GENERALIZED INVERSE GAUSSIAN DISTRIBUTIONS AND THEIR MANOVA CONNECTIONS

Ronald W. Butler[1]

Department of Statistics

Colorado State University

Fort Collins, CO 80523

February 1, 1996

[1]Research supported by National Science Foundation Grant DMS-93.04274

Abstract

The matrix generalized inverse Gaussian distribution (MGIG) is shown to arise as a conditional distribution of components of a Wishart distribution. In the special scalar case, the characterization refers to members of the class of generalized inverse Gaussian distributions (GIGs) and includes the inverse Gaussian distribution among others.

Keywords and phrases. Generalized hyperbolic distribution, generalized inverse Gaussian distribution, inverse Gaussian distribution, matrix generalized hyperbolic distribution, matrix generalized inverse Gaussian distribution

AMS 1991 *subject classifications. Primary:* 62H05,62E10; *Secondary:* 62H10,62E15

Running title: GENERALIZED INVERSE GAUSSIANS

1 Introduction

The matrix generalized inverse Gaussian (MGIG) distribution was introduced in Barndorff-Nielsen et.al.(1982) as a distribution over the space of symmetric $(p \times p)$ positive definite matrices $\{W : W > 0\}$. Suppose Φ and Ψ are symmetric non-negative definite matrices $(\Phi \geq 0, \Psi \geq 0)$ and $\lambda \in \Re$. Then the $\mathrm{MGIG}_p(\Phi, \Psi, \lambda)$ density is

$$f(W) = a_p(\Phi, \Psi, \lambda)^{-1} |W|^{\lambda - \frac{1}{2}(p+1)} \operatorname{etr}\left[-\frac{1}{2}\left(\Phi W^{-1} + \Psi W\right)\right] 1_{W>0} \tag{1}$$

where the norming constant can be expressed in terms of a matrix Bessel function of the second kind, B_λ in Herz(1955, sec. 5), as

$$a_p(\Phi, \Psi, \lambda) = 2^{-p\lambda} |\Phi|^\lambda B_\lambda\left(\frac{1}{4}\Psi\Phi\right).$$

The domain of variation for parameters Φ and Ψ with a fixed value of λ is

$$\left.\begin{array}{lll} \{\Phi \geq 0, \Psi > 0\} & \text{if} & \lambda \geq \frac{1}{2}p \\ \{\Phi > 0, \Psi > 0\} & \text{if} & -\frac{1}{2}(p-1) \leq \lambda < \frac{1}{2}p \\ \{\Phi > 0, \Psi \geq 0\} & \text{if} & \lambda < -\frac{1}{2}(p-1) \end{array}\right\} \quad (p \geq 2). \tag{2}$$

Convergence of the density for any $\lambda \in \Re$ when $\Phi > 0, \Psi > 0$ is noted by Herz(1955, p.506). For the setting in which $\Phi = 0$, the density is Wishart and requires $\lambda \geq \frac{1}{2}p$ for integrability. When $\Psi = 0$, the density is inverted Wishart requiring $\lambda < \frac{1}{2}(1-p)$ for integrability.

The special scalar case $p = 1$ is the generalized inverse Gaussian distribution GIG(ϕ, ψ, λ) extensively studied by Jørgensen(1982). Its density has the form in (1) with norming constant

$$a_1(\phi, \psi, \lambda) = 2K_\lambda\left(\sqrt{\phi\psi}\right)(\phi/\psi)^{\frac{\lambda}{2}} \tag{3}$$

where K_λ is the standard notation of Abramowitz & Stegun(1972) for the modified Bessel function of the third kind. Its domain of variation for parameters is:

$$\begin{aligned} &\{\phi \geq 0, \psi > 0\} \quad \text{if} \quad \lambda > 0 \\ &\{\phi > 0, \psi > 0\} \quad \text{if} \quad \lambda = 0 \\ &\{\phi > 0, \psi \geq 0\} \quad \text{if} \quad \lambda < 0, \end{aligned} \tag{4}$$

which is not exactly (2) when evaluated at $p = 1$.

