Introduction to probability theory and stochastic processes

Cameron Ray Smith[1], Aviv Bergman[1,2,3,4]

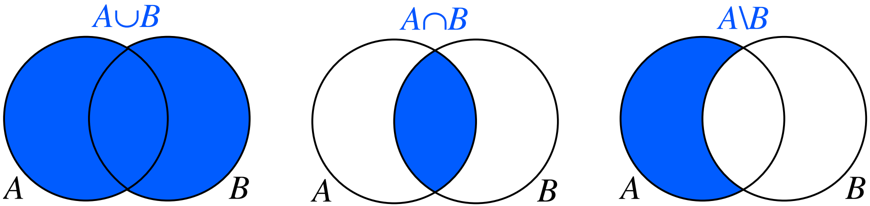
[1]Department of Systems and Computational Biology
[2]Dominick P. Purpura Department of Neuroscience
[3]Department of Pathology
Albert Einstein College of Medicine
[4]Santa Fe Institute

Albert Einstein College of Medicine

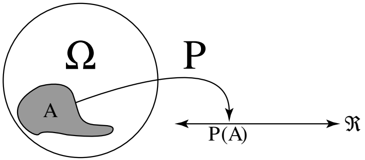
August 17, 2012

Outline

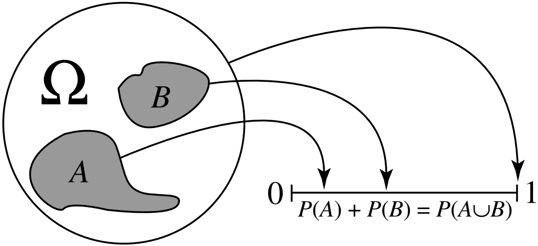
sets



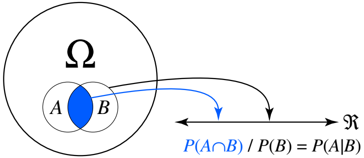
probability measure



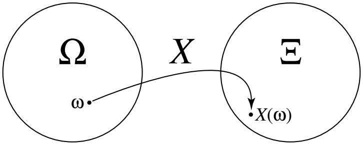
probability axioms



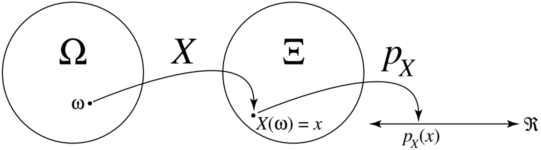
conditional probability



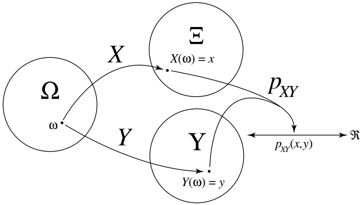
random variables



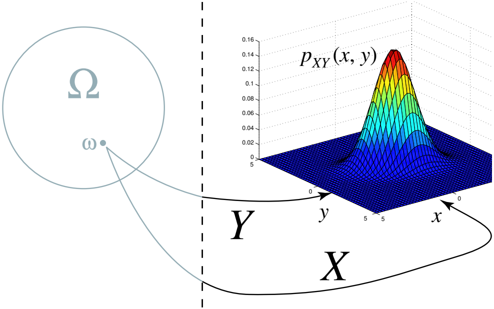
probability densities



joint probability densities



the reality



- For the general theory of measure spaces, we first need a *measurable space* (Ω, Σ), that is a set equipped with a collection Σ of **measurable sets** complete under certain operations. Then this becomes a measure space (Ω, Σ, μ) by throwing in a function μ from Σ to a space of values (such as the real line) that gets along with the set-theoretic operations that Σ has. If E is a measurable set, then $\mu(E)$ is called the **measure** of E with respect to μ. [1, 2]

1. Given a set Ω,
2. a σ-**algebra** is a collection of subsets of Ω that is closed under complementation, countable unions, and countable intersections.
3. A **measurable space**, by the usual modern definition, is a set Ω equipped with a σ-algebra Σ.
4. The elements of Σ are called the **measurable sets** of Ω (or more properly, the measurable subsets of (Ω, Σ)).

