Information Theory Notes

Charles Zheng and Yuval Benjamini

January 28, 2016

1 Proof of Key Result

We give the proof with minimal context or motivation. See the paper for more details.

Fix integer $K \geq 2$. Let $p^{[d]}(x,y)$ be a sequence of probability density functions, where x is of dimension $p^{[d]}$ and y is of dimension $q^{[d]}$. Let $p^{[d]}(x)$ and $p^{[d]}(y)$ denote the marginal densities, and let

$$p^{[d]}(y|x) = p^{[d]}(x,y)/p^{[d]}(y).$$

Let $(X^{([d],i)}, Y^{([d],i)})$ be iid random variates from $p^{[d]}(x,y)$ for $i = 0, \ldots, K-1$; we will supress the superscripts $[d]$ and/or (i) when convenient. Recall the definitions of entropy,

$$H(X) = -\int p(x) \log p(x) dx,$$

and mutual information

$$I(X;Y) = \int p(x,y) \log \frac{p(x,y)}{p(x)p(y)} dxdy.$$

Furthermore, define the K-class average Bayes error as

$$\text{ABE}_K \Pr[p(Y^{(0)}|X^{(0)}) < \max_{i=1}^{K-1} p(Y^{(0)}|X^{(i)})].$$

Define

$$u^{[d]}(x,y) = \log p^{[d]}(x,y) - \log p^{[d]}(x) - \log p^{[d]}(y),$$

and define

$$\ell_{ij}^{[d]} = \log p(y^{(i)}|x^{(j)})$$

Theorem. Let $p^{[d]}(x, y)$ be a sequence of joint densities for $d = 1, 2, \ldots$ as given above. Further assume that

A1. $\lim_{d\to\infty} I(X^{[d]}; Y^{[d]}) = \iota < \infty$.

A2. There exists a sequence of scaling constants $a_{ij}^{[d]}$ and $b_{ij}^{[d]}$ such that the random vector $(a_{ij}\ell_{ij}^{[d]} + b_{ij}^{[d]})_{i,j=1,\ldots,K}$ converges in distribution to a multivariate normal distribution.

A3. There exists a sequence of scaling constants $a^{[d]}$, $b^{[d]}$ such that

$$a^{[d]}u(X^{(1)}, Y^{(2)}) + b^{[d]}$$

converges in distribution to a univariate normal distribution.

A4. For all $i \neq k$,

$$\lim_{d\to\infty} \mathrm{Cov}[u(X^{(i)}, Y^{(j)}), u(X^{(k)}, Y^{(j)})] = 0.$$

Then for ABE_K as defined above, we have

$$\lim_{d\to\infty} \mathrm{ABE}_K = \pi_K(\iota)$$

where

$$\pi_K(\iota) = 1 - \int_{\mathbb{R}} \phi(z - \sqrt{2\iota})(1 - \Phi(z))^{K-1} dz$$

where ϕ and Φ are the standard normal density function and cumulative distribution function, respectively.

Proof.

For $i = 1, \ldots, K-1$, define

$$Z_i = \log p(Y^{(0)}|X^{(i)}) - \log p(Y^{(0)}|X^{(0)}).$$

Then, we claim that $\vec{Z} = (Z_1, \ldots, Z_{K-1})$ converges in distribution to

$$\vec{Z} \sim N\left(-2\iota, \begin{bmatrix} 4\iota & 2\iota & \cdots & 2\iota \\ 2\iota & 4\iota & \cdots & 2\iota \\ \vdots & \vdots & \ddots & \vdots \\ 2\iota & 2\iota & \cdots & 4\iota \end{bmatrix}\right).$$

Combining the claim with the lemma in the following section yields the desired result.

To prove the claim, it suffices to derive the limiting moments

$$\mathbf{E}[Z_i] \to -2\iota,$$

$$\mathrm{Cov}[Z_i] \to 4\iota,$$

$$\mathrm{Cov}[Z_i, Z_j] \to 2\iota,$$

for $i \neq j$, since then assumption A2 implies the existence of a multivariate normal limiting distribution with the given moments.

