Bayesian Experimental Design for Stochastic Dynamical Models

Gavin J Gibson
School of Mathematical & Computer Sciences and Maxwell Institute for Mathematical Sciences, Heriot-Watt University, Edinburgh, UK

Joint work with: Alex Cook, Luis Carrasco, *Dept. Stats, NUS, Singapore*, Chris Gilligan, *Department of Plant Sciences, Cambridge University, UK*

Acknowledgements: BBSRC

Outline

- Bayesian inference/epidemic models
- Microcosm experiments
- Design problem
- Bayesian approach
- Approximations and implementation
- Results

Stochastic epidemic model (SEIR):
Finite population size N, partitioned into susceptible (S), exposed (E), infectious (I) and removed (R) classes.

$S \rightarrow E$: If j is in state S at time t, then

$$\Pr(j \text{ is exposed in } (t, t+dt)) = \beta I(t)dt$$

$E \rightarrow I$: $T_E^j \sim \pi_{\theta_E}^E$ (random time in E)

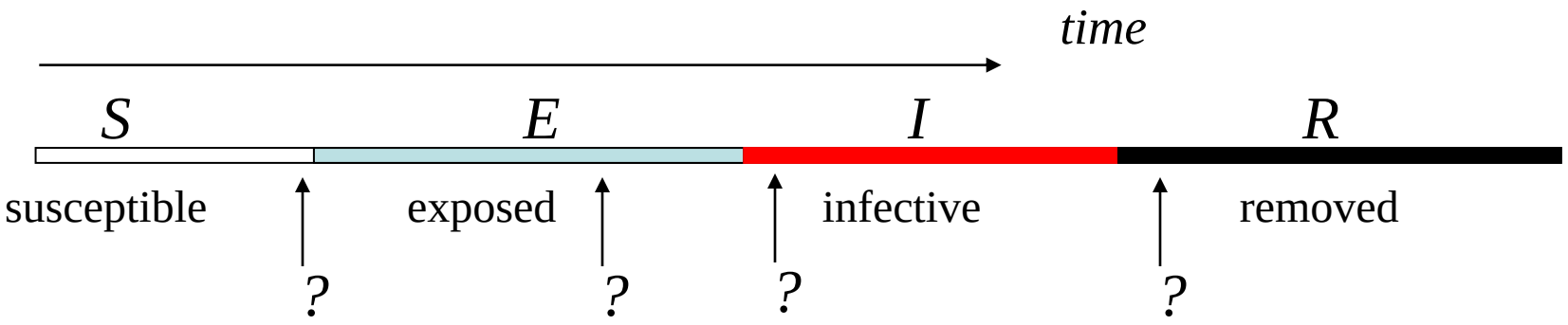
$I \rightarrow R$: $T_I^j \sim \pi_{\theta_I}^I$ (random time in I)

Parameters: $\theta = (\beta, \theta_E, \theta_I)$

Partially observed epidemics

SEIR model for an epidemic in closed, homogeneously mixing population:

History of individual i.



Diagnostic tests:

- *Limited frequency of tests*
- *Certain states indistinguishable* (e.g. *S* & *E*)
- *False negative and positive results*

Bayesian inference in a nutshell…

Given $\boldsymbol{y}$, likelihood principle says all evidence about θ contained in likelihood (assuming model)

$L(\boldsymbol{\theta}|\boldsymbol{y}) = \Pr(\boldsymbol{y}|\boldsymbol{\theta})$ *"probability of the observations given* $\boldsymbol{\theta}$*"*.

Bayesian approach. Represent prior beliefs on $\boldsymbol{\theta}$ as density $\pi(\boldsymbol{\theta})$. Update in light of $\boldsymbol{y}$ using Bayes' Theorem

$\pi(\boldsymbol{\theta}|\boldsymbol{y}) \propto \pi(\boldsymbol{\theta})L(\boldsymbol{\theta}|\boldsymbol{y})$.

When $\boldsymbol{\theta}$ *multivariate, inference on individual parameters made using marginal posterior density.*

Bayesian inference for epidemic models

•Let ***y*** represent observed data, ***z*** represent complete data (times and nature of all events).

•Pr(***z***|θ) usually tractable, but

$$\Pr(y \mid \theta) = \int \Pr(z \mid \theta) dz$$

Integrate over all z consistent with y

Solution: Consider 'hidden' aspects, ***x***, as additional unknown parameters. Investigate the joint posterior density $\pi(\theta, \boldsymbol{x} \mid \boldsymbol{y})$.

Now, by Bayes' Theorem

$$\pi(\theta, \boldsymbol{x} \mid \boldsymbol{y}) \propto \pi(\theta)\pi(\boldsymbol{x}, \boldsymbol{y} \mid \theta)$$

Likelihood for augmented data (tractable)

Use Markov chain methods to generate samples from $\pi(\boldsymbol{\theta}, \boldsymbol{x}|\boldsymbol{y})$, where

$$\pi(\boldsymbol{\theta}, \boldsymbol{x}\,|\boldsymbol{y}) \propto \pi(\boldsymbol{\theta})\pi(\boldsymbol{x}, \boldsymbol{y}|\theta)$$

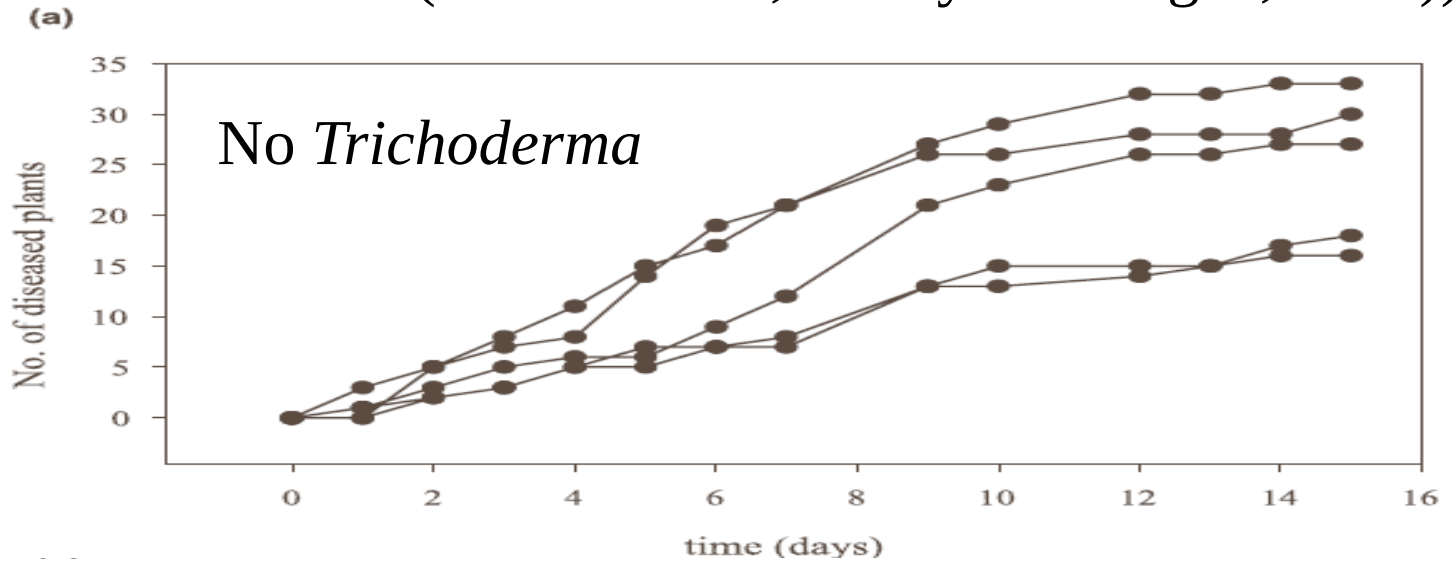
- Construct Markov chain with this stationary distribution.
- Iterates by proposing & accepting/rejecting changes to the current state $(\boldsymbol{\theta}_i, \boldsymbol{x}_i)$ to obtain $(\boldsymbol{\theta}_{i+1}, \boldsymbol{x}_{i+1})$.

