Two Lectures:
Bayesian Inference in a Nutshell
and
The Perils & Promise of Statistics With Large Data Sets & Complicated Models

Bayesian-Frequentist Cross-Fertilization

Tom Loredo

Dept. of Astronomy, Cornell University

http://www.astro.cornell.edu/staff/loredo/

Lecture 1

- Statistical inference
- Bayesian inference with parametric models
- Simple examples: Coins, Gaussians
- Suggested reading

Inference

Inference: the act of passing from one proposition, statement or judgment considered as true to another whose truth is believed to follow from that of the former (Webster)

Do premises $A, B, \ldots \quad \rightarrow$ *hypothesis,* H*?*

Deductive inference: Premises allow definite determination of truth/falsity of H (syllogisms, symbolic logic, Boolean algebra)

$$B(H|A, B, \ldots) = 0 \quad \text{or} \quad 1$$

Inductive inference: Premises bear on truth/falsity of H, but don't allow its definite determination (weak syllogisms, analogies)

A, B, C, D share properties x, y, z; E has properties x, y
$\rightarrow E$ probably has property z

Statistical Inference

Quantify the strength of inductive inferences from data (D, facts) and models (other premises) to hypotheses about the phenomena producing the data.

Quantify via *probabilities*, or averages calculated using probabilities. $\mathcal{F}$ and $\mathcal{B}$ use probabilities very differently for this.

Inference With Parametric Models

Models M_i ($i = 1$ to N), each with parameters θ_i, each imply a *sampling dist'n* (conditional predictive dist'n for possible data):

$$p(D|\theta_i, M_i)$$

The θ_i dependence when we fix attention on the observed data is the *likelihood function*:

$$\mathcal{L}_i(\theta_i) \equiv p(D_{\text{obs}}|\theta_i, M_i)$$

We may be uncertain about i (model uncertainty) or θ_i (parameter uncertainty).

Parameter Estimation

Premise $=$ choice of model (pick specific i)
$\rightarrow$ What can we say about θ_i?

Model Uncertainty

- Model comparison: Premise $= \{M_i\}$

  $\rightarrow$ What can we say about i?

- Model adequacy/GoF: Premise $= M_1$
  $\rightarrow$ Is M_1 adequate?

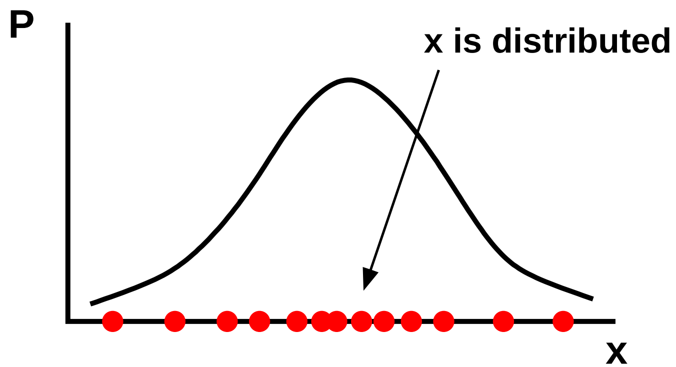
Hybrid Uncertainty

Models share some common params: $\theta_i = \{\phi, \eta_i\}$
$\rightarrow$ What can we say about ϕ?
(Systematic error is an example)

Rival Quantifications

Frequentist

- Devise procedure to choose among H_i using D. Apply it to D_{obs}.
- Report long-run performance (e.g., how *often* it is correct, how "far" the choice is from the truth *on average*).
- Probabilities are rates/proportions/frequencies.

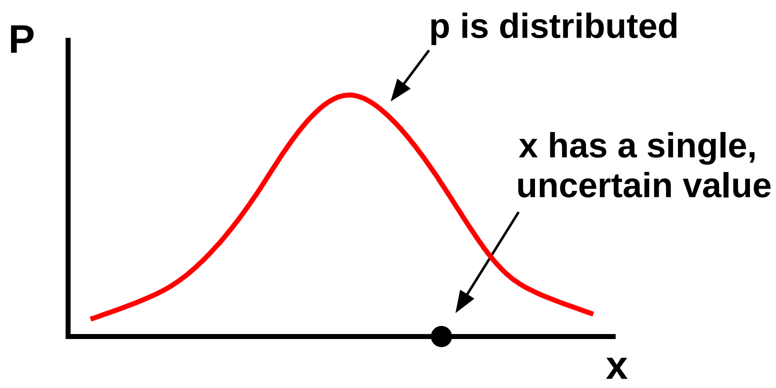


Bayesian

- Calculate probabilities for hypotheses given D_{obs} and modelling premises, using rules of probability theory.
- Probability is used as an abstract generalization of Boolean algebra:

$$B(H|A, B, \ldots) \in \{0, 1\} \quad \rightarrow \quad P(H|A, B, \ldots) \in [0, 1]$$

Probability theory provides the "calculus of inference."



The Bayesian Recipe

Assess hypotheses by calculating their probabilities $p(H_i|\ldots)$ conditional on known and/or presumed information using the rules of probability theory.

