ST5215 Advanced Statistical Theory, Lecture 3

HUANG Dongming

National University of Singapore

18 Aug 2020

Review

Last time

- Exchange limit with measure
- Properties of integrals
- Exchange limit with integration (MCT, Fatou's Lemma, DCT)
- Change of variable
- Fubini's theorem
- Absolute continuity and Radon-Nikodym theorem
- Moments

Recap: inverse image

Let $f : \Omega \to \Lambda$ be a function and B be a subset of Λ.
$f^{-1}(B)$ is called the inverse image of B and is defined as

$$f^{-1}(B) = \{\omega \in \Omega : f(\omega) \in B\}$$

Recap: inverse image

Let $f : \Omega \to \Lambda$ be a function and B be a subset of Λ.
$f^{-1}(B)$ is called the inverse image of B and is defined as

$$f^{-1}(B) = \{\omega \in \Omega : f(\omega) \in B\}$$

- $f^{-1}(\emptyset) = \emptyset$
- $f^{-1}(\Lambda) = \Omega$

Recap: inverse image

Let $f : \Omega \to \Lambda$ be a function and B be a subset of Λ.
$f^{-1}(B)$ is called the inverse image of B and is defined as

$$f^{-1}(B) = \{\omega \in \Omega : f(\omega) \in B\}$$

- $f^{-1}(\emptyset) = \emptyset$
- $f^{-1}(\Lambda) = \Omega$

Example:

- $\Omega = \mathbb{Z}$, $\mathcal{F} = 2^{\Omega}$.
- Let $g(x) = 0$ for all $x \in \Omega$. Then

$$g^{-1}(A) = \begin{cases} \Omega, & \text{if } 0 \in A, \\ \emptyset, & \text{if } 0 \notin A. \end{cases} \tag{1}$$

Recap: inverse image

Let $f : \Omega \to \Lambda$ be a function and B be a subset of Λ.
$f^{-1}(B)$ is called the inverse image of B and is defined as

$$f^{-1}(B) = \{\omega \in \Omega : f(\omega) \in B\}$$

- $f^{-1}(\emptyset) = \emptyset$
- $f^{-1}(\Lambda) = \Omega$

Example:

- $\Omega = \mathbb{Z}$, $\mathcal{F} = 2^{\Omega}$.
- Let $g(x) = 0$ for all $x \in \Omega$. Then

$$g^{-1}(A) = \begin{cases} \Omega, & \text{if } 0 \in A, \\ \emptyset, & \text{if } 0 \notin A. \end{cases} \tag{1}$$

Recap: σ-field generated by a measurable function

Let f be a measurable function from $(\Omega, \mathcal{F})$ to $(\Lambda, \mathcal{G})$.
The σ-field generated by f is $\sigma(f) = \sigma(\{f^{-1}(B) : B \in \mathcal{G}\})$.

Recap: σ-field generated by a measurable function

Let f be a measurable function from $(\Omega, \mathcal{F})$ to $(\Lambda, \mathcal{G})$.
The σ-field generated by f is $\sigma(f) = \sigma(\{f^{-1}(B) : B \in \mathcal{G}\})$.
In probability, when X is a r.v., $\sigma(X)$ is usually interpreted as "the information in X", or "the events related to X"

Recap: σ-field generated by a measurable function

Let f be a measurable function from $(\Omega, \mathcal{F})$ to $(\Lambda, \mathcal{G})$.
The σ-field generated by f is $\sigma(f) = \sigma(\{f^{-1}(B) : B \in \mathcal{G}\})$.
In probability, when X is a r.v., $\sigma(X)$ is usually interpreted as "the information in X", or "the events related to X"
Example:

- $\Omega = \mathbb{Z}$, $\mathcal{F} = 2^{\Omega}$.
- Let $g(x) = 0$ for all $x \in \Omega$. Then $\sigma(g) = \{\emptyset, \Omega\}$

Recap: σ-field generated by a measurable function

Let f be a measurable function from $(\Omega, \mathcal{F})$ to $(\Lambda, \mathcal{G})$.
The σ-field generated by f is $\sigma(f) = \sigma(\{f^{-1}(B) : B \in \mathcal{G}\})$.
In probability, when X is a r.v., $\sigma(X)$ is usually interpreted as "the information in X", or "the events related to X"
Example:

- $\Omega = \mathbb{Z}$, $\mathcal{F} = 2^{\Omega}$.
- Let $g(x) = 0$ for all $x \in \Omega$. Then $\sigma(g) = \{\emptyset, \Omega\}$
- Let $f(x) = |x|$ for all $x \in \Omega$. Then f is a Borel function on $(\Omega, \mathcal{F})$.
- $f^{-1}(\{0\}) = \{0\}$, $f^{-1}(\{1\}) = \{-1, 1\}$, ..., $f^{-1}(\{n\}) = \{-n, n\}$, ...
- $\sigma(f) = \sigma(\{\{0\}, \{-1, 1\}, \ldots, \{-n, n\}, \ldots\})$
- Find an explicit expression of $\sigma(f)$

Recap: σ-field generated by a measurable function

Let f be a measurable function from $(\Omega, \mathcal{F})$ to $(\Lambda, \mathcal{G})$.
The σ-field generated by f is $\sigma(f) = \sigma(\{f^{-1}(B) : B \in \mathcal{G}\})$.
In probability, when X is a r.v., $\sigma(X)$ is usually interpreted as "the information in X", or "the events related to X"
Example:

- $\Omega = \mathbb{Z}$, $\mathcal{F} = 2^{\Omega}$.
- Let $g(x) = 0$ for all $x \in \Omega$. Then $\sigma(g) = \{\emptyset, \Omega\}$
- Let $f(x) = |x|$ for all $x \in \Omega$. Then f is a Borel function on $(\Omega, \mathcal{F})$.
- $f^{-1}(\{0\}) = \{0\}$, $f^{-1}(\{1\}) = \{-1, 1\}$, ..., $f^{-1}(\{n\}) = \{-n, n\}$, ...
- $\sigma(f) = \sigma(\{\{0\}, \{-1, 1\}, \ldots, \{-n, n\}, \ldots\})$
- Find an explicit expression of $\sigma(f)$:
  the collection of sets of the form $\{\pm n : n \in A\}$ where $A \subset \mathbb{N}$, including $\emptyset$

Recap: Lebesgue integral and Riemann integral

- Notations: $\int f \, \mathrm{d}\nu = \int f(\omega) \, \mathrm{d}\nu(\omega) = \int f(\omega)\nu(\, \mathrm{d}\omega)$

Recap: Lebesgue integral and Riemann integral

- Notations: $\int f \, \mathrm{d}\nu = \int f(\omega) \, \mathrm{d}\nu(\omega) = \int f(\omega)\nu(\, \mathrm{d}\omega)$
- For Lebesgue measure m, $\int f \, \mathrm{d}m = \int f(x)m(\, \mathrm{d}x)$

Recap: Lebesgue integral and Riemann integral

- Notations: $\int f \, \mathrm{d}\nu = \int f(\omega) \, \mathrm{d}\nu(\omega) = \int f(\omega)\nu(\, \mathrm{d}\omega)$
- For Lebesgue measure m, $\int f \, \mathrm{d}m = \int f(x)m(\, \mathrm{d}x)$
- You can write $\int_A f(x)m(\, \mathrm{d}x) = \int_A f(x) \, \mathrm{d}x$,
  (i.e., Lebesgue integral $=$ Riemann integral), if A is a closed interval or finite unions of closed intervals, and f is Riemann integrable on A

Recap: Lebesgue probability density function

Suppose P is a probability measure on $(\Omega, \mathcal{F})$ and X is a random variable.

