AI2613 随机过程 2020–2021 春季学期

Lecture 8 – Poisson Process (II)

2021 年 4 月 19 日

Lecturer: 张驰豪 *Scribe*: 杨宽

Today we are going to talk about some interesting properties and applications of the Poisson process. The Poisson process has three well-known transformations: thinning, superposition and conditioning. We will discuss thinning and conditioning today, and leave superposition in the exercise. After the two transformations we will introduce the Poisson approximation.

1 Thinning

We first review the definition of Poisson process. A Poisson process with rate λ is a sequence of random variables $N(t)_{t\geq 0}$ s.t. (i) $N(0)=0$, (ii) $N(t+s)-N(s)\sim \mathrm{Pois}(\lambda t)$, and (iii) $N(t)$ has independent increments. We also introduced another constructive definition in the last lecture. Let $\tau_i \sim \mathrm{Exponential}(\lambda)$ and $T_n=\sum_{i=1}^n \tau_i$. Then $N(t)=\max\{n : T_n \leq t\}$ is a Poisson process.

Now consider the example of customers coming into the restaurant. Sometimes we have a more detailed characterization of customers, such as the gender. We associate an independent and identically distributed (i.i.d.) random variable Y_i with each arrival, and then use the value of Y_i to separate the Poisson process into several.

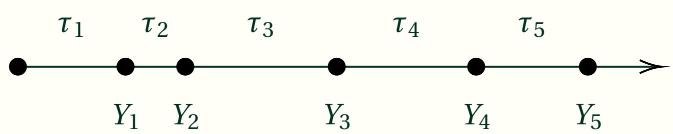

Figure 1: Poisson process with random variables associated with arrivals

Formally, suppose that $Y_i \in \mathbb{N}$ and are i.i.d. random variables. Let $p_j = \mathbf{Pr}[Y_i = j]$. For all $j \in \mathbb{N}$ (or $\mathrm{Range}(Y_i)$), let $N_j(t)$ denote the number of arrivals that have arrived by time t with exactly value j. Then $\{N_j(t)\}$ is called the *thinning* of a Poisson process.

The following properties for thinning might be a bit surprising.

Theorem 1. *$N_j(t)$ are independent rate $p_j\lambda$ Poisson processes. Namely,*

1. *For all j, $N_j(t)$ is a Poisson process with rate $p_j\lambda$;*

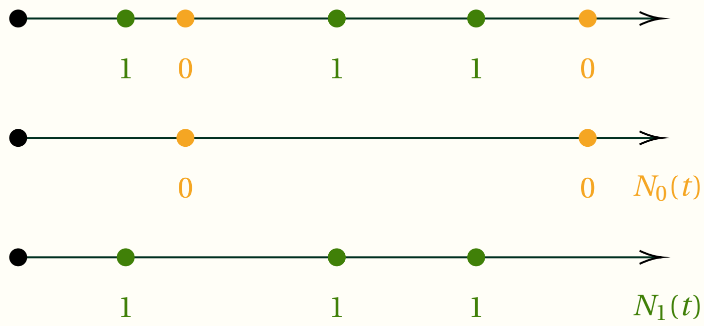

Figure 2: Thinning

2. *All $N_j(t)$ are mutually independent.*

Remark. Intuitively, we probably can understand that the resulting processes are Poissons. But the independence is a real "surprise". The following example explains why this lemma is surprising. Assume that the customers coming into a restaurant is a Poisson process, and each customer is male or female independently with probability 1/2 and 1/2 respectively. In fact we can assume that we flip coins to determine whether arriving customers are male or female. So intuitively one might think that a large number of men (such as 40) arriving in one hour would indicates a large volume of business and hence a larger than normal number of women arriving. However this theorem tells us that the number of men arriving and the number of women arriving are independent.

