Lec 8

$N(t)$. poisson with rate λ.

Each arrivial associated with a iid. r.v. $Y_i \in \mathbb{N}$.

Ex. Y_i : # of people in the i-th group.

$\forall j \quad P_j = \Pr[Y_i = j]$.

$N_j(t)$ # of times with $Y_i = j$ before time t.

Thm. $N_j(t)$ are independent Poisson process with rate λP_j.

pf. Only need to prove that $Y_i \in \{0,1\}$.

Only need to check the distribution of

$$X_i = N_i(t+s) - N_i(s).$$

$$P[X_1 = j, X_2 = k] = e^{-\lambda t} \cdot \frac{(\lambda t)^{j+k}}{(j+k)!} \cdot \binom{j+k}{j} P_1^j P_2^k$$

$$= e^{-\lambda t} \cdot \frac{(\lambda t)^j \cdot (\lambda t)^k}{j! \cdot k!} P_1^j P_2^k$$

$$= e^{-\lambda P_1 t} \cdot \frac{(\lambda P_1 t)^j}{j!} \cdot e^{-\lambda P_2 t} \frac{(\lambda P_2 t)^k}{k!} \qquad (P_1 + P_2 = 1)$$

Example

Restaurant. rate 10 per hour.

40% male, 60% female.

Prob of exactly 12 male, 10 female. in 2 hours.

male in 2 hours Poisson (8).

female in 2 hours Poisson (12).

$$e^{-8} \cdot \frac{8^{12}}{12!} \cdot e^{-12} \cdot \frac{12^{10}}{10!} \approx 0.505\%.$$

Example of two Editors

Two editor read 300-page manuscript.

A 100 typos. B. 120 typos. 80 in common.

Typos follow Poisson (λ). A. P_A. B. P_B.

Estimates λ. P_A. P_B.

X_0 : neither found $\sim$ Poisson $\left(\mu(1-p_1)(1-p_2)\right)$.

X_1 : Only A found $\sim$ Poisson $(\mu p_1 (1-p_2))$.

X_2 : Only B found $\sim$ Poisson $(\mu (1-p_1) p_2)$.

X_3 : Both found $\sim$ Poisson $(\mu p_1 p_2)$.

* $\mu = 300\lambda$.

* $A \cup B = 100 + 120 - 80 = 140$.

* neither found $300\lambda - 140$.

* Only A : 20

* Only B : 40

* Both : 80

$300 \cdot \frac{8}{15} \lambda = 80$

$\frac{p_2}{1-p_2} = 4$

$$\left.\begin{array}{l} 300\lambda p_1 p_2 = 80 \\ 300\lambda(1-p_1)p_2 = 40 \\ 300\lambda p_1(1-p_2) = 20 \end{array}\right\} \Rightarrow \begin{array}{l} p_1 = 2/3 \\ p_2 = 4/5 \\ \lambda = 1/2. \end{array}$$

Why choose mean?

Maximum Likelihood 极大似然估计.

Given data x. Compute $\underset{\theta}{\operatorname{argmax}}\, p(x|\theta)$.

$$p(x|\theta) = e^{-\theta} \cdot \frac{\theta^x}{x!} \Leftrightarrow \underbrace{\log p(x|\theta)}_{g(\theta) \triangleq} = -\theta + x\log\theta - \log x!$$

$$g'(\theta) = \frac{x}{\theta} - 1 = 0 \Rightarrow \theta = x.$$

Coupon Collector Revisited

n coupons.

each with prob. P_i.

N - # of coupons collected until a complete set.

What is $\mathbb{E}[N]$.

N_j : # of coupons needed until meet type j.

$N = \max N_j$.

Suppose coupons are collected according to Poisson (1)

$N_j(t) \sim \text{Poisson}(P_j)$. independent!

X_j. time to get the first type j coupon.

$$\Pr[\underbrace{\max_j X_j}_{X} < t] = \prod_{j=1}^{n} \Pr[X_j < t]$$

$$= \prod_{j=1}^{n} (1 - e^{-P_j t}).$$

$$X = \sum_{i=1}^{N} T_i$$

$$\mathbb{E}[X] = \mathbb{E}[N] \cdot \mathbb{E}[T_i]$$

$$= \mathbb{E}[N].$$

Wald's equation.

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

$$\mathbb{E}[X] \stackrel{!}{=} \int_0^\infty P(X > t)\, dt = \int_0^\infty \left(1 - \prod_{j=1}^{n} (1 - e^{-P_j t})\right) dt.$$

$$X \geq 0 \Rightarrow X = \int_0^X dt = \int_0^\infty \mathbb{1}[t \leq X]\, dt \Rightarrow \mathbb{E}X = \mathbb{E}\int_0^\infty \mathbb{1}[X \geq t]\, dt$$

Conditioning

$T_1, T_2, \ldots$ arrivial times of a Poisson process with rate λ.

$U_1, \ldots, U_n$ ind. uniform on $[0, t]$.

$V_1 < V_2 \cdots < V_n$ rearrange of $U_1, \ldots, U_n$.

