STAT 810 Probability Theory I

Chapter 5: Integration and Expectation

Dr. Dewei Wang
Associate Professor
Department of Statistics
University of South Carolina
deweiwang@stat.sc.edu

5.1.1 Simple Functions

On $(\Omega, \mathcal{B}, P)$, say $X : \Omega \mapsto \mathbb{R}$ is **simple** if it has a finite range. Such a function can always be written in the form

$$X(\omega) = \sum_{i=1}^{k} a_i I_{A_i}(\omega),$$

where $a_i \in \mathbb{R}$ and $A_i \in \mathcal{B}$, $i = 1, \ldots, k$ are disjoint and $\cup_{i=1}^{k} A_i = \Omega$. Then

$$\sigma(X) = \sigma(A_i : i = 1, \ldots, k) = \{\cup_{i \in I} A_i : I \subset \{1, \ldots, k\}\}.$$

Let $\mathcal{E}$ be the set of all simple functions on Ω. We have

1. $\mathcal{E}$ is a vector space; i.e., (i) if $X \in \mathcal{E}$, then $\alpha X \in \mathcal{E}$ for $\alpha \in \mathbb{R}$; (ii) if $X, Y \in \mathcal{E}$, then $X + Y \in \mathcal{E}$.
2. If $X, Y = \sum_j b_j I_{B_j} \in \mathcal{E}$, then $XY = \sum_{i,j} a_i b_j I_{A_i \cap B_j} \in \mathcal{E}$.
3. If $X, Y \in \mathcal{E}$, then $X \vee Y = \sum_{i,j} (a_i \vee b_j) I_{A_i \cup B_j} \in \mathcal{E}$ and $X \wedge Y = \sum_{i,j} (a_i \wedge b_j) I_{A_i \cap B_j}$

Any measurable function can be approximated by a simple function.

Theorem 5.1.1 (Measurability Theorem)

Suppose $X(\omega) \geq 0$ for all ω. Then $X \in \mathcal{B}/\mathcal{B}(\mathbb{R})$ iff there exist simple functions $X_n \in \mathcal{E}$ and

$$0 \leq X_n \uparrow X.$$

Proof: Because taking limits preserves measurability, every simple function is measurable, thus $X \in \mathcal{B}/\mathcal{B}(\mathbb{R})$. Conversely, define

$$X_n = \sum_{k=1}^{n2^n} \left(\frac{k-1}{2^n} \right) I_{[\frac{k-1}{2^n} \leq X \leq \frac{k}{2^n}]} + nI_{[X \geq n]}.$$

Because X is measurable, $X_n \in \mathcal{E}$. Also $X_n \leq X_{n+1}$ and if $X(\omega) < \infty$, then for large n, $|X(\omega) - X_n(\omega)| \leq 2^{-n} \to 0$ (Note that if $\sup_\omega |X(\omega)| < \infty$, then $\sup_\omega |X(\omega) - X_n(\omega)| \to 0$). If $X(\omega) = \infty$, then $X_n(\omega) = n \to \infty$.

Suppose $X : (\Omega, \mathcal{B}) \mapsto (\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}))$ where $\bar{\mathbb{R}} = [-\infty, \infty]$ (in stochastic modeling, we often deal with waiting time for an event to happen, If the event never occurs, then the return time is infinite). Define

$$E(X) = \int_\Omega X dP \text{ or } \int_\Omega X(\omega) P(d\omega),$$

as the *Lebesgue-Stieltjes* integral of X with respect to P.

Suppose X is a simple random variable of the form

$$X = \sum_{i=1}^{n} a_i I_{A_i}$$

where $|a_i| < \infty$, $\{A_i\}$ are mutually exclusive, and $\cup_i A_i = \Omega$. Then

$$E(X) = \int X dP = \sum_{i=1}^{k} a_i P(A_i).$$

5.2.1 Expectation of Simple Functions

Below are some simple properties (HW 5-1: prove these properties)

1. $E(1) = 1$, $E(I_A) = P(A)$.
2. If $X \geq 0$ and $X \in \mathcal{E}$, then $E(X) \geq 0$.
3. Linearity: if $X, Y \in \mathcal{E}$, then $E(\alpha X + \beta Y) = \alpha E(X) + \beta E(Y)$ for $\alpha, \beta \in \mathbb{R}$.
4. Monotonicity: if $X \leq Y \in \mathcal{E}$, then $E(X) \leq E(Y)$.
5. If $X_n, X \in \mathcal{E}$, either $X_n \uparrow X$ or $X_n \downarrow X$, then $E(X_n) \uparrow E(X)$ or $E(X_n) \downarrow E(X)$.

Let $\mathcal{E}_+$ collect all the non-negative valued simple functions, and define

$$\bar{\mathcal{E}}_+ = \{X \geq 0 : X : (\Omega, \mathcal{B}) \mapsto (\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}))\}$$

to be non-negative, measurable functions with domain Ω. If $X \in \bar{\mathcal{E}}_+$ and $P[X = \infty] > 0$, define $E(X) = \infty$.
Otherwise by Theorem 5.1.1, we may find $X_n \in \mathcal{E}_+$, such that

$$0 \leq X_n \uparrow X.$$

We call $\{X_n\}$ the **approximating sequence** to X. The sequence $\{E(X_n)\}$ is non-decreasing by monotonicity of expectations applied to $\mathcal{E}_+$. Since limits of monotone sequences always exist, we conclude that $\lim_{n\to\infty} E(X_n)$ exists and define

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

This extends expectation from $\mathcal{E}$ to $\bar{\mathcal{E}}_+$.

Proposition 5.2.1 (Well definition)

If $X_n, Y_m \in \mathcal{E}_+$ and $X_n \uparrow X$, $Y_m \uparrow X$, then

$$\lim_{n\to\infty} E(X_n) = \lim_{m\to\infty} E(Y_m).$$

Proof: We prove that if $\lim_{n\to\infty} \uparrow X_n \leq \lim_{m\to\infty} \uparrow Y_m$, then $\lim_{n\to\infty} \uparrow E(X_n) \leq \lim_{m\to\infty} \uparrow E(Y_m)$.

Note that since $\lim_{m\to\infty} Y_m \geq \lim_{n\to\infty} X_n \geq X$, $\mathcal{E}_+ \ni X_n \wedge Y_m \uparrow X_n \in \mathcal{E}_+$ as $m \to \infty$. By monotonicity of expectation on $\mathcal{E}_+$, $E(X_n) = \lim_{m\to\infty} \uparrow E(X_n \wedge Y_m) \leq \lim_{m\to\infty} E(Y_m)$ holds for all n, which completes the proof of $\lim_{n\to\infty} \uparrow E(X_n) \leq \lim_{m\to\infty} \uparrow E(Y_m)$.

5.2.3 Basic Properties of Expectation

For expectation on $\bar{\mathcal{E}}_+$:

1. $0 \le E(X) \le \infty$ and if $X \le Y \in \bar{\mathcal{E}}_+$, then $E(X) \le E(Y)$.
   Proof: Find approximating sequences in $\mathcal{E}_+$: $X_n \uparrow X$, $Y_m \uparrow Y$. Then $X = \lim_{n\to\infty} \uparrow X_n \le \lim_{m\to\infty} \uparrow Y_m = Y$. We have proved that
   $E(X) = \lim_{n\to\infty} \uparrow E(X_n) \le \lim_{m\to\infty} \uparrow E(Y_m) = E(Y)$.

