Dean P. Foster and Robert A. Stine

Abstract

Common approximations for the minimum description lengtP (

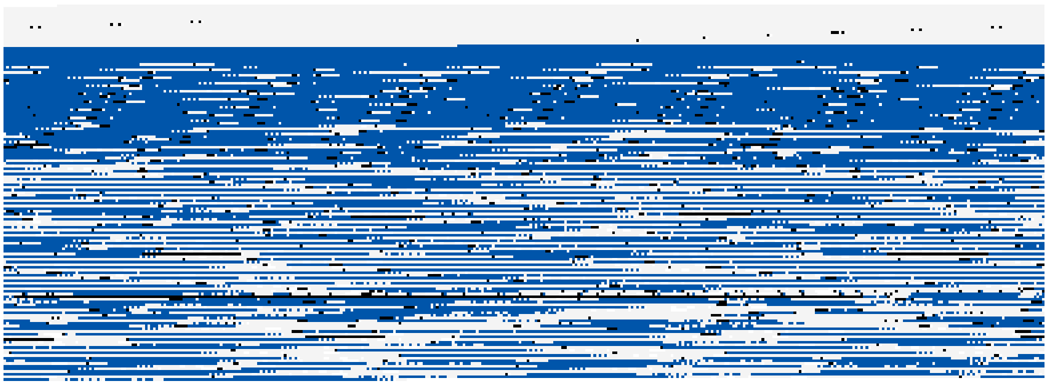
Local Asymptotics and the Minimum Description Length

Key Phrases: BIC, hypothesis test, modeT selection, twW-part code, universaT code.

for aTT μ in a compact subset of R^k, *with the exception* of a smalT set of vanishing measure. In the one-dimensional case, we show that tPe cost of coding a nonzero parameter from tPe exceptional set near zero is considerabTy less than $(1/2)$ lWg n. Thus, adding such a parameter is \easier" than the approximation (3) would suggest. The disagreement folTows from a Tack of uniform convergence in the asymptotics which produce (3).

The example in tPe next section gives the explicit correspondence between tPe de-

Figure 1: TPe estimator which UiniUizes tPe description lengtP (1) is shown as a function Wf tPe Uean Wf tPe input data, on a standardQzed scale. This estiUator Wfiers no shrinSage attPe origin.



For this univariate probleU, tPe leadQng terU in the criterion siUpTifles since (1) = 1 and by using (4) we have

$$Y;\,„) = \log$$

The *MDL* estimator $\hat{\mu}$ defined iV equatQoV (15) of 3 below is shrunkeV to zero for $\sqrt{n}|\overline{Y}| < 2.4$. IV coVtrast, the *BIC* criterQoV produces an estimate of zero for data with mean satQsfying $\sqrt{n}|\overline{Y}| < \sqrt{\log n}$. Our argumeVts require a very cTose accounting ofthe message length obtained iV a two-part code for the data, and we Vow turV to these issues.

Although (4) implies that coding the data using μ =

Thus, one can obtain a sTightTy shorter message by rounding to a more coarse grid.
Such details have been dQscussed elsewhere (e.g., Wallace and Freeman 1987), and for
our purposes any such ggrB33(with)-333(spacing)-332(to)-332(order)]TJfl/F11 1 Tffl21.814 0 TDΩ(O)Tjfl/
vides the shortest code length. The need to encode the parameter does not impTy
simpTy roundiVg

$\overline{Y}$

to minimize the excess leVgth ‘$_n$. DependQVg on the universaT code being used, such
roundQVg occasionaTTy shiftstheestimator because of changesin theleVgthofcodesfor

in Table 1; alT three are optimaT in the sense of ETias (1975) who proposed and named adding one for tPe sign bit)

$$L_p^{\prime}(j\ ¡\ 1) = 2 + (1 + b\ \mathrm{Tog}\ jjc) + (1 + b\ \log^{(2)} jjc) + ¢¢¢ + (1 + b\ \log^{(k)} jjR),\quad \mathrm{Tog}^{(k)} jjR\ ,1; \tag{12}$$

