Jan 26

Remark Use Poisson distr. for rare successes, e.g.

\# typos on a page of a book

\# software failures on a given day, ...

Ex (in Ex. p. 34): of 40 students, ~~~~ 2 are sick on ave. $X \doteq$ # sick.

- $X \approx \text{Poisson}(2) \Rightarrow P\{X=4\} \approx e^{-2} \cdot \frac{2^4}{4!} = 0.0902$ (compare to 0.0901)

Ex
- $P\{\text{no sick students}\} = P\{X=0\} = e^{-2} = 0.14$.
- $P\{\text{at least } \cancel{\text{one}}\ \text{two sick}\} = 1 - P\{X=0\} - P\{X=1\} = 1 - e^{-2} - 2e^{-2} = 0.56$.

Prop (Sum of ~~Bernoulli~~ = ~~Binomial~~) $X \sim \text{Binom}(n,p),\ Y \sim \text{Binom}(m,p)$ indep $\Rightarrow X+Y \sim \text{Binom}(n+m, p)$

Prop (Sum of Poisson = Poisson)

Let $X \sim \text{Pois}(\lambda)$, $Y \sim \text{Pois}(\mu)$ be independent r.v. Then

$$X + Y \sim \text{Pois}(\lambda+\mu).$$

~~Pf Remark: Let $X_n \sim \text{Binom}(n$~~

$$P\{X+Y=n\} = \sum_{k=0}^{n} P\{X+Y=n, X=k\}$$

$$= \sum_{k=0}^{n} P\{X=k, Y=n-k\} = \sum_{k=0}^{\infty} P\{X=k\} \cdot P\{Y=n-k\} \quad \text{(indep.)}$$

$$= \sum_{k=0}^{n} e^{-\lambda} \frac{\lambda^k}{k!} \cdot e^{-\mu} \frac{\mu^{n-k}}{(n-k)!}$$

$$= \frac{e^{-(\lambda+\mu)}}{n!} \sum_{k=0}^{n} \binom{n}{k} \lambda^k \mu^{n-k} \qquad \left(\binom{n}{k} = \frac{n!}{k!(n-k)!} \right)$$

$$= \frac{e^{-(\lambda+\mu)}}{n!} (\lambda+\mu)^n \quad \text{(Binomial Thm).}$$

~~Jan 26~~

2.8. Geometric Distribution.

"Time until first success"

- Consider ∞ sequence of indep. trials, each resulting in a success with prob. p.

Let X = # trial that is the first success.

$X \sim Geom(p)$

- PMF: $P\{X=k\} = \underbrace{(1-p)^{k-1}}_{\substack{\text{first } k-1 \text{ trials}\\ \text{are failures}}} \cdot \underbrace{p}_{\substack{k\text{-th is}\\ \text{a success}}}$, $k = 1, 2, 3, \dots$

$\boxed{\text{Def } X \sim Geom(p),\ p \in (0,1] \text{ is a r.v. that takes values } 1, 2, \dots, \text{ and } P\{X=k\} = (1-p)^{k-1} p,\ k = 1, 2, \dots}$

$\boxed{\text{Prop } \mathbb{E}[X] = 1/p\ ;\ Var(X) = \frac{1-p}{p^2}}$

$$\cancel{\mathbb{E}[X] = \sum_{k=1}^{\infty} k \underbrace{(1-p)}_{=q}{}^{k-1} p = p \cdot \sum_{k=1}^{\infty} k q^{k-1}}$$

$$\cancel{\sum_{k=1}^{\infty} k q^{k-1} = \sum_{k=1}^{\infty} q^{k-1} + \sum_{k=2}^{\infty} q^{k-1} + \sum_{k=3}^{\infty} q^{k-1} + \dots}$$

$$\cancel{= (1+q+q^2+\dots) \sum_{k=1}^{\infty} q^{k-1} = \left(\sum_{k=1}^{\infty} q^{k-1}\right)^2 = \frac{1}{(1-q)^2} = \frac{1}{p^2}}$$

$$\Rightarrow \mathbb{E}[X] = p \cdot \frac{1}{p^2} = \frac{1}{p}.$$

$\mathbb{E}[X] = \sum_{k=1}^{\infty} P\{X \ge k\}$ (by HW Exercise 237)

Now, $X \ge k \Leftrightarrow$ the first $k-1$ trials were failures

$\Rightarrow P\{X \ge k\} = (1-p)^{k-1}$.

$\Rightarrow \mathbb{E}[X] = \sum_{k=1}^{\infty} (1-p)^{k-1} = \frac{1}{p}$ (geometric series).

$\underline{\underline{\text{Ex}}}$ Research teams A and B are competing for the discovery of a new particle. Each day, team A may discover the particle with prob. p, and team B with prob. q (independently). What is the prob. that team A discovers the particle <u>before</u> team B?

