ghyp: A package on generalized hyperbolic distributions

Preliminary draft

Wolfgang Breymann
David Lüthi

Institute of Data Analysis and Process Design
www.idp.zhaw.ch

E-mail: david.luethi@zhaw.ch

March 7, 2008

In this document the R package ghyp is described in detail. Basically, the density functions of the generalized hyperbolic distribution and its special cases and fitting procedures. Some code chunks indicate how the package ghyp can be used.

Contents

1 Introduction

The origin of this package goes back to the first authors' years at RiskLab, when he worked together with Alexander McNeil to develop an algorithm for fitting multivariate generalized hyperbolic distributions. Accordingly, the functionality of this package partly overlaps McNeil's library QRMlib (McNeil, 2005). However, there are quite some differences in the implementation. From the user's point of view, one of the most important may be that one can choose between different parametrizations. In addition, with the present library it is possible to fit multivariate as well as univariate generalized hypberbolic distributions.

2 Definition

Facts about generalized hyperbolic (GH) distributions are cited according to McNeil, Frey, and Embrechts (2005) chapter 3.2.

The random vector $\mathbf{X}$ is said to have a multivariate GH distribution if

$$\mathbf{X} \stackrel{\mathrm{d}}{=} \mu + W\gamma + \sqrt{W}A\mathbf{Z} \tag{2.1}$$

where

(i) $\mathbf{Z} \sim N_k(\mathbf{0}, I_k)$

(ii) $A \in \mathbb{R}^{d \times k}$

(iii) $\mu, \gamma \in \mathbb{R}^d$

(iv) $W \geq 0$ is a scalar-valued random variable which is independent of $\mathbf{Z}$ and has a Generalized Inverse Gaussian distribution, written $GIG(\lambda, \chi, \psi)$ (see appendix C).

Note that the conditional distribution of $\mathbf{X}|W = w$ is normal,

$$\mathbf{X}|W = w \sim N_d(\mu + w\,\gamma, w\Sigma). \tag{2.2}$$

2.1 Expected value and variance

The expected value and the variance are given by

$$\begin{aligned} \mathrm{E}(\mathbf{X}) &= \mu + \mathrm{E}(W)\gamma & (2.3)\\ \mathrm{var}(\mathbf{X}) &= \mathrm{E}(\mathrm{cov}(\mathbf{X}|W)) + \mathrm{cov}(\mathrm{E}(\mathbf{X}|W)) & (2.4)\\ &= \mathrm{var}(W)\gamma\gamma' + \mathrm{E}(W)\Sigma \end{aligned}$$

where $\Sigma = AA'$.

2.2 Density

Since the conditional distribution of $\mathbf{X}$ given W is Gaussian with mean $\mu + W\gamma$ and variance $W\Sigma$ the GH density can be found by mixing $\mathbf{X}|W$ with respect to W.

$$
\begin{aligned}
f_{\mathbf{X}}(\mathbf{x}) &= \int_0^\infty f_{\mathbf{X}|W}(\mathbf{x}|w)\, f_W(w)\, \mathrm{d}w \qquad (2.5)\\
&= \int_0^\infty \frac{e^{(\mathbf{x}-\mu)'\Sigma^{-1}\gamma}}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}} w^{\frac{d}{2}}} \exp\left\{-\frac{\mathrm{Q}(\mathbf{x})}{2w} - \frac{\gamma'\Sigma^{-1}\gamma}{2/w}\right\} f_W(w)dw \\
&= \frac{(\sqrt{\psi/\chi})^\lambda(\psi+\gamma'\Sigma^{-1}\gamma)^{\frac{d}{2}-\lambda}}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}\mathrm{K}_\lambda(\sqrt{\chi\psi})} \times \frac{\mathrm{K}_{\lambda-\frac{d}{2}}(\sqrt{(\chi+\mathrm{Q}(\mathbf{x}))(\psi+\gamma'\Sigma^{-1}\gamma)})\ e^{(\mathbf{x}-\mu)'\Sigma^{-1}\gamma}}{(\sqrt{(\chi+\mathrm{Q}(\mathbf{x}))(\psi+\gamma'\Sigma^{-1}\gamma)})^{\frac{d}{2}-\lambda}},
\end{aligned}
$$

where the relation (B.2) for the modified Bessel function of the third kind $\mathrm{K}_\lambda(\cdot)$ (B.1) is used and $\mathrm{Q}(\mathbf{x})$ denotes the mahalanobis distance $\mathrm{Q}(\mathbf{x}) = (\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)$. The domain of variation of the parameters λ, χ and ψ is given in appendix C.

2.3 Linear transformations

The GH class is closed under linear operations:
If $\mathbf{X} \sim \mathrm{GH}_d(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$ and $\mathbf{Y} = B\mathbf{X} + \mathbf{b}$, where $B \in \mathbb{R}^{k \times d}$ and $\mathbf{b} \in \mathbb{R}^k$, then $Y \sim \mathrm{GH}_k(\lambda, \chi, \psi, B\mu + \mathbf{b}, B\Sigma B', B\gamma)$.

3 Special cases of the generalized hyperbolic distribution

The GH distribution contains several special cases known under special names.

- If $\lambda = \frac{d+1}{2}$ the name generalized is dropped and we have a multivariate hyperbolic (hyp) distribution. The univariate margins are still GH distributed. Inversely, when $\lambda = 1$ we get a multivariate GH distribution with hyperbolic margins.
- If $\lambda = -\frac{1}{2}$ the distribution is called Normal Inverse Gauss (NIG).
- If $\chi = 0$ and $\lambda > 0$ one gets a limiting case which is known amongst others as Variance Gamma (VG) distribution.
- If $\psi = 0$ and $\lambda < -1$ one gets a limiting case which is known as a skewed Student-t distribution.

Further information about the special cases and the necessary formulas to fit these distributions to data can be found in the appendixes C and D. The parameter constraints for the special cases in different parametrizations are described in the following chapter.

4 Parametrization

There are several alternative parametrizations for the GH distribution. In this package the user can choose between three of them. Appendix G.1 explains how to use.
The following table describes the parameter ranges for each parametrization and each special case. Clearly, the dispersion matrices Σ and Δ have to fulfill the usual conditions for covariance matrices, i.e., symmetry and positive definiteness. We denote the set of all feasible covariance matrices in $\mathbb{R}^{d\times d}$ with $\mathbb{R}^{\Sigma}$. Furthermore, let $\mathbb{R}^{\Delta} = \{A \in \mathbb{R}^{\Sigma} : |A| = 1\}$.

