Probability and Computing

Randomized Algorithms and Probabilistic Analysis

Michael Mitzenmacher
Eli Upfal

CAMBRIDGE

Stochastic Analysis of Dynamic Processes

Eli Upfal

Brown University

"Today's Important Applications are not Programs"

"Today's Important Applications are not Programs"

- Dynamic Processes:
    - Process runs forever (or for very long time).
- Static (batch) processes:
    - Process terminates with an output.

"Today's Important Applications are not Programs"

- Dynamic Processes:
  - Process runs forever (or for very long time).
  - Input is continuously injected to the system.

- Static (batch) processes:
  - Process terminates with an output.
  - All data is available at the beginning of the computation.

"Today's Important Applications are not Programs"

- Dynamic Processes:
  - Process runs forever (or for very long time).
  - Input is continuously injected to the system.
  - Irrevocable decisions made online based on current input, before observing future input items.
- Static (batch) processes:
  - Process terminates with an output.
  - All data is available at the beginning of the computation.

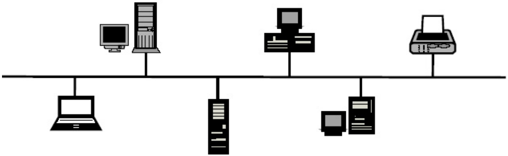
Examples of Dynamic Processes

- Contention-resolution protocols - Ethernet, Aloha protocol.
- Routing protocols - Internet, parallel computation.
- Virtual-memory management - Paging, Prefetching.
- Load balancing protocols - time sharing, distributed scheduling.

Examples of Dynamic Processes

- Contention-resolution protocols - Ethernet, Aloha protocol.
- Routing protocols - Internet, parallel computation.
- Virtual-memory management - Paging, Prefetching.
- Load balancing protocols - time sharing, distributed scheduling.
- Markovian Decision Processes - Computational finance (equity and commodity trading), stochastic planning, etc.

Contention Resolution Protocols



Ethernet Channel:

- One request can be satisfied at a given time.
- If more than one request is submitted, no request is satisfied.
- A sender can detect if its request is satisfied.
- A satisfied request occupies the resource for one step.
- No other communication/coordination between the senders.

Protocol: based on history of success/failure decide when to try to submit next.

Analysis of Dynamic Processes

- Performance measures: Instead of the run-time we are interested in the long term (steady state) performance of the process:
  - Stability: The (expected) number of requests/jobs in the systems is bounded;
  - Utility; What fraction of requests is actually executed (on average)?
  - Efficiency: How long does it take (on average) to process a request?

Input Model

Performance depends on the timing of input arrival:

- Standard worst-case analysis - meaningless in most cases
- Competitive (online) analysis
- Restricted input:
  - Deterministic bounds on input stream - Adversarial Queuing Theory
  - Stochastic input: Assuming some probability distribution or statistical properties on input distribution

Stochastic Analysis

- Assume some underlying distribution on possible input sequences.

Stochastic Analysis

- Assume some underlying distribution on possible input sequences.
- Input distribution can vary in time.

Stochastic Analysis

- Assume some underlying distribution on possible input sequences.
- Input distribution can vary in time.
- Known or unknown input distribution.

Stochastic Analysis

- Assume some underlying distribution on possible input sequences.
- Input distribution can vary in time.
- Known or unknown input distribution.
- Input distribution is characterized by some statistical properties (bounds on moments - stochastic adversarial analysis)

Stochastic Analysis

- Assume some underlying distribution on possible input sequences.
- Input distribution can vary in time.
- Known or unknown input distribution.
- Input distribution is characterized by some statistical properties (bounds on moments - stochastic adversarial analysis)
- Study:
  - Stability: The (expected) number of requests/jobs in the systems is bounded;
  - Loss rate: Expected fraction of requests that are actually executed?
  - Recover-time: Expected time to recover from a bad state.
  - Efficiency: Expected cost/time to process a request?

Related Questions Are Studied in ...

- Queuing theory;
- Markovian decision process.
- Theory of infinite stochastic processes;
- Dynamic systems.
- Control theory.

Stochastioc Analysis

Tools: modeling temporal relations between random variables.

Course Outline

Course Outline

- Martingales
- Application: sampling web search results

Course Outline

- Martingales
- Application: sampling web search results
- Drift criteria
- Application: load balancing protocol

Course Outline

- Martingales
- Application: sampling web search results
- Drift criteria
- Application: load balancing protocol
- Poisson process
- Equilibrium through differential equations
- Application: the supermarket model (dynamic balanced allocation)
- Application: analyzing P2P network protocol

Course Outline

- Martingales
- Application: sampling web search results
- Drift criteria
- Application: load balancing protocol
- Poisson process
- Equilibrium through differential equations
- Application: the supermarket model (dynamic balanced allocation)
- Application: analyzing P2P network protocol
- Dynamic analysis through pure combinatorial arguments

Stochastic Performance Measures

λ : Arrival rate - expected number of new requests in a unit time interval.
$N(t)$: Number of requests in the system at time t.
$W(t)$: The waiting + execution time of a request that entered the system at time t.
Stability: We say that the system is stable if the expected number of requests in the system is bounded with respect to time,

$$\lim_{t \to \infty} E[N(t)] < \infty.$$

Little's Formula

$$N = \lim_{t\to\infty} \frac{1}{t} \sum_{i=1}^{t} N(i)$$

Assume $\alpha(t)$ arrivals during $[0, t]$, in times $\tau_1, \ldots, \tau_{\alpha(t)}$. Then

$$W = \lim_{t\to\infty} \frac{\sum_{i=1}^{\alpha(t)} W(\tau_i)}{\alpha(t)}.$$

Little's Equation: If N is bounded then the system is stable and

$$N = \lambda W.$$

“Proof”

The expected number of new arrivals the interval $[0, t]$ is λt .
The expected number of departures is $\frac{N}{W}t$.
In the limit

$$\lambda t = \frac{N}{W}t$$

or

$$\lambda W = N.$$

[non-lattice, with probability 1]

Martingales

A simple but powerful technique for modeling stationary stochastic processes.

gambler's ruin

Consider a sequence of independent, fair gambling games:

- In each round, a player wins a dollar with probability $1/2$ or loses a dollar with probability $1/2$.
- The player quits the game when she either loses ℓ_1 or wins ℓ_2 dollars.
- What is the probability that the player wins ℓ_2 dollars before losing ℓ_1 dollars?
- The expected change in the "fortune" of the player is always 0

Martingales

Definition

A sequence of random variables $Z_0, Z_1, \ldots$ is a *martingale* with respect to the sequence $X_0, X_1, \ldots$ if for all $n \geq 0$ the following conditions hold:

- Z_n is a function of $X_0, X_1, \ldots X_n$;
- $\mathbf{E}[|Z_n|] < \infty$;
- $\mathbf{E}[Z_{n+1} \mid X_0, \ldots, X_n] = Z_n$.

