Astro 193 : Feb 25

• Follow up
  • reduced $\chi^2$ : mean = 1 and variance = $2/\nu$
• non-linear least-squares fitting (contd.)
• Working through Homework 5
• Bayesian Probability Theory
• MCMC basics
• Homework 6
Homework 5

1. Generate Poisson distributions using rejection sampling for $\lambda=1,10,100$

2. Compute means for 10 consecutive random deviates and demonstrate the Central Limit Theorem
Probability

Reading

T. Loredo (1990) monograph, chapters 1-3
*From Laplace to Supernova 1987A: Bayesian Inference in Astrophysics*
Gelman et al., *Bayesian Data Analysis*, chapter 1
Probability Definitions

- Frequency of occurrence of an event
- Degree of belief in a proposition
Axioms of Probability Theory

• \( p(A \text{ or not } A) = p(A) + p(\text{not } A) = 1 \)

• \( p(A \text{ and } B) = p(B) \ p(A \text{ given } B) \equiv p(A) \ p(B \text{ given } A) \)
Axioms of Probability Theory

- $p(A + \overline{A}) = p(A) + p(\overline{A}) = 1$
- $p(A \cap B) = p(B) \, p(A \mid B) \equiv p(A) \, p(B \mid A)$
(Alt) Sum Rule

\[ p(A + B) = 1 - p(\overline{A} \overline{B}) = 1 - p(\overline{A}) \cdot p(\overline{B} | \overline{A}) \]

\[ = 1 - p(\overline{A}) \left( 1 - p(B | \overline{A}) \right) = 1 - p(\overline{A}) + p(\overline{A}) \cdot p(B | \overline{A}) \]

\[ = p(A) + p(\overline{A}B) = p(A) + p(B) \cdot p(\overline{A} | B) \]

\[ = p(A) + p(B) \left( 1 - p(A | B) \right) = p(A) + p(B) - p(B) \cdot p(A | B) \]

\[ = p(A) + p(B) - p(AB) \]
Bayes’ Theorem

\[ p(AB|C) = p(A|BC) \ p(B|C) = p(B|AC) \ p(A|C) \]

\[ p(A|BC) = p(B|AC) \ p(A|C) / p(B|C) \]

\[ p(\theta|D \ I) = p(D|\theta \ I) \ p(\theta|I) / p(D|I) \]
prior, likelihood, posterior

\[ p(\theta|D \ I) = p(D|\theta \ I) \ p(\theta|I) / p(D|I) \]

*a priori* distribution: \( p(\theta|I) \)

likelihood: \( p(D|\theta \ I) \)

*a posteriori* distribution: \( p(\theta|D \ I) \)
prior

• Unfairly maligned as “subjective”, but actually a mechanism to explicitly encode your assumptions

• When your data are weak, your prior beliefs don’t change; when your data are strong, your prior beliefs don’t matter.

• You update your prior belief with new data, using Bayes’ Theorem. Lets you daisy-chain analyses.

• When your prior is informative, takes more data to make a large change.

• Technically, the biggest difference between likelihood analysis and Bayesian analysis: converts $p(D|\theta I)$ to $p(\theta|D I)$
marginalization

• multi-dimensional parameters \( \Theta = \{ \theta_i \} \)

• joint posterior: \( p(\Theta) d\Theta \)

• integrate over “nuisance” parameters

\[
p(\theta_1) d\theta_1 = \int_{\theta_2} p(\theta_1, \theta_2) \, d\theta_2 \, d\theta_1
\]
uncertainty

• $p(\Theta|D)$ describes the uncertainty on $\Theta$

• Usually reported as 68% or 90% central intervals (always say what they are!)

• No guarantee of good coverage properties (because of priors), unlike frequentist confidence intervals (“the true value is contained 90% of the time for CIs calculated in this manner when the experiment is repeated”)
## frequentist vs Bayesian

<table>
<thead>
<tr>
<th>Bayesian</th>
<th>Frequentist</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data are fixed and parameters have uncertainties</td>
<td>The true parameter is fixed, and data are realizations from different experiments</td>
</tr>
<tr>
<td>Uncertainty is range in parameter values that encompass a certain probability</td>
<td>Confidence interval overlaps the true value a certain fraction of the time</td>
</tr>
<tr>
<td>Prior assumptions are explicit, but can be arbitrary</td>
<td>You have to be aware of assumptions and limits of applicability</td>
</tr>
<tr>
<td>Allows daisy-chaining of analyses</td>
<td>Must be careful about stopping rules</td>
</tr>
<tr>
<td>No guarantee of good coverage</td>
<td>Confidence interval defined to produce good coverage</td>
</tr>
</tbody>
</table>
frequentist vs Bayesian

• use whichever technique as appropriate

• Bayesian: non-repeatable experiments (many in Astronomy), make assumptions explicit via prior, daisy-chain analyses, hierarchical modeling, etc.

• Frequentist: repeatable experiments, large datasets, vast toolkit for highly specific problems, non-parametric tests, hypothesis tests, least-squares fitting, goodness-of-fit, etc.
MCMC Basics