Recap :- $L(\theta)$ and $\hat{\theta}$ for one and more than one observation.

19th Jan, 2024

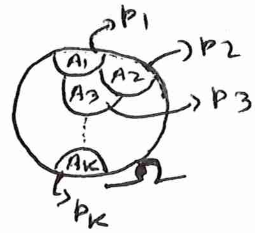
Section 2.2 :- Likelihoods based on frequency tables

Frequency tables :

They are used to present results of an experiment by counting the number of occurances in each event.

- n independent trials
- k mutually exclusive events
- Ω is partitioned by the events $A_1, A_2, \cdots, A_k$
- probabilities $p_1, p_2, \cdots, p_k$ are constant.

$$p_i = \mathbb{P}(A_i \text{ occurs}) \; ; \; i = 1, 2, \cdots, k.$$



Frequency tables looks like,

Event	A_1	A_2	....	A_k
obs. freq	x_1	x_2	....	x_k

$\longrightarrow \bigcup_{i=1}^{k} A_i = \Omega$

$\longrightarrow \sum_{i=1}^{k} x_i = n$

$$(X_1, X_2, \cdots, X_k) \sim \text{Multinomial}\,(n, p_1, p_2, \cdots, p_k),$$

where $\sum_{i=1}^{k} p_i = 1$ and $p_1, p_2, \cdots, p_k \in [0,1]$

Note:- p_i's are functions (sometimes) of a more relevant parameter θ, i.e., $p_i = p_i(\theta)$, $i = 1, 2, \cdots, k$.

We can use the distribution of the frequencies to :

a) Estimate $\hat{\theta}$ using $L(p_1(\theta), p_2(\theta), \cdots, p_k(\theta))$

b) Compare $x_1, x_2, \cdots, x_k$ with their estimated expected values $E(\hat{x}_i) = n\hat{p}_i = n\hat{p_i(\theta)} = np_i(\hat{\theta})$

for $i = 1, 2, \cdots, k$.

Let's Look at the Example 2.2.1, of the Complete Lecture Notes (Pg ~ 18)

The given table,

$Y_i \rightarrow$	# hits required to fracture	1	2	3	≥ 4	Total
$X_i \rightarrow$	# specimens	112	36	22	30	200

Let, Y = # hits needed to break the plastic

$\theta = \mathbb{P}$(not fracturing)

$Y_1, Y_2, \cdots, Y_{200}$ is our random sample

where, $Y \sim$ Neg. Bin $(r = 1, p = 1 - \theta)$

or, $Y \sim Geo(p = 1 - \theta)$

The pmf of Y is : $f(y; p) = p(1-p)^{y-1}; \; p \in [0,1]$

in terms of θ : $f(y;\theta) = \theta^{y-1}(1-\theta)^{\cancel{y-1}}; \; \theta \in [0,1]$

with $L(\theta; y)$, we can find $\hat{\theta} = 1 - 1/\bar{y}$.

But we donot have all the data necessary to evaluate $\hat{\theta}$. So, we use $L(\theta; \underset{\sim}{x})$ instead.

Let's express our p_i's in terms of θ:

$$p_1 = \mathbb{P}(Y=1) = 1-\theta.$$

$$p_2 = \mathbb{P}(Y=2) = \theta(1-\theta)$$

$$p_3 = \mathbb{P}(Y=3) = \theta^2(1-\theta)$$

$$\begin{aligned} p_4 = \mathbb{P}(Y \geq 4) &= 1 - \mathbb{P}(Y \leq 3) \\ &= 1-(1-\theta) - \theta(1-\theta) - \theta^2(1-\theta) \\ &= \theta^3 \end{aligned}$$

The Multinomial pmf is then:

$$f(\underset{\sim}{x}; \theta) = \binom{n}{x_1, x_2, x_3, x_4} (1-\theta)^{x_1} \cdot [\theta(1-\theta)]^{x_2} \cdot [\theta^2(1-\theta)]^{x_3} \cdot [\theta^3]^{x_4}$$

$$L(\underset{\sim}{x}; \theta) = (1-\theta)^{x_1+x_2+x_3} \theta^{x_2+2x_3+3x_4} \; ; \theta \in [0,1]$$

Find the MLE $\hat{\theta}$.

Here, $x_1+x_2+x_3 = 170$; $x_2+2x_3+3x_4 = 170$

$$\therefore L(\underset{\sim}{x}; \theta) = (1-\theta)^{170} \theta^{170} \; ; \theta \in [0,1]$$

Now, $\ell(\underset{\sim}{x}; \theta) = 170 \log(\theta) + 170 \log(1-\theta)$;

$\theta \in (0,1)$

* no boundaries.

$$l'(\theta) = \frac{170}{\theta} - \frac{170}{1-\theta}$$

Equating to Zero, $l'(\theta) = 0$

we will get, $\hat{\theta} = 1/2$

i.e, $\boxed{\text{The MLE of } \theta \text{ is } \hat{\theta} = 1/2}$

(Do the 2nd derivative test)

We can substitute this value into the p_i's

$$\hat{p}_1 = 1/2 \;;\; \hat{p}_2 = 1/4 \;;\; \hat{p}_3 = 1/8 \;;\; \hat{p}_4 = 1/8$$

Now we can compute the expected frequencies:

$$E(X_i) = n\hat{p}_i$$

$y_i \to$ # hits required	1	2	3	≥ 4	Total
$X_i \to$ # specimens	112	36	22	30	200
$\hat{E}(X_i) \to$ Estimated (E(# specimens))	100	50	25	25	200
$\lvert X_i - \hat{E}(X_i) \rvert$	12	14	3	5	0

So, there is some variation between the freq's and their expected values, especially for the 1st and 2nd events in the table.
It may be the case that the model, $Y \sim geo(1-\theta)$ is not a good fit.

↳ We will briefly revisit this example in chapter 3 as well and we will discuss Goodness of fit tests in chapter 4.

↳ But next class, we'll look at Section 2.5 (Relative Likelihood functions).