

On the Multivariate t Distribution

Michael Roth

Division of Automatic Control

E-mail: roth@isy.liu.se

17th April 2013

Report no.: LiTH-ISY-R-3059

Address:

Department of Electrical Engineering

Linköpings universitet

SE-581 83 Linköping, Sweden

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

AUTOMATIC CONTROL
REGLERTEKNIK
LINKÖPINGS UNIVERSITET



Abstract

This technical report summarizes a number of results for the multivariate t distribution which can exhibit heavier tails than the Gaussian distribution. It is shown how t random variables can be generated, the probability density function (pdf) is derived, and marginal and conditional densities of partitioned t random vectors are presented. Moreover, a brief comparison with the multivariate Gaussian distribution is provided. The derivations of several results are given in an extensive appendix.

Keywords: Student's t distribution, heavy tails, Gaussian distribution.

On the Multivariate t Distribution

Michael Roth

April 17, 2013

CONTENTS

1	Introduction	1
2	Representation and moments	2
3	Probability density function	3
4	Affine transformations and marginal densities	3
5	Conditional density	4
6	Comparison with the Gaussian distribution	4
6.1	Gaussian distribution as limit case	5
6.2	Tail behavior	5
6.3	Dependence and correlation	7
6.4	Probability regions and distributions of quadratic forms	7
6.5	Parameter estimation from data	8
A	Probability theory background and detailed derivations	8
A.1	Remarks on notation	8
A.2	The transformation theorem	9
A.3	Utilized probability distributions	9
A.4	Derivation of the t pdf	11
A.5	Useful matrix computations	14
A.6	Derivation of the conditional t pdf	15
A.7	Mathematica code	17

1 INTRODUCTION

This technical report summarizes a number of results for the multivariate t distribution [2, 3, 7] which can exhibit heavier tails than the Gaussian distribution. It is shown how t random variables can be generated, the probability density function (pdf) is derived, and marginal and conditional densities of partitioned t random vectors are presented. Moreover, a brief comparison with the multivariate Gaussian distribution is provided. The derivations of several results are given in an extensive appendix. The reader is assumed to have a working knowledge of probability theory and the Gaussian distribution in particular. Nevertheless, important techniques and utilized probability distributions are briefly covered in the appendix.

Any t random variable is described by its degrees of freedom parameter ν (scalar and positive), a mean μ (dimension d), and a symmetric matrix parameter Σ (dimension $d \times d$). In general, Σ is not the covariance of X . In case Σ is positive definite, we can write the pdf of a t random vector X (dimension d) as

$$\text{St}(x; \mu, \Sigma, \nu) = \frac{\Gamma(\frac{\nu+d}{2})}{\Gamma(\frac{\nu}{2})} \frac{1}{(\nu\pi)^{\frac{d}{2}}} \frac{1}{\sqrt{\det(\Sigma)}} \left(1 + \frac{1}{\nu}(x - \mu)^T \Sigma^{-1} (x - \mu) \right)^{-\frac{\nu+d}{2}}. \quad (1.1)$$

Equation (1.1) is presented in, for instance, [2, 3, 7].

2 REPRESENTATION AND MOMENTS

Multivariate t random variables can be generated in a number of ways [7]. We here combine Gamma and Gaussian random variables and show some properties of the resulting quantity without specifying its distribution. The fact that the derived random variable admits indeed a t distribution with pdf (1.1) will be postponed to Section 3. The use of the more general Gamma distribution instead of a chi-squared distribution [3] shall turn out to be beneficial in deriving more involved results in Section 5.

Let $V > 0$ be a random variable that admits the Gamma distribution $\text{Gam}(\alpha, \beta)$ with shape parameter $\alpha > 0$, and rate (or inverse scale) parameter $\beta > 0$. The related pdf is given in (A.3). Furthermore, consider the d -dimensional, zero mean, random variable Y with Gaussian distribution $\mathcal{N}(0, \Sigma)$. Σ is the positive definite covariance matrix of Y . The related pdf is shown in (A.2). Given that V and Y are independent of one another, the following random variable is generated

$$X = \mu + \frac{1}{\sqrt{V}}Y. \quad (2.1)$$

with a $d \times 1$ vector μ . Realizations of X are scattered around μ . The spread is intuitively influenced by the pdf of V , and it can be easily seen that for $V = v$ we have $X|v \sim \mathcal{N}(\mu, \frac{1}{v}\Sigma)$, a Gaussian with scaled covariance matrix.

In the limit $\alpha = \beta \rightarrow \infty$, the Gamma pdf (A.3) approaches a Dirac impulse centered at one, which results in $V = 1$ with probability 1. It follows (without further specification of the distribution) that X in (2.1) then reduces to a Gaussian random variable.

With mere knowledge of (2.1) we can derive moments of X . Trivially taking the expectation reveals $E(X) = \mu$. For higher moments, it is convenient work with $W = \frac{1}{\sqrt{V}}$ which admits an inverse Gamma distribution $\text{IGam}(\alpha, \beta)$. The random variable W has the pdf (A.5) and the expected value $E(W) = \frac{\beta}{\alpha-1}$ for $\alpha > 1$. The covariance computations

$$\begin{aligned} \text{cov}(X) &= E(X - E(X))(X - E(X))^T \\ &= E(WY Y^T) \\ &= \frac{\beta}{\alpha-1} \Sigma, \quad \alpha > 1 \end{aligned} \quad (2.2)$$

reveal that Σ is in general not a covariance matrix, and finite only if $\alpha > 1$. Higher moments can in principle be computed in a similar manner, again with conditions for existence.

3 PROBABILITY DENSITY FUNCTION

The probability density function of X in (2.1) can be derived using a transformation theorem that can be found in many textbooks on multivariate statistics, for instance [4] or [9]. The theorem is presented, in a basic version, in Appendix A.2. We immediately present the main result but show detailed calculations in Appendix A.4.

Let $V \sim \text{Gam}(\alpha, \beta)$ and $Y \sim \mathcal{N}(0, \Sigma)$. Then $X = \mu + \frac{1}{\sqrt{V}}Y$ admits a t distribution with pdf

$$p(x) = \frac{\Gamma(\alpha + \frac{d}{2})}{\Gamma(\alpha)} \frac{1}{(2\beta\pi)^{\frac{d}{2}}} \frac{1}{\sqrt{\det(\Sigma)}} \left(1 + \frac{\Delta^2}{2\beta}\right)^{-\frac{d}{2} - \alpha} \quad (3.1)$$

$$= \text{St}(x; \mu, \frac{\beta}{\alpha}\Sigma, 2\alpha) \quad (3.2)$$

where $\Delta^2 = (x - \mu)^T \Sigma^{-1} (x - \mu)$.

