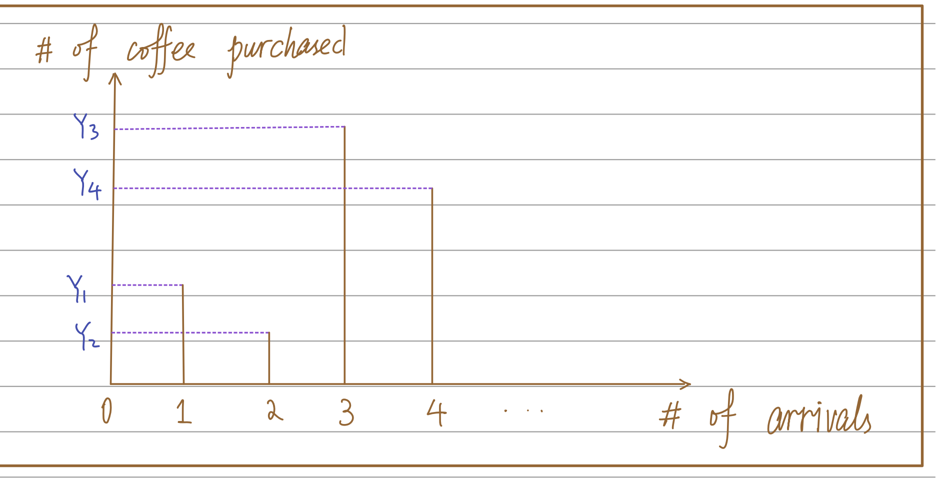
Lecture 21:

Compound Poisson Process, Thinning

Part I. Compound Poisson Process

Let us come back to a homogeneous Poisson Process with rate constant λ. For the i-th arrival to the Tim Hortons at DC, let Y_i be the number of coffee drinks purchased by the i-th consumer.

Q: What is the expectation of the total number of coffee drinks sold in a 1-hour time interval? What is the variance of it?



Theorem 21.1. Let $Y_1, Y_2, \cdots$ be i.i.d. and N be an independent non-negative integer-valued random variable such that $\mathbb{E}|Y_i|, \mathbb{E}N < \infty$. Define

$$S = \begin{cases} Y_1 + \cdots + Y_N, & \text{if } N \geq 1; \\ 0, & \text{if } N = 0. \end{cases}$$

Then the Wald's identity theorem implies

(i). $\mathbb{E}S = \mathbb{E}N \cdot \mathbb{E}Y_i$;

(ii). If $\mathbb{E}(Y_i^2), \mathbb{E}(N^2) < \infty$, then

$$var(S) = \mathbb{E}N \cdot var(Y_i) + var(N) \cdot (\mathbb{E}Y_i)^2.$$

(iii). In particular, if $N = \text{Poisson}(\lambda)$ and $\mathbb{E}(Y_i^2) < \infty$, then

$$var(S) = \lambda\, \mathbb{E}(Y_i^2).$$

Example 21.1. Suppose that the number of customers at an LCBO in a day has a Poisson distribution with mean 160 and that each customer spends an average \$30 with a standard deviation \$10. Then the average revenue for the day is $160 \times \$30 = \$4800$, with a standard deviation $\sqrt{160 \times [(\$30)^2 + (\$10)^2]} = \$400$.

Part II. Thinning.

Let $N(t)$ be the number of students arriving at the Tim Hortons at SLC by time t. Let $N_j(t)$ be the number of arrivals by t who bought j cups of coffee, i.e., $N_j(t) = \#\{i \le N(t) : Y_i = j\}$.

Q: Assume that arrivals follows Poisson Process with rate λ/min.

(a). how many arrivals on average for a 10-min interval?

A: 10λ.

(b). In 10 mins, how many arrivals on average bought exactly two cups of coffee?

