JOURNAL OF AIRCRAFT
Vol. 45, No. 6, November–December 2008

Engineering Notes

ENGINEERING NOTES are short manuscripts describing new developments or important results of a preliminary nature. These Notes should not exceed 2500 words (where a figure or table counts as 200 words). Following informal review by the Editors, they may be published within a few months of the date of receipt. Style requirements are the same as for regular contributions (see inside back cover).

Transformed Normal Probability Density Functions for Parameter Estimation

Marco W. Soijer*
EADS Defence and Security, 85077 Manching, Germany

DOI: 10.2514/1.39299

Introduction

PARAMETER-ESTIMATION methods (as used, for example, in model-identification techniques for aircraft performance and control assessment) almost always depend on probability density functions that characterize the statistics of the model uncertainties and measurement errors. In well-known applications such as the Kalman filter or least-squares estimators, probability density functions are reduced to a value for the mean and variance of the distribution, thus assuming the distribution to be normal [1–3]. Kalman filters and least-squares estimators are special cases of maximum-likelihood or, even more generically, Bayesian estimation [4]. The general Bayesian or maximum-likelihood estimators depend on a description of the applicable probability density functions, but do not require these to be normal [5]. Nevertheless, most widespread methods make the assumption of normality for all random variables: for example, the dynamic system-identification methods described by Maine and Iliff [6] or those enumerated in the AGARD lecture series on rotorcraft system identification [7].

In applications in which an assumed normal probability density function does not appropriately describe the true statistics of the random variable, parameter-estimation results may be improved by using more realistic density functions for the random variables at hand. However, the popularity of the normal probability density function is caused by its quadratic nature, which allows for numerically efficient or even closed-form solution of the optimization problem. It is desirable to maintain as much of these characteristics as possible when formulating alternative probability density functions. This Note presents a probability density function that is directly derived from the normal density function; as such, the normal density function is a special case of the form presented here, which is referred to as the transformed normal probability density function.

Definition

With the goal to model the probability densities of the aerodynamic model parameters and the measurement corrections realistically, a probability density function is introduced that is based on the standard normal distribution, but allows for two important additional features: The probability density function can be asymmetric about its maximum (which then generally does not coincide with the parameter's mean). The bell shape of the normal distribution can be adjusted to have a more uniform distribution over a limited accuracy range and near-zero probability outside that range. The first feature is most important for any model parameter that has an a priori estimate with known asymmetric uncertainties. The second feature is more related to the modeling of measurement corrections. Within a single data recording that is used for parameter estimation (for example, a flight-test maneuver), measured parameters are usually expected to be biased. Typical examples are the angle of attack or the Mach number. The unknown biases can be realistically be limited, but as a stochastic variable, they do not have the characteristics of a normal distribution. Instead, biases occur within a limited range with almost equal probability over the specified range; outside the range, the probability drops rapidly.

The functional description starts with the expression for the normal probability density function:

$$p_n(x) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{(-\frac{1}{2}(\frac{x-\mu_x}{\sigma_x})^2)} \tag{1}$$

where x is the random variable, μ_x is the value of x for which the probability density function reaches its maximum (which, in this case, equals the mean of x), and σ_x is the standard deviation of x. The transformed normal probability density function is constructed from the normal density function by adjusting the order of the exponentiation, separating the function in left- and right-hand tolerance sides, and correcting the scale factor accordingly:

$$p_{\text{tn}}(x) = \frac{1}{A_d(\tau_R + \tau_L)} e^{(-\frac{1}{2}(\frac{x-\mu_x}{\tau})^{2d})} \tag{2}$$

where d is the order of the transformed function; A_d represents the order-dependent scale factor; τ is the tolerance; and τ_L and τ_R represent the left and right tolerances, respectively, with $\tau_L > 0$ and $\tau_R > 0$. The tolerances are chosen as $\tau = \tau_L$ for $x < \mu_x$ and $\tau = \tau_R$ otherwise. The order d must be a positive integer; the multiplication by 2 in the exponent ensures that the function is decreasing in both directions away from μ_x.

With the appropriate substitution of A_d, Eq. (2) defines a valid probability density function. For this to be true, the function must be nonnegative and its integral over the real line must be 1 [6]. When A_d is positive, the function is obviously nonnegative. What remains to be shown is that the function's integral over the real line is finite and independent of τ_L, τ_R, and μ_x, for a value of A_d to exist that scales the integral to unity. First, the integral is split in one part below and one part above the maximum μ_x:

$$\int_{-\infty}^{\infty} p_{\text{tn}}(x)\mathrm{d}x = \frac{1}{A_d(\tau_R + \tau_L)} \times \left[\int_{-\infty}^{\mu_x} e^{(-\frac{1}{2}(\frac{x-\mu_x}{\tau_L})^{2d})}\mathrm{d}x + \int_{\mu_x}^{\infty} e^{(-\frac{1}{2}(\frac{x-\mu_x}{\tau_R})^{2d})}\mathrm{d}x\right] \tag{3}$$

In the second integral, ξ is introduced as $(x - \mu_x)/\tau_R$. Thus, $dx = \tau_R d\xi$:

Received 23 June 2008; accepted for publication 11 July 2008.