We show how members of both of these classes of distributions are characterized as conditional distributions of components of a Wishart distribution. More specifically, if

$$W = \begin{pmatrix} W_{xx} & W_{xy} \\ W_{yx} & W_{yy} \end{pmatrix}$$

is Wishart, then the conditional distribution of W_{xx} given W_{xy} is either MGIG or GIG as specified in Theorems 1 and 2 and Corollaries 1-3. These characterizations also arise when based on the normal variates that underlie the Wishart distribution. Three important special cases of GIG(ϕ, ψ, λ) admit to characterization when W_{xx} is a scalar. The first of these is the inverse Gaussian(ϕ, ψ) or GIG$(\phi, \psi, -\frac{1}{2})$

distribution. Chhikara & Folks(1989), Seshadri(1993), and Johnson, Kotz & Balakrishnan(1994, chap.15) give extensive discussion of this distribution. The other two characterized are the positive hyperbolic(ϕ, ψ) or GIG($\phi, \psi, 1$) distribution, and the hyperbola or GIG($\phi, \psi, 0$) distribution, both studied in Barndorff-Nielsen(1978) and Barndorff-Nielsen & Blæsild(1980).

The marginal distribution of W_{xy} is shown to have a variance-mean mixture distribution as described in Barndorff-Nielsen, Kent & Sørensen(1982). When W_{xy} is either a row or column vector, its distribution is a generalized hyperbolic distribution (Barndorff-Nielsen,1978); if it is a matrix then we introduce its distribution as a matrix generalized hyperbolic distribution.

The conditional distribution of W_{xx} given W_{xy} when W is either inverted Wishart or MGIG(Φ, Ψ, λ) with $\Phi \neq 0$ is generally not in the MGIG class of distributions but forms a new class containing the MGIG class.

2 Matrix Generalized Inverse Gaussian

Suppose W has a Wishart$_{p+q}(N,\Sigma)$ distribution on the space of $(p+q) \times (p+q)$ symmetric positive definite ($W > 0$) matrices with density

$$f(W) = b(p+q, N)\, |\Sigma|^{-\frac{1}{2}N}\, |W|^{\frac{1}{2}(N-p-q-1)}\, \mathrm{etr}\left(-\frac{1}{2}\Sigma^{-1}W\right) 1_{W>0} \qquad (5)$$

where $b(p+q,N)$ is the norming constant, $\Sigma>0$ and symmetric, and $N\geq p+q$ and real-valued. Partition W and Σ conformably as

$$W=\begin{pmatrix} W_{xx} & W_{xy} \\ W_{yx} & W_{yy} \end{pmatrix}, \qquad \Sigma=\begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix}$$

where W_{xx} and Σ_{xx} are $(p\times p)$, W_{xy} and Σ_{xy} are $(p\times q)$, etc.

Theorem 1 . *If W is Wishart$_{p+q}(N,\Sigma)$ with $N\geq p+q$ and real-valued, and $\Sigma>0$, then the conditional distribution of W_{xx} given W_{xy} is $MGIG_p[W_{xy}\Sigma_{yy.x}^{-1}W_{yx},\Sigma_{xx.y}^{-1},\frac{1}{2}(N-q)]$ where $\Sigma_{yy.x}=\Sigma_{yy}-\Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}$ and $\Sigma_{xx.y}=\Sigma_{xx}-\Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}$.*

Proof. Start with the density of W and compute the marginal density of (W_{xx},W_{xy}) as follows. Use the block inverse of Σ as

$$\Sigma^{-1}=\begin{pmatrix} \Sigma_{xx.y}^{-1} & -\Sigma_{xx}^{-1}\Sigma_{xy}\Sigma_{yy.x}^{-1} \\ -\Sigma_{yy.x}^{-1}\Sigma_{yx}\Sigma_{xx}^{-1} & \Sigma_{yy.x}^{-1} \end{pmatrix},$$

and the identity $|W|=|W_{xx}|\times|W_{yy.x}|$, where $W_{yy.x}=W_{yy}-W_{yx}W_{xx}^{-1}W_{xy}$, to write the density in (5) as

$$\begin{aligned} f(W) \;=\; & b(p+q,N)\,|\Sigma|^{-\frac{1}{2}N}\left(|W_{xx}|\,|W_{yy.x}|\right)^{\frac{1}{2}(N-p-q-1)} \\ & \operatorname{etr}\left[-\frac{1}{2}\left(\Sigma_{xx.y}^{-1}W_{xx}+\Sigma_{yy.x}^{-1}W_{yy}\right)+\Sigma_{xx}^{-1}\Sigma_{xy}\Sigma_{yy.x}^{-1}W_{yx}\right]1_{W>0}. \end{aligned} \tag{6}$$

