

Bayesian Learning

Lecture 12 - Model evaluation

Bertil Wegmann

Department of Computer and Information Science
Linköping University



- Model evaluation - Posterior predictive analysis

Models - why?

- We now know how to **compare** models.
- But how do we know if any given model is 'any good'?
- George Box: '**All models are false, but some are useful**'.

What is your model for?

■ Prediction.

- ▶ Interpretation not a concern
- ▶ Black-box approach may be ok.
- ▶ Model averaging may be a good idea.

■ Abstraction to aid in thinking about a phenomena.

- ▶ Prediction accuracy of less concern.
- ▶ Model averaging may be a bad idea.

■ Model as a compact description of a complex phenomena.

- ▶ Computational cost of model evaluation may be a concern.
- ▶ Online/real-time analysis.

Posterior predictive analysis

- If $p(y|\theta)$ is a 'good' model, then the data actually observed should not differ 'too much' from simulated data from $p(y|\theta)$.
- Bayesian: simulate data from the **posterior predictive distribution**:

$$p(y^{rep}|y) = \int p(y^{rep}|\theta)p(\theta|y)d\theta.$$

- Difficult to compare y and y^{rep} because of dimensionality.
- Solution: compare **low-dimensional statistic** $T(y, \theta)$ to $T(y^{rep}, \theta)$.
- Evaluates the full probability model consisting of both the likelihood *and* prior distribution.

Posterior predictive analysis

- **Algorithm** for simulating from the posterior predictive density $p[T(y^{rep})|y]$:
 - 1 Draw a $\theta^{(1)}$ from the posterior $p(\theta|y)$.
 - 2 Simulate a data-replicate $y^{(1)}$ from $p(y^{rep}|\theta^{(1)})$.
 - 3 Compute $T(y^{(1)})$.
 - 4 Repeat steps 1-3 a large number of times to obtain a sample from $T(y^{rep})$.
- We may now compare the observed statistic $T(y)$ with the distribution of $T(y^{rep})$.
- **Posterior predictive p-value:** $\Pr[T(y^{rep}) \geq T(y)]$
- Informal graphical analysis.

Posterior predictive analysis - Examples

- Ex. 1. Model: $y_1, \dots, y_n | \mu, \sigma^2 \stackrel{iid}{\sim} N(\mu, \sigma^2)$. $T(y) = \max_i |y_i|$.
- Ex. 2. ARIMA-process. $T(y)$ may be the autocorrelation function.
- Ex. 3. Poisson regression. $T(y)$ frequency distribution of the response counts. Proportions of zero counts.

Posterior predictive analysis - Normal model, max statistic

