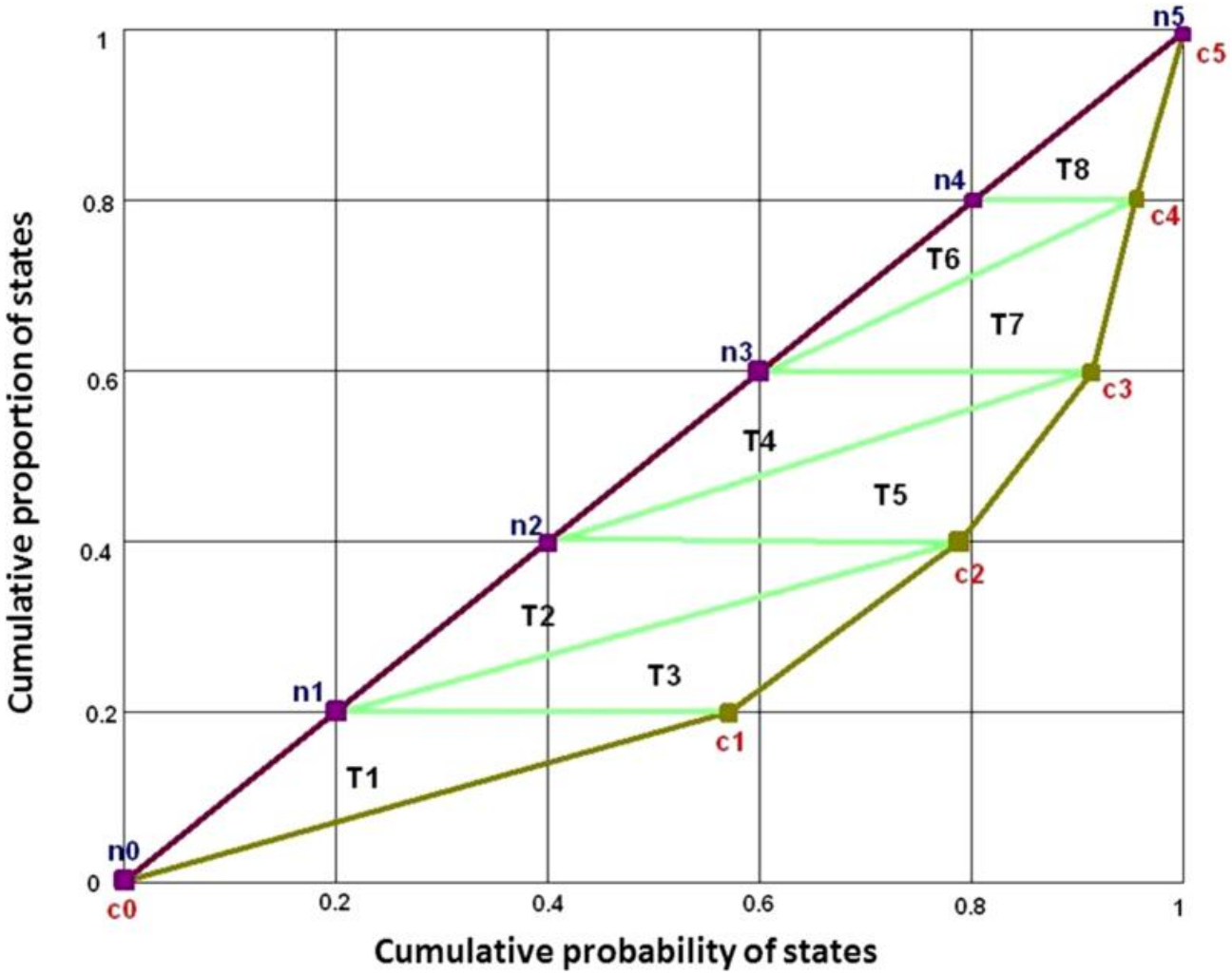
Supplementary material for:

Marcot, B. G. 2012. Metrics for evaluating performance and uncertainty of Bayesian network models. Ecological Modelling 230:50-62.

Appendix. Exact and approximate calculations of the Gini coefficient with discrete BN model outcome states.

An example plot of a cumulative frequency distribution "Lorenz curve" is shown in Appendix Figure A, derived from a posterior probability distribution (PPD) of N = 5 outcome states {*c1* ... *c5*} rank-ordered in decreasing probability values [0.57, 0.21, 0.12, 0.05, 0.05] (see text for further explanation). Note that the point *c0* is added to connect the curve to the origin. A corresponding line of total uncertainty (perfect evenness) $\overline{n_0 n_5}$ is likewise plotted as the positive diagonal.



Appendix Figure A. Example Lorenz curve *c* for a posterior probability distribution of 5 states.

Because BN models typically entail discrete-state variables, the area subtended between the Lorenz curve *c* and the line of total uncertainty *n* can be exactly calculated by summing the

area of the individual scalene triangles as shown. In general, with *N* number of states, the number of such triangles formed by a discrete Lorenz curve is (2*N* – 2).

The area of a scalene triangle with side lengths a, b, c, is $\sqrt{s(s-a)(s-b)(s-c)}$, where $s = (a+b+c)/2$. Because each axis of the Lorenz curve derived from a PPD of a BN model spans [0,1], applying the standard distance formula results in the length of $\overline{n_0 n_5} = \sqrt{2}$, so that the line segments $\overline{n_0 n_1} = \overline{n_1 n_2} = \cdots = \overline{n_4 n_5} = \frac{\sqrt{2}}{N}$, or for *N* = 5 in this example, each segment length ≈ 0.2828. Lengths of the horizontal segments $\overline{n_i c_i}$ are each calculated as the simple difference $c_i - n_i$. Lengths of the hypotenuses and additional sides of each triangle are calculated from the distance formula, for example, for any two consecutive points *a*,*b* in $\{c_0, c_1, c_2, \cdots c_N\}$, the length of each segment $\overline{c_a c_b}$ on the Lorenz curve is

$$\sqrt{(c_b - c_a)^2 + \frac{1}{N^2}} \ .$$

Then, the sum of all areas of the scalene triangles is the exact value of the Gini coefficient, which, when doubled, represents the degree of certainty [0,1] in the overall outcome PPD. In this example, the Gini coefficient = 0.48, which suggests only a very moderate level of certainty (values of 0 and 1 represent complete uncertainty and complete certainty, respectively). This value could then be compared to values generated under different input scenarios or from different models with the same or different number of states *N*.

For continuous variables, the Lorenz curve becomes a continuous function and the Gini coefficient can be approximated as the area subtended by a standard Beta distribution (i.e., scaled to range [0,1]),

$$f(x) = \frac{x^{p-1}(1-x)^{q-1}}{B(p,q)} \ ,$$

where $0 \leq x \leq 1$; $p, q > 0$; $B(p,q)$ is the beta function

$$B(p,q) = \int_0^1 t^{p-1}(1-t)^{q-1} dt \ \ ;$$

and *p*,*q* are shape parameters. Further, the standard deviation of a standard Beta distribution is

$$SD = \sqrt{\frac{pq}{(p+q)^2(p+q+1)}} \ .$$

The area under the Beta distribution (as an estimator of the Gini coefficient), as well as its shape parameters *p*,*q* and SD values, all can be used as indices of uncertainty (the spread of posterior probabilities) by comparing their values among alternative models or scenarios.