Obviously, (3.1) is a t pdf in disguise. For $\alpha = \beta = \frac{\nu}{2}$, (3.2) turns into the familiar (1.1). Intermediate steps to obtain (3.1) and (3.2) include introduction of an auxiliary variable and subsequent marginalization. Intermediate results reveal that β and Σ only appear as product $\beta\Sigma$. Consequently, β can be set to any positive number without altering (3.2), if the change is compensated for by altering Σ . This relates to the fact that t random variables can be constructed using a wide range of rules [7]. The parametrization using α and β is used to derive the conditional pdf of a partitioned t random variable in Appendix A.6.

4 AFFINE TRANSFORMATIONS AND MARGINAL DENSITIES

The following property is common to both t and Gaussian distribution. Consider a $d_z \times d_x$ matrix A and a d_z -dimensional vector b . Let X be a t random variable with pdf (1.1). Then, the random variable $Z = AX + b$ has the pdf

$$p(z) = \text{St}(z; A\mu + b, A\Sigma A^T, \nu). \quad (4.1)$$

The result follows from (2.1) and the formulas for affine transformations of Gaussian random variables [2]. Here, A must be such that $A\Sigma A^T$ is invertible for the pdf to exist. The degrees of freedom parameter ν is not altered by affine transformations.

There is an immediate consequence for partitioned random variables $X^T = [X_1^T \ X_2^T]$ which are composed of X_1 and X_2 , with respective dimensions d_1 and d_2 . From the joint pdf

$$p(x) = p(x_1, x_2) = \text{St}\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}; \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^T & \Sigma_{22} \end{bmatrix}, \nu\right), \quad (4.2)$$

the marginal pdf of either X_1 or X_2 is easily obtained by applying a linear transformation. Choosing $A = \begin{bmatrix} 0 & I \end{bmatrix}$, for instance, yields the marginal density of X_2

$$p(x_2) = \text{St}(x_2; \mu_2, \Sigma_{22}, \nu). \quad (4.3)$$

5 CONDITIONAL DENSITY

The main result of this section is that X_1 given $X_2 = x_2$ admits another t distribution if X_1 and X_2 are jointly t distributed with pdf (4.2). The derivations are straightforward but somewhat involved, and therefore provided in Appendix A.6. The conditional density is given by

$$p(x_1|x_2) = \frac{p(x_1, x_2)}{p(x_2)} = \text{St}(x_1; \mu_{1|2}, \Sigma_{1|2}, \nu_{1|2}), \quad (5.1)$$

where

$$\nu_{1|2} = \nu + d_2, \quad (5.2)$$

$$\mu_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \quad (5.3)$$

$$\Sigma_{1|2} = \frac{\nu + (x_2 - \mu_2)^T \Sigma_{22}^{-1} (x_2 - \mu_2)}{\nu + d_2} (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T). \quad (5.4)$$

These expressions are in accordance with the formulas given in [6] (without derivation) and [3] (derivation is given as exercise). The book [7] misleadingly states that the conditional distribution is in general not t . However, a “generalized” t distribution is presented in Chapter 5 of [7], which provides the property that the conditional distribution is again t . The derivations provided here show that there is no need to consider a generalization of the t for such a closedness property.

Obviously, conditioning increases the degrees of freedom. The expression for $\mu_{1|2}$ is the same as the conditional mean in the Gaussian case [2]. The matrix parameter $\Sigma_{1|2}$, in contrast to the Gaussian case, is also influenced by the realization x_2 . By letting ν go to infinity, we can recover the Gaussian conditional covariance: $\lim_{\nu \rightarrow \infty} \Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T$.

6 COMPARISON WITH THE GAUSSIAN DISTRIBUTION

The Gaussian distribution appears to be the most popular distribution among continuous multivariate unimodal distributions. Motivation for this is given by the many convenient properties: simple marginalization, conditioning, parameter estimation from data. We here compare the t distribution with the Gaussian to assess whether some of the known Gaussian results generalize. After all, the t distribution can be interpreted as generalization of the Gaussian (or conversely, the Gaussian as special case of the t distribution).

6.1 Gaussian distribution as limit case

It is a well known result [2] that, as the degrees of freedom ν tend to infinity, the t distribution converges to a Gaussian distribution. This fact can be established, for instance, using (2.1) and the properties of the Gamma distribution. Another encounter with this result is described in Appendix A.4.

6.2 Tail behavior

One reason to employ the t distribution is its ability to generate *outliers*, which facilitates more accurate modeling of many real world scenarios. In contrast to the Gaussian, the t distribution has *heavy tails*. To illustrate this, we consider the 2-dimensional example with parameters

$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1 & -0.4 \\ -0.4 & 1 \end{bmatrix}, \quad \nu = 5. \quad (6.1)$$

Figure 1 contains 25 samples each, drawn from the Gaussian $\mathcal{N}(\mu, \Sigma)$ and the t distribution $\text{St}(\mu, \Sigma, \nu)$. Close to μ , there is no obvious difference between Gaussian and t samples. More t samples are scattered further from the mean, with one outlier being far off. The illustrated phenomenon can be explained by the t distribution having heavier tails: the pdf does not decay as quickly when moving away from the mean.

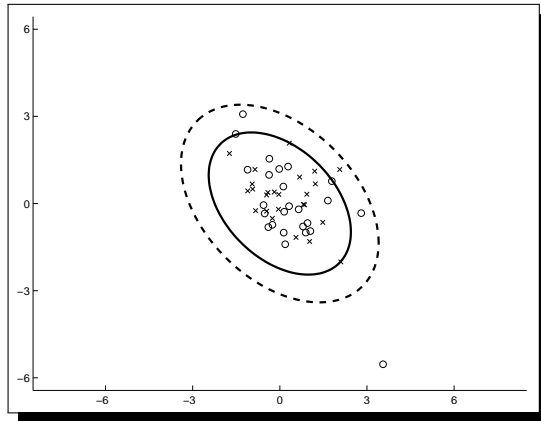


Figure 1: 25 random samples each from the 2-dimensional distributions $\mathcal{N}(\mu, \Sigma)$ (crosses) and $\text{St}(\mu, \Sigma, \nu)$ (circles) with the parameters given in (6.1). The ellipses illustrate regions that contain 95 percent of the probability mass, solid for the Gaussian and dashed for the t distribution.

