LTCC, MTP: Conditional concepts.

Let ξ be a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ with definite $\mathbb{E}\xi \Leftrightarrow$ for

$$\xi = \xi_+ - \xi_-, \qquad \xi_\pm = \min(\pm\xi, 0)$$

either $\quad \mathbb{E}\,\xi_+ < \infty$ or $\mathbb{E}\,\xi_- < \infty$ (or both).

Consider set function

$$Q(A) = \int_A \xi \, d\mathbb{P} \qquad A \in \mathcal{F}.$$

We may write $Q = Q_+ - Q_-$, where

$$Q_\pm = \int_A \xi_\pm \, d\mathbb{P}.$$

We have for disjoint $A_1, A_2, \ldots \in \mathcal{F}$, $A = \cup A_n$

$$Q_+(A) = \mathbb{E}(\xi_+ 1_A) = \mathbb{E}\Big(\sum_n \xi_+ 1_{A_n}\Big) = \text{by monotone convergence}$$

$$\sum_n \mathbb{E}(\xi_+ 1_{A_n}) = \sum_n Q_+(A_n)$$

$\Rightarrow$ Q_+ is σ-additive, same way Q_- σ-additive

hence Q σ-additive.

Q is a <u>signed</u> measure on $(\Omega, \mathcal{F})$

Moreover,

$$\mathbb{P}(A) = 0 \implies Q(A) = 0 \quad \text{for } A \in \mathcal{F}$$

We say: Q is absolutely continuous w.r.t. $\mathbb{P}$, written $Q \ll \mathbb{P}$.

Converse is true:

Theorem (Radon–Nikodym) On measurable space $(\Omega, \mathcal{F})$, let

μ be σ-finite measure

$\nu = \nu_+ - \nu_-$ signed measure, s.t.
either $\nu_+(\Omega) < \infty$ or $\nu_-(\Omega) < \infty$.

If $\nu \ll \mu$ then there exists $\mathcal{F}$-measurable function f such that

$$\nu(A) = \int_A f(\omega)\, \mu(d\omega).$$

Up to a set of zero μ-measure, such f is unique.

Notation: $f = \frac{d\nu}{d\mu}$ – Radon-Nikodym derivative (or density).

Example X r.v. on $(\Omega, \mathcal{F}, \mathbb{P})$, $F(x) = \mathbb{P}(X \le x)$.

A density function satisfying

$$F(x) = \int_{-\infty}^{x} f(y)\,dy$$

exists if and only if

$\lambda(B) = 0,\ B \in \mathcal{B}(\mathbb{R}) \implies \mathbb{P}(X \in B) = 0.$

Lebesgue measure

(2a)

<u>Example</u> (Discrete measures). $\Omega = \mathbb{N}$, $X(j) = j$

$$\left.\begin{array}{l} P = (p_j,\ j \in \mathbb{N}) \\ Q = (q_j,\ j \in \mathbb{N}) \end{array}\right\}\ \text{probability mass functions}$$

$$\mathbb{E}_P\, g(X) = \sum_{j=1}^{\infty} g(j)\, p_j$$

$$\mathbb{E}_Q\, g(X) = \sum_{j=1}^{\infty} g(j)\, q_j$$

Now suppose $p_j = 0 \Longrightarrow q_j = 0$, and let

$$\xi(j) = \begin{cases} \dfrac{q_j}{p_j} & \text{if } p_j \neq 0 \\ \text{undefined, or any value} & \text{if } p_j = 0. \end{cases} \qquad (*)$$

Then

$$\mathbb{E}_Q\, g(X) = \sum_{j=1}^{\infty} g(j)\, \xi_j\, p_j = \mathbb{E}_P[\xi\, g(X)]$$

Condition $(*)$ means $P \gg Q$, and the r.v. ξ is the Radon-Nikodym derivative.

$$\xi = \frac{dQ}{dP}.$$

<u>Example</u> (To be detailed) Girsanov's Theorem on abs. continuity of BM with drift w.r.t. standard BM.

