Lecture 6: Fundamentals of Observational Cosmology (II)

Gong-Bo Zhao

Abstract

This lecture introduces the statistical language used throughout observational cosmology. We review Bayes' theorem, priors and posteriors, Gaussian distributions, covariance matrices, the central limit theorem, χ^2 fitting, confidence regions, and marginalization. The goal is not to turn this course into a statistics course, but to give students a working toolkit for reading modern cosmology papers.

Learning goals

After this lecture, students should be able to:

- state Bayes' theorem and identify posterior, likelihood, prior, and evidence;
- compute sample mean, variance, covariance, and correlation coefficient;
- explain why Gaussian errors appear so often in astronomy;
- write a χ^2 function for both uncorrelated and correlated data;
- distinguish best-fit values, confidence regions, priors, and marginalized constraints.

1 Bayes' theorem

Bayes' theorem states

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}.$$

An easy proof is

$$P(B \mid A)P(A) = P(B \cap A) = P(A \cap B) = P(A \mid B)P(B).$$

In cosmology we usually write

$$P(\boldsymbol{\theta} \mid D) = \frac{\mathcal{L}(D \mid \boldsymbol{\theta})\, \pi(\boldsymbol{\theta})}{Z},$$

where

- $P(\boldsymbol{\theta} \mid D)$ is the **posterior**;
- $\mathcal{L}(D \mid \boldsymbol{\theta})$ is the **likelihood**;
- $\pi(\boldsymbol{\theta})$ is the **prior**;
- $Z = P(D)$ is the **evidence**.

For parameter estimation, the posterior is usually the main object of interest. For model comparison, the evidence also matters.

A simple example

Suppose a company makes phones in three factories:

$$P(A) = 0.35, \qquad P(B) = 0.35, \qquad P(C) = 0.30.$$

The defect probabilities are

$$P(D \mid A) = 0.015, \qquad P(D \mid B) = 0.010, \qquad P(D \mid C) = 0.020.$$

Then

$$P(D) = \sum_i P(D \mid i)P(i) = 0.015 \times 0.35 + 0.010 \times 0.35 + 0.020 \times 0.30 = 0.01475.$$

Therefore,

$$P(A \mid D) = \frac{0.015 \times 0.35}{0.01475} \approx 0.356,$$

$$P(B \mid D) = \frac{0.010 \times 0.35}{0.01475} \approx 0.237, \qquad P(C \mid D) = \frac{0.020 \times 0.30}{0.01475} \approx 0.407.$$

This example is simple, but it already contains the same logic used in Bayesian cosmological inference.

2 Mean, variance, covariance, and correlation

For a dataset $\{X_i\}_{i=1}^N$, the sample mean is

$$\bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i.$$

The sample variance is

$$s_X^2 = \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \bar{X})^2.$$

For two datasets $\{X_i\}$ and $\{Y_i\}$, the sample covariance is

$$\mathrm{Cov}(X, Y) = \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \bar{X})(Y_i - \bar{Y}),$$

and the correlation coefficient is

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{s_X s_Y}.$$

Interpretation.

- Variance measures the spread of one variable.
- Covariance measures whether two variables fluctuate together.
- Correlation is the dimensionless version of covariance, with values between -1 and 1.

3 The central limit theorem

The central limit theorem says that sums of many independent random variables tend toward a Gaussian distribution under broad conditions. This is one of the reasons Gaussian statistics appear so often in astronomy: detector noise, calibration uncertainties, and averages over many contributions often become approximately Gaussian even when the microscopic processes are not.

4 Gaussian distributions

A one-dimensional Gaussian with mean μ and variance σ^2 is

$$p(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right].$$

For a data vector $\mathbf{x}$ with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}$, the multivariate Gaussian is

$$p(\mathbf{x}) = \frac{\exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{C}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]}{\sqrt{(2\pi)^N |\mathbf{C}|}}.$$

This formula appears constantly in cosmology because many likelihoods are approximated as Gaussian in either the data space or the parameter space.

5 From Gaussian likelihood to χ^2

If the covariance is parameter-independent, maximizing the Gaussian likelihood is equivalent to minimizing

$$\chi^2 = (\mathbf{d} - \mathbf{m})^T \mathbf{C}^{-1}(\mathbf{d} - \mathbf{m}),$$

where $\mathbf{d}$ is the data vector and $\mathbf{m}(\boldsymbol{\theta})$ is the model prediction.

Uncorrelated case

If the data points are independent, then $\mathbf{C}$ is diagonal and

$$\chi^2 = \sum_{i=1}^{N} \frac{[d_i - m_i(\boldsymbol{\theta})]^2}{\sigma_i^2}.$$

Correlated case

If the off-diagonal elements of $\mathbf{C}$ are nonzero, we must use the full matrix expression. Ignoring correlations can make constraints look artificially tight.

Reduced χ^2. A common diagnostic is

$$\chi^2_\nu \equiv \frac{\chi^2}{N_{\text{data}} - N_{\text{par}}},$$

where N_{data} is the number of data points and N_{par} is the number of fitted parameters. Values near unity often indicate a reasonable fit, but only if the error model is realistic.

6 Confidence regions

For a one-parameter Gaussian problem:

- 68.3% confidence corresponds to $\Delta\chi^2 = 1$;
- 95.4% confidence corresponds to $\Delta\chi^2 = 4$;
- 99.7% confidence corresponds to $\Delta\chi^2 = 9$.

