ST5215 Advanced Statistical Theory, Lecture 7

HUANG Dongming

National University of Singapore

3 Sep 2020

Overview

Last time

- Sufficiency
- Factorization theorem

Today

- Minimal sufficiency

P6 in Lect 5: Existence of conditional distributions

Theorem

Suppose

- X *is a random n-vector on a probability space* $(\Omega, \mathcal{F}, P)$ *, and*
- Y *is measurable from* $(\Omega, \mathcal{F})$ *to* $(\Lambda, \mathcal{G})$.

Then there exists a function $P_{X|Y}(B \mid y)$ *on* $\mathcal{B}^n \times \Lambda$ *such that*

1. $P_{X|Y}(\cdot \mid y)$ *is a probability measure on* $(\mathcal{R}^n, \mathcal{B}^n)$ *for any fixed* $y \in \Lambda$,
2. $P_{X|Y}(B \mid y) = P[X \in B \mid Y = y]$ *a.s.* P_Y *for any fixed* $B \in \mathcal{B}^n$.

Remark.

- By definition, for any $B \in \mathcal{B}^n$, $P[X \in B \mid Y] := E[I_{X \in B} \mid Y]$ is a random variable on $(\Omega, \sigma(Y))$ and can be represented as $h_B(Y)$.
- This theorem ensures that for almost every fixed $y \in \Lambda$, we can find a probability measure $P_{X|Y}(\cdot \mid y)$ such that the equation $P_{X|Y}(B \mid y) = h_B(y)$ holds.

P8 in Lect 4: Properties of conditional expectation

All r.v.s. are integrable on the probability space $(\Omega, \mathcal{F}, P)$, and $\mathcal{G}$ is a sub-σ-field of $\mathcal{F}$.

- 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 $\mathbb{E}X^2 < \infty$, then $\{\mathbb{E}(X \mid \mathcal{G})\}^2 \leq \mathbb{E}(X^2 \mid \mathcal{G})$ a.s.
- (Fatou's lemma). If $X_n \geq 0$ for any n, then $\mathbb{E}\left(\liminf_n X_n \mid \mathcal{G}\right) \leq \liminf_n \mathbb{E}\left(X_n \mid \mathcal{G}\right)$ a.s.
- (Dominated convergence theorem). If $|X_n| \leq Y$ for any n and $X_n \rightarrow_{\text{a.s.}} X$, then $\mathbb{E}\left(X_n \mid \mathcal{G}\right) \rightarrow_{\text{a.s.}} \mathbb{E}(X \mid \mathcal{G})$
- All the integral-inequalities we saw before have conditional versions

P12 in Lecture 4: Conditional p.d.f.

Theorem

Suppose

- X *is a random n-vector,* Y *is a random m-vector*
- (X, Y) *has a p.d.f.* $f(x, y)$ *w.r.t.* $\nu \times \lambda$ *(*ν *on* $\mathcal{B}^n$, λ *on* $\mathcal{B}^m$*, both* σ*-finite).*

Let $f_Y(y) = \int f(x, y) d\nu(x)$ *be the marginal p.d.f. of* Y *w.r.t.* λ*, and* $A = \{y \in \mathcal{R}^m : f_Y(y) > 0\}$.
Then for any fixed $y \in A$*, the p.d.f. of* $P_{X|Y=y}$ *w.r.t.* ν *is given by*

$$f_{X|Y}(x \mid y) = \frac{f(x, y)}{f_Y(y)}. \tag{1}$$

Furthermore, if $g(x, y)$ *is a Borel function on* $\mathcal{R}^{n+m}$ *and* $\mathbb{E}|g(X, Y)| < \infty$*, then*

$$\mathbb{E}[g(X, Y) \mid Y] = \int g(x, Y) f_{X|Y}(x \mid Y) \, \mathrm{d}\nu(x), \text{a.s.} \tag{2}$$

