Lecture 4

Wednesday, January 7, 2026 12:39 PM

Definition 2.1.2. Let $X \in \mathcal{L}^1$ be a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$, and let $\mathcal{G} \subset \mathcal{F}$ be a finite σ-algebra generated by the partition $\{A_k\}_{k=1,\ldots,n}$. Then

$$\mathbb{E}(X|\mathcal{G}) \equiv \sum_{k=1}^{n} \mathbb{E}(X|A_k)\, I_{A_k},$$

where $\mathbb{E}(X|A) = \mathbb{E}(XI_A)/\mathbb{P}(A)$ if $\mathbb{P}(A) > 0$ and we may define $\mathbb{E}(X|A)$ arbitrarily if $\mathbb{P}(A) = 0$. The conditional expectation thus defined is unique up to a.s. equivalence, i.e., any two random variables $Y, \tilde{Y}$ that satisfy the definition obey $Y = \tilde{Y}$ a.s. (why?).

Remark 2.1.3. $X \in \mathcal{L}^1$ ensures that $\mathbb{E}(X|A_k)$ is well defined and finite.

Definition 2.3.1 (Kolmogorov). Let $X \in \mathcal{L}^1$ and let $\mathcal{F}$ be any σ-algebra. Then $\mathbb{E}(X|\mathcal{F})$ is, by definition, the unique $\mathcal{F}$-measurable random variable that satisfies the relation $\mathbb{E}(I_A X) = \mathbb{E}(I_A \mathbb{E}(X|\mathcal{F}))$ for all events $A \in \mathcal{F}$.

This is usually taken as the starting point in developing conditional expectations, but with your knowledge of martingales it should now be evident that this is just the limiting case of the familiar discrete conditional expectation, but in disguise. On the other hand, this definition is often much easier to deal with: the definition itself does not involve taking any limits (the limits are only involved in proving *existence* of an object that satisfies the definition!)

(*) As you see, the $\mathbb{E}(X|Y)$ doesn't depend on the realized value of Y, rather depend on "the partitions" generated by this realization of Y.

The following property makes precise the idea that $\mathbb{E}(X|\mathcal{F})$ can be interpreted as the *best estimate* of X given the information $\mathcal{F}$. In other words, we give the conditional expectation (a probabilistic concept) a *statistical* interpretation.

Proposition 2.3.3 (Least squares property). *Let $X \in \mathcal{L}^2$. Then $\mathbb{E}(X|\mathcal{F})$ is the least-mean-square estimate of X given $\mathcal{F}$, i.e., $\mathbb{E}(X|\mathcal{F})$ is the unique $\mathcal{F}$-measurable random variable that satisfies $\mathbb{E}((X - \mathbb{E}(X|\mathcal{F}))^2) = \min_{Y \in \mathcal{L}^2(\mathcal{F})} \mathbb{E}((X - Y)^2)$, where $\mathcal{L}^2(\mathcal{F}) = \{Y \in \mathcal{L}^2 : Y \text{ is } \mathcal{F}\text{-measurable}\}$.*

Beside its statistical interpretation, you can also interpret this result *geometrically*: the conditional expectation $\mathbb{E}(X|\mathcal{F})$ is the *orthogonal projection* of $X \in \mathcal{L}^2$ onto the linear subspace $\mathcal{L}^2(\mathcal{F}) \subset \mathcal{L}^2$ with respect to the inner product $\langle X, Y \rangle = \mathbb{E}(XY)$.

Proof. First, note that $\mathbb{E}((X-Y)^2)$ is finite for any $Y \in \mathcal{L}^2(\mathcal{F})$ by Hölder's inequality. Clearly $\mathbb{E}((X-Y)^2) = \mathbb{E}((X - \mathbb{E}(X|\mathcal{F}) + \mathbb{E}(X|\mathcal{F}) - Y)^2)$. But $\Delta = \mathbb{E}(X|\mathcal{F}) - Y$ is $\mathcal{F}$-measurable by construction, so using the elementary properties of the conditional expectation (and Hölder's inequality to show that $X\Delta \in \mathcal{L}^1$) we see that $\mathbb{E}(\mathbb{E}(X|\mathcal{F})\Delta) = \mathbb{E}(X\Delta)$. Thus $\mathbb{E}((X-Y)^2) = \mathbb{E}((X - \mathbb{E}(X|\mathcal{F})^2) + \mathbb{E}(\Delta^2)$. [Recall the geometric intuition: $\Delta \in \mathcal{L}^2(\mathcal{F})$ and $\mathbb{E}(X|\mathcal{F})$ is the orthogonal projection of X onto $\mathcal{L}^2(\mathcal{F})$, so $H = X - \mathbb{E}(X|\mathcal{F}) \perp \mathcal{L}^2(\mathcal{F})$, and thus $\langle H, \Delta\rangle = \mathbb{E}(H\Delta) = 0$.] The least squares property follows, as $\mathbb{E}(\Delta^2) \geq 0$.

