Introduction to Bayesian inference for astronomy, 1

Tom Loredo
Cornell Center for Astrophysics and Planetary Science,
Carl Sagan Institute,
& Dept. of Statistics and Data Science, Cornell U.
http://hosting.astro.cornell.edu/~loredo/

CASt Summer School

Grant support from NSF AAG & Statistics
AST-2206339, DMS-2210790

This summer school content is being maintained on GitHub:

`https://github.com/tloredo/SummerSchool2025-IntroBayes/`

Besides lecture notes:

- *Resources*: Annotated list of selected texts, tutorials, software on Bayesian data analysis for astronomers, and other scientists
- *Supplement*: Details and extra topics not covered in the lecture

Q&A:

- Use the scheduled live Q&A session.
- Also post questions to Slack during or after viewing the lectures; I'll monitor Slack through this week. (Check already-posted questions before posting.)
- Feel free to email me with questions about the lecture content, or to discuss Bayesian ideas relevant to your work.

Agenda

❶ **The big picture: Statistics and the scientific method**

❷ **Quantifying uncertainty with probability**

❸ **Motivating example: $\bar{x} \pm \sigma/\sqrt{N}$ via Monte Carlo**
Confidence intervals vs. credible intervals

Agenda

1. **The big picture: Statistics and the scientific method**
2. **Quantifying uncertainty with probability**
3. **Motivating example: $\bar{x} \pm \sigma/\sqrt{N}$ via Monte Carlo**
   - Confidence intervals vs. credible intervals

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

Scientists *argue!*

Argument $\equiv$ Collection of statements comprising an act of reasoning from *premises* to a *conclusion*

A key goal of science: Explain or predict *quantitative measurements* (data!)

Data analysis: Constructing and appraising arguments that reason from data to interesting scientific conclusions (explanations, predictions)

The role of data

Data do not speak for themselves!

"No body of data tells us all we need to know about its own analysis."
— John Tukey, *EDA*

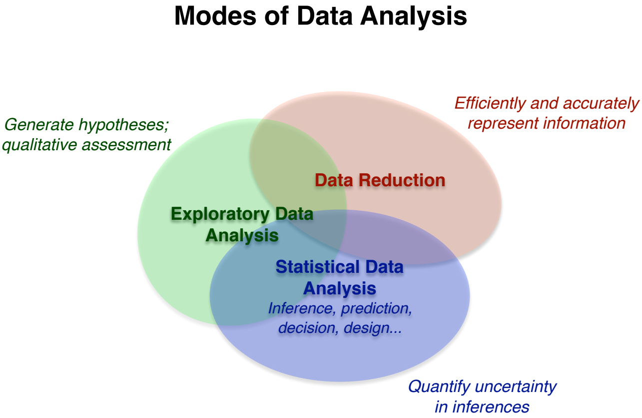
We don't just *tabulate* data, we *analyze* data

We gather data so they may speak for or against existing hypotheses, and guide the formation of new hypotheses

A key role of data in science is to be among the premises in scientific arguments—the evidence for evidence-based reasoning

Data analysis

Building & Appraising Arguments Using Data



Inference: Learning about the *data generating process* (population, signals...) from observed data—just one of several interacting modes of analyzing data

Statistical inference as a style of reasoning

See *MSMA* (Feigelson & Babu 2012), § 1.1.3, "Statistics and science":

For better or worse, statistical inference has provided an entirely new *style of reasoning*. The quiet statisticians have changed our world—not by discovering new facts or technical developments but by changing the ways we reason, experiment, and form our opinions about it.

—*Ian Hacking, philosopher of science (in Science '84)*

[From "Trial by number; Karl Pearson's chi-square test…" in *Science '84*, vol. 5, 69–71 AAAS (1984)]

Two main styles of statistical reasoning

Statistics has not yet aged into a stable discipline with complete agreement on foundations. All the statisticians mentioned here [Pearson, Galton, Fisher] assumed that *the key to probability lay in the relative frequency with which different kinds of events occur*. . . . But what does that mean? Some say nothing, for probability is concerned with subjective degrees of belief, and that subjective approach only gives a *reasonable degree of certainty*. Work emanating from F. P. Ramsey in England, Bruno de Finetti in Italy, and L. J. Savage in the United States has turned such subjectivity into a serious scientific approach. Today we have vigorous, sometimes violent, disagreement on these matters, but perhaps battles about first principles are less important than the large-scale application of many competing methods.