2. E is linear: For $\alpha > 0$ and $\beta > 0$, $E(\alpha X + \beta Y) = \alpha E(X) + \beta E(Y)$.
   Proof: $\mathcal{E}_+ \ni X_n + Y_m \uparrow X + Y$.
   $E(X+Y) = \lim_{n\to\infty} E(X_n + Y_n) = \lim_{n\to\infty}(E(X_n) + E(Y_n)) = \lim_{n\to\infty} E(X_n) + \lim_{n\to\infty} E(Y_n) = E(X) + E(Y)$.
   For $\alpha > 0$, $\alpha X_n \uparrow \alpha X$, thus
   $E(\alpha X) = \lim_{n\to\infty} E(\alpha X_n) = \lim_{n\to\infty} \alpha E(X_n) = \alpha E(X)$.
3. **Monotone Convergence Theorem (MCT)** If $0 \le X_n \uparrow X$, then $E(X_n) \uparrow E(X)$ (interchange of expectations and limits).

Proof of MCT: For each $X_n \in \bar{\mathcal{E}}_+$, find an approximating sequence $Y_m^{(n)} \in \mathcal{E}_+$ such that $Y_m^{(n)} \uparrow X_n$ as $m \to \infty$. Define $Z_m = \vee_{n \le m} Y_m^{(n)}$. Note that $\{Z_m\}$ is non-decreasing. Next observe that for $n \le m$, $Y_m^{(n)} \le \bigvee_{j \le m} Y_m^{(j)} = Z_m \le \bigvee X_j = X_m$. Thus, for all n

$$X_n = \lim_{m \to \infty} Y_m^{(n)} \le \lim_{m \to \infty} Z_m \le \lim_{m \to \infty} X_m = X.$$

Therefore $X = \lim_{n \to \infty} X_n = \lim_{m \to \infty} Z_m$. Thus, we have $\{Z_m\}$ as an approximating sequence in $\mathcal{E}_+$ of X. Thus $\lim_{m \to \infty} \uparrow E(Z_m) = E(X)$. Furthermore, we have $E(X_n) = \lim_{m \to \infty} \uparrow E(Y_m^{(n)}) \le \lim_{m \to \infty} \uparrow E(Z_m) = E(X) \le \lim_{m \to \infty} \uparrow E(X_m)$ for each n. Taking limit on n, we have $\lim_{n \to \infty} E(X_n) \le E(X) \le \lim_{m \to \infty} E(X_m)$.

5.2.3 Basic Properties of Expectation on $\bar{\mathcal{E}}_+$

We now further extend the definition of $E(X)$ beyond $\bar{\mathcal{E}}_+$. For a random variable X, define

$$X^+ = X \vee 0, \quad X^- = (-X) \vee 0.$$

We have $X^{\pm} \geq 0$, $X = X^+ - X^-$, $|X| = X^+ + X^-$ and

$$X \in \mathcal{B}/\mathcal{B}(\mathbb{R}) \text{ iff both } X^{\pm} \in \mathcal{B}/\mathcal{B}(\mathbb{R}).$$

We call X **quasi-integrable** if at least one of $E(X^+)$ and $E(X^-)$ is finite. In this case, define

$$E(X) = E(X^+) - E(X^-).$$

If both $E(X^+)$ and $E(X^-)$ are finite, call X **integrable**. This is the case of $E|X| < \infty$. The set of integrable random variables is denoted by L_1 or $L_1(P) = \{X : E|X| < \infty\}$. If both $E(X^+)$ and $E(X^-)$ are infinite, then $E(X)$ **does not exist**.

5.2.3 Basic Properties of Expectation (Summary)

$$X \in \mathcal{E},\ E(X) = \sum_i a_i P(A_i)$$

$$\downarrow$$

$$X \in \bar{\mathcal{E}}_+\text{: By } \mathcal{E}_+ \ni X_n \uparrow X,\ E(X) = \lim_{n\to\infty} E(X_n)$$

$$\downarrow$$

$$\text{General } X\text{: } E(X) = E(X^+) - E(X^-).$$

Example 5.2.1 (Heavy Tails)

Let X's density be $f(x)$, then X's expectation, if exists, is $E(X) = \int xf(x)dx$.

If $f(x) = x^{-1}I(x > 1)$. Then $E(X)$ exists and $E(X) = \infty$.

If $f(x) = 0.5|x|^{-2}I(|x| > 1)$, then $E(X^+) = E(X^-) = \infty$ and $E(X)$ does not exist.

The same conclusion would hold if f were the Cauchy density; i.e., $f(x) = 1/\{\pi(1 + x^2)\}$ for $x \in \mathbb{R}$.

For expectation of any random variable:

1. If X is integrable, then $P[X = \pm\infty] = 0$.
   Proof: if $P[X = \infty] > 0$, then $E(X^+) = \infty$ and X is not integrable.

2. If $E(X)$ exists, $E(cX) = cE(X)$. If either $E(X^+) < \infty$ and $E(Y^+) < \infty$ or $E(X^-) < \infty$ and $E(Y^-) < \infty$, then $X + Y$ is quasi-integrable and $E(X + Y) = E(X) + E(Y)$.
   Proof: We only prove the last equation. It is based on
   $(X + Y)^+ - (X + Y)^- = X + Y = X^+ - X^- + Y^+ - Y^-$
   which implies $(X + Y)^+ + X^- + Y^- = (X + Y)^- + X^+ + Y^+$.
   Taking expectation, we have
   $E(X+Y)^+ + E(X^-) + E(Y^-) = E(X+Y)^- + E(X^+) + E(Y^+)$.
   Rearranging completes the proof.

For expectation of any random variable:

3. If $X \geq 0$, then $E(X) \geq 0$. If $X, Y \in L_1$ and $X \leq Y$, then $E(X) \leq E(Y)$.
   Proof: $Y - X \geq 0 \implies E(Y - X) \geq 0$.
   $|Y - X| \leq |Y| + |X|$, thus $Y - X \in L_1$. Then by 2, we have $E(Y - X) = E(Y) - E(X)$.
4. Suppose $\{X_n\}$ is a sequence of random variables such that $X_n \in L_1$ for some n, if either $X_n \uparrow X$ or $X_n \downarrow X$, then $E(X_n) \uparrow E(X)$ or $E(X_n) \downarrow E(X)$.
   Proof: Focus on $X_n \uparrow X$. Then $X_n^- \downarrow X^-$ so $E(X^-) < \infty$.
   Then $0 \leq X_n^+ = X_n + X_n^- \leq X_n + X_1^- \uparrow X + X_1^-$. By MCT, $0 \leq E(X_n + X_1^-) \uparrow E(X + X_1^-)$. Because $X_n \in L_1$, we have $E(X_n + X_1^-) = E(X_n) + E(X_1^{)}$ Further because $E(X^-) < \infty$ and $E(X_1^-) < \infty$, by 2, we have
   $E(X + X_1^-) = E(X) + E(X_1^-)$. Thus
   $\lim_{n\to\infty}\{E(X_n) + E(X_1^-)\} = E(X) + E(X_1^-)$; i.e.,
   $\lim_{n\to\infty} E(X_n) = E(X)$. (HW 5-2: Prove it for $X_n \downarrow X$)

For expectation of any random variable:

