Expectation: Limits, computation

Integration to the limit: already proved

The limit–integral interchange results were proved in the integration section.

Since $\mathbb{E}[\cdot]$ *is* an integral, they transfer verbatim:

- Fatou's lemma (for $X_n \geq 0$)
- Monotone convergence theorem (for $0 \leq X_n \uparrow X$)
- Dominated convergence theorem (if $X_n \to X$ a.s. and $|X_n| \leq Y$, $\mathbb{E}[Y] < \infty$)
- Bounded convergence theorem (dominated by a constant)

Important. We will *not* re-prove these results; today we focus on how they are used for expectations.

Fatou, MCT, DCT, BCT (Expectation Form)

MCT. If $0 \le X_n \uparrow X$, then

$$\mathbb{E}[X_n] \uparrow \mathbb{E}[X].$$

Fatou. If $X_n \ge 0$, then

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

DCT. If $X_n \to X$ a.s., $|X_n| \le Y$, and $\mathbb{E}[Y] < \infty$, then

$$\mathbb{E}[X_n] \to \mathbb{E}[X].$$

BCT. (DCT with a constant bound.) If $|X_n| \le M$, then $\mathbb{E}[X_n] \to \mathbb{E}[X]$.

A useful strengthening beyond DCT (optional depth)

Sometimes we can prove $\mathbb{E}[h(X_n)] \to \mathbb{E}[h(X)]$ without a *pointwise* dominating random variable.

Theorem (integration to the limit via truncation). Suppose $X_n \to X$ a.s. Let g, h be continuous with:

- $g \geq 0$ and $g(x) \to \infty$ as $|x| \to \infty$,
- $\dfrac{|h(x)|}{g(x)} \to 0$ as $|x| \to \infty$,
- $\mathbb{E}[g(X_n)] \leq K < \infty$ for all n.

Then

$$\mathbb{E}[h(X_n)] \to \mathbb{E}[h(X)].$$

Key idea: truncate to $|X_n| \leq M$, use BCT on the truncated part, then control tails using (iii).

Intuition: what replaces domination?

This theorem gives convergence of expectations without a single dominating random variable.

What replaces domination?

- Tail control via $g(X_n)$ (uniform bound on $\mathbb{E}[g(X_n)]$)
- h grows strictly slower than g at infinity

Condition (iii),

$$\sup_n \mathbb{E}[g(X_n)] < \infty,$$

prevents mass from escaping to infinity.

Condition (ii),

$$\frac{|h(x)|}{g(x)} \to 0 \quad (|x| \to \infty),$$

makes the contribution from large values negligible.

The proof follows a simple idea.

Step 1: Truncate. Fix M and write

$$\bar{X}_n := X_n \mathbf{1}_{\{|X_n| \le M\}}.$$

On $\{|X_n| \le M\}$, $h(\bar{X}_n)$ is bounded and $h(\bar{X}_n) \to h(\bar{X})$ a.s., so BCT applies.

Step 2: Control the tails. Outside $\{|X_n| \le M\}$,

$$|h(X_n)| \le \varepsilon_M \, g(X_n), \qquad \varepsilon_M := \sup_{|x|>M} \frac{|h(x)|}{g(x)} \to 0.$$

Uniform control of $\mathbb{E}[g(X_n)]$ makes the tail contribution vanish.

A common choice is

$$g(x) = |x|^p, \qquad h(x) = x, \qquad p > 1.$$

If $\sup_n \mathbb{E}|X_n|^p < \infty$ and $X_n \to X$ a.s., then

$$\mathbb{E}[X_n] \to \mathbb{E}[X].$$

Interpretation: an L^p bound with $p > 1$ prevents mass from escaping to infinity.

(Closely related to *uniform integrability*.)

Computing expectation via change of variables

Let X be a random element of $(S, \mathcal{S})$ with distribution

$$\mu(A) = P(X \in A).$$

If $f \geq 0$ or $\mathbb{E}|f(X)| < \infty$, then

$$\mathbb{E}[f(X)] = \int_S f(y)\,\mu(dy).$$

Interpretation. Expectations can be computed entirely from the distribution of X. Law of the unconscious statistician (LOTUS).

Remark. Writing $\mu = P \circ X^{-1}$,

$$\int_\Omega f(X(\omega))\,dP(\omega) = \int_S f(y)\,d(P \circ X^{-1})(y).$$

Proof sketch: extension argument

The identity is proved by extending from simple cases.

1. **Indicator functions.** If $f = \mathbf{1}_B$ with $B \in \mathcal{S}$, then

$$\mathbb{E}[\mathbf{1}_B(X)] = P(X \in B) = \mu(B) = \int_S \mathbf{1}_B(y)\,\mu(dy).$$

2. **Simple functions.** For $f = \sum_{m=1}^{n} c_m \mathbf{1}_{B_m}$, the result follows by linearity of expectation and integration.
3. **Nonnegative functions.** For $f \geq 0$, define

$$f_n(x) := \left(\frac{\lfloor 2^n f(x) \rfloor}{2^n}\right) \wedge n.$$

   Then f_n is simple and $f_n \uparrow f$. (Visualization of the approximation: Desmos plot, courtesy of Noah.) Apply the monotone convergence theorem to both sides.
4. **Integrable functions.** Write $f = f^+ - f^-$ and apply the nonnegative case to each part.

Product Measures and Iterated Integrals

Why product measures?

Many identities in probability involve two variables:

- an outcome $\omega \in \Omega$,
- a real variable (time, threshold, level).