Probability Theory Axioms:

'OR' (sum rule)
$$P(H_1 + H_2|I) = P(H_1|I) + P(H_2|I) - P(H_1, H_2|I)$$

'AND' (product rule)
$$P(H_1, D|I) = P(H_1|I)\,P(D|H_1, I) = P(D|I)\,P(H_1|D, I)$$

Three Important Theorems

Normalization

For *exclusive, exhaustive* H_i

$$\sum_i P(H_i|\cdots) = 1$$

Bayes's Theorem

$$P(H_i|D_{\text{obs}}, I) = P(H_i|I)\,\frac{P(D_{\text{obs}}|H_i, I)}{P(D_{\text{obs}}|I)}$$

posterior $\propto$ prior $\times$ likelihood

norm. const. $P(D_{\text{obs}}|I) =$ prior predictive

Marginalization

Note that for exclusive, exhaustive $\{B_i\}$,

$$\begin{aligned}\sum_i P(A, B_i|I) &= \sum_i P(B_i|A, I)P(A|I) = P(A|I) \\ &= \sum_i P(B_i|I)P(A|B_i, I)\end{aligned}$$

$\rightarrow$ We can use $\{B_i\}$ as a “basis” to get $P(A|I)$.

Example: Take $A = D_{\text{obs}}$, $B_i = H_i$; then

$$\begin{aligned}P(D_{\text{obs}}|I) &= \sum_i P(D_{\text{obs}}, H_i|I) \\ &= \sum_i P(H_i|I)P(D_{\text{obs}}|H_i, I)\end{aligned}$$

prior predictive for D_{obs} = Average likelihood for H_i
(*aka* “marginal likelihood”)

Probability and Frequency

Frequency from probability

Bernoulli's laws of large numbers: In repeated trials, given $P(\text{success}|\ldots)$, predict

$$\frac{N_{\text{success}}}{N_{\text{total}}} \to P \quad \text{as} \quad N \to \infty$$

Probability from frequency

Bayes's "An Essay Towards Solving a Problem in the Doctrine of Chances" $\rightarrow$ First use of Bayes's theorem

Interpreting abstract probabilities

A Thermal Analogy

Intuitive notion	*Quantification*	*Calibration*
Hot, cold	Temperature, T	Cold as ice $= 273$K Boiling hot $= 373$K
uncertainty	Probability, P	Certainty $= 0$, 1 $p = 1/36$: plausible as “snake’s eyes” $p = 1/1024$: plausible as 10 heads

Example: Coin Flipping!

Parameter Estimation

$M =$ Assumed independence of flips

$H_i =$ Statements about a, the probability for heads on the next flip $\rightarrow$ seek $p(a|D, M)$

$D =$ Sequence of results from N previous flips:

THTHHHTHHHHT ($n = 8$ heads in $N = 12$ flips)

Likelihood:

$$
\begin{aligned}
p(D|a, M) &= p(\text{tails}|a, M) \times p(\text{heads}|a, M) \times \cdots \\
&= a^n (1-a)^{N-n} \\
&= \mathcal{L}(a)
\end{aligned}
$$

Prior:

Starting with no information about a beyond its definition, use as an "uninformative" prior $p(a|M) = 1$. Justifications:

- Intuition: Don't prefer any a interval to any other of same size
- Bayes's justification: "Ignorance" means that before doing the N flips, we have no preference for how many will be heads:

$$P(n\,\text{heads}|M) = 1/N \rightarrow p(a|M) = 1$$

Consider this a *convention*—an assumption added to M to make the problem well posed.

Prior Predictive:

$$
\begin{aligned}
p(D|M) &= \int da\; a^n (1-a)^{N-n} \\
&= B(n+1, N-n+1) = \frac{n!(N-n)!}{(N+1)!}
\end{aligned}
$$

A *Beta integral*, $B(a,b) \equiv \int dx\, x^{a-1}(1-x)^{b-1} = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$.

Posterior:

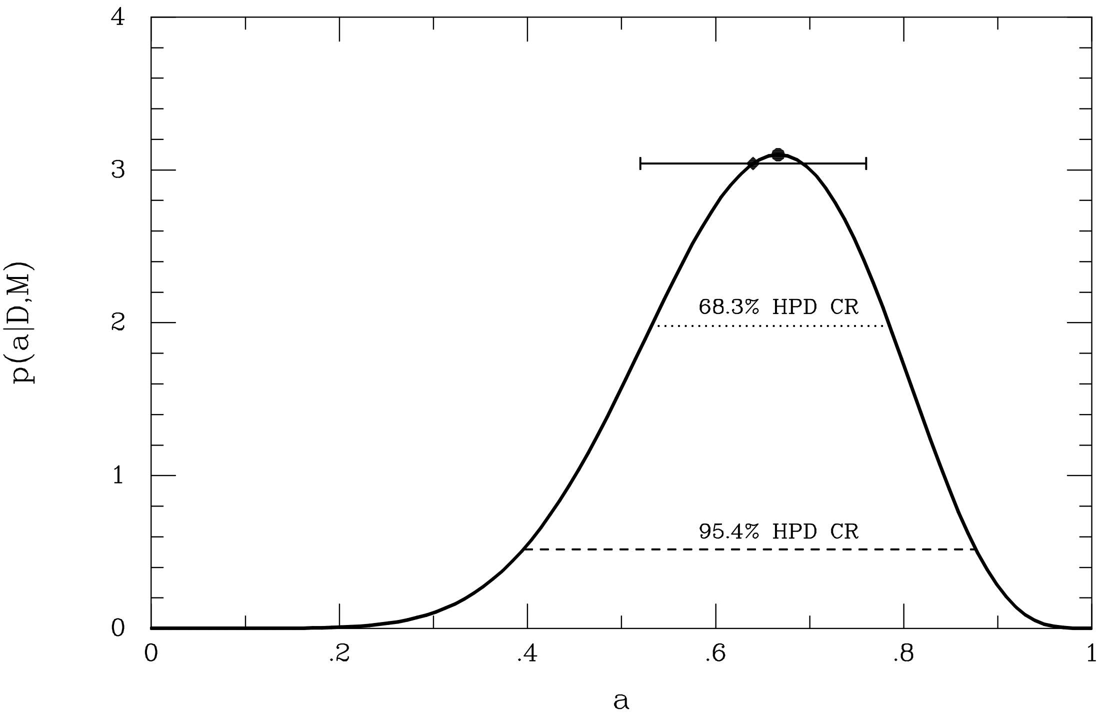
$$p(a|D,M) \;=\; \frac{(N+1)!}{n!(N-n)!}a^n(1-a)^{N-n}$$

