Conditional expectation; Stochastic processes

September 21, 2006

©2006 by Christopher A. Sims. This document may be reproduced for educational and research purposes, so long as the copies contain this notice and are retained for personal use or distributed free.

Measurable functions

A function $f : S \mapsto \mathbb{R}^k$ is **$\mathcal{F}$-measurable** if and only if for every open set B in $\mathbb{R}^k$, $f^{-1}(B)$ is in $\mathcal{F}$.

Note that this looks much like one definition of a continuous function — for f to be continuous, it must be that $f^{-1}(B)$ is open for every open B. So continuous functions are always measurable with respect to the Borel field. *Example* 1. $S = \{1, 2, 3, 4, 5\}$. $\mathcal{F}$ generated by $\{1, 3\}, \{2, 4\}$. $\mathcal{F}$ consists of $\emptyset, \{1, 3\}, \{2, 4\}, \{1, 3, 5\}, \{2, 4, 5\}, \{5\}, \{1, 2, 3, 4, 5\}$. Then the identity function $f(\omega) = \omega$ is not $\mathcal{F}$-measurable, but the function $f(\omega) = \omega \bmod 2$ (i.e. f is 1 for odd arguments, 0 for even arguments) is $\mathcal{F}$-measurable.

A function **integrable** w.r.t. a measure μ defined on a σ-field $\mathcal{F}$ is an $\mathcal{F}$-measurable function f for which $\int f d\mu$ is finite.

Conditional expectation

Suppose we have a probability triple $(S, \mathcal{F}, P)$ and in addition another σ-field $\mathcal{G}$ contained in $\mathcal{F}$. If f is a P-integrable function, then it has a **conditional expectation** with respect to $\mathcal{G}$, defined as a G-measurable function $E[f \mid \mathcal{G}]$ such that

$$(\forall A \in \mathcal{G}) \int_A f(\omega) P(d\omega) = \int_A E[f \mid \mathcal{G}](\omega) P(d\omega) \; .$$

Note that

- The conditional expectation always exists.
- It is not unique, but two such conditional expectation functions can differ only on a set of probability zero.

- Common special case 1: $S = \mathbb{R}^2$, $\mathcal{F}$ is the Borel field on it, and points in S are indexed as pairs (x, y) of real numbers. $\mathcal{G}$ is the σ-field generated by all subsets of S of the form $\{(x, y) \mid a < x < b\}$, where a and b are real numbers. In this case sets in $\mathcal{G}$ are defined entirely in terms of restrictions on x, with y always unrestsricted. A $\mathcal{G}$-measurable function will be a function of x alone, therefore. In this case, we would usually write $E[f(x, y) \mid x]$ instead of the more general notation

$$E[f(x, y) \mid \mathcal{G}](x) \; .$$

- Common special case 2: $\mathcal{G}$ is the σ-field generated by the single subset A of S. (I.e., $\{\emptyset, A, A^c, S\}$). Then a $\mathcal{G}$-measurable function must be constant on A and also constant on A^c. The value of $E[f \mid \mathcal{G}](\omega)$ for $\omega \in A$ then is what is usually written as $E[f \mid A]$.

Stochastic processes

Definition 1. A **stochastic process** is a probability measure on a space of functions $\{X_t\}$ that map an index set $\mathbb{K}$ to $\mathbb{R}^n$ for some n. The index set is $\mathbb{R}$, or some subset of it.

Stochastic processes with $\mathbb{R}$ or $\mathbb{R}^+$ as index set are called **continuous-time** processes. Those with $\mathbb{Z}$ or $\mathbb{Z}^+$ as index set are called **discrete-time** processes.

An ordinary random vector $X = \{X_i,\, i = 1, \ldots, k\}$ with values in $\mathbb{R}^k$ is a special case of a discrete time process. Instead of $\mathbb{Z}$ as an index set, it has the finite set of integers $1, \ldots, k$ as index set.

There are generalizations of this idea. If the index set is a subset of $\mathbb{R}^2$, we have a spatial process. These are useful in analysis of data that may vary randomly over a geographical region.

Probability-triple defnition

An ordinary random variable X is defined as an $\mathcal{F}$-measurable function $X(\omega)$ mapping S from a probability space $(S, \mathcal{F}, P)$ to the real line. That is $X : S \mapsto \mathbb{R}$. A random vector is $X : S \mapsto \mathbb{R}^k$. A one-dimensional continuous time stochastic process is formally $X : S \mapsto \mathbb{R}^{\mathbb{R}}$, and a one-dimensional discrete-time process is formally $X : S \mapsto \mathbb{R}^{\mathbb{Z}}$.

This formalism, with the underlying space S, allows us to consider many different random variables and stochastic processes on the same S, and thus to model stochastic relationships among processes and random variables.

If we are dealing only with a single discrete (say) stochastic process, it is easier to take S to be $\mathbb{R}^{\mathbb{Z}}$ itself, so that the function on S defining the process is just the identity function.

σ-fields for stochastic processes

- Our definition of a measurable function assumes that we have a well defined class of open sets on the space in which the function takes its values. For ordinary random variables and vectors, taking their values in $\mathbb{R}^k$, the open sets are the obvious ones.

- What is the class of open sets in $\mathbb{R}^{\mathbb{R}}$ or $\mathbb{R}^{\mathbb{Z}}$? There is no unique way to choose open sets in these spaces. The standard class of open sets in these spaces for our purposes is the **cylinder sets**. These are sets of the form

$$\left\{X \in \mathbb{R}^{\mathbb{K}} \mid X_t \leq a\right\} ,$$

where t is some element of $\mathbb{K}$ and a is an element of $\mathbb{R}$ (for a one-dimensional process).

