Statistics and Probability Letters 104 (2015) 141–145

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Spherically symmetric multivariate beta family kernels

Tarn Duong

Sorbonne Paris City, University Paris-North — Paris 13, Computer Science Laboratory (LIPN), CNRS UMR 7030, F-93430, Villetaneuse, France

ARTICLE INFO

Article history:
Received 30 April 2015
Received in revised form 13 May 2015
Accepted 13 May 2015
Available online 28 May 2015

Keywords:
Bounded support
Integral on unit ball
Optimal kernel
Polynomial family
Quadratic form

ABSTRACT

The generalisation of univariate beta kernels to the multivariate spherically symmetric case is considered. By integrating the powers of quadratic forms over the unit ball, we exhibit closed form expressions, based on ratios of beta functions, for analysing these kernels.

1. Spherically symmetric kernels

For kernel estimation methods, the kernel function is usually a symmetric probability density function. A common choice is the family of kernels based on the beta density. The rth beta kernel is, for $r \geq 0$,

$$K(x; r) = c_r(1 - x^2)^r \mathbf{1}\{x \in [-1, 1]\}.$$

The normalising constant c_r ensures that the integral of the kernel is one, $c_r = (2r + 1)!/[2^{2r+1}(r!)^2] = 1/B(r + 1, 1/2)$ where $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$ is the beta function, e.g. from Marron and Nolan (1988). For $r = 0, 1, 2, 3$, the resulting kernels are known as the uniform, Epanechnikov (or quadratic), biweight (or quartic) and triweight kernels respectively. The popularity of this family stems from their desirable mathematical properties. The uniform kernel can be considered the simplest kernel, the Epanechnikov kernel is optimal in a mean integrated squared error (MISE) sense (Epanechnikov, 1969), and the biweight and triweight kernels possess the minimal integrated squared gradient and squared curvature respectively (Terrell, 1990). The normal kernel $\phi(x) = (2\pi)^{-1/2} \exp(-x^2/2)$, which is also widely used, is not strictly a member of the beta family class, but is the limiting case as $r \to \infty$, see Marron and Nolan (1988).

For multivariate data, there are two main ways to obtain multivariate kernels from these univariate kernels. The product kernel is, as its name suggests, the product of the marginal univariate kernels. The spherically symmetric kernel is obtained by substituting $x = (\boldsymbol{x}^T\boldsymbol{x})^{1/2}$, where $\boldsymbol{x} = (x_1, \ldots, x_d)$. For the beta family kernels, these are

$$K^P(\boldsymbol{x}; r) = c_r^d \prod_{i=1}^{d} (1 - x_i^2)^r \mathbf{1}\{|x_i| \leq 1\}$$

$$K^S(\boldsymbol{x}; r) = c_r^S (1 - \boldsymbol{x}^T\boldsymbol{x})^r \mathbf{1}\{\boldsymbol{x}^T\boldsymbol{x} \leq 1\}.$$

E-mail address: duong@lipn.univ-paris13.fr.

http://dx.doi.org/10.1016/j.spl.2015.05.012

Table 1
Spherically symmetric beta family kernels. The normalising constant c_r^S, the second moment $m_2(K^S(\cdot; r))$ and the square integral $R(K^S(\cdot; r))$ for $d = 1, 2, 3$ and $r = 0, 1, 2, 3$.

	r	c_r^S			$m_2(K^S(\cdot; r))$			$R(K^S(\cdot; r))$		
		$d = 1$	$d = 2$	$d = 3$	$d = 1$	$d = 2$	$d = 3$	$d = 1$	$d = 2$	$d = 3$
Uniform	0	1/2	$1/\pi$	$3/(4\pi)$	1/3	1/4	1/5	1/2	$1/\pi$	$3/(4\pi)$
Epanechnikov	1	3/4	$2/\pi$	$15/(8\pi)$	1/5	1/6	1/7	3/5	$4/(3\pi)$	$15/(14\pi)$
Biweight	2	15/16	$3/\pi$	$105/(32\pi)$	1/7	1/8	1/9	5/7	$9/(5\pi)$	$35/(22\pi)$
Triweight	3	35/32	$4/\pi$	$315/(64\pi)$	1/9	1/10	1/11	350/429	$16/(7\pi)$	$315/(143\pi)$

