of interest one is forced for mathematical reasons to take $\mathcal{F}$ to be strictly smaller than $\mathcal{P}(\Omega)$.
- The "likelihood" of occurrence to events in $\mathcal{F}$ is assigned through a *probability measure* $\mathbb{P}$ defined on $\mathcal{F}$.

Discrete probability models

A case of particular importance arises when Ω is *countable*, in which case it is customary to take $\mathcal{F} = \mathcal{P}(\Omega)$ where $\mathcal{P}(\Omega)$ denotes the power set of Ω (sometimes also denoted 2^{Ω}). In that setting, specifying $\mathbb{P}$ on $(\Omega, \mathcal{P}(\Omega))$ is equivalent to specifying

$$\{\mathbb{P}[\{\omega\}], \ \omega \in \Omega\}.$$

This is a straightforward consequence of the σ-additivity of probability measures and the fact that

$$F = \cup_{\omega \in F}\{\omega\}, \quad F \in \mathcal{P}(\Omega).$$

Uniform probability assignments

Let Ω be an arbitrary set to be used as the sample space of a probabilistic experiment $\mathcal{E}$ where outcomes are equally likely to occur – According to $\mathcal{E}$ an element of Ω is selected at *random* as the saying goes, or more accurately, uniformly.

- Uniform probability measure on a discrete set with $|\Omega| < \infty$ assigns the same probability of occurrence to any outcome. Thus, take $\mathbb{P}[\{\omega\}] = p$ for all $\omega \in \Omega$, so that

$$\mathbb{P}[F] = \sum_{\omega \in F} \mathbb{P}[\{\omega\}] = |F|p, \quad F \in \mathcal{P}(\Omega)$$

whence

$$p = \frac{1}{|\Omega|}$$

upon taking $F = \Omega$. Finally we get

$$\mathbb{P}[F] = \frac{|F|}{|\Omega|}, \quad F \in \mathcal{P}(\Omega).$$

- What happens when Ω is countable with $|\Omega| = \infty$? We should still have $\mathbb{P}[\{\omega\}] = p$ for all $\omega \in \Omega$. In that case it still follows that

$$\mathbb{P}[F] = |F|p, \quad \begin{array}{c} F \in \mathcal{P}(\Omega) \\ |F| < \infty. \end{array}$$

Therefore, we get

$$|F|p \leq 1, \quad \begin{array}{c} F \in \mathcal{P}(\Omega) \\ |F| < \infty \end{array}$$

and this implies $p = 0$ (because we can select a sequence $\{F_n,\ n = 1, 2, \ldots\}$ of subsets of Ω such that $|F_n| = n$ for all $n = 1, 2, \ldots$). A contradiction immediately arises since

$$1 = \mathbb{P}[\Omega] = \sum_{\omega \in \Omega} p = 0!$$

It is not possible to have a uniform probability measure on a discrete set with $|\Omega| = \infty$!

- What happens when Ω is uncountable? For the purpose of defining probability measures on non-countable sets Ω, in general it is not possible to take $\mathcal{F} = \mathcal{P}(\Omega)$. In other words, in the non-countable case, it is not possible to assign a likelihood of occurrence (through a probability measure satisfying the axioms (P1)-(P2)) to every subset of Ω! The difficulties involved will be on two examples to be discussed shortly, namely infinite coin tosses of a fair coin and selecting a point at random in the interval $[0, 1]$.

Simple consequences of the definitions (F1)-(F5) and (P1)-(P2)

- Complementarity:

$$\mathbb{P}[E^c] = 1 - \mathbb{P}[E], \quad E \in \mathcal{F}$$

- Generalizing additivity:

$$\mathbb{P}[E \cup F] = \mathbb{P}[E] + \mathbb{P}[F] - \mathbb{P}[E \cap F], \quad E, F \in \mathcal{F}$$

- Monotonicity:

$$\mathbb{P}[E] \leq \mathbb{P}[F], \qquad \begin{array}{c} E \subseteq F \\ E, F \in \mathcal{F} \end{array}$$

Bounds

The following elementary bounds are often used; they can be established by induction:

- Boole's inequality (also known as union bound): With countable index set I,

$$\mathbb{P}[\cup_{i \in I} E_i] \leq \sum_{i \in I} \mathbb{P}[E_i]$$

- Bonferroni's inequality: With finite index set I,

$$\mathbb{P}[\cup_{i \in I} E_i] \geq \sum_{i \in I} \mathbb{P}[E_i] - \sum_{i,j \in I:\ i<j} \mathbb{P}[E_i \cap E_j]$$

Continuity properties of $\mathbb{P}$

Consider a sequence $\{E_n,\ n = 1, 2, \ldots\}$ of events in $\mathcal{F}$.

- If the sequence is monotone increasing in the sense that

$$E_n \subseteq E_{n+1}, \quad n = 1, 2, \ldots$$

then

$$\lim_{n\to\infty} \mathbb{P}[E_i] = \mathbb{P}[\cup_{i=1}^{\infty} E_i].$$

We have a continuity result for $\mathbb{P}$ if we define $\lim_{n\to\infty} E_i \equiv \cup_{i=1}^{\infty} E_i$ in the sense that $\lim_{n\to\infty} \mathbb{P}[E_i] = \mathbb{P}[\lim_{n\to\infty} E_i]$.

- If the sequence is monotone decreasing in the sense that

$$E_{n+1} \subseteq E_n, \quad n = 1, 2, \ldots$$

then

$$\lim_{n\to\infty} \mathbb{P}[E_i] = \mathbb{P}[\cap_{i=1}^{\infty} E_i].$$

We have a continuity result for $\mathbb{P}$ if we define $\lim_{n\to\infty} E_i \equiv \cap_{i=1}^{\infty} E_i$ in the sense that $\lim_{n\to\infty} \mathbb{P}[E_i] = \mathbb{P}[\lim_{n\to\infty} E_i]$.

The sequence $\{E_n,\ n = 1, 2, \ldots\}$ is monotone increasing (resp. decreasing) if and only if the complementary sequence $\{E_n^c,\ n = 1, 2, \ldots\}$ is monotone decreasing (resp. increasing).

Independence

Consider $\{E_i,\ i \in I\}$ where I is an arbitrary index set.

- Pairwise independence: The events $\{E_i,\ i \in I\}$ are said to be *pairwise independent* if the conditions

$$\mathbb{P}[E_i \cap E_j] = \mathbb{P}[E_i]\,\mathbb{P}[E_j], \qquad \begin{array}{c} i \neq j \\ i, j \in I \end{array}$$

hold. When I is finite, this is a set of $\frac{|I|(|I|-1)}{2}$ conditions.

- Mutual independence (with I finite): The events $\{E_i,\ i \in I\}$ are said to be *mutually independent* if

$$\mathbb{P}[\cap_{j\in J} E_j] = \prod_{j\in J} \mathbb{P}[E_j], \qquad \begin{array}{c} J \subset I \\ |J| > 0 \end{array}$$

This represents $2^{|I|} - (|I| + 1)$ conditions.

- Mutual independence (with I arbitrary): The events $\{E_i,\ i \in I\}$ are said to be *mutually independent* if for each *finite* subset $J \subseteq I$ with $0 < |J| < \infty$, the events $\{E_j,\ j \in J\}$ are mutually independent, namely

$$\mathbb{P}\left[\cap_{j \in J} E_j\right] = \prod_{j \in J} \mathbb{P}\left[E_j\right], \qquad \begin{array}{c} J \subseteq I \\ 0 < |J| < \infty. \end{array}$$

Borel-Cantelli lemmas

Let $\{A_n,\ n = 1, 2, \ldots\}$ be a collection of events in $\mathcal{F}$. We write

$$[A_n \text{ i.o. }] = \cap_{n=1}^{\infty} \left(\cup_{m \geq n} A_m\right)$$

- If
$$\sum_{n=1}^{\infty} \mathbb{P}\left[A_n\right] < \infty,$$
then it is always the case that
$$\mathbb{P}\left[A_n \text{ i.o. }\right] = 0$$

- When the events $\{A_n,\ n = 1, 2, \ldots\}$ are mutually independent, if
$$\sum_{n=1}^{\infty} \mathbb{P}\left[A_n\right] = \infty,$$
then
$$\mathbb{P}\left[A_n \text{ i.o. }\right] = 1.$$

Limsup and liminf, and limits

Let $\{A_n,\ n = 1, 2, \ldots\}$ be a collection of events in $\mathcal{F}$. Define

$$\limsup_{n \to \infty} A_n = \cap_{n=1}^{\infty} \left(\cup_{m \geq n} A_m\right) = \cap_{n=1}^{\infty} \bar{A}_n$$

with

$$\bar{A}_n = \cup_{m \geq n} A_m, \quad n = 1, 2, \ldots$$

Similarly,

$$\liminf_{n \to \infty} A_n = \cup_{n=1}^{\infty} \left(\cap_{m \geq n} A_m\right) = \cup_{n=1}^{\infty} \underline{\mathbf{A}}_n$$

with
$$\underline{\mathbf{A}}_n = \cap_{m \geq n} A_m, \quad n = 1, 2, \ldots$$

We have the memnonic notation

$$\limsup_{n \to \infty} A_n = [A_n \text{ infinitely often (i.o.) }]$$

and
$$\liminf_{n \to \infty} A_n = [\text{ eventually all } A_n]$$

Obviously, for each $n = 1, 2, \ldots$ we have

$$\underline{\mathbf{A}}_n \subseteq \bar{A}_n$$

with
$$\bar{A}_{n+1} \subseteq \bar{A}_n \quad \text{[Monotone decreasing]}$$

and
$$\underline{\mathbf{A}}_n \subseteq \underline{\mathbf{A}}_{n+1} \quad \text{[Monotone increasing]}$$

By continuity of $\mathbb{P}$ it follows that

$$\mathbb{P}\left[\limsup_{n \to \infty} A_n\right] = \lim_{n \to \infty} \mathbb{P}\left[\bar{A}_n\right]$$

and
$$\mathbb{P}\left[\liminf_{n \to \infty} A_n\right] = \lim_{n \to \infty} \mathbb{P}\left[\underline{\mathbf{A}}_n\right]$$

The collection $\{A_n, \ n = 1, 2, \ldots\}$ will be said to converge if

$$\limsup_{n \to \infty} A_n = \liminf_{n \to \infty} A_n,$$

in which case
$$\lim_{n \to \infty} A_n \equiv \limsup_{n \to \infty} A_n = \liminf_{n \to \infty} A_n$$

and
$$\lim_{n \to \infty} \mathbb{P}\left[A_n\right] = \mathbb{P}\left[\lim_{n \to \infty} A_n\right].$$

This holds without any monotonicity assumption on the collection $\{A_n, \ n = 1, 2, \ldots\}$.

Conditional probabilities

With A and B events in $\mathcal{F}$ such that $\mathbb{P}[B] > 0$, define the *conditional probability of A given B* by

$$\mathbb{P}[A|B] =\equiv \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}.$$

When $\mathbb{P}[B] = 0$ it is customary to take $\mathbb{P}[A|B]$ to be arbitrary in $[0, 1]$.

However, when $\mathbb{P}[B] > 0$ define the mapping $\mathbb{Q}_B : \mathcal{F} \to [0, 1]$ by

$$\mathbb{Q}_B(A) \equiv \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}, \quad A \in \mathcal{F}.$$

It is easy to show that $\mathbb{Q}_B : \mathcal{F} \to \mathbb{R}_+$ is a probability measure on $\mathcal{F}$. Incidentally it is this fact that is often invoked to justify that $\mathbb{P}[\ \cdot\ |B]$ be selected as a probability measure on $\mathcal{F}$ when $\mathbb{P}[B] = 0$.

The relation

$$\mathbb{P}[A|B]\,\mathbb{P}[B] = \mathbb{P}[A \cap B], \quad A \in \mathcal{F}$$

is always true regardless of $\mathbb{P}[B] > 0$ or not.

Three easy consequences

With I a countable index set, let $\{B_i,\ i \in I\}$ be events in $\mathcal{F}$ that form a partition of Ω, i.e.,

$$B_i \cap B_j = \emptyset, \quad \begin{matrix} i, j \in I \\ i \neq j \end{matrix} \quad \text{and} \quad \cup_{i \in I} B_i = \Omega$$

- Law of total probabilities: Because $A = \cup_{i \in I}(A \cap B_i)$, we have

$$\begin{aligned} \mathbb{P}[A] &= \sum_{i \in I} \mathbb{P}[A \cap B_i] \\ &= \sum_{i \in I} \mathbb{P}[A|B_i]\,\mathbb{P}[B_i], \quad A \in \mathcal{F}. \end{aligned} \tag{1}$$

  Put differently,

$$\mathbb{P}[A] = \sum_{i \in I} \mathbb{Q}_{B_i}(A)\mathbb{P}[B_i], \quad A \in \mathcal{F}.$$

- Bayes' rule (From prior probabilities to posterior probabilities): Consider any event A in $\mathcal{F}$ such that $\mathbb{P}[A] > 0$. For each k in I, we have

$$\begin{aligned} \mathbb{P}[B_k|A] &= \frac{\mathbb{P}[B_k \cap A]}{\mathbb{P}[A]} \\ &= \frac{\mathbb{P}[A \cap B_k]}{\sum_{i \in I} \mathbb{P}[A \cap B_i]} \\ &= \frac{\mathbb{P}[A|B_k]\,\mathbb{P}[B_k]}{\sum_{i \in I} \mathbb{P}[A|B_i]\,\mathbb{P}[B_i]}. \end{aligned} \tag{2}$$

- Modeling sequential decision making: If I is a finite set, say $I = \{1, \ldots, n\}$, we have

$$\mathbb{P}[A_1 \cap \ldots \cap A_n] = \prod_{i=2}^{n} \mathbb{P}[A_i | A_1 \cap \ldots \cap A_{i-1}] \cdot \mathbb{P}[A_1].$$

Associated σ-fields

Let I denote an index set (not necessarily countable). If $\{\mathcal{F}_i,\ i \in I\}$ is a collection of σ-fields on Ω, then the σ-field $\wedge_{i\in I}\mathcal{F}_i$ defined by

$$\wedge_{i\in I}\mathcal{F}_i \equiv \{E \in \mathcal{P}(\Omega) : E \in \mathcal{F}_i,\ i \in I\}$$

is also a σ-field on Ω (sometimes referred to as the intersection σ-field). It is the largest σ-field on Ω that is contained in each of the σ-fields $\{\mathcal{F}_i,\ i \in I\}$.

Let $\mathcal{G}$ denote a collection of subsets of Ω, so $\mathcal{G} \subseteq \mathcal{P}(\Omega)$. The notation $\sigma(\mathcal{G})$ will be used to denote the *smallest* σ-field on Ω that contains $\mathcal{G}$: The σ-field $\sigma(\mathcal{G})$ can be interpreted as

$$\sigma(\mathcal{G}) \equiv \wedge_{a\in A}\mathcal{G}_a$$

where $\{\mathcal{G}_a,\ a \in A\}$ is the non-empty collection of all the σ-fields on Ω which contain $\mathcal{G}$, namely

$$\mathcal{G} \subseteq \mathcal{G}_\alpha, \quad a \in A.$$

This collection is not empty because $\mathcal{P}(\Omega)$ is a σ-field that contains $\mathcal{G}$.

Fact: Let $\mathcal{G}_1$ and $\mathcal{G}_2$ denote two collection of subsets of Ω such that $\mathcal{G}_1 \subseteq \mathcal{G}_2$, then it holds that

$$\sigma(\mathcal{G}_1) \subseteq \sigma(\mathcal{G}_2).$$

Terminology: Consider $\mathcal{G}$ and $\mathcal{F}$ two collections of subsets of Ω with $\mathcal{G} \subseteq \mathcal{F}$. If $\mathcal{F}$ is a σ-field on Ω and

$$\mathcal{F} = \sigma(\mathcal{G}),$$

then we say that $\mathcal{G}$ *generates* $\mathcal{F}$, or equivalently, $\mathcal{G}$ is a *generating family* (or a *generator*) for $\mathcal{F}$.

Example 1: Infinite coin tosses

The experiment $\mathcal{E}$ consists in repeating a coin toss under ”identical and independent conditions” with a fair coin (so that the likelihood of occurrence of Head is the same as that of Tail). It is convenient to take the sample space Ω to be $\{0,1\}^{\mathbb{N}_0}$, i.e.,

$$\Omega = \{\omega = (\omega_1, \omega_2, \ldots) : \ \omega_k \in \{0,1\},\ k = 1, 2, \ldots\}$$

with the understanding that $\omega_k = 1$ (resp. $\omega_k = 0$) if the k^{th} toss yields Head (resp. Tail).

Note that Ω has the same cardinality as the unit interval $[0,1]$ (hence is uncountable). How should we construct $\mathcal{F}$ (and $\mathbb{P}$)?

- It is natural to require that for *any* $n = 1, 2, \ldots$, any collection of outcomes determined by the first n tosses should be an event in $\mathcal{F}$ – After all one should expect that the model we are seeking to construct would also contain a model for *each* of the finite toss experiments. In particular, with any *given* binary sequence $(b_1, \ldots, b_n)$ of length n, consider

$$F_n(b_1, \ldots, b_n) \equiv \left\{ \omega = (\omega_1, \omega_2, \ldots) \in \Omega : \begin{array}{c} \omega_k = b_k \\ k = 1, \ldots, n \end{array} \right\}. \tag{3}$$

It is plain that $\mathcal{F}$ must at least contain these events, i.e.,

$$F_n(b_1, \ldots, b_n) \in \mathcal{F} \tag{4}$$

Fairness (which is essentially a uniformity condition) requires that

$$\mathbb{P}\left[F_n(b_1, \ldots, b_n)\right] = 2^{-n} \tag{5}$$

since $\mathbb{P}\left[F_n(b_1, \ldots, b_n)\right]$ should not depend on $(b_1, \ldots, b_n)$ and there are 2^n distinct sets of the form (3). Note also that

$$\cup_{(b_1, \ldots, b_n) \in \{0,1\}^n} F_n(b_1, \ldots, b_n) = \Omega.$$

- It is therefore natural to take

$$\mathcal{F} = \sigma \left(F_n(b_1, \ldots, b_n) : \begin{array}{c} b_1, \ldots, b_n \in \{0, 1\} \\ n = 1, 2, \ldots \end{array} \right) = \sigma\left(\mathcal{G}\right)$$

where the generator $\mathcal{G}$ is the collection

$$\mathcal{G} \equiv \left\{ F_n(b_1, \ldots, b_n) : \begin{array}{c} b_1, \ldots, b_n \in \{0, 1\} \\ n = 1, 2, \ldots \end{array} \right\}.$$

- The σ-field $\mathcal{F}$ so defined is very large/rich but does *not* coincide with $\mathcal{P}(\Omega)$. It does however contain some events that do not depend on a given finite number of tosses, e.g.,

$$\begin{aligned} F &= \left\{ \omega = (\omega_1, \omega_2, \ldots) \in \Omega : \begin{array}{c} \text{A even number of tosses needed} \\ \text{before observing Head} \end{array} \right\} \\ &= \cup_{k=1}^{\infty} E_{2k} \end{aligned}$$

where for each $k = 1, \ldots$ we have defined

$$E_k \equiv \left\{ \omega = (\omega_1, \omega_2, \ldots) \in \Omega : \begin{array}{c} \omega_1 = \ldots = \omega_{k-1} = 0 \\ \text{and} \\ \omega_k = 1 \end{array} \right\}.$$

Note that $E_k = F_k(0, \ldots, 0, 1)$.