5. **Modulus Inequality.** If $X \in L_1$, $|E(X)| \leq E(|X|)$.
   Proof:
   $|E(X)| = |E(X^+) - E(X^-)| \leq E(X^+) + E(X^-) = E(|X|)$.
6. **Variance and Covariance.** Suppose $X^2 \in L_1$ (or $X \in L_2$), then $\mathrm{Var}(X) = E(X - E(X))^2 = E(X^2) - (E(X))^2$. For $X, Y \in L_2$,
   $\mathrm{Cov}(X, Y) = E((X - E(X))(Y - E(Y))) = E(XY) - E(X)E(Y)$.
   $\mathrm{Cov}(X, Y) = 0$ defines that X and Y are **uncorrelated** If $X \perp Y$ and $X, Y \in L_2$, then $\mathrm{Cov}(X, Y) = 0$.
   If $X_1, \ldots, X_n \in L_2$ are uncorrelated, then
   $\mathrm{Var}(\sum_i X_i) = \sum_i \mathrm{Var}(X_i)$.
   Also if $Y_1, \ldots, Y_m \in L_2$, $a_i, b_i \in \mathbb{R}$, we have
   $\mathrm{Cov}(\sum_i a_i X_i, \sum_j b_j Y_j) = \sum_i \sum_j a_i b_j \mathrm{Cov}(X_i, Y_j)$.

For expectation of any random variable:

7. **Markov inequality.** Suppose $X \in L_1$. For any $\lambda > 0$, $P[|X| \geq \lambda] \leq \lambda^{-1} E(|X|)$.
   Proof: observe $1 \times I_{[|X| \geq \lambda]} \leq \frac{|X|}{\lambda}$ then take expectations.
8. **Chebychev inequality.** Suppose $X \in L_1$. For any $\lambda > 0$, $P[|X - E(X)| \geq \lambda] \leq \mathrm{Var}(X)/\lambda^2$.
   Proof: follows from the Markov's inequality.
9. **WLLN.** Let $\{X_n\}$ be iid with finite mean μ and variance σ^2. Then for any $\epsilon > 0$, $\lim_{n\to\infty} P[|\bar{X}_n - \mu| > \epsilon] = 0$, where $\bar{X}_n = n^{-1} \sum_i X_i$.
   Proof: using Chebyshev yields
   $P[|\bar{X}_n - \mu| > \epsilon] \leq \epsilon^{-2} \mathrm{Var}(\bar{X}_n) = \frac{\mathrm{Var}(X_i))}{n\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \to 0$.

Theorem 5.3.1 (MCT)

If $0 \leq X_n \uparrow X$, then $0 \leq E(X_n) \uparrow E(X)$.

Corollary 5.3.1 (Series versions of MCT)

if $\xi_j \geq 0$ are non-negative random variables for $n \geq 1$, then

$$E(\sum_{j=1}^{\infty} \xi_j) = \sum_{j=1}^{\infty} E(\xi_j).$$

Proof: $X_n = \sum_{j=1}^{n} \xi_j$ and $X = \sum_{j=1}^{\infty} \xi_j$. Apply MCT.

Theorem 5.3.2 (Fatou Lemma)

If $0 \le X_n$, then

$$E(\liminf_{n\to\infty} X_n) \le \liminf_{n\to\infty} E(X_n).$$

More generally, if there exists $Z \in L_1$ and $X_n \ge Z$, then

$$E(\liminf_{n\to\infty} X_n) \le \liminf_{n\to\infty} E(X_n).$$

Proof: $\liminf_{n\to\infty} X_n = \sup_{n\ge1} \inf_{k\ge n} X_k$. Thus if $X_n \ge 0$, then

$$E(\liminf_{n\to\infty} X_n) = E(\lim_{n\to\infty} \uparrow (\inf_{k\ge n} X_k)) = \lim_{n\to\infty} \uparrow E(\inf_{k\ge n} X_k) \le \liminf E(X_n).$$

For $X_n \ge Z$, we consider $X_n - Z \ge 0$.

Corollary 5.3.2 (More Fatou)

If $X_n \le Z$ where $Z \in L_1$, then

$$E(\limsup_{n\to\infty} X_n) \ge \limsup_{n\to\infty} E(X_n).$$

Proof: We have $-X_n \ge -Z \in L_1$. Then

$$E(\liminf(-X_n)) \le \liminf_{n\to\infty} E(-X_n),$$

so that

$$E(-\liminf_{n\to\infty}(-X_n)) \ge -\liminf_{n\to\infty}(-E(X_n)).$$

It completes the proof because $-\liminf_{n\to\infty}(-X_n) = \limsup_{n\to\infty} X_n$ and $-\liminf_{n\to\infty}(-E(X_n)) = \limsup_{n\to\infty} E(X_n)$.

Canonical Example

Nasty things could happened when interchanging limits and integrals: Let $(\Omega, \mathcal{B}, P) = ([0,1], \mathcal{B}([0,1]), \lambda)$, where λ is Lebesgue measure. Define

$$X_n = n^2 I_{(0,1/n)}$$

For any $\omega \in [0,1]$, $I_{(0,1/n)}(\omega) \to 0$, so $X_n \to 0$. However, $EX_n = n^2(1/n) = n \to \infty$. So

$$E(\liminf_{n\to\infty} X_n) = 0 \leq \liminf_{n\to\infty} E(X_n) = \infty,$$

and

$$E(\limsup_{n\to\infty} X_n) = 0 \not\geq \limsup_{n\to\infty} E(X_n) = \infty.$$

Corollary 5.3.2 failed because there is no $Z \in L_1$ such tat $X_n \leq Z$. (Dominating condition is important!)

Theorem 5.3.3 (Dominated Convergence Theorem (DCT)) If $X_n \to X$ and there exists a dominating random variable $Z \in L_1$ such that

$$|X_n| \leq Z,$$

then

$$E(X_n) \to E(X).$$

Proof: We have $-Z \leq X \leq Z$. Thus, we can apply Theorem 5.3.2 and Corollary 5.3.2.

$$\begin{aligned} E(X) = E(\liminf_{n\to\infty} X_n) &\leq \liminf_{n\to\infty} E(X_n) \\ &\leq \limsup_{n\to\infty} E(X_n) \leq E(\limsup_{n\to\infty} X_n) = E(X). \end{aligned}$$

5.4 Indefinite Integrals

Definition 5.4.1

If $X \in L_1$, we define

$$\int_A XdP = E(X \cdot I_A)$$

and call $\int_A XdP$ the integral of X over A. Call X the **integrand**.

Suppose $X \geq 0$, we have (HW 5-3: prove these)

1. $0 \leq \int_A XdP \leq E(X)$.
2. $\int_A XdP = 0$ iff $P(A \cap [X > 0]) = 0$.
3. If $\{A_n : n \geq 1\}$ is a sequence of disjoint events $\int_{\cup_n A_n} Xdp = \sum_{n=1}^{\infty} \int_{A_n} Xdp$.
4. If $A_1 \subset A_2$, then $\int_{A_1} Xdp \leq \int_{A_2} Xdp$.
5. Suppose $X \in L_1$ and $\{A_n\}$ is a monotone sequence of events. If $A_n \uparrow A$, then $\int_{A_n} Xdp \uparrow \int_A XdP$; while if If $A_n \downarrow A$, then $\int_{A_n} Xdp \downarrow \int_A XdP$.

