

**The random uncertainty distribution
of measurements of a complex quantity
with partially-correlated real and imaginary parts
(or any pair of partially-correlated variables)**

T E Hodgetts

Division of Electrical Science
National Physical Laboratory
% DRA Malvern, St Andrews Road
Malvern, Worcs, WR14 3PS
United Kingdom

ABSTRACT

This report presents a rigorous derivation, from first principles, of the random uncertainty distribution of a sample of measurements of a complex number, allowing for partial linear correlation between the complex number's real and imaginary parts, and assuming that the underlying law of joint distribution of the real and imaginary parts is the law of Gaussian normal correlation. This is the equivalent in two dimensions of the "Student's t" distribution in one dimension. The principal result is an expression for the probability that the underlying true value of the complex number is in the neighbourhood of any particular point in the complex plane, in terms of the means of the sample's real and imaginary parts, their standard deviations and their correlation coefficient. This probability density distribution is such that the contours of equal probability density are ellipses centred on the sample's two-dimensional mean, and a simple closed-form expression is given for the probability that the true value is inside any particular ellipse, which relates directly to the customary expression by confidence levels of uncertainties in measurement.

© Crown Copyright 1995

ISSN 0143-7305

National Physical Laboratory
Teddington, Middlesex, UK, TW11 0LW

Extracts from this report may be reproduced
provided that the source is acknowledged.

Approved on behalf of Chief Executive, NPL
by Dr T G Blaney, Head, Division of Electrical Science

CONTENTS

	Page
INTRODUCTION	.. 1
PART I	... 1
PART II	... 17
PART III	48
PART IV . .	73
CONCLUSIONS	88
REFERENCES	.. 89
APPENDIX	90

INTRODUCTION

In theoretical work on the statistics of measurements, it is normally assumed that the result of a repeated measurement is a single number, and the purpose of applying a statistical treatment is to provide an estimate of that number's uncertainty. Some theory exists for analysing correlations between several quantities measured simultaneously; but the customary treatment is not rigorous, since it involves introducing the Gaussian (normal) error distribution without taking any account of the underlying assumptions – which are often falsified by circumstances, in particular if the number of repetitions of the measurement is not very large. This paper presents a rigorous analysis, from first principles, of the probability distribution of repeated measurements of a complex number, allowing for partial linear correlation between the number's real and imaginary parts. Measurements of complex numbers are continually being made in applications of electrical standards, because of their use as "phasors" or "vectors" in RF and AC theory, and this work was undertaken to help improve the accuracy of such measurements on microwave network analysers; but the mathematics is completely general and will apply in any measurement context where complex numbers arise naturally, or indeed to any pair of partially-correlated variables.

To demonstrate the consistency of this work, it is necessary to show how the familiar statistical distributions follow from the same underlying assumptions as are made to derive the new probability distribution. I begin, therefore, by examining the derivations of the binomial theorem, the normal distribution and the "Student's t" distribution, following the treatment given by Jeffreys [1] (whose book has guided the logical development of this work). This part of the paper (Part I) may be skipped by readers who are uninterested in its subject matter.

PART I

Elementary probabilities illustrated by the binomial theorem

I consider first the simplest possible type of repeated "measurement", which consists of performing some procedure which gives an answer to a "yes/no" question – such as tossing a coin and asking "was it a head?", or throwing a dice and asking "was it a six?". It is assumed that a "yes/no" answer is always given to the question; I exclude coins which fail to come down when tossed, or which land edgewise. It is also assumed that an answer does not depend on previous answers. It then follows that, if the underlying or prior probability of obtaining the answer "yes" is y , and that of obtaining the answer "no" is x , then $y+x=1$; in the customary jargon, the two answers are mutually exclusive and exhaustive. Furthermore, the joint probability that $(\ell+m)$ "measurements" will give "no" ℓ times and "yes" m times, in any particular specified order, is $x^\ell y^m$. The number of possible orders is determined from the theory of combinations and permutations as $(\ell+m)!/(\ell!m!)$, and the prior probability of any of these orders is the same as that of any other; so, by the usual addition rule for exclusive probabilities,

$$P(\ell \& m) = \frac{(\ell+m)!}{\ell!m!} x^\ell y^m \quad (1)$$

This expression is a typical term in the binomial expansion of $(x+y)^{\ell+m}$. If $(\ell+m)$ is chosen and ℓ is allowed to run over all the possible values from 0 to $(\ell+m)$, with m varying correspondingly, then the sum of all the mutually exclusive values of $P(\ell \& m)$ is $(x+y)^{\ell+m}$, which is 1 for any number of repetitions; this is to be expected, since all the possibilities have been exhausted.

The line of argument just given contains a number of imperfectly-stated assumptions; I have introduced the ideas of prior probability, independent trials, equally-likely alternatives, exclusiveness, exhaustiveness, and the addition and multiplication rules for combining elementary probabilities into compound ones. Jeffreys [1] presents a very detailed discussion of these, referring them all back to a consistent set of axioms and conventions. I shall not try to reproduce all his material, but some general comments are in order here; more will be said as the arguments progress.

Probability theory is not, and cannot be, completely rigorous in the sense of proceeding by inexorable logic from a few unchallengeable assumptions; it makes predictions which can be used to test the assumptions from which they were deduced. Eqn (1) above is such a prediction, or can be made so with a little more work; and its validity (for the kind of repeated measurement to which it is intended to apply) is not seriously disputed by any of the schools of the philosophy of statistics, which is a good practical indication that the philosophical debates about such matters as prior probability are irrelevant in such simple cases. The problem becomes rather more difficult when the number of possible outcomes (or answers to the generalised question) becomes infinite, as in the form of a continuous variable; and more will be said about this in due course. In the meantime, it should be noted that the derivation of eqn (1) does not depend on assuming any particular values for x and y , only that the values do not change as the experiment proceeds (and sum to 1 for consistency). If, then, eqn (1) is used to predict the values of ℓ and m arising out of a long series (or even a short series) of repeated measurements, the values of x and y can be inferred from the measurements and compared with a priori predictions – such as $1/2$ and $1/2$ when looking for heads, or $5/6$ and $1/6$ when looking for sixes. Such comparisons test both the physics underlying a probability law (such as eqn (1)) and the physics underlying the predicted values of the law's parameters; so probability theory is a branch of theoretical physics, and the role of mathematical logic within it is not to prove that it is 'true' but to prove that it is consistent. This hybrid dependence on logic and experimental observations is summed up by the much-quoted and variously-translated remark attributed to Lippman by Poincare (see Whittaker and Robinson [2]): "Everybody believes in the exponential law of errors: the experimenters, because they think it can be proved by mathematics; and the mathematicians, because they believe it has been established by observation".

The normal distribution (de Moivre's limit of the binomial law)

Moving on from these general remarks, I consider next what happens to the binomial law when the number of measurements, $(\ell+m)$, in eqn (1) becomes large. Suppose that neither of the underlying probabilities x and y is very small (if this is not true, the binomial law still possesses a limiting form, but the limit is the Poisson distribution; the interested reader may pursue this in [1], for instance). It then follows that $(\ell+m)$, ℓ and m all become large together, and may be well-approximated using Stirling's asymptotic expansion (e.g. Jeffreys and Jeffreys [3]):

$$\ln z! = \ln \Gamma(z+1) = (z+\frac{1}{2}) \ln z - z + \frac{1}{2} \ln(2\pi) + O(z^{-1}) \quad (2)$$

From the definitions of the underlying probabilities, it is to be expected that ℓ will be near the fraction x of the total number of measurements $(\ell+m)$, and similarly for y ; I write

$$n \equiv \ell+m \quad (3a); \quad \ell = nx + \delta \quad (3b); \quad m = ny - \delta \quad (3c)$$

where n is introduced for convenience, and eqn (3c) follows from (3a) and (3b) since $x+y=1$. Now, from eqns (2) and (3),

$$\ln (\ell+m)! = (\ell+m+\frac{1}{2}) \ln n - (\ell+m) + \frac{1}{2} \ln(2\pi) + O(n^{-1}) \quad (3d)$$

$$\ln m! = (m + \frac{1}{2}) \ln m - m + \frac{1}{2} \ln(2\pi) + O(n^{-1})$$

$$\ln \ell! = (\ell + \frac{1}{2}) \ln \ell - \ell + \frac{1}{2} \ln(2\pi) + O(n^{-1})$$

and so from eqn (1)

$$\ln P(\ell \& m) = \ell \ln \left(\frac{nx}{\ell} \right) + m \ln \left(\frac{ny}{m} \right) - \frac{1}{2} \ln \left(\frac{2\pi \ell m}{n} \right) + O(n^{-1})$$

$$\left(\ell + \frac{1}{2} \right) \ln \left(\frac{\ell}{nx} \right) - \left(m + \frac{1}{2} \right) \ln \left(\frac{m}{ny} \right) - \frac{1}{2} \ln(2\pi nxy) + O(n^{-1}) \quad (3g)$$

It then follows, on substituting for ℓ and m from eqns (3) and using the well-known expansion

$$\ln(1+z) = z - \frac{1}{2} z^2 + O(z^3),$$

that

$$\begin{aligned} \ln P(\ell \& m) = & - \left[nx + \delta + \frac{1}{2} \right] \left[\frac{\delta}{1} - \frac{1}{2} \left(\frac{\delta}{nx} \right)^2 \right] - \left[ny - \delta + \frac{1}{2} \right] \left[\frac{\delta}{ny} - \frac{1}{2} \left(\frac{\delta}{ny} \right)^2 \dots \right] \\ & + \frac{1}{2} \ln(2\pi nxy) + O(n^{-1}) + O(n^{-2}\delta^3) \end{aligned}$$

$$- \frac{1}{2} \frac{\delta^2}{nx} - \frac{1}{2} \frac{\delta^2}{ny} - \frac{1}{2} \ln(2\pi nxy) + O(n^{-1}) + O(n^{-2}\delta^3) + O(n^{-1}\delta) \quad (4)$$

In eqn (4), strictly speaking, δ can only take discrete values separated by unity, since ℓ and m in eqns (3) are integers. However, for large n , the difference between the values of $\ln P$ for successive values of δ is of order $(n^{-1}\delta)$ and must be small for all important values of δ , since if δ is much larger than $n^{1/2}$ then the corresponding value of $\ln P$ is much smaller than the maximum value it attains near $\delta=0$ (not exactly at $\delta=0$, will be seen shortly). The discrete distribution may therefore be replaced by the corresponding probability density distribution; this replacement is explicitly presented here because it occurs implicitly several times later on. The same restriction, δ not much larger than $n^{1/2}$, is sufficient to make the $O(\)$ terms in eqn (4) generally negligible, both relatively and absolutely. Hence, after exponentiating both sides of eqn (4), it may be written (after slight simplification) as de Moivre's limiting form

$$P(\delta) = (2\pi nxy)^{-1/2} \exp(-\delta^2/(2nxy)) \quad (5)$$

where the probability of finding δ to be between two given values is obtained by integrating $P(\delta)$ between them, in conformity with the customary definition of a probability density distribution. It can be seen that this integration gives 1 (representing the whole probability of all possible values of δ) only if the limits of integration are extended to $\pm \infty$; this is not strictly possible, since the true limits (from eqns (3)) are $+ny$ and $-nx$, but these limits are far outside the significant range

defined by " δ not much larger than $n^{1/2}$ ". Trivial normalisation errors of this kind often occur when changing between finite discrete and infinite continuous distributions, as will be seen later.

So far, the analysis in this section has followed Jeffreys [1]. I now extend it slightly, as an introduction to the next section, by examining some of the apparently negligible $O(\)$ terms in eqn (4).

Firstly, consider the terms $O(n^{-1})$, which come from eqn (2). If I choose to neglect terms of order (z^{-3}) in this equation, but not larger terms, I must include explicitly the first term of the asymptotic remainder series, which is $+1/(12z)$. After making the appropriate corrections to eqns (3d) – (3f), reviewing the derivation of eqn (4) gives

$$\begin{aligned} \ln P(\ell \& m) = & - \left[nx + \delta + \frac{1}{2} \right] \left[\frac{\delta}{nx} - \frac{1}{2} \left(\frac{\delta}{nx} \right)^2 \right] \\ & \left[ny - \delta + \frac{1}{2} \right] \left[-\frac{\delta}{ny} - \frac{1}{2} \left(\frac{\delta}{ny} \right)^2 \right] - \frac{1}{2} \ln(2\pi nxy) \\ & + \frac{1}{12n} \left[1 - \frac{1}{x} \left(1 + \frac{\delta}{nx} \right)^{-1} - \frac{1}{y} \left(1 - \frac{\delta}{ny} \right)^{-1} \right] \\ & + O(n^{-3}) + O(n^{-2}\delta^3) \end{aligned}$$

or

$$\begin{aligned} \ln P(\ell \& m) = & \frac{1}{2} \frac{\delta^2}{nx} - \frac{1}{2} \frac{\delta^2}{ny} - \frac{1}{2} \ln(2\pi nxy) \\ & - \frac{1}{2} \frac{\delta}{nx} + \frac{1}{2} \frac{\delta}{ny} \\ & + \frac{1}{12n} \left[1 - \frac{1}{x} - \frac{1}{y} \right] + O(n^{-2}\delta) \\ & + O(n^{-3}) + O(n^{-2}\delta^3) \end{aligned} \tag{6}$$

In the third line of eqn (6), there are terms independent of δ which contribute to the normalising factor in the revised form of eqn (5); as before, it is clear that changing to an infinite continuous form of a distribution induces a normalising error, but this analysis shows that the error is much larger than appeared at first sight – in eqn (5) it was exponentially small, but now it is of order (n^{-1}) . This is a general result; it will occur again.

Secondly, consider the terms $O(n^{-1}\delta)$ in eqn (4); these are explicitly presented in the second line of eqn (6) (so avoiding re-deriving eqn (4) twice). Allowing for these, it appears that the maximum of $P(\ell \& m)$ (or of $\ln P(\ell \& m)$) is not at $\delta=0$, but is at the zero of the derivative of the right-hand side of eqn (6) with respect to δ , which (neglecting the smaller $O(\)$ terms) is given by

$$\frac{\delta_m}{n} \left[\frac{x+y}{xy} \right] + \frac{1}{2n} \left[\frac{x-y}{xy} \right] = 0 + O(n^{-2})$$

or

$$\delta_m = \frac{1}{2} (x-y) + O(n^{-1}) \quad (7)$$

Eqn (7) shows that in general there is a natural offset involved in δ . It is connected with the asymmetry of the general binomial ($x \neq y$); this has disappeared in eqn (5), which is therefore somewhat unsatisfactory since it does not tell the whole story. A partial solution to this difficulty is obtained by inserting eqn (7) into eqn (6), giving

$$\ln P(\ell \& m) = -\frac{(\delta - \delta_m)^2}{2nxy} - \frac{1}{2} \ln(2\pi nxy) + \frac{1}{24n} \left[-10 + \frac{1}{xy} \right] + O(n^{-1}) \quad (8)$$

which equation has somewhat unsatisfactory symmetry (and a new form for the term $O(n^{-1})$ which controls the normalising error). There is also an arbitrary and radical solution to the asymmetry problem; if x is set equal to y , the problem disappears (as does δ_m). The consequences of this will be examined shortly; but first there is a little to say about eqn (8), which may be transformed into

$$P(\delta) = (2\pi nxy)^{-1/2} \exp(-(\delta - \delta_m)^2 / (2nxy)) \quad (9)$$

(corresponding to eqn (5)) on resetting the normalising error to be exponentially small. This form raises a difficulty over consistency, because the expectation or natural mean value of δ is now given by

$$\int_{-\infty}^{\infty} \delta P(\delta) d\delta = \int_{-\infty}^{\infty} \delta (2\pi nxy)^{-1/2} \exp(-(\delta - \delta_m)^2 / (2nxy)) d\delta$$

which (neglecting exponential errors and higher order terms written as $O(\)$ in eqn (8)) is δ_m , not zero. This contradicts the implication of the definitions in eqns (3), that ℓ and m ought to tend respectively to (nx) and (ny) as n tends to infinity; at least, the spirit of the definitions is violated, although the "letter of the law" is observed since (ℓ/n) and (m/n) correctly approach x and y (respectively) as n tends to infinity. (The nature of the approach to the expected values x and y is discussed by Jeffreys [1], who shows that it is considerably more complicated than it appears at first sight, although he does not note the difficulty I have just described.) However, as already remarked, the difficulty disappears if x is set equal to y , when δ_m is zero by eqn (7). I consider next the derivation of the normal distribution in its usual form by the aid of this trick.

The normal distribution (limit of multiple independent small perturbations)

Suppose that some quantity has a true value of λ , but that every measurement of it is affected by perturbations, n in number, each of which can affect any particular measurement by $+\epsilon$ or $-\epsilon$ with equal probability $\frac{1}{2}$ and each of which is independent of all the others. This is a case of the binomial law; the reason for choosing equal probabilities for $+\epsilon$ and $-\epsilon$ has already been discussed in terms of symmetry, and the restriction that ϵ is the same for every perturbation will be partly relaxed later. Using some of the notation of the previous section, if ℓ of the perturbations of a particular measurement are positive and the remaining $(n-\ell)$ are negative, then the measured value will be $\lambda + \ell\epsilon - (n-\ell)\epsilon$ or $\lambda + (2\ell - n)\epsilon$; and for large n the probability of obtaining a particular value ℓ will be given by eqns (9) and (3) with $x=y=\frac{1}{2}$ (so that $\delta = \ell - \frac{1}{2}n$ and $\delta_m = 0$), which is

$$P(\ell) = \left(\frac{2}{\pi n} \right)^{1/2} \exp(-2(\ell - \frac{1}{2}n)^2 / n) \quad (10)$$

Then the probability of finding ℓ in the range ℓ_1 to ℓ_2 inclusive ($\ell_2 > \ell_1$) is

$$P(\ell_1 \leq \ell \leq \ell_2) = \sum_{\ell=\ell_1}^{\ell_2} \left(\frac{2}{\pi n} \right)^{1/2} \exp(-2(\ell - \frac{1}{2}n)^2/n) \quad (11a)$$

which is also the probability that the measured value (which will be called z) will lie in the range $z_1 = \lambda + (2\ell_1 - n)\epsilon \leq z \leq \lambda + (2\ell_2 - n)\epsilon = z_2$, so

$$P(z_1 \leq z \leq z_2) = \sum_{z=z_1}^{z_2} \left(\frac{2}{\pi n} \right)^{1/2} \exp(-(z-\lambda)^2/(2n\epsilon^2))$$

Suppose that the range z_1 to z_2 is much larger than ϵ but much smaller than $\epsilon\sqrt{n}$; the sum in eqns (11) then approximates to

$$P(z_1 \leq z \leq z_2) = \int_{z_1}^{z_2} \left(\frac{2}{\pi n} \right)^{1/2} \exp\left[-\frac{(z-\lambda)^2}{2n\epsilon^2}\right] \left(\frac{dz}{2\epsilon}\right)$$

since the number of values of ℓ covering this range is $(1+(z_2-z_1)/2\epsilon)$. If ϵ is of such magnitude that $\epsilon\sqrt{n}$ is finite (neither very large nor very small), then the probability "density" within the range z_1 to z_2 is essentially constant, and so

$$P(z_1 \leq z < (z_1 + dz)) = (2\pi)^{-1/2} (\epsilon\sqrt{n})^{-1} \exp\left[-\frac{(z-\lambda)^2}{2(\epsilon\sqrt{n})^2}\right] dz$$

which is normally written in the form

$$P(z) dz = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(z-\lambda)^2}{2\sigma^2}\right] dz$$

This is the so-called normal distribution, usually credited to Gauss but apparently [1,2] due to Laplace. As before, it is considered to extend to infinity, although this is strictly only possible if n is allowed to tend to infinity (which is also required for exact normalisation). The parameter σ is usually called the standard deviation; it is in fact the root-mean-square deviation, as follows from using eqn (12b) to evaluate the mean square deviation (also called the second moment or the variance), which is

$$\int_{-\infty}^{\infty} (z-\lambda)^2 P(z) dz = \sigma^2 \quad (12c)$$

The artificial and restrictive conditions on the behaviour of the perturbations can be considerably relaxed by appealing to the various forms of the central limit theorem, and to the reproductive property of the normal distribution (the resultant sum of any number of normal distributions is another normal distribution, which allows groups of perturbations having different values of ϵ to be compounded together while preserving the normal form). Jeffreys [1] presents a careful discussion of the central limit theorem, exposing some bogus arguments which have been presented as simple "proofs" of it (eg in [2]). In fact, as remarked earlier, the normal distribution cannot be proved to hold in a real situation without considering the physics as well as the mathematics; but it is much more generally valid than would follow from the argument presented above. My reason for presenting it in that way is that the argument extends immediately to the case of two correlated variables, which is what I really want to discuss; and this case cannot be

safely inferred from the central limit theorem because all forms of it assume that the perturbations are essentially independent of each other, which disallows correlation at the outset.

The primitive distribution for two correlated variables

I proceed now to apply the argument of the last section to the case of two correlated variables. Suppose that two quantities u and v are to be measured simultaneously, both having (for convenience) true values of zero (which can be shifted later as required). Suppose further that there are $(m+n)$ independent perturbations, each affecting u by $+\alpha$ or $-\alpha$ and v by $+\beta$ or $-\beta$, with equal probabilities $\frac{1}{2}$; and that m of the perturbations are constrained to affect u and v in the same sense (both positive or both negative) while the other n perturbations affect u and v in opposite senses (one positive and the other negative), without compromising the "independence" of the perturbations in the sense of the last section. (Jeffreys [1] does not consider in his presentation how the "independence" is to be uncompromised, and it is best if the reader does not do so either at present; more will be said later.) If, for a particular measurement, p of the m "same sense" perturbations affect u and v positively while q of the n "opposite sense" perturbations affect u positively and v negatively, then the measured values are given in terms of p and q by

$$\begin{aligned} u &= p\alpha + (m-p)(-\alpha) + q\alpha + (n-q)(-\alpha) = (2p-m)\alpha + (2q-n)\alpha \\ v &= p\beta + (m-p)(-\beta) + q(-\beta) + (n-q)\beta = (2p-m)\beta - (2q-n)\beta \end{aligned} \quad (13b)$$

and by the argument of the previous section (see eqn (10)) the corresponding binomial probability is

$$P(p, q) = \frac{2}{\pi(mn)^{1/2}} \exp \left\{ -\frac{2}{m} \left(p - \frac{1}{2}m\right)^2 - \frac{2}{n} \left(q - \frac{1}{2}n\right)^2 \right\} \quad (14)$$

As in the previous section, the sum of this over a range of discrete but closely spaced values of p and q can be replaced by an integral over the corresponding essentially continuous ranges of u and v ; the sum and the integral are both double, so the result depends on the Jacobian determinant of the transformation, which is

$$\frac{\partial(u, v)}{\partial(p, q)} = 8\alpha\beta \quad (15a)$$

from eqns (13), which also give

$$p - \frac{1}{2}m = \frac{1}{4} \left[\frac{u}{\alpha} + \frac{v}{\beta} \right]$$

$$q - \frac{1}{2}n = \frac{1}{4} \left[\frac{u}{\alpha} - \frac{v}{\beta} \right]$$

Using eqns (15), by analogy with eqns (11) and (12a), eqn (14) becomes

$$P(u, v) du dv = \frac{1}{4\pi\alpha\beta\sqrt{mn}} \exp \left\{ -\frac{1}{8m} \left[\frac{u}{\alpha} + \frac{v}{\beta} \right]^2 - \frac{1}{8n} \left[\frac{u}{\alpha} - \frac{v}{\beta} \right]^2 \right\} du dv$$

and with the new parameters

$$\sigma = \alpha(m+n)^{1/2} \quad (16b)$$

$$\tau = \beta(m+n)^{1/2} \quad (16c)$$

$$\rho = \alpha\beta(m-n)/(\sigma\tau) \quad (16d)$$

this becomes, analogously to eqn (12b) (and with $1 > \rho > -1$),

$$P(u,v)dudv = \frac{(1-\rho^2)^{-1/2}}{2\pi\sigma\tau} \exp\left\{-\frac{1}{2(1-\rho^2)} \left[\frac{u^2}{\sigma^2} - \frac{2\rho uv}{\sigma\tau} + \frac{v^2}{\tau^2}\right]\right\} dudv \quad (16e)$$

This is the distribution of normal correlation. According to Jeffreys [1] it was actually discovered by inference from observed frequencies, which should encourage the reader to believe that it is valid in more plausible circumstances than this crude derivation would suggest. The assumption that the perturbations are, simultaneously, "independent" and have precisely-specifiable effects on one variable relative to the other is the most crude part of the analysis. However, it is less dubious in the present application than in general, because it amounts to assuming that "independence" only applies to one of u and v – when the perturbations of one are inferred from a measurement, the other is fixed – and this is quite plausible when u and v are the real and imaginary parts of a complex number, which are a pair of mathematical Siamese twins. Moreover, there are several general points in favour of this distribution. If only one of the variables is measured, its expected distribution is obtainable by integrating over the range of the other one; the normal distribution is then recovered, with σ as the standard deviation for u or τ as the standard deviation for v . Similarly, if one variable is fixed, the distribution of the other is found to be normal, with (for non-zero ρ) a smaller standard deviation than before and a shift of the expected value from zero by an amount proportional to ρ (both in magnitude and in sign). If ρ is zero, eqn (16e) reduces to the product of the two normal distributions which correspond to the measurement of either variable alone, so the uncorrelated case is included in the correlated one. Also, the normal correlation distribution has its own reproductive property, at least in the following carefully-defined sense: if the quantities u and v are subject to several sets of perturbations, with each set correlated within itself but independent of the other sets, then if each set acting alone would produce a distribution of normal correlation, the resultant sum of the sets reproduces that distribution.

These considerations are comforting; but the main justification for the use of the two-dimensional normal correlation distribution is the same as that for the use of its one-dimensional precursor, the normal distribution: it appears to work in practice (as shown by the manner of its discovery) and it is convenient to manipulate in theory (very few other distributions have a reproductive property, for instance). In the words of Jeffreys [1], "...everything that can be said against the normal law of error for one variable can be said twice against this ... generalization to two variables ... on the other hand the chief thing that can be said in favour of the normal law, that of all laws that are anywhere near the truth it is far the easiest to apply, can also be said with greater force of normal correlation".

One other property of the distribution of normal correlation will be remarked here, and used frequently later without further comment: the argument of the exponential factor in it is negative-definite. This is obvious from eqn (16a), and almost obvious from eqn (16e).

The "Student's t" distribution of samples from the normal distribution

I now begin to consider the consequences of the normal distribution. Since it represents the infinite limit of the resultant sum of a large number of perturbations, it might be thought that the arithmetic mean of a finite number of measurements (a "sample") of a normally-distributed quantity would necessarily also be normally-distributed (the total number of perturbations being essentially infinite in either case); this is implicitly assumed by Whittaker and Robinson [2], for instance. In fact, the probability of obtaining a particular sample mean may be distributed differently with respect to the true value of the quantity being measured, depending on the assumptions made. Consider a sample of n measurements of a quantity which is normally distributed with true value λ and standard deviation σ . I introduce the notation in which, for instance, " $P(dx|\lambda)$ " means "the probability that the measurable x lies between x and $(x+dx)$, subject to the given parameter value λ "; then eqn (12b) becomes

$$P(dx|\lambda, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\lambda)^2}{2\sigma^2}\right] dx \quad (17)$$

In this notation, the joint probability of obtaining the n measured values (strictly, values in infinitesimal ranges enclosing each of them) is

$$P(dx_1 dx_2 \dots dx_n | \lambda, \sigma) = \frac{1}{(2\pi)^{n/2} \sigma^n} \exp\left\{-\frac{\sum_{r=1}^n (x_r - \lambda)^2}{2\sigma^2}\right\} dx_1 \dots dx_n$$

I write

$$\bar{x} \equiv \frac{\sum_{r=1}^n x_r}{n}$$

and

$$s \equiv \left[\sum_{r=1}^n (x_r - \bar{x})^2 / n \right]^{1/2} \quad (18c)$$

so that $\bar{x}$ is the mean of the sample and s is the sample standard deviation (note that this is defined with the divisor n , not $(n-1)$ as is frequently done). Also

$$\begin{aligned} \sum_{r=1}^n (x_r - \lambda)^2 &= \sum_{r=1}^n \left\{ (x_r - \bar{x}) + (\bar{x} - \lambda) \right\}^2 \\ &= \sum_{r=1}^n (x_r - \bar{x})^2 + 2(\bar{x} - \lambda) \sum_{r=1}^n (x_r - \bar{x}) + \sum_{r=1}^n (\bar{x} - \lambda)^2 \\ &= ns^2 + 2(\bar{x} - \lambda)(n\bar{x} - n\bar{x}) + n(\bar{x} - \lambda)^2 \\ &= ns^2 + n(\bar{x} - \lambda)^2 \end{aligned} \quad (18d)$$

so

$$P(dx_1 dx_2 \dots dx_n | \lambda, \sigma) = \frac{1}{(2\pi)^{n/2} \sigma^n} \exp \left\{ -\frac{n[s^2 + (\bar{x} - \lambda)^2]}{2\sigma^2} \right\} dx_1 \dots dx_n \quad (18e)$$

The right-hand side of this equation contains (apart from the infinitesimals and the parameters λ, σ) only $\bar{x}$, s and n ; the details of the observations have disappeared, to be replaced by quantities which characterise the sample as a whole. It is therefore natural to ask for the corresponding joint probability $P_n(d\bar{x}ds|\lambda, \sigma) (\equiv P(d\bar{x}ds|\lambda, \sigma, n))$, which is formally obtained by considering a pair of infinitesimal ranges enclosing $\bar{x}$ and s and performing an $(n-2)$ - fold integration of eqn (18e) over the ranges of the x_r which give values in the chosen infinitesimal ranges of $\bar{x}$ and s . This requires an algebraic transformation which may be expressed as follows. If the x_r are regarded

as the rectangular co-ordinates of a point in n -dimensional space, then the quantity $\sum_{r=1}^n (x_r - \lambda)^2$ in eqn (18a) is the distance of this point from the point whose co-ordinates are all λ . This property is preserved after any general rotation of the axes of the co-ordinates; if such a rotation transforms the x_r into corresponding x'_i ($i=1, 2, \dots, n$), then the relation between the two sets of co-ordinates is

$$x'_i = \sum_{r=1}^n a_{ir} x_r \quad (i=1, 2, \dots, n)$$

where the a_{ir} are the elements of an $(n \times n)$ orthogonal matrix, so that

$$\sum_i a_{ir}^2 = 1 \quad (19b)$$

$$\sum_r a_{ir}^2 = 1$$

$$\sum_r a_{ir} a_{jr} = 0$$

for all i and r , and with $j \neq i$ in eqn (19d). Let x'_1 be chosen as

$$x'_1 = \left[\sum_{r=1}^n x_r \right] / \sqrt{n} = \bar{x} \sqrt{n}$$

for which $a_{1r} = 1/\sqrt{n}$, which is consistent with eqn (19c). Using this with eqn (19a), it follows that the point whose co-ordinates are all λ is transformed into the point $(\lambda\sqrt{n}, 0, 0, \dots, 0)$, since

$$\sum_i x_i'^2 = \sum_r x_r^2. \text{ Then}$$

$$\sum_r (x_r - \lambda)^2 = \sum_{i \neq 1} (x'_i)^2 + (x'_1 - \lambda\sqrt{n})^2$$

(the distance-preserving property), and from eqn (18d)

$$\sum_{r=1}^n (x_r - \lambda)^2 - n(\bar{x} - \lambda)^2 = \sum_{r=1}^n (x_r - \lambda)^2 - (x'_1 - \lambda\sqrt{n})^2 = ns^2$$

so

$$\sum_{i \neq 1} x_i'^2 = ns^2$$

Also, the Jacobian determinant of the transformation from the x_r to the x_i' is the determinant of the matrix whose elements are the a_{ir} , and is 1 by the orthogonal properties (eqns (19b)–(19d)). Therefore, from eqns (18a) or (18e),

$$P(dx'_1 dx'_2 \dots dx'_n | \lambda, \sigma) = P(dx_1 dx_2 \dots dx_n | \lambda, \sigma)$$

$$\begin{aligned} & \frac{1}{(2\pi)^{n/2} \sigma^n} \exp \left\{ -\frac{ns^2}{2\sigma^2} - \frac{(x'_1 - \lambda\sqrt{n})^2}{2\sigma^2} \right\} dx_1 dx_2 \dots dx_n \\ & \frac{1}{(2\pi)^{n/2} \sigma^n} \exp \left\{ -\frac{ns^2}{2\sigma^2} - \frac{(x'_1 - \lambda\sqrt{n})^2}{2\sigma^2} \right\} dx'_1 dx'_2 \dots dx'_n \\ & \frac{1}{(2\pi)^{n/2} \sigma^n} \exp \left\{ -\frac{(x'_1 - \lambda\sqrt{n})^2}{2\sigma^2} \right\} dx'_1 \cdot \exp \left\{ -\frac{ns^2}{2\sigma^2} \right\} dx'_2 \dots dx'_n \end{aligned} \quad (20)$$

From eqn (19e), the first factor here is proportional to

$$\exp \left\{ -\frac{n}{2\sigma^2} (\bar{x} - \lambda)^2 \right\} d\bar{x} \quad (21a)$$

From eqn (19g), the result of integrating the second factor here ($n-2$) times is a one-dimensional differential form, representing an infinitesimal hyperspherical shell of hyperradius s , which is obtained by transforming the co-ordinates $x'_2, x'_3, \dots, x'_n$ into hyperspherical polar co-ordinates and integrating over the entire ranges of the Eulerian angles in the polar representation. The Jacobian determinant of this transformation contains a factor s^{n-2} (compare $\rho d\rho$ for plane polar co-ordinates, $\rho^2 d\rho$ for ordinary spherical polar co-ordinates, etc) which is preserved through the angular integrations; the second factor in eqn (20) is therefore proportional to

$$s^{n-2} \exp \left\{ -\frac{ns^2}{2\sigma^2} \right\} ds$$

and so eqn (20) becomes (with a suffix introduced for later convenience)

$$P_n(d\bar{x}ds|\lambda, \sigma) = P(dx'_1 dx'_2 \dots dx'_n | \lambda, \sigma) \\ \propto \frac{1}{\sigma} \exp \left\{ -\frac{n}{2\sigma^2} (\bar{x} - \lambda)^2 \right\} d\bar{x} \cdot \frac{s^{n-2}}{\sigma^{n-1}} \exp \left\{ -\frac{ns^2}{2\sigma^2} \right\} ds$$

where the constant of proportionality is fixed by integrating over all allowed values of $\bar{x}$ ($-\infty$ to $+\infty$) and of s (0 to ∞). This joint probability distribution is of relatively little interest in itself; what is of real interest is the pair of distributions representing $\bar{x}$ in terms of σ and of s . The first of these is obtained simply by integrating eqn (22) over all s and normalising the result, and is

$$P_n(d\bar{x}|\lambda, \sigma) = \sqrt{\frac{n}{2\pi}} \frac{1}{\sigma} \exp \left\{ -\frac{n}{2\sigma^2} (\bar{x} - \lambda)^2 \right\} d\bar{x}$$

which is evidently a normal distribution with mean λ and standard deviation $(\sigma/\sqrt{n})$. This is the basis of the custom of averaging repeated measurements of a quantity which is known (or supposed) to be normally distributed; the expectation of a sample mean is the true value and is more tightly bound to that value than the individual measurements are. However, this result is obtained by implicitly measuring $(\bar{x} - \lambda)$ in units of σ ; if instead I choose to measure it in units of s , a striking difference appears. Let

$$\bar{x} - \lambda = sz \tag{23b}$$

and transform from variables $\bar{x}$ and s to variables z and s , for which the Jacobian determinant is s . Now, from eqn (22),

$$P_n(dzds|\lambda, \sigma) = P_n(d\bar{x}ds|\lambda, \sigma) \\ \propto \frac{s^{n-2}}{\sigma^n} \exp \left\{ -\frac{ns^2 z^2}{2\sigma^2} - \frac{ns^2}{2\sigma^2} \right\} \cdot s dz ds \\ \left(\frac{s}{\sigma} \right)^{n-1} \exp \left\{ -\frac{n}{2} \left(\frac{s}{\sigma} \right)^2 (1+z^2) \right\} d \left(\frac{s}{\sigma} \right) dz$$

which on integrating over all s (or (s/σ)) gives

$$P_n(dz|\lambda, \sigma) \propto (1+z^2)^{-n/2} dz$$

as first obtained (according to Jeffreys [1], from whom as usual much of this argument is taken) by Gosset, who wrote under the nom de plume of "Student" which has become firmly attached to this distribution. It is evidently not normal, although it tends to become so for large n (since then most of the probability is concentrated at small values of z , for which $(1+z^2)$ may be written approximately as $\exp(z^2)$). I shall make further use of the form (23c) as it stands, but it is commonly re-written by replacing s and z by $s_{\bar{x}}$ and t , where

$$s_{\bar{x}} \equiv \left\{ \frac{\sum_{r=1}^n (x_r - \bar{x})^2}{n(n-1)} \right\}^{1/2}$$

which is the conventional estimate of the standard error of a mean (including the divisor $(n-1)$ mentioned earlier in this section) and t is defined by

$$\bar{x} - \lambda = s_{\bar{x}} t \quad (24b)$$

(compare eqn (23b)), so that

$$s_{\bar{x}} = s/(n-1)^{1/2}$$

and

$$t = (n-1)^{1/2} z$$

then

$$P_n(dt|\lambda, \sigma) \propto \left[1 + \frac{t^2}{n-1} \right]^{-n/2} dt \quad (24e)$$

which is the "Student's t " distribution in its customary form.