For the presented two-dimensional case, we can compute both Gaussian and t pdf on a grid and visualize the results. Figure 2 contains heat maps that illustrate the pdf logarithms $\ln \mathcal{N}(x; \mu, \Sigma)$ and $\ln \text{St}(x; \mu, \Sigma, \nu)$, where the logarithmic scale helps highlight the difference in tail behavior. The coloring is the same in Fig. 2a and Fig. 2b, but Fig. 2a contains a large proportion of white ink which corresponds to values below -15 . While the densities appear similar close to μ , the Gaussian drops quickly with increasing distance whereas the t pdf shows a less rapid decay. The connection between

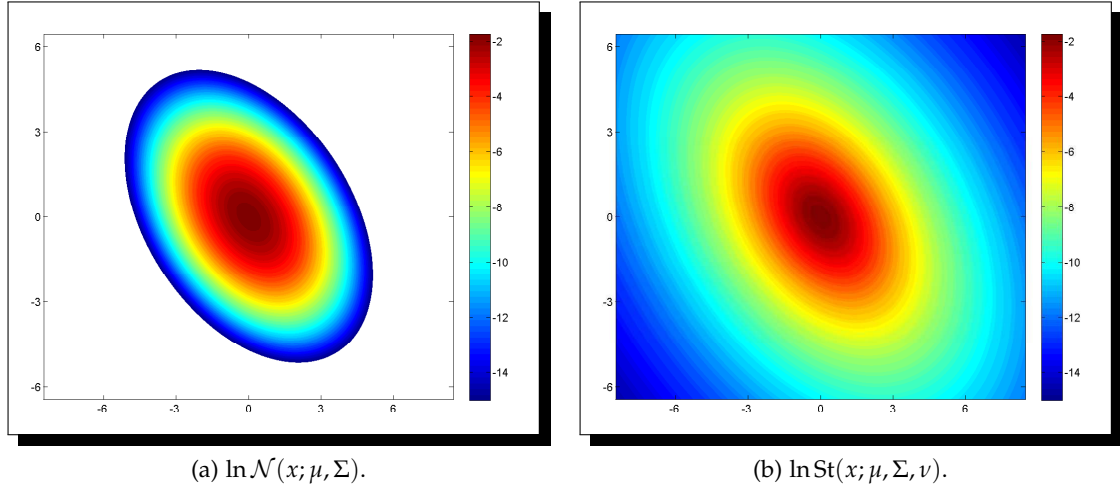


Figure 2: Heat maps of $\ln \mathcal{N}(x; \mu, \Sigma)$ and $\ln \text{St}(x; \mu, \Sigma, \nu)$ for a two-dimensional x . The logarithmic presentation helps highlight the difference in tail behavior. White areas in (a) correspond to values below -15 . The parameters are given by (6.1).

the occurrence of samples that lie on ellipses that are further away from μ and the corresponding values of the pdf is apparent.

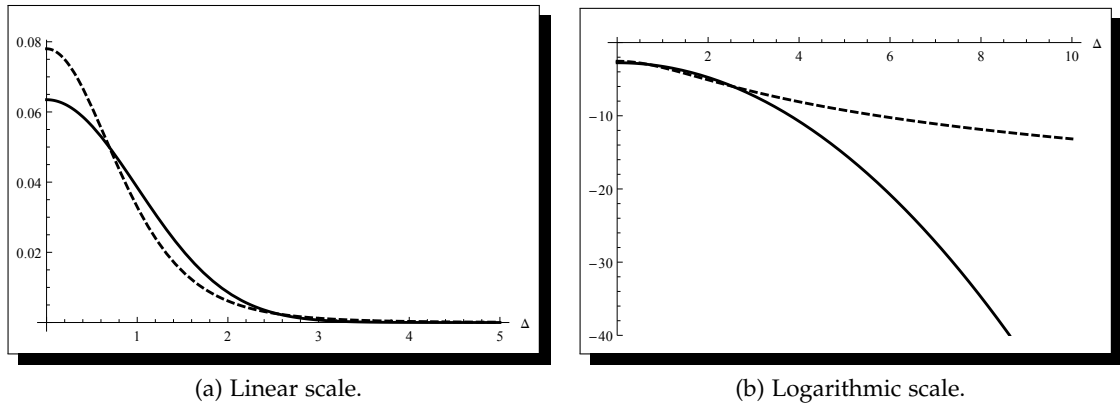


Figure 3: Gaussian pdf (solid) and t pdf (dashed) with $d = 3$ and $\nu = 3$, as functions of the Mahalanobis distance Δ .

The general d -dimensional case is a bit trickier to illustrate. An investigation of the density expression (1.1) and (A.2), however, reveals that both can be written as functions of a scalar quantity $\Delta^2 = (x - \mu)^T \Sigma^{-1} (x - \mu)$. A constant Δ^2 corresponds to an ellipsoidal region (an ellipse for $d = 2$) with constant pdf, as seen in Figure 2. Δ^2 can be interpreted as a squared weighted distance and is widely known as squared Mahalanobis distance [2]. Having understood the distance dependence, it is reasonable to plot each pdf as function of Δ . Furthermore, Σ can be set to unity without loss of generality when comparing $\mathcal{N}(x; \mu, \Sigma)$ and $\text{St}(x; \mu, \Sigma, \nu)$, because of the common factor $\det(\Sigma)^{-1/2}$. Figure 3 contains the Gaussian and t pdf as functions of Δ with $\Sigma = I_d$. The shape of the curves is determined by ν and d only. For $d = 1$, such illustrations coincide with the univariate densities (over the positive ray). Here, we display the functions for $d = 3$ and $\nu = 3$ instead but provide the

mathematica code snippet that can be used to produce interactively manipulable plots for arbitrary d and ν in Section A.7. For the chosen parameters, the t pdf appears more peaked towards the mean (small Δ in Fig. 3a). The logarithmic display (Fig. 3b) again shows the less rapid decay of the t pdf with increasing Δ , commonly known as heavy tails.

6.3 Dependence and correlation

Gaussian random variables that are not correlated are by default independent: the joint pdf is the product of marginal densities. This is not the case for arbitrary other distributions, as we shall now establish for the t distribution. Consider the a partitioned t random vector X with pdf (4.2). For $\Sigma_{12} = 0$ we see from the covariance result (2.2) that X_1 and X_2 are not correlated. The joint pdf, however,

$$p(x_1, x_2) \propto \left(1 + \frac{1}{\nu}(x_1 - \mu_1)^T \Sigma_{11}^{-1}(x_1 - \mu_1) + \frac{1}{\nu}(x_2 - \mu_2)^T \Sigma_{22}^{-1}(x_2 - \mu_2) \right)^{-\frac{d}{2} - \frac{\nu}{2}}$$

does not factor trivially. Consequently, uncorrelated t random variables are in general not independent. Furthermore, the product of two t densities is in general not a t pdf (unless we have $p(x_1|x_2)p(x_2)$).

6.4 Probability regions and distributions of quadratic forms

The following paragraphs make heavy use of the results provided in Appendix A.3.

