The Laplace Transform with Applications in Machine Learning

Zhang Zhihua

July 17, 2013

Outline

Backgrounds

- The Laplace transform is a classical mathematical theory and has also received wide applications in science and engineering. The Laplace transform is named after mathematician and astronomer Pierre-Simon Laplace, who used a similar transform (now called z transform) in his work on probability theory. The current widespread use of the transform came about soon after World War II although it had been used in the 19th century by Abel, Lerch, Heaviside, and Bromwich.

Backgrounds

- The Laplace transform is essentially an integral transform. In probability theory and statistics, the Laplace transform is defined as expectation of a random variable.
- This lecture will provide an detailed illustration for the application of the Laplace transform in machine learning. We will see that the Laplace transform helps us establish a bridge of connecting some machine learning methods from different approaches.

The Laplace Transform

- Suppose that f is a real-or complex-valued function of the variable $t \geq 0$. The Laplace transform of function $f(t)$ is defined by

$$\phi(s) = \int_0^{\infty} \exp(-ts) f(t) dt \triangleq \lim_{z \to \infty} \int_0^{z} \exp(-ts) f(t) dt$$

whenever the limit exists. If the limit does not exist, the integral is said to diverge and there is no Laplace transform defined for f. Here s is a real or complex parameter. However, we always assume that f is a real function and s is a real nonnegative number. Usually, the above equation is also called the Laplace integral.

The Laplace Transform

- Suppose that $\{a_n\}$ is a sequence of real numbers. The Laplace transform is a Dirichlet series of the form

$$\phi(s) = \sum_{n=1}^{\infty} a_n \exp(-t_n s) = \lim_{N \to \infty} \sum_{n=1}^{N} a_n \exp(-t_n s),$$

  where the discrete set $\{t_n\}$ of exponents corresponds to the continuous variable t in the integral.

- Both the Laplace integral and the Dirichlet series can be included in the Laplace-Stieltjes integral:

$$\phi(s) = \int_0^{\infty} \exp(-st) dF(t).$$

  If $F(t)$ is absolutely continuous, it becomes the Laplace integral; if $F(t)$ is a step-function, it is the Dirichlet series.

The Laplace Transform

- We are especially interested in that $F(x)$ is a probability measure concentrated on $[0, \infty)$. In this case, if X is a random variable with probability distribution F, then the Laplace transform is given by the expectation; namely,

$$\phi(s) = \mathbb{E}(\exp(-sX)).$$

By abuse of language, we also speak of "the Laplace transform of the random variable X." Note that replacing s by $-s$ gives the moment generating function of X.

Example 1

- We consider the Laplace transform of a generalized inverse Gaussian (GIG) distribution. The density of the GIG distribution is defined as

$$p(X = x) = \frac{(\alpha/\beta)^{\gamma/2}}{2K_\gamma(\sqrt{\alpha\beta})} x^{\gamma-1} \exp(-(\alpha x + \beta x^{-1})/2), \ x > 0,$$

  where $K_\gamma(\cdot)$ represents the modified Bessel function of the second kind with the index γ. We denote this distribution by $\mathrm{GIG}(\gamma, \beta, \alpha)$. It is well known that its special cases include the gamma distribution $\mathrm{Ga}(\gamma, \alpha/2)$ when $\beta = 0$ and $\gamma > 0$, the inverse gamma distribution $\mathrm{IG}(-\gamma, \beta/2)$ when $\alpha = 0$ and $\gamma < 0$, the inverse Gaussian distribution when $\gamma = -1/2$, and the hyperbolic distribution when $\gamma = 0$.

Example 1

- It is direct to obtain the Laplace transform of $\mathrm{GIG}(\gamma, \beta, \alpha)$ as

$$\phi(s) = \frac{K_\gamma(\sqrt{(\alpha+2s)\beta})}{K_\gamma(\sqrt{\alpha\beta})}(\alpha/(\alpha+2s))^{\gamma/2}, \quad s \geq 0. \qquad (1)$$

Especially, the Laplace transform of the inverse Gaussian distribution (i.e., $\gamma = -1/2$) is

$$\phi(s) = \exp(-\sqrt{(\alpha+2s)\beta} + \sqrt{\alpha\beta})$$

due to $K_{1/2}(z) = K_{-1/2}(z) = \sqrt{\frac{\pi}{2z}}\exp(-z)$.