	$(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$-Parametrization					
	λ	χ	ψ	μ	Σ	γ
ghyp	$\lambda \in \mathbb{R}$	$\chi > 0$	$\psi > 0$	$\mu \in \mathbb{R}^d$	$\Sigma \in \mathbb{R}^{\Sigma}$	$\gamma \in \mathbb{R}^d$
hyp	$\lambda = \frac{d+1}{2}$	$\chi > 0$	$\psi > 0$	$\mu \in \mathbb{R}^d$	$\Sigma \in \mathbb{R}^{\Sigma}$	$\gamma \in \mathbb{R}^d$
NIG	$\lambda = -\frac{1}{2}$	$\chi > 0$	$\psi > 0$	$\mu \in \mathbb{R}^d$	$\Sigma \in \mathbb{R}^{\Sigma}$	$\gamma \in \mathbb{R}^d$
t	$\lambda < 0$	$\chi > 0$	$\psi = 0$	$\mu \in \mathbb{R}^d$	$\Sigma \in \mathbb{R}^{\Sigma}$	$\gamma \in \mathbb{R}^d$
VG	$\lambda > 0$	$\chi = 0$	$\psi > 0$	$\mu \in \mathbb{R}^d$	$\Sigma \in \mathbb{R}^{\Sigma}$	$\gamma \in \mathbb{R}^d$

	$(\lambda, \overline{\alpha}, \mu, \Sigma, \gamma)$-Parametrization				
	λ	$\overline{\alpha}$	μ	Σ	γ
ghyp	$\lambda \in \mathbb{R}$	$\overline{\alpha} > 0$	$\mu \in \mathbb{R}^d$	$\Sigma \in \mathbb{R}^{\Sigma}$	$\gamma \in \mathbb{R}^d$
hyp	$\lambda = \frac{d+1}{2}$	$\overline{\alpha} > 0$	$\mu \in \mathbb{R}^d$	$\Sigma \in \mathbb{R}^{\Sigma}$	$\gamma \in \mathbb{R}^d$
NIG	$\lambda = \frac{1}{2}$	$\overline{\alpha} > 0$	$\mu \in \mathbb{R}^d$	$\Sigma \in \mathbb{R}^{\Sigma}$	$\gamma \in \mathbb{R}^d$
t	$\lambda = -\frac{\nu}{2} < -1$	$\overline{\alpha} = 0$	$\mu \in \mathbb{R}^d$	$\Sigma \in \mathbb{R}^{\Sigma}$	$\gamma \in \mathbb{R}^d$
VG	$\lambda > 0$	$\overline{\alpha} = 0$	$\mu \in \mathbb{R}^d$	$\Sigma \in \mathbb{R}^{\Sigma}$	$\gamma \in \mathbb{R}^d$

	$(\lambda, \alpha, \mu, \Delta, \delta, \beta)$-Parametrization					
	λ	α	δ	μ	Δ	β
ghyp	$\lambda \in \mathbb{R}$	$\alpha > 0$	$\delta > 0$	$\mu \in \mathbb{R}^d$	$\Delta \in \mathbb{R}^{\Delta}$	$\beta \in \{\mathbf{x} \in \mathbb{R}^d : \alpha^2 - \mathbf{x}'\Delta\mathbf{x} > 0\}$
hyp	$\lambda = \frac{d+1}{2}$	$\alpha > 0$	$\delta > 0$	$\mu \in \mathbb{R}^d$	$\Delta \in \mathbb{R}^{\Delta}$	$\beta \in \{\mathbf{x} \in \mathbb{R}^d : \alpha^2 - \mathbf{x}'\Delta\mathbf{x} > 0\}$
NIG	$\lambda = -\frac{1}{2}$	$\alpha > 0$	$\delta > 0$	$\mu \in \mathbb{R}^d$	$\Delta \in \mathbb{R}^{\Delta}$	$\beta \in \{\mathbf{x} \in \mathbb{R}^d : \alpha^2 - \mathbf{x}'\Delta\mathbf{x} > 0\}$
t	$\lambda < 0$	$\alpha = \sqrt{\beta'\Delta\beta}$	$\delta > 0$	$\mu \in \mathbb{R}^d$	$\Delta \in \mathbb{R}^{\Delta}$	$\beta \in \mathbb{R}^d$
VG	$\lambda > 0$	$\alpha > 0$	$\delta = 0$	$\mu \in \mathbb{R}^d$	$\Delta \in \mathbb{R}^{\Delta}$	$\beta \in \{\mathbf{x} \in \mathbb{R}^d : \alpha^2 - \mathbf{x}'\Delta\mathbf{x} > 0\}$

Internally, he package `ghyp` uses the $(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$-parametrization. However, fitting is done in the $(\lambda, \overline{\alpha}, \mu, \Sigma, \gamma)$-parametrization since this parametrization does not necessitate additional constraints to eliminate the redundant degree of freedom. Consequently, what cannot be represented by the $(\lambda, \overline{\alpha}, \mu, \Sigma, \gamma)$-parametrization cannot be fitted (see section 4.2).

4.1 $(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$-Parametrization

The $(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$-parametrization is obtained as the normal mean-variance mixture distribution when $W \sim GIG(\lambda, \chi, \psi)$. This parametrization has a drawback of an identification problem. Indeed, the distributions $\mathrm{GH}_d(\lambda,\chi,\psi,\mu,\Sigma,\gamma)$ and $\mathrm{GH}_d(\lambda,\chi/k,k\psi,\mu,k\Sigma,k\gamma)$ are identical for any $k>0$. Therefore, an identifying problem occurs when we start to fit the parameters of a GH distribution. This problem could be solved by introducing a suitable contraint. One possibility is to require the determinant of the dispersion matrix Σ to be 1.

4.2 $(\lambda,\overline{\alpha},\mu,\Sigma,\gamma)$-Parametrization

There is a more elegant way to eliminate the degree of freedom. We simply constrain the expected value of the mixing variable W to be 1. This makes the interpretation of the skewness parameters γ easier and in addition, the fitting procedure becomes faster (see 5.1).
We define

$$\mathrm{E}(W)=\sqrt{\frac{\chi}{\psi}}\frac{\mathrm{K}_{\lambda+1}(\sqrt{\chi\psi})}{\mathrm{K}_{\lambda}(\sqrt{\chi\psi})}=1. \tag{4.1}$$

and set

$$\overline{\alpha}=\sqrt{\chi\psi}. \tag{4.2}$$

It follows that

$$\psi=\overline{\alpha}\,\frac{\mathrm{K}_{\lambda+1}(\overline{\alpha})}{\mathrm{K}_{\lambda}(\overline{\alpha})}\ \text{ and }\ \chi=\frac{\overline{\alpha}^2}{\psi}=\overline{\alpha}\,\frac{\mathrm{K}_{\lambda}(\overline{\alpha})}{\mathrm{K}_{\lambda+1}(\overline{\alpha})}. \tag{4.3}$$

The drawback of the $(\lambda,\overline{\alpha},\mu,\Sigma,\gamma)$-parametrization is that it does not exist in the case $\overline{\alpha}=0$ and $\lambda\in[-1,0]$. This is the case of a Student-t distribution with non-existing variance. Note that the $(\lambda,\overline{\alpha},\mu,\Sigma,\gamma)$-parametrization yields to a slightly different parametrization for the special case of a Student-t distribution (see section D.1 for details). The limit of the equations (4.3) as $\overline{\alpha}\downarrow 0$ can be found in (D.3) and (D.9).

4.3 $(\lambda,\alpha,\mu,\Delta,\delta,\beta)$-Parametrization

When the GH distribution was introduced in Barndorff-Nielsen (1977), the following parametrization for the multivariate case was used.

$$f_{\mathbf{X}}(\mathbf{x})=\frac{(\alpha^2-\beta'\Delta\beta)^{\lambda/2}}{(2\pi)^{\frac{d}{2}}\sqrt{|\Delta|}\,\delta^{\lambda}\,\mathrm{K}_{\lambda}(\delta\sqrt{\alpha^2-\beta'\Delta\beta})}\times\frac{\mathrm{K}_{\lambda-\frac{d}{2}}(\alpha\sqrt{\delta^2+(\mathbf{x}-\mu)'\Delta^{-1}(\mathbf{x}-\mu)})\ e^{\beta'(\mathbf{x}-\mu)}}{(\alpha\sqrt{\delta^2+(\mathbf{x}-\mu)'\Delta^{-1}(\mathbf{x}-\mu)})^{\frac{d}{2}-\lambda}}, \tag{4.4}$$

where the determinant of Δ is constrained to be 1. In the univariate case the above expression reduces to

$$f_{\mathbf{X}}(\mathbf{x})=\frac{(\alpha^2-\beta^2)^{\lambda/2}}{\sqrt{2\pi}\,\alpha^{\lambda-\frac{1}{2}}\,\delta^{\lambda}\,\mathrm{K}_{\lambda}(\delta\sqrt{\alpha^2-\beta^2})}\times\mathrm{K}_{\lambda-\frac{1}{2}}(\alpha\sqrt{\delta^2+(x-\mu)^2})\ e^{\beta(x-\mu)}, \tag{4.5}$$

which is the most widely used parametrization of the GH distribution in literature.