Suppose $T : (\Omega, \mathcal{B}) \mapsto (\Omega', \mathcal{B}')$ is a measurable map. P is a probability measure on $\mathcal{B}$. The induced probability measure on $\mathcal{B}'$ is

$$P' = P \circ T^{-1}; \text{ i.e., } P'(A') = P(T^{-1}(A')),\ A' \in \mathcal{B}'.$$

Example

$\Omega = \{(a, b) : a, b = 1, \ldots, 6\}$: tossing two dices.
$T(a, b) = \max(a, b) : \Omega \mapsto \Omega'$
$\Omega' = \{m : m = 1, \ldots, 6\}$: the max of the two dices.
Let $A' = \{m = 2\}$, then $P'(\{m = 2\}) = P(\{(1, 2), (2, 1), (2, 2)\})$.

Suppose $X' : (\Omega', \mathcal{B}') \mapsto (\mathbb{R} : \mathcal{B}(\mathbb{R}))$ is a random variable, and the induced probability by X' is $P_{X'}$, where $P_{X'}(B) = P'(X'^{-1}(B)), B \in \mathcal{B}(\mathbb{R})$.

$$(\Omega, \mathcal{B}, P) \overset{T}{\to} (\Omega', \mathcal{B}', P') \overset{X'}{\to} (\mathbb{R}, \mathcal{B}(\mathbb{R}), F')$$

where $F'(A) = P' \circ X'^{-1}(A) = P \circ T^{-1} \circ X'^{-1}(A)$ for $A \in \mathcal{B}(\mathbb{R})$.

Theorem 5.5.1 (Transformation Theorem)

Suppose $X' : (\Omega', \mathcal{B}') \mapsto (\mathbb{R} : \mathcal{B}(\mathbb{R}))$ is a random variable. We know $X' \circ T : \Omega \mapsto \mathbb{R}$ is also a random variable by composition.

(i) If $X' \geq 0$, then

$$\int_{\Omega'} X'(\omega')P'(d\omega') = \int_{\Omega} X'(T(\omega))P(d\omega), \text{ or } E'(X') = E(X' \circ T),$$

where E' is the expectation operator computed with respect to P'.

(ii) We have

$$X' \in L_1(P') \text{ iff } X' \circ T \in L_1(P)$$

in which case

$$\int_{T^{-1}(A')} X'(T(\omega))P(d\omega) = \int_{A'} X'(\omega')P'(d\omega').$$

Proof. (i) Start with X as an indicator function (a), proceeding to X as a simple function (b) and concluding with X being general (c).

(a): Suppose $X'(\omega') = I_{A'}(\omega')$ for $A' \in \mathcal{B}'$. Then $X'(T(\omega)) = I(T(\omega) \in A') = I(\omega \in T^{-1}(A')) = I_{T^{-1}A'}(\omega)$. Thus

$$\int_\Omega X'(T(\omega))P(d\omega) = \int_\Omega I_{T^{-1}A'}(\omega)P(d\omega) = P(T^{-1}(A'))$$
$$= P'(A') = \int_{\Omega'} I_{A'}(\omega')P'(d\omega') = \int_{\Omega'} X'(\omega')P'(d\omega').$$

(b) Let X' be simple: $X'(\omega') = \sum_{t=1}^k a'_t I_{A'}(\omega')$. Then $X'(T(\omega)) = \sum_{t=1}^k a'_t I_{A'_t}(T(\omega)) = \sum_{t=1}^k a'_t I_{T^{-1}A'_t}(\omega)$. Then everything follows.

Proof continued. (c) Let $X' \geq 0$ which is measurable. There exists an approximating sequence $X'_n \uparrow X'$. By MCT, $E'(X'_n) \uparrow E'(X')$. Also $X'_n \circ T \uparrow X' \circ T$. Then by MCT: $E(X'_n \circ T) \uparrow E(X' \circ T)$. Thus

$$\begin{aligned} \int_\Omega X'(T(\omega))P(d\omega) &= \lim_{n\to\infty} \uparrow \int_\Omega X'_n(T(\omega))P(d\omega) \\ &= \lim_{n\to\infty} \uparrow \int_{\Omega'} X'_n(\omega')P'(d\omega') \\ &= \int_{\Omega'} X'(\omega')P'(d\omega'). \end{aligned}$$

THe proof of (ii) is similar by using $X' I_{A'}$.

5.5.1 Expectation is Always an Integral on R

Let X be a random variable on $(\Omega, \mathcal{B}, P)$ and define the induced probability measure on $(\mathbb{R}, \mathcal{B}(R))$ by

$$F = P \circ X^{-1}, \text{ or } F(A) = P \circ X^{-1}(A) = P[X \in A].$$

The distribution function of X is $F(x) = P[X \leq x]$. Using the Transformation Theorem allows us to compute the abstract integral

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

as

$$E(X) = \int_{\mathbb{R}} x F(dx),$$

which is an integral on $\mathbb{R}$.

5.5.1 Expectation is Always an Integral on R

Corollary 5.5.1 HW 5-4: prove it

(i) If X is an integrable random variable with distribution F, then

$$E(X) = \int_{\mathbb{R}} xF(dx).$$

(ii) Suppose $X : (\Omega, \mathcal{B}) \mapsto (\mathbb{E}, \mathcal{E})$ is a random element of $\mathbb{E}$ with distribution $F = P \circ X^{-1}$ and suppose

$$g : (\mathbb{E}, \mathcal{E}) \mapsto (\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$$

is a non-negative measurable function. The expectation of $g(X)$ is

$$E(g(X)) = \int_{\Omega} g(X(\omega))P(d\omega) = \int_{\mathbb{E}} g(x)F(dx).$$

5.5.2 Densities

Let $\boldsymbol{X} : (\Omega, \mathcal{B}) \mapsto (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ be a random vector on $(\Omega, \mathcal{B}, P)$ with distribution F. We say $\boldsymbol{X}$ or F is **absolutely continuous** (AC) if there exists a non-negative function

$$f : (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k)) \mapsto (\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$$

such that

$$F(A) = \int_A f(\boldsymbol{x}) d\boldsymbol{x}$$

where $d\boldsymbol{x}$ stands for Lebesgue measure and the integral is a Lebesgue-Stieltjes integral.

Proposition 5.5.2

Let $g : (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k)) \mapsto (\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$ be a non-negative measurable function. Suppose $\boldsymbol{X}$ is a random vector with distribution F which is AC with density f, then

$$E(g(\boldsymbol{X})) = \int_{\mathbb{R}} g(\boldsymbol{x}) f(\boldsymbol{x}) d\boldsymbol{x}.$$

5.6 The Riemann vs Lebesgue Integral

We always use Riemann integrals to compute expectations using densities. How does the Riemann integral compare with the Lebesgue integral?

Theorem 5.6.1 (Riemann and Lebesgue)

Suppose $f : (a, b] \mapsto \mathbb{R}$ and

(a) f is $\mathcal{B}((a, b])/\mathcal{B}(\mathbb{R})$ measurable,

(b) f is Riemann-integrable on $(a, b]$.

Let λ be the Lebesgue measure on $(a, b]$. Then

(i) $f \in L_1([a, b], \lambda)$. In fact f is bounded.

(ii) The Riemann integral of f equals the Lebesgue integral.

However, a function could have Lebesgue integral but not Riemann integral. In fact, for a function to be Riemann-integrable, it is necessary and sufficient that the function be bounded and continuous almost everywhere.

