ST5215 Advanced Statistical Theory, Lecture 5

HUANG Dongming

National University of Singapore

25 Aug 2020

Review

Last time

- Properties of conditional expectation
- Conditional distribution
- Statistical models

Today

- Exponential families
- Statistics

Recap: 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 $A_1, \ldots, A_n \in \mathcal{C}$,

$$P(A_1 \cap \cdots \cap A_n) = P(A_1) \cdots P(A_n). \tag{1}$$

- (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 $X_1, \ldots, X_n$ are said to be independent iff $\sigma(X_1), \ldots, \sigma(X_n)$ are independent.

Example: $C_1 = \{A_1, A_2\}, C_2 = \{A_3, A_4\}$. Then C_1 and C_2 are independent iff 1). A_1 and A_3 are independent, 2). A_1 and A_4 are independent, 3). A_2 and A_3 are independent, 4). A_2 and A_4 are independent.

Recap: 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 $\mathbb{E}(X \mid \mathcal{A})$, is a random variable satisfying the following two conditions:

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

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

Definition (Conditional expectation in L^2 sense)

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

1. Define $L^2(\Omega, \mathcal{F}, P)$ to be the collection of all Borel functions f on $(\Omega, \mathcal{F})$ such that $\int f^2 \, \mathrm{d}P < \infty$. (Inner product $\langle X, Y \rangle := \mathbb{E}(XY)$ makes it a *Hilbert space*)
2. Let $\mathcal{G}$ be a sub-σ-field of $\mathcal{F}$. For any r.v. $X \in L^2(\Omega, \mathcal{F}, P)$, the conditional expectation of X given $\mathcal{G}$, denoted by $\mathbb{E}(X \mid \mathcal{G})$, is defined as the orthogonal projection of X onto the closed subspace $L^2(\Omega, \mathcal{G}, P)$.

- Orthogonal projection: $\langle X - \mathbb{E}(X \mid \mathcal{G}), Z \rangle = 0$ for all $Z \in L^2(\Omega, \mathcal{G}, P)$
- One can prove that $\mathbb{E}(X \mid \mathcal{G}) = \underset{f \in L^2(\Omega, \mathcal{G}, P)}{\textit{arg min}} \; \mathbb{E}(X - f)^2$
- For any $B \in \mathcal{G}$, let $Z = I_B$

$$\int_B X \, \mathrm{d}P = \int_B \mathbb{E}(X \mid \mathcal{G}) \, \mathrm{d}P, \tag{2}$$

so this definition is the same as the one defined before (of course, only when $\mathbb{E}X^2 < \infty$)

Recap: 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 $B \in \mathcal{B}^n$, $P[X \in B \mid Y]$ is a random variable on $(\Omega, \sigma(Y))$ and can be represented as $h(Y)$.
- 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 $P_{X|Y}(B \mid y) = h(y)$ holds.

Recap: Statistical models

- A *statistical model* is a set of assumptions on the population P, and is often expressed as

$$P \in \mathcal{P} = \{Q : Q \textit{ satisfies some conditions}\} \tag{3}$$

- *A parametric model* refers to the assumption that the population P is in a parametric family $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$, where $\Theta \subset \mathcal{R}^d$
- A parametric family $\{P_\theta : \theta \in \Theta\}$ is said to be *identifiable* if and only if $\theta_1 \neq \theta_2$ and $\theta_1, \theta_2 \in \Theta$ imply $P_{\theta_1} \neq P_{\theta_2}$
- Let $\mathcal{P}$ be a family of populations and ν a σ-finite measure on $(\Omega, \mathcal{F})$. If $P \ll \nu$ for all $P \in \mathcal{P}$, then we say $\mathcal{P}$ is *dominated* by ν. In this case, $\mathcal{P}$ can be identified by the family of densities $\{\frac{\mathrm{d}P}{\mathrm{d}\nu} : P \in \mathcal{P}\}$.

Exercise

- Suppose $P_\theta(X > x) = \exp(-x/\theta)$, for $\theta > 0, x > 0$. Show that $\{P_\theta\}$ is dominated.
- Suppose P_θ is a point mass at θ, i.e., $P_\theta(X = \theta) = 1$, for any $\theta \in \mathcal{R}$. Show that $\{P_\theta\}$ cannot be dominated.
  (Hint: Suppose it is dominated by ν. Find an uncountable set A s.t. $\nu(A)$ is finite. Then "exhaust" the measure of A by choosing a countable set.)

