Fieller's Problem

STA4522

March 16, 2021

1 Introduction

- a famous problem where various approaches run into difficulties

- initially presented in *Fieller (1954) Some problems in interval estimation. JRSSB, 16, 2, 175–185* and to a certain extent is still unsolved

- the model: $x = (x_1, \ldots, x_m) \overset{i.i.d.}{\sim} N(\mu, \sigma_0^2)$ ind. of $y = (y_1, \ldots, y_n) \overset{i.i.d.}{\sim} N(\nu, \sigma_0^2)$ so means are unknown but the variance is known

- subsequently we'll weaken the assumptions so that the variances are unknown and possibly unequal

- the problem: want to make inference about $\psi = \Psi(\mu, \nu) = \mu/\nu$

- after observing the data (x, y) we want to provide an estimate $\psi(x, y)$ of Ψ and an assessment of the accuracy of the estimate

- also want to assess whether there is evidence in favor of or against $H_0 : \Psi(\mu, \nu) = \psi_0$ for some specified value ψ_0 together with an assessment of the strength of the evidence

- for both problems we also need to specify δ, the difference that matters

- what are some practical examples of such a problem?

- inferences only depend on the mss $(\bar{x}, \bar{y})$ where $\bar{x} \sim N(\mu, \sigma_0^2/n)$ ind. of $\bar{y} \sim N(\nu, \sigma_0^2/m)$

- since the mss is a complete sufficient statistics checking for prior-data conflict also only depends on the mss as ancillaries are independent of the mss (Basu's theorem)

- model checking depends on aspects of the data beyond the mss (like residuals)

- we'll carry along an example where we know the correct answers to demonstrate the methodology

Example

- we set $m = n = 10, \mu = 20, \nu = 10, \sigma_0^2 = 1$ so $\psi = 20/10 = 2$

- the following data was generated

$$x = (19.02, 20.00, 21.13, 19.60, 21.1, 20.18, 21.23, 18.98, 20.48, 20.15)$$
$$y = (10.09, 11.06, 10.10, 10.82, 11.45, 10.55, 11.45, 10.78, 11.14, 9.55)$$

- so $\bar{x} = 20.188, \bar{y} = 10.699$

- for the hypothesis assessment problem consider $H_0 : \Psi(\mu, \nu) = 2$

- take $\delta = 0.2$, could use smaller values too although this will tend to increase necessary Monte Carlo sample sizes to get smooth density estimates

- it is to be remembered that δ is to be specified as part of the application as it represents the accuracy an investigator wants when determining the true value of ψ and this affects how much data they need to collect (so δ is not arbitrary and is not determined by numerical considerations)

2 Eliciting and computing the prior of ψ

- want an elicitation algorithm for μ and ν that takes into account that something is known about ψ as we generally would have some information on this quantity as otherwise we wouldn't be making inference about it

- it is okay to allow 0 as a possible (probable) value for μ but not for both μ and ν as that would suggest the possibility that ψ is not defined (if it isn't defined why are we making inferenece about it?)

- take conjugate priors $\mu \sim N(\mu_0, \tau_{10}^2)$ ind. of $\nu \sim N(\nu_0, \tau_{20}^2)$

- we need to specify $(\mu_0, \tau_{10}^2, \nu_0, \tau_{20}^2)$

- one approach: specify (m_1, m_2) s.t. the true value of $\mu \in (m_1, m_2)$ with virtual certainty, say prior prob. $\gamma = 0.99$

- then put $\mu_0 = (m_1 + m_2)/2$ and solve

$$\Phi((m_2 - \mu_0)/\tau_{10}) - \Phi((m_1 - \mu_0)/\tau_{10}) = \gamma$$

for τ_{10}

- with $(m_1, m_2) = (10, 25)$, then $\mu_0 = ?, \tau_{10} = ?$

- suppose it is known that $\psi \in (r_1, r_2)$ with virtual certainty and we specify a value $\psi_0 \in (r_1, r_2)$ (the hypothesized value makes sense when the inference is hypothesis assesment) and then take $\nu_0 = \mu_0/\psi_0$ so with $(r_1, r_2) = (1, 3), \nu_0 = ?$

- then with virtual certainty $\nu \in (m_1/r_2, m_2/r_1)$ so τ_{20} satisfies

$$\Phi((m_2/r_1 - \nu_0)/\tau_{20}) - \Phi((m_1/r_2 - \nu_0)/\tau_{20}) = \gamma$$

giving $\tau_{20} = ?$

- so now we can generate samples $(\mu_1, \ldots, \mu_N) \overset{i.i.d.}{\sim} N(\mu_0, \tau_{10}^2)$ and $(\nu_1, \ldots, \nu_N) \overset{i.i.d.}{\sim} N(\nu_0, \tau_{20}^2)$ and so obtain a sample $(\psi_1, \ldots, \psi_N) \overset{i.i.d.}{\sim} \pi_\Psi$, where $\psi_i = \mu_i/\nu_i$ from the prior for Ψ

- generally the Monte Carlo sample size N will be something like $10^4, 10^5$ etc.