The conditional p.d.f. is given by $f_{X|Y}$

- After we proved the second equation, we can claim that **the p.d.f. of $P_{X|Y=y}$ is given by the first equation**
- Need to show: for any $B \in \mathcal{B}^n$, $P_{X|Y=y}(B) = \int_B f_{X|Y}(x \mid y)\, \mathrm{d}\nu(x)$
- Define $g(x, y) = I_B(x)$
- Define $\psi_B(y) = \int_B f_{X|Y}(x \mid y)\, \mathrm{d}\nu(x)$
- The theorem's second equation implies $E(I_B(X) \mid Y) = \psi_B(Y)$ a.s.
- By the definition of $P(X \in B \mid Y = y)$, we have $P(X \in B \mid Y = y) = \psi_B(y)$
- By the definition of $P_{X|Y=y}(B)$, we have $P_{X|Y=y}(B) = P(X \in B \mid Y = y) = \psi_B(y)$, where RHS equals to $\int_B f_{X|Y}(x \mid y)\, \mathrm{d}\nu(x)$
- Since this holds for any $B \in \mathcal{B}^n$, we conclude that the RN derivative of $P_{X|Y=y}(\cdot)$ is $f_{X|Y}(\cdot \mid y)$.

Recap: Sufficiency and factorization

Let X be a sample from an unknown population $P \in \mathcal{P}$, where $\mathcal{P}$ is a family of populations.

- A statistic $T(X)$ is said to be *sufficient* for $P \in \mathcal{P}$ iff the conditional distribution of X given T is known (does not depend on P)
- *Factorization Theorem*: Further suppose that $\mathcal{P}$ is a family of probability measures on $(\mathcal{R}^n, \mathcal{B}^n)$ dominated by a σ-finite measure ν.

  Then a statistic $T(X)$ is sufficient for $P \in \mathcal{P}$ **if and only if** there are nonnegative Borel functions $h(x)$ on $(\mathcal{R}^n, \mathcal{B}^n)$ and $g_P(t)$ on the range of T such that

$$\frac{\mathrm{d}P}{\mathrm{d}\nu}(x) = g_P(T(x))h(x).$$

Sufficiency of Order statistics

Let $X = (X_1, \ldots, X_n)$ be i.i.d. according to an unknown continuous distribution F

- By the continuity assumptions, the X_i's are distinct with probability 1 (Left for exercise)
- Let $T = \left(X_{(1)}, \ldots, X_{(n)}\right)$ where $X_{(1)} < \cdots < X_{(n)}$ denotes the ordered observations
- Given T, the only possible values for X are the $n!$ vectors $\left(X_{(i_1)}, \cdots, X_{(i_n)}\right)$, where $i_1, \ldots, i_n$ is a permutation of $1, \ldots, n$
- By symmetry, each of these has conditional probability $1/n!$
- This conditional distribution is independent of F, so T is sufficient
- *Generating fake samples:* A random vector $\tilde{X}$ with the same distribution as X can be obtained from T by labeling the n coordinates of T at random

Exercise

Let X be a sample from $P \in \mathcal{P}$, where $\mathcal{P}$ is a family of distributions on the Borel σ -field on $\mathcal{R}^n$.

Show that if $T(X)$ is a sufficient statistic for $P \in \mathcal{P}$ and $T = \psi(S)$, where ψ is measurable and $S(X)$ is another statistic, then $S(X)$ is sufficient for $P \in \mathcal{P}$

Solution.

First suppose $\mathcal{P}$ is dominated by a σ-finite measure ν.

- By the factorization theorem,

$$\frac{\mathrm{d}P}{\mathrm{d}\nu}(x) = g_P(T(x))h(x)$$

  where h is a Borel function of x and $g_P(t)$ is a Borel function of t

- If $T = \psi(S)$, then

$$\frac{\mathrm{d}P}{\mathrm{d}\nu}(x) = g_P(\psi(S(x)))h(x)$$

  and, by the factorization theorem again, $S(X)$ is sufficient for $P \in \mathcal{P}$

Exercise (Cont.)

Now consider the general case.