4.4 Switching between different parametrizations

The following formulas can be used to switch from the $(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$-parametrization to the $(\lambda, \alpha, \mu, \Delta, \delta, \beta)$-parametrization and vice versa:

$$\begin{aligned} \Delta = |\Sigma|^{-\frac{1}{d}}\Sigma \quad &, \quad \beta = \Sigma^{-1}\gamma \\ \delta = \sqrt{\chi|\Sigma|^{\frac{1}{d}}} \quad &, \quad \alpha = \sqrt{|\Sigma|^{-\frac{1}{d}}(\psi + \gamma'\Sigma^{-1}\gamma)} \end{aligned} \tag{4.6}$$

$$\Sigma = \Delta, \quad \gamma = \Delta\beta, \quad \chi = \delta^2, \quad \psi = \alpha^2 - \beta'\Delta\beta. \tag{4.7}$$

To switch from the $(\lambda, \overline{\alpha}, \mu, \Sigma, \gamma)$-parametrization to the $(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$-parametrization formulas (4.3) can be used.

5 Fitting generalized hyperbolic distributions to data

Numerical optimizers can be used to fit univariate GH distributions to data by means of maximum likelihood estimation. Multivariate GH distributions can be fitted with algorithms based on the expectation-maximazion (EM) scheme.

5.1 EM-Scheme

Assume we have iid data $\mathbf{x}_1, \ldots, \mathbf{x}_n$ and parameters represented by $\Theta = (\lambda, \overline{\alpha}, \mu, \Sigma, \gamma)$. The problem is to maximize

$$\ln L(\Theta; \mathbf{x}_1, \ldots, \mathbf{x}_n) = \sum_{i=1}^{n} \ln f_{\mathbf{X}}(\mathbf{x}_i; \Theta). \tag{5.1}$$

This problem is not easy to solve due to the number of parameters and necessity of maximizing over covariance matrices. We can proceed by introducing an augmented likelihood function

$$\ln \tilde{L}(\Theta; \mathbf{x}_1, \ldots, \mathbf{x}_n, w_1, \ldots, w_n) = \sum_{i=1}^{n} \ln f_{\mathbf{X}|W}(\mathbf{x}_i | w_i; \mu, \Sigma, \gamma) + \sum_{i=1}^{n} \ln f_W(w_i; \lambda, \overline{\alpha}) \tag{5.2}$$

and spend the effort on the estimation of the latent mixing variables w_i coming from the mixture representation (2.2). This is where the EM algorithm comes into play.

E-step: Calculate the conditional expectation of the likelihood function (5.2) given the data $\mathbf{x}_1, \ldots, \mathbf{x}_n$ and the current estimates of parameters $\Theta^{[k]}$. This results in the objective function

$$Q(\Theta; \Theta^{[k]}) = \mathrm{E}(\ln \tilde{L}(\Theta; \mathbf{x}_1, \ldots, \mathbf{x}_n, w_1, \ldots, w_n) | \mathbf{x}_1, \ldots, \mathbf{x}_n; \Theta^{[k]}). \tag{5.3}$$

M-step: Maximize the objective function with respect to Θ to obtain the next set of estimates $\Theta^{[k+1]}$.

Alternating between these steps yields to the maximum likelihood estimation of the parameter set Θ.

In practice, performing the E-Step means maximizing the second summand of (5.2) numerically. The log density of the GIG distribution (see C.1) is

$$\ln f_W(w) = \frac{\lambda}{2}\ln(\psi/\chi) - \ln(2\mathrm{K}_\lambda\sqrt{\chi\psi}) + (\lambda-1)\ln w - \frac{\chi}{2}\frac{1}{w} - \frac{\psi}{2}w. \tag{5.4}$$

When using the $(\lambda, \overline{\alpha})$-parametrization this problem is of dimension two instead of three as it is in the (λ, χ, ψ)-parametrization As a consequence the performance increases.
Since the w_i's are latent one has to replace w, $1/w$ and $\ln w$ with the respective expected values in order to maximize the log likelihood function. Let

$$\eta_i^{[k]} := \mathrm{E}(w_i \,|\, \mathbf{x}_i; \Theta^{[k]}), \;\; \delta_i^{[k]} := \mathrm{E}(w_i^{-1} \,|\, \mathbf{x}_i; \Theta^{[k]}), \;\; \xi_i^{[k]} := \mathrm{E}(\ln w_i \,|\, \mathbf{x}_i; \Theta^{[k]}). \tag{5.5}$$

We have to find the conditional density of w_i given $\mathbf{x}_i$ to be able to calculate these quantities (see (E.1)).

5.2 MCECM estimation

In the `R` implementation a modified EM scheme is used, which is called multi-cycle, expectation, conditional estimation (MCECM) algorithm (McNeil, Frey, and Embrechts, 2005; McNeil, 2005). The different steps of the MCECM algorithm are sketched as follows:

(1) Select reasonable starting values for $\Theta^{[k]}$. For example $\lambda = 1$, $\overline{\alpha} = 1$, μ is set to the sample mean, Σ to the sample covariance matrix and γ to a zero skewness vector.

(2) Calculate $\chi^{[k]}$ and $\psi^{[k]}$ as a function of $\overline{\alpha}^{[k]}$ using (4.3).

(3) Use (5.5), (C.2) and (E.1) to calculate the weights $\eta_i^{[k]}$ and $\delta_i^{[k]}$. Average the weights to get

$$\bar{\eta}^{[k]} = \frac{1}{n}\sum_{i=1}^{n}\eta_i^{[k]} \text{ and } \bar{\delta}^{[k]} = \frac{1}{n}\sum_{i=1}^{n}\delta_i^{[k]}. \tag{5.6}$$

(4) If an asymmetric model is to be fitted set γ to $\mathbf{0}$, else set

$$\gamma^{[k+1]} = \frac{1}{n}\frac{\sum_{i=1}^{n}\delta_i^{[k]}(\bar{\mathbf{x}} - \mathbf{x}_i)}{\bar{\eta}^{[k]}\bar{\delta}^{[k]} - 1}. \tag{5.7}$$

(5) Update μ and Σ:

$$\mu^{[k+1]} = \frac{1}{n}\frac{\sum_{i=1}^{n}\delta_i^{[k]}(\mathbf{x}_i - \gamma^{[k+1]})}{\bar{\delta}^{[k]}} \tag{5.8}$$

$$\Sigma^{[k+1]} = \frac{1}{n}\sum_{i=1}^{n}\delta_i^{[k]}(\mathbf{x}_i - \mu^{[k+1]})(\mathbf{x}_i - \mu^{[k+1]})' - \bar{\eta}^{[k]}\gamma^{[k+1]}\gamma^{[k+1]\,\prime}. \tag{5.9}$$

(6) Set $\Theta^{[k,2]} = (\lambda^{[k]}, \overline{\alpha}^{[k]}, \mu^{[k+1]}, \Sigma^{[k+1]}, \gamma^{[k+1]})$ and calculate weights $\eta_i^{[k,2]}$, $\delta_i^{[k,2]}$ and $\xi_i^{[k,2]}$ using (5.5), (C.3) and (C.2).