A sequence of random variables $Z_0, Z_1, \ldots$ is called martingale when it is a martingale with respect to itself. That is, $\mathbf{E}[|Z_n|] < \infty$, and $\mathbf{E}[Z_{n+1} \mid Z_0, \ldots, Z_n] = Z_n$.

Example - Gambling

Consider a sequence of fair games.
$X_i =$ be the amount the gambler wins on the i-th game (X_i negative if the gambler loses).
$Z_i =$ be the gambler's total winnings at the end of the i-th game.
$\mathbf{E}[X_i] = 0$, and

$$\mathbf{E}[Z_{i+1} \mid X_1, X_2, \ldots, X_i] = Z_i + \mathbf{E}[X_{i+1}] = Z_i.$$

Thus, $Z_1, Z_2, \ldots, Z_n$ is a martingale with respect to the sequence $X_1, X_2, \ldots, X_n$.
The sequence is a martingale even if the ammount bet at each game is different and dependends upon previous results.

Doob martingale

Let $X_0, X_1, \ldots, X_n$ be a sequence of random variables, and let Y be a random variable with $\mathbf{E}[|Y|] < \infty$. Then

$$Z_i = \mathbf{E}[Y \mid X_0, \ldots, X_i], \qquad i = 0, 1, \ldots, n,$$

gives a martingale with respect to $X_0, X_1, \ldots, X_n$, since

$$\begin{aligned}\mathbf{E}[Z_{i+1} \mid X_0, \ldots, X_i] &= \mathbf{E}[\mathbf{E}[Y \mid X_0, \ldots, X_{i+1}] \mid X_0, \ldots, X_i] \\ &= \mathbf{E}[Y \mid X_0, \ldots X_i] \\ &= Z_i.\end{aligned}$$

If Y is fully determined by $X_1, \ldots, X_n$, then

$$\mathbf{E}[Y] = Z_0, Z_1, Z_2, \ldots, Z_i = \mathbf{E}[Y \mid X_0, \ldots, X_i], \ldots, Z_n = Y.$$

Balls and Bins

We throw m balls independently and uniformly at random into n bins.
$X_i =$ the random variable representing the bin that the ith ball falls into.
$F =$ the number of empty bins after the m balls are thrown.
The sequence

$$Z_i = \mathbf{E}[F \mid X_1, \ldots, X_i]$$

is a Doob martingale.

Example: Edge Exposure Martingale

Let G be a random graph from $G_{n,p}$. (n vertices, each possible edge exists with probability p independent of other possible edges.) Label the $m = \binom{n}{2}$ possible edge slots in some arbitrary order.

$$X_j = \begin{cases} 1 & \text{if there is an edge in the } j\text{-th edge slot,} \\ 0 & \text{otherwise.} \end{cases}$$

Consider any finite-valued function F defined over graphs; for example, let $F(G)$ be the size of the largest independent set in G. Let $Z_0 = \mathbf{E}[F(G)]$, and

$$Z_i = \mathbf{E}[F(G) \mid X_1, \ldots, X_i], \quad i = 1, \ldots, m.$$

The sequence $Z_0, Z_1, \ldots, Z_m$ is a Doob martingale that represents the conditional expectations of $F(G)$ as we reveal whether each edge is in the graph, one edge at a time.

Lemma

If the sequence $Z_0, Z_1, \ldots, Z_n$ *is a martingale with respect to* $X_0, X_1, \ldots, X_n$, *then for all* $0 \le i \le n$,

$$\mathbf{E}[Z_n] = \mathbf{E}[Z_0].$$

Proof.

Since $Z_0, Z_1, \ldots$ is a martingale with respect to $X_0, X_1, \ldots, X_n$,

$$Z_i = \mathbf{E}[Z_{i+1} \mid X_0, \ldots, X_i].$$

Taking the expectation of both sides and using the definition of conditional expectation, we have

$$\mathbf{E}[Z_i] = \mathbf{E}[\mathbf{E}[Z_{i+1} \mid X_0, \ldots, X_i]] = \mathbf{E}[Z_{i+1}].$$

Repeating this argument, we have $\mathbf{E}[Z_n] = \mathbf{E}[Z_0]$. □

Tail Inequalities for Martingales

Theorem

Azuma-Hoeffding inequality *Let* $X_0, \ldots, X_n$ *be a martingale such that*

$$|X_k - X_{k-1}| \leq c_k.$$

Then for all $t \geq 0$ *and any* $\lambda > 0$*,*

$$\Pr(|X_t - X_0| \geq \lambda) \leq 2\mathrm{e}^{-\lambda^2/(2\sum_{k=1}^{t} c_k^2)}.$$

Corollary

Let $X_0, X_1, \ldots$ *be a martingale such that for all* $k \geq 1$,

$$|X_k - X_{k-1}| \leq c.$$

Then for all $t \geq 1$ *and* $\lambda > 0$,

$$\Pr(|X_t - X_0| \geq \lambda c\sqrt{t}) \leq 2e^{-\lambda^2/2}.$$

proof

We use Markov Inequality (Chernoff's style)

$$\Pr(|X_t - X_0| \geq \lambda) \leq \frac{\mathbf{E}[\mathrm{e}^{\alpha(X_t - X_0)}]}{e^{\lambda}}.$$

