Stochastic Process 1

Rui Gong

February 8, 2024

Contents

1 Poisson Process

1.1 Poisson Approximation to Binomial

Given a Poisson random random variable $Y \sim \text{Poisson}(\lambda)$ with pdf

$$\Pr[Y = k] = \frac{e^{-\lambda}\lambda^k}{k!}, \forall k \in N_0 = \{0, 1 \ldots\}.$$

The probablity of a binomial random variable being k is

$$\binom{n}{k} p^k (1-p)^{n-k}$$

Theorem 1.1. Given $p \to 0$, $np \to \lambda$, we have that

$$\binom{n}{k} p^k (1-p)^{n-k} \to \frac{e^{-\lambda}\lambda^k}{k!}$$

Proof.

$$\begin{aligned}
\binom{n}{k} p^k (1-p)^{n-k} &= \binom{n}{k} (\frac{np}{n})^k (1 - \frac{np}{n})^{n-k} \\
&= \frac{n!}{k!\,(n-k)!} \left(\frac{np}{n}\right)^k \left(1 - \frac{np}{n}\right)^n \left(1 - \frac{np}{n}\right)^{-k} \\
&= \frac{1}{k!} \underbrace{\frac{n(n-1)\ldots(n-k+1)}{n^k}}_{\to 1} \underbrace{(np)^k}_{\to \lambda^k} \underbrace{\left(1 - \frac{np}{n}\right)^n}_{\to e^{-\lambda}} \underbrace{\left(1 - \frac{np}{n}\right)^{-k}}_{\to 1} \\
&\to \frac{1}{k!} \lambda^k e^{-\lambda}
\end{aligned}$$

□

With that, we consider three different binomial random variables:

$$\begin{aligned}
X_n &\sim \text{Binomial}(n, p_n), p_n \to 0, np_n \to \lambda > 0, \text{ as } n \to \infty. \\
Z_p &\sim \text{Binomial}(n(p), p), p \to 0, n(p)p \to \lambda > 0, \text{ as } p \to 0. \\
N_x &\sim \text{Binomial}(n(x), p(x)), p(x) \to 0, n(x) \to \lambda > 0, \text{ as } x.
\end{aligned}$$

For example, if $X_n \sim \text{Binomial}(n, 2/n)$, then we expect

$$\binom{n}{k} p^k (1-p)^{n-k} \to \frac{e^{-2}2^k}{k!}$$

1.2 Total Variance Distance

Let $X_1, \ldots, X_n$ be n independent Bernoulli random variable, where $\mathbb{E}[X_i] = p_i$.

Given $S = \sum X_i$ and $T \sim \text{Poisson}(\lambda = \sum p_i)$, how close are these two distributions? Or, how to measure the closedness?

	0	1	1	3
$\Pr[X=k]$	$5/10$	$3/10$	$1/10$	$1/10$
$\Pr[Y=k]$	$2/10$	$1/10$	$1/10$	$6/10$

Table 1: Discrete Distribution Distance

Definition 1.2. Given two random variables X, Y, (which shares the sample space), we have the *total variance distance* defined as

$$d_{TV}(X,Y) = \sup_A |\Pr[X \in A] - \Pr[Y \in A]|$$

where A is a Borel set defined with respect to the sample space σ-algebra.

Example 1.3. Consider the $A = \{0,1\}$, we have

$$|\Pr[X \in A] - \Pr[Y \in A]| = 3/10 + 2/10 = 1/2.$$

Given $A = \{3\}$,

$$|\Pr[X \in A] - \Pr[Y \in A]| = |1/10 - 6/10| = 1/2.$$

If X, Y take values in a countable set E,

$$\begin{aligned} d_{TV}(X,Y) &= \sum_{i \in E} (\Pr[X=i] - \Pr[Y=i])^+ \\ &= \sum_{i \in E} (\Pr[Y=i] - \Pr[X=i])^+ \\ &= \frac{1}{2} \sum_{i \in E} |\Pr[Y=i] - \Pr[X=i]| \end{aligned}$$

Proposition 1.4. Given two random variables, we have

$$d_{TV}(X,Y) \le \Pr[X \ne Y]$$

Proof. For any A,

$$\begin{aligned} &|\Pr[X \in A] - \Pr[Y \in A]| \\ =&|\Pr[X \in A, Y \in A] + \Pr[X \in A, Y \notin A] - \Pr[Y \in A, X \in A] - \Pr[Y \in A, X \notin A]| \\ =&|\Pr[X \in A, Y \notin A] - \Pr[Y \in A, X \notin A]| \\ \le& \max\{\Pr[X \in A, Y \notin A], \Pr[Y \in A, X \notin A]\} \le \Pr[X \ne Y]. \end{aligned}$$

□

Let $X_1, \ldots, X_n$ be independent Bernoulli random variables with $\mathbb{E}[X_i] = p_i$. Let $S := \sum X_i$, and $T \sim \text{Poisson}(\lambda := p_1 + \ldots + p_n)$. Then $\mathbb{E}[S] = \mathbb{E}[\sum X_i] = \sum \mathbb{E}[X_i] = \sum p_i = \lambda$. Let $Y_1, \ldots, Y_n$ be independent Poissson random variables with $\mathbb{E}[Y_i] = p_i$. Then $T = \sum Y_i \sim \text{Poisson}(\lambda)$. We also have

$$[S \ne T] \subseteq \underbrace{[X_1 \ne Y_1]}_{B_1} \cup [X_2 \ne Y_2] \cup \ldots [X_n \ne Y_n]$$

And hence

$$\begin{aligned}\Pr[S \neq T] &\subseteq \Pr[B_1 \cup \ldots \cup B_n] \\ &\subseteq \Pr[B_1] + \ldots + \Pr[B_n] \\ &\leq p_1^2 + \ldots + p_n^2\end{aligned}$$

where $\Pr[X_i = Y_i] = 1 - p + pe^{-p}, \Pr[X_i \neq Y_i] = p - pe^{-p} \leq p(1 - (1 - p + p^2/2! + \ldots) = p(p - p^2/2! + \ldots) \leq p^2$.

