Bayesians and Frequentists have long been ambivalent toward each other. The concept of “Prior” remains the center of this 250 years old tug-of-war: frequentists view prior as a weakness that can cloud the final inference, whereas Bayesians view it as a strength to incorporate expert knowledge into the data analysis. So, the question naturally arises, how can we develop a Bayes-frequentist consolidated data analysis workflow that enjoys the best of both worlds?

To develop a “defendable and defensible” Bayesian learning model, we have to go beyond blindly ‘turning the crank’ based on a “go-as-you-like” [approximate guess] prior. A lackluster attitude towards prior modeling could lead to disastrous inference, impacting various fields from clinical drug development to presidential election forecasts. The real questions are: How can we uncover the blind spots of the conventional wisdom-based prior? How can we develop the science of prior model-building that combines both data and science [DS-prior] in a *testable* manner – a double-yolk Bayesian egg? Unfortunately, these questions are outside the scope of business-as-usual Bayesian modus operandi and require new ideas. In the following, we demonstrate how to prepare the “Bayesian omelet” — the operational part — using the R package BayesGOF.

Our model-building approach proceeds sequentially as follows:

1. it starts with a scientific (or empirical) parametric prior \(g(\theta;\alpha,\beta)\),
2. inspects the adequacy and the remaining uncertainty of the elicited prior using a graphical exploratory tool,
3. estimates the necessary “correction” for assumed \(g\) by looking at the data,
4. generates the final statistical estimate \(\hat \pi(\theta)\), and
5. executes macro and micro-level inference.

Our algorithmic solution yields answers to all five of the phases using *one single* algorithm, which we will now demonstrate for rat tumor data. The rat tumor data consists of observations of endometrial stromal polyp incidence in \(k=70\) groups of rats. For each group, \(y\_i\) is the number of rats with polyps and \(n\_i\) is the total number of rats in the experiment. The dataset is available in the R package BayesGOF.

The Rat-data model: \(y\_i\,\overset{{\rm ind}}{\sim}\,\mbox{Binomial}( n\_i, \theta\_i)\), \(i=1,\ldots,k\), where the unobserved parameters \(\theta=(\theta\_1,\ldots,\theta\_k)\) are independent realizations from the *unknown* \(\pi(\theta)\).

**Step 1.** We begin by finding the starting parameter values for parametric conjugate \(g \sim Beta(\alpha, \beta)\):

library(BayesGOF)

set.seed(8697)

data(rat)

###Use MLE to determine starting values

rat.start <- gMLE.bb(rat$y, rat$n)$estimate

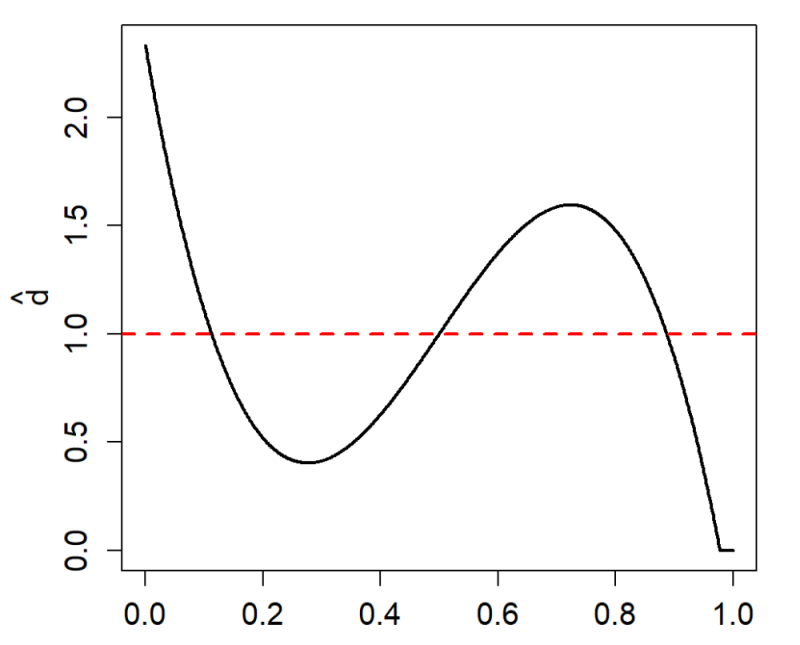
We use our starting parameter values to run the main DS.prior function:

rat.ds <- DS.prior(rat, max.m = 6, rat.start, family = "Binomial")

Next we will discuss how to interpret and use this rat.ds object for exploratory Bayes modeling and prior uncertainty quantification.

**Step 2.** We display the U-function to quantify and characterize the uncertainty of the a priori selected \(g\):

plot(rat.ds, plot.type = "Ufunc")



The deviations from the uniform distribution (the red dashed line) indicates that our initial selection for \(g\), \(\text{Beta}(\alpha = 2.3,\beta = 14.1)\), is incompatible with the observed data and requires repair; the data indicate that there are, in fact, two different groups of incidence in the rats.