To bound for $\mathbf{E}[\mathrm{e}^{\alpha(X_t - X_0)}]$ we define

$$Y_i = X_i - X_{i-1}, \ i = 1, \ldots, t.$$

Note that $|Y_i| \leq c_i$, and since $X_0, X_1, \ldots$ is a martingale,

$$\mathbf{E}[Y_i \mid X_0, X_1, \ldots, X_{i-1}] = \mathbf{E}[X_i - X_{i-1} \mid X_0, X_1, \ldots, X_{i-1}]$$

$$= \mathbf{E}[X_i \mid X_0, X_1, \ldots, X_{i-1}] - X_{i-1} = 0.$$

Consider

$$\mathbf{E}[\mathrm{e}^{\alpha Y_i} | X_0, X_1, \ldots, X_{i-1}].$$

Writing

$$Y_i = -c_i \frac{1 - Y_i/c_i}{2} + c_i \frac{1 + Y_i/c_i}{2},$$

and using the convexity of $\mathrm{e}^{\alpha Y_i}$ we have that

$$\begin{aligned} \mathrm{e}^{\alpha Y_i} &\leq \frac{1 - Y_i/c_i}{2}\mathrm{e}^{-\alpha c_i} + \frac{1 + Y_i/c_i}{2}\mathrm{e}^{\alpha c_i} \\ &= \frac{\mathrm{e}^{\alpha c_i} + \mathrm{e}^{-\alpha c_i}}{2} + \frac{Y_i}{2c_i}\left(\mathrm{e}^{\alpha c_i} - \mathrm{e}^{-\alpha c_i}\right). \end{aligned}$$

Since $\mathbf{E}[Y_i \mid X_0, X_1, \ldots, X_{i-1}] = 0$, we have

$$\begin{aligned} &\mathbf{E}[\mathrm{e}^{\alpha Y_i} | X_0, X_1, \ldots, X_{i-1}] \\ \le\ &\mathbf{E}\left[\tfrac{\mathrm{e}^{\alpha c_i} + \mathrm{e}^{-\alpha c_i}}{2} + \tfrac{Y_i}{2c_i}\left(\mathrm{e}^{\alpha c_i} - \mathrm{e}^{-\alpha c_i}\right) | X_0, X_1, \ldots, X_{i-1}\right] \\ =\ &\tfrac{\mathrm{e}^{\alpha c_i} + \mathrm{e}^{-\alpha c_i}}{2} \le \mathrm{e}^{(\alpha c_i)^2/2}. \end{aligned}$$

Using the Taylor series expansion of e^x to find

$$\frac{\mathrm{e}^{\alpha c_i} + \mathrm{e}^{-\alpha c_i}}{2} \le \mathrm{e}^{(\alpha c_i)^2/2},$$

Since $Y_i = X_i - X_{i-1}$,

$$\begin{aligned}
\mathbf{E}\left[e^{\alpha(X_t - X_0)}\right] &= \mathbf{E}\left[\prod_{i=1}^{t-1} e^{\alpha Y_i}\right] \\
&= \mathbf{E}\left[\prod_{i=1}^{t-2} e^{\alpha Y_i}\right] \mathbf{E}[e^{\alpha Y_{t-1}} | X_0, X_1, \ldots, X_{t-2}] \\
&\leq \mathbf{E}\left[\prod_{i=1}^{t-2} e^{\alpha Y_i}\right] e^{(\alpha c_t)^2/2} \\
&\leq e^{\alpha^2 \sum_{i=1}^{t} c_i^2/2}.
\end{aligned}$$

$$\begin{aligned}
\Pr(X_t - X_0 \geq \lambda) &= \Pr(\mathrm{e}^{\alpha(X_t - X_0)} \geq \mathrm{e}^{\alpha\lambda}) \\
&\leq \frac{\mathbf{E}[\mathrm{e}^{\alpha(X_t - X_0)}]}{\mathrm{e}^{\alpha\lambda}} \\
&\leq \mathrm{e}^{\alpha^2 \sum_{i=1}^{t} c_i^2 / 2 - \alpha\lambda} \\
&\leq \mathrm{e}^{-\frac{\lambda^2}{2\sum_{k=1}^{t} c_k^2}},
\end{aligned}$$

by setting $\alpha = \lambda / \sum_{k=1}^{t} c_k^2$.

Application: Balls and Bins

We throw m balls independently and uniformly at random into n bins.
$X_i =$ the random variable representing the bin that the ith ball falls into.
$F =$ the number of empty bins after the m balls are thrown.
The sequence

$$Z_i = \mathbf{E}[F \mid X_1, \ldots, X_i]$$

is a Doob martingale and $|Z_i - Z_{i-1}| < 1$
By the Azuma-Hoeffding inequality

$$\Pr(|F - \mathbf{E}[F]| \geq \epsilon) \leq 2e^{-2\epsilon^2/m}.$$

$$\mathbf{E}[F] = n\left(1 - \frac{1}{n}\right)^m.$$

Application: Chromatic Number

The *chromatic number* $\chi(G)$ of a graph 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.
Let G be a random grapg in $G_{n,p}$.
Let G_i be the random subgraph of G induced by the set of vertices $1, \ldots, i$.
Let $Z_0 = \mathbf{E}[\chi(G)]$, and

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

Since a vertex uses no more than one new color, $|Z_i - Z_{i-1}| \leq 1$.

$$\Pr(|\chi(G) - \mathbf{E}[\chi(G)]| \geq \lambda\sqrt{n}) \leq 2\mathrm{e}^{-2\lambda^2}.$$

This result holds even without knowing $\mathbf{E}[\chi(G)]$.

Stopping Times

Lemma

If the sequence $Z_0, Z_1, \ldots, Z_n$ *is a martingale with respect to* $X_0, X_1, \ldots, X_n$*, then for all* $0 \leq i \leq n$*,*

$$\mathbf{E}[Z_n] = \mathbf{E}[Z_0].$$

Proof.

