Risk-averse design under uncertainty

Michael U. Gutmann

michael.gutmann@ed.ac.uk

School of Informatics, University of Edinburgh

28 March 2025

Acknowledgements

Presentation is based on the pre-print:

Risk-averse optimization of genetic circuits under uncertainty
Michal Kobiela, Diego A. Oyarzún, Michael U. Gutmann
`https://doi.org/10.1101/2024.11.13.623219`

Coding and simulations done by Michal. Most figures shown here were produced by him.

Contents

Program

Problem statement

Proposed methodology

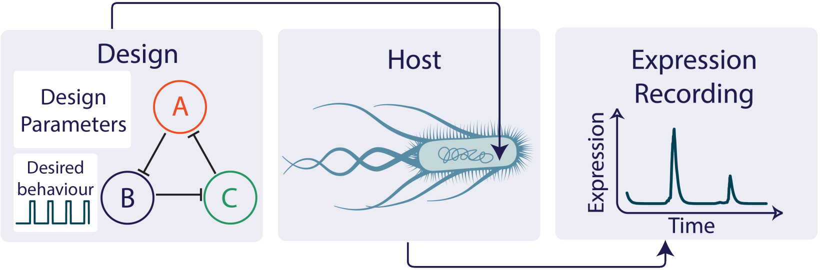
Example: design of a repressilator

Motivation

- Designing systems such that they have some prescribed properties is a ubiquitous problem.
- Mathematical/computer models of the system are often used to accelerate the design, e.g.
  - Hospitals: models are used to predict patient flow and optimise staffing.
  - Traffic management: models are used to optimise road networks and reduce congestion.
  - Synthetic biology: models are used to design genetic circuits.
- Design is framed as an optimisation problem.
- Issue: Models are not perfect and model-based designs may not achieve the predicted performance in reality.

Example: Design of a genetic circuit

- Genes form circuits that can be modelled as dynamical systems. Similar to circuits in electrical engineering.
- A network of three genes that inhibit (suppress) each other defines an oscillator, called "repressilator".
- Issue in genetic circuit design: performance on the computer may not be indicative of performance in-vivo.

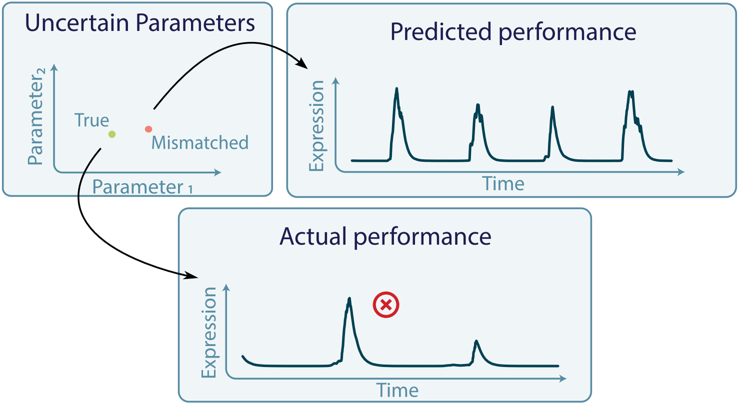


A case for design under uncertainty

- Hypothesis: gap between designed and actual performance is due to uncertainty (limited knowledge) about the real system.
- Two sources of uncertainty:
  1. Irreducible (aleatoric) uncertainty: modelled as system noise $\boldsymbol{\omega}$ with distribution $p(\boldsymbol{\omega})$.
  2. Reducible (epistemic) uncertainty: modelled as system parameters $\boldsymbol{\eta}$ with prior distribution $p(\boldsymbol{\eta})$.
- Both types of uncertainty need to be considered in the design; otherwise we design the system under wrong assumptions.

A case for design under uncertainty

- Reconsider the example of the design of a repressilator.
- Assuming wrong values for the model parameters results in a design that works on the computer but not in reality.