Example 1

- The Laplace transform of gamma distribution $\mathrm{Ga}(\gamma, 2/\alpha)$ with shape γ and scale $2/\alpha$ (i.e., $\beta = 0$) is

$$\phi(s) = \left[\frac{\alpha}{\alpha+2s}\right]^{\gamma},$$

which can be also followed from (1) by using the fact $\lim_{z \to 0} \frac{2K_{\gamma}(z)}{\Gamma(\gamma)} \left[\frac{z}{2}\right]^{\gamma} = 1$ for $\gamma > 0$.

Example 2

- Let X be α-stable distribution of the density

$$p(X = x) = \frac{1}{\pi x} \sum_{k=1}^{\infty} \frac{\Gamma(1+k\alpha)}{k!} (-1)^{k+1} x^{-k\alpha} \sin(k\alpha\pi), \quad x > 0$$

for $0 < \alpha < 1$. The α-stable distribution has Laplace transform

$$\phi(s) = \exp(-s^{\alpha}), \quad s \geq 0.$$

Example 2

- Note that the $\frac{1}{2}$-stable distribution becomes Lévy distribution $\mathrm{Lv}(0, 1/2)$ where $\mathrm{Lv}(u, \sigma^2)$ represents that x follows an Lévy distribution of the form

$$p(x) = \frac{\sigma}{\sqrt{2\pi}}(x-u)^{-3/2} \exp\left(-\frac{\sigma^2}{2(x-u)}\right), \text{ for } x \geq u.$$

  We can obtain that the Laplace transform of $\mathrm{Lv}(0, \sigma^2)$ is $\exp(-\sigma(2s)^{1/2})$. This shows that $\exp(-s^{1/2})$ is the Laplace transform of Lévy distribution $\mathrm{Lv}(0, 1/2)$.

Example 3

We see several discrete random variables.

- First, let X be a Poisson distribution of intensity λ. That is,

$$\Pr(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \text{ for } k = 0, 1, 2, \ldots$$

The transform of Poisson distribution $\mathrm{Po}(\lambda)$ is $\exp[-\lambda(1 - \exp(-s))]$.

Example 4

- Second, we let random variable X follow negative binomial distribution $\mathrm{NB}(\xi, q)$ with $\xi > 0$ and $0 < q < 1$. That is, the probability mass function of X is

$$\Pr(X = k) = \frac{\Gamma(k+\xi)}{k!\Gamma(\xi)} q^{\xi}(1-q)^{k}. \tag{2}$$

The Laplace transform is

$$\phi(s) = q^{\xi}[1-(1-q)\exp(-s)]^{-\xi}, \quad s \geq 0.$$

It is worth noting that for any fixed nonzero constant α, αX also follows negative binomial distribution $\mathrm{NB}(\xi, q)$ but it takes values on $\{k\alpha; k = 0, 1, \ldots\}$. In this case, the Laplace transform of αX is

$$\phi(s) = q^{\xi}[1-(1-q)\exp(-s\alpha)]^{-\xi}, \quad s \geq 0. \tag{3}$$

Uniqueness

Theorem (Uniqueness)

Distinct probability distributions have distinct Laplace transforms.

The uniqueness is only with respect to a probability measure. For any function, the theorem would be no longer satisfied.

Continuity

Theorem

Let F_n be a probability distribution with Laplace transform ϕ_n, for $n = 1, 2, \ldots$
If $F_n \to F$ where F is a possibly defective distribution with Laplace transform ϕ, then $\phi_n(s) \to \phi(s)$ for $s > 0$.
Conversely, if the sequence $\{\phi_n(s)\}$ converges for each $s > 0$ to a limit $\phi(s)$, then ϕ is the Laplace transform of some possibly defective distribution F such that $F_n \to F$. Furthermore, the limit F is not defective iff $\phi(s) \to 1$ as $s \to 0$.

