Lectures 09 ~ 11

2.5 Geometric distribution

Ex) Rolling a die: $Y \triangleq$ # of rolls to have the first six

$$P_Y(1) = \frac{1}{6}$$

$$P_Y(2) = \frac{5}{6} \times \frac{1}{6}$$

The first outcome is not six

the second outcome is six

$$P_Y(k) = \left(\frac{5}{6}\right)^{k-1} \frac{1}{6} \quad \text{for } k \geq 1$$

If we repeat the Bernoulli experiment, how many trials do we need to get the first one?

Def: A r.v. Y is a geometric random variable with parameter p if

$$P_Y(k) = \begin{cases} (1-p)^{k-1} p & \text{for } k \geq 1 \\ 0 & \text{otherwise.} \end{cases}$$

* Memoryless property

$$P(Y > k) = P\{\text{no one occurs in the first } k \text{ trials}\} = (1-p)^k$$

$$P(Y > k+d \mid Y > d) = \frac{P\{L > k+d, Y > d\}}{P\{Y > d\}} = \frac{P\{L > k+d\}}{P\{Y > d\}} = \frac{(1-p)^{k+d}}{(1-p)^d} = (1-k)^k = P(Y > k)$$

# of trials to get the first one is the same, no matter how many trials you have conducted.

Mean and variance of a geometric R.V.

$$E[Y] = \sum_{k=1}^{\infty} k \cdot P_Y(k) = \sum_{k=1}^{\infty} k \cdot p(1-p)^{k-1} = p \sum_{k=1}^{\infty} k(1-p)^{k-1} = p \cdot \frac{1}{p^2} = \frac{1}{p}$$

Note that $\displaystyle\sum_{k=1}^{\infty} x^k = \frac{1}{1-x}$ for $|x|<1$.

Differentiate $\Rightarrow \displaystyle\sum_{k=1}^{\infty} \frac{d}{dx} x^k = \frac{d}{dx}\left(\frac{1}{1-x}\right)$

$\Rightarrow \displaystyle\sum_{k=1}^{\infty} k x^{k-1} = \frac{1}{(1-x)^2}$ take $x = 1-p$

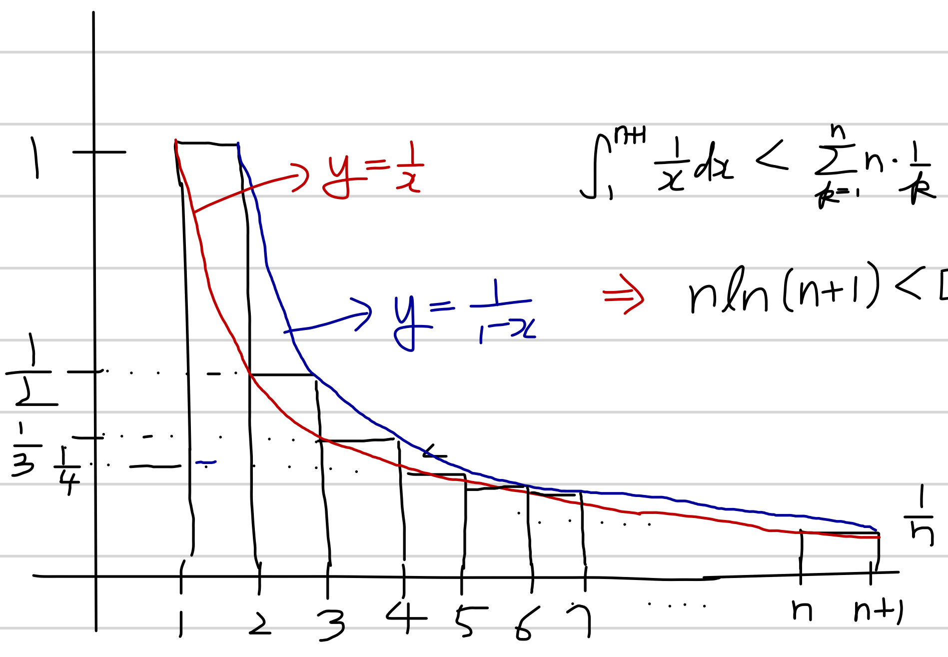
Ex) Coupon Collection Problem

- A box contains six coupons numbered $1, 2, \cdots, 6$.
- You pick one coupon at a time and return the coupon to the box.
- Y = # of picks until you collect all six different coupons.