For a Gaussian random variable $X \sim \mathcal{N}(\mu, \Sigma)$, it is well known [3] that the quadratic form (or squared Mahalanobis distance) $\Delta^2 = (X - \mu)^T \Sigma^{-1}(X - \mu)$ admits a chi-squared distribution $\chi^2(d)$. This can be used to define symmetric regions around μ in which a fraction, say P percent, of the probability mass is concentrated. Such a P percent probability region contains all x for which

$$(x - \mu)^T \Sigma^{-1}(x - \mu) \leq \delta_P^2, \quad \text{with} \quad \int_0^{\delta_P^2} \chi^2(r; d) dr = P/100. \quad (6.2)$$

The required δ_P^2 can be obtained from the inverse cumulative distribution function of $\chi^2(d)$, which is implemented in many software packages.

For $X \sim \text{St}(\mu, \Sigma, \nu)$ similar statements can be made. The involved computations, however, are slightly more involved. Recall from (2.1) that X can be generated using a Gaussian random vector $Y \sim \mathcal{N}(0, \Sigma)$ and the Gamma random variable $V \sim \text{Gam}(\frac{\nu}{2}, \frac{\nu}{2})$. Furthermore, observe that V can be written as $V = W/\nu$ with a chi-squared random variable $W \sim \chi^2(\nu)$, as explained in Appendix A.3. Using

$$X = \mu + \frac{1}{\sqrt{V}}Y = \mu + \frac{1}{\sqrt{W/\nu}}Y, \quad (6.3)$$

we can re-arrange the squared form

$$\Delta^2 = (X - \mu)^T \Sigma^{-1} (X - \mu) = d \frac{Y^T \Sigma^{-1} Y / d}{W / \nu} \quad (6.4)$$

and obtain a ratio of two scaled chi-squared variables, since $Y^T \Sigma^{-1} Y \sim \chi^2(d)$. Knowing that such a ratio admits an F distribution, we have

$$\frac{1}{d} (X - \mu)^T \Sigma^{-1} (X - \mu) \sim \mathcal{F}(d, \nu) \quad (6.5)$$

with pdf given by (A.6). Following the steps that were taken for the Gaussian case, we can derive the symmetric P percent probability region for $X \sim \text{St}(\mu, \Sigma, \nu)$ as the set of all x for which

$$(x - \mu)^T \Sigma^{-1} (x - \mu) \leq dr_P, \quad \text{with} \quad \int_0^{r_P} \mathcal{F}(r; d, \nu) dr = P/100. \quad (6.6)$$

Again, the required r_P can be obtained from an inverse cumulative distribution function, that is available in many software packages.

Figure 1 contains two ellipses, a solid for the Gaussian and a dashed for the t distribution. Both illustrate regions which contain 95 percent of the probability mass, and have been obtained with the techniques explained in this section. It can be seen that the 95 percent ellipse for the t distribution stretches out further than the Gaussian ellipse, which of course relates to the heavy tails of the t distribution.

6.5 Parameter estimation from data

Methods for finding the parameters of a Gaussian random variables from a number of independent and identically distributed realizations are well known. A comprehensive treatment of *maximum likelihood* and *Bayesian* methods is given in [2]. The solutions are available in closed form and comparably simple. The maximum likelihood estimate for the mean of a Gaussian, for instance, coincides with the average over all realizations.

For the t distribution, there are no such simple closed form solutions. However, iterative maximum likelihood schemes based on the *expectation maximization* algorithm are available. The reader is referred to [8] for further details.

A PROBABILITY THEORY BACKGROUND AND DETAILED DERIVATIONS

A.1 Remarks on notation

We make a distinction between the distribution a random variable admits and its pdf. For example, $X \sim \mathcal{N}(\mu, \Sigma)$ denotes that X admits a Gaussian distribution with parameters μ and Σ . In contrast, $p(x) = \mathcal{N}(x; \mu, \Sigma)$ (with the extra argument) denotes the Gaussian probability density function. Furthermore, random variables are capitalized whereas the (dummy) argument of a pdf is lower

case. Where required, the pdf carries an extra index that clarifies the corresponding random variable, for instance $p_X(x)$ in Section A.2.

A.2 The transformation theorem

We here introduce a powerful tool which we plainly refer to as the *transformation theorem*. Under some conditions, it can be used to find the pdf of a random variable that is transformed by a function. We only present the theorem in its most basic form, but refer the reader to [4, 9] for further details. For clarity purposes all probability density functions carry an extra index.

Transformation theorem: Let X be a scalar random variable with pdf $p_X(x)$, that admits values from some set D_X . Let f in $y = f(x)$ be a one-to-one mapping for which an inverse g with $x = g(y)$ exists, with an existing derivative $g'(y)$ that is continuous. Then, the random variable $Y = f(X)$ has the pdf

$$p_Y(y) = \begin{cases} p_X(g(y))|g'(y)|, & y \in D_Y, \\ 0, & \text{otherwise,} \end{cases} \quad (\text{A.1})$$

with $D_Y = \{f(x) : x \in D_X\}$.

The theorem is presented in [9], Ch.3, or [4], Ch.1, including proofs and extensions to cases when $f(x)$ is not a one-to-one mapping (bijection). The occurrence of $|g'(y)|$ is the result of a change of integration variables, and in the multivariate case $|g'(y)|$ is merely replaced by the determinant of the Jacobian matrix of $g(y)$. In case Y is of lower dimension than X , auxiliary variables need to be introduced and subsequently marginalized. An example of the entire procedure is given for the derivation of the multivariate t pdf in Section A.4.

A.3 Utilized probability distributions

We here present some well known probability distributions that are used in this document. All of these are covered in a plethora of text books. We adopt the presentation of [2] and [3].

Gaussian distribution

A d -dimensional random variable X with Gaussian distribution $\mathcal{N}(\mu, \Sigma)$ is characterized by its mean μ and covariance matrix Σ . For positive definite Σ , the pdf is given by

$$\mathcal{N}(x; \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{d}{2}}} \frac{1}{\sqrt{\det(\Sigma)}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right). \quad (\text{A.2})$$

Information about the many convenient properties of the Gaussian distribution can be found in almost any of the references at the end of this document. A particularly extensive account is given in [2]. We shall therefore only briefly mention some aspects. Gaussian random variables that are uncorrelated are also independent. The marginal and conditional distribution of a partitioned Gaussian random vector are again Gaussian, which relates to the Gaussian being its own conjugate

prior in Bayesian statistics [2]. The formulas for marginal and conditional pdf can be used to derive the Kalman filter from a Bayesian perspective [1].