- The induced measure $P_X = P \circ X^{-1}$ is also called the law of X or the distribution of X. It is a measure on $(\mathcal{R}, \mathcal{B})$.
- Give an example of P_X such that $P_X \ll m$
- Give an example of P_X such that P_X is not absolutely continuous w.r.t. m

Recap: convergence theorems

Exercise:

- Let $f_n(x) = nI_{(0,1/n)}$ be a sequence of functions on $(\mathcal{R}, \mathcal{B})$.
- Find f such that $f_n \to f$ m-a.e.
- Can you conclude that $\int f_n \,\mathrm{d}m \to \int \,\mathrm{d}f \,\mathrm{d}m$? Why?

Take the poll if you have finished.

Jensen's Inequality

- A set $A \subset \mathcal{R}^d$ is *convex* if for any $x, y \in A$, $tx + (1-t)y \in A$.

Jensen's Inequality

- A set $A \subset \mathcal{R}^d$ is *convex* if for any $x, y \in A$, $tx + (1-t)y \in A$.
- A real function f is *convex* on A if

$$f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2) \tag{2}$$

holds for any $x_1, x_2 \in A$ and $t \in [0, 1]$. If the inequality holds with $\leq$ replaced by $<$, then f is *strictly convex*.

Jensen's Inequality

- A set $A \subset \mathcal{R}^d$ is *convex* if for any $x, y \in A$, $tx + (1-t)y \in A$.
- A real function f is *convex* on A if

$$f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2) \tag{2}$$

  holds for any $x_1, x_2 \in A$ and $t \in [0, 1]$. If the inequality holds with $\leq$ replaced by $<$, then f is *strictly convex*.

Theorem

For a random vector and a convex function φ,

$$\varphi(\mathbb{E}X) \leq \mathbb{E}\varphi(X). \tag{3}$$

If φ is strictly convex and $\varphi(X)$ is not a constant, then $\varphi(\mathbb{E}X) < \mathbb{E}\varphi(X)$.

- If φ is twice differentiable, then the convexity of φ is implied by the positive semi-definiteness of its Hessian (or second derivative if φ is univariate) φ''.

Proof:

- A well known fact: For any $z \in A$, there exists a vector $c_z \in \mathcal{R}^d$ such that
$$\varphi(x) \geq \varphi(z) + c_z^\top (x - z), \forall x \in A. \tag{4}$$

Proof:

- A well known fact: For any $z \in A$, there exists a vector $c_z \in \mathcal{R}^d$ such that
$$\varphi(x) \geq \varphi(z) + c_z^\top (x - z), \forall x \in A. \tag{4}$$

- Substitute $z = \mathbb{E}X$ and $x = X$, and then take expectation on both sides. $\square$

Proof:

- A well known fact: For any $z \in A$, there exists a vector $c_z \in \mathcal{R}^d$ such that

$$\varphi(x) \geq \varphi(z) + c_z^\top (x - z), \forall x \in A. \tag{4}$$

- Substitute $z = \mathbb{E}X$ and $x = X$, and then take expectation on both sides. $\square$

Remark: c_z in the proof is called the sub-gradient of φ at z; when φ is differentiable at z, c_z is unique and equals to the derivative.
The hyperplane defined by $(x, y = \varphi(z) + c_z^\top (x - z))$ is called the supporting hyperplane. It is tangent to the surface of φ (i.e., $(x, y = \varphi(x))$). They intersect at $(x = z, y = \varphi(z))$.

Examples

Suppose X is a nonconstant positive r.v., then

- $(\mathbb{E}X)^{-1} < \mathbb{E}(X^{-1})$ (since $f(x) = x^{-1}$ is convex for $x > 0$)
- $\mathbb{E}(\log X) < \log(\mathbb{E}X)$ (since $f(x) = \log(x)$ is concave, i.e. $-\log(x)$ is convex)

Examples

Suppose X is a nonconstant positive r.v., then

- $(\mathbb{E}X)^{-1} < \mathbb{E}(X^{-1})$ (since $f(x) = x^{-1}$ is convex for $x > 0$)
- $\mathbb{E}(\log X) < \log(\mathbb{E}X)$ (since $f(x) = \log(x)$ is concave, i.e. $-\log(x)$ is convex)
- Let f and g be positive integrable functions on a measure space with a σ -finite measure ν. If $\int f \, \mathrm{d}\nu \geq \int g \, \mathrm{d}\nu > 0$, we can show that

$$\int f \log\left(\frac{f}{g}\right) \, \mathrm{d}\nu \geq 0$$

Examples

Suppose X is a nonconstant positive r.v., then

- $(\mathbb{E}X)^{-1} < \mathbb{E}(X^{-1})$ (since $f(x) = x^{-1}$ is convex for $x > 0$)
- $\mathbb{E}(\log X) < \log(\mathbb{E}X)$ (since $f(x) = \log(x)$ is concave, i.e. $-\log(x)$ is convex)
- Let f and g be positive integrable functions on a measure space with a σ -finite measure ν. If $\int f \, \mathrm{d}\nu \geq \int g \, \mathrm{d}\nu > 0$, we can show that

$$\int f \log\left(\frac{f}{g}\right) \, \mathrm{d}\nu \geq 0$$

1. Let $h = f / \int f \, \mathrm{d}\nu$, a p.d.f. w.r.t. ν. Let $\mathbb{E}$ is the expectation w.r.t. h

Examples

Suppose X is a nonconstant positive r.v., then

- $(\mathbb{E}X)^{-1} < \mathbb{E}(X^{-1})$ (since $f(x) = x^{-1}$ is convex for $x > 0$)
- $\mathbb{E}(\log X) < \log(\mathbb{E}X)$ (since $f(x) = \log(x)$ is concave, i.e. $-\log(x)$ is convex)
- Let f and g be positive integrable functions on a measure space with a σ -finite measure ν. If $\int f \, \mathrm{d}\nu \geq \int g \, \mathrm{d}\nu > 0$, we can show that

$$\int f \log\left(\frac{f}{g}\right) \, \mathrm{d}\nu \geq 0$$

1. Let $h = f / \int f \, \mathrm{d}\nu$, a p.d.f. w.r.t. ν. Let $\mathbb{E}$ is the expectation w.r.t. h
2. By Jensen's inequality, $\mathbb{E} \log(g/f) \leq \log(\mathbb{E}(g/f))$

