Exchange limit and integration

Theorem 1.1
Let \( f_1, f_2, \ldots \) be a sequence of Borel functions on \((\Omega, \mathcal{F}, \nu)\).

(i) (Fatou’s lemma). If \( f_n \geq 0 \), then
\[
\int \lim \inf_n f_n \, d\nu \leq \lim \inf_n \int f_n \, d\nu.
\]

(ii) (Dominated convergence theorem). If \( \lim_{n \to \infty} f_n = f \) a.e. and there exists an integrable function \( g \) such that \( |f_n| \leq g \) a.e., 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
\[
\int \lim_{n \to \infty} f_n \, d\nu = \lim_{n \to \infty} \int f_n \, d\nu.
\]
Proof of Theorem 1.1
Part (i) and part (iii) are equivalent.

**Proof of Part (ii):**
By the condition, \( g + f_n \geq 0 \) and \( g - f_n \geq 0 \)
By Fatou’s lemma and the fact that \( \lim_{n \to \infty} f_n = f \),

\[
\int (g + f) \, d\nu = \int \liminf_n (g + f_n) \, d\nu \leq \liminf_n \int (g + f_n) \, d\nu
\]
\[
\int (g - f) \, d\nu = \int \liminf_n (g - f_n) \, d\nu \leq \liminf_n \int (g - f_n) \, d\nu
\]
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
\( \int gd\nu \) in the above inequalities.
Then

\[
\int fd\nu \leq \liminf_n \int f_n \, d\nu \leq \limsup_n \int f_n \, d\nu \leq \int fd\nu
\]
Proof of part (iii):

By Proposition 1.6 (i), \( \int f_n d\nu \) is increasing, hence there exists \( \lim_{n \to \infty} \int f_n d\nu \leq \int f d\nu \). It is left to show \( \lim_{n \to \infty} \int f_n d\nu \geq \int f d\nu \).

Case 1: Suppose there is a simple function \( \phi \) such that \( 0 \leq \phi \leq f \) and \( \nu(A_\phi) = \infty \), \( A_\phi = \{\phi > 0\} \).

Let \( a = \min_{\omega \in A_\phi} \phi(\omega) \) and \( b = a - \epsilon > 0 \) for some \( \epsilon > 0 \). Define \( A_n = \{f_n > b\} \). Then \( A_n \) is increasing and \( A_\phi \subset \bigcup A_n \).

\[
\int f_n d\nu \geq \int_{A_n} f_n d\nu \geq b \nu(A_n) \to \nu(\bigcup A_n) \geq \nu(A_\phi) = \infty.
\]

Thus, \( \lim_{n \to \infty} \int f_n d\nu \geq \infty \geq \int f d\nu \).

Case 2: Suppose for any \( 0 \leq \phi \leq f \), \( \nu(A_\phi) < \infty \).

Claim: There is a \( B \subset A_\phi \) with \( \nu(B) < \epsilon \) such that \( f_n \to f \) uniformly on \( A_\phi \cap B^c \).
If the claim is true then

\[
\int f_n \, d\nu \geq \int_{A_\phi \cap B^c} f_n \, d\nu \rightarrow \int_{A_\phi \cap B^c} f \, d\nu = \int_{A_\phi \cap B^c} \phi \, d\nu
\]

\[
= \int \phi \, d\nu - \int_B \phi \, d\nu \geq \int \phi \, d\nu - \epsilon \max_{\omega \in A_\phi} \phi(\omega).
\]

Since both \(\epsilon\) and \(\phi\) are arbitrary, \(\lim_{n \to \infty} \int f_n \, d\nu \geq \int f \, d\nu\).

