The Measurement of Statistical Evidence
Lecture 5 - part 1

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

2021

- recap
- probability model $(\Omega, \mathcal{F}, P)$
- principles of inference concerning unknown value of $\omega \in \Omega$

  1. **Principle of Conditional Probability***: initial belief that $\omega \in A \in \mathcal{F}$ is measured by $P(A)$ and after observing that $\omega \in C$ (via a known information generator), where $P(C) > 0$, then belief that $\omega \in A$ is measured by $P(A \mid C) = P(A \cap C)/P(C)$.*
  2. **Principle of Evidence***: if $P(A \mid C) > P(A)$, then the observation that C is true is evidence in favor of A being true, if $P(A \mid C) < P(A)$, then the observation that C is true is evidence against A being true, and $P(A \mid C) = P(A)$ is neither evidence in favor nor evidence against A being true.*
  3. **Principle of the Relative Belief Ratio***: when a numerical measure of evidence is required this is given by the relative belief ratio $RB(A \mid C) = P(A \mid C)/P(A)$ (provided $P(A) > 0$).*

- valid measures of evidence satisfy the principle of evidence

Properties of RB

1. (Savage-Dickey) $RB(A\,|\,C) = \frac{P(A\cap C)}{P(A)P(C)} = RB(C\,|\,A)$
2. $RB(A^c\,|\,C) = \frac{1-P(A)RB(A\,|\,C)}{1-P(A)} \begin{array}{ll} <1 & \text{if } RB(A\,|\,C)>1 \\ >1 & \text{if } RB(A\,|\,C)<1 \\ =1 & \text{if } RB(A\,|\,C)=1 \end{array}$
3. $0 \le RB(A\,|\,C) \le \frac{1}{P(A)}$, lower bound attained when $P(A\,|\,C)=0$ (e.g. $A\cap C=\phi$), upper bound attained when $P(A\,|\,C)=1$ (e.g. $C\subset A$) and so no universal scale
4. $RB(A\cap B\,|\,C) = \frac{RB(A\,|\,B\cap C)RB(B\,|\,C)}{RB(A\,|\,B)} =$
$$\begin{cases} \frac{RB(A\,|\,C)RB(B\,|\,C)}{RB(A\,|\,B)} & \text{when } A, B \text{ cond. ind. given } C \\ RB(A\,|\,C)RB(B\,|\,C) & \text{when } A, B \text{ cond. ind. given } C \text{ and ind.} \end{cases}$$
5. if $\Omega = \cup_{i=1}^k A_i, A_i\cap A_j=\phi$ when $i\ne j, P(A_i)>0$ for all i,
$$1 = \frac{P(\Omega\,|\,C)}{P(\Omega)} = RB(\cup_{i=1}^k A_i\,|\,C) = \sum_{i=1}^k RB(A_i\,|\,C)P(A_i)$$
Proof: $1 = RB(\cup_{i=1}^k A_i\,|\,C) = \frac{P(\cup_{i=1}^k A_i\,|\,C)}{P(\cup_{i=1}^k A_i)} = \sum_{i=1}^k RB(A_i\,|\,C)P(A_i)$

Can it happen that when $A \subset B$ then $RB(A \mid C) > 1$ but $RB(B \mid C) < 1$?

$$\begin{aligned} 1 &= RB(A \mid C)P(A) + RB(A^c \cap B \mid C)P(A^c \cap B) + \\ &\quad RB(B^c \mid C)P(B^c) \\ &= RB(B \mid C)P(B) + RB(B^c \mid C)P(B^c) \text{ so} \\ RB(B \mid C) &= RB(A \mid C)\frac{P(A)}{P(B)} + RB(A^c \cap B \mid C)\frac{P(A^c \cap B)}{P(B)} \\ &= RB(A \mid C)P(A \mid B) + RB(A^c \cap B \mid C)P(A^c \cap B \mid B) \; < 1 \text{ iff} \\ 0 &\leq P(A \mid B) < \frac{1 - RB(A^c \cap B \mid C)P(A^c \cap B \mid B)}{RB(A \mid C)} \end{aligned}$$

