Conditional expectation is one of the most useful tools of probability. The Radon-Nikodym theorem enables us to construct conditional expectations.

**Definition 1.** A measure $\mu$ on $(\Omega, F)$ is $\sigma$-finite iff there is $A_n \in F$ satisfying $A_n \uparrow \Omega$ where $\mu(A_n) < \infty$ for all $n$.

**Definition 2.** Let $\mu, \nu$ be measures on $(\Omega, F)$. Then $\nu$ is absolutely continuous with respect to $\mu$ ($\nu \ll \mu$) iff for all $A \in F$, $\mu(A) = 0$ means $\nu(A) = 0$.

For instance, let $(\Omega, F, P)$ be a probability space and let $X: \Omega \to [0, \infty)$ be a random variable. Let $\nu(A) = \int_A X dP$ for all $A \in F$. Then $\nu$ is a $\sigma$-finite measure which is absolutely continuous ($\nu \ll P$).

**Proof.** Let $A_n = \{X_n \leq n\}$, which increases to $\Omega$. Then $\nu(A_n) \leq \int n dP = n < \infty$ implies that $\nu$ is $\sigma$-finite. Now let $A$ satisfy $P(A) = 0$. Then $X1_A = 0$ a.s. implies that $\int X1_A dP = \int 0 dP = 0 = \nu(A)$. Hence $\nu \ll P$.

The Radon-Nikodym theorem says that the converse of the above phenomenon is true.

**Theorem 4.1** (Radon-Nikodym). Let $\nu, \mu$ be $\sigma$-finite measures on a measure space $(\Omega, F)$ such that $\nu \ll \mu$. Then there is a random variable $X: \Omega \to [0, \infty)$ such that, for all $D \in F$, $\nu(D) = \int_D X d\mu$. Write $X = \frac{d\nu}{d\mu}$. If $X'$ is another random variable satisfying the previous condition, then $X = X'$ almost surely.

**Proof.** Refer to Durrett, Chapters A.4.5 and A.4.6

Radon-Nikodym gives us existence of conditional expectations. Write $D \subset \sigma F$ if $D$ is a sub $\sigma$-field of $F$.

**Theorem 4.2.** Suppose $D \subset \sigma F$ and $X$ is integrable on $(\Omega, F)$. Then there is a unique, up to $P$-null sets, random variable $E(X|D)(\omega)$ which is i) $D$-measurable and ii) for all $D \in D$, we have:

$$\int_D E(X|D) dP = \int_D X dP$$

**Call $E(X|D)$ the conditional expectation of $X$ given $D$.**

**Proof.** Case 1: Firstly, suppose $X$ is a non-negative random variable. Define $\nu$ on sub $\sigma$-algebra $D$ by $\nu(D) = \int_D X dP$. We have $\nu \ll P$ where $P$ is the measure on the original probability space restricted to $D$. Hence, Radon Nikodym says that there is a random variable $E(X|D) > 0$ such that for all sets $D \in D$, we have $\int_D X dP = \int_D E(X|D) dP$.

Case 2: In the general case, suppose $X = X^+ - X^-$. Apply Case 1 to each of $X^+, X^-$ and define $E(X|D) = E(X^+|D) - E(X^-|D)$. Then the resulting random variable is $D$-measurable.
since each \( x - y \) is a measurable function whenever \( x, y \) are, and furthermore ii) holds by the linearity of the integral.

Uniqueness: Suppose \( \tilde{E}(X|\mathcal{D}) = E(X|\mathcal{D}) \) is a random variable satisfying conditions i/ii. We need to show that they are equal almost surely. Consider the set \( D_n = \{ \omega : (\tilde{E}(X|\mathcal{D}) - E(X|\mathcal{D}))\omega \geq \frac{1}{n} \} \), which is \( \mathcal{D} \)-measurable. Then,

\[
P(D_n) \leq \int_{D_n} (\tilde{E}(X|\mathcal{D}) - E(X|\mathcal{D})) dP = 0
\]

since each of \( \tilde{E}(X|\mathcal{D}) \) and \( E(X|\mathcal{D}) \) integrate like \( X \) over \( \mathcal{D} \)-measurable sets. Hence, \( D_n \uparrow \{ \tilde{E}(X|\mathcal{D}) > E(X|\mathcal{D}) \} \) means that \( P(\tilde{E}(X|\mathcal{D}) > E(X|\mathcal{D})) = \lim P(D_n) = 0 \). Hence, \( \tilde{E}(X|\mathcal{D}) \leq E(X|\mathcal{D}) \) almost surely, and by symmetry, the reverse inequality holds a.s.. Hence, \( \tilde{E}(X|\mathcal{D}) = E(X|\mathcal{D}) \), almost surely. \( \square \)