*Military Air Systems; currently Leader, Electronic Warfare Acquisition Team, Defence Electronics, Woerthstrasse 85, 89077 Ulm, Germany; marco.soijer@eads.com. Senior Member AIAA.

$$\int_{\mu_x}^{\infty} e^{(-\frac{1}{2}(\frac{x-\mu_x}{\tau_R})^{2d})}\mathrm{d}x = \tau_R \int_0^{\infty} e^{(-\frac{1}{2}\xi^{2d})}\mathrm{d}\xi \tag{4}$$

The result of the integral on the right-hand side of Eq. (4) is only dependent on d. As a contribution to Eq. (3), it is scaled with τ_R. The right-hand side of Eq. (4) is split once more. The part for $\xi < 1$ is majorized with 1, as the exponent is always negative. The part for $\xi > 1$ is majorized by multiplying the integrand with ξ^{2d-1}, which is always larger than 1. Thus,

$$\int_{\mu_x}^{\infty} e^{(-\frac{1}{2}(\frac{x-\mu_x}{\tau_R})^{2d})}\mathrm{d}x < \tau_R\left(\int_0^1 1\mathrm{d}\xi + \int_1^{\infty} \xi^{2d-1} e^{(-\frac{1}{2}\xi^{2d})}\mathrm{d}\xi\right) \tag{5}$$

The right-hand side of Eq. (5) can easily be resolved:

$$\int_{\mu_x}^{\infty} e^{(-\frac{1}{2}(\frac{x-\mu_x}{\tau_R})^{2d})}\mathrm{d}x < \tau_R\left([\xi]_0^1 + \left[\frac{-1}{d} e^{(-\frac{1}{2}\xi^{2d})}\right]_1^{\infty}\right) = \tau_R\left(1 + \frac{1}{d\sqrt{e}}\right) \tag{6}$$

which proves that the second integral of Eq. (3) is finite. The same procedure can be applied to the first integral in Eq. (3). Now A_d is selected as

$$\int_0^{\infty} e^{(-\frac{1}{2}\xi^{2d})}\mathrm{d}\xi$$

A combination of Eqs. (3) and (4) shows that

$$\int_{-\infty}^{\infty} p_{\mathrm{tn}}(x)\mathrm{d}x = 1$$

which completes the proof that Eq. (2) is a valid probability density function. Table 1 shows numerically obtained values of A_d for the first five function orders d. Note that the value for $d = 1$ equals $\sqrt{\pi/2}$, as it is known from the standard normal probability density function, when $\tau_R = \tau_L = \sigma_x$ is substituted into Eq. (2).

Characteristics

The function order d is used to increase the flattening of the function about its maximum value and to steepen the slopes. Typical values for the order d are 1 through 4; higher values have little qualitative effect, but may complicate numerical analysis as the function's gradient approaches zero close to and far away from the maximum, which makes it difficult for an optimization procedure to determine in what direction the maximum is found. Figure 1 shows three transformed normal probability density functions with orders 1, 2, and 4, respectively, where all τ are unity and $\mu_x = 0$.

When $d = 1$ and $\tau_L = \tau_R$, the transformed normal probability density function equals the normal probability density function with $\sigma_x = \tau$. For function orders larger than 1, or for asymmetric functions where $\tau_L \neq \tau_R$, the tolerance no longer equals the standard deviation of the random variable. However, Eq. (2) shows that, similar to the standard deviation in case of the original normal probability density function, the left and right tolerances determine at which distance from the maximum at $x = \mu_x$ that the probability density has dropped to a value of $1/\sqrt{e}$ (approximately 60.7%) of the maximum. Figure 2 shows some examples of transformed normal probability density functions with different right tolerances, where the left tolerance and the function order are 1 and $\mu_x = 0$.

Table 1 Scale factors for the transformed normal probability density function

d	A_d
1	1.2533
2	1.0779
3	1.0413
4	1.0270
5	1.0196

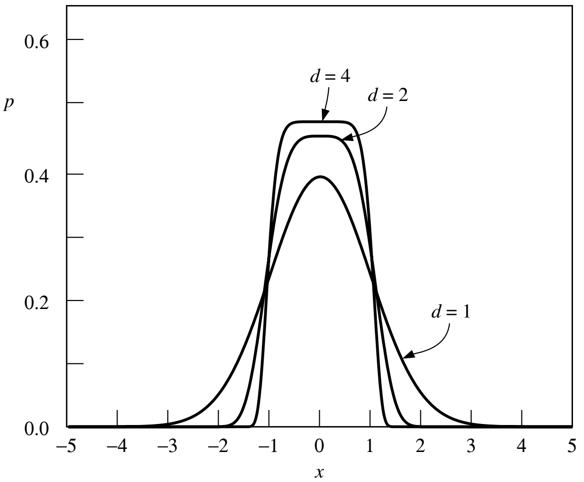

Fig. 1 Symmetric transformed normal density functions for various orders.