A *Beta distribution*. Summaries:

- Best-fit: $\hat{a} = \frac{n}{N} = 2/3$; $\langle a \rangle = \frac{n+1}{N+2} \approx 0.64$
- Uncertainty: $\sigma_a = \sqrt{\frac{(n+1)(N-n+1)}{(N+2)^2(N+3)}} \approx 0.12$

  Find credible regions numerically, or with incomplete beta function

Note that the posterior depends on the data only through n, not the N binary numbers describing the sequence. n is a (minimal) *Sufficient Statistic*.



p(a|D,M)
4
3
2
1
0
0
.2
.4
.6
.8
1
a
68.3% HPD CR
95.4% HPD CR

Model Comparison: Fair Flips?

M_1: $a = 1/2$

M_2: $a \in [0, 1]$ with flat prior.

Maximum Likelihoods:

$$M_1: \qquad p(D|M_1) = \frac{1}{2^N} = 2.44 \times 10^{-4}$$

$$M_2: \quad \mathcal{L}(2/3) \;=\; \left(\frac{2}{3}\right)^n \left(\frac{1}{3}\right)^{N-n} = 4.82 \times 10^{-4}$$

$$\frac{p(D|M_1)}{p(D|\hat{a}, M_2)} = 0.51$$

Maximum likelihoods favor M_2 (biased flips).

Bayes Factor (ratio of model likelihoods):

$$p(D|M_1) = \frac{1}{2^N}; \qquad \text{and} \qquad p(D|M_2) = \frac{n!(N-n)!}{(N+1)!}$$

$$\begin{aligned} \to B_{12} \equiv \frac{p(D|M_1)}{p(D|M_2)} &= \frac{(N+1)!}{n!(N-n)!2^N} \\ &= 1.57 \end{aligned}$$

Bayes factor (odds) favors M_1 (fair flips).

Note that for $n = 6$, $B_{12} = 2.93$; for this small amount of data, we can never be very sure the coin is fair.

If $n = 0$, $B_{12} \approx 1/315$; if $n = 2$, $B_{12} \approx 1/4.8$; for extreme data, 12 flips *can* be enough to lead us to strongly suspect the coin flipping is not fair.

Aside: Model Comparison

$I = (M_1 + M_2 + \ldots)$ — Specify a set of models.
$H_i = M_i$ — Hypothesis chooses a model.

Posterior probability for a model:

$$
\begin{aligned}
p(M_i|D,I) &= p(M_i|I)\frac{p(D|M_i,I)}{p(D|I)} \\
&\propto p(M_i)\mathcal{L}(M_i)
\end{aligned}
$$

But $\mathcal{L}(M_i) = p(D|M_i) = \int d\theta_i\, p(\theta_i|M_i)p(D|\theta_i, M_i)$.

