Lecture 2: Integration theory and Radon-Nikodym derivative

a.e. and a.s. statements

A statement holds a.e. ν if it holds for all ω in N^c with $\nu(N) = 0$.
If ν is a probability, then a.e. may be replaced by a.s.

Proposition 1.6

Let $(\Omega, \mathscr{F}, \nu)$ be a measure space and f and g be Borel functions.

(i) If $f \leq g$ a.e., then $\int f d\nu \leq \int g d\nu$, provided that the integrals exist.

(ii) If $f \geq 0$ a.e. and $\int f d\nu = 0$, then $f = 0$ a.e.

Proof of (ii)

Let $A = \{f > 0\}$ and $A_n = \{f \geq n^{-1}\}$, $n = 1, 2, ...$
Then $A_n \subset A$ for any n and $\lim_{n\to\infty} A_n = \cup A_n = A$ (why?).
By Proposition 1.1(iii), $\lim_{n\to\infty} \nu(A_n) = \nu(A)$.
Using part (i) and Proposition 1.5, we obtain that, for any n,

$$n^{-1}\nu(A_n) = \int n^{-1} I_{A_n} d\nu \leq \int f I_{A_n} d\nu \leq \int f d\nu = 0$$

Exchange limit and integration

$\{f_n : n = 1, 2, ...\}$: a sequence of Borel functions.
Can we exchange the limit and integration, i.e.,

$$\int \lim_{n\to\infty} f_n d\nu = \lim_{n\to\infty} \int f_n d\nu?$$

Example 1.7

Consider $(\mathscr{R}, \mathscr{B})$ and the Lebesgue measure.
Define $f_n(x) = nI_{[0,n^{-1}]}(x)$, $n = 1, 2, ...$
Then $\lim_{n\to\infty} f_n(x) = 0$ for all x but $x = 0$.
Since a single point has Lebesgue measure 0, $\int \lim_{n\to\infty} f_n(x)dx = 0$.
On the other hand, $\int f_n(x)dx = 1$ for any n and $\lim_{n\to\infty} \int f_n(x)dx = 1$.

Theorem 1.1

Let $f_1, f_2, ...$ be a sequence of Borel functions on $(\Omega, \mathscr{F}, \nu)$.

(i) (Fatou's lemma). If $f_n \geq 0$, then $\int \liminf_n f_n d\nu \leq \liminf_n \int f_n d\nu$.

(ii) (Dominated convergence theorem). If $\lim_{n\to\infty} f_n = f$ a.e. and $|f_n| \leq g$ a.e. for integrable g, then $\int \lim_{n\to\infty} f_n d\nu = \lim_{n\to\infty} \int f_n d\nu$.

(iii) (Monotone convergence theorem). If $0 \leq f_1 \leq f_2 \leq \cdots$ and $\lim_{n\to\infty} f_n = f$ a.e., then $\int \lim_{n\to\infty} f_n d\nu = \lim_{n\to\infty} \int f_n d\nu$.

Partial proof of Theorem 1.1

Part (i) and part (iii) are equivalent (exercise)

See the textbook for a proof of part (iii).

We now prove part (ii) (the DCT) using Faton's lemma (part (iii))

By the condition, $g + f_n \geq 0$ and $g - f_n \geq 0$
By Faton's lemma and the fact that $\lim_n f_n = f$,

$$\int (g+f)d\nu = \int \liminf_n (g+f_n)d\nu \leq \liminf_n \int (g+f_n)d\nu$$

$$\int (g-f)d\nu = \int \liminf_n (g-f_n)d\nu \leq \liminf_n \int (g-f_n)d\nu$$

The last expression is the same as

$$\int (f-g)d\nu \geq \limsup_n \int (f_n-g)d\nu$$

Since g is integrable, all integrals are finite and we can cancel $\int g d\nu$ in the above inequalities.
Then

$$\int f d\nu \leq \liminf_n \int f_n d\nu \leq \limsup_n \int f_n d\nu \leq \int f d\nu$$

Example

Let $f_n(x) = \frac{n}{x+n}$, $x \in \Omega = [0,1]$, $n = 1, 2, ...$
Then $\lim_n f_n(x) = 1$.
To apply the DCT, note that $0 \leq f_n(x) \leq 1$.
To apply the MCT, note that $0 \leq f_n(x) \leq f_{n+1}(x)$.
Hence, $\lim_n \int f_n(x)dx = \int \lim_n f_n(x)dx = \int dx = 1$.

