Investigating the Generalized t-distribution

A. Verster and D.J, de Waal

Introduction

The t distribution is commonly used to model data with heavy tails, see for example Beirlant *et al*. (2004). In some cases it can be argued that the tail on the t distribution might be too heavy or too light for a specific situation. In such cases the Generalized t distribution might be considered as a more appropriate distribution to model the data. The Generalized t distribution, as we will show, allows for situations where the tails are heavier or lighter than the usual t distribution. In this paper we introduce the Generalized t distribution and compare it to the t distribution by means of a practical application. The real data set considered here is the total rainfall in Bloemfontein, South Africa, for the month February from 1970 to 2011. The data is shown in Figure 1. In Section 1 the t distribution is fitted to the February data and the 95th quantile is estimated from the posterior predictive density. In Section 2 the Generalized t distribution is introduced and fitted to the same data set. Again the 95th quantile is estimated using the posterior predictive density and compared to the quantile obtained with the t distribution.

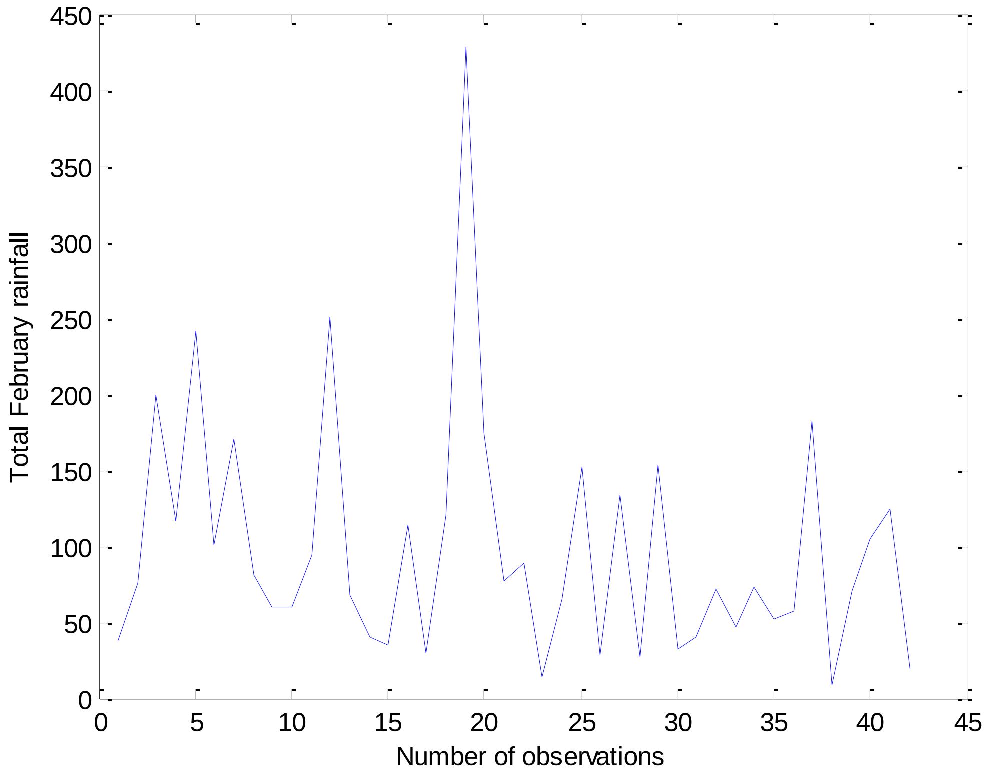

Figure 1: February rainfall for 1970 to 2011

1 Fitting the t distribution

The t distribution with density function, given by the following equation, is fitted to the February rainfall data (X) using the function dfittool in the statistical package Matlab

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\left(\frac{1}{\sigma\sqrt{\pi\nu}}\right)^{\frac{1}{2}}\left(1+\frac{(x-\mu)^2}{\sigma^2\nu}\right)^{-\frac{\nu+1}{2}}, \qquad -\infty \leq x \leq \infty \tag{1}$$

where μ is the location parameter, σ is the scale parameter and ν is the degrees of freedom. The parameters are estimated as follows: $\hat{\mu} = 107.884$, $\hat{\sigma} = 45.0887$ and $\hat{\nu} = 6.77934$. The t distribution with the estimated parameters are plotted on the histogram of the data and shown in Figure 2.