The integration of this expression dW_{yy} over $\{W_{yy}:W>0\}$ with (W_{xx},W_{xy}) fixed is the same as the integration $dW_{yy.x}$ over $\{W_{yy.x}>0\}$ with (W_{xx},W_{xy}) fixed. Upon substitution $W_{yy} = W_{yy.x} + W_{yx}W_{xx}^{-1}W_{xy}$ into (6) we get

$$\begin{aligned} f(W_{xx}, W_{xy}) &= b(p+q, N)\,|\Sigma|^{-\frac{1}{2}N}\,|W_{xx}|^{\frac{1}{2}(N-p-q-1)} \\ &\quad \text{etr}\left[-\frac{1}{2}\left(\Sigma_{xx.y}^{-1}W_{xx} + \Sigma_{yy.x}^{-1}W_{yx}W_{xx}^{-1}W_{xy}\right) + \Sigma_{xx}^{-1}\Sigma_{xy}\Sigma_{yy.x}^{-1}W_{yx}\right] \\ &\quad \int_{W_{yy.x}>0} |W_{yy.x}|^{\frac{1}{2}(N-p-q-1)}\,\text{etr}\left(-\frac{1}{2}\Sigma_{yy.x}^{-1}W_{yy.x}\right) dW_{yy.x}\,. \end{aligned}$$

The integral is $b(q, N-p)^{-1}\,|\Sigma_{yy.x}|^{\frac{1}{2}(N-p)}$ and the result follows when we condition upon W_{xy} as fixed. ■

All members of the $\text{MGIG}_p\,(\Phi, \Psi, \lambda)$ family of distributions do not have this Wishart characterization. To admit to such characterization, parameter λ must be restricted to $\lambda \geq \frac{1}{2}p$ in order to fulfill the requirement of a full rank Wishart distribution, $N \geq p+q$. Matrix Ψ must be of full rank p. Finally, along with this, the rank of Φ can be completely arbitrary in the range $\{0, ..., p\}$; this follows by taking q as the rank of Φ and finding a $(p \times q)$ matrix W_{xy} of rank q such that $\Phi = W_{xy}\Sigma_{yy.x}^{-1}W_{yx}$. Thus we have that the $\text{MGIG}_p\,(\Phi, \Psi, \lambda)$ distributions with $\lambda \geq \frac{1}{2}p$, full rank Ψ, and arbitrary rank Φ are characterized.

The settings of negative powers for λ and non-full rank Ψ are characterized by inverting W_{xx}.

Corollary 1 . *The conditional distribution of* W_{xx}^{-1} *given* W_{xy} *is* $MGIG_p[\Sigma_{xx.y}^{-1}, W_{xy}\Sigma_{yy.x}^{-1}W_{yx}, -\frac{1}{2}(N-q)]$.

Proof. . Transform $W_{xx} \to W_{xx}^{-1}$ using the Jacobian $|W_{xx}^{-1}|^{-(p+1)}$. ■

This provides a characterization of the $\mathrm{MGIG}_p(\Phi, \Psi, \lambda)$ distribution with $\lambda \leq -\frac{1}{2}p$, Φ of full rank p, and Ψ of arbitrary rank $q \in \{0, ..., p\}$. We summarize the subclass of the $\mathrm{MGIG}_p(\Phi, \Psi, \lambda)$ family that can be characterized as conditional distributions involving Wishart components in the first two lines:

$$\begin{aligned}
W_{xx}|W_{xy} &\sim \mathrm{MGIG}_p(\Phi, \Psi, \lambda) \quad \forall_{\Phi \geq 0}, \forall_{\Psi > 0} \quad \text{if} \quad \lambda \geq \tfrac{1}{2}p \\
W_{xx}^{-1}|W_{xy} &\sim \mathrm{MGIG}_p(\Phi, \Psi, \lambda) \quad \forall_{\Phi > 0} \forall_{\Psi \geq 0} \quad \text{if} \quad \lambda \leq -\tfrac{1}{2}p \\
X^T X \,|\, X^T Y &\sim \mathrm{MGIG}_p(\Phi, \Psi, \lambda) \quad \forall_{\Phi > 0} \forall_{\Psi > 0} \quad \text{if} \quad \lambda \in \{0, \pm\tfrac{1}{2}, \pm 1, ...\}
\end{aligned} \tag{7}$$

The characterizations of MGIG_p in the last line are derived from a different approach using the multivariate normal variables that underlie the Wishart distribution used in the first two lines. The use of such normal variates gives additional characterizations when $\lambda \in \left\{0, \pm\frac{1}{2}, ..., \pm\frac{1}{2}(p-1)\right\}$ and thus allows for the relaxation of the constraint $N \geq p + q$ in Theorem 1. The restriction to half-integer values of λ, however, is an unavoidable consequence of the dimensionality of the normals involved. The following result connects the $\mathrm{MGIG}_p(\Phi, \Psi, \lambda)$ distribution to multivariate normal variables.

Theorem 2 . *Suppose X and Y are $(N \times p)$ and $(N \times q)$ respectively and the N rows of (X, Y) are i.i.d. $N_{p+q}(0, \Sigma)$. If $N \geq p$, then the conditional distribution of $X^T X$ given $X^T Y$ is $MGIG_p[X^T Y \Sigma_{yy.x}^{-1} Y^T X, \Sigma_{xx.y}^{-1}, \frac{1}{2}(N-q)]$. The conditional distribution of $(X^T X)^{-1}$ given $X^T Y$ is $MGIG_p[\Sigma_{xx.y}^{-1}, X^T Y \Sigma_{yy.x}^{-1} Y^T X, -\frac{1}{2}(N-q)]$.*

Proof. . The conditional distribution of X^TY given X is $\mathrm{N}_{pq}(X^TX\Sigma_{xx}^{-1}\Sigma_{xy}, \Sigma_{yy.x} \otimes X^TX)$ where the first entry is the $(p \times q)$ mean and the second entry is the covariance of $\mathrm{vec}(X^TY)$, the stacked columns of X^TY. The conditional dependence on X is through X^TX and, since the conditional distribution of X given X^TX is uniform over $\{X : X^TX \text{ is fixed}\}$, the conditional distribution of X^TY given X^TX is the same normal distribution. The marginal distribution of X^TX is $\mathrm{Wishart}_p(N, \Sigma_{xx})$ when $N \geq p$, so the joint density of X^TY and X^TX can be written down. This appears in Appendix A where the conditional distribution of X^TX given X^TY is shown to be MGIG. ■

The MGIG settings in the last row of (7) are characterized by Theorem 2. For fixed p, choose $q \geq p$ and choose $N \geq p$ such that $\frac{1}{2}(N-q)$ is any value of $\lambda \in \{0, \pm\frac{1}{2}, \pm 1, ...\}$.

3 Generalized Inverse Gaussian

The $\mathrm{GIG}(\phi, \psi, \lambda)$ distribution occurs in the scalar case with $p = 1$ where w_{xx} and x^Tx are scalars and w_{xy} and x^TY are $(1 \times q)$ vectors. We catalog the characterizations of distributions from this class in terms of Wishart and normal components.

Corollary 2 . *In the Wishart setting of Theorem 1 with $p = 1$, the conditional distribution of w_{xx} given w_{xy} is $GIG\Big[w_{xy}\Sigma_{yy.x}^{-1}w_{yx}, \Sigma_{xx.y}^{-1}, \frac{1}{2}(N-q)\Big]$. The following subclass of $GIG(\phi, \psi, \lambda)$ admits to characterization:*

$$w_{xx}|w_{xy} \sim GIG(\phi,\psi,\lambda) \quad \forall_{\phi\geq 0},\forall_{\psi>0} \quad \textit{if} \quad \lambda \geq \tfrac{1}{2}$$

$$w_{xx}^{-1}|w_{xy} \sim GIG(\phi,\psi,\lambda) \quad \forall_{\phi>0},\forall_{\psi\geq 0} \quad \textit{if} \quad \lambda \leq -\tfrac{1}{2}.$$

The $\lambda = 1$ case is the positive hyperbolic distribution and the $\lambda = -\frac{1}{2}$ case is the inverse Gaussian distribution.

