Example

- $f_X(u) = \begin{cases} \frac{3}{4}(1-u^2) & if \quad -1 \leq u \leq 1 \\ 0 & else \end{cases}$
- Find $F_X(u) = \int_{-1}^{u} \frac{3}{4}(1-u^2)\,du$

$$= \left[\frac{3}{4}u - \frac{u^3}{4}\right]_{-1}^{u} = \left(\frac{3}{4}u - \frac{u^3}{4}\right) - \left[\frac{-3}{4} + \frac{1}{4}\right]$$

$$= \frac{3}{4}u - \frac{u^3}{4} + \frac{1}{2}.$$

$$F_x(u) \begin{cases} 0 & u \leq -1 \\ \frac{3}{4}u - \frac{u^3}{4} + \frac{1}{2} & -1 \leq u \leq 1 \\ 1 & u \geq 1 \end{cases}$$

Caveats

- Problems in the slide will be independent from midterm problems
  - $P(p_1 | p_1' \text{ in slide}) = P(p_1)$

- Review problems are mostly from homework previous years
  - Midterm will be slightly easier than homework in general

- All numbers will be replaced by symbols in the slide
  - In midterm, you may need to compute

- We will cover top-K options from the Slido survey
  - Survey does not cover all topics
  - You still need to review all topics by yourself

Agenda & Survey result

- Bernoulli Process
  - Bernoulli/ Binomial/ Poisson/ Geometry/ Neg. Bi. Review

- Counting

- Markov Inequality

- Binary Hypothesis Testing (BHT)

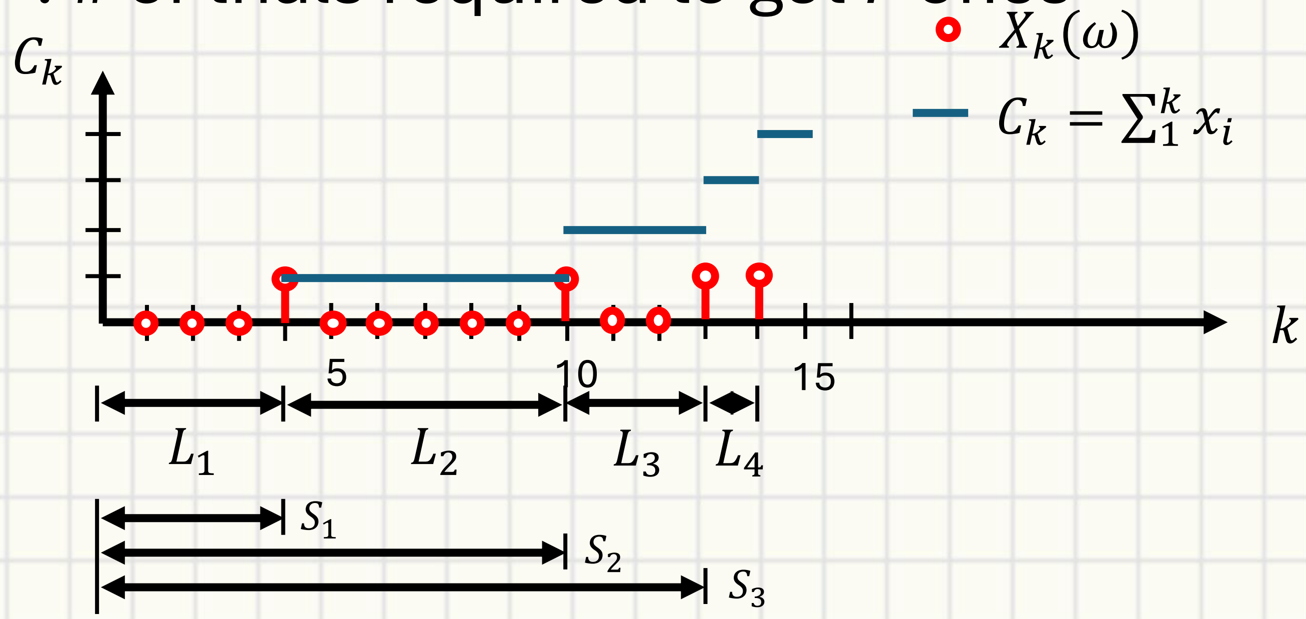
Bernoulli Process & Distributions

Bernoulli Process

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

- Toss unfair coin many times

1. $X_k \sim Bern(p)$
2. $C_k \sim Bi(\overset{n=k.}{k}, p) = \sum_k X_i$
3. $L_k \sim Geo(p)$
4. $S_r \sim NB(r, p) = \sum_1^r L_r$: # of trials required to get r ones