Example 1'

Let us return to Example 1. As we see earlier, $\mathrm{GIG}(\gamma, \beta, \alpha)$ converges to $\mathrm{Ga}(\gamma, 1/(2\alpha))$ as $\beta \to 0$. Theorem 2 implies that

$$\lim_{\beta \to 0} \frac{K_\gamma(\sqrt{(\alpha + 2s)\beta})}{K_\gamma(\sqrt{\alpha\beta})} (\alpha/(\alpha + 2s))^{\gamma/2} = \Big[\frac{\alpha}{\alpha + 2s}\Big]^\gamma.$$

In fact, the limit can also obtained from that $K_\gamma(z) \sim \frac{\Gamma(\gamma)}{2}(\frac{2}{z})^\gamma$ for $\gamma > 0$ as $z \to 0$. Additionally, recall that

$$\lim_{\alpha \to 0} \exp(-\sqrt{(\alpha + 2s)\beta} + \sqrt{\alpha\beta}) = \exp(-\sqrt{2s\beta}),$$

which implies that inverse Gaussian distribution $\mathrm{GIG}(-1/2, \beta, \alpha)$ converges to Lévy distribution $\mathrm{Lv}(0, \beta)$ as $\alpha \to 0$.

Example 3'

We now see random variable αX in Example 3. Let $q = \frac{\xi}{\xi+\lambda}$ and $\alpha = \frac{\xi}{\xi+1}$. Taking the limit of $\xi \to \infty$, we have

$$\lim_{\xi \to \infty} q^{\xi}[1 - (1-q)\exp(-\alpha s)]^{-\xi} = \exp[-\lambda(1 - \exp(-s))].$$

Theorem 2 then shows that $\lim_{\xi \to \infty} \mathrm{NB}(\xi, \xi/(\xi+\lambda)) = \mathrm{Po}(\lambda)$. On the other hand, let $\alpha = q$. For fixed ξ, we have

$$\lim_{\alpha \to 0} q^{\xi}[1 - (1-q)\exp(-\alpha s)]^{-\xi} = \left[\frac{1}{1+s}\right]^{\xi}.$$

We then know that αX converges in distribution to a gamma distribution with shape ξ and scale 1. That is, for any $u > 0$

$$\lim_{\alpha \to 0} \Pr(\alpha X \leq u) = \int_0^u \frac{1}{\Gamma(\xi)} t^{\xi-1} e^{-t} dt.$$

Remarks

There are other types of transforms related to the Laplace transform. For example, the Stieltjes transform is

$$g(s) = \int_0^\infty \frac{1}{s+x} dF(x), \quad s > 0. \tag{4}$$

Note that $\frac{1}{s+x} = \int_0^\infty e^{-su} e^{-xu} du$. Using the Fubini theorem, we have

$$g(s) = \int_0^\infty e^{-su} \int_0^\infty e^{-xu} dF(x) du,$$

which is a double Laplace transform.

Definition

An important application of the Laplace transform lies in the definition of a completely monotone function. Let $f \in C^\infty(0, \infty)$ with $f \geq 0$. We say f is completely monotone if $(-1)^n f^{(n)} \geq 0$ for all $n \in \mathbb{N}$, and a Bernstein function if $(-1)^n f^{(n)} \leq 0$ for all $n \in \mathbb{N}$. Thus, a nonnegative function $f : (0, \infty) \to \mathbb{R}$ is a Bernstein function iff f' is a completely monotone function.

Properties

- For any two completely monotone (or Bernstein) functions $f_1(x)$ and $f_2(x)$, it is clearly seen that $a_1 f_1(x) + a_2 f_2(x)$ where a_1 and a_2 are nonnegative constants is also completely monotone (or Bernstein).
- Moreover, $f_1(x)f_2(x)$ are also completely monotone when $f_1(x)$ and $f_2(x)$ are completely monotone, while $f_1(f_2(x))$ is Bernstein when when $f_1(x)$ and $f_2(x)$ are Bernstein.
- However, $f_1(f_2(x))$ is completely monotone if f_1 is completely monotone and f_2 is nonnegative with a completely monotone derivative.

Properties

Theorem

Let ϕ and ψ be functions from $(0, \infty)$ to $\mathbb{R}$, and $(x \wedge y)$ denote $\min(x, y)$.

