

Deviance information criterion

From Wikipedia, the free encyclopedia

The **deviance information criterion (DIC)** is a hierarchical modeling generalization of the AIC (Akaike information criterion) and BIC (Bayesian information criterion, also known as the Schwarz criterion). It is particularly useful in Bayesian model selection problems where the posterior distributions of the models have been obtained by Markov chain Monte Carlo (MCMC) simulation. Like AIC and BIC it is an asymptotic approximation as the sample size becomes large. It is only valid when the posterior distribution is approximately multivariate normal.

Define the deviance as $D(\theta) = -2 \log(p(y|\theta)) + C$, where y are the data, θ are the unknown parameters of the model and $p(y|\theta)$ is the likelihood function. C is a constant that cancels out in all calculations that compare different models, and which therefore does not need to be known.

The expectation $\bar{D} = \mathbf{E}^\theta[D(\theta)]$ is a measure of how well the model fits the data; the larger this is, the worse the fit.

There are two calculations in common usage for the effective number of parameters of the model. The first, as described in Spiegelhalter et al. (2002, p. 587) is $p_D = \bar{D} - D(\bar{\theta})$, where $\bar{\theta}$ is the expectation of θ .

The second, as described in Gelman et al. (2004, p. 182) is $p_D = p_V = \frac{1}{2} \widehat{\text{var}}(D(\theta))$. The larger the effective number of parameters is, the *easier* it is for the model to fit the data, and so the deviance needs to be penalized.

The deviance information criterion is calculated as

$$DIC = p_D + \bar{D},$$

or equivalently as

$$DIC = D(\bar{\theta}) + 2p_D.$$

From this later form, the connection with Akaike's information criterion is evident.

The idea is that models with smaller DIC should be preferred to models with larger DIC. Models are penalized both by the value of \bar{D} , which favors a good fit, but also (in common with AIC and BIC) by the effective number of parameters p_D . Since \bar{D} will decrease as the number of parameters in a model increases, the p_D term compensates for this effect by favoring models with a smaller number of parameters.

The advantage of DIC over other criteria in the case of Bayesian model selection is that the DIC is easily calculated from the samples generated by a Markov chain Monte Carlo simulation. AIC and BIC require calculating the likelihood at its maximum over θ , which is not readily available from the MCMC simulation. But to calculate DIC, simply compute \bar{D} as the average of $D(\theta)$ over the samples of θ , and $D(\bar{\theta})$ as the value of D evaluated at the average of the samples of θ . Then the DIC follows directly from these approximations. Claeskens and Hjort (2008, Ch. 3.5) show that the DIC is large-sample equivalent to the natural model-robust version of the AIC.

In the derivation of DIC, it is assumed that the specified parametric family of probability distributions that generate future observations encompasses the true model. This assumption does not always hold, and it is desirable to consider model assessment procedures in that scenario. Also, the observed data are used both to construct the posterior distribution and to evaluate the estimated models. Therefore, DIC tends to select over-fitted models. Recently, these issues are resolved by Ando (2007), Bayesian predictive information criterion, BPIC. Ando (2010, Ch. 8) provided a discussion of various Bayesian model selection criteria. To avoid the over-fitting problems of DIC, Ando (2011) developed Bayesian model selection criteria from a predictive view point. The criterion is calculated as

$$IC = \bar{D} + 2p_D = -2\mathbf{E}^\theta[\log(p(y|\theta))] + 2p_D.$$

The first term is a measure of how well the model fits the data, while the second term is a penalty on the model complexity. Note, that the p in this expression is the predictive distribution rather than the likelihood above.

See also

- Akaike information criterion (AIC)
- Bayesian information criterion (BIC)
- Bayesian predictive information criterion (BPIC)
- Focused information criterion (FIC)
- Hannan–Quinn information criterion
- Kullback–Leibler divergence
- Jensen–Shannon divergence

References

- Ando, Tomohiro (2007). "Bayesian predictive information criterion for the evaluation of hierarchical Bayesian and empirical Bayes models". *Biometrika* **94** (2): 443–458. doi:10.1093/biomet/asm017 (https://dx.doi.org/10.1093%2Fbiomet%2Fasm017).
- Ando, T. (2010). *Bayesian Model Selection and Statistical Modeling*, CRC Press. Chapter 7.
- Ando, Tomohiro (2011). "Predictive Bayesian Model Selection". *American Journal of Mathematical and Management Sciences* **31** (1-2): 13–38. doi:10.1080/01966324.2011.10737798 (https://dx.doi.org/10.1080%2F01966324.2011.10737798).
- Claeskens, G, and Hjort, N.L. (2008). *Model Selection and Model Averaging*, Cambridge. Section 3.5.
- Gelman, Andrew; Carlin, John B.; Stern, Hal S.; Rubin, Donald B. (2004). *Bayesian Data Analysis: Second Edition*. Texts in Statistical Science. CRC Press. ISBN 1-58488-388-X. LCCN 2003051474 (http://lccn.loc.gov/2003051474). MR 2027492 (https://www.ams.org/mathscinet-getitem?mr=2027492).
- van der Linde, A. (2005). "DIC in variable selection", *Statistica Neerlandica*, 59: 45-56. doi:10.1111/j.1467-9574.2005.00278.x (http://dx.doi.org/10.1111/j.1467-9574.2005.00278.x)
- Spiegelhalter, David J.; Best, Nicola G.; Carlin, Bradley P.; van der Linde, Angelika (2002). "Bayesian measures of model complexity and fit (with discussion)". *Journal of the Royal Statistical Society, Series B* **64** (4): 583–639. doi:10.1111/1467-9868.00353 (https://dx.doi.org/10.1111%2F1467-9868.00353). JSTOR 3088806 (https://www.jstor.org/stable/3088806). MR 1979380 (https://www.ams.org/mathscinet-getitem?mr=1979380).
- Spiegelhalter, David J.; Best, Nicola G.; Carlin, Bradley P.; van der Linde, Angelika (2014). "The deviance information criterion: 12 years on (with discussion)". *Journal of the Royal Statistical Society, Series B* **76** (3): 485–493.

Retrieved from "https://en.wikipedia.org/w/index.php?"

title=Deviance_information_criterion&oldid=633766905"

Categories: Bayesian statistics | Regression variable selection | Model selection

- This page was last modified on 14 November 2014, at 04:42.
- Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.