Module overview

Stochastic calculus and process introduces a rigorous mathematical description based upon set theory to describe what is intuitively perceived as randomness.

- Probability theory
- Random variables
- Probability densities
- Conditional expectations
- Stochastic processes
- Brownian motion
- Diffusion
- Karhunen-Loève theorem, relation to singular value decomposition

Definition. *A σ-algebra on a set A is a collection of subsets $\mathcal{S} \subseteq 2^A$ that includes A, and is closed under complement and is closed under countable unions.*

Definition. *A probability space $(\Omega, \mathcal{F}, P)$ consists of:*

a) *a sample space Ω, a set of all possible outcomes of a random trial;*

b) *a σ-algebra $\mathcal{F}$ of measurable subsets of Ω whose elements are events about which it is possible to obtain information*

c) *a probability measure $P: \mathcal{F} \to [0,1]$, with $P(A)$ interpreted as the probability that event $A \in \mathcal{F}$ occurs. For $P(A) = 1$, it is said that event A occurs almost surely.*

Example. The open subsets of $\mathbb{R}$ or $[0,1]$ form a σ-algebra called the *Borel algebra* $\mathcal{B}$ on $\mathbb{R}$ or $[0,1]$. With $\Omega = [0,1]$, $\mathcal{F}$ the Borel algebra on $[0,1]$, and P given by (Lebesgue) interval length, $(\Omega, \mathcal{F}, P)$ forms a probability space.

A function $X\colon \Omega \to \mathbb{R}$, with $\mathcal{F}$ a σ-algebra on Ω is said to be $\mathcal{F}$-*measurable*, if $X^{-1}(A) \in \mathcal{F}$ for all $A \in \mathcal{B}$ in $\mathbb{R}$.

A *random variable* on probability space $(\Omega, \mathcal{F}, P)$ is a real-valued, $\mathcal{F}$-measurable function $X\colon \Omega \to \mathbb{R}$. A random variable is a real-valued quantity that can be measured from the outcome of a random trial.

The *expected value* of random variable X is $\mathbb{E}[X]$, also called the *mean value*, $\mu = \mathbb{E}[X]$. It is assumed that X is integrable, hence $\mathbb{E}[X] < \infty$.

The *variance* is the expected value of the squared deviation from the mean, $\sigma^2 = \mathbb{E}[(X-\mu)^2]$, with σ the *standard deviation*.

The covariance of two random variables X_1, X_2 with means μ_1, μ_2, and standard deviations σ_1, σ_2 is $\operatorname{cov}(X_1, X_2) = \mathbb{E}[(X_1 - \mu_1)(X_2 - \mu_2)]$. The correlation is

$$\operatorname{cor}(X_1, X_2) = \frac{\operatorname{cov}(X_1, X_2)}{\sigma_1 \sigma_2}$$

The expectation operator is a linear functional

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y], a, b \in \mathbb{R}.$$

The expectation operator can be expressed as an integral over probability measure

$$\mathbb{E}[X] = \int_{\Omega} X(\omega)\,\mathrm{d}P(\omega)$$

The probability of an event A is the expectation of its indicator function 1_A

$$P(A) = \mathbb{E}[1_A], 1_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases}$$

Let $P, Q: \mathcal{F} \to [0,1]$ be measures on the same σ-algebra $\mathcal{F}$, and sample space Ω.

Q is *absolutely continuous* w.r.t. P if there exists an integrable random variable $f: \Omega \to \mathbb{R}$, called the *density* of Q w.r.t. P, such that $\forall A \in \mathcal{F}$

$$Q(A) = \int_A f(\omega) \mathrm{d}P(\omega), \mathrm{or}, \mathrm{d}Q = f\,\mathrm{d}P.$$

The expectations w.r.t. P, Q are related by

$$\mathbb{E}^Q[X] = \int_\Omega X\,\mathrm{d}Q(\omega) = \int_\Omega f \cdot X \mathrm{d}P(\omega) = \mathbb{E}^P[fX].$$

The probability measures P, Q are *singular* if $\exists A \in \mathcal{F}$ such that $P(A) = 1$, $Q(A) = 0$.

The distribution function $F\colon \mathbb{R} \to [0,1]$ of a random variable $X\colon \Omega \to \mathbb{R}$ is defined as $F(x) = P\{X \leqslant x\}$.

A random variable is *continuous* if its distribution function is absolutely continuous w.r.t. Lebesgue measure.

