Notes from Problems Class 3.

Will discuss the idea of (Jeffreys) - Lindley's Paradox.

Consider a null hypothesis $H_0 : \theta = \theta_0$. Construct a mixed prior of the form

$$\pi(\theta) = p_0 \mathbb{I}_{(\theta_0)}(\theta) + (1-p_0)\pi_1(\theta)$$

where $p_0 = \mathbb{P}(\theta = \theta_0)$ and $\pi_1(\theta) = f_1(\theta \mid \theta \neq \theta_0)$

(prior density under H_1)

The posterior probability of $\theta = \theta_0$ is

$$\mathbb{P}(\theta = \theta_0 \mid x) = \frac{\mathbb{P}(\theta = \theta_0) f_X(x \mid \theta = \theta_0)}{f(x)}$$

$$= \frac{\mathbb{P}(\theta = \theta_0) f_X(x \mid \theta = \theta_0)}{\mathbb{P}(\theta = \theta_0) f_X(x \mid \theta = \theta_0) + \mathbb{P}(\theta \neq \theta_0) f_X(x \mid \theta \neq \theta_0)}$$

$$= \frac{1}{1 + \dfrac{\mathbb{P}(\theta \neq \theta_0)}{\mathbb{P}(\theta = \theta_0)} \dfrac{f_X(x \mid \theta \neq \theta_0)}{f_X(x \mid \theta = \theta_0)}}$$

$$= [1 + y]^{-1} \quad \text{where} \quad y = \frac{1-p_0}{p_0} \frac{f_X(x \mid \theta \neq \theta_0)}{f_X(x \mid \theta = \theta_0)}$$

Ratio of likelihoods.

$$\left[\text{Bayes factor: } \frac{\mathbb{P}(\theta = \theta_0 \mid x)}{\mathbb{P}(\theta \neq \theta_0 \mid x)} \Bigg/ \frac{\mathbb{P}(\theta = \theta_0)}{\mathbb{P}(\theta \neq \theta_0)} = \frac{f_X(x \mid \theta = \theta_0)}{f_X(x \mid \theta \neq \theta_0)} \right]$$

and $f_X(x \mid \theta \neq \theta_0) = \displaystyle\int_{\Theta_1} f_X(x \mid \theta, \theta \neq \theta_0) f_1(\theta \mid \theta \neq \theta_0)\, d\theta.$

Example.

Consider $H_0: \theta = 0$ and $H_1: \theta \neq 0$. Suppose that $X | \theta \sim N(\theta, \sigma^2)$ for σ^2 known and for $\pi_1(\theta)$ let $\theta \sim N(0, \sigma_0^2)$ where σ_0^2 is known.

$$f_X(x | \theta = 0) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{ -\frac{1}{2\sigma^2} x^2 \right\}$$

$$f_X(x | \theta \neq 0) = \int_{-\infty}^{\infty} \underbrace{f_X(x | \theta, \theta \neq 0)}_{N(\theta,\, \sigma^2)} \underbrace{f_1(\theta | \theta \neq \theta_0)}_{N(0,\, \sigma_0^2)} \, d\theta$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{ -\frac{1}{2\sigma^2} (x - \theta)^2 \right\} \frac{1}{\sqrt{2\pi}\,\sigma_0} \exp\left\{ -\frac{1}{2\sigma_0^2} \theta^2 \right\} d\theta$$

$$= \frac{(\sigma^2 + \sigma_0^2)^{-1/2}}{\sqrt{2\pi}} \exp\left\{ - \frac{x^2}{2(\sigma^2 + \sigma_0^2)} \right\}$$

i.e. $X | \theta \neq 0 \sim N(0, \sigma^2 + \sigma_0^2)$

[In essence we have $\theta \sim N(0, \sigma_0^2)$, $X | \theta \sim N(\theta, \sigma^2)$,

so $E(X) = E(E(X|\theta)) = E(\theta) = 0$

$Var(X) = E(Var(X|\theta)) + Var(E(X|\theta))$

$= E(\sigma^2) + Var(\theta)$

$= \sigma^2 + \sigma_0^2$]

Hence,
$$\frac{f_X(x | \theta \neq 0)}{f_X(x | \theta = 0)} = \left(\frac{\sigma^2}{\sigma^2 + \sigma_0^2} \right)^{1/2} \exp\left\{ -\frac{1}{2} \left(\frac{1}{\sigma_0^2 + \sigma^2} - \frac{1}{\sigma^2} \right) x^2 \right\}$$

$$= \left(\frac{\sigma^2}{\sigma^2 + \sigma_0^2} \right)^{1/2} \exp\left\{ + \frac{1}{2} \frac{\sigma_0^2 x^2}{(\sigma_0^2 + \sigma^2)\sigma^2} \right\}$$

Hence, $$y = \frac{1-p_0}{p_0}\left(\frac{\sigma^2}{\sigma^2+\sigma_0^2}\right)^{1/2} \exp\left\{\frac{\sigma_0^2 x^2}{2(\sigma_0^2+\sigma^2)\sigma^2}\right\}$$

In testing $H_0: \theta = \theta_0$ versus $H_1: \theta \neq \theta_0$. A dispersed prior of $H_1: \theta \neq \theta_0$ is sometimes proposed.

