Constrained Optimization and Calibration for Deterministic and Stochastic Simulation Experiments

Herbie Lee & Bobby Gramacy

University of California, Santa Cruz

& University of Chicago Booth School of Business

Additional collaborators:

Genetha Gray (Sandia National Laboratories)

Crystal Linkletter (Brown University)

Thanks to AIM, NSF, RAND for various support

`http://www.ams.ucsc.edu/~herbie`

Outline

1. Optimization and Expected Improvement
2. Constrained Optimization
3. Integrated Expected Conditional Improvement
4. Examples

Optimization and Calibration

For concreteness, we focus on minimization here

$$\text{Find } x^* = \arg\min_{x \in \mathcal{X}} f(x)$$

Variety of possible approaches:

- Gradient-based methods
- Simulated annealing
- Genetic algorithms
- Surrogate modeling

Surrogate Modeling

- Use a statistical model as a fast approximation to the unknown objective function
- Traditional model is a Gaussian Process (GP)
  - A random process with any finite collection of points having a joint multivariate normal distribution, with covariance depending on the displacement between the points
- Fit model and find minimum
- Or search iteratively

Improvement Function

$$I(x) = \max\{f_{\min} - f(x), 0\}$$

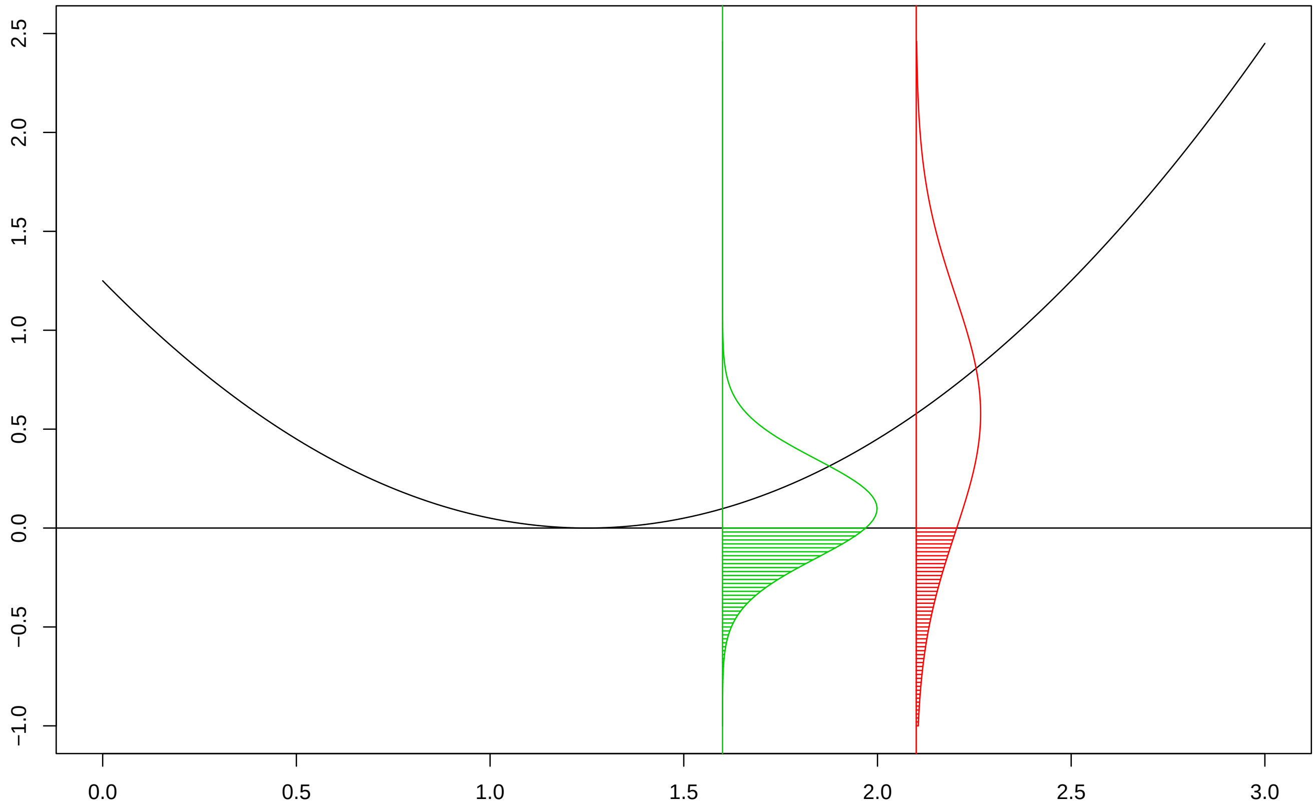
- $f_{\min} = \min\{f(x_1), \ldots, f(x_N)\}$ is the current known minimum
- How much lower is a new point?
- No improvement if the new point is not lower