I complete this section by deriving the normalising factor in eqn (23c); this takes a form which will be required later, which is my justification for doing it, but it also has an interesting consequence which is relevant here. Using eqn (23c), write

$$P_n(dz|\lambda, \sigma) = K_n (1+z^2)^{-n/2} dz \quad (25a)$$

and let $z = \tan \phi$ ($-\pi/2 \leq \phi \leq \pi/2$), $dz = \sec^2 \phi d\phi$; then

$$1 = \int_{-\infty}^{\infty} P_n(dz|\lambda, \sigma) = K_n \int_{-\infty}^{\infty} (1+z^2)^{-n/2} dz$$

$$K_n \int_{-\pi/2}^{\pi/2} (\cos \phi)^{n-2} d\phi$$

where n must be at least 2 for convergence (if $n=1$, $s=0$, so the analysis fails anyway).

This integral is a standard one, whose value is easily proved as follows. (I do this, rather than quoting a reference book, because it leads to an elegant proof of another result.) Write

$$\bar{I}_n \equiv \int_{-\pi/2}^{\pi/2} (\cos \phi)^n d\phi \quad (26a)$$

then, integrating by parts, and assuming temporarily that $n \geq 2$,

$$\begin{aligned} \bar{I}_n &= \int_{-\pi/2}^{\pi/2} (\cos \phi)^{n-1} \cos \phi d\phi \\ &= \left[(\cos \phi)^{n-1} \sin \phi \right]_{-\pi/2}^{\pi/2} + \int_{-\pi/2}^{\pi/2} (n-1) (\cos \phi)^{n-2} \sin^2 \phi d\phi \\ &= \int_{-\pi/2}^{\pi/2} (n-1) (\cos \phi)^{n-2} \sin^2 \phi d\phi \quad (\text{since } \cos \phi = 0 \text{ at both limits}) \\ &= (n-1) \int_{-\pi/2}^{\pi/2} (\cos \phi)^{n-2} (1 - \cos^2 \phi) d\phi \\ &= (n-1) \left\{ \int_{-\pi/2}^{\pi/2} (\cos \phi)^{n-2} d\phi - \int_{-\pi/2}^{\pi/2} (\cos \phi)^n d\phi \right\} \\ &= (n-1) \{ \bar{I}_{n-2} - \bar{I}_n \} \end{aligned}$$

so

$$\bar{I}_n = \left(\frac{n-1}{n} \right) \bar{I}_{n-2} \quad \text{for } n \geq 2 \quad (26c)$$

If n is even, write $n = 2n'$ ($n' \geq 1$); then

$$\begin{aligned} \bar{I}_{2n'} &= \left(\frac{n' - \frac{1}{2}}{n'} \right) \bar{I}_{2n'-2} \\ &= \left(\frac{(n' - \frac{1}{2})(n' - \frac{3}{2}) \dots \frac{1}{2}}{n'(n'-1) \dots 1} \right) \bar{I}_0 = \left(\frac{(n' - \frac{1}{2})! / (-\frac{1}{2})!}{n'!} \right) \bar{I}_0 \end{aligned}$$

$$\frac{(n' - \frac{1}{2})!}{n'!} \sqrt{\pi}$$

since $-\frac{1}{2}! = \Gamma(\frac{1}{2}) = \sqrt{\pi}$ and $\bar{I}_0 = \int_{-\pi/2}^{\pi/2} 1.d\phi = \pi$

If n is odd, write $n = 2n' + 1$ ($n' \geq 1$); then

$$\begin{aligned} \bar{I}_{2n'+1} &= \left[\frac{n'}{n' + \frac{1}{2}} \right] \bar{I}_{2n'-1} \\ &= \left[\frac{n'(n'-1)\dots\dots\dots 1}{(n' + \frac{1}{2})(n' - \frac{1}{2})\dots\dots\dots \frac{3}{2}} \right] \bar{I}_1 = \left[\frac{n'!}{(n' + \frac{1}{2})! / (\frac{1}{2})!} \right] \bar{I}_1 \\ &\quad \frac{n'!}{(n' + \frac{1}{2})!} \sqrt{\pi} \end{aligned}$$

since

$$(\frac{1}{2})! = \frac{1}{2} (-\frac{1}{2})! = \frac{1}{2}\sqrt{\pi} \quad \text{and} \quad \bar{I}_1 = \int_{-\pi/2}^{\pi/2} \cos \phi \, d\phi = 2.$$

On substituting $n' = \frac{1}{2}n$ into eqn (26d), and $n' = \frac{1}{2}(n-1)$ into eqn (26e), both these equations reduce to

$$\bar{I}_n = \frac{(\frac{1}{2}n - \frac{1}{2})!}{(\frac{1}{2}n)!} \sqrt{\pi} \quad (26f)$$

in which the restriction $n \geq 2$ may be discarded, since this equation also gives the correct values of $\bar{I}_0$ and $\bar{I}_1$ with the aid of the results quoted above; and, from eqns (25b) and (26a), the normalising factor K_n in eqn (25a) is $(\bar{I}_{n-2})^{-1}$. Combining eqns (26a) and (26f) now gives

$$\int_{-\pi/2}^{\pi/2} (\cos \phi)^n \, d\phi = \frac{(\frac{1}{2}n - \frac{1}{2})!}{(\frac{1}{2}n)!} \sqrt{\pi} \quad (27)$$

for all relevant (non-negative integral) values of n . This result will be needed in Part II of this paper. In the meantime, I consider the customary measure of a distribution's spread, viz its second moment or mean square deviation from the expectation. (The expectation itself is clearly λ , corresponding to $z=0$, since the distribution is symmetrical in z ; but n must be at least 3 for the formal integral expression for the expectation to be properly convergent at its infinite limits.)

From eqns (25a) and (23b), the mean square deviation in units of s^2 is

$$\begin{aligned}
 \int_{-\infty}^{\infty} z^2 P_n(dz|\lambda, \sigma) &= K_n \int_{-\infty}^{\infty} (1+z^2)^{-n/2} dz \cdot z^2 \\
 &= K_n \int_{-\infty}^{\infty} (1+z^2)(1+z^2)^{-n/2} dz - \int_{-\infty}^{\infty} (1+z^2)^{-n/2} dz \\
 &= K_n \left\{ K_{n-2}^{-1} - K_n^{-1} \right\} \quad \text{from eqn (25b)} \\
 &= \left\{ \bar{I}_{n-4} - \bar{I}_{n-2} \right\} / \bar{I}_{n-2} \quad \text{from eqns (25b) and (26a)} \\
 &= 1/(n-3) \quad \text{from eqn (26b)}
 \end{aligned}$$

and so, from eqns (18c) and (24a), the mean square deviation is

$$\frac{\sum_{r=1}^n (x_r - \bar{x})^2}{n(n-3)} = s_{\bar{x}}^2 \left[\frac{n-1}{n-3} \right] \quad (28)$$

with $n \geq 4$.

The conventional divisor $(n-1)$ is inserted in the definition of $s_{\bar{x}}$ to (supposedly) provide an unbiased estimate of the standard deviation of the underlying normal distribution of sample means (which is $(\sigma/\sqrt{n})$, from eqn (23a)). The appearance of the factor $((n-1)/(n-3))$ in eqn (28) shows that this procedure cannot be used blindly; its effectiveness depends on the use which is made of the result, particularly if n is small. This point will be considered again in the reprise of this section at the beginning of Part II, which now follows.

PART II

Introduction

In Part I of this paper I presented a review of established probability theory (mostly according to the treatment by Jeffreys [1]) with two aims: i) to introduce, and as far as possible to justify, the normal distribution and the normal correlation distribution; and ii) to introduce a derivation of the "Student's t" distribution, which bears the same relationship to my new distribution as the normal distribution bears to the normal correlation distribution, and is therefore a suitable vehicle for introducing most of the necessary prior probability forms. As illustrated particularly by the derivation of the "Student's t" distribution, all this previous analysis assumes that the probability distribution relevant to a problem, together with any parameters occurring in it, can be inferred or known from first principles (assisted by the great generality of the normal distribution and its correlated form); and then the expected results of measurement are predicted from the specified distribution. I turn now to the logical inverse of this procedure: to predict the parameters in a postulated general law, given the results of some measurements which are expected to conform to it.

The laws of distribution discussed in Part I, including the normal distribution and its correlated form, are essentially what Jeffreys [1] calls "likelihood" functions; that is, they express the probability that the observed measurement values should have occurred, given particular values of the parameters in the postulated law (or infinitesimal value ranges of them, to be precise) and any previous information. Using the multiplication rule for concurrent elementary probabilities, I can combine the "likelihood" of any particular measurement with the probability of particular parameter values due to all that measurement's predecessors; but to escape an infinite regress, one must have a starting point, representing the relative probability of different parameter values in the absence of any measurement information at all. The starting-point is known as the prior probability, and has been a source of much controversy among the philosophers of statistics, although it is tacitly accepted by everybody in its most elementary form: the "equally likely" assumption for binomial samples (eg from coin-tossing or dice-throwing, as discussed in Part I), which is essential to proving the binomial law of distribution. The chief problem is that the natural extension of the "equally likely" assumption to continuous variables, taking the prior probability of a parameter λ as being distributed proportionally to its corresponding differential $d\lambda$, is only satisfactory for "location" parameters (those which specify the "centre" of a distribution); it does not work for "scale" parameters (those which express a distribution's "spread"), nor can it be convincingly used directly for a correlation coefficient such as ρ in eqns (16). Jeffreys [1] treats this problem in two stages, first considering a standard deviation σ in isolation from other parameters and demonstrating that its prior probability is distributed proportionally to $(d\sigma/\sigma)$, and then embedding this result in a general theory of transformation invariance which can treat parameters such as the correlation coefficient ρ without resorting to heuristics (such as the classification into "location" and "scale" parameters). I shall review these results, with some minor extensions to obtain the joint prior probability distribution which I need, before deriving my new probability distribution and examining its consequences.

Sampling the normal distribution

Consider first the case of a sample drawn from measurements conforming to the normal law, with a known standard deviation σ and an unknown true value λ . Then the prior probability distribution of λ is proportional to $d\lambda$; and the appropriate joint "likelihood" for a sample of n measured values $x_1, x_2, \dots, x_n$ (strictly, values in infinitesimal ranges enclosing each of these) has been derived in Part I and is given by eqn (18e) with the parameter λ dropped on the left-hand side

$$P(dx_1 dx_2 \dots dx_n | \sigma, \theta) = \frac{1}{(2\pi)^{n/2} \sigma^n} \exp \left\{ -\frac{n[s^2 + (\bar{x} - \lambda)^2]}{2\sigma^2} \right\} dx_1 \dots dx_n$$

with the mean and standard deviation of the sample, $\bar{x}$ and s , defined by eqns (18b) and (18c), and θ introduced as an abbreviation for the sample data $x_1, x_2, \dots, x_n$. The "likelihood" and the prior probability are now combined by the multiplication rule, giving the so-called "posterior probability". In this, the differentials $dx_1 \dots dx_n$ serve merely to define a "posterior probability density" in the neighbourhood of a particular λ and θ for a given σ , $\bar{x}$ and s , and can be absorbed into the proportionality factor in the prior probability (alternatively, the implied integration over $x_1 \dots x_n$ can be regarded as being postponed indefinitely, and the differentials eventually cancel out since only relative probabilities have physical significance). Then, on absorbing also the constant factor $(2\pi)^{n/2}$,

$$P_n(d\lambda | \sigma, \theta) \propto \sigma^{-n} \exp \left\{ -\frac{ns^2}{2\sigma^2} \right\} \exp \left\{ -\frac{n(\bar{x} - \lambda)^2}{2\sigma^2} \right\} d\lambda$$

which, after normalisation to determine the proportionality constant, gives

$$P_n(d\lambda | \sigma, \theta) = \sqrt{\frac{n}{2\pi}} \frac{1}{\sigma} \exp \left\{ -\frac{n}{2\sigma^2} (\lambda - \bar{x})^2 \right\} d\lambda \quad (29c)$$

This is the exact analogue of eqn (23a) in Part I, but these two equations are logical inverses of each other: eqn (23a) expresses the distribution of predictions of a sample mean $\bar{x}$ in terms of a known λ (and σ), while eqn (29c) expresses the distribution of predictions of the true value λ in terms of a known $\bar{x}$ (and σ). In the last step I have reversed the sign of $(\bar{x} - \lambda)$, to emphasise this point! As in Part I, it is concluded that the best estimate of λ from the sample is its mean $\bar{x}$, and that the variability of samples is less than that of individual measurements by the factor $\sqrt{n}$ which divides σ . However, this analysis assumes that σ is known but λ is not, which is physically rather unlikely. I consider next the more usual case, where only the normal law is assumed and σ and λ are both unknown.

The "Student's t" distribution revisited

If the standard deviation σ and the true value λ are unknown, and are taken to have no direct dependence on each other (this point will be considered in a later section), then their joint prior probability is the product of the two distributions already discussed, which is $(d\lambda)(d\sigma/\sigma)$. The joint "likelihood" for a sample of n measured values is the same as before (eqn (29a)) except for dropping σ from the data on the left-hand side; hence eqn (29b) becomes

$$P_n(d\lambda d\sigma | \theta) \propto \sigma^{-n-1} \exp \left\{ -\frac{ns^2}{2\sigma^2} \right\} \exp \left\{ -\frac{n(\lambda - \bar{x})^2}{2\sigma^2} \right\} d\lambda d\sigma$$

$$P_n(d\lambda|\theta) \propto d\lambda \int_0^\infty \sigma^{-n-1} \exp \left\{ -\frac{n}{2\sigma^2} [s^2 + (\lambda - \bar{x})^2] \right\} d\sigma$$

On scaling σ using the dummy variable u defined by

$$u \equiv \frac{n}{2\sigma^2} [s^2 + (\lambda - \bar{x})^2],$$

eqn (30b) becomes

$$P_n(d\lambda|\theta) \propto d\lambda \left(\frac{2}{n} \right)^{n/2} [s^2 + (\lambda - \bar{x})^2]^{-n/2} \int_0^\infty u^{n/2} \exp(-u) \left(\frac{du}{u} \right)$$

where the numerical factor immediately after $(d\lambda)$ is included for formal consistency. The integral on the right-hand side is a number, independent of s , λ and $\bar{x}$; hence, using the transformation given by eqn (23b) and the results of eqns (23c), (25) and (27),

$$P_n(d\lambda|\theta) = \left[\frac{1}{\sqrt{\pi}} \right] \frac{\left(\frac{1}{2}n-1 \right)!}{\left(\frac{1}{2}n-\frac{3}{2} \right)!} s^{n-1} [s^2 + (\lambda - \bar{x})^2]^{-n/2} d\lambda \quad (30d)$$

This is identical in form to "Student's" original result (eqn (25a)) and has taken over its name (at least when written like eqn (24e)) because it is the practically-useful law (see Jeffreys [1] for a discussion of the mutual implications). If the prior probability distribution of σ is taken proportional to $(d\sigma)$ instead of $(d\sigma/\sigma)$, (n) in eqn (30d) must be replaced by $(n-1)$, as follows from inserting an extra factor (σ) in eqn (30a) and proceeding through the same steps. This apparently trivial change reduces the equation to nonsense, thereby providing an excellent argument against the "obvious" $(d\sigma)$ prior distribution! For instance, on making the substitution

$$\lambda - \bar{x} = s \tan \phi \quad (-\pi/2 \leq \phi \leq \pi/2)$$

(after eqns (23b) and (25)) with $n=2$, eqn (30d) gives

$$P_2(d\lambda|\theta) \rightarrow P(d\phi|\theta) = \frac{1}{\pi} d\phi$$

for two measurements, which is perfectly well-behaved, and leads to the interesting conclusion that there is probability $\frac{1}{2}$ that the true value lies between the two measurements (because, for $n=2$, s is the distance of either of the measurements from their mean (from eqn (18c)); and so (from eqn (30e)) setting λ equal to each measurement in turn gives $\phi = -\pi/4$ for the smaller measurement and $\phi = \pi/4$ for the larger one; which are the limits of integration in eqn (30f)). This is consistent with a simple-minded binomial view of the two-measurement experiment; it is as likely as not that a given measurement will be above the true value instead of below it (the necessary symmetry of the distribution is implied by the normal law on which the "likelihood" is based), so the probability that two measurements will bracket the true value (one above, one below) is the same as the probability of obtaining different results when tossing a coin twice. In contrast, replacing (n) by $(n-1)$ in eqn (30d) and then setting $n=2$ gives a factor in the denominator of $(-1)!$, which is infinite; and correspondingly the functional part of eqn (30d) reduces to $[s^2 + (\lambda - \bar{x})^2]^{-1/2} d\lambda$, which is not integrable. This result is also obtained by setting $n=1$ in eqn (30d), corresponding to the $(d\sigma/\sigma)$ distribution with a single measurement, but it is not then unreasonable since one

measurement clearly gives no information about a distribution's "spread" or "scale"; moreover, when $n=1$, $s=0$, so eqn (30c) simply reduces (for one measurement x_1 , with $\bar{x}=x_1$) to

$$P_1(d\lambda|x_1) \propto |\lambda-x_1|^{-1} d\lambda$$

which is similar enough to $(d\sigma/\sigma)$ to be acceptable, with the distribution symmetrical about the one measurement. However, on inserting an extra factor (σ) in eqn (30a) and then setting $n=1$, $s=0$, $\bar{x}=x_1$, it reduces to

$$P'_1(d\lambda d\sigma|x_1) \propto \sigma^{-1} \exp\left\{-\frac{(\lambda-x_1)^2}{2\sigma^2}\right\} d\lambda d\sigma$$

and on scaling σ using the dummy variable u defined by

$$u \equiv \frac{(\lambda-x_1)^2}{2\sigma^2}$$

this equation becomes, with integration over all σ (or u),

$$P'_1(d\lambda|x_1) \propto d\lambda \int_0^\infty \left(\frac{d\sigma}{\sigma}\right) \exp\left\{-\frac{(\lambda-x_1)^2}{2\sigma^2}\right\} = d\lambda \int_0^\infty \frac{1}{2} \left(\frac{du}{u}\right) \exp(-u)$$

The integral here is a pure number; it is divergent at its lower limit, but if it is regularised by truncating the upper range of σ , the upper limit of σ will have to be taken proportional to $(\lambda-x_1)$ – since there is nothing else to set its scale – and then the lower limit of u will be a pure number greater than zero. It follows that, when the prior probability of σ is assumed to be distributed proportional to $(d\sigma)$, and whether the integral in eqn (31b) is regularised or not,

$$P'_1(d\lambda|x_1) \propto d\lambda$$

which is plainly unacceptable, since it states that the "posterior probability" of λ after one measurement is distributed in the same way as the prior probability before any measurement, and therefore implies that one measurement contributes no information whatever. Accordingly, I take the prior distribution $(d\sigma/\sigma)$ as demonstrated, and proceed to the next stage.

Invariance theory and generalised prior probability distributions

Following Jeffreys [1], I now consider how to formulate a general prior probability distribution for laws containing parameters (such as eqn (16e) in Part I) so as to express the fundamental logical requirement that equivalent propositions must have the same probability: that is, that the prior probability associated with a particular region in the co-ordinate space of the parameters must be the same as the prior probability associated with the corresponding region after any physically-meaningful transformation of the parameters. (For example, in the normal correlation distribution expressed by eqn (16e), the parameters σ , τ and ρ can be transformed by rotating the axes of the variables u and v ; this will be discussed later). This formulation is made as follows. If there are two laws according to which the chances of a variable x being less than a given value are P and P' , then the integrals I_w and J are invariants, where w is a positive real parameter and

$$I_w = \int |(dP')^{1/w} - (dP)^{1/w}|^w$$

$$J = \int \log \left[\frac{dP'}{dP} \right] d(P' - P)$$

with the Stieltjes definition of integration (taking δP , $\delta P'$ for the same interval of x , forming the approximating sums and allowing all the x -intervals to tend to zero; see Jeffreys and Jeffreys [3]). To apply these to the problem in hand, set $p_r = \delta P_r$ and $p_r' = \delta P_r'$ for the x -interval δx_r , let p_r depend on a set of parameters α_i (i being an index from 1 to m) and be differentiable with respect to all the α_i , and let p_r' be the result of changing each α_i in p_r to a corresponding $(\alpha_i + \delta\alpha_i)$, where the $\delta\alpha_i$ are all small; then to the second order in small quantities (using the tensor summation convention on indices i and k), eqn (32b) gives

$$\begin{aligned} J &= \lim_{\delta x_r \rightarrow 0} \sum_r \left[\frac{1}{p_r} \right] \left[\frac{\partial p_r}{\partial \alpha_i} \delta\alpha_i \right] \left[\frac{\partial p_r}{\partial \alpha_k} \delta\alpha_k \right] \\ &= g_{ik} \delta\alpha_i \delta\alpha_k \end{aligned}$$

where

$$g_{ik} = \lim_{\delta x_r \rightarrow 0} \sum_r \left[\frac{1}{p_r} \right] \left[\frac{\partial p_r}{\partial \alpha_i} \right] \left[\frac{\partial p_r}{\partial \alpha_k} \right]$$

and similarly, to the second order in small quantities,

$$I_2 = \frac{1}{4} g_{ik} \delta\alpha_i \delta\alpha_k$$

Eqns (33) show that the invariants J and $(4I_2)$ have the form of the square of a distance element in general curvilinear co-ordinates (as in the Riemannian geometry of curved space applied to general relativity by Einstein and Levi-Civita, for example). If the parameters α_i are transformed to a corresponding set α_j' (with finite changes, not infinitesimal ones), the invariants J and $(4I_2)$ remain the same, so

$$g_{ik} \delta\alpha_i \delta\alpha_k = J = g'_{j\ell} \delta\alpha_j' \delta\alpha_\ell'$$

where

$$g'_{j\ell} = g_{ik} \frac{\partial \alpha_i}{\partial \alpha_j'} \frac{\partial \alpha_k}{\partial \alpha_\ell'}$$

using the tensor summation convention also on indices j and ℓ . Now, using the matrix multiplication rule and the theorem that the determinant of a square matrix product is equal to the product of the determinants of the separate square matrices, it follows from eqn (34b) that

$$\|g'_{j\ell}\| \quad \|g_{ik}\| \quad \left\| \frac{\partial \alpha_j}{\partial \alpha'_i} \right\| \quad \left\| \frac{\partial \alpha_k}{\partial \alpha'_\ell} \right\|$$

where the tensor summation convention is no longer required, and where the second and third determinants on the right-hand side are equal (since when they are evaluated by expansion they differ only in their dummy indices, which disappear when the terms of the expansions are summed). Each of these equal determinants will be recognised as the Jacobian determinant of the transformation of parameters from the set α_i to the set α'_j (apart possibly from a sign which cancels because the Jacobian appears twice), as shown (e.g.) in Jeffreys and Jeffreys [3]; and the transformation of a multiple integral over the α_i to one over the α'_j (expressed in standard form with all limits in both integrals increasing) is therefore (using the modulus of the Jacobian determinant)

$$\begin{aligned} d\alpha_1 d\alpha_2 \dots d\alpha_m & \left\| \frac{\partial \alpha_i}{\partial \alpha'_j} \right\| \cdot d\alpha'_1 d\alpha'_2 \dots d\alpha'_m \\ & \left[\left\| \frac{\partial \alpha_i}{\partial \alpha'_j} \right\| \right]^2 \right]^{1/2} \cdot d\alpha'_1 d\alpha'_2 \dots d\alpha'_m \\ & \left[\frac{\|g'_{j\ell}\|}{\|g_{ik}\|} \right]^{1/2} d\alpha'_1 d\alpha'_2 \dots d\alpha'_m \end{aligned} \quad (34d)$$

using eqn (34c) in the last step. It immediately follows from eqn (34d) that

$$\|g_{ik}\|^{1/2} d\alpha_1 d\alpha_2 \dots d\alpha_m \quad \|g'_{j\ell}\|^{1/2} d\alpha'_1 d\alpha'_2 \dots d\alpha'_m$$

and this invariant multiple differential form is a candidate for the prior probability of the parameter sets α_i or α'_j . The square roots are always real, because the original invariant integrals I_w and J are known to be positive-definite and so (from eqns (33a) and (33c)) the determinant of the g_{ik} is positive-definite also; it is assumed that the permitted transformations from the α_i to the α'_j exclude any singular forms. The results have been stated for only one variable x in the Stieltjes integrations, but they extend immediately to several variables (a fact which will be used later).

This argument does not, of course, prove that the candidate prior probability form given by eqn (35) is actually the correct one; however, according to Jeffreys [1], no other such invariant form is known, and the invariance is required by logical consistency (as remarked at the beginning of this section). I therefore follow Jeffreys in adopting the candidate form and verifying it by tests of consistency with previous work – particularly the "Student's t " distribution, used already as a reference standard.

Tests of the invariant prior probability distribution against previous cases

I consider first the normal distribution, for which I choose to evaluate the invariant I_2 (in most of the cases discussed by Jeffreys [1] he evaluates both J and $(4I_2)$ to verify their second-order agreement as stated above, but I refer the reader to his book for this). In the notation of the last section, with one-dimensional Stieltjes integration, I write (from eqn (12b))

$$p_r = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x_r-\lambda)^2}{2\sigma^2}\right\} dx_r$$

and similarly

$$p'_r = \frac{1}{\sigma'\sqrt{2\pi}} \exp\left\{-\frac{(x_r-\lambda')^2}{2\sigma'^2}\right\} dx_r$$

so from eqn (32a)

$$\begin{aligned} I_2 &= \int_{-\infty}^{\infty} [\sqrt{p'_r} - \sqrt{p_r}]^2 \\ &= \int_{-\infty}^{\infty} \frac{1}{\sigma'\sqrt{2\pi}} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x_r-\lambda')^2}{4\sigma'^2} - \frac{(x_r-\lambda)^2}{4\sigma^2}\right\} dx_r \\ &= 2 \int_{-\infty}^{\infty} \frac{2}{\sqrt{2\pi\sigma\sigma'}} \exp\left\{-\frac{(x_r-\lambda')^2}{4\sigma'^2} - \frac{(x_r-\lambda)^2}{4\sigma^2}\right\} dx_r \end{aligned}$$

since

$$\int_{-\infty}^{\infty} p'_r = \int_{-\infty}^{\infty} p_r = 1$$

Now

$$\begin{aligned} \frac{(x_r-\lambda')^2}{4\sigma'^2} + \frac{(x_r-\lambda)^2}{4\sigma^2} &= \frac{\{\sigma^2(x_r-\lambda')^2 + \sigma'^2(x_r-\lambda)^2\}}{4\sigma'^2\sigma^2} \\ &= \frac{\{(\sigma^2 + \sigma'^2)x_r^2 - 2(\lambda'\sigma^2 + \lambda\sigma'^2)x_r + (\lambda'^2\sigma^2 + \lambda^2\sigma'^2)\}}{4\sigma'^2\sigma^2} \\ &= \frac{(\sigma^2 + \sigma'^2)}{4\sigma'^2\sigma^2} \left[x_r - \frac{(\lambda'\sigma^2 + \lambda\sigma'^2)}{(\sigma^2 + \sigma'^2)} \right]^2 + \left[\frac{(\lambda'\sigma^2 + \lambda\sigma'^2)^2}{4\sigma'^2\sigma^2(\sigma^2 + \sigma'^2)} - \frac{(\lambda'^2\sigma^2 + \lambda^2\sigma'^2)}{4\sigma'^2\sigma^2} \right] \end{aligned}$$

and

$$\begin{aligned}
 & \left[\frac{(\lambda' \sigma^2 + \lambda \sigma'^2)^2}{4\sigma'^2 \sigma^2 (\sigma^2 + \sigma'^2)} - \frac{(\lambda'^2 \sigma^2 + \lambda^2 \sigma'^2)}{4\sigma'^2 \sigma^2} \right] = \frac{(\lambda' \sigma^2 + \lambda \sigma'^2)^2 - (\sigma^2 + \sigma'^2)(\lambda'^2 \sigma^2 + \lambda^2 \sigma'^2)}{4\sigma'^2 \sigma^2 (\sigma^2 + \sigma'^2)} \\
 & = \frac{\lambda'^2 \sigma^4 + 2\lambda' \lambda \sigma^2 \sigma'^2 + \lambda^2 \sigma'^4 - \lambda'^2 \sigma^4 - \lambda'^2 \sigma^2 \sigma'^2 - \lambda^2 \sigma^2 \sigma'^2 - \lambda^2 \sigma'^4}{4\sigma'^2 \sigma^2 (\sigma^2 + \sigma'^2)} \\
 & \frac{\sigma^2 \sigma'^2 (2\lambda' \lambda - \lambda'^2 - \lambda^2)}{4\sigma'^2 \sigma^2 (\sigma^2 + \sigma'^2)} = - \frac{(\lambda' - \lambda)^2}{4(\sigma^2 + \sigma'^2)} \quad (37c)
 \end{aligned}$$

Using eqns (37b) and (37c), the integral in eqn (37a) becomes

$$\begin{aligned}
 & \int_{-\infty}^{\infty} \frac{2}{\sqrt{2\pi\sigma\sigma'}} \exp \left\{ - \frac{(\lambda' - \lambda)^2}{4(\sigma^2 + \sigma'^2)} \exp \left\{ - \frac{(\sigma^2 + \sigma'^2)}{4\sigma^2 \sigma'^2} \left[x_r - \frac{(\lambda' \sigma^2 + \lambda \sigma'^2)}{(\sigma^2 + \sigma'^2)} \right]^2 \right\} \right\} dx_r \\
 & 2\sqrt{2} \left[\frac{\sigma\sigma'}{\sigma^2 + \sigma'^2} \right]^{1/2} \exp \left\{ - \frac{(\lambda' - \lambda)^2}{4(\sigma^2 + \sigma'^2)} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \left[\frac{\sigma^2 + \sigma'^2}{2\sigma^2 \sigma'^2} \right]^{1/2} \exp \left\{ - \frac{(\sigma^2 + \sigma'^2)}{4\sigma^2 \sigma'^2} x_r'^2 \right\} dx_r' \right\} \\
 & \text{(writing } x_r' = x_r - \frac{(\lambda' \sigma^2 + \lambda \sigma'^2)}{(\sigma^2 + \sigma'^2)}) \\
 & 2\sqrt{2} \left[\frac{\sigma'}{\sigma} + \frac{\sigma}{\sigma'} \right]^{-1/2} \exp \left\{ - \frac{(\lambda' - \lambda)^2}{4(\sigma^2 + \sigma'^2)} \right\} \quad (37d)
 \end{aligned}$$

since the last integral is simply the total probability (unity) of a normal distribution with standard deviation $(2\sigma^2 \sigma'^2 / (\sigma^2 + \sigma'^2))^{1/2}$ and mean zero (in Stieltjes variable x_r') or mean $(\lambda' \sigma^2 + \lambda \sigma'^2) / (\sigma^2 + \sigma'^2)$ (in Stieltjes variable x_r). Substituting eqn (37d) into (37a), and including a factor of 4 as suggested by eqn (33c), now gives

$$(4I_2) = 8 \left[1 - \sqrt{2} \left[\frac{\sigma'}{\sigma} + \frac{\sigma}{\sigma'} \right]^{-1/2} \exp \left\{ - \frac{(\lambda' - \lambda)^2}{4(\sigma^2 + \sigma'^2)} \right\} \right] \quad (37e)$$

To obtain the lowest order infinitesimal form, I write the natural forms

$$\sigma' = \sigma + d\sigma \quad (38a)$$

$$\lambda' = \lambda + d\lambda \quad (38b)$$

$$\frac{(\lambda' - \lambda)^2}{4(\sigma^2 + \sigma'^2)} = -\frac{(d\lambda)^2}{8\sigma^2} (1 + O(d\sigma/\sigma))$$

and

$$\exp\left\{-\frac{(\lambda' - \lambda)^2}{4(\sigma^2 + \sigma'^2)}\right\} = 1 - \frac{(d\lambda)^2}{8\sigma^2} + \text{smaller terms}$$

while

$$\frac{\sigma'}{\sigma} + \frac{\sigma}{\sigma'} = 1 + \frac{d\sigma}{\sigma} + \left(1 + \frac{d\sigma}{\sigma}\right)^{-1} = 2 + \frac{(d\sigma)^2}{\sigma^2} + \text{smaller terms}$$

so

$$\left(\frac{\sigma'}{\sigma} + \frac{\sigma}{\sigma'}\right)^{-1/2} = 2^{-1/2} \left(1 + \frac{(d\sigma)^2}{2\sigma^2}\right)^{-1/2} = 2^{-1/2} \left(1 - \frac{(d\sigma)^2}{4\sigma^2}\right) + \text{smaller terms} \quad (38d)$$

Hence the lowest order infinitesimal form of eqn (37e) becomes

$$(4I_2) = 8 \left[-\sqrt{2} \cdot 2^{-1/2} \left(1 - \frac{(d\sigma)^2}{4\sigma^2} \dots\right) \left(1 - \frac{(d\lambda)^2}{8\sigma^2} \dots\right) \right] \\ \left(\frac{d\lambda}{\sigma}\right)^2 + 2 \left(\frac{d\sigma}{\sigma}\right)^2$$

Consider first the cases of constant σ or constant λ . For these, the infinitesimal form in eqn (38e) reduces to $(d\lambda/\sigma)^2$ or $2(d\sigma/\sigma)^2$, and the matrix of elements g_{ik} in eqns (33)–(35) reduces to a single element $1/\sigma^2$ (apart from a numerical proportionality constant, which is eventually set by normalisation as previously). Hence in both cases $\|g_{ik}\|^{1/2} = 1/\sigma$, which is constant for constant σ ; and, from eqn (35), the candidate prior probability is proportional to $(d\lambda)$ for constant σ or $(d\sigma/\sigma)$ for constant λ . These results agree with the previous ones. However, if λ and σ vary concurrently, then the matrix of elements g_{ik} is of order (2×2) , with $g_{11} = g_{22} = 1/\sigma^2$, $g_{12} = g_{21} = 0$ (neglecting a constant factor as before), and then $\|g_{ik}\|^{1/2} = 1/\sigma^2$, so that the candidate prior probability from eqn (35) is proportional to $(d\lambda d\sigma/\sigma^2)$. If this distribution is adopted, (n) in eqn (30d) must be replaced by $(n+1)$, as follows from inserting an extra divisor (σ) in eqn (30a) and proceeding through the same steps; and, as shown by the discussion following eqn (30d), tampering with the exponent n is sufficient to make nonsense of this equation.

