Modelling and optimisation of group dose-response challenge experiments

David Price

Supervisors: Nigel Bean, Joshua Ross, Jonathan Tuke

July 9, 2013

THE UNIVERSITY of ADELAIDE

Experimental Design

- An experiment is a scientific procedure undertaken to make a discovery, test a hypothesis or demonstrate a known fact
  - Procedure
  - Subjects
  - Time(s) to observe experiment

Optimal Experimental Design

Statistical Criterion

- Given the log-likelihood

$$\ell = \log(L(\boldsymbol{\theta}))$$

The Fisher information is

$$\mathcal{I}_{i,j} = \mathbb{E}\left[\left(\frac{\partial \ell}{\partial \theta_i}\right)\left(\frac{\partial \ell}{\partial \theta_j}\right)\right]$$

- Related to the variance of parameter estimates

Optimality Criterion

Many different optimality criteria that look at minimising the variance of the parameter estimates.

E-Optimality: Maximise the smallest eigenvalue of $\mathcal{I}$
Minimise variance of parameter estimate with largest variance

A-Optimality: Maximise trace of $\mathcal{I}$
Minimise average variance of parameter estimates

D-Optimality: Maximise determinant of $\mathcal{I}$
Minimise the generalised variance of the parameter estimates

Frequentist Experimental Design

- Choose design that satisfies chosen criterion
- Locally optimal
- Require knowledge about the parameter in order to determine the optimal design

Frequentist Example

- Exponential lifetimes; $F(t) = 1 - \exp(-t\theta)$

$$\mathcal{I}(\theta, t) = \frac{t^2 \exp(-t\theta)}{1 - \exp(-t\theta)}$$

$$\frac{\partial \mathcal{I}}{\partial t} = \frac{t \exp(-t\theta)(2 - 2\exp(-t\theta) - t\theta)}{(1 - \exp(-t\theta))^2}$$

- Maximised when $t \approx 1.5936/\theta$

Bayesian Optimal Experimental Design

- Allows incorporation of prior knowledge into design
- Choose a utility function that we wish to maximise
- Expected Kullback-Leibler divergence

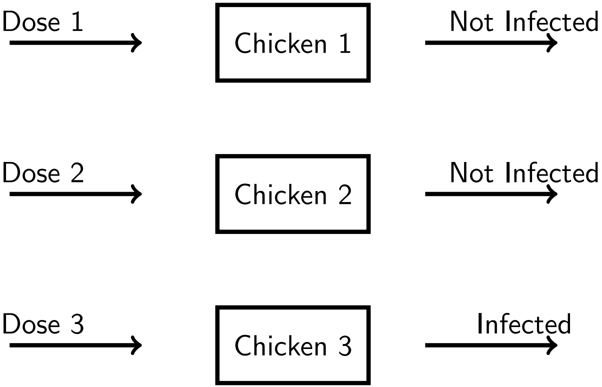
$$U(d) = \int\int \log\left(\frac{p(\theta \mid \boldsymbol{y}, d)}{p(\theta)}\right) p(\boldsymbol{y}, \theta \mid d) d\boldsymbol{y} d\theta$$

- Maximise our gain in information

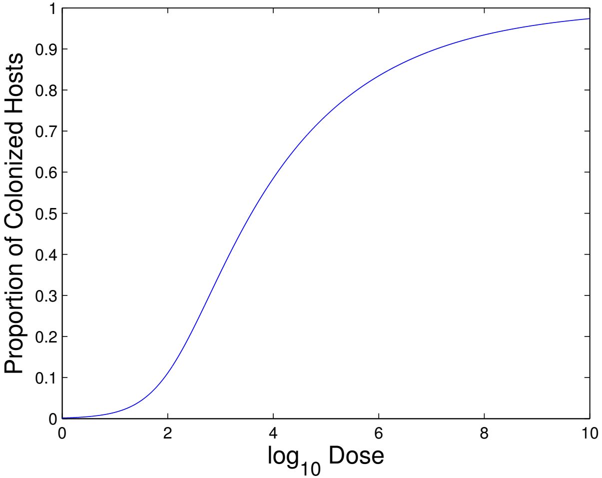
Optimal Experimental Design for Markov Chains

- Why Markov Chains?
- Open field of research
  - Becker and Kersting [1983]
  - Cook *et al.* [2008]
  - Pagendam and Pollett [2010]
  - Pagendam and Ross [2013]
  - Pagendam and Pollett [2013]

