Welcome Page

Latest News

Contact us/BUGS list

WinBUGS

New WinBUGS examples

FAQs

DIC

GeoBUGS

PKBUGS

Running from other software

BUGS resources online

WinBUGS development site

OpenBUGS site

DIC: Deviance Information Criterion

DIC (Deviance Information Criterion) is a Bayesian method for model comparison that WinBUGS can calculate for many models.

Full details of DIC can be found in Spiegelhalter DJ, Best NG, Carlin BP and Van der Linde A, "Bayesian Measures of Model Complexity and Fit (with Discussion)", Journal of the Royal Statistical Society, Series B, 2002 64(4):583-616.

Some slides on DIC from a presentation in February 2006 can be downloaded.

DIC has its own Wikipedia page.

Frequently Asked Questions about DIC

1. What is the 'deviance' that goes into DIC?
2. What is pD?
3. What is DIC?
4. What is the connection between DIC and AIC?
5. What is the connection between DIC and BIC?
6. When are AIC, DIC and BIC appropriate?
7. Can DIC be negative?
8. Can pD be negative?
9. How do I compare different DICs?
10. How does DIC depend on the parameterisation used?
11. How does the pD in WInBUGS 1.4 compare to the pV in Andrew Gelman's bugs.R function and R2WInBUGS?
12. Can DIC be used to compare alternative prior

Frequently Asked Questions about DIC

1. **What is the 'deviance' that goes into DIC?**

   This is exactly the same as if the node `deviance` had been monitored when running WinBUGS. This deviance is defined as - 2 * log(likelihood), where 'likelihood' is defined as p(y | &theta) including all the normalising constants: y comprises all stochastic nodes given values (i.e. data), and &theta comprises the immediate stochastic parents of y.

   'Stochastic parents' are the stochastic nodes upon which the distribution of y depends, when collapsing over all logical relationships. For example, if `y ~ dnorm(mu,tau)`, and if `tau` is a function of a parameter `phi` over which a prior distribution has been placed, then the likelihood is defined as a function of `phi`.

   Return to the top

   ---

2. **What is pD?**

   Using the notation in the WinBUGS output for the DIC tool,

   `Dbar` is the posterior mean of the deviance,

   `Dhat` is a point estimate of the deviance obtained by substituting in the posterior means &theta .bar: thus Dhat = - 2 * log(p(y | &theta .bar)).

   `pD` is 'the effective number of parameters', and is given by

$$pD = Dbar - Dhat.$$

Thus pD is the posterior mean of the deviance minus the deviance of the posterior means. In normal hierarchical models, pD = tr(H) where H is the 'hat' matrix that maps the observed data to their fitted values.

Return to the top

---

3. **What is DIC?**

   DIC is the 'Deviance Information Criterion', and is given by

   $$DIC = Dbar + pD = Dhat + 2\ pD.$$

   The model with the smallest DIC is estimated to be the model that would best predict a replicate dataset which has the same structure as that currently observed.

   Return to the top

---

4. **What is the connection between DIC and AIC?**

   DIC is intended as a generalisation of Akaike's Information Criterion (AIC). For non-hierarchical models with little prior information, pD should be approximately the true number of parameters. AIC requires counting parameters and hence any intermediate level ('random-effects') parameters need to be integrated out. A recent 'conditional AIC' by Vaida and Blanchard (2005) focuses on the random effects in normal hierarchical models and uses tr(H) as the effective number of parameters, and so again matches DIC.

Return to the top

---

5. **What is the connection between DIC and BIC?**

   DIC differs from Bayes factors and BIC in both form and aims. BIC attempts to ientify the 'true' model, DIC is not based on any assumption of a 'true' model and is concerend with short-term predictive ability. BIC requires specification of the number of parameters, while DIC estimates the effective number of parameters. BIC provides a procedure for model averaging, DIC does not.

   Two 'cross-overs' have been suggested. First, using pD as the number of parameters in a BIC for hierarchical models. Second, using exp(-DIC/2) as model weights for model averaging, just as 'Akaike weights' have been used. Neither procedure appears to have a theoretical justification, although using DIC weights for model averaging may well have good empirical predictive properties.