Lemma 5.6.1 (Integral Comparison Lemma) HW 5-5: prove it

Suppose X and X' are random variables on $(\Omega, \mathcal{B}, P)$ and suppose $X \in L_1$.

(a) If $P[X = X'] = 1$, then $X' \in L_1$ and $E(X) = E(X')$.

(b) $P[X = X'] = 1$ iff $\int_A X dP = \int_A X' dP$ for all $A \in \mathcal{B}$.

Example 5.6.1 (Riemann and Lebesgue)

Set $\Omega = [0, 1]$ and $P = \lambda =$Lebesgue measure. Let $X(s) = I_{\mathbb{Q}}(s)$ for $s \in \Omega$, where $\mathbb{Q}$ collects the rational real numbers. Then

$$\lambda(\mathbb{Q}) = \lambda(\cup_{r\in\mathbb{Q}}\{r\}) = \sum_{r\in\mathbb{Q}} \lambda(\{r\}) = 0.$$

Thus $\lambda([X = 1]) = 0$ and $\lambda([X = 0]) = 1 - 0 = 1$. Then by Lemma 5.6.1, $E(X) = E(0) = 0$. What about using Riemann integral to calculate $E(X) = \int_{[0,1]} X(s)ds$? No matter how fine we partition the $[0, 1]$, there always exits rational number in a sub-interval. Thus the upper Riemann approximating sum is always 1 while the lower one is always 0. Thus the Riemann integral does not exist but the Lebesgue integral does and is equal to 0.

5.7 Product Spaces, Independence, Fubini Theorem

Let Ω_1, Ω_2 be two sets. Define the **product space**

$$\Omega_1 \times \Omega_2 = \{(\omega_1, \omega_2) : \omega_i \in \Omega_i, i = 1, 2\}$$

and define the **coordinate** or **projection** maps $\pi_i : \Omega_1 \times \Omega_2 \mapsto \Omega_i$, $i = 1, 2$, by

$$\pi_i(\omega_1, \omega_2) = \omega_i$$

If $A \subset \Omega_1 \times \Omega_2$ define

$$A_{\omega_1} = \{\omega_2 : (\omega_1, \omega_2) \in A\} \subset \Omega_2$$
$$A_{\omega_2} = \{\omega_1 : (\omega_1, \omega_2) \in A\} \subset \Omega_1.$$

A_{ω_i} is called the **section** of A at ω_i.

(i) If $A \subset \Omega_1 \times \Omega_2$, then $(A^c)_{\omega_1} = (A_{\omega_1})^c$.

(ii) If, for an index set T, we have $A_\alpha \subset \Omega_1 \times \Omega_2$, for all $\alpha \in T$, then

$$(\cup_\alpha A_\alpha)_{\omega_1} = \cup_\alpha (A_\alpha)_{\omega_1}, \quad (\cap_\alpha A_\alpha)_{\omega_1} = \cap_\alpha (A_\alpha)_{\omega_1}.$$

Let X be a function with domain $\Omega_1 \times \Omega_2$ and range S. Define the **section** of X as

$$X_{\omega_1}(\omega_2) = X(\omega_1, \omega_2)$$

so

$$X_{\omega_1} : \Omega_2 \mapsto S.$$

We think of ω_1 as fixed and the section is a function of varying ω_2. Call X_{ω_1} the **section** of X at ω_1.

(i) $(I_A)_{\omega_1} = I_{A_{\omega_1}}$

(ii) If $S = \mathbb{R}^k$ for some $k \geq 1$ and if for $i = 1, 2$ we have $X_i : \Omega_1 \times \Omega_2 \mapsto S$, then

$$(X_1 + X_2)_{\omega_1} = (X_1)_{\omega_1} + (X_2)_{\omega_1}.$$

(iii) Suppose S is a metric space, $X_n : \Omega_1 \times \Omega_2 \mapsto S$ and $\lim_{n\to\infty} X_n$ exists. Then

$$\lim_{n\to\infty} (X_n)_{\omega_1} = (\lim_{n\to\infty} X_n)_{\omega_1}.$$

A **rectangle** in $\Omega_1 \times \Omega_2$ is a subset of $\Omega_1 \times \Omega_2$ of the form $A_1 \times A_2$ where $A_i \in \Omega_i$, $i = 1, 2$. We call A_1 and A_2 the **sides** of the rectangle. The rectangle is **empty** if at least one of the sides is empty.

Suppose $(\Omega_i, \mathcal{B}_i)$ are two measurable spaces $(i = 1, 2)$. A rectangle is called **measurable** if it is of the form $A_1 \times A_2$ where $A_i \in \mathcal{B}_i$, for $i = 1, 2$. An important **fact:** The class of measurable rectangles is a semi-algebra which we call **RECT**.

We now define a σ-algebra on $\Omega_1 \times \Omega_2$ to be the smallest σ-algebra containing RECT. We denote it by $\mathcal{B}_1 \times \mathcal{B}_2$ and call it the **product** σ-algebra. Thus

$$\mathcal{B}_1 \times \mathcal{B}_2 = \sigma(\text{RECT}).$$

If $\Omega_1 = \Omega_2 = \mathbb{R}$, then

$$\begin{aligned}\mathcal{B}_1 \times \mathcal{B}_2 &= \sigma(A_1 \times A_2 : A_i \in \mathcal{B}(\mathbb{R}), i = 1, 2) \\ &= \sigma(\{I_1 \times I_2 : I_i \text{ is of form } (a, b], i = 1, 2\}).\end{aligned}$$

Lemma 5.7.1 (Sectioning Sets)

Sections of measurable sets are measurable. If $A \in \mathcal{B}_1 \times \mathcal{B}_2$, then for all $\omega \in \Omega_1$,

$$A_{\omega_1} \in \mathcal{B}_2.$$

Proof: Define $\mathcal{C}_{\omega_1} = \{A \subset \Omega_1 \times \Omega_2 : A_{\omega_1} \in \mathcal{B}_2\}$. We prove $\mathcal{C}_{\omega_1} \supset \mathcal{B}_1 \times \mathcal{B}_2 = \sigma(\text{RECT})$. Known RECT is a π-system, by Dynkin's Theorem (2.2.2), it suffices to show that $\mathcal{C}_{\omega_1}$ is a Dynkin's system and RECT$\subset \mathcal{C}_{\omega_1}$.

If $A \in$RECT and $A = A_1 \times A_2$, $A_i \in \mathcal{B}_i$ for $i = 1, 2$, then $A_{\omega_1} = \{\omega_2 : (\omega_1, \omega_2) \in A_1 \times A_2\}$ which equals to $A_2 \in \mathcal{B}_2$ if $\omega_1 \in A_1$ or $\emptyset \in \mathcal{B}_2$ otherwise. Thus $A_{\omega_1} \in \mathcal{C}_{\omega_1}$. It concludes RECT$\subset \mathcal{C}_{\omega_1}$.

Proof continued: We now show $\mathcal{C}_{\omega_1}$ is a Dynkin's system.

(i) $\Omega_1 \times \Omega_2 \in \text{RECT} \subset \mathcal{C}_{\omega_1}$.

(ii) If $A \in \mathcal{C}_{\omega_1}$, then $(A^c)_{\omega_1} = (A_{\omega_1})^c \in \mathcal{B}_2$ because $A_{\omega_1} \in \mathcal{B}_2$. Thus, $A^c \in \mathcal{C}_{\omega_1}$.

