Technical report from Automatic Control at Linköpings universitet

Properties and approximations of some matrix variate probability density functions

Karl Granström, Umut Orguner

Division of Automatic Control

E-mail: karl@isy.liu.se, umut@eee.metu.edu.tr

16th December 2011

Report no.: LiTH-ISY-R-3042

Address:
Department of Electrical Engineering
Linköpings universitet
SE-581 83 Linköping, Sweden

WWW: http://www.control.isy.liu.se

Abstract

This report contains properties and approximations of some matrix valued probability density functions. Expected values of functions of generalised Beta type II distributed random variables are derived. In two Theorems, approximations of matrix variate distributions are derived. A third theorem contain a marginalisation result.

Keywords: Extended target, random matrix, Kullback-Leibler divergence, inverse Wishart, Wishart, generalized Beta.

Properties and approximations of some matrix variate probability density functions

Karl Granström, Umut Orguner

2013-02-15

Abstract

This report contains properties and approximations of some matrix valued probability density functions. Expected values of functions of generalised Beta type II distributed random variables are derived. In two Theorems, approximations of matrix variate distributions are derived. A third theorem contain a marginalisation result.

1 Some matrix variate distributions

1.1 Wishart distribution

Let $\mathbb{S}^d_{++}$ be the set of symmetric positive definite $d \times d$ matrices. The random matrix $X \in \mathbb{S}^d_{++}$ is Wishart distributed with degrees of freedom $n > d-1$ and $d \times d$ scale matrix $N \in \mathbb{S}^d_{++}$ if it has probability density function (pdf) [1, Definition 3.2.1]

$$p(X) = \mathcal{W}_d\left(X\,;\,n, N\right) \tag{1}$$

$$= \frac{|X|^{\frac{n-d-1}{2}}}{2^{\frac{nd}{2}}\Gamma_d\left(\frac{n}{2}\right)|N|^{\frac{n}{2}}}\operatorname{etr}\left(-\frac{1}{2}N^{-1}X\right), \tag{2}$$

where, for $a > \frac{d-1}{2}$, the multivariate gamma function, and its logarithm, can be expressed in terms of the ordinary gamma function as [1, Theorem 1.4.1]

$$\Gamma_d(a) = \pi^{d(d-1)}\prod_{i=1}^{d}\Gamma\left(a-(i-1)/2\right), \tag{3a}$$

$$\log\Gamma_d(a) = d(d-1)\log\pi + \sum_{i=1}^{d}\log\Gamma\left(a-(i-1)/2\right). \tag{3b}$$

Let A_{ij} denote the i,j:th element of a matrix A. The expected value and covariance of X are [1, Theorem 3.3.15]

$$\mathrm{E}[X_{ij}] = nN_{ij}, \tag{4}$$

$$\mathrm{Cov}(X_{ij}, X_{kl}) = n(N_{ik}N_{jl} + N_{il}N_{jk}). \tag{5}$$

1.2 Inverse Wishart distribution

The random matrix $X \in \mathbb{S}^d_{++}$ is inverse Wishart distributed with degrees of freedom $v > 2d$ and inverse scale matrix $V \in \mathbb{S}^d_{++}$ if it has pdf [1, Definition 3.4.1]

$$p(X) = \mathcal{IW}_d\left(X\,;\, v, V\right) \tag{6}$$

$$= \frac{2^{-\frac{v-d-1}{2}}|V|^{\frac{v-d-1}{2}}}{\Gamma_d\left(\frac{v-d-1}{2}\right)|X|^{\frac{v}{2}}} \operatorname{etr}\left(-\frac{1}{2}X^{-1}V\right). \tag{7}$$

The expected value and covariance of X are [1, Theorem 3.4.3]

$$\mathrm{E}[X_{ij}] = \frac{V_{ij}}{v-2d-2}, \quad v-2d-2>0, \tag{8}$$

$$\mathrm{Cov}(X_{ij}, X_{kl}) = \frac{2(v-2d-2)^{-1}V_{ij}V_{kl} + V_{ik}V_{jl} + V_{il}V_{jk}}{(v-2d-1)(v-2d-2)(v-2d-4)}, \quad v-2d-4>0. \tag{9}$$

1.3 Generalized matrix variate Beta type II distribution

Let $\mathbb{S}^d_{+}$ be the set of symmetric positive semi-definite $d \times d$ matrices. The random matrix $X \in \mathbb{S}^d_{++}$ is generalized matrix variate Beta type II distributed with matrix parameters $\Psi \in \mathbb{S}^d_{+}$ and $\Omega > \Psi$, and scalar parameters a and b, if it has pdf [1, Definition 5.2.4]

$$p(X) = \mathcal{GB}^{II}_d\left(X; a, b, \Omega, \Psi\right) \tag{10}$$

$$= \frac{|X-\Psi|^{a-\frac{d+1}{2}}|\Omega+X|^{-(a+b)}}{\beta_d(a,b)|\Omega+\Psi|^{-b}}, \quad X > \Psi \tag{11}$$

where, for $a > \frac{d-1}{2}$ and $b > \frac{d-1}{2}$, the multivariate beta function is expressed in terms of the multivariate gamma function as [1, Theorem 1.4.2]

$$\beta_d(a,b) = \frac{\Gamma_d(a)\Gamma_d(b)}{\Gamma_d(a+b)}. \tag{12}$$

Let $\mathbf{0}_d$ be a $d \times d$ all zero matrix. If $\Psi = \mathbf{0}_d$, the first and second order moments of X are [1, Theorem 5.3.20]

$$\mathrm{E}[X_{ij}] = \frac{2a}{2b-d-1}\Omega_{ij} \tag{13}$$

$$\mathrm{E}[X_{ij}X_{kl}] = \frac{2a}{(2b-d)(2b-d-1)(2b-d-3)}\left[\left\{2a(2b-d-2)+2\right\}\Omega_{ij}\Omega_{kl} + (2a+2b-d-1)(\Omega_{jl}\Omega_{ik} + \Omega_{il}\Omega_{kj})\right], \quad 2b-d-3>0 \tag{14}$$

2 Expected values of the $\mathcal{GB}_d^{II}$-distribution

This appendix derives some expected values for the matrix variate generalized beta type-II distribution.

