4 Poisson Approximation and Poisson Process

4.1 Limit of Binomial distributions

The m-balls-into-n-bins model is the following simple random process: throwing m balls into n bins uniformly at random. We've already met the model in previous lectures. Today we continue to discuss the model with some new tools.

Let X_i be the number of balls in the i-th bin. It is easy to find that X_i follows the **binomial distribution**

$$\Pr[X_i = r] = \binom{m}{r}\left(\frac{1}{n}\right)^r\left(1-\frac{1}{n}\right)^{m-r}.$$

Fix r, and let $m, n \to \infty$. It yields that

$$\begin{aligned}\Pr[X_i = r] &\approx \frac{m^r}{r!}n^{-r}\mathrm{e}^{-m/n}\\ &= \frac{1}{r!}\left(\frac{m}{n}\right)^r \mathrm{e}^{-m/n}.\end{aligned}$$

The distribution is known as the **Poisson distribution** with mean $\lambda = m/n$.

Definition 4.1. Poisson distribution

A random variable X is said to follow a *Poisson distribution* with mean λ, denoted $X \sim \text{Pois}(\lambda)$, if

$$\Pr[X = k] = \frac{\lambda^k}{k!}\cdot \mathrm{e}^{-\lambda}.$$

We can verify that a Poisson distribution is indeed a distribution, since

$$\sum_{k=0}^{\infty}\Pr[X = k] = \mathrm{e}^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = \mathrm{e}^{-\lambda}\mathrm{e}^{\lambda} = 1.$$

Moreover, the expectation of a Poisson with mean λ is indeed λ:

$$\mathbb{E}_{X\sim\text{Pois}(\lambda)}[X] = \sum_{k=0}^{\infty} k\cdot\frac{\lambda^k}{k!}\cdot \mathrm{e}^{-\lambda} = \lambda\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)}\cdot \mathrm{e}^{-\lambda} = \lambda\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}\cdot \mathrm{e}^{-\lambda} = \lambda.$$

Remark 4.1.

Let λ be a fixed *constant*. A Poisson distribution $\text{Pois}(\lambda)$ is the limit of binomial distributions $\text{Binom}(n, \lambda/n)$ as $n \to \infty$. If n is sufficiently large but $np = O(1)$, $\text{Binom}(n, p)$ is approximately $\text{Pois}(np)$.

Proposition 4.1.
Suppose that $X_1, X_2, \dots, X_n$ are n mutually independent random variables, where $X_i \sim \text{Pois}(\lambda_i)$. Then

$$\left(\sum_{i=1}^{n} X_i\right) \sim \text{Pois}\left(\sum_{i=1}^{n} \lambda_j\right).$$

In particular, if $X_1, X_2, \dots, X_n$ are i.i.d. Poissons with mean λ, then

$$\left(\sum_{i=1}^{n} X_i\right) \sim \text{Pois}(n\lambda).$$

Proof. We only prove the case of two r.v.s. It is easy to extend the result to finite number of r.v.s.

Suppose that $X_1 \sim \text{Pois}(\lambda_1)$ and $X_2 \sim \text{Pois}(\lambda_2)$ are independent. We calculate the distribution of $X_1 + X_2$ directly:

$$\begin{aligned}
\Pr[X_1 + X_2 = n] &= \sum_{m=0}^{n} \Pr[X_1 = m \wedge X_2 = n - m] \\
&= \sum_{m=0}^{n} \Pr[X_1 = m] \cdot \Pr[X_2 = n - m] \\
&= \sum_{m=0}^{n} \frac{\lambda_1^m}{m!} \cdot \mathrm{e}^{-\lambda_1} \cdot \frac{\lambda_2^{n-m}}{(n-m)!} \cdot \mathrm{e}^{-\lambda_2} \\
&= \mathrm{e}^{-(\lambda_1+\lambda_2)} \sum_{m=0}^{n} \frac{\lambda_1^m \lambda_2^{n-m}}{n!} \cdot \binom{n}{m} \\
&= \frac{(\lambda_1 + \lambda_2)^n}{n!} \cdot \mathrm{e}^{-(\lambda_1+\lambda_2)}.
\end{aligned}$$

□

4.2 Poisson Approximation

In the balls-into-bins model, each X_i has identical distribution $\text{Binom}(m, 1/n)$. But they are highly correlated, since $\sum X_i = m$. This is the main difficulty to analyze the model in many situations. Surprisingly, if we replace these binomial random variables with independent Poisson variables, the distributions turn out to be the same under some condition.

