Martingales in finite probability space

Lecturer: Dr. Hong-Gwa Yeh
Department of Mathematics
National Central University
hgyeh@math.ncu.edu.tw

Finite Probability Space

Ω = the set of all possible outcomes of a random experiment (Ω is called a sample space, and $\omega \in \Omega$ is called a sample point) note: we only consider $|\Omega| < \infty$.

σ-algebra: $\mathcal{F} \subseteq 2^{\Omega}$ is called a σ-algebra if

① $\Omega \in \mathcal{F}$ ② $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$ ③ $A_1, A_2, \cdots \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

Ex: Toss a coin 3 times. $\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$

Let $\mathcal{F}_0 = \{\phi, \Omega\}$,

$\mathcal{F}_1 = \{\phi, \Omega, A_H, A_T\}$, where $A_H = \{HHH, HHT, HTH, HTT\}$

EX (continued) After the random experiment, you are not told the outcome, but you are told, for every set in $\mathcal{F}_1$ whether or not the outcome is in that set. For example, you would be told that the outcome is not in ϕ, and is in Ω. Moreover, you might be told that outcome is not in A_H, but is in A_T. In effect, You have been told that the first toss was a T.

We interpret the σ-algebra $\mathcal{F}_1$ as a record of the "information of the first toss"

$(\Omega, \mathcal{F}, \mathcal{P})$

Probability measure: let $\mathcal{F}$ be a σ-algebra on Ω.

$\mathcal{P}: \mathcal{F} \to [0,1]$ is called a probability measure on $(\Omega, \mathcal{F})$

① $\mathcal{P}(\Omega) = 1$ ② If $A_1, A_2, \dots$ is a sequence of disjoint sets in $\mathcal{F}$

then $\mathcal{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathcal{P}(A_i)$

probability space: $(\Omega, \mathcal{F}, \mathcal{P})$ is called a probability space.

a σ-algebra on Ω — a probability measure on $(\Omega, \mathcal{F})$

Remark: Here we only consider the case that $|\Omega| < \infty$.

$\alpha(D)$

called block (or atom)

decomposition of Ω: i.e. a collection $\mathcal{D} = \{D_1, D_2, \cdots, D_n\}$ of sets in Ω

(assume $|\Omega| < \infty$)

s.t. ① $D_i \neq \phi \ \forall i$ ② $D_1 + D_2 + \cdots + D_n = \Omega$

Fact: let $\mathcal{F} = \{ S : S \text{ is a union of sets in } \mathcal{D} \} \cup \{\phi\}$.

Then $\mathcal{F}$ is a σ-algebra, and is called the σ-algebra generated by $\mathcal{D}$, and is denoted by $\alpha(\mathcal{D})$.

Fact: Let $\mathcal{B}$ be a σ-algebra of subset of a finite space Ω.

Then $\exists!$ decomposition $\mathcal{D}$ of Ω s.t. $\mathcal{B} = \alpha(\mathcal{D})$

Hint: see Shiryayev p13. Let $\mathcal{D} = \{ D \in \mathcal{B} : D \neq \phi,\ D \cap B = D \text{ or } \phi \text{ for any } B \in \mathcal{B} \}$.

This $\mathcal{D}$ will meet our need.

$\mathcal{G}$-measurable

Note*: $\exists$ a one-to-one correspondence between σ-algebras and decomposition of a finite space Ω.

- Let $\mathcal{D} = \{D_1, \cdots, D_k\}$ be a decomposition of Ω. $|\Omega| < \infty$

  Let $\eta = \eta(\omega)$ be a function on Ω.

Def: η is said to be $\mathcal{D}$-measurable if η has the form

$$\eta(\omega) = \sum_{i=1}^{k} y_i I_{D_i}(\omega)$$

i.e. η takes constant values on the blocks of $\mathcal{D}$.

Remark: In above def. we also said that η is $\mathcal{G}$-measurable where $\mathcal{G}$ is the σ-algebra $\alpha(\mathcal{D})$

Random Variable X, $\mathscr{D}_X$

Def: For a finite probability space $(\Omega, \mathscr{F}, P)$, we say that X is a random variable on it if X is an $\mathscr{F}$-measurable real-valued function defined on Ω.

