36-752 Advanced Probability Overview **Spring 2018**

Lecture 10: March 1

Lecturer: Alessandro Rinaldo *Scribes: Wanshan Li*

Note: *LaTeX template courtesy of UC Berkeley EECS dept.*

10.1 Conditional Expectation

Given a probability space $(\Omega, \mathcal{F}, P)$, let $\mathcal{C} \subseteq \mathcal{F}$ be a sub-σ-field of $\mathcal{F}$, and a fixed set $A \in \mathcal{F}$. Our goal is to define the conditional probability $P(A|\mathcal{C})$. The point is that, $\mathcal{C}$ provides us additional information, so $P(A|\mathcal{C})$ would be different from $P(A)$.

First we consider of a special case $C = \sigma(B_1, \cdots, B_n)$ where $\{B_1, \cdots, B_n\}$ is a partition of Ω. The additional information here is that, for any $\omega \in \Omega$, one knows whether $\omega \in B_k$ or not.

Define $f : \Omega \to \mathbb{R}$ by

$$f(\omega) = \begin{cases} \frac{P(A \cap B_k)}{P(B_k)}, & \text{if } \omega \in B_k \text{ and } P(B_k) > 0 \\ c_k, & \text{if } \omega \in B_k \text{ and } P(B_k) = 0, \end{cases} \tag{10.1}$$

where $c_k \in \mathbb{R}$ can be any constant. Now we define the conditional probability, as a real-valued function on Ω, by $\Pr(A|\mathcal{C})(\omega) = f(\omega)$. The following fact shows that our definition is reasonable in a way:

$$P(A \cap B_k) = P(A|B_k)P(B_k) = \int_{B_k} \Pr(A|\mathcal{C})(\omega) dP(\omega).$$

Now let $\mathcal{C}$ be a generic sub-σ-field of $\mathcal{F}$. We can create a measure ν on $(\Omega, \mathcal{C})$, given by

$$\nu(B) = P(A \cap B).$$

If we can find some $\mathcal{C}$-measurable function f, such that (Notice that P is originally defined on $(\Omega, \mathcal{F})$, but here we can treat it as a probability measure on $(\Omega, \mathcal{C})$ as $\mathcal{C} \subseteq \mathcal{F}$)

$$\nu(B) = P(A \cap B) = \int_B f(\omega) dP(\omega),$$

then we define function f to be the *conditional probability of A given* $\mathcal{C}$, and denote it as $f = \Pr(A|\mathcal{C})$. Thus by our definition

1) $\Pr(A|\mathcal{C})(\cdot)$ is $\mathcal{C}$-measurable.
2) $\forall B \in \mathcal{C}$, $\nu(B) = P(A \cap B) = \int_B \Pr(A|\mathcal{C})(\omega) dP(\omega)$.

Remark There exists many *versions* of $\Pr(A|\mathcal{C})(\cdot)$, but by property 2), these versions are equal to each other $a.s.[P]$.

Let $X(\omega) = \mathbb{1}_A(\omega)$, then we may want to write

$$\Pr(A|\mathcal{C}) = \mathbb{E}(X|\mathcal{C}).$$

By generalizing X from an indicator function to any random variable we can get the definition of the conditional expectation.

Definition 10.1. *Given a probability space $(\Omega, \mathcal{F}, P)$, let $\mathcal{C} \subseteq \mathcal{F}$ be a sub-σ-field of $\mathcal{F}$, and X an $\mathcal{F}/\mathcal{B}$-measurable random variable with $\mathbb{E}|X| < \infty$. The conditional expectation of X given $\mathcal{C}$ is any real valued function $h : \Omega \to \mathbb{R}$, such that*

1) h is $\mathcal{C}$-measurable.

2) $\int_B h(\omega)dP(\omega) = \int_B X(\omega)dP(\omega)$, $\forall B \in \mathcal{C}$.

h is denoted as $\mathbb{E}[X|\mathcal{C}]$.

