Chernoff Bound on Average Error Probability

Aleksandar Dogandžić

April 14, 2017

Contents

READING: [Fukunaga 1990], [Johnson 2013, §5.1.6], [Van Trees et al. 2013, §2.4], [Cover and Thomas 2006, §11.9], [Poor 1994].

Consider testing

$$\begin{aligned} &\mathbb{H}_0 : \Theta \in \mathrm{sp}_\Theta(0) \quad \text{versus} \\ &\mathbb{H}_1 : \Theta \in \mathrm{sp}_\Theta(1) \end{aligned}$$

For the 0–1 loss, the Bayes' rule that minimizes the average error probability

$$\begin{aligned} \text{av. error probability} = \Pr(\mathbb{H}_0) &\int_{\mathcal{X}_1} f_{\boldsymbol{X}\,|\,\Theta}(\boldsymbol{x} \,|\, \mathbb{H}_0)\, \mathrm{d}\boldsymbol{x} \\ &+ \Pr(\mathbb{H}_1) \int_{\mathcal{X}_0} f_{\boldsymbol{X}\,|\,\Theta}(\boldsymbol{x} \,|\, \mathbb{H}_1)\, \mathrm{d}\boldsymbol{x} \end{aligned} \tag{1}$$

is the maximum *a posteriori* (MAP) decision rule:

$$\mathcal{X}_1^\star = \big\{\boldsymbol{x} \mid f(\boldsymbol{x} \,|\, \mathbb{H}_0)\Pr(\mathbb{H}_0) - f(\boldsymbol{x} \,|\, \mathbb{H}_1)\Pr(\mathbb{H}_1) < 0\big\}. \tag{2}$$

see (7b) in handout `Bayesdet`

In many applications, we *may not be able to obtain* a simple closed-form expression for the minimum average error probability, but we *can find an upper bound for it* as follows:

$$\begin{aligned} \text{min av. error prob.} &= \Pr(\mathbb{H}_0) \int_{\mathcal{X}_1^\star} f_{\boldsymbol{X}\,|\,\Theta}(\boldsymbol{x} \,|\, \mathbb{H}_0)\, \mathrm{d}\boldsymbol{x} \\ &\quad + \Pr(\mathbb{H}_1) \int_{\mathcal{X}_0^\star} f_{\boldsymbol{X}\,|\,\Theta}(\boldsymbol{x} \,|\, \mathbb{H}_1)\, \mathrm{d}\boldsymbol{x} \\ &= \int_{\mathcal{X}} \min\big\{f(\boldsymbol{x} \,|\, \mathbb{H}_0)\Pr(\mathbb{H}_0), f(\boldsymbol{x} \,|\, \mathbb{H}_1)\Pr(\mathbb{H}_1)\big\}\, \mathrm{d}\boldsymbol{x} \\ &\le \int_{\mathcal{X}} [f(\boldsymbol{x} \,|\, \mathbb{H}_0)\Pr(\mathbb{H}_0)]^\lambda [f(\boldsymbol{x} \,|\, \mathbb{H}_1)\Pr(\mathbb{H}_1)]^{1-\lambda}\, \mathrm{d}\boldsymbol{x} \end{aligned}$$

plug the definition of $\mathcal{X}_1^\star$ into (2)

$0 \le \lambda \le 1$

which is the *Chernoff bound on the minimum average error probability*. Here, we have used the fact that

$$\min\{a, b\} \le a^\lambda b^{1-\lambda}, \quad \text{for} \quad 0 \le \lambda \le 1, \quad a, b \ge 0.$$

Conditionally I.I.D. Measurements

SUPPOSE that $(X[n])_{n=0}^{N-1}$ are conditionally independent, identically distributed (i.i.d.) given $\Theta = \theta$, following $f_{X\,|\,\Theta}(x[n]\,|\,\theta)$. Then, the Chernoff bound becomes