Def: Let X be a rv having the distinct values $x_1, x_2, \cdots, x_k$ with positive probabilities i.e. $X = \sum_{i=1}^{k} x_i I_{D_i}(\omega)$, where $D_i = \{\omega \in \Omega : X(\omega) = x_i\}$.
We define the decomposition $\mathscr{D}_X \overset{def}{=} \{D_1, D_2, \cdots, D_k\}$.
If $X_1, X_2, \cdots, X_m$ are rvs then decomposition $\mathscr{D}_{X_1, X_2, \cdots, X_m}$ is defined in the same way.

Note: $\alpha(\mathscr{D}_X)$ is the smallest σ-algebra over which X is measurable.

$P(A \mid \mathscr{D})$

- $(\Omega, \mathcal{F}, P)$ is a finite prob. space. $A \in \mathcal{F}$.
- $\mathscr{D} = \{D_1, \dots, D_k\}$ is a decomposition of Ω with $\begin{cases} D_i \in \mathcal{F} & \forall i \\ P(D_i) > 0 & \forall i. \end{cases}$

Def: Define the $P(A \mid \mathscr{D}) : \Omega \longrightarrow \mathbb{R}$ as follows

$$P(A \mid \mathscr{D})(\omega) = \sum_{i=1}^{k} P(A \mid D_i)\, I_{D_i}(\omega)$$

Facts:
(1) $A \cap B = \phi \Rightarrow P(A \cup B \mid \mathscr{D}) = P(A \mid \mathscr{D}) + P(B \mid \mathscr{D})$

(2) $P(A \mid \{\Omega\}) = P(A)$ constant rv

(3) $\mathcal{E}\, P(A \mid \mathscr{D}) = P(A)$

(4) $P(A \mid \mathscr{D})$ is $\mathscr{D}$-measurable and hence is $\mathcal{F}$-measurable
i.e. $P(A \mid \mathscr{D})$ is a rv.

$\mathcal{E}(X|\mathcal{D})$

- $(\Omega, \mathcal{F}, P)$ a finite prob. space
- $X : \Omega \to \{x_1, \cdots, x_n\}$ a rv (distinct)
- $\mathcal{D} = \{D_1, \cdots, D_k\}$ a decomposition of Ω s.t. $D_i \in \mathcal{F}$, $P(D_i) > 0$, $\forall i$.

Def: Def the function $\mathcal{E}(X|\mathcal{D}) : \Omega \to \mathbb{R}$ as follows

$$\mathcal{E}(X|\mathcal{D})(\omega) = \sum_{i=1}^{n} x_i P(X = x_i | \mathcal{D})(\omega)$$

Fact: (2) $\mathcal{E}(X|\mathcal{D})$ is $\mathcal{D}$-measurable and hence $\mathcal{F}$-measurable.

i.e. $\mathcal{E}(X|\mathcal{D})$ is a rv.

(1) $\mathcal{E}(X|\mathcal{D}) = \sum_{j=1}^{k} \left(\sum_{i=1}^{n} x_i P(X = x_i | D_j) \right) I_{D_j}$

$= \sum_{j=1}^{k} \mathcal{E}(X|D_j) I_{D_j}$

the average value of X on the block D_j

(3) $\mathcal{E}\left(\mathcal{E}(X|\mathcal{D}) \right) = \mathcal{E}X$

$$\mathcal{P}(A \mid \mathfrak{I}),\ \mathcal{E}(X \mid \mathfrak{I})$$

Remark: If $\mathfrak{I} = \alpha(\mathcal{D})$ then

a $\sigma-$algebra of a finite space

$\mathcal{P}(A \mid \mathcal{D})$ is also denoted by $\mathcal{P}(A \mid \mathfrak{I})$,

and $\mathcal{E}(A \mid \mathcal{D})$ is also denoted by $\mathcal{E}(A \mid \mathfrak{I})$.

Warning: we only use this notation for finite probability space.

