36-705 11/07/2020

Lecture Notes 6: Likelihood Function - Review

· Likelihood fn:-

$x^n = (x_1, \ldots, x_n)$

$p(x^n; \theta) = p(x_1, \ldots, x_n; \theta) \qquad \theta \in \Theta$

$L: \Theta \longrightarrow [0, \infty) \qquad L(\theta) \equiv L(\theta; x^n) = p(x^n; \theta)$

we view:- where x^n is fixed, θ varies in Θ

· IID data:- $L(\theta) = \prod_{i=1}^{n} p(x_i; \theta)$

(*) In general, $\int L(\theta)\, d\theta \neq 1$

log-like $\ell(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log p(x_i; \theta)$

(*) nuances of distinction between likelihood, probability

- still not entirely clear.
- WIKI.
  - Goodness of fit of a statistical model to a sample of data for given values of the unknown parameters.
  - formed from joint p.d. of the sample, but viewed and used as a function of the parameters only.
  - Hence random variables are treated as fixed at the observed values
  - $L(\theta)$ describes a hypersurface whose peak, if it exists, represents the combination of model parameter values that maximise the probability of drawing the sample obtained.
  - shape and curvature of hypersurface can represent info about estimate stability.
  - likelihood is the probability that data is observed when the true value of the parameter is θ, p.d. over data, not over param (?)

Fisher: Probability vs likelihood → 2 measures of rational belief

Probability → knowing the population we can express our incomplete knowledge of, or expectation of, the sample in terms of probability

Likelihood → knowing the sample, we can express our incomplete knowledge of the population in terms of likelihood.

(*) interpretation of likelihood,
if you go deeper, is not fully resolved (frequentist vs Bayesian vs likelihoodist vs AIC)

Example 1 (A0)

$X = (X_1, X_2, X_3) \sim \text{Multinomial}(n, p)$ typo in lecture notes -!

$p = (p_1, p_2, p_3) = (\theta, \theta, 1-2\theta)$

$$p(x;\theta) = \binom{n}{x_1, x_2, x_3} p_1^{x_1} p_2^{x_2} p_3^{x_3} \propto \theta^{x_1+x_2}(1-2\theta)^{x_3}$$

Multinomial coeff =

$$\binom{n}{k_1, k_2, k_3} = \frac{n!}{k_1!k_2!k_3!}$$

$x = (1, 3, 2)$.

$$L(\theta) = \frac{6!}{1!3!2!}\theta^1\theta^3(1-2\theta)^2 \propto \theta^4(1-2\theta)^2$$

$x = (2, 2, 2)$

$$L(\theta) = \frac{6!}{2!2!2!}\theta^2\theta^2(1-2\theta)^2 \propto \theta^4(1-2\theta)^2$$

- likelihood same for both datasets.

(A1) - Proport. calc.

$X_1, \ldots, X_n \sim N(\mu, 1)$

$$L(\mu) = p(x^n;\mu) = \prod_{i=1}^{n} p(x_i,\mu) = \prod_{i=1}^{n} \frac{1}{(2\pi)^{1/2}} \exp\left\{-\frac{1}{2}(x_i-\mu)^2\right\}$$

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

$$= \frac{1}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\left((x_i-\bar{x})+(\bar{x}-\mu)\right)^2\right\}$$ - separating out μ terms.

$$= \frac{1}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\bar{x})^2 - \frac{1}{2}\sum_{i=1}^{n}(\bar{x}-\mu)^2\right\}$$

$$\Rightarrow L(\mu) = \frac{1}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\bar{x})^2\right\} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\bar{x}-\mu)^2\right\}$$

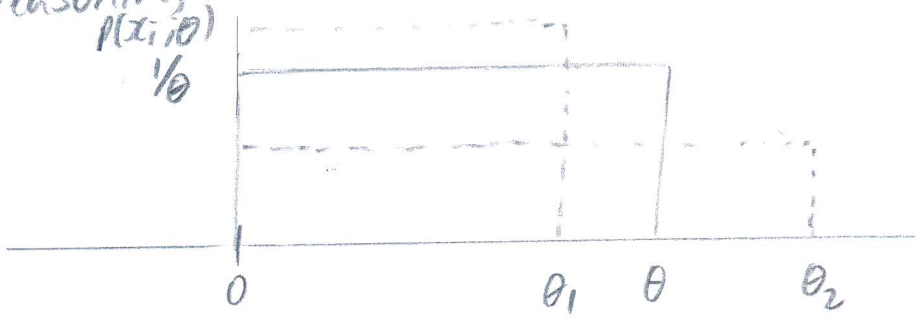
$$\propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\bar{x}-\mu)^2\right\}$$