To prove that the minimum is unique, suppose that Y_* is another $\mathcal{F}$-measurable random variable that minimizes $\mathbb{E}((X-Y)^2)$. Then $\mathbb{E}((X-Y_*)^2) = \mathbb{E}((X - \mathbb{E}(X|\mathcal{F}))^2)$. But by

the general formula above, $\mathbb{E}((X-Y_*)^2) = \mathbb{E}((X-\mathbb{E}(X|\mathcal{F}))^2) + \mathbb{E}((\mathbb{E}(X|\mathcal{F}) - Y_*)^2)$. It follows that we must have $\mathbb{E}((\mathbb{E}(X|\mathcal{F}) - Y_*)^2) = 0$; this implies that $\mathbb{E}(X|\mathcal{F}) = Y_*$ a.s. □

Final remark about "Conditional Expectation":

we have three ways to think about Cond. Expec.

I) (Kolmogorov): most useful to use in theory, hard to interpret
showing uniqueness (a.s.) is easy, Existence is hard (R.-N. Thm)

II) (mean squared estimator): Follows directly by properties of $\mathcal{L}_2$
but only applies to r.v. in $\mathcal{L}_2$ ($\mathbb{E}(|X|^2)<\infty$) and
not trivial to generalize it to $X \in \mathcal{L}_1$ (only $\mathbb{E}(|X|)<\infty$).

III) (limit of discrete cond. expectation):
with "martingale conv. thm" we can easily show
the existence but not the uniqueness; harder to work with
[only natural for separable σ-algebras (generated by countable
collection of events (or discrete σ-algebras).]

Continuous time Continuous-path stochastic processes:

Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Def: $\{X_t, t \in [0,T]\}$ is a stochastic process on $(\Omega, \mathcal{F}, \mathbb{P})$ if each X_t is a r.v. on Ω.

Def: $\{X_t\}$ is called measurable process if

$$X : [0,T] \times \Omega \to \mathbb{R} \text{ is } \mathcal{B}([0,T]) \times \mathcal{F}\text{-measurable}$$

So, $Y : \Omega \to \mathbb{R}$, $Y(\omega) = \int_0^T X_t(\omega)\, dt$ is a well-defined r.v.

Now, under mild assumptions of Fubini or Tonelli, we get

$$\mathbb{E}[Y] = \mathbb{E}\left[\int_0^T X_t(\omega)\, dt\right] = \int_0^T E[X_t]\, dt \quad !$$

Def: A filteration is a collection of sub σ-algebras $\{\mathcal{F}_t \subset \mathcal{F} \,;\, t \in \mathbb{T}\}$ such that $\mathcal{F}_t \subset \mathcal{F}_s$ if $s > t$.

Def: $\{X_t : t \in \mathbb{T}\}$ is a continuous-path process if $t \mapsto X_t(\omega)$ is continuous for almost all $\omega \in \Omega$.

Definition 2.4.5. Let X_t be a stochastic process on some filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ and time set $\mathbb{T} \subset [0, \infty[$ (e.g., $\mathbb{T} = [0, T]$ or $[0, \infty[$). Then X_t is called *adapted* if X_t is $\mathcal{F}_t$-measurable for all t, is called *measurable* if the random variable $X_\cdot : \mathbb{T} \times \Omega \to \mathbb{R}$ is $\mathcal{B}(\mathbb{T}) \times \mathcal{F}$-measurable, and is called *progressively measurable* if $X_\cdot : [0,t] \cap \mathbb{T} \times \Omega \to \mathbb{R}$ is $\mathcal{B}([0,t] \cap \mathbb{T}) \times \mathcal{F}_t$-measurable for all t.

