Lecture 5

Poisson Approximation for Binomial Distribution

Definitions

You are probably familiar with the binomial, Multinomial, and Poisson distributions. Here we provide their definitions.

Binomial distribution: A *Binomial* random variable X with parameters (m, q) is given by

$$X \sim \mathbf{Bin}(m, q), \quad \mathbf{Pr}[X = k] = \binom{m}{k} q^k (1-q)^{m-k},$$

for $k = 0, 1, \ldots, m$. This distribution arises when we count the number of successes in a fixed number of independent trials, where each trial has the same probability of success.

Multinomial distribution: The Multinomial distribution generalizes the binomial distribution to scenarios with more than two outcomes. We have m independent trials, where each trial can result in one of n possible outcomes with probabilities $p_1, p_2, ..., p_n$ (where $\sum_{i=1}^n p_i = 1$). Let X_i be the number of times outcome i occurs in the m trials. Then, the random vector $(X_1, X_2, ..., X_n)$ follows a multinomial distribution. This is equivalent to drawing m samples from a distribution p over $[n]$, where X_i represents the frequency of element i in the sample set.

Poisson distribution: *Poisson* random variable Y with parameter λ is given by

$$Y \sim \mathbf{Poi}(\lambda), \quad \mathbf{Pr}[Y = k] = \frac{e^{-\lambda}\lambda^k}{k!},$$

for $k = 0, 1, 2, \ldots$. This distribution arises when we count the number of events that occur over a fixed period of time or space, given a constant rate of occurrence and independence between occurrences.

A well-known approximation result states that a binomial distribution can be approximated with a Poisson distribution under certain conditions. This approximation provides a useful simplification in probability applications. In this lecture, we will explore the connections between these two distributions from various aspects and prove some interesting properties about these two distributions.

Convergence in the limit

The binomial distribution converges to the Poisson distribution in the limit. Specifically, when the number of trials m is large and the success probability q is small, the binomial distribution with parameters m and q converges to the Poisson distribution. This concept is formalized in the following theorem.

Theorem 1. *Let $X_m \sim \mathbf{Bin}(m, q)$, where q depends on m in such a way that $m\,q = \lambda$ (a constant). Then for any fixed $k \geq 0$,*

$$\lim_{m\to\infty} \mathbf{Pr}[X_m = k] \;=\; \frac{e^{-\lambda}\,\lambda^k}{k!} = \mathbf{Pr}_{Y\sim\mathbf{Poi}(\lambda)}[Y = k]\,.$$

Sketch of Proof. Here is an informal proof sketch. The following shows that these probabilities are roughly the same.

$$\begin{aligned}
\mathbf{Pr}[X_m = k] &= \binom{m}{k} q^k\,(1-q)^{m-k} \\
&= \frac{m!}{k!\,(m-k)!}\, q^k\,(1-q)^{m-k} \\
&= \frac{m(m-1)\cdots(m-k+1)\,(m-k)!}{k!\,(m-k)!}\, q^k\,(1-q)^{m-k} \\
&\approx \frac{(m\,q)^k}{k!}\,(1-q)^m \qquad\qquad (m \approx m-i \text{ when } m \text{ is large compared to } i.) \\
&\approx \frac{(m\,q)^k}{k!}\,e^{-m\,q} \\
&= \frac{\lambda^k\,e^{-\lambda}}{k!} = \mathbf{Pr}_{Y\sim\mathbf{Poi}(\lambda)}[Y = k]\,,
\end{aligned}$$

where we used the approximation $(1-q)^m \approx e^{-mq}$ for large m, since:

$$\lim_{m\to\infty} \frac{(1-\lambda/m)^m}{e^{-\lambda}} = 1\,.$$

For a more rigorous proof, see Theorem 5.5 and its proof in [MU05].

From Poissons to Multinomials

A nice property of Poisson distributions is that they are closed under addition. That is, suppose we have two independent Poisson random variables: $Y_1 \sim \mathbf{Poi}(\lambda_1)$ and $Y_2 \sim \mathbf{Poi}(\lambda_2)$.