Example 1.8 (Interchange of differentiation and integration)

Let $(\Omega, \mathscr{F}, \nu)$ be a measure space and, for any fixed $\theta \in \mathscr{R}$, let $f(\omega, \theta)$ be a Borel function on Ω.
Suppose that $\partial f(\omega, \theta)/\partial \theta$ exists a.e. for $\theta \in (a, b) \subset \mathscr{R}$ and that $|\partial f(\omega, \theta)/\partial \theta| \leq g(\omega)$ a.e., where g is an integrable function on Ω.
Then, for each $\theta \in (a, b)$, $\partial f(\omega, \theta)/\partial \theta$ is integrable and, by Theorem 1.1(ii),

$$\frac{d}{d\theta} \int f(\omega, \theta) d\nu = \int \frac{\partial f(\omega, \theta)}{\partial \theta} d\nu.$$

Theorem 1.2 (Change of variables)

Let f be measurable from $(\Omega, \mathscr{F}, \nu)$ to $(\Lambda, \mathscr{G})$ and g be Borel on $(\Lambda, \mathscr{G})$. Then

$$\int_\Omega g \circ f d\nu = \int_\Lambda g d(\nu \circ f^{-1}),$$

i.e., if either integral exists, then so does the other, and the two are the same.

Remarks

- For Riemann integrals, $\int g(y)dy = \int g(f(x))f'(x)dx$, $y = f(x)$.
- For a random variable X on $(\Omega, \mathscr{F}, P)$, $EX = \int_\Omega XdP = \int_{\mathscr{R}} xdP_X$, $P_X = P \circ X^{-1}$
- Let Y be a random vector from Ω to $\mathscr{R}^k$ and g be Borel on $\mathscr{R}^k$.
    - Example: $Y = (X_1, X_2)$ and $g(Y) = X_1 + X_2$.
    - $E(X_1 + X_2) = EX_1 + EX_2$ (why?) $= \int_{\mathscr{R}} xdP_{X_1} + \int_{\mathscr{R}} xdP_{X_2}$.
    - We need to handle two integrals involving P_{X_1} and P_{X_2}.
    - On the other hand, $E(X_1 + X_2) = \int_{\mathscr{R}} xdP_{X_1+X_2}$ involving one integral w.r.t. $P_{X_1+X_2}$, which is not easy to obtain unless we have some knowledge about the joint c.d.f. of (X_1, X_2).

Theorem 1.3 (Fubini's theorem)

Let ν_i be a σ-finite measure on $(\Omega_i, \mathscr{F}_i)$, $i = 1, 2$, and f be a Borel function on $\prod_{i=1}^{2}(\Omega_i, \mathscr{F}_i)$ with $f \geq 0$ or $\int |f| \nu_1 \times \nu_2 < \infty$.
Then

$$g(\omega_2) = \int_{\Omega_1} f(\omega_1, \omega_2) d\nu_1$$

exists a.e. ν_2 and defines a Borel function on Ω_2 whose integral w.r.t. ν_2 exists, and

$$\int_{\Omega_1 \times \Omega_2} f(\omega_1, \omega_2) d\nu_1 \times \nu_2 = \int_{\Omega_2} \left[\int_{\Omega_1} f(\omega_1, \omega_2) d\nu_1 \right] d\nu_2.$$

Extensions to $\prod_{i=1}^{k}(\Omega_i, \mathscr{F}_i)$ is straightforward.
Fubini's theorem is very useful in

1. evaluating multi-dimensional integrals (exchanging the order of integrals);
2. proving a function is measurable;
3. proving some results by relating a one dimensional integral to a multi-dimensional integral

Radon-Nikodym derivative

Absolutely continuous

Let λ and ν be two measures on a measurable space $(\Omega, \mathcal{F}, \nu)$.
We say λ is *absolutely continuous* w.r.t. ν and write $\lambda \ll \nu$ iff

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

Let f be a nonnegative Borel function and

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

Then λ is a measure and $\lambda \ll \nu$.
Computing $\lambda(A)$ can be done through integration w.r.t. a well-known measure.
$\lambda \ll \nu$ is also almost sufficient for the existence of f with $\lambda(A) = \int_A f d\nu, \quad A \in \mathcal{F}$.

Theorem 1.4 (Radon-Nikodym theorem)

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

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

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

Remarks

- 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 $(\Omega, \mathscr{F})$ and $\int_A f d\nu = 0$ for any $A \in \mathscr{F}$, then $f = 0$ a.e.