Since $Z_0, Z_1, \ldots$ is a martingale with respect to $X_0, X_1, \ldots, X_n$,

$$Z_i = \mathbf{E}[Z_{i+1} \mid X_0, \ldots, X_i].$$

Taking the expectation of both sides and using the definition of conditional expectation, we have

$$\mathbf{E}[Z_i] = \mathbf{E}[\mathbf{E}[Z_{i+1} \mid X_0, \ldots, X_i]] = \mathbf{E}[Z_{i+1}].$$

Repeating this argument, we have $\mathbf{E}[Z_n] = \mathbf{E}[Z_0]$. □

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 $Z_0, Z_1, \ldots, Z_n$ (or independent of Z_k for $k \geq n+1$).

Theorem

Martingale Stopping Theorem *If $Z_0, Z_1, \ldots$ is a martingale with respect to $X_1, X_2, \ldots$, and T is a stopping time for $X_1, X_2, \ldots$, then*

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

whenever one of the following holds:

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

gambler's ruin

Consider a sequence of independent, fair gambling games.
In each round, a player wins a dollar with probability $1/2$ or loses a dollar with probability $1/2$.
$X_i =$ the amount won on the i-th game.
$Z_i =$ the total won by the player after i games ($Z_0 = 0$).
Assume that the player quits the game when she either loses ℓ_1 or wins ℓ_2 dollars.
What is the probability that the player wins ℓ_2 dollars before losing ℓ_1 dollars?

T be the first time the player has either won ℓ_2 or lost ℓ_1.
Then T is a stopping time for $X_1, X_2, \ldots$.
The sequence $Z_0, Z_1, \ldots$ is a martingale, and the values of the Z_i's are bounded.
Applying the martingale stopping theorem we have

$$\mathbf{E}[Z_T] = \mathbf{E}[Z_0] = 0.$$

Let q be the probability that the gambler quits playing after winning ℓ_2 dollars.

$$\mathbf{E}[Z_T] = \ell_2 q - \ell_1(1-q) \;\; = \;\; 0,$$

giving

$$q \;\; = \;\; \frac{\ell_1}{\ell_1 + \ell_2},$$

Application: A Ballot Theorem

Two candidates run for an election. Candidate A obtains a votes, and candidate B obtains $b < a$ votes.

The votes are counted in a random order, chosen uniformly at random from all permutations on the $a + b$ votes.

Show that the probability that candidate A is always ahead in the count is $\frac{a-b}{a+b}$.

Wald's Equation

Theorem

Wald's equation *Let* $X_1, X_2, \ldots$ *be non-negative, 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*

$$\mathbf{E}\left[\sum_{i=1}^{T} X_i\right] = \mathbf{E}[T] \cdot \mathbf{E}[X].$$

In fact Wald's equation holds more generally; there are different proofs of the equality that do not require the random variables $X_1, X_2, \ldots$ to be non-negative.

proof

For $i \geq 1$, let

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

The sequence $Z_1, Z_2, \ldots$ is a martingale with respect to $X_1, X_2, \ldots$, and $\mathbf{E}[Z_1] = 0$.
Now, $\mathbf{E}[T] < \infty$ and

$$\mathbf{E}[|Z_{i+1} - Z_i| \mid X_1, \ldots, X_i] = \mathbf{E}[|X_{i+1} - \mathbf{E}[X]|] \leq 2\mathbf{E}[X].$$

Hence we can apply the martingale stopping theorem to compute

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

We now find

$$\begin{aligned}
\mathbf{E}[Z_T] &= \mathbf{E}\left[\sum_{j=1}^{T}(X_j - \mathbf{E}[X])\right] \\
&= \mathbf{E}\left[\sum_{j=1}^{T} X_j - T\mathbf{E}[X]\right] \\
&= \mathbf{E}\left[\sum_{j=1}^{T} X_j\right] - \mathbf{E}[T]\cdot\mathbf{E}[X] \\
&= 0,
\end{aligned}$$

which gives the result.

Simple example

Consider a gambling game in which a player first rolls one standard die. If the outcome of the roll is X then she rolls X new standard dice and her gain Z is the sum of the outcomes of the X dice.
What is the expected gain of this game?
For $1 \leq i \leq X$, let Y_i be the outcome of the i-th die in the second round. Then

$$\mathbf{E}[Z] = \mathbf{E}\left[\sum_{i=1}^{X} Y_i\right].$$

Applying Wald's equality we obtain

$$\mathbf{E}[Z] = \mathbf{E}[X] \cdot \mathbf{E}[Y_i] = \left(\frac{7}{2}\right)^2 = \frac{49}{4}.$$

Example: Ethernet Protocol

- n servers communicating through a shared channel.
- Time is divided into *discrete* slots.
- At each time slot, any server that has a packet can transmit it through the channel.
- If exactly one packet is sent at that time, the transmission is successfully completed. If more than one packet is sent, then none are successful (and the senders detect the failure).
- Packets are stored in the server's buffer until they are successfully transmitted.
- Servers follow the following simple protocol: at each time slot, if the server's buffer is not empty, then with probability $\frac{1}{n}$ it attempts to send the first packet in its buffer.

Assume that servers have an infinite sequence of packets in their buffers. What is the expected number of time slots until each server successfully sends at least one packet?

$N =$ be the number of packets succesfully sent until each server successfully sends at least one packet.

$t_i =$ the time slot in which the i-th packet is successfully transmitted ($t_0 = 0$)

Let $r_i = t_i - t_{i-1}$.

$T =$ the number of time slots until each server successfully sends at least one packet.

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

The probability that a packet is successfully sent in a given time slot is given by

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

The r_i's each have a geometric distribution with parameter p, so $\mathbf{E}[r_i] = \frac{1}{p} \approx \mathrm{e}$.
Given that a packet was successfully sent at a given time slot, the sender of that packet is uniformly distributed among the n servers, independent of previous steps. By the coupon collector's analysis $\mathbf{E}[N] = nH(n) = n \ln n + O(n)$.
We use the Wald's equality to compute

$$\begin{aligned}\mathbf{E}[T] &= \mathbf{E}\left[\sum_{i=1}^{N} r_i\right] \\ &= \mathbf{E}[N] \cdot \mathbf{E}[r_i] \\ &= \frac{nH(n)}{p},\end{aligned}$$

which is about $\mathrm{e} n \ln n$.

stochastic counting process

Consider a sequence of events occurring at random times. Let $N(t)$ denote the number of events in interval $[0, t]$. The process $\{N(t),\ t \geq 0\}$ is a *stochastic counting process*.

