# The Menu of the Method

Petar Maymounkov
petar@csail.mit.edu

**Abstract**

This writing is my personal concise and fairly incomplete collection of daily-used probabilistic and other techniques. In essence, a lot of it consists of literal copies of frequently read sections of "The Probabilistic Method" by Spencer and Alon together with *italicized* comments and clarifications by myself. In addition, I have included excerpts from various lecture notes. This document is in preliminary and evolving form.

## 1 Basic Method

### 1.1 Linearity of Expectation

Definition of expectation:

$$\mu = \mathbb{E}[X] = \sum_x x \mathbb{P}\mathrm{r}[X = x]$$

Say $X = c_1 X_1 + \cdots + c_m X_m$, then "linearity of expectation" postulates:

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_i c_i X_i\right] = \sum_i c_i \mathbb{E}[X_i]$$

*Linearity of Expectation is such a powerful tool that I find it necessary to mention a few tricks from the literature that are remarkably powerful.*

**Lemma 1.1** (Expectation Taking Rules)**.** *Let $A$ and $B$ be arbitrary, not necessarily numeric, random variables, let $X$ be an indicator random variable, and let $f$ and $g$ be arbitrary functions. Then:*

1. *If $f(\cdot) = g(\cdot)$, then $\mathbb{E}[f(\cdot)] = \mathbb{E}[g(\cdot)]$*
2. $\mathbb{E}_B\Big[\mathbb{E}_A[A|B]\Big] = \mathbb{E}_A[A]$*, or simply* $\mathbb{E}\Big[\mathbb{E}[A|B]\Big] = \mathbb{E}[A]$
3. $\mathbb{E}_B\Big[\mathbb{Pr}_A[A|B]\Big] = \mathbb{Pr}_A[A]$*, or simply* $\mathbb{E}\Big[\mathbb{Pr}[A|B]\Big] = \mathbb{Pr}[A]$
4. $\mathbb{E}_X\Big[\mathbb{I}[X=1]\Big] = \mathbb{E}_X[X]$*, or simply* $\mathbb{E}\Big[\mathbb{I}[X=1]\Big] = \mathbb{E}[X]$

***Example*** 1.1. *This example which uses all three of the above equalities is taken from a beautiful paper of Bollobás and Riordan on scale-free graphs [BR02]. Let $G_1^n$ be a random undirected graph with vertex set $[n]$, generated as follows. Vertices are added one at a time. When added, vertex $j$ has 1 incident edge whose other end is connected to some random vertex $g_j \leq j$ with probability proportional to that vertex' current degree, i.e.:*

$$\mathbb{Pr}\big[g_j = i | G_1^{j-1}\big] = \begin{cases} d_{j-1,i}/(2j-1), & \text{if } j > i, \\ 1/(2j-1), & \text{if } j = i, \end{cases}$$

*where $d_{j-1,i}$ denotes the degree of vertex $i$ in $G_1^{j-1}$. Our goal will be to compute $\mathbb{Pr}[g_j = i]$ in $G_1^n$. Taking the expectation on both sides of this equation for the case $j > i$ produces:*

$$\mathbb{Pr}[g_j = i] = \frac{\mathbb{E}[d_{j-1,i}]}{2j-1}$$

*For $j > i$, we have $d_{j,i} = d_{j-1,i} + \mathbb{I}[g_j = i]$, so taking expectations gives:*

$$\mathbb{E}[d_{j,i}|G_1^{j-1}] = d_{j-1,i} + \frac{d_{j-1,i}}{2j-1} = \left(1 + \frac{1}{2j-1}\right) d_{j-1,i}$$

*Take expectations again! Get:*

$$\mathbb{E}[d_{j,i}] = \left(1 + \frac{1}{2j-1}\right) \mathbb{E}[d_{j-1,i}]$$

*Solve this recurance, observing that $\mathbb{E}[d_{i,i}] = 2i/(2i-1)$, to get:*

$$\mathbb{E}[d_{j,i}] = O(\sqrt{j/i})$$

*Finally, substitute this above to arrive at the final result:*

$$\mathbb{Pr}[g_j = i] = O(1/\sqrt{ij})$$

*Notice that taking expectations helped us find a closed formula for an event in a complex sequentially-defined random process. This scenario bears connections to random processes whose evolution is defined via differential equations. In particular, we could have described the graph model by:*

$$\frac{\Delta d_i}{\Delta j} = \frac{d_i}{2j}$$

## 1.2 Markov's Inequality

**Theorem 1.1** (The Markov Inequality)**.**

$$\mathbb{Pr}\big[X \geq t\mu\big] \leq \frac{1}{t} \quad \textit{equivalently} \quad \mathbb{Pr}\big[X \geq \lambda\big] \leq \frac{\mu}{\lambda}$$

Perhaps the most common application of the expectation uses that there must exist a point in the probability space for which $X \geq \mathbb{E}[X]$ and a point for which $X \leq \mathbb{E}[X]$.