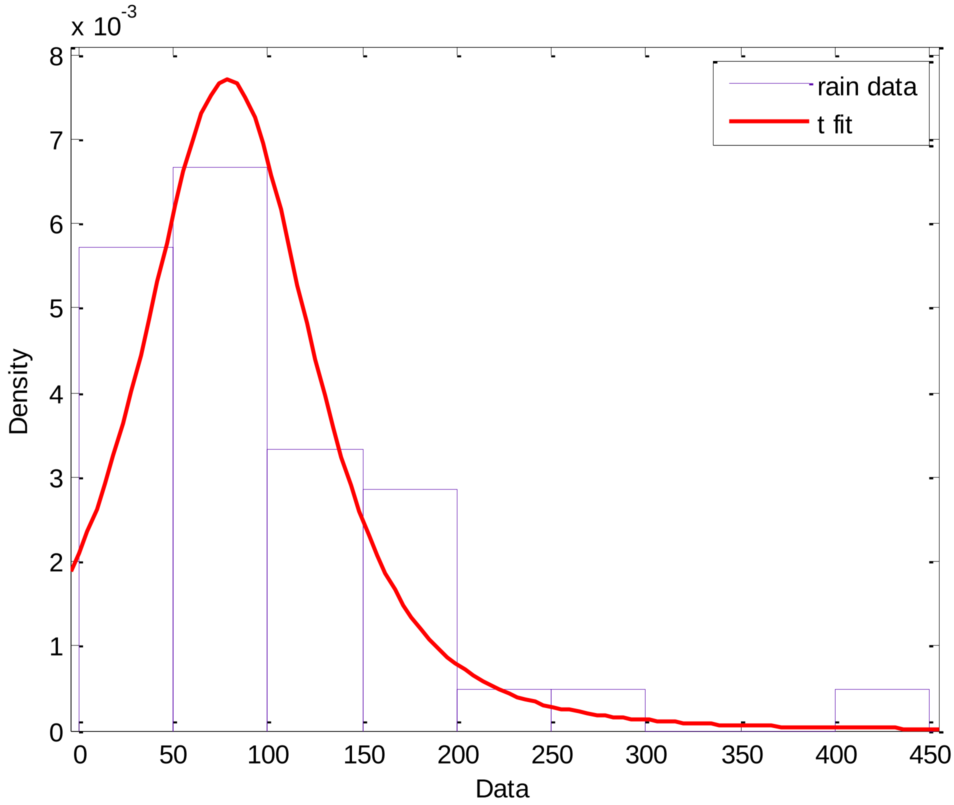

Figure 2: t distribution fitted on the February rainfall

From Figure 2 the t distribution seems to be a reasonable fit to the data although the tail of the distribution might be too heavy or too light. This will be discussed in the next section.

The posterior predictive density of the t distribution can be used to estimate high quantiles. The posterior predictive density of the t distribution for a future observation, X_{n+1}, is derived as follows:

$$U^2 = \frac{(n-1)n}{(n^2-1)}\frac{(X_{n+1}-\overline{x})^2}{s^2} \sim F(1, n-1) \tag{2}$$

where $s^2 = \frac{\sum(x-\overline{x})^2}{n-1}$ and $\overline{x} = \sum_{i=1}^{n} x_i$ for $n = 42$ observations (Geisser, 1993). Therefore $U \sim t_{n-1}(0,1)$ follows a Student's t distribution with mean 0 and variance 1. The 95th quantile is predicted by first calculated U using the Matlab function $1 - tinv(0.95, n-1)$. After U is calculated it is substituted into (2) and the 95th quantile $Q(0.95)x_{n+1}$ is predicted as 206.9705. The posterior predictive density of X_{n+1} for future observations is shown in Figure 3.

