The Bayesian Approach to Combination of Evidence

Jim Berger

Duke University
Statistical and Applied Mathematical Sciences Institute

DIMACS
Rutgers
May 6-7, 2010

My Philosophy Towards Combining Evidence

- If information were precise, there would be no issue:
  - each additional piece of information would constrain the solution space;
  - with enough information the answer would eventually become apparent.
- But almost no information is precise in the inferential sense (except, perhaps, that coming from a precise theory, and even then the theory was likely not originally "precise"); thus the real question is how to combine information involving uncertainties.
- Dominant answer (through 250 years of study, including study of hundreds of alternatives): probability theory, the instantiation of which in combining evidence is Bayesian analysis.
- Caveat: computational and real-time processing considerations may entail utilization of many other techniques.

Some Issues in Bayesian Combination of Evidence

- Expert information is easily combined with other information through prior elicitation, but in many contexts elicitation is difficult or unwanted (FDA device division); one can then still often use Bayesian combination of evidence through
  - hierarchical modeling;
  - objective Bayes.
- Senstivity or Robustness
- Computation, computation, computation

A Medical Diagnosis Example (with Mossman, 2001)

The Medical Problem:

- Within a population, $p_0 = Pr(\text{Disease } D)$.
- A diagnosic test results in either a Positive (P) or Negative (N) reading.
- $p_1 = Pr(P \mid \text{patient has } D)$.
- $p_2 = Pr(P \mid \text{patient does not have } D)$.

It follows from Bayes theorem that

$$\theta = Pr(D \mid P) = \frac{p_0 p_1}{p_0 p_1 + (1 - p_0) p_2}.$$

The Statistical Problem: The $p_i \in (0, 1)$ are unknown, but three independent medical studies are reported in the literature, yielding $X_i \sim \text{Binomial}(n_i, p_i),\ i = 0, 1, 2.$

Goal: find a $100(1 - \alpha)\%$ confidence set for θ.

Suggested Objective Bayes Solution:

- Assign p_i the Jeffreys-rule prior (Beta(1/2,1/2) distribution)

$$\pi(p_i) \propto p_i^{-1/2}(1-p_i)^{-1/2}.$$

- By Bayes theorem, the posterior distribution of p_i given the data, x_i, is

$$\pi(p_i \mid x_i) = \frac{p_i^{-1/2}(1-p_i)^{-1/2} \times \begin{pmatrix} n \\ x_i \end{pmatrix} p_i^{x_i}(1-p_i)^{n_i - x_i}}{\int p_i^{-1/2}(1-p_i)^{-1/2} \times \begin{pmatrix} n \\ x_i \end{pmatrix} p_i^{x_i}(1-p_i)^{n_i - x_i}\, dp_i},$$

which is the Beta$(x_i + \frac{1}{2}, n_i - x_i + \frac{1}{2})$ distribution;

- the joint posterior distribution of p_0, p_1, and p_2 is (by independence)

$$\pi(p_0 \mid x_0)\pi(p_1 \mid x_1)\pi(p_2 \mid x_2)\,,$$

- which determines $\pi(\theta \mid x_0, x_1, x_2)$, the posterior distribution of

$$\theta = Pr(D \mid P) = \frac{p_0 p_1}{p_0 p_1 + (1-p_0)p_2}\,.$$

Computational implementation for determining the confidence set for θ:

One can simply compute the desired confidence set (formally, the $100(1-\alpha)\%$ equal-tailed posterior credible set) by

- drawing random p_i from the Beta$(x_i + \frac{1}{2}, n_i - x_i + \frac{1}{2})$ distributions, $i = 0, 1, 2$;
- computing the associated $\theta = p_0 p_1 / [p_0 p_1 + (1 - p_0) p_2]$;
- repeating this process $10,000$ times;
- using the $\frac{\alpha}{2}\%$ upper and lower percentiles of these generated θ to form the desired confidence limits.

$n_0 = n_1 = n_2$	(x_0, x_1, x_2)	95% confidence interval
20	(2,18,2)	(0.107, 0.872)
20	(10,18,0)	(0.857, 1.000)
80	(20,60,20)	(0.346, 0.658)
80	(40,72,8)	(0.808, 0.952)

Table 1: The 95% equal-tailed posterior credible interval for $\theta = Pr(D \mid P) = \frac{p_0 p_1}{p_0 p_1 + (1-p_0)p_2}$, for various values of the n_i and x_i.

Consider the frequentist percentage of the time that the 95% Bayesian credible sets for $\theta = Pr(D \mid P) = \frac{p_0 p_1}{p_0 p_1 + (1-p_0) p_2}$ miss on the left and on the right (ideal would be 2.5% each) for the indicated parameter values when $n_0 = n_1 = n_2 = 20$.

(p_0, p_1, p_2)	O-Bayes	Log Odds	Gart-Nam	Delta
$(\frac{1}{4}, \frac{3}{4}, \frac{1}{4})$	2.86,2.71	1.53,1.55	2.77,2.57	2.68,2.45
$(\frac{1}{10}, \frac{9}{10}, \frac{1}{10})$	2.23,2.47	0.17,0.03	1.58,2.14	0.83,0.41
$(\frac{1}{2}, \frac{9}{10}, \frac{1}{10})$	2.81,2.40	0.04,4.40	2.40,2.12	1.25,1.91

Adjusting for Multiple Testing

San Jose Mercury News

mercurynews.com WEST VALLEY 102 Friday, September 25, 2009 THE NEWSPAPER OF SILICON VALLEY 75 cents

AIDS MILESTONE

New path for HIV vaccine

Some in study protected from infection, but trial raises more questions

By Karen Kaplan and Thomas H. Maugh II
Los Angeles Times

Hours after HIV researchers announced the achievement of a milestone that had eluded them for a quarter of a century, reality began to set in: Tangible progress could take another decade.

A Thai and American team announced early Thursday in Bangkok that they had found a combination of vaccines providing modest protection against infection with the virus that causes AIDS, unleashing excitement worldwide. The idea of a vaccine to prevent infection with the human immunodeficiency virus, HIV, had long been frustrating and fruitless.

But by Thursday afternoon, initial euphoria gave way to a more sober assessment. There is still a very long way to go before reaching the goal of producing a vaccine that reliably shields people from HIV.

Some researchers questioned whether the apparent 31 percent reduction in infections was a sta-

See **VACCINE**, *Page 14*

A researcher during the Thai phase III HIV Vaccine Trial, also known as RV 144, tests the "prime-boost" combination of two vaccines.
ASSOCIATED PRESS

Hypotheses and Data:

- Alvac had shown no effect
- Aidsvax had shown no effect

Question: Would Alvac as a primer and Aidsvax as a booster work?

