Approximating the Distribution for Sums of Products of Normal Variables

Robert Ware[1] and Frank Lad[2]

Abstract

We consider how to calculate the probability that the sum of the product of variables assessed with a Normal distribution is negative. The analysis is motivated by a specific problem in electrical engineering. To resolve the problem, two distinct steps are required. First, we consider ways in which we can assess the distribution for the product of two Normally distributed variables. Three different methods are compared: a numerical methods approximation, which involves implementing a numerical integration procedure on MATLAB, a Monte Carlo construction and an approximation to the analytic result using the Normal distribution. The second step considers how to assess the distribution for the sum of the products of two Normally distributed variables by applying the Convolution Formula. Finally, the two steps are combined to compute the distribution for the sum of products of Normally distributed variables, and thus to calculate the probability that this sum of products is negative. The problem is also approached directly, using a Monte Carlo approximation.

Key Words: Product of Normal variables; Numerical integration; Differential continuous phase frequency shift keying; Convolutions; Moment generating function.

1 Introduction

The work discussed in this Report originated from a problem posed by Griffin (2000) who was interested in calculating the probability that the sum of the product of variables assessed with a Normal distribution was negative. Previous work involving the distribution of the product of two Normally distributed variables has been undertaken by Craig (1936), who was the first to determine the algebraic form of the moment-generating function of the product. Aroian et al. (1978) proved that, under certain conditions, the product of two Normally distributed variables approaches the standardised Pearson type III distribution. Cornwell et al. (1978) described the numerical evaluation of the product. Conradie and Gupta (1987) presented basic

[1]Research Fellow, School of Population Health, The University of Queensland

[2]Research Associate, Department of Mathematics and Statistics, University of Canterbury

distributional results of the quadratic forms of p-variate complex Normal distributions. Their results were developed in terms of characteristic functions. However, these results were found to be too theoretical to be easily transferred to a problem as applied as the one under investigation.

While the work of Craig (1936) is relatively old, it is not at all well-known among statisticians. Indeed we did not even learn of it until we had rediscovered it ourselves in developing this Thesis. At that time we also learned of the researches of Aroian et al. (1978) and Cornwell et al. (1978). Nonetheless, recent advances in computing power and graphics have allowed us to make several useful advances on the consulting problem that had been posed.

To calculate the probability that the sum of products of variables assessed as having a Normal distribution is negative, two distinct steps are required. The first step considers ways in which we can assess the distribution for the product of two Normally distributed variables. The second step involves identifying the distribution when summing a number of these products together.

In Section 2, the process of differential continuous phase frequency shift keying is briefly introduced as it was studied by Griffin (2000). The relevance of the distribution of sums of products of Normally distributed variables is recognised. In Section 3 we assess the distribution for $Y = X_1X_2$, the product of two independent Normally distributed variables. We compare three different methods: a numerical methods approximation, which involves implementing a numerical integration procedure on MATLAB, a Monte Carlo construction and an approximation to the analytic result using the Normal distribution. The numerical integration procedure involves the joint distribution of Y and X_2. We discover that $f(y, x_2)$ has a simple singularity at $(0, 0)$, and discuss the resulting consequences for the numerical integration. The numerical integration is implemented via MATLAB and Maple subroutines, eliminating the need for evaluation via statistical tables. New graphics are presented to aid understanding of the shape of the distribution Y. We undertake a specific analysis of the skewness of the product of two Normally distributed variables when the multiplicands are correlated. Section 4 considers how to assess the distribution for the sum of the products of two Normally distributed variables by applying the Convolution Formula. This technique is demonstrated using the products previously obtained via numerical integration. A computational procedure for approximating the required distribution using convolutions is developed. In Section 5 the methods of Sections 3 and 4 are combined to compute the distribution for the sum of products of Normally distributed variables, and thus to calculate the probability that this sum of products is negative. We also approach this problem directly, using a Monte Carlo approximation. Finally, in Section 6 a summary of the Report is presented.

2 Introduction to Differential Continuous Phase Frequency Shift Keying

Differential continuous phase frequency shift keying is a procedure for transmitting and decoding a signal that has been intentionally disturbed by noise. Interest centres on how accurately the receiver can decode the original signal.

In a typical encoding problem the original signal, called $s(t)$, is differentially encoded and then transmitted. During transmission over a channel, Gaussian noise, $w(t)$, is added to the signal so that the received signal, $y(t)$, is a combination of the transmitted signal and noise, i.e. $y(t) = s(t)+w(t)$. The ratio of $s(t)$ to the standard deviation of $w(t)$ is called the "signal to noise ratio". The estimate of the original signal is called the hypothesised signal, and is denoted $\hat{s}(t)$. The receiver uses a decoder to minimise the squared Euclidean distance between the transmitted signal and the hypothesised signal. The performance of the receiver can be characterised by the probability of error between $s(t)$ and $\hat{s}(t)$. As is common in problems of electrical engineering, the problem is expressed via complex valued functions and the practical solution is determined by the real component of the complex solution. In the problem posed by Griffin (2000), interest centres on the probability that the real component of the error metric between the transmitted and hypothesised signals is less than 0, that is, $P[Re(M_e(s, \hat{s})) < 0]$.

The transmitted signal consists of a finite number of received signals, say N of them, so that

$$\mathbf{y} = \begin{bmatrix} y_{-\xi+1} \\ y_{-\xi+2} \\ \vdots \\ y_t \\ \vdots \\ y_{-\xi+N} \end{bmatrix} = \begin{bmatrix} s_{-\xi+1} + w_{-\xi+1} \\ s_{-\xi+2} + w_{-\xi+2} \\ \vdots \\ s_t + w_t \\ \vdots \\ s_{-\xi+N} + w_{-\xi+N} \end{bmatrix} = \mathbf{s} + \mathbf{w} \tag{1}$$

where ξ is a positive integer. In the case of the problem posed, the error metric is achieved as the real component of a complex expression and is representable as

$$M_e(s, \hat{s}) = 2\sum_{t=1}^{N\xi} \left(y_{R,t} y_{R,t-\xi} b_{R,t} + y_{R,t} y_{I,t-\xi} b_{I,t} - y_{I,t} y_{R,t-\xi} b_{I,t} + y_{I,t} y_{I,t-\xi} b_{R,t} \right). \tag{2}$$

The coefficients $b_{R,t}$ and $b_{I,t}$ are real constants, predetermined along with the means of $y_{R,t}$ and $y_{I,t}$ by the signal that is sent. Furthermore the added noise is constructed so that

$$\begin{bmatrix} y_{R,t} \\ y_{I,t} \\ y_{R,t-\xi} \\ y_{I,t-\xi} \end{bmatrix} \sim N \left(\begin{bmatrix} \mu_{R,t} \\ \mu_{I,t} \\ \mu_{R,t-\xi} \\ \mu_{I,t-\xi} \end{bmatrix}, \begin{bmatrix} \sigma^2_{\mathbf{w}} & 0 & 0 & 0 \\ 0 & \sigma^2_{\mathbf{w}} & 0 & 0 \\ 0 & 0 & \sigma^2_{\mathbf{w}} & 0 \\ 0 & 0 & 0 & \sigma^2_{\mathbf{w}} \end{bmatrix} \right).$$

The relative size of the different values of μ and $\sigma^2_{\mathbf{w}}$ are determined by the signal to noise ratio. For problems of this type it is usual practise to set all values of μ equal to 1, and then set $\sigma^2_{\mathbf{w}}$ to achieve the required signal to noise ratio for the problem under investigation. N is usually given the value 2 or 3, ξ is set to be some value of 2^k, where $k \geq 0$, and $b_{R,t}$ and $b_{I,t}$ are both set to 1.

Our interest centres on finding the probability that the real component of the error metric is negative. In Sections 3 and 4 our focus will be on the shape of the distribution of the error metric. Thus, we shall disregard the coefficient "2" from Equation 2.

Essentially, the problem posed by $P\left[Re\left(M_e(s, \hat{s})\right) < 0\right]$ requires that we be able to compute probabilities for the sum of products of Normal variables. We now turn to a study of this problem in a general context.

3 The Simplest Problem of the Product of Two Normal Variables

The form of the simplest problem we consider is a bivariate Normal distribution with independent variables:

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}\right).$$

The joint density of X_1 and X_2 is

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - \frac{1}{2}\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2} \qquad \text{for all} \quad x_1, x_2 \in \mathcal{R} \tag{3}$$

Our interest centres on the distribution of the product X_1X_2. To simplify notation, for the remainder of this Report we let $Y = X_1X_2$.

3.1 A Numerical Methods Approximation

The conditional distribution of $Y \mid (X_2 = x_2)$, and the distribution of X_2, are:

$$\begin{aligned} Y \mid X_2 = x_2 &\sim N(x_2\mu_1, x_2^2\sigma_1^2), \\ X_2 &\sim N(\mu_2, \sigma_2^2). \end{aligned}$$

Thus we can find the joint density of X_2 and Y:

$$\begin{aligned} f(y, x_2) &= f(y \mid x_2)f(x_2) \\ &= \frac{1}{\sqrt{2\pi}|x_2|\sigma_1} e^{-\frac{1}{2x_2^2\sigma_1^2}(y - x_2\mu_1)^2} \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{1}{2\sigma_2^2}(x_2-\mu_2)^2} \\ &= \frac{1}{2\pi|x_2|\sigma_1\sigma_2} e^{-\frac{1}{2\sigma_1^2}\left(\frac{y}{x_2} - \mu_1\right)^2 - \frac{1}{2\sigma_2^2}(x_2-\mu_2)^2}. \end{aligned} \tag{4}$$

To find the marginal density $f(y)$, we need to integrate $f(y \mid x_2)f(x_2)$ with respect to x_2. In other words we solve

$$\begin{aligned} f(y) &= \int_{-\infty}^{\infty} f(y \mid x_2)f(x_2)dx_2 \\ &= \int_{-\infty}^{\infty} f(y, x_2)dx_2. \end{aligned} \tag{5}$$

This integration can be undertaken using the numerical integration procedure on the mathematical computer package MATLAB called “quad8”, which works by using an adaptive recursive Newton Cotes 8 panel rule. A numeric value of $\int f(y, x_2)dx_2$ is obtained for an array of points in the domain of Y. These points are all extremely close to one another and essentially cover all realistic possibilities for y. We consider the density as essentially uniform on these tight intervals. These integrated values are then summed and normalised. The same numerical integration can be performed on the computer package Maple. The calling sequence "evalf(Int)" invokes a Clenshaw-Curtis quadrature method. As in MATLAB, a numeric value of $\int f(y, x_2)dx_2$ is obtained for an array of points in the domain of Y. We shall compare the results of these two integration procedures.

3.1.1 Case 1: $\mu_1 = 1, \mu_2 = 0.5, \sigma^2 = 1, cov(X_1, X_2) = 0$.

Consider the case

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N\left(\begin{bmatrix} 1 \\ 0.5 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right).$$

The joint density of X_1 and X_2 is called a circular Normal density, and by Equation 3

$$f(x_1, x_2) = \frac{1}{2\pi} e^{-\frac{1}{2}\left(x_1^2 - 2x_1 + x_2^2 - x_2 + \frac{5}{4}\right)}. \tag{6}$$

We can view $f(x_1, x_2)$ in three dimensions, Figure 1, as well as by continuous contours in two dimensions, Figure 2. Figure 1 shows that $f(x_1, x_2)$ is a bivariate Normal surface with a maximum at $(1, 0.5)$. In Figure 2 the contours of constant probability density for $f(x_1, x_2)$ are drawn at 0.005, 0.055, 0.105 and 0.155. The isoproduct lines are shown for $\mathrm{x}_1 x_2 = \pm 1, \pm 2, \pm 3, \pm 4, \pm 5$.

The joint density $f(y, x_2)$ can be easily calculated. By Equation 4 we have

$$f(y, x_2) = \frac{1}{2\pi |x_2|} e^{-\frac{1}{2}\left(\frac{y^2}{x_2^2} - \frac{2y}{x_2} + x_2^2 - x_2 + \frac{5}{4}\right)}. \tag{7}$$

A study of $f(y, x_2)$ shows us that:

1. The domain for (X_2, Y) is $\mathcal{R}^2 - \{(0, y) \mid y \neq 0\}$, because $y = x_1 x_2$ and if $x_2 = 0$, y must be 0.
2. When x_2 and y are both large, $f(y, x_2) \approx 0$.
3. At pairs $(y, x_2) \neq (0, 0)$ the density will be what it is — varying real values.

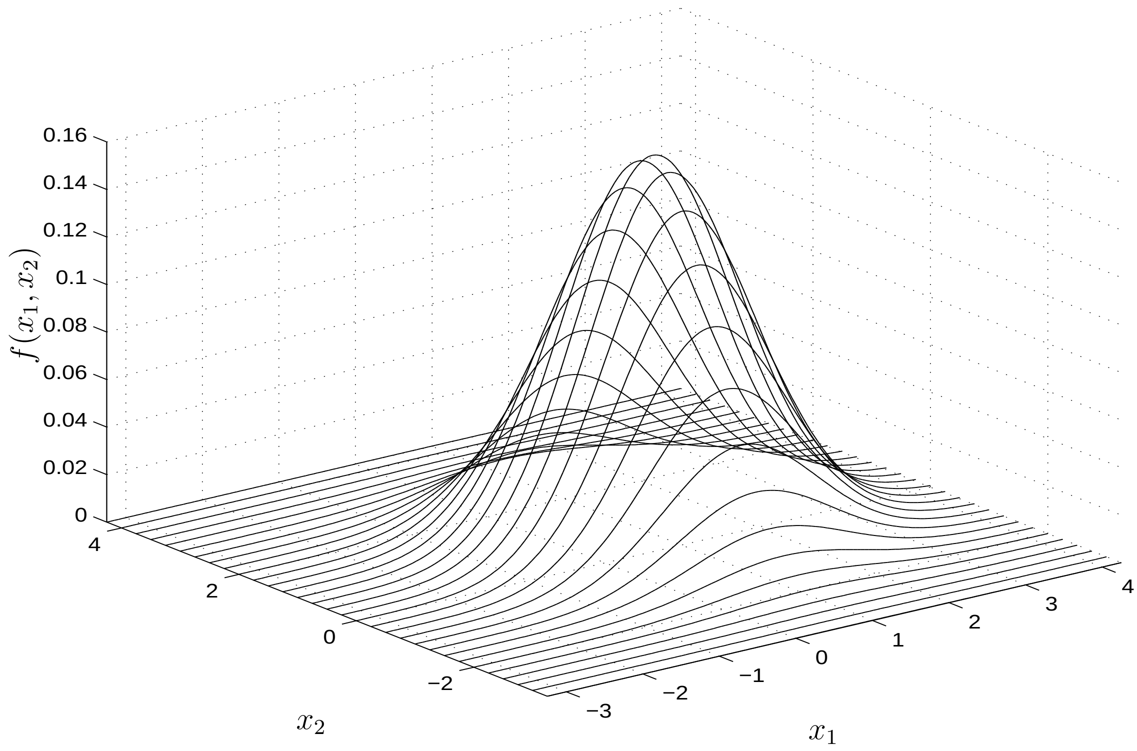

Figure 1: Joint density of $X_1 \sim N(1,1)$ and $X_2 \sim N(0.5,1)$ displayed in three dimensions.

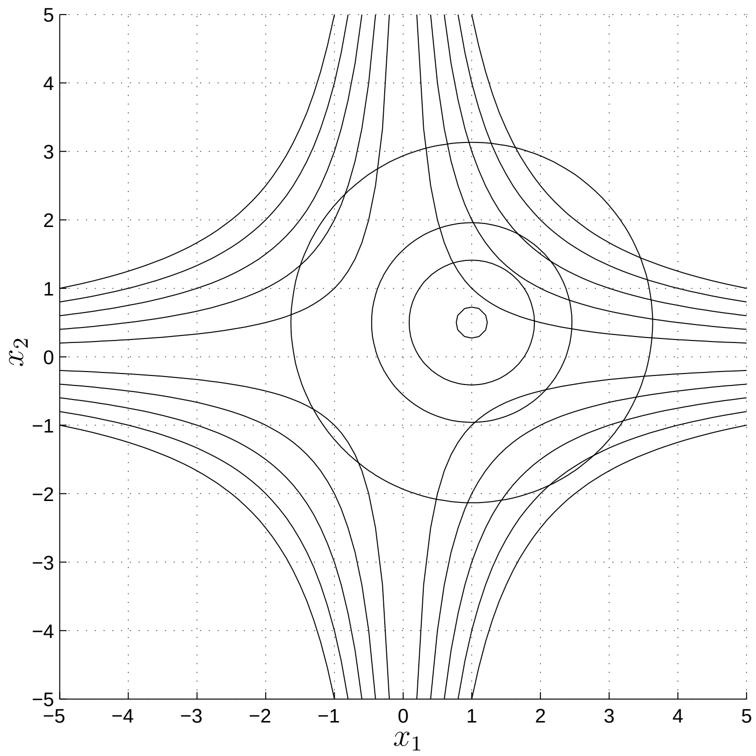

Figure 2: Constant density contours for the joint density $f(x_1, x_2)$ displayed in two dimensions, along with the isoproduct lines for X_1X_2.

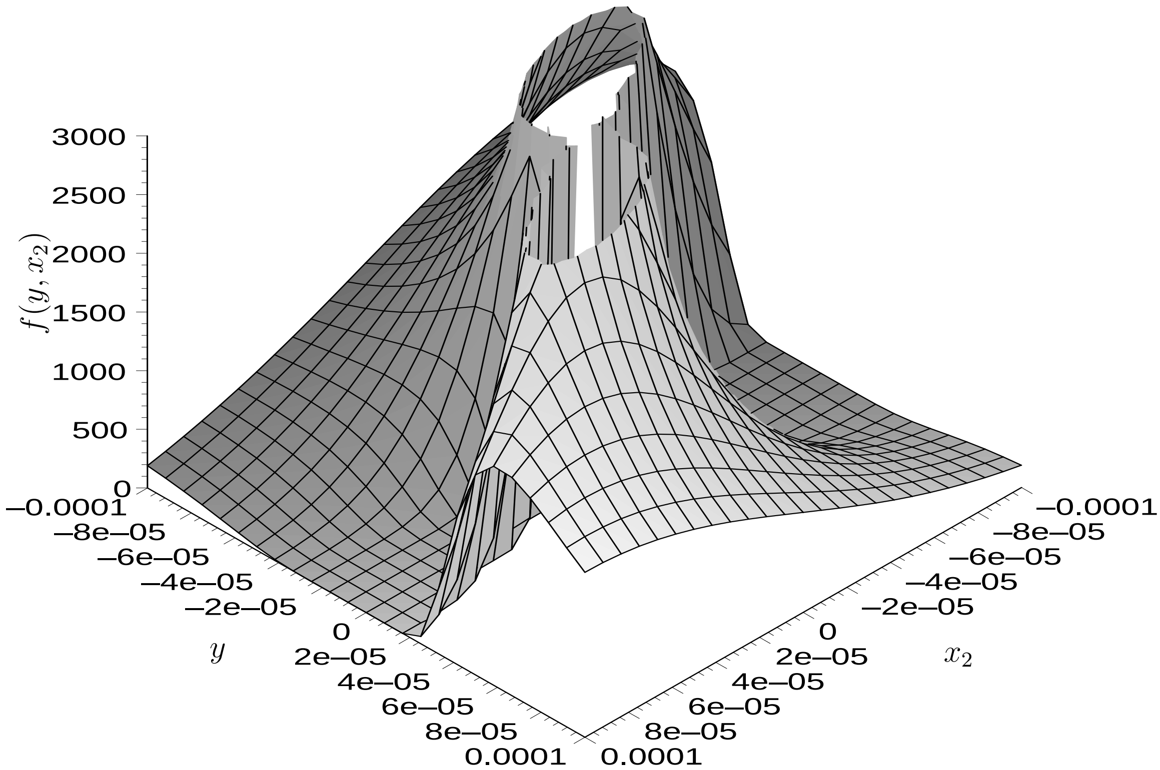

Figure 3: The joint density $f(y, x_2)$ displayed in three dimensions for Case 1. For clarity the density is only displayed for $f(y, x_2) < 3000$.