The Poisson Process

Definition

A *Poisson process* with parameter (or rate) λ is a stochastic counting process $\{N(t),\ t \geq 0\}$ such that:

1. $N(0) = 0$.
2. The process has independent and stationary increments. That is, for any $t, s > 0$, the distribution of $N(t+s) - N(s)$, is identical to the distribution of $N(t)$, and for any two disjoint intervals $[t_1, t_2]$ and $[t_3, t_4]$, the distribution of $N(t_2) - N(t_1)$ is independent of the distribution of $N(t_4) - N(t_3)$.
3. $\lim_{t \to 0} \frac{\Pr(N(t)=1)}{t} = \lambda$. That is, the probability of an event in a short interval t is proportional to λt.
4. $\lim_{t \to 0} \frac{\Pr(N(t) \geq 2)}{t} = 0$. That is, the probability of more than one event is a short interval t tends to 0.

Theorem

Let $\{N(t) \mid t \geq 0\}$ *be a Poisson process. Then for any* $t, s \geq 0$ *and any integer* $n \geq 0$,

$$P_n(t) = \Pr(N(t+s) - N(s) = n) = \mathrm{e}^{-\lambda t}\frac{(\lambda t)^n}{n!}.$$

$P_n(t)$ is well defined since the distribution of $N(t+s) - N(s)$ depends only on t and is independent of s.
To compute $P_0(t)$:

$$P_0(t+h) = P_0(t)P_0(h)$$

$$\begin{aligned} & \frac{P_0(t+h) - P_0(t)}{h} = P_0(t)\frac{P_0(h) - 1}{h} \\ = \quad & P_0(t)\frac{1 - \Pr(N(h) = 1) - \Pr(N(h) \geq 2) - 1}{h} \\ = \quad & P_0(t)\frac{-\Pr(N(h) = 1) - \Pr(N(h) \geq 2)}{h}, \end{aligned}$$

$$
\begin{aligned}
P_0'(t) &= \lim_{h\to 0} \frac{P_0(t+h) - P_0(t)}{h} \\
&= \lim_{h\to 0} P_0(t) \frac{-\Pr(N(h)=1) - \Pr(N(h)\geq 2)}{h} \\
&= -\lambda P_0(t).
\end{aligned}
$$

$$P_0(t) = C\mathrm{e}^{-\lambda t}$$

Since $P_0(0) = 1$ we conclude that

$$P_0(t) = \mathrm{e}^{-\lambda t}. \qquad (1)$$

For $n \geq 1$

$$
\begin{aligned}
P_n(t+h) &= \sum_{k=0}^{n} P_{n-k}(t) P_k(h) \\
&= P_n(t)P_0(h) + P_{n-1}(t)P_1(h) + \\
&\qquad \sum_{k=2}^{n} P_{n-k}(t) \Pr(N(h) = k).
\end{aligned}
$$

Computing the first derivative of $P_n(t)$ we get

$$
\begin{aligned}
P_n'(t) &= \lim_{h \to 0} \frac{P_n(t+h) - P_n(t)}{h} \\
&= \lim_{h \to 0} \frac{1}{h} \Big(P_n(t)(P_0(h) - 1) + P_{n-1}(t)P_1(h) + \\
&\qquad \sum_{k=2}^{n} P_{n-k}(t) \Pr(N(h) = k) \Big) \\
&= -\lambda P_n(t) + \lambda P_{n-1}(t).
\end{aligned}
$$

To solve

$$P_n'(t) = -\lambda P_n(t) + \lambda P_{n-1}(t)$$

we write

$$e^{\lambda t}(P_n'(t) + \lambda P_n(t)) = e^{\lambda t}\lambda P_{n-1}(t),$$

which gives

$$\frac{d}{dt}(e^{\lambda t}P_n(t)) = \lambda e^{\lambda t}P_{n-1}(t). \qquad (2)$$

$$\frac{d}{dt}(e^{\lambda t}P_1(t)) = \lambda e^{\lambda t}P_0(t) = \lambda$$

implying

$$P_1(t) = (\lambda t + c)e^{-\lambda t}.$$

Since $P_1(0) = 0$ we conclude that

$$P_1(t) = \lambda t e^{-\lambda t}. \qquad (3)$$

We continue by induction on n to prove that for all $n \geq 0$,

$$P_n(t) = \mathrm{e}^{-\lambda t} \frac{(\lambda t)^n}{n!}.$$

Using (2) and the induction hypothesis gives

$$\frac{d}{dt}(\mathrm{e}^{\lambda t} P_n(t)) = \lambda \mathrm{e}^{\lambda t} P_{n-1}(t) = \frac{\lambda^n t^{n-1}}{(n-1)!}.$$

Integrating and using the fact that $P_n(0) = 0$ gives the result.

Interarrival Distribution

Let X_1 be the time of the first event of the Poisson process, and X_n be the interval of time between the $(n-1)$-st and the n-th event.

Theorem

X_1 has an exponential distribution with parameter λ.

Proof.

$$\Pr(X_1 > t) = \Pr(N(t) = 0) = \mathrm{e}^{-\lambda t}.$$

Thus,

$$F(X_1) = 1 - \Pr(X_1 > t) = 1 - \mathrm{e}^{-\lambda t}.$$

□

Theorem

The random variables $X_i, \ i = 1, 2, \ldots$ *are independent, identically distributed, exponential random variables with parameter* λ*.*

Proof.

$$
\begin{aligned}
& \Pr(X_i > t_i \mid (X_0 = t_0) \cdots \cap (X_{i-1} = t_{i-1})) \\
= \; & \Pr(N(\sum_{k=0}^{i} t_k) - N(\sum_{k=0}^{i-1} t_k) = 0) = \mathrm{e}^{-\lambda t_i}.
\end{aligned}
$$

□

Combining and Splitting Poisson Processes

Theorem

Let $N_1(t)$ and $N_2(t)$ be independent Poisson processes with parameters λ_1 and λ_2, respectively. Then $N_1(t) + N_2(t)$ is a Poisson process with parameter $\lambda_1 + \lambda_2$, and each event for the process $N_1(t) + N_2(t)$ arises from the process $N_1(t)$ with probability $\frac{\lambda_1}{\lambda_1+\lambda_2}$.