(1) A necessary and sufficient condition that ϕ is completely monotone is that

$$\phi(s) = \int_0^\infty e^{-su} dF(u),$$

where $F(u)$ is non-decreasing and the integral converges for $s \in (0, \infty)$. Furthermore, $F(u)$ is a probability distribution, iff $\phi(0) = 1$.

(2) ψ is a Bernstein function iff the mapping $s \to \exp(-t\psi(s))$ is completely monotone for all $t \geq 0$.

Properties

Theorem

Let ϕ and ψ be functions from $(0, \infty)$ to $\mathbb{R}$, and $(x \wedge y)$ denote $\min(x, y)$*.*

(3) ψ is a Bernstein function iff it admits the representation of

$$\psi(s) = a + bs + \int_{(0,\infty)} (1 - \exp(-su))\nu(du), \tag{5}$$

where $a, b \geq 0$ and ν is a Lévy measure on $(0, \infty)$ such that $\int_{(0,\infty)} (1 \wedge u)\nu(du) < \infty$. Furthermore, the triplet (a, b, ν) uniquely determines ψ.

Remarks

- The representation in (5) is also well known as the Lévy-Khintchine formula. This formula implies that $\lim_{s\to 0+} \psi(s) = a$ and $\lim_{s\to\infty} \frac{\psi(s)}{s} = \lim_{s\to\infty} \psi'(s) = b$. Let ν be a Borel measure defined on $\mathbb{R}^d - \{\mathbf{0}\} = \{\mathbf{x} \in \mathbb{R}^d : \mathbf{x} \neq \mathbf{0}\}$. We call ν a Lévy measure if

$$\int_{\mathbb{R}^d - \{\mathbf{0}\}} (\|\mathbf{z}\|_2^2 \wedge 1)\nu(d\mathbf{z}) < \infty. \tag{6}$$

Example 5

- The following elementary functions are completely monotone, namely,

$$\frac{1}{(a+s)^{\nu}}, \; e^{-bs}, \; \text{and} \; \ln(1+\frac{c}{s}),$$

  where $a \geq 0$, $b \geq 0$, $\nu \geq 0$ and $c > 0$.

- The following functions are then Bernstein, which are

$$s^{\alpha}, \; 1-\exp(-\beta s), \; \text{and} \; \log(1+\gamma s),$$

  where $\alpha \in (0,1]$, $\beta > 0$, $\gamma > 0$. Consider that

$$\frac{s}{1+s} = 1-\exp(-\log(1+s)),$$

  which is also Bernstein based on the composition property.

Example 6

- We now have from Theorem 4 that e^{-bs} can be represented as the Laplace transform of some probability distribution, but so cannot $\ln(1 + \frac{c}{s})$. Moreover, the functions ϕ given in Examples 1, 2 and 3 are completely monotone.

Example 7

- It is immediate that $\psi(s) = s^\alpha$ for $\alpha \in (0, 1)$ is a Bernstein function of s in $(0, \infty)$. Moreover, it is also directly verified that
$$s^\alpha = \frac{\alpha}{\Gamma(1-\alpha)} \int_0^\infty (1 - e^{-su}) \frac{dx}{u^{1+\alpha}},$$
which implies that
$$\nu(du) = \frac{\alpha}{\Gamma(1-\alpha)} \frac{du}{u^{1+\alpha}}.$$
Obviously, $\psi(s) = s$ is also Bernstein. It is an extreme case, because we have that $a = 0$, $b = 1$ and $\nu(du) = \delta_0(u)du$ (the Dirac Delta measure).

