Last lecture

Distributions

- Bernoulli (Ch 2.4.3)
- Binomial (Ch 2.4.4)

Geometric distribution (Ch 2.5)

- Definition, mean and variance

Agenda

Correction on Best-of-K

Geometric distribution (Ch 2.5)

- Memoryless property

Bernoulli Process (Ch 2.6)

- Definition
- Connect to distributions

Poisson Distribution (Ch 2.7)

Maximum Likelihood Estimation (MLE)

Binomial Example – Best of K

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Team A and B play "Best of 7" games

- No tie, whoever wins 4 games out of 7 is the match winner
- E.g. $w_i = \{A, A, A, B, A\}$: the winner is A
- Let p denotes A's win rate per game
- Y denotes the number of games played, $p_Y(k) = ?$

Geometric Distributions

Property – Memoryless property

For geometric series, failing 10 times will not affect the 11-th trial

- $P\{L > k + n | L > n\} =$

- Called "memoryless property"
- What's the expected total number to get the first 1 after getting {0,0,0,0}?

Game – Push the luck (simplified Incan-Gold)

Start a game with infinite rounds and 0 points

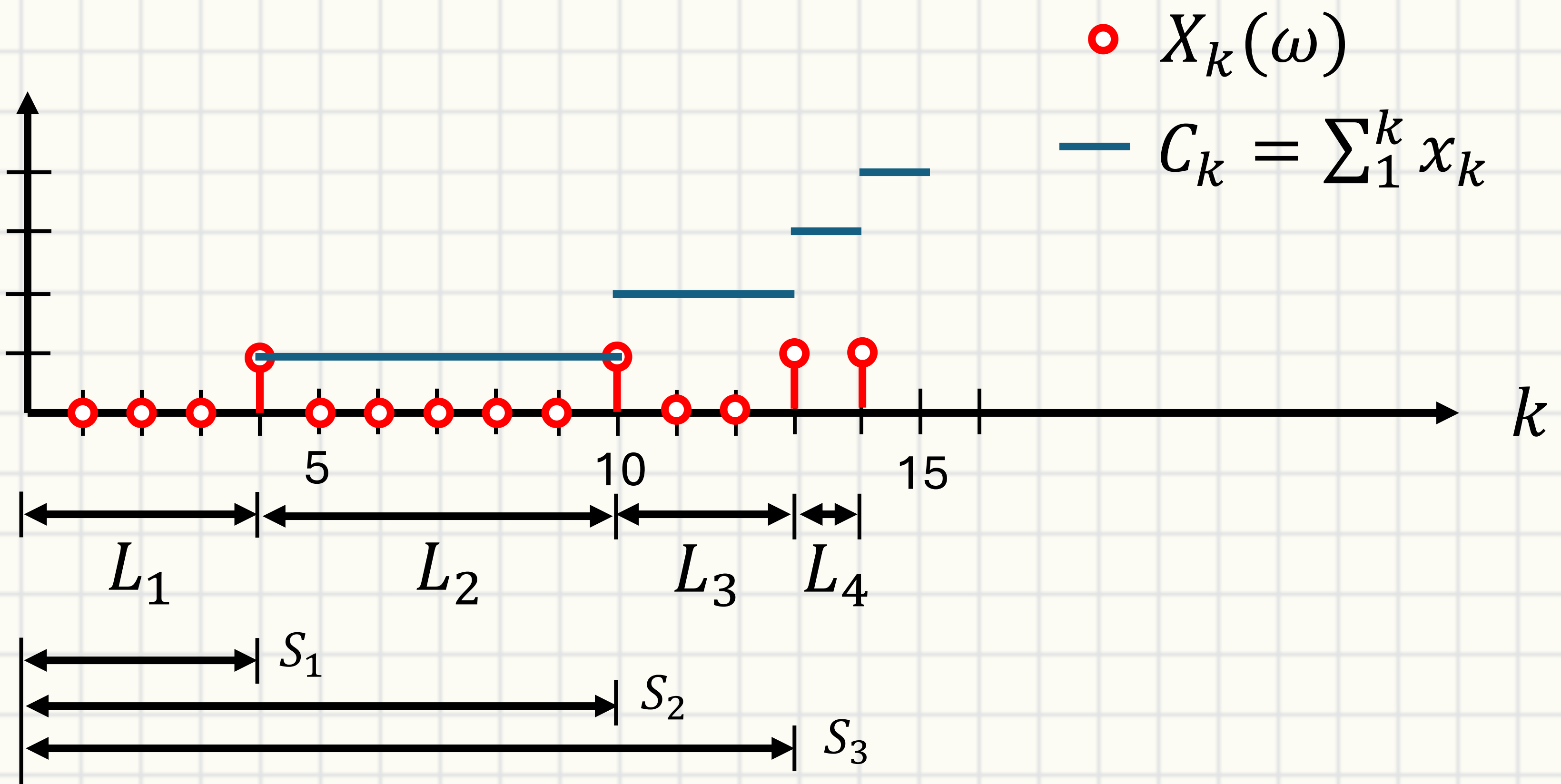
- 2 Actions per round – Go or Keep
- Go
  - $p = \frac{2}{3}$ win 1 point
  - Otherwise, lose all points
- Keep
  - Deposit the current point and end the game
- What's the best strategy?

