The impact of correlations on the optimum location measure and its variance for samples drawn from Gaussian populations

F. Masci, 11/09/2011

We explore the optimal location measure (or mean) μ and its variance $v_\mu = \sigma^2_\mu$ for two and three random variables where correlations may be present between any data pairs, i.e., $\rho_{12} \neq 0$, $\rho_{23} \neq 0$, and/or $\rho_{13} \neq 0$. **Note**: if $\rho_{ij} = 0$ for all i, j where $i \neq j$, we get the familiar: $v_\mu = \sigma^2/3$ for $N = 3$. Why $N \leq 3$? Because this is small and manageable enough for illustrating what happens in general.

We make use of an error-covariance matrix Ω_N for N random variables. If their population priors (σ^2) are all the same, this can be conveniently written as a correlation matrix:

$$\Omega_N = \sigma^2 \begin{pmatrix} 1 & \rho_{12} & \rho_{13} & \cdots & \rho_{1N} \\ \rho_{21} & 1 & \rho_{23} & \cdots & \rho_{2N} \\ \rho_{31} & \rho_{32} & 1 & \cdots & \rho_{3N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{N1} & \rho_{N2} & \rho_{N3} & \cdots & 1 \end{pmatrix} \tag{1}$$

The covariance or correlation matrix is always symmetric since $\rho_{ij} = \rho_{ji}$ (or $\Omega = \Omega^T$). Furthermore, for a covariance matrix to correspond to a physically plausible multivariate probability distribution for all variables *jointly*, it must be positive definite, i.e.,

$$X^T \Omega_N X > 0 \quad \forall \quad \text{real-valued vectors } X \neq 0. \tag{2}$$

Equivalently, since the inverse of a positive-definite matrix is also positive definite, it follows that

$$\chi^2 = X^T \Omega_N^{-1} X > 0 \quad \forall \quad \text{real-valued vectors } X \neq 0. \tag{3}$$

The conditions defined by Equation (2) or (3) are equivalent to ensuring that *all* eigenvalues of Ω_N be non-zero, positive, and real-valued. The eigenvalues represent the variances along the principal axes of the multivariate probability distribution in a diagonalized system of new random variables, i.e., with all correlations removed. These new variables are linear combinations of the initially correlated variables and the eigenvectors give the directions of the (orthogonal) principal axes. Geometrically, the error-ellipsoids (surfaces of equal density), or any vector of the N variables in the two systems are related by a $N \times N$ rotation matrix Q. This is commonly referred to as the eigendecomposition matrix since covariance matrices in the diagonalized and rotated (or correlated) system are related by:

$$\Omega_N = Q \Omega_N^D Q^T. \tag{4}$$

The determinant of Ω_N (designated $|\Omega_N|$) is equal to the product of all its eigenvalues and is preserved under rotation. Incidentally, the volume of an error-ellipsoid is $\propto \sqrt{|\Omega_N|}$ and hence, it is no surprise this measure is found in the general form of the multivariate Gaussian probability *density* function:

$$f_N(X) = \frac{e^{-\chi^2/2}}{(2\pi)^{N/2}\sqrt{|\Omega_N|}}, \tag{5}$$

where χ^2 is defined in Equation (3) with $\boldsymbol{X} = \boldsymbol{X'} - \boldsymbol{\mu_X'}$ for some real data vector $\boldsymbol{X'}$.

A statement often encountered in the literature is that the condition for positive-definiteness (Eq. 2 or 3) is always satisfied if $|\Omega_N| > 0$. I have discovered this not to be true because a positive determinant hides the possibility that an even number of eigenvalues can be negative, rendering the covariance matrix unphysical. I.e., such a covariance matrix can never arise for real correlated variates in nature, unless of course one estimates the sample (co)variances (rather than using priors) and measurement error makes a specific covariance structure implausible. More commonly, the assignment of arbitrary correlation coefficients to generate test covariance matrixes will always require one to check for plausibility before use. I have only observed the occurrence of two negative eigenvalues starting with $N = 4$. None were found for $N \leq 3$. A proof is in preparation. Therefore, $|\Omega_N| > 0$ alone is not a sufficient condition to ensure a positive definite covariance matrix.