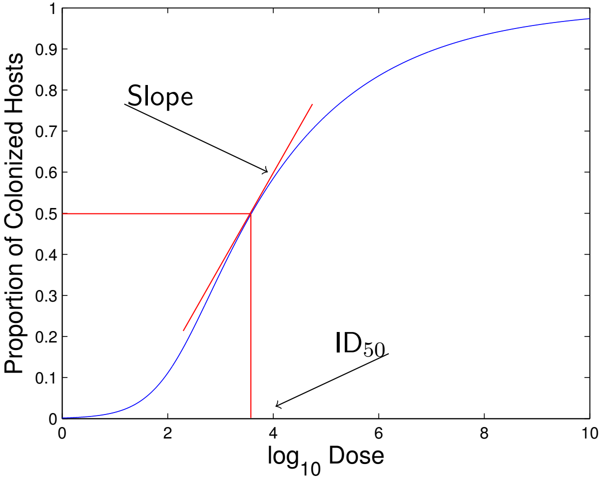
Dose-response challenge experiments



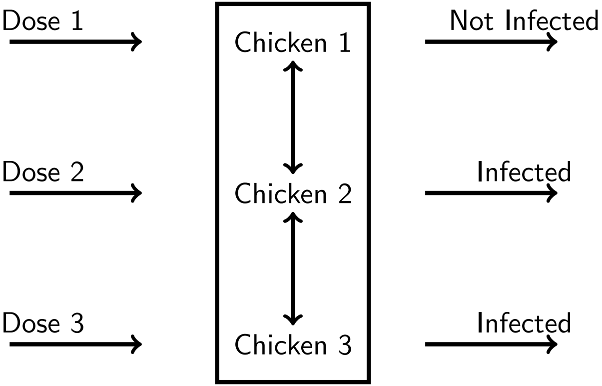
Dose-response relationship



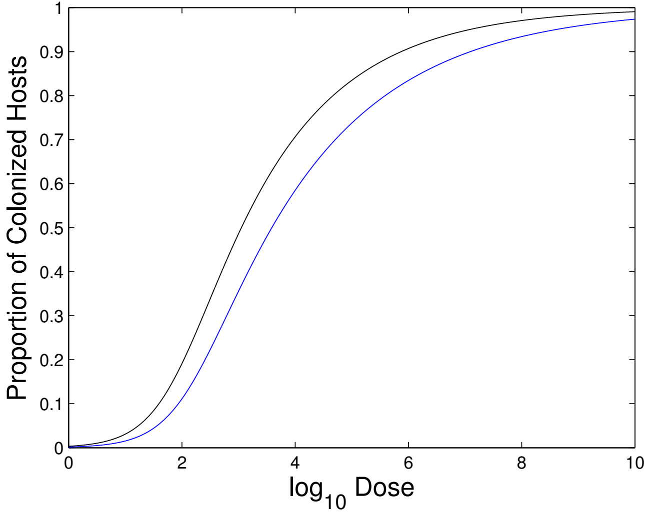
Dose-response relationship



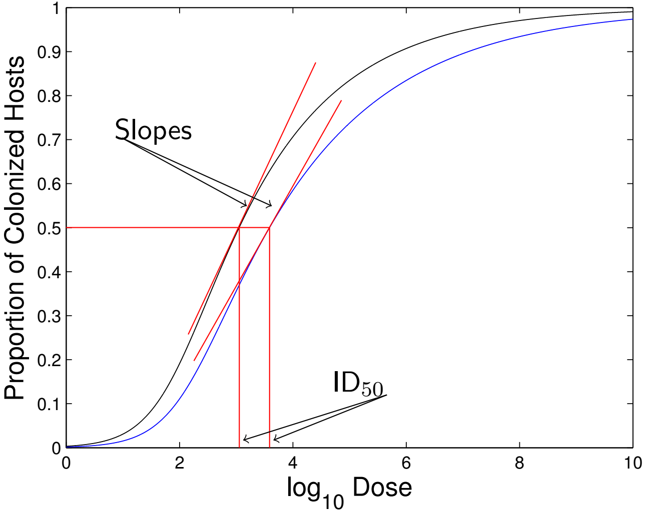
Group dose-response experiment with transmission