- Suppose that $S(X)$ is not sufficient for $P \in \mathcal{P}$, i.e., there exist at least two measures P_1 and $P_2 \in \mathcal{P}$ s.t. the conditional distributions of X given $S(X)$ under P_1 and P_2 are different $\quad (*)$
- Let $\mathcal{P}_0 = \{P_1, P_2\}$. It is dominated by the measure $(P_1 + P_2)$
- Since $T(X)$ is sufficient for $P \in \mathcal{P}$, it is also sufficient for $P \in \mathcal{P}_0$ because $\mathcal{P}_0 \subset \mathcal{P}$
- By the previously proved result, $S(X)$ is sufficient for $P \in \mathcal{P}_0$
- Hence, the conditional distributions of X given $S(X)$ under P_1 and P_2 are the same, which contradicts with $(*)$

Information reduction

- If $T(X) = \psi(S(X))$ and T is sufficient, then S is also sufficient
  - Knowledge of S implies knowledge of T and hence permits reconstruction of the original data
  - T provides a greater reduction of the data than S unless ψ is one-to-one
- Can we find the sufficient statistic that provides the greatest reduction?

Definition (Minimal sufficiency)

Let T be a sufficient statistic for $P \in \mathcal{P}$. T is called a minimal sufficient statistic if and only if, for any other statistic S sufficient for $P \in \mathcal{P}$, there is a measurable function ψ such that $T = \psi(S)$ $\mathcal{P}$-a.s.

- Notation: If a statement holds except for outcomes in an event A satisfying $P(A) = 0$ for all $P \in \mathcal{P}$, then we say that the statement holds $\mathcal{P}$-a.s.

- Minimal sufficient statistics are **unique** (almost surely)
  - If both T and S are minimal sufficient statistics (for a family $\mathcal{P}$), then there is a one-to-one measurable function ψ such that $T = \psi(S)$ $\mathcal{P}$-a.s.
- Minimal sufficient statistics **exist** under weak conditions
  - e.g., $\mathcal{P}$ contains distributions on $\mathcal{R}^n$ dominated by a σ-finite measure (Bahadur, 1957)

Theorems for finding minimal sufficient statistics

We have several useful tools for checking minimal sufficiency.
In the following slides, $\mathcal{P}$ is a family of distributions on $\mathcal{R}^n$.

Theorem (A)

Suppose $\mathcal{P}_0 \subset \mathcal{P}$ and $\mathcal{P}_0$-a.s. implies $\mathcal{P}$-a.s.
If T is sufficient for $P \in \mathcal{P}$ and minimal sufficient for $P \in \mathcal{P}_0$, then T is minimal sufficient for $P \in \mathcal{P}$.

Proof:

- If S is sufficient for $\mathcal{P}$, then it is also sufficient for $\mathcal{P}_0$.
- Thus, $T = \psi(S)$ $\mathcal{P}_0$-a.s. for a measurable function ψ
- Then $T = \psi(S)$ $\mathcal{P}$-a.s. since $\mathcal{P}_0$-a.s. implies $\mathcal{P}$-a.s by assumption.

Theorem (B)

Suppose that $\mathcal{P}$ contains p.d.f.'s $f_0, f_1, \ldots$ w.r.t. a σ-finite measure.

- *Define $f_\infty(x) = \sum_{i=0}^{\infty} c_i f_i(x)$, where $c_i > 0$ and $\sum_{i=0}^{\infty} c_i = 1$. Let $T_i(x) = f_i(x)/f_\infty(x)$ when $f_\infty(x) > 0$. Then $T(X) = (T_0(X), T_1(X), \ldots)$ is minimal sufficient for $\mathcal{P}$.*
- *If $\{x : f_i(x) > 0\} \subset \{x : f_0(x) > 0\}$ for all i, then $T(X) = (f_1(x)/f_0(x), f_2(x)/f_0(x), \ldots)$ is minimal sufficient for $\mathcal{P}$.*

Proof: The argument for the second case is the same.

- The construction of f_∞ assures that $f_\infty(X) > 0$ $\mathcal{P}$-a.s.
- Let $g_i(T) = T_i$. Then $f_i(x) = g_i(T(x))f_\infty(x)$. By Factorization theorem, T is sufficient for $\mathcal{P}$.
- Suppose $S(X)$ is another sufficient statistic. By Factorization theorem, $f_i(x) = \tilde{g}_i(S(x))h(x)$ for all i and some $\tilde{g}_i$ and h.
- Then $T_i(x) = \tilde{g}_i(S(x))/\sum_{j=0}^{\infty} c_j \tilde{g}_j(S(x))$ when $f_\infty(x) > 0$.
- Thus, T is minimal sufficient for $\mathcal{P}$.

