Lecture - 16

Hypergeometric r.v.

Urn → N balls → population

m white

$N-m$ black

Select a sample of n balls, without replacement

X: no. of white balls selected

$$P(X=i) = \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}}$$

e.g. Count wildlife

Estimate the no. of lions in Gir.

N = total no. of lions, unknown, we want to estimate this

Initially capture & mark m lions

After 3–6 months, you capture n lions.

You count how many of them are marked?

Let this be i

$N = \cancel{n+m}\ \binom{nm}{i}$ ✓

$P(X=i) =$

$m = 50, \quad n = 30, \quad i = 3$

$P(X=0) =$

$P(X=1) =$

$P(X=2) =$

$\boxed{P(X=3)} =$ should have the maximum likelihood or maximum probability

$\vdots$

$P(X=30) =$

$$P_i(N) = \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}}$$

$$\frac{P_i(N)}{P_i(N-1)} = \frac{\cancel{\binom{m}{i}}\binom{N-m}{n-i}\binom{N-1}{n}}{\binom{N}{n}\cancel{\binom{m}{i}}\binom{N-1-m}{n-i}}$$

$$= \frac{(N-m)(N-n)}{N(N-m-n+i)}$$

this will be ≥ 1 if

$$(N-m)(N-n) > N(N-m-n+i)$$

$$\boxed{N \leq \frac{mn}{i}}$$

$$p_i(2) \geq p_i(1) \checkmark$$

$$p_i(3) \geq p_i(2) \checkmark$$

$$\vdots$$

$$p_i\left(\frac{mn}{i}\right) \geq p_i\left(\frac{mn}{i}-1\right) \checkmark$$

After this, it starts decreasing

e.g. The no. of traffic accidents in Berkley, California on 10 randomly chosen non raining days in 1998 is

4, 0, 6, 5, 2, 1, 2, 0, 4, 3

Use this data to estimate the probability that there are ≤ 2 accidents on a random day.

$\frac{5}{10} = 0.5$

$$\lambda = \frac{4+0+6+5+2+1+2+0+4+3}{10}$$

$$\lambda = 2.7$$

$$P(X \leq 2) = P(X=0) + P(X=1) + P(X=2)$$

$$= e^{-\lambda}\left(\frac{\lambda^0}{0!} + \frac{\lambda^1}{1!} + \frac{\lambda^2}{2!}\right)$$

$$= e^{-2.7}\left(1 + 2.7 + \frac{2.7^2}{2}\right)$$

$$= \boxed{0.4936}$$

Poisson random variable

Let $X_1, X_2, \cdots, X_n$ are independent Poisson random variables, each having mean λ. Determine the maximum likelihood estimator of λ.

Sol: likelihood function

$$f(x_1, x_2, \cdots, x_n \mid \lambda) = \frac{e^{-\lambda} \lambda^{x_1}}{x_1!} \frac{e^{-\lambda} \lambda^{x_2}}{x_2!} \cdots \frac{e^{-\lambda} \lambda^{x_n}}{x_n!}$$

$$\log f = -n\lambda + \left(\sum x_i\right)\log\lambda + \log c$$

$$\frac{d\log f}{d\lambda} = -n + \frac{\sum x_i}{\lambda} = 0$$

$$\boxed{\lambda = \frac{\sum x_i}{n}}$$

Let $X_1, X_2, \cdots, X_n$ are independent normal random variables, each with unknown mean μ and unknown standard deviation σ.

likelihood function

$$f(x_1, x_2, \cdots, x_n \mid \mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[\frac{-(x_i-\mu)^2}{2\sigma^2}\right]$$

$$= \frac{1}{(2\pi)^{n/2}} \cdot \frac{1}{\sigma^n} \exp\left[\frac{-\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}\right]$$

$$\log f = -\frac{n}{2}\log(2\pi) - n\log\sigma - \sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma^2}$$

Maximize this over σ and μ

$$\frac{\partial f}{\partial \mu} = 0 \qquad \sum_{i=1}^{n}\frac{(x_i-\mu)}{2\sigma^2} = 0$$

$$\frac{\partial f}{\partial \sigma} = 0 \qquad -\frac{n}{\sigma} + \sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^3} = 0$$

$$\sum (x_i - \mu) = 0$$

$$\Rightarrow \boxed{\mu = \frac{\sum x_i}{n}}$$

$$\sum_{i=1}^{n} \frac{(x_i - \mu)^2}{\sigma^3} = \frac{n}{\sigma}$$

$$\boxed{\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n}}$$

Uniform random variable.

2, 3, 5, 7, 1, 4, 10

what is your estimate for the mean?

$x_1, x_2, \ldots, x_n$ are samples from a uniform distribution $(0, \theta)$, where θ is unknown.

likelihood function = joint density = ?

$f(x_1, x_2, \ldots, x_n | \theta) =$

$\underbrace{\frac{1}{\theta} \cdot \frac{1}{\theta} \cdots \frac{1}{\theta}}_{n \text{ times}}$

$= \frac{1}{\theta^n}$

Maximize this over θ

As θ increases, f decreases

You want to choose minimim possible value of θ.

$$\theta = \max\{x_1, x_2, \cdots, x_n\}$$

$$\text{mean} = \frac{\theta}{2}$$

$$\boxed{= \frac{\max\{x_1, x_2, \cdots, x_n\}}{2}}$$