# Relations between Kullback-Leibler distance and Fisher information

Anand G. Dabak
Texas Instruments
DSP R&D Center
Dallas, Texas
dabak@ti.com

Don H. Johnson*
Dept. of Electrical & Computer Engineering
Rice University
Houston, Texas
dhj@rice.edu

**Abstract**

The Kullback-Leibler distance between two probability densities that are parametric perturbations of each other is related to the Fisher information. We generalize this relationship to the case when the perturbations may not be small and when the two densities are non-parametric.

**Index Terms**

Kullback-Leibler distance, Fisher information

EDICS: 2-INFO

## I. Introduction

CONSIDER a parametric density $p(\omega, \theta)$ defined over a probability space $\Omega$ parametrized by $\theta \in \mathbb{R}$. The Kullback-Leibler distance between $p(\omega, \theta_0)$ and $p(\omega, \theta_1)$ is given by [2, 3, 6]

$$\mathcal{D}(p(\omega,\theta_1)\|p(\omega,\theta_0)) = \int_\Omega p(\omega,\theta_1)\ln\frac{p(\omega,\theta_1)}{p(\omega,\theta_0)}\,d\omega$$

When $\theta_1 = \theta_0 + \delta\theta$ with $\delta\theta$ a perturbation, the Kullback-Leibler distance is proportional to the density's Fisher information [6].

$$\mathcal{D}(p(\omega,\theta_0+\delta\theta)\|p(\omega,\theta_0)) \overset{\delta\theta\to 0}{=} \frac{1}{2}F(\theta_0)(\delta\theta)^2 \tag{1}$$

where $F(\theta)$ is the Fisher information [5, Page 158] of $p(\omega, \theta)$ with respect to the parameter $\theta$.

$$F(\theta) = \int_{\mathbf{\Omega}} p(\omega,\theta)\left(\frac{1}{p(\omega,\theta)}\frac{\partial p(\omega,\theta)}{\partial\theta}\right)^2 d\omega \tag{2}$$

Said another way, equation (1) means that the second derivative of the Kullback-Leibler distance equals the Fisher information.

$$\left.\frac{\partial^2\mathcal{D}(p(\omega,\theta)\|p(\omega,\theta_0))}{\partial\theta^2}\right|_{\theta=\theta_0} = F(\theta_0) \tag{3}$$

Note that this relation (to within a constant of proportionality) applies to all Ali-Silvey distances [1] and others as well. In this correspondence we generalize the relation between Kullback-Leibler distance and Fisher information when the condition $\delta\theta$ small may not hold and when we do not have parametric densities.

## II. Results

Consider two probability density functions $p_0(\omega)$ and $p_1(\omega)$ defined on a probability space $\Omega$. As mentioned above they could be arbitrary densities, not necessarily defined by an underlying parametric density. The only condition required in subsequent results is that the second and third moments ($n = 2, 3$) of the log-likelihood ratio with respect to $p_0$ and $p_1$ are finite.

$$\begin{aligned} &\int_\Omega \left(\ln \frac{p_1(\omega)}{p_0(\omega)}\right)^n p_1(\omega)\, d\omega < \infty \\ &\int_\Omega \left(\ln \frac{p_0(\omega)}{p_1(\omega)}\right)^n p_0(\omega)\, d\omega < \infty \end{aligned} \tag{4}$$

Employing the Cauchy-Schwartz inequality, we find that

$$\int_\Omega \left|\ln \left(\frac{p_1(\omega)}{p_0(\omega)}\right)\right| p_1(\omega)\ d\omega \leq \left[\int_\Omega \left(\ln \frac{p_1(\omega)}{p_0(\omega)}\right)^2 p_1(\omega)\, d\omega\right]^{1/2} \left[\int_\Omega p_1(\omega)\, d\omega\right]^{1/2} < \infty \tag{5}$$

which means that our second-moment conditions imply that $\mathcal{D}(p_1\|p_0) < \infty$; similar considerations show that $\mathcal{D}(p_0\|p_1) < \infty$. Because the Kullback-Leibler distances are finite, our second-moment conditions mean that $p_0(\omega)$ and $p_1(\omega)$ have common support: $\forall \omega \in \Omega$, $p_0(\omega) = 0 \implies p_1(\omega) = 0$ and vice versa. Hence, the following parametric density is well defined.

$$\begin{aligned} p_t(\omega) &= \frac{[p_0(\omega)]^{1-t}\,[p_1(\omega)]^t}{N_t} \\ N_t &= \int_\Omega [p_0(\omega)]^{1-t}\,[p_1(\omega)]^t\ d\omega,\ 0 \leq t \leq 1 \end{aligned} \tag{6}$$

The density $p_t$ is well known in the literature as the exponential twist density [2]. The normalizing function $N_t$ is a strictly convex function and $0 < N_t \leq 1$ over $0 \leq t \leq 1$ [2, 3]. With $t$ the parameter of the density, $p_t$ can be considered a curve on the manifold of probability densities connecting $p_0$ and $p_1$, which are arbitrary save for conditions (4). This curve starts at $p_0$ with the curve's parameter $t$ equaling zero and ending at $p_1$ with $t = 1$. When $\Omega$ is a simplex, $p_t$ is the geodesic connecting the two densities [3]. Under the second-moment conditions (4), $p_t$ is the geodesic even when $\Omega$ is not a simplex [4]. However, for the present correspondence, this fact is not used.