The Study: Conducted in Thailand with 16,395 individuals from the general (not high-risk) population:

- 71 HIV cases reported in the 8198 individuals receiving placebos
- 51 HIV cases reported in the 8197 individuals receiving the treatment

The test that was performed:

- Let p_1 and p_2 denote the probability of HIV in the placebo and treatment populations, respectively.
- Test $H_0 : p_1 = p_2$ versus $H_1 : p_1 \neq p_2$
- Normal approximation okay, so

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{\sigma}_{\{\hat{p}_1 - \hat{p}_2\}}}} = \frac{.00926 - .00641}{.00140} = 2.04$$

  is approximately $\mathrm{N}(\theta, 1)$, where $\theta = (p_1 - p_2)/(.00140)$.
  We thus test $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$, based on z.

- Observed $z = 2.04$, so the p-value is 0.04.

Questions:

- Is the p-value useable as a direct measure of vaccine efficacy?
- Should the fact that there were two previous similar trials be taken into account?

Bayesian Analysis of the Single Trial:

Prior distribution:

- $Pr(H_i)$ = prior probability that H_i is true, $i = 0, 1$,
- On $H_1 : \theta > 0$, let $\pi(\theta)$ be the prior density for θ.

Note: H_0 must be believable (at least approximately) for this to be reasonable (i.e., no fake nulls).

Subjective Bayes: choose these based on personal beliefs

Objective (or default) Bayes: choose

- $Pr(H_0) = Pr(H_1) = \frac{1}{2}$,
- $\pi(\theta) = \text{Uniform}(0, 6.46)$, which arises from assigning
  - uniform for p_2 on $0 < p_2 < p_1$,
  - plug in for p_1 .

Posterior probability of hypotheses:

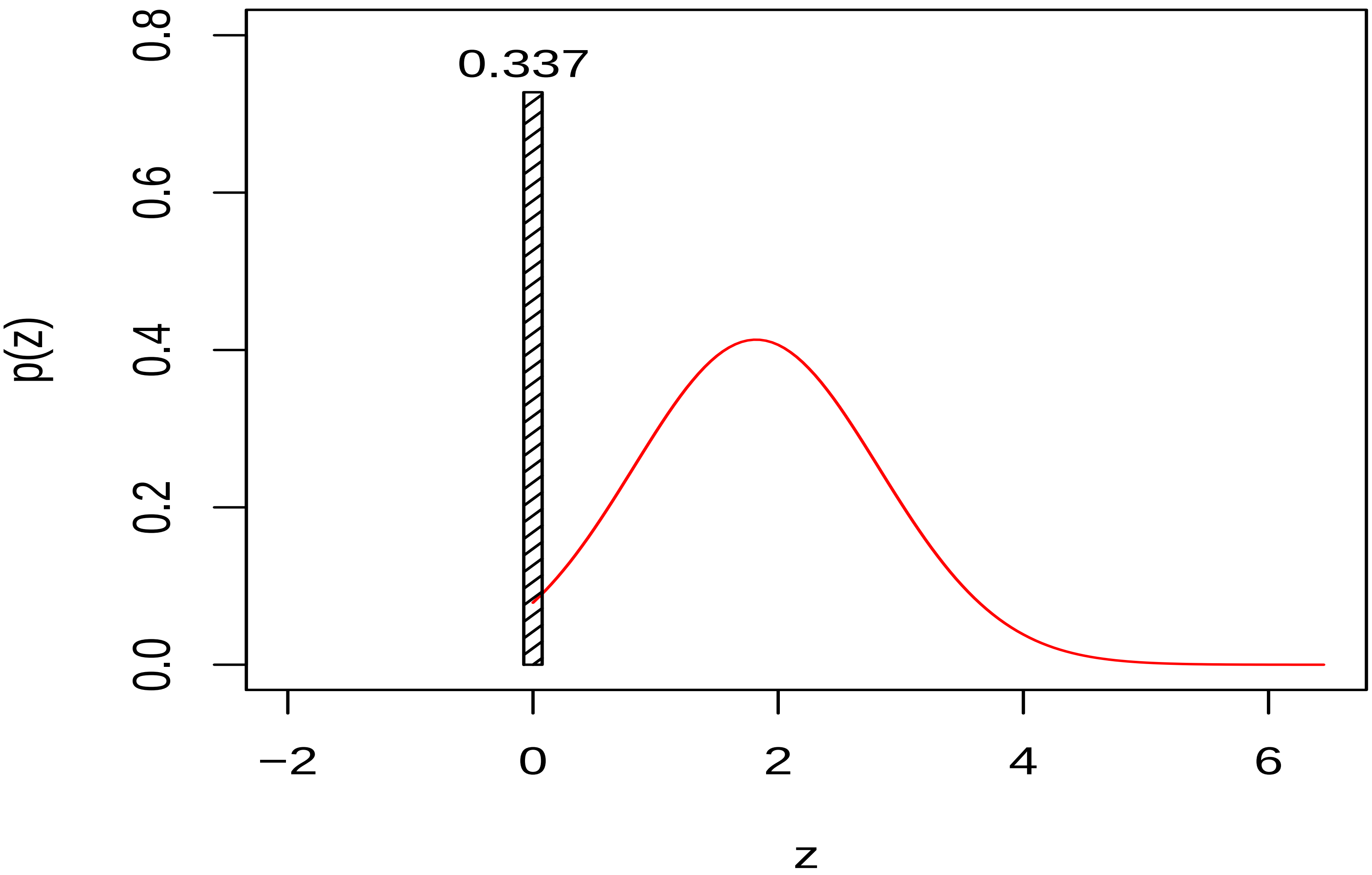
$$
\begin{aligned}
Pr(H_0|z) &= \text{probability that } H_0 \text{ true, given data } z \\
&= \frac{f(z \,|\, \theta = 0)\, Pr(H_0)}{Pr(H_0)\, f(x \,|\, \theta = 0) + Pr(H_1) \int_0^\infty f(z \,|\, \theta)\pi(\theta) d\theta}
\end{aligned}
$$

For the objective prior, $Pr(H_0 \mid z = 1.82) \approx 0.337$ (recall, p-value $\approx .04$)

Posterior density on $H_1 : \theta > 0$ is

$$\pi(\theta|z = 1.82, H_1) \propto \pi(\theta) f(1.82 \mid \theta) = (0.413) e^{-\frac{1}{2}(1.82-\theta)^2}$$