*Note that Markov's inequality is most specific when $t = 1$, i.e. it gives "the most exact" bound in this case. In particular, at $t = 1$ the inequality can overestimate the tail probability by at most a factor of 2. For any $t > 1$ the inequality could possibly overestimate the tail probability by an arbitrarily large factor. This property of the Markov inequality permeats into the Chebyshev and Chernoff bounds.*

*The Chebyshev Inequality (Theorem 2.1) gives (in the same sense) the most exact bound for the probability of the tails that are one standard deviation away from the mean, and a possibly grossly overestimated bound for more distant tails.*

*In contrast, the Chernoff Bound (Section 3) cleverly applies the Markov inequality near $t = 1$ for any input parameters, and thus it achieves a very tight tail bound.*

***Example*** 1.2. Szele (1943) showed that there exists a tournament $T$ with $n$ players and at least $n!/2^{n-1}$ Hamiltonian paths. Let $X$ be the number of Hamiltonian paths in the random tournament. For each permutation $\pi$ let $X_\pi$ be the indicator that $\pi$ gives a Hamiltonian path, i.e. $(\pi(i), \pi(i+1)) \in T$ for $1 \leq \imath < n$. Then $X = \sum_\pi X_\pi$ and:

$$\mathbb{E}[X] = \sum_\pi \mathbb{E}[X_\pi] = n!/2^{n-1}$$

Some tournament has at least $\mathbb{E}[X]$ Hamiltonian paths.

## 2 The Second Moment

**Variance and Covariance:** Definition:

$$\begin{aligned}\sigma^2 = \mathbb{V}\mathtt{ar}[X] &= \mathbb{E}\big[(X - \mathbb{E}[X])^2\big] \\ &= \mathbb{E}[X^2] - \mathbb{E}[X]^2\end{aligned}$$

For $X = X_1 + \cdots + X_n$:

$$\mathbb{V}\mathtt{ar}[X] = \sum_i \mathbb{V}\mathtt{ar}[X_i] + \sum_{i \neq j} \mathbb{C}\mathtt{ov}[X_i, X_j], \quad \text{where}$$

$$\mathbb{C}\mathtt{ov}[Y, Z] = \mathbb{E}[YZ] - \mathbb{E}[Y]\mathbb{E}[Z]$$

**Theorem 2.1** (Chebyshev's Inequality)**.** *For any $\lambda$:*

$$\mathbb{P}\mathtt{r}[|X - \mu| \geq \lambda\sigma] \leq \frac{1}{\lambda^2}$$

Note that Chebyshev's inequality is not too specific. To the contrary, consider the case when $X$ is normally distributed with mean $\mu$ and standard deviation $\sigma$, then for large $\lambda$ asymptotically:

$$\mathbb{P}\mathtt{r}[|X - \mu| \geq \lambda\sigma] \sim \sqrt{\frac{2}{\pi}} \frac{e^{-\lambda^2/2}}{\lambda} \ll \frac{1}{\lambda^2}$$

*The reason for this lies in the nature of the bounds that the Markov Inequality (Theorem 1.1) gives.*

When $X = X_1 + \cdots + X_m$ is a sum of indicator random variables with corresponding probabilities $p_i$:

$$\mathbb{V}\mathtt{ar}[X_i] = p_i(1-p_i) \leq p_i = \mathbb{E}[X_i]$$
$$\mathbb{V}\mathtt{ar}[X] \leq \mathbb{E}[X] + \sum_{i \neq j} \mathbb{C}\mathtt{ov}[X_i, X_j]$$

## 2.1 Bounding Towards and Away from Zero

Let $X$ be integral and nonnegative. If $\mathbb{E}[X] < 1$ we can bound the probability that $X > 0$:

$$\mathbb{P}\mathtt{r}[X > 0] \leq \mathbb{E}[X]$$

Asymptotically, if $\mathbb{E}[X] = o(1)$, then $X = 0$ almost always.

On the contrary, suppose $\mathbb{E}[X] \to \infty$, then:

**Theorem 2.2.**

$$\mathbb{P}\mathtt{r}[X = 0] \leq \frac{\mathbb{V}\mathtt{ar}[X]}{\mathbb{E}[X]^2}$$

And in particular:

**Corollary 2.1.** *If $\mathbb{V}\mathtt{ar}[X] = o(\mathbb{E}[X]^2)$ then $X > 0$ almost always.*

A slight generalization of Theorem 2.2 gives:

**Theorem 2.3.**

$$\mathbb{P}\mathtt{r}\left[|X - \mathbb{E}[X]| \geq \epsilon \mathbb{E}[X]\right] \leq \frac{\mathbb{V}\mathtt{ar}[X]}{\epsilon^2 \mathbb{E}[X]^2}$$

Which amplifies Corollary 2.1 in asymptotic terms:

**Corollary 2.2.** *If $\mathbb{V}\mathtt{ar}[X] = o(\mathbb{E}[X]^2)$ then $X \sim \mathbb{E}[X]$ almost always.*

**Bounding for Indicators:** Suppose again $X = X_1 + \cdots + X_m$ are indicator random variables for events $A_i$. For indices $i, j$ write $i \sim j$ if $i \neq j$

and the events $A_i$ and $A_j$ are not independent. Set (the sum over ordered pairs):

$$\Delta = \sum_{i \sim j} \mathbb{Pr}[A_i \wedge A_j]$$

Note that when $i \sim j$,

$$\mathbb{Cov}[X_i, X_j] = \mathbb{E}[X_i X_j] - E[X_i]E[X_j] \leq \mathbb{E}[X_i X_j] = \mathbb{Pr}[A_i \wedge A_j]$$

and, respectively, that when $i \neq j$ and not $i \sim j$ then $\mathbb{Cov}[X_i, X_j] = 0$. Thus:

$$\mathbb{Var}[X] \leq \mathbb{E}[X] + \Delta$$

**Corollary 2.3.** *If $\mathbb{E}[X] \to \infty$ and $\Delta = o(\mathbb{E}[X]^2)$ then $X > 0$ almost always. Furthermore, $X \sim \mathbb{E}[X]$ almost always.*

When $\Delta > \mu$ and the indicator random variables pertain to monotone decreasing events the extended Janson Inequality 7.3 gives a significantly stronger bound.