(7) Maximise the second summand of (5.2) with respect to λ, χ and ψ to complete the calculation of $\Theta^{[k,2]}$ and go back to step (2). Note that the objective function must calculate χ and ψ in dependence of λ and $\overline{\alpha}$ using relation (4.3).

A Shape of the univariate generalized hyperbolic distribution

`ghyp, 2007, Institute of Data Analysis and Process Design, GPL`

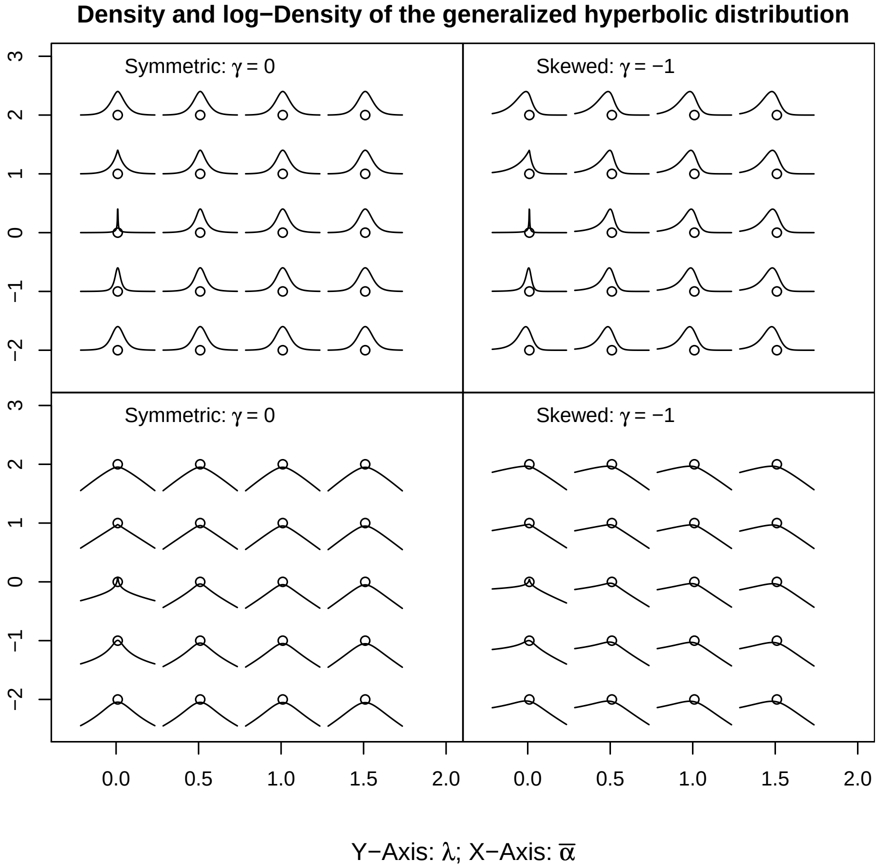


Figure A.1: The shape of the univariate generalized hyperbolic density drawn with different shape parameters $(\lambda, \overline{\alpha})$. The location and scale parameter μ and σ are set to 0 and 1, respectively. The skewness parameter γ is 0 in the left column and -1 in the right column of the graphics array.

B Modified Bessel function of the third kind

The modified Bessel function of the third kind appears in the GH as well as in the GIG density (2.5, C.1). This function has the integral representation

$$K_\lambda(x) = \frac{1}{2}\int_0^\infty w^{\lambda-1} \exp\left\{-\frac{1}{2}x\left(w + w^{-1}\right)\right\} \mathrm{d}w\,, \qquad x > 0. \tag{B.1}$$

The substitution $w = x\sqrt{\chi/\psi}$ can be used to obtain the following relation, which allows one to bring the GH density (2.5) into a closed-form expression.

$$\int_0^\infty w^{\lambda-1} \exp\left\{-\frac{1}{2}\left(\frac{\chi}{w} + w\psi\right)\right\} \mathrm{d}w = 2\left(\frac{\chi}{\psi}\right)^{\frac{\lambda}{2}} K_\lambda(\sqrt{\chi\psi}) \tag{B.2}$$

When calculating the densities of the special cases of the GH density we can use the asymtotic relations for small arguments x

$$\mathrm{K}_\lambda(x) \sim \Gamma(\lambda)\, 2^{\lambda-1} x^{-\lambda} \quad \text{as} \quad x \downarrow 0 \quad \text{and} \quad \lambda > 0 \tag{B.3}$$

and

$$\mathrm{K}_\lambda(x) \sim \Gamma(-\lambda)\, 2^{-\lambda-1} x^{\lambda} \quad \text{as} \quad x \downarrow 0 \quad \text{and} \quad \lambda < 0. \tag{B.4}$$

(B.4) follows from (B.3) and the observation that the Bessel function is symmetric with respect to the index λ. An asymptotic relation for large arguments x is given by

$$\mathrm{K}_\lambda(x) \sim \sqrt{\frac{\pi}{2}}\, x^{-\frac{1}{2}}\, e^{-x} \quad \text{as} \quad x \to \infty. \tag{B.5}$$

We refer to Abramowitz and Stegun (1972) for further information on Bessel functions.

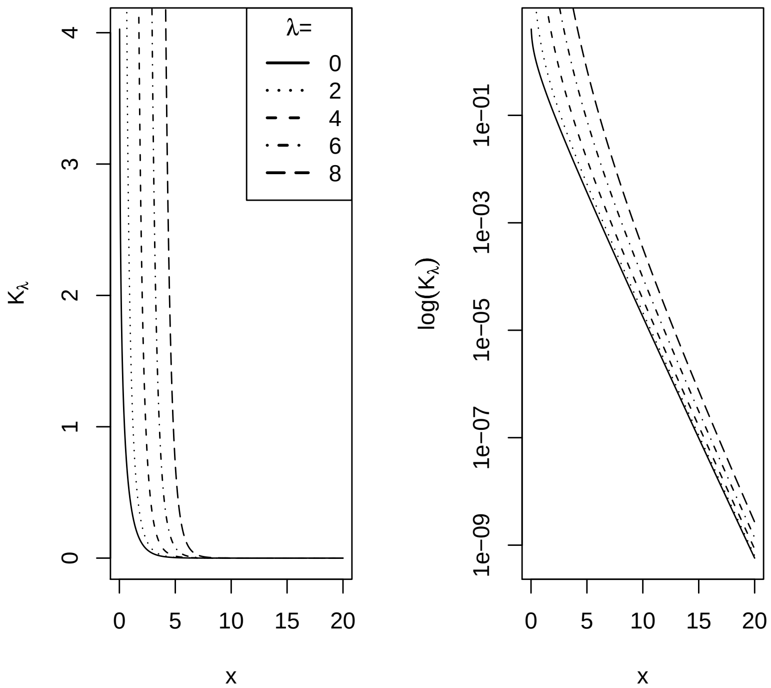


Figure B.1: The modified Bessel function of the third kind drawn with different indices λ.