Hence,

$$d_{TV}(S,T) \leq \Pr[S \neq T] \leq \sum_{i=1}^{n} p_i^2.$$

Consider $X_1 \sim \text{Bernoulli}(p_1 = 1/5), X_2 \sim \text{Bernoulli}(p_1 = 1/6), X_3 \sim \text{Bernoulli}(p_3 = 1/10)$, $S = X_1 + X_2 + X_3$ and $T \sim \text{Poisson}(\lambda = \frac{7}{15})$. Then if estimate T by S, for example,

$$\Pr[S \text{ is an odd number}] \approx \Pr[T \text{ is an odd number}]$$

the probability of getting an error is at most

$$(1/5)^2 + (1/6)^2 + (1/10)^2$$

by letting A be the set of odd numbers.

1.3 Probablity Axioms

Consider the sample space Ω, the set of events $\mathcal{F}$ and the probablity P, where

$$\begin{aligned}&\Omega : \text{sample spaces - set of all outcomes} \\ &\mathcal{F} : \text{all events} \\ &P : \mathcal{F} \to [0,1]\end{aligned}$$

. Then we can write a random variable X_1 as:

$$X_1 : \Omega \to \mathbb{R}$$

and an event as

$$B_1 = [X_1 \neq Y_1] = [w \in \Omega | X_1(w) \neq Y_1(w)].$$

Definition 1.5. Event Axioms:

E.1 $\Omega \in \mathcal{F}$

E.2 $A \in \mathcal{F} \implies A^C \in \mathcal{F}$

E.3 $A_1, A_2, \ldots \in \mathcal{F} \implies A_1 \cup A_2 \cup \ldots \in \mathcal{F}$

Definition 1.6. Probability Axioms:

P.1 $A \in \mathcal{F} \implies P(A) \geq 0$

P.2 Countable additivity. $A_1, A_2, \ldots$ being disjoint events, then $P(A_1 \cup A_2 \cup \ldots) = \sum_{i=1} P(A_i)$.

P.3 $P(\Omega) = 1$.

Example 1.7. $X = (\Omega, \mathcal{F}) \to \mathbb{R}$, and let B be a Borel set. Then we can write $\{X = 3\} = \{w \in \Omega : X(w) = 3\} \in \mathcal{F}$. Similarly, $P(X \in B) \in \mathcal{F}$.

$\Omega = \{a, b, c\}$, $\mathcal{F} = \{\emptyset, \Omega, \{a\}, \{b, c\}\}$. Given $X(a) = 1, X(b) = 2, X(c) = 3$, we have

$$[X = 3] = [w \in \Omega : X(w) = 3] = [c]$$

which is not in the event, so X is not a random variable. If $X(b) = 3$, then X is a random variable.

Given $X \sim \text{Exp}(\lambda), Y \sim \text{Exp}(\mu)$ and $X \perp Y$, then

$$\begin{aligned} &P(X > s, Y - X > t | X < Y) \\ =&P(X > s | X < Y) P(Y - X > t | X < Y) \end{aligned}$$

$\lambda \to \lambda \implies (1 + \frac{\lambda_n}{n})^n \to e^{\lambda}$. $f(h) = o(h) \implies f(h)/h \to 0$ as $h \to 0$.

Fix x, a function f is differentiable at x iff there exists a number $f'(x)$ such that

$$\begin{aligned} f(x + h) &= f(x) + hf'(x) + o(h) \\ \frac{f(x + h) - f(x)}{h} &= f'(x) + o(h)/h, h \to 0 \end{aligned}$$

For example, if we want to show $n \log(1 + \frac{\lambda_n}{n}) \to \lambda$. Take $h_n = \lambda_n/n$, $x = 1$.

$$n \log(1 + h_n) = nh_n + nO(h_n) = nh_n + \frac{\lambda_n}{h_n} O(h_n)$$

where $\log(1 + h) = \log(1) + h + o(h)$. Then as $n \to \infty$, we have $h_n \to 0$, $nh_n \to \lambda$, $n \log(1 + \lambda_n/n) \to \lambda$.

Suppose X is nonnegative, integer-valued random variable $P(X = k) = p_k$ for $k = 0, 1, 2, \ldots$, then

$$G(s) = \mathbb{E}[s^X] = \sum_{k=0}^{\infty} p_k s^k = p_0 + p_1 s + p_2 s^2 + \ldots$$

and $G(s) < \infty$ for $|s| < R$. Then we have

$$\begin{aligned} G'(s) &= \sum_{k=0}^{\infty} k p_k s^{k-1} = \mathbb{E}[X s^{X-1}] \\ G'(1) &= \mathbb{E}[X] \\ G''(s) &= \sum_{k=0}^{\infty} k(k-1) p_k s^{k-2} = \mathbb{E}[X(X-1) s^{X-2}] \\ G''(1) &= \mathbb{E}[X(X-1)] = \mathbb{E}[X^2 - X] = \mathbb{E}[X^2] - \mathbb{E}[X] \\ \mathbb{E}[X^2] &= G''(1) + G'(1) \\ var(X) &= G''(1) + G'(1) - [G'(1)]^2 = E[X^2] - E[X]^2 \end{aligned}$$

$$G(0) = p_0, G'(0) = p_1, \frac{G''(0)}{2} = p_2.$$

Let X, Y be independent nonnegative, integer-value random variable.

$$\begin{aligned} T =& X + Y \\ \mathbb{E}[S^T] =& \mathbb{E}[s^{X+Y}] = \mathbb{E}[s^X s^Y] = \mathbb{E}[s^X]\mathbb{E}[s^Y] \end{aligned}$$

Example 1.8. Let $X_1, \ldots, X_n$ be i.i.d. Bernoulli random variable.

$$\begin{aligned} T =& X_1 + \ldots + X_n \\ \mathbb{E}[s^T] =& \mathbb{E}[s^{X_1+\ldots+X_n}] = (\mathbb{E}[s^X])^n = (1 - p + ps)^n \\ \mathbb{E}[S^{X_1}] =& s^0(1-p) + sp \end{aligned}$$

Let $X_n \sim \text{Binomial}(n, p_n)$, $p_n \to 0$, $np_n \to \lambda$, $n \to \infty$.

$$\begin{aligned} G_n(s) &= \mathbb{E}[s^{X_n}] \\ &= (1 - p_n + p_n s)^n \\ &= \left(1 - \frac{np_n}{n} + \frac{np_n s}{n}\right)^n \\ &= \left(1 - \frac{np_n(1-s)}{n}\right)^n \to e^{-\lambda(1-s)} \end{aligned}$$

as $n \to \infty$.