Exponential families

Definition

A parametric family $\{P_\theta : \theta \in \Theta\}$ dominated by a σ-finite measure ν on $(\Omega, \mathcal{E})$ is called an *exponential family* iff

$$f_\theta(\omega) = \frac{\mathrm{d}P_\theta}{\mathrm{d}\nu}(\omega) = \exp\left\{[\eta(\theta)]^\top T(\omega) - \xi(\theta)\right\} h(\omega), \qquad \omega \in \Omega, \tag{4}$$

where T is a random p-vector, η is a function from Θ to $\mathcal{R}^p$, h is a nonnegative Borel function on $(\Omega, \mathcal{E})$, and

$$\xi(\theta) = \log\left\{\int_\Omega \exp\{[\eta(\theta)]^\top T(\omega)\} h(\omega)\, \mathrm{d}\nu(\omega)\right\}. \tag{5}$$

- T and h are functions of ω only
- ξ and η are functions of θ only

Exponential families (Cont.)

The representation of an exponential family is not unique:

1. Let D be a $p \times p$ nonsingular matrix. Transforming $\tilde{\eta} = D\eta(\theta)$ and $\tilde{T} = D^{-\top} T$ gives another representation for the same family
2. Another measure that dominates the family also changes the representation:
   Define $\lambda(A) = \int_A \, h \, \mathrm{d}\nu$ for any $A \in \mathcal{F}$, then we can represent the same exponential family with densities w.r.t. λ

$$\frac{dP_\theta}{d\lambda}(\omega) = \exp\left\{[\eta(\theta)]^\tau T(\omega) - \xi(\theta)\right\}$$

The canonical form

Reparametrize the family by $\eta = \eta(\theta)$, so that

$$f_\eta(\omega) = \exp\{\eta^\top T(\omega) - \zeta(\eta)\} h(\omega) \tag{6}$$

where $\zeta(\eta) = \log\left\{\int_\Omega \exp\{\eta^\top T(\omega)\} h(\omega)\, \mathrm{d}\nu(\omega)\right\}$.

- This is the *canonical form* for the family (still not unique)
- η is called the *natural parameter*
- The *natural parameter space*: $\Xi = \{\eta(\theta) : \theta \in \Theta\} \subset \mathcal{R}^p$
- An exponential family in its canonical form is called a *natural exponential family*
- *Full rank*: if Ξ contains an open set

Example: Binomial distribution

The Binomial distributions $\{\mathrm{Binom}(\theta, n) : \theta \in (0, 1)\}$ is an exponential family. Here the density of $\mathrm{Binom}(\theta, n)$ w.r.t. the counting measure is

$$
\begin{aligned}
f_\theta(x) &= \binom{n}{x} \theta^x (1-\theta)^{n-\theta} \\
&= \exp\left\{ x \log \frac{\theta}{1-\theta} + n \log(1-\theta) \right\} \binom{n}{x}, \quad x \in \Omega = \{0, 1, \dots, n\}
\end{aligned}
$$

- $T(x) = x$
- $\eta(\theta) = \log \frac{\theta}{1-\theta}$
- $\xi(\theta) = -n \log(1-\theta)$
- $h(x) = \binom{n}{x}$
- $\Theta = (0, 1)$

Example: Binomial distribution (Canonical form)

We can turn it into its canonical form. Let $\eta = \log \frac{\theta}{1-\theta}$. The density becomes

$$f_\eta(x) = \exp\{\eta x - n \log(1 + e^\eta)\} \binom{n}{x}, \qquad \forall x \in \Omega = \{0, 1, \ldots, n\}. \quad (7)$$

The parameter space is $\Xi = \mathcal{R}$.