C Generalized Inverse Gaussian distribution

The density of a Generalized Inverse Gaussian (GIG) distribution is given as

$$f_{GIG}(w) = \left(\frac{\psi}{\chi}\right)^{\frac{\lambda}{2}} \frac{w^{\lambda-1}}{2K_\lambda(\sqrt{\chi\psi})} \exp\left\{-\frac{1}{2}\left(\frac{\chi}{w} + \psi w\right)\right\}, \tag{C.1}$$

with parameters satisfying

$$\begin{aligned}
&\chi > 0, \psi \geq 0, \quad \lambda < 0 \\
&\chi > 0, \psi > 0, \quad \lambda = 0 \\
&\chi \geq 0, \psi > 0, \quad \lambda > 0\ .
\end{aligned}$$

The GIG density contains the Gamma (Γ) and Inverse Gamma ($\mathrm{I}\Gamma$) densities as limiting cases. If $\chi = 0$ and $\lambda > 0$ then X is gamma distributed with parameters λ and $\frac{1}{2}\psi$ ($\Gamma(\lambda, \frac{1}{2}\psi)$).
If $\psi = 0$ and $\lambda < 0$ then X has an inverse gamma distribution with parameters $-\lambda$ and $\frac{1}{2}\chi$ ($\mathrm{I}\Gamma(-\lambda, \frac{1}{2}\chi)$).

The n-th moment of a GIG distributed random variable can be found to be

$$\mathrm{E}(X^n) = \left(\frac{\chi}{\psi}\right)^{\frac{n}{2}} \frac{\mathrm{K}_{\lambda+n}(\sqrt{\chi\psi})}{\mathrm{K}_\lambda(\sqrt{\chi\psi})}. \tag{C.2}$$

Furthermore

$$\mathrm{E}(\ln X) = \left.\frac{\mathrm{dE}(X^\alpha)}{\mathrm{d}\alpha}\right|_{\alpha=0}. \tag{C.3}$$

Numerical calculations may be performed with the integral representation as well. In the `R` package `ghyp` the derivative construction is implemented.

C.1 Gamma distribution

When $\chi = 0$ and $\lambda > 0$ the GIG distribution reduces to the gamma distribution defined as

$$f_W(w) = \frac{\beta^\alpha}{\Gamma(\alpha)} w^{\alpha-1} \exp\{-\beta w\}.$$

The expected value and the variance are

$$\mathrm{E}(X) = \frac{\beta}{\alpha} \quad \text{and} \quad \mathrm{var}(X) = \frac{\alpha}{\beta^2}, \tag{C.4}$$

respectively. The expected value of the logarithm is $\mathrm{E}(\ln X) = \psi(\alpha) - \ln(\beta)$ where $\psi(\cdot)$ is the digamma function. We will see that this value is not needed to fit a multivariate variance gamma distribution (see E.3).

C.2 Inverse gamma distribution

When $\psi = 0$ and $\lambda < 0$ the GIG distribution reduces to the gamma distribution defined as

$$f_W(w) = \frac{\beta^\alpha}{\Gamma(\alpha)} w^{-\alpha-1} \exp\left\{-\frac{\beta}{w}\right\}.$$

The expected value and the variance are

$$\mathrm{E}(X) = \frac{\beta}{\alpha - 1} \quad \text{and} \quad \mathrm{var}(X) = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}, \tag{C.5}$$

and exist provided that $\alpha > 1$ and $\alpha > 2$, respectively. The expected value of the logarithm is $\mathrm{E}(\ln X) = \ln(\beta) - \psi(\alpha)$. This value is required in order to fit a symmetric multivariate Student-t distribution by means of the MCECM algorithm (see E.2).

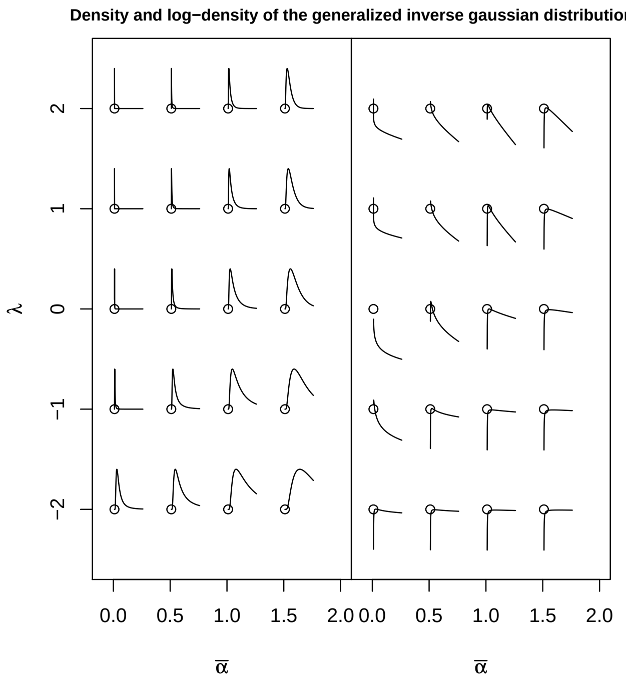


Figure C.1: The density and the log-density of the generalized inverse gaussian distribution drawn with different shape parameters $(\lambda, \overline{\alpha})$. See (4.3) for the transformation from $\overline{\alpha}$ to (χ, ψ).

D Densities of the special cases of the GH distribution

As mentioned in section 3 the GH distribution contains several special cases. In what follows the densities of the special cases are derived. In the case of a hyperbolic or normal inverse gaussian distribution we simply fix the parameter λ either to $(d+1)/2$ or -0.5.

D.1 Student-t distribution

With relation (B.4) it can be easily shown that when $\psi \to 0$ and $\lambda < 0$ the density of a GH distribution results in

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{\chi^{-\lambda}(\gamma'\Sigma^{-1}\gamma)^{\frac{d}{2}-\lambda}}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}\Gamma(-\lambda)2^{-\lambda-1}} \times \frac{\mathrm{K}_{\lambda-\frac{d}{2}}(\sqrt{(\chi+\mathrm{Q}(\mathbf{x}))\gamma'\Sigma^{-1}\gamma})e^{(\mathbf{x}-\mu)'\Sigma^{-1}\gamma}}{(\sqrt{(\chi+\mathrm{Q}(\mathbf{x}))\gamma'\Sigma^{-1}\gamma})^{\frac{d}{2}-\lambda}}. \tag{D.1}$$

As $\gamma \to 0$ we obtain again with relation (B.4) the symmetric multivariate Student-t density

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{\chi^{-\lambda}\Gamma(-\lambda+\frac{d}{2})}{\pi^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}\Gamma(-\lambda)} \times (\chi+\mathrm{Q}(\mathbf{x}))^{\lambda-\frac{d}{2}}. \tag{D.2}$$

We switch to the Student-t parametrization and set the degree of freedom $\nu = -2\lambda$ [1]. Because $\psi = 0$ the transformation of $\overline{\alpha}$ to χ and ψ (see 4.3) reduces to

$$\chi = \overline{\alpha}\ \frac{\mathrm{K}_\lambda(\overline{\alpha})}{\mathrm{K}_{\lambda+1}(\overline{\alpha})} \overset{\overline{\alpha}\to 0}{\longrightarrow} 2\ (-\lambda-1) = \nu - 2. \tag{D.3}$$