Tower Property

Thm If two σ-algebra $\mathcal{F}, \mathcal{G}$ have $\mathcal{F} \subseteq \mathcal{G}$ then

$$\mathcal{E}(\mathcal{E}(X|\mathcal{G})|\mathcal{F}) = \mathcal{E}(X|\mathcal{F})$$

Fact (1) If X is $\mathcal{D}$-measurable for a decomposition $\mathcal{D}$ of Ω then $\mathcal{E}(X|\mathcal{D}) = X$ and $\mathcal{E}(XY|\mathcal{D}) = X\,\mathcal{E}(Y|\mathcal{D})$

"taking out what is known"

(2) If X is independent of decomposition $\mathcal{D}$ (i.e. $\forall D_i \in \mathcal{D}$, X and I_{D_i} are independent) then $\mathcal{E}(X|\mathcal{D}) = \mathcal{E}X$

pf of (2): Say $\mathcal{D} = \{D_1, \dots, D_k\}$. Then

$$\mathcal{E}(X|\mathcal{D}) = \sum_{i=1}^{k} \mathcal{E}(X|D_i) I_{D_i} = \sum_{i=1}^{k} \mathcal{E}X\, I_{D_i} = \mathcal{E}X.$$

$P(A|X_1, X_2, \ldots, X_M)$

- $(\Omega, \mathcal{F}, P)$ is a finite prob. space. $A \in \mathcal{F}$
- rv $X: \Omega \to \{x_1, \ldots, x_k\}$ with $P(X = x_i) > 0 \ \forall i$

Def: $P(A|X) \overset{\text{def}}{=\!=} P(A|\mathcal{D}_X)$

$P(A|X_1, X_2, \ldots X_m) \overset{\text{def}}{=\!=} P(A|\mathcal{D}_{X_1, X_2, \ldots X_m})$

Facts: (1) $P(A|X)(\omega) = \sum_{i=1}^{k} P(A|X = x_i)\, I_{\{X = x_i\}}(\omega)$

(2) $P(A|X_1, X_2, \cdots, X_m)(\omega)$

$$= \sum_{y_1, \ldots, y_m} P(A|X_1 = x_1, \ldots X_m = y_m)\, I_{\{X_1 = x_1, \ldots X_m = y_m\}}(\omega)$$

$E(X|Y_1, Y_2, \ldots, Y_M)$

Def: In a finite probability space, we define rvs

$$\mathcal{E}(X|Y) \overset{def}{=\!=} \mathcal{E}(X|\mathcal{D}_Y)$$

$$\mathcal{E}(X|Y_1, Y_2, \cdots, Y_m) \overset{def}{=\!=} \mathcal{E}(X|\mathcal{D}_{Y_1, Y_2, \cdots, Y_m})$$

Fact: The rv $\mathcal{E}(X|Y)$ is the rv $f(Y)$ such that $f(y) = \mathcal{E}(X|Y=y)$.

Pf: Let $Y: \Omega \to \{y_1, \cdots, y_K\}$ with $P(Y=y_i) > 0 \ \forall i$.

$$\mathcal{E}(X|Y) = \mathcal{E}(X|\mathcal{D}_Y) = \sum_{j=1}^{K} \mathcal{E}(X|D_j) I_{D_j}, \text{ where } D_j = \{Y = y_j\}$$

Done!

Remark: We can generalize the above fact to the rv $\mathcal{E}(X|Y_1, \cdots, Y_m)$.

Example for $\mathcal{E}(X \mid Y_1, Y_2, \cdots, Y_m)$

Example Consider independent throws of an unbiased 6-side die. For $1 \le i \le 6$, let X_i denote the number of times the value i appears in n throws of the die. Then

$$\mathcal{E}(X_1 \mid X_2) = \frac{n - X_2}{5}, \quad \mathcal{E}(X_1 \mid X_2, X_3) = \frac{n - X_2 - X_3}{4}.$$

Pf: $\mathcal{E}(X_1 + X_2 + \cdots + X_6 \mid X_2 = \alpha, X_3 = \beta) = n$

$$\Longrightarrow 4\,\mathcal{E}(X_1 \mid X_2 = \alpha, X_3 = \beta) = n - \alpha - \beta \Rightarrow \mathcal{E}(X_1 \mid X_2 = \alpha, X_3 = \beta) = \frac{n - \alpha - \beta}{4}.$$

$$\Longrightarrow \mathcal{E}(X_1 \mid X_2, X_3) = \frac{n - X_2 - X_3}{4}.$$

QED

A filter in finite prob. spaces

- Given a finite probability space $(\Omega, 2^{\Omega}, P)$.