Examples

Suppose X is a nonconstant positive r.v., then

- $(\mathbb{E}X)^{-1} < \mathbb{E}(X^{-1})$ (since $f(x) = x^{-1}$ is convex for $x > 0$)
- $\mathbb{E}(\log X) < \log(\mathbb{E}X)$ (since $f(x) = \log(x)$ is concave, i.e. $-\log(x)$ is convex)
- Let f and g be positive integrable functions on a measure space with a σ -finite measure ν. If $\int f \, \mathrm{d}\nu \geq \int g \, \mathrm{d}\nu > 0$, we can show that

$$\int f \log\left(\frac{f}{g}\right) \, \mathrm{d}\nu \geq 0$$

1. Let $h = f / \int f \, \mathrm{d}\nu$, a p.d.f. w.r.t. ν. Let $\mathbb{E}$ is the expectation w.r.t. h
2. By Jensen's inequality, $\mathbb{E}\log(g/f) \leq \log(\mathbb{E}(g/f))$
3. LHS is $\mathbb{E}\log(g/f) = \int \log\left(\frac{g}{f}\right) h \, \mathrm{d}\nu = \int \log\left(\frac{g}{f}\right) f \, \mathrm{d}\nu / \int f \, \mathrm{d}\nu$

Examples

Suppose X is a nonconstant positive r.v., then

- $(\mathbb{E}X)^{-1} < \mathbb{E}(X^{-1})$ (since $f(x) = x^{-1}$ is convex for $x > 0$)
- $\mathbb{E}(\log X) < \log(\mathbb{E}X)$ (since $f(x) = \log(x)$ is concave, i.e. $-\log(x)$ is convex)
- Let f and g be positive integrable functions on a measure space with a σ -finite measure ν. If $\int f \,\mathrm{d}\nu \geq \int g \,\mathrm{d}\nu > 0$, we can show that

$$\int f \log\left(\frac{f}{g}\right) \,\mathrm{d}\nu \geq 0$$

1. Let $h = f / \int f \,\mathrm{d}\nu$, a p.d.f. w.r.t. ν. Let $\mathbb{E}$ is the expectation w.r.t. h
2. By Jensen's inequality, $\mathbb{E}\log(g/f) \leq \log(\mathbb{E}(g/f))$
3. LHS is $\mathbb{E}\log(g/f) = \int \log\left(\frac{g}{f}\right) h \,\mathrm{d}\nu = \int \log\left(\frac{g}{f}\right) f \,\mathrm{d}\nu / \int f \,\mathrm{d}\nu$
4. RHS is $\log(\mathbb{E}(g/f)) = \log\left(\int \frac{g}{f} h \,\mathrm{d}\nu\right) = \log\left(\frac{\int g \,\mathrm{d}\nu}{\int f \,\mathrm{d}\nu}\right) \leq 0$

Theorem (Chebyshev)

Let X be a random variable and φ be a nonnegative and nondecreasing function on $[0, \infty)$ and $\varphi(-t) = \varphi(t)$ for all real t. Then, for each constant $t \geq 0$,

$$\varphi(t)P(|X| \geq t) \leq \int_{\{|X| \geq t\}} \varphi(X)\, \mathrm{d}P \leq \mathbb{E}\varphi(X). \tag{5}$$

Theorem (Chebyshev)

Let X be a random variable and φ be a nonnegative and nondecreasing function on $[0, \infty)$ and $\varphi(-t) = \varphi(t)$ for all real t. Then, for each constant $t \geq 0$,

$$\varphi(t)P(|X| \geq t) \leq \int_{\{|X| \geq t\}} \varphi(X) \, \mathrm{d}P \leq \mathbb{E}\varphi(X). \tag{5}$$

Proof: $\varphi(X) \geq \varphi(X)1_{\{|X| \geq t\}} \geq \varphi(t)1_{\{|X| \geq t\}}$

Theorem (Chebyshev)

Let X be a random variable and φ be a nonnegative and nondecreasing function on $[0, \infty)$ and $\varphi(-t) = \varphi(t)$ for all real t. Then, for each constant $t \geq 0$,

$$\varphi(t)P(|X| \geq t) \leq \int_{\{|X| \geq t\}} \varphi(X) \, \mathrm{d}P \leq \mathbb{E}\varphi(X). \tag{5}$$

Proof: $\varphi(X) \geq \varphi(X)1_{\{|X| \geq t\}} \geq \varphi(t)1_{\{|X| \geq t\}}$

Examples:

- $\varphi(t) = |t|$, we have Markov's inequality

$$P(|X| \geq t) \leq \frac{\mathbb{E}|X|}{t}, \quad \forall t > 0. \tag{6}$$

Theorem (Chebyshev)

Let X be a random variable and φ be a nonnegative and nondecreasing function on $[0, \infty)$ and $\varphi(-t) = \varphi(t)$ for all real t. Then, for each constant $t \geq 0$,

$$\varphi(t)P(|X| \geq t) \leq \int_{\{|X| \geq t\}} \varphi(X)\, \mathrm{d}P \leq \mathbb{E}\varphi(X). \tag{5}$$

Proof: $\varphi(X) \geq \varphi(X)1_{\{|X| \geq t\}} \geq \varphi(t)1_{\{|X| \geq t\}}$

Examples:

- $\varphi(t) = |t|$, we have Markov's inequality

$$P(|X| \geq t) \leq \frac{\mathbb{E}|X|}{t}, \quad \forall t > 0. \tag{6}$$

- $\varphi(t) = t^2$ and X is replaced with $X - \mu$ where $\mu = \mathbb{E}X$, we obtain the classic Chebyshev' inequality:

$$P(|X - \mu| \geq t) \leq \frac{\sigma_X^2}{t^2}. \tag{7}$$

Examples

- Let $X \in \{a_1, \ldots, a_n\}$ and $P(X = a_i) = 1/n$.
- Let $\varphi(x) = x^2$, convex.

Examples

- Let $X \in \{a_1, \ldots, a_n\}$ and $P(X = a_i) = 1/n$.
- Let $\varphi(x) = x^2$, convex.
- $\varphi(\mathbb{E}X) = \left(\frac{1}{n}\sum_{i=1}^n a_i\right)^2$ and $\mathbb{E}\varphi(X) = \frac{1}{n}\sum_{i=1}^n a_i^2$

Then Jensen's inequality implies that

$$\left(\frac{1}{n}\sum_{i=1}^n a_i\right)^2 \leq \frac{1}{n}\sum_{i=1}^n a_i^2. \tag{8}$$

Hölder's inequality

Suppose $1/p + 1/q = 1$ and X, Y are r.v.s

$$\mathbb{E}|XY| \leq (\mathbb{E}|X|^p)^{1/p}(\mathbb{E}|Y|^q)^{1/q} \tag{9}$$

- If $1/p + 1/q = 1$, then we say p and q are *Hölder conjugate* of each other.
- If both $\mathbb{E}|X|^p$ and $\mathbb{E}|Y|^q$ are finite, then the equality holds if and only if $|X|^p$ and $|Y|^q$ are linearly dependent (i.e., $a|X|^p = b|Y|^q$ a.s., for some $a, b \geq 0$ and not both zero)
- Sketch of proof:
  - Discuss: (1). If RHS $= \infty$, (2). If $E|X|^p = 0$ or $E|Y|^q = 0$, (3). If both $E|X|^p$ and $E|Y|^q$ are in $(0, \infty)$
  - For (3), use Young's inequality:

    $$ab \leq \frac{a^p}{p} + \frac{b^q}{q}$$

    for all nonnegative a and b, where equality is achieved if and only if $a^p = b^q$.