$$\begin{aligned}\text{Chernoff bound} &= \int_{\mathcal{X}} \Big[\Pr(\mathbb{H}_0) \prod_{n=0}^{N-1} f(\boldsymbol{x}\,|\,\mathbb{H}_0)\Big]^{\lambda} \Big[\Pr(\mathbb{H}_1) \prod_{n=0}^{N-1} f(\boldsymbol{x}\,|\,\mathbb{H}_1)\Big]^{1-\lambda} \\ &= [\Pr(\mathbb{H}_0)]^{\lambda}[\Pr(\mathbb{H}_1)]^{1-\lambda} \prod_{n=0}^{N-1}\Big\{\int_{\mathcal{X}} [f(x[n]\,|\,\mathbb{H}_0)]^{\lambda}[f(x[n]\,|\,\mathbb{H}_1)]^{1-\lambda}\,\mathrm{d}x[n]\Big\} \\ &= [\Pr(\mathbb{H}_0)]^{\lambda}[\Pr(\mathbb{H}_1)]^{1-\lambda}\Big\{\int_{\mathcal{X}} [f(x\,|\,\mathbb{H}_0)]^{\lambda}[f(x\,|\,\mathbb{H}_1)]^{1-\lambda}\,\mathrm{d}x\Big\}^{N}\end{aligned}$$

and, therefore,

$$\begin{aligned}\frac{1}{N}\ln(\text{min av. error prob.}) \le{}& \frac{1}{N}\ln\Big\{[\Pr(\mathbb{H}_0)]^{\lambda}[\Pr(\mathbb{H}_1)]^{1-\lambda}\Big\} \\ &+ \ln\int_{\mathcal{X}} [f(x\,|\,\mathbb{H}_0)]^{\lambda}[f(x\,|\,\mathbb{H}_1)]^{1-\lambda}\,\mathrm{d}x. \qquad \forall\lambda \in [0,1]\end{aligned}$$

If $\Pr(\mathbb{H}_0) = \Pr(\mathbb{H}_1) = 1/2$ (which is often of interest when evaluating average error probabilities), we can state that for N conditionally i.i.d. measurements given $\Theta = \theta$ and simple hypotheses

$$\text{min av. error probability} \to w(N)\exp\Big(-N\underbrace{\Big\{-\min_{\lambda\in[0,1]}\int_{\mathcal{X}} [f(x\,|\,\mathbb{H}_0)]^{\lambda}[f(x\,|\,\mathbb{H}_1)]^{1-\lambda}\,\mathrm{d}x\Big\}}_{\text{Chernoff information for a single observation}}\Big)$$

as $N \nearrow +\infty$, where $w(N)$ is a slowly-varying function compared with the exponential term:

$$\lim_{N\nearrow+\infty}\frac{\ln w(N)}{N} = 0.$$

Note that the Chernoff information in the exponent term of the above minimum error-probability expression quantifies the asymptotic behavior of the minimum average error probability.

We now give a useful result for evaluating a class of Chernoff bounds, taken from [Fukunaga 1990].

Lemma 1. Consider $f_1(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}\,|\,\boldsymbol{\mu}_1, \Sigma_1)$ and $f_2(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}\,|\,\boldsymbol{\mu}_2, \Sigma_2)$. Then

$$\int [f_1(\boldsymbol{x})]^{\lambda}[f_2(\boldsymbol{x})]^{1-\lambda}\,\mathrm{d}\boldsymbol{x} = \exp[-g(\lambda)]$$

where

$$\begin{aligned}g(\lambda) = {}& \frac{\lambda(1-\lambda)}{2}(\boldsymbol{\mu}_2-\boldsymbol{\mu}_1)^T[\lambda\Sigma_1 + (1-\lambda)\Sigma_2]^{-1}(\boldsymbol{\mu}_2-\boldsymbol{\mu}_1) \\ &+0.5\ln\Big[\frac{\det[\lambda\Sigma_1 + (1-\lambda)\Sigma_2]}{\det(\Sigma_1)^{\lambda}\det(\Sigma_2)^{1-\lambda}}\Big].\end{aligned}$$

☞ HERE, we bound average error probability, whereas [Poor 1994; Van Trees et al. 2013] bound conditional error probabilities using similar tools.

References

Cover, Thomas M and Joy A Thomas (2006). *Elements of Information Theory*. 2nd ed. New York: Wiley (cit. on p. 1).

Fukunaga, Keinosuke (1990). *Introduction to Statistical Pattern Recognition*. 2nd ed. Boston, MA: Academic Press Inc. (cit. on pp. 1, 2).

Johnson, Don H. (2013). *Statistical Signal Processing*. Lecture notes. Rice Univ., Houston, TX (cit. on p. 1).

Poor, H. Vincent (1994). *An Introduction to Signal Detection and Estimation*. 2nd ed. New York: Springer (cit. on pp. 1, 3).

Van Trees, Harry L., Kristine L. Bell, and Zhi Tian (2013). *Detection, Estimation, and Modulation Theory, Part I*. 2nd ed. New York: Wiley (cit. on pp. 1, 3).