Densities with respect to measures

1 Introduction to measures and integration

- Reference: Mathematic Statistics by Shao, Jun (2nd Ed.)
- σ-fields
  - σ-field (Definition 1.1 in Section 1.1.1)
  - $\sigma(\mathcal{C})$: the σ-field generated by a collection $\mathcal{C}$ (the smallest σ-field containing $\mathcal{C}$)
  - Borel σ-field $\mathcal{B}(\Omega)$: the σ-field generated by open sets in Ω
- Measure related definitions
  - Measurable space
  - Measure (Definition 1.2 in Section 1.1.1)
  - Example. δ_c measure ($c \in R$). For $A \in \mathcal{B}(R)$,

$$\delta_c(A) = \begin{cases} 1 & \text{if } c \in A \\ 0 & \text{if } c \notin A \end{cases}$$

- Uniqueness
  - Theorem 10.3 (Billingsley 1986) Suppose that μ_1 and μ_2 are measures on $\sigma(\mathcal{P})$, where $\mathcal{P}$ is a π-system, and suppose they are σ-finite on $\mathcal{P}$. If μ_1 and μ_2 agree on $\mathcal{P}$, then they agree on $\sigma(\mathcal{P})$.
  - Definition. Suppose that $\mathcal{C}$ is a collection of some subsets of Ω. $\mathcal{C}$ is π-system if it is closed under finite intersections.
  - Definition. Suppose that $\mathcal{C}$ is a collection of some subsets of Ω. A measure μ is σ-finite on $\mathcal{C}$ if there exists $\{A_k\}$: a sequence of sets in $\mathcal{C}$ such that

$$\Omega = \cup_k A_k \text{ and } \mu(A_k) < \infty \text{ for all } k.$$

  - Lebesgue measure
  - Product measure
- Some properties of a measure
  - Monotonicity
  - Subadditivity
  - Continuity
- Measurable functions.
  - Definition of a measurable function (Definition 1.3 in Section 1.1.2).

- Random variable. A random variable on a probability space $(\Omega, \mathcal{F}, P)$ is a measurable function from $(\Omega, \mathcal{F})$ to $(R, \mathcal{B}(R))$.
- Distribution of X: the probability measure $P_X = P \circ X^{-1}$ on $(R, \mathcal{B}(R))$.

- Indicator function I_A is measurable from $(\Omega, \mathcal{F})$ to $(R, \mathcal{B}(R))$ if $A \in \mathcal{F}$.

- Suppose that a function f: $R^m \to R^k$ is continuous on R^m. Then f is measurable from $(R^m, \mathcal{B}(R^m))$ to $(R^k, \mathcal{B}(R^k)$.

- Suppose that $f_1, \ldots, f_k$ are measurable from $(\Omega, \mathcal{F})$ to $(R, \mathcal{B}(R))$, let $f = (f_1, \ldots, f_k)$, then f is measurable from $(\Omega, \mathcal{F})$ to $(R^k, \mathcal{B}(R^k))$.

- Compositions of measurable functions are measurable (Proposition 1.4 (iv))

- Approximation property.

  Fact 1 Suppose that f is measurable from $(\Omega, \mathcal{F})$ to $(R, \mathcal{B}(R))$, then (i) and (ii) hold.

  (i) Suppose that $f \geq 0$. Then there exists $\{f_n\}$: a sequence of real-valued simple functions such that $0 \leq f_n \leq f_{n+1} < \infty$ and $\lim_{n\to\infty} f_n = f$.

  (ii) Let
  $$f^+(w) = \begin{cases} f(w) & \text{if } f(w) > 0; \\ 0 & \text{otherwise,} \end{cases}$$
  and
  $$f^-(w) = \begin{cases} -f(w) & \text{if } f(w) < 0; \\ 0 & \text{otherwise.} \end{cases}$$
  Then $f = f^+ - f^-$ and f^+ and f^- are measurable from $(\Omega, \mathcal{F})$ to $(R, \mathcal{B}(R))$.