**Step 3a.** Extract the parameters for the nonparametrically *corrected* prior \(\hat{\pi}\):

rat.ds

## $g.par

## alpha beta

## 2.304768 14.079707

##

## $LP.coef

## LP1 LP2 LP3

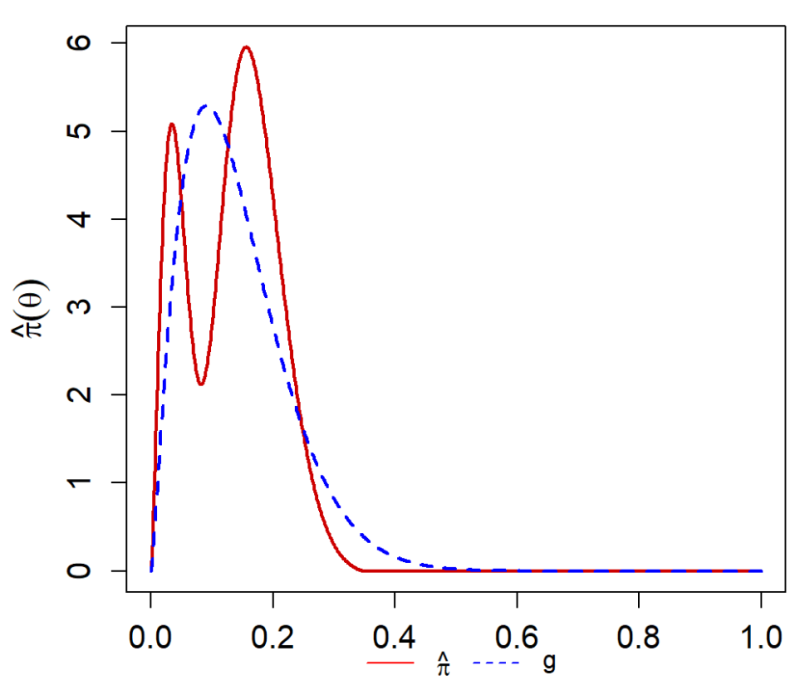
## 0.0000000 0.0000000 -0.5040361

Therefore, our estimated DS(G.m) prior is given by: \[\hat{\pi}(\theta) = g(\theta; \alpha,\beta)\Big[1 – 0.52T\_3(\theta;G) \Big].\]

The DS-prior has a unique two-component structure that combines parametric \(g\), and a nonparametric \(d\) (which we call the *U-function*). Here \(T\_j(\Theta;G)\), \(j = 1,\ldots,m\) are a specialized orthonormal basis given by \(\text{Leg}\_j[G(\Theta)]\), members of LP-class of rank-polynomials. Note that \({\rm DS}(G,m=0) \equiv g(\theta;\alpha,\beta)\). The truncation point \(m\) reflects the *concentration* of true unknown \(\pi\) around the pre-selected \(g\).

**Step 3b.** Plot the estimated DS prior \(\hat{\pi}\) along with the original parametric \(g\):

plot(rat.ds, plot.type = "DSg")



**MacroInference**

The term “MacroInference” aims to answer the following question: How to combine \(k\) binomial parameters to come up with an overall, macro-level aggregated statistical behavior of \(\theta\_1,\ldots,\theta\_k\)? This is often important in applied analysis, as the limited sample size of a single study hardly provides adequate evidence for a definitive conclusion.

**Step 4.** Here we are interested in the *overall* macro-level inference by combining the \(k=70\) parallel studies. The group-specific modes along with their SEs can be computed as follows:

rat.macro.md <- DS.macro.inf(rat.ds, num.modes = 2 , iters = 25, method = "mode")