Proof. For convenience we assume that $Y_i \in \{0,1\}$. Then the following calculation concludes the independence and the distribution of $N_j(t)$ at the same time.

$$\begin{aligned}
\mathbf{Pr}[N_0(t) = j \wedge N_1(t) = k] &= \mathbf{Pr}[N_0(t) = j \wedge N_0(t) + N_1(t) = j+k] \\
&= \mathbf{Pr}[N_0(t) + N_1(t) = j+k] \cdot \mathbf{Pr}[N_0(t) = j \mid N_0(t) + N_1(t) = j+k] \\
&= \mathrm{e}^{-\lambda t} \cdot \frac{(\lambda t)^{j+k}}{(j+k)!} \cdot \binom{j+k}{j} \cdot p_0^j \cdot p_1^k \\
&= \mathrm{e}^{-\lambda t} \cdot \frac{(\lambda t)^{j+k}}{j! \cdot k!} \cdot p_0^j \cdot p_1^k \\
&= \mathrm{e}^{-p_0 \lambda t} \cdot \frac{(p_0 \lambda t)^j}{j!} \cdot \mathrm{e}^{-p_1 \lambda t} \cdot \frac{(p_1 \lambda t)^k}{k!}.
\end{aligned}$$ □

Example 1. Customers come into a restaurant at times of a Poisson process with rate 10 per hour. Suppose that 40% of customers are male and 60% of customers are female. What is the distribution of the number of male customers in 2 hours?

By independence the number of customers has a Poisson distribution with rate 8.

Example 2 (Maximum Likelihood Estimation). Now we consider the *maximum likelihood estimation* of the Poisson process. Suppose there are two editors reading a book of 300 pages. Editor A finds 100 typos in the book, and editor B finds 120 typos, 80 of which are in common. Suppose that the author's typos follow a Poisson process with some unknown rate λ per page, while the two editors catch typos with unknown probabilities of success p_A and p_B respectively. Our goal is to determine λ, p_A and p_B.

Clearly, there are four types of typos.

1. Type 1: neither of editors found, with probability $q_1 \triangleq (1-p_A)(1-p_B)$;
2. Type 2: only editor A found, with probability $q_2 \triangleq p_A(1-p_B)$;
3. Type 3: only editor B found, with probability $q_3 \triangleq p_B(1-p_A)$;
4. Type 4: both of editors found, with probability $q_4 \triangleq p_A p_B$.

So the occurrence of type-i typos follows an independent Poisson process with rate $q_i\lambda$.

Let $\mathcal{A}$ and $\mathcal{B}$ be the sets of typos that editors A and B found. We claim that by the maximum likelihood estimation we have $|\mathcal{A}\setminus(\mathcal{A}\cap\mathcal{B})| = 300q_2\lambda$, $|\mathcal{B}\setminus(\mathcal{A}\cap\mathcal{B})| = 300q_3\lambda$ and $|\mathcal{A}\cap\mathcal{B}| = 300q_4\lambda$, which yield that

$$\begin{aligned} p_A(1-p_B)\lambda &= 20/300\,, \\ p_B(1-p_A)\lambda &= 40/300\,, \\ p_A p_B\lambda &= 80/300\,. \end{aligned}$$

Thus we have $p_A = 2/3$, $p_B = 4/5$ and $\lambda = 1/2$.

Finally we prove our claim. Suppose $X \sim \mathrm{Pois}(\theta)$ with some unknown θ. Then given x, our goal (*maximum likelihood estimation*) is to find $\arg\max_\theta p(x\mid\theta) \triangleq \mathbf{Pr}[X = x \mid X \sim \mathrm{Pois}(\theta)]$. Note that

$$p(x\mid\theta) = \mathrm{e}^{-\theta}\cdot\frac{\theta^x}{x!}\,.$$

So the *likelihood function* is $\mathcal{L}(\theta) \triangleq p(x\mid\theta)$ and the *log-likelihood function* is

$$\ell(\theta) \triangleq \ln\mathcal{L}(\theta) = -\theta + x\ln\theta - \ln(x!)\,.$$

Since $\ell'(\theta) = -1 + x/\theta$, it is easy to verify that $\arg\max_\theta p(x\mid\theta) = \arg\max_\theta \ell(\theta) = x$.

Example 3 (Coupon Collector Problem for Non-Uniform Coupons). Suppose that the rates of all types of coupons are not identical in the coupon collector problem. Specifically, suppose that there are n distinct types of coupons and each purchase gives a coupon of type-i independently with probability p_i. Let N be the random variable that denotes the number of purchases until collecting all n types of coupons. Our goal is to compute $\mathbb{E}[N]$.

