VI Pure Likelihood Theory

- minimal requirements
- ingredients: data x and the model $M = \{f_\theta : \theta \in \Theta\}$
- discrete case $f_\theta(x)$ = probability that x occurs when θ is true.
- basic idea: θ values that give higher probability to x are preferred to those that give lower probabilities
- likelihood preference ordering (a total ordering on Θ)

  $\theta_1 \preceq \theta_2$ (θ_1 is not preferred to θ_2) whenever $f_{\theta_1}(x) \le f_{\theta_2}(x)$
- note - preference ordering remains the same if we replace the function $f_\cdot(x) : \Theta \to [0,\infty)$ by $c f_\cdot(x)$
- so likelihood function is defined as any function $L(\cdot \mid x) : \Theta \to [0,\infty)$ s.t. $L(\theta \mid x) = c f_\theta(x)$ for some $c > 0$ $\forall \theta \in \Theta$
- note the set of ratios $\dfrac{L(\theta_1 \mid x)}{L(\theta_2 \mid x)}$ for $\theta_1, \theta_2 \in \Theta$

  remains invariant under the choice of c and conversely $\left(\dfrac{g_1(\theta)}{g_1(\theta_1)} = \dfrac{g_2(\theta)}{g_2(\theta_1)} \ \forall \theta \therefore g_1 = c g_2 \right.$ where $\left. c = g_1(\theta_1)/g_2(\theta_1) \right)$

- continuous case $P_\theta(B_\delta(x)) \approx f_\theta(x)\,\mathrm{Vol}(B_\delta(x))$
  so again use $f_\theta(x)$ or $c f_\theta(x)$ for some $c>0$ for preference ordering.

- <u>estimation of θ</u>

- given likelihood preference ordering the value $\theta_{MLE}(x)$ which maximizes $L(\theta \mid x)$ is the most preferred value and is the <u>maximum likelihood estimate</u> of θ

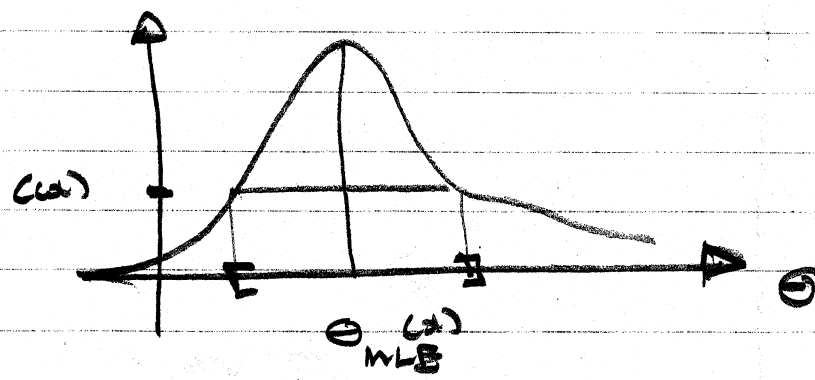
eg $x = (x_1, \ldots, x_n) \sim \text{Bernoulli}(\theta)$
$\theta \in [0,1]$

$\theta_{MLE}(x) = \bar{x}$.

- for error assessment we quote a <u>likelihood region</u>

$C(x) = \{\theta : L(\theta|x) \geq c(x)\}$

for some $c(x) \in [0, \infty)$



$C(x)$

since if $\theta_1 \in C(x)$ and $L(\theta_2|x) \geq L(\theta_1|x)$
then we must have $\theta_2 \in C(x)$

- how to choose $c(x)$?

<u>law of the likelihood</u>

the ratio $L(\theta_1|x)/L(\theta_2|x)$ measures
the strength of the evidence supporting
θ_1 over θ_2.

- choose $c(x) = (1-\gamma)\, L(\theta_{MLE}(x)|x)$
for some $\gamma \in [0,1]$

- then $C_\gamma(x) = \{\theta : L(\theta|x)/L(\theta_{MLE}(x)|x) \geq 1-\gamma\}$

a γ-LR

= set of all θ values supported
$(1-\gamma)$ as much as best supported
value.

