A GENERALIZATION OF A VON MISES-FISHER CONDITIONAL DISTRIBUTION RESULT.

CARSTEN WIUF

Department of Statistics
Institute of Mathematics
University of Aarhus
Denmark

1: Introduction.

The purpose of this paper is to extend and unify some results given separately in the literature.

To be more specific we shall be concerned with orthogeodesic exponential composite transformation models and the similarity between the distribution $F(\chi, \psi)$ (χ denoting the group parameter and ψ the index parameter) of a variable x and the conditional distribution of the maximum likelihood estimator $\hat{\chi}$ of χ given the maximum likelihood estimator $\hat{\psi}$ of ψ under repeated sampling.

It is known that in some cases $\hat{\chi}|\hat{\psi}$ follows a $F(\chi, \phi)$ distribution with ϕ being some function of ψ, $\hat{\psi}$ and n (sampling size), i.e. the distribution of $\hat{\chi}|\hat{\psi}$ is of the same type as x.

This is for example the case if x_i, $i = 1, ..., n$ is a sample from the von Mises-Fisher distribution $C_d(\mu, \kappa)$. Then we have that the conditional distribution of the maximum likelihood estimate $\hat{\mu}$ ($= x./\|x.\|$) of the direction parameter μ given the resultant length $\|x.\|$ also is distributed according to a von Mises-Fisher distribution: $\hat{\mu} \,|\, \|x.\| \sim C_d(\mu, \|x.\|\kappa)$. However the maximum likelihood estimate $\hat{\kappa}$ of κ is in unique correspondence with the resultant length $\|x.\|$, and hence the distribution of $\hat{\mu}|\hat{\kappa}$ is given by $C_d(\mu, \|x.\|\kappa)$. A similar result holds if x is distributed according to a hyperboloid distribution (Jensen (1981)). Moreover a number of τ-parallel exponential composite transformation models (see Barndorff-Nielsen and Blæsild (1983a)) exhibit this property. Here (χ, ψ) equals (θ^1, τ_2) for some similar partitions (θ^1, θ^2) of the canonical parameter θ and (τ_1, τ_2) of the mean value parameter τ. In these cases

$\hat{\chi}$ and $\hat{\psi}$ are stochastical independent and we then have that $\hat{\chi}$ follows a $F(\chi, \phi)$-distribution if x follows a $F(\chi, \psi)$-distribution.

As proved in Barndorff-Nielsen and Blæsild (1983a) the result on τ-parallel models is not limited to composite transformation models. The conditions given in this paper to ensure stochastical independence of $\hat{\chi}$ and $\hat{\psi}$ are (slightly) different from those given in Barndorff-Nielsen and Blæsild (1983a).

We will give two different but similar sets of conditions that ensure a similarity of the distributions of $\hat{\chi}|\hat{\psi}$ and x. The first set of conditions is the one that resembles the conditions in Barndorff-Nielsen and Blæsild (1983a), and the other set is easily seen to be fulfilled for both the von Mises-Fisher model and the hyperboloid model. The proofs of the distribution similarity under the two sets of conditions are quite similar to each other.

In Barndorff-Nielsen (1988) it is shown that the conditional distribution of $\hat{\chi}$ given $\hat{\psi}$, under repeated sampling and under some conditions similar to those given in the present paper is of the same type (in some sense) for all n (sampling size). This should be compared to the result given here, however we are mainly interested (in contrast to Barndorff-Nielsen (1988)) in the connection between the distribution of x and the distribution of $\hat{\chi}|\hat{\psi}$, not just the distribution of $\hat{\chi}|\hat{\psi}$. Moreover 'the same type' will have a more strict meaning in the present paper, than in Barndorff-Nielsen (1988). This will be commented on in the example in the end of the paper.

Finally we will derive expressions for the distribution of the m.l.e. $\hat{\psi}$ of ψ under the two sets of conditions.

2: Preliminaries and Basic Assumptions.

This section contains an introduction to the models of interest in this paper: orthogeodesic exponential composite transformation models (OECT-Models).

Group theoretical concepts will be assumed known (see e.g. Barndorff-Nielsen et al. (1989)), and similarly we will not go into a discussion of the concept of an orthogeodesic model - all we need is the 'structure theorem' for orthogeodesic exponential models (see Barndorff-Nielsen and Blæsild (1993) or Wiuf (1994a)).

If μ is a measure on a measure space $(\mathfrak{X}, \mathcal{A})$ and f is a measurable function from $\mathfrak{X}$ into another measure space $(\mathfrak{Y}, \mathcal{B})$, we let $f\mu$ denote the image measure of f under μ, i.e. $f\mu(B) = \mu(f^{-1}(B))$, $B \in \mathcal{B}$.

Let $\mathcal{M} = (\mathfrak{X}, \mathcal{P}, \mathcal{A})$ be a statistical model. Assume that G is a group acting on $\mathfrak{X}$ and moreover assume that $G\mathcal{P} = \{\, gP \,|\, P \in P,\, g \in G\} \subseteq \mathcal{P}$. Hence an action of G on $\mathcal{P}$ is induced, given by $g : \mathcal{P} \to \mathcal{P}$ and $P \mapsto gP$. If $\mathcal{P}$ is parameterized by $\omega \in \Omega$, i.e. $\mathcal{P} = \{\, \mathcal{P}_\omega \,|\, \omega \in \Omega \,\}$, this action can of course be expressed as a function of ω, with $g\omega$ given by

$$P_{g\omega} = gP_\omega. \tag{1}$$

Following Barndorff-Nielsen et al. (1989) we define:

Definition 2.1: $\mathcal{M}$ with the above mentioned properties is called a composite transformation model. If G acts transitively on $\mathcal{P}$, then $\mathcal{M}$ is called a transformation model. □

We let $\mathcal{M} = (\mathfrak{X}, \mathcal{P}, \mathcal{A}, \mathcal{G})$ denote a composite transformation model.

Let $\mathcal{P}$ be parameterized by $\omega \in \Omega$. In the following we will assume that $\mathcal{P}$ is of constant orbit type, i.e. that there exists a set of orbit representatives $\Psi \subset \Omega$ and a subgroup K of G, such that the isotropic group of ψ is K for all $\psi \in \Psi$, i.e. $G_\psi = \{g \in G \,|\, g\psi = \psi\} = K$. Let G/K denote the set of left cosets $gK = \{gk \,|\, k \in K\}$.

Lemma 2.1: ω and (ψ, gK) is in one-to-one correspondence.

Proof: If $\omega \in \Omega$ there exists a unique ψ, such that $\omega \in G\psi = \{g\psi \,|\, g \in G\}$. If there exists gK and $\tilde{g}K$, such that $\omega = gK\psi = \tilde{g}K\psi$ it follows that $g^{-1}\tilde{g} \in K$. Hence $gK = gg^{-1}\tilde{g}K = \tilde{g}K$ and ω determines a unique ψ and gK. Oppositely ψ and gK of course determine a unique ω. □

The last assumptions to be made concern the geometric structure of $\mathcal{P}$: assume that $\mathcal{P}$ is a differentiable product manifold $X \times \Psi$ of dimension d covered by a single chart $\Omega \subseteq R^d$, such that X (of dimension d_χ) is diffeomorphic to G/K and Ψ (of dimension d_ψ) is diffeomorphic to the set of orbit representatives.