Cauchy-Schwarz inequality: when $p = q = 2$, we have

$$\mathbb{E}|XY| \leq \sqrt{\mathbb{E}X^2}\sqrt{\mathbb{E}Y^2}. \tag{10}$$

Cauchy-Schwarz inequality: when $p = q = 2$, we have

$$\mathbb{E}|XY| \leq \sqrt{\mathbb{E}X^2}\sqrt{\mathbb{E}Y^2}. \tag{10}$$

- This also implies that $|\mathrm{Cov}(XY)| \leq \sigma_X \sigma_Y$ and hence the correlation between X and Y are between -1 and 1
  - Use the inequality for $X - \mathbb{E}X$ and $Y - \mathbb{E}Y$

Minkowski's inequality

$$(\mathbb{E}|X+Y|^p)^{1/p} \leq (\mathbb{E}|X|^p)^{1/p} + (\mathbb{E}|Y|^p)^{1/p}, p \geq 1 \tag{11}$$

Proof:

- Let $q = p/(p-1)$ so that $1/p + 1/q = 1$.
- Decompose: $|X+Y|^p \leq |X+Y|^{p-1}(|X|+|Y|)$
- Use Hölder's inequality for $|X|$ and $|X+Y|^{p-1}$ with p and q
- Divide both sides by $[\mathbb{E}|X+Y|^p]^{\frac{p-1}{p}}$.

Related inequalities

- (Generalization of Hölder's inequality). For $0 < p < 1$ and $q = -p/(1-p)$
$$E|XY| \geqslant (E|X|^p)^{1/p} (E|Y|^q)^{1/q}$$
- (Generalization of Minkowski's inequality). $\left(E\left(\sum_{j=1}^n |X_j|\right)^r\right)^{1/r} > \sum_{j=1}^n (E|X_j|^r)^{1/r}$ for $0 < r < 1$

Left for the next tutorial. (Hint: Use Hölder's inequality)

Lyapunov's inequality

For $0 < s < t$,

$$(\mathbb{E}|X|^s)^{1/s} \leq (\mathbb{E}|X|^t)^{1/t}. \tag{12}$$

Lyapunov's inequality

For $0 < s < t$,

$$(\mathbb{E}|X|^s)^{1/s} \leq (\mathbb{E}|X|^t)^{1/t}. \tag{12}$$

Proof:
Let $p = \frac{t}{s} > 1$. Let $q = p/(p-1)$ so that $1/p + 1/q = 1$.
By Hölder's inequality,

$$\begin{aligned}
\mathbb{E}\left[|X|^s\right] &= \mathbb{E}\left[|X|^s \cdot 1\right] \\
&\leq \mathbb{E}\left[|X|^{sp}\right]^{1/p} \mathbb{E}\left[1^q\right]^{1/q} \\
&= \mathbb{E}\left[|X|^t\right]^{s/t},
\end{aligned}$$

i.e., $(\mathbb{E}|X|^s)^{1/s} \leq (\mathbb{E}|X|^t)^{1/t}$. $\square$

Characteristic function and moment generating function

Definition

Let X be a random d-vector.

- The characteristic function (ch.f.) φ of X of P_X is defined as

$$\phi_X(t) = \mathbb{E}e^{\sqrt{-1}t^\top X},$$

  where $e^{it^\top X} = \cos(t^\top X) + \sqrt{-1}\sin(t^\top X)$.

- The moment generating function (MGF, m.g.f.) of X or P_X is defined as

$$\psi_X(t) = \mathbb{E}e^{t^\top X}.$$

- ϕ_X is complex-valued and always well defined. $|\phi_X| \leq 1$, and uniformly continuous

Characteristic function and moment generating function

Definition

Let X be a random d-vector.

- The characteristic function (ch.f.) φ of X of P_X is defined as

$$\phi_X(t) = \mathbb{E}e^{\sqrt{-1}t^\top X},$$

  where $e^{it^\top X} = \cos(t^\top X) + \sqrt{-1}\sin(t^\top X)$.

- The moment generating function (MGF, m.g.f.) of X or P_X is defined as

$$\psi_X(t) = \mathbb{E}e^{t^\top X}.$$

- ϕ_X is complex-valued and always well defined. $|\phi_X| \leq 1$, and uniformly continuous
- $\psi_X(t)$ is nonnegative, but may be infinite for some t

Characteristic function and moment generating function

Definition

Let X be a random d-vector.

- The characteristic function (ch.f.) φ of X of P_X is defined as

$$\phi_X(t) = \mathbb{E}e^{\sqrt{-1}t^\top X},$$

  where $e^{it^\top X} = \cos(t^\top X) + \sqrt{-1}\sin(t^\top X)$.

- The moment generating function (MGF, m.g.f.) of X or P_X is defined as

$$\psi_X(t) = \mathbb{E}e^{t^\top X}.$$

- ϕ_X is complex-valued and always well defined. $|\phi_X| \leq 1$, and uniformly continuous
- $\psi_X(t)$ is nonnegative, but may be infinite for some t
  - e.g., if P_X is the Cauchy distribution, then $\psi_X(t) = \infty$ for all $t \neq 0$

Characteristic function and moment generating function

Definition

Let X be a random d-vector.

- The characteristic function (ch.f.) φ of X of P_X is defined as

$$\phi_X(t) = \mathbb{E}e^{\sqrt{-1}t^\top X},$$

where $e^{it^\top X} = \cos(t^\top X) + \sqrt{-1}\sin(t^\top X)$.

- The moment generating function (MGF, m.g.f.) of X or P_X is defined as

$$\psi_X(t) = \mathbb{E}e^{t^\top X}.$$

- ϕ_X is complex-valued and always well defined. $|\phi_X| \leq 1$, and uniformly continuous
- $\psi_X(t)$ is nonnegative, but may be infinite for some t
  - e.g., if P_X is the Cauchy distribution, then $\psi_X(t) = \infty$ for all $t \neq 0$
- $\phi_X(0) = \psi_X(0) = 1$, $\phi_{-X}(t) = \overline{\phi_X(t)}$, $\psi_{-X}(t) = \psi_X(-t)$.

Properties

If the m.g.f. is finite in a neighborhood of $0 \in \mathcal{R}^k$, then

- moments of X of any order are finite,
- $\phi_X(t)$ can be obtained by replacing t in $\psi_X(t)$ by $\sqrt{-1}t$.

Properties

If the m.g.f. is finite in a neighborhood of $0 \in \mathcal{R}^k$, then

- moments of X of any order are finite,
- $\phi_X(t)$ can be obtained by replacing t in $\psi_X(t)$ by $\sqrt{-1}t$.

ch.f. and m.g.f. determine distributions uniquely (Theorem 1.6 in JS)

- if $\phi_X(t) = \phi_Y(t)$ for all t, then $P_X = P_Y$
- if $\psi_X(t) = \psi_Y(t) < \infty$ for all t in a neighborhood of 0, then $P_X = P_Y$.

