MA4546: Introduction to Stochastic Process

Xiang Zhou

Department of Mathematics

Grading Policy

- closed-book final exam: 70%;
- coursework: 30%
  - 10% : one midterm test
  - 10% : average of two best marks from several in-class quizzes.
  - 10% : take-home assignments (`Homework` . Marked with $\star$ is optional)

penalty for late submission of homework: (usually two weeks are given for each assignment and the solution are released online within 1-2 days after submission deadline.)

- before the release of answers: 10% subtraction from you original score.
- after the release of answers: 50% subtraction from your original score.

No makeup for midterm test or quiz. For the justified and approved medical reason for the absence of the midterm test, 10% part may be added to 70% for the final test.

TEXTBOOK:
`Introduction to modeling and analysis of stochastic systems` [CityU Library holds electronic resource]
author : Kulkarni, Vidyadhar G. New York : Springer Science+Business Media, LLC, 2011. 2nd ed.

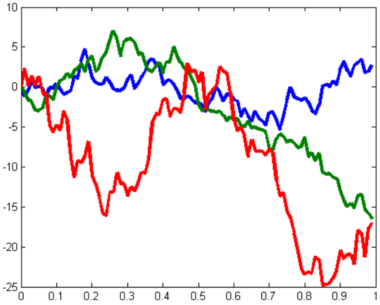
reference book (optional) `Understanding Markov Chains: Examples and Applications` author : Privault, Nicolas Springer Undergraduate Mathematics Series. 2013 (ebook link: `http://link.springer.com/book/10.1007/978-981-4451-51-2`)

Our Plan

- Chapter 1: Review of Probability (1w)
- Chapter 2: Discrete-Time Markov Models (4~5 w)
- Chapter 3: Poisson Processes (2.5w)
- Chapter 4*: Continuous-Time Markov Models (2.5w)
- Chapter 7*: Brownian Motion (2w)

  *: partial coverage of the textbook

Stochastic process is time-dependent description (process) of random phenomena.



香港城市大學
City University of Hong Kong

Deterministic vs Stochastic, which one do you like ?

- Terminology : *stochastic, random, probability, chance, uncertainty, unpredictable*
- flip coins: HTTTHTHHTHHTHHT ...
- gambling: hong kong horse racing, max six, ...
- thermal fluctuation (statistical physics): temperature = randomness in a large population of atoms = entropy = complexity of the microscopic world
- incomplete information: weather predication; financial market; risk analysis
- application: financial engineering, statistical physics, weather forecast, risk analysis, statistics, geology, data sciences,

What is "probability" ?

Chapter 1, (part i) Review of Probability Theory

Probability space: Kolmogorov's axioms

A probability model is a triplet $(\Omega, \mathscr{F}, \Pr)$ (*Andrey Kolmogorov 1930s*):

- sample space: Ω.
- set of events of interest: $\mathscr{F} \subset 2^{\Omega}$ (all subsets of Ω).
- probability (measure) of these events: $\Pr : \mathscr{F} \to [0,1]$, which satisfies
  - $\Pr(\Omega) = 1, \Pr(\emptyset) = 0$
  - **countable additivity**: For any countable collection $\{A_i\}, i = 1,2,3,\cdots$, of pairwise disjoint sets ($A_i \cap A_j = \emptyset$ if $i \neq j$):

$$\Pr(\bigcup_i A_i) = \sum_i \Pr(A_i)$$

$\mathscr{F}$, a set of some subsets of Ω on which $\Pr$ is defines, satisfies

- $\Omega \in \mathscr{F}$
- closed under complements: if $A \in \mathscr{F}$,then also $(A^c) \in \mathscr{F}$
- closed uner **countable unions:** if $\{A_i\} \in \mathscr{F}$, then also $(\bigcup_i A_i) \in \mathscr{F}$.

The collection of sets satisfying the above conditions is called a σ-algebra.

