Last lecture

Geometric distribution (Ch 2.5)

- Memoryless property

Bernoulli Process (Ch 2.6)

- Definition
- Connect to distributions

Poisson Distribution (Ch 2.7)

Agenda

Maximum Likelihood Estimation (MLE) (Ch 2.8)

- Not focus

Markov and Chebychev inequalities (Ch 2.9)

- Markov inequality
- Chebychev inequality
- Confidence interval

Binary Hypothesis Testing (Ch 2.11)

- Definition
- Likelihood table
- Maximum likelihood decision rule

Maximum Likelihood Estimation (MLE)

Not in the exam

Definition

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

- $\text{argmax}_\theta \, p_\theta(k)$ ↗ w
- If I get $\{H, H, H, T, H\}$ out of unfair coin toss, what's $p^* = \frac{4}{5}$

↳ $p^4(1-p) = p(w)$

Likelihood $p_\theta(k)$

$\text{argmax}_p \, p^4(1-p)$

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

MLE

- Find θ that "Maximize" the (log)-likelihood given k
- $\frac{\partial P_\theta(k)}{\partial \theta} = 0$ $\quad \frac{\partial \log P}{\partial \theta} = 0$

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.

$k=$

$4 \cdot \frac{3+\sqrt{3}}{6} > 1$

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

$P_\theta(k) = (1-\theta)(1-2\theta)4\theta = 4(2\theta^3 - 3\theta^2 + \theta)$

$\frac{\partial P}{\partial \theta} = 0 \Rightarrow 6\theta^2 - 6\theta + 1 = 0 \qquad \theta = \frac{3-\sqrt{3}}{6}$

#1 Midterm Reminder

- March 2 (Mon.) 7-8:30 PM @ 1002/1013 ECEB
  - Conflict - March 3 8-9:30AM @ 3081 ECEB
  - Please register the conflict exam by next Friday
    - https://forms.illinois.edu/sec/665800842
- All topics ~~in~~ Ch 2.
  until
- 1 Letter size HAND-WRITTEN notes
- No calculator
- Exam will be scanned and graded on Gradescope

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 $c > 0$

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

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

bucket u by $>$ or $<$ c

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 $\frac{200}{5} = 40$
- Markov inequality
  - $E[Y] = \frac{200}{100} = 2$
  - $P\{Y \geq 5\} \leq \frac{E[Y]}{c} = \frac{2}{5}$

c

$$\#\text{bins} = 100 \times P\{Y \geq 5\} = 100 \times \frac{2}{5}$$

$$= 40$$

Chebyshev Inequality

Markov

$P\{Y \geq c\} \leq \frac{\mu_Y}{c}$

$Y = (X - \mu_X)^2$

$P\{(X-\mu_X)^2 \geq d^2\} \leq \frac{Var(X)}{d^2}$

Give information regarding $Var(X)$

If X is a RV, for d > 0

- $P\{|X - \mu_X| \geq d\} \leq \frac{\sigma_X^2}{d^2}$

$d = a\sigma_x$

- $P\{|X - \mu_X| \geq a\sigma_X\} \leq \frac{1}{a^2}$
- 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
- $X \sim Bi(n = 200, p)$
- $\hat{p} = \frac{X}{n} \neq p$

Cheb.

- $P\{|X - np| \geq a\sigma\} \leq \frac{1}{a^2}$

complement

$P\{|X - np| < a\sigma\} \geq 1 - \frac{1}{a^2}$

- $P\left\{\left|\frac{X}{n} - p\right| \leq \frac{a\sigma}{n}\right\} \geq 1 - \frac{1}{a^2}$

$p \in \hat{p} \pm \frac{a\sigma}{n}$

- $\left(\hat{p} - a\sqrt{\frac{p(1-p)}{n}}, \hat{p} + a\sqrt{\frac{p(1-p)}{n}}\right)$ is called **Confidence interval**