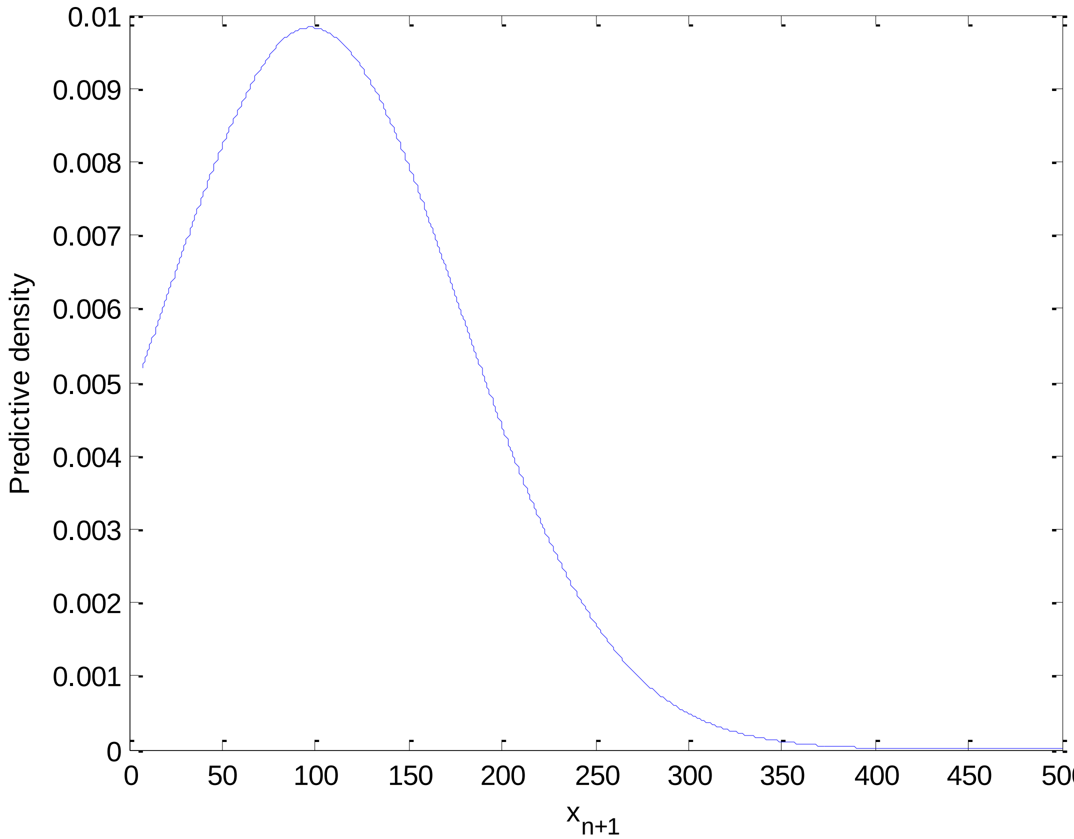

Figure 3: The posterior predictive density of the t distribution for future observations X_{n+1}.

2 Introducing the Generalized t distribution

It might be that the tail of the t distribution is too heavy or too light for the given data set, therefore we introduce the Generalized t (Gt) distribution to investigate this aspect. We start this section with a brief discussion on the Generalized Gamma and Generalized Beta Type II distribution before we introduce the Gt distribution.

Assume that Y is Generalized Gamma distributed then the distribution function of $Y \sim GGAM(k, \mu, \sigma)$ is given in (3) (Beirlant *et al*. 2002)

$$F(x|k, \mu, \sigma) = P(X < x) = \frac{1}{\Gamma(k)} \int_0^{\nu} e^{-u} u^{k-1} du \ , x > 0 \tag{3}$$

where

$$\nu = (x e^{\alpha})^{\frac{1}{\beta}} \tag{4}$$

or

$$\nu = \exp\left(\psi(k) + \frac{\log(x)+\mu}{\sigma} \sqrt{\psi'(k)}\ \right). \tag{5}$$