$X \sim \text{Poisson}(\lambda)$,

$$G(s) = \mathbb{E}(s^X) = \sum_{k=0}^{\infty} s^k P(x = k) = \sum_{k=0}^{\infty} s^k \frac{e^{-\lambda}\lambda^k}{\lambda!} = e^{-\lambda}e^{\lambda s} = e^{-\lambda(1-s)}.$$

1.4 Cumulative Distribution Function (c.d.f.)

Definition 1.9. Given a random variable X, its *cumulative distribution function (c.d.f.)* is defined as

$$F(t) := P(X \leq t), -\inf < t < \inf.$$

Given a Borel set A, we have

$$F(A) = P(X \in A)$$

For example, $P(X \in (a, b]) = F(b) - F(a)$.

Definition 1.10 (Convergence in distribution)**.** Let X_n be a sequence of random variables, X be a random variable. Let F_n be the cdf of X_n and F be the cdf of X. We can X_n converges to X in distribution (written as $X_n \xRightarrow{D} X$, or $X_n \to X$), if

$$\begin{aligned} & F_n(t) \to F(t), \forall t \in \mathcal{C}(F)(\text{ the continuous domain of } F) \\ & \mathbb{E}[h(X_n)] \to \mathbb{E}[h(X)], \forall \text{ continuous function of } h \end{aligned}$$

Definition 1.11. We say X_n converges to X in (total) variation if $d_{TV}(X_n, X) \to 0$ as $n \to \infty$.

Example 1.12. Let X_n constant random variable $1/n$ and $X = 0$. For every n, we have

$$d_{TV}(X_n, X) = \sup_A |F_n(A) - F_X(A)|$$

and $P(X_n = 0) = 0$, $P(X = 0) = 1$, so X_n does not converge to X in variation.

Given that $\mathcal{C}(F_X) = (-\infty, 0) \cup (0, \infty)$, we have for every $t \in \mathcal{C}(F_X)$, and for all large n,

$$\begin{cases} F_n(t) = 1, \text{ if } t \in (0, \infty) \\ F_n(t) = 0, \text{ if } t \in (-\infty, 0) \end{cases}$$

Hence, $F_n(t)$ converges to $F_X(t)$ for every $t \in \mathcal{C}(F)$, so $X_n \overset{D}{\Longrightarrow} X$.

Example 1.13. If you have an n-sided die labelled $1/n, 2/n, \ldots, n/n$. Then notice that

$$X_n \overset{D}{\Longrightarrow} U \sim \text{Uniform}(0, 1)$$

because if we consider any $t \in (0, 1)$, $F_U(t) = t$, and $F_n(t) = \frac{k}{n}$ where $(k-1)/n < t \leq k/n$. As $n \to \infty$, k/n converges to t.

Again X_n does not converge to X in variation. Let Q be the set of rational numbers. $P(X \in Q) = 0$ because Q has measure zero, but $P(X_n \in Q) = 1$. Hence $d_{TV}(X_n, X) = 1$ for every n.

1.4.1 Geometric Distribution to Exponential, the Memoryless variables

Let $T_n \sim \text{Geo}\, p_n$, then $\Pr[T_n = k] = (1-p_n)^{k-1} p_n, k = 1, 2, \ldots$, $\Pr[T_n > k] = (1-p_n)^k$, and $\Pr[T_n > k + j | T_n > k] = \Pr[T_n > j]$. Also, $\mathbb{E}[T_n] = 1/p_n$.

And let $X \sim \text{Exp}(\lambda)$, $f_x(t) = \lambda e^{-\lambda t}, t \geq 0$, $\Pr[X > t] = e^{-\lambda t}, \Pr[X > t + s | X > t] = \Pr[X > s]$, $\mathbb{E}[X] = 1/\lambda$.

We will show

$$\frac{T_n}{n} \overset{D}{\to} X \sim \text{Exp}(\lambda).$$

First, let F_n be the c.d.f. of T_n and F_x be the c.d.f. of X. We needto show that $F_n(t) \to F_x(T)$ for all $t \in C(X)$.

Proof.

$$\begin{aligned} F_n(t) &= \Pr\left[\frac{T_n}{n} > t\right] = \Pr[T_n > nt] = \Pr[T_n > \lfloor nt \rfloor] \\ &= (1 - \frac{np_n}{n})^{\lfloor nt \rfloor} = (1 - \frac{\lambda_n}{n})^{\lfloor nt \rfloor} \to e^{-\lambda t} \end{aligned}$$

where $\lambda_n := np_n \to \lambda$ as $n \to \infty$, and the convergence to $e^{-\lambda t}$ is by squeeze theorem. $\square$

1.5 Point Process

Consider $N \sim \text{Poisson}(\lambda)$ and let $X_1, \ldots, X_N$ be i.i.d. $\text{Bernoulli}(p)$. Define $Y = \sum_{i=1}^{N} X_i$. Then if consider N, for each of its count, it has p chances to be 1 and $(1-p)$ to be 0, where we can split N into two Poisson distribution

$$
\begin{aligned}
Y &\sim \text{Poisson}(\lambda p) \\
Z &\sim \text{Poisson}(\lambda(1-p))
\end{aligned}
$$

where $Z := N - Y$ and we have $Z \perp Y$ (seen that in homework 1).