Def: A filter is a nested sequence of σ-algebras in Ω s.t. $\{\phi, \Omega\} = \mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_n = 2^{\Omega}$

Remark: In above definition, if we have decompositions $\mathcal{D}_0 \cdots \mathcal{D}_n$ s.t. $\mathcal{D}_0 = \{\Omega\}$, $\alpha(\mathcal{D}_i) = \mathcal{F}_i$, $\forall i$, and $\mathcal{D}_n = \{\{\omega\} : \omega \in \Omega\}$

Then sometimes we write the above filter as

$$\{\Omega\} = \mathcal{D}_0 \preceq \mathcal{D}_1 \preceq \cdots \preceq \mathcal{D}_n = \{\{\omega\} : \omega \in \Omega\}$$

Martingale (I) general setting

Def: Given a finite p.s. $(\Omega, 2^\Omega, P)$ with a filter $\{\phi, \Omega\} = \mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_n = 2^\Omega$, a sequence of rvs $X_0, X_1, \cdots, X_n$ is called a martingale w.r.t. the filter $\mathcal{F}_0, \mathcal{F}_1, \cdots, \mathcal{F}_n$ if $\mathcal{E}(X_{k+1} | \mathcal{F}_k) = X_k$ for each $k = 0, 1, 2, \cdots, n-1$.

Notation: Sometimes we use $(X_k, \mathcal{F}_k)_{k=0}^{n}$ to denoted the above martingale.

Fact: (1) X_k is $\mathcal{F}_k$-measurable, $k = 0, 1, 2, \ldots, n$

(2) $\mathcal{E}X_k = \mathcal{E}X_0$ for each $k = 1, 2, \cdots, n$.

Martingale (II) special case

Def: A sequence of rvs $X_0, X_1, \cdots, X_m$ is called a martingale if for $0 \le i \le m$, $\mathcal{E}(X_{i+1} \mid X_0, X_1, \cdots, X_i) = X_i$.

Recall: $\mathcal{E}(X_{i+1} \mid X_0, X_1, \cdots, X_i) \stackrel{\text{def}}{=} \mathcal{E}(X_{i+1} \mid \mathcal{D}_{X_0, X_1, \cdots, X_i})$

Note: In the above definition, $X_0, X_1, \cdots X_m$ is a martingale w.r.t. the filter $\alpha(\mathcal{D}_{X_0}), \alpha(\mathcal{D}_{X_0, X_1}), \alpha(\mathcal{D}_{X_0, X_1, X_2}), \cdots, \alpha(\mathcal{D}_{X_0, X_1, \cdots, X_{m-1}}), 2^{\Omega}$ in the general setting.