- typically based on this sample we would determine the interval $(\psi_{\min}, \psi_{\max}) =$ (min. sample value, max. sample value) and divide $(\psi_{\min}, \psi_{\max})$ up into $\lceil(\psi_{\max} - \psi_{\min})/\delta\rceil$ subintervals of length δ,

$$(\psi_{\min} - \delta/2, \psi_{\min} + \delta/2], \ldots, (\psi_{\max} - \delta/2, \psi_{\max} + \delta/2]$$

and take the midpoints of these subintervals as gridpoints (find the subinterval which contains ψ_0)

- **problem -** the prior distribution of Ψ can be extremely long-tailed so a suggestion is to calculate the ecdf $\hat{F}_\Psi$ of Ψ and take

$$(\psi_{\min}, \psi_{\max}) = (\hat{F}_\Psi^{-1}(0.0005), \hat{F}_\Psi^{-1}(0.9995))$$

and then subdivide (so truncating by ignoring 0.001 of the probability in the tails)

- the truncation can be avoided by first transforming from ψ to $\psi_{\text{mod}} = G(\psi)$ where G is a long-tailed cdf (like Cauchy or even sub-Cauchy), transform the above grid to

$$(G(\psi_{\min} - \delta/2), G(\psi_{\min} + \delta/2)], \ldots, (G(\psi_{\max} - \delta/2), G(\psi_{\max} + \delta/2)]$$

and add the tail intervals $[0, G(\psi_{\min} - \delta/2)]$ and $(G(\psi_{\max} + \delta/2), 1]$ with corresponding gridpoints $G(\psi_{\min} - \delta/2)/2$ and $(G(\psi_{\max} + \delta/2) + 1)/2$, respectively and do all the following calculations based on the ψ_{mod} sample $\psi_{\text{mod},1}, \ldots, \psi_{\text{mod},N}$ where $\psi_{\text{mod},i} = G(\psi_i)$ and at the end transform back (recall relative belief inferences are invariant to such transformations) but this step is not required with the current data example

- then record the proportion of Monte Carlo sample values of ψ in each subinterval, divide these proportions by δ (or length of corresponding interval when using ψ_{mod}) to get the density histogram estimate of the prior density of ψ and plot these values against the midpoints to see what the prior density of ψ looks like

- note - you will use exactly the same subintervals when estimating the posterior density of ψ

- Figure 1 is based on a simluation of $N = 10^5$ and the elicited values previously discussed

3 Inference about ψ

- the posterior density of ψ is computed in exactly the same way (use the same δ and grid) by generating from the posterior distributions of μ and ν whose posteriors are normal (please specify these)

- Figure 2 is based on a simluation of $N = 10^5$ and both the prior and posterior densities are on the same plot and note that the posterior is much more concentrated

- dividing the posterior by the prior gives the relative belief ratio $RB_\Psi(\cdot \,|\, x, y)$ which is plotted in Figure 3

- **note** - sometimes when you do this calculation you will get NaN's (not a number) when the both the prior and posterior contents of subintervals are 0 but these are ignorable and can be dealt with by just setting the RB for such a subinterval equal to 0 (recall that if a prior probability is effectively 0 then a

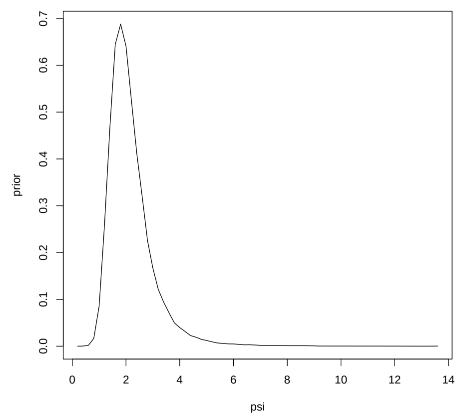

Figure 1: The prior density of ψ.