Expected Improvement

$$\mathbb{E}\{I(x)\} = (f_{\min} - \hat{z}_N(x))\Phi\left(\frac{f_{\min} - \hat{z}_N(x)}{\hat{\sigma}_N(x)}\right) + \hat{\sigma}_N(x)\phi\left(\frac{f_{\min} - \hat{z}_N(x)}{\hat{\sigma}_N(x)}\right)$$

- Expectation is with respect to the predictive distribution of the surrogate
- Balances expected improvement because the predicted **mean** is small with expected improvement because the **variability** is large
- Can minimize iteratively by choosing next sampling location as the point with largest expected improvement (Jones et al., 1998)

Expected Improvement



Constrained Optimization

A more general optimization problem:

$$Z(x) = f(x) + \varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, \eta^2)$$

$$C(x) = c(x + \varepsilon_c) = \mathbb{I}_{\{x+\varepsilon_c \in C\}} \in \{0, 1\}$$

$$x^* = \arg \min_{x \in \{x : C(x) = 1\}} f(x)$$

- C is the constraint region
- Allows noise in both the function and the constraint
- Both f and c may be unknown, with f expensive to evaluate

Types of Constraints

- Known — c or C is known a priori
- Unknown — c is unknown, but f still returns a valid value
- Hidden — c is unknown, and f does not return a valid value for $x \notin C$

Known constraints present an obvious optimization strategy.
Hidden constraints are somewhat straightforward.

Fun with unknown constraints

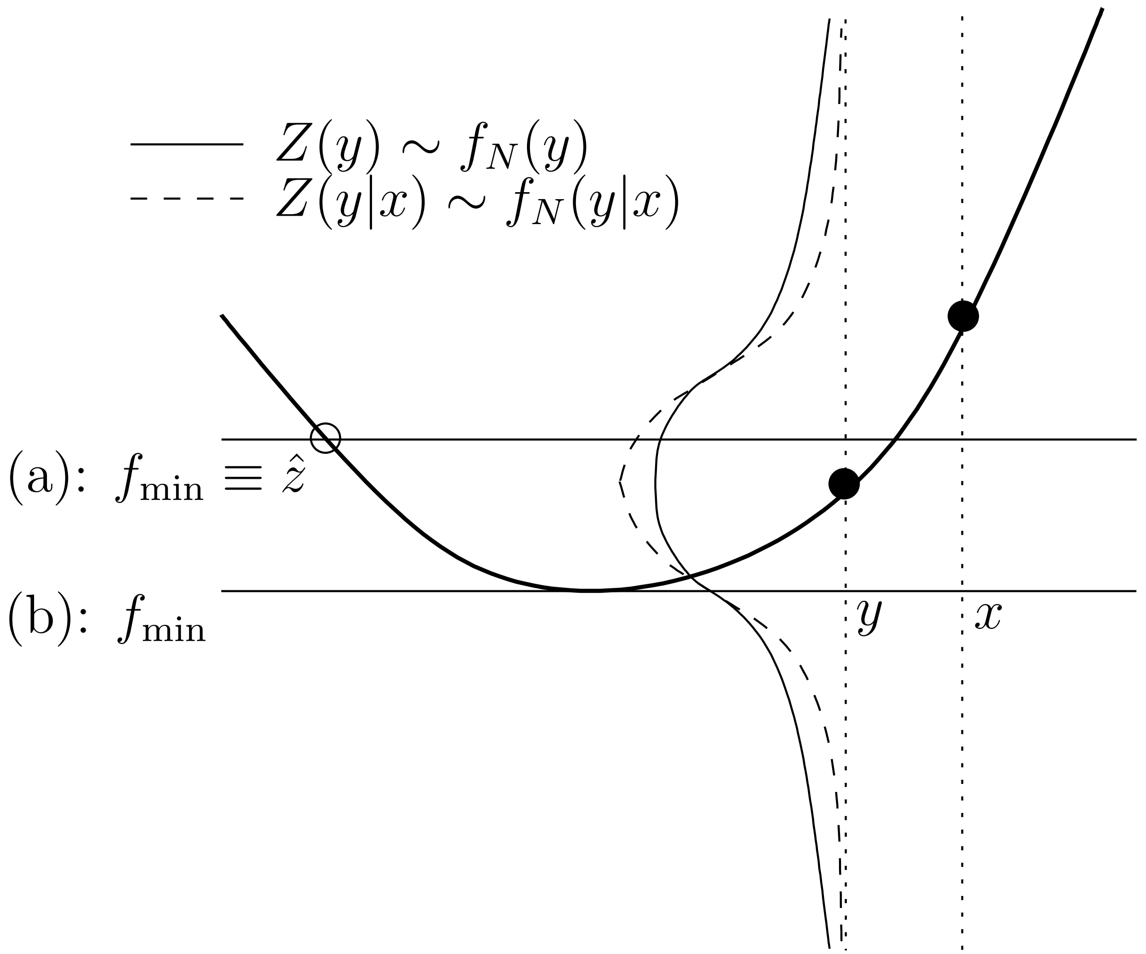
- Evaluating f for $x \notin C$ still yields information about f
- When do we learn enough about f to make it worth choosing a new point x which we think may not be in C?
- We need a criterion which quantifies how much we expect to learn about the minimum, which may be a point other than the one at which we are evaluating f