If \( Y \) is a random vector, define \( E(X|Y) = E(X|\sigma(Y)) \). Furthermore, if \( B \) is an event with \( P(B) > 0 \), then \( E(X|B) = \frac{1}{P(B)} \int_B X dP \in \mathbb{R} \).

The following are some examples of conditional expectation, with an intention of relating the abstract definition of conditional expectation to more concrete examples.

- Let \( \Omega = B_1 \cup \ldots \cup B_N \) where each \( B_i \in \mathcal{F} \) satisfy \( P(B_i) > 0 \). Let \( \mathcal{D} = \sigma(\{B_1, \ldots, B_N\}) \). An element of the \( \sigma \)-field \( D \) is a finite disjoint union of the \( \{B_i\} \).

Claim: \( E(X|\mathcal{D})(\omega) = \sum_{i=1}^{N} E(X|B_i)1_{B_i}(\omega) \).

Proof. Let \( Z(\omega) = \sum_{i=1}^{N} E(X|B_i)1_{B_i}(\omega) \). Since \( B_i \in \mathcal{D} \), then \( Z \) is a \( \mathcal{D} \)-measurable function. It remains to verify property (ii) of conditional expectation. To begin, take one of the \( B_j \) s.

Then

\[
\int_{B_j} Z dP = E(X|B_j)P(B_j) = \int_{B_j} E(X|B_j)dP = \frac{P(B_j)}{P(B)} \int_{B_j} X dP = \int_{B_j} X dP
\]

by definition of conditioning of an event of positive probability. Now recall that the elements of \( \mathcal{D} \) are finite disjoint unions of the \( B_j \). Let \( D = \bigcup_{k=1}^{m} B_{ik} \). Then by linearity and the previous observation:

\[
\int_{D} Z dP = \sum_{k=1}^{m} \int_{B_{ik}} Z dP = \sum_{k=1}^{m} \int_{B_{ik}} X dP = \int_{D} X dP
\]

By uniqueness of conditional expectations, then \( E(X|\mathcal{D}) = Z \) almost surely. \( \square \)

- Suppose \( X \) is a \( \mathcal{D} \)-measurable function. Then here, we “know everything”, and then \( E(X|\mathcal{D})(\omega) = X(\omega) \) almost surely. Property i is verified by assumption, and ii follows trivially.

- Suppose random variables \( D \) and \( X \) are independent (which occurs if sigma fields \( \sigma(D), \sigma(X) \) are.) Here we “know nothing”.

Claim: \( E(X|D) = E(X) \) almost surely.

2
Proposition 4.3. Let \( X_1, X_2 \) be integrable random variables and \( D \subset \sigma \). Then \( \sigma \subset \mathcal{F} \) be sub-\( \sigma \)-fields.

a. For all \( a_1, a_2 \in \mathbb{R} \): \( E(a_1X_1 + a_2X_2|D) = a_1E(X_1|D) + a_2E(X_2|D) \) almost surely.

b. If \( X_1 \leq X_2 \) almost surely, then \( E(X_1|D) \leq E(X_2|D) \) almost surely.

c. \( E(E(X_1|\mathcal{E})|D) = E(X_1|D) \) almost surely.

Proof. 1. Suppose \( Z = a_1E(X_1|D) + a_2E(X_2|D) \). Then \( Z \) is \( D \)-measurable by linearity. Suppose \( D \in \mathcal{D} \). By linearity of the integral and since \( E(X_1|D), E(X_2|D) \) are conditional expectations, then:

\[
\int_D Z \, dP = \int_D E(X_1|D) \, dP + \int_D E(X_2|D) \, dP = a_1 \int_D X_1 \, dP + a_2 \int_D X_2 \, dP = \int_D (a_1X_1 + a_2X_2) \, dP
\]

means that \( Z = a_1E(X_1|D) + a_2E(X_2|D) \) almost surely.