Basics of Probability Theory

- **De Morgan's laws**

$$(\cup_i A_i)^c = \cap_i (A_i^c), \qquad (\cap_i A_i)^c = \cup_i (A_i^c)$$

- **Inclusion-Exclusion Principle**

$$\Pr\left(\cup_{i=1}^n A_i\right) = \sum_{i=1}^n \Pr(A_i) - \sum_{i<j} P(A_i A_j) + \sum_{i<j<k} P(A_i A_j A_k)$$
$$\cdots (-1)^{n+1} P\left(\cap_{i=1}^n A_i\right)$$

example:

$n=2$, $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(AB)$

$$\begin{aligned} n=3, \quad \Pr(A\cup B\cup C) &= \Pr(A) + \Pr(B\cup C) - \Pr(A(B\cup C)) \\ &= \Pr(A) + \Pr(B\cup C) - \Pr((AB)\cup(AC)) \\ &= \Pr(A) + \Pr(B) + \Pr(C) - \Pr(BC) - \Big(\Pr(AB) + \Pr(AC) - \Pr(ABC)\Big) \\ &= \Pr(A) + \Pr(B) + \Pr(C) - \Pr(BC) - \Pr(AB) - \Pr(AC) + \Pr(ABC). \end{aligned}$$

- Given a subset $A \subset \Omega$, define an **indicator function** $\Omega \to \{0,1\}$

$$1_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A \\ 0, & \text{if } \omega \notin A \end{cases}$$

Random variable and its distribution

- A random variable (r.v.) X is a function $X : \Omega \to \mathbb{R} = (-\infty, \infty)$ *
- notation convention : capitalized letters for r.v.s; low case letters for specific numerical values.
- for each set $A \in \mathscr{B}$

$$\Pr(\{X \in A\}) \triangleq \Pr(\{\omega \in \Omega : X(\omega) \in A\}) = \Pr(X^{-1}(A))$$

- This induces a new probability measure $\mathbb{F}_X(A) \triangleq \Pr(X \in A),\ \forall A \in \mathscr{B}$ on $(\mathbb{R}, \mathscr{B})$, which is the *distribution (law)* of the r.v. X. We have a new probability space $(\mathbb{R}, \mathscr{B}, \mathbb{F}_X)$.
- In particular, for the representative subset $A = (-\infty, x]$, we can define Cumulative Distribution Function ("CDF") $F(x) = \mathbb{F}((-\infty, x)) = \Pr(X \leq x)$, which is also sometimes written as F_X to show the underlying r.v. is X.
- If a Probability Density Function ("PDF") exists, then†

$$F(x) = \int_{-\infty}^{x} p(x')dx', \qquad p(x) = F'(x).$$

---

*strictly speaking, X is a "measurable" function from the triplet $(\Omega, \mathscr{F}, \Pr)$ to $((-\infty, \infty), \mathscr{B})$ such that $X^{-1}(E) \triangleq \{\omega \in \Omega : X(\omega) \in E\} \in \mathscr{F}$ for all $E \in \mathscr{B}$, where the Borel set $\mathscr{B}$ is the smallest σ-algebra including all open intervals of $(-\infty, \infty)$.

†strictly speaking, this integration is Lebesgue integration.

Functions of Random Variable

Let X be a random variable and g be a function $\mathbb{R} \to \mathbb{R}$. Then $Y(\omega) = g(X(\omega))$ is another random variable. What is the law of Y ?

$$\mathbb{F}_Y = \mathbb{F}_X \circ g^{-1}$$

because the law of Y is calculated as follows: for any (measurable) event $E \in \mathscr{F}$,

$$\begin{aligned}\mathbb{F}_Y(E) &\triangleq \Pr(Y \in E) = \Pr(\{g(X(\omega)) \in E\}) \\ &= \Pr(X(\omega) \in g^{-1}(E)) = \mathbb{F}_X(g^{-1}(E))\end{aligned}$$

where g^{-1} is the inverse of g defined as

$$g^{-1}(A) \triangleq \left\{x \in \mathbb{R} : g(x) \in A\right\}, \quad \forall A \subset \mathbb{R}.$$

Example $Y = X^2$. Then CDF of Y is
$F_Y(y) = \Pr(Y \le y) = \Pr(X^2 \le y) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$ for all $y \ge 0$.