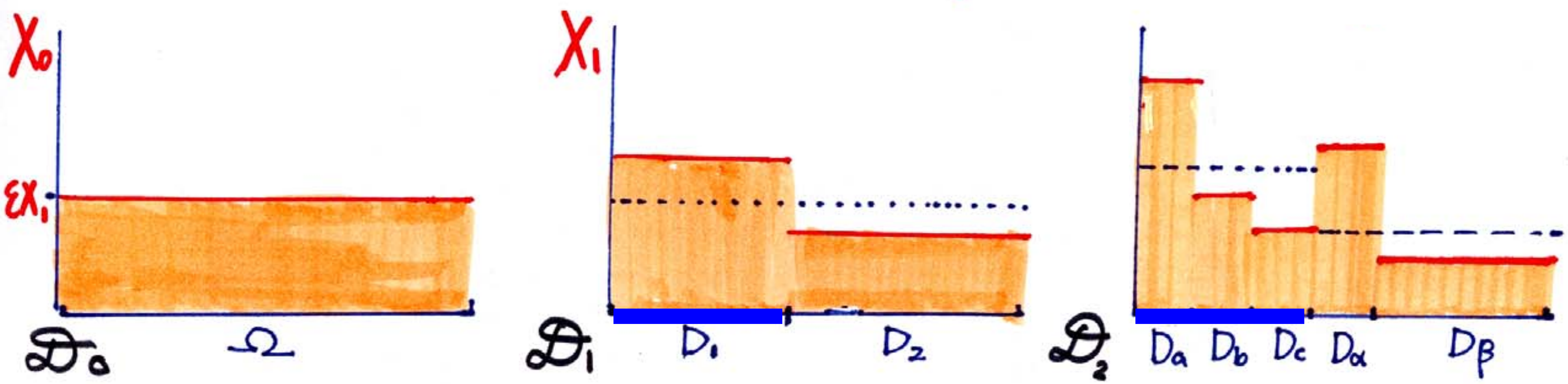
Recall

Suppose $\Im_k = \alpha(\ \left\{ D_{k_1}, D_{k2}, \cdots, D_{k_t} \right\}\ ).$

Then $E(X_{k+1} \mid \Im_k) = X_k$ implies

$$X_k = \sum_{i=1}^{t} E(X_{k+1} \mid D_{k_i}) I_{D_{k_i}}$$

An illustration of a martingale



- We assume that $\alpha(\mathcal{D}_0) = \mathfrak{I}_0$. $\alpha(\mathcal{D}_1) = \mathfrak{I}_1$, $\alpha(\mathcal{D}_2) = \mathfrak{I}_2$.
- The values of X_i are indicated by the red lines.

$X_0 = \mathcal{E}X_1 I_\Omega$

Note $\mathfrak{I}_0 \subseteq \mathfrak{I}_1 \subseteq \mathfrak{I}_2$.

$X_1 = \mathcal{E}(X_2 | D_1) I_{D_1} + \mathcal{E}(X_2 | D_2) I_{D_2}$

$X_2 = \mathcal{E}(X_3 | D_a) I_{D_a} + \mathcal{E}(X_3 | D_b) I_{D_b} + \mathcal{E}(X_3 | D_c) I_{D_c}$

$+ \mathcal{E}(X_3 | D_\alpha) I_{D_\alpha} + \mathcal{E}(X_3 | D_\beta) I_{D_\beta}$

Doob Martingales

- Let $(\Omega, 2^{\Omega}, \mathcal{P})$ be a finite p.s. with a filter $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_n$

Thm Let X be a rv on $(\Omega, 2^{\Omega}, \mathcal{P})$. Define $X_i = \mathcal{E}(X \mid \mathcal{F}_i)$ for $i = 0, 1, 2, \dots, n$. Then $(X_i, \mathcal{F}_i)_{i=0}^{n}$ is a martingale.

Pf: For $k = 0, 1, 2, \dots, n-1$, $\mathcal{E}(X_{k+1} \mid \mathcal{F}_k) = \mathcal{E}(\mathcal{E}(X \mid \mathcal{F}_{k+1}) \mid \mathcal{F}_k) = \mathcal{E}(X \mid \mathcal{F}_k)$ by Tower Thm.

Ex: Toss a fair coin three times. Let $X_i = 1$ if ith toss is head, $X_i = 0$ otherwise, $i = 1, 2, 3$. Let $f(X_1, X_2, X_3) = \sum_{i=1}^{3} X_i$.

The Doob process. $Y_0 = \mathcal{E}(f(\underset{\sim}{X}) \mid \mathcal{F}_0) = \mathcal{E} f(\underset{\sim}{X}) = \frac{3}{2}$, where $\underset{\sim}{X} = (X_1, X_2, X_3)$

$Y_1 = \mathcal{E}(f(\underset{\sim}{X}) \mid \mathcal{D}_{X_1}) = \mathcal{E}(\sum_{i=1}^{3} X_i \mid X_1) = X_1 + 1$

$Y_2 = \mathcal{E}(f(\underset{\sim}{X}) \mid \mathcal{D}_{X_1, X_2}) = \mathcal{E}(f(\underset{\sim}{X}) \mid X_1, X_2) = X_1 + X_2 + \frac{1}{2}$

