`brr`: An R package for Bayesian analysis of the rate ratio

Stéphane Laurent

July 7, 2015

This is the companion paper of the **brr** package. We present the Bayesian model implemented in **brr** and the usage of this package. The present paper also addresses some extensions of the procedures available in the literature.

Contents

1. Introduction

The **brr** R package implements the Bayesian analysis of the *two Poisson samples model* with the *semi-conjugate family of priors* introduced by Laurent & Legrand ([13]). It is not available on CRAN yet, but it is available at **https://github.com/stla/brr**, where the instructions for installation are given.

The statistical model of interest is the *two Poisson samples model*, given by two independent observations

$$\begin{cases} x \sim \mathscr{P}(\lambda S) \\ y \sim \mathscr{P}(\mu T) \end{cases}$$

where $\lambda > 0$ and $\mu > 0$ are the unknown incidence rates and $S, T > 0$ are the known observation-opportunity sizes. When S and T are some durations, they are sometimes called the times at risk. For short, we will say the *sample sizes* throughout the paper.

This model can also be used in the case of two independent series of observations $(x_i)_{i=1}^m$ and $(y_i)_{i=1}^n$ assuming

$$\begin{cases} x_i \sim_{i.i.d.} \mathscr{P}(\lambda S_i) \\ y_i \sim_{i.i.d.} \mathscr{P}(\mu T_i) \end{cases}$$

by setting $S = \sum S_i$ and $T = \sum T_i$, and by using as observations the sufficient statistics $x = \sum x_i$ and $y = \sum y_i$.

Statistical inference on the rate ratio $\phi := \frac{\lambda}{\mu}$ $\phi = \lambda / \mu$ $\phi := \frac{\lambda}{\mu}, \phi := \frac{\lambda}{\mu}, \phi := \frac{\lambda}{\mu}$, also called the *relative risk*, has a large broad of applications. For instance, it is used in vaccine efficacy studies. To study the effect of a vaccine, clinical statisticians randomly assign some subjects to a vaccine group and others to a placebo group. Having this example in mind, we will say that, throughout this paper, the first group (with sample size S) is the treated group and the second group (with sample size T) is the control group. After a follow-up, one observes the numbers x and y of individuals who developed the disease in the vaccine group and the placebo group respectively. Here, the sample sizes S and T are the person-times at risk, defined as the sum of the follow-up times of all individuals in the corresponding group. The unknown parameters λ and μ are interpreted as the disease incidence rates in the vaccine group and the placebo group respectively. The vaccine efficacy is then defined by $VE = 1 - \phi \in\ ] - \infty, 1]$.

The semi-conjugate family of priors is particularly appropriate when the parameter of main interest is the relative risk ϕ. Three different types of prior distributions are available in the semi-conjugate family, which are distinguished according to their informational aspect:

- *the reference prior*, also called the *non-informative prior*: to be used in the case when one does not want to include some prior information about the incidence rates;

- *a semi-reference prior*, also called a *semi-informative prior*: to be used in a case when one wants to include some prior information about the incidence rate μ while one does not want to include some information about the relative risk ϕ;

- *a fully informative prior*, also called an *informative prior*: to be used in the case when one wants to include some prior information about all parameters.

The reference prior is derived from Bernardo's theory of objective Bayesian statistics, further developed by Berger & Bernardo ([1], [2]). Bernardo advocates the use of reference priors as standards for scientific communications. His arguments are based on solid theoretical foundations as well as judicious philosophical considerations. The reference prior is relative to the choice of the parameter of interest, here the relative risk ϕ, and it is intended to yield a posterior distribution which best reflects the information about ϕ brought by the data and only the data. In fact, there is *no prior distribution* in this framework, there is only a posterior distribution. The reference prior is somehow an artificial prior distribution, whose only role is to yield the reference posterior distribution by a formal application of Bayes' therorem. The fascinating story of this theory is beyond the scope of this paper. The interested reader is referred to [3], [8], [10].

Following a procedure proposed by Berger & Sun in [4], the semi-reference prior is defined in the same way as the reference prior after choosing a prior distribution on μ and then averaging the model over μ with respect to this subjective prior distribution on. Thus, in this framework, there is a prior distribution on μ, but not on ϕ.

In Section 2 we present the usage **brr** after recalling the semi-conjugate family of prior distributions.

In addition to the reference posterior distribution, Bernardo later developed *integrated objective Bayesian statistics*. The reader is referred to [8], [9], [10], [11] for details and historical references. Integrated objective Bayesian statistics provide a Bayesian decision-theoretic solution to the usual problems of inference summaries: point estimation, region estimation (credibility intervals), and hypotheses testing. Following the Bayesian decision theory, each of these procedures is based on a loss function, and the one proposed by Bernardo, namely the *intrinsic discrepancy loss*, is derived from information-theoretic considerations. These procedures are implemented in **brr** and will be presented in Section 3. In addition to the procedures based on the intrinsic discrepancy loss as it is originally defined, **brr** includes similar procedures based on another intrinsic discrepancy loss, more appropriate for the case of a semi-informative prior.

Finally, in Section 4, we present some possible extensions of the semi-conjugate family to the case of several control groups.

2. Usage of **brr**

We show in this section how to use **brr**. We mainly present the user-friendly way to use it. The main points to remember are:

- create a **brr** object with the **Brr** function, in which you set the prior parameters a, b, c, d, the sample sizes S and T, and the observed counts x and y;

- do not set the prior parameters if you want to use the non-informative prior.

We use the pipe operator `%>%` from the **magrittr** package for our illustrations.

```
> library(magrittr)
```

2.1. The semi-conjugate family

Before speaking about **brr** itself, let us recall the natural semi-conjugate family of prior distributions introduced in [13] and the corresponding posterior distributions.

It involves the *Beta prime* distribution and the *Beta distribution of the second kind*. The Beta prime distribution $\mathscr{B}'(c,d)$ with shape parameters $c,d > 0$ is the law of the $(0,+\infty)$-valued random variable $U(1-U)^{-1}$ where U is a random variable following the Beta distribution $\mathscr{B}(c,d)$ on $(0,1)$. The Beta distribution of second kind $\mathscr{B}_2(c,d;\tau)$ with shape parameters $c,d > 0$ and scale parameter τ is the distribution $\tau \times \mathscr{B}'(c,d)$ (that is to say the law of a $\mathscr{B}'(c,d)$ random variable multiplied by τ). We use the notation $\tau \times \mathscr{B}'(c,d)$ below, which has the advantage to show that τ is the scale parameter.

Now the semi-conjugate family is described below.

For any positive numbers a,b,c,d, assign the followng independent prior distributions on μ and $\phi = \lambda/\mu$:

$$\mu \sim \mathscr{G}(a,b) \quad and \quad \phi \sim \frac{T+b}{S} \times \mathscr{B}'(c,d).$$

Then the joint posterior on (μ,ϕ) distribution is given by the semi-conjugacy rule

$$(\mu \mid \phi, x, y) \sim \mathscr{G}(a + x + y, b + \phi S + T) \quad and \quad (\phi \mid x, y) \sim \frac{T+b}{S} \times \mathscr{B}'(c + x, a + d + y).$$

Note that the posterior is also a proper distribution when $d = 0$ or $b = 0$. The following results are derived in [13]:

- *when $a = c = 0.5$ and $b = d = 0$, the prior is the reference prior when ϕ is the parameter of interest (we will simply call it the reference prior, since it will always be understood that ϕ is the parameter of interest);*

- *when $a, b > 0$, $c = 0.5$ and $d = 0$, the prior is the semi-reference prior when the arbitrary Gamma prior distribution $\mathscr{G}(a,b)$ is assigned on μ.*

The prior and posterior predictive distributions of the counts x and y are provided in [14].

The following functions related to the semi-conjugate family are implemented in the **brr** package.