  Note. In Fact 1, $(R, \mathcal{B}(R))$ can be replaced by $(\overline{R}, \overline{\mathcal{B}})$, where $\overline{R} = R \cup \{\infty, -\infty\}$ and $\overline{\mathcal{B}} = \sigma(\mathcal{B}(R) \cup \{\{\infty\}, \{-\infty\}\})$.

- Definition of integration. (Definition 1.4 in Section 1.2.1)
  - Integrable functions.
  - Integration over a set.
  - Example 1. $\Omega = \{1, 2, 3, 4, 5, 6\}$. $X(\omega) = \omega$. P: measure on $(\Omega, 2^\Omega)$ such that $P(\{\omega\}) = 1/6$ for $\omega \in \Omega$. Find $\int X dP$.

- Basic properties of integration
  - Linearity (Proposition 1.5 in Section 1.2.1)
  - Monotonicity (Proposition 1.6(i) in Section 1.2.1)

- If $f \geq 0$ ν-a.e. and $\int f d\nu = 0$, then $f = 0$ ν-a.e. (Proposition 1.6(ii) in Section 1.2.1).
- $\nu(A) = 0$ implies that $\int_A f d\nu = 0$.

- Limits of integrals (Theorem 1.1 and Example 1.8 in Section 1.2.1)
  - Dominated convergence theorem
  - Monotone convergence theorem
- Fubini's theorem (Theorem 1.3 in Section 1.2.1)
- Suppose that f is Riemann integrable on a finite interval I with endpoints a and b, where $a < b$. Let λ be the Lebesgue measure on $(R, \mathcal{B}(R))$. Then $\int_I f d\lambda = \int_a^b f(x)dx$.
- Suppose that Ω is a countable set and ν is a measure on $(\Omega, 2^\Omega)$. Then for a nonnegative f that is measurable from $(\Omega, 2^\Omega)$ to $(R, \mathcal{B}(R))$,

$$\int f d\nu = \sum_{\omega \in \Omega} f(\omega)\nu(\{\omega\}).$$

- Change of Variable (Theorem 1.2 in Section 1.2.1)

$$\int (g \circ f) d\mu = \int g d(\mu \circ f^{-1}).$$

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

$$E(X) = \int X(\omega) dP(\omega) = \int x dP_X(x),$$

where $P_X = P \circ X^{-1}$ is the distribution of X.

- Absolute continuity (Equation (1.19) in Section 1.2.2)
- Radon-Nikodym Theorem (Theorem 1.4 in Section 1.2.2).
- For a random variable X with distirbution P_X, if P_X has a density f with respect to some measure ν, then we say that X has a density f with respect to ν.
- Example 2. Suppose that $Z \sim N(0, 1)$ and

$$X = \begin{cases} Z & \text{if } Z \geq 5; \\ 5 & \text{if } Z < 5. \end{cases}$$

Then X has a density with respect to $\lambda + \delta_5$, where λ is the Lebesgue measure on $(R, \mathcal{B}(R))$.

- Calculus with Radon Nikodym derivatives (Proposition 1.7 (i)-(iii))

– For $f \geq 0$,

$$\int f\mu = \int f \frac{d\mu}{d\nu} d\nu.$$

– sum of a density of μ_1 and a density of μ_2 (w.r.t the same measure) is a density of $(\mu_1 + \mu_2)$.

– Chain rule

- Example 3. Suppose that X is a random variable with a Lebesgue density f_X, and $f_X(x) = 0$ for $x \leq 0$. Let $Y = X^2$ and let

  $$g(y) = \frac{f_X(\sqrt{y})}{2\sqrt{y}} I_{(0,\infty)}(y)$$

  for $y \in R$. Then g is a Lebesgue density of Y.