<u>Definition</u> Two measures μ, ν on $(\Omega, \mathcal{F})$ are mutually singular if there exist $A, B \in \mathcal{F}$ s.t.

$\mu(A^c) = 0$, $\nu(B^c) = 0$, and $A \cap B = \emptyset$.

<u>Example</u> $\Omega = \{0,1\}^\infty$ $(\omega_1, \omega_2, \ldots)$

$\mathbb{P}_1$ coin-tossing with success prob. p_1

$\mathbb{P}_2$ — " — p_2, $p_1 \neq p_2$

$$A = \{(\omega_1, \omega_2, \ldots) : \lim_{n \to \infty} \frac{\omega_1 + \ldots + \omega_n}{n} = p_1\}$$

$$B = \{ \quad — " — \quad = p_2\}$$

By the Law of Large Numbers,

$\mathbb{P}_1$ and $\mathbb{P}_2$ are mutually singular.

<u>Theorem</u> (Lebesgue decomposition) For σ-finite measures μ and ν on $(\Omega, \mathcal{F})$ there exists a unique (up to null sets) decomposition

$$\nu = \nu_a + \nu_s$$

where
$$\begin{cases} \nu_a \ll \mu \\ \nu_a, \nu_s \text{ mutually singular,} \\ \mu, \nu_s \text{ mutually singular.} \end{cases}$$

For ξ r.v. on $(\Omega, \mathcal{F}, \mathbb{P})$, and $\mathcal{G} \subset \mathcal{F}$ sub-σ-algebra we wish to define conditional expectation

$$\mathbb{E}(\xi \mid \mathcal{G}).$$

Simple situation: $\mathcal{G}$ is generated by $A_1, \dots, A_n$, disjoint partition $\bigcup_{i=1}^{n} A_i = \Omega$; $A_i \cap A_j = \emptyset$, $i \neq j$

For $B \in \mathcal{F}$ $\quad \mathbb{P}(B \mid A_i) = \dfrac{\mathbb{P}(B \cap A_i)}{\mathbb{P}(A_i)}$

$$\mathbb{P}(B \mid \mathcal{G}) = \sum_{i=1}^{n} \mathbb{P}(B \mid A_i) \mathbb{1}_{A_i} \quad \text{random variable}$$

$$\mathbb{E}\left[\mathbb{P}(B \mid \mathcal{G})\right] = \sum_{i=1}^{n} \mathbb{P}(B \mid A_i)\, \mathbb{P}(A_i) = \mathbb{P}(B)$$

now let

$$\xi = \sum_{j=1}^{m} x_j \mathbb{1}_{B_j},$$

define

$$\mathbb{E}\left[\xi \mid \mathcal{G}\right] = \sum_{j=1}^{m} x_j \mathbb{P}(B_j \mid \mathcal{G}).$$

Then ξ is $\mathcal{G}$-measurable: on A_i assumes value

$$\sum_{j=1}^{m} x_j \mathbb{P}(B_j \mid A_i)$$

and satisfies

$$\mathbb{E}\left[\mathbb{E}(\xi \mid \mathcal{G})\right] = \mathbb{E}\xi$$

But we need to also consider conditional probability / expectation w.r.t. events of zero probability

Example Choose $\xi \sim$ Uniform $[0,1]$, given $\xi = x$ toss a coin w.p. x for "heads" n times

$$\mathbb{P}(k \text{ heads} \mid \underbrace{\xi = x}_{\text{event of } 0 \text{ prob.}}) = \binom{n}{k} x^k (1-x)^{n-k} \quad \text{(logically!)}$$

Definition For ξ on $(\Omega, \mathcal{F}, \mathbb{P})$, $\mathcal{G} \subset \mathcal{F}$ we define

$\mathbb{E}(\xi_+ \mid \mathcal{G})$ by conditions

(a) $\mathbb{E}(\xi_+ \mid \mathcal{G})$ is $\mathcal{G}$-measurable

(b) for $A \in \mathcal{G}$

$$\mathbb{E}[\xi_+ 1_A] = \mathbb{E}[\mathbb{E}(\xi_+ \mid \mathcal{G}) \cdot 1_A]$$

$$\int_A \xi_+ \, d\mathbb{P} = \int_A \mathbb{E}(\xi_+ \mid \mathcal{G}) \, d\mathbb{P}.$$

Similarly $\mathbb{E}(\xi_- \mid \mathcal{G})$.