- Any density function, distribution function, quantile function, and sampling function of a prior distribution has a name starting with, respectively, **dprior_**, **qprior_**, **pprior_** and **rprior_**, followed by the name of the parameter or the name of an observation for a prior predictive distribution. For example **dprior_mu**, **dprior_phi**, **dprior_lambda**, **dprior_x**, **dprior_y**. The **pprior_lambda** function fails in many cases (see Appendix A). The **qprior_lambda** function is not available. In the same way, functions whose name starts with **sprior_** generate a summary of the distribution. In addition, there are the functions **[d/p/q/r/s]prior_x_given_y** for the conditional prior predictive distribution of x given a value of y.

- Similarly, any density function, distribution function, quantile function, sampling function and summary function of a posterior distribution has a name starting with, respectively, **dpost_**, **qpost_**, **ppost_**, **rpost_** and **spost**, followed by the name of the parameter or the name of an observation for a posterior predictive distribution. Functions **ppost_mu**, **qpost_mu**, **ppost_lambda** and **qpost_lambda** functions are not available.

- As we will see in the next section, all these functions can be more conveniently called by setting the parameters in a **brr** object and using functions **dprior**, **dpost**, etc.

Some distributions, such as the prior and posterior distributions of λ, and the prior and posterior predictive distribution of x, are not standard distributions. These special distributions and their implementation or partial implementation in the **brr** package are the subject of Appendix A.

2.2. The ergonomics of **brr**

The **Brr** function is intended to offer a convenient usage of the **brr** package. For example, start by specifying the parameters a and b of the Gamma prior distribution on μ:

```
> model <- Brr(a=5, b=10)
```

The **model** variable created by this way is a **brr** object, and it can be used to call more conveniently the functions related to the Bayesian inference. For example, instead of calling **dprior_mu(.,a=5,b=10)**, **pprior_mu(.,a=5,b=10)**, etc, you can simply use **dprior(model,"mu",.)**, **pprior(model,"mu",.)**, etc:

```
> pprior(model, "mu", 1)

[1] 0.9707473

> pprior_mu(1, a=5, b=10)
```

```
[1] 0.9707473
```

In the same way, **pprior(model,"phi",.)** is a wrapper for the **pprior_phi** function. For the moment it crashes because the required parameters are not specified yet:

```
> try(pprior(model, "phi", 1)) %>% cat

Error in brr_generic("pprior", model, parameter, ...) :
  Missing parameters. You must supply b, c, d, S, T.
```

As previously said, the variable **model** is a **brr** object. This is just a function returning the list of specified parameters when it is called without arguments:

```
> model() %>% unlist

 a  b
 5 10
```

But if it is called with arguments, then these arguments are added to the list:

```
> model <- model(c=3, d=4, S=10, T=10)
> model() %>% unlist

 c  d  S  T  a  b
 3  4 10 10  5 10
```

Once a, b, c, d, S, T are given in the **brr** object, all available functions related to prior distributions can be called using **dprior**, **pprior**, **qprior**, **rprior** and **sprior**, for example:

```
> pprior(model, "phi", 1)

[1] 0.3196159

> dprior(model, "lambda", 1)

[1] 0.3954872

> sprior(model, "phi") %>% unlist

      mode      mean        sd        Q1        Q2        Q3
 0.8000000 2.0000000 2.0000000 0.8446125 1.4566624 2.4762582
```

To get an overall summary of the model, apply, as usual for a R user, the **summary** function to the **brr** object:

```
> summary(model)

Type of prior distribution: informative prior

*Prior distribution on µ*:   Gamma(a=5,b=10)
```

mode	mean	sd	Q1	Q2	Q3
0.4	0.5	0.2236	0.3369	0.4671	0.6274

```
*Prior distribution on φ*:   Beta2(c=3,d=4,scale=2)
```

mode	mean	sd	Q1	Q2	Q3
0.8	2	2	0.8446	1.457	2.476

```
*Sample sizes*
  S (treated group): 10
  T (control group): 10

*Observed counts*
  x (treated group): not supplied yet
  y (control group): not supplied yet

*Posterior distribution on φ*:
  a, b, c, d, S, T, x and y must be supplied
```

To use the non-informative prior, simply do not set parameters *a*, *b*, *c* and *d*, or set them to **NULL**. Setting new values to a **brr** object replaces the old ones:

```
> model_noninfo <- model(a=NULL, b=NULL, c=NULL, d=NULL)
> model_noninfo() %>% unlist

 S  T
10 10
```

As said in the introduction, the non-informative prior is not a real prior distribution. Therefore, there is nothing to get from functions **dprior**, **pprior**, etc. There only are posterior distributions, obviously requiring the observations *x* and *y* in addition to the samples sizes *S* and *T*:

```
> model_noninfo <- model_noninfo(x=5, y=10)
> spost(model_noninfo, "phi") %>% unlist

      mode      mean        sd        Q1        Q2        Q3
0.3913043 0.5789474 0.3279396 0.3512455 0.5085074 0.7259461
```

Estimates and credibility intervals for the relative risk are given with functions **coef** and **confint**:

```
> coef(model_noninfo)

Estimates of φ

mode : 0.3913043
mean : 0.5789474
median : 0.5085074
intrinsic : 0.5104077
intrinsic2 : NA

> confint(model_noninfo)

95%-credibility intervals about φ
```

interval	lwr	upr
equi-tailed	0.1631	1.405
equi-tailed*	0.1631	1.405
hpd	0.1009	1.218
intrinsic	0.1454	1.313
intrinsic2	NA	NA

The meaning of **intrinsic** and **intrinsic2** will be explained in Section 3. The outputs in the **intrinsic2** lines are always **NA** when one uses the non-informative prior.

Descarding these lines for the moment, the **confint** function returns three credibility intervals. The *equi-tailed* and *hpd* intervals are well-known in Bayesian statistics. The *equi-tailed** credibility interval is introduced in [13] as a correction of the equi-tailed interval in the case when the posterior distribution of φ has its mode at 0. In

other cases, it is exactly the same as the equi-tailed interval. The case of the correction occurs when $c + x \leq 1$, which happens when one uses the non-informative prior and gets the count $x = 0$ in the control group.

Predictions are provided by the **predict** function. This requires to set the values of the new sample sizes in the **brr** object, under the names **Snew** and **Tnew**, but these values are automatically set to be **S** and **T** if they are not provided by the user:

```
> predict(model_noninfo)
```

Predictions and 95%-credibility prediction intervals

```
+-------+--------+----------+-------+-------+
|  obs  |  size  |  median  |  lwr  |  upr  |
+=======+========+==========+=======+=======+
| xnew  |   10   |    5     |   1   |  13   |
+-------+--------+----------+-------+-------+
| ynew  |   10   |    10    |   3   |  20   |
+-------+--------+----------+-------+-------+
```

Alternatively, instead of setting the new sample sizes **Snew** and **Tnew** in the **brr** object, it is possible to pass them as arguments of the **predict** function:

```
> predict(model_noninfo, Snew=20, Tnew=20)
```

Predictions and 95%-credibility prediction intervals

```
+-------+--------+----------+-------+-------+
|  obs  |  size  |  median  |  lwr  |  upr  |
+=======+========+==========+=======+=======+
| xnew  |   20   |    10    |   2   |  24   |
+-------+--------+----------+-------+-------+
| ynew  |   20   |    20    |   8   |  38   |
+-------+--------+----------+-------+-------+
```

2.3. Comparison with frequentist inference

It is interesting to compare the results of the Bayesian inference with a frequentist inference. A traditional frequentist way for statistical inference on the relative risk ϕ consists in considering the conditional distribution of x given the sum $x + y$, which is binomial with size parameter $n = x + y$ and proportion parameter

$$\theta = \frac{\lambda S}{\lambda S + \mu T} = \frac{\phi S/T}{\phi S/T + 1}.$$

Then one gets inference on ϕ by inferring on θ in the binomial model and finally back-transforming to get inference on ϕ.