The normalising constant c_r^S is less straightforward to compute as it involves the integration of the powers of the quadratic form $\boldsymbol{x}^T\boldsymbol{x}$, with only $c_0^S = 1/v_d$ and $c_1^S = (d+2)/(2v_d)$ currently known in closed form, where $v_d = \pi^{d/2}/\Gamma((d+2)/2)$ is the hyper-volume of unit d-ball $B_d \equiv B_d(\mathbf{0}, 1) = \{\boldsymbol{x} : \boldsymbol{x}^T\boldsymbol{x} \le 1\}$, see Fukunaga and Hostetler (1975). We state two preliminary results in Theorem 1 which will facilitate our goal of a closed form characterisation of a spherically symmetric rth beta family kernel $K^S(\cdot; r)$ in Theorem 2.

Theorem 1. *For $r \ge 0$ and $d \ge 1$,(i) the partial sum of the alternating series of binomial coefficients and rational functions is*

$$\sum_{i=0}^{r} \binom{r}{i} \frac{(-1)^i}{d+2i} = 2^r r! \Big/ \prod_{i=0}^{r} (d+2i) = \tfrac{1}{2} B(r+1, d/2);$$

(ii) *the integral of the rth power of the quadratic form $\boldsymbol{x}^T\boldsymbol{x}$ over the unit d-ball B_d is*

$$\int_{B_d} (\boldsymbol{x}^T\boldsymbol{x})^r \, d\boldsymbol{x} = \frac{d v_d}{d+2r}$$

where $v_d = \pi^{d/2}/\Gamma((d+2)/2)$ is the hyper-volume of B_d.

Sacks and Ylvisaker (1981) exhibited a similar result to Theorem 1(ii) but did not provide a proof.

In addition to the normalising constant for a kernel K, the two other important quantities which characterise it are its second moment $m_2(K)\mathbf{I}_d = \int_{B_d} \boldsymbol{x}\boldsymbol{x}^T K(\boldsymbol{x}) \, d\boldsymbol{x}$ where $\mathbf{I}_d$ is the $d \times d$ identity matrix, and its square integral $R(K) = \int_{B_d} K(\boldsymbol{x})^2 \, d\boldsymbol{x}$.

Theorem 2. *For the spherically symmetric rth beta family kernel $K^S(\boldsymbol{x}; r) = c_r^S (1 - \boldsymbol{x}^T\boldsymbol{x})^r \mathbf{1}\{\boldsymbol{x} \in B_d\}$, the normalising constant, the second moment and the square integral are*

$$c_r^S = \frac{1}{v_d 2^r r!} \prod_{i=1}^{r} (d+2i) = \frac{2}{d v_d B(r+1, d/2)}$$

$$m_2(K^S(\cdot; r)) = \frac{1}{d+2r+2}$$

$$R(K^S(\cdot; r)) = \frac{(2r)!}{v_d (r!)^2} \prod_{i=1}^{r} \frac{(d+2i)}{(d+2r+2i)} = \frac{2B(2r+1, d/2)}{d v_d B(r+1, d/2)^2}.$$

We compute the coefficients in Theorem 2 explicitly for $d = 1, 2, 3$ and $r = 0, 1, 2, 3$ in Table 1.

For $d = 1$, these constants have an alternate form since $B(r+1, 1/2) = 2^{2r+1} B(r+1, r+1)$ using the identity $\Gamma(r+1/2) = (2r)!/(2^{2r} r!)\Gamma(1/2)$, i.e., $c_r = 1/[2^{2r+1} B(r+1, r+1)]$, $m_2(K(\cdot; r)) = 1/(2r+3)$ and $R(K(\cdot; r)) = B(2r+1, 2r+1)/[2B(r+1, r+1)^2]$.