$Y_3 = \mathcal{E}(f(\underset{\sim}{X}) \mid \mathcal{D}_{X_1, X_2, X_3}) = X_1 + X_2 + X_3$

$\{\Omega\} = \mathcal{D}_0 \preccurlyeq \mathcal{D}_{X_1} \preccurlyeq \mathcal{D}_{X_1, X_2} \preccurlyeq \mathcal{D}_{X_1, X_2, X_3} = 2^{\Omega}$, where

$$\begin{cases} \mathcal{D}_1 = \{ \{X_1 = 1\}, \{X_1 = 0\} \} \\ \mathcal{D}_2 = \{ \{X_1 = 1, X_2 = 0\}, \{X_1 = 1, X_2 = 1\}, \\ \qquad \{X_1 = 0, X_2 = 1\}, \{X_1 = 0, X_2 = 0\} \} \end{cases}$$

Edge Exposure Martingale

- Consider random graph space $\mathcal{G}_{n,p}$. Lebel the $\binom{n}{2} \stackrel{\text{def}}{=} m$ possible edges with the sequence $1, 2, 3, \cdots, m$.

  Define the rv's $I_j(\omega) = \begin{cases} 1 & \text{if edge } j \text{ appears in } \omega \\ 0 & \text{o.w.} \end{cases}$

  Consider any real-valued function F over $\mathcal{G}_{n,p}$, e.g. the clique number

  The edge exposure martingale is defined to be the sequence of rvs $X_0, X_1, X_2, \cdots, X_m$ s.t.

$$X_0 = \mathcal{E}(F)$$
$$X_1 = \mathcal{E}(F \mid I_1)$$
$$\vdots$$
$$X_{m-1} = \mathcal{E}(F \mid I_1, I_2, \cdots, I_{m-1})$$
$$X_m = F(G_{n,p})$$

Note: $X_0, X_1, \cdots, X_m$ is a Doob martingale.

Vertex Exposure Martingale

- In the same setting as in the edge exposure martingale.

Let $I_{xy}(\omega) = \begin{cases} 1 & \text{if edge } xy \text{ appears in } \omega \\ 0 & \text{o.w.} \end{cases}$

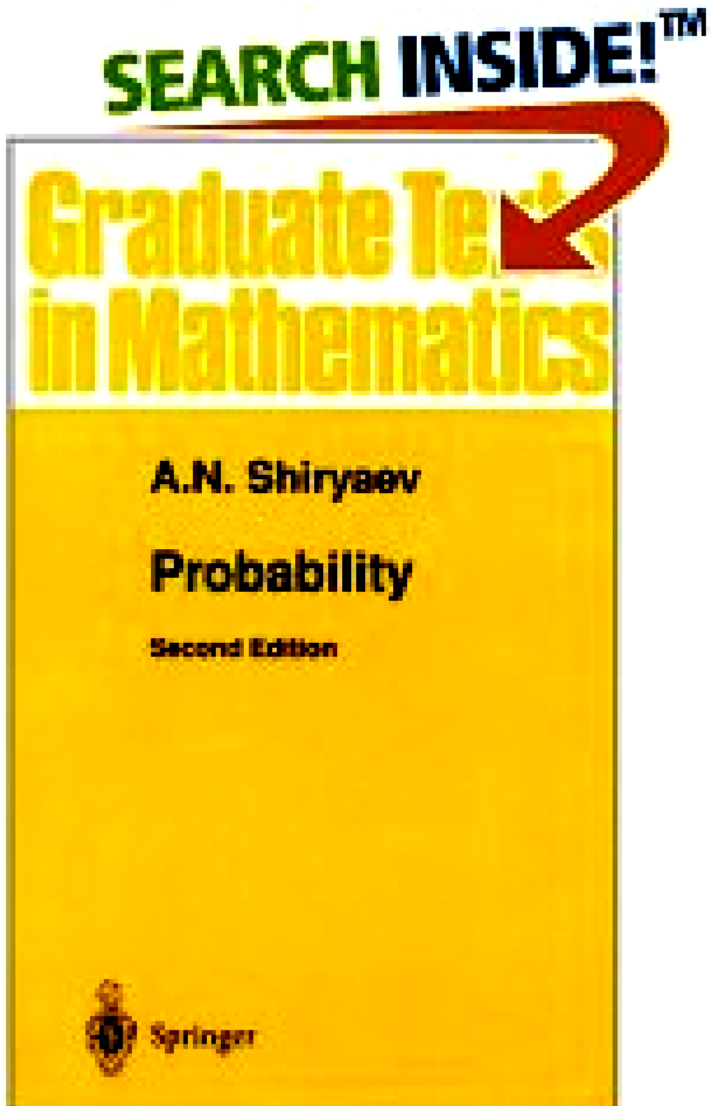
The vertex exposure martingale is defined to be the sequence of rvs $Y_1, Y_2, \cdots, Y_n$ s.t.