Problem statement

- We assume we are given:
  1. a stochastic generative model of the system:

     $$\mathbf{x} = g(\boldsymbol{\omega}, \boldsymbol{\eta}, \boldsymbol{\phi}), \quad \boldsymbol{\omega} \sim p(\boldsymbol{\omega}) \tag{1}$$

     where $\mathbf{x}$ describes the system, $\boldsymbol{\eta}$ are unknown system parameters not under our control, and $\boldsymbol{\phi}$ are design parameters that we can control.
  2. a prior belief $p(\boldsymbol{\eta})$
  3. data $\mathcal{D}_o = (\mathbf{x}_o, \boldsymbol{\phi}_o)$ on the system from previous design attempts with $\boldsymbol{\phi}_o$
  4. a loss function $\tilde{L}(\mathbf{x}) = L(\boldsymbol{\omega}, \boldsymbol{\eta}, \boldsymbol{\phi})$ that measures how good $\mathbf{x}$ is.
- Goal: Find designs $\boldsymbol{\eta}$ that are expected to work well even in unfavourable conditions.

Program

Problem statement

Proposed methodology

Example: design of a repressilator

Overview of our approach

Our approach consists of three steps:

1. Bayesian inference to update our belief about $\boldsymbol{\eta}$ given $\mathcal{D}_o$
2. Thompson sampling to obtain a set of promising designs.
3. Risk-management to select suitable designs.

Step 1: Bayesian inference

- Failed designs are oftentimes written off.
- We instead learn from them using Bayesian inference, updating prior to posterior belief $p(\boldsymbol{\eta}|\mathcal{D}_o)$.
- Depending on the model, we can compute the posterior with MCMC, variational inference, or simulation-based (likelihood-free) inference.
- Won't go further into this part of the pipeline.

Step 2: Thompson sampling

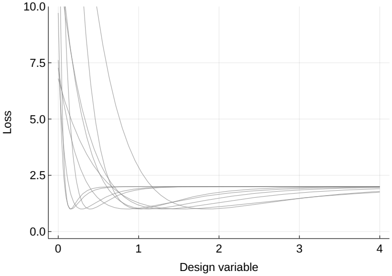
- The loss $\tilde{L}(\mathbf{x}) = L(\boldsymbol{\omega}, \boldsymbol{\eta}, \phi)$ measures how good a specific $\mathbf{x}$ is.
- It is random due to the randomness of $\boldsymbol{\omega}$ and $\boldsymbol{\eta}$.
- Let us average over the system noise $\boldsymbol{\omega}$ but pushforward the randomness of $\boldsymbol{\eta}$.
- This results in a stochastic process with realisations

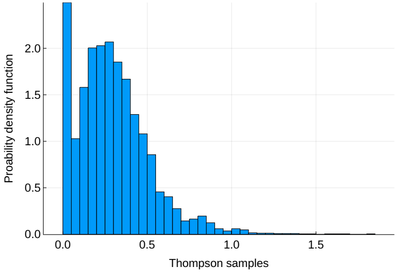
$$L_i(\phi) = \mathbb{E}_{\omega}\left[L(\boldsymbol{\omega}, \boldsymbol{\eta}_i, \phi)\right], \qquad \boldsymbol{\eta}_i \sim p(\boldsymbol{\eta}|\mathcal{D}_o) \tag{2}$$

- Thompson samples ϕ_i are defined as $\phi_i = \operatorname{argmin}_{\phi} L_i(\phi)$
- They are random variables: Uncertainty about the system translates into uncertainty about the design.
- Named after William R Thompson (1887-1972).
  (On the Likelihood That One Unknown Probability Exceeds Another in View of the Evidence of Two Samples. Biometrika (1933))

Simple example of Thompson sampling

- Let $x = g(\omega, \boldsymbol{\eta}, \phi) = \eta_1 \exp(-\phi\eta_2) + \omega$, with $\boldsymbol{\eta} = (\eta_1, \eta_2)$, $\phi \geq 0$, and the loss is the squared distance between x and 1.0.
- Gamma distributions for η_1 and η_2, standard normal for ω.
- Left: Loss as a function of ϕ for different realisations of η_1 and η_2.
- Right: Distribution of the Thompson samples