Sacks and Ylvisaker (1981) state that the spherically symmetric Epanechnikov kernel $K^S(\cdot; 1)$ is optimal in a MISE sense rather than its product kernel counterpart. So the efficiencies of the other kernels can be expressed in the ratio $[C(K^S(\cdot; 1))/C(K^S(\cdot; r))]^{(d+4)/4}$ where $C(K) = [R(K)^4 m_2(K)^{2d}]^{1/(d+4)}$, see Wand and Jones (1995, p. 103). These efficiency ratios are less than one, and they can be interpreted as follows: to achieve the same MISE as the optimal Epanechnikov kernel with sample size n, the kernel $K^S(\cdot; r)$ requires a sample size of $n[C(K^S(\cdot; r))/C(K^S(\cdot; 1))]^{(d+4)/4}$. From Theorem 2, the efficiency for an rth beta kernel is

$$\left[\frac{C(K^S(\cdot; 1))}{C(K^S(\cdot; r))}\right]^{(d+4)/4} = \frac{R(K^S(\cdot; 1)) m_2(K^S(\cdot; 1))^{d/2}}{R(K^S(\cdot; r)) m_2(K^S(\cdot; r))^{d/2}} = \frac{B(3, d/2) B(r+1, d/2)^2}{B(2, d/2)^2 B(2r+1, d/2)} \left[\frac{d+2r+2}{d+4}\right]^{d/2}$$

which are explicitly calculated in Table 2, including the normal kernel ϕ as $m_2(\phi) = 1$, $R(\phi) = (4\pi)^{-d/2}$, thus extending Wand and Jones (1995, Table 2.1, p. 31) for $d > 1$. As d increases, the efficiency decreases, though we note, as is the case for univariate data, the loss of efficiency is small.

Table 2
Efficiencies for spherically symmetric beta family kernels. The efficiency measure is $[C(K^S(\cdot;1))/C(K^S(\cdot;r))]^{(d+4)/4}$ for $d = 1, 2, 3$ and $r = 0, 1, 2, 3, \infty$.

	r	Efficiency		
		$d = 1$	$d = 2$	$d = 3$
Uniform	0	0.930	0.889	0.862
Epanechnikov	1	1.000	1.000	1.000
Biweight	2	0.994	0.988	0.982
Triweight	3	0.987	0.972	0.958
Normal	∞	0.951	0.889	0.820

2. Proofs

Proof of Theorem 1. (i) Let the induction hypothesis be the theorem statement

$$\sum_{i=0}^{r}\binom{r}{i}\frac{(-1)^i}{d+2i} = 2^r r! \Big/ \prod_{i=0}^{r}(d+2i).$$

Incrementing r, we have

$$\sum_{i=0}^{r+1}\binom{r+1}{i}\frac{(-1)^i}{d+2i} = \frac{1}{d} + \sum_{i=1}^{r}\binom{r+1}{i}\frac{(-1)^i}{d+2i} + \frac{(-1)^{r+1}}{d+2r+2} = \frac{1}{d} + \sum_{i=1}^{r}\left[\binom{r}{i} + \binom{r}{i-1}\right]\frac{(-1)^i}{d+2i} + \frac{(-1)^{r+1}}{d+2r+2}$$
$$= \sum_{i=0}^{r}\binom{r}{i}\frac{(-1)^i}{d+2i} + \sum_{i=1}^{r+1}\binom{r}{i-1}\frac{(-1)^i}{d+2i} = \sum_{i=0}^{r}\binom{r}{i}\frac{(-1)^i}{d+2i} - \sum_{i=0}^{r}\binom{r}{i}\frac{(-1)^i}{d+2+2i}.$$

The first summation on the right hand side is the induction hypothesis as stated, and the second summation is the induction hypothesis with $d + 2$ replacing d, i.e.,

$$\sum_{i=0}^{r+1}\binom{r+1}{i}\frac{(-1)^i}{d+2i} = \frac{2^r r!}{\prod_{i=0}^{r}(d+2i)} - \frac{2^r r!}{\prod_{i=0}^{r}(d+2+2i)} = \frac{2^r r!(d+2+2r-d)}{\prod_{i=0}^{r+1}(d+2i)} = \frac{2^{r+1}(r+1)!}{\prod_{i=0}^{r+1}(d+2i)}$$

which completes the induction process.

