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 $\Psi$
- as part of this determine $\delta$, 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

- $x = (x_1, \ldots, x_n) \sim \mathcal{N}(\mu, \sigma_0^2)$ with $\mu \in R^1, \sigma_0^2$ known and $\pi$ a $\mathcal{N}(\mu_0, \tau_0^2)$ dist.
- specify interval $(m_1, m_2)$ that contains the true $\mu$ with virtual certainty $\gamma$ (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 $\tau_0$
- 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 $\bar{x} \sim N(\mu, \sigma^2_0/n)$ ind. of $\bar{y} \sim N(\nu, \sigma^2_0/m)$ and $\psi = \Psi(\mu, \nu) = \mu/\nu$
- $\mu \sim N(\mu_0, \tau^2_{10})$ ind. of $\nu \sim N(\nu_0, \tau^2_{20})$ and want to assess $H_0 : \Psi(\mu, \nu) = \psi_0$
- you could apply the previous elicitation algorithm to each of $\mu$ and $\nu$ but presumably something is known about $\psi$ (else why make inference about it)
- so perhaps use the previous algorithm to obtain $(\mu_0, \tau^2_{10})$, via interval $(m_1, m_2)$, and specify interval $(r_1, r_2)$ that contains the true value of $\psi$ with virtual certainty (also contains $\psi_0$ 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 $(\nu_0, \tau^2_{20})$ with $\nu_0 = \mu_0/\psi_0$ and $\tau_{20}$ 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 $\pi$ is selected which supposedly represents this ignorance

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

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

- when the prior is improper, then $\pi(\theta)f_\theta(x)$ does not correspond to a joint probability distribution for $(\theta, x)$ 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 $\theta$

- but in the improper prior case this is not an application of $R_1$ 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($\theta$) then Jeffreys prior is $\theta \sim \text{beta}(1/2, 1/2)$ with infinite singularities at 0 and 1 and it seems unlikely that this represents "ignorance"

- empirical Bayes also violates $R_1$ since the prior depends on $x$

- 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


- 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 \( \theta \) 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


\[
M_T(m_T(t \mid U(x)) \leq m_T(T(x) \mid U(x)) \mid U(x)) \rightarrow \Pi(\pi(\theta) \leq \pi(\theta_{true}))
\]
- when there is prior-data conflict there is a lack or robustness to the prior


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


**Example** - *location normal*

- \( T(x) = \bar{x} \mid \mu \sim N(\mu, \sigma^2_0/n) \) and \( \mu \sim N(\mu_0, \tau^2_0) \) so

\[
\bar{x} \sim N(\mu_0, \tau^2_0 + \sigma^2_0/n)
\]

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

\[
m_T(t) = (\tau^2_0 + \sigma^2_0/n)^{-1/2} \varphi((\tau^2_0 + \sigma^2_0/n)^{-1/2} (t - \mu_0))
\]

\[
M_T(m_T(t) \leq m_T(\bar{x})) = M_T((t - \mu_0)^2 \geq (\bar{x} - \mu_0)^2)
\]

\[
= 2[1 - \Phi(|\bar{x} - \mu_0| / (\tau^2_0 + \sigma^2_0/n)^{1/2})] \rightarrow 2[1 - \Phi(|\mu_{true} - \mu_0| / \tau_0)]
\]
7. Inference

- relative belief inferences to answer $E$ and/or $H$ about $\Psi$
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?**