2.1 Expected value of the inverse

Let U be matrix variate beta type-II distributed with pdf [1, Definition 5.2.2]

$$p(U) = \mathcal{B}_d^{II}(U;\ a,\ b) \tag{15}$$

$$= \frac{|U|^{a-\frac{d+1}{2}}|\mathbf{I}_d + U|^{-(a+b)}}{\beta_d(a,b)} \tag{16}$$

where $a > \frac{d-1}{2}$, $b > \frac{d-1}{2}$, and $\mathbf{I}_d$ is a $d \times d$ identity matrix. Then U^{-1} has pdf [1, Theorem 5.3.6]

$$p(U^{-1}) = \mathcal{B}_d^{II}\left(U^{-1};\ b,\ a\right). \tag{17}$$

Let $X = \Omega^{1/2}U\Omega^{1/2}$ where $\Omega \in \mathbb{S}_{++}^d$. The pdf of (X) is [1, Theorem 5.2.2]

$$p(X) = \mathcal{GB}_d^{II}(X;\ a,\ b,\ \Omega,\ \mathbf{0}_d) \tag{18}$$

and subsequently the pdf of $X^{-1} = \Omega^{-1/2}U^{-1}\Omega^{-1/2}$ is

$$p(X^{-1}) = \mathcal{GB}_d^{II}\left(X^{-1};\ b,\ a,\ \Omega^{-1},\ \mathbf{0}_d\right) \tag{19}$$

The expected value of X^{-1} is [1, Theorem 5.3.20]

$$\mathrm{E}\left[X^{-1}\right] = \frac{2b}{2a-d-1}\Omega^{-1}. \tag{20}$$

2.2 Expected value of the log-determinant

Let y be a univariate random variable. The moment generating function of y is defined as

$$\mu_y(s) \triangleq \mathrm{E}_y\left[e^{sy}\right], \tag{21}$$

and the expected value of y is given in terms of $\mu_y(s)$ as

$$\mathrm{E}[y] = \left.\frac{d\mu_y(s)}{ds}\right|_{s=0}. \tag{22}$$

Let $y = \log|X|$, where $p(X) = \mathcal{B}_d^{II}(X;\ a,\ b)$. The moment generating function of y is

$$\mu_y(s) = \mathrm{E}\left[|X|^s\right] = \int |X|^s p(X)\, dX \tag{23a}$$

$$= \int |X|^s \beta_d^{-1}(a,b)|X|^{a-\frac{1}{2}(d+1)}|\mathbf{I}_d + X|^{-(a+b)} dX \tag{23b}$$

$$=\beta_d^{-1}(a,b)\beta_d(a+s,b-s)$$

$$\times \int \beta_d^{-1}(a+s,b-s)|X|^{a+s-\frac{1}{2}(d+1)}|\mathbf{I}_d + X|^{-(a+s+b-s)} dX \tag{23c}$$

$$=\frac{\beta_d(a+s,b-s)}{\beta_d(a,b)} \int \mathcal{B}_d^{II}(X;\ a+S,\ b-s)\, dX \tag{23d}$$

$$=\frac{\beta_d(a+s,b-s)}{\beta_d(a,b)} = \frac{\Gamma_d(a+s)\Gamma_d(b-s)}{\Gamma_d(a+s+b-s)} \frac{\Gamma_d(a+b)}{\Gamma_d(a)\Gamma_d(b)} \tag{23e}$$

$$=\frac{\Gamma_d(a+s)\Gamma_d(b-s)}{\Gamma_d(a)\Gamma_d(b)}, \tag{23f}$$

The expected value of y is

$$\mathrm{E}[y] = \mathrm{E}[\log|X|] \tag{24a}$$

$$= \left.\frac{d\big(\Gamma_d(a+s)\Gamma_d(b-s)\big)}{ds}\right|_{s=0} \frac{1}{\Gamma_d(a)\Gamma_d(b)} \tag{24b}$$

$$= \left.\left(\frac{\frac{d\Gamma_d(a+s)}{ds}}{\Gamma_d(a+s)} + \frac{\frac{d\Gamma_d(b-s)}{ds}}{\Gamma_d(b-s)}\right)\right|_{s=0} \tag{24c}$$

$$= \left.\left(\frac{d\log\Gamma_d(a+s)}{ds} + \frac{d\log\Gamma_d(b-s)}{ds}\right)\right|_{s=0} \tag{24d}$$

$$= \left.\left(\sum_{i=1}^{d} \frac{d\log\Gamma(a+s-(i-1)/2)}{ds} + \frac{d\log\Gamma(b-s-(i-1)/2)}{ds}\right)\right|_{s=0} \tag{24e}$$

$$= \sum_{i=1}^{d} \psi_0\left(a-(i-1)/2\right) - \psi_0\left(b-(i-1)/2\right), \tag{24f}$$

where $\psi_0(\cdot)$ is the digamma function, also called the polygamma function of order zero. If $p(Y) = \mathcal{GB}_d^{II}(Y;\ a,\ b,\ \Omega,\ \mathbf{0}_d)$, then $Z = \Omega^{-1/2}Y\Omega^{-1/2}$ has pdf $\mathcal{B}_d^{II}(Z;\ a,\ b)$ [1, Theorem 5.2.2]. It then follows that

