CONDITIONAL EXPECTATION; STOCHASTIC PROCESSES

1. Three examples of stochastic processes

(I) X_t has three possible time paths. With probability .5 $X_t \equiv t$, with probability .25 $X_t \equiv -t$, and with probability .25 $X_t \equiv 0$.

(II) For any collection of time indexes $\tau = \{t_1, \ldots, t_n\}$, $\{X_{t_1}, \ldots, X_{t_n}\} \sim N(0, \Sigma)$, where the i, j'th element of Σ is 0 for $|i - j| > 1$, 1 for $|i - j| = 1$, and .5 for $|i - j| = 1$.

(III) $X_0 = 5$, $X_t \mid X_{t-1} \sim N(X_{t-1}, 1)$, all $t > 0$.

2. What is a stochastic process?

- Examples illustrate three ways of thinking about them:
  - Probability distributions over time paths
  - A rule for producing the joint distribution of random variables indexed by time
  - A rule for generating the next random variable in a sequence of them from the values of previous ones (recursive)
- What is t? In the first example, it could be either $\mathbb{R}$ or $\mathbb{Z}$. In the second, it might appear that it could also be either $\mathbb{R}$ or $\mathbb{Z}$. However, the rule does not work for $\mathbb{R}$: For some collections of τ values, it produces "covariance matrices" that are not positive definite. In the third example, the rule as presented only applies to $\mathbb{Z}^+$.
- We usually mean, by "stochastic processes", the case with $\mathbb{R}$ or Z or some subset thereof as the domain of the index t. There is a literature on stochastic processes where t is an abstract mathematical "group". There is also a literature with t two-dimensional or three-dimensional (spatial processes.)

3. Can we connect these approaches?

- From example 1, we can generate characterizations of either of the other two types.
- From example 3, we can generate an example 2 type representation, but type 1 is hard.
- From example 2, a type 1 representation is also hard, and even going to a type 3 representation in a systematic way is hard.
- This is all to motivate introducing a bunch of abstract notation.

Date: December 15, 2010.

4. 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.

5. 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 unrestricted. 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 $E[f \mid A]$.

6. Consciousness-expanding special case

$$(x, y) \sim N(0, I)$$
$$E[y^2 \mid x = 0] = 1, \text{ obviously (since } y \mid x \sim N(0,1), \text{ all } x)$$
$$\theta = \arcsin(y/\sqrt{(x^2 + y^2)}.$$
$$E[y^2 \mid \theta = \pi/2] \neq 1$$

The set $\{(x, y) \mid x = 0\}$ is the same as $\{(x, y) \mid \theta = \pi/2\}$.
The σ-field generated by x-intervals is different from that generated by θ-intervals.
Geometry: Pie-slices vs. ribbons.

7. 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.

8. Probability-triple definition

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.

9. 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).

10. 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.

11. 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 $s \leq t$.
- 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 a function of the history of X up to time t.

12. 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}.

13. 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 $\{X_{t_1}, \dots, X_{t_n}\}$ for any set of elements $t_1, \dots, t_n$ 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.

14. Gaussian MA processes

- A useful class of processes: Let $\{a_i,\ i = -\infty, \dots \infty\}$ 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, $i = 1, \dots, m$, is jointly normal, as it consists of linear combinations of normal variables.
- $$\operatorname{Cov}(X_t, X_s) = \sum_{v=-\infty}^{\infty} a_v a_{v+s-t}' \,.$$
- Here we are treating a_i as defined for all i, positive or negative. When such a model emerges from a behavioral model, though, most commonly $a_i = 0$ for $i < 0$. Also, for tractability one often encounters **finite-order** MA processes, in which also $a_i = 0$ for $i > k$, for some k.

15. Stationarity; The autocovariance function

- Note that for these Gaussian MA processes, $\operatorname{Cov}(X_t, X_s)$ 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

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

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 $\{t_1, \dots, t_n\} \subset \mathbb{K}$ and any $s \in \mathbb{K}$, the joint distribution of $X_{t_1} \dots X_{t_n}$ is the same as that of $\{X_{t_1+s}, \dots, X_{t_n+s}\}$, is called a **stationary** process.

16. 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.

17. UNIQUENESS FOR R_X, FOR a?

- If $\mathrm{Var}(X) < \infty$, and X is stationary, then $R_X(t)$ is uniquely defined for all t.
- If X is also Gaussian, then its mean together with R_X are all we need to define the joint distribution of $\left\{X_{t_j}\right\}$ for any finite collection of time indexes $\{t_j\}$.
- $\therefore$ the mean and R_X uniquely determine a Gaussian process.

18. WHAT CAN BE AN R_X?

- $R_X(0)$ must be positive semi-definite.
- For univariate X $R_X(0) \geq R_X(t)$ for all t. In the multivariate case $R_X(0) \geq (R_X(t) + R_X(-t))/2$ (in the sense that the difference is p.s.d.). But these conditions are only necessary.
- The full requirement is that for any finite collection $\{t_j\}$ the matrix with i, j'th block $R_X(t_i - t_j)$ must be positive semi-definite.
- If R_X is square-summable, then it is an autocovariance function if and only if its Fourier transform is everywhere positive semi-definite.
- But not every stationary process with finite variance has a square-summable R_X.