(iii) If $A_n \in \mathcal{C}_{\omega_1}$ (meaning $(A_n)_{\omega_1} \in \mathcal{B}$) with $\{A_n\}$ disjoint. Then $(\cup_n A_n)_{\omega_1} = \cup_n (A_n)_{\omega_1} \in \mathcal{B}_2$, thus $\cup_n A_n \in \mathcal{C}_{\omega_1}$.

This completes the proof.

Corollary 5.7.1 (Sectioning Sets)

Sections of measurable function are measurable. That is if

$$X : (\Omega_1 \times \Omega_2, \mathcal{B}_1 \times \mathcal{B}_2) \mapsto (S, \mathcal{S})$$

then

$$X_{\omega_1} \in \mathcal{B}_2.$$

Proof: Since X is $\mathcal{B}_1 \times \mathcal{B}_2/\mathcal{S}$ measurable, we have for $\Lambda \in \mathcal{S}$ that $X^{-1}(\Lambda) = \{(\omega_1, \omega_2) : X(\omega_1, \omega_2) \in \Lambda\} \in \mathcal{B}_1 \times \mathcal{B}_2$. Therefore, by Lemma 5.7.1, $(X^{-1}(\Lambda))_{\omega_1} \in \mathcal{B}_2$. we note

$$\begin{aligned}(X^{-1}(\Lambda))_{\omega_1} &= \{\omega_2 : X(\omega_1, \omega_2) \in \Lambda\} \\ &= \{\omega_2 : X_{\omega_1}(\omega_2) \in \Lambda\} = (X_{\omega_1})^{-1}(\Lambda).\end{aligned}$$

Transition Functions

Call a function

$$K(\omega_1, A_2) : \Omega_1 \times \mathcal{B}_2 \mapsto [0, 1]$$

a **transition function** (or **transition kernel**) if

(i) for each ω_1, $K(\omega_1, \cdot)$ is a probability measure on $\mathcal{B}_2$, and

(ii) for each $A_2 \in \mathcal{B}_2$, $K(\cdot, A_2)$ is $\mathcal{B}_1/\mathcal{B}([0, 1])$ measurable.

We interpret $K(\omega_1, A_2)$ as the conditional probability given ω_1, the result transits to A_2.

Theorem 5.8.1

Let P_1 be a probability measure on $\mathcal{B}_1$, and suppose

$$K : \Omega_1 \times \mathcal{B}_2 \mapsto [0, 1]$$

is a transition function. Then K and P_1, uniquely determine a probability on $\mathcal{B}_1 \times \mathcal{B}_2$ via the formula

$$P(A_1 \times A_2) = \int_{A_1} K(\omega_1, A_2) P_1(d\omega_1).$$

for all $A_1 \times A_2 \in$RECT.

Interpretation: $P(A_1 \times A_2) = P(A_2|A_1)P(A_1)$.

Proof of Theorem 5.8.1: Again, we specified P on the semi-algebra RECT. We need to show P is a valid probability measure on $\mathcal{B}_1 \times \mathcal{B}_2 = \sigma(\text{RECT})$. This can be done by applying the Combo Extension Theorem 2.4.3. It requires us to check P is a σ-additive set function mapping RECT to $[0, 1]$ such that $P(\Omega_1 \times \Omega_2) = 1$.

Because, for each ω_1, $K(\omega_1, \cdot)$ is a probability measure on $\mathcal{B}_2$,

$$K(\omega_1, \Omega_2) = 1, \forall \omega_1 \in \Omega_1.$$

Because P_1 is probability measure on $\mathcal{B}_1$,

$$\begin{aligned} P(\Omega_1 \times \Omega_2) &= \int_{\Omega_1} K(\omega_1, \Omega_2) P_1(d\omega_1) \\ &= \int_{\Omega_1} P_1(d\omega_1) = P_1(\Omega_1) = 1. \end{aligned}$$

Proof of Theorem 5.8.1 (continued): Now we show P is σ-additive on RECT. Let $\{A^{(n)} = A_1^{(n)} \times A_2^{(n)} : n \geq 1\}$ be disjoint elements of RECT whose union is in RECT (i.e., $\cup_{n=1}^{\infty}(A_1^{(n)} \times A_2^{(n)}) = A_1 \times A_2$). We need to show

$$P(A_1 \times A_2) = \sum_{n=1}^{\infty} P(A_1^{(n)} \times A_2^{(n)}).$$

Because $K(\omega_1, \cdot)$ is a probability measure on $\mathcal{B}_2$, for any $A_2 \in \mathcal{B}_2$, $K(\omega_1, A_2) = \int_{\Omega_2} I_{A_2}(\omega_2) K(\omega_1, d\omega_2)$.

$$\begin{aligned} P(A_1 \times A_2) &= \int_{\Omega_1} I_{A_1}(\omega_1) K(\omega_1, A_2) P_1(d\omega_1) \\ &= \int_{\Omega_1} I_{A_1}(\omega_1) \int_{\Omega_2} I_{A_2}(\omega_2) K(\omega_1, d\omega_2) P_1(d\omega_1) \\ &= \int_{\Omega_1} \int_{\Omega_2} I_{A_1}(\omega_1) I_{A_2}(\omega_2) K(\omega_1, d\omega_2) P_1(d\omega_1). \end{aligned}$$

Proof of Theorem 5.8.1 (continued): Because $\cup_{n=1}^{\infty}(A_1^{(n)} \times A_2^{(n)}) = A_1 \times A_2$, $I_{A_1}(\omega_1)I_{A_2}(\omega_2) = I_{A_1 \times A_2}(\omega_1, \omega_2) = \sum_n I_{A_1^{(n)}}(\omega_1)I_{A_2^{(n)}}(\omega_2)$.
Continued, we have

$$\begin{aligned}
P(A_1 \times A_2) &= \int_{\Omega_1}\int_{\Omega_2}\sum_n I_{A_1^{(n)}}(\omega_1)I_{A_2^{(n)}}(\omega_2)K(\omega_1, d\omega_2)P_1(d\omega_1) \\
\text{by MTC} &= \int_{\Omega_1}\sum_n I_{A_1^{(n)}}(\omega_1)\int_{\Omega_2} I_{A_2^{(n)}}(\omega_2)K(\omega_1, d\omega_2)P_1(d\omega_1) \\
\text{by MTC} &= \sum_n \int_{\Omega_1} I_{A_1^{(n)}}(\omega_1)K(\omega_1, A_2^{(n)})P_1(d\omega_1) \\
&= \sum_n \int_{A_1^{(n)}} K(\omega_1, A_2^{(n)})P_1(d\omega_1) \\
&= \sum_n P(A_1^{(n)} \times A_2^{(n)}).
\end{aligned}$$

Special case. Suppose for some probability measure P_2 on $\mathcal{B}_2$ that $K(\omega_1, A_2) = P_2(A_2)$. Then the previously defined P satisfies

$$P(A_1 \times A_2) = P_1(A_1)P_2(A_2).$$

We denote this P by $P_1 \times P_2$ and call P **product measure**. Define σ-algebra ins $\Omega_1 \times \Omega_2$ by $\mathcal{B}_1^{\#} = \{A_1 \times \Omega_2 : A_1 \in \mathcal{B}_1\}$ and $\mathcal{B}_2^{\#} = \{\Omega_1 \times A_2 : A_2 \in \mathcal{B}_2\}$. With respect to the product measure P, we have

$$\mathcal{B}_1^{\#} \perp \mathcal{B}_2^{\#}$$

because $P(A_1 \times \Omega_2 \cap \Omega_1 \times A_2) = P(A_1 \times A_2) = P_1(A_1)P_2(A_2) = P(A_1 \times \Omega_2)P(\Omega_1 \times A_2)$.

