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): 40 students, 2 are sick on ave. $X = \# \text{sick}$.

- $X \sim \text{Poisson}(2) \implies P\{X=4\} = e^{-2} \cdot \frac{2^4}{4!} = 0.0902$ (compare to 0.0901).
- $P\{\text{no sick student}\} = P\{X=0\} = e^{-2} = 0.14$.
- $P\{\text{at least two sick}\} = 1 - P\{X=0\} - P\{X=1\} = 1 - 2e^{-2} - 2e^{-2} = 0.56$.

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

Prop (Sum of Poisson = Poisson): Let $X \sim \text{Pois}(\lambda)$, $Y \sim \text{Pois}(\mu)$ be independent r.v. Then

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

For $X \sim \text{Binomial}(n, p)$, $Y \sim \text{Binomial}(m, p)$,

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

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

(indep.)

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

$= e^{-(\lambda+\mu)} \left(\frac{n}{k!(n-k)!}\right) (\lambda+\mu)^n$ (Binomial Theorem)
2.8. Geometric Distribution

"Time until first success"

- Consider an infinite sequence of independent trials, each resulting in a success with probability $p$.
- Let $X = \# \text{ trials}$ that is the first success.
- $X \sim \text{Geom}(p)$

**PMF:** $P(X = k) = (1-p)^{k-1}p$, $k = 1, 2, 3, \ldots$

**Def:** $X \sim \text{Geom}(p)$, $p \in (0,1)$ is a r.v. that takes values $1, 2, \ldots$, and

$$P(X = k) = (1-p)^{k-1}p, \quad k = 1, 2, \ldots$$

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

$$E[X] = \sum_{k=1}^{\infty} k(1-p)^{k-1}p = \frac{1}{p^2} \sum_{k=1}^{\infty} \frac{k(1-p)^{k-1}}{p} = \frac{1}{p^2} \frac{d}{dp} \sum_{k=1}^{\infty} (1-p)^{k-1}$$

$$= \frac{1}{p^2} \frac{d}{dp} \left( \frac{1}{1-p} \right) = \frac{1}{p^2} \frac{1}{(1-p)^2} = \frac{1}{p - 1} = \frac{1}{p}$$

$$E[X] = \sum_{k=1}^{\infty} P(X \geq k) \quad \text{(by KW Exercise 2.3.7)}$$

Now, $X \geq k \iff$ the first $k-1$ trials were failures

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

$$\Rightarrow E[X] = \sum_{k=1}^{\infty} (1-p)^{k-1} = \frac{1}{p} \quad \text{(geometric series)}.$$
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 before team B?

Let $X =$ day team A discovers particle, $X \sim \text{Geom}(p)$ 
$Y =$ day team B discovers particle, $Y \sim \text{Geom}(q)$, indep.

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

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

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

$$= \sum_{k=1}^{\infty} (1-q)^{k-1} \cdot (1-p)^{p-1}$$

$$= \left[ \frac{p(1-q)}{1-(1-p)(1-q)} \right] \left( \approx \frac{p}{p+q} \text{ if } p, q \text{ small} \right)$$
Ex (Coupon Collecting Problem)

There are \( n \) different types of coupons. Each time one obtains a coupon, it is equally likely to be of any type.

Compute the expected number of the different coupons among \( N \) collected.

Applications: Clinical trials - study side effects of a drug.

- \( X \)

\[ Y = N - X, \] \# of uncollected coupons.

\[ E(Y) = ? \]

\[ Y = Y_1 + Y_2 + \ldots + Y_n, \] where \( Y_i = \begin{cases} 1, & \text{coupon of } i^{\text{th}} \text{ type is not collected} \\ 0, & \end{cases} \]

\[ E(Y) = \sum_{i=1}^{n} E(Y_i) = \sum_{i=1}^{n} P \]

\[ P = P \{ \text{coupon of } i^{\text{th}} \text{ type is not collected} \} = \left( 1 - \frac{1}{n} \right)^N. \]

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

\[ E(X) = n - n(1 - \frac{1}{n})^N \]

E.g. if \( N = n \) \( \Rightarrow \) \( E(X) = (1 - \frac{1}{e})n = 0.63n \).

Thus: 63% of collection after collecting \( n \) coupons.

Asymptotic Analysis: \( n \to \infty, N = tn \)

\[ E(Y) \approx n e^{-N/n} = n e^{-t}. \]

\[ E(Y) < 1 \text{ for } t > \log(n) \text{ for } N = n \log(n). \]

Should expect a complete collection in time \( N = n \log(n) \).

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

What is the expected number of coupons one needs to collect before obtaining a complete set of all \( n \) types of coupons?

\[ X = X_0 + X_1 + \ldots + X_{n-1}, \]
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 a new type with prob.

\[ p_i = \frac{n-i}{n} \quad \text{## new types} \]
\[ \Rightarrow X_i \sim \text{Geom } (p_i) \quad \Rightarrow E[X_i] = \frac{1}{p_i}. \]

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

Asymptotic analysis:

\[ 1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n} \approx \ln n \quad \text{(harmonic series)} \]

\[ E[X] \approx n \ln n \quad \text{(log oversampling)} \]

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

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

Example: \( n = 50, \ E[X] = 22.5 \)

Remark. [Erdős– Rényi '61]. \( P\{X < n \ln n + \ln \} \rightarrow \exp(-e^t), \ n \rightarrow \infty \ (t > 0) \).