Absolutely continuous

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

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

Let $f$ be a nonnegative Borel function and

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

Then $\lambda$ 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

$$\lambda(A) = \int_A fd\nu, \quad A \in \mathcal{F}.$$
Theorem 1.4 (Radon-Nikodym theorem)

Let \( \nu \) and \( \lambda \) be two measures on \((\Omega, \mathcal{F})\) and \( \nu \) be \( \sigma \)-finite. If \( \lambda \ll \nu \), then there exists a nonnegative Borel function \( f \) on \( \Omega \) such that

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

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

Remarks

- The function \( f \) is called the Radon-Nikodym derivative or density of \( \lambda \) w.r.t. \( \nu \) 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.
Theorem 1.4 (Radon-Nikodym theorem)

Let \( \nu \) and \( \lambda \) be two measures on \((\Omega, \mathcal{F})\) and \( \nu \) be \( \sigma \)-finite. If \( \lambda \ll \nu \), then there exists a nonnegative Borel function \( f \) on \( \Omega \) such that

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

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

Remarks

- The function \( f \) is called the Radon-Nikodym derivative or density of \( \lambda \) w.r.t. \( \nu \) 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.
Probability density function

If \( \int f d\nu = 1 \) for an \( f \geq 0 \) a.e. \( \nu \), then \( \lambda \) is a probability measure and \( f \) is called its \textit{probability density function} (p.d.f.) w.r.t. \( \nu \). For any probability measure \( P \) on \((\mathbb{R}^k, \mathcal{B}^k)\) 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 \( \nu \), then \( f \) is also called the p.d.f. of \( F \) or \( X \) w.r.t. \( \nu \).

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 \( p_n, n = 1, 2, \ldots \), be a sequence of positive numbers such that \( \sum_{n=1}^{\infty} p_n = 1 \). Then

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

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 \( a_{n+1} \), \( n = 1, 2, \ldots \).

Such a c.d.f. is called a \textit{discrete} c.d.f.
Probability density function

If $\int f d\nu = 1$ for an $f \geq 0$ a.e. $\nu$, then $\lambda$ is a probability measure and $f$ is called its probability density function (p.d.f.) w.r.t. $\nu$.

For any probability measure $P$ on $(\mathbb{R}^k, \mathcal{B}^k)$ 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 $\nu$, then $f$ is also called the p.d.f. of $F$ or $X$ w.r.t. $\nu$.

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 $p_n, n = 1, 2, \ldots$, be a sequence of positive numbers such that $\sum_{n=1}^{\infty} p_n = 1$.

Then

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

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 $a_{n+1}$, $n = 1, 2, \ldots$.

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 \( \nu \) 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. \( \nu \).

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 \textit{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 \( x \in \mathbb{R} \).
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,

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

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 $\mathbb{R}$.
$f$ 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): \( \pi \)- and \( \lambda \)-system (Exercise 5)

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

\( \mathcal{C} \) is a \( \pi \)-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 \( \lambda \)-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 \( \bigcup_i B_i \in \mathcal{D} \) (need to verify)

If \( B \in \mathcal{D} \), then

\[
P(B^c) = 1 - P(B) = 1 - \int_B f dm = \int f dm - \int I_B f dm
\]

\[
= \int (1 - I_B) f dm = \int I_{B^c} f dm = \int_{B^c} f dm.
\]

This shows \( B^c \in \mathcal{D} \).
Proof of (1): $\pi$- and $\lambda$-system (Exercise 5)

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

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

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

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

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

The result follows (i.e., $\sigma(\mathcal{C}) \subset \mathcal{D}$) if we can show $\mathcal{D}$ is a $\lambda$-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 $\bigcup B_i \in \mathcal{D}$ (need to verify)

If $B \in \mathcal{D}$, 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 $B^c \in \mathcal{D}$. 
If $B_i \in \mathcal{D}$ and $B_i$’s are disjoint, then

$$
\int_{\bigcup_i B_i} f \, dm = \int l_{\bigcup_i B_i} f \, dm = \int \sum_i l_{B_i} f \, dm = \sum_i \int l_{B_i} f \, dm
$$

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

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

**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

\[
\int_{\bigcup_i B_i} fdm = \int \mathbb{I}_{\bigcup_i B_i} fdm = \int \sum_i \mathbb{I}_{B_i} fdm = \sum_i \int \mathbb{I}_{B_i} fdm
\]

\[
= \sum_i \int_{B_i} fdm = \sum_i P(B_i) = P(\bigcup_i B_i).
\]

Thus, \( \bigcup_i B_i \in \mathcal{D} \).

**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 $\nu$ be a $\sigma$-finite measure on a measure space $(\Omega, \mathcal{F})$.
All other measures discussed in (i)-(iii) are defined on $(\Omega, \mathcal{F})$.

(i) If $\lambda$ 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 \( \nu \) be a \( \sigma \)-finite measure on a measure space \((\Omega, \mathcal{F})\).
All other measures discussed in (i)-(iii) are defined on \((\Omega, \mathcal{F})\).

(i) If \( \lambda \) 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

\[
\frac{d(\lambda_1 + \lambda_2)}{d\nu} = \frac{d\lambda_1}{d\nu} + \frac{d\lambda_2}{d\nu} \quad \text{a.e. } \nu.
\]
Proposition 1.7 (continued)

(iii) (Chain rule). If $\tau$ is a measure, $\lambda$ is a $\sigma$-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 $\lambda$ and $\nu$ 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 $\nu_i$ be $\sigma$-finite, $i = 1, 2$. Let $\lambda_i$ be a $\sigma$-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

$$\frac{d(\lambda_1 \times \lambda_2)}{d(\nu_1 \times \nu_2)}(\omega_1, \omega_2) = \frac{d\lambda_1}{d\nu_1}(\omega_1) \frac{d\lambda_2}{d\nu_2}(\omega_2) \quad \text{a.e. } \nu_1 \times \nu_2.$$
Proof of Proposition 1.7(i)

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

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

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

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

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

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

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