Lecture 10 - 9/14

Wednesday, September 14, 2022 9:19 AM

Review:

① Poisson distribution: $Poi(\lambda)$, $P_X(k) = \frac{e^{-\lambda}\lambda^k}{k!}$

- it estimates Binomial distribution $Bi(n,p)$, where

  n is very large, p is very small, $np = O(1)$

- e.g. urbana-champaign has 42,000 people. Each person visits U of I credit union with probability 10^{-5}, independent of others, during hours of 8am to 9am

  $\rightsquigarrow$ # of visits to U of I credit union during 8am to 9am is a $Poi(0.42)$ random variable.

② Maximum likelihood estimator:

we are given a realization $X=k$. We know X belongs to a familly of distributions with unknown parameter:

e.g.: X is a $Bi(n,p)$ random variable where n is known but p is not known.

Our goal is to estimate the parameter, so that the probability of observing $X=k$ is maximized.

e.g.: $\hat{P}_{ML} = \arg\max_{p\in[0,1]} P(X=k)$ $\rightsquigarrow$ if our observation was $X\in A$ then $\hat{P}_{ML} = \arg\max_{p\in[0,1]} P(X\in A)$

③ Markov inequality: for non-negative random variable X and constant $c>0$

$$P(X \geq c) \leq \frac{E[X]}{c}$$

Today:

① Chebychev's inequality

② Confidence intervals

③ law of total probability and Bayes' formula

---

① Chebyshev's inequality.

Let $Y = |X - E[X]|$. Using markov inequality, we have

$$P(Y \geq a) = P(Y^2 \geq a^2) \leq \frac{E[Y^2]}{a^2}$$

In particular,

$$P\left(\left|X - E[X]\right| \geq a\right) \leq \frac{Var(X)}{a^2}$$

and

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

---

② Confidence interval:

We can use chebychev's inequality to get confidence interval. In particular, by chebychev's inequality, we are

$1 - \frac{Var(X)}{...}$ confident that $|X - E[X]| < a$, i.e., We can use chebychev's inequality to get confidence interval. In particular, by chebychev's inequality, we are $1-\frac{Var(X)}{a^2}$ confident that $|X-E[X]|<a$, i.e.,

$$P(-a < X-E[X] < a) \leqslant 1-\frac{var(X)}{a^2}$$

Main example: Given a large population, each individual agrees with the new policy with probability p independent of others. We sample n individuals at random, and # of people who was in favor of the policy was X. We estimat p using $\hat{p}=\frac{X}{n}$.

$$P(|X-np| \geqslant a\sigma_X) \leqslant \frac{1}{a^2}$$

Notice that $\sigma_X=\sqrt{np(1-p)}$. Hence,

$$P\left(\left|\frac{X}{n}-p\right| \geqslant a\sqrt{\frac{p(1-p)}{n}}\right) \leqslant \frac{1}{a^2}$$

or equivalenty

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

In here,

- $\hat{p}=\frac{X}{n}$ is our estimate and is random
- $\hat{p}$ is point estimator of p
- $\left(\hat{p}-a\sqrt{\frac{p(1-p)}{n}}, \hat{p}+a\sqrt{\frac{p(1-p)}{n}}\right)$ is interval estimator of p

Notice that interval estimator depends on p (which is unknown)! Using $p(1-p)\leqslant\frac{1}{4}$ (why!), we can use the following interval estimator:

- $\left(\hat{p}-\frac{a}{2\sqrt{n}}, \hat{p}+\frac{a}{2\sqrt{n}}\right)$ is interval estimator of p

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

- $1-\frac{1}{a^2}$ is the confidence level (we are $1-\frac{1}{a^2}$ confident that p is in within $\frac{a}{2\sqrt{n}}$ of point estimator $\hat{p}=\frac{X}{n}$.)
- $\frac{a}{2\sqrt{n}}$ is also called half-width

important note: p is fixed but unknown, $\hat{p}$ and $\left(\hat{p}-\frac{a}{2\sqrt{n}}, \hat{p}+\frac{a}{2\sqrt{n}}\right)$ are random and based on our observation of X.

---

③ law of total probability and Bayes' formula:

- Suppose that $\Omega = E_1\cup E_2\cup\cdots\cup E_n$, where $E_i\in\mathcal{F}$ and $P(E_i)>0$. Consider an event $A\in\mathcal{F}$

$$\begin{aligned} P(A) &= P(A\cap\Omega) \\ &= P(A\cap(E_1\cup E_2\cup\cdots\cup E_n)) \\ &= \sum_{i=1}^{n} P(A\cap E_i) = \sum_{i=1}^{n} P(E_i)P(A|E_i) \end{aligned}$$

$$= P(A \cap (E_1 \cup E_2 \cup \cdots \cup E_n))$$
$$= \sum_{i=1}^{n} P(A \cap E_i) = \sum_{i=1}^{n} P(E_i)\, P(A|E_i)$$

This is called low of total probability.

- Using conditional probability, for any $A, B \in \mathcal{F}$ with $P(A)>0$ and $P(B)>0$ we have:

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

This is called Bayes' formula.

- Combining Bayes's formula and conditional probability:

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

Problem 3.31.

Let $S = X_1 + X_2 + X_3 + X_4$. Suppose that X_is are independent $Ber(p)$ random variables:

$$P(S=2) = \binom{4}{2} p^2(1-p)^2$$

$$P(S=2 \cap X_1=1) = P(X_2+X_3+X_4 = 1 \cap X_1=1) = \binom{3}{1} p(1-p)^2 \cdot p$$
$$P(S=2 | X_1=1) = P(X_2+X_3+X_4 = 1 | X_1=1) = P(X_2+X_3+X_4 = 1) = \binom{3}{1} p(1-p)^2$$

$$P(X_1=1 | S=2) = \frac{P(S=2|X_1=1)\, P(X_1=1)}{P(S=2)} = \frac{\binom{3}{1} p(1-p)^2 \cdot p}{\binom{4}{2} p^2(1-p)^2}$$