- **note** - γ is not a probability

- **hypothesis assessment**

- we want to assess $H_0 : \theta = \theta_0$

- compute the relative likelihood $\rho_{\theta_0}(x) = L(\theta_0|x)/L(\hat{\theta}(x)|x)$

- if $\rho_{\theta_0}(x)$ is large then we have evidence for H_0 and if small then we have evidence against H_0

- but what value gives evidence in favor?

- there is no intrinsic definition of evidence

eg $x = (x_1, \ldots, x_n) \overset{iid}{\sim} N(\theta, 1)$

- $\theta \in \mathbb{R}$.

- $L(\theta|x) = \prod_{i=1}^{n} \exp\{-\frac{1}{2}(x_i - \theta)^2\}$

$$= \exp\{-\tfrac{1}{2}\sum_{i=1}^{n}(x_i - \theta)^2\}$$

$$\sum_{i=1}^{n}(x_i-\theta)^2 = \sum_{i=1}^{n}(x_i - \bar{x} + \bar{x} - \theta)^2$$

$$= \sum_{i=1}^{n}(x_i-\bar{x})^2 + (\bar{x}-\theta)\sum_{i=1}^{n}(x_i - \bar{x})$$

$$+ n(\bar{x}-\theta)^2$$

$$= \sum_{i=1}^{n} (x_i - \bar{x})^2 + n(\bar{x} - \theta)^2$$

$$= \exp\{-\tfrac{1}{2}\sum_{i=1}^{n} (x_i - \bar{x})^2\} \exp\{-\tfrac{n}{2}(\bar{x} - \theta)^2\}$$

$\therefore$ $\bar{x}$ is sufficient (Factorization)

and since $\theta_{MLE}(x) = \theta$ we have that $\bar{x}$ is a MSS.

- take $L(\theta \mid x) = \exp\{-\tfrac{n}{2}(\bar{x} - \theta)^2\}$

so $L(\theta_{MLE}(x) \mid x) = 1$

$$C_\gamma(x) = \{\theta : \exp\{-\tfrac{n}{2}(\bar{x} - \theta)^2\} \geq 1 - \gamma\}$$

$$= \{\theta : (\bar{x} - \theta)^2 \leq -\tfrac{2}{n}\ln(1-\gamma)\}$$

$$= \left[\bar{x} - \sqrt{-\tfrac{2}{n}\ln(1-\gamma)},\ \bar{x} + \sqrt{-\tfrac{2}{n}\ln(1-\gamma)}\right]$$

- relative likelihood $\exp\{-\tfrac{n}{2}(\bar{x} - \theta_0)^2\} = P_{\theta_0}(x)$

– what about inference for $\alpha = \Psi(\theta)$?

eg $x = (x_1, \ldots, x_n) \sim N(\mu, \sigma^2)$ location-scale normal model

$\theta = (\mu, \sigma^2) \in \mathbb{R} \times (0, \infty)$

$\alpha = \sigma/\mu$ = coefficient of variation

a unit free measure of variability

– $L(\mu, \sigma^2 \mid x) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\{-(\frac{x_i - \mu}{\sigma})^2\}$

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

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

– for fixed σ^2 $L(\cdot, \sigma^2 \mid x)$ is maximized at $\mu = \bar{x}$ and as this ind of σ^2 we need only max

$L(\bar{x}, \sigma^2 \mid x) = (\sigma^2)^{-\frac{n}{2}} \exp\{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \bar{x})^2\}$

$\log L(\bar{x}, \sigma^2 \mid x) = -\frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \bar{x})^2$

$\frac{\partial \log L(\bar{x}, \sigma^2 \mid x)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (x_i - \bar{x})^2$

$= 0$

implies that $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = s^2$

and $\left.\frac{\partial^2 \log L(\bar{x}, \sigma^2 \mid x)}{\partial(\sigma^2)^2}\right|_{\sigma^2 = s^2}$

$$= \left.\left(\frac{n}{2\sigma^4} - \frac{n}{\sigma^6}s^2\right)\right|_{\sigma^2 = s^2}$$

$$= \frac{n}{2s^4} - \frac{n}{s^4} = \frac{-n}{2s^4} < 0$$

$\therefore$ we have proved that

$$(\mu, \sigma^2)_{MLE}(x) = (\bar{x}, s^2)$$

- also by factorization $(\bar{x}, s^2)$ is sufficient and since we can compute $(\bar{x}, s^2)$ from any LF we have that it is a MSS.
- but how do we estimate and assess hypotheses about $\psi = \sigma/\mu$?
- intuitively - estimate by $s/\bar{x}$ but then what about error assessment and how do we assess $H_0 : \psi = \psi_0$?