Probability density function

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 $(\mathscr{R}^k, \mathscr{B}^k)$ 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.

Example 1.10 (continued)

The corresponding probability measure is

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

where $\mathscr{F}$ = 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.
A discrete p.d.f. f corresponds to a discrete c.d.f. F and the value $f(x)$ is the jump size of F at $x \in \mathscr{R}$.

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,

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

Let P be the probability measure corresponding to F.
Then

$$P(A) = \int_A f dm \quad \text{for any } A \in \mathscr{B}, \tag{1}$$

where m is the Lebesgue measure on $\mathscr{R}$.
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.

How do we prove (1)?

Proof of (1): π- and λ-system (Exercise 5)

Let $\mathscr{C} = \{(-\infty, x] : x \in \mathscr{R}\}$

$\mathscr{C}$ is a π-system: $A \in \mathscr{C}$ and $B \in \mathscr{C}$ imply $A \cap B \in \mathscr{C}$.

$\sigma(\mathscr{C}) = \mathscr{B}$

Let $\mathscr{D} = \{A \in \mathscr{B} : P(B) = \int fdm\}$

$\mathscr{C} \subset \mathscr{D}$.

The result follows (i.e., $\sigma(\mathscr{C}) \subset \mathscr{D}$) if we can show $\mathscr{D}$ is a λ-system:

$\emptyset \in \mathscr{D}$ (obvious)

$B \in \mathscr{D}$ implies $B^c \in \mathscr{D}$ (need to verify)

$B_i \in \mathscr{D}$ and B_i's are disjoint imply $\cup_i B_i \in \mathscr{D}$ (need to verify)

If $B \in \mathscr{D}$, then

$$P(B^c) = 1 - P(B) = 1 - \int_B fdm = \int fdm - \int I_B fdm$$

$$= \int (1 - I_B) fdm = \int I_{B^c} fdm = \int_{B^c} fdm.$$

This shows $B^c \in \mathscr{D}$.

If $B_i \in \mathscr{D}$ and B_i's are disjoint, then

$$\int_{\cup_i B_i} f dm = \int I_{\cup_i B_i} f dm = \int \sum_i I_{B_i} f dm = \sum_i \int I_{B_i} f dm$$

$$= \sum_i \int_{B_i} f dm = \sum_i P(B_i) = P(\cup_i B_i).$$

Thus, $\cup_i B_i \in \mathscr{D}$.

Example 1.11 (continued)

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 $\varepsilon > 0$, 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 $\sum[F(b_i) - F(a_i)] < \varepsilon$.
Absolute continuity is weaker than differentiability, but is stronger than continuity.

Remarks

- A p.d.f. w.r.t. Lebesgue measure is called a Lebesgue p.d.f.
- Note that every c.d.f. is differentiable a.e. Lebesgue measure (Chung, 1974, Chapter 1).
- Some c.d.f. does not have Lebesgue p.d.f.

Proposition 1.7 (Calculus with Radon-Nikodym derivatives)

Let ν be a σ-finite measure on a measure space $(\Omega, \mathscr{F})$.
All other measures discussed in (i)-(iii) are defined on $(\Omega, \mathscr{F})$.

(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

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

Proposition 1.7 (continued)

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

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

Proof of Proposition 1.7(i)

- If $f = I_B$ is an indicator function, then

$$\int f d\lambda = \int_B d\lambda = \lambda(B) = \int_B \frac{d\lambda}{d\nu} d\nu = \int f \frac{d\lambda}{d\nu} d\nu$$

- If $f = \sum_j a_j I_{B_j j} \geq 0$ (a nonnegative simple function), then

$$\int f d\lambda = \int \sum_j a_j I_{B_j} d\lambda = \sum_j a_j \int I_{B_j} d\lambda = \sum_j a_j \int I_{B_j} \frac{d\lambda}{d\nu} d\nu$$

$$= \int \sum_j a_j I_{B_j} \frac{d\lambda}{d\nu} d\nu = \int f \frac{d\lambda}{d\nu} d\nu$$

- For general $f \geq 0$, there exists an increasing sequence of nonnegative simple functions $\varphi_k \to f$ and

$$\int f d\lambda = \lim_k \int \varphi_k d\lambda = \lim_k \int \varphi_k \frac{d\lambda}{d\nu} d\nu = \int f \frac{d\lambda}{d\nu} d\nu$$