# Conditional expectation and statistics

Salem Said

Nonlinear Signal Processing Lab
University of Melbourne

Introduction to filtering 2010

# Organization/Plan

**Organization**

- Four topics
- Weekly session (Friday)
- Each topic covered in two or four sessions
- No sessions in September

**Plan**

- Conditional expectation and statistics
- Discrete time martingales
- Brownian stochastic calculus
- Optimal filtering equations

# Goals

**Learning**

- Sequel to Jonathan's course, "Measure theoretic probability"
- More compact and applicable language of mathematical probability
- A closer knowledge of bibliography on probability and statistics

**Applications**

- Statistics : point estimation
- Probability : Martingales
- Stochastic calculus : Itô integral
- Filtering : Derivation and general form

# Conditional expectation and statistics

**20 August**

- Introduction and examples :
  1. Kolmogorov's (measure theoretic) axioms
  2. Three way of interpreting probability ?
  3. First example of conditional expectation
  4. Random variables
  5. Expectation
  6. Distribution of a random variable
  7. Independence
- Inequalities and convergence

**27 August**

- Conditional expectation
- Point estimation

# 1-Kolmogorov axioms

Kolmogorov, Foundations of probability, 1933 : “This task would have been a rather hopeless one before the introduction of Lebesgue’s theories ...”

**Probability space** : Let $(\Omega, \mathcal{A}, \mathbb{P})$ ...

- $\Omega$ sample space, probability space : set of possible outcomes
- $\mathcal{A} \subset 2^{\Omega}$ a $\sigma$-field : set of “measurable” events
  - $\Omega, \phi \in \mathcal{A}$
  - $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$
  - $(A_n, n \in \mathbb{N}) \subset \mathcal{A} \Rightarrow \cup_{n \in \mathbb{N}} A_n \in \mathcal{A}$
- $\mathbb{P} : \mathcal{A} \to [0, 1]$ probability measure : how to choose it ?
  - $\mathbb{P}(\Omega) = 1$, $\mathbb{P}(\phi) = 0$
  - $(A_n, n \in \mathbb{N})$ pairwise disjoint, $\mathbb{P}(\cup_n A_n) = \sum_n \mathbb{P}(A_n)$

**Continuity of measure** : $A_n \uparrow A \Rightarrow \mathbb{P}(A_n) \uparrow \mathbb{P}(A)$

**Random variables** : A class of functions $X : \Omega \to \mathbb{R}$

# 2-Three interpretations of probability

- Classical (*a priori*) probability
  - Principle of indifference, equal probabilities (Laplace)
  - Criticism : problem of a loaded coin
- Subjective probability
  - Probability is a degree of belief (De Finetti, Keynes)
  - Criticism : Idealist concept, not measurable
- Frequentist probability
  - Probability is the limit of a relative frequency
  - "mass phenomena and repetitive events" –Von Mises
  - A measurable quantity !

Mathematical probability does not give a way of choosing the probability measure. It might even consider several probability measures at the same time : In mathematical finance, real world vs. risk neutral probability.

## 3-First example of conditional expectation

Consider a dice throwing experiment : $\Omega = \{1, \ldots, 6\}$ **Paradigm** : $\sigma$-field = information available to an observer **Three cases** : (The notion of filtration)

- No knowledge of result : $\mathcal{F}_1 = \{\phi, \Omega\}$
- Knows if result even/odd : $\mathcal{F}_2 = \{\phi, \{1,3,5\}, \{2,4,6\}, \Omega\}$
- Full knowledge on result : $\mathcal{F}_3 = 2^\Omega$

**Fair distribution of gain** : $X(\omega) = \omega, \omega = 1, \ldots, 6$

- $X_1(\omega) = p_1 X(1) + \ldots + p_6 X(6)$ for all $\omega$
- According to $\omega$ even or odd $i, j, k = 1, 3, 5$ or $2, 4, 6$

$$X_2(\omega) = [p_i X_i + p_j X_j + p_k X_k]/[p_i + p_j + p_k]$$

- $X_3(\omega) = X(\omega)$

$X_1, X_2, X_3$ are conditional expectation of $X$ given $\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3$

# 4-Random variables

Consider a function $X : \Omega \to \mathbb{R}$. $X$ is called measurable if

$$\sigma(X) \subset \mathcal{A} \qquad \sigma(X) = X^{-1}(\mathcal{B})$$

$\mathcal{B}$ Borel $\sigma$-algebra. Equivalent condition

$$X^{-1}(\{t|t \leq x\}) \in \mathcal{A} \qquad x \in \mathbb{R}$$

$X$ is called a random variable if $X$ is measurable **Measurability depends on** $\mathcal{A}$ : Consider $\mathcal{F}_1 \subset \mathcal{F}_2$

- $X$ is $\mathcal{F}_1$-measurable $\Rightarrow$ $X$ is $\mathcal{F}_2$-measurable
- Equivalence does not hold
- $\mathcal{F}_1 = \mathcal{F}_2$ *iff* equivalence holds

