A Note on Log-Concavity

Matthias Seeger
Max Planck Institute for Biological Cybernetics
P.O. Box: 21 69, 72012 Tuebingen, Germany
seeger@tuebingen.mpg.de

May 14, 2007

Abstract

This is a small observation concerning scale mixtures and their log-concavity.

A function $f(\boldsymbol{x}) \geq 0,\ \boldsymbol{x} \in \mathbb{R}^n$ is called *log-concave* if

$$f\left(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}\right) \geq f(\boldsymbol{x})^{\lambda} f(\boldsymbol{y})^{1-\lambda} \tag{1}$$

for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n,\ \lambda \in [0,1]$. Log-concavity is important in applied Bayesian Statistics, since a distribution with a log-concave density is easy to treat with many different approximate inference techniques. For example, log-concavity implies unimodality. Log-concave distributions over few variables can be sampled from using a generic Markov chain Monte Carlo technique called adaptive rejection sampling [3]. For certain approximate inference techniques such as expectation propagation [5, 6], log-concavity of all sites means that the algorithm can be implemented in a numerically stable manner and tends to converge quickly, while in the absense of log-concavity it can fail badly.

Many well-known densities are log-concave, for example the Gaussian or the Gamma $\propto x^{a-1}e^{-bx}\mathrm{I}_{\{x>0\}}$, the latter for $a \geq 1$. In the Bayesian context it is important to note that exponential family densities are in general log-concave in their natural parameters (but not necessarily in their data argument).

A very important result concerning log-concave functions has been given by Prékopa (see [1], Sect. 1.8). Namely, if $f : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \to \mathbb{R}$ is (jointly) log-concave, so is $g(\boldsymbol{x}) = \int f(\boldsymbol{x}, \boldsymbol{y}) d\boldsymbol{y}$, where the integral is over all of $\mathbb{R}^{n_2}$.

In this note, we are interested in scale mixture distributions [4] $P(x) = \mathrm{E}[N(x|h,s)]$ with some mixing distribution $P(s)$ over the variance. We give a sufficient condition for $P(x)$ to be log-concave, and show that the condition is not necessary. Since concavity is preserved under linear transformations, we can assume w.l.o.g. that $h = 0$.

We have that

$$P(x) = \int \exp\left(-(1/2)x^2/s - (1/2)\log s + \log P(s) + C\right)\, ds.$$

First, $(x,s) \mapsto x^2/s$ is convex on $\mathbb{R} \times \mathbb{R}_{>0}$ (see [2], Sect. 3.1.5). Therefore, if $\log P(s) - (1/2)\log s$ is concave there, we see that the function

$$g(x,s) = P(s)s^{-1/2}e^{-(1/2)x^2/s}\mathrm{I}_{\{s>0\}}$$

is log-concave on $\mathbb{R}^2$. To see that, invoke Eq. 1 for $\boldsymbol{x} = (x, s_x)$, $\boldsymbol{y} = (y, s_y)$. If $s_x \le 0$ or $s_y \le 0$, the r.h.s. is zero. Otherwise, $\lambda s_x + (1-\lambda)s_y > 0$, and the inequality follows from the concavity of $\log g$.

We have shown that if $P(s)s^{-1/2}$ is log-concave on $\mathbb{R}$, then the scale mixture $P(x) = \mathrm{E}[N(x|h, s)]$ is log-concave as well.

This condition is not necessary. The Laplace density $P(x) = \tau e^{-\tau|x|}$ is a scale mixture under the exponential $P(s) = \lambda e^{-\lambda s}\mathrm{I}_{\{s>0\}}$, with $\lambda = \tau^2/2$ [4, 7]. However, $\log P(s) - (1/2)\log s = -\lambda s - (1/2)\log s$ is strictly convex for $s > 0$.

References

[1] V. Bogachev. *Gaussian Measures.* Mathematical Surveys and Monographs. American Mathematical Society, 1998.

[2] S. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge University Press, 2002. Available online at `www.stanford.edu/~boyd/cvxbook.html`.

[3] W. R. Gilks and P. Wild. Adaptive rejection sampling for Gibbs sampling. *Applied Statistics*, 41(2):337–348, 1992.

[4] T. Gneiting. Normal scale mixtures and dual probability densities. *J. Statist. Comput. Simul.*, 59:375–384, 1997.

[5] Thomas Minka. Expectation propagation for approximate Bayesian inference. In J. Breese and D. Koller, editors, *Uncertainty in Artificial Intelligence 17.* Morgan Kaufmann, 2001.

[6] Manfred Opper and Ole Winther. Gaussian processes for classification: Mean field algorithms. *Neural Computation*, 12(11):2655–2684, 2000.

[7] T. Park and G. Casella. The Bayesian Lasso. Technical report, University of Florida, 2005.