A necessary and sufficient condition for Ω_N to be positive-definite is that all its eigenvalues be > 0. This automatically ensures $|\Omega_N| > 0$ and hence Ω_N is non-singular. Instead of laboriously computing all the eigenvalues, the following check will ensure this, known as "Sylvester's Criterion" (from *http://mathworld.wolfram.com/PositiveDefiniteMatrix.html*):

****** The definition of positive definiteness is equivalent to the requirement that the determinants associated with all upper-left submatrices are positive.*

Optimal measures of location and its noise-variance for Gaussian-distributed variables

Appendix I gives a derivation of the optimal location estimate and its variance (in the maximum likelihood sense) for the combination of N correlated *Gaussian*-distributed variables, each drawn from either the same or different Gaussian population:

$$\hat{\mu} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij} x_i}{\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}}, \tag{6}$$

$$v_\mu = \left[\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}\right]^{-1}, \tag{7}$$

where $w_{ij} \equiv \Omega_{ij}^{-1}$, i.e., the weights are the elements of the *inverse* covariance matrix in Equation (1). When correlations are absent, Equations (6) and (7) reduce to the more familiar inverse-variance weighted mean and variance expressions with the weights determined exclusively by the prior variances: $w_i \equiv 1/\sigma_i^2$.

Furthermore, when correlations are absent and all measurements are drawn from the same Gaussian population with prior noise-variance σ^2 (weights are all the same), the optimal location measure becomes the well known "arithmetic mean" $\mu = \Sigma_i x_i / N$ with uncertainty $\sigma_\mu = \sqrt{v_\mu} = \sigma/\sqrt{N}$.

In general, if all N measurements have the *same* prior variance and the *same* non-zero correlations are present between all mutual variables ($0 < |\rho_{ij}| < 1$), the optimal *Gaussian* location measure (Eq. 6) will not be affected. This has to do with the "ellipsoidal symmetry" of a multivariate Gaussian distribution and the fact that only *linear* correlations can be represented. As a side note, the following memo could be of interest: http://web.ipac.caltech.edu/staff/fmasci/home/statistics_refs/UncorrelatedButDependent2011.pdf
If however the measurements have *different* prior variances σ_i^2 and/or correlations ρ_{ij}, the weights in Eq. (6) will not cancel and the location measure will now depend on them (with the σ_i^2 becoming coupled to the ρ_{ij}). For $N = 2$, the optimum location measure will always be independent of ρ (for equal prior variances) because only one such ρ exists (see example below). For $N > 2$, all the ρ_{ij} need to be equal for the same rule to apply.

Two correlated random variables

For example, let's explore the $N = 2$ case with $\rho = \rho_{12} = \rho_{21}$ and a constant prior σ^2. The covariance matrix and its inverse are given by:

$$\Omega_2 = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

$$\Omega_2^{-1} = \frac{1}{\sigma^2\left(1-\rho^2\right)} \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix} \equiv \frac{\sigma^2}{|\Omega_2|} \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix}.$$

With

$$w_{11} = w_{22} = \frac{1}{\sigma^2\left(1-\rho^2\right)},$$

$$w_{12} = w_{21} = \frac{-\rho}{\sigma^2\left(1-\rho^2\right)},$$

Equations (6) and (7) then give:

$$\hat{\mu} = \frac{\left(w_{11}+w_{12}\right)x_1 + \left(w_{21}+w_{22}\right)x_2}{w_{11}+w_{12}+w_{21}+w_{22}} = \frac{\left[\sigma^2\left(1-\rho^2\right)\right]^{-1}\left[\left(1-\rho\right)x_1 + \left(1-\rho\right)x_2\right]}{\left[\sigma^2\left(1-\rho^2\right)\right]^{-1}\left(2-2\rho\right)} = \frac{x_1+x_2}{2}.$$

$$v_\mu = \frac{1}{w_{11}+w_{12}+w_{21}+w_{22}} = \frac{|\Omega_2|}{2\sigma^2\left(1-\rho\right)} = \frac{\sigma^4\left(1-\rho^2\right)}{2\sigma^2\left(1-\rho\right)} = \frac{\sigma^2}{2}\left(1+\rho\right).$$

Therefore, when the measurements have the same prior variance σ_i^2, the optimal measure of location $\hat{\mu}$ reduces to the arithmetic mean and correlations will not impact its value. This will also be true for all $N > 2$ only if $\rho_{ij} = constant$ for all $i \neq j$ and the σ_i^2 are all equal. If the σ_i^2 are not all equal, then $N = 2$ gives:

$$\hat{\mu} = \frac{\left(\sigma_2^2 - \rho\sigma_1\sigma_2\right)x_1 + \left(\sigma_1^2 - \rho\sigma_1\sigma_2\right)x_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2},$$

$$v_\mu = \frac{\sigma_1^2\sigma_2^2\left(1-\rho^2\right)}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}.$$

These reduce to the above expressions when $\sigma_1^2 = \sigma_2^2$.