The function **rr_interval_binomial** in the **brr** provides the confidence interval of ϕ obtained by this way, and by taking the exact binomial confidence interval of θ (the one provided by the well-known **binom.test** function). Another function, **rr_interval_SK**, provides the Sahai & Khurshid confidence interval which is studied in [13] for comparison with the Bayesian inference based on the reference posterior distribution. The function **rr_intervals** returns both frequentist confidence intervals. The **with** function is useful to apply these functions from a **brr** object:

```
> with(model_noninfo(), rr_intervals(x, y, S, T))

$binomial
[1] 0.1340969 1.6054984

$SK
[1] 0.1585795 1.4280107

attr(,"level")
[1] 0.95
```

As shown in [13], the Bayesian credibility intervals based on the reference posterior distribution achieve a very good frequentist matching property: their frequentist coverage is close to their credibility interval. For example a 95%-credibility interval can be considered as a 95%-confidence interval.

In fact, after replacing x with $x - 0.5$ and y with $y + 0.5$, the upper bound of the binomial confidence interval becomes the same as the one of the equi-tailed credibility interval:

```
> with(model_noninfo(),
+      rr_interval_binomial(x-.5, y+.5, S, T)
+ )

[1] 0.1079239 1.4047929
```

And after replacing x with $x + 0.5$ and y with $y - 0.5$, the lower bound of the binomial confidence interval becomes the same as the one of the equi-tailed credibility interval:

```
> with(model_noninfo(),
+      rr_interval_binomial(x+.5, y-.5, S, T)
+ )

[1] 0.1631324 1.8342167
```

10

2.4. Graphical outputs

We take the **model_noninfo** example to illustrate this section, after renaming it to a more convenient name:

```
> model <- model_noninfo
```

As we have seen, the function **dpost(model,"phi",.)** (for example) is convenient to evaluate the posterior density of ϕ. It can be used as follow to draw the density:

```
> phi <- seq(0.5, 2, length.out=101)
> phi %>% { plot(., dpost(model, "phi", .), type="l") }
```

But **brr** provides a custom **plot** function for **brr** objects. For example a graphic equivalent to the one generated by the previous code, but with automatic limits and aesthetics, can be generated as follows (it is shown on Figure 1):

```
> par(mar=c(4,1,0,0))
> plot(model, dpost(phi), cex.lab=2, cex.axis=2)
```
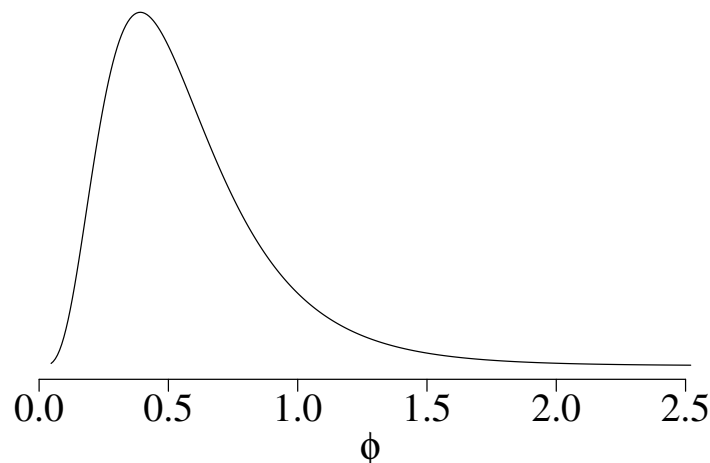

Figure 1: Output of **plot(model, dpost(phi))**

Similary, **plot(model, dprior(lambda))** displays the curve of the prior density of λ, **plot(model, pprior(mu))** displays the curve of the prior distribution function of μ, etc.

The **plot** function also generates automatic bar plots for the predictive distributions, but the new sample size must be included in the **brr** object. For example (see Figure 2):

```
> model <- model(Snew=10)
> par(mar=c(4,4,0,0))
> plot(model, dpost(x), cex.lab=2, cex.names=2, cex.axis=2)
```
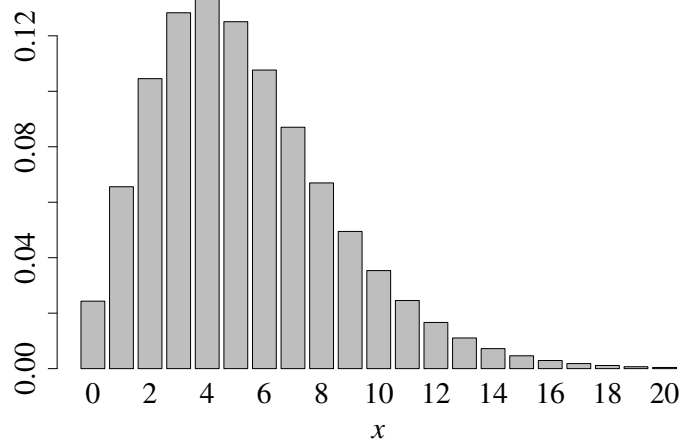
Figure 2: Output of `plot(model, dpost(x))`

3. Intrinsic hypothesis testing and intrinsic credible intervals

The *intrinsic discrepancy loss*, has been introduced by Bernardo and Rueda in [6], and further considered by Bernardo and Juarez in [7]. It can be used for inference summaries such as estimation but also for the problem of Bayesian hypothesis testing ([6, 8, 11]), and the problem of producing Bayesian credible regions ([9, 11]). All these procedures are developed in the objective perspective, when one uses the reference prior, but it is possible to use them with a subjective prior.

These procedures in the case of the two Poisson samples model, and their implementation in **brr**, are explained in subsection 3.1. All necessary calculations are provided in [15].

In subsection 3.2 we present another intrinsic discrepancy which is more appropriate in the case when one uses a semi-informative prior. All inference procedures that follow from it are implemented in **brr**. The necessary calculations are provided in Appendix B.

The **brr** functions related to this section are listed below, where **intrinsic(2)** means either **intrinsic** (first intrinsic discrepancy) or **intrinsic2** (second intrinsic discrepancy):

- **intrinsic(2)_discrepancy**, computing the intrinsic discrepancy loss;

- **intrinsic(2)_phi0**, computing the posterior loss at a given value ϕ_0 of ϕ;

- **intrinsic(2)_estimate**, computing the intrinsic estimate of ϕ, defined as the value of ϕ at which the posterior loss is minimal;

- **intrinsic(2)_interval**, computing the intrinsic credibility interval about ϕ;

12

- `intrinsic(2)_H0`, computing the posterior loss for intrinsic hypothesis testing.

3.1. First intrinsic discrepancy

The case of the two Poisson samples model with the semi-conjugate family of priors is studied in [15]. For this model with ϕ as the parameter of interest, the intrinsic discrepancy loss is the quantity $\ell(\phi_0; \mu, \phi)$ interpreted as a measure of incompatibility of using ϕ_0 as a proxy for the relative risk when the true values of the parameters are ϕ and μ, and it is defined by

$$\ell(\phi_0; \mu, \phi) := \inf_{\mu_0 > 0} \min\{\delta(\mu, \phi \mid \mu_0, \phi_0), \delta(\mu_0, \phi_0 \mid \mu, \phi)\},$$

where $\delta(\mu, \phi \mid \mu_0, \phi_0)$ is the Kullback-Leibler divergence from the sampling distribution at (μ, ϕ) to the sampling distribution at (μ_0, ϕ_0).