It is clear that $\sum_{j=1}^{n} p_j = 1$. Let N_j be the random variable that denotes the number of purchases until collecting type j. Then N_j has a *geometric distribution* with parameter p_j and $N = \max_{1 \le j \le n} N_j$.

However it is difficult to compute the distribution of the maximum of N_j since they are not independent. We now consider another case: the maximum of *independent* exponential random variables.

Suppose that the coupons are collected at times chosen according to a Poisson process with rate $\lambda = 1$, where each arrival of this Poisson process brings a type-j coupon independently with probability p_j. Let $\{P_j(t)\}$ be the thinning of this Poisson process, X_j be the first time to meet a type-j coupon, and $X = \max_{1 \le j \le n} X_j$. Note that X_j has an exponential distribution of rate p_j. By the independence of thinning, we have

$$\begin{aligned}
\mathbf{Pr}[X \le t] &= \mathbf{Pr}[X_1 \le t \wedge X_2 \le t \wedge \cdots \wedge X_n \le t] \\
&= \prod_{j=1}^{n} \mathbf{Pr}\left[X_j \le t\right] \\
&= \prod_{j=1}^{n} \left(1 - \mathrm{e}^{-p_j t}\right),
\end{aligned}$$

which implies that $\mathbf{Pr}[X > t] = 1 - \prod_{j=1}^{n}\left(1 - \mathrm{e}^{-p_j t}\right)$.

We now claim that for any nonnegative random variable X, it holds that

$$\mathbb{E}[X] = \int_0^{\infty} \mathbf{Pr}[X > t]\,\mathrm{d}t.$$

Proof of our claim. It is a double-counting. We first prove the discrete version where $X \in \mathbb{N}$:

$$\mathbb{E}[X] = \sum_{t=0}^{\infty} t \cdot \mathbf{Pr}[X = t] = \sum_{t=0}^{\infty}\sum_{s=0}^{t-1} \mathbf{Pr}[X = t] = \sum_{s=0}^{\infty}\sum_{t=s+1}^{\infty} \mathbf{Pr}[X = t] = \sum_{s=0}^{\infty} \mathbf{Pr}[X > s].$$

The proof of continuous case is almost the same.

$$\mathbb{E}[X] = \mathbb{E}\left[\int_0^{X} 1\,\mathrm{d}t\right] = \mathbb{E}\left[\int_0^{\infty} \mathbb{1}_{[X>t]}\,\mathrm{d}t\right].$$

Fubini's theorem, or Tonelli's theorem justifies exchanging the order of expectation and integration. Hence,

$$\mathbb{E}[X] = \int_0^{\infty} \mathbb{E}\left[\mathbb{1}_{[X>t]}\right]\mathrm{d}t = \int_0^{\infty} \mathbf{Pr}[X > t]\,\mathrm{d}t. \qquad \square$$

Remark. Note that in our proof we need Fubini's theorem or Tonelli's theorem. The conclusions of these two theorems are identical, but the assumptions are different. In fact, there are various alternative statements. Here we introduce a simple one.

Theorem 2 (Fubini's Theorem). *Let $A \times B \subseteq \mathbb{R}^2$ and suppose that $f : A \times B \to \mathbb{R}$ is a measurable function such that either $f \geq 0$ throughout $A \times B$ or*

$$\int_{A\times B} |f| \,\mathrm{d}(x,y) < \infty,$$

Then it follows that $\int_{A\times B} f(x,y)\,\mathrm{d}(x,y)$ can be evaluated by way of an iterated integral in either order, that is,

$$\int_A \Big(\int_B f(x,y)\,\mathrm{d}y\Big)\,\mathrm{d}x = \int_B \Big(\int_A f(x,y)\,\mathrm{d}x\Big)\,\mathrm{d}y.$$

Now let's continue our analysis of the coupon collector problem. Using our claim, we conclude that

$$\mathbb{E}[X] = \int_0^\infty \mathbf{Pr}[X > t]\,\mathrm{d}t = \int_0^\infty 1 - \prod_{j=1}^n \left(1 - \mathrm{e}^{-p_j t}\right)\mathrm{d}t.$$

Finally, we relate $\mathbb{E}[X]$, the expected time until collecting all types of coupons in the Poisson process, to our goal $\mathbb{E}[N]$ in the coupon collector problem.
Let τ_i denote the time between the $i-1$-th arrival of coupons and the i-th arrival of coupons. It is clear that

$$X = \sum_{i=1}^N \tau_i,$$

where τ_i has an exponential distribution with rate $\lambda = 1$. Taking the expectation of the both sides we have

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^N \tau_i\right].$$

If N is a constant, then we have $\mathbb{E}[X] = \sum_{i=1}^N \mathbb{E}[\tau_i] = N\cdot\mathbb{E}[\tau_i]$ immediately. However, N is a random variable now.