Theorem (C)

Suppose that $\mathcal{P}$ contains p.d.f.'s f_P w.r.t. a σ-finite measure ν and that there exists a sufficient statistic $T(X)$ such that for any possible values x and y of X and a measurable function ϕ,

$$f_P(x) = f_P(y)\phi(x, y), \ \forall P \in \mathcal{P} \Rightarrow T(x) = T(y). \tag{3}$$

Then $T(X)$ is minimal sufficient for $\mathcal{P}$.

Proof: See the textbook (Theorem 2.3.iii). Here is the basic idea.

- Suppose S is sufficient. We need to show T is a function of S.
- By factorization theorem, we have $f_P(x) = g_P(S(x))h(x)$, ν-a.e. , for any $P \in \mathcal{P}$.
- Let $A = \{x : h(x) > 0\}$. Then $P(A^c) = 0$ for any $P \in \mathcal{P}$.
- For any two values x and y in A s.t. $S(x) = S(y)$, for all $P \in \mathcal{P}$,

$$f_P(x) = g_P(S(x))h(x) = g_P(S(y))h(y) \times \frac{h(x)}{h(y)} = f_P(y) \times \frac{h(x)}{h(y)}, \tag{4}$$

  which implies $T(x) = T(y)$. So $T(x) = \psi(S(x))$, $\mathcal{P}$-a.s.

Example: Normal minimal sufficient statistic

- Let $X_1, \ldots, X_n$ be iid $N\left(\mu, \sigma^2\right)$, both μ and σ^2 unknown.
- Let $\mathbf{x}$ and $\mathbf{y}$ denote two sample points
- Let $\left(\bar{x}, s_{\mathbf{x}}^2\right)$ and $\left(\bar{y}, s_{\mathbf{y}}^2\right)$ be the sample means and variances corresponding to the $\mathbf{x}$ and $\mathbf{y}$ samples, respectively
- The ratio of densities is

$$\frac{f\left(\mathbf{x} \mid \mu, \sigma^2\right)}{f\left(\mathbf{y} \mid \mu, \sigma^2\right)} = \frac{\left(2\pi\sigma^2\right)^{-n/2} \exp\left(-\left[n(\bar{x}-\mu)^2 + (n-1)s_{\mathbf{x}}^2\right] / \left(2\sigma^2\right)\right)}{\left(2\pi\sigma^2\right)^{-n/2} \exp\left(-\left[n(\bar{y}-\mu)^2 + (n-1)s_{\mathbf{y}}^2\right] / \left(2\sigma^2\right)\right)}$$

$$= \exp\left(\left(2\sigma^2\right)^{-1}\left[-n\left(\bar{x}^2 - \vec{y}^2\right) + 2n\mu(\bar{x}-\bar{y}) - \right.\right.$$

$$\left.\left.(n-1)\left(s_{\mathbf{x}}^2 - s_{\mathbf{y}}^2\right)\right]\right)$$

- This ratio will be constant as a function of μ and σ^2 if and only if $\bar{x} = \bar{y}$ and $s_{\mathbf{x}}^2 = s_{\mathbf{y}}^2$
- Thus, by Theorem C, $\left(\bar{X}, S^2\right)$ is a minimal sufficient statistic for $\left(\mu, \sigma^2\right)$

Example: Exponential family

Let $\mathcal{P} = \{f_\theta : \theta \in \Theta\}$ be a p-dimensional exponential family with p.d.f.'s

$$f_\theta(x) = \exp\{[\eta(\theta)]^\tau T(x) - \xi(\theta)\}h(x).$$

- By Factorization Theorem, $T(X)$ is sufficient for $\theta \in \Theta$
- If there exists $\Theta_0 = \{\theta_0, \theta_1, \ldots, \theta_p\} \subset \Theta$ such that the vectors $\eta_i = \eta(\theta_i) - \eta(\theta_0)$, $i = 1, \ldots, p$, are linearly independent in $\mathcal{R}^p$,
  - This is true if the corresponding natural exponential family is of full rank
- then T is also minimal sufficient
  - Method 1: combining Theorems A and B
  - Method 2: using Theorem C