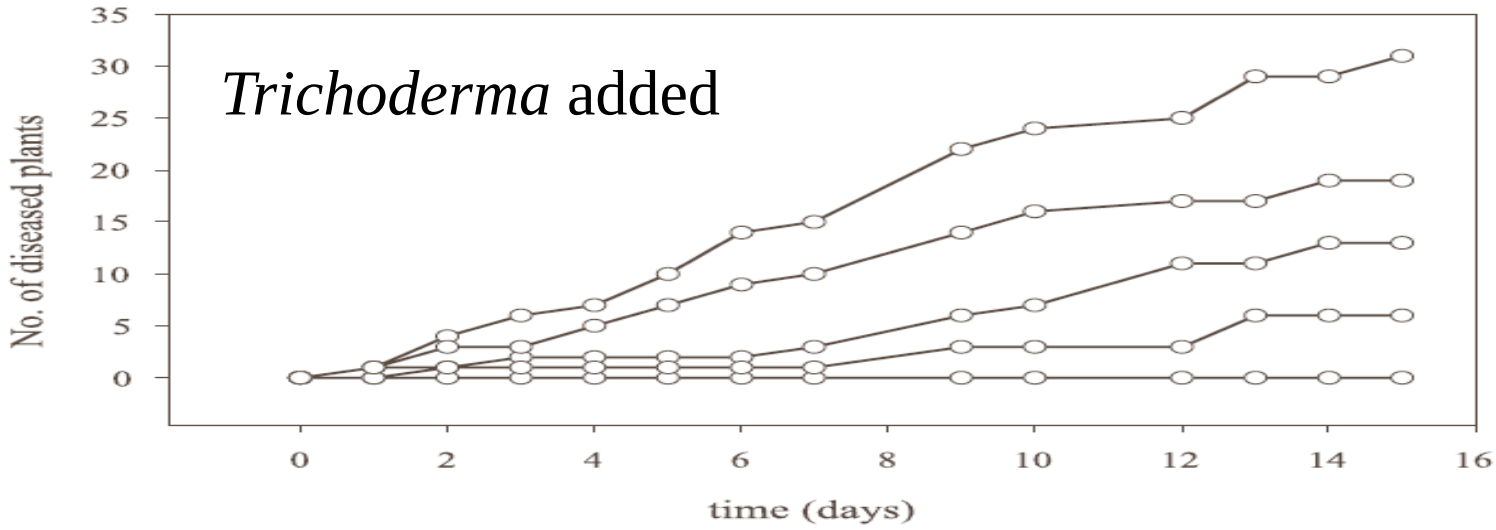
Updates to (some components of) $\boldsymbol{\theta}$ can often be carried out by Gibb's steps.

Updates to $\boldsymbol{x}$, usually require M-H and RJ type approaches.
For recent examples see Forrester *et al.* (2007), Chis-Ster & Ferguson (2009), Jewell & Roberts (2009).

Example: Microcosm experiments (N = 50)

R solani in radish (Kleczkowski, Bailey & Gilligan, 1996))





How does Trichoderma affect the system?

Simple model: SEI with primary, secondary infection + quenching (Gibson *et al.*, PNAS, 2004)

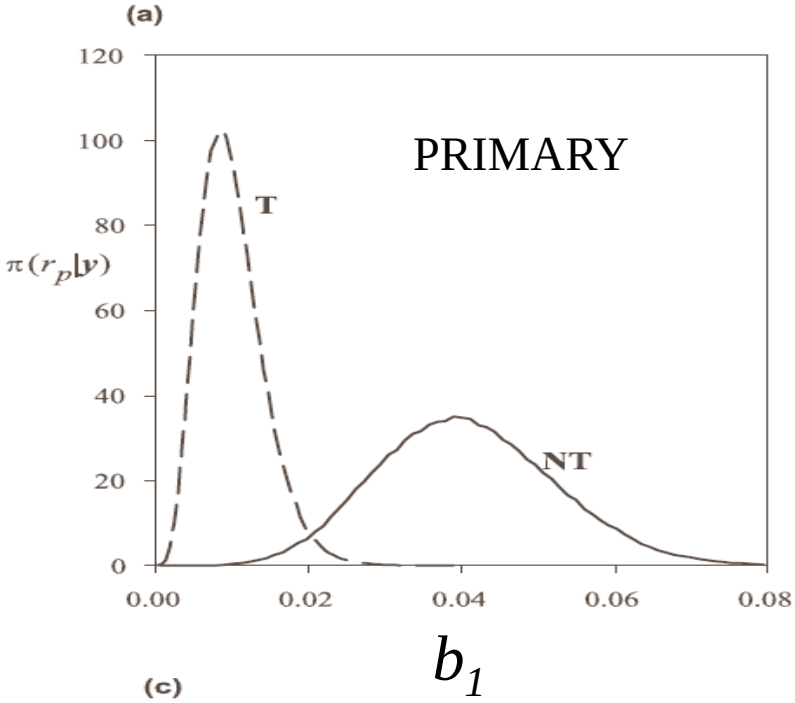
For susceptible individual at time t, the probability of becoming exposed in the time interval $[t, t+\mathrm{d}t)$ is

$$R(t)\mathrm{d}t = (b_1 + b_2 I(t))\exp\{-rt\}\mathrm{d}t.$$

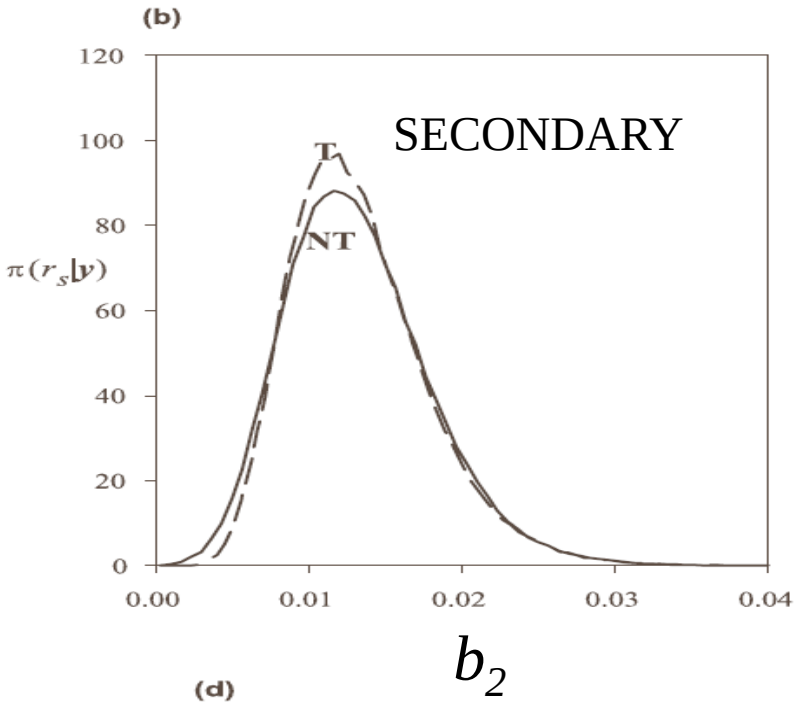
Exposed individuals become infectious at fixed time μ after exposure.