- Measure Theory tells us that that there exists a *unique* probability measure $\mathbb{P}$ on $\mathcal{F}$ so that (5) holds for all $n = 1, 2, \ldots$.

Example 2: Selecting a point at random in the interval $[0, 1]$
A particularly important case is that of equipping the *non*-countable interval $[0, 1]$ into a measurable space on which probabilities can be defined through a probability measure. This corresponds to the random experiment where you select at random a point in the finite interval $[0, 1]$, so here $\Omega = [0, 1]$. Intuitively we could proceed as follows to define $\mathcal{F}$ and $\mathbb{P}$ (denoted here λ for Lebesgue measure).

A well-known fact of topology on $\mathbb{R}$: Any open subset U in $\mathbb{R}$ can be expressed as the union of a *countable* collection of *non-overlapping* open intervals, i.e., there exists a countable collection $\{J_i, i \in I\}$ of *open intervals* of $\mathbb{R}$ such that

$$U = \cup_{i \in I} J_i \quad \text{with} \quad J_k \cap J_\ell = \emptyset, \quad \begin{array}{c} k \neq \ell \\ k, \ell \in I \end{array} \tag{6}$$

To define the appropriate σ-field $\mathcal{F}$ and the probability measure λ on it, it is natural to proceed as follows:

(i) For any interval $I_{\alpha,\beta} = [\alpha, \beta] \subseteq \Omega$, set

$$\lambda([\alpha, \beta]) = \beta - \alpha.$$

(ii) Thus, with $\alpha = \beta$, $\lambda(\{\alpha\}) = 0$ for all α in Ω.

(iii) It follows that

$$\lambda((\alpha, \beta)) = \beta - \alpha$$

by continuity of probability measures applied to the sequence of open intervals $\{(\alpha - \frac{1}{n}, \beta + \frac{1}{n}),\ n = 1, 2, \ldots\}$.

(iv) Union of countable collections of open intervals should be in $\mathcal{F}$. Therefore, by the well-known fact (6) we see that every open set $U \subseteq (0,1)$ is in $\mathcal{F}$ with

$$\lambda(U) = \sum_{i \in J} \lambda(J_i)$$

where the notation and the assumptions are the ones used in (6).

(v) Therefore, since a set F of $[0,1]$ is closed if and only F^c is open, we conclude that every closed set $F \subseteq (0,1)$ is also in $\mathcal{F}$ with $\lambda(F) = 1 - \lambda(F^c)$.

(vi) Any countable union of open subsets should be in $\mathcal{F}$

(vii) Any countable intersection of closed subsets should be in $\mathcal{F}$

(viii) .....

This leads to defining $\mathcal{F}$ as

$$\mathcal{F} = \sigma\left(\mathcal{I}([0,1])\right)$$

where $\mathcal{I}([0,1])$ denotes the collection of all *open intervals* contained in $[0,1]$. The σ-field $\sigma\left(\mathcal{I}([0,1])\right)$ is called the *Borel* σ-field on $[0,1]$ and is denoted by $\mathcal{B}([0,1])$. However, by the well-known fact (6) it follows readily that we have also the characterization

$$\mathcal{B}([0,1]) = \sigma\left(\mathcal{O}([0,1])\right)$$

where $\mathcal{O}([0,1])$ denotes the collection of all *open sets* contained in $[0,1]$.

Borel σ-fields

More generally, with I denoting an *interval* (closed or open or neither, finite or not), we define

$$\mathcal{B}(I) \equiv \sigma\left(\mathcal{O}(I)\right)$$

where $\mathcal{O}(I)$ denotes the collection of all *open sets* contained in I. The σ-field $\sigma\left(\mathcal{O}(I)\right)$ is called the *Borel* σ-field on I and is denoted by $\mathcal{B}(I)$.

This notion can be further extended: With A denoting a subset of $\mathbb{R}^p$ for some positive integer p, we write

$$\mathcal{B}(A) \equiv \sigma\left(\mathcal{O}(A)\right)$$

where $\mathcal{O}(A)$ denotes the collection of all *open sets* contained in A. In particular,

$$\mathcal{B}(\mathbb{R}^p) \equiv \sigma\left(\mathcal{O}(\mathbb{R}^p)\right)$$

where $\mathcal{O}(\mathbb{R}^p)$ denotes the collection of all *open sets* contained in $\mathbb{R}^p$.

Note that the general definition of a Borel σ-field uses the collection of open sets as a generator for in higher-dimensions there are no intervals!

Definitions and some simple facts

Consider mappings $g : \Omega_a \to \Omega_b$ and $h : \Omega_b \to \Omega_c$ where Ω_a, Ω_b and Ω_c are arbitrary sets (possibly identical).

Let $\mathcal{B}$ be a collection of subsets of Ω_b (so $\mathcal{B} \subseteq \mathcal{P}(\Omega_b)$). With

$$g^{-1}(\mathcal{B}) \equiv \left\{ g^{-1}(F_b) : \ F_b \in \mathcal{B} \right\},$$

it is always the case that

$$g^{-1}\left(\sigma(\mathcal{B})\right) = \sigma\left(g^{-1}(\mathcal{B})\right). \tag{7}$$

Proof. The collection $g^{-1}\left(\sigma(\mathcal{B})\right)$ is a σ-field on Ω_a, and it contains $g^{-1}(\mathcal{B})$, hence the inclusion

$$\sigma\left(g^{-1}(\mathcal{B})\right) \subseteq g^{-1}\left(\sigma(\mathcal{B})\right).$$

To establish the reverse inclusion, consider the collection $\mathcal{B}_g$ of subsets of Ω_b defined by

$$\mathcal{B}_g \equiv \left\{ F_b \subseteq \Omega_b : \ g^{-1}(F_b) \in \sigma\left(g^{-1}(\mathcal{B})\right) \right\}.$$

It is plain that $\mathcal{B}_g$ is a σ-field on Ω_b; as it obviously contains $\mathcal{B}$, it must also contain $\sigma(\mathcal{B})$ and the inclusion

$$g^{-1}\left(\sigma(\mathcal{B})\right) \subseteq \sigma\left(g^{-1}(\mathcal{B})\right).$$

follows.

Define the mapping $h \circ g : \Omega_a \to \Omega_c$ obtained by *composing* g with h through

$$(h \circ g)(\omega_a) = h(g(\omega_a)), \quad \omega \in \Omega_a$$

If $\mathcal{C}$ be a collection of subsets of Ω_c (so $\mathcal{C} \subseteq \mathcal{P}(\Omega_c)$), then

$$(h \circ g)^{-1}(\mathcal{C}) = g^{-1}\left(h^{-1}(\mathcal{C})\right)$$

Borel mappings

Consider an arbitrary set S equipped with a σ-field $\mathcal{S}$. A mapping $g : S \to \mathbb{R}^p$ is said to be a *Borel* mapping if the conditions

$$g^{-1}(B) \in \mathcal{S}, \quad B \in \mathcal{B}(\mathbb{R}^p) \tag{8}$$

are all satisfied where

$$g^{-1}(B) \equiv \{s \in S : \ g(s) \in B\} .$$

Fact: If $g : S \to \mathbb{R}^p$ and $h : \mathbb{R}^p \to \mathbb{R}^q$ are Borel mappings, then the composition mapping $h \circ g : S \to \mathbb{R}^q$ is also a Borel mapping.

Proof. This is a simple consequence of the fact that

$$(h \circ g)^{-1}(B) = g^{-1}\left(h^{-1}(B)\right), \quad B \in \mathcal{B}(\mathbb{R}^q).$$

Thus, $h^{-1}(B)$ is an element of $\mathcal{B}(\mathbb{R}^q)$ by the Borel measurability of h, whence $(h \circ g)^{-1}(B)$ s an element of $\mathcal{B}(\mathbb{R}^p)$ by the Borel measurability of g.

An important fact: Let $\mathcal{G}$ denote a collection of subsets of $\mathbb{R}^p$ which generates the Borel σ-field $\mathcal{B}(\mathbb{R}^p)$, i.e.,

$$\mathcal{B}(\mathbb{R}^p) = \sigma(\mathcal{G}) . \tag{9}$$

It holds that the mapping $g : S \to \mathbb{R}^p$ is a Borel mapping if and only if the *weaker* set of conditions

$$g^{-1}(E) \in \mathcal{S}, \quad E \in \mathcal{G} \tag{10}$$

holds.

Proof. One implication is trivial since the conditions (10) constitute a subset of the conditions (8). To prove the reverse implication consider the collection $\mathcal{E}_g$ given by

$$\mathcal{E}_g \equiv \left\{E \subseteq \mathbb{R}^p : \ g^{-1}(E) \in \mathcal{S}\right\} .$$

The collection $\mathcal{E}_g$ is a σ-field on $\mathbb{R}^p$ because $\mathcal{S}$ is a σ-field on S. *Under* condition (10) we note the inclusion $\mathcal{G} \subseteq \mathcal{E}_g$, hence $\sigma(\mathcal{G}) \subseteq \mathcal{E}_g$ and the conditions (8) all hold since $\sigma(\mathcal{G}) = \mathcal{B}(\mathbb{R}^p)$ by assumption.

There are many generators known for the Borel σ-field $\mathcal{B}(\mathbb{R}^p)$. For instance, we have (9) with

- $\mathcal{G} = \mathcal{R}_{\rm open}(\mathbb{R}^p)$ where $\mathcal{R}_{\rm open}(\mathbb{R}^p)$ is the collection of all *finite open rectangles*, i.e.,

$$\mathcal{R}_{\rm open}(\mathbb{R}^p) \equiv \left\{ I_1 \times \ldots \times I_p, \quad \begin{array}{c} I_k \in \mathcal{I}(\mathbb{R}) \\ k = 1, \ldots, p \end{array} \right\}$$

where

$$\mathcal{I}(\mathbb{R}) = \{(a,b) : \ a, b \in \mathbb{R}\}$$

Use the following fact: For any open set U in $\mathbb{R}^p$ there exists a countable family of *open* rectangles $\{R_i, \ i \in I\}$ in $\mathcal{R}_{\rm open}(\mathbb{R}^p)$ with countable I such that $U = \cup_{i \in I} R_i$. It is the analog of a similar fact encountered in one dimension.

- $\mathcal{G} = \mathcal{R}_{\rm SW}(\mathbb{R}^p)$ where $\mathcal{R}_{\rm SW}(\mathbb{R}^p)$ is the collection of all *closed Southwest rectangles*, i.e.,

$$\mathcal{R}_{\rm SW}(\mathbb{R}^p) \equiv \left\{ I_1 \times \ldots \times I_p, \quad \begin{array}{c} I_k = (-\infty, a_k] \\ a_k \in \mathbb{R} \\ k = 1, \ldots, p \end{array} \right\}.$$

It follows from the discussion above that a mapping $g : S \to \mathbb{R}^p$ is a Borel mapping if the seemingly weaker conditions

$$\left\{ s \in S : \ g(s) \in \prod_{i=1}^p (-\infty, a_k] \right\} \in \mathcal{S}, \quad (a_1, \ldots, a_p) \in \mathbb{R}^p$$

all hold. Equivalently, a mapping $g : S \to \mathbb{R}^p$ is a Borel mapping if

$$\{s \in S : \ g_k(s) \leq a_k, \ k = 1, \ldots, p\} \in \mathcal{S}, \quad (a_1, \ldots, a_p) \in \mathbb{R}^p$$

where it is understood that

$$g(s) = (g_1(s), \ldots, g_p(s)), \quad s \in S.$$

It is now plain that for each $k = 1, \ldots, p$, the component mapping $g_k : S \to \mathbb{R}$ is also a Borel mapping – Just take $a_\ell = \infty$ for all $\ell = 1, \ldots, k$ different from k. Conversely, since

$$\{s \in S : \ g_k(s) \leq a_k, \ k = 1, \ldots, p\} = \cap_{k=1}^p \{s \in S : \ g_k(s) \leq a_k\}$$

for arbitrary $(a_1, \ldots, a_p)$ in $\mathbb{R}^p$, we see that the mapping $g : S \rightarrow \mathbb{R}^p$ is a Borel mapping if and only if each of the component mappings $g_1 : S \rightarrow \mathbb{R}, \ldots, g_p : S \rightarrow \mathbb{R}$ is a Borel mapping.

Most (if not all) mappings $\mathbb{R}^p \rightarrow \mathbb{R}^q$ encountered in applications are Borel mappings. Furthermore, any *continuous* mapping $\mathbb{R}^p \rightarrow \mathbb{R}^q$ can be shown to be a Borel mapping!

Random variables

Given a probability triple $(\Omega, \mathcal{F}, \mathbb{P})$, a mapping $X : \Omega \to \mathbb{R}^p$ is a *random variable* (rv) if

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}, \quad B \in \mathcal{B}(\mathbb{R}^p).$$

In other words, the mapping $X : \Omega \to \mathbb{R}^p$ is a rv if it is a Borel mapping $X : \Omega \to \mathbb{R}^p$ – Here $S = \Omega$ and $\mathcal{S} = \mathcal{F}$. We shall often write $[X \in B]$ in lieu of $X^{-1}(B)$ and $\mathbb{P}[X \in B]$ for $\mathbb{P}[[X \in B]]$.

In view of the earlier discussion the mapping $X : \Omega \to \mathbb{R}^p$ is a rv if and only

$$\{\omega \in \Omega : \; X_k(\omega) \le a_k, \; k = 1, \ldots, p\} \in \mathcal{F}, \quad (a_1, \ldots, a_p) \in \mathbb{R}^p$$

where it is understood that

$$X(\omega) = (X_1(\omega), \ldots, X_p(\omega)), \quad \omega \in \Omega.$$

This last condition can also be rewritten as

$$\cap_{k=1}^p [X_k \le a_k] \in \mathcal{F}, \quad (a_1, \ldots, a_p) \in \mathbb{R}^p$$

It is now plain that for each $k = 1, \ldots, p$, the component mapping $X_k : \Omega \to \mathbb{R}$ is also a rv – Just take $a_\ell = \infty$ for all $\ell = 1, \ldots, k$ different from k. Here as well, we conclude that the mapping $X : \Omega \to \mathbb{R}^p$ is a rv if and only if each of the component mappings $X_1 : \Omega \to \mathbb{R}, \ldots, X_p : \Omega \to \mathbb{R}$ is a rv.

Probability distribution functions

The *probability distribution (function)* $F_X : \mathbb{R}^p \to [0, 1]$ of the rv X is defined by

$$\begin{aligned} F_X(x) &\equiv \mathbb{P}[X \in (-\infty, x_1] \times \ldots \times (-\infty, x_p]] \\ &= \mathbb{P}[X_1 \le x_1, \ldots, X_p \le x_p], \quad x = (x_1, \ldots, x_p) \in \mathbb{R}^p. \end{aligned} \tag{11}$$

with the notation $X = (X_1, \ldots, X_p)$.

It turns out that there is as much probabilistic information in the probability distribution $F_X : \mathbb{R}^p \to [0, 1]$ of the rv X as in

$$\{\mathbb{P}[X \in B], \; B \in \mathcal{B}(\mathbb{R}^p)\}$$

In fact, knowledge of $F_X : \mathbb{R}^p \to \mathbb{R}$ allows a *unique* reconstruction of

$$\{\mathbb{P}[X \in B], \; B \in \mathcal{B}(\mathbb{R}^p)\}.$$

Properties of F_X (Case $p = 1$): It is easy to see that the following properties hold:

- Monotonicity:
$$F_X(x) \leq F_X(y), \quad x, y \in \mathbb{R}$$
- Right-continuous:
$$\lim_{y \downarrow x} F_X(y) = F_X(x), \quad x \in \mathbb{R}$$
- Left limit exists:
$$\lim_{y \uparrow x} F_X(y) = F_X(x-) \quad \text{with} \quad \mathbb{P}\left[X = x\right] = F_X(y) - F_X(x-), \quad x \in \mathbb{R}$$
- Behavior at infinity: Monotonically
$$\lim_{x \to -\infty} F_X(x) = 0 \quad \text{and} \quad \lim_{x \to \infty} F_X(x) = 1$$

A *probability distribution (function)* on $\mathbb{R}$ is any mapping $F : \mathbb{R} \to [0, 1]$ such that

- Monotonicity:
$$F(x) \leq F(y), \quad x, y \in \mathbb{R}$$
- Right-continuous:
$$\lim_{y \downarrow x} F(y) = F(x), \quad x \in \mathbb{R}$$
- Left limit exists:
$$\lim_{y \uparrow x} F(y) = F(x-) \quad x \in \mathbb{R}$$
- Behavior at infinity: Monotonically
$$\lim_{x \to -\infty} F(x) = 0 \quad \text{and} \quad \lim_{x \to \infty} F(x) = 1$$

Important fact: Any rv $X : \Omega \to \mathbb{R}$ generates a probability distribution function $F_X : \mathbb{R} \to [0, 1]$. Conversely, for any probability distribution function $F : \mathbb{R} \to [0, 1]$, there exists a probability triple $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}})$ and a rv $\tilde{X} : \widetilde{\Omega} \to \mathbb{R}$ defined on it such that
$$\widetilde{\mathbb{P}}\left[\widetilde{X} \leq x\right] = F(x), \quad x \in \mathbb{R}$$

This is the basis of Monte-Carlo simulation. There exists a multi-dimensional analog to this fact.