$$\mathrm{E}[\log|Y|] = \mathrm{E}\left[\log|\Omega^{1/2}\Omega^{-1/2}Y\Omega^{-1/2}\Omega^{1/2}|\right] \tag{25a}$$

$$= \mathrm{E}\left[\log|\Omega^{1/2}| + \log|\Omega^{-1/2}Y\Omega^{-1/2}| + \log|\Omega^{1/2}|\right] \tag{25b}$$

$$= \mathrm{E}\left[\log|\Omega| + \log|\Omega^{-1/2}Y\Omega^{-1/2}|\right] \tag{25c}$$

$$= \log|\Omega| + \mathrm{E}[\log|Z|] \tag{25d}$$

$$= \log|\Omega| + \sum_{i=1}^{d}\Big[\psi_0\left(a-(i-1)/2\right) - \psi_0\left(b-(i-1)/2\right)\Big]. \tag{25e}$$

3 Approximating a $\mathcal{GB}_d^{II}$-distribution with an $\mathcal{IW}_d$-distribution

This section presents a theorem that approximates a $\mathcal{GB}_d^{II}$-distribution with an $\mathcal{IW}_d$-distribution.

3.1 Theorem 1

Theorem 1. *Let* $p(X) = \mathcal{GB}_d^{II}(X;\ a,\ b,\ \Omega,\ \mathbf{0}_d)$*, and let* $q(X) = \mathcal{IW}_d(X\,;\, v, V)$ *be the minimizer of the Kullback-Leibler (*KL*) divergence between* $p(X)$ *and* $q(X)$ *among all* $\mathcal{IW}_d$*-distributions, i.e.*

$$q(X) \triangleq \underset{q(\,\cdot\,)=\mathcal{IW}_d(\,\cdot\,)}{\arg\min} \ \mathrm{KL}\left(p(X)\,||q(X)\right). \tag{26}$$

Then V *is given as*

$$V = \frac{(v-d-1)(2a-d-1)}{2b}\Omega, \tag{27}$$

and v *is the solution to the equation*

$$\begin{aligned}&\sum_{i=1}^{d}\left[\psi_0\left(\frac{2a+1-i}{2}\right)-\psi_0\left(\frac{2b+1-i}{2}\right)\right.\\&\left.+\psi_0\left(\frac{v-d-i}{2}\right)\right]-d\log\left(\frac{(v-d-1)(2a-d-1)}{4b}\right)=0,\end{aligned} \tag{28}$$

where $\psi_0(\,\cdot\,)$ *is the digamma function (a.k.a. the polygamma function of order 0).* □

3.2 Proof of Theorem 1

The density $q(X)$ is given as

$$q(X) \triangleq \underset{q(X)}{\arg\min}\ \mathrm{KL}\left(p(X)\,||q(X)\right) \tag{29a}$$

$$=\underset{q(X)}{\arg\max} \int p(X)\log\left(q(X)\right)dX \tag{29b}$$

$$\begin{aligned}=\underset{q(X)}{\arg\max} \int p(X)\Bigg[&-\frac{(v-d-1)d}{2}\log 2+\frac{v-d-1}{2}\log|V|\\&-\log\Gamma_d\left(\frac{v-d-1}{2}\right)-\frac{v}{2}\log|X|+\mathrm{Tr}\left(-\frac{1}{2}X^{-1}V\right)\Bigg]dX\end{aligned} \tag{29c}$$

$$\begin{aligned}=\underset{q(X)}{\arg\max}\ &-\frac{(v-d-1)d}{2}\log 2+\frac{v-d-1}{2}\log|V|\\&-\log\Gamma_d\left(\frac{v-d-1}{2}\right)-\frac{v}{2}\mathrm{E}_p\left[\log|X|\right]+\mathrm{Tr}\left(-\frac{1}{2}\mathrm{E}_p\left[X^{-1}\right]V\right)\end{aligned} \tag{29d}$$

$$=\underset{q(X)}{\arg\max}\ f(v,V) \tag{29e}$$

Differentiating the objective function $f(v, V)$ with respect to V gives

$$\frac{df(v,V)}{dV} = \frac{v-d-1}{2} V^{-1} - \frac{1}{2} \mathrm{E}_p \left[X^{-1} \right]. \tag{30}$$

Setting to zero and solving for V gives

$$V = (v-d-1)\, \mathrm{E}_p \left[X^{-1} \right]^{-1} = \frac{(v-d-1)(2a-d-1)}{2b} \Omega \tag{31}$$

where the expected value is calculated based on a result derived in Section 2. Differentiating the objective function with respect to v gives

$$\begin{aligned} \frac{df(v,V)}{dv} &= -\frac{d}{2} \log 2 + \frac{1}{2} \log |V| - \frac{d \log \Gamma_d \left(\frac{v-d-1}{2} \right)}{dv} - \frac{1}{2} \mathrm{E}_p \left[\log |X| \right] && \text{(32a)} \\ &= -\frac{d}{2} \log 2 + \frac{1}{2} \log |V| - \frac{1}{2} \sum_{i=1}^{d} \psi_0 \left(\frac{v-d-i}{2} \right) - \frac{1}{2} \mathrm{E}_p \left[\log |X| \right]. && \text{(32b)} \end{aligned}$$