Definition 1.14. A *point process* on $[0, \infty)$ is a mapping, assigning each Borel set $J \subseteq [0, \infty)$, a nonnegative extended integer valuesd r.v. $N(J) = N_J$, so that if $J_1, J_2, \ldots,$ are disjoint, then

$$
N(\cup_i J_i) = \sum_i N(J_i)
$$

A counting process associated with N (family of random variables), $N(t) = N_t$ for $t \geq 0$ where $N(t) = N((0, t])$ for $t > 0$. By convention, the sample paths are right continuous.

Definition 1.15. A *Poisson point process* with intensity $\lambda > 0$ is a point process with:

a) If $J_1, J_2, \ldots,$ are nonoverlapping intervals, then $N(J_1), N(J_2), \ldots,$ are independent.

b) $N(J) \sim \text{Poisson}(\lambda |J|)$ where J is the length of the interval J.

Given a Poissson Point Process above, let $0 = T_0 < T_1 \leq T_2 \leq T_3 \leq \ldots$ be the time i^{th} customer arrives and $\tau_n = T_n - T_{n-1}$. Then $\tau_1, \tau_2, \ldots,$ are i.i.d. $\exp(\lambda)$.

Example 1.16. Let $N(t)$ be the number of customers arriving during $(0, t]$ and $N \sim \text{Poisson}(5)$. The probability of 0 arrivals up to time 2 is

$$
\Pr[N(2) = 0] = e^{-5(2)} = e^{-10}
$$

While the probability of k arrivals up up time 2 is

$$
\Pr[N(2) = k] = \frac{e^{-10} 10^k}{k!}.
$$

Consider

$$
\begin{aligned}
&\{N(5) = 7 | N(2) = 1\} \\
&\{N((2, 5]) = 6 | N(2) = 1\} \\
&\Pr[N(5) - N(2) = 6 | N(2) = 1] \\
=&\Pr[N(5) - N(2) = 6] \\
=&\Pr[N((2, 5]) = 6] \\
=&\Pr[N(3) = 6]
\end{aligned}
$$

We can also consider

$$\Pr[T_2 > 5.8|T_1 = 3.7] = \Pr[\tau_2 > 2.1|\tau_1 = 3.7] = e^{-\lambda(2.1)}$$

If you look at the store a 100 min, when will the next customer arrive?
We expect $\frac{1}{\lambda} = \frac{1}{5}\text{hr} = 12\text{min}$.

$$\begin{aligned} \Pr[X_1 > t] = \Pr[N(t) = 0] &= e^{-\lambda t}, t \geq 0 \\ \Pr[X_2 > t|X_1 = s] &= \Pr[N((s, s+t]) = 0|X_1 = s] \\ &= \Pr[N((s, s+t]) = 0] \\ &= e^{-\lambda t} \end{aligned}$$

1.6 Bernoulli and Poisson

Let $X_1, X_2, \ldots$, be Bernoulli Process with $p \in (0, 1)$.
Question:

a) Is $\Pr[X_n = k|T = n]$ equal $\Pr[X_T = k|T = n]$? Yes.
Let $A = \{w \in \Omega : X_n(w) = k\}, B = \{w \in \Omega : T(w) = n\}, C = \{w \in \Omega : X_{T(w)}(w) = k\}$ and $A \cap B = \{w \in \Omega : X_n(w) = k, T(w) = n\}, C \cap B = \{w \in \Omega : X_{T(w)}(w) = k, T(w) = n\}$, which implies $\Pr[A \cap B]/\Pr[B] = \Pr[C \cap B]/\Pr[B]$

b) Is $\Pr[X_n = k|T = n]$ equal to $\Pr[X_n = k]$? No. e.g. $T := \min\{n : X_n = 1\}$, and $\Pr[X_n = 1|T = n] = 1, \Pr[X_n = 1] = p$.
e.g. $X_i \sim \text{Exp}(\lambda)$ where $X_1, X_2, \ldots$, are event times.

$$\begin{aligned} \Pr[X_2 > t|X_1 = s] &= \Pr[N((s, s+t]) = 0|X_1 = s] \\ &= \Pr[N(s, s+t] = 0] \text{ by independent increment} \\ &= \Pr[N(X_1, X_1 + t] = 0|X_1 = s] \end{aligned}$$

But then let $T := \min\{r : N(r, r+t] = 10\}$. We have

$$\Pr[N(T, T+t) = 0|T = 3.87] = 0, \Pr[N(3.87, 3.87 + t] = 0] = e^{-\lambda t}$$

Definition 1.17. Let $0 = T_0 < T_1 = \tau_1 \leq T_2 = \tau_1 + \tau_2 \leq \ldots$ be the *occurence times* of a Poisson process which are the successive times $N(t)$ jumps. Let $\tau_1, \tau_2, \ldots$ be the *interoccurence time*, where $\tau_i := T_i - T_{i-1}$.

Theorem 1.18 (Interoccurence Time Theorem)**.**

(A) Interoccurence times $\tau_1, \tau_2, \ldots$, of a Poisson process with rate λ are i.i.d. $\text{Exp}(\lambda)$

(B) Let $Y_1, Y_2, \ldots$, be i.i.d. $\text{Exp}(\lambda)$.

$$N(t) := \max\{n : \sum_{i=1}^{n} Y_i \leq t\} \implies \{N(t)\}_{t \geq 0} \text{ is a Poisson counting process with rate } \lambda > 0$$

Example 1.19. Consider Bernoulli processes $\{X_k^m\}_{k\in\frac{\mathbb{N}}{m}}$ withe paremeter $p_m \in (0,1)$. Then $\tau_1^m = T_1^m = \min\{n \in \frac{\mathbb{N}}{m} = X_n^m = 1\}$ Then $m\tau_1^m \sim \text{Geo}(p_m)$. Let $T_2^m = \min\{n > T_1^m : X_n^m = 1\}$ and $\tau_2^m = T_2^m - T_1^m$, then $m\tau_2^m \sim \text{Geo}(p_m)$ as well. Then

$$N^m(t_1) \sim \text{Binomial}(?\,, p_m)$$

Useful later: $\{T_1 \geq t_1, T_2 \geq t_2\} \iff \{N(t_1) \geq 1, N(t_2) \geq 2\}$.

