ESAIM: PS 28 (2024) 195–226
https://doi.org/10.1051/ps/2024005

ESAIM: Probability and Statistics
www.esaim-ps.org

DEFORMED NORMAL DISTRIBUTIONS

Tore Langholm and Harald Totland*

Abstract. It is well known that the probability distribution of the product of two normal variables itself approaches a normal distribution if one of the factor variables has a low coefficient of variation. A related, but apparently somewhat neglected problem, is to find the conditional distribution of a factor variable when the product of two independent normal variables is known. In the present paper we indicate why this conditional distribution also tends towards a normal distribution under similar conditions, and demonstrate numerically that this is indeed the case. Power series expansions for the mean and variance of the conditional distribution are also presented, which hold some problems of convergence but nevertheless provide good approximations when one coefficient of variation is low. Finally, a simplified version is presented of an actual application in object tracking, yielding an approximation for the probability distribution of the distance out to an object when the relative azimuth is known. The power series expansions shown in this paper are most conveniently developed in a more abstract setting, yielding results about the more general notion of a deformed normal distribution.

Mathematics Subject Classification. 60E05, 62Exx.

Received June 30, 2023. Accepted March 19, 2024.

1. Introduction

As we shall be using the term, a deformed normal distribution is a probability distribution where the density function is the product of a normal probability density and a second function. Typically, but not necessarily, the second function may reflect additional information obtained *a posteriori*, yielding a modified, conditional distribution. A situation of this type arises when the product of two independent and normally distributed random variables is known, and a quick estimate for the conditional probability distribution of one of the component variables is needed. The concept of multiplicative errors provides a simple example. If items with normally distributed weights are measured on a scale where the relative error is normally distributed, then a given reading will be the product of the figure of interest – the actual weight – and another normal variable, and (provided more readings are not available) the best one can do is to find a conditional distribution for the actual weight given this reading. Another setting of this type is described in Section 10, where an object has traveled an unknown length along a circular path of unknown curvature, where this length and curvature are independently and normally distributed. In this case, the total change of direction is given as the product of the two variables, and can be observed in situations where the values of the individual components are unknown. In navigation and tracking it may be crucial to have quick recourse to an estimate for the distance traveled, in

Keywords and phrases: Normal distribution, product distribution, multiplicative error, heat equation, generalized Weierstrass transform, tracking.

Norwegian Defence University College, Royal Norwegian Naval Academy, Bergen, Norway.

* Corresponding author: htotland@mil.no

some cases for a very large number of different, hypothetical scenarios. In recent publications [1, 2] we used the results obtained in Section 7 in a setting somewhat similar to the above, but with a more complex model for the movement of objects.

When the probability distribution for two independent, normal variables is given, and the product of their outcome values is known, it is straightforward to find a mathematical expression for the conditional probability density function of either one of the component variables. However, complicated integrals are involved that are not well suited for straightforward analysis and may be too demanding to compute in situations where many such computations need to be carried out in a short time span, perhaps also on limited hardware. To find a practical solution, then, one may explore the extent to which it is possible to approximate such conditional probability distributions by normal probability distributions with the parameters obtained by low-degree polynomials in the observed product value. In the following sections, this is lifted to the more general level of deformed normal distributions, and then in Section 5 and onward the results are evaluated in the more specific setting of known product values.

2. General notions and results

Below we assume the usual understanding of a probability density function on the reals as any non-negative, measurable function p such that the integral $\int_{\mathbb{R}} p(x)\mathrm{d}x$ exists and equals unity. In the sequel we shall suppress the domain of integration $\mathbb{R}$ and just write $\int p(x)\mathrm{d}x$ for the integral over the reals. We shall also use the notation

$$g(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

for the probability density function of a normally distributed variable $X \sim \mathcal{N}(\mu,\sigma)$ with mean μ and standard deviation σ.

Definition 2.1. Suppose the random variable X has a probability density function that equals

$$p_X(x) = \frac{1}{\Theta}\, f(x)\, g(x;\mu,\sigma) \tag{2.1}$$

for some positive constant Θ and some non-negative function f on the reals. We then say that X has the normal distribution $\mathcal{N}(\mu,\sigma)$ deformed by f, and write $X \sim \mathcal{N}(\mu,\sigma;f)$.[1]

Deformed normal distributions turn up in situations with a normal variable $X \sim \mathcal{N}(\mu,\sigma)$ and some event A with a probability $P(A|X=x) = f(x)$ that depends on the value of X. If A has non-zero probability, Bayes' rule yields the identity

$$p_{X|A}(x) = \frac{P(A|X=x)\cdot g(x;\mu,\sigma)}{P(A)},$$

so $X|A \sim \mathcal{N}(\mu,\sigma;f)$ provided f is measurable. For instance, if some characteristic of the individuals in a population originally followed a normal distribution, but the population has since been reduced, with a probability of remaining in the population that depends on the characteristic in question, then this probability, as a function of the values of the relevant characteristic, takes the role of a deforming function in the new, resulting distribution.

When Y is another continuous (or in particular a normal) variable and A is the event $X\cdot Y = z$, we obtain the situation described in the introduction, with the conditional distribution for X given that $XY = z$. In this case, however, there is a slight difference, since $P(XY=z) = 0$ for any z, and the deforming function takes a different form, as discussed in Section 5 and onward.

[1] In the special case where f is itself the probability density function for a normal variable, the results here are of limited use, since it is then well known that X is normal as well. In the present terminology, a normal distribution deformed by a normal density is normal. The same applies to the case where $f(x) = 1/g(x;\mu',\sigma')$ and $\sigma' > \sigma$.

Deformed normal distributions are not necessarily conditional distributions, however. Consider an example where a certain characteristic of the individuals born into a population is normally distributed, but then the individuals remain in the population for an expected time period that depends on the value of this characteristic. Now looking at the individuals of such a population at any given moment in time, we find that the characteristic in question has a deformed normal distribution, with life expectancy as the deforming function.

It follows from Definition 2.1 that f is a measurable function if $X \sim \mathcal{N}(\mu, \sigma; f)$. It also follows that

$$\Theta = \int f(x) g(x; \mu, \sigma) \mathrm{d}x,$$

and, in particular, that the integral exists and is nonzero. As this expression for Θ is indexed by μ (and σ), it makes sense to consider the derivatives of Θ with respect to μ. Taking μ to be a parameter not occurring in f,[2] such differentiation is governed by $g(x; \mu, \sigma)$ alone, yielding some useful, general connections between the moments of a deformed normal distribution and the derivatives of Θ. The relevant results are recorded in Theorem 2.2.

In order for $\frac{\partial \Theta}{\partial \mu}$ to exist, Θ will have to be defined for all μ in some open interval. Such a condition is stated explicitly in the theorem. This is necessary, as our notion of a deformed normal distribution is very general, allowing *any* distribution into the fold, since $p(x)$ can be written as $\frac{p(x)}{g(x;0,1)} g(x; 0, 1)$. Such *ad hoc* constructions may not be stable, however, in the sense that an arbitrarily small shift in the expectation parameter may yield an infinite Θ. For instance, the Cauchy distribution can be considered as the deformed normal distribution with density function $\frac{1}{\pi(1+x^2) g(x;0,1)} g(x; 0, 1)$ and $\Theta = 1$, while $\int \frac{1}{\pi(1+x^2) g(x;0,1)} g(x; \delta, 1) \mathrm{d}x$ is infinite for any $\delta \neq 0$. One sufficient condition on non-negative, measurable f that rules out such examples and ensures that $\int f(x) g(x; \mu, \sigma) \mathrm{d}x$ has a finite positive value for all real μ, is the existence of a, b, and $c < \frac{1}{2\sigma^2}$ such that $0 < \int_{-a}^{a} f(x) \mathrm{d}x < \infty$ and $f(x) \leq b \mathrm{e}^{cx^2}$ for all x with $|x| > a$.

Note that Theorem 2.2 only requires a weaker, local "stability condition". The proof of the theorem uses two of the lemmas in Appendix A, where we show that such stability ensures that the relevant integrals converge uniformly for the μ in an open interval, which is needed to guarantee differentiability of the integrals with respect to μ.

In part 2 of the theorem, L^m denotes repeated application of the linear differential operator L, while M_m in part 3 is the m-th central moment.

Theorem 2.2. *Suppose $X \sim \mathcal{N}(\mu, \sigma; f)$, and suppose Θ exists for all μ in an open interval. Then all moments of X exist, Θ and all the moments are infinitely differentiable functions of μ, and the following identities hold.*[3]

1. $$E[X^m] = \frac{1}{\Theta} \sum_{n=0}^{m} \frac{\partial^n \Theta}{\partial \mu^n} \cdot \binom{m}{n} \sum_{k=0}^{\lfloor \frac{m-n}{2} \rfloor} \binom{m-n}{2k} (2k-1)!! \mu^{m-n-2k} \sigma^{2n+2k}$$
2. $$E[X^m] = L\left(E[X^{m-1}]\right) = L^m(1), \quad \text{where} \quad L = E[X] + \sigma^2 \frac{\partial}{\partial \mu} = \mu + \sigma^2 \frac{\partial \ln \Theta}{\partial \mu} + \sigma^2 \frac{\partial}{\partial \mu}$$
3. $$M_m = \sigma^2 \frac{\partial M_{m-1}}{\partial \mu} + (m-1) M_2 M_{m-2}, \quad \text{where} \quad M_m = E\left[(X - E[X])^m\right]$$

Proof. Lemma A.2 yields the existence of $E[X] = \frac{1}{\Theta} \int x f(x) g(x; \mu, \sigma) \mathrm{d}x$ and infinite differentiability of Θ, and in turn infinite differentiability of $E[X]$. Thus all moments are expectation values of polynomials in X with infinitely differentiable coefficients, and hence Lemma A.2 applies to all.

[2] This means excluding functions of the form $f(x, \mu)$ in (2.1), but does not prevent Θ from being rewritten as $\int f(\mu - x) g(x; 0, \sigma) \mathrm{d}x$ or $\int f(\mu + \sigma x) g(x; 0, 1) \mathrm{d}x$.

[3] Here, $\lfloor\ \rfloor$ is the floor function. The double factorial $n!!$ of a positive integer n is defined as the product of all positive integers that are less than or equal to n and have the same parity as n, while $0!!$ and $(-1)!!$ are defined as 1.

Since $g(x;\mu,\sigma)$ is proportional to $\mathrm{e}^{-\frac{z^2}{2}}$ with $z=\frac{x-\mu}{\sigma}$, we use the (probabilist's) Hermite polynomials, defined by $H_n(x)=(-1)^n\mathrm{e}^{\frac{x^2}{2}}\frac{\mathrm{d}^n}{\mathrm{d}x^n}\mathrm{e}^{-\frac{x^2}{2}}$, to find

$$\frac{\partial^n g}{\partial\mu^n}=\left(-\frac{1}{\sigma}\right)^n\frac{\partial^n g}{\partial z^n}=\frac{1}{\sigma^n}H_n(z)g(x;\mu,\sigma),$$

and hence, due to Lemma A.2,

$$\frac{\partial^n\Theta}{\partial\mu^n}=\frac{1}{\sigma^n}\int H_n(z)f(x)g(x;\mu,\sigma)\mathrm{d}x.$$

Writing $x^m=(\sigma z+\mu)^m=\sum_{n=0}^{m}\binom{m}{n}z^n\mu^{m-n}\sigma^n$ and using z for x in the well-known relation

$$x^n=\sum_{k=0}^{\lfloor\frac{n}{2}\rfloor}\binom{n}{2k}(2k-1)!!H_{n-2k}(x),$$

we get

$$E[X^m]=\frac{1}{\Theta}\sum_{n=0}^{m}\sum_{k=0}^{\lfloor\frac{n}{2}\rfloor}\binom{m}{n}\binom{n}{2k}(2k-1)!!\mu^{m-n}\sigma^{2n-2k}\frac{\partial^{n-2k}\Theta}{\partial\mu^{n-2k}}.$$

Rearranging the sums with $n'=n-2k$, noting that $0\le 2k=n-n'\le m-n'$, dropping the mark on n', and using the relation $\binom{m}{n+2k}\binom{n+2k}{2k}=\binom{m}{n}\binom{m-n}{2k}$, we obtain the formula in part 1.

For a polynomial $p(x,\mu)$ in x where the coefficients are infinitely differentiable functions of μ, differentiation of

$$\Theta\cdot E[p(X,\mu)]=\int p(x,\mu)f(x)g(x;\mu,\sigma)\mathrm{d}x$$

with respect to μ yields, with the use of Lemma A.2,

$$\frac{\partial\Theta}{\partial\mu}E[p(X,\mu)]+\Theta\frac{\partial}{\partial\mu}E[p(X,\mu)]=\int\left(\frac{\partial}{\partial\mu}p(x,\mu)\right)f(x)g(x;\mu,\sigma)\mathrm{d}x+\int p(x,\mu)\frac{x-\mu}{\sigma^2}f(x)g(x;\mu,\sigma)\mathrm{d}x,$$

and hence

$$\left(\mu+\sigma^2\frac{\partial}{\partial\mu}\ln\Theta\right)E\left[p(X,\mu)\right]+\sigma^2\frac{\partial}{\partial\mu}E\left[p(X,\mu)\right]=\sigma^2E\left[\frac{\partial}{\partial\mu}p(X,\mu)\right]+E\left[Xp(X,\mu)\right].$$

Setting $p(x,\mu)=1$ gives the identity

$$E[X]=\mu+\sigma^2\frac{\partial}{\partial\mu}\ln\Theta,$$

in accordance with part 1. Thus,

$$E\left[Xp(X,\mu)\right] + \sigma^2 E\left[\frac{\partial}{\partial\mu}p(X,\mu)\right] = E[X]E\left[p(X,\mu)\right] + \sigma^2\frac{\partial}{\partial\mu}E\left[p(X,\mu)\right]. \tag{2.2}$$

Now the choice $p(x,\mu) = x^{m-1}$ yields the recursion relation

$$E[X^m] = \left(E[X] + \sigma^2\frac{\partial}{\partial\mu}\right)E[X^{m-1}] = \left(\mu + \sigma^2\frac{\partial \ln\Theta}{\partial\mu} + \sigma^2\frac{\partial}{\partial\mu}\right)E[X^{m-1}],$$

from which part 2 follows.

Finally, using $p(x,\mu) = (x - E[X])^{m-1}$ in (2.2) yields

$$E\left[(X-E[X])^m\right] = \sigma^2\frac{\partial}{\partial\mu}E\left[(X-E[X])^{m-1}\right] + (m-1)\cdot\sigma^2\frac{\partial}{\partial\mu}E[X]\cdot E\left[(X-E[X])^{m-2}\right],$$

which in the case $m = 2$ reduces to

$$\mathrm{Var}[X] = \sigma^2\frac{\partial}{\partial\mu}E[X].$$

Combination of the two last equations yields part 3. □

Corollary 2.3. *Under the assumptions of Theorem 2.2, the following identities hold.*

1. $E[X] = \mu + \sigma^2\dfrac{\partial}{\partial\mu}\ln\Theta$
2. $\mathrm{Var}[X] = \sigma^2 + \sigma^4\dfrac{\partial^2}{\partial\mu^2}\ln\Theta = \sigma^2\dfrac{\partial}{\partial\mu}E[X]$

Proof. Part 1 and the second half of part 2 already appeared in the proof of Theorem 2.2, and combine into the first half of part 2. Alternatively, the equations $E[X] = \mu + \frac{\sigma^2}{\Theta}\frac{\partial\Theta}{\partial\mu}$ and

$$E[X^2] = \mu^2 + \sigma^2 + \frac{2\mu\sigma^2}{\Theta}\frac{\partial\Theta}{\partial\mu} + \frac{\sigma^4}{\Theta}\frac{\partial^2\Theta}{\partial\mu^2}$$

from part 1 of Theorem 2.2 yield

$$\mathrm{Var}[X] = E[X^2] - E[X]^2 = \sigma^2 + \frac{\sigma^4}{\Theta}\frac{\partial^2\Theta}{\partial\mu^2} - \frac{\sigma^4}{\Theta^2}\left(\frac{\partial\Theta}{\partial\mu}\right)^2 = \sigma^2 + \sigma^4\frac{\partial^2}{\partial\mu^2}\ln\Theta.$$

□

Note that in the case where f is constant (and thus Θ is the same constant), part 1 of Theorem 2.2 reduces to the well-known formula

$$E[X_1^m] = \sum_{k=0}^{\lfloor\frac{m}{2}\rfloor}\binom{m}{2k}(2k-1)!!\mu^{m-2k}\sigma^{2k}$$

for the moments of an undeformed normal distribution with random variable $X_1 \sim \mathcal{N}(\mu, \sigma)$, while part 2 becomes

$$E[X_1^m] = \left(\mu + \sigma^2 \frac{\partial}{\partial \mu}\right)^m 1.$$

Notice also the resemblance of the formulas involving $\ln \Theta$ in Corollary 2.3 to formulas from statistical mechanics relating the expectation value and variance of the energy to the partition function.