Question. Is it true that

$$\mathbb{E}\left[\sum_{i=1}^N \tau_i\right] = \mathbb{E}[N]\cdot\mathbb{E}[\tau_i]$$

if N is a random variable?

The answer is true if $\tau_1, \tau_2, \ldots$ are i.i.d. random variables with finite mean, N is independent of $(\tau_1, \tau_2, \ldots)$ and N has finite expectation as well. Actually, this is a simple case of the Wald's equation, which we will introduce later. Here we only give a simple proof of our case.

Note that τ_i are i.i.d. exponential random variables with rate 1, and N is indpendent of $(\tau_1, \tau_2, \ldots)$, so

$$\mathbb{E}[X \mid N = n] = \mathbb{E}[\tau_1 + \cdots + \tau_N \mid N = n] = \mathbb{E}[\tau_1 + \cdots + \tau_n] = N \cdot \mathbb{E}[\tau_i] = N .$$

Applying the law of total expectation, we have

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid N]] = \mathbb{E}[N] ,$$

and hence conclude that

$$\mathbb{E}[N] = \int_0^\infty \mathbf{Pr}[X > t] \, \mathrm{d}t = \int_0^\infty 1 - \prod_{j=1}^n \left(1 - \mathrm{e}^{-p_j t}\right) \mathrm{d}t .$$

Now we verify that the expectation of purchases is n times the n-th harmonic number in the uniform coupon collector problem.

$$\begin{aligned}
\mathbb{E}[N] &= \int_0^\infty 1 - \left(1 - \mathrm{e}^{-t/n}\right)^n \mathrm{d}t \\
&= \int_0^\infty \left(1 - \left(1 - \mathrm{e}^{-t/n}\right)^n\right) \cdot \left(-n \cdot \mathrm{e}^{t/n}\right) \mathrm{d}\mathrm{e}^{-t/n} \\
&= n \int_0^1 \frac{1}{x} - \frac{(1-x)^n}{x} \, \mathrm{d}x \\
&= n \sum_{k=1}^n \int_0^1 \frac{(1-x)^{k-1}}{x} - \frac{(1-x)^k}{x} \, \mathrm{d}x \\
&= n \sum_{k=1}^n \int_0^1 (1-x)^{k-1} \, \mathrm{d}x \\
&= n \sum_{k=1}^n \frac{1}{k} .
\end{aligned}$$

2 Conditioning

Let $T_1, T_2, T_3, \ldots$ be the arrival times of a Poisson process with rate λ. Suppose that $N(t) = n$, that is, $0 < T_1 < T_2 < \cdots < T_n \le t < T_{n+1}$. Let $U_1, U_2, \ldots, U_n$ be n numbers independently and uniformly at random chosen from $[0, t]$, and then $V_1, V_2, \ldots, V_n$ be n nubmers by rearranging $U_1, U_2, \ldots, U_n$ in increasing order. Conditioned on $N(t) = n$, we have the following theorem.

Theorem 3. *The vector $(T_1, T_2, \ldots, T_n)$ has the same distribution as $(V_1, V_2, \ldots, V_n)$.*