Theorem 4.2.
The joint distribution of $(X_1, \dots, X_n)$ is the same as the joint distribution of $(Y_1, \dots, Y_n)$ conditioned on $\sum_{i=1}^{n} Y_i = m$, where $Y_i \sim \text{Pois}(\lambda)$ are independent Poisson random variables with an arbitrary rate λ.

Proof. We first give the distribution of $(X_1, \dots, X_n)$. Note that the number of all possible ways to put balls into bins is n^m, and the number of the permutations of a multiset of r_1 1's, r_2 2's, ..., and r_n n's is

$$\binom{r_1 + r_2 + \cdots + r_n}{r_1, r_2, \dots, r_n} \triangleq \frac{(r_1 + r_2 + \cdots + r_n)!}{r_1!\, r_2! \cdots r_n!}.$$

Hence for all $r_1, r_2, \dots, r_n \in \mathbb{N}$ such that $\sum r_i = m$, we have

$$\Pr[X_1 = r_1, X_2 = r_2, \dots, X_n = r_n] = \frac{1}{n^m} \cdot \frac{m!}{r_1!\, r_2! \cdots r_n!}.$$

Now we show that $(Y_1, \dots, Y_n)$ has the same conditional distribution.

$$\begin{aligned}
&\Pr\left[Y_1 = r_1, Y_2 = r_2, \dots, Y_n = r_n \mid \sum Y_i = m\right] \\
&= \frac{\Pr\left[Y_1 = r_1, Y_2 = r_2, \dots, Y_n = r_n\right]}{\Pr\left[\sum Y_i = m\right]} \\
&= \frac{\prod_{i=1}^{n} \Pr[Y_i = r_i]}{\Pr\left[\sum Y_i = m\right]} \qquad = \frac{\prod_{i=1}^{n} \mathrm{e}^{-\lambda} \cdot \frac{\lambda^{r_i}}{r_i!}}{\mathrm{e}^{-n\lambda} \cdot \frac{(n\lambda)^m}{m!}} \qquad = \frac{1}{n^m} \cdot \frac{m!}{r_1!\, r_2! \cdots r_n!}. \qquad \square
\end{aligned}$$

The theorem tells us that one might try to use independent Poisson variables to replace correlated binomial variables when studying the balls-into-bins model. However, in order to apply the theorem, we still need to handle the condition. Sometimes we can simply drop the condition since it happens with reasonable probability.

Theorem 4.3. Poisson Approximation
Let $f : \mathbb{N}^n \to \mathbb{N}$ be an arbitrary function, $Y_1, Y_2, \dots, Y_n$ be n independent Poisson r.v.s with rate $\lambda = m/n$, i.e., $Y_i \sim \mathsf{Pois}(m/n)$. Then we have

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

Remark 4.2.
The power of this theorem is to bound the expectation of *any* function of X_i by the expectation of the function of *independent* Poisson variables. Usually, we let the function indicate some event, so the expectation of the indicator is the probability that the event happens. For example, let f be an indicator function of some event $\mathcal{B}(X_1, \dots, X_n)$. Then

$$\begin{aligned}
\Pr[\mathcal{B}(X_1, \dots, X_n)] = \mathbb{E}[f(X_1, \dots, X_n)] &\le \mathrm{e}\sqrt{m} \cdot \mathbb{E}[f(Y_1, \dots, Y_n)] \\
&= \mathrm{e}\sqrt{m} \cdot \Pr[\mathcal{B}(Y_1, \dots, Y_n)].
\end{aligned}$$

As an application of this theorem, we consider the MaxLoad problem in the balls-into-bins model with $m = n$. Recall that in the last lecture, we showed that $\Pr[\max X_i > k] \le o(1/n)$ if $k = O(\log n / \log\log n)$. Today we analyze the lower bound.

Theorem 4.4. Max Load
Let $X = \max X_i$ in the m-balls-into-n-bins model with $m = n$. Then there exists two constant $c_1, c_2 > 0$ such that

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

Proof. We only prove the lower bound here, i.e., we show that there exists $c > 0$ s.t.

