KL-Divergences of Normal, Gamma, Dirichlet and Wishart densities

W.D. Penny

Wellcome Department of Cognitive Neurology,
University College London.

March 30, 2001

1 Kullback-Liebler Divergence

For densities $q(x)$ and $p(x)$ the KL-divergence from q to p is

$$KL(q;p) = \int q(x) \log \frac{q(x)}{p(x)} dx \tag{1}$$

2 Normal

For univariate Normal densities $q(x) = N(x; \mu_q, \sigma_q^2)$ and $p(x) = N(x; \mu_p, \sigma_p^2)$ the KL-divergence is

$$KL_{N_1}(\mu_q, \sigma_q; \mu_p, \sigma_p) = 0.5 \log \frac{\sigma_p^2}{\sigma_q^2} + \frac{\mu_q^2 + \mu_p^2 + \sigma_q^2 - 2\mu_q \mu_p}{2\sigma_p^2} - 0.5 \tag{2}$$

The multivariate Normal density is given by

$$\mathsf{N}(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}^{-1}) = (2\pi)^{-d/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}-\boldsymbol{\mu})\right) \tag{3}$$

The KL divergence for Normal densities $q(\boldsymbol{x}) = \mathsf{N}(\boldsymbol{x}; \boldsymbol{\mu}_q, \boldsymbol{\Sigma}_q^{-1})$ and $p(\boldsymbol{x}) = \mathsf{N}(\boldsymbol{x}; \boldsymbol{\mu}_p, \boldsymbol{\Sigma}_p^{-1})$ is

$$\begin{aligned} KL_N(\boldsymbol{\mu}_q, \boldsymbol{\Sigma}_q; \boldsymbol{\mu}_p, \boldsymbol{\Sigma}_p) &= 0.5 \log \frac{|\boldsymbol{\Sigma}_p|}{|\boldsymbol{\Sigma}_q|} + 0.5 Tr(\boldsymbol{\Sigma}_p^{-1} \boldsymbol{\Sigma}_q) \\ &+ 0.5 (\boldsymbol{\mu}_q - \boldsymbol{\mu}_p)^T \boldsymbol{\Sigma}_p^{-1} (\boldsymbol{\mu}_q - \boldsymbol{\mu}_p) - \frac{d}{2} \end{aligned} \tag{4}$$

where $|\boldsymbol{\Sigma}_p|$ denotes the determinant of the matrix $\boldsymbol{\Sigma}_p$.

3 Gamma

The Gamma density is defined as

$$Ga(x; b, c) = \frac{1}{\Gamma(c)} \frac{x^{c-1}}{b^c} \exp\left(\frac{-x}{b}\right) \tag{5}$$

For Gamma densities $q(x) = Ga(\boldsymbol{x}; b_q, c_q)$ and $p(x) = Ga(\boldsymbol{x}; b_p, c_p)$ the KL-divergence is

$$\begin{aligned} KL_{Ga}(b_q, c_q; b_p, c_p) &= (c_q - 1)\Psi(c_q) - \log b_q - c_q - \log\Gamma(c_q) \\ &+ \log\Gamma(c_p) + c_p \log b_p - (c_p - 1)(\Psi(c_q) + \log b_q) + \frac{b_q c_q}{b_p} \end{aligned} \tag{6}$$

4 Dirichlet

A Dirichlet density is given by

$$Dir(\boldsymbol{\pi}; \boldsymbol{\lambda}) = \frac{\Gamma(\sum_{s=1}^{m} \lambda_s)}{\prod_{s=1}^{m} \Gamma(\lambda_s)} \prod_{s=1}^{m} \pi_s^{\lambda_s - 1} \tag{7}$$

where λ_s is the sth element of $\boldsymbol{\lambda}$ and $\Gamma(x)$ is the Gamma function [2].