Gamma distribution

A scalar random variable X admits the Gamma distribution $\text{Gam}(\alpha, \beta)$ with shape parameter $\alpha > 0$ and rate (or inverse scale) parameter $\beta > 0$ if it has the pdf [2,3]

$$\text{Gam}(x; \alpha, \beta) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x), & x > 0, \\ 0, & \text{otherwise.} \end{cases} \quad (\text{A.3})$$

$\Gamma(z) = \int_0^\infty e^{-t} t^{z-1} dt$ is the Gamma function. Two parametrizations of the Gamma distribution exist in literature: the rate version that we adopt, and a version that uses the inverse rate (called scale parameter) instead. Hence, care must be taken to have the correct syntax in e.g. software packages.

Using the transformation theorem of Section A.2 and $X \sim \text{Gam}(\alpha, \beta)$, we can derive that $cX \sim \text{Gam}(\alpha, \frac{\beta}{c})$. Mean and variance of $X \sim \text{Gam}(\alpha, \beta)$ are $E(X) = \frac{\alpha}{\beta}$ and $\text{var}(X) = \frac{\alpha}{\beta^2}$, respectively. For $\alpha = \beta \rightarrow \infty$ the support of the pdf collapses and $\text{Gam}(x; \alpha, \beta)$ approaches a Dirac impulse centered at one.

chi-squared distribution

A special case of the Gamma distribution for $\alpha = \frac{\nu}{2}, \beta = \frac{1}{2}$ is the chi-squared distribution $\chi^2(\nu)$ with degrees of freedom $\nu > 0$ and pdf

$$\chi^2(x; \nu) = \text{Gam}(x; \frac{\nu}{2}, \frac{1}{2}) = \begin{cases} \frac{2^{-\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})} x^{\frac{\nu}{2}-1} \exp(-\frac{1}{2}x), & x > 0, \\ 0, & \text{otherwise.} \end{cases} \quad (\text{A.4})$$

From the scaling property of the Gamma distribution we have that $X/\nu \sim \text{Gam}(\frac{\nu}{2}, \frac{\nu}{2})$ for $X \sim \chi^2(\nu)$, which relates to a common generation rule for t random variables [3,7].

Inverse Gamma distribution

In order to compute the moments of t random vectors, it is required to work with the inverse of a Gamma random variable. We shall here present the related pdf, and even sketch the derivation as it is not contained in [2] or [3].

Let $Y \sim \text{Gam}(\alpha, \beta)$. Then its inverse $X = 1/Y$ has an inverse Gamma distribution with parameters $\alpha > 0$ and $\beta > 0$, and pdf

$$\text{IGam}(x; \alpha, \beta) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp(-\beta/x), & x > 0, \\ 0, & \text{otherwise.} \end{cases} \quad (\text{A.5})$$

The result follows from the transformation theorem of Section A.2. Using $g(x) = 1/x$, we obtain

$$\begin{aligned} p_X(x) &= p_Y(g(x))|g'(x)| \\ &= \text{Gam}(x^{-1}; \alpha, \beta)x^{-2} \end{aligned}$$

which gives (A.5). The pdf for one set of parameters is illustrated in Figure 4. The mean of

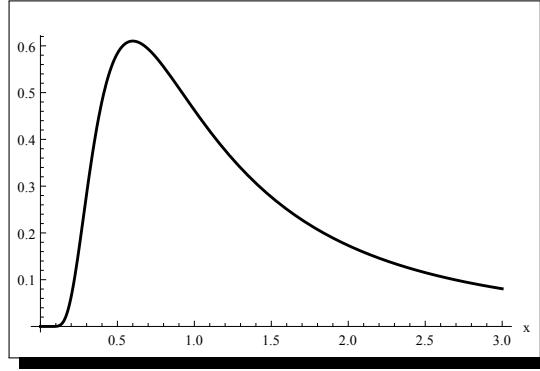


Figure 4: The pdf of the inverse of a Gamma random variable, $\text{IGam}(x; 1.5, 1.5)$.

$X \sim \text{IGam}(\alpha, \beta)$ is $E(X) = \frac{\beta}{\alpha-1}$ and is finite for $\alpha > 1$. Similar conditions are required for higher moments to exist.

F distribution

We conclude this presentation of univariate densities (A.3)-(A.5) with the F distribution [3]. Consider the (scaled) ratio of two chi-squared random variables $Z = \frac{X/a}{Y/b}$ with $X \sim \chi^2(a)$ and $Y \sim \chi^2(b)$. Recall that this is also a ratio of two Gamma random variables. Z then has an F distribution [3] with $a > 0$ and $b > 0$ degrees of freedom, with pdf

$$\mathcal{F}(z; a, b) = \begin{cases} \frac{\Gamma(\frac{a+b}{2})}{\Gamma(\frac{a}{2})\Gamma(\frac{b}{2})} a^{\frac{a}{2}} b^{\frac{b}{2}} \frac{z^{\frac{a}{2}-1}}{(b+az)^{\frac{a+b}{2}}}, & z > 0, \\ 0, & \text{otherwise.} \end{cases} \quad (\text{A.6})$$

An illustration of the pdf for one set of parameters is given in Figure 5. The mean value of $Z \sim \mathcal{F}(a, b)$ is $E(Z) = \frac{b}{b-2}$, and exists for $b > 2$. The mode, i.e. the maximum of $\mathcal{F}(z; a, b)$, is located at $\frac{(a-2)b}{a(b+2)}$, again for $b > 2$. The median can be computed but the expression is not as conveniently written down as mean and mode.

A.4 *Derivation of the t pdf*

In this appendix we make use of the transformation theorem of Appendix A.2 to derive the pdf of the random vector in (2.1). This technique involves the introduction of an auxiliary variable, a change of integration variables, and subsequent marginalization of the auxiliary variable. In the following calculations, we merely require (2.1) and the joint pdf of V and Y .

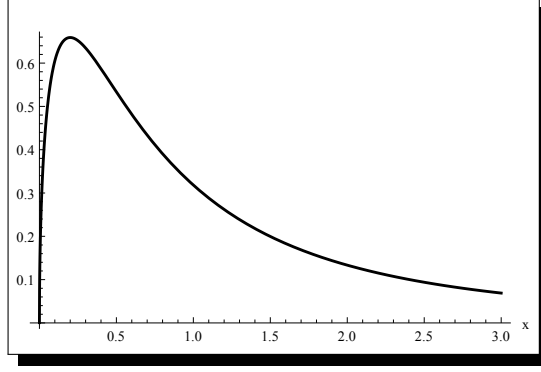


Figure 5: The pdf of a scaled ratio of chi-squared random variables, $\mathcal{F}(x; 3, 3)$.