Example 8

We consider the following function:

$$\psi_\alpha(s) = \begin{cases} \log(1+\gamma s) & \alpha = 0, \\ \frac{1}{\alpha}\left[1 - \left(1+(1-\alpha)\gamma s\right)^{-\frac{\alpha}{1-\alpha}}\right] & \alpha \in (0,1), \\ 1 - \exp(-\gamma s) & \alpha = 1 \end{cases}$$

for $\gamma > 0$. This function is Bernstein for any $\alpha \in [0,1]$. In this case, we have that $a = 0$ and $b = 0$. The corresponding Lévy measure is

$$\nu(du) = \begin{cases} \frac{1}{u}\exp(-\frac{u}{\gamma})du & \alpha = 0, \\ \frac{\gamma[(1-\alpha)\gamma]^{-\frac{1}{1-\alpha}}}{\Gamma(1/(1-\alpha))} u^{\frac{\alpha}{1-\alpha}-1}\exp\left(-\frac{1}{(1-\alpha)\gamma}u\right)du & \alpha \in (0,1), \\ \delta_\gamma(u)du & \alpha = 1. \end{cases}$$

Introduction

- Machine learning methods based on the reproducing kernel theory have received successful applications, and they become a classical tool in machine learning and data mining. There is an equivalent relationship between a reproducing kernel Hilbert space and a positive definite kernel. We here exploit completely monotone functions in positive and negative definite kernels.

Definition

Definition

Let $\mathcal{X} \subset \mathbb{R}^d$ be a nonempty set. A function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is said to be symmetric if $k(\mathbf{x}, \mathbf{y}) = k(\mathbf{y}, \mathbf{x})$ for all $\mathbf{x}$ and $\mathbf{y} \in \mathcal{X}$. A symmetric function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is called a positive definite kernel iff

$$\sum_{i,j=1}^{m} a_i a_j k(\mathbf{x}_i, \mathbf{x}_j) \geq 0$$

for all $m \in \mathbb{N}$, $\{\mathbf{x}_1, \ldots, \mathbf{x}_m\} \subset \mathcal{X}$ and $\{a_1, \ldots, a_m\} \subset \mathbb{R}$.

Definition

Definition

Let $\mathcal{X} \subset \mathbb{R}^d$ be a nonempty set. A function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is said to be symmetric if $k(\mathbf{x}, \mathbf{y}) = k(\mathbf{y}, \mathbf{x})$ for all $\mathbf{x}$ and $\mathbf{y} \in \mathcal{X}$. A symmetric function k is called a conditionally positive kernel iff

$$\sum_{i,j=1}^{m} a_i a_j k(\mathbf{x}_i, \mathbf{x}_j) \geq 0$$

for all $m \geq 2$, $\{\mathbf{x}_1, \ldots, \mathbf{x}_m\} \subset \mathcal{X}$ and $\{a_1, \ldots, a_m\} \subset \mathbb{R}$ with $\sum_{j=1}^{m} a_j = 0$.

Remarks

- It is worth pointing out that if $\mathcal{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$ is a finite set, then k is positive definite iff the $m \times m$ matrix $\mathbf{K} = [k(\mathbf{x}_i, \mathbf{x}_j)]$ is positive semidefinite.
- We call a kernel k *negative definite* if $-k$ is conditionally positive definite.

Properties

Theorem

Let μ be a probability measure on the half-line $\mathbb{R}_+$ such that $0 < \int_0^\infty u d\mu(u) < \infty$. Then $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+$ is negative definite iff the Laplace transform of tk (i.e., $\int_0^\infty \exp(-utk) d\mu(u)$) is positive definite for all $t > 0$.

Taking $\mu(du) = \delta_1(u)du$, we obtain the following theorem, which illustrates a connection between positive and negative definite kernels.

Theorem (Schoenberg correspondence)

For function $k : \mathbb{R}^d \to \mathbb{R}$, it is negative definite iff $\exp(-tk)$ is positive definite for all $t > 0$.

Definition

Especially, we are interested in the definition that $k(\mathbf{x}_i, \mathbf{x}_j) = \psi(\mathbf{x}_i - \mathbf{x}_j)$. In this case, we say that $\psi : \mathcal{X} \to \mathbb{R}$ is positive definite if kernel $k(\mathbf{x}_i, \mathbf{x}_j) \triangleq \psi(\mathbf{x}_i - \mathbf{x}_j)$ is positive definite, and ψ is conditionally positive definite if $k(\mathbf{x}_i, \mathbf{x}_j) \triangleq \psi(\mathbf{x}_i - \mathbf{x}_j)$ is conditionally positive definite. In this regard, $\psi : \mathcal{X} \to \mathbb{R}$ is said to be symmetric if $\psi(-\mathbf{x}) = \psi(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{X}$. Moreover, we always have $\psi(\mathbf{0}) = \psi(\mathbf{x} - \mathbf{x}) \geq 0$ if ψ is positive definite.