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

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

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

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

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

Note that

$$\ln(k+1)! = \sum_{r=2}^{k+1} \ln r < (k-2)\ln k + \ln 2 + \ln(k+1) < k \ln k\,.$$

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

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

for sufficiently large n. It follows that

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

and thus

$$\mathbb{E}[f(Y_1, Y_2, \dots, Y_n)] \le \mathrm{e}^{-n/(\mathrm{e}(k+1)!)} < \mathrm{e}^{-\mathrm{e}\cdot\log n} = n^{-\mathrm{e}}\,. \qquad \square$$

Another example is coupon collector. Since it can be thought of as balls-and-bins problem, we can use Poisson approximation to obtain much stronger results.

Recall that we have seen

$$\Pr[X > n\ln n + cn] < \mathrm{e}^{-c}$$

in the last lecture, which is close to this result as $1 - x \approx \mathrm{e}^{-x}$ when x is sufficiently small.

> **Theorem 4.5.**
> $\Pr[X > n\log n + cn] = 1 - \mathrm{e}^{-\mathrm{e}^{-c}}$ as $n \to \infty$.

Proof. Suppose we throw $m = n\log n + cn$ balls. Let X_i be the number of the i-th type, Y_i be the Poisson approximation of X_i, and $Y = \sum Y_i$.

It's clear that $Y_i \sim$ Pois$(\log n + c)$. So we can get $\Pr[Y_i = 0] = \frac{\mathrm{e}^{-c}}{n}$. Since Y_i's are independent and have identical distribution, it follows that

$$\begin{aligned}
\Pr[\nexists Y_i = 0] &= \Big(1 - \Pr[Y_i \ne 0]\Big)^n \\
&= \left(1 - \frac{\mathrm{e}^{-c}}{n}\right)^n \\
&\approx \mathrm{e}^{-\mathrm{e}^{-c}}\,.
\end{aligned}$$

The remaining part is to show the Poisson approximation is accurate. Note that we can not directly apply Corollary 8 here, because it will multiply $\mathrm{e}\sqrt{m}$ to the probability, making the bound too loose. In fact we need some more powerful tool, which will discuss later. $\square$

Finally we give the proof of Poisson approximation.

Proof of Theorem 4.3. Applying Theorem 4.2, it is easy to see that

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

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

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

4.3 Poisson Process

Suppose that there exists a restaurant. How can we predict the nubmer of tomorrow's customers based on the number of customers in the past several days?

For instance, we assume that the number of customers in the past five days are: 100, 120, 80, 75 and 110. A natural idea is to use the average number (e.g., 97 in our instance) of the past. However, with probability $1/2$ or even greater, the restaurant may not prepare sufficient food.

To analyze the distribution of the number of customers, we should make some assumptions first. Assume that there are n slots in a day. Every slot is sufficiently small such that at most one customer comes into the restaurant in a slot and the probability of coming in each slot is p independently of each other. Now let's compute the distribution of the number of customers X_n (where we denote $p\cdot n$ by λ):

$$\begin{aligned}\Pr[X_n=k] &= \binom{n}{k}\cdot p^k\cdot(1-p)^{n-k}\\ &= \binom{n}{k}\cdot\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k}\overset{n\to\infty}{\longrightarrow}\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}.\end{aligned}$$

So X_n is a Poisson variable.

The Poisson distribution with mean λ can be used to describe the distribution of the number of customers when the average number of customers per unit time (e.g., per day) is λ. If we count the number of customers over a period of time (for example, from time t_1 to time t_2), assuming that t_2-t_1 is an integer and the arrival of customers is still uniform, then according to our conclusion about the sum of multiple Poisson random variables, the number of customers arriving during this period should follow the $\mathsf{Pois}((t_2-t_1)\lambda)$ distribution. We can use the Poisson process to describe the number of customers arriving over a period of time.

Definition 4.2. Poisson process
A *Poisson process* $\{N(s):\, s\geq 0\}$ with rate λ satisfies that

1. $N(0)=0$;
2. $\forall\, t,s\geq 0$, $N(t+s)-N(s)\sim \mathsf{Pois}(\lambda\cdot t)$;
3. $\forall\, t_0\leq t_1\leq\cdots\leq t_n$, $N(t_1)-N(t_0), N(t_2)-N(t_1),\dots,N(t_n)-N(t_{n-1})$ are mutually independent.

Why does this process exist?
In fact, the Poisson process has another *constructive* definition.