$\psi(k) = \frac{d}{dk}\log(\Gamma(k))$ and $\psi'(k) = \frac{d}{dk}\psi(k)$ are the digamma and trigamma functions respectively where $-\psi(1) = 0.5772$ is the well known Euler's constant and $\psi'(1) = \frac{\pi^2}{6}$. The parameters μ and σ can be expressed in terms of α and β as $\mu = \alpha - \beta\psi(k)$ and $\sigma = \beta\sqrt{\psi'(k)}$. From (3) $V \sim Gam(k, 1)$ and it is known that $E(\log(V)) = \psi(k)$. The distribution (3) is constructed such that $E(log(Y)) = -\mu$ and $ar(log(Y)) = \sigma^2$. The $GGAM(k, \mu, \sigma)$ can be rewritten as $GGAM(k, a, \beta)$ with parameter β called the tail index and parameter $a = e^{\psi(k)+\mu/\beta} = \mathrm{e}^{\alpha/\beta}$ called the scale parameter. The parameter space $\Omega = \{-\infty < \mu < \infty, \sigma > 0, k > 0\}$ changes to $\Omega = \{a > 0, \beta > 0, k > 0\}$ and V is rewritten as $V = aY^{1/\beta} \sim GAM(k, 1)$ where $\mathrm{a} = e^{\frac{\mu}{\beta}+\psi(k)}$. The density function of Y is then given by

$$f(y|k, a, \beta) = \frac{a^k}{\beta\Gamma(k)} e^{-ay^{1/\beta}} y^{\frac{k}{\beta}-1}, \ y > 0. \tag{6}$$

We will refer to (6) as the GGAM density, $Y \sim GGAM(k, a, \beta)$, where k is the shape parameter, β the tail index and a the scale parameter.

Assume now that $Z_1 \sim GGAM(k_1, a_1, \beta)$ and $Z_2 \sim GGAM(k_2, a_2, \beta)$ where Z_1 and Z_2 are independent, then $W = \left(\frac{a_1}{a_2}\right)^{\beta} \frac{X}{Y}$ is Generalized Beta Type II (GBET2) distributed denoted by $W \sim GBET2(k_1, k_2, \beta)$. Let $U = W^{\frac{1}{\beta}}$, U is then expressed as the quotient of two GGAM distributions and is Beta Type 2 (BET2) distributed denoted by $U \sim BET2(k_1, k_2)$. The probability density of U is given by

$$f(u) = \frac{\Gamma(k_1+k_2)}{\Gamma(k_1)\Gamma(k_2)} u^{k_1-1}(1+u)^{-(k_1+k_2)}, u > 0. \tag{7}$$

The probability density of $W \sim GBET2(k_1, k_2, \beta)$ follows from (7) as

$$f(w) = \frac{\Gamma(k_1+k_2)}{\beta\Gamma(k_1)\Gamma(k_2)} \left(\mathrm{w}^{\frac{1}{\beta}}\right)^{\mathrm{k_1}-\beta} (1+\mathrm{w}^{\frac{1}{\beta}})^{-(\mathrm{k_1}+\mathrm{k_2})}, w > 0 \tag{8}$$

where the Jacobian is $J(u \rightarrow w) = \frac{1}{\beta} u^{\frac{1}{\beta}-1} = \frac{1}{\beta} w^{1-\beta}$.

It can now be argued that $T = \left(\frac{a_1}{a_2}\right)^{\beta/2} \left|\sqrt{\frac{Z_1}{Z_2}}\right|$ is Generalized |t| distributed denoted by $T \sim Gt(k_2, \beta)$. Further, $S = T^{\frac{2}{\beta}}$ is again the quotient of two GGAM distributions and is therefore Beta Type 2 (BET2) distributed with parameters $k_1 = \frac{1}{2}$ and k_2.

The probability density of $T \sim Gt(k_2, \beta)$ follows as

$$f(t) = \frac{2\Gamma\left(\frac{1}{2}+k_2\right)}{\beta\sqrt{\pi}\Gamma(k_2)} \left(t^{\frac{2}{\beta}}\right)^{\frac{1-\beta}{2}} (1 + t^{\frac{2}{\beta}})^{-\left(\frac{1}{2}+k_2\right)}, t > 0 \qquad (9)$$

where the Jacobian is $J(s \rightarrow t) = \frac{2}{\beta} t^{\frac{2}{\beta}-1}$. t is an absolute standardized variable, $t = \left|\frac{x-\mu}{\sigma}\right|$, where x is the observations, μ is the mean and σ is the standard deviation of the observations.