Using **vague uniform priors** we obtain posteriors for the model parameters. These suggest *Trichoderma* acts on primary infection process.

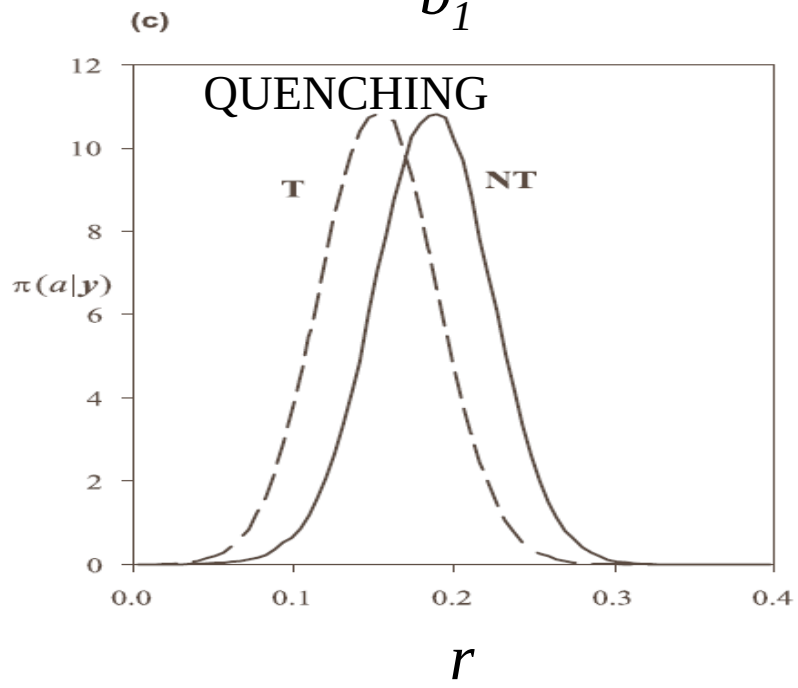
HERIOT WATT UNIVERSITY


(a)
PRIMARY
T
NT
$\pi(r_p|y)$
120
100
80
60
40
20
0
0.00
0.02
0.04
0.06
0.08

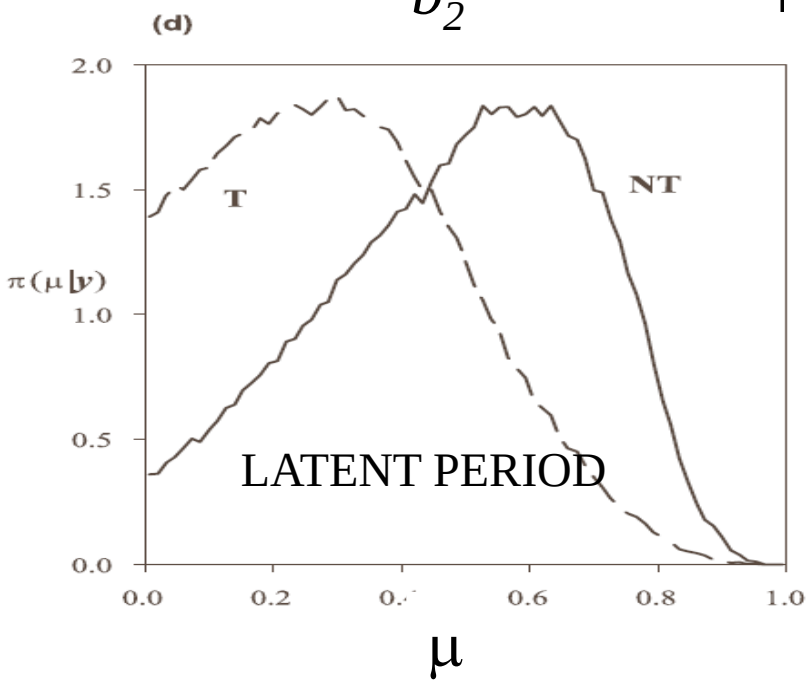
b_1


(b)
SECONDARY
T
NT
$\pi(r_s|y)$
120
100
80
60
40
20
0
0.00
0.01
0.02
0.03
0.04

b_2

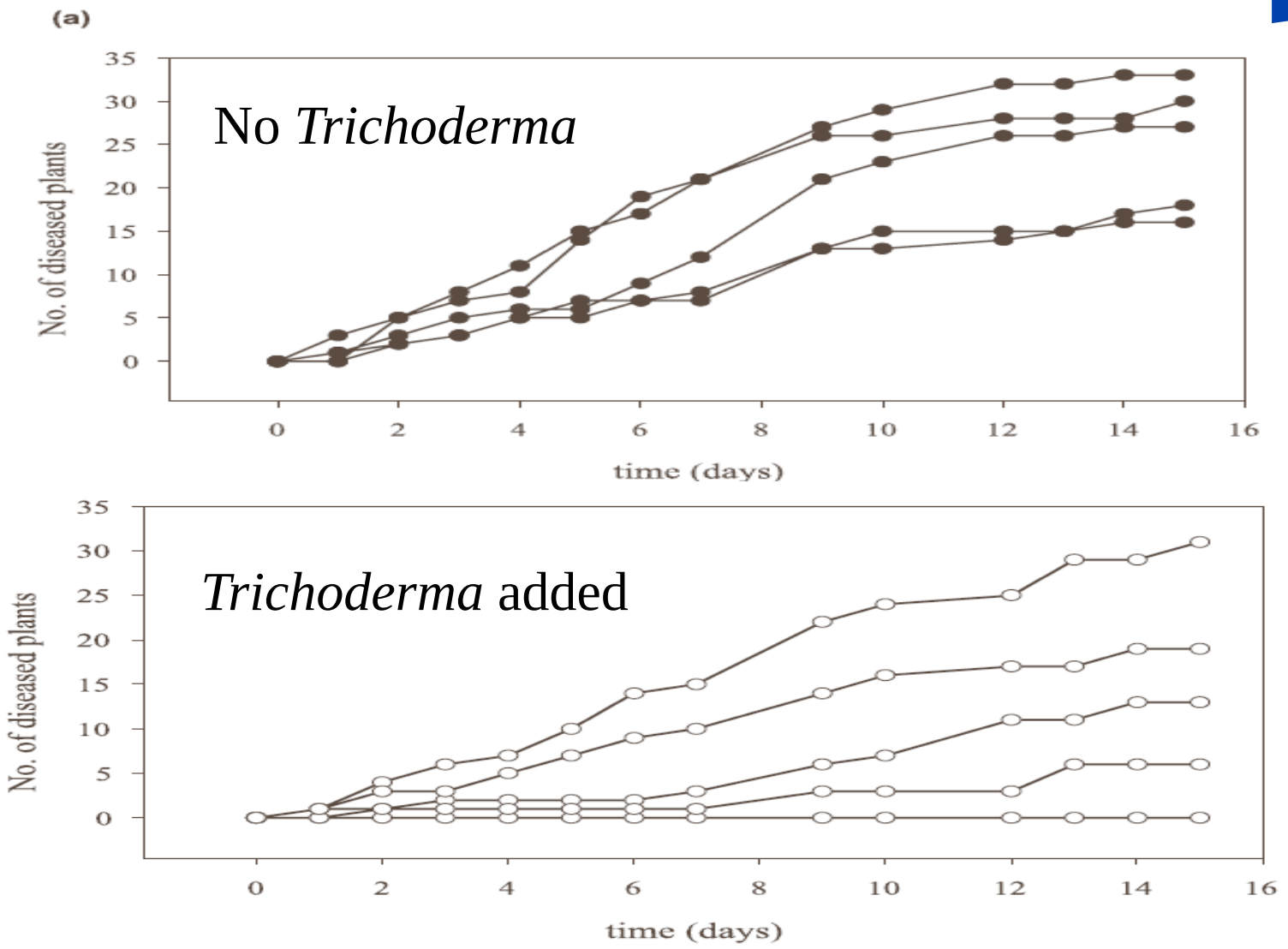

(c)
QUENCHING
T
NT
$\pi(a|y)$
12
10
8
6
4
2
0
0.0
0.1
0.2
0.3
0.4

r


(d)
T
NT
LATENT PERIOD
$\pi(\mu|y)$
2.0
1.5
1.0
0.5
0.0
0.0
0.2
0.6
0.8
1.0

μ

Sampling was intensive in this experiment



How should we design future experiments on this system?

Identifying optimal 'sparse' sampling schemes

Bayesian framework (Muller, 1999) for identifying optimal designs.

$\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta})$ - current belief regarding $\boldsymbol{\theta}$.

$\boldsymbol{z}' \sim \pi(\boldsymbol{z}' | \boldsymbol{\theta})$ – 'complete' future realisation of process.

$d(\boldsymbol{z}')$ – censored/filtered version of $\boldsymbol{z}'$ arising from design d chosen from some suitable space of designs.

$U(d(\boldsymbol{z}'))$ utility function quantifying usefulness of $d(\boldsymbol{z}')$

Aim: Select d to maximise the expectation of $U(d(\boldsymbol{z}'))$

Optimal designs as posterior modes

HERIOT WATT UNIVERSITY

•Treat the design, d, as an additional variable and assign joint distribution to $(\boldsymbol{\theta}, \mathbf{z}', d)$ with density