Proof. Take $\widetilde{\Omega} = [0,1]$, $\widetilde{\mathcal{F}} = \mathcal{B}([0,1])$ and $\widetilde{\mathbb{P}} = \lambda$. Define the rv $\tilde{X} : \widetilde{\Omega} \to \mathbb{R}$ by setting

$$\tilde{X}(\widetilde{\omega}) = F^{-}(\widetilde{\omega}), \quad \widetilde{\omega} \in [0,1]$$

where $F^{-} : [0,1] \to [-\infty, \infty]$ is the *generalized inverse* of F given by

$$F^{-}(u) = \inf\left(x \in \mathbb{R} : \ u \leq F(x)\right), \quad 0 \leq u \leq 1.$$

with the understanding that $F^{-}(u) = \infty$ if the defining set is empty, i.e., $F(x) < u$ for all x in $\mathbb{R}$.

Discrete distributions

A rv $X : \Omega \to \mathbb{R}^p$ is a *discrete* rv if there exists a *countable* subset $S \subseteq \mathbb{R}^p$ such that

$$\mathbb{P}\left[X \in S\right] = 1.$$

Note that

$$\mathbb{P}\left[X \in B\right] = \sum\nolimits_{x \in S \cap B} \mathbb{P}\left[X = x\right], \quad B \in \mathcal{B}(\mathbb{R}^p).$$

It is often more convenient to characterize the distributional properties of the rv X through its probability mass function (pmf) of the rv X given by

$$\boldsymbol{p}_X \equiv (p_X(x), \ x \in S)$$

with

$$p_X(x) = \mathbb{P}\left[X = x\right], \quad x \in S.$$

Well-known examples of discrete rvs (and of their distributions) include:

(i) Bernoulli $\mathrm{Ber(p)}$ (with $0 \leq p \leq 1$)

(ii) Binomial $\mathrm{Bin(n;p)}$ (with $n = 1, 2, \ldots$ and $0 \leq p \leq 1$)

(iii) Poisson $\mathrm{Poi}(\lambda)$ (with $\lambda > 0$)

(iv) Geometric $\mathrm{Geo(p)}$ (with $0 \leq p \leq 1$)

Absolutely continuous distributions

A rv $X : \Omega \to \mathbb{R}^p$ is an (absolutely) continuous rv if there exists a Borel mapping $f_X : \mathbb{R}^p \to \mathbb{R}_+$ such that

$$\mathbb{P}\left[X_i \leq x_i,\ i = 1, \ldots, p\right] = \int_{-\infty}^{x} f_X(\xi) d\xi, \quad x = (x_1, \ldots, x_p) \in \mathbb{R}^p.$$

Well-known examples of continuous rvs (and of their distributions) include:

(i) Uniform $\mathrm{U}(a, b)$ (with $a < b$ in $\mathbb{R}$)

(ii) Exponential $\mathrm{Exp}(\lambda)$ (with $\lambda > 0$)

(iii) Gaussian $\mathrm{N}(m, \sigma^2)$ (with m, σ in $\mathbb{R}$)

(iv) Cauchy $\mathrm{C}(m, a)$ (with m, a in $\mathbb{R}$)

Properties of F_X when $p \geq 1$

- Monotonicity needs to be modified and now reads

$$\mathbb{P}\left[x_k < X_k \leq y_k\right] \geq 0, \qquad \begin{array}{c} x_k < y_k \\ x_k, y_k \in \mathbb{R} \\ k = 1, \ldots, p \end{array}$$

  with the understanding that the quantity $\mathbb{P}\left[x_k < X_k \leq y_k\right]$ is expressed solely in terms of $F_X : \mathbb{R}^p \to [0, 1]$.

- Right-continuous:

$$\lim_{y \downarrow x} F_X(y) = F_X(x), \quad x \in \mathbb{R}^p$$

  with the understanding that $y_k \downarrow x_k$ for each $k = 1, \ldots, p$.

- Left limit exists:

$$\lim_{y \uparrow x} F_X(y) = F_X(x-) \quad \text{with} \quad \mathbb{P}\left[X = x\right] = F_X(y) - F_X(x-), \quad x \in \mathbb{R}^p$$

  with the understanding that $y_k \uparrow x_k$ for each $k = 1, \ldots, p$.

- Behavior at infinity:

$$\lim_{\min(x_k,\ k=1,\ldots,p)\to-\infty} F_X(x) = 0$$

and

$$\lim_{\min(x_k,\ k=1,\ldots,p)\to\infty} F_X(x) = 1$$

Independence of rvs

Consider a collection of rvs $\{X_i,\ i \in I\}$ which are all defined on some probability triple $(\Omega, \mathcal{F}, \mathbb{P})$. Assume that for each i in I, the rv X_i is a $\mathbb{R}^{p_i}$-valued rv for some positive integer p_i.

With I finite, we shall say that the rvs $\{X_i,\ i \in I\}$ are *mutually independent* if for each selection of B_i in $\mathcal{B}(R^{p_i})$ for each i in I, the events

$$\{[X_i \in B_i],\ i \in I\}$$

are mutually independent. It is easy to see that this is equivalent to requiring

$$\mathbb{P}\left[\cap_{i\in I}[X_i \in B_i]\right] = \prod_{i\in I} \mathbb{P}\left[X_i \in B_i\right], \qquad \begin{array}{c} B_i \in \mathcal{B}(R^{p_i}) \\ i \in I. \end{array}$$

More generally, with I arbitrary (and possibly uncountable), the rvs $\{X_i,\ i \in I\}$ are said to be mutually independent if for every finite subset $J \subseteq I$, the rvs $\{X_j,\ j \in J\}$ are mutually independent!

Product spaces

Some facts: Consider two arbitrary sets Ω_a and Ω_b (possibly identical). Let $\mathcal{A}$ and $\mathcal{B}$ denote non-empty collections of subsets of Ω_a and Ω_b, respectively. While the collection $\mathcal{A} \times \mathcal{B}$ is usually *not* a σ-field on $\Omega_a \times \Omega_b$, even when $\mathcal{A}$ and $\mathcal{B}$ are themselves σ-fields, it can be shown that

$$\sigma\left(\mathcal{A} \times \mathcal{B}\right) = \sigma\left(\sigma\left(\mathcal{A}\right) \times \sigma\left(\mathcal{B}\right)\right).$$

Consider the probability triples $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1), \ldots, (\Omega_p, \mathcal{F}_p, \mathbb{P}_p)$. Their *Cartesian* product is the set Ω defined by

$$\Omega \equiv \Omega_1 \times \ldots \times \Omega_n.$$

We introduce the collection $\mathcal{F}_1 \times \ldots \times \mathcal{F}_p$ of subsets of Ω given by

$$\mathcal{F}_1 \times \ldots \times \mathcal{F}_p = \left\{ F_1 \times \ldots \times F_p, \quad \begin{array}{c} F_k \in \mathcal{F}_k, \\ k = 1, \ldots, p \end{array} \right\}.$$

We write

$$\mathcal{F}_1 \otimes \ldots \otimes \mathcal{F}_p = \otimes_{k=1}^p \mathcal{F}_k = \sigma \left(\mathcal{F}_1 \times \ldots \times \mathcal{F}_p \right).$$

Note that

$$\sigma \left(\mathcal{F}_1 \times \ldots \times \mathcal{F}_p \right) = \sigma \left(\sigma \left(\mathcal{F}_1 \right) \times \ldots \times \sigma \left(\mathcal{F}_p \right) \right).$$

The *product* probability measure $\mathbb{P}$ is defined on $\otimes_{k=1}^p \mathcal{F}_k$ as follows: For any rectangle

$$R = F_1 \times \ldots \times F_p, \qquad \begin{array}{c} F_k \in \mathcal{F}_k \\ k = 1, \ldots, p \end{array}$$

set

$$\mathbb{P}\left[R\right] = \prod_{k=1}^p \mathbb{P}_k[F_k]. \tag{12}$$

So far, $\mathbb{P}$ is defined only on $\mathcal{F}_1 \times \ldots \times \mathcal{F}_p$. However, Measure Theory guarantees that there exists a *unique* probability measure on the σ-field

$$\sigma \left(\sigma \left(\mathcal{F}_1 \right) \times \ldots \times \sigma \left(\mathcal{F}_p \right) \right)$$

such that (12) holds.

An important modeling fact: Under $\mathbb{P}$, the events

$$\begin{aligned} E_1 &= A_1 \times \Omega_2 \times \ldots \times \Omega_p \\ E_2 &= \Omega_1 \times A_2 \times \ldots \times \Omega_p \\ \vdots &= \vdots \\ E_p &= \Omega_1 \times \Omega_2 \times \ldots \times A_p \end{aligned}$$

are *mutually independent* with

$$\mathbb{P}\left[E_k\right] = \mathbb{P}_k[A_k], \quad k = 1, \ldots, p.$$

Taking limits ______________________________

Consider the sequence of $\mathbb{R}$-valued rvs $\{X_n, \ n = 1, 2, \ldots\}$ which are *all* defined on the *same* probability triple $(\Omega, \mathcal{F}, \mathbb{P})$. The following mappings $\Omega \to [-\infty, \infty]$ are rvs in the extended sense:

The supremum mapping $\Omega \to [-\infty, \infty]$ defined by

$$\omega \to \sup_{n \geq 1} X_m(\omega), \quad \omega \in \Omega$$

The infimum mapping $\Omega \to [-\infty, \infty]$ defined by

$$\omega \to \inf_{n \geq 1} X_m(\omega), \quad \omega \in \Omega.$$

The limsup mapping $\Omega \to [-\infty, \infty]$ defined by

$$\omega \to \limsup_{n \to \infty} X_n(\omega), \quad \omega \in \Omega.$$

The liminf mapping $\Omega \to [-\infty, \infty]$ defined by

$$\omega \to \liminf_{n \to \infty} X_n(\omega), \quad \omega \in \Omega.$$

It follows that

$$\Omega^\star \equiv \left[\liminf_{n \to \infty} X_n = \limsup_{n \to \infty} X_n \right] \in \mathcal{F}$$

and on $\Omega^\star$, it holds that $\lim_{n \to \infty} X_n$ exists (possibly as an element in $[-\infty, \infty]$), and is the common value assumed by $\liminf_{n \to \infty} X_n$ and $\limsup_{n \to \infty} X_n$.

When $\mathbb{P}[\Omega^\star] = 1$ it is customary to say that the sequence $\{X_n, \ n = 1, 2, \ldots\}$ *converges almost surely* (a.s.) (under $\mathbb{P}$), and we write

$$\lim_{n \to \infty} X_n \quad \mathbb{P}\text{-a.s.}$$

In that case, for any rv $X : \Omega \to \mathbb{R}$ such that

$$X(\omega) = \lim_{n \to \infty} X_n(\omega), \quad \omega \in \Omega^\star$$

we shall write

$$\lim_{n \to \infty} X_n = X \quad \mathbb{P}\text{-a.s.}$$

Such a rv X always exists when $\mathbb{P}[\Omega^\star] = 1$ but is not unique. Existence is immediate since we can always take

$$X(\omega) \equiv \begin{cases} \liminf_{n \to \infty} X_n(\omega) = \limsup_{n \to \infty} X_n(\omega) & \text{if } \omega \in \Omega^\star \\ \\ Z(\omega) & \text{if } \omega \notin \Omega^\star \end{cases}$$

where $Z : \Omega \to \mathbb{R}$ is some arbitrary rv, and non-uniqueness is obvious.

Simple rvs

A rv $X : \Omega \to \mathbb{R}$ is a *simple* variable if

$$X = \sum_{k \in I} a_k \mathbf{1}\left[A_k\right]$$

where (i) I is a *finite* index set, (ii) $\{a_k, \ k \in I\}$ are scalars (not necessarily distinct) and (iii) the subsets $\{A_k, \ k \in I\}$ form an $\mathcal{F}$*-partition* of Ω, i..e., the subsets $\{A_k, \ k \in I\}$ are all in $\mathcal{F}$ with

$$\cup_{k \in I} A_k = \Omega \quad \text{ and } \quad A_k \cap A_\ell = \emptyset \quad \begin{array}{c} k \neq \ell \\ \ell, k \in I. \end{array}$$

This representation is not necessarily unique. In many arguments it is customary to assume that the values $\{a_k, \ k \in I\}$ are *distinct* scalars and that the events $\{A_k, \ k \in I\}$ forming the $\mathcal{F}$-partition are all non-empty, in which case $\{X(\omega), \ \omega \in \Omega\} = \{a_k, \ k \in I\}$ and

$$A_k = \left[X = a_k\right], \quad k \in I.$$

We refer to this representation as the *generic* representation of the simple rv. There is no loss of generality in using the generic representation as will shortly become apparent.

Fact: For any rv $X : \Omega \to \mathbb{R}_+$, there exists a monotonically increasing sequence of simple rvs $\{X_n, \ n = 1, 2, \ldots\}$ such that

$$X_n \leq X_{n+1} \leq X, \quad n = 1, 2, \ldots$$

and

$$\lim_{n \to \infty} X_n = X.$$

For instance, for each $n = 1, 2, \ldots$, define the simple rv $X_n : \Omega \to \mathbb{R}_+$ by

$$X_n = \sum_{m=0}^{n-1} \sum_{k=0}^{2^n - 1} \left(m + k 2^{-n}\right) \mathbf{1}\left[m + k 2^{-n} < X \leq m + (k+1) 2^{-n}\right] \tag{13}$$

Expectation of rvs

Consider a probability triple $(\Omega, \mathcal{F}, \mathbb{P})$, and let $X : \Omega \to \mathbb{R}$ denote an $\mathbb{R}$-valued rv defined on this probability triple. The operation of *expectation* associates with every $\mathbb{R}$-valued rv $X : \Omega \to \mathbb{R}$ a value in $[-\infty, \infty]$, denoted $\mathbb{E}\left[X\right]$; this value can be interpreted as an *average value* for X as *weighted* by its probability distribution F_X. The definition given shortly is guided by the following considerations:

(i) When defined, the quantity $\mathbb{E}[X]$ is *uniquely* determined by the probability distribution $F_X : \mathbb{R} \to [0,1]$ of the rv X.

(ii) The expectation of the indicator function of an event $[X \in B]$ should coincide with its probability, namely

$$\mathbb{E}[\mathbf{1}[X \in B]] = \mathbb{P}[X \in B], \quad B \in \mathcal{B}(\mathbb{R}).$$

(iii) The expectation of an $\mathbb{R}_+$-valued rv is always well defined (although it could be infinite) while the value of a bounded rv is also well defined

(iv) When defined, the quantity $\mathbb{E}[X]$ is defined *independently* of the type of distribution F_X, say discrete or absolute continuous, but its definition will coincide with the usual elementary definitions given in elementary courses in Probability Theory.

(v) The operation is expected to be *linear* under very broad conditions, namely for $\mathbb{R}$-valued rvs X and Y, we have

$$\mathbb{E}[\alpha X + \beta Y] = \alpha \mathbb{E}[X] + \beta \mathbb{E}[Y], \quad \alpha, \beta \in \mathbb{R}$$

whenever the involved quantities on the right handsome are well defined.

(vi) The operation is *monotone* in that $\mathbb{E}[X] \geq 0$ if $X \geq 0$. More generally, for two rvs X and Y such that $X \leq Y$, it is desired that $\mathbb{E}[X] \leq \mathbb{E}[Y]$ whenever these expectations are well defined

The expectation operation is basically defined as a Lebesgue-Stieltjes integral (either under $\mathbb{P}$ or under $\mathbb{P}_X$). We shall write alternatively,

$$\mathbb{E}[X] = \int_\Omega X(\omega) d\mathbb{P}(\omega)$$

and

$$\mathbb{E}[X] = \int_\mathbb{R} x dF_X(x).$$

Important special cases

(i) Moments: With $r = 1, 2, \ldots$, we define the r^{th} moment of X by

$$m_r \equiv \mathbb{E}\left[X^r\right]$$

while for any $r \geq 0$ the *absolute* moment of X is well defined and give by

$$\mu_r \equiv \mathbb{E}\left[|X|^r\right].$$

(ii) Transforms: For $\mathbb{R}^p$-valued rv X,

– The characteristic function of X:

$$\Phi_X(\theta) \equiv \mathbb{E}\left[e^{i\theta' X}\right], \quad \theta \in \mathbb{R}^p$$

– The moment generating function of X:

$$M_X(t'X) \equiv \mathbb{E}\left[e^{t'X}\right], \quad t \in \mathbb{R}^p$$

– The probability generating function of X of a $\mathbb{N}$-valued rv X,

$$G_X(z) = \mathbb{E}\left[z^X\right], \quad z \in$$

Defining expectations

It is a three step process:

- Step 1: For indicator rvs and for simple rvs
- Step 2: For non-negative rvs through an approximation argument in terms of simple rvs
- Step 3: For arbitrary rvs by noting the decomposition

$$X = X^+ - X^-.$$

Simple rvs

A. *Indicator* rvs: With $X = \mathbf{1}[A]$ for some A in $\mathcal{F}$, set

$$\mathbb{E}[X] = \mathbb{E}[\mathbf{1}[A]] \equiv \mathbb{P}[A].$$

B. *Simple* rvs: With simple rv X given by

$$X = \sum_{k \in I} a_k \mathbf{1}\left[A_k\right],$$

we have

$$\mathbb{E}\left[X\right] \equiv \sum_{k \in I} a_k \mathbb{P}\left[A_k\right].$$

This definition is *independent* of the representation used: If the simple rv $X : \Omega \to \mathbb{R}$ admits the two representations

$$X = \sum_{k \in I} a_k \mathbf{1}\left[A_k\right] \quad \text{and} \quad X = \sum_{\ell \in J} b_\ell \mathbf{1}\left[B_\ell\right],$$

then

$$\sum_{k \in I} a_k \mathbb{P}\left[A_k\right] = \sum_{\ell \in J} b_\ell \mathbb{P}\left[B_\ell\right]$$

and $\mathbb{E}\left[X\right]$ is this common value.

Proof: There is no loss of generality in a assuming that there are no duplications in the set of values $\{a_k,\ k \in I\}$. Since $\{a_k,\ k \in I\} = \{b_\ell,\ \ell \in J\}$, for each k in I we must have

$$\begin{aligned} \sum_{\ell \in J:\ b_\ell = a_k} b_\ell \mathbb{P}\left[B_\ell\right] &= a_k \cdot \left(\sum_{\ell \in J:\ b_\ell = a_k} \mathbb{P}\left[B_\ell\right] \right) \\ &= a_k \mathbb{P}\left[A_k\right] \end{aligned} \tag{14}$$

as we use the fact that

$$\cup_{\ell \in J:\ b_\ell = a_k} B_\ell = A_k.$$

C. *Linearity:* If X and Y are simple rvs, then for every scalars α and β, the rv $\alpha X + \beta Y$ is a simple rv and we have

$$\mathbb{E}\left[\alpha X + \beta Y\right] = \alpha \mathbb{E}\left[X\right] + \beta \mathbb{E}\left[Y\right]$$

D. *Monotonicity:* If X and Y are simple rvs such that $X \leq Y$, then

$$\mathbb{E}\left[X\right] \leq \mathbb{E}\left[Y\right].$$

In particular, if $X \geq 0$, then $\mathbb{E}[X] \geq 0$.

