Conditional Probability and Expectation
A Review
January 7, 2003

Notation: Let \((\Omega, \mathcal{F}, \mathbb{P})\) be a probability space and let \(\mathcal{G} \subseteq \mathcal{F}\) be a sub sigma-algebra of \(\mathcal{F}\).

Existence Theorem. If \(X : \Omega \to \mathbb{R}\) is any non-negative or integrable random variable, then there is an essentially unique \(\mathcal{G}\)-measurable random variable \(Z : \Omega \to \mathbb{R}\) for which

\[
\mathbb{P}(1_A X) = \mathbb{P}(1_A Z) \tag{\ast}
\]

for all \(A \in \mathcal{G}\). If \(X\) is integrable, then \(Z\) may be chosen to have finite values.

Proof. Omitted

Any \(\mathcal{G}\)-measurable \(Z\) for which \(\ast\) holds is called a version of the conditional expectation of \(X\) given \(\mathcal{G}\) and denote by

\[
Z = \mathbb{P}(X|\mathcal{G}) \text{ a.e.}
\]

Remarks 1. Relation (\ast) is called the defining relation. In conventional notation, it is

\[
\int_A Zd\mathbb{P} = \int_A Xd\mathbb{P} \tag{\ast}
\]

for \(A \in \mathcal{G}\).

2. In applications, one has often to guess the answer; \(\ast\) and the measurability condition then provide a method for checking.

Examples 1. If \(X\) is \(\mathcal{G}\)-measurable, then \(Z = X\).

2. If \(X\) and \(\mathcal{G}\) are independent, then \(Z = \mathbb{P}(X)\) is constant. For any constant is \(\mathcal{G}\)-measurable and

\[
\mathbb{P}(1_A X) = \mathbb{P}(A)\mathbb{P}(X) = \mathbb{P}(1_A Z),
\]

verifying \(\ast\).

3. Let \(\{G_1, \cdots, G_m\}\) be a measurable partition for which \(\mathbb{P}(G_i) > 0\) for all \(i = 1, \cdots, m\); and let \(\mathcal{G} = \sigma\{G_1, \cdots, G_m\}\). If \(X\) is measurable and either non-negative or integrable, let

\[
Z = \sum_{i=1}^{m} \frac{\mathbb{P}(1_{G_i} X)}{\mathbb{P}(G_i)} 1_{G_i}.
\]

Then \(Z\) is \(\mathcal{G}\)-measurable. To see that \(\ast\) hold, let \(H \in \mathcal{G}\). Then there is a subset \(J \subseteq \{1, \cdots, m\}\) for which \(H = \cup_{j \in J} G_j\) in which case

\[
\mathbb{P}(1_H Z) = \sum_{i=1}^{m} \frac{\mathbb{P}(1_{G_i} X)}{\mathbb{P}(G_i)} \mathbb{P}(G_i \cap H) = \cdots = \sum_{i \in J} \mathbb{P}(1_{G_i} X) = \mathbb{P}(1_H X).
\]
4. Let \((\Theta, B)\) be a second measurable space; let \(T : \Omega \to \Theta\) be a measurable transformation; and let \(e : \Theta \to \mathbb{R}\) be the conditional expectation of \(X\) given \(T\). Then

\[ Z = e \circ T \]

is the conditional expectation of \(X\) given \(\mathcal{G} = \sigma(T)\).

**Algebraic Properties.**

a) **Monotonicity.** If \(\mathbb{P}X_1\) and \(\mathbb{P}X_2\) are defined and \(X_1 \leq X_2\) a.e., then \(\mathbb{P}(X_1|\mathcal{G}) \leq \mathbb{P}(X_2|\mathcal{G})\) a.e.

b) **Linearity.** If \(X_1\) and \(X_2\) are integrable and \(\alpha_1, \alpha_2 \in \mathbb{R}\), then

\[ \mathbb{P}(\alpha_1 X_1 + \alpha_2 X_2|\mathcal{G}) = \alpha_1 \mathbb{P}(X_1|\mathcal{G}) + \alpha_2 \mathbb{P}(X_2|\mathcal{G}) \text{ a.e.} \]

*Proof.* To see b), for example, let \(Z_i = \mathbb{P}(X_i|\mathcal{G})\), \(i = 1, 2\) and \(Z = \alpha_1 Z_1 + \alpha_2 Z_2\). Then \(Z\) is \(\mathcal{G}\)-measurable; and if \(A \in \mathcal{G}\), then

\[ \mathbb{P}(1_A Z) = \alpha_1 \mathbb{P}(1_A Z_1) + \alpha_2 \mathbb{P}(1_A Z_2) = \alpha_1 \mathbb{P}(1_A X_1) + \alpha_2 \mathbb{P}(1_A X_2) = \mathbb{P}(1_A X). \]

So, \(Z\) satisfies the defining relation.

