Martingales and Stopping Times

Use of martingales in obtaining bounds and analyzing algorithms

Paris Siminelakis

School of Electrical and Computer Engineering at the National Technical Univercity of Athens

24/3/2010

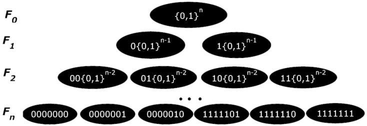
Outline

Filtration

- A *σ-field* $(\Omega, \mathbb{F})$ consists of a sample space Ω and a collection of subsets $\mathbb{F}$ satisfying the following conditions :
  1. Contains the empty set ($\emptyset \in \mathbb{F}$).
  2. Is closed under complement($\mathcal{E} \in \mathbb{F} \Rightarrow \overline{\mathcal{E}} \in \mathbb{F}$).
  3. Is closed under countable union and intersection.
- Given the *σ-field* $(\Omega, \mathbb{F})$ with $\mathbb{F} = 2^{\Omega}$, a filter (sometimes also called a *filtration*) is a nested sequence $\mathbb{F}_0 \subseteq \mathbb{F}_1 \subseteq \ldots \subseteq \mathbb{F}_n$ of subsets of 2^{Ω} such that :
  1. $\mathbb{F}_0 = \{\emptyset, \Omega\}$(*no information)*.
  2. $\mathbb{F}_n = 2^{\Omega}$(*full information)*.
  3. for $0 \leq i \leq$ n, $(\Omega, \mathbb{F}_i)$ is a σ-field(*partial information)*.
- Essentially a filter is a sequence of *σ-fields* such that each new *σ-field* corresponds to the additional information that becomes available at each step and thus the further refinement of the sample space Ω.

Filtration-Examples

- Binary Strings : Consider w a binary string size n. A filter for the sample space $\Omega = \{0,1\}^n$ could be the sequence of sets $\mathbb{F}_i$ such that each set corresponds to the partitioning of the sample space according to the first i bits.



- Americans: Let Ω be the sample space of all Americans.Define the random variable X, denoting the weight of a randomly chosen American. A filter with respect to Ω could be:
  - $\mathbb{F}_0$ is the trivial σ-*field*(no information - no partition).
  - $\mathbb{F}_1$ is the σ-*field* generated by partioning Ω according to sex.
  - $\mathbb{F}_2$ is the σ-*field* generated by the refinement of the previous partion into sets of different heights.
  - $\mathbb{F}_3$ is the further refinement based on age.
  - $\mathbb{F}_4$ is the partition into sigleton sets

Conditional Expectation

- The expectation of a random variable X conditioned on an event A can be viewed as a function of a random variable Y which takes constant real values for every different outcome of A. In other words :

$$\mathbf{E}[X|A] = \mathbf{E}[X|Y] = \mathbf{E}[X|Y = y] = f(y)$$

- If the outcome of the event A or equivalently the value of the variable Y is not known then the conditional expectation itself is a random variable, denoted f(Y).

Americans

Consider the example about Americans. We saw that we can define a filter on the sample space by partitioning appropriately the sample space. Let $\mathbb{F}_0 \subseteq \mathbb{F}_1 \subseteq \ldots \subseteq \mathbb{F}_4$ be the filter we mentioned earlier.Define $X_i = \mathbf{E}[X|\mathbb{F}_i]$, for $0 \leq i \leq 4$. Then :

- $X_0 = \mathbf{E}[X]$ denotes the average weight of an American.
- $X_1 = \mathbf{E}[X|\mathbb{F}_1]$ denotes the average weight of Americans as a function of their sex.
- $X_2 = \mathbf{E}[X|\mathbb{F}_2]$ denotes the average weight as a function of their sex and height.
- $X_3 = \mathbf{E}[X|\mathbb{F}_3]$ denotes the average weight as a function of their sex, height and age.
- Whereas $X_4 = \mathbf{E}[X|\mathbb{F}_4] = X$ corresponds to the weight of an individual American.

6-Sided Unbiased die

- Consider n independent throws of an unbiased 6-sided die. For $0 \leq i \leq 6$, let X_i denote the number of times the value i appears. Then :

$$\mathbf{E}[X_1|X_2] = \frac{n - X_2}{6 - 1}$$

$$\mathbf{E}[X_1|X_2 X_3] = \frac{n - X_2 - X_3}{4}$$

- These equations define the expected value of the random variable X_1 given the number of times 2 and 3 appear. Of course the variables X_2, X_3 are random themselves if they are not given.