In the normal setting of Theorem 2 with x as $(N \times 1)$ and Y as $(N \times q)$, the conditional distribution of $x^T x$ given $x^T Y$ is $GIG\left[x^T Y \Sigma_{yy.x}^{-1} Y^T x, \Sigma_{xx.y}^{-1}, \frac{1}{2}(N-q)\right]$. This characterizes the entire subclass with half integer values for λ :

$$x^T x|x^T Y \sim GIG(\phi,\psi,\lambda) \quad \textit{for} \quad \lambda \in \left\{\tfrac{1}{2},1,...\right\} \qquad \textit{and} \quad \phi \geq 0, \psi > 0$$

$$x^T x|x^T Y \sim GIG(\phi,\psi,0) \quad \textit{for} \quad \lambda = 0 \qquad \textit{and} \quad \phi > 0, \psi > 0$$

$$x^T x|x^T Y \sim GIG(\phi,\psi,\lambda) \quad \textit{for} \quad \lambda \in \{-\tfrac{1}{2},-1,...\} \quad \textit{and} \quad \phi > 0, \psi \geq 0.$$

The list includes the inverse Gaussian ($\lambda = -\frac{1}{2}$), hyperbola ($\lambda = 0$), and hyperbolic ($\lambda = 1$) distributions.

Simple geometrical interpretations for each of the special distributions result from Corollary 2 by taking $N = 1$.

Corollary 3 . *Suppose $x, y_1, y_2,$and y_3 are i.i.d. $N(0,\sigma^2)$ variates. Then in $\Re^2, \Re^3$, and $\Re^4$,*

$$x^2|\, xy_1 \sim \textit{hyperbola or } GIG[\sigma^{-2}(xy_1)^2, \sigma^{-2}, 0]$$

$$x^2|\, xy_1, xy_2 \sim \textit{inverse Gaussian or } GIG[\sigma^{-2}x^2(y_1^2+y_2^2), \sigma^{-2}, -\tfrac{1}{2}]$$

$$x^{-2}|\, xy_1, xy_2, xy_3 \sim \textit{positive hyperbolic or } GIG[\sigma^{-2}, \sigma^{-2}x^2(y_1^2+y_2^2+y_3^2), 1].$$

The conditioning here is along hyperbolic curves. In $\Re^2$ as in the first row, Blæsild(1979) conditions upon a different set of hyperbolic curves in the same context of a bivariate normal to get the generalized hyperbolic distribution.

4 Generalized Hyperbolic

The marginal density of W_{xy} can be expressed in terms of the norming constant for the MGIG distribution and is a matrix generalized hyperbolic distribution.

Corollary 4 . *If $N \geq q$ then the marginal density of W_{xy} on $\Re^{pq}$ is*

$$f(W_{xy}) = 2^{-\frac{1}{2}p(N+q)}\pi^{-\frac{1}{4}p(p+2q-1)}\left\{\prod_{j=1}^{p}\Gamma\left[\frac{1}{2}(N+1-j)\right]\right\}^{-1}|\Sigma_{xx}|^{-\frac{1}{2}N}|\Sigma_{yy.x}|^{-\frac{1}{2}p} \quad (8)$$
$$\operatorname{etr}\left(\Sigma_{xx}^{-1}\Sigma_{xy}\Sigma_{yy.x}^{-1}W_{yx}\right)a_p\left(W_{xy}\Sigma_{yy.x}^{-1}W_{yx},\Sigma_{xx.y}^{-1},\frac{1}{2}(N-q)\right),$$

a density we shall call a matrix generalized hyperbolic distribution. When $p=1$, the density of row vector w_{xy} on $\Re^q$ is given as