Example 9

Consider the function $\psi(\mathbf{x}) = \|\mathbf{x}\|_2^2$. Let $\mathbf{x}_1, \ldots, \mathbf{x}_m \in \mathbb{R}^d$, and let $\mathbf{A} = [a_{ij}]$ be the $m \times m$ matrix with $a_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|_2^2$. Given any real numbers $\beta_1, \ldots, \beta_m$ such that $\sum_{j=1}^m \beta_j = 0$, we have

$$\sum_{i,j=1}^m \beta_i\beta_j\|\mathbf{x}_i - \mathbf{x}_j\|_2^2 = \sum_{i,j=1}^m \beta_i\beta_j(\|\mathbf{x}_i\|_2^2 + \|\mathbf{x}_j\|_2^2 - 2\mathbf{x}_i^T\mathbf{x}_j)$$
$$= -2\sum_{i,j=1}^m (\beta_i\mathbf{x}_i)^T(\beta_j\mathbf{x}_j) = -2\left\|\sum_{i=1}^m \beta_i\mathbf{x}_i\right\|_2^2 \le 0,$$

which implies that $\|\mathbf{x} - \mathbf{y}\|_2^2$ is negative definite. It then follows from Theorem 8 that $\exp(-t\|\mathbf{x} - \mathbf{y}\|_2^2)$ for all $t > 0$ is positive definite; that is, the Gaussian RBF kernel is positive definite.

Example 10

Consider the function $\psi(\mathbf{x}) = \|\mathbf{x}\|_1 = \sum_{j=1}^{p} |x_j|$. Note that

$$\exp(-t|x|) = \int_0^\infty \exp(-t^2 u|x|^2) \mathrm{Lv}(u|0, 1/2) du \qquad (7)$$

for all $t > 0$. It then follows from the negative definiteness of $|x - y|^2$ and Theorem 7 that $\exp(-t|x - y|)$ is positive definite for all $t > 0$. Again using Theorem 8, we obtain that $|x - y|$ is negative definite. Accordingly, we have that $\|\mathbf{x} - \mathbf{y}\|_1$ $(= \sum_{j=1}^{p} |x_j - y_j|)$ is negative definite and that $\exp(-t\|\mathbf{x} - \mathbf{y}\|_1)$ for all $t > 0$ is positive definite.

Distance Functions

It is well known that a positive definite kernel can induce a kernel matrix, which is applied to nonlinear machine learning methods such as kernel principal component analysis, generalized Fisher discriminant analysis, spectral clustering. We now show that a negative definite kernel can then define an Euclidean distance matrix, which has been used in multidimensional scaling (MDS).

Distance Matrices

Definition

An $m \times m$ matrix $\mathbf{D} = [d_{rs}]$ is called a distance matrix if $d_{rs} = d_{sr}$, $d_{rr} = 0$ and $d_{rs} \geq 0$. An $m \times m$ distance matrix $\mathbf{D} = [d_{rs}]$ is said to be Euclidean if there exists a configuration of points in some Euclidean space whose interpoint distances are given by $\mathbf{D}$; that is, if for some $p \in \mathbb{N}$, there exist points $\mathbf{z}_1, \ldots, \mathbf{z}_m \in \mathbb{R}^p$ such that $d_{rs}^2 = \|\mathbf{z}_r - \mathbf{z}_s\|_2^2 = (\mathbf{z}_r - \mathbf{z}_s)^T (\mathbf{z}_r - \mathbf{z}_s)$.

Main Result

Theorem

Let $\mathbf{D} = [d_{rs}]$ be an $m \times m$ distance matrix and define $\mathbf{B} = \mathbf{H}_m \mathbf{A} \mathbf{H}_m$ where $\mathbf{H}_m = \mathbf{I}_m - \frac{1}{m}\mathbf{1}_m \mathbf{1}_m^T$ (the centering matrix) and $\mathbf{A} = [a_{rs}]$ with $a_{rs} = -\frac{1}{2} d_{rs}^2$. Then $\mathbf{D}$ is Euclidean iff $\mathbf{B}$ is positive semidefinite.