$\sigma_X^2 = np(1-p)$

Closer look – confidence interval

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

$\to 1 - \frac{1}{25}$

- Before starting the poll, if we take $a = 5$,
  - we have $1 - \frac{1}{25} =$ 96% confidence that p will locate at this interval
  - But we don't know $p(1-p)$? Replace it with a loose bound $p(1-p) \leq 0.25$
- $\boxed{P\left\{p \in \left(\hat{p} - \frac{a}{2\sqrt{n}}, \hat{p} + \frac{a}{2\sqrt{n}}\right)\right\} \geq 1 - \frac{1}{a^2}}$

Example

We want to do an opinion poll of size n for a policy.

- X is # of positive votes
- $\hat{p} = \frac{X}{n}$ be the estimate of the support rate. error wrt $\hat{p}$
- If we want true p within 0.1 with 96% confidence
- How many participants n do we need?

1. $P\left\{p \in \left[\hat{p} - \boxed{\frac{a}{2\sqrt{n}}}, \hat{p} + \frac{a}{2\sqrt{n}}\right]\right\} \geq 1 - \frac{1}{a^2}$

0.1 96%

2. $1 - \frac{1}{a^2} = 96\%$ $a = 5.$

3. $\frac{5}{2\sqrt{n}} = 0.1$ $\sqrt{n} = 25$ $n = 625$ $\square$

Example

We want to do guess the phone busy rate p from survey size n

- X is # of busy lines
- $\hat{p} = \frac{X}{n}$ be the estimate of the support rate.
- If we want true p within 0.05 with 99% confidence
- How many phones n do we need to check?

#4151495

0.05 99%

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

$a = 10$

$\frac{10}{2\sqrt{n}} = \frac{1}{20}$ $\sqrt{n} = 100$ $n = 10000$

Hypothesis Testing

Binary Hypothesis Testing with Discreet Observations

Given two hypotheses H_1 and H_0 where $H_0 = H_1^C$

- E.g., $H_1 \triangleq$ “Patient has tumor(s)”; $H_0 \triangleq$ “Patient has no tumor”
- Decision rule
  - Decide H_1 or H_0 given X
  - E.g., $X \triangleq$”Suspect circles in ultrasound scan”
  - How can we pick the best rule?

Maximum Likelihood Table

- Table showing likelihood of two hypotheses

$P(X \mid H_1)$

	$X = 0$	$X = 1$	$X = 2$	$X \geq 3$
H_1	0	0.1	$\underline{0.3}$	$\underline{0.6}$
H_0	$\underline{0.4}$	$\underline{0.3}$	0.2	0.1

$\Sigma = 1$

- Decision rule can be shown on the table by underscore on each column

False alarm and missing

	$X=0$	$X=1$	$X=2$	$X=3$
H_1	0	0.1	0.3	0.6
H_0	0.4	0.3	0.2	0.1

- $P_{false\ alarm} = P\{\text{Claim } H_1 \mid H_0\} \overset{\downarrow}{=} 0.3 + 0.2 + 0.1 = 0.6$
- $P_{miss} = P\{\text{Claim } H_0 \mid H_1\} = 0$

	$X=0$	$X=1$	$X=2$	$X=3$
H_1	0	0.1	0.3	0.6
H_0	0.4	0.3	0.2	0.1

$P_{false\ alarm} = 0.2 + 0.1 = 0.3 \downarrow$

$P_{miss} = 0 + 0.1 = 0.1 \uparrow$

Maximum Likelihood (ML) decision rule

Pick whichever is higher per column!

	$X = 0$	$X = 1$	$X = 2$	$X = 3$
H_1	0	0.1	0.3	0.6
H_0	0.4	0.3	0.2	0.1

✓ balance P_{miss} & $P_{false\ alarm}$

What's the problem?

if $P(H_0) >> P(H_1)$ or vice versa

$\Rightarrow$ $P_{false\ alarm}$ is much more important than P_{miss}