V and Y are independent, and hence their joint density is the product of individual pdfs

$$p_{V,Y}(v, y) = p_V(v)p_Y(y).$$

We now introduce the pair U and X , which are related to V and Y through

$$X = \mu + \frac{1}{\sqrt{V}}Y; \quad U = V.$$

Here, it is X that we are actually interested in. U is an auxiliary variable that we have to introduce in order to get a one to one relation between (U, X) and (V, Y) :

$$Y = \sqrt{U}(X - \mu); \quad V = U.$$

According to the transformation theorem of Appendix A.2, the density of (U, X) is given by

$$p_{U,X}(u, x) = p_V(u)p_Y(\sqrt{u}(x - \mu)) |\det(J)| \quad (\text{A.7})$$

with J as the Jacobian of the variable change relation

$$J = \begin{bmatrix} \partial y / \partial x^T & \partial y / \partial u \\ \partial v / \partial x^T & \partial v / \partial u \end{bmatrix} = \begin{bmatrix} \sqrt{u}I_d & \frac{1}{2\sqrt{u}}(x - \mu) \\ 0 & 1 \end{bmatrix}.$$

With J being upper triangular, the determinant simplifies to $\det(J) = \det(\sqrt{u}I_d) = u^{\frac{d}{2}}$. We next write down each factor of the joint density (A.7)

$$\begin{aligned} p_Y(\sqrt{u}(x - \mu)) &= \frac{1}{(2\pi)^{\frac{d}{2}}} \frac{1}{\sqrt{\det(\Sigma)}} \exp\left(-\frac{1}{2}\sqrt{u}(x - \mu)^T \Sigma^{-1} \sqrt{u}(x - \mu)\right) \\ &= u^{-\frac{d}{2}} \mathcal{N}(x; \mu, \Sigma/u), \\ p_V(u) &= \text{Gam}(u; \alpha, \beta), \end{aligned}$$

and combine them according to (A.7) to obtain the joint density

$$p_{u,x}(u, x) = \text{Gam}(u; \alpha, \beta) \mathcal{N}(x; \mu, \Sigma/u).$$

From the above expression we can now integrate out u to obtain the sought density

$$p_X(x) = \int_0^\infty \text{Gam}(u; \alpha, \beta) \mathcal{N}(x; \mu, \Sigma/u) du. \quad (\text{A.8})$$

The alert reader might at this stage recognize the integral representation of the t pdf [2] for the case $\alpha = \beta = \frac{\nu}{2}$. As we proceed further and compute the integral, the connections become more evident. Also, we see that if the Gamma part reduces to a Dirac impulse $\delta(u - 1)$ for $\alpha = \beta \rightarrow \infty$, the result of the integration is a Gaussian density.

To simplify the upcoming calculations we introduce the squared Mahalanobis distance [2] $\Delta^2 = (x - \mu)^T \Sigma^{-1} (x - \mu)$ and proceed:

$$\begin{aligned} p_X(x) &= \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} u^{\alpha-1} \exp(-\beta u) \frac{1}{(2\pi)^{\frac{d}{2}}} \frac{1}{\sqrt{\det(\Sigma/u)}} \exp\left(-\frac{u}{2} \Delta^2\right) du \\ &\propto \int_0^\infty u^{\frac{d}{2} + \alpha - 1} \exp\left(-u\left(\beta + \frac{\Delta^2}{2}\right)\right) du \\ &= \left(\beta + \frac{\Delta^2}{2}\right)^{-\frac{d}{2} - \alpha} \Gamma\left(\alpha + \frac{d}{2}\right). \end{aligned}$$

Arranging terms and re-introducing the constant factor eventually yields the result of 3.1:

$$\begin{aligned} p_X(x) &= \frac{\Gamma\left(\alpha + \frac{d}{2}\right)}{\Gamma(\alpha)} \frac{\beta^\alpha}{(2\pi)^{\frac{d}{2}}} \frac{1}{\sqrt{\det(\Sigma)}} \left(\beta + \frac{\Delta^2}{2}\right)^{-\frac{d}{2} - \alpha} \\ &= \frac{\Gamma\left(\alpha + \frac{d}{2}\right)}{\Gamma(\alpha)} \frac{1}{(2\beta\pi)^{\frac{d}{2}}} \frac{1}{\sqrt{\det(\Sigma)}} \left(1 + \frac{\Delta^2}{2\beta}\right)^{-\frac{d}{2} - \alpha}. \end{aligned} \quad (\text{A.9})$$

Although not visible at first sight, the parameters β and Σ are related. To show this we investigate the parts of (A.9) where β and Σ occur,

$$\begin{aligned} \frac{\Delta^2}{\beta} &= (x - \mu)^T (\beta \Sigma)^{-1} (x - \mu), \\ \sqrt{\det(\Sigma)} \beta^{\frac{d}{2}} &= \sqrt{\det(\beta \Sigma)}, \end{aligned} \quad (\text{A.10})$$

and see that the product of both determines the density. This means that any random variable generated using (2.1) with $V \sim \text{Gam}\left(\alpha, \frac{\beta}{c}\right)$ and $Y \sim \mathcal{N}(y; 0, c\Sigma)$ admits the same distribution for any $c > 0$. Consequently we can choose $c = \frac{\beta}{\alpha}$ and thereby have $V \sim \text{Gam}(\alpha, \alpha)$ without loss of generality. More specifically, we can replace $\alpha = \beta = \frac{\nu}{2}$ in (A.9) and recover the t pdf [2, 3, 7] of (1.1).

A.5 Useful matrix computations

We here present results that are used in the derivation of the conditional pdf of a partitioned t random vector, as presented in Section A.6. However, similar expressions show up when working with the Gaussian distribution [2].

Consider the a partitioned random vector X that admits the pdf (1.1) with

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^T & \Sigma_{22} \end{bmatrix}, \quad (\text{A.11})$$

where X_1 and X_2 have dimension d_1 and d_2 , respectively. Mean and matrix parameter are composed of blocks of appropriate dimensions. Sometimes it is more convenient to work with the inverse matrix parameter $\Sigma^{-1} = \Lambda$, another block matrix:

$$\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^T & \Sigma_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{12}^T & \Lambda_{22} \end{bmatrix}. \quad (\text{A.12})$$

Using formulas for block matrix inversion [5] we can find the expressions

$$\Lambda_{11} = (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T)^{-1}, \quad (\text{A.13})$$

$$\begin{aligned} \Lambda_{12} &= -(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T)^{-1}\Sigma_{12}\Sigma_{22}^{-1} \\ &= -\Lambda_{11}\Sigma_{12}\Sigma_{22}^{-1}, \end{aligned} \quad (\text{A.14})$$

$$\Sigma_{22} = (\Lambda_{22} - \Lambda_{12}^T\Lambda_{11}^{-1}\Lambda_{12})^{-1}. \quad (\text{A.15})$$