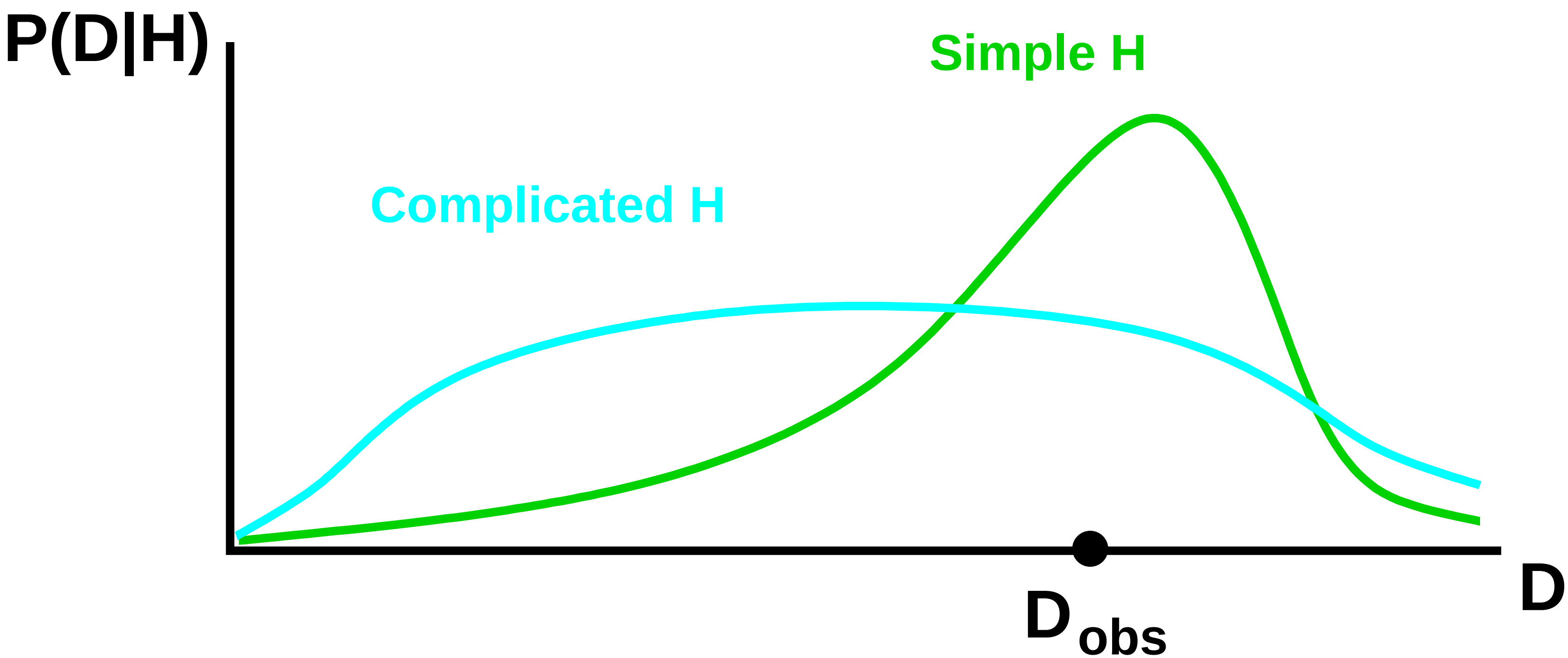
Likelihood for model $=$ Average likelihood for its parameters

$$\mathcal{L}(M_i) = \langle \mathcal{L}(\theta_i) \rangle$$

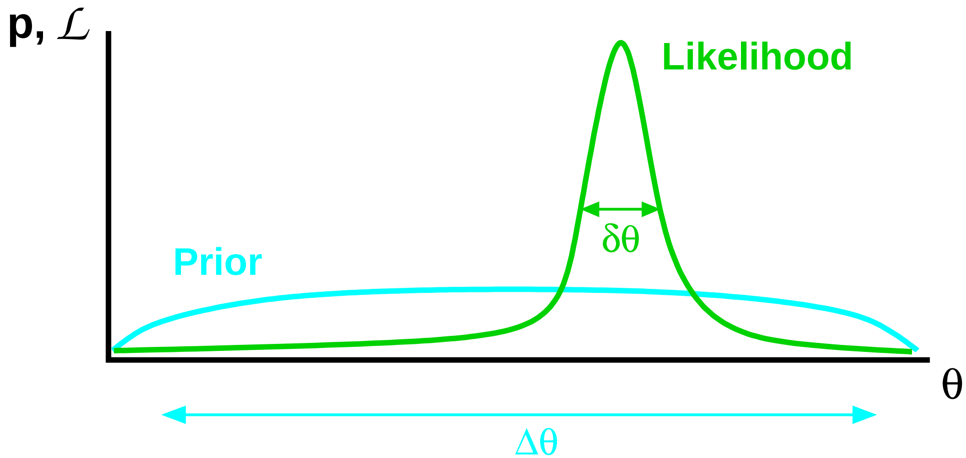
An Automatic Occam's Razor

Predictive probabilities can favor simpler models:

$$p(D|M_i) = \int d\theta_i \; p(\theta_i|M) \; \mathcal{L}(\theta_i)$$



The Occam Factor:



$$
\begin{aligned}
p(D|M_i) &= \int d\theta_i \; p(\theta_i|M) \; \mathcal{L}(\theta_i) \approx p(\hat{\theta}_i|M)\mathcal{L}(\hat{\theta}_i)\delta\theta_i \\
&\approx \mathcal{L}(\hat{\theta}_i)\frac{\delta\theta_i}{\Delta\theta_i} \\
&= \text{Maximum Likelihood} \times \text{Occam Factor}
\end{aligned}
$$

Models with more parameters often make the data more probable— *for the best fit*.

Occam factor penalizes models for “wasted” volume of parameter space.

Coin Flipping: Binomial Distribution

Suppose $D = n$ (number of heads in N flips), rather than the actual sequence. What is $p(a|n, M)$?

Likelihood:

Let $S =$ a sequence of flips with n heads.

$$\begin{aligned} p(n|a, M) &= \sum_S p(S|a, M) p(n|S, a, M) \\ &= a^n (a - a)^{N-n} C_{n,N} \end{aligned}$$

$C_{n,N} =$ # of sequences of length N with n heads.

$$\rightarrow p(n|a, M) = \frac{N!}{n!(N-n)!} a^n (1-a)^{N-n}$$

The *Binomial Distribution* for n given a, N.

About the Prior. . .