$$E[Y] = 1 + \frac{6}{5} + \frac{6}{4} + \frac{6}{3} + \frac{6}{2} + \frac{6}{1} = \sum_{k=1}^{6} 6 \cdot \frac{1}{k}$$

- If there are n coupons,

$$E[Y] = \sum_{k=1}^{n} n \cdot \frac{1}{k} = \text{Area of the following rectangles}$$



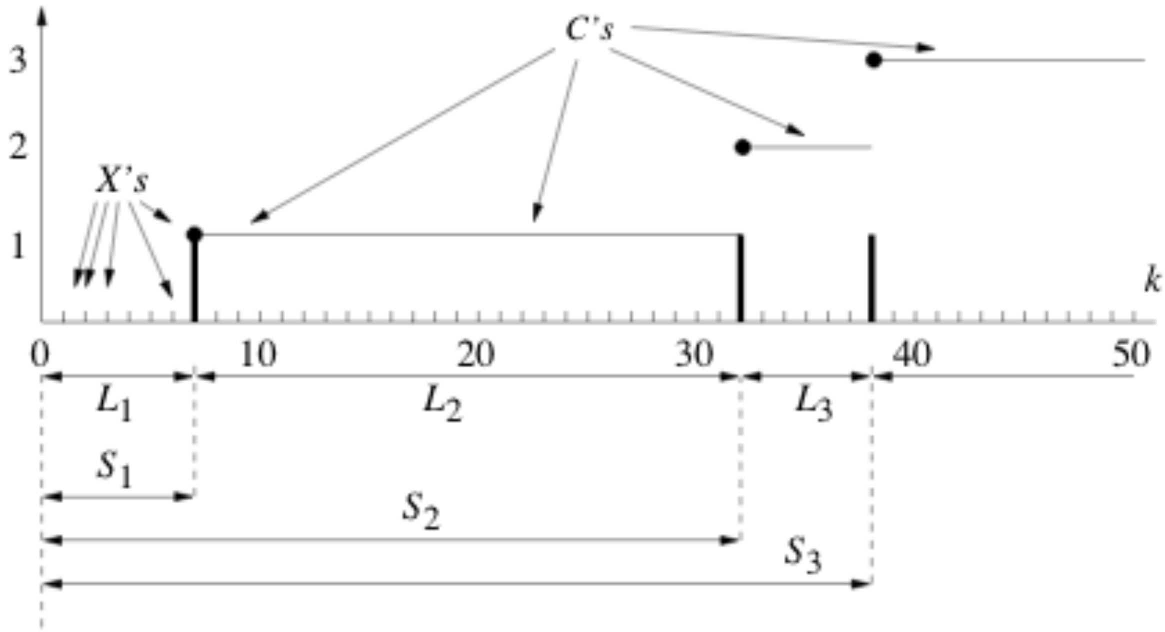
Area of the first block!

$$\int_1^{n+1} \frac{1}{x}dx < \sum_{k=1}^{n} n \cdot \frac{1}{k} < 1 + \int_2^{n+1} \frac{1}{x-1} dx$$

$$\Rightarrow n\ln(n+1) < E[Y] < n(1+\ln(n))$$

2.6 Bernoulli process and the negative binomial distribution

Def: A Bernoulli process is an infinite sequence of independent random variables.



$L_1, L_2, L_3, \cdots$: Geometric random variables.

C_n : # of ones by the n-th trial.

$\Rightarrow$ Binomial R.V. with (n, p).

S_r : Negative binomial R.V.

Q: In the Bernoulli process, how many trials are needed to get r ones?

Def: A random variable X is a negative binomial R.V. with parameters r and p if

$$p_X(k) = \begin{cases} \binom{k-1}{r-1} p^r (1-p)^{k-r} & \text{for } k \geq r. \\ 0 & \text{otherwise} \end{cases}$$

In the Bernoulli process with parameter p, the number of trials until we get r ones is a negative binomial (r, p).