Furthermore, we have for the determinant of Σ

$$\begin{aligned} \det\left(\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^T & \Sigma_{22} \end{bmatrix}\right) &= \det(\Sigma_{22}) \det(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T) \\ &= \frac{\det(\Sigma_{22})}{\det(\Lambda_{11})}. \end{aligned} \quad (\text{A.16})$$

We next investigate the quadratic form in which x enters the t pdf (1.1). Following [2] we term Δ^2 squared Mahalanobis distance. Inserting the parameter blocks (A.11) and grouping powers of x_1 yields

$$\begin{aligned} \Delta^2 &= (x - \mu)\Lambda(x - \mu) \\ &= x_1^T \Lambda_{11} x_1 - 2x_1^T (\Lambda_{11} \mu_1 - \Lambda_{12} (x_2 - \mu_2)) \\ &\quad + \mu_1^T \Lambda_{11} \mu_1 - 2\mu_1^T \Lambda_{12} (x_2 - \mu_2) + (x_2 - \mu_2)^T \Lambda_{22} (x_2 - \mu_2). \end{aligned}$$

With the intention of establishing compact quadratic forms we introduce the quantities

$$\begin{aligned} \mu_{1|2} &= (\Lambda_{11})^{-1} (\Lambda_{11} \mu_1 - \Lambda_{12} (x_2 - \mu_2)) \\ &= \mu_1 - \Lambda_{11}^{-1} \Lambda_{12} (x_2 - \mu_2) \end{aligned} \quad (\text{A.17})$$

$$\Delta_{1*}^2 = (x_1 - \mu_{1|2})^T \Lambda_{11} (x_1 - \mu_{1|2}) \quad (\text{A.18})$$

and proceed. Completing the square in x_1 and subsequent simplification yields

$$\begin{aligned}
\Delta^2 &= \Delta_{1*}^2 - \mu_{1|2}^T \Lambda_{11} \mu_{1|2} \\
&\quad + \mu_1^T \Lambda_{11} \mu_1 - 2\mu_1^T \Lambda_{12} (x_2 - \mu_2) + (x_2 - \mu_2)^T \Lambda_{22} (x_2 - \mu_2) \\
&= \Delta_{1*}^2 + (x_2 - \mu_2)^T \left(\Lambda_{22} - \Lambda_{12}^T \Lambda_{11}^{-1} \Lambda_{12} \right) (x_2 - \mu_2) \\
&= \Delta_{1*}^2 + \Delta_2^2,
\end{aligned} \tag{A.19}$$

where we used (A.15) and introduced $\Delta_2^2 = (x_2 - \mu_2)^T \Sigma_{22}^{-1} (x_2 - \mu_2)$.

A.6 Derivation of the conditional t pdf

We here derive the conditional pdf of X_1 given that $X_2 = x_2$, as presented in Section 5. Although basic probability theory provides the result as ratio of joint and marginal pdf, the expressions take a very complicated form. Step by step, we disclose that the resulting pdf is again Student's t . In the course of the derivations, we make heavy use of the results from appendix A.5. We furthermore work on the α, β parametrization of the t pdf given in (3.1).

The conditional pdf is given as ratio of joint (X_1 and X_2) and marginal (X_2 only) pdf

$$\begin{aligned}
p(x_1|x_2) &= \frac{p(x_1, x_2)}{p(x_2)} \\
&= \frac{\Gamma(\alpha + \frac{d}{2})}{\Gamma(\alpha)} \frac{1}{(2\beta\pi)^{\frac{d}{2}}} \frac{1}{\sqrt{\det(\Sigma)}} \left(1 + \frac{\Delta^2}{2\beta} \right)^{-\frac{d}{2}-\alpha} \\
&\quad \left(\frac{\Gamma(\alpha + \frac{d_2}{2})}{\Gamma(\alpha)} \frac{1}{(2\beta\pi)^{\frac{d_2}{2}}} \frac{1}{\sqrt{\det(\Sigma_{22})}} \left(1 + \frac{\Delta_2^2}{2\beta} \right)^{-\frac{d_2}{2}-\alpha} \right)^{-1}
\end{aligned} \tag{A.20}$$

where we combined (4.2), (4.3), and (3.1). For the moment, we ignore the normalization constants and focus on the bracket term of $p(x_1, x_2)$ (which is colored in red in (A.20)). Using the result (A.19) we can rewrite it

$$\begin{aligned}
\left(1 + \frac{\Delta^2}{2\beta} \right)^{-\frac{d}{2}-\alpha} &= \left(1 + \frac{\Delta_2^2}{2\beta} + \frac{\Delta_{1*}^2}{2\beta} \right)^{-\frac{d}{2}-\alpha} \\
&= \left(1 + \frac{\Delta_2^2}{2\beta} \right)^{-\frac{d}{2}-\alpha} \left(1 + \frac{\Delta_{1*}^2}{2\beta + \Delta_2^2} \right)^{-\frac{d}{2}-\alpha}.
\end{aligned}$$

The bracket expressions (both red and blue) in (A.20) now simplify to

$$\begin{aligned}
&\left(1 + \frac{\Delta_2^2}{2\beta} \right)^{-\frac{d}{2}-\alpha} \left(1 + \frac{\Delta_{1*}^2}{2\beta + \Delta_2^2} \right)^{-\frac{d}{2}-\alpha} \left(1 + \frac{\Delta_2^2}{2\beta} \right)^{\frac{d_2}{2}+\alpha} \\
&= \left(1 + \frac{\Delta_2^2}{2\beta} \right)^{-\frac{d_1}{2}} \left(1 + \frac{\Delta_{1*}^2}{2\beta + \Delta_2^2} \right)^{-\frac{d}{2}-\alpha}.
\end{aligned} \tag{A.21}$$

The remaining bits of (A.20), i.e. the ratio of normalization constants can be written as

$$\frac{\Gamma(\alpha + \frac{d}{2}) (2\beta\pi)^{\frac{d_2}{2}} \sqrt{\det(\Sigma_{22})}}{\Gamma(\alpha + \frac{d_2}{2}) (2\beta\pi)^{\frac{d}{2}} \sqrt{\det(\Sigma)}} = \frac{\Gamma(\alpha + \frac{d_2}{2} + \frac{d_1}{2})}{\Gamma(\alpha + \frac{d_2}{2})} \frac{1}{(2\beta\pi)^{\frac{d_1}{2}}} \sqrt{\det(\Lambda_{11})}, \quad (\text{A.22})$$

where we used the determinant result (A.16). Combining the above results (A.21) and (A.22) gives the (still complicated) expression

$$p(x_1|x_2) = \frac{\Gamma(\alpha + \frac{d_2}{2} + \frac{d_1}{2}) \sqrt{\det(\Lambda_{11})}}{\Gamma(\alpha + \frac{d_2}{2}) (2\beta\pi)^{\frac{d_1}{2}}} \left(1 + \frac{\Delta_2^2}{2\beta}\right)^{-\frac{d_1}{2}} \left(1 + \frac{\Delta_{1*}^2}{2\beta + \Delta_2^2}\right)^{-\frac{d}{2} - \alpha},$$

