The Radon Nikodym Theorem and Conditional Expectations

1 Background and Notation

In this section we discuss some of the important results of probability theory. A *probability space* is a triple $(\Omega, \mathcal{A}, P)$, where Ω is a sample space, $\mathcal{A}$ is a σ-field of distinguishable events, and P is a probability measure. A function $X{:}\Omega \rightarrow \mathbb{R}$ is said to be *measruable* if the inverse images or Borel sets are elements of $\mathcal{A}$, that is, $X^{-1}(B) \in \mathcal{A}\ \forall B \in \mathcal{B}(\mathbb{R})$. If this condition holds, then X is called a random variable. The *expectation* of X is given by

$$EX = \int_\Omega X(\omega) dP(\omega). \tag{1}$$

A *signed measure* is a function ϕ that obeys all of the properties of probability measures except that it may take positive or negative values and need not sum to unity; that is, ϕ possesses the following properties:

$$\begin{aligned} \phi(\emptyset) &= 0 \\ \phi\big(\cup_{i=1}^{\infty} E_i\big) &= \sum_{i=1}^{\infty} \phi(A_i) \end{aligned}$$

for all pair-wise disjoint sets $\{E_i\}$. In addition, for our development we will assume that $|\phi(\Omega)| < \infty$.

As usual, we define the *indicator function* I_E as

$$I_E(\omega) = \begin{cases} 1 & \text{if } \omega \in E \\ 0 & \text{if } \omega \notin E \end{cases}. \tag{2}$$

2 The Radon Nikodym Theorem

Theorem 1 *Let $(\Omega, \mathcal{A}, P)$ be a probability space and let ϕ be a signed measure such that $P(E) = 0$ implies $\phi(E) = 0$. Then there exists a unique random variable H such that*

$$\phi(E) = \int_E H(\omega) dP(\omega) = E_P(HI_E), \tag{3}$$

where E_P denotes expectation with respect to the probability P.

Proof: Let us first restrict ϕ to be non-negative. Let $R = P + \phi$ and define the linear functional L by

$$LX = \int_\Omega X(\omega) d\phi(\omega). \tag{4}$$

Now let us recall the Riesz Representation Theorem (e.g., see the notes for ECEn 671, page 97), which says that every linear functional can be represented as in inner product. Let us define the inner product

$$\langle X, Y\rangle = \int_\Omega X(\omega)Y(\omega)dR(\omega) = \int_\Omega X(\omega)Y(\omega)dP(\omega) + \int_\Omega X(\omega)Y(\omega)d\phi(\omega) \tag{5}$$

(You may want to review the claim that the sum of two inner products is an inner product). Thus, according to the Riesz Representation Theorem, there exists a random variable Z such that

$$\int_\Omega X(\omega)d\phi(\omega) = \langle X, Z\rangle = \int_\Omega X(\omega)Z(\omega)dR(\omega) \tag{6}$$

for all random variables X.

Now let $E \in \mathcal{A}$ such that $P(E) > 0$, and set $X = I_E$. The left-hand side of (6) is

$$\int_\Omega I_E(\omega)d\phi(\omega) = \phi(E). \tag{7}$$

We now observe that, since $P(E) = 0$ implies $\phi(E) = 0$, it follows that $\phi(E) \le P(E)$ for all $E \in \mathcal{A}$. Thus, we have

$$\int_\Omega I_E(\omega)Z(\omega)dR(\omega) = \phi(E) < \phi(E) + P(E) = R(E) \tag{8}$$

since $P(E) > 0$. Thus,

$$\frac{1}{R(E)}\int_E Z(\omega)dR(\omega) < 1. \tag{9}$$

The next step is to show that $0 \le Z(\omega) < 1$ almost surely. To do this, we need the following Lemma.

Lemma 2 *If* $0 \le \frac{1}{R(E)}\int_E Z(\omega)dR(\omega) < 1$ *for every* $E \in \mathcal{A}$*, then* $0 \le Z(\omega) < 1$ *almost surely.*

Proof: Let A be an interval in $[0,1]^c$ with center α and radius $r > 0$, and let $E = \{\omega \in \Omega\colon Z(\omega) \in A\}$. Thus, every point in A is separated from points in $[0,1]$ by a distance greater than r.