Example: Exponential distribution

The exponential distributions $\{E(a,\theta) : \theta > 0\}$ for a fixed $a \in \mathcal{R}$ is an exponential family. Here the density of $E(a,\theta)$ is

$$f_\theta(x) = \theta^{-1}\exp\{-(x-a)/\theta\}, \quad \text{for } x > a. \tag{8}$$

- We can rewrite it as

$$f_\theta(x) = \exp\{-x/\theta + a/\theta - \log\theta\} I_{(a,\infty)}(x) \tag{9}$$

- $T(x) = x$
- $\eta(\theta) = -1/\theta$
- $\xi(\theta) = -a/\theta + \log\theta$
- $h(x) = I_{(a,\infty)}(x)$.
- Note: if a is not fixed, then it is not an exponential family
- To turn it into a natural family, reparametrize $\eta = -1/\theta$ and $\Xi = (-\infty, 0)$; it is of full rank

Example: Normal distribution

The normal family $\left\{N\left(\mu, \sigma^2\right) : \mu \in \mathcal{R}, \sigma > 0\right\}$ is an exponential family. Here $N\left(\mu, \sigma^2\right)$ has Lebesgue density

$$\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

- We rewrite the density as

$$\frac{1}{\sqrt{2\pi}}\exp\left\{\frac{\mu}{\sigma^2}x - \frac{1}{2\sigma^2}x^2 - \frac{\mu^2}{2\sigma^2} - \log\sigma\right\}$$

- $T(x) = \left(x, -x^2\right), \eta(\theta) = \left(\frac{\mu}{\sigma^2}, \frac{1}{2\sigma^2}\right)$
- $\theta = \left(\mu, \sigma^2\right)$
- $\xi(\theta) = \frac{\mu^2}{2\sigma^2} + \log\sigma$
- $h(x) = 1/\sqrt{2\pi}$
- To turn it into a natural family, reparametrize $\eta = (\eta_1, \eta_2) = \left(\frac{\mu}{\sigma^2}, \frac{1}{2\sigma^2}\right)$ and $\Xi = \mathcal{R} \times (0, \infty)$; it is of full rank

Example: Normal distribution (Cont.)

A subfamily $\left\{N\left(\mu, \mu^2\right) : \mu \in \mathcal{R}, \mu \neq 0\right\}$ is also an exponential family

- $\eta = \left(\frac{1}{\mu}, \frac{1}{2\mu^2}\right)$
- $\Xi = \left\{(x, y) : y = 2x^2, x \in \mathcal{R}, y > 0\right\}.$
- It is not of full rank

Properties

Exercise: show the following result.
Suppose $X_i \sim f_i$ independently and each f_i is in an exponential family, then the joint distribution of $X_1, \ldots, X_n$ is again in an exponential family.

Properties (Cont.)

- For an exponential family P_θ, there is a nonzero measure λ such that $\frac{\mathrm{d}P_\theta}{\mathrm{d}\lambda}(\omega) > 0$ for all ω (λ-a.e.) and θ.
- Use this property to show that some families of distributions are not exponential families.

Example (Uniform distribution)

Let $U(0,\theta)$ denote the uniform distribution on $(0,\theta)$. Let $\mathcal{P} = \{U(0,\theta) : \theta \in \mathcal{R}_+\}$. Show that this family is not an exponential family.

- $\Omega = \mathcal{R}$
- If this is an exponential family, then $\frac{\mathrm{d}P_\theta}{\mathrm{d}\lambda}(\omega) > 0$ for all θ, all $\omega \in \mathcal{R}$ for some measure λ
- For any $t > 0$, there is a $\theta < t$ such that $P_\theta([t,\infty)) = 0$
- Then $\lambda([t,\infty)) = 0$ for any $t > 0$, and further $\lambda((0,\infty)) = 0$
- Also, for any $\theta > 0$, $P_\theta((-\infty,0]) = 0$, which implies $\lambda((-\infty,0]) = 0$
- Then $\lambda(\mathcal{R}) = 0$