Before deriving the limiting moments, note the following identities. Let $X' = X^{(1)}$ and $Y = Y^{(0)}$.

$$\mathbf{E}[e^{u(X',Y)}] = \int p(x)p(y)e^{u(x,y)}dxdy = \int p(x,y)dxdy = 1.$$

Therefore, from assumption A3 and the formula for gaussian exponential moments, we have

$$\lim_{d\to\infty} \mathbf{E}[u(X',Y)] - \frac{1}{2}\mathrm{Var}[u(X',Y)] = 0.$$

Let $\sigma^2 = \lim_{d\to\infty} \mathrm{Var}[u(X',Y)]$. Meanwhile, by applying assumption A2,

$$\begin{aligned}
\lim_{d\to\infty} I(X;Y) &= \lim_{d\to\infty} \int p(x,y)u(x,y)dxdy = \lim_{d\to\infty} \int p(x)p(y)e^{u(x,y)}u(x,y)dxdy \\
&= \lim_{d\to\infty} \mathbf{E}[e^{u(X,Y')}u(X,Y')] \\
&= \int_{\mathbb{R}} e^z z \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(z+\sigma^2/2)^2}{2\sigma^2}} \text{ (applying A2)} \\
&= \int_{\mathbb{R}} z \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(z-\sigma^2/2)^2}{2\sigma^2}} \\
&= \frac{1}{2}\sigma^2.
\end{aligned}$$

Therefore,

$$\sigma^2 = 2\iota,$$

and

$$\lim_{d\to\infty} \mathbf{E}[u(X',Y)] = -\iota.$$

Once again by applying A2, we get

$$\begin{aligned}
\lim_{d\to\infty} \mathrm{Var}[u(X,Y)] &= \lim_{d\to\infty} \int (u(x,y)-\iota)^2 p(x,y)dxdy \\
&= \lim_{d\to\infty} \int (u(x,y)-\iota)^2 e^{u(x,y)} p(x)p(y)dxdy \\
&= \lim_{d\to\infty} \mathbf{E}[(u(X',Y)-\iota)^2 e^{u(X',Y)}] \\
&= \int (z-\iota)^2 e^z \frac{1}{\sqrt{4\pi\iota}} e^{-\frac{(z+\iota)^2}{4\iota}} dz \text{ (applying A2)} \\
&= \int (z-\iota)^2 \frac{1}{\sqrt{4\pi\iota}} e^{-\frac{(z-\iota)^2}{4\iota}} dz \\
&= 2\iota.
\end{aligned}$$

We now proceed to derive the limiting moments. We have

$$\begin{aligned}
\lim_{d\to\infty} \mathbf{E}[Z] &= \lim_{d\to\infty} \mathbf{E}[\log p(Y|X') - \log p(Y|X)] \\
&= \lim_{d\to\infty} \mathbf{E}[u(X',Y) - u(X,Y)] = -2\iota.
\end{aligned}$$

Also,

$$\begin{aligned}
\lim_{d\to\infty} \mathrm{Var}[Z] &= \lim_{d\to\infty} \mathrm{Var}[u(X',Y) - u(X,Y)] \\
&= \lim_{d\to\infty} \mathrm{Var}[u(X',Y)] + \mathrm{Var}[u(X,Y)] \text{ (using assumption A4)} \\
&= 4\iota,
\end{aligned}$$

and similarly

$$\begin{aligned}
\lim_{d\to\infty} \mathrm{Cov}[Z_i, Z_j] &= \lim_{d\to\infty} \mathrm{Var}[u(X,Y)] \text{ (using assumption A4)} \\
&= 2\iota.
\end{aligned}$$

This concludes the proof. $\square$.

2 Lemma

Lemma. *Suppose* $(Z_*, Z_1, \ldots, Z_{K-1})$ *are jointly multivariate normal, with* $\boldsymbol{E}[Z_* - Z_1] = \alpha$, $Var(Z_*) = \beta$, $Cov(Z_*, Z_i) = \gamma$, $Var(Z_i) = \delta$, *and*

$Cov(Z_i, Z_j) = \epsilon$ for all $i, j = 1, \ldots, K-1$. Then, letting

$$\mu = \frac{\boldsymbol{E}[Z_* - Z_i]}{\sqrt{\frac{1}{2}\,Var(Z_i - Z_j)}} = \frac{\alpha}{\sqrt{\delta - \epsilon}},$$

$$\nu^2 = \frac{Cov(Z_* - Z_i, Z_* - Z_j)}{\frac{1}{2}\,Var(Z_i - Z_j)} = \frac{\beta + \epsilon - 2\gamma}{\delta - \epsilon},$$

we have

$$\begin{aligned}\Pr[Z_* < \max_{i=1}^{K-1}] &= \Pr[W < M_{K-1}] \\ &= 1 - \int - \frac{1}{\sqrt{2\pi\nu^2}} e^{-\frac{(w-\mu)^2}{2\nu^2}} (1 - \Phi(w))^{K-1} dw,\end{aligned}$$

where $W \sim N(\mu, \nu^2)$ and M_{K-1} is the maximum of $K-1$ independent standard normal variates, which are independent of W.

[ADD PROOF OF LEMMA HERE]