Example

Claim: X is symmetric about 0 iff its ch.f. ϕ_X is real-valued.

- Symmetry: A random vector X is *symmetric* about 0 iff X and $-X$ have the same distribution.

Example

Claim: X is symmetric about 0 iff its ch.f. ϕ_X is real-valued.

- Symmetry: A random vector X is *symmetric* about 0 iff X and $-X$ have the same distribution.

$\Rightarrow$:

- If X and $-X$ have the same distribution, then $\phi_X(t) = \phi_{-X}(t)$.

Example

Claim: X is symmetric about 0 iff its ch.f. ϕ_X is real-valued.

- Symmetry: A random vector X is *symmetric* about 0 iff X and $-X$ have the same distribution.

$\Rightarrow$:

- If X and $-X$ have the same distribution, then $\phi_X(t) = \phi_{-X}(t)$.
- But $\phi_{-X}(t) = \phi_X(-t) = \overline{\phi_X(t)}$, so $\phi_X(t) = \overline{\phi_X(t)}$, thus, ϕ_X is real-valued.

Example

Claim: X is symmetric about 0 iff its ch.f. ϕ_X is real-valued.

- Symmetry: A random vector X is *symmetric* about 0 iff X and $-X$ have the same distribution.

$\Rightarrow$:

- If X and $-X$ have the same distribution, then $\phi_X(t) = \phi_{-X}(t)$.
- But $\phi_{-X}(t) = \phi_X(-t) = \overline{\phi_X(t)}$, so $\phi_X(t) = \overline{\phi_X(t)}$, thus, ϕ_X is real-valued.
- Here we used $\sin(-t^\tau X) = -\sin(t^\tau X)$ and $\cos(t^\tau X) = \cos(-t^\tau X)$. $\square$

Example

Claim: X is symmetric about 0 iff its ch.f. ϕ_X is real-valued.

- Symmetry: A random vector X is *symmetric* about 0 iff X and $-X$ have the same distribution.

$\Rightarrow$:

- If X and $-X$ have the same distribution, then $\phi_X(t) = \phi_{-X}(t)$.
- But $\phi_{-X}(t) = \phi_X(-t) = \overline{\phi_X(t)}$, so $\phi_X(t) = \overline{\phi_X(t)}$, thus, ϕ_X is real-valued.
- Here we used $\sin(-t^\tau X) = -\sin(t^\tau X)$ and $\cos(t^\tau X) = \cos(-t^\tau X)$. $\square$

$\Leftarrow$:

- If ϕ_X is real-valued, $\phi_X(t) = \overline{\phi_X(t)} = \phi_{-X}(t)$.

Example

Claim: X is symmetric about 0 iff its ch.f. ϕ_X is real-valued.

- Symmetry: A random vector X is *symmetric* about 0 iff X and $-X$ have the same distribution.

$\Rightarrow$:

- If X and $-X$ have the same distribution, then $\phi_X(t) = \phi_{-X}(t)$.
- But $\phi_{-X}(t) = \phi_X(-t) = \overline{\phi_X(t)}$, so $\phi_X(t) = \overline{\phi_X(t)}$, thus, ϕ_X is real-valued.
- Here we used $\sin(-t^\tau X) = -\sin(t^\tau X)$ and $\cos(t^\tau X) = \cos(-t^\tau X)$. $\square$

$\Leftarrow$:

- If ϕ_X is real-valued, $\phi_X(t) = \overline{\phi_X(t)} = \phi_{-X}(t)$.
- Since a ch.f. uniquely determines a distribution (Theorem 1.6), X and $-X$ must have the same distribution.

Independence

Definition

Let $(\Omega, \mathcal{E}, P)$ be a probability space.

- (Independent events) The events in a subset $\mathcal{C} \subset \mathcal{E}$ are said to be *independent* iff for any positive n and distinct events $A_1, \ldots, A_n \in \mathcal{C}$,

$$P(A_1 \cap \cdots \cap A_n) = P(A_1) \cdots P(A_n). \tag{13}$$

Independence

Definition

Let $(\Omega, \mathcal{E}, P)$ be a probability space.

- (Independent events) The events in a subset $\mathcal{C} \subset \mathcal{E}$ are said to be *independent* iff for any positive n and distinct events $A_1, \ldots, A_n \in \mathcal{C}$,

$$P(A_1 \cap \cdots \cap A_n) = P(A_1) \cdots P(A_n). \tag{13}$$

- (Independent collections) Collections $\mathcal{C}_i \subset \mathcal{E}$, $i \in \mathcal{I}$ are independent iff events in a collection of the form $\{A_i \in \mathcal{C}_i : i \in \mathcal{I}\}$ are independent.

Independence

Definition

Let $(\Omega, \mathcal{E}, P)$ be a probability space.

- (Independent events) The events in a subset $\mathcal{C} \subset \mathcal{E}$ are said to be *independent* iff for any positive n and distinct events $A_1, \dots, A_n \in \mathcal{C}$,

$$P(A_1 \cap \cdots \cap A_n) = P(A_1) \cdots P(A_n). \tag{13}$$

- (Independent collections) Collections $\mathcal{C}_i \subset \mathcal{E}$, $i \in \mathcal{I}$ are independent iff events in a collection of the form $\{A_i \in \mathcal{C}_i : i \in \mathcal{I}\}$ are independent.
- (Independent random variables): random variables $X_1, \dots, X_n$ are said to be independent iff $\sigma(X_1), \dots, \sigma(X_n)$ are independent.

Properties

- To check whether random variables $X_1, \ldots, X_n$ are independent, only need to check whether for any real numbers a_i's,

$$P(X_1 \leq a_1, \cdots, X_n \leq a_n) = P(X_1 \leq a_1) \cdots P(X_n \leq a_n). \qquad (14)$$

Properties

- To check whether random variables $X_1, \ldots, X_n$ are independent, only need to check whether for any real numbers a_i's,

$$P(X_1 \leq a_1, \cdots, X_n \leq a_n) = P(X_1 \leq a_1) \cdots P(X_n \leq a_n). \tag{14}$$

- If $(X_1, \ldots, X_n)$ has a joint p.d.f. f w.r.t. a product measure $\nu_1 \times \cdots \times \nu_n$ defined on $\mathcal{B}^n$, then $X_1, \ldots, X_n$ are independent if and only if

$$f(x_1, \ldots, x_n) = f_1(x_1) \cdots f_n(x_n), \quad (x_1, \ldots, x_n) \in \mathcal{R}^n$$

where f_i is the p.d.f. of X_i w.r.t. $\nu_i, i = 1, \ldots, n$.

Properties (Cont.)

- (Lemma 1.1 in JS). Let $X_1, \ldots, X_n$ be independent random variables. Then random variables $g(X_1, \ldots, X_k)$ and $h(X_{k+1}, \ldots, X_n)$ are independent, where g and h are Borel functions and k is an integer between 1 and n.

Properties (Cont.)