If $\min\left(\mathbb{E}(\xi_+ \mid \mathcal{G}), \mathbb{E}(\xi_- \mid \mathcal{G})\right) < \infty$ a.s.

we set

$$\mathbb{E}(\xi \mid \mathcal{G}) = \mathbb{E}(\xi_+ \mid \mathcal{G}) - \mathbb{E}(\xi_- \mid \mathcal{G}).$$

Why cond. expectation exists? (6)

Suppose $\xi \geq 0$, consider

$$Q(A) = \int_A \xi \, d\mathbb{P} \qquad A \in \mathcal{G} \subset \mathcal{F}$$

— this is a σ-additive measure on $(\Omega, \mathcal{G}) \Rightarrow$ by the Radon-Nikodym theorem

$$Q(A) = \int_A f \, d\mathbb{P}$$ for some $\mathcal{G}$-measurable nonnegative function f, so we set $\underline{f =: \mathbb{E}(\xi | \mathcal{G})}$, that is

$$\mathbb{E}(\xi | \mathcal{G}) = \frac{dQ}{d\mathbb{P}}.$$

— $\mathbb{E}(\xi | \mathcal{G})$ is a random variable, defined uniquely up to sets of zero probability.

We write $\mathbb{E}(\xi \mid \text{some data})$ to mean $\mathbb{E}(\xi \mid \sigma(\text{this data}))$, e.g. for r.v.'s ξ, η

$$\mathbb{E}(\xi | \eta) = \mathbb{E}(\xi | \sigma(\eta))$$

for instance $\mathbb{E}(\xi | \xi) = \xi$; ($\xi$ is $\sigma(\xi)$-meas)

Some properties:

(i) $\mathbb{E}(\xi \mid \mathcal{G}) = \xi$ if ξ is $\mathcal{G}$-measurable

(ii) $\mathbb{E}(\xi \eta \mid \mathcal{G}) = \xi\, \mathbb{E}(\eta \mid \mathcal{G})$ if ξ is $\mathcal{G}$-measurable

(iii) $\mathcal{G}_1 \subset \mathcal{G}_2 \Rightarrow \mathbb{E}\big(\mathbb{E}(\xi \mid \mathcal{G}_2) \mid \mathcal{G}_1\big) = \mathbb{E}(\xi \mid \mathcal{G}_1)$ a.s

(tower property)

(iii') $\mathcal{G}_1 \supset \mathcal{G}_2$

$$\mathbb{E}\big(\mathbb{E}(\xi \mid \mathcal{G}_2) \mid \mathcal{G}_1\big) = \mathbb{E}(\xi \mid \mathcal{G}_2) \text{ a.s.}$$

(iv) $\mathbb{E}\big(\mathbb{E}(\xi \mid \mathcal{G})\big) = \mathbb{E}\xi$.

Convergence results for $\mathbb{E}$ extend to $\mathbb{E}(\ldots \mid \mathcal{G})$.

Theorem (Dominated convergence). Suppose $|\xi_n| \le \eta$

$\mathbb{E}\eta < \infty$ and $\xi_n \to \xi$ a.s. $(n \to \infty)$.

Then $\mathbb{E}(\xi_n \mid \mathcal{G}) \to \mathbb{E}(\xi \mid \mathcal{G})$ a.s.