Martingales

Martingales originally referred to systems of betting in which a player doubled his stake each time he lost a bet.

Definition

Let $(\Omega, \mathbb{F}, \mathbf{Pr})$ be a propability space with a filter $\mathbb{F}_0$,$\mathbb{F}_1, \ldots$. Suppose that $X_0, X_1, \ldots$ are random variables such that for all $i \geq 0, X_i$ is $\mathbb{F}_i$ measurable (*constant over each block in the partition generating* $\mathbb{F}_i$). The sequence $X_0, \ldots, X_n$ is a martingale provided that for all $i \geq 0$

$$\mathbf{E}[X_{i+1}|\mathbb{F}_i] = X_i$$

Edge Exposure Martingale

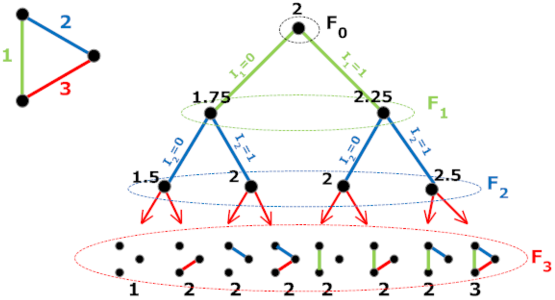
- Let G be a random graph on the vertex set $V = \{1, \ldots, n\}$ obtained by intepedently choosing to include each possible edge with propability p. The underlying propability space is called $\mathcal{G}_{n,p}$.
- Arbitarily label the $m = n(n-1)/2$ edges with the sequence $1, \ldots, m$. For $1 \leq i \leq m$, define the inidcator random variable I_j which takes value 1 if edge j is present in G, and has value 0 otherwise. These indicator variables are independent and each takes value 1 with propability p.
- Consider any real-valued function F defined over the space of all graphs, e.g., the clique number. The edge exposure martingale is defined to be the sequence of random variables $X_0, \ldots, X_m$ such that :

$$X_k = \mathbf{E}[F(G)|I_1, \ldots, I_k]$$

while $X_0 = \mathbf{E}[F(G)]$ and $X_m = F(G)$.

Edge Exposure Martingale

Consider that $m = n = 3$,and $F(G) = chromatic\ number$ we will show that the sequence $X_0, \ldots, X_m$ is indeed a martingale. Specifically that the following property holds : $\mathbf{E}[X_{i+1}|\mathbb{F}_i] = X_i$.



Vertex Exposure Martingale

- Let again consider the propability space $\mathcal{G}_{n,p}$ mentioned earlier.
- For $1 \leq i \leq n$,let E_i be the set of all possible edges with both end-points in $\{1, \ldots, n\}$.Define the indicator random variables I_j for all $j \in E_i$.
- Again consider any real valued fucntion F(G). The vertex exposure martingale is defined to be the sequence of of random variables $X_0, \ldots, X_n$ such that :

$$X_{i+1} = \mathbf{E}[F(G)|I_j \forall j \in E_i]$$

- The vertex exposure martingale reveals the induced graph G_i generated by only the first i nodes.

Expected Running Time

Let T be the running time of a randomized algorithm $\mathcal{A}$ that uses a total of n random bits, on a specific input. Clearly T is a random variable whose value depends in the random bits.

- Observe that T is $\mathbb{F}_n$ measurable, but in general is not $\mathbb{F}_i$ measurable for $i < n$.
- Define the conditional expectation $T_i = \mathbf{E}[T|\mathbb{F}_i]$ where $\mathbb{F}_i$ is the *σ-field* with the i-first bits known. Observe that $T_0 = \mathbf{E}[T]$ and $T_n = T$.
- T_i is a function of the values of the first i random bits denoting the expected running time for a random choice of the remaining n-i bits. The sequence of random variables $T_0, T_1, \ldots, T_n$ is a martingale.

Why martingales are usefull?

- We have seen various example of filters and the corresponding martingales. They have the nasty habit to come up in a variety of applications.
- We may view the *σ-field* sequence $\mathbb{F}_0 \subseteq \mathbb{F}_1 \subseteq \ldots \subseteq \mathbb{F}_n$ as representing the evolution of the algorithm, with each succesive *σ-field* providing more information about the behaviour of the algorithm.
- The random variables $T_0, T_1, \ldots, T_n$ represent the changing expectation of the running time as more information is revealed about the random choices. As we will see later, if it can be shown that the absolute difference $|T_i - T_{i-1}|$ is suitably bounded, then the random variable T_n behaves like T_0 in the limit.
- We mainly utilize martingales in obtaining concentration bounds.