Remark

- $f = \mathbb{E}[X|\mathcal{C}]$ means f is a version of $\mathbb{E}[X|\mathcal{C}]$.
- By 2) in the definition, if h_1 and h_2 are two versions of $\mathbb{E}[X|\mathcal{C}]$, then $h_1(\omega) = h_2(\omega)$, $a.s.[P]$. Conversely, if h_1 is a version of $\mathbb{E}[X|\mathcal{C}]$ and $h_1(\omega) = h_2(\omega)$, $a.s.[P]$, then h_2 is also a version of $\mathbb{E}[X|\mathcal{C}]$.
- If $\mathcal{C} = \{\varnothing, \Omega\}$, then $\mathbb{E}[X|\mathcal{C}] = \mathbb{E}[X]$.
- If X itself is $\mathcal{C}/\mathcal{B}$ measurable, then $X = \mathbb{E}[X|\mathcal{C}]$.
- If $X = a,\ a.s.$, then $\mathbb{E}[X|\mathcal{C}] = a,\ a.s.$

10.1.1 Two perspectives

RN derivative One may ask "Does this function exist?". The answer is "Yes", and one can demonstrate this by using *RN derivative*. Assume $X \geq 0,\ a.s.$, the sketched proof is: define $\nu(B) = \int_B X(\omega)dP(\omega)$, then ν is a measure on $(\Omega, \mathcal{C})$. By the RN theorem, $\exists h$, which is $\mathcal{C}$-measurable and $\forall B \in \mathcal{C}$,

$$\nu(B) = \int_B h(\omega)dP(\omega).$$

Then by definition, the RN derivative h is the conditional expectation $\mathbb{E}[X|\mathcal{C}]$.

Projection An alternative perspective is to think of $\mathbb{E}[X|\mathcal{C}]$ as a "projection". Given a r.v. X on $(\Omega, \mathcal{F}, P)$ s.t. $\mathbb{E}X^2 < \infty$ and $\mathcal{C} \subseteq \mathcal{F}$. Consider $L^2(\Omega, \mathcal{C}, P)$, a Hilbert space of r.v.'s that are $\mathcal{C}$-measurable and L_2. Then one can show that, the $\mathcal{C}$-measurable random variable Z is the conditional expectation of X *if and only if* Z is the orthogonal projection of X onto $L^2(\Omega, \mathcal{C}, P)$, that is

$$\mathbb{E}[W(X - Z)] = 0,\ \forall W \in L^2(\Omega, \mathcal{C}, P),$$

or equivalently,

$$Z = \operatorname*{argmin}_{W \in L^2(\Omega, \mathcal{C}, P)} \mathbb{E}(X - W)^2.$$

Now from this perspective, if we let $\mathcal{C} = \sigma(Y)$ where Y is an r.v. on $(\Omega, \mathcal{F}, P)$, then by theorem 39 in notes,

$$\mathbb{E}[X|Y] \triangleq \mathbb{E}[X|\mathcal{C}] = \underset{\text{meas. function } g,\ s.t.\ \mathbb{E}[g(Y)]^2<\infty}{\text{argmin}} \mathbb{E}(X - g(Y))^2.$$

Recall that the usual machinery of defining $\mathbb{E}[X|Y]$ is

$$\mathbb{E}[X|Y] = g(Y), \text{ where } g(y) = \int_{\mathbb{R}} x f_{X|Y}(x,y)dy = \int_{\mathbb{R}} x \frac{f_{X,Y}(x,y)}{f_Y(y)} dy.$$

Example 10.2. *Let* $X_1, X_2 \overset{i.i.d.}{\sim}$ Unifrom$(0,1)$*, and* $Y = \max\{X_1, X_2\}$*,* $X = X_1$*. Then one version of* $\mathbb{E}[X|Y]$ *is* $h(Y) = \frac{3}{4}Y$*. In addition, another version can be given by*

$$h_1(Y) = \begin{cases} \frac{3}{4}Y, & \text{if } Y \text{ is irrational,} \\ 0, & \text{otherwise.} \end{cases}$$

