Poissonization

Ayush Vora, Nathan Harms

September 2025

1 Chernoff Bound

The Chernoff bound is a pair of inequalities that bound a random variable using its moment generating function.

1.1 Obtaining the Bounds

Let X be a random variable, and $M_X(t)$ be the moment generating function (MGF) of X. We can start with Markov's inequality, and apply it to a function of X, specifically e^{tX} as shown:

$$\begin{aligned}\mathbb{P}[X \geq a] &= \mathbb{P}[e^{tX} \geq e^{ta}] \text{ for } t > 0 \\ &\leq \mathbb{E}[e^{tX}]/e^{ta} \text{ by Markov's Inequality} \\ &= M_X(t)e^{-ta} \text{ by definition of MGF}\end{aligned}$$

Similarly, we see that:

$$\begin{aligned}\mathbb{P}[X \leq a] &= \mathbb{P}[e^{tX} \geq e^{ta}] \text{ for } t < 0 \\ &\leq \mathbb{E}[e^{tX}]/e^{ta} \\ &= M_X(t)e^{-ta}\end{aligned}$$

Because we have these bounds for all values of t, we can choose the tightest bound by choosing t such that $M_X(t)e^{-ta}$ is minimized.

1.2 Bounds for the "Gambler" Random Variable

Suppose a gambler plays n rounds of a game, where each round is independent of each other. Say there is an equal chance of either winning or losing \$1. Let the gambler's winnings of round i be X_i. Then, the total winnings would be $X = \sum_{i=1}^{n} X_i$. We can generate a bound on the winnings. First, note that $M_{X_i}(t) \leq e^{t^2/2}$. We can start by finding $M_X(t)$:

$$
\begin{aligned}
M_X(t) &= \mathbb{E}[e^{tX}] \\
&= \mathbb{E}[e^{t\sum_i X_i}] \\
&= \mathbb{E}\left[\prod_{i=1}^{n} e^{tX_i}\right] \\
&= \prod_{i=1}^{n} \mathbb{E}[e^{tX_i}] \text{ by independence} \\
&\leq \prod_{i=1}^{n} e^{t^2/2} \text{ using } M_{X_i}(t) \\
&= e^{nt^2/2}
\end{aligned}
$$

We can now use this in finding our bound for X:

$$
\begin{aligned}
\mathbb{P}[X \geq a] &\leq M_X(t)e^{-ta} \\
&\leq e^{nt^2/2}e^{-ta} \\
&= e^{nt^2/2-ta}
\end{aligned}
$$

The minimum of $e^{nt^2/2-ta}$, or, equivalently, the minimum of $nt^2/2-ta$ can be found using calculus. We find that it is minimized when $t = a/n$. Substituting these values above, we get that:

$$\mathbb{P}[X \geq a] \leq e^{n(a/n)^2/2-(a/n)a} = e^{-a^2/2n}$$

Using this, we can go back to our example. Say our gambler wants to play 10 games, and wants to make a profit of at least \$6. Then the probability that occurs is:

$$\mathbb{P}[X \geq 6] \leq e^{-6^2/(2\times 10)} = e^{-9/5} \approx 0.1653$$

The exact probability is $56/1024 \approx 0.0547$.

1.3 Exercises

1. Show that our bound in section 1.2 is stronger than that of Chebyshev's inequality.

2. (a) Find the upper Chernoff bound for the Poisson variable $Y \sim \text{Poi}(\lambda)$. You may use that the $M_Y(t) = e^{\lambda(e^t-1)}$.

   (b) Show that this bound is stronger than that of Markov's inequality.

3. In section 1.2, we state that $M_{X_i}(t) \leq e^{t^2/2}$. Show that this is the case. (Hint: use the power series definition of e^x.)

2 Poissonization

The following is heavily inspired by Maryam Aliakbarpour's lecture notes in "Hidden Gems of Sublinear Algorithms".