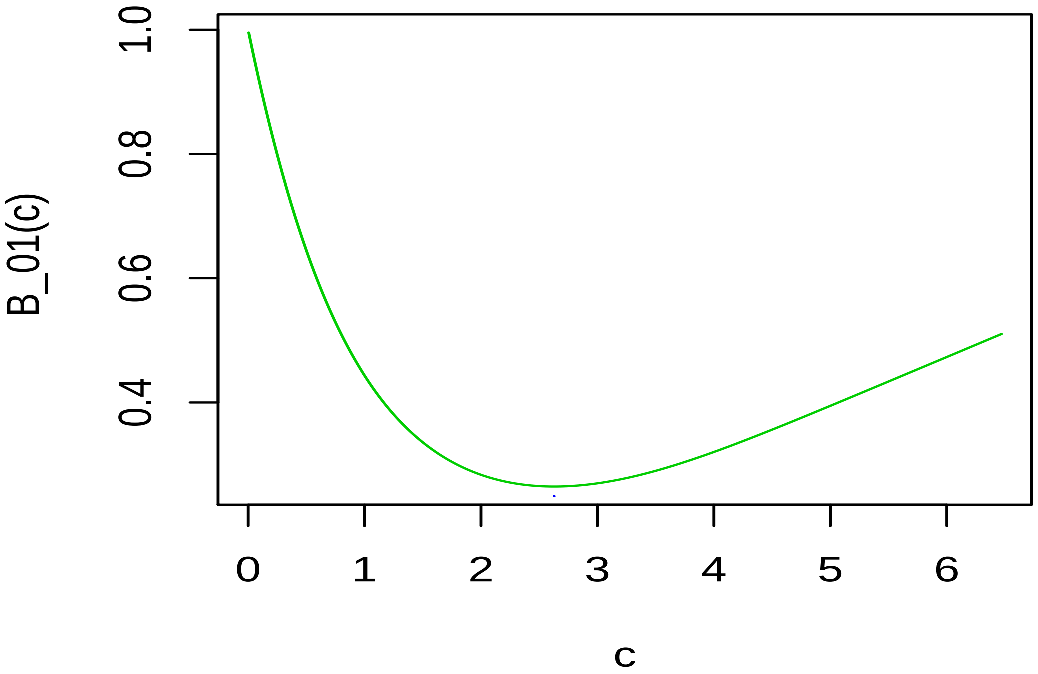
for $0 < \theta < 6.46$.


0.337
p(z)
0.0
0.2
0.4
0.6
0.8
−2
0
2
4
6
z

Robust Bayes: Report the *Bayes factor* (the odds of H_0 to H_1) as a function of $\pi_C(\theta) \equiv \text{Uniform}(0, C)$:

$$B_{01}(C) = \frac{\text{likelihood of H}_0\text{ for observed data}}{\text{average likelihood of H}_1} = \frac{\frac{1}{\sqrt{2\pi}} e^{-(1.82-\theta)^2/2}}{\int_0^C \frac{1}{\sqrt{2\pi}} e^{-(1.82-\theta)^2/2} C^{-1} d\theta}$$

.



Note: $\min_C B_{01}(C) = 0.265$ (while $B_{01}(6.46) = 0.51$).

Incorporation information from multiple tests: To adjust for the two previous similar failed trials, the (exchangeable) Bayesian solution

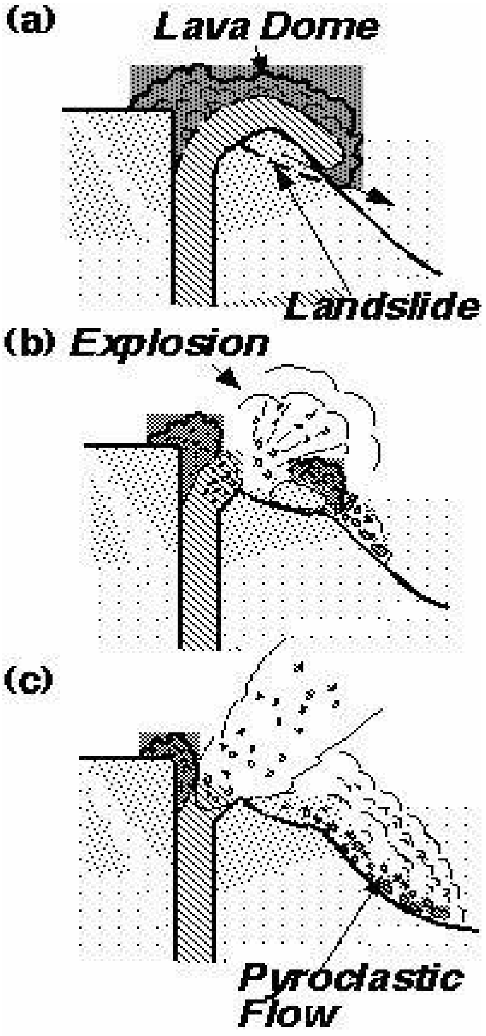
- assigns each trial common unknown probability p of success, with p having a uniform distribution;
- computes the resulting posterior probability that the current trial exhibits no efficacy

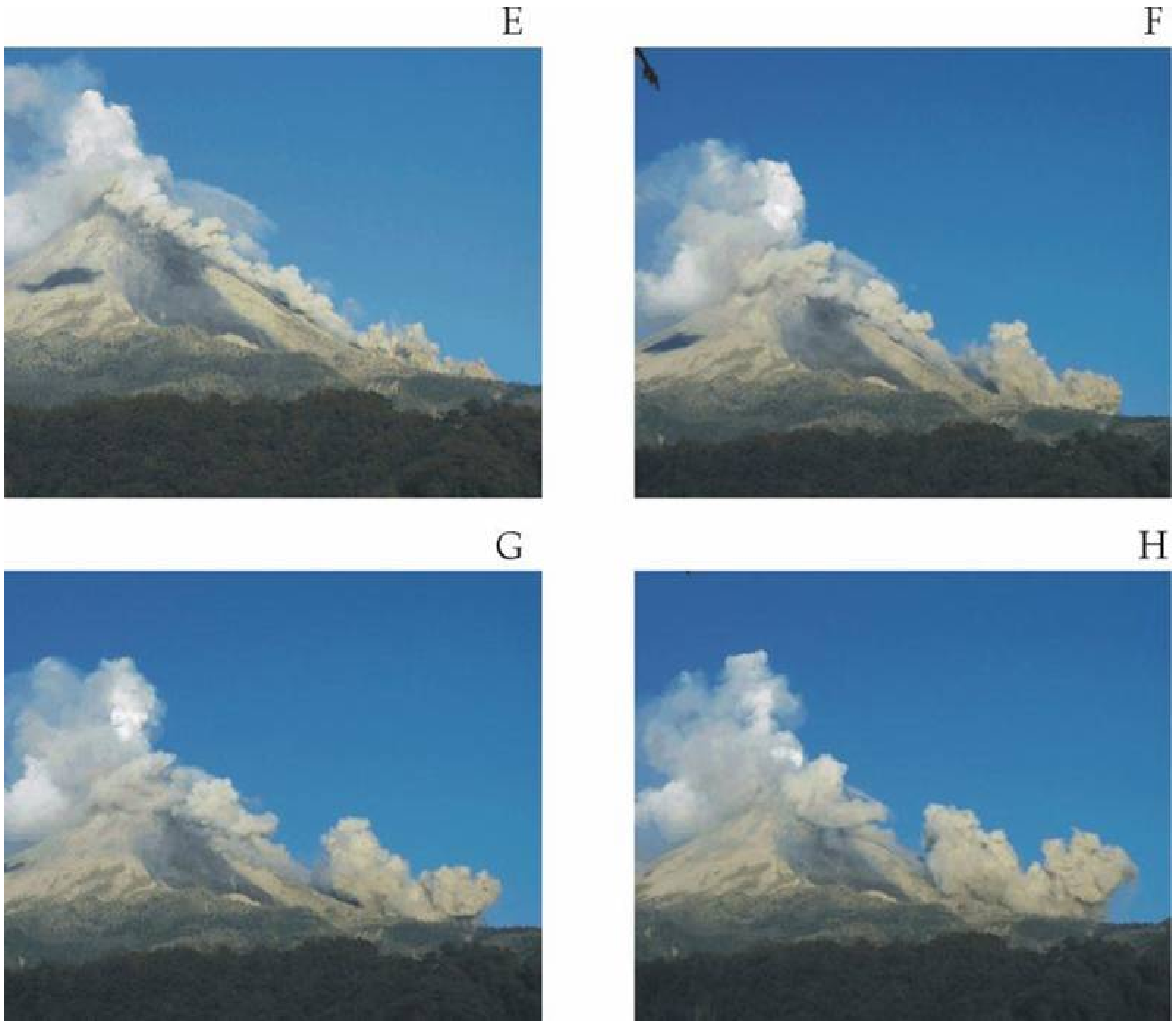
$$Pr(H_0 \mid x_1, x_2, x_3) = \left(1 + \frac{B_{01}(x_1)B_{01}(x_2) + B_{01}(x_1) + B_{01}(x_2) + 3}{3B_{01}(x_1)B_{01}(x_2) + B_{01}(x_1) + B_{01}(x_2) + 1} \times \frac{1}{B_{01}(x_3)}\right)^{-1}$$

