CLASSICAL VERSUS COHERENT CONDITIONAL PROBABILITIES

Pietro Rigo
University of Pavia

The Mathematics of Subjective Probability
Milano, september 3-5, 2018

Classical (Kolmogorovian) conditional probabilities

Let $(\Omega, \mathcal{A}, P)$ be a probability space and $\mathcal{G} \subset \mathcal{A}$ a sub-σ-field.

A **regular conditional distribution (rcd)** is a map Q on $\Omega \times \mathcal{A}$ such that

(i) $Q(\omega, \cdot)$ is a probability on $\mathcal{A}$ for $\omega \in \Omega$

(ii) $Q(\cdot, A)$ is $\mathcal{G}$-measurable for $A \in \mathcal{A}$

(iii) $P(A \cap B) = \int_B Q(\omega, A)\, P(d\omega)$ for $A \in \mathcal{A}$ and $B \in \mathcal{G}$

An rcd can fail to exist. However, it exists and is a.s. unique under mild conditions ($\mathcal{A}$ countably generated and P perfect)

In the standard framework, thus, conditioning is with respect to a σ-field $\mathcal{G}$ and not with respect to an event H.

What does it mean ?

According to the usual interpretation: For each $B \in \mathcal{G}$, we now whether B is true or false. This naive interpretation is dangerous.

Example 1 Let $X = \{X_t : t \geq 0\}$ be a process adapted to a filtration $\mathcal{F} = \{\mathcal{F}_t : t \geq 0\}$. Suppose

$P(X = x) = 0$ for each path x and

$\{A \in \mathcal{A} : P(A) = 0\} \subset \mathcal{F}_0$.

In this case,

$\{X = x\} \in \mathcal{F}_0$ for each path x.

But then we can stop. We already know the X-path at time 0 !

Example 2 (Borel-Kolmogorov paradox) Suppose

$$\{X = x\} = \{Y = y\}$$

for some random variables X and Y. Let Q_X and Q_Y be rcd's given $\sigma(X)$ and $\sigma(Y)$. Then,

$$P(\cdot\,|X = x) = Q_X(\omega, \cdot) \text{ and } P(\cdot\,|Y = y) = Q_Y(\omega, \cdot)$$

where $\omega \in \Omega$ meets $X(\omega) = x$ and $Y(\omega) = y$. Hence it may be that

$$P(\cdot\,|X = x) \neq P(\cdot\,|Y = y) \text{ even if } \{X = x\} = \{Y = y\}.$$

Example 3 For the naive interpretation to make sense, Q should be **proper**, i.e.

$$\boxed{Q(\omega, \cdot) = \delta_\omega \text{ on } \mathcal{G} \text{ for almost all } \omega}.$$

But Q needs not be proper. In fact, properness of Q essentially amounts to $\mathcal{G}$ countably generated

Conditional 0-1 laws

An rcd Q is 0-1 on $\mathcal{G}$ if

$$\boxed{Q(\omega,\cdot) \in \{0,1\} \text{ on } \mathcal{G} \text{ for almost all } \omega}$$

Why to focus on such a 0-1 law ?

- It is a (natural) consequence of properness

- It is equivalent to

  $$\boxed{\mathcal{A} \text{ independent } \mathcal{G}, \text{ under } Q(\omega,\cdot), \text{ for almost all } \omega}$$

- It is basic for integral representation of invariant measures

- It is not granted. It typically fails if $\{A \in \mathcal{A} : P(A) = 0\} \subset \mathcal{G}$

Theorem 1

Let $\mathcal{G}_n \subset \mathcal{A}$ be a sub-σ-field and Q_n an rcd given $\mathcal{G}_n$.

