The Kullback-Leibler divergence between a Poisson distribution and a geometric distribution

Frank Nielsen
Frank.Nielsen@acm.org

October 4, 2021

Abstract

This note illustrates how to apply the generic formula of the Kullback-Leibler divergence between two densities of two different exponential families [2].

This column is also available as the file `KLPoissonGeometricDistributions.pdf`.

It is well-known that the Kullback-Leibler between two densities P_{θ_1} and P_{θ_2} of the same exponential family amounts to a reverse Bregman divergence between the corresponding natural parameters for the Bregman generator set to the cumulant function $F(\theta)$ [1]:

$$D_{\mathrm{KL}}[P_{\theta_1} : P_{\theta_2}] = {B_F}^*(\theta_1 : \theta_2) = B_F(\theta_2 : \theta_1) := F(\theta_2) - F(\theta_1) - (\theta_2 - \theta_1) \cdot \nabla F(\theta_1).$$

The following formula for the Kullback-Leibler divergence (KLD) between two densities P_θ and $Q_{\theta'}$ of two different exponential families $\mathcal{P}$ (with cumulant function $F_\mathcal{P}$) and $\mathcal{Q}$ (with cumulant function $F_\mathcal{Q}$) was reported in [2] (Proposition 5):

$$D_{\mathrm{KL}}[P_\theta : Q_{\theta'}] = F_\mathcal{Q}(\theta') + F_\mathcal{P}^*(\eta) - E_{P_\theta}[t_\mathcal{Q}(x)] \cdot \theta' + E_{P_\theta}[k_\mathcal{P}(x) - k_\mathcal{Q}(x)]. \quad (1)$$

When $\mathcal{P} = \mathcal{Q}$ (and $F = F_\mathcal{P} = F_\mathcal{Q}$), we recover the reverse Fenchel-Young divergence which corresponds to the reverse Bregman divergence:

$$D_{\mathrm{KL}}[P_\theta : P_{\theta'}] = F(\theta') + F^*(\eta) - \eta \cdot \theta' =: Y_{F,F^*}(\theta' : \eta) = Y_{F^*,F}(\eta : \theta').$$

Consider the KLD between a Poisson probability mass function (pmf) and a geometric pmf. The canonical decomposition of the Poisson and geometric pmfs are summarized in Table 1.

Thus we calculate the KLD between two geometric distributions Q_{p_1} and Q_{p_2} as

$$\begin{aligned} D_{\mathrm{KL}}[Q_{p_1} : Q_{p_2}] &= B_{F_\mathcal{Q}}(\theta(p_2) : \theta(p_1)), \\ &= F_\mathcal{Q}(\theta(p_2)) - F_\mathcal{Q}(\theta(p_1)) - (\theta(p_2) - \theta(p_1))\eta(p_1), \end{aligned}$$

That is, we have

$$\boxed{D_{\mathrm{KL}}[Q_{p_1} : Q_{p_2}] = \log\left(\frac{p_1}{p_2}\right) - \left(1 - \frac{1}{p_1}\right) \log \frac{1-p_1}{1-p_2}}.$$

The following code in Maxima (`https://maxima.sourceforge.io/`) check the above formula.

```
Geometric(x,p):=((1-p)**x)*p;
nbterms:50;
KLGeometricSeries(p1,p2):=sum((Geometric(x,p1)*log(Geometric(x,p1)/Geometric(x,p2))),x,0,nbterms);
KLGeometricFormula(p1,p2):=log(p1/p2)-log((1-p2)/(1-p1))*((1/p1)-1);
p1:0.2;
p2:0.6;
float(KLGeometricSeries(p1,p2));
float(KLGeometricFormula(p1,p2));
```

	Poisson family $\mathcal{P}$	Geometric family $\mathcal{Q}$
support	$\mathbb{N} \cup \{0\}$	$\mathbb{N} \cup \{0\}$
base measure	counting measure	counting measure
ordinary parameter	rate $\lambda > 0$	success probability $p \in (0,1)$
pmf	$\frac{\lambda^x}{x!} \exp(-\lambda)$	$(1-p)^x p$
sufficient statistic	$t_{\mathcal{P}}(x) = x$	$t_{\mathcal{Q}}(x) = x$
natural parameter	$\theta(\lambda) = \log \lambda$	$\theta(p) = \log(1-p)$
cumulant function	$F_{\mathcal{P}}(\theta) = \exp(\theta)$	$F_{\mathcal{Q}}(\theta) = -\log(1-\exp(\theta))$
	$F_{\mathcal{P}}(\lambda) = \lambda$	$F_{\mathcal{Q}}(p) = -\log(p)$
auxiliary measure term	$k_{\mathcal{P}}(x) = -\log x!$	$k_{\mathcal{Q}}(x) = 0$
moment parameter $\eta = E[t(x)]$	$\eta = \lambda$	$\eta = \frac{e^\theta}{1-e^\theta} = \frac{1}{p} - 1$
negentropy (convex conjugate) $(F^*(\eta) = \theta \cdot \eta - F(\theta))$	$F^*_{\mathcal{P}}(\eta(\lambda)) = \lambda \log \lambda - \lambda$	$F^*_{\mathcal{Q}}(\eta(p)) = \left(1 - \frac{1}{p}\right) \log(1-p) + \log p$

Table 1: Canonical decomposition of the Poisson and the geometric discrete exponential families.

Evaluating the above code, we get:

```
(%o7) 1.673553688712277
(%o8) 1.673976433571672
```

Thus we have the KLD between a Poisson pmf p_λ and a geometric pmf q_p is equal to

$$\begin{aligned} D_{\mathrm{KL}}[P_\lambda : Q_p] &= F_{\mathcal{Q}}(\theta') + F^*_{\mathcal{P}}(\eta) - E_{P_\theta}[t_{\mathcal{Q}}(x)] \cdot \theta' + E_{P_\theta}[k_{\mathcal{P}}(x) - k_{\mathcal{Q}}(x)], \quad (2) \\ &= -\log p + \lambda \log \lambda - \lambda \log(1-p) - E_{P_\lambda}[\log x!] \quad (3) \end{aligned}$$

Since $E_{p_\lambda}[-\log x!] = -\sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k \log(k!)}{k!}$, we have

$$\boxed{D_{\mathrm{KL}}[P_\lambda : Q_p] = -\log p + \lambda \log \frac{\lambda}{1-p} - \lambda - \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k \log(k!)}{k!}}$$

We check in Maxima the above formula:

```
Poisson(x,lambda):=(lambda**x)*exp(-lambda)/x!;
KLseries(lambda,p):=sum((Poisson(x,lambda)*log(Poisson(x,lambda)/Geometric(x,p))),x,0,nbterms);
KLformula(lambda,p):=-log(p)+lambda*log(lambda)-lambda-lambda*log(1-p)
-sum(exp(-lambda)*(lambda**x)*log(x!)/x!,x,0,nbterms);
lambda:5.6;
p:0.3;
float(KLseries(lambda,p));
float(KLformula(lambda,p));
```

Evaluating the above code, we get

```
(%o14) 0.9378529269681795
(%o15) 0.9378529269681785
```

References

[1] Arindam Banerjee, Srujana Merugu, Inderjit S Dhillon, Joydeep Ghosh, and John Lafferty. Clustering with Bregman divergences. *Journal of machine learning research*, 6(10), 2005.

[2] Frank Nielsen. On a Variational Definition for the Jensen-Shannon Symmetrization of Distances Based on the Information Radius. *Entropy*, 23(4):464, 2021.