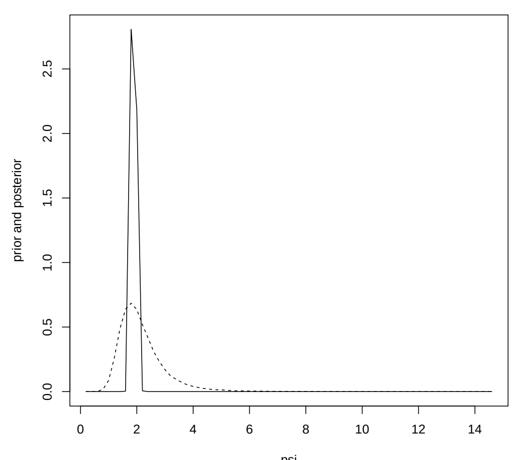

Figure 2: Plots of the prior (- - -) and the posterior (__) densities of ψ.

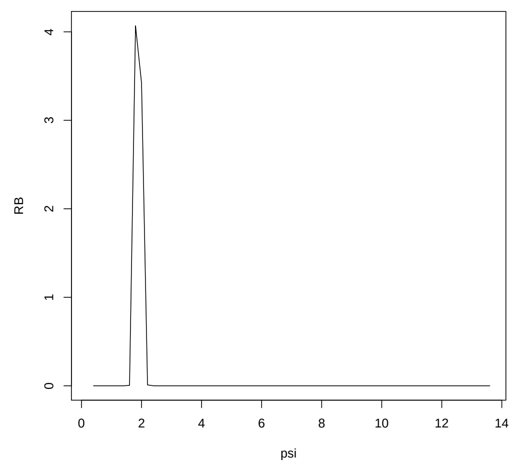

Figure 3: Plot of the relative belief ratio of ψ.

posterior will effectively be 0 unless you have prior-data conflict and then you need to modify the prior)

- the next step is to compute the relative belief estimate $\psi(x)$ =? and the plausible region $Pl_\Psi(x, y)$ =? and note that the estimate is pretty accurate for a relatively small amount of data

- also we want to assess the hypothesis $H_0 : \Psi(\theta) = 2$

- for this we need $RB_\Psi(2 \,|\, x, y)$ and for this you need to find the subinterval in the discretization that contains 2 and record the relative belief ratio for that subinterval and here we obtain $RB_\Psi(2 \,|\, x, y)$ =? which gives evidence ? for H_0

- also it is necessary to calibrate this evidence by computing the strength

$$\Pi(RB_\Psi(\psi \,|\, x, y) \le RB_\Psi(2 \,|\, x, y) \,|\, x, y) =?$$

which is given by the sum of the posterior contents of the subintervals for which the inequality is satisfied

Note

- inferences about ψ are reasonable (actually pretty accurate) for this data, but for the Cox problems A and B problems may ensue

- this is because for the data stated in Fraser, Reid and Lin (2018) (note treat this data as $y_1 = \bar{y}, y_2 = \bar{x}$ and $\sigma_0^2/n = 1$) it is possible that $\nu = 0$ in A (so $\psi = \infty$ is possible) and $\mu = \nu = 0$ in B (so ψ is undefined is possible) both of which cause problems

- we will come back to these issues but for the project just get the inferences working (fill in the question marks) and especially do the bias calcaluations

- to see what the implications of the anomalous situations just described (does it really make sense to make inferences about ψ when it can be infinite or undefined? are there relevant applications?) for relative belief it might be helpful to work out closed form exapressions for the prior and posterior densities of ψ

4 Bias

- to compute the biases for hypothesis assessment we need to compute

$$M(RB_\Psi(\psi_0 \,|\, \bar{x}, \bar{y}) \leq 1 \,|\, \psi) \tag{1}$$

for various values of ψ where $(\bar{x}, \bar{y})$ is generated its from the conditional prior given ψ

- to compute the biases for estimation we need to be able to compute (1) for values of $\psi_0 \sim \pi_\Psi$ and then average

- so it is necessary to

1. be able to generate $(\bar{x}, \bar{y})$ from its conditional prior given $\psi, M(\cdot \,|\, \psi)$
2. be able to compute $RB_\Psi(\psi \,|\, \bar{x}, \bar{y})$ as efficiently as possible

- we consider each of these problems in turn

1. Generating $(\bar{x}, \bar{y})$ from $M(\cdot \,|\, \psi)$

- make the change of variable $(\mu, \nu) \to (\psi, \nu)$ then joint prior 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 $M(\cdot \,|\, \psi)$ generate sequentially as follows