—*Ian Hacking (1984)*

Entry points for literature comparing Bayesian and frequentist approaches:

- Jaynes (1976): Confidence Intervals vs Bayesian Intervals (article # 32)
- Diaconis & Skyrms (2017): *Ten Great Ideas about Chance* (See *Resources*); focuses on foundations rather than applications
- Some TL papers on foundations/comparisons:
    - The promise of Bayesian inference for astrophysics (1992); unabridged version at BIPS
    - On the future of astrostatistics: statistical foundations and statistical practice (2012)
    - Bayesian astrostatistics: A backward look to the future (2013)

Agenda

1. The big picture: Statistics and the scientific method
2. **Quantifying uncertainty with probability**
3. Motivating example: $\bar{x} \pm \sigma/\sqrt{N}$ via Monte Carlo
   - Confidence intervals vs. credible intervals

Fundamental principle

"The most fundamental principle of the statistical paradigm,
its starting point,
is that variation may be described by probability."

Fundamental principle

~~*"The most fundamental principle of the statistical paradigm,*~~
~~*its starting point,*~~
~~*is that variation may be described by probability."*~~

Fundamental principle

The most fundamental principle of the statistical paradigm,
its starting point,
is that **uncertainty** *may be described by probability.*

Fundamental principle

The most fundamental principle of the statistical paradigm,
its starting point,
is that **uncertainty** *may be described by probability.*

An important corollary is that, in some settings
—most notably, for **IID replications**,
and for **exchangeable sequences**—
individual-case probability and multi-case expected variation
are intimately linked.

The “classical” understanding of probability

Pierre Simon Laplace (1819)

Probability theory is nothing but *common sense reduced to calculation*.

James Clerk Maxwell (1850)

They say that Understanding ought to work by the rules of right reason. These rules are, or ought to be, contained in Logic, but the actual science of *Logic is conversant at present only with things either certain, impossible, or entirely doubtful*, none of which (fortunately) we have to reason on. Therefore *the true logic of this world is the calculus of Probabilities*, which takes account of the magnitude of the probability which is, or ought to be, in a reasonable man’s mind.

Harold Jeffreys (1931)

If we like there is no harm in saying that a *probability expresses a degree of reasonable belief*. . . . ‘Degree of confirmation’ has been used by Carnap, and possibly avoids some confusion. But whatever verbal expression we use to try to convey the primitive idea, this expression cannot amount to a definition. *Essentially the notion can only be described by reference to instances where it is used.* It is intended to express *a kind of relation between data and consequence* that habitually arises in science and in everyday life, and the reader should be able to recognize the relation from examples of the circumstances when it arises.

Probability theory as (generalized) logic

"Logic can be defined as *the analysis and appraisal of arguments*"
—Gensler, *Intro to Logic*

Build arguments with *propositions* and logical *operators/connectives*

- *Propositions:* Statements that may be true or false

$$\begin{aligned} \mathcal{P} :& \quad \text{Universe can be modeled with } \Lambda\text{CDM} \\ A :& \quad \Omega_{\rm tot} \in [0.9, 1.1] \\ B :& \quad \Omega_\Lambda \text{ is not } 0 \\ \overline{B} :& \quad \text{"not } B\text{," i.e., } \Omega_\Lambda = 0 \end{aligned}$$

  Events in freq. PT are propositions about outcomes in repeated trials

- *Connectives:*

$$\begin{aligned} A \wedge B \text{ or } A, B :& \quad A \text{ and } B \text{ are } \textit{both} \text{ true} \\ A \vee B :& \quad A \text{ or } B \text{ is true, or both are} \end{aligned}$$

Arguments

Argument: Assertion that an *hypothesized conclusion*, H, follows from *premises*, $\mathcal{P} = \{A, B, C, \ldots\}$ (take "," = "and")

Notation:

$H|\mathcal{P}$: Premises $\mathcal{P}$ imply H
H may be deduced from $\mathcal{P}$
H follows from $\mathcal{P}$
H is true given that $\mathcal{P}$ is true

Deductive logic applies when we can *reason with certainty*; can model this with *Boolean algebra* over $\{0, 1\}$ (False, True)

Classical/Bayesian PT applies when we must *reason amidst uncertainty*; it quantifies degree of certainty on a $[0, 1]$ scale, providing a mathematical model for *inductive reasoning*

Probability as argument strength

$$P(H|\mathcal{P}) \equiv \text{strength of argument } H|\mathcal{P}$$

$$\begin{aligned} P &= 1 \rightarrow \text{Argument is \textit{deductively valid}} \\ &= 0 \rightarrow \text{Premises imply } \overline{H} \\ &\in (0,1) \rightarrow \text{Degree of deducibility/entailment} \end{aligned}$$

Mathematical model for inductive reasoning

$$\text{`AND' (product rule):} \quad P(A \wedge B|\mathcal{P}) = P(A|\mathcal{P})\, P(B|A \wedge \mathcal{P})$$
$$= P(B|\mathcal{P})\, P(A|B \wedge \mathcal{P})$$