*Note that if you are able to compute $\Delta$ in a specific application, you are most likely also able to compute the exact value of $\sum_{i \sim j} \mathbb{Cov}[X_i, X_j]$, which may give a better asymptotic in the above lemma.*

**Bounding for Symmetric Indicators:** Say $X_1, \ldots, X_m$ are symmetric if for every $i \neq j$ there is an automorphism of the underlying probability space that sends event $A_i$ to event $A_j$. In this instance, we write:

$$\Delta = \sum_{i \sim j} \mathbb{Pr}[A_i \wedge A_j] = \sum_i \mathbb{Pr}[A_i] \sum_{j \sim i} \mathbb{Pr}[A_j | A_i]$$

and since under the symmetry assumption the inner summation is independent of $i$, we set:

$$\Delta^* = \sum_{j \sim i} \mathbb{Pr}[A_j | A_i]$$

where $i$ is any fixed index. Then:

$$\Delta = \Delta^* \sum_i \mathbb{Pr}[A_i] = \Delta^* \mathbb{E}[X]$$

**Corollary 2.4.** *If $\mathbb{E}[X] \to \infty$ and $\Delta^* = o(\mathbb{E}[X])$ then $X > 0$ almost always. Furthermore $X \sim \mathbb{E}[X]$ almost always.*

# 3 Chernoff Bounds

## 3.1 Bound for Unbiased Variables

**Theorem 3.1.** *Let $X_1, \ldots, X_n$ be mutually independent with*

$$\mathbb{Pr}[X_i = +1] = \mathbb{Pr}[X_i = -1] = 1/2$$

*and set $X = X_1 + \cdots + X_n$, then for any $a > 0$:*

$$\mathbb{Pr}[X > a] < e^{-a^2/2n}$$

*The proof of this theorem simply considers the variables $Y_i = e^{\lambda X_i}$ and applies the Markov inequality to $e^{\lambda X}$, where its effect is highly amplified (viewed from the point of view of $X$). This transformation is useful only because $\mathbb{E}[Y_i Y_j] = \mathbb{E}[Y_i]\mathbb{E}[Y_j]$ due to the independence condition.*

*Unlike the Chebyshev inequality, the Chernoff bound is achieved by applying the Markov inequality close to the point when it is precise (by picking the parameter $\lambda$ appropriately for any fixed choice of $n$ and $a$). Hence the Chernoff bound tends to be very tight, although a bit worse than an equivalent application of the Central Limit Theorem, which however holds only when $n \to \infty$.*

*I believe that the Chernoff bound can be made perfectly tight to match the Central Limit Theorem exactly at the expense of formulaic convenience. This is most likely an unworthy effort from an asymptotic standpoint.*

**Corollary 3.1.** *Under the assumptions of the previous theorem:*

$$\mathbb{Pr}\big[|S_n| > a\big] < 2e^{-a^2/2n}$$

### 3.2 Bounds for Bernoulli Variables

**Theorem 3.2.** *Let $X = X_1 + \cdots + X_n$ be a sum of independent Bernoulli random variables. Set $\mu = \mathbb{E}[X]$. Then:*

$$\begin{aligned}
\mathbb{Pr}[X \leq (1-\epsilon)\mu] &\leq e^{-\epsilon^2\mu/2} && \text{for } 0 < \epsilon \\
\mathbb{Pr}[X \geq (1+\epsilon)\mu] &\leq e^{-\epsilon^2\mu/3} && \text{for } \epsilon < 1 \\
\mathbb{Pr}[X \geq (1+\epsilon)\mu] &\leq e^{-\epsilon^2\mu/4} && \text{for } 1 < \epsilon < 2e-1 \\
\mathbb{Pr}[X \geq (1+\epsilon)\mu] &\leq 2^{-(1+\epsilon)\mu} && \text{for } 2e-1 < \epsilon
\end{aligned}$$

### 3.3 Bounds for Mixed Bernoulli Variables*

This theorem is due to Janson [Jan02]:

**Theorem 3.3.** *Let $X = X_1 + \cdots + X_n$ where $X_i$ are independent Bernoulli random variables with $\mathbb{Pr}[X_i = 1] = p_i$. Let $\mu = \mathbb{E}[X] = p_1 + \cdots + p_n$. Then for $t \geq 0$ we have:*

$$\begin{aligned}
\mathbb{Pr}[X \geq \mu + t] &\leq \exp\left(-\frac{t^2}{2(\mu + t/3)}\right), \quad \text{and} \\
\mathbb{Pr}[X \leq \mu - t] &\leq \exp\left(-\frac{t^2}{2\mu}\right)
\end{aligned}$$

## 4 Basic Inequalities

**Bonferroni Inequalities:** (The Taylor expansion of probabilities) To prepare the grounds, let $B_1, \ldots, B_m$ be events with corresponding indicator random variables $X = X_1, \ldots, X_m$. Define:

$$S^{(r)} = \sum_{i_1,\ldots,i_r} \mathbb{Pr}[B_{i_1} \wedge \cdots \wedge B_{i_r}]$$

where the sum is over all sets $\{i_1, \ldots, i_r\} \subseteq \{1, \ldots, m\}$. The Inclusion-Exclusion Principle gives that:

$$\mathbb{Pr}[X = 0] = \mathbb{Pr}[\overline{B_1} \wedge \cdots \wedge \overline{B_m}] = 1 - S^{(1)} + S^{(2)} - \cdots + (-1)^r S^{(r)} \ldots$$

This is to be interpreted that, in general, the inclusion-exclusion formula alternately over- and underestimates $\mathbb{Pr}[X = 0]$. So, for example, the union bound is a special case of the Inclusion-Exclusion Principle when the formula is truncated after the $S^{(1)}$ term.