If a random variable is continuous with distribution function F, then F is differentiable and $p(x) = F'(x)$ is the probability density function of X, and

$$P(X \in A) = \int_A p(x)\,\mathrm{d}x$$

For any $f\colon \mathbb{R} \to \mathbb{R}$, Borel-measurable

$$\mathbb{E}[f(X)] = \int_{-\infty}^{\infty} f(x)\,p(x)\,\mathrm{d}x$$

Joint probability densities

$X_1, \ldots, X_n \colon \Omega \to \mathbb{R}$ are jointly continuous if there exists a joint probability density function $p(x_1, \ldots, x_n)$ such that

$$P(X_1 \in A_1, \ldots, X_n \in A_n) = \int_A p(x_1, \ldots, x_n)\, \mathrm{d}x_1 \ldots \mathrm{d}x_n$$

with $A = A_1 \times \cdots \times A_n$.

Gaussian probability densities:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-(x-\mu) \frac{1}{2\sigma^2} (x-\mu) \right]$$

$$p(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2} (\det C)^{1/2}} \exp[-(\boldsymbol{x}-\boldsymbol{\mu})^T C^{-1} (\boldsymbol{x}-\boldsymbol{\mu})]$$

Independent random variables

$X_1, \dots, X_n\colon \Omega \to \mathbb{R}$ are *independent* if

$$P\{X_1 \in A_1, \dots, X_n \in A_n\} = P\{X_1 \in A_1\} \cdot \dots \cdot P\{X_n \in A_n\}$$

For any $f_1, \dots, f_n\colon \mathbb{R} \to \mathbb{R}$: $\mathbb{E}[f_1(X_1) \cdot \dots \cdot f_n(X_n)] = \mathbb{E}[f(X_1)] \cdot \dots \cdot \mathbb{E}[f_n(X_n)]$

For jointly continuous random variables the probability density can be factorized

$$p(x_1, \dots, x_n) = p_1(x_1) \cdot \dots \cdot p_n(x_n)$$

If $p_1 = \dots = p_n$, $X_1, \dots, X_n$ are *independent, identically distributed random variables*, (*iid random variables*).

Each random variable defines a different coordinate axis in probability space

Gaussian random variables are independent if the covariance matrix is diagonal.

Linear transformations of Gaussian random variables are Gaussian.

Often, there is interest in the expectation of a random variable in probability space $(\Omega, \mathcal{F} \subseteq 2^{\Omega}, P)$ after certain prior events have taken place. This is translated into set-theoretic formulations of probability theory through subsets of σ-algebras. Recall that σ-algebras on A are subsets of 2^A that allow a measurement operation: one can unambigously assign a number that reflects the size of each member of the σ-algebra.

With $\mathcal{E} \subseteq 2^A$, denote by $\sigma(\mathcal{E})$ the smallest σ-algebra that contains $\mathcal{E}$.

Consider an infinite sequence of coin flips. The probability space is $\Omega = \{0,1\}^{\infty}$, or $\Omega = \{(x_1, x_2, \ldots), x_i \in \{0,1\}\}$. Suppose that after n flips, a change is desired in the betting strategy for successive flips. The relevant σ-algebra is

$$\mathcal{F}_n = \{A \times \{0,1\}^{\infty}, A \in \{0,1\}^n\} \cup \varnothing \subseteq 2^{\Omega}, \mathcal{F}_n \text{ coarser than } \mathcal{F}_{n+1},$$

which is the collection of subsets of Ω decided by the first n flips. and the inclusions $\mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_{\infty}$ hold, with $\mathcal{F}_{\infty}$ the smallest σ-algebra that contains all the others, and $\mathcal{F}_{n+1} = \sigma(\mathcal{F}_n)$. $(\Omega, \mathcal{F}_0 = \bigcup_{i \in \mathbb{N}} \mathcal{F}_n, P_0 = |A^{(n)}| / 2^n)$ is a probability space.

$X: \Omega \to \mathbb{R}$ a random variable in probability space $(\Omega, \mathcal{F}, P)$. Consider another σ-algebra, $\mathcal{G} \subset \mathcal{F}$. The *conditional expectation* of X given $\mathcal{G}$ (i.e., some event in $\mathcal{G}$ has occured), denoted as $Y = \mathbb{E}[X \,|\, \mathcal{G}]$, is itself a $\mathcal{G}$-measurable random variable $Y: \Omega \to \mathbb{R}$, such that for all $\mathcal{G}$-measurable random variables Z

$$\mathbb{E}[\mathbb{E}[X \,|\, \mathcal{G}]\, Z] = \mathbb{E}[YZ] = \mathbb{E}[XZ].$$