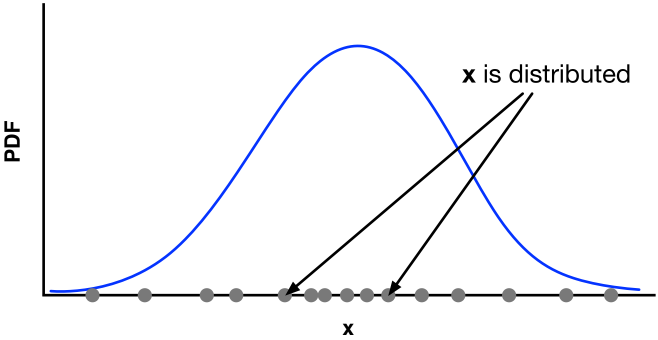
$$\text{`OR' (sum rule):} \quad P(A \vee B|\mathcal{P}) = P(A|\mathcal{P}) + P(B|\mathcal{P})$$
$$-P(A \wedge B|\mathcal{P})$$

$$\text{`NOT':} \quad P(\overline{A}|\mathcal{P}) = 1 - P(A|\mathcal{P})$$

Interpreting PDFs

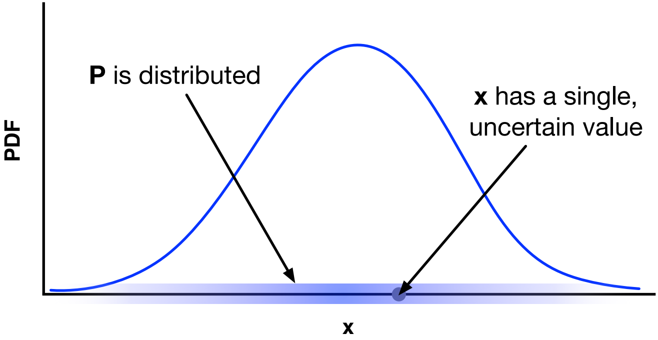
Frequentist

Probabilities are always (limiting) rates/proportions/frequencies that *quantify variability* in a sequence of trials. $p(x)$ describes how the *values of x* would be distributed among *infinitely many trials*:



Bayesian

Probability *quantifies uncertainty* in an inductive inference. $p(x)$ describes how *probability* is distributed over the possible values x might have taken in *the single case before us*:



Probability & frequency in IID settings

Consider a setting where we assign the same probability to many independent outcomes (flips of a coin, rolls of a die, star/galaxy classifications, searches for an Earth around a G dwarf. . .):

- If the probability is high, we expect the outcomes to occur frequently
- If the probability is low, we expect the outcomes to occur rarely

In IID repeated trial settings, it seems there should be a relationship between single-trial probability and multiple-trial (relative) frequency

Frequency from probability

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

$$\frac{n_{\text{success}}}{N_{\text{total}}} \to \alpha \quad \text{as} \quad N_{\text{total}} \to \infty$$

If $P(\text{success}|\ldots)$ does not change from sample to sample, it may be interpreted as the expected relative frequency

Probability from frequency

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

Probability for success in next trial of IID sequence:

$$\mathrm{E}(\alpha) \to \frac{n_{\text{success}}}{N_{\text{total}}} \quad \text{as} \quad N_{\text{total}} \to \infty$$

If $P(\text{success}|\ldots)$ does not change from sample to sample, it may be estimated using relative frequency data

Aside: Twiddle notation for the normal dist'n

$$\mathrm{Norm}(x;\mu,\sigma) \equiv \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{\sigma^2}\right]$$

Frequentist

random *fixed but unknown*

$$p(\underbrace{x}_{\text{random}} ; \underbrace{\mu,\sigma}_{\text{fixed but unknown}}) = \mathrm{Norm}(x;\mu,\sigma)$$

$$x \sim \mathcal{N}(\mu,\sigma^2)$$

"*x is distributed* as normal with mean..."

"random" = "varies unpredictably in repeated trials"

Bayesian

random *random or known*

$$p(x \mid \mu,\sigma) = \mathrm{Norm}(x|\mu,\sigma)$$

$$x \sim \mathcal{N}(\mu,\sigma^2)$$

"*The probability for x is distributed* as normal with mean..."

"random" = "uncertain in the case at hand"

Bayesian inference

- Bayesian inference uses probability theory to *quantify the strength of data-based arguments* (i.e., a more abstract view than restricting PT to describe variability in repeated "random" experiments)
- A different approach to *all* statistical inference problems (i.e., not just another method in the list: BLUE, linear regression, least squares/χ^2 minimization, maximum likelihood, ANOVA, survival analysis, LDA classification . . .)
- Focuses on *deriving consequences of modeling assumptions* rather than *devising and calibrating procedures*

Frequentist vs. Bayesian statements

"The data D_{obs} support hypothesis H ... "

Frequentist assessment

"H was selected with a procedure that's right 95% of the time over a set $\{D_{hyp}\}$ that includes D_{obs}."

Probabilities are properties of *procedures*, not of particular results. Guaranteed long-run performance is the *sine qua non*.