Let the measures in $\mathcal{P}$ be dominated by a measure μ on $(\mathfrak{X}, \mathcal{A})$ with the μ densities being strictly positive. Since $P_{g\omega}$ has density w.r.t. μ and $g\mu$, and the densities are strictly positive, then the measures μ and $g\mu$ are equivalent. We then have from (1) that

$$\frac{dP_{g\omega}}{d\mu}(x) = \frac{dgP_\omega}{dg\mu}(x)\frac{dg\mu}{d\mu}(x) = \frac{dP_\omega}{d\mu}(g^{-1}x)\frac{dg\mu}{d\mu}(x). \tag{2}$$

Assume now that $\mathcal{M}$ is an exponential model of dimension d with open canonical parameter space, i.e. that the densities have the following form

$$\frac{dP_\theta}{d\mu}(x) = \exp\{\theta^\rho t_\rho(x) - \kappa(\theta) - \varphi(x)\} \tag{3}$$

for θ in an open subset Θ of R^d and $x \in \mathfrak{X}$. Here we have adopted the Einstein summation convention with ρ, σ etc. denoting indices of θ, τ and t. Moreover assume that $\mathcal{M}$ is orthogeodesic relative to the parameterization $\omega = (\chi, \psi)$ in the following sense (see Barndorff-Nielsen and Blæsild (1993) or Wiuf (1994a)):

There exists scalars $\alpha(\psi)$ and $\gamma(\chi)$, vectors $B_\rho(\chi)$ and $D^\rho(\chi)$, matrices $A^i_\rho(\chi)$ and $C^\rho_i(\chi)$, such that the following conditions are satisfied:

(a) $\theta^\rho(\chi,\psi) = \psi^i C^\rho_i(\chi) + D^\rho(\chi)$
(b) $\tau_\rho(\chi,\psi) = \alpha_{/j}(\psi) A^j_\rho(\chi) + B_\rho(\chi)$
(c) $\kappa(\chi,\psi) = \alpha(\psi) + \gamma(\chi) + \psi^i C^\rho_i(\chi) B_\rho(\chi)$
(d) $A^j_\rho(\chi) C^\rho_i(\chi) = \delta^j_i$
(e) $A^j_\rho(\chi) C^\rho_{i/a}(\chi) = 0$
(f) $A^j_\rho(\chi) D^\rho_{/a}(\chi) = 0$
(g) $B_{\rho/a}(\chi) C^\rho_i(\chi) = 0$
(h) $\gamma_{/a}(\chi) = B_\rho(\chi) D^\rho_{/a}(\chi)$.

Indices i, j etc. denote generic components of ψ and indices a, b denote generic components of χ. Moreover $/\rho$ respectively $/a$ and $/j$ denote differentiation w.r.t. θ^ρ respectively χ^a and ψ^j, e.g. $\alpha_{/j}(\psi) = \partial\alpha/\partial\psi^j(\psi)$. The vectors $B_\rho(\chi)$ and $D^\rho(\chi)$, matrices $A^i_\rho(\chi)$ and $C^\rho_i(\chi)$ and scalars $\alpha(\psi)$ and $\gamma(\chi)$ are called an orthogeodesic representation of $\mathcal{M}$.

In Barndorff-Nielsen and Blæsild (1993) one moreover requires variation independence of χ and ψ, but this is automatically fulfilled for χ and ψ in the group set-up mentioned above. According to the structure theorem for exponential (composite) transformation models (see e.g. Barndorff-Nielsen et al. (1989) or Wiuf (1994b)) the matrices are uniquely determined.

The class of models fulfilling the assumptions (a)-(h) is quite large, e.g. the family of von Mises-Fisher distributions, the family of hyperboloid distributions and the family of normal distributions are all examples of this kind. In Wiuf (1994b) some further 2-dimensional examples are given.

3: The Conditional Distribution of $\hat{\chi}$ given $\hat{\psi}$ and the Distribution of $\hat{\psi}$.

In this section we will let g denote both the action of G on the sample space $\mathfrak{X}$ and the action of G on the parameter space of $\mathcal{M}$ as well as the induced action of G on X, i.e. $g\chi = \tilde{\chi}$ if $g(\chi, \psi) = (\tilde{\chi}, \psi)$.

We need a result concerning the maximum likelihood estimate of ω, which is not bounded to OECT-Models. The proposition can be found more or less in existing textbooks, e.g. Barndorff-Nielsen (1988).

Proposition 3.1: Assume that $\mathcal{M}$ is a composite transformation model (not necessarily exponential) parameterized by $\omega \in \Omega$, and assume moreover that the maximum likelihood estimate (m.l.e.) $\hat{\omega}$ of ω exists and is unique with probability 1, then

(a) $\hat{\omega}$ is G-equivariant, i.e. $\hat{\omega}(x) = \hat{\omega}(y) \Rightarrow \hat{\omega}(gx) = \hat{\omega}(gy)$.
(b) The action of G on Ω induced by $\hat{\omega}$, i.e. $g(\hat{\omega}) = \hat{\omega}(gx)$ if $\hat{\omega} = \hat{\omega}(x)$ and $g \in G$, equals the natural action of G on ω defined by (1), i.e. $g(\hat{\omega}) = g\hat{\omega}$.
(c) In the parameterization $\omega = (\chi, \psi)$ (if it exists) $\hat{\psi}$ is G-invariant, i.e. the distribution of $\hat{\psi}$ depends on ψ only.

Proof: (a) Put $p(x; \omega) = \frac{dP_\omega}{d\mu}(x)$ and $r(x; g) = \frac{dg\mu}{d\mu}(x)$, then (2) can be rewritten as

$$p(gx; g\omega) = p(x; \omega) r(gx; g). \tag{4}$$

Furthermore, letting $\hat{\omega} = \hat{\omega}(x)$ we have that

$$p(x; \hat{\omega}) = p(gx; g\hat{\omega}) \frac{1}{r(gx; g)} \leq p(gx; \hat{\omega}(gx)) \frac{1}{r(gx; g)} = p(x; g^{-1}\hat{\omega}(gx)) \leq p(x; \hat{\omega}),$$

where the two inequality signs are a consequence of the definition of the m.l.e. and the two equalities follow from (4). Since we assumed that m.l.e. $\hat{\omega}$ is unique, we have from the above that $\hat{\omega} = g^{-1}\hat{\omega}(gx)$ or $g\hat{\omega} = \hat{\omega}(gx)$. Hence if $\hat{\omega}(x) = \hat{\omega}(y)$ then $\hat{\omega}(gx) = g\hat{\omega}(x) = g\hat{\omega}(y) = \hat{\omega}(gy)$, which proves (a).
(b) Shown in the proof of (a).
(c) Since ψ is invariant under the natural action of G on Ω, we have from (b) that

$$\hat{\psi}(x) = \hat{\psi} = g\hat{\psi} = \hat{\psi}(gx),$$

i.e. $\hat{\psi}$ is G-invariant, and hence the distribution of $\hat{\psi}$ depends on ψ only. $\square$

Assume the model $\mathcal{M}$ fulfills the assumptions (a)-(h) made in section 2, i.e. that $\mathcal{M}$ is an OECT-Model relative to the parameterization $\omega = (\chi, \psi)$.

Besides the above proposition we need a lemma concerning the relation between the matrix D and the function φ in the exponential representation (3) of $\mathcal{P}$. Put $\frac{dg\mu}{d\mu}(x) = \exp\{-\eta(x; g)\}$. Then we have

Lemma 3.1: Assume that $\mathcal{M}$ is an OECT-Model relative to the parameterization $\omega = (\chi, \psi)$ as defined above. Then the following statements are equivalent:

(a) $\varphi(x) = \varphi(g^{-1}x) + \eta(x; g) \quad \forall x \in \mathfrak{X}, \; g \in G$

(b) $D(\chi) = 0 \quad \forall \chi \in X$

In particular if μ is G-invariant then condition (a) means that φ depends on the G-orbits of $\mathfrak{X}$ only, and if G acts transitively on $\mathfrak{X}$, then (a) states that φ is constant.

Proof: See Wiuf (1994b). $\square$

The assumption $D = 0$ is fulfilled quite often, e.g. all three examples mentioned in the previous section fulfill this assumption. If $D = 0$ and $d_\psi = 1$ then $\mathcal{M}$ is a proper exponential dispersion model as defined in Barndorff-Nielsen and Jørgensen (1991).