Important here is the Kullback-Leibler distance between two densities $p_u$ and $p_t$ on the geodesic.

$$\mathcal{D}(p_t\|p_u) = \int_\Omega p_t(\omega) \ln \frac{p_t(\omega)}{p_u(\omega)}\, d\omega, \quad 0 \leq t, u \leq 1 \tag{7}$$

*Result 1:* Under conditions (4), if we define the Fisher information of $p_t(\omega)$ at $t$ as

$$F(t) = \int_\Omega \left(\frac{d\ \ln p_t(\omega)}{dt}\right)^2 p_t(\omega)\, d\omega \tag{8}$$

then $F(t) < \infty$ and $F'(t)$ exists $\forall t \in [0, 1]$.

To prove that the Fisher information is always finite, we find that the derivative $\frac{d \ln p_t(\omega)}{dt}$ equals

$$\frac{d \ln p_t(\omega)}{dt} = \ln \frac{p_1(\omega)}{p_0(\omega)} - \int_\Omega p_t(\omega) \ln \frac{p_1(\omega)}{p_0(\omega)}\, d\omega$$

Substituting into equation (8) and simplifying gives

$$F(t) = \int_\Omega p_t(\omega) \left(\ln \frac{p_1(\omega)}{p_0(\omega)}\right)^2 d\omega - \ \left(\int_\Omega p_t(\omega) \ln \frac{p_1(\omega)}{p_0(\omega)}\, d\omega\right)^2 \tag{9}$$

Let $\Omega^+$ denote the set of all $\omega \in \Omega$ such that $p_0(\omega) \geq p_1(\omega)$. Similarly, let $\Omega^-$ denote the set of all $\omega \in \Omega$ such that $p_0(\omega) < p_1(\omega)$. The first integral in (9) equals,

$$\int_\Omega p_t(\omega)\left(\ln \frac{p_1(\omega)}{p_0(\omega)}\right)^2 d\omega = \frac{1}{N_t}\int_{\Omega^+} p_0(\omega)\left(\ln \frac{p_1(\omega)}{p_0(\omega)}\right)^2 \left(\frac{p_1(\omega)}{p_0(\omega)}\right)^t d\omega + \frac{1}{N_t}\int_{\Omega^-} p_1(\omega)\left(\ln \frac{p_1(\omega)}{p_0(\omega)}\right)^2 \left(\frac{p_0(\omega)}{p_1(\omega)}\right)^{(1-t)} d\omega$$

Notice that over $\Omega^+$, $\frac{p_1(\omega)}{p_0(\omega)} \leq 1$ and over $\Omega^-$, $\frac{p_0(\omega)}{p_1(\omega)} < 1$. Thus, using the second-moment conditions (4) and the fact that $N_t > 0$ for $\forall t \in [0,1]$ gives us

$$\int_\Omega p_t(\omega)\left(\ln \frac{p_1(\omega)}{p_0(\omega)}\right)^2 d\omega < \infty$$

Similarly, the second part of the right-hand side of equation (9) is also finite. Thus $F(t) < \infty$, proving the first part of the result. The differentiability of the Fisher information follows because the derivative can be taken inside the integrals in (9) and $p_t$ is differentiable. The derivative $F'(t)$ is finite if we assume the third-moment condition in (4).□

The following three results relate the Kullback-Leibler distance between densities on the geodesic (7) and the Fisher information (9).