Conditional Improvement

$$I(y|x) = \max\{f_{\min} - Z(y|x), 0\}$$

- $f_{\min} = \min \mathbb{E}\{Z(\cdot)\}$ is now defined as the minimum of the posterior predictive over all locations
- $Z(y|x) \sim F_N(y|x)$, the predictive distribution of the response $Z(y)$ at a *reference* location y under the surrogate model f_N given that the *candidate* location x is added into the design
- $I(y) - I(y|x)$ measures how much we learn about $f(y)$ by evaluating $f(x)$
- Expected conditional improvement (ECI) is $\mathbb{E}\{I(y|x)\}$

Defining the minimum of a noisy function



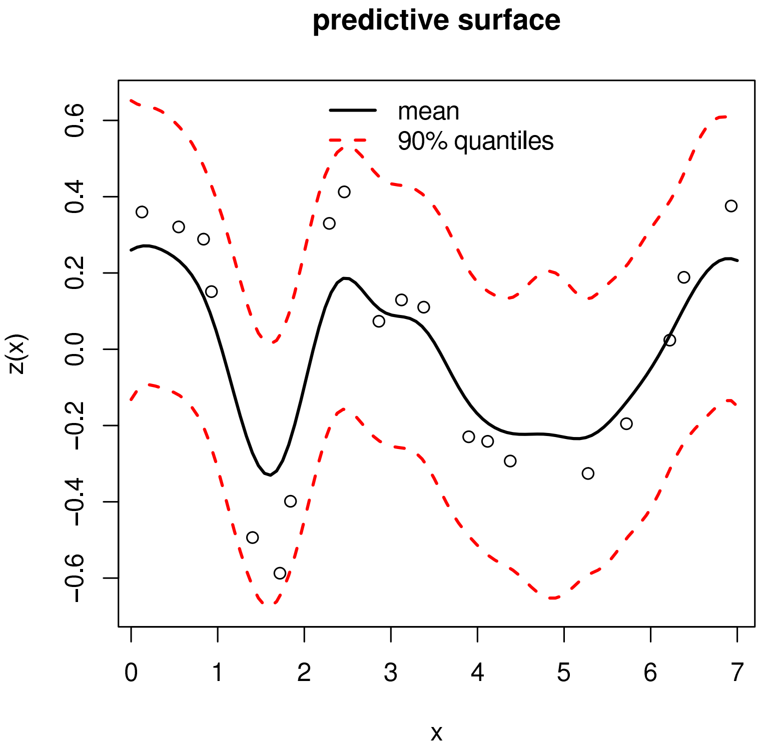
Integrated Expected Conditional Improvement

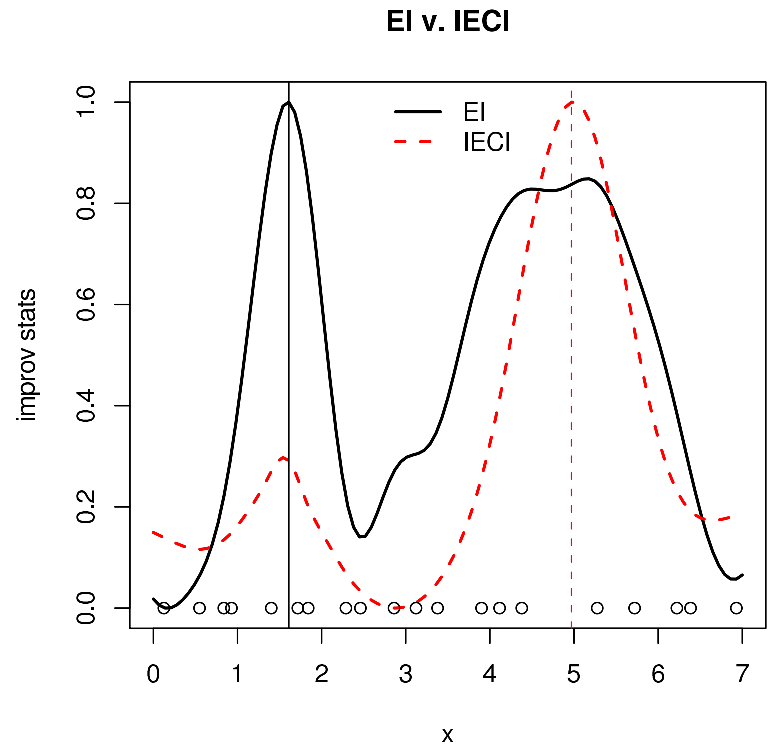
$$\begin{aligned}\mathbb{E}_g\{I(x)\} &= -\int_{\mathcal{X}} \mathbb{E}\{I(y|x)\} g(y)\, dy \\ \text{or}\ \ \mathbb{E}_g\{I(x)\} &= \int_{\mathcal{X}} (\mathbb{E}\{I(y)\} - \mathbb{E}\{I(y|x)\}) g(y)\, dy\end{aligned}$$

