Conditional expectation and statistics

Salem Said

Nonlinear Signal Processing Lab
University of Melbourne

Introduction to filtering 2010

Conditional expectation and statistics

20 August

- Introduction and examples
- Inequalities and convergence

27 August

- Conditional expectation
  1. Definition, uniqueness issue
  2. Elementary properties
  3. Convergence properties
  4. Conditional probability
  5. Gaussian random variables
- Point estimation

Definition, uniqueness issue

In all generality, conditioning is with respect to $\mathcal{F}$

- $\mathbb{E}[X|\mathcal{F}]$ as projection
- $\mathbb{E}[X|\mathcal{F}]$ as Radon-Nikodym derivative

Definition : $X_{\mathcal{F}}$ is $\mathcal{F}$-measurable and for all Y, $\mathcal{F}$-measurable

$$\mathbb{E}[X|\mathcal{F}] = X_{\mathcal{F}} \text{ such that } \mathbb{E}[(X - X_{\mathcal{F}})Y] = 0$$

Definition as projection : $\mathbb{E}|X|^2 < \infty$

$$X_{\mathcal{F}} = \pi(X) \qquad \pi : L^2(\mathcal{A}) \to L^2(\mathcal{F})$$

Extension : $\mathbb{E}|X| < \infty$

$$X_n = X \wedge n \qquad X_{\mathcal{F}} = \lim_n \mathbb{E}[X_n|\mathcal{F}]$$

Extension requires continuity property

Definition, uniqueness issue

In all generality, conditioning is with respect to $\mathcal{F}$
Definition : For all Y $\mathcal{F}$-measurable

$$\mathbb{E}[X|\mathcal{F}] = X_{\mathcal{F}} \text{ such that } \mathbb{E}[(X - X_{\mathcal{F}})Y] = 0$$

Discrete example : A Euclidean projection
Conditioning random variable :
If $\mathcal{F} = \sigma(Y)$ then (all f and all $u \in \mathbb{R}$)

$$\mathbb{E}[(X - X_{\mathcal{F}})f(Y)] \Leftrightarrow \mathbb{E}\left[(X - X_{\mathcal{F}})e^{iuX}\right] = 0$$

Interpretation : When $X \in L^2(\mathcal{A})$, closest element in $L^2(\mathcal{F})$
Uniqueness : $X'_{\mathcal{F}} = X_{\mathcal{F}}$ a.s. *iff*

$$X'_{\mathcal{F}} = \mathbb{E}[X|\mathcal{F}]$$

Elementary properties

Linearity : If $Y_1 = \mathbb{E}[X_1|\mathcal{F}]$ and $Y_2 = \mathbb{E}[X_2|\mathcal{F}]$ then

$$a_1 Y_1 + a_2 Y_2 = \mathbb{E}[a_1 X_1 + a_2 X_2|\mathcal{F}]$$

Positivity : If $X \geq 0$ a.s. and $Y = \mathbb{E}[X|\mathcal{F}]$ then $Y \geq 0$ a.s.
Contraction : $|\mathbb{E}[X|\mathcal{F}]| \leq \mathbb{E}[|X||\mathcal{F}]$
Projection : If $\mathcal{F}_1 \subset \mathcal{F}_2$ then

$$\mathbb{E}[\mathbb{E}[X|\mathcal{F}_2]|\mathcal{F}_1] = \mathbb{E}[X|\mathcal{F}_1]$$

Factorization : If Y is $\mathcal{F}$-measurable

$$\mathbb{E}[XY|\mathcal{F}] = Y\mathbb{E}[X|\mathcal{F}]$$

Independence : If X, Y independent

$$\mathbb{E}[g(X, Y)|Y] = \int g(x, Y) dF_X(x)$$

Convergence properties

Conditional monotone convergence :
$X_n \uparrow X$ a.s. and $Y_n = \mathbb{E}[X_n|\mathcal{F}] \Rightarrow Y_n \uparrow Y$ where $Y = \mathbb{E}[X|\mathcal{F}]$
Conditional dominated convergence :

- (just) dominated convergence
- $X_n \to X$ in $L^1(\mathcal{A}) \Rightarrow Y_n \to Y$ in $L^1(\mathcal{F})$

To recover monotone convergence :

- Convergence in L^1 implies convergence in probability
- Increasing convergence in probability implies convergence a.s.
  More contraction properties : In L^2, ...

