Notes on Randomized Algorithms

Mahdi Zamani
Visa Research, Palo Alto, CA
mzamani@visa.com

1 Basics

Poisson Trials. A sequence of *independent* 0-1 random variables $(X_1, ..., X_n)$, where for every $i \in [n]$,

$$\begin{aligned} \Pr(X_i \text{ succeeds}) &= \Pr(X_i = 1) = p_i \\ \Pr(X_i \text{ fails}) &= \Pr(X_i = 0) = 1 - p_i \end{aligned}$$

$$0 \leq p_i \leq 1.$$

Bernoulli Trials. A special case of Poisson trials, where the independent random variables have the same distribution, i.e., for every $i \in [n]$,

$$\begin{aligned} \Pr(X_i = 1) &= p \\ \Pr(X_i = 0) &= 1 - p \end{aligned}$$

$$0 \leq p \leq 1.$$

Binomial Distribution. The distribution of the number of successes in a sequence of Bernoulli trials, e.g., the number of heads in a sequence of n coin flips. A binomial random variable X with parameters n and p, denoted by $B(n, p)$, is defined by the following probability distribution on $x = 0, 1, ..., n$:

$$\Pr(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$\mathbf{E}[X] = np.$$

The binomial distribution can also be used to approximate the probability that x items are marked in a sample of size n from an *infinite* pool of items, where the probability of each item being marked is p. This is similar to sampling with replacement from the pool. If sampling is done without replacement (*i.e.*, the sample size relative to the population size is not negligible), then the hypergeometric distribution (see below) can offer a better approximation.

Geometric Distribution. The distribution of the number of trials in a Bernoulli process until we get a success, e.g., the number of times to flip a coin until it lands on heads. A geometric random variable X with parameter p is defined by the following probability distribution:

$$\Pr(X = n) = p(1-p)^{n-1}$$

$$\mathbf{E}[X] = 1/p.$$

Poisson Distribution. A Poisson random variable X with parameter μ is defined by the following probability distribution on $x = 0, 1, ..., n$:

$$\Pr(X = x) = \frac{e^{-\mu}\mu^x}{x!}.$$

- The Poisson distribution expresses the probability of a given number of events occurring in a fixed interval of time or space if these events occur with a known constant rate and independently of the time since the last event.
- When throwing m balls into n bins, the distribution of the number of balls in a bin is approximately Poisson with $\mu = m/n$ which is exactly the average number of balls per bin.

Exponential Distribution. The exponential distribution is *memoryless* because the past has no impact on its future behavior.

Bounds on Poisson Random Variables. Let X be a Poisson random variable with parameter μ.

1. If $x > \mu$, then
$$\Pr(X \geq x) \leq \frac{e^{-\mu}(e\mu)^x}{x^x}.$$
2. If $x < \mu$, then
$$\Pr(X \leq x) \leq \frac{e^{-\mu}(e\mu)^x}{x^x}.$$

Hypergeometric Distribution. In binomial distribution, the probability of success, p, remains the same for each trial because the trials are assumed to be independent of each other. The hypergeometric distribution happens when sampling is performed from a finite population without replacement thus making trials dependent on each other. When the sample size is small relative to the population size, the random variable could still be approximated by the binomial distribution, but if the sample size is large the effect on p will be large so the trials become very dependent. In such cases, the hypergeometric distribution gives a much better approximation. The hypergeometric distribution calculates the probability of having x marked elements when randomly selecting a sample of size m without replacement from a population of n elements containing t marked elements:

$$\Pr(X = x) = \frac{\binom{t}{x}\binom{n-t}{m-x}}{\binom{n}{m}}.$$

The expected value of X is $\mathbf{E}[X] = \frac{mt}{n}$.

Properties of Expectations.