We are now ready to prove a conditional distribution result under repeated sampling that is a unification of two well known results given earlier in the literature: the facts that if x follows a von Mises-Fisher distribution then also the conditional distribution $\hat{\chi}|\hat{\psi}$ of $\hat{\chi}$ given $\hat{\psi}$ follows a von Mises-Fisher distribution with the same location parameter, and if x follows a hyperboloid distribution then $\hat{\chi}|\hat{\psi}$ too follows a hyperboloid distribution with the same location parameter (Jensen (1981)). Moreover we derive an other but similar result for OECT-Models fulfilling slightly different assumptions - a result that contains a well known distributional result for the normal family.

This second result is a consequence of condition (a) in the below theorem 3.1. This condition and the consequence of it resemble a result given in Barndorff-Nielsen and Blæsild (1983) on τ-parallel models, which include the normal family as a special case.

Let $\mathcal{M}^n = (\mathfrak{X}^n, \mathcal{P}^n, \mathcal{A}^\backslash)$ be the model consisting of n-fold product measures of identical measures from $\mathcal{P}$, i.e the measure P_θ^n in $\mathcal{P}^\setminus$ has density w.r.t. μ^n given by

$$\frac{dP_\theta^n}{d\mu^n}(\bar{x}) = \exp\{n\theta^\rho \bar{t}_\rho(\bar{x}) - n\kappa(\theta) - \bar{\varphi}(\bar{x})\}, \tag{5}$$

where $\bar{x} = (x_1, ...x_n)$, $\bar{t}(\bar{x}) = \frac{1}{n}\sum_j t(x_j)$ and $\bar{\varphi}(\bar{x}) = \frac{1}{n}\sum_j \varphi(x_j)$.

If a statistic z follows a distribution in the model $\mathcal{M}$ we let $F(\chi, \psi)$ denote the distribution of z.

Let $(*)$ be the statement:

$(*)$ If Q_1 and Q_2 are two G-invariant measures on $(\mathfrak{X}, \mathcal{A})$, then there exists a constant $c > 0$, such that $Q_1 = cQ_2$.

Finally let $(\dagger)$ be the statement:

$(\dagger)$ μ is G-invariant, and the marginal measure $(\hat{\chi}, \hat{\psi})\mu^n$ of $(\hat{\chi}, \hat{\psi})$ under μ^n is a product measure $\nu^n = \nu_X^n \otimes \nu_\Psi^n$ of a measure ν_X^n on $(X, \mathcal{B}(\mathcal{X}))$ and a measure ν_Ψ^n on $(\Psi, \mathcal{B}(\Psi))$, which both are sum-finite.

Both statements are under weak regularity conditions fulfilled (see e.g. Barndorff-Nielsen et al. (1989)), and it causes no restrictions in practice. The regularity conditions to ensure $(*)$ to hold concern the action of G on $\mathfrak{X}$ and are of topological nature. The condition $(\dagger)$ is similarly fulfilled if the action induced on $X \times \Psi$ by $(\hat{\chi}, \hat{\psi})$ satisfy some topological conditions and if K (the isotropic group of ψ, $\psi \in \Psi$) is a regular subgroup of G. If $\mathcal{M}$ is a standard transformation model (see Barndorff-Nielsen et al. (1989)) then $(\dagger)$ is satisfied per assumption.

Theorem 3.1: Assume that $\mathcal{M}$ is an OECT-Model relative to the parameterization $\omega = (\chi, \psi)$ as defined above, $(\mathfrak{X}, \mathcal{A}) = (\mathcal{X}, \mathcal{B}(\mathcal{X}))$, $D = 0$ and that $\bar{t}(\mathfrak{X}^n) = \tau(\Theta)$. Moreover assume that either '$0 \in \Theta$ and $(*)$' is fulfilled (remember '$0 \in \Theta$' can always be fulfilled, but not necessarily with $D = 0$), or that '$(*)$ and $(\dagger)$' is fulfilled.

If

(a) (1) $t_\rho(x) = B_\rho(x)$
(2) $A_\rho^j(x) = A_\rho^j$ constant,

then $\hat{\chi}$ and $\hat{\psi}$ are stochastically independent and $\hat{\chi} \sim F(\chi, n\psi)$.

And if

(b) (1) $t_\rho(x) = c_j A^j_\rho(x)$

(2) $$A(x) = \left\{ \begin{array}{cccc} A[1](x) & 0 & ... & 0 \\ 0 & A[2](x) & ... & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & ... & A[d_\psi](x) \end{array} \right\}$$

(3) $$C(x) = \left\{ \begin{array}{cccc} C[1](x) & 0 & ... & 0 \\ 0 & C[2](x) & ... & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & ... & C[d_\psi](x) \end{array} \right\}$$

(4) $B_\rho(x) = 0,$

where $c = (c_j)_j$ is a constant vector, $A[j](x)$ is a column vector and $C[i](x)$ is a row vector, such that $A[j](x)^T,\ C[j](x) \in R^{d_j}$ for $j = 1, ..., d_\psi$ and $\sum_1^{d_\psi} d_j = d$, then $\hat{\chi}|\hat{\psi} \sim F(\chi, n\phi)$ with $\phi^i = \psi^i \frac{\alpha_{/i}(\hat{\psi})}{c_i}$.

Note 1: The condition $(\mathfrak{X}, \mathcal{A}) = (\mathcal{X}, \mathcal{B}(\mathcal{X}))$ is quite often fulfilled even though it involves both the sample space $\mathfrak{X}$ and a part of the parameter space.

Note 2: The assumption $D = 0$ is essential since if this is not fulfilled then the conditional distribution of $\hat{\chi}$ given $\hat{\psi}$ will not resemble the distribution of x (D will be replaced by nD).

Note 3: '$\bar{t}(\mathfrak{X}^\mathfrak{n}) = \tau(\Theta)$' implies that $\mathcal{M}^\setminus$ is steep (in the sense of Barndorff-Nielsen (1978)), and that the m.l.e. $(\hat{\chi}, \hat{\psi})$ of (χ, ψ) exists and is unique with probability 1 and is the solution to

$$\bar{t} = E_{\theta(\hat{\chi},\hat{\psi})} t = \alpha_{/j}(\hat{\psi}) A^j_\rho(\hat{\chi}) + B_\rho(\hat{\chi})$$

(according to the orthogeodesic condition (b)).

Note 4: Condition (a) implies that the model is τ-parallel (Barndorff-Nielsen and Blæsild (1983a)).

Note 5: The conditions (b)(2) and (b)(3) are trivially fulfilled if $d_\psi = 1$. The conditions are also satisfied if $\mathcal{M}$ is of the form $(\mathfrak{X}_1 \times \mathfrak{X}_2, \mathcal{P}_\infty \otimes \mathcal{P}_\in, \mathcal{A}_\infty \otimes \mathcal{A}_\in, \mathcal{G}_\infty \times \mathcal{G}_\in)$ with $\mathcal{P}_\infty \otimes \mathcal{P}_\in = \{\mathcal{P}_\infty \otimes \mathcal{P}_\in | \mathcal{P}_\rangle \in \mathcal{P}_\rangle, \rangle = \infty, \in\}$ and if $\mathcal{M}_\rangle = (\mathfrak{X}_\mathfrak{i}, \mathcal{P}_\rangle, \mathcal{A}_\rangle, \mathcal{G}_\rangle)$, $i = 1, 2$ fulfill the conditions.

Note 6: The densities have the form (3), and if we put $d\tilde{\mu} = \exp\{-\varphi\}d\mu$ then

$$\frac{dP_\theta}{d\tilde{\mu}}(x) = \exp\{\theta^\rho t_\rho(x) - \kappa(\theta)\}.$$

Hence using lemma 3.1 we see that $D = 0$ implies that $\eta = 0$ and $\tilde{\mu}$ is G-invariant.