Plugging in the values for λ and ν, the densities take the form

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{(\nu-2)^{\frac{\nu}{2}}(\gamma'\Sigma^{-1}\gamma)^{\frac{\nu+d}{2}}}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}\Gamma(\frac{\nu}{2})2^{\frac{\nu}{2}-1}} \times \frac{\mathrm{K}_{\frac{\nu+d}{2}}(\sqrt{(\nu-2+\mathrm{Q}(\mathbf{x}))\gamma'\Sigma^{-1}\gamma})\ e^{(\mathbf{x}-\mu)'\Sigma^{-1}\gamma}}{(\sqrt{(\nu-2+\mathrm{Q}(\mathbf{x}))\gamma'\Sigma^{-1}\gamma})^{\frac{\nu+d}{2}}} \tag{D.4}$$

and for the symmetric case as $\gamma \to 0$

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{(\nu-2)^{\frac{\nu}{2}}\Gamma(\frac{\nu+d}{2})}{\pi^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}\Gamma(\frac{\nu}{2})(\nu-2+\mathrm{Q}(\mathbf{x}))^{\frac{\nu+d}{2}}}. \tag{D.5}$$

It is important to note that the variance does not exist in the symmetric case for $\nu \leq 2$ while for the asymmetric case the variance does not exist for $\nu \leq 4$. This is because the variance of an asymmetric GH distribution involves the variance of the mixing distribution. In case of a Student-t distribution, the mixing variable w is inverse gamma distributed and has finite variance only if $\beta > 2$ which corresponds to $\lambda < -2$, i.e. $\nu > 4$ (see C.5).

[1] Note that the $(\lambda, \overline{\alpha}, \mu, \Sigma, \gamma)$ parametrization yields to a slightly different Student-t parametrization: In this package the parameter Σ denotes the variance in the multivariate case and the standard deviation in the univariate case. Thus, set $\sigma = \sqrt{\nu/(\nu-2)}$ in the univariate case to get the same results as with the standard `R` implementation of the Student-t distribution.

Alternatively, in the univariate case, this can be seen by the fact that the Student-t density has regularly varying tails. For $x \to \infty$, one obtains

$$f_X(x) = L(x)\, x^{-\nu-1}, \quad \text{for } \gamma = 0 \tag{D.6}$$

$$f_X(x) = L(x)\, x^{-\frac{\nu}{2}-1}, \quad \text{for } \gamma > 0, \tag{D.7}$$

where $L(x)$ denotes a slowly varying function at ∞. The asymptotic relation for the modified Bessel function of the third kind (B.5) was applied to (D.4) to arrive at (D.7).

D.2 Variance gamma distribution

Relation (B.3) can be used again to show that for $\chi \to 0$ and $\lambda > 0$ the density of the GH distribution results in

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{\psi^{\lambda}(\psi + \gamma'\Sigma^{-1}\gamma)^{\frac{d}{2}-\lambda}}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}\Gamma(\lambda)2^{\lambda-1}} \times \frac{\mathrm{K}_{\lambda-\frac{d}{2}}(\sqrt{\mathrm{Q}(\mathbf{x})(\psi + \gamma'\Sigma^{-1}\gamma)})\; e^{(\mathbf{x}-\mu)'\Sigma^{-1}\gamma}}{(\sqrt{\mathrm{Q}(\mathbf{x})(\psi + \gamma'\Sigma^{-1}\gamma)})^{\frac{d}{2}-\lambda}}. \tag{D.8}$$

In the case of a variance gamma distribution the transformation of $\overline{\alpha}$ to χ and ψ (see 4.3) reduces to

$$\psi = \overline{\alpha}\, \frac{\mathrm{K}_{\lambda+1}(\overline{\alpha})}{\mathrm{K}_{\lambda}(\overline{\alpha})} = 2\lambda \tag{D.9}$$

Thus, the density is

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{2\,\lambda^{\lambda}(2\lambda + \gamma'\Sigma^{-1}\gamma)^{\frac{d}{2}-\lambda}}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}\Gamma(\lambda)} \times \frac{\mathrm{K}_{\lambda-\frac{d}{2}}(\sqrt{\mathrm{Q}(\mathbf{x})(2\lambda + \gamma'\Sigma^{-1}\gamma)})\; e^{(\mathbf{x}-\mu)'\Sigma^{-1}\gamma}}{(\sqrt{\mathrm{Q}(\mathbf{x})(2\lambda + \gamma'\Sigma^{-1}\gamma)})^{\frac{d}{2}-\lambda}}. \tag{D.10}$$

A limiting case arises when $\mathrm{Q}(\mathbf{x}) \to 0$, that is when $\mathbf{x} - \mu \to 0$. As long as $\lambda - \frac{d}{2} > 0$ relation (B.3) can be used to verify that the density reduces to

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{\psi^{\lambda}(\psi + \gamma'\Sigma^{-1}\gamma)^{\frac{d}{2}-\lambda}\,\Gamma(\lambda - \frac{d}{2})}{2^{d}\,\pi^{\frac{d}{2}}\,|\Sigma|^{\frac{1}{2}}\,\Gamma(\lambda)}. \tag{D.11}$$

By replacing ψ with 2λ the limiting density is obtained in the $(\lambda, \overline{\alpha}, \mu, \Sigma, \gamma)$-parametrization. [2]

For $\lambda - \frac{d}{2} \leq 0$ the density diverges. [3]

[2] The numeric implementation in R uses spline interpolation for the case where $\lambda - \frac{d}{2} > 0$ and $\mathrm{Q}(\mathbf{x}) < \epsilon$.

[3] The current workaround in R simply sets observations where $\mathrm{Q}(\mathbf{x}) < \epsilon$ to ϵ when $\lambda - \frac{d}{2} \leq 0$.

E Conditional density of the mixing variable W

Performing the E-Step of the MCECM algorithm requires the calculation of the conditional expectation of w_i given $\mathbf{x}_i$. In this section the conditional density is derived.

E.1 Generalized hyperbolic, hyperbolic and NIG distribution

The mixing term w is GIG distributed. By using (2.5) and (C.1) the density of w_i given $\mathbf{x}_i$ can be calculated to be again the GIG density with parameters $(\lambda - \frac{d}{2}, \mathrm{Q}(\mathbf{x}) + \chi, \psi + \gamma'\Sigma^{-1}\gamma)$.

$$\begin{aligned} f_{w|\mathbf{x}}(w) &= \frac{f_{\mathbf{X},W}(\mathbf{x}, w)}{f_{\mathbf{X}}(\mathbf{x})} \\ &= \frac{f_{\mathbf{X}|W}(\mathbf{x}) f_{GIG}(w)}{\int_0^\infty f_{\mathbf{X}|W}(\mathbf{x}) f_{GIG}(w) \mathrm{d}w} \\ &= \left(\frac{\gamma'\Sigma^{-1}\gamma + \psi}{\mathrm{Q}(\mathbf{x}) + \chi}\right)^{0.5(\lambda - \frac{d}{2})} \times \\ &\quad \frac{w^{\lambda - \frac{d}{2} - 1} \exp\left\{-\frac{1}{2}\left(\frac{\mathrm{Q}(\mathbf{x})+\chi}{w} + w\,(\gamma'\Sigma^{-1}\gamma + \psi)\right)\right\}}{2\, K_{\lambda - \frac{d}{2}}(\sqrt{(\mathrm{Q}(\mathbf{x}) + \chi)\,(\gamma'\Sigma^{-1}\gamma + \psi)})} \end{aligned} \tag{E.1}$$