**Jensen Inequality:** If $f$ is convex, $\sum_k \lambda_k = 1$ and $\lambda_k \in [0, 1]$, then:

$$f\left(\sum_k \lambda_k x_k\right) \leq \sum_k \lambda_k f(x_k), \quad \text{also}$$
$$f\big(\mathbb{E}[X]\big) \leq \mathbb{E}\big[f(X)\big]$$

**Cauchy-Schwarz Inequality:**

$$\left(\sum_{i=1}^{n} a_i b_i\right)^2 \leqq \left(\sum_{i=1}^{n} a_i^2\right)\left(\sum_{i=1}^{n} b_i^2\right)$$

This inequality is also summarized by its vector form (after taking the square roots on both sides):

$$|a \cdot b| \leq |a||b|$$

Often the case $b_i = 1$, for all $i$, will be useful, i.e.:

$$\left(\sum_{i=1}^{n} a_i\right)^2 \leqq n\left(\sum_{i=1}^{n} a_i^2\right)$$

# 5 Bounding Techniques

### Union Bound

$$\mathbb{Pr}[\vee_i A_i] \leq \sum_i \mathbb{Pr}[A_i]$$

**Conditional and Conditional Union Bounds**

Generally:

$$\mathbb{Pr}[A|B \wedge C] \geq \mathbb{Pr}[A \wedge B|C]$$

In particular:

$$\begin{aligned}\mathbb{Pr}[A|B] &\geq \mathbb{Pr}[A \wedge B] \\ &= \mathbb{Pr}[A] - \mathbb{Pr}[A \wedge \overline{B}]\end{aligned}$$

When $B = \wedge_i B_i$ the above gives:

$$\mathbb{Pr}[A| \wedge_i B_i] \geq \mathbb{Pr}[A] - \sum_i \mathbb{Pr}[A \wedge \overline{B_i}]$$

**Common Triggering Events**

Suppose $A_1, \ldots, A_n$ and $B_1, \ldots, B_m$ are two sets of events, and let it be that $\mathbb{Pr}[A_i] = \mathbb{Pr}\big[\bigvee_{j \in D_i} B_j\big]$ for some $D_i \subseteq [m]$.

$$\mathbb{Pr}\Big[\bigvee_i A_i\Big] \leq \mathbb{Pr}\Big[\bigvee_j B_j\Big] \leq \sum_j \mathbb{Pr}[B_j]$$

# 6 Martingales and Tight Concentration

A martingale is a sequence of random variables $X_0, \ldots, X_m$ such that:

$$\mathbb{E}[X_{i+1}|X_i, \ldots, X_0] = X_i \tag{6.1}$$

**Theorem 6.1** (Azuma's Inequality)**.** *Let $0 = X_0, \ldots, X_m$ be a martingale with*

$$|X_{i+1} - X_i| \leq 1$$

*for all $0 \leq i < m$. Let $\lambda > 0$ be arbitrary. Then:*

$$\mathbb{Pr}\big[X_m > \lambda\sqrt{m}\big] < e^{-\lambda^2/2}$$

**Corollary 6.1.** *Let $c = X_0, \ldots, X_m$ be a martingale with*

$$|X_{i+1} - X_i| \leq 1$$

*for all $0 \leq i < m$. Then:*

$$\mathbb{P}\mathtt{r}\left[|X_m - c| > \lambda\sqrt{m}\right] < 2e^{-\lambda^2/2}$$

## 6.1 Edge and Vertex Exposure

*Two most typical and very instructive applications of martingales are the edge and vertex exposure martingales, described below. They demonstrate how to setup martingales so that the martingale property "automatically" holds, thus allowing you to focus your attention (in the design) entirely on the numerical aspect of the random process at hand.*

***The Edge Exposure Martingale:*** Let $G \sim G(n,p)$ be a random graph, and let $f$ be a graph theoretic function. Label the potential edges $\{i,j\} \subset [n]$ of $G$ by $e_1, \ldots, e_m$ with $m = \binom{n}{2}$ in any desirable order. And let the indicator random variables $E_1, \ldots, E_m$ be such that $E_i = 1$ if $e_i \in G$, and $E_i = 0$ if $e_i \notin G$. Define the edge exposure martingale $X_1, \ldots, X_m$ as:

$$X_i = \mathbb{E}[f(G)|E_i, \ldots, E_1]$$

Notice that $X_0 = \mathbb{E}[f(G)]$ and $X_m = f(G)$ are constants. *Also notice that the martingale property (6.1) holds precisely because the $E_i$ variables are independent.*

***The Vertex Exposure Martingale:*** Let $G \sim G(n,p)$ be a random graph, and let $f$ be a graph theoretic function. Label the potential edges $\{i,j\} \subset [n]$ of $G$ as $e_{ij}$, and associate them with the corresponding indicator random variables $E_{ij}$. Define $X_1, \ldots, X_n$ by:

$$X_i = \mathbb{E}[f(G)|E_{xy} \text{ for } x, y \leq i]$$

Once again $X_0 = \mathbb{E}[f(G)]$ and $X_m = f(G)$ are constants. Also notice that the vertex exposure martingale can be viewed as a subsequence of the edge exposure martinagle. *This is formalized in the next paragraph.*