Clearly, if $\mathbf{A}$ is conditionally positive semidefinite, then $\mathbf{H}_m \mathbf{A} \mathbf{H}_m$ is positive semidefinite. Thus, we directly have the following corollary.

Corollary

Let $g(s)$ be Bernstein on $(0, \infty)$ with $g(0+) = 0$. Then for any vector $\mathbf{x}_1, \ldots, \mathbf{x}_m \in \mathbb{R}^d$, the matrix $\left[\sqrt{g(\|\mathbf{x}_i - \mathbf{x}_j\|_\alpha^\alpha)}\right]$ for any $\alpha \in \{1, 2\}$ is an Euclidean distance matrix.

Examples 11

The functions $\sqrt{1+s}-1$, $\log(1+s)$ and $1-\exp(-s)$ are Bernstein on $(0,\infty)$ and equal to zero at the origin. Thus, for any distinct vectors $\mathbf{x}_1,\ldots,\mathbf{x}_m \in \mathbb{R}^d$, both the matrices $\left[\sqrt{1+\|\mathbf{x}_i-\mathbf{x}_j\|_1}-1\right]$ and $\left[\sqrt{1+\|\mathbf{x}_i-\mathbf{x}_j\|_2^2}-1\right]$, $\left[\log(1+\|\mathbf{x}_i-\mathbf{x}_j\|_1)\right]$ and $\left[\log(1+\|\mathbf{x}_i-\mathbf{x}_j\|_2^2)\right]$, and $\left[1-\exp(-\|\mathbf{x}_i-\mathbf{x}_j\|_1)\right]$ and $\left[1-\exp(-\|\mathbf{x}_i-\mathbf{x}_j\|_2^2)\right]$ are negative semidefinite and nonsingular. Denote these matrices by $\mathbf{A}$. Then $-\mathbf{H}_m\mathbf{A}\mathbf{H}_m$ is an Euclidean distance matrix. Furthermore, the rank of $-\mathbf{H}_m\mathbf{A}\mathbf{H}_m$ is $m-1$. This implies that $p=m-1$ (the dimension of the Euclidean space).

Introduction

We consider the application of completely monotone functions in scale mixture distributions, which in turn have wide use in applied and computational statistics. We particularly study scale mixtures of exponential power (EP) distributions.

Exponential Power (EP) Distributions

A univariate random variable $X \in \mathbb{R}$ is said to follow an EP distribution if the density is specified by

$$p(X=x) = \frac{\eta^{-1/q}}{2^{\frac{q+1}{q}}\Gamma(\frac{q+1}{q})}\exp(-\frac{1}{2\eta}|x-u|^q) = \frac{q}{2}\frac{(2\eta)^{-\frac{1}{q}}}{\Gamma(\frac{1}{q})}\exp(-\frac{1}{2\eta}|x-u|^q),$$

with $\eta > 0$. It is typically assumed that $0 < q \leq 2$. The distribution is denoted by $\mathrm{EP}(x|u,\eta,q)$. There are two classical special cases: the Gaussian distribution arises when $q=2$ (denoted $N(x|u,\eta)$) and the Laplace distribution arises when $q=1$ (denoted $L(x|u,\eta)$). As for the case that $0<q<2$, the corresponding density is called a bridge distribution. Moreover, we will only consider the setting that $u=0$.

Scale Mixture of EP Distributions

Let $\tau = 1/(2\eta)$ be a random variable with probability density $p(\tau)$. The marginal distribution of x is

$$f_X(x) = \int_0^\infty \mathrm{EP}(b|0, 1/(2\tau), q)p(\tau)d\tau.$$

We call the resulting distribution a scale mixture of exponential power distributions. This distribution is symmetric about 0 and can be denoted as $\phi(|x|^q)$. Especially, when $q = 2$, it is a scale mixture of normal distributions. When $q = 1$, it is called a scale mixture of Laplace distributions.