Following Bayesian decision-theoretic principles, the posterior expected intrinsic discrepancy (for short, the posterior loss) $\bar{\ell}(\phi_0 \mid x, y)$ is obtained by integrating $\ell(\phi_0; \mu, \phi)$ over the posterior distribution of (μ, ϕ). As shown in [15], this can be expressed as a one-dimensional integral. The posterior loss $\bar{\ell}(\phi_0 \mid x, y)$ is evaluated in the **brr** package by the `intrinsic_phi0` function. Figure 3 shows the curve $\phi_0 \mapsto \bar{\ell}(\phi_0 \mid x, y)$ when we use the reference prior and we observe the two counts $x = 10$ and $y = 20$ with sample sizes $S = T = 10000$.
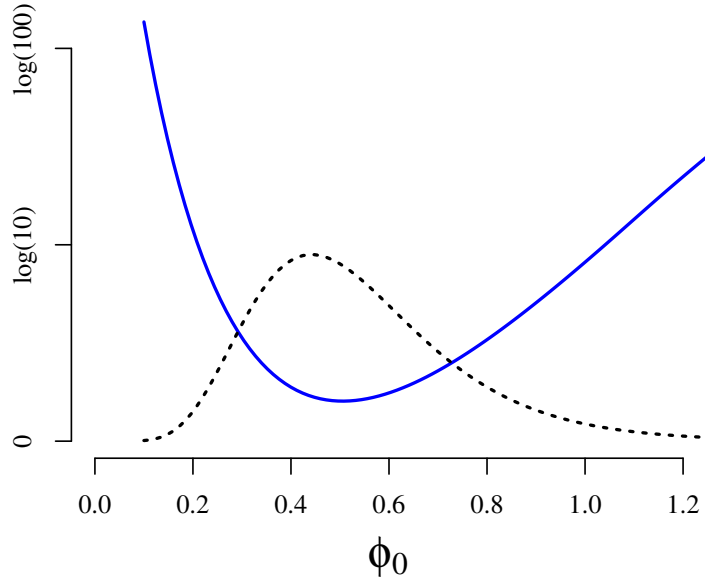

Figure 3: Expected posterior intrinsic loss (blue) and posterior density of ϕ (black, dotted).

13

Intrinsic estimate. The posterior loss $\bar{\ell}(\phi_0 \mid x, y)$ is a convex function of ϕ_0. The value of ϕ_0 where it achieves its minimum is the *intrinsic estimate* of the relative risk ϕ. The simplest way to get it with the **brr** is to call the **coef** function on a **brr** object. The example below is in the same situation as the one of Figure 3.

```
> model <- Brr(x=10, y=20, S=10000, T=10000)
> coef(model)

Estimates of ϕ

mode : 0.4418605
mean : 0.5384615
median : 0.5042123
intrinsic : 0.5055056
intrinsic2 : NA
```

Typically, the intrinsic estimate is greater than the mode, and close to the median. The **intrinsic2** line in the output of **coef(model)** returns the intrinsic estimate based on the second intrinsic discrepancy (object of the next subsection). It is **NA** in this example, because the second intrinsic discrepancy only makes sense when one uses an informative prior on μ.

Intrinsic credible interval. The intrinsic $100(1 - \alpha)\%$-credibility interval is defined as the $100(1 - \alpha)\%$-credibility interval made of values of ϕ_0 having the lower posterior loss $\bar{\ell}(\phi_0 \mid x, y)$. An example is given on Figure 4, on which the intrinsic credible interval is shown by the green dotted segment. The figure also shows the construction of this interval. It is easy to compute it by optimization when the inverse cumulative function F^{-1} of the posterior distribution of ϕ is available, following a way similar to the one used for the computation if the HPD interval. Indeed, every $100(1 - \alpha)\%$-credibility interval has form $[F^{-1}(\epsilon), F^{-1}(\epsilon + 1 - \alpha)]$ for some value of $\epsilon \in [0, \alpha]$, and among these intervals, the intrinsic credible interval is the one whose bounds $F^{-1}(\epsilon)$ and $F^{-1}(\epsilon + 1 - \alpha)$ have the same posterior loss.

Intrinsic hypothesis testing. Given a hypothesis $H_0 \colon \{\phi \in \Phi_0\}$, the intrinsic discrepancy from H_0 to (μ, ϕ) is defined as the minimal value of the intrinsic discrepancy $\ell(\phi_0; \mu, \phi)$ when ϕ_0 runs over H_0:

$$\ell(H_0; \mu, \phi) = \min_{\phi_0 \in \Phi_0} \ell(\phi_0; \mu, \phi)$$

Thene intrinsic hypothesis test (called the *Bayesian reference criterion* in case of the reference prior) is defined by the rejection rule:

$$\begin{cases} \text{reject } H_0 & \text{if } \bar{\ell}(H_0 \mid x, y) > \ell^* \\ \text{do not reject } H_0 & \text{if } \bar{\ell}(H_0 \mid x, y) \leq \ell^* \end{cases}$$

where ℓ^* is a given threshold and $\bar{\ell}(\Phi_0 \mid x, y)$ is the posterior expected value of the intrinsic discrepancy from Φ_0 to (ϕ, μ), or, for short, the posterior loss.
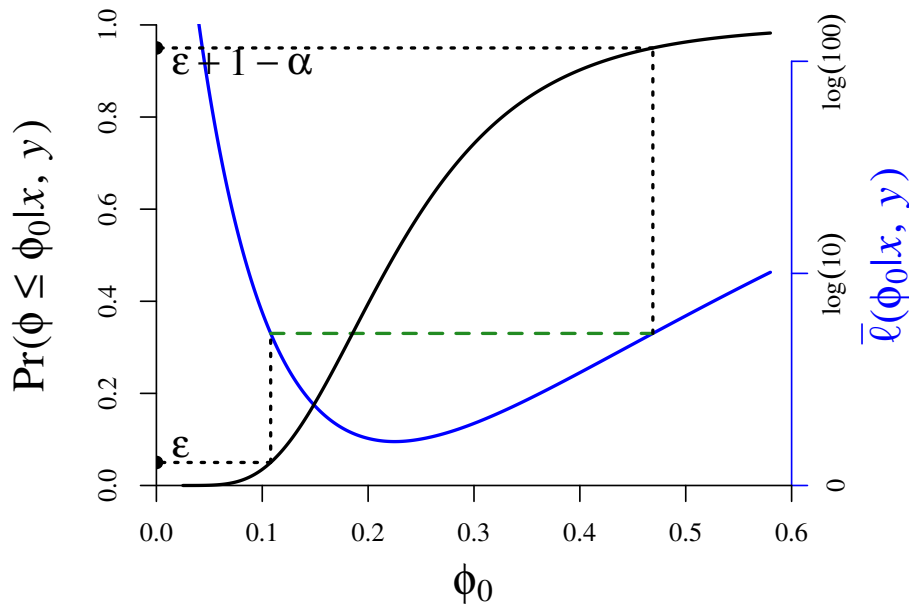
Figure 4: Construction of the intrinsic credible interval.

Bernardo ([8, 11]) proposes to consider threshold values $\ell^* = \log 10, \log 100, \log 1000$ as mild, strong, and very strong evidence against H_0, respectively.

The **brr** package performs the intrinsic hypothesis test for one-sided hypotheses $H_0\colon \{\phi > \phi^*\}$ or $H_0\colon \{\phi < \phi^*\}$ with the **intrinsic_H0** function. For the example of Figure 3, we get a value lower than $\log 10$ for the hypothesis $H_0\colon \{\phi > 1\}$, thereby not allowing to reject H_0 with the mild evidence threshold $\ell^* = \log 10$.

```
> model <- Brr(x=10, y=20, S=10000, T=10000)
> with(model(),
+       intrinsic_H0(phi.star=1, alternative="less", x=x, y=y,S=S, T=T
+ ) %>% exp

[1] 8.115875
```

For comparison, the conditional frequentist method yields a p-value about 10%:

```
> with( model(), binom.test(x,x+y)$p.value )

[1] 0.09873715
```

3.2. Second intrinsic discrepancy

As said before, the intrinsic discrepancy was originally developed to be used with the non-informative reference prior, but nothing prevents one to use it with an informative subjective prior.

15

However, if one uses a semi-informative prior (informative about μ, non-informative about ϕ), then, following the principles of objective Bayesian theory, the intrinsic discrepancy should be derived from the marginal model, that is to say the model having ϕ as unique parameter, obtained after averaging the original two-parameters model with respect to the subjective prior distribution $\mathscr{G}(a,b)$ on μ.