4. The joint density has a simple singularity at $(y, x_2) = (0, 0)$. The direction from which the domain point $(y, x_2) = (c, 0)$ is approached does not affect the fact that $\lim_{x_2 \to 0} f(y, x_2)$ does not exist. To see this, let $y = kx_2$ for some real k. Then Equation 7 becomes

$$f(y, x_2) = \frac{1}{2\pi |x_2|} e^{-\frac{1}{2}\left(k^2 - 2k + x_2^2 - x_2 + \frac{5}{4}\right)}. \tag{8}$$

In this form it is easy to see that $f(y, x_2) \to \infty$ as $x_2 \to 0$ for every real k. Moreover, it is evident directly from Equation 8 that for any fixed value of $y \neq 0$, $\lim_{x_2 \to 0} f(y, x_2) = 0$. (Remember that points $(y, x_2) = (y, 0)$ for $y \neq 0$ are missing from the domain of (Y, X_2).)

Let us now examine the density $f(y, x_2)$ as produced by Maple in Figure 3 and MATLAB in Figure 4. Figure 3 displays a standard three dimensional picture of the function, while Figure 4 shows the isodensity contours. The larger contours correspond to low density values, the smaller contours correspond to high density values.

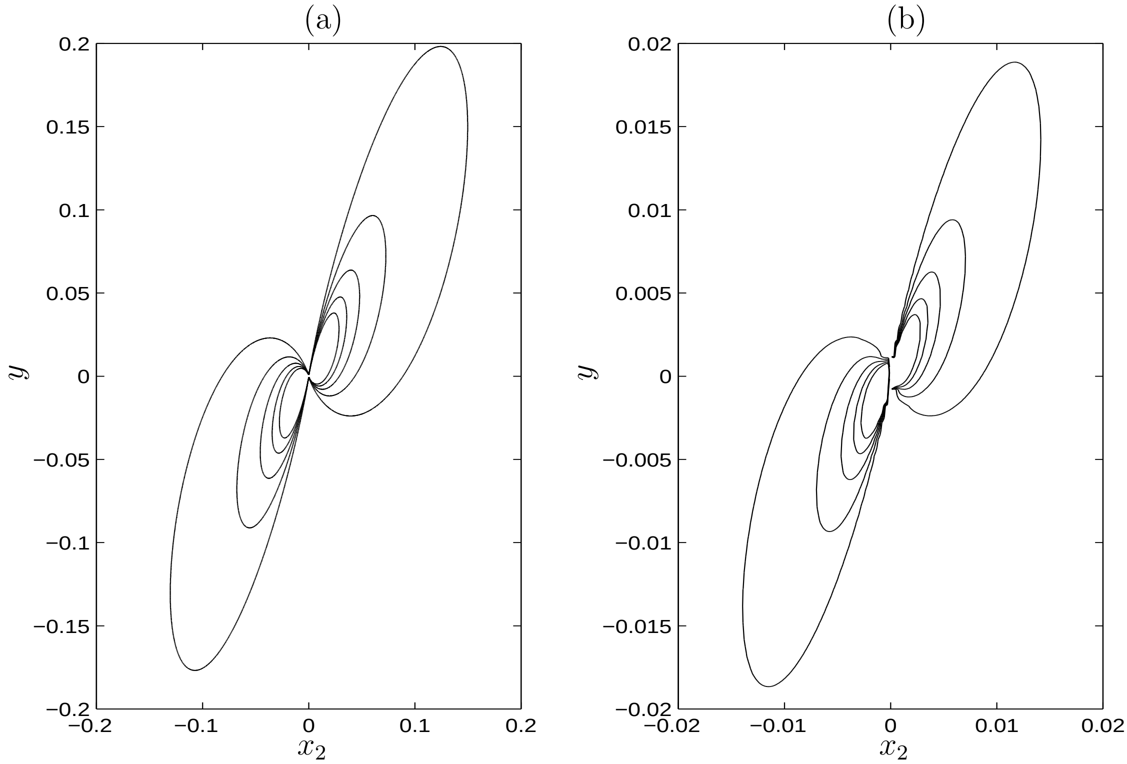

Figure 4: Contours for the joint density $f\,(y, x_2)$ displayed in two dimensions for Case 1. In (a) contours are plotted at $f(y, x_2) = 1, 2, 3, 4, 5$. In (b) contours are plotted at $f(y, x_2) = 10, 20, 30, 40, 50$.

A consequence of $f(y, x_2)$ being undefined when $x_2 = 0$ and $y \neq 0$ is that to integrate $f(y, x_2)$ numerically we need to separate the integral into two domains for $X_2 : x_2 \in (-\infty, 0)$ and $x_2 \in (0, \infty)$. Once the positive and negative halves have been integrated individually, they are added together and normalised to give the marginal density $f(y)$. In MATLAB it took 10 minutes to complete this numerical integration using interval widths of 0.01. Figure 5 demonstrates that the marginal density is neither Normal or symmetric. Notice there is a slight fluctuation in $f(y)$ at $y \approx -2.3$. Similar, and far more marked, fluctuations occur in Case 2, and are discussed in the next subsection.

The density has an asymptote at $y = 0$, but of course we cannot display it graphically with the computer. There is a limit to how finely we can break up the domain of X_2 numerically. But because of our analysis of the $f(y, x_2)$ function we know that if we could do the numerical integration more finely, the area of the inner region around $y = 0$ would continue to increase. This is due to the simple singularity

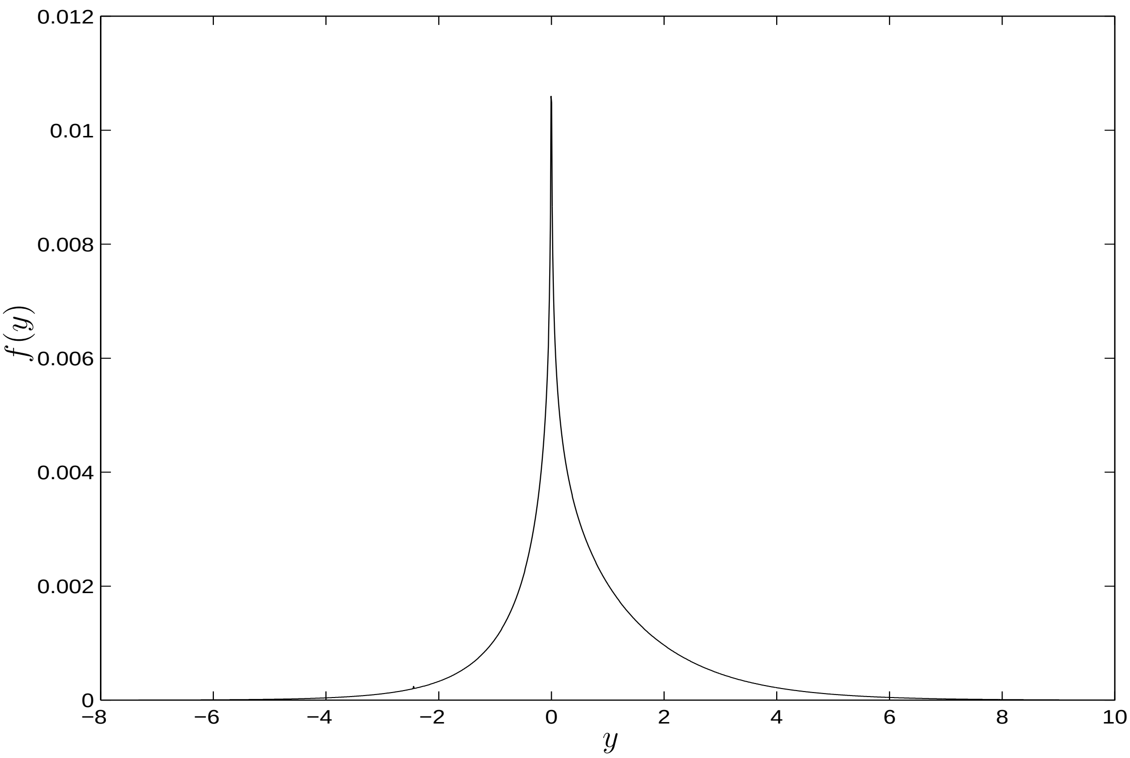

Figure 5: The marginal density $f(y)$ obtained via numerical integration for Case 1. The function was integrated between -13 and 21, with interval widths of 0.01.

of $f(y, x_2)$ at $(y, x_2) = (0, 0)$.

3.1.2 Case 2: $\mu_1 = 5, \mu_2 = 2, \sigma^2 = 1, cov(X_1, X_2 = 0)$.

The density studied in the previous subsection was complicated by the extreme behaviour of $f(y, x_2)$ around the origin and the asymptote in $f(y)$ at $y = 0$. We shall now re-illustrate the method used in the previous subsection using values of μ that give a 'nicer' joint density $f(y, x_2)$.

We shall consider the case

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N\left(\begin{bmatrix} 5 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right).$$

From Equations 3 and 4 we calculate the joint densities $f(x_1, x_2)$ and $f(y, x_2)$ to be

$$f(x_1, x_2) = \frac{1}{2\pi} e^{-\frac{1}{2}\left(x_1^2 - 10x_1 + x_2^2 - 4x_2 + 29\right)} \tag{9}$$

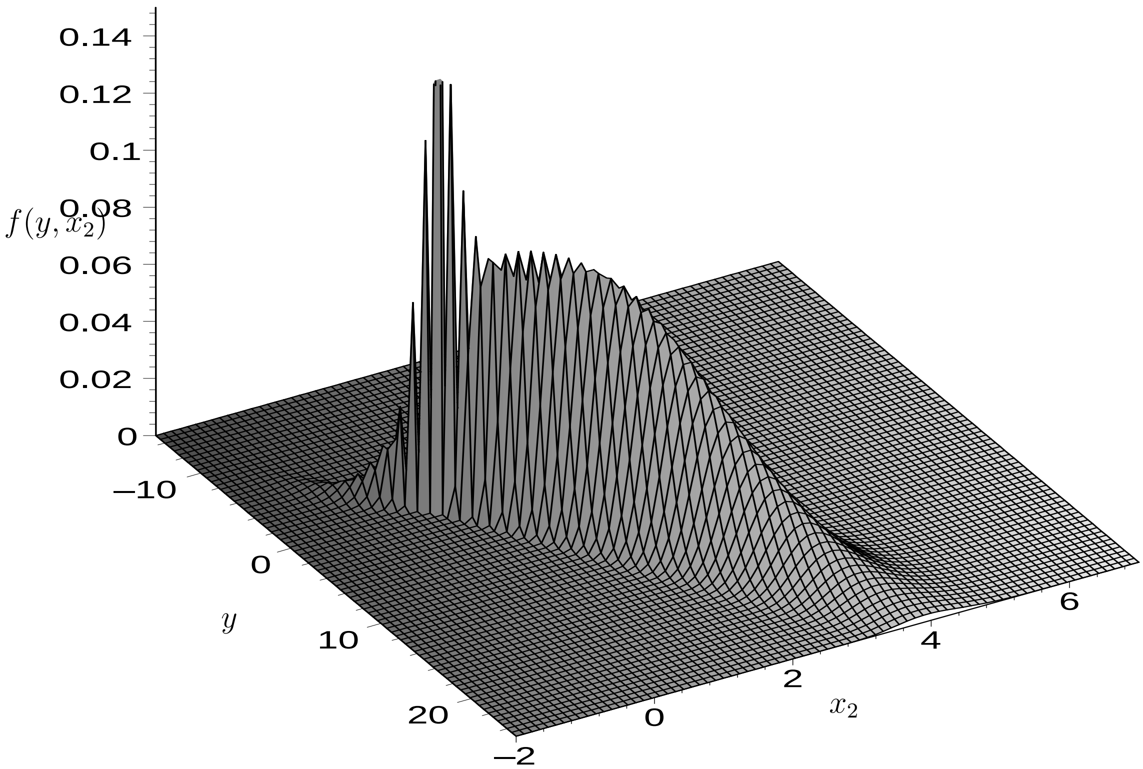

Figure 6: The joint density $f(y, x_2)$ displayed in three dimensions for Case 2.

and

$$f(y, x_2) = \frac{1}{2\pi |x_2|} e^{-\frac{1}{2}\left(\frac{y^2}{x_2^2} - \frac{10y}{x_2} + x_2^2 - 4x_2 + 29\right)}. \tag{10}$$

Of course Equation 9 shows the density $f(x_1, x_2)$ is a bivariate Normal surface which has a maximum at $(5, 2)$. The contours of $f(x_1, x_2)$ are circles.

The joint density $f(y, x_2)$ is displayed in three dimensions in Figure 6. We can see that, as in Case 1, there is a singularity at the origin. The analysis of the function we studied in Case 1 did not depend on the values of μ_1 and μ_2 at this point. Figure 7 displays $f(y, x_2)$ in two dimensions. Contour lines are drawn at $f(y) = 0.01, 0.03, 0.05, 0.07, 0.09$. The larger isodensity contours correspond to the lower density values and vice versa. Notice there are no contours lines displayed when $y > 0$ and $x_2 < 0$, or when $y < 0$ and $x_2 > 0$. The largest value of $f(y, x_2)$ in either of these quadrants is 7.9×10^{-6}.

To find the marginal density $f(y)$ we implement the same numerical integration procedure on MATLAB that was described in the previous subsection. We split $f(y, x_2)$ into negative and positive domains for X_2 and integrate the function over

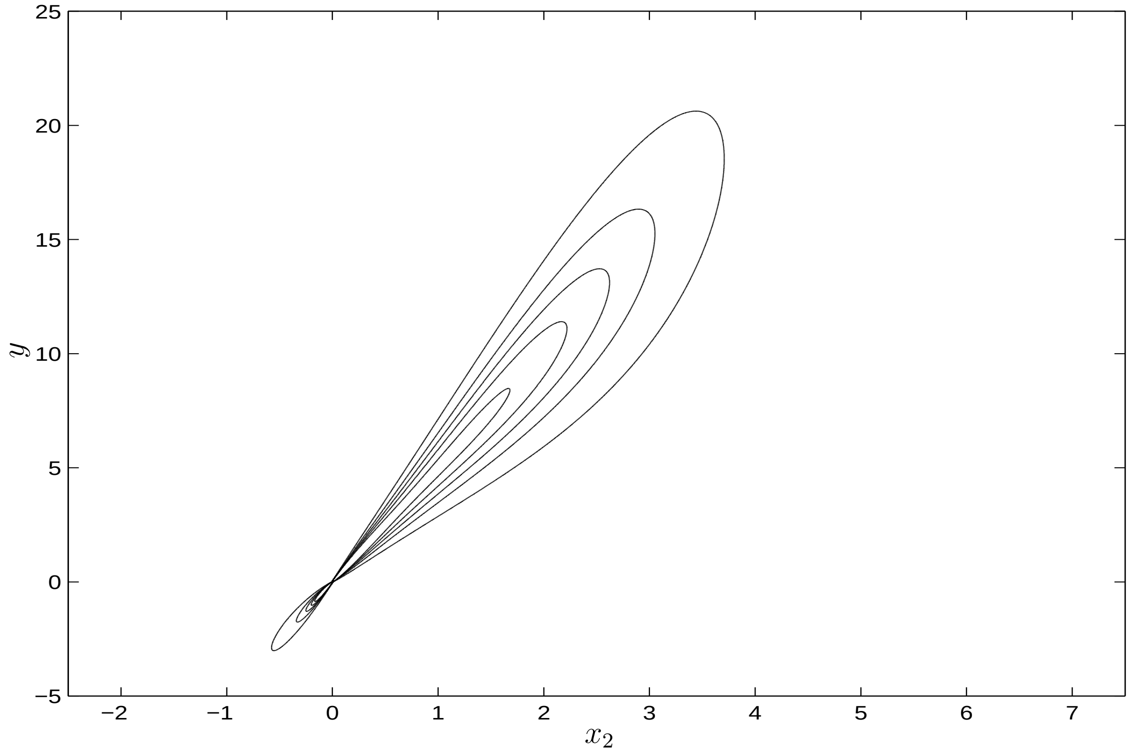

Figure 7: Contours for the joint density $f(y, x_2)$ displayed in two dimensions for Case 2. Contours are plotted at $0.01, 0.03, 0.05, 0.07, 0.09$.

each domain separately. By combining the two domains and normalising, we can find a numerical approximation of $f(y)$. Figure 8 shows that although $f(y)$ is non-symmetric and non-Normal, it is far less so than in the previous problem. It took 200 minutes to complete the numerical integration with interval widths of 0.001.

