1 Definition of Conditional Expectation

1.1 General definition

Recall the definition of conditional probability associated with Bayes' Rule

$$\mathbb{P}(A|B) \equiv \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

For a discrete random variable X we have

$$\mathbb{P}(A) = \sum_x \mathbb{P}(A, X = x) = \sum_x \mathbb{P}(A|X = x)\mathbb{P}(X = x)$$

and the resulting formula for conditional expectation

$$\begin{aligned}
\mathbb{E}(Y|X = x) &= \int_\Omega Y(\omega)\mathbb{P}(dw|X = x) \\
&= \frac{\int_{X=x} Y(\omega)\mathbb{P}(dw)}{\mathbb{P}(X = x)} \\
&= \frac{\mathbb{E}(Y\mathbf{1}_{(X=x)})}{\mathbb{P}(X = x)}
\end{aligned}$$

We would like to extend this to handle more general situations where densities don't exist or we want to condition on very "complicated" sets.

Definition 1 *Given a random variable Y with $\mathbb{E}|Y| < \infty$ defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and some sub-σ-field $\mathcal{G} \subset \mathcal{A}$ we will define the* **conditional expectation** *as the almost surely unique random variable $\mathbb{E}(Y|\mathcal{G})$ which satisfies the following two conditions*

1. *$\mathbb{E}(Y|\mathcal{G})$ is $\mathcal{G}$-measurable*
2. *$\mathbb{E}(YZ) = \mathbb{E}(\mathbb{E}(Y|\mathcal{G})Z)$ for all Z which are bounded and $\mathcal{G}$-measurable*

Remark: one could replace 2. in the previous definition with:

$$\forall G \in \mathcal{G}, \quad \mathbb{E}(Y\mathbf{1}_G) = \mathbb{E}(\mathbb{E}(Y|\mathcal{G})\mathbf{1}_G).$$

Proof of existence and unicity

- **Existence** Using linearity, we need only consider $X \geq 0$. Define a measure Q on $\mathcal{F}$ by $Q(A) = \mathbb{E}[X\mathbf{1}_A]$ for $A \in \mathcal{F}$. This is trivially absolutely continuous with respect to $P_{|\mathcal{F}}$, the restriction of P to F. Let $\mathbb{E}[X|\mathcal{F}]$ be the Radon-Nikodym derivative of Q with respect to $P_{|\mathcal{F}}$. The Radon-Nikodym derivative is $\mathcal{F}$-measurable by construction and so provides the desired random variable.

- **Unicity:** If Y_1, Y_2 are two $\mathcal{F}$-measurable random variables with $\mathbb{E}[Y_1\mathbf{1}_A] = \mathbb{E}[Y_2\mathbf{1}_A]$ for all $A \in \mathcal{F}$, then $Y_1 = Y_2$, a.s., or conditional expectation is unique up to a.s. equivalence.

For $\mathcal{G} = \sigma(X)$ when X is a discrete variable, the space Ω is simply partitioned into disjoint sets $\Omega = \sqcup G_n$. Our definition for the discrete case gives

$$\begin{aligned}
\mathbb{E}(Y|\sigma(X)) &= \mathbb{E}(Y|X) \\
&= \sum_n \frac{\mathbb{E}(Y\mathbf{1}_{X=x_n})}{\mathbb{P}(X=x_n)}\mathbf{1}_{X=x_n} \\
&= \sum_n \frac{\mathbb{E}(Y\mathbf{1}_{G_n})}{\mathbb{P}(G_n)}\mathbf{1}_{G_n}
\end{aligned}$$

which is clearly $\mathcal{G}$-measurable. In general for $\mathcal{G} = \sigma(X)$:

Definition 2 Conditional expectation of Y given X
Let $(\Omega, \mathcal{A}, P)$ be a probability space, $Y \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$ and X another random variable defined on $(\Omega, \mathcal{A}, P)$. Define then $E(Y \mid X)$ the conditional expectation of Y given X as $E(Y \mid \sigma(X))$.

Proposition 3 *Let $(\Omega, \mathcal{A})$ be a measurable space,*

$$Y \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$$

and X another real-valued random variable defined on $(\Omega, \mathcal{A}, P)$. As $X = f(Y)$, where f is measurable, real-valued function if and only if $\sigma(X) \subset \sigma(Y)$, we get that $E(Y \mid X)$ is a measurable function of X.

