

Stat 709: Mathematical Statistics

Lecture 5

Jun Shao

Department of Statistics
University of Wisconsin
Madison, WI 53706, USA

Lecture 5: 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 (Ω, \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 .

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

Theorem 1.4 (Radon-Nikodym theorem)

Let ν and λ be two measures on (Ω, \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 .

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 (Ω, \mathcal{F}) and $\int_A f d\nu = 0$ for any $A \in \mathcal{F}$, 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 $(\mathcal{R}^k, \mathcal{B}^k)$ 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 $a_1 < a_2 < \dots$ be a sequence of real numbers and let $p_n, n = 1, 2, \dots$, be a sequence of positive numbers such that $\sum_{n=1}^{\infty} p_n = 1$.

Then

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

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} , $n = 1, 2, \dots$

Such a c.d.f. is called a *discrete* c.d.f.

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 $(\mathcal{R}^k, \mathcal{B}^k)$ 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 $a_1 < a_2 < \dots$ be a sequence of real numbers and let $p_n, n = 1, 2, \dots$, be a sequence of positive numbers such that $\sum_{n=1}^{\infty} p_n = 1$.

Then

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

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} , $n = 1, 2, \dots$

Such a c.d.f. is called a *discrete* c.d.f.

Example 1.10 (continued)

The corresponding probability measure is

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

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 $f(a_i) = p_i$, $i = 1, 2, \dots$

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 $x \in \mathcal{R}$.

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,

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

Let P be the probability measure corresponding to F .

Then

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

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.

How do we prove (1)?

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

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

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

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

Let $\mathcal{D} = \{A \in \mathcal{B} : P(B) = \int f dm\}$

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

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

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

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

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

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

$$\begin{aligned} P(B^c) &= 1 - P(B) = 1 - \int_B f dm = \int f dm - \int I_B f dm \\ &= \int (1 - I_B) f dm = \int I_{B^c} f dm = \int_{B^c} f dm. \end{aligned}$$

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

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

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

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

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

Let $\mathcal{D} = \{A \in \mathcal{B} : P(B) = \int f dm\}$

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

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

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

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

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

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

$$\begin{aligned} P(B^c) &= 1 - P(B) = 1 - \int_B f dm = \int f dm - \int I_B f dm \\ &= \int (1 - I_B) f dm = \int I_{B^c} f dm = \int_{B^c} f dm. \end{aligned}$$

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

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

$$\begin{aligned}\int_{\cup_i B_i} f dm &= \int I_{\cup_i B_i} f dm = \int \sum_i I_{B_i} f dm = \sum_i \int I_{B_i} f dm \\ &= \sum_i \int_{B_i} f dm = \sum_i P(B_i) = P(\cup_i B_i).\end{aligned}$$

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

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.

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

$$\begin{aligned}\int_{\cup_i B_i} f dm &= \int I_{\cup_i B_i} f dm = \int \sum_i I_{B_i} f dm = \sum_i \int I_{B_i} f dm \\ &= \sum_i \int_{B_i} f dm = \sum_i P(B_i) = P(\cup_i B_i).\end{aligned}$$

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

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 (Ω, \mathcal{F}) .

All other measures discussed in (i)-(iii) are defined on (Ω, \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 $\lambda_i, i = 1, 2$, are measures and $\lambda_i \ll \nu$, then $\lambda_1 + \lambda_2 \ll \nu$ and

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

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 (Ω, \mathcal{F}) .

All other measures discussed in (i)-(iii) are defined on (Ω, \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

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

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 $(\Omega_i, \mathcal{F}_i, \nu_i)$ 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

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