Estimating the Bayes decision error using generalized entropies

Martin Horáček and Igor Vajda

Institute of Computer Science AS CR, Center of Biomedical Informatics, Prague, Czech Republic

Introduction

This paper deals with the tightness of relationship between the average minimal posterior probability of error (the posterior Bayes error)

$$e_B(X|Z) = 1 - E_Z \max_{x\in\mathcal{X}} P(X=x|Z=z) = 1 - E_Z \max_{x\in\mathcal{X}} p(x|z)$$

given random variables X and Z, where X is discrete with values in space $\mathcal{X}$, $\mathrm{card}(\mathcal{X}) = r < \infty$ and Z is absolutely continuous with respect to Lebesgue measure, and the family of conditional f-entropies

$$H_f(X|Z) = E_Z\left[\sum_{x\in\mathcal{X}} p^2(x|z) f(1/p(x|z)) + f(0)\left(1 - \sum_{x\in\mathcal{X}} p^2(x|z)\right)\right],$$

where f is a convex function on the interval $[0,\infty)$, strictly convex in $f(1)$ satisfying $f(1)=0$, $f(0)<\infty$ and the function $[f(x)-f(0)]/x$ is concave on the interval $[1,\infty)$. This family was introduced by Zvárová [1].
We study the tightness of this relationship by the means of average inaccuracy, with which one can estimate the posterior Bayes error when only the value of the conditional f-entropy is known. We deal with the hypothesis of Morales and Vajda [2], who stated, on the basis of numerical computations, that the quadratic entropy has the most tight relation to the posterior Bayes error among α-power entropies in the sense of the average inaccuracy. We approach the hypothesis of Morales and Vajda analytically and we show that when a close approximation of average inaccuracy is used, the quadratic entropy has the most tight relation to the posterior Bayes error in the family of f-entropies.

Methods

Morales and Vajda measured tightness between entropies and the posterior Bayes error using average difference between the maximal and minimal possible value of posterior Bayes error $e(X|Z)$ when $r = \mathrm{card}(\mathcal{X})$ and a f-entropy are given (Morales and Vajda considered only α-power entropies with $f(x) = (\alpha-1)^{-1}x(1-x^{(1-\alpha)}),\ \alpha>0$), i.e. by the means of average inaccuracy

$$AI_r(e_B|H_f) = \frac{1}{a_{r,f}}\int_0^{a_{r,f}}\left[e^+_{B,f}(H) - e^-_{B,f}(H)\right]\mathrm{d}H,$$

where

$$e^+_{B,f}(H) = \max_{X,Z:H_f(X|Z)=H} e_B(X|Z),\quad e^-_{B,f}(H) = \min_{X,Z:H_f(X|Z)=H} e_B(X|Z)$$

and

$$a_{r,f} = \max_{X,Z} H_f(X|Z) = H_f(1/r,\dots,1/r) = f(r)/r + f(0)(1-1/r).$$

It could be shown that $e^\pm_{B,f}$ are both continuous strictly increasing functions, $e^+_{B,f}(0) = e^-_{B,f}(0)$ and $e^+_{B,f}(a_{r,f}) = e^-_{B,f}(a_{r,f})$, and therefore $AI_r(e_B|H_f)$ can be also computed by the means of inverse functions

$$H^+_f(e_B) = \max_{X,Z:e_B(X|Z)=e_B} H_f(X|Z),\quad H^-_f(e_B) = \min_{X,Z:e_B(X|Z)=e_B} H_f(X|Z),$$

as

$$AI_r(e_B|H_f) = \frac{1}{a_{r,f}}\int_0^{\frac{r-1}{r}}\left[H^+_f(e_B) - H^-_f(e_B)\right]\mathrm{d}e_B.$$