Filtrations

- On a probability space $(S, \mathcal{F}, P)$, a **filtration** is a class $\{\mathcal{F}_t\}$ of σ-fields indexed by the index set $\mathbb{K}$ such that for each $s < t \in \mathbb{K}$, $\mathcal{F}_s \subset \mathcal{F}_t$ and $\mathcal{F}_t \subset \mathcal{F}$ for all $t \in \mathbb{K}$.

- The interpretation of a filtration is that $\mathcal{F}_t$ is the collection of all events that are verifiable at t. The increase in the size of $\mathcal{F}_t$ as t increases reflects the accumulation of information over time.

- A common example of a filtration: We have a stochastic process $\{X_t\}$ defined on $(S, \mathcal{F}, P)$ and we define $\mathcal{F}_t$ to be the σ-field generated by inverse images of sets of the form $X_s(\omega) < a$ for any real number a and any $s \leq t$. Then events in $\mathcal{F}_t$ can be verified to have occurred or not by observation of X_s for $s \leq t$. $\mathcal{F}_t$ can be thought of as the class of events verifiable at time t by observation of the history of X_s up to time t. An $\mathcal{F}_t$-measurable random variable is then a function of the history of X up to time t.

Prediction

- Combining the notion of a filtration with that of a conditional expectation, we can form
$$E[Z \mid \mathcal{F}_t] = E_t[Z] \, .$$

- These are two notations for the same thing. Both are "the conditional expectation of Z given information at t". The latter notation is a shorthand used when there is only one filtration to think about.

- When $\mathcal{F}_t$ is defined in terms of the stochastic process X as in the previous section, there is a third common notation for this same concept:
$$E[Z \mid \{X_s, \, s \leq t\}] \, .$$

- When the random variable Z is X_{t+v} for $v > 0$, then $E[X_{t+v} \mid \mathcal{F}_t]$ is the minimum variance v-period ahead predictor (or forecast) for X_{t+v}.

The i.i.d. Gaussian processes

- There is a second, equivalent, way to define a stochastic process. Specify a rule for defining the joint distribution of the finite collection of random variables $\{X_{t_1}, \ldots, X_{t_n}\}$ for any set of elements $t_1, \ldots, t_n$ of $\mathbb{K}$.

- Of course the joint distributions have to be consistent. For example, I can't specify that $\{X_1, X_2\}$ form $N(0, I)$ random vector, while $\{X_2, X_4\}$ form a $N(0, 2I)$ random vector, since the variances of X_2 in the two distributions conflict.

- A simple stochastic process that is a building block for many others: $\{X_t\}$ are i.i.d. $N(0, 1)$ for $t \in \mathbb{Z}$. Or, more generally, $\{X_t\}$ are i.i.d. $N(0, I)$ random vectors.

Gaussian MA processes

- A useful class of processes: Let $\{a_i,\, i = -\infty, \ldots \infty\}$ be a set of real $n \times n$ matrices, let $\{\varepsilon_t\}$ be an n-dimensional i.i.d. $N(0, I)$ process, and define

$$X_t = \sum_{i=-\infty}^{\infty} a_i \varepsilon_{t-i} \,.$$

- We know finite linear combinations of normal variables are themselves normal. So long as $\sum a_i a_i' < \infty$,

$$\lim_{k,\ell \to \infty} \sum_{-k}^{\ell} a_i \varepsilon_{t-i}$$

  is well defined both as a limit in probability and a limit in mean square and is normal.

- Then any finite collection of X_{t_i}'s, $i = 1, \ldots, m$, is jointly normal, as it consists of linear combinations of normal variables.

- $$\mathrm{Cov}(X_t, X_s) = \sum_{v=-\infty}^{\infty} a_v a'_{v+s-t} \, .$$

- Here we are treating a_i as defined for all i, positive or negative, but with $a_i = 0$ except for $0 \leq i \leq k$.

Stationarity; The autocovariance function

- Note that for these Gaussian MA processes, $\mathrm{Cov}(X_t, X_s)$ depends only on $t - s$. That is, it depends only on the distance in time between the X's, not on their absolute location in time. We write

$$\mathrm{Cov}(X_t, X_{t-v}) = R_X(v)$$

  and call R_X the **autocovariance function** (sometimes abbreviated acf) for X.

- Note that $R_X(s) = R_X(-s)'$. Of course if $m = 1$, this becomes $R_X(s) = R_X(-s)$.

- Since this is a Gaussian process, its covariances (and mean, always zero) fully determine its joint distributions. A process that, like this one, has the property that for any $\{t_1, \ldots, t_n\} \subset \mathbb{K}$ and any $s \in \mathbb{K}$, the joint distribution of $X_{t_1} \ldots X_{t_n}$ is the same as that of $\{X_{t_1+s}, \ldots, X_{t_n+s}\}$, is called a **stationary** process.

Qualitative behavior of MA processes

- Time paths of MA processes tend to be qualitatively similar to $\{a_s\}$, considered as a function of s.

- If the a's are all of the same sign and smooth, the time paths of X will tend to be smooth. If the a's oscillate, the X's will tend to oscillate, and at about the same frequency.

Uniqueness for R_X, for a?

The fundamental MA representation

The likelihood for an MA process