Theorem 1.20 (The law of small numbers for Bernoulli Process)**.** Let $\{X_r^m\}_{r\in\mathbb{N}/m}$ be a Bernoulli Process with parameter p_m indexed by multipliers of $\mathbb{N}/m$. Let $N^m(t)$ be the corresponding counting process. If $mp_m \to \lambda > 0$, then the counting process N^m converge in distribution to the counting process of a Poisson process with rate $\lambda > 0$ in the following sense:

$$\forall n, 0 = t_0 < t_1 < \ldots < t_n, (N^m(t_1), \ldots, N^m(t_n)) \xrightarrow{D} (N(t_1), \ldots, N(t_n))$$

Proof of Interoccurence Time Theorem.

a) We showed in the previous section that for a geometric r.v. with p_n with $np_n \to \lambda$. $T_n/n \xrightarrow{D} \text{Exp}(\lambda)$. And we have seen that the interoccurence times of Bernoulli $\{X_k^m\}_{k\in\mathbb{N}/n}$ are geometric, $\Delta_k^m = N^m(t_k) - N^m(t_{k-1}) \sim \text{Binomial}(m(t_k - t_{k-1}) \pm 1, p_m)$ where $\pm$ considers the rounding of $m(t_k - t_{k-1})$. And this converges in distribution to $\Delta_k \sim \text{Poisson}(\lambda(t_k - t_{k-1}))$. Thus the occurence time of $N^m(t)$ converges to $N(t)$ in distribution. Thus, the iteroccurence time of X_k^m, which is the iteroccurence time of $N^m(t)$, converging to $\text{Exp}(\lambda)$ implies that the interoccurence time of $N(t)$ converges to $\text{Exp}(\lambda)$.

b) With a Poisson process with rate λ, and let τ_i be its interoccurence times, and we know $\tau_i \overset{iid}{\sim} \text{Exp}(\lambda)$. Let Y_i be another sequence of i.i.d. expenentials with λ. Then since τ_i and Y_i have the same joint distribution, we also have

$$(\tau_1, \tau_1 + \tau_2, \ldots, \sum_{i=1}^n \tau_i \overset{D}{=} (Y_1, Y_1 + Y_2, \ldots, \sum_{i=1}^n Y_i)$$

But $(Y_1, Y_1 + Y_2, \ldots, \sum_{i=1}^n Y_i)$ determines the joint distribution of the occurrence time of $N(t)$. That is, the occurence times of $N(t)$ are the occurence times of a Poisson distribution. So $N(t)$ is Poisson.

$\square$

Given B), now we can simulate Poisson with $U_i \overset{D}{\sim} \text{Uniform}([0,1])$ and have $\tau_i = -\frac{1}{\lambda}\log(1 - U_i)$. However, if the actual $\lambda > \mu$ and we simulate with μ, then we have

$$\tilde{\tau}_i = -\frac{1}{\mu}\log(1 - U_i) \overset{D}{=} \frac{\lambda}{\mu}\tau_k$$

Theorem 1.21 (Generalized Thinning Theorem)**.** Let $N \sim \text{Poisson}(\lambda)$, X_i be iid r.v. with $\Pr[X_i = k] = p_k, k = 1, \ldots, m$ and $\sum_{i=1}^m p_k = 1$. And N independs X_i for all i. Let $N_k = \sum_{j=1}^N \mathbb{1}_{\{X_j = k\}}$. e.g:

$$
\begin{array}{cccccc}
m=3 & x_1 & x_2 & x_3 & x_4 & x_5 \\
N=5 & 2 & 3 & 3 & 1 & 2
\end{array}
$$

then $N_1 = 1$, $N_2 = 2$, $N_3 = 2$, $N_1 + N_2 + N_3 = N$.
then we have $N_1, \ldots, N_m$ are independent Poisson r.v. with $E[N_k] = \lambda p_k$. (You can consider this as splitting a Poisson process into m different ones with probability p_k.)
And we have

$$
\begin{aligned}
\Pr[N_1 = j_1, N_2 = j_2, \ldots, N_m = j_m] &= \Pr[N = j_1 + \ldots + j_m, N_1 = j_1, \ldots, N_m = j_m] \\
&= \underbrace{\Pr[N = j_1 + \ldots + j_m]}_{\text{Poisson}} \underbrace{\Pr[N_1 = j_1, \ldots, N_m = j_m | N = \sum_{i=1}^{m} j_i]}_{\text{multinomial}} \\
&= \frac{e^{-\lambda}\lambda^{j_1+\ldots+j_m}}{(j_1+\ldots+j_m)!} \binom{j_1+\ldots+j_m}{j_1, \ldots, j_m} p_1^{j_1} \ldots p_m^{j_m} \\
&= \prod_{i=1}^{m} \frac{e^{-\lambda p_i}\lambda^{j_i}}{j_i!}
\end{aligned}
$$

Scecond Construction Let $m_1, m_2, \ldots$ be iid $\mathrm{Poisson}(\lambda)$. Let $U_1, U_2, \ldots$ be iid $\mathrm{Uniform}(0,1)$ such that $(m_1, m_2, \ldots)$ independs $(U_1, U_2, \ldots)$. Put points at $U_1, \ldots, U_{m_1}$ if $m_1 > 0$. Put points at $1 + U_{m_1+1}, \ldots, 1 + U_{m_2}$ if $M_2 > 0$ and so on.

Claim 1.21.1. Above points form a Poisson point process (THM 7 of UChichago Notes).

Proof. $0 = t_1 < t_1 < \ldots < t_n = 1, J_k = (t_{k-1}, t_k] \implies p_k = t_k - t_{k-1}$. $N(J_1), \ldots, N(J_n)$ independent Poisson $\mathbb{E}[N(J_k)] = \lambda p_k = \lambda |J_k|$. □

Definition 1.22. Poisson point process on $\mathbb{R}^k$ with mean measure Λ is a point process on $\mathbb{R}^k$ with

1. $J_1, J_2, \ldots$ disjoint Borel sets in $\mathbb{R}^k$; $N(J_1), N(J_2), \ldots$ are independent.
2. $N(J_k) \sim \mathrm{Poisson}(\Lambda(J_k))$

Proposition 1.23. To show a point process is a Poisson point process, it suffices to verify the conditions above for rectangles J, J_i with sides parallel to the coordinate axes.