Since any f-entropy is a Schur-concave function, the bounds $H^\pm_f(e_B)$ satisfy

$$H^+_f(e_B) = (1-e_B)^2 f\left(\frac{1}{1-e_B}\right) + \frac{e_B^2}{r-1} f\left(\frac{r-1}{e_B}\right) + e_B f(0)\left(2 - e_B\frac{r}{r-1}\right)$$

when $e\in[0,(r-1)/r]$ and the lower bound is a piecewise linear function with values

$$H^-_f(e_B) = (1-e_B) f\left(\frac{1}{1-e_B}\right) + f(0)e_B$$

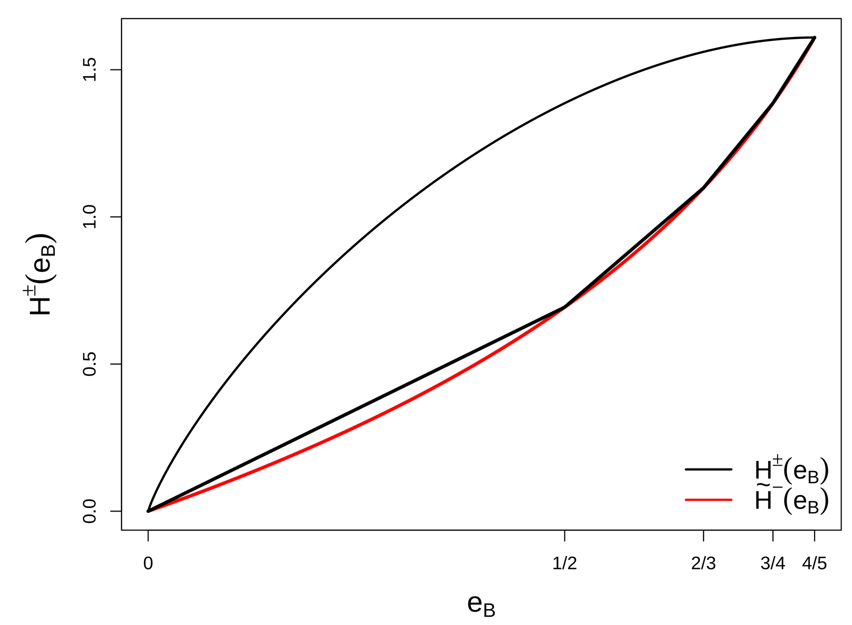
when $e_B = \frac{k}{k+1},\ k\in\{0,\dots,r-1\}$ (see e.g. Horáček [3] and Vajda and Vašek [4]). We suggest to approximate the piecewise linear function $H^-_f(e_B)$ by a function defined as

$$\tilde{H}^-_f(e_B) = (1-e_B) f\left(\frac{1}{1-e_B}\right) + f(0)e_B$$

on the whole interval $[0,(r-1)/r]$. This alternative lower bound $\tilde{H}^-_f(e_B)$ is again a continuous strictly increasing function, the inverse function $\tilde{e}^+_{B,f}(H)$ is well defined and we can approximate the average inaccuracy by

$$\widetilde{AI}_r(e_B|H_f) = \frac{1}{a_{r,f}}\int_0^{\frac{r-1}{r}}\left[H^+_f(e_B) - \tilde{H}^-_f(e_B)\right]\mathrm{d}e_B.$$

An example of bounds $H^+_f(e_B)$, $H^-_f(e_B)$ and $\tilde{H}^-_f(e_B)$ for Shannon's entropy $H_f(X) = -\sum_{x\in\mathcal{X}} p(x)\ln p(x)$, $r=5$, is shown in the following figure.



Results

We show that under the described approximation, the quadratic entropy $H_2(X) = 1 - \sum_{x\in\mathcal{X}} p^2(x)$ has the tightest relation to the posterior Bayes error, i.e.

$$\widetilde{AI}_r(e_B|H_2) \le \widetilde{AI}_r(e_B|H_f)$$

for any f-entropy H_f and any given integer $r\ge 2$.
Let us first consider only those functions f that satisfy $f(0)=0$. Then it follows that $f(x)>0$ when $x>1$ and

$$\begin{aligned}\widetilde{AI}_r(e_B|H_f) &= \frac{1}{a_{r,f}}\int_0^{\frac{r-1}{r}}\left[H^+_f(e_B) - \tilde{H}^-_f(e_B)\right]\mathrm{d}e_B\\ &= \frac{r}{f(r)}\int_0^{\frac{r-1}{r}}(1-e_B)^2 f\left(\frac{1}{1-e_B}\right) + \frac{e_B^2}{r-1} f\left(\frac{r-1}{e_B}\right) - (1-e_B) f\left(\frac{1}{1-e_B}\right)\mathrm{d}e_B\\ &= \frac{r}{f(r)}\int_0^{\frac{r-1}{r}} e_B\left[f\left(\frac{r-1}{e_B}\right)\frac{e_B}{r-1} - f\left(\frac{1}{1-e_B}\right)(1-e_B)\right]\mathrm{d}e_B.\end{aligned}$$