The t distribution is a special case of the Gt with mean 0 and $2k_2$ degrees of freedom if $\beta = 1$. Figure 4 shows the Gt density with $k_2 = 2$ and different values of β from 0.2 to 2 with increments of 0.2.

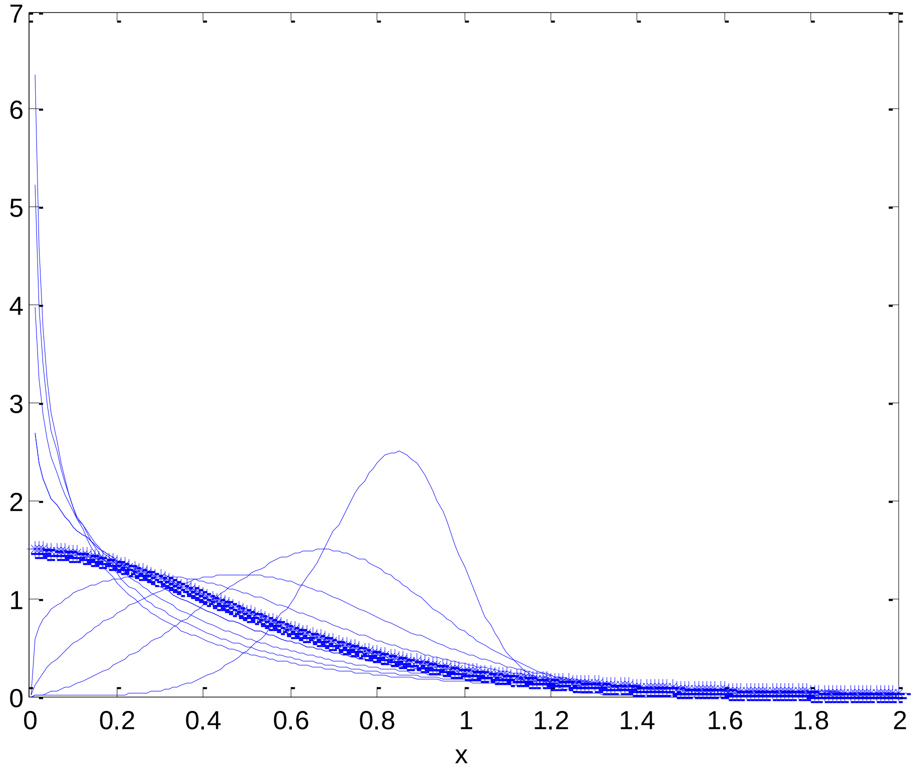

Figure 4: The Gt densities for $k_2 = 2$ and $\beta = 0.2{:}\,0.2{:}\,2$.

The density of |t| where $\beta = 1$ can be recognized from Figure 4 printed in a thick line with (*) as the positive part of the symmetric density of the ordinary t_{2k_2}. The tail of the densities become heavier as β increases. We therefore classify the Gt distribution with tails heavier than the t distribution (HTT) when $\beta > 1$ or less than the t distribution (LTT) if $\beta < 1$.

A Bayesian approach is considered for estimating the two parameters of the Gt jointly. The joint likelihood of k_2 and β is given by

$$like(k_2, \beta|\underline{t}) = \left(\frac{2\Gamma\left(\frac{1}{2}+k_2\right)}{\beta\sqrt{\pi}\Gamma(k_2)}\right)^n \prod_{i=1}^n \left(t_i^{\frac{2}{\beta}}\right)^{\frac{1-\beta}{2}} \left(1 + t_i^{\frac{2}{\beta}}\right)^{-\left(\frac{1}{2}+k_2\right)}. \tag{10}$$