- Integrate w.r.t. density $g(y)$
  - Unconstrained, $g(y) \propto 1$ gives a general global criterion
  - For known constraints, could use $g(y) \propto C(y)$
  - For unknown constraints, $g(y) = \mathbb{P}(C(y) = 1)$
- Choose next point x to maximize IECI
- Estimate IECI via Monte Carlo approximation

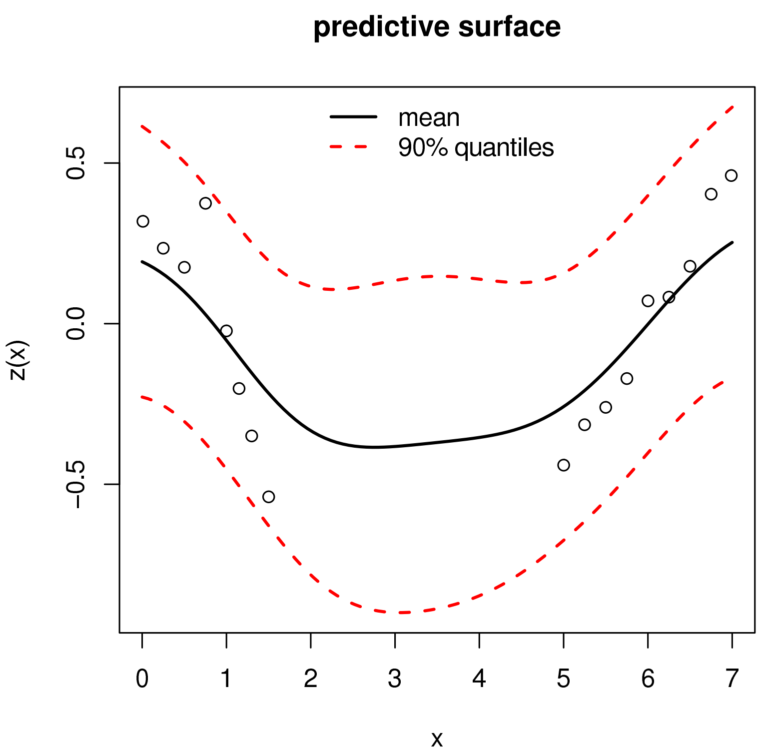
Example of EI vs. IECI

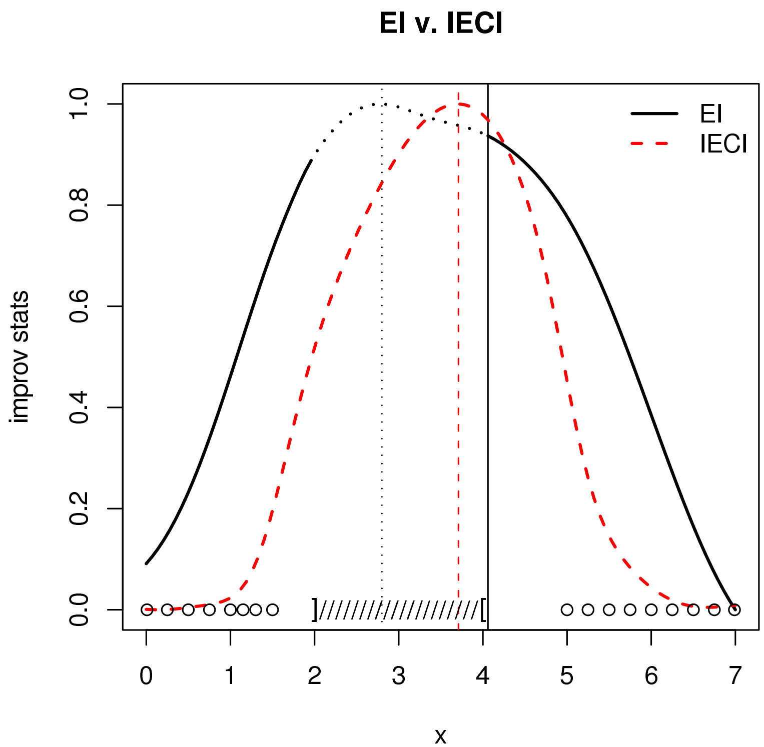
$$Z(x) = \sin(x) + 2.55\phi_{0.45}(x-3) + N(0, 0.15^2) \text{ for } x \in [0, 7]$$