Lipschitz Condition

- Let $f : D_1 \times D_2 \times \ldots \times D_n \to \mathbb{R}$ be a real valued function with n arguments from possible distinct domains. The function f is said to satisfy the *Lipschitz Condition* if for any $x_1 \in D_1, x_2 \in D_2, \ldots, x_n \in D_n$, any $i \in \{1, \ldots, n\}$ and any $y \in D_i$

$$|f(x_1, \ldots, x_i, \ldots, x_n) - f(x_1, \ldots, y_i, \ldots, x_n)| \leq c$$

- Basically a function satisfies the *Lipschitz Condition* if an arbitary change in the value of any one argument does not change the value of the function by more than a constant c.

Azuma-Hoeffding Inequality

Theorem

Let $(Y, \mathbb{F})$ be a martingale, and suppose that there exists a sequence $K_1, K_1, \ldots, K_n$ of real numbers such that $|Y_i - Y_{i-1}| \leq K_n$ for all i (bounded difference condition). Then :

$$\mathbb{P}(|Y_n - Y_0| \geq x) \leq 2exp(-\frac{1}{2}x^2 / \sum_{i=1}^{n} K_i^2), \qquad x > 0$$

Proof I

- We begin the proof with an elementary inequality that stems from the convexity of $g(d) = e^{\psi d}$.

$$e^{\psi d} \leq \frac{1}{2}(1-d)e^{-\psi} + \frac{1}{2}(1+d)e^{+\psi} \qquad |d| \leq 1.$$

- Applying this to a random variable D having mean 0 and $|D| \leq 1$ we obtain

$$\mathbf{E}(e^{\psi D}) \leq \frac{1}{2}(e^{-\psi} + e^{+\psi}) \leq e^{\frac{1}{2}\psi^2}. \qquad (1)$$

- By applying the Markov Inequality we have :

$$\mathbb{P}(Y_n - Y_0 \geq x) \leq e^{-\theta x}\mathbf{E}(e^{\theta(Y_n - Y_0)}). \qquad (2)$$

- Writing $D_n = Y_n - Y_{n-1}$, we have that :

$$\mathbf{E}(e^{\theta(Y_n - Y_0)}) = \mathbf{E}(e^{\theta(Y_{n-1} - Y_0)}e^{\theta D_n}).$$

Proof II

- Conditioning on $\mathbb{F}_{n-1}$, using the fact that $Y_{n-1} - Y_0$ is $\mathbb{F}_{n-1}$-measurable and applying (1) to the random variable D_n/K_n, we obtain :

$$\mathbf{E}(e^{\theta(Y_n - Y_0)}|\mathbb{F}_{n-1}) = e^{\theta(Y_{n-1}-Y_0)}\mathbf{E}(e^{\theta D_n}|\mathbb{F}_{n-1}) \leq e^{\theta(Y_{n-1}-Y_0)}exp(\frac{1}{2}\theta^2 K_n^2)$$

- Taking expectation of the above inequality, using the fact that $\mathbf{E}[\mathbf{E}[X|Y]] = \mathbf{E}[X]$ and then iterating we find that :

$$\mathbf{E}(e^{\theta(Y_n-Y_0)}) \leq \mathbf{E}(e^{\theta(Y_{n-1}-Y_0)})exp(\frac{1}{2}\theta^2 K_n^2) \leq exp(\frac{1}{2}\theta^2 \sum_{i=1}^{n} K_i^2)$$

- Applying (2)which is known as the *Bernstein Inequality* we obtain :

$$\mathbb{P}(Y_n - Y_0 \geq x) \leq exp(-\theta x + \frac{1}{2}\theta^2 \sum_{i=1}^{n} K_i^2)$$

Proof III

- Suppose $x > 0$, the value that minimizes the exponent is $\theta = x/\sum_{i=1}^{n} K_i^2$. Thus we have :

$$\mathbb{P}(Y_n - Y_0 \geq x) \leq exp(-\frac{1}{2}x^2 \sum_{i=1}^{n} K_i^2)$$

- The same argument is valid with $Y_n - Y_0$ replaced with $Y_0 - Y_n$, and the claim of the theorem follows by adding the two bounds together.
- The Azuma-Hoeffding inequality can be generalized if $a_k \leq Y_k - Y_{k-1} \leq b_k$ to yield :

$$\mathbb{P}(|Y_n - Y_0| \geq x) \leq 2exp(-2x^2/\sum_{i=1}^{n}(b_k - ak)^2), \qquad x > 0$$

- The application of the Azuma-Hoeffding inequality is sometimes called "the method of bounded differences".