To resolve this difficulty, I follow the argument given by Jeffreys [1]. The candidate prior probability expressed in eqn (35) represents an invariant under a general non-singular transformation from the set of parameters α_i to the set α'_i ; but the logical properties of a particular problem may restrict this generality. In a one-dimensional estimation problem, where the aim is to predict one mean λ (or standard deviation σ), it is usually found that the unknown λ and σ are each capable of taking any independent values within ranges which are so wide that they can be treated as essentially infinite (or semi-infinite for σ), so neither of λ and σ gives any appreciable information about the other. It then follows, as discussed earlier, that the joint prior probability

for λ and σ should be taken (by the multiplication rule) to be the product of the prior probabilities for each of λ and σ considered alone; the judgement, on "physical" grounds, that λ and σ are essentially mutually irrelevant is sufficient to restrict the possible transformations from (λ, σ) to (λ', σ') so that the general rule is restricted into a special one. I consider now an extension of this judgement of irrelevance; since the difficulty only arises when λ varies concurrently with the other parameter, I examine the law of partial correlation, which is a generalisation of eqn (16e) to many variables. This equation, in the notation used from eqn (17) onwards, is

$$P(dudv|\rho, \sigma, \tau) \propto \exp \left\{ -\frac{1}{2} \left[\frac{u^2}{\sigma^2(1-\rho^2)} - \frac{2\rho uv}{\sigma\tau(1-\rho^2)} + \frac{v^2}{\tau^2(1-\rho^2)} \right] \right\} dudv$$

which generalises to

$$P(d\hat{x}_1 d\hat{x}_2 \dots d\hat{x}_n | \alpha_{ik}, \sigma_i) \propto \exp \left\{ -\frac{1}{2} \sum_i \sum_k \frac{\alpha_{ik}}{\sigma_i \sigma_k} \hat{x}_i \hat{x}_k \right\} \prod_j d\hat{x}_j$$

where the $\hat{x}_i$ are the measurable quantities (distinguished by the $\hat{}$ from the actual results of a set of measurements), the σ_i are parameters similar to the standard deviation σ (or τ), the α_{ik} are numerical coefficients similar to the correlation coefficient ρ (usually $\alpha_{ii} = 1$, $|\alpha_{ik}| \leq 1$ for $i \neq k$, so that the double sum in the exponential is positive-definite), and the implied parameters similar to the mean λ are fixed (at zero for convenience, as in eqns (13)–(16)). Then it is clear from similarity considerations that the invariant J , expressed as an n -dimensional integral over all the $\hat{x}_i$ and taken to the lowest order of infinitesimals, is quadratic in the $(d\sigma_i/\sigma_i)$, and that the corresponding $\|g_{ik}\|$ will be of the form $A\alpha_{ik}/(1/\sigma_i^2)$, where A is a factor depending only on the α_{ik} ; these conclusions will be verified later for the two-dimensional case with which this paper is chiefly concerned. The candidate prior probability is then proportional to $A^{1/2} \prod d\alpha_{ik} \prod (d\sigma_i/\sigma_i)$, in which form there are no troublesome extra divisors of σ_i . The mean or "location" parameters like λ can now be included in the prior probability form, using the judgement of essential irrelevance, by taking their individual prior probabilities proportional to $(d\lambda)$ (as before) and applying the multiplication rule. This complete prior probability is invariant under those transformations of the λ -parameters which amount merely to "rescaling" consequent upon transformations of the σ_i and α_{ik} , and those are the only transformations of the λ -parameters which I need to consider, as will be shown shortly.

General form of the invariant prior probability distribution in two dimensions

I consider now the distribution of normal correlation, for which I choose to evaluate the invariant J (since this distribution is the simplest special case of partial correlation, which has just been discussed in outline in terms of J). In the notation of the last two sections, with two-dimensional Stieltjes integration, I write (from eqn (16e))

$$p_r = \frac{1}{2\pi\sigma\tau(1-\rho^2)^{1/2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\frac{(u_r - \lambda)^2}{\sigma^2} - \frac{2\rho(u_r - \lambda)(v_r - \mu)}{\sigma\tau} + \frac{(v_r - \mu)^2}{\tau^2} \right] \right\} du_r dv_r \quad (40a)$$

where non-zero means λ and μ have been introduced for the variables u and v (respectively), by analogy with eqns (12b) and (17); in the notation used from eqn (17) onwards, the right-hand side of eqn (40a) without the r -suffixes is also $P(dudv|\lambda, \mu, \sigma, \tau, \rho)$, a result which will be used later. Similarly

$$p'_r = \frac{1}{2\pi\sigma'\tau'(1-\rho'^2)^{1/2}} \exp \left\{ -\frac{1}{2(1-\rho'^2)} \left[\frac{(u_r-\lambda')^2}{\sigma'^2} - \frac{2\rho'(u_r-\lambda')(v_r-\mu')}{\sigma'\tau'} + \frac{(v_r-\mu')^2}{\tau'^2} \right] \right\} du_r dv_r \quad (40b)$$

Hence, from eqn (32b), and remembering that $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p_r = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p'_r = 1$

$$\begin{aligned} J &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left\{ \log(p'_r) - \log(p_r) \right\} \{p'_r - p_r\} \\ &\quad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi} \left[-\log \left(\frac{\sigma'\tau'}{\sigma\tau} \right) - \frac{1}{2} \log \left(\frac{1-\rho'^2}{1-\rho^2} \right) + \right. \\ &\quad \left. + \frac{1}{2(1-\rho^2)} \left[\frac{(u_r-\lambda)^2}{\sigma^2} - \frac{2\rho(u_r-\lambda)(v_r-\mu)}{\sigma\tau} + \frac{(v_r-\mu)^2}{\tau^2} \right] \right. \\ &\quad \left. - \frac{1}{2(1-\rho'^2)} \left[\frac{(u_r-\lambda')^2}{\sigma'^2} - \frac{2\rho'(u_r-\lambda')(v_r-\mu')}{\sigma'\tau'} + \frac{(v_r-\mu')^2}{\tau'^2} \right] \right\} \times \{p'_r - p_r\} \\ &\quad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi} \left[\frac{1}{2(1-\rho^2)} \left[\frac{(u_r-\lambda)^2}{\sigma^2} - \frac{2\rho(u_r-\lambda)(v_r-\mu)}{\sigma\tau} + \frac{(v_r-\mu)^2}{\tau^2} \right] \right. \\ &\quad \left. - \frac{1}{2(1-\rho'^2)} \left[\frac{(u_r-\lambda')^2}{\sigma'^2} - \frac{2\rho'(u_r-\lambda')(v_r-\mu')}{\sigma'\tau'} + \frac{(v_r-\mu')^2}{\tau'^2} \right] \right\} \times \\ &\quad \times \frac{1}{\sigma'\tau'(1-\rho'^2)^{1/2}} \exp \left[-\frac{1}{2(1-\rho'^2)} \left[\frac{(u_r-\lambda')^2}{\sigma'^2} - \frac{2\rho'(u_r-\lambda')(v_r-\mu')}{\sigma'\tau'} + \frac{(v_r-\mu')^2}{\tau'^2} \right] \right] \\ &\quad \left. - \frac{1}{\sigma\tau(1-\rho^2)^{1/2}} \exp \left[-\frac{1}{2(1-\rho^2)} \left[\frac{(u_r-\lambda)^2}{\sigma^2} - \frac{2\rho(u_r-\lambda)(v_r-\mu)}{\sigma\tau} + \frac{(v_r-\mu)^2}{\tau^2} \right] \right] \right\} du_r dv_r \quad (41a) \end{aligned}$$

This integral is less formidable than it appears to be at first sight; it can be resolved into the sum of a small number of integrals of special forms, such that each of them can be evaluated from the integral of the distribution of normal correlation by differentiation under the integral signs. Some of these intermediate results are required later for another purpose, so I will derive them systematically.

I begin by verifying that $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p_r$ is indeed 1, using the alternative $P(\cdot)$ notation as mentioned after eqn (40a); in this notation (dropping the r -suffixes where they are not logically necessary)

$$\begin{aligned}
 P(dv|\lambda, \mu, \sigma, \tau, \rho) &= \int_{u=-\infty}^{u=\infty} P(dudv|\lambda, \mu, \sigma, \tau, \rho) \\
 &= \int_{u=-\infty}^{u=\infty} \frac{1}{2\pi\sigma\tau(1-\rho^2)^{1/2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\frac{(u-\lambda)^2}{\sigma^2} - \frac{2\rho(u-\lambda)(v-\mu)}{\sigma\tau} + \frac{(v-\mu)^2}{\tau^2} \right] \right\} dudv \\
 &= \int_{u=-\infty}^{u=\infty} \frac{1}{2\pi\sigma\tau(1-\rho^2)^{1/2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\left[\frac{(u-\lambda)}{\sigma} - \rho \frac{(v-\mu)}{\tau} \right]^2 + (1-\rho^2) \frac{(v-\mu)^2}{\tau^2} \right] \right\} dudv \\
 &= \int_{u=-\infty}^{u=\infty} \frac{1}{2\pi\tau} \exp \left[-\frac{(v-\mu)^2}{2\tau^2} \right] \exp \left[-\frac{1}{2} \left[\frac{u}{\sigma(1-\rho^2)^{1/2}} - \left[\frac{\lambda}{\sigma(1-\rho^2)^{1/2}} + \rho \frac{(v-\mu)}{\tau(1-\rho^2)^{1/2}} \right] \right]^2 \right] d \left[\frac{u}{\sigma(1-\rho^2)^{1/2}} \right] dv \\
 &= \frac{1}{\tau\sqrt{2\pi}} \exp \left[-\frac{(v-\mu)^2}{2\tau^2} \right] dv \tag{41b}
 \end{aligned}$$

so clearly

$$\begin{aligned}
 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(dudv|\lambda, \mu, \sigma, \tau, \rho) &= \int_{-\infty}^{\infty} P(dv|\lambda, \mu, \sigma, \tau, \rho) \\
 \int_{-\infty}^{\infty} \frac{1}{\tau\sqrt{2\pi}} \exp \left[-\frac{(v-\mu)^2}{2\tau^2} \right] dv &= 1 \tag{41c}
 \end{aligned}$$

as required. Eqn (41b) states that in circumstances of normal correlation the distribution of either variable without regard to the other is normal about its own mean with its own standard deviation, as indicated when the normal correlation distribution was derived (eqn (16e) and following text). For my purposes an equally important result is obtained by using the multiplication rule with eqns (41b) and (40a), thus (noting the intermediate forms of eqn (41b))

$$P(du|\lambda, \mu, \sigma, \tau, \rho, v) = \frac{P(dudv|\lambda, \mu, \sigma, \tau, \rho)}{P(dv|\lambda, \mu, \sigma, \tau, \rho)}$$

$$\frac{\frac{1}{2\pi\sigma\tau(1-\rho^2)^{1/2}} \exp \left[-\frac{1}{2(1-\rho^2)} \left\{ \left[\frac{(u-\lambda)}{\sigma} - \rho \frac{(v-\mu)}{\tau} \right]^2 + (1-\rho^2) \frac{(v-\mu)^2}{\tau^2} \right\} \right]}{\frac{1}{\tau\sqrt{2\pi}} \exp \left[-\frac{(v-\mu)^2}{2\tau^2} \right]} dudv$$

$$\frac{1}{\sigma(1-\rho^2)^{1/2}\sqrt{2\pi}} \exp \left[-\frac{1}{2\sigma^2(1-\rho^2)} \left[(u-\lambda) - \frac{\rho\sigma}{\tau} (v-\mu) \right]^2 \right] du$$

so the distribution of u for a given v is also normal, but the parameters are different from those in the distribution of u without regard to v ; from eqn (41b) and the comment following eqn (41c) (or by interchanging the parameters in eqn (41b)) the distribution of u without regard to v has mean λ and standard deviation σ , but the distribution of u for a given v has mean shifted from λ to $\left[\lambda + \frac{\rho\sigma}{\tau}(v-\mu) \right]$ and standard deviation contracted by the factor $(1-\rho^2)^{1/2}$. These results will be used later to help in understanding the restrictions on transformations of the " λ -parameters" (here λ and μ) as outlined in the last section.

In the meantime, I note from eqn (41c) and the intermediate forms of eqn (41b) that

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi} \exp \left[-\frac{1}{2(1-\rho^2)} \left[\frac{(u-\lambda)}{\sigma} - \rho \frac{(v-\mu)}{\tau} \right]^2 + (1-\rho^2) \frac{(v-\mu)^2}{\tau^2} \right] dudv$$

$$= \sigma\tau (1-\rho^2)^{1/2}$$

I denote the exponential factor in eqn (41e) by E ; then this equation is simply

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi} E dudv = \sigma\tau (1-\rho^2)^{1/2} \quad (41f)$$

Differentiating eqns (41e) or (41f) partially by σ , τ and ρ in turn gives

$$\begin{aligned}
& \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2(1-\rho^2)} \left[\frac{(u-\lambda)}{\sigma} - \rho \frac{(v-\mu)}{\tau} \right] \left[\frac{2(u-\lambda)}{\sigma} \right] \frac{E}{2\pi} \, du dv \\
& \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sigma \frac{\partial}{\partial \sigma} \left[\frac{E}{2\pi} \right] \, du dv = \sigma \frac{\partial}{\partial \sigma} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{E}{2\pi} \, du dv \\
& \sigma \frac{\partial}{\partial \sigma} \left\{ \sigma \tau (1-\rho^2)^{1/2} \right\} = \sigma \tau (1-\rho^2)^{1/2} ; \\
& \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2(1-\rho^2)} \left[\frac{(v-\mu)}{\tau} - \rho \frac{(u-\lambda)}{\sigma} \right] \left[\frac{2(v-\mu)}{\tau} \right] \frac{E}{2\pi} \, du dv \\
& \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \tau \frac{\partial}{\partial \tau} \left[\frac{E}{2\pi} \right] \, du dv = \tau \frac{\partial}{\partial \tau} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{E}{2\pi} \, du dv \\
& = \tau \frac{\partial}{\partial \tau} \left\{ \sigma \tau (1-\rho^2)^{1/2} \right\} = \sigma \tau (1-\rho^2)^{1/2} \tag{41h}
\end{aligned}$$

using the form for E analogous to that in eqn (41e) but with (u and v), (σ and τ) and (λ and μ) interchanged; and, using the symmetrical form for E as in eqns (40a) and (41a),

$$\begin{aligned}
& \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\rho}{(1-\rho^2)} \left[\frac{(u-\lambda)^2}{\sigma^2} - \frac{2\rho(u-\lambda)(v-\mu)}{\sigma\tau} + \frac{(v-\mu)^2}{\tau^2} \right] + \frac{(u-\lambda)(v-\mu)}{\sigma\tau} \frac{E}{2\pi} \, du dv \\
& \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (1-\rho^2) \frac{\partial}{\partial \rho} \left[\frac{E}{2\pi} \right] \, du dv = (1-\rho^2) \frac{\partial}{\partial \rho} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{E}{2\pi} \, du dv \\
& (1-\rho^2) \frac{\partial}{\partial \rho} \left\{ \sigma \tau (1-\rho^2)^{1/2} \right\} = -\rho \sigma \tau (1-\rho^2)^{1/2}
\end{aligned}$$

Since the limits of integration are constant and the integrands vanish exponentially at them, the differentiations under the integral signs are valid provided the integrands are everywhere finite and continuous; this is true provided $\sigma \neq 0$, $\tau \neq 0$ and $\rho \neq \pm 1$ (or $(1-\rho^2) \neq 0$), all of which are necessary to make the normal correlation distribution be physically realistic and mathematically sensible. Now, I add together eqns (41g) and (41h), multiply the sum by ρ , and add the result to eqn (41i); then, on taking the factor ρ inside the integral signs and combining the resulting three integrals into one, I obtain

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\frac{(u-\lambda)(v-\mu)}{\sigma\tau} \right] \frac{E}{2\pi} du dv = \rho\sigma\tau(1-\rho^2)^{1/2}$$

Consider now the final form of J in eqn (41a). When the factors in the outer pairs of brace brackets are multiplied out, four terms appear; two of these contain the exponential factor E , while the other two can be obtained from the first two by interchanging the primed and unprimed parameters everywhere (including inside the factor E). Hence, on separating the double integral over the four terms into the sum of four double integrals over each term individually, it immediately follows that

$$J = \bar{J} + \bar{J}'$$

where $\bar{J}'$ is obtained from $\bar{J}$ by interchanging the primed and unprimed parameters everywhere, and (on dropping the unnecessary r suffixes in eqn (41a))

$$\begin{aligned} \bar{J} = & - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\sigma\tau(1-\rho^2)^{3/2}} \left[\frac{(u-\lambda)^2}{\sigma^2} - \frac{2\rho(u-\lambda)(v-\mu)}{\sigma\tau} + \frac{(v-\mu)^2}{\tau^2} \right] \frac{E}{2\pi} du dv \\ & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\sigma\tau(1-\rho^2)^{1/2}(1-\rho'^2)} \left[\frac{(u-\lambda')^2}{\sigma'^2} - \frac{2\rho'(u-\lambda')(v-\mu')}{\sigma'\tau'} + \frac{(v-\mu')^2}{\tau'^2} \right] \frac{E}{2\pi} du dv \end{aligned}$$

The first of these integrals is trivial to evaluate; adding together eqns (41g) and (41h) and combining their integral left-hand sides gives (as for eqn (41j))

$$\begin{aligned} & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{(1-\rho^2)} \left[\frac{(u-\lambda)^2}{\sigma^2} - \frac{2\rho(u-\lambda)(v-\mu)}{\sigma\tau} + \frac{(v-\mu)^2}{\tau^2} \right] \frac{E}{2\pi} du dv \\ & 2\sigma\tau(1-\rho^2)^{1/2} \end{aligned} \quad (42c)$$

and the first integral in eqn (42b) is now obtained by dividing eqn (42c) by $(2\sigma\tau(1-\rho^2)^{1/2})$, showing that its value is 1. Hence, on expanding the square-bracketed factor in the second integral in eqn (42b) (from here on, integrals will be scaled, separated or combined without special comment),

$$\begin{aligned} \bar{J} = & -1 + (2\sigma\sigma'^2\tau(1-\rho^2)^{1/2}(1-\rho'^2))^{-1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (u-\lambda')^2 \left[\frac{E}{2\pi} \right] du dv \\ & - (\sigma\sigma'\tau\tau'(1-\rho^2)^{1/2}(1-\rho'^2))^{-1} \rho' \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (u-\lambda')(v-\mu') \left[\frac{E}{2\pi} \right] du dv \\ & + (2\sigma\tau\tau'^2(1-\rho^2)^{1/2}(1-\rho'^2))^{-1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (v-\mu')^2 \left[\frac{E}{2\pi} \right] du dv \end{aligned}$$

Two more special integrals are required to evaluate the integral terms in this equation. First, I multiply eqn (41g) by $(\sigma^2(1-\rho^2))$ and eqn (41j) by $(\rho\sigma^2)$, and add the results; then

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (u-\lambda)^2 \left[\frac{E}{2\pi} \right] du dv = \sigma^3 \tau (1-\rho^2)^{1/2}$$

Second, by analogy, or by multiplying eqn (41h) by $(\tau^2(1-\rho^2))$ and eqn (41j) by $(\rho\tau^2)$ and adding the results,

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (v-\mu)^2 \left[\frac{E}{2\pi} \right] du dv = \sigma \tau^3 (1-\rho^2)^{1/2}$$

Now, consider the third integral in eqn (42d); from eqn (42f) this becomes

$$\begin{aligned} & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (v-\mu')^2 \left[\frac{E}{2\pi} \right] du dv \\ & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (v-\mu)^2 \left[\frac{E}{2\pi} \right] du dv + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\mu-\mu')(2v-\mu-\mu') \left[\frac{E}{2\pi} \right] du dv \\ & \sigma \tau^3 (1-\rho^2)^{1/2} + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\mu-\mu')(2v-\mu-\mu') \left[\frac{E}{2\pi} \right] du dv \\ & \sigma \tau^3 (1-\rho^2)^{1/2} + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\mu-\mu')^2 \left[\frac{E}{2\pi} \right] du dv \\ & + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} 2(\mu-\mu')(v-\mu) \left[\frac{E}{2\pi} \right] du dv \\ & \sigma \tau^3 (1-\rho^2)^{1/2} + \sigma \tau (1-\rho^2)^{1/2} (\mu-\mu')^2 + 2(\mu-\mu') \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (v-\mu) \left[\frac{E}{2\pi} \right] du dv \end{aligned}$$

using eqn (41f) at the last stage. The last integral here vanishes because of its internal antisymmetry; for substituting for E from eqn (41e) gives

$$\begin{aligned} & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (v-\mu) \left[\frac{E}{2\pi} \right] du dv \\ & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (v-\mu) \left[\frac{1}{2\pi} \right] \exp \left[-\frac{1}{2(1-\rho^2)} \left[\frac{(u-\lambda)}{\sigma} - \rho \frac{(v-\mu)}{\tau} \right]^2 + (1-\rho^2) \frac{(v-\mu)^2}{\tau^2} \right] du dv \end{aligned}$$

$$\frac{\sigma(1-\rho^2)^{1/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (v-\mu) \exp \left[-\frac{(v-\mu)^2}{2\tau^2} \right] dv \quad (42h)$$

using the last three lines of eqn (41b) at the last stage, and this integral is clearly zero because of its antisymmetry about the point $v=\mu$ (or by changing the integration variable from v to $((v-\mu)^2)$ and performing the integration with a break at $v=\mu$). By analogy, using the symmetrical form for E as in eqns (40a) and (41a) and then interchanging (u and v), (σ and τ) and (λ and μ),

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (u-\lambda) \left[\frac{E}{2\pi} \right] dudv = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (v-\mu) \left[\frac{E}{2\pi} \right] dudv = 0 \quad (42i)$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (v-\mu')^2 \left[\frac{E}{2\pi} \right] dudv = \sigma\tau(1-\rho^2)^{1/2} (\tau^2 + (\mu-\mu')^2)$$

(from eqns (42g) and (42h)), and

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (u-\lambda')^2 \left[\frac{E}{2\pi} \right] dudv = \sigma\tau(1-\rho^2)^{1/2} (\sigma^2 + (\lambda-\lambda')^2)$$

The last two of these are the first and third integrals in eqn (42d). It only remains to use eqns (41j) and (42i) to evaluate the second integral in eqn (42d); thus

$$\begin{aligned} & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (u-\lambda')(v-\mu') \left[\frac{E}{2\pi} \right] dudv \\ & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (u-\lambda)(v-\mu) \left[\frac{E}{2\pi} \right] dudv \\ & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (-\lambda'v - \mu'u + \lambda'\mu' + \lambda v + \mu u - \lambda\mu) \left[\frac{E}{2\pi} \right] dudv \\ & = \rho\sigma^2\tau^2(1-\rho^2)^{1/2} \\ & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [(\lambda-\lambda')(v-\mu) + \mu(\lambda-\lambda') + (\mu-\mu')(u-\lambda) + \lambda(\mu-\mu') + \lambda'\mu' - \lambda\mu] \left[\frac{E}{2\pi} \right] dudv \\ & = \rho\sigma^2\tau^2(1-\rho^2)^{1/2} + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [\mu(\lambda-\lambda') + \lambda(\mu-\mu') + \lambda'\mu' - \lambda\mu] \left[\frac{E}{2\pi} \right] dudv \\ & = \rho\sigma^2\tau^2(1-\rho^2)^{1/2} + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\lambda-\lambda')(\mu-\mu') \left[\frac{E}{2\pi} \right] dudv \\ & = \sigma\tau(1-\rho^2)^{1/2} (\rho\sigma\tau + (\lambda-\lambda')(\mu-\mu')) \quad (42\ell) \end{aligned}$$

using eqn (41f) at the last stage. Finally, I substitute eqns (42j), (42k) and (42ℓ) into eqn (42d), giving

$$\begin{aligned}\bar{J} = & -1 + (2\sigma'^2(1-\rho'^2))^{-1} (\sigma^2 + (\lambda-\lambda')^2) \\ & - (\sigma'\tau'(1-\rho'^2))^{-1} \rho'(\rho\sigma\tau + (\lambda-\lambda')(\mu-\mu')) \\ & + (2\tau'^2(1-\rho'^2))^{-1} (\tau^2 + (\mu-\mu')^2)\end{aligned}\quad (42m)$$

and so, from eqn (42a) and the text following it,

$$\begin{aligned}J = & -2 + \frac{((\sigma/\sigma')^2 + (\tau/\tau')^2 - 2\rho\rho'(\sigma/\sigma')(\tau/\tau'))}{2(1-\rho'^2)} \\ & + \frac{((\sigma'/\sigma)^2 + (\tau'/\tau)^2 - 2\rho\rho'(\sigma'/\sigma)(\tau'/\tau))}{2(1-\rho^2)} \\ & + (\lambda-\lambda')^2 ((2\sigma'^2(1-\rho'^2))^{-1} + (2\sigma^2(1-\rho^2))^{-1}) \\ & + (\mu-\mu')^2 ((2\tau'^2(1-\rho'^2))^{-1} + (2\tau^2(1-\rho^2))^{-1}) \\ & + (\lambda-\lambda')(\mu-\mu') (\rho'(\sigma'\tau'(1-\rho'^2))^{-1} + \rho(\sigma\tau(1-\rho^2))^{-1})\end{aligned}\quad (43a)$$

To obtain the lowest-order infinitesimal form of this expression for J , I recall the natural forms (eqns (38a) and (38b)) used for the normal distribution

$$\sigma' = \sigma + d\sigma \quad (43b)$$

$$\lambda' = \lambda + d\lambda$$

and their obvious analogues

$$\tau' = \tau + d\tau$$

$$\mu' = \mu + d\mu$$

while for ρ I follow Jeffreys [1] and introduce Fisher's transformation, writing

$$\rho = \tanh \zeta$$

with the natural counterparts

$$\rho' = \tanh \zeta'$$

and

$$\zeta' = \zeta + d\zeta$$

although this transformation will not be applied everywhere. Then, since

$$(1-\rho^2)^{-1} = \cosh^2 \zeta$$

(with the obvious extension to primed parameters), and

$$(\sigma/\sigma')^2 = \left[1 + \frac{d\sigma}{\sigma}\right]^{-2} = 1 - 2\left[\frac{d\sigma}{\sigma}\right] + 3\left[\frac{d\sigma}{\sigma}\right]^2 + \text{smaller terms}$$

with various similar results, it follows that (to second-order infinitesimals)

$$\begin{aligned} J &= -2 + \frac{1}{2} \cosh^2 \zeta' \left[1 - 2\left[\frac{d\sigma}{\sigma}\right] + 3\left[\frac{d\sigma}{\sigma}\right]^2 + 1 - 2\left[\frac{d\tau}{\tau}\right] + 3\left[\frac{d\tau}{\tau}\right]^2 \right. \\ &\quad \left. - 2 \tanh \zeta \tanh \zeta' \left[1 - \left[\frac{d\sigma}{\sigma}\right] + \left[\frac{d\sigma}{\sigma}\right]^2 \right] \left[1 - \left[\frac{d\tau}{\tau}\right] + \left[\frac{d\tau}{\tau}\right]^2 \right] \right] \\ &\quad + \frac{1}{2} \cosh^2 \zeta \left[1 + 2\left[\frac{d\sigma}{\sigma}\right] + \left[\frac{d\sigma}{\sigma}\right]^2 + 1 + 2\left[\frac{d\tau}{\tau}\right] + \left[\frac{d\tau}{\tau}\right]^2 \right. \\ &\quad \left. - 2 \tanh \zeta \tanh \zeta' \left[1 + \frac{d\sigma}{\sigma} \right] \left[1 + \frac{d\tau}{\tau} \right] \right] \\ &\quad + (\sigma^2(1-\rho^2))^{-1} (d\lambda)^2 + (\tau^2(1-\rho^2))^{-1} (d\mu)^2 - 2\rho(\sigma\tau(1-\rho^2))^{-1} (d\lambda)(d\mu) \\ &\equiv J_{\sigma\tau\zeta} + J_{\lambda\mu} \end{aligned}$$

where

$$J_{\lambda\mu} = (\sigma^2(1-\rho^2))^{-1} (d\lambda)^2 + (\tau^2(1-\rho^2))^{-1} (d\mu)^2 - 2\rho(\sigma\tau(1-\rho^2))^{-1} (d\lambda)(d\mu)$$

and $J_{\sigma\tau\zeta}$ is the rest of J in eqn (44a). Then, from eqn (43h),

$$\begin{aligned} J_{\sigma\tau\zeta} &= -2 + \frac{1}{2} \cosh^2 \zeta' \left[1 - 2\left[\frac{d\sigma}{\sigma}\right] + 3\left[\frac{d\sigma}{\sigma}\right]^2 + 1 - 2\left[\frac{d\tau}{\tau}\right] + 3\left[\frac{d\tau}{\tau}\right]^2 \right. \\ &\quad \left. - 2 \tanh^2 \zeta \left[1 + \operatorname{sech} \zeta \operatorname{csch} \zeta (d\zeta) - \operatorname{sech}^2 \zeta (d\zeta)^2 \right] \left[1 - \left[\frac{d\sigma}{\sigma}\right] + \left[\frac{d\sigma}{\sigma}\right]^2 \right] \left[1 - \left[\frac{d\tau}{\tau}\right] + \left[\frac{d\tau}{\tau}\right]^2 \right] \right] \\ &\quad + \frac{1}{2} \cosh^2 \zeta \left[1 + 2\left[\frac{d\sigma}{\sigma}\right] + \left[\frac{d\sigma}{\sigma}\right]^2 + 1 + 2\left[\frac{d\tau}{\tau}\right] + \left[\frac{d\tau}{\tau}\right]^2 \right. \\ &\quad \left. - 2 \tanh^2 \zeta \left[1 + \operatorname{sech} \zeta \operatorname{csch} \zeta (d\zeta) - \operatorname{sech}^2 \zeta (d\zeta)^2 \right] \left[1 + \left[\frac{d\sigma}{\sigma}\right] \right] \left[1 + \left[\frac{d\tau}{\tau}\right] \right] \right] \end{aligned}$$

$$\begin{aligned}
&= -2 + \cosh^2 \zeta' \left[1 - \left(\frac{d\sigma}{\sigma} \right) + \frac{3}{2} \left(\frac{d\sigma}{\sigma} \right)^2 - \left(\frac{d\tau}{\tau} \right) + \frac{3}{2} \left(\frac{d\tau}{\tau} \right)^2 \right. \\
&\quad - \tanh^2 \zeta \left[1 + \operatorname{sech} \zeta \operatorname{csch} \zeta (d\zeta) - \operatorname{sech}^2 \zeta (d\zeta)^2 \right. \\
&\quad \quad \left(\frac{d\sigma}{\sigma} \right) - \operatorname{sech} \zeta \operatorname{csch} \zeta (d\zeta) \left(\frac{d\sigma}{\sigma} \right) + \left(\frac{d\sigma}{\sigma} \right)^2 \\
&\quad \quad - \left(\frac{d\tau}{\tau} \right) - \operatorname{sech} \zeta \operatorname{csch} \zeta (d\zeta) \left(\frac{d\tau}{\tau} \right) + \left(\frac{d\tau}{\tau} \right)^2 \\
&\quad \quad \left. \left. + \left(\frac{d\sigma}{\sigma} \right) \left(\frac{d\tau}{\tau} \right) \right] \right] \\
&+ \cosh^2 \zeta \left[1 + \left(\frac{d\sigma}{\sigma} \right) + \frac{1}{2} \left(\frac{d\sigma}{\sigma} \right)^2 + \left(\frac{d\tau}{\tau} \right) + \frac{1}{2} \left(\frac{d\tau}{\tau} \right)^2 \right. \\
&\quad - \tanh^2 \zeta \left[1 + \operatorname{sech} \zeta \operatorname{csch} \zeta (d\zeta) - \operatorname{sech}^2 \zeta (d\zeta)^2 \right. \\
&\quad \quad + \left(\frac{d\sigma}{\sigma} \right) + \operatorname{sech} \zeta \operatorname{csch} \zeta (d\zeta) \left(\frac{d\sigma}{\sigma} \right) \\
&\quad \quad + \left(\frac{d\tau}{\tau} \right) + \operatorname{sech} \zeta \operatorname{csch} \zeta (d\zeta) \left(\frac{d\tau}{\tau} \right) \\
&\quad \quad \left. \left. + \left(\frac{d\sigma}{\sigma} \right) \left(\frac{d\tau}{\tau} \right) \right] \right] \tag{44c}
\end{aligned}$$

neglecting infinitesimals of higher than second order (as always). In this, I write

$$\cosh^2 \zeta' \equiv \cosh^2 \zeta + \{ \cosh^2 \zeta' - \cosh^2 \zeta \}$$

where the expression in braces is a first-order infinitesimal (so that second-order infinitesimals can be disregarded when it is multiplied into the round-bracketed co-factor of $(\cosh^2 \zeta')$ in eqn (44c)). Then