Properties

	$Bern(p)$	$Bin(n,p)$	$Poi(\lambda)$	$Geo(p)$	$NB(r,p)$
Relation	X_i	$\sum_n X_i$	$\sum_n X_i, n \to \infty$	Y_i	$\sum_n Y_i$
Mean	p	np	$\lambda = np$.	$\frac{1}{p}$.	$\frac{r}{p}$
Variance	$p(1-p)$	$np(1-p)$	λ.	$\frac{1-p}{p^2}$.	$\frac{r(1-p)}{p^2}$
pmf	$p_X(1) = p$ $p_X(0) = 1-p$	$\binom{n}{k} p^k (1-p)^{n-k}$	$\frac{e^{-\lambda} \lambda^k}{k!}$	$(1-p)^{k-1} p$	$\binom{k-1}{r-1} p^r (1-p)^{k-r}$
Example	Toss a coin	Toss n times	Large n small p	Until first H	Until r^{th} H
Special		$(p+q)^n$		Memoryless	

Poisson

	$Bern(p)$	$Bin(n,p)$	$Poi(\lambda)$	$Geo(p)$	$NB(r,p)$
Mean	p	np	λ	$1/p$	r/p
Variance	$p(1-p)$	$np(1-p)$	λ	$(1-p)/p^2$	$r(1-p)/p^2$
pmf	-	$\binom{n}{k} p^k (1-p)^{n-k}$	$e^{-\lambda}\lambda^k/k!$	$(1-p)^{k-1}p$	$\binom{k-1}{r-1}(1-p)^{k-r}p^r$

Alice misspelled the word with $p \ll 1$ probability. She wrote an essay of n words.

- Let E denotes the number of misspelled words
- $P\{E \leq 2\}$? $\sum_{k=0}^{2} P_E(k) = \sum_{k=0}^{2} \frac{e^{-\lambda}\lambda^k}{k!} = \sum_{k=0}^{2} \frac{e^{-(np)}(np)^k}{k!}$
- $P\{E > 3 | E \geq 1\}$?
- For $E = e$ misspelled words, Alice will be fined $2e^2$, expected fine = ?

$$1 - P\{\underbrace{E \in \{1,2,3\}}_{A} \,|\, \underbrace{E \geq 1}_{B}\} = 1 - \frac{\overbrace{\sum_{k=1}^{3} P_E(k)}^{P(A \cap B)}}{\underbrace{1 - P_E(0)}_{P(B)}}$$

$$E[2E^2] = 2\,E[E^2] = 2(\lambda + \underline{\lambda^2})$$

$$E[E^2 - \underbrace{\mu_E^2}_{\lambda^2}] = \sigma_E^2 = \lambda$$

$$E[E^2] = \lambda + \lambda^2$$

$$P(A|B) = \frac{P(AB)}{P(B)}$$

Binomial

	$Bern(p)$	$Bin(n,p)$	$Poi(\lambda)$	$Geo(p)$	$NB(r,p)$
Mean	p	np	λ	$1/p$	r/p
Variance	$p(1-p)$	$np(1-p)$	λ	$(1-p)/p^2$	$r(1-p)/p^2$
pmf	-	$\binom{n}{k} p^k (1-p)^{n-k}$	$e^{-\lambda}\lambda^k/k!$	$(1-p)^{k-1}p$	$\binom{k-1}{r-1}(1-p)^{k-r}p^r$

An airplane has s seats but sell $t > s$ tickets. Each customer has probability p to show.

- P{Everyone has seats}
- P{Everyone has seats| 2 already changed the ticket}?

$$1 - \text{fail case} = 1 - \sum_{k=s+1}^{t} P_X(k)$$

↳ $X \sim Bi(t,p)$

$$1 - \sum_{k=s+1}^{t-2} P_X(k)$$

↳ $X \sim Bi(t-2,p)$

Geometry

	$Bern(p)$	$Bin(n,p)$	$Poi(\lambda)$	$Geo(p)$	$NB(r,p)$
Mean	p	np	λ	$1/p$	r/p
Variance	$p(1-p)$	$np(1-p)$	λ	$(1-p)/p^2$	$r(1-p)/p^2$
pmf	-	$\binom{n}{k} p^k (1-p)^{n-k}$	$e^{-\lambda}\lambda^k/k!$	$(1-p)^{k-1}p$	$\binom{k-1}{r-1}(1-p)^{k-r}p^r$

Play a Roulette wheel

- $\Omega = \{00, 0, 1, \ldots .36\}$ $\|\Omega\| = 38$
- Always bet "small" $S = \{1,2, \ldots 15\}$ $p = \frac{15}{38}$
- P(Lose first 3 bets)
- P(First win on 5^{th} bet)
- P(First win on 6^{th} bet | Lose on first bet)

$(1-p)^3$,

$(1-p)^4 p$,

starting from 2nd round

$P_2(5) = (1-p)^4 p$

Neg. Bin.