**Convex Functions.** A function \(\psi : \mathbb{R} \to \mathbb{R}\) is convex iff

\[ \psi[\alpha x + (1 - \alpha) y] \leq \alpha \psi(x) + (1 - \alpha) \psi(y) \]

for all \(x, y \in \mathbb{R}\) and \(0 \leq \alpha \leq 1\). In this case, there is a countable collection \((a_i, b_i), i = 1, 2, \cdots\) for which

\[ \psi(x) = \sup_i a_i x + b_i \]

for all \(x \in \mathbb{R}\). See Appendix C.

A sufficient condition for \(\psi\) to be convex is that \(\psi''(x) \geq 0\) for all \(x\).

**Examples:** \(\psi(x) = x^2\) and \(\psi(x) = |x|\)

**Jensen’s Inequality.** If \(X\) is integrable, \(\psi\) is convex, and \(\psi \circ X\) is either non-negative or convex, then \(\mathbb{P}[\psi(X)|\mathcal{G}] \geq \psi[\mathbb{P}(X|\mathcal{G})] \text{ a.e. for any sub sigma-algebra } \mathcal{G}\).

*Proof.* Let \(a_i\) and \(b_i\) be as in (†). Then

\[ \mathbb{P}(\psi \circ X|\mathcal{G}) \geq \mathbb{P}(a_i X_1 + b_i|\mathcal{G}) = a_i \mathbb{P}(X|\mathcal{G}) + b_i \text{ a.e.} \]

for all \(i\). So,

\[ \mathbb{P}(\psi \circ X|\mathcal{G}) \geq \sup_i a_i \mathbb{P}(X|\mathcal{G}) + b_i = \psi[\mathbb{P}(X|\mathcal{G})], \text{ a.e.} \]

as asserted.

*Example:* \(\mathbb{P}(X|\mathcal{G}) \leq \mathbb{P}(X|\mathcal{G})\).

**Smoothing Properties:** Let \(\mathcal{G}_1 \subseteq \mathcal{G}_2\) be sub sigma-algebras of \(\mathcal{F}\) and let \(X\) be either non-negative or integrable. Then

\[ \mathbb{P}[\mathbb{P}(X|\mathcal{G}_1)|\mathcal{G}_2] = \mathbb{P}(X|\mathcal{G}_1) = \mathbb{P}[\mathbb{P}(X|\mathcal{G}_2)|\mathcal{G}_1] \text{ a.e.} \]
Proof. Let \( Z_i = \mathbb{P}(X|G_i), \ i = 1, 2. \) For the first equality, \( Z_1 \) is \( G_2 \)-measurable and, therefore, \( \mathbb{P}(Z_1|G_2) = Z_1. \) For the second, let \( A \in G_1. \) Then \( A \in G_2 \) and, therefore,
\[
\mathbb{P}(1_A Z_2) = \mathbb{P}(1_A X) = \mathbb{P}(1_A Z_1).
\]
That is, \( \mathbb{P}(Z_2|G_1) = Z_1. \)

The Convergence Theorems: Let \( X, X_1, X_2, \ldots \) be measurable and let \( G \) be a sub sigma-algebra.

The Monotone Convergence Theorem: If \( 0 \leq X_1 \leq X_2 \leq \cdots \uparrow X \) a.e., then
\[
\lim_{n \to \infty} \mathbb{P}(X_n|G) = \mathbb{P}(X|G) \text{ a.e.}
\]

Fatou’s Lemma. If \( X_1, X_2, \ldots \) are non-negative, then
\[
\mathbb{P}(\lim \inf_{n \to \infty} X_n|G) \leq \lim \inf_{n \to \infty} \mathbb{P}(X_n|G) \text{ a.e.}
\]

The Dominated Convergence Theorem. If \( \lim_{n \to \infty} X_n = X \) a.e. and \( \sup_n |X_n| \) is integrable, then
\[
\lim_{n \to \infty} \mathbb{P}(X_n|G) = \mathbb{P}(X|G) \text{ a.e.}
\]