EI vs. IECI with Constraint





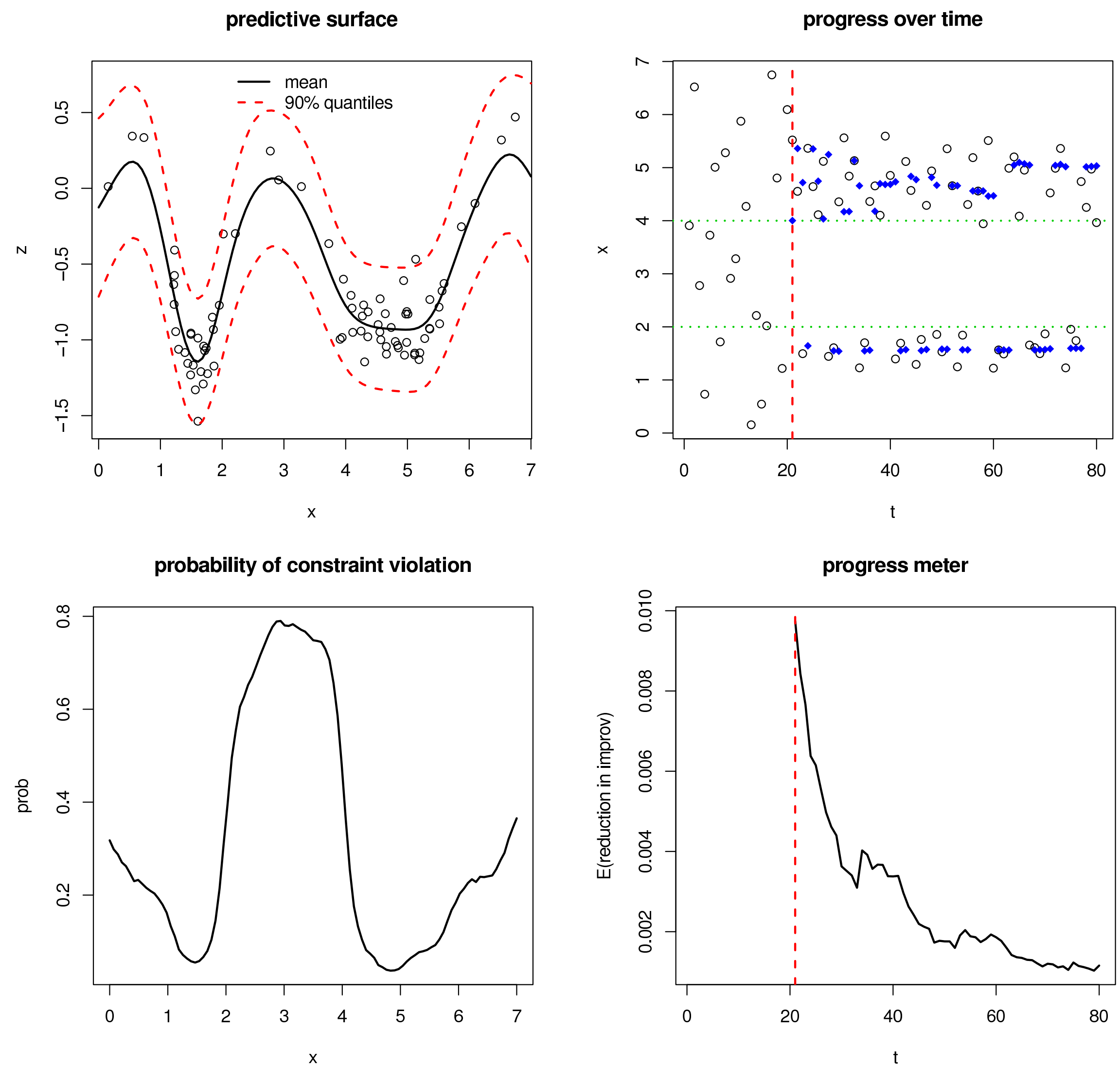
Finding the Minimum IECI

- Approximate search by looking at a Latin Hypercube design of randomly chosen candidate locations
- Evaluate IECI at candidate locations and find the minimum
- Also consider an "oracle" point at the minimum of the predictive mean

Estimating the Unknown Constraint

- Use a surrogate model for the probability $\mathbb{P}(C(y) = 1)$
- We use a classification GP
- Can fit with particle learning for efficient updates
- R package **plgp**

1-D Example



predictive surface
mean
90% quantiles
z
x
progress over time
x
t
probability of constraint violation
prob
x
progress meter
E(reduction in improv)
t