Proof: The proofs under the two different assumptions '$0 \in \Theta$ and $(*)$' and '$(*)$ and $(\dagger)$' are almost identical.

Assume (a) and '$0 \in \Theta$ ($D = 0$) and $(*)$': Let $Q^n_{\chi,\psi}$ denote the marginal measures of $(\hat{\chi}, \hat{\psi})$ corresponding to the parameter value (χ, ψ). Assuming that $\kappa(0) = 0$ (which can be done without loss of generality) and using the orthogeodesic conditions (a), (c) and (d), and that $\bar{t} = \alpha_{/j}(\hat{\psi})A^j_\rho(\hat{\chi}) + B_\rho(\hat{\chi})$ (note 3) we see by inserting (a) in (5) that $Q^n_{\chi,\psi}$ has density w.r.t. Q^n_0 given by

$$\frac{dQ^n_{\chi,\psi}}{dQ^n_0}(\hat{\chi}, \hat{\psi}) = \exp\{n\psi^i C^\rho_i(\chi)t_\rho(\hat{\chi}) - n\alpha(\psi) + n\psi^i\alpha_{/i}(\hat{\psi}) - n\psi^i C^\rho_i(\chi)B_\rho(\chi)\}.$$

From proposition 3.1 we have that the marginal distribution of $\hat{\psi}$ depends on ψ only, and hence we see that the conditional distribution $Q^n_{\chi,\psi}(\cdot|\hat{\psi})$ of $\hat{\chi}$ given $\hat{\psi}$ has density w.r.t. the conditional distribution $Q^n_0(\cdot|\hat{\psi})$ of $\hat{\chi}$ given $\hat{\psi}$ under Q^n_0 given by

$$\frac{dQ^n_{\chi,\psi}(\cdot|\hat{\psi})}{dQ^n_0(\cdot|\hat{\psi})}(\hat{\chi}) = \frac{1}{q^n(\hat{\psi};\psi)}\exp\{n\psi^i C^\rho_i(\chi)t_\rho(\hat{\chi}) - n\alpha(\psi) + n\psi^i\alpha_{/i}(\hat{\psi}) - n\psi^i C^\rho_i(\chi)B_\rho(\chi)\}, \tag{6}$$

where we have used (d) $C^\rho_i(\chi)A^j_\rho = \delta^j_i$, and where $q^n(\hat{\psi};\psi)$ denotes the density of $\hat{\psi}$ w.r.t the marginal density of $\hat{\psi}$ under Q^n_0. Note that x has density w.r.t. P_0 given by

$$\frac{dP_{\chi,\psi}}{dP_0}(x) = \exp\{\psi C^\rho_i(\chi)t_\rho(x) - \alpha(\psi) - \psi C^\rho_i(\chi)B_\rho(\chi)\}. \tag{7}$$

Since the representation of x is minimal, it follows from lemma 3.1 ($D = 0$), (6) and (7) that P_0 and $Q^n_0(\cdot|\hat{\psi})$ are both G-invariant measures, since the φ function in (6) and (7) both are constant. Hence according to $(*)$ there exists a constant c dependent on $\hat{\psi}$ and n such that $Q^n_0(\cdot|\hat{\psi}) = c(\hat{\psi};n)P_0$. This constant is however 1, since both $Q^n_0(\cdot|\hat{\psi})$ and P_0 are probability measures. Inserting P_0 in (6) we obtain

$$\frac{dQ^n_{\chi,\psi}(\cdot|\hat{\psi})}{dP_0}(\hat{\chi}) = \frac{1}{q^n(\hat{\psi};\psi)}\exp\{n\psi^i C^\rho_i(\chi)t_\rho(\hat{\chi}) - n\alpha(\psi) + n\psi^i\alpha_{/i}(\hat{\psi}) - n\psi^i C^\rho_i(\chi)B_\rho(\chi)\}. \tag{8}$$

Integrating (8) over $\mathfrak{X}$ w.r.t. P_0 we get (using (7))

$$1 = \int_{\mathfrak{X}} \frac{dQ^n_{\chi,\psi}(\cdot|\hat{\psi})}{dP_0}(\hat{\chi})\, dP_0 =$$

$$\frac{1}{q^n(\hat{\psi};\psi)} \exp\{n\psi^i \alpha_{/i}(\hat{\psi})\} \int_{\mathfrak{X}} \exp\{n\psi^i C_i^\rho(\chi)[t_\rho(\hat{\chi}) - B_\rho(\chi)] - n\alpha(\psi)\} dP_0 =$$

$$\frac{1}{q^n(\hat{\psi};\psi)} \exp\{n\psi^i \alpha_{/i}(\hat{\psi}) - n\alpha(\psi) + \alpha(n\psi)\}. \tag{9}$$

Using (9) we can write (8) as

$$\frac{dQ_{\chi,\psi}(\cdot|\hat{\psi})}{dP_0}(\hat{\chi}) = \exp\{n\psi^i C_i^\rho(\chi) t_\rho(\hat{\chi}) - \alpha(n\psi) - n\psi^i C_i^\rho(\chi) B_\rho(\chi)\}.$$

The right side does not depend on $\hat{\psi}$ and hence $\hat{\chi}$ and $\hat{\psi}$ are stochastically independent, and comparing with (7) we see that $\hat{\chi} \sim F(\chi, n\psi)$. This proves the first part of the theorem under the condition '$0 \in \Theta$ $(D = 0)$ and $(*)$'.

Assume (b) and '$0 \in \Theta$ $(D = 0)$ and $(*)$': Start by noting that $c_j \neq 0$ for all j. Oppositely assume there exists j such that $c_j = 0$, say $j = 1$. Then (b)(1) and (b)(2) imply that $t(x) = (0, t_2(x))$ with 0 denoting a d_1-dimensional zero vector and $t_2(x)$ being of dimension $(d - d_1)$. Hence $t(\mathfrak{X}) \subseteq \mathfrak{R}^{\mathfrak{d}-\mathfrak{d}_1}$ in contradiction to the assumption that the representation of $\mathcal{M}$ is minimal. This means $c_j \neq 0$ for all j. Then note that

$$\psi^i C_i^\rho(\chi) A_\rho^j(x) \alpha_{/j}(\hat{\psi}) = \left(\psi^1 C[1](\chi), ..., \psi^{d_\psi} C[d_\psi](\chi)\right) \begin{pmatrix} \alpha_{/1}(\hat{\psi}) A[1](x) \\ \vdots \\ \alpha_{/d_\psi}(\hat{\psi}) A[d_\psi](x) \end{pmatrix} =$$

$$\left(\psi^1 C[1](\chi), ..., \psi^{d_\psi} C[d_\psi](\chi)\right) \begin{pmatrix} \frac{\alpha_{/1}(\hat{\psi})}{c_1} c_1 A[1](x) \\ \vdots \\ \frac{\alpha_{/d_\psi}(\hat{\psi})}{c_{d_\psi}} c_{d_\psi} A[d_\psi](x) \end{pmatrix} = \phi^i C_i^\rho(\chi) t_\rho(x),$$

where $\phi^i = \psi^i \frac{\alpha_{/i}(\hat{\psi})}{c_i}$ for $i = 1, ..., d_\psi$. Using this equality together with condition (b)(4) we find, by an argument as above, that

$$\frac{dQ^n_{\chi,\psi}(\cdot|\hat{\psi})}{dQ^n_0(\cdot|\hat{\psi})}(\hat{\chi}) = \frac{1}{q^n(\hat{\psi};\psi)} \exp\{n\phi^i C_i^\rho(\chi) t_\rho(\hat{\chi}) - n\alpha(\psi)\}. \tag{10}$$

Moreover analogous to (7) we have that x has density w.r.t. P_0 given by

$$\frac{dP_{\chi,\psi}}{dP_0}(x) = \exp\{\psi^i C_i^\rho(\chi) t_\rho(x) - \alpha(\psi)\},$$

and that P_0 and $Q_0^n(\cdot|\hat{\psi})$ are identical G-invariant measures on $\mathfrak{X}$ (according to $(*)$). Hence we can write

$$\frac{dQ^n_{\chi,\psi}(\cdot|\hat{\psi})}{dP_0}(\hat{\chi}) = \frac{1}{q^n(\hat{\psi};\psi)} \exp\{n\phi^i C_i^\rho(\chi) t_\rho(\hat{\chi}) - n\alpha(\psi)\}.$$

Applying the same integration argument as above we derive that

$$1 = \frac{1}{q^n(\hat{\psi};\psi)} \exp\{-n\alpha(\psi) + \alpha(n\phi)\}, \tag{11}$$

and

$$\frac{dQ^n_{\chi,\psi}(\cdot|\hat{\psi})}{dP_0}(\hat{\chi}) = \exp\{n\phi^i C_i^\rho(\chi) t_\rho(\hat{\chi}) - \alpha(n\phi)\}.$$

Hence $\hat{\chi}|\hat{\psi} \sim F(\chi, n\phi)$, and the proof of the second part of the theorem under the condition '$0 \in \Theta$ $(D = 0)$ and $(*)$' is completed.