A **measure space** is a **measurable space** equipped with a **measure**. There are many different types of measures parametrized by the type of their codomains. Let (Ω, Σ) be a measurable space. A **probability measure** on Ω (due to Kolmogorov) is a function μ from the collection Σ of measurable sets to the unit interval $[0, 1]$ such that:

1. The measure of the empty set is zero: $\mu(\emptyset) = 0$;
2. The measure of the entire space is one: $\mu(\Omega) = 1$;
3. Countable additivity: $\mu(\bigcup_{i=1}^{\infty} S_i) = \sum_{i=1}^{\infty} \mu(S_i)$ whenever the S_i are mutually disjoint sets—disjoint. (Part of the latter condition is the requirement that the sum on the right-hand side must converge.)

It is sometimes stated (but in fact follows from the previous) that:

- Finitary additivity: $\mu(S \cup T) = \mu(S) + \mu(T)$ whenever S and T are disjoint.
- μ is increasing: $\mu(A) \leq \mu(B)$ if $A \subseteq B$.

Measures can be thought of in terms of integrals and densities are defined in terms of measures. Let A be, for example, one of the measurable sets from the collection of measurable sets, Σ, of our sample space Ω.

- $\mu(A) = \int_A dx$ or $\mu(A) = \int_A p(x)dx$
- $\mu(A)$ represents the mass of A which can be interpreted geometrically as an *abstract volume* or probabilistically as *the probability mass of the event "random variable X takes a value within A"*
- A **density** can then be defined intuitively as a function that transforms some measure μ_1 into a measure μ_2 by pointwise reweighting on the sample space Ω. Thus, densities are always relative measures.
- $d\mu_2(x) = f(x)d\mu_1(x)$ or $\frac{d\mu_2}{d\mu_2}(x) = f(x)$

Does a density always exist?

- A density function f is thus a function that is integrated to obtain information in terms of measure μ_2 from information in terms of measure μ_1.
- $\mu_2(A) = \int_A d\mu_2(x) = \int_A f(x) d\mu_1(x)$ is not defined if $\mu_1(A) = 0$ and $\mu_2(A) \neq 0$.
- If this is never the case for all $A \in \Sigma$, then μ_2 is referred to as *absolutely continuous* with respect to μ_1 and this relationship is written $\mu_2 \ll \mu_1$.
- This conclusion is formalized in the **Radon-Nikodym theorem** which states that μ_2 has a density with respect to μ_1 if and only if $\mu_2 \ll \mu_1$.

. . . so the answer is . . . no, which is the reason for going through all this abstract stuff

Outline

Intuitively, **stochastic processes** are ∞-dimensional probability distributions.

- In most applications, stochastic processes model systems that evolve randomly in time, which is likely the origin of the word *process* in the name, but stochastic process are not restricted to the temporal metaphor. Think about *order* as a generalization of time.
- The order of this evolution can be described through the use of an index and an index set respectively $t \in T$.

Definition

Consider a random experiment with sample space Ω, a σ-algebra Σ, a base probability measure $\mu : \Sigma \to [0, 1]$, and a collection of random variables X_t indexed by a set T. A **stochastic process** *is then defined by the set $\{X_t, t \in T\}$.*

- This definition can be specialized to the case of discrete or continuous stochastic processes by taking the index set to be $T \in \mathbb{N}$ or $T \in \mathbb{R}_+$ respectively.

In the discrete case imagine indexing by natural numbers, $\mathbb{N}$, such that the process could be represented as a sequence $\{X_n, n = 0, 1, 2, \ldots\}$.

How should we understand the continuous case?