19. LINEARLY REGULAR AND LINEARLY DETERMINISTIC PROCESSES

A stationary Gaussian process is **linearly regular** iff $E_t X_{t+s} \to E[X_t]$ as $s \to \infty$. It is **linearly deterministic** iff $E_t X_{t+s} = X_{t+s}$ for all s, t.

20. The fundamental MA representation: the Wold representation

Suppose X_t is a stationary Gaussian vector-valued stochastic process with finite variance. Then $X_t = \tilde{X}_t + \bar{X}_t$, where $\tilde{X}_t$ is linearly regular and $\bar{X}_t$ is linearly deterministic. Furthermore if $X_t - E_{t-1}X_t = \varepsilon_t$, we can write $\tilde{X}_t = \sum_{s=0}^{\infty} A_s \varepsilon_{t-s}$ with $\sum A_s A_s' < \infty$. If $\mathrm{Var}(\varepsilon_t)$ is non-singular, the A_s matrices are uniquely determined. ε_t is referred to as the **innovation** in X_t. A necessary and sufficient condition for a square-summable $\{A_s, s = 0, \dots, \infty\}$ to be the coefficients of a fundamental MA representation of a process is that $\sum_0^\infty A_s z^s = 0 \Rightarrow |z| > 1$ (i.e. "all roots outside the unit circle").

21. Ergodicity

A stationary stochastic process X_t is **ergodic** iff

$$\frac{1}{T}\sum_{1}^{T} f(X_t) \xrightarrow[T\to\infty]{a.s.} E[f(X_t)]$$

for any bounded measurable f.

A Gaussian process is ergodic in the mean (strictly ergodic?) iff

$$\frac{1}{T}\sum_{s=-T}^{T} \frac{T-|s|}{T} R_X(s) \xrightarrow[T\to\infty]{} 0\,.$$

A sufficient condition is obviously $\sum |R_X| < \infty$, but there are other cases.

22. Examples

- An i.i.d. Gaussian process is linearly regular.
- A stationary periodic process (e.g.,

  $$X_t = \begin{cases} \varepsilon_1 & t \text{ odd} \\ \varepsilon_2 & t \text{ even} \end{cases}$$

  $$\varepsilon_1, \varepsilon_2 \sim N(0, I)\,.$$

  is linearly deterministic. ($X_t = X_{t \bmod 2}$)
- The sum of the previous two examples is a stationary process.
- $R_X(t) = 2\sin(\pi t/2)/(\pi t)$ (with $R_X(0) = 1$) defines a linearly deterministic process, though with any finite span of data, only imperfect prediction is possible.
- Usually we think of linearly regular processes as ergodic and linearly deterministic processes as non-ergodic, but these are different concepts. $R_X(t) = (-1)^t$ defines a linearly deterministic, yet ergodic, Gaussian process.

23. The frequency domain

- **Fourier transforms**: If $f : \mathbb{Z} \mapsto^k$ is a vector-valued function of time t, its Fourier transform is
$$\tilde{f}(\omega) = \sum_{-\infty}^{\infty} f(t) e^{-i\omega t} .$$
- Of course this infinite sum might not converge. If f is square-summable, its FT $\tilde{f}$ is well-defined and
$$\frac{1}{2w\pi} \int_{-\pi}^{\pi} e^{i\omega t} \tilde{f}(\omega) d\omega = f(t).$$
So there is a one-one correspondence between square-summable f's and square-integrable $\tilde{f}$'s.
- If f is allowed to be complex valued, so square-summability means that ff' is summable (and f' is the complex conjugate of the transpose of f), then every square-summable f maps to a square-integrable $\tilde{f}$ and vice versa.
- If f takes values in $\mathbb{R}^k$, $\tilde{f}(\omega) = \mathrm{Conj}(\tilde{f}(-\omega)) = \bar{\tilde{f}}(-\omega)$. Every square-integrable function on $[-\pi, \pi]$ that satisfies this condition is the FT of a real-valued square-summable sequence f.
- Convolution — polynomial multiplication — frequency domain multiplication
$$\sum f(s) g(t-s) = f * g(t) \; . \quad \widetilde{f * g} = \tilde{f} \tilde{g}$$
- For LR process X, the **spectral density** $S_X(\omega) = \tilde{R}_X(\omega) = \tilde{A}(\omega) \tilde{A}(\omega)'$.

24. Properties of the spectral density

- Since R_X is real, $S_X(\omega) = S_X(-\omega)'$.
- $S_X(\omega)$ is p.s.d. for all ω.
- For a linearly regular $X(t) = A * \varepsilon(t)$, $R_X(t) = \sum A(s) A'(s-t)$ so $S_X = \tilde{A}\tilde{A}'$. Since A is square-summable , S_X is integrable (but not necessarily square-integrable). Every function S on $[-\pi, \pi]$ such that S is integrable, $S(\omega) = S(-\omega)'$ and $S(\omega)$ is p.s.d. for each $\omega \in [-\pi, \pi]$, is the spectral density of some real-valued, finite-variance stationary process.

25. Frequency domain criteria for linear regularity

- But not every process with an integrable, p.s.d. spectral density is linearly regular.
- The spectral density of a linearly regular process whose innovation process has a non-singular covariance matrix must satisfy in addition:
$$\int_{-\pi}^{\pi} \log |S_x(\omega)| \, d\omega > -\infty .$$

- Note that this does not imply $|S_X(\omega)| > 0$ for all ω. It does imply that there can be no interval of ω values of non-zero length over which $|S_X(\omega)|$ is identically zero.

26. Finite-order AR processes