where $B_{01}(x_i)$ is the Bayes factor of “no effect” to “effect” for trial i.

The result is $Pr(H_0 \mid x_1, x_2, x_3) = 0.54$.

Combining information from deterministic and statistical models: Example - risk from pyroclastic flows





Plymouth, the former capital of Montserrat

We combine use of computer models and statistical models to assess the risk of a volcanic hazard. We compute

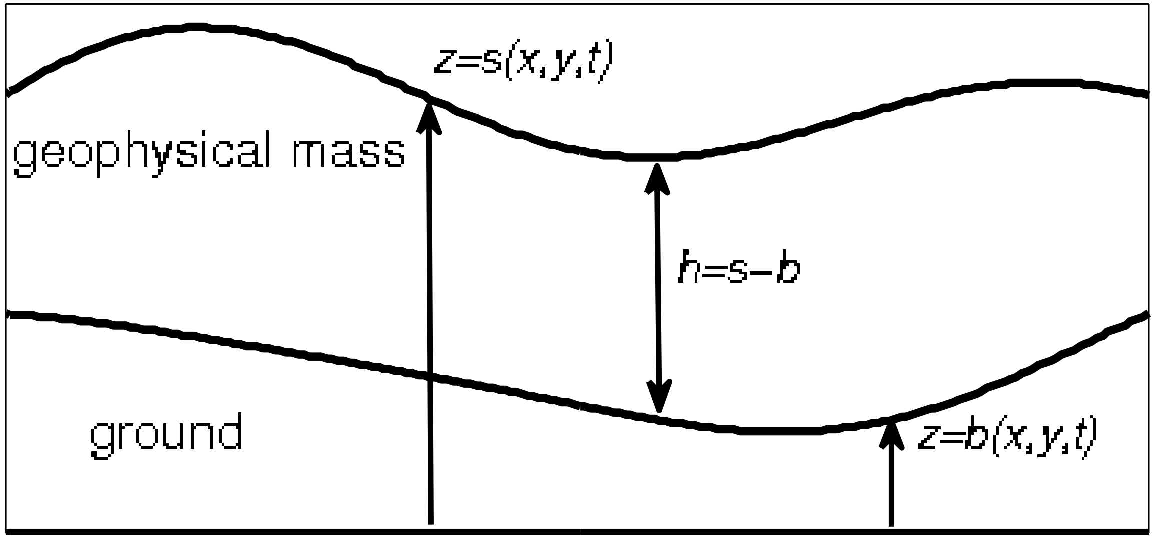
Pr (a catastrophic event occurs in the next T years)

at specified locations, utilizing

- computer implementations of mathematical models of flows to allow extrapolation to unseen situations;
- statistical models for needed stochastic inputs to the computer model, appropriate for rare events;
- a computational strategy for rare events, based on development of adaptive approximations to the computer model.

The Geophysical/Math Model

- Use 'thin layer' modeling $\rightsquigarrow$ system of PDE for the flow depth and the depth-averaged momenta.
- Main feature: Incorporates topographical data from GIS.



The Computer Model Implementation

TITAN2D (U Buffalo) computes solution to the math model

- Stochastic inputs whose randomness is the basis of the risk uncertainty:
  - x_1 = initial volume V (size of initial flow),
  - x_2 = initial angle φ (direction of initial flow).
- x_3 = basal friction coefficient b (friction at interface of flow and ground).
- Other inputs: internal friction,initial velocity (speed and direction) $\rightsquigarrow$ kept fixed for the moment.
- Output: flow height and depth-averaged velocity at every grid point at every time step; we will focus on the maximum flow height at each grid point.
- Each run takes about 1 hour, so that a Gaussian process needs to be developed for most of the computations.