Example 2.4. In the simple case where $f(x) = x^2$, we have $\Theta = \mu^2 + \sigma^2$, and thus, from Corollary 2.3,

$$E[X] = \mu + \sigma^2 \frac{2\mu}{\mu^2 + \sigma^2}, \qquad \mathrm{Var}[X] = \sigma^2 + \sigma^4 \frac{2(\sigma^2 - \mu^2)}{(\mu^2 + \sigma^2)^2},$$

which in the case $\mu = 0$ reduces to $E[X] = 0$, $\mathrm{Var}[X] = 3\sigma^2$.

Now we define

$$t = \tfrac{1}{2}\sigma^2$$

and consider Θ as a function of μ and t,

$$\Theta(\mu, t) = \frac{1}{\sqrt{4\pi t}} \int f(x) \mathrm{e}^{-\frac{(x-\mu)^2}{4t}} \mathrm{d}x. \tag{2.3}$$

Thus, Θ is the convolution of f with the heat kernel, *i.e.*, a generalized Weierstrass transform of f. This means that Θ is a solution of the one-dimensional heat equation for an infinite rod with μ as the space variable,[4]

$$\frac{\partial \Theta}{\partial t} = \frac{\partial^2 \Theta}{\partial \mu^2}, \tag{2.4}$$

with the initial condition

$$\Theta(\mu, 0) = f(\mu). \tag{2.5}$$

(Here, to avoid subtleties, f is assumed to be twice differentiable.) Conversely, any non-negative solution $u(x, t)$ of the heat equation $u_t = u_{xx}$ defined for $t \geq 0$ corresponds to a deformed normal distribution given by (2.1) where $f(x) = u(x, 0)$ and $\Theta = u(\mu, \frac{1}{2}\sigma^2)$.

For later use we note that, as a result of (2.4), the function $v = \ln \Theta$ satisfies the partial differential equation

$$v_t = v_{\mu\mu} + (v_\mu)^2. \tag{2.6}$$

Example 2.5. If $f(x) = 1 + \sin x$, then $\mathcal{N}(0, 1; f)$ is the distribution with density function $p(x) = (1 + \sin x) \cdot g(x; 0, 1)$, since in this case the normalizing constant equals 1. The density function is shown in Figure 1. The mean can be seen to lie somewhere to the right of zero, while the variance appears to be somewhat less than 1. To get at the exact values, we first consider the general case $\mathcal{N}(\mu, \sigma; f)$ for arbitrary μ and σ, and calculate

$$\Theta = 1 + \mathrm{e}^{-t} \sin \mu = 1 + \exp\left(-\tfrac{1}{2}\sigma^2\right) \sin \mu,$$

[4] Equation (2.4) can easily be derived from (2.3), noting that Lemma A.2 has an analogue that applies to differentiation of the integral with respect to t.

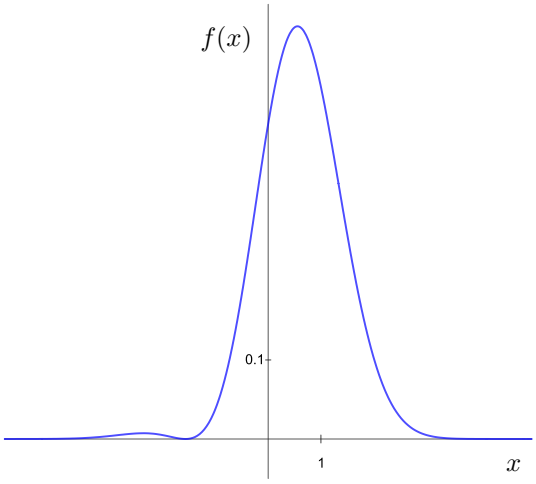

FIGURE 1. Density function of the distribution $\mathcal{N}(0,1;f)$, where $f(x) = 1 + \sin x$.

which clearly satisfies (2.4) and (2.5). Using Corollary 2.3, and setting $\mu = 0$, $\sigma = 1$, gives

$$E[X] = \mu + \frac{\sigma^2 \cos\mu}{\mathrm{e}^{\frac{\sigma^2}{2}} + \sin\mu} = \frac{1}{\sqrt{e}} \quad \text{and} \quad \mathrm{Var}[X] = \sigma^2 - \sigma^4 \frac{\mathrm{e}^{\frac{\sigma^2}{2}} \sin\mu + 1}{(\mathrm{e}^{\frac{\sigma^2}{2}} + \sin\mu)^2} = 1 - \frac{1}{e}. \tag{2.7}$$

3. Series expansions

When the deforming function is an entire function, and thus representable by a power series with infinite convergence radius, power series expansions can also be obtained for Θ and the moments of X. The procedure involves an integration over the reals, removing x from the power expansion and introducing powers of σ^2 instead. The basic, relevant result is stated in Lemma A.3 of Appendix A. The first part of Theorem 3.1 below is an immediate consequence of this lemma, while part 2 requires a short proof. Note the difference between the assumptions made in Theorem 2.2 and Theorem 3.1. In Theorem 2.2, the deforming function f must be such that Θ exists for all μ in an open interval, but f itself need not be smooth, or even continuous (at any point). In contrast, Theorem 3.1 requires f to be an entire function, but is only stated for some specific value of μ (and σ), and may even apply in cases where f is an (entire) function for which Θ only exists for this single value of μ.

Theorem 3.1. *Let $X \sim \mathcal{N}(\mu, \sigma; f)$, where f is an entire function.*

1. *The following identity holds if the series to the right converges.*

$$\Theta = \sum_{n=0}^{\infty} \frac{f^{(2n)}(\mu)}{n!\, 2^n} \sigma^{2n}.$$

2. *Suppose $E[X^m] = \frac{1}{\Theta} \int x^m f(x) g(x;\mu,\sigma)\mathrm{d}x$ exists, or that m is even. Then the identity below holds whenever the series to the right converges.*

$$E[X^m] = \mu^m + \frac{1}{\Theta} \sum_{n=1}^{\infty} \sigma^{2n} \sum_{k=1}^{\min(m,2n)} \binom{m}{k} \frac{(2n-1)!!}{(2n-k)!} \mu^{m-k} f^{(2n-k)}(\mu).$$

Proof. Let $F(x)$ be $(x^m - \mu^m)f(x)$. Then also F is an entire function, and if m is even, F is eventually (in both x-directions) non-negative. Moreover,

$$\begin{aligned} F^{(n)}(x) &= \sum_{k=0}^{n} \binom{n}{k} (x^m - \mu^m)^{(k)} f^{(n-k)}(x) \\ &= (x^m - \mu^m)f^{(n)}(x) + \sum_{k=1}^{\min(m,n)} \binom{n}{k} \frac{m!}{(m-k)!} x^{m-k} f^{(n-k)}(x), \end{aligned}$$

so

$$F^{(n)}(\mu) = \sum_{k=1}^{\min(m,n)} \binom{m}{k} \frac{n!}{(n-k)!} \mu^{m-k} f^{(n-k)}(\mu).$$

The result now follows by noting that $\frac{1}{\Theta}\int F(x)g(x;\mu,\sigma)\mathrm{d}x = E[X^m] - \mu^m$, and applying Lemma A.3 to the function F, to obtain the identity

$$\int F(x)g(x;\mu,\sigma)\mathrm{d}x = \sum_{n=1}^{\infty} \frac{\sigma^{2n}}{(2n)!!} \sum_{k=1}^{\min(m,2n)} \binom{m}{k} \frac{(2n)!}{(2n-k)!} \mu^{m-k} f^{(2n-k)}(\mu),$$

where the term with $n = 0$ vanishes since $F(\mu) = 0$. □

Note that $\Theta = \sum_{n=0}^{\infty} \frac{1}{n!} f^{(2n)}(\mu) t^n$, which satisfies the heat equation (2.4) when differentiated termwise.

Corollary 3.2. *Let $X \sim \mathcal{N}(\mu, \sigma; f)$, where f is an entire function.*

1. *Suppose $E[X]$ exists. Then $E[X] = \mu + \dfrac{\sigma^2}{\Theta} \displaystyle\sum_{n=0}^{\infty} \frac{f^{(2n+1)}(\mu)}{n!\, 2^n} \sigma^{2n}$ if the series converges.*
2. *If the series below converge, then $\mathrm{Var}[X]$ exists and the identity holds.*

$$\mathrm{Var}[X] = \sigma^2 + \frac{\sigma^4}{\Theta} \sum_{n=0}^{\infty} \frac{f^{(2n+2)}(\mu)}{n!\, 2^n} \sigma^{2n} - \frac{\sigma^4}{\Theta^2} \left(\sum_{n=0}^{\infty} \frac{f^{(2n+1)}(\mu)}{n!\, 2^n} \sigma^{2n} \right)^2$$

Proof. Part 1 follows immediately from the previous theorem. For part 2, assume that both series appearing in the formula for $\mathrm{Var}[X]$ converge. It then follows from Theorem 3.1 that $E[X^2]$ exists and is given by

$$E[X^2] = \mu^2 + \frac{\sigma^2}{\Theta} \sum_{n=0}^{\infty} \frac{f^{(2n)}(\mu)}{n!\, 2^n} (2n+1)\sigma^{2n} + \frac{2\mu\sigma^2}{\Theta} \sum_{n=0}^{\infty} \frac{f^{(2n+1)}(\mu)}{n!\, 2^n} \sigma^{2n}$$

if the first series is convergent (as the second one is assumed to be). Now straightforward manipulations yield

$$\sum_{n=0}^{\infty} \frac{f^{(2n)}(\mu)}{n!\, 2^n} (2n+1)\sigma^{2n} = \sum_{n=0}^{\infty} \frac{f^{(2n)}(\mu)}{n!\, 2^n} \sigma^{2n} + \sigma^2 \sum_{n=0}^{\infty} \frac{f^{(2n+2)}(\mu)}{n!\, 2^n} \sigma^{2n},$$

where the last series converges by assumption. By part 1 of Theorem 3.1, the series $\sum_{n=0}^{\infty} \frac{f^{(2n)}(\mu)}{n!\, 2^n} \sigma^{2n}$ can be replaced by Θ if it converges. To see that it does, rewrite it as $f(\mu) + \sum_{n=0}^{\infty} \frac{\sigma^2}{2n+2} \cdot \frac{f^{(2n+2)}(\mu)}{n!\, 2^n} \sigma^{2n}$, which converges

by Abel's test since its terms are the termwise products of the monotone and bounded sequence $\{\frac{\sigma^2}{2n+2}\}$ and a convergent series. Hence, $E[X^2]$ does exist and is given by

$$E[X^2] = \mu^2 + \sigma^2 + \frac{\sigma^4}{\Theta}\sum_{n=0}^{\infty}\frac{f^{(2n+2)}(\mu)}{n!\,2^n}\sigma^{2n} + \frac{2\mu\sigma^2}{\Theta}\sum_{n=0}^{\infty}\frac{f^{(2n+1)}(\mu)}{n!\,2^n}\sigma^{2n},$$

while $E[X]$ (which exists because $E[X^2]$ does) is given by the series in part 1 (which converges by assumption). From this, part 2 follows by the identity $\mathrm{Var}[X] = E[X^2] - E[X]^2$. □

Note that the formulas in Corollary 3.2 are in accordance with the formulas $E[X] = \mu + \sigma^2\frac{\partial \ln\Theta}{\partial\mu}$ and $\mathrm{Var}[X] = \sigma^2 + \sigma^4\frac{\partial^2\ln\Theta}{\partial\mu^2}$ from Corollary 2.3 combined with termwise differentiation of the series for Θ from Theorem 3.1.

Example 3.3. In Example 2.5, the derivatives $f'(0), f''(0), f'''(0)$, *etc.*, follow the pattern 1, 0, −1, 0, 1, 0, *etc.*, while $f(0)$ itself equals 1, hence it can be seen from Corollary 3.2 that

$$E[X] = \sum_{n=0}^{\infty}\frac{(-1)^n}{n!\,2^n} = \mathrm{e}^{-\frac{1}{2}} \quad\text{and}\quad \mathrm{Var}[X] = 1 + 0 - \left(\sum_{n=0}^{\infty}\frac{(-1)^n}{n!\,2^n}\right)^2 = 1 - \mathrm{e}^{-1},$$

reproducing the results in (2.7).

If the series for Θ and $E[X^m]$ given in Theorem 3.1 both converge, then they can be combined into a power series in σ^2 for $E[X^m]$ where all coefficients are independent of σ, that converges for sufficiently small σ. For example, if $f(\mu) \neq 0$, then the power series for Θ has a nonzero constant term, and a multiplicatively inverse power series can be found. Theorem 3.1 and Corollary 3.2 can then be combined to provide power series in σ^2 for $E[X]$ and $\mathrm{Var}[X]$.

Proposition 3.4. *Let $X \sim \mathcal{N}(\mu, \sigma; f)$, where f is an entire function, and $f(\mu) \neq 0$. Under the assumptions of part 2 of Corollary 3.2, and writing a_n for $f^{(n)}(\mu)$, the following identities hold for sufficiently small σ.*

$$E[X] = \mu + \frac{a_1}{a_0}\sigma^2 + \frac{1}{2}\left(\frac{a_3}{a_0} - \frac{a_1a_2}{a_0^2}\right)\sigma^4 + \frac{1}{8}\left(\frac{a_5}{a_0} - \frac{a_1a_4}{a_0^2} - \frac{2a_2a_3}{a_0^2} + \frac{2a_1a_2^2}{a_0^3}\right)\sigma^6 + O(\sigma^8)$$
$$\mathrm{Var}[X] = \sigma^2 + \left(\frac{a_2}{a_0} - \frac{a_1^2}{a_0^2}\right)\sigma^4 + \frac{1}{2}\left(\frac{a_4}{a_0} - \frac{a_2^2}{a_0^2} - \frac{2a_1a_3}{a_0^2} + \frac{2a_1^2a_2}{a_0^3}\right)\sigma^6 + O(\sigma^8)$$

Note that these series are related by the formula $\mathrm{Var}[X] = \sigma^2\frac{\partial E[X]}{\partial\mu}$ from Corollary 2.3.

4. Convergence issues

The conditions in the results of the previous section are, in general, all necessary. Even if the Taylor series for f around μ converges everywhere, the above sums for the moments may not. Take for example the entire function $f(x) = \mathrm{e}^{-x^4}$. Then $\int f(x)g(x;\mu,\sigma)\mathrm{d}x$ converges to a finite value for all $\sigma > 0$ and all μ, hence by Theorem 2.2 all the moments of $X \sim \mathcal{N}(\mu, \sigma; f)$ exist. Also, the Taylor series for f around $x = 0$ equals $\sum_{n=0}^{\infty}\frac{1}{n!}\left(-x^4\right)^n$, which converges everywhere, but plugging this into Theorem 3.1 with $\mu = 0$ yields a series expansion for Θ,

$$\sum_{n=0}^{\infty}\frac{(-1)^n\cdot(4n)!}{(2n)!\cdot 2^{2n}\cdot n!}\sigma^{4n} = \sum_{n=0}^{\infty}\frac{(-1)^n\cdot(4n-1)!!}{n!}\sigma^{4n}, \tag{4.1}$$

which diverges for all $\sigma \neq 0$. Similar examples of entire functions f can be given for which $E[X]$ and the corresponding series may converge or diverge independently.