10.1.2 Properties

Some basic properties of conditional expectation coincide with expectation, including

1) Linearity. If $\mathbb{E}[X]$, $\mathbb{E}[Y]$, and $\mathbb{E}[X+Y]$ all exist, then $\mathbb{E}[X|\mathcal{C}] + \mathbb{E}[Y|\mathcal{C}]$ is a version of $\mathbb{E}[X+Y|\mathcal{C}]$.
2) Monotonicity. If $X_1 \leq X_2$ *a.s.*, then $\mathbb{E}[X_1|\mathcal{C}] \leq \mathbb{E}[X_2|\mathcal{C}]$ *a.s.*
3) Jensen's inequality. Let $\mathbb{E}(X)$ be finite. If ϕ is a convex function and $\phi(X) \in L^1$, then $\mathbb{E}[\phi(X)|MC] \geq \phi(\mathbb{E}[X|\mathcal{C}])$ *a.s.*
4) Convergence theorems: monotone convergence theorem, dominant convergence theorem.

Theorem 10.3 (Convergence theorem)**.** *Let* $\mathcal{C}$ *be a sub-*σ*-field of* $\mathcal{F}$*.*

1) (Monotone) If $0 \leq X_n \leq X$ *a.s. for all* n *and* $X_n \to X$ *a.s., then* $\mathbb{E}[X_n|\mathcal{C}] \to \mathbb{E}[X|\mathcal{C}]$*.*

2) (Dominant) If $X_n \to X$ *a.s. and* $|X_n| \leq Y$ *a.s., where* $Y \in L^1$*, then* $\mathbb{E}[X_n|\mathcal{C}] \to \mathbb{E}[X|\mathcal{C}]$*.*

Proposition 10.4 (Tower property of conditional expectation)**.** *If sub-*σ*-fields* $\mathcal{C}_1 \subseteq \mathcal{C}_2 \subseteq \mathcal{F}$*, and* $\mathbb{E}|X| < \infty$*, then* $\mathbb{E}[X|\mathcal{C}_1]$ *is a version of* $\mathbb{E}[\mathbb{E}[X|C_2]|\mathcal{C}_1]$*. In particular,* $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|\mathcal{C}]]$ *(taking* $\mathcal{C} = \{\varnothing, \Omega\}$*).*

10.2 Regular Conditional Probability

Notice that $\Pr(\cdot|\mathcal{C})(\cdot)$ is a function defined on $\mathcal{F} \times \Omega$.

- By definition, for $A \in \mathcal{F}$, $\Pr(A|\mathcal{C})(\cdot)$ is a version of $\mathbb{E}[\mathbb{1}_A|\mathcal{C}](\cdot)$.
- We would like $\forall \omega \in \Omega$, $\Pr(\cdot|\mathcal{C})(\omega)$ to be a probability measure on $(\Omega, \mathcal{F})$.

It is easy to see that $\Pr(A|\mathcal{C})(\cdot) \in [0,1]$ *a.s.*$[P]$ as a function of ω on $(\Omega, \mathcal{C}, P)$. We can also prove that it is countably additive *a.e.*$[P]$:

Proposition 10.5. *If* $\{A_n\}_{n=1}^{\infty}$ *is a sequence of disjoint* $\mathcal{F}$*-measurable sets, then*

$$W(\omega) = \sum_{n=1}^{\infty} \Pr(A_n|\mathcal{C})(\omega)$$

is a version of $\Pr(\bigcup_{n=1}^{\infty} A_n|\mathcal{C})$.