Conditional probability

For one event :

$$\mathbb{P}(A|\mathcal{F}) = \mathbb{E}[\mathbf{1}_A|\mathcal{F}] \qquad A \in \mathcal{A}$$

Consistency problem : For given $A_n \in \mathcal{A}$ with $A_n \uparrow A$

$$1 = \mathbb{P}[\Omega|\mathcal{F}] \qquad 0 = \mathbb{P}[\phi|\mathcal{F}] \qquad \mathbb{P}[A_n|\mathcal{F}] \uparrow \mathbb{P}[A|\mathcal{F}]\text{a.s.}$$

Precise statement : Convergence depends on A_n.
Regular conditional distribution : For random variable X,

$$F_X(x|\mathcal{F}) = \mathcal{P}[X \leq x|\mathcal{F}] \qquad \text{increasing and } \textit{càdlàg} \text{ paths}$$

Uniqueness : Regular conditional distribution is unique !

Gaussian random variables

Gaussian couple : $(X, Y) \hookrightarrow \mathcal{N}(\mu, \Sigma)$
Uncorrelated couple : $-K = \Sigma_{12}\Sigma_{22}^{-1}$

$$X' = X + KY \qquad Y' = Y$$

Direct computation :

$$\mathbb{E}[X|Y] = \mathbb{E}[X] - K(Y - \mathbb{E}[Y])$$

Projection property : K any matrix

$$\min_{Z=KY} \mathbb{E}|X - Z|^2$$

More general/easy than Bayes formula

Conditional expectation and statistics

20 August

- Introduction and examples
- Inequalities and convergence

27 August

- Conditional expectation
- Point estimation
  1. Sufficiency, minmality
  2. Ancillarity and completeness
  3. Basu's theorem
  4. Blackwell-Rao theorem

Sufficiency, minimality

Statistical model : $X : \Omega \to \mathbb{R}$ fixed

$$\sigma(X) \text{ does not depend on } F_X \in \{F_\theta\}_{\theta \in \Theta}$$

Statistics : Y such that Y is $\sigma(X)$-measurable,

$$Y = T(X) \qquad \sigma(Y) \subset \sigma(X)$$

Sufficiency : Y sufficient

$$F_X(x|Y) \text{ same for all } \theta \in \Theta$$

Example : X is sufficient $F_X(x|X) = \mathbf{1}_{[X,\infty[}(x)$
Minimality : U minimal

$$Y \text{ sufficient } \Rightarrow \sigma(U) \subset \sigma(Y)$$

Ancillarity, completeness

Ancillarity : Y ancillary

$$F_Y(y) \text{ same for all } \theta \in \Theta$$

Completeness : Y

$$\mathbb{E}[g(Y)] \text{ same for all } \theta \Rightarrow g = \mathrm{const.}$$

Example : $X = (X_1, X_2)$ with $X_1, X_2 \hookrightarrow \mathcal{N}(x, 1)$

$$T_1(X) = \frac{1}{\sqrt{2}}(X_1 + X_2) \qquad T_2(X) = \frac{1}{\sqrt{2}}(X_1 - X_2)$$

Basu's theorem : ($T_1(X)$ suff. compl., $T_2(X)$ ancillary)
Any sufficient complete statistics is independent of any ancillary statistics

Basu's theorem

Any sufficient complete statistics is independent of any ancillary statistics

Proof : Y_1 suff. compl., Y_2 ancillary

- $Y_2 = T_2(X)$,
- Sufficiency $\Rightarrow \mathbb{E}[f(Y_2)|Y_1] = g(Y_1)$ same for all $\theta \in \Theta$
- Ancillarity $\Rightarrow \mathbb{E}[f(Y_2)] = \mathbb{E}[g(Y_1)]$ same for all $\theta \in \Theta$
- Completeness $\Rightarrow g = \mathrm{const.}$
- Conclusion : $\mathbb{E}[f(Y_2)|Y_1] = \mathbb{E}[f(Y_2)]$

Example :

For a Gaussian family ($\theta = \textit{average}$) the sample mean and sample variance are independent

Blackwell-Rao theorem

How to improve a statistic : $Y = T(X)$,

$$Z \text{ sufficient } \Rightarrow Y \mapsto Y' = \mathbb{E}[Y|Z] \text{ same for all } \theta \in \Theta$$

Improved error : For all $\theta \in \Theta$

$$\mathbb{E}|Y' - \theta|^2 \leq \mathbb{E}|Y - \theta|^2$$

Importance of sufficiency : $Y' = g(Z)$ does not depend on θ
Blackwell-Rao process : Repeat !
Example :
(http://en.wikipedia.org/wiki/Rao%E2%80%93Blackwell_theorem)

- $X_1, \ldots, X_k$ Independent, Poisson with parameter λ
- $X_1 + \ldots + X_k$ sufficient for λ
- How to estimate $\mathbb{P}(X_1 = 0)$?