A different limitation on the use of the results in the previous section arises from the fact that the function f used in a deformed normal distribution may have a Taylor series at μ that converges to $f(x)$ only on an interval around μ. In this case, the infinite power series expansions in σ^2 of Θ, $E[X]$, $\mathrm{Var}[X]$, *etc.*, may be incorrect or divergent. For example, the function $f(x) = \frac{1}{1+x^2}$ has a Taylor series $\sum_{n=0}^{\infty} \left(-x^2\right)^n$ at $x = \mu = 0$ with a finite radius of convergence, while the corresponding series expansion $\sum_{n=0}^{\infty} (-1)^n (2n-1)!!\, \sigma^{2n}$ for Θ diverges for all $\sigma \neq 0$.

However, in all such cases, partial sums may still yield usable approximations, even if arbitrarily large errors will sooner or later turn up when sufficiently many terms are included in the sum.[5] To see the reason for this, note that in all results in Section 3, the assumption about convergence to $f(x)$ everywhere is essentially used to obtain identities of the form

$$\int x^m \cdot f(x) \cdot g(x;\mu,\sigma)\,\mathrm{d}x = \int x^m \cdot \sum_{n=0}^{\infty} \frac{f^{(n)}(\mu)}{n!}(x-\mu)^n \cdot g(x;\mu,\sigma)\,\mathrm{d}x.$$

Now if the radius of convergence to $f(x)$ is wide relative to σ, such that some Taylor polynomial $s_k(x) = \sum_{n=0}^{k} \frac{f^{(n)}(\mu)}{n!}(x-\mu)^n$ gives a good approximation to $f(x)$ for all x within several multiples of σ around μ, then the values of x for which $s_k(x)$ is *not* a good approximation may contribute very little to the integrals, as both $x^m f(x)$ and $x^m s_k(x)$ are being multiplied by exceedingly small values of $g(x;\mu,\sigma)$. With $m+k = 2n$ or $m+k = 2n+1$, the error introduced is $O(\sigma^{2n+2})$, and thus, for a given n, the relative error will be negligible for sufficiently small σ. Consequently, even in cases where the Taylor series converges to $f(x)$ only on a finite interval, contrary to the assumption in Theorem 3.1 and Corollary 3.2, the results of Section 3 can still be used by truncating the series and using big O notation.

We encounter an example of this in Section 6 and onward, where short partial sums, although based on divergent series, yield excellent approximations. In standard terminology, these divergent power series in σ^2 are asymptotic expansions at $\sigma^2 = 0$.

5. Product of a normal variable with another variable

When the product $Z = XY$ of a normally distributed random variable X with another, independent continuous random variable Y is known and this product is nonzero, $X|Z = z$ is naturally equipped with a deformed normal distribution. To see this, consider first the general case for arbitrary independent continuous random variables X, Y. For a fixed $x \neq 0$, the linear function $Z = xY$ of Y has, in general, the probability density function $\frac{1}{|x|}p_Y(\frac{z}{x})$ (*cf.* [3] pp. 37–38), where p_Y is the probability density function for Y. Multiplying this with the probability density function for X, one obtains the general expression $\frac{1}{|x|}p_Y(\frac{z}{x})p_X(x)$ for the joint probability density function for $Z = XY$ and X when X, Y are independent. (See also [3] pp. 60–61.) Now if $X \sim \mathcal{N}(\mu_X, \sigma_X)$, the joint probability density function for X and Z is

$$p_{X,Z}(x,z) = \frac{1}{|x|} \cdot p_Y\left(\frac{z}{x}\right) \cdot g(x;\mu_X,\sigma_X),$$

[5] For example, if $\sigma = 0.2$, then the series (4.1), although divergent, provides an approximation to the correct value $\Theta \approx 0.9953$. The partial sum with three terms already has four correct digits, and ten terms give ten correct digits. The terms a_n first decrease in absolute value towards the minimum value $a_{40} \approx 10^{-18}$, but then steadily increase; for example, $a_{100} \approx 10^{-4}$, while a_{130}, as well as the corresponding partial sum, are larger than 10^{10}.

and the marginal density function for Z is found by integrating this with respect to x over the set of all reals. Hence the density function for X when $Z = z$ can be written as

$$p_{X|z}(x) = \frac{1}{\Theta} f(x) g(x; \mu_X, \sigma_X), \tag{5.1}$$

where

$$\Theta = p_Z(z) = \int f(x) g(x; \mu_X, \sigma_X) \mathrm{d}x, \tag{5.2}$$

$$f(x) = \frac{1}{|x|} p_Y\left(\frac{z}{x}\right), \tag{5.3}$$

and $f(0)$ is defined arbitrarily. If $z \neq 0$, the integral for Θ exists and is nonzero. To see this, note that since $\Theta = p_Z(z)$, the role of the variables X, Y in $Z = XY$ can be switched to yield the identity $\Theta = \int \frac{1}{|y|} g(\frac{z}{y}; \mu_X, \sigma_X) p_Y(y) \mathrm{d}y$, where $\frac{1}{|y|} g(\frac{z}{y}; \mu_X, \sigma_X)$ is positive for $y \neq 0$ as well as bounded. Moreover, the measurability of p_Y implies that f is measurable as well. Thus, $p_{X|z}$ is the probability density function of a variable $X|z$ that has the deformed normal distribution $\mathcal{N}(\mu_X, \sigma_X; f)$, with f given by (5.3).

In the case $z = 0$, (5.1) does not define a probability density function, because, as (5.2) and (5.3) show, Θ vanishes if $p_Y(0) = 0$, and else, the integral for Θ diverges. This divergence reflects a general feature of product distributions, and in Appendix B we show – under quite weak assumptions on the variables X and Y – that with diminishing $|z|$, $X|z$ collapses towards a constant at 0, while $E[\,|X| \mid z\,]$ and $\mathrm{Var}[X|z]$ vanish. In many cases, however, this plays out only for values of z in a very narrow interval around 0, as a consequence of the underlying divergence being logarithmic (*cf.* Appendix C). In the cases to be considered in the rest of this paper, we shall see that the special properties of $p_{X|z}$ near $z = 0$ can be ignored for practical purposes.

6. Product of two normal variables

The probability distribution of the product $Z = XY$ of two normal variables X and Y was studied by Craig [4] and Aroian [5]. In this and the subsequent sections we look instead at the conditional probability distribution of one of the factor variables, say X, when the value of XY is known. Thus we proceed to the special case of the situation in the previous section, now with both of the independent variables being normally distributed, *i.e.*, $X \sim \mathcal{N}(\mu_X, \sigma_X)$ and $Y \sim \mathcal{N}(\mu_Y, \sigma_Y)$.[6] Based on the results in Sections 2, 3, and 5, as well as the assumption of a small coefficient of variation for one of the variables, we find approximation formulas for the expectation and variance of $X|z$ and $Y|z$.

As shown in Section 5, the conditional probability density function $p_{X|z}(x)$ for $z \neq 0$ and $x \neq 0$ is given by (5.1)–(5.3), the latter equation now yielding the deforming function

$$f(x) = \frac{1}{|x|} \cdot g\left(\frac{z}{x}; \mu_Y, \sigma_Y\right) = \frac{1}{\sqrt{2\pi}\sigma_Y} \frac{1}{|x|} \mathrm{e}^{-\frac{1}{2}\left(\frac{z}{x\sigma_Y} - \frac{\mu_Y}{\sigma_Y}\right)^2}. \tag{6.1}$$

The conditional probability density function $p_{Y|z}(y)$ is obtained in a completely similar manner. The factor Θ appearing in the densities for the two resulting deformed normal distributions is the same in both cases, as it is equal to the marginal probability density function for Z, *i.e.*, $p_Z(z)$.

[6] The results in this section may be compared with the much simpler case $Z = X + Y$, where $X|z$ is normally distributed with mean $\frac{\mu_X \sigma_Y^2 + (z - \mu_Y)\sigma_X^2}{\sigma_X^2 + \sigma_Y^2}$ and standard deviation $\frac{\sigma_X \sigma_Y}{\sqrt{\sigma_X^2 + \sigma_Y^2}}$. This follows by observing that when X and Y are normal and independent, $(X, X + Y)$ has a bivariate normal distribution with covariance σ_X^2, and using standard results about the conditional distribution of one variable in a bivariate normal distribution when the value of the other is given.

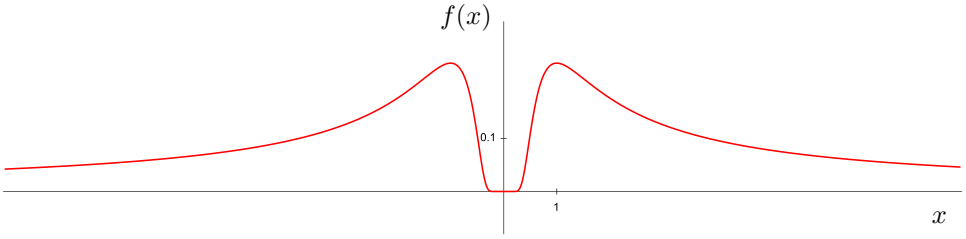

FIGURE 2. Graph of the function $f(x) = \frac{1}{|x|} g(\frac{z}{x}; 0, \sigma_Y)$, with global maxima of $\frac{1}{\sqrt{2\pi e}|z|}$ at $x = \pm\frac{z}{\sigma_Y}$, shown here with $\sigma_Y = 1$, $z = 1$.

With the definition $f(0) = 0$, f and $p_{X|z}$ are smooth functions (the graph of f is shown in Figure 2 for $\mu_Y = 0$, in which case it is symmetric). However, f is not analytic at $x = 0$, because, regarded as a complex function, it has a singularity at 0. Thus, the Taylor series of f at $x = \mu_X$ has a convergence radius of $|\mu_X|$. Hence the results in Section 3 will not be directly applicable. Indeed, with this f, the series for Θ appearing in Theorem 3.1 is, in fact, divergent for nonzero σ_X (*cf.* Appendices D and E). Nevertheless, the previous results can be used as guides to obtain good approximations under certain conditions, as discussed in Section 4.

Suppose X has a low coefficient of variation $\frac{\sigma_X}{|\mu_X|}$. Then the approximations will be based on series expansions involving the derivatives of f. For a function of the form

$$f(x) = \frac{c}{|x|} \cdot \mathrm{e}^{-\frac{1}{2}\left(\frac{a}{x} - b\right)^2},$$

the n-th derivative is given by

$$f^{(n)}(x) = \frac{1}{x^n} Q_n\left(\frac{a}{x}, b\right) f(x), \tag{6.2}$$

where $Q_n(u, v)$ is the polynomial of degree $2n$ in two variables defined recursively by

$$Q_0(u, v) = 1, \quad Q_n(u, v) = \left(u^2 - uv - n - u\frac{\partial}{\partial u}\right) Q_{n-1}(u, v).$$

In the particular f under consideration, $c = \frac{1}{\sqrt{2\pi}\,\sigma_Y}$, $a = \frac{z}{\sigma_Y}$, and $b = \frac{\mu_Y}{\sigma_Y}$. With the dimensionless parameters

$$\omega = \frac{z}{\mu_X\,\sigma_Y}, \qquad \kappa = \frac{\mu_Y}{\sigma_Y}, \qquad \lambda = \frac{\sigma_X}{|\mu_X|}, \tag{6.3}$$

we have $f^{(n)}(\mu_X) = \frac{1}{\mu_X^n} f(\mu_X) Q_n(\omega, \kappa)$, and thus, by Theorem 3.1 and (6.2),

$$p_Z(z) = \Theta = f(\mu_X)\left[1 + \tfrac{1}{2}Q_2(\omega, \kappa)\lambda^2 + \tfrac{1}{8}Q_4(\omega, \kappa)\lambda^4 + \tfrac{1}{48}Q_6(\omega, \kappa)\lambda^6 + O(\lambda^8)\right]. \tag{6.4}$$

Note that $f(\mu_X) = g(z; \mu_X\mu_Y, |\mu_X|\sigma_Y)$, which means that (6.4) gives the probability density of $Z = XY$ as that of the normal distribution $\mathcal{N}(\mu_X\mu_Y, |\mu_X|\sigma_Y)$ deformed by the sum in brackets.[7]

[7]With the variable $W = \frac{Z}{\mu_X\sigma_Y}$, (6.4) may be rewritten as $p_W(\omega) = g(\omega; \kappa, 1)F(\omega, \kappa, \lambda)$, where F is the sum in brackets. This is the probability density function of a normal distribution deformed by a function that depends on the mean (κ) of the normal distribution. Theorems 2.2 and 3.1 may be adapted to include such cases, where (2.1) is generalized to $p_X(x) = \frac{1}{\Theta} g(x; \mu, \sigma)F(x, \mu)$, by treating the latter μ as a constant.

Writing $f(\mu_X)$ as $\frac{1}{|\mu_X|\sigma_Y} g(\omega;\kappa,1)$ and using the series expansion $\ln(1+S) = S - \frac{1}{2}S^2 + \frac{1}{3}S^3 - \dots$ for the logarithm of the sum in brackets in (6.4), we get

$$\ln\Theta = -\ln(\sqrt{2\pi}|\mu_X|\sigma_Y) - \tfrac{1}{2}(\omega-\kappa)^2 + \tfrac{1}{2}Q_2\lambda^2 + \tfrac{1}{8}(Q_4 - Q_2^2)\lambda^4 + \tfrac{1}{48}(Q_6 - 3Q_2Q_4 + 2Q_2^3)\lambda^6 + O(\lambda^8). \quad (6.5)$$

Although this expression contains products of λ^{2n} and polynomials in ω and κ of degree $4n$ (like $Q_{2n}(\omega,\kappa)$), after simplification, the remaining polynomial has degree $2n+2$ (*cf.* Appendix E).

To obtain expressions for the expectation and variance of $X|z$ that include terms up to fourth order in λ, only terms up to second order in λ are needed in the expression for $\ln\Theta$, *i.e.*,

$$\ln\Theta = C - \ln(|\mu_X|\sigma_Y) - \tfrac{1}{2}\omega^2 + \omega\kappa - \tfrac{1}{2}\kappa^2 + \left(\tfrac{1}{2}\omega^4 - \omega^3\kappa + \tfrac{1}{2}\omega^2\kappa^2 - \tfrac{5}{2}\omega^2 + 2\omega\kappa + 1\right)\lambda^2 + O(\lambda^4). \quad (6.6)$$

Applying Corollary 2.3 (and keeping in mind the μ_X-dependence of both ω and λ), we arrive at

$$E[X|z] = \mu_X\left[1 + (\omega^2 - \omega\kappa - 1)\lambda^2 + (-3\omega^4 + 5\omega^3\kappa - 2\omega^2\kappa^2 + 10\omega^2 - 6\omega\kappa - 2)\lambda^4 + O(\lambda^6)\right], \quad (6.7)$$

$$\mathrm{Var}[X|z] = \sigma_X^2\left[1 + (-3\omega^2 + 2\omega\kappa + 1)\lambda^2 + (21\omega^4 - 30\omega^3\kappa + 10\omega^2\kappa^2 - 50\omega^2 + 24\omega\kappa + 6)\lambda^4 + O(\lambda^6)\right]. \quad (6.8)$$

Higher order terms may be found by including more terms from the series for $\ln\Theta$, where the fourth- and sixth-order terms are

$$\left(-\tfrac{3}{2}\omega^6 + 4\omega^5\kappa - \tfrac{7}{2}\omega^4\kappa^2 + \omega^3\kappa^3 + \tfrac{47}{4}\omega^4 - 18\omega^3\kappa + \tfrac{13}{2}\omega^2\kappa^2 - \tfrac{37}{2}\omega^2 + 10\omega\kappa + \tfrac{5}{2}\right)\lambda^4 \quad (6.9)$$

and

$$\begin{aligned}&\left(\tfrac{13}{2}\omega^8 - 22\omega^7\kappa + \tfrac{55}{2}\omega^6\kappa^2 - 15\omega^5\kappa^3 + 3\omega^4\kappa^4 - \tfrac{443}{6}\omega^6 + 168\omega^5\kappa\right.\\ &\quad\left. - 123\omega^4\kappa^2 + \tfrac{86}{3}\omega^3\kappa^3 + 228\omega^4 - 287\omega^3\kappa + \tfrac{165}{2}\omega^2\kappa^2 - \tfrac{353}{2}\omega^2 + 74\omega\kappa + \tfrac{37}{3}\right)\lambda^6.\end{aligned} \quad (6.10)$$

To find expressions for the expectation and variance of $Y|z$ with the same number of terms as for $X|z$ above, we need to include the fourth-order term (6.9) of the series for $\ln\Theta$, and, for the variance, the sixth-order term (6.10) as well. Using Corollary 2.3 again, but this time with $\mu = \mu_Y$ and $\sigma = \sigma_Y$, we obtain

$$\begin{aligned}E[Y|z] &= \frac{z}{\mu_X}\left[1 + (-\omega^2 + \omega\kappa + 2)\lambda^2\right.\\ &\qquad\left. + (4\omega^4 - 7\omega^3\kappa + 3\omega^2\kappa^2 - 18\omega^2 + 13\omega\kappa + 10)\lambda^4 + O(\lambda^6)\right], \quad (6.11)\\ \mathrm{Var}[Y|z] &= \frac{z^2}{\mu_X^2}\lambda^2\left[1 + (-7\omega^2 + 6\omega\kappa + 13)\lambda^2\right.\\ &\qquad\left. + (55\omega^4 - 90\omega^3\kappa + 36\omega^2\kappa^2 - 246\omega^2 + 172\omega\kappa + 165)\lambda^4 + O(\lambda^6)\right]. \quad (6.12)\end{aligned}$$

The polynomials in ω and κ appearing in (6.7), (6.8), (6.11), and (6.12), display some regularities that carry on to higher powers of λ^2 and are discussed in Appendix E. Note that since $\sigma_X^2 = \mu_X^2\lambda^2$ and $\frac{z}{\mu_X} = \sigma_Y\omega$, the corresponding expansions for the dimensionless variables $\hat{X} = \frac{X}{\mu_X}$ and $\hat{Y} = \frac{Y}{\sigma_Y}$ can be written entirely in terms of the dimensionless parameters ω, κ, and λ. Similar approximations for the moments $E[X^m|z]$ and $E[Y^m|z]$ may be found by using part 2 of Theorem 2.2 combined with termwise differentiation.