After transformation $z = \frac{r}{r-1}e_B$ we have

$$\widetilde{AI}_r(e_B|H_f) = \frac{r}{f(r)}\int_0^1\left(\frac{r-1}{r}\right)^2 z\left[f\left(\frac{r}{z}\right)\frac{z}{r} - f\left(\frac{r}{r-zr+z}\right)\frac{r-zr+z}{r}\right]\mathrm{d}z. \tag{1}$$

Since f is convex, $z\in(0,1]$ and $r\ge 1$, we get following inequality

$$\begin{aligned}\frac{f(r/z)-f(1)}{r/z-1} &\ge \frac{f(r)-f(1)}{r-1},\\ \frac{f(r/z)}{r/z-1} &\ge \frac{f(r)}{r-1},\\ \frac{f(r/z)z}{f(r)} &\ge \frac{r-z}{r-1}.\end{aligned} \tag{2}$$

Similarly the formula

$$\frac{f[r/(z-zr+r)]-f(1)}{r/(z-zr+r)-1} \le \frac{f(r)-f(1)}{r-1}$$

can be converted into

$$\frac{f[r/(z-zr+r)](z-rz+r)}{f(r)} \le z. \tag{3}$$

Inserting inequalities (2) and (3) into the equality (1) we get

$$\begin{aligned}\widetilde{AI}_r(e_B|H_f) &\ge \int_0^1\left(\frac{r-1}{r}\right)^2 z\left(\frac{r-z}{r-1} - z\right)\mathrm{d}z\\ &= \int_0^1\left(\frac{r-1}{r}\right) z(1-z)\mathrm{d}z = \frac{1}{6}\frac{r-1}{r} = \widetilde{AI}_r(e_B|H_2).\end{aligned}$$

Now let us consider also the situation when $f(0)=0$. Then $\widetilde{AI}_r(e_B|H_f)$ satisfies

$$\begin{aligned}\widetilde{AI}_r(e_B|H_f) = &\frac{1}{a_{r,f}}\int_0^1\left(\frac{r-1}{r}\right)^2 z\left[f\left(\frac{r}{z}\right)\frac{z}{r} - f\left(\frac{r}{r-zr+z}\right)\frac{r-zr+z}{r}\right]\mathrm{d}z+\\ &+\frac{1}{a_{r,f}}f(0)\frac{r-1}{r}\widetilde{AI}_r(e_B|H_2).\end{aligned}$$

We can use the inequalities (2) and (3) again (multiplied by $f(r)/r$ to avoid potential division by zero or negative number) to receive

$$\begin{aligned}\widetilde{AI}_r(e_B|H_f) &\ge \frac{1}{a_{r,f}}\left[\left(\frac{f(r)}{r} + f(0)\frac{r-1}{r}\right)\widetilde{AI}_r(e_B|H_2)\right]\\ &= \frac{r}{f(r)+(r-1)f(0)}\left[\frac{f(r)+f(0)(r-1)}{r}\widetilde{AI}_r(e_B|H_2)\right] = \widetilde{AI}_r(e_B|H_2).\end{aligned}$$

References

[1] J. Zvárová. On Measures of Statistical Dependence. *Časopis pro pěstování matematiky*, 99:15-29, 1974.
[2] D. Morales and I. Vajda. *Generalized Information Criteria for Optimal Bayes Decisions.* Research report 2274, ÚTIA AV ČR, v.v.i, 2010.
[3] M. Horáček. *Míry biodiverzity a jejich aplikace. (Measures of biodiversity and their applications.)* Master thesis, Charles university, Prague, supervisor J. Zvárová, 2009.
[4] I. Vajda and K. Vašek. Majorization, concave entropies and comparison of experiments. *Problems of Control and Information Theory*, 14:105-115, 1985.

Acknowledgements

This work was supported by the project 1M06014 of the Ministry of Education, Youth and Sports of the Czech Republic. The authors also thank J. Zvárová and M. Studený for various useful suggestions during the preparation of the paper.

Contact address

M. Horáček, Institute of Computer Science AS CR, Pod Vodárenskou věží 2, 182 07 Prague, Czech Republic, e-mail: horacek@euromise.cz