When the numerical integration procedure was implemented on MATLAB results showed fluctuations near $y = 0$, $y \approx 4$, $y \approx 8$, $y \approx 12$, and $y \approx 17$. The fluctuations around $y = 0$ are shown in Figure 9, where $f(y)$ is evaluated at intervals of 10^{-5}. Intuition, as well as our own analysis, suggests that these fluctuations are produced due to the computational limitations of MATLAB and the discrete nature of the variable, rather than because they accurately represent the density $f(y)$. To test this speculation the same numerical integration was implemented on Maple. The Maple output for Case 2 is shown in Figure 10. The fluctuations did not occur in the Maple output. For practical purposes the fluctuations produced in the MATLAB output are irrelevant because the numerical results are so satisfactory. Moreover, as we shall soon see, they are quite accurate when compared to the crudeness of Monte

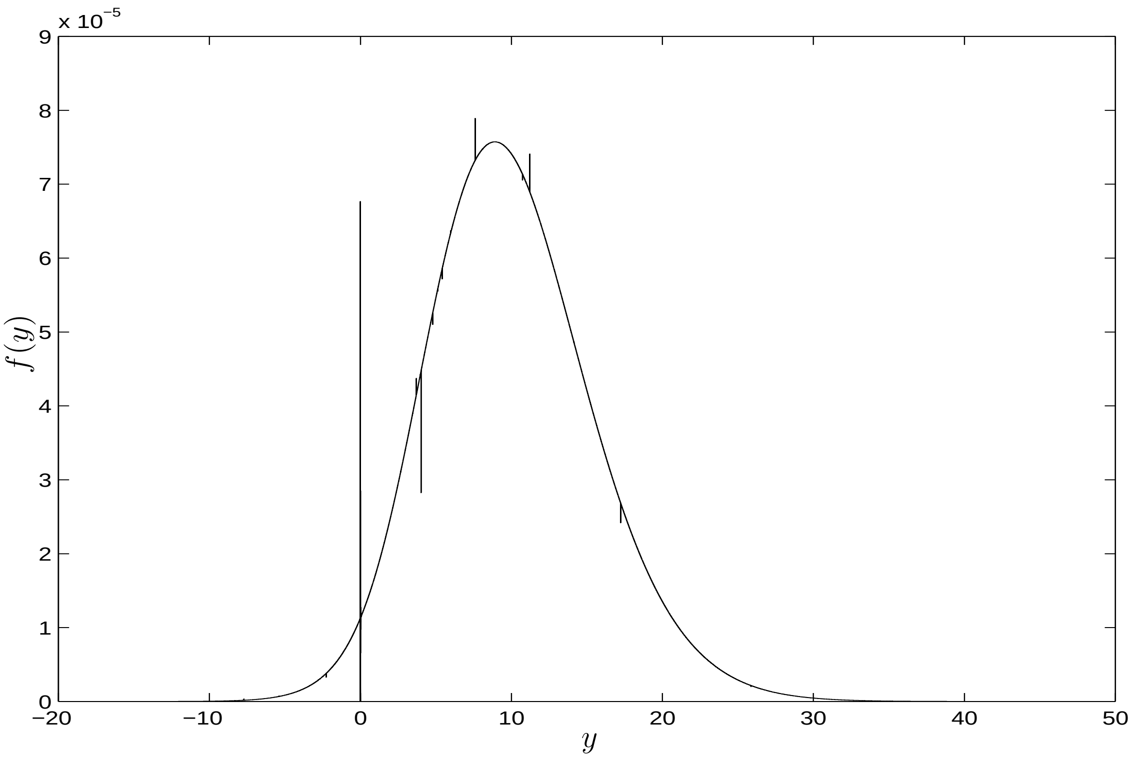

Figure 8: The marginal density $f(y)$ obtained via numerical integration for Case 2. The function was integrated between -15 and 43, with interval widths of 0.001. Notice the fluctuations that occur at irregular intervals along $f(y)$.

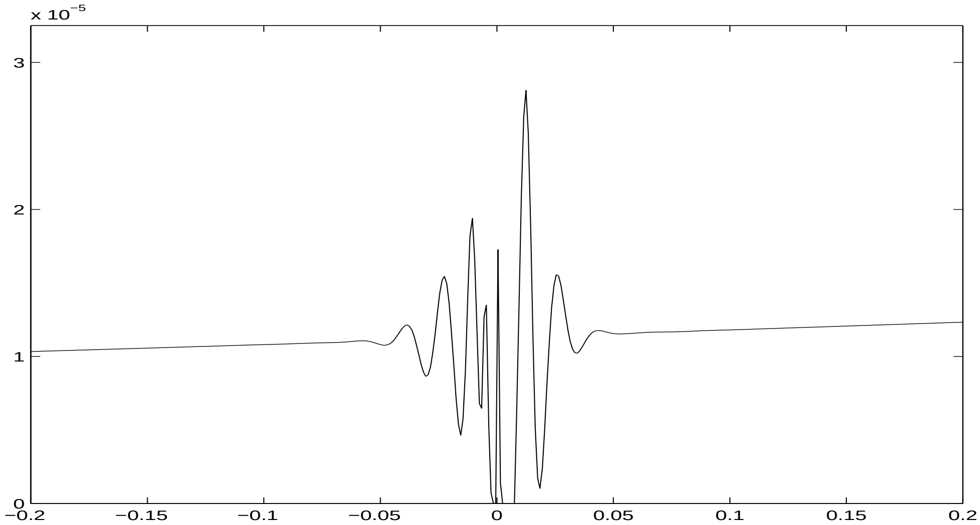

Figure 9: A close up view of $f(y)$ around $y = 0$ for Case 2. Notice the fluctuations near $y = 0$. The function was numerically integrated with interval widths of 10^{-5}.

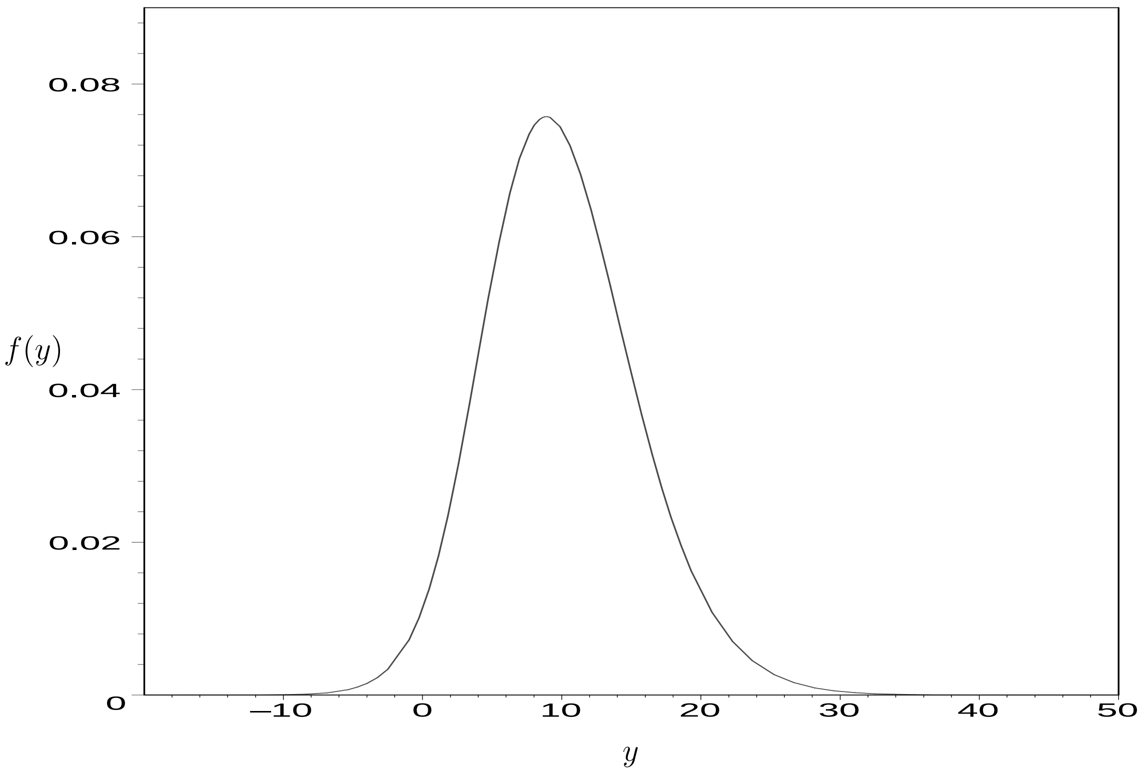

Figure 10: Numerically integrated density $f(y)$ produced using Maple for Case 2. Notice there is no irregular behaviour along $f(y)$.

Carlo methods.

3.1.3 An Open Question

While I have resolved the computation of the density for the product of two Normal variables in a way that can be applied to any bivariate Normal, there is an open question that I have not had time to resolve. At what (μ_1, μ_2) configurations does the density for Y peak at $y = 0$ and at what configurations is it away from 0? Is it always unimodal?

Rather than confront that question here, let us turn to a Monte Carlo resolution of the same specific problems we have already resolved.

3.2 A Monte Carlo Construction

A Monte Carlo method can be used to simulate $f(y)$. Generate two vectors, $\mathbf{X}_1$ and $\mathbf{X}_2$, each of length N, that consist of random variables drawn from $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$. A new vector, $\mathbf{Y}$, can be obtained by multiplying $\mathbf{X}_1$ and $\mathbf{X}_2$

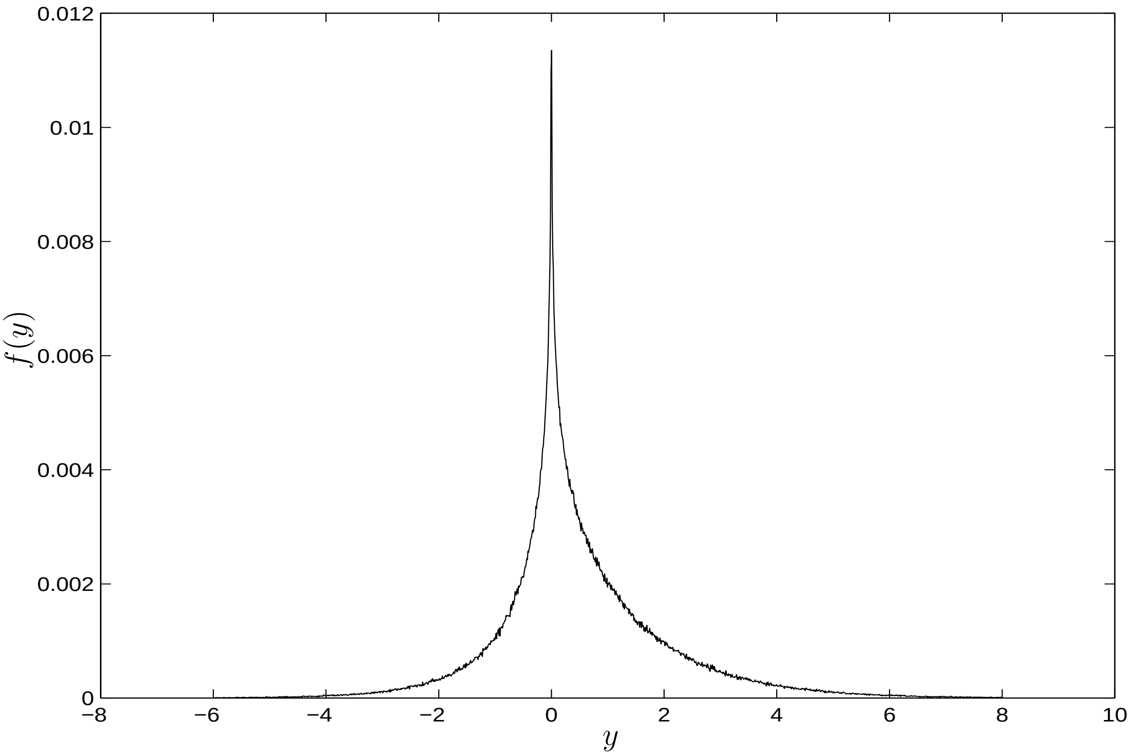

Figure 11: The marginal density, $f(y)$, approximated using a Monte Carlo simulation for Case 1. The number of elements drawn to form $\mathbf{X}_1$ and $\mathbf{X}_2$ was $1,000,000$. The bins have width 0.01

element-wise. The constituent elements of $\mathbf{Y}$ approximate a random sample from $f(y)$. To approximate the probability density function $f(y)$, sort the elements of $\mathbf{Y}$ into small, evenly spaced intervals and then normalise.

3.2.1 Case 1: $\mu_1 = 1, \mu_2 = 0.5, \sigma^2 = 1, cov(X_1, X_2) = 0$.

Two vectors of length $1,000,000$ were obtained. The first contained elements drawn from $X_1 \sim N(1,1)$ and the second contained elements from $X_2 \sim N(0.5,1)$. The vectors were element-wise multiplied to form $\mathbf{Y}$, a vector of length $1,000,000$. The elements of $\mathbf{Y}$ constituted a random sample from $f(y)$. The elements were sorted into intervals of width 0.01. The resulting histogram is shown in Figure 11. The time taken to generate the histogram in MATLAB was approximately 5 seconds. It can be seen that Figure 5 and Figure 11 are approximately the same shape and cover the same domain, but that the Monte Carlo approximation is cruder than the numerical integration.

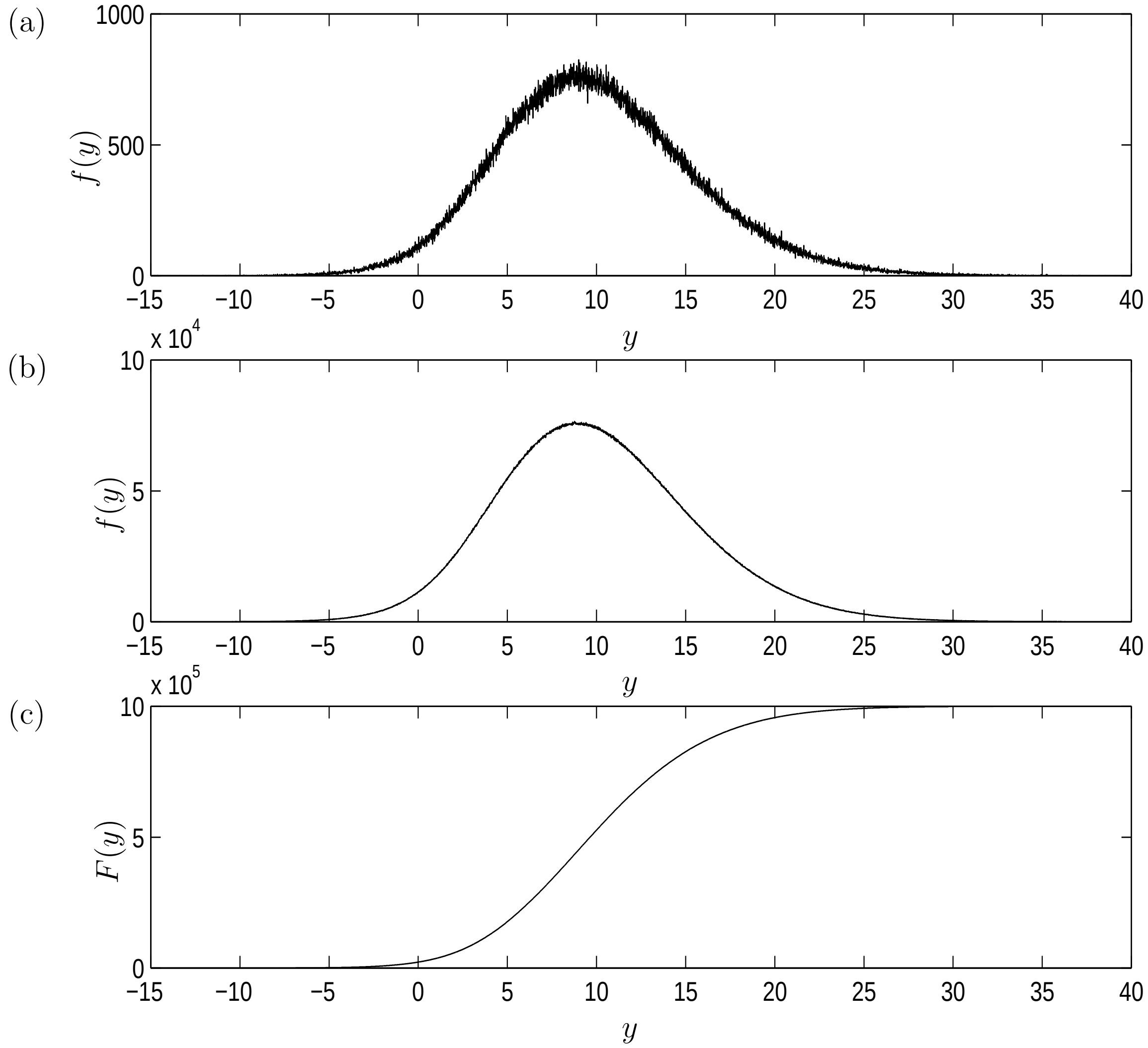

Figure 12: The marginal density, $f(y)$, and marginal cumulative density, $F(y)$, obtained by Monte Carlo simulation for Case 2. $1,000,000$ **((a), (c)) and** $100,000,000$ **(b) elements were sorted into bins of width** 0.01**.**

3.2.2 Case 2: $\mu_1 = 5, \mu_2 = 2, \sigma^2 = 1, cov(X_1, X_2) = 0$.

Two random vectors were generated and multiplied in a similar fashion to the previous subsection. The length of each vector was 1,000,000. The random samples were drawn from $X_1 \sim N(5,1)$ and $X_2 \sim N(2,1)$. After element-wise multiplication the elements of $\mathbf{Y}$ were sorted into intervals of width 0.01. A histogram representing $f(y)$ is shown in Figure 12(a). The largest negative non-zero value of $\mathbf{Y}$ was at $y = -16.65$, while the largest positive non-zero value was at $y = 47.74$.

Although the numerical integration and Monte Carlo method produce densities of similar shapes, it is clear the approximation displayed in Figure 12(a) is far cruder than that displayed in Figure 8.

The immediate difference between the two marginal densities obtained by a Monte Carlo method is that $f(y)$ has far higher variation in Case 2 than it does in Case 1. The number of simulated random variables and the widths of the bins that the values are sorted into remain the same from case to case, but in Case 2 the non-zero domain of Y is much larger. Consequently, in Case 2 there will be a smaller number of observations per interval width, and thus the variation between contiguous bins will be larger. A smoother Monte Carlo approximation can be obtained by increasing the number of elements drawn from X_1 and X_2 and/or increasing the bin width. To approximate $f(y)$ by a histogram containing $100,000,000$ observations took approximately 26 hours in MATLAB. The speed of the program was retarded by the size of the swap space on the computer system in the Department of Mathematics and Statistics, University of Canterbury, Christchurch. The resulting histogram is shown in Figure 12(b). Notice that the histogram comprising 100,000,000 observations is considerably smoother, but it still contains a surprising amount of fuzzy resolution compared to either Figure 8 or Figure 10, and these use bins that are ten times smaller!

Figure 12(c) shows the cumulative density function, $F(y)$, corresponding to the marginal density with $1,000,000$ elements. The cumulative density functions corresponding to $f(y)$ with $1,000,000$ and $100,000,000$ elements are more similar than their probability density functions are, as the fluctuations are 'evened out', e.g. compare the relative smoothness of Figure 12(a) and Figure 12(c).

3.3 An Approximation to the Analytic Result using the Normal Distribution

A third way to approximate $f(y)$ is by calculating the first two moments of Y, and then finding a distribution whose parameters match the moments of Y. We shall derive the moment-generating function for Y, and show that Y can be approximated by a Normal curve under certain conditions.

3.3.1 The Product of Two Correlated Normally Distributed Variables

Craig (1936) was the first to study the form of the distribution of the product of two Normally distributed variables. He investigated the product $Z = \frac{X_1 X_2}{\sigma_1 \sigma_2}$, parameterising his results in terms of ratios $\delta_1 = \mu_1/\sigma_1$, $\delta_2 = \mu_2/\sigma_2$ and ρ, the correlation coefficient, and determining the algebraic form of the moment-generating function. Aroian (1947) showed that Z is asymptotically Normal if either δ_1 or δ_2 or both approach infinity. Aroian et al. (1978) distinguished six different cases for Z depending on what is known about the parameters δ_1, δ_2 and ρ. The cases are: (1) $\delta_1 = 0$, δ_2, ρ; (2) $\delta_1 = \delta_2$, ρ; (3) $\delta_1 = \delta_2 = 0$, ρ; (4) $\delta_1 = \delta_2 = \delta$, $\rho = 1$; (5) $\rho = 0$, and (6) $\delta_1 \neq \delta_2$, $\rho = 1$.