The log of the MDI prior, $log\,\pi(k_2, \beta) \propto E(\log(f(T)))$, on k_2 and β under the Gt distribution is given by

$$\begin{aligned} &log\pi(k_2, \beta) = \\ &log\left(\Gamma\left(\frac{1}{2} + k_2\right)\right) - log(\Gamma(k_2)) - log(\beta) \\ &+ \left(\frac{1-\beta}{2}\right) E\left(log\left(T^{\frac{2}{\beta}}\right)\right) - \left(\frac{1}{2} + k_2\right) E\left(log\left(1 + T^{\frac{2}{\beta}}\right)\right),\ k_2 > 0, \beta > 0. \end{aligned} \tag{11}$$

Equation (11) follows since $S = T^{\frac{2}{\beta}}$ is distributed Beta2$\left(\frac{1}{2}, k_2\right)$, $E(log(S)) = \psi\left(\frac{1}{2}\right) - \psi(k_2)$ and $E(log(1 + S)) = \psi\left(\frac{1}{2} + k_2\right) - \psi(k_2)$. Also $log(a_1) = \psi\left(\frac{1}{2}\right)$ and $log(a_2) = \psi(k_2)$. The prior becomes

$$\begin{aligned} &\pi(k_2, \beta) \\ &= \frac{\Gamma\left(\frac{1}{2}+k_2\right)}{\beta\Gamma(k_2)} exp\{\left(\frac{1-\beta}{2}\right)\left((\psi\left(\frac{1}{2}\right) - \psi(k_2)\right) - \left(\frac{1}{2} + k_2\right)(\psi(\frac{1}{2} + k_2) - \psi(k_2))\}. \end{aligned} \tag{12}$$

The joint posterior distribution of k_2 and β is then given as

$$\pi(k_2, \beta|\underline{t}) \propto like(k_2, \beta|\underline{t})\pi(k_2, \beta). \tag{13}$$

To simulate a set of (k_2, β)'s from the posterior (13), we make use of the Gibbs sampling method by simulating alternatively k_2 from its conditional density function given β fix, then simulating β from its conditional density given the selected k_2. This process is repeated a large number of times. Figure 5 shows a scatter plot of the simulated k against β for the February rainfall data. The means of the simulated $k's$ and $\beta's$ are calculated as 1.3776 and 0.7784 respectively and is considered as the estimates of k and β. In Figure 6 the density function of the Gt with the estimated parameters are plotted on the histogram of the standardized February rainfall data. Note that $\beta < 1$ and therefore indicating a lighter tail than the t distribution, the t distribution is therefore not appropriate for estimating high February rainfall tail quantiles, because the tail of the t distribution is too heavy.

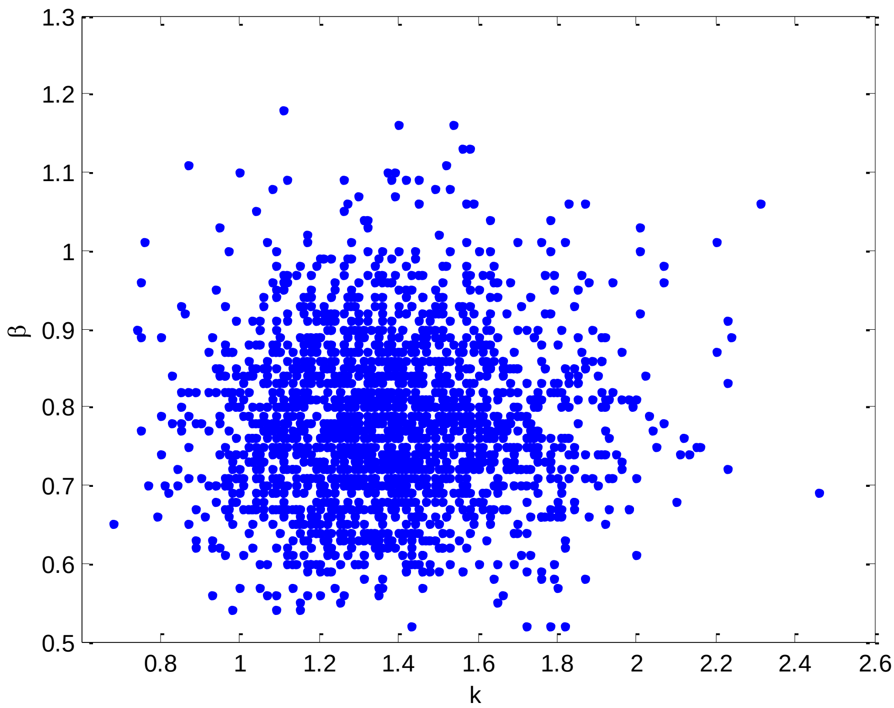

Figure 5: Plot of 2000 simulated (k_2, β) values simulated through the Gibbs sampler