E. If X is a simple rv, then we have the inequality

$$|\mathbb{E}[X]| \leq \mathbb{E}[|X|].$$

Proof: With $X = \sum_{k \in I} a_k \mathbf{1}[A_k]$, elementary algebra shows that

$$\left| \sum_{k \in I} a_k \mathbb{P}[A_k] \right| \leq \sum_{k \in I} |a_k| \mathbb{P}[A_k].$$

F. If X is a simple rv with $X \geq 0$, then $\mathbb{E}[X] = 0$ implies $X = 0$ $\mathbb{P}$-a.s.

Proof: The condition $X \geq 0$ implies $a_k \geq 0$ for all k in I. Now, if

$$\sum_{k \in I} a_k \mathbb{P}[A_k] = 0,$$

then for all k in I we must have $a_k \mathbb{P}[A_k] = 0$. Thus, either $\mathbb{P}[A_k] = 0 > 0$ in which case $a_k = 0$, or $\mathbb{P}[A_k] = 0$. The set $A_0 = \cup_{k \in I:\, \mathbb{P}[A_k]=0} \mathbb{P}[A_k]$ has probability measure zero, i.e., $\mathbb{P}[A_0] = 0$, and $X = 0$ on the event $\Omega - A_0$ (which has probability measure 1) – Thus, $X = 0$ $\mathbb{P}$-a.s.

Non-negative rvs

Consider a rv $X : \Omega \rightarrow \mathbb{R}_+$ and let the monotone sequence of simple rvs $\{X_n,\ n = 1, 2, \ldots\}$ given by (13) and which approximate X from below. We define

$$\mathbb{E}[X] \equiv \lim_{n \rightarrow \infty} \mathbb{E}[X_n].$$

Note that $\mathbb{E}[X]$ always exists as an element in $[0, +\infty]$ due to the fact that the sequence $\{\mathbb{E}[X_n],\ n = 1, 2, \ldots\}$ is increasing in $\mathbb{R}_+$.

The limit is *independent* of the approximating sequence: If $\{X_n,\ n = 1, 2, \ldots\}$ and $\{Y_n,\ n = 1, 2, \ldots\}$ are two monotone sequences of simple $\mathbb{R}_+$-valued rvs which approximate X from below, i.e.,

$$X_n \leq X_{n+1} \leq X \quad \text{and} \quad Y_n \leq Y_{n+1} \leq X, \quad n = 1, 2, \ldots$$

with

$$\lim_{n\to\infty} X_n = X \quad \text{and} \quad \lim_{n\to\infty} Y_n = X.$$

Then,

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \lim_{n\to\infty} \mathbb{E}[Y_n],$$

and the common value is $\mathbb{E}[X]$.

H. For non-negative rvs X and Y (so both $\mathbb{E}[X]$ and $\mathbb{E}[Y]$ exist in $[0,\infty]$), the comparison $X \leq Y$ implies

$$\mathbb{E}[X] \leq \mathbb{E}[Y].$$

Proof: For each $n = 1, 2, \ldots$, define the simple "staircase" rvs $X_n, Y_n : \Omega \to \mathbb{R}_+$ defined according to (13), namely

$$X_n = \sum_{m=0}^{n-1} \sum_{k=0}^{2^n-1} \left(m + k2^{-n}\right) \mathbf{1}\left[m + k2^{-n} < X \leq m + (k+1)2^{-n}\right] \tag{15}$$

and

$$Y_n = \sum_{m=0}^{n-1} \sum_{k=0}^{2^n-1} \left(m + k2^{-n}\right) \mathbf{1}\left[m + k2^{-n} < X \leq m + (k+1)2^{-n}\right] \tag{16}$$

For each $n = 1, 2, \ldots$, we have $X_n \leq Y_n$ whence $\mathbb{E}[X_n] \leq \mathbb{E}[Y_n]$. Let n go to infinity in this inequality. We conclude by noting that $\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X]$ and $\lim_{n\to\infty} \mathbb{E}[Y_n] = \mathbb{E}[Y]$.

General case

Consider a rv $X : \Omega \to \mathbb{R}$. We introduce the $\mathbb{R}_+$-valued rvs X^+ and X^- given by

$$X^+ \equiv \max(0, X) \quad \text{and} \quad X^- \equiv \max(0, -X).$$

It is plain that

$$X = X^+ - X^- \quad \text{and} \quad |X| = X^+ + X^-.$$

Note that $\mathbb{E}[X^+]$ and $\mathbb{E}[X^-]$ are both well defined (possibly infinite).

We now define

$$\mathbb{E}[X] \equiv \mathbb{E}\left[X^+\right] - \mathbb{E}\left[X^-\right]$$

provided at least one of the quantities $\mathbb{E}[X^+]$ and $\mathbb{E}[X^-]$ is finite. Thus, three cases are possible:

We have (i) $\mathbb{E}[X]$ finite if both $\mathbb{E}[X^+]$ and $\mathbb{E}[X^-]$ are finite, (ii) $\mathbb{E}[X] = \infty$ if $\mathbb{E}[X^+] = \infty$ and $\mathbb{E}[X^-]$ is finite, and (iii) $\mathbb{E}[X] = -\infty$ if $\mathbb{E}[X^-] = \infty$ and $\mathbb{E}[X^+]$ is finite. When $\mathbb{E}[X^+] = \mathbb{E}[X^-] = \infty$, the expectation of X does not exist.

In general, $\mathbb{E}[X]$ *exists* and is *finite* if and only if $\mathbb{E}[|X|]$ is *finite*. In particular, when X is bounded, i.e., there exists a constant $M > 0$ such that $|X| \leq M$ $\mathbb{P}$-a.s., then

$$|X_n| \leq M \quad \mathbb{P} - \text{a.s.}$$

for every $n = 1, 2, \ldots$, it is plain that $\mathbb{E}[X]$ exists and is finite with $-M \leq \mathbb{E}[X] \leq M$.

G. If $\mathbb{E}[X]$ exists, then for any scalar α in $\mathbb{R}$, $\mathbb{E}[\alpha X]$ exists and

$$\mathbb{E}[\alpha X] = \alpha \mathbb{E}[X].$$

H. If both $\mathbb{E}[|X|] < \infty$ and $\mathbb{E}[|Y|] < \infty$, then

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$

K. If both $\mathbb{E}[X]$ and $\mathbb{E}[Y]$ exist, then $X \leq Y$ implies

$$\mathbb{E}[X] \leq \mathbb{E}[Y]$$

Proof: It is a simple consequence of the observation that $X^+ \leq Y^+$ and $Y^- \leq X^-$.

L. If $\mathbb{E}[X]$ exists, then

$$|\mathbb{E}[X]| \leq \mathbb{E}[|X|].$$

Proof: Note that

$$\mathbb{E}[X] = \mathbb{E}\left[X^+\right] - \mathbb{E}\left[X^-\right]$$

so that

$$|\mathbb{E}[X]| \leq \mathbb{E}\left[X^+\right] + \mathbb{E}\left[X^-\right] = \mathbb{E}[|X|].$$

M. If $\mathbb{E}[X]$ exists, then for every A in $\mathcal{F}$, $\mathbb{E}[\mathbf{1}[A]\, X]$ exists; it is finite if $\mathbb{E}[X]$ is finite.

N. If $X = 0$ a.s., then $\mathbb{E}[X] = 0$

P. If $X = Y$ a.s. with $\mathbb{E}[|X|] < \infty$, then $\mathbb{E}[|Y|] < \infty$ and $\mathbb{E}[X] = \mathbb{E}[Y]$

Q. If $X \geq 0$ and $\mathbb{E}[X] = 0$, then $X = 0$ a.s.

As pointed earlier, with rv $X : \Omega \to \mathbb{R}^p$ we have

$$
\begin{array}{lll}
\mathcal{E} & : (\Omega, \mathcal{F}, \mathbb{P}) & \int_\Omega X(\omega) d\mathbb{P}(\omega) \\
\mathcal{E}_X & : (\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p), \mathbb{P}_X) & \int_{\mathbb{R}} x dP_X(x).
\end{array}
$$

The expectation operation is basically defined as a Lebesgue-Stieltjes integral (either under $\mathbb{P}$ or under $\mathbb{P}_X$). We shall write alternatively (with $p = 1$),

$$\mathbb{E}[X] = \int_\Omega X(\omega) d\mathbb{P}(\omega)$$

and

$$\mathbb{E}[X] = \int_{\mathbb{R}} x dP_X(x) = \int_{\mathbb{R}} x dF_X(x).$$

Change of variable formula

Consider an $\mathbb{R}^p$-valued rv $X : \Omega \to \mathbb{R}^p$. With Borel mapping $g : \mathbb{R}^p \to \mathbb{R}$, it holds that

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}^p} g(x) dF_X(x)$$

with the understanding that if one of the quantities is well defined, so is the other and their values coincide.

Proof: If $g : \mathbb{R}^p \to \mathbb{R}$ is of the form

$$g(x) = \mathbf{1}[x \in B], \quad x \in \mathbb{R}^p$$

for some Borel set B in $\mathcal{B}(\mathbb{R}^p)$, then

$$\mathbb{E}[g(X)] = \mathbb{P}[X \in B] = \mathbb{P}_X[B] = \mathbb{E}_X[g(\cdot)] = \int_{\mathbb{R}^p} g(x) dF_X(x)$$

Assume now that $g : \mathbb{R}^p \to \mathbb{R}$ is simple in the sense that

$$g(x) = \sum_{i \in I} g_i \mathbf{1}\left[x \in B_i\right], \quad x \in \mathbb{R}^p$$

Then,

$$\begin{aligned}
\mathbb{E}\left[g(X)\right] &= \mathbb{E}\left[\sum_{i \in I} g_i \mathbf{1}\left[X \in B_i\right]\right] \\
&= \sum_{i \in I} g_i \mathbb{E}\left[\mathbf{1}\left[X \in B_i\right]\right] \\
&= \sum_{i \in I} g_i \mathbb{P}\left[X \in B_i\right] \\
&= \sum_{i \in I} g_i \int_{\mathbb{R}^p} \mathbf{1}\left[B_i\right](x) dF_X(x) \\
&= \int_{\mathbb{R}^p} g(x) dF_X(x) \qquad (17)
\end{aligned}$$

If $g : \mathbb{R}^p \to \mathbb{R}_+$, then we generate the sequence of simple mappings $\{g_n,\ n = 1, 2, \ldots\}$ where for each $n = 1, 2, \ldots$, the Borel mapping $g_n : \mathbb{R}^p \to \mathbb{R}$ is given by

$$g_n(x) = \sum_{m=0}^{n-1} \sum_{k=0}^{2^n - 1} \frac{k}{2^n} \mathbf{1}\left[\frac{k}{2^n} < x \leq \frac{k+1}{2^n}\right], \quad x \in \mathbb{R}^p$$

We already have

$$\mathbb{E}\left[g_n(X)\right] = \int_{\mathbb{R}^p} g_n(x) dF_X(x), \quad n = 1, 2, \ldots$$

and the conclusion

$$\mathbb{E}\left[g(X)\right] = \int_{\mathbb{R}^p} g(x) dF_X(x),$$

follows by the Monotone Convergence Theorem (under $\mathbb{P}$ and $\mathbb{P}_X$).

In the general case $g : \mathbb{R}^p \to \mathbb{R}$, write

$$g(x) = g(x)^+ - g(x)^-, \quad x \in \mathbb{R}^p$$

and by linearity, we get

$$\mathbb{E}\left[g(X)\right] = \mathbb{E}\left[g(X)^+\right] - \mathbb{E}\left[g(X)^-\right]$$

Convergence results for expectations

Consider a sequence $\{X, Y, Z, X_n,\ n = 1, 2, \ldots\}$ of $\mathbb{R}$-valued rvs.

Monotone Convergence Theorem

With $Y \leq X_n \leq X_{n+1}$ for all $n = 1, 2, \ldots$ where $\mathbb{E}\left[Y\right] > -\infty$, we have

$$\lim_{n\to\infty} \mathbb{E}\left[X_n\right] = \mathbb{E}\left[\lim_{n\to\infty} X_n\right] \tag{18}$$

monotonically. With $X \geq X_n \geq X_{n+1}$ for all $n = 1, 2, \ldots$ where $\mathbb{E}\left[X\right] < \infty$, we have

$$\lim_{n\to\infty} \mathbb{E}\left[X_n\right] = \mathbb{E}\left[\lim_{n\to\infty} X_n\right] \tag{19}$$

monotonically

An important consequence of the Monotone Convergence Theorem is as follows: Let $\{X_n,\ n = 1, 2, \ldots\}$ denote a sequence of $\mathbb{R}_+$-valued rvs. It follows from the Monotone Convergence Theorem that

$$\mathbb{E}\left[\sum_{n=1}^{\infty} X_n\right] = \sum_{n=1}^{\infty} \mathbb{E}\left[X_n\right]$$

This is because, with

$$S_n = \sum_{k=1}^{n} X_k, \quad n = 1, 2, \ldots$$

non-negativity implies $0 \leq S_n \leq S_{n+1}$ for all $n = 1, 2, \ldots$, whence

$$\lim_{n\to\infty} \mathbb{E}\left[S_n\right] = \mathbb{E}\left[\lim_{n\to\infty} S_n\right]$$

by (18). By linearity, we have

$$\mathbb{E}\left[S_n\right] = \sum_{k=1}^{n} \mathbb{E}\left[X_k\right]$$

so that $\lim_{n\to\infty} \mathbb{E}\left[S_n\right] = \sum_{n=1}^{\infty} \mathbb{E}\left[X_n\right]$, while $\lim_{n\to\infty} S_n = \sum_{n=1}^{\infty} X_n$.

Fatou's Lemma

With $X_n \geq Y$ for all $n = 1, 2, \ldots$ where $\mathbb{E}[Y] > -\infty$, we have

$$\mathbb{E}\left[\liminf_{n\to\infty} X_n\right] \leq \liminf_{n\to\infty} \mathbb{E}[X_n]. \tag{20}$$

With $X_n \leq Y$ for all $n = 1, 2, \ldots$ where $\mathbb{E}[Y] < \infty$, we have

$$\limsup_{n\to\infty} \mathbb{E}[X_n] \leq \mathbb{E}\left[\limsup_{n\to\infty} X_n\right]. \tag{21}$$

Counterexample: Take $\Omega = \mathbb{R}$, $\mathcal{F} = \mathcal{B}(\mathbb{R})$ and the rvs $\{X_n,\ n = 1, 2, \ldots\}$ are given by

$$X_n(\omega) = \begin{cases} 0 & \text{if } \omega \notin [\frac{1}{n}, \frac{2}{n}] \\ \\ -n & \text{if } \omega \in [\frac{1}{n}, \frac{2}{n}] \end{cases}, \qquad \begin{array}{c} \omega \in \Omega \\ n = 2, 3, \ldots \end{array}$$

Bounded Convergence Theorem

Assume that there exists a rv $X : \Omega \to \mathbb{R}$ such that $\lim_{n\to\infty} X_n = X$. If there exists $M > 0$ such that for each $n = 1, 2, \ldots$,

$$|X_n| \leq M,$$

then

$$\mathbb{E}\left[\lim_{n\to\infty} X_n\right] = \lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X]. \tag{22}$$

Dominated Convergence Theorem

Assume that there exists a rv $X : \Omega \to \mathbb{R}$ such that $\lim_{n\to\infty} X_n = X$. If there exists a rv $Y : \Omega \to \mathbb{R}_+$ such that

$$|X_n| \leq Y$$

for all $n = 1, 2, \ldots$ with $\mathbb{E}[Y] < \infty$, then

$$\mathbb{E}\left[\lim_{n\to\infty} X_n\right] = \lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X] \tag{23}$$

Independence and expectations

Consider two rvs $X, Y : \Omega \to \mathbb{R}_+$ which are assumed to be *independent*.

- If X and Y are simple rvs, then the product rv XY is also a simple rv and the relation
$$\mathbb{E}\left[XY\right] = \mathbb{E}\left[X\right]\mathbb{E}\left[Y\right]$$
holds

- If X and Y are non-negative rvs, then $\mathbb{E}\left[X\right]$ and $\mathbb{E}\left[Y\right]$ are always defined (although possibly infinite) and the relation
$$\mathbb{E}\left[XY\right] = \mathbb{E}\left[X\right]\mathbb{E}\left[Y\right]$$
still holds.

- In the general case,
$$\mathbb{E}\left[XY\right] = \mathbb{E}\left[X\right]\mathbb{E}\left[Y\right]$$
whenever $\mathbb{E}\left[|X|\right]$ and $\mathbb{E}\left[|Y|\right]$ are finite.

Fact: Consider rvs the $X_1 : \Omega \to \mathbb{R}^{p_1}$, ..., $X_k : \Omega \to \mathbb{R}^{p_k}$ which are mutually independent. With Borel mappings $g_1 : \mathbb{R}^{p_1} \to \mathbb{R}$, ..., $g_k : \mathbb{R}^{p_k} \to \mathbb{R}$, define the rvs
$$Y_\ell = g_\ell(X_\ell), \quad \ell = 1, \ldots, k.$$

The $\mathbb{R}$-valued rvs $Y_1, \ldots, Y_k$ are mutually independent, and
$$\mathbb{E}\left[\prod_{\ell=1}^{k} Y_\ell\right] = \prod_{\ell=1}^{k} \mathbb{E}\left[Y_\ell\right]$$
whenever $\mathbb{E}\left[|Y_\ell|\right] < \infty$ for all $\ell = 1, \ldots, k$.

A definition

Let $\mathcal{D}$ be a σ-field on Ω contained in $\mathcal{F}$, i.e., $\mathcal{D} \subseteq \mathcal{F}$. A rv $\mathbb{R}^p$-valued rv $X : \Omega \to \mathbb{R}^p$ is said to be *$\mathcal{D}$-measurable* if

$$[X \in B] \in \mathcal{D}, \quad B \in \mathcal{B}(\mathbb{R}^p).$$

This definition is often used when the σ-field $\mathcal{D}$ is itself generated by some rv $Y : \Omega \to \mathbb{R}^q$; this σ-field is denoted $\sigma(Y)$ and is defined

$$\sigma(Y) \equiv \left\{ Y^{-1}(C) : \ C \in \mathcal{B}(\mathbb{R}^q) \right\}$$

as expected.