The interarrival times for the two processes are independent. Let T_1 and T_2 be the times for the first arrival for N_1 and N_2, respectively.

$$\begin{aligned}
& \Pr((T_1 \le x) \cap (T_2 \le y)) \\
= & \Pr((N_1(x) \ge 1) \cap (N_2(y) \ge 1)) \\
= & \Pr(N_1(x) \ge 1) \Pr(N_2(y) \ge 1) \\
= & \Pr(T_1 \le x) \Pr(T_2 \le y).
\end{aligned}$$

The interarrival time for $N_1(t) + N_2(t)$ is exponentially distributed with parameter $\lambda_1 + \lambda_2$, and hence $N_1(t) + N_2(t)$ is a Poisson process with parameter $\lambda_1 + \lambda_2$.

Each event for $N_1(t) + N_2(t)$ comes from the process $N_1(t)$ with probability $\frac{\lambda_1}{\lambda_1+\lambda_2}$.

Theorem

Suppose that we have a Poisson process $N(t)$ *with rate* λ*. Each event is independently labeled as being Type 1 with probability* p *and Type 2 with probability* $1 - p$*. Then the Type 1 events form a Poisson process* $N_1(t)$ *of rate* λp*, the Type 2 events form a Poisson process* $N_2(t)$ *of rate* $\lambda(1 - p)$*, and the two Poisson processes are independent.*

Proof:

We show that the Type 1 events in fact form a Poisson process.

$$\begin{aligned}
\Pr(T > t) &= \sum_{k=0}^{\infty} \Pr(N_1(t) = 0 \mid N(t) = k) \cdot \Pr(N(t) = k) \\
&= \sum_{k=0}^{\infty} (1-p)^k \frac{\mathrm{e}^{-\lambda t}(\lambda t)^k}{k!} \\
&= \mathrm{e}^{-\lambda t} \sum_{k=0}^{\infty} \frac{(\lambda t(1-p))^k}{k!} \\
&= \mathrm{e}^{-\lambda t}\mathrm{e}^{\lambda t(1-p)} = \mathrm{e}^{-(\lambda p)t}.
\end{aligned}$$

The interarrival distribution of Type 1 events is exponential with parameter λp, and therefore $N_1(t)$ is a Poisson process.

To show independence,

$$
\begin{aligned}
& \Pr((N_1(t) = m) \cap (N_2(t) = n)) \\
= \; & \Pr((N(t) = m + n) \cap (N_2(t) = n)) \\
= \; & \frac{\mathrm{e}^{-\lambda t}(\lambda t)^{m+n}}{(m+n)!} \binom{m+n}{n} p^m (1-p)^n \\
= \; & \frac{\mathrm{e}^{-\lambda t}(\lambda t)^{m+n}}{m!n!} p^m (1-p)^n \\
= \; & \frac{\mathrm{e}^{-\lambda tp}(\lambda tp)^m}{m!} \frac{\mathrm{e}^{-\lambda t(1-p)}(\lambda t(1-p))^n}{n!} \\
= \; & \Pr(N_1(t) = m) \cdot \Pr(N_2(t) = n).
\end{aligned}
$$

Conditional Arrival Time Distribution

Theorem

Given that $N(t) = n$, the n arrival times have the same distribution as the order statistics of n independent random variables with uniform distribution over $[0, t]$.

Proof.

$$
\begin{aligned}
\Pr(X_1 < s | N(t) = 1) &= \frac{\Pr((X_1 < s) \cap (N(t) = 1))}{\Pr(N(t) = 1)} \\
&= \frac{\Pr((N(s) = 1) \cap (N(t) - N(s) = 0))}{\Pr(N(t) = 1)} \\
&= \frac{\left(\lambda s \mathrm{e}^{-\lambda s}\right) \mathrm{e}^{-\lambda(t-s)}}{\lambda t \mathrm{e}^{-\lambda t}} \\
&= \frac{s}{t}.
\end{aligned}
$$

□

Discrete-space Continuous Time Markov Processes

Definition

A continuous time random process $\{X_t \mid t \geq 0\}$, is *Markovian* (or is called a *Markov process*) if for all $s, t \geq 0$:

$$\Pr(X(s+t) = x \mid X(u), 0 \leq u \leq t) =$$

$$\Pr(X(s+t) = x \mid X(t) = y),$$

and this probability is independent of the time t.

A discrete-space continuous time Markov process can be expressed as a combination of two random processes:

1. A transition matrix $P = (p_{i,j})$; where $p_{i,j}$ is the probability that the next state is j given that the current state is i. The matrix P is called the *embedded* or *skeleton* Markov chain of the corresponding Markov process.
2. A vector of parameters $(\theta_1, \theta_2, \ldots)$, such that the distribution of time the process spends in state i before moving to the next step is exponential with parameter θ_i.

stationary distribution

$P_{j,i}(t) =$ the probability of being in state i at time t when starting from state j at time 0.

$$\lim_{t\to\infty} P_{j,i}(t) = \pi_i.$$

If the initial state (at $t = 0$) j is chosen from the stationary distribution, then the probability of being in state i at time t is π_i for all $t > 0$.
To determine the stationary distribution:

$$\begin{aligned}
P'_{j,i}(t) &= \lim_{h\to 0} \frac{P_{j,i}(t+h) - P_{j,i}(t)}{h} \\
&= \lim_{h\to 0} \frac{\sum_k P_{j,k}(t) P_{k,i}(h) - P_{j,i}(t)}{h} \\
&= \lim_{h\to 0} \left(\sum_{k\neq i} \frac{P_{k,i}(h)}{h} P_{j,k}(t) - \frac{1 - P_{i,i}(h)}{h} P_{j,i}(t) \right).
\end{aligned}$$

$$\lim_{h\to 0} \frac{P_{k,i}(h)}{h} = \theta_k p_{k,i}.$$

$$\lim_{h\to 0} \frac{1 - P_{i,i}(h)}{h} = \theta_i(1 - p_{i,i}).$$

$$\begin{aligned}
&\lim_{h\to 0} \left(\sum_{k\neq i} \frac{P_{k,i}(h)}{h} P_{j,k}(t) - \frac{1 - P_{i,i}(h)}{h} P_{j,i}(t) \right) \\
= &\sum_{k\neq i} \theta_k p_{k,i} P_{j,k}(t) - P_{j,i}(t)(\theta_i - \theta_i p_{i,i}) \\
= &\sum_{k} \theta_k p_{k,i} P_{j,k}(t) - \theta_i P_{j,i}(t).
\end{aligned}$$