In the marginal model, the sampling distribution of the pair (x,y) of the two counts is a bivariate negative binomial distribution, having ϕ as unique unknown parameter (a fortiori the parameter of interest). The derivation of the intrinsic discrepancy for this model is easier than its derivation for the original two-parameters model. We denote it by $\ell_{a,b}$ to empasize it is derived after the prior distribution $\mathscr{G}(a,b)$ has been fixed. It is defined by

$$\ell_{a,b}(\phi_0;\phi) := \min\{\delta(\phi \mid \phi_0), \delta(\phi_0 \mid \phi)\},$$

where $\delta(\phi \mid \phi_0)$ is the Kullback-Leibler divergence from the sampling distribution at ϕ to the sampling distribution at ϕ_0. Its derivation is provided in Appendix B.

The posterior expected loss $\bar{\ell}_{a,b}(\phi_0 \mid x,y)$ is then defined by averaging the intrinsic discrepancy $\ell_{a,b}(\phi_0;\phi)$ over ϕ with respect to its posterior distribution. Then the intrinsic estimate and the intrinsic credible interval are defined exactly as in the case of the two-parameters model. Intrinsic hypothesis testing of a hypothesis $H_0 \colon \{\phi \in \Phi_0\}$ is also similar to the case of the two-parameters model. It is based on the intrinsic discrepancy $\ell_{a,b}(H_0;\phi)$ from H_0 to ϕ, defined as the minimal value of the intrinsic discrepancy $\ell_{a,b}(\phi_0;\phi)$ when ϕ_0 runs over H_0. Then the posterior expected loss $\bar{\ell}_{a,b}(H_0 \mid x,y)$ is obtained by averaging the intrinsic discrepancy $\ell_{a,b}(H_0;\phi)$ with respect to the posterior distribution of ϕ, and then the test is carried out by comparing the posterior expected loss $\bar{\ell}_{a,b}(H_0 \mid x,y)$ to some threshold ℓ^*.

4. Extension to several treated groups

Consider now the three samples Poisson model in the situation when there are two treated groups:

$$\begin{cases} x_1 \sim \mathscr{P}(\lambda_1 S_1) \\ x_2 \sim \mathscr{P}(\lambda_2 S_2) \\ y \sim \mathscr{P}(\mu T) \end{cases}$$

and denote $\phi_1 = \frac{\lambda_1}{\mu}$ and $\phi_2 = \frac{\lambda_2}{\mu}$ the two relative risks.

We give three different semi-conjugate families of prior distributions for this model. As we said in the introduction, the reference prior for a model with two parameters relies on the choice of a parameter of interest. When there are more than two parameters, it relies on the choice of an interest ordering between parameters. The first semi-conjugate family contains the reference prior when the ordering (from the main parameter of interest to the main nuisance parameter) is ϕ_1, ϕ_2, μ. The second one contains the reference prior when the ordering is ϕ_1/ϕ_2, ϕ_2, μ. The third one, up to a caveat, contains the reference prior when the ordering is (ϕ_1,ϕ_2), μ.

These semi-conjugate families are not implemented in **brr**. In fact, we will see that the second control group can be dropped when we are interested in ϕ_1 only. And the control group can be dropped when we are interested in ϕ_1/ϕ_2 only. These facts are in agreement with the intution. And because of them, we do not see some motivating reasons to implement these semi-conjugate families in **brr**.

We explain the derivation of the reference prior for the first family. However our derivation is a bit heuristic because we do not check some technical details. The derivation is the same and it is not more difficult for the second family and we do not show the derivation. The method is a bit less easy for the third family because the parameter of interest (ϕ_1, ϕ_2) is two-dimensional, and we will provide the derivationfor this family.

4.1. First semi-conjugate family: (ϕ_1, ϕ_2, μ) ordering

For positive numbers a, b, c_1, d_1, c_2, d_2, assign the independent prior distributions on μ and (ϕ_1, ϕ_2) given by

$$\mu \sim \mathscr{G}(a, b), \quad (\phi_2 \mid \phi_1) \sim \frac{\phi_1 S_1 + T + b}{S_2} \times \mathscr{B}'(c_2, d_2) \quad and \quad \phi_1 \sim \frac{T + b}{S_1} \times \mathscr{B}'(c_1, d_1).$$

Then the joint posterior on (μ, ϕ_2, ϕ_1) is given by

$$(\mu \mid \phi_1, \phi_2, x_1, x_2, y) \sim \mathscr{G}(a + x_1 + x_2 + y, \phi_1 S_1 + \phi_2 S_2 + T + b),$$

$$(\phi_2 \mid \phi_1, x_1, x_2, y) \sim \frac{\phi_1 S_1 + T + b}{S_2} \times \mathscr{B}'(c_2 + x_2, a + d_2 + x_1 + y) \quad and$$

$$(\phi_1 \mid x_1, x_2, y) \sim \frac{T + b}{S_1} \times \mathscr{B}'(c_1 + x_1, a + d_1 + y).$$

In particular :

- *when $a = c_1 = c_2 = 0.5$ and $b = d_1 = d_2 = 0$, the prior is the reference prior when ϕ_1 is parameter of first interest and ϕ_2 the parameter of second interest;*

- *when $a, b > 0$, $c_1 = c_2 = 0.5$ and $d_1 = d_2 = 0$, the prior is the semi-reference prior with information on μ when ϕ_1 is parameter of first interest and ϕ_2 the parameter of second interest;*

- *when $a, b, c_2, d_2 > 0$, $c_1 = 0.5$ and $d_1 = 0$, the prior is the semi-reference prior with information on μ and ϕ_2 (when, a fortioti, ϕ_1 is parameter of interest).*

Observe that the posterior distribution of ϕ_1 does not involve anything related to the second control group: it does not involve x_2, S_2, nor the prior parameters c_2 and d_2. In the non-informative case, it is the same as the posterior distribution that would be obtained with the two Poisson samples model by dropping the second control group. But the joint posterior of (μ, ϕ_1) would be different. The conditional posterior distribution $(\mu \mid \phi_1, x_1, x_2, y)$ belongs to the same family as the marginal posterior distribution

$(\mu \mid x, y)$ in the two Poisson samples model. The predictive distributions of x and y would be different too.

The reference prior is derived in parallel to the derivation of the semi-conjugate family as follows, in three steps. We do not provide the calculations. At each step, they are similar to the calculations for the case of only one control group, which are given in detail in [13].

- **First step: μ given (ϕ_1, ϕ_2).**

  The likelihood of μ conditional to ϕ_1 and ϕ_2 satisfies

  $$L(\mu \mid \phi_1, \phi_2, y, x_1, x_2) \propto \exp\{-\mu(\phi_1 S_1 + \phi_2 S_2 + T)\} \mu^{y+x_1+x_2},$$

  from which it is straightforward to see the conditional conjugate family $(\mu \mid \phi_1, \phi_2) \sim \mathscr{G}(a, b)$. The conditional reference prior is the Jeffreys prior given by the square root of the Fisher information associated to this conditional likelihood, which is the likelihood of the rate of a Poisson likelihood, and it is known that this Jeffreys prior is $\mathscr{G}(a, b)$ with $a = \frac{1}{2}$ and $b = 0$.

- **Second step: ϕ_2 given ϕ_1.**

  At this step, one consider the likelihood of ϕ_1 conditional to ϕ_2, obtained by integrating the original likelihood with respect to a $\mathscr{G}(a, b)$ distribution on μ. One gets, after easy calculations

  $$L(\phi_2 \mid \phi_1, y, x_1, x_2) \propto \frac{\phi_2^{x_2}}{(1 + r_1\phi_1 + r_2\phi_2)^{a+y+x_1+x_2}}$$

  with $r_i = \frac{S_i}{T+b}$. It is then straightforward to see that the conditional prior distribution $(\phi_2 \mid \phi_1) \sim \frac{\phi_1 S_1 + T + b}{S_2} \times \mathscr{B}'(c_2, d_2)$ has the announced conjugacy property.