2.1 Motivation

Suppose we have m balls, and each ball must go in one of n bins. Let the probability that a ball goes into bin i be p_i. We want to look into how many balls are in each bin. Let us define the number of balls in bin i as:

$$X_i \sim \mathrm{Bin}(m, p_i)$$

However, these X_i can sometimes lead to issues when doing further analysis. This is because the X_i's are not independent of each other (shown below). **Poissonization** is a technique that can translate these dependent random variables into independent Poisson random variables, assuming we are okay with getting an approximate answer.

Theorem 2.1. The outcomes X_i are dependent on each other.

Proof. First, notice that because we have exactly m balls, and every ball is in a bin, we have that:

$$\sum_{i=1}^{n} X_i = m$$

and that m is constant. Because the variance of a constant is 0, it follows:

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \mathrm{Var}(m) = 0$$

Now, for the sake of contradiction, assume that X_i are independent. Then from the properties of independent random variables and binomial random variables, we have that:

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}\left(X_i\right) = \sum_{i=1}^{n} m p_i (1 - p_i) \neq 0$$

This is a contradiction. □

2.2 Benefits of the Poisson Distribution

The Poisson distribution can help us solve this issue! To start off, recall that if $Y_1 \sim \mathrm{Poi}(\lambda_1)$ and $Y_2 \sim \mathrm{Poi}(\lambda_2)$, then $Y_1 + Y_2 \sim \mathrm{Poi}(\lambda_1 + \lambda_2)$.

The main goal of Poissonization is to generate a different set of independent outcomes $Y_i \sim \mathrm{Poi}(m p_i)$ as a "replacement" for the dependent X_i's. We will show how we generate these Y_i in section 2.3. First, we will show that our new random variables, conditioned on a set sum is identical to original distribution.

Let $k_1, ..., k_n \in \mathbb{N}$ such that $\sum_{i=1}^n k_i = k$.

Theorem 2.2. The joint distribution of $(Y_1, ..., Y_n)$ conditioned on $\sum_{i=1}^n Y_i = k$ is identical to the joint distribution of $(X_1, ..., X_n)$.

To show that they are identical, it is sufficient to show that their probability mass functions (PMF) are equal.

Proof. Because the joint distribution of $(X_1, ..., X_n)$ is multinomial, the PMF of it is:

$$p_{X_1,...,X_n}(k_1, ..., k_n) = \frac{k!}{n^k \prod_{i=1}^n k_i!}$$

The joint conditional PMF of $(Y_1, ..., Y_n)$ is:

$$\begin{aligned}
p_{Y_1,...,Y_n \mid \sum Y_i = k}(k_1, ..., k_n) &= \frac{p_{Y_1,...,Y_n}(k_1, ..., k_n)}{\mathbb{P}[\sum_i Y_i = k]} \\
&= \frac{\prod_{i=1}^n p_{Y_i}(k_i)}{\mathbb{P}[\sum Y_i = k]} \text{ by independence of random variables} \\
&= \frac{\prod_{i=1}^n p_{Y_i}(k_i)}{p_{\sum Y_i}(k)} \text{ as } \sum_{i=1}^n Y_i \sim \text{Poi}(m) \\
&= \frac{k!}{e^{-m} m^k} \prod_{i=1}^n \frac{e^{-m/n}(m/n)^{k_i}}{k_i!} \text{ by Poisson PMF} \\
&= \left(\frac{k!}{e^{-m} m^k}\right) \left(e^{-m/n}\right)^n \left(\frac{m}{n}\right)^k \left(\frac{1}{\prod_{i=1}^n k_i!}\right) \\
&= \frac{k!}{n^k \prod_{i=1}^n k_i!} \\
&= p_{X_1,...,X_n}(k_1, ..., k_n)
\end{aligned}$$

as needed.

□

This shows us that we can successfully approximate the X_i's using the Y_i's. This leaves us one main question: How do we generate these Y_i's?