**Information theory** : $\sigma(X_1) \subset \sigma(X_2)$ *iff* $H(X_1|X_2) = 0$

# 4-Random variables

Let $A_1, \ldots, A_n \in \mathcal{A}$ and $A_i \cap A_j = \phi$ for $i, j = 1, \ldots, n$

**Simple random variable**

$$X = \sum_i x_i \mathbf{1}_{A_i} \qquad \sigma(X) \subset \mathcal{A}$$

**Discretization**

$$X_n = k2^{-n} \text{ over } \{k2^{-n} \leq X < (k+1)2^{-n}\}$$

**Important result** :$X$ measurable *iff* $X$ limit of simple $X_n$

- $X_1, X_2$ measurable $\Rightarrow X_1 + X_2$ measurable
- $X_1 \wedge X_2$ and $X_1 \vee X_2$ measurable
- Limit (and limit operations) of random variables is a random variable

# 5-Expectation

**Simple random variable**

$$\mathbb{E}[X] = \sum_i x_i \mathbb{P}(A_i)$$

**Definition for positive $X$** ($X$ is said integrable)

$$\mathbb{E}[X] = \sup \mathbb{E}[Y] < \infty \qquad Y \leq X \text{ simple}$$

**Approximation**

$$\mathbb{E}[X] = \lim_n \mathbb{E}[X_n]$$

$X_1, X_2$ integrable

- $\mathbb{E}[aX_1 + bX_2] = a\mathbb{E}[X_1] + b\mathbb{E}[X_2]$
- $\mathbb{P}(X < 0) = 0 \Rightarrow \mathbb{E}[X] \geq 0$
- Respects uniform convergence, but much more !

# 6-Distribution of a random variable

**Definition of c.d.f.**

$$F(x) = \mathbb{P}(X \leq x) = \lim_{t \downarrow x} F(t) \qquad \lim_{s \uparrow x} F(s) = \mathbb{P}(X < x)$$

$F$ is *càdlàg* : right continuous with left limits

$$\mathbb{E}[g(X)] = \int g(x) dF(x)$$

Decomposition of $F$ : absolutely continuous + atomic + diffuse

$$\mathbb{E}[g(X)] = \int g(x) f(x) dx \qquad F \text{ absolutely continuous}$$

**Functional representation**

$$\sigma(X_1) \subset \sigma(X_2) \Leftrightarrow X_1 = g(X_2) \qquad g \text{ regression function}$$

# 6-Distribution of a random variable

**Definition of c.d.f.**

$$F(x) = \mathbb{P}(X \leq x) = \lim_{t \downarrow x} F(t) \qquad \lim_{s \uparrow x} F(s) = \mathbb{P}(X < x)$$

$F$ is *càdlàg* : right continuous with left limits

$$\mathbb{E}[g(X)] = \int g(x) dF(x)$$

Decomposition of $F$ : absolutely continuous + atomic + diffuse

$$\mathbb{E}[g(X)] = \int g(x) f(x) dx \qquad F \text{ absolutely continuous}$$

**Inverse function trick**

$$X = F^{-1}(U) \qquad U \text{ uniform in } [0, 1]$$

# 7-Independence

$\mathcal{F}_1, \ldots, \mathcal{F}_n$ are said independent if

$$\mathbb{P}(A_1 \cap \ldots A_n) = \mathbb{P}(A_1) \ldots \mathbb{P}(A_n) \text{ for } A_1 \in \mathcal{F}_1, \ldots, A_n \in \mathcal{F}_n$$

Equivalent condition ($X_i$ is $\mathcal{F}_i$-measurable, $i = 1, \ldots, n$)

$$\mathbb{E}[X_1 \ldots X_n] = \mathbb{E}X_1 \ldots \mathbb{E}[X_n]$$

Independence of random variables,

$$\mathbb{E}[g_1(X_1)g_2(X_2)] = \mathbb{E}[g_1(X_1)]\mathbb{E}[g_2(X_2)]$$

Equivalent condition Kac's theorem

$$\mathbb{E}[\exp(iu_1X_1 + iu_2X_2)] = \mathbb{E}[\exp(iu_1X_1)]\mathbb{E}[\exp(iu_2X_2)]$$

**Three concepts** : Independence $\neq$ pairwise independence $\neq$ orthogonality

# I-Conditional expectation and statistics

### 20 August

- Introduction and examples
- Inequalities and convergence :
    1. Chebyshev inequality
    2. Norm inequalities
    3. Convergence in probability
    4. Almost sure convergence
    5. Monotone convergence
    6. Dominated convergence
    7. Jensen's inequality

### 27 August

- Conditional expectation
- Point estimation

# 1-Chebyshev and norm

**Chebyshev inequality** : Bound on tail distribution
($f(X) \geq 0$ integrable)

$$a\mathbb{P}(f(X) \geq a) \leq \mathbb{E}[f(X)] \qquad a^2\mathbb{P}(|X| \geq a) \leq \mathbb{E}|X|^2$$