  The corresponding Fisher information obtained after calculations is

  $$I(\phi_2 \mid \phi_1) \propto \frac{1}{\phi_2(1 + r_1\phi_1 + r_2\phi_2)} \propto \frac{\phi_2^{-1}}{1 + \frac{r_2}{r_1\phi_1+1}\phi_2},$$

  thereby giving the Jeffreys prior $(\phi_2 \mid \phi_1) \sim \frac{\phi_1 S_1 + T + b}{S_2} \times \mathscr{B}'(\frac{1}{2}, 0)$.

- **Third step: ϕ_1.**

  This time we integrate the previous likelihood with respect to the distribution of $(\phi_2 \mid \phi_1)$, and get after calculations

  $$L(\phi_1 \mid y, x_1, x_2) \propto \frac{\phi_1^{x_1}}{(1 + r_1\phi_1)^{a+y+x_1}}$$

  wherefrom it is straightfoward to see that the prior distribution $\phi_1 \sim \frac{T+b}{S_1} \times \mathscr{B}'(c_1, d_1)$ has the announced conjugacy property.

  The corresponding Fisher information is, after some calculations,

  $$I(\phi_1) \propto \frac{1}{\phi_1(1 + r_1\phi_1)},$$

18

yielding the Jeffreys prior $\phi_1 \sim \frac{T+b}{S_1} \times \mathcal{B}'(\frac{1}{2}, 0)$.

As we announced, our derivation of the reference prior is a bit heuristic. Firstly, we should check that at each step that the regularity conditions hold in order to take the Jeffreys prior. More importantly, the heuristic nature of our derivation mainly lies in the fact that at the second and third steps, we treat the conditional reference prior of the previous step as if it were a proper distribution. In order to properly do the derivation, we should use an increasing sequence of compact sets approximating the parameter space (see [5]).

4.2. Second semi-conjugate family: $(\phi_1/\phi_2, \phi_2, \mu)$ ordering

As we saw, there is no interest to use the previous semi-conjugate family if one is interested in ϕ_1 only. Posterior probabilities of hypotheses such as $\phi_1 > \phi_2$ or more generally $\phi_1 > k\phi_2$ are relevant quantities for comparing the two relative risks. But since $\phi_1 > k\phi_2$ is equivalent to $\frac{\phi_1}{\phi_2} > k$, one should consider the ratio $\rho = \frac{\phi_1}{\phi_2}$ as the parameter of interest if one is interested in such probabilities. The reference prior for the interest ordering (ρ, ϕ_2, μ) belongs to the following semi-conjugate family.

For positive numbers a, b, c_1, d_1, c_2, d_2, assign the independent prior distributions on μ and (ρ, ϕ_2) given by

$$\mu \sim \mathcal{G}(a, b), \quad (\phi_2 \mid \rho) \sim \frac{T+b}{\rho S_1 + S_2} \times \mathcal{B}'(c_2, d_2) \quad and \quad \rho \sim \frac{S_2}{S_1} \times \mathcal{B}'(c_1, d_1).$$

Then the joint posterior on (μ, ϕ_2, ρ) is given by

$$(\mu \mid \rho, \phi_2, x_1, x_2, y) \sim \mathcal{G}(a + x_1 + x_2 + y, \phi_1 S_1 + \phi_2 S_2 + T + b),$$

$$(\phi_2 \mid \rho_1, x_1, x_2, y) \sim \frac{T+b}{S_2 + \rho S_1} \times \mathcal{B}'(c_2 + x_1 + x_2, a + d_2 + y) \quad and$$

$$(\rho \mid x_1, x_2, y) \sim \frac{S_2}{S_1} \times \mathcal{B}'(c_1 + x_1, c_2 + d_1 + x_1).$$

In particular :

- *when $a = c_1 = c_2 = 0.5$ and $b = d_1 = d_2 = 0$, the prior is the reference prior when ρ is parameter of first interest and ϕ_2 the parameter of second interest;*

- *when $a, b > 0$, $c_1 = c_2 = 0.5$ and $d_1 = d_2 = 0$, the prior is the semi-reference prior with information on μ when ρ is parameter of first interest and ϕ_2 the parameter of second interest;*

- *when $a, b, c_2, d_2 > 0$, $c_1 = 0.5$ and $d_1 = 0$, the prior is the semi-reference prior with information on μ and ϕ_2 (when, a fortiori, ρ is parameter of interest).*

For this semi-conjugate family, the posterior distribution of ρ does not involve anything related to the treated group: it does not involve y, T, nor the prior parameters a and b.

4.3. Third semi-conjugate family: $((\phi_1, \phi_2), \mu)$ ordering

In the third semi-conjugate family, the prior and posterior distributions of (ϕ_1, ϕ_2) is a bivariate Beta distribution of the second kind. Say that a pair of random variables (ϕ_1, ϕ_2) has the bivariate Beta distribution of the second kind $\mathscr{B}_2(c_1, c_2, d; \tau_1, \tau_2)$ with shape parameters c_1, c_2, d, and scale parameters τ_1 and τ_2, when it has the distribution of $\left(\tau_1 \frac{G_1}{G_0}, \tau_2 \frac{G_2}{G_0}\right)$ where G_0, G_1, G_2 are independent random variables with laws $G_0 \sim \mathscr{G}(1, d)$, $G_1 \sim \mathscr{G}(1, c_1)$, $G_2 \sim \mathscr{G}(1, c_2)$. Then the marginals distributions are $\phi_1 \sim \mathscr{B}_2(c_1, d, \tau_1)$ and $\phi_2 \sim \mathscr{B}_2(c_2, d, \tau_2)$, and the conditional distribution of ϕ_2 given ϕ_1 is $(\phi_2 \mid \phi_1) \sim \mathscr{B}_2\left(c_2, c_1 + d, \tau_2\left(1 + \frac{\phi_1}{\tau_1}\right)\right)$.

The third semi-conjugate family is the following one.

For positive numbers a, b, c_1, d_1, c_2, d_2, assign the independent prior distributions on μ and (ϕ_1, ϕ_2) given by

$$\mu \sim \mathscr{G}(a, b), \quad (\phi_1, \phi_2) \sim \mathscr{B}_2\left(c_1, c_2, d; \frac{T+b}{S_1}, \frac{T+b}{S_2}\right)$$

Then the joint posterior on (μ, ϕ_1, ϕ_1) is given by

$$(\mu \mid \phi_1, \phi_2, x_1, x_2, y) \sim \mathscr{G}(a + x_1 + x_2 + y, \phi S_1 + \phi_2 S_2 + T + b),$$

$$(\phi_1, \phi_2 \mid x_1, x_2, y) \sim \mathscr{B}_2\left(c_1 + x_1, c_2 + x_2, d + a + y; \frac{T+b}{S_1}, \frac{T+b}{S_2}\right)$$

There is a permissibility problem in the derivation of the reference prior using the pair (ϕ_1, ϕ_2) as the parameter of interest. Taking the square root of the determinant of the Fisher information $I(\phi_1, \phi_2)$ as the Jeffreys prior, we find a non-permissible improper prior distribution, in the sense that the formal application of Bayes' formula with this prior distribution possibly yields an improper posterior distribution.

We provide some details. After integrating the likelihood with respect to the $\mathscr{G}(a, b)$ of μ, one gets the following likelihood of (ϕ_1, ϕ_2):

$$L(\phi_1, \phi_2 \mid y, x_1, x_2) \propto \frac{\phi_1^{x_1} \phi_2^{x_2}}{(1 + r_1\phi_1 + r_2\phi_2)^{a+y+x_1+x_2}}$$

with $r_i = \frac{S_i}{T+b}$. In fact we can generalize without cost to the case of more than two control groups. The likelihood would be

$$L(\phi_1, \phi_2, \ldots \mid y, x_1, x_2, \ldots) \propto \frac{\prod_k \phi_k^{x_k}}{(1 + \sum r_k\phi_k)^{a+y+\sum x_k}},$$

and the following calculations are given for this more general case.