Step 3: Risk management

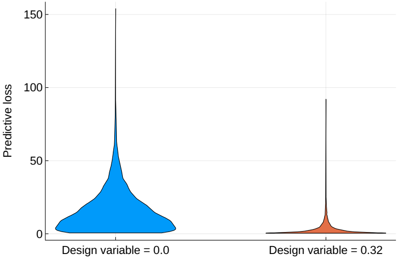
- Each Thompson sample ϕ_i has an associated predictive distribution of the loss,

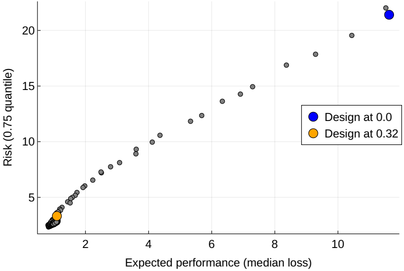
$$L(\boldsymbol{\omega}, \boldsymbol{\eta}, \phi_i), \quad \boldsymbol{\omega} \sim p(\boldsymbol{\omega}), \boldsymbol{\eta} \sim p(\boldsymbol{\eta}|\mathcal{D}_o) \tag{3}$$

- Distribution reflects system noise and our uncertainty about the system.
- The distribution can be used to rank the Thompson samples.
- We reduce the distribution to two statistics:
  - Median as a robust, risk-neutral, measure of expected performance
  - 0.75 quantile as a measure of risk, measuring how much worse the performance is when performing worse than the median.
- Pick design that achieves desired trade-off.
- Other choices possible, e.g. for optimistic design.

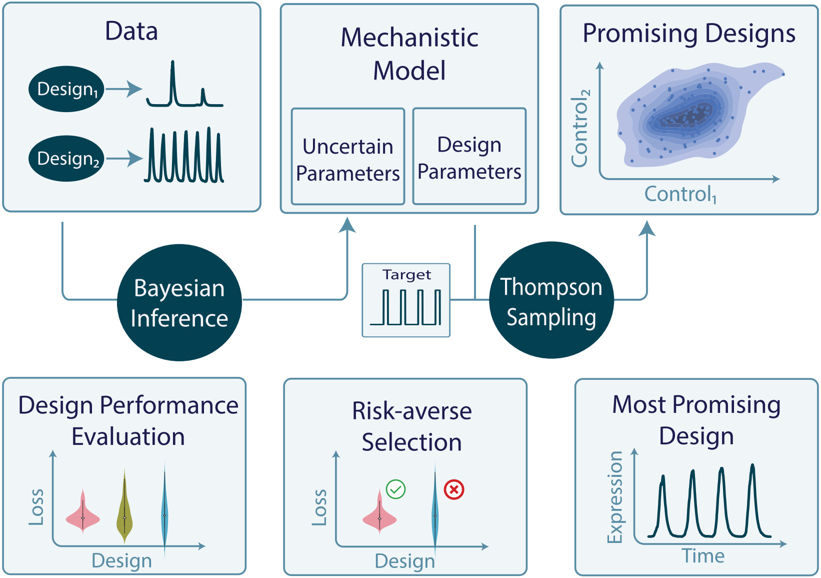
Back to the example

- Left: Predictive loss for two Thompson samples
- Right: Mapping out expected performance vs risk for different values of ϕ.
- No trade-off between performance and risk in this example.





Summary of our approach



Program

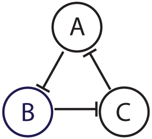
Problem statement

Proposed methodology

Example: design of a repressilator

Model

- The model $\mathbf{x} = g(\boldsymbol{\omega}, \boldsymbol{\eta}, \boldsymbol{\phi})$ is implicitly defined via the solution to a stochastic differential equation for the protein expressions p_1, p_2, p_3.

