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

$\mu_{Pois} =$ $\lambda = np.$

$Var(Pois) = \sigma^2_{Pois} =$ $np(\underbrace{1-p}_{\to 1.})$

$= \lambda.$

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 likely 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

$p = 80\%$. θ.

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

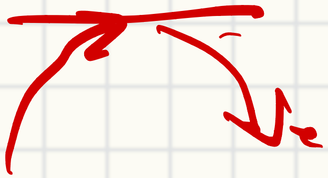
- Guess p? $\frac{k}{n}$ why?
- $p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$ $\quad X \sim Bi(n, p)$

  $P(k|p)$
- $\frac{dp_X(k)}{dp} = \left(\frac{k}{p} - \frac{n-k}{1-p}\right) p^k (1-p)^{n-k} = (k - np)p^{k-1}(1-p)$

  $= 0$ when $k = np$.

  extreme $p = \frac{k}{n}$

  $p < \frac{k}{n}$ $+$ $\quad k > \frac{k}{n}$ $-$
- Max at $\frac{k}{n}$

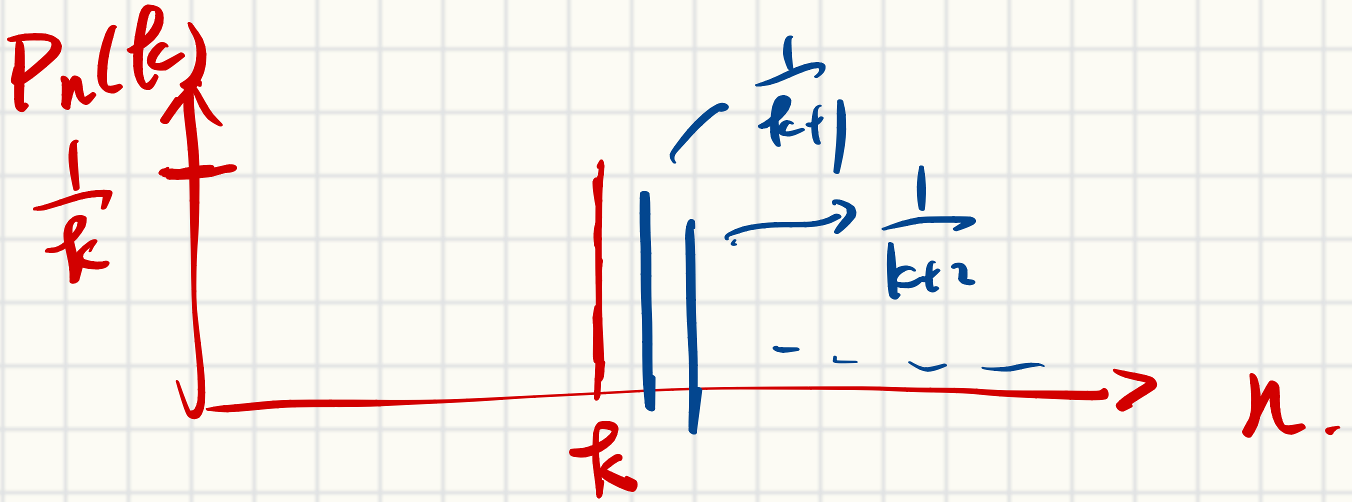


Example – Unknown interval

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

- $p_n(k) = \begin{cases} \frac{1}{n} & n \ge k \\ 0 & n < k \end{cases}$

MLE$(k) \Rightarrow n = k$.



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.

$\frac{-b \pm \sqrt{b^2-4ac}}{2a}$

$\sqrt{12} = 2\sqrt{3}$

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

$P_x = \underline{(1-\theta)(1-2\theta)} 4\theta = 8\theta^3 - 12\theta^2 + 4\theta = 4(2\theta^3 - 3\theta^2 + \theta)$

$\frac{dP_x}{d\theta} = \underset{a}{6}\theta^2 - \underset{b}{6}\theta + \underset{c}{1} = 0$

$\theta = \frac{6 \pm \sqrt{36-24}}{12} = \frac{3 \pm \sqrt{3}}{6}$

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

$\frac{60}{100}$ $\frac{40}{100}$ $P_Y(5)$

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
  - $E[Y] =$ balls per bin $\frac{200}{100} = 2$
  - $P\{Y \geq 5\} \leq \frac{E[Y]}{c} = \frac{2}{5} = 0.4$

$c=5$

$\Rightarrow \underbrace{100}_{\text{bins}} \times \underbrace{0.4}_{P_{bin}\{\geq 5\}} = 40$ bins

$P\{Y=5\}$
$= P\{Y \geq 5\} - P\{Y \geq 4\}$

Chebychev Inequality

Give information regarding $Var(X)$

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

$(X-\mu)^2$

$Var(X)$

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