Dose-response relationship with transmission



Dose-response relationship with transmission



Modelling dose-response challenge experiments

- Conlan *et al.* [2011] first to account for transmission
- Two-stage process; dose-response and transmission
- Need to take into account latency period of dose-response
- Create SEIR model (Susceptible, Exposed, Infected, Removed)

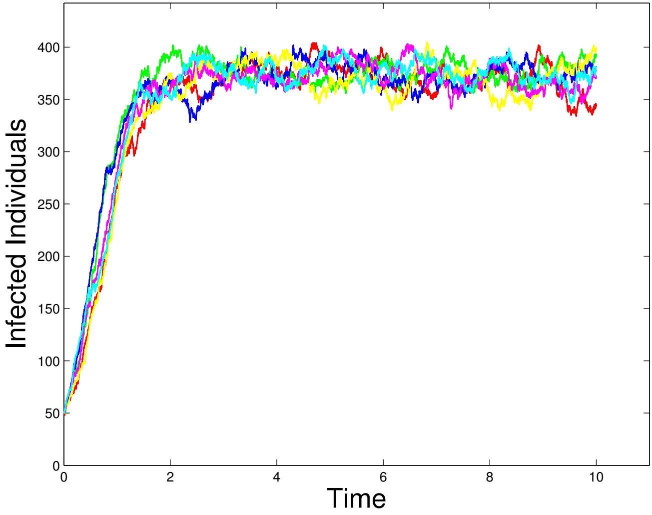
Optimal Design for some epidemic models

- Pagendam [2010] and Pagendam and Pollett [2013] looked at optimal design for experimental epidemics
- Locally optimal design of the SIS epidemic
- Likelihood evaluation computationally expensive

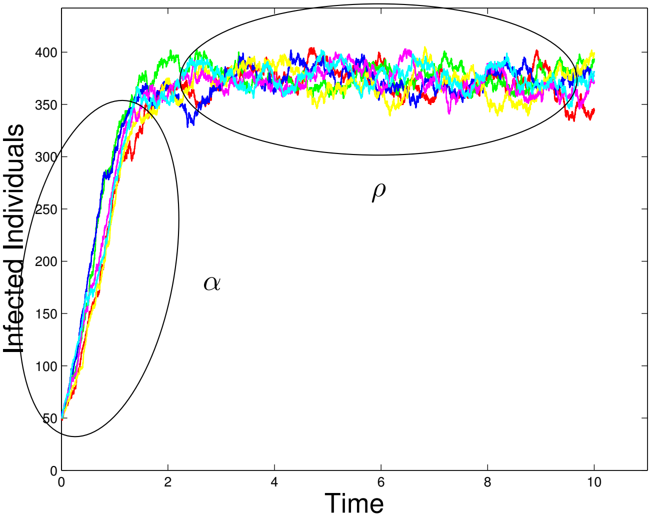
Recall: SIS epidemic

- State space is number of infected individuals $(0, \ldots, N)$
- Transition rates are
  - $q_{i,i+1} = \frac{\beta i(N-i)}{N}$
  - $q_{i,i-1} = \mu i$
- Estimate parameters (ρ, α), where $\rho = \frac{\mu}{\beta}$ and $\alpha = \beta - \mu$
- Nice physical interpretation

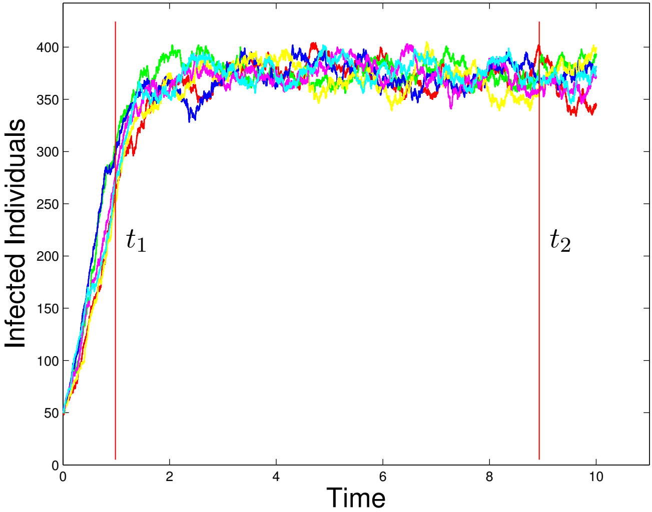
An SIS epidemic, $\alpha = 3$, $\rho = 0.25$



