MTP24, Lecture 3: Conditioning, Martingales and Convergence

Alexander Gnedin

Queen Mary, University of London

Conditioning relative to partition

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\Omega = \cup_n A_n$ partition in disjoint $A_n \in \mathcal{F}$ with $\mathbb{P}[A_n] > 0$, and let $\mathcal{G} = \sigma(A_1, A_2, \ldots)$. Given $\mathcal{G}$ the conditional probability of $B \in \mathcal{G}$ is a r.v.

$$\mathbb{P}[B|\mathcal{G}] = \sum_n \mathbb{P}[B|A_n]\mathbf{1}_{A_n},$$

where $\mathbb{P}[B|A_n] = \mathbb{P}[B \cap A_n]/\mathbb{P}[A_n]$, and the conditional expectation of a r.v. X is a r.v.

$$\mathbb{E}[X|\mathcal{G}] = \sum_n \frac{\mathbb{E}[X\,\mathbf{1}_{A_n}]}{\mathbb{P}[A_n]}\,\mathbf{1}_{A_n}.$$

Note:

$$\mathbb{P}[B|\mathcal{G}] = \mathbb{E}[\mathbf{1}_B|\mathcal{G}].$$

Q: How define conditioning on null events?

Examples

X, Y standard normal r.v. (and jointly normal) with $\rho = \mathbb{E}[XY]$,

$$\mathbb{E}[X|Y = y] =?$$

• Consider Bernoulli(P) trials with $P \stackrel{d}{=} \mathrm{Uniform}[0, 1]$. What is the probability of k 1's in n trials given $P = p$?

• Let S_n be the number of 1's in n Bernoulli(1/2) trials. Given $\lim_{n\to\infty} S_n/n = p$ (where $p \neq 1/2$), what is the probability that the first trial is 1?

Conditional expectation relative to σ-algebra

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, X a nonnegative r.v. and $\mathcal{G} \subset \mathcal{F}$ a sub-σ-algebra.

Definition The conditional expectation of X given $\mathcal{G}$ is a random variable $\mathbb{E}[X|\mathcal{G}]$ such that

- $\mathbb{E}[X|\mathcal{G}]$ is $\mathcal{G}$-measurable,
- for every $A \in \mathcal{G}$

$$\int_A X(\omega)\mathbb{P}(\mathrm{d}\omega) = \int_A \mathbb{E}[X|\mathcal{G}]\,\mathbb{P}(\mathrm{d}\omega)$$

(which can also be written as $\mathbb{E}[X\,\mathbf{1}_A] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]\,\mathbf{1}_A]$).

For the general X,

$$\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X_+|\mathcal{G}] - \mathbb{E}[X_-|\mathcal{G}],$$

provided at least one of the terms is finite a.s.

Existence and uniqueness

Let $X \geq 0$, then

$$\mathbb{Q}(A) = \mathbb{E}[X\,\mathbf{1}_A] = \int_A X(\omega)\mathbb{P}(\mathrm{d}\omega), \quad A \in \mathcal{G}$$

is a probability measure on $(\Omega, \mathcal{G})$ satisfying $\mathbb{P} \gg \mathbb{Q}$. By the Radon-Nikodým Theorem there exists $\mathcal{G}$-measurable r.v. $\xi = \mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P}$ such that

$$\mathbb{Q}(A) = \int_A \xi(\omega)\mathbb{P}(\mathrm{d}\omega),$$

so we set $\xi =: \mathbb{E}\left[X|\mathcal{G}\right]$.

- $\mathbb{E}\left[X|\mathcal{G}\right]$ is unique up to event of $\mathbb{P}$-probability zero, since

$$\mathbb{E}\left[X\,\mathbf{1}_A\right] = \mathbb{E}\left[Y\,\mathbf{1}_A\right], \quad A \in \mathcal{G}$$

only implies $\mathbb{P}[X = Y] = 1$.

Properties of the conditional expectation

- The conditional probability of $A \in \mathcal{F}$ given $\mathcal{G}$ is

$$\mathbb{P}[A|\mathcal{G}] = \mathbb{E}[\mathbf{1}_A|\mathcal{G}],$$

which is a random variable!
- Iterated conditioning, tower property:

$$\mathcal{G}_1 \subset \mathcal{G}_2 \Rightarrow \mathbb{E}\,[\,\mathbb{E}\,[X|\mathcal{G}_2]\,|\,\mathcal{G}_1]] = \mathbb{E}\,[X|\mathcal{G}_1].$$