which we shall now slowly shape into a recognizable t pdf. Therefore, we introduce $\alpha_{1|2} = \alpha + \frac{d_2}{2}$, and recall from (A.18) that $\Delta_{1*}^2 = (x_1 - \mu_{1|2})^T \Lambda_{11} (x_1 - \mu_{1|2})$:

$$\begin{aligned} p(x_1|x_2) &= \frac{\Gamma(\alpha_{1|2} + \frac{d_1}{2}) \sqrt{\det(\Lambda_{11})}}{\Gamma(\alpha_{1|2}) (2\beta\pi)^{\frac{d_1}{2}}} \left(1 + \frac{\Delta_2^2}{2\beta}\right)^{-\frac{d_1}{2}} \left(1 + \frac{\Delta_{1*}^2}{2\beta + \Delta_2^2}\right)^{-\frac{d_1}{2} - \alpha_{1|2}} \\ &= \frac{\Gamma(\alpha_{1|2} + \frac{d_1}{2})}{\Gamma(\alpha_{1|2})} \frac{\sqrt{\det(\Lambda_{11})}}{(\pi(2\beta + \Delta_2^2))^{\frac{d_1}{2}}} \left(1 + \frac{(x_1 - \mu_{1|2})^T \Lambda_{11} (x_1 - \mu_{1|2})}{2\beta + \Delta_2^2}\right)^{-\frac{d_1}{2} - \alpha_{1|2}} \end{aligned}$$

Comparison with (3.2) reveals that this is also a t distribution in α, β parametrization with

$$\begin{aligned} \beta_{1|2}^* &= \frac{1}{2}(2\beta + \Delta_2^2) = \beta + \frac{1}{2}(x_2 - \mu_2)^T \Sigma_{22}^{-1} (x_2 - \mu_2), \\ \Sigma_{1|2}^* &= \Lambda_{11}^{-1} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{12}^T. \end{aligned}$$

Obviously, $\alpha_{1|2} \neq \beta_{1|2}^*$. Because, however, only the product $\beta_{1|2}^* \Sigma_{1|2}^*$ enters a t pdf, the parameters can be adjusted as discussed in Section 3 and Appendix A.4. Converting back to ν parametrization, as in (1.1), we obtain $p(x_1|x_2) = \text{St}(x_1; \mu_{1|2}, \Sigma_{1|2}, \nu_{1|2})$ with

$$\begin{aligned} \nu_{1|2} &= 2\alpha_{1|2} = 2\alpha + d_2 = \nu + d_2, \\ \mu_{1|2} &= \mu_1 - \Lambda_{11}^{-1} \Lambda_{12} (x_2 - \mu_2) \\ &= \mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (x_2 - \mu_2), \\ \Sigma_{1|2} &= \frac{\beta_{1|2}^*}{\alpha_{1|2}} \Sigma_{1|2}^* \\ &= \frac{\beta + \frac{1}{2}(x_2 - \mu_2)^T \Sigma_{22}^{-1} (x_2 - \mu_2)}{\alpha + \frac{d_2}{2}} (\Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{12}^T) \\ &= \frac{\nu + (x_2 - \mu_2)^T \Sigma_{22}^{-1} (x_2 - \mu_2)}{\nu + d_2} (\Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{12}^T), \end{aligned}$$

which are the parameters provided in Section 5.


A.7 Mathematica code

The following code snippet can be used to re-produce Figure 3b. Moreover, by using mathematica's recently developed Manipulate function, the means to interactively change d and ν via two slider buttons are provided. More information can be found on www.wolfram.com/mathematica/.

```
Manipulate[Plot[{Log[1/(E^(Delta^2/2)*(2*Pi)^(d/2))],  
  Log[Gamma[(d + nu)/2]/(Gamma[nu/2]*((Pi*nu)^(d/2)*  
    (Delta^2/nu + 1)^((d + nu)/2))]}], {Delta, 0, 10},  
  PlotStyle -> {{Thick, Black}, {Thick, Black, Dashed}},  
  PlotRange -> {-40, 0}, AxesLabel -> {"Delta", ""},  
  {d, 1, 10, 1, Appearance -> "Labeled"},  
  {nu, 3, 100, 1, Appearance -> "Labeled"}]
```

REFERENCES

- [1] B. D. Anderson and J. B. Moore. *Optimal Filtering*. Prentice Hall, June 1979.
- [2] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, Aug. 2006.
- [3] M. H. DeGroot. *Optimal Statistical Decisions*. Wiley-Interscience, WCL edition, Apr. 2004.
- [4] A. Gut. *An Intermediate Course in Probability*. Springer, second edition, June 2009.
- [5] T. Kailath, A. H. Sayed, and B. Hassibi. *Linear Estimation*. Prentice Hall, Apr. 2000.
- [6] G. Koop. *Bayesian Econometrics*. Wiley-Interscience, July 2003.
- [7] S. Kotz and S. Nadarajah. *Multivariate t Distributions and Their Applications*. Cambridge University Press, Feb. 2004.
- [8] G. J. McLachlan and T. Krishnan. *The EM Algorithm and Extensions*. Wiley-Interscience, second edition, Mar. 2008.
- [9] C. R. Rao. *Linear Statistical Inference and its Applications*. Wiley-Interscience, second edition, Dec. 2001.

	Avdelning, Institution Division, Department Division of Automatic Control Department of Electrical Engineering	Datum Date 2013-04-17
	Språk Language <input type="checkbox"/> Svenska/Swedish <input checked="" type="checkbox"/> Engelska/English <input type="checkbox"/> _____	Rapporttyp Report category <input type="checkbox"/> Licentiatavhandling <input type="checkbox"/> Examensarbete <input type="checkbox"/> C-uppsats <input type="checkbox"/> D-uppsats <input checked="" type="checkbox"/> Övrig rapport <input type="checkbox"/> _____
URL för elektronisk version http://www.control.isy.liu.se		LiTH-ISY-R-3059
Titel On the Multivariate t Distribution Title		
Författare Michael Roth Author		
Sammanfattning Abstract <p>This technical report summarizes a number of results for the multivariate t distribution which can exhibit heavier tails than the Gaussian distribution. It is shown how t random variables can be generated, the probability density function (pdf) is derived, and marginal and conditional densities of partitioned t random vectors are presented. Moreover, a brief comparison with the multivariate Gaussian distribution is provided. The derivations of several results are given in an extensive appendix.</p>		
Nyckelord Keywords Student's t distribution, heavy tails, Gaussian distribution.		