The rcd Q is 0-1 on $\mathcal{G}$ if

- The "big" σ-field $\mathcal{A}$ is countably generated
- Q_n is 0-1 on $\mathcal{G}_n$ for each n
- $E(1_A|\mathcal{G}_n) \to E(1_A|\mathcal{G})$ a.s. for $A \in \mathcal{A}$ and $\mathcal{G} \subset \limsup_n \mathcal{G}_n$

Note that, by martingale convergence, the last condition is automatically true if the sequence $\mathcal{G}_n$ is monotonic

Examples

Let S be a Polish space, $\mathcal{B} = \text{Borel}(S)$, and

$(\Omega, \mathcal{A}) = (S^\infty, \mathcal{B}^\infty)$

Theorem 1 applies to

Tail σ-field: $\mathcal{G} = \cap_n \sigma(X_n, X_{n+1}, \ldots)$

where X_n is a sequence of real random variables

Symmetric σ-field:

$\mathcal{G} = \{B \in \mathcal{B}^\infty : B = f^{-1}(B)$ for each finite permutation $f\}$

In particular,

Theorem 1 $\Rightarrow$ de Finetti's theorem

Open problem: Theorem 1 does not apply to the **shift-invariant σ-field:**

$$\mathcal{G} = \{B \in \mathcal{B}^\infty : B = s^{-1}(B)\}$$

where $s(x_1, x_2, \ldots) = (x_2, x_3, \ldots)$ is the shift

Disintegrability

The notion of disintegrability makes sense in both the classical and the coherent frameworks

Let $\Pi \subset \mathcal{A}$ be a partition of Ω. P is disintegrable on Π if

$$\boxed{P(A) = \int_{\Pi} P(A|H)\, P^*(dH)}$$

for each $A \in \mathcal{A}$, where

- $P(\cdot|H)$ is a finitely additive probability (f.a.p.) on $\mathcal{A}$ such that

  $$\boxed{P(H|H) = 1}$$

- P^* is a f.a.p. on the power set of Π

Let $\mathcal{A}_0$ denote the σ-field on Ω generated by the maps

$\omega \mapsto P[A|H(\omega)]$ for all $A \in \mathcal{A}$

- If $\mathcal{A}_0 \subset \mathcal{A}$, one trivially obtains

  $P(A) = \int_\Pi P(A|H)\, P^*(dH) = P^*\{H \in \Pi : H \subset A\}$

  for all $A \in \mathcal{A}_0$. Thus, P^* essentially agrees with $P|\mathcal{A}_0$ and, with a slight abuse of notation, one can write

  $P(A) = \int_\Pi P(A|H)\, P(dH)$.

  The reason for involving P^* is that, in general, $\mathcal{A}_0$ needs not be included in $\mathcal{A}$

- The pair $(P^*, P(\cdot|\cdot))$ is said to be a disintegration for P. It is called a σ-additive disintegration if P^* is σ-additive on $\mathcal{A}_0$ and $P(\cdot|H)$ is σ-additive on $\mathcal{A}$ for each $H \in \Pi$

Theorem 2

Given a partition Π of Ω, let

$G = \{(x, y) \in \Omega \times \Omega : x \sim y\}$.

Then, P admits a σ-additive disintegration on Π whenever

- $(\Omega, \mathcal{A})$ is nice (e.g. a standard space)
- G is a Borel subset of $\Omega \times \Omega$

Remark: G is actually a Borel set if Π is the partition in the atoms of the tail, or the symmetric, or the shift invariant σ-fields

Remark: The condition on G can be relaxed. Indeed, it suffices G coanalytic, or else G analytic and every member of Π a G_δ or an F_σ set.

Open problem: If $\Omega = [0, 1]$, $\mathcal{A}$ the Borel σ-field and P the Lebesgue measure, is P disintegrable on *any* partition $\Pi \subset \mathcal{A}$?

Note: The answer is actually yes under suitable axioms of set theory. Under the Martin axiom, for instance, the Lebesgue measure on $[0, 1]$ admits a σ-additive disintegration on every Borel partition

Coherent (de Finettian) conditional probabilities

A different notion of conditioning is as follows.

Let

$P(\cdot|\cdot) : \mathcal{A} \times \mathcal{G} \to R$.