Special case continued. Suppose $X_i : (\Omega_i, \mathcal{B}_i) \mapsto (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is a random variable on Ω_t for $i = 1, 2$. Define on $\Omega_1 \times \Omega_2$ the functions

$$X_1^{\#}(\omega_1, \omega_2) = X_1(\omega_1), \quad X_2^{\#}(\omega_1, \omega_2) = X_2(\omega_2)$$

with respect to $P = P_1 \times P_2$. The variables $X_1^{\#}$ and $X_2^{\#}$ are independent because

$$\begin{aligned}
&P[X_1^{\#} \le x, X_2^{\#} \le y] \\
&= P_1 \times P_2(\{(\omega_1, \omega_2) : X_1(\omega_1) \le x, X_2(\omega_2) \le y\}) \\
&= P_1 \times P_2(\{\omega_1 : X_1(\omega_1) \le x\} \times \{\omega_2 : X_2(\omega_2) \le y\}) \\
&= P_1[X_1 \le x]P_2[X_2 \le y] = P_1[X_1 \le x]P_2(\Omega_2)P_1(\Omega_1)P_2[X_2 \le y] \\
&= P([X_1 \le x] \times \Omega_2)P(\Omega_1 \times [X_2 \le y]) \\
&= P(\{(\omega_1, \omega_2) : X_1(\omega_1) \le x\})P(\{(\omega_1, \omega_2) : X_2(\omega_2) \le y\}) \\
&= P[X_1^{\#} \le x]P[X_2^{\#} \le y].
\end{aligned}$$

Independence is automatically built into the model by construction when using product measure.

Theorem 5.9.1

Let P_1 be a probability measure on $(\Omega_1, \mathcal{B}_1)$ and suppose $K : \Omega_1 \times \mathcal{B}_1 \mapsto [0,1]$ is a transition kernel. Define P on $(\Omega_1 \times \Omega_2, \mathcal{B}_1 \times \mathcal{B}_2)$ by $P(A_1 \times A_2) = \int_{A_1} K(\omega_1, A_2) P_1(d\omega_1)$. Assume $X : (\Omega_1 \times \Omega_2, \mathcal{B}_1 \times \mathcal{B}_2) \mapsto (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and suppose $X \geq 0$ (X is integrable). Then

$$Y(\omega_1) = \int_{\Omega_2} K(\omega_1, d\omega_2) X_{\omega_1}(\omega_2)$$

has the properties

(a) Y is well defined.

(b) $Y \in \mathcal{B}_1$

(c) $Y \geq 0$ ($Y \in L_1(P_1)$).

and furthermore

$$\int_{\Omega_1 \times \Omega_2} X dP = \int_{\Omega_1} Y(\omega_1) P_1(d\omega_1)$$
$$= \int_{\Omega_1} \left[\int_{\Omega_2} K(\omega_1, d\omega_2) X_{\omega_1}(\omega_2) \right] P_1(d\omega_1). \qquad (1)$$

Interpretation: When calculating $\int\int h(\omega_1,\omega_2)f(\omega_1,\omega_2)d\omega_2 d\omega_1$, we can have $f(\omega_1,\omega_2) = f_{2|1}(\omega_2|\omega_1)f_1(\omega_1)$ (joint equals conditional times marginal). Then

$$\begin{aligned}
&\int\int h(\omega_1,\omega_2)f(\omega_1,\omega_2)d\omega_2 d\omega_1 \\
= &\int\int h(\omega_1,\omega_2)f_{2|1}(\omega_2|\omega_1)f_1(\omega_1)d\omega_2 d\omega_1 \\
= &\int \underbrace{\int h(\omega_1,\omega_2)f_{2|1}(\omega_1|\omega_2)d\omega_2}_{Y(\omega_1)} \underbrace{f_2(\omega_2)d\omega_1}_{P_1(d\omega_1)}.
\end{aligned}$$

Proof of Theorem 5.9.1: We only show (1) under the assumption $X \geq 0$. Start with the indicator function $X = I_{A_1 \times A_2}$, where $A_1 \times A_2 \in$RECT. Then $\int_{\Omega_1 \times \Omega_2} X dP = \int_{A_1 \times A_2} dP = P(A_1 \times A_2)$. And

$$\begin{aligned}\int_{\Omega_1} Y(\omega_1) P_1(d\omega_1) = & \int_{\Omega_1} [\int_{\Omega_2} K(\omega_1, d\omega_2) I_{A_1}(\omega_1) I_{A_2}(\omega_2)] P_1(d\omega_1) \\ = & \int_{A_1} K(\omega_1, A_2) P_1(d\omega_1) = P(A_1 \times A_2).\end{aligned}$$

Thus (1) holds for indicators of measurable rectangles. Let

$$\mathcal{C} = \{A \in \mathcal{B}_1 \times \mathcal{B}_2 : \text{(1) holds for } X = I_A\},$$

and we know RECT$\subset \mathcal{C}$. We claim $\mathcal{C}$ is a Dynkin system.

Proof of Theorem 5.9.1, continued: We check $\mathcal{C}$ is a Dynkin system:

(i) $\Omega_1 \times \Omega_2 \in \mathcal{C}$.

(ii) If $A \in \mathcal{C}$, $A^c \in \mathcal{C}$. Because for $X = I_{A^c}$, we have

$$\begin{aligned}
\int_{\Omega_1 \times \Omega_2} X dP &= P(A^c) = 1 - P(A) \\
&= 1 - \int_{\Omega_1} \int_{\Omega_2} K(\omega_1, d\omega_2) I_{A_{\omega_1}}(\omega_2) P_1(d\omega_1) \\
&= \int_{\Omega_1} \int_{\Omega_2} K(\omega_1, d\omega_2)(1 - I_{A_{\omega_1}}(\omega_2)) P_1(d\omega_1) \\
&= \int_{\Omega_1} \int_{\Omega_2} K(\omega_1, d\omega_2) I_{(A_{\omega_1})^c}(\omega_2) P_1(d\omega_1) \\
&= \int_{\Omega_1} \int_{\Omega_2} K(\omega_1, d\omega_2) I_{(A^c)_{\omega_1}}(\omega_2) P_1(d\omega_1) \\
&= \int_{\Omega_1} \int_{\Omega_2} K(\omega_1, d\omega_2) X_{\omega_1}(\omega_2) P_1(d\omega_1).
\end{aligned}$$

Proof of Theorem 5.9.1, continued: We check $\mathcal{C}$ is a Dynkin system:

(iii) If $A_n \in \mathcal{C}$, and $\{A_n : n \geq 1\}$ are disjoint events, then $\cup_n A_n \in \mathcal{C}$. Because if $X = I_{\cup_n A_n}$,

$$\int_{\Omega_1 \times \Omega_2} X dP = P(\cup_n A_n) = \sum_n P(A_n)$$

$$= \sum_n \int_{\Omega_1} \int_{\Omega_2} K(\omega_1, d\omega_2) I_{(A_n)_{\omega_1}}(\omega_2) P_1(d\omega_1)$$

$$\text{by MCT} = \int_{\Omega_1} \int_{\Omega_2} K(\omega_1, d\omega_2) \sum_n I_{(A_n)_{\omega_1}}(\omega_2) P_1(d\omega_1)$$

$$= \int_{\Omega_1} \int_{\Omega_2} K(\omega_1, d\omega_2) I_{(\cup_n A_n)_{\omega_1}}(\omega_2) P_1(d\omega_1).$$

Then we have show $\mathcal{C}$ is a Dynkin system and the π-system RECT$\subset \mathcal{C}$. Thus $\sigma(\text{RECT}) = \mathcal{B}_1 \times \mathcal{B}_2 \subset \mathcal{C}$; i.e., for any $A \in \mathcal{B}_1 \times \mathcal{B}_2$, (1) holds for $X = I_A$.