$$\begin{aligned}
J_{\sigma\tau\zeta} &= -2 + \{ \cosh^2 \zeta' - \cosh^2 \zeta \} \times \\
&\quad \times \left[\operatorname{sech}^2 \zeta \left[1 - \left(\frac{d\sigma}{\sigma} \right) - \left(\frac{d\tau}{\tau} \right) - \tanh \zeta (d\zeta) \right] \right] \\
&+ \cosh^2 \zeta \left[1 - \left(\frac{d\sigma}{\sigma} \right) + \frac{3}{2} \left(\frac{d\sigma}{\sigma} \right)^2 - \left(\frac{d\tau}{\tau} \right) + \frac{3}{2} \left(\frac{d\tau}{\tau} \right)^2 \right. \\
&\quad \left. + 1 + \left(\frac{d\sigma}{\sigma} \right) + \frac{1}{2} \left(\frac{d\sigma}{\sigma} \right)^2 + \left(\frac{d\tau}{\tau} \right) + \frac{1}{2} \left(\frac{d\tau}{\tau} \right)^2 \right]
\end{aligned}$$

$$\begin{aligned}
& \tanh^2 \zeta \left[1 + \operatorname{sech} \zeta \operatorname{csch} \zeta (d\zeta) - \operatorname{sech}^2 \zeta (d\zeta)^2 \right. \\
& \quad - \left(\frac{d\sigma}{\sigma} \right) - \operatorname{sech} \zeta \operatorname{csch} \zeta (d\zeta) \left(\frac{d\sigma}{\sigma} \right) + \left(\frac{d\sigma}{\sigma} \right)^2 \\
& \quad \left(\frac{d\tau}{\tau} \right) - \operatorname{sech} \zeta \operatorname{csch} \zeta (d\zeta) \left(\frac{d\tau}{\tau} \right) + \left(\frac{d\tau}{\tau} \right)^2 \\
& \quad + \left(\frac{d\sigma}{\sigma} \right) \left(\frac{d\tau}{\tau} \right) \\
& \quad + \operatorname{sech} \zeta \operatorname{csch} \zeta (d\zeta) - \operatorname{sech}^2 \zeta (d\zeta)^2 \\
& \quad + \left(\frac{d\sigma}{\sigma} \right) + \operatorname{sech} \zeta \operatorname{csch} \zeta (d\zeta) \left(\frac{d\sigma}{\sigma} \right) \\
& \quad + \left(\frac{d\tau}{\tau} \right) + \operatorname{sech} \zeta \operatorname{csch} \zeta (d\zeta) \left(\frac{d\tau}{\tau} \right) \\
& \quad \left. + \left(\frac{d\sigma}{\sigma} \right) \left(\frac{d\tau}{\tau} \right) \right] \\
& - 2 + \left\{ \cosh^2 \zeta' - \cosh^2 \zeta \right\} \times \\
& \times \left[\operatorname{sech}^2 \zeta \left(1 - \left(\frac{d\sigma}{\sigma} \right) - \left(\frac{d\tau}{\tau} \right) - \tanh \zeta (d\zeta) \right) \right] \\
& + \cosh^2 \zeta \left[2 + 2 \left(\frac{d\sigma}{\sigma} \right)^2 + 2 \left(\frac{d\tau}{\tau} \right)^2 \right. \\
& \quad - \tanh^2 \zeta \left[2 + 2 \operatorname{sech} \zeta \operatorname{csch} \zeta (d\zeta) - 2 \operatorname{sech}^2 \zeta (d\zeta)^2 \right. \\
& \quad \left. \left. + \left(\frac{d\sigma}{\sigma} \right)^2 + \left(\frac{d\tau}{\tau} \right)^2 + 2 \left(\frac{d\sigma}{\sigma} \right) \left(\frac{d\tau}{\tau} \right) \right] \right] \\
& \left\{ \cosh^2 \zeta' - \cosh^2 \zeta \right\} \left[\operatorname{sech}^2 \zeta \left(1 - \left(\frac{d\sigma}{\sigma} \right) - \left(\frac{d\tau}{\tau} \right) - \tanh \zeta (d\zeta) \right) \right] \\
& + (2 \cosh^2 \zeta - \sinh^2 \zeta) \left[\left(\frac{d\sigma}{\sigma} \right)^2 + \left(\frac{d\tau}{\tau} \right)^2 \right] \\
& - 2 \tanh \zeta (d\zeta) + 2 \tanh^2 \zeta (d\zeta)^2 - 2 \sinh^2 \zeta \left(\frac{d\sigma}{\sigma} \right) \left(\frac{d\tau}{\tau} \right) \tag{44d}
\end{aligned}$$

and finally, on expanding the expression in braces to second-order as

$$\begin{aligned}
 \left\{ \cosh^2 \zeta' - \cosh^2 \zeta \right\} &\equiv 2 \cosh \zeta \sinh \zeta (d\zeta) + (\cosh^2 \zeta + \sinh^2 \zeta) (d\zeta)^2, \\
 J_{\sigma\tau\zeta} &= \left(2 \tanh \zeta (d\zeta) + (1 + \tanh^2 \zeta) (d\zeta)^2 \right) \cosh^2 \zeta \operatorname{sech}^2 \zeta \left[1 - \left(\frac{d\sigma}{\sigma} \right) - \left(\frac{d\tau}{\tau} \right) - \tanh \zeta (d\zeta) \right] \\
 &\quad + \left(2 \cosh^2 \zeta - \sinh^2 \zeta \right) \left[\left(\frac{d\sigma}{\sigma} \right)^2 + \left(\frac{d\tau}{\tau} \right)^2 \right] \\
 &\quad - 2 \tanh \zeta (d\zeta) + 2 \tanh^2 \zeta (d\zeta)^2 - 2 \sinh^2 \zeta \left(\frac{d\sigma}{\sigma} \right) \left(\frac{d\tau}{\tau} \right) \\
 &= 2 \tanh \zeta (d\zeta) + (1 + \tanh^2 \zeta) (d\zeta)^2 \\
 &\quad - 2 \tanh \zeta \left(\frac{d\sigma}{\sigma} \right) (d\zeta) - 2 \tanh \zeta \left(\frac{d\tau}{\tau} \right) (d\zeta) - 2 \tanh^2 \zeta (d\zeta)^2 \\
 &\quad + \left(2 \cosh^2 \zeta - \sinh^2 \zeta \right) \left[\left(\frac{d\sigma}{\sigma} \right)^2 + \left(\frac{d\tau}{\tau} \right)^2 \right] \\
 &\quad - 2 \tanh \zeta (d\zeta) + 2 \tanh^2 \zeta (d\zeta)^2 - 2 \sinh^2 \zeta \left(\frac{d\sigma}{\sigma} \right) \left(\frac{d\tau}{\tau} \right) \\
 &= \left(2 \cosh^2 \zeta - \sinh^2 \zeta \right) \left(\frac{d\sigma}{\sigma} \right)^2 + \left(2 \cosh^2 \zeta - \sinh^2 \zeta \right) \left(\frac{d\tau}{\tau} \right)^2 \\
 &\quad + (1 + \tanh^2 \zeta) (d\zeta)^2 - 2 \sinh^2 \zeta \left(\frac{d\sigma}{\sigma} \right) \left(\frac{d\tau}{\tau} \right) \\
 &\quad - 2 \tanh \zeta \left(\frac{d\sigma}{\sigma} \right) (d\zeta) - 2 \tanh \zeta \left(\frac{d\tau}{\tau} \right) (d\zeta) \tag{44e}
 \end{aligned}$$

Reduction of the invariant prior probability distribution from general to special form

I now proceed to determine the matrix of elements g_{ik} (as defined in eqns (33)–(35)) corresponding to the invariant J defined by eqns (44a), (44b) and (44e). Since J has separated completely into two infinitesimal forms, one involving only $d\sigma$, $d\tau$ and $d\zeta$, the other involving only $d\lambda$ and $d\mu$, with no cross-products, it follows that the corresponding matrix of g_{ik} has zeros in the elements corresponding to products of $(d\sigma, d\tau \text{ or } d\zeta)$ with $(d\lambda \text{ or } d\mu)$; so the matrix is quasi-diagonal, and its determinant (required in eqn (35) for the candidate prior probability form) is simply the product of the determinants of the two sub-matrices of g_{ik} corresponding to $J_{\sigma\tau\zeta}$ and $J_{\lambda\mu}$.

Consider first the sub-matrix of g_{ik} corresponding to $J_{\sigma\tau\zeta}$. Remarking that $g_{ik} = g_{ki}$ (from eqn (33b)), and identifying the various elements from eqns (33a) and (44e), it follows that

$$\|g_{ik}\| = \begin{vmatrix} (1+\tanh^2\zeta) & (-\sigma^{-1}\tanh\zeta) & (-\tau^{-1}\tanh\zeta) \\ (-\sigma^{-1}\tanh\zeta) & \sigma^{-2}(2\cosh^2\zeta-\sinh^2\zeta) & (-\sigma^{-1}\tau^{-1}\sinh^2\zeta) \\ (-\tau^{-1}\tanh\zeta) & (-\sigma^{-1}\tau^{-1}\sinh^2\zeta) & \tau^{-2}(2\cosh^2\zeta-\sinh^2\zeta) \end{vmatrix} \quad (45a)$$

$[d\zeta] \qquad [d\sigma] \qquad [d\tau]$

where the bottom line ($[d\zeta]$, etc) denotes the order of the columns of the matrix (the order of the rows is naturally the same by symmetry). The first term of this determinant's conventional expansion along its top row is

$$\begin{aligned} & (1+\tanh^2\zeta) \left\{ \sigma^{-2} (2\cosh^2\zeta-\sinh^2\zeta) \tau^{-2} (2\cosh^2\zeta-\sinh^2\zeta) \right. \\ & \quad \left. - (-\sigma^{-1}\tau^{-1}\sinh^2\zeta)^2 \right\} \\ & = (1+\tanh^2\zeta) \left\{ \sigma^{-2}\tau^{-2} \left[(2\cosh^2\zeta-\sinh^2\zeta)^2 - (\sinh^2\zeta)^2 \right] \right\} \\ & \quad \sigma^{-2}\tau^{-2} (1+\tanh^2\zeta) [2\cosh^2\zeta] [2\cosh^2\zeta-2\sinh^2\zeta] \end{aligned}$$

(on factorising the "difference of two squares" in the square brackets)

$$\begin{aligned} & = \sigma^{-2}\tau^{-2} (1+\tanh^2\zeta) [4\cosh^2\zeta] [\cosh^2\zeta-\sinh^2\zeta] \\ & = 4\sigma^{-2}\tau^{-2}\cosh^2\zeta (1+\tanh^2\zeta) \end{aligned} \quad (45b)$$

since $(\cosh^2\zeta - \sinh^2\zeta) \equiv 1$ (compare eqn (43i)). The third term of the expansion of the determinant in eqn (45a) is

$$\begin{aligned} & (-\tau^{-1}\tanh\zeta) \left\{ (-\sigma^{-1}\tanh\zeta) (-\sigma^{-1}\tau^{-1}\sinh^2\zeta) \right. \\ & \quad \left. - (-\tau^{-1}\tanh\zeta) \left[\sigma^{-2} (2\cosh^2\zeta-\sinh^2\zeta) \right] \right\} \\ & \quad -\sigma^{-2}\tau^{-2}\tanh\zeta \left\{ (\tanh\zeta) [\sinh^2\zeta + (2\cosh^2\zeta-\sinh^2\zeta)] \right\} \end{aligned} \quad (45c)$$

while the second term of the expansion, including the required extra minus sign, is

$$\begin{aligned} & -(-\sigma^{-1}\tanh\zeta) \left\{ (-\sigma^{-1}\tanh\zeta) \left[\tau^{-2}(2\cosh^2\zeta-\sinh^2\zeta) \right] \right. \\ & \quad \left. (-\tau^{-1}\tanh\zeta) (-\sigma^{-1}\tau^{-1}\sinh^2\zeta) \right\} \end{aligned}$$

$$- \sigma^{-2} \tau^{-2} \tanh \zeta \left\{ (\tanh \zeta) \left[(2 \cosh^2 \zeta - \sinh^2 \zeta) + \sinh^2 \zeta \right] \right\}$$

which is the same as expression (45c); hence the algebraic sum of the second and third terms is twice their common value, or

$$\begin{aligned} & 2 \left\{ -\sigma^{-2} \tau^{-2} \tanh \zeta \right\} \left\{ (\tanh \zeta) \left[\sinh^2 \zeta + (2 \cosh^2 \zeta - \sinh^2 \zeta) \right] \right\} \\ &= -2 \sigma^{-2} \tau^{-2} \tanh^2 \zeta \left[2 \cosh^2 \zeta \right] \\ & 4 \sigma^{-2} \tau^{-2} \cosh^2 \zeta \left[-\tanh^2 \zeta \right] \end{aligned} \quad (45d)$$

So the algebraic sum of the three terms of the determinant $\|g_{ik}\|$ in eqn (45a) is the sum of expressions (45b) and (45d), or

$$\begin{aligned} \|g_{ik}\| &= 4 \sigma^{-2} \tau^{-2} \cosh^2 \zeta \left\{ (1 + \tanh^2 \zeta) + [-\tanh^2 \zeta] \right\} \\ & 4 \sigma^{-2} \tau^{-2} \cosh^2 \zeta \end{aligned}$$

and the factor contributed by $J_{\sigma\tau\zeta}$ to the candidate prior probability form (eqn (35)) is

$$\|g_{ik}\|^{1/2} d\sigma d\tau d\zeta = 2 \left(\frac{d\sigma}{\sigma} \right) \left(\frac{d\tau}{\tau} \right) \cosh \zeta d\zeta$$

I recall now the forms of the invariant J and the candidate prior probability relating to the law of partial correlation (of which eqns (40) and following are a special case), as described in eqns (39) and the text following them. It was asserted in that description that J would be quadratic in the various $(d\sigma_i/\sigma_i)$ – here $(d\sigma/\sigma)$ and $(d\tau/\tau)$ – and that the corresponding candidate prior probability would be proportional to $A^{1/2} \prod d\alpha_{ik} \prod (d\sigma_i/\sigma_i)$ with A depending only on the α_{ik} ; clearly these conclusions are verified, the only α_{ik} here being ζ transformed from ρ . As also expected, the prior probability form (46) contains no troublesome extra divisors of σ or τ . To maintain this desirable state, it is now necessary to justify taking the mean or "location" parameters – here λ and μ – as essentially irrelevant to σ , τ and ρ , and to each other.

I begin the justification by determining what extra factors would appear in the candidate prior probability form (eqns (35) or (46)) in the general case. From $J_{\lambda\mu}$ (eqn (44b)) and the form of J (eqn (44a)), the determinant of the matrix of g_{ik} would include a factor of the determinant of the sub-matrix of g_{ik} corresponding to $J_{\lambda\mu}$, which is (recalling that $g_{ik} = g_{ki}$)

$$\begin{aligned} \|g_{ik}\| & \begin{vmatrix} (\sigma^2(1-\rho^2))^{-1} & -\rho(\sigma\tau(1-\rho^2))^{-1} \\ -\rho(\sigma\tau(1-\rho^2))^{-1} & (\tau^2(1-\rho^2))^{-1} \end{vmatrix} \\ & \begin{vmatrix} [d\lambda] & [d\mu] \end{vmatrix} \end{aligned} \quad (47a)$$

where the bottom line ($[d\lambda]$, etc) denotes the matrix's column order as in eqn (45a). This determinant can be evaluated at sight and is

$$\|g_{ik}\| = (\sigma^2\tau^2(1-\rho^2))^{-1}$$

so the factor which $J_{\lambda\mu}$ would contribute to the candidate prior probability form is

$$\|g_{ik}\|^{1/2} d\lambda d\mu = (\sigma\tau(1-\rho^2)^{1/2})^{-1} d\lambda d\mu$$

In the analysis of the normal distribution, which has one variable, an extra factor of σ^{-1} appears in the prior probability form (at eqn (38e) and in the following text). In the present case, where there are two variables, factors of σ^{-1} and τ^{-1} would be expected in the absence of correlation; and these would be the "rescaling" factors of the parameters λ and μ , as described in the text following eqns (39). In the presence of correlation, however, the distribution of one variable for a definite value of the other is both contracted ("rescaled") and shifted, as shown by eqn (41d) and the discussion surrounding it, and it must therefore be expected that one of the standard deviations σ and τ (or their product) will be contracted by the same factor; so the "rescaling" factor is changed from $\sigma^{-1}\tau^{-1}$ (which is $(\sigma\tau)^{-1}$) to $(\sigma\tau(1-\rho^2)^{1/2})^{-1}$, as in eqn (47c).

This argument is plausible, but not entirely convincing. In particular, the reasoning from eqn (41d) gives no indication why one should shift and contract only λ and σ , leaving μ and τ alone – or, if (u and v) and (σ and τ) and (λ and μ) are interchanged in the derivation of eqn (41d), why one should shift and contract only μ and τ , leaving λ and σ alone; these changes are both unsymmetrical, and some kind of symmetrical transformation would be more natural. To meet this objection, I re-examine the definition of J . From eqn (33a),

$$g_{ik} d\alpha_i d\alpha_k \quad (48a)$$

(using the tensor summation convention, and discarding the redundant distinction between small quantities $\delta\alpha_i$ and infinitesimals $d\alpha_i$); then, in the notation of eqns (47),

$$J = (d\lambda d\mu) (G_{ik}) (d\lambda d\mu)^T \quad (48b)$$

where G_{ik} is the matrix of the g_{ik} in eqn (47a) (the matrix itself, not its determinant) and the superscript T represents a matrix or vector transposition. The matrix G_{ik} can be reduced to diagonal form by a single "rotation" matrix R (see Jeffreys and Jeffreys [3]; the reason for the name will appear shortly); thus

$$R \equiv \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} \quad (48c)$$

where ϕ is chosen so that the matrix product $R^T G_{ik} R$ is diagonal. The explicit condition for this is obtained by actually multiplying out the product, giving

$$\begin{bmatrix} g'_{11} & g'_{12} \\ g'_{21} & g'_{22} \end{bmatrix} \equiv (G'_{ik}) \equiv \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}$$

$$\begin{bmatrix} g_{11}\cos^2\phi + g_{22}\sin^2\phi + g_X\sin 2\phi & g_X\cos 2\phi + \frac{1}{2}(g_{22}-g_{11})\sin 2\phi \\ g_X\cos 2\phi + \frac{1}{2}(g_{22}-g_{11})\sin 2\phi & g_{11}\sin^2\phi + g_{22}\cos^2\phi - g_X\sin 2\phi \end{bmatrix}$$

where g_X is the common value of g_{12} and g_{21} . By inspection, this symmetry is preserved, and the common value of g_{12}' and g_{21}' can be reduced to zero by choosing any ϕ such that (as described in Wilkinson [4])

$$\tan 2\phi = 2g_X/(g_{11}-g_{22})$$

This result will reappear later, when I have shown that my new distribution predicts that the curves of equal probability density derived from it (represented on a plane with λ and μ as rectangular co-ordinates) are a family of similar and concentric ellipses, inclined to the axes of λ and μ by one of the possible angles ϕ . In the meantime, I write the result $g_{12}' = g_{21}' = 0$ into the identity part of eqn (48d), giving

$$\begin{bmatrix} g_{11}' & 0 \\ 0 & g_{22}' \end{bmatrix} = R^T G_{ik} R$$

or

$$R \begin{bmatrix} g_{11}' & 0 \\ 0 & g_{22}' \end{bmatrix} R^T = (R^T)^{-1} \begin{bmatrix} g_{11}' & 0 \\ 0 & g_{22}' \end{bmatrix} R^{-1}$$

(by inspection of eqn (48c))

$$(R^T)^{-1} (R^T G_{ik} R) R^{-1} = G_{ik} \quad (48g)$$

from eqn (48f); so, from eqn (48b),

$$J = (d\lambda \ d\mu) R \begin{bmatrix} g_{11}' & 0 \\ 0 & g_{22}' \end{bmatrix} R^T (d\lambda \ d\mu)^T \quad (48h)$$

I write into this the definition

$$(d\lambda' \ d\mu') \equiv (d\lambda \ d\mu) R \quad (48i)$$

which gives

$$J = (d\lambda' \ d\mu') \begin{bmatrix} g_{11}' & 0 \\ 0 & g_{22}' \end{bmatrix} (d\lambda' \ d\mu')^T$$

$$= g'_{11}(d\lambda')^2 + g'_{22}(d\mu')^2$$

and, using eqn (48c),

$$d\lambda' = (\cos\phi) d\lambda + (\sin\phi) d\mu \quad (48k)$$

$$d\mu' = -(\sin\phi) d\lambda + (\cos\phi) d\mu$$

The last two results say that $d\lambda'$ and $d\mu'$ are obtained from $d\lambda$ and $d\mu$ by rotating the axes of λ and μ in the standard (anti-clockwise) direction through angle ϕ (which is why R is called a "rotation" matrix). In terms of these new parameters, eqn (48j) says that J reduces to the sum of two independent squared terms, as in eqn (38e), and the discussion following that equation shows that the appropriate candidate prior probability form is $\|g_{ik}'\|^{1/2} d\lambda' d\mu'$ or $(g'_{11} g'_{22})^{1/2} d\lambda' d\mu'$. From eqn (48c), the determinants of R and R^T are both unity; hence, from eqn (48f), $\|g_{ik}'\|^{1/2} = \|g_{ik}\|^{1/2}$, and so, from eqn (47c), the determinant factor which is to be discarded can be justified directly without apportioning the "rescaling" in it artificially between λ and μ (or even λ' and μ'). In fact, there are an infinite number of ways of transforming λ and μ to give forms consisting of independent squared terms (see Jeffreys and Jeffreys [3]), including more general transformations which genuinely change the scales of the parameters instead of merely rotating them to change their orientation (the transformation implied by eqn (41d) is of this kind); however, it is always found that the determinant factor is left unchanged if the transformations are physically consistent.

Two further points will be made before I leave this discussion. Firstly, from the discussion in the last paragraph taken with eqn (47c), it follows both that $\|g_{ik}'\|^{1/2} = \|g_{ik}\|^{1/2}$ and that the prior probability may be represented by either $\|g_{ik}'\|^{1/2} d\lambda' d\mu'$ or $\|g_{ik}\|^{1/2} d\lambda d\mu$. For these to be consistent, there must be a sense in which $(d\lambda' d\mu')$ is the same thing as $(d\lambda d\mu)$, at least for normalised integrations over prior probability such as those used in the second derivation of the "Student's t " distribution. In general terms, this result follows from the discussion surrounding eqns (34); the Jacobian of the transformation from $(d\lambda$ and $d\mu)$ (or $(\lambda$ and $\mu)$) to $(d\lambda'$ and $d\mu')$ (or $(\lambda'$ and $\mu')$) is the determinant of R (strictly, of R^T), which is 1. For two dimensions, a specific illustration is provided by using the vector notation which is commonly employed to state Stokes' theorem (e.g. in [3]); in those terms, $d\lambda$ and $d\mu$ are (infinitesimal) orthogonal vectors in a plane, and each element of the integrand is an (infinitesimal) vector in the third dimension, along the positive normal to the plane of integration. It is then simply required that the vector cross-products $(d\lambda \times d\mu)$ and $(d\lambda' \times d\mu')$ shall be equal, which follows immediately from eqns (48k) and (48l). Secondly, when $(\lambda$ and $\mu)$ are transformed to $(\lambda'$ and $\mu')$, there is a corresponding transformation from $(\sigma, \tau$ and $\rho)$ to $(\sigma', \tau'$ and $\rho')$. The elements g_{ik}' will be the same functions of $(\sigma', \tau'$ and $\rho')$ that the g_{ik} are of $(\sigma, \tau$ and $\rho)$; comparing g_{12} or g_{21} with $g'_{12} = g'_{21} = 0$ gives $\rho' = 0$ as the representation of the rotation R through the angle ϕ , and similarly the other elements give (from eqn (47a))

$$\sigma'^{-2} = g'_{11} \quad (49a)$$

$$\tau'^{-2} = g'_{22} \quad (49b)$$

which make sense when eqns (48j) and (38e) are compared. To obtain the two-dimensional version of the second term of eqn (38e), I attach primes to σ and τ in eqn (44e), and combine $\rho' = 0$ with eqn (43g) to give $\tanh \zeta' = 0$, $\sinh \zeta' = 0$, $\cosh \zeta' = 1$, $d\zeta' = 0$ (since ρ' is fixed); then eqn (44e) reduces to

$$J_{\sigma', \tau', \zeta'} = 2 \left[\frac{d\sigma'}{\sigma'} \right]^2 + 2 \left[\frac{d\tau'}{\tau'} \right]^2$$

which completes the extension of eqn (38e) to two dimensions. It follows that $(\lambda'$ and $\mu')$ can be treated as essentially irrelevant to $(\sigma'$ and $\tau')$, using the argument following eqn (38e); then, fixing ρ' at zero provides the necessary "restriction" (in the sense used in that argument) on the transformations of $\lambda, \mu, \sigma, \tau$ and ρ to λ', μ', σ' and τ' , but otherwise $(\lambda$ and $\mu)$ may be treated as essentially irrelevant to $(\sigma, \tau$ and $\rho)$ in the previous sense of the word.

By all these arguments, then, I have justified discarding the determinant factor arising out of $J_{\lambda\mu}$, which is $(\sigma\tau(1-\rho^2)^{1/2})^{-1}$ (as in eqn (47c)), and simply taking the prior probability distribution for λ and μ as proportional to $(d\lambda)$ and $(d\mu)$ individually, or (by the multiplication rule) as proportional to $(d\lambda d\mu)$ for both parameters together; and so (by the multiplication rule again) I take the prior probability for $\lambda, \mu, \sigma, \tau$ and ρ concurrently as the product of the prior probability for λ and μ considered alone (without σ, τ and ρ) and the prior probability for σ, τ and ρ considered alone (without λ and μ). This latter prior probability is the one arising out of $J_{\sigma\tau\zeta}$, given in eqn (46); hence, on absorbing a factor of 2 into the proportionality constant, I can now write

$$\text{prior probability } (\lambda, \mu, \sigma, \tau, \rho) \propto d\lambda d\mu \left[\frac{d\sigma}{\sigma} \right] \left[\frac{d\tau}{\tau} \right] \cosh \zeta d\zeta$$

I complete this section by transforming ζ back to ρ to give the final form of the prior probability which my new distribution requires. From eqn (43f),

$$\frac{d\rho}{d\zeta} = \frac{d}{d\zeta} \left[\frac{\sinh \zeta}{\cosh \zeta} \right] = (\cosh \zeta)^{-2} \quad (50b)$$

and so

$$\begin{aligned} \cosh \zeta d\zeta &= \cosh \zeta \left[\frac{d\rho}{d\zeta} \right]^{-1} d\rho = \cosh^3 \zeta d\rho \\ &= (\cosh^2 \zeta)^{3/2} d\rho \\ &= (1-\rho^2)^{-3/2} d\rho \end{aligned}$$

from eqn (43i), since $(\cosh \zeta)$ is necessarily positive. Hence

$$\text{prior probability } (\lambda, \mu, \sigma, \tau, \rho) \propto d\lambda d\mu \left[\frac{d\sigma}{\sigma} \right] \left[\frac{d\tau}{\tau} \right] (1-\rho^2)^{-3/2} d\rho \quad (50d)$$

This result is given by Jeffreys [1], on whose analysis I have based Part II of this paper; but he appears unconvinced by his own mathematical logic, which is why I have reviewed the derivation at such length. I have also discussed in much more detail the relationships between $(\sigma, \tau$ and $\rho)$ and $(\sigma', \tau'$ and zero- $\rho')$ to justify the most difficult step in the argument, which is the discarding of the determinant factor arising out of $J_{\lambda\mu}$. I now proceed to apply the derived prior probability form, expression (50d).

The posterior probability distribution for normal correlation sampling

Consider now the case of a sample drawn from measurements which are assumed to conform to the distribution of normal correlation in its most general form (see eqn (40a) and the text following it)

$$P(du dv | \lambda, \mu, \sigma, \tau, \rho) = \frac{1}{2\pi\sigma\tau(1-\rho^2)^{1/2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\frac{(u-\lambda)^2}{\sigma^2} - \frac{2\rho(u-\lambda)(v-\mu)}{\sigma\tau} + \frac{(v-\mu)^2}{\tau^2} \right] \right\} du dv \quad (51a)$$

where the parameters $\lambda, \mu, \sigma, \tau$ and ρ are unknown. Their joint prior probability is given by the expression (50d) just derived (at great length!). The appropriate joint "likelihood" of obtaining n measured value-pairs $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ – strictly, values in infinitesimal ranges enclosing each of them – is given from eqn (51a), by extension of eqns (18a), (18e) and (29a), as

$$P(dx_1 dx_2 \dots dx_n dy_1 dy_2 \dots dy_n | \theta) = \frac{1}{(2\pi)^n (\sigma\tau)^n (1-\rho^2)^{n/2}} dx_1 dx_2 \dots dx_n dy_1 \dots dy_n \times \\ \times \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\frac{\sum_{i=1}^n (x_i - \lambda)^2}{\sigma^2} - \frac{2\rho \sum_{i=1}^n (x_i - \lambda)(y_i - \mu)}{\sigma\tau} + \frac{\sum_{i=1}^n (y_i - \mu)^2}{\tau^2} \right] \right\}$$

which becomes, using the definitions of the sample parameters

$$\bar{x} \equiv \frac{\sum_{i=1}^n x_i}{n}$$

$$s \equiv \left[\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} \right]^{1/2} \quad (52b)$$

$$\bar{y} \equiv \frac{\sum_{i=1}^n y_i}{n}$$

$$t \equiv \left[\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n} \right]^{1/2} \quad (52d)$$

$$r \equiv \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n\sigma\tau}$$

with $(-1 \leq r \leq 1)$ by applying the triangle inequality to these definitions (compare eqns (16) and (18), and note the divisors of n instead of $(n-1)$),

$$\begin{aligned} P(dx_1 \dots dx_n dy_1 \dots dy_n | \theta) &= \frac{1}{(2\pi)^n (\sigma\tau)^n (1-\rho^2)^{n/2}} dx_1 \dots dx_n dy_1 \dots dy_n \times \\ &\times \exp \left\{ \frac{1}{2(1-\rho^2)} \left[\frac{n[s^2 + (\bar{x}-\lambda)^2]}{\sigma^2} + \frac{n[t^2 + (\bar{y}-\mu)^2]}{\tau^2} \right. \right. \\ &\quad \left. \left. - \frac{2\rho n[r\sigma\tau + (\bar{x}-\lambda)(\bar{y}-\mu)]}{\sigma\tau} \right] \right\} \end{aligned}$$

(In this step, eqn (18d) gives directly for x and analogously for y)

$$\sum_{i=1}^n (x_i - \lambda)^2 = n[s^2 + (\bar{x}-\lambda)^2] \quad \text{and} \quad \sum_{i=1}^n (y_i - \mu)^2 = n[t^2 + (\bar{y}-\mu)^2]$$

and by the method of eqn (18d)

$$\begin{aligned} \sum_{i=1}^n (x_i - \lambda)(y_i - \mu) &= \sum_{i=1}^n \left\{ (x_i - \bar{x}) + (\bar{x} - \lambda) \right\} \left\{ (y_i - \bar{y}) + (\bar{y} - \mu) \right\} \\ &= n[r\sigma\tau + (\bar{x}-\lambda)(\bar{y}-\mu)] \end{aligned}$$

using the definitions in eqns (52).) Now this joint "likelihood" is combined with the joint prior probability by the multiplication rule to give the joint "posterior probability", and (as in the text following eqn (29a)) the numerical factor $(2\pi)^n$ and the differential factors $dx_1 \dots dx_n dy_1 \dots dy_n$ may be absorbed into the proportionality factor in the prior probability. The joint "posterior probability" of the unknown parameters $\lambda, \mu, \sigma, \tau$ and ρ , written in the notation of eqns (29b) and (30), is then (from expressions (50d) and (53a))

$$\begin{aligned} P_n(d\lambda d\mu d\sigma d\tau d\rho | \theta) &\propto d\lambda d\mu d\sigma d\tau d\rho (\sigma\tau)^{-n-1} (1-\rho^2)^{(-3-n)/2} \times \\ &\times \exp \left[\frac{n}{2(1-\rho^2)} \left[\frac{(\lambda - \bar{x})^2}{\sigma^2} + \frac{s^2}{\sigma^2} + \frac{(\mu - \bar{y})^2}{\tau^2} + \frac{t^2}{\tau^2} - \frac{2\rho}{\sigma\tau} ((\lambda - \bar{x})(\mu - \bar{y}) + r\sigma\tau) \right] \right] \end{aligned} \quad (54)$$

As for eqns (23) and (30) in the descriptions of the "Student's t" distribution, I now obtain the joint probability distribution of λ and μ by integrating expression (54) over all possible values of σ , τ and ρ , and normalising the result. As this is an exercise in pure mathematics and does not depend on probability logic, I shall perform it and examine the consequences of its result in Part III (which now follows); the analysis of probability is now essentially complete. From here on, the material presented is original, apart from the documented mathematical results used in the derivations.

PART III**Fisher's substitution; the first integration**

I begin the integration of expression (54) by introducing a substitution given by Jeffreys [1] as due to Fisher (used by him to study correlation distributions); in this, the variables of integration σ and τ are replaced by α and β with the following defining equations

$$\frac{st}{\sigma\tau} = \alpha$$

$$\frac{s\tau}{\sigma t} = \exp(\beta)$$

from which follow immediately

$$\frac{s^2}{\sigma^2} = \alpha \exp(\beta)$$

$$\frac{t^2}{\tau^2} = \alpha \exp(-\beta)$$

$$\frac{\partial(\sigma, \tau)}{\partial(\alpha, \beta)} = \frac{st}{2\alpha^2}$$

the last of these equations giving the Jacobian determinant of the two-dimensional substitution. On absorbing various factors of s and t into the proportionality factor in eqn (54), I can write (including the appropriate changes to the limits of integration)

$$\begin{aligned} P_n(d\lambda d\mu d\rho | \theta) &= \int_0^\infty d\sigma \int_0^\infty d\tau \left[\frac{P_n(d\lambda d\mu d\rho d\sigma d\tau | \theta)}{d\sigma d\tau} \right] \\ &= \int_0^\infty d\alpha \int_{-\infty}^\infty d\beta \frac{\partial(\sigma, \tau)}{\partial(\alpha, \beta)} \left[\frac{P_n(d\lambda d\mu d\sigma d\tau d\rho | \theta)}{d\sigma d\tau} \right] \\ &\propto \int_{-\infty}^\infty d\beta d\lambda d\mu d\rho (1-\rho^2)^{(-3-n)/2} \times \\ &\times \int_0^\infty d\alpha \exp \left\{ \frac{n\alpha}{(1-\rho^2)} \left[\frac{(\lambda-\bar{x})^2}{2s^2} \exp(\beta) + \frac{(\mu-\bar{y})^2}{2t^2} \exp(-\beta) + \right. \right. \\ &\quad \left. \left. + \cosh\beta - \frac{\rho}{st} (\lambda-\bar{x})(\mu-\bar{y}) - \rho r \right] \right\} \alpha^n. \end{aligned}$$

$$\begin{aligned}
& \propto \int_{-\infty}^{\infty} d\beta \, d\lambda d\mu d\rho \, (1-\rho^2)^{(n-3)/2} \times \\
& \times \left[\frac{(\lambda-\bar{x})^2}{2s^2} \exp(\beta) + \frac{(\mu-\bar{y})^2}{2t^2} \exp(-\beta) + \cosh\beta - \frac{\rho}{st} (\lambda-\bar{x})(\mu-\bar{y}) - \rho r \right]^{-n} \quad (56)
\end{aligned}$$

the integration over α being performed by integrating by parts $(n-1)$ times (each time integrating the exponential factor and differentiating the power of α) and then integrating directly, absorbing various numerical factors into the proportionality factor as usual.