Bayesian assessment

"The strength of the chain of reasoning from the model and D_{obs} to H is 0.95, on a scale where 1 = certainty."

Probabilities are associated with arguments that directly refer only to *actually observed data*.

Long-run performance must be separately evaluated (and is typically good by frequentist criteria).

Aside: Bayesian lingo

- *Bayesian probability theory*: Assigning and manipulating probabilities interpreted as argument strength/degree of belief/entailment strength
- *Bayesian inference*: Computing posterior probabilities for hypotheses (e.g., statements about parameters), or posterior predictive probabilities for future data
- *Bayesian decision theory*: Making optimal decisions given data, accounting for consequences of good or bad decisions (utility or loss functions); includes experimental design
- *Bayesian data analysis (BDA)*: Any mode of data analysis using a Bayesian interpretation of probability, including inference & decision/design, but also prediction, model checking (predictive tests), EDA/visualization, data reduction...

- *Bayesian statistics*: Informal catch-all for any one or more of the above
- *Bayesian workflow*: Emerging term of art for best practices, including model checking and adjustment/refinement

We'll discuss only Bayesian inference here,
but the other areas are also important

Agenda

1. The big picture: Statistics and the scientific method
2. Quantifying uncertainty with probability
3. **Motivating example: $\bar{x} \pm \sigma/\sqrt{N}$ via Monte Carlo**
   Confidence intervals vs. credible intervals

A Simple (?) confidence region

Problem

Estimate the location parameter (mean, μ) of a Gaussian distribution from a set of N IID samples $D = \{x_i\}$. Report a region summarizing the uncertainty.

Here assume std dev'n σ is *known*; we are uncertain only about μ.

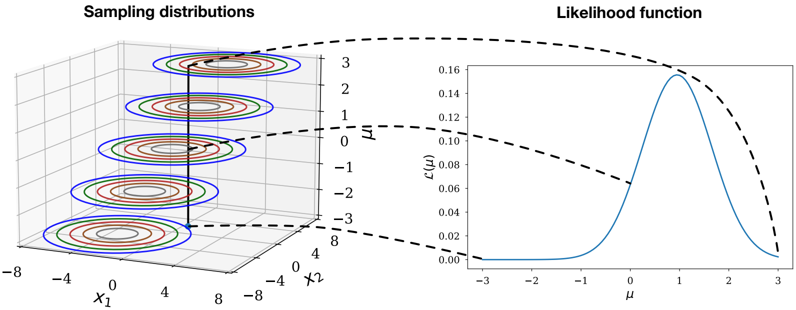
Model

The *sampling distribution* for *any* $(x_1, \ldots, x_N)$ has PDF

$$\begin{aligned} p(x_1, \ldots, x_N|\mu) &= \prod_i^N \frac{1}{\sigma\sqrt{2\pi}} e^{-(x_i-\mu)^2/2\sigma^2}; \qquad \sigma = 1 \\ &\propto e^{-\chi^2(\mu)/2} \end{aligned}$$

This gives the *likelihood function*, $\mathcal{L}(\mu)$ if we set $(x_1, \ldots)$ to the *observed values*. Units $= 1/[x]^N$.

Sampling distributions and likelihood function



The likelihood function shows how well each of the candidate sampling distributions—labeled by the parameter, μ—predicts the observed data (x_1, x_2); *predictive function* or *prognostic function* might have been less confusing.

The *likelihood for the parameter* is the (sampling) *probability for the observed data*; "likelihood for the data" is incorrect usage—it entirely misses the point of likelihood!

Fisher on likelihood

"If we need a word to characterise this relative property of different values of [a parameter] p, I suggest that we may speak without confusion of the likelihood of one value of p being thrice the likelihood of another, bearing always in mind that *likelihood is not here used loosely as a synonym of probability*, but simply to express the relative frequencies with which such values of the hypothetical quantity p would in fact yield the observed sample." (Fisher 1922)

"Likelihood also *differs from probability* in that it is a differential element, and is *incapable of being integrated*: it is assigned to a particular point of the range of variation, not to a particular element [interval]." (Fisher 1922)

"...the integration with respect to [parameter] m is illegitimate and has no definite meaning..." (Fisher 1912)

Classes of variables—the two spaces

- μ is the unknown we seek to estimate—the *parameter*. The *parameter space* is the space of possible values of μ—here the real line (perhaps bounded). *Hypothesis space* is a more general term.
- A particular set of N data values $D = \vec{x} = (x_1, \ldots, x_N)$ is a *sample*. The *sample space* is the N-dimensional space of possible samples. The *observed* data correspond to a single point in this space.

A statistical model connects the two spaces:

- Specifying μ lets us *predict* D (via the sampling dist'n).
- Specifying $D = D_{\rm obs}$ lets us *infer* μ.