Example 1.24. Let T_i be the occurence times of a Poisson process on $[0, \infty)$ with rate λ. Let S_j be the iid rv with CDF F. S_j, T_i are indep. Then we have $J = [t_1, t_2] \times [s_1, s_2]$. So $N(J) = \lambda(t_2 - t_1) \times (s_2 - s_1)$, where $J' \cap J = \emptyset$ implies $N(J)$ independent $N(J')$.

For a Poisson Point Process on $\mathbb{R}$ with rate $\lambda > 0$, then given $t > 0$, we have

$$
\begin{aligned}
\Pr[N(0,t] = 0] &= e^{-\lambda t} \\
\Pr[N(-t,0] = 0] &= e^{-\lambda t} \\
\Pr[N(-t,t] = 0] &= e^{-2\lambda t}
\end{aligned}
$$

Given 2 Poisson Processes on $[0, \infty)$ with $N \sim \mathrm{Poisson}(\lambda)$, $M \sim \mathrm{Poisson}(\mu)$, $\lambda > \mu$, how can we comply them so $N(J) \geq M(J)$ for every Borel set J?

1. Superposition: Consider M as above and $L \sim \mathrm{Poisson}(\lambda - \mu)$, which are independent, then take the superposition (a process made of all success of M, L) so we get another $\mathrm{Poisson}(\lambda)$.

2. Decomposition: With the N above, for each success of N, split it to M with probability μ/λ, and L with $(1 - \mu/\lambda)$, then M and L are independent Poisson Processes and M is what's required.

Consider N, M with the distributions above, let T_1 be the time of first success of N, then what's the probability that $M(T_1) = k$? If we directly compute it, it will be

$$\Pr[M(T_1) = k] = \int_0^\infty \Pr[M(T_1) = k | T_1 = s] \underbrace{\lambda e^{-\lambda s}}_{\Pr[T_1 = s]} ds$$

which is that easy to compute. But we can consider $N + M \sim \mathrm{Poisson}(\lambda + \mu)$. And split its success to N, M with probability $\frac{\lambda}{\mu+\lambda}$ and $\frac{\mu}{\mu+\lambda}$ respectively. Then T_1 is the time when a success is splitted to N the first time. That is, $M(T_1 = k)$ can be considered as a geometric process with k failure and one success, so

$$\Pr[M(T_1) = k] = \left(\frac{\mu}{\lambda + \mu}\right)^k \left(\frac{\lambda}{\lambda + \mu}\right)$$

Let $\{N(t)\}_{t \geq 0}$ be a counting process on $[0, \infty)$. Prove or disprove: If $N(t) \sim \mathrm{Poisson}(\lambda t)$ for all $t > 0$, then N is a Poisson Process.
Let T_i be the occurence times and τ_i be the interoccurence times as before. Then $T_n = \tau_1 + \ldots + \tau_n$. If τ_i are independent $\mathrm{Exp}(\lambda)$, we know $T_n \sim \mathrm{Erlang}(n, \lambda)$, so $\mathbb{E}[T_n] = n/\lambda$ and

$$F_n(t) = \Pr[T_n \leq t] = \Pr[N(t) \geq n] = \sum_{k=n}^{\infty} \frac{e^{-\lambda t (\lambda t)^k}}{k!} = 1 - \sum_{k=0}^{n-1} \frac{e^{-\lambda t (\lambda t)^k}}{k!}$$

so if $T_1, T_2, \ldots,$ have the "right" distribution, then $N(t)$ will be $\mathrm{Poisson}(\lambda t)$. What if we don't have the independence? Consider $T_i := F_i^{-1}(U)$ where F_i is the cdf of $\mathrm{Erlang}(i, \lambda)$ and $U \sim \mathrm{Uniform}(0, 1)$. Then it's not hard to see that each $T_i \sim \mathrm{Erlang}(i, \lambda)$, however, once T_1 is given, we can compute U_1 and hence all $T_2, T_3, \ldots$ are know, so the process with T_i being the occurence time is not a Poisson.

limits of expectation and expectation of limits

Theorem 1.25 (Monotone Convergence Theorem)**.** Let $\{X_n\}_{n \geq 1}$ be a sequence of random variables such that for all $n \geq 1$,

$$0 \leq X_n \leq X_{n+1}, \text{ Probably a.s.},$$

then

$$\mathbb{E}[\lim_{n \to \infty} X_n] = \lim_{n \to \infty} \mathbb{E}[X_n].$$

Theorem 1.26 (Deominated Convergence Theorem)**.** Let $\{X_n\}_{n\geq 1}$ be a sequence of random variable such that for all ω outside a set $\mathcal{N}$ of null probability there exists $\lim_{n\to\infty} X_n(\omega)$ and such that for all $n \geq 1$

$$|X_n| \leq Y, \text{ Probably a.s.},$$

where Y is some integrable random variable. Then

$$\mathbb{E}[\lim_{n\to\infty} X_n] = \lim_{n\to\infty} \mathbb{E}[X_n].$$

Example 1.27 ("Counter Example")**.** Suppose we are rolling a fair dice independently. Every time we get 6, we lost all money, otherwise, we double the current amount. Starting with $X_0 = 100$, we have

$$X_n = \begin{cases} 100 * 2^n, & \text{with prob } (5/6)^n \\ 0, & \text{with prob } 1-(5/6)^n \end{cases}$$

$$\begin{aligned} \mathbb{E}[X_n] &= 100 * (5/3)^n \\ \lim_{n\to\infty} \mathbb{E}[X_n] &= \infty \\ \mathbb{E}[\lim_{n\to\infty} X_n] &= 0 \end{aligned}$$

where the last inequality is by $\lim_{n\to\infty} \Pr[X_n = 1] = 0$ and $\lim_{n\to\infty} \Pr[X_n = 0] = 1$, so $X_n \to 0$ almost surely.