Important fact: Assume the σ-field $\mathcal{D}$ to be generated by some rv $Y : \Omega \to \mathbb{R}^q$, so that $\mathcal{D} = \sigma(Y)$:

(i) For any Borel mapping $g : \mathbb{R}^q \to \mathbb{R}$, the rv $X = g(Y)$ is $\mathcal{D}$-measurable.

(ii) Conversely, any $\mathcal{D}$-measurable rv $X : \Omega \to \mathbb{R}$ can be written in the form $X = g(Y)$ for some Borel mapping $g : \mathbb{R}^q \to \mathbb{R}$.

(i) The conclusion is immediate from the fact that

$$\begin{aligned} [X \in B] &= [g(Y) \in B] \\ &= [Y \in g^{-1}(B)] \in \mathcal{D}, \quad B \in \mathcal{B}(\mathbb{R}). \end{aligned} \tag{24}$$

since Y is $\mathcal{D}$-measurable and $g^{-1}(B)$ belongs to $\mathcal{B}(\mathbb{R}^p)$ by the Borel measurability of g.

(ii) Conversely, assume that the rv $X : \Omega \to \mathbb{R}$ is $\mathcal{D}$-measurable. The proof proceeds in three standard steps:

- First assume that $X = \mathbf{1}[D]$ for some D in $\sigma(Y)$, in which case $D = [Y \in C]$ for some C in $\mathcal{B}(\mathbb{R}^q)$. It is now plain that $X = g_C(Y)$ with Borel mapping $g_C : \mathbb{R}^q \to \mathbb{R}$ given by

$$g_C(y) = \begin{cases} 0 & \text{if } y \notin C \\ \\ 0 & \text{if } y \in C. \end{cases}$$

The desired conclusion is readily seen to hold for simple $\mathcal{D}$-measurable rvs of the form

$$X = \sum_{i \in I} a_i \mathbf{1}\left[D_i\right]$$

where I is a countable index, $\{a_i,\ i \in I\}$ are scalars and $\{D_i,\ i \in I\}$ form a $\mathcal{D}$-partition of Ω. Indeed, we have $X = g(Y)$ with Borel mapping $g : \mathbb{R}^q \to \mathbb{R}$ given by

$$g(y) = \sum_{i \in I} a_i g_{C_i}$$

where for each i in I, we have $D_i = [Y \in C_i]$ for C_i in $\mathcal{B}(\mathbb{R}^p)$.

- For any non-negative $\mathcal{D}$-measurable rv $X : \Omega \to \mathbb{R}_+$, we introduce the usual sequence of simple rvs $\{X_n,\ n = 1, 2, \ldots\}$ given by

$$X_n = \sum_{m=0}^{n-1} \sum_{k=0}^{2^n-1} \frac{k}{2^n} \mathbf{1}\left[\frac{k}{2^n} < X \leq \frac{k+1}{2^n}\right], \quad n = 1, 2, \ldots$$

with $\lim_{n\to\infty} X_n = X$. Obviously, the simple rvs $\{X_n,\ n = 1, 2, \ldots\}$ are all $\mathcal{D}$-measurable, hence by the last part of the proof, for each $n = 1, 2, \ldots$, there exists a Borel mapping $g_n : \mathbb{R}^q \to \mathbb{R}$ such that

$$X_n = g_n(Y), \quad n = 1, 2, \ldots$$

and that

$$X(\omega) = \lim_{n\to\infty} X_n(\omega) = \lim_{n\to\infty} g_n(Y(\omega)), \quad \omega \in \Omega.$$

Now define the subset $L \subseteq \mathbb{R}^q$ by

$$L \equiv \{y \in \mathbb{R}^q : \ \lim_{n\to\infty} g_n(y) \text{ exists in } \mathbb{R}\}$$

The set L is a Borel subset of $\mathbb{R}^q$, whence the mapping $g : \mathbb{R}^q \to \mathbb{R}$ given by

$$g(y) \equiv \begin{cases} \lim_{n\to\infty} g_n(y) & \text{if } y \in L \\ \\ 0 & \text{if } y \notin L \end{cases}$$

is a Borel mapping $\mathbb{R}^q \to \mathbb{R}$. By construction it is plain that $X = g(Y)$ since $Y(\omega)$ lies in L for each ω in Ω.

- The case of any $\mathcal{D}$-measurable rv $X : \Omega \to \mathbb{R}$ is handled in the usual manner: Just write
$$X = X^+ - X^-$$
and apply the last conclusion to each of the rvs X^+ and X^-. In particular, there exist Borel mappings $g_+ : \mathbb{R}^q \to \mathbb{R}$ and $g_- : \mathbb{R}^q \to \mathbb{R}$ such that $X^+ = g_+(Y)$ and $X^- = g_-(Y)$. The desired Borel mapping $g : \mathbb{R}^q \to \mathbb{R}$ is simply
$$g(y) = g_+(y) - g_-(y), \quad y \in \mathbb{R}^q.$$

Conditional distributions and condition expectations

Let D be an event in $\mathcal{F}$, and let $X : \Omega \to \mathbb{R}$ be an $\mathbb{R}$-valued. With $\mathbb{P}[D] > 0$, we can define the *conditional probability distribution* of X given D, namely

$$\mathbb{P}[X \leq x | D] = \frac{\mathbb{P}[[X \leq x] \cap D]}{\mathbb{P}[D]}, \quad x \in \mathbb{R}.$$

It is now possible to define the conditional expectation of X given D; it is simply the expectation of the rv X evaluated under the conditional probability measure $\mathbb{Q}_D : \mathcal{F} \to [0,1]$; it will be denoted $\mathbb{E}[X|D]$. It is easy to see from the definition of expectation that

$$\mathbb{E}[X|D] = \frac{\mathbb{E}[\mathbf{1}[D]X]}{\mathbb{P}[D]}.$$

This quantity exists as soon as $\mathbb{E}[X]$ is well defined. This fact and the last expression for it can be seen by the usual three step process: First for indicator rvs and simple rvs, then for non-negative rvs and finally for arbitrary rvs (by the standard decomposition).

When $\mathbb{P}[D] = 0$, it is will be convenient to take $\mathbb{Q}_D : \mathcal{F} \to [0,1]$ to be an arbitrary probability measure on $(\Omega, \mathcal{F})$ – To be revisited. However, regardless of the choice we will always have

$$\mathbb{E}[\mathbf{1}[D]X] = \mathbb{E}[X|D] \cdot \mathbb{P}[D]. \tag{25}$$

Consider an $\mathcal{F}$-partition $\{D_i,\ i \in I\}$ where I is a countable index set, i.e.,

$$D_i \cap D_j = \emptyset, \quad \begin{matrix} i \neq j \\ i, j \in I \end{matrix}$$

and

$$\cup_{i \in I} D_i = \Omega.$$

It is plain that

$$\sum_{i \in I} \mathbf{1}[D_i] = 1. \tag{26}$$

Throughout the events in the partition are assumed to be non-empty.

Set $\mathcal{D} = \sigma(D_i,\ i \in I)$. Note that any element D of $\mathcal{D}$ is necessarily of the form

$$D = \cup_{j \in J} D_j \tag{27}$$

for some countable subset $J \subseteq I$ (possibly empty if $D = \emptyset$ or $J = I$ if $D = \Omega$). The decomposition

$$\sum_{j \in IJ} \mathbf{1}[D_i] = \mathbf{1}[D]. \tag{28}$$

will be used on several occasions.

Fact: Consider a rv $X : \Omega \to \mathbb{R}^p$ which is $\mathcal{D}$-measurable where $\mathcal{D} = \sigma(D_i,\ i \in I)$. For each i in I, the rv X is constant on the event D_i so that the set $\{X(\omega),\ \omega \in \Omega\}$ is a countable set of points in $\mathbb{R}^p$.

Proof: For each x in $\mathbb{R}^p$, the $\mathcal{D}$-measurability of X implies that $[X = x]$ is an element in $\mathcal{D}$. The result immediately follows since any element D of $\mathcal{D}$ is necessarily of the form (27) for some countable subset $J \subseteq I$.

An important definition

Consider a rv $X : \Omega \to \mathbb{R}^p$. For each B in $\mathcal{B}(\mathbb{R}^p)$ define the rv

$$\begin{aligned} \mathbb{P}[X \in B|\mathcal{D}] &\equiv \sum_{i \in I} \mathbb{P}[X \in B|D_i]\, \mathbf{1}[D_i] \\ &= \sum_{i \in I} \mathbb{Q}_{D_i}[X \in B] \mathbf{1}[D_i]. \end{aligned} \tag{29}$$

This defines the *conditional probability* of $[X \in B]$ *given* $\mathcal{D}$. This rv is clearly a $\mathcal{D}$-measurable rv in the sense that

$$[\mathbb{P}[X \in B|\mathcal{D}] \in C] \in \mathcal{D}, \quad C \in \mathcal{B}(\mathbb{R})$$

as we note that

$$\begin{aligned}[\mathbb{P}[X \in B|\mathcal{D}] \in C] &= \cup_{i\in I}\left([\mathbb{P}[X \in B|\mathcal{D}] \in C] \cap D_i\right) \\ &= \cup_{i\in I}\left([\mathbb{P}[X \in B|D_i] \in C] \cap D_i\right) \\ &= \cup_{i\in I:\ \mathbb{P}[X\in B|D_i]\in C} D_i.\end{aligned}$$

We also observe that

$$\begin{aligned}\mathbb{E}[\mathbb{P}[X \in B|\mathcal{D}]] &= \mathbb{E}\left[\sum_{i\in I} \mathbb{P}[X \in B|D_i]\,\mathbf{1}[D_i]\right] \\ &= \sum_{i\in I} \mathbb{P}[X \in B|D_i]\,\mathbb{E}[\mathbf{1}[D_i]] \\ &= \sum_{i\in I} \mathbb{P}[X \in B|D_i]\,\mathbb{P}[D_i] \\ &= \mathbb{P}[X \in B]\end{aligned}$$

by the law of total probability.

Consider a rv $X : \Omega \to \mathbb{R}$ We can now also define the *conditional expectation* of X *given* $\mathcal{D}$ as

$$\mathbb{E}[X|\mathcal{D}] \equiv \sum_{i\in I} \mathbb{E}[X|D_i]\,\mathbf{1}[D_i]$$

where $\mathbb{E}[X|D_i]$ is the expectation of X under the conditional probability distribution of X given D_i.

The rv $\mathbb{E}[X|\mathcal{D}]$ is clearly a $\mathcal{D}$-measurable rv in the sense that

$$[\mathbb{E}[X|\mathcal{D}] \in C] \in \mathcal{D}, \quad C \in \mathcal{B}(\mathbb{R}).$$

Indeed we have

$$\begin{aligned}[\mathbb{E}[X \in B|\mathcal{D}] \in C] &= \cup_{i\in I}\left([\mathbb{E}[X \in B|\mathcal{D}] \in C] \cap D_i\right) \\ &= \cup_{i\in I}\left([\mathbb{E}[X \in B|D_i] \in C] \cap D_i\right) \\ &= \cup_{i\in I:\ \mathbb{E}[X\in B|D_i]\in C} D_i.\end{aligned}$$

Consider two $\mathcal{F}$-partitions of Ω, say $\{D_i,\ i \in I\}$ and $\{D'_k,\ k \in K\}$. We say that the partition $\{D'_k,\ k \in K\}$ is a *refinement* of $\{D_i,\ i \in I\}$ if for every i in I it holds that

$$D_i = \cup_{k \in K_i} D'_k$$

for some non-empty subset $K_i \subseteq K$ – Note that $\{K_i,\ i \in I\}$ is a partition of K. The inclusion $\mathcal{D} \subseteq \mathcal{D}'$ obviously holds where $\mathcal{D} = \sigma(D_i,\ i \in I)$ and $\mathcal{D}' = \sigma(D'_k,\ k \in K)$.

Iterated conditioning (I)

For any rv $X : \Omega \to \mathbb{R}$ with $\mathbb{E}[|X|] < \infty$, it holds that

$$\mathbb{E}[\mathbb{E}[X|\mathcal{D}]] = \mathbb{E}[X] \tag{30}$$

Proof: It suffices to consider $X : \Omega \to \mathbb{R}_+$. We have

$$\begin{aligned}
\mathbb{E}[\mathbb{E}[X|\mathcal{D}]] &= \mathbb{E}\left[\sum_{i \in I} \mathbb{E}[X|D_i] \mathbf{1}[D_i]\right] \\
&= \sum_{i \in I} \mathbb{E}[\mathbb{E}[X|D_i] \mathbf{1}[D_i]] \\
&= \sum_{i \in I} \mathbb{E}[X|D_i] \mathbb{P}[D_i] \\
&= \sum_{i \in I:\ \mathbb{P}[D_i] > 0} \frac{\mathbb{E}[\mathbf{1}[D_i] X]}{\mathbb{P}[D_i]} \mathbb{P}[D_i] \\
&= \sum_{i \in I:\ \mathbb{P}[D_i] > 0} \mathbb{E}[\mathbf{1}[D_i] X] \\
&= \mathbb{E}\left[\left(\sum_{i \in I:\ \mathbb{P}[D_i] > 0} \mathbf{1}[D_i]\right) X\right] \\
&= \mathbb{E}[X] \qquad (31)
\end{aligned}$$

upon using (28) together with the fact that $\sum_{i \in I:\ \mathbb{P}[D_i] = 0} \mathbf{1}[D_i] = 0$ $\mathbb{P}$-a.s.

Iterated conditioning (II)

For any rv $X : \Omega \to \mathbb{R}$ with $\mathbb{E}[|X|] < \infty$, it holds that

$$\mathbb{E}[\mathbb{E}[X|\mathcal{D}] |\mathcal{D}'] = \mathbb{E}[X|\mathcal{D}] \quad \mathbb{P}\text{-a.s.} \tag{32}$$

Proof: It suffices to consider $X : \Omega \to \mathbb{R}_+$. With

$$\mathbb{E}\left[X|\mathcal{D}\right] = \sum_{i\in I} \mathbb{E}\left[X|D_i\right] \mathbf{1}\left[D_i\right],$$

we have

$$\begin{aligned}\mathbb{E}\left[\mathbb{E}\left[X|\mathcal{D}\right]|\mathcal{D}'\right] &= \mathbb{E}\left[\sum_{i\in I} \mathbb{E}\left[X|D_i\right] \mathbf{1}\left[D_i\right]|\mathcal{D}'\right] \\ &= \sum_{i\in I} \mathbb{E}\left[X|D_i\right] \mathbb{E}\left[\mathbf{1}\left[D_i\right]|\mathcal{D}'\right]. \end{aligned} \tag{33}$$

For each i in I it holds

$$\begin{aligned}\mathbb{E}\left[\mathbf{1}\left[D_i\right]|\mathcal{D}'\right] &= \sum_{k\in K} \mathbb{E}\left[\mathbf{1}\left[D_i\right]|D'_k\right] \mathbf{1}\left[D'_k\right] \\ &= \sum_{k\in K:\ \mathbb{P}\left[D'_k\right]>0} \frac{\mathbb{P}\left[D_i \cap D'_k\right]}{\mathbb{P}\left[D'_k\right]} \mathbf{1}\left[D'_k\right] \quad \mathbb{P}\text{-a.s.} \\ &= \sum_{k\in K_i:\ \mathbb{P}\left[D'_k\right]>0} \frac{\mathbb{P}\left[D'_k\right]}{\mathbb{P}\left[D'_k\right]} \mathbf{1}\left[D'_k\right] \\ &= \sum_{k\in K_i:\ \mathbb{P}\left[D'_k\right]>0} \mathbf{1}\left[D'_k\right] \\ &= \sum_{k\in K_i} \mathbf{1}\left[D'_k\right] \quad \mathbb{P}\text{-a.s.} \\ &= \mathbf{1}\left[D_i\right]. \end{aligned} \tag{34}$$

We have used repeatedly the fact that

$$\mathbb{P}\left[\cup_{k\in K_i:\ \mathbb{P}\left[D'_k\right]=0} D'_k\right] = \sum_{k\in K_i:\ \mathbb{P}\left[D'_k\right]=0} \mathbb{P}\left[D'_k\right] = 0.$$

From this fact we conclude that

$$\mathbb{E}\left[\mathbb{E}\left[X|\mathcal{D}\right]|\mathcal{D}'\right] = \sum_{i\in I} \mathbb{E}\left[X|D_i\right] \mathbf{1}\left[D_i\right] = \mathbb{E}\left[X|\mathcal{D}\right] \quad \mathbb{P}\text{-a.s.}$$

Iterated conditioning (III)

For any rv $X : \Omega \to \mathbb{R}$ with $\mathbb{E}[|X|] < \infty$, it holds that

$$\mathbb{E}[\mathbb{E}[X|\mathcal{D}'] | \mathcal{D}] = \mathbb{E}[X|\mathcal{D}] \quad \mathbb{P}\text{-a.s.} \tag{35}$$

Proof: It suffices to consider $X : \Omega \to \mathbb{R}_+$. With

$$\begin{aligned} \mathbb{E}[X|\mathcal{D}'] &= \sum_{k \in K} \mathbb{E}[X|D'_k] \mathbf{1}[D'_k] \\ &= \sum_{k \in K:\ \mathbb{P}[D'_k] > 0} \mathbb{E}[X|D'_k] \mathbf{1}[D'_k] \quad \mathbb{P}\text{-a.s.} \end{aligned} \tag{36}$$

since

$$\mathbb{P}\left[\cup_{k \in K:\ \mathbb{P}[D'_k] = 0} D'_k\right] = \sum_{k \in K:\ \mathbb{P}[D'_k] = 0} \mathbb{P}[D'_k] = 0.$$

It follows that

$$\mathbb{E}\left[\sum_{k \in K} \mathbb{E}[X|D'_k] \mathbf{1}[D'_k] | \mathcal{D}\right] = \mathbb{E}\left[\sum_{k \in K:\ \mathbb{P}[D'_k] > 0} \mathbb{E}[X|D'_k] \mathbf{1}[D'_k] \Big| \mathcal{D}\right].$$

Therefore, we have

$$\begin{aligned} \mathbb{E}[\mathbb{E}[X|\mathcal{D}'] | \mathcal{D}] &= \mathbb{E}\left[\sum_{k \in K} \mathbb{E}[X|D'_k] \mathbf{1}[D'_k] \Big| \mathcal{D}\right] \\ &= \mathbb{E}\left[\sum_{k \in K:\ \mathbb{P}[D'_k] > 0} \mathbb{E}[X|D'_k] \mathbf{1}[D'_k] \Big| \mathcal{D}\right] \\ &= \sum_{k \in K:\ \mathbb{P}[D'_k] > 0} \mathbb{E}[X|D'_k] \mathbb{E}[\mathbf{1}[D'_k] | \mathcal{D}]. \end{aligned} \tag{37}$$