Example *An important lesson about measuring evidence.*

- a murder is committed and it is known that an adult member of a town with m adult citizens committed the crime and assume uniform beliefs before evidence is obtained

- evidence obtained $C =$ "a person of ethnic origin a committed this crime" and there are $m_1 < m$ adult members of this ethnic group in the town so $P(C) = \frac{m_1}{m}$

- suppose the town contains a university with n adult students of which n_1 are of ethnic origin a

$P(\text{"university student committed the crime"}) = P(B) = \frac{n}{m}$ and so

$$\begin{aligned} P(B \mid C) &= \frac{P(B \cap C)}{P(C)} = \frac{n_1 / m}{m_1 / m} = \frac{n_1}{m_1} \\ RB(B \mid C) &= \frac{P(B \mid C)}{P(B)} = \frac{n_1 / m_1}{n / m} = \frac{n_1 / n}{m_1 / m} < 1 \text{ when } \frac{n_1}{n} < \frac{m_1}{m} \end{aligned}$$

$P(\text{"university student of ethnic origin } a \text{ committed the crime"})$

$= P(A) = \frac{n_1}{m}$ and

$$\begin{aligned} P(A \mid C) &= \frac{P(A \cap C)}{P(C)} = \frac{P(A)}{P(C)} = \frac{n_1 / m}{m_1 / m} = \frac{n_1}{m_1} \\ RB(A \mid C) &= \frac{P(A \mid C)}{P(A)} = \frac{n_1 / m_1}{n_1 / m} = \frac{m}{m_1} > 1 \end{aligned}$$

- so $A \subset B$ and C is evidence if favor of A being true but, if relatively few students of ethnic origin a, then evidence against B

- in the statistical context with $(\{f_\theta : \theta \in \Omega\}, \pi, x)$ and interest in $\psi = \Psi(\theta)$

- let A_ϵ be nbds of ψ s.t. $A_\epsilon \overset{nicely}{\longrightarrow} \{\psi\}$ as $\epsilon \to 0$, then

$$RB_\Psi(\psi \,|\, x) \overset{def.}{=} \lim_{\epsilon \to 0} \frac{\Pi_\Psi(A_\epsilon \,|\, x)}{\Pi_\Psi(A_\epsilon)} \overset{conditions}{=} \frac{\pi_\Psi(\psi \,|\, x)}{\pi_\Psi(\psi)}$$

whenever π_Ψ is positive and continuous at ψ

- $RB(\psi \,|\, x) > 1$ says there is evidence in favor of ψ, $RB(\psi \,|\, x) < 1$ says there is evidence against ψ and $RB(\psi \,|\, x) = 1$ says there is no evidence either way

- the values $RB(\psi \,|\, x)$ order the possible values of $\psi \in \Psi$ but only when the values in Ψ are similar in nature (so they can be compared)

- this ordering is the same for any 1-1 transformation of $RB(\psi \,|\, x)$ such as $\log RB(\psi \,|\, x) = \log \pi_\Psi(\psi \,|\, x) - \log \pi_\Psi(\psi)$

- also if $\xi = \Xi(\psi)$ is a "smooth" bijection, then

$$RB_{\Xi}(\xi \,|\, x) = \frac{\pi_{\Xi}(\xi \,|\, x)}{\pi_{\Xi}(\xi)} = \frac{\pi_{\Psi}(\Xi^{-1}(\xi) \,|\, x) J_{\Xi}(\Xi^{-1}(\xi))}{\pi_{\Psi}(\Xi^{-1}(\xi)) J_{\Xi}(\Xi^{-1}(\xi))} = RB_{\Psi}(\Xi^{-1}(\xi) \,|\, x)$$

so inferences based on the relative belief ratio are invariant under repars

E : the estimate of $\psi = \Psi(\theta)$ is given by
$\psi(x) = \sup\{RB_{\Psi}(\psi \,|\, x) : \psi \in \Psi\}$ (as this maximizes the evidence in favor) and the accuracy of this estimate is assessed via the size and posterior content of the *plausible region (positive evidence region)*

$$Pl_{\Psi}(x) = \{\psi : RB_{\Psi}(\psi \,|\, x) > 1\}$$

and call $\psi(x)$ the *relative belief estimate*

- so if $Pl(x)$ is "small" and $\Pi_{\Psi}(Pl_{\Psi}(x) \,|\, x)$ is high, then we have an accurate estimate of ψ