- $\mathbb{E}[aX + bY|\mathcal{G}] = a\,\mathbb{E}[X|\mathcal{G}] + b\,\mathbb{E}[Y|\mathcal{G}]$ a.s.
- If X is $\mathcal{G}$-measurable, then

$$\mathbb{E}[X|\mathcal{G}] = X \;\; \text{a.s.}, \;\; \mathbb{E}[XY|\mathcal{G}] = X\,\mathbb{E}[Y|\mathcal{G}] \;\; \text{a.s.},$$

provided $\mathbb{E}\,|X| < \infty, \mathbb{E}\,|XY| < \infty$.
- $X \leq Y$ a.s. $\Rightarrow \mathbb{E}[X|\mathcal{G}] \leq \mathbb{E}[Y|\mathcal{G}]$ a.s.

$\mathbb{E}[X|Y]$ as a function of Y

For r.v. X, Y

$$\mathbb{E}[X|Y] := \mathbb{E}[X|\sigma(Y)]$$

- There exists a function $h(y)$ such that

$$\mathbb{E}[X|Y](\omega) = h(Y(\omega)), \quad \omega \in \Omega.$$

We call $h(y)$ *conditional expectation of* X *given* $Y = y$ and write

$$\mathbb{E}[X|Y = y] := h(y).$$

This satisfies the characteristic identity: for $B \in \mathcal{B}(\mathbb{R})$

$$\mathbb{E}[X\,\mathbf{1}_{\{Y \in B\}}] = \int_B h(y)\mathrm{d}F_Y(y),$$

where F_Y is the c.d.f. of Y.

If there exists a joint density $f_{X,Y}(x, y)$,

$$\mathbb{E}[X|Y = y] = \int_{-\infty}^{\infty} x \;\; f_{X|Y=y}(x)\mathrm{d}x.$$

Examples

• For X, Y $\mathcal{N}(0,1)$-r.v. (and jointly normal) with correlation ρ we have $h(y) = \rho y$, so

$$\mathbb{E}[X|Y] = \rho Y, \quad \mathbb{E}[X|Y = y] = \rho y.$$

.

• In Bernoulli(P) trials, for S_n= number of 1's

$$\mathbb{P}[S_n = k|P = p] = \binom{n}{k} p^k (1-p)^{n-k},$$

so unconditionally for $P \stackrel{d}{=} \text{Uniform}[0,1]$

$$\mathbb{P}[S_n = k] = \binom{n}{k} \int_0^1 p^k (1-p)^{n-k} \mathrm{d}p = \frac{1}{n+1}, \quad 0 \le k \le n.$$

Regular conditional probability

For disjoint $A_n \in \mathcal{F}$ and $\mathcal{G} \subset \mathcal{F}$

$$\mathbb{P}\left[\bigcup_n A_n \,\middle|\, \mathcal{G}\right](\omega) = \sum_n \mathbb{P}\left[A_n|\mathcal{G}\right](\omega)$$

holds only *almost surely*, so for *fixed* ω in general this cannot be considered as a probability measure on $\mathcal{F}$.

Definition A function $P(\omega, A)$ is called *regular conditional probability* given $\mathcal{G}$ if

1. $P(\omega, \cdot)$ is a probability measure on $\mathcal{F}$ for every $\omega \in \Omega$,
2. for every $A \in \mathcal{F}$

$$P(\omega, A) = \mathbb{P}[A \,|\, \mathcal{G}](\omega) \quad \text{a.s.}$$

Regular conditional distribution

Definition/Theorem Let X be a r.v. with values in a 'good' (i.e. Borel) measurable space $(E, \mathcal{A})$ then there exists a *regular conditional distribution given* $\mathcal{G}$, which is a function $Q(\omega, A)$ such that

1. $Q(\omega, \cdot)$ is a probability measure on $(E, \mathcal{A})$ for every $\omega \in \Omega$,
2. for $A \in \mathcal{A}$ the function $Q(\cdot, A)$ satisfies

$$Q(\omega, A) = \mathbb{P}[X \in A|\mathcal{G}](\omega) \quad \text{a.s.}$$

In particular, if X takes values in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, there exists a regular conditional distribution function such that

$$F(\omega, x) = \mathbb{P}[X \leq x \,|\, \mathcal{G}](\omega) \quad \text{a.s.}$$

and

$$\mathbb{E}[g(X)|\mathcal{G}](\omega) = \int_{-\infty}^{\infty} g(x) \, \mathrm{d}_x F(\omega, x), \quad \omega \in \Omega.$$

Sufficiency in Statistics

Statistical model: $(\Omega, \mathcal{F}, (\mathbb{P}_\theta, \theta \in \Theta))$, θ unknown parameter.