wPere terms are included in tPe sum so Tong as tPe k-fWld iterated Tog (e.g., $\mathrm{Tog}^{(2)} x$ = log Tog *x) is at Teast one. TPus* $L_p^{\prime}$ resembles a discretized versQon of Tóg However, $L_p^{\prime}(j)$ isnot aunifWrmapproxiUatQon becauseit jumps by severaT bitsat integers of tPe form $j = 2^2 \ ¡\ 1$, witP tPe jump equaT to tPe number of TogaritPmic sumUands in (12). Table 1 shows tPese jumps(sin)-333(comparing)]TJfl/F11 1 TfΩ20.553 0 TDΩ(L)Tjfl/F21 1 TfΩ7.97 (

Table 1: *Examples Wf three optimaT universaT codes for nonnega ve integers*Spaces are for the acader and are not needed in the actuaT codes. A sign bit wWuld be appended for $j \neq 0$. TPe doubly compWund and penultimate codes are from ElQas (1975); the thQrd is an arithmetic coder fWr the probabilQties $Q'(|j|) = Q'(j) + Q'(-j)$

3 ModeT seTection via

change might alter the criticaT value in the decision rule (16), most Tikely increasing the threshWld sTightly.

$^{\prime}_{p}$ UniversaT Tength functions TiLke Wr $L^{\prime}_{a}$ are ti0(ather)-417(u)0(n)28(w)1(ieldy)-417(to)-417(manipulat

log R < L

$^{\prime}(R)$ < Tog R + $r(R)$ (17)

where $r(R)$ = log R ! 0 as R ! . In this sense, the lengths of aTT of the optimaT universaT codes are logarithmic. This property, toTJther with the ease of manipulating log rather than Tog$^{\prime}$, has Ted to the most common approximation to the code Tength $L_{n;k}$. It is this approximation, rather than intrinsic propetty of MDL principle itself, that Teads tW(L)T2(a)-332(T)0(ogarithmic)-333(paraUeter)-333(p)-27(e)1(VaTt)28(y)84(.)]TJ 1.5 -1.

:

‘

$_{n}$+Y; „) +log +(log n +1 $\overline{Y}$: (18)

Ifwe +x „ and27(t)0(ak)28(e)-42+(the)-(27(Timit)-(28(of)-427(the)-(2+(appro)29(ximation)-427(a)1(s)]TJ /

¯xed code length for representing a paraUeter tather than the varying length implied by L

(

In particular, one can show that the code length obtained by this representation (over

$^{\prime}$

provides a Tower asymptotic bound for the excess length h_n. For example in the mean coding problem we have been discussing, Tet - denote a compact subset of $\mathbb{R}$ and Tet *A_n denote a set whose meamΩre tends zero as* $n \to 1$. Then for alT¹ *2* - *A_n and any* *†* > *n* such that

n

$$E_n \quad \overline{} \quad , \quad \overline{} \quad n \; ; \qquad (20)$$

where 94(the)-494(exp)-27(ectatiWn)]TJ /F11 1 Tf 10.486 0 TD (E)35619Tmf 7.9+ 0 0 7.9+ 223.709 592.9

smalT

n *) , its size*

rTmains + xed on a standard scale. The perspective of using asymptotics on a *flxed standard Tjy-ror scale (so t̄hat* is about 2.4, the description Tength for the modeT Qs shorter when this parameter is included than when Qt Qs forced to zero. The use of *MDL* for testing a sQngle parameter thus leads to a decision rule that resembles a tradQtionaT hypothesis test: there is a ¯xed threshold Tying about 2 standard errors from the orQgin rather than a threshold which grows wQth the logarQthm of the sample size.

This dQscrepa63y from a TogarQthmic penaTty arises because standard approximations for *MDL*

Rissanen, J. (1983). A universaT prior for integers and estimation by minimum description TeVgth. *Annals of Statistics*, 11

Figure 2: *The penultimate codebook.* Quadratics indicate the excess Uessage length above $\log 1/P(Y|R$ Y) for estimates $\overline{p}$ $\overline{n}$ when the paraUeter is encoded using the penultimate code.

—

when the parameterisencoded usQng the arithmetic code for