$$f(w_{xy}) = 2^{-\frac{1}{2}(N+q)+1}\pi^{-\frac{1}{2}q}\Gamma\left(\frac{N}{2}\right)^{-1}\Sigma_{xx}^{-\frac{1}{2}N}|\Sigma_{yy.x}|^{-\frac{1}{2}}$$
$$\exp\left(\Sigma_{xx}^{-1}\Sigma_{xy}\Sigma_{yy.x}^{-1}w_{yx}\right)K_{\frac{1}{2}(N-q)}\left(\sqrt{w_{xy}\Sigma_{yy.x}^{-1}w_{yx}\Sigma_{xx.y}^{-1}}\right)$$
$$\left(w_{xy}\Sigma_{yy.x}^{-1}w_{yx}\Sigma_{xx.y}\right)^{\frac{1}{4}(N-q)}.$$

The transformed vector $v_{xy} = w_{xy}\Sigma_{yy.x}^{-\frac{1}{2}}$ has a q-dimensional generalized hyperbolic distribution as in Barndorff-Nielsen, Kent & Sørensen(1982).

Proof. The density of W_{xy} follows from the computations in Appendix A when norming constants are retained.

The density for w_{xy} follows from (3). The transformed variable v_{xy} relates to the generalized hyperbolic density in equation (2.5) of Barndorff-Nielsen, Kent, & Sørensen by taking in their notation: $r = q$; $\mu = 0$; $u = x^T x \sim \text{Gamma}(\frac{1}{2}N, \frac{1}{2}\Sigma_{xx}^{-1})$ so that $\lambda = \frac{1}{2}N$, $\delta^2 = 0$ and $\kappa^2 = \Sigma_{xx}^{-1}$; $\beta = \Sigma_{xx}^{-1}\Sigma_{xy}\Sigma_{yy.x}^{-\frac{1}{2}}$; and $\Delta = I_q$. The assumption of $\delta^2 = 0$ presumes that the limiting value of $K_\lambda(x)$ as $x \downarrow 0$ is used in their (2.5) with the form

$$\lim_{x \downarrow 0} x^\lambda K_\lambda(x) = \Gamma(\lambda)\, 2^{\lambda - 1} \qquad (\lambda > 0)\,.$$

In addition, use $\alpha^2 = \kappa^2 + \beta\Delta\beta^T = \Sigma_{xx}^{-1} + \Sigma_{xx}^{-1}\Sigma_{xy}\Sigma_{yy.x}^{-1}\Sigma_{yx}\Sigma_{xx}^{-1} = \Sigma_{xx.y}^{-1}$. ■

The density in (8) is a normal variance-mean mixture distribution as discussed in Barndorff-Nielsen, Kent & Sørensen. It is the marginal distribution of $X^T Y$ when $X^T Y \mid X^T X \sim \text{N}_{pq}(X^T X \Sigma_{xx}^{-1}\Sigma_{xy}, \Sigma_{yy.x} \otimes X^T X)$ and the mixing distribution is $X^T X \sim \text{Wishart}_p(N, \Sigma_{xx})$.

5 Generalizing MGIG

If $W \sim \text{MGIG}_{p+q}(\Phi, \Psi, \lambda)$ then the conditional distribution of W_{xx} given W_{xy} is not in the MGIG class when $\Phi_{xx} \neq 0$ or $\Phi_{xy} \neq 0$ as, for example, when W is inverted Wishart.

Theorem 3 *Suppose* $W \sim MGIG_{p+q}(\Phi, \Psi, \lambda)$. *Then the conditional density of* $(p \times p)$ *block* W_{xx} *given* $(p \times q)$ *block* W_{xy} *is*

$$\begin{aligned} f(W_{xx}|W_{xy}) \quad \propto \quad & |W_{xx}|^{\lambda - \frac{1}{2}(p+q+1)} \operatorname{etr}\left[-\frac{1}{2} W_{xx} \Psi_{xx}\right] \\ & \operatorname{etr}\left\{-\frac{1}{2} W_{xx}^{-1}\left(\Phi_{xx} + W_{xy}\Psi_{yy}W_{yx}\right)\right\} \\ & a_q\left(\Phi_{yy} + W_{yx}W_{xx}^{-1}\Phi_{xx}W_{xx}^{-1}W_{xy} - 2W_{yx}W_{xx}^{-1}\Phi_{xy}, \Psi_{yy}, \lambda - \frac{1}{2}p\right) \end{aligned} \tag{9}$$

with proportionality in W_{xx} *and all matrix blocks for* Φ *and* Ψ *conformable with* W.