Assume now (a) or (b) and '$(*)$ and $(\dagger)$': From proposition 3.1 we know that the m.l.e. $(\hat{\chi},\hat{\psi})$ is G-equivariant, and hence we conclude that the image measure $\nu^n = (\hat{\chi},\hat{\psi})\mu^n$ of $(\hat{\chi},\hat{\psi})$ under μ^n is G-invariant, since μ and hence too μ^n are G-invariant. Choose $B \in \mathcal{B}(\Psi)$ such that $0 < \nu^n_\Psi(B) < \infty$ (this can be done due to sum-finiteness of ν^n_Ψ). We then have

$$g\nu^n_X(A)\nu^n_\Psi(B) = \nu^n_X(g^{-1}A)\nu^n_\Psi(B) = \nu^n(g^{-1}A \times B) = \nu^n(A \times B) = \nu^n_X(A)\nu^n_\Psi(B),$$

since $g^{-1}(A \times B) = g^{-1}A \times B$ (proposition 3.1). Dividing by $\nu^n_\Psi(B)$ we see that

$$g\nu^n_X(A) = \nu^n_X(A),$$

i.e. ν^n is G-invariant on $(X, \mathcal{B}(\mathcal{X}))$ $(= (\mathfrak{X}, \mathcal{A}))$. But μ is too G-invariant on $(X, \mathcal{B}(\mathcal{X}))$, and hence from $(*)$ we conclude that $\nu^n_X = c^n\mu$ for some constant c^n. Again since μ is G-invariant and since $D = 0$ we have from lemma 3.1 that $\varphi = 0$, and hence that the marginal distribution $Q^n_{\chi,\psi}$ of $(\hat{\chi},\hat{\psi})$ under $P^n_{\chi,\psi}$ has density w.r.t. ν^n which takes the form

$$\frac{dQ^n_{\chi,\psi}}{d\nu^n}(\hat{\chi},\hat{\psi}) = \exp\{n\psi^i C_i^\rho(\chi)\bar{t}_\rho(\hat{\chi},\hat{\psi}) - n\alpha(\psi) - n\psi^i C_i^\rho(\chi) B_\rho(\chi)\}.$$

Note that the marginal distribution $\hat{\psi}Q^n_{\chi,\psi}$ of $\hat{\psi}$ under $Q^n_{\chi,\psi}$ has density w.r.t. ν^n_Ψ since

$$\hat{\psi}Q^n_{\chi,\psi}(B) = \int_{X\times B} \frac{dQ^n_{\chi,\psi}}{d\nu^n} d\nu^n_X \otimes \nu^n_\Psi = \int_B q^n(\hat{\psi};\psi) d\nu^n_\Psi$$

for some function $q^n(\hat{\psi};\psi)$ dependent of $\hat{\psi}$ and ψ ($q^n(\hat{\psi};\psi)$ does not depend on χ since according to proposition 3.1 the distribution of $\hat{\psi}$ does only depend on ψ).

From Hoffmann-Jørgensen (1994) 6.11 we conclude that the conditional distribution $Q_{\chi,\psi}(\cdot|\hat{\psi})$ of $\hat{\chi}$ given $\hat{\psi}$ has density w.r.t. the G-invariant measure ν^n_X on $(X, \mathcal{B}(\mathcal{X}))$ (here we use sum-finiteness of both ν^n_X and ν^n_Ψ). But as noted earlier $\nu^n_X = c^n\mu$ for some constant c^n and hence the conditional density can be given w.r.t. μ. Moreover since $\varphi = 0$ then the distribution of x is given by

$$\frac{dP_{\chi,\psi}}{d\mu}(x) = \exp\{\psi^i C^\rho_i(\chi) t_\rho(x) - \alpha(\psi) - \psi^i C^\rho_i(\chi) B_\rho(\chi)\}.$$

Performing exactly the same calculations as in the two proofs above we conclude that under the assumption '($*$) and (†)' the two statements under condition (a) and (b) are valid as well. This completes the proof of theorem 3.1. □

Let us remark that the above arguments concerning G-invariant measures and conditional densities more or less might be found in e.g. Barndorff-Nielsen et al. (1989).

We moreover remark that from the proof of theorem 3.1 we have (see (9) and (11)) that

$$1 = \frac{c^n}{q^n(\hat{\psi};\psi)} \exp\{n\psi^i \alpha_{/i}(\hat{\psi}) - n\alpha(\psi) + \alpha(n\psi)\}, \tag{12}$$

and

$$1 = \frac{c^n}{q^n(\hat{\psi};\psi)} \exp\{-n\alpha(\psi) + \alpha(n\phi)\} \tag{13}$$

for some constant c^n, which under the assumption '$0 \in \Theta$ and ($*$)' is 1, and might be different from 1 under '($*$) and (†)', and where $q^n(\hat{\psi};\psi)$ is the density of $\hat{\psi}$ w.r.t $\hat{\psi}Q^n_0$ under '$D = 0$ and ($*$)' and w.r.t. ν^n_Ψ under '($*$) and (†)'.

Theorem 3.2: Assume that the conditions in theorem 3.1 are fulfilled. Let π be the measure $\hat{\psi}Q^n_0$ if '$D = 0$ and ($*$)' is satisfied and the measure ν^n_Ψ if '($*$) and (†)' is satisfied. Moreover put $\pi_\psi = \hat{\psi}Q^n_{\chi,\psi}$ (independent of χ according to proposition 3.1).

If (a) is satisfied then the family $\mathcal{M}_{\hat{\psi}} = (\Psi, \mathcal{P}_{\hat{\psi}}, \mathcal{B}(\Psi))$, $\mathcal{P}_{\hat{\psi}} = \{\pi_\psi | \psi \in \Psi\}$, is an exponential model with minimal representation given by

$$\frac{d\pi_\psi}{d\pi}(\hat{\psi}) = c^n \exp\{n\psi^i \alpha_{/i}(\hat{\psi}) - n\alpha(\psi) + \alpha(n\psi)\},$$

and moreover $n\Psi \subseteq \Psi$.

If (b) is satisfied then the family $\mathcal{M}_{\hat{\psi}} = (\Psi, \mathcal{P}_{\hat{\psi}}, \mathcal{B}(\Psi))$, $\mathcal{P}_{\hat{\psi}} = \{\pi_\psi | \psi \in \Psi\}$, has densities w.r.t. π given by

$$\frac{d\pi_\psi}{d\pi}(\hat{\psi}) = c^n \exp\{\alpha(n\phi) - n\alpha(\psi)\}$$

with $\phi^i = \psi^i \frac{\alpha_{/i}(\hat{\psi})}{c_i}$, and moreover $\Phi = \{\phi | \phi^i = \psi^i \frac{\alpha_{/i}(\hat{\psi})}{c_i},\ \psi, \hat{\psi} \in \Psi\} \subseteq \Psi$.

