Lecture 5: Radon-Nikodym derivative

Let $(\Omega, \mathcal{F}, \nu)$ be a measure space and f be a nonnegative Borel function. Note that

$$\lambda(A) = \int_A f d\nu, \quad A \in \mathcal{F}$$

is a measure satisfying

$$\nu(A) = 0 \quad \text{implies} \quad \lambda(A) = 0.$$

(we say λ is *absolutely continuous* w.r.t. ν and write $\lambda \ll \nu$).

Comupting $\lambda(A)$ can be done through integration w.r.t. a well-known measure

$\lambda \ll \nu$ is also almost sufficient.

Theorem 1.4 (Radon-Nikodym theorem). Let ν and λ be two measures on $(\Omega, \mathcal{F})$ and ν be σ-finite. If $\lambda \ll \nu$, then there exists a nonnegative Borel function f on Ω such that

$$\lambda(A) = \int_A f d\nu, \quad A \in \mathcal{F}.$$

Furthermore, f is unique a.e. ν, i.e., if $\lambda(A) = \int_A g d\nu$ for any $A \in \mathcal{F}$, then $f = g$ a.e. ν.

The function f is called the Radon-Nikodym *derivative* or *density* of λ w.r.t. ν and is denoted by $d\lambda/d\nu$.

Consequence: If f is Borel on $(\Omega, \mathcal{F})$ and $\int_A f d\nu = 0$ for any $A \in \mathcal{F}$, then $f = 0$ a.e.

If $\int f d\nu = 1$ for an $f \geq 0$ a.e. ν, then λ is a probability measure and f is called its *probability density function* (p.d.f.) w.r.t. ν. For any probability measure P on $(\mathcal{R}^k, \mathcal{B}^k)$ corresponding to a c.d.f. F or a random vector X, if P has a p.d.f. f w.r.t. a measure ν, then f is also called the p.d.f. of F or X w.r.t. ν.

Example 1.10 (Discrete c.d.f. and p.d.f.). Let $a_1 < a_2 < \cdots$ be a sequence of real numbers and let p_n, $n = 1, 2, ...$, be a sequence of positive numbers such that $\sum_{n=1}^{\infty} p_n = 1$. Then

$$F(x) = \begin{cases} \sum_{i=1}^{n} p_i & a_n \leq x < a_{n+1}, \quad n = 1, 2, ... \\ 0 & -\infty < x < a_1. \end{cases}$$

is a *stepwise* c.d.f. It has a jump of size p_n at each a_n and is flat between a_n and a_{n+1}, $n = 1, 2, ...$ Such a c.d.f. is called a *discrete* c.d.f. The corresponding probability measure is

$$P(A) = \sum_{i: a_i \in A} p_i, \quad A \in \mathcal{F},$$

where $\mathcal{F}$ = the set of all subsets (power set).
Let ν be the counting measure on the power set. Then

$$P(A) = \int_A f d\nu = \sum_{a_i \in A} f(a_i), \quad A \subset \Omega,$$

where $f(a_i) = p_i$, $i = 1, 2, \ldots$. That is, f is the p.d.f. of P or F w.r.t. ν. Hence, any discrete c.d.f. has a p.d.f. w.r.t. counting measure. A p.d.f. w.r.t. counting measure is called a *discrete* p.d.f.

Example 1.11. Let F be a c.d.f. Assume that F is differentiable in the usual sense in calculus. Let f be the derivative of F. From calculus,

$$F(x) = \int_{-\infty}^{x} f(y)dy, \quad x \in \mathcal{R}.$$

Let P be the probability measure corresponding to F.
Then $P(A) = \int_A f dm$ for any $A \in \mathcal{B}$, where m is the Lebesgue measure on $\mathcal{R}$.
f is the p.d.f. of P or F w.r.t. Lebesgue measure.
Radon-Nikodym derivative is the same as the usual derivative in calculus.

A continuous c.d.f. may not have a p.d.f. w.r.t. Lebesgue measure.
A necessary and sufficient condition for a c.d.f. F having a p.d.f. w.r.t. Lebesgue measure is that F is *absolute continuous* in the sense that for any $\epsilon > 0$, there exists a $\delta > 0$ such that for each finite collection of disjoint bounded open intervals (a_i, b_i), $\sum(b_i - a_i) < \delta$ implies $\sum[F(b_i) - F(a_i)] < \epsilon$.
Absolute continuity is weaker than differentiability, but is stronger than continuity.
Note that every c.d.f. is differentiable a.e. Lebesgue measure (Chung, 1974, Chapter 1).

A p.d.f. w.r.t. Lebesgue measure is called a Lebesgue p.d.f.

Proposition 1.7 (Calculus with Radon-Nikodym derivatives). Let ν be a σ-finite measure on a measure space $(\Omega, \mathcal{F})$. All other measures discussed in (i)-(iii) are defined on $(\Omega, \mathcal{F})$.
(i) If λ is a measure, $\lambda \ll \nu$, and $f \geq 0$, then

$$\int f d\lambda = \int f \frac{d\lambda}{d\nu} d\nu.$$

(Notice how the $d\nu$'s "cancel" on the right-hand side.)
(ii) If λ_i, $i = 1, 2$, are measures and $\lambda_i \ll \nu$, then $\lambda_1 + \lambda_2 \ll \nu$ and

$$\frac{d(\lambda_1 + \lambda_2)}{d\nu} = \frac{d\lambda_1}{d\nu} + \frac{d\lambda_2}{d\nu} \quad \text{a.e. } \nu.$$

(iii) (Chain rule). If τ is a measure, λ is a σ-finite measure, and $\tau \ll \lambda \ll \nu$, then

$$\frac{d\tau}{d\nu} = \frac{d\tau}{d\lambda}\frac{d\lambda}{d\nu} \quad \text{a.e. } \nu.$$

In particular, if $\lambda \ll \nu$ and $\nu \ll \lambda$ (in which case λ and ν are *equivalent*), then

$$\frac{d\lambda}{d\nu} = \left(\frac{d\nu}{d\lambda}\right)^{-1} \quad \text{a.e. } \nu \text{ or } \lambda.$$

(iv) Let $(\Omega_i, \mathcal{F}_i, \nu_i)$ be a measure space and ν_i be σ-finite, $i = 1, 2$. Let λ_i be a σ-finite measure on $(\Omega_i, \mathcal{F}_i)$ and $\lambda_i \ll \nu_i$, $i = 1, 2$. Then $\lambda_1 \times \lambda_2 \ll \nu_1 \times \nu_2$ and

$$\frac{d(\lambda_1 \times \lambda_2)}{d(\nu_1 \times \nu_2)}(\omega_1, \omega_2) = \frac{d\lambda_1}{d\nu_1}(\omega_1)\frac{d\lambda_2}{d\nu_2}(\omega_2) \quad \text{a.e. } \nu_1 \times \nu_2.$$