Let N be a Poisson on $[0,\infty)$ with rate λ. Let $T \geq 0$ be a r.v. such that N, T are independent. If we know the distribution of $N(T)$, can we determine the distribution of T? First consider the probability generating function (p.g.f.) of a Poisson $X \sim \text{Poisson}(\lambda)$, we have

$$G(s) = \mathbb{E}[s^X] = \sum_{k=0}^{\infty} s^k \Pr[X = k] = \sum_{k=0}^{\infty} s^k \frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda}e^{\lambda s} = e^{-\lambda(1-s)}$$

Or let x be nonnegative, integer-valued r.v.m the Laplac-Stieltjes Transformation of X is

$$L(s) = \mathbb{E}[e^{-sX}] = \int_0^{\infty} e^{-st} dF(t) = \int_0^{\infty} e^{-st} F(dt)$$

note this formula prevent us from worrying about the continuity of X by $F(t)$.
Recall the moment generating function (m.g.f.) $m(\theta) = \mathbb{E}[e^{(\theta t}]$. We give some examples,

Example 1.28.

1. When $\Pr[T = t] = 1$, we have $\mathbb{E}[e^{-sT}] = e^{-s}$.

2. When $T \sim \text{Bernoulli}(p)$,

$$L(s) = \mathbb{E}[e^{-sT}] = (1-p) * 1 + p * e^{-s} = \int_{[0,\infty)} e^{-st} dF(t)$$

3. $T \sim \text{Binomial}(n, p)$. $T = X_1 + \ldots + X_n$.

$$\begin{aligned}
L(S) &= \mathbb{E}[e^{-sT}] \\
&= \int_{[0,\infty)} e^{-st} d(t) \\
&= \mathbb{E}[e^{-s(x_1+\ldots+x_n}] \\
&= \mathbb{E}[e^{-sx_1} \ldots e^{-sx_n}] \\
&= \mathbb{E}[e^{-sx_1}] \ldots \mathbb{E}[e^{-sx_n}] \\
&= (1 - p + pe^{-s})^n
\end{aligned}$$

4. Let $X \sim \text{Exp}(\lambda)$, we have

$$\mathbb{E}[e^{-sX}] = \int_0^\infty e^{-st} \lambda e^{-\lambda t} dt = \frac{\lambda}{s+\lambda}.$$

Lemma 1.29. Given a $N(T) \sim \text{Poisson}(\lambda)$, we have $L(s) = G(1 - s/\lambda)$.

Proof.

$$\begin{aligned}
G(s) &= \mathbb{E}[z^{N(T)}] \\
&= \mathbb{E}[\mathbb{E}[z^{N(T)}|T]] \\
&= \mathbb{E}[e^{-\lambda T(1-z)}] \\
&= L(\lambda(1-s))
\end{aligned}$$

where the second last inequality is by

$$G(z) = \mathbb{E}[z^{N(T)}] = \sum_{k=0}^\infty z^k \frac{e^{-\lambda t}(\lambda t)^k}{k!} = e^{-\lambda t(1-z)}.$$

And then let $s = \lambda(1-z)$, we are done. □

Thus, when $N(T) \sim \text{Poisson}(\lambda t)$,

$$L(s) = G(1 - s/\lambda) = e^{-\lambda t(1-(1-s/\lambda))} = e^{-st}$$

so $\Pr[T = t] = 1$.

Theorem 1.30 (Not gonna prove)**.** Like p.g.f. and m.g.f., $L(s)$ uniquely corresponds to a random distribution.

Example 1.31. Let $\Pr[N(T) = k] = \rho^k(1-\rho)^k, k = 0, 1, \ldots$. Then

$$G(z) = \mathbb{E}[z^{N(T)}] = \sum_{k=0}^\infty z^k \rho^k (1-\rho) = \frac{1-\rho}{1-\rho z}.$$

$$\begin{aligned}
L(s) = \mathbb{E}[e^{-sT}] = G(1 - s/\lambda) &= \frac{1-\rho}{1-\rho(1-s/\lambda)} \\
= \frac{1-\rho}{1-\rho+\rho s/\lambda} &= \frac{\frac{\lambda}{\rho}(1-\rho)}{\frac{\lambda}{\rho}(1-\rho)+s}
\end{aligned}$$

which shows that $T \sim \text{Exp}(\frac{\lambda}{\rho}(1-\rho))$.

2 Markov-Chain

Let $X_0, X_1, \ldots$ be discrete-time stochastic processes and let the state space be countable.

$$\Pr[X_0 = i_0, \ldots, X_n = i_n], \forall n,\ i_0, \ldots, i_n \in \text{ state space.}$$

1. Markov Property:
$$\Pr[\underbrace{X_{n+1} = j}_{\text{future}} | \underbrace{X_n = i_n}_{\text{present}}, \underbrace{\ldots, X_0 = i_0}_{\text{past}}] = \Pr[X_{n+1} = j | X_n = i_n]$$

2. Time Homogeneity:
$$\Pr[X_{n+1} = j | X_n = i] = \Pr[X_1 = j | X_0 = i] = \Pr(i, j)$$

Definition 2.1. $X_0, X_1, \ldots$ is a *discrete-time Markov chain (DTMC)* if $X_0, X_1, \ldots$ has the two properties above.

Example 2.2. Let $X_0, X_1, \ldots$ be an independent Bernoulli process with parameter p. Then the state space is $\{0, 1\}$.

$$\begin{aligned}
\Pr[X_{n+1} = j | X_n = i_n, \ldots, X_0 = i_0] &= \Pr[X_{n+1} = j] \\
\Pr[X_{n+1} = j | X_n = i_n] &= \Pr[X_{n+1} = j] \\
\Pr[X_{n+1} = j | X_n = j] &= \Pr(i, j).
\end{aligned}$$

This forms a really special DTMC, basically every r.v. are i.i.d.. Its transition matrix looks like

$$P = \begin{bmatrix} 1-p & p & \ldots \\ 1-p & p & \ldots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

where the rows represent the ”from” and columns represent the ”to”. That is, $[P]_{ij} = \Pr(i, j)$.