- note: size is not invariant but recall our specification of the difference that matters δ as this is not invariant either

- invariance allows doing calculations in a convenient parameterization and then transforming back

- note - when $\Psi(\theta) = \theta$, then $RB(\theta \,|\, x) = \frac{\pi(\theta \,|\, x)}{\pi(\theta)} = \frac{\pi(\theta) f_\theta(x)}{m(x)\pi(\theta)} = \frac{f_\theta(x)}{m(x)}$ and so $\theta(x)$ is the MLE and $Pl_\Psi(x) = \{\theta : \frac{f_\theta(x)}{m(x)} > 1\}$ is a likelihood region but note you can't multiply $RB(\theta \,|\, x)$ by a constant and retain its interpretation in terms of evidence

- also, there is the *Savage-Dickey ratio* result

$$
\begin{aligned}
RB_\Psi(\psi \,|\, x) &= \frac{\pi_\Psi(\psi \,|\, x)}{\pi_\Psi(\psi)} = \frac{1}{\pi_\Psi(\psi)} \int_{\Psi^{-1}\{\psi\}} \pi(\theta \,|\, x) J_\Psi(\theta)\, d\theta \\
&= \frac{1}{\pi_\Psi(\psi)} \int_{\Psi^{-1}\{\psi\}} \frac{f_\theta(x)\pi(\theta)}{m(x)} J_\Psi(\theta)\, d\theta \\
&= \frac{1}{m(x)} \int_{\Psi^{-1}\{\psi\}} f_\theta(x) \frac{\pi(\theta) J_\Psi(\theta)}{\pi_\Psi(\psi)}\, d\theta \\
&= \frac{1}{m(x)} \int_{\Psi^{-1}\{\psi\}} f_\theta(x) \pi(\theta \,|\, \psi)\, d\theta = \frac{m(x \,|\, \psi)}{m(x)}
\end{aligned}
$$

- note - $m(x \,|\, \psi) = \int_{\Psi^{-1}\{\psi\}} f_\theta(x)\pi(\theta \,|\, \psi)\, d\theta$ is an integrated likelihood and so generally $\psi(x)$ is an MLE (unlike profiling)

- note - $\psi(x)$ depends on using $RB_\Psi(\cdot\,|\,x)$ to order the possible values but $Pl_\Psi(x)$ is the same no matter what valid measure of evidence is used which suggests that other estimates based on a valid measure of evidence could be used, as they all have the same accuracy as provided by $Pl_\Psi(x)$

- also, instead of quoting $Pl_\Psi(x)$ for assessing accuracy, a γ*-relative belief region*

$$C_{\Psi,\gamma}(x) = \{\psi : H_\Psi(RB_\Psi(\psi\,|\,x)\,|\,x) \geq 1-\gamma\},$$

where $H_\Psi(\cdot\,|\,x)$ is the posterior cdf of $RB_\Psi(\cdot\,|\,x)$ so $\Pi_\Psi(C_{\Psi,\gamma}(x)\,|\,x) \geq \gamma$, can be quoted **but** it is necessary that $\gamma \leq \Pi_\Psi(Pl_\Psi(x)\,|\,x)$ otherwise $C_{\Psi,\gamma}(x)$ will contain values of ψ for which there is evidence against

- so a relevant γ can only be determined after seeing the data

H : to assess $H_0 : \Psi(\theta) = \psi_0$ quote $RB_\Psi(\psi_0 \,|\, x)$ to determine if there is evidence in favor (> 1) or against (< 1)

- to measure the *strength* of the evidence quote

$$\Pi_\Psi(RB_\Psi(\psi_0 \,|\, x) \leq RB_\Psi(\psi_0 \,|\, x) \,|\, x)$$

which gives posterior belief that the true value has evidence no greater than that for the hypothesized value