E.2 Student-t distribution

The mixing term w is IΓ distributed. Again the conditional density of w_i given $\mathbf{x}_i$ results in the GIG density. The equations (2.5) and (C.5) were used. The parameters of the GIG density are $(\lambda - \frac{d}{2}, \mathrm{Q}(\mathbf{x}) + \chi, \gamma'\Sigma^{-1}\gamma)$. When γ becomes 0 the conditional density reduces to the IΓ density with parameters $(\frac{d}{2} - \lambda, \frac{\mathrm{Q}(\mathbf{x})+\chi}{2})$.

$$\begin{aligned} f_{w|\mathbf{x}}(w) &= \frac{f_{\mathbf{X},W}(\mathbf{x}, w)}{f_{\mathbf{X}}(\mathbf{x})} \\ &= \frac{f_{\mathbf{X}|W}(\mathbf{x}) f_{\mathrm{I}\Gamma}(w)}{\int_0^\infty f_{\mathbf{X}|W}(\mathbf{x}) f_{\mathrm{I}\Gamma}(w) \mathrm{d}w} \\ &= \left(\frac{\gamma'\Sigma^{-1}\gamma}{\mathrm{Q}(\mathbf{x}) + \chi}\right)^{0.5(\lambda - \frac{d}{2})} \times \frac{w^{\lambda - \frac{d}{2} - 1} \exp\left\{-\frac{1}{2}\left(\frac{\mathrm{Q}(\mathbf{x})+\chi}{w} + w\,\gamma'\Sigma^{-1}\gamma\right)\right\}}{2\, K_{\lambda - \frac{d}{2}}(\sqrt{(\mathrm{Q}(\mathbf{x}) + \chi)\,\gamma'\Sigma^{-1}\gamma})} \end{aligned} \tag{E.2}$$

E.3 Variance gamma distribution

The mixing term w is Γ distributed. By using (2.5) and (C.4) the density of w_i given $\mathbf{x}_i$ can be calculated to be again the GIG density with parameters $(\lambda - \frac{d}{2}, \mathrm{Q}(\mathbf{x}), \psi + \gamma'\Sigma^{-1}\gamma)$.

$$\begin{aligned} f_{w|\mathbf{x}}(w) &= \frac{f_{\mathbf{X},W}(\mathbf{x},w)}{f_{\mathbf{X}}(\mathbf{x})} \\ &= \frac{f_{\mathbf{X}|W}(\mathbf{x}) f_{\Gamma}(w)}{\int_0^\infty f_{\mathbf{X}|W}(\mathbf{x}) f_{\Gamma}(w) \mathrm{d}w} \\ &= \left(\frac{\gamma'\Sigma^{-1}\gamma + \psi}{\mathrm{Q}(\mathbf{x})}\right)^{0.5(\lambda - \frac{d}{2})} \times && \text{(E.3)} \\ & \quad \frac{w^{\lambda - \frac{d}{2} - 1} \exp\left\{-\frac{1}{2}\left(\frac{\mathrm{Q}(\mathbf{x})}{w} + w\left(\gamma'\Sigma^{-1}\gamma + \psi\right)\right)\right\}}{2\, K_{\lambda - \frac{d}{2}}(\sqrt{\mathrm{Q}(\mathbf{x})\left(\gamma'\Sigma^{-1}\gamma + \psi\right)})} && \text{(E.4)} \end{aligned}$$

F Distribution objects

In the package `ghyp` we follow an object-oriented programming approach and introduce distribution objects. There are mainly four reasons for that:

1. Unlike most distributions the GH distribution has quite a few parameters which have to fulfill some consistency requirements. Consistency checks can be performed uniquely when an object is initialized.

2. Once initialized the common functions belonging to a distribution can be called conveniently by passing the distribution object. A repeated input of the parameters is avoided.

3. Distributions returned from fitting procedures can be directly passed to, e.g., the density function since fitted distribution objects add information to the distribution object and consequently inherit from the class of the distribution object.

4. Generic method dispatching can be used to provide a uniform interface to, e.g., calculate the expected value `mean(distribution.object)`. Additionally, one can take advantage of generic programming since `R` provides virtual classes and some forms of polymorphism.

See appendix G for several examples and G.2 for particular examples concerning the object-oriented approach.

G Examples

This section provides examples of distribution objects and the object-oriented approach as well as fitting to data and portfolio optimization.

G.1 Initializing distribution objects

This example shows how GH distribution objects can be initialized by either using the $(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$, the $(\lambda, \overline{\alpha}, \mu, \Sigma, \gamma)$ or the $(\lambda, \alpha, \mu, \Delta, \delta, \beta)$-parametrization.

```
> library(ghyp)
> ghyp(lambda = -2, alpha.bar = 0.5, mu = 10, sigma = 5, gamma = 1)

Asymmetric Generalized Hyperbolic Distribution:


Parameters:
   lambda alpha.bar        mu     sigma     gamma
     -2.0       0.5      10.0       5.0       1.0

> ghyp.ad(lambda = -2, alpha = 1, beta = 0.5, mu = 10, delta = 1)

Asymmetric Generalized Hyperbolic Distribution:


Parameters:
lambda  alpha  delta   beta     mu
  -2.0    1.0    1.0    0.5   10.0

> ghyp(lambda = -2, chi = 5, psi = 0.1, mu = 10:11, sigma = diag(5:6),
+     gamma = -1:0)

Asymmetric Generalized Hyperbolic Distribution:


Parameters:
lambda    chi    psi
  -2.0    5.0    0.1


mu:
[1] 10 11


sigma:
     [,1] [,2]
[1,]    5    0
[2,]    0    6


gamma:
[1] -1  0
```