Example 2.3. Let $X_0, X_1, \ldots \sim \text{Bernoulli}(p), p \in (0, 1)$. $Y_n = X_n + X_{n+1} \in \{0, 1, 2\}$. Is $Y_0, Y_1, \ldots$ a Markov Chain? No.

$$\begin{aligned}
\Pr[Y_2 = 0 | Y_1 = 1, Y_0 = 0] = 0 \\
\Pr[Y_2 = 0 | Y_1 = 1, Y_0 = 2] = 1 - p
\end{aligned}$$

because $Y_0 = 0, Y_1 = 1$ implies that $X_2 = 1, X_0 = X_1 = 0$, first probability is the probability that $X_3 = -1$ and the second one is the probability that $X_3 = 0$.
What can we add to make it a DTMC?
Acquire more information. Let $Z_n = (X_n, Y_n)$, then we consider

$$\Pr[Z_{n+1} = (j_1, j_2) | Z_n = (i_1, i_2), Z_{n-1} = (k_{n-1}, \ell_{n-1}), \ldots, Z_0 = (k_0, \ell_0)]$$

And the transition matrix is

	(0,0)	(0,1)	(1,0)	(1,1)
(0,0)	1-p	p	0	0
(0,1)	0	0	1-p	p
(1,0)	1-p	p	0	0
(1,1)	0	0	1-p	p

$M/M/1$ **Queue** Consider an $M/M/1$ queue, which is the queue with customers arriving according to $\text{Poisson}(\lambda)$, service time following i.i.d. $\exp(\mu)$ with 1 server. The model records the number of customers whenever a process (arrival or service) is done. For example, if we have $X_0 = 0$ and the next event is finishing a service, $X_1 = 0$, if it's a customer arrival, $X_1 = 1$. This model is also called the birth and death model, basically we add one when we have a birth and minus one when we have a death. Since the moment starts, we can only have "deaths" (or departures) until the first arrival. That is, given $X_n = 0$, the probability that $X_{n+1} = 0$ is the probability that

$$\Pr[D < A] = \frac{\mu}{\lambda + \mu}$$

where $D \sim \exp(\mu)$ is the service time and $A \sim \exp(\lambda)$ is the interoccurence time of $\text{Poisson}(\lambda)$ (i.e. the arrival time). Similarly, given $X_n = 0$, the probability that $X_{n+1} = 1$ is the probability that the customer arrives before the service time. So the transition matrix looks like

$$\begin{bmatrix} \frac{\mu}{\lambda+\mu} & \frac{\lambda}{\lambda+\mu} & 0 & \dots & \dots & \\ \frac{\mu}{\lambda+\mu} & 0 & \frac{\lambda}{\lambda+\mu} & 0 & \dots & \\ 0 & \frac{\mu}{\lambda+\mu} & 0 & \frac{\lambda}{\lambda+\mu} & 0 & \dots \end{bmatrix}$$

where rows and columns are from 0 to infinity.

We can also consider $X_n :=$ number of customers in the system just before n-th arrival. For example, given $X_n = 0$, the probability $X_{n+1} = 0$ is $\frac{\mu}{\lambda+\mu}$, because $X_n = 0$, so between n-th and $n+1$th arrival, there is at most one customer in the system, and we have the probability $\frac{\mu}{\lambda+\mu}$ to finish the service before $n+1$-th arrival, otherwise, with probability $\frac{\lambda}{\mu+\lambda}$, we still have a customer in the system when $n+1$-th customer arrives.

Another way of considering this is treating the arrivals as a geometric distribution with $\frac{\lambda}{\lambda+\mu}$ success rate. For example, if $X_n = 1$. That means between n and $n+1$ arrivals, there are 2 customers in the system, and we do the geometric experiment. The probability that there is no customer in the system when $n+1$th customer arrives is the probability we "fail" at least twice before the "success". Similarly, the probability that there is one customer in the system when $n+1$th customer arrives is the prob that we "fail" exactly once before the first success, and so on. So the transition matrix looks like:

$$\begin{bmatrix} \frac{\mu}{\lambda+\mu} & \frac{\lambda}{\lambda+\mu} & 0 & \dots & \dots \\ \left(\frac{\mu}{\lambda+\mu}\right)^2 & \frac{\mu\lambda}{(\lambda+\mu)^2} & \frac{\lambda}{\lambda+\mu} & 0 & \dots \\ \left(\frac{\mu}{\lambda+\mu}\right)^3 & \left(\frac{\mu}{\lambda+\mu}\right)^2 \frac{\lambda}{(\lambda+\mu)} & \frac{\mu\lambda}{(\lambda+\mu)^2} & \frac{\lambda}{\mu+\lambda} & \dots \end{bmatrix}$$

$M/M/1/3$ **Queue** Consider the $M/M/1/3$ queue where the 3 means the capacity of the system. Let $Y_n :=$ number of customers in the system just after the n-th departure, so now the state space is $\{0, 1, 2\}$. Then let's say $X_n = 0$, then the probability $X_{n+1} = 1$ is the probability that there is an arrival between n-th and $n+1$-th departures. In other words, for $n+1$-th departure to happen, there has to be an arrival, so the probability is actually the probability that the $(n+1)$-th departure happen before any arrivals except for the necessary one, which is $\frac{\mu}{\lambda+\mu}$, similar to other cases. So the transition matrix looks like:

$$\begin{bmatrix} \frac{\mu}{\mu+\lambda} & \frac{\lambda\mu}{\mu+\lambda} & \left(\frac{\lambda}{\lambda+\mu}\right)^2 \\ \frac{\mu}{\mu+\lambda} & \frac{\lambda\mu}{\mu+\lambda} & \left(\frac{\lambda}{\lambda+\mu}\right)^2 \\ 0 & \frac{\mu}{\mu+\lambda} & \frac{\lambda}{\mu+\lambda} \end{bmatrix}$$