Proof: Follows directly from (12) and (13). □

The similarity in distribution between the variable x and the conditional variable $\hat{\chi}$ given $\hat{\psi}$ under repeated sampling is closely related to the concept (strong) reproductivity as defined and discussed in Barndorff-Nielsen and Blæsild (1983b) (see too Barndorff-Nielsen and Blæsild (1983a)). These models are all τ-parallel models, and as seen from Barndorff-Nielsen and Blæsild (1983b) the models fulfilling condition (a) in theorem 3.1 are all strongly reproductive. However the group theoretical set-up is not present in Barndorff-Nielsen and Blæsild (1983b), and they derive a result (as mentioned earlier) which seems to be more general than theorem 3.1 (a). It could be interesting to see if a similar (generalization) holds for theorem 3.1 (b) in a non-group theoretical set-up. As seen in the proof of theorem 3.1 uniqueness of a G-invariant measure plays an important role.

Example 3.1: Let us introduce a family of distributions, which includes the family $\mathcal{GIG}_c = \{\ GIG_c(\phi, \zeta) \mid \phi > 0,\ \zeta > 0\ \}$ of generalized inverse Gaussian distributions with index $c \in R$. We also call this class for the class of generalized inverse Gaussian distributions, but with index $c \in R$ and μ, $\nu \in R_+$, and denote the class by $\mathcal{GIG}_c^{\mu,\nu}$. The density of the distribution in $\mathcal{GIG}_c^{\mu,\nu}$ with parameters (ϕ, ζ) is given w.r.t. the Lebesgue measure λ on R_+ by

$$a_c^{\mu,\nu}(\phi,\zeta) x^{c-1} \exp\{-\frac{1}{2}(\phi x^{-\mu} + \zeta x^{\nu})\}, \qquad x \in R_+. \tag{14}$$

Hence $\mathcal{GIG}_c^{\mu,\nu}$ has two parameters ϕ and ζ, and the domain D_c of variation of (ϕ, ζ) is varying with c, but not with μ and ν. However we will here take the domain to be the largest set contained in all D_c, $c \in R$. This means

$$(\phi, \zeta) \in R_+ \times R_+. \tag{15}$$

$\mathcal{GIG}_c^{\mu,\nu}$ is an open exponential family of order two with the parameter space given by (15), and (14) is seen to be a minimal representation of $\mathcal{GIG}_c^{\mu,\nu}$. As shown in an appendix the norming constant $a_c^{\mu,\nu}(\phi, \zeta)$ takes the form

$$a_c^{\mu,\nu}(\phi,\zeta)^{-1} = \frac{2}{\nu}\sum_{n=0}^{\infty}(\frac{\phi}{2})^n \frac{1}{n!}\sum_{k=0}^{n}(-1)^k \binom{n}{k} K_{\varepsilon_{kn}}(\sqrt{\phi\zeta})\left(\frac{\phi}{\zeta}\right)^{\frac{1}{2}\varepsilon_{kn}}, \tag{16}$$

with $\varepsilon_{kn} = \frac{c}{\nu} - n + k(1 - \frac{\mu}{\nu})$. Here $K_\varepsilon(\cdot)$ is the modified Bessel-function of the third kind with index ε. In the case $\mu = \nu$ (16) reduces to

$$a_c^{\mu,\mu}(\phi,\zeta) = \frac{\mu(\zeta/\phi)^{c/(2\mu)}}{2K_{c/\mu}(\sqrt{\phi\zeta})}.$$

In particular, if $\mu = \nu = 1$ (which corresponds to the family of generalized inverse Gaussian distributions with index $c \in R$) then

$$a_c^{1,1}(\phi,\zeta) = \frac{(\zeta/\phi)^{c/2}}{2K_c(\sqrt{\phi\zeta})}.$$

Moreover we have that $a_c^{\mu,\nu}(\phi, \zeta)$ fulfills the following relation (see appendix)

$$\nu a_{c/\nu}^{\mu/\nu,1}(\phi,\zeta) = a_c^{\mu,\nu}(\phi,\zeta) = \mu a_{-c/\mu}^{\nu/\mu,1}(\zeta,\phi),$$

and hence we are able to calculate the norming constant for all values of $(\phi, \zeta, c, \mu, \nu)$, if we know the value of the constant for all values of (ϕ, ζ, c) and $\mu \leq 1$ and $\nu = 1$. With

$$\theta(\chi,\psi) = (\frac{1}{2\nu}\phi, \frac{1}{2\mu}\zeta) = (\psi\chi^\mu, \psi\chi^{-\nu}), \tag{17}$$

and

$$t^T(x) = (-\nu x^{-\mu}, -\mu x^\nu). \tag{18}$$

the density in (14) takes the following form

$$a_c^{\mu,\nu}(\chi,\psi)x^{c-1}\exp\{-\psi(\chi^\mu \nu x^{-\mu} + \chi^{-\nu}\mu x^\nu)\}. \tag{19}$$

We will in the sequel write $P^c_{\chi,\psi}$ to denote the density in (19) and $a_c(\chi,\psi)$ for short instead of $a^{\mu,\nu}_c(\psi,\chi)$. The reason for using a notation involving c only, and not μ and ν, is that the value of c turns out to be important geometrically, in contrast to μ and ν, and hence it ought not be suppressed in the notation.

One easily sees that $\mathcal{GIG}^{\mu,\nu}_c$ is a composite transformation model for all values of c, μ and ν with ψ denoting the index parameter and χ the group parameter. The group G is given by $G = R_+$ and the actions by

$$g(\chi,\psi) = (g\chi,\psi) \qquad \text{and} \qquad gx = gx.$$

The norming constant is seen to fulfill (due to the transformation model property)

$$a_c(\chi,\psi) = \chi^{-c} a_c(1,\psi). \tag{20}$$

Hence if $c = 0$ then the norming constant is a function of ψ (and μ, ν) only. Formula (20) implies that the cumulant transform of the canonical observator t (see (18)) is

$$\kappa(\chi,\psi) = c\log\chi - \log a_c(1,\psi) = c\log\chi + \alpha_c(\psi), \tag{21}$$

where we have put $\alpha_c(\psi) = -\log a_c(1,\psi)$. Using (17) and (21) we obtain the following expression of the mean value $\tau(\chi,\psi)$ of the canonical parameter $t^T(x) = (-\nu x^{-\mu}, -\mu x^{\nu})$ (see (18))

$$\begin{aligned}
\tau^T(\chi,\psi) &= \frac{d}{d\theta}\kappa(\chi(\theta),\psi(\theta)) = \left[c\frac{d}{d\chi}\log\chi, \frac{d}{d\psi}\alpha_c(\psi)\right]\frac{d(\chi,\psi)^T}{d\theta} \\
&= \frac{1}{\mu+\nu}\left[\left(\frac{c}{\psi} + \nu\alpha_{c/\psi}(\psi)\right)\chi^{-\mu}, \left(-\frac{c}{\psi} + \mu\alpha_{c/\psi}(\psi)\right)\chi^{\nu}\right].
\end{aligned} \tag{22}$$

From the orthogeodesic conditions (a)-(c) we conclude that $\mathcal{GIG}^{\mu,\nu}_c$ can not be orthogeodesic relative to the parameterization (χ,ψ) unless $c = 0$. In Wiuf (1994b) it is shown that there do not exist any orthogeodesic parameterization (ξ,φ) of $\mathcal{GIG}^{1,1}_c = \mathcal{GIG}_c$ (i.e. with $\mu = \nu = 1$), such that φ is an index parameter and ξ the group parameter. The author conjecture that this also is the case for arbitrary choice of μ and ν.