- no general solution but use the following

Def The <u>profile likelihood function</u> for $\psi = \Psi(\theta)$ based on model M is given by

$$L_\Psi(\psi|x) = \sup_{\theta \in \Psi^{-1}\{\psi\}} L(\theta|x)$$

- so to compute $L_\psi(\psi|x)$ we need to find the max of $L(\theta|x)$ for θ s.t. $\Psi(\theta) = \psi$

eg <u>location-scale normal</u>

- $L(\mu, \sigma^2|x) = (\sigma^2)^{-\frac{n}{2}} \exp\{-\frac{n}{2\sigma^2}(\bar{x}-\mu)^2\} \times$
$\exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \bar{x})^2\}$

- $\Psi(\mu, \sigma^2) = \sigma^2$

$\therefore \quad L_\Psi(\sigma^2|x) = (\sigma^2)^{-\frac{n}{2}} \exp\{-\frac{n}{2\sigma^2} s^2\}$

- $\Psi(\mu, \sigma^2) = \mu$

$$\log L(\mu, \sigma^2|x) = -\frac{n}{2}\log\sigma^2 - \frac{n}{2\sigma^2}(\bar{x}-\mu)^2 - \frac{n}{2\sigma^2}s^2$$

$$\frac{\partial}{\partial\sigma^2} = -\frac{n}{2\sigma^2} + \frac{n}{2\sigma^4}(\bar{x}-\mu)^2 + \frac{n}{2\sigma^4}s^4 = 0$$

$-\sigma^2 + (\bar{x}-\mu)^2 + s^2 = 0$

so $\sigma^2_{MLE}(\mu) = s^2 + (\bar{x}-\mu)^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2$

$\therefore L_\Psi(\mu|x) = \left(s^2 + (\bar{x}-\mu)^2\right)^{-\frac{n}{2}} \exp\{-\frac{n}{2}\}$

- $\Psi(\mu,\sigma^2) = \sigma/\mu \qquad \mu = \sigma/\psi$

$L(\mu,\sigma^2|x) = (\sigma^2)^{-\frac{n}{2}} \exp\{-\frac{n}{2\sigma^2}(\bar{x}-\sigma/\psi)^2\} \times$

$\times \quad \exp\{-\frac{n}{2\sigma^2}s^2\}$

$L_\Psi(\psi|x) = ?$ solve numerically

- for inference about $\psi = \Psi(\theta)$ use $L_\Psi(\psi|x)$ just as you would use a LF for θ

- assume $\theta_{MLE}(x)$ is unique.

<u>Lemma</u> If for each ψ, $\theta_\psi(x) \in \Psi^{-1}\{\psi\}$ exists uniquely such that $L_\Psi(\psi|x) = L(\theta_\psi(x)|x)$ then $\psi_{MLE}(x) = \Psi(\theta_{MLE}(x))$ and

$C_{\Psi,\gamma}(x) = \Psi C_\gamma(x)$

γ-profile LR for ψ $\qquad$ γ-LR for θ

Proof: We have $L_\Psi(\psi \mid x) = \sup_{\theta \in \Psi^{-1}\{\psi\}} L(\theta \mid x)$
$\leq \sup_{\theta \in \Theta} L(\theta \mid x) = L(\theta_{MLE}(x) \mid x)$ and so
$L_\Psi(\psi_{MLE}(x) \mid x) \leq L(\theta_{MLE}(x) \mid x)$.
Also for $\theta \in \Theta$, $L(\theta \mid x) \leq L_\Psi(\Psi(\theta) \mid x)$
and so $L(\theta_{MLE}(x) \mid x) \leq L_\Psi(\psi_{MLE}(x) \mid x)$.
Therefore $L(\theta_{MLE}(x) \mid x) = L_\Psi(\psi_{MLE}(x) \mid x)$.
Now $\theta_{\psi_{MLE}(x)}(x)$ is such that $L_\Psi(\psi_{MLE}(x) \mid x)$
$= L(\theta_{\psi_{MLE}(x)}(x) \mid x)$ and so $\theta_{\psi_{MLE}(x)}(x) = \theta_{MLE}(x)$
and $\Psi(\theta_{MLE}(x)) = \psi_{MLE}(x)$.