Generalized inverse of CDF and its application

- If F is the CDF of a r.v. X, the **quantile function or generalized inverse** of F is defined as
$$F^{-}(v) \triangleq \inf\{x : F(x) \geq v\}.$$
$F^{-}(0.5)$ is called the *median*.
- Let F be a monotonically nondecreasing function taking value $[0,1]$. Define it generalized inverse
$$F^{-}(v) = \inf\{x : F(x) \geq v\}$$
Show that if a random variable X has a uniform distribution $U[0,1]$, then the CDF of the random variable $Y \triangleq F^{-}(X)$ is exactly the function F.
- *proof:* $\Pr(Y \leq y) = \Pr(F^{-}(X) \leq y) = \Pr(X \leq F(y)) = F(y)$.
- This fact is used to generate some random variable using the inverse transform sampling-method.

example The exponential distribution, Exp(λ), has CDF $F(x) = 1 - e^{-\lambda x}$ for all $x \geq 0$. So, $F^{-}(v) = -\frac{1}{\lambda}\log(1-v)$ for all $v \in (0,1)$. Thus the formula $-\frac{1}{\lambda}\log(1-X)$ will generate Exp(λ) random variable, by generating a uniform distribution random variable $X \sim U(0,1)$ first.

Moment-generating function

- The moment-generating function of a r.v. X is
$$M_X(t) := \mathrm{E}[e^{tX}] = \int e^{tx} p_X(x)dx, \quad t \in \mathbb{R},$$
wherever this expectation exists. By Taylor expansion,
$$e^{tX} = 1 + tX + (tX)^2/2 + (tX)^3/3! + \dots$$
Then
$$M_X(t) := 1 + t\,\mathrm{E}[X] + t^2\,\mathrm{E}[X^2]/2 + t^3\,\mathrm{E}[X^3]/3! + \dots$$
and the m-th moment is
$$\boxed{\mathrm{E}[X^m] = \frac{d^m M_x(t)}{dt^m}|_{t=0}.}$$
- $M_X(-\lambda) = \int e^{-\lambda x} p_X(x)dx$ is just the Laplace transform of the pdf $p_X(x)$.

Characteristic function

- The **characteristic function** of a r.v. X is a complex-valued function defined by ($\mathbf{i}^2 = -1$)

$$\varphi_X(t) := M_X(\mathbf{i}t) = \mathrm{E}\left[e^{\mathbf{i}tX}\right] = \int e^{\mathbf{i}tx} p_X(x)dx, \quad t \in \mathbb{R},$$

  wherever this expectation exists. This is the Fourier transform of the pdf p_X.

- For any two random variables X_1, X_2, X_1, X_2 both have the same probability distribution if and only if $\varphi_{X_1} = \varphi_{X_2}$. The m-th moment is

$$\boxed{\mathrm{E}[X^m] = (-\mathbf{i})^m \frac{d^m \varphi_x(t)}{dt^m}|_{t=0}.}$$

- $X_1, \ldots, X_n$ are independent r.v.s, *if and only if* for any constants $a_1, \ldots, a_n$, the characteristic function of the linear combination of the X_i's is

$$\varphi_{a_1 X_1 + \cdots + a_n X_n}(t) = \varphi_{X_1}(a_1 t) \cdots \varphi_{X_n}(a_n t)$$

Normal distribution

Read Section 7.1 and 7.2 in the TEXTBOOK

Conditional Probability

- The conditional probability of an event A given that an event B has occurred is
$$\Pr(A|B) = \frac{\Pr(AB)}{\Pr(B)}$$
- Fix B, $\Pr(\cdot|B)$ is a *probability measure*.
- Fix A, let $B = \{\omega\}$, then $\Pr(A|\omega)$ is a *random variable* , i.e., a function from Ω to $[0,1]$.