The goodness of the approximations presented in this section depends on the parameter settings and may be analyzed as in Sections 7–9, where (6.7) and (6.8) are further investigated. As indicated by the increasing coefficients in the series for $\ln\Theta$, both λ as well as the products $|\omega|\lambda$ and $|\kappa|\lambda$ must be small for the terms of the series to become small, and thus for the approximation to work. The last of these requirements comes

from the condition $|\omega\kappa|\lambda^2 \ll 1$ combined with the requirement that the approximation should work well around $z = \mu_X\mu_Y$, or $\omega = \kappa$, and means that the coefficient of variation for X, $\lambda = \frac{\sigma_X}{|\mu_X|}$, is small compared with the one for Y, $\frac{1}{|\kappa|} = \frac{\sigma_Y}{|\mu_Y|}$.

However, what will not be apparent from the polynomial approximations above for Θ, $\ln\Theta$, and the conditional expectations and variances, is a behavior that arises when $|\omega|$ takes on extremely small values, typically far smaller than 10^{-1000}. As discussed in Section 5, the integral for Θ does not converge for $z = 0$. It follows from Theorem B.1 of Appendix B that when z approaches 0, $E[|X|^n \mid z]$ vanishes, and that the cumulative distribution function $P(X \le x \mid Z = 0)$, when defined as the limit of $P(X \le x \mid |Z| \le |z|)$ as $|z|$ approaches 0, equals the unit step function. This can also be inferred from (5.1), (5.2), and (6.1), from which it follows that with diminishing $|z|$, the graph of $p_{X|z}$ evolves into two spikes approaching the y-axis while the rest of the graph approaches zero and falls off exponentially. Thus, the conditional distribution for $X|Z = 0$ may be regarded as a degenerate distribution with probability density function $p_{X|Z=0}(x)$ equal to the delta function $\delta(x)$.

For small λ, however, all this only plays out in an exceedingly narrow interval for ω around 0, as long as $|\kappa|$ is small compared to $\frac{1}{\lambda}$. We show in Appendix C, in the case $\mu_Y = 0$, that for small values of λ (such that X for many purposes can be approximated with a random variable $\tilde{X}$ similar to X but with zero probability around 0), Θ initially stabilizes around the value $g(0; 0, \sigma_Y) \cdot E\left[\frac{1}{|\tilde{X}|}\right]$ when z approaches 0, and does not depart from here by a factor of more than $1 + 10^{-17}$ as long as λ stays below 0.1 and $|\omega|$ stays above 10^{-2000}.

7. Product of two normal variables – special case

This section investigates in more detail the conditional probability distribution of X when the product $Z = XY$ is known, in the case where Y has zero mean, *i.e.*, $\mu_Y = 0 = \kappa$. Then, as shown in Section 6, $X|z \sim \mathcal{N}(\mu_X, \sigma_X; f)$ with the deforming function given by (6.1), which now simplifies to

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma_Y}\frac{1}{|x|}\mathrm{e}^{-\frac{1}{2}\left(\frac{z}{x\sigma_Y}\right)^2}.$$

In this simpler case, the approximating polynomials for $E[X|z]$ and $\mathrm{Var}[X|z]$ can be given more explicitly. In Appendix D we show that the function $h(x) = \frac{1}{|x|}\mathrm{e}^{\frac{a}{x^2}}$ has the n-th derivative

$$h^{(n)}(x) = (-1)^n \frac{1}{x^n} P_n\left(\frac{a}{x^2}\right) h(x),$$

where $P_n(x)$ is an n-th degree polynomial with positive integer coefficients, given by the equation

$$P_n(x) = \sum_{k=0}^{n} \begin{bmatrix} n+1 \\ k+1 \end{bmatrix} 2^k T_k(x). \tag{7.1}$$

Here, $\begin{bmatrix} n \\ k \end{bmatrix}$ denote unsigned Stirling numbers of the first kind and $T_k(x)$ is the Touchard polynomial of degree k. Thus,

$$f^{(n)}(\mu_X) = (-1)^n \frac{1}{\mu_X^n} P_n\left(-\tfrac{1}{2}\omega^2\right) f(\mu_X).$$

When this is plugged into part 1 of Theorem 3.1 and part 1 of Corollary 3.2, the factor $f(\mu_X)$ in $f^{(n)}(\mu_X)$ cancels out, and $E[X|z]$ is approximately equal to a finite sum obtained from truncation of the power series

expansion of

$$\mu_X \left(1 - \lambda^2 \frac{\sum_{n=0}^{\infty} \frac{1}{n!\,2^n} P_{2n+1}\left(-\frac{1}{2}\omega^2\right) \lambda^{2n}}{\sum_{n=0}^{\infty} \frac{1}{n!\,2^n} P_{2n}\left(-\frac{1}{2}\omega^2\right) \lambda^{2n}} \right). \tag{7.2}$$

Using standard power series manipulation, one now obtains the partial expansion

$$\begin{aligned} E[X|z] = \mu_X \big[& 1 + (\omega^2 - 1)\lambda^2 + (-3\omega^4 + 10\omega^2 - 2)\lambda^4 \\ & + (15\omega^6 - 94\omega^4 + 111\omega^2 - 10)\lambda^6 + O(\lambda^8) \big], \end{aligned} \tag{7.3}$$

in accordance with (6.7). Applying part 2 of Corollary 2.3 to Eq. (7.3) yields

$$\begin{aligned} \mathrm{Var}[X|z] = \sigma_X^2 \big[& 1 + (-3\omega^2 + 1)\lambda^2 + (21\omega^4 - 50\omega^2 + 6)\lambda^4 \\ & + (-165\omega^6 + 846\omega^4 - 777\omega^2 + 50)\lambda^6 + O(\lambda^8) \big]. \end{aligned} \tag{7.4}$$

Below we consider how these polynomials fare when used to estimate the true values $E[X|z]$ and $\mathrm{Var}[X|z]$. As expected, the approximations deteriorate somewhat for larger values of λ and $|\omega|$. The behavior at values of $|\omega|$ below 10^{-2000} discussed in the end of Section 6 is not shown in either of the figures with values of ω along one of the axes. For example, the graphs in Figures 3 and 4 for conditional expectation and variance as functions of ω indicate nicely rounded curves around $\omega = 0$, and this is indeed the behavior that is seen as long as $|\omega|$ is kept above extremely low values.

Figure 3 compares $E[X|z]$ to the approximations obtained by the first finite, partial sums from (7.3), in the case that λ equals 0.03. A visible difference between expectation and fourth-degree approximation turns up at about $\omega = 7$. Such an ω-value is quite a bit out in the distribution, as with the present values of λ the corresponding variable $\frac{Z}{\mu_X\,\sigma_Y}$ will have a standard deviation very close to 1.[8] Similarly, Figure 4 compares the standard deviation of $X|z$ with the approximations obtained by taking the square root of partial sums for the variance in (7.4).

When looking to evaluate the accuracy of these approximations at various values of ω and λ, one should keep in mind that the variations in $E[X|z]$ and $SD[X|z]$ for different values of z are not very large to begin with. In Figure 3, for instance, $E[X|z]$ never differs from $E[X]$ by more than about 7%. Still, an estimation error for $E[X|z]$ at this level would be entirely unacceptable in this case, as it would exceed both σ_X and all possible values of $SD[X|z]$. What is required, then, are errors in the estimation of $E[X|z]$ that are small, not as a percentage of their own, correct values, but in relation to $SD[X|z]$. Figure 5 shows a contour plot for $\frac{err}{SD[X|z]}$, where err is the absolute value of the difference between $E[X|z]$ and the approximation obtained by the fourth-degree polynomial in λ. Here we see for instance that as long as $|\omega| < 4$ and $\lambda < 0.03$, the error in the estimate of $E[X|z]$ is within one tenth of a percentage of $SD[X|z]$. Similarly, Figure 6 shows the relative error in the estimation of $SD[X|z]$ using the fourth-degree polynomial in λ for $\mathrm{Var}[X|z]$.

8. Approximation by normal distribution

To the naked eye, a plot of the probability density function for $X|z$ looks exactly like a Bell curve, provided ω is moderate and $\lambda \ll 1$. Visible differences eventually show up when $|\omega|$ and λ are allowed to grow, but for many purposes a normal approximation is more than sufficient when $|\omega|$ and λ stay below given bounds.

[8] When X, Y are independent and $\mu_Y = 0$, then $\mathrm{Var}[Z] = (\sigma_X^2 + \mu_X^2)\,\sigma_Y^2$, so $\mathrm{Var}\left[\frac{Z}{\mu_X\,\sigma_Y}\right] = 1 + \lambda^2$.

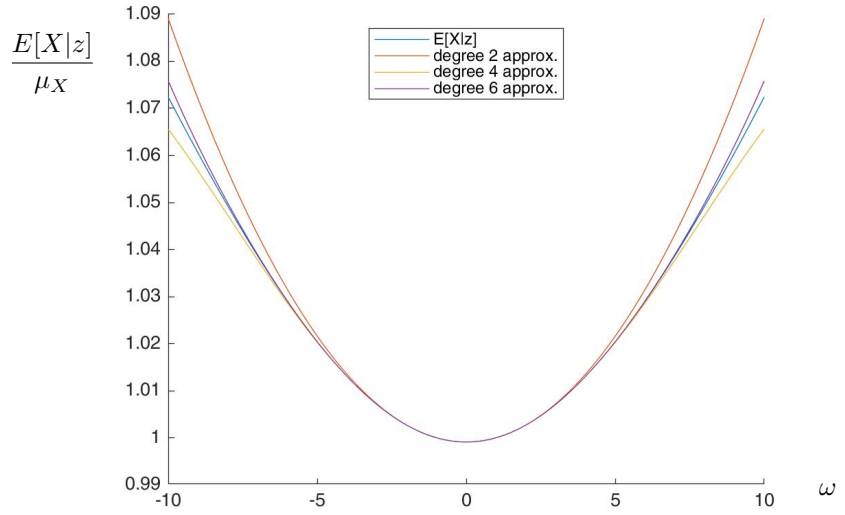

FIGURE 3. Conditional expectation of X given z as a function of ω, together with approximations using second, fourth and sixth-degree polynomials in λ, in the case when $\lambda = 0.03$. The graph for $E[X|z]$ is the second from below on the two sides, while the second-degree approximation overestimates the most, the fourth-degree approximation underestimates, and the sixth-degree approximation is seen to overestimate somewhat at the margins of the plot. A unit on the vertical axis corresponds to μ_X.

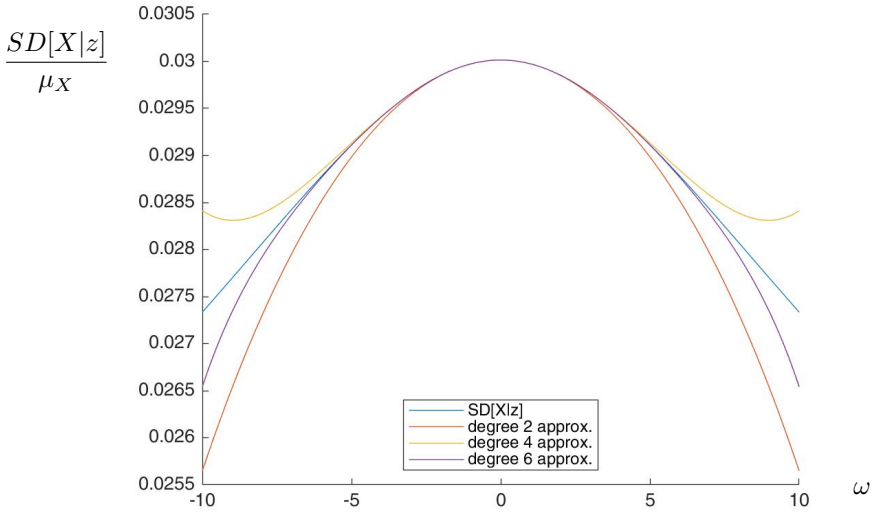

FIGURE 4. Conditional standard deviation of X given z as a function of ω, together with approximations using (square roots of) second, fourth and sixth-degree polynomials in λ, in the case when $\lambda = 0.03$. The graph for $SD[X|z]$ is the second from above on the two sides, while the second-degree approximation underestimates the most, the fourth-degree approximation overestimates, and the sixth-degree approximation is seen to underestimate somewhat at the margins of the plot. A unit on the vertical axis corresponds to μ_X.

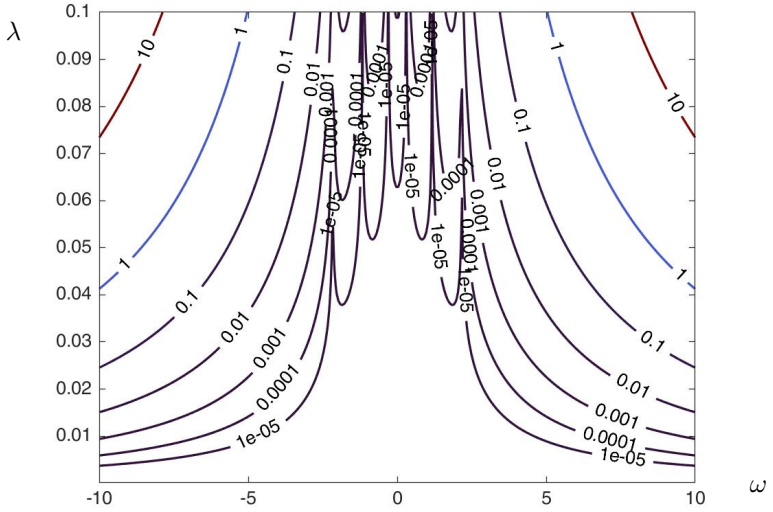

FIGURE 5. Contour plot for the error when estimating $E[X|z]$ by the fourth-degree polynomial, with $|\omega| < 10$ and $\lambda < 0.1$. The values in the diagram are those obtained by dividing the absolute value of the error by $SD[X|z]$. To avoid clutter, powers below 10^{-5} are not shown.