   Return to the top

---

6. **When are AIC, DIC and BIC appropriate?**

   In hierarchical models, these three techniques are essentially answering different prediction problems.

   Suppose the three levels of our model concerned *classes* within *schools* within a *country*. Then

   1. if we were interested in predicting results of future *classes* in those actual schools, then DIC is appropriate (ie the random effects themselves are of interest);
   2. if we were interested in predicting results of future *schools* in that country, then marginal-likelihood methods such as AIC are appropriate (ie the population parameters are of interest);
   3. if we were interested in predicting results for

a new *country*, then BIC/ Bayes factors are appropriate (ie the 'true' underlying model is of interest).

This suggests that BIC/Bayes factors may in many circumstances be inappropriate measures by which to compare models. See these slides for more details.

Return to the top

---

7. **Can DIC be negative?**

   Yes; the deviance can be negative, since a probability density can be greater than 1 (if it has a range of less than 1, say). See these slides for an example.

Return to the top

---

8. **Can pD be negative?**

   For sampling distributions that are log-concave in their stochastic parents, pD is guaranteed to be positive (provided the simulation has converged). However, in other circumstances it is theoretically possible to get negative values. We have obtained negative pD's in the following situations:

   1. with non-log-concave likelihoods (e.g. Student-t distributions) when there is substantial conflict between prior and data;
   2. when the posterior distribution for a parameter is extremely asymmetric, or symmetric and bimodal, or in other situations where the posterior mean is a very poor summary statistic and gives a very large deviance.

   See these slides for an example.

Return to the top

---

9. **How do I compare different DICs?**

   The minimum DIC estimates the model that will make the best short-term predictions, in the same spirit as Akaike's criterion.

   It is difficult to say what would constitute an *important* difference in DIC. Very roughly, differences of more than 10 might definitely rule out the model with the higher DIC, differences between 5 and 10 are substantial, but if the difference in DIC is, say, less than 5, and the models make very different inferences, then it could be misleading just to report the model with the lowest DIC.

Return to the top

---

10. **How does DIC depend on the parameterisation used?**

    Different values of Dhat (and hence pD and DIC) can be obtained depending on the parameterisation used for the prior distribution. For example, consider the precision tau (1 / variance) of a normal distribution. The two priors

```
tau ~ dgamma(0.001, 0.001)
```

    and

```
  log.tau ~ dunif(-10, 10);
log(tau) <- log.tau
```

    are essentially identical but will give slightly different results for Dhat: for the first prior the stochastic parent is tau and hence the posterior mean of tau is substituted in Dhat, while in the second parameterisation the stochastic parent is log.tau and hence the posterior mean of log(tau) is substituted in Dhat.

Unfortunately this can produce a conflict in interest: from the DIC perspective it would be best to express a prior on a parameterisation that is more likely to have approximate posterior normality, and hence a transformation to the real line (as described above for log.tau) may be appropriate. However this may not be the most intuitive parameterisation on which to express an informative prior. See these slides for a full worked example.

Return to the top

11. **How does the pD in WInBUGS 1.4 compare to the pV in Andrew Gelman's bugs.R function and R2WInBUGS?**

Suppose we have a non-hierarchical model with weak prior and K parameters. Then the posterior variance of the deviance is approximately approx 2K.

Thus with negligible prior information, half the variance of the deviance is an estimate of the number of free parameters in the model. This estimate generally turns out to be remarkably robust and accurate.

This might suggest using pV = (variance of the deviance)/2 as an estimate of the effective number of parameters in a model. This was originally tried in a working paper by Spiegelhalter et al (1997), and has since been suggested by Gelman et al (2004). It is currently used in R2WinBUGS, largely as the full DIC values cannot be extracted when remotely running WinBUGS 1.4.1.

pV is invariant to parameterisation, robust and trivial to calculate.

Working through distribution theory for simple Normal random-effects model with K groups suggests pV is approximately pD(2 - pD/K), although many assumptions have been made. Hence we may expect pV to be larger than pD when there is moderate shrinkage.

The slides contain examples where pV does seem rather high.