Roles for probability

- Both frequentist and Bayesian approaches require probability distributions on the sample space—*sampling distributions*.
- The frequentist perspective denies the legitimacy of probability distributions on the parameter space (the parameter value isn't random in the sense of stochastically varying). Inference is handled by describing the long-run performance of statistical procedures that produce data-based statements about parameters.
- The Bayesian approach allows probability distributions over parameter space. Inference corresponds to *learning*, going from a *prior distribution* in parameter space (prior to consideration of the observed data) to a *posterior distribution* (accounting for observed data).

Standard inferences for a normal mean

Inference takes us from the sample space to the parameter space.

Let $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$.

Point estimator (function): $\hat{\mu}(\vec{x}) = \bar{x}$.
Point estimate (scalar): $\hat{\mu}(\vec{x}_{\rm obs}) = \bar{x}_{\rm obs}$.

Uncertainty quantification:

- "Standard error" (rms error) is $\sigma/\sqrt{N}$
- "1σ" interval: $\bar{x} \pm \sigma/\sqrt{N}$ with conf. level CL $= 68.3\%$
- "2σ" interval: $\bar{x} \pm 2\sigma/\sqrt{N}$ with CL $= 95.4\%$

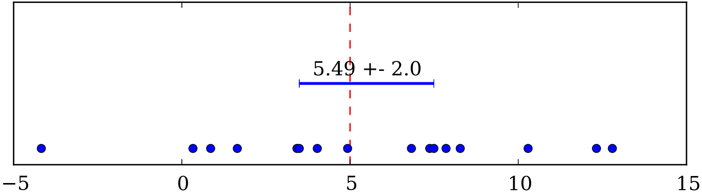
How are the uncertainty quantification results found?

Take a Monte Carlo perspective—for both frequentist & Bayesian approaches.

Some simulated data

Take $\mu = 5$ and $\sigma = 4$ and $N = 16$, so $\sigma/\sqrt{N} = 1$.

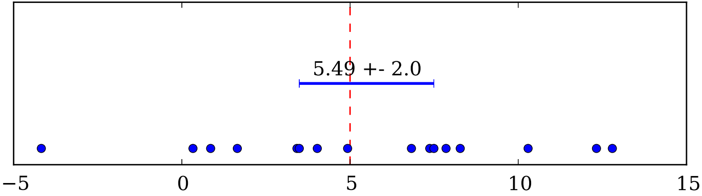
What is the CL associated with this interval?



Some simulated data

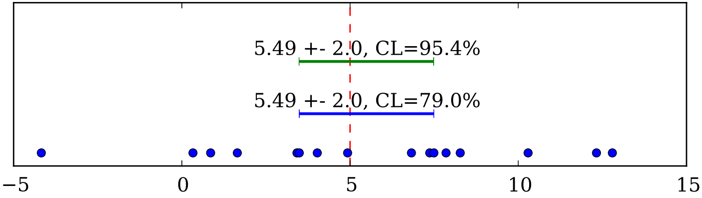
Take $\mu = 5$ and $\sigma = 4$ and $N = 16$, so $\sigma/\sqrt{N} = 1$.

What is the CL associated with this interval?



The confidence level for this interval is 79.0%.

Two intervals



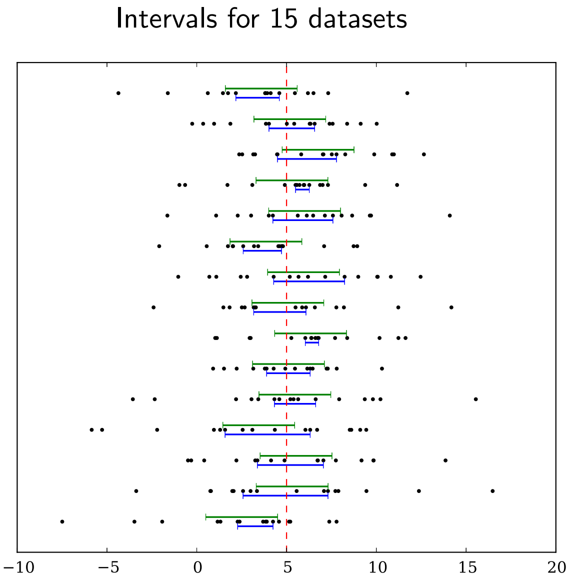
- Green interval: $\bar{x} \pm 2\sigma/\sqrt{N}$
- Blue interval: Let $x_{(k)} \equiv k$'th order statistic
  Report $[x_{(6)}, x_{(11)}]$ (i.e., leave out 5 outermost each side)

The point

The (frequentist) confidence level is a property of the ***procedure****, not of the particular interval reported for a given dataset.*

Rigorous frequentist probabilities are thus *prior probabilities*, computed without reference to the observed data—but they are priors for statistics, not for hypotheses.

