Last lecture

Geometric distribution (Ch 2.5)

- Property – memoryless

Bernoulli Process (Ch 2.6)

- Definition
- Properties
- Negative binomial distribution

Poisson Process (Ch 2.6)

- Definition

$$\mu_{Pois} =$$

$$Var(Pois) = \sigma_{Pois}^2 =$$

Agenda

Maximum Likelihood Estimation (MLE) (Ch 2.8)

- Definition
- Motivation and Examples
- Method

Markov and Chebychev inequalities (Ch 2.9)

- Markov inequality
- Chebychev inequality
- Confidence interval

Maximum Likelihood Estimation (MLE)

Definition

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

- $\text{argmax}_\theta\, p_\theta(k)$
- 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

Motivation

In real cases, we often do not know the parameters

- Mean of binomial/ Poisson
- But we can measure
- Not limited to distributions…

Examples

- Estimate no-show rate in flight
- Estimate the mean time failure
- Estimate the win-rate of a bandit machine
- Fitting a curve (e.g. income curve)

Example – Unfair coin

We have an unfair coin of p probability getting H. If we toss $n = 1000$ times and get k heads

- Guess p?
- $p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$
- $\frac{dp_X(k)}{dp} = \left(\frac{k}{p} - \frac{n-k}{1-p}\right) p^k (1-p)^{n-k} = (k-np)p^{k-1}(1-p)$
- Max at

Example – Unknown interval

Draw a number between $[1, n]$ where n is an unknown parameters. If we observe k being drawn. Find the MLE of n

- $p_n(k) = \begin{cases} \end{cases}$

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**