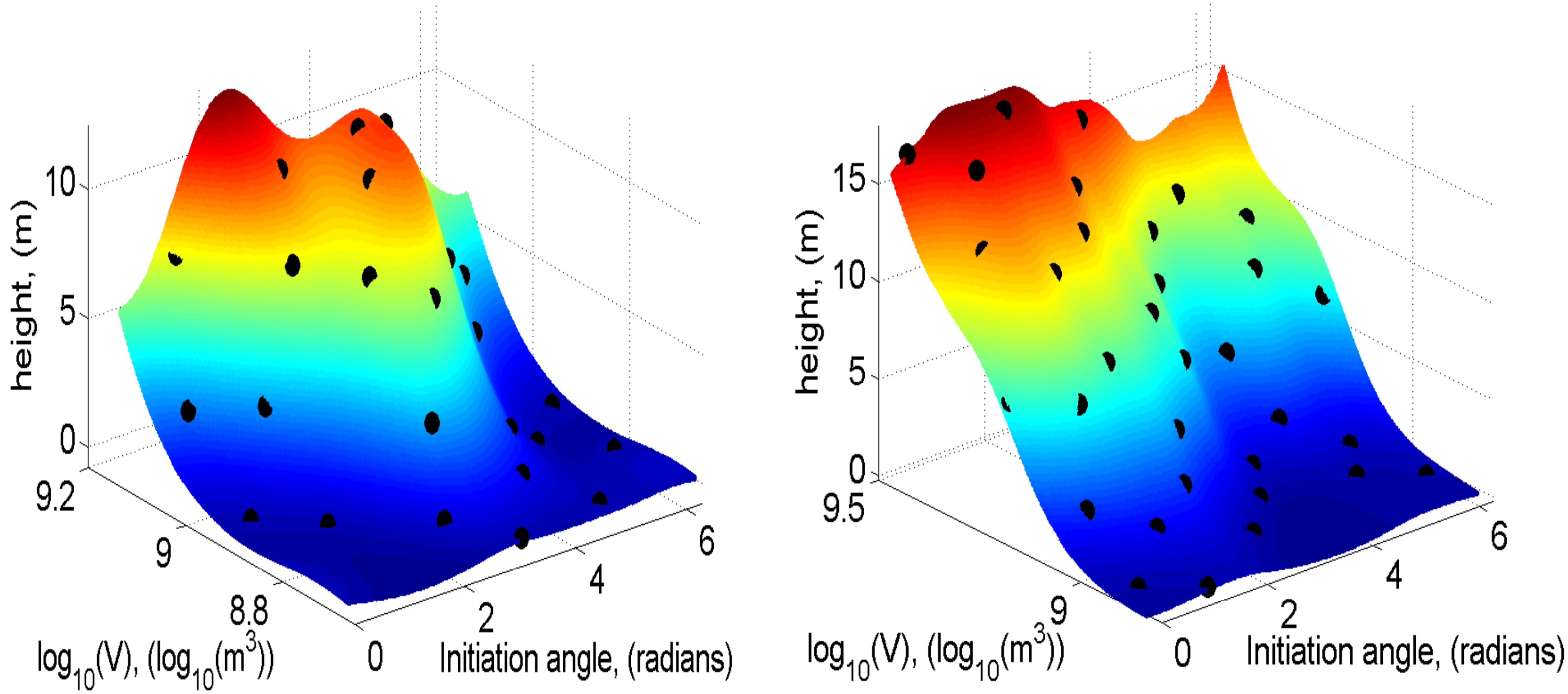


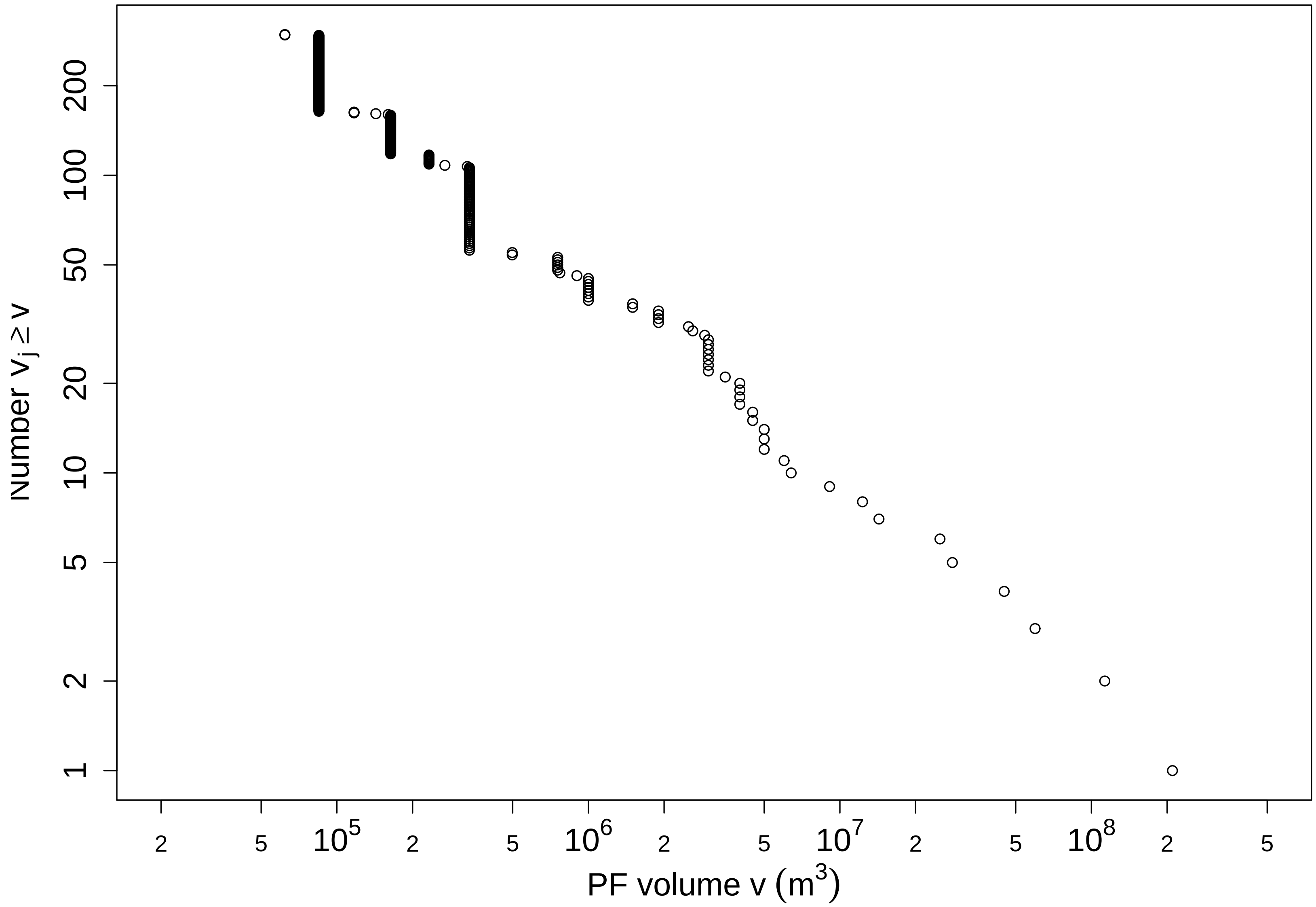
Figure 1: Median of the emulator, transformed back to the original space. Left: Plymouth, Right: Bramble Airport. Black points: max-height simulation outputs at design points.

Risk Assessment: Probability of Catastrophe

- Use the emulator to determine the *critical region* $\mathcal{X}_C$ of input values that would lead to a catastrophe.
- Determine the distribution of the input variables (V, φ, b) to compute

$$\Pr(\text{at least one } (V, \varphi, b) \in \mathcal{X}_C \text{ in the next } t \text{ years }) .$$

  – The distribution of b is found from auxiliary data concerning information about pyroclastic flow runouts and runs of TITAN2D for those flows.
  – The distribution of the stochastic inputs (V, φ, b) is found from field data, using objective Bayesian analysis.