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

where

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

and so the only difficulty is to generate from $\pi(\cdot\,|\,\psi)$

Note the expression for $\nu_0(\psi)$ changed from previous version

- note that $\pi(\cdot\,|\,\psi)$ is close to a normal density but for the factor $|\nu|$

- transforming $\nu \to z = (\nu - \nu(\psi))\,/\tau_{20}(\psi)$ we need to be able to generate z from a density

$$g(z) \propto |a + bz|\varphi(z) \propto |z - z_0|\varphi(z)$$

where $a = \nu(\psi), b = \tau_{20}(\psi)$ and $z_0 = -a/b$ and note $b > 0$

- now using $d\varphi(z)/dz = -z\varphi(z)$

$$\int_{-\infty}^{z_0} -(z - z_0)\varphi(z)\,dz = z_0\Phi(z_0) + \varphi(z_0)$$
$$\int_{z_0}^{\infty} (z - z_0)\varphi(z)\,dz = -z_0(1 - \Phi(z_0)) + \varphi(z_0)$$

$$g(z) = p(z_0)I_{(-\infty,z_0]}(z)g_1(z) + (1 - p(z_0)I_{(z_0,\infty)}(z)g_0(z)$$
$$g_1(z) = \frac{(z_0 - z)\varphi(z)}{z_0\Phi(z_0) + \varphi(z_0)} \text{ when } z \le z_0$$
$$g_0(z) = \frac{(z - z_0)\varphi(z)}{-z_0(1 - \Phi(z_0)) + \varphi(z_0)} \text{ when } z > z_0$$
$$p(z_0) = \frac{z_0\Phi(z_0) + \varphi(z_0)}{-z_0 + 2(z_0\Phi(z_0) + \varphi(z_0))}$$

so generate z from g_1 with prob. $p(z_0)$ and otherwise generate from g_0

- generate from g_1 via inversion where for $z \le z_0$

$$G_1(z) = \int_{-\infty}^{z} g_1(x)\,dx = \frac{z_0\Phi(z) + \varphi(z)}{z_0\Phi(z_0) + \varphi(z_0)}$$

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

- to start the bisection set $z_{up} = z_0$ and iteratively evaluate $G_1(-i|z_0|)$ for $i = 0, 1, \ldots$ until $G_1(-i|z_0|) \le u$ and set $z_{low} = -i|z_0|$ as this guarantees $G_1(z_{low}) \le u \le G_1(z_{up}) = 1$ so bisection will work

- to generate from g_0 proceed similarly where for $z > z_0$

$$G_0(z) = \int_{z_0}^{z} g_0(x)\,dx = \frac{z_0(\Phi(z) - \Phi(z_0)) + (\varphi(z_0) - \varphi(z))}{-z_0(1 - \Phi(z_0)) + \varphi(z_0)}$$

and now start bisection with $z_{low} = z_0$ and iteratively evaluate $G_0(i|z_0|)$ until $u \le G_0(i|z_0|)$ and then set $z_{up} = i|z_0|$

- for given z put $\nu = \nu(\psi) + \tau_{20}(\psi)z$ to get the appropriately generated value of ν

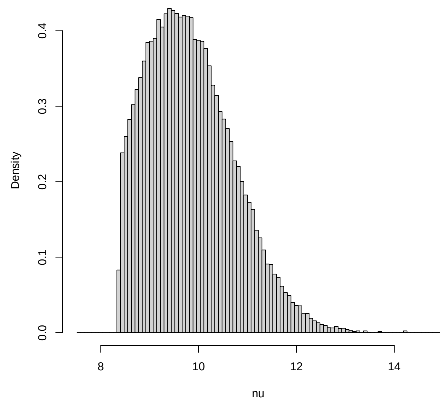

Figure 4: Density histogram of a sample of 10^5 from the conditional prior of ν given $\psi_0 = 2$.