Performance of intervals



Confidence interval for a normal mean

Suppose we have a sample of $N = 5$ values x_i,

$$x_i \sim N(\mu, 1)$$

We want to estimate μ, including some *quantification of uncertainty* in the estimate: an interval *with a probability attached*

Frequentist approaches: method of moments, BLUE, least-squares/χ^2, maximum likelihood

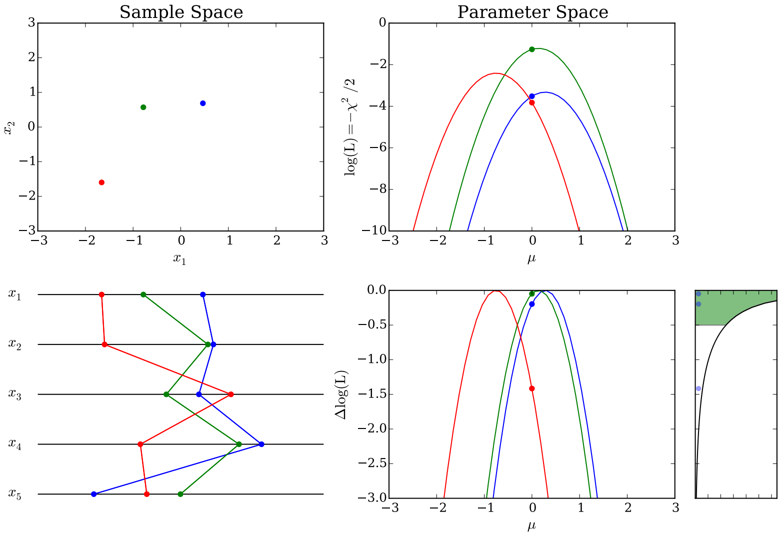
Focus on likelihood (equivalent to χ^2 here); this is closest to Bayes:

$$\begin{aligned}
\mathcal{L}(\mu) &\equiv p(\{x_i\}|\mu) \\
&= \prod_i \frac{1}{\sigma\sqrt{2\pi}} e^{-(x_i-\mu)^2/2\sigma^2}; \qquad \sigma = 1 \\
&\propto e^{-\chi^2(\mu)/2}
\end{aligned}$$

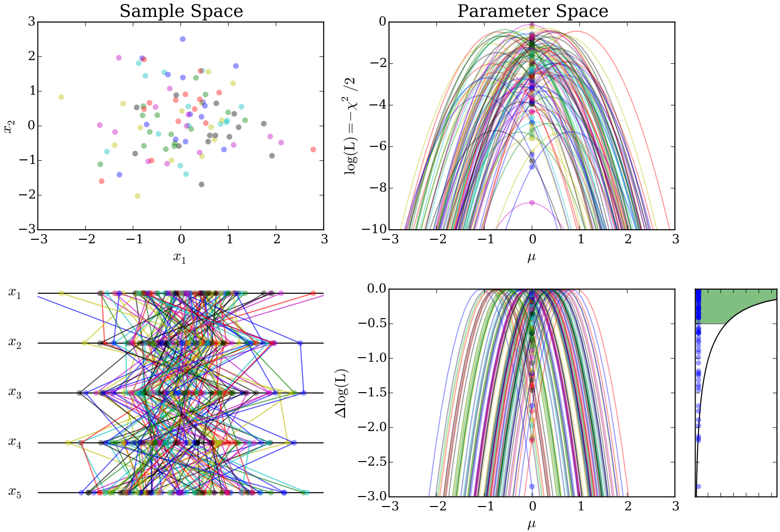
Estimate μ from maximum likelihood (minimum χ^2)
Define an interval and its coverage frequency from the $\mathcal{L}(\mu)$ curve

Construct an interval procedure for known μ

Likelihoods for 3 simulated data sets, $\mu = 0$

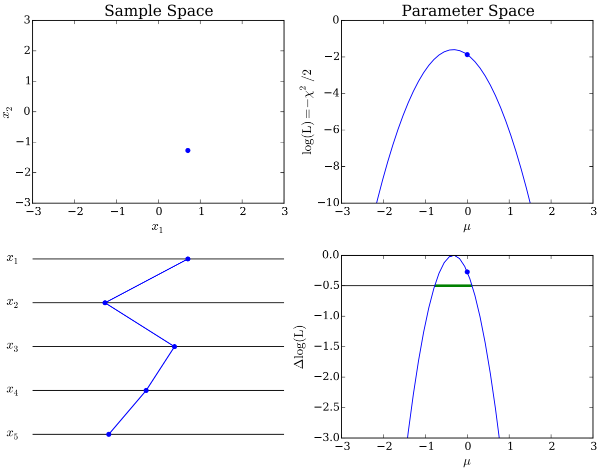


Likelihoods for 100 simulated data sets, $\mu = 0$



Careful! This is for $\mu = 0$, but μ will be unknown.
Luckily, the $\Delta \log(\mathcal{L})$ dist'n is independent of μ.