In this normal example, the prior information is contained in σ_0^2, the variance.

Consider $\sigma_0^2 \to \infty$ then $y \to 0$ and $\mathbb{P}(\theta = 0 \mid x) \to 1$ for all x

- What this really says is that $f_1(\theta \mid \theta = \theta_0) \propto 1$ doesn't work
- Can't just apply Bayes theorem when using an improper prior.

If I use a proper prior then σ_0^2 is FINITE. In this case, as $|x| \to \infty$ $y \to \infty$ and $\mathbb{P}(\theta \neq 0 \mid x) \to 1$.

$(\, \mathbb{P}(\theta = 0 \mid x) = [1+y]^{-1} \,)$

The Bayesian analysis behaves reasonably.

We now consider n observations so $x = (x_1, \dots, x_n)$, and that these are iid.

Perform analogous calculations but with $x \mapsto \bar{x}$, $\sigma^2 \mapsto \frac{\sigma^2}{n}$ gives:

$$y_n = \frac{1-p_0}{p_0}\left(\frac{\sigma^2}{\sigma^2+n\sigma_0^2}\right)^{1/2} \exp\left\{\frac{\sigma_0^2 n^2 \bar{x}^2}{2(n\sigma_0^2+\sigma^2)\sigma^2}\right\}$$

$$= \frac{1-p_0}{p_0}\left(\frac{\sigma^2}{\sigma^2+n\sigma_0^2}\right)^{1/2} \exp\left\{\frac{\sigma_0^2}{2(n\sigma_0^2+\sigma^2)}\left(\frac{n^2\bar{x}^2}{\sigma^2}\right)\right\}$$

$$= \frac{1-p_0}{p_0}\left(\frac{\sigma^2}{\sigma^2+n\sigma_0^2}\right)^{1/2} \exp\left\{\frac{n\sigma_0^2}{2(n\sigma_0^2+\sigma^2)}\left(\frac{\sqrt{n}\,\bar{x}}{\sigma}\right)^2\right\}$$

$z = \frac{\sqrt{n}\,\bar{x}}{\sigma}$

If $\bar{x} > 0$ then as $n \to \infty$, $y_n \to \infty$ and $\mathbb{P}(\theta = 0 \mid \bar{x}) = \frac{1}{1+y_n}$ $\to 0$. This is the expected outcome. With enough data and $\bar{x} > 0$ I can identify precisely whether $\theta = 0$.

$$p(\bar{x}; \theta_0) = \mathbb{P}(|\bar{X}| \geq \bar{x} \mid \theta_0)$$ ($\theta_0 = 0$ here)

Under H_0, $\frac{\sqrt{n}\bar{X}}{\sigma} \sim N(0,1)$ so if $z = \frac{|\bar{x}|}{\sigma/\sqrt{n}} > 1.96$ we would reject H_0 at the 5% significance level (for a two-sided test).

We now think about keeping z fixed as we increase our number of observations. Then we see that $y \to 0$ and hence $\mathbb{P}(\theta = 0 \mid x) \to 1$ as $n \to \infty$ for any z.

This is the main idea of the seeming paradox:

For a normal model $N(\theta, \sigma^2)$ with known variance σ^2
$H_0: \theta = \theta_0$ versus $H_1: \theta \neq \theta_0$

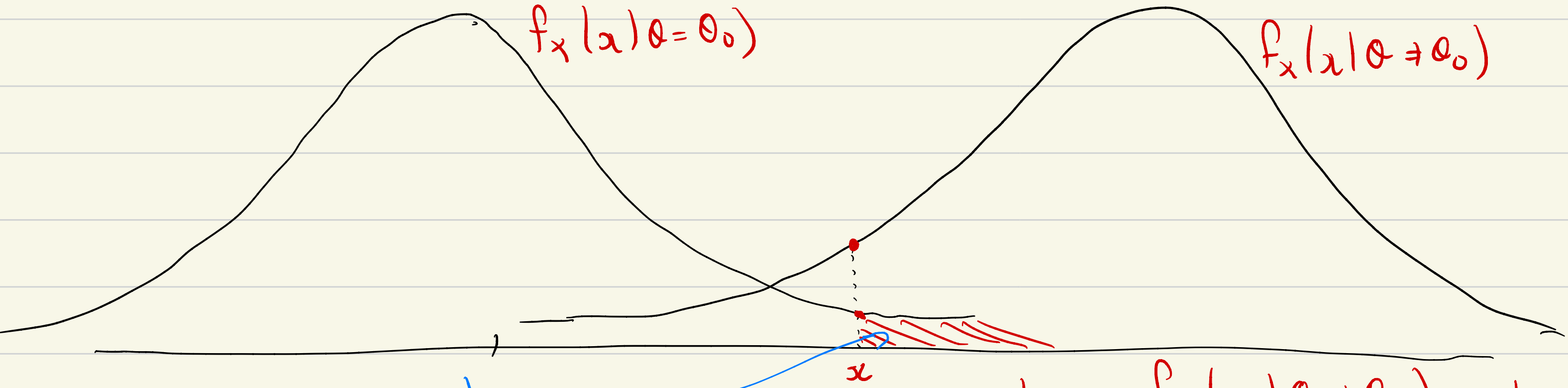
Assume $\mathbb{P}(\theta \in H_0) > 0$ and any regular prior on $\{\theta \neq \theta_0\}$ then for any $\alpha \in [0,1]$, can find a sample size $N(\alpha)$ and iid data x such that:

① The sample mean $\bar{x}$ is different from θ_0 at level α
(Classical model would reject H_0)

② $\mathbb{P}(\theta = \theta_0 \mid X) \geq 1 - \alpha$
(Bayesian model supports H_0 rather than H_1).

We get different conclusions depending upon the school we take.

Small n. $\quad H_0: \theta = 0 \qquad H_1: \theta = 1$

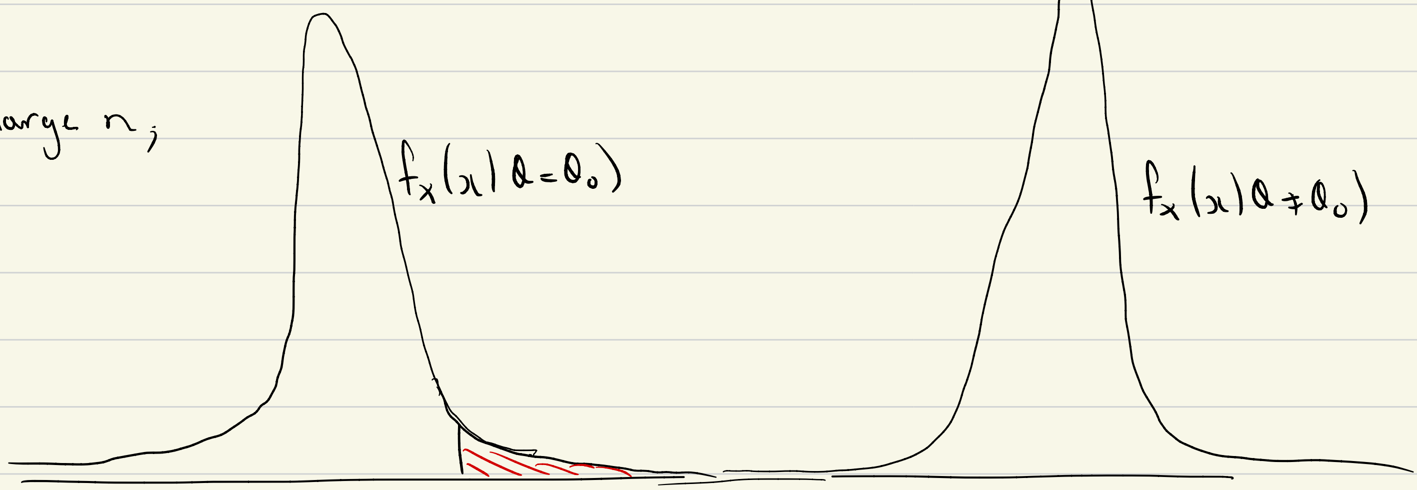


p-value.
This doesn't account for H_1.

$$y_n = \frac{1-p_0}{p_0}\,\frac{f_X(x|\theta\neq\theta_0)}{f_X(x|\theta=\theta_0)} \geq \frac{1-p_0}{p_0}$$

Calculation of $\mathbb{P}(\theta=\theta_0|x)$ involves the likelihood ratio of the other values of θ

For large n;



In this case $\quad y_n = \dfrac{1-p_0}{p_0}\,\dfrac{f_X(x|\theta\neq\theta_0)}{f_X(x|\theta=\theta_0)} \leq \dfrac{1-p_0}{p_0}$

The classical and the Bayesian approach ask different questions.

- If I reject Ho using a p-value, so a small p-value, then Ho is a poor explanation for the observation
- Bayesian approach: Ho is a better explanation for the data x than H_1.

What information do I need to compute the probability that say the null hypothesis is true when I have a p-value ?

Suppose $p(x; H_0) \leq u$, what can I say about my posterior probability ?

$$\mathbb{P}(H_0 \mid p(X; H_0) \leq u) = \frac{\mathbb{P}(p(X; H_0) \leq u \mid H_0)\, \mathbb{P}(H_0)}{\mathbb{P}(p(X; H_0) \leq u)}$$

$$\leq \frac{u\, \mathbb{P}(H_0)}{\mathbb{P}(p(X; H_0) \leq u)}$$ (by uniformity of the p-value)

NB, $\mathbb{P}(p(X; H_0) \leq u) = \mathbb{P}(p(X; H_0) \leq u \mid H_0)\, \mathbb{P}(H_0) + \mathbb{P}(p(X; H_0) \leq u \mid H_1)\, \mathbb{P}(H_1)$

I need to know the distribution of the p-value under the alternate hypothesis.