A prior for the *parameters* implies a prior predictive distribution for the *data*:

$$p(D|M) = \int d\theta\; p(D|\theta, M)$$

Calculating this can give insight about just what the prior is saying. E.g., Bayes's flat prior for binomial:

$$p(n_A|M_1) = \frac{1}{N_{\rm tot} + 1}$$

Log-flat (scale invariant) prior for Poisson counts:

$$p(n|M) = \frac{1}{n}$$

This only makes sense if $n = 0$ *is not in the sample space*.

Posterior:

$$p(a|n,M) = \frac{\frac{N!}{n!(N-n)!}a^n(1-a)^{N-n}}{p(n|M)}$$

$$\begin{aligned} p(n|M) &= \frac{N!}{n!(N-n)!}\int da\ a^n(1-a)^{N-n} \\ &= \frac{1}{N+1} \end{aligned}$$

$$\rightarrow p(a|n,M) = \frac{(N+1)!}{n!(N-n)!}a^n(1-a)^{N-n}$$

Same result as when data specified the actual sequence.

Stopping Rules in Inference

Consider repeated sampling, $\theta = p(\text{type A}|M_i)$

$a_p = 0.1$, prediction to be tested

Data $m = 5$ type A in $N = 96$ total samples

M_1: Binomial sampling (n_A is random)

$$p(n_A|a, M_1) = \frac{N_{\rm tot}!}{n_A!(N_{\rm tot} - n_A)!} a^{n_A} (1-a)^{N_{\rm tot} - n_A}$$

M_2: Negative binomial sampling ($N_{\rm tot}$ is random)

$$p(N_{\rm tot}|a, M_1) = \frac{(N_{\rm tot} - 1)!}{(n_A - 1)!(N_{\rm tot} - n_A)!} a^{n_A} (1-a)^{N_{\rm tot} - n_A}$$

Frequentist hypothesis tests

Significance level for data if binomial:

$$\begin{aligned}\alpha_1 &= \sum_{n_A=0}^{m} p(n_A|a_p, M_1) \\ &= 0.07 \\ &= 0.12 \quad (\chi^2 \text{ approx})\end{aligned}$$

Significance level for data if negative binomial:

$$\begin{aligned}\alpha_2 &= \sum_{N_{\rm tot}=0}^{N} p(N_{\rm tot}|a_p, M_2) \\ &= 0.03\end{aligned}$$

Can reach different conclusions from same data!

Especially troubling for observational (vs. experimental) science

Bayesian analysis

$$p(D|a_p, M_i) = C_i(n,N)\; a_p^n (1-a_p)^{N-n}$$

For comparison, take $M_0 : \theta$ has some other value, $p(a|M_0) = 1$:

$$p(D|M_0) = C_i(n,N) \int da\; a^n (1-a)^{N-m}$$

Bayes factor

$$B = \frac{p(D|M_i)}{p(D|M_0)} = \frac{\int da\; a^n (1-a)^{N-m}}{a_p^n (1-a_p)^{N-n}}$$

$C_i(n,N)$ *drops out!*

Edwards, Lindman & Savage 1966

Suppose that you collect data of any kind whatsoever—not necessarily Bernoullian, nor identically distributed, nor independent of each other given the parameter λ—stopping only when the data thus far collected satisfy some criterion of a sort that is sure to be satisfied sooner or later, then the import of the sequence of n data actually observed will be exactly the same as it would be had you planned to take exactly n observations in the first place. It is not even necessary that you stop according to a plan. You may stop when tired, when interrupted by the telephone, when you run out of money, when you have the casual impression that you have enough data to prove your point, and so on. The one proviso is that the moment at which your observation is interrupted must not in itself be any clue to λ that adds anything to the information in the data already at hand. *A man who wanted to know how frequently lions watered at a certain pool was chased away by lions before he actually saw any of them watering there; in trying to conclude how many lions do water there he should remember why his observation was interrupted when it was. We would not give a facetious example had we been able to think of a serious one.*

Lessons from Coin Flip Analyses

- Sufficiency: Calculation of the likelihood identified a sufficient statistic
- Only dependence of likelihood on *parameters* matters
- Demonstration of “Occam factors”:
  - Simpler model can be favored
  - Experiments can have limited strength

Inference With Gaussians

Gaussian PDF:

$$p(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \text{ over } [-\infty, \infty]$$