Proof. See Appendix B. ∎

When $\Phi_{xx} = 0 = \Phi_{xy}$, then this is proportional to the $\text{MGIG}_p\left(\Phi_{xx} + W_{xy}\Psi_{yy}W_{yx}, \Psi_{xx}, \lambda - \frac{1}{2}q\right)$ density.

A Appendix Proof of Theorem 2:

Denote $W_{xy} = X^T Y$ and $W_{xx} = X^T X$. Since $W_{xy}|W_{xx} \sim \text{N}_{pq}(W_{xx}\Sigma_{xx}^{-1}\Sigma_{xy}, \Sigma_{yy.x} \otimes W_{xx})$ and $W_{xx} \sim \text{Wishart}_p(N, \Sigma_{xx})$, the joint density of W_{xy}, W_{xx} is

$$\begin{aligned} f(W_{xy}, W_{xx}) \quad \propto \quad & |W_{xx}|^{-\frac{1}{2}q} \operatorname{etr}\left\{-\frac{1}{2}u^T\left(\Sigma_{yy.x} \otimes W_{xx}\right)^{-1} u\right\} \\ & |W_{xx}|^{\frac{1}{2}(N-p-1)} \operatorname{etr}\left(-\frac{1}{2}\Sigma_{xx}^{-1} W_{xx}\right) \end{aligned}$$

where $u = \operatorname{vec}(W_{xy} - W_{xx}\Sigma_{xx}^{-1}\Sigma_{xy})$. Using Lemma 2.2.3(iii) of Muirhead(1982), write

$$u^T\left(\Sigma_{yy.x} \otimes W_{xx}\right)^{-1} u = \operatorname{tr}\left\{\Sigma_{yy.x}^{-1}\left(W_{xy} - W_{xx}\Sigma_{xx}^{-1}\Sigma_{xy}\right)^T W_{xx}^{-1}\left(W_{xy} - W_{xx}\Sigma_{xx}^{-1}\Sigma_{xy}\right)\right\}$$

so the joint density can be expressed as

$$f(W_{xy}, W_{xx}) \;\propto\; |W_{xx}|^{\frac{1}{2}(N-p-q-1)} \operatorname{etr}\left\{-\frac{1}{2}\left[\Sigma_{xx}^{-1} + \Sigma_{xx}^{-1}\Sigma_{xy}\Sigma_{yy.x}^{-1}\Sigma_{yx}\Sigma_{xx}^{-1}\right] W_{xx}\right\}$$
$$\operatorname{etr}\left\{-\frac{1}{2}\left(W_{xy}\Sigma_{yy.x}^{-1}W_{yx}W_{xx}^{-1}\right) + \Sigma_{yy.x}^{-1}W_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}\right\}.$$

Fixing W_{xy} and noting that $\Sigma_{xx.y}^{-1} = \Sigma_{xx}^{-1} + \Sigma_{xx}^{-1}\Sigma_{xy}\Sigma_{yy.x}^{-1}\Sigma_{yx}\Sigma_{xx}^{-1}$ then $W_{xx}|W_{xy} \sim \text{MGIG}\left[W_{xy}\Sigma_{yy.x}^{-1}W_{yx}, \Sigma_{xx.y}^{-1}, \frac{1}{2}(N-q)\right]$.