- (Lemma 1.1 in JS). Let $X_1, \ldots, X_n$ be independent random variables. Then random variables $g(X_1, \ldots, X_k)$ and $h(X_{k+1}, \ldots, X_n)$ are independent, where g and h are Borel functions and k is an integer between 1 and n.
- We usually use $X \perp Y$ to denote that X and Y are independent.

Properties (Cont.)

- (Lemma 1.1 in JS). Let $X_1, \ldots, X_n$ be independent random variables. Then random variables $g(X_1, \ldots, X_k)$ and $h(X_{k+1}, \ldots, X_n)$ are independent, where g and h are Borel functions and k is an integer between 1 and n.
- We usually use $X \perp Y$ to denote that X and Y are independent.
- If $X \perp Y$, $\mathbb{E}(XY) = (\mathbb{E}X)(\mathbb{E}Y)$.

Properties (Cont.)

- (Lemma 1.1 in JS). Let $X_1, \ldots, X_n$ be independent random variables. Then random variables $g(X_1, \ldots, X_k)$ and $h(X_{k+1}, \ldots, X_n)$ are independent, where g and h are Borel functions and k is an integer between 1 and n.
- We usually use $X \perp Y$ to denote that X and Y are independent.
- If $X \perp Y$, $\mathbb{E}(XY) = (\mathbb{E}X)(\mathbb{E}Y)$.

Exercises:

- Find an example that $\mathbb{E}(XY) = (\mathbb{E}X)(\mathbb{E}Y)$ but X and Y are not independent

Properties (Cont.)

- (Lemma 1.1 in JS). Let $X_1, \ldots, X_n$ be independent random variables. Then random variables $g(X_1, \ldots, X_k)$ and $h(X_{k+1}, \ldots, X_n)$ are independent, where g and h are Borel functions and k is an integer between 1 and n.
- We usually use $X \perp Y$ to denote that X and Y are independent.
- If $X \perp Y$, $\mathbb{E}(XY) = (\mathbb{E}X)(\mathbb{E}Y)$.

Exercises:

- Find an example that $\mathbb{E}(XY) = (\mathbb{E}X)(\mathbb{E}Y)$ but X and Y are not independent
  - Let $X \sim \mathit{Unif}(-1, 1)$ and $Y = X^2$.

Properties (Cont.)

- (Lemma 1.1 in JS). Let $X_1, \ldots, X_n$ be independent random variables. Then random variables $g(X_1, \ldots, X_k)$ and $h(X_{k+1}, \ldots, X_n)$ are independent, where g and h are Borel functions and k is an integer between 1 and n.
- We usually use $X \perp Y$ to denote that X and Y are independent.
- If $X \perp Y$, $\mathbb{E}(XY) = (\mathbb{E}X)(\mathbb{E}Y)$.

Exercises:

- Find an example that $\mathbb{E}(XY) = (\mathbb{E}X)(\mathbb{E}Y)$ but X and Y are not independent
  - Let $X \sim \mathit{Unif}(-1, 1)$ and $Y = X^2$.
- If X, Y, Z are r.v.'s such that $X \perp Y$, $X \perp Z$ and $Z \perp Y$, can we conclude that X, Y, Z are independent? Why?

Properties (Cont.)

- (Lemma 1.1 in JS). Let $X_1, \ldots, X_n$ be independent random variables. Then random variables $g(X_1, \ldots, X_k)$ and $h(X_{k+1}, \ldots, X_n)$ are independent, where g and h are Borel functions and k is an integer between 1 and n.
- We usually use $X \perp Y$ to denote that X and Y are independent.
- If $X \perp Y$, $\mathbb{E}(XY) = (\mathbb{E}X)(\mathbb{E}Y)$.

Exercises:

- Find an example that $\mathbb{E}(XY) = (\mathbb{E}X)(\mathbb{E}Y)$ but X and Y are not independent
  - Let $X \sim Unif(-1, 1)$ and $Y = X^2$.
- If X, Y, Z are r.v.'s such that $X \perp Y$, $X \perp Z$ and $Z \perp Y$, can we conclude that X, Y, Z are independent? Why?
  - Let $P(X = 1) = P(X = -1) = 0.5$, let Y be independent with X and have the same distribution. Let $Z = XY$.

Properties (Cont.)

- (Lemma 1.1 in JS). Let $X_1, \ldots, X_n$ be independent random variables. Then random variables $g(X_1, \ldots, X_k)$ and $h(X_{k+1}, \ldots, X_n)$ are independent, where g and h are Borel functions and k is an integer between 1 and n.
- We usually use $X \perp Y$ to denote that X and Y are independent.
- If $X \perp Y$, $\mathbb{E}(XY) = (\mathbb{E}X)(\mathbb{E}Y)$.

Exercises:

- Find an example that $\mathbb{E}(XY) = (\mathbb{E}X)(\mathbb{E}Y)$ but X and Y are not independent
  - Let $X \sim Unif(-1, 1)$ and $Y = X^2$.
- If X, Y, Z are r.v.'s such that $X \perp Y$, $X \perp Z$ and $Z \perp Y$, can we conclude that X, Y, Z are independent? Why?
  - Let $P(X = 1) = P(X = -1) = 0.5$, let Y be independent with X and have the same distribution. Let $Z = XY$.
  - Can check that $P(X = 1, Y = 1) = P(Y = 1, Z = 1) = P(X = 1, Z = 1) = 0.5^2$, but $P(X = 1, Y = 1, Z = 1) = 0.25 \neq 0.5^3$

Conditional Expectation

Definition

- Let X be an integrable random variable on $(\Omega, \mathcal{F}, P)$.
- Let $\mathcal{A}$ be a sub-σ-field of $\mathcal{F}$.

The *conditional expectation* of X given $\mathcal{A}$, denoted by $E(X \mid \mathcal{A})$, is a random variable satisfying the following two conditions:

1. $E(X \mid \mathcal{A})$ is measurable from $(\Omega, \mathcal{A})$ to $(\mathcal{R}, \mathcal{B})$
2. $\int_C E(X \mid \mathcal{A})dP = \int_C XdP$ for any $C \in \mathcal{A}$

Such $E(X \mid \mathcal{A})$ exists and is unique.

- Uniqueness: Suppose both f and g satisfies the conditions, then $\mathbb{P}(f \neq g) = 0$.
- Note that if f is measurable w.r.t. $\mathcal{A}$ then it is also measurable w.r.t. $\mathcal{F}$ since $\mathcal{A} \subset \mathcal{F}$

Proof

- Define $\lambda(A) = \int_A f \, \mathrm{d}P$ for any $C \in \mathcal{A}$.
- λ is a measure on $(\Omega, \mathcal{A})$ and $\lambda \ll P|_{\mathcal{A}}$
  - $P|_{\mathcal{A}}$ is the *restriction of the measure P on $\mathcal{A}$*, meaning that it has the same image of P but is now only define on $\mathcal{A}$ rather than Ω
- Then $E(X \mid \mathcal{A}) = \frac{\mathrm{d}\lambda}{\mathrm{d}P|_{\mathcal{A}}}$ exists and is unique