Reduction to algebraic form

I can now remove the exponentials and the fractional power from eqn (56) (although some irrationality will reappear in other forms). It is convenient first to introduce normalised forms of λ and μ , which will be transformed back at the end of the work; thus

$$\begin{aligned}
\lambda' & \equiv \frac{(\lambda-\bar{x})}{s} \\
\mu' & \equiv \frac{(\mu-\bar{y})}{t}
\end{aligned} \quad (57b)$$

(These should not be confused with the use of λ' and μ' to discuss rotations in the previous Part of this paper.) Then, absorbing factors of s and t (and a power of 2) into the proportionality factor,

$$\begin{aligned}
P_n(d\lambda' d\mu' d\rho|\theta) & \propto \int_{-\infty}^{\infty} d\beta \, d\lambda' d\mu' d\rho \, (1-\rho^2)^{(n-3)/2} \times \\
& \times \left[(\lambda'^2+1) \exp(\beta) + (\mu'^2+1) \exp(-\beta) - 2\rho(\lambda' \mu' + r) \right]^{-n} \quad (58a)
\end{aligned}$$

Next, I substitute

$$\exp(\beta) = \gamma \quad (58b)$$

$$d\gamma = \gamma d\beta \quad (58c)$$

$$(0 \leq \gamma < \infty)$$

then

$$\begin{aligned}
P_n(d\lambda' d\mu' d\rho|\theta) & \propto \int_0^\infty d\gamma \, d\lambda' d\mu' d\rho \, (1-\rho^2)^{(n-3)/2} \times \\
& \times \gamma^{n-1} \left[\gamma^2(\lambda'^2+1) + (\mu'^2+1) - 2\gamma\rho(\lambda' \mu' + r) \right]^{-n}
\end{aligned}$$

I wish also to integrate over all ρ , for which the range is from (-1) to 1 (see eqns (16) and the text following them); hence

$$\begin{aligned} P_n(d\lambda' d\mu'|\theta) &= \int_{-1}^1 d\rho \left[\frac{P_n(d\lambda' d\mu' d\rho|\theta)}{d\rho} \right] \\ &\propto d\lambda' d\mu' \int_{-1}^1 d\rho (1-\rho^2)^{(n-3)/2} \times \\ &\quad \times \int_0^\infty d\gamma \gamma^{n-1} \left[\gamma^2(\lambda'^2+1) + (\mu'^2+1) - 2\gamma\rho(\lambda'\mu'+r) \right]^{-n} \end{aligned}$$

I now define

$$\hat{I}_n \equiv \int_0^\infty d\gamma \gamma^{n-1} \left[\gamma^2(\lambda'^2+1) + (\mu'^2+1) - 2\gamma\rho(\lambda'\mu'+r) \right]^{-n}$$

When this integration has been performed, I can reduce eqn (58e) to a convenient trigonometric form by a substitution for ρ , which will remove the obvious fractional power (albeit at the cost of another complication). First, however, I must determine $\hat{I}_n$.

The second integration

The right-hand side of eqn (59) depends in a very simple way on the parameter r ; by inspection

$$\begin{aligned} \frac{\partial \hat{I}_n}{\partial r} &= \int_0^\infty d\gamma \gamma^{n-1} \left[\gamma^2(\lambda'^2+1) + (\mu'^2+1) - 2\gamma\rho(\lambda'\mu'+r) \right]^{-n-1} (2n\rho\gamma) \\ &= 2n\rho \hat{I}_{n+1} \end{aligned}$$

from which it immediately follows that

$$\hat{I}_n = \frac{(2\rho)^{1-n}}{(n-1)!} \left(\frac{\partial}{\partial r} \right)^{n-1} \hat{I}_1 \quad (60b)$$

with

$$\hat{I}_1 = \int_0^\infty d\gamma \left[\gamma^2(\lambda'^2+1) + (\mu'^2+1) - 2\gamma\rho(\lambda'\mu'+r) \right]^{-1} \quad (60c)$$

This last integral is of an elementary kind, and may be evaluated by factorising the denominator of the integrand and resolving into partial fractions; there is a complication, however, because the zeros of the denominator are complex. I write

$$\left[\gamma^2(\lambda'^2+1) + (\mu'^2+1) - 2\gamma\rho(\lambda'\mu'+r) \right] \equiv (\lambda'^2+1)(\gamma-\gamma_1)(\gamma-\gamma_2)$$

where γ_1, γ_2 are the zeros and can be written down from the well-known formula for solving a quadratic equation, thus

$$\gamma_{1,2} = \frac{\rho(\lambda'\mu'+r) \pm \sqrt{\rho^2(\lambda'\mu'+r)^2 - (\lambda'^2+1)(\mu'^2+1)}}{(\lambda'^2+1)} \quad (61b)$$

These are complex because the argument of the square root is essentially negative. To see this, notice first that the term $(\lambda'^2+1)(\mu'^2+1)$ is independent of the signs of λ' and μ' , and the term $\rho^2(\lambda'\mu'+r)^2$ depends only on the product of the signs of λ' and μ' . Without loss of generality, therefore, I may treat λ', μ' and r as all positive for the purpose of maximising the square root's argument; for if λ' and μ' have the same sign, $\lambda'\mu'$ is positive and must be added to a positive r to maximise it, while if λ' and μ' have opposite signs their product must be correspondingly added to a negative r , for which

$$(\lambda'\mu'+r)^2 = (-\lambda'\mu'-r)^2 = (|\lambda'\mu'|+|r|)^2$$

so that $\lambda'\mu'$ and r have the same effective values when both are negative as when both are positive. The maximum value of r is 1, as is the maximum value of ρ^2 (strictly, this is its upper limit which it is not allowed to attain); hence the maximum value of

$$\rho^2(\lambda'\mu'+r)^2 - (\lambda'^2+1)(\mu'^2+1)$$

is

$$\begin{aligned} (\lambda'\mu'+1)^2 - (\lambda'^2+1)(\mu'^2+1) &= 2\lambda'\mu' - \lambda'^2 - \mu'^2 \\ &= -(\lambda' - \mu')^2 \end{aligned}$$

which is zero if $\lambda' = \mu'$ and otherwise is obviously negative, and since ρ^2 cannot strictly attain its limit of 1, the square root's argument cannot attain zero. Accordingly, I write

$$\gamma_0 \equiv \frac{\rho(\lambda'\mu'+r)}{(\lambda'^2+1)}$$

$$\gamma_3 \equiv \frac{\left((\lambda'^2+1)(\mu'^2+1) - \rho^2(\lambda'\mu'+r)^2 \right)^{1/2}}{(\lambda'^2+1)} \quad (61d)$$

where γ_3 is real and positive, so that from eqn (61b)

$$\gamma_{1,2} = \gamma_0 \pm i\gamma_3 \quad (61e)$$

where i is the square root of (-1) (there should be no confusion between this and the use of i elsewhere as an integral index). Inserting eqns (61a) and (61e) into eqn (60c) gives

$$\begin{aligned}
\hat{I}_1 &= \int_0^{\infty} d\gamma (\lambda'^2 + 1)^{-1} (\gamma - \gamma_0 - i\gamma_3)^{-1} (\gamma - \gamma_0 + i\gamma_3)^{-1} \\
&= (\lambda'^2 + 1)^{-1} \int_0^{\infty} d\gamma \left\{ \frac{1}{\gamma - \gamma_0 + i\gamma_3} - \frac{1}{\gamma - \gamma_0 - i\gamma_3} \right\} (2i(-\gamma_3))^{-1} \\
&= \frac{i}{2\gamma_3(\lambda'^2 + 1)} \left[\log \frac{\gamma - \gamma_0 + i\gamma_3}{\gamma - \gamma_0 - i\gamma_3} \right]_0^{\infty}
\end{aligned}$$

Since γ is real, the numerator of the logarithm's argument is its denominator's complex conjugate; so the argument always has magnitude 1 and the value of the logarithm is simply i times the argument's phase, which in turn is twice the phase of the argument's numerator, provided the phases are chosen consistently with the requirement of mathematical continuity. Let the magnitude of $(\gamma - \gamma_0 + i\gamma_3)$ be M and its phase be ϕ , so that

$$M \cos \phi = \gamma - \gamma_0$$

$$M \sin \phi = \gamma_3 \quad (61h)$$

At the upper limit of integration, γ tends to $+\infty$ and so therefore does M , but $(M \sin \phi)$ remains fixed; hence $(\sin \phi)$ tends to zero. I choose $\phi = 0$, giving the principal value of the logarithm, at this upper limit (any value consistent with $(\sin \phi = 0)$ would do, but this is the most convenient). As γ decreases from $+\infty$ towards zero, M decreases while $(M \sin \phi)$ remains fixed and positive; hence $(\sin \phi)$ and ϕ increase, and must do so continuously. If γ_0 is negative, so that $(\gamma - \gamma_0)$ remains positive as γ decreases to zero, ϕ must remain less than $(\frac{1}{2}\pi)$ to keep $(\cos \phi)$ positive, and the value of ϕ at the lower limit of integration ($\gamma=0$) is the value satisfying the three requirements

$$0 < \phi < \frac{1}{2}\pi; \quad M \cos \phi = -\gamma_0; \quad M \sin \phi = \gamma_3$$

which is $\phi = \tan^{-1}(-\gamma_3/\gamma_0)$ (since γ_0 is here negative), taking $(\tan^{-1})$ in its usual range. If γ_0 is positive, $(\gamma - \gamma_0)$ passes through zero as γ decreases from $+\infty$; at this point $(\cos \phi)$ is also zero, so ϕ must be $(\frac{1}{2}\pi)$, since ϕ increases from zero as γ decreases from $+\infty$ (and there have not been any previous zeros of $(\cos \phi)$). As γ decreases further to zero, ϕ increases further; but it must remain between $(\frac{1}{2}\pi)$ and (π) , since $(\cos \phi)$ never vanishes again (so ϕ cannot cross $(\frac{1}{2}\pi)$) and $(\sin \phi)$ never vanishes at all for finite M and fixed non-zero $(M \sin \phi)$ (so ϕ cannot cross (π)). The value of ϕ at the lower limit of integration ($\gamma = 0$) is then the value satisfying the three requirements

$$(\frac{1}{2}\pi) < \phi < \pi; \quad M \cos \phi = -\gamma_0; \quad M \sin \phi = \gamma_3$$

which is $\phi = \pi + \tan^{-1}(-\gamma_3/\gamma_0)$ (since γ_0 is positive here), again taking $(\tan^{-1})$ in its usual range. Now, let δ be an arbitrary positive real number; then $\delta = \tan(\tan^{-1}\delta)$, so

$$\frac{1}{\delta} \frac{1}{\tan(\tan^{-1}\delta)} = \cot(\tan^{-1}\delta) = \tan(\frac{1}{2}\pi - \tan^{-1}\delta) \quad (61i)$$

and since the argument of the last tangent in eqn. (61i) is in the usual range of $(\tan^{-1})$, the equation becomes

$$\tan^{-1}\left(\frac{1}{\delta}\right) = \frac{1}{2}\pi - \tan^{-1}\delta \quad (61j)$$

When γ_0 is negative, $\phi = \tan^{-1}(-\gamma_3/\gamma_0)$, and $(-\gamma_0/\gamma_3)$ is positive. Let this be δ ; then (from eqn. (61j))

$$\begin{aligned} \phi &= \tan^{-1}(-\gamma_3/\gamma_0) = \tan^{-1}\left[\frac{1}{-\gamma_0/\gamma_3}\right] = \tan^{-1}\left[\frac{1}{\delta}\right] \\ &= \frac{1}{2}\pi - \tan^{-1}\delta = \frac{1}{2}\pi - \tan^{-1}(-\gamma_0/\gamma_3) = \frac{1}{2}\pi + \tan^{-1}(\gamma_0/\gamma_3) \end{aligned}$$

Now, if δ is an arbitrary negative real number, then eqn. (61i) becomes

$$\begin{aligned} \frac{1}{\delta} &= \frac{1}{\tan(\tan^{-1}\delta)} = \cot(\tan^{-1}\delta) = -\cot(\tan^{-1}(-\delta)) \\ &= -\tan\left(\frac{1}{2}\pi - \tan^{-1}(-\delta)\right) \end{aligned} \quad (61l)$$

and since the argument of the last tangent in eqn. (61l) is in the usual range of $(\tan^{-1})$, the equation becomes

$$\begin{aligned} \tan^{-1}\left[\frac{1}{\delta}\right] &= -\tan^{-1}\left[-\frac{1}{\delta}\right] = -\tan^{-1}\left(\tan\left(\frac{1}{2}\pi - \tan^{-1}(-\delta)\right)\right) \\ &= -\left(\frac{1}{2}\pi - \tan^{-1}(-\delta)\right) = -\frac{1}{2}\pi + \tan^{-1}(-\delta) \end{aligned}$$

When γ_0 is positive, $\phi = \pi + \tan^{-1}(-\gamma_3/\gamma_0)$, and $(-\gamma_0/\gamma_3)$ is negative. If this is chosen for δ , then (from eqn. (61m))

$$\begin{aligned} \phi &= \pi + \tan^{-1}(-\gamma_3/\gamma_0) = \pi + \tan^{-1}\left[\frac{1}{-\gamma_0/\gamma_3}\right] = \pi + \tan^{-1}\left[\frac{1}{\delta}\right] \\ \frac{1}{2}\pi + \tan^{-1}(-\delta) &= \frac{1}{2}\pi + \tan^{-1}(\gamma_0/\gamma_3) \end{aligned} \quad (61n)$$

which is eqn. (61k) again; this equation is therefore true for both positive and negative γ_0 . It is also true when $\gamma_0 = 0$, since then the point $(\gamma - \gamma_0) = 0$ at which $(\cos \phi)$ vanishes and ϕ attains $(\frac{1}{2}\pi)$ is also the lower limit of the integration (see the argument for positive γ_0 in the text preceding eqn. (61i)). It follows that, for all values of γ_0 , the change in the phase ϕ of the numerator of the argument of the logarithm in eqn. (61f) is $0 - (\frac{1}{2}\pi + \tan^{-1}(\gamma_0/\gamma_3))$ (taking γ now as moving in the correct sense, from 0 to $+\infty$); the phase change of the whole argument is twice this, or $(-\pi - 2\tan^{-1}(\gamma_0/\gamma_3))$; and the change in the value of the logarithm itself is i times this, or $-i(\pi + 2\tan^{-1}(\gamma_0/\gamma_3))$ (both results from the text following eqn. (61f)). So, from eqn. (61f),

$$\hat{I}_1 = (\lambda'^2 + 1)^{-1} \gamma_3^{-1} \left(\frac{1}{2}\pi + \tan^{-1}(\gamma_0/\gamma_3) \right)$$

As this result has required much delicate argument about the phases of complex logarithms, it is desirable to check it using the trigonometric substitution suggested by the form of eqn. (62). From the first line of eqn. (61f),

$$\hat{I}_1 = \int_0^{\infty} d\gamma (\lambda'^2 + 1)^{-1} \left((\gamma - \gamma_0)^2 + \gamma_3^2 \right)^{-1}$$

I write into this

$$\gamma = \gamma_0 + \gamma_3 \tan \phi' \quad (63b)$$

$$d\gamma = \gamma_3 (1 + \tan^2 \phi') d\phi' \quad (63c)$$

where ϕ' is the principal value of $\tan^{-1}(\tan \phi')$ with $(\tan \phi')$ defined by eqn. (63b); then the limits $\gamma = 0$ and $\gamma \rightarrow +\infty$ become $\phi' = \tan^{-1}(-\gamma_0/\gamma_3)$ and $\phi' = \frac{1}{2}\pi$ (since γ_3 is positive) and eqn. (63a) becomes

$$\begin{aligned} \hat{I}_1 &= \int_{\tan^{-1}(-\frac{\gamma_0}{\gamma_3})}^{\frac{1}{2}\pi} d\phi' \gamma_3 (1 + \tan^2 \phi') (\lambda'^2 + 1)^{-1} (\gamma_3^2 \tan^2 \phi' + \gamma_3^2)^{-1} \\ &= \int_{\tan^{-1}(-\frac{\gamma_0}{\gamma_3})}^{\frac{1}{2}\pi} (\lambda'^2 + 1)^{-1} \gamma_3^{-1} d\phi' \\ &= (\lambda'^2 + 1)^{-1} \gamma_3^{-1} \left[\phi' \right]_{\tan^{-1}(-\frac{\gamma_0}{\gamma_3})}^{\frac{1}{2}\pi} \end{aligned} \quad (63d)$$

from which eqn. (62) follows immediately. In principle, therefore, this section could be considerably shortened. It serves, however, as an illustration of the fact that most of the results in this paper were obtained by systematic analysis without the use of "short cuts"; indeed, in most cases I have looked for a "short cut" and failed to find one.

It remains now to substitute the definitions of γ_0 and γ_3 from eqns. (61) into eqn. (62). Before doing this, I make a small change which will shortly lead to a simplification; writing $\phi'' = \tan^{-1}(\gamma_0/\gamma_3)$ gives

$$\begin{aligned}\tan \phi'' &= \gamma_0/\gamma_3; & \sec^2 \phi'' &= 1 + \tan^2 \phi'' = 1 + (\gamma_0/\gamma_3)^2; \\ \sec \phi'' &= (\sec^2 \phi'')^{\frac{1}{2}} = \left((\gamma_3^2 + \gamma_0^2)/\gamma_3^2 \right)^{\frac{1}{2}} = (\gamma_3^2 + \gamma_0^2)^{\frac{1}{2}}/\gamma_3\end{aligned}$$

(since γ_3 is positive, and ϕ'' is the principal value of an arctangent and therefore lies between $(-\frac{1}{2}\pi)$ and $(\frac{1}{2}\pi)$ so that $\sec \phi''$ is also positive), so

$$\sin \phi'' = \tan \phi'' / \sec \phi'' = \gamma_0/(\gamma_3^2 + \gamma_0^2)^{\frac{1}{2}} \quad \text{723.38}$$

so that

$$\tan^{-1}(\gamma_0/\gamma_3) = \phi'' = \sin^{-1}(\sin \phi'') = \sin^{-1}(\gamma_0/(\gamma_3^2 + \gamma_0^2)^{\frac{1}{2}}) \quad (64a)$$

with the same principal value range for the arcsine as for the arctangent. From eqns. (61c) and (61d),

$$(\gamma_3^2 + \gamma_0^2)^{\frac{1}{2}} = \left((\mu'^2 + 1)/(\lambda'^2 + 1) \right)^{\frac{1}{2}}$$

and, on substituting into eqn. (62) from eqns. (64a), (64b) and (61c), with the further result (for positive γ_3)

$$\gamma_3 \equiv (\gamma_3^2 + \gamma_0^2)^{\frac{1}{2}} \left\{ 1 - (\gamma_0/(\gamma_3^2 + \gamma_0^2)^{\frac{1}{2}})^2 \right\}^{\frac{1}{2}}$$

I obtain

$$\begin{aligned}\hat{I}_1 &= (\lambda'^2 + 1)^{-1} \left((\mu'^2 + 1)/(\lambda'^2 + 1) \right)^{-\frac{1}{2}} \left\{ 1 - \left(\rho(\lambda' \mu' + r)/((\lambda'^2 + 1)(\mu'^2 + 1))^{\frac{1}{2}} \right)^2 \right\}^{-\frac{1}{2}} \times \\ &\quad \times \left[\frac{1}{2}\pi + \sin^{-1} \left(\rho(\lambda' \mu' + r)/((\lambda'^2 + 1)(\mu'^2 + 1))^{\frac{1}{2}} \right) \right] \\ &= \left((\lambda'^2 + 1)(\mu'^2 + 1) \right)^{-\frac{1}{2}} \left\{ 1 - \left[\frac{\rho(\lambda' \mu' + r)}{((\lambda'^2 + 1)(\mu'^2 + 1))^{\frac{1}{2}}} \right]^2 \right\}^{-\frac{1}{2}} \times \\ &\quad \times \left[\frac{1}{2}\pi + \sin^{-1} \left[\frac{\rho(\lambda' \mu' + r)}{((\lambda'^2 + 1)(\mu'^2 + 1))^{\frac{1}{2}}} \right] \right]\end{aligned}$$

The form of this suggests that I should introduce a "normalised" form of r , r' , defined by

$$r' \equiv \frac{(\lambda' \mu' + r)}{\left((\lambda'^2 + 1)(\mu'^2 + 1) \right)^{\frac{1}{2}}}$$

From the argument following eqn. (61b), the maximum value of

$$(\lambda' \mu' + r)^2 - (\lambda'^2 + 1)(\mu'^2 + 1)$$

is zero; hence the maximum value of

$$\frac{(\lambda' \mu' + r)^2}{(\lambda'^2 + 1)(\mu'^2 + 1)} - 1$$

which from the definition (66) is $(r'^2 - 1)$, is also zero. Hence

$$-1 \leq r' \leq 1$$

as for r , the limits being attainable in principle since r (unlike ρ) can also attain them (but this will be discussed again later). Eqn. (60b) requires $\hat{I}_1$ to be differentiated repeatedly by r ; it is easy to allow for this, however. I write

$$\hat{I}_1 = Cg(\rho r') = Cg(\rho(\lambda' \mu' + r)/((\lambda'^2 + 1)(\mu'^2 + 1))^{\frac{1}{2}})$$

where C is independent of r (and therefore of r' when the substitution is made), while $g(\rho r')$ is a function of only one variable, such that

$$g(z) = (1 - z^2)^{\frac{1}{2}} (\frac{1}{2}\pi + \sin^{-1} z)$$

for any z . Then

$$\begin{aligned} \frac{\partial}{\partial r} \hat{I}_1 &= \frac{\partial}{\partial r} \left\{ Cg(\rho(\lambda' \mu' + r)/((\lambda'^2 + 1)(\mu'^2 + 1))^{\frac{1}{2}}) \right\} \\ &= ((\lambda'^2 + 1)(\mu'^2 + 1))^{-\frac{1}{2}} \rho Cg'(\rho(\lambda' \mu' + r)/((\lambda'^2 + 1)(\mu'^2 + 1))^{\frac{1}{2}}) \\ &= ((\lambda'^2 + 1)(\mu'^2 + 1))^{-\frac{1}{2}} \rho Cg'(\rho r') \end{aligned}$$

where g' denotes the derivative of the function g of one variable. This is of the same form as that in eqn. (67a), since $((\lambda'^2 + 1)(\mu'^2 + 1))^{-\frac{1}{2}} \rho C$ is independent of r and g' is another function of only one variable; so the procedure may be repeated indefinitely to give

$$\left(\frac{\partial}{\partial r} \right)^{n-1} \hat{I}_1 = ((\lambda'^2 + 1)(\mu'^2 + 1))^{-\frac{1}{2}(n-1)} \rho^{n-1} Cg^{(n-1)'}(\rho r') \quad (67d)$$

Similarly

$$\frac{\partial}{\partial r'} \hat{I}_1 = \frac{\partial}{\partial r'} \{ Cg(\rho r') \} = \rho Cg'(\rho r')$$

which is also of the same form as that in eqn. (67a), so on repetition

$$\left(\frac{\partial}{\partial r'} \right)^{n-1} \hat{I}_1 = \rho^{n-1} Cg^{(n-1)'}(\rho r') \quad (67f)$$

Hence, combining eqns. (67d) and (67f),

$$\left[\frac{\partial}{\partial r} \right]^{n-1} \hat{I}_1 = ((\lambda'^2 + 1)(\mu'^2 + 1))^{-\frac{1}{2}(n-1)} \left[\frac{\partial}{\partial r'} \right]^{n-1} \hat{I}_1$$

Substituting eqns. (67g), (65) and (66) into eqn. (60b) now gives

$$\hat{I}_n = \frac{(2\rho)^{1-n}}{(n-1)!((\lambda'^2 + 1)(\mu'^2 + 1))^{\frac{1}{2}n}} \left[\frac{\partial}{\partial r'} \right]^{n-1} \left\{ (1 - (\rho r')^2)^{-\frac{1}{2}} \left(\frac{1}{2}\pi + \sin^{-1}(\rho r') \right) \right\}$$

which, on substituting into eqn. (58e) with the definition (59), and absorbing numerical factors into the proportionality factor, gives

$$P_n(d\lambda' d\mu' | \theta) \propto d\lambda' d\mu' ((\lambda'^2 + 1)(\mu'^2 + 1))^{-\frac{1}{2}n} \times \\ \times \int_{-1}^1 d\rho (1 - \rho^2)^{\frac{1}{2}(n-3)} \rho^{1-n} \left[\frac{\partial}{\partial r'} \right]^{n-1} \left\{ (1 - (\rho r')^2)^{-\frac{1}{2}} \left(\frac{1}{2}\pi + \sin^{-1}(\rho r') \right) \right\}$$

As indicated in the text following eqn. (58e) and the definition (59), I can now make a trigonometric substitution to simplify this, although the fractional powers will later reappear in a different form. Writing $\rho = \sin \theta'$ into eqn. (68a) puts it into the following form (with the usual range for the implied arcsine)

$$P_n(d\lambda' d\mu' | \theta) \propto d\lambda' d\mu' ((\lambda'^2 + 1)(\mu'^2 + 1))^{-\frac{1}{2}n} \hat{I}'_n \quad (68b)$$

where

$$\hat{I}'_n \equiv \int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} d\theta' (\cos \theta')^{n-2} (\sin \theta')^{1-n} \left[\frac{\partial}{\partial r'} \right]^{n-1} \left\{ (1 - (r' \sin \theta')^2)^{-\frac{1}{2}} \left(\frac{1}{2}\pi + \sin^{-1}(r' \sin \theta') \right) \right\}$$

This completes the second integration, and states the form of the third one.

Preliminaries to the third integration

The integral $\hat{I}'_n$ is apparently singular, because of the factor $(\sin \theta')^{1-n}$ in eqn. (68c) or ρ^{1-n} in eqn. (68a). The resolution of this difficulty is provided by eqns. (67e) and (67f), which show that each of the $(n-1)$ partial differentiations with respect to r' contributes a factor of ρ or $(\sin \theta')$ to the integrand, just cancelling the singularity. This result also shows, however, that it is impossible to take any of the partial differentiations outside the integral; they must all be performed in place. If n is very small there is apparently also a problem due to the factor $(\cos \theta')^{n-2}$; but it turns out that n must be at least 3 for all this analysis to be physically meaningful, which is sufficient to solve the problem.

Now, from eqn. (68c),

$$\begin{aligned}
\hat{I}'_n &= \int_0^{\frac{1}{2}\pi} d\theta' (\cos\theta')^{n-2} (\sin\theta')^{1-n} \left[\frac{\partial}{\partial r'} \right]^{n-1} \left\{ (1 - (r'^2 \sin^2 \theta'))^{-\frac{1}{2}} \left(\frac{1}{2}\pi + \sin^{-1}(r' \sin \theta') \right) \right\} \\
&+ \int_{-\frac{1}{2}\pi}^0 d\theta' (\cos\theta')^{n-2} (\sin\theta')^{1-n} \left[\frac{\partial}{\partial r'} \right]^{n-1} \left\{ (1 - (r'^2 \sin^2 \theta'))^{-\frac{1}{2}} \left(\frac{1}{2}\pi + \sin^{-1}(r' \sin \theta') \right) \right\} \\
&= \int_0^{\frac{1}{2}\pi} d\theta' (\cos\theta')^{n-2} (\sin\theta')^{1-n} \left[\frac{\partial}{\partial r'} \right]^{n-1} \left\{ (1 - r'^2 \sin^2 \theta')^{-\frac{1}{2}} \left(\frac{1}{2}\pi + \sin^{-1}(r' \sin \theta') \right) \right\} \\
&+ \int_0^{\frac{1}{2}\pi} d\theta' (-1)^{n-1} (\cos\theta')^{n-2} (\sin\theta')^{1-n} \left[\frac{\partial}{\partial r'} \right]^{n-1} \left\{ (1 - r'^2 \sin^2 \theta')^{-\frac{1}{2}} \left(\frac{1}{2}\pi - \sin^{-1}(r' \sin \theta') \right) \right\} \quad (68d)
\end{aligned}$$

on replacing θ' by $(-\theta')$ in the second integral of the first line. This equation takes different forms for odd and even n ; thus, if n is odd,

$$\hat{I}'_n = \pi \int_0^{\frac{1}{2}\pi} d\theta' (\cos\theta')^{n-2} (\sin\theta')^{1-n} \left[\frac{\partial}{\partial r'} \right]^{n-1} \left\{ (1 - r'^2 \sin^2 \theta')^{-\frac{1}{2}} \right\} \quad (69a)$$

and if n is even.

$$\hat{I}'_n = 2 \int_0^{\frac{1}{2}\pi} d\theta' (\cos\theta')^{n-2} (\sin\theta')^{1-n} \left[\frac{\partial}{\partial r'} \right]^{n-1} \left\{ (1 - r'^2 \sin^2 \theta')^{-\frac{1}{2}} \sin^{-1}(r' \sin \theta') \right\}$$

I can drop the numerical factors (π and 2) outside the integrals here, since they can be absorbed into the proportionality factor in eqn. (68b). It remains to evaluate the integrals in eqns. (69), which will be taken as defining $\hat{I}'_n$ from now on.

Odd number of measurement pairs in the sample

I consider first $\hat{I}'_n$ for odd n , defined by the integral in eqn. (69a), and I write into it

$$w = \sin^2 \theta' \quad (70a)$$

$$dw = 2 \sin \theta' \cos \theta' d\theta' \quad (70b)$$

$$(0 \leq w \leq 1)$$

Then, writing also odd $n = 2m + 1$, and discarding another numerical factor (of 2),

$$\hat{I}'_{2m+1} = \int_0^1 \frac{dw}{(w(1-w))^{\frac{1}{2}}} (1-w)^{m-\frac{1}{2}} w^{-m} \left[\frac{\partial}{\partial r'} \right]^{2m} \left\{ (1 - r'^2 w)^{-\frac{1}{2}} \right\} \quad (70c)$$

Now $(1-r'^2w)^{-\frac{1}{2}}$ may be expanded as a power series using the binomial theorem, thus

$$\begin{aligned}
 (1-r'^2w)^{-\frac{1}{2}} &= 1 + \left(-\frac{1}{2}\right)(-r'^2w) + \frac{\left(-\frac{1}{2}\right)\left(-\frac{3}{2}\right)}{2!}(-r'^2w)^2 + \frac{\left(-\frac{1}{2}\right)\left(-\frac{3}{2}\right)\left(-\frac{5}{2}\right)}{3!}(-r'^2w)^3 \dots \\
 &= 1 + \left[\frac{1}{2}\right](r'^2w) + \frac{\left(\frac{1}{2}\right)\left(\frac{3}{2}\right)}{2!}(r'^2w)^2 + \frac{\left(\frac{1}{2}\right)\left(\frac{3}{2}\right)\left(\frac{5}{2}\right)}{3!}(r'^2w)^3 \dots \\
 &= \sum_{j=0}^{\infty} \frac{\Gamma(j+\frac{1}{2})}{\Gamma(\frac{1}{2})} \frac{(r'^2w)^j}{j!}
 \end{aligned}$$

so

$$\begin{aligned}
 \left[\frac{\partial}{\partial r'}\right]^2 \left\{ (1-r'^2w)^{-\frac{1}{2}} \right\} &= \left[\frac{\partial}{\partial r'}\right]^2 \left\{ \sum_{j=0}^{\infty} \frac{\Gamma(j+\frac{1}{2})}{\Gamma(\frac{1}{2})} \frac{(r'^2w)^j}{j!} \right\} \\
 &= \sum_{j=1}^{\infty} \frac{\Gamma(j+\frac{1}{2})}{\Gamma(\frac{1}{2})} (2j)(2j-1)w \frac{(r'^2w)^{j-1}}{j!} \\
 &= w \sum_{j'=0}^{\infty} \frac{\Gamma(j'+\frac{3}{2})}{\Gamma(\frac{1}{2})} \frac{(2j'+2)!}{(2j')!} \frac{(r'^2w)^{j'}}{(j'+1)!} \quad (71b)
 \end{aligned}$$

and by repetition m times

$$\begin{aligned}
 \left[\frac{\partial}{\partial r'}\right]^{2m} \left\{ (1-r'^2w)^{-\frac{1}{2}} \right\} &= w^m \sum_{j=m}^{\infty} \frac{\Gamma(j+\frac{1}{2})}{\Gamma(\frac{1}{2})} (2j)(2j-1)\dots(2j-2m+1) \frac{(r'^2w)^{j-m}}{j!} \\
 &= w^m \sum_{j'=0}^{\infty} \frac{\Gamma(j'+m+\frac{1}{2})}{\Gamma(\frac{1}{2})} \frac{(2j'+2m)!}{(2j')!} \frac{(r'^2w)^{j'}}{(j'+m)!} \\
 &= \frac{\Gamma(m+\frac{1}{2})}{\Gamma(\frac{1}{2})} \frac{(2m)!}{m!} w^m \sum_{j=0}^{\infty} \frac{\Gamma(j+m+\frac{1}{2})}{\Gamma(m+\frac{1}{2})} \frac{(2j+2m)!}{(2m)!} \frac{1}{(2j)!} \frac{m!}{(j+m)!} (r'^2w)^j
 \end{aligned}$$

where the numerical factor (which as usual may be absorbed into the proportionality factor) has been chosen to make the "zeroth" term of the sum be 1, and the prime has been dropped from j' as no longer necessary. I now want to represent the sum by a hypergeometric function $F(a,b;c;z)$, which is defined by the following series expansion for $|z| < 1$ (see Copson [5])

$$F(a,b;c;z) \equiv \sum_{j=0}^{\infty} \frac{\Gamma(a+j)}{\Gamma(a)} \frac{\Gamma(b+j)}{\Gamma(b)} \frac{\Gamma(c)}{\Gamma(c+j)} \frac{z^j}{j!}$$

sum is the same for both sums for all $j (\geq 1)$. I have already used (from Copson [5]) the relation

$$\Gamma(x+1) = x\Gamma(x) \quad (71e)$$

and using it again gives the ratio of the j^{th} term to the $(j-1)^{\text{th}}$ term in the sum in eqn. (71c) as

$$\frac{(j+m-\frac{1}{2})(2j+2m)(2j+2m-1)}{(2j)(2j-1)(j+m)} (r'^2 w) \quad (71f)$$

while the same ratio in the sum in eqn. (71d) is

$$\frac{(a+j-1)(b+j-1)}{(c+j-1)(j)} z \quad (71g)$$

Cancelling $(j+m)$ and various factors of 2 between the numerator and denominator of expression (71f), and then equating it to expression (71g), gives

$$\frac{(j+m-\frac{1}{2})^2}{j(j-\frac{1}{2})} (r'^2 w) = \frac{(a+j-1)(b+j-1)}{j(c+j-1)} z \quad (71h)$$

for all $j (\geq 1)$; so immediately

$$a = b = (m+\frac{1}{2}); \quad c = \frac{1}{2} \quad z = r'^2 w \quad (71i)$$

and the sum in eqn. (71c) is $F((m+\frac{1}{2}), (m+\frac{1}{2}); \frac{1}{2}; (r'^2 w))$. On substituting into eqn. (70c) from eqn. (71c) and this last result, then, I obtain

$$\hat{I}'_{2m+1} \propto \int_0^1 dw w^{-\frac{1}{2}} (1-w)^{m-1} F(m+\frac{1}{2}, m+\frac{1}{2}; \frac{1}{2}; r'^2 w) \quad (71j)$$

Integration in hypergeometric form

The integral expression in eqn. (71j) is given as a standard form in Gradshteyn and Ryzhik [6], but there appears to be a misprint confusing the notation; the general result is correctly stated in Erdelyi [7] (where it is attributed to Bateman) with an indication of the method of proof. In view of the confusion, I shall work through the method to confirm that the result is correct.