Now suppose $R(E) > 0$. Then

$$\begin{aligned}
\left|\frac{1}{R(E)}\int_E Z(\omega)dR(\omega) - \alpha\right| &= \frac{1}{R(E)}\left|\int_E (Z(\omega)-\alpha)dR(\omega)\right| \\
&\le \frac{1}{R(E)}\int_E \big|(Z(\omega)-\alpha)\big|dR(\omega) \\
&\le \frac{1}{R(E)}\int_E r\,dR(\omega) \\
&\le r,
\end{aligned}$$

which is impossible, since $\frac{1}{R(E)}\int_E Z(\omega)dR(\omega) < 1$. Thus, $R(E) = 0$ and $0 \le Z(\omega) < 1$ almost surely. Without loss of generality, we may thus assume that $Z(\omega) < 1$ for every ω. □

Continuing with the proof of the theorem, let us re-write (6) as

$$\int_{\Omega} X(\omega)d\phi(\omega) = \int_{\Omega} X(\omega)Z(\omega)d\phi(\omega) + \int_{\Omega} X(\omega)Z(\omega)dP(\omega) \tag{10}$$

or

$$\int_{\Omega}[1 - Z(\omega)]X(\omega)d\phi(\omega) = \int_{\Omega} X(\omega)Z(\omega)dP(\omega). \tag{11}$$

Now let $X = (1 + Z + Z^2 + \cdots + Z^n)I_E$, and obtain

$$\int_{E}[1 - Z^{n+1}(\omega)]d\phi(\omega) = \int_{E}[1 + Z(\omega) + \cdots + Z^n(\omega)]dP(\omega). \tag{12}$$

Now let $n \to \infty$. Since $0 \leq Z(\omega) < 1$, the left-hand side of (12) converges to $\phi(E)$ and the integrand of the right-hand size converges to a random variable H. Thus,

$$\phi(E) = \int_{E} H(\omega)dP(\omega). \tag{13}$$

We may extend the proof to the general signed case by decomposing a signed measure ϕ into positive and negative components, that is, let

$$\phi^+(E) = \begin{cases} \phi(E) & \text{if } \phi(E) \geq 0 \\ 0 & \text{if } \phi(E) < 0 \end{cases} \tag{14}$$

and

$$\phi^-(E) = \begin{cases} -\phi(E) & \text{if } \phi(E) < 0 \\ 0 & \text{if } \phi(E) \geq 0 \end{cases} \tag{15}$$

Then apply the above result to obtain

$$\phi^+(E) = \int_{E} H^+(\omega)dP(\omega). \tag{16}$$

and

$$\phi^-(E) = \int_{E} H^-(\omega)dP(\omega). \tag{17}$$

Letting $H = H^+ - H^-$ obtains the desired result. □

3 Conditional Expectation

Let Y be a random variable over $(\Omega, \mathcal{A}, P)$, and let $\sigma\{Y\}$ denote the σ-field *generated* by Y; that is, $\sigma\{Y\}$ is the smallest σ-field that contains the inverse images of the form $Y^{-1}(B)$, $\forall B \in \mathcal{B}(\mathbb{R})$. We may thus define the probability space $(\Omega, \sigma\{Y\}, Q)$, where Q is defined over elements of $\sigma\{Y\}$ such that

$$Q(E) = P(E) \; \forall E \in \sigma\{Y\} \tag{18}$$

Notice that Q is not defined over all elements of $\mathcal{A}$, since it will generally happen that $\sigma\{Y\}$ is a proper subset of $\mathcal{A}$. Such a σ-field is called a *sub-σ-field*.

Now define the function

$$\phi(G) = E_P X I_G = \int_G X(\omega) dP(\omega) \ \forall G \in \sigma\{Y\}. \tag{19}$$

Clearly, $\phi(\emptyset) = 0$. Let $\{G_i\}$ be a pairwise disjoint set such that $G_i \in \sigma\{Y\}$, $i = 1, 2, \ldots$. Then

$$\begin{aligned} \phi\big(\cup_{i=1}^{\infty}\big) &= E_P\big(X I_{\cup G_i}\big) \\ &= E_P\big(X \sum_{i=1}^{\infty} I_{G_i}\big) \\ &= \sum_{i=1}^{\infty} E_P\big(X I_{G_i}\big) \\ &= \sum_{i=1}^{\infty} \phi(G_i). \end{aligned}$$

Thus, ϕ is a signed measure.