***Generally, The Independent Disclosure Martingale:*** *The construction of the edge and vertex exposure martingales are generalized in the following way. Let $Y = (Y_1, \ldots, Y_m)$ be a vector of independent random variables, and let $f(Y)$ be any real-valued function. Also let $\varnothing = W_0 \subseteq W_1 \subseteq \cdots \subseteq W_n = \{Y_i\}$ be a gradation of the set of variables $Y_i$. Define a martingale $X_0, \ldots, X_n$ as:*

$$X_i = \mathbb{E}\big[f(Y)|W_i\big]$$

*As before, $X_0 = \mathbb{E}[f(Y)]$ and $X_n = f(Y)$ are constants. The martingale property (6.1) holds since (using the independence of the $Y_i$'s and linearity of expectation):*

$$\begin{aligned}
\mathbb{E}\big[X_{i+1}|X_i\big] &= \mathbb{E}\Big[\mathbb{E}\big[f(Y)|W_{i+1}\big]\big|W_i\Big] \\
&= \sum_{W_{i+1}\setminus W_i} \mathbb{Pr}\big[W_{i+1}\setminus W_i\big]\mathbb{E}\big[f(Y)|W_{i+1}\big] \\
&= \sum_{W_{i+1}\setminus W_i} \mathbb{Pr}\big[W_{i+1}\setminus W_i\big] \sum_{W_n\setminus W_{i+1}} \mathbb{Pr}\big[W_n\setminus W_{i+1}\big] f(Y) \\
&= \sum_{W_n\setminus W_i} \mathbb{Pr}\big[W_n\setminus W_i\big] f(Y) \\
&= \mathbb{E}\big[f(Y)|W_i\big] \\
&= X_i
\end{aligned}$$

*Note that the $Y_i$'s don't have to be indicator variables, in fact, they don't even have to be real-valued. E.g., the $Y_i$'s can be assigned to disjoint sets of potential edges of $G \sim G(n,p)$ where the value of each $Y_i$ represents the outcome for the entire set of corresponding edges.*

***The Lipschitz Condition:***

A graph theoretic function $f$ is said to satisfy the edge Lipschitz condition if whenever $H$ and $H'$ differ in only one edge then $|f(H) - f(H')| \leq 1$. It satisfies the vertex Lipschitz condition if whenever $H$ and $H'$ differ at only one vertex, $|f(H) - f(H')| \leq 1$.

**Theorem 6.2.** *When $f$ satisfies the edge Lipschitz condition, the corresponding martingale satisfies $|X_{i+1} - X_i| \leq 1$. When $f$ satisfies the vertex Lipschitz condition the corresponding vertex exposure martingale satisfies $|X_{i+1} - X_i| \leq 1$.*

***Example*** 6.1. Shamir and Spencer (1987): Let $n, p$ be arbitrary and let $c = \mathbb{E}[\chi(G)]$ where $G \sim G(n,p)$. Then

$$\mathbb{P}\mathrm{r}\left[|\chi(G) - c| > \lambda\sqrt{n-1}\right] < 2e^{-\lambda^2/2}$$

To prove this, consider the vertex exposure martingale $X_1, \ldots, X_n$ with $f(G) = \chi(G)$. A single vertex can always be given a new color so the Lipschitz condition applies. Now, just apply Corollary 6.1. Notice that when $\lambda \to \infty$ such that $\lambda = o(n)$, this result shows that $\chi(G)$ is tightly concentrated around its mean, and this was shown without any knowledge of what the actual mean is.

# 7 The Poisson Paradigm

*When $X$ is the sum of many rare "mostly independent" random variables and $\mu = \mathbb{E}[X]$ we would like to say that $X$ is close to a Poisson distribution with mean $\mu$ and, in particular, $\mathbb{P}\mathrm{r}[X = 0]$ is nearly $e^{-\mu}$. This rough statement is called the "Poisson Paradigm."*

**Definition 7.1.** *A non-negative integral random variable $X$ is said to be Poisson-distributed with mean $\mu > 0$ if:*

$$\mathbb{P}\mathrm{r}[X = t] = \frac{\mu^t}{t!}e^{-\mu}$$

## 7.1 Basic Facts about Poisson Variables

*The results in this section will be concerned with approximating various probabilistic processes by Poisson-distributed variables. Hence, it seems appropriate to preface this section with some basic tools for dealing with Poisson variables. We begin with a Chernoff bound:*

**Theorem 7.1.** *Let $X$ have Poisson distribution with mean $\mu$. For $\epsilon > 0$:*

$$\mathbb{P}\mathrm{r}[X \leq \mu(1-\epsilon)] \leq e^{-\epsilon^2\mu/2}$$
$$\mathbb{P}\mathrm{r}[X \geq \mu(1+\epsilon)] \leq \left[e^{\epsilon}(1+\epsilon)^{-(1+\epsilon)}\right]^{\mu}$$

A couple other useful facts:

**Lemma 7.1.** *If $X$ and $Y$ are Poisson random variables with means, respectively, $\mu_X$ and $\mu_Y$, then $Z = X + Y$ is Poisson-distributed with mean $\mu_Z = \mu_X + \mu_Y$.*

**Lemma 7.2.** *Let $X$ be Poisson-distributed with mean $\mu$ and let $p_i = \mathbb{Pr}[X = i]$. Then for the generating function of $X$ we have:*

$$p(x) = \sum_{i=0}^{\infty} p_i x^i = e^{c(x-1)}$$

## 7.2 Janson's Inequalities

*Janson's Inequality is used to prove that at least one of many rare mostly independent events happens with high probability, when the events are monotone decreasing and we can compute their pairwise correlation.*