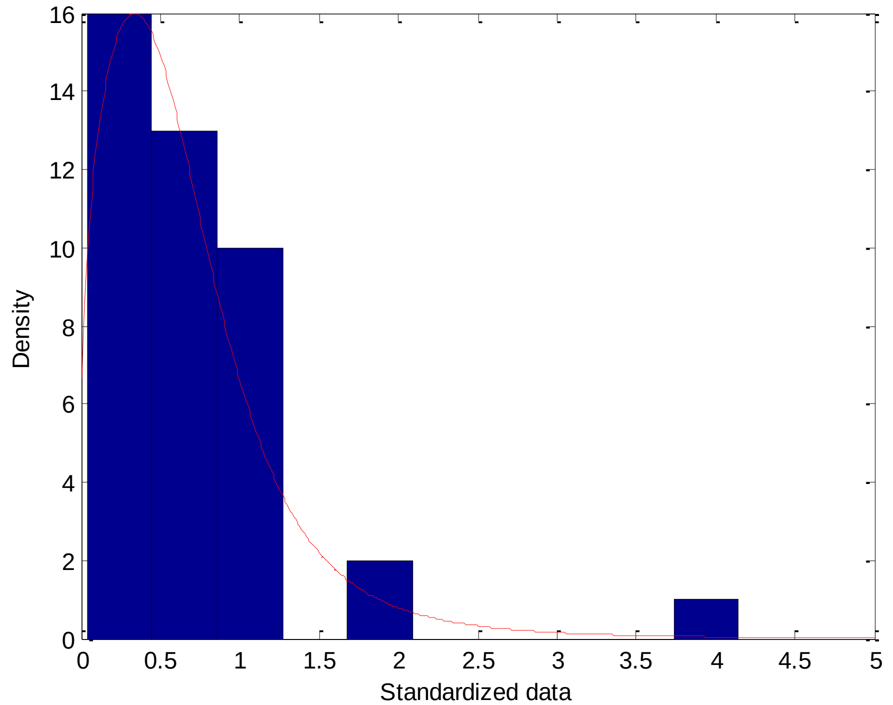

Figure 6: The Gt with estimated parameters plotted on the histogram of the standardized data.

The posterior predictive density of a future observation T_{n+1}, given the data is given as follows:

$$p(t_{n+1}|\underline{t}) = E_{k_2,\beta|\underline{t}} f(t_{n+1}|k_2,\beta). \tag{14}$$

Since (14) cannot be solved explicitly the posterior predictive density can be simulated by taking the mean of the densities at t_{n+1} for a large number of k_2 and β values simulated from the posterior. Equation (14) simplifies then to

$$\hat{p}(t_{n+1}|\underline{t}) = \frac{1}{m}\sum_{j=1}^{m} f(t_{n+1}|k_{2_j},\beta_j) \tag{15}$$

where *m* is the large number of simulated k_2 and β values. The posterior predictive density of the Gt is shown in Figure 7.

