The Measurement of Statistical Evidence
Lecture 6 - part 1

Michael Evans
University of Toronto
http://www.utstat.utoronto.ca/mikevans/sta4522/STA4522.html

2021

Statistical Reasoning

- here is a sequence of steps to statistical reasoning concerning **E** and/or **H** about object of interest Ψ

1. choose a model $\{f_\theta : \theta \in \Theta\}$
2. choose a prior π (elicitation)
3. measure bias and select the amount of data to collect to avoid bias (design)
4. collect the data x
5. check the model against x (modify if necessary)
6. check the prior against x (modify if necessary)
7. derive the inferences (based on principles of inference to be discussed)

- we discuss 3, 2 and 6 today and based on the ingredients

$$(\{f_\theta : \theta \in \Theta\}, \pi, x)$$

Bias

- is it possible that the ingredients have been chosen such that the answer to **E** or **H** is a foregone conclusion with high prior probability?

- yes, but we can measure this bias and control it and in doing so we are led to a resolution of the conflict between Bayesian and frequentist statistics

- bias calculations are necessary as part of assessing the quality of a study

> *Would you accept the results of a statistical analysis that reported evidence against (in favor of)* $H_0 : \Psi(\theta) = \psi_0$ *if the prior probability of obtaining such evidence was* ≈ 1 *when* H_0 *was true (false)?*

- bias calculations only depend on the principle of evidence $\mathbf{R}_2$

Example - *location-normal (Jeffreys-Lindley paradox)*

- $\bar{x} \sim N(\mu, \sigma_0^2/n)$ and $\mu \sim N(\mu_0, \tau_0^2)$ then

$$RB(\mu \,|\, \bar{x}) = \left(1 + \frac{n\tau_0^2}{\sigma_0^2}\right)^{1/2} \times$$
$$\exp\left\{-\frac{1}{2}\left(1 + \frac{\sigma_0^2}{n\tau_0^2}\right)^{-1}\left(\frac{\sqrt{n}(\bar{x}-\mu)}{\sigma_0} + \frac{\sigma_0(\mu_0-\mu)}{\sqrt{n}\tau_0^2}\right)^2 + \frac{(\mu_0-\mu)^2}{2\tau_0^2}\right\}$$

and for $H_0 : \mu = \mu_0$ then for fixed $\sqrt{n}(\bar{x} - \mu_0)/\sigma_0$ we have $RB(\mu_0 \,|\, \bar{x}) \to \infty$ as $\tau_0^2 \to \infty$ even when μ_0 is false (also $RB = BF$ here when BF based on sharp prior)

- could have classical p-value $2(1 - \Phi(\sqrt{n}|\bar{x} - \mu_0|/\sigma_0)) \approx 0$ so a contradiction between frequentism and Bayes

- but the strength satisfies

$$\Pi(RB(\mu \,|\, \bar{x}) \leq RB(\mu_0 \,|\, \bar{x}) \,|\, \bar{x}) \to 2(1 - \Phi(\sqrt{n}|\bar{x} - \mu_0|/\sigma_0))$$

as $\tau_0^2 \to \infty$ so evidence in favor is very weak in this situation (partial resolution)

- general resolution of Jeffreys-Lindley: measure and control bias

H: bias for $H_0 : \Psi(\theta) = \psi_0$

bias against: $M(RB_\Psi(\psi_0 \,|\, X) \leq 1 \,|\, \psi_0) =$ *prior probability of not getting evidence in favor of* H_0 *when it is true.*
bias in favor: $\sup_{\psi : d(\psi, \psi_0) > \delta} M(RB_\Psi(\psi_0 \,|\, X) \geq 1 \,|\, \psi) =$ *maximum prior probability of not getting evidence against* H_0 *when it is meaningfully false.*

- in location-normal: need to calculate $M(RB(\mu_0 \,|\, \bar{X}) \leq 1 \,|\, \mu_0)$ and $M(RB(\mu_0 \,|\, \bar{X}) \geq 1 \,|\, \mu)$ and recall that the joint distribution of $(\mu, \bar{x})$ is

$$\bar{x} \,|\, \mu \sim N(\mu, \sigma_0^2/n), \mu \sim N(\mu_0, \tau_0^2)$$

and so both biases can be easily computed via simulation since

$$\sup_{\mu : d(\mu, \mu_0) > \delta} M(RB(\mu_0 \,|\, \bar{X}) \geq 1 \,|\, \mu) = \sup_{\mu = \mu_0 \pm \delta} M(RB(\mu_0 \,|\, \bar{X}) \geq 1 \,|\, \mu)$$

- **note** - bias against $\to 0$ and bias in favor $\to 1$ as $\tau_0^2 \to \infty$ so the real explanation for the strange behavior is that a diffuse prior is injecting bias in favor of H_0

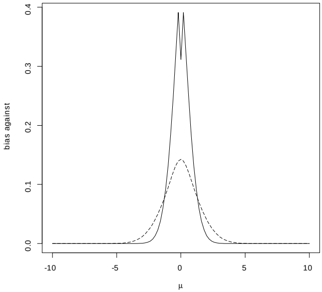

Figure: Plot of bias against $H_0 = \{\mu\}$ with a $N(0,1)$ prior (- - -) and a $N(0, 0.01)$ prior (—) with $n = 5, \sigma_0 = 1$.