Properties of natural exponential families

Let $\mathcal{P}$ be a natural exponential family with PDF

$$f_\eta(x) = \exp\{\eta^\top T(x) - \zeta(\eta)\}h(x) \tag{10}$$

- Let $T = (Y, U)$ and $\eta = (\vartheta, \varphi)$, where Y and ϑ have the same dimension. Then Y has the PDF

$$f_\eta(y) = \exp\{\vartheta^\top y - \zeta(\eta)\} \tag{11}$$

  w.r.t. a σ-finite measure depending on φ.
  In particular, T has a p.d.f. in a natural exponential family.
- Furthermore, the conditional distribution of Y given $U = u$ has the p.d.f. (w.r.t. a σ -finite measure depending on u)

$$f_{\vartheta,u}(y) = \exp\left\{\vartheta^\tau y - \zeta_u(\vartheta)\right\}$$

  which is in a natural exponential family indexed by ϑ

- If η_0 is an interior point of the natural parameter space, then the MGF $\psi_{\eta_0}(t)$ of $T(X)$ (with $P = P_{\eta_0}$) is finite in a neighborhood of $t = 0$ and is given by

$$\psi_{\eta_0}(t) = \exp\{\zeta(\eta_0 + t) - \zeta(\eta_0)\}. \tag{12}$$

- Furthermore, if f is a Borel function satisfying $\int |f| dP_{\eta_0} < \infty$ then the function

$$\int f(\omega) \exp\left\{\eta^\tau T(\omega)\right\} h(\omega) d\nu(\omega)$$

is infinitely often differentiable in a neighborhood of η_0, and the derivatives may be computed by differentiation under the integral sign.

Example (MGF of binomial distribution)

Recall that

- the canonical form of $\text{Binom}(n, e^\eta/(1+e^\eta))$ is given by

$$f_\eta(x) = \exp\{\eta x - n\log(1+e^\eta)\}\binom{n}{x}, \qquad \forall x \in \Omega = \{0, 1, \ldots, n\} \tag{13}$$

- $\zeta(\eta) = n\log(1+e^\eta)$
- $T(x) = x$

$$\begin{aligned}\psi_{\eta_0}(t) = \exp\{\zeta(\eta_0+t) - \zeta(\eta_0)\} &= \exp\{n\log(1+e^{\eta_0+t}) - n\log(1+e^{\eta_0} \\ &= \left(\frac{1+e^{\eta_0}e^t}{1+e^{\eta_0}}\right)^n = (1-\theta+\theta e^t)^n\end{aligned}$$

since $\theta = e^\eta/(1+e^\eta)$.

Statistics

A **statistic** $T(X)$ is a measurable function of sample X.

- $T(X)$ only depends on X
- T is a known function: $T(X)$ is a known value whenever X is known.
- Trivial statistics: X itself, any constant
- Nontrivial statistic: $T(X)$ is simpler than that of X but contains some information about X. For instance, X may be a random n -vector and $T(X)$ may be a random p-vector with $1 \leq p \leq n$
- Some examples are:
  - sample mean: $\bar{X} = \frac{1}{n} \sum_i X_i$
  - sample variance: $S^2 = \frac{1}{n-1} \sum_i (X_i - \bar{x})^2$
  - order statistics, $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$
  - sample median: middle value of ordered statistics
  - sample minimum: $X_{(1)}$
  - sample maximum: $X_{(n)}$

Example: Sample mean and sample variance

Let $X_1, \ldots, X_n$ be i.i.d. sample from P and $X = (X_1, \ldots, X_n)$.

- Sample mean: $\bar{X} = \frac{1}{n} \sum_i X_i$
- Sample variance: $S^2 = \frac{1}{n-1} \sum_i (X_i - \bar{x})^2$

Moments

- Assume that P has a finite mean denoted by μ. Then

$$E\bar{X} = \mu$$

  - If P is in a parametric family $\{P_\theta : \theta \in \Theta\}$, then $E\bar{X} = \int x dP_\theta = \mu(\theta)$ for some function $\mu(\cdot)$. Even if the form of μ is known, $\mu(\theta)$ may still be unknown when θ is unknown.

- Assume now that P has a finite variance denoted by σ^2. Then

$$\mathrm{Var}(\bar{X}) = \sigma^2/n.$$

  - If P is in a parametric family, $\mathrm{Var}(\bar{X})$ equals $\sigma^2(\theta)/n$ for some function $\sigma^2(\cdot)$.

Example: Sample mean and sample variance (Cont.)