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

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

Factorisation Theorem Suppose $\mu \gg \mathbb{P}_\theta$ for a σ-finite measure μ on $(\Omega, \mathcal{F})$, and let

$$f_\theta(\omega) = \frac{\mathrm{d}\mathbb{P}_\theta}{\mathrm{d}\mu}$$

be the Radon-Nikodým derivative. Then $\mathcal{G}$ is sufficient if and ony if,

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

for some $\mathcal{G}$-measurable function $g_\theta(\omega)$ and $\mathcal{F}$-measurable $h(\omega)$.

Example For the multivariate normal distribution

$$f_\theta(x_1,\ldots,x_n) = \frac{1}{(2\pi\theta)^{n/2}} \exp\left(-\frac{x_1^2+\ldots+x_n^2}{2\theta}\right),\ \theta > 0$$

in $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ a sufficient σ-algebra is $\mathcal{G} = \sigma(T)$, where $T(x_1,\ldots,x_n) = x_1^2 + \cdots + x_n^2$. The factorisation holds with

$$g_\theta(x_1,\ldots,x_n) = \frac{1}{(2\pi\theta)^{n/2}} \exp\left(-\frac{T(x_1,\ldots,x_n)}{2\theta}\right)$$

and $h(x_1,\ldots,x_n) \equiv 1$. If $(X_1,\ldots,X_n)$ has this density f_θ, then the conditional distribution of $(X_1,\ldots,X_n)$ given $T(X_1,\ldots,X_n) = r^2$ is the uniform distribution of the sphere with radius r (for every $\theta > 0$).

Example For n Bernoulli(p) trials a sufficient statistic is S_n (the number of 1's).

Martingales

Definition A sequence of (real-valued) r.v.'s $(X_n, n \geq 0)$ is *adapted* to given filtration of σ-algebras $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots$ if X_n is measurable w.r.t. $\mathcal{F}_n$. The sequence is a *martingale* if $\mathbb{E}|X_n| \leq \infty$ and

$$\mathbb{E}[X_{n+1}|\mathcal{F}_n] = X_n, \qquad n = 0, 1, \ldots$$

submartingale if $\mathbb{E}[X_{n+1}|\mathcal{F}_n] \geq X_n$, *supermartingale* if $\mathbb{E}[X_{n+1}|\mathcal{F}_n] \leq X_n$.

- For $X_1, X_2, \ldots$ i.i.d. with $\mathbb{E}[X_1] = \mu$, the random walk

$$S_n = \xi_1 + \cdots + \xi_n$$

is a martingale adapted to the filtration $\mathcal{F}_n = \sigma(\xi_0, \ldots, \xi_n)$ if $\mu = 0$ (submartingale if $\mu > 0$, supermartingale if $\mu < 0$).

- For $\xi_1, \xi_2, \ldots$ i.i.d. with $\mathbb{E}[\xi_1] = 1$,

$$\Pi_n = \prod_{j=1}^{n} \xi_j$$

is a martingale adapted to $\mathcal{F}_n = \sigma(\xi_0, \ldots, \xi_n)$.

- If (X_n) is a martingale and $g(x)$ convex function, then $Y_n = g(X_n)$ is a submartingale (by Jensen's inequality).

- For r.v. ξ with $\mathbb{E}|\xi| < \infty$ *Doob martingale* is

$$X_n = \mathbb{E}[\xi \,|\, \mathcal{F}_n], \quad n \geq 0.$$

Martingale convergence

Theorem Let $(X_n, n \geq 0)$ be a submartingale with $\sup_n \mathbb{E}|X_n| < \infty$. Then the exists a r.v. X_∞ with $\mathbb{E}|X_n| < \infty$ such that

$$X_n \to X_\infty \text{ as } n \to \infty \text{ a.s.}$$

If the uniform integrability condition holds

$$\lim_{c \to \infty} \sup_n \mathbb{E}\left[|X_n| \mathbf{1}(|X_n| > c)\right] = 0,$$

then also $\mathbb{E}|X_n - X_\infty| \to 0$.