I consider then the integral

$$\int_0^1 w^{\gamma-1} (1-w)^{c-\gamma-1} F(a, b; \gamma; zw) dw \quad (72a)$$

where $c > \gamma > 0$ and $|z| < 1$ (these restrictions guarantee convergence, taking all the parameters as real; the first restriction corresponds to the restriction on the powers of $(\sin \theta')$ and $(\cos \theta')$ in eqn. (68c), as discussed in the text following it). I substitute into expression (72a) the series expansion (71d); the integral is then

$$\frac{\Gamma(\gamma)}{\Gamma(a)\Gamma(b)} \int_0^1 w^{\gamma-1} (1-w)^{c-\gamma-1} \sum_{j=0}^{\infty} \frac{\Gamma(a+j)\Gamma(b+j)}{\Gamma(\gamma+j)} \frac{(zw)^j}{j!} dw \quad (72b)$$

As j tends to infinity, the limit of the ratio of successive terms of the infinite series becomes (zw) , which is strictly less than 1 in magnitude over the whole range of integration (including the

endpoints), since $|z| < 1$ by choice. The series therefore converges uniformly with respect to w over the whole range of integration, so I can invert the order of integration and summation (see the discussion of a similar integral involving a hypergeometric function by Copson [5]) to give expression (72b) as

$$\frac{\Gamma(\gamma)}{\Gamma(a)\Gamma(b)} \sum_{j=0}^{\infty} \frac{\Gamma(a+j)\Gamma(b+j)}{\Gamma(\gamma+j)} \frac{z^j}{j!} \int_0^1 w^{\gamma-1+j} (1-w)^{c-\gamma-1} dw \quad (72c)$$

The integral inside the sum here is a convergent Eulerian integral of the first kind, or Beta function, and is shown by Copson [5] to have the value

$$\int_0^1 w^{(\gamma+j)-1} (1-w)^{(c-\gamma)-1} dw = \frac{\Gamma(\gamma+j)\Gamma(c-\gamma)}{\Gamma((\gamma+j)+(c-\gamma))} = \frac{\Gamma(\gamma+j)\Gamma(c-\gamma)}{\Gamma(c+j)} \quad (72d)$$

so expression (72c) becomes

$$\begin{aligned} & \frac{\Gamma(\gamma)}{\Gamma(a)\Gamma(b)} \sum_{j=0}^{\infty} \frac{\Gamma(a+j)\Gamma(b+j)}{\Gamma(c+j)} \Gamma(c-\gamma) \frac{z^j}{j!} \\ & \quad \frac{\Gamma(\gamma)\Gamma(c-\gamma)}{\Gamma(c)} \sum_{j=0}^{\infty} \frac{\Gamma(a+j)}{\Gamma(a)} \frac{\Gamma(b+j)}{\Gamma(b)} \frac{\Gamma(c)}{\Gamma(c+j)} \frac{z^j}{j!} \\ & = \frac{\Gamma(\gamma)\Gamma(c-\gamma)}{\Gamma(c)} F(a, b; c; z) \end{aligned}$$

from the definition (71d); and this is the value of the integral expression (72a).

The factor multiplying the hypergeometric function in the last expression is itself a Beta function (corresponding to expression (72d) with $j = 0$), but this appears to be a mere curiosity with no deep significance. A rather more useful result is obtained by reversing the substitution ($w = \sin^2 \theta'$) (eqn. (70a)) into eqn. (72d); then on putting $(\gamma + j) = \frac{1}{2}$, $(c - \gamma) = \frac{1}{2}(n + 1)$, the equation for the Beta function integral becomes

$$2 \int_0^{\frac{1}{2}\pi} (\cos \theta')^n d\theta' = \frac{\Gamma(\frac{1}{2})\Gamma(\frac{1}{2}n + \frac{1}{2})}{\Gamma(\frac{1}{2}n + 1)}$$

which is a re-statement of eqn. (27).

Returning from this little diversion, I now choose to set $\gamma = \frac{1}{2}$, $c = (m + \frac{1}{2})$, $a = b = (m + \frac{1}{2})$ and $z = r'^2$ into expression (72a); these values make the integral the same as the one in eqn. (71j), and so the integral in eqn. (71j) is then given by expression (72e), provided the conditions on expression (72a) ($c > \gamma > 0$, $|z| < 1$) are satisfied. I then have

$$\begin{aligned} \hat{I}'_{2m+1} & \propto \frac{\Gamma(\frac{1}{2})\Gamma(m)}{\Gamma(m+\frac{1}{2})} F(m+\frac{1}{2}, m+\frac{1}{2}; m+\frac{1}{2}; r'^2) \\ & \propto F(m+\frac{1}{2}, m+\frac{1}{2}; m+\frac{1}{2}; r'^2) \end{aligned} \quad (74a)$$

on absorbing the numerical factor into the proportionality factor as usual. The conditions on expression (72a) now become $(m+\frac{1}{2}) > \frac{1}{2} > 0$, $r'^2 < 1$. The first of these reduces to $m > 0$, which from the definition of m (eqns. (70)) means that the smallest allowable odd n is 3; I have already discussed this in the text preceding eqn. (68d). The condition $r'^2 < 1$ will shortly become obviously necessary for another reason, but it adds another note to a theme which has been running through all this work: whether r' , r and ρ can attain their upper limits or not. Since it is always possible to put a straight line through two points in a plane, a sample of two measured value-pairs always exhibits perfect correlation; hence n of the sample must be at least 3 to make this analysis physically sensible. However, if n is at least 3, the correlation will always be imperfect (except by extraordinary coincidence); hence $|r|$ will be less than 1, and so will $|r'|$ be (from the definition (66)). The strict restriction on r is therefore harmless, adding nothing which would not be required anyway; while the physical restriction on n automatically satisfies the condition on m .

Now I consider eqn. (74a). From the definition (71d),

$$\begin{aligned}
 F(m+\frac{1}{2}, m+\frac{1}{2}; m+\frac{1}{2}; r'^2) &= \sum_{j=0}^{\infty} \frac{\Gamma(m+\frac{1}{2}+j)}{\Gamma(m+\frac{1}{2})} \frac{\Gamma(m+\frac{1}{2}+j)}{\Gamma(m+\frac{1}{2})} \frac{\Gamma(m+\frac{1}{2})}{\Gamma(m+\frac{1}{2}+j)} \frac{(r'^2)^j}{j!} \\
 &= \sum_{j=0}^{\infty} \frac{\Gamma(m+\frac{1}{2}+j)}{\Gamma(m+\frac{1}{2})} \frac{(r'^2)^j}{j!} \\
 &= 1 + (m+\frac{1}{2})(r'^2) + \frac{(m+\frac{1}{2})(m+\frac{3}{2})}{2!} (r'^2)^2 + \frac{(m+\frac{1}{2})(m+\frac{3}{2})(m+\frac{5}{2})}{3!} (r'^2)^3 \\
 &\quad + \frac{(-m-\frac{1}{2})(-m-\frac{3}{2})}{2!} (-r'^2)^2 + \frac{(-m-\frac{1}{2})(-m-\frac{3}{2})(-m-\frac{5}{2})}{3!} (-r'^2)^3 \\
 &= (1-r'^2)^{-m-\frac{1}{2}}
 \end{aligned} \tag{74b}$$

by the binomial theorem. Hence, since $(-m-\frac{1}{2}) = -\frac{1}{2}(2m+1) = -\frac{1}{2}n$ for odd n , from eqns. (74) and (68)

$$\begin{aligned}
 P_n(d\lambda' d\mu' | \theta) &\propto d\lambda' d\mu' ((\lambda'^2+1)(\mu'^2+1))^{-\frac{1}{2}n} (1-r'^2)^{-\frac{1}{2}n} \\
 &\quad d\lambda' d\mu' ((\lambda'^2+1)(\mu'^2+1)(1-r'^2))^{-\frac{1}{2}n} \\
 &\quad d\lambda' d\mu' ((\lambda'^2+1)(\mu'^2+1) - (\lambda'\mu' + r)^2)^{-\frac{1}{2}n} \\
 &\quad d\lambda' d\mu' ((1-r^2) + \lambda'^2 + \mu'^2 - 2r\lambda'\mu')^{-\frac{1}{2}n}
 \end{aligned} \tag{75}$$

for odd n , using the definition (66) of r' . (I note again that r'^2 must be strictly less than 1, since the substitution of eqn. (74b) plainly demands it.)

Earlier in this paper I used a discussion of the "Student's t" distribution to show that the prior probability distribution for the standard deviation(s) σ (and τ) ought to be taken as $(d\sigma/\sigma)$ rather than the plausible $(d\sigma)$ (see the text surrounding eqns. (30) and (31)). The result (75), and some of its immediate consequences, can be used in the same way to check that the prior probability distribution for the correlation coefficient ρ really ought to be taken as $((1-\rho^2)^{-3/2}d\rho)$ (as in eqns. (50)), rather than the plausible $(d\rho)$ – which Jeffreys [1] seems to believe, in spite of presenting himself the essence of all the arguments I have given for preferring the more complicated form. The main part of the necessary analysis will be performed later, but I mention it now because it is obviously necessary to know what the result (75) would be if the prior probability for ρ were given by the form $(d\rho)$. It is quite easy to trace this change through the last few sections; if I insert an extra factor of $(1-\rho^2)^{3/2}$ into eqns. (50), cancelling the factor $(1-\rho^2)^{-3/2}$ already there, this is carried through eqns. (54), (56), (58) and (68a), then (in the form $(\cos \theta)^3$) through the other eqns. (68) and eqns. (69), then (in the form $(1-w)^{3/2}$) through eqn. (70c) to arrive in eqn. (71j). The new integral in this last equation is evaluated as described in the text preceding eqn. (74a), with the chosen value of the parameter c changed appropriately from $(m+\frac{1}{2})$ to $(m+2)$, and eqn. (74a) then becomes (apart from a numerical factor which is absorbed as usual into the proportionality factor)

$$\langle \hat{I}'_{2m+1} \rangle \propto F(m+\frac{1}{2}, m+\frac{1}{2}; m+2; r'^2)$$

This cannot be written in closed form, but it can usefully be transformed. The necessary transformation appears as a by-product of the treatment of $\hat{I}'_n$ for even n , to which I now turn, postponing consideration of eqn. (76).

Even number of measurement pairs in the sample

I consider now $\hat{I}'_n$ for even n , defined by the integral in eqn. (69b), and I write into it the substitutions (70a) and (70b). Then, writing also even $n = 2m$, and discarding another numerical factor (of 2) as before,

$$\hat{I}'_{2m} = \int_0^1 \frac{dw}{(w(1-w))^{\frac{1}{2}}} (1-w)^{m-1} w^{\frac{1}{2}-m} \left[\frac{\partial}{\partial r'} \right]^{2m-1} \left\{ (1-r'^2 w)^{-\frac{1}{2}} \sin^{-1}(r' w^{\frac{1}{2}}) \right\}$$

It is not at all obvious what should be done next; in fact, the original development of this work was held up for several weeks while I searched for a way to proceed. The problem, in terms of the treatment of the odd case (eqns. (71)), is that the quantity to be differentiated is now not one elementary function but the product of two elementary functions, whose separate power series expansions cannot be combined into one power series by any elementary method; it does not even help to observe that $(1-z^2)^{-\frac{1}{2}} = d(\sin^{-1}z)/dz$, so that the problem can be reduced to that of squaring the power series for $(\sin^{-1}z)$. When I finally discovered an effective way to treat the problem, its solution turned out to be "obvious" with a great deal of hindsight – that is, it could be quickly verified using obscure properties of the hypergeometric function which one would never consider without prior knowledge of their relevance. However, this method of proof is unilluminating; there is a deep connection between the integrands in eqns. (69) which it leaves unexplored. Accordingly, I present a simplified systematic treatment based on my original investigation.

The idea of what follows is to find a "simple" (meaning linear) differential equation satisfied by the function in the brace brackets in eqns. (69b) and (77), which may be conveniently written as $((1-z^2)^{-\frac{1}{2}} \sin^{-1}z)$; then the differential equation can be solved in series, giving the wanted expansion directly, using the classical method of Picard and Frobenius (see Jeffreys and Jeffreys [3]). It is well known that the method of solution in series is much easier to apply if the differential equation is at most of second order; moreover, the hypergeometric equation is of this kind, and its first solution is just the hypergeometric function which has appeared several times

already in the discussion of an odd number of measurement pairs. Hence I look for a second-order linear differential equation satisfied by $((1-z^2)^{-1/2} \sin^{-1}z)$, which I will call $f(z)$; then, using primes to denote derivatives,

$$f(z) \equiv (1-z^2)^{-1/2} \sin^{-1}z$$

$$f'(z) = (1-z^2)^{-1} + z(1-z^2)^{-3/2} \sin^{-1}z \quad (78b)$$

$$f''(z) = 3z(1-z^2)^{-2} + (1+2z^2)(1-z^2)^{-5/2} \sin^{-1}z$$

with $(\sin^{-1}z)' = (1-z^2)^{-1/2}$ as remarked above. To find an identically-zero linear combination of these, I note that only $f'(z)$ and $f''(z)$ contain terms without $(\sin^{-1}z)$ as a factor, so I begin by choosing the combination $((1-z^2)f''(z) - 3zf'(z))$ to eliminate them; then

$$(1-z^2)f''(z) - 3zf'(z) = ((1+2z^2) - 3z \cdot z)(1-z^2)^{-3/2} \sin^{-1}z$$

$$(1-z^2)^{-1/2} \sin^{-1}z$$

and so, combining eqns. (78d) and (78a),

$$(1-z^2)f''(z) - 3zf'(z) - f(z) = 0$$

Now, by the very way this equation was constructed, it is also satisfied by the function $(1-z^2)^{-1/2}$. For, when eqns. (78a)–(78c) are substituted into the left-hand side of eqn. (78e) and similar terms are collected together, the terms with and without a factor of $(\sin^{-1}z)$ separately vanish identically; but the co-factor of the terms with a factor of $(\sin^{-1}z)$ is simply the result of applying the differential operator of the left-hand side of eqn. (78e), which is $((1-z^2)(d/dz)^2 - 3z(d/dz) - 1 \times)$, to the co-factor of $(\sin^{-1}z)$ in $f(z)$, which is $(1-z^2)^{-1/2}$. Hence I have found two obviously independent solutions of eqn. (78e), and the general solution must be a linear combination of them – which includes the function $g(z)$ in eqn. (67b), from which are deduced the later eqns. (67), (68) and (69); and it is easy to verify the general solution, by writing $f(z) = (1-z^2)^{-1/2}f(z)$ and substituting this into eqn. (78e). The appearance of two apparently unconnected functions in the brace brackets in eqns. (69) is now easily explained: they are a complementary pair; $(1-z^2)^{-1/2}$ (putting $r'\sin\theta' = z$) is even and $((1-z^2)^{-1/2} \sin^{-1}z)$ is odd, and this is the unique partition of the general solution of eqn. (78e) yielding particular solutions with these properties (apart from numerical scaling factors, which are essentially irrelevant because the differential equation is linear). A quirk is that the even brace-bracketed function appears in the equation for I_n' with odd n , and vice versa, in eqns. (69). The reason for this is that both integrands in eqns. (69) are even, the odd parts having dropped out when the range was changed from $(-\frac{1}{2}\pi$ to $\frac{1}{2}\pi)$ (in eqn. (68c)) to $(0$ to $\frac{1}{2}\pi)$ (in eqn. (68d)); and so, since the operator $((\sin\theta')^{1-n} (\partial/\partial r')^{n-1})$ is $(\partial/\partial z)^{n-1}$ and is even for odd n and vice versa, so must its operand be.

I can now find the series expansion of $((1-z^2)^{-1/2} \sin^{-1}z)$ by solving eqn. (78e) in series and picking out the odd solution. Since the equation is not hypergeometric, there is no obvious way of predicting its solutions from the hypergeometric function in the general case – although the even solution, $(1-z^2)^{-1/2}$, is $F(\frac{1}{2}, \frac{1}{2}; 1; z^2)$ (by putting $m = 0$ in eqn. (74b) and replacing r' by z), which can also be written $F(\frac{1}{2}, c; c; z^2)$ by noting (from the intermediate lines of eqn. (74b)) that two pairs of gamma-functions cancel identically when $b = c$ in $F(a, b; c; z^2)$. I therefore originally carried out the tedious but standard series-solution process (as described in Jeffreys and Jeffreys [3]),

obtaining the result

$$(1-z^2)^{-\frac{1}{2}} \sin^{-1} z = z F(1, 1; 3/2; z^2) \quad (79)$$

after using the method exemplified in eqns. (71) to express the series in hypergeometric form. On comparing these two solutions with each other and with the standard solutions of the hypergeometric equation, which (as shown in Copson [5]) are $F(a, b; c; z)$ and $z^{1-c} F(1+a-c, 1+b-c; 2-c; z)$, and noting that c may be taken as $\frac{1}{2}$, it does become rather plausible (if still not obvious) that z in eqn. (78e) should be replaced by $\bar{z}^{\frac{1}{2}}$ (introducing the bar to avoid confusion). Then, since

$$\frac{d}{dz} = \frac{d}{d(\bar{z}^{\frac{1}{2}})} \left[\frac{d(\bar{z}^{\frac{1}{2}})}{dz} \right]^{-1} \frac{d}{d\bar{z}} = 2\bar{z}^{\frac{1}{2}} \frac{d}{d\bar{z}} \quad (80a)$$

eqn. (78e) becomes

$$(1-\bar{z}) \left[2\bar{z}^{\frac{1}{2}} \frac{d}{d\bar{z}} \right]^2 \bar{f}(\bar{z}) - 3\bar{z}^{\frac{1}{2}} \left[2\bar{z}^{\frac{1}{2}} \frac{d}{d\bar{z}} \right] \bar{f}(\bar{z}) - \bar{f}(\bar{z}) = 0$$

or

$$\bar{z}(1-\bar{z}) \frac{d^2 \bar{f}}{d\bar{z}^2} + \left(\frac{1}{2} - 2\bar{z} \right) \frac{d\bar{f}}{d\bar{z}} - \frac{1}{4} \bar{f} = 0 \quad (80b)$$

which is the hypergeometric equation with the parameters $a = b = c = \frac{1}{2}$ (see Copson [5]) and therefore has the solutions $F(\frac{1}{2}, \frac{1}{2}; \frac{3}{2}; \bar{z})$ and $\bar{z}^{\frac{1}{2}} F(1, 1; 3/2; \bar{z})$ which have already been identified, since setting $\bar{z} = z^2$ makes them be $F(\frac{1}{2}, \frac{1}{2}; \frac{3}{2}; z^2)$ and $z F(1, 1; 3/2; z^2)$ as above (eqns. (74b) and (79)). (In the general theory of the hypergeometric function there are certain difficulties when $a = b$, but they do not affect the results here, as the direct solution in series shows.)

Finally, at the deepest level to which I shall go in exploiting the properties of the hypergeometric function, I quote one of the results deduced by Kummer from studying transformations of the hypergeometric equation into itself using the elementary bilinear transformations of the cross-ratio group; these change the independent variable $\bar{z}$ through the sequence $\bar{z} \rightarrow (1-\bar{z}) \rightarrow 1/(1-\bar{z}) \rightarrow (-\bar{z})/(1-\bar{z}) \rightarrow 1-1/\bar{z} \rightarrow 1/\bar{z} \rightarrow \bar{z}$, which is obtained by alternately subtracting from unity and inverting (see Erdelyi [7] or Piaggio [8]). The result I want is

$$F(a, b; c; \bar{z}) = (1-\bar{z})^{c-a-b} F(c-a, c-b; c; \bar{z}) \quad (81a)$$

This is stated without comment by Erdelyi [7]; Piaggio [8] remarks that making the substitution $\bar{f}(\bar{z}) = (1-\bar{z})^k \bar{F}(\bar{z})$ transforms the hypergeometric equation for $\bar{f}(\bar{z}) = F(a, b; c; \bar{z})$ into another hypergeometric equation for $\bar{F}(\bar{z})$ (provided $k = c-a-b$) with parameters such that the two solutions $F(a, b; c; \bar{z})$ and $\bar{z}^{1-c} F(1+a-c, 1+b-c; 2-c; \bar{z})$ of the original equation become (for the new equation)

$$\begin{aligned} \bar{F}(\bar{z}) &= \text{a linear combination of } F(c-a, c-b; c; \bar{z}) \text{ and of} \\ &\quad \bar{z}^{1-c} F(1-a, 1-b; 2-c; \bar{z}) \end{aligned} \quad (81b)$$

from which eqn. (81a) follows, on inserting the factor $(1-\bar{z})^k$ (to change $\bar{F}(\bar{z})$ back to $\bar{f}(\bar{z})$) and considering the behaviour of the four solutions at $\bar{z} = 0$. Then, on writing $\bar{z} = z^2$ as before, eqn. (81a) gives obviously

$$F(a,b;c;z^2) = (1-z^2)^{c-a-b} F(c-a,c-b;c;z^2) \quad (81c)$$

It is now very simple to verify the series expansion of $((1-z^2)^{-1/2} \sin^{-1}z)$. From eqn. (71a) and the remark that $(1-z^2)^{-1/2} = d(\sin^{-1}z)/dz$, I obtain using term-by-term integration (with $\sin^{-1}0 = 0$)

$$\begin{aligned} \sin^{-1}z &= \int dz \sum_{j=0}^{\infty} \frac{\Gamma(j+\frac{1}{2})}{\Gamma(\frac{1}{2})} \frac{z^{2j}}{j!} = \sum_{j=0}^{\infty} \frac{\Gamma(j+\frac{1}{2})}{\Gamma(\frac{1}{2})} \frac{z^{2j+1}}{(2j+1)j!} \\ &= z \sum_{j=0}^{\infty} \frac{\Gamma(j+\frac{1}{2})}{\Gamma(\frac{1}{2})} \frac{z^{2j}}{(2j+1)j!} \end{aligned}$$

The "zeroth" terms of the sums in eqn. (82a) and eqn. (71d) (the hypergeometric function) are both 1; and the ratio of the j^{th} term to the $(j-1)^{\text{th}}$ term in eqn. (71d) has already been identified as expression (71g), while the corresponding ratio of terms in eqn. (82a) is (using eqn. (71e))

$$\frac{(j-\frac{1}{2})(2j-1)}{j(2j+1)} z^2 = \frac{(j-\frac{1}{2})^2}{(j+\frac{1}{2})j} \frac{z}{z}$$

Hence the sum in eqn. (82a) is (comparing expressions (71g) and (82b)) simply $F(\frac{1}{2}, \frac{1}{2}; 3/2; z^2)$ or $F(\frac{1}{2}, \frac{1}{2}; 3/2; z^2)$, and so eqn. (82a) becomes

$$\sin^{-1}z = z F(\frac{1}{2}, \frac{1}{2}; 3/2; z^2)$$

so

$$(1-z^2)^{-1/2} \sin^{-1}z = z \left\{ (1-z^2)^{-1/2} F(\frac{1}{2}, \frac{1}{2}; 3/2; z^2) \right\} \quad (82d)$$

However, the expression in brace brackets in eqn. (82d) is the right-hand side of eqn. (81c), with the parameters $a = 1$, $b = 1$, $c = 3/2$; and so, from the left-hand side of eqn. (81c), eqn. (79) follows immediately.

Eqn. (81c) will also be used when I come to discuss eqn. (76) and its counterpart for even n , which will be derived in due course.

I can now return to eqn. (77), and discuss it in the same way that eqn. (70c) was discussed, following the sequence of eqns. (71). In the first place, from eqns. (79) and (71d) I have

$$\begin{aligned} (1-r'^2 w)^{-1/2} \sin^{-1}(r' w^{1/2}) &= (r' w^{1/2}) \sum_{j=0}^{\infty} \frac{\Gamma(j+1)}{\Gamma(1)} \frac{\Gamma(j+1)}{\Gamma(1)} \frac{\Gamma(\frac{3}{2})}{\Gamma(j+\frac{3}{2})} \frac{(r'^2 w)^j}{j!} \\ &= (r' w^{1/2}) \sum_{j=0}^{\infty} \frac{\Gamma(\frac{3}{2})}{\Gamma(j+\frac{3}{2})} j! (r'^2 w)^j \end{aligned} \quad (83a)$$

which is the result I obtained originally by solving eqn. (78e) in series. So

$$\begin{aligned}
\frac{\partial}{\partial r'} \left\{ (1-r'^2 w)^{-\frac{1}{2}} \sin^{-1}(r' w^{\frac{1}{2}}) \right\} &= \frac{\partial}{\partial r'} (w^{\frac{1}{2}}) \sum_{j=0}^{\infty} \frac{\Gamma(\frac{3}{2})}{\Gamma(j+\frac{3}{2})} j! w^j r'^{2j+1} \\
&= w^{\frac{1}{2}} \sum_{j=0}^{\infty} \frac{\Gamma(\frac{3}{2})}{\Gamma(j+\frac{3}{2})} (2j+1) j! (r'^2 w)^j
\end{aligned} \tag{83b}$$

then, comparing with eqns. (71b) and (71c),

$$\begin{aligned}
\left[\frac{\partial}{\partial r'} \right]^2 \left[\frac{\partial}{\partial r'} \right] \left\{ (1-r'^2 w)^{-\frac{1}{2}} \sin^{-1}(r' w^{\frac{1}{2}}) \right\} &= \left[\frac{\partial}{\partial r'} \right]^2 w^{\frac{1}{2}} \sum_{j=0}^{\infty} \frac{\Gamma(\frac{3}{2})}{\Gamma(j+\frac{3}{2})} (2j+1) j! (r'^2 w)^j \\
&= w^{\frac{1}{2}} \sum_{j=1}^{\infty} \frac{\Gamma(\frac{3}{2})}{\Gamma(j+\frac{3}{2})} (2j+1)(2j)(2j-1) j! w (r'^2 w)^{j-1} \\
&= w^{\frac{1}{2}} \sum_{j'=0}^{\infty} \frac{\Gamma(\frac{3}{2})}{\Gamma(j'+\frac{5}{2})} (2j'+3)(2j'+2)(2j'+1) (j'+1)! (r'^2 w)^{j'}
\end{aligned} \tag{83c}$$

and by repetition (m-1) times

$$\begin{aligned}
\left[\frac{\partial}{\partial r'} \right]^{2m-1} \left\{ (1-r'^2 w)^{-\frac{1}{2}} \sin^{-1}(r' w^{\frac{1}{2}}) \right\} \\
&= w^{m-\frac{1}{2}} \sum_{j=m-1}^{\infty} \frac{\Gamma(\frac{3}{2})}{\Gamma(j+\frac{3}{2})} (2j+1)(2j)(2j-1) \dots (2j-2m+3) j! (r'^2 w)^{j-m+1} \\
&= w^{m-\frac{1}{2}} \sum_{j'=0}^{\infty} \frac{\Gamma(\frac{3}{2})}{\Gamma(j'+m+\frac{1}{2})} \frac{(2j'+2m-1)!}{(2j')!} (j'+m-1)! (r'^2 w)^{j'} \\
&= \frac{\Gamma(\frac{3}{2})}{\Gamma(m+\frac{1}{2})} (2m-1)! (m-1)! w^{m-\frac{1}{2}} \sum_{j=0}^{\infty} \frac{\Gamma(m+\frac{1}{2})}{\Gamma(j+m+\frac{1}{2})} \frac{(2j+2m-1)!}{(2m-1)!} \frac{(j+m-1)!}{(m-1)!} \frac{(r'^2 w)^j}{(2j)!}
\end{aligned} \tag{83d}$$

where the numerical factor (which may be absorbed into the proportionality factor) has been chosen to make the "zeroth" term of the sum be 1, and the prime has been dropped from j' as no longer necessary. I now want (as usual) to represent the sum by a hypergeometric function (as given in definition (71d)), also having a "zeroth" term of 1, so the representation will be valid if I can choose the parameters a , b and c and the variable z of the hypergeometric function so that the ratio of the j^{th} term to the $(j-1)^{\text{th}}$ term is the same in the sum in eqn. (83d) as in the hypergeometric definition (71d) for all $j (\geq 1)$. The ratio in the definition is already given by expression (71g); using eqn. (71e), the ratio in the sum in eqn. (83d) is

$$\frac{(2j+2m-1)(2j+2m-2)(j+m-1)}{(j+m-\frac{1}{2})(2j)(2j-1)} (r'^2 w)$$

Cancelling $(j+m-\frac{1}{2})$ and various factors of 2 between the numerator and denominator of expression (83e), and then equating it to expression (71g), gives

$$\frac{(j+m-1)^2}{j(j-\frac{1}{2})} (r'^2 w) = \frac{(a+j-1)(b+j-1)}{j(c+j-1)} z$$

for all $j (\geq 1)$; so immediately

$$a = b = m; \quad c = \frac{1}{2}; \quad z = r'^2 w$$

and the sum in eqn. (83d) is $F(m, m; \frac{1}{2}; r'^2 w)$. On substituting into eqn. (77) from eqn. (83d) and this last result, then, I obtain

$$\hat{I}'_{2m} \propto \int_0^1 dw w^{-\frac{1}{2}} (1-w)^{m-\frac{3}{2}} F(m, m; \frac{1}{2}; r'^2 w)$$

Integration in hypergeometric form revisited

The integral expression in eqn. (83h) is of the same form as the one in eqn. (71j), both being particular cases of the general expression (72a) whose value is expression (72e); hence, apart from a numerical factor which is absorbed as usual, $\hat{I}'_{2m}$ is given by setting in expression (72e) the parameter values $\gamma = \frac{1}{2}$, $c = m$, $a = b = m$, $z = r'^2$, which make the integrals in eqns. (72a) and (83h) the same. I then have

$$\begin{aligned} \hat{I}'_{2m} &\propto \frac{\Gamma(\frac{1}{2})\Gamma(m-\frac{1}{2})}{\Gamma(m)} F(m, m; m; r'^2) \\ &\propto F(m, m; m; r'^2) \end{aligned}$$

In this case, the conditions on expression (72a) (stated immediately after it) become $r'^2 < 1$ (which has already been discussed) and $m > \frac{1}{2} > 0$; the latter condition reduces to $m > \frac{1}{2}$, which from the definition of m here (in the statement of eqn. (77)) means that the smallest allowable even n is 2 as far as this integration is concerned (see the text preceding eqn. (68d)). In fact, the smallest allowable n (even or odd) is 3, as already shown in the discussion following eqn. (74a); this result will be confirmed by later mathematical developments.

Now consider eqn. (84a), and compare it with its analogue, eqn. (74a); the treatment of this latter equation in eqn. (74b) extends immediately to the present case, on replacing $(m+\frac{1}{2})$ by m , and so

$$F(m, m; m; r'^2) = (1-r'^2)^{-m} \quad (84b)$$

Here, with even n , $(-m) = (-\frac{1}{2}n)$, so the analysis of eqn. (75) can be taken over unchanged to give

$$P_n(d\lambda' d\mu' | \theta) \propto d\lambda' d\mu' ((1-r'^2)^{\lambda'^2 + \mu'^2 - 2r\lambda'\mu'})^{-\frac{1}{2}n}$$

for all (odd or even) n sufficiently large for the total probability to be represented by a convergent integral expression (so that it is well-behaved and can be normalised). The discussion following eqn. (75), concerning the effect of taking the prior probability distribution for the correlation

coefficient ρ as the plausible ($d\rho$) instead of the calculated $((1-\rho^2)^{-3/2}d\rho)$, can also be taken over unchanged (apart from replacing the later equation numbers by their analogues presented in the section before this one) down to the conclusion that an extra factor of $(1-w)^{3/2}$ appears in eqn. (71j) and its analogue eqn. (83h); the new integral form of eqn. (83h) is then evaluated in the same way as the old one, with the chosen value of the parameter c changed appropriately from m to $(m+3/2)$, and the equation then becomes (analogously to eqn. (76))

$$\langle \hat{I}'_{2m} \rangle \propto F(m, m; m + \frac{3}{2}; r'^2)$$

On inserting $m = \frac{1}{2}n$ into this, and $m = \frac{1}{2}(n-1)$ into eqn. (76) for the case of odd n (as in the definitions preceding eqns. (77) and (70c)), I obtain for both even and odd n

$$\langle \hat{I}'_n \rangle \propto F(\frac{1}{2}n, \frac{1}{2}n; \frac{1}{2}n + \frac{3}{2}; r'^2)$$

This can usefully be transformed, using eqn. (81a) with $a = b = \frac{1}{2}n$, $c = (\frac{1}{2}n + 3/2)$, $\bar{z} = r'^2$, to

$$\langle \hat{I}'_n \rangle \propto (1-r'^2)^{\frac{3}{2} - \frac{1}{2}n} F(\frac{3}{2}, \frac{3}{2}, \frac{1}{2}n + \frac{3}{2}; r'^2) \quad (8)$$

and on combining this with eqns. (68), the result analogous to eqns. (75) and (85) is

$$\langle P_n(d\lambda' d\mu' | \theta) \rangle \propto d\lambda' d\mu' ((1-r^2) + \lambda'^2 + \mu'^2 - 2r\lambda'\mu')^{-\frac{1}{2}n} (1-r'^2)^{\frac{3}{2}} F(\frac{3}{2}, \frac{3}{2}, \frac{1}{2}n + \frac{3}{2}; r'^2) \quad (86d)$$

According to the discussions following eqns. (75) and (85), the form of the prior probability distribution for ρ can be checked by investigating the normalisation of the probability distributions given by eqns. (85) and (86d). As this normalisation is obviously desirable anyway, I now carry it out.