2.3 Generating Independent Outcomes

Let P be a probability distribution over $[n]$. Let $p_i = \mathbb{P}_{S \sim P}[S = i]$. Assume that we draw m samples $S_1, ..., S_m \sim P$. These samples can be thought of as the location of a ball from our example above. Let X_i be a random variable indicating how many samples S_j take the value i. Specifically:

$$X_i = \sum_{j=1}^{m} \mathbb{1}[S_j = i] \sim \mathrm{Bin}(m, p_i)$$

Similar to how we established above, these X_i are not independent. To generate independence, do the following:

1. Let $\hat{m} \sim \mathrm{Poi}(m)$. Our value $\hat{m}$ will be an approximation for m. Draw this value.
2. Draw $\hat{m}$ samples $S_1, ..., S_{\hat{m}}$ from P.
3. Let $Y_i = \sum_{j=1}^{\hat{m}} \mathbb{1}[S_j = i]$.

Following the process above, we can generate Y_i that are independent Poisson random variables. To use the theorem above, we only require to show that $Y_i \sim \mathrm{Poi}(mp_i)$

Theorem 2.3. $Y_i \sim \mathrm{Poi}(mp_i)$

Proof. It is sufficient to show that $\mathbb{P}[Y_i = k] = \frac{e^{mp_i}(mp_i)^k}{k!}$, the PMF of the Poisson distribution.

$$\begin{aligned}
\mathbb{P}[Y_i = k] &= \sum_{t=0}^{n} \mathbb{P}[\hat{m} = t]\mathbb{P}[Y_i = k|\hat{m} = t] \text{ by law of total probability} \\
&= \sum_{t=0}^{k-1} \mathbb{P}[\hat{m} = t]\mathbb{P}[Y_i = k|\hat{m} = t] + \sum_{t=k}^{n} \mathbb{P}[\hat{m} = t]\mathbb{P}[Y_i = k|\hat{m} = t] \\
&= \sum_{t=k}^{n} \mathbb{P}[\hat{m} = t]\mathbb{P}[Y_i = k|\hat{m} = t] \text{ See note (1)} \\
&= \sum_{t=k}^{n} \left(\frac{e^{-m}m^t}{t!}\right)\left(\binom{t}{k} p_i^k (1-p_i)^{t-k}\right) \text{ by Poisson and Binomial PMF.} \\
&= \frac{e^{-m}p_i^k}{k!} \sum_{t=k}^{n} \frac{m^t(1-p_i)^{t-k}}{(t-k)!} \\
&= \frac{e^{-m}p_i^k m^k}{k!} \sum_{t=k}^{n} \frac{m^{t-k}(1-p_i)^{t-k}}{(t-k)!} \\
&= \frac{e^{-m}p_i^k m^k e^{m(1-p_i)}}{k!} \text{ By power series definition of } e^x \\
&= \frac{e^{mp_i}(mp_i)^k}{k!}
\end{aligned}$$

as needed.

(1): We can think of t as the number of balls that we have, and k as the number of balls in bin i. If we consider the case where we have $t < k$, that would be saying that we have more balls in bin i than we have total balls. The probability of this happening is 0. Hence, $\sum_{t=0}^{k-1} \mathbb{P}[\hat{m} = t]\mathbb{P}[Y_i = k|\hat{m} = t] = 0$. $\square$

2.4 Use in Randomized Algorithms

This situation of counting the number of instances given a sample shows up often in randomized algorithms. Doing statistical analysis on these may be difficult due to the dependence between the random variables. However, Poissonization solves this problem for us. In particular, if we have an algorithm $\mathcal{A}(P, m, \delta)$ that does the following:

1. Pull m samples from distribution P. Call these samples $S_1, ...S_m$.
2. Let $X_i = \sum_{j=1}^{m} \mathbb{1}(S_j = i)$.
3. Run function $\mathcal{A}^*(X_1, ..., X_n)$ with $\mathbb{P}(\text{Fail}) \leq \delta$. We can think of $\mathcal{A}^*$ as the "remaining code" of our algorithm.