2. Since \( X_2 - X_1 \geq 0 \) almost surely, then by a, it suffices to show that \( E(X_2 - X_1|D) \geq 0 \) almost surely. Let \( D_n = \{ \omega : E(X_2 - X_1|D) \leq -\frac{1}{n} \} \) which is in \( \mathcal{D} \). Then

\[
\int_{D_n} E(X_2 - X_1|D) \, dP = \int_{D_n} X_2 - X_1 \, dP \geq 0
\]

by assumption, means that \( P(D_n) = 0 \) for all \( n \). Hence \( P(\bigcup_n D_n) = 0 \) and hence \( E(X_2 - X_1|D) \geq 0 \) almost surely.

3. By definition \( Z = E(X_1|D) \). Furthermore, since sets which are \( D \)-measurable are also \( \mathcal{E} \)-measurable, then by the definition of conditional expectation and the previous observation:

\[
\int_D E(E(X|\mathcal{E})|D) \, dP = \int_D E(X|\mathcal{E}) \, dP = \int_D X \, dP = \int_D Z \, dP
\]

Proposition 4.4. Assume \( X, XZ \) are integrable random variables with \( Z \) \( D \)-measurable and \( D \subset \sigma \). Then:

a. \( E(ZX|D) = ZE(X|D) \)

b. \( \int_\Omega ZX \, dP = \int_\Omega ZE(X|D) \, dP \)
Proof. $a \rightarrow b$ is immediate since $\Omega$ is always in a sigma-field. To prove $a$, we apply the usual argument when developing Lebesgue integrals of functions. Firstly, suppose $Z = 1_{D'}$ where $D, D' \in \mathcal{D}$. Then since $D \cap D' \in \mathcal{D}$:

$$\int_D 1_{D'} X \, dP = \int_{D \cap D'} X \, dP = \int_{D \cap D'} E(X|\mathcal{D}) \, dP = \int_D 1_{D'} E(X|\mathcal{D}) \, dP$$

By linearity, $a$ holds when $Z$ is simple and $a$ holds for non-negative $Z$ by the monotone convergence theorem. For general $Z$, $a$ holds once we write $Z = Z^+ - Z^-$.

\[ \Box \]

**Note 4.5.** If $X_1 \leq X_2$ almost surely, then $E(X_1|\mathcal{D}) \leq E(X_2|\mathcal{D})$ almost surely. Hence if $X_1 = X_2$ almost surely, then $E(X_1|\mathcal{D}) = E(X_2|\mathcal{D})$ almost surely.

The convergence theorems for random variables apply for conditional expectations.

**Theorem 4.6.** Let $\mathcal{D} \subset^\sigma F$. Then:

a (MCT) Let $X_n$ be non-negative random variables increasing to an r.v. $X$ with $E(X) < \infty$. Then $E(X_n|\mathcal{D}) \uparrow E(X|\mathcal{D})$ almost surely.

b (Fatou’s Lemma) Let $X_n \geq 0$ be integrable, and $\liminf X_n$ be integrable. Then $E(\liminf X_n|\mathcal{D}) \leq \liminf E(X_n|\mathcal{D})$ almost surely.

c (DCT) Assume $X_n \rightarrow X$ almost surely, where $|X_n| \leq Y$ and $E(Y) < \infty$ for some random variable $Y$. Then $E(X_n|\mathcal{D}) \rightarrow E(X|\mathcal{D})$ almost surely.

Proof. a We have, almost surely, that $E(X|\mathcal{D}) \geq E(X_{n+1}|\mathcal{D}) \geq E(X_n|\mathcal{D})$ for all $n$. Let $D'$ be the set of $\omega$ for which this property holds ($P(D') = 1$), and $U(\omega) = \lim_{n \rightarrow \infty} E(X_n|\mathcal{D})(\omega)1_{D'}(\omega)$. Each $E(X_n|\mathcal{D})1_{D'}(\omega)$ is also a conditional expectation since $E(X_n|\mathcal{D})1_{D'}(\omega) = E(X_n|\mathcal{D})$ almost surely. Next, $E(X_n|\mathcal{D}) \uparrow U$ pointwise by definition, means that for $D \in \mathcal{D}$ and by the monotone convergence theorem:

$$\int_D E(X_n|\mathcal{D}) \, dP = \int_D X_n \, dP \rightarrow \int_D U \, dP = \int_D X \, dP$$