Connection with Chernoff Bound

- Let $Z_1, \dots, Z_n$ be independent random variables that take values 0 or 1 each with propability p.
- The random variable $S = \sum_{i=1}^{n} Z_i$ has the binomial distribution with parameters n and p.
- Define a maritngale sequence $X_0, \dots, X_n$ by setting $X_0 = \mathbf{E}[S]$,and, for $1 \leq i \leq n$, $X_i = \mathbf{E}[S|Z_1, \dots, Z_i]$. It is clear that $|X_i - X_{i-1}| \leq 1$, since fixing the value of one variable Z_i can only affect the expected value of the sum by at most 1.
- It follows that the propability that S deviates from its expected value is bounded by :

$$\mathbb{P}(|X_n - X_0| \geq x) \leq 2exp(-\frac{x^2}{2n})$$

- Which is a weaker result than can be inferred from the Chernoff bound approach.

Bin Packing

Given n items with random sizes $X = (X_1, \ldots, X_n)$ uniformly distributed in the interval [0,1] and unlimited collection of unit size bins. The problem is to find the minimum number of bins required to store all the items,denoted B_n. It can be shown that B_n grows approximately linearly in n : $\mathbf{E}[B_n] \to c \cdot n$. How close is B_n to its mean value :

- Define for $i \leq n$, $Y_i = \mathbf{E}(B_n|\mathbb{F}_i)$, where $\mathbb{F}_i$ is the $\sigma - field$ generated by $X_1, \ldots, X_i$.
- It easily seen that $(Y, \mathbb{F})$ is a martingale. Because the items are distributed between [0,1] we derive that $|Y_i - Y_{i-1}| \leq 1$.
- Applying the Azuma inequalitywith $\sum_{i=1}^{n} K_i^2 = n$, we get :

$$\mathbb{P}(|Y_n - Y_0| \geq x) \leq 2exp(-\frac{1}{2}x^2/n)$$

- setting $x = \epsilon n$ we see that the chance that B_n deviates from its mean by ϵn decays exponentially in n.

Chromatic Number

Given a random graph G in $\mathcal{G}_{n,p}$, the chromatic number $\chi(G)$ is the minimum number of colors needed in order to color all vertices of the graph so that no adjacent vertices have the same color. We use the vertex exposure martingale to obtain a concentration result for $\chi(G)$.

- Let G_i be the random subgraph induced by the set of vertices $1, \ldots, i$, let $Z_0 = \mathbf{E}[\chi(G)]$ and let :

$$Z_i = \mathbf{E}[\chi(G)|G_1, \ldots, G_i].$$

- Since a vertex uses no more than one new color, again we have that the gap between Z_i and Z_{i-1} is at most 1. Applying the *Azuma-Hoeffding inequality*,we obtain :

$$\mathbb{P}(|Z_n - Z_0| \geq \lambda\sqrt{n}) \leq 2exp(-\lambda^2/2)$$

- The result holds without knowing the mean. We must note that by using the generalized version of the inequality we obtained a better bound.

Pattern Matching

Let $X = (X_1, \ldots, X_n)$ be a sequence of characters chosen independently and uniformly at random from an alphabet Σ, where $|\Sigma| = s$. Let $B = (B_1, \ldots, B_k)$ be a fixed string of k characters from Σ. Let F be the number of occurences of B in the random string X.The expectation of F is $\mathbf{E}[F] = (n-k+1)(\frac{1}{s})^k$

- Define the martingale sequence $Z_i = \mathbf{E}[F|X_1, \ldots, X_i]$.
- Since each character in the string X cannot participate in no more than k possible matches, we have that the function F satisfies the *lipschitz condition* for bound k. Thus we have that : $|Z_i - Z_{i-1}| \leq k$.
- By applying the general Azuma-Hoeffding inequality we have :

$$\mathbb{P}(|Z_n - Z_0| \geq \epsilon) \leq 2exp(-\epsilon^2/2nk^2)$$

Balls and Bins

Suppose we are throwing m balls independently and uniformly at random at n bins. Let X_i denote the random variable representing the bin into which the ith ball falls. Let F be the number of empty bins after the m balls are thrown. Then the sequence: $Z_i = \mathbf{E}[F|X_1, \ldots, X_i]$ is a martingale.