2-D example

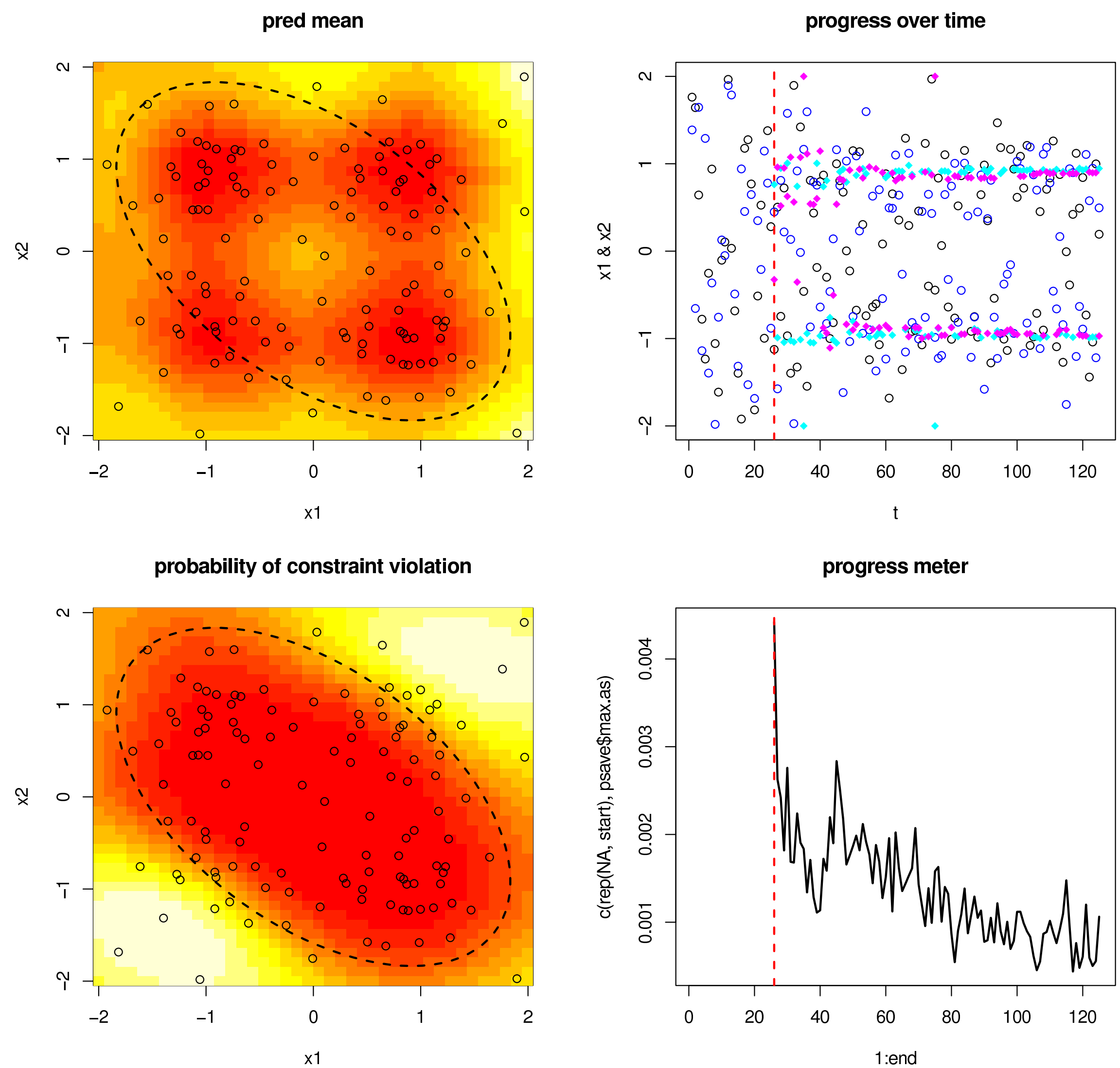
$$f(x_1, x_2) = -w(x_1)w(x_2), \qquad \text{where}$$

$$w(x) = \exp\left(-(x-1)^2\right) + \exp\left(-0.8(x+1)^2\right) - 0.05\sin\left(8(x+0.1)\right)$$

C is given by the 95% contour of a bivariate normal distribution centered at the origin, with correlation -0.5 and variance 0.75^2

Global minimum is at $(x_1, x_2) = (-1.408, -1.408) \notin C$

2-D Results



pred mean
x1
x2
progress over time
t
x1 & x2
probability of constraint violation
x1
x2
progress meter
1:end
c(rep(NA, start), psave$max.as)