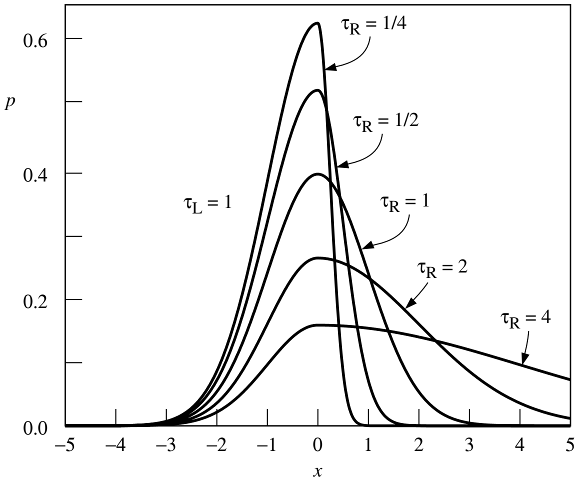

Fig. 2 Transformed normal density functions of order 1 for various asymmetric tolerances.

Use in Numeric Optimization

The transformed probability density function does not preserve all the favorable characteristics of the original normal probability density function for optimization problems. The discontinuity in τ at the probability maximum at μ_x make it almost impossible to use the function for a closed-form solution as it is available for the Gaussian distribution in a least-squares estimator.

However, typical linear and nonlinear parameter-estimation techniques (for example, the maximum-likelihood estimator [4,6]) do not provide a closed-form solution, but include a numeric optimization procedure. A joint Gaussian probability density function is obtained by multiplying several functions for different parameters, and the maximum of the multivariate function is sought. A well-known strategy for finding the maximum is to take the logarithm of the joint density function to differentiate with respect to all variables and to find the values for which the gradient equals zero. This converts the multiplication of scaled exponent functions into a simpler summation of differentiated exponents only. This technique can also be applied when the transformed normal probability density function is used. Therefore, it is possible to integrate this type of function with well-known parameter-estimation techniques without complication.

Conclusions

With the function order d and the recognition of distinct left and right tolerances, the transformed normal probability density function offers two additional degrees of freedom with respect to the Gaussian probability density function. The following cases taken from the practice of aerodynamic model identification at the flight-test center of EADS Defence and Security, Military Air Systems, are a few suggestions of how these additional degrees of freedom with respect to the Gaussian model can be used:

1) Tolerance parameters (for example, aerodynamic stability or control coefficients) may be known or may have to be considered with different uncertainties above and below the nominal or previously estimated value. Stability analyses for static longitudinal stability may, for example, consider a larger tolerance in the positive direction on the pitch stability C_{m_α} than in the negative direction. This can be modeled using different left and right tolerances.

2) Boundary conditions on parameters in an estimation process can be modeled smoothly. Consider, for example, a necessarily positive variable with a certain expected mean and a large standard deviation compared with the mean. Using a Gaussian distribution will produce negative estimates; introducing a boundary condition that forces the parameter to be positive will produce an unrealistic clutter of estimates just above zero. An asymmetric tolerance can keep the estimate away from zero while allowing large deviations above the expected mean.

3) Estimates for measurement biases are bounded by the known accuracy of the measurement, where biases within the accuracy range are rather uniformly distributed. Using a Gaussian distribution will either produce estimates that generally tend to be too small or, when the standard deviation is enlarged, will regularly produce estimates that are known to be incorrect, as they are well outside the accuracy range. Higher-order transformed probability density functions can model these biases correctly.

References

[1] Kalman, R. E., and Bucy, R. S., "New Results in Linear Filtering and Prediction Theory," *Journal of Basic Engineering*, Vol. 63, Mar. 1961, pp. 95–107.

[2] Gelb, A. (ed.), *Applied Optimal Estimation*, MIT Press, Cambridge, MA, 1974, pp. 27–34, 102–119.

[3] Nahi, N. E., *Estimation Theory and Applications*, Wiley, New York, 1969, pp. 31–57, 108–115.

[4] Eykhoff, P., *System Identification: Parameter and State Estimation*, Wiley, London, 1974, pp. 137–149.

[5] Sage, A. P. and Melsa, J. L., *Estimation Theory with Applications to Communications and Control*, McGraw–Hill, New York, 1971, pp. 175–249.

[6] Maine, R. E., and Iliff, K. W., "Identification of Dynamic Systems: Theory and Formulation," NASA RP 1138, Feb. 1985, pp. 23, 45–72.

[7] *Rotorcraft System Identification*, AGARD Lecture Series, Vol. LS 178, AGARD, Neuilly-sur-Seine, France, Oct. 1991, pp. 1.4, 5.1–5.9, 6.1–6.4, 7.1–7.9.