In particular for $\forall B \in \mathcal{G}$, choose $Z = 1_B$, and the following must hold $\mathbb{E}[\mathbb{E}[X \,|\, \mathcal{G}]\, 1_B] = \int_\Omega \mathbb{E}[X \,|\, \mathcal{G}]\, 1_B \,\mathrm{d}P = \int_B Y \,\mathrm{d}P = \int_\Omega X\, 1_B \,\mathrm{d}P = \int_B X \,\mathrm{d}P$

In essence, Y corresponds to the average of X over the finer σ-algebra $\mathcal{F}$ to obtain a random variable measurable w.r.t. coarser σ-algebra $\mathcal{G}$, as exemplified by:

Consider $\mathcal{G} = \{\varnothing, B, B^c, \Omega\}$: $\mathbb{E}[X \,|\, \mathcal{G}] = p 1_B + q 1_{B^c}$, with $p = \int_B X \,\mathrm{d}P / P(B)$, $q = \int_{B^c} X \,\mathrm{d}P / P(B^c)$.

Consider random variables $Y_1, \ldots, Y_n$. The *conditional expectation* of X given $Y_1, \ldots, Y_n$ is the random variable Z defined by

$$Z = \mathbb{E}[X \,|\, Y_1, \ldots, Y_n] = \mathbb{E}[X \,|\, \sigma(Y_1, \ldots, Y_n)].$$

For continuous random variables $X_1, \ldots, X_m, Y_1, \ldots, Y_n$ with a joint probability density $p(x_1, \ldots, x_m, y_1, \ldots, y_n)$, the following holds

$$p(x_1, \ldots, x_m | y_1, \ldots, y_n) = \frac{p(x_1, \ldots, x_m, y_1, \ldots, y_n)}{p_Y(y_1, \ldots, y_n)},$$

with the *marginal density* p_Y defined as

$$p_Y(y_1, \ldots, y_n) = \int_{\mathbb{R}^m} p(x_1, \ldots, x_m, y_1, \ldots, y_n) \, \mathrm{d}x_1 \ldots \mathrm{d}x_m$$

The main motivation for introducing previous concepts is to set the framework for discussing *stochastic processes*, i.e., collections of random variables. In particular, the interest here is in continuous-time random processes $X\colon [0,\infty)\times\Omega \to \mathbb{R}$.

- Fix $\omega\in\Omega$, $X^{\omega}\colon[0,\infty)\to\mathbb{R}$, $X^{\omega}\colon t\to X(t,\omega)$ is a *sample function*
- Fix $t\in[0,\infty)$, $X_t\colon\Omega\to\mathbb{R}$, $X_t\colon\omega\to X(t,\omega)$ is a collection of random variables indexed by t

For each t, let $p(x,t)$ be the PDF of $X_t(\omega)$. However, such one-point PDFs *do not* characterize relationships at different times in a stochastic process.

Example. $X_t=1$ or $X_t=-1$ at all times t with probability $1/2$, is different from Y_t in which Y_s, Y_t are iid for $s\neq t$, and $Y_t=\pm 1$ equiprobably

$$\mathbb{E}[X_t]=0, \mathbb{E}[Y_t]=0, \mathbb{E}[X_sX_t]=1, \mathbb{E}[Y_sY_t]=\delta_{s,t}$$

Joint probability densities for stochastic processes

Define relationships between different times in a stochastic process by considering n times, $0 \leqslant t_1 < t_2 < \cdots < < t_n$, and $A_1, A_2, \ldots, A_n$, Borel subsets of $\mathbb{R}$. Event E

$$E = \{\omega \in \Omega : X_{t_j}(\omega) \in A_j, 1 \leqslant j \leqslant n, j \in \mathbb{N}\}$$

has probability

$$P(E) = \int_A p(x_n, t_n; \ldots; x_1, t_1)\, \mathrm{d}x_1 \ldots \mathrm{d}x_n, A = A_1 \times \cdots \times A_n \subset \mathbb{R}^n,$$

$p(x_n, t_n; \ldots; x_1, t_1)$ is the joint probability density for random variables $X_{t_1}, \ldots, X_{t_n}$.

Consistency condition

$$\int_{\mathbb{R}} p(x_n, t_n; \ldots; x_1, t_1)\, \mathrm{d}x_i = p(x_n, t_n; \ldots; x_{i+1}, t_{i+1}; x_{i-1}, t_{i-1}; \ldots; x_1, t_1)$$

A stochastic process described by PDFs at countably many times is *separable*.