Number $v_j \geq v$
200
100
50
20
10
5
2
1
2
5
10^5
2
5
10^6
2
5
10^7
2
5
10^8
2
5
PF volume v (m^3)

Probability of a catastrophic event

It follows that, for any fixed $t > 0$, the number of catastrophic PF's (those with $V_i > \Psi(\varphi_i)$) in t years is Poisson with (conditional) mean

$$\begin{aligned} E\big(\#\text{ catastrophic PFs in } t \text{ yrs } \mid \alpha, \lambda\big) &= \int_0^{2\pi} \int_{\Psi(\varphi)}^{\infty} [\lambda \epsilon^{-\alpha} t] \frac{f(v \mid \alpha)}{2\pi} dv d\varphi \\ &= \frac{t\,\lambda}{2\pi} \int_0^{2\pi} \Psi(\varphi)^{-\alpha}\, d\varphi\,, \end{aligned}$$

$$\Pr(\text{At least one CPF in } t \text{ yrs} \mid \alpha, \lambda) = 1 - \exp\left[-\frac{t\,\lambda}{2\pi} \int_0^{2\pi} \Psi(\varphi)^{-\alpha}\, d\varphi \right],$$

$$\begin{aligned} P(t) &\equiv \Pr(\text{At least one CPF in } t \text{ yrs} \mid \text{data}) \\ &= 1 - \iint \exp\left[-\frac{t\,\lambda}{2\pi} \int_0^{2\pi} \Psi(\varphi)^{-\alpha}\, d\varphi \right] \pi(\alpha, \lambda \mid \text{data})\, d\alpha\, d\lambda\,, \end{aligned}$$

where $\pi(\alpha, \lambda \mid \text{data})$ is the posterior distribution of (α, λ) given the data.

Computing the probabilities of catastrophe

To compute Pr(at least one catastrophic event in t years | data) for a range of t, an importance sampling estimate is

$$P(t) \cong 1 - \frac{\sum_i \exp\left[- \frac{t\,\lambda_i\,\widehat{\Psi}(\alpha_i)}{2\pi}\right] \frac{\pi^*(\alpha_i,\lambda_i)}{f_I(\alpha_i,\lambda_i)}}{\sum_i \frac{\pi^*(\alpha_i,\lambda_i)}{f_I(\alpha_i,\lambda_i)}},$$

- where $\widehat{\Psi}(\alpha)$ is an MC estimate of $\int_0^{2\pi} \Psi(\varphi)^{-\alpha}\, d\varphi$ based on draws $\varphi_i \sim Un(0, 2\pi)$;
- $\pi^*(\alpha, \lambda)$ is the un-normalized posterior;
- (α_i, λ_i) are drawn from the importance sampling density $f_I(\alpha, \lambda) = t_2(\alpha, \lambda \mid \widehat{\mu}, \widehat{\Sigma}, 3)$, with d.f. 3, mean $\widehat{\mu}^t = (\hat{\alpha}, \hat{\lambda})$, and scale $\widehat{\Sigma}$ = inverse of observed Fisher information matrix.