- so when $RB_\Psi(\psi_0 \,|\, x) > 1$ and strength ≈ 1, then there is strong evidence in favor of H_0 and if $RB_\Psi(\psi_0 \,|\, x) < 1$ and strength ≈ 0, then there is strong evidence against H_0

- note - whether or not there is evidence in favor or against is independent of the valid measure of evidence used but the strength is dependent on this

- alternatively, to measure strength of the evidence you could quote $Pl_\Psi(x)$ and $\Pi_\Psi(Pl_\Psi(x) \,|\, x)$ when $RB_\Psi(\psi_0 \,|\, x) > 1$ since $\psi_0 \in Pl_\Psi(x)$ and it is now the natural "estimate" of ψ or when $RB_\Psi(\psi_0 \,|\, x) < 1$, quote the *implausible region* $Im_\Psi(x) = \{\psi : RB_\Psi(\psi \,|\, x) < 1\}$ and $\Pi_\Psi(Im_\Psi(x) \,|\, x)$

Bayes Factors

- often misdefined

- for probability model $(\Omega, \mathcal{F}, P)$ and $A \in \mathcal{F}$ with $0 < P(A) < 1$ and having observed that $C \in \mathcal{F}$ is true with $P(XC) > 0$, the *Bayes factor in favor of* A is

$$BF(A \mid C) = \frac{P(A \mid C)}{P(A^c \mid C)} \Big/ \frac{P(A)}{P(A^c)} = \frac{\text{posterior odds in favor of } A}{\text{prior odds in favor of } A}$$

Lemma $BF(A \mid C) > (<, =)\ 1$ iff $P(A \mid C) > (<, =)\ P(A)$.
Proof:

$$1 < BF(A \mid C) \text{ iff } \frac{P(A)}{1 - P(A)} < \frac{P(A \mid C)}{1 - P(A \mid C)} \text{ iff } \frac{1}{P(A)} > \frac{1}{P(A \mid C)}$$

- so the BF satisfies the principle of evidence

- note - $BF(A \mid C) = \frac{P(A \mid C)}{P(A)} \Big/ \frac{P(A^c \mid C)}{P(A^c)} = \frac{RB(A \mid C)}{RB(A^c \mid C)}$ but $RB(A \mid C) > 1$ iff $RB(A^c \mid C) < 1$ so it is not a comparison of the evidence for A with the evidence for A^c and you can't express the BF in terms of the RB

- why do we want to compare odds as opposed to probabilities anyway?

- the continuous (Bayesian) case: when $\Pi_\Psi(\{\psi_0\}) = 0$ the *BF* is not defined

- for the *RB* we defined this in terms of a limit of nbds A_ε converging to ψ_0 so $RB_\Psi(\psi_0 \,|\, x) = \lim_{\varepsilon \to 0} \Pi_\Psi(A_\varepsilon \,|\, x)/\Pi_\Psi(A_\varepsilon)$

- the natural thing to do then with the *BF* is to define it as

$$\begin{aligned} BF_\Psi(\psi_0 \,|\, x) &= \lim_{\varepsilon \to 0} \frac{\Pi_\Psi(A_\varepsilon \,|\, x)}{\Pi_\Psi(A_\varepsilon^c \,|\, x)} \Big/ \frac{\Pi_\Psi(A_\varepsilon)}{\Pi_\Psi(A_\varepsilon^c)} = \lim_{\varepsilon \to 0} \frac{\Pi_\Psi(A_\varepsilon \,|\, x)}{\Pi_\Psi(A_\varepsilon)} \Big/ \frac{\Pi_\Psi(A_\varepsilon^c \,|\, x)}{\Pi_\Psi(A_\varepsilon^c)} \\ &= \frac{RB_\Psi(\psi_0 \,|\, x)}{RB_\Psi(\{\psi_0\}^c \,|\, x)} = RB_\Psi(\psi_0 \,|\, x) \text{ when } \Pi_\Psi(\{\psi_0\}) = 0 \end{aligned}$$

and so the *BF* and *RB* would agree in the continuous case

- but that is not what is recommended in the continuous case where $\Pi_\Psi(\{\psi_0\}) = 0$