This proposition means that **given** a sequence of disjoint $\mathcal{F}$-measurable sets $\{A_n\}_{n=1}^{\infty}$, for $[P]a.e.\ \omega$, we have

$$\sum_{n=1}^{\infty} \Pr(A_n|\mathcal{C})(\omega) = \Pr(\bigcup_{n=1}^{\infty} A_n|\mathcal{C})(\omega).$$

In general, however, for the collection of functions $\{\Pr(A|\mathcal{C})(\cdot)\} : A \in \mathcal{F}\}$ and a **given** $\omega \in \Omega$, $\Pr(\cdot|\mathcal{C})(\omega)$ is not necessarily countably additive, and therefore is not a probability measure. Even in the sense of $a.s.[P]$ (with respect to $\omega \in \Omega$), $\Pr(\cdot|\mathcal{C})(\omega)$ is not necessarily a probability measure.

The intuition is, for a given $\{A_n\}_{n=1}^{\infty}$ of disjoint $\mathcal{F}$-measurable sets, to make $\sum_{n=1}^{\infty} \Pr(A_n|\mathcal{C})(\omega) = \Pr(\bigcup_{n=1}^{\infty} A_n|\mathcal{C})(\omega)$ (in the sense of $a.s.[P]$), we can only allow $\Pr(\bigcup_{n=1}^{\infty} A_n|\mathcal{C})(\omega) \neq \sum_{n=1}^{\infty} \Pr(A_n|\mathcal{C})(\omega)$ for ω in a P-measure-0 set $N(\{A_n\}_{n=1}^{\infty}) \subseteq \Omega$.

Therefore, to make $\Pr(\cdot|\mathcal{C})(\omega)$ a probability measure (also in the sense of $a.s.[P]$), we want

$$P(N(\{A_n\}_{n=1}^{\infty})) = 0$$

to hold $a.s.[P]$ (w.r.t. ω) over all possible choices of sequence $\{A_n\}_{n=1}^{\infty}$. To ensure this, We need

$$P(\bigcup_{\{A_n\}} N(\{A_n\}_{n=1}^{\infty})) = 0.$$

However, since there are uncountably many sequences $\{A_n\}_{n=1}^{\infty}$, this may not necessarily hold. When this nontrivial property holds, we call $\Pr(\cdot|\mathcal{C})(\cdot) : \mathcal{A} \times \Omega \to [0,1]$ a regular conditional probability.

Definition 10.6 (Regular conditional probability)**.** *Given a probability space* $(\Omega, \mathcal{F}, P)$*. Let* $\mathcal{A} \subseteq \mathcal{F}$ *be a sub-*σ*-field. We say that the function* $\Pr(\cdot|\mathcal{C})(\cdot) : \mathcal{A} \times \Omega \to [0,1]$ *is a regular conditional probability (rcd) if*

1) $\forall A \in \mathcal{A}$, $\Pr(A|\mathcal{C})(\cdot)$ *is a version of* $\mathbb{E}[\mathbb{1}_A|\mathcal{C}]$.
2) *For* $[P]a.e.\ \omega \in \Omega$, $\Pr(\cdot|\mathcal{C})(\omega)$ *is a probability measure on* $(\Omega, \mathcal{A})$.

10.2.1 Regular Conditional Distribution

Let $\mathcal{A} = \sigma(X)$ for some r.v. X that is $\mathcal{F}/\mathcal{B}$ measurable. For each $B \in \mathcal{B}$, let

$$\mu_{X|\mathcal{C}}(B)(\omega) = \Pr(X^{-1}(B)|\mathcal{C})(\omega).$$

Then function $\mu_{X|\mathcal{C}}(\cdot|\mathcal{C})(\cdot) : \mathcal{B} \times \Omega \to [0,1]$ is called a regular conditional distribution of X given $\mathcal{C}$ when

1) $\forall B \in \mathcal{B}$, $\mu_{X|\mathcal{C}}(B|\mathcal{C})(\cdot)$ is a version of $\mathbb{E}[\mathbb{1}_{X \in B}|\mathcal{C}]$.
2) For $[P]a.e.\ \omega \in \Omega$, $\mu_{X|\mathcal{C}}(\cdot|\mathcal{C})(\omega)$ is a probability measure on $(\mathbb{R}, \mathcal{B})$.