Proposition 4 *Let $(\Omega, \mathcal{A}, P)$ be a probability space, and X and Y two independent random variables such that Y is P-integrable. Then $E(Y \mid X) = E(Y)$, P-almost surely.*

Do not mix this notion with the following:

1.2 Couples of random variables with p.d.f.

Proposition 5 *Let (X, Y) be a couple of real-valued random variables with p.d.f. $f_{X,Y}(x, y)$ w.r.t. the Lebesgue measure on $\mathbb{R}^2$. Denote the respective marginal p.d.f. of X and Y as $f_X(x)$ and $f_Y(y)$. Consider $f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$. Then almost surely*

$$\forall C \in \mathcal{B}, P(X \in C \mid Y = y) = \int_C f_{X|Y}(x \mid y)dx.$$

If besides X is P-integrable, then

$$E(X \mid Y = y) = \int_{\mathbb{R}} x f_{X|Y}(x \mid y)dx.$$

If $g : \mathbb{R}^2 \to \mathbb{R}$ is a measurable function such that $g(X, Y)$ is integrable, then

$$E(g(X, Y) \mid Y = y) = \int_{\mathbb{R}} g(x, y) f_{X|Y}(x \mid y)dx.$$

Remarks: As soon as $f_Y(y) > 0$, this defines the distribution of X given that $Y = y$, described by p.d.f $f_{X|Y}(x \mid y)$, which is nonnegative and of integral 1.
If X and Y are independent, $f_{X|Y} = f_X$ and $f_{Y|X} = f_Y$. To make the link with $\mathbb{E}[X|Y]$ would require to introduce the concept of regular conditional distribution.
Equation (5) may be useful to compute the mathematical expectation of $g(X, Y)$ as

$$E(g(X, Y)) = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} g(x, y) f_{X|Y}(x \mid y)dx \right) f_Y(y)dy.$$

2 Properties of Conditional Expectation

2.1 Conditional expectation

$\mathbb{E}(\cdot|\mathcal{G})$ may be seen as an operator on random variables that transforms $\mathcal{A}$-measurable variables into $\mathcal{G}$-measurable ones.

Let us recall the basic properties of conditional expectation:

1. $\mathbb{E}(\cdot|\mathcal{G})$ is positive:
$$Y \geq 0 \rightarrow \mathbb{E}(Y|\mathcal{G}) \geq 0)$$

2. $\mathbb{E}(\cdot|\mathcal{G})$ is linear:
$$\mathbb{E}(aX + bY|\mathcal{G}) = a\mathbb{E}(X|\mathcal{G}) + b\mathbb{E}(Y|\mathcal{G})$$

3. $\mathbb{E}(\cdot|\mathcal{G})$ is a projection:
$$\mathbb{E}(\mathbb{E}(X|\mathcal{G})|\mathcal{G}) = \mathbb{E}(X|\mathcal{G})$$

4. More generally, the "tower property". If $\mathcal{H} \subset \mathcal{G}$ then
$$\mathbb{E}(\mathbb{E}(X|\mathcal{G})|\mathcal{H}) = \mathbb{E}(X|\mathcal{H}) = \mathbb{E}(\mathbb{E}(X|\mathcal{H}) \mid \mathcal{G})$$
Proof: The right equality holds because $\mathbb{E}[X|\mathcal{H}]$ is $\mathcal{H}$- measurable, hence $\mathcal{G}$-measurable. To show the left equality, let $A \in \mathcal{H}$. Then since A is also in $\mathcal{G}$,
$$\mathbb{E}[\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}]\mathbf{1}_A] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]\mathbf{1}_A] = \mathbb{E}[X\mathbf{1}_A] = \mathbb{E}[\mathbb{E}[X|\mathcal{H}]\mathbf{1}_A].$$
Since both sides are $\mathcal{H}$- measurable, the equality follows.

5. $\mathbb{E}(\cdot|\mathcal{G})$ commutes with multiplication by $\mathcal{G}$-measurable variables:
$$\mathbb{E}(XY|\mathcal{G}) = \mathbb{E}(X|\mathcal{G})Y \text{ for } \mathbb{E}|XY| < \infty \text{ and } Y\mathcal{G}\text{measurable}$$
Proof: If $A \in \mathcal{G}$, then for any $B \in \mathcal{G}$,
$$\mathbb{E}[\mathbf{1}_A\mathbb{E}[X|\mathcal{G}]\mathbf{1}_B] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]\mathbf{1}_{A\cap B}] = \mathbb{E}[X\mathbf{1}_{A\cap B}] = \mathbb{E}[(\mathbf{1}_A X)\mathbf{1}_B].$$
Since $\mathbf{1}_A\mathbb{E}[X|\mathcal{G}]$ is $\mathcal{G}$-measurable, this shows that the required equality holds when $Y = \mathbf{1}_A$ and $A \in \mathcal{G}$. Using linearity and taking limits shows that the equality holds whenever Y is $\mathcal{G}$-measurable and X and XY are integrable.