Definition 4.3. Exponential distribution
The probability density function of the exponential distribution with rate $\lambda > 0$ is given by

$$f(x) = \begin{cases} \lambda \mathrm{e}^{-\lambda x}, & x \geq 0; \\ 0, & \text{otherwise}. \end{cases}$$

So the cumulative probability function of $X \sim \text{Exponential}(\lambda)$ is

$$F_X(x) = \Pr[X \leq k] = \int_{-\infty}^{k} f(x)\,\mathrm{d}x = 1 - e^{-\lambda k}.$$

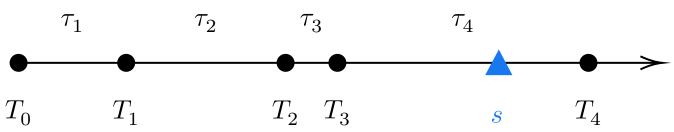
Then the following proposition gives another definition of the Poisson process.

We omit the proof here. If you are interested, you can refer to any textbook on stochastic processes.

Theorem 4.6.
Suppose that $\tau_1, \tau_2, \ldots, \tau_n, \ldots$ is a sequence of independent random variables that each r.v. has an exponential distribution with rate λ (i.e., $\tau_i \sim \text{Exponential}(\lambda)$). Let $T_n = \sum_{i=1}^{n} \tau_i$ and

$$N(s) \triangleq \max\{n : T_n \leq s\}.$$

Then $N(s)$ is a Poisson process with rate λ.



Now we discuss some properties of the exponential distribution.

Proposition 4.7.
Let $X \sim \text{Exponential}(\lambda)$. Then $\mathbb{E}[X] = 1/\lambda$.

Proof. We calculate the expectation directly.

$$\begin{aligned} \mathbb{E}[X] &= \int_0^\infty t \cdot \lambda \mathrm{e}^{-\lambda t}\,\mathrm{d}t = -\int_0^\infty t\,\mathrm{d}\mathrm{e}^{-\lambda t} \\ &= \left(-t \cdot \mathrm{e}^{-\lambda t}\right)\Big|_0^\infty + \int_0^\infty \mathrm{e}^{-\lambda t}\,\mathrm{d}t = -\frac{1}{\lambda} \cdot \left(\mathrm{e}^{-\lambda t}\right)\Big|_0^\infty = -1/\lambda. \end{aligned}$$ □

Remark 4.3.
We think of the τ_n as times between arrivals of customers at the restaurant, so $T_n = \tau_1 + \cdots + \tau_n$ is the arrival time of the n-th customer, and $N(s)$ is the number of arrivals by time s. So $1/\lambda$ measures the average of times between arrivals — the average of τ_i is $1/\lambda$.
Note that in Definition 4.2, the rate λ measures the average increments of arrivals in units of time. Intuitively, the quantity measured by λ in Theorem 4.6 is consistent with the quantity measured by λ in Definition 4.2.

Proposition 4.8.
Let $X \sim \text{Exponential}(\lambda)$. Then $\mathbb{E}[X^2] = 2/\lambda^2$ and thus $\mathbf{Var}[X] = 1/\lambda^2$.

Proof. Again, we calculate it directly.

$$
\begin{aligned}
\mathbb{E}[X^2] = \int_0^\infty t^2 \cdot \lambda e^{-\lambda t}\, dt &= -\int_0^\infty t^2\, de^{-\lambda t} \\
&= (-t^2 \cdot e^{-\lambda t})\Big|_0^\infty + \int_0^\infty e^{-\lambda t}\, dt^2 \\
&= 2\int_0^\infty t \cdot e^{-\lambda t}\, dt = \mathbb{E}[X] \cdot 2/\lambda = 2/\lambda^2 .
\end{aligned}
$$
□

Moreover, the exponential distribution has the following property (that may be a bit surprising). This property somewhat explains the *mutual independence* in Definition 4.2.

Theorem 4.9. Lack of Memory Property
Let $X \sim \text{Exponential}(\lambda)$. Then for all $t, s > 0$,

$$
\Pr[X > t + s \mid X > s] = \Pr[X > t] .
$$

Proof. It is easy to verify that

$$
\begin{aligned}
\Pr[X > t + s \mid X > s] = \frac{\Pr[X > t + s \wedge X > s]}{\Pr[X > s]} &= \frac{\Pr[X > t + s]}{\Pr[X > s]} \\
= \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} &= e^{-\lambda t} = \Pr[X > t] .
\end{aligned}
$$
□

4.4 Thinning of Poisson Process

We now introduce an important property of Poisson processes—**thining**.

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.