P(the r-th one comes out at the k-th trial)

$= P$(one comes out at the k-th trial) independent

$\times P$($r-1$ ones come out at the first $(k-1)$ trials)

$$= p \times \binom{k-1}{r-1} p^{r-1} (1-p)^{k-r} = \binom{k-1}{r-1} p^{r} (1-p)^{k-r}$$

$$E[X] = \sum_{k=r}^{\infty} k \cdot \binom{k-1}{r-1} p^{r} (1-p)^{k-r}$$

$$= \sum_{k=r}^{\infty} \frac{k \cdot (k-1) \cdot (k-2) \cdots (k-r+1)}{(r-1)(r-2) \cdots 2 \cdot 1} p^{r} (1-p)^{k-r}$$

$$= r \sum_{k=r}^{\infty} \binom{k}{r} p^{r} (1-p)^{k-r} = r p^{r} \sum_{k=r}^{\infty} \binom{k}{r} (1-p)^{k-r}$$

Maclaurin series

$$= r p^{r} \sum_{k'=0}^{\infty} \binom{k'-r}{r} (1-p)^{k'}$$

$$= r p^{r} \cdot (1-(1-p))^{-(r+1)} = \frac{r}{p}$$

EXERCISE: $\mathrm{Var}(X) = \dfrac{r(1-p)}{p^{2}}$

2.7 Poisson R.V.

$$\text{Binomial}(n,p) \xrightarrow[\substack{n\to\infty \\ p\to 0 \\ \text{s.t. } np=\lambda \text{ (constant)}}]{} \text{Poisson}(\lambda)$$

Def: A random variable X is a Poisson r.v. with parameter λ if

$$P_X(k) = \begin{cases} \dfrac{e^{-\lambda}\lambda^k}{k!} & \text{for } k \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Poisson r.v. X is a nice approximation of a binomial r.v. with (n,p) when n is large and $\lambda = np$.

Ex) Spaceshuttle consists of 200,000 components. Each component is defective with probability 0.00001, independently of others. If there are five or more defective components, the spaceshuttle will be in trouble after the launch.

P(Trouble)?

$X \triangleq$ # of defective components
$\sim \text{Binomial}(200000, 0.00001)$

$$P(\text{Trouble}) = P(X \geq 5) = 1 - P(X<5)$$
$$= 1 - \sum_{k=0}^{4} \binom{200{,}000}{k} 0.00001^k \cdot 0.99999^{200000-k}$$

$$\approx 1 - P(Y<5) \text{ where } Y \sim \text{Poisson}(\overbrace{20000 \cdot 0.00001}^{\Rightarrow \lambda=2})$$
$$= 1 - \sum_{k=0}^{4} \frac{e^{-2}2^k}{k!} = 1 - e^{-2}\left(1 + \frac{2}{1} + \frac{4}{2} + \frac{8}{6} + \frac{16}{24}\right)$$

For $X \sim \text{Binomial}(n, p)$ and $Y \sim \text{Poisson}(\lambda)$ where $(\lambda = np)$

Why $p_X(k) \to p_Y(k)$ as $n \to \infty$?

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k} = \frac{n(n-1)\cdots(n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

$$= \frac{\lambda^k}{k!} \underbrace{\left(\frac{n(n-1)\cdots(n-k+1)}{n^k}\right)}_{\to 1 \text{ as } n \to \infty} \underbrace{\left(1 - \frac{\lambda}{n}\right)^n}_{\to e^{-\lambda} \text{ as } n \to \infty} \underbrace{\left(1 - \frac{\lambda}{n}\right)^{-k}}_{\to 1 \text{ as } n \to \infty}$$

$$= \frac{\lambda^k e^{-\lambda}}{k!} = p_Y(k)$$

* Exercise: Show $\sum_{k=0}^{\infty} p_Y(k) = 1$

Mean and Variance

$$E[Y] = \sum_{k=0}^{\infty} k \frac{\lambda^k e^{-\lambda}}{k!} = \sum_{k=1}^{\infty} \frac{\lambda^k e^{-\lambda}}{(k-1)!} = \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1} e^{-\lambda}}{(k-1)!} \xrightarrow{k' = k-1} = \lambda \sum_{k'=0}^{\infty} \frac{\lambda^{k'} e^{-\lambda}}{k'!}$$

$$= \underline{\underline{\lambda}}$$

Exercise: $\text{Var}(Y) = \lambda$.

2.8. Maximum likelihood parameter estimation

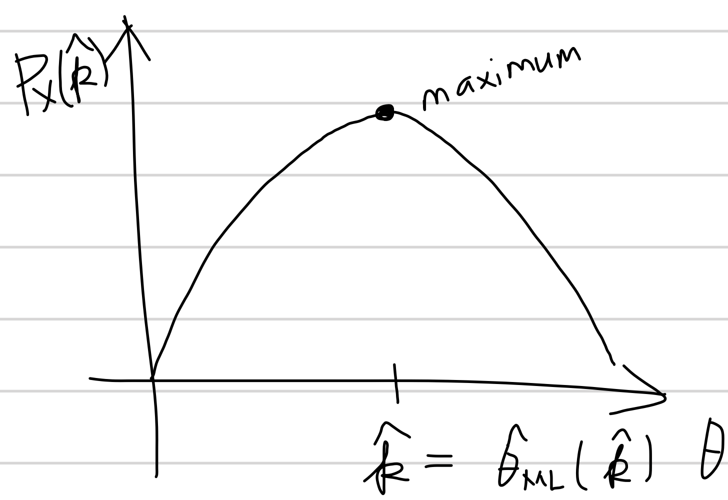
For a given sample of a R.V. X $\Longrightarrow$ how to estimate a parameter of X?

