Bayesian fitting of multivariate mixture models via Gibbs sampling

Georg Schnabel
My research domain

Models
- TALYS
- INCL

Experiments

Evaluation method
- Linearized method
- Monte Carlo method

Simulation
- MCNP, GEANT4, ...
- Perturbation Theory
- Total Monte Carlo

Evaluation method output:
- ENDF + Cov
- ENDF #1
- El. Prod. #1
- El. Prod. #2
- ...
Multivariate mixture models

\[ \rho(\vec{x}) = \sum_{k=1}^{K} \pi_k \rho_k(\vec{x}) \]

Goldfish (~12cm)

Blacktip reef shark (~1.6m)
Areas of application

Speech to text translation

Hello, this is a test!

Face recognition

Hans

Image denoising

Unsupervised learning: clustering (alternative to k-means)

Super resolution
Example: Detector events

\[ \rho(\mathbf{x}) = \pi_1 U(\mathbf{x}) + \pi_2 N(\mathbf{x} | \mu, \Sigma) \]
Bayesian statistics

Hypotheses
H1: It rained yesterday
H2: The sun shined yesterday

Observations
O1: The ground is wet now
O2: The ground is dry now

Which hypothesis is true or at least more plausible?
Desiderata

(I) Degrees of Plausibility are represented by real numbers.

(II) Qualitative Correspondence with common sense.

(IIIa) If a conclusion can be reasoned out in more than one way, then every possible way must lead to the same result.

(IIIb) The robot always takes into account all of the evidence it has relevant to a question. It does not arbitrarily ignore some of the information, basing its conclusions only on what remains. In other words, the robot is completely non-ideological.

(IIIc) The robot always represents equivalent states of knowledge by equivalent plausibility assignments. That is, if in two problems the robot’s state of knowledge is the same (except perhaps for the labeling of the propositions), then it must assign the same plausibilities in both.

Cited from “Probability Theory: The Logic of Science” by E. T. Jaynes, 1995
Bayesian update equation

According to Cox (1946 & 1961): These desiderata uniquely determine a rule set for inference, which happens to be the Kolmogorov axioms of probability!

\[ \rho(H \mid O) = \frac{1}{\rho(O)} \rho(O \mid H) \rho(H) \]

- **Posterior** $\rho(H \mid O)$ is our updated knowledge taking into account the observation $O$, e.g., if $O = O_1$,
  - $\rho(H_1 \mid O_1) = 1$
  - $\rho(H_2 \mid O_1) = 0$

- **Likelihood** $\rho(O \mid H)$ is the plausibility to observe $O$ given $H$ is true, e.g.,
  - $\rho(O_1 \mid H_1) = 1$

- **Prior** $\rho(H)$ is a probability distribution reflecting prior belief, e.g.,
  - $\rho(H_1) = 0.3$ and $\rho(H_2) = 0.7$

- $H$ is an hypothesis, e.g.,
  - $H \in \{H_1, H_2\}$
    - $H_1$: It rained
    - $H_2$: It didn’t

- $O$ is a specific observation, e.g.,
  - $O \in \{O_1, O_2\}$
    - $O_1$: Ground wet
    - $O_2$: Ground dry
Back to our mixture model

\[ \rho(\vec{x}) = \pi_1 N(\vec{x} | \vec{\mu}_1, \Sigma_1) + \pi_2 N(\vec{x} | \vec{\mu}_2, \Sigma_2) + \pi_3 \text{Exp}(\vec{x} | \vec{a}) \]

Observations \( O = \{ \vec{x}_1, \vec{x}_2, \ldots, \vec{x}_N \} \)

\[ \rho(H | O) = \frac{1}{\rho(O)} \rho(O | H) \rho(H) \]

\[ \rho(\vec{x} | \pi_1, \pi_2, \pi_3, \vec{\mu}_1, \vec{\mu}_2, \Sigma_1, \Sigma_2, \vec{a}) \]

Choice of prior

- Dirichlet
- Uniform
- Jeffreys
- Unif.
Non-informative / “Objective” priors

Jeffreys prior: \[ \rho(\Sigma) \propto (\det(\Sigma))^{-(1+d)/2} \]