Now taking the limit as $t \to \infty$ gives

$$\begin{aligned} & \lim_{t\to\infty} P'_{j,i}(t) \\ = \; & \lim_{t\to\infty} \sum_k \theta_k p_{k,i} P_{j,k}(t) - \theta_i P_{i,i}(t) = \sum_k \theta_k p_{k,i} \pi_k - \theta_i \pi_i . \end{aligned}$$

If the process has a stationary distribution, it must be that

$$\lim_{t\to\infty} P'_{j,i}(t) = 0.$$

$$\pi_i \theta_i = \sum_k \pi_k \theta_k p_{k,i}.$$

Application: $M/M/1$ Queue

- FIFO queue.
- Customers arrive to a queue according to a Poisson process with parameter λ.
- One server.
- The service times for the customers are independent and exponentially distributed with parameter μ.
- Let $M(t)$ be the number of customers in the queue at time t.
- The process $\{M(t) \mid t \geq 0\}$ defines a continuous-time Markov process. $P_k(t) = \Pr(M(t) = k)$

$$
\begin{aligned}
\frac{dP_0(t)}{dt} &= \lim_{h\to 0} \frac{P_0(t+h) - P_0(t)}{h} \\
&= \lim_{h\to 0} \frac{P_0(t)(1-\lambda h) + P_1(t)\mu h - P_0(t)}{h} \\
&= -\lambda P_0(t) + \mu P_1(t), \qquad (4)
\end{aligned}
$$

and for $k \geq 1$,

$$
\begin{aligned}
&\frac{dP_k(t)}{dt} = \lim_{h\to 0} \frac{P_k(t+h) - P_k(t)}{h} \\
&= \lim_{h\to 0} \frac{P_k(t)(1-\lambda h - \mu h) + P_{k-1}(t)\lambda h + P_{k+1}(t)\mu h - P_k(t)}{h} \\
&= -(\lambda + \mu) P_k(t) + \lambda P_{k-1}(t) + \mu P_{k+1}(t).
\end{aligned}
$$

In equilibrium

$$\frac{dP_k(t)}{dt} = 0 \quad \text{for} \quad k = 0, 1, 2, \ldots.$$

$$\pi_0 = 1 - \frac{\lambda}{\mu}$$

$$\pi_k = \left(1 - \frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^k.$$

Queue Parameters

- The number of customers in the system $+1$ has geometric distribution with parameter $1 - \frac{\lambda}{\mu}$
- Expected number of customer in the system

$$L = \frac{1}{1 - \frac{\lambda}{\mu}} - 1 = \frac{\lambda}{\mu - \lambda}.$$

- Let W be the expected time a customer spends in the system.

$$W = \frac{L}{\lambda} = \frac{1}{\mu - \lambda}$$

Little's Formula

$$N = \lim_{t \to \infty} \frac{1}{t} \sum_{i=1}^{t} N(i)$$

Assume $\alpha(t)$ arrivals during $[0, t]$, in times $\tau_1, \ldots, \tau_{\alpha(t)}$. Then

$$W = \lim_{t \to \infty} \frac{\sum_{i=1}^{\alpha(t)} W(\tau_i)}{\alpha(t)}.$$

Little's Equation: If N is bounded then the system is stable and

$$N = \lambda W.$$

“Proof”

The expected number of new arrivals the interval $[0, t]$ is λt .
The expected number of departures is $\frac{N}{W}t$.
In the limit

$$\lambda t = \frac{N}{W}t$$

or

$$\lambda W = N.$$

[non-lattice, with probability 1]

Little's Result

Let W be the expected time a customer spends in the system.
$\alpha(t) =$ total number of arrivals up to time t.
$\beta(t) =$ total time spent by all customers in the system up to time t.

$$W_t \lambda_t = \frac{\beta(t)}{\alpha(t)} \frac{\alpha(t)}{t} = \frac{\beta(t)}{t} = L_t$$

Assume that the following limits exist:
$\lim_{t \to \infty} W_t = W$
$\lim_{t \to \infty} \lambda_t = \lambda$
$\lim_{t \to \infty} L_t = L$
Then

$$L = \lambda W$$

$M/M/1/K$ Queue in Equilibrium

An $M/M/1/K$ queue is an $M/M/1$ queue with bounded queue size. If a customer arrives while the queue already has K customers this customer leaves the system instead of joining queue.

$$\pi_k = \begin{cases} \pi_0(\frac{\lambda}{\mu})^k & \text{for } k \leq K \\ 0 & \text{for } k > K \end{cases}$$

and

$$\pi_0 = \frac{1}{\sum_{k=0}^{K}(\frac{\lambda}{\mu})^k}.$$

$M/M/\infty$ Queue

Customers arrive to the coffee shop according to a Poisson process with parameter λ and stay for interval with exponential distribution with parameter μ (all independent).

$$\pi_0\lambda = \pi_1\mu,$$

and for $k \geq 1$

$$\pi_k(\lambda + k\mu) = \pi_{k-1}\lambda + \pi_{k+1}(k+1)\mu. \tag{5}$$

$$\begin{aligned} \pi_{k+1}(k+1)\mu &= \pi_k(\lambda + k\mu) - \pi_{k-1}\lambda \\ &= \pi_k\lambda + \pi_k k\mu - \pi_{k-1}\lambda. \end{aligned}$$

$$\pi_k k\mu = \pi_{k-1}\lambda,$$

$$\pi_{k+1} = \frac{\lambda}{\mu(k+1)} \pi_k.$$

$$\pi_k = \pi_0 \left(\frac{\lambda}{\mu}\right)^k \frac{1}{k!},$$

$$1 = \sum_{k=0}^{\infty} \pi_0 \left(\frac{\lambda}{\mu}\right)^k \frac{1}{k!} = \pi_0 \mathrm{e}^{\lambda/\mu}.$$

$\pi_0 = \mathrm{e}^{-\lambda/\mu}$ and

$$\pi_k = \frac{\mathrm{e}^{-\lambda/\mu}(\lambda/\mu)^k}{k!},$$

The equilibrium distribution is the discrete Poisson distribution with parameter λ/μ.