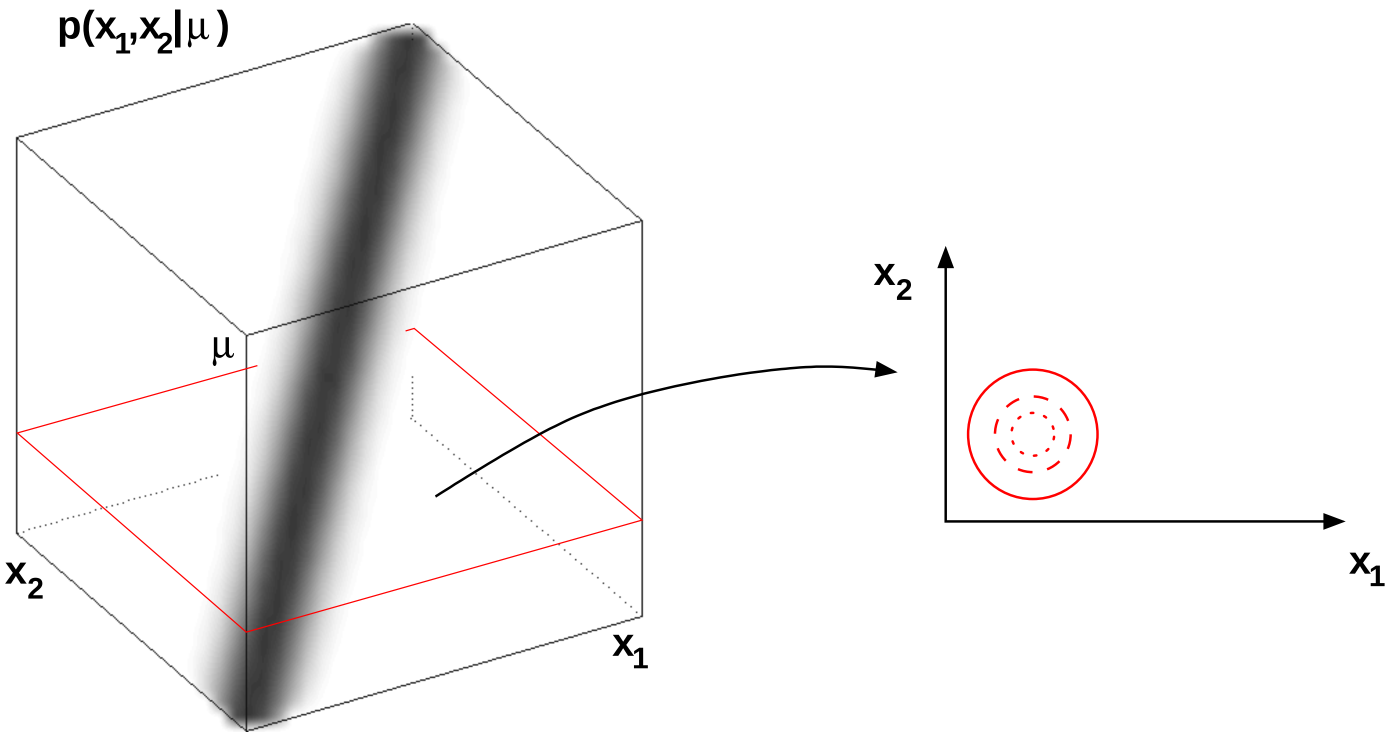
Common abbreviated notation: $x \sim N(\mu, \sigma^2)$

Parameters:

$$\begin{aligned} \mu &= \langle x \rangle \equiv \int dx\; x\, p(x|\mu,\sigma) \\ \sigma^2 &= \langle (x-\mu)^2 \rangle \equiv \int dx\; (x-\mu)^2\, p(x|\mu,\sigma) \end{aligned}$$

A Frequentist Confidence Region

Infer μ : $x_i = \mu + \epsilon_i;$ $p(x_i|\mu, M) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right]$



68% confidence region: $\bar{x} \pm \sigma/\sqrt{N}$

Monte Carlo Algorithm:

1. Pick a null hypothesis, $\mu = \mu_0$
2. Draw $x_i \sim N(\mu_0, \sigma^2)$ for $i = 1$ to N
3. Find $\bar{x}$; check if $\mu_0 \in \bar{x} \pm \sigma/\sqrt{N}$
4. Repeat $M >> 1$ times; report fraction (≈ 0.683)
5. *Hope result is independent of μ_0!*

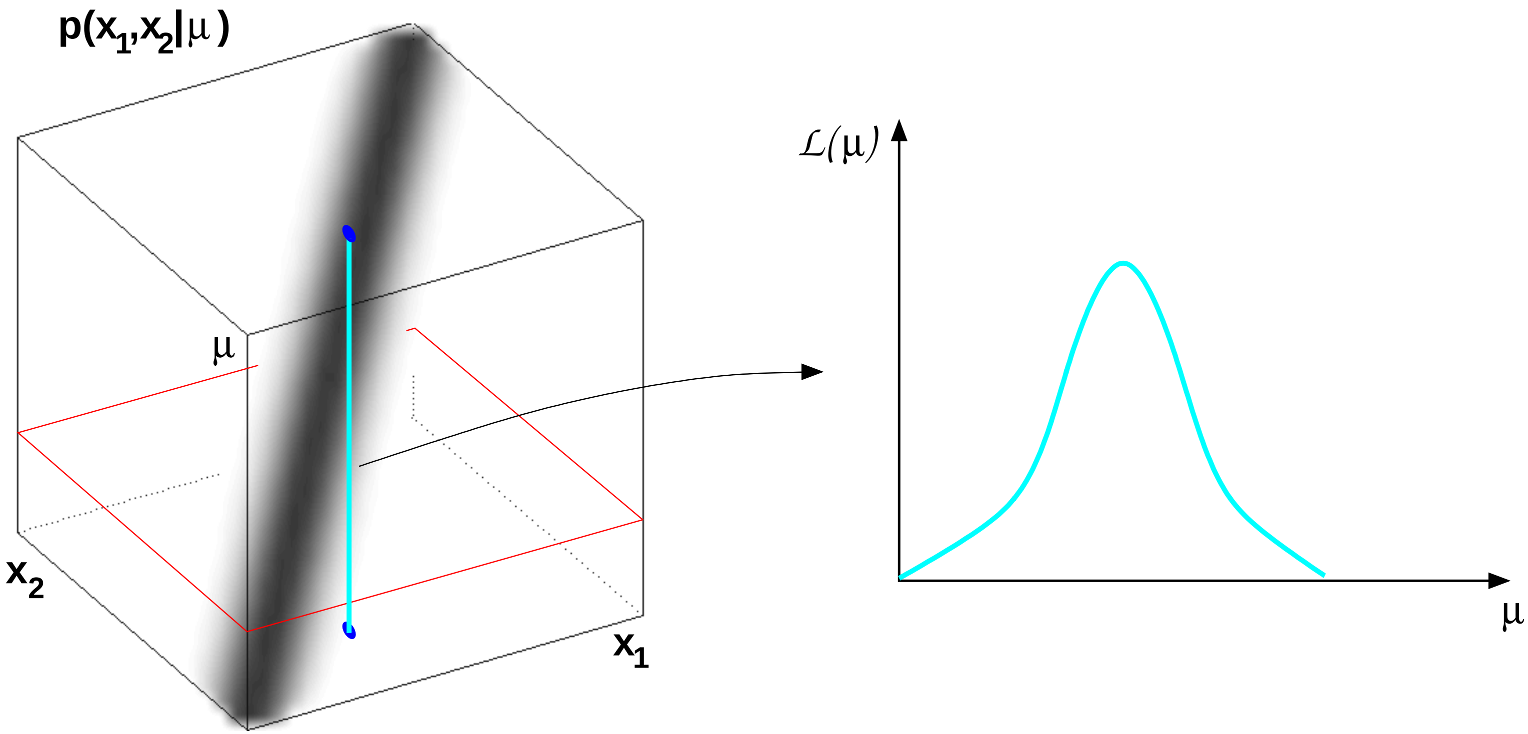
A Monte Carlo calculation of the N-dimensional integral:

$$\int dx_1 \frac{e^{-\frac{(x_1-\mu)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}} \cdots \int dx_N \frac{e^{-\frac{(x_N-\mu)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}} \times [\mu_0 \in \bar{x} \pm \sigma/\sqrt{N}]$$

$$= \int d(\text{angles}) \int_{\bar{x}-\sigma/\sqrt{N}}^{\bar{x}+\sigma/\sqrt{N}} d\bar{x} \exp\left[-\frac{(\bar{x}-\mu)^2}{2(\sigma/\sqrt{N})^2}\right] \cdots \approx 0.683$$

A Bayesian Credible Region

Infer μ : Flat prior; $\mathcal{L}(\mu) \propto \exp\left[-\frac{(\bar{x}-\mu)^2}{2(\sigma/\sqrt{N})^2}\right]$



68% credible region: $\bar{x} \pm \sigma/\sqrt{N}$

68% credible region: $\bar{x} \pm \sigma/\sqrt{N}$

$$\frac{\int_{\bar{x}-\sigma/\sqrt{N}}^{\bar{x}+\sigma/\sqrt{N}} d\mu \exp\left[-\frac{(\bar{x}-\mu)^2}{2(\sigma/\sqrt{N})^2}\right]}{\int_{-\infty}^{\infty} d\mu \exp\left[-\frac{(\bar{x}-\mu)^2}{2(\sigma/\sqrt{N})^2}\right]} \approx 0.683$$

Equivalent to a Monte Carlo calculation of a 1-d integral:

1. Draw μ from $N(\bar{x}, \sigma^2/N)$ (i.e., prior $\times \mathcal{L}$)
2. Repeat $M >> 1$ times; histogram
3. Report most probable 68.3% region

This simulation uses hypothetical *hypotheses* rather than hypothetical *data*.

Informative Conjugate Prior:

Use a Gaussian prior, $\mu \sim N(\mu_0, w_0^2)$

Posterior:

Remains Gaussian with

$$\begin{aligned} \hat{\mu} &= \frac{\bar{d}}{1 + \frac{w^2}{w_0^2}} + \frac{\mu_0}{1 + \frac{w_0^2}{w^2}} \\ w' &= w \frac{1}{\sqrt{1 + w^2/w_0^2}} \end{aligned}$$

“Principle of stable estimation:” The prior affects inferences only when data are not informative.

Estimating a Normal Mean: Unknown σ

Model: $d_i = \mu + \epsilon_i$, $\epsilon_i \sim N(0, \sigma^2)$, σ is *unknown*

Parameter space: (μ, σ); seek $p(\mu|D, \sigma, M)$

Likelihood:

$$\begin{aligned} p(D|\mu,\sigma,M) &= \frac{1}{\sigma^N (2\pi)^{N/2}} \exp\left(-\frac{Nr^2}{2\sigma^2}\right) \exp\left(-\frac{N(\mu-\bar{d})^2}{2\sigma^2}\right) \\ &\propto \frac{1}{\sigma^N} e^{-Q/2\sigma^2} \end{aligned}$$

where $\quad Q = N\left[r^2 + (\mu - \bar{d})^2\right]$

Uninformative Priors:

Assume priors for μ and σ are independent.

Translation invariance $\Rightarrow p(\mu) \propto C$, a constant.

Scale invariance $\Rightarrow p(\sigma) \propto 1/\sigma$.