**Norm inequalities** :

$$\mathbb{E}|X_1 + X_2| \leq \mathbb{E}|X_1| + \mathbb{E}|X_1| \qquad \mathbb{E}|X_1 + X_2|^2 \leq 2\mathbb{E}|X_1|^2 + 2\mathbb{E}|X_2|^2$$

General proof : Check for simple random variables
**Cauchy-Schwarz** : (Correlation coefficient $\leq 1$)

$$\mathbb{E}[X_1 X_2]^2 \leq \mathbb{E}|X_1|^2\mathbb{E}|X_2|^2 \qquad \mathbb{E}|X_1|^2, \mathbb{E}|X_2|^2 < \infty$$

Relation between $L^2$ and $L^1$ norms

$$(\mathbb{E}|X|)^2 = (\mathbb{E}1.|X|)^2 \leq \mathbb{E}|X|^2$$

The $L^2$ norm is stronger

# 2-Convergence in probability

**Definition** : $X_n \xrightarrow{P} X$

$$\mathbb{P}(|X_n - X| \geq \sigma) \to 0 \text{ for } \sigma > 0$$

**Relation to convergence in the mean** :

$$\mathbb{E}|X_n - X| \geq \sigma \mathbb{E}(|X_n - X| \geq \sigma)$$

**Relation to convergence in the square mean** :

$$\mathbb{E}|X_n - X|^2 \geq \mathbb{E}|X_n - X| \geq \sigma \mathbb{E}(|X_n - X| \geq \sigma)$$

Uniformly continuous function $f$

$$X_n \xrightarrow{P} X \Rightarrow f(X_n) \xrightarrow{P} f(X)$$

A complete metric $d(X, Y) = \mathbb{E}[|X - Y| \wedge 1]$

# 3-Almost sure convergence

**Definition** : $X_n \to X$ a.s.

$$\mathbb{P}(\cup_N A_N) = 0 \qquad A_N = \cap_{n\geq 1} \cup_{m\geq n} \{|X_m - X_m| \geq 1/N\}$$

Borel-Cantelli reasoning : $\mathbb{P}(A_N) =$

$$\lim_n \mathbb{P}(\cup_{m\geq n}\{|X_m - X_m| \geq 1/N\}) \leq \sum_{m\geq n} \mathbb{P}(|X_m - X_m| \geq 1/N)$$

Stronger than convergence in probability

$$X_n \xrightarrow{a.s} X \Rightarrow X_n \xrightarrow{P} X$$

**Counter example** : Almost sure convergence is not topological

# 4-Monotone convergence

**Statement** : Even if $\mathbb{E}|X| = +\infty$, if $\mathbb{E}|X_n| < \infty$

$$X_n \uparrow X \Rightarrow \mathbb{E}X_n \uparrow \mathbb{E}X_n$$

**Proof** : Clear when $X_n$ are simple !
**Example** :

$$\mathbb{E}[X \wedge n] \uparrow \mathbb{E}[X]$$

**Example** : Continuity of characteristic function

$$\mathbb{E}\left|e^{u_1 X} - e^{u_2 X}\right| \leq \mathbb{E}\left|1 - e^{i(u_1 - u_2)X}\right| = 0$$

**Note** : If $X_n \overset{P}{\to} X$ and increasing then $X_n \uparrow X$ a.s.

# 5-Dominated convergence

**Statement** : $|X_n| \leq Y, \mathbb{E}|Y| < \infty$

$$X_n \to X \Rightarrow \lim_n \mathbb{E}X_n = \mathbb{E}X_n$$

$Y$ controls large values of $X_n$

$$\mathbb{E}[X_n \mathbf{1}_A] \leq \mathbb{E}[Y \mathbf{1}_A] \qquad A \in \mathcal{A}$$

**Egorov's theorem** : For $\delta > 0$ there exists $A_\delta \in \mathcal{A}$, $\mathbb{P}(A_\delta^c) < \delta$ and $X_n$ converge uniformly on $A_\delta$.

**Proof** : Uniform convergence is applied locally

$$\mathbb{E}[X_n - X] \leq \mathbb{E}[X_n - X; A_\delta] + \mathbb{E}[X_n - X; A_\delta^c]$$

**Extension :** Uniform integrability

# 6-Jensen's inequality

**Statement** : Convex function $f$

$$f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)]$$

**Example** : $f(x) = |x|^p, p \geq 1$
**Example** : $f(x) = -\log(x)$ appears in information theory
**Proof** : Clear when $X$ is simple

$$f(\mathbb{E}[X]) = f(\sum x_i \mathbb{P}(A_i)) \leq \sum \mathbb{P}(A_i) f(x_i)$$

To go on, use continuity of $f$ and dominated convergence. **Other proof :** $f$ is defined at $\mathbb{E}[X]$, by convexity

$$f(X) \geq f(\mathbb{E}[X]) + K(X - \mathbb{E}[X]) \qquad K \in \mathbb{R}$$