G.2 Object-oriented approach

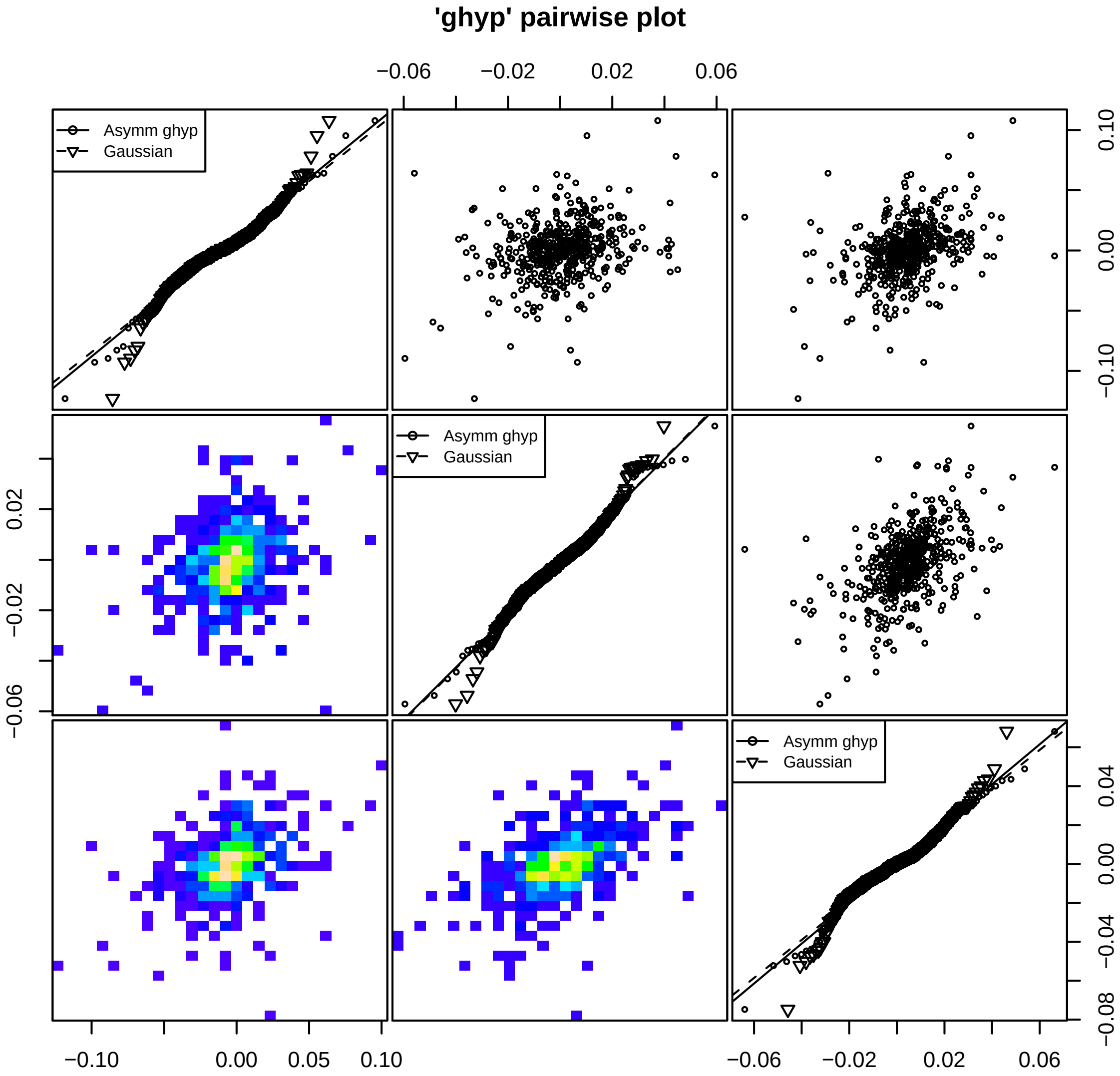
First of all a GH distribution object is initialized and a consistency check takes place. The second command shows how the initialized distribution object is passed to the density function. Then a Student-t distribution is fitted to the daily log-returns of the Novartis stock. The fitted distribution object is passed to the quantile function. Since the fitted distribution object inherits from the distribution object this constitutes no problem. The generic methods *hist*, *mean* and *vcov* are defined for distribution objects inheriting from classes ”ghypuv” and ”ghypbase”, respectively.

```
## Consistency check when initializing a GH distribution object.
## Valid input:
univariate.ghyp.object <- ghyp(lambda = -2, alpha.bar = 0.5,
                               mu = 10, sigma = 5, gamma = 1)

## Passing a distribution object to the density function
dghyp(10:14,univariate.ghyp.object)

## Passing a fitted distribution object to the quantile function
fitted.ghyp.object <- fit.tuv(smi.stocks[,"Novartis"], silent = T)
qghyp(c(0.01,0.05),fitted.ghyp.object)

## Generic method dispatching: the histogram method
hist(fitted.ghyp.object, legend.cex = 0.7)

## Generic programming:
mean(fitted.ghyp.object)     ## fitted.ghyp.object extends "ghypuv"
                             ## which extends "ghypbase"

vcov(univariate.ghyp.object) ## univariate.ghyp.object extends "ghypbase"
```

G.3 Fitting generalized hyperbolic distributions to data

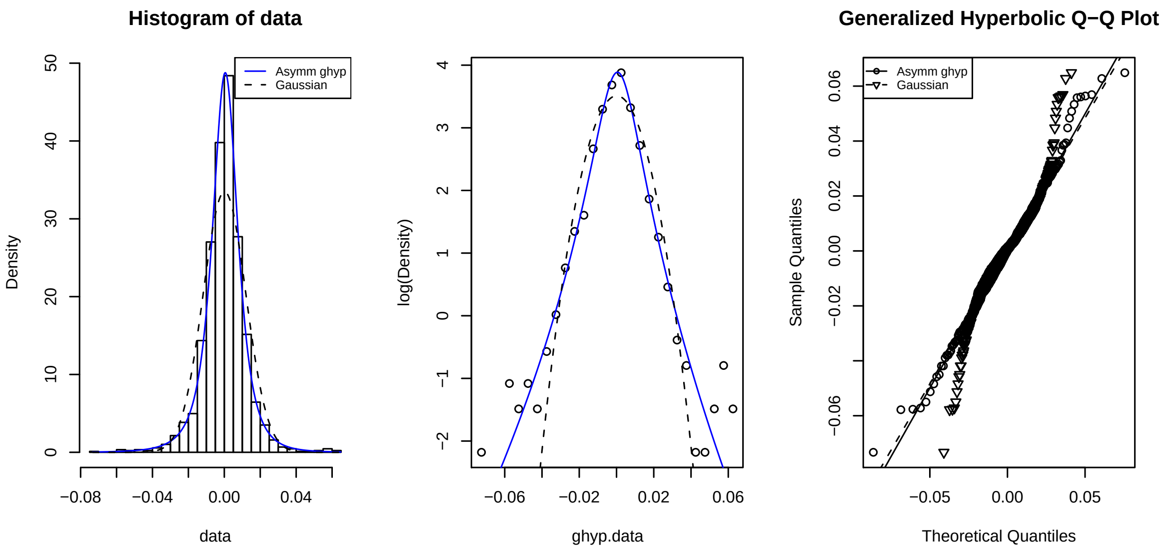
A multivariate GH distribution is fitted to the daily returns of three swiss blue chips: Credit Suisse, Nestle and Novartis. A *pairs* plot is drawn in order to do some graphical diagnostics of the accuracy of the fit.

```
> data(smi.stocks)
> fitted.returns.mv <- fit.ghypmv(data = smi.stocks[1:500, c("CS",
+     "Nestle", "Novartis")], silent = TRUE)
> pairs(fitted.returns.mv, cex = 0.5, legend.cex = 0.8, nbins = 30)
```



In the following part daily log-returns of the SMI are fitted to the GH distribution. Again, some graphical verification is done to check the accuracy of the fit.

```
> fitted.smi.returns <- fit.ghypuv(data = smi.stocks[, c("SMI")],
+     silent = TRUE)
> par(mfrow = c(1, 3))
> hist(fitted.smi.returns, ghyp.col = "blue", legend.cex = 0.7)
> hist(fitted.smi.returns, log.hist = T, nclass = 30, plot.legend = F,
+     ghyp.col = "blue")
> qqghyp(fitted.smi.returns, plot.legend = T, legend.cex = 0.7)
```



References

Abramowitz, M. and I. A. Stegun (1972). *Handbook of mathematical functions* (9. Dover printing ed.). Dover books on mathematics. New York, NY: Dover.

Barndorff-Nielsen, O. E. (1977). Exponentially decreasing distributions for the logarithm of particle size. *Proceedings of the royal Society London A 353*, 401–419.

McNeil, A. J. (2005). Free S-Plus and R library designed to accompany the book quantitative risk management: Concepts, techniques and tools. http://cran.r-project.org/web/packages/QRMlib/index.html.

McNeil, A. J., R. Frey, and P. Embrechts (2005). *Quantitative risk management: Concepts, techniques and tools.* Princeton: Princeton University Press.