- If X is a random variable and a is a constant, then $\mathbf{E}[aX] = a\mathbf{E}[X]$. Thus, $\mathbf{E}[\mathbf{E}[X]] = \mathbf{E}[X]$, because $\mathbf{E}[X]$ is a constant.
- Linearity of expectation: For any set of random variables $X_1, ..., X_n$, we have $\mathbf{E}[\sum_{i=1}^n X_i] = \sum_{i=1}^n \mathbf{E}[X_i]$.

2 Concentration Bounds

Union Bound (Boole's Inequality). The probability that at least one event in a set of events happens is no greater than the sum of the probabilities of the individial events, i.e., for a set of events $A_1, ..., A_n$,

$$\Pr\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \Pr(A_i).$$

Markov's Inequality. Let X be a random variable that assumes only non-negative values. Then, for all $a > 0$,

$$\Pr(X \geq a) \leq \frac{\mathbf{E}[X]}{a}.$$

Variance. The variance of a random variable X is defined as $\mathbf{Var}[X] = \mathbf{E}[X^2] - (\mathbf{E}[X])^2$.

If X is a binomial random variable, then $\mathbf{Var}[X] = np(1-p)$.

If X is a geometric random variable, then $\mathbf{Var}[X] = (1-p)/p^2$.

Chebyshev's Inequality. For any $a > 0$,

$$\Pr(|X - \mathbf{E}[X]| \geq a) \leq \frac{\mathrm{Var}[X]}{a^2}.$$

Chernoff Bounds. Let $X_1, ..., X_n$ be independent Poisson trials such that $\Pr(X_i) = p_i$. Let $X = \sum_{i=1}^{n} X_i$, $\mu = \mathbf{E}[X]$, and $\mu_L \leq \mu \leq \mu_H$. The following Chernoff bounds hold:

For $\delta > 0$,	$\Pr(X \geq (1+\delta)\mu) \leq \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}$	[?, Page 64]
For $0 < \delta < 1$,	$\Pr(X \leq (1-\delta)\mu) \leq \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{\mu}$	[?, Page 66]
For $0 < \delta < 1$,	$\Pr(X \geq (1+\delta)\mu) \leq e^{-\mu\delta^2/3}$	[?, Page 64]
For $0 < \delta < 1$,	$\Pr(X \leq (1-\delta)\mu) \leq e^{-\mu\delta^2/2}$	[?, Page 66]
For $0 < \delta < 1$,	$\Pr(\lvert X - \mu\rvert \geq \delta\mu) \leq 2e^{-\mu\delta^2/3}$	[?, Page 67]
For $\delta > 0$,	$\Pr(X < \mu - \delta),\ \Pr(X > \mu + \delta) \leq e^{-2\delta^2/n}$	[?, Page 6]
For $\delta > 0$,	$\Pr(X < \mu_L - \delta),\ \Pr(X > \mu_H + \delta) \leq e^{-2\delta^2/n}$	[?, Page 8]
For $\delta > 2e\mu$,	$\Pr(X > \delta) \leq 2^{-\delta}$	[?, Page 7]
For $\delta > 1$,	$\Pr(X \geq (1+\delta)\mu) \leq e^{-\mu\delta/3}$	Wikipedia

Convex Function. A function $f : \mathbb{R} \to \mathbb{R}$ is said to be convex if and only if, for any x_1, x_2 and $0 \leq \lambda \leq 1$,

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2).$$

Jensen's Inequality. If f is a convex function, then $\mathbf{E}[f(X)] \geq f(\mathbf{E}[X])$.

Conditional Expectation. A conditional expectation is not a scalar value; it is a random variable defined as

$$\mathbf{E}[X|Y] = \{\mathbf{E}[X|Y = y] \text{ with } \Pr(Y = y)\}.$$

In other words, $\mathbf{E}[X|Y]$ is a function that maps every $y \in Y$ to a $x \in X$. It is easy to see that

$$\mathbf{E}[\mathbf{E}[X|Y]] = \mathbf{E}[X].$$

Also,

$$\mathbf{E}[\mathbf{E}[X|Y, Z]|Z] = \mathbf{E}[X|Z].$$

3 Martingales

A martingale is a sequence of random variables used to model an event that is a function of a sequence of past events. Martingales are useful for events where the knowledge of past events does not allow to predict the expected value or the actual probability of the event.