Example (Cont.): Method 1

Let $\mathcal{P} = \{f_\theta : \ \theta \in \Theta\}$ be a p-dimensional exponential family with p.d.f.'s

$$f_\theta(x) = \exp\{[\eta(\theta)]^\tau T(x) - \xi(\theta)\}h(x).$$

If there exists $\Theta_0 = \{\theta_0, \theta_1, \ldots, \theta_p\} \subset \Theta$ such that the vectors $\eta_i = \eta(\theta_i) - \eta(\theta_0)$, $i = 1, \ldots, p$, are linearly independent in $\mathcal{R}^p$, then T is minimal sufficient.

- Let $\mathcal{P}_0 = \{f_\theta : \ \theta \in \Theta_0\}$
- Note that the set $\{x : \ f_\theta(x) > 0\}$ does not depend on θ
- It follows from Theorem B with f_{θ_0} that

$$S(X) = \Big(\exp\{\eta_1^\tau T(x) - \xi_1\}, \ldots, \exp\{\eta_p^\tau T(x) - \xi_p\}\Big)$$

  is minimal sufficient for $\theta \in \Theta_0$
- Since η_i's are linearly independent, there is a one-to-one measurable function ψ such that $T(X) = \psi(S(X))$ a.s. $\mathcal{P}_0$
- Hence, T is minimal sufficient for $\theta \in \Theta_0$
- It is easy to see that $\mathcal{P}_0$-a.s. implies $\mathcal{P}$-a.s.
- By Theorem A, T is minimal sufficient for $\theta \in \Theta$.

Example (Cont.): Method 2

Let $\mathcal{P} = \{f_\theta : \ \theta \in \Theta\}$ be a p-dimensional exponential family with p.d.f.'s

$$f_\theta(x) = \exp\{[\eta(\theta)]^\tau T(x) - \xi(\theta)\}h(x).$$

If there exists $\Theta_0 = \{\theta_0, \theta_1, \ldots, \theta_p\} \subset \Theta$ such that the vectors $\eta_i = \eta(\theta_i) - \eta(\theta_0)$, $i = 1, \ldots, p$, are linearly independent in $\mathcal{R}^p$, then T is minimal sufficient.

- If there exists a function $\phi(x, y)$, such that any $x, y \in \{x : h(x) > 0\}$

  $$f_\theta(x) = f_\theta(y)\phi(x, y), \forall\theta \in \Theta,$$

  then by restricting on Θ_0, we have

  $$\exp\{[\eta(\theta_i) - \eta(\theta_0)]^\tau [T(x) - T(y)\} = 1, \forall i = 1, \ldots, p,$$

  which implies $T(x) = T(y)$ because $\{\eta_i\}_{i=1}^p$ are linearly independent
- Since T is sufficient, by Theorem C, T is minimal sufficient.

Example: Uniform with Varying Location

Let $X_1, \ldots, X_n \sim P_\theta = U(\theta, \theta+1)$ for $\theta \in \mathcal{R}$, where $n > 1$.

- This is a location family, with the location parameter θ
- The joint Lebesgue p.d.f. is

$$f_\theta(x) = \prod_{i=1}^{n} I_{(\theta,\theta+1)}(x_i) = I_{(x_{(n)}-1, x_{(1)})}(\theta), \qquad x = (x_1, \ldots, x_n) \in \mathcal{R}^n.$$

  - because $\theta < x_{(1)} \le \cdots \le x_{(n)} < \theta + 1 \Leftrightarrow I_{(x_{(n)}-1, x_{(1)})}(\theta) = 1$
- By Factorization Theorem, $T = (X_{(1)}, X_{(n)})$ is sufficient for θ

We will show that $T = (X_{(1)}, X_{(n)})$ is minimal sufficient by using Theorem C. As an exercise, you can also try to use Theorems A+B to prove this result.