These are the familiar 1σ, 2σ, and 3σ intervals. In integral form,

$$\frac{\int_0^1 e^{-x^2/2}\,\mathrm{d}x}{\int_0^\infty e^{-x^2/2}\,\mathrm{d}x} = 68.3\%, \qquad \frac{\int_0^2 e^{-x^2/2}\,\mathrm{d}x}{\int_0^\infty e^{-x^2/2}\,\mathrm{d}x} = 95.5\%, \qquad \frac{\int_0^3 e^{-x^2/2}\,\mathrm{d}x}{\int_0^\infty e^{-x^2/2}\,\mathrm{d}x} = 99.7\%.$$

For *two* parameters, the corresponding values are

$$\Delta\chi^2 = 2.30, \qquad 6.18, \qquad 11.83.$$

This is why a "1σ contour" in two dimensions is not obtained by simply taking independent $\pm 1\sigma$ intervals on each parameter.

7 General χ^2 distribution cases

For k degrees of freedom, the χ^2 distribution is

$$f(x;k) = \frac{x^{k/2-1} e^{-x/2}}{2^{k/2}\Gamma(k/2)}, \qquad x > 0.$$

Its cumulative distribution function is

$$F(x;k) = \frac{\gamma\left(\frac{k}{2}, \frac{x}{2}\right)}{\Gamma(k/2)},$$

where $\gamma(s,t)$ is the lower incomplete gamma function.

Useful special cases are

$$F(x;1) = \operatorname{erf}\left(\sqrt{\frac{x}{2}}\right), \qquad F(x;2) = 1 - e^{-x/2}.$$

Therefore, for two variables one finds

$$\Delta\chi^2 = 2.30, \qquad 6.18, \qquad 11.83$$

for confidence levels of approximately 68%, 95%, and 99%, respectively.

8 Priors and marginalization

A prior expresses information or assumptions we have before analyzing the current dataset. Priors can be:

- flat over a finite range;
- Gaussian around an external measurement;
- physically motivated, such as requiring $\Omega_\mathrm{m} > 0$.

Marginalization means integrating over nuisance parameters:

$$P(\theta_1 \mid D) = \int P(\theta_1, \theta_2, \ldots \mid D)\,\mathrm{d}\theta_2 \mathrm{d}\theta_3 \cdots .$$

This is essential in cosmology because not every parameter in the likelihood is of direct scientific interest.

9 Common statistical misunderstandings

- A best-fit point is not the same as a robust constraint.
- A narrow posterior can still be biased if the model is wrong.
- "Gaussian errors" are an approximation, not a law of nature.
- Priors matter most when the data are weak or parameters are strongly degenerate.

10 Summary

The key statistical ideas used in cosmology are surprisingly compact:

- Bayes' theorem tells us how data and priors combine.
- Covariance matrices describe both uncertainty and correlation.
- Gaussian likelihoods lead naturally to χ^2 fitting.
- Confidence regions depend on how many parameters are being varied.
- Marginalization is how we remove nuisance parameters consistently.

These concepts will be used repeatedly in later lectures on supernovae, BAO, RSD, and the CMB.

Suggested reading

- Dodelson and Schmidt, chapters on statistical inference.
- Trotta, *Bayes in the sky*, for a readable cosmology-focused review.
- Any introductory statistics text for the central limit theorem and Gaussian distributions.

References from the 2025 version

References

[1] A. G. Riess *et al.* [Supernova Search Team], "Observational evidence from supernovae for an accelerating universe and a cosmological constant," *Astron. J.* **116**, 1009 (1998), doi:10.1086/300499, [astro-ph/9805201].

[2] S. Perlmutter *et al.* [Supernova Cosmology Project Collaboration], "Measurements of Omega and Lambda from 42 high redshift supernovae," *Astrophys. J.* **517**, 565 (1999), doi:10.1086/307221, [astro-ph/9812133].

[3] W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, "Numerical Recipes in FORTRAN: The Art of Scientific Computing."

Homework

1. **Bayes practice.** Reproduce the phone-factory calculation in the lecture notes. Then change the defect rate of factory C from 0.020 to 0.012 and recompute $P(A \mid D)$, $P(B \mid D)$, and $P(C \mid D)$.

2. **Sample statistics.** For the data vectors

$$X = (2, 4, 5, 7, 8), \qquad Y = (1, 3, 4, 6, 8),$$

compute the sample mean, variance, covariance, and correlation coefficient.

3. **From likelihood to χ^2.** Starting from the one-dimensional Gaussian likelihood, show that maximizing the likelihood is equivalent to minimizing

$$\chi^2 = \sum_i \frac{(d_i - m_i)^2}{\sigma_i^2}$$

when the data points are independent.

4. **Correlated data.** Consider two measurements with covariance matrix

$$\mathbf{C} = \begin{pmatrix} 4 & 1.2 \\ 1.2 & 1 \end{pmatrix}.$$

Compute the correlation coefficient and write the full matrix expression for χ^2.

5. **Confidence regions.** Explain in a few sentences why the 68% contour in two parameters corresponds to $\Delta\chi^2 = 2.30$ rather than $\Delta\chi^2 = 1$.

6. **Optional coding task.** Draw 10^5 realizations of the sum of 20 uniformly distributed random numbers and show numerically that the histogram is close to a Gaussian.