For example, suppose that we are throwing m balls independently and uniformly at random into n bins. Let X_i be the random variable representing the bin into which the i-th ball falls, and F be the number of empty bins after the m balls are thrown. Then, the sequence

$$Z_i = \mathrm{E}[F|X_0, ..., X_i]$$

is a martingale. This model allows us to find a concentration bound for F using Azuma's inequality (which we will describe later):

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

Definition 1. *A sequence of random variables $Z_0, Z_1, ...$ is a* martingale with respect to a sequence $X_0, X_1...$, *if for all $i \geq 0$, the following conditions hold:*

1. *Z_i is a function of $X_0, ..., X_i$;*
2. *$\mathbf{E}[|Z_i|] < \infty$; and*
3. *$\mathbf{E}[Z_{i+1}|X_0, ..., X_i] = Z_i$.*

Definition 2. *A sequence of random variables $Z_0, Z_1, ...$ is called a* martingale *when it is a martingale with respect to itself, that is*

1. *$\mathbf{E}[|Z_i|] < \infty$*
2. *$\mathbf{E}[Z_{i+1}|Z_0, ..., Z_i] = Z_i$*

Definition 3. *Let $X_0, ..., X_n$ be a sequence of random variables and Y be a random variable that depends on the X_i's. Then, for $i = 0, 1, ...n$,*

$$Z_i = \mathbf{E}[Y | X_0, ..., X_i]$$

is a martingale with respect to $X_0, ..., X_n$ and is called a Doob martingale.

Definition 4. *A random variable T is a stopping time for $\{Z_i\}$ if the event $T = n$ depends only on $Z_0, ..., Z_n$.*

Theorem 1. [Martingale Stopping Time Theorem] *If $Z_0, Z_1, ...$ is a martingale with respect to $X_0, X_1, ...$ and T is a stopping time for $X_1, X_2, ...,$ then*

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

if one of the following holds:

1. *$|Z_i| < c$ for some constant c.*
2. *T is bounded.*
3. *$\mathbf{E}[T] < \infty$ and there exists a c such that $E[Z_{i+1} - Z_i | X_1, ..., X_i] < c$.*

Theorem 2. [Azuma's Inequality] *Let $X_0, ..., X_n$ be a martingale such that $|X_k - X_{k-1}| \leq c_k$. Then, for all $t \geq 0$ and any $\lambda > 0$,*

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

Theorem 3. [Azuma's Inequality – Tighter Bound] *Let $X_0, ..., X_n$ be a martingale such that*

$$B_k \leq X_k - X_{k-1} \leq B_k + d_k$$

for some constants d_k and for some random variables B_k. Then, for all $t \geq 0$ and any $\lambda > 0$,

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

Definition 5. *[Informal] A* dependency graph *for a set of events $E_1, ..., E_n$ is graph with a vertex for each event, where there is an edge between two vertices iff they have some dependency with each other. In other words, if the two vertices are mutually independent, then there is no edge between them.*

Theorem 4. [Lovasz Local Lemma] *Let $E_1, ..., E_n$ be a set of events and*

1. *for all i, $\Pr(E_i) \leq p$;*
2. *the degree of the dependency graph of $E_1, ..., E_n$ is at most d;*
3. *$ep(d + 1) \leq 1$.*

Then,

$$\Pr\Big(\bigcap_{i=1}^{n} \bar{E}_i\Big) > 0.$$

4 Useful bounds

The following bounds follow from the Taylor series expansion of e^x:

$$1 + x \leq e^x$$

$$1 - x \leq e^{-x}$$

$$\text{For } x \geq 0: \; 1 + x \geq e^x(1 - x^2)$$

$$\text{For } x \geq 0: \; 1 - x \geq e^{-x} - x^2/2$$

References