Definition

- The conditional probability of $B \in \mathcal{F}$ given $\mathcal{A}$ is defined to be $P(B \mid \mathcal{A}) = E\left(I_B \mid \mathcal{A}\right)$
- Let Y be measurable from $(\Omega, \mathcal{F}, P)$ to $(\Lambda, \mathcal{G})$. The conditional expectation of X given Y is defined to be $E(X \mid Y) = E[X \mid \sigma(Y)]$

Definition

- The conditional probability of $B \in \mathcal{F}$ given $\mathcal{A}$ is defined to be $P(B \mid \mathcal{A}) = E\left(I_B \mid \mathcal{A}\right)$
- Let Y be measurable from $(\Omega, \mathcal{F}, P)$ to $(\Lambda, \mathcal{G})$. The conditional expectation of X given Y is defined to be $E(X \mid Y) = E[X \mid \sigma(Y)]$

Remark: $E(X \mid Y)$ is the "expectation" of X given the information provided by $\sigma(Y)$

Definition

- The conditional probability of $B \in \mathcal{F}$ given $\mathcal{A}$ is defined to be $P(B \mid \mathcal{A}) = E\left(I_B \mid \mathcal{A}\right)$
- Let Y be measurable from $(\Omega, \mathcal{F}, P)$ to $(\Lambda, \mathcal{G})$. The conditional expectation of X given Y is defined to be $E(X \mid Y) = E[X \mid \sigma(Y)]$

Remark: $E(X \mid Y)$ is the "expectation" of X given the information provided by $\sigma(Y)$

Lemma

Let Y be measurable from $(\Omega, \mathcal{F})$ to $(\Lambda, \mathcal{G})$ and Z a function from $(\Omega, \mathcal{F})$ to $\mathcal{R}^k$. If Z is Borel on $(\Omega, \sigma(Y))$, then there is a Borel function h from $(\Lambda, \mathcal{G})$ such that $Z = h \circ Y$

Definition

- The conditional probability of $B \in \mathcal{F}$ given $\mathcal{A}$ is defined to be $P(B \mid \mathcal{A}) = E\left(I_B \mid \mathcal{A}\right)$
- Let Y be measurable from $(\Omega, \mathcal{F}, P)$ to $(\Lambda, \mathcal{G})$. The conditional expectation of X given Y is defined to be $E(X \mid Y) = E[X \mid \sigma(Y)]$

Remark: $E(X \mid Y)$ is the "expectation" of X given the information provided by $\sigma(Y)$

Lemma

Let Y be measurable from $(\Omega, \mathcal{F})$ to $(\Lambda, \mathcal{G})$ and Z a function from $(\Omega, \mathcal{F})$ to $\mathcal{R}^k$. If Z is Borel on $(\Omega, \sigma(Y))$, then there is a Borel function h from $(\Lambda, \mathcal{G})$ such that $Z = h \circ Y$

- We can write $E(X \mid Y) = h \circ Y$ where h is a Borel function on $(\Lambda, \mathcal{G})$.
- Define $h(y)$ to be *the conditional expectation of X given $Y = y$*, and is denoted by $E(X \mid Y = y)$.

Properties of conditional expectation

- linearity: $\mathbb{E}(aX + bY \mid \mathcal{G}) = a\mathbb{E}(X \mid \mathcal{G}) + b\mathbb{E}(X \mid \mathcal{G})$ a.s.
- If $X = c$ a.s. for a constant c, then $\mathbb{E}(X \mid \mathcal{G}) = c$ a.s.
- monotonicity: if $X \leq Y$, then $\mathbb{E}(X \mid \mathcal{G}) \leq \mathbb{E}(Y \mid \mathcal{G})$ a.s.
- if $\mathcal{G} = \{\emptyset, \Omega\}$ (a trivial σ-field), then $\mathbb{E}(X \mid \mathcal{G}) = \mathbb{E}(X)$
- tower property: if $\mathcal{H} \subset \mathcal{G}$ is a σ-field, (so that $\mathcal{H} \subset \mathcal{G} \subset \mathcal{F}$), then

$$\mathbb{E}(X \mid \mathcal{H}) = \mathbb{E}\{\mathbb{E}(X \mid \mathcal{G}) \mid \mathcal{H}\}. \tag{15}$$

  - if $\mathcal{H} = \{\emptyset, \Omega\}$, then $\mathbb{E}(X) = \mathbb{E}\{\mathbb{E}(X \mid \mathcal{G})\}$.
- if $\sigma(Y) \subset \mathcal{G}$ and $\mathbb{E}|XY| < \infty$, then $\mathbb{E}(XY \mid \mathcal{G}) = Y\mathbb{E}(X \mid \mathcal{G})$
  - since $\sigma(Y) \subset \mathcal{G}$, information about Y is contained in $\mathcal{G}$, and thus, Y is kind of "known" given the information $\mathcal{G}$.
- if $\mathbb{E}X^2 < \infty$, then $\{\mathbb{E}(X \mid \mathcal{G})\}^2 \leq \mathbb{E}(X^2 \mid \mathcal{G})$ a.s.

Tutorial

1. Suppose that $\{f_n\}$ is a sequence of Borel functions on a measure space $(\Omega, \mathcal{F}, \nu)$ and $f_n(\omega) \to f(\omega)$ for $\omega \in \Omega$ with $\nu(\Omega) < \infty$. For any $\epsilon > 0$, define $A_k = \{x \in \Omega : |f_k(x) - f(x)| \geq \epsilon\}$, show that
$$\lim_{j \to \infty} \nu(\cup_{k=j}^{\infty} A_k) = 0. \tag{16}$$
2. Prove the Egoroff's theorem:
Suppose that $\{f_n\}$ is a sequence of Borel functions on a measure space $(\Omega, \mathcal{F}, \nu)$, $f_n(\omega) \to f(\omega)$ ν-a.e., and $\nu(\Omega) < \infty$. Show that for any $\eta > 0$, there is a $B \in \mathcal{F}$ with $\nu(B) < \eta$ such that $f_n(\omega) \to f(\omega)$ uniformly on B^c, i.e., for any small $\epsilon > 0$, one can find an $N_\eta \in \mathbb{N}$ such that $|f_n(\omega) - f(\omega)| < \epsilon$, $\forall n \geq N_\eta$ and $\forall \omega \in B^c$. lo
3. Prove the monotone convergence theorem:
If $0 \leq f_1 \leq \cdots$ and $\lim_n f_n = f$ a.e., then
$$\int \lim_n f_n \, \mathrm{d}\nu = \lim_n \int f_n \, \mathrm{d}\nu. \tag{17}$$

Ex 1

Suppose that $\{f_n\}$ is a sequence of Borel functions on a measure space $(\Omega, \mathcal{F}, \nu)$ and $f_n(\omega) \to f(\omega)$ for $\omega \in \Omega$ with $\nu(\Omega) < \infty$. For any $\epsilon > 0$, define $A_k = \{x \in \Omega : |f_k(x) - f(x)| \geq \epsilon\}$, show that

$$\lim_{j \to \infty} \nu(\cup_{k=j}^{\infty} A_k) = 0. \tag{18}$$

Ex 1

Suppose that $\{f_n\}$ is a sequence of Borel functions on a measure space $(\Omega, \mathcal{F}, \nu)$ and $f_n(\omega) \to f(\omega)$ for $\omega \in \Omega$ with $\nu(\Omega) < \infty$. For any $\epsilon > 0$, define $A_k = \{x \in \Omega : |f_k(x) - f(x)| \geq \epsilon\}$, show that

$$\lim_{j\to\infty} \nu(\cup_{k=j}^{\infty} A_k) = 0. \tag{18}$$

Proof: Note that $\cup_{k=j}^{\infty} A_k$ is decreasing (in j) and every $\omega \in \cap_{j=1}^{\infty} \cup_{k=j}^{\infty} A_k$, $f_n(\omega)$ does not converge to $f(\omega)$. So

$$\nu(\cap_{j=1}^{\infty} \cup_{k=j}^{\infty} A_k) = 0. \tag{19}$$

By monotonicity of measures and $\nu(\Omega) < \infty$, we have $\lim_{j\to\infty} \nu(\cup_{k=j}^{\infty} A_k) = 0$.