Thm. Condition on $N(t) = n$, $(T_1, \ldots, T_n)$ has the same distribution as $(V_1, \ldots, V_n)$.

pf. $\forall\, t_1 < t_2 < \cdots < t_n$.

$$\Pr\left[\bigwedge_{i=1}^{n} T_i = t_i \,\middle|\, N(t) = n\right]$$

$$= \Pr[\tau_1 = t_1,\ \tau_2 = t_2 - t_1,\ \ldots,\ \tau_n = t_n - t_{n-1},\ \tau_{n+1} > t - t_n] \Big/ \Pr[N(t) = n]$$

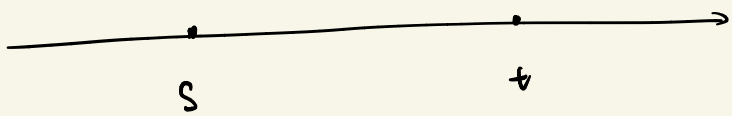
$$= \lambda e^{-\lambda t_1} \cdot \lambda e^{-\lambda(t_2 - t_1)} \cdots \lambda e^{-\lambda(t_n - t_{n-1})} \cdot e^{-\lambda(t - t_n)} \Big/ e^{-\lambda t} \frac{(\lambda t)^n}{n!}$$

$$= \frac{\lambda^n e^{-\lambda t} \cdot n!}{e^{-\lambda t} \cdot \lambda^n \cdot t^n} = \frac{n!}{t^n} = 1 \Big/ \frac{t^n}{n!}$$

$$\left| \{ (t_1, \ldots, t_n) \mid 0 < t_1 < t_2 < \cdots < t_n < t \} \right| = t^n / n!$$

Cor. If $s < t$ and $0 \le m < n$, then

$$p\left(N(s) = m \mid N(t) = n \right) = \binom{n}{m} \left(\frac{s}{t}\right)^m \left(1 - \frac{s}{t}\right)^{n-m}.$$



$$\sim \mathrm{Binom}\left(n, \frac{s}{t}\right).$$

Poisson Approximation

A sister thm.

m-balls-into-n bins. max load.

Each bin $X_i^{(m)} \sim \mathrm{Binom}(m, \frac{1}{n})$. subject to $\sum X_i = m$.

Let $Y_1, \dots, Y_n \sim \mathrm{Poisson}(\lambda)$ independently.

Thm. distribution of $(X_1, \dots, X_n)$

$= $ dist. of $(Y_1, \dots, Y_n)$ conditioned on $\sum Y_i = m$.

Pf. Fix $\vec{a} = (a_1, \dots, a_n) \geq 0$ such that $\sum a_i = m$.

$$\Pr[\vec{Y} = \vec{a} \mid \sum a_i = m] = \frac{\prod_{i=1}^{n} \Pr[Y_i = a_i]}{\Pr[\sum a_i = m]}$$

$$= \frac{\prod_{i=1}^{n} e^{-\lambda} \cdot \frac{\lambda^{a_i}}{a_i!}}{e^{-\lambda n} \cdot \frac{(\lambda n)^m}{m!}} = \frac{1}{n^m} \cdot \frac{m!}{a_1! \cdot a_2! \cdots a_n!}$$

$$= \Pr[\vec{X} = \vec{a}].$$

<u>Thm</u>. $\forall f : \mathbb{N}^n \to \mathbb{N}$.

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

Pf. $E[f(Y_1, \dots, Y_n)] = \sum_{k=0}^{\infty} E[f(Y_1, \dots, Y_n) \mid \sum Y_i = k] \cdot \Pr[\sum Y_i = k]$

$$\geq E[f(Y_1, \dots, Y_n) \mid \sum Y_i = m] \cdot \Pr[\sum Y_i = m]$$

$$= E[f(X_1, \dots, X_n)] \cdot \boxed{\Pr[\sum Y_i = m]}.$$

Let $\lambda = \frac{m}{n}$. $\sum Y_i \sim \text{Poisson}(m) \Rightarrow \square = e^{-m} \cdot \frac{m^m}{m!}$

$$\geq \frac{1}{e\sqrt{m}}.$$

Application

Maxload when $m = n$. $X = \max X_i$

$$\exists\, c_1, c_2.\quad \Pr\left[\frac{c_1 \log n}{\log\log n} \le X \le \frac{c_2 \log n}{\log\log n}\right] = 1 - O\left(\frac{1}{n}\right)$$

Pf. $\Pr[X > a] = \Pr[\exists i.\ X_i > a]$

$$\le n \cdot \Pr[X_1 > a] \le n \cdot \binom{n}{a} \cdot \left(\frac{1}{n}\right)^a \le n\left(\frac{e}{a}\right)^a$$

$$\left(\frac{e}{a}\right)^a \le \frac{c}{n^2} \iff 2\log n + a \cdot (1 - \log a) \le c'$$

$$a \log a \ge c \cdot \log n \quad \checkmark$$

Now prove $\Pr[X < a'] = O\left(\frac{1}{n}\right)$.

By Poisson Approx.

$$\Pr[Y < a'] = (\Pr[Y_1 < a'])^n$$

$$= (1 - \Pr[Y_1 \ge a'])^n \le (1 - \Pr[Y_1 = a'])^n = \left(1 - \frac{1}{ea'!}\right)^n$$

$$\le e^{-\frac{n}{ea'!}} \le e^{-\frac{n}{a'!}}$$