A finite universe set $\Omega$ is given. And a random subset $R \subset \Omega$ is chosen, with $\mathbb{Pr}[r \in R] = p$. Also, a collection of subsets $A_i \subset \Omega$ with $i \in I$ is fixed, and event $B_i$ is defined to be $A_i \subset R$, with corresponding indicator variables $X_i$ and $X = \sum_i X_i$. Write $i \sim j$ whenever $i \neq j$ and $A_i \cap A_j \neq \emptyset$. *The goal of Janson's inequalities is to upper-bound the probability $\mathbb{Pr}[\wedge_i \overline{B_i}]$ that none of the events occur. (The lower bound is trivial, because the events are monotone decreasing, hence less interesting.)*

*The goal of the Janson inequalities is interesting because when the underlying events are monotone decreasing, for any two non-independent events we have $\mathbb{Pr}[\overline{B_i} | \wedge_j \overline{B_j}] > \mathbb{Pr}[\overline{B_i}]$, and therefore to say the least $\mathbb{Pr}[\wedge_i \overline{B_i}] > \prod_i \mathbb{Pr}[\overline{B_i}]$.*

Set:

$$\Delta = \sum_{i \sim j} \mathbb{Pr}[B_i \wedge B_j]$$

$$M = \prod_{i \in I} \mathbb{Pr}[\overline{B_i}]$$

$$\mu = E[X] = \sum_{i \in I} \mathbb{Pr}[B_i]$$

**Theorem 7.2** (The Janson Inequality)**.** *If* $\mathbb{Pr}[B_i] \leq \epsilon$*, then:*

$$M \leq \mathbb{Pr}\big[\wedge_{i\in I}\overline{B_i}\big] \leq Me^{\frac{1}{1-\epsilon}\frac{\Delta}{2}}, \text{ and}$$
$$M \leq \mathbb{Pr}\big[\wedge_{i\in I}\overline{B_i}\big] \leq e^{-\mu+\frac{\Delta}{2}}$$

The two inequalities are very similar, since:

$$\mathbb{Pr}[\overline{B_i}] = 1 - \mathbb{Pr}[B_i] \leq e^{-\mathbb{Pr}[B_i]}$$

and so:

$$M \leq e^{-\mu}$$

Furthermore, oftentimes $M \sim e^{-\mu}$, and in particular when $\epsilon = o(1)$ and $\epsilon\mu = o(1)$.

*I believe Janson's first inequality could be made more precise by involving higher order correlations like* $\Delta_3 = \sum_{i\sim j\sim k}\mathbb{Pr}[B_i \wedge B_j \wedge B_k]$.

Note that the second Janson inequality makes no reference to $\epsilon$. *Judging from its proof, it is not clear that it could benefit from higher order correlation information like* $\Delta_3$.

***Example*** 7.1.

When $\Delta \geq 2\mu$ the inequalities are useless, however, for $\Delta$ slightly less, it is improved by:

**Theorem 7.3** (The Extended Janson Inequality)**.** *Under the further assumption that* $\Delta \geq \mu$*:*

$$\mathbb{Pr}\big[\wedge_{i\in I}\overline{B_i}\big] \leq e^{-\frac{\mu^2}{2\Delta}}$$

The extended Janson Inequality is a "boosted" version of the second Janson Inequality.

Note that when applicable the extended Janson inequality gives significantly stronger bounds than the variance method for bounding indicators (Corollary 2.3), which achieves:

$$\mathbb{Pr}[\wedge_i\overline{B_i}] \leq \frac{\mathbb{Var}[X]}{\mathbb{E}[X]^2} \leq \frac{\mu+\Delta}{\mu^2}$$

In particular, when $\mu \to \infty$, $\mu \ll \Delta$, and $\gamma = \mu^2/\Delta \to \infty$, Chebyshev's upper bound on $\mathbb{Pr}[\wedge_i \overline{B_i}]$ is roughly $\gamma^{-1}$ while Janson's upper bound is roughly $e^{-\gamma}$.

(The above two theorems are stated in [NA00] Section 8.1 and proven in Section 8.2.)

Janson's Inequalities, in fact, apply to any collection of target events $B_i$ for which the following hold:

$$\mathbb{Pr}[B_i | \wedge_{j \in J} \overline{B_j}] \leq \mathbb{Pr}[B_i]$$

valid for all index sets $J \subset I$, $i \notin J$ and

$$\mathbb{Pr}\Big[B_i \Big| B_k \wedge \bigwedge_{j \in J} \overline{B_j}\Big] \leq \mathbb{Pr}[B_i | B_k]$$

valid for all index sets $J \subset I$, $i, k \notin J$. ***Have to verify:*** *Observe that these inequalities hold if we simply replace all $B_i$ with $C_i = \overline{B_i}$. In other words, Janson's inequalities can also be used to lower- and upper-bound $\mathbb{Pr}[\wedge_i B_i] = \mathbb{Pr}[\wedge_i \overline{C_i}]$.*

(These generalizations are given in [NA00] Section 8.7, and in fact they use the same proof of Section 8.2.)