Three correlated random variables

For the general case of unequal variances and non-zero correlations, $N = 3$ gives:

$$\hat{\mu} = \frac{\left\{\begin{array}{c} \left(\left[1-\rho_{23}^2\right]\sigma_2^2\sigma_3^2 - \rho_{12}\sigma_3^2\sigma_1\sigma_2 - \rho_{13}\sigma_2^2\sigma_1\sigma_3 + \rho_{12}\rho_{23}\sigma_2^2\sigma_1\sigma_3 + \rho_{13}\rho_{23}\sigma_3^2\sigma_1\sigma_2\right)x_1 + \\ \left(\left[1-\rho_{13}^2\right]\sigma_1^2\sigma_3^2 - \rho_{12}\sigma_3^2\sigma_1\sigma_2 - \rho_{23}\sigma_1^2\sigma_2\sigma_3 + \rho_{12}\rho_{13}\sigma_1^2\sigma_2\sigma_3 + \rho_{13}\rho_{23}\sigma_3^2\sigma_1\sigma_2\right)x_2 + \\ \left(\left[1-\rho_{12}^2\right]\sigma_1^2\sigma_2^2 - \rho_{13}\sigma_2^2\sigma_1\sigma_3 - \rho_{23}\sigma_1^2\sigma_2\sigma_3 + \rho_{12}\rho_{13}\sigma_1^2\sigma_2\sigma_3 + \rho_{12}\rho_{23}\sigma_2^2\sigma_1\sigma_3\right)x_3 \end{array}\right\}}{\left\{\begin{array}{c} \sigma_1^2\sigma_2^2\left(1-\rho_{12}^2\right) + \sigma_1^2\sigma_3^2\left(1-\rho_{13}^2\right) + \sigma_2^2\sigma_3^2\left(1-\rho_{23}^2\right) - 2\sigma_1\sigma_2\sigma_3\left(\rho_{12}+\rho_{13}+\rho_{23}\right) + \\ 2\left(\rho_{12}\rho_{13}\sigma_1^2\sigma_2\sigma_3 + \rho_{12}\rho_{23}\sigma_2^2\sigma_1\sigma_3 + \rho_{13}\rho_{23}\sigma_3^2\sigma_1\sigma_2\right) \end{array}\right\}}$$

$$v_\mu = \frac{\sigma_1^2\sigma_2^2\sigma_3^2\left(1-\rho_{12}^2-\rho_{13}^2-\rho_{23}^2+2\rho_{12}\rho_{13}\rho_{23}\right)}{\left\{\begin{array}{c} \sigma_1^2\sigma_2^2\left(1-\rho_{12}^2\right) + \sigma_1^2\sigma_3^2\left(1-\rho_{13}^2\right) + \sigma_2^2\sigma_3^2\left(1-\rho_{23}^2\right) - 2\sigma_1\sigma_2\sigma_3\left(\rho_{12}+\rho_{13}+\rho_{23}\right) + \\ 2\left(\rho_{12}\rho_{13}\sigma_1^2\sigma_2\sigma_3 + \rho_{12}\rho_{23}\sigma_2^2\sigma_1\sigma_3 + \rho_{13}\rho_{23}\sigma_3^2\sigma_1\sigma_2\right) \end{array}\right\}}, \tag{8}$$

If the $N = 3$ sample were drawn from the same Gaussian population with prior variance σ^2, Equation (8) reduces to:

$$v_\mu = \sigma^2 \frac{1-\rho_{12}^2-\rho_{13}^2-\rho_{23}^2+2\,\rho_{12}\rho_{13}\rho_{23}}{3-2\left(\rho_{12}+\rho_{13}+\rho_{23}\right)+2\left(\rho_{12}\rho_{13}+\rho_{12}\rho_{23}+\rho_{13}\rho_{23}\right)-\rho_{12}^2-\rho_{13}^2-\rho_{23}^2} \tag{9}$$

We explored Eq. (9) in some detail by first selecting arbitrary values of ρ_{ij} and then checking them for joint-plausibility, i.e., such that they made Ω_3 positive definite. This ensured $0 < v_\mu < \sigma^2$ and that Ω_3 was a plausible covariance matrix for some multivariate normal distribution. We used Sylvester's Criterion (see above) to derive the following constraint equations:

$$1-\rho_{12}^2 > 0 \quad \& \tag{10}$$

$$1-\rho_{12}^2-\rho_{13}^2-\rho_{23}^2+2\,\rho_{12}\rho_{13}\rho_{23} > 0 \tag{11}$$

