Lecture 11 - 9/16

Review:

① Maximum likelihood estimator.

- There is a random variable X that its distribution belongs to a parametrized familly of distributions:

  e.g. X is a $Poi(\lambda)$ random variable

- We observe that X belongs to A after runing an experiment.

  e.g. number of customers who showed up at bank from 8am to 9 am was either 1 or 2.

  $A = \{1,2\}$ and our observation is $X \in A$ or equivalently $1 \leq X \leq 2$.

- We want to estimate the unknown parameter so that $P(X \in A)$ is maximized.

$$\text{e.g. } P(X \in A) = P(X=1) + P(X=2)$$

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

$$\frac{d}{d\lambda}P(X \in A) = e^{-\lambda} - \lambda e^{-\lambda} - e^{-\lambda}\frac{\lambda^2}{2} + \lambda e^{-\lambda}$$

$$= e^{-\lambda}\left(1 - \frac{\lambda^2}{2}\right)$$

$\frac{d}{d\lambda} P(X \in A) = 0 \Rightarrow \lambda^2 = 2 \Rightarrow \lambda = \pm\sqrt{2}$, $\sqrt{2}$ is acceptable.

notice that $P(X \in A)$ is maximized for $\lambda = \sqrt{2}$ since

$$\lambda < \sqrt{2} \Rightarrow \frac{d}{d\lambda} P(X \in A) > 0 \Rightarrow \text{increasing in } \lambda$$

$$\lambda > \sqrt{2} \Rightarrow \frac{d}{d\lambda} P(X \in A) < 0 \Rightarrow \text{decreasing in } \lambda$$

② Markov and Chebyche's inequality:

Markov: X is non-negative $c > 0$, $P(X \geq c) \leq \frac{E[X]}{c}$

In HW4. part (b): $P(X > 200) \leq \frac{np}{200}$, so if want to make sure that $P(X > 200) \leq 0.25$, we need to have

$$\frac{np}{200} \leq 0.25 \Rightarrow np \leq 50 \Rightarrow n \leq \frac{50}{0.9} = 55.55 \text{ !!!}$$

which is useless since for $n = 200$, we have $P(X > 200) = 0$!!!

Chebychev's inequality: $P(|X - E[X]| \geq a\sigma_X) \leq \frac{1}{a^2}$, $\sigma_X = \sqrt{Var(X)}$

- equivalently, $P(|X - E[X]| < a\sigma_X) \geq 1 - \frac{1}{a^2}$

  or

$$P\left(X \in \left(E[X] - a\sigma_X, E[X] + a\sigma_X\right)\right) \geq 1 - \frac{1}{a^2}$$

i.e., we are $1 - \frac{1}{a^2}$ confident that $E[X] - a\sigma_X < X < E[X] + a\sigma_X$.

③ Confidence interval.

- Suppose that each individual is in favor of policy A with probability p.
- We sample n individuals, X of them were in favor of policy A
- We do not know p.

(i) $\hat{p} = \frac{X}{n}$ is the point estimate of p

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

- $1 - \frac{1}{a^2}$ is the confidence level
- $\frac{a}{2\sqrt{n}}$ is called half-width
- $\left(\hat{p} - \frac{a}{2\sqrt{n}}, \hat{p} + \frac{a}{2\sqrt{n}}\right)$ is interval estimator of p

e.g. we want to estimate p within 0.1 with 96% confidence. How large n should be?

96% confidence $\Rightarrow 1 - \frac{1}{a^2} = \frac{96}{100}$

within 0.1 $\Rightarrow$ within 0.1 of point estimate $\hat{p}$

$\Rightarrow \frac{a}{2\sqrt{n}} = 0.1$

$\Big\}$ together gives a bound for n.

④ law of total probability and Bayes' formula:

Suppose that $E_1, E_2, \dots, E_n$ is a partition of Ω. Suppose that $E_i \in \mathcal{F}$ and $P(E_i) > 0$.

- $P(A) = \sum_{i=1}^{n} P(A \cap E_i) = \sum_{i=1}^{n} P(E_i) P(A|E_i)$

- suppose that $P(A) > 0$ and $P(B) > 0$

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

- Combining Bayes' formula and law of total probability

. Combining Bayes' formula and law of total probability

$$P(E_i|A) = \frac{P(A|E_i)P(E_i)}{P(A|E_1)P(E_1)+\cdots+P(A|E_n)P(E_n)}$$

Today:

① law of total probability for expectation

② Examples

① law of total probability for expectation.