Aroian et al. (1978) also proved that if $\delta_1 = \delta_2 = \delta$, then as $\delta \to \infty$, the standardised distribution of Z approaches the standardised Pearson type III distribution

$$f(z) = c\left(1 + \frac{\alpha_3 z}{2}\right)^{\left(-4/\alpha_3^2\right)-1} exp\left(-\frac{2z}{\alpha_3}\right), \qquad z \geq -\frac{2}{\alpha_3}, \tag{11}$$

where α_3 is the measure of skewness and

$$c = \left(\frac{4}{\alpha_3^2}\right)^{\left(4/\alpha_3^2\right)-\frac{1}{2}} \left[\Gamma\left(\frac{4}{\alpha_3^2}\right)\right]^{-1} exp\left(-\frac{4}{\alpha_3^2}\right).$$

That is, if $W = (Z - \mu_Z)/\sigma_Z$, then as μ/σ increases to infinity, $F(w)$ approaches the standard Pearson type III distribution with mean zero and standard deviation one. Cornwell et al. (1978) describe the numerical evaluation of Z.

3.3.2 The Moment-Generating Function of the Product of Two Correlated Normally Distributed Variables

Assume $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$ and that X_1 and X_2 have correlation ρ. Define $X_1 = X_0 + Z_1$ and $X_2 = X_0 + Z_2$, where

$$\begin{bmatrix} X_0 \\ Z_1 \\ Z_2 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \rho\sigma_1\sigma_2 & 0 & 0 \\ 0 & \sigma_1^2 - \rho\sigma_1\sigma_2 & 0 \\ 0 & 0 & \sigma_2^2 - \rho\sigma_1\sigma_2 \end{bmatrix}\right).$$

We have decomposed X_1 and X_2 into independent summands, one of which is shared between them.

To find the moment-generating function of $Y = X_1X_2 = (X_0+Z_1)(X_0+Z_2)$, we know that

$$
\begin{aligned}
M_Y(t) &= \int_{-\infty}^{\infty} e^{ty} f(y)dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{t(x_0+z_1)(x_0+z_2)} f(x_0,z_1,z_2)dx_0dz_1dz_2 \\
&= \frac{1}{\sqrt{2\pi\rho\sigma_1\sigma_2}} \frac{1}{\sqrt{2\pi\left(\sigma_1^2-\rho\sigma_1\sigma_2\right)}} \frac{1}{\sqrt{2\pi\left(\sigma_2^2-\rho\sigma_1\sigma_2\right)}} \\
&\quad \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{t(x_0+z_1)(x_0+z_2)-\frac{1}{2}\frac{(x_0-0)^2}{\rho\sigma_1\sigma_2}-\frac{1}{2}\frac{(z_1-\mu_1)^2}{\sigma_1^2-\rho\sigma_1\sigma_2}-\frac{1}{2}\frac{(z_2-\mu_2)^2}{\sigma_2^2-\rho\sigma_1\sigma_2}} dx_0dz_1dz_2 \\
&= \left(\sqrt{(1-\rho\sigma_1\sigma_2t)^2-\sigma_1^2\sigma_2^2t^2}\right)^{-1} e^{\frac{\mu_1\mu_2t+\frac{1}{2}\left(\mu_1^2\sigma_2^2+\mu_2^2\sigma_1^2-2\rho\mu_1\mu_2\sigma_1\sigma_2\right)t^2}{(1-\rho\sigma_1\sigma_2t)^2-\sigma_1^2\sigma_2^2t^2}}. \qquad (12)
\end{aligned}
$$

See Appendix A for complete exposition.

An MGF of this form was found by Craig (1936), who calculated $M_Z(t)$, where $Z = \frac{X_1X_2}{\sigma_1\sigma_2}$, to be

$$
M_Z(t) = \left(\sqrt{\left[1-(1+\rho)t\right]\left[1+(1-\rho)t\right]}\right)^{-1} e^{\frac{2\delta_1\delta_2t+\left(\delta_1^2+\delta_2^2-2\rho\delta_1\delta_2\right)t^2}{2[1-(1+\rho)t][1+(1-\rho)t]}}, \qquad (13)
$$

Note that this result is written only in terms of the ratios δ_1 and δ_2, which are proportional to the reciprocals of the coefficient of variation, and ρ, the correlation coefficient.

To confirm the equivalence of Equations 12 and 13, observe that

$$
\begin{aligned}
M_Z(t) &= M_{\frac{X_1X_2}{\sigma_1\sigma_2}}(t) \\
&= M_Y\left(\frac{t}{\sigma_1\sigma_2}\right). \qquad (14)
\end{aligned}
$$

By replacing every "t" in Equation 12 with "$\frac{t}{\sigma_1\sigma_2}$" it is soon seen that the two MGFs are identical.

3.3.3 Moments of the Product of Two Correlated Normally Distributed Variables

The moment-generating function can be used to find moments about the origin of Y. By differentiating $M_Y(t)$ and evaluating at $t = 0$ we can find as many moments as required. These moments can be used to calculate the mean, variance and skewness of the product of two correlated Normal variables. Using Maple we find that:

$$
\begin{aligned}
E(Y) &= \mu_1\mu_2 + \rho\sigma_1\sigma_2, &(15)\\
V(Y) &= \mu_1^2\sigma_2^2 + \mu_2^2\sigma_1^2 + \sigma_1^2\sigma_2^2 + 2\rho\mu_1\mu_2\sigma_1\sigma_2 + \rho^2\sigma_1^2\sigma_2^2, &(16)\\
\alpha_3(Y) &= \frac{6\sigma_1\sigma_2\left(\mu_1\mu_2\sigma_1\sigma_2\left(\rho^2+1\right) + \rho\left(\mu_1^2\sigma_2^2 + \mu_2^2\sigma_1^2\right)\right) + 2\rho\sigma_1^3\sigma_2^3\left(3+\rho^2\right)}{\left(\mu_1^2\sigma_2^2 + \mu_2^2\sigma_1^2 + \sigma_1^2\sigma_2^2\left(1+\rho^2\right) + 2\rho\mu_1\mu_2\sigma_1\sigma_2\right)^{3/2}}. &(17)
\end{aligned}
$$

Craig (1936) found equivalent moments, again using his $(\delta_1, \delta_2, \rho)$ parameterisation.

3.3.4 Special Cases

We shall investigate three special cases for the product of two correlated Normal variables. First, we shall examine what happens when $\rho = 0$. In this case the Normal variables are independent. Second, we shall consider $\mu_1 = \mu_2 = 0$, $\sigma_1 = \sigma_2 = 1$ and $\rho = 1$. Here we are considering the square of a standard Normal distribution. Conventional theory tells us that we can expect to obtain the MGF and moments of a Chi-Square distribution. Finally, we examine what happens as the ratio of μ/σ changes over different values of ρ. This is achieved by setting μ_1 and μ_2 to μ and holding σ_1 and σ_2 constant at 1. ρ varies between -1 and 1. Note that our three cases are equivalent to, in order, cases 5, 3 and 2 from Aroian et al. (1978).

Case I: $\rho = 0$

When $\rho = 0$, X_1 and X_2 are independent. The moment-generating function of Y can be written as

$$
M_Y(t) = \left(\sqrt{1 - \sigma_1^2\sigma_2^2 t^2}\right)^{-1} e^{\frac{\mu_1\mu_2 t + \frac{1}{2}\left(\mu_1^2\sigma_2^2 + \mu_2^2\sigma_1^2\right)t^2}{1 - \sigma_1^2\sigma_2^2 t^2}}. \tag{18}
$$

The mean, variance and skewness are:

$$
\begin{aligned}
E(Y) &= \mu_1\mu_2, &(19)\\
V(Y) &= \mu_1^2\sigma_2^2 + \mu_2^2\sigma_1^2 + \sigma_1^2\sigma_2^2, &(20)\\
\alpha_3(Y) &= \frac{6\mu_1\mu_2\sigma_1^2\sigma_2^2}{\left(\mu_1^2\sigma_2^2 + \mu_2^2\sigma_1^2 + \sigma_1^2\sigma_2^2\right)^{3/2}}. &(21)
\end{aligned}
$$

The moments of Y can be found more quickly by observing that $E(Y^r) = E(X_1^r)E(X_2^r)$, and the moments of the Normal distribution are well-known. The values of $E(Y)$, $V(Y)$ and $\alpha_3(Y)$ obtained using this method are the same as the values produced using the moment-generating function; see Appendix B for details. The distribution of the product of two independent Normal variables corresponds to Cases 1 and 2 as discussed in the previous subsections, where we had $cov(X_1, X_2) = 0$, and as such we shall consider this problem in greater depth in Section 3.3.5.

Case II: $\mu_1 = \mu_2 = 0$**,** $\sigma_1 = \sigma_2 = 1$**,** $\rho = 1$

When two perfectly correlated standard Normal variables are multiplied together, the moment-generating function, mean, variance and skewness of their product, Y, can be written as:

$$
\begin{aligned}
M_Y(t) &= (1-2t)^{-\frac{1}{2}}, && (22)\\
E(Y) &= 1, && (23)\\
V(Y) &= 2, && (24)\\
\alpha_3(Y) &= 2\sqrt{2}. && (25)
\end{aligned}
$$

It is well known that if $U_1, U_2, \ldots, U_\nu$ are independent standard Normal variables, then $\sum_{i=1}^{\nu} U_i^2$ has a Chi-square distribution with ν degrees of freedom. Thus it is no surprise that the MGF and moments are identical to those from a Chi-Square distribution with one degree of freedom.

Note that when $\rho = -1$, we obtain

$$M_Y(t) = (1+2t)^{-\frac{1}{2}}, \tag{26}$$

the MGF of a "negative Chi-square" distribution with one degree of freedom.

Case III: $\mu_1 = \mu_2 = \mu$**,** $\sigma_1 = \sigma_2 = 1$

In this case the moment-generating function of Y can be written as

$$M_Y(t) = \left(\sqrt{(1-\rho t)^2 - t^2}\right)^{-1} e^{\frac{\mu^2 t(1+t-\rho t)}{(1-\rho t)^2 - t^2}}. \tag{27}$$

The expected value of Y is

$$E(Y) = \mu^2 + \rho. \tag{28}$$

Observe that for any distinct value of ρ, $E(Y)$ increases at an increasing rate as ratio μ/σ increases absolutely. However, for any specific value of μ/σ, $E(Y)$ increases linearly on $\rho \in [-1, 1]$. Both Case I and Case II occur as special cases in Figure 13.

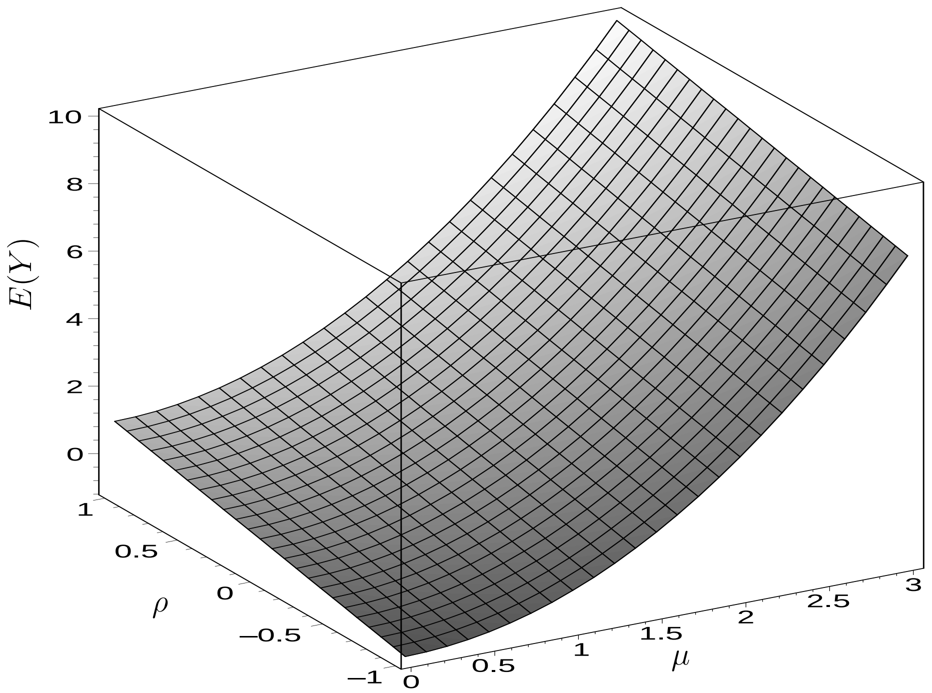

Figure 13: Expected value of Y for Case III.

When $\rho = 0$ we observe that the expected value of Y is μ^2, which is the value given in Equation 19 with $\mu_1 = \mu_2 = \mu$. Remember that for Case III we have set $\sigma^2 = 1$, so when $\mu = 0$ and $\rho = 1$ we have the Chi-square case. From Figure 13 we can see that this gives an expected value of 1. Note that if the graph were to include values of μ that are less than zero, the graph would be symmetric about $\mu = 0$. Note that for Case III we could write any "μ" or "μ/σ" term as "δ", but for the sake of clarity we shall retain "μ" for the remainder of this subsection.

The variance of Y is

$$V(Y) = 2\mu^2 (1 + \rho) + \rho^2 + 1. \tag{29}$$

Figure 14 shows how $V(Y)$ changes as μ/σ varies over different values of ρ. As ratio μ/σ increases the variance will get larger increasingly quickly, except for when $\rho = -1$, when this is the case the variance is always 2. If μ/σ is held constant, the variance increases as ρ increases; the larger μ/σ is specified to be, the faster $V(Y)$ increases. As with the expected value, if the graph was plotted over negative values of μ, it would be symmetric about $\mu = 0$.

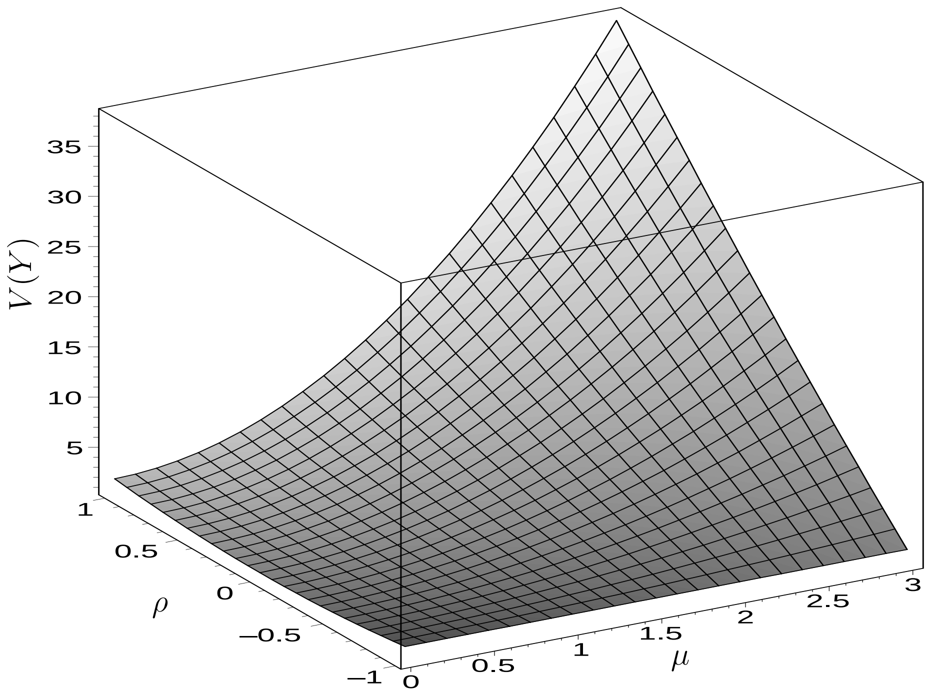

Figure 14: Variance of Y for Case III.

The skewness of Y is

$$\alpha_3(Y) = \frac{6\mu^2\left(\rho+1\right)^2 + 2\rho\left(3+\rho^2\right)}{\left(2\mu^2\left(\rho+1\right)+\rho^2+1\right)^{3/2}}. \tag{30}$$

Figure 15 shows how the skewness changes as μ/σ varies over different values of ρ. Observe that when $\rho = 0$, $\alpha_3(Y) = 0$ when $\mu/\sigma = 0$, but as the ratio increases the skewness rises rapidly, until it is at its maximum when $\mu/\sigma = 1$, as the ratio continues to increase $\alpha_3(Y)$ gradually decreases to 0. When $\mu = 0$ and $\rho = 1$ the skewness is $2\sqrt{2}$, and when $\mu = 0$ and $\rho = -1$ the skewness is $-2\sqrt{2}$. In general, $\alpha_3 \to 0$ as $\mu/\sigma \to \infty$. This is because in Equation 30 we have a μ^2 term over a μ^3 term. The closer $\mid \rho \mid$ is to one the slower the approach of $\alpha_3(Y)$ to 0. The sole exception to this limit is when $\rho = -1$, whenever this is the case the skewness is a constant $-2\sqrt{2}$. As with $E(Y)$ and $V(Y)$, if $\alpha_3(Y)$ had been plotted over negative values of μ it would be symmetric about $\mu = 0$.

3.3.5 The Product Of Two Independent Normally Distributed Variables

After investigating the distribution of the product of two correlated Normally distributed variables we now turn our attention back to Case I, where the variables are

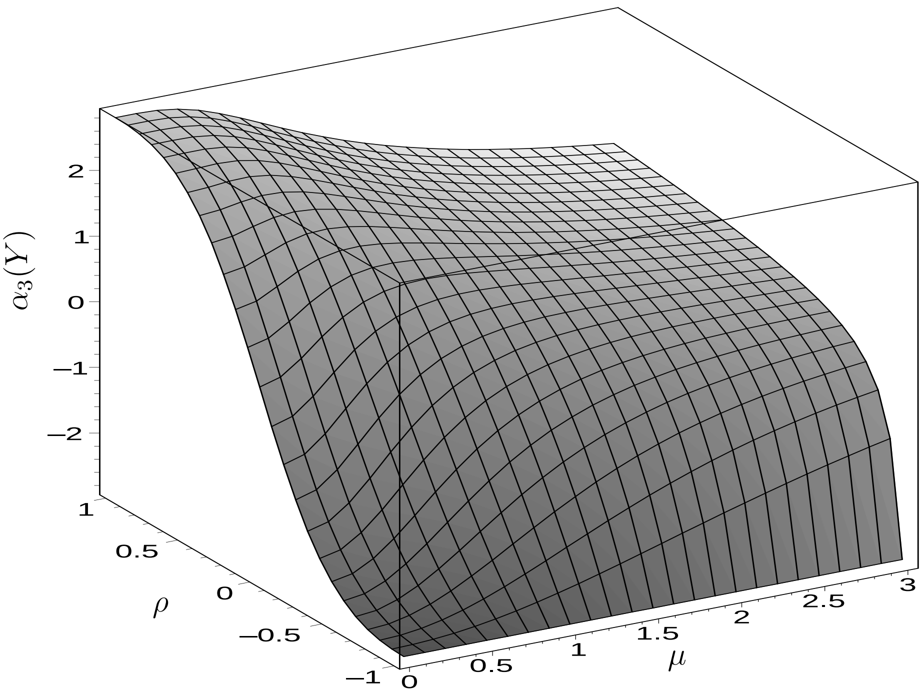

Figure 15: Skewness of Y for Case III.

independent. Remember that in both Case 1 and Case 2 of Sections 3.1 and 3.2 we had $cov(X_1, X_2) = 0$.