- The function $F = f(X_1, \ldots, Xm)$ satisfies the Lipschitz Condition with bound 1. Because changing the bin where the ith ball was, will either decrease F by 1(relocate to an empty bin), increase F by 1 (relocate to a non-empty bin) or stay the same.
- Applying the Azuma inequality we obtain :

$$\mathbb{P}(|Z_n - Z_0| \geq x) \leq 2exp(-\frac{1}{2}x^2/m)$$

- This result can be improved by taken more care in bounding the difference $|Z_i - Z_{i-1}|$.

Occupancy Revised

In the Balls and Bins setting we will obtain tighter concentration bounds. Let $Z_0, \ldots, Z_m$ be the martingale sequence defined earlier. Define $z(Y, t)$ as the expectation of Z given that Y bins are empty at time t. The propability that none of these bins does not receive a ball during the last $m - t$ time units is $(1 - 1/n)^{m-t}$.

- By linearity of expectation, we obtain that the number of these bins that remain empty is given by :

$$\mathbf{E}[Z|Y_t] = z(Y,t) = Y_t(1 - \frac{1}{n})^{m-t}$$

  where the random variable Y_t denotes the number of empty bins at time t.

- Then for the martingale sequence we have :

$$Z_{t-1} = z(Y_{t-1}, t-1) = Y_{t-1}(1 - \frac{1}{n})^{m-t+1}$$

Analysis I

Suppose we are at time $t-1$, so that the values of Y_{t-1}, Z_{t-1} are determined. At time t there are two possibilities :

1. With propability $1 - Y_{t-1}/n$, the tth ball goes into a currently non-empty bin. Then $Y_t = Y_{t-1}$ and we have :

$$Z_t = z(Y_t, t) = z(Y_{t-1}, t) = Y_{t-1}(1 - \frac{1}{n})^{m-t}$$

2. With propability Y_{t-1}/n, the tth ball goes into a currently empty bin. Then $Y_t = Y_{t-1} - 1$ and we have :

$$Z_t = z(Y_t, t) = z(Y_{t-1} - 1, t) = (Y_{t-1} - 1)(1 - \frac{1}{n})^{m-t}$$

Analysis II

We will focus on the difference random variable $\Delta_t = Z_t - Z_{t-1}$.
The distribution of Δ_t can be characterized as follows :

1. With propability $1 - Y_{t-1}/n$, the value of Δ_t is :

$$\delta_1 = Y_{t-1}(1 - \frac{1}{n})^{m-t} - Y_{t-1}(1 - \frac{1}{n})^{m-t+1} = \frac{Y_{t-1}}{n}(1 - \frac{1}{n})^{m-t}$$

2. With propability Y_{t-1}/n, the value of Δ_t is :

$$\begin{aligned}\delta_2 &= (Y_{t-1} - 1)(1 - \frac{1}{n})^{m-t} - Y_{t-1}(1 - \frac{1}{n})^{m-t+1} \\ &= Y_{t-1}(1 - \frac{1}{n})^{m-t}(1 - (1 - \frac{1}{n})) - (1 - \frac{1}{n})^{m-t} \\ &= -(1 - \frac{Y_{t-1}}{n})(1 - \frac{1}{n})^{m-t}\end{aligned}$$

Analysis III

- Observing that $0 \leq Y_{t-1} \leq n$, and using δ_1 and δ_2 for the upper and lower bound respectively we obtain :

$$-(1-\frac{1}{n})^{m-t} \leq \Delta_t \leq (1-\frac{1}{n})^{m-t}$$

- For $1 \leq i \leq m$, we set $c_t = (1-\frac{1}{n})^{m-t}$, and we have that $|Z_t - Z_{t-1}| \leq c_t$. Consequently :

$$\sum_{t=1}^{m} c_t^2 = \frac{1-(1-1/n)^{2m}}{1-(1-1/n)^2} = \frac{n^2-\mu^2}{2n-1}$$

where we used the geometric series sum and the expected value $\mu = n(1-1/n)^m$.

- Invoking Azuma-Hoeffding inequality now gives :

$$\mathbb{P}(|Z_n - \mu| \geq \lambda) \leq 2exp(-\frac{\lambda^2(n-1/2)}{n^2-\mu^2})$$

Traveling Salesman Problem

We consider a randomized version of the problem where we have a set of independent and uniformly distributed points $P_1 = (x_1, y_1), P_2 = (x_1, y_1)$, $P_3 = (x_1, y_1), \dots, P_n = (x_n, y_n)$ in the unit square $[0, 1]^2$.