*Result 2:* Derivatives of the Kullback-Leibler distance with respect to the first argument's parameter depend on the Fisher information.

$$\frac{\partial \mathcal{D}(p_t \| p_u)}{\partial t} = (t-u)F(t) \quad \forall t, u \in [0,1] \tag{10}$$

$$\left.\frac{\partial^2 \mathcal{D}(p_t \| p_u)}{\partial t^2}\right|_{t=u} = F(u) \quad \forall u \in [0,1] \tag{11}$$

To show this, consider

$$\mathcal{D}(p_t \| p_u) = \int_\Omega (t-u)p_t(\omega)\ln\left(\frac{p_1(\omega)}{p_0(\omega)}\right) d\omega - \ln\frac{N_t}{N_u}$$

Differentiating both sides with respect to $t$,

$$\frac{\partial \mathcal{D}(p_t \| p_u)}{\partial t} = \int_\Omega (t-u)\frac{dp_t(\omega)}{dt}\ln\frac{p_1(\omega)}{p_0(\omega)} d\omega + \int_\Omega p_t(\omega)\ln\frac{p_1(\omega)}{p_0(\omega)} d\omega - \frac{d\ln N_t}{dt}$$

We find that $\frac{dp_t(\omega)}{dt} = p_t(\omega)\left(\ln\frac{p_1(\omega)}{p_0(\omega)} - \frac{d\ln N_t}{dt}\right)$ and that $\frac{d\ln N_t}{dt} = \int p_t(\omega)\ln\frac{p_1(\omega)}{p_0(\omega)} d\omega$, which gives

$$\frac{\partial \mathcal{D}(p_t \| p_u)}{\partial t} = (t-u)\int_\Omega p_t(\omega)\ln\frac{p_1(\omega)}{p_0(\omega)}\left(\ln\frac{p_1(\omega)}{p_0(\omega)} - \frac{d\ln N_t}{dt}\right) d\omega \tag{12}$$

Comparing this expression with (9), which gives us (10). Evaluating the derivative of (10) yields

$$\frac{\partial^2 \mathcal{D}(p_t \| p_u)}{\partial t^2} = F(t) + (t-u)F'(t)$$

Evaluating at $t = u$ gives the result (11) that the second derivative of the Kullback-Leibler distance equals the Fisher information, thereby generalizing (3). □

Note that results (10) and (11) describe relationships between Fisher information and derivatives with respect to the geodesic curve parameter of the *first* argument of the Kullback-Leibler distance. The Kullback-Leibler distance is generally not a symmetric function of its arguments and is not a symmetric function of densities along the geodesic.

*Result 3:* The integral form of the differential results 2 is

$$\mathcal{D}(p_t \| p_0) = \int_0^t uF(u)\, du \tag{13}$$

Integrating equation (10) and noting $\mathcal{D}(p_0 \| p_0) = 0$ proves this result. □

Thus the Kullback-Leibler information between any two densities satisfying equation (4) is related to the integral of the product of the Fisher information and the parameter $t$ along the geodesic curve in equation (6).

*Result 4:* The sum of the Kullback-Leibler distances between $p_0$ and $p_1$, known as the $J$-divergence [5], equals the integral of the Fisher information along the geodesic connecting $p_0$ and $p_1$.[1]

$$\mathcal{D}(p_0 \| p_1) + \mathcal{D}(p_1 \| p_0) = \int_0^1 F(t)\ dt \tag{14}$$

To show this result, reparametrize equation (6) with $u = 1 - t$ and use a derivation similar to above to yield

$$\mathcal{D}(p_0 \| p_1) = \int_0^1 uF(1-u)\, du = \int_0^1 (1-t)F(t)\, dt$$

Adding (13) gives the result. □

## III. Conclusions

The fundamental relation (3) between the Kullback-Leibler distance and Fisher information applies when we consider densities having a common parameterization. This result also applies when $\theta$ represents a parameter vector, with the second mixed partial of the Kullback-Leibler distance equaling the corresponding term of the Fisher information matrix. Here, we have generalized (3) to the case of non-parametric densities by considering the behavior of the Kullback-Leibler distance along the geodesic connecting two densities. In addition, we have found new properties relating the Kullback-Leibler distance to the integral of the Fisher information along the geodesic path between two densities. Because the Fisher information corresponds to the Riemannian metric on the manifold of probability measures, we see that its integral along the geodesic is the $J$-divergence. Unfortunately, this quantity cannot be construed to be the distance between $p_0$ and $p_1$ [4].

[1] Acknowledgement to Srinath Hosur, Texas Instruments, for pointing out this equality.

## References

[1] S.M. Ali and D. Silvey. A general class of coefficients of divergence of one distribution from another. *J. Roy. Stat. Soc., Ser. B*, 28: 131–142, 1966.

[2] J.A. Bucklew. *Large Deviation Techniques in Decision, Simulation and Estimation*. John Wiley & Sons, 1990.

[3] N.N. Čencov. *Statistical Decision Rules and Optimal Inference*, volume 14. American Mathematical Society, Providence, Rhode Island, 1972.

[4] A G. Dabak. *A Geometry for Detection Theory*. PhD thesis, Rice University, Houston, Tx, 1992.

[5] H. Jeffreys. *Theory of Probability*. Oxford University Press, 1948.

[6] S. Kullback. *Information Theory and Statistics*. Wiley, New York, 1959.