Recall that

$$E[X] = \sum_i u_i P_X(u_i) = \sum_i u_i P(X=u_i)$$

We can define conditional mean of a random variable in the same way:

$$E[X|A] = \sum_i u_i P(X=u_i|A)$$

e.g. Let X denote a $Geo(p)$ random variable, i.e., $P_X(k) = P(X=k) = (1-p)^{k-1}p \quad k \geq 1$.

Given $X > 1$. calculate $E[X|X>1]$:

$$E[X|X>1] = \sum_{k=1}^{\infty} k\, P(X=k|X>1) = \sum_{k=2}^{\infty} k\, P(X=k|X>1)$$

$$= \sum_{k=2}^{\infty} k \cdot \frac{(1-p)^{k-1}p}{1-p} = \sum_{k=2}^{\infty} k(1-p)^{k-2}p$$

$$= \sum_{\ell=1}^{\infty} (\ell+1)(1-p)^{\ell-1}p = 1 + \frac{1}{p}$$

Similarly, we have LOTUS:

$$E[g(X)|A] = \sum_i g(u_i) P(X=u_i|A)$$

Suppose that $E_1, E_2, \ldots, E_n$ is a partition of Ω. Suppose that $E_i \in \mathcal{F}$ and $P(E_i) > 0$.

We have,

$$E[X] = \sum_i u_i P(X=u_i)$$

$$= \sum_i u_i \sum_{j=1}^{n} P(X=u_i|E_j)P(E_j)$$

$$= \sum_i u_i \sum_{j=1}^{n} P(X=u_i \mid E_j)\, P(E_j)$$

$$= \sum_{j=1}^{n} P(E_j) \sum_i u_i P(X=u_i \mid E_j)$$

$$\Rightarrow E[X] = \sum_{j=1}^{n} P(E_j)\, E[X \mid E_j]$$

More generally : $E[g(X)] = \sum_{j=1}^{n} P(E_j)\, E[g(X) \mid E_j]$

---

(2) Examples:

Example 2.10.6 Let $0 < p < 0.5$. Suppose there are two biased coins. The first coin shows heads with probability p and the second coin shows heads with probability q, where $q = 1 - p$. Consider the following two stage experiment. First, select one of the two coins at random, with each coin being selected with probability one half, and then flip the selected coin n times. Let X be the number of times heads shows. Compute the pmf, mean, and standard deviation of X.

C_1 = event that coin 1 is selected. C_2 = event that coin 2 is selected

$$p_X(k) = P(X=k) = P(C_1)P(X=k \mid C_1) + P(C_2)P(X=k \mid C_2) = \frac{1}{2}\cdot\binom{n}{k} p^k (1-p)^{n-k} + \frac{1}{2}\cdot\binom{n}{k} q^k (1-q)^{n-k}$$

$$E[X] = P(C_1)E[X \mid C_1] + P(C_2)E[X \mid C_2] = \frac{1}{2}\cdot np + \frac{1}{2}\cdot nq = \frac{n}{2}$$

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2, \quad E[X^2] = P(C_1)E[X^2 \mid C_1] + P(C_2)E[X^2 \mid C_2] = \frac{1}{2}\cdot(n^2p^2 + np(1-p)) + \frac{1}{2}(n^2q^2 + nq(1-q))$$

$$\Rightarrow \sigma_X = \sqrt{\mathrm{Var}(X)} = \sqrt{\frac{1}{2}(n^2p^2 + n^2q^2) + npq - \frac{n^2}{4}}$$

2.33. [**Which airline was late?**]
Three airlines fly out of the Bloomington airport:

- American has five flights per day; 20% depart late,
- AirTrans has four flights per day; 5% depart late,
- Delta has nine flights per day; 10% depart late.

(a) What fraction of flights flying out of the Bloomington airport depart late?

(b) Given that a randomly selected flight departs late (with all flights over a long period of time being equally likely to be selected) what is the probability the flight is an American flight?

Let X_A = number of delayed flight of american, X_T = number of delayed flight of Air Tans, X_D = number of

Let X_A = number of delayed flight of american, X_T = number of delayed flight of Air Tans, X_D = number of delayed flight of Delta. Notice that X_A is $Bi(5, 0.2)$, X_T is $Bi(4, 0.05)$, and X_D is $Bi(9, 0.1)$.