Let X and Y be random variables over the probability space $(\Omega, \mathcal{A}, P)$.

Theorem 3 *The Radon-Nikodym derivative of Q with respect to P is orthogonal to every function of Y.*

Proof: Since ϕ is a signed measure, by the Radon Nikodym theorem there exists a random variable $H \in \sigma\{Y\}$ (i.e., H is a function of Y, such that

$$\phi(G) = \int_G X(\omega) dP(\omega) = \int_G H(\omega) dP(\omega) = E H I_G. \tag{20}$$

Rearranging, we obtain

$$E(X - H) I_G = 0, \tag{21}$$

which establishes that $X - H$ is orthogonal to I_G for every $G \in \sigma\{Y\}$.

Now let g be a piecewise constant function of Y; that is, g is of the form

$$g(Y) = \sum_i a_i I_{A_i} \tag{22}$$

where $\{a_i\}$ is a set of constants and $\{G_i\}$ is a set of pairwise disjoint sets in $\sigma\{Y\}$. Then

$$\begin{aligned} E[(X - H) g(Y)] &= E[(X - H) \sum_i a_i I_{A_i}] \\ &= \sum_i a_i E[(X - H) I_{A_i}] \\ &= 0. \end{aligned}$$

Now let g be an arbitrary measurable function over $\sigma\{Y\}$, and let g_i denote a sequence of piecewise constant functions such that

$$\lim_{i\to\infty} g_i(x) = g(x). \tag{23}$$

Now let g^+ and g^- be such that $g(x) = g^+(x) + g^-(x)$, where

$$g^+(x) = \begin{cases} g(x) & g(x) \geq 0 \\ 0 & g(x) < 0 \end{cases} \tag{24}$$

and

$$g^-(x) = \begin{cases} -g(x) & g(x) < 0 \\ 0 & g(x) \geq 0 \end{cases}. \tag{25}$$

Applying Lebesgue's Monotone Convergence Theorem to g^+ and g^- and combining the results yields

$$0 = \lim_{i\to\infty} E[(X - H)g_i(Y)] = E[(X - H)g(Y)]. \tag{26}$$

□

The random variable H is called the *conditional expectation* of X with respect to Y. Since H is a function of Y, it is more convenient to express this random variable as $H = h(Y)$.

To motivate this terminology, let us now suppose that X and Y possess a joint density function f_{XY}. Also, let g be any function of Y. Then orthogonality requires that

$$E[Xg(Y)] = E[h(Y)g(Y)]. \tag{27}$$

Expressing this relationship in terms of densities and using the fact that the joint density can be factored into the product of the conditional density of X given $Y = y$ and the marginal density of Y, we have $f_{XY}(x, y) = f_{X|Y}(x|y)f_Y(y)$. Thus,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xg(y)f_{XY}(x, y)dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(y)g(y)f_{XY}(x, y)dxdy \tag{28}$$

or, equivalently,

$$\int_{-\infty}^{\infty} g(y)\left[\int_{-\infty}^{\infty} xf_{X|Y}(x|y)dx\right]f_Y(y)dy = \int_{-\infty}^{\infty} h(y)g(y)\left[\int_{-\infty}^{\infty} f_{XY}(x, y)dx\right]dy. \tag{29}$$

But $\int_{-\infty}^{\infty} f_{XY}(x, y)dx = f_Y(y)$, thus we have

$$\int_{-\infty}^{\infty} g(y)\left[\int_{-\infty}^{\infty} xf_{X|Y}(x|y)dx\right]f_Y(y)dy = \int_{-\infty}^{\infty} h(y)g(y)f_Y(y)dy. \tag{30}$$

Since this function must hold for all functions g, it follows that

$$h(y) = \int_{-\infty}^{\infty} xf_{X|Y}(x|y)dx, \tag{31}$$

the conditional expectation of X given $Y = y$.