For the Dirichlet density with $q(\boldsymbol{\pi}) = Dir(\boldsymbol{\pi}; \boldsymbol{\lambda}_q)$ and $p(\boldsymbol{\pi}) = Dir(\boldsymbol{\pi}; \boldsymbol{\lambda}_p)$ where the number of states is m and the KL-divergence is

$$\begin{aligned} KL_{Dir}(\lambda_q; \lambda_p) &= \log \frac{\Gamma(\lambda_{qt})}{\Gamma(\lambda_{pt})} + \sum_{s=1}^{m} \log \frac{\Gamma(\lambda_p(s))}{\Gamma(\lambda_q(s))} \\ &+ \sum_{s=1}^{m} [\lambda_q(s) - \lambda_p(s)][\Psi(\lambda_q(s)) - \Psi(\lambda_{qt})] \end{aligned} \tag{8}$$

where

$$\begin{aligned} \lambda_{qt} &= \sum_{s=1}^{m} \lambda_q(s) \\ \lambda_{pt} &= \sum_{s=1}^{m} \lambda_p(s) \end{aligned} \tag{9}$$

5 Wishart

The Wishart distribution is given by ([1], page 85)

$$\mathsf{Wi}(\boldsymbol{\Gamma}; a, \boldsymbol{B}) = \frac{1}{Z(a, \boldsymbol{B})} |\boldsymbol{\Gamma}|^{(a-d-1)/2} \exp\left[-\frac{1}{2} Tr(\boldsymbol{B}\boldsymbol{\Gamma})\right] \tag{10}$$

where

$$Z(a, \boldsymbol{B}) = 2^{ad/2} |\boldsymbol{B}|^{-a/2} \Gamma_d(a/2) \tag{11}$$

the generalised gamma function is given by

$$\Gamma_d(a/2) = \pi^{d(d-1)/4} \prod_{j=1}^{d} \Gamma((a - j + 1)/2) \tag{12}$$

which is valid for $a > d - 1$. The entropy and KL-divergence of a Wishart can be defined in terms of the integral

$$L(a, \boldsymbol{B}) = \int \mathsf{Wi}(\boldsymbol{\Gamma}; a, \boldsymbol{B}) \log |\boldsymbol{\Gamma}| d\boldsymbol{\Gamma} \tag{13}$$

$$
\begin{aligned}
&= \log \tilde{\mathbf{\Gamma}}(a, B) \\
&= \sum_{i=1}^{d} \Psi((a_s + 1 - i)/2) - \log |\boldsymbol{B}_s| + d \log 2
\end{aligned}
$$

The entropy of $q(\mathbf{\Gamma}) = \mathsf{Wi}(\mathbf{\Gamma}; q, \boldsymbol{Q})$ is then given by

$$
H_W(q, \boldsymbol{Q}) = -\left(\frac{q-d-1}{2}\right) L(q, \boldsymbol{Q}) + \frac{qd}{2} + \log Z(q, \boldsymbol{Q}) \tag{14}
$$

The KL-Divergence between densities $q(\mathbf{\Gamma}) = \mathsf{Wi}(\mathbf{\Gamma}; q, \boldsymbol{Q})$ and $p(\mathbf{\Gamma}) = \mathsf{Wi}(\mathbf{\Gamma}; p, \boldsymbol{P})$ is given by

$$
\begin{aligned}
KL_W(q, \boldsymbol{Q}; p, \boldsymbol{P}) &= \left(\frac{q-d-1}{2}\right) L(q, \boldsymbol{Q}) \qquad (15) \\
&\quad - \left(\frac{p-d-1}{2}\right) L(p, \boldsymbol{P}) - \frac{qd}{2} + \frac{q}{2} Tr(\boldsymbol{P}\boldsymbol{Q}^{-1}) + \log \frac{Z(p, \boldsymbol{P})}{Z(q, \boldsymbol{Q})}
\end{aligned}
$$

References

[1] R.J. Muirhead. *Aspects of Multivariate Statistical Theory.* John Wiley, 1982.

[2] W. H. Press, S.A. Teukolsky, W.T. Vetterling, and B.V.P. Flannery. *Numerical Recipes in C.* Cambridge, 1992.