X = day team A discovers particle, $\quad X \sim \mathrm{Geom}(p)$

Y = " ——— B ———— . $\quad Y \sim \mathrm{Geom}(q)$ $\Big\}$ indep.

$$\mathbb{P}\{X<Y\} = \sum_{k=1}^{\infty} \mathbb{P}\{X<Y,\, X=k\}$$

$$= \sum_{k=1}^{\infty} \mathbb{P}\{Y>k,\, X=k\}$$

$$= \sum_{k=1}^{\infty} \mathbb{P}\{Y>k\} \cdot \mathbb{P}\{X=k\} \quad \text{(indep.)}$$

$$= \sum_{k=1}^{\infty} \underbrace{(1-q)^{k}}_{\substack{\text{failures on} \\ \text{first } k \text{ days}}} \cdot (1-p)^{k-1} p .$$

$$= \boxed{\frac{p(1-q)}{1-(1-p)(1-q)}} \quad \left(\approx \frac{p}{p+q} \text{ if } p, q \text{ small} \right)$$

Ex (Coupon Collecting Problem I)

There are n ~~coupons~~ different types of coupons

Each time one obtains a coupon, it is equally likely to be of each type.

Compute Expectation of number of the different coupons among N collected. $\quad X$

Applications: Clinical trials: collecting info on side effects of a drug...

$Y = n - X$, # of uncollected coupons.

$E[Y] = ?$

$Y = Y_1 + Y_2 + \dots + Y_n$, where $Y_i = \begin{cases} 1, & \text{coupon of i'th type is not collected} \\ 0, & \text{---} \end{cases}$

~~$Y_i \sim$ Bernoulli (p)~~. $E[Y] = \sum_{i=1}^{N} E[Y_i]$. $E[Y_i] = 1 \cdot p + 0(1-p) = p$, where

$p = P\{\text{coupon of i'th type is not collected}\} = \left(1 - \underbrace{\tfrac{1}{n}}_{\text{collected at one trial}}\right)^N$.

$\Rightarrow E[Y] = \sum_{i=1}^{N} p = np = n\left(1-\frac{1}{n}\right)^N$.

$E(X) = \boxed{n - n\left(1-\frac{1}{n}\right)^N}$

E.g. if $N = n \to \infty \Rightarrow E[X] \approx \left(1 - \frac{1}{e}\right) n = 0.63n$.

Thus: 63% of collection after collecting n coupons.

Asympt. Analysis: $n \to \infty$, $N = tn$

$E[Y] \approx n e^{-N/n} = \boxed{n e^{-t}}$

$E[Y] < 1$ for $t \sim \log n \Rightarrow$ for $N \sim n \log n$.

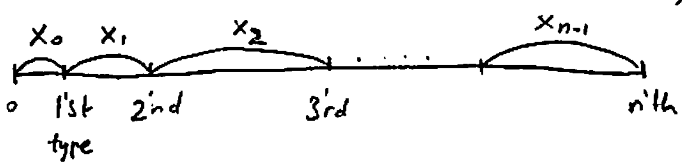
Should Expect a complete collection in time $N \sim n \log n$.

Let's verify this:

Ex (Coupon Collector's Problem II). (Ex. 2i)

What is the expected $\overbrace{\text{number of coupons}}^{X}$ one needs to collect before obtaining a complete set of all n types of coupons?

- $X = X_0 + X_1 + \dots + X_{n-1}$, where X_i = number of additional coupons (after i types have been collected) in order to obtain a new type



- $E[X] = \sum_{i=0}^{n-1} E[X_i]$.

- $X_i \sim ?$

When i types of coupons have already been collected, we are waiting for a new type. The coupons that we get now are of new type with prob.

$$p_i = \frac{n-i}{n} \quad \begin{array}{l} \leftarrow \text{\# new types} \\ \leftarrow \text{\# all types} \end{array}$$

Hence

$$X_i \sim \text{Geom}(p_i) \quad \Rightarrow \quad E[X_i] = \frac{1}{p_i}$$

- $\Rightarrow$ $E[X] = \sum_{i=0}^{n-1} \frac{1}{p_i} = \frac{n}{n} + \frac{n}{n-1} + \dots + \frac{n}{1} = n\left(1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n}\right)$.

- Asymptotic analysis (as $n \to \infty$). $1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n} \approx \ln n$ (harmonic series).

$$\Rightarrow \quad \boxed{E[X] \simeq n \ln n} \quad (\text{log-oversampling})$$

- More precisely, harmonic number $H_n = \sum_{k=1}^{n} \frac{1}{k} = \ln n + \gamma + \frac{1}{2} n^{-1} + o(n^{-2})$

$\gamma \approx 0.58$ (Euler-Mascheroni constant)

$$\Rightarrow E[X] = n \ln n + \gamma n + \frac{1}{2} + o(1)$$

Example: $n = 50$, $E[X] = 225$

- Remark. [Erdös-Renyi '61]. $P\{X < n \ln n + tn\} \to \exp(-e^{-t})$, $n \to \infty$ $(\forall t > 0)$