Joint Posterior for μ, σ:

$$p(\mu, \sigma|D, M) \propto \frac{1}{\sigma^{N+1}} e^{-Q/2\sigma^2}$$

Marginal Posterior:

$$p(\mu|D,M) \propto \int d\sigma \; \frac{1}{\sigma^{N+1}} e^{-Q/2\sigma^2}$$

Let $\tau = \frac{Q}{2\sigma^2}$ so $\sigma = \sqrt{\frac{Q}{2\tau}}$ and $|d\sigma| = \tau^{-3/2}\sqrt{\frac{Q}{2}}$

$$\begin{aligned} \Rightarrow p(\mu|D,M) &\propto 2^{N/2} Q^{-N/2} \int d\tau \; \tau^{\frac{N}{2}-1} e^{-\tau} \\ &\propto Q^{-N/2} \end{aligned}$$

Write $Q = Nr^2\left[1 + \left(\frac{\mu-\bar{d}}{r}\right)^2\right]$ and normalize:

$$p(\mu|D,M) = \frac{\left(\frac{N}{2}-1\right)!}{\left(\frac{N}{2}-\frac{3}{2}\right)!\sqrt{\pi}}\frac{1}{r}\left[1+\frac{1}{N}\left(\frac{\mu-\bar{d}}{r/\sqrt{N}}\right)^2\right]^{-N/2}$$

"Student's t distribution," with $t = \frac{(\mu-\bar{d})}{r/\sqrt{N}}$

A "bell curve," but with power-law tails

Large N:

$$p(\mu|D,M) \sim e^{-N(\mu-\bar{d})^2/2r^2}$$

Key Operational Distinctions

- The role of subjectivity:
  - $\mathcal{F}$: Implicitly subjective — choice of null & statistic
  - $\mathcal{B}$: Explicitly subjective — choice of alternatives, priors
- Design of procedures:
  - $\mathcal{F}$ is a solution-characterization approach
  - $\mathcal{B}$ is a problem-solving approach
- Conditioning:
  - $\mathcal{F}$ conditions on hypotheses, $\mathcal{B}$ conditions on D_{obs}
- Types of integrals/sums:
  - $\mathcal{F}$ requires integrals over sample/data space
  - $\mathcal{B}$ requires integrals over hypothesis/parameter space

Provocative Reflections

Philip Dawid (2000)

What is the principal distinction between Bayesian and classical statistics? It is that Bayesian statistics is fundamentally boring. There is so little to do: just specify the model and the prior, and turn the Bayesian handle. There is no room for clever tricks or an alphabetic cornucopia of definitions and optimality criteria. I have heard people use this ‘dullness’ as an argument against Bayesianism. One might as well complain that Newton’s dynamics, being based on three simple laws of motion and one of gravitation, is a poor substitute for the richness of Ptolemy’s epicyclic system.

All my experience teaches me that it is invariably more fruitful, and leads to deeper insights and better data analyses, to explore the consequences of being a ‘thoroughly boring Bayesian’.

David Lindley (2000)

The philosophy places more emphasis on model construction than on formal inference. . . I do agree with Dawid that ‘Bayesian statistics is fundamentally boring’. . . My only qualification would be that the theory may be boring but the applications are exciting.

Suggested Reading

See my web site for links to online tutorials and other resources:

`http://www.astro.cornell.edu/staff/loredo/bayes/`

Recent books targeted toward physical scientists and engineers:

- Ed Jaynes's *Probability Theory: The Logic of Science*
  A deep and definitive text; very readable and lengthy treatment emphasizing important conceptual issues; no treatment of modern computational techniques; not the place to go for a quick, practical introduction!
- Devinder Sivia's *Data Analysis: A Bayesian Tutorial*
  Practical introduction targeting advanced undergrads and grad students; very clear treatment of basic principles; somewhat quirky "Cambridge school" terminology; little treatment of modern computational techniques (but a new edition is forthcoming)
- Prasenjit Saha's *Principles of Data Analysis*
  Short, informal intro, inexpensive, free online version:
  `http://ankh-morpork.maths.qmw.ac.uk/~saha/book/`

- Phil Gregory's *Bayesian Logical Data Analysis for the Physical Sciences*
  New in Summer 2005
- David MacKay's *Information Theory, Inference, and Learning Algorithms*
  Targeted to computer scientists and the machine learning community; emphasizes information theory issues, but full of original insights valuable to physicists; includes very good intro to modern Monte Carlo techniques; free for online viewing (not printing) at `http://www.inference.phy.cam.ac.uk/mackay/itila/book.html`
- Chen, Shao, Ibrahim's *Monte Carlo methods in Bayesian computation*
  Targeted to engineers; focuses on algorithms; `FORTRAN` software available at authors' site `http://merlot.stat.uconn.edu/~mhchen/mcbook/`; a colleague reports numerous errors in formulae—be sure to check the errata on the web site

There is a wealth of new books on Bayesian inference by statisticians. Two recent and highly regarded books are: *Bayesian Theory* by J. Bernardo & A. Smith; and *The Bayesian Choice* by C. Robert. A very accessible introduction is *Bayesian Statistics* (3rd edn.) by P. Lee. See the ISBA web site for a list of dozens of other choices: `http://www.bayesian.org/`. Two noteworthy recent review articles are: D. V. Lindley (2000), "The philosophy of statistics" (*The Statistician*, **49**, 293–337); J. Bernardo (2003), "Bayesian Statistics" (*Encyclopedia of Life Support Systems*, paper 60 at `http://www.uv.es/~bernardo/publications.html`).