We know that the MGF of $W \sim N(\mu, \sigma^2)$ is

$$M_W(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}, \tag{31}$$

giving $E(W) = \mu$ and $V(W) = \sigma^2$. A study of Equation 18 shows that as $\delta_1 \to \infty$ and $\delta_2 \to \infty$, the coefficient of the exponential term will have a decreasing influence on $M_Y(t)$. In fact, as δ_1 and δ_2 increase,

$$M_Y(t) \to e^{\mu_1\mu_2 t + \frac{1}{2}\left(\mu_1^2\sigma_2^2 + \mu_2^2\sigma_1^2\right)t^2}. \tag{32}$$

We recognise that as δ_1 and δ_2 both increase without bound, $M_Y(t)$ converges to a Normal moment-generating function of Y with mean $\mu_1\mu_2$ and variance $\mu_1^2\sigma_2^2 + \mu_2^2\sigma_1^2$. In Equation 18 it was shown that the analytic result of the product of two Normal distributions is not a Normal distribution, however Equation 32 shows that the limit of $M_Y(t)$ is Normally distributed. In other words, the product of $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ tends towards a $N(\mu_1\mu_2, \mu_1^2\sigma_2^2 + \mu_2^2\sigma_1^2)$ distribution as δ_1 and δ_2 increase.

This knowledge allows us to investigate a third approximation method in our quest to assess the distribution for the product of two independent Normal variables. When δ_1 and δ_2 are large, the distribution of the product of two independent Normal variables will be approximately Normal. Since the Normal distribution is fully specified by its first two moments, we can calculate the mean and the variance of Y and approximate $f(y)$ by forming a Normal density with these parameters.

As we have already seen, the mean and variance for Case I are:

$$\begin{aligned} E(Y) &= \mu_1\mu_2, \\ V(Y) &= \sigma_1^2\sigma_2^2 + \mu_1^2\sigma_2^2 + \mu_2^2\sigma_1^2 \\ &= \sigma_1^2\sigma_2^2\left(1+\delta_1^2+\delta_2^2\right). \end{aligned} \tag{33}$$

The expression for $V(Y)$ obtained in Equation 33 differs from the variance of Y deduced from Equation 32. In Equation 33 an analytic result was calculated, while in Equation 32 we took the limit of $M_Y(t)$ as $\delta_1 \to \infty$ and $\delta_2 \to \infty$. It is clear that under these conditions the $\sigma_1^2\sigma_2^2$ term becomes negligible. In the two Cases we shall consider in this Section we assume that X_1 and X_2 have equal variance, that is, $\sigma^2 = \sigma_1^2 = \sigma_2^2$, so

$$\begin{aligned} V(Y) &= \sigma^2\left(\sigma^2+\mu_1^2+\mu_2^2\right) \\ &= \sigma^4\left(1+\delta_1^2+\delta_2^2\right). \end{aligned} \tag{34}$$

We can approximate the distribution of the product of two independent Normally distributed variables with a $N(\mu_1\mu_2, \sigma^4\left(1+\delta_1^2+\delta_2^2\right))$ distribution. The approximation will improve as δ_1 and δ_2 become large, since the limit of the distribution is $N(\mu_1\mu_2, \sigma^4\left(\delta_1^2+\delta_2^2\right))$.

Before moving on we shall briefly consider what happens to $M_Y(t)$ as δ_1 and δ_2 decrease to 0. Now the moment-generating function of Y is certainly not Normal. A study of $M_Y(t)$ shows us that as $\delta_1 \to 0$ and $\delta_2 \to 0$,

$$M_Y(t) \to \left(1-\sigma_1^2\sigma_2^2t^2\right)^{-1/2}. \tag{35}$$

A quick study of $M_Y(t)$ shows that all its odd moments are equal to zero. Consequently Y is symmetric about the origin. Also $V(Y) \to \sigma_1^2\sigma_2^2$ and the kurtosis of the limit of Y is 9.

3.3.6 The Skewness of the Analytic Result for Independent Variables

One way to predict the adequacy of the Normal approximation is by considering the skewness of Y from the analytic result. We know the skewness of Case I is

$$\begin{aligned} \alpha_3(Y) &= \frac{6\mu_1\mu_2\sigma_1^2\sigma_2^2}{(\sigma_1^2\sigma_2^2+\mu_1^2\sigma_2^2+\mu_2^2\sigma_1^2)^{3/2}} \\ &= \frac{6\delta_1\delta_2}{(\sigma_1\sigma_2)^{-3}(\sigma_1^2\sigma_2^2+\mu_1^2\sigma_2^2+\mu_2^2\sigma_1^2)^{3/2}}. \end{aligned} \tag{36}$$

The skewness of Y depends on the ratios δ_1 and δ_2. Clearly $\alpha_3(Y) \to 0$ as $\delta_1 \to \infty$ and $\delta_2 \to \infty$. The adequacy of the approximating Normal curve is related to the size of $\alpha_3(Y)$ for large δ_1, δ_2 values. As the skewness decreases the approximation improves. Note that $\alpha_3(Y)$ will always be skewed in the direction of the mean of Y, $\mu_1\mu_2$. The skewness will be largest when $\frac{d\alpha_3(Y)}{d\delta_1} = \frac{d\alpha_3(Y)}{d\delta_2} = 0$. Differentiating $\alpha_3(Y)$ with respect to δ_1 and δ_2, setting the two derivatives to zero and solving simultaneously we find that $\alpha_3(Y)$ has extreme points at $(\delta_1, \delta_2) = (\pm 1, \pm 1)$. By differentiating again we find that the maximum value of $\alpha_3(Y)$ occurs at $(-1, -1)$ and $(1, 1)$. The maximum value of $\alpha_3(Y)$ is 1.15 (2dp). The minimum value of $\alpha_3(Y)$, which is -1.15 (2dp), occurs at $(-1, 1)$ and $(1, -1)$. Thus the skewness of Y is largest when $\mu_1 = \sigma_1$ and $\mu_2 = \sigma_2$. When X_1 and X_2 have equal variance the skewness will be largest when $\mu_1 = \mu_2 = \sigma$.

The advantage of this method of approximation compared to the methods studied in Section 3.1 and Section 3.2 is that it is far quicker. Calculating the two numbers that represent μ and σ using straightforward formulae is simpler and faster than either numerically integrating a function over narrow interval widths, or running a Monte Carlo simulation a large number of times.

3.3.7 Case 1: $\mu_1 = 1, \mu_2 = 0.5, \sigma^2 = 1, cov(X_1, X_2) = 0$.

In this Case $\delta_1 = 1$, $\delta_2 = 0.5$ and $\alpha_3(Y) = 0.\dot{8}$. The small size of δ_1 and δ_2 result in $f(y)$ being significantly skewed. This suggests that it is not appropriate to approximate $f(y)$ with a Normal density for Case 1. This suspicion is reinforced by a study of Sections 3.1.1 and 3.2.1, where it is demonstrated that the density of Y is clearly not Normal. In fact, since $\alpha_3(Y) = 0.\dot{8}$ and $\mu_1 = \sigma_1$, this is one of the least suitable (δ_1, δ_2) configurations for us to approximate with a Normal distribution.

The parameters of our Normal approximation to the analytic result are:

$$\begin{aligned} \mu &= \mu_1\mu_2 \\ &= 0.5 \qquad (37) \\ \text{and} \qquad \sigma^2 &= \sigma^2\left(\mu_1^2+\mu_2^2+\sigma^2\right) \\ &= 2.25. \qquad (38) \end{aligned}$$

To approximate $f(y)$ we could use a $N(0.5, 2.25)$ density but in this case that approximation would clearly be inappropriate.

3.3.8 Case 2: $\mu_1 = 5, \mu_2 = 2, \sigma^2 = 1, cov(X_1, X_2) = 0.$

In Case 2 we have $\delta_1 = 5$, $\delta_2 = 2$ and $\alpha_3(Y) = 0.37$ (2dp). The skewness here is less than the skewness in Case 1 so, although the approximation to $f(y)$ is non-Normal, it is far closer to being Normal than in Case 1. The parameters of the approximating Normal distribution are $\mu = 10$ and $\sigma^2 = 30$. To approximate $f(y)$ we use a $N(10, 30)$ density. This is shown by the green line in Figure 16.

3.3.9 An Open Question

In this Subsection it has been shown that the limiting distribution of the product of two independent Normal distributions is also a Normal distribution. However the question remains: When is the ratio of mean to variance large enough for $f(y)$ to be adequately approximated by a Normal density? Is it the individual ratios μ_1/σ and μ_2/σ that determine the adequacy of the Normal approximation to $f(y)$, or is the combined ratio $\mu_1\mu_2/\sigma^2$ more important? Is there a critical value of $\alpha_3(Y)$ below which a Normal approximation is justified when ratios μ_1/σ and μ_2/σ are large?

Rather than confront those questions here, we shall put the Normal approximation aside until Section 5.

3.4 Comparison of Approximation Methods

The distribution of the product of two independent Normal densities has been approximated using three different methods. In Section 3.1, Section 3.2 and Section 3.3

we approximated the density of Y by numerical integration, by a Monte Carlo construction and via a Normal distribution. These three approximation methods were studied for the Cases where (μ_1, μ_2) was equal to $(1, 0.5)$, Case 1, and $(5, 2)$, Case 2. For each Case we set $\sigma^2 = 1$.

In Case 1 the approximations of $f(y)$ produced by numerical integration and Monte Carlo simulation are of the same shape and cover similar domains — compare Figure 5 and Figure 11. Remember that the numerical integration has interval widths of 0.01 and for the Monte Carlo simulation $N = 1,000,000$. The major difference between the two methods is that the Monte Carlo approximation is far cruder than the approximation attained through numerical integration. The Normal approximation, a $N(0.5, 2.25)$ curve, is clearly a different shape to the other two $f(y)$ approximations. The shape difference of the Normal approximation is not surprising considering the high value of $\alpha_3(Y)$.

In Case 2 the similarity of the approximations obtained via numerical integration, blue, and Monte Carlo simulation, red, can be seen in Figure 16. The numerical integration has interval widths of 0.001 and for the Monte Carlo simulation $N = 100,000,000$. Notice how much more accurate the numerical integration is. The Normal approximation, green, is much closer to the other two approximations than it was in Case 1. We expect that as μ_1 and μ_2 increase relative to σ^2 the approximations of $f(y)$ obtained via numerical integration and Monte Carlo methods will become increasingly similar to a $N(\mu_1\mu_2, \sigma^2(\mu_{X_1}^2 + \mu_{X_2}^2 + \sigma^2))$ density.

Table 1 shows that in both Cases the mean and variance calculated from $f(y)$ are very similar. This is true for Case 1 even though the shape of the Normal approximation is very different to the shape of the other two approximations. In both Cases the skewness of the numerical integration and Monte Carlo approximations is very close to the exact known skewness of the product distribution. The skewness of the Normal approximation will always be zero, suggesting that in both of these two Cases it is a comparatively poor method of approximation.

In Case 1 there is more difference between the three different mean and variance estimates than there is in Case 2. If the numerical integration takes place using interval widths of 0.001 rather than 0.01, the approximate values of $E(Y)$ and $V(Y)$ are 0.5000 and 2.2499 respectively. If the Monte Carlo simulation takes place with

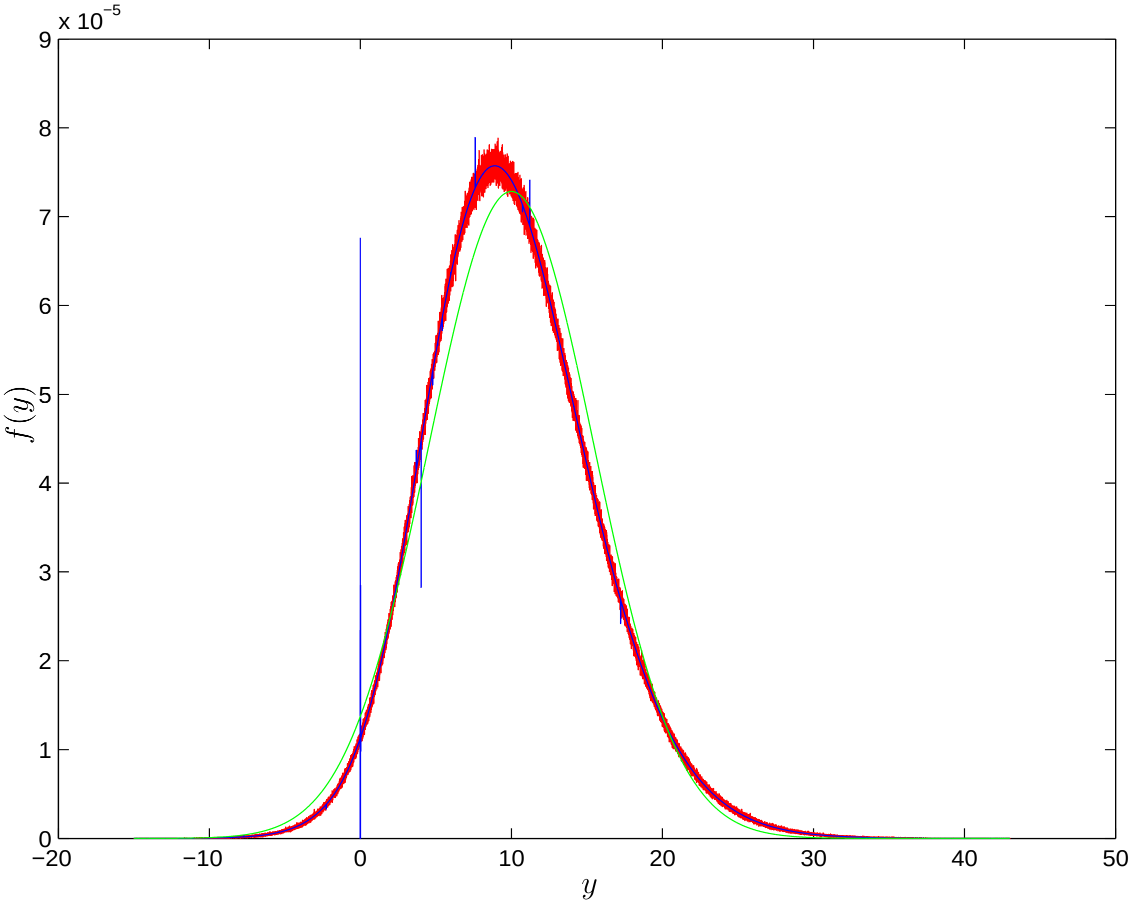

Figure 16: A comparison of the three different approximations of $f(y)$ for Case 2. The Monte Carlo (red) and numerical integration (blue) approximations have interval widths of 0.001. The Normal approximation to the analytic result is green.

a larger value of N, say with $100,000,000$ simulations rather than $1,000,000$, then $E(Y) = 0.4999$ and $V(Y) = 2.2500$.

Case	Method of Approximation	Mean	Variance	Skewness
Case 1	Numerical Integration	0.5004	2.2522	0.8894
	Monte Carlo	0.4978	2.2418	0.8941
	Normal	0.5	2.25	0
	Exact Moment Values	0.5	2.25	0.8889
Case 2	Numerical Integration	10.0000	29.9943	0.3649
	Monte Carlo	9.9992	29.9952	0.3653
	Normal	10	30	0
	Exact Moment Values	10	30	0.3651

Table 1: A comparison of the mean and variance of the three approximate distributions. The exact known values of the first three moments of the product distribution are included for comparison. The Normal approximation uses the exact known values of the first two moments of the product distribution. Values are recorded to 4dp.

4 Sums of Products of Two Normal Variables

To calculate the error metric, $M_e(\mathbf{s}, \hat{\mathbf{s}})$, we need to investigate the density for a sum of several products of Normal variables. In Section 3 we found how to obtain the density for the product of two Normal variables. In this section we shall consider how to find the density for the sum of several products. One way is to use the theory of convolutions.

4.1 Convolutions of Numerical Integration

The theory of convolutions is one method that can be used to find the distribution of a sum of random variables. If we have two random variables, Y_1 and Y_2, whose joint density is given by $f(y_1, y_2)$, we can write the joint density of $Z = Y_1 + Y_2$ and Y_2 as

$$f(z, y_2) = f(y_1, y_2) \left| \frac{dy_1}{dz} \right|. \tag{39}$$

Since

$$\left| \frac{dy_1}{dz} \right| = 1, \tag{40}$$

Equation 39 becomes

$$f(z, y_2) = f(z - y_2, y_2). \tag{41}$$

Summing over y_2 gives

$$f(z) = \sum_{y_2} f(z - y_2, y_2). \tag{42}$$

Formula 42 is called the Convolution Formula. In the case of independent variables, as we study here, Equation 42 reduces to

$$f(z) = \sum_{y_2} f(z - y_2) f(y_2). \tag{43}$$

In practical terms, the Convolution Formula involves obtaining vectors which contain a grid of possible values of Y_1 and Y_2 along with the density valuations at each respective point. Computationally the procedure for implementing this convolution process is as follows,

1. Select a value of Z.

2. Find pairs of (y_1, y_2) that sum to z.

3. Evaluate $f(\mathbf{y}_1)$ and $f(\mathbf{y}_2)$ over the vectors $\mathbf{y}_1$ and $\mathbf{y}_2$ that support the sum.

4. Flip the vector $f(\mathbf{y}_2)$, so the orders of $\mathbf{y}_1$ and $\mathbf{y}_2^F$ are appropriate to generate the sum in each component.

5. Find the component product of $f(\mathbf{y}_1)$ and the flipped $f(\mathbf{y}_2)$.

6. Sum the product vector.

Once this process is repeated for all possible values of Z, the vector containing $f(z)$ values is normalised. The resulting mass function is an approximation of the density.