Then, the sum of the two random variables is also a Poisson random variable, with the parameter being the sum of the individual parameters:

$$Y_1 + Y_2 \;\sim\; \mathbf{Poi}(\lambda_1 + \lambda_2).$$

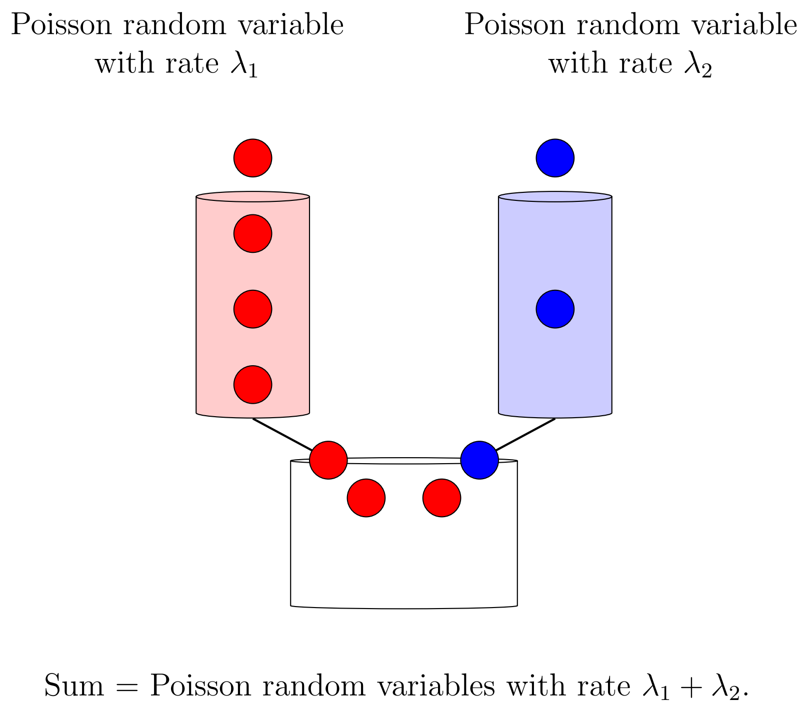

Figure 1: Imagine two pipes pouring streams of red and blue balls into a pool. Each pipe follows an independent Poisson process with rates λ_1 and λ_2, respectively. The total number of balls in the pool also follows a Poisson process, with a combined rate of $\lambda_1 + \lambda_2$. If the pool contains exactly m balls, the number of red balls follows a binomial distribution: $\mathbf{Bin}\left(m, \frac{\lambda_1}{\lambda_1+\lambda_2}\right)$.

An interesting question is: if we fix the sum of two independent Poisson random variables (i.e., $Y_1 + Y_2 = m$), what would be the contribution of the first versus the second? It turns out that given $Y_1 + Y_2 = m$, the distribution of Y_1 is a binomial distribution: $\mathbf{Bin}\Big(m, \frac{\lambda_1}{\lambda_1+\lambda_2}\Big)$. More formally, we have the following lemma which shows this interesting connection between the Poisson and binomial distributions:

Lemma 1. *Suppose we have two independent Poisson random variables:* $Y_1 \sim \mathbf{Poi}(\lambda_1)$ *and* $Y_2 \sim \mathbf{Poi}(\lambda_2)$*. Then, condition of* $Y_1 + Y_2 = m$*, for any* $k \in \{0, 1, \ldots, m\}$*, we have:*

$$\mathbf{Pr}\big[Y_1 = k \;\big|\; Y_1 + Y_2 = m\big] \;=\; \mathbf{Pr}[X = k], \quad \textit{for } X \sim \mathbf{Bin}\Big(m, \frac{\lambda_1}{\lambda_1 + \lambda_2}\Big).$$

Proof. We claim that, conditioning on $Y_1 + Y_2 = m$, the distribution of Y_1 is exactly the same as a $\mathbf{Bin}\big(m, \frac{\lambda_1}{\lambda_1+\lambda_2}\big)$. Indeed, for $k \leq m$, we have:

$$\begin{aligned}
\mathbf{Pr}\big[Y_1 = k \,\big|\, Y_1 + Y_2 = m\big] &= \frac{\mathbf{Pr}[Y_1 = k \text{ and } Y_1 + Y_2 = m]}{\mathbf{Pr}[Y_1 + Y_2 = m]} \\
&= \frac{\mathbf{Pr}[Y_1 = k \text{ and } Y_2 = m - k]}{\mathbf{Pr}_{Y \sim \mathbf{Poi}(\lambda_1+\lambda_2)}[Y = m]} \\
&= \frac{\dfrac{e^{-\lambda_1}\,\lambda_1^k}{k!} \cdot \dfrac{e^{-\lambda_2}\,\lambda_2^{m-k}}{(m-k)!}}{\dfrac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^m}{m!}} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}\,\lambda_1^k\,\lambda_2^{m-k}}{k!\,(m-k)!} \cdot \frac{m!}{e^{-(\lambda_1+\lambda_2)}\,(\lambda_1+\lambda_2)^m} \\
&= \binom{m}{k}\left(\tfrac{\lambda_1}{\lambda_1+\lambda_2}\right)^k\left(\tfrac{\lambda_2}{\lambda_1+\lambda_2}\right)^{m-k}.
\end{aligned}$$