A: Compound Poisson Process

Let $Z_i = \begin{cases} 1, & Y_i = 2; \\ 0, & Y_i \neq 2 \end{cases}$ (i.e. i-th arrival bought 0, 1, 3, ... cups of coffee)

Then $N_2(t) = \#\{ i \le N(t) : Y_i = 2 \}$

$$= \sum_{i=1}^{N(t)} Z_i = Z_1 + Z_2 + \cdots + Z_{N(t)}$$

By Theorem 21.1,

$$\mathbb{E}[N_2(10)] = \mathbb{E}[N(10)] \cdot \mathbb{E}Z = 10\lambda \cdot (1 \cdot \mathbb{P}(Z_i = 1) + 0 \cdot \mathbb{P}(Z_i = 0))$$

$$= 10\lambda \cdot \mathbb{P}(Y_i = 2).$$

(c). How many arrivals in 10 mins bought exactly two cups of coffee?

Theorem 21.2. If $\{N(t): t \geq 0\}$ is a Poisson Process with rate $\lambda(r)$ and for each arrival i, Y_i is i.i.d and have the same distributions with Y. Then $\{N_j(t): t \geq 0\}$ are independent Poisson Process with rate $\lambda(r) \cdot \mathbb{P}(Y=j)$, where $N_j(t) := \#\{i \leq N(t): Y_i = j\}$.

Answer of (c): By Theorem 21.2., $\{N_2(t): t \geq 0\}$ is a Poisson Process with rate $\lambda \cdot \mathbb{P}(Y=2)$. Thus,

$$N_2(10) = N_2(10) - N_2(0) = \text{Poisson}\left(\int_0^{10} \lambda \, \mathbb{P}(Y=2)\, dt\right)$$

$$= \text{Poisson}(10 \cdot \lambda \cdot \mathbb{P}(Y=2)).$$

Remark 21.1. One can check that the expectation of the answer in (c) coincides with the answer in (b).

Remark 21.2. Theorem 21.2 says not only we have mean $\int \lambda(r)\,\mathbb{P}(Y=j)\,dr$, but we know it is a Poisson Process which carries Poisson distributions (not necessarily exponential / gamma distributions if we have non-homogeneous Poisson Processes).

Example 21.2. Ellen catches fish at times of a Poisson Process with rate 2 per hour. 40% of the fish are salmon, while 60% are trout.

Q: What is the probability she will catch exactly one salmon and two trout if she fishes for 2.5 hours?

A: The total number of fish she catches in 2.5 hours follows Poisson(5), so the numbers of salmon and trout are independent Poisson with means 2 and 3. Thus, the probability is

$$e^{-2} \cdot \frac{2^1}{1!} \cdot e^{-3} \cdot \frac{3^2}{2!} = 9 \cdot e^{-5}.$$

Example 21.3. Two copy editors read a 300-page manuscript. The first found 100 typos, the second found 120, and their lists contain 80 erross in common. Suppose that the author's typos follow a Poisson Process with some unknown rate λ per page, while the two copy editors catch errors with unknown probabilities of success p_1 and p_2.

Q: What is approximalely λ? p_1? p_2? and the total number of undiscovered typos?

A: Let X_0 be the number of typos neither found; let X_1 and X_2 be the number of typos found only by editor 1 and only by editor 2, respectively; let X_3 be the number of typos that both found. Then

$$X_1 = 20, \quad X_2 = 40, \quad X_3 = 80.$$

If we let $\mu = 300\lambda$, then X_i are independent Poisson with mean

$$\mu(1-p_1)(1-p_2), \quad \mu p_1(1-p_2), \quad \mu(1-p_1)p_2, \quad \mu p_1 p_2.$$

Thus, $$p_1 = \frac{\mathbb{E}X_3}{\mathbb{E}(X_2+X_3)} \doteq \frac{80}{40+80} = \frac{2}{3};$$

$$p_2 = \frac{\mathbb{E}X_3}{\mathbb{E}(X_1+X_3)} \doteq \frac{80}{100} = \frac{4}{5};$$

$$\mu = \frac{\mathbb{E}X_3}{p_1 p_2} \doteq \frac{80}{\frac{2}{3}\cdot\frac{4}{5}} = 150;$$

$$\lambda = \frac{\mu}{300} \doteq \frac{1}{2}.$$

$$\mathbb{E}X_0 = \mu(1-p_1)(1-p_2) \doteq 150\cdot\frac{1}{3}\cdot\frac{1}{5} = 10.$$

This is the end of this lecture!