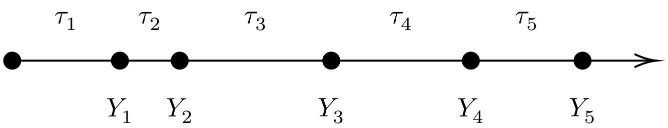


Figure 4.1: Poisson process with random variables associated with arrivals

Formally, suppose that $Y_i \in \mathbb{N}$ and are i.i.d. random variables. Let $p_j = \Pr[Y_i = j]$. For all $j \in \mathbb{N}$ (or $\text{Range}(Y_i)$), 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.

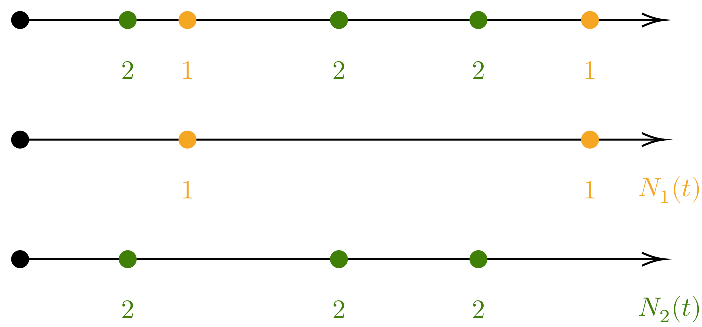


Figure 4.2: Thinning

Theorem 4.10. Thining
$N_j(t)$ are independent rate $p_j\lambda$ Poisson processes. Namely,

1. For all j, $N_j(t)$ is a Poisson process with rate $p_j\lambda$;
2. All $N_j(t)$ are mutually independent.

Remark 4.4.
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 $Y_i \in \{1,2\}$. Then the following calculation concludes the independence and the distribution of $N_j(t)$ at the same time.

$$\begin{aligned}
\Pr[N_1(t)=r_1 \wedge N_2(t)=r_2] &= \Pr[N_1(t)=r_1 \wedge N_1(t)+N_2(t)=r_1+r_2] \\
&= \Pr[N_1(t)+N_2(t)=r_1+r_2]\cdot\Pr[N_1(t)=r_1 \mid N_1(t)+N_2(t)=r_1+r_2] \\
&= \mathrm{e}^{-\lambda t}\cdot\frac{(\lambda t)^{r_1+r_2}}{(r_1+r_2)!}\cdot\binom{r_1+r_2}{r_1}\cdot p_1^{r_1}\cdot p_2^{r_2} \\
&= \mathrm{e}^{-\lambda t}\cdot\frac{(\lambda t)^{r_1+r_2}}{r_1!\cdot r_2!}\cdot p_1^{r_1}\cdot p_2^{r_2} \\
&= \mathrm{e}^{-p_1\lambda t}\cdot\frac{(p_1\lambda t)^{r_1}}{r_1!}\cdot\mathrm{e}^{-p_2\lambda t}\cdot\frac{(p_2\lambda t)^{r_2}}{r_2!}.
\end{aligned}$$
□

This property has many interesting applications. We introduce the analysis of non-uniform coupon collector.

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 $\sum_{i=1}^{n} p_i = 1$. Let N_i be the random variable that denotes the number of purchases until collecting type i. Then N_i has a *geometric distribution* with parameter p_i and $N = \max_{1\le i\le n} N_i$.

However it is difficult to compute the distribution of the maximum of N_i 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 the Poisson process brings a type-i coupon independently with probability p_i. Let $\{P_i(t)\}$ be the thinning of this Poisson process, X_i be the first time to meet a type-i coupon, and $X = \max_{1\le i\le n} X_i$. Note that X_i has an exponential distribution of rate p_i. By the independence of thinning, we have

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

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

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

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

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

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

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,

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

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: 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 $f \ge 0$ 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

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

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. Intuitively one may hope that $\mathbb{E}[X] = \mathbb{E}[N]$ since N is the number of arrivals of time period X in the Poisson process with rate $\lambda = 1$. But we need a rigorous proof.

Let τ_i denote the time between the $(i-1)$-th arrival and the i-th arrival in the Poisson process. 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, clearly we have $\mathbb{E}[X] = \sum_{i=1}^{N} \mathbb{E}[\tau_i] = N$ immediately. However, N is a random variable now.

Question: If N is a random variable, is it true that

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

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

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

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

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