Example: Uniform with Varying Location (Cont.)

$$f_\theta(x) = \prod_{i=1}^{n} I_{(\theta,\theta+1)}(x_i) = I_{(x_{(n)}-1, x_{(1)})}(\theta), \qquad x = (x_1, \dots, x_n) \in \mathcal{R}^n.$$

- For any x, y in the possible range of sample, and ψ measurable such that

$$f_\theta(x) = f_\theta(y)\psi(x, y), \quad \forall \theta \in \mathcal{R} \tag{5}$$

- If $x_{(1)} < y_{(1)}$
  - We can choose θ such that $x_{(1)} < \theta < y_{(1)}$ and $y_{(n)} < \theta + 1$
  - Then$f_\theta(x) = 0$ and $f_\theta(y) = 1$. This shows $\psi(x, y) = 0$
  - Then (5) implies that $f_\theta(x) = 0$ for all $\theta \in \mathcal{R}$, which cannot be true.
- So $x_{(1)} \geq y_{(1)}$. Similarly, $x_{(n)} \leq y_{(n)}$.
- If $x_{(1)} > y_{(1)}$
  - We can choose θ such that $y_{(1)} < \theta < x_{(1)}$ and $x_{(n)} < \theta + 1$,
  - Then $f_\theta(x) = 1$ and $f_\theta(y) = 0$, which contradicts (5).
- So $x_{(1)} \leq y_{(1)}$. Similarly, $x_{(n)} \geq y_{(n)}$
- So (5) implies $T(x) = T(y)$. By Theorem C, T is minimal sufficient

Tutorial: P1$\sim$3 in Homework 1

1. If $f : \mathcal{R} \mapsto \mathcal{R}$ is a continuous function, then it is Borel measurable
2. Ex 1.6.12 in JS
   Let ν and μ be two measures on $(\Omega, \mathcal{F})$ such that $\nu(A) = \mu(A)$ for any $A \in \mathcal{C}$, where $\mathcal{C} \subset \mathcal{F}$ and $\mathcal{C}$ is a π-system (i.e., if A and B are in $\mathcal{C}$, then so is $A \cap B$). Assume that there are $A_i \in \mathcal{C}, i = 1, 2, \ldots,$ such that $\cup A_i = \Omega$ and $\nu(A_i) < \infty$ for all i. Show that $\nu(A) = \mu(A)$ for any $A \in \sigma(\mathcal{C})$. This proves the uniqueness part of Proposition 1.3 . (Hint: show that $\{A \in \sigma(\mathcal{C}) : \nu(A) = \mu(A)\}$ is a σ -field.)
3. Ex 1.6.23 in JS
   Let $\nu_i, i = 1, 2,$ be measures on $(\Omega, \mathcal{F})$ and f be Borel. Show that
   $$\int f d(\nu_1 + \nu_2) = \int f d\nu_1 + \int f d\nu_2$$
   i.e., if either side of the equality is well defined, then so is the other side, and the two sides are equal.

Ex 1

If $f : \mathcal{R} \mapsto \mathcal{R}$ is a continuous function, then it is Borel measurable
Proof:

- We use Tutorial Problem 2 in Lecture 2.

Proposition

Let $(\Omega, \mathcal{F})$ be a measurable space. $f : \Omega \mapsto \mathcal{R}$ is Borel if and only if $f^{-1}(a, \infty) \in \mathcal{F}$ for all $a \in \mathcal{R}$.

- If f is continuous, then $f^{-1}(a, \infty)$ is an open subset of $\mathcal{R}$, which is in $\mathcal{B}$. Therefore, f is Borel measurable.

Ex 1.6.12 in JS

Let ν and μ be two measures on $(\Omega, \mathcal{F})$ such that $\nu(A) = \mu(A)$ for any $A \in \mathcal{C}$, where $\mathcal{C} \subset \mathcal{F}$ and $\mathcal{C}$ is a π-system (i.e., if A and B are in $\mathcal{C}$, then so is $A \cap B$). Assume that there are $A_i \in \mathcal{C}, i = 1, 2, \ldots,$ such that $\cup A_i = \Omega$ and $\nu(A_i) < \infty$ for all i. Show that $\nu(A) = \mu(A)$ for any $A \in \sigma(\mathcal{C})$.