- Invariance of posterior under reparametrization
- Indifference about orientation of distribution
- Indifference about scale (self-similarity)
Taming the distribution beast

\[ \rho(\tilde{x}') = \pi_1 N(\tilde{x} | \bar{\mu}_1, \Sigma_1) + \pi_2 N(\tilde{x} | \bar{\mu}_2, \Sigma_2) + \pi_3 \text{Exp}(\tilde{x} | \bar{a}) \]

\[ \rho(\pi_1, \pi_2, \pi_3, \bar{\mu}_1, \bar{\mu}_2, \Sigma_1, \Sigma_2, \bar{a} | \tilde{x}_1, \ldots \tilde{x}_N) \propto \prod_{i=1}^{N} \rho(x_i | \pi_1, \pi_2, \pi_3, \bar{\mu}_1, \bar{\mu}_2, \Sigma_1, \Sigma_2, \bar{a}) \times \]

\[ \rho(\pi_1, \pi_2, \pi_3) \rho(\bar{\mu}_1) \rho(\bar{\mu}_2) \rho(\Sigma_1) \rho(\Sigma_2) \rho(\bar{a}) \]

**Optimization**: Find maximum of posterior (MAP estimate)

**Sampling**: Draw a sample from the posterior distribution
The expectation, standard deviation, covariances, correlations, etc. are defined as integrals over a probability distribution. If we are able to sample from the distribution, we can approximately solve the respective integrals by empirical averages:

$$
\int f(z) \rho(z \mid \vec{x}_1, \ldots, \vec{x}_N) \approx \frac{1}{N} \sum_{i=1}^{N} f(z_i)
$$
Gibbs sampling

\[ \rho(\vec{\mu}_1, \Sigma_1, \vec{\mu}_2, \Sigma_2, z_1, \ldots, z_N \mid \vec{x}_1, \ldots, \vec{x}_N) \]

Start with random assignment of \( \mu_1, \Sigma_1, \mu_2, \Sigma_2 \)

\[ \rho(z_1, \ldots, z_N \mid \vec{\mu}_1, \Sigma_1, \vec{\mu}_2, \Sigma_2, \vec{x}_1, \ldots \vec{x}_N) \]

\[ \rho(\vec{\mu}_1, \Sigma_1, \vec{\mu}_2, \Sigma_2 \mid z_1, \ldots, z_N, \vec{x}_1, \ldots \vec{x}_N) \]
In action!

$$\rho(\mathbf{x}) = \pi_1 N(\mathbf{x} | \mu_1, \Sigma_1) + \pi_2 N(\mathbf{x} | \mu_2, \Sigma_2) + \pi_3 \text{Exp}(\mathbf{x} | \tilde{\alpha})$$

**Green ellipses**
True component specification

**Blue ellipses**
Posterior expectation of component specifications

Animated sampling from posterior distribution: Link
Some traceplots

\[ \rho(\vec{x}) = \pi_1 N(\vec{x} \mid \vec{\mu}_1, \Sigma_1) + \pi_2 N(\vec{x} \mid \vec{\mu}_2, \Sigma_2) + \pi_3 \text{Exp}(\vec{x} \mid \vec{\alpha}) \]
Comparison Estimate vs Truth

<table>
<thead>
<tr>
<th>props</th>
<th>true</th>
<th>est</th>
<th>sd</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \pi_1 )</td>
<td>0.048</td>
<td>0.040</td>
<td>0.008</td>
</tr>
<tr>
<td>( \pi_2 )</td>
<td>0.368</td>
<td>0.381</td>
<td>0.016</td>
</tr>
<tr>
<td>( \pi_3 )</td>
<td>0.582</td>
<td>0.579</td>
<td>0.015</td>
</tr>
</tbody>
</table>

True \( \Sigma_1 \):

- \([10, 0]\)
- \([0, 40]\)

Est \( \Sigma_1 \):

- \([9.9, 4.95]\)
- \([4.95, 28.3]\)

Sd \( \Sigma_1 \):

- \([6.0, 5.64]\)
- \([5.64, 15.25]\)

\[
\rho(\vec{x}) = \pi_1 N(\vec{x} | \vec{\mu}_1, \Sigma_1) + \pi_2 N(\vec{x} | \vec{\mu}_2, \Sigma_2) + \pi_3 \text{Exp}(\vec{x} | \vec{\alpha})
\]
Summary

- Multivariate mixture models are useful in many domains
- Bayesian statistics is a framework for learning under uncertainty
- Sampling based approaches enable the extraction of various kinds of information from the posterior distribution (=quantified updated knowledge), e.g. credible intervals, expectations, standard deviations, ...

Code package available at:
https://github.com/gschnabel/mvMixDist

Interactive demonstration online:
https://nddemos.shinyapps.io/bayfitmixdist/