- A route is a permutation π of $\{1, \dots, n\}$. The total length of the journey is :

$$d(\pi) = \sum_{i=1}^{n-1} |P_{\pi(i+1)} - P_{\pi(i)}| + |P_{\pi(n)} - P_{\pi(1)}|$$

- The shortest tour has length $D_n = min_\pi d(\pi)$.We are interested in finding how close is D_n to its mean. We set $Y_i = \mathbf{E}[D_n|\mathbb{F}_i]$ for $i \leq n$ where $\mathbb{F}_i$ is the σ-field generated by $P_1, \dots, P_i$.As before,$(Y, \mathbb{F})$ is a martingale and $Y_n = D_n$, $Y_0 = \mathbf{E}(D_n)$.

Analysis I

We will try to obtain a bounding difference condition for D_n. Let $D_n(i)$ be the minimal-length tour through all points except i, and note that $\mathbf{E}[D_n(i)|\mathbb{F}_i] = \mathbf{E}[D_n(i)|\mathbb{F}_{i-1}]$.

- The vital inequality is :

$$D_n(i) \leq D_n \leq D_n(i) + 2Z_i, \qquad i \leq n-1$$

  where Z_i is the shortest distance from P_i to one of the points $P_{i+1}, \ldots, P_n$.

- It is obvious that $D_n \geq D_n(i)$ .Since every tour of the n points includes a tour of all the points except i. For the second inequality we argue that,let P_j be the closest point to P_i amongst the set $\{P_{i+1}, \ldots, P_n\}$,a (sub-optimal) tour could be when we arrive at P_j to visit P_i and return.

- We must note that because the space is continuous we can come arbitarily close to P_j without visiting it. Thus we have a valid tour.

Analysis II

- Taking conditional expectations of the previous inequality we obtain:

$$\mathbf{E}[D_n(i)|\mathbb{F}_{i-1}] \leq Y_{i-1} \leq \mathbf{E}[D_n(i)|\mathbb{F}_{i-1}] + 2\mathbf{E}[Z_i|\mathbb{F}_{i-1}]$$

$$\mathbf{E}[D_n(i)|\mathbb{F}_i] \leq Y_i \leq \mathbf{E}[D_n(i)|\mathbb{F}_i] + 2\mathbf{E}[Z_i|\mathbb{F}_i]$$

- Manipulating the above inequalities and using the fact $D_n(i)$ is independent of point i. We have :

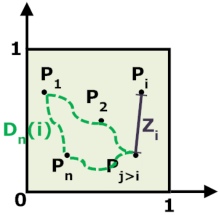
$$|Y_i - Y_{i-1}| \leq 2max\{\mathbf{E}[Z_i|\mathbb{F}_i], \mathbf{E}[Z_i|\mathbb{F}_{i-1}]\} \qquad i \leq n-1 \qquad (1)$$

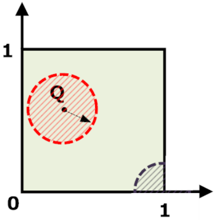
- We need to estimate the right hand side here.

Analysis III

- Let $Q \in [0,1]^2$ and let $Z_i(Q)$ be the shortest distance from Q to the closest of $n-i$ random points. If $Z_i(Q) > r$ then no point lies within the circle $C(r,Q)$.Note that the largest possible distance between two points in the square is $\sqrt{2}$.
- There exists $c(=\pi/4)$ such that for all $r \in (0,\sqrt{2}]$, the intersection of $C(r,Q)$ with the unit square has area at least cr^2.Therefore :

$$\mathbb{P}(Z_i(Q) > r) \leq (1-cr^2)^{n-i}, \qquad 0 < r \leq \sqrt{2}.$$





Analysis IV

- Integrating over x, using the $(1+x) < e^x$ inequality and $\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X > x)dx$ we have for a constant C:

$$\mathbf{E}(Z_i(Q)) \leq \int_0^{\sqrt{2}} (1 - cr^2)^{n-i} dr \leq \int_0^{\sqrt{2}} e^{-cr^2(n-i)} dr < \frac{C}{\sqrt{n-i}} (2)$$