This is exactly the PMF of a $\mathbf{Bin}\big(m, \frac{\lambda_1}{\lambda_1+\lambda_2}\big)$. Hence, we have:

$$\mathbf{Pr}\big[Y_1 = k \,\big|\, Y_1 + Y_2 = m\big] \;=\; \mathbf{Pr}[X = k], \quad X \sim \mathbf{Bin}\Big(m, \tfrac{\lambda_1}{\lambda_1+\lambda_2}\Big).$$

□

By induction, one can verify that similar statement is true when we have n Poisson random variables: If $Y_i \sim \mathbf{Poi}(\lambda_i)$ for all $i \in [n]$, then we have:

$$\sum_{i=1}^{n} Y_i \sim \mathbf{Poi}\left(\sum_{i=1}^{n} \lambda_i\right).$$

Moreover, condition on that the sum is m, for each Y_i and $k \in \{0, 1, \ldots, m\}$, we have :

$$\mathbf{Pr}\left[Y_i = k \,\middle|\, \sum_{i=1}^{n} Y_i = m\right] \;=\; \mathbf{Pr}[X = k], \quad X \sim \mathbf{Bin}\Big(m, \tfrac{\lambda_i}{\sum_{i=1}^{n}\lambda_i}\Big).$$

From Multinomials to Poissons: Poissonization Method

Consider a distribution p over $[n]$. Suppose we draw m samples from p. Let X_i be the frequency of element $i \in [n]$. We stated earlier that X_i's are coming from a multinomial distribution. We can also view each X_i individually as a binomial random variable:

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

Note that X_i's here are not independent since we know their sum is fixed: $\sum_i X_i = m$. This is an issue if we want to analyze a statistic involving X_i, as we saw in the case of number collisions in the previous lecture, where the analysis of the variance gets complicated when random variables are not independent. In such cases, one applies a "Poissonization" trick to handle the dependency more easily - replacing a fixed-sample-size model by an approximate Poisson model, for which the frequency of elements becomes a Poisson random variable and independent of the rest of the frequencies, making certain calculations more tractable.

Based on our intuition from the previous part, in this case, we know we need to make the total number of samples a Poisson random variable. Let's try this approach and make the number of samples, instead of m, a random variable $\widehat{m} \sim \mathbf{Poi}(m)$:

- Draw a (random) number $\widehat{m} \sim \mathbf{Poi}(m)$.
- Then, draw $\widehat{m}$ i.i.d. samples $s_1, \ldots, s_{\widehat{m}}$ from p.
- For each $i \in [n]$, let
$$Y_i = \#\{\text{indices } j : s_j = i\}.$$

We claim each Y_i is distributed as $\mathbf{Poi}(m\, p_i)$, and moreover the Y_i's are independent. Compare this to the standard binomial model: $X_i \sim \mathbf{Bin}(m,\ p_i)$, where $\sum_{i=1}^n X_i = m$ is fixed.

Lemma 2. *If Y_i's are generated according to the process we described earlier, then we have:*

$$Y_i \sim \mathrm{Poi}(m\, p_i)\,,$$

and they are independent of each other.

Proof. Note that if we fix $\widehat{m}$, Y_i's form a random variable from a multinomial distribution. Therefore, we obtain:

$$\begin{aligned}\mathbf{Pr}\big[Y_i = k\big] &= \sum_{t=0}^{\infty} \mathbf{Pr}\big[Y_i = k \;\big|\; \widehat{m} = t\big]\ \mathbf{Pr}\big[\widehat{m} = t\big] && \text{(By law of total probability)}\\ &= \sum_{t=k}^{\infty} \binom{t}{k} p_i^k \,(1-p_i)^{t-k}\, e^{-m}\, \frac{m^t}{t!}\,.\end{aligned}$$