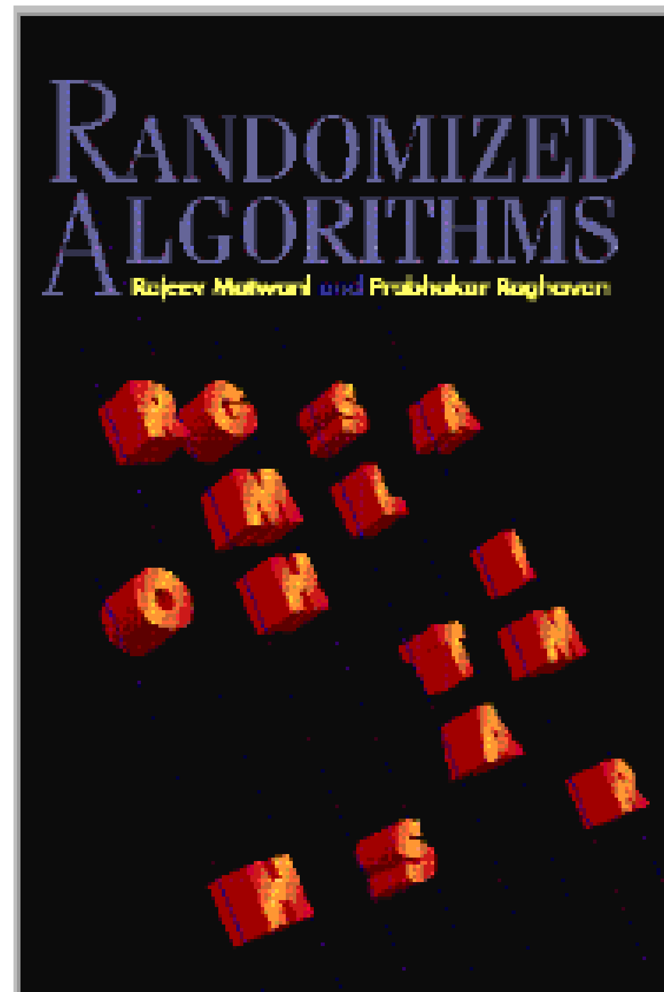
$$Y_1 = \mathcal{E}(F)$$
$$Y_2 = \mathcal{E}(F \mid I_{xy},\ \{x,y\} \in \left[\tbinom{[2]}{2}\right])$$
$$Y_3 = \mathcal{E}(F \mid I_{xy},\ \{x,y\} \in \left[\tbinom{[3]}{2}\right])$$
$$\vdots \qquad\qquad \vdots$$
$$Y_{n-1} = \mathcal{E}(F \mid I_{xy},\ \{x,y\} \in \left[\tbinom{[n-1]}{2}\right])$$
$$Y_n = F(G_{n,p})$$

Note $Y_1, Y_2, \cdots, Y_n$ is a Doob Martingale.

By ordering the edge appropriately the vertex exposure martingale is a subsequence of the edge exposure martingale.

References


SEARCH INSIDE!™
Graduate Texts in Mathematics
A.N. Shiryaev
Probability
Second Edition
Springer


RANDOMIZED ALGORITHMS
Rajeev Motwani and Prabhakar Raghavan