**Proof of the claim:** Let \(\delta_k\) be a sequence of positive numbers converging to zero. Let \(B_{nk} = \cap_{m \geq n} \{\omega : |f_m(\omega) - f(\omega)| < \delta_k\}\). Then \(\cup_n B_{nk} = A_\phi\) and \(\nu(B_{nk}) \to \nu(\cup_n B_{nk}) = \nu(A_\phi)\). There is \(n_k\) large enough such that \(\nu(B_{nk}) > \nu(A_\phi) - \epsilon/2^{k+1}\). Let \(B^c = \cap_{k=1}^\infty B_{nk,k}\). Then \(f_n \to f\) uniformly on \(B^c\). On the other hand, \(\nu(B) = \nu(\cup_{k=1}^\infty B^c_{nk,k}) \leq \sum_{k=1}^\infty \nu(B^c_{nk,k}) \leq \epsilon\).

All the sets and their complements are taken within \(A_\phi\).
Change of variables

**Theorem 1.2**
Let $f$ be measurable from $(\Omega, \mathcal{F}, \nu)$ to $(\Lambda, \mathcal{G})$ and $g$ be Borel on $(\Lambda, \mathcal{G})$. Then

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

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

▶ For a random variable $X$ on $(\Omega, \mathcal{F}, P)$,

$$EX = \int_{\Omega} XdP = \int_{\mathbb{R}} xdP_X, \quad P_X = P \circ X^{-1}$$

▶ Notation: If $F$ is the c.d.f. of $P_X$,

$$\int_{\Omega} f \circ XdP = \int f(x)dP_X = \int f(x)dF(x) = \int fdF.$$
Fubini’s theorem (Theorem 1.3)

Let $\nu_i$ be a $\sigma$-finite measure on $(\Omega_i, \mathcal{F}_i)$, $i = 1, 2$, and let $f$ be a Borel function on $\prod_{i=1}^2(\Omega_i, \mathcal{F}_i)$. Suppose that either $f \geq 0$ 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. $\nu_2$ and defines a Borel function on $\Omega_2$ whose integral w.r.t. $\nu_2$ exists, and

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

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.
Example 1.9
Let $\Omega_1 = \Omega_2 = \{0, 1, 2, \ldots\}$, and $\nu_1 = \nu_2$ be the counting measure. A function $f$ on $\Omega_1 \times \Omega_2$ defines a double sequence. If either $f \geq 0$ or $\int |f| \, d\nu_1 \times \nu_2 < \infty$, then

$$\int fd\nu_1 \times \nu_2 = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} f(i, j) = \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} f(i, j)$$

(by Theorem 1.3 and Example 1.5). Thus, a double series can be summed in either order, if it is summable or $f \geq 0$. 
Example: Exercise 47

Let $X$ and $Y$ be random variables such that the joint c.d.f. of $(X, Y)$ is $F_X(x)F_Y(y)$, where $F_X$ and $F_Y$ are marginal c.d.f.’s.

Let $Z = X + Y$.

We want to show that

$$F_Z(z) = \int F_Y(z - x) dF_X(x).$$

Note that

$$F_Z(z) = \int_{x+y\leq z} dF_X(x) dF_Y(y)$$

$$= \int \left( \int_{y\leq z-x} dF_Y(y) \right) dF_X(x)$$

$$= \int F_Y(z - x) dF_X(x),$$

where the second equality follows from Fubini’s theorem.
Radon-Nikodym derivative

Absolute continuity

Let $\lambda$ and $\nu$ be two measures on a measurable space $(\Omega, \mathcal{F}, \nu)$. We say $\lambda$ is 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

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

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

Chen Zehua

ST5215: Advanced Statistical Theory
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 *probability density function* (p.d.f.) w.r.t. $\nu$. 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 $\nu$, then $f$ is also called the p.d.f. of $F$ or $X$ w.r.t. $\nu$. 
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 $(\mathcal{R}^k, \mathcal{B}^k)$ 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 $\nu$, 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 \( f(a_i) = p_i, \quad i = 1, 2, \ldots \).

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 \( x \in \mathcal{R} \).
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

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

(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.
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.
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

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