Setting the result equal to zero gives

$$\begin{aligned} 0 =& \log |V| - d \log 2 - \sum_{i=1}^{d} \psi_0 \left(\frac{v-d-i}{2} \right) - \mathrm{E}_p \left[\log |X| \right] && \text{(33a)} \\ =& \log |V| - d \log 2 - \sum_{i=1}^{d} \psi_0 \left(\frac{v-d-i}{2} \right) - \log |\Omega| \\ & - \sum_{i=1}^{d} \left[\psi_0 \left(a - \frac{1}{2}(i-1) \right) - \psi_0 \left(b - \frac{1}{2}(i-1) \right) \right] && \text{(33b)} \end{aligned}$$

where the expected value of $\log |X|$ is derived in Section 2. Inserting V from (31) gives

$$\begin{aligned} 0 =& \log |\Omega| + d \log \left(\frac{(v-d-1)(2a-d-1)}{2b} \right) - d \log 2 - \sum_{i=1}^{d} \psi_0 \left(\frac{v-d-i}{2} \right) \\ & - \log |\Omega| - \sum_{i=1}^{d} \left[\psi_0 \left(\frac{2a+1-i}{2} \right) - \psi_0 \left(\frac{2b+1-i}{2} \right) \right] && \text{(34)} \\ =& d \log \left(\frac{(v-d-1)(2a-d-1)}{4b} \right) \\ & - \sum_{i=1}^{d} \left[\psi_0 \left(\frac{2a+1-i}{2} \right) - \psi_0 \left(\frac{2b+1-i}{2} \right) + \psi_0 \left(\frac{v-d-i}{2} \right) \right] && \text{(35)} \end{aligned}$$

which is the equation for v in the theorem.

3.3 Corollary to Theorem 1

Corollary 1. *A closed form solution for v can be obtained using only* (27) *together with matching the first order moments. The expected values of the* *densities $p(\cdot)$ and $q(\cdot)$ are [1, Theorems 5.3.20, 3.4.3]*

$$\mathrm{E}_p[X] = \frac{2a}{2b-d-1}\Omega, \tag{36a}$$

$$\mathrm{E}_q[X] = \frac{V}{v-2d-2} = \frac{v-d-1}{v-2d-2}\frac{2a-d-1}{2b}\Omega. \tag{36b}$$

Equating the expected values and solving for v gives

$$v = (d+1)\frac{\frac{2a-d-1}{2b} - 2\frac{2a}{2b-d-1}}{\frac{2a-d-1}{2b} - \frac{2a}{2b-d-1}}. \tag{37}$$

□

3.4 Remarks to Theorem 1

The equations for V (27) and v (28) in Theorem 1 correspond to matching the expected value of X^{-1} and $\log|X|$,

$$\mathrm{E}_q\left[X^{-1}\right] = \mathrm{E}_p\left[X^{-1}\right], \tag{38a}$$

$$\mathrm{E}_q\left[\log|X|\right] = \mathrm{E}_p\left[\log|X|\right]. \tag{38b}$$

Notice that in Theorem 1, substituting a value for v into (27) gives the analytical solution for V. The parameter v can be found by applying a numerical root-finding algorithm to (28), see e.g. [2, Section 5.1]. Examples include Newton-Raphson or modified Newton algorithms, see e.g. [2, Section 5.4], for more alternatives see e.g. [2, Chapter 5]. In the following corollary, we supply an alternative to root-finding to obtain a value for v.

Matching the expected values, as in Corollary 1, can be seen as an approximation of matching the expected values of the log determinant (38b). Indeed, with numerical simulations one can show that the v given by (37) is approximately equal to the solution of (28), the difference is typically on the order of one tenth of a degree of freedom.

References [3, 1, 4] contain discussions about using moment matching to approximate a $\mathcal{GB}_d^{II}$-distribution with a $\mathcal{IW}_d$-distribution. Theorem 1 defines an approximation by minimising the KL divergence, which results in matching the expected values (38). The KL criterion is well-known in the literature for its moment-matching characteristics, see e.g. [5, 6].

4 Approximating the density of $\mathbb{V}_{\mathbf{x}}$ with a $\mathcal{W}_d$-distribution

This section shows how the distribution of a matrix valued function of the kinematical target state $\mathbf{x}$ can be approximated with a Wishart distribution.

4.1 Theorem 2

Theorem 2. *Let $\mathbf{x}$ be Gaussian distributed with mean m and covariance P, and let $\mathbb{V}_{\mathbf{x}} \triangleq \mathbb{V}(\mathbf{x}) \in \mathbb{S}_{++}^{n_x}$ be a matrix valued function of $\mathbf{x}$. Let $p(\mathbb{V}_{\mathbf{x}})$ be the density of $\mathbb{V}_{\mathbf{x}}$ induced by the random variable $\mathbf{x}$, and let $q(\mathbb{V}_{\mathbf{x}}) = \mathcal{W}_d(\mathbb{V}_{\mathbf{x}}\,; s, S)$*

be the minimizer of the KL*-divergence between* $p(\mathbb{V}_{\mathbf{x}})$ *and* $q(\mathbb{V}_{\mathbf{x}})$ *among all* $\mathcal{W}$*-distributions, i.e.*

$$q(\mathbb{V}_{\mathbf{x}}) \triangleq \underset{q(\,\cdot\,)=\mathcal{W}(\,\cdot\,)}{\arg\min} \;\; \mathrm{KL}\left(p\left(\mathbb{V}_{\mathbf{x}}\right)||q\left(\mathbb{V}_{\mathbf{x}}\right)\right). \tag{39}$$

Then S *is given as*

$$S = \frac{1}{s}\mathbb{C}_{II} \tag{40}$$

and s *is the solution to the equation*

$$d\log\left(\frac{s}{2}\right) - \sum_{i=1}^{d}\psi_0\left(\frac{s-i+1}{2}\right) + \mathbb{C}_I - \log|\mathbb{C}_{II}| = 0 \tag{41}$$

where $\mathbb{C}_I \triangleq \mathrm{E}\left[\log|\mathbb{V}_{\mathbf{x}}|\right]$ *and* $\mathbb{C}_{II} \triangleq \mathrm{E}\left[\mathbb{V}_{\mathbf{x}}\right]$. □