Apply to observed sample



Report the green region, reporting CL as the coverage calculated for ensemble of hypothetical data (green region, previous slide)

Credible interval for a normal mean

Recall the likelihood, $\mathcal{L}(\mu) \equiv p(D_{\rm obs}|\mu)$, is a probability for the observed data, but *not* for the parameter μ (wrong PDF units)

Convert likelihood to a probability distribution over μ via *Bayes's theorem* (changes units from per-unit-D to per-unit-μ):

$$\begin{aligned} p(\mu, D) &= p(\mu)p(D|\mu) \\ &= p(D)p(\mu|D) \\ \rightarrow p(\mu|D) &= p(\mu)\frac{p(D|\mu)}{p(D)}, \quad \text{Bayes's th.} \end{aligned}$$

$$\Rightarrow p(\mu|D_{\rm obs}) \quad \propto \quad \pi(\mu)\mathcal{L}(\mu) \qquad \text{(prior} \times \text{like.)}$$

$p(\mu|D_{\rm obs})$ is called the *posterior probability distribution for μ*

This requires a prior probability density, $\pi(\mu)$, often taken to be constant over the allowed region if there is no significant information available (or sometimes constant w.r.t. some reparameterization motivated by a symmetry in the problem)

Gaussian problem posterior distribution

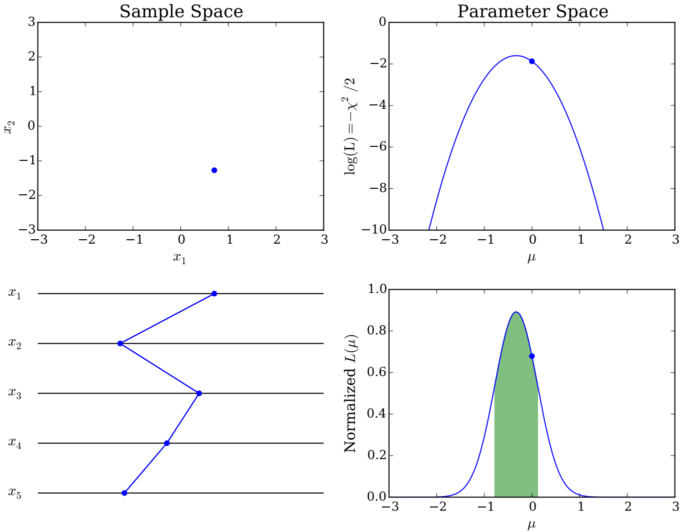
For the Gaussian example, a bit of algebra ("complete the square") gives:

$$\begin{aligned}
\mathcal{L}(\mu) &\propto \prod_i \exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right] \\
&\propto \exp\left[-\frac{1}{2}\sum_i \frac{(x_i-\mu)^2}{\sigma^2}\right] \\
&\propto \exp\left[-\frac{(\mu-\bar{x})^2}{2(\sigma/\sqrt{N})^2}\right]
\end{aligned}$$

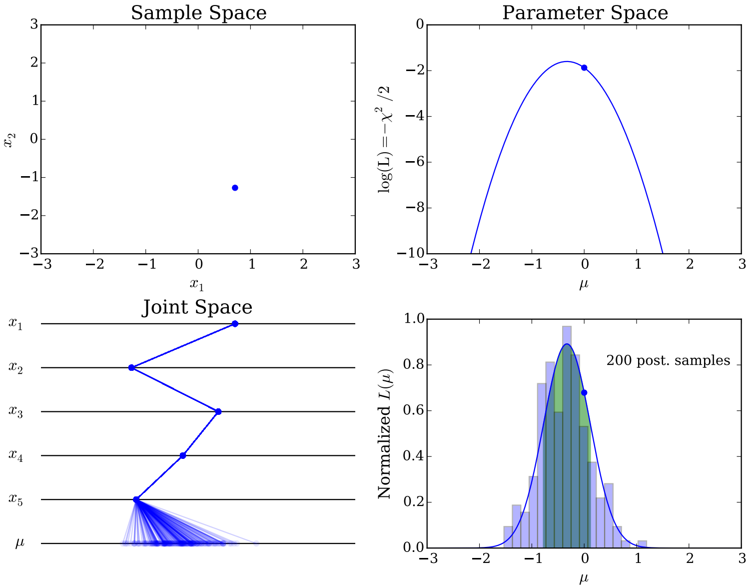
The likelihood is Gaussian in μ.
Flat prior $\rightarrow$ posterior density for μ is $\mathcal{N}(\bar{x}, \sigma^2/N)$.