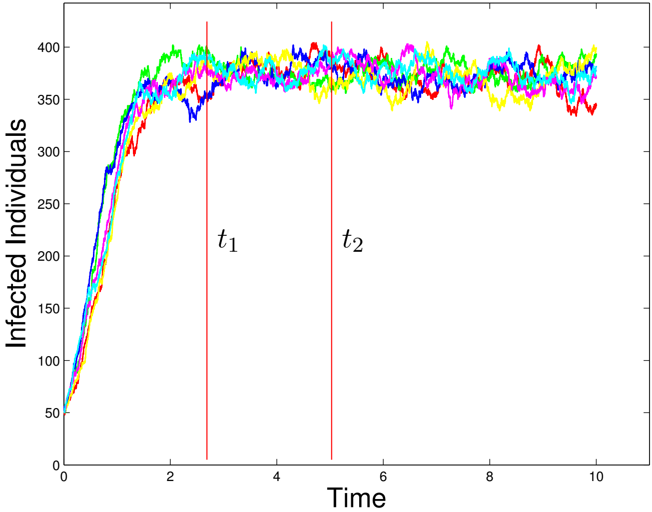
An SIS epidemic, $\alpha = 3$, $\rho = 0.25$



An SIS epidemic, $\alpha = 3$, $\rho = 0.25$



An SIS epidemic, $\alpha = 3$, $\rho = 0.25$



Overestimating α

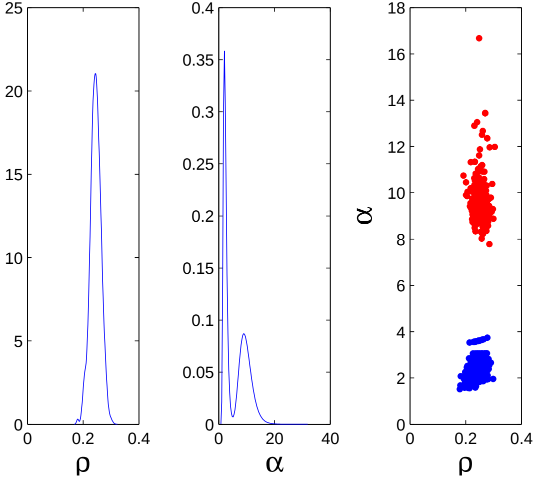
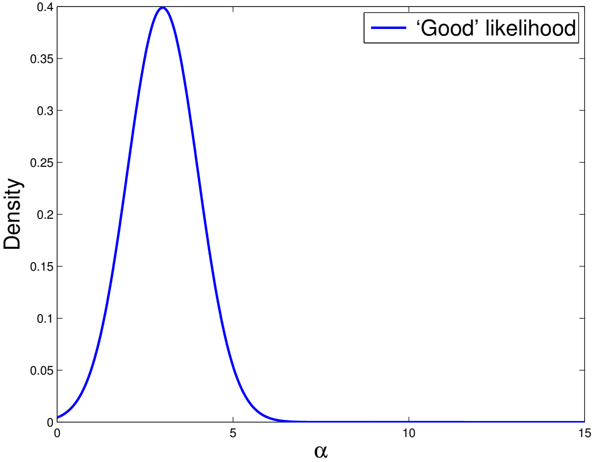


Figure: Densities of MLE's for (ρ, α). True values are $(0.25, 3)$.

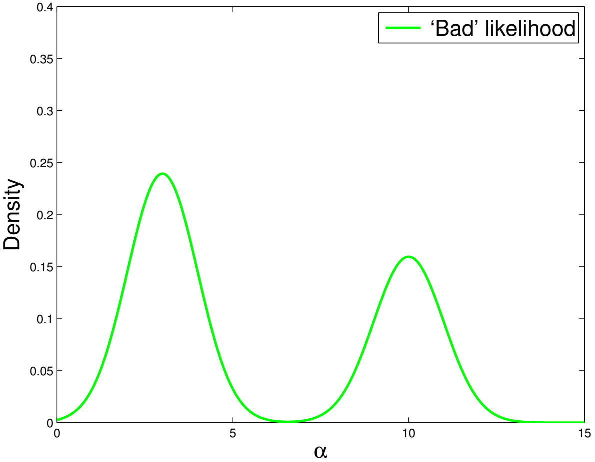
Does Bayes have a problem?

- Kullback-Leibler divergence looks to maximise the difference between the prior and posterior
- What if our posterior distribution for the 'bad' design is "further away" than the posterior for the 'good' design?

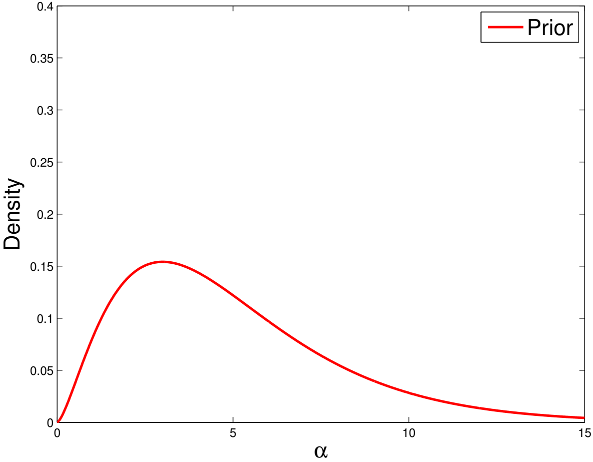
Example



Example



Example



Example

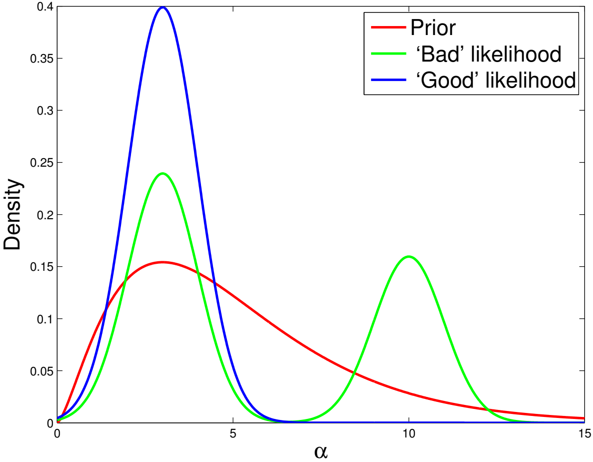


Figure: Example prior and two likelihood functions

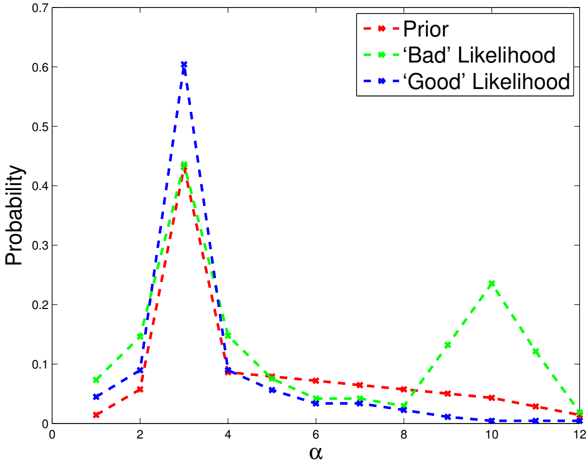


Figure: Discretised example prior and likelihood functions

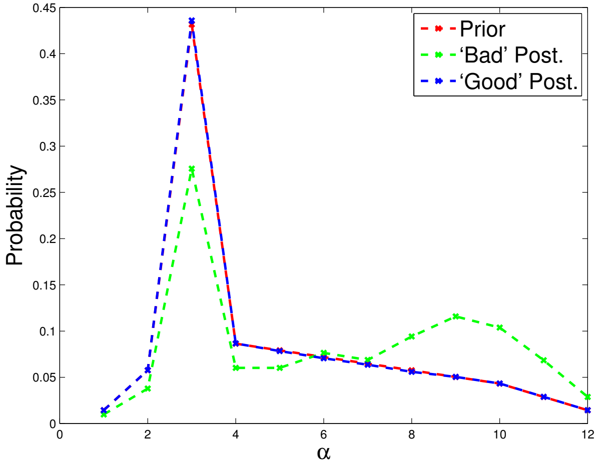


Figure: Discretised example prior and posterior distributions

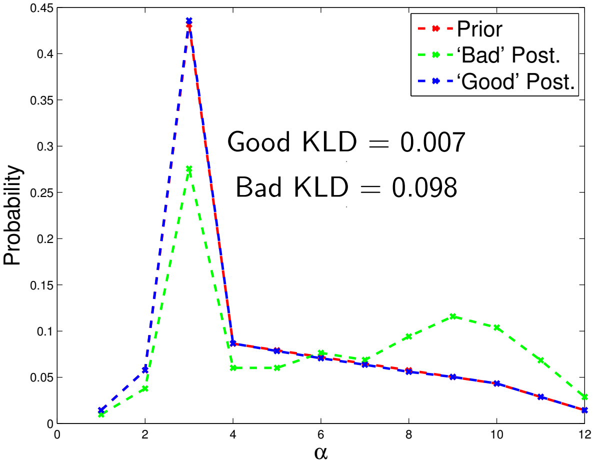


Figure: Discretised example prior and posterior distributions

Warning!

- Kullback-Leibler divergence is not a "black box"!
- Check:
  - Posterior
  - Prior
  - Design
- Care needs to be taken when using KLD for Bayesian Optimal design

Research Aims

- Investigate Kullback-Leibler divergence further
- Intractable likelihood for SEIR model
- Compare results of different design approaches for the SIS epidemic
- Move on to developing the SEIR model and applying these methods to that model

Thank you

- Supervisors:
  - Prof. Nigel Bean,
  - Dr Joshua Ross, and
  - Dr Jonathan Tuke
- Daniel Pagendam for correspondence.
- Everyone for listening!

References

G. Becker and G. Kersting. Design problems for the pure birth process. *Advances in Applied Probability*, 15(2):255–273, 1983.

A. Conlan, J. Line, C. Hiett, K. Coward, P. Van Diemen, M. Stevens, M. Jones, J. Gog, and D. Maskell. Transmission and dose-response experiments for social animals: a reappraisal of the colonization biology of *Campylobacter jejuni* in chickens. *J.R.S. Interface*, 8:1720–1735, 2011.

A. R. Cook, G. J. Gibson, and C. A. Gilligan. Optimal observation times in experimental epidemic processes. *Biometrics*, 64(3):860–868, 2008.

T. G. Kurtz. Solutions of ordinary differential equations as limits of pure jump markov processes. *Journal of Applied Probability*, 7(1):49–58, 1970.

D. E. Pagendam. *Experimental Design and Inference for Population Models*. PhD thesis, School of Mathematics and Physics, The University of Queensland, 2010.

D. E. Pagendam and P. K. Pollett. Locally optimal designs for the simple death process. *Journal of Statistical Planning and Inference*, 140(11), 2010.

D. E. Pagendam and P. K. Pollett. Optimal design of experimental epidemics. *Journal of Statistical Planning and Inference*, 143(3):563–572, 2013.

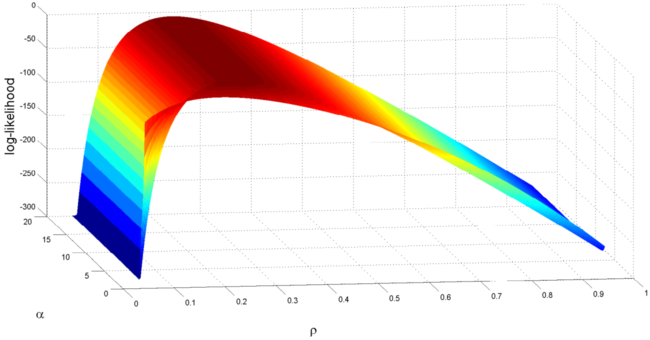
D. E. Pagendam and J. V. Ross. Optimal use of GPS transmitter for estimating species migration rates. *Ecological Modelling*, 249:37–41, 2013.

Gaussian Diffusion Approximation of the Likelihood

- Matrix exponentials computationally inefficient, especially as population size grows
- Kurtz [1970]
- The expected value of the SIS process over time, follows the deterministic trajectory.
- Σ is the covariance matrix, $\boldsymbol{y}$ is the observed number of infected at the observation times, and $\boldsymbol{m}$ is the corresponding mean number of infected at those times
- Very computationally efficient

$$L(\theta; \boldsymbol{y} \mid y_0) = (2\pi)^{-n/2} |N\Sigma|^{-1/2} \exp\left(-\frac{1}{2} (\boldsymbol{y} - N\boldsymbol{m}) \frac{\Sigma^{-1}}{N} (\boldsymbol{y} - N\boldsymbol{m})^T \right)$$

Likelihood for ‘bad’ design



Likelihood for ‘good’ design