Conditional probability

$$\mathbb{P}(A \mid \mathcal{G}) = \mathbb{E}(\mathbb{1}_A \mid \mathcal{G}) \text{ is a random variable.}$$

Example ξ, η jointly distributed with density

$$f_{\xi,\eta}(x,y).$$

Let

$$f_{\xi|\eta=y}(x) = \frac{f_{\xi,\eta}(x,y)}{f_\eta(y)}.$$

$\sigma(\eta)$

Then $\mathbb{E}(\xi|\eta)$ is a random variable, obtained by computing

$$\underbrace{\mathbb{E}(\xi|\eta=y) = \int_{-\infty}^{\infty} x\, f_{\xi|\eta=y}(x)\,dx}_{}$$

$\hat{h}(y)$ " some function

then setting $\mathbb{E}(\xi|\eta) = \hat{h}(\eta)$.

More generally

$$\mathbb{E}(g(\xi)|\eta=y) = \int_{-\infty}^{\infty} g(x) f_{\xi|\eta=y}(x)\,dx$$

$\hat{h}(y)$ "

$$\underline{\mathbb{E}(g(\xi)|\eta) = \hat{h}(\eta)}$$

Sufficiency in Statistics.

Setup: $(\Omega, \mathcal{F}, (\mathbb{P}_\theta, \theta \in \Theta))$

family of prob. measures
θ "unknown" parameter.

Definition A sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$ is *sufficient* for family $(\mathbb{P}_\theta, \theta \in \Theta)$, if for all $A \in \mathcal{F}$

$$\mathbb{P}_\theta (A \mid \mathcal{G})(\omega) = P(A, \omega) \quad (\mathbb{P}_\theta\text{-a.s.})$$

does not depend on θ.

Thm (Factorisation Thm). Let $\mathbb{P}_\theta$'s be abs. cont. with respect to σ-finite measure μ on $(\Omega, \mathcal{F})$,

$$f_\theta(\omega) = \frac{d\mathbb{P}_\theta}{d\mu}(\omega) \qquad \text{R-N derivative.}$$

A σ-algebra $\mathcal{G} \subset \mathcal{F}$ is sufficient if

$$f_\theta(\omega) = g_\theta(\omega) \cdot h(\omega)$$

where $g_\theta(\omega)$ is $\mathcal{G}$-measurable.

<u>Example</u> $\Omega = \mathbb{R}^n$, $\mathcal{F} = \mathcal{B}(\mathbb{R}^n)$, $\theta > 0$

Let λ n-dim. Lebesgue measure

$$\frac{d\mathbb{P}_\theta}{d\lambda}(\omega) = \begin{cases} \frac{1}{\theta^n} & \text{if } 0 \le x_i \le \theta, \ i = 1, \dots, n \\ 0 & \text{othw.} \end{cases}$$

$\omega = (x_1, \dots, x_n)$

under this $\mathbb{P}_\theta$ $X_1, \dots, X_n$ i.i.d. Uniform $[0, \theta]$.

Then

$$f_\theta(\omega) = g_\theta(T(\omega))\, h(\omega) \quad \text{for}$$

$$h(x_1, \dots, x_n) = 1(x_1, \dots, x_n \ge 0)$$

$$g_\theta(t) = \frac{1}{\theta^n} 1(0 \le t \le \theta)$$

$$T(x_1, \dots, x_n) = \max(x_1, \dots, x_n)$$

$\mathcal{G} = \sigma(T)$ sufficient σ-algebra

T is sufficient statistic.

<u>Example</u> Exponential family $\Omega = \mathbb{R}^n$

$\omega = (x_1, \dots, x_n)$

$$\mathbb{P}_\theta(d\omega) = \mathbb{P}_\theta(dx_1) \dots \mathbb{P}_\theta(dx_n)$$

$$\mathbb{P}_\theta(dx) = \alpha(\theta)\, e^{\beta(\theta) S(x)} \gamma(x)\, \lambda(dx), \quad x \in \mathbb{R}$$

$$\mathbb{P}_\theta(dx_1 \dots dx_n) = \alpha^n(\theta)\, e^{\beta(\theta)[S(x_1) + \dots + S(x_n)]} \gamma(x_1) \dots \gamma(x_n)\, dx_1 \dots dx_n$$

$$T(x_1, \dots, x_n) = \cancel{S(x)}\ S(x_1) + \dots + S(x_n)$$

suff. statistic.

Example: Exponential(θ)-distribution.