Exact normalisation of the true probability distribution

As in earlier results (eqns. (23)–(30)), the proportionality factor in eqn. (85) is determined by integrating over the whole ranges of λ' and μ' ; since the original variables λ and μ (to which I shall return in due course) range from $-\infty$ to $+\infty$, so do λ' and μ' , from eqns. (57) – provided that the sample parameters s and t are non-zero, which from eqns. (52) requires $n \geq 2$. Now, I write the proportionality factor as $\bar{K}_n$; then from eqn. (85), since the total probability is 1,

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} d\lambda' d\mu' \left[\frac{P_n(d\lambda' d\mu' | \theta)}{d\lambda' d\mu'} \right] \\ \int_{-\infty}^{\infty} d\lambda' \int_{-\infty}^{\infty} d\mu' \bar{K}_n ((1-r^2) + \lambda'^2 + \mu'^2 - 2r\lambda'\mu')^{-\frac{1}{2}n}$$

I take the integration over μ' first; then for any λ' I can "complete the square" and write

$$\mu' = r\lambda' + ((\lambda'^2 + 1)(1-r^2))^{\frac{1}{2}} \tan \phi$$

$$d\mu' = ((\lambda'^2 + 1)(1 - r^2))^{\frac{1}{2}} \sec^2 \bar{\phi} d\bar{\phi}$$

with $-\frac{1}{2}\pi \leq \bar{\phi} \leq \frac{1}{2}\pi$; then eqn. (87a) becomes

$$\bar{K}_n^{-1} = \int_{-\infty}^{\infty} d\lambda' \int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} ((\lambda'^2 + 1)(1 - r^2))^{-\frac{1}{2}n + \frac{1}{2}} (\cos \bar{\phi})^{n-2} d\bar{\phi} \quad (87d)$$

for suitable values of n . Again, making the simplest form of the same substitution, I can write

$$\lambda' = \tan \bar{\phi}' \quad (87e)$$

$$d\lambda' = \sec^2 \bar{\phi}' d\bar{\phi}' \quad (87f)$$

(compare eqns. (30e) and (57a)) with $-\frac{1}{2}\pi \leq \bar{\phi}' \leq \frac{1}{2}\pi$; then eqn. (87d) becomes

$$\bar{K}_n^{-1} = (1 - r^2)^{-\frac{1}{2}n + \frac{1}{2}} \int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} d\bar{\phi}' (\cos \bar{\phi}')^{n-3} \int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} d\bar{\phi} (\cos \bar{\phi})^{n-2}$$

for suitable n , which now clearly means $n \geq 3$. The integrals here can be written down from eqn. (73), remembering that

$$2 \int_0^{\frac{1}{2}\pi} (\cos \theta')^n d\theta' = \int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} (\cos \theta')^n d\theta' \quad (87h)$$

because $(\cos \theta')$ and its integral powers are even functions of θ' ; thus

$$\bar{K}_n^{-1} = (1 - r^2)^{-\frac{1}{2}n + \frac{1}{2}} \left[\frac{\Gamma(\frac{1}{2})\Gamma(\frac{1}{2}n-1)}{\Gamma(\frac{1}{2}n-\frac{1}{2})} \right] \left[\frac{\Gamma(\frac{1}{2})\Gamma(\frac{1}{2}n-\frac{1}{2})}{\Gamma(\frac{1}{2}n)} \right]$$

From eqn. (71e) with $x = \frac{1}{2}n-1$, $\Gamma(\frac{1}{2}n) = (\frac{1}{2}n-1)\Gamma(\frac{1}{2}n-1)$; and for $n \geq 3$ the argument $(\frac{1}{2}n-1) > 0$, so all the gamma functions have positive arguments and are finite. Hence, since $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, eqn. (87i) becomes

$$\bar{K}_n^{-1} = (1 - r^2)^{-\frac{1}{2}n + \frac{1}{2}} \left[\frac{2\pi}{n-2} \right] \quad (87j)$$

and eqn. (85) becomes

$$P_n(d\lambda' d\mu' | \theta) = \frac{(n-2)}{2\pi(1-r^2)^{\frac{1}{2}}} \left[1 + \frac{\lambda'^2 + \mu'^2 - 2r\lambda'\mu'}{1-r^2} \right]^{-\frac{1}{2}n} d\lambda' d\mu' \quad (88)$$

with a very strong resemblance to eqns. (25) and (26).

It has been asserted several times that n must be at least 3 for meaningful results to be obtained. In fact, however, the critical or degenerate case is $n = 2$: that is, not the ordinary case of lowest order ($n = 3$) but the next one below it. From eqns. (52), all the sample parameters are well-defined for $n = 2$, and s and t are (normally) non-zero; the degeneracy consists in the fact that, when $n = 2$, r must be ± 1 (perfect correlation in the sample). The whole of the analysis presented in Part III so far is valid for $n = 2$ and $r = \pm 1$, except that r'^2 can now attain 1 when $\lambda' = \pm \mu'$ with the same sign as r (and r'); however, eqns. (84) are still true, regarded as limits in the case $r'^2 = 1$,

and eqn. (85) then assumes the form

$$P_2(d\lambda' d\mu' | \theta) \propto d\lambda' d\mu' (\lambda' \mp \mu')^{-2} \quad (89)$$

which is finite when r^2 is less than 1 and correctly approaches infinity as r^2 approaches 1. Now, the distribution (89) is not integrable through its "line of singularity" in the (λ, μ) -plane, just as the degenerate one-dimensional distribution (30g) is not integrable through its point of singularity on the (λ) -line. As discussed in the text surrounding eqns. (30), the one-dimensional result is perfectly reasonable; one measurement gives a definite (if crude) estimate of the underlying distribution's mean, but not only does it not give any information about the "spread" of the distribution, it also predicts (in effect) that there is no spread: $s = 0$, so the distribution must be (in a sense) infinitely tightly bound to the sample mean. Similarly, in the two-dimensional case, two measurement pairs give definite (if crude) estimates for the underlying distribution's twin means and standard deviations, but not only do they give no information about the variation of correlation within the distribution, they predict (in effect) that there is no room for variations in correlation: $r = \pm 1$, so the correlation must be perfect, and a "line of singularity" is to be expected. In the next section, I analyse the results of this kind given by the alternative distribution in eqns. (86), and I show that they are well-behaved not only for $n \geq 3$ (as expected) but also at $n = 2$: which, by the argument just given, I assert is unphysical and wrong. It is interesting to note the paradoxical nature of this reasoning; I condemn the alternative distribution, not because it predicts an extraordinary result, but because it fails to predict an extraordinary result in a case where the "natural physics" of the situation demands one (and duly gets it from the true distribution).

Approximate normalisation of the alternative distribution, and its disproof

Consider first the behaviour of the distribution (86d) for $n > 3$. From the definition (71d), the hypergeometric series is geometrically convergent for all values of its argument z less than 1 in magnitude. The only relevant value outside this range is $r^2 = 1$, for which the series gives $F(3/2, 3/2; \frac{1}{2}n+3/2; 1)$ provided this converges. According to Copson [5],

$$F(a, b; c; 1) = \frac{\Gamma(c) \Gamma(c-a-b)}{\Gamma(c-a) \Gamma(c-b)} \quad (90a)$$

provided the real part of $(c-a-b)$ is strictly positive; this requires $(\frac{1}{2}n-3/2) > 0$ or $n > 3$ (which is the reason for excluding $n = 3$ for the moment). Moreover, for all $n > 3$ the four gamma functions in eqn. (90a) have positive arguments for the required values $a=b=3/2$ and $c=\frac{1}{2}n+3/2$, so they are all finite and non-zero, and so $F(3/2, 3/2; \frac{1}{2}n+3/2; 1)$ is finite. Hence $(1-r^2)^{3/2} F(3/2, 3/2; \frac{1}{2}n+3/2; r^2)$ is finite for all allowable values of r^2 ; it is also non-negative, since the terms of the series definition (71d) are all positive except when $z = 0$, and then the series reduces to 1. Hence the co-factor of the differentials $(d\lambda' d\mu')$ in eqn. (86d) is non-negative, and the integral over all λ' and μ' is positive; also, the integral is finite, because its integrand is everywhere bounded above by

$$((1-r^2) + \lambda'^2 + \mu'^2 - 2r\lambda'\mu')^{-\frac{1}{2}n} G_n \quad (90b)$$

where G_n is the greatest value of $(1-r^2)^{3/2} F(3/2, 3/2; \frac{1}{2}n+3/2; r^2)$ in the allowable range $0 \leq r^2 \leq 1$, and the integral over expression (90b) is merely a constant multiple of the convergent integral in eqns. (87). Hence the alternative distribution can be normalised for $n > 3$, which leaves its status open.

Next, consider the behaviour of the distribution (86d) for $n = 3$. As before, $(1-r^2)^{3/2} F(3/2, 3/2; \frac{1}{2}n+3/2; r^2)$ is finite and non-negative for all values of r^2 other than 1; however, the condition for $F(a, b; c; 1)$ to be finite, that $(c-a-b)$ has a real part greater than zero and the right-

hand side of eqn. (90a) is finite, now fails, since $(\frac{1}{2}n+3/2) - 3/2 - 3/2 = 0$ for $n = 3$. However, for $n = 3$, the relevant function is $F(a,b;(a+b);r^2)$ for $a=b=3/2$; and according to Copson [5] and Erdelyi [7],

$$\lim_{x \rightarrow 1^-} \left[\frac{F(a,b;a+b;x)}{-\log(1-x)} \right] = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$$

so, since r'^2 is real and non-negative,

$$\begin{aligned} & \lim_{r'^2 \rightarrow 1^-} \left[(1-r'^2)^{\frac{3}{2}} F\left(\frac{3}{2}, \frac{3}{2}; 3; r'^2\right) \right] \\ & \lim_{r'^2 \rightarrow 1^-} \left[-\log(1-r'^2) \cdot (1-r'^2)^{\frac{3}{2}} \cdot \frac{\Gamma(3)}{\Gamma(\frac{3}{2})\Gamma(\frac{3}{2})} \right] \\ & = 0 \end{aligned}$$

It follows that $(1-r'^2)^{3/2} F(3/2, 3/2; 3; r'^2)$ is finite and non-negative for all allowable values of r'^2 including 1; the rest of the argument of the last paragraph then applies, including the bounding expression (90b) for the integrand of the normalising integral; and so the alternative distribution can also be normalised for $n = 3$, and again its status is open.

Finally, however, I consider the distribution (86d) for $n = 2$; or, to be precise, I consider the form obtained by combining eqns. (68) with eqn. (86b) instead of eqn. (86c), which is (for $n = 2$)

$$\langle P_2(d\lambda' d\mu' | \theta) \rangle \propto d\lambda' d\mu' ((\lambda'^2+1)(\mu'^2+1))^{-1} F(1, 1; \frac{5}{2}; r'^2)$$

The hypergeometric function here satisfies the conditions for eqn. (90a) to be valid at $r'^2 = 1$, and it is geometrically convergent everywhere else. Hence, when eqn. (90e) is integrated, the integrand is everywhere non-negative and is bounded above by

$$((\lambda'^2+1)(\mu'^2+1))^{-1} G'_2$$

where G'_2 is the greatest value of $F(1,1;(5/2);r'^2)$ in the allowable range $0 \leq r'^2 \leq 1$ (actually at the upper limit, but it is not necessary to notice this). Hence the integral over expression (90e) converges, because the integral over expression (90f) does so (on making the substitutions (87e) and (87f) for λ' and similarly for μ'). It follows that the alternative distribution of probability, based on the simple prior probability distribution $(d\rho)$ for correlation, can be normalised for $n = 2$: and therefore must be rejected, since this is wrong by the closing arguments of the last section. This does not, of course, prove that the calculated prior probability distribution $((1-\rho^2)^{-3/2}d\rho)$ is correct – the arguments used to derive it must be held to do that – but it does definitely disprove the simple plausible distribution (which Jeffreys [1] seems to prefer, in spite of presenting himself most of the arguments I have used to prove the more complicated form). Accordingly, I assert on the basis of the arguments presented above that eqn. (88), taken together with the definitions (57) and (52), gives the joint probability distribution of the unknown means λ and μ of measurements conforming to the normal correlation distribution, when the information about the underlying distribution is comprised in a sample θ . I consider the consequences and applications of this new distribution in the final Part of this paper, which now follows.

PART IV

The new distribution's asymptotic form and general shape

I consider first the behaviour of eqn. (88) for large n . As with the original "Student's t " distribution (compare eqn. (23c)), I can write for large n that $(1+x)^{-\frac{1}{2}n} \simeq \exp(-\frac{1}{2}nx)$, since these two expressions agree if $|x| \ll 1$ – and if this condition is not satisfied, the "probability density" in eqn. (88) is relatively negligible. Then eqn. (88) becomes (recalling the definitions (57))

$$P_n(d\lambda' d\mu'|\theta) \simeq \frac{n}{2\pi st(1-r^2)^{\frac{1}{2}}} \exp \left[-\frac{n}{2(1-r^2)} [\lambda'^2 + \mu'^2 - 2r\lambda'\mu'] \right] d\lambda' d\mu'$$

$$\simeq \frac{n}{2\pi st(1-r^2)^{\frac{1}{2}}} \exp \left[-\frac{n}{2(1-r^2)} \left[\frac{(\lambda-\bar{x})^2}{s^2} + \frac{2r(\lambda-\bar{x})(\mu-\bar{y})}{st} + \frac{(\mu-\bar{y})^2}{t^2} \right] \right] d\lambda d\mu \quad (91)$$

which is simply a modified form of the original normal correlation distribution (eqn. (51a)) from which the "likelihood" was derived at the end of Part II; the modifications consist of replacing each underlying distribution parameter by its corresponding sample parameter and then contracting the result in both dimensions by the familiar "sampling factor" of $\sqrt{n}$ (as in the one-dimensional case (compare the comments on eqns. (23a) and (29c)). This is another verification of consistency, apart from its direct application to a large sample.

Next, I consider the shape of the distribution. The co-factor of $(d\lambda d\mu)$ or $(d\lambda' d\mu')$ in eqns. (88) or (91), which is often called the probability density, is constant whenever the values of λ and μ , or λ' and μ' , are such that

$$\lambda'^2 + \mu'^2 - 2r\lambda'\mu' \quad \text{or} \quad \frac{(\lambda-\bar{x})^2}{s^2} + \frac{(\mu-\bar{y})^2}{t^2} - \frac{2r(\lambda-\bar{x})(\mu-\bar{y})}{st}$$

is constant. A set of values satisfying such a relation defines an ellipse in the (λ, μ) -plane, and varying the constant produces a family of ellipses which are similar (all have the same aspect ratio), similarly-situated (all have the same inclination to the axes of λ and μ) and concentric (all are centred on the point $\lambda' = 0 \mid \mu' = 0$ or $\lambda = \bar{x} \mid \mu = \bar{y}$, which is both the maximum and the centre of symmetry of the probability density function). From these properties (particularly the last one) it follows that the best estimate of the mean values of the underlying distribution is, as is natural, $\lambda = \bar{x}$ and $\mu = \bar{y}$. I will now verify these elliptical properties and discuss their relation to the transformations examined earlier in Part II.

Probability ellipses and changes of axes

From here on I shall abandon the use of λ' and μ' according to the definitions (57), because I wish to revert to their use in Part II (in the text surrounding eqns. (48) and (49)) to denote the results of rotations of axes. I therefore begin by inserting the definitions (57) into eqn. (88) to obtain my new distribution in the form

$$P_n(d\lambda d\mu|\theta) = \frac{(n-2)}{2\pi st(1-r^2)^{\frac{1}{2}}} \frac{1}{(1-r^2)} \left[\frac{(\lambda-\bar{x})^2}{s^2} + \frac{(\mu-\bar{y})^2}{t^2} - \frac{2r(\lambda-\bar{x})(\mu-\bar{y})}{st} \right]^{-\frac{1}{2}n} d\lambda d\mu \quad (92)$$

Now I shift the origin of co-ordinates to the centre of the distribution at the point $(\bar{x}, \bar{y})$, then in an obvious notation the curves of constant probability density are given by the family of equations

$$\left[\frac{(\Delta\lambda)^2}{s^2} + \frac{(\Delta\mu)^2}{t^2} - \frac{2r(\Delta\lambda)(\Delta\mu)}{st} \right] = \text{a constant (say } C)$$

or, in matrix and vector form,

$$(\Delta\lambda \quad \Delta\mu) \begin{bmatrix} s^{-2} & -r(st)^{-1} \\ -r(st)^{-1} & t^{-2} \end{bmatrix} (\Delta\lambda \quad \Delta\mu)^T = C \quad (93b)$$

On comparing this with eqns. (48b) and following, it is immediately obvious that the axes of $(\Delta\lambda, \Delta\mu)$ can be rotated through an angle ϕ which satisfies eqn. (48e), thus (provided the ellipse is not a circle, in which case ϕ is indeterminate)

$$\tan(2\phi) = 2g_X/(g_{11}-g_{22}) = -2r(st)^{-1}/(s^{-2}-t^{-2})$$

$$2rst/(s^2-t^2)$$

and (by analogy with eqns. (48k) and (48l)) the corresponding new co-ordinates are

$$\Delta\lambda' = (\cos\phi) \Delta\lambda + (\sin\phi) \Delta\mu$$

$$\Delta\mu' = (-\sin\phi) \Delta\lambda + (\cos\phi) \Delta\mu \quad (93e)$$

with an anti-clockwise rotation from $(\Delta\lambda, \Delta\mu)$ to $(\Delta\lambda', \Delta\mu')$. Eqn. (93c) has four solutions within the standard range of angles, spaced 90° apart; these correspond to aligning positive $\Delta\lambda'$ (say) with each of the four semi-axes of the ellipse in turn. (A condition for selecting the semi-major pair will be given shortly.) To prove that eqn. (93a) actually is an ellipse, I note that eqn. (93b) in the primed co-ordinates is simply (from eqn. (48j); note that C is unchanged)

$$(\Delta\lambda' \quad \Delta\mu') \begin{bmatrix} s'^{-2} & 0 \\ 0 & t'^{-2} \end{bmatrix} (\Delta\lambda' \quad \Delta\mu')^T = C$$

or

$$\frac{(\Delta\lambda')^2}{s'^2} + \frac{(\Delta\mu')^2}{t'^2} = C$$

which is indeed an ellipse, centred at the origin (of both primed and unprimed co-ordinates), with its semi-axes aligned with the axes of the primed co-ordinates and of lengths $s'\sqrt{C}$ and $t'\sqrt{C}$ (e.g. Jeffreys and Jeffreys [3]), where it is natural to choose the $\Delta\lambda'$ -direction as the major axis. The simplest way to determine s' and t' is by noting (from Wilkinson [4]) that the rotation from unprimed to primed co-ordinates does not change the eigenvalues of the embedded (2×2) -matrix (in the notation of eqns. (48), G_{ik} and $(R^T G_{ik} R)$ have the same eigenvalues); hence the two matrices

$$\begin{bmatrix} s^{-2} & -r(st)^{-1} \\ -r(st)^{-1} & t^{-2} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} s'^{-2} & 0 \\ 0 & t'^{-2} \end{bmatrix} \quad (93h)$$

have the same eigenvalues. (Making the choice of s' as the semi-major axis now requires (using eqns. (48d) and (93c)) that $\sin(2\phi)$ has the same sign as r .) It is also known (e.g. [4]) that the sum of the eigenvalues of a matrix is equal to the sum of its diagonal elements, and that the product of the eigenvalues of a matrix is equal to its determinant; hence the sums of diagonal elements of the matrices (93h) are equal, and so are their determinants. These equalities give, respectively,

$$s'^{-2} + t'^{-2} = s^{-2} + t^{-2}$$

and

$$s'^{-2}t'^{-2} = s^{-2}t^{-2}(1-r^2)$$

then s'^{-2} and t'^{-2} are the roots of the quadratic equation

$$X^2 - (s'^{-2} + t'^{-2})X + (s'^{-2}t'^{-2}) = 0 \quad (93k)$$

which, from eqns. (93i) and (93j), is

$$X^2 - (s^{-2} + t^{-2})X + s^{-2}t^{-2}(1-r^2) = 0$$

This can be solved by the well-known formula. I note that, since the right-hand side of eqn. (93j) is strictly positive ($r^2 < 1$), the product of s'^{-2} and t'^{-2} is strictly positive, so they are both strictly positive or both strictly negative; but if they were both strictly negative, the right-hand side of eqn. (93i) would be strictly negative, which is impossible; hence they are both strictly positive. This, of course, is as it should be. It is also interesting to consider whether the ellipse can degenerate into a circle, for which the condition is $s'^{-2} = t'^{-2}$ or the two roots of eqn. (93k) being equal. This can only happen if the quadratic equation's discriminant vanishes, that is if $(b^2 - 4ac) = 0$ when eqn. (93k) is rewritten as $(aX^2 + bX + c) = 0$, or

$$(s^{-2} + t^{-2})^2 - 4(s^{-2}t^{-2}(1-r^2)) = 0$$

$$(s^{-2} - t^{-2})^2 + 4s^{-2}t^{-2}r^2 \quad (93m)$$

which requires both $(s^{-2} = t^{-2})$, or $s = t$, and $r = 0$. (This double condition makes eqn. (93c) break down, as stated in the text preceding it.) It follows that the probability density cannot be transformed into a uniform circular distribution unless it is already distributed that way, which again is physically sensible.

By these considerations I have shown that the probability density is distributed in such a way that the curves of equal density are ellipses and are all centred at the shifted origin $(\bar{x}, \bar{y})$. These ellipses are also similar (since the ratio $s'\sqrt{C}:t'\sqrt{C}$ is the same for all C and depends only on s , t and r through eqn. (93k)) and are similarly-situated (since ϕ likewise depends only on s , t and r through eqn. (93c), and not on C). It remains to discuss how to calculate the probability that the unknown means of the underlying distribution lie in any particular region of the (λ, μ) -plane. For regions of arbitrary shape this can only be done by numerical integration of eqn. (92), about which there is really nothing more to be said. However, random uncertainties are usually assessed with reference to simple regions centred on the best estimates of the measurands (here $\bar{x}$ and $\bar{y}$). There are two plausible choices of region in the present case: an ellipse of constant probability

density (with orientation and aspect ratio determined from eqns (93)) or a circle (ignoring the angular variation of the probability density). I discuss the circular region first, as this is currently used for uncertainty assessment.

Probability within a circular region

To specify the probability that the true means (λ, μ) lie within a radius η_P of the estimated means $(\bar{x}, \bar{y})$, I introduce polar co-ordinates in the form

$$\lambda = \bar{x} + \eta \cos \psi$$

$$\mu = \bar{y} + \eta \sin \psi$$

with the usual two-dimensional differential form

$$d\lambda d\mu \equiv d(\lambda - \bar{x}) d(\mu - \bar{y}) = \eta d\eta d\psi \quad (0 \leq \psi \leq 2\pi, \eta \geq 0)$$

Then, from eqn (92),

$$P_n(\eta \leq \eta_P) \equiv \int_0^{\eta_P} \int_0^{2\pi} P_n(d\eta d\psi | \theta) \quad \left\{ \begin{array}{l} \int_0^{2\pi} \int_0^{\eta_P} \frac{(n-2)}{2\pi s t (1-r^2)^{\frac{1}{2}}} \left[1 + \frac{1}{(1-r^2)} \left[\frac{\cos^2 \psi}{s^2} + \frac{\sin^2 \psi}{t^2} - 2r \frac{\sin \psi \cos \psi}{st} \right] \right]^{\frac{1}{2}(n-2)} \eta d\eta d\psi \end{array} \right. \quad (95)$$

The simplest representation of this is achieved by performing the integration over η , which can be done at sight after noticing that $(\eta d\eta) = \frac{1}{2} d(\eta^2)$; then (since $n \geq 3$)

$$P_n(\eta \leq \eta_P) = \int_0^{2\pi} \left[1 + \frac{1}{(1-r^2)} \left[\frac{\cos^2 \psi}{s^2} + \frac{\sin^2 \psi}{t^2} - 2r \frac{\sin \psi \cos \psi}{st} \right] \right]^{\frac{1}{2}(n-2)} d\psi \quad (96a)$$

For purely numerical purposes this is quite satisfactory as it stands: that is to say, a good numerical integration routine will easily calculate it for given η_P , s , t and r . It is equally easy to calculate the complementary form. If η_P is allowed to tend to infinity, the region of integration extends to the whole (λ, μ) or (η, ψ) plane, and since the distribution (92) has been normalised, it follows immediately that

$$1 = P_n(\eta \leq \infty) = \int_0^{2\pi} \frac{(1-r^2)^{\frac{1}{2}}}{2\pi st} \cdot \frac{1}{\left[\frac{\cos^2 \psi}{s^2} + \frac{\sin^2 \psi}{t^2} - 2r \frac{\sin \psi \cos \psi}{st} \right]} d\psi$$

and so

$$P_n(\eta \geq \eta_P) = 1 - P_n(\eta \leq \eta_P)$$

$$\int_0^{2\pi} \frac{(1-r^2)^{\frac{1}{2}}}{2\pi st} \frac{\left\{ 1 + \frac{1}{(1-r^2)} \left[\eta_P^2 \left(\frac{\cos^2 \psi}{s^2} + \frac{\sin^2 \psi}{t^2} - 2r \frac{\sin \psi \cos \psi}{st} \right) \right] \right\}^{-\frac{1}{2}(n-2)}}{\left[\frac{\cos^2 \psi}{s^2} + \frac{\sin^2 \psi}{t^2} - 2r \frac{\sin \psi \cos \psi}{st} \right]} d\psi$$

Eqn (96b) can be directly verified in a suggestive way, thus:

$$\begin{aligned} & \frac{(1-r^2)^{\frac{1}{2}}}{2\pi st} \int_0^{2\pi} \left[\frac{\cos^2 \psi}{s^2} + \frac{\sin^2 \psi}{t^2} - 2r \frac{\sin \psi \cos \psi}{st} \right]^{-1} d\psi \\ & \frac{(1-r^2)^{\frac{1}{2}}}{2\pi st} \int_0^{2\pi} \left[\frac{1 + \cos 2\psi}{s^2} + \frac{-\cos 2\psi}{t^2} - \frac{r \sin 2\psi}{st} \right]^{-1} d\psi \\ & \frac{(1-r^2)^{\frac{1}{2}}}{2\pi st} \int_0^{2\pi} \left[\frac{+\cos \psi}{s^2} + \frac{1 - \cos \psi}{t^2} - \frac{r \sin \psi}{st} \right]^{-1} d\psi \end{aligned} \quad (96d)$$

which admits the standard substitution $z = \exp(i\psi)$, $dz = iz d\psi$ with $i = \sqrt{-1}$ (e.g. Jeffreys and Jeffreys [3], Copson [5]) reducing it to the integral of a rational function of z – the reciprocal of a quadratic polynomial, in general – around the unit circle in the complex z plane. On rejecting the unphysical possibilities $s=0$, $t=0$, $r^2=1$, the integrand is found to have one simple pole inside the unit circle and its other pole outside, so the contour integration is elementary. However, if $s=t$ and $r=0$, giving the uniform circular distribution mentioned after eqn (93m), the rational function integrand degenerates to the reciprocal of z , and the other eqns (96) degenerate also, giving

$$P_n(\eta \leq \eta_P) = \frac{\eta_P^2}{(s \& t)^2}^{-\frac{1}{2}(n-2)}$$

where $(s \& t)$ denotes the common value of s and t in this special case. An exact form like this will appear later when I consider elliptical probability regions. In the meantime, I note regretfully that

this contour integration technique is not useful for general values of η_P , s , t and r , because the integrand in eqn (96a) has a branch point for odd n , and for large even n it is prohibitively complicated to determine recursively the residues of the corresponding rational function's poles; likewise the possible recursive representation in terms of complete elliptic integrals is impracticably complicated to handle. However, if η_P is not too large relative to s and t , it is possible to obtain a result of some value by expanding the numerator of the integrand in eqn (96a) as a power series in (η_P^2) and then integrating term by term, as in the next section.

Probability within a circular region of moderate radius

I define an auxiliary parameter h by

$$h \equiv \left((t^2 - s^2)^2 + (2rst)^2 \right)^{1/4}$$

noting that $(t^2 - s^2)^2 + (2rst)^2 = (t^2 + s^2)^2 - 4s^2t^2(1 - r^2)$ and so, for all physically realistic s , t and r , $0 \leq h < (t^2 + s^2)^{1/2}$, the lower bound being only attainable for the uniform circular distribution with $s=t$ and $r=0$. I also introduce an auxiliary angle ψ_0 defined by

$$h \cos \psi_0 = (t^2 - s^2)$$

$$h \sin \psi_0 = 2rst$$

(With these definitions it follows from eqns (93c) and (93h) that $(2\phi) = \pi - \psi_0$; see the Appendix.) Then

$$\begin{aligned} & \frac{\cos^2 \psi}{s^2} + \frac{\sin^2 \psi}{t^2} - 2r \frac{\sin \psi \cos \psi}{st} = \frac{1 + \cos 2\psi}{2s^2} + \frac{1 - \cos 2\psi}{2t^2} - \frac{r \sin 2\psi}{st} \\ & \left(\frac{1}{2s^2} + \frac{1}{2t^2} \right) + \cos 2\psi \left(\frac{1}{2s^2} - \frac{1}{2t^2} \right) - \sin 2\psi \left(\frac{r}{st} \right) \\ & = (2s^2t^2)^{-1} \left[(t^2 + s^2) + \{ \cos 2\psi (t^2 - s^2) - \sin 2\psi (2rst) \} \right] \\ & = (2s^2t^2)^{-1} \left[(t^2 + s^2) + \{ \cos 2\psi (h \cos \psi_0) - \sin 2\psi (h \sin \psi_0) \} \right] \\ & = (2s^2t^2)^{-1} \left[(t^2 + s^2) + h \cos (2\psi + \psi_0) \right] \end{aligned} \quad (97d)$$

using eqns (97b) and (97c). Hence, from eqn (96c), on introducing for later convenience the definition

$$\overline{\eta_P} \equiv \eta_P^2$$

it follows that

$$\begin{aligned} P_n(\eta^2 \leq \overline{\eta_P}) &= 1 - \frac{(1-r^2)^{\frac{1}{2}}}{2\pi st} \int_0^{2\pi} \frac{\left\{ 1 + \frac{\overline{\eta_P} [(t^2+s^2) + h \cos(2\psi+\psi_0)]}{2s^2t^2(1-r^2)} \right\}^{(1-\frac{1}{2}n)}}{(2s^2t^2)^{-1} [(t^2+s^2) + h \cos(2\psi+\psi_0)]} d\psi \\ &= \frac{1}{2\pi st(1-r^2)^{\frac{1}{2}}} \int_0^{2\pi} \frac{\left\{ 1 + \frac{\overline{\eta_P} [(t^2+s^2) + h \cos(\psi+\frac{1}{2}\psi_0)]}{2s^2t^2(1-r^2)} \right\}^{(1-\frac{1}{2}n)}}{(2s^2t^2(1-r^2))^{-1} [(t^2+s^2) + h \cos(\psi+\frac{1}{2}\psi_0)]} d\psi \\ &= 1 - \frac{1}{2\pi st(1-r^2)^{\frac{1}{2}}} \int_0^{2\pi} \frac{\left\{ 1 + \frac{\overline{\eta_P} [(t^2+s^2) + h \cos\psi]}{2s^2t^2(1-r^2)} \right\}^{(1-\frac{1}{2}n)}}{(2s^2t^2(1-r^2))^{-1} [(t^2+s^2) + h \cos\psi]} d\psi \end{aligned}$$

Writing $P_n(\eta^2 \leq \overline{\eta_P})$ as the functional form $\overline{P_n}(\overline{\eta_P})$ for convenience, I can immediately expand this in a Taylor-Maclaurin series as

$$\overline{P_n}(\overline{\eta_P}) = \sum_{k=1}^{\infty} \overline{P_n}'^{(k)}(0) \frac{\overline{\eta_P}^k}{k!}$$

where $\overline{P_n}'^{(k)}$ denotes the k^{th} derivative of $\overline{P_n}(\overline{\eta_P})$ with respect to $\overline{\eta_P}$, $\overline{P_n}'^{(k)}(0)$ is that derivative's value at $\overline{\eta_P} = 0$, and the series starts at $k=1$ because $\overline{P_n}(0)$ is zero from eqn (96a) (or indeed from common sense). The derivatives may be evaluated by differentiating k times under the integral sign in eqn (98a) and then setting $\overline{\eta_P} = 0$ (the resulting integrands are all everywhere finite, since $h < (t^2+s^2)$ and so the denominators never vanish); the result of this procedure is

$$\begin{aligned} \overline{P_n}'^{(k)}(0) &= - \frac{1}{2\pi st(1-r^2)^{\frac{1}{2}}} \int_0^{2\pi} \frac{[(t^2+s^2) + h \cos\psi]^{(k-1)} (1-\frac{1}{2}n)(1-\frac{1}{2}n-1)\dots(1-\frac{1}{2}n-(k-1))}{(2s^2t^2(1-r^2))^{(k-1)}} d\psi \\ &= \frac{(-1)^{(k+1)} (\frac{1}{2}n+k-2)(\frac{1}{2}n+k-3)\dots(\frac{1}{2}n-1)}{(st(1-r^2)^{\frac{1}{2}})^k} \frac{1}{2\pi} \int_0^{2\pi} \left\{ \frac{(t^2+s^2) + h \cos\psi}{2st(1-r^2)^{\frac{1}{2}}} \right\}^{(k-1)} d\psi \quad (98c) \end{aligned}$$

The expression in the brace brackets here may be written as

$$\frac{1}{2\pi} \int_0^{2\pi} [\alpha + (\alpha^2 - 1)^{\frac{1}{2}} \cos \psi]^{(k-1)} d\psi \quad (98d)$$

using the definition (97a), where obviously

$$\alpha \equiv (t^2 + s^2) / (2st(1 - r^2)^{\frac{1}{2}})$$

and, according to Jeffreys and Jeffreys [3] or Copson [5], expression (98d) is the Legendre polynomial of order $(k-1)$ and argument α , which I write (following [3]) as $p_{k-1}(\alpha)$ (the customary notation $P_{k-1}(\alpha)$ would be too confusing here). Then, since $p_{k-1}(\alpha) = (-1)^{(k-1)} p_{k-1}(-\alpha)$ (from [3] or [5]), eqn (98c) becomes (using also the Γ -function recurrence, eqn (71e))

$$\begin{aligned} P_n^{(k)}(0) &= \frac{(-1)^{(k+1)} (\frac{1}{2}n+k-2)(\frac{1}{2}n+k-3)\dots(\frac{1}{2}n-1)}{(st(1-r^2)^{\frac{1}{2}})^k} p_{k-1} \left[\frac{(t^2+s^2)}{2st(1-r^2)^{\frac{1}{2}}} \right] \\ &= \frac{(\frac{1}{2}n+k-2)(\frac{1}{2}n+k-3)\dots(\frac{1}{2}n-1)}{(st(1-r^2)^{\frac{1}{2}})^k} p_{k-1} \left[\frac{-(t^2+s^2)}{2st(1-r^2)^{\frac{1}{2}}} \right] \\ &= \frac{\Gamma(\frac{1}{2}n+k-1)}{\Gamma(\frac{1}{2}n-1)} (st(1-r^2)^{\frac{1}{2}})^{-k} p_{k-1} \left[\frac{t^2+s^2}{2st(1-r^2)^{\frac{1}{2}}} \right] \end{aligned}$$

and so, from eqns (98b) and (97e),

$$P_n(\eta \leq \eta_P) = \sum_{k=1}^{\infty} \frac{\Gamma(\frac{1}{2}n+k-1)}{k! \Gamma(\frac{1}{2}n-1)} p_{k-1} \left[\frac{t^2+s^2}{2st(1-r^2)^{\frac{1}{2}}} \right] \left[\frac{\eta_P^2}{st(1-r^2)^{\frac{1}{2}}} \right]^k$$

The Legendre polynomials all have the same argument and may be evaluated using the (apparently stable) recurrence relation (from [3])

$$(k+1) p_{k+1}(-\alpha) + (2k+1) \alpha p_k(-\alpha) + k p_{k-1}(-\alpha) = 0 \quad (98h)$$

using the notation of the definition (98e), with the starting values (also from [3])

$$p_0(-\alpha) = \quad (98i)$$

$$p_1(-\alpha) = -\alpha$$

Unfortunately this result is not very useful. As remarked at the end of the last section, it depends ultimately on expanding the brace-bracketed part of the integrand of eqns (96a) or (98a) as a binomial series in powers of

$$\eta_P^2 [(t^2+s^2) + h \cos \psi] / (2s^2 t^2 (1-r^2)) \quad (98k)$$

and the series converges only if expression (98k) is less than 1 throughout the range of ψ ; using

the definition of h , it follows that η_P must be less than the smaller of s and t even when r is zero, and must be much smaller still when r^2 is nearly 1. A more fundamental objection is that this approach assumes that it is appropriate to calculate $P_n(\eta \leq \eta_P)$ given η_P , s , t and r , whereas uncertainty estimates are normally made by calculating η_P from given s , t and r and a specified value of $P_n(\eta \leq \eta_P)$ – typically 0.95 or "95%". In the next section I attack this problem by inverting the order of integration in eqn (95).