- Since the random variables $\mathbf{E}[Z_i|\mathbb{F}_i], \mathbf{E}[Z_i|\mathbb{F}_{i-1}]$ are each smaller than $C/\sqrt{n-i}$ we have : $|Y_i - Y_{i-1}| \leq 2C/\sqrt{n-i}$ for $i \leq n-1$. For the case $i = n$, we use the trivial bound $|Y_n - Y_{n-1}| \leq 2\sqrt{2}$. Applying the Azuma-Hoeffding Inequality, we obtain :

$$\mathbb{P}(|D_n - \mathbf{E}(D_n)| \geq x) \leq 2exp(-\frac{x^2}{2(8 + \sum_{i=1}^{n-1} 4C^2/i}$$

$$\leq exp(-Ax^2/logn), x > 0.$$

Analysis V

- It can be shown that $\frac{1}{\sqrt{n}}\mathbb{E}(D_n) \to \tau$ as $n \to \infty$ so using the previous result :

$$\mathbf{P}(|D_n - \tau\sqrt{n}| \geq \epsilon\sqrt{n}) \leq 2exp(-\frac{B\epsilon^2 n}{logn}) \qquad \epsilon > 0$$

for some positive constant B and all large n.

Stopping Times

Consider again the betting martingale we saw at the beginning. Due to the martingale property if the number of games is initially fixed then the expected gain from the sequence of games is zero.

- Suppose now that the number of games is not fixed. What happens if the gambler plays a random number of games or even better according to a strategy?
- For example a gambler could be playing until he doubles his original assets. There are many strategies that one can conjure but not all of them are possible to quantify and analyze.

Stopping Times

Definition

A non-negative, integer-valued random variable T is a stopping time for the sequence $\{Z_n, n \geq 0\}$ if the event T=n depends only on the value of the random variables $Z_1, \ldots, Z_n$.

- Essentially a stopping time corresponds to a strategy for determining when to stop a sequence based only on the outcomes seen so far.
- A stopping time could be the first time the gamble has won at least 100 dollars or lost 50 dollars.
- Letting T be the last time the gambler wins before he loses would not be a stopping time since determining whether T=n cannot be done without knowing Z_{n+1}.

Martingale Stopping Theorem

In order to fully utilize the martingale property, we need to characterize conditions on the stopping time T that maintain the property $\mathbb{E}[Z_T] = \mathbb{E}[Z_0]$.

Theorem

if $Z_0, Z_1, \ldots$ is a martingale with respect to $X_1, X_2, \ldots$ and if T is a stopping time for $X_1, X_2, \ldots$ then:

$$\mathbb{E}[Z_T] = \mathbb{E}[Z_0]$$

whenever one of the following holds :

- *the Z_i are bounded.*
- *T is bounded.*
- *$\mathbb{E}[T] < \infty$, and there is a constant c such that $\mathbb{E}[|Z_{i+1} - Z_i| \; |X_1, \ldots, X_i] < c$*

Betting Strategy

We will use the martinale stopping theorem to derive a simple solution to the gambler's ruin problem. Let $Z_0 = 0$, let X_i be the amount won on the ith game and Z_i be the total amount won after i games.Assume that the player quits the game when has either won W or lost L. What is the propability that he wins W dollars before he loses L?

- Let T be the first time has either won W or lost L. Then T is a stopping time for the sequence $X_1, X_2, \ldots$.
- The sequence $Z_1, Z_2, \ldots$ is a martingale and the values are clearly bounded.
- let q be the propability first winning W.We apply the Martingale Stopping Theorem :

$$\mathbb{E}[Z_T] = \mathbb{E}[Z_0] = 0 \qquad \textit{and} \qquad \mathbb{E}[Z_T] = W \cdot q - L(1 - q)$$

$$q = \frac{L}{W + L}$$

Wald's Equation

Wald's equation deals with the expectation of the sum of independent random variables in the case where the number of random variables being summed is itself a random variable.