Law of Total Probability and Bayes' rule

Let $E_1, E_2, E_3, \cdots$ be a set of mutually exclusive and exhaustive events, *i.e.*,

$$E_i \cap E_j = \emptyset \quad \text{if} \quad i \neq j, \text{ and } \quad \cup_{i \geq 1} E_i = \Omega.$$

Then, for any event $E \in \mathscr{F}$,

Law of Total Probability

$$\Pr(E) = \sum_{i \geq 1} \Pr(E \cap E_i) = \sum_{i \geq 1} \Pr(E|E_i)\Pr(E_i)$$

Bayes' Rule

$$\Pr(E_i|E) = \frac{\Pr(E|E_i)\Pr(E_i)}{\Pr(E)} = \frac{\Pr(E|E_i)\Pr(E_i)}{\sum_{i \geq 1} \Pr(E \mid E_i)\Pr(E_i)}$$

Independence

- Events A and B are said to be independent of each other if $\Pr(AB) = \Pr(A)\Pr(B)$. If $\Pr(B) > 0$, then $\Pr(A|B) = \Pr(A)$ if A and B are independent.
- Two r.v.s X and Y are said to be independent of each other if $\Pr(X \in A, Y \in B) = \Pr(X \in A)\Pr(Y \in B)$ for all set A and B, i.e., $F_{X,Y}(x,y) = F_X(x) \cdot F_Y(y)$ or $p_{X,Y}(x,y) = p_X(x)p_Y(y)$ if pdfs exist.
- Mutual(joint) Independence
  $\{E_1, E_2, \cdots, E_n\}$ is said to be mutually independent if for any subset $S \subset \{1, 2, \cdots, n\}$,
  $$\Pr(\bigcap_{i \in S} E_i) = \prod_{i \in S} \Pr(E_i).$$
  i.e., *any* selection of events from this collection is independent.

Expectation

- expectation (with respect to r.v. X with pdf $p_X(x)$):

$$\mathrm{E}(g(X)) = \int g(x) p_X(x) dx$$

- When we write the notation E, we implicitly assume the underlying distribution $p_X(x)$ is clear to the reader. It is also sometime to write $\mathrm{E}_X[\cdot]$ to explicitly point out the underlying distribution for r.v. X.
- For indicator function, $\mathrm{E}(1_A(X)) = \Pr(X \in A)$.
- Variance

$$\begin{aligned}\mathrm{var}(g(X)) &= \mathrm{E}(g^2(X)) - (\mathrm{E}\, g(X))^2 \\ &= \int g^2(x) p_X(x) dx - \left(\int g(x) p_X(x) dx\right)^2\end{aligned}$$

Conditional Expectation

- Conditional probability : $\Pr(X \in A | Y \in B) = \frac{\Pr(X \in A, Y \in B)}{\Pr(Y \in B)}$ for two r.v.s X, Y and two events A, B.
- Conditional pdf: pdf of X given $Y = y$ is *

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$$

- Conditioned expectation given $Y = y$

$$\mathrm{E}(X|Y = y) = \int x \, p_{X|Y}(x|y) dx$$

View the above as a function $h(y)$, then

$$\mathrm{E}(X|Y) \triangleq h(Y) \text{ which is a random variable}$$

- For r.v. X and events A,B, $\mathrm{E}[X, A|B]$ means $\mathrm{E}[X \cdot 1_A | B] = \mathrm{E}[X \cdot 1_A \cdot 1_B]/\Pr(B)$, which is equal to $\mathrm{E}[X|A,B]\Pr(A|B) = \mathrm{E}[X \cdot 1_A \cdot 1_B]/\Pr(AB) \times (\Pr(AB)/\Pr(B))$.

---

* $p_{X,Y}(x,y)$ is called the joint pdf of (X,Y) and $p_Y(y) \triangleq \int p_{X,Y}(x,y)dx$ is called the marginal distribution of Y.