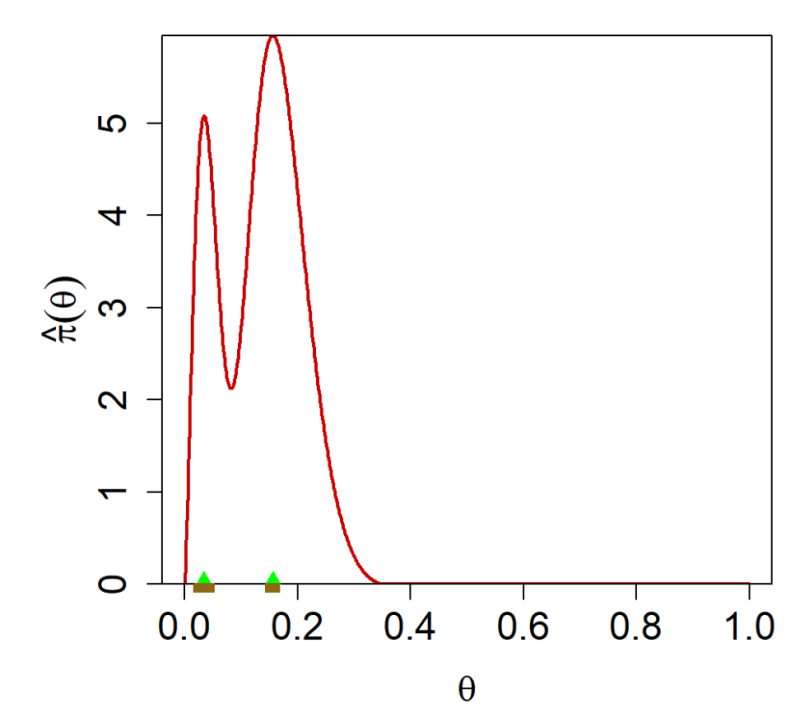
rat.macro.md

## 1SD Lower Limit Mode 1SD Upper Limit

## [1,] 0.0161 0.0340 0.0520

## [2,] 0.1442 0.1562 0.1681

plot(rat.macro.md)

[](http://a6.typepad.com/6a0105360ba1c6970c01b8d2def3ae970c-pi)

**MicroInference**

“Microinference” refers to the process of using information from historical studies to improve the estimates of one or more studies of particular interest. This is known as “borrowing strength” in Bayesian inference literature. It is noteworthy to mention that the classical Stein’s shrinkage does not work for rat data due to the presence of multiple partially exchangeable studies. Our adaptive (or selective) shrinkage technology selectively borrows strength from ‘similar’ experiments in an automated manner, by answering the important question: *where to shrink?*.

**Step 5.** In addition to the earlier \(k=70\) studies for the rat tumor data, we have a current experimental study that shows \(y\_{71}=4\) out of \(n\_{71}=14\) rats developed tumors. The following code performs the desired microinference for \(\theta\_{71}\) (posterior distribution along with its mean and mode):

rat.y71.micro <- DS.micro.inf(rat.ds, y.0 = 4, n.0 = 14)

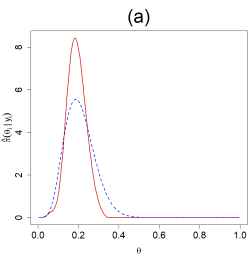
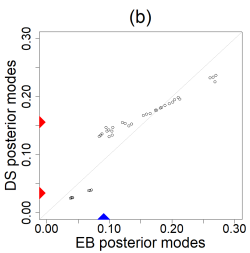
rat.y71.micro

## Posterior summary for y = 4, n = 14:

## Posterior Mean = 0.1897

## Posterior Mode = 0.1833

## Use plot(x) to generate posterior plot

[](http://a1.typepad.com/6a0105360ba1c6970c01b7c954a0f9970b-pi)[](http://a7.typepad.com/6a0105360ba1c6970c01b8d2def3cf970c-pi)

The left plot (a) compares the posterior distributions for the parametric \(g\) (blue) and the DS posterior (red). The right plot (b) compares our adaptive shrinkage with Stein’s estimates. The vertical red triangles indicate the modes of the DS prior, while the blue triangle is the mode of the parametric \(g\). For additional real-data examples, please see below:

**I. Illustration using rat tumor data (Binomial Family)**

The rat tumor data consists of observations of endometrial stromal polyp incidence in k=70k=70 groups of rats. For each group, yiyi is the number of rats with polyps and nini is the total number of rats in the experiment. Here we describe the analysis of rat tumor data using Bayes-DS(G,m)DS(G,m) modeling.

**Pre-Inferential Modeling**

**Step 1.** We begin by finding the starting parameter values for g∼Beta(α,β)g∼Beta(α,β) by MLE:

**library**(BayesGOF)

**set.seed**(8697)

**data**(rat)

*###Use MLE to determine starting values*

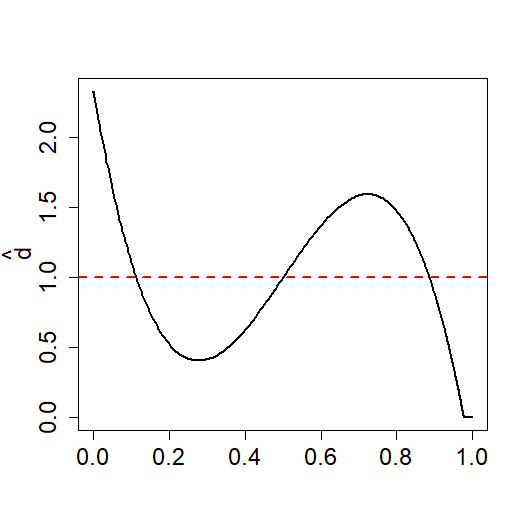
rat.start <- **gMLE.bb**(rat$y, rat$n)$estimate

We use our starting parameter values to run the main function:

rat.ds <- **DS.prior**(rat, max.m = 6, rat.start, family = "Binomial")

**Step 2.** We display the U-function to quantify and characterize the uncertainty of the a priori selected gg:

**plot**(rat.ds, plot.type = "Ufunc")



The deviations from the uniform distribution (the red dashed line) indicate that our initial selection for gg, Beta(α=2.3,β=14.1)Beta(α=2.3,β=14.1), is incompatible with the observed data and requires repair; the data show that there are, in fact, two different groups of incidence in the rats.

**Step 3a.** Extract the parameters for the *corrected* prior πˆπ^:

rat.ds

## $g.par

## alpha beta

## 2.304768 14.079707

##

## $LP.coef

## LP1 LP2 LP3

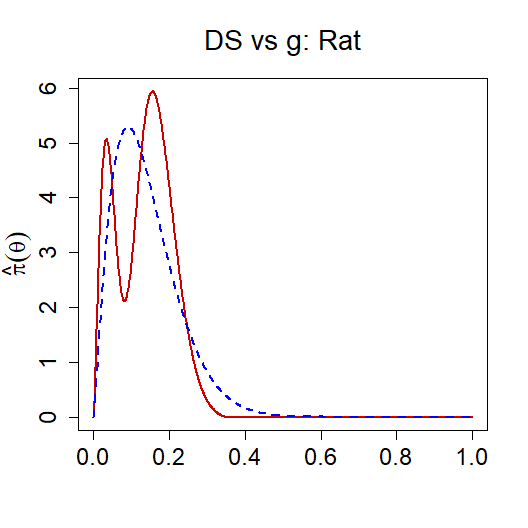
## 0.0000000 0.0000000 -0.5040361

Therefore, the DS prior πˆπ^ given gg is:

πˆ(θ)=g(θ;α,β)[1−0.52T3(θ;G)]π^(θ)=g(θ;α,β)[1−0.52T3(θ;G)]

**Step 3b.** We can now plot the estimated DS prior πˆπ^ along with the original parametric gg:

**plot**(rat.ds, plot.type = "DSg", main = "DS vs g: Rat")



**MacroInference**

**Step 4.** Here we are interested in the *overall* macro-level inference by combining the k=70k=70 parallel studies. The group-specific modes along with their standard errors are computed as follows:

rat.macro.md <- **DS.macro.inf**(rat.ds, num.modes = 2 , iters = 15, method = "mode")

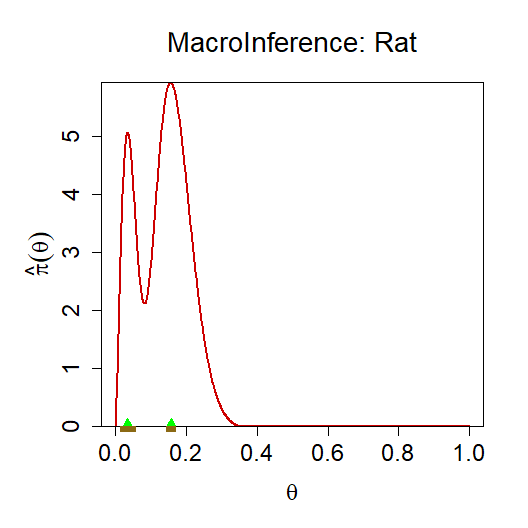
rat.macro.md

## 1SD Lower Limit Mode 1SD Upper Limit

## [1,] 0.0136 0.0340 0.0544

## [2,] 0.1431 0.1562 0.1692

**plot**(rat.macro.md, main = "MacroInference: Rat")



**MicroInference**

**Step 5.** Given an additional study θ71θ71 where y71=4y71=4 and n71=14n71=14, the goal is to estimate the probability of a tumor for this new clinical study. The following code performs the desired microinference to find the posterior distribution πLP(θ71|y71,n71)πLP(θ71|y71,n71) along with its mean and mode:

rat.y71.micro <- **DS.micro.inf**(rat.ds, y.0 = 4, n.0 = 14)

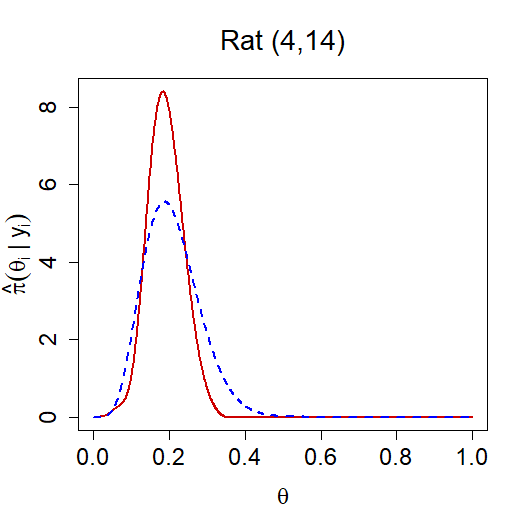
rat.y71.micro

## Posterior Mean = 0.1904

## Posterior Mode = 0.1832

## Use plot(x) to generate posterior plot

**plot**(rat.y71.micro, main = "Rat (4,14)")



**Finite Bayes**

The previous microinferece step ignores the randomness in the estimated hyperparameters (see step 3a). To take into account this extra variability (of the hyperparameter estimates in our DS(G,m) model) we perform finite Bayes adjustment to our microinference for θ71θ71. We also report the 90% Bayesian credible interval based on the finite-sample “corrected” posterior of θ71θ71.

rat.y71.FB <- **DS.Finite.Bayes**(rat.ds, y.0 = 4, n.0 = 14, iters = 15)

rat.y71.FB

## Posterior Mean = 0.193

## Posterior Mode = 0.1742

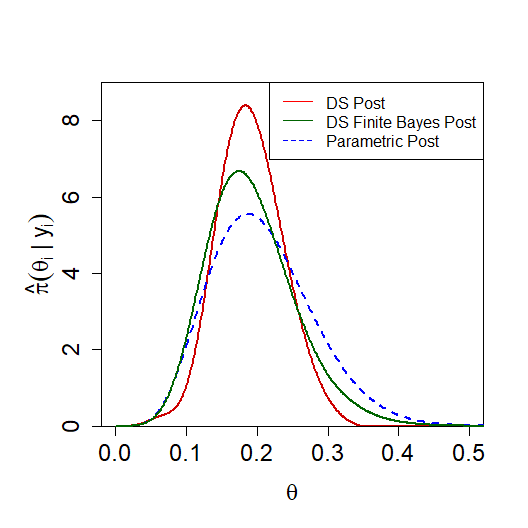
## Lower Bound Credible Interval = 0.1001

## Upper Bound Credible Interval = 0.3033

## Use plot(x) to generate finite Bayes plot

**plot**(rat.y71.FB, xlim = **c**(0,.5), ylim = **c**(0, 9))

**legend**("topright", **c**("DS Post", "DS Finite Bayes Post", "Parametric Post"), col = **c**("red", "darkgreen", "blue"), lty = **c**("solid","solid","dashed"), cex = 1)



**II. Comparison of**L2L2**and maximum entropy representations using galaxy data (Normal Family)**

This demonstration will compare the results using the default L2L2 representation of the estimated DS prior πˆπ^ to the maximum entropy π˘π˘ representation. We will use the galaxy data, which consists of k=324k=324 observed rotation velocities yiyi and their uncertainties of Low Surface Brightness (LSB) galaxies.

**Step 1.** We begin by finding the starting parameter values for g∼Normal(μ,τ2)g∼Normal(μ,τ2) by MLE:

**data**(galaxy)

gal.start <- **gMLE.nn**(galaxy$y, galaxy$se, method = "DL")$estimate

**Step 2.** Use our starting parameters to run the main function for two disinct cases:

* LP.type = “L2” estimates the DS prior in the L2L2 representation. This is the default case.
* LP.type = “MaxEnt” estimates the DS prior in terms of the maximum entropy representation.

gal.ds.L2 <- **DS.prior**(galaxy[,**c**(1,2)], max.m = 6, g.par = gal.start, family = "Normal", LP.type = "L2")

gal.ds.ME <- **DS.prior**(galaxy[, **c**(1,2)], max.m = 6, g.par = gal.start, family = "Normal", LP.type = "MaxEnt")

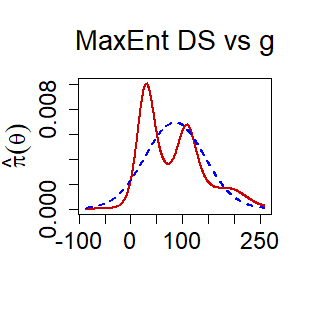
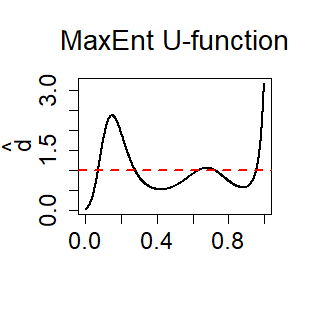
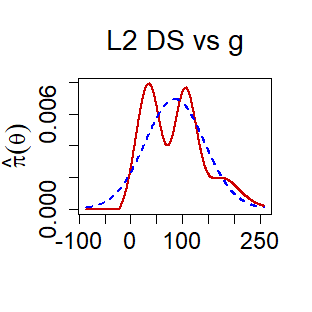
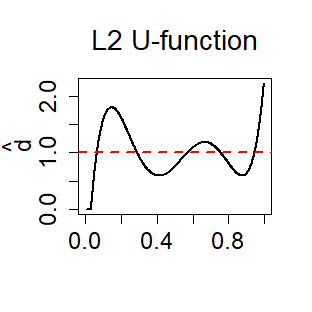
**Step 3.** We plot the U-functions and DS priors for both cases for comparison.

**plot**(gal.ds.L2, plot.type = "Ufunc", main = "L2 U-function")

**plot**(gal.ds.L2, plot.type = "DSg", main = "L2 DS vs g")

**plot**(gal.ds.ME, plot.type = "Ufunc", main = "MaxEnt U-function")

**plot**(gal.ds.ME, plot.type = "DSg", main = "MaxEnt DS vs g")



The U-function plots maintain a similar structure, but show some slight differences with respect to the two peaks. The maximum entropy representation shows the first peak as more narrow, while the second peak is not as pronounced as the L2L2 representation. We can see how those slight changes in the U-functions influence the resulting DS-prior in the “DS vs g” plots. The maximum entropy representation has a more significant first peak, smaller second peak, and smoother tails.

**Step 4.** Extract the parameters for the *corrected* priors: πˆπ^ and π˘π˘:

gal.ds.L2

## $g.par

## mu tau^2

## 85.51316 3304.41491

##

## $LP.coef

## LP1 LP2 LP3 LP4 LP5

## 0.0000000 0.0000000 0.2131113 -0.1967060 0.4069104

gal.ds.ME

## $g.par

## mu tau^2

## 85.51316 3304.41491

##

## $LP.coef

## LP(ME)0 LP(ME)1 LP(ME)2 LP(ME)3 LP(ME)4 LP(ME)5

## -0.1529718 0.0000000 0.0000000 0.2553139 -0.2765342 0.4560368

The DS prior in its L2L2 representation is:

πˆ(θ)=g(θ;μ,τ2)[1+0.21T3(θ;G)−0.20T4(θ;G)+0.41T5(θ;G)]π^(θ)=g(θ;μ,τ2)[1+0.21T3(θ;G)−0.20T4(θ;G)+0.41T5(θ;G)]

.

The DS prior in its maximum entropy representation is:

π˘(θ)=g(θ;μ,τ2)exp[−0.15+0.26T3(θ;G)−0.28T4(θ;G)+0.46T5(θ;G)]π˘(θ)=g(θ;μ,τ2)exp⁡[−0.15+0.26T3(θ;G)−0.28T4(θ;G)+0.46T5(θ;G)]

**III. Illustration using arsenic data (Normal Family)**

For this example, we will focus on the macroinference for the arsenic data set. The arsenic data set details the measurements of the level of arsenic in oyster tissue from k=28k=28 laboratories. We will use the maximum entropy representation for this example.

**Pre-Inferential Modeling**

**Step 1.** We begin by finding the starting parameter values for g∼Normal(μ,τ2)g∼Normal(μ,τ2) by MLE:

**data**(arsenic)

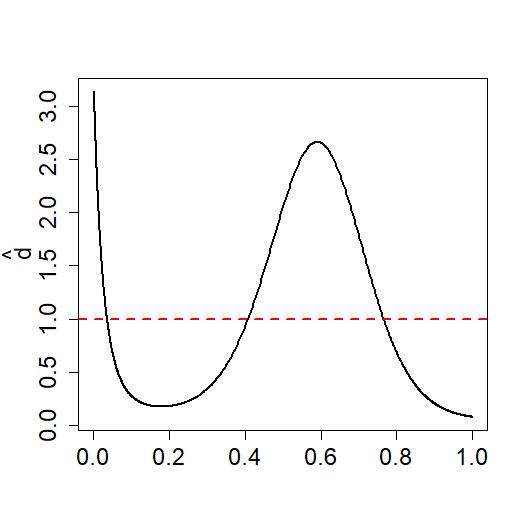
arsn.start <- **gMLE.nn**(arsenic$y, arsenic$se, method = "DL")$estimate

We use our starting parameter values to run the main DS.prior function:

arsn.ds <- **DS.prior**(arsenic, max.m = 4, arsn.start, family = "Normal", LP.type = "MaxEnt")

**Step 2.** We display the U-function to quantify and characterize the uncertainty of the a priori selected gg:

**plot**(arsn.ds, plot.type = "Ufunc")



**Step 3.** We now extract the parameters for the *corrected* prior π˘π˘ and plot it, along with the original gg:

arsn.ds

## $g.par

## mu tau^2

## 13.220522 3.407165

##

## $LP.coef

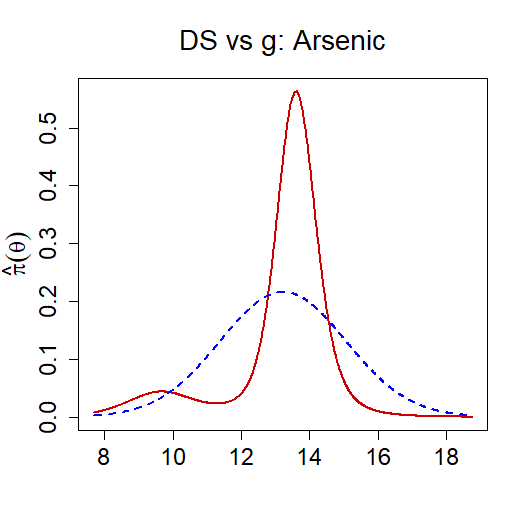
## LP(ME)0 LP(ME)1 LP(ME)2 LP(ME)3 LP(ME)4

## -0.4780663 0.0000000 -0.6503293 -0.7091040 0.4174869

The resulting equation for π˘(θ)π˘(θ) is:

π˘(θ)=g(θ;μ,τ2)exp[−0.48−0.65T2(θ;G)−0.71T3(θ;G)+0.42T4(θ;G)]π˘(θ)=g(θ;μ,τ2)exp⁡[−0.48−0.65T2(θ;G)−0.71T3(θ;G)+0.42T4(θ;G)]

**plot**(arsn.ds, plot.type = "DSg", main = "DS vs g: Arsenic")



**MacroInference**

**Step 4.** We now execute the macroinference to find a global estimate to summarize the k=28k=28 studies.

arsn.macro <- **DS.macro.inf**(arsn.ds, num.modes = 2, iters = 20, method = "mode")

arsn.macro

## 1SD Lower Limit Mode 1SD Upper Limit

## [1,] 8.4138 9.6674 10.9210

## [2,] 13.3589 13.6030 13.8471

Based on our results, we find two significant modes. Therefore, the prior shows structured heterogeneity and requires both modes to describe the distribution and its two groups. In this case, though, we are looking to estimate the consensus value of the measurand and its uncertainty. Therefore, we would select the dominant mode, which requires the macroinference results:

*#plot(arsn.macro, main = "MacroInference: Arsenic Data")*

**par**(mar=**c**(5,5.2,4,2)+0.3) *#changes left margin to make large labels fit*

**plot**(arsn.macro$prior.fit$theta.vals, arsn.macro$prior.fit$ds.prior,

type = "l", xlim = **c**(8,18.5),

lwd = 2, col = "red3",

xlab = **expression**(theta), ylab = "", font.main = 1,

cex.lab=1.45, cex.axis=1.5, cex.main= 1.75, cex.sub=1.5,

main = "MacroInference: Arsenic Data")

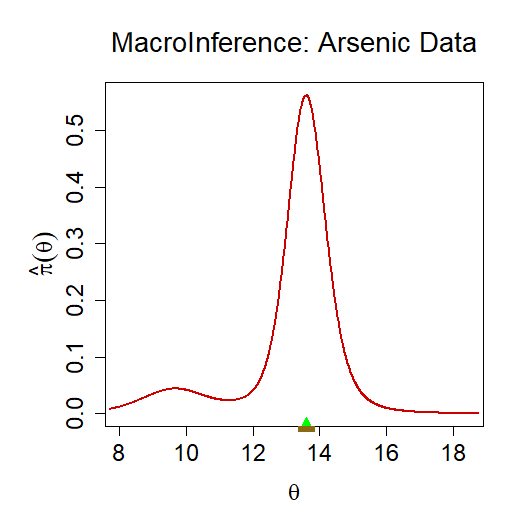
**title**(ylab = **expression**(**paste**(**hat**(pi)(theta))), line = 2.3, cex.lab=1.45)

**points**(arsn.macro$model.modes[2],-.02, col = "green", pch = 17, cex = 1.5)

**axis**(1, at=**seq**(arsn.macro$model.modes[2]-arsn.macro$mode.sd[2],

arsn.macro$model.modes[2]+arsn.macro$mode.sd[2],

length= (3-2) \* 20),tick=TRUE, col="goldenrod4", labels = F, tck=-0.015)



The plot shows the dominant mode and our consensus value for the measurand is 13.6 with a standard error (using smooth bootstrap) of 0.25. This mode is resistant to any extreme low measurements and achieves a slightly more accurate result than the standard PEB estimate (13.22±0.2513.22±0.25).

**IV. Illustration using child illness data (Poisson Family)**

The next example will conduct microinference on the child illness data. The child illness data comes from a study where researchers followed k=602k=602 pre-school children in north-east Thailand, recording the number of times (yiyi) a child became sick during every 2-week period for over three years. In particular, we want to compare posterior distributions for the number of children who became sick 1, 3, 5, and 10 times during a two week period. We will use the L2L2 representation for this example.

**Pre-Inferential Modeling**

**Step 1.** We begin by finding the starting parameter values for g∼Gamma(α,β)g∼Gamma(α,β) by MLE:

**data**(ChildIll)

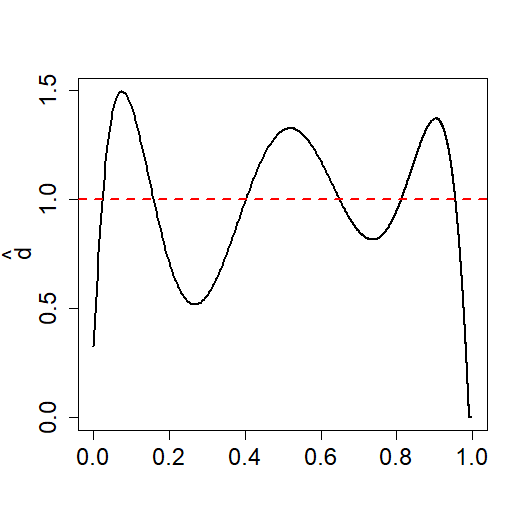
child.start <- **gMLE.pg**(ChildIll)

We use our starting parameter values to run the main DS.prior function for the Poisson family:

child.ds <- **DS.prior**(ChildIll, max.m = 8, child.start, family = "Poisson")

**Step 2.** We display the U-function to quantify and characterize the uncertainty of the selected gg:

**plot**(child.ds, plot.type = "Ufunc")



**Step 3.** We now extract the parameters for the *corrected* prior πˆπ^:

child.ds

## $g.par

## alpha beta

## 1.060878 4.193337

##

## $LP.coef

## LP1 LP2 LP3 LP4 LP5 LP6

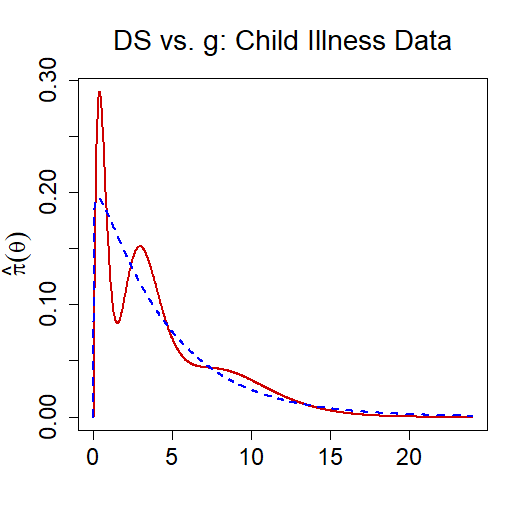
## 0.0000000 0.0000000 -0.1259159 0.0000000 0.0000000 -0.2797667

The DS prior πˆπ^ given gg is:

πˆ(θ)=g(θ;α,β)[1−0.13T3(θ;G)−0.28T6(θ;G)].π^(θ)=g(θ;α,β)[1−0.13T3(θ;G)−0.28T6(θ;G)].

We can plot πˆπ^, along with gg:

**plot**(child.ds, plot.type = "DSg", main = "DS vs. g: Child Illness Data")



**MicroInference**

**Step 4.** The plot shows some very interesting behavior in πˆπ^. We want to explore the posterior distributions for y=1,3,5,10y=1,3,5,10. For those results, we use the microinference functions.

child.micro.1 <- **DS.micro.inf**(child.ds, y.0 = 1)

child.micro.3 <- **DS.micro.inf**(child.ds, y.0 = 3)

child.micro.5 <- **DS.micro.inf**(child.ds, y.0 = 5)

child.micro.10 <- **DS.micro.inf**(child.ds, y.0 = 10)

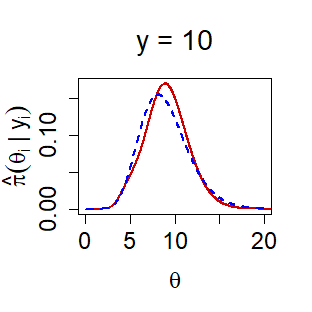
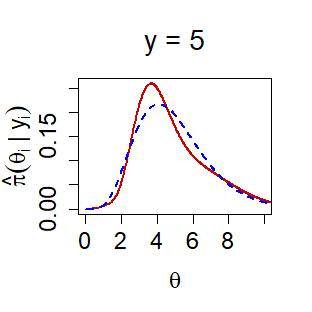
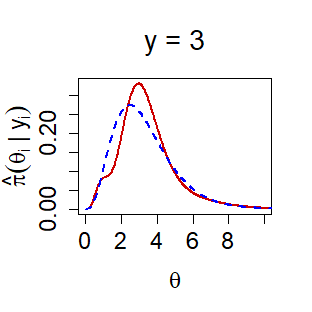
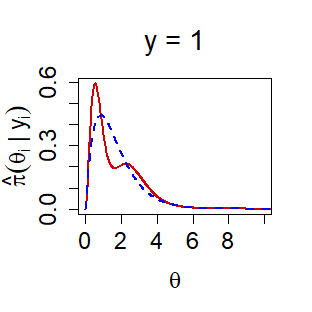
By plotting the posterior distributions we see how the distributions change based on the number of times a child is ill. The plots for each of the four microinferences are shown below.

**plot**(child.micro.1, xlim = **c**(0,10), main = "y = 1")

**plot**(child.micro.3, xlim = **c**(0,10), main = "y = 3")

**plot**(child.micro.5, xlim = **c**(0,10), main = "y = 5")

**plot**(child.micro.10, xlim = **c**(0,20), main = "y = 10")



**Conclusion**

All most all modern scientific research utilizes domain-knowledge and data to come up with breakthrough results. But the fundamental problem of how to fuse these “approximate” scientific prior knowledge with the data at hand is not a settled issue even 250 years after the discovery of the Bayes law. Bayesian modeling via goodness-of-fit technology, synthesized in the R package BayesGOF, allows us to determine a *scientific* prior that is consistent with the *data* at hand, in a systematic and principled way.