Distributions

- If P is in a parametric family, we can often find the distribution of $\bar{X}$.
  - if P is $N\left(\mu, \sigma^2\right)$ then $\bar{X}$ is $N\left(\mu, \sigma^2/n\right)$;
  - if P is the exponential distribution $E(0, \theta)$, $n\bar{X}$ has the gamma distribution $\Gamma(n, \theta)$
- Usually hard to find the exact form of the distribution of $\bar{X}$ if P is not in a parametric family
  - one can use the *Central Limit Theorem* to approximate the distribution of $\bar{X}$ by $N\left(\mu, \sigma^2/n\right)$ where μ and σ^2 are the mean and variance of P (assumed to be finite)

Tutorial

1. Suppose f and g are independent and identically distributed. Show that
$$E(f \mid f+g) = (f+g)/2, \text{a.s.} \tag{14}$$
2. Suppose $F(x)$ is a continuous CDF of P, where P is a probability measure on $(\mathcal{R}, \mathcal{B})$. Show that $\int F(x)\,\mathrm{d}P(x) = 1/2$
3. Suppose ν is a σ-finite measure on $(\Omega, \mathcal{F})$, f is a nonnegative measurable function and $\alpha > 0$. Show that
$$\int f^\alpha\,\mathrm{d}\nu = \alpha \int_0^\infty t^{\alpha-1}\nu(f > t)\,\mathrm{d}t \tag{15}$$
4. Suppose ν and ϕ are finite measures on $(\Omega, \mathcal{F})$. Show that there exist two measures ϕ_c and ϕ_s such that
    1. $\phi = \phi_c + \phi_s$,
    2. $\phi_c \ll \nu$, and
    3. there exists $N \in \mathcal{F}$ such that $\phi_s(N) = \nu(N^c) = 0$. (Note that in this case, we denote $\phi_s \perp \nu$ and say that ϕ_s and ν are singular with each other.)

    (Hint: make use of $\frac{\mathrm{d}\phi}{\mathrm{d}(\phi+\nu)}$ by Radon-Nikodym theorem)

Ex 1

Suppose f and g are independent and identically distributed. Show that

$$E(f \mid f+g) = (f+g)/2, \text{a.s.} \tag{16}$$

Proof:

- Define a mapping $\phi(x,y) = (x, x+y)$. It is easy to show $(f,g) \stackrel{d}{=} (g,f)$ (here $\stackrel{d}{=}$ means that the two sides have the same distribution), and thus $\phi(f,g) \stackrel{d}{=} \phi(g,f)$. That is

$$(f, f+g) \stackrel{d}{=} (g, f+g) \tag{17}$$

- We then have for any Borel function h on $\mathcal{R}$,

$$\int f \; h(f+g) \; \mathrm{d}P = \int g \; h(f+g) \; \mathrm{d}P, \tag{18}$$

and thus for any $A \in \sigma(f+g)$

$$\int_A f \; \mathrm{d}P = \int_A g \; \mathrm{d}P. \tag{19}$$

Ex 1 (Cont.)

- Therefore, any $A \in \sigma(f+g)$

$$\int_A E(f \mid f+g)\,\mathrm{d}P = \int_A E(g \mid f+g)\,\mathrm{d}P, \tag{20}$$

  which implies that $E(f \mid f+g) = E(g \mid f+g)$ a.s.
- But $E(f \mid f+g) + E(g \mid f+g) = E(f+g \mid f+g) = f+g$.
- So $E(f \mid f+g) = (f+g)/2$ a.s.

Ex 2

Suppose $F(x)$ is a continuous CDF of P, where P is a probability measure on $(\mathcal{R}, \mathcal{B})$. Show that $\int F(x)\, \mathrm{d}P(x) = 1/2$

Proof: Suppose P is the probability of r.v. X, i.e. $F(x) = P(X \leq x)$.

We need a useful result:

Lemma

Let X be a random variable having a continuous c.d.f. F. Then $Y = F(X)$ has the uniform distribution $Unif\,(0,1)$.

Here $Unif\,(0,1)$ is defined as the probability on $\mathcal{R}$ with Lebesgue p.d.f. $I_{(0,1)}$, whose CDF is $F(x) = x \cdot I_{(0,1)}(x) + I[1,\infty)(x)$ and is continuous

Using the lemma and change of variable:

$$\int F(x)\, \mathrm{d}P(x) = \int Y\, \mathrm{d}P_Y = \int_0^1 y\, \mathrm{d}y = 1/2 \tag{21}$$

Ex 2 (Cont.)