- note - (an interesting consequence) that it must be true that $0 < -z(1 - \Phi(z)) + \varphi(z)$ for every z and this implies the well-known *Mills ratio inequality*

$$\frac{1-\Phi(z)}{\varphi(z)} < \frac{1}{z} \text{ when } z \geq 0 \text{ and } \frac{\Phi(z)}{\varphi(z)} < \frac{1}{|z|} \text{ when } z \leq 0$$

which gives useful bounds on tail probabilities for the normal distribution when $|z|$ is large

Example

- suppose $\psi_0 = 2$ so, using the elicited values of $(\mu_0, \tau_{10}^2, \nu_0, \tau_{20}^2)$

$$\tau_{20}^2(2) = \left(\frac{4}{\tau_{10}^2} + \frac{1}{\tau_{20}^2}\right) =?$$

$$\nu_0(2) = \tau_{20}^2(\psi)\left(\frac{2\mu_0}{\tau_{10}^2} + \frac{\nu_0}{\tau_{20}^2}\right) =?$$

- Figure 4 is density histogram of a sample of 10^5 from $\pi(\cdot \mid \psi_0)$

2. Computing $RB_\Psi(\psi \mid \bar{x}, \bar{y})$

- for bias against H_0, namely $M(RB_\Psi(\psi_0 \mid \bar{x}, \bar{y}) \leq 1 \mid \psi_0)$, we need to calculate $RB_\Psi(\psi_0 \mid \bar{x}, \bar{y})$ for each generated value $(\bar{x}, \bar{y})$ from $M(\cdot \mid \psi)$ and compare it with 1

- for computing $RB_\Psi(\psi_0 \mid \bar{x}, \bar{y})$ there are two ways to proceed in this problem, either numerically or exactly

- the exact method is much more computationally efficient but often the numerical approach is necessary and it is worth recalling that a high degree of accuracy is not required fro the numerical answers when computing biases

Numerical approach

- for this you should have the prior content of the subinterval in the discretization that contains ψ_0 from the inference computations and denote this subinterval by (ψ_{low}, ψ_{up})

- for the posterior contents of this interval you need to compute this in another loop for each generated value of $(\bar{x}, \bar{y})$ and then compute $RB_\Psi(\psi_0 \,|\, \bar{x}, \bar{y})$ as was done in the inference computations

- do this repeatedly and record the proportion of times $RB_\Psi(\psi_0 \,|\, \bar{x}, \bar{y}) \leq 1$ as the estimate of $M(RB_\Psi(\psi_0 \,|\, \bar{x}, \bar{y}) \leq 1 \,|\, \psi_0)$

- for the bias in favor this computation is done for $M(RB_\Psi(\psi_0 \,|\, \bar{x}, \bar{y}) > 1 \,|\, \psi_{low})$ and $M(RB_\Psi(\psi_0 \,|\, \bar{x}, \bar{y}) > 1 \,|\, \psi_{up})$ with the maximum of these values being the bias in favor

- for the bias against for estimation generate a sample $\psi_{01}, \ldots, \psi_{0N^*} \sim \pi_\Psi$ and compute and average the values $M(RB_\Psi(\psi_{0i} \,|\, \bar{x}, \bar{y}) \leq 1 \,|\, \psi_{0i})$

- for the bias in favor for estimation generate a sample $\psi_{01}, \ldots, \psi_{0N^*} \sim \pi_\Psi$ and compute and average the values $M(RB_\Psi(\psi_{0i} \,|\, \bar{x}, \bar{y}) > 1 \,|\, \psi_{0i})$

Closed form approach

- an exact formula can be determined for $RB_\Psi(\psi \,|\, \bar{x}, \bar{y})$ which can be used in the bias calculations to make the computatiions more efficient

- so when $\mu \sim N(\mu_0, \tau_{10}^2)$ ind. of $\nu \sim N(\nu_0, \tau_{20}^2)$, then some calculation (posted on the website) determines the prior density of $\psi = \mu/\nu$ as

$$\pi_\Psi(\psi) = \sqrt{\frac{2}{\pi}} \frac{1}{\tau_{10}\tau_{20}} \left(\frac{\psi^2}{\tau_{10}^2} + \frac{1}{\tau_{20}^2} \right)^{-1} \times$$
$$\exp\left\{ -\frac{1}{2} \left(\frac{\psi^2}{\tau_{10}^2} + \frac{1}{\tau_{20}^2} \right)^{-1} \left[\frac{(\mu_0 - \nu_0\psi)^2}{\tau_{10}^2\tau_{20}^2} \right] \right\} \times$$
$$\left\{ \begin{array}{c} \varphi\left(\left(\frac{\psi^2}{\tau_{10}^2} + \frac{1}{\tau_{20}^2} \right)^{-1/2} \left(\frac{\mu_0}{\tau_{10}^2\psi} + \frac{\nu_0}{\tau_{20}^2} \right) \right) + \\ \left(\frac{\psi^2}{\tau_{10}^2} + \frac{1}{\tau_{20}^2} \right)^{-1/2} \left(\frac{\mu_0}{\tau_{10}^2\psi} + \frac{\nu_0}{\tau_{20}^2} \right) \left(\Phi\left(\left(\frac{\psi^2}{\tau_{10}^2} + \frac{1}{\tau_{20}^2} \right)^{-1/2} \left(\frac{\mu_0}{\tau_{10}^2\psi} + \frac{\nu_0}{\tau_{20}^2} \right) \right) - \frac{1}{2} \right) \end{array} \right\}$$