	$Bern(p)$	$Bin(n,p)$	$Poi(\lambda)$	$Geo(p)$	$NB(r,p)$
Mean	p	np	λ	$1/p$	r/p
Variance	$p(1-p)$	$np(1-p)$	λ	$(1-p)/p^2$	$r(1-p)/p^2$
pmf	-	$\binom{n}{k} p^k (1-p)^{n-k}$	$e^{-\lambda}\lambda^k/k!$	$(1-p)^{k-1}p$	$\binom{k-1}{r-1}(1-p)^{k-r}p^r$

$X \sim NB(r,p)$ 5 10 14

- P(r success occurs before m loss)?

$$P\{X < r+m.\} = \sum_{k=r}^{r+m-1} P_X(k)$$

Counting

Counting

If outcomes $x \in \Omega$ are equally likely

- $P(A) = \frac{||A||}{||\Omega||}$ is the ratio of event/ sample space cardinality
- Simple counting - $||\Omega|| = ||\Omega_A|| \times ||\Omega_B||$
  - Roll a die and toss a coin
  - w/ replacement
- Overcounting
  - Permutation (order matters ↙ no replacement
    - Draw 5 cards one by one
  - Combination (order doesn't matter
    - Draw 5 cards at once

Permutation

Order k out of n different things

- $\overset{k\ \text{terms}}{\underbrace{n \times (n-1) \times \cdots \times (n-k+1)}} = \frac{n!}{(n-k)!}$
- E.g. Draw 5 cards out of 52 one by one
- If there are m_i things of the same type i chosen
  - Divided by $m_i!$
  - Order $ILLINI$

$$\frac{6!}{3!\,2!}$$

Combination

Choose n out of k different things

- $\binom{n}{k} = \frac{n!}{k!(n-k)!}$
- Default thinking in counting problems unless mentioned ordered, one-by-one,etc

- Mindset
    - Enumerate some combinations
        - Are they counted?
        - Are they overcounted?

Lottery Example

Lottery pick 3 different numbers from 1 to 10. The host draws 3 winning numbers.

- $P\{\underline{\text{Exactly 2 numbers match}}\}$ A

$$\frac{\|A\|}{\|\Omega\|} = \frac{\binom{\boxed{3}}{2} \times 7}{\binom{10}{3}}$$

winning

select

8 9 10

$\underline{8 \quad 9}$ (1-7)

Markov/ Chebyshev Inequality

Inequalities

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

- Valid only for none negative outcomes
- Large outcomes are unlikely

Chebyshev: $P\{|X - \mu_X| \geq a\sigma_X\} \leq \frac{1}{a^2}$

- Should not deviates from mean too much

Confidence Interval: $P\left\{p \in \left(\hat{p} - \frac{a}{2\sqrt{n}}, \hat{p} + \frac{a}{2\sqrt{n}}\right)\right\} \geq 1 - \frac{1}{a^2}$

- How should we decide n if we want the estimate to be accurate.

Inequality Example

$L \sim Geo(p)$ and $X \sim Poi(\lambda)$ are independent , $Y = X + L$

- Bound $P\{Y \geq 2p\}$ using Markov

$\frac{2}{p}$

$C = \frac{2}{p}$

$\mu_Y = \lambda + \frac{1}{p}$

$\sigma_Y = \lambda + \frac{1-p}{p^2}$

- Bound $P\{Y < \frac{p+\lambda}{2}\}$ using Chebyshev

$\frac{1}{p}$

$$P\{Y \geq \frac{2}{p}\} \leq \frac{\lambda + \frac{1}{p}}{\frac{2}{p}} = \frac{\lambda p + 1}{2}$$

$$P\left\{\left|Y - \left(\lambda + \frac{1}{p}\right)\right| \geq \frac{\lambda + \frac{1}{p}}{2}\right\} \rightarrow \underset{\sim}{a\sigma}$$

$$a = \frac{\lambda + \frac{1}{p}}{2\left(\lambda + \frac{1-p}{p^2}\right)}$$

Binary Hypothesis Testing

Two types of table

$\pi_1 = 0.2$ $\pi_0 = 0.8$

$P(X\|H)$	$X=0$	$X=1$	$X=2$	$X=3$
H_1	0	0.1	0.3	0.6
H_0	0.4	0.3	0.2	0.1

$\times \pi_1$

$\times \pi_0$

$P(H,X)$	$X=0$	$X=1$	$X=2$	$X=3$
H_1	0	0.02	0.06	
H_0	0.32			

Likelihood Table $P(X|H)$

- Good for ML rule
- Row sum to 1
- Column Ratio = LRT $= \Lambda$
- None chosen row sum gives p_{miss} and $p_{false\ alarm}$

Joint Prob. Table $P(X,H)$

- Good for MAP rule
- Table sum to 1
- None chosen table sum gives p_{error}

Example

$P(X|H_1) \sim \underline{Poi(\lambda_1)}$ and $\underline{P(X|H_0) \sim Poi(\lambda_0)}$, assume $\frac{\pi_0}{\pi_1} = \pi$

- Describe ML and MAP rule
- Find corresponding p_e

ML $>?$ 1

$$\Lambda(k) = \frac{e^{-\lambda_1}\lambda_1^{k}}{e^{-\lambda_2}\lambda_2^{k}}$$

MAP $>?$ π

$k > k^*_{ML}$ claim H_1

$$P_e = \sum_{k=0}^{k^*} \frac{e^{-\lambda_1}\lambda_1^{k}}{k!} + \sum_{k=k^*+1}^{\infty} \frac{e^{-\lambda_0}\lambda_0^{k}}{k!}$$

Real-Time QA