For each k in K such that $\mathbb{P}[D'_k] > 0$, we have

$$\mathbb{E}[X|D'_k] = \frac{\mathbb{E}[\mathbf{1}[D'_k] X]}{\mathbb{P}[D'_k]}$$

and

$$
\begin{aligned}
\mathbb{E}\left[\mathbf{1}\left[D'_k\right]|\mathcal{D}\right] &= \sum_{i\in I} \mathbb{P}\left[D'_k|D_i\right]\mathbf{1}\left[D_i\right] \\
&= \sum_{i\in I:\ \mathbb{P}[D_i]>0} \mathbb{E}\left[\mathbf{1}\left[D'_k\right]|D_i\right]\mathbf{1}\left[D_i\right] \quad \mathbb{P}\text{-a.s.} \\
&= \sum_{i\in I:\ \mathbb{P}[D_i]>0} \frac{\mathbb{P}\left[D'_k\cap D_i\right]}{\mathbb{P}\left[D_i\right]}\mathbf{1}\left[D_i\right] \quad \mathbb{P}\text{-a.s.} \qquad (38)
\end{aligned}
$$

Note that for each i in I, we have

$$
\mathbb{P}\left[D'_k\cap D_i\right] = \begin{cases} \mathbb{P}\left[D'_k\right] & \text{if } k\in K_i \\ \\ 0 & \text{if } k\notin K_i. \end{cases}
$$

With this in mind, collecting earlier expressions we conclude that

$$
\begin{aligned}
&\mathbb{E}\left[\mathbb{E}\left[X|\mathcal{D}'\right]|\mathcal{D}\right] \\
&= \sum_{k\in K:\ \mathbb{P}\left[D'_k\right]>0} \left(\sum_{i\in I:\ \mathbb{P}[D_i]>0} \frac{\mathbb{P}\left[D'_k\cap D_i\right]}{\mathbb{P}\left[D_i\right]}\mathbf{1}\left[D_i\right] \right) \mathbb{E}\left[X|D'_k\right] \quad \mathbb{P}\text{-a.s.} \\
&= \sum_{k\in K:\ \mathbb{P}\left[D'_k\right]>0} \left(\sum_{i\in I:\ \mathbb{P}[D_i]>0} \frac{\mathbb{P}\left[D'_k\cap D_i\right]}{\mathbb{P}\left[D_i\right]}\mathbf{1}\left[D_i\right] \right) \frac{\mathbb{E}\left[\mathbf{1}\left[D'_k\right]X\right]}{\mathbb{P}\left[D'_k\right]} \\
&= \sum_{i\in I:\ \mathbb{P}[D_i]>0} \left(\sum_{k\in K:\ \mathbb{P}\left[D'_k\right]>0} \frac{\mathbb{P}\left[D'_k\cap D_i\right]}{\mathbb{P}\left[D_i\right]} \cdot \frac{\mathbb{E}\left[\mathbf{1}\left[D'_k\right]X\right]}{\mathbb{P}\left[D'_k\right]} \right) \mathbf{1}\left[D_i\right] \\
&= \sum_{i\in I:\ \mathbb{P}[D_i]>0} \left(\sum_{k\in K_i:\ \mathbb{P}\left[D'_k\right]>0} \frac{\mathbb{P}\left[D'_k\cap D_i\right]}{\mathbb{P}\left[D_i\right]} \cdot \frac{\mathbb{E}\left[\mathbf{1}\left[D'_k\right]X\right]}{\mathbb{P}\left[D'_k\right]} \right) \mathbf{1}\left[D_i\right] \\
&= \sum_{i\in I:\ \mathbb{P}[D_i]>0} \left(\sum_{k\in K_i:\ \mathbb{P}\left[D'_k\right]>0} \frac{\mathbb{P}\left[D'_k\right]}{\mathbb{P}\left[D_i\right]} \cdot \frac{\mathbb{E}\left[\mathbf{1}\left[D'_k\right]X\right]}{\mathbb{P}\left[D'_k\right]} \right) \mathbf{1}\left[D_i\right] \\
&= \sum_{i\in I:\ \mathbb{P}[D_i]>0} \frac{1}{\mathbb{P}\left[D_i\right]} \left(\sum_{k\in K_i:\ \mathbb{P}\left[D'_k\right]>0} \mathbb{E}\left[\mathbf{1}\left[D'_k\right]X\right] \right) \mathbf{1}\left[D_i\right]
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{i \in I:\ \mathbb{P}[D_i]>0} \frac{1}{\mathbb{P}[D_i]} \mathbb{E}\left[\left(\sum_{k \in K_i:\ \mathbb{P}\left[D'_k\right]>0} \mathbf{1}\left[D'_k\right] \right) X \right] \mathbf{1}\left[D_i\right] \\
&= \sum_{i \in I:\ \mathbb{P}[D_i]>0} \frac{\mathbb{E}\left[\mathbf{1}\left[D_i\right] X\right]}{\mathbb{P}\left[D_i\right]} \mathbf{1}\left[D_i\right] \\
&= \mathbb{E}\left[X|\mathcal{D}\right] \quad \mathbb{P}\text{-a.s.} \qquad (39)
\end{aligned}
$$

as we use the fact noted earlier that

$$
\sum_{k \in K_i:\ \mathbb{P}\left[D'_k\right]>0} \mathbf{1}\left[D'_k\right] = \mathbf{1}\left[D_i\right] \quad \mathbb{P}\text{-a.s.}
$$

for each i in I.

An important fact

For any rv $X : \Omega \to \mathbb{R}$ with $\mathbb{E}\left[|X|\right] < \infty$, it holds that

$$
\mathbb{E}\left[\mathbf{1}\left[D\right] \mathbb{E}\left[X|\mathcal{D}\right]\right] = \mathbb{E}\left[\mathbf{1}\left[D\right] X\right], \quad D \in \mathcal{D},
$$

or equivalently,

$$
\mathbb{E}\left[\mathbf{1}\left[D\right] \left(X - \mathbb{E}\left[X|\mathcal{D}\right]\right)\right] = 0, \quad D \in \mathcal{D}. \qquad (40)
$$

Proof: Given the structure of the σ-field $\mathcal{D}$, it suffices to show this equality for $D = D_i$ as i ranges over I, namely

$$
\mathbb{E}\left[\mathbf{1}\left[D_i\right] \mathbb{E}\left[X|\mathcal{D}\right]\right] = \mathbb{E}\left[\mathbf{1}\left[D_i\right] X\right],
$$

Since

$$
\mathbb{E}\left[X|\mathcal{D}\right] = \sum_{j \in I} \mathbb{E}\left[X|D_j\right] \mathbf{1}\left[D_j\right] \qquad (41)
$$

we get

$$
\begin{aligned}
\mathbb{E}\left[\mathbf{1}\left[D_i\right] \mathbb{E}\left[X|\mathcal{D}\right]\right] &= \mathbb{E}\left[\mathbf{1}\left[D_i\right] \left(\sum_{j \in I} \mathbb{E}\left[X|D_j\right] \mathbf{1}\left[D_j\right] \right)\right] \\
&= \sum_{j \in I} \mathbb{E}\left[X|D_j\right] \mathbb{E}\left[\mathbf{1}\left[D_i\right] \mathbf{1}\left[D_j\right]\right] \\
&= \mathbb{E}\left[X|D_j\right] \mathbb{P}\left[D_i\right] \qquad (42)
\end{aligned}
$$

since $\mathbb{P}[D_i \cap D_j] = \delta_{ij}\mathbb{P}[D_i]$. If $\mathbb{P}[D_i] > 0$, then

$$\mathbb{E}[X|D_j]\,\mathbb{P}[D_i] = \frac{\mathbb{E}[\mathbf{1}[D_i]\,X]}{\mathbb{P}[D_i]} \cdot \mathbb{P}[D_i] = \mathbb{E}[\mathbf{1}[D_i]\,X]$$

as desired. On the other hand, if $\mathbb{P}[D_i] = 0$, then $\mathbb{E}[\mathbf{1}[D_i]\,\mathbb{E}[X|\mathcal{D}]] = 0$ while $\mathbb{E}[\mathbf{1}[D_i]\,X] = 0$, again as desired.

Another important fact

Consider rvs $X, Z : \Omega \to \mathbb{R}$ with $\mathbb{E}[|X|] < \infty$ and $\mathbb{E}[|XZ|] < \infty$. If Z is a $\mathcal{D}$-measurable rv, then

$$\mathbb{E}[ZX|\mathcal{D}] = Z\mathbb{E}[X|\mathcal{D}] \quad \mathbb{P}\text{-a.s.}$$

We begin by noting that

$$\begin{aligned}\mathbb{E}[ZX|\mathcal{D}] &= \sum_{i\in I} \mathbb{E}[ZX|D_i]\,\mathbf{1}[D_i] \\ &= \sum_{i\in I:\ \mathbb{P}[D_i]>0} \frac{\mathbb{E}[\mathbf{1}[D_i]\,ZX]}{\mathbb{P}[D_i]}\mathbf{1}[D_i] \quad \mathbb{P}\text{-a.s.}\end{aligned} \tag{43}$$

First consider the case when Z is an indicator rv, say $Z = \mathbf{1}[D]$ with $D = \cup_{j\in J} D_j$ for some $J \subseteq I$. Pick i in I with $\mathbb{P}[D_i] > 0$. Since $\mathbf{1}[D] = \sum_{j\in J} \mathbf{1}[D_j]$ we then have

$$\begin{aligned}\mathbb{E}[\mathbf{1}[D_i]\,ZX] &= \mathbb{E}[\mathbf{1}[D_i]\,\mathbf{1}[D]\,X] \\ &= \mathbb{E}[\mathbf{1}[D_i \cap D]\,X] \\ &= \sum_{j\in J} \mathbb{E}[\mathbf{1}[D_i \cap D_j]\,X] \\ &= \begin{cases} \mathbb{E}[\mathbf{1}[D_i]\,X] & \text{if } i \in J \\ \\ 0 & \text{if } i \notin J. \end{cases}\end{aligned} \tag{44}$$

It follows that

$$\mathbb{E}[ZX|\mathcal{D}] = \sum_{i\in I:\ \mathbb{P}[D_i]>0} \frac{\mathbb{E}[\mathbf{1}[D_i]\,ZX]}{\mathbb{P}[D_i]}\mathbf{1}[D_i]$$

$$
\begin{aligned}
&= \sum_{i \in I:\ \mathbb{P}[D_i]>0} \frac{\mathbb{E}\left[\mathbf{1}\left[D_i\right] X\right]}{\mathbb{P}\left[D_i\right]} \delta(i \in J) \mathbf{1}\left[D_i\right] \\
&= \left(\sum_{i \in I:\ \mathbb{P}[D_i]>0} \frac{\mathbb{E}\left[\mathbf{1}\left[D_i\right] X\right]}{\mathbb{P}\left[D_i\right]} \mathbf{1}\left[D_i\right] \right) \mathbf{1}\left[D\right] \\
&= \mathbb{E}\left[X|\mathcal{D}\right] \mathbf{1}\left[D\right]. \qquad (45)
\end{aligned}
$$

This establishes the result when Z is the indicator rv of a $\mathcal{D}$-measurable event.

If Z is a simple $\mathcal{D}$-measurable rv, say

$$
Z = \sum_{k \in K} b_k \mathbf{1}\left[D_k\right]
$$

then the result follows by linearity of expectation.

When $Z \geq 0$, write

$$
ZX = Z(X^+ - X^-) = ZX^+ - ZX^-
$$

and recall that the usual staircase approximations $\{Z_n,\ n = 1, 2, \ldots\}$ to the rv Z are simple non-negative rvs which are $\mathcal{D}$-measurable with $\lim_{n\to\infty} Z_n = Z$. Since the result holds for simple non-negative rvs which are $\mathcal{D}$-measurable, we get

$$
\mathbb{E}\left[Z_n X|\mathcal{D}\right] = Z_n \mathbb{E}\left[X|\mathcal{D}\right], \quad n = 1, 2, \ldots
$$

Letting n go to infinity in these relations yield the result by the Dominated Convergence Theorem applied to each of the terms $\mathbb{E}\left[Z_n X \mathbf{1}\left[D_i\right]\right]$ for each i in I.

In the general case, write $ZX = (Z^+ - Z^-)X = Z^+X - Z^-X$, and note that both rvs Z^+ and Z^- are $\mathcal{D}$-measurable rvs. The proof then proceeds in the usual manner by applying the previous step to both terms $\mathbb{E}\left[Z^+ X \mathbf{1}\left[D_i\right]\right]$ and $\mathbb{E}\left[Z^- X \mathbf{1}\left[D_i\right]\right]$ for each i in I.

A uniqueness result

Let $Z : \Omega \to \mathbb{R}$ be a $\mathcal{D}$-measurable rv such that $\mathbb{E}\left[|Z|\right] < \infty$. If

$$
\mathbb{E}\left[\mathbf{1}\left[D\right] Z\right] = 0, \quad D \in \mathcal{D}
$$

then $Z = 0$ $\mathbb{P}$-a.s.

Proof: Apply the condition to the $\mathcal{D}$-measurable events $D_+ = [Z > 0]$ and

$D_- = [Z < 0]$, hence $\mathbb{E}[\mathbf{1}[D_+]Z] = 0$ and $\mathbb{E}[\mathbf{1}[D_-]Z] = 0$. Noting that $\mathbf{1}[D_+]Z \geq 0$ and $-\mathbf{1}[D_-]Z \geq 0$, we get

$$\mathbf{1}[D_\pm]Z = 0 \quad \mathbb{P}\text{-}a.s.$$

and the conclusion follows from the obvious decomposition $Z = Z\mathbf{1}[D_-] + Z\mathbf{1}[Z_+]$.

Conditioning with respect to a discrete rv

We briefly discuss how $\mathcal{F}$-partitions are induced by discrete rvs, and how this ultimately relates to conditional expectations with respect to such rvs: Consider a discrete rv $Y : \Omega \to \mathbb{R}^q$. By definition there exists a countable subset $S \subseteq \mathbb{R}^p$ such that $\mathbb{P}[Y \in S] = 1$. For ease of notation, with I countable we shall use the representation $S = \{y_i,\ i \in I\}$ where the elements are *distinct* and each of the events $\{[Y = y_i],\ i \in I\}$ is *non*-empty. So far we can only assert that the event

$$\Omega_Y \equiv \cup_{i \in I}[Y = y_i]$$

has probability one, or equivalently, that the complement Ω_Y^c has zero probability. Nothing precludes the set of values

$$\{Y(\omega),\ \omega \notin \Omega_Y\}$$

to form an *uncountable* set. Only when that set is empty, will the collection $\{[Y = y_i],\ i \in I\}$ be an $\mathcal{F}$-partition of Ω.

To remedy this difficulty, with b an element *not* in S, now define the rv $Y_b : \Omega \to \mathbb{R}^q$ given by

$$Y_b(\omega) \equiv \begin{cases} Y(\omega) & \text{if } \omega \in \Omega_Y \\ \\ b & \text{if } \omega \notin \Omega_Y. \end{cases}$$

The collection $\{[Y = b],\ [Y = y_i],\ i \in I\}$ is now an $\mathcal{F}$-partition of Ω. The following facts are easy consequences from the following observation

$$\mathbb{P}[Y \neq Y_b] \leq \mathbb{P}[\Omega_Y^c] = 0.$$

- The rvs Y and Y_b have the same probability distribution under $\mathbb{P}$.

- If $X : \Omega \to \mathbb{R}^p$ is another rv, the pairs (X, Y) and (X, Y_b) have the same probability distribution under $\mathbb{P}$.

- Consider a Borel mapping $h : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$ such that $\mathbb{E}\left[|h(X,Y)|\right] < \infty$. With
$$S_b = \{y_i, \in I;\ b\} = S \cup \{b\},$$
and $\mathcal{D}_b = \sigma([Y = b],\ [Y = y_i],\ i \in I)$, we note that
$$\begin{aligned}
&\mathbb{E}\left[h(X,Y_b)|\mathcal{D}_b\right] \\
&= \sum_{y \in S_b} \mathbb{E}\left[h(X,Y_b)|Y_b = y\right] \mathbf{1}\left[Y_b = y\right] \\
&= \sum_{y \in S} \mathbb{E}\left[h(X,Y_b)|Y_b = y\right] \mathbf{1}\left[Y_b = y\right] + \mathbb{E}\left[h(X,Y_b)|Y_b = b\right] \mathbf{1}\left[Y_b = b\right] \\
&= \sum_{y \in S} \frac{\mathbb{E}\left[\mathbf{1}\left[Y_b = y\right] h(X,Y_b)\right]}{\mathbb{P}\left[Y_b = y\right]} \mathbf{1}\left[Y_b = y\right] + \mathbb{E}\left[h(X,Y_b)|Y_b = b\right] \mathbf{1}\left[Y_b = b\right] \\
&= \sum_{y \in S} \frac{\mathbb{E}\left[\mathbf{1}\left[Y_b = y\right] h(X,y)\right]}{\mathbb{P}\left[Y_b = y\right]} \mathbf{1}\left[Y_b = y\right] + \mathbb{E}\left[h(X,Y_b)|Y_b = b\right] \mathbf{1}\left[Y_b = b\right] \\
&= \sum_{y \in S} \frac{\mathbb{E}\left[\mathbf{1}\left[Y = y\right] h(X,y)\right]}{\mathbb{P}\left[Y = y\right]} \mathbf{1}\left[Y = y\right] + \mathbb{E}\left[h(X,Y_b)|Y_b = b\right] \mathbf{1}\left[Y_b = b\right]
\end{aligned}$$
It follows that
$$\mathbb{E}\left[h(X,Y_b)|\mathcal{D}_b\right] = \sum_{y \in S} \frac{\mathbb{E}\left[\mathbf{1}\left[Y = y\right] h(X,y)\right]}{\mathbb{P}\left[Y = y\right]} \mathbf{1}\left[Y = y\right] \quad \mathbb{P}\text{-a.s.}$$

- In light of this last calculation, with Borel mapping $h : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$ such that $\mathbb{E}\left[|h(X,Y)|\right] < \infty$, for distinct values $b \neq c$ in $\mathbb{R}^q$, we have
$$\mathbb{E}\left[h(X,Y_b)|\mathcal{D}_b\right] = \mathbb{E}\left[h(X,Y_c)|\mathcal{D}_c\right] \quad \mathbb{P}\text{-}a.s.$$
where we use the notation $\mathcal{D}_b = \sigma([Y = b],\ [Y = y_i],\ i \in I)$ and $\mathcal{D}_c = \sigma([Y = c],\ [Y = y_i],\ i \in I)$.