- in general, both biases converge to 0 as the amount of data $n \to \infty$ and so bias can be controlled by design

> **Choose priors via elicitation, don't choose arbitrarily diffuse priors in an attempt to be "conservative", and design to avoid bias.**

E: bias for estimating Ψ

bias against *the prior prob. that true value is not in* $Pl_\Psi(x)$,
$E_{\Pi_\Psi}(M(\psi \notin Pl_\Psi(X)\,|\,\psi)) = E_{\Pi_\Psi}(M(RB_\Psi(\psi\,|\,X) \leq 1\,|\,\psi))$

- so $1 - E_{\Pi_\Psi}(M(\psi \notin Pl_\Psi(X)\,|\,\psi))$ is the prior coverage probability (confidence) of $Pl_\Psi(x)$

- typically there exist a $\psi_0 = \arg\sup M(RB_\Psi(\psi\,|\,X) \leq 1\,|\,\psi)$

- then $M(\psi \in Pl_\Psi(X)\,|\,\psi) \geq 1 - M(RB_\Psi(\psi_0\,|\,X) \leq 1\,|\,\psi_0)$ and a "pure" frequentist confidence when $\Psi(\theta) = \theta$ otherwise like a random effects model where random effects are the nuisance parameters

bias in favor: *the prior prob. that a meaningfully false value is not in the implausible region* $\mathrm{Im}_\Psi(x) = \{\psi : RB_\Psi(\psi_0\,|\,x) < 1\}$,
$E_{\Pi_\Psi}\left(\sup_{\psi:d_\Psi(\psi,\psi_0)\geq\delta} M(\psi_0 \notin Im_\Psi(X)\,|\,\psi)\right) =$
$E_{\Pi_\Psi}\left(\sup_{\psi:d_\Psi(\psi,\psi_0)\geq\delta} M(RB_\Psi(\psi_0\,|\,X) \geq 1\,|\,\psi)\right)$

- like the probability that confidence region covers a false value

- both biases converge to 0 with increasing amounts of data

Example - *location-normal*

$$PI(x) = \bar{x} \pm w(\bar{x}, n, \sigma_0^2, \mu_0, \tau_0^2) \text{ where}$$

$$w(\bar{x}, n, \sigma_0^2, \mu_0, \tau_0^2)$$

$$= \frac{\sigma_0}{\sqrt{n}} \left(1 + \frac{n\tau_0^2}{\sigma_0^2}\right)^{-\frac{1}{2}} \left\{ \left(1 + \frac{n\tau_0^2}{\sigma_0^2}\right) \log\left(1 + \frac{n\tau_0^2}{\sigma_0^2}\right) + \left(\frac{\bar{x} - \mu_0}{\sigma_0/\sqrt{n}}\right)^2 \right\}^{\frac{1}{2}}$$

n	$\tau_0 = 1$	$\tau_0 = 0.5$
5	0.625 (0.893)	0.491 (0.807)
10	0.499 (0.925)	0.389 (0.854)
20	0.393 (0.949)	0.312 (0.893)
50	0.281 (0.969)	0.231 (0.933)
100	0.215 (0.979)	0.181 (0.954)

Table: Expected half-widths (coverages) of the plausible interval when using a $N(\mu_0, \tau_0^2)$ prior for different sample sizes n and $\sigma_0^2 = 1$.

n	$(\mu_0, \tau_0) = (0, 1), \delta = 1.0$	$(\mu_0, \tau_0) = (0, 1), \delta = 0.5$
5	0.451	0.798
10	0.185	0.690
20	0.025	0.486
50	0.000	0.131
100	0.000	0.009

Table: Average bias in favor for estimation when using a $N(0, \tau_0^2)$ prior for different sample sizes n and difference δ.

Inferences are Bayesian and based on the evidence in the observed data while assessment of the reliability (or quality) of the inferences is frequentist and considers the possible data values that could occur a priori.