4.2 Proof of Theorem 2

The density $q\left(\mathbb{V}_{\mathbf{x}}\right)$ is

$$\begin{aligned}
q\left(\mathbb{V}_{\mathbf{x}}\right) =& \underset{q(\mathbb{V}_{\mathbf{x}})}{\arg\min}\ \mathrm{KL}\left(p\left(\mathbb{V}_{\mathbf{x}}\right)||q\left(\mathbb{V}_{\mathbf{x}}\right)\right) && \text{(42a)}\\
=& \underset{q(\mathbb{V}_{\mathbf{x}})}{\arg\max}\ \int p\left(\mathbb{V}_{\mathbf{x}}\right)\log\left(q\left(\mathbb{V}_{\mathbf{x}}\right)\right)d\mathbb{V}_{\mathbf{x}} && \text{(42b)}\\
=& \underset{q(\mathbb{V}_{\mathbf{x}})}{\arg\max}\ \int p\left(\mathbb{V}_{\mathbf{x}}\right)\Bigg[-\frac{sd}{2}\log 2 - \log\Gamma_d\left(\frac{s}{2}\right)\\
& -\frac{s}{2}\log|S| + \frac{s-d-1}{2}\log|\mathbb{V}_{\mathbf{x}}| + \mathrm{Tr}\left(-\frac{1}{2}S^{-1}\mathbb{V}_{\mathbf{x}}\right)\Bigg]d\mathbb{V}_{\mathbf{x}} && \text{(42c)}\\
=& \underset{q(\mathbb{V}_{\mathbf{x}})}{\arg\max}\ \int p\left(\mathbf{x}\right)\Bigg[-\frac{sd}{2}\log 2 - \log\Gamma_d\left(\frac{s}{2}\right)\\
& -\frac{s}{2}\log|S| + \frac{s-d-1}{2}\log|\mathbb{V}_{\mathbf{x}}| + \mathrm{Tr}\left(-\frac{1}{2}S^{-1}\mathbb{V}_{\mathbf{x}}\right)\Bigg]d\mathbf{x} && \text{(42d)}\\
=& \underset{q(\mathbb{V}_{\mathbf{x}})}{\arg\max}\ \mathrm{E}_{\mathbf{x}}\Bigg[-\frac{sd}{2}\log 2 - \log\Gamma_d\left(\frac{s}{2}\right) - \frac{s}{2}\log|S|\\
& + \frac{s-d-1}{2}\log|\mathbb{V}_{\mathbf{x}}| + \mathrm{Tr}\left(-\frac{1}{2}S^{-1}\mathbb{V}_{\mathbf{x}}\right)\Bigg] && \text{(42e)}\\
=& -\frac{sd}{2}\log 2 - \log\Gamma_d\left(\frac{s}{2}\right) - \frac{s}{2}\log|S|\\
& + \frac{s-d-1}{2}\mathrm{E}_{\mathbf{x}}\left[\log|\mathbb{V}_{\mathbf{x}}|\right] + \mathrm{Tr}\left(-\frac{1}{2}S^{-1}\,\mathrm{E}_{\mathbf{x}}\left[\mathbb{V}_{\mathbf{x}}\right]\right). && \text{(42f)}
\end{aligned}$$

Let $\mathbb{C}_I = \mathrm{E}_{\mathbf{x}}\left[\log|\mathbb{V}_{\mathbf{x}}|\right]$ and $\mathbb{C}_{II} = \mathrm{E}_{\mathbf{x}}\left[\mathbb{V}_{\mathbf{x}}\right]$. This results in

$$\begin{aligned}
q\left(\mathbb{V}_{\mathbf{x}}\right) =& \underset{q(\mathbb{V}_{\mathbf{x}})}{\arg\max}\ -\frac{sd}{2}\log 2 - \log\Gamma_d\left(\frac{s}{2}\right) - \frac{s}{2}\log|S| + \frac{s-d-1}{2}\mathbb{C}_I + \mathrm{Tr}\left(-\frac{1}{2}S^{-1}\mathbb{C}_{II}\right)\\
=& \underset{q(\mathbb{V}_{\mathbf{x}})}{\arg\max}\ f\left(s,S\right). && \text{(43)}
\end{aligned}$$

Differentiating the objective function $f(s, S)$ with respect to S, setting the result equal to zero and multiplying both sides by 2 gives

$$-sS^{-1} + S^{-1}\mathbb{C}_{II}S^{-1} = 0 \quad \Leftrightarrow \quad S = \frac{1}{s}\mathbb{C}_{II}. \tag{44}$$

Note that the expected value for $\mathbb{V}_{\mathbf{x}}$ under the Wishart distribution $q(\,\cdot\,)$ is $sS = \mathbb{C}_{II}$. Thus the expected value under $q(\,\cdot\,)$ is correct regardless of the parameter s. Differentiating the objective function $f(s, S)$ in (43) with respect to s gives

$$\frac{df(s,S)}{ds} = -\frac{d}{2}\log 2 - \frac{1}{2}\sum_{i=1}^{d}\psi_0\left(\frac{s-i+1}{2}\right) - \frac{1}{2}\log|S| + \frac{1}{2}\mathbb{C}_I \tag{45}$$

$$= \frac{d}{2}\log\frac{s}{2} - \frac{1}{2}\sum_{i=1}^{d}\psi_0\left(\frac{s-i+1}{2}\right) - \frac{1}{2}\log|\mathbb{C}_{II}| + \frac{1}{2}\mathbb{C}_I \tag{46}$$

where we substituted S with (44) to obtain (46). Equating the result to zero and multiplying both sides by 2 gives (41) in Theorem 2.