Note that it is impossible to have $Y_i > \widehat{m}$. Thus, the first k terms in the above sum are zero. We continue our calculation by expanding the binomial coefficient:

$$\begin{aligned}
\mathbf{Pr}\big[Y_i = k\big] &= p_i^k\, e^{-m} \sum_{t=k}^{\infty} \frac{m^t\,(1-p_i)^{\,t-k}}{k!\,(t-k)!} \\
&= \frac{p_i^k}{k!}\, e^{-m} \sum_{u=0}^{\infty} \frac{m^{k+u}\,(1-p_i)^{\,u}}{u!} && \text{(Change of variable to } u \coloneqq t-k\text{)} \\
&= \frac{p_i^k\, m^k}{k!}\, e^{-m} \sum_{u=0}^{\infty} \frac{\big[m(1-p_i)\big]^u}{u!} \\
&= \frac{p_i^k\, m^k}{k!}\, e^{-m} e^{m(1-p_i)} && \text{(Since } \textstyle\sum_{n=0}^{\infty} \frac{x^n}{n!} = e^x \text{ for all } x \in \mathbb{R}\text{)} \\
&= \frac{(p_i\, m)^k}{k!}\, e^{-p_i m},
\end{aligned}$$

which is exactly the PMF of a $\mathbf{Poi}(p_i\, m)$.

Moreover, a closer look at the above derivation shows that $\mathbf{Pr}\big[Y_i = k\big]$ involves no terms from Y_j ($j \neq i$), meaning the probability of $Y_i = k$ stays the same regardless of all other Y_j's. One can make a similar calculation for each Y_j. Therefore, one can imply that the Y_i's are independent $\mathbf{Poi}(p_i\, m)$ variables. □

Poissonization and De-Poissonization

One way to leverage the Poisson approximation is to treat binomial random variables as if they were Poisson. This allows us to exploit the independence of Poisson variables and simplify calculations in the "Poisson world". We can then translate the results back to the binomial world. This technique is particularly useful for the classical "balls and bins" problem.

To illustrate, consider throwing m balls into n bins uniformly at random. For each bin i, the number of balls in that bin, denoted by X_i, follows a binomial distribution:

$$X_i \;\sim\; \mathbf{Bin}\Big(m,\, \tfrac{1}{n}\Big).$$

We can approximate X_i with a Poisson random variable Y_i:

$$Y_i \;\sim\; \mathbf{Poi}\Big(\tfrac{m}{n}\Big).$$

Both X_i and Y_i have the same mean, m/n, but their distributions differ. The strategy is to perform calculations in the Poisson world, treating X_i as if it were Y_i, and then translate the results back to the binomial world. This is often easier due to the independence and other convenient properties of Poisson variables. The following theorems provide bounds for translating results between the two worlds.

Theorem 2. *Let $f(x_1, \ldots, x_n)$ be a nonnegative function. For the balls-and-bins setup, the* *expectation in the binomial world is bounded by the expectation in the Poisson world as follows:*

$$\mathbf{E}[f(X_1,\dots,X_n)] \;\leq\; e\sqrt{m}\,\mathbf{E}[f(Y_1,\dots,Y_n)]$$

Corollary 3. *If an event $\mathcal{A}$ occurs with probability q in the Poisson world, its probability in the binomial world is at most*

$$q\,e\sqrt{m}\,.$$

Proof. Define an indicator function f that equals 1 if event $\mathcal{A}$ occurs and 0 otherwise. The expectations of f in the binomial and Poisson worlds correspond to the probabilities of $\mathcal{A}$ in those worlds. Applying Theorem 1 yields the desired bound. □

Theorem 3. *Let $f(x_1,\dots,x_n)$ be a nonnegative function whose expectation is monotonic in the number of balls ($m = \sum_{i=1}^{n} X_i$). Then,*

$$\mathbf{E}[f(X_1,\dots,X_n)] \;\leq\; 2\,\mathbf{E}[f(Y_1,\dots,Y_n)]\,.$$

This theorem offers a tighter bound than Theorem 2 when the expectation of the function is monotonic.

Corollary 4. *Let $\mathcal{A}$ be an event whose probability is either monotonically increasing or monotonically decreasing as we vary the number of balls m. If $\mathbf{Pr}[\mathcal{A}$ in Poisson world$] \leq q$, then $\mathbf{Pr}[\mathcal{A}$ in binomial world$] \leq 2\,q$.*

Here, "monotonically increasing" means that adding more balls can only make $\mathcal{A}$ more likely. And, similarly for "monotonically decreasing" means that adding more balls can only make $\mathcal{A}$ less likely.

For the proof of these theorems, see Section 5.4 of [MU05].

Bibliographic note:

The content of this lecture was drawn from Section 5.4 of our reference book [MU05].

References

[MU05] Michael Mitzenmacher and Eli Upfal. *Probability and Computing: Randomized Algorithms and Probabilistic Analysis.* Cambridge University Press, New York, NY, USA, 2005.