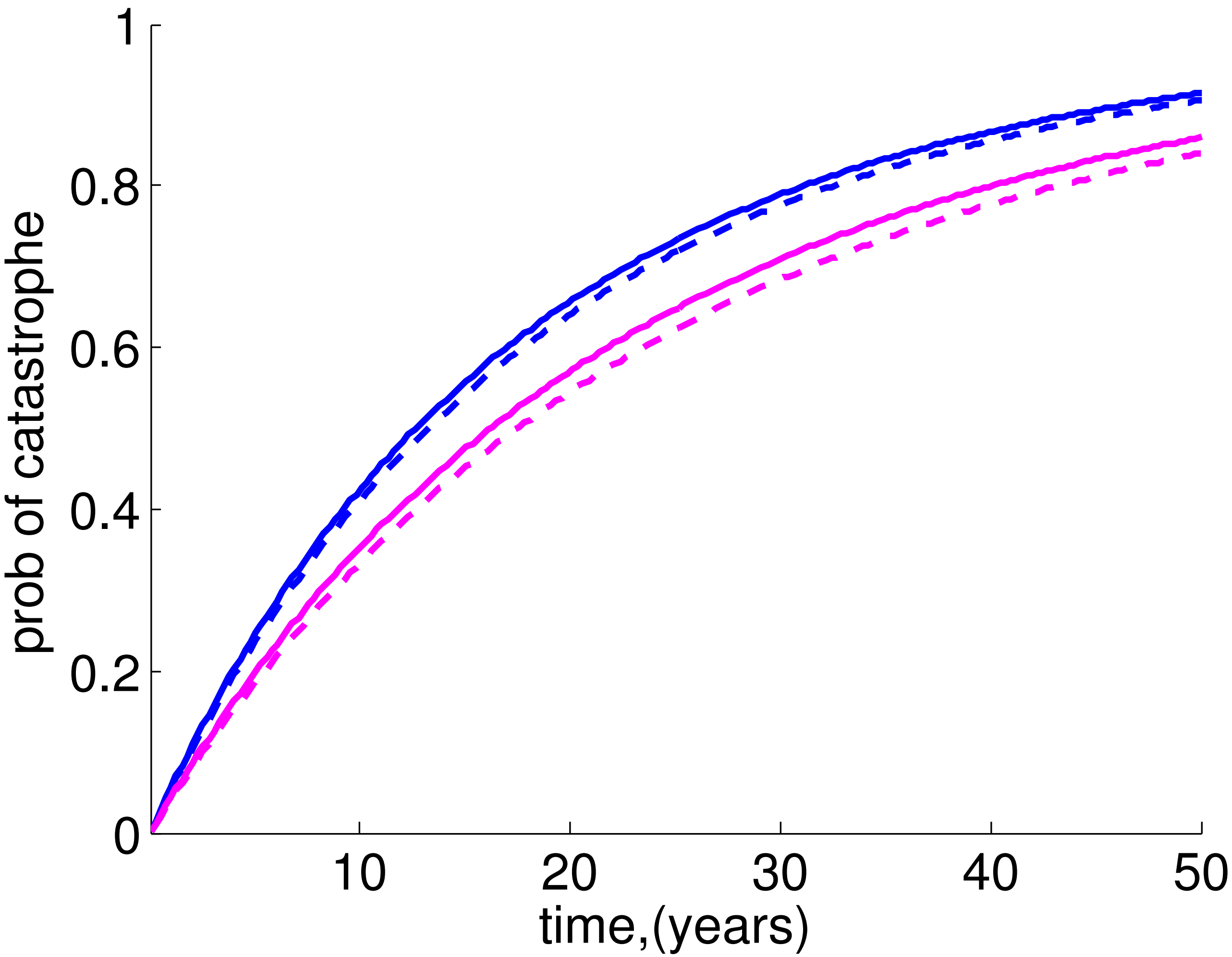


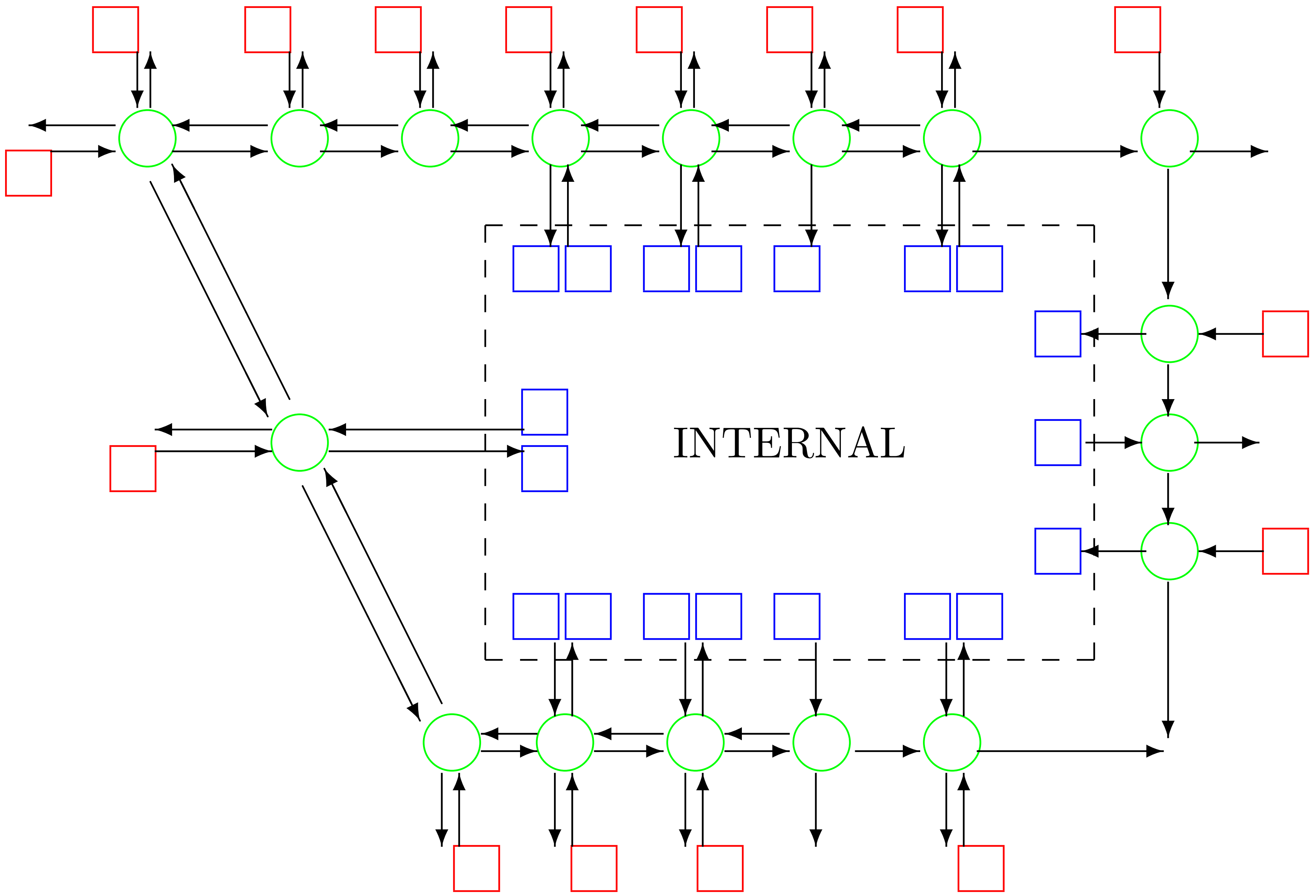
Figure 2: $P(t)$ at Plymouth (higher curves) and Airport (lower curves). Solid (dashed) $\leadsto$ computed with the upper (lower) 75% confidence bands. Reference priors lead to overlapping curves.

A Traffic Microsimulator

- CORSIM is a vehicular traffic microsimulator.
- Application: use CORSIM to model a 44-intersection neighborhood in Chicago.
- Data: Vehicle counts (many inaccurate) from a 1-hour period during rush hour (9am-10am) on a single day.
- Goal: Solve the 'inverse problem' (or parameter estimation problem) of determining needed inputs for CORSIM from this data.

Needed CORSIM inputs and data

- Needed CORSIM inputs:
  - *Demands* (λ): 16 Means of exponential inter-arrival time distributions, at entry points to the network.
  - 84 *turning probabilities* ($\mathbf{P}$).
- Data: Types of counts ($\mathbf{C}$):
  - Demands: Observer counts of # vehicles entering network at the 16 entry locations during the hour.
  - Turning Counts: Observer counts of # vehicles turning at each intersection during 20-minute intervals.
  - Video Counts: Counts from a video recording of the # vehicles passing through central intersections.
- Goal: Obtain $\pi(\lambda, \mathbf{P} \mid \mathbf{C})$, the posterior distribution of the inputs $\lambda, \mathbf{P}$ given the data.



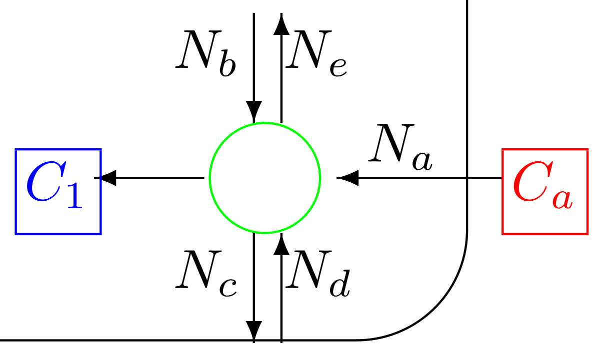
Locations of Observer counts, Video counts, and Turning counts.

Probabilistic Structure and Latent Counts

- Each network link has unobserved latent number of vehicles N_i.
- Observed Demand Counts $C_i \sim \text{Poisson}(b_i N_i)$.

  Observer bias $b_i \sim \text{Gamma}(\alpha, \beta)$ with $\pi(\alpha, \beta) = 1_{\alpha < 2\beta}$

  True Demand counts $N_i \sim \text{Poisson}(\lambda_i)$ with $\pi(\lambda_i) \propto \lambda_i^{-1}$
- Latent turning counts $(N_{iL}, N_{iT}, N_{iR}) \sim \text{MN}(N_i | P_{iL}, P_{iT}, P_{iR})$ with $\pi(P_{iL}, P_{iT}, P_{iR}) \propto (P_{iL} P_{iT} P_{iR})^{-\frac{1}{2}}$
- Observed turning counts: independent observations coming from the same multinomial distribution as the latent counts.
- Network restrictions (27): $N_{bT} + N_{aL} = N_c (= N_{cL} + N_{cT} + N_{cR})$.
- Video restrictions (10):
  $C_1 = N_{dL} + N_{aT} + N_{bR}$.