4.3 Corollary to Theorem 2

Corollary 2. *$\mathbb{C}_I$ and $\mathbb{C}_{II}$ can be calculated using a Taylor series expansion of $\mathbb{V}_{\mathbf{x}}$ around $\mathbf{x} = m$. A third order expansion yields*

$$\mathbb{C}_I \approx \log|\mathbb{V}_m| + \sum_{i=1}^{n_x}\sum_{j=1}^{n_x} \left.\frac{d^2 \log|\mathbb{V}_{\mathbf{x}}|}{d\mathbf{x}_j d\mathbf{x}_i}\right|_{\mathbf{x}=m} P_{ij}, \tag{47a}$$

$$\mathbb{C}_{II} \approx \mathbb{V}_m + \sum_{i=1}^{n_x}\sum_{j=1}^{n_x} \left.\frac{d^2 \mathbb{V}_{\mathbf{x}}}{d\mathbf{x}_j d\mathbf{x}_i}\right|_{x=m} P_{ij}. \tag{47b}$$

In (47) the i:th element of the vector $\mathbf{x}$ and the i,j:th element of the matrix P are $\mathbf{x}_i$ and P_{ij}, respectively. Moreover, the matrix $\mathbb{V}_m$ is the function $\mathbb{V}_{\mathbf{x}}$ evaluated at the mean m of the random variable $\mathbf{x}$. □

The expected values taken via second order Taylor expansions of $\log|\mathbb{V}_{\mathbf{x}}|$ and $\mathbb{V}_{\mathbf{x}}$ around $\mathbf{x} = m$ are

$$\begin{aligned}
\mathrm{E}_{\mathbf{x}}[\log|\mathbb{V}_{\mathbf{x}}|] \approx \mathrm{E}_{\mathbf{x}}\Bigg[& \log|\mathbb{V}_m| + \sum_{i=1}^{n_x} \left.\frac{d\log|\mathbb{V}_{\mathbf{x}}|}{dx_i}\right|_{\mathbf{x}=m} (\mathbf{x}_i - m_i) \\
& + \sum_{i=1}^{n_x}\sum_{j=1}^{n_x} \left.\frac{d^2\log|\mathbb{V}_{\mathbf{x}}|}{d\mathbf{x}_j d\mathbf{x}_i}\right|_{\mathbf{x}=m} (\mathbf{x}_i - m_i)(\mathbf{x}_j - m_j)\Bigg] \\
= & \log|\mathbb{V}_m| + \sum_{i=1}^{n_x}\sum_{j=1}^{n_x} \left.\frac{d^2\log|\mathbb{V}_{\mathbf{x}}|}{d\mathbf{x}_j d\mathbf{x}_i}\right|_{\mathbf{x}=m} P_{ij} && \text{(48a)} \\
\triangleq & \mathbb{C}_I, && \text{(48b)}
\end{aligned}$$

and

$$
\begin{aligned}
\mathrm{E}_{\mathbf{x}}\left[\mathbb{V}_{\mathbf{x}}\right] \approx \mathrm{E}_{\mathbf{x}} \Bigg[& \mathbb{V}_m + \sum_{i=1}^{n_x} \left.\frac{d\mathbb{V}_{\mathbf{x}}}{d\mathbf{x}_i}\right|_{\mathbf{x}=m} \left(\mathbf{x}_i - m_i\right) \\
& + \sum_{i=1}^{n_x} \sum_{j=1}^{n_x} \left.\frac{d^2\mathbb{V}_{\mathbf{x}}}{d\mathbf{x}_j d\mathbf{x}_i}\right|_{\mathbf{x}=m} \left(\mathbf{x}_i - m_i\right)\left(\mathbf{x}_j - m_j\right) \Bigg] \\
= & \mathbb{V}_m + \sum_{i=1}^{n_x} \sum_{j=1}^{n_x} \left.\frac{d^2\mathbb{V}_{\mathbf{x}}}{d\mathbf{x}_j d\mathbf{x}_i}\right|_{\mathbf{x}=m} P_{ij} \qquad (49a) \\
\triangleq & \mathbb{C}_{II}. \qquad (49b)
\end{aligned}
$$

Note that this is equal to a third order Taylor series expansion, because the addition of the third order Taylor series terms would not change the results above because all odd central moments of the Gaussian density are zero. Hence, the error of the above approximation is of the order $\mathcal{O}\left(E_p\|\mathbf{x}-m\|^4\right)$, i.e. the error of a third order Taylor series expansion.

4.4 Remarks to Theorem 2

The equations for S (40) and s (41) in Theorem 2 correspond to matching the expected values of $\mathbb{V}_{\mathbf{x}}$ and $\log|\mathbb{V}_{\mathbf{x}}|$,

$$
\begin{aligned}
\mathrm{E}_q\left[\mathbb{V}_{\mathbf{x}}\right] &= \mathrm{E}_p\left[\mathbb{V}_{\mathbf{x}}\right], \qquad (50a) \\
\mathrm{E}_q\left[\log|\mathbb{V}_{\mathbf{x}}|\right] &= \mathrm{E}_p\left[\log|\mathbb{V}_{\mathbf{x}}|\right]. \qquad (50b)
\end{aligned}
$$

Similarly to (28), numerical root-finding can be used to calculate a solution to (41). Note that using a moment matching approach similar to Corollary 1 to find a value for s is not advisable, since this would lead to further approximations (because the true distribution $p(\mathbb{V}_{\mathbf{x}})$ is unknown), and would possibly require a more complicated numerical solution. For $s > d-1$ and any $S \in \mathbb{S}_{++}^d$ there is a unique root to (41).