(a) Suppose that we pick a random flight.

$$P(\{\text{random flight is late}\}) = P(A \text{ is selected})\, P(\text{late departure} \mid A \text{ is selected}) + P(T \text{ is selected})\, P(\text{late departure} \mid T \text{ is selected}) + P(D \text{ is selected})\, P(\text{late departure} \mid D \text{ is selected})$$

$$= \frac{5}{18} \times 0.2 + \frac{4}{18} \times 0.05 + \frac{9}{18} \times 0.1$$

(b) $P(A \text{ is selected} \mid \text{selected one departures late})$

$$= \frac{P(\text{selected one departures late} \mid A \text{ is selected})\, P(A \text{ is selected})}{P(\text{selected one departures late})}$$

$$= \frac{0.2 \times \frac{5}{18}}{\frac{5}{18} \times 0.2 + \frac{4}{18} \times 0.05 + \frac{9}{18} \times 0.1}$$

2.29. **[Estimation of signal amplitude for Poisson observation]**
The number of photons X detected by a particular sensor over a particular time period is assumed to have the Poisson distribution with mean $1 + a^2$, where a is the amplitude of an incident field. It is assumed $a \geq 0$, but otherwise a is unknown.

(a) Find the maximum likelihood estimate, $\widehat{a}_{ML}$, of a for the observation $X = 6$.

(b) Find the maximum likelihood estimate, $\widehat{a}_{ML}$, of a given that it is observed $X = 0$.

(a) . X is distributed as $Poi(1 + a^2)$

. We observe $X = 6$

. $\hat{a}_{ML} = \underset{a \geq 0}{\arg\max}\, P(X = 6) = \underset{a \geq 0}{\arg\max}\, e^{-(1+a^2)} \frac{(1+a^2)^6}{6!}$

$$\frac{d}{dp} e^{-(1+a^2)} (1+a^2)^6 = -2a e^{-(1+a^2)} (1+a^2)^6 + e^{-(1+a^2)}\, 6 \times 2a \times (1+a^2)^5$$

$$= (1+a^2)^5 e^{-(1+a^2)} \times 2a\,(6 - (1+a^2))$$

$$\frac{d}{dp} e^{-(1+a^2)} (1+a^2)^6 = 0 \Rightarrow 6 = 1 + a^2 \Rightarrow a = \pm\sqrt{5}, \quad \hat{a}_{ML} = \sqrt{5} \text{ is acceptable.}$$

You should also check that it is a local maximum.

(b) Following the same logic

$$\frac{d}{dp} e^{-(1+a^2)} = -2a e^{-(1+a^2)} = 0 \Rightarrow \hat{a}_{ML} = 0$$

Of course, to maximize probability of observing $X=0$ parameter $\hat{a}_{ML}$ should be zero

2.27. [**Maximum likelihood estimation and the Poisson distribution**]
An auto insurance company wishes to charge monthly premiums based on an individual's risk factor. It defines the risk factor as the probability p that individual is involved in a auto accident during a trip. Assume that whether an accident occurred on one trip is independent of accidents occurring on others, i.e., the insurance company assumes that drivers are reckless and don't learn to be cautious after being in an accident. The insurance company assumes that each driver will be driving 120 trips a month.

(a) Determine the maximum likelihood estimate of the risk factor $\hat{p}_{ML}$ if no accidents are reported by a driver in a month. Repeat for the cases when the driver reports 1, 2 and 3.

(b) Assume that the actual value of $p = 0.01$. Compute the approximate values of $P\{X = k\}$ for $k = 0, 1, 2, 3$ using the Poisson approximation to the binomial distribution, and compare those approximations to the actual probabilities computed using the binomial distribution.

(a) X = number of accidents in 120 trips, X is $Bi(120, p)$

For a Binomial r.v. $Bi(n,p)$, $P(X=k)$ maximized when

$$\frac{d}{dp}\binom{n}{k} p^k (1-p)^{n-k} = 0 \Rightarrow k p^{k-1}(1-p)^{n-k} - (n-k)p^k(1-p)^{n-k-1} = 0 \Rightarrow k(1-p) = (n-k)p \Rightarrow p = \frac{k}{n}$$

Which is a maximum. (see Example 2.8.1 in book)