What do these definitions mean? Adapted you know; measurable means that

$$Y_t = \int_0^t X_s \, ds \quad \text{is well defined and } \mathcal{F}\text{-measurable,}$$

and progressively measurable means that

$$Y_t = \int_0^t X_s \, ds \quad \text{is well defined and } \mathcal{F}_t\text{-measurable;}$$

in particular, progressive measurability guarantees that the process Y_t is adapted.[3]

Ask them to recall Tonelli and Fubini.

Definition 1.6.5. Given two measurable spaces $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$, we define the *product space* $\Omega_1 \times \Omega_2 \equiv \{(\omega_1, \omega_2) : \omega_1 \in \Omega_1,\ \omega_2 \in \Omega_2\}$. The *product σ-algebra* on $\Omega_1 \times \Omega_2$ is defined as $\mathcal{F}_1 \times \mathcal{F}_2 \equiv \sigma\{\rho_1, \rho_2\}$, where $\rho_i : \Omega_1 \times \Omega_2 \to \Omega_i$ are the projection maps $\rho_1 : (\omega_1, \omega_2) \mapsto \omega_1$ and $\rho_2 : (\omega_1, \omega_2) \mapsto \omega_2$.

Theorem 1.6.6. *Let $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ and $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ be two probability spaces, and denote by $\rho_1 : \Omega_1 \times \Omega_2 \to \Omega_1$ and $\rho_2 : \Omega_1 \times \Omega_2 \to \Omega_2$ the projection maps.*

1. *There exists a unique probability measure $\mathbb{P}_1 \times \mathbb{P}_2$ on $\mathcal{F}_1 \times \mathcal{F}_2$, called the product measure, under which the law of ρ_1 is $\mathbb{P}_1$, the law of ρ_2 is $\mathbb{P}_2$, and ρ_1 and ρ_2 are independent Ω_1- and Ω_2-valued random variables, respectively.*
2. *The construction is in some sense unique: suppose that on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ are defined an S-valued random variable X (with law μ_X) and a T-valued random variable Y (with law μ_Y). Then the law of the $S \times T$-valued random variable (X, Y) is the product measure $\mu_X \times \mu_Y$.*
3. *If $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ is $\mathcal{F}_1 \times \mathcal{F}_2$-measurable, then $f(\omega_1, \cdot) : \Omega_2 \to \mathbb{R}$ is $\mathcal{F}_2$-measurable for all ω_1, and $f(\cdot, \omega_2) : \Omega_1 \to \mathbb{R}$ is $\mathcal{F}_1$-measurable for all ω_2.*
4. ***(Tonelli)** If $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ is $\mathcal{F}_1 \times \mathcal{F}_2$-measurable and $f \geq 0$ a.s., then*

   $$\mathbb{E}_{\mathbb{P}_1 \times \mathbb{P}_2}(f) = \mathbb{E}_{\mathbb{P}_2}\left[\int f(\omega_1, \omega_2)\, \mathbb{P}_1(d\omega_1)\right] = \mathbb{E}_{\mathbb{P}_1}\left[\int f(\omega_1, \omega_2)\, \mathbb{P}_2(d\omega_2)\right].$$

   In particular, these expressions make sense (the inner integrals are measurable).
5. ***(Fubini)** The previous statement still holds for random variables f that are not necessarily nonnegative, provided that $\mathbb{E}(|f|) < \infty$.*

Tonelli and Fubini theorems are useful for interpreting

$$\mathbb{E}_{\mathbb{P}_1 \times \mathbb{P}_2}[f(X_1, X_2)] = E_{\mathbb{P}_1}[E_{\mathbb{P}_2}[f(X_1, X_2)]]$$

For continuous-path $\mathcal{F}_t$-adapted processes $\{X_t\}$

1) measurable = progressively measurable.

2) $\inf_t X_t$, $\sup_t X_t$, $\liminf_t X_t$, $\limsup_t X_t$ are all well-defined measurable r.v.. (Proof: $\inf_t X_t = \inf_{t \in \mathbb{Q}} X_t$)

3) **Lemma 2.4.8.** *Let X_t be an a.s. nonnegative stochastic process with continuous sample paths. Then $\mathbb{E}(\liminf_t X_t) \leq \liminf_t \mathbb{E}(X_t)$. If there is a $Y \in \mathcal{L}^1$ such that $X_t \leq Y$ a.s. for all t, then $\mathbb{E}(\limsup_t X_t) \geq \limsup_t \mathbb{E}(X_t)$.*

Proof: by contradiction assume $\exists$ a subseq. t_n otherwise!