Bayesian model selection in cosmology

Martin Kilbinger
Institut d’Astrophysique de Paris (IAP)

Shanghai Normal University
3 August 2009
Model selection

• **Traditional parameter estimation:**
  Q: For a specific model with \( n \) parameters which is the most likely (best-fit) parameter and confidence interval given the data?

• **Model selection:**
  Q: Which of two or more models with parameters \( n_1, n_2, \ldots \) is the most likely to fit the data?

• **Examples in cosmology:**
  ★ Cosmological constant \( \Lambda \) vs. dark energy vs. modified gravity
  ★ Flat vs. curved
  ★ Primordial fluctuations: scale-free (\( n_s=1 \)) vs. \( n_s=\text{const} \) vs. running \( n_s(k) \)

• **Other applications:** Cluster profile reconstruction, exo-planets, nuisance parameters, ...
Bayesian evidence

• Bayes’ theorem

\[
p(\theta|d, m) = \frac{\mathcal{L}(d|\theta, m) \pi(\theta|m)}{E(d|m)}
\]

\[m : \text{model}\]
\[d : \text{data}\]
\[\theta : \text{model parameter}\]

• Posterior is normalised

\[E(d|m) = \int d^n \theta \mathcal{L}(d|\theta, m) \pi(\theta|m)\]

• Bayes factor

\[B_{12} = \frac{E(d|m_1)}{E(d|m_2)}\]
Posterior odds

How can we interpret the evidence?

• Bayes’ theorem again:

\[ p(m|d) \propto E(d|m)\pi(m) \]

• Ratio for two models \( m_1 \) and \( m_2 \):

\[
\frac{p(m_1|d)}{p(m_2|d)} = B_{12} \frac{\pi(m_1)}{\pi(m_2)}
\]

Posterior odds

<table>
<thead>
<tr>
<th>\log_{10} B_{12}</th>
<th>Odds</th>
<th>Strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 ... 0.5</td>
<td>1 ... 3</td>
<td>weak</td>
</tr>
<tr>
<td>0.5 ... 1</td>
<td>3 ... 10</td>
<td>substantial</td>
</tr>
<tr>
<td>1 ... 2</td>
<td>10 ... 100</td>
<td>strong</td>
</tr>
<tr>
<td>&gt;2</td>
<td>&gt;100</td>
<td>decisive</td>
</tr>
</tbody>
</table>

Jeffreys’ scale
Approximations to the Evidence

- **BIC (Bayesian Information Criterion)** [Schwarz 1987]
  \[
  \text{BIC} = -2 \ln \mathcal{L}_{\text{max}} + k \ln N_{\text{data}}
  \]
  - likelihood @ maximum
  - penalty for large parameter space (Occam’s razor)

(Similar: AIC, DIC)
Problem: Penalty independent whether parameters constrained by data or not

- **Laplace approximation**: likelihood Gaussian, priors large and uniform
  [Lazarides, Ruiz de Austri & Trotta 2004, Heavens, Kitching & Verde 2007]
  \[
  E \approx (2\pi)^{n/2} |F|^{-1/2} (\Delta \theta_1 \ldots \Delta \theta_n)^{-1} \mathcal{L}_{\text{max}}
  \]
  - volume allowed by data
  - initial volume (prior)
  - Occam’s razor
CMB+SNla+BAO:
Dark energy, curvature

- Base model: flat ΛCDM, $n_{\text{par}}=3$ \((\Omega_m, \Omega_b, h)\)
  Data: pure geometrical probes (WMAP5 distance priors, SNla, BAO)

- Three models of dark energy:
  \[
  w = -1 \quad \text{ΛCDM} \\
  w = w_0 \quad w\text{CDM} \\
  w = w_0 + w_1(1 - a) \quad w(z)\text{CDM}
  \]

- Flat ($\Omega_K=0$) vs. curved ($\Omega_K\neq0$)
Priors

- Exclude phantom energy \((w \geq -1)\), require accelerated expansion for \(z < z_{\text{acc}} = 0.5 (w < -1/3)\)

- Upper boundary for curvature: empty Universe, \(\Omega_k < 1\), lower boundary \(\Omega_k > -1\)
Bayes’ factor: dark energy, curvature
Prior to the analysis of the raw text, there was an incorrect mention of the variables and concepts. Let's correct and refine the text:

**Priomordial Universe**

- **Density perturbation power spectrum**
  
  \[ P_\delta \propto k^{n_s+1/2 \alpha_s \ln(k/k_0)} \]

  \( n_s \): spectral index (\( n_s = 1 \): scale-free, Harrison-Zel’dovich spectrum)

  \( \alpha_s \): ‘running’ of spectral index

- **Tensor perturbation (gravitational waves) power spectrum**
  
  \[ P_t \propto k^{n_t} \]

- **Standard model** has \( n_s = \text{const}, \alpha_s = n_t = 0, r=\text{tensor/scalar}=0 \)
Priors

- Slow-roll approximation of inflation

\[ \epsilon = \frac{m_{P1}^2}{4\pi} \left( \frac{H'}{H} \right)^2 ; \]

\[ \ell \lambda_H = \left( \frac{m_{P1}^2}{4\pi} \right)^\ell \frac{(H')^{\ell-1}}{H^\ell} \frac{d^{\ell+1}H}{d\phi^{\ell+1}} ; \ell \leq 1, \]

depend on potential \( V \) of scalar field \( \Phi \) causing inflation.

- Slow-roll condition: \( \epsilon \ll 1, \quad |\ell \lambda_H| \ll 1 \)

- Choose: \( \epsilon \leq 0.1, \quad -0.1 < \frac{1}{\lambda_H} = \eta < 0.1, \quad \ell \lambda_H = 0 \) for \( \ell > 1 \)
Priors

• Relations between slow-roll and power-spectrum parameters

\[ n_s = 1 + 2\eta - 4\epsilon - 2(1 + C)\epsilon^2 - \frac{1}{2}(3 - 5C)\epsilon\eta; \]

\[ r = 16\epsilon [1 + 2C(\epsilon - \eta)]; \]

\[ \alpha_s = \frac{\epsilon}{1 - \epsilon} (8\epsilon + 10\eta); \]

\[ n_t = -2\epsilon - (3 + C)\epsilon^2 + (1 + C)\epsilon\eta. \]

\[ C = 4(\ln 2 + \gamma) - 5 \approx -0.7296. \]
Bayes’ factor (CMB+SNIa+BAO)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r$</td>
<td>0</td>
<td>1.83</td>
</tr>
<tr>
<td>$(\ln r)$</td>
<td>-80</td>
<td>0.604</td>
</tr>
</tbody>
</table>
Constraints (CMB+SNIa+BAO)