Probability within a circular region of arbitrary radius

I now rewrite eqn (95) in the form

$$P_n(\eta \leq \eta_P) = \int_0^{\eta_P} d\eta \, \eta \left[\int_0^{2\pi} \frac{(n-2)}{2\pi st(1-r^2)^{\frac{1}{2}}} \left(1 + \frac{\eta^2}{(1-r^2)} \left[\frac{\cos^2\psi}{s^2} + \frac{\sin^2\psi}{t^2} - 2r \frac{\sin\psi \cos\psi}{st} \right] \right)^{-\frac{1}{2}n} d\psi \right] \quad (99a)$$

where from eqn (97d)

$$\begin{aligned} \left[\frac{\cos^2\psi}{s^2} + \frac{\sin^2\psi}{t^2} - 2r \frac{\sin\psi \cos\psi}{st} \right] &= (2s^2t^2)^{-1} ((t^2+s^2) + h \cos(2\psi+\psi_0)) \\ &= (2s^2t^2)^{-1} ((t^2+s^2) + h (\cos^2(\psi+\frac{1}{2}\psi_0) - \sin^2(\psi+\frac{1}{2}\psi_0))) \\ &= (2s^2t^2)^{-1} ((t^2+s^2+h) - 2h \sin^2(\psi+\frac{1}{2}\psi_0)) \end{aligned}$$

Hence the square-bracketed inner integral part of eqn (99a) may be written as

$$\begin{aligned} & \frac{(n-2)}{2\pi st(1-r^2)^{\frac{1}{2}}} \int_0^{2\pi} \frac{2s^2t^2(1-r^2) + \eta^2(t^2+s^2+h) - 2\eta^2h \sin^2(\psi+\frac{1}{2}\psi_0)}{2s^2t^2(1-r^2)}^{-\frac{1}{2}n} d\psi \\ &= \frac{(n-2)}{2\pi st(1-r^2)^{\frac{1}{2}}} + \frac{\eta^2(t^2+s^2+h)}{2s^2t^2(1-r^2)} \int_0^{2\pi} \left(1 - \frac{2\eta^2h \sin^2(\psi+\frac{1}{2}\psi_0)}{2s^2t^2(1-r^2) + \eta^2(t^2+s^2+h)} \right)^{-\frac{1}{2}n} d\psi \\ &= \frac{(n-2)}{2\pi st(1-r^2)^{\frac{1}{2}}} \left(1 + \frac{\eta^2(t^2+s^2+h)}{2s^2t^2(1-r^2)} \int_0^{2\pi} \left(1 - \frac{2\eta^2h \sin^2\psi}{2s^2t^2(1-r^2) + \eta^2(t^2+s^2+h)} \right)^{-\frac{1}{2}n} d\psi \right) \\ &= \frac{(n-2)}{\pi st(1-r^2)^{\frac{1}{2}}} \left(1 + \frac{\eta^2(t^2+s^2+h)}{2s^2t^2(1-r^2)} \int_0^{\frac{1}{2}\pi} \left(1 - \frac{2\eta^2h \sin^2\psi}{2s^2t^2(1-r^2) + \eta^2(t^2+s^2+h)} \right)^{-\frac{1}{2}n} 2d\psi \right) \end{aligned}$$

In the last integral here I make the substitutions (after eqns (70))

$$w = \sin^2 \psi \quad (99d)$$

$$dw = 2 \sin \psi \cos \psi \, d\psi$$

with $(0 \leq w \leq 1)$, which transform the integral into the form

$$\int_0^1 dw \, w^{-\frac{1}{2}} (1-w)^{-\frac{1}{2}} \left[1 - \frac{2\eta^2 h w}{2s^2 t^2 (1-r^2) + \eta^2 (t^2 + s^2 + h)} \right]^{-\frac{1}{2}n} \quad (99e)$$

Using the binomial expansion in reverse (compare eqn (74b)), I can now write

$$\begin{aligned} \frac{2\eta^2 h w}{2s^2 t^2 (1-r^2) + \eta^2 (t^2 + s^2 + h)} &= F \left[\frac{1}{2}n, \frac{1}{2}n; \frac{1}{2}n; \frac{2\eta^2 h w}{2s^2 t^2 (1-r^2) + \eta^2 (t^2 + s^2 + h)} \right] \\ &= F \left[\frac{1}{2}n, c; c; \frac{2\eta^2 h w}{2s^2 t^2 (1-r^2) + \eta^2 (t^2 + s^2 + h)} \right] \end{aligned}$$

on recalling the argument preceding eqn (79). I now choose $c = \frac{1}{2}$ in this last hypergeometric expression, and substitute back into the integral expression (99f) which then becomes

$$\int_0^1 dw \, w^{-\frac{1}{2}} (1-w)^{-\frac{1}{2}} F \left[\frac{1}{2}n, \frac{1}{2}; \frac{1}{2}; \left[\frac{2\eta^2 h}{2s^2 t^2 (1-r^2) + \eta^2 (t^2 + s^2 + h)} \right] w \right] \quad (99h)$$

Now I can compare this form with the integral expression (72a), with which I can identify it by choosing the parameters in expression (72a) to be $z = 2\eta^2 h / (2s^2 t^2 (1-r^2) + \eta^2 (t^2 + s^2 + h))$, $\gamma = \frac{1}{2}$, $a = \frac{1}{2}n$, $b = \frac{1}{2}$ and $c = 1$. The conditions stated after expression (72a) are satisfied; clearly $c > \gamma > 0$, and $|z| < 1$ because (with the usual physical restrictions) $\eta^2 \geq 0$, $h \geq 0$, $(t^2 + s^2) > h$ and $(2s^2 t^2 (1-r^2)) > 0$, so that $0 \leq z < 1$. Hence, from expressions (72a) and (72e), the value of expression (99h) is

$$\frac{\Gamma(\frac{1}{2})\Gamma(\frac{1}{2})}{\Gamma(1)} F \left[\frac{1}{2}n, \frac{1}{2}; 1; \frac{2\eta^2 h}{2s^2 t^2 (1-r^2) + \eta^2 (t^2 + s^2 + h)} \right]$$

and since $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ and $\Gamma(1) = 1$, substituting expression (99i) back into eqn (99c) gives

$$\frac{(n-2)}{st(1-r^2)^{\frac{1}{2}}} \left[1 + \frac{\eta^2 (t^2 + s^2 + h)}{2s^2 t^2 (1-r^2)} \right]^{-\frac{1}{2}n} F \left[\frac{1}{2}n, \frac{1}{2}; 1; \frac{2\eta^2 h}{2s^2 t^2 (1-r^2) + \eta^2 (t^2 + s^2 + h)} \right]$$

The principal argument of the last hypergeometric function (z , in the notation of the last paragraph) is strictly less than 1 in magnitude, so in principle the function can be evaluated from its series expansion (71d), which reduces eqn (99a) to a numerical integration over η ; but, if n is large and η is not small, the convergence of the series is formidably slow, because the first parameter contributes to the k^{th} term a factor $(\frac{1}{2}n)(\frac{1}{2}n+1)\dots(\frac{1}{2}n-1+k)$. If n is not very large, this

difficulty may be dealt with by using the Gaussian recurrence relation (e.g. Copson [5], Erdelyi [7]; compare eqn (98h))

$$\begin{aligned} k(z-1)F(k+1, \frac{1}{2}; 1; z) + ((2k-1) - (k-\frac{1}{2})z)F(k, \frac{1}{2}; 1; z) \\ + (1-k)F(k-1, \frac{1}{2}; 1; z) = 0 \end{aligned} \quad (99k)$$

which allows $F(\frac{1}{2}n, \frac{1}{2}; 1; z)$ to be built up from $F(1, \frac{1}{2}; 1; z)$ (if n is even) or from $F(\frac{1}{2}, \frac{1}{2}; 1; z)$ and $F(-\frac{1}{2}, \frac{1}{2}; 1; z)$ (if n is odd). The last two hypergeometric functions may be evaluated from the series expansion (71d) with reasonable facility, provided the values of s , t and r are not close to extreme cases; while the function $F(1, \frac{1}{2}; 1; z)$ is $(1-z^2)^{-1}$, from the discussion preceding eqn (79). (As in the previous section, the recurrence relation (99k) appears to be stable.) The square-bracketed inner integral part of eqn (99a) may therefore be regarded as determined by eqn (99j), and so the value of η_P for given s , t and r and specified $P_n(\eta \leq \eta_P)$ can be determined by successive approximations in eqn (99a), provided n is not too large for the recurrence relation (99k) to be conveniently applied; the most logical choice of successive-approximation method seems to be the well-known Newton-Raphson algorithm, since the derivative of $P_n(\eta \leq \eta_P)$ with respect to η_P is simply the integrand in eqn (99a) evaluated at the upper limit of integration (η_P).

If n is very large, eqn (99a) approaches a kind of Gaussian limit, as the one-dimensional "Student's t " distribution does. Suppose that $\eta_P^2 < 2s^2t^2(1-r^2)/(nh)$; this is not as restrictive as it seems, provided that r^2 is not very close to 1, as will become clear. Then the hypergeometric function in eqn (99j) approaches its degenerate or confluent form, and I can write (e.g. Copson [5], Erdelyi [7]) for any z of order (n^{-1}) or smaller

$$F(\frac{1}{2}n, \frac{1}{2}; 1; z) = {}_1F_1(\frac{1}{2}; 1; \frac{1}{2}nz) + O(n^{-1}) \quad (100a)$$

and similarly

$$+ \frac{\eta^2(t^2+s^2+h)}{2s^2t^2(1-r^2)} \exp \left[-\frac{1}{2}nz \right] = \exp \left[-\frac{n\eta^2(t^2+s^2+h)}{4s^2t^2(1-r^2)} \right] + O(n^{-1}) \quad (100b)$$

and from [5] or Jeffreys and Jeffreys [3]

$${}_1F_1(\frac{1}{2}; 1; \frac{1}{2}nz) = \exp(\frac{1}{2}nz) I_0(\frac{1}{2}nz) \quad (100c)$$

where I_0 is the modified Bessel function of the first kind and of order 0. The required z here is (from eqn (99j)) $2\eta^2h/(2s^2t^2(1-r^2)+\eta^2(t^2+s^2+h))$, which for all η in the range 0 to η_P may be written (using the restriction on η_P^2 in terms of n) as $\eta^2h/(s^2t^2(1-r^2))$ + relative error of (n^{-1}) , provided that s and t are not nearly equal. Then eqn (99j) reduces, with relative error of (n^{-1}) , to

$$\frac{n}{st(1-r^2)^{\frac{1}{2}}} \exp \left[-\frac{n\eta^2(t^2+s^2)}{4s^2t^2(1-r^2)} \right] I_0 \left[\frac{n\eta^2h}{4s^2t^2(1-r^2)} \right] \quad (100d)$$

and so eqn (99a) becomes

$$P_n(\eta \leq \eta_P) \simeq \int_0^{\eta_P} d\eta \eta \frac{n}{st(1-r^2)^{\frac{1}{2}}} \exp \left[-\frac{n\eta^2(t^2+s^2)}{4s^2t^2(1-r^2)} \right] I_0 \left[\frac{n\eta^2h}{4s^2t^2(1-r^2)} \right] \quad (100e)$$

The appearance here of the function I_0 is consistent with the researches into random electronic noise distributions by Rice [9]. If r^2 is not too close to 1, the maximum argument of the I_0 factor is about $\frac{1}{2}$; so (from the series expansion of I_0 in Jeffreys and Jeffreys [3] or Copson [5]) the I_0 factor varies only between 1 and about 11/16 over the range of integration in eqn (100e), and can be dropped in the first approximation. Its co-factor will be recognised as the probability density of a two-dimensional Gaussian distribution, which may be integrated at sight to give

$$P_n(\eta \leq \eta_P) \simeq \frac{2st(1-r^2)^{\frac{1}{2}}}{(t^2+s^2)} \left[1 - \exp \left[-\frac{n\eta_P^2(t^2+s^2)}{4s^2t^2(1-r^2)} \right] \right] \quad (100f)$$

From the definition (97a) of h (and its discussion) it is clear that the factor in front of the square brackets here is less than unity, and so therefore is the whole expression, as it should be. Moreover the two main factors are both $O(1)$, provided s and t are not very different; from this it follows that the restriction on η_P^2 is not very significant, because the range over which it applies includes most of the probability (this is somewhat similar to the behaviour of the binomial law as it tends to the normal distribution, described in Part I). Also, the "scaling parameter" in the distribution (100f) is clearly proportional to $1/\sqrt{n}$, a result which has appeared several times before in Parts I and II and this Part.

In the derivation of eqns (100) I ruled out of consideration all extreme combinations of values of s , t and r , including $s=t$ and $r=0$ (the uniform circular distribution). In fact, however, the degenerate form of eqn (100f) for this case, which is

$$P_n(\eta \leq \eta_P) \simeq 1 - \exp \left[-\frac{n\eta_P^2}{2(s+t)^2} \right] \quad (100g)$$

(using the notation of eqn (96e)), is clearly the proper form for large n of eqn (96e), using the well-known exponential limit as expressed by eqn (100b). This comes about because eqn (96e) is itself a special case of the distribution within regions bounded by ellipses of constant probability density, and I now derive this distribution to complete my mathematical analysis.

Probability within a region bounded by an ellipse of constant probability density

I treat this by introducing a parametric form similar to that of eqns (94); thus

$$\lambda = \bar{x} + sR \left\{ (\frac{1}{2}(1+r))^{\frac{1}{2}} \cos\psi' + (\frac{1}{2}(1-r))^{\frac{1}{2}} \sin\psi' \right\} \quad (101a)$$

$$\mu = \bar{y} + tR \left\{ (\frac{1}{2}(1+r))^{\frac{1}{2}} \cos\psi' - (\frac{1}{2}(1-r))^{\frac{1}{2}} \sin\psi' \right\} \quad (101b)$$

or, using the notation of eqns (93),

$$\Delta\lambda \equiv \lambda - \bar{x} = sR \cos(\psi' - \psi'_0) \quad (101c)$$

$$\Delta\mu \equiv \mu - \bar{y} = tR \cos(\psi' + \psi'_0) \quad (101d)$$

where

$$\tan \psi'_0 = ((1-r)/(1+r))^{\frac{1}{2}} \quad (101e)$$

with $0 < \psi'_0 < \frac{1}{2}\pi$ (allowing for negative values of r). Using either of these forms, as ψ' increases from 0 to 2π the point (λ, μ) proceeds clockwise (the natural direction, if not the mathematicians' choice) once around a closed curve satisfying eqn (93a) (which has been proved to be an ellipse), with the constant C in that equation having the value $(1-r^2)R^2$, so that the semi-axes of the ellipse are $s'(1-r^2)^{\frac{1}{2}}R$ and $t'(1-r^2)^{\frac{1}{2}}R$ in the notation of eqns (93). I can write these as $s''R$ and $t''R$, where (by transforming eqns (93k) and (93l)) the quantities s''^2 and t''^2 are the roots of the quadratic equation

$$(s^{-2}t^{-2})X^2 - (s^{-2}+t^{-2})X + (1-r^2) = 0 \quad (101f)$$

or

$$X^2 - (s^2+t^2)X + s^2t^2(1-r^2) = 0 \quad (101g)$$

On substituting the value for C of eqn (93a) into eqn (92), I obtain

$$P_n(d\lambda d\mu|\theta) = \frac{(n-2)}{2\pi st(1-r^2)^{\frac{1}{2}}} \{1 + R^2\}^{-\frac{1}{2}n} d\lambda d\mu \quad (102a)$$

In terms of the Jacobian determinant, I have here (using eqns (101c) and (101d))

$$\begin{aligned} d\lambda d\mu &= \frac{\partial(\lambda, \mu)}{\partial(R, \psi')} dR d\psi' = \left[\left(\frac{\partial\lambda}{\partial R} \right) \left(\frac{\partial\mu}{\partial\psi'} \right) - \left(\frac{\partial\mu}{\partial R} \right) \left(\frac{\partial\lambda}{\partial\psi'} \right) \right] dR d\psi' \\ &= \left[\left\{ s \cos(\psi' - \psi'_0) \right\} \left\{ -tR \sin(\psi' + \psi'_0) \right\} - \left\{ t \cos(\psi' + \psi'_0) \right\} \left\{ -sR \sin(\psi' - \psi'_0) \right\} \right] dR d\psi' \\ &= -sR \left[\cos(\psi' - \psi'_0) \sin(\psi' + \psi'_0) - \sin(\psi' - \psi'_0) \cos(\psi' + \psi'_0) \right] dR d\psi' \\ &= -sR \left[\sin \left\{ (\psi' + \psi'_0) - (\psi' - \psi'_0) \right\} \right] dR d\psi' \\ &= -sR \sin(2\psi'_0) dR d\psi' \end{aligned} \quad (102b)$$

From the specification of ψ'_0 at eqn (101e), $\sin(2\psi'_0)$ is positive throughout the range of ψ'_0 (except at the unattainable/unphysical endpoints where it vanishes), and so the Jacobian in eqn (102b) is essentially negative. This is due to the fact that ψ' increases clockwise, and so (R, ψ') is a left-handed co-ordinate set while (λ, μ) is a right-handed one, which is inconsistent; to restore consistency I change the sign in eqn (102b). Also, from eqn (101e)

$$\begin{aligned}
\sin(2\psi'_0) &= 2\sin\psi'_0\cos\psi'_0 = 2\tan\psi'_0\cos^2\psi'_0 \\
&= 2\tan\psi'_0\{(\cos^2\psi'_0 + \sin^2\psi'_0)/\cos^2\psi'_0\}^{-1} \\
&= 2\tan\psi'_0(1 + \tan^2\psi'_0)^{-1} = (1-r^2)^{\frac{1}{2}}
\end{aligned} \tag{102c}$$

Substituting back eqn (102c) and the altered-sign form of eqn (102b) into eqn (102a) gives then

$$P_n(dRd\psi'|\theta) = \frac{(n-2)}{2\pi} \{1 + R^2\}^{-\frac{1}{2}n} R dR d\psi' \tag{102d}$$

so

$$\begin{aligned}
P_n(R \leq R_P) &\equiv \int_0^{2\pi} \int_0^{R_P} P_n(dRd\psi'|\theta) \\
&= \int_0^{2\pi} \int_0^{R_P} \frac{(n-2)}{2\pi} \{1 + R^2\}^{-\frac{1}{2}n} R dR d\psi' \\
&= \left[-\{1 + R^2\}^{-\frac{1}{2}(n-2)} \right]_0^{R_P} = 1 - \{1 + R_P^2\}^{-\frac{1}{2}(n-2)}
\end{aligned} \tag{102e}$$

remembering that n is at least 3; and obviously the complementary form is given by

$$P_n(R \geq R_P) \equiv 1 - P_n(R \leq R_P) = \{1 + R_P^2\}^{-\frac{1}{2}(n-2)} \tag{102f}$$

with the correct normalisation verified at sight. When $r=0$ and $s=t=(s\&t)$, from eqn (101g) $s'^2=t'^2=(s\&t)^2$, so the ellipses of constant probability density degenerate into circles of radii $(s\&t)R$; then, on comparison with eqns (94), $\eta=(s\&t)R$, $\eta_P=(s\&t)R_P$, and eqn (102e) reduces to eqn (96e).

Eqns (102e) and (102f) are quite astonishingly simple, considering that I have derived them from the five-dimensional differential form (54). In my view, this simplicity is a strong argument for specifying two-dimensional uncertainties in terms of the ellipses of constant probability density, rather than using circular or rectangular regions as is customarily done; but this is a matter for practical measurement specialists to debate. The exact results are so simple that it is hardly worth using an approximate exponential form for large n ; but for formal completeness I note that, just as the limit expressed by eqn (100b) allows eqn (96e) to be reduced to eqn (100g), so in the same way eqn (102e) reduces (on remarking that R_P^2 must be $O(n^{-1})$ for $P_n(R \leq R_P)$ to be significantly less than 1) to

$$P_n(R \leq R_P) = 1 - \exp\{-\tfrac{1}{2}nR_P^2\} + O(n^{-1}) \tag{102g}$$

$$- (\exp\{-\tfrac{1}{2}nR_P^2\})(1 + R_P^2 + \tfrac{1}{2}nR_P^4) + O(n^{-2}) \tag{102h}$$

where the first approximation shows, as always, that the scale factor of the distribution for large n and constant total probability is proportional to $1/\sqrt{n}$. This exponential form is similar to the Rayleigh distribution.

Eqn (102e) is not entirely original, having been almost derived by Wilks [10]. He obtained an expression equivalent to the result of integrating eqn (102d) with respect to ψ' from 0 to (2π) – as in eqn (102e) – but failed to notice that the integration over R can also be performed in closed form; this is understandable, since he was concerned to derive the distribution of R (or a generalisation of it) not only for measurement pairs but for any multiplets of measurements (triplets, quadruplets, etc.). However, Wilks' derivation relies on the idea of "maximum likelihood" (e.g. Jeffreys [1], Kendall and Stuart [11]), and is therefore neither rigorous nor strictly accurate (except when n is assumed to be very large, an assumption which I have carefully avoided from the start of this work); so the agreement must be regarded as no more than an encouraging coincidence. Moreover, there is no reliable way of arguing back from the distribution of R (eqn (102e)) to the joint distribution of λ and μ (eqn (92)); so even if Wilks' derivation is admitted, the fundamental conclusion of this paper does not follow, and the present independent analysis is still necessary. For a detailed philosophical and analytical comparison of "maximum likelihood" with the so-called "Bayesian method" (which I have essentially used) and certain other possibilities, the reader is referred to Jeffreys [1] and particularly to Kendall and Stuart [11].

CONCLUSIONS

In this paper, I have presented a derivation from first principles of the probability distribution of a sample of measurements of a complex number when the underlying law of joint distribution of the complex number's real and imaginary parts is the distribution of normal correlation, and I have shown that the derivation is valid under the same general conditions which make the Gaussian normal law be the underlying law of distribution of one measurable quantity. With the five principal parameters of the sample ($\overline{x}, \overline{y}, s, t$ and r) defined by eqns (52) from the individual measurement pairs (x_i, y_i ; $i=1$ to n), where n is the number of measurement pairs in the sample, the joint probability that the mean values in the underlying law are (λ, μ) is given by eqn (92), which is normalised to unity total probability. The best estimate of these underlying mean values is the sample's joint mean $(\overline{x}, \overline{y})$, and the distribution of probability is such that equal values of probability density lie on an ellipse.

The ellipses of constant probability density are all centred at the point $(\overline{x}, \overline{y})$ in the (x, y) or (λ, μ) plane; they are similar (all having the same aspect ratio or eccentricity); and they are similarly situated (every one has its major axis inclined to the x -axis by the angle ϕ (measured anti-clockwise from the x -axis to the major axis) given by eqns (97b) and (97c), or by eqn (93c) with $\sin(2\phi)$ having the same sign as r). (See also the Appendix.) If the semi-major and semi-minor axes of any such ellipse are taken as $(s''R_p)$ and $(t''R_p)$ respectively, where the quantities s'' and t'' (which fix the common aspect ratio) are the square roots of the roots of the quadratic equation (101g) and R_p is a parametric scaling factor, then the probability that both mean values in the underlying law lie within the ellipse parametrised by R_p is given by the expression (102e), which is far simpler than any other relevant law of distribution. The probability that both mean values in the underlying law lie within a circle of radius η_p centred at $(\overline{x}, \overline{y})$ can also be expressed in various ways simpler than two-dimensional numerical integration, notably the one-dimensional numerical integral (96a) and (for sufficiently small η_p) the convergent infinite series (98g) with parameters determined recursively from eqns (98h)–(98j).

I consider that the most logical and satisfactory way of specifying random uncertainties of measurements tractable by my analysis is by quoting the dimensions and orientation of that ellipse of constant probability density within which there is a specified probability (e.g. "95%") of finding the mean values of the underlying distribution; and I hope that the existence of this analysis will stimulate discussion among measurement specialists about this proposed specification and other possible methods of applying my theoretical results.

REFERENCES

- [1] JEFFREYS,H. Theory of probability, International Series of Monographs on Physics, Oxford University Press, 1948 (2nd edition; also 3rd edition, 1961).
 - [2] WHITTAKER,E. and ROBINSON,G. The calculus of observations, Glasgow, Blackie and Son, 1944 (4th edition).
 - [3] JEFFREYS,H. and JEFFREYS,B.S. Methods of mathematical physics, Cambridge University Press, 1946 (1st edition; also 3rd edition, 1956, reprinted 1972).
 - [4] WILKINSON,J.H. The algebraic eigenvalue problem, Monographs on Numerical Analysis, Oxford University Press, 1965.
 - [5] COPSON,E.T. An introduction to the theory of functions of a complex variable, Oxford University Press, 1935 (reprinted 1960).
 - [6] GRADSHTEYN,I.S. and RYZHIK,I.M. Tables of integrals, series and products, London and New York, Academic Press, 1965.
 - [7] ERDELYI,A. (editor), Higher transcendental functions, vol. 1, Bateman Manuscript Project at California Institute of Technology, New York and London, McGraw-Hill, 1953.
 - [8] PIAGGIO,H.T.H. An elementary treatise on differential equations and their applications, London, G Bell and Sons, 1928 (reprinted 1949).
 - [9] RICE,S.O. Mathematical analysis of random noise. Bell System Tech. J. (in several parts), 1944.
- WILKS,S.S. Mathematical statistics, Princeton University Press, 1943.
- KENDALL,M.G. and STUART,A. The advanced theory of statistics, vol. 2 (of 3): "Inference and relationship", London, Griffin and Co., 1961.

APPENDIX

The determination of the angle ϕ between the x-axis of the measurements and the major axis of the ellipses of constant probability density has been presented very briefly in the main text. As this angle is essential to the natural specification of the uncertainty with which a measurement sample's joint mean estimates the joint true mean of the underlying distribution, I present a detailed derivation here.

Compare eqns (93) and (48). The angle ϕ is defined by eqns (93d) and (93e), or eqns (48k) and (48l). It has the property that, when its corresponding rotation transformation (using the matrix definition (48c)) is applied to the matrix of the g_{ik} in eqn (48d), that matrix becomes the matrix of the g'_{ik} in eqn (48d). The matrix elements g'_{12} and g'_{21} are both zero, for which the necessary and sufficient condition is eqn (48e); when the values of the g_{ik} specified in eqns (93) are inserted into eqn (48e), it becomes eqn (93c), which depends only on ratios of matrix elements, so the scaling from s' and t' in eqns (93) to s'' and t'' in the text of eqns (101) does not affect it.

Now, I remark that the semi-axes s'' and t'' (with the scaling) or s' and t' (without it) must be different, unless the ellipses degenerate into circles (for which eqn (93m) shows that $s=t$ and $r=0$, and eqn (93c) breaks down – as is reasonable, since a circle has no preferred axes). The angle ϕ is defined in eqns (93d) and (93e) as the angle between the $\Delta\lambda$ -direction (which is the original x-axis) and the $\Delta\lambda'$ -direction, which I want to be the common major axis of the ellipses. Satisfying eqn (93c) makes the matrix elements g'_{12} and g'_{21} both zero, which (from eqns (48j), (93f) and (93g)) is sufficient to align the $\Delta\lambda'$ -direction with one or other of the lines of semi-axes; to achieve alignment specifically with the common major axis, however, the matrix element g'_{11} in eqn (48j), which has the value s'^{-2} in eqn (93f), must be the matrix element corresponding to the major axis, since it is obviously the matrix element corresponding to the $\Delta\lambda'$ -direction. In the non-degenerate or truly elliptical general case, the semi-major axis (which is proportional to s' , from eqn (93g)) is by definition bigger than the semi-minor axis (which is proportional to t' , from eqn (93g)). Hence, in eqn (93f) s'^{-2} is smaller than t'^{-2} ; or, in eqn (48d) g'_{11} is smaller than g'_{22} . These two matrix elements can be evaluated, using eqn (48d), from the matrix elements g_{11} , g_{22} and $g_{12}=g_{21}=g_X$; which, from eqn (93b) compared with eqns (48b) and (48d), are respectively proportional to s^{-2} , t^{-2} and $(-r(st)^{-1})$. The condition $g'_{11} < g'_{22}$ is not affected by constants of proportionality or scaling factors (provided they are positive, which they obviously are from the text of eqns (101)); so I may disregard the proportionality and write simply into eqn (48d)

$$g_{11} = s^{-2} \quad (\text{A1a}); \quad g_{22} = t^{-2} \quad (\text{A1b}); \quad g_X = -r(st)^{-1} \quad (\text{A1c})$$

Then, from eqn (48d), with eqns (A1a)–(A1c) and (93c),

$$\begin{aligned} 0 &< (g'_{22} - g'_{11}) \\ &= (g_{11}\sin^2\phi + g_{22}\cos^2\phi - g_X\sin 2\phi) - (g_{11}\cos^2\phi + g_{22}\sin^2\phi + g_X\sin 2\phi) \\ &= (g_{22} - g_{11})(\cos^2\phi - \sin^2\phi) - 2g_X\sin 2\phi = (g_{22} - g_{11})\cos 2\phi - 2g_X\sin 2\phi \\ &((g_{22} - g_{11}) - 2g_X\tan 2\phi) \cos 2\phi \end{aligned}$$

$$\begin{aligned}
& \left[\left[\frac{s^2-t^2}{s^2t^2} \right] + \left[\frac{4r^2}{s^2-t^2} \right] \right] \cos 2\phi \\
&= (st)^{-2} \left((s^2-t^2)^2 + 4r^2s^2t^2 \right) \left[\frac{\cos 2\phi}{s^2-t^2} \right] \quad (\text{A2a})
\end{aligned}$$

In this inequality I substitute from the definition (97a), recalling that the quantity h cannot vanish in the non-degenerate case; then

$$(st)^{-2} h^2 \left[\frac{\cos 2\phi}{s^2-t^2} \right] > 0 \quad (\text{A2b})$$

or

$$\cos 2\phi = K_C (s^2-t^2) \quad (\text{A2c})$$

where K_C is strictly positive (note that $\cos 2\phi$ and (s^2-t^2) can vanish, as long as they vanish (or not) together, and as long as r does not also vanish to avoid the degenerate case). On multiplying eqn (A2c) by eqn (93c), I have

$$\begin{aligned}
\sin 2\phi &= \cos 2\phi \tan 2\phi = K_C (s^2-t^2) (2rst/(s^2-t^2)) \\
&2rstK_C \quad (\text{A2d})
\end{aligned}$$

provided (for the moment) that $(s^2-t^2) \neq 0$ to justify the cancellation at the last stage. I can immediately remove this restriction. When $(s^2-t^2)=0$, $\cos 2\phi=0$ from eqn (A2c), so this equation no longer determines K_C . Hence I can choose a strictly positive K_C to satisfy eqn (A2d), provided $\sin 2\phi$ and r have the same sign; and when $\cos 2\phi$ vanishes, the third line of eqn (A2a) reduces to

$$0 < -2g_X \sin 2\phi = 2r(st)^{-1} \sin 2\phi \quad (\text{A2e})$$

so r and $\sin 2\phi$ do indeed have the same sign (and are both forbidden to vanish, which is to be expected since in this special case r cannot vanish without causing degeneracy and no sine can vanish simultaneously with its cosine). Hence eqns (A2c) and (A2d) both hold, whether $(s^2-t^2)=0$ or not; and (2ϕ) then follows by rectangular-to-polar conversion, the ambiguity of π radians or 180° in ϕ having no significance since either direction along the major axis will do.

Finally, I make this appendix logically complete by proving that $(2\phi)=\pi-\psi_0$, using the notation and text of eqns (97). I remark that $K_C=h^{-1}$; on squaring and adding eqns (A2c) and (A2d), this result follows immediately from eqn (97a), remembering that K_C is strictly positive and h is non-negative by definition. Eqn (A2d) then becomes identical with (97c), while eqn (A2c) becomes identical with (97b) except for a reversal of sign; so (2ϕ) and ψ_0 have identical sines and sign-reversed cosines, and the stated relationship between these angles then follows.

...the ... of ...
...the ... of ...
...the ... of ...

...the ... of ...
...the ... of ...
...the ... of ...

...the ... of ...
...the ... of ...
...the ... of ...

...the ... of ...
...the ... of ...
...the ... of ...

...the ... of ...
...the ... of ...
...the ... of ...

...the ... of ...
...the ... of ...
...the ... of ...

...the ... of ...
...the ... of ...
...the ... of ...

...the ... of ...
...the ... of ...
...the ... of ...