Theorem

$$\mathrm{E}(g(Y)|Y) = g(Y)$$

$$\mathrm{E}(Xg(Y)|Y) = g(Y)\,\mathrm{E}(X|Y)$$

double expectation theorem

$$\mathrm{E}(\mathrm{E}(X|Y)) = \mathrm{E}(X)$$

$\mathrm{E}(X|Y=y)$ is a function of y, say $h(y)$. This function h maps any possible value of r.v. Y ("information" of Y, or σ-algebra generated by Y) into a real number. Its expectation (w.r.t to r.v. Y) is

$$\mathrm{E}(\mathrm{E}(X|Y)) = \mathrm{E}(h(Y)) = \int h(y)p_Y(y)dy$$
$$= \int\int x\,p_{X|Y}(x|y)dx p_Y(y)dy = \int\int x\,p_{X,Y}(x,y)dxdy$$
$$= \int xp_X dx = \mathrm{E}(X)$$

Conditional Expectation as an optimal prediction/projection operator (*optional*) *

$\mathrm{E}(X|Y) = h(Y)$: Optimal approximation of X by using a function of $h(Y)$

Theorem

$$\mathrm{E}[|X - h(Y)|^2] = \min_{g \text{ is a function}} \mathrm{E}(|X - g(Y)|^2)$$

where the function $h(y)$ is the conditional expectation $h(y) = \mathrm{E}(X|Y = y)$.

Proof (exercise) . (hint) first show the orthogonality $\mathrm{E}[(X - h(Y))f(Y)] = 0$ for any function f using the double expectation theorem. Then use the triangular equality $(x - g)^2 = (x - h)^2 + (g - h)^2 - 2(x - h)(g - h)$.

This means the conditional expectation is a projection of X onto the linear space of all functions $g(Y)$ in L_2 sense.

*The content marked with " *optional* " in notes means that these parts are at advance level and will not be covered in any test or quiz (but possible in Homework).

application (optional)

Variance decomposition formula

For two r.v.s X and Y,

$$\mathrm{var}(X) = \mathrm{E}(\mathrm{var}(X|Y)) + \mathrm{var}(\mathrm{E}(X|Y))$$

* proof: Since $\mathrm{var}(X|Y) = \mathrm{E}(X^2|Y) - (\mathrm{E}(X|Y))^2$, take expectation on both sides (a function of y) for Y, then from double expectation theorem,

$$\mathrm{E}[\mathrm{var}(X|Y)] = \mathrm{E}[\mathrm{E}(X^2|Y)] - \mathrm{E}[(\mathrm{E}(X|Y))^2] = \mathrm{E}(X^2) - \mathrm{E}[h^2(Y)]$$

where $h(y) = \mathrm{E}(X|y)$. On the other hand,

$$\mathrm{var}(\mathrm{E}(X|Y)) = \mathrm{var}\, h(Y) = \mathrm{E}\, h^2(Y) - (\mathrm{E}\, h(Y))^2$$
$$= \mathrm{E}\, h^2(Y) - (\mathrm{E}[\mathrm{E}(X|Y)])^2 = \mathrm{E}\, h^2(Y) - (\mathrm{E}\, X)^2.$$

---

*If we write explicitly the underling r.v.s for each var and E, then the conclusion is

$$\mathrm{var}_X(X) = \mathrm{E}_Y(\mathrm{var}_X(X|Y)) + \mathrm{var}_Y(\mathrm{E}_X(X|Y)).$$

Random Sums of i.i.d r.v.s

Let $\{X_n : n = 1,2,3,\cdots\}$ be a sequence of iid * random variables with common expectation $\mathrm{E}\,X$ and variance $\mathrm{var}\,X$, and let N be a nonnegative integer-valued random variable that is independent of $\{X_n\}$. Let $Z = \sum_{n=1}^{N} X_n$. Then

$$\boxed{\mathrm{E}(Z) = \mathrm{E}(X)\,\mathrm{E}(N)} \qquad \text{(Wald Identity)}$$

$$\boxed{\mathrm{var}(Z) = \mathrm{E}(N)\,\mathrm{var}(X) + (\mathrm{E}(X))^2\,\mathrm{var}(N).}$$

