Bayesian Model Goodness of Fit

DIC, AIC, BIC,
PPL, MSPE
Residuals
Predictive residuals

Overall Measures of GOF

- **Deviance**: this measures the overall likelihood of the model given a parameter vector

$$D(\boldsymbol{\theta}) = -2\,log\,L(\boldsymbol{\theta})$$

This can be averaged over all L iterations of a sampler to give

$$\bar{D} = \frac{1}{L}\sum_{l=1}^{L} D(\boldsymbol{\theta}^{l}) \text{ where the}$$

$\boldsymbol{\theta}^{l}$ is the l th iteration parameters

Deviance Information Criterion

- The deviance is a measure of GOF but is not adjusted for parameterisation.
- The deviance information criterion (DIC) is adjusted for number of parameters
- The formula is given as

$DIC = \bar{D} + pD$ where pD is

the effective number of parameters, estimated as

$pD = \bar{D} - D(\hat{\boldsymbol{\theta}}) = Dbar - Dhat$ and

$\hat{\boldsymbol{\theta}}$ is the posterior mean parameter vector estimates

DIC notes

- Note that while Dbar can be computed in the sampler
- Dhat can only be computed at the end
- pD can be negative and this is a **major** problem
- DIC can also be negative but this is not a problem
- DIC is only a relative measure : lower values better
- DIC difference of at least 2 - 3 are need for a better model (i.e. model 1: DIC= 124.0 ; model 2: DIC= 120.0 means that model 2 is preferred)
- pD is a secondary criterion of GOF (parsimony)

DIC, AIC BIC

- DIC is related to AIC and is equal to AIC when models with only fixed effects are fitted
- AIC : Akaike Information Criterion

$$AIC = -2\,log\,L(\boldsymbol{\theta}) + 2p = D(\boldsymbol{\theta}) + 2p$$

where $\boldsymbol{\theta}$ is an estimated vector (eg ML)

and p is the number of parameters

Bayesian Information Criterion

- BIC:
  - Like AIC except the penalty is a function of the sample size

$$BIC = -2 \log L(\boldsymbol{\theta}) + p \ln n = D(\boldsymbol{\theta}) + p \ln n$$

where $\boldsymbol{\theta}$ is an estimated vector (eg ML) and p is the number of parameters

Random Effects and AIC/BIC

- AIC and BIC assumes that you can estimate the number of parameters in the model
- This is not clear in the case of random effects
- A single random intercept involves a variance parameter for the (usually) Gaussian prior distribution BUT there are N different effects estimated.
- Should the number of parameters contributed be ℕ or ℕ+𝟙, or just 𝟙 ??
- The DIC tries to estimate this within the sampler
- Hence AIC and BIC are difficult to compute as P is unclear

DIC estimation

- DIC is usually estimated from McMC samplers.
- It requires convergence !!
- Can have Dhat > Dbar and so pD can be negative .
- This implies over-dispersion in the sampler.
- Alternative estimation of pD
- Can use a crude (approximate) estimator which is always positive :

$$\widehat{pD} = \widehat{var}(D(\boldsymbol{\theta})) / 2$$

Posterior Predictive Loss

- **Posterior Predictive loss** (PPL)

This is the average loss measured between the observed data and the data predicted from the posterior predictive distribution :

$$r_i^p = y_i - y_i^{Pred}$$

where y_i^{Pred} is a value of from the predictive distribution

MSPE: mean square predictive error is just this averaged.

Predictive Distribution

a new value of y is y^* :

$$f(y^*|\mathbf{y}) = \int_{\boldsymbol{\theta}} f(y^*|\mathbf{y},\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}$$

where $p(\boldsymbol{\theta}|\mathbf{y})$ is the posterior distribution

and $f(y^*|\mathbf{y},\boldsymbol{\theta})$ is the likelihood of y^*

This is just th expected value of the likelihood over the posterior.

To get an approximate value

we can generate samples from $p(\boldsymbol{\theta}|\mathbf{y})$

and then use these in $f(y^*|\mathbf{y},\boldsymbol{\theta})$ and average.

This approximates $\int_{\boldsymbol{\theta}} f(y^*|\mathbf{y},\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}$.

and average them to get predictive y values

WinBUGS code

- {
- y[i]~dpois(mu[i])
- ypred[i]~dpois(mu[i])

will generate predictive distribution values averaged over the sampler

- R[i]<-y[i]-ypred[i] # predictive residual
- Sqr[i]<-pow(R[i],2)
- }
- mspe<-mean(Sqr[])

Residuals

- Can get item-wise diagnostics also
  - Bayesian residuals
  - Predictive residuals
  - Standardization also

$$r_i = y_i - \mu_i$$

$$r_i^p = y_i - y_i^*$$