- For each $\omega \in \Omega$, consider $X_t(\omega) = g_\omega(t)$. $g_\omega(t)$ can then be thought of as a function of t that realizes or samples from the stochastic process.
- For any given t, X_t is a random variable, thus to completely describe the stochastic process we need a description of the joint family of random variables $\{X_t, t \in T\}$ as opposed to just the individual random variables as if they were independent.
- For any discrete subset of times $\{t_1, \ldots, t_n\}$ such that $t_1 < \cdots < t_n$ and associated $\{x_1, \ldots, x_n\}$ we must determine $P(X_{t_1} \leq x_1, \ldots, X_{t_n} \leq x_n)$

How should we understand the continuous case?

- The **Kolmogorov extension theorem** ensures that the potentially infinite distribution $\{X_t, t \in T\}$ where, for example $T \in \mathbb{R}_+$, i.e. $T = [0, \infty)$, can be generated from the finite-dimensional families defined by $P(X_{t_1} \leq x_1, \ldots, X_{t_n} \leq x_n)$ [3]

The necessary conditions are

- Exchangeability: for all permutations π of $1, \ldots, n$ and $x_1, \ldots, x_n$, $P(X_{t_1} \leq x_1, \ldots, X_{t_n} \leq x_n) = P(X_{t_{\pi(1)}} \leq x_{\pi(1)}, \ldots, X_{t_{\pi(n)}} \leq x_{\pi(n)})$
- Extendability: for all $x_1, \ldots, x_n$ and $t_{n+1}, \ldots, t_{n+m}$, $P(X_{t_1} \leq x_1, \ldots, X_{t_n} \leq x_n) = P(X_{t_1} \leq x_1, \ldots, X_{t_n} \leq x_n, X_{t_{n+1}} < \infty, \ldots, X_{t_{n+m}} < \infty)$

How should we understand the continuous case?

- Given that the conditions are satisfied then there exists a probability space $(\Omega, \Sigma, \mathbb{P})$ with an associated stochastic process $X_t : T \times \Omega \to \mathbb{R}^n$ with the families $X_{t_1}, \ldots, X_{t_n}$ as finite-dimensional marginal distributions.

Markov processes

The Chapman-Kolmogorov equation

The Master equation

The Fokker-Planck Equation

Stochastic differential equations

Outline

Population genetics

TODO: topics to cover

Determinism and stochasticity in gene expression

- Elowitz [4]
- Paulsson [5, 6, 7]
- Van Oudenaarden [8, 9]

Outline

[1] David Insua, Fabrizio Ruggeri, and Mike Wiper.
Bayesian Analysis of Stochastic Process Models.
Wiley, 2012.

[2] NLab.
Measure space.
http://ncatlab.org/nlab/show/measure+space.

[3] Bernt Ø ksendal.
Stochastic Differential Equations: An Introduction with Applications.
Springer, 2010.

[4] Peter S Swain, Michael B Elowitz, and Eric D Siggia.
Intrinsic and extrinsic contributions to stochasticity in gene expression.
PNAS, 99(20):12795–800, October 2002.

[5] Johan Paulsson.
Summing up the noise in gene networks.
Nature, 427(6973):415–8, January 2004.

[6] Andreas Hilfinger and Johan Paulsson.
Separating intrinsic from extrinsic fluctuations in dynamic biological systems.
PNAS, 2011(9), July 2011.

[7] Ioannis Lestas, Glenn Vinnicombe, and Johan Paulsson.
Fundamental limits on the suppression of molecular fluctuations.
Nature, 467(7312):174–178, September 2010.

[8] Mukund Thattai and Alexander van Oudenaarden.
Stochastic gene expression in fluctuating environments.
Genetics, 167(1):523–30, May 2004.

[9] B. Munsky, G. Neuert, and A. van Oudenaarden.
Using Gene Expression Noise to Understand Gene Regulation.
Science, 336(6078):183–187, April 2012.