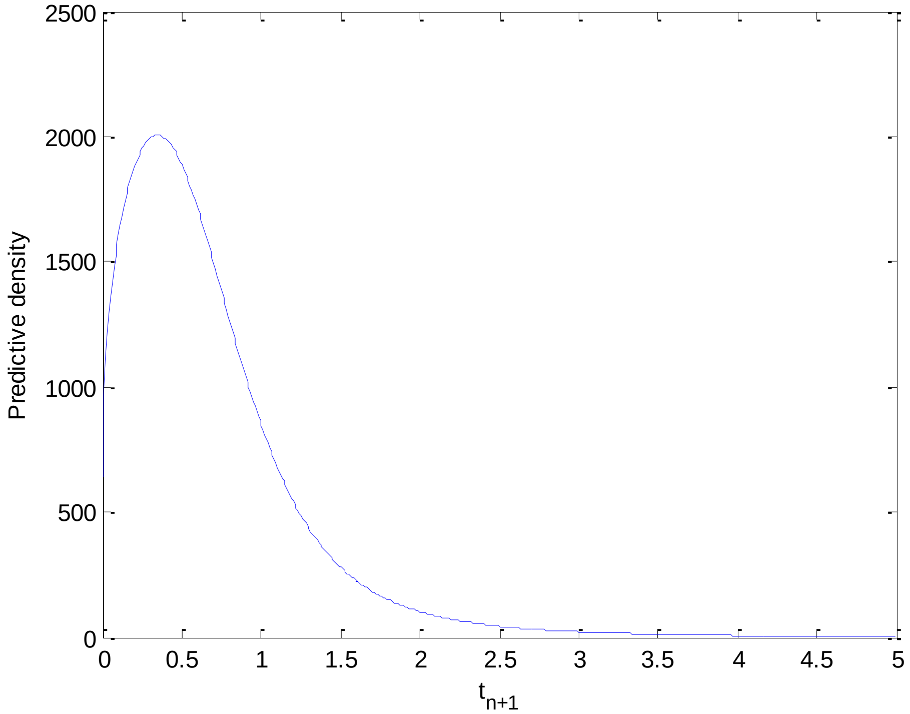

Figure 6: The posterior predictive density of the Gt distribution for future observations T_{n+1}.

A posterior predictive quantile function (16) can be used to predict a high pth quantile

$$Q_p(\underline{t}) = E_{k_2,\beta|\underline{t}} Q_p(t_{n+1}|k_2,\beta). \tag{16}$$

Equation (16) is solved by doing a similar simulation where a large number of k_2 and β values are simulated from the posterior and substituted into (17)

$$\hat{Q}_p = \frac{1}{m}\sum_{j=1}^{m} Q_p(t_{n+1}|k_{2_j},\beta_j). \tag{17}$$

The 95th quantile can easily be estimated by Matlab using the function $q_{Beta} = betainv(0.95, \frac{1}{2}, \underline{k}_2)$. q_{Beta} is then transformed to $q_{BetT2} = \frac{q_{Beta}}{1-q_{Beta}}$ and finally $q_{Gt} = q_{BetaT2}{}^{\frac{\beta}{2}}$=1.7913. Since T is a standardized variable the quantile must be rescaled as follows $q_{Gt}\sigma + \mu = 240.8594$. The 95th quantile obtained is smaller than the quantile predicted with the t distribution. This was expected since $\beta < 1$. Other predicted tail quantiles are shown in Table 1.

Table 1: Comparing predicted tail quantile for the Gt and the t distributions.

	Gt distribution	t distribution
p = 0.975	279.2683	285.7616
p = 0.95	240.8594	260.9705
p = 0.9	208.4054	233.8001

Conclusion

The Gt distribution can be used to model data with a heavier or lighter tail than the t-distribution as illustrated through the February rainfall data. $\beta < 1$ indicates a lighter tail than the t-distribution and $\beta > 1$ indicates a heavier tail than the t-distribution. The two parameters of the Gt can be estimated through Gibbs sampling when considering a Bayesian approach. The posterior predictive density can be estimated as well as posterior predicted quantiles.

Reference

Beirlant J, Goedgebeur Y, Segers J, Teugels J (2004) Statistics of extremes. Theory and applications. Wiley, Chichester.

Beirlant J., de Waal D.J. and Teugels J.L. (2002): The Generalised Burr-Gamma family of distributionswith applications in Extreme value Analysis. *Limit theorems in Probability and Statsitics I (I. Berkes, E. Csaki, M. Csorgo, eds.). Budapest, pp 113-132.*

Geisser S (1993) Predictive Inference: An Introduction. Chapman & Hall.