Proof. For any $0 < t_1 < t_2 < \cdots < t_n \le t$, by independence we have

$$\begin{aligned}\mathbf{Pr}\left[\bigwedge_{i=1}^{n} T_i = t_i \,\middle|\, N(t) = n\right] &= \frac{\mathbf{Pr}[\tau_1 = t_1, \tau_2 = t_2 - t_1, \dots, \tau_n = t_n - t_{n-1}, \tau_{n+1} > t - t_n]}{\mathbf{Pr}[N(t) = n]} \\ &= \frac{\lambda \mathrm{e}^{-\lambda t_1} \cdot \lambda \mathrm{e}^{-\lambda(t_2 - t_1)} \cdot \ \cdots \ \cdot \lambda \mathrm{e}^{-\lambda(t_n - t_{n-1})} \cdot \mathrm{e}^{-\lambda(t - t_n)}}{\mathrm{e}^{-\lambda t}(\lambda t)^n / n!} \\ &= \frac{\lambda^n \mathrm{e}^{-\lambda t}}{\mathrm{e}^{-\lambda t}(\lambda t)^n / n!} = \frac{n!}{t^n}.\end{aligned}$$

Note that the volume of the simplex $S = \{(t_1, t_2, \dots, t_n) : 0 < t_1 < t_2 < \ldots < t_n \le t\}$ is $t^n / n!$. So $(T_1, T_2, \dots, T_n)$ has a uniform distribution over S.

It is easy to see that $(V_1, V_2, \dots, V_n)$ also has a uniform distribution over S, since for any permutation $\pi = (\pi_1, \dots, \pi_n)$ of $[n]$,

$$\mathbf{Pr}\left[U_{\pi_1} < U_{\pi_2} < \cdots < U_{\pi_n}\right] = \frac{1}{n!},$$

and thus

$$\mathbf{Pr}\left[\bigwedge_{i=1}^{n} V_i = t_i \,\middle|\, U_{\pi_1} < U_{\pi_2} < \cdots < U_{\pi_n}\right] = \frac{\mathbf{Pr}\left[\bigwedge_{i=1}^{n} U_{\pi_i} = t_i\right]}{\mathbf{Pr}\left[U_{\pi_1} < U_{\pi_2} < \cdots < U_{\pi_n}\right]} = \frac{1/t^n}{1/n!} = \frac{n!}{t^n}.$$

Remark. Here we should point out that the argument to show the volume of S only works for finite set. Since every infinite subset of $\mathbb{R}^n$ has a one-to-many mapping into itself, please **DO NOT** compute the volume by mapping points to another set. Actually it is easy to compute the volume of S by an induction on n:

$$\begin{aligned}\mathrm{Vol}(S) &= \int_{0 < t_1 < t_2 < \cdots < t_n \le t} 1 \,\mathrm{d}(t_1, t_2, \dots, t_n) \\ &= \int_0^t \Big(\int_{0 < t_1 < t_2 < \cdots < t_{n-1} < t_n} 1 \,\mathrm{d}(t_1, t_2, \dots, t_n) \Big) \mathrm{d}t_n \\ &= \int_0^t \frac{t_n^n}{(n-1)!} \,\mathrm{d}t_n = \frac{t^n}{n!}.\end{aligned}$$

□

Corollary 4. *Suppose that* $0 \le s < t$ *and* $m \le n$. *Then*

$$\mathbf{Pr}[N(s) = m \mid N(t) = n] = \binom{n}{m} \cdot \left(\frac{s}{t}\right)^m \cdot \left(1 - \frac{s}{t}\right)^{n-m}.$$

Proof. The probability is the same as the probability that exactly m of n independently and uniformly chosen numbers are in $(0, s]$. □

3 Poisson Approximation

Example 4 (Balls-into-bins, Maxload). There are m balls and n bins. Each ball is independently and uniformly put into a bin. So for each bin, the expected number of balls in this bin is m/n, and the number has a binomial distribution with parameter $1/n$. However, the number of balls in these bins are not independent. So the problem is how to analyze the joint distribution.

Formally, let X_i denote the number of balls in the i-th bin. Then we have $X_i \sim \mathsf{Binom}(m, 1/n)$ and $\sum_{i=1}^{n} X_i = m$.

Theorem 5. *The distribution of $(X_1, \ldots, X_n)$ is the same as the distribution of $(Y_1, \ldots, Y_n)$ conditioned on $\sum_{i=1}^{n} Y_i = m$, where $Y_i \sim \mathsf{Pois}(\lambda)$ are independent Poisson random variables with an arbitrary rate λ.*

The proof is straightforward. But this theorem has many important applications. For example, let's consider the following *maxload* problem.

Assume that $m = n$, and let $X = \max_{1 \le i \le n} X_i$. Applying Theorem 5 we will show the following result.