For all $n \geq 1$, $c_1, \ldots, c_n \in R$, $A_1, \ldots, A_n \in \mathcal{A}$ and $B_1, \ldots, B_n \in \mathcal{G} \setminus \emptyset$, define

$G(\omega) = \sum_{i=1}^{n} c_i \, 1_{B_i}(\omega) \, \{1_{A_i}(\omega) - P(A_i|B_i)\}$.

Then, $P(\cdot|\cdot)$ is coherent if

$\sup_{\omega \in B} G(\omega) \geq 0$ where $B = \cup_{i=1}^{n} B_i$

Such a definition has both merits and drawbacks. In particular, contrary to the classical case:

- The conditioning is now with respect to events,

- $P(B|B) = 1$,

- For fixed B, $P(\cdot|B)$ is "only" a f.a.p.,

- Disintegrability on Π is not granted, where Π is the partition of Ω in the atoms of $\mathcal{G}$

Bayesian inference

$(\mathcal{X}, \mathcal{E})$ sample space, $(\Theta, \mathcal{F})$ parameter space,

$\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ statistical model,

A **prior** is a f.a.p. π on the power set of Θ,

A **posterior** for π is any collection $\mathcal{Q} = \{Q_x : x \in \mathcal{X}\}$ such that

- Q_x is a f.a.p. on $\mathcal{F}$ for each $x \in \mathcal{X}$
- $Q_x(d\theta)\, m(dx) = P_\theta(dx)\, \pi(d\theta)$ on $\mathcal{E} \otimes \mathcal{F}$

  for some f.a.p. m on the power set of $\mathcal{X}$

The posterior $\mathcal{Q}$ is σ-additive if

- Q_x is σ-additive for each $x \in \mathcal{X}$ and m is σ-additive on the σ-field generated by the map $x \mapsto Q_x$

Theorem 3

Fix a measurable function T on $\mathcal{X}$ (a statistic) such that

$P_\theta(T = t) = 0$ for all θ and t

Under mild conditions, for any prior π, there is a posterior $\mathcal{Q}$ for π such that

$T(x) = T(y) \;\Rightarrow\; Q_x = Q_y$

Moreover, $\mathcal{Q}$ is σ-additive if the prior π is σ-additive on $\mathcal{F}$

Interpretation:

In a subjective framework, the condition

$$T(x) = T(y) \;\Rightarrow\; Q_x = Q_y$$

means that T is **sufficient** for $\mathcal{Q}$. Suppose you start with a prior π, describing your feelings on θ, and a statistic T, describing how different samples affect your inference on θ. Theorem 3 states that, whatever π and T (provided $P_\theta(T = t) = 0$) there is a posterior $\mathcal{Q}$ for π which makes T sufficient. In addition, $\mathcal{Q}$ can be taken to be σ-additive if the prior π is σ-additive on $\mathcal{F}$

Point estimation

Suppose $\Theta \subset R$ and $d : \mathcal{X} \to \Theta$ is an estimate of θ.

Theorem 4:

Under mild conditions, if the prior π is null on compacta, there is a posterior $\mathcal{Q}$ for π such that $\int \theta^2 Q_x(d\theta) < \infty$ and

$$\boxed{E_{\mathcal{Q}}(\theta|x) = \int \theta \, Q_x(d\theta) = d(x)}$$

Interpretation: The above condition means that d is optimal under square error loss. Suppose you start with a measurable map $d : \mathcal{X} \to \Theta$, to be regarded as your estimate of θ. Theorem 4 states that, if the prior π vanishes on compacta, there is a posterior $\mathcal{Q}$ for π which makes d optimal

Remark: A prior π vanishing on compacta may look strange. In fact, it is exactly what happens with most improper priors

Warning: Abusing terminology, in Theorem 4, I said "a posterior $\mathcal{Q}$ for π". Instead, the equation

$$Q_x(d\theta)\, m(dx) = P_\theta(dx)\, \pi(d\theta)$$

holds on $\mathcal{E} \times \mathcal{F}$ (i.e., on the measurable rectangles) but not necessarily on the product σ-field $\mathcal{E} \otimes \mathcal{F} = \sigma(\mathcal{E} \times \mathcal{F})$