Now suppose $\theta \in C_\gamma(x)$ so $L(\theta \mid x)$
$\geq (1-\gamma) L(\theta_{MLE}(x) \mid x)$ which implies
$L_\Psi(\Psi(\theta) \mid x) = \sup_{\theta' \in \Psi^{-1}\Psi(\theta)} L(\theta' \mid x)$
$\geq (1-\gamma) L(\theta_{MLE}(x) \mid x) = (1-\gamma) L(\psi_{MLE}(x) \mid x)$
which implies $\Psi(\theta) \in C_{\Psi,\gamma}(x)$ and $\Psi C_\gamma(x) \subseteq C_{\Psi,\gamma}(x)$.

Now suppose $\psi \in C_{\Psi,\gamma}(x)$ so

$L_{\Psi}(\psi|x) = L(\theta_\psi(x)|x) \geq (1-\gamma) L_{\Psi}(\psi_{MLE}(x)|x)$
$= (1-\gamma) L(\theta_{MLE}(x)|x)$ and so $\theta_\psi(x) \in C_\gamma(x)$
which implies $C_{\Psi,\gamma}(x) \subseteq \Psi C_\gamma(x)$.

- for assessing hypothesis $H_0 : \Psi(\theta) = \psi_0$
compute relative profile likelihood

$$P_{\psi_0}(x) = \frac{L_{\Psi}(\psi_0|x)}{L_{\Psi}(\psi_{MLE}(x)|x)}$$

- important property of all likelihood inference

- <u>invariance</u>

- suppose $\beta = \beta(\psi) = \beta(\Psi(\theta)) = (\beta\circ\Psi)(\theta)$
and suppose β is 1-1

eg location-scale normal

$\psi = \Psi(\mu, \sigma^2) = \mu$

$\beta(\psi) = \psi^3$

- by Lemma $C_{\beta,\gamma}(x) = (\beta\circ\Psi)\, C_\gamma(x)$
$= \beta\left(\Psi C_\gamma(x)\right) = \beta\, C_{\Psi,\gamma}(x)$

and by 1-1 property $C_{\Psi,\gamma}(x) = \beta^{-1} C_{\beta,\gamma}(x)$

eg (cont'd)

so inferences about μ^3 are just transformations of inferences about μ and conversely

<u>problem</u>

- likelihood inferences about θ are based on the likelihood preference ordering and the law of the likelihood
- but in general $L_\Psi(\psi \mid x)$ is not a likelihood

eg <u>profile LF is not a LF</u>

- $\mathcal{X} = \{1, 2\}$ $\Theta = \{0, 1, 2\}$

$\theta \backslash x$	1	2
0	1/2	1/2
1	1/3	2/3
2	1/5	4/5

- $\psi = \Psi(\theta) = I_{\{0,1\}}(\theta) \in \{0, 1\}$

$L_{\Pi}(0 \mid 1) = 1/4$, $L_{\Pi}(0 \mid 2) = 4/5$

$L_{\Pi}(1 \mid 1) = 1/2$, $L_{\Pi}(1 \mid 2) = 2/3$

- so if L_{Π} is to be a LF there must be ~~statistic~~ T defined on $\mathcal{X}$ s.t. $L_{\Pi}(\cdot \mid x)$ is a LF based on observing $T(x)$

- but $L_{\Pi}(\cdot \mid 1)$ and $L_{\Pi}(\cdot \mid 2)$ are not proportional so T must be 1-1 but this implies that T has same model and so ~~same~~ LF as model

- various proposals have been made to fix this problem and in particular different types of LF's defined but nothing has worked (so far)

- Summary for pure likelihood theory

- positives - intuitively reasonable
  - always an answer
  - somewhat evidence-based

- negatives - no definition of evidence for or against only comparative
  - no justification for inference for marginal parameter
  - no model checking