Problem 1: Reparameterization of the Posterior The 37 linear restrictions essentially mean that there are 37 extra parameters in the model. Write the restrictions as

$$\boxed{\Gamma\mathbf{N}=\Lambda} \begin{cases} \Gamma\text{: (37 by 127) matrix of \{-1,0,1\} coefficients} \\ \mathbf{N}\text{: (127 by 1) matrix of parameters} \\ \Lambda\text{: (37 by 1) matrix of known coefficients (video or zeros)} \end{cases}$$

To reparameterize,

- find a singular (37 by 37) $\Gamma^\star$ and (37 by 90) $\mathbf{B}$ (with corresponding partitions $\mathbf{N}_1$ and $\mathbf{N}_2$ of $\mathbf{N}$) such that $\boxed{\Gamma^\star\mathbf{N}_1 + \mathbf{B}\mathbf{N}_2 = \Lambda}$;
- replace $\mathbf{N}_1$ in the likelihood with $\boxed{\mathbf{N}_1 = (\Gamma^\star)^{-1}\Lambda - (\Gamma^\star)^{-1}\mathbf{B}\mathbf{N}_2}$.

Computation of $\Gamma^\star$ is a fast and easy preprocessing step.
Existence of (at least one) $\Gamma^\star$ is guaranteed by the nontriviality of the restrictions.

Problem 2: Iteration-dependent support of the N_i Cheaply finding the exact support of each of the 90 remaining N_i is crucial for efficient MCMC. To compute the support

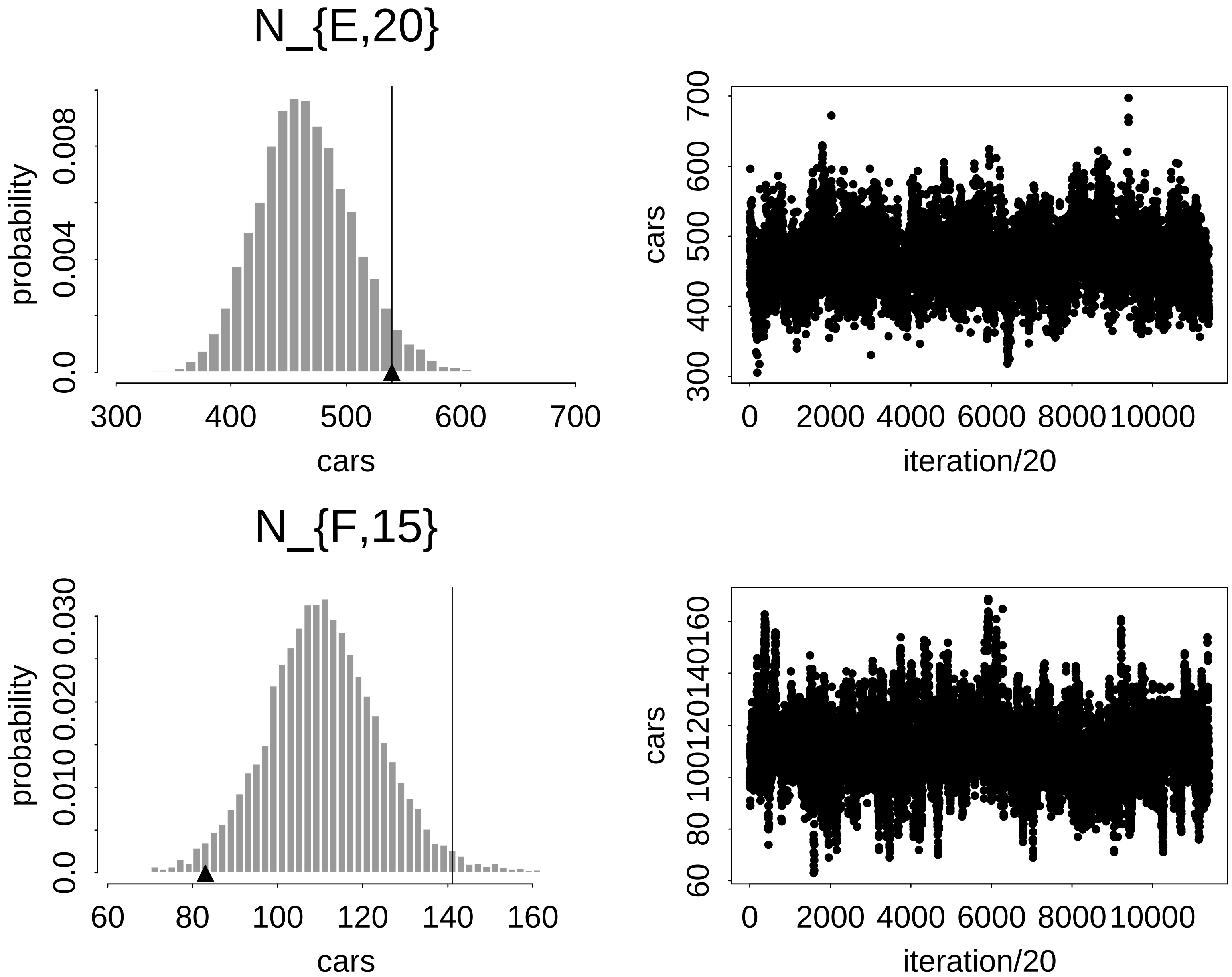
- note that nonnegativity of counts $\equiv$ nonnegativity of each multinomial factorial argument in the likelihood:
  $[N_i|\ldots] \propto$
  $$\frac{(f_1(N_{-i},C)+a_1N_i)!\ f_6(P)^{N_i}}{(f_2(N_{-i},C)+a_2N_i)!\ (f_3(N_{-i},C)+a_3N_i)!\ (f_4(N_{-i},C)-a_4N_i)!\ (f_5(N_{-i},C)-a_5N_i)!},$$
  where f_1, f_2, f_3 (+ sign) *can* be negative, and f_4, f_5 (- sign) *are* non-negative;

- also, G.L.B.= -min$(0,\frac{f_1}{a_1},\frac{f_2}{a_2},\frac{f_3}{a_3})$, L.U.B.= +min$(\frac{f_4}{a_4},\frac{f_5}{a_5})$, with the existence of support in the previous iteration guaranteeing that L.U.B. $\geq$ G.L.B;

- which together imply $N_i \in \left[\text{-min}(0,\frac{f_1}{a_1},\frac{f_2}{a_2},\frac{f_3}{a_3})\ ,+\text{min}(\frac{f_4}{a_4},\frac{f_5}{a_5})\right]$.

Problem 3: MCMC starting values for the latent counts

- A starting value of the 90-dimensional N must satisfy $\Gamma \mathbf{N} = \Lambda$, $\mathbf{N} \geq 0$, and all N's are integer-valued.
- Since the demand inputs to the network are unknown, this is actually easy to ensure for our case.
- Indeed, a simple algorithm can be created which provides a satisfactory starting value for any set of data.



Posterior distributions of two of the latent demands.

Final Comments

- Bayesian combination of evidence (of all types) is conceptually straightforward.
- It is often done through hierarchical modeling and utilization of objective Bayesian methods.
- Subjective Bayesian analysis is a crucial addition if one wishes to incorporate expert opinion.
- Computational issues can require the need of many types of approximation.

Thanks!