Bayesian credible region

Normalize the likelihood for the observed sample; report the region that includes 68.3% of the normalized likelihood; green shows the *highest posterior density (HPD) region*:



Posterior sampling: Credible region via Monte Carlo (MCMC, ABC)



Posterior summaries

- Posterior mean is $\langle\mu\rangle \equiv \int \mathrm{d}\mu\, \mu\, p(\mu|D_{\text{obs}}) = \bar{x}$
- Posterior mode is $\hat{\mu} = \bar{x}$
- Posterior std dev'n is $\sigma/\sqrt{N}$
- $\bar{x} \pm \sigma/\sqrt{N}$ is a 68.3% *credible region*:

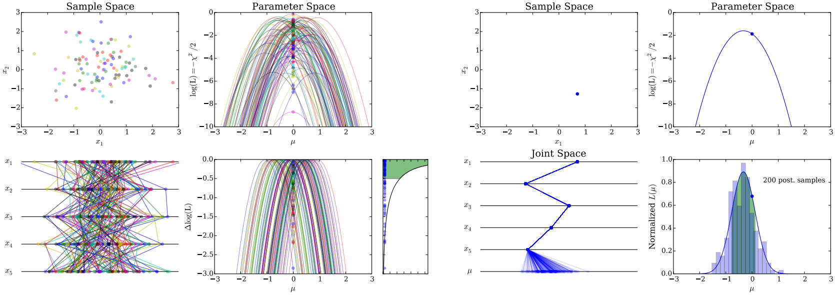
$$\int_{\bar{x}-\sigma/\sqrt{N}}^{\bar{x}+\sigma/\sqrt{N}} \mathrm{d}\mu\; p(\mu|D_{\text{obs}}) \approx 0.683$$

- $\bar{x} \pm 2\sigma/\sqrt{N}$ is a 95.4% credible region

The credible regions above are *highest posterior density* credible regions (HPD regions). These are the smallest regions with a specified probability content.

These reproduce familiar frequentist results, but this is a *coincidence* due to special properties of Gaussians.

Confidence vs. credible regions



Find lower/upper functions of the data that give desired values to:

$$\text{Coverage}(\mu) = \int \mathrm{d}^N\vec{x}\; p(\vec{x}|\mu)\; \mathbb{I}\left[l(\vec{x}) < \mu < u(\vec{x})\right]$$

$$\text{CredLev}(\vec{x}_{\rm obs}) = \int \mathrm{d}\mu\; p(\mu|\vec{x}_{\rm obs})\; \mathbb{I}\left[L(\vec{x}_{\rm obs}) < \mu < U(\vec{x}_{\rm obs})\right]$$

When the approaches differ

Both approaches report $\mu \in [\bar{x} - \sigma/\sqrt{N}, \bar{x} + \sigma/\sqrt{N}]$, and assign 68.3% to this interval (*with very different meanings!*)

This matching is a *coincidence*!

When might results differ? ($\mathcal{F}$ = frequentist, $\mathcal{B}$ = Bayes)

- If $\mathcal{F}$ procedure doesn't use likelihood directly
- If $\mathcal{F}$ procedure properties depend on params (e.g., nonlinear models; need to find pivotal quantities)
- If likelihood shape varies strongly between datasets (conditional inference, ancillary statistics, recognizable subsets)
- If there are extra uninteresting parameters (*nuisance parameters*; adjusted profile likelihood, conditional inference)
- If $\mathcal{B}$ uses important prior information

Also, for a different task—comparison of parametric models—the approaches are *qualitatively* different (significance tests, *p*-values, & information criteria vs. Bayes factors)

Bayesian and Frequentist inference

Brad Efron, ASA President (2005)

> The 250-year debate between Bayesians and frequentists is unusual among philosophical arguments in actually having *important practical consequences*.... The physicists I talked with were really bothered by our 250 year old Bayesian-frequentist argument. Basically there's only one way of doing physics but there seems to be at least two ways to do statistics, and *they don't always give the same answers*....
>
> Broadly speaking, Bayesian statistics dominated 19th Century statistical practice while the 20th Century was more frequentist. What's going to happen in the 21st Century?...I strongly suspect that statistics is in for a burst of new theory and methodology, and that this burst will feature a combination of Bayesian and frequentist reasoning....

Roderick Little, ASA President's Address (2005)

Pragmatists might argue that good statisticians can get sensible answers under Bayes or frequentist paradigms; indeed maybe two philosophies are better than one, since they provide more tools for the statistician's toolkit. . . . I am discomforted by this "inferential schizophrenia." Since the Bayesian (B) and frequentist (F) philosophies *can differ even on simple problems*, at some point decisions seem needed as to which is right. I believe our credibility as statisticians is undermined when we cannot agree on the fundamentals of our subject. . . .

An assessment of strengths and weaknesses of the frequentist and Bayes systems of inference suggests that *calibrated Bayes*. . . captures the strengths of both approaches and provides a roadmap for future advances.

[Calibrated Bayes = Bayesian inference within a specified space of models + frequentist-based model checking; Andrew Gelman et al. use *Bayesian data analysis* similarly]

(see TL's arXiv:1208.3035 for discussion/references)