Hence, $U = E(X|\mathcal{D})$ almost surely, implies that $E(X_n|\mathcal{D}) \uparrow E(X|\mathcal{D})$ almost surely.

b The proofs of Fatou’s Lemma and the Dominated Convergence Theorem are analogous to those of the classical ones.

\[ \Box \]

Now, we will examine what it means to condition on a random variable. Recall $E(X|Y)(\omega) = E(X|\sigma(Y))(\omega)$ where $X, Y$ are random variables. Recall that $\sigma(Y) = \{Y^{-1}(B) : B \subset S\}$ if $Y$ takes values in $(S, S)$.

**Proposition 4.7.** Let $Y : (\Omega, F) \rightarrow (S, S)$ be a random vector. Then, a random variable $Z$ is $\sigma(Y)$-measurable iff $Z = \phi(Y)$ for some measurable function $\phi : (S, S) \rightarrow (\mathbb{R}, B(\mathbb{R}))$.

Proof. ($\leftarrow$) is trivial since $Y$ is $\sigma(Y)$-measurable.

For ($\rightarrow$), let $D \in \sigma(Y)$. Then $D = Y^{-1}(B)$ for some $B \in S$. Hence, $1_D = 1_B(Y)$. We will construct $\phi$ using this observation.

Firstly, suppose $Z$ be $\sigma(Y)$-measurable and simple. That is, let $Z = \sum_{i=1}^N \alpha_i 1_{B_i}$. By observation, $Z = \sum_{i=1}^N \alpha_i 1_{B_i}(Y) = \phi(Y)$, where $\phi(y) = \sum_{i=1}^N \alpha_i 1_{B_i}(y)$. 

4
Now let $Z \geq 0$. Take $\{Z_n\}$ as a sequence of discrete skeletons converging to $Z$, where each $Z_n$ are $\sigma(Y)$-measurable. By the previous case, there is a $\phi_n$ such that $Z_n = \phi_n(Y)$ for each $n$. Redefine, $\tilde{Z}_n = \max_{k \leq n} Z_k = \max_{k \leq n} \phi_k(Y) = \tilde{\phi}_n(Y)$. Then $\tilde{\phi}_n$ increase pointwise to a measurable map $\tilde{\phi}$ satisfying $Z = \tilde{\phi}(Y)$. Now, $\tilde{\phi}$ by take infinite values on a measure zero set so we get rid of those by redefining, $\phi(x) = \tilde{\phi}1_{\mathbb{R}}$ which also makes $Z = \phi(Y)$.

For the general case, let $Z \in \mathbb{R}$. Separate into positive and negative parts by $Z = Z^+-Z^-$. Setting $\phi(x) = \phi_+(x) - \phi_-(x)$ makes $Z = \phi(Y)$.

**Note 4.7.** If $\phi, \tilde{\phi} : (S, \mathcal{S}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ are two such maps satisfying $\phi(Y) = Z, \tilde{\phi}(Y) = Z$, then $\phi = \tilde{\phi}$ $P_Y$-almost surely. That is, $P_Y(\phi \neq \tilde{\phi}) = P(\phi(Y) \neq \tilde{\phi}(Y)) = 0$.

**Definition 3.** Let $Y : (\Omega, \mathcal{F}) \to (S, \mathcal{S})$ and $X$ be an integrable random variable. Then $h(y) = E(X|Y = y)$ is the unique, up to $P_Y$-null sets map such that $h : (S, \mathcal{S}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ satisfies $h(Y) = E(X|Y)$.

Call $h(y)$, the conditional expectation of $X$ given $\{Y = y\}$. When $Y$ is a random variable with a probability density function function, we call $h(y)$ the **conditional density**. For instance, if $X, Y$ are real valued, then we claim that:

$$E(X|Y = y) = \lim_{\epsilon \to 0} \frac{E(X1_{Y \in [y,y+\epsilon)})}{P(Y \in [y, y+\epsilon])}$$

for almost every $y$.

This was proved in 419.