We can modify this algorithm to ensure that the variables for $\mathcal{A}^*$ are independent.

Theorem 2.4. If there exists an algorithm $\mathcal{A}(P, m, \delta)$ that uses m samples from P where $\mathbb{P}(\text{Fail}) \leq \delta$, then there exists an algorithm $\mathcal{A}'(P, m, \delta)$ that uses $\text{Poi}(2m)$ samples from P where $\mathbb{P}(\text{Fail}) \leq 2\delta$ given that $m \geq \frac{\ln(\delta)}{\ln(2)-1}$.

Proof. Consider the following algorithm $\mathcal{A}'(P, m, \delta)$:

1. Let $m' \sim \text{Poi}(2m)$.
2. If $m' < m$, fail.
3. Pull m' samples from P, call them $S_1, ..., S_{m'}$.
4. Let $Y_i = \sum_{j=1}^{m'} \mathbb{1}(S_j = i)$.
5. Run function $\mathcal{A}^*(Y_1, ..., Y_n)$ with $\mathbb{P}(\text{Fail}) \leq \delta$.

This algorithm fails when either $\mathcal{A}^*$ fails, or $m' < m$. We will show that this probability is less than 2δ.

$$
\begin{aligned}
\mathbb{P}(\mathcal{A}^* \text{ fails } \cup m' < m) &\leq \mathbb{P}(\mathcal{A}^* \text{ fails}) + \mathbb{P}(m' < m) \text{ by Union Bound} \\
&= \delta + \mathbb{P}(m' < m) \\
&\leq \delta + \mathbb{P}(m' \leq m) \\
&\leq \delta + (m/2m)^{-m} e^{m-2m} \text{ by Chernoff Bound} \\
&= \delta + (e/2)^{-m} \\
&\leq \delta + \delta \text{ for } m \geq \frac{\ln(\delta)}{\ln(2) - 1} \\
&= 2\delta
\end{aligned}
$$

□

We can also do this proof in the other direction, where we have an algorithm $\mathcal{B}(P, m, \delta)$ that uses $m' \sim \text{Poi}(m)$ samples with failure rate δ. We can construct an algorithm $\mathcal{B}'(P, m, \delta)$ that uses $2m$ samples with failure rate 2δ.

Theorem 2.5. If there exists an algorithm $\mathcal{B}(P, m, \delta)$ that uses $m' \sim \text{Poi}(m)$ samples from P where $\mathbb{P}(\text{Fail}) \leq \delta$, then there exists an algorithm $\mathcal{B}'(P, m, \delta)$ that uses $2m$ samples from P where $\mathbb{P}(\text{Fail}) \leq 2\delta$ given that $m \geq \frac{\ln(\delta)}{1-\ln(4)}$.

Proof. We can construct it similar to the algorithms in the above proof. We will focus on showing the error probability.

$$\begin{aligned}
\mathbb{P}(B^* \text{ fails } \cup m' > 2m) &\leq \mathbb{P}(B^* \text{ fails}) + \mathbb{P}(m' > 2m) \text{ by Union Bound} \\
&= \delta + \mathbb{P}(m' > 2m) \\
&\leq \delta + \mathbb{P}(m' \geq 2m) \\
&\leq \delta + (2m/m)^{-2m} e^{2m-m} \text{ by Chernoff Bound} \\
&= \delta + (e/4)^m \\
&\leq \delta + \delta \text{ for } m \geq \frac{\ln(\delta)}{1-\ln(4)} \\
&= 2\delta
\end{aligned}$$

□

2.5 Exercises

1. In the fourth equality of Theorem 2.3, we assume that Y_i conditioned on $\hat{m}$ follows $\text{Bin}(t, p_i)$. Explain why we assume so.