The Measurement of Statistical Evidence
Lecture 6 - part 2

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

2021

1. Choosing the model

- unfortunately not too much to say beyond the obvious fact that the model has to be able to capture the real-world meaning of the object of interest Ψ

- as part of this determine δ, the accuracy that matters

- it is fair to say, however, that after the data x the model is the most important part of the whole program of statistical reasoning

- also need discussion of how to modify a model when it fails

2. Choosing the prior

- the "right" way to choose a prior is via elicitation

- this means you use what you know about the true value

- **note** - there isn't only one right way to carry out an elicitation

- also presumably you collect enough data so that the prior doesn't dominate the inferences (bias calculations)

Example *location normal*

- $x = (x_1, \ldots, x_n) \overset{i.i.d.}{\sim} N(\mu, \sigma_0^2)$ with $\mu \in R^1, \sigma_0^2$ known and π a $N(\mu_0, \tau_0^2)$ dist.

- specify interval (m_1, m_2) that contains the true μ with virtual certainty γ (e.g. $\gamma = 0.99$)

- put $\mu_0 = (m_1 + m_2)/2$ and then solve
$\Phi((m_2 - \mu_0)/\tau_0) - \Phi((m_1 - \mu_0)/\tau_0) = \gamma$ for τ_0

- if you take (m_1, m_2) too short then risk prior-data conflict and bias against and if you take it too long then you will need a large sample size to avoid bias in favor

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 $H_0 : \Psi(\mu, \nu) = \psi_0$
- you could apply the previous elicitation algorithm to each of μ and ν but presumably something is known about ψ (else why make inference about it)
- so perhaps use the previous algorithm to obtain (μ_0, τ_{10}^2), via interval (m_1, m_2), and specify interval (r_1, r_2) that contains the true value of ψ with virtual certainty (also contains ψ_0 say $\psi_0 = (r_1 + r_2)/2$
- then, provided r_1, r_2 are of the same sign (say positive) $r_1 \leq \mu/\nu \leq r_2$ iff $\mu/r_2 \leq \nu \leq \mu/r_1$ so $m_1/r_2 \leq \nu \leq m_2/r_1$ with virtual certainty and determine (ν_0, τ_{20}^2) with $\nu_0 = \mu_0/\psi_0$ and τ_{20} satisfying $\Phi((m_2/r_1 - \nu_0)/\tau_{20}) - \Phi((m_1/r_2 - \nu_0)/\tau_{20}) = \gamma$

Improper Priors

- sometimes individuals claim complete ignorance about a quantity that takes values in an infinite region and so a prior π is selected which supposedly represents this ignorance

- such priors are often chosen by a default rule and they are improper

- e.g., Jeffreys prior $\pi(\theta) \propto \left|\det(E_\theta(\frac{\partial^2 \log f_\theta(x)}{\partial\theta_i \partial\theta_j})\right|^{1/2}$

- when the prior is improper, then $\pi(\theta) f_\theta(x)$ does not correspond to a joint probability distribution for (θ, x) even when $\int_\Theta \pi(\theta) f_\theta(x)\, d\theta$ is finite, yet when it is finite, $\pi(\theta \,|\, x) = \pi(\theta) f_\theta(x) / \int_\Theta \pi(\theta) f_\theta(x)\, d\theta$ is called the posterior of θ

- but in the improper prior case this is not an application of $\mathbf{R}_1$ the conditionality principle, so what "principle" is being applied?

- even when Jeffreys prior is finite the prior is questionable as a representative of ignorance, e.g., Bernoulli(θ) then Jeffreys prior is $\theta \sim$ beta$(1/2, 1/2)$ with infinite singularities at 0 and 1 and it seems unlikely that this represents "ignorance"

- empirical Bayes also violates $\mathbf{R}_1$ since the prior depends on x

3. Measuring bias, 4. Data collection and 5. Model checking

- as already discussed measure biases and use these step to decide on the data collection to obtain x

- there are many methods for model checking based on x but those based on the conditional distributions given a mss $T(x)$ or based on an ancillary statistic $U(x)$ seem the most principled and these can be based on a p-value as there are no alternatives

6. Checking for prior-data conflict

Evans and Moshonov (2006) Checking for prior-data conflict. Bayesian Analysis, 1, 4, 893-914.

- for mss $T(x)$ and ancillary compute

$$M_T(m_T(t \mid U(x)) \leq m_T(T(x) \mid U(x)) \mid U(x))$$

and this serves to locate $T(x)$ in its conditional prior distribution so if this prob. is small there is an indication of a prior-data conflict

- recall that the distribution of the data for a given value of $T(x)$ does not involve θ so this can tell us nothing about whether the prior is contradicted by the data and similarly conditioning on $U(x)$ removes the variation due to U when making this assessment

Evans and Jang (2011) A limit result for the prior predictive applied to checking for prior-data conflict. Statistics and Probability Letters, 81, 1034-1038.

$$M_T(m_T(t \mid U(x)) \leq m_T(T(x) \mid U(x)) \mid U(x)) \to \Pi(\pi(\theta) \leq \pi(\theta_{true}))$$

- when there is prior-data conflict there is a lack or robustness to the prior

Al-Labadi and Evans (2017) Optimal robustness results for some Bayesian procedures and the relationship to prior-data conflict. Bayesian Analysis 12, 3, 702-728.

- what to do when there is prior-data conflict?

Evans and Jang (2011). Weak informativity and the information in one prior relative to another. Statistical Science, 26, 3, 423-439.

Example - *location normal*
- $T(x) = \bar{x} \,|\, \mu \sim N(\mu, \sigma_0^2/n)$ and $\mu \sim N(\mu_0, \tau_0^2)$ so $\bar{x} \sim N(\mu_0, \tau_0^2 + \sigma_0^2/n)$ and note that since $\bar{x}$ is a complete mss, Basu's theorem says it is independent of any ancillary statistic so no need for conditioning

$$\begin{aligned} m_T(t) &= \left(\tau_0^2 + \sigma_0^2/n\right)^{-1/2} \varphi\left(\left(\tau_0^2 + \sigma_0^2/n\right)^{-1/2}(t - \mu_0)\right) \\ M_T(m_T(t) \le m_T(\bar{x})) &= M_T((t - \mu_0)^2 \ge (\bar{x} - \mu_0)^2) \\ &= 2[1 - \Phi(|\bar{x} - \mu_0| / \left(\tau_0^2 + \sigma_0^2/n\right)^{1/2})] \to 2[1 - \Phi(|\mu_{true} - \mu_0| / \tau_0)] \end{aligned}$$

7. Inference

- relative belief inferences to answer **E** and/or **H** about Ψ

Summary

- the central core concept in statistics is the idea that data contains evidence concerning answers to **E** and **H**

- **thesis**: to build a sound theory of statistical reasoning it is necessary to give a clear characterization of statistical evidence and how to quantify it

- the prominent, commonly used approaches to statistics fail in this regard

- the approach via relative belief:

1. answers and resolves a variety of paradoxes and doesn't (seem to) introduce new ones,
2. is relatively simple,
3. is a whole theory of statistical reasoning where the individual parts are all inter-related and agrees with basic scientific principles like falsifiability to support objectivity,
4. unifies aspects of Bayesian and frequentist thinking as each plays a key role.

Is it correct?