## 7.3 Brun's Sieve

- Hidden parameter $n$ (for asymptotics)
- Events $B_1, \ldots, B_m$ with indicators $X = X_1 + \cdots + X_m$
- Let $S^{(r)} = \sum \mathbb{Pr}\big[B_{i_1} \wedge \cdots \wedge B_{i_r}\big]$

**Theorem 7.4** (Brun's Sieve)**.** *Suppose there is a constant $\mu$ such that for every fixed $r$:*

$$\mathbb{E}\big[X^{(r)}/r!\big] = S^{(r)} \to \mu^r/r!.$$

*Then for every $t$:*

$$\mathbb{Pr}[X = t] \to \frac{\mu^t}{t!} e^{-\mu}.$$

# 8 Bernoulli

- $B(n, x)$ – Bernoulli random variable with $n$ trials and probability of success $x$
- $\mathbb{Pr}[B(n, x) = j] = \binom{n}{j} x^j (1 - x)^{n-j}$
- $\mathbb{E}[B(n, x)] = nx$
- $\mathbb{Var}[B(n, x)] = nx(1 - x)$

# 9 Markov Chains

# 10 Linear Algebra

## 10.1 Hermitian (Symmetric) Matrices

A matrix $A$ is Hermitian if and only if $A = A^*$. An easy consequence of this is:

**Theorem 10.1** (Theorem 4.1.3 of [RH99])**.** *Let $A \in M_n$ be Hermitian. Then*

*(a) $x^*Ax$ is real for all $x \in \mathbb{C}^n$;*

*(b) All the eigenvalues of $A$ are real; and*

*(c) $S^*AS$ is Hermitian for all $S \in M_n$*

In view of this theorem, we adopt the convention of labeling the eigenvalues of a Hermitian matrix in ascending (non-decreasing) order:

$$\lambda_{\min} = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{n-1} \leq \lambda_n = \lambda_{\max}$$

**Theorem 10.2** (Rayleigh-Ritz, Theorem 4.2.2 of [RH99]). *Let $A \in M_n$ be Hermitian, and let the eigenvalues of $A$ be ordered as above. Then:*

$$\lambda_1 x^* x \leq x^* A x \leq \lambda_n x^* x \quad \textit{for all} \quad x \in \mathbb{C}^n$$

$$\lambda_{\max} = \lambda_n = \max_{x \neq 0} \frac{x^* A x}{x^* x} = \max_{x^* x = 1} x^* A x$$

$$\lambda_{\min} = \lambda_1 = \min_{x \neq 0} \frac{x^* A x}{x^* x} = \min_{x^* x = 1} x^* A x$$

## 10.2 Positive Semidefinite Matrices

**Theorem 10.3** (Theorem 7.2.10 of [RH99]). *Let $G \in M_k$ be the Gram matrix of the vectors $\{w_1, \ldots, w_k\} \subseteq \mathbb{C}^n$ with respect to a given inner product $\langle \cdot, \cdot \rangle$, and let $W = [w_1 \ w_2 \ \ldots \ w_k] \in M_{n,k}$. Then:*

*(a) $G$ is positice semidefinite;*

*(b) $G$ is non-singular if and only if the vectors $w_1, \ldots, w_k$ are independent;*

*(c) There exists a positive definite matrix $A \in M_n$ such that $G = W^* A W$; and*

*(d)* $\operatorname{rank}(G) = \operatorname{rank}(W) =$ *maximum number of independent vectors in the set $\{w_1, \ldots, w_k\}$.*

**Corollary 10.1** (Exercise of [RH99]). *Let $\langle x, y \rangle = y^* x$ be the usual Euclidean inner product. Then (using the notation from above) $A = I$ and the maximum number of independent vectors in $\{w_1, \ldots, w_k\}$ is exactly equal to* $\operatorname{rank}(G)$.

**Corollary 10.2** (Corollary 7.2.11 of [RH99]). *Let $A \in M_n$ be a given matrix. Then $A$ is positive semidefinite with rank $r \leq n$ if and only if there is a set of vectors $S = \{w_1, \ldots, w_k\} \subseteq \mathbb{C}^n$ containing exactly $r$ independent vectors such that $A$ is the Gram matrix of $S$ with respect to the Euclidean inner product.*

# 11 Pseudorandomness

## 11.1 Prime Numbers

Let $\pi(x)$ be the number of primes up to and including $x$, then:

**Theorem 11.1** (Prime Number Theorem)**.**

$$\pi(x) \sim \frac{x}{\ln x}$$

Let $p_n$ be the $n$-th prime, then:

**Corollary 11.1.**

$$p_n \sim n \ln n$$

## 11.2 Quadratic Residues

### 11.2.1 Basic Properties

*Quadratic residues in finite fields are a useful tool for building quasi- or pseudo-random objects, so we mention their basic properties.*

**Definition 11.1** (Legendre's Symbol)**.** *For $a \in \mathbb{F}_p$, define the Legendre Symbol as:*

$$\left(\frac{a}{p}\right) = \begin{cases} +1, & \text{if } a \text{ is a quadratic residue,} \\ -1, & \text{if } a \text{ is a non-resiude} \end{cases}$$

**Definition 11.2** (Quadratic Character)**.** *For $a \in \mathbb{F}_p$, define the quadratic character of $a$ as:*

$$\chi(a) = a^{(p-1)/2}$$

**Lemma 11.1.**

$$\chi(a) = \left(\frac{a}{p}\right)$$

**Lemma 11.2.** *In $\mathbb{F}_p$ there are exactly $(p-1)/2$ non-zero quadratic residues and $(p-1)/2$ non-zero non-residues.*