4.1.1 Case 1: $\mu_1 = 1, \mu_2 = 0.5, \sigma^2 = 1, cov(X_1, X_2) = 0$.

In Section 3.1.1 we used numerical integration to obtain an approximation of the marginal density $f(y)$, where $X_1 \sim N(1, 1)$, $X_2 \sim N(0.5, 1)$ and $cov(X_1, X_2) = 0$. The resultant density was shown in Figure 5. The problem we address here is how to determine the density for the sum of N independent generations of such a variable Y. This is achieved sequentially by first convoluting $f(y)$ with itself, and then

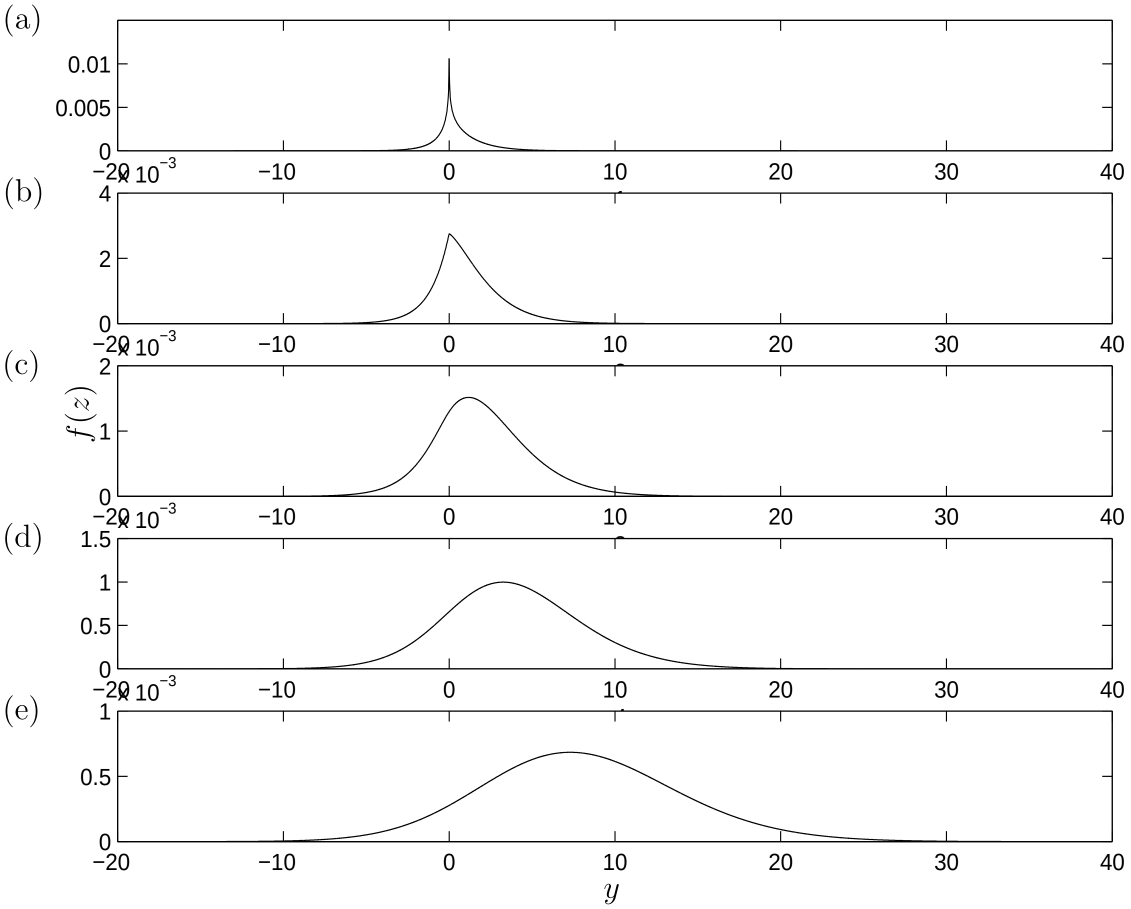

Figure 17: A sequence of convolutions based on the numerical integration obtained in Section 3.1. The panels represent $f_{Z_i}(z)$ for $i = 1, \ldots, 5$.

convoluting $f(y)$ with the achieved density for the convoluted sum. The density $f(y)$ can be convoluted either with itself or with another density any number of times. To display quickly how the distribution of the sum converges to a Normal we shall convolute the sum densities with themselves a few times.

We shall convolute $Y_i \sim f_{Y_i}(y)$ with itself. The densities $f_{Z_j}(z) = f_{\left[\sum_{i=1}^{2^{j-1}} Y_i\right]}(z)$, $j = 1, \ldots, 5$, obtained through a series of convolutions are demonstrated in Figure 17. The top panel displays the unadulterated $f_{Z_1}(z) = f_Y(y)$ density for comparison. Note the density featured in the top panel is the density displayed in Figure 5. The second panel displays $f_{Z_2}(z)$. The density still retains a maximum at 0, but the curve is noticeably less 'peaked'. The third panel demonstrates $f_{Z_3}(z) = f_{\sum_{i=1}^{4} Y_i}(y)$. The density now has a mode greater than 0, and although it is still slightly skewed, it looks much more like a Normal density. The fourth and

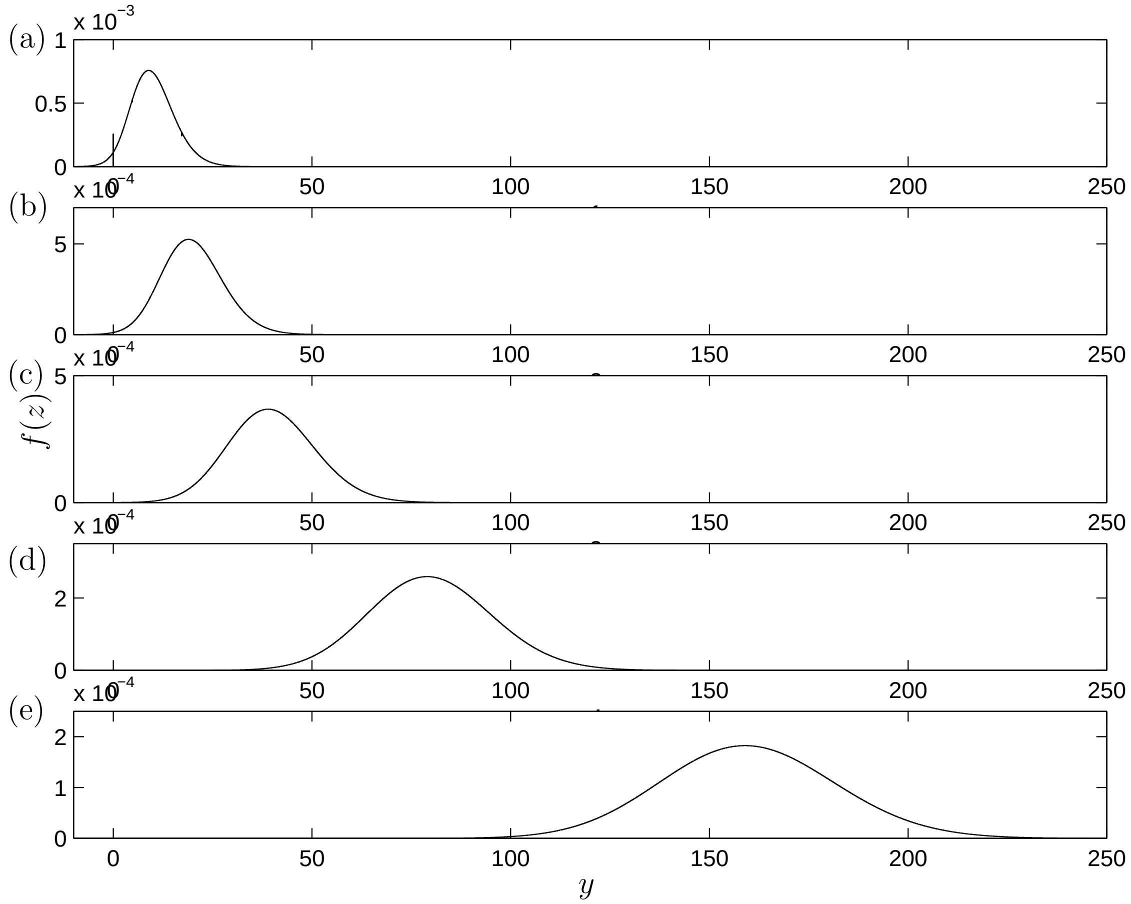

Figure 18: A sequence of convolutions based on the numerical integration obtained in Section 3.1. The panels represent $f_{Z_i}(z)$ for $i = 1, \dots, 5$.

fifth panels show $f_{Z_4}(z) = f_{\sum_{i=1}^{8} Y_i}(y)$ and $f_{Z_5}(z) = f_{\sum_{i=1}^{16} Y_i}(y)$. We can see that as the number of summed variables increases, $f_Z(z)$ approaches normality, even though the original $f_Y(y)$ is far from a Normal density. Of course such is expected from the Central Limit Theorem. However the convolution procedure allows us to compute exact distribution values for non-Normal distributions of sums of fewer variables. The only non-Normal parts of $f_{Z_5}(z)$ are the extreme tails, in the top and bottom 2%.

4.1.2 Case 2: $\mu_1 = 5, \mu_2 = 2, \sigma^2 = 1, cov(X_1, X_2) = 0$.

In Section 3.1.1 we used a numerical integration procedure to attain an approximation of the marginal density $f(y)$, where $X_1 \sim N(5, 1)$, $X_2 \sim N(2, 1)$ and $cov(X_1, X_2) = 0$. The density is displayed in Figure 8. Figure 18 displays a series of convolutions $f_{Z_j}(z)$, $j = 1, \ldots, 5$. Note that $f_{Z_1}(z) = f(y)$. The density in the top panel is the approximation of the marginal density which was attained in Section 3.1.2. As with Case 1, which was described in the previous Subsection, $f(y)$ is convoluted with itself $1, 2, 4, 8$ and 16 times. Once again, although $f(y)$ is clearly non-Normal, $f_{Z_j}(z)$ quickly approaches normality as j increases.

5 Calculation of the Error Metric

The problem that instigated this research involves finding the probability that the real component of the error metric between transmitted and hypothesised signals is negative. See Section 2 for full details. The form of the error metric that we shall concern ourselves with for the remainder of this Report is

$$M_e(s, \hat{s}) = \sum_{t=1}^{N\xi} \left(y_{R,t} y_{R,t-\xi} b_{R,t} + y_{R,t} y_{I,t-\xi} b_{I,t} - y_{I,t} y_{R,t-\xi} b_{I,t} + y_{I,t} y_{I,t-\xi} b_{R,t} \right). \tag{44}$$

Calculation of $M_e(s, \hat{s})$ involves computing the density for the sum of four or more products of variables, each assessed as having a Normal distribution.

We investigate three procedures to approximate the error metric. One way to obtain $M_e(s, \hat{s})$ is by computing the densities of products of independent Normally distributed variables via numerical integration and then using the Convolution Formula to identify the density of the sum of a series of these products, that is, by combining the processes introduced in Sections 3.1 and 4. Another way to approximate $M_e(s, \hat{s})$ is to implement a Monte Carlo simulation. A third way to approximate the distribution of the error metric is with a Normal distribution whose parameters are set to the mean and variance of $M_e(s, \hat{s})$. This approximation is only appropriate when $\alpha_3(Y)$ is small.

5.1 Properties of the Error Metric

5.1.1 Correlation of Products of Normal Variables

In Section 4 we studied the distribution of sums of products of independent Normal variables, where each product is uncorrelated with any other product. However, a study of Equation 44 shows us that when calculating the distribution of the error metric, some of the products of Normal variables will be correlated. In fact, if we assess

$$\begin{bmatrix} Y_0 \\ Y_1 \\ Y_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_0 \\ \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & 0 & 0 \\ 0 & \sigma_1^2 & 0 \\ 0 & 0 & \sigma_2^2 \end{bmatrix}\right),$$

then the covariance of Y_0Y_1 and Y_1Y_2 is

$$\begin{aligned}
cov(Y_0Y_1, Y_1Y_2) &= E[(Y_0Y_1 - \mu_0\mu_1)(Y_1Y_2 - \mu_1\mu_2)] \\
&= E(Y_0Y_1^2Y_2) - \mu_0\mu_1E(Y_1Y_2) - \mu_1\mu_2E(Y_0Y_1) + \mu_0\mu_1^2\mu_2 \\
&= \int_{-\infty}^{\infty} y_0f(y_0)dy_0 \int_{-\infty}^{\infty} y_1^2f(y_1)dy_1 \int_{-\infty}^{\infty} y_2f(y_2)dy_2 \\
&\quad -\mu_0\mu_1 \int_{-\infty}^{\infty} y_1f(y_1)dy_1 \int_{-\infty}^{\infty} y_2f(y_2)dy_2 \\
&\quad -\mu_1\mu_2 \int_{-\infty}^{\infty} y_0f(y_0)dy_0 \int_{-\infty}^{\infty} y_1f(y_1)dy_1 + \mu_0\mu_1^2\mu_2 \\
&= \mu_0\left(\mu_1^2 + \sigma_1^2\right)\mu_2 + \mu_0\mu_1^2\mu_2 - \mu_0\mu_1^2\mu_2 - \mu_0\mu_1^2\mu_2 \\
&= \sigma_1^2\mu_0\mu_2, \qquad (45)
\end{aligned}$$

which is non-zero when both $\mu_0 \neq 0$ and $\mu_2 \neq 0$.

5.1.2 Mean and Variance of the Error Metric

Standard distribution theory tells us that for random variables X_0 and X_1, $E(X_0 + X_1) = E(X_0) + E(X_1)$ and $V(X_0 + X_1) = V(X_0) + V(X_1) + 2cov(X_0, X_1)$. Since, from Section 3.3, $E(Y_tY_{t+1}) = \mu_t\mu_{t+1}$ and $V(Y_tY_{t+1}) = \sigma_1^2\sigma_2^2 + \mu_1^2\sigma_2^2 + \mu_2^2\sigma_1^2$, we have

$$\begin{aligned}
E\left(M_e(s,\hat{s})\right) &= E\left(\sum_{t=1}^{N\xi}(y_{R,t}y_{R,t-\xi}b_{R,t} + y_{R,t}y_{I,t-\xi}b_{I,t} - y_{I,t}y_{R,t-\xi}b_{I,t} + y_{I,t}y_{I,t-\xi}b_{R,t})\right) \\
&= \sum_{i=1}^{N\xi}(b_{R,t}E(Y_{R,t}Y_{R,t-\xi}) + b_{I,t}E(Y_{R,t}Y_{I,t-\xi}) - b_{I,t}E(Y_{R,t-\xi}Y_{I,t}) \\
&\quad +b_{R,t}E(Y_{I,t}Y_{I,t-\xi})) \\
&= \sum_{t=1}^{N\xi}(\mu_{R,t}\mu_{R,t-\xi}b_{R,t} + \mu_{R,t}\mu_{I,t-\xi}b_{I,t} - \mu_{I,t}\mu_{R,t-\xi}b_{I,t} + \mu_{I,t}\mu_{I,t-\xi}b_{R,t}) \qquad (46)
\end{aligned}$$

and

$$V\left(M_e(s,\hat{s})\right) = V\left(\sum_{t=1}^{N\xi}(y_{R,t}y_{R,t-\xi}b_{R,t} + y_{R,t}y_{I,t-\xi}b_{I,t} - y_{I,t}y_{R,t-\xi}b_{I,t} + y_{I,t}y_{I,t-\xi}b_{R,t})\right)$$

$$
\begin{aligned}
&= \sum_{t=1}^{N\xi} \Big(b_{R,t}^2 V(Y_{R,t}Y_{R,t-\xi}) + b_{I,t}^2 V(Y_{R,t}Y_{I,t-\xi}) \\
&\quad + b_{I,t}^2 V(Y_{I,t}Y_{R,t-\xi}) + b_{R,t}^2 V(Y_{I,t}Y_{I,t-\xi}) \\
&\quad + 2b_{R,t}b_{I,t}\left(cov(Y_{R,t}Y_{R,t-\xi}, Y_{I,t-\xi}Y_{R,t}) + cov(Y_{R,t}Y_{I,t-\xi}, Y_{I,t-\xi}Y_{I,t})\right. \\
&\quad \left. - cov(Y_{R,t}Y_{R,t-\xi}, Y_{R,t-\xi}Y_{I,t}) - cov(Y_{I,t}Y_{R,t-\xi}, Y_{I,t-\xi}Y_{I,t})\right) \\
&\quad + 2b_{R,t}b_{R,t+\xi}(cov(Y_{R,t}Y_{R,t-\xi}, Y_{R,t+\xi}Y_{R,t}) + cov(Y_{I,t}Y_{I,t-\xi}, Y_{I,t+\xi}Y_{I,t})) \\
&\quad + 2b_{R,t}b_{I,t+\xi}(cov(Y_{I,t}Y_{I,t-\xi}, Y_{I,t}Y_{R,t+\xi}) - cov(Y_{R,t}Y_{R,t-\xi}, Y_{R,t}Y_{I,t+\xi})) \\
&\quad + 2b_{I,t}b_{R,t+\xi}(cov(Y_{R,t}Y_{I,t-\xi}, Y_{R,t}Y_{R,t+\xi}) - cov(Y_{R,t-\xi}Y_{I,t}, Y_{I,t}Y_{I,t+\xi})) \\
&\quad + 2b_{I,t}b_{I,t+\xi}(-cov(Y_{I,t-\xi}Y_{R,t}, Y_{R,t}Y_{I,t+\xi}) - cov(Y_{R,t-\xi}Y_{I,t}, Y_{I,t}Y_{R,t+\xi}))\Big) \\
&= \sum_{t=1}^{N\xi} \Big(b_{R,t}^2 \left(\sigma_{R,t-\xi}^2\sigma_{R,t}^2 + \mu_{R,t-\xi}^2\sigma_{R,t}^2 + \mu_{R,t}^2\sigma_{R,t-\xi}^2 + \sigma_{I,t-\xi}^2\sigma_{I,t}^2\right. \\
&\qquad \left. + \mu_{I,t-\xi}^2\sigma_{I,t}^2 + \mu_{I,t}^2\sigma_{I,t-\xi}^2\right) + b_{I,t}^2\left(\sigma_{R,t}^2\sigma_{I,t-\xi}^2 + \mu_{R,t}^2\sigma_{I,t-\xi}^2\right. \\
&\qquad \left. + \mu_{I,t-\xi}^2\sigma_{R,t}^2 + \sigma_{R,t-\xi}^2\sigma_{I,t}^2 + \mu_{I,t}^2\sigma_{R,t-\xi}^2 + \mu_{R,t-\xi}^2\sigma_{I,t}^2\right) \\
&\qquad + 2b_{R,t}b_{I,t}\left(\sigma_{R,t}^2\mu_{R,t-\xi}\mu_{I,t+\xi} + \sigma_{I,t-\xi}^2\mu_{R,t}\mu_{I,t}\right. \\
&\qquad \left. - \sigma_{R,t-\xi}^2\mu_{R,t}\mu_{I,t} - \sigma_{I,t}^2\mu_{R,t-\xi}\mu_{I,t-\xi}\right) \\
&\qquad + 2b_{R,t}b_{R,t+\xi}\left(\sigma_{R,t}^2\mu_{R,t-\xi}\mu_{R,t+\xi} + \sigma_{I,t}^2\mu_{I,t-\xi}\mu_{I,t+\xi}\right) \\
&\qquad + 2b_{R,t}b_{I,t+\xi}\left(\sigma_{I,t}^2\mu_{I,t-\xi}\mu_{R,t+\xi} - \sigma_{R,t}^2\mu_{R,t-\xi}\mu_{I,t+\xi}\right) \\
&\qquad + 2b_{I,t}b_{R,t+\xi}\left(\sigma_{R,t}^2\mu_{I,t-\xi}\mu_{R,t+\xi} - \sigma_{I,t}^2\mu_{R,t-\xi}\mu_{I,t+\xi}\right) \\
&\qquad + 2b_{I,t}b_{I,t+\xi}\left(-\sigma_{R,t}^2\mu_{I,t-\xi}\mu_{I,t+\xi} - \sigma_{I,t}^2\mu_{R,t-\xi}\mu_{R,t+\xi}\right)\Big). \qquad (47)
\end{aligned}
$$

When $\mu_{R,t} = \mu_{I,t} = \mu$, $b_{R,t} = b_{I,t} = 1$ and $\sigma_{R,t}^2 = \sigma_{I,t}^2 = \sigma^2$ for all t, the expectation and variance of the error metric reduce to

$$
\begin{aligned}
E\left(M_e(s,\hat{s})\right) &= 2N\xi\mu^2 \qquad (48) \\
\text{and} \quad V\left(M_e(s,\hat{s})\right) &= 4N\xi\sigma^2\left(\sigma^2 + 2\mu^2\right) \qquad (49)
\end{aligned}
$$

5.2 Calculation of the Error Metric using Numerical Integration and Convolutions

To use numerical integration and convolutions to identify the density of the error metric via the processes discussed in the preceding Sections, each product must be uncorrelated with any other product. A study of Equation 44 shows that we cannot rearrange $M_e(s, \hat{s})$ so that this will be the case. However, an inspection of Equations 48 and 49 shows that, whether we are summing independent or correlated variables, the expectation and variance of the error metric will be the same. Simulations suggest that this relationship does not hold for any higher moments. Thus, although we cannot expect the approximation of the error metric attained by using our numerical integration and convolution procedure to be as accurate as a Monte Carlo approximation, the two methods should produce densities that are reasonably similar. This is because, as $N\xi$ increases, both distributions will approach Normality (by the Central Limit Theorem) and will have the same mean and variance (by Equations 48 and 49).