In order to apply the theory developed in the present paper we therefore restrict our considerations to the case $c = 0$. Using (20) and putting $P_{\chi,\psi} = P^0_{\chi,\psi}$ and $a(\psi) = a_0(1,\psi)$ the densities in $\mathcal{GIG}^{\mu,\nu}_0$ take the form

$$\frac{dP_{\chi,\psi}}{d\lambda}(x) = a(\psi)\frac{1}{x}\exp\{-\psi(\chi^{\mu}\nu x^{-\mu} + \chi^{-\nu}\mu x^{\nu})\}.$$

Moreover we see that $\mathcal{GIG}_0^{\mu,\nu}$ is orthogeodesic relative to the parameterization (χ, ψ) and that an orthogeodesic representation A, B, C, D, $\alpha(\cdot)$ and $\gamma(\cdot)$ can be chosen to be

$$A^T(\chi) = (\tfrac{\nu}{\mu+\nu}\chi^{-\mu}, \tfrac{\mu}{\mu+\nu}\chi^{\nu}) \quad B^T(\chi) = (0,0) \quad C(\chi) = (\chi^{\mu}, \chi^{-\nu}) \quad D(\chi) = (0,0)$$

$$\alpha(\psi) = -\log a(\psi) \quad \gamma(\chi) = 0.$$

Let us note that the model $(\mathcal{GIG}_0^{\mu,\nu})^n$ consisting of n-fold product measures of identical measures from $\mathcal{GIG}_0^{\mu,\nu}$ is a regular exponential family (since $\mathcal{GIG}_0^{\mu,\nu}$ is regular), and hence steep (in the sense of Barndorff-Nielsen (1978)), i.e. $\tau(\Theta) = \operatorname{int} C$, with C denoting the closed convex support of $\bar{t}(\bar{x}) = \frac{1}{n}\sum_1^n t(x_i)$. Moreover note that $\bar{t}(\mathfrak{X}^\mathfrak{n}) = \bar{\mathfrak{t}}(\mathfrak{R}_+^\mathfrak{n}) = \mathfrak{R}_- \times \mathfrak{R}_-$, and hence we see that

$$\tau(\Theta) = \operatorname{int} C = R_- \times R_- = \bar{t}(\mathfrak{X}^\mathfrak{n}).$$

This means that the conditions in theorem 3.1: $(\mathfrak{X}, \mathcal{A}) = (\mathcal{X}, \mathcal{B}(\mathcal{X})) = (\mathcal{R}_+, \mathcal{B}(\mathcal{R}_+))$, $D = 0$ and $\tau(\Theta) = \bar{t}(\mathfrak{X}^\mathfrak{n})$ all are fulfilled.

Furthermore we have that R_+ acts transitively on $\mathfrak{X} = \mathfrak{R}_+$ and the map $\Gamma : R_+ \times R_+ \to R_+ \times R_+$, $\Gamma(g, x) = (gx, x)$ is proper, i.e. Γ is continuous and the inverse image $\Gamma^{-1}(C)$ of a compact set $C \subset R_+ \times R_+$ is compact. From Barndorff-Nielsen et al. (1989) we then conclude that any two R_+-invariant measures on $(R_+, \mathcal{B}(\mathcal{R}_+))$ are identical modulo a constant factor, i.e. $(*)$ is fulfilled.

Moreover $d\mu(x) = \frac{1}{x}d\lambda(x)$ is R_+-invariant and a few considerations show that the conditions in theorem 5.4 in Barndorff-Nielsen et al. (1989) are satisfied as well. We conclude them from theorem 5.4 in Barndorff-Nielsen et al. (1989) that the marginal measure of the m.l.e. $(\hat{\chi}, \hat{\psi})$ of (χ, ψ) is a product measure of the form in $(\dagger)$, i.e. $(\dagger)$ is fulfilled. This means that the basic assumptions in theorem 3.1 are satisfied.

Moreover we have that

$$t^T(x) = (-\nu x^{-\mu}, -\mu x^{\nu}) = -(\mu + \nu)A^T(x),$$

and the model fulfills condition (b) in theorem 3.1.

We find the following expression of the m.l.e. $\hat{\chi}$ of χ:

$$\hat{\chi} = \left(\frac{\sum_1^n x_i^{\nu}}{\sum_1^n x_i^{-\mu}}\right)^{1/(\mu+\nu)},$$

and the m.l.e. $\hat{\psi}$ of ψ fulfills

$$n\frac{d}{d\psi}\log K_0(\hat{\psi}) = n\frac{d}{d\psi}\alpha(\hat{\psi}) = -(\mu + \nu)\left(\sum_1^n x_i^{\nu}\right)^{\frac{\mu}{\mu+\nu}}\left(\sum_1^n x_i^{-\mu}\right)^{\frac{\nu}{\mu+\nu}}.$$

Since $\hat{\psi}$ is in unique correspondence with $\hat{\pi} = \alpha_{/\psi}(\hat{\psi}) = -\frac{\mu+\nu}{n}\left(\sum_1^n x_i^{\nu}\right)^{\frac{\mu}{\mu+\nu}}\left(\sum_1^n x_i^{-\mu}\right)^{\frac{\nu}{\mu+\nu}}$ we might conditioning on $\hat{\pi}$ instead of $\hat{\psi}$ and we conclude from theorem 3.1 that the conditional distribution of $\hat{\chi} = \left(\frac{\sum_1^n x_i^{\nu}}{\sum_1^n x_i^{-\mu}}\right)^{1/(\mu+\nu)}$ given $\hat{\pi}$ has density w.r.t. Lebesgue measure λ on R_+ given by

$$\frac{dP_{\chi,\psi}}{d\lambda}(x) = a(\psi)\frac{1}{x}\exp\{-n\phi(\chi^{\mu}\nu x^{-\mu} + \chi^{-\nu}\mu x^{\nu})\} \tag{23}$$

with $\phi = -\frac{1}{\mu+\nu}\psi\hat{\pi} = \frac{\psi}{n}\left(\sum_1^n x_i^{\nu}\right)^{\frac{\mu}{\mu+\nu}}\left(\sum_1^n x_i^{-\mu}\right)^{\frac{\nu}{\mu+\nu}}$. If the function $\alpha(\cdot)$ was known an expression for the distribution of $\hat{\pi}$ could be derived from theorem 3.2.

Finally let us comment on the case $c \neq 0$. As noted before $\mathcal{GIG}_c^{\mu,\nu}$ is not orthogeodesic relative to the parameterization (χ, ψ), and hence the theory developed here do not apply. However a simularity in the distribution of $\hat{\chi}|\hat{\psi}$ under reapeted sampling will occur as pointed out by Barndorff-Nielsen (1989) p.102, but the index parameter c will vary with n (see e.g. Jørgensen (1980) in the case $\mu = \nu = 1$), and hence the simularity is not of the same strict type as discussed here. □

4: Appendix.

In this appendix we will derive an expression for the norming constant $a_c^{\mu,\nu}(\phi,\zeta)$ appearing in the densities of the distributions in $\mathcal{GIG}_c^{\mu,\nu}$ (see (14)). These densities are given w.r.t. Lebesgue measure λ on R_+ by

$$a_c^{\mu,\nu}(\phi,\zeta)x^{c-1}\exp\{-\frac{1}{2}(\phi x^{-\mu} + \zeta x^{\nu})\}, \qquad x \in R_+,$$

where

$$(\phi, \zeta, c, \mu, \nu) \in R_+ \times R_+ \times R \times R_+ \times R_+.$$

In the following $K_{\varepsilon}(\cdot)$ will denote the modified Besssel function of the third kind with index ε.