Health Policy Example

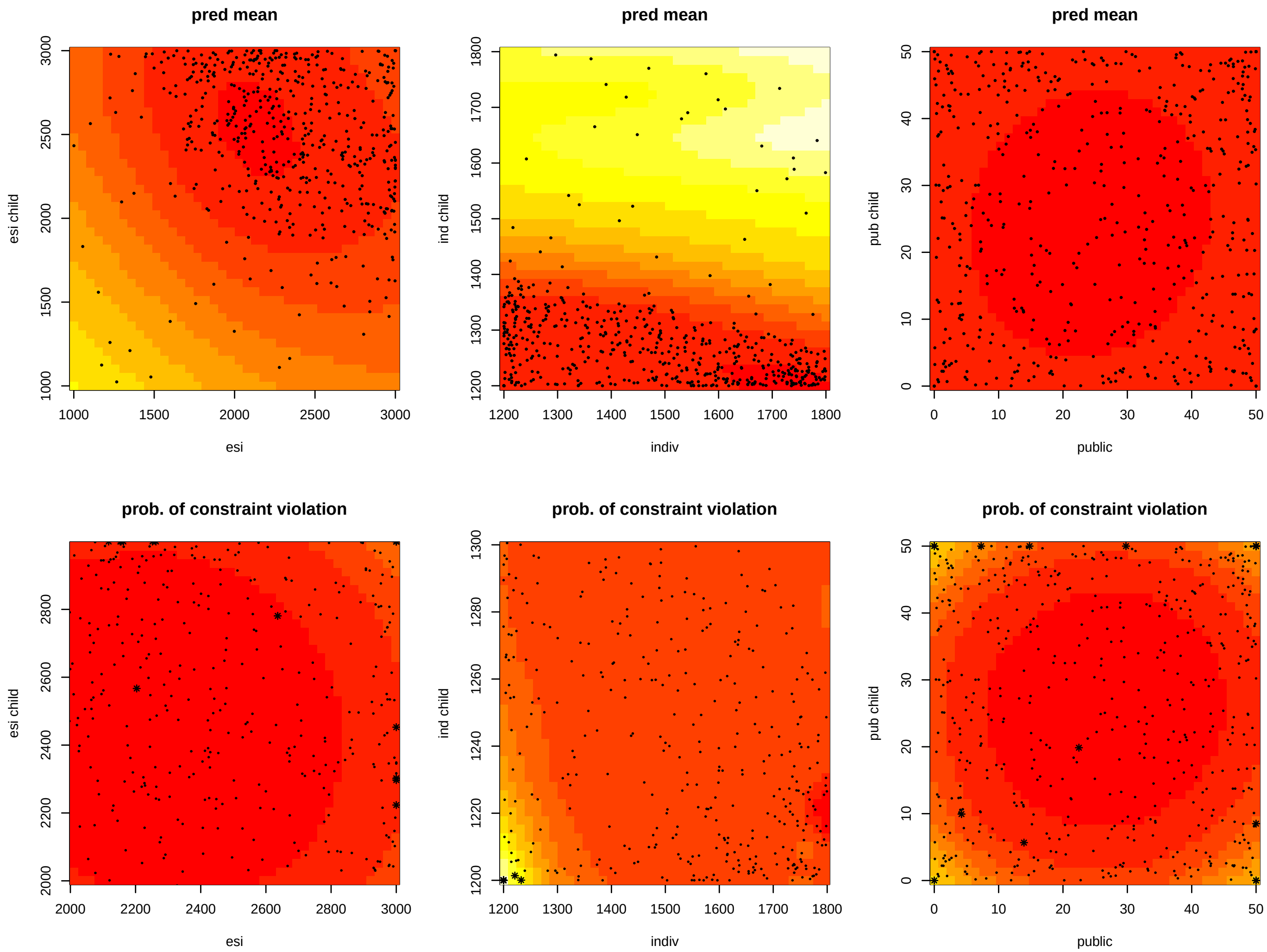
Application to a Health Care Policy Simulator (COMPARE) at RAND (Girosi et al., 2009)

- Microsimulation Model
- Agent-based simulator to model health insurance decisions of individuals
- Allows various policy interventions
- Utility maximization at the agent level
- Statistical issues in calibration and optimization
- Can be run deterministically or stochastically