  In other words, although the two conditional expectation rvs are *not* necessarily identical (as mappings $\Omega \to \mathbb{R}$), they are equal to each other except on a set of zero probability measure (under $\mathbb{P}$). As this notion defines an equivalence relation on rvs,[1] we write $\mathbb{E}\left[h(X,Y)|Y\right]$ (or sometimes $\mathbb{E}\left[h(X,Y)|\sigma(Y)\right]$) to denote *any* representative in the equivalence class.

[1]The rvs $U, V : \Omega \to \mathbb{R}^p$ are $\mathbb{P}$-equivalent if $U = V$ $\mathbb{P}$-a.s., namely $\mathbb{P}\left[U \neq V\right] = 0$.

- One standard representative in that class of $\mathbb{P}$-equivalent rvs is given by

$$\sum_{y\in S} \mathbb{E}\left[h(X,Y)|Y=y\right] \mathbf{1}\left[Y=y\right] \tag{46}$$

Note that all the terms in (46) are well defined in terms of Y! It is *convenient* to use this expression when *representing* the conditional expectation of $h(X,Y)$ given the discrete rv Y.

- Next, observe that

$$\begin{aligned}
& \sum_{y\in S} \mathbb{E}\left[h(X,Y)|Y=y\right] \mathbf{1}\left[Y=y\right] \\
&= \sum_{y\in S} \frac{\mathbb{E}\left[\mathbf{1}\left[Y=y\right] h(X,Y)\right]}{\mathbb{P}\left[Y=y\right]} \mathbf{1}\left[Y=y\right] \\
&= \sum_{y\in S} \frac{\mathbb{E}\left[\mathbf{1}\left[Y=y\right] h(X,y)\right]}{\mathbb{P}\left[Y=y\right]} \mathbf{1}\left[Y=y\right] \\
&= \sum_{y\in S} \mathbb{E}\left[h(X,y)|Y=y\right] \mathbf{1}\left[Y=y\right] \qquad (47)
\end{aligned}$$

This last expression suggests introducing the mapping $\widehat{h} : \mathbb{R}^q \to \mathbb{R}$ given by

$$\widehat{h}(y) = \begin{cases} \mathbb{E}\left[h(X,y)|Y=y\right] & \text{if } y \in S \\ \\ h^\star(y) & \text{if } y \notin S \end{cases} \tag{48}$$

where $h^\star : \mathbb{R}^q \to \mathbb{R}$ is an arbitrary Borel mapping such that $\mathbb{E}\left[|h^\star(Y)|\right] < \infty$. This definition is always well posed, and produces a Borel mapping $\mathbb{R}^q \to \mathbb{R}$.

With this notation we conclude that

$$\begin{aligned}
& \sum_{y\in S} \mathbb{E}\left[h(X,Y)|Y=y\right] \mathbf{1}\left[Y=y\right] \\
&= \sum_{y\in S} \mathbb{E}\left[h(X,y)|Y=y\right] \mathbf{1}\left[Y=y\right] \\
&= \sum_{y\in S} \widehat{h}(y) \mathbf{1}\left[Y=y\right]
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{y \in S} \widehat{h}(Y) \mathbf{1}\left[Y = y\right] \\
&= \widehat{h}(Y) \left(\sum_{y \in S} \mathbf{1}\left[Y = y\right] \right) \\
&= \widehat{h}(Y) \quad \mathbb{P}\text{-a.s.} \qquad (49)
\end{aligned}
$$

since

$$\sum_{y \in S} \mathbf{1}\left[Y = y\right] = \mathbf{1}\left[Y \in S\right] = 1 \quad \mathbb{P}\text{-a.s.}$$

- Symbolically, this last discussion can be summarized as follows:

$$
\begin{aligned}
\mathbb{E}\left[h(X,Y)|Y\right] &= \left(\mathbb{E}\left[h(X,Y)|Y = y\right]\right)_{y=Y} \\
&= \left(\mathbb{E}\left[h(X,y)|Y = y\right]\right)_{y=Y} \quad \mathbb{P}\text{-a.s.} \qquad (50)
\end{aligned}
$$

Consider the rvs $X : \Omega \to \mathbb{R}^p$ and $Y : \Omega \to \mathbb{R}^q$ under the following assumptions: (i) The rvs are *independent*, and (ii) the rv $Y : \Omega \to \mathbb{R}^q$ is a discrete rv.

- With Borel mapping $h : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$ such that $\mathbb{E}\left[|h(X,Y)|\right] < \infty$, define the mapping $\widehat{h} : \mathbb{R}^q \to \mathbb{R}$ given by

$$\widehat{h}(y) = \mathbb{E}\left[h(X,y)\right], \quad y \in \mathbb{R}^q.$$

  This definition is always well posed, and produces a Borel mapping $\mathbb{R}^q \to \mathbb{R}$. It holds that

$$\mathbb{E}\left[h(X,Y)|Y\right] = \widehat{h}(Y) \quad \mathbb{P}\text{-}a.s.$$

- An important special case: With Borel mapping $g : \mathbb{R}^p \to \mathbb{R}$ such that $\mathbb{E}\left[|g(X)|\right] < \infty$, we have

$$\mathbb{E}\left[g(X)|Y\right] = \mathbb{E}\left[g(X)\right] \quad \mathbb{P}\text{-}a.s.$$

  as expected.

The general definition of conditional expectations

Let $\mathcal{D}$ be a sub-σ-field of $\mathcal{F}$. Consider a rv $X : \Omega \to \mathbb{R}$ such that $\mathbb{E}\left[|X|\right] < \infty$.

(i) (Existence) There exists a $\mathcal{D}$-measurable rv $Z : \Omega \to \mathbb{R}$ with $\mathbb{E}[|Z|] < \infty$ such that

$$\mathbb{E}[\mathbf{1}[D](X - Z)] = 0, \quad D \in \mathcal{D} \tag{51}$$

(ii) (Uniqueness) If the $\mathcal{D}$-measurable rvs $Z_1, Z_2 : \Omega \to \mathbb{R}$ with $\mathbb{E}[|Z_1|] < \infty$ and $\mathbb{E}[|Z_2|] < \infty$ both satisfy (51), then $Z_1 = Z_2$ $\mathbb{P}$-a.s.

Existence is a consequence of the celebrated Radon-Nikodym Theorem. Condition (51) is often used in the equivalent form

$$\mathbb{E}[\mathbf{1}[D] X] = \mathbb{E}[\mathbf{1}[D] Z], \quad D \in \mathcal{D} \tag{52}$$

The $\mathcal{D}$-measurable rvs with finite expectation satisfying (51) form an equivalence class (under the $\mathbb{P}$-a.s. equivalence); any one of its representatives will be denoted by $\mathbb{E}[X|\mathcal{D}]$.

As a general rule, in order to establish that two $\mathcal{D}$-measurable rvs are representative of the conditional expectations $\mathbb{E}[X|\mathcal{D}]$, one typically proceeds as follows: Two $\mathcal{D}$-measurable rvs Z_1 and Z_2 are identified with $\mathbb{E}[|Z_1|] < \infty$ and $\mathbb{E}[|Z_2|] < \infty$ such that

$$\mathbb{E}[\mathbf{1}[D] Z_i] = \mathbb{E}[\mathbf{1}[D] X], \quad \begin{array}{c} i = 1, 2 \\ D \in \mathcal{D} \end{array}$$

It follows that

$$\mathbb{E}[\mathbf{1}[D](Z_1 - Z_2)] = 0, \quad D \in \mathcal{D}$$

whence Z_1 and Z_2 are both representatives of $\mathbb{E}[X|\mathcal{D}]$ with $Z_1 = Z_2$ $\mathbb{P}$-a.s.

Properties of conditional expectations

Many of the properties of conditional expectations are given below. They are stated here under *finiteness* assumptions on the expectations of the rvs involved. However, the assumptions can be somewhat weakened when the rvs are non-negative or when the expectations are simply assumed to exist.

It is noteworthy that in spite of an *existential* definition (51), all these properties (some of which have already been established when the σ-field is generated by an $\mathcal{F}$-partition under a *constructive* definition) will be seen to be direct consequences of the constraint (51). As the reader will note, in this more abstract setting the proofs are much simpler than the proofs based on the constructive approach.

When everything is said and done, we can interpret the conditional expectation of X given $\mathcal{D}$ (through any of its representatives) as providing a form of expectation of the rv X, not with respect to the probability measure $\mathbb{P}$, but with respect to the conditional probability distribution of X given $\mathcal{D}$. Therefore it is not surprising that the operation that associates with X, the conditional expectation of X given $\mathcal{D}$ behaves essentially like a regular expectation. In particular, one expects versions of the various convergence theorems (such as the Monotone Convergence Theorem, the Dominated Convergence Theorem and the Bounded Convergence Theorem) to hold for conditional expectations. Here we shall omit a discussion of these topics.

Linearity

Consider rvs $X, Y : \Omega \to \mathbb{R}$ with $\mathbb{E}[|X|] < \infty$ and $\mathbb{E}[|Y|] < \infty$. For any α and β in $\mathbb{R}$, we have

$$\mathbb{E}[\alpha X + \beta Y|\mathcal{D}] = \alpha\mathbb{E}[X|\mathcal{D}] + \beta\mathbb{E}[Y|\mathcal{D}] \quad \mathbb{P}\text{-a.s.}$$

Proof: Pick D in $\mathcal{D}$. By linearity and (52) we have

$$\begin{aligned}
&\mathbb{E}[\mathbf{1}[D] \cdot \mathbb{E}[\alpha X + \beta Y|\mathcal{D}]] \\
&\quad = \mathbb{E}[\mathbf{1}[D] \cdot (\alpha X + \beta Y)] \\
&\quad = \alpha\mathbb{E}[\mathbf{1}[D] X] + \beta\mathbb{E}[\mathbf{1}[D] Y] \\
&\quad = \alpha\mathbb{E}[\mathbf{1}[D] \cdot \mathbb{E}[X|\mathcal{D}]] + \beta \cdot \mathbb{E}[\mathbf{1}[D] \mathbb{E}[Y|\mathcal{D}]] \\
&\quad = \mathbb{E}[\mathbf{1}[D] \cdot (\alpha\mathbb{E}[X|\mathcal{D}] + \beta\mathbb{E}[Y|\mathcal{D}])]
\end{aligned}$$

The rv $\alpha\mathbb{E}[X|\mathcal{D}] + \beta\mathbb{E}[Y|\mathcal{D}]$ being $\mathcal{D}$-measurable, we get the desired result by the $\mathbb{P}$-a.s. uniqueness of conditional expectation.

Monotonicity

Consider rvs $X, Y : \Omega \to \mathbb{R}$ with $\mathbb{E}[|X|] < \infty$ and $\mathbb{E}[|Y|] < \infty$. Whenever $X \leq Y$ we have

$$\mathbb{E}[X|\mathcal{D}] \leq \mathbb{E}[Y|\mathcal{D}] \quad \mathbb{P}\text{-a.s.}$$

Proof: Pick D in $\mathcal{D}$. Using (52) we find

$$\begin{aligned}
\mathbb{E}[\mathbf{1}[D] \mathbb{E}[X|\mathcal{D}]] &= \mathbb{E}[\mathbf{1}[D] X] \\
&\leq \mathbb{E}[\mathbf{1}[D] Y] \\
&= \mathbb{E}[\mathbf{1}[D] \mathbb{E}[Y|\mathcal{D}]] \qquad (53)
\end{aligned}$$

because $X \leq Y$, and therefore

$$\mathbb{E}\left[\mathbf{1}\left[D\right]\left(\mathbb{E}\left[Y|\mathcal{D}\right] - \mathbb{E}\left[X|\mathcal{D}\right]\right)\right] \geq 0.$$

Now pick the event $D_- = [\mathbb{E}\left[Y|\mathcal{D}\right] - \mathbb{E}\left[X|\mathcal{D}\right] < 0]$. The event D_- being $\mathcal{D}$-measurable event, the last inequality applies to yield

$$0 \leq \mathbb{E}\left[\mathbf{1}\left[D_-\right]\left(\mathbb{E}\left[Y|\mathcal{D}\right] - \mathbb{E}\left[X|\mathcal{D}\right]\right)\right] \leq 0,$$

whence

$$\mathbb{E}\left[\mathbf{1}\left[D_-\right]\left(\mathbb{E}\left[Y|\mathcal{D}\right] - \mathbb{E}\left[X|\mathcal{D}\right]\right)\right] = 0.$$

But $\mathbf{1}\left[D_-\right]\left(\mathbb{E}\left[Y|\mathcal{D}\right] - \mathbb{E}\left[X|\mathcal{D}\right]\right) \leq 0$ by construction, and the zero expectation constraint implies $\mathbf{1}\left[D_-\right]\left(\mathbb{E}\left[Y|\mathcal{D}\right] - \mathbb{E}\left[X|\mathcal{D}\right]\right) = 0$ $\mathbb{P}$-a.s. Therefore,

$$\mathbb{E}\left[Y|\mathcal{D}\right] - \mathbb{E}\left[X|\mathcal{D}\right] = \mathbf{1}\left[D_-^c\right]\left(\mathbb{E}\left[Y|\mathcal{D}\right] - \mathbb{E}\left[X|\mathcal{D}\right]\right) \quad \mathbb{P}\text{-a.s.}$$

and the result follows since $\mathbb{E}\left[Y|\mathcal{D}\right] - \mathbb{E}\left[X|\mathcal{D}\right] \geq 0$ on D_-^c.

Iterated conditioning (I) For any rv $X : \Omega \to \mathbb{R}$ with $\mathbb{E}\left[|X|\right] < \infty$, it holds that

(54) $$\mathbb{E}\left[\mathbb{E}\left[X|\mathcal{D}\right]\right] = \mathbb{E}\left[X\right]$$

When read from right to left this equality gives rise to the idea of *preconditioning* as a way to compute certain expectations.

Proof: Use (51) with $D = \Omega$.

Iterated conditioning (II): Let $\mathcal{D}$ and $\mathcal{D}'$ be two sub-σ-fields of $\mathcal{F}$ with $\mathcal{D} \subseteq \mathcal{D}'$. For any rv $X : \Omega \to \mathbb{R}$ with $\mathbb{E}\left[|X|\right] < \infty$, it holds that

(55) $$\mathbb{E}\left[\mathbb{E}\left[X|\mathcal{D}\right]|\mathcal{D}'\right] = \mathbb{E}\left[X|\mathcal{D}\right], \quad \mathbb{P}\text{-a.s.}$$

Proof: Pick D in $\mathcal{D}'$. Using (52) (for the rv $\mathbb{E}\left[X|\mathcal{D}\right]$ and σ-field $\mathcal{D}'$) we get

$$\mathbb{E}\left[\mathbf{1}\left[D\right]\mathbb{E}\left[\mathbb{E}\left[X|\mathcal{D}\right]|\mathcal{D}'\right]\right] = \mathbb{E}\left[\mathbf{1}\left[D\right]\mathbb{E}\left[X|\mathcal{D}\right]\right].$$

The conclusion follows from the fact that the rv $\mathbb{E}[X|\mathcal{D}]$ is $\mathcal{D}$-measurable, hence $\mathcal{D}'$-measurable.

Iterated conditioning (III): Let $\mathcal{D}$ and $\mathcal{D}'$ be two sub-σ-fields of $\mathcal{F}$ with $\mathcal{D} \subseteq \mathcal{D}'$. For any rv $X : \Omega \to \mathbb{R}$ with $\mathbb{E}[|X|] < \infty$, it holds that

$$\mathbb{E}[\mathbb{E}[X|\mathcal{D}']|\mathcal{D}] = \mathbb{E}[X|\mathcal{D}] \quad \mathbb{P}\text{-a.s.} \tag{56}$$

Proof: Pick D in $\mathcal{D}$. We have

$$\begin{aligned} \mathbb{E}[\mathbf{1}[D]\,\mathbb{E}[\mathbb{E}[X|\mathcal{D}']|\mathcal{D}]] &= \mathbb{E}[\mathbf{1}[D]\,\mathbb{E}[X|\mathcal{D}']] \\ &= \mathbb{E}[\mathbf{1}[D]\,X] \\ &= \mathbb{E}[\mathbf{1}[D]\,\mathbb{E}[X|\mathcal{D}]]\,. \end{aligned}$$

Here we have used (52) three times, first for the rv $\mathbb{E}[X|\mathcal{D}']$ and the σ-field $\mathcal{D}$, then for the rv X and the σ-field $\mathcal{D}'$, and finally for the rv X and the σ-field $\mathcal{D}$.

An important property

For any rvs $X, Z : \Omega \to \mathbb{R}$ with $\mathbb{E}[|X|] < \infty$ and $\mathbb{E}[|ZX|] < \infty$, we have

$$\mathbb{E}[ZX|\mathcal{D}] = Z\mathbb{E}[X|\mathcal{D}] \quad \mathbb{P}\text{-a.s.}$$

whenever the rv Z is $\mathcal{D}$-measurable.

Proof: Pick D in $\mathcal{D}$. Using (51) for the rvs ZX we get

$$\mathbb{E}[\mathbf{1}[D]\,\mathbb{E}[ZX|\mathcal{D}]] = \mathbb{E}[\mathbf{1}[D]\,ZX]\,.$$

Let Z be a simple $\mathcal{D}$-measurable rv, say of the form

$$Z = \sum_{k \in K} b_k \mathbf{1}[D_k]$$

where the events $\{D_k, \in K\}$ are all in $\mathcal{D}$. The events $\{D_k \cap D, \in K\}$ are also all in $\mathcal{D}$, hence

$$\mathbb{E}[\mathbf{1}[D]\,ZX] = \sum_{k \in K} b_k \mathbb{E}[\mathbf{1}[D]\,\mathbf{1}[D_k]\,X]$$

$$
\begin{aligned}
&= \sum_{k\in K} b_k \mathbb{E}\left[\mathbf{1}\left[D\cap D_k\right] X\right] \\
&= \sum_{k\in K} b_k \mathbb{E}\left[\mathbf{1}\left[D\cap D_k\right] \mathbb{E}\left[X|\mathcal{D}\right]\right] \\
&= \mathbb{E}\left[\left(\sum_{k\in K} b_k \mathbf{1}\left[D\cap D_k\right]\right) \mathbb{E}\left[X|\mathcal{D}\right]\right] \\
&= \mathbb{E}\left[\mathbf{1}\left[D\right]\left(\sum_{k\in K} b_k \mathbf{1}\left[D_k\right]\right) \mathbb{E}\left[X|\mathcal{D}\right]\right] \\
&= \mathbb{E}\left[\mathbf{1}\left[D\right] Z \mathbb{E}\left[X|\mathcal{D}\right]\right].
\end{aligned}
$$

It follows that

$$
\mathbb{E}\left[\mathbf{1}\left[D\right]\mathbb{E}\left[ZX|\mathcal{D}\right]\right] = \mathbb{E}\left[\mathbf{1}\left[D\right] Z\mathbb{E}\left[X|\mathcal{D}\right]\right]
$$

The rv $Z\mathbb{E}\left[X|\mathcal{D}\right]$ being $\mathcal{D}$-measurable, we have the desired result by the $\mathbb{P}$-a.s. uniqueness of conditional expectation.