Theorem 6 (Maxload). *There exists two constant $c_1, c_2 > 0$ such that*

$$\mathbf{Pr}\left[c_1 \cdot \frac{\log n}{\log\log n} < X < c_2 \cdot \frac{\log n}{\log\log n}\right] = 1 - o(1/n)\,.$$

Proof of the upper bound. We first prove the upper bound of X as a warm-up, that is, we are going to show that

$$\mathbf{Pr}\left[X \ge c \cdot \frac{\log n}{\log\log n}\right] = o(1/n)$$

for some $c > 0$. For convenience, let $k = c\log n/\log\log n$. Using the union bound we have

$$\begin{aligned}
\mathbf{Pr}[X \ge k] &= \mathbf{Pr}[\exists\, i \text{ s.t. } X_i \ge k] \\
&\le n \cdot \mathbf{Pr}[X_i \ge k] \;\le n\binom{n}{k}\left(\frac{1}{n}\right)^k \\
&\le n \cdot \left(\frac{\mathrm{e}n}{k}\right)^k \cdot n^{-k} \;\le \frac{n \cdot \mathrm{e}^k}{k^k}\,.
\end{aligned}$$

It is sufficient to show that $n \cdot \mathrm{e}^k/k^k < 1/n^{1+\varepsilon}$ for some $\varepsilon > 0$. Taking the logarithms to both sides, it is equivalent to

$$\log n + k - k\log k < -(1+\varepsilon)\log n\,.$$

Hence $k \log k > 3 \log n$ suffices. Note that

$$
\begin{aligned}
k \log k &= c \cdot \frac{\log n}{\log\log n} \cdot (\log\log n - \log\log\log n + \log c) \\
&> c \log n \left(1 - \frac{\log\log\log n}{\log\log n}\right) > \frac{c}{2} \cdot \log n .
\end{aligned}
$$

So we complete our proof by letting $c > 6$. □

The proof of the lower is a bit more complicated. Now introduce a powerful tool.

Theorem 7. *Let $f : \mathbb{N}^n \to \mathbb{N}$ be an arbitrary function, $Y_1, Y_2, \ldots, Y_n$ be n independent Poisson random variables with rate $\lambda = m/n$, i.e., $Y_i \sim \mathrm{Pois}(m/n)$. Then we have*

$$\mathbb{E}\left[f(X_1, X_2, \ldots, X_n)\right] \leq \mathrm{e}\sqrt{m} \cdot \mathbb{E}\left[f(Y_1, Y_2, \ldots, Y_n)\right] .$$

Remark. Note that the difference between the statement of this theorem and Theorem 5 is that we no longer need conditioning on $\sum Y_i = m$.

The power of this theorem is to bound the expectation of *any* function of X_i by the expectation of the function of *independent* Poissons. Usually, we let the function indicate some event, so the expectation of the indicator is the probability that the event happens.

As an application of this theorem, we first prove the lower bound of X in Theorem 6.

Proof of the lower bound. We are now ready to show that there exists $c > 0$ s.t.

$$\mathbf{Pr}\left[X \leq c \cdot \frac{\log n}{\log\log n}\right] = o(1/n) .$$

Let $k = c \log n / \log\log n$ and f be the indicator function such that $f(x_1, x_2, \ldots, x_n) = 1$ if $x_i \leq k$ for all i and $f = 0$ otherwise. Namely,

$$f(x_1, x_2, \ldots, x_n) = \mathbb{1}_{\max_{1 \leq i \leq n} x_i \leq k} .$$

Hence, $\mathbb{E}\left[f(Z_1, Z_2, \ldots, Z_n)\right] = \mathbf{Pr}[\max_{1 \leq i \leq n} Z_i \leq k]$ for any random variables $Z_1, Z_2, \ldots, Z_n$. Since $Y_1, Y_2, \ldots, Y_n$ are independent Poissons, we have

$$
\begin{aligned}
\mathbb{E}\left[f(Y_1, Y_2, \ldots, Y_n)\right] &= \mathbf{Pr}\left[\max_{1 \leq i \leq n} Y_i \leq k\right] \\
&= \mathbf{Pr}[Y_1 \leq k \wedge Y_2 \leq k \wedge \cdots \wedge Y_n \leq k] \\
&= \mathbf{Pr}[Y_1 \leq k] \cdot \mathbf{Pr}[Y_2 \leq k] \cdot \cdots \cdot \mathbf{Pr}[Y_n \leq k] \\
&\leq \left(1 - \mathbf{Pr}[Y_i = k+1]\right)^n \\
&= \left(1 - \frac{1}{(k+1)! \cdot \mathrm{e}}\right)^n \\
&\leq \mathrm{e}^{-n/(\mathrm{e}(k+1)!)} .
\end{aligned}
$$