Proofs. For \( a) \), let \( Z_n = \mathbb{P}(X_n|G). \) Then \( 0 \leq Z_1 \leq Z_2 \leq \cdots \leq a.e. \) by monotonicity. So \( Z := \lim_{n \to \infty} Z_n \) exists a.e and is \( G \)-measurable; and if \( A \in G \), then
\[
\mathbb{P}(AX) = \lim_{n \to \infty} \mathbb{P}(AX_n) = \lim_{n \to \infty} \mathbb{P}(AZ_n) = \mathbb{P}(AZ),
\]
using the Monotone Convergence Theorem for unconditional expectations (twice). So, \( Z = \mathbb{P}(X|G). \)

For \( b) \), let \( Y_m = \inf_{n \geq m} X_n \) for \( m = 1, 2, \ldots \) and \( Y = \lim_{m \to \infty} Y_m = \lim \inf_{n \to \infty} X_n. \) If \( n \geq m \), then \( \mathbb{P}(X_n|G) \geq \mathbb{P}(Y_m|G) \text{ a.e.} \) So,
\[
\inf_{n \geq m} \mathbb{P}(X_n|G) \geq \mathbb{P}(Y_m|G)
\]
and, therefore,
\[
\lim \inf_{n \to \infty} \mathbb{P}(X_n|G) \geq \lim_{m \to \infty} \mathbb{P}(Y_m|G) = \mathbb{P}(Y|G),
\]
by \( a) \).

For \( c) \), let \( Y_n = |X_n - X|. \) Then
\[
|\mathbb{P}(X_n|G) - \mathbb{P}(X|G)| \leq |\mathbb{P}(X_n - X|G)| \leq \mathbb{P}(Y_n|G).
\]
So, it suffices to show that \( \lim_{n \to \infty} \mathbb{P}(Y_n|G) = 0. \) Let \( W = \sup_n Y_n. \) Then \( \mathbb{P}(W) \leq 2\mathbb{P}(\sup_n |X_n|) < \infty. \) Now \( W - Y_n \geq 0 \) and \( W - Y_n \to W \text{ a.e.} \) So,
\[
\mathbb{P}(W|G) = \mathbb{P}(\lim_{n \to \infty} W - Y_n|G)
\leq \lim \inf_{n \to \infty} \mathbb{P}(W - Y_n|G)
= \mathbb{P}(W|G) - \lim \sup_{n \to \infty} \mathbb{P}(Y_n|G) \text{ a.e.}
\]
and, therefore,
\[
\lim \sup_{n \to \infty} \mathbb{P}(Y_n|G) = 0 \text{ a.e.}
\]
Corollary. If $X$ is non-negative, then

$$IP(WX|\mathcal{G}) = WIP(X|\mathcal{G})$$

for every non-negative $\mathcal{G}$-measurable $W$. If $X$ is integrable, then (!) holds for all $\mathcal{G}$-measurable $W$ for which $IP(|WX|) < \infty$.

Proof. Approximate by simple functions.

More on (*). Here is a sufficient condition for (*).

Problem 1. Let $\mathcal{G}_0$ is a $\pi$-system for which $\Omega \in \mathcal{G}_0$ and suppose that $\mathcal{G} = \sigma(\mathcal{G}_0)$. If $X$ and $Z$ integrable random variable for which (*) holds for all $A \in \mathcal{G}_0$, then (*) holds for all $A \in \mathcal{G}$

Conditional Independence. Sigma-algebras $\mathcal{G}_1$ and $\mathcal{G}_2$ are said to be conditionally independent given a third sigma-algebra $\mathcal{H}$ if

$$IP(G_1 \cap G_2|\mathcal{H}) = IP(G_1|\mathcal{H})IP(G_2|\mathcal{H}) \ a.e.$$ 

for every choice of $G_1 \in \mathcal{G}_1$ and $G_2 \in \mathcal{G}_2$.

Joints. If $\mathcal{G}_1$ and $\mathcal{G}_2$ are sigma-algebras, then

$$\mathcal{G}_1 \vee \mathcal{G}_2 = \sigma\{G_1 \cap G_2 : G_1 \in \mathcal{G}_1, G_2 \in \mathcal{G}_2\}.$$ 

Problem 2. Show that $\mathcal{G}_1$ and $\mathcal{G}_2$ are conditionally independent given $\mathcal{H}$ iff

$$P(G_2|\mathcal{G}_1 \vee \mathcal{H}) = P(G_2|\mathcal{H}) \ a.e.$$ 

for every $G_2 \in \mathcal{G}_2$. 

4