  Note. Let λ be the Lebesgue measure on $(R, \mathcal{B}(R))$. To verifty that g is a Lebesgue density of Y is to verifty that

  $$\int_{-\infty}^{t} g(x) d\lambda(x) = P(Y \leq t) \tag{1}$$

  for $t \in R$. For $t > 0$,

  $$\int_{-\infty}^{t} g(y) d\lambda(y)$$

2 Conditional expectations and conditional distributions

- Definition of $E(X|\mathcal{A})$ (Definition 1.6 in Section 1.4.1).

  – For a nonnegative random variable X, $E(X|\mathcal{A})$ is a Radon-Nikodym derivative.

- From now on, whenever we write $E(X|\mathcal{A})$, it is assumed that $E|X| < \infty$.

- Suppose that Y is a random vector of dimension m on a probability space $(\Omega, \mathcal{F}, P)$, then $\sigma(Y)$ is the σ-field $\{Y^{-1}(B) : B \in B(R^m)\}$.

  – $E(X|Y) = E(X|\sigma(Y))$.

- Fact 2 Suppose that Y is a random vector on a probability space $(\Omega, \mathcal{F}, P)$. Suppose that Z is a random variable on $(\Omega, \mathcal{F}, P)$ such that Z is measurable from $(\Omega, \sigma(Y)$ to $(R, \mathcal{B}(R))$, then $Z = h(Y)$ for some function h.

- Example 4. Suppose that $\Omega = \{1, 2, 3, 4\}$. P is a measure on $(\Omega, 2^\Omega)$ such that $P(\{k\}) = 1/4$ for $k \in \Omega$. Suppose that $X(k) = k$ for $k \in \Omega$ and $Y(1) = 4$, $Y(2) = 5$, $Y(3) = Y(4) = 6$. Find $E(X|Y)$.

- Given a σ-field $\mathcal{A} \Leftrightarrow$ given the information that whether A occurs for every $A \in \mathcal{A}$.
- Some facts following from the definition of a conditional expection.
  - Suppose that X is measurable from $(\Omega, \mathcal{A}_0)$ to $(R, \mathcal{B})$, where $\mathcal{A}_0$ is a sub-σ-field of $\mathcal{A}$. Then $E(X|\mathcal{A}) = X$.
  - If $\mathcal{A} = \{\emptyset, \Omega\}$, then $E(X|\mathcal{A}) = E(X)$.
- Properties of conditional expectations (Proposition 1.10 or Proposition 1.12 in the first edition).
- $E(X|\mathcal{A})$ is the "best guess" of X given the information of occurrences of events in $\mathcal{A}$ in the following sense

$$\int (X - E(X|\mathcal{A}))^2 dP \leq \int (X - Y)^2 dP \tag{2}$$

for all Y: measurable from $(\Omega, \mathcal{A})$ to $(R, \mathcal{B})$.

Fact 3 Suppose that X is a random variable on $(\Omega, \mathcal{F}, P)$ with $E|X| < \infty$, and $\mathcal{A}_1$ and $\mathcal{A}_2$ are sub-σ-fields of $\mathcal{F}$. If $\sigma(\sigma(X) \cup \mathcal{A}_1)$ and $\mathcal{A}_2$ are independent, then

$$E(X|\sigma(\mathcal{A}_1 \cup \mathcal{A}_2)) = E(X|\mathcal{A}_1) \text{ a.s.}$$

The proof of Fact 3 is based on the result that

$$\int_{A_1 \cap A_2} E(X|\mathcal{A}_1) dP = \int_{A_1 \cap A_2} X dP \text{ for } A_1 \in \mathcal{A}_1 \text{ and } A_2 \in \mathcal{A}_2,$$

which can be established from the following fact:

Fact 4 Suppose that X is a nonnegative random variable on $(\Omega, \mathcal{F}, P)$ and $\mathcal{A}_2$ is a sub-σ-field of $\mathcal{F}$. If $\mathcal{A}_2$ is independent of $\sigma(X)$, then for $A_2 \in \mathcal{A}_2$,

$$E(X I_{A_2}) = P(A_2) E(X).$$

The proof of Fact 4 is based on Proposition 1.10 (vii).