Using the 2-gamma function $\Gamma_2(a) = \int_0^\infty x^{a-1}\exp(-x^2/2)\,dx = 2^{a/2-1}\Gamma(a/2)$, the product $\prod_{i=0}^{r}(d+2i)$ can be expressed as a ratio of 2-gamma functions from Díaz and Pariguan (2007, Proposition 6.2), and hence gamma functions,

$$\prod_{i=0}^{r}(d+2i) = \frac{\Gamma_2(d+2r+2)}{\Gamma_2(d)} = \frac{2^{(d+2r)/2}\Gamma((d+2r+2)/2)}{2^{d/2-1}\Gamma(d/2)} = \frac{2^{r+1}\Gamma((d+2r+2)/2)}{\Gamma(d/2)}. \tag{1}$$

Therefore

$$\frac{2^r r!}{\prod_{i=0}^{r}(d+2i)} = \frac{2^r\Gamma(r+1)\Gamma(d/2)}{2^{r+1}\Gamma(r+1+d/2)} = \tfrac{1}{2}B(r+1, d/2).$$

(ii) We follow the approach of Folland (2001) who connects integrals of $(\boldsymbol{x}^T\boldsymbol{x})^r$ over the unit ball with the more tractable integrals of $(\boldsymbol{x}^T\boldsymbol{x})^r\exp(-\frac{1}{2}\boldsymbol{x}^T\boldsymbol{x})$ over the entire Euclidean space:

$$\begin{aligned}\int_{\mathbb{R}^d}(\boldsymbol{x}^T\boldsymbol{x})^r\exp(-\tfrac{1}{2}\boldsymbol{x}^T\boldsymbol{x})\,d\boldsymbol{x} &= 2^{(d+2r)/2}\int_{\mathbb{R}^d}(\boldsymbol{x}^T\boldsymbol{x})^r\exp(-\boldsymbol{x}^T\boldsymbol{x})\,d\boldsymbol{x}\\ &= 2^{(d+2r)/2}\int_{S_d}\left[\int_0^\infty(x^2\boldsymbol{y}^T\boldsymbol{y})^r\exp(-x^2)x^{d-1}\,dx\right]d\sigma(\boldsymbol{y})\\ &= 2^{(d+2r)/2}\int_0^\infty x^{d+2r-1}\exp(-x^2)\,dx\int_{S_d}(\boldsymbol{y}^T\boldsymbol{y})^r\,d\sigma(\boldsymbol{y})\\ &= 2^{(d+2r-2)/2}\Gamma((d+2r)/2)\int_{S_d}(\boldsymbol{y}^T\boldsymbol{y})^r\,d\sigma(\boldsymbol{y}).\end{aligned} \tag{2}$$

The second equality follows from the result from Folland (2001) that the integral of a function f over $\mathbb{R}^d$ can be evaluated as an iterated integral over $[0, 1] \times S_d$

$$\int_{\mathbb{R}^d} f(\boldsymbol{x})\,d\boldsymbol{x} = \int_{S_d}\left[\int_0^\infty f(x\boldsymbol{y})x^{d-1}\,dx\right]d\sigma(\boldsymbol{y}),$$

using the change of variables $\boldsymbol{y} = \boldsymbol{x}/x$ where $x = \|\boldsymbol{x}\| = (\boldsymbol{x}^T\boldsymbol{x})^{1/2}$, and $d\sigma$ is the surface area element on the unit sphere $S_d = \partial B_d$. The last equality follows from the definition of the gamma function $\Gamma(a) = 2\int_0^\infty x^{2a-1}\exp(-x^2)\,dx$,

$$\int_0^\infty x^{d+2r-1}\exp(-x^2)\,dx = \tfrac{1}{2}\Gamma((d+2r)/2).$$

Folland (2001) also demonstrated the relationship between the integral of $(\boldsymbol{x}^T\boldsymbol{x})^r$ over the unit ball and the unit sphere:

$$\int_{B_d}(\boldsymbol{x}^T\boldsymbol{x})^r\,d\boldsymbol{x} = \frac{1}{d+2r}\int_{S_d}(\boldsymbol{x}^T\boldsymbol{x})^r\,d\sigma(\boldsymbol{x})$$

which when substituted into a rearranged Eq. (2) yields

$$\int_{B_d}(\boldsymbol{x}^T\boldsymbol{x})^r\,d\boldsymbol{x} = \frac{1}{(d+2r)2^{(d+2r-2)/2}\Gamma((d+2r)/2)}\int_{\mathbb{R}^d}(\boldsymbol{x}^T\boldsymbol{x})^r\exp(-\tfrac{1}{2}\boldsymbol{x}^T\boldsymbol{x})\,d\boldsymbol{x}. \tag{3}$$

The integral on the right hand side of Eq. (3) can be expressed as the expected value of $(\mathbf{Z}^T\mathbf{Z})^r$ where $\mathbf{Z}$ is a standard normal random variable

$$\nu_r \equiv \mathbf{E}[(\mathbf{Z}^T\mathbf{Z})^r] = (2\pi)^{-d/2}\int_{\mathbb{R}^d}(\boldsymbol{x}^T\boldsymbol{x})^r\exp(-\tfrac{1}{2}\boldsymbol{x}^T\boldsymbol{x})\,d\boldsymbol{x}$$

as extensively studied, for example by Holmquist (1996) and the references therein. To evaluate this ν_r, Chacón et al. (2011, Corollary 7) establish the recurrence relation $\nu_{r+1} = (d+2r)\nu_r$, which immediately implies that

$$\nu_r = d(d+2)\cdots(d+2(r-1)) = \prod_{i=0}^{r-1}(d+2i)$$

as $\nu_0 = 1$. Thus the numerator is $\int_{\mathbb{R}^d}(\boldsymbol{x}^T\boldsymbol{x})^r\exp(-\frac{1}{2}\boldsymbol{x}^T\boldsymbol{x})\,d\boldsymbol{x} = (2\pi)^{d/2}\prod_{i=0}^{r-1}(d+2i)$. For the gamma function in the denominator

$$\Gamma((d+2r)/2) = \frac{d+2(r-1)}{2}\Gamma((d+2(r-1))/2) = \cdots = \frac{\Gamma((d+2)/2)}{2^{r-1}}\prod_{i=1}^{r-1}(d+2i)$$

by repeatedly applying the recursive property of the gamma function, $\Gamma(a+1) = a\Gamma(a)$ for any $a > 0$. Combining these alternative expressions for the numerator and denominator in Eq. (3), we obtain

$$\int_{B_d}(\boldsymbol{x}^T\boldsymbol{x})^r\,d\boldsymbol{x} = \frac{1}{(d+2r)}\frac{\pi^{d/2}\prod_{i=0}^{r-1}(d+2i)}{\Gamma((d+2)/2)\prod_{i=1}^{r-1}(d+2i)} = \frac{d\nu_d}{d+2r}. \qquad \square$$

Proof of Theorem 2. To obtain the closed form characterisations for $K^S(\boldsymbol{x}; r)$, we appeal to the binomial expansion of $(1-\boldsymbol{x}^T\boldsymbol{x})^r$ and Theorem 1 to evaluate the factorial-like products and the alternating partial sums. The normalisation constant c_r^S is the reciprocal of

$$\int_{B_d}[1-(\boldsymbol{x}^T\boldsymbol{x})]^r\,d\boldsymbol{x} = \sum_{i=0}^r(-1)^i\binom{r}{i}\int_{B_d}(\boldsymbol{x}^T\boldsymbol{x})^i\,d\boldsymbol{x} = d\nu_d\sum_{i=0}^r\binom{r}{i}\frac{(-1)^i}{d+2i} = \frac{d\nu_d 2^r r!}{\prod_{i=0}^r(d+2i)} = \frac{d\nu_d}{2}B(r+1, d/2).$$