$$\phi(\boldsymbol{\theta}, \mathbf{z}', d) \propto \pi(\boldsymbol{\theta})\pi(\mathbf{z}'|\boldsymbol{\theta})U(d(\mathbf{z}'))$$

•Integrating *w.r.t.* $\mathbf{z}'$, to obtain $\phi(d|\boldsymbol{\theta}) \propto E(U(d)|\boldsymbol{\theta})$, the expectation of $U(d)$ conditional on $\boldsymbol{\theta}$.

•Integrating with respect to $\boldsymbol{\theta}$ gives $\phi(d) \propto E(U(d))$.

•Optimal design is mode of $\phi(d)$. **In theory** we could investigate ϕ using MCMC methods.

Potential difficulties….

Natural for Bayesian to base utility on the posterior density of parameters.

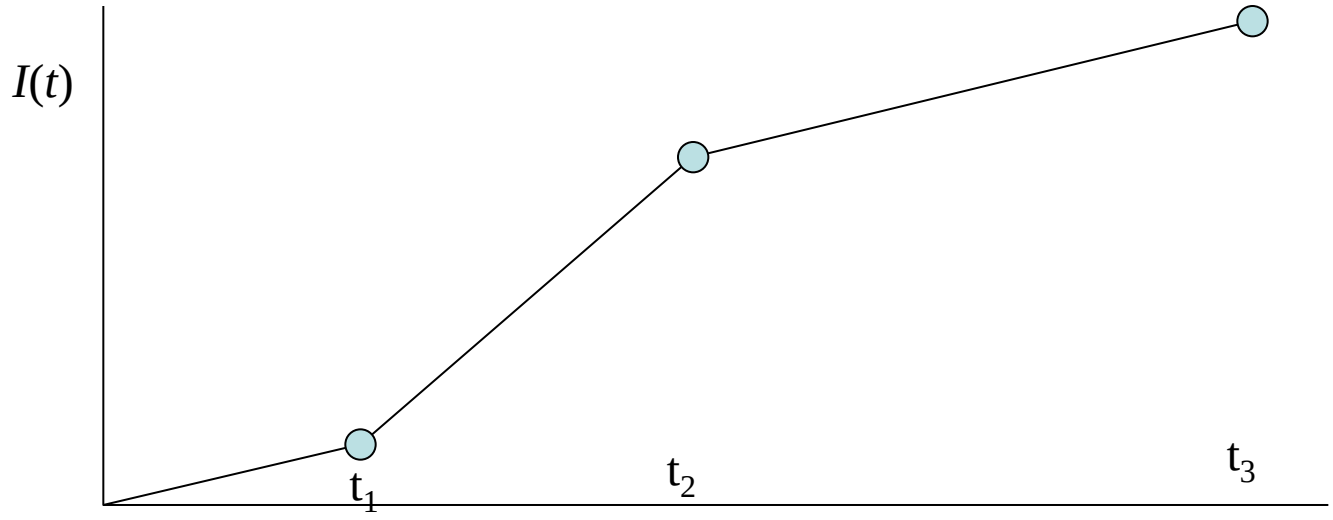
Example: Kullback-Leibler divergence between prior and posterior. Let $\boldsymbol{y} = d(\boldsymbol{z}')$

May be nasty!

Then $U(d(\boldsymbol{z}')) = \int \pi(\boldsymbol{\theta} \mid \boldsymbol{y}) \log\left(\frac{\pi(\boldsymbol{\theta} \mid \boldsymbol{y})}{\pi(\boldsymbol{\theta})}\right) d\boldsymbol{\theta}$

Recall $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ is usually difficult to obtain – particularly so if sampling infrequent.

Consider problem of selecting $m = 3$ observation times for microcosm experiments. Focus on a **simpler model** in which the latent period is assumed to be zero.



For this experiment $\boldsymbol{y} = (I(t_1), I(t_2), I(t_3))$, $\boldsymbol{\theta} = (b_1, b_2, r)$.

Aim to approximate $\int \pi(\boldsymbol{\theta} \mid \boldsymbol{y}) \log\left(\frac{\pi(\boldsymbol{\theta} \mid \boldsymbol{y})}{\pi(\boldsymbol{\theta})} \right) d\boldsymbol{\theta}$

First approximation: Approximate prior belief with a discrete parameter space by drawing sample from π.

Second approximation: Approximate $L(\theta)$ using moment-closure methods

Basic idea: $\mathrm{L}(\theta) = \mathrm{P}(I(t_1) = y_1|\theta)\times\mathrm{P}(I(t_2) = y_2 \mid I(t_1) = y_1, \theta)$
$\times\ \mathrm{P}(I(t_k) = y_k \mid I(t_{k\text{-}1}) = y_{k\text{-}1}, \theta)$ …….

Approximate *each term* using moment-closure (*e.g.* Krishnarajah *et al.*, 2005). For $t > t_i$, $J(t) = I(t) - I(t_i)$.

Set up system of ODEs for evolution of moments of $J(t)|I(t_i)$.

Resulting system:

$$\frac{d}{dt}E(J(t)|I(t_i)) = \alpha_1(\boldsymbol{\theta}, I(t_i)) + \beta_1(\boldsymbol{\theta}, I(t_i))E(J(t)|I(t_i)) + \gamma_1(\boldsymbol{\theta}, I(t_i))E(J^2(t)|I(t_i))$$

$$\frac{d}{dt}E(J^2(t)|I(t_i)) = \alpha_2(\boldsymbol{\theta}, I(t_i)) + \beta_2(\boldsymbol{\theta}, I(t_i))E(J(t)|I(t_i)) + \gamma_2(\boldsymbol{\theta}, I(t_i))E(J^2(t)|I(t_i)) + \delta_2(\boldsymbol{\theta}, I(t_i))E(J^3(t)|I(t_i))$$

Due to nonlinearity, DE for moment *j* contains term of higher order. **Q:** How can we close the system?

A: Assume a particular distributional form.

Here we assume that

$$J(t)|I(t_i) \sim \text{BetaBin}(S(t_i), \alpha(t), \beta(t)).$$

$\alpha(t)$ and $\beta(t)$ are determined by the first two moments.

$E(J^3(t))$ is then a function of $\alpha(t)$ and $\beta(t)$, and hence the first two moments. Substituting this function into the differential equation for $E(J^2(t))$ allows the system to be closed and easily solved numerically.

Evidence (Krishnarajah *et al.*, BMB 2005) indicates that the BetaBin distribution provides a good approximation for the SI model considered here.

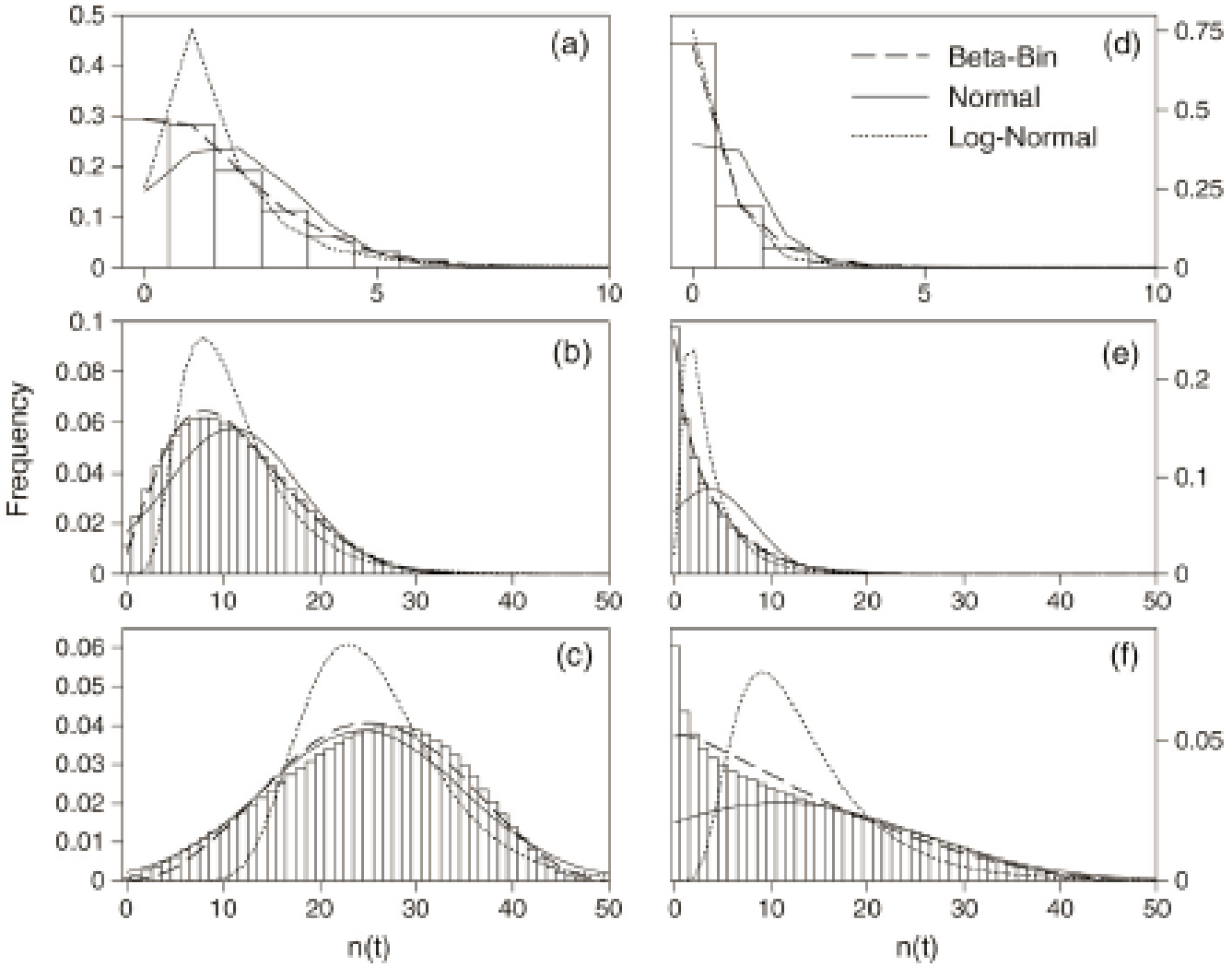
Comparison of moment-closure estimate with exact probability function.

No *Trichoderma* *Trichoderma*

$t = 1$

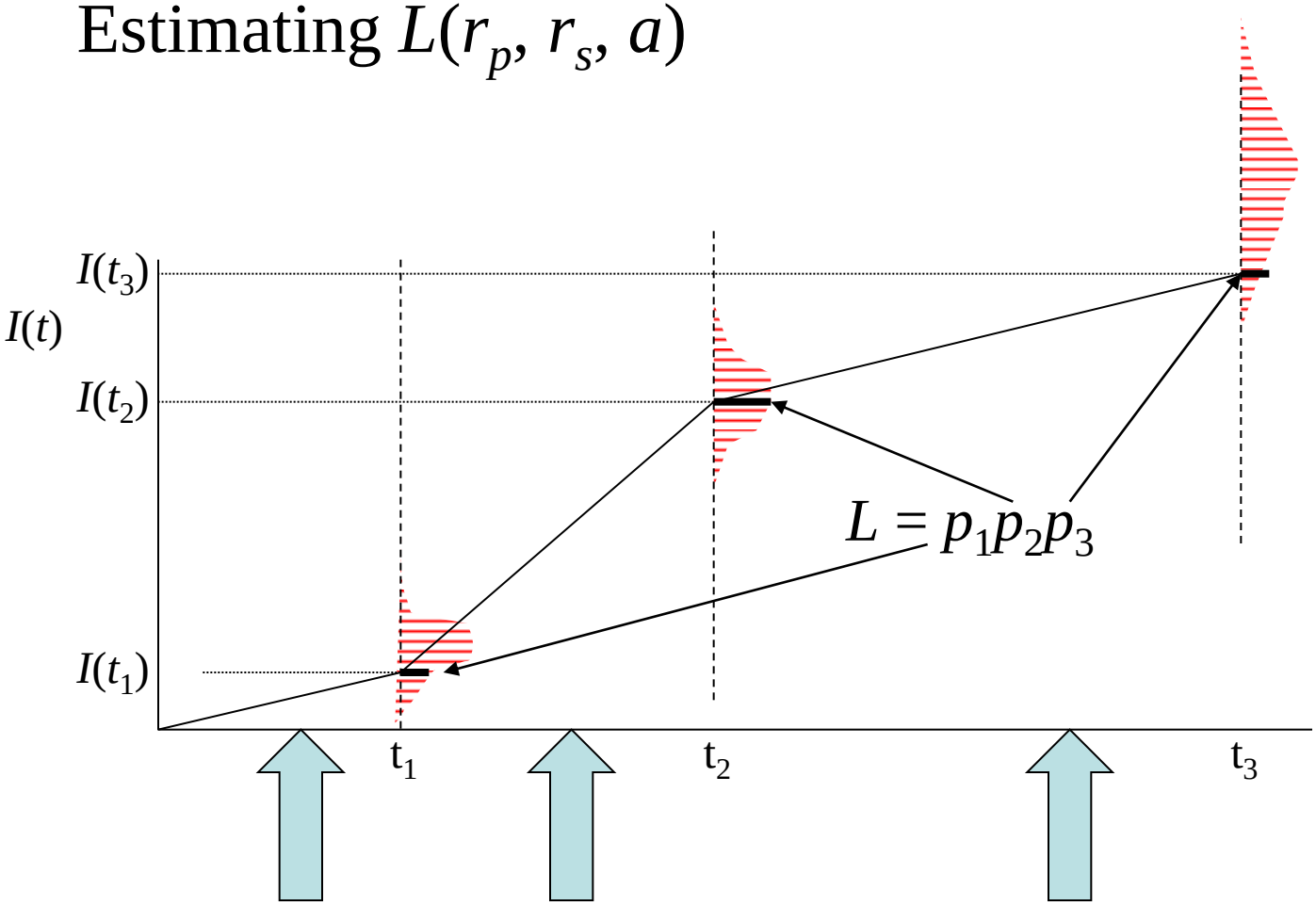
$t = 5$

$t = 15$



From Krishnarajah et al (2005)

Estimating $L(r_p, r_s, a)$



Integrate ODE to get distribution for $I(t_1)$, given I(0) = 0.

Integrate ODE to get distribution for $I(t_2)$, given $I(t_1)$.

Integrate ODE to get distribution for $I(t_3)$, given $I(t_2)$.

Full algorithm for experimental design

Recall we wish to draw from joint density

$$\phi(\boldsymbol{\theta}, \mathbf{z}', d) \propto \pi(\boldsymbol{\theta})\pi(\mathbf{z}'|\boldsymbol{\theta})U(d(\mathbf{z}'))$$

where d represents a set of m distinct sampling times arranged between $t = 0$ and t_{max}.

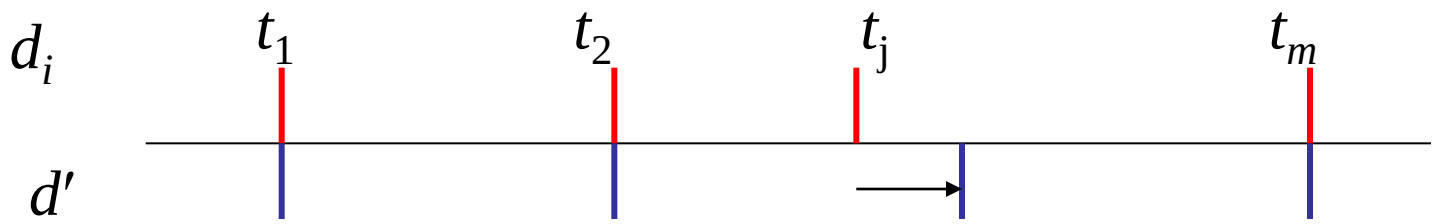
$\boldsymbol{\theta} = (r_p, r_s, a)$ and is a priori uniform over sample from continuous π.

$\mathbf{z}'$ represents complete process and comprises the infection times for all individuals in the population.

Outline of steps for updating $(\theta_i, \mathbf{z}_i, d_i)$

1. Propose new $(\theta', \mathbf{z}')$ by drawing θ' from the prior and simulating realisation $\mathbf{z}'$ from the model.
   Accept with probability $min\{1, U(d_i(\mathbf{z}'))/U(d_i(\mathbf{z}_i))\}$, otherwise **reject**.

2. Propose changes to sampling times *e.g.* using M-H methods. Proposed d' can be d_i with perturbation applied to one of sampling times.



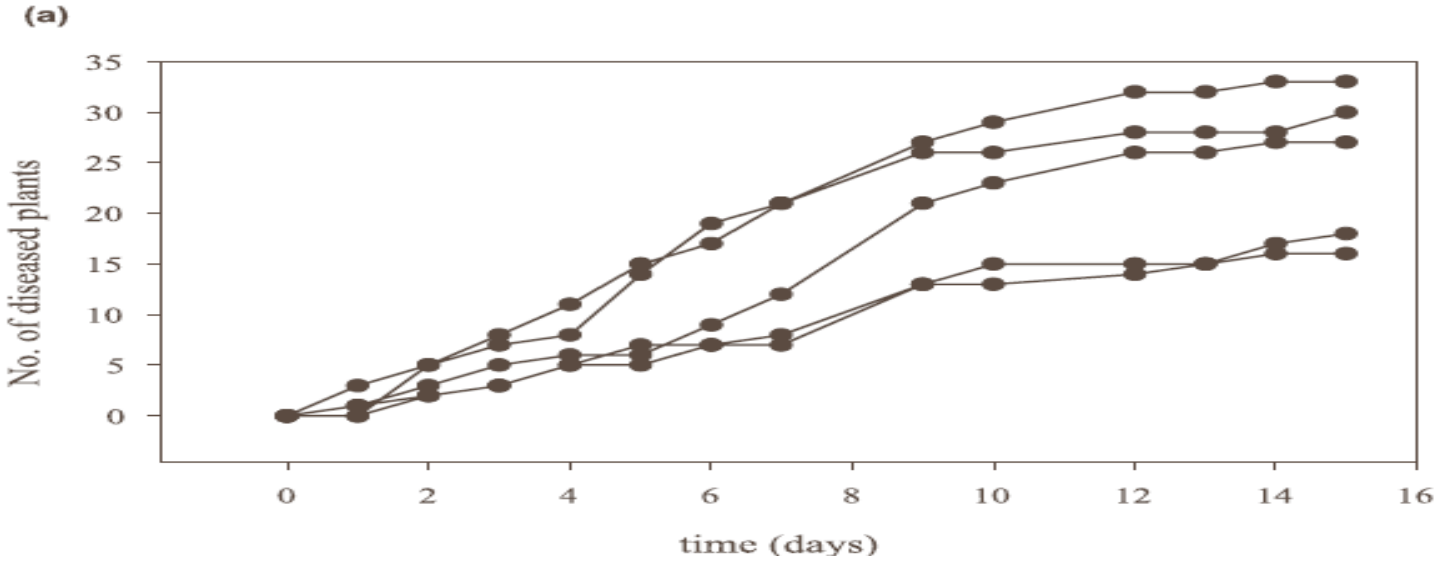
Accept with prob. $\min\{1, U(d'(\mathbf{z}_{i+1}))/U(d_i(\mathbf{z}_{i+1}))\}$

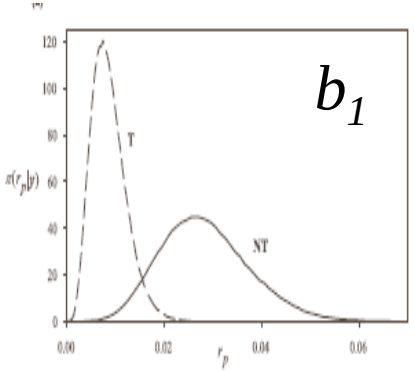
Some points to note:

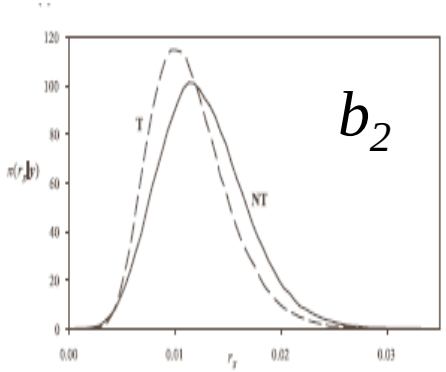
•The algorithm identifies optimal design for a single-replicate experiment. Optimal designs for multi-replicate experiments look broadly similar.

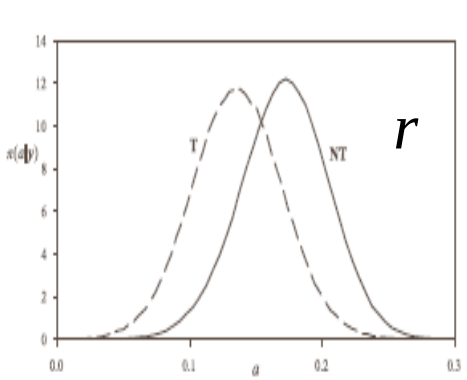
•The utility can be 'sharpened' by adapting the algorithm to propose k independent $(\boldsymbol{\theta}, \boldsymbol{z})$ combinations. Now $\phi(\boldsymbol{d}) \propto [E(U(d(\boldsymbol{z}))]^k$.

Designing future experiments for the *R solani* system (without *Trichoderma*) (Cook *et al*, *Biometrics*, 2008)**.**









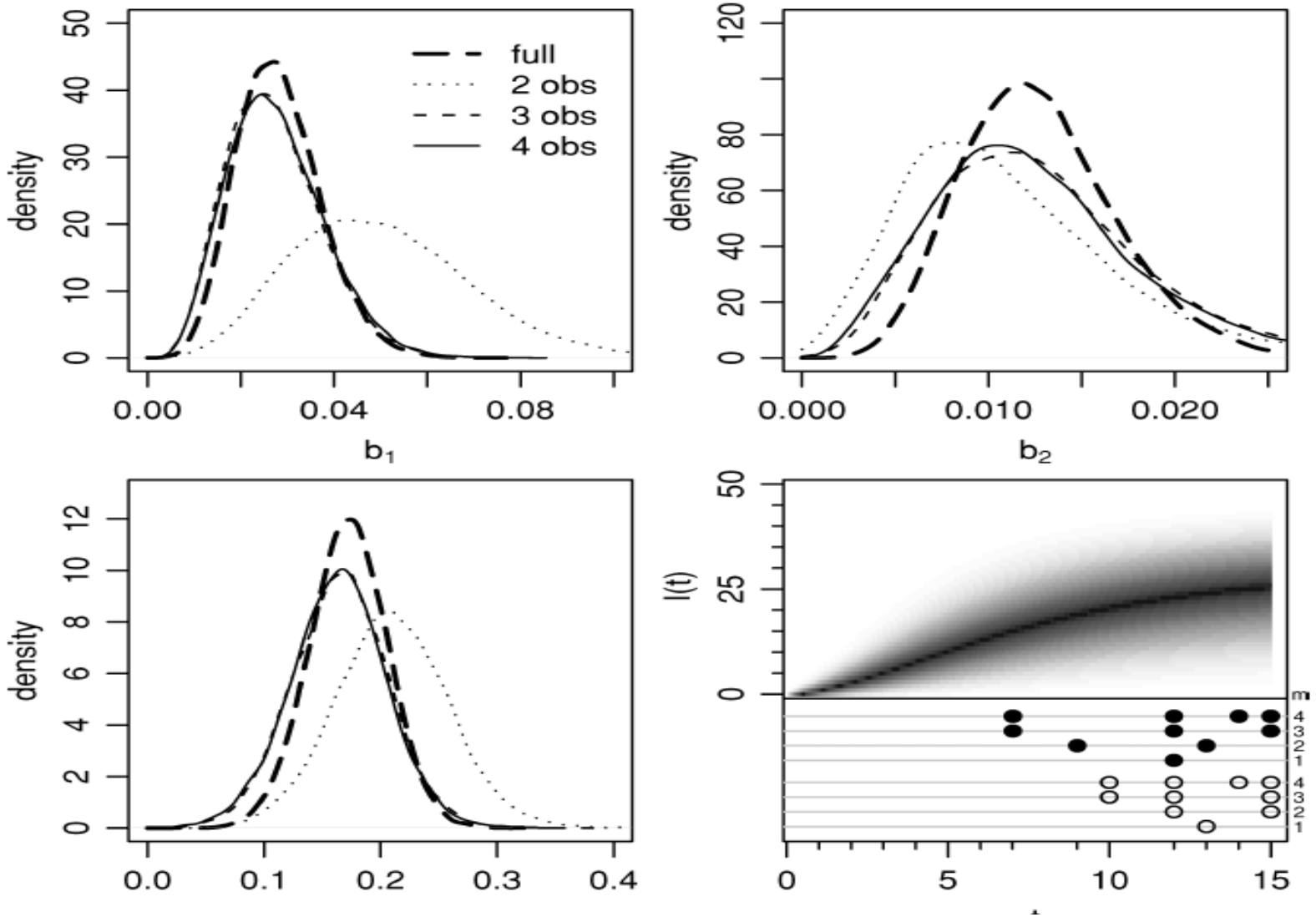
Sample from joint posterior gives new prior, π'.

Two scenarios considered….

Progressive design: How should you repeat experiment to maximise expected information change *w.r.t.* π'? (Uses π' to propose new $\boldsymbol{z}$'s and π' *as prior* in utility calculation.)

Confirmatory (pedagogic) design: How should you repeat experiment to maximise expected change w.r.t. π? (Uses π' to propose new $\boldsymbol{z}$'s and π *as prior* in utility calculation.)

Designs constrained to be subset of sampling times in original experiment.



Marginal posteriors for b_1, b_2, r using optimal pedagogic designs {9, 13}, {7, 12, 15}, {7, 12, 14, 15} and full set of sampling times. Bottom right shows optimal pedagogic (black) and progressive (white) designs and distribution of epidemic curves.

Extensions and current work (Cook *et al*, in prep)

- More general utility functions, taking account of cost of experiment
- Number of replicates as a design parameter
- More efficient approximations
- Use of direct simulation approach
- Larger experimental systems
- Imperfect diagnostic tests

Design of further rhizoctonia experiments

Population size: 160 (v. 50)

Model: SI with 'quenching' as before

Design parameters: Number of replicates ($N_r \leq 4$), sampling times (t_1, t_2, …, t_k), restricted to (2, 4, 6, 8, …., 16)

Simplifying assumptions: Observation times the same for each replicate, restricted to times {2, 4, 6, …, 16}, diagnostic tests assumed perfect (initially)

Cost and Information Gain

Cost of experiment:

$$C(d) = 0.25 \times k \times N_r + N_r$$

N_r = number of replicates, k = number of sampling times
(Cost of setting up replicate equivalent to 4 observations)

Benefit:

$$b(X_D, D, B) = \log\det V(\theta, B) - \log\det V(\theta \mid X_D, D, B)$$

Where $V(\theta,..)$ denotes the covariance matrix of θ, D denotes the design and B prior beliefs.

Alternative approximations to likelihood:

Based on gamma approximation to generalised Erlang distribution

$T_{j,i}$ = time of j^{th} infection after t_i.

For chosen model

$T_{j,i} \sim \text{Exp}(\lambda_1) + \text{Exp}(\lambda_2) \ldots. + \text{Exp}(\lambda_j)$ (*independent*)

Approximate with Gamma of same mean and variance.

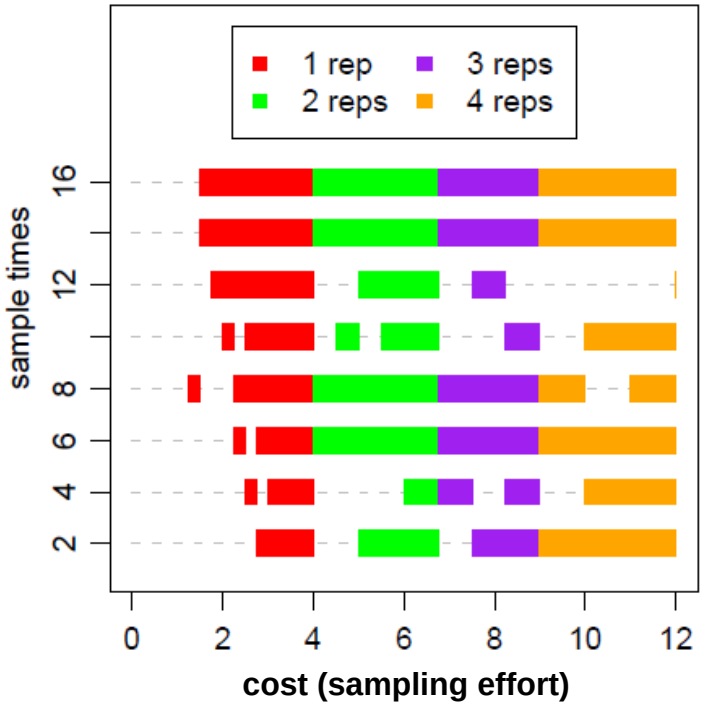
Some preliminary results using direct sampling algorithm:

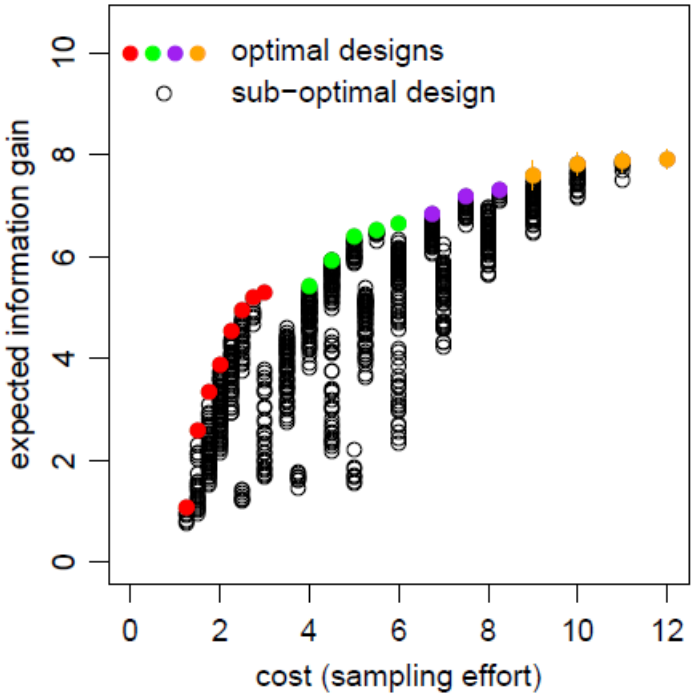
•For each design D simulate 1000 experiments and estimate expected benefit.

•Identify optimal designs for each different cost level.

•Prior for θ for a given treatment estimated from preliminary replicate.

Optimal design for a given cost





NB. Results sensitive to balance of costs between setting up replicates and numbers of sampling times.

Summing up

• Bayesian methods have role to play in design of experiments for stochastic epidemic models when number of replicates is constrained to be low.

•Potential for adaptive designs.

•Challenges in extending to spatio-temporal modelling where designs include locations of host or inoculum.

•Important to identify appropriate design criteria taking account of purpose of data collection.

Some references:

Gibson, G. J., Kleczkowski, A. & Gilligan, C. A. 2004. A Bayesian analysis of botanical epidemics using stochastic compartmental models. *Proc. Natl Acad. Sci. (USA)* **101**, 12120–12124.

Cook, A. R., Gibson, G. J. & Gilligan, C. A. 2008. Optimal observation times in experimental epidemic processes. *Biometrics* DOI:10.1111/j.1541-0420.2007.00931.x.

Krishnarajah, I., Cook, A., Marion, G. & Gibson, G. 2005. Novel moment closure approximations in stochastic epidemics. *Bull. Math. Biol.* **67**, 855–873.

Cook, A. R., Carrasco, L. R., Gibson, G. J. & Gilligan, C. A. 2011. Optimal design of epidemic sampling schemes: a Bayesian cost-benefit approach, in prep.

Verdinelli, I. & Kadane, J. B. 1992. Bayesian designs for maximizing information and outcome. *J. Am. Stat. Assoc.* **87**, 510–515.

Müller, P. 1999. Simulation-based optimal design. *Bayesian Statistics* **6**, 459–474.