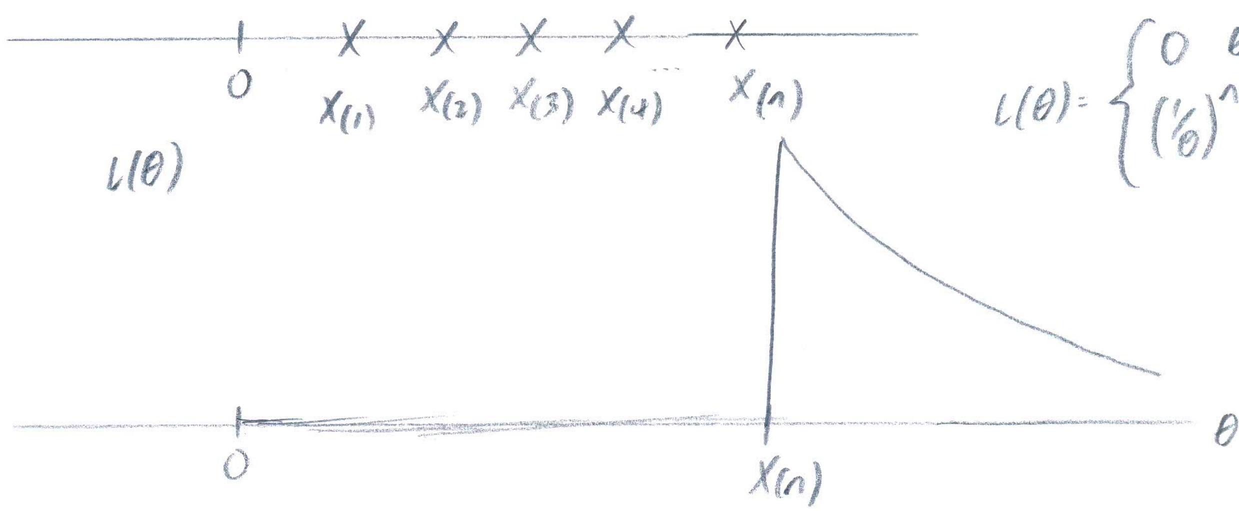
- dropping terms with no μ
- likelihood equivalent up to const. of proport.

- As required.

example 5 (p12)

(*) nuances of diagram a little unclear.
- in particular, reasoning about how we view likelihood function of getting a particular dataset, at a particular value of param θ.
- the reasoning missed on written notes; redo here.





$$L(\theta) = \begin{cases} 0 & \theta < X_{(n)} \\ \left(\frac{1}{\theta}\right)^n & \theta > X_{(n)} \end{cases}$$

- consider various values of θ (in particular θ_1 and θ_2).
- consider $L(\theta_1) = \prod_{i=1}^{n} p(x_i;\theta_1)$ and $L(\theta_2)$.

- $L(\theta_1)$ corresponds to situation where $0 < X_{(n)} < \theta$; $L(\theta_2)$ to situation where $\theta > X_{(n)}$

$L(\theta_1) = 0$ as all terms $p(x_i;\theta_1)$ where $x_i > \theta_1$ will be 0.

i.e. probability of data points which are outside the rage of $unif(0,\theta)$ will be 0; and hence entire likelihood will be 0.

$L(\theta_2) = \left(\frac{1}{\theta}\right)^n$

- Okay, all clarified

Theorem 4:

- $x^n \sim y^n$ if $L_{X_{(1)},\dots,X_{(n)}}(\theta) \propto L_{Y_{(1)},\dots,Y_{(n)}}(\theta)$ (i.e. $L(\theta|y^n) \propto L(\theta|x^n)$)
- The partition induced by $\sim$ is the minimal sufficient partition

Proof: OS1

(*) Note the relation between equivalence class, constant of proportionality, and equivalent statistics.

(*) Need to tighten understanding of this.

---

2. Likelihood, Sufficiency, Likelihood Principle

- Likelihood function is a minimal sufficient statistic (MSS)
- If we define the equivalence relation:-

  $x^n \sim y^n$ when $L(\theta;x^n) \propto L(\theta;y^n)$ then the resulting partition is minimal sufficient.

ux: Illustration of why it is not entirely correct to say the likelihood function contains all information on the data:-

- $C = \{c_1, \dots, c_N\}$ - finite set of constants
- $c_j \in \{0,1\}$ - assume for conv.
- Let $\theta = \frac{1}{N}\sum_{j=1}^{N} c_j$
- Suppose we want to estimate θ.

· let $S_1, S_2, \ldots, S_n \sim \text{Bernoulli}(\pi)$, where π is known

· If $S_i = 1$, you get to see c_i, if $S_i = 0$, you do not (survey sampling).

likelihood function: $L(\theta) = \prod_{i=1}^{n} \pi^{S_i}(1-\pi)^{1-S_i}$

$$= \pi^S(1-\pi)^{n-S} \qquad \text{where } S = \sum_{i=1}^{N} S_i$$

(O1): unknown parameter θ does not appear in likelihood

- No unknown params in likelihood - π is known.

· likelihood fn contains no information.

· ~~But~~ we can estimate θ

· let
$$\hat{\theta} = \frac{1}{N\pi}\sum_{j=1}^{N} c_j S_j \qquad \text{O/S 2}$$

$$\mathbb{E}[\hat{\theta}] = \mathbb{E}\left[\frac{1}{N\pi}\sum_{j=1}^{N} c_j S_j\right] = \mathbb{E}\left[\qquad\right] = \mathbb{E}[\theta]$$

via Hoeffding: (what is underlying r.v.?)

$$?(|\hat{\theta} - \theta| > \epsilon) \leq 2e^{-2n\epsilon^2\pi^2} \qquad (?) \quad \text{O/S 3}$$

· $\hat{\theta}$ is close to θ with high probability

(O/S 4): do not understand why this illustrates point.

(*) There is some contention on the proposition above which LW takes a stance on.