We will show that the norming constant $a_c^{\mu,\nu}(\phi,\zeta)$ is given by

$$a_c^{\mu,\nu}(\phi,\zeta)^{-1} = \frac{2}{\nu}\sum_{n=0}^{\infty}(\frac{\phi}{2})^n\frac{1}{n!}\sum_{k=0}^{n}(-1)^k\binom{n}{k}K_{\varepsilon_{kn}}(\sqrt{\phi\zeta})\left(\frac{\phi}{\zeta}\right)^{\frac{1}{2}\varepsilon_{kn}}, \tag{24}$$

with $\varepsilon_{kn} = \frac{c}{\nu} - n + k(1 - \frac{\mu}{\nu})$. Moreover we will note that, if $\mu = \nu$ then (24) reduces to

$$a_c^{\mu,\mu}(\phi, \zeta) = \frac{\mu(\zeta/\phi)^{c/(2\mu)}}{2K_{c/\mu}(\sqrt{\phi\zeta})}. \tag{25}$$

If we substitute y for x^ν we obtain the following expression for $a_c^{\mu,\nu}(\phi, \zeta)$:

$$\begin{aligned} a_c^{\mu,\nu}(\phi, \zeta)^{-1} &= \int_{R_+} x^{c-1} \exp\{-\frac{1}{2}(\phi x^{-\mu} + \zeta x^\nu)\} dx \\ &= \int_{R_+} \frac{1}{\nu} y^{\frac{c}{\nu}-1} \exp\{-\frac{1}{2}\phi y^{-\frac{\mu}{\nu}} - \frac{1}{2}\zeta y\} dy. \end{aligned}$$

Subtracting and adding the term $\frac{1}{2}\phi y^{-1}$ to the exponent in the exponential part yields

$$\begin{aligned} a_c^{\mu,\nu}(\phi, \zeta)^{-1} &= \int_{R_+} \frac{1}{\nu} y^{\frac{c}{\nu}-1} \exp\{-\frac{1}{2}(\phi y^{-1} + \zeta y)\} \exp\{\frac{1}{2}\phi(y^{-1} - y^{-\frac{\mu}{\nu}})\} dy \\ &= \frac{2K_{c/\nu}(\sqrt{\phi\zeta})}{\nu} (\frac{\phi}{\zeta})^{\frac{c}{2\nu}} \int_{R_+} f(y; \phi, \zeta, \frac{c}{\nu}) \exp\{\frac{1}{2}\phi(y^{-1} - y^{-\frac{\mu}{\nu}})\} dy \\ &= \frac{2K_{c/\nu}(\sqrt{\phi\zeta})}{\nu} (\frac{\phi}{\zeta})^{\frac{c}{2\nu}} E_{\phi,\zeta,\frac{c}{\nu}} \exp\{\frac{1}{2}\phi(Y^{-1} - Y^{-\frac{\mu}{\nu}})\}, \end{aligned} \tag{26}$$

where

$$f(y; \phi, \zeta, \frac{c}{\nu}) = \frac{1}{2K_{c/\nu}(\sqrt{\phi\zeta})} (\frac{\zeta}{\phi})^{\frac{c}{2\nu}} y^{\frac{c}{\nu}-1} \exp\{-\frac{1}{2}(\phi y^{-1} + \zeta y)\}$$

denotes the density function of a generalized inverse Gaussian distributed variable Y with parameters (ϕ, ζ) and index $\frac{c}{\nu}$. If we expand $\exp\{\frac{1}{2}\phi(y^{-1} - y^{-\frac{\mu}{\nu}})\}$ into powers of y we obtain

$$\begin{aligned} \exp\{\frac{1}{2}\phi(y^{-1} - y^{-\frac{\mu}{\nu}})\} &= \sum_{n=0}^{\infty} (\frac{\phi}{2})^n \frac{1}{n!} (y^{-1} - y^{-\frac{\mu}{\nu}})^n \\ &= \sum_{n=0}^{\infty} (\frac{\phi}{2})^n \frac{1}{n!} \sum_{k=0}^{n} (-1)^k \binom{n}{k} y^{k(1-\frac{\mu}{\nu})-n}. \end{aligned} \tag{27}$$

All moments of a generalized inverse Gaussian distributed variable Y with parameters (ϕ, ζ) and index d are known, and we have the following expression for the moments (see e.g. Jørgensen (1980))

$$E_{\phi,\zeta,d} Y^\varepsilon = \frac{K_{d+\varepsilon}(\sqrt{\phi\zeta})}{K_d(\sqrt{\phi\zeta})} (\frac{\phi}{\zeta})^{\frac{\varepsilon}{2}}.$$

Inserting the moments into (26) using (27) and the theorem of Beppo-Levi (see e.g. Hoffmann-Jørgensen (1994) 3.7) results in

$$a_c^{\mu,\nu}(\phi,\zeta)^{-1} = \frac{2}{\nu}\sum_{n=0}^{\infty}(\frac{\phi}{2})^n\frac{1}{n!}\sum_{k=0}^{n}(-1)^k\binom{n}{k}K_{\varepsilon_{kn}}(\sqrt{\phi\zeta})\left(\frac{\phi}{\zeta}\right)^{\frac{1}{2}\varepsilon_{kn}},$$

with $\varepsilon_{kn} = \frac{c}{\nu} - n + k(1-\frac{\mu}{\nu})$. This proves (24). In particular if $\mu = \nu$ then we obtain

$$a_c^{\mu,\mu}(\phi,\zeta) = \frac{\mu(\zeta/\phi)^{c/(2\mu)}}{2K_{c/\mu}(\sqrt{\phi\zeta})},$$

which is equal to (25). This follows from the fact that if $\mu = \nu$ then the Bessel functions become independent of k, and summation over terms involving k results in

$$\sum_{k=0}^{n}(-1)^k\binom{n}{k} = \begin{cases} 1 & n=0 \\ 0 & n>0. \end{cases}$$

The expression of the norming constant in the case $\mu = \nu$ is also known in the literature (see e.g. Jørgensen (1980)).
Finally let us note that a substitution of x into $y = x^\nu$, respectively $y = x^{-\mu}$ in the integral expression of $a_c^{\mu,\nu}(\phi,\zeta)$ yields the following relations

$$\nu a_{c/\nu}^{\mu/\nu,1}(\phi,\zeta) = a_c^{\mu,\nu}(\phi,\zeta) = \mu a_{-c/\mu}^{\nu/\mu,1}(\zeta,\phi),$$

which however do not give us much information on the norming constant.

5: References.

Barndorff-Nielsen, O.E. (1978): Information and Exponential Families. Wiley, Chichester.

Barndorff-Nielsen, O.E. (1988): Parametric Statistical Models and Likelihood. Lecture Notes in Statistics 50. Springer-Verlag.

Barndorff-Nielsen, O.E. and Blæsild, P. (1983a): Exponential Models with Affine Dual Foliations. Ann. Statist., 11 753-769.

Barndorff-Nielsen, O.E. and Blæsild, P. (1983b): Reproductive Exponential Families. Ann. Statist., 11 770-782.

Barndorff-Nielsen, O.E. and Blæsild, P. (1993): Orthogeodesic Models. Ann. Statist., 21 1018-1039.

Barndorff-Nielsen, Blæsild, P. and Eriksen, P.S. (1989): Decomposition and Invariance of Measures, and Statistical Transformation Models. Lecture Notes in Statistics 58. Springer-Verlag.

Barndorff-Nielsen, O.E. and Jørgensen, B. (1991): Proper Dispersion Models. Research Reports no.200. Department of Statistics. University of Aarhus.

Hoffmann-Jørgensen, J. (1994): Probability with a View towards Statistics. Volume 1. Chapman and Hall.

Jensen, J.L. (1981): On the Hyperboloid Distribution. Scand. J. Statist. 8 193-206.

Jørgensen, B. (1980): Statistical Properties of the Inverse Gaussian Distribution. Memoirs no.4. Department of Statistics. University of Aarhus.

Wiuf, C. (1994a): Dual Spaces and β-Orthogeodesic Models. Research Reports no.286. Department of Statistics. University of Aarhus.

Wiuf, C. (1994b): Geometric Characterization of 2-dimensional Orthogeodesic Exponential Composite Transformation Models. Unpublished report.