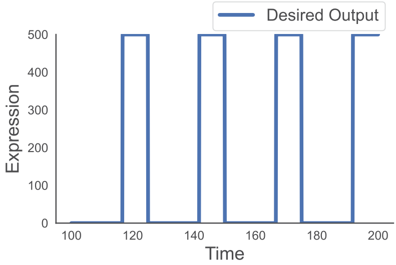


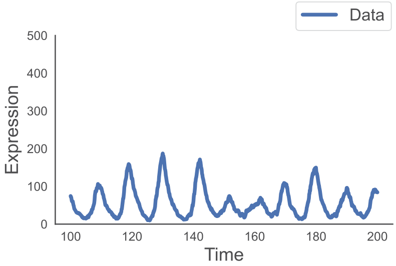
$$dp_i = \left(\underbrace{\frac{\alpha}{1 + \left(\frac{p_{i-1}}{K}\right)^n}}_{\text{Production}} - \underbrace{\gamma \cdot p_i}_{\text{Degradation}} \right) dt + \underbrace{\sqrt{\frac{\alpha}{1 + \left(\frac{p_{i-1}}{K}\right)^n}} dW_{i1} + \sqrt{\gamma \cdot p_i} dW_{i2}}_{\text{Noise}},$$

- Uncertain parameters $\boldsymbol{\eta} = (\gamma, n)$: degradation rate, Hill coefficient
- Design (controllable) parameters $\boldsymbol{\phi} = (\alpha, K)$: production rate and dissociation constant

Design goal and data

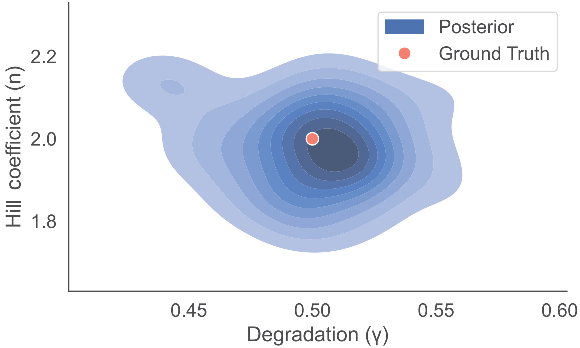
- Design goal: 4 oscillations during 100 minutes (left).
- Failed attempt: $\phi = (334, 10)$. Used as data $\mathcal{D}_o$ to update our belief about η (right)





Step 1: Bayesian inference

- Prior on η: Uniform on (0, 10).
- Posterior obtained with SMC-ABC.
- For invariance to temporal shifts of the trajectories, we matched the absolute value of the Fourier spectrum of simulated and observed data.

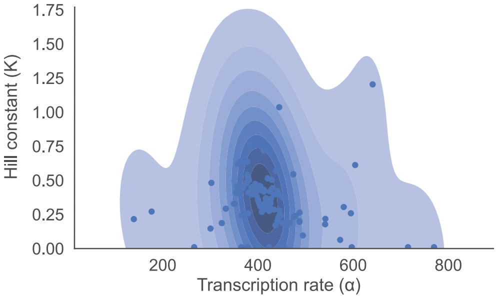


Step 2: Thompson sampling

- For each Thompson sample ϕ_i, we minimise the L_2 distance between the spectrum of the simulated and target trajectory.

$$L_i(\phi) = \mathbb{E}_\omega \left[\frac{1}{N} \sum_{n=0}^{N} (|F_{(\eta_i,\phi)}[n]| - |F_{target}[n]|)^2 \right], \quad \boldsymbol{\eta}_i \sim p(\boldsymbol{\eta}|\mathcal{D}_o)$$

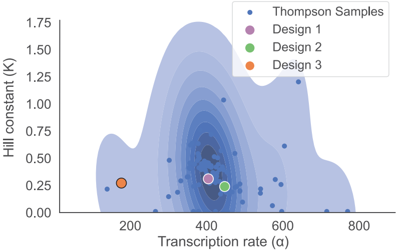
- Distribution of Thompson samples.

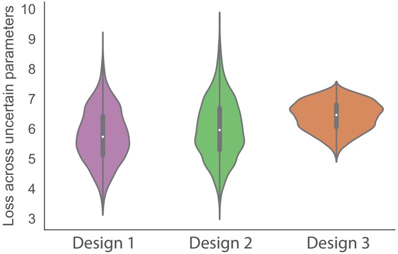


- To assess the Thompson samples ϕ_i, we evaluate their predictive loss distribution, using

$$L(\omega, \eta, \phi_i), \quad \omega \sim p(\omega), \eta \sim p(\eta|\mathcal{D}_o) \tag{4}$$

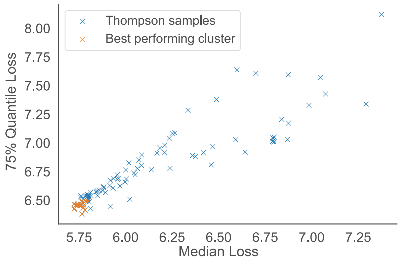
- Examples

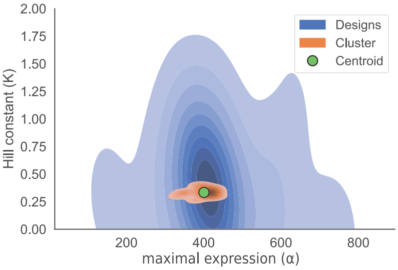




Step 3: Risk management

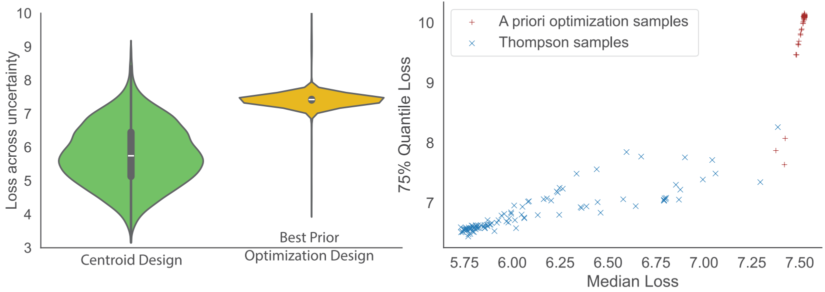
- We summarise each predictive loss distribution in terms of their median (expected performance) and 75% quantile (risk).
- Enables informed decision about suitable designs.
- Cluster marked in orange are most promising designs. Centroid as single best one.





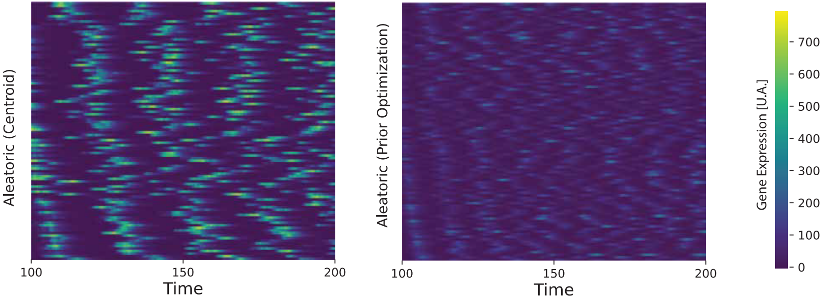
Evaluation

- Is there a benefit in using the failed design attempts to learn more about the systems?
- Comparison between Thompson samples for $\boldsymbol{\eta} \sim p(\boldsymbol{\eta}|\mathcal{D}_o)$ and $\boldsymbol{\eta} \sim p(\boldsymbol{\eta})$ indicates large predicted performance difference.



Evaluation

- The centroid design point is *predicted* to be best. How good is it when deployed?
- Left fig: Design achieves four oscillations in the specified time-interval under ground truth $\boldsymbol{\eta}$ parameters.
- Right fig: This is not the case for best design under the prior.



Conclusions

- Talk was on designing systems with the help of models.
- Models are not perfect and model-based designs may not achieve the predicted performance in reality.
- Considered the case where the gap between designed and actual performance is due to uncertainty (limited knowledge) about the real system.
- Proposed an uncertainty-aware design methodology by combining:
  - Bayesian inference
  - Thompson sampling
  - Risk management
- Application in synthetic biology.
- Preprint with more results: `https://www.biorxiv.org/node/4220646`