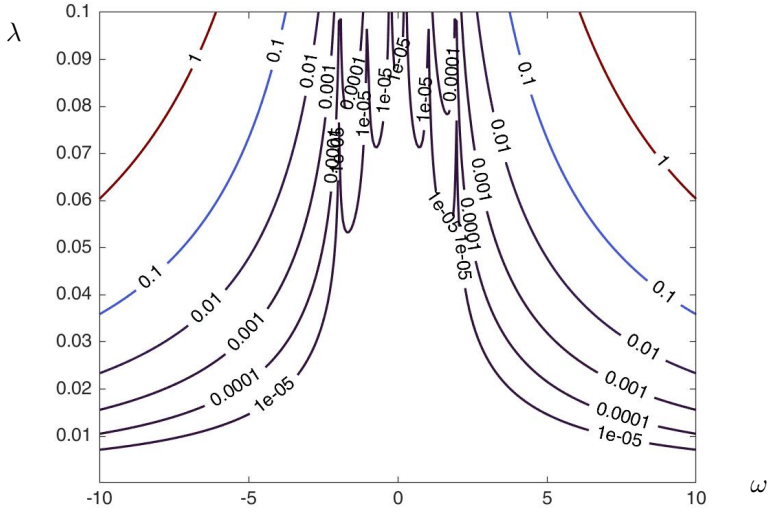

FIGURE 6. Contour plot for the error when estimating $SD[X|z]$ by the square root of the fourth-degree polynomial in λ for $\mathrm{Var}[X|z]$, with $|\omega| < 10$ and $\lambda < 0.1$. The values in the diagram are those obtained by dividing the absolute value of the error by $SD[X|z]$. To avoid clutter, powers below 10^{-5} are not shown.

A somewhat informal explanation for this behavior can be seen from the expression $\frac{1}{|x|} \cdot g(\frac{z}{x}; 0, \sigma_Y) \cdot g(x; \mu_X, \sigma_X)$ for the joint probability density of X and Z. When this is divided by the marginal density for Z at z, one obtains an expression for the density function of $X|z$ of the form

$$\frac{C}{|x|} \mathrm{e}^{-\frac{1}{2}\left(\frac{z}{x\sigma_Y}\right)^2} \cdot \mathrm{e}^{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2},$$

where C is some constant indexed by z and the parameters $\mu_X, \sigma_X, \sigma_Y$. This expression can be written as

$$\frac{C}{|\mu_X|}\,\mathrm{e}^{-\frac{1}{2}f(x)},$$

where

$$f(x) = 2\ln\left|\frac{x}{\mu_X}\right| + \frac{z^2}{\sigma_Y^2 x^2} + \left(\frac{x-\mu_X}{\sigma_X}\right)^2.$$

Approximating the first two terms of $f(x)$ by their second-order Taylor polynomial at $x = \mu_X$, and writing y for $\frac{x-\mu_X}{\mu_X}$, gives

$$f(x) \approx 2y - y^2 + \omega^2(1 - 2y + 3y^2) + \frac{1}{\lambda^2}y^2 = \left(\frac{1}{\lambda^2} + 3\omega^2 - 1\right)\left(y + \frac{1-\omega^2}{\frac{1}{\lambda^2} + 3\omega^2 - 1}\right)^2 + C'.$$

Thus, the density of $X|z$ is approximately

$$C''\,\mathrm{e}^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

where

$$\mu = \mu_X\left(1 + \frac{\omega^2 - 1}{1 + (3\omega^2 - 1)\lambda^2}\lambda^2\right), \qquad \sigma^2 = \frac{\sigma_X^2}{1 + (3\omega^2 - 1)\lambda^2}.$$

These expressions for μ and σ^2 have the same second-degree polynomial approximations in λ as those found in the previous section.

The above argument provides no precise indication of how well the conditional distribution $X|z$ can be approximated with a normal distribution. This will of course depend on the values of ω and λ. In Figure 7, a contour plot is shown for the Bhattacharyya distance between the distribution for $X|z$ and a normal distribution with the same expectation and variance, for a range of values of ω and λ. Note that the Bhattacharyya distance stays below approximately 10^{-6} for any $|\omega| \leq 5$ and $\lambda \leq 0.03$.

The values for the expectation and variance of $X|z$ used in Figure 7 were obtained by costly, numerical computations. To continue the evaluation of the polynomial approximations for these parameters, obtained in the previous section, we include also a similar plot in Figure 8, showing the Bhattacharyya distance between the conditional distribution and the normal distribution with parameters given by the fourth-degree polynomial approximations. Note that Figures 7 and 8 are nearly identical at the bottom quarter or so, where the polynomial parameter approximations are best.

9. Goodness of approximation

The Bhattacharyya distance between two continuous distributions with density functions p_1 and p_2, is defined as $-\ln(BC)$, where the Bhattacharyya coefficient BC is given by the integral $\int \sqrt{p_1(x) \cdot p_2(x)}\mathrm{d}x$. This distance is infinite if no event is assigned a positive likelihood by both distributions, while the distance is zero if, and – ignoring some niceties – only if, the two are identical.

When one of the two distributions is an approximation to the other, and the task is to determine how safe it is to use this approximation, then the Bhattacharyya distance between the two may not provide all the answers. In a situation where, for instance, a 1% error in the density value is acceptable, the crucial figure may be the likelihood that the error stays within this bound. Hence for any given probability density p_1 and approximation p_2 and some bound α, let R_α be the probability, according to p_1, of the event $\frac{|p_1(X) - p_2(X)|}{p_1(X)} \leq \alpha$. A stricter

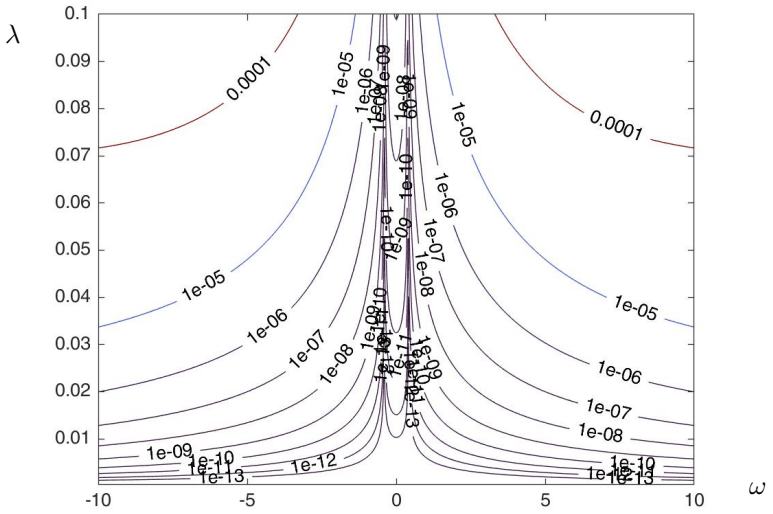

FIGURE 7. Contour plot of Bhattacharyya distance between the distribution for $X|z$ and the normal distribution with the same mean and variance.

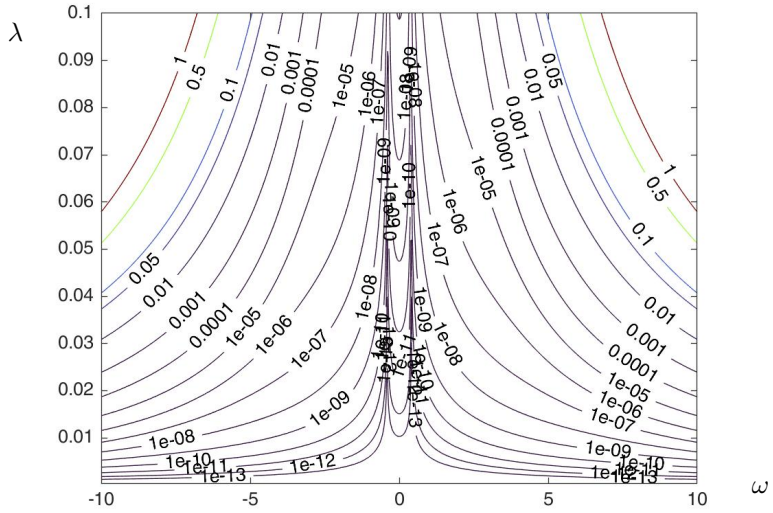

FIGURE 8. Contour plot of Bhattacharyya distance between the distribution for $X|z$ and the normal distribution with the mean and variance estimated by the fourth-degree polynomials in λ.

variant of this measure can also be defined (*cf.* [1, 2]), focusing on the part of the probability space with the highest density values,[9] but we restrict attention to the simpler measure R_α here.

[9]If β is the least density value such that $\frac{|p_1(x)-p_2(x)|}{p_1(x)} \leq \alpha$ whenever $p_1(x) \geq \beta$, then the stricter measure returns the likelihood, according to p_1, of the event $p_1(X) \geq \beta$. This likelihood never exceeds R_α, and if the relative error increases monotonically with lower probability density values, the two measures coincide.

Informally, one can think of an infinite process that inspects the entire probability space, starting with areas of high probability density and moving through ever lower density values until an error exceeding α is found. The proportion of the probability mass traversed at that point is returned. Hence areas of lower probability density but acceptable error are not counted, the rationale being that they are less relevant, or even that such peripheral dips in the error rate should be considered spurious.

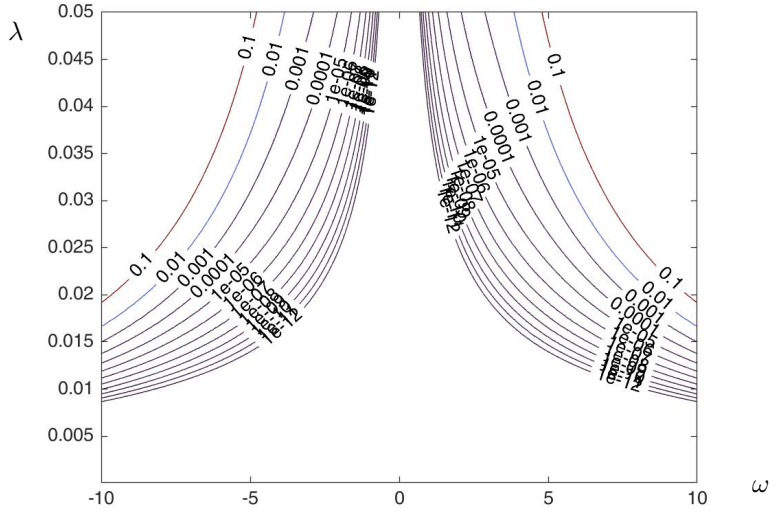

FIGURE 9. Contour plot of $1 - R_{0.05}$ when the distribution for $X|z$ is approximated by a normal distribution with parameters given by the fourth-degree polynomials. To avoid clutter, contour lines above 0.01 and below 10^{-12} are not included.

When the Bhattacharyya distance is known, a lower bound for R_α can be inferred, but this bound is typically well below the actual value. To provide a better picture of the quality and usefulness of the approximation to the distribution of $X|z$ by a normal distribution with fourth-degree polynomial approximations for the parameters, we therefore include in Figures 9–11 some contour plots for this measure as well, for $\alpha = 0.05$, 0.01, and 0.001, respectively. Since we primarily foresee an application of these results in cases where λ stays below 0.05, the vertical axes are restricted accordingly in these plots.

It can be seen from Figure 10 that as long as X's coefficient of variation is within 0.04 and the standardized product value ω is less than 3 in absolute value, the error, when approximating the probability density for $X|z$ by the probability density of the normal distribution with parameters obtained by the fourth-degree polynomial approximations, will stay below 1% in 99% of the distribution. With a better resolution, one would also be able to read off the value $R_{0.01} \approx 0.994$ when $\lambda = 0.04$ and $\omega = \frac{\pi}{1.2} \approx 2.6$. This will be used in the example in the next section.

10. Example: Distance when azimuth is known

This section presents an example which requires the conditional distribution of one factor variable when the product of two independent, normal variables is given. An object has moved an unknown distance from its original position, along an unknown trajectory. It is assumed that the trajectory has maintained a constant turn radius, or, equivalently, a constant curvature. It is also assumed that the original direction of movement is known.

The situation is depicted in Figure 12, with the original position in the origin and the original direction of movement going upwards. The object has moved along the red curve, and its current position is indicated by the star. L is the length of the completed trajectory and R is the turn radius. The model is based on the variable $L \sim \mathcal{N}(\mu_L, \sigma_L)$ for the trajectory length and the curvature variable $K \sim \mathcal{N}(0, \sigma_K)$. Positive values of K correspond to right turns, negative values to left turns, while zero corresponds to movement along a straight line in the original direction. Except in the latter case, $R = \frac{1}{|K|}$. D is the distance out to the object, measured from its original position, while α is the azimuth; the angle out to the object from the original position, relative to its original direction of movement. The angle φ measures the total change of direction of the object, or, equivalently, the angle spanned by the two radiuses from the center of curvature out to the original and final

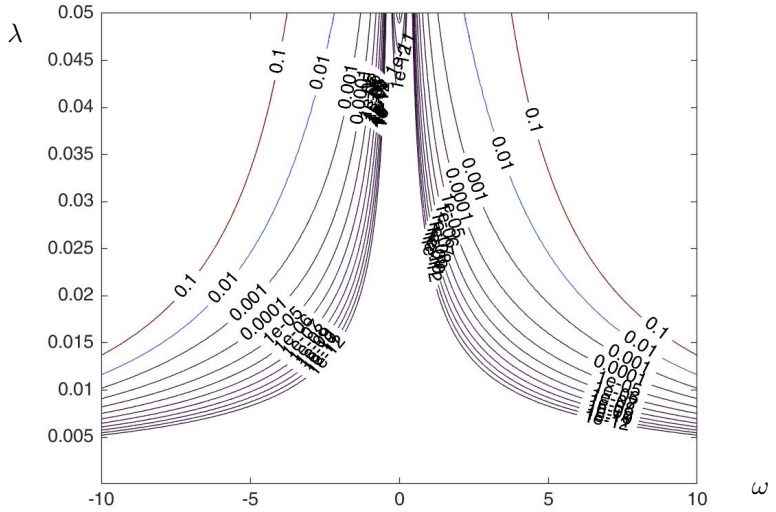

FIGURE 10. Contour plot of $1 - R_{0.01}$ when the distribution for $X|z$ is approximated by a normal distribution with parameters given by the fourth-degree polynomials. To avoid clutter, contour lines above 0.01 and below 10^{-12} are not included.

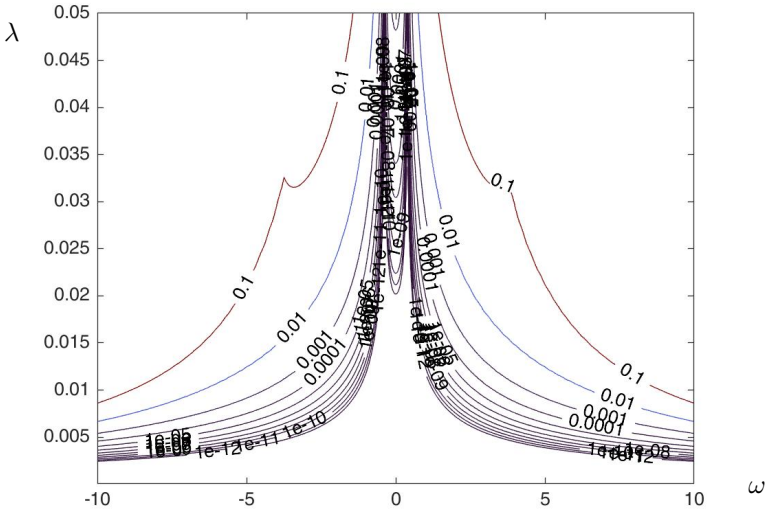

FIGURE 11. Contour plot of $1 - R_{0.001}$ when the distribution for $X|z$ is approximated by a normal distribution with parameters given by the fourth-degree polynomials. To avoid clutter, contour lines above 0.01 and below 10^{-12} are not included.

positions. Under these assumptions, the following identities can be proved by elementary trigonometry:

$$2\alpha = \varphi = L \cdot K, \qquad D = L \cdot \operatorname{sinc} \alpha.$$

Continuing the example, suppose an observer has remained in the original position and is able to measure the azimuth α but not the distance D. The product $\varphi = L \cdot K$ is directly obtained from α as 2α, and an approximation to the conditional distribution for $L|\varphi$ can then be found by the methods described in Section 7.