Taking traces of the definition of $m_2(K^S(\cdot; r))$, we have

$$m_2(K^S(\cdot; r)) = \frac{c_r^S}{d}\int_{B_d}(\boldsymbol{x}^T\boldsymbol{x})(1-\boldsymbol{x}^T\boldsymbol{x})^r\,d\boldsymbol{x} = \frac{c_r^S}{d}\sum_{i=0}^r(-1)^i\binom{r}{i}\int_{B_d}(\boldsymbol{x}^T\boldsymbol{x})^{i+1}\,d\boldsymbol{x}$$

$$= \nu_d c_r^S\sum_{i=0}^r\binom{r}{i}\frac{(-1)^i}{d+2+2i} = \frac{\nu_d\prod_{i=1}^r(d+2i)}{\nu_d 2^r r!}\frac{2^r r!}{\prod_{i=0}^r(d+2+2i)} = \frac{1}{d+2r+2}.$$

For $R(K^S(\cdot; r))$, similarly we have

$$
\begin{aligned}
R(K^S(\cdot; r)) &= (c_r^S)^2 \int_{B_d} (1 - \boldsymbol{x}^T\boldsymbol{x})^{2r}\, d\boldsymbol{x} = (c_r^S)^2 \sum_{i=0}^{2r} (-1)^i \binom{2r}{i} \int_{B_d} (\boldsymbol{x}^T\boldsymbol{x})^i\, d\boldsymbol{x} \\
&= (c_r^S)^2 d v_d \sum_{i=0}^{2r} \binom{2r}{i} \frac{(-1)^i}{d+2i} = \left[\frac{\prod_{i=1}^{r}(d+2i)}{(v_d 2^r r!)} \right]^2 \frac{v_d 2^{2r}(2r)!}{\prod_{i=1}^{2r}(d+2i)} \\
&= \frac{(2r)!}{v_d (r!)^2} \frac{\prod_{i=1}^{r}(d+2i)}{\prod_{i=r+1}^{2r}(d+2i)} = \frac{(2r)!}{v_d (r!)^2} \prod_{i=1}^{r} \frac{(d+2i)}{(d+2r+2i)} \\
&= \frac{4}{d^2 v_d^2 B(r+1, d/2)^2} \frac{d v_d B(2r+1, d/2)}{2} = \frac{2}{d v_d} \frac{B(2r+1, d/2)}{B(r+1, d/2)^2}. \qquad \square
\end{aligned}
$$

Acknowledgement

This research has been supported by the Square Predict project within a 'Projet investissement d'avenir (PIA)' 2013–2016 grant provided by the French government.

References

Chacón, J.E., Duong, T., Wand, M.P., 2011. Asymptotics for general multivariate kernel density derivative estimators. Statist. Sinica 21, 807–840.
Díaz, R., Pariguan, E., 2007. On hypergeometric functions and Pochhammer k-symbol. Divulg. Mat. 15, 179–192.
Epanechnikov, V.A., 1969. Non-parametric estimation of a multivariate probability density. Theory Probab. Appl. 14, 153–158.
Folland, G., 2001. How to integrate a polynomial over a sphere. Amer. Math. Monthly 108, 446–448.
Fukunaga, K., Hostetler, L., 1975. The estimation of the gradient of a density function, with applications in pattern recognition. IEEE Trans. Inform. Theory 21, 32–40.
Holmquist, B., 1996. Expectations of products of quadratic forms in normal variables. Stoch. Anal. Appl. 14, 149–164.
Marron, J., Nolan, D., 1988. Canonical kernels for density estimation. Statist. Probab. Lett. 7, 195–199.
Sacks, J., Ylvisaker, D., 1981. Asymptotically optimum kernels for density estimation at a point. Ann. Statist. 9, 334–346.
Terrell, G.R., 1990. The maximal smoothing principle in density estimation. J. Amer. Statist. Assoc. 85, 470–477.
Wand, M.P., Jones, M.C., 1995. Kernel Smoothing. Chapman and Hall/CRC, London.