Remark. This result is often used to prove the uniqueness of measures. To prove this result, we first introduce the notion of λ-*system*

Definition

A collection $\mathcal{L}$ of subsets of Ω is a λ-*system* if

1. $\Omega \in \mathcal{L}$
2. If $A, B \in \mathcal{L}$ and $A \subset B$, then $B \setminus A \in \mathcal{L}$
3. If $A_n \in \mathcal{L}$ for all n and $A_n \uparrow$, then $\cup_n A_n \in \mathcal{L}$

It is easy to see that a σ-field is a λ-system. The reverse is not true.

Proof of Ex 2 by Dynkin's π-λ Theorem

We first solve the homework by using the following result.

Theorem (Dynkin's π-λ Theorem)

Suppose $\mathcal{C}$ is a π-system and $\mathcal{L}$ is a λ system that contains $\mathcal{C}$, then $\sigma(\mathcal{C}) \subset \mathcal{L}$.

- Fixed any $B \in \mathcal{C}$ with $\nu(B) < \infty$. Define

$$\mathcal{G}_B = \{A \in \mathcal{F} : \nu(A \cap B) = \mu(A \cap B)\}. \tag{6}$$

  Then $\mathcal{C} \subset \mathcal{G}_B$.
- By additivity and monotonicity of measures, we can see that $\mathcal{G}_B$ is a λ-system.
- By the theorem, we have $\sigma(\mathcal{C}) \subset \mathcal{G}_B$.
- Since B is arbitrary, we conclude that for any $A \in \sigma(\mathcal{C})$ and any $B \in \mathcal{C}$ with $\nu(B) < \infty$, $\nu(A \cap B) = \mu(A \cap B)$

- Now fixed any $B \in \sigma(\mathcal{C})$ with $\nu(B) < \infty$. We can apply the same argument before to conclude that for any $A \in \sigma(\mathcal{C})$,
  $\nu(A \cap B) = \mu(A \cap B)$ $\quad (*)$
- Let A_i's be the sequence given in the statement of the exercise. Define $B_n = \cup_{i=1}^n A_i$. Then $B_n \in \sigma(\mathcal{C})$, $B_n \uparrow \Omega$, and $\nu(B_n) \leq \sum_{i=1}^n \nu(A_i) < \infty$.
- For any $A \in \sigma(\mathcal{C})$,

$$
\begin{aligned}
\nu(A) &= \lim_n \nu(A \cap B_n) && (7)\\
&= \lim_n \mu(A \cap B_n) && (8)\\
&= \mu(B_n), && (9)
\end{aligned}
$$

  where the first and third equalities are due to monotonicity of measures, and the second is because $B_n \in \sigma(C)$, $\nu(B_n) < \infty$, $A \in \sigma(C)$, and the result $(*)$

"σ-field $\Leftrightarrow$ (π & λ) system"

To prove the theorem, we need the following result.

Proposition

If $\mathcal{A}$ is a π-system and is a λ-system, then it is a σ-field.

- Easy to see that $\Omega \in \mathcal{A}$ and if $B \in \mathcal{A}$ then $B^c = \Omega \setminus B \in \mathcal{A}$ by definition of λ-system
- For a sequence of $B_n \in \mathcal{A}$, then

$$C_n := \bigcup_{i=1}^{n} B_i = \left(\bigcap_{i=1}^{n} B_i^c \right)^c \in \mathcal{A},$$

and C_n is increasing. So $\bigcup\limits_{n=1}^{\infty} B_n = \bigcup\limits_{n=1}^{\infty} C_n \in \mathcal{A}$

We now prove the theorem

Theorem (Dynkin's π-λ Theorem)

Suppose $\mathcal{C}$ is a π-system and $\mathcal{L}$ is a λ system that contains $\mathcal{C}$, then $\sigma(\mathcal{C}) \subset \mathcal{L}$.