Scale Mixture of EP Distributions

Let $s = |x|^q$. Then we can write $f_X(s)$ as follows

$$\int_0^{\infty} \frac{\tau^{1/q}}{2\Gamma(\frac{q+1}{q})} \exp(-\tau s) p(\tau) d\tau, \tag{8}$$

which we now denote by $\phi(s)$. Clearly, $\phi(s)$ is the Laplace transform of $\frac{\tau^{1/q}}{2\Gamma(\frac{q+1}{q})} p(\tau)$. This implies that $\phi(s)$ is completely monotone on $(0, \infty)$.

Main Result

Theorem

Let X have a density function $f_X(x)$, which is a function of $|x|^q$ with $0 < q \leq 2$ (denoted $\phi(|x|^q)$). Then $\phi(|x|^q)$ can be represented as a scale mixture of exponential power distributions iff $\phi(s)$ is completely monotone on $(0, \infty)$.

The Laplace Distribution

The marginal density of b is thus defined by

$$\frac{\sqrt{\alpha}}{2}\exp(-\sqrt{\alpha}|b|) = \int_0^{+\infty} N(b|0,\eta)\mathrm{Ga}(\eta|\gamma,\alpha/2)d\eta.$$

The Bridge Distribution

The marginal density of b is thus defined by

$$\frac{\alpha}{4}\exp(-\sqrt{\alpha|b|}) = \int_0^{+\infty} L(b|0,\eta)\mathrm{Ga}(\eta|3/2,\alpha/2)d\eta.$$

Generalized t distributions

We t consider a scale mixtures of exponential power $\mathrm{EP}(b|u, \eta, q)$ with inverse gamma $\mathrm{IG}(\eta|\tau/2, \tau/(2\lambda))$. They are called *generalized t distributions*. The density is

$$
\begin{aligned}
&\frac{q}{2} \frac{\Gamma(\frac{\tau}{2}+\frac{1}{q})}{\Gamma(\frac{\tau}{2})\Gamma(\frac{1}{q})} \Big(\frac{\lambda}{\tau}\Big)^{\frac{1}{q}} \Big(1 + \frac{\lambda}{\tau}|b|^q\Big)^{-(\frac{\tau}{2}+\frac{1}{q})} \\
&= \int \mathrm{EP}(b|0, \eta, q)\mathrm{IG}(\eta|\tau/2, \tau/(2\lambda))d\eta,
\end{aligned}
$$

where $\tau > 0$, $\lambda > 0$ and $q > 0$. Clearly, when $q = 2$ the generalized t distribution becomes to a t-distribution. Moreover, when $\tau = 1$, it is the Cauchy distribution.

Generalized Double Pareto Distributions

When $q = 1$, it is also called called the resulting distributions *generalized double Pareto distributions*. The densities are given as follows:

$$\frac{\lambda}{4}\Big(1+\frac{\lambda|b|}{\tau}\Big)^{-(\tau/2+1)} = \int_0^\infty L(b|0,\eta)\mathrm{IG}(\eta|\tau/2,\tau/(2\lambda))d\eta,\ \lambda > 0, \tau > 0.$$

Furthermore, consider $\tau = 1$, such that $\eta \sim \mathrm{IG}(\eta|1/2, 1/(2\lambda))$. We obtain

$$p(b) = \frac{\lambda}{4}(1 + \lambda|b|)^{-3/2}.$$

EP-GIG

We now develop a family of distributions by mixing the exponential power $\mathrm{EP}(b|0, \eta, q)$ with the generalized inverse Gaussian $\mathrm{GIG}(\eta|\gamma, \beta, \alpha)$. The marginal density of b is thus defined by

$$p(b) = \int_0^{+\infty} \mathrm{EP}(b|0, \eta, q)\mathrm{GIG}(\eta|\gamma, \beta, \alpha) d\eta.$$

We refer to this distribution as the EP-GIG. The density is

$$p(b) = \frac{K_{\frac{\gamma q - 1}{q}}(\sqrt{\alpha(\beta + |b|^q)})}{2^{\frac{q+1}{q}} \Gamma(\frac{q+1}{q}) K_\gamma(\sqrt{\alpha\beta})} \frac{\alpha^{1/(2q)}}{\beta^{\gamma/2}} [\beta + |b|^q]^{(\gamma q - 1)/(2q)}.$$

Thanks

Questions & Comments!