Ex 2

Suppose that $\{f_n\}$ is a sequence of Borel functions on a measure space $(\Omega, \mathcal{F}, \nu)$, $f_n(\omega) \to f(\omega)$ ν-a.e., and $\nu(\Omega) < \infty$. Show that for any $\eta > 0$, there is a $B \in \mathcal{F}$ with $\nu(B) < \eta$ such that $f_n(\omega) \to f(\omega)$ uniformly on B^c, i.e., for any small $\epsilon > 0$, one can find an $N_\eta \in \mathbb{N}$ such that $|f_n(\omega) - f(\omega)| < \epsilon$, $\forall n \geq N_\eta$ and $\forall \omega \in B^c$.

Ex 2

Suppose that $\{f_n\}$ is a sequence of Borel functions on a measure space $(\Omega, \mathcal{F}, \nu)$, $f_n(\omega) \to f(\omega)$ ν-a.e., and $\nu(\Omega) < \infty$. Show that for any $\eta > 0$, there is a $B \in \mathcal{F}$ with $\nu(B) < \eta$ such that $f_n(\omega) \to f(\omega)$ uniformly on B^c, i.e., for any small $\epsilon > 0$, one can find an $N_\eta \in \mathbb{N}$ such that $|f_n(\omega) - f(\omega)| < \epsilon$, $\forall n \geq N_\eta$ and $\forall \omega \in B^c$.
Proof:

- We will use Ex 1 with $\epsilon = 1/i$ and denote the sequence of set by $A_k(\frac{1}{i}), k = 1, 2, \ldots$.

Ex 2

Suppose that $\{f_n\}$ is a sequence of Borel functions on a measure space $(\Omega, \mathcal{F}, \nu)$, $f_n(\omega) \to f(\omega)$ ν-a.e., and $\nu(\Omega) < \infty$. Show that for any $\eta > 0$, there is a $B \in \mathcal{F}$ with $\nu(B) < \eta$ such that $f_n(\omega) \to f(\omega)$ uniformly on B^c, i.e., for any small $\epsilon > 0$, one can find an $N_\eta \in \mathbb{N}$ such that $|f_n(\omega) - f(\omega)| < \epsilon$, $\forall n \geq N_\eta$ and $\forall \omega \in B^c$.

Proof:

- We will use Ex 1 with $\epsilon = 1/i$ and denote the sequence of set by $A_k(\frac{1}{i}), k = 1, 2, \ldots$.
- For any given $\eta > 0$, for each $i \in \mathbb{N}$, we can find a j_i such that

$$\nu\left(\cup_{k=j_i}^{\infty} A_k(\frac{1}{i})\right) < \frac{\eta}{2^i}. \tag{20}$$

Ex 2

Suppose that $\{f_n\}$ is a sequence of Borel functions on a measure space $(\Omega, \mathcal{F}, \nu)$, $f_n(\omega) \to f(\omega)$ ν-a.e., and $\nu(\Omega) < \infty$. Show that for any $\eta > 0$, there is a $B \in \mathcal{F}$ with $\nu(B) < \eta$ such that $f_n(\omega) \to f(\omega)$ uniformly on B^c, i.e., for any small $\epsilon > 0$, one can find an $N_\eta \in \mathbb{N}$ such that $|f_n(\omega) - f(\omega)| < \epsilon$, $\forall n \geq N_\eta$ and $\forall \omega \in B^c$.

Proof:

- We will use Ex 1 with $\epsilon = 1/i$ and denote the sequence of set by $A_k(\frac{1}{i}), k = 1, 2, \ldots$.
- For any given $\eta > 0$, for each $i \in \mathbb{N}$, we can find a j_i such that

$$\nu\left(\cup_{k=j_i}^{\infty} A_k(\frac{1}{i})\right) < \frac{\eta}{2^i}. \tag{20}$$

- Let $B_\eta = \cup_{i=1}^{\infty} \cup_{k=j_i}^{\infty} A_k(\frac{1}{i})$, then

$$\nu(B_\eta) \leq \sum_{i=1}^{\infty} \frac{\eta}{2^i} = \eta$$

Ex 2

Suppose that $\{f_n\}$ is a sequence of Borel functions on a measure space $(\Omega, \mathcal{F}, \nu)$, $f_n(\omega) \to f(\omega)$ ν-a.e., and $\nu(\Omega) < \infty$. Show that for any $\eta > 0$, there is a $B \in \mathcal{F}$ with $\nu(B) < \eta$ such that $f_n(\omega) \to f(\omega)$ uniformly on B^c, i.e., for any small $\epsilon > 0$, one can find an $N_\eta \in \mathbb{N}$ such that $|f_n(\omega) - f(\omega)| < \epsilon$, $\forall n \geq N_\eta$ and $\forall \omega \in B^c$.

Proof:

- We will use Ex 1 with $\epsilon = 1/i$ and denote the sequence of set by $A_k(\frac{1}{i}), k = 1, 2, \ldots$.
- For any given $\eta > 0$, for each $i \in \mathbb{N}$, we can find a j_i such that

$$\nu\left(\cup_{k=j_i}^{\infty} A_k(\frac{1}{i})\right) < \frac{\eta}{2^i}. \tag{20}$$

- Let $B_\eta = \cup_{i=1}^{\infty} \cup_{k=j_i}^{\infty} A_k(\frac{1}{i})$, then

$$\nu(B_\eta) \leq \sum_{i=1}^{\infty} \frac{\eta}{2^i} = \eta$$

- We can check that f_n's uniformly converges to f on B^c.

Ex 3

Prove the monotone convergence theorem:
If $0 \leq f_1 \leq \cdots$ and $\lim_n f_n = f$ a.e., then

$$\int \lim_n f_n \, \mathrm{d}\nu = \lim_n \int f_n \, \mathrm{d}\nu. \tag{21}$$

Ex 3

Prove the monotone convergence theorem:
If $0 \leq f_1 \leq \cdots$ and $\lim_n f_n = f$ a.e., then

$$\int \lim_n f_n \,\mathrm{d}\nu = \lim_n \int f_n \,\mathrm{d}\nu. \tag{21}$$

Proof: We can prove this result using the same argument in Ex 3 in Lecture 1's tutorial.