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, which is the goal of the paper, [“Bayesian Modeling via Goodness of Fit”](https://arxiv.org/pdf/1802.00474.pdf). In the following, we demonstrate how to prepare the “Bayesian omelet” — the operational part — using the R package [BayesGOF](https://cran.r-project.org/web/packages/BayesGOF/index.html).

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

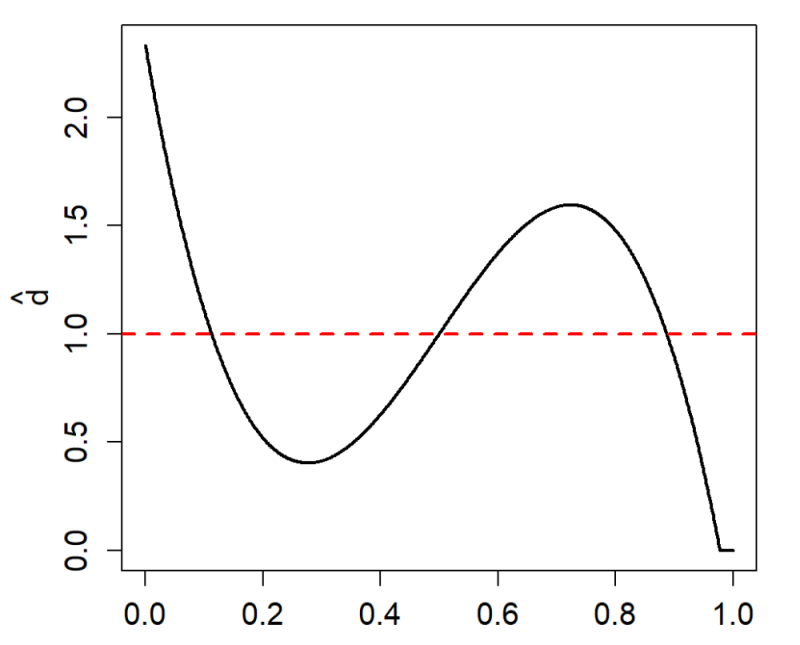
We use our starting parameter values to run the main DS.prior function:

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

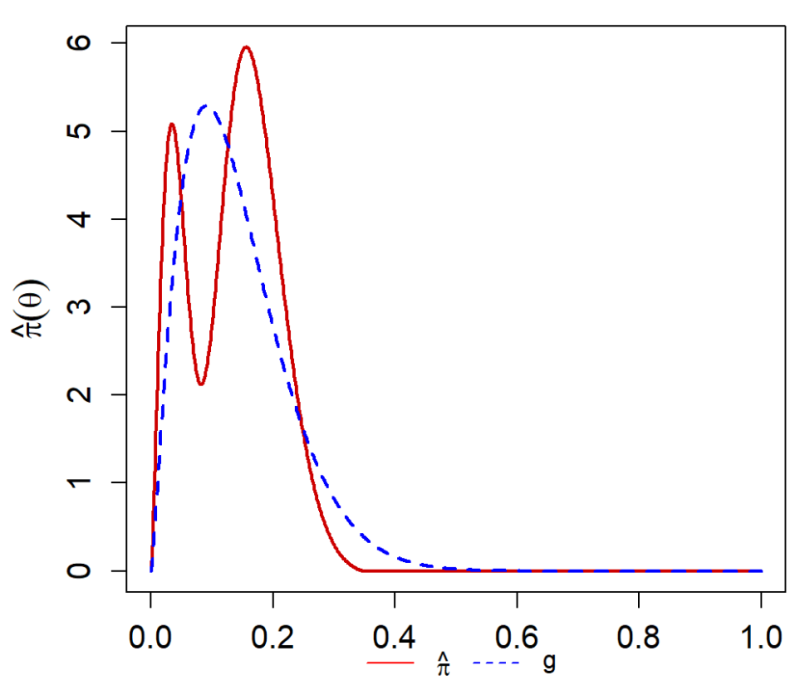
## 0.0000000 0.0000000 -0.5040361

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

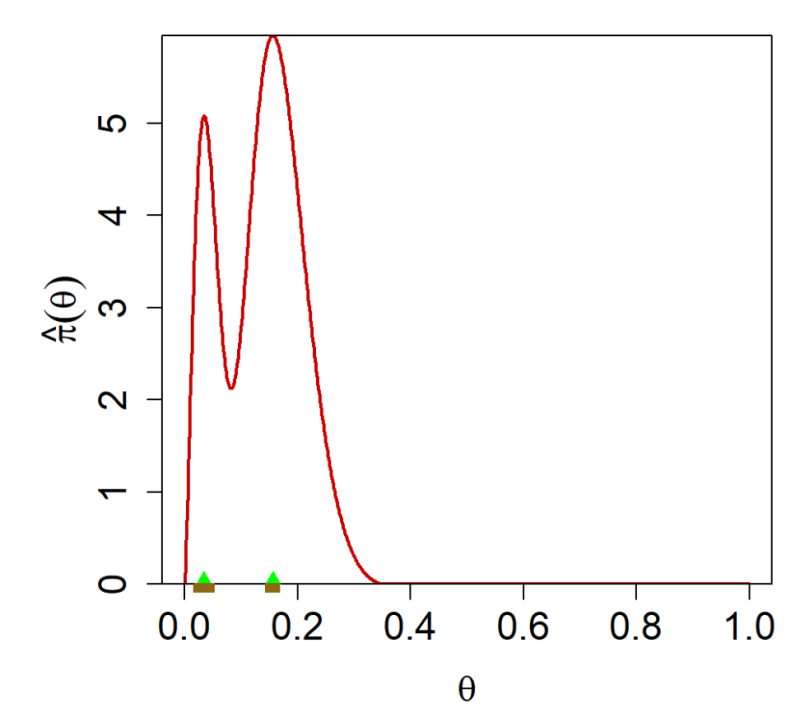
[rat.macro.md](http://rat.macro.md)

## 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://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)

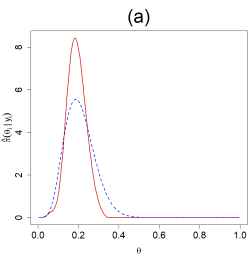
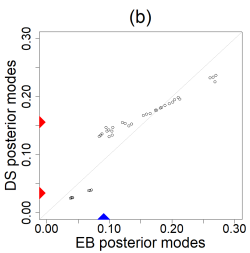
rat.y71.micro

## 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 the [package vignette](https://cran.r-project.org/web/packages/BayesGOF/vignettes/vignette_BayesGOF.html).

**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 (for more details, see [“Bayes’ Theorem in the 21st Century”](http://science.sciencemag.org/content/340/6137/1177) and [“Statistical thinking for 21st century scientists”](http://advances.sciencemag.org/content/3/6/e1700768.full)). Bayesian modeling via goodness-of-fit technology, synthesized in the R package [BayesGOF](https://cran.r-project.org/web/packages/BayesGOF/index.html), allows us to determine a *scientific* prior that is consistent with the *data* at hand, in a systematic and principled way.