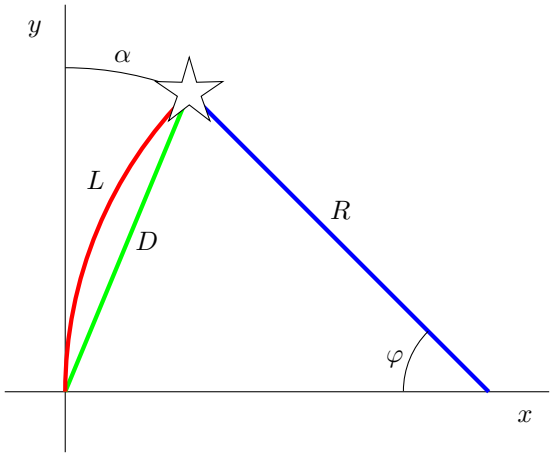

FIGURE 12. Stochastic model based on two independent, normal variables; trajectory length L along a circular path, and curvature $K = \pm\frac{1}{R}$ of this circular path. The product of the variables equals the completed rotation φ, which is also twice the azimuth α.

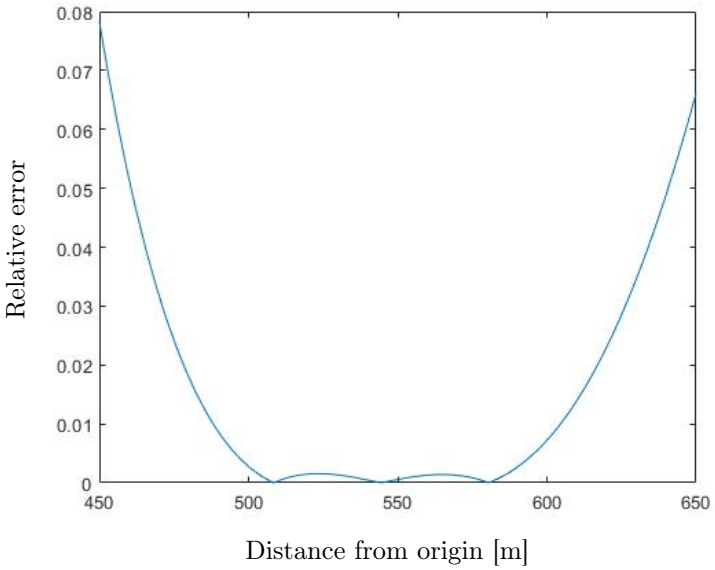

FIGURE 13. Relative error of probability density values when a normal distribution with parameters based on fourth-degree polynomial estimates is used to approximate the actual distribution for $D|\alpha = \frac{\pi}{4}$ as outlined in Figure 12, when μ_L, σ_L and σ_K are set to 600 m, 24 m and $10^{-3}\,\mathrm{m}^{-1}$, respectively.

But when α is given, $D = L \cdot \operatorname{sinc}\alpha$ is just a linear function of L, and hence the distribution of $D|\alpha$ is readily obtained from the distribution of $L|\varphi$.

Now suppose the length of the completed trajectory has an expected value μ_L and a standard deviation σ_L of 600 m and 24 m, respectively, while the trajectory's curvature has a standard deviation of $\sigma_K = 10^{-3}\,\mathrm{m}^{-1}$, yielding a turn radius of 1000 m when the curvature is one standard deviation out from zero, and then imagine an azimuth of $\alpha = 45° = \frac{\pi}{4}$ is observed. In this case $\lambda = \frac{\sigma_L}{\mu_L} = 0.04$. The known product value (z) is $\varphi = 2\alpha = \frac{\pi}{2}$,

and the corresponding standardized value (ω) is $\frac{\varphi}{\mu_L\,\sigma_K} = \frac{\pi}{1.2} \approx 2.6$. Using the fourth-degree polynomial approximations for the parameters and suppressing units, this yields a normal approximation for the distribution of $L|\varphi = \frac{\pi}{2}$ with mean 605.51 and standard deviation 23.64, and hence a normal approximation for the distribution of $D|\alpha = \frac{\pi}{4}$ with mean $\operatorname{sinc}\alpha \cdot 605.51 \approx 545$ and standard deviation $\operatorname{sinc}\alpha \cdot 23.64 \approx 21$.

When the density function of the latter, normal distribution is plotted in the same, regular paper-sized diagram as a plot of the actual probability density function for $D|\alpha = \frac{\pi}{4}$, the curves fall right on top of each other without any visible difference. The relative match is good in the inner part of the distribution, and only starts to slip in the outer part where neither function is distinguishable from zero. A plot of the unsigned, relative error is more instructive, and is shown in Figure 13. The relative error stays below 1% for values of D between approximately 488 and 605. The probability mass contained in this interval is approximately 0.994, *i.e.*, in accordance with what we already saw from Figure 10. At the same time, it can be seen that the relative error never exceeds 8% in the whole interval between 450 and 650, which contains a probability mass of approximately $1 - 4 \cdot 10^{-6}$.

Appendix A. Uniform convergence lemmas

The first two lemmas of this appendix are used in Section 2 in the proof of Theorem 2.2, while Lemma A.3 is used in Section 3 in the proof of Theorem 3.1.

Lemma A.1. *Let f be a non-negative, measurable function on $\mathbb{R}$, and suppose $\int f(x)g(x;\mu,\sigma)\mathrm{d}x$ converges for all μ in an open interval I. Let $p(x,\mu)$ be a polynomial in x where the coefficients are functions in μ that are continuous over I, and let J be a closed subinterval of I. Then there exists a measurable function $h(x)$ such that $g(x;\mu,\sigma)\cdot|p(x,\mu)| \le h(x)$ for all $\mu \in J$ and $x \in \mathbb{R}$, and the integral $\int f(x)h(x)\mathrm{d}x$ converges.*

Proof. It suffices to consider $p(x,\mu)$ of the form $p(\mu)x^n$, as the general version is a trivial consequence. Let $J = [\alpha,\beta]$, and choose some $a < \alpha$ and $b > \beta$, $a, b \in I$. We start by showing that for any $\mu \in [\alpha,\beta]$, the function $g(x;\mu,\sigma)|p(\mu)x^n|$ is bounded by $g(x;b,\sigma)$ for any x above some fixed limit which does not depend on μ. The inequality required is equivalent to

$$2x \ge b + \mu + \frac{2\sigma^2}{b-\mu}\left(n\ln|x| + \ln|p(\mu)|\right),$$

which, given the assumptions $\alpha \le \mu \le \beta < b$, follows from

$$x > b + \frac{\sigma^2}{b-\beta}\left(n\ln|x| + \max_{\alpha\le\mu\le\beta}\left(\ln|p(\mu)|\right)\right).$$

Since x sooner or later outruns any linear function of $\ln|x|$, this inequality is satisfied by all x beyond some limit.

A limit for x below which $g(x;\mu,\sigma)|p(x,\mu)|$ is bounded by $g(x;a,\sigma)$ is found in a similar manner. Hence $g(x;\mu,\sigma)|p(x,\mu)|$ is bounded by $g(x;a,\sigma) + g(x;b,\sigma)$ for all $\mu \in [\alpha,\beta]$ and all x except those in a finite interval. The supremum of $g(x;\mu,\sigma)|p(x,\mu)|$ for x in this interval and for $\mu \in [\alpha,\beta]$ is finite, while the infimum of $g(x;a,\sigma)+g(x;b,\sigma)$ for these x is positive; hence for some constant C, the function $g(x;\mu,\sigma)|p(x,\mu)|$ is bounded by $h(x) = C\left(g(x;a,\sigma) + g(x;b,\sigma)\right)$ for all $\mu \in [\alpha,\beta]$ and all $x \in \mathbb{R}$. □

Lemma A.2. *Let f be a non-negative, measurable function on $\mathbb{R}$, and suppose $\int f(x)g(x;\mu,\sigma)\mathrm{d}x$ converges for all μ in an open interval I. Let $p(x,\mu)$ be a polynomial in x where the coefficients are functions in μ that are infinitely differentiable over I. Then $\int f(x)p(x,\mu)g(x;\mu,\sigma)\mathrm{d}x$ exists and is infinitely differentiable with respect to μ, with n-th derivative $\int f(x)\frac{\partial^n}{\partial\mu^n}(p(x,\mu)g(x;\mu,\sigma))\mathrm{d}x$.*

Proof. Modifying the statement, we claim that if the coefficients of $p(x,\mu)$ are only required to be differentiable with continuous derivatives, then $\int f(x)p(x,\mu)g(x;\mu,\sigma)\mathrm{d}x$ exists and is differentiable with respect to μ, with

derivative $\int f(x)\frac{\partial}{\partial\mu}(p(x,\mu)g(x;\mu,\sigma))\mathrm{d}x$. The lemma follows from repeated application of this simpler statement, which we now prove.

Existence follows from Lemma A.1. Moreover, $\frac{\partial}{\partial\mu}(p(x,\mu)g(x;\mu,\sigma)) = p_1(x,\mu)g(x;\mu,\sigma)$ for some polynomial[10] $p_1(x,\mu)$ in x where the coefficients are continuous functions of μ in all of I. Then by Lemma A.1, for any closed subinterval $J=[\alpha,\beta]$ there is a function $h(x)$ such that the integral $\int f(x)h(x)\mathrm{d}x$ exists, and such that

$$\left|\frac{\partial}{\partial\mu}(f(x)p(x,\mu)g(x;\mu,\sigma))\right| = f(x)\left|\frac{\partial}{\partial\mu}(p(x,\mu)g(x;\mu,\sigma))\right| \le f(x)h(x)$$

for all $\mu\in J$. By the sufficient criterion for differentiability and for interchanging differentiation and integration given in [6], Theorem 2.27(b), it follows that for all $\mu\in(\alpha,\beta)$, $\int f(x)p(x,\mu)g(x;\mu,\sigma)\mathrm{d}x$ is differentiable with respect to μ, with derivative equal to $\int\frac{\partial}{\partial\mu}(f(x)p(x,\mu)g(x;\mu,\sigma))\mathrm{d}x$. Since $[\alpha,\beta]\subset I$ was arbitrary, we are done. □

In Lemma A.3 below, note that an integral part of the result is the admissibility, under the given assumptions, of distributing the improper integral over the infinite sum, when $f(x)$ is replaced by its Taylor series around μ, since $\int(x-\mu)^n g(x;\mu,\sigma)\mathrm{d}x = (n-1)!!\,\sigma^n$ for even n, while the integral vanishes if n is odd, and $n! = n!!(n-1)!!$.

Lemma A.3. *Suppose f is a real entire function. Then the identity*

$$\int f(x)g(x;\mu,\sigma)\mathrm{d}x = \sum_{n=0}^{\infty}\frac{f^{(2n)}(\mu)\sigma^{2n}}{(2n)!!}$$

holds whenever both sides converge to finite values. If f is non-negative outside of some finite interval, then convergence of the integral follows from convergence of the sum.

Proof. Suppose the lemma is true for the special case where $\mu=0$ and $\sigma=1$. To prove the general case from this, let $h(u)$ be $f(\sigma u+\mu)$. Then $\int f(x)g(x;\mu,\sigma)\mathrm{d}x = \int h(u)g(u;0,1)du$ and $f^{(2n)}(\mu)\sigma^{2n} = h^{(2n)}(0)$.

Thus, we only need to show that if the Maclaurin series of f converges (pointwise) to $f(x)$ for all $x\in\mathbb{R}$, then

$$\int f(x)g(x;0,1)\mathrm{d}x = \sum_{n=0}^{\infty}\frac{f^{(2n)}(0)}{(2n)!!} \tag{A.1}$$

whenever (a) both sides converge to finite values or (b) the series converges and f is non-negative for all x with sufficiently large absolute value. In any of these two cases, assume that the series in (A.1) converges; it will then suffice to show that the integral

$$\int_{-a}^{a} f(x)g(x;0,1)\mathrm{d}x = \int_{-a}^{a}\sum_{n=0}^{\infty}\frac{f^{(n)}(0)}{n!}x^n g(x;0,1)\mathrm{d}x$$

converges to the right-hand side of (A.1) when a approaches infinity. Since the Maclaurin series of f is assumed to have infinite convergence radius, it converges uniformly on the finite interval $[-a,a]$. Hence so does the series $\sum_{n=0}^{\infty}\frac{f^{(n)}(0)}{n!}x^n g(x;0,1)$, which means that

$$\begin{aligned}\int_{-a}^{a}\sum_{n=0}^{\infty}\frac{f^{(n)}(0)}{n!}x^n g(x;0,1)\mathrm{d}x = \sum_{n=0}^{\infty}\int_{-a}^{a}\frac{f^{(n)}(0)}{n!}x^n g(x;0,1)\mathrm{d}x &= \sum_{n=0}^{\infty}\frac{f^{(n)}(0)}{n!}\int_{-a}^{a}x^n g(x;0,1)\mathrm{d}x \\ &= \sum_{n=0}^{\infty}\frac{f^{(2n)}(0)}{(2n)!}\int_{-a}^{a}x^{2n}g(x;0,1)\mathrm{d}x.\end{aligned}$$

[10] This polynomial will depend on the fixed parameter σ, but that does not affect the argument.

Consider the last integral. By induction on n and integration by parts, it can be shown that

$$\int_{-a}^{a} x^{2n} g(x;0,1)\mathrm{d}x = (2n-1)!! \left(\int_{-a}^{a} g(x;0,1)\mathrm{d}x - \sqrt{\frac{2}{\pi}}\, \mathrm{e}^{\frac{-a^2}{2}} \sum_{k=0}^{n-1} \frac{a^{2k+1}}{(2k+1)!!} \right). \tag{A.2}$$

Hence the improper integral in (A.1) equals the limit of $\alpha - \beta$ when a approaches infinity, where

$$\alpha = \sum_{n=0}^{\infty} \frac{f^{(2n)}(0)}{(2n)!!} \int_{-a}^{a} g(x;0,1)\mathrm{d}x, \quad \beta = \sum_{n=0}^{\infty} \frac{f^{(2n)}(0)}{(2n)!!} \sqrt{\frac{2}{\pi}}\, \mathrm{e}^{\frac{-a^2}{2}} \sum_{k=0}^{n-1} \frac{a^{2k+1}}{(2k+1)!!}.$$

Now α approaches the right-hand side of (A.1) when a approaches infinity; it remains to show that β vanishes. To prove this, we first note that

$$\sum_{k=0}^{\infty} \frac{a^{2k+1}}{(2k+1)!!} = \sqrt{\frac{\pi}{2}}\, \mathrm{e}^{\frac{a^2}{2}} \int_{-a}^{a} g(x;0,1)\mathrm{d}x. \tag{A.3}$$

This follows from dividing (A.2) by $(2n-1)!!$ and noting that $\int_{-a}^{a} x^{2n} g(x;0,1)\mathrm{d}x < 2a \cdot a^{2n} g(0;0,1)$, and hence that the ratio on the left-hand side vanishes in the limit where n goes to infinity.[11]

We now write

$$\beta = \sum_{n=0}^{\infty} a_n b_n, \quad a_n = \sqrt{\frac{2}{\pi}}\, \mathrm{e}^{\frac{-a^2}{2}} \sum_{k=0}^{n-1} \frac{a^{2k+1}}{(2k+1)!!}, \quad b_n = \frac{f^{(2n)}(0)}{(2n)!!}.$$

Since the terms of β are the termwise products of a sequence $\{a_n\}$ that is monotone and bounded (by its limit $\int_{-a}^{a} g(x;0,1)\mathrm{d}x < 1$) and a convergent series $\sum b_n$, β is convergent by Abel's test. We will show that for any $\varepsilon > 0$, there exists $A > 0$ such that $|\beta| < \varepsilon$ for all $a \geq A$. Choose N such that

$$\left| \sum_{k=N}^{n} b_k \right| < \frac{\varepsilon}{3}$$

for all $n \geq N$. Next, choose $A > 0$ such that

$$\left| \sum_{n=0}^{N-1} a_n b_n \right| < \frac{\varepsilon}{3}$$

for all $a \geq A$; this is possible because the finite sum is a polynomial in a multiplied by $\mathrm{e}^{\frac{-a^2}{2}}$. Define $B_n = \sum_{k=N}^{n} b_k$. For any value of a, and $M \geq N$,

$$\left| \sum_{n=N}^{M} a_n b_n \right| = \left| \sum_{n=N}^{M-1} B_n(a_n - a_{n+1}) + B_M a_M \right| \leq \sum_{n=N}^{M-1} |B_n|\,|a_n - a_{n+1}| + |B_M|\, a_M < \frac{\varepsilon}{3} \cdot |a_N - a_M| + \frac{\varepsilon}{3} \cdot 1 < \frac{2\varepsilon}{3},$$

since $0 < a_n < 1$. Thus, $\left| \sum_{n=N}^{\infty} a_n b_n \right| \leq \frac{2\varepsilon}{3}$, and hence for all $a \geq A$, $|\beta| = \left| \sum_{n=0}^{\infty} a_n b_n \right| < \varepsilon$. □

[11] Equation (A.3) is equivalent to formulas for the error function given in [7] and [8], which are derived in different ways.