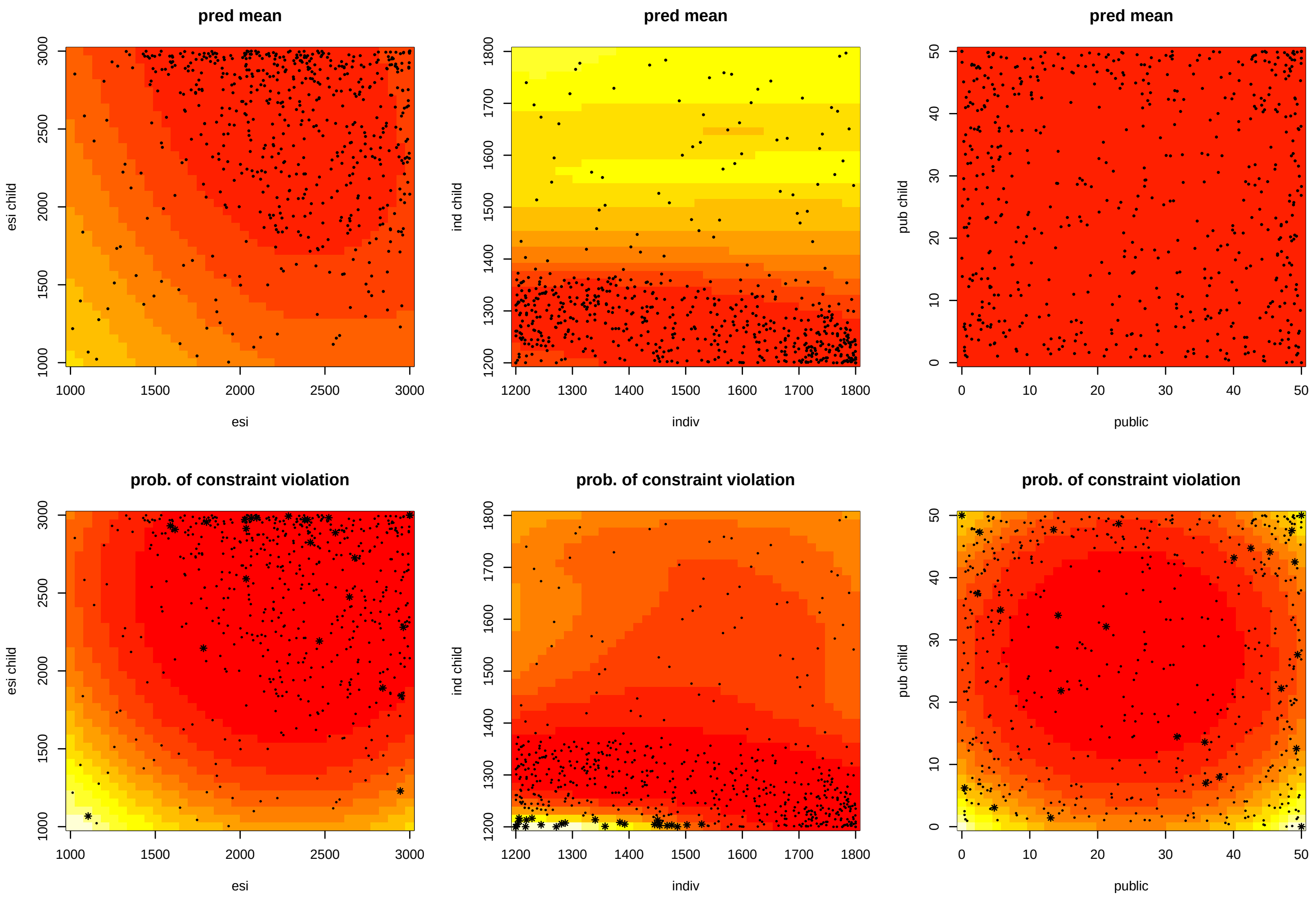
Calibration Details

- Six calibration parameters, relating to utilities for adults and for children on each of ESI, individual, and public programs
- Constraints relate to estimated elasticities of policy choice to price information
- Objective function is a combination of absolute errors in predicted counts and squares of predicted elasticities
$Z(\mathbf{x}) = \alpha_1 \sum_{j=1}^{4} |y_{aj} - \hat{y}_{aj}| + \alpha_2 \sum_{j=1}^{4} |y_{cj} - \hat{y}_{cj}| + \sum_{k=1}^{4} \alpha_{3k} y_{ek}^2 \mathbb{I}_{\{|y_{ek}|>1\}}$

Deterministic Model Calibration



Stochastic Model Calibration Results



Conclusions

- Constrained optimization can be quite difficult
- Use of statistical surrogate models can speed up optimization
- IECI is more general than EI
- Applies for deterministic and stochastic simulators

References

Girosi, F., Cordova, A., Eibner, C., Gresenz, C. R., Keeler, E., Ringel, J., Sullivan, J., Bertko, J., Buntin, M. B., and Vardavas, R. (2009). "Overview of the COMPARE Microsimulation Model." Tech. Rep. WR-650, RAND.

Gramacy, R. and Polson, N. (2010). "Particle learning of Gaussian process models for sequential design and optimization." Tech. Rep. arXiv:0909.5262, University of Cambridge.

Gramacy, R. B. and Lee, H. K. H. (2008). "Bayesian treed Gaussian process models with an application to computer modeling." *Journal of the American Statistical Association*, 103, 1119–1130.

— (2009). "Adaptive Design and Analysis of Supercomputer Experiment." *Technometrics*, 51, 2, 130–145.

Gramacy, R. B. and Taddy, M. A. (2010). "Categorical Inputs, Sensitivity Analysis, Optimization and Importance Tempering with `tgp` Version 2, an R Package for Treed Gaussian Process Models." *Journal of Statistical Software*, 33, 6, 1–48.

Jones, D., Schonlau, M., and Welch, W. J. (1998). "Efficient Global Optimization of Expensive Black Box Functions." *Journal of Global Optimization*, 13, 455–492.

Lee, H., Gramacy, R., Linkletter, C., and Gray, G. (2010). "Optimization Subject to Hidden Constraints via Statistical Emulation." Tech. Rep. UCSC-SOE-10-10, University of California, Santa Cruz, Department of Applied Mathematics and Statistics.

Santner, T. J., Williams, B. J., and Notz, W. I. (2003). *The Design and Analysis of Computer Experiments*. New York, NY: Springer-Verlag.

Taddy, M., Lee, H. K. H., Gray, G. A., and Griffin, J. D. (2009). "Bayesian Guided Pattern Search for Robust Local Optimization." *Technometrics*, 51, 389–401.