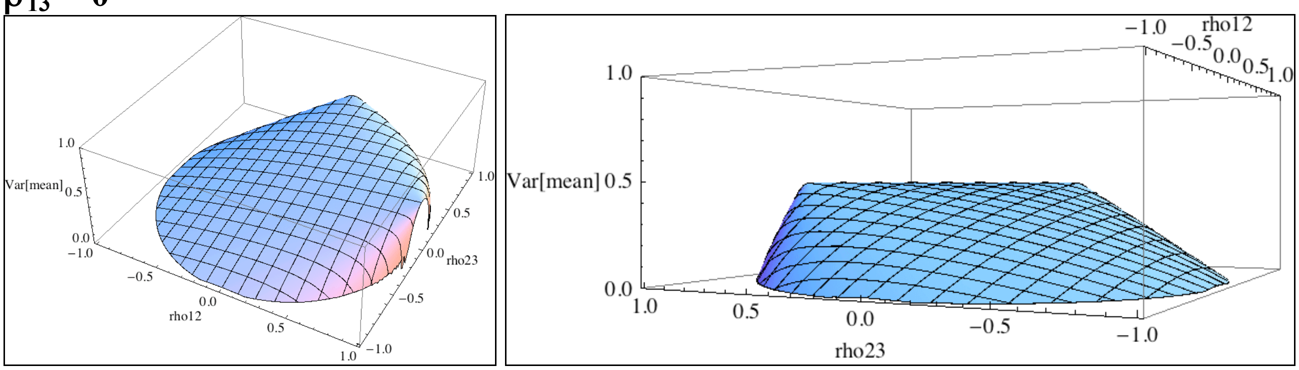
The first equation implies $|\rho_{12}| < 1$ and is enforced by requiring $|\rho_{12}| < 0.99$. This specific upper bound was chosen to minimize precision and round-off error in general for all ρ_{ij} ($i \neq j$). Therefore, we are left with the constraint in Eq. (11):

$$\rho_{12}^2 + \rho_{13}^2 + \rho_{23}^2 < 1 + 2\ \rho_{12}\rho_{13}\rho_{23}. \tag{12}$$

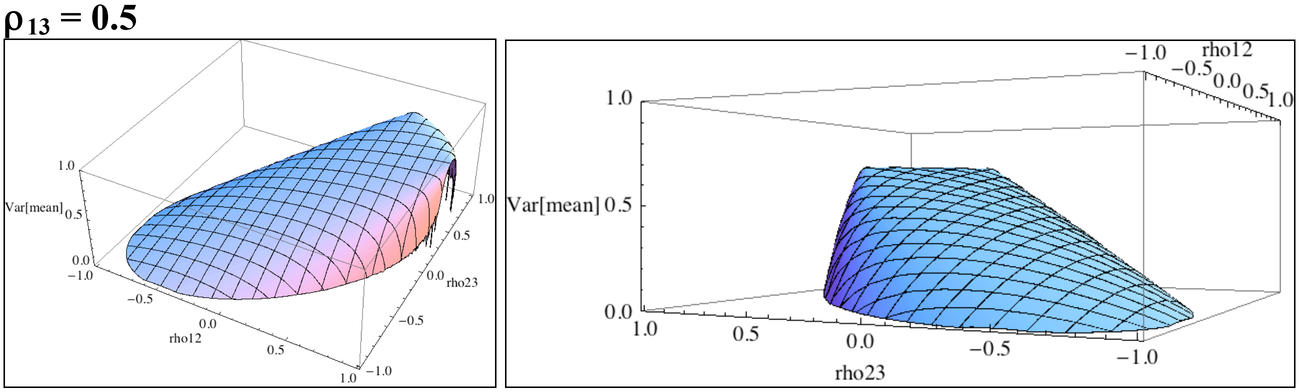
Below we show plots of Equation (9) for five different fixed values of ρ_{13} and assuming a constant prior of $\sigma^2 = 1$. Two orientations are shown for each ρ_{13}. The dependency of v_μ on the size of the correlations can differ considerably. Qualitatively, if all three data pairs (or at least two depending on their ρ_{ij}) are positively correlated ($\rho_{ij} > 0$), then in general $v_\mu > \sigma^2/3$, i.e., more than that expected if all correlations were zero. If all (or at least two depending on their ρ_{ij}) are anti-correlated ($\rho_{ij} < 0$), then in general $v_\mu < \sigma^2/3$. Regions with no data are forbidden, i.e., these do not satisfy Equations (10) and (11).

In closing, what happens if we relax the requirement of a positive-definite Ω_3, i.e., that all three eigenvalues be > 0? