Overview of Modern Bayesian Statistical Methods

(actually, five vignettes of modern Bayesian methods and modern takes on old-fashioned Bayesian methods)

Jim Berger, Duke University

Statistics and Exoplanets
Honolulu
August 3, 2015

Outline

- Bayesian model uncertainty
- Multiple testing
- Hiearchical modeling
- Bayes and big data
- A new Bayesian/frequentist synthesis in testing

Bayesian Model Uncertainty

Formulation

- Let $\mathcal{M}_l, l = 1, \ldots, k$, denote the models under consideration.
  - Example: $\mathcal{M}_l$ is the model that there are l planets in a star system.
- Data $\boldsymbol{X}$ has density $f_l(\boldsymbol{x} \mid \boldsymbol{\theta}_l)$ under M_l, with $\boldsymbol{\theta}_l$ unknown.
  - Example: $\boldsymbol{X}$ could be (e.g.) radial velocity measurements or transit data, and $\boldsymbol{\theta}_l$ would be the relevant system and planet parameters.
- Determine prior distributions $\pi_l(\boldsymbol{\theta}_l)$ for the parameters.
  - Example: As more is learned about exoplanets and star systems, these evolve from *objective* (vague) priors to *evidence-based* ('subjective' is not really an appropriate name) proper priors.
- Compute the marginal likelihood of model $\mathcal{M}_l$,

$$m_l(\boldsymbol{x}) = \int f_l(\boldsymbol{x} \mid \boldsymbol{\theta}_l)\pi_l(\boldsymbol{\theta}_l)\, d\boldsymbol{\theta}_l\,.$$

- Choose prior probabilities $P(\mathcal{M}_l)$ for the k models.
  - Example: This is an interesting issue for exoplanets, since we only have one observation of l (and even then l has been debated).
  - A standard objective choice is $P(\mathcal{M}_l) = 1/k$, but this can be bad in scenarios, such as variable selection, since it does not account for multiple testing.
- Compute the posterior probabilities of the models (Bayes theorem) as

$$P(\mathcal{M}_l \mid \boldsymbol{x}) = \frac{P(\mathcal{M}_l) m_l(\boldsymbol{x})}{\sum_{i=1}^{k} P(\mathcal{M}_i) m_i(\boldsymbol{x})} \,.$$

  - Example: Many years ago, for HD73526 based on 18 radial-velocity observations, $m_0(\boldsymbol{x}) = 5.9 \times 10^{-50}$, $m_1(\boldsymbol{x}) = 4.5 \times 10^{-41}$, and $m_2(\boldsymbol{x}) = 1.6 \times 10^{-42}$. If $P(\mathcal{M}_l) = 1/3$, then

$$P(\mathcal{M}_0 \mid \boldsymbol{x}) \approx 0, \quad P(\mathcal{M}_1 \mid \boldsymbol{x}) \approx 0.97, \quad P(\mathcal{M}_2 \mid \boldsymbol{x}) \approx 0.03 \,.$$

    Later, after 30 observations,

$$P(\mathcal{M}_0 \mid \boldsymbol{x}) \approx 0, \quad P(\mathcal{M}_1 \mid \boldsymbol{x}) \approx 0, \quad P(\mathcal{M}_2 \mid \boldsymbol{x}) \approx 1 \,.$$

Conventional prior and analysis for variable selection in the normal linear model

S. Bayarri, Jim Berger, A. Forte and G. García-Donato Annals of Statistics, 2013

Full Model: For $i = 1, \ldots, n$,

$$x_i = \sum_{j=1}^{k} z_{ij}\ \theta_j + \epsilon_i\ ,$$

where the z_{ij} are specified covariates and ϵ_i are 0 mean Gaussian errors.

Submodels: Any model of this form with some of the θ_j set equal to 0.

The paper:

- Developed desiderata (2 types of consistency and invariance, etc.) that should be satisfied by choices of priors for model uncertainty.
- Found an objective assignment of prior distributions to model parameters that was consistent with these desiderata.
- Produced software to implement the procedure.

R package BayesVarSel (Garcia-Donato and Forte 12-12-12)

- freely available at CRAN (for sequential or parallel computation)
- programmed in `C` using GNU-gsl libraries
- priors allowed:
  - prior.betas (i.e., the $\pi_l(\boldsymbol{\theta}_l)$):
    - "Robust" (*the recommended prior from the paper*),
    - "Liangetal",
    - "gZellner"
    - "ZellnerSiow".
  - prior.models (i.e., the $P(\mathcal{M}_l)$):
    - "Constant" (*Bad, since it will not adjust for the multiple testing.*)
    - "Jeffreys": Assign probability $1/(k+1)$ to all models of a given size, and divide this probability up among all models of that size. (*Good, because it does an appropriate adjustment for the multiple testing.*)

Common Bayesian outputs

- The *posterior inclusion probability* for variable i is

$$p_i = \sum_{l\,:\text{ variable } i \text{ is in } \mathcal{M}_l} P(\mathcal{M}_l \mid \boldsymbol{x}),$$

  the overall posterior probability that variable i is in the model (or that $\theta_i \neq 0$).

- The *median probability model* is the model consisting of those variables whose posterior inclusion probability is at least 1/2.
  - Surprisingly, this is typically a better model for prediction than the highest probability model.

- The Bayesian *model averaged predictor* (the best possible predictor from a Bayesian perspective) of $\boldsymbol{x}$ for a new vector of covariates $\boldsymbol{z}^*$ is

$$\hat{\boldsymbol{x}} = \boldsymbol{z}^*\ \hat{\boldsymbol{\theta}}, \quad \text{where} \quad \hat{\boldsymbol{\theta}} = \sum_l P(\mathcal{M}_l \mid \boldsymbol{x})\hat{\boldsymbol{\theta}}_l\,,$$

  with $\hat{\boldsymbol{\theta}}_l$ being the posterior mean under model $\mathcal{M}_l$.

Example: Infant obesity

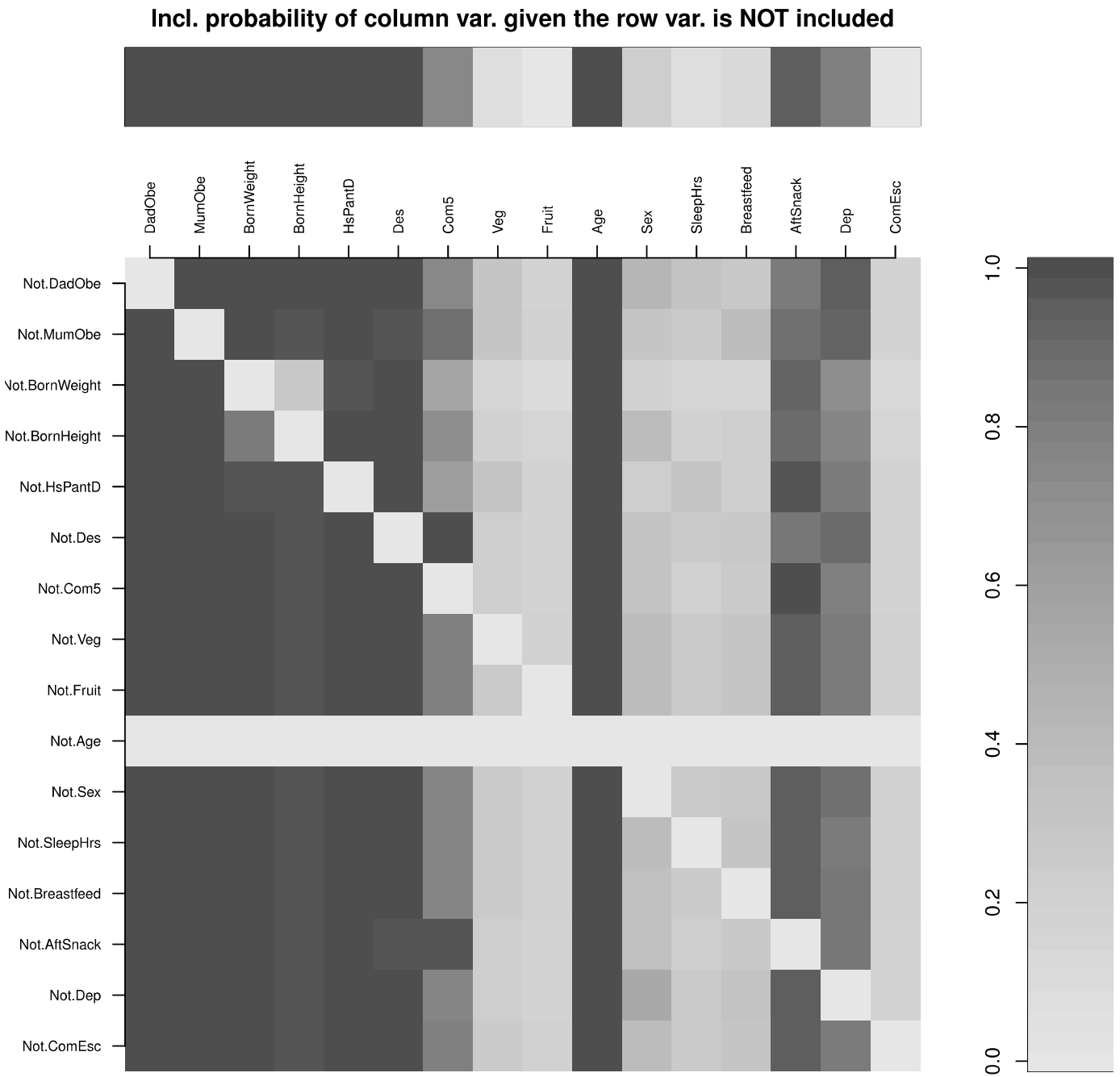
- Understand which factors affect *infant obesity*
- Response variable: Body Mass index (BMI)
- Set of 16 explanatory covariates with 1002 observations.
- There are 65,536 models.

```
> analRSB
Call:
Bvs(formula = "BMI~.", data = obesidad, prior.betas = "r", prior.models = "SB",
    n.keep = 1000)

This is the result for a model selection problem with 16 covariates and 1002 observations
The potential covariates are:
DadObe MumObe BornWeight BornHeight HsPantD Des Com5 Veg Fruit Age Sex SleepHrs Breastfeed AftSnack Dep ComEsc
It has take 23.36641 seconds to compute.

The 10 most probable models and their probabilities are:
   Intercept DadObe MumObe BornWeight BornHeight HsPantD Des Com5 Veg Fruit Age Sex SleepHrs Breastfeed AftSnack Dep ComEsc       prob
1          *      *      *          *          *       *   *    *                *                                 *   *        0.09410687
2          *      *      *          *          *       *   *    *                *   *                             *   *        0.04221871
3          *      *      *          *          *       *   *                     *                             *   *        0.03758388
4          *      *      *          *          *       *   *    *                *                 *           *   *        0.03180562
5          *      *      *          *          *       *   *    *                *             *               *   *        0.02792746
6          *      *      *          *          *       *   *    *   *            *                             *   *        0.02719485
7          *      *      *          *          *       *   *    *                *   *                 *       *   *        0.02041098
8          *      *      *          *          *       *   *    *                *                             *   *      * 0.01979366
9          *      *      *          *          *       *   *    *        *   *   *                             *   *        0.01922485
10         *      *      *          *          *       *   *    *   *            *   *                         *   *        0.01607668
>
```

```
> summary(analRSB)
Estimates and Inclusion Probabilities:
Note: If a covariate has a * in HPM(MPM) indicates that it is
 included in the Higest Probability Model (Median Probability Model)
                Estimate Incl.prob. HPM MPM
(Intercept) 23.662281637  1.0000000   *   *
DadObe       2.567126181  1.0000000   *   *
MumObe       2.097477902  1.0000000   *   *
BornWeight   0.001200232  0.9975935   *   *
BornHeight  -0.190916969  0.9908898   *   *
HsPantD      0.315063906  0.9963068   *   *
Des         -1.811662017  0.9950241   *   *
Com5        -0.647506038  0.8032369   *   *
Veg         -0.061650687  0.2962966
Fruit       -0.008038054  0.2368901
Age          0.478363586  1.0000000   *   *
Sex          0.147028315  0.4187302
SleepHrs    -0.034031386  0.3001681
Breastfeed  -0.103945758  0.3396248
AftSnack    -1.477718905  0.9525155   *   *
Dep         -0.137623013  0.8423909   *   *
ComEsc       0.022540222  0.2433688
>
```


Incl. probability of column var. given the row var. is NOT included
DadObe
MumObe
BornWeight
BornHeight
HsPantD
Des
Com5
Veg
Fruit
Age
Sex
SleepHrs
Breastfeed
AftSnack
Dep
ComEsc
Not.DadObe
Not.MumObe
Not.BornWeight
Not.BornHeight
Not.HsPantD
Not.Des
Not.Com5
Not.Veg
Not.Fruit
Not.Age
Not.Sex
Not.SleepHrs
Not.Breastfeed
Not.AftSnack
Not.Dep
Not.ComEsc
1.0
0.8
0.6
0.4
0.2
0.0

Multiple Testing

I. Introduction to the multiple testing problem

In a recent talk about the drug discovery process, the following numbers were given in illustration.

- 10,000 relevant compounds were screened for biological activity.
- 500 passed the initial screen and were studied in vitro.
- 25 passed this screening and were studied in Phase I animal trials.
- 1 passed this screening and was studied in a Phase II human trial.

This could be nothing but noise, if screening was done based on 'significance at the 0.05 level.'

If no compound had any effect,

- about $10,000 \times 0.05 = 500$ would initially be significant at the 0.05 level;
- about $500 \times 0.05 = 25$ of those would next be significant at the 0.05 level;
- about $25 \times 0.05 = 1.25$ of those would next be significant at the 0.05 level
- the 1 that went to Phase II would fail with probability 0.95.

A Bayesian multiple testing example:

- Suppose $X_i \sim N(x_i \mid \mu_i, \sigma^2)$, $i = 1, \ldots, m$, are observed.
- It is desired to test $H_i^0 : \mu_i = 0$ versus $H_i^1 : \mu_i \neq 0$, $i = 1, \ldots, m$, and *any* H_i^0 could be true or false regardless of the others.
- The simplest 'objective' probability assignment is $Pr(H_i^0) = Pr(H_i^1) = 0.5$, independently, for all i.
- This does *not* control for multiple testing; indeed, each test is then done completely independently of the others.

Note: The same holds in any model selection problem such as variable selection: *use of equal probabilities for all models typically does not induce any multiplicity adjustment.*

Inducing multiplicity control (Scott and Berger, 2006 JSPI; other, more sophisticated full Bayesian analyses are in Gönen et. al. (03), Do, Müller, and Tang (02), Newton et all. (01), Newton and Kendziorski (03), Müller et al. (03), Guindani, M., Zhang, S. and Mueller, P.M. (2007), ...; many empirical Bayes such as Efron and Tibshirani (2002), Storey, J.D., Dai, J.Y and Leek, J.T. (2007), Efron (2010))

- Suppose $X_i \sim N(x_i \mid \mu_i, \sigma^2)$, $i = 1, \ldots, m$, are observed, σ^2 known, and test $H_i^0 : \mu_i = 0$ versus $H_i^1 : \mu_i \neq 0$.
- If the hypotheses are viewed as exchangeable, let p denote the common prior probability of H_i^1, and *assume p is unknown* with a uniform prior distribution. *This does provide multiplicity control.*
- Complete the prior specification, e.g.
  - Assume that the nonzero μ_i follow a $N(0, V)$ distribution, with V unknown.
  - Assign V the objective (proper) prior density $\pi(V) = \sigma^2/(\sigma^2 + V)^2$.

- Then the posterior probability that $\mu_i \neq 0$ is

$$p_i = 1 - \frac{\int_0^1 \int_0^1 p \prod_{j \neq i} \left(p + (1-p)\sqrt{1-w}\, e^{w {x_j}^2/(2\sigma^2)} \right) dpdw}{\int_0^1 \int_0^1 \prod_{j=1}^m \left(p + (1-p)\sqrt{1-w}\, e^{w {x_j}^2/(2\sigma^2)} \right) dpdw}.$$

- $(p_1, p_2, \ldots, p_m)$ can be computed numerically; for large m, it is most efficient to use importance sampling, with a common importance sample for all p_i.

Example: Consider the following ten 'signal' observations:
-8.48, -5.43, -4.81, -2.64, -2.40, 3.32, 4.07, 4.81, 5.81, 6.24

- Generate $n = 10, 50, 500,$ and 5000 $N(0,1)$ noise observations.
- Mix them together and try to identify the signals.

	The ten 'signal' observations										#noise
n	-8.5	-5.4	-4.8	-2.6	-2.4	3.3	4.1	4.8	5.8	6.2	$p_i > .6$
10	1	1	1	.94	.89	.99	1	1	1	1	1
50	1	1	1	.71	.59	.94	1	1	1	1	0
500	1	1	1	.26	.17	.67	.96	1	1	1	2
5000	1	1.0	.98	.03	.02	.16	.67	.98	1	1	1

Table 1: The posterior probabilities of being nonzero for the ten 'signal' means.

Note 1: The penalty for multiple comparisons is automatic.

Note 2: **Theorem:** $E[\#i : p_i > .6 \mid \text{all } \mu_j = 0] = O(1)$ as $m \to \infty$, so the Bayesian procedure exerts medium-strong control over false positives. (In comparison, $E[\#i :$ Bonferroni rejects $\mid$ all $\mu_j = 0] = \alpha$.)

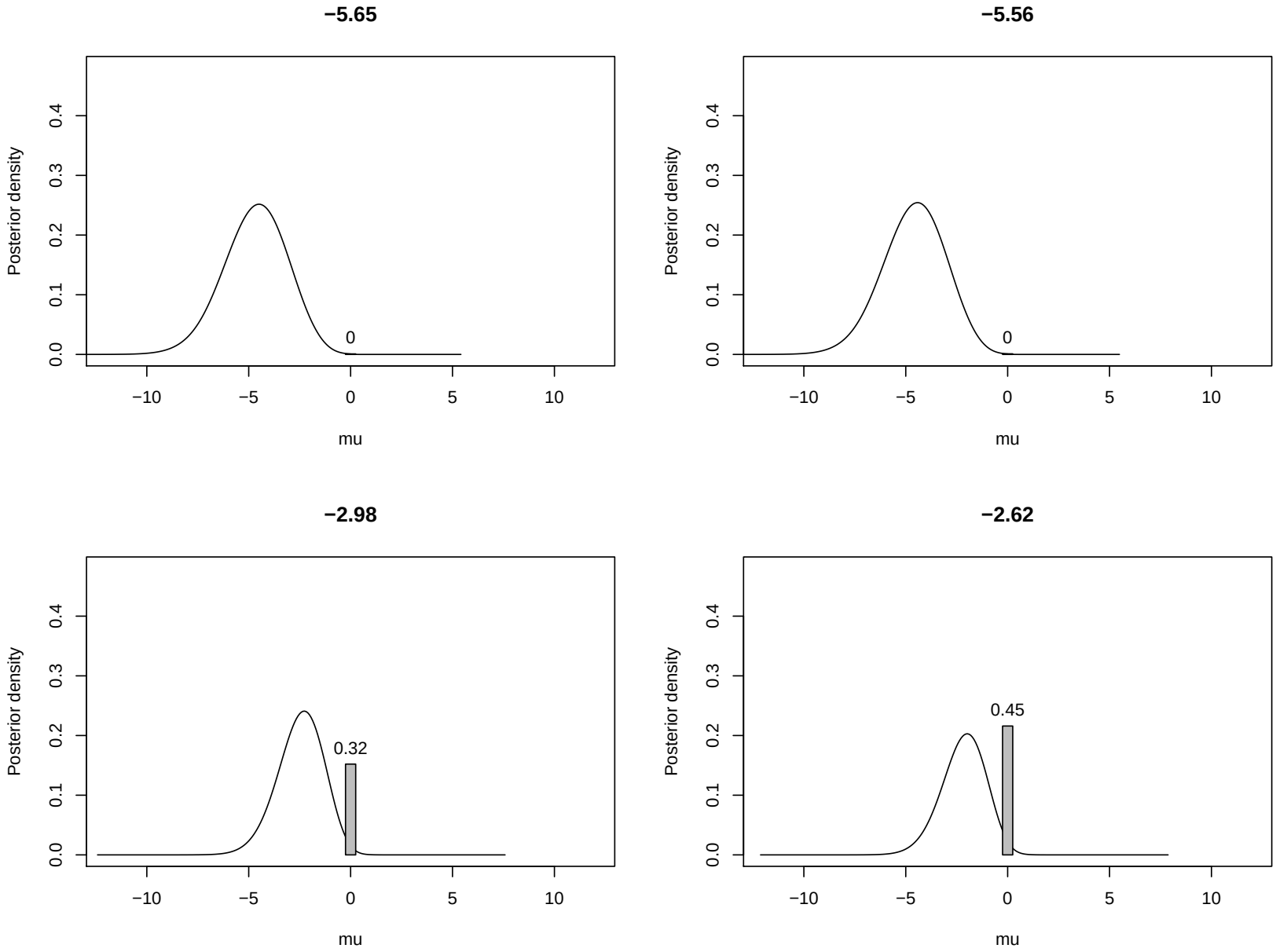


Figure 1: For four of the observations, $1 - p_i = \Pr(\mu_i = 0 \,|\, \boldsymbol{y})$ (the vertical bar), and the posterior densities for $\mu_i \neq 0$.

Hierarchical Modeling

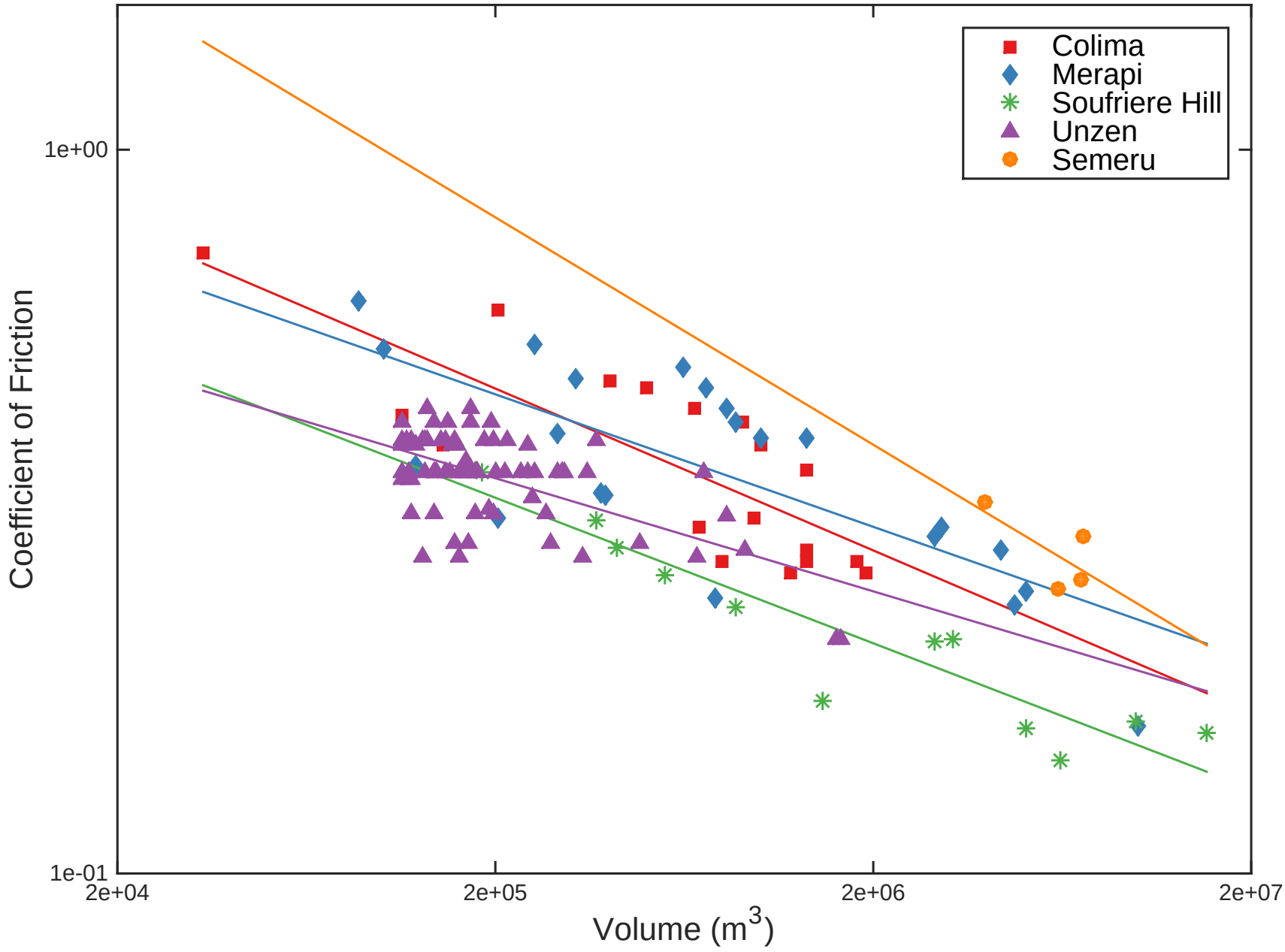
Example: Predicting the hazard from volcanic pyroclastic flows

To predict how far dangerous proclastic flows on volcanoes will go, it is necessary to learn the relationship between the volume of the flow, v, and the friction x of the flow with the ground, modeled on Mountain i as

$$x = \alpha_i + \theta_i v + \epsilon, \qquad \epsilon \overset{iid}{\sim} N(0, \sigma^2)$$

where α_i and θ_i are the unknown intercept and slope of the regression line for Mountain i.

Data is available from pyroclastic flows on 5 mountains, the data points given in the next figure along with least squares fits to each of the five regression lines.

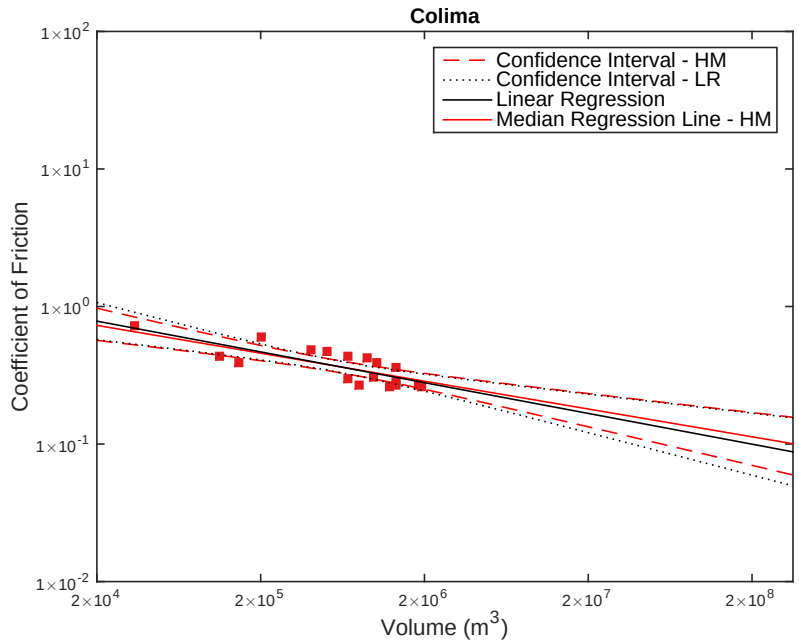

Colima
Merapi
Soufriere Hill
Unzen
Semeru
1e+00
1e-01
Coefficient of Friction
2e+04
2e+05
2e+06
2e+07
Volume (m3)

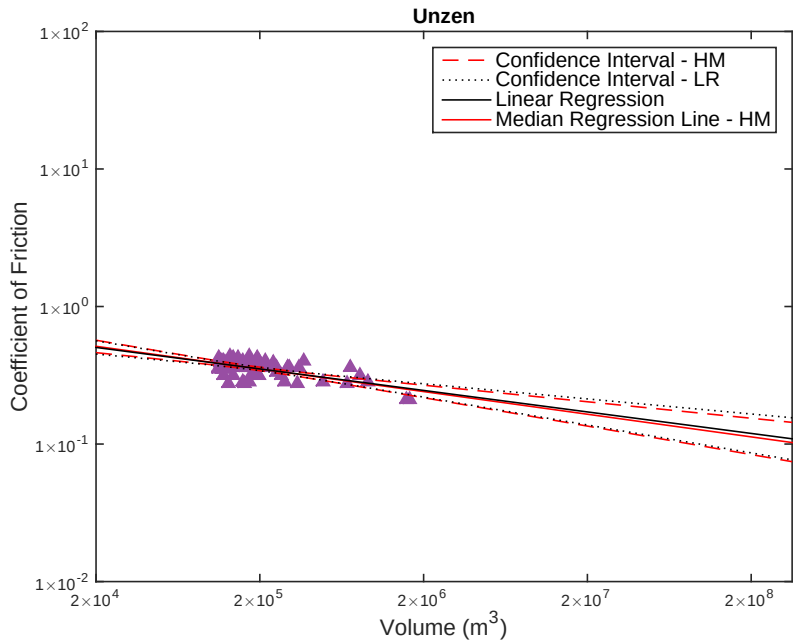
Hierarchical Bayes modeling:

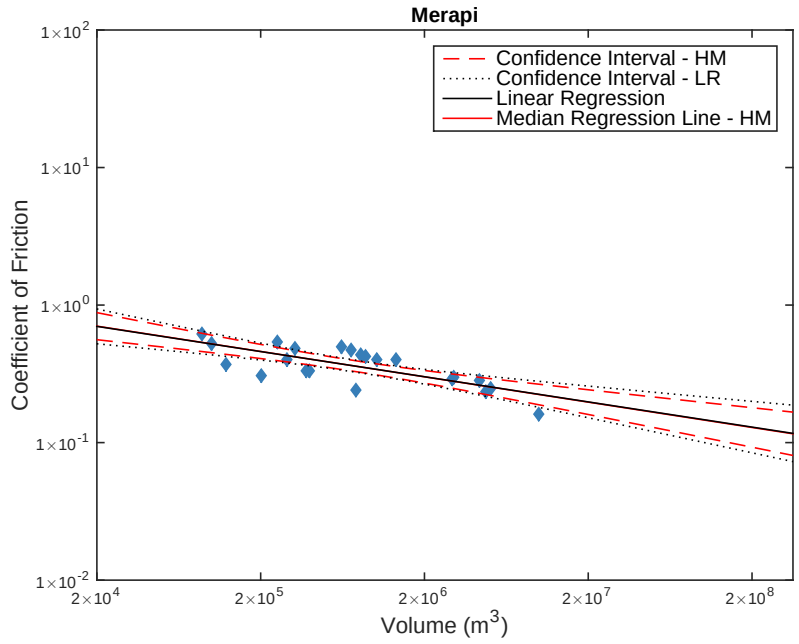
- It is believed that the intercepts α_i are unrelated and, hence, are simply assigned the usual objective prior $\pi(\alpha_i) = 1$.
- It is believed that the slopes θ_i are related, and are modeled as θ_i are i.i.d. $N(\cdot \mid \xi, \tau^2)$.
- The variance σ^2 is assigned the usual objective prior $\pi(\sigma^2) = 1/\sigma^2$.
- The hyperparameters ξ and τ^2 are assigned the objective prior $\pi(\xi, \tau^2) = 1/\tau$.

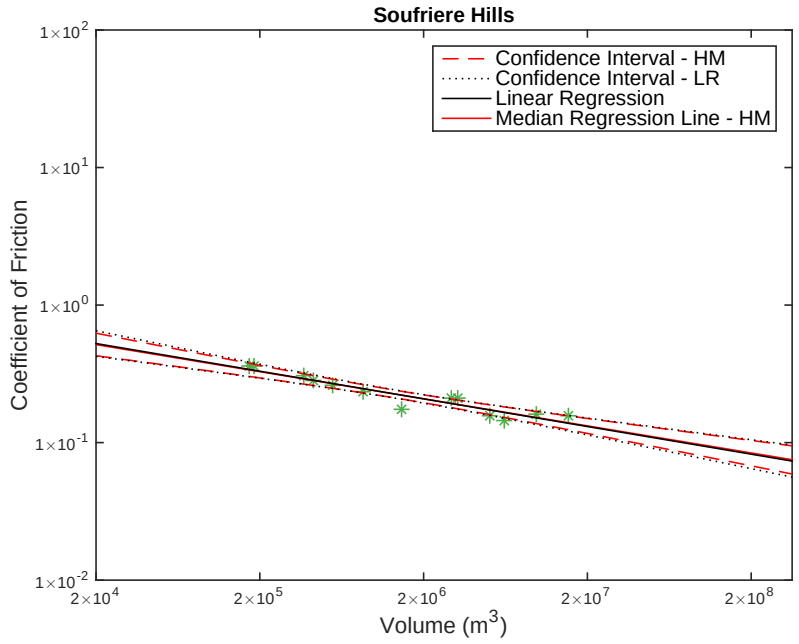
Analysis:

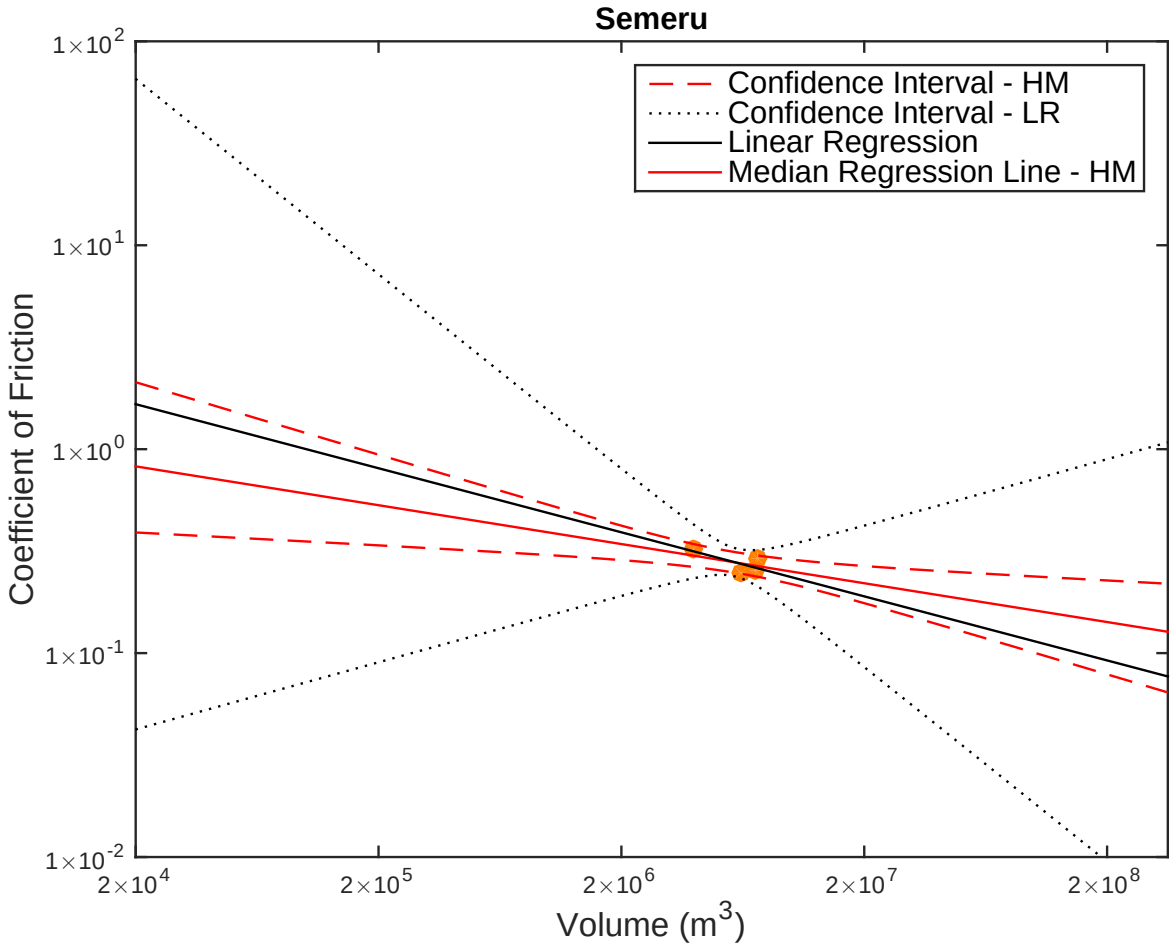
- The posterior distribution given the data is proportional to the product of the data likelihood and this prior.
- Gibbs sampling was done to obtain samples from this posterior.
- From these samples, the mean regression line and 95% confidence bands for the regression lines were found for each mountain, and are given in the next figures, along with the answers for ordinary linear regression.


Colima
Confidence Interval - HM
Confidence Interval - LR
Linear Regression
Median Regression Line - HM
Coefficient of Friction
Volume (m3)


Unzen
Confidence Interval - HM
Confidence Interval - LR
Linear Regression
Median Regression Line - HM
Coefficient of Friction
Volume (m3)


Merapi
Confidence Interval - HM
Confidence Interval - LR
Linear Regression
Median Regression Line - HM
Coefficient of Friction
Volume (m3)


Soufriere Hills
Confidence Interval - HM
Confidence Interval - LR
Linear Regression
Median Regression Line - HM
Coefficient of Friction
Volume (m3)



For this mountain, there were only 4 data points, so the ordinary linear regression gives very wide confidence bands (the dotted lines).

Hierarchical Bayes analysis gives much narrow bands because it uses the knowledge about the θ_i from the other mountains (often called "borrowing strength.")

Bayes and (sort of) Big Data

Example: emulating a compute model of pyroclastic flow

Emulation (approximation) of simulators

- Simulators – complex computer models of processes – often take hours to weeks for a single run, making it infeasible to use the simulator directly in intensive computations.
- A crucial need is thus development of an *emulator*, a fast simulator approximation (also called surrogate models, meta models, response surface approximations, ...).
- Emulators are typically statistical in nature, most often Gaussian processes.

Example: TITAN2D – can simulate pyroclastic flow on Montserrat

Inputs: *Initial conditions:* $V =$ flow volume; $\varphi =$ flow direction;
Model parameters: $\delta_{bed} =$ basal friction ; $\delta_{int} =$ internal friction.

Background of the application: The simulator, TITAN2D, for given inputs V, φ, δ_{bed}, and δ_{int}, predicts the pyroclastic flow, $y^M(\boldsymbol{x}, t \mid V, \varphi, \delta_{bed}, \delta_{int})$, over a large ($10^9$) space-time grid of points $(\boldsymbol{x}, t)$. Each run of TITAN2D takes two hours.

TITAN2D is run at $m = 2048$ vectors of inputs, yielding the 'data' $\boldsymbol{y}^D$ which can be a matrix of size up to 2048×10^9.

Based on $\boldsymbol{y}^D$, we need to construct the emulator, which needs to

- predict the output of TITAN2D at inputs other than the initial 2048;
- provide an (accurate) assessment of the accuracy of the prediction.

Note that the emulator is developed on the same grid of space time points $(\boldsymbol{x}, t)$ on which TITAN2D is run; we thus switch notation and write $y_j^M(V, \varphi, \delta_{bed}, \delta_{int}) = y^M(\boldsymbol{x}, t \mid V, \varphi, \delta_{bed}, \delta_{int})$, with the coordinates $j = 1, \ldots, k$ (k as large as 10^9) being the grid points.

The simplest possible (simultaneous) Gaussian process emulator:

An *independent* real Gaussian process is assigned to each coordinate $y_j^M(V, \varphi, \delta_{bed}, \delta_{int})$, with

- prior mean functions of the regression form $\Psi(\boldsymbol{x})\ \boldsymbol{\theta}_j$, where $\Psi(\boldsymbol{x})$ is a *common* l-vector of given basis functions and the $\boldsymbol{\theta}_j$ are *differing* unknown regression coefficients;
- *differing* unknown prior variances σ_j^2;
- *common* estimated correlation parameters $\hat{\boldsymbol{\beta}}$.
- The particular type of Gaussian process is not really important, but here we use a separable process with power exponential correlations.

The surprising result is that the emulator mean (the best prediction of the simulator), at a new vector of inputs, $(V^*, \varphi^*, \delta^*_{bed}, \delta^*_{int})$, is of the form

$$\boldsymbol{h}(V^*, \varphi^*, \delta^*_{bed}, \delta^*_{int})\ \boldsymbol{y}^D\,, \tag{1}$$

where computation of the $m \times 1$ vector $\boldsymbol{h}(V^*, \varphi^*, \delta^*_{bed}, \delta^*_{int})$ requires only $m^2 = 2048^2$ computations. Thus the overall computational cost for large k (e.g. 10^9) is only mk.

- The emulator is an interpolator so, when $(V^*, \varphi^*, \delta^*_{bed}, \delta^*_{int})$ equals one of the initial runs of the simulator, the emulator will return the exact values from the simulator.
- As the emulator mean is just a weighted average of the actual simulator runs, it hopefully captures some of the dynamics of the process.
- Computation of all the emulator variances is $O(m^2 k)$, but one rarely needs to compute all of them.
- Even if a (separable) spatial Gaussian process is also placed on the space time locations $(\boldsymbol{x}, t)$, the emulator posterior mean is that in (1).

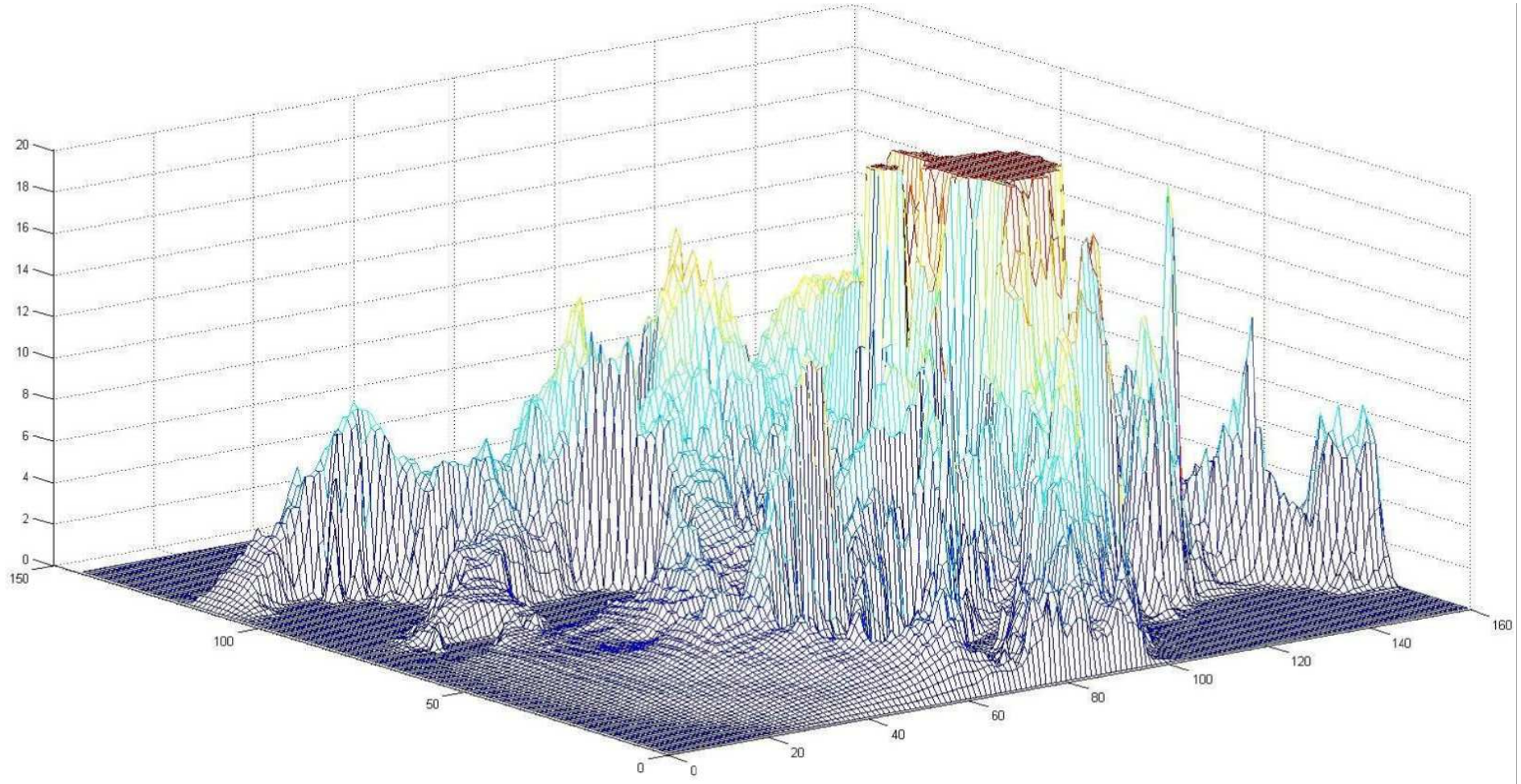


Figure 2: The mean of the emulator of 'maximum flow height over time' from TITAN2D, at 24,000 spatial locations over Montserrat and for new input values $V^* = 10^{7.462}$, $\varphi^* = 2.827$, $\delta^*_{bed} = 11.111$, and $\delta^*_{int} = 27.7373$.

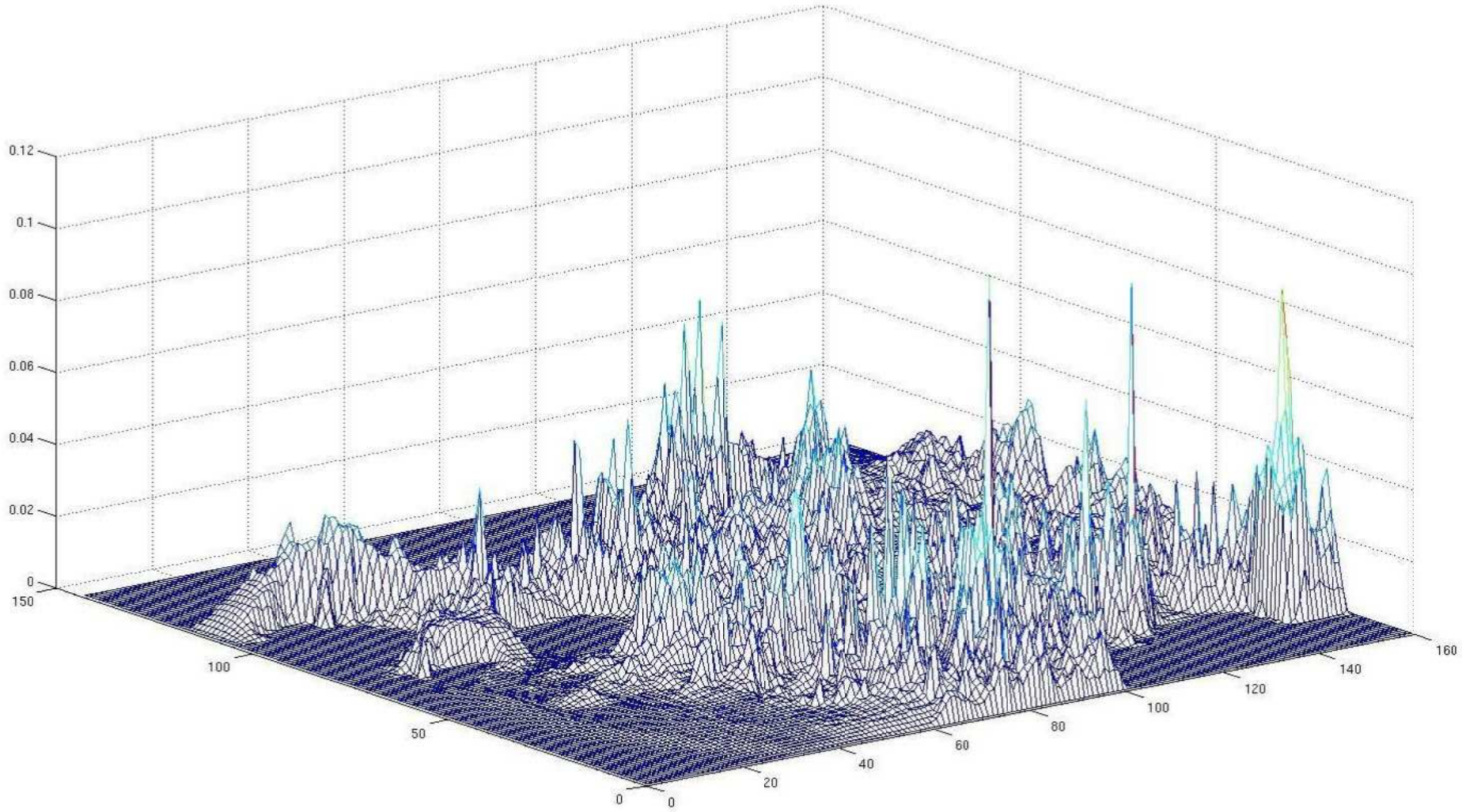


Figure 3: Variance of the emulator of 'maximum flow height over time' from TITAN2D, at 24,000 spatial locations over Montserrat and for new input values $V^* = 10^{7.462}$, $\varphi^* = 2.827$, $\delta^*_{bed} = 11.111$, and $\delta^*_{int} = 27.7373$.

A New Bayes/Frequentist Synthesis in Testing

Setup:

We observe data $\boldsymbol{x}$ from the density $f(\boldsymbol{x} \mid \theta)$ and wish to test

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta \neq \theta_0 \, .$$

- Suppose a frequentist rejection region $\mathcal{R}$ is specified.
- Let $\alpha = Pr(\mathcal{R} \mid \theta_0)$ and $(1 - \beta(\theta)) = Pr(\mathcal{R} \mid \theta)$ be the Type I error and power corresponding to the rejection region $\mathcal{R}$.
- Let π_0 and $\pi_1 = 1 - \pi_0$ be the prior probabilities of H_0 and H_1, respectively.
- Let $\pi(\theta)$ be the prior density of θ under H_1 (this could just be a point mass at a point for which power is to be evaluated).
  - Then $(1 - \bar{\beta}) = \int (1 - \beta(\theta))\pi(\theta) d\theta$ is the average power wrt the prior $\pi(\theta)$.
  - And $m(\boldsymbol{x}) = \int f(\boldsymbol{x} \mid \theta)\pi(\theta) d\theta$ is the marginal likelihood of the data $\boldsymbol{x}$ under the prior $\pi(\theta)$.

Pre-experimental analysis (not new):

The pre-experimental probability of incorrectly rejecting H_0 is then $\pi_0\alpha$, while the pre-experimental probability of correctly rejecting H_0 is $\pi_1(1-\bar{\beta})$.

Definition: The *pre-experimental odds of correct to incorrect rejection of* H_0 are

$$\begin{aligned} O_{pre} &= \frac{\pi_1}{\pi_0} \times \frac{(1-\bar{\beta})}{\alpha} \\ &\equiv O_P \times O_R \\ &\equiv [\text{prior odds of } H_1 \text{ to } H_0] \times [\textit{rejection odds} \text{ of } H_1 \text{ to } H_0]\,. \end{aligned}$$

Reporting of the rejection odds, O_R, recognizes the crucial role of power in understanding the strength of evidence in rejecting, and does so in a simple way (reducing the evidence to a single number).

average power	0.05	0.25	0.50	0.75	1.0	0.01	0.25	0.50	0.75	1.0
type I error	0.05	0.05	0.05	0.05	0.05	0.01	0.01	0.01	0.01	0.01
O_R	1	5	10	15	20	1	25	50	75	100

Post-experimental odds analysis (not new):

Once the data is at hand a Bayesian would focus on the posterior odds of H_1 to H_0 given by

$$\begin{aligned} O_{post} &= \frac{\pi_1}{\pi_0} \times \frac{m(\boldsymbol{x})}{f(\boldsymbol{x} \mid \theta_0)} \\ &\equiv O_P \times B_{10}(\boldsymbol{x})\,, \end{aligned}$$

where $B_{10}(\boldsymbol{x})$ is the Bayes factor (or weighted likelihood ratio) of H_0 to H_1.

Lemma (new): *The frequentist expectations of* $B_{10}(\boldsymbol{x})$ *and* $B_{01}(\boldsymbol{x}) = 1/B_{10}(\boldsymbol{x})$ *over the rejection region, conditional on the respective hypotheses, are*

$$E[B_{10}(\boldsymbol{x}) \mid H_0, \mathcal{R}] = O_R \quad \textit{and} \quad E[B_{01}(\boldsymbol{x}) \mid H_1^*, \mathcal{R}] = [O_R]^{-1}\,,$$

where H_1^* *refers to the marginal alternative model with density* $m(\boldsymbol{x})$.

The first identity guarantees that, under H_0, the "average of the reported Bayes factors when rejecting" equals the actual rejection odds O_E, so $B_{10}(\boldsymbol{x})$ *is a completely valid frequentist report.* It is also better than the unconditional O_R, in that it reflects the error arising for the actual data.

Example: Genome-wide Association Studies (GWAS)

- Early genomic epidemiological studies almost universally failed to replicate (estimates of the replication rate are as low as 1%), because they were doing multiple testing at 'ordinary p-values'.
- A very influential paper in Nature (2007) by the Wellcome Trust Case Control Consortium proposed cutoff $p < 5 \times 10^{-7}$.
  - Found 21 genome/disease associations; 20 have been replicated.
- Bayesian argument for the cutoff:
  - *Pre-experimental odds of true to false rejection* = *prior odds* $\times \frac{(1-\bar{\beta})}{\alpha}$.
  - For the GWAS study, they wanted pre-experimental odds of 10 : 1 for true to false rejections; estimated *prior odds* $= \frac{1}{100{,}000}$ and $(1 - \bar{\beta}) = 0.5$. Solving gave $\alpha = 5 \times 10^{-7}$.
    (They stated the prior odds could vary by a factor of 10.)
  - Some authors of the paper argued that it is better to report the Bayes factors $B_{10}(\boldsymbol{x})$, and the posterior odds = *prior odds* $\times B_{10}(\boldsymbol{x})$. These ranged between $\frac{1}{10}$ and 10^{68} for the 21 claimed associations.

Thanks!