Next the proof proceeds in the usual manner: Consider an non-negative $\mathcal{D}$-measurable Z and approximate it by a sequence of simple non-negative $\mathcal{D}$-measurable $\{Z_n,\ n=1,2,\ldots\}$ with $\lim_{n\to\infty} Z_n = Z$ monotonically from below. By the first part of the proof we have

$$
\mathbb{E}\left[\mathbf{1}\left[D\right]\mathbb{E}\left[Z_n X|\mathcal{D}\right]\right] = \mathbb{E}\left[\mathbf{1}\left[D\right] Z_n X\right] = \mathbb{E}\left[\mathbf{1}\left[D\right] Z_n\mathbb{E}\left[X|\mathcal{D}\right]\right], \quad n=1,2,\ldots
$$

Assume first the rv $X:\Omega\to\mathbb{R}$ to be non-negative. Let n go to infinity. Applying the Monotone Convergence Theorem we get

$$
\mathbb{E}\left[\mathbf{1}\left[D\right] ZX\right] = \mathbb{E}\left[\mathbf{1}\left[D\right] Z\mathbb{E}\left[X|\mathcal{D}\right]\right],
$$

whence

$$
\mathbb{E}\left[\mathbf{1}\left[D\right]\mathbb{E}\left[ZX|\mathcal{D}\right]\right] = \mathbb{E}\left[\mathbf{1}\left[D\right] Z\mathbb{E}\left[X|\mathcal{D}\right]\right]
$$

by the definition of conditional expectation. The desired result follows by uniqueness since the rv $Z\mathbb{E}\left[X|\mathcal{D}\right]$ is $\mathcal{D}$-measurable. To remove the non-negativity condition on X we use the decomposition $X = X^+ - X^-$, and apply the earlier result to each of the rvs $\mathbb{E}\left[ZX^+|\mathcal{D}\right]$ and $\mathbb{E}\left[ZX^-|\mathcal{D}\right]$.

Finally, to handle the case of an arbitrary $\mathcal{D}$-measurable Z, we write

$$
ZX = \left(Z_+ - Z_-\right) X = Z_+ X - Z_- X
$$

and apply the earlier result to rvs $\mathbb{E}\left[Z^+X|\mathcal{D}\right]$ and $\mathbb{E}\left[Z^-X|\mathcal{D}\right]$.

Independence and conditional expectation (I): Consider a rv $X : \Omega \to \mathbb{R}$ with $\mathbb{E}[|X|] < \infty$. If the rv X is independent of the σ-field $\mathcal{D}$, then

$$\mathbb{E}[X|\mathcal{D}] = \mathbb{E}[X] \quad \mathbb{P}\text{-a.s.}$$

Here, the independence of the rv X from the σ-field $\mathcal{D}$ means that for each D in $\mathcal{D}$, the rvs X and $\mathbf{1}[D]$ are independent.

Proof: By independence we note that

$$\mathbb{E}[\mathbf{1}[D]X] = \mathbb{P}[D]\mathbb{E}[X] = \mathbb{E}[\mathbf{1}[D]\mathbb{E}[X]], \quad D \in \mathcal{D}.$$

Therefore,

$$\mathbb{E}[\mathbf{1}[D](X - \mathbb{E}[X])] = 0, \quad D \in \mathcal{D}$$

and the conclusion follows since the defining condition (51) holds for the constant rv $\mathbb{E}[X]$ (which is clearly $\mathbb{D}$-measurable).

Independence and conditional expectation (II): Consider a Borel mapping $h : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$, and rvs $X : \Omega \to \mathbb{R}^p$ and $Y : \Omega \to \mathbb{R}^q$ such that $\mathbb{E}[|h(X,Y)|] < \infty$. Define the mapping $\widehat{h} : \mathbb{R}^q \to \mathbb{R}$ given by

$$\widehat{h}(y) = \mathbb{E}[h(X,y)], \quad y \in \mathbb{R}^q.$$

This definition is always well posed, and produces a Borel mapping $\mathbb{R}^q \to \mathbb{R}$.

If the rv X is independent of the σ-field $\mathcal{D}$ and the rv Y is $\mathcal{D}$-measurable, then

$$\mathbb{E}[h(X,Y)|\mathcal{D}] = \widehat{h}(Y) \quad \mathbb{P}\text{-a.s.}$$

Proof: This conclusion is easy to check when

$$h(x,y) = \sum_{\ell=1}^{L} g_\ell(x)h_\ell(y), \quad \begin{array}{l} x \in \mathbb{R}^p, \\ y \in \mathbb{R}^q \end{array} \tag{57}$$

where for each $\ell = 1, \ldots, L$ the mappings $g_\ell : \mathbb{R}^p \to \mathbb{R}$ and $h_\ell : \mathbb{R}^q \to \mathbb{R}$ are Borel with $\mathbb{E}[|g_\ell(X)|] < \infty$ and $\mathbb{E}[|h_\ell(Y)|] < \infty$. The independence of the rvs X

and Y implies that of the rvs $g_\ell(X)$ and $h_\ell(Y)$, whence the foregoing integrability conditions imply that $\mathbb{E}\left[|g_\ell(X)h_\ell(Y)|\right] = \mathbb{E}\left[|g_\ell(X)|\right]\mathbb{E}\left[|h_\ell(Y)|\right] < \infty$.

Indeed, all the equalities being understood in the $\mathbb{P}$-a.s. sense, we have

$$\begin{aligned}
\mathbb{E}\left[h(X,Y)|\mathcal{D}\right] &= \mathbb{E}\left[\sum_{\ell=1}^{L} g_\ell(X)h_\ell(Y)|\mathcal{D}\right] \\
&= \sum_{\ell=1}^{L} \mathbb{E}\left[g_\ell(X)h_\ell(Y)|\mathcal{D}\right] \\
&= \sum_{\ell=1}^{L} h_\ell(Y)\mathbb{E}\left[g_\ell(X)|\mathcal{D}\right] \\
&= \sum_{\ell=1}^{L} h_\ell(Y)\mathbb{E}\left[g_\ell(X)\right] \qquad (58)
\end{aligned}$$

while here we have

$$\begin{aligned}
\widehat{h}(y) &= \mathbb{E}\left[\sum_{\ell=1}^{L} g_\ell(X)h_\ell(y)\right] \\
&= \sum_{\ell=1}^{L} \mathbb{E}\left[g_\ell(X)\right] h_\ell(y), \quad y \in \mathbb{R}^q.
\end{aligned}$$

Indeed,

$$\mathbb{E}\left[h(X,Y)|\mathcal{D}\right] = \widehat{h}(Y) \quad \mathbb{P}\text{-a.s.}$$

The general case is a consequence of the following fact: For every Borel mapping $h : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$ there exists a sequence $\{h_n, \ n = 1, 2, \ldots\}$ of Borel mappings $\mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$ of the form (57) such that

$$\lim_{n\to\infty} h_n(x,y) = h(x,y), \qquad \begin{array}{c} x \in \mathbb{R}^p, \\ y \in \mathbb{R}^q \end{array}$$

Details are omitted.

More generally, *without independence* we have

$$\mathbb{E}\left[h(X,Y)|\mathcal{D}\right] = \mathbb{E}\left[\sum_{\ell=1}^{L} g_\ell(X)h_\ell(Y)|\mathcal{D}\right]$$

$$
\begin{aligned}
&= \sum_{\ell=1}^{L} \mathbb{E}\left[g_\ell(X)h_\ell(Y)|\mathcal{D}\right] \\
&= \sum_{\ell=1}^{L} h_\ell(Y)\mathbb{E}\left[g_\ell(X)|\mathcal{D}\right] \\
&= \left(\sum_{\ell=1}^{L} h_\ell(y)\mathbb{E}\left[g_\ell(X)|\mathcal{D}\right]\right)_{y=Y} \\
&= \left(\mathbb{E}\left[\sum_{\ell=1}^{L} h_\ell(y)g_\ell(X)|\mathcal{D}\right]\right)_{y=Y} \\
&= \left(\mathbb{E}\left[h(X,y)|\mathcal{D}\right]\right)_{y=Y} \qquad (59)
\end{aligned}
$$

The absolutely continuous case

Consider rvs $X : \Omega \to \mathbb{R}^p$ and $Y : \Omega \to \mathbb{R}^q$. If the rv Y is absolutely continuous, then

$$\mathbb{P}\left[Y = y\right] = 0, \quad y \in \mathbb{R}^q$$

sine

$$\int_{\{y\}} f_Y(\eta)d\eta = 0.$$

As a result, for each y in $\mathbb{R}^q$ we *cannot* define the conditional probabilities

$$\mathbb{P}\left[X \in B|Y = y\right] = \frac{\mathbb{P}\left[X \in B, Y = y\right]}{\mathbb{P}\left[Y = y\right]}, \quad B \in \mathcal{B}(\mathbb{R}^p)$$

A natural approach

With y in $\mathbb{R}^q$. The ball centered at y with radius $\varepsilon > 0$ is denoted by $B_\varepsilon(y)$, i.e.,

$$B_\varepsilon(y) \equiv \left\{\eta \in \mathbb{R}^q : \|\eta - y\| \leq \varepsilon\right\}.$$

Pick y in $\mathbb{R}^q$ such that $f_Y(y) > 0$ and assume there exists $\varepsilon_0 > 0$ such that

$$\mathbb{P}\left[Y \in B_\varepsilon(y)\right] > 0, \quad 0 < \varepsilon \leq \varepsilon_0.$$

The basic idea is as follows: Pick B in $\mathcal{B}(\mathbb{R}^p)$. Whatever definition is given to the conditional probability

$$\mathbb{P}\left[X \in B|Y = y\right],$$

it should be compatible with the limiting value

$$\lim_{\varepsilon \downarrow 0} \mathbb{P}\left[X \in B | Y \in B_\varepsilon(y)\right]$$

if it exists.

With this in mind we note that

$$\begin{aligned} \mathbb{P}\left[X \in B | Y \in B_\varepsilon(y)\right] &= \frac{\mathbb{P}\left[[X \in B] \cap B_\varepsilon(y)\right]}{\mathbb{P}\left[B_\varepsilon(y)\right]} \\ &= \frac{\int_{B \times B_\varepsilon(y)} f_{XY}(\xi, \eta) d\xi d\eta}{\int_{B_\varepsilon(y)} f_Y(\eta) d\eta} \\ &= \frac{\int_B \left(\int_{B_\varepsilon(y)} f_{XY}(\xi, \eta) d\eta \right) d\xi}{\int_{B_\varepsilon(y)} f_Y(\eta) d\eta} \\ &= \int_B \left(\frac{\int_{B_\varepsilon(y)} f_{XY}(\xi, \eta) d\eta}{\int_{B_\varepsilon(y)} f_Y(\eta) d\eta} \right) d\xi \end{aligned} \tag{60}$$

Note that

$$\lim_{\varepsilon \downarrow 0} \int_{B_\varepsilon(y)} f_{XY}(\xi, \eta) d\eta = 0, \quad \xi \in \mathbb{R}^p$$

and

$$\lim_{\varepsilon \downarrow 0} \int_{B_\varepsilon(y)} f_Y(\eta) d\eta = 0.$$

However, in many cases of interest in applications, we find that these limits have the same rate of convergence so that the limit

$$\lim_{\varepsilon \downarrow 0} \frac{\int_{B_\varepsilon(y)} f_{XY}(\xi, \eta) d\eta}{\int_{B_\varepsilon(y)} f_Y(\eta) d\eta}$$

in fact exists. This is analogous to the situation handled by L'Hospital's rule when the indeterminate form $\frac{0}{0}$ arises. Indeed note that under broad conditions it holds

$$\lim_{\varepsilon \downarrow 0} \frac{\int_{B_\varepsilon(y)} f_{XY}(\xi, \eta) d\eta}{\lambda(B_\varepsilon(y))} = f_{XY}(\xi, y), \quad \xi \in \mathbb{R}^p$$

and

$$\lim_{\varepsilon \downarrow 0} \frac{\int_{B_\varepsilon(y)} f_Y(\eta) d\eta}{\lambda(B_\varepsilon(y))} = f_Y(y).$$

where $\lambda(B_\varepsilon(y))$ denotes the Lebesgue measure of the ball $B_\varepsilon(y)$. It now follows that

$$\begin{aligned} \lim_{\varepsilon \downarrow 0} \frac{\int_{B_\varepsilon(y)} f_{XY}(\xi,\eta)d\eta}{\int_{B_\varepsilon(y)} f_Y(\eta)d\eta} &= \lim_{\varepsilon \downarrow 0} \frac{\frac{\int_{B_\varepsilon(y)} f_{XY}(\xi,\eta)d\eta}{\lambda(B_\varepsilon(y))}}{\frac{\int_{B_\varepsilon(y)} f_Y(\eta)d\eta}{\lambda(B_\varepsilon(y))}} \\ &= \frac{f_{XY}(\xi,y)}{f_Y(y)}, \quad \xi \in \mathbb{R}^p \end{aligned} \tag{61}$$

This suggests

$$\begin{aligned} \lim_{\varepsilon \downarrow 0} \mathbb{P}\left[X \in B | Y \in B_\varepsilon(y)\right] &= \lim_{\varepsilon \downarrow 0} \int_B \left(\frac{\int_{B_\varepsilon(y)} f_{XY}(\xi,\eta)d\eta}{\int_{B_\varepsilon(y)} f_Y(\eta)d\eta} \right) d\xi \\ &= \int_B \lim_{\varepsilon \downarrow 0} \left(\frac{\int_{B_\varepsilon(y)} f_{XY}(\xi,\eta)d\eta}{\int_{B_\varepsilon(y)} f_Y(\eta)d\eta} \right) d\xi \\ &= \int_B \frac{f_{XY}(\xi,y)}{f_Y(y)} d\xi \end{aligned} \tag{62}$$

assuming that the interchange of limit and integration is permissible.

With y in $\mathbb{R}^q$, define the mapping $f_{X|Y}(\cdot|y) : \mathbb{R}^p \to \mathbb{R}_+$ by

$$f_{X|Y}(x|y) \equiv \begin{cases} \frac{f_{XY}(x,y)}{f_Y(y)} & \text{if } f_Y(y) > 0 \\ \\ g(x) & \text{if } f_Y(y) = 0 \end{cases}$$

where the Borel mapping $g : \mathbb{R}^p \to \mathbb{R}_+$ is a probability density density, hence satisfies

$$\int_{\mathbb{R}^p} g(x)dx = 1.$$

Consider a Borel mapping $u : \mathbb{R}^p \to \mathbb{R}$ such that that $\mathbb{E}\left[|u(X)|\right] < \infty$, and pick a Borel set C in $\mathcal{B}(\mathbb{R}^q)$. Note that

$$\mathbb{P}\left[[Y \in C] \cap [f_Y(Y) = 0]\right] = 0$$

since

$$\mathbb{P}\left[f_Y(Y) = 0\right] = \int \eta \in \mathbb{R}^q) : \ f_Y(\eta) = 0 f_Y(\eta)d\eta = 0.$$

With
$$C_Y^+ \equiv \{\eta \in \mathbb{R}^q : \ f_Y(\eta) > 0\},$$
this becomes
$$\mathbb{P}\left[Y \notin C_Y^+\right] = \mathbb{P}\left[f_Y(Y) = 0\right] = 0.$$

We find
$$\begin{aligned}\mathbb{E}\left[\mathbf{1}\left[Y \in C\right] u(X)\right] &= \mathbb{E}\left[\mathbf{1}\left[Y \in C, f_Y(Y) > 0\right] u(X)\right] \\ &= \int_{\mathbb{R}^p \times (C \cap C_Y^+)} u(\xi) f_{XY}(\xi, \eta) d\xi d\eta \\ &= \int_{C \cap C_Y^+} \left(\int_{\mathbb{R}^p} u(\xi) f_{XY}(\xi, \eta) d\xi \right) d\eta\end{aligned}$$
by Fubini's Theorem.

If $f_Y(\eta) > 0$, then
$$\begin{aligned}\int_{\mathbb{R}^p} u(\xi) f_{XY}(\xi, \eta) d\xi &= \int_{\mathbb{R}^p} u(\xi) f_{X|Y}(\xi|\eta) f_Y(\eta) d\xi \\ &= \left(\int_{\mathbb{R}^p} u(\xi) f_{X|Y}(\xi|\eta) d\xi \right) f_Y(\eta) \\ &= \widehat{u}(\eta) f_Y(\eta) \qquad (63)\end{aligned}$$
as we define $\widehat{u} : \mathbb{R}^q \to \mathbb{R}$ given by
$$\widehat{u}(y) = \int_{\mathbb{R}^p} u(\xi) f_{X|Y}(\xi|y) d\xi, \quad y \in \mathbb{R}^p.$$
It can be shown that the mapping $\widehat{u} \to \mathbb{R}$ is well defined and Borel.

It follows that
$$\begin{aligned}\mathbb{E}\left[\mathbf{1}\left[Y \in C\right] u(X)\right] &= \int_{C \cap C_Y^+} \widehat{u}(\eta) f_Y(\eta) d\eta \\ &= \int_C \widehat{u}(\eta) f_Y(\eta) d\eta \\ &= \mathbb{E}\left[\mathbf{1}\left[Y \in C\right] \widehat{u}(Y)\right]. \qquad (64)\end{aligned}$$
Recalling that $\sigma(Y) = \{Y \in C, \ C \in \mathcal{B}(\mathbb{R}^p)\}$, we conclude that
$$\mathbb{E}\left[u(X)|\sigma(Y)\right] = \widehat{u}(Y), \quad \mathbb{P}\text{-a.s.}$$