Proof of Theorem 5.9.1, continued: If $X = \sum_{i=1}^{k} a_i I_{A_i}$, where $A_i \in \mathcal{B}_1 \times \mathcal{B}_2$. It is easy to check (1) holds.

For arbitrary $X \geq 0$, denoted its approximating sequence by $X_n \uparrow X$. By monotone convergence, $\int_{\Omega_1\times\Omega_2} X_n dP \uparrow \int_{\Omega_1\times\Omega_2} X dP$. We know (1) holds for each X_n; i.e.,

$$
\begin{aligned}
& \lim_{n\to\infty} \uparrow \int_{\Omega_1\times\Omega_2} X_n dP \\
&= \lim_{n\to\infty} \uparrow \int_{\Omega_1} \int_{\Omega_2} K(\omega_1, d\omega_2)(X_n)_{\omega_1}(\omega_2) P_1(d\omega_1) \\
\text{by MCT} &= \int_{\Omega_1} [\lim_{n\to\infty} \uparrow \int_{\Omega_2} K(\omega_1, d\omega_2)(X_n)_{\omega_1}(\omega_2)] P_1(d\omega_1) \\
\text{by MCT} &= \int_{\Omega_1} [\int_{\Omega_2} K(\omega_1, d\omega_2) \lim_{n\to\infty} \uparrow (X_n)_{\omega_1}(\omega_2)] P_1(d\omega_1) \\
&= \int_{\Omega_1} [\int_{\Omega_2} K(\omega_1, d\omega_2) X_{\omega_1}(\omega_2)] P_1(d\omega_1).
\end{aligned}
$$

Theorem 5.9.2 Fubini Theorem

Let $P = P_1 \times P_2$ be the product measure. If X is $\mathcal{B}_1 \times \mathcal{B}_2$ measurable and is either non-negative or integrable with respect to P, then

$$\int_{\Omega_1 \times \Omega_2} X dP = \int_{\Omega_1} [\int_{\Omega_2} X_{\omega_1}(\omega_2) P_2(d\omega_2)] P_1(d\omega_1)$$
$$= \int_{\Omega_2} [\int_{\Omega_1} X_{\omega_2}(\omega_1) P_1(d\omega_1)] P_2(d\omega_2).$$

Proof: Let $K(\omega_1, A_2) = P_2(A_2)$. Then P_1 and K determine $P = P_1 \times P_2$ on $\mathcal{B}_1 \times \mathcal{B}_2$ and

$$\begin{aligned}\int_{\Omega_1 \times \Omega_2} XdP &= \int_{\Omega_1} [\int_{\Omega_2} K(\omega_1, d\omega_2) X_{\omega_1}(\omega_2)] P_1(d\omega_1) \\ &= \int_{\Omega_1} [\int_{\Omega_2} P_2(d\omega_2) X_{\omega_1}(\omega_2)] P_1(d\omega_1).\end{aligned}$$

Also let $\tilde{K}(\omega_2, A_1) = P_1(A_1)$ be a transition kernel with $\tilde{K} : \Omega_2 \times \mathcal{B}_1 \mapsto [0, 1]$. Then $\tilde{K}$ and P_2 also determine $P = P_1 \times P_2$ and we have

$$\begin{aligned}\int_{\Omega_1 \times \Omega_2} XdP &= \int_{\Omega_2} [\int_{\Omega_1} \tilde{K}(\omega_2, d\omega_1) X_{\omega_2}(\omega_1)] P_2(d\omega_2) \\ &= \int_{\Omega_2} [\int_{\Omega_1} P_1(d\omega_1) X_{\omega_2}(\omega_1)] P_2(d\omega_2).\end{aligned}$$

Example 5.9.2

Let $X_i \geq 0$, $i = 1, 2$, be two independent random variables. Then

$$E(X_1 X_2) = E(X_1)E(X_2).$$

Proof: Let $\boldsymbol{X} = (X_1, X_2)$, $g(x_1, x_2) = x_1 x_2$, F_i the distribution of X_i. Then $P \circ \boldsymbol{X}^{-1}(A_1 \times A_2) = P[(X_1, X_2) \in A_1 \times A_2] = P[X_1 \in A_1, X_2 \in A_2] = P_1[X_1 \in A_1]P_2[X_2 \in A_2] = F_1(A_1)F_2(A_2) = F_1 \times F_2(A_1 \times A_2)$. So $P \circ \boldsymbol{X}^{-1}$ and $F_1 \times F_2$ agree on RECT and hence on $\mathcal{B}(\text{RECT}) = \mathcal{B}_1 \times \mathcal{B}_2$. From Corollary 5.5.1 we have

$$E(X_1 X_2) = E(g(\boldsymbol{X})) = \int_{\mathbb{R}_+^2} g(\boldsymbol{x}) P \circ \boldsymbol{X}^{-1}(d\boldsymbol{x}) = \int_{\mathbb{R}_+^2} g d(F_1 \times F_2)$$

$$\text{by Fubini} = \int_{\mathbb{R}_+} x_2 [\int_{\mathbb{R}_+} x_1 F_1(dx_1)] F_2(dx_2)$$

$$= E(X_1) \int x_2 F_2(dx_2) = E(X_1)E(X_2).$$

Example 5.9.3 (Convolution)

Suppose X_1 and X_2 are two independent random variables with distributions F_1, F_2. The distribution of $X_1 + X_2$ is given by **convolution** $F_1 * F_2$. For $x \in \mathbb{R}$,

$$P[X_1+X_2 \le x] = F_1*F_2(x) = \int_{\mathbb{R}} F_1(x-u)F_2(du) = \int_{\mathbb{R}} F_2(x-u)F_1(du).$$

Let $\boldsymbol{X} = (X_1, X_2)$ which has distribution $F_1 \times F_2$ and set

$$g(x_1, x_2) = I_{\{(u,v):u+v\le x\}}(x_1, x_2), \quad (x_1, x_2) \in \mathbb{R}^2$$

Then $P[X_1 + X_2 \le x] = E(g(\boldsymbol{X})) = \int_{\mathbb{R}^2} gd(F_1 \times F_2)$

Fubini $= \int_{\mathbb{R}} [\int_{\mathbb{R}} I_{\{(u,v):u+v\le x\}}(x_1, x_2)F_1(dx_1)]F_2(dx_2)$

$$= \int_{\mathbb{R}} [\int_{\mathbb{R}} I_{\{v:v\le x-x_2\}}(x_1)F_1(dx_1)]F_2(dx_2) = \int_{\mathbb{R}} F_1(x-x_2)F_2(dx_2).$$

Other HW 5 problems: Section 5.10, Q5-7, Q9-12, Q14-Q16, Q18, Q20, Q22, Q25, Q30-31, Q36.