After calculations, the Fisher information matrix $I(\phi_1, \phi_2, \ldots)$ is found to be proportional to $(1 + \sum r_k\phi_k)^{-1} M$ where the (i, j)-th entry of the matrix M is

$$M_{i,j} = \begin{cases} \dfrac{S_i\left[1 + \sum_{k \neq i} r_k\phi_k\right]}{\phi_i} & \text{if } i = j \\ -S_i S_j & \text{if } i \neq j. \end{cases}$$

The Jeffreys prior is the square root of the determinant of the Fisher information matrix, and this determinant is given by the following lemma.

Lemma 1. *Let n be a positive integer and let $a_1, \ldots, a_n$, $\lambda_1, \ldots, \lambda_n$ and K be positive real numbers. Let $M = (m_{i,j})$ be the $n \times n$-matrix with diagonal entries*

$$m_{i,i} = \frac{a_i}{\lambda_i}\left(K + \sum_{j \neq i} a_j \lambda_j\right)$$

and off-diagonal entries $m_{i,j} = -a_i a_j$, $i \neq j$. Then one has

$$\det M = \frac{K\left(K + \sum_{i=1}^{n} a_j \lambda_j\right)^{n-1}}{\prod_{i=1}^{n} \lambda_i} \prod_{i=1}^{n} a_i.$$

Proof. One has $M = D - A$ where the matrix A has (i,j)-th entry $a_i a_j$, and D is the diagonal matrix with diagonal $(d_1, \ldots, d_n)$ where

$$d_i = \left(K + \sum_{j=1}^{n} a_j \lambda_j\right)\frac{a_i}{\lambda_i}.$$

The (i,j)-th entry of the matrix $H := AD^{-1}$ is $h_{i,j} = a_i a_j d_j^{-1}$. This implies by standard linear algebra that the characteristic polynomial of H is $X^{n-1}(X - \sum_{i=1}^{n} h_{i,i})$. Therefore

$$\det M = \det(I - AD^{-1})\det D = \left(1 - \sum_{i=1}^{n} h_{i,i}\right)\prod_{i=1}^{n} d_i,$$

and the announced result follows. □

Applying this lemma in the case of two treated groups gives the Jeffreys prior

$$(\phi_1, \phi_2) \sim \mathscr{B}_2\left(c_1 = \tfrac{1}{2}, c_2 = \tfrac{1}{2}, d = -\tfrac{1}{2}; \frac{T+b}{S_1}, \frac{T+b}{S_2}\right).$$

The problem with this improper prior is that it yields, through a formal application of Bayes' formula, a possibly improper posterior distribution: this happens in the case when $a + y \leq \tfrac{1}{2}$, which occurs when $a = \tfrac{1}{2}$ and one observes $y = 0$. For more than two groups, this is even worse, because $d = -\frac{n-1}{2}$ in the general case, where n is the number of control groups.

A. Special distributions

Some distributions related to the semi-conjugate are not standard. They are presented in this section, but we do not give every detail about them. For more details, the reader is referred to [14] and the help provided in the **brr** package. This section is intended to say a word about their implementation, and also to help the reading of the article [14] and to provide a complement of this article.

These distributions involve some hypergeometric functions. The **brr** package uses the implementations of these functions in the **gsl** package when they are available. We use the Pochhammer symbol $(a)_x$ in this section, defined by

$$(a)_x = \frac{\Gamma(a+x)}{\Gamma(a)},$$

which is the ascending factorial when x is a positive integer:

$$(a)_x = a(a+1)\dots(a+x-1).$$

The **brr** package also uses the **gsl** package for Γ-related functions such as the Pochhammer symbol.

A.1. Continuous distributions

The prior and posterior distributions on λ as well as the posterior distribution on μ involve the Tricomi hypergeometric function U defined by

$$U(\alpha, \beta, z) = \frac{1}{\Gamma(\alpha)} \int_0^\infty e^{-zt} t^{\alpha-1} (1+t)^{\beta-\alpha-1}\, dt.$$

This function has name **hyperg_U** in the **gsl** package.

The prior distribution on λ is called a Gamma-Beta2 distribution in the **brr** package, where Beta2 denotes the Beta distribution of second kind. The Gamma-Beta2 distribution is a mixture distribution of Gamma distributions with a Beta2 mixing distribution on the rate parameter. If a is the shape of the Gamma distribution, c and d the shape parameters of the Beta2 distribution and τ its scale parameter, then its density as a function of $x > 0$ is

$$\frac{(a)_\alpha}{B(d,c)} \tau^a x^{a-1} U(a+c, a-d+1, \tau x).$$

Its cumulative distribution function involves the Meijer G-function because of the identity

$$\int_0^y x^{a-1} U(\alpha, \beta, x)\, dx = \frac{1}{\Gamma(\alpha)\Gamma(\alpha-\beta+1)} G_{2,3}^{2,2} \left(y \,\Big|\, \begin{matrix} 1, a-\alpha+1 \\ a, a-\beta+1, 0 \end{matrix} \right).$$

Meijer G-functions are not available in R. It is possible to express the Meijer function $G_{2,3}^{2,2}$ with the help of the ${}_2F_2$ hypergeometric function. The **pGB2** function in the **brr** uses this fact to evaluate the cdf by calling the **genhypergeo** function of the **hypergeo** package. The evaluation fails for many cases.

An implementation of Meijer G-functions is provided in Pyhton by the **mpmath** package. Below we give a R function to evaluate the integral above by calling Python with the **rPython** package. It also uses **lngamma** from the **gsl** package.

```
> library(rPython); library(gsl)
> python.exec("from mpmath import *")
> python.exec( "def I(a,alpha,beta,x):
+                return float(meijerg(
+                    [[1,-alpha+a+1],[]],
+                    [[a,-beta+a+1],[0]],
+                    x))" )
> pGB2_python <- function(q, a, c, d, tau){
+    exp(lngamma(c)-lngamma(d)-lngamma(a))*
+      python.call("I", a, a+c, a-d+1, tau*q)
+ }
> # example:
> a <- 3; c <- 2; d <- 2.5; tau <- 2
> pGB2_python(1, a, c, d, tau)

[1] 0.3339994

> pGB2(1, a, c, d, tau)

[1] 0.3339994
```

The posterior distribution of λ and the posterior distribution of μ belong to the family of so-called Gamma-Inverse Beta distributions in **brr**. The pdf of these distributions as a function of $x > 0$ have the following form:

$$\frac{\rho^a}{\Gamma(a)} \frac{\Gamma(\alpha + \beta)}{\Gamma(\beta)} x^{a-1} e^{-\rho z} U(\alpha, a - \beta + 1, \rho x)$$

The cdf of these distributions involve the Meijer function $G_{2,1}^{2;3}$ because of the equality

$$\int_0^y \exp(-x) x^{a-1} U(\alpha, \beta, x)\, dx = G_{2,3}^{2,1} \left(y \left| \begin{array}{l} 1, a + \alpha - \beta + 1 \\ a, a - \beta + 1, 0 \end{array} \right. \right).$$

Below we give a R function to evaluate the integral above by calling Python with the **rPython** package.

```
> python.exec( "def J(a,alpha,beta,x):
+                return float(meijerg([[1],
+                    [alpha-beta+a+1]],
+                    [[a,-beta+a+1],[0]],
+                    x))" )
> pGIB_python <- function(x,a,alpha,beta,rho=1){
+    exp(lnpoch(beta,alpha)-lngamma(a))*x^(a-1)*
+      python.call("J", a, alpha, a-beta+1, rho*x)
+ }
> a <- 1; alpha <- 2; beta <- 9; rho <- 2
> pGIB_python(1,a,alpha,beta,rho)
```

```
[1] 0.9120724

> ecdf(rGIB(1e6,a,alpha,beta,rho))(1)

[1] 0.912312
```

A.2. Discrete distributions

The conditional prior predictive distribution of x given y is a Beta-Negative binomial distribution. Its probability masses are given by

$$p(x \mid y) = \frac{(a+y)_x}{x!} \frac{B(c+x, a+y+d)}{B(c,d)}.$$

This distribution is available as a generalized hypergeometric distribution in the **SuppDists** package. By noting that

$$p(x \mid y) = \frac{(a)_x}{x!} \frac{B(c+x, a+d)}{B(c,d)} \frac{(a+x)_y (a+d)_y}{(a)_y (c+x+a+d)_y}$$

and knowing that

$$p(y) = \frac{(a)_y}{y!} \frac{b^a T^y}{(b+T)^{a+y}},$$

one gets by summation the prior predictive distribution of x:

$$p(x) = \left(\frac{b}{b+T} \right)^a \frac{B(c+x, a+d)}{B(c,d)} \frac{(a)_x}{x!} {}_2F_1 \left(a+x, a+d, c+x+a+d, \frac{T}{b+T} \right).$$

where ${}_2F_1$ is the Gauss hypergeometric function. The **brr** package uses the implementation of ${}_2F_1$ offered by the **gsl** package.