6. $\mathbb{E}(\cdot|\mathcal{G})$ respects monotone convergence:

$$0 \leq X_n \uparrow X \implies \mathbb{E}(X_n|\mathcal{G}) \uparrow \mathbb{E}(X|\mathcal{G})$$

7. If ϕ is convex (in particular if $\phi(x) = x^2$) and $\mathbb{E}|\phi(X)| < \infty$ then a conditional form of Jensen's inequality holds:

$$\phi(\mathbb{E}(X|\mathcal{G}) \leq \mathbb{E}(\phi(X)|\mathcal{G})$$

8. $\mathbb{E}(\cdot|\mathcal{G})$ is a continuous contraction of $\mathbf{L}^p$ for $p \geq 1$:

$$\|\mathbb{E}(X|\mathcal{G})\|_p \leq \|X\|_p$$

and

$$X_n \xrightarrow{\mathbf{L}^2} X \text{ implies } \mathbb{E}(X_n|\mathcal{G}) \xrightarrow{\mathbf{L}^2} \mathbb{E}(X|\mathcal{G})$$

9. Repeated Conditioning. For $\mathcal{G}_0 \subset \mathcal{G}_1 \subset \ldots$, $\mathcal{G}_\infty = \sigma(\cup\mathcal{G}_i)$, and $X \in \mathbf{L}^p$ with $p \geq 1$ then

$$\mathbb{E}(X|\mathcal{G}_n) \xrightarrow{a.s.} \mathbb{E}(X|\mathcal{G}_\infty)$$
$$\mathbb{E}(X|\mathcal{G}_n) \xrightarrow{\mathbf{L}^p} \mathbb{E}(X|\mathcal{G}_\infty)$$

10. Best approximation property:
Suppose that the random variable X is square-integrable, but not measurable with respect to $\mathcal{G}$. That is, the information in $\mathcal{G}$ does not completely determine the values of X. The conditional expectation, $Y = E[X \mid \mathcal{G}]$, has the property that it is the best approximation to X among functions measurable with respect to Y, in the least squares sense. That is, if $\tilde{Y}$ is $\mathcal{G}$-measurable, then

$$\mathbb{E}\left[(\tilde{Y} - X)^2\right] \geq \mathbb{E}\left[(Y - X)^2\right] .$$

It thus realizes the orthogonal projection of X onto a convex closed subset of a Hilbert space. This predicts the variance decomposition theorem that we shall see in a further section.

2.2 Conditional variance

Definition 6 *Let X be a square-integrable, real-valued random variable defined on a probability space $(\Omega, \mathcal{A}, P)$, and let $\mathcal{F}$ be a sub-σ-algebra of $\mathcal{A}$. Define the* **conditional variance of** X **given** $\mathcal{F}$ *(denoted by $Var(X \mid \mathcal{F})$) as the random variable $E((X - E(X \mid \mathcal{F}))^2 \mid \mathcal{F})$.*
Define also the conditional variance of X given a real-valued random variable Y defined on $(\Omega, \mathcal{A}, P)$ (denoted by $Var(X \mid Y)$) as the random variable $E((X - E(X \mid Y))^2 \mid Y)$.

Proposition 7 *$Var(X \mid \mathcal{F})$ and $Var(X \mid Y)$ are well- defined, almost surely nonnegative and finite.*

$$Var(X \mid \mathcal{F}) = E(X^2 \mid \mathcal{F}) - E(X \mid \mathcal{F})^2,$$

and

$$Var(X \mid Y) = E(X^2 \mid Y) - E(X \mid Y)^2.$$

Proposition 8 Variance decomposition formula
Let (X, Y) be a couple of random variables defined on a probability space $(\Omega, \mathcal{A}, P)$, such that X is square-integrable. Then

$$Var(X) = E(Var(X \mid Y)) + Var(E(X \mid Y)).$$

This may be very useful in non-life insurance to find the variance of a compound distribution.

Proof:

- $\mathrm{Var}(X \mid Y) = E(X^2 \mid Y) - (E(X \mid Y))^2$.
- $E[\mathrm{Var}(X \mid Y)] = E[E(X^2 \mid Y)] - E[(E(X \mid Y))^2]$.
- $E[E(X^2 \mid Y)] = E[X^2]$.
- $E[\mathrm{Var}(X \mid Y)] = E[X^2] - E[(E(X \mid Y))^2]$.
- $\mathrm{Var}(E(X \mid Y)) = E[(E(X \mid Y))^2] - (E[E(X \mid Y)])^2$.
- $E[E(X \mid Y)] = E[X]$.
- Hence $\mathrm{Var}(E(X \mid Y)) = E[(E(X \mid Y))^2] - (E[X])^2$.