B Appendix Proof of Theorem 3:

The proof is similar to that of Theorem 1 so we focus on those aspects that are different. We need to integrate the density of W in (1) over $\{W_{yy.x} > 0\}$ but first we write

$$\operatorname{etr}\left(-\frac{1}{2}\Phi W^{-1}\right) = \operatorname{etr}\left[-\frac{1}{2}\left(\Phi_{xx}W_{xx.y}^{-1} + \Phi_{yy}W_{yy.x}^{-1}\right) + \Phi_{xy}W_{yy.x}^{-1}W_{yx}W_{xx}^{-1}\right] \qquad (10)$$

and reexpress $W_{xx.y}^{-1}$ in terms of $W_{yy.x}$. Use

$$W_{xx.y}^{-1} = W_{xx}^{-1} + W_{xx}^{-1}W_{xy}W_{yy.x}^{-1}W_{yx}W_{xx}^{-1}$$

to reexpress $W_{xx.y}^{-1}$ in (10) so that

$$\operatorname{etr}\left(-\frac{1}{2}\Phi W^{-1}\right) \;=\; \operatorname{etr}\left(-\frac{1}{2}\Phi_{xx}W_{xx}^{-1}\right)$$
$$\operatorname{etr}\left[-\frac{1}{2}W_{yy.x}^{-1}\left(\Phi_{yy} + W_{yx}W_{xx}^{-1}\Phi_{xx}W_{xx}^{-1}W_{xy} - 2W_{yx}W_{xx}^{-1}\Phi_{xy}\right)\right].$$

Using this expression, write the integration of (1) as

$$
\begin{aligned}
f\left(W_{xx}, W_{xy}\right) \;\propto\;& \left|W_{xx}\right|^{\lambda-\frac{1}{2}(p+q+1)} \operatorname{etr}\left\{-\frac{1}{2}\left[W_{xx}^{-1}\left(\Phi_{xx}+W_{xy} \Psi_{yy} W_{yx}\right)+W_{xx} \Psi_{xx}\right]\right\} \\
& \int_{W_{yy.x}>0}\left|W_{yy.x}\right|^{\lambda-\frac{1}{2}(p+q+1)} \operatorname{etr}\left\{-\frac{1}{2} W_{yy.x} \Psi_{yy}\right\} \\
& \quad \operatorname{etr}\left\{-\frac{1}{2} W_{yy.x}^{-1}\left(\Phi_{yy}+W_{yx} W_{xx}^{-1} \Phi_{xx} W_{xx}^{-1} W_{xy}-2 W_{yx} W_{xx}^{-1} \Phi_{xy}\right)\right\} d W_{yy.x}.
\end{aligned}
$$

The integral is the norming constant of MGIG which gives the expression in (9).

References

Abramowitz, M. and Stegun, I. (1972). *Handbook of Mathematical Functions.* Dover, New York.

Barndorff-Nielsen, O. (1978). Hyperbolic distributions and distributions on hyperbolae. *Scand. J. Statist.* **5** 151-157.

Barndorff-Nielsen, O. and Blæsild, P. (1983). Hyperbolic distributions. *Encyclopedia of Statistical Sciences* **3**. Kotz, S., Johnson, N.L. & Read, C. (editors). Wiley, New York.

Barndorff-Nielsen, O., Blæsild, P., Jensen, J.L. and Jørgensen, B. (1982) . Exponential transformation models. *Proc. R. Soc. Lond.* **A 379** 41-65.

Barndorff-Nielsen, O., Kent, J. and Sørensen, M. (1982). Normal variance-mean mixtures and z distributions. *Int. Statist. Rev.* **50** 145-159.

Blæsild, P. (1979). Conditioning with conic sections in the two-dimensional normal distribution. *Ann. Statist.* **7** 659-670.

Chhikara, R.S. and Folks, J.L. (1989). *The Inverse Gaussian Distribution.* Marcel Dekker, New York.

Herz, C.S. (1955). Bessel functions of matrix argument. *Ann. Math.* **61** 474-523.

Johnson, N.L., Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions, Vol. 1.* Wiley, New York.

Jørgensen, B. (1982). Statistical Properties of the Generalized Inverse Gaussian Distribution, *Lecture Notes in Statistics No. 9.* Springer-Verlag, New York.

Muirhead, R.J. (1982). *Aspects of Multivariate Statistical Theory.* Wiley, New York.

Seshadri, V. (1993). *The Inverse Gaussian Distribution: A Case Study in Exponential Families.* Oxford University Press, New York.