Proof: Note that from independence between N and $\{X_n\}$,

$$\mathrm{E}(Z|N=k) = \mathrm{E}(\sum_{n=1}^{N} X_n|N=k) = \mathrm{E}(\sum_{n=1}^{k} X_n|N=k) = \mathrm{E}(\sum_{n=1}^{k} X_n) = k\,\mathrm{E}(X)$$

So, the function $h : k \mapsto \mathrm{E}(Z|N=k)$ is a linear function $h(k) = k\cdot\mathrm{E}(X)$. Therefore, the random variable $\mathrm{E}(Z|N) = h(N) = N\cdot\mathrm{E}(X)$. By double expectation theorem, $\mathrm{E}(Z) = \mathrm{E}(\mathrm{E}(Z|N)) = \mathrm{E}(h(N)) = \mathrm{E}(X)\,\mathrm{E}(N)$.

*independent and identically distributed

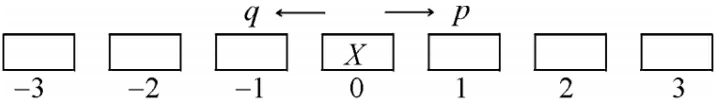
Next, We calculate the mapping g defined as $k \mapsto \mathrm{E}(Z^2|N=k)$

$$
\begin{aligned}
\mathrm{E}(Z^2|N=k) &= \mathrm{E}((\sum_{n=1}^{k} X_n)^2\,|N=k) \\
&= \mathrm{E}((\sum_{n=1}^{k} X_n)^2) = \mathrm{E}(\sum_{n=1}^{k} X_n^2 + \sum_{1\le m,n\le k, m\ne n} X_m X_n) \\
&= k\,\mathrm{E}(X_n^2) + (k^2-k)\,\mathrm{E}(X_n)\,\mathrm{E}(X_m) \\
&= k^2(\mathrm{E}\,X)^2 + k\,\mathrm{var}(X).
\end{aligned}
$$

So, $g(k)$ is the quadratic function. $\mathrm{E}(Z^2|N) = g(N)$.

$$
\begin{aligned}
\mathrm{var}(Z) &= \mathrm{E}(\mathrm{E}(Z^2|N)) - (\mathrm{E}(Z))^2 = \mathrm{E}(g(N)) - (\mathrm{E}\,N\cdot\mathrm{E}\,X)^2 \\
&= \mathrm{E}[N^2(\mathrm{E}\,X)^2 + N\cdot\mathrm{var}\,X] - (\mathrm{E}\,N\cdot\mathrm{E}\,X)^2 \\
&= \mathrm{E}(N^2)\cdot(\mathrm{E}\,X)^2 + \mathrm{E}\,N\cdot\mathrm{var}\,X - (\mathrm{E}\,N\cdot\mathrm{E}\,X)^2 \\
&= \mathrm{var}\,N\cdot(\mathrm{E}\,X)^2 + \mathrm{E}\,N\cdot\mathrm{var}\,X.
\end{aligned}
$$

Chapter 1, (part ii) Random Walk Model



- Define the iid (Bernoulli) random variable $Z_i = \begin{cases} +1, & \text{with prob } p \\ -1, & \text{with prob } q = 1-p. \end{cases}$.
- What is expectation and variance of Z_i ?

$$\mathrm{E}[Z_i] = -1 \times q + 1 \times p = p - q := \mu$$

$$\begin{aligned} \mathrm{var}[Z_i] &= \mathrm{E}\, Z_i^2 - \mu^2 = 1^2 \times q + 1^2 \times p - \mu^2 \\ &= 1 - \mu^2 = (1+\mu)(1-\mu) = 2p * 2q = 4pq \end{aligned}$$

- Let $X_0 = 0$ then the random walk (X_n) is the random sequence

$$\boxed{X_n = \sum_{i=1}^{n} Z_i}$$

The case of $p = q = 1/2$ is called symmetric random walk.

