A note on the Chernoff information between Bernoulli distributions

Frank Nielsen
Sony Computer Science Laboratories Inc

November 2024

The probability mass function of a random variable X following a Bernoulli distribution $\mathrm{Ber}(p)$ with probability of success $p \in (0,1)$ is:

$$\Pr(X = x) = f_p(x) = p^x(1-p)^{1-x}, \quad x \in \mathcal{X} = \{0,1\}.$$

The Chernoff information [1] between two Bernoulli distributions f_{p_1} and f_{p_2} is

$$C(f_{p_1}, f_{p_2}) = \max_{\alpha\in(0,1)} -\log\left(f_{p_1}(0)^\alpha f_{p_2}^{1-\alpha}(0) + f_{p_1}(1)^\alpha f_{p_2}^{1-\alpha}(1)\right).$$

The optimal exponent α maximizing the skew Bhattacharyya distance

$$B_\alpha(f_{p_1}, f_{p_2}) = -\log\left(f_{p_1}(0)^\alpha f_{p_2}^{1-\alpha}(0) + f_{p_1}(1)^\alpha f_{p_2}^{1-\alpha}(1)\right)$$

is called the Chernoff information.

The family $\{f_p \ : \ p \in (0,1)\}$ of Bernoulli distributions form a discrete exponential family of order 1 with natural parameter $\theta = \log(p/(1-p))$ (and $\lambda(\theta) = \frac{e^\theta}{1+e^\theta}$), cumulant function $F(\theta) = \log(1+e^\theta)$, moment parameter $\eta(\theta) = \frac{e^\theta}{1+e^\theta}$ (with $\eta(p) = p$ and $\theta(\eta) = \log\frac{\eta}{1-\eta}$), convex conjugate (negentropy) $F^*(\eta) = \eta\log\eta + (1-\eta)\log(1-\eta) = -H(\mathrm{Ber}(p))$.

Assume $p_1 \neq p_2$ (otherwise $C(f_{p_1}, f_{p_2}) = 0$). A closed-form formula for the Chernoff exponent of uni-order exponential families was reported in [2]:

$$\alpha^*(\theta_1, \theta_2) = \frac{F^{*\prime}\left(\frac{\Delta_F}{\Delta_\theta}\right) - \theta_2}{\Delta_\theta},$$

with $\Delta_\theta = \theta_1 - \theta_2$ and $\Delta_F = F(\theta_1) - F(\theta_2)$.

Applying this generic formula to the case of Bernoulli distributions on the ordinary parameterization, we get the Chernoff exponent between Bernoulli distributions:

$$\alpha^*(p_1, p_2) = \frac{\log\left(-\left(\frac{\log\left(\frac{p_1-1}{p_2-1}\right)(p_2-1)}{p_2\log\left(\frac{p_2}{p_1}\right)}\right)\right)}{\log\left(\frac{p_1(p_2-1)}{(p_1-1)p_2}\right)},$$

and the Chernoff information in closed-form as

$$C(f_{p_1}, f_{p_2}) = B_{\alpha^*(p_1,p_2)}(f_{p_1}, f_{p_2}).$$

Let $a=\frac{p_1}{p_2}$ and $b=\frac{1-p_1}{1-p_2}$. Then the Chernoff exponent can be rewritten as

$$\alpha^*(p_1,p_2)=\frac{\log\left(\frac{\log b}{\log a}\left(1-\frac{1}{p_2}\right)\right)}{\log\frac{a}{b}}. \tag{1}$$

Let f_{p^*} be the Chernoff distribution where $p^*=p_{\alpha^*}$ corresponds to the parameter of the Bernoulli distribution obtained after normalization of $f_{p_1}^{\alpha^*}f_{p_2}^{1-\alpha^*}$ (e-geodesic). We have

$$p_\alpha=\frac{\left(\frac{p_1}{1-p_1}\right)^\alpha\left(\frac{p_2}{1-p_2}\right)^{1-\alpha}}{1+\left(\frac{p_1}{1-p_1}\right)^\alpha\left(\frac{p_2}{1-p_2}\right)^{1-\alpha}}.$$

We check that we have

$$C(f_{p_1},f_{p_2})=D_{\mathrm{KL}}(f_{p^*}:f_{p_1})=D_{\mathrm{KL}}(f_{p^*}:f_{p_2}),$$

where D_{KL} denotes the Kullback-Leibler divergence:

$$D_{\mathrm{KL}}(f_{p_1},f_{p_2})=f_{p_1}(0)\log\frac{f_{p_1}(0)}{f_{p_2}(0)}+f_{p_1}(1)\log\frac{f_{p_1}(1)}{f_{p_2}(1)}=p_1\log\frac{p_1}{p_2}+(1-p_1)\log\frac{1-p_1}{1-p_2}.$$

Example 1 *Let $p_1=0.1$ and $p_2=0.2$. We have $\alpha^*\simeq 0.47612450297278$, $p^*\simeq 0.14524435432427263$, and $C(f_{p_1},f_{p_2})\simeq 0.0101245165799591$.*

Example 2 *Let $p_1=0.1$ and $p_2=0.7$. We have $\alpha^*\simeq 0.466076329662342$, $p^*\simeq 0.360848806714530$, and $C(f_{p_1},f_{p_2})\simeq 0.244321377719848$.*

Note that the Chernoff information is known in closed-form between univariate normal distributions [3] but not for arbitrary d-variate normal distributions.

A code snippet in MAXIMA (`https://maxima.sourceforge.io/`) to calculate the Chernoff exponent:

```
kill(all);
lambda2theta(p):= log(p/(1-p));
theta2lambda(theta):=exp(theta)/(1+exp(theta));
F(theta):=log(1+exp(theta));
Gprime(eta):=log(eta/(1-eta));
deltaF(theta1,theta2):=F(theta2)-F(theta1);
deltaTheta(theta1,theta2):=theta2-theta1;
alpha(theta1,theta2):=(Gprime(deltaF(theta1,theta2)
/deltaTheta(theta1,theta2))-theta2)/(theta1-theta2);
alpha(lambda2theta(p1),lambda2theta(p2));
ratsimp(%);logcontract(%);tex(%);
```

References

[1] Herman Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *The Annals of Mathematical Statistics*, pages 493–507, 1952.

[2] Frank Nielsen. An information-geometric characterization of Chernoff information. *IEEE Signal Processing Letters*, 20(3):269–272, 2013.

[3] Frank Nielsen. Revisiting Chernoff information with likelihood ratio exponential families. *Entropy*, 24(10):1400, 2022.