- rather it is recommended that the prior Π be changed by specifying
(i) a prob. $p \in (0,1)$
(ii)a conditional prior $\Pi^*(\cdot \mid H_0)$ for $\theta \in H_0 = \Psi^{-1}\{\psi_0\}$
(iii) a conditional prior $\Pi^*(\cdot \mid H_0^c)$ for $\theta \in H_0^c$ (which is typically Π)
then use the prior (sometimes called a *sharp prior*)

$$\Pi^* = p\Pi^*(\cdot \mid H_0) + (1-p)\Pi^*(\cdot \mid H_0^c)$$

as then

$$BF(H_0 \mid x) = \frac{\Pi^*(H_0 \mid x)}{\Pi^*(H_0^c \mid x)} \Big/ \frac{\Pi^*(H_0)}{\Pi^*(H_0^c)}$$
$$= \frac{p\int_{H_0} f_\theta(x)\Pi^*(d\theta \mid H_0)}{(1-p)\int_{H_0^c} f_\theta(x)\Pi^*(d\theta \mid H_0^c)} \Big/ \frac{p}{1-p} = \frac{m(x \mid H_0)}{m(x \mid H_0^c)}$$

which is a likelihood ratio

- in general $BF(H_0 \mid x) \neq RB_\Psi(\psi_0 \mid x)$ and $BF(H_0 \mid x)$ suffers from "information inconsistency" which $RB_\Psi(\psi_0 \mid x)$ does not

Example *location-scale normal*

- $\{N(\mu, \sigma^2) : \mu \in R^1, \sigma^2 > 0\}, \Psi(\mu, \sigma^2) = \mu,\ H_0 : \Psi(\mu, \sigma^2) = \mu_0$ and sample of n giving

$$L(\mu, \sigma^2 \,|\, x) = (2\pi\sigma^2)^{-n/2} \exp\{-[n(\bar{x} - \mu)^2 + s^2]/2\sigma^2$$

and prior (see Example 5.3.1 for elicitation of hyperparameters)

$$\begin{aligned} \mu \,|\, \sigma^2 &\sim N(\mu_0, \tau_0^2 \sigma^2) \\ \frac{1}{\sigma^2} &\sim \text{gamma}(\alpha_0, \beta_0) \\ \mu &\sim \mu_0 + \sqrt{\tau_0^2 \beta_0 / \alpha_0}\, t_{2\alpha_0} \end{aligned}$$

- then

$$\begin{aligned} RB_\Psi(\mu \,|\, x) &= \frac{\int_0^\infty L(\mu, \sigma^2 \,|\, x)(\tau_0^2 \sigma^2)^{-1/2} \varphi\left(\frac{\mu - \mu_0}{(\tau_0^2 \sigma^2)^{1/2}}\right) \pi(1/\sigma^2)\, d(1/\sigma^2)}{m(x)\pi(\mu)} \\ RB_\Psi(\mu_0 \,|\, x) &= \frac{\int_0^\infty L(\mu_0, \sigma^2 \,|\, x)(\tau_0^2 \sigma^2)^{-1/2} \varphi(0)\, \pi(1/\sigma^2)\, d(1/\sigma^2)}{m(x)\pi(\mu_0)} \end{aligned}$$

- with sharp prior

$$\begin{aligned}\mu \,|\, \sigma^2 &\sim p\delta_{\mu_0} + (1-p)N(\mu_0, \tau_0^2\sigma^2) \\ \frac{1}{\sigma^2} &\sim \text{gamma}(\alpha_0, \beta_0) \\ \mu &\sim p\mu_0 + (1-p)\pi(\mu)\end{aligned}$$

so

$$BF(H_0 \,|\, x) = \frac{\int_0^\infty L(\mu_0, \sigma^2 \,|\, x)\pi(1/\sigma^2)\, d(1/\sigma^2)}{m(x)}$$