- This is also a gambling model: Each bet is 1 dollar. Win prob is p at each round n. Then X_n is the money at time n.

Distribution of X_n

$$\mathrm{E}\,X_n = \mathrm{E}(\sum_i Z_i) = \sum_i \mathrm{E}(Z_i) = n\mu$$

$$\begin{aligned}
\mathrm{var}(X_n) &= \mathrm{E}(\sum_i Z_i)^2 - (\mathrm{E}\sum_i Z_i)^2 = \mathrm{E}\left[(\sum_i Z_i)(\sum_j Z_j)\right] - (\sum_i \mathrm{E}\,Z_i)(\sum_j \mathrm{E}\,Z_i) \\
&= \mathrm{E}(\sum_i Z_i^2) + \mathrm{E}\left[\sum_{i\neq j} Z_i Z_j\right] - \sum_i (\mathrm{E}\,Z_i)^2 - \sum_{i\neq j}(\mathrm{E}\,Z_i)(\mathrm{E}\,Z_j) \quad \because \{X_i\} \text{ indept.} \\
&= \sum_i \left[\mathrm{E}(Z_i^2) - (\mathrm{E}\,Z_i)^2\right] = \sum_i \mathrm{var}(Z_i) = 4npq
\end{aligned}$$

$$\Pr(X_1 = 1) = p,\ \Pr(X_1 = -1) = q,$$

$$\Pr(X_2 = 2) = p^2,\ \Pr(X_2 = 0) = 2pq,\ \Pr(X_2 = -2) = q^2,$$

$$\Pr(X_3 = 3) = p^3,\ \Pr(X_3 = 1) = 3p^2 q,\ \Pr(X_3 = -1) = 3pq^2,\ \Pr(X_3 = -3) = q^3,$$

$$\cdots$$

Martingale (*optional*)

Definition

A stochastic process $M(t)$, where time t is continuous ($t \in \mathbb{R}$) or discrete ($t = 0, 1, 2, \cdots$), adapted to a filtration $\mathscr{F} = (\mathscr{F}_t : t \in \mathbb{R})$ is a martingale if for any t, $\mathrm{E}|M(t)| < +\infty$ and for any s and t with $0 \le s < t \le T$,

$$\mathrm{E}(M(t)|\mathscr{F}_s) = M(s).$$

- The filtration $\mathscr{F} = (\mathscr{F}_t : t \in \mathbb{R})$ is a stream of information: the known information up to time t is denoted as $\mathscr{F}_t$. So, $\mathscr{F}_s \subset \mathscr{F}_t$, $\forall s \le t$.
- adaptive: $M(t) \in \mathscr{F}_t$, i.e., the set $\{M(t) \le a\} \in \mathscr{F}_t$ for all real number a.
- random walk. $\mu := \mathrm{E}\, Z_1 = p - q = 2p - 1$. Then $(X_n - \mu n)_{n \in \mathbb{Z}}$ is a martingale since

$$\mathrm{E}|X_n - \mu n| \le \mathrm{E}|X_n| + \mu n \le \sum_{i=1}^{n} \mathrm{E}|Z_1| = \mu n < \infty$$

$$\mathrm{E}(X_{n+1} - \mu(n+1) \mid Z_n, \cdots, Z_1) = \mathrm{E}(X_n - \mu n + Z_{n+1} - \mu \mid Z_n, \cdots, Z_1)$$
$$= \mathrm{E}(X_n - \mu n \mid Z_n, \cdots, Z_1) + \mathrm{E}(Z_{n+1} - \mu) = \mathrm{E}(X_n - \mu n \mid Z_n, \cdots, Z_1) = X_n - \mu n.$$

Predictable process (*optional*)