To justify exchanging orders of integration, we need a measure on a *product space*.

Product measures: setup

Let $(X, \mathcal{A}, \mu_1)$ and $(Y, \mathcal{B}, \mu_2)$ be σ-finite measure spaces.

Define the product space

$$X \times Y := \{(x, y) : x \in X,\ \ y \in Y\}.$$

Rectangles. For $A \in \mathcal{A}$ and $B \in \mathcal{B}$, the set

$$A \times B := \{(x, y) : x \in A,\ \ y \in B\}$$

is called a *rectangle*. Let

$$\mathcal{S} := \{A \times B :\ A \in \mathcal{A},\ B \in \mathcal{B}\}.$$

Product σ-field.

$$\mathcal{A} \otimes \mathcal{B} := \sigma(\mathcal{S}).$$

Rectangles form a semi-algebra

The class $\mathcal{S}$ of rectangles behaves well under basic set operations.

If $A, C \in \mathcal{A}$ and $B, D \in \mathcal{B}$, then:

- **Intersection:** $(A \times B) \cap (C \times D) = (A \cap C) \times (B \cap D)$.
- **Difference:** $(A \times B) \setminus (C \times D)$ is a *finite disjoint union* of rectangles.[1]
- **Complement:** $(A \times B)^c = (A^c \times Y) \;\cup\; (X \times B^c)$.

So $\mathcal{S}$ is a *semi-algebra*, and generates $\mathcal{A} \otimes \mathcal{B}$.

[1] You can write it explicitly as $\big((A \setminus C) \times B\big) \;\dot\cup\; \big((A \cap C) \times (B \setminus D)\big)$.

Theorem (Product measure). If μ_1 and μ_2 are σ-finite, then there exists a unique measure μ on $\mathcal{A} \otimes \mathcal{B}$ such that

$$\mu(A \times B) = \mu_1(A)\mu_2(B) \quad \text{for all } A \in \mathcal{A},\ B \in \mathcal{B}.$$

Notation: $\mu = \mu_1 \times \mu_2$.

Big picture: we define μ first on rectangles, then extend to all of $\mathcal{A} \otimes \mathcal{B}$ (using the extension theorem).

Proof idea: why $\mu(A \times B) = \mu_1(A)\mu_2(B)$ is consistent

Define a *premeasure* on rectangles by

$$\mu_0(A \times B) := \mu_1(A)\mu_2(B).$$

To apply the extension theorem, we must check **countable additivity on rectangles**: if

$$A \times B = \biguplus_{i=1}^{\infty} (A_i \times B_i),$$

then

$$\mu_1(A)\mu_2(B) = \sum_{i=1}^{\infty} \mu_1(A_i)\mu_2(B_i).$$

Key trick (slice by x). For fixed $x \in X$, define $I(x) := \{i : \ x \in A_i\}$. Then

$$\mathbf{1}_A(x)\, \mu_2(B) = \sum_{i=1}^{\infty} \mathbf{1}_{A_i}(x)\, \mu_2(B_i).$$

Integrate both sides over x with respect to μ_1 to get the desired

Extension: n-fold products (optional)

By induction, if $(X_m, \mathcal{F}_m, \mu_m)$ are σ-finite for $m = 1, \ldots, n$, there is a unique measure on the product space

$$X_1 \times \cdots \times X_n$$

satisfying, on rectangles,

$$\mu(A_1 \times \cdots \times A_n) = \prod_{m=1}^{n} \mu_m(A_m).$$

In particular, taking $(X_m, \mathcal{F}_m, \mu_m) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ recovers Lebesgue measure on $\mathbb{R}^n$.

Tonelli's theorem (nonnegative case)

Theorem (Tonelli). If $f \geq 0$ is measurable on $X \times Y$, then

$$\int_{X\times Y} f \, d(\mu_1 \times \mu_2) = \int_X \left(\int_Y f(x,y)\, \mu_2(dy) \right) \mu_1(dx) = \int_Y \left(\int_X f(x,y)\, \mu_1(dx) \right) \mu_2(dy).$$

No integrability assumptions are required; integrals may be $+\infty$.

Theorem (Fubini). If

$$\int_{X\times Y} |f|\, d(\mu_1 \times \mu_2) < \infty,$$

then the iterated integrals exist and are equal:

$$\int_{X\times Y} f = \int_X \int_Y f = \int_Y \int_X f.$$

Key distinction: Tonelli for $f \geq 0$, Fubini for $\int |f| < \infty$.

Let $Z \geq 0$ be a nonnegative random variable.

Claim.

$$\mathbb{E}[Z] = \int_0^\infty P(Z > t)\, dt.$$

This is a direct application of Tonelli on $\Omega \times \mathbb{R}_+$.

Proof via Tonelli

Define $f(\omega, t) := \mathbf{1}_{\{Z(\omega)>t\}}$ on $\Omega \times \mathbb{R}_+$.

For each ω,

$$\int_0^\infty f(\omega, t)\, dt = \int_0^\infty \mathbf{1}_{\{Z(\omega)>t\}}\, dt = Z(\omega).$$

Since $f \geq 0$, Tonelli gives

$$\mathbb{E}[Z] = \int_\Omega \left(\int_0^\infty f(\omega, t)\, dt \right) dP(\omega) = \int_0^\infty \left(\int_\Omega f(\omega, t)\, dP(\omega) \right) dt.$$

But $\int_\Omega f(\omega, t)\, dP(\omega) = P(Z > t)$, hence

$$\mathbb{E}[Z] = \int_0^\infty P(Z > t)\, dt.$$