Return to the top

12. **Can DIC be used to compare alternative prior distributions?**

    Yes, provided these are genuine prior distributions expressed independently of the data. A smaller DIC can always be 'fiddled' by selecting a prior that matches the observed data.

    Usually an informative prior distribution will give a smaller DIC than a 'vague' prior, although this may not occur if there is conflict between the informative prior and the data.

    Return to the top

13. **Why is DIC greyed out?**

    DIC is currently greyed out in WinBUGS when one of the stochastic parents is a discrete node. The formal basis for DIC relies on approximate posterior normality for the parameter estimates and requires a plug-in estimate of each stochastic parent - for discrete nodes it is not clear which estimate to use.

    Return to the top

14. **How can I calculate DIC for mixture distributions?**

    Celeux et al (2006) explore a wide range of options for constructing an appropriate DIC for missing data problems, including mixture models. The paper will appear in Bayesian Analysis with discussion.

    Currently any of these options have to be implemented individually in WinBUGS.

    Return to the top

15. **What improvements to DIC could be made?**

It would be better if WinBUGS used the posterior mean of appropriate functions (ie transformations to the real line) of the `direct parameters' (eg those that appear in the WinBUGS distribution syntax) to give a 'plug-in' deviance, rather than the posterior means of the stochastic parents.

For example, for a Normal likelihood, we could specify that the posterior mean of log(tau) is always used, regardless of the function of tau on which a prior distribution has actually been placed.

Unfortunately this turns out not to be so straightforward to implement.

Users are free to calculate this themselves: they can dump out posterior means of appropriate functions of `direct' parameters in the likelihood, then calculate the plug-in deviance outside WinBUGS or by reading posterior means in as data and checking the value for the deviance node in the Node/Info menu.

Return to the top

16. **Are there any bugs in the DIC code?**

    There is a bug which means that if you close the DIC tool wondow, you cannot open it again during that run!

    Return to the top

17. **If Dbar measures 'lack of fit', why can it increase when I add a covariate?**

    Suppose Yi is assumed to be N(0,1) under model 1, and consider a covariate xi with mean 0 and which is uncorrelated with Y. Then it is straightforward to show that fitting a more complex model 2: Yi ~ N(b xi,1) leads to Dbar increasing by 1. The crucial idea is that Dbar should perhaps not really be considered a measure of fit (in spite of the title of Spiegelhalter et al (2002)!). Fit might better be measured by Dhat. As emphasised by van der Linde (2005) (also available from here), Dbar is more a measure of model 'adequacy', and already incorporates a degree of penalty for complexity.

18. **How and why does WinBUGS partition DIC and p_D?**

WinBUGS separately reports the contribution to Dbar, p_D and DIC for each separate node or array, together with a Total. This enables the individual contributions from different parts of the model to be assessed.

In some circumstances some of these contributions may need to be ignored and removed from the Total. For example, in the following model:

```
for(i in 1:N) {
Y[i] ~ dnorm(mu, tau)
}
tau <- 1/pow(sigma, 2)
sigma ~ dnorm(0, 1) I(0, )
mu ~ dunif(-100, 100)
```

where Y is observed data, then the DIC tool will give DIC partitioned into Y, sigma and the Total. Clearly in this case, there should be no contribution from sigma, but because of the lower bound specified using the I(,) notation in the prior, WinBUGS treats sigma as if it were an observed but censored stochastic node when deciding what things to report in the DIC table.

In another situation, we might genuinely have censored data, e.g.

```
for(i in 1:N) {
Y[i] ~ dnorm(mu, tau)I(Y.cens[i], )
}
```

where Y is unknown but Y.cens is the observed lower bound on Y.

WinBUGS has no way of knowing that in the first case, sigma should be ignored in the DIC whereas in the second case Y should be included in DIC. (This is as much a problem of how the BUGS language rather confusingly represents censoring, truncation and bounds using the same notation as it is to do with how DIC displayed, but hopefully it illustrates the ambiguity and why the only option we could think of was to report each DIC contribution separately and let the user pick out the relevant bits).

Return to the top

For queries about DIC, please mail bugs@mrc-bsu.cam.ac.uk