Let's play the gambling with *random* ante: the stake to bet at round n is H_n dollars. Here (H_n) is another stochastic process. Then the money at time n is

$$\boxed{Y_n := \sum_{i=1}^{n} H_i Z_i = \sum_{i=1}^{n} H_i (X_i - X_{i-1}).}$$

The natural requirement is that H_n is known *before* the start of round n, i.e.,

$$H_n \in \mathscr{F}_{n-1} = \text{information of } (Z_1, \cdots, Z_{n-1})^*.$$

Remark: Here we do not need independence of $(H_n : n \in \mathbb{Z})$ and $(Z_n : n \in \mathbb{Z})$. The strategy H_n can depend on $Z_1, \cdots, Z_{n-1}$, but not on Z_n!
Let $p = q$, then we already know that (X_n) is a martingale. Show that (Y_n) is also a martingale w.r.t. $(\mathscr{F}_n)$ if $|H_n| < C$ for any n;

$$\begin{aligned}
\mathrm{E}(Y_{n+1}|\mathscr{F}_n) &= \mathrm{E}(Y_n|\mathscr{F}_n) + \mathrm{E}(H_{n+1} Z_{n+1}|\mathscr{F}_n) \\
(\because \text{ predictable }) &= \mathrm{E}(Y_n|\mathscr{F}_n) + H_{n+1}\, \mathrm{E}(Z_{n+1}|\mathscr{F}_n) \\
&= Y_n + H_{n+1} \cdot 0 = Y_n
\end{aligned}$$

*We call (H_n) is *predictable* w.r.t. $\mathscr{F} = (\mathscr{F}_n)$ if $H_n \in \mathscr{F}_{n-1}$

stochastic integration (*optional*)

stochastic integration (discrete form)

Assume that the stochastic process (X_n) is a martingale w.r.t. the filtration $(\mathscr{F}_n)$ and (H_n) is **predictable** to $(\mathscr{F}_n)$ and $\sup_n |H_n| < \infty$ a.e. Then, define the stochastic integration

$$\boxed{Y_n := \sum_{i=1}^{n} H_i (X_i - X_{i-1})} \quad \sim \int H dX$$

Theorem

(Y_n) is also a martingale w.r.t. $(\mathscr{F}_n)$.

Proof:

$$\begin{aligned} \mathrm{E}(Y_{n+1}|\mathscr{F}_n) &= \mathrm{E}(Y_n|\mathscr{F}_n) + \mathrm{E}(H_{n+1}(X_{n+1} - X_n)|\mathscr{F}_n) \\ (\because \text{ predictable }) &= \mathrm{E}(Y_n|\mathscr{F}_n) + H_{n+1}\,\mathrm{E}((X_{n+1} - X_n)|\mathscr{F}_n) \\ &= Y_n + H_{n+1}(\mathrm{E}(X_{n+1} \mid \mathscr{F}_n) - X_n) \\ &= Y_n + H_{n+1} \cdot 0 = Y_n \end{aligned}$$

Exercises [†]

1. If two r.v.s X and Y are independent, then
$$p_{X|Y}(x|y) = p_X(x), \qquad \mathrm{E}(X|Y) = \mathrm{E}\,X.$$
2. If $\{X_1, X_2, \cdots, X_n\}$ is mutually independent, then $\mathrm{var}(\sum_i X_i) = \sum_i \mathrm{var}(X_i)$
3. Find the moment-generating and characteristic functions for the following distributions: Bernoulli distribution Bern(p), Poisson distribution Poi(λ), exponential distribution Exp(λ), normal distribution $N(\mu, \sigma^2)$. *
4. Suppose that $X = (X_1, X_2)$ is a two dimensional Gaussian random variable with mean $\mu = (\mu_1, \mu_2)$ and the covariance matrix $\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$. What is the conditional pdf $p(x_1|x_2)$ of X_1 given $X_2 = x_2$? For what value of ρ, X_1 and X_2 are independent ?
5. Show that if (X_t) is a martingale, then its expectation $\mathrm{E}\,X_t$ is independent of time t.
6. For the random walk defined above, find the value of a positive number σ such that $(X_n - \mu n)^2 - \sigma^2 n$ is a martingale.

---

*click for online answer

[†] not Homework , no need to submit. But you are encouraged to solve.