Bernoulli Process

Bernoulli Process Definition

An infinite sequence $X_1, X_2 \ldots$ s.t. $X_k \sim Bern(p)$

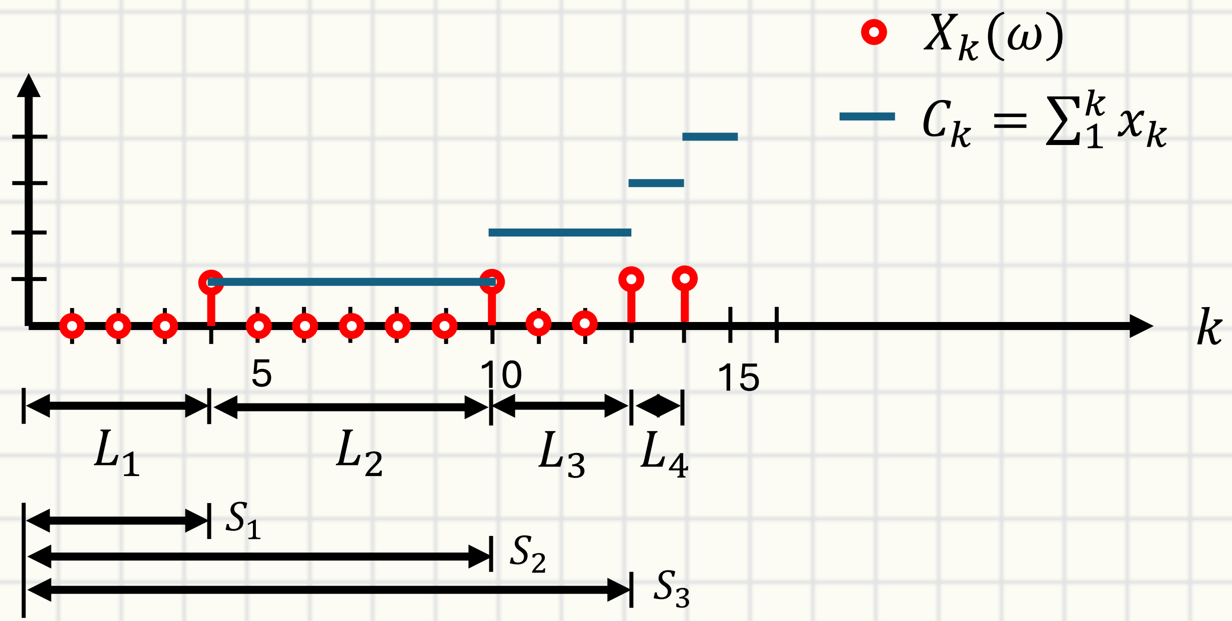
- ω is a possible outcome of the sequence
- $X_k(\omega)$ is called a “ ” of outcome ω



Bernoulli Process Definition

Observe that a Bernoulli process can be defined by

1. $X_k \sim Bern(p)$
2. $C_k \sim B(k, p)$
3. $L_k \sim L(p)$
4. $S_r = \sum_1^r L_r$: # of trials required to get r ones

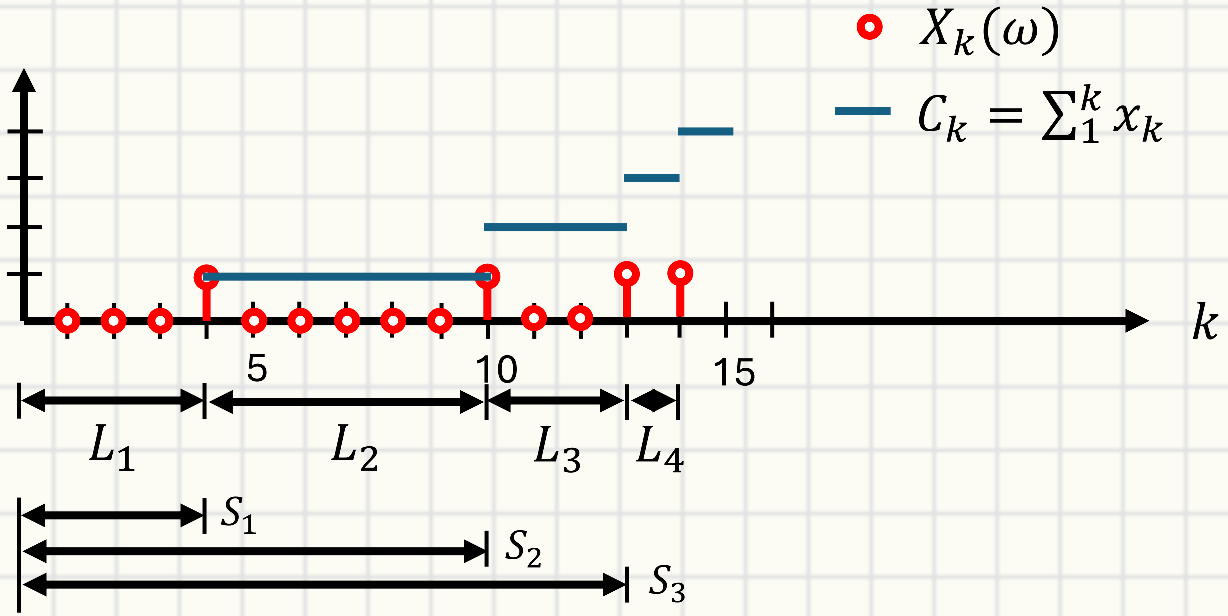


S_r - Negative Binomial Distribution

What is the pmf of S_r with parameter (r, p)?