Note that $\ln(k+2) - \ln k = \ln(1+2/k) < 2/k$, so for $k \geq 2$,

$$\ln(k+1)! = \sum_{i=1}^{k+1} \ln i < \int_1^{k+2} \ln x \, \mathrm{d}x = (k+2)\ln(k+2) - k - 1 < (k+2)\log - k + 3 .$$

Plugging in $k = c \log n / \log\log n$ and letting $c = 1$, we have

$$\begin{aligned} \ln(k+1)! &< \frac{c\log n + 2\log\log n}{\log\log n}\big(\log\log n + \log c - \log\log\log n\big) - \frac{c\log n}{\log\log n} + 3 \\ &< \log n + 2\log\log n - \log n / \log\log n \; < \log n - \log\log n - 2 \end{aligned}$$

for sufficiently large n. It follows that

$$\mathrm{e} \cdot (k+1)! < \frac{n}{\mathrm{e} \cdot \log n}$$

and thus

$$\mathbb{E}\left[f(Y_1, Y_2, \ldots, Y_n)\right] \leq \mathrm{e}^{-n/(\mathrm{e}(k+1)!)} < \mathrm{e}^{-\mathrm{e}\cdot\log n} = n^{-\mathrm{e}} .$$

Combining with Theorem 7 it concludes that

$$\mathbf{Pr}\left[X \leq \frac{\log n}{\log\log n}\right] = \mathbb{E}\left[f(X_1, \ldots, X_n)\right] \leq \mathrm{e}\sqrt{n} \cdot \mathbb{E}\left[f(Y_1, \ldots, Y_n)\right] < \frac{\mathrm{e}\sqrt{n}}{n^{\mathrm{e}}}$$

for sufficiently large n. □

As the last part of today's lecture, let's prove Theorem 7.

Proof of Theorem 7. Applying Theorem 5, it is easy to see that

$$\begin{aligned} \mathbb{E}\left[f(Y_1, Y_2, \ldots, Y_n)\right] &= \sum_{k=0}^{\infty} \mathbb{E}\left[f(Y_1, Y_2, \ldots, Y_n) \mid \sum Y_i = k\right] \cdot \mathbf{Pr}\left[\sum Y_i = k\right] \\ &\geq \mathbb{E}\left[f(Y_1, Y_2, \ldots, Y_n) \mid \sum Y_i = m\right] \cdot \mathbf{Pr}\left[\sum Y_i = m\right] \\ &= \mathbb{E}\left[f(X_1, X_2, \ldots, X_n)\right] \cdot \mathbf{Pr}\left[\sum Y_i = m\right] . \end{aligned}$$

Now it is sufficient to verify that $\mathbf{Pr}[\sum Y_i = m] \geq 1/(\mathrm{e}\sqrt{m})$ if $Y_i \sim \mathrm{Pois}(m/n)$. The proof is straightforward by *Stirling's approximation*, since $\sum Y_i \sim \mathrm{Pois}(m)$ and thus

$$\mathbf{Pr}\left[\sum Y_i = m\right] = \mathrm{e}^{-m} \cdot \frac{m^m}{m!} > \mathrm{e}^{-m} \frac{m^m}{\mathrm{e}^{1/(12n)}\sqrt{2\pi m}(m/\mathrm{e})^m} > \frac{1}{\mathrm{e}\sqrt{m}} .$$ □

Remark. Theorem 6 estimates the maxload of the balls-into-bins model, which is very useful in computer science. For example, in the Erdős-Rényi random graph model, using the same technique we can show that the maximum degree of a random graph $\mathcal{G}(n, d/n)$ is $\Theta(\log n / \log\log n)$ with high probability, even though the average degree d is a constant.