- note that when $\mu_0 = \nu_0 = 0$ then π_Ψ is a (rescaled) Cauchy density

- it is a formidable looking formula but there is actually a fair bit of symmetry so it is not hard to write an R function for it as the following shows

$$\begin{aligned} \pi_\Psi(\psi) = \sqrt{\frac{2}{\pi}} \frac{\tau_{20}^2(\psi)}{\pi\tau_{10}\tau_{20}} \exp\left\{ -\frac{1}{2} \frac{\tau_{20}^2(\psi)(\mu_0 - \nu_0\psi)^2}{\tau_{10}^2\tau_{20}^2} \right\} \times \\ \left\{ \varphi\left(\frac{\nu_0(\psi)}{\tau_{20}(\psi)} \right) + \frac{\nu_0(\psi)}{\tau_{20}(\psi)} \left(\Phi\left(\frac{\nu_0(\psi)}{\tau_{20}(\psi)} \right) - \frac{1}{2} \right) \right\} \end{aligned} \tag{2}$$

where

$$\tau_{20}^2(\psi) = \left(\frac{\psi^2}{\tau_{10}^2} + \frac{1}{\tau_{20}^2} \right)^{-1} \text{ and } \nu_0(\psi) = \tau_{20}^2(\psi) \left(\frac{\mu_0}{\tau_{10}^2}\psi + \frac{\nu_0}{\tau_{20}^2} \right)$$

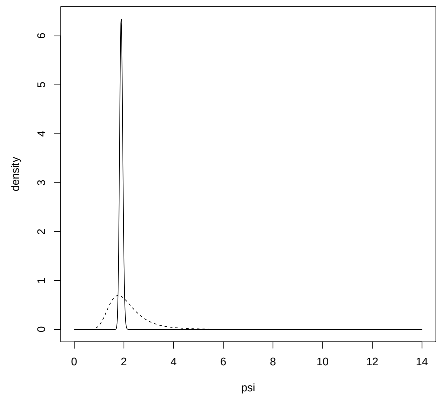

Figure 5: Plots of the exact prior (- - -) and the exact posterior (__) densities of ψ.

-note - there was a typo in the last term inside the curly brackets in the first version I put up and this expression is more numerically stable even with the correction, also the previous formula for $\nu_0(\psi)$ was wrong

- the expression inside $\{\}$ of (2) implies the inequality

$$\varphi(z) + z\,(\Phi(z) - 1/2) > 0 \text{ for all } z$$

but this can also be deduced from the Mills ratio inequality

- for the posterior we have $\mu \,|\, \bar{x} \sim N(\mu_x, \tau_{1x}^2)$ ind. of $\nu \,|\, \bar{y} \sim N(\nu_y, \tau_{2y}^2)$ and so the posterior density of ψ will take exactly the same form as the prior but with $(\mu_0, \tau_{10}^2, \nu_0, \tau_{20}^2)$ replaced by $(\mu_x, \tau_{1x}^2, \nu_y, \tau_{2y}^2)$

- Figure 5 is a plot of the exact prior and posterior of ψ which is verysimilar to Figure 2 but much smoother as the latter was based on a discretization into subintervals of length 0.2 and about 75 plotted points while the former is based on 1000 plotted points

- Figure 6 is plot of the exact relative belief ratio of ψ

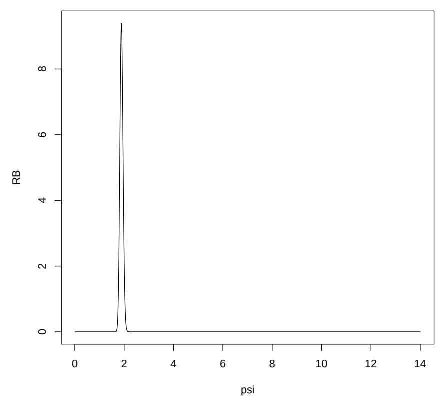

Figure 6: Plot of the exact relative belief ratio.