Example *Fieller's problem*

- mss $\bar{x} \sim N(\mu, \sigma_0^2/n)$ ind. of $\bar{y} \sim N(\nu, \sigma_0^2/m)$ and $\psi = \Psi(\mu, \nu) = \mu/\nu$
- $\mu \sim N(\mu_0, \tau_{10}^2)$ ind. of $\nu \sim N(\nu_0, \tau_{20}^2)$ and want to assess $H_0 : \psi = 2$
- need to choose a relevant δ (let's take $\delta = 0.2$ and use absolute error)
- let $m = n, \sigma_0^2 = 1$ and suppose $\mu_{true} = 2\nu_{true}$ where $\nu_{true} = 10$ and generate the samples x and y
- for bias against need to condition on $H_0 = \{(\mu, \nu) : \mu = \psi\nu\}$
- make the change of variable $(\mu, \nu) \to (\psi, \nu)$ then joint density of $(\psi, \nu, \bar{x}, \bar{y})$ is proportional to

$$|\nu| \exp\left\{-\frac{1}{2}\left[\frac{n(\bar{x} - \psi\nu)^2 + m(\bar{y} - \nu)^2}{\sigma_0^2} + \frac{(\psi\nu - \mu_0)^2}{\tau_{10}^2} + \frac{(\nu - \nu_0)^2}{\tau_{20}^2}\right]\right\}$$

- so to generate $(\bar{x}, \bar{y})$ from $M(\cdot \mid \psi)$ generate

$$\begin{aligned} \bar{x} \mid (\psi, \nu, \bar{y}) &\sim N(\psi\nu, \sigma_0^2/n) \\ \bar{y} \mid (\psi, \nu) &\sim N(\nu, \sigma_0^2/m) \\ \nu \mid \psi &\sim \pi(\cdot \mid \psi) \end{aligned}$$

where

$$\begin{aligned} \pi(\nu \mid \psi) &\propto |\nu| \exp\left\{-\frac{1}{2}\left[\frac{(\psi\nu - \mu_0)^2}{\tau_{10}^2} + \frac{(\nu - \nu_0)^2}{\tau_{20}^2}\right]\right\} \\ &\propto |\nu| \exp\left\{-\frac{1}{2}\frac{(\nu - \nu(\psi))^2}{\tau_{20}^2(\psi)}\right\} \text{ where} \\ \tau_{20}^2(\psi) &= \left(\frac{\psi^2}{\tau_{10}^2} + \frac{1}{\tau_{20}^2}\right)^{-1} \text{ and } \nu(\psi) = \tau_{20}^2(\psi)\left(\frac{\psi\mu_0}{\tau_{10}^2} + \frac{\nu_0}{\tau_{20}^2}\right) \end{aligned}$$

- transforming $\nu \to z = (\nu - \nu(\psi))/\tau_{20}(\psi)$ we need to generate z from a density $g(z) \propto |a + bz|\varphi(z)$ where $a = \nu(\psi)$, $b = \tau_{20}(\psi)$

- now with $b > 0$ then $a + bz \geq 0$ iff $z \geq -a/b$ so density is

$$
\begin{aligned}
g(z) &= p(a,b)I_{(-\infty,-a/b]}(z)g_1(z) + (1 - p(a,b))I_{(-a/b,\infty)}(z)g_0(z) \\
g_1(z) &= \frac{-(a+bz)\varphi(z)}{-a\Phi(-a/b)) + b\varphi(-a/b)} \text{ when } a + bz \leq 0 \\
g_0(z) &= \frac{(a+bz)\varphi(z)}{a(1-\Phi(-a/b)) + b\varphi(-a/b)} \text{ when } a + bz \geq 0 \\
p(a,b) &= \frac{-a\Phi(-a/b)) + b\varphi(-a/b)}{a(1-2\Phi(-a/b)) + 2b\varphi(-a/b)}
\end{aligned}
$$

so generate z from g_1 with prob. $p(a,b)$ and otherwise generate from g_0

- generate from g_1 via inversion where for $z \leq -a/b$

$$G_1(z) = \int_{-\infty}^{z} g_1(x)\,dx = \frac{-a\Phi(z) + bz\varphi(z)}{-a\Phi(-a/b)) + b\varphi(-a/b)}$$

so gen. $u \sim U(0,1)$ and solve $G_1(z) = u$ for z by bisection and similarly for g_0

- also need to calculate $RB_{\Psi}(\psi_0 \,|\, \bar{x}, \bar{y})$ for each generated $(\bar{x}, \bar{y})$ and then compare it with 1

- for this you should have the prior content of $(\psi_0 - \delta/2, \psi_0 + \delta/2)$ from the inference computations

- for the posterior contents of this interval you need to compute this in another loop for each generated $(\bar{x}, \bar{y})$ and approx.

$$RB_{\Psi}(\psi_0 \,|\, \bar{x}, \bar{y}) \approx \frac{\Pi((\psi_0 - \delta/2, \psi_0 + \delta/2) \,|\, \bar{x}, \bar{y})}{\Pi((\psi_0 - \delta/2, \psi_0 + \delta/2))}$$

- really only need one decimal place accuracy for the bias computations