Appendix B. General behavior of X when $|XY|$ approaches 0

This appendix and the next present complementary results about the conditional distribution of X when $|XY|$ is known to be small. Theorem B.1 below essentially states – under quite general terms – that the conditional distribution of $X||XY| < |z|$ collapses to a constant at 0 when z vanishes, or, to put it differently, that all the probability mass of the variable squeezes in towards 0. When the variables have continuous probability density functions, the only additional requirement for the results to apply, beyond mutual independence of X and Y, is that X have positive probability density at 0.

Independent, normal variables clearly meet this condition, but to complement the results we also show in Appendix C that when both variables are normal and X has a low coefficient of variation, this convergence behavior is only evident in such extremely narrow (and thus improbable) intervals around $z = 0$ that they can safely be ignored for most practical purposes.

In the statement of the theorem, note that the existence of $E[X^n]$ is trivially guaranteed in the case when $n = 0$, hence the result about $E[\,|X| \mid |XY| < \delta\,]$ is valid even for such probability distributions as the Cauchy distribution, where $E[X]$ does not exist for the unconditioned variable.

The proof uses the fact that the n-th moment $E[X^n]$ exists if an only if the n-th absolute moment $E[|X^n|]$ exists. Convergence of the conditional absolute moments to zero of course imply the corresponding result for the conditional moments themselves, but the result is stronger when stated as below.

Theorem B.1. *Let X, Y be independent, continuous random variables with probability density functions p_X and p_Y. Suppose p_X has a positive, lower bound on some proper interval containing 0, and that p_Y has a finite, upper bound on an interval containing 0 as an interior point. Then the following statements hold.*

1. $\lim_{\delta\to 0^+} P\left(|X| > a \mid |XY| < \delta\right) = 0$ *for all* $a > 0$.
2. *For any integer* $n \geq 0$, *if* $E[X^n]$ *exists, then* $\lim_{\delta\to 0^+} E\left[|X|^{n+1} \mid |XY| < \delta\right] = 0$.

Proof. Since the relevant properties are preserved when X and Y are scaled up or down by constant (positive or negative) factors, we may, without loss of generality, assume that p_X has the lower bound $\alpha > 0$ on the interval $[0, 1]$ and that p_Y has the upper bound $\beta \geq 0$ on the interval $[-1, 1]$.

We first prove part 2. Suppose that $E[X^n]$ exists, and let $\delta > 0$.

$$E\left[\,|X|^{n+1} \mid |XY| < \delta\,\right] = \frac{\displaystyle\int_{-\infty}^{\infty}\int_{-\frac{\delta}{|x|}}^{\frac{\delta}{|x|}} |x|^{n+1} p_X(x) p_Y(y)\mathrm{d}y\,\mathrm{d}x}{\displaystyle\int_{-\infty}^{\infty}\int_{-\frac{\delta}{|y|}}^{\frac{\delta}{|y|}} p_X(x) p_Y(y)\mathrm{d}x\,\mathrm{d}y}. \tag{B.1}$$

The numerator can be written as $2\delta\int_{-\infty}^{\infty} |x|^n p_X(x)\left\langle p_Y\right\rangle_x \mathrm{d}x$, where $\left\langle p_Y\right\rangle_x$ denotes the average value of p_Y on the interval $\left[-\frac{\delta}{|x|}, \frac{\delta}{|x|}\right]$. Since $p_Y(y) \leq \beta$ for $-1 \leq y \leq 1$, the average value is no greater than $\max\{\frac{1}{2}, \beta\}$ for any x, and hence the numerator in (B.1) is less than $\delta \cdot (1 + 2\beta)E[|X|^n]$, *i.e.*, bounded from above.

The denominator in (B.1) is at least

$$\int_{-\infty}^{-\delta}\int_{0}^{\frac{\delta}{|y|}} p_X(x)p_Y(y)\mathrm{d}x\,\mathrm{d}y + \int_{\delta}^{\infty}\int_{0}^{\frac{\delta}{y}} p_X(x)p_Y(y)\mathrm{d}x\,\mathrm{d}y \geq \delta\cdot\alpha\cdot\left(\int_{-\infty}^{-\delta}\frac{p_Y(y)}{|y|}\mathrm{d}y + \int_{\delta}^{\infty}\frac{p_Y(y)}{y}\mathrm{d}y\right),$$

which is positive for sufficiently small δ. Thus, after reducing the fraction,

$$E\left[|X|^{n+1} \mid |XY| < \delta\right] \leq \frac{\frac{2}{\alpha}\int_{-\infty}^{\infty} |x|^n p_X(x)\left\langle p_Y\right\rangle_x \mathrm{d}x}{\int_{-\infty}^{-\delta}\frac{p_Y(y)}{|y|}\mathrm{d}y + \int_{\delta}^{\infty}\frac{p_Y(y)}{y}\mathrm{d}y}.$$

If the denominator grows to infinity as δ diminishes, then, since the numerator is bounded, the fraction will tend to zero. If, on the other hand, the denominator is bounded from above, then it has a positive limit, and the integral $\int_{-t}^{t}\frac{p_Y(y)}{|y|}\mathrm{d}y$ exists for all t and tends to 0 as t approaches zero. In this case, the numerator has the limit 0 as δ approaches 0. To see this, note first that that part of the integral in the numerator where $|x| \leq \sqrt{\delta}$, evaluates to less than $\delta^{\frac{n}{2}}\left(\frac{1}{2}+\beta\right)\int_{-\sqrt{\delta}}^{\sqrt{\delta}} p_X(x)\mathrm{d}x$ and therefore tends to 0 with diminishing δ. In the rest of the integral, where $|x| \geq \sqrt{\delta}$, the factor $\left\langle p_Y\right\rangle_x$ in the integrand equals $\int_{-t}^{t}\frac{p_Y(y)}{2t}\mathrm{d}y \leq \int_{-t}^{t}\frac{p_Y(y)}{2|y|}\mathrm{d}y \leq \int_{-\sqrt{\delta}}^{\sqrt{\delta}}\frac{p_Y(y)}{2|y|}\mathrm{d}y$ with $t = \frac{\delta}{|x|} \leq \sqrt{\delta}$. As this bound is independent of x, it can be factored out of the surrounding integral, yielding the product of one integral over x and one over y, where the former is finite by the existence of the relevant moment, and the latter vanishes with δ.

Part 1 may be proven in a similar (but simpler) way. Alternatively, suppose part 1 is wrong. Then there is an $\varepsilon > 0$ and arbitrarily small δ such that $P\left(|X| > a \mid |XY| < \delta\right) \geq \varepsilon$. This implies that $E\left[\,|X| \mid |XY| < \delta\,\right] \geq a \cdot \varepsilon$ for arbitrarily small δ, contradicting part 2. □

Appendix C. Behavior of X when XY approaches 0 for normal X, Y

It follows from the results in Appendix B that for normal variables X and Y, the conditional absolute moments of X given $|XY| < \delta$ vanish with δ. This is in accordance with the identity $E[\,|X|^n \mid XY = z\,] = \frac{M_n}{\Theta}$ for independent X, Y such that $X \sim \mathcal{N}(\mu_X, \sigma_X)$ and $Y \sim \mathcal{N}(\mu_Y, \sigma_Y)$, where $\Theta = \int \frac{1}{|x|} g(\frac{z}{x}; \mu_Y, \sigma_Y) g(x; \mu_X, \sigma_X)\mathrm{d}x$ and $M_n = \int \frac{|x|^n}{|x|} g(\frac{z}{x}; \mu_Y, \sigma_Y) g(x; \mu_X, \sigma_X)\mathrm{d}x$. Here M_n is bounded by $g(0; 0, \sigma_Y) \cdot E[|X|^{n-1}]$ for any z, while Θ grows to infinity with diminishing z. To see this, we may zoom in on the interval $(|z|, 1)$ for $0 < |z| < 1$, in which interval any value of x yields values of $g(\frac{z}{x}; \mu_Y, \sigma_Y)$ and $g(x; \mu_X, \sigma_X)$ at least equal to $g(-1; |\mu_Y|, \sigma_Y)$ and $g(-1; |\mu_X|, \sigma_X)$, respectively, both limits independent of z. Hence $\Theta > c\int_{|z|}^{1}\frac{1}{x}\mathrm{d}x = -c\ln|z|$, which grows to infinity when z approaches 0.

This growth is extremely slow, however, and we show in the proposition below that with diminishing values of $|z|$ (and provided $\frac{\sigma_X}{|\mu_X|}$ is small), Θ first stabilizes around a certain value Θ_0, where it remains for all but the most unthinkably small values of $|z|$. For simplicity, we assume that $\mu_Y = 0$, as in Section 7. For positive μ_X, Θ_0 is defined by the equation

$$\Theta_0 = g(0; 0, \sigma_Y)\int_{\frac{\mu_X}{10}}^{\infty}\frac{1}{x} g(x; \mu_X, \sigma_X)\mathrm{d}x.$$

If μ_X is negative, Θ_0 is defined in a symmetric manner by replacing the factor $\frac{1}{x}$ by $\frac{1}{|x|}$ and integrating from $-\infty$ to $\frac{\mu_X}{10}$. Thus Θ_0 is independent of z, and is obtained from Θ by cutting off the integration at $\frac{\mu_X}{10}$ and replacing $g(\frac{z}{x}; 0, \sigma_Y)$ by $g(0; 0, \sigma_Y)$. It is essentially equal to $g(0; 0, \sigma_Y)E\left[\frac{1}{|\tilde{X}|}\right]$, where $\tilde{X}$ is a random variable similar to X, but where the density function is set to 0 for values of x beyond a tenth of μ_X.[12]

[12] To make this a proper probability density function it would have to be scaled up, but – when $\frac{\sigma_X}{|\mu_X|}$ stays below 0.1 – not more than with a factor of approximately $1 + 10^{-19}$, given the exceedingly low likelihood that a random variable will end up more than nine standard deviations to the left.

The proposition is stated using the parameters ω and λ defined in (6.3). Then

$$\Theta = \frac{1}{|\mu_X|\sigma_Y}\int_{-\infty}^{\infty}\frac{1}{|x|}g(\tfrac{\omega}{x};0,1)g(x;1,\lambda)\mathrm{d}x \quad\text{and}\quad \Theta_0 = \frac{1}{\sqrt{2\pi}|\mu_X|\sigma_Y}\int_{0.1}^{\infty}\frac{1}{x}g(x;1,\lambda)\mathrm{d}x.$$

Proposition C.1. *Let Θ and Θ_0 be as above, i.e., with $\mu_Y = 0$, and suppose $\lambda \le 0.1$ and $10^{-2000} \le |\omega| \le 0.1$. Then $\left(1-\omega^2\right)\cdot\Theta_0 < \Theta < \Theta_0\cdot\left(1+10^{-17}\right)$.*

Proof. The left part of the proposition follows from the inequality $A > (1-\omega^2)B$, where $A = \int_{0.1}^{\infty}\mathrm{e}^{-\frac{\omega^2}{2x^2}}\frac{1}{x}g(x;1,\lambda)\mathrm{d}x$ and $B = \int_{0.1}^{\infty}\frac{1}{x}g(x;1,\lambda)\mathrm{d}x$. To prove this, let A_1, A_2, A_3 and B_1, B_2, B_3 be the corresponding integrals over the intervals $(0.1, \sqrt{0.5})$, $(\sqrt{0.5}, 1)$, and $(1,\infty)$. Since $\mathrm{e}^{-\frac{\omega^2}{2x^2}}$ is a monotonically increasing function for positive x, and $\mathrm{e}^{-a} > 1-a$ for positive a, it follows that $A_1 > (1-50\omega^2)B_1$, $A_2 > (1-\omega^2)B_2$, and $A_3 > (1-0.5\omega^2)B_3$. Hence it suffices to show that $A_1 + A_3 > (1-\omega^2)(B_1+B_3)$, which again follows if we can show that B_3 is always at least 98 times larger than B_1.

To this end, note that for any $\lambda \le 0.1$, $B_1 \le \int_{0.1}^{\sqrt{0.5}}\frac{1}{x}g(x;1,0.1)\mathrm{d}x < 0.003$, and that $B_3 = \int_1^{\infty}\frac{1}{x}g(x;1,\lambda)\mathrm{d}x > \int_1^{1.2}\frac{1}{x}g(x;1,\lambda)\mathrm{d}x > \frac{1}{1.2}\int_1^{1.2}g(x;1,\lambda)\mathrm{d}x$, which for any $\lambda \le 0.1$ is larger than 0.3, *i.e.*, more than 100 times larger than B_1.

Let I and J be the integrands $\frac{1}{|x|}g(\frac{\omega}{x};0,1)g(x;1,\lambda)$ and $\frac{1}{|x|}g(x;1,\lambda)$, respectively; the right inequality of the proposition is proved by showing that $\int_{-\infty}^{\infty}I\mathrm{d}x < (1+10^{-17})\cdot\frac{1}{\sqrt{2\pi}}\int_{0.1}^{\infty}J\mathrm{d}x$. Since $I < \frac{1}{\sqrt{2\pi}}J$ for all x, $\int_{0.1}^{\infty}I\mathrm{d}x < \frac{1}{\sqrt{2\pi}}\int_{0.1}^{\infty}J\mathrm{d}x$, so it only remains to show that the latter quantity outsizes $\int_{-\infty}^{0.1}I\mathrm{d}x$ to the specified degree.

A lower limit for $\int_{0.1}^{\infty}J\mathrm{d}x$ can be found by noting that for $x \ge 0.1$, the function $\frac{1}{x}$ lies above its tangent at $x=1$, where it has the value 1, while the second factor of J is symmetric about $x=1$. Thus, $\int_{0.1}^{1.9}J\mathrm{d}x > \int_{0.1}^{1.9}g(x;1,\lambda)\mathrm{d}x$, which is close to 1 for all $\lambda \le 0.1$. Since $\frac{1}{\sqrt{2\pi}} \approx 0.40$, we have $\frac{1}{\sqrt{2\pi}}\int_{0.1}^{\infty}J\mathrm{d}x > 0.39$.

To obtain an upper limit for $\int_{-\infty}^{0.1}I\mathrm{d}x$, we consider the constituent intervals $(-\infty,-0.001)$, $(-0.001, 0.001)$, and $(0.001, 0.1)$. It is clear that $\int_{-\infty}^{-0.001}I\mathrm{d}x < g(0;0,1)\int_{-\infty}^{0.001}J\mathrm{d}x$, and similarly, $\int_{0.001}^{0.1}I\mathrm{d}x < g(0;0,1)\int_{0.001}^{0.1}J\mathrm{d}x$. For $\lambda = 0.1$, these upper limits evaluate to less than 10^{-21} and $5.2\cdot10^{-19}$, respectively. As the function $g(a;b,\lambda)$ diminishes with diminishing values of λ when $\lambda < |a-b|$, these limits hold for all $\lambda \le 0.1$.

Finally, $\int_{-0.001}^{0.001}I\mathrm{d}x < \frac{1}{\sqrt{2\pi}}g(0.001;1,\lambda)\int_{-0.001}^{0.001}\frac{1}{|x|}\mathrm{e}^{-\frac{\omega^2}{2x^2}}\mathrm{d}x$. The integral to the right equals[13] $E_1(5\cdot10^5\cdot\omega^2)$, which for $|\omega| \ge 10^{-2000}$ is less than 9197, while the product in front is smaller than $3.4\cdot10^{-22}$ for $\lambda = 0.1$, and diminishes with smaller λ. Hence when both $|\omega| \ge 10^{-2000}$ and $\lambda \le 0.1$, the bound on the right is less than $3.13\cdot10^{-18}$.