The Gauss hypergeometric function is also involved in the posterior predictive distributions of x and y.

These special discrete distributions are presented in [14]. The **brr** implementations of the cumulative distribution functions simply perform the sum of the probability masses.

B. Intrinsic discrepancy for the marginal model

Set $\psi = \frac{S}{T+b} \phi$ to lighten the notation. The marginal likelihood $L_{a,b}(\psi \mid x, y)$ is such that (see [13] for calculations if needed)

$$L_{a,b}(\psi \mid x, y) \underset{\psi}{\propto} \frac{\psi^x}{(\psi+1)^{x+y+a}}.$$

Therefore

$$\log \frac{L_{a,b}(x, y \mid \psi)}{L_{a,b}(x, y \mid \psi_0)} = xM + (y + a)N$$

with

$$M = M(\phi, \phi_0) = \log \frac{\psi}{\psi_0} + \log \frac{\psi_0 + 1}{\psi + 1},$$

and

$$N = N(\phi, \phi_0) = \log \frac{\psi_0 + 1}{\psi + 1}.$$

The expectations of x and y in the marginal model are

$$\mathbb{E}_\phi(x) = \phi S \times \frac{a}{b} \qquad \text{and} \qquad \mathbb{E}_\phi(y) = T \times \frac{a}{b}.$$

One gets the Kullback-Leibler divergence

$$\begin{aligned}
\delta(\phi \mid \phi_0) &= \mathbb{E}_\phi \left[\log \frac{L_{a,b}(x, y \mid \psi)}{L_{a,b}(x, y \mid \psi_0)} \right] \\
&= \frac{a}{b} \left(M \phi S + (T + b)N \right) \\
&= \frac{a}{b}(T + b)(M\psi + N).
\end{aligned}$$

The *intrinsic discrepancy* is defined by

$$\ell(\phi_0, \phi) := \min\{\delta(\phi \mid \phi_0), \delta(\phi_0 \mid \phi)\}$$

and one can prove that

$$\ell(\phi_0, \phi) = \begin{cases} \delta(\phi \mid \phi_0) & \text{if } \phi \leq \phi_0 \\ \delta(\phi_0 \mid \phi) & \text{if } \phi \geq \phi_0. \end{cases}$$

Indeed, remark that $M(\phi, \phi_0) = -M(\phi_0, \phi)$ and $N(\phi, \phi_0) = -N(\phi_0, \phi)$, hence

$$\delta(\phi \mid \phi_0) - \delta(\phi_0 \mid \phi) = \frac{a}{b}(T + b)\left((\psi + \psi_0)M + 2N\right)$$

and

$$\begin{aligned}
\frac{d}{d\psi_0}\left((\psi + \psi_0)M + 2N\right) &= \psi\left(\frac{1}{\psi_0 + 1} - \frac{1}{\psi_0}\right) + \psi_0\left(\frac{1}{\psi_0 + 1} - \frac{1}{\psi_0}\right) + M + \frac{2}{\psi_0 + 1} \\
&= \frac{\psi_0 - \psi}{\psi_0(\psi_0 + 1)} + M.
\end{aligned}$$

Because of

$$M(\phi, \phi_0) = \log\left(\frac{\psi}{\psi_0}\frac{\psi_0 + 1}{\psi + 1}\right) = \log\left(1 + \frac{\psi - \psi_0}{\psi_0(\psi + 1)}\right),$$

the inequality $\phi \geq \phi_0$ holds if and only if $M(\phi, \phi_0) \leq \frac{\psi - \psi_0}{\psi_0(\psi + 1)}$, and this is equivalent to

$$\frac{d}{d\psi_0} \left((\psi + \psi_0)M + 2N \right) \leq \frac{\psi_0 - \psi}{\psi_0(\psi_0 + 1)} + \frac{\psi - \psi_0}{\psi_0(\psi + 1)}$$

$$= -\frac{(\psi - \psi_0)^2}{\psi_0(\psi_0 + 1)(\psi + 1)} \leq 0.$$

References

[1] Bernardo, J.M., 1979. Reference posterior distributions for Bayesian inference (with discussion). J. R. Statist. Soc., B, 41, 113–148.

[2] Berger, J.O., Bernardo, J.M., 1992b. On the development of reference priors. Bayesian Statistics 4 (Bernardo, J.M., Berger, J.O., Dawid, A.P., and Smith, A.F.M., eds.) Oxford: University Press, 35–60 (with discussion).

[3] Bernardo JM and Smith AFM. 1994. *Bayesian Theory*. Chichester: Wiley.

[4] Berger JO and Sun D. 1998. Reference priors with partial information. *Biometrika* 85(1):55–71.

[5] J. M. Bernardo, J. M. Ramon (1998). An introduction to Bayesian reference analysis: inference on the ratio of multinomial parameters. *J. Roy. Statist. Soc. D*, **47**, 101–135.

[6] Bernardo JM and Rueda R. 2002. Bayesian hypothesis testing: A reference approach. *Internat. Statist. Rev.* 70:351–372.

[7] Bernardo JM and Juarez MA. 2003. Intrinsic estimation. In *Bayesian Statistics* 7, eds. Bernardo JM, Bayarri MJ, Berger JO, Dawid AP, Heckerman D, Smith AFM and West M, 465–476. Oxford: Oxford University Press.

[8] Bernardo JM. 2005a. Reference Analysis. In *Handbook of Statistics* 25, eds. Dey DK and Rao CR, 17–90. Amsterdam: Elsevier.

[9] Bernardo JM. 2005b. Intrinsic Credible Regions: an Objective Bayesian approach to Interval Estimation (with discussion). *Test* 14:317–384.

[10] Bernardo JM. 2009. Modern Bayesian Inference: Foundations and Objective Methods. In *Philosophy of Statistics*, eds. Bandyopadhyay P and Forster M. Amsterdam: Elsevier.

[11] Bernardo JM. 2010. Integrated objective Bayesian Estimation and hypothesis testing (with discussion). In *Bayesian Statistics* 9, eds. Bernardo JM, Bayarri MJ, Berger JO, Dawid AP, Heckerman D, Smith AFM and West M. Oxford: Oxford University Press.

[12] R. K. S. Hankin 2006. Special functions in R: introducing the gsl package. R News 6(4)

[13] Laurent S and Legrand C. 2011. A Bayesian framework for the ratio of two Poisson rates in the context of vaccine efficacy trials. ESAIM, Probability & Statistics 16 (2012), 375–398.

[14] Laurent S. 2012. *Some Poisson mixtures distributions with a hyperscale parameter*. Brazilian Journal of Probability and Statistics 26 (2012), 265–278.

[15] Laurent S. (2012). Intrinsic Bayesian inference on a Poisson rate and on the ratio of two Poisson rates. Journal of Statistical Planning and Inference 142 (2012), 2656–2671.

[16] R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.