A study of Equation 44 shows that to assess the distribution for the error metric, we first need to find the density of the product of two Normally distributed variables via numerical integration. Then we convolute $N\xi$ of these individual densities of products of Normal variables. Since each of our $Y_{R,t}$ and $Y_{I,t}$ terms is Normally distributed with mean μ and variance σ^2, we merely find the density for the product of two $N(\mu, \sigma^2)$ densities using numerical integration, and then use the Convolution Formula to find the density of the sum of $N\xi$ of these products. Finally, we calculate $P(M_e(s, \hat{s})) < 0$ by forming the cumulative sum of $M_e(s, \hat{s})$ and finding its value at 0.

5.3 Calculation of the Error Metric using a Monte Carlo Construction

A Monte Carlo construction can be used to simulate the error metric directly from Equation 44. We can approximate the distribution of the error metric by implementing Monte Carlo simulations on either MATLAB or WinBUGS. A value of $M_e(s, \hat{s})$ can be calculated for any set of received Y values. If a large number of sets of Y values are randomly drawn, and if $M_e(s, \hat{s})$ is calculated for each of the sets, then we have a large number of random samples drawn from the distribution of the error metric, our target distribution. An estimate of $P[M_e(s, \hat{s}) < 0]$ is obtained by counting the number of random samples which have taken on a negative value, and dividing by the total number of samples.

The density of the error metric is approximated by sorting the samples into intervals and normalising. Scott (1992) showed that the shape of a histogram estimate depends on how the bins are defined. The smaller the bin widths, the more accurate the estimates are. However, narrow bins tend to introduce extra variability. In Section 3.2 we discussed how finely inaccurate the Monte Carlo approximation is compared to the numerically integrated and convoluted approximation, viz. Figures 5 and 11. This is a common problem for researchers who are interested in studying the density of the target distribution (Hoti et al., 2002). One way to alleviate this is to 'smooth' the data using the average of M different histograms, all based on the same bin width, h, but using different equally spaced sideways shifts (Scott, 1985). When the number of histograms is increased to infinity the bin width of the averaged shifted histogram vanishes (since $h/M \to 0$ as $M \to \infty$) and the estimate becomes the kernel density estimator (KDE),

$$\hat{f}_{KDE}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \tag{50}$$

where $K(u)$ is a kernel smoothing function, n is the number of draws from the target density and $h > 0$ is a smoothing parameter. The smoothing parameter plays a role similar to that of the bin widths of the histogram. When comparing the approximated histogram with the KDE, we can view the KDE as the average of infinitely many histograms with bin width h.

MATLAB programs to implement kernel density estimates are available from the Rolf Nevanlinna Institute's internet site[3]. The smoothing function used in Examples 1 and 2 is $K(u) = (1 - |u|)_+$, where $(u)_+ = u$ if $u > 0$ and zero otherwise. This is known as the triangle kernel function. To compare the approximated histogram with the KDE, we can view the KDE as the average of infinitely many histograms with bin width h.

The density of the error metric can be directly simulated using WinBUGS. Y values are generated and the relevant statistics are directly obtained. If the variable under investigation is assessed as continuous, as it is in this case, WinBUGS automatically plots a smoothed KDE of the target density. The KDE is calculated using a default h value of 0.2.

[3] www.rni.helsinki.fi/∼fjh

5.4 Calculation of the Error Metric using Normal Approximations

If the values of $\alpha_3(Y)$ are sufficiently small we can assess the distribution of the error metric by approximating the product of two Normal distributions as another Normal distribution. This approximation will improve as $N\xi$ increases, since the Central Limit Theorem tells us that the distribution of the error metric will become increasingly Normal as the number of sums of products increases. In Section 5.1 we showed that, for our problem, the mean and variance of $M_e(s,\hat{s})$ are $2N\xi\mu^2$ and $4N\xi\sigma^2\left(\sigma^2+2\mu^2\right)$ respectively.

5.5 Example 1

To illustrate the discussion of the previous two subsections, and to show how to calculate $P[M_e(s,\hat{s})<0]$ in practice, let $N=2$, $\xi=1$, $b_{R,i}=b_{I,i}=[1 \quad 1]$ and

$$\begin{bmatrix} Y_{R,0} \\ Y_{I,0} \\ Y_{R,1} \\ Y_{I,1} \\ Y_{R,2} \\ Y_{I,2} \end{bmatrix} \sim N\left(\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}\right).$$

The form of the error metric is, from Equation 44,

$$\begin{aligned} M_e(s,\hat{s}) &= \sum_{t=1}^{2}\left(y_{R,t}y_{R,t-1}b_{R,t}+y_{R,t}y_{I,t-1}b_{I,t}-y_{I,t}y_{R,t-1}b_{I,t}+y_{I,t}y_{I,t-1}b_{R,t}\right) \\ &= y_{R,0}y_{R,1}+y_{R,1}y_{I,0}-y_{R,0}y_{I,1}+y_{I,0}y_{I,1}+y_{R,1}y_{R,2}+y_{R,2}y_{I,1} \\ &\quad -y_{R,1}y_{I,2}+y_{I,1}y_{I,2}. \end{aligned} \tag{51}$$

5.5.1 Example 1: Numerical Integration and Convolutions

We use numerical integration, as described in Section 3.1, to approximate the density of the product of two $N(1,1)$ distributions. Each product is integrated between limits of -8.0005 and 17.0005, with interval width 0.001.

To compute density $M_e(s,\hat{s})$, we sum the eight products of Normal variables using the Convolution Formula. $M_e(s,\hat{s})$ has values recorded on the interval

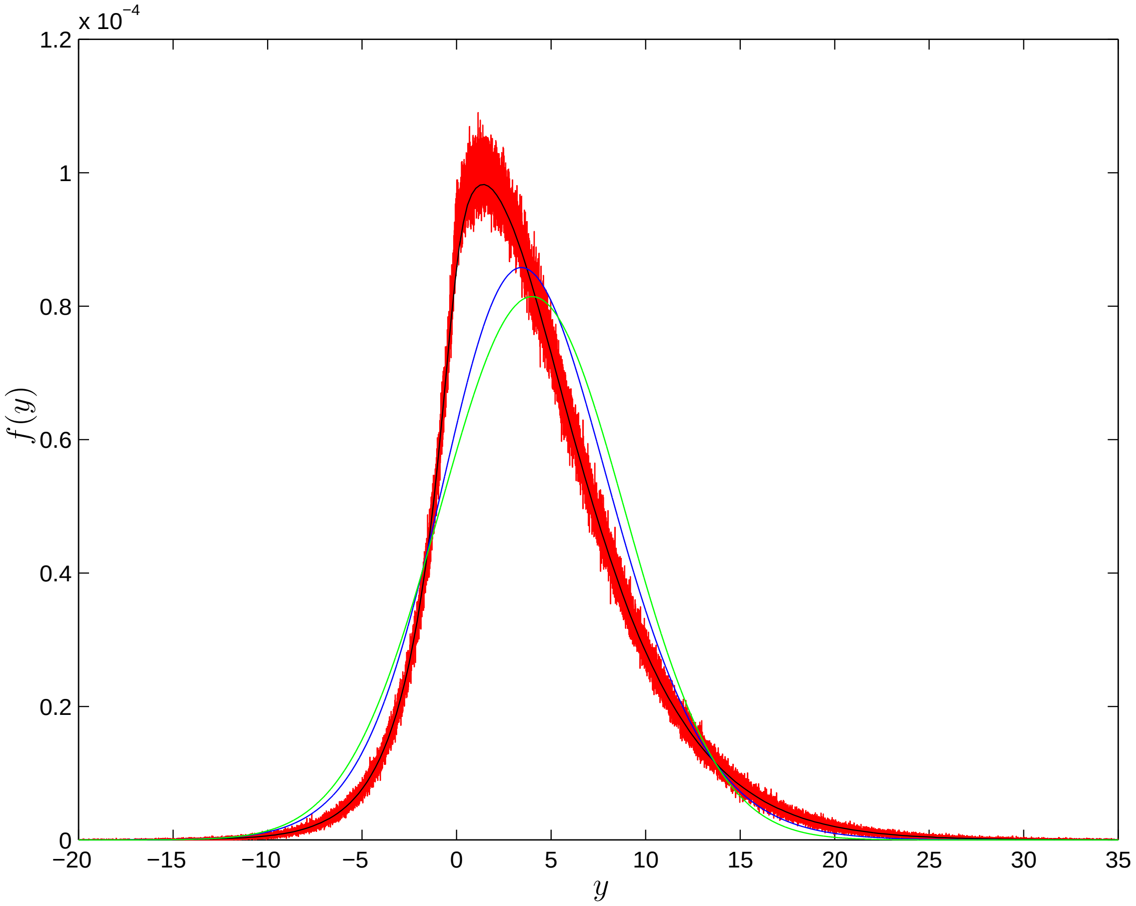

Figure 19: The density of the error metric approximated by numerical integration and convolutions (blue), a Monte Carlo method (red), a kernel density estimate (black) and a Normal density (green) for Example 1.

$(-82.004, 118.004)$. In Figure 19 this approximation of $M_e(s, \hat{s})$ is represented by the smooth blue line.

5.5.2 Example 1: Monte Carlo using MATLAB

By directly using a Monte Carlo construction we can approximate the density of the error metric. To simulate the Y vectors needed we randomly select $15,000,000$ observations for each $Y_{R,t}$ and $Y_{I,t}$, $t = 0, 1, 2$. We chose to select $15,000,000$ observations because that is the maximum number that can currently be stored in the memory of a computer operating in the Department of Mathematics and Statistics, University of Canterbury. A value of the error metric is calculated for each of these $15,000,000$ sets of Y values. By sorting these values into fine intervals we obtain an approximation of the density of the error metric. The simulated values of $M_e(s, \hat{s})$ were sorted into bins of width 0.001. Figure 19 shows this approximation as the jagged red line. To estimate $P[M_e(s, \hat{s}) < 0]$ we calculate the proportion of our $15,000,000$ simulations that produce a negative value of $M_e(s, \hat{s})$.

The KDE was computed using a smoothing parameter of $h = 0.4$. It is displayed in Figure 19 as the smooth black line. As expected the KDE bisects the approximating histogram.

5.5.3 Example 1: Monte Carlo using WinBUGS

The density of the error metric can be directly simulated using WinBUGS. $5,000,000$ sets of Y values were generated and the mean, variance, skewness and $P[M_e(s, \hat{s}) < 0]$ were directly obtained. These statistics are displayed in Table 2. The KDE computed using WinBUGS has the same shape and coverage as the KDE generated using MATLAB.

5.5.4 Example 1: Normal Approximation

In the previous section we showed that, for our examples, the expectation and variance of the error metric can be represented as $E\left(M_e(s, \hat{s})\right) = 2N\xi\mu^2$ and $V\left(M_e(s, \hat{s})\right) = 4N\xi\sigma^2\left(\sigma^2 + 2\mu^2\right)$. Since $Y_{R,t} \sim N(1, 1)$ and $Y_{I,t} \sim N(1, 1)$ for $t = 0, 1, 2$, $N = 2$ and $\xi = 1$, the density of the error metric is approximated by a Normal density with mean 4 and variance 24.

Recall that in Section 3.3 we found that the skewness of the product of two independent Normal variables, $\alpha_3(Y_1Y_2)$, is greatest when $\mu_1 = \sigma_1$ and $\mu_2 = \sigma_2$. In this example each individual $\alpha_3(Y)$ value is the largest possible. Although the Central Limit Theorem tells us that the distribution of $M_e\left(s, \hat{s}\right)$ will approach Normality as the number of summed variables increases, in this case we are only summing eight products of Normals. Thus, although the resulting density is less non-Normal than the original product, it is still far from Normal itself.

	NI	MC MAT	KDE	MC WinB	Normal
Mean	4.0032	3.9993	3.9985	3.9996	4
Variance	23.9955	24.0006	24.1588	23.9648	24
Skewness	0.2067	0.8165	0.8166	0.8135	0
$P[M_e(s, \hat{s}) < 0]$	0.2007	0.1840	0.1881	0.1851	0.2071

Table 2: Comparison of statistics for the different error metric approximations for Example 1.

5.5.5 Example 1: Results

Figure 19 displays the four densities produced by our different approximation methods. The density of the error metric is clearly non-Normal. The KDE (black) is a smoothed version of the Monte Carlo (red), and both of these densities are quite different from the numerically integrated and convoluted (blue) and Normal (green) densities. Given the high skewness of each product of Normal variables this is hardly surprising.

Table 2 shows that the mean, variance and skewness for the error metric, and the probability of observing a value less than 0, are very similar for both Monte Carlo approximations and the smoothed KDE. Again, this is not a surprise when we consider that both Monte Carlo estimates were formed using the same process on different computer packages, and that the KDE is a smoothed estimate of the Monte Carlo data. The numerically integrated and convoluted and Normal approximations have a similar mean and variance to the other methods, but both underestimate the skewness.

When we consider the probability of observing a value of the error metric less than zero, all five methods are reasonably similar, but examination of Figure 19 suggests this may be a matter of luck! The numerical integration and Normal densities both give higher probabilities to large negative values and smaller probabilities to values slightly below zero.

5.6 Example 2

In the second example we assume that $N = 2$, $\xi = 1$, $b_{R,i} = b_{I,i} = [1 \quad 1]$ and

$$\begin{bmatrix} Y_{R,0} \\ Y_{I,0} \\ Y_{R,1} \\ Y_{I,1} \\ Y_{R,2} \\ Y_{I,2} \end{bmatrix} \sim N \left(\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0.1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.1 \end{bmatrix} \right).$$

The form of the error metric is the same as in Equation 51,

$$\begin{aligned} M_e(s, \hat{s}) &= y_{R,0}y_{R,1} + y_{R,1}y_{I,0} - y_{R,0}y_{I,1} + y_{I,0}y_{I,1} + y_{R,1}y_{R,2} + y_{R,2}y_{I,1} \\ &\quad - y_{R,1}y_{I,2} + y_{I,1}y_{I,2}. \end{aligned}$$

5.6.1 Example 2: Numerical Integration and Convolutions

Numerical integration is used to identify the density of the product of two $N(1, 0.1)$ distributions. The product is integrated between limits -1 and 4 with interval widths of 0.001, and is convoluted to produce an approximate density for $M(s, \hat{s})$. This is displayed in Figure 20 as the smooth blue line.

5.6.2 Example 2: Monte Carlo using MATLAB

A Monte Carlo construction can be used to simulate an error metric directly. 15,000,000 values were simulated for each of the six Normally distributed variables. An approximating histogram is shown in Figure 20 as the jagged red line. A KDE, displayed as a smooth black line, was computed using $h = 0.2$.

5.6.3 Example 2: Monte Carlo using WinBUGS

WinBUGS was used to directly obtain the density of the error metric. $5,000,000$ sets of Y values were generated and the mean, variance, skewness and $P[M_e(s, \hat{s}) < 0]$ were calculated. They are displayed in Table 3.

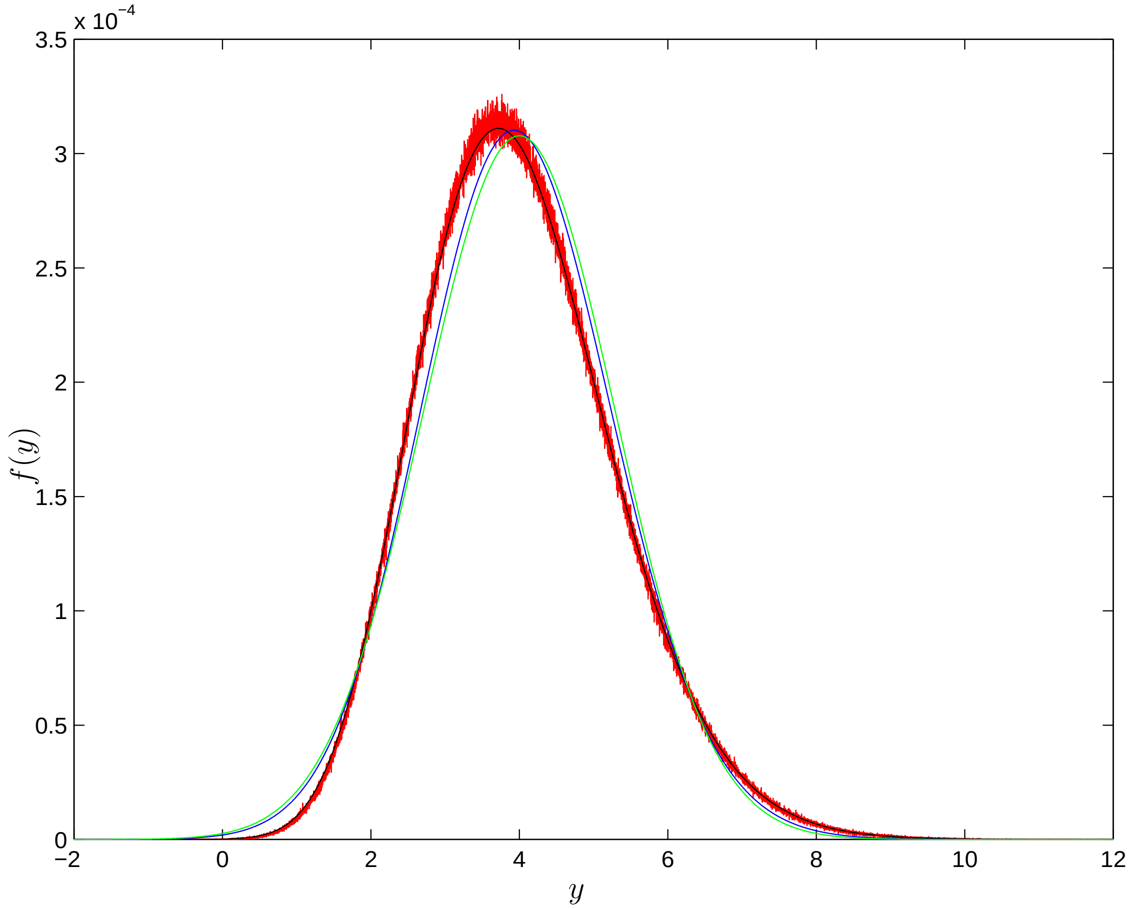

Figure 20: The density of the error metric approximated by numerical integration and convolutions (blue), a Monte Carlo method (red), a kernel density estimate (black) and a Normal density (green) for Example 2.