- \# of trials required to get r ones

- $p_S(n) =$



Poisson Distribution

Poisson Distribution $\boldsymbol{Pois(\lambda)}$

A binomial distribution with large n, small p, and $\lambda = np$

- Example – Misspelled words in a document
  - Many number of words n
  - Small misspelled rate p

- When we care about the "rate" np
  - We only have the mean np
  - We know p is small

- $p_X(k) =$

Poisson Distribution $\boldsymbol{Pois(\lambda)}$

- Why $p_X(k) = \frac{e^{-\lambda}\lambda^k}{k!}$?
  - $p_X(k) \propto \frac{\lambda^k}{k!}$
  - $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$
  - $\mu_x =$
  - $\sigma_x^2 =$

Poisson Distribution Example

Consider a wireless link features bit error rate $1e^{-4}$

- $X \triangleq$ # of error bits in 1 Byte, $P_X(0) =?$

- $Y \triangleq$ # of error bits in 10MB , $P_Y(10) =?$

Poisson Distribution Example

Consider a wireless link features bit error rate $1e^{-4}$

- Each packet (N bits) is governed by an error correction algorithm that can correct up to 2 bits
- If we want packet error rate lower than 0.01, what's the maximum packet size N?

- $P_X(0) + P_X(1) + P_X(2) \geq 0.99$
- $e^{-\lambda}\left(1 + \lambda + \frac{\lambda^2}{2}\right) \geq 0.99$
- $\lambda = Np \leq 0.436$
- $N \leq 4360$

Maximum Likelihood Estimation (MLE)

Not in the exam

Definition

How a distribution is of parameter θ given the observation k.

- $\operatorname{argmax}_{\theta} p_{\theta}(k)$
- If I get $\{H, H, H, T, H\}$ out of unfair coin toss, what's p

Likelihood $p_{\theta}(k)$

- $P(k|\theta)$ for different θ
- How likely there will be $1R2B$ if I draw $\{R, R, B\}$

MLE

- Find θ that "Maximize" the likelihood given k

Example – Special Lottery

In the first draw, the customer has a probability of θ to win (W) and $(1 - \theta)$ to lose (L).
For each L ticket drawn in a sequence, the winning rate is doubled.
E.g. If Alice draws $\{L, L\}$, she has the probability 4θ to draw a W ticket.

Estimate θ if Alice draw $\{L, L\ W\}$

Markov and Chebychev Inequalities

Markov Inequality

What if we only know $E[Y]$ or $Var(Y)$?

- Can we know more?

Markov inequality – If Y is a non-negative RV, for $c > 0$

- $P\{Y \geq c\} \leq \frac{E[Y]}{c}$

$$
\begin{aligned}
E[Y] &= \sum_i u_i p_Y(u_i) \\
&= \sum_{u_i < c} u_i p_Y(u_i) + \sum_{u_i \geq c} u_i p_Y(u_i) \\
&\geq \sum_{u_i < c} 0 \times p_Y(u_i) + \sum_{u_i \geq c} c p_Y(u_i) \\
&= c \sum_{u_i \geq c} p_Y(u_i) = cP(Y \geq c)
\end{aligned}
$$

Equality holds iif $p_Y(0) + p_Y(c) = 1$

Example

Through 200 balls into 100 bins randomly. At most how many bins can contain $c \geq 5$ balls?

- Intuitive solution

- Markov inequality
  - $E[Y] =$
  - $P\{Y \geq 5\} \leq$

Chebychev Inequality

Give information regarding $Var(X)$

If X is a RV, for $\mathrm{d} > 0$

- $P\{|X - \mu_X| \geq d\} \leq \frac{\sigma_X^2}{d^2}$
- $P\{|X - \mu_X| \geq a\sigma_X\} \leq \frac{1}{a^2}$
- Proof - Extension of Markov inequality

Confidence Interval

How close is our estimate $\hat{p}$ to the real parameter p

- Do a poll of 200 people - X denotes # of people agree
- $X \sim Bi(n = 200, p)$
- $\hat{p} =$
- $P\{|X - np| \geq a\sigma\} \leq \frac{1}{a^2}$
- $P\left\{\left|\frac{X}{n} - p\right| \leq \frac{a\sigma}{n}\right\} \geq \frac{1}{a^2}$
- $(\hat{p} - a\sqrt{\frac{p(1-p)}{n}}, \hat{p} + a\sqrt{\frac{p(1-p)}{n}})$ is called **Confidence interval**