Special cases of Fact 3:

- Proposition 1.11 in Section 1.4.2 (Proposition 1.14 in the first edition). Note: $\sigma((Y_1, Y_2)) = \sigma(\sigma(Y_1) \cup \sigma(Y_2))$.
- Suppose that X is a random variable on $(\Omega, \mathcal{F}, P)$ with $E|X| < \infty$ and Y is a measurable function from $(\Omega, \mathcal{F})$ to a measurable space. Suppose that $\sigma(X)$ and $\sigma(Y)$ are independent. Then

$$E(X|Y) = E(X) \text{ a.s.}$$

- Conditional distributions are random probability measures.

- Definition of a random probability measure on $(R^n, \mathcal{B}(R^n))$. Suppose that $(\Omega, \mathcal{F}, P)$ is a probability space and μ is a function on $\mathcal{B}(R^n) \times \Omega$ satisfying (i) and (ii):

  (i) For every $\omega \in \Omega$, $\mu(\cdot, \omega)$ is a probability measure on $(R^n, \mathcal{B}(R^n))$.

  (ii) For every $B \in \mathcal{B}(R^n)$, let $X(\omega) = \mu(B, \omega)$. Then X is a random variable on $(\Omega, \mathcal{F}, P)$.

  Then μ is a random probability measure on $(R^n, \mathcal{B}(R^n))$ with respect to the probability space $(\Omega, \mathcal{F}, P)$.

- Existence of conditional distributions (Theorem 1.7 (i); Theorem 1.7 in the first edition). Suppose that X and Y are random vectors on $(\Omega, \mathcal{F}, P)$ and take values in R^n and R^m respectively. Let $P_Y = P \circ Y^{-1}$ be the distribution of Y. Then there exists a random probability measure μ on $(R^n, \mathcal{B}(R^n))$ with respect to the probability space $(R^m, \mathcal{B}(R^m), P_Y)$ such that

$$P((X, Y) \in B \times C) = \int_C \mu(B, y) dP_Y(y) \text{ for all } B \in \mathcal{B}(R^n), C \in \mathcal{B}(R^m). \quad (3)$$

  $\{\mu(\cdot, y) : y \in R^m\}$ is called a version of the conditional distribution of X given Y. We denote $\mu(\cdot, y)$ by $P_{X|Y}(\cdot|y)$ or $P_{X|Y=y}$.

- Conditional expectation as expectation with respect to conditional distribution. Suppose that g is measurable from $(R^n \times R^m, \mathcal{B}(R^{n+m}))$ to $(R, \mathcal{B}(R))$. Suppose that g is nonnegative. Let

$$h(y) = \int g(x, y) dP_{X|Y=y}(x)$$

  for $y \in R^m$, then $E(g(X, Y)|Y) = h(Y)$. Note that if $E(|g(X, Y)|) < \infty$, then h can be defined P_Y-a.e and we still have $E(g(X, Y)|Y) = h(Y)$.

- Suppose that X and Y are random vectors of dimensions n and m respectively. Suppose that (X, Y) has a density $f_{X,Y}$ with respect to a product measure $\mu \times \nu$, where μ and ν are measures on $(R^n, \mathcal{B}(R^n)$ and $(R^m, \mathcal{B}(R^m))$ respectively. Let

$$f_Y(y) = \int f_{X,Y}(x, y) d\mu(x)$$

  for $y \in R^m$, then f_Y is the PDF of Y with respect to ν. Let $R_Y = \{y \in R^m : f_Y(y) > 0\}$. For $y \in R_Y$, define

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

for $x \in R^n$. For $y \in R_Y$, let

$$\mu(A, y) = \int_A f_{X|Y=y}(x) d\mu(x)$$

for $A \in \mathcal{B}(R^n)$, then (3) holds and $\{\mu(\cdot, y) : y \in R_Y\}$ is a version of the conditional distribution X given Y. Thus $\{f_{X|Y=y} : y \in R_Y\}$ is called a version of the conditional density of X given Y with respect to μ.