5.6.4 Example 2: Normal Approximation

Equations 48 and 49 tell us that $E(M(s,\hat{s})) = 4$ and $V(M(s,\hat{s})) = 1.68$. We approximate the density of the error metric with a $N(4, 1.68)$ density. This is shown in Figure 20 as the green line.

On this occasion we expect the Normal approximation will more accurately reflect the density of the error metric. Each individual product of Normals is less skewed than in Example 1, since $\mu = \frac{1}{\sqrt{10}}\sigma$.

5.6.5 Example 2: Results

Figure 20 shows that the densities produced have much more in common than they did in the previous example. The approximations obtained through numerical inte-

	NI	MC MAT	KDE	MC WinB	Normal
Mean	4.0023	4.0002	3.9993	3.9993	4
Variance	1.6755	1.6805	1.8463	1.6810	1.68
Skewness	0.1183	0.4404	0.4393	0.4396	0
$P[M_e(s, \hat{s}) < 0]$	0.0007	0.0002	0.0001	0.0002	0.0010

Table 3: Comparison of statistics for the different error metric approximations for Example 2.

gration and convolutions and the Normal density are very similar. Although they are still noticeably different from the Monte Carlo and KDE approximations, the four densities are much more similar than they were in Example 1. This is due to the individual densities of products of two $N(1, 0.1)$ variables being more Normal themselves, since the μ/σ ratio is larger.

In both of the examples we have studied $N\xi$ has been small. In fact it has been 2, the smallest specification possible. In an example with a larger value of $N\xi$, which contains more densities to sum, we can expect the different approximation methods to produce increasingly similar results.

Table 3 shows how similar the statistics for the different approximations are. As expected, the difference between the different means and variances is small. The KDE and the two Monte Carlo approximations are still more skewed, but the skewness is roughly half what it was in Example 1. For all approximation methods $P[M_e(s, \hat{s}) < 0]$ is very small, although the numerically integrated and convoluted and Normal approximations still give higher probabilities to negative values.

6 Summary

We began this Report by briefly introducing the process of Differential Continuous Phase Frequency Shift Keying. The error metric was introduced and we stated our problem in full detail. We discovered that the calculation of the error metric involves two distinct steps: we must assess the distribution for the product of two Normally distributed variables, and we must identify the distribution of the sum of a number of these products of Normal distributions. Sections 3 and 4 discuss these two processes in depth.

In Section 3 the density of the product of two independent Normal variables is approximated using three different methods. We considered a numerical methods approximation, which consisted of implementing numerical integration procedures on both MATLAB and Maple. We considered a Monte Carlo construction. We investigated the moment generating function of the product of two correlated Normal variables, and showed how the mean, variance and skewness change for three special cases. We also showed that as the ratio of μ to σ increases, the distribution of the product of two independent Normal variables tends towards a Normal distribution. These three approximation methods were compared over two examples. It was shown that although both the numerical integration and the Monte Carlo simulation produced densities of the same shape, and similar statistics, the numerical integration was finely accurate compared to the crudeness of the Monte Carlo method. It was shown that it is inappropriate to use a Normal density as an approximation when μ/σ is small.

Section 4 considered how to identify the density of the sum of the products of two Normal variables. The Convolution Formula was introduced, and the computational steps used to identify the density of a sum of random variables outlined. The technique was demonstrated using the product Normal densities obtained in Section 3.

In Section 5 the theory of the preceding Sections was combined. The density of the error metric was approximated using numerical integration and the Convolution Formula, and $P[M_e(s, \hat{s}) < 0]$ was computed. A Monte Carlo method was used to approximate the density of the error metric, as was a Normal density. The densities were shown to have the same shape and similar statistics, but the density of $(M_e(s, \hat{s}))$ approximated using numerical integration and convolutions is more accurate.

Acknowledgements

This work was undertaken as part of Robert Ware's doctoral thesis. He was partially supported by the University of Canterbury Keith Laugesen Scholarship and a

Mathematics and Statistics Departmental Scholarship. Thank you to Doris Barnard and Alistair Smith for many helpful comments.

References

Aroian, L. A. (1947). The Probability Function of a Product of Two Normally Distributed Variables. *Ann. Math. Statist.*, 18:265–271.

Aroian, L. A., Taneja, V. S., and Cornwell, L. W. (1978). Mathematical Forms of the Distribution of the Product of Two Normal Variables. *Commun. Statist.-Theor. Meth.*, 7(2):164–172.

Conradie, W. and Gupta, A. (1987). Quadratic Forms in Complex Normal Variates: Basic Results. *Statistica*, 47:73–84.

Cornwell, W., Aroian, L. A., and Taneja, V. S. (1978). Numerical Evaluation of the Distribution of the Product of Two Normal Variables. *J. Statist. Comput. Simul.*, 7:123–131.

Craig, C. C. (1936). On the Frequency Function of xy. *Ann. Math. Statist.*, 7:1–15.

Griffin, A. (2000). *Coding CPFSK for differential demodulation.* PhD thesis, University of Canterbury.

Hoti, F. J., Sillanpää, M. J., and Holmström, L. (2002). A Note on Estimating the Posterior Density of a Quantitative Trait Locus From a Markov Chain Monte Carlo Sample. *Genetic Epidemiology*, 22:369–376.

Scott, D. W. (1985). Averaged shifted histograms: effective nonparametric density estimators in several dimensions. *Annals of Statistics*, 13:1024–1040.

Scott, D. W. (1992). *Multivariate density estimation.* New York: John Wiley and Sons.

Appendix A: The Moment-Generating Function of the Product of Two Correlated Normally Distributed Variables

Assume $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$ and that X_1 and X_2 have correlation ρ. If we define $X_1 = X_0 + Z_1$ and $X_2 = X_0 + Z_2$, where

$$\begin{bmatrix} X_0 \\ Z_1 \\ Z_2 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \rho\sigma_1\sigma_2 & 0 & 0 \\ 0 & \sigma_1^2 - \rho\sigma_1\sigma_2 & 0 \\ 0 & 0 & \sigma_2^2 - \rho\sigma_1\sigma_2 \end{bmatrix}\right).$$

We have decomposed X_1 and X_2 into independent summands of which one is shared between them.

To find the moment-generating function of $Y = X_1X_2 = (X_0 + Z_1)(X_0 + Z_2)$, we know that

$$\begin{aligned} M_Y(t) &= \int_{-\infty}^{\infty} e^{ty} f(y)dy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{t(x_0+z_1)(x_0+z_2)} f(x_0, z_1, z_2)dx_0dz_1dz_2 \\ &= \frac{1}{\sqrt{2\pi\rho\sigma_1\sigma_2}} \frac{1}{\sqrt{2\pi\left(\sigma_1^2 - \rho\sigma_1\sigma_2\right)}} \frac{1}{\sqrt{2\pi\left(\sigma_2^2 - \rho\sigma_1\sigma_2\right)}} \\ &\quad \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{t(x_0+z_1)(x_0+z_2) - \frac{1}{2}\frac{(x_0-0)^2}{\rho\sigma_1\sigma_2} - \frac{1}{2}\frac{(z_1-\mu_1)^2}{\sigma_1^2-\rho\sigma_1\sigma_2} - \frac{1}{2}\frac{(z_2-\mu_2)^2}{\sigma_2^2-\rho\sigma_1\sigma_2}} dx_0dz_1dz_2 \quad (52) \end{aligned}$$

To undertake the triple integration in Equation 52, we will need to reorganise the exponent of the exponential term. Let

$$q_1 = t\left(x_0 + z_1\right)\left(x_0 + z_2\right) - \frac{1}{2}\left(\frac{\left(x_0 - 0\right)^2}{\rho\sigma_1\sigma_2} - \frac{\left(z_1 - \mu_1\right)^2}{\sigma_1^2 - \rho\sigma_1\sigma_2} - \frac{\left(z_2 - \mu_2\right)^2}{\sigma_2^2 - \rho\sigma_1\sigma_2}\right). \quad (53)$$

If we define

$$q_2 = -\frac{1}{2}\left(\frac{x_0 - c_1\left(z_1 + z_2\right) - c_2}{c_3}\right)^2 - \frac{1}{2}\left(\frac{z_1 - c_4z_2 - c_5}{c_6}\right)^2 - \frac{1}{2}\left(\frac{z_2 - c_7}{c_8}\right)^2 + c_9, \quad (54)$$

we can rewrite q_1 in the form of q_2 by equating terms c_1 to c_9 with some part of q_1. The moment-generating function of Y can then be written in the form

$$M_Y(t) = \frac{1}{\sqrt{\rho\sigma_1\sigma_2}} \frac{1}{\sqrt{\sigma_1^2 - \rho\sigma_1\sigma_2}} \frac{1}{\sqrt{\sigma_2^2 - \rho\sigma_1\sigma_2}} c_3c_6c_8e^{c_9}. \quad (55)$$

By rewriting q_1 and q_2 as

$$\begin{aligned} q_1 &= x_0^2\left(t-\frac{1}{2\rho\sigma_1\sigma_2}\right)+x_0\left(z_1+z_2\right)(t)+x_0\left(0\right)+z_1^2\left(-\frac{1}{2(\sigma_1^2-\rho\sigma_1\sigma_2)}\right) \\ &\quad +z_1\left(\frac{\mu_1}{\sigma_1^2-\rho\sigma_1\sigma_2}\right)+z_2^2\left(-\frac{1}{2(\sigma_2^2-\rho\sigma_1\sigma_2)}\right)+z_2\left(\frac{\mu_2}{\sigma_2^2-\rho\sigma_1\sigma_2}\right)+z_1z_2\left(t\right) \\ &\quad +\left(-\frac{\mu_1^2}{2(\sigma_1^2-\rho\sigma_1\sigma_2)}-\frac{\mu_2^2}{2(\sigma_2^2-\rho\sigma_1\sigma_2)}\right) \end{aligned} \tag{56}$$

and

$$\begin{aligned} q_2 &= x_0^2\left(-\frac{1}{2c_3^2}\right)+x_0\left(z_1+z_2\right)\left(\frac{c_1}{c_3^2}\right)+x_0\left(\frac{c_2}{c_3^2}\right)+z_1^2\left(-\frac{c_1^2}{2c_3^2}-\frac{1}{2c_6^2}\right) \\ &\quad +z_1\left(-\frac{c_1c_2}{c_3^2}+\frac{c_5}{c_6^2}\right)+z_2^2\left(-\frac{c_1}{2c_3^2}-\frac{c_4}{2c_6^2}-\frac{1}{2c_8^2}\right)+z_2\left(-\frac{c_1c_2}{c_3^2}-\frac{c_4c_5}{c_6^2}+\frac{c_7}{c_8^2}\right) \\ &\quad +z_1z_2\left(-\frac{c_1^2}{c_3^2}+\frac{c_4}{c_6^2}\right)+\left(-\frac{c_2^2}{2c_3^2}-\frac{c_5^2}{2c_6^2}-\frac{c_7^2}{2c_8^2}+c_9\right), \end{aligned} \tag{57}$$

and then solving on Maple, we get the results:

$$\begin{aligned} c_1 &= \frac{t\rho}{1-2t\rho\sigma_1\sigma_2}, \\ c_2 &= 0, \\ c_3 &= \sqrt{\frac{\rho}{1-2t\rho\sigma_1\sigma_2}}, \\ c_4 &= \frac{t\sigma_1(t\rho\sigma_1\sigma_2-1)(\sigma_1-\rho\sigma_2)}{t^2\rho\sigma_1^3\sigma_2-(t\rho\sigma_1\sigma_2-1)^2}, \\ c_5 &= \frac{\mu_1(2t\rho\sigma_1\sigma_2-1)}{t^2\rho\sigma_1^3\sigma_2-(t\rho\sigma_1\sigma_2-1)^2}, \\ c_6 &= \sqrt{\frac{2t\rho\sigma_1^2\sigma_2\left(\sigma_1-\rho\sigma_2^2\right)-\sigma_1\left(\sigma_1+\rho\sigma_2\right)}{t^2\rho\sigma_1^3\sigma_2-(t\rho\sigma_1\sigma_2-1)^2}}, \\ c_7 &= \frac{t\rho\sigma_1\sigma_2\left(t\rho\sigma_1\sigma_2\left(\mu_1+\mu_2\right)-t\left(\mu_1\sigma_2^2+\mu_2\sigma_1^2\right)-\mu_1-2\mu_2\right)+t\mu_1\sigma_2^2+\mu_2}{(t\rho\sigma_1\sigma_2-1)^2-t^2\sigma_1^2\sigma_2^2}, \\ c_8 &= \sqrt{\frac{t^2\rho^2\sigma_1^2\sigma_2^2\left(\sigma_1^2-\rho\sigma_1\sigma_2+\sigma_2^2\right)+t\rho\sigma_1\sigma_2^2\left(2\rho\sigma_1-t\sigma_1^2\sigma_2-2\sigma_2\right)+\sigma_2\left(\sigma_2-\rho\sigma_1\right)}{(t\rho\sigma_1\sigma_2-1)^2-t^2\sigma_1^2\sigma_2^2}}, \\ c_9 &= \frac{t^2\left(\mu_1^2\sigma_2^2+\mu_2^2\sigma_1^2-2\mu_1\mu_2\sigma_1\sigma_2\rho\right)+2t\mu_1\mu_2}{2\left((t\rho\sigma_1\sigma_2-1)^2-t^2\sigma_1^2\sigma_2^2\right)}. \end{aligned}$$

Also

$$c_3c_6c_8=\sqrt{\frac{\rho\sigma_1\sigma_2\left(\sigma_1^2-\rho\sigma_1\sigma_2\right)\left(\sigma_2^2-\rho\sigma_1\sigma_2\right)}{(t\rho\sigma_1\sigma_2-1)^2-t^2\sigma_1^2\sigma_2^2}}. \tag{58}$$

To find the moment-generating function of Y, substitute c_3, c_6, c_8 and c_9 into

Equation 55,

$$
\begin{aligned}
M_Y(t) &= \frac{1}{\sqrt{\rho\sigma_1\sigma_2\left(\sigma_1^2-\rho\sigma_1\sigma_2\right)\left(\sigma_2^2-\rho\sigma_1\sigma_2\right)}}\sqrt{\frac{\rho\sigma_1\sigma_2(\sigma_1^2-\rho\sigma_1\sigma_2)(\sigma_2^2-\rho\sigma_1\sigma_2)}{\left(t\rho\sigma_1\sigma_2-1\right)^2-t^2\sigma_1^2\sigma_2^2}} \\
&\quad e^{\frac{t^2\left(\mu_1^2\sigma_2^2+\mu_2^2\sigma_1^2-2\mu_1\mu_2\sigma_1\sigma_2\rho\right)+2t\mu_1\mu_2}{2\left(\left(t\rho\sigma_1\sigma_2-1\right)^2-t^2\sigma_1^2\sigma_2^2\right)}} \\
&= \left(\sqrt{\left(t\rho\sigma_1\sigma_2-1\right)^2-t^2\sigma_1^2\sigma_2^2}\right)^{-1} e^{\frac{\mu_1\mu_2 t+\frac{1}{2}\left(\mu_1^2\sigma_2^2+\mu_2^2\sigma_1^2-2\rho\mu_1\mu_2\sigma_1\sigma_2\right)t^2}{(1-t\rho\sigma_1\sigma_2)^2-\sigma_1^2\sigma_2^2t^2}}. \qquad (59)
\end{aligned}
$$

Appendix B: Moments

The moment-generating function can be used to find moments about the origin of Y. By differentiating $M_Y(t)$ and evaluating at $t=0$ we can find as many moments as required. These moments can be used to calculate the mean, variance and skewness of the product of two correlated Normal variables. Using Maple we find that

$$
\begin{aligned}
M_Y(0) &= 1, \\
M_Y'(0) &= \mu_1\mu_2+\rho\sigma_1\sigma_2, \\
M_Y''(0) &= \left(\mu_1^2+\sigma_1^2\right)\left(\mu_2^2+\sigma_2^2\right)+4\mu_1\mu_2\rho\sigma_1\sigma_2+2\rho^2\sigma_1^2\sigma_2^2, \\
M_Y'''(0) &= \left(\mu_1^3+3\mu_1\sigma_1^2\right)\left(\mu_2^3+3\mu_2\sigma_2^2\right)+9\rho\sigma_1\sigma_2\left(\mu_1^2+\sigma_1^2\right)\left(\mu_2^2+\sigma_2^2\right) \\
&\quad +6\rho^2\sigma_1^2\sigma_2^2\left(\rho\sigma_1\sigma_2+3\mu_1\mu_2\right).
\end{aligned}
$$

Thus

$$
\begin{aligned}
E(Y) &= \mu_1\mu_2+\rho, && (60) \\
V(Y) &= \mu_1^2\sigma_2^2+\mu_2^2\sigma_1^2+\sigma_1^2\sigma_2^2+2\mu_1\mu_2\rho\sigma_1\sigma_2+\rho^2\sigma_1^2\sigma_2^2, && (61) \\
\alpha_3(Y) &= \frac{6\sigma_1\sigma_2\left(\mu_1\mu_2\sigma_1\sigma_2\left(\rho^2+1\right)+\rho\left(\mu_1^2\sigma_2^2+\mu_2^2\sigma_1^2\right)\right)+2\rho\sigma_1^3\sigma_2^3\left(3+\rho^2\right)}{\left(\mu_1^2\sigma_2^2+\mu_2^2\sigma_1^2+\sigma_1^2\sigma_2^2\left(1+\rho^2\right)+2\rho\mu_1\mu_2\sigma_1\sigma_2\right)^{3/2}}. && (62)
\end{aligned}
$$