4.5 Proof of unique root to (41)

To prove that (41) has a unique solution, we will first show that the optimization function (43), which is what leads to (41), is strictly concave. From the definition of the Wishart distribution we have $s > d-1$ so the problem at hand is to show that (43) is strictly concave with respect to s for $s > d-1$ and for any $S \in \mathbb{S}_{++}^d$. Let $S \in \mathbb{S}_{++}^d$ be arbitrary and define the function $q(s)$ as

$$
q(s) = -\left(\mathbb{A}\frac{s}{2} + \log\Gamma_d\left(\frac{s}{2}\right)\right) + \mathbb{B} \qquad (51)
$$

where $\mathbb{A}$ and $\mathbb{B}$, defined as

$$
\begin{aligned}
\mathbb{A} &= \log(2) + \log|S| - \mathbb{C}_I, \qquad (52) \\
\mathbb{B} &= -\frac{d+1}{2}\mathbb{C}_I + \mathrm{Tr}\left(-\frac{1}{2}S^{-1}\mathbb{C}_{II}\right), \qquad (53)
\end{aligned}
$$

are constants with respect to s. Note that the equation under investigation, i.e. (41), is nothing but the equation

$$\frac{\mathrm{d}}{\mathrm{d}s} q(s) = 0. \tag{54}$$

A second order condition for strict concavity of a function is that the second order derivative is < 0 on the function's domain, see e.g. [7, Section 3.1.4]. We thus need $\mathrm{d}^2 q(s)/\mathrm{d}s^2 < 0$ for $s > d-1$. The second order derivative of $q(s)$ is

$$\frac{\mathrm{d}^2 q(s)}{\mathrm{d}s^2} = -\frac{1}{4}\sum_{i=1}^{d}\sum_{k=0}^{\infty}\frac{1}{\left(\left(\frac{s}{2}-\frac{i-1}{2}\right)+k\right)^2} \tag{55}$$

where we have used the series expansion of the second order derivative of the function $\log(\Gamma(s))$, see e.g., [8, Equation (1)]. For all $s > d-1$ we have $\mathrm{d}^2 q(s)/\mathrm{d}s^2 < 0$, and thus the function $q(s)$ is strictly concave.

It now easy to see that

$$\lim_{s \to d-1} q(s) = -\infty. \tag{56}$$

Similarly we have

$$\lim_{s \to \infty} q(s) = -\infty \tag{57}$$

since $\log\Gamma(s)$ (and hence $\log\Gamma_d(s)$) grows much faster than s as s goes to infinity. Moreover, the function $q(s)$ is both differentiable and bounded from above in the interval $(d-1, \infty)$. Therefore, we can conclude that there exists a local maximum of the function $q(s)$ in the interval $(d-1,\infty)$ where (41) must be satisfied. This local maximum is unique due to strict concavity.

5 Marginalising $\mathcal{IW}(X|V)\mathcal{W}(V)$ over V

This section presents a result that is similar to the following property [1, Problem 5.33]: if $p(S|\Sigma) = \mathcal{W}_d(S\,;\, n, \Sigma)$ and $p(\Sigma) = \mathcal{IW}_d(\Sigma\,;\, m, \Psi)$ then the marginal density of S is

$$p(S) = \mathcal{GB}_d^{II}\left(X;\ \frac{n}{2},\ \frac{m-d-1}{2}, \Psi,\ \mathbf{0}_d\right). \tag{58}$$

Theorem 3. *Let $p(X|V) = \mathcal{IW}_d(X\,;\, v, V/\gamma)$ and let $p(V) = \mathcal{W}_d(V\,;\, s, S)$. The marginal for X is*

$$p(X) = \mathcal{GB}_d^{II}\left(X;\ \frac{s}{2},\ \frac{v-d-1}{2}, \frac{S}{\gamma},\ \mathbf{0}_d\right). \tag{59}$$