- The smallest λ-system that contains $\mathcal{C}$ (a.k.a. *the lambda-system generated by* $\mathcal{C}$), denoted by $\lambda(\mathcal{C})$, is defined as the the intersection of all λ-systems that contains $\mathcal{C}$, i.e.,

$$\lambda(\mathcal{C}) = \bigcap_{\substack{\mathcal{A}:\mathcal{A}\supset\mathcal{C},\\ \mathcal{A} \text{ is a } \lambda\text{-system}}} \mathcal{A}. \tag{10}$$

- Easy to check that $\lambda(\mathcal{C})$ is a λ-system and is contained by any λ-system that contains $\mathcal{C}$
- We will show that $\sigma(\mathcal{C}) = \lambda(\mathcal{C})$
- $\sigma(\mathcal{C}) \supset \lambda(\mathcal{C})$ because a σ-field is a λ-system
- If we can **show** $\lambda(\mathcal{C})$ **is a** π**-system**, then by the last proposition, $\lambda(\mathcal{C})$ is a σ-field, and thus $\sigma(\mathcal{C}) \subset \lambda(\mathcal{C})$

Showing $\lambda(\mathcal{C})$ is a π-system

- Fix any $A \in \mathcal{C}$, define

$$\mathcal{L}_A := \{B \in \lambda(\mathcal{C}) : A \cap B \in \lambda(\mathcal{C})\}. \tag{11}$$

- We can check that $\mathcal{L}_A$ is a λ-system and $\mathcal{C} \subset \mathcal{L}_A$. So $\lambda(\mathcal{C}) \subset \mathcal{L}_A$

We conclude that if $A \in \mathcal{C}$ and $B \in \lambda(\mathcal{C})$ then $A \cap B \in \lambda(\mathcal{C})$.

- Fix any $B \in \lambda(\mathcal{C})$, let

$$\mathcal{G}_B = \{A \in \lambda(\mathcal{C}) : A \cap B \in \lambda(\mathcal{C})\}. \tag{12}$$

- We can check that $\mathcal{G}_B$ is a λ-system, and $\mathcal{C} \subset \mathcal{G}_B$. So $\lambda(\mathcal{C}) \subset \mathcal{G}_B$

We conclude that if $A \in \lambda(\mathcal{C})$ and $B \in \lambda(\mathcal{C})$ then $A \cap B \in \lambda(\mathcal{C})$. That is, $\lambda(\mathcal{C})$ is a π-system.

Ex 1.6.23 in JS

Let $\nu_i, i = 1, 2$, be measures on $(\Omega, \mathcal{F})$ and f be Borel. Show that

$$\int f d(\nu_1 + \nu_2) = \int f d\nu_1 + \int f d\nu_2$$

i.e., if either side of the equality is well defined, then so is the other side, and the two sides are equal.

Proof: Use canonical method

- The case of simple functions is straightforward to prove by using the definition
- The case of nonnegative functions can be proved using approximation of simple functions
- For general Borel function f, since the LHS is well-defined, we know that $\int f_+ d(\nu_1 + \nu_2)$ or $\int f_- d(\nu_1 + \nu_2)$ is finite
  - WLOG, assume $\int f_- d(\nu_1 + \nu_2)$ is finite
  - By the result for nonnegative functions, we have $\int f_- d(\nu_1 + \nu_2) = \int f_- d\nu_1 + \int f_- d\nu_2$, so both $\int f_- d\nu_1$ and $\int f_- d\nu_2$ are finite.

Ex 1.6.23 in JS (Cont.)

Therefore,

$$\begin{aligned}\int fd\,(\nu_1+\nu_2) &= \int f_+ d\,(\nu_1+\nu_2) - \int f_- d\,(\nu_1+\nu_2)\\ &= \int f_+ d\nu_1 + \int f_+ d\nu_2 - (\int f_- d\nu_1 + \int f_- d\nu_2)\\ &= (\int f_+ d\nu_1 - \int f_- d\nu_1) + (\int f_+ d\nu_2 - \int f_- d\nu_2)\\ &= \int fd\nu_1 + \int fd\nu_2\end{aligned}$$