$\Rightarrow$ Given $X=0$, $\hat{p}_{ML} = 0$

Given $X=1$, $\hat{p}_{ML} = \frac{1}{120}$

Given $X=2$, $\hat{p}_{ML} = \frac{2}{120}$

Given $X=3$, $\hat{p}_{ML} = \frac{3}{120}$

(b) using Poissson estimation, $\lambda = np = 120 \cdot 0.01 = 1.2$

$P(X=0) = e^{-\lambda} = e^{-1.2}$, $P(X=1) = e^{-\lambda}\frac{\lambda}{} = 1.2 \times e^{-1.2}$, $P(X=2) = e^{-\lambda}\frac{\lambda^2}{} = e^{-1.2} \times \frac{(1.2)^2}{}$

(b) using Poissson estimation, $\lambda = np = 120 \cdot 0.01 = 1.2$

$P(X=0) = e^{-\lambda} = e^{-1.2}$, $P(X=1) = e^{-\lambda}\frac{\lambda}{1} = 1.2 \times e^{-1.2}$, $P(X=2) = e^{-\lambda}\frac{\lambda^2}{2!} = e^{-1.2} \times \frac{(1.2)^2}{2}$

2.25. **[Ultimate verdict]**

Suppose each time a certain defendant is given a jury trial for a particular charge (such as trying to sell a seat in the US Senate), an innocent verdict is given with probability q_I, a guilty verdict is given with probability q_G, and a mistrial occurs with probability q_M, where q_I, q_G, and q_M are positive numbers that sum to one. Suppose the prosecutors are determined to get a guilty or innocent verdict, so that after any number of consecutive mistrials, another trial is given. The process ends immediately after the first trial with a guilty or innocent verdict; appeals are not considered. Let T denote the total number of trials required, and let I denote the event that the verdict for the final trial is innocent.

(a) Find $P(I|T = 1)$. Express your answer in terms of q_I and q_G.

(b) Find the pmf of T.

(c) Find $P(I)$. Express your answer in terms of q_I and q_G.

(d) Compare your answers to parts (a) and (c). For example, is one always larger than the other?

(a) $P(I|T=1) = \dfrac{P(I \cap T=1)}{P(T=1)} = \dfrac{q_I}{1-q_M} = \dfrac{q_I}{q_I+q_G}$

(b) $P(T=k) = P($ first k-1 trials were misstrials & last one was successful $)$

$= q_M^{k-1} \cdot (1-q_M)$

Notice that T is $Geo(1-q_M)$ randomvariable

(c) $P(I) = \sum_{K=1}^{\infty} P(I \cap T=K) = \sum_{K=1}^{\infty} (1-q_M)^{K-1} q_I = q_I \cdot \dfrac{1}{1-q_M} = \dfrac{q_I}{q_I+q_G}$

(d) Notice that $P(I) = \dfrac{q_I}{q_I+q_G}$. It is important to note that $P(I|T=1) + P(G|T=1) = 1 = P(I) + P(G)$

2.28. **[Scaling of a confidence interval]**
Suppose the fraction of people in Tokyo in favor of a certain referendum will be estimated by a poll. A confidence interval based on the Chebychev bound will be used (i.e. the interval is centered at $\widehat{p}$ with width $\frac{a}{\sqrt{n}}$ and confidence level $1 - \frac{1}{a^2}$, for some constant a, where $\widehat{p}$ is the fraction of the n people sampled that are in favor of the referendum.)

(a) Suppose the width of the confidence interval would be 0.1 for sample size $n = 300$ and some given confidence level. How many samples would be needed instead to yield a confidence interval that has only half the width, for the same level of confidence?

(b) What is the confidence level for the test of part (a)?

(c) Keeping the width of the confidence interval at 0.1 as in (a), how many samples would be required for a 96% confidence level?

(a) width $= 0.1 \Rightarrow \frac{a}{\sqrt{n}} = 0.1$ New width $= \frac{0.1}{2} = \frac{a}{\sqrt{n_{new}}}$ $\Rightarrow n_{new} = 4n = 1200$

confidence level $= 1 - \frac{1}{a^2}$ same confidence level $\Rightarrow$ same a

(b) $\frac{a}{\sqrt{n}} = 0.1 \Rightarrow a = 0.1 \times \sqrt{300} = \sqrt{3} \Rightarrow$ confidence level $= 1 - \frac{1}{3} = \frac{2}{3}$

(c) width $= 0.1 = \frac{a}{\sqrt{n}}$

$1 - \frac{1}{a^2} = 0.96 \Rightarrow a = 5$ $\Rightarrow \sqrt{n} = 50 \Rightarrow n = 2500$