Summing up, we see that as long as $\lambda \le 0.1$ and $10^{-2000} \le |\omega| \le 0.1$, the three integrals of I taken together yield less than $3.7\cdot10^{-18}$, and hence less than 10^{-17} of 0.39. □

Appendix D. Derivatives and series expansions

This appendix investigates the general function $h(x) = \frac{1}{|x|}\mathrm{e}^{\frac{a}{x^2}}$, that appears in (6.1), in more detail, establishing the results about its derivatives used in Section 7. We are primarily interested in the case that a is negative, but the results hold for all a. As claimed in Section 7, one ingredient in the expression for the n-th derivative of h is the Touchard polynomials $T_n(x) = \sum_{k=0}^{n}\left\{{n \atop k}\right\}x^k$, where $\left\{{n \atop k}\right\}$ is the Stirling number of the second kind. Similarly, $\left[{n \atop k}\right]$ is the unsigned Stirling number of the first kind.

[13] $E_1(x) = -\mathrm{Ei}(-x)$, where Ei is the exponential integral; *cf.* [9], definition 5.1.1, and formula 5.1.11, which implies that $E_1(x) < -\gamma - \ln x + x$ for positive x, where γ is the Euler–Mascheroni constant.

Proposition D.1. *The n-th derivative of the function $h(x) = \frac{1}{|x|}\mathrm{e}^{\frac{a}{x^2}}$ is given by*

$$h^{(n)}(x) = (-1)^n \frac{1}{x^n} h(x) P_n\left(\frac{a}{x^2}\right), \tag{D.1}$$

where $P_n(x)$ is the n-th degree polynomial

$$P_n(x) = \sum_{k=0}^{n} \begin{bmatrix} n+1 \\ k+1 \end{bmatrix} 2^k T_k(x). \tag{D.2}$$

Proof. It is easily seen that $P_n(x)$ in (D.1) must satisfy

$$P_0(x) = 1, \quad P_n(x) = \left(2x + n + 2x\frac{\mathrm{d}}{\mathrm{d}x}\right) P_{n-1}(x). \tag{D.3}$$

We show by induction that the solution of (D.3) is given by (D.2). The latter equation gives $P_0(x) = 1$, as required, since $\begin{bmatrix} 1 \\ 1 \end{bmatrix} = 1$ and $T_0(x) = 1$. Suppose (D.2) is correct for $n-1$. Then, using the well-known recurrence relation $T_n(x) = (x + x\frac{\mathrm{d}}{\mathrm{d}x})T_{n-1}(x)$, we get

$$\begin{aligned} P_n(x) &= \left(2\left(x + x\frac{\mathrm{d}}{\mathrm{d}x}\right) + n\right) \sum_{k=0}^{n-1} \begin{bmatrix} n \\ k+1 \end{bmatrix} 2^k T_k(x) \\ &= \sum_{k=0}^{n-1} \begin{bmatrix} n \\ k+1 \end{bmatrix} 2^{k+1} T_{k+1}(x) + \sum_{k=0}^{n-1} n \begin{bmatrix} n \\ k+1 \end{bmatrix} 2^k T_k(x) \\ &= \sum_{k=1}^{n} \begin{bmatrix} n \\ k \end{bmatrix} 2^k T_k(x) + \sum_{k=0}^{n-1} n \begin{bmatrix} n \\ k+1 \end{bmatrix} 2^k T_k(x). \end{aligned}$$

Since $\begin{bmatrix} n \\ 0 \end{bmatrix} = \begin{bmatrix} n \\ n+1 \end{bmatrix} = 0$, both sums can be extended to run from 0 to n. Finally, using the well-known recurrence relation $\begin{bmatrix} n+1 \\ k+1 \end{bmatrix} = n\begin{bmatrix} n \\ k+1 \end{bmatrix} + \begin{bmatrix} n \\ k \end{bmatrix}$, we end up with the right-hand side of (D.2). □

From (D.3) it is clear that $P_n(x)$ has positive integer coefficients. The same is true for the polynomials $\hat{P}_n(x) = P_n(\frac{x}{2})$, since $\hat{P}_n(x) = (x + n + 2x\frac{\mathrm{d}}{\mathrm{d}x})\hat{P}_{n-1}(x)$.

In order to investigate the convergence properties of the series for Θ in Section 7, we express the n-th derivative of h in a different way than above. ${}_pF_q$ denotes the generalized hypergeometric function, with ${}_2F_2(a_1, a_2; b_1, b_2; c)$ being defined as $\sum_{n=0}^{\infty} \frac{(a_1)_n \cdot (a_2)_n}{(b_1)_n \cdot (b_2)_n} \frac{c^n}{n!}$, where the Pochhammer symbol $(a)_n$ denotes rising factorials, *i.e.*, $(a)_0 = 1$ while $(a)_n = a \cdot (a+1)_{n-1}$ for positive integers n.

Proposition D.2. $h^{(n)}(x) = (-1)^n \cdot \frac{n!}{x^n|x|} \cdot {}_2F_2\left(\frac{n+1}{2}, \frac{n+2}{2}; \frac{1}{2}, 1; \frac{a}{x^2}\right)$.

Proof. This follows from Proposition D.1 and the relation

$${}_2F_2(\tfrac{n+1}{2}, \tfrac{n+2}{2}; \tfrac{1}{2}, 1; x) = \frac{\mathrm{e}^x}{n!} P_n(x), \tag{D.4}$$

which we show by induction. By definition of ${}_pF_q$, the left-hand side of (D.4) equals e^x for $n = 0$, as does the right-hand side, since $P_0(x) = 1$. Suppose (D.4) is true for $n-1$. Then, using the recurrence relation (D.3), and,

in the last step, a basic property of generalized hypergeometric functions,

$$\begin{aligned}\tfrac{1}{n!}\mathrm{e}^x P_n(x) &= \tfrac{1}{n!}\mathrm{e}^x\left(n+2x+2x\tfrac{\mathrm{d}}{\mathrm{d}x}\right)P_{n-1}(x) = \tfrac{1}{n!}\left(n+2x\tfrac{\mathrm{d}}{\mathrm{d}x}\right)\mathrm{e}^x P_{n-1}(x)\\ &= \tfrac{2}{n}\left(\tfrac{n}{2}+x\tfrac{\mathrm{d}}{\mathrm{d}x}\right){}_2F_2\left(\tfrac{n}{2},\tfrac{n+1}{2};\tfrac{1}{2},1;x\right) = {}_2F_2\left(\tfrac{n}{2}+1,\tfrac{n+1}{2};\tfrac{1}{2},1;x\right).\end{aligned}$$

□

From Proposition D.2 it follows that for any nonzero constant μ, $h(x)$ has the Taylor series $\frac{1}{|\mu|}\sum_{n=0}^{\infty} {}_2F_2\left(\frac{n+1}{2},\frac{n+2}{2};\frac{1}{2},1;\frac{a}{\mu^2}\right)\left(1-\frac{x}{\mu}\right)^n$ around μ. This series has a finite convergence radius of $|\mu|$, as mentioned in Section 7 for the case where $\mu=\mu_X$ and f is proportional to h with $a<0$. One might, therefore, question the convergence of the series for Θ in this case. From (6.1), Theorem 3.1, and Proposition D.2, this series is

$$\sum_{n=0}^{\infty}\frac{f^{(2n)}(\mu_X)}{(2n)!}(2n-1)!!\,\sigma_X^{2n} = \frac{1}{\sqrt{2\pi}|\mu_X|\sigma_Y}\sum_{n=0}^{\infty}{}_2F_2\left(n+\tfrac{1}{2},n+1;\tfrac{1}{2},1;-\tfrac{1}{2}\omega^2\right)(2n-1)!!\,\lambda^{2n}. \tag{D.5}$$

The series obviously diverges for all $\lambda\neq 0$ in the case that $\omega=0$, since then ${}_2F_2(\ldots)=1$. It can be shown that for nonzero λ, the series (D.5) for Θ is divergent for almost all values of ω. We omit the proof, that is based on (a) knowledge of the radius of convergence of the Taylor series discussed above, (b) the relation $(1+\frac{x}{a}\frac{\mathrm{d}}{\mathrm{d}x})\,{}_2F_2(a,b;c,d;x) = {}_2F_2(a+1,b;c,d;x)$, and (c) equation (D.4). The possible exceptions, *i.e.*, the values of ω for which the series might converge, constitute a set that has measure zero, and that we conjecture to be empty. Indeed, if the series were to converge for any nonzero λ and ω, then as a necessary condition ${}_2F_2\left(n+\frac{1}{2},n+1;\frac{1}{2},1;-\frac{1}{2}\omega^2\right)$ would have to converge to zero as n grows towards infinity, but instead all simulations indicate an oscillating behavior around zero with ever increasing amplitudes and periods.

Appendix E. Properties of the general series expansions

Here we prove some properties of the power series expansions for $\ln\Theta$ and the conditional expectations and variances that are discussed in Section 6 for the general case with nonzero κ. The first terms of these series are given in (6.5)–(6.12) and were derived from the power series for Θ that appears in truncated form in (6.4). It can be shown, similarly to the case where $\kappa=0$ discussed after (D.5), that for nonzero λ and every value of κ, the power series for Θ diverges for (at least) almost all values of ω; in other words, it is an asymptotic expansion in λ^2. The infinite series appearing below should therefore be regarded as formal power series.

Proposition E.1. *In the power series in λ^2 for $\ln\Theta$, the coefficient of λ^{2n} for $n\geq 1$ is a polynomial in ω and κ of total degree $2n+2$, where the coefficient of $\omega^i\kappa^j$ is nonzero if and only if $i+j$ is even and $i\geq j$, and has sign $(-1)^{\frac{i-j}{2}}$.*

Proof. We prove instead that the corresponding expansion in $\frac{1}{2}\lambda^2$ has the same properties; clearly, that result is equivalent to the proposition. With $t=\frac{1}{2}\sigma_X^2$ and $\frac{1}{2}\lambda^2=\mu_X^{-2}t$, the series for $\ln\Theta$ can be written as

$$\alpha = C-\ln(|\mu_X|\sigma_Y)-\tfrac{1}{2}\kappa^2+\sum_{n=0}^{\infty}A_n(\omega,\kappa)\mu_X^{-2n}t^n.$$

The discussion in Section 6 shows that every A_n is a polynomial in ω and κ; we have to show that the rest of the statements in the proposition are true for all A_n with $n\geq 1$. With termwise differentiation, the series α for $\ln\Theta$ satisfies (2.6), since, as noted after Theorem 3.1, the series for Θ satisfies (2.4). We use (2.6) to establish a

recursion relation for the coefficients of α. Now,

$$\frac{\partial\alpha}{\partial t} = \sum_{n=1}^{\infty} n\, A_n(\omega,\kappa)\mu_X^{-2n} t^{n-1} = \sum_{n=0}^{\infty}(n+1)A_{n+1}(\omega,\kappa)\mu_X^{-2n-2}t^n,$$

$$\frac{\partial\alpha}{\partial\mu_X} = \sum_{n=0}^{\infty} B_n(\omega,\kappa)\mu_X^{-2n-1}t^n \quad \text{with} \quad B_n(\omega,\kappa) = -\omega\frac{\partial A_n}{\partial\omega} - 2n\, A_n(\omega,\kappa) - \delta_{n,0},$$

$$\frac{\partial^2\alpha}{\partial\mu_X^2} = \sum_{n=0}^{\infty} C_n(\omega,\kappa)\mu_X^{-2n-2}t^n \quad \text{with} \quad C_n(\omega,\kappa) = -\omega\frac{\partial B_n}{\partial\omega} - (2n+1)B_n(\omega,\kappa),$$

$$\left(\frac{\partial\alpha}{\partial\mu_X}\right)^2 = \sum_{n=0}^{\infty}\sum_{m=0}^{n} B_m(\omega,\kappa)B_{n-m}(\omega,\kappa)\mu_X^{-2n-2}t^n,$$

where the exponents in the last equation follow from $[-2m-1]+[-2(n-m)-1] = -2n-2$ and $m+(n-m) = n$. Equation (6.6) shows that $A_0(\omega,\kappa) = -\frac{1}{2}\omega^2 + \omega\kappa$, and hence the statements in the proposition are true for A_0, except that the constant term is zero. Let $k \geq 0$, and suppose all the statements are true for every A_n with $n \leq k$, with an exception for the vanishing constant term of A_0. Then the expressions above show that B_n and C_n are polynomials in ω and κ for which the statements are true as well, except that the terms of B_n have the opposite sign. Using (2.6) and comparing the coefficients of t^k gives

$$(k+1)A_{k+1} = C_k + \sum_{m=0}^{k} B_m B_{k-m}.$$

C_k has terms of all relevant kinds up to degree $2k+2$. The products B_mB_{k-m} produce terms of the kind $\omega^{i_1}\kappa^{j_1}\cdot\omega^{i_2}\kappa^{j_2} = \omega^i\kappa^j$ where $i+j$ is even and $i = i_1+i_2 \geq j_1+j_2 = j$, and with sign $(-1)^{\frac{i_1-j_1}{2}+1}\cdot(-1)^{\frac{i_2-j_2}{2}+1} = (-1)^{\frac{i-j}{2}}$. Since every appearance of $\omega^i\kappa^j$ comes with this sign, none of the terms on the right-hand side cancel. The leading terms have total degree $2m+2+2(k-m)+2 = 2(k+1)+2$, and all the relevant ones are present in the products; for example, B_0B_k generates terms of the kind $\omega^2\cdot\omega^{2k+2} = \omega^{2k+4}$, $\omega^2\cdot\omega^{2k+1}\kappa = \omega^{2k+3}\kappa$, …, $\omega\kappa\cdot\omega^{k+1}\kappa^{k+1} = \omega^{k+2}\kappa^{k+2}$. Thus, the proposition follows by complete induction. □

Proposition E.2. *In the power series in λ^2 for $E[X|z]$, $\mathrm{Var}[X|z]$, $E[Y|z]$, and $\mathrm{Var}[Y|z]$, the coefficient of λ^{2n} in the bracket $[1+\ldots]$ is a polynomial in ω and κ of total degree $2n$, where the coefficient of $\omega^i\kappa^j$ is nonzero if and only if $i+j$ is even and $i \geq j$.*

Proof. Using Corollary 2.3 and the identities

$$\frac{\partial\omega}{\partial\mu_X} = -\frac{\omega}{\mu_X}, \quad \frac{\partial\lambda}{\partial\mu_X} = -\frac{\lambda}{\mu_X}, \quad \sigma_X^2\frac{1}{\mu_X} = \mu_X\lambda^2, \quad \sigma_Y^2\frac{\partial\kappa}{\partial\mu_Y} = \sigma_Y = \frac{z}{\mu_X}\omega^{-1},$$

the assertions follow from Proposition E.1 by inspection. □

Acknowledgements. We thank the anonymous referees for valuable comments and suggestions.

References

[1] T. Langholm and H. Totland, Approximating banana distributions based on constant, unknown acceleration and turn rate. *Global Oceans 2020: Singapore – U.S. Gulf Coast*. MS, USA (2020) 1–8. https://doi.org/10.1109/ieeeconf38699.2020.9389326.

[2] T. Langholm and H. Totland, Banana distributions based on stochastic polar coordinates. *Necesse* **6** (2021) 86–101. Modified and extended version of [1]. https://hdl.handle.net/11250/2832592

[3] M. Fisz, Probability Theory and Mathematical Statistics, 3rd edn. Wiley (1963).

[4] C.C. Craig, On the frequency function of *xy*. *Ann. Math. Stat.* **7** (1936) 1–15.

[5] L.A. Aroian, The probability function of the product of two normally distributed variables. *Ann. Math. Stat.* **18** (1947) 265–271.
[6] G.B. Folland, Real Analysis. Modern Techniques and Their Applications, 2nd edn. Wiley (1999).
[7] F.S. Acton, Numerical Methods that Work, updated and revised edn. Mathematical Association of America (1990).
[8] J. Soch, Solution for the indefinite integral of the standard normal probability density function. Preprint (2016). arXiv:1512.04858v3
[9] M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables*. National Bureau of Standards, Applied Mathematics Series (1964).