Topic 1 Principles for Statistical Inference - Problems Class.

$\mathrm{Ev}(\xi, x)$

Strong likelihood principle.

If ξ_1, ξ_2 two experiments $f_{X_1}(x_1|\theta) = c(x_1, x_2) f_{X_2}(x_2|\theta)$ for all θ then
$\mathrm{Ev}(\xi_1, x_1) = \mathrm{Ev}(\xi_2, x_2)$.

Note: if $\xi_1 = \xi_2$ then $f_X(x_1|\theta) = c(x_1, x_2) f_X(x_2|\theta)$

$$L_X(\theta; x_1) = c(x_1, x_2) L_X(\theta; x_2)$$

$$\Rightarrow \frac{L_X(\theta; x_1)}{L_X(\theta; x_2)} = c(x_1, x_2).$$

$$\text{SLP} \Rightarrow \mathrm{Ev}(\xi, x_1) = \mathrm{Ev}(\xi, x_2).$$

i.e. $\mathrm{Ev}(\xi, x)$ should depend on ξ, x only through $L_X(\theta; x)$.
This is a function of θ for fixed x. Thus, if your method of inference depends on values of $x \in \mathcal{X}$ other than those you have observed you will violate the SLP.

What is the principle saying?

Suppose $f_X(x|\theta) = c(x, y) f_X(y|\theta)$ for observed x, y.

If
$$L_X(\theta_2; x) = 2 L_X(\theta_1; x) \text{ then}$$
$$L_X(\theta_2; y) = 2 L_X(\theta_1; y)$$

and so it doesn't matter whether I observe x or y.

Example (Berger and Wolpert (1988) - Example 10).

Consider two experiments: $\mathcal{X} = \{1, 2, 3\}$ and $\Theta = \{0, 1\}$.
$\mathcal{E}_1, \mathcal{E}_2$ constructed over the same Θ.

$\mathcal{E}_1$

	1	2	3
$f_{X_1}(x_1 \mid \theta = 0)$	0.9	0.05	0.05
$f_{X_1}(x_1 \mid \theta = 1)$	0.09	0.055	0.855

$\mathcal{E}_2$

	1	2	3
$f_{X_2}(x_2 \mid \theta = 0)$	0.26	0.73	0.01
$f_{X_2}(x_2 \mid \theta = 1)$	0.026	0.803	0.171

If I observe $x_1 = 1 = x_2$.

$$L_{X_1}(\theta; 1) = f_{X_1}(1 \mid \theta) = 0.9\,\mathbb{I}_{\{\theta=0\}} + 0.09\,\mathbb{I}_{\{\theta=1\}}$$

$$= 0.9\,(\underbrace{\mathbb{I}_{\{\theta=0\}} + 0.1\,\mathbb{I}_{\{\theta=1\}}}_{\text{function of } \theta})$$

$$L_{X_2}(\theta; 1) = f_{X_2}(1 \mid \theta) = 0.26\,\mathbb{I}_{\{\theta=0\}} + 0.026\,\mathbb{I}_{\{\theta=1\}})$$

$$= 0.26\,(\underbrace{\mathbb{I}_{\{\theta=0\}} + 0.1\,\mathbb{I}_{\{\theta=1\}}}_{\text{function of } \theta})$$

I have $L_{X_1}(\theta; 1) = c(1,1)\,L_{X_2}(\theta; 1)$ and so if the SLP is asserted then

$$Ev(\mathcal{E}_1, 1) = Ev(\mathcal{E}_2, 1).$$

Consider $H_0: \theta = 0$ versus $H_1: \theta = 1$

Accept H_0 if I observe $x = 1$ and reject otherwise.

involving values $x \in \mathcal{X}$ which are not observed.

In $\mathcal{E}_1$ $\quad \mathbb{P}(\text{Reject } H_0 \mid H_0 \text{ true}) = \overbrace{f_{X_1}(2 \mid \theta = 0) + f_{X_1}(3 \mid \theta = 0)}$
$= 0.1$

$\mathbb{P}(\text{Accept } H_0 \mid H_1 \text{ true}) = f_{X_1}(1 \mid \theta = 1) = 0.09.$

In $\mathcal{E}_2$ $\quad \mathbb{P}(\text{Reject } H_0 \mid H_0 \text{ true}) = f_{X_2}(2 \mid \theta = 0) + f_{X_2}(3 \mid \theta = 0)$
$= 0.74$

$\mathbb{P}(\text{Accept } H_0 \mid H_1 \text{ true}) = 0.026.$

The classicist would report very different information from each experiment, leading to a violation of the SLP.

Example 2: Bernoulli trials, parameter θ. (θ is the probability of a success).

$\mathcal{E}_1$: carry out n trials and let X denote the total number of successes

$$X \mid \theta \sim \text{Bin}(n, \theta)$$

fixed trials, count successes $\quad f_X(x \mid \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x} \quad x = 0, 1, \ldots, n$

$\mathcal{E}_2$: count the total number Y of trials up to and including the rth successes.

$$Y \mid \theta \sim \text{NBin}(r, \theta)$$

$$f_Y(y \mid \theta) = \binom{y-1}{r-1} \theta^r (1-\theta)^{y-r}$$

fixed successes, count the number of trials. $\quad y = r, r+1, \ldots$

Think about these as having different stopping rules.

binomial: always take the next observation (Bernoulli trial) until $n = 12$

negative-binomial: always take the next observation (Bernoulli trial) until I've seen three successes.