Second Approach

$M(t) =$ number of customers at time t, assuming $M(0) = 0$.
Let $N(t) =$ the total number of customers that arrived in the interval $[0, t]$.

$$\Pr(M(t) = j) = \sum_{n=0}^{\infty} \Pr(M(t) = j \mid N(t) = n) \mathrm{e}^{-\lambda t} \frac{(\lambda t)^n}{n!}. \quad (6)$$

If a customer arrived at time x, the probability that she is still in at time t is $\mathrm{e}^{-\mu(t-x)}$.
the arrival time of an arbitrary customer is uniform on $[0, t]$.
The probability that an arbitrary customer is still in at time t is given by

$$p = \int_0^t \mathrm{e}^{-\mu(t-x)} \frac{dx}{t} = \frac{1}{\mu t}(1 - \mathrm{e}^{-\mu t}).$$

Since the events for different users are independent, for $j \leq n$

$$\Pr(M(t) = j \mid N(t) = n) = \binom{n}{j} p^j (1-p)^{n-j}.$$

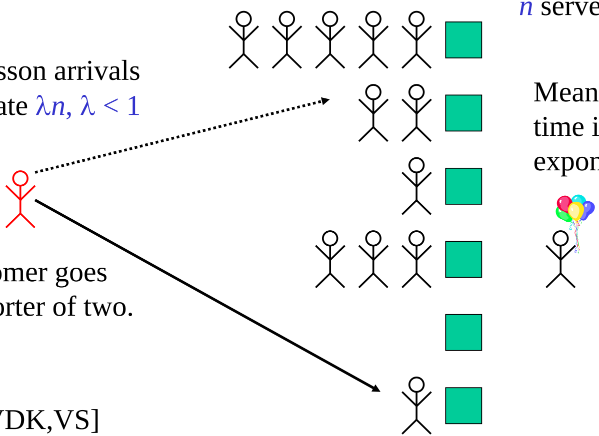
$$\begin{aligned}
\Pr(M(t) = j) &= \sum_{n=j}^{\infty} \binom{n}{j} p^j (1-p)^{n-j} \mathrm{e}^{-\lambda t} \frac{(\lambda t)^n}{n!} \\
&= \mathrm{e}^{-\lambda t} \frac{(\lambda tp)^j}{j!} \sum_{n=j}^{\infty} \frac{(\lambda t(1-p))^{n-j}}{(n-j)!} \\
&= \mathrm{e}^{-\lambda t} \frac{(\lambda tp)^j}{j!} \sum_{m=0}^{\infty} \frac{(\lambda t(1-p))^{m}}{(m)!} \\
&= \mathrm{e}^{-\lambda t} \frac{(\lambda tp)^j}{j!} \mathrm{e}^{\lambda t(1-p)} \\
&= \mathrm{e}^{-\lambda tp} \frac{(\lambda tp)^j}{j!}.
\end{aligned}$$

Thus, $M(t)$ has a Poisson distribution with parameter λtp.

Supermarket Model

n servers

Poisson arrivals at rate λn, $\lambda < 1$



Mean service time is 1, exponential

Customer goes to shorter of two.

[M,VDK,VS]

- $s_k =$ fraction of queues with at least k customers.
- The state of the system at time t: $(s_0(t), s_1(t), s_2(t), \ldots)$
- Fraction of queues with k customers is

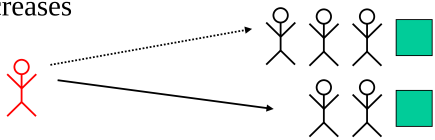
$$S_k - S_{k+1}$$

- Probability that the smallest of d random choices has $k-1$ customers

$$S_{k-1}^d - S_k^d$$

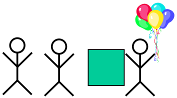
Setting Differential Equations

rate s_k increases



$$(\lambda n\, dt)\left(s_{k-1}^2 - s_k^2\right)/n$$

rate s_k decreases



$$n\left(s_k - s_{k+1}\right)(dt)/n$$

Expected behavior of process

$$\begin{cases} \frac{ds_i}{dt} = \lambda(s_{i-1}^d - s_i^d) - (s_i - s_{i+1}) & \text{for } i \geq 1 \\ s_0 = 1 \end{cases} \tag{7}$$

In equilibrium (fixed point), for all i we have

$$\frac{ds_k}{dt} = 0$$

Summing

$$\sum_{i \geq k} \frac{ds_i}{dt} = \lambda(s_{i-1}^d - s_i^d) - (s_i - s_{i+1}) = 0$$

we get

$$-\lambda s_k^2 + s_{k+1} = 0$$

which gives

$$\pi_k = \lambda^{2^k - 1}.$$

Comparison:

- Choosing a random queue:
  - Each queue is an $M/M/1$ queue
  - For each queue $\pi_k = \lambda^k$
  - Expected maximum queue length: $\frac{\log n}{\log \log n} + O(1)$
- Choosing the best of d random choices:
  - For each queue $\pi_k = \lambda^{2^k - 1}$
  - Expected maximum queue length $\frac{\log \log n}{\log d} + O(1)$

Whats Missing?

Why do the differential equations describe the random process?

Kurtzs Theorem

Over fixed time intervals and for fixed finite dimensional processes, the deviation of the random process from the solution to the differential equations obeys Chernoff-like bounds.

$$\Pr(\sup_{t,i} |s_i(t) - \hat{s}_i(t)| \geq \epsilon) \leq e^{-cn\epsilon^2}$$

Problem: Kurtz's Theorem (generally stated) requires fixed finite dimensional spaces