Proof of Lemma:

- Define the inverse of F by $F^{-1}(t) = \inf\{x : F(x) > t\}$
- One can show that $F(x) < t \Leftrightarrow F^{-1}(t) > x$
- For any $y \in (0, 1)$,

$$\begin{aligned} P(Y < y) &= P(F(X) < y) \\ &= P(F^{-1}(y) > X) \\ &= \lim_n P(X \le F^{-1}(y) - 1/n) \\ &= \lim_n F(F^{-1}(y) - 1/n) \\ &= F(F^{-1}(y)) = y \end{aligned}$$

  - The 3rd equality is due to continuity of P
  - The 5th equality is due to the continuity of F
  - The 6th equality is due to the definition of F^{-1} and the continuity of F

Ex 3

Suppose ν is a σ-finite measure on $(\Omega, \mathcal{F})$, f is a nonnegative measurable function and $\alpha > 0$. Show that

$$\int f^\alpha \, \mathrm{d}\nu = \alpha \int_0^\infty t^{\alpha-1} \nu(f > t) \, \mathrm{d}t \tag{22}$$

Proof:

- For any $x \in \mathcal{R}_+$, we have

$$x^\alpha = \int_{(0,\infty)} \alpha t^{\alpha-1} I_{t<x} \, \mathrm{d}m(t) \tag{23}$$

- Apply Fubini's theorem to the bivariate function $\alpha t^{\alpha-1} I_{t<f(\omega)}$ of (ω, t) w.r.t. $\nu \times m$.
- See also page 13 of Lecture 2.

Ex 4

Suppose ν and ϕ are finite measures on $(\Omega, \mathcal{F})$. Show that there exist two measures ϕ_c and ϕ_s such that

1. $\phi = \phi_c + \phi_s$,
2. $\phi_c \ll \nu$, and
3. there exists $N \in \mathcal{F}$ such that $\phi_s(N) = \nu(N^c) = 0$.

Proof:

- Note that $\phi \ll (\phi + \nu)$, so Radon-Nikodym derivative $\frac{\mathrm{d}\phi}{\mathrm{d}(\phi+\nu)}$ exists.
- For any $A \in \mathcal{F}$, we have $0 \leqslant \int_A \frac{d\phi}{d(\phi+\nu)} d(\phi + \nu) = \phi(A) \leqslant \phi(A) + \nu(A) = \int_A 1 \mathrm{d}(\phi + \nu)$, which implies that $0 \leq \frac{\mathrm{d}\phi}{\mathrm{d}(\phi+\nu)} \leq 1$ a.e.
- Let $N = \{\omega \in \Omega : \frac{\mathrm{d}\phi}{\mathrm{d}(\phi+\nu)}(\omega) < 1\}$, $\phi_c(A) = \phi(A \cap N)$, and $\phi_s(A) = \phi(A \cap N^c)$. Then $\phi = \phi_c + \phi_s$.

Ex 4 (Cont.)

To show that $\phi_c \ll \nu$:
For any $A \in \mathcal{F}$ such that $\nu(A) = 0$, it holds that

$$\begin{aligned} &\int_{A\cap N} \left[1 - \frac{\mathrm{d}\phi}{\mathrm{d}(\phi+\nu)}\right] \mathrm{d}(\phi+\nu) \\ =&(\phi+\nu)(A\cap N) - \phi(A\cap N) \\ =&\nu(A\cap N) = 0. \end{aligned}$$

- But $[1 - \frac{\mathrm{d}\phi}{\mathrm{d}(\phi+\nu)}](\omega) > 0$ for any $\omega \in N$
- So $(\phi+\nu)(A\cap N) = 0$
- Thus, $\phi_c(A) = 0$.

Ex 4 (Cont.)

To show $\nu(N^c) = 0$:

$$\begin{aligned}\phi(N^c) &= \int_{N^c} \frac{\mathrm{d}\phi}{\mathrm{d}(\phi+\nu)}\ \mathrm{d}(\phi+\nu) \\ &= \int_{N^c} 1\ \mathrm{d}(\phi+\nu) \\ &= \phi(N^c) + \nu(N^c),\end{aligned}$$

which implies that $\nu(N^c) = 0$ because $\phi(N^c)$ is a finite number.
Remark. The result of this exercise can be generalized to σ-finite measures, and is known as *Lebesgue decomposition*.