Suppose we observe $x = r = 3$ and $y = n = 12$.
In both cases, I've seen the same number of successes, 3, in the same number of trials, 12.

a). Consider $H_0: \theta = \frac{1}{2}$ versus $H_1: \theta < \frac{1}{2}$.

ξ_1: small values of x suggest H_1 rather than H_0.
ξ_2: large values of y suggest H_1 rather than H_0.

Consider $\mathbb{P}(X \leq 3 \mid \theta = \frac{1}{2}) = \sum_{x=0}^{3} \binom{12}{x} \left(\frac{1}{2}\right)^{12} = \frac{299}{4096} = 0.0730.$

$$\mathbb{P}(Y \geq 12 \mid \theta = \tfrac{1}{2}) = 1 - \mathbb{P}(Y \leq 11 \mid \theta = \tfrac{1}{2})$$
$$= 1 - \sum_{y=3}^{11} \binom{y-1}{2} \left(\frac{1}{2}\right)^{y} = \frac{67}{2048} = 0.0327.$$

Let $Ev(\xi_1, x)$ be significance test at 5%. Then we would <u>NOT</u> reject as $0.0730 > 0.05$
If $Ev(\xi_2, y)$ be significance test at 5% then we would H_0 as $0.0327 < 0.05$.

Inference using p-value depends on the stopping rule.

As SLP $\Rightarrow$ SRP then we also have a strong violation of the SLP.

$$f_X(x \mid \theta) = \binom{12}{3} \theta^3 (1-\theta)^9 \propto \theta^3 (1-\theta)^9$$
$$f_Y(y \mid \theta) = \binom{11}{2} \theta^3 (1-\theta)^9 \propto \theta^3 (1-\theta)^9$$
as functions of θ

b). Jeffreys prior is a noninformative prior / automatic.

$$\pi_X(\theta) \propto \sqrt{I_X(\theta)}$$

expectation calculated using $f_X(x|\theta)$ i.e. assuming θ is known

when $$I_X(\theta) = -E\left(\frac{\partial^2}{\partial\theta^2}\log f_X(x|\theta)\,\Big|\,\theta\right)$$

Fisher Information.

$X|\theta \sim Bin(n, \theta)$

$$\Rightarrow \log f_X(x|\theta) = \log\binom{n}{x} + x\log\theta + (n-x)\log(1-\theta)$$

$$\frac{\partial^2}{\partial\theta^2}\log f_X(x|\theta) = \frac{-x}{\theta^2} - \frac{(n-x)}{(1-\theta)^2}$$

$$I_X(\theta) = \frac{E(X|\theta)}{\theta^2} + \frac{(n - E(X|\theta))}{(1-\theta)^2}$$

$$= \frac{n}{\theta} - \frac{n}{1-\theta} = \frac{n}{\theta(1-\theta)}$$

The Jeffreys prior for X is $\pi_X(\theta) \propto \sqrt{\frac{n}{\theta(1-\theta)}} \propto \theta^{-\frac{1}{2}}(1-\theta)^{-1/2}$

i.e. $\theta \sim Beta(1/2, 1/2)$.

I can calculate the Jeffreys prior for the negative binomial.

In this case, $$I_Y(\theta) = \frac{r}{\theta^2} + \frac{\overset{= \frac{r}{\theta}}{E(Y|\theta)} - r}{(1-\theta)^2}$$

$$= \frac{r}{\theta^2} + \frac{r}{\theta(1-\theta)} = \frac{r}{\theta^2(1-\theta)}$$

The Jeffreys prior for Y is $\pi_Y(\theta) \propto \sqrt{\frac{r}{\theta^2(1-\theta)}} \propto \theta^{-1}(1-\theta)^{-1/2}$

This is the improper Beta$(0, 1/2)$.

	Prior	Prop Likelihood	Posterior
X	Beta$(1/2, 1/2)$	$\theta^x(1-\theta)^{n-x}$	Beta$(1/2+x, 1/2+n-x)$
Y	Beta$(0, 1/2)$	$\theta^r(1-\theta)^{y-r}$	Beta$(r, 1/2+y-r)$.

If $x = r = 3$ and $y = n = 12$ then:

$\theta \mid x \sim$ Beta$(7/2, 19/2)$ and $\theta \mid y \sim$ Beta$(3, 19/2)$.

Thus, the posteriors are DIFFERENT. Jeffreys prior will violate the SRP and SLP. The Jeffreys prior depends on the FORM of the data and so utilises values of the sample space which are not observed.

(iii) In a proper Bayesian analysis, the prior distribution for θ should be the same in BOTH models

e.g. $\theta \sim$ Beta(α, β) in both cases leads to

$\theta \mid x \sim$ Beta$(\alpha + \boxed{x}, \beta + \boxed{n-x})$

$\theta \mid y \sim$ Beta$(\alpha + \underbrace{\boxed{r}}_{\text{successes}}, \beta + \underbrace{\boxed{y-r}}_{\text{failures}})$

and so when $x = r$ and $y = n$ then $\theta \mid x \sim$ Beta$(\alpha + x, \beta + n - x)$

$\theta \mid y \sim$ Beta$(\alpha + x, \beta + n - x)$

Thus, inferences drawn on the posterior will be the same.