Theorem

Let $X_1, X_2, \ldots$ be nonnegative, independent, identically distributed random variables with distribution X. Let T be a stopping time for this sequence. If T and X have bounded expectation, then :

$$\mathbb{E}[\sum_{i=1}^{T} X_i] = \mathbb{E}[T] \cdot \mathbb{E}[X]$$

Proof I

For $i \geq 1$, let :

$$Z_i = \sum_{j=1}^{i} (X_j - \mathbb{E}[X]).$$

- The sequence $Z_1, Z_2, \ldots$ is a martingale for $X_1, X_2, \ldots$ and $\mathbb{E}[Z_1] = 0$
- Now, $\mathbb{E}[T] < \infty$(by definition) and

$$\mathbb{E}[|\Delta Z_i| \mathbb{F}_i] = \mathbb{E}[|X_{i+1} - \mathbb{E}[X]|] \leq 2\mathbb{E}[X].$$

- Hence we can apply the martingale stopping theoremto compute :

$$\mathbb{E}[Z_T] = \mathbb{E}[Z_1] = 0.$$

Proof II

We now find by linearity of expectation :

$$\begin{aligned}\mathbb{E}[Z_T] &= \mathbb{E}[\sum_{j=1}^{T}(X_j - \mathbb{E}[X])] \\ &= \mathbb{E}[(\sum_{j=1}^{T}(X_j) - T\mathbb{E}[X]] \\ &= \mathbb{E}[(\sum_{j=1}^{T}(X_j)] - \mathbb{E}[T]\mathbb{E}[X]] \\ &= 0.\end{aligned}$$

which gives the result.

Las Vegas Algorithms

Wald's equation can arise in the analysis of Las Vegas algorithms, which always give the right answer but have variable running times.

- In a Las Vegas algorithm we often repeatedly perform some randomized subroutine that may or may not return the right answer.
- We then use some determenistic checking subroutine to determine whether or not the answer is correct; If it is correct then it terminates, otherwise the algorithm runs the subroutine again.
- If N is the number of trials until a correct answer is found and if X_i is the running time for the two subroutines (*randomized routine and determenistic checking routine*).Then as long as X_i are independent and identically distributed with distribution X, Wald's equation gives that the expected running time of the algorithm is:

$$\mathbb{E}[\sum_{i=1}^{T} X_i] = \mathbb{E}[T] \cdot \mathbb{E}[X]$$

Server Routing

- Consider a set of n servers communicating through a shared channel.Time is divided in time slots, and at each one any server that needs to send a packet can transmit through the channel.
- If exactly one packet is sent at that time, the transmission is completed. If there are more than one, none is succesfull. Packets not sent, are stored in the server's buffer until they are transmitted. Servers follow the following protocol :

Randomized Protocol

At each time slot, if the server's buffer is not empty then with propability $1/n$ it attempts to send the first package in its buffer.

- What is the expected number of time slots used until all servers have sent at least one packet ?

Server Routing

Let N be the number of packets succesfully sent until each server has successfully sent at least one packet. Let t_i be the time slot in which the ith succesfully transmitted packet is sent. Starting from time $t_0 = 0$, and let $r_i = t_i - t_{i-1}$.

- Then T, the number of time slots until each server successfully sends at least one packet, is given by :

$$T = \sum_{i=1}^{N} r_i.$$

- We see that N is independent of r_i, and N is bounded in expectation;thus is a stopping time.

Server Routing

- The propability that a packet is successfully sent in a given time slot is :

$$p = \binom{n}{1}(\frac{1}{n})(1-\frac{1}{n})^{n-1} \approx e^{-1}$$

- The r_i each have a geometric distribution with parameter p, so :

$$\mathbb{E}[r_i] = 1/p \approx e.$$

- The sender of a succesfully transimited packet is uniformly distributed amongst the n servers, independent of previous steps. Using the analysis of the Coupon Collector's problem we deduce that $\mathbb{E}[n] = nH(n) = n\ lnn + On$.
- We now use Wald's identity to compute :

$$\mathbb{E}[T] = \mathbb{E}[\sum_{i=1}^{N}]r_i] = \mathbb{E}[N]\mathbb{E}[r_i] = \frac{nH(n)}{p} \approx en\ lnn.$$

Concluding Remarks

- Using martingales we can obtain bounds even under complex dependencies between the random variables.
- It is not necessary to know the mean value inorder to seek concentration results.
- Appropriately defining martingales and using their properties finds great application in analyzing and designing randomized algorithms,e.g *Las Vegas algorithms*.

Further Reading

R.Motwani, P.Raghavan
Randomized Algorithms.
Cambridge University Press, 1995.

M.Mitzenmacher, E.Upfal
Propability and Computing
Cambridge University Press, 2005.

G.Grimmet, D.Stirzaker
Propability and Random Processes
Oxford University Press, 2001.

M.Habib,C.McDiarmid,J.Ramirez-Alfonsin,B.Reed
Propabilistic Methods
for Algorithmic Descrete Mathematics
Springer, 1998