□

5.1 Proof of Theorem 3

We have $p(X)$ given as

$$
\begin{aligned}
p(X) &= \int p(X|V)p(V)\mathrm{d}V = \int \mathcal{IW}_d\left(X\,;\,v,\frac{V}{\gamma}\right)\mathcal{W}_d\left(V\,;\,s,S\right)\mathrm{d}V && \text{(60a)}\\
&= \int \left\{2^{\frac{(v-d-1)d}{2}}\Gamma_d\left(\frac{v-d-1}{2}\right)|X|^{\frac{v}{2}}\right\}^{-1}\left|\frac{V}{\gamma}\right|^{(v-d-1)/2}\mathrm{etr}\left(-0.5X^{-1}\frac{V}{\gamma}\right)\\
&\quad\times\left\{2^{\frac{sd}{2}}\Gamma_d\left(\frac{s}{2}\right)|S|^{\frac{s}{2}}\right\}^{-1}|V|^{\frac{s-d-1}{2}}\,\mathrm{etr}(-0.5S^{-1}V)\mathrm{d}V && \text{(60b)}\\
&= \left\{\Gamma_d\left(\frac{v-d-1}{2}\right)\Gamma_d\left(\frac{s}{2}\right)|X|^{\frac{v}{2}}|S|^{\frac{s}{2}}\right\}^{-1}2^{-\frac{(v+s-d-1)d}{2}}\gamma^{-\frac{(v-d-1)d}{2}}\\
&\quad\times\int|V|^{\frac{v+s-2d-2}{2}}\,\mathrm{etr}\left(-0.5\left(\frac{X^{-1}}{\gamma}+S^{-1}\right)V\right)\mathrm{d}V && \text{(60c)}\\
&= \left\{\Gamma_d\left(\frac{v-d-1}{2}\right)\Gamma_d\left(\frac{s}{2}\right)|X|^{\frac{v}{2}}|S|^{\frac{s}{2}}\right\}^{-1}2^{-\frac{(v+s-d-1)d}{2}}\gamma^{-\frac{(v-d-1)d}{2}}\\
&\quad\times 2^{\frac{sd}{2}}\Gamma_d\left(\frac{v+s-d-1}{2}\right)\left|\left(\frac{X^{-1}}{\gamma}+S^{-1}\right)^{-1}\right|^{\frac{v+s-d-1}{2}}\\
&\quad\times\int\mathcal{W}_d\left(V\,;\,v+s-d-1,\frac{X^{-1}}{\gamma}+S^{-1}\right)\mathrm{d}V && \text{(60d)}\\
&= \left\{\Gamma_d\left(\frac{v-d-1}{2}\right)\Gamma_d\left(\frac{s}{2}\right)|X|^{\frac{v}{2}}|S|^{\frac{s}{2}}\right\}^{-1}2^{-\frac{(v+s-d-1)d}{2}}\gamma^{-\frac{(v-d-1)d}{2}}\\
&\quad\times 2^{\frac{(v+s-d-1)d}{2}}\Gamma_d\left(\frac{v+s-d-1}{2}\right)\left|(\gamma X)^{-1}+S^{-1}\right|^{-\frac{v+s-d-1}{2}} && \text{(60e)}\\
&= \left\{\Gamma_d\left(\frac{v-d-1}{2}\right)\Gamma_d\left(\frac{s}{2}\right)|X|^{\frac{v}{2}}|S|^{\frac{s}{2}}\right\}^{-1}\gamma^{-\frac{(v-d-1)d}{2}}\\
&\quad\times\Gamma_d\left(\frac{v+s-d-1}{2}\right)\left|X^{-1}\left(\frac{S}{\gamma}+X\right)S^{-1}\right|^{-\frac{v+s-d-1}{2}} && \text{(60f)}\\
&= \frac{\Gamma_d\left(\frac{s+v-d-1}{2}\right)}{\Gamma_d\left(\frac{v-d-1}{2}\right)\Gamma_d\left(\frac{s}{2}\right)}\gamma^{-\frac{(v-d-1)d}{2}}\frac{\left(\left|X^{-1}\right|\left|\frac{S}{\gamma}+X\right|\left|S^{-1}\right|\right)^{-\frac{v+s-d-1}{2}}}{|X|^{\frac{v}{2}}|S|^{\frac{s}{2}}} && \text{(60g)}\\
&= \frac{1}{\beta_d\left(\frac{s}{2},\frac{v-d-1}{2}\right)}\gamma^{-\frac{(v-d-1)d}{2}}\frac{\left|\frac{S}{\gamma}+X\right|^{-\frac{v+s-d-1}{2}}}{|X|^{\frac{s-d-1}{2}}|S|^{\frac{v-d-1}{2}}} && \text{(60h)}\\
&= \frac{|X|^{\frac{s-d-1}{2}}\left|X+\frac{S}{\gamma}\right|^{-\frac{s+v-d-1}{2}}}{\beta_d\left(\frac{s}{2},\frac{v-d-1}{2}\right)\left|\frac{S}{\gamma}\right|^{\frac{v-d-1}{2}}} && \text{(60i)}
\end{aligned}
$$

which, by [1, Theorem 5.2.2], is the probability density function for

$$
\mathcal{GB}_d^{II}\left(X;\ \frac{s}{2},\ \frac{v-d-1}{2},\frac{S}{\gamma},\ \mathbf{0}_d\right). \tag{61}
$$

References

[1] A. K. Gupta and D. K. Nagar, *Matrix variate distributions*, ser. Chapman & Hall/CRC monographs and surveys in pure and applied mathematics. Chapman & Hall, 2000.

[2] J. Stoer and R. Bulirsch, *Introduction to Numerical Analysis*, 2nd ed. New York: Springer-Verlag, 1993.

[3] J. W. Koch, "Bayesian approach to extended object and cluster tracking using random matrices," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 44, no. 3, pp. 1042–1059, Jul. 2008.

[4] J. Lan and X.-R. Li, "Tracking of extended object or target group using random matrix – part I: New model and approach," in *Proceedings of the International Conference on Information Fusion*, Singapore, Jul. 2012, pp. 2177–2184.

[5] C. M. Bishop, *Pattern recognition and machine learning*. New York, USA: Springer, 2006.

[6] T. Minka, "A family of algorithms for approximate Bayesian inference," Ph.D. dissertation, Massachusetts Institute of Technology, Jan. 2001.

[7] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.

[8] M. Merkle, "Logarithmic convexity and inequalities for the gamma function," *Journal of Mathematical Analysis and Applications*, no. 203, pp. 369–380, 1996.

Avdelning, Institution
Division, Department

Division of Automatic Control
Department of Electrical Engineering

Datum
Date

2011-12-16

Språk
Language

☐ Svenska/Swedish
☒ Engelska/English
☐ ______

Rapporttyp
Report category

☐ Licentiatavhandling
☐ Examensarbete
☐ C-uppsats
☐ D-uppsats
☒ Övrig rapport
☐ ______

ISBN
—

ISRN
—

Serietitel och serienummer
Title of series, numbering

ISSN
1400-3902

LiTH-ISY-R-3042

URL för elektronisk version

`http://www.control.isy.liu.se`

Titel
Title

Properties and approximations of some matrix variate probability density functions

Författare
Author

Karl Granström, Umut Orguner

Sammanfattning
Abstract

This report contains properties and approximations of some matrix valued probability density functions. Expected values of functions of generalised Beta type II distributed random variables are derived. In two Theorems, approximations of matrix variate distributions are derived. A third theorem contain a marginalisation result.

Nyckelord
Keywords

Extended target, random matrix, Kullback-Leibler divergence, inverse Wishart, Wishart, generalized Beta.