**Lemma 11.3.** *The number $-1$ is a quadratic residue of primes of the form $4k+1$ and a non-residue of primes of the form $4k+3$, i.e.:*

$$\left(\frac{-1}{p}\right) = (-1)^{(p-1)/2}$$

### 11.2.2 Pseudorandom Tournaments from Quadratic Residues

For a tournament $T$ on $n$ vertices and a permuation $\pi$ of the vertices $c(\pi, T)$ is defined to be the number of edges in $T$ that agree with the order of vertices enforced by $\pi$. Furthermore, the natural characteristic $c(T)$ of $T$ is defined as:

$$c(T) = \max_{\pi} c(\pi, T)$$

Clearly, $c(T) \geq \frac{1}{2}\binom{n}{2}$ for any tournament $T$. It can further be shown that $c(T) \geq \frac{1}{2}\binom{n}{2} + O(n^{3/2})$ for any tournament $T$. On the other hand, simple probabilistic arguments exhibit the existence of tournaments for which $c(T) \leq (1 + o(1))\frac{1}{2}\binom{n}{2}$.

Using the random-like structure of quadratic residues one can build explicit tournaments that have the properties of probabilistic ones. The techniques used are of general interest in building pseudo-random objects.

Let $p$ be any prime with $p \equiv 3 (\mathrm{mod}\, 4)$

# 12 Asymptotics and bounds

## 12.1 Taylor Series

$$f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(\alpha)}{k!}(x-\alpha)^k$$

$$f(x) = f(0) + \frac{f'(0)}{1!}x + \frac{f''(0)}{2!}x^2 + \frac{f'''(0)}{3!}x^3 + \cdots$$

**Common functions**

$$e^x = 1 + x + \cdots + \frac{x^n}{n!} + \cdots = \sum_{i=0}^{\infty} \frac{x^i}{i!}$$

$$\ln(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \cdots - \frac{x^k}{k} - \cdots$$

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots + (-1)^{k-1}\frac{x^{2k-1}}{(2k-1)!} + \cdots$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots + (-1)^k \frac{x^{2k}}{(2k)!} + \cdots$$

$$(1+x)^\alpha = \sum_{k=0}^{\infty} \binom{\alpha}{k} x^k, \quad \text{for } |x| < 1$$

## 12.2 Asymptotic Approximations

**Stirling's Formula:**

$$n! \approx \sqrt{2\pi n}\ \left(\frac{n}{e}\right)^n \left(1 + O\left(\frac{1}{n}\right)\right)$$

**Derangements**

**Definition 12.1.** *A derangement is a permutation $\pi \in S_n$ with no fixed points, i.e. no $i$ such that $\pi(i) = i$.*

**Lemma 12.1.** *The number of derangements in $S_n$ (also called the subfactorial function) is denoted by $!n$ and:*

$$!n = \left[\frac{n!}{e}\right],$$

*where $[\cdot]$ is the nearest integer function.*

**Asymptotics of Combinations:**
For fixed $k$:

$$\binom{n}{k} \approx n^k/k!$$

For $k = pn$, where $p \in (0, 1)$:

$$\binom{n}{pn} \approx 2^{n\left(H(p)+o(1)\right)}$$

Where $H(p) = -p \ln p - (1-p)\ln(1-p)$ is the entropy function.

**Some inequalities used in asymptotic reductions:**

- $\left(1 - \frac{1}{n}\right)^n \lessapprox e^{-1}$, approximation for $n \to \infty$, inequality for $n > 0$
- $\left(1 + \frac{1}{n}\right)^n \lessapprox e$, approximation for $n \to \infty$, inequality for $n > 0$
- $\ln(1-x) \lessapprox -x$, approximation for $x \to 0$, inequality for all $x$
- $1 - x \lessapprox e^{-x}$, approximation for $x \to 0$, inequality for all $x$
- $\ln(1+x) \lessapprox e^{x}$, approximation for $x \to 0$, inequality for all $x$
- $1 + x \lessapprox e^{x}$, approximation for $x \to 0$, inequality for all $x$
- $1 - e^{-x} \lessapprox x$, approximation for $x \to 0$, inequality for all $x$

## 12.3 Identities

**Basic:** Let $x_1, \ldots, x_n$ be algebraic variables:

$$\sum_{\substack{S \subseteq [n] \\ |S|=m}} \prod_{i \in S} x_i = \left(\sum_{i=1}^{n} x_i\right)^m$$

**Euler's Pentagonal Theorem:**

$$\prod_{k=1}^{\infty}(1-x^k) = 1 - x - x^2 + x^5 + x^7 - x^{12} - x^{15} + \cdots$$
$$= \sum_{k=-\infty}^{+\infty} (-1)^k x^{k(3k+1)/2}$$

### 12.4 Integration

$$\int e^{ax+b}dx = \frac{e^{ax+b}}{a}$$

## 13 ToDo

1. Downward-biased Random Walks with Restarts: Approximating 3-SAT
2. Tighter Chernoff bound from Janson's survey on Concentration
3. Minimax Principle in Linear Algebra
4. Minimax Principle in competitive algorithms/games analysis
5. Geometry: Relationship between $L_1$ and $L_2$
6. Markov: Coupling, Path Coupling, Relaxation Time

## References


[BR02] Béla Bollobás and Oliver Riordan. The diameter of a scale-free random graph. *Combinatorica*, 24:5–34, 2002.

[Jan02] Svante Janson. On concentration of probability. *Combinatorics, Probability and Computing*, 11, 2002.

[NA00] Joel Spencer Noga Alon. *The Probabilistic Method*. John Wiley and Sons, Inc., 2000.

[RH99] Charles Johnson Robert Horn. *Matrix Analysis*. Cambridge University Press, 1999.