Consider $X\colon [0,\infty) \times \Omega \to \mathbb{R}$ a stochastic process on $(\Omega, \mathcal{F}, P)$. Inspired by the infinite coin-flip example, consider a coarser σ-algebra

$$\mathcal{F}_t = \sigma(X_s\colon 0 \leqslant s \leqslant t)$$

For $0 \leqslant s < t$, $\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}$. The family of all such σ-algebras $\{\mathcal{F}_t\colon 0 \leqslant t < \infty\}$ is a *filtration* of $\mathcal{F}$. $\mathcal{F}_t$ is the collection of events observed up to time t, and $\mathbb{E}[X|\mathcal{F}_t]$ is the expectation of X based on observations up to time t.

A stochastic process is said to be a *Markov process* if $\forall 0 \leqslant s < t$ and any Borel-measurable function $f\colon \mathbb{R} \to \mathbb{R}$ with finite $\mathbb{E}[f(X_t)]$ we have

$$\mathbb{E}[f(X_t)|\mathcal{F}_s] = \mathbb{E}[f(X_t)|X_s].$$

Read this as stating that the expectation over all possible histories is the same as that for the history that actually occured, hence the process has no memory of how it arrived at time t.

Characterize a Markov process by its finite-dimensional PDFs at times $0 \leqslant t_1 < \cdots < t_m < t_{m+1} < \cdots < t_n$. Introduce notation $r_i = (x_i, t_i)$The conditional PDF that $X_i = x_i$ for $m+1 \leqslant i \leqslant n$ given that $X_i = x_i$ for $1 \leqslant i \leqslant m$ is

$$p(r_n, \ldots, r_{m+1} | r_m, \ldots, r_1) = \frac{p(r_n, \ldots, r_{m+1}, r_m, \ldots, r_1)}{p(r_m, \ldots, r_1)}.$$

For a Markov process $p(r_{n+1} | r_n, \ldots, r_1) = p(r_{n+1} | r_n)$, hence

$$p(r_n, r_{n-1}, \ldots, r_2 | r_1) = p(r_n | r_{n-1}) \cdot \ldots \cdot p(r_2 | r_1),$$

and satisfy the *Chapman-Kolmogorov relation*

$$p(x, t | y, s) = \int_{\mathbb{R}} p(x, t | z, r) p(z, r | t, s) \, \mathrm{d}z, \forall s < r < t,$$

meaning the process *must* pass through some z at time r going from y, s to x, t.

Brownian motion (Wiener process)

$B(t, \omega)$ is a *Brownian motion* or *Wiener process* if:

1. $B(0, \omega) = 0$, $\forall \omega \in \Omega$
2. $\forall\, 0 \leqslant t_1 < t_2 < \cdots < t_n$, the increments $B_{t_2} - B_{t_1}, \ldots, B_{t_n} - B_{t_{n-1}}$ are independent random variables
3. $\forall\, 0 \leqslant s < t < \infty$, $B_t - B_s$ is a Gaussian random variable with mean 0 and vraiance $t - s$
4. Sample paths $B^\omega \colon [0, \infty) \to \mathbb{R}$ are continuous functions $\forall \omega \in \mathbb{R}$

With $A_0, A_1, \ldots, A_n, \ldots$ iid Gaussian random variables, $A_i \sim \mathcal{N}(0, 1)$

$$B(t) = \frac{1}{\sqrt{\pi}} \left(A_0 t + 2 \sum_{k=1}^{\infty} A_k \frac{\sin(kt)}{k} \right)$$

The transition density is $p(x, t | y, 0) = \mathcal{N}(y, \sqrt{t})$, satisfying $p_t = \frac{1}{2} p_{xx}$.

The Wiener expression of a Brownian motion as a Fourier series,

$$B(t) = \frac{1}{\sqrt{\pi}}\left(A_0 t + 2\sum_{k=1}^{\infty} A_k \frac{\sin(kt)}{k}\right)$$

can be generalized to an arbitrary stochastic process of zero mean, $\mathbb{E}[X_t] = 0$, through the Karhunen-Loève theorem

$$X_t = \sum_{k=1}^{\infty} Z_k u_k(t), Z_k = \int X_t u_k(t)\mathrm{d}t, \mathbb{E}[Z_k]0, \mathbb{E}[Z_i Z_j] = \delta_{ij}\lambda_j$$

The singular value decomposition carries out a discrete version,

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}(\omega_1) & \dots & \mathbf{x}(\omega_N) \end{pmatrix} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T, \mathbf{x} = \begin{pmatrix} x(t_1) & \dots & x(t_n) \end{pmatrix}^T$$

with $\mathbf{\Sigma} = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$, $\mathbf{U} = \begin{pmatrix} \mathbf{u}_1 & \dots & \mathbf{u}_n \end{pmatrix}$, $\mathbf{u}_k = \begin{pmatrix} u_k(t_1) & \dots & u_k(t_n) \end{pmatrix}^T$.