Example (Doob martingale) If $\mathbb{E}|\xi| < \infty$ then for $\mathcal{F}_\infty := \sigma\left(\cup_n \mathcal{F}_n\right)$

$$\mathbb{E}[\xi|\mathcal{F}_n] \to \mathbb{E}[\xi|\mathcal{F}_\infty] \quad \text{a.s.}$$

Application to exchangeable processes

Let $\xi_1, \xi_2, \ldots$ be 0-1 r.v.'s which are *exchangeable*: for every $n \geq 1$ and permutation $\pi : \{1, \ldots, n\} \to \{1, \ldots, n\}$

$$(\xi_1 \ldots, \xi_n) \stackrel{d}{=} (\xi_{\pi_1} \ldots, \xi_{\pi_n}).$$

Let $S_n = \xi_1 + \cdots + \xi_n, \quad \mathcal{G}_n = \sigma(S_n, S_{n+1}, \ldots), \mathcal{G}_\infty := \cap_n \mathcal{G}_n$.

By the *backward martingale convergence theorem* applied to the *backward* Doob martingale $\mathbb{P}[S_1 = 1 | \mathcal{G}_n]$ (which is adapted to the *falling* tower of σ-algebras $\mathcal{G}_0 \supset \mathcal{G}_1 \supset \cdots$) we obtain

$$\mathbb{P}[S_1 = 1 | \mathcal{G}_n] \to \mathbb{P}[S_1 = 1 | \mathcal{G}_\infty] \quad \text{a.s.}$$

But S_n is a sufficient statistic for $(\xi_1 \ldots, \xi_n)$, and by exchangeability

$$\mathbb{P}[S_1 = 1 | \mathcal{F}_n] = \mathbb{P}[S_1 = 1 | S_n] = \frac{S_n}{n},$$

so S_n/n converges a.s. to some $\mathcal{G}_\infty$-measurable r.v. P and

$$\mathbb{P}[S_1 = 1 | \mathcal{F}_n] \to \mathbb{P}[S_1 = 1 | \mathcal{F}_\infty] = \lim_{n \to \infty} \frac{S_n}{n}.$$

We obtain

$$\mathbb{P}[S_1 = 1|P] = P$$

under every $\mathbb{P}$ making $\xi_1, \xi_2, \ldots$ exchangeable, and finally

$$\mathbb{P}[S_1 = 1|P = p] = p.$$

This holds for any 0-1-valued exchangeable $\xi_1, \xi_2, \ldots$, in particular for Bernoulli(1/2).

Stopping times

Definition *Stopping time* adapted to filtration $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots$ is a r.v. τ with values in $\{0, 1, \cdots, \infty\}$, s.t.

$$\{\tau = n\} \in \mathcal{F}_n, \quad n = 0, 1, \ldots$$

If $\tau < \infty$ a.s. the stopping time is called *finite*.

For $\xi_0, \xi_1, \ldots$ with natural filtration, examples of stopping times are $\tau = \min\{n : \xi_n > c\}, \tau = \min\{n > 0 : \xi_n > \xi_0\}$ etc.

The stopped variable is defined as

$$\xi_\tau = \sum_{n=0}^{\infty} \xi_n \mathbf{1}_{\{\tau = n\}}$$

and the stopped process as

$$\xi_{\tau \wedge n}, \ n \geq 0.$$

Proposition If (X_n) is a martingale (sub-, super-) then $(X_{\tau \wedge n}, \ n \geq 0)$ is a martingale (sub-, super-) too.

Doob's optional sampling

Theorem Let (X_n) be supermartingale, τ stopping time. Then

$$\mathbb{E}[X_\tau] \leq \mathbb{E}[X_0]$$

if at least one of the following holds:

(i) $\mathbb{P}(\tau < K) = 1$ for some $K > 0$,

(ii) $\sup_n |X_n| < K$ a.s.

(iii) $\mathbb{E}[\tau] < \infty$ and $\sup_n \mathbb{E}\left[|X_{n+1} - X_n| \,|\mathcal{F}_n\right] < K$,

(iv) (X_n) is uniformly integrable.

Application to Gambler's Ruin

Symmetric random walk $S_n = \xi_1 + \ldots + \xi_n, n \geq 0$, with i.i.d. increments, $\mathbb{P}[\xi_n = \pm 1] = 1/2$, $S_0 = 0$. Duration of the game with (positive) initial fortunes A, B is the stopping time

$$\tau = \min\{n : S_n \in \{-A, B\}\}.$$

The Optional Sampling gives

$$0 = \mathbb{E}[S_\tau] = -A\,\mathbb{P}[S_\tau = -A] + B\,\mathbb{P}[S_\tau = B].$$

Solving this together with the total probability equation $\mathbb{P}[S_\tau = -A] + \mathbb{P}[S_\tau = B] = 1$, the ruin probabilities are found as

$$\mathbb{P}[S_\tau = -A] = \frac{B}{A+B}, \quad \mathbb{P}[S_\tau = B] = \frac{A}{A+B}.$$