Ex) If a sample of a Poisson R.V. X with unknown parameter θ is $\hat{k}$, what is the maximum likelihood estimate $\hat{\theta}_{ML}(\hat{k})$?

$\hat{\theta}_{ML}(\hat{k})$ = the θ that maximizes $P_X(\hat{k})$.

$$P_X(\hat{k}) = \frac{\theta^{\hat{k}} e^{-\theta}}{\hat{k}!} \Longrightarrow \frac{d}{d\theta}\left(\frac{\theta^{\hat{k}} e^{-\theta}}{\hat{k}!}\right) = \frac{1}{\hat{k}!}\left(\hat{k}\,\theta^{\hat{k}-1} e^{-\theta} - \theta^{\hat{k}} e^{-\theta}\right)$$

$$= \frac{\theta^{\hat{k}-1} e^{-\theta}}{\hat{k}!}\left(\hat{k} - \theta\right) \longrightarrow \begin{cases} \le 0 & \text{if } \theta \ge \hat{k} \\ > 0 & \text{if } \theta < \hat{k} \end{cases}$$



$\therefore$ $P_X(\hat{k})$ is maximized when θ is given by $\hat{k}$.

2.9 Markov's and Chebychev's inequalities.

Def: Markov's inequality: For a non-negative r.v. X, and $c>0$,

$$P(X \geq c) \leq \frac{E[X]}{c}.$$

Proof: $E[X] = \sum_{k} k \cdot p_X(k) = \underbrace{\sum_{k: k<c} k \cdot p_X(k)}_{\geq 0} + \sum_{k: k\geq c} \underbrace{k}_{\geq c} p_X(k)$

$$\geq 0 + \sum_{k: k\geq c} c \cdot p_X(k) = c \sum_{k: k\geq c} p_X(k) = c \cdot P(X \geq k)$$

$$\Rightarrow P(X \geq k) \leq \frac{E[X]}{c}$$

Def: Chebychev inequality: For a R.V. X with mean μ and variance σ^2

$$P(|X-\mu| \geq d) \leq \frac{\sigma^2}{d^2}$$

Let $Y=(X-\mu)^2$ and apply Y to Markov's inequality

$$\Rightarrow P(Y \geq d^2) \leq \frac{E[Y]}{d^2}$$

Since $E(Y) = \sigma^2$ and $P(Y \geq d^2) = P(|X-\mu| \geq d)$,

$$\Rightarrow P(|X-\mu| \geq d) \leq \frac{\sigma^2}{d^2}$$

Ex) Spaceshuttle with 200,000 components (Continued)

Q1) Using Markov's inequality, find the upper bound of P(TROUBLE)

Let X = # of defective components

$$P(X \geq 5) \leq \frac{E[X]}{5} = \frac{2}{5}$$

Q2) Using Chebychev's inequality, find the upper bound of P(trouble)

$$P(X \geq 5) = P(X - \underset{\text{mean}}{2} \geq 3) = P(|X-2| \geq 3) \leq \frac{Var(X)}{3^2} = \frac{np(1-p)}{9} \approx \frac{2}{9}$$

Q3) suppose the probability that a component is defective is unknown.
If you inspect the spaceshuttle before the launch, you found 8 components out of 200,000 components are defective.

What is the ML estimator of $\hat{p}$ (prob. of a defective component)

$$p_X(8) = \binom{200,000}{8} p^8 (1-p)^{200,000-8}$$

What p maximize $p_X(8)$?

$$\frac{d}{dp} p_X(8) = \binom{200,000}{8} \cdot \left[8 \cdot p^7 (1-p)^{200,000-8} - (200,000-8) p^8 (1-p)^{200000-9}\right]$$

$$= \binom{200,000}{8} p^7 (1-p)^{200,000-9} \left[8(1-p) - (200,000-8)p\right]$$

$$= \binom{200,000}{8} p^7 (1-p)^{200,000-9} (8 - 200,000p)$$