# STSCI 4780/5780
# Decision theory & experimental design

Tom Loredo, CCAPS & SDS, Cornell University

# Agenda

1. **Decision theory**

2. **Experimental design**

3. **Recap**

# Naive Decision Making

A Bayesian analysis results in probabilities for two hypotheses:

$$p(H_1|I) = 5/6; \qquad\qquad p(H_2|I) = 1/6$$

Equivalently, the odds favoring $H_1$ over $H_2$ are

$$O_{12} = 5$$

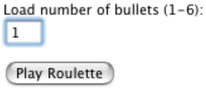
We must base future actions on either $H_1$ or $H_2$.

Which should we choose?

Naive decision maker: *Choose the most probable,* $H_1$

# Naive Decision Making—Deadly!

*Russian Roulette*




$H_1$ = Chamber is empty; $H_2$ = Bullet in chamber

What is your choice now?

*Decisions should depend on consequences!*

Unattributed JavaScript at `http://www.javascriptkit.com/script/script2/roulette.shtml`

# Bayesian Decision Theory

*Decisions depend on consequences*

Might bet on an improbable outcome provided the payoff is large if it occurs and/or the loss is small if it doesn't

*Utility and loss functions*

Compare consequences via *utility* quantifying the benefits of a decision, or via *loss* quantifying costs

$$\text{Utility} = U(a, o) \qquad \begin{aligned} a &= \text{Choice of action (decide b/t these)} \\ o &= \text{Outcome (what we are uncertain of)} \end{aligned}$$

Loss $L(a, o) = U_{\max} - U(a, o)$

## *Russian Roulette Utility*

Suppose you're offered \$6,000 to play

| | **Outcomes** | |
|---|---|---|
| **Actions** | Empty (*click*) | Bullet (*BANG!*) |
| *Play* | \$6,000 | −\$Life |
| *Pass* | 0 | 0 |

## *Uncertainty & expected utility*

We are uncertain of what the outcome will be
$\rightarrow$ Expected utility *averages over outcomes*:

$$\mathbb{E}U(a) = \sum_{\text{outcomes}} P(o|\ldots)\, U(a, o)$$

The best action *maximizes the expected utility*:

$$\hat{a} = \arg\max_a \mathbb{E}U(a)$$

I.e., minimize expected loss.

Axiomatized: von Neumann & Morgenstern; Ramsey, de Finetti, Savage

## *Russian Roulette Expected Utility*

| | **Outcomes** | | |
|---|---|---|---|
| **Actions** | Empty (*click*) | Bullet (*BANG!*) | $\mathbb{E}U$ |
| *Play* | \$6,000 | $-$\$Life | \$5000$-$\$Life/6 |
| *Pass* | 0 | 0 | 0 |

As long as \$Life $>$ \$30,000, *don't play!*

# Decision theory and parametric models

Decision theory can motivate specific posterior summaries:

- Point estimates for parameters ("Bayes estimators"):
  - Posterior median: best for absolute error loss
  - Posterior mean: best for squared error loss
  - Posterior mode: best for 0/1 loss ("all or nothing" prize)
- HPD regions best if penalize regions for increasing size

For model choice, *explanatory* vs. *predictive* criteria can lead to different choices

- They trade off bias vs. variance differently
- E.g., AIC comes from a *predictive* criterion, and BIC/Bayes factors from an *explanatory* criterion

# Agenda

1. Decision theory
2. **Experimental design**
3. Recap

# Experimental design as decision making

When we perform an experiment we have choices of actions:

- What sample size to use
- What times or locations to probe/query
- Whether to do one sensitive, expensive experiment or several less sensitive, less expensive experiments
- Whether to stop or continue a sequence of trials
- ...

We must choose amidst uncertainty about the data we may obtain and the resulting consequences for our experimental results

$\Rightarrow$ Seek a principled approach for optimizing experiments, accounting for all relevant uncertainties

# Bayesian experimental design

Actions $= \{e\}$, possible experiments (sample sizes, sample times/locations, stopping criteria . . . ).

Outcomes $= \{d_e\}$, values of future data from experiment $e$.

Utility measures value of $d_e$ for achieving experiment goals, possibly accounting for the cost of the experiment.

Choose the experiment that maximizes

$$\mathbb{E}U(e) = \sum_{d_e} p(d_e| \ldots)\, U(e, d_e)$$

To predict $d_e$ we must consider various hypotheses, $H_i$, for the data-producing process $\rightarrow$ Average over $H_i$ uncertainty:

$$\mathbb{E}U(e) = \sum_{d_e} \left[ \sum_{H_i} p(H_i| \ldots) p(d_e|H_i, \ldots) \right] U(e, d_e)$$

# Information-based utility

Many scientific studies do not have a single, clear-cut goal.

Broad goal: Learn/explore, with resulting information made available for a variety of future uses.

Example: Astronomical measurement of orbits of minor planets or exoplanets

- Use to infer physical properties of a body (mass, habitability)
- Use to infer distributions of properties among the population (constrains formation theories)
- Use to predict future location (collision hazard; plan future observations)

Motivates using a "general purpose" utility that measures *what is learned* about the $H_i$ describing the phenomenon

Lindley (1956, 1972) and Bernardo (1979) advocated using $\mathcal{I}(D)$ as such a general-purpose utility

## *MaxEnt sampling for parameter estimation*

Setting:

- We have specified a model, $M$, with uncertain parameters $\theta$
- We have data $D \to$ current posterior $p(\theta|D, M)$
- The entropy of the noise distribution doesn't depend on $\theta$

$$\to \mathbb{E}\mathcal{I}(e) \quad = \quad \text{Const} - \sum_{d_e} p(d_e|D, I) \log p(d_e|D, I)$$

*Maximum entropy sampling.*
(Sebastiani & Wynn 1997, 2000)
*To learn the most, sample where you know the least.*

# Scientific method

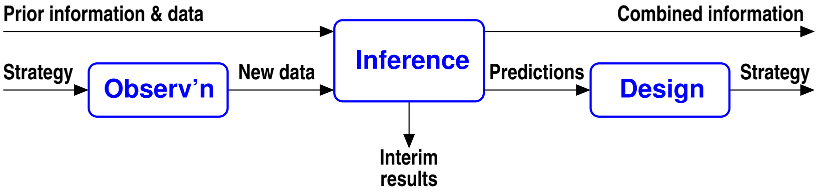
*Science is more than a body of knowledge; it is a way of thinking.*
*The method of science, as stodgy and grumpy as it may seem,*
*is far more important than the findings of science.*

—Carl Sagan

*Classic hypothetico-deductive approach*

- Form hypothesis (based on past observation/experiment)
- Devise experiment to test predictions of hypothesis
- Perform experiment
- Analysis $\rightarrow$
    - Devise new hypothesis if hypothesis fails
    - Devise new experiment if hypothesis corroborated

# Bayesian Adaptive Exploration



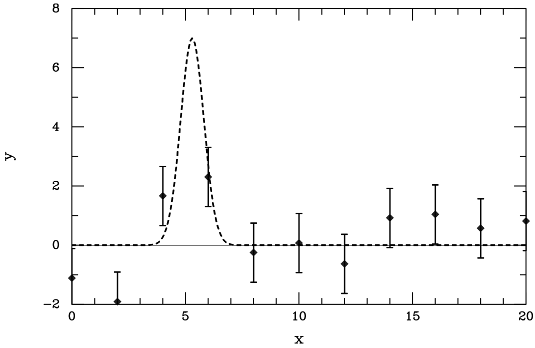

- Observation — Gather new data based on observing plan
- Inference — Interim results via posterior sampling
- Design — Predict future data; explore where expected information from new data is greatest

# Locating a bump

Object is 1-d Gaussian of unknown loc'n, amplitude, and width. True values:

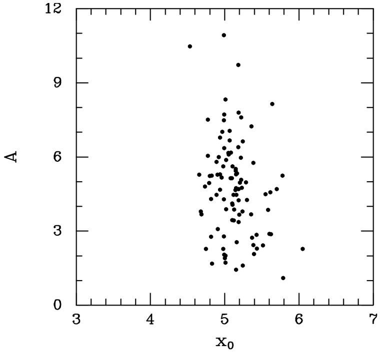
$$x_0 = 5.2, \qquad \text{FWHM} = 0.6, \qquad A = 7$$

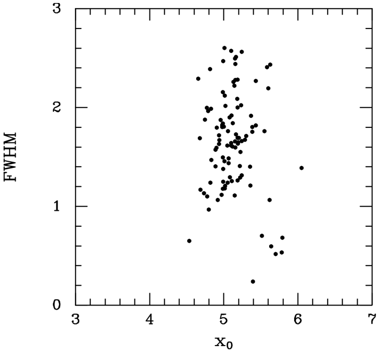
Initial scan with crude ($\sigma = 1$) instrument provides 11 equispaced observations over $[0, 20]$. Subsequent observations will use a better ($\sigma = 1/3$) instrument.



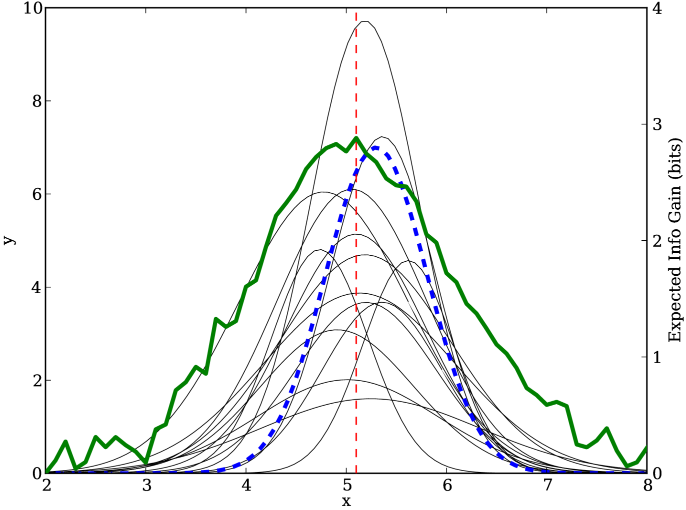
# Cycle 1 Interim Inferences

Generate $\{x_0, FWHM, A\}$ via posterior sampling.

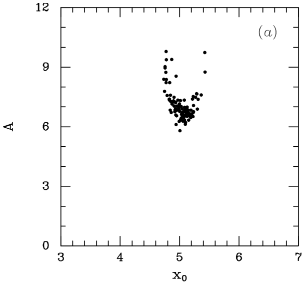


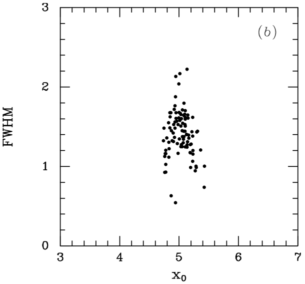



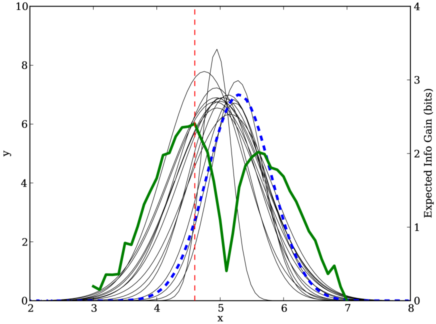
# Cycle 1 Design: Predictions, Entropy



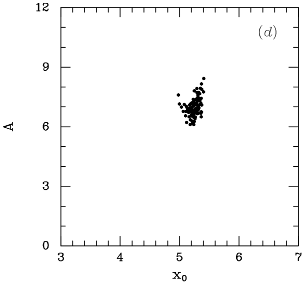
# Cycle 2: Inference, Design

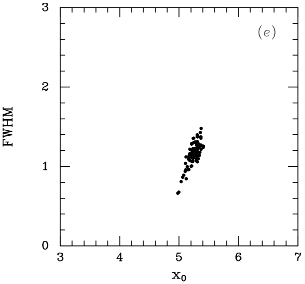


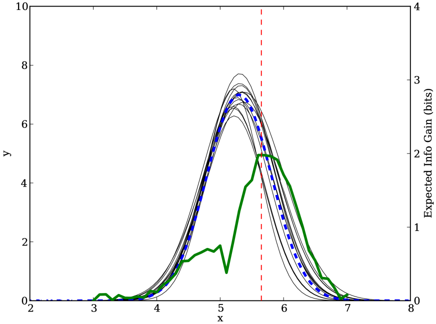



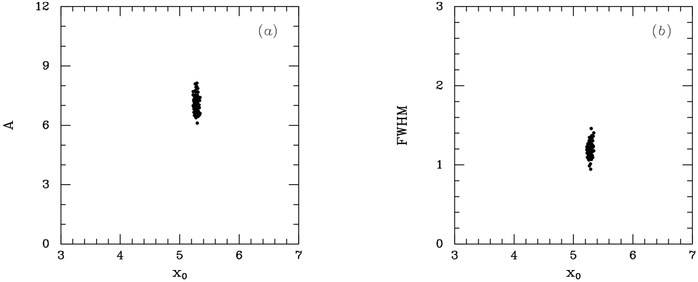
# Cycle 3: Inference, Design



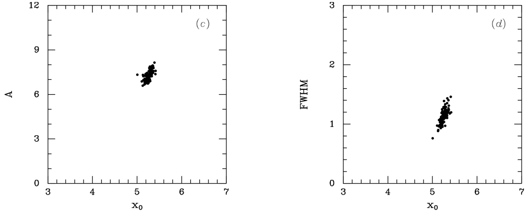



# Cycle 4: Inferences




## Inferences from *non-optimal* datum



# Agenda

1. Decision theory
2. Experimental design
3. **Recap**

# Recap: Bayesian inference in one slide

*Probability as generalized logic*

Probability quantifies the *strength of arguments*

To appraise hypotheses, calculate probabilities for arguments from data and modeling assumptions to each hypothesis

Use *all* of probability theory for this

*Bayes's theorem*

$$p(\text{Hypothesis} \mid \text{Data}) \quad \propto \quad p(\text{Hypothesis}) \times p(\text{Data} \mid \text{Hypothesis})$$

Data *change* the support for a hypothesis $\propto$ ability of hypothesis to *predict* the data

*Law of total probability*

$$p(\text{Hypothes}\underline{\textbf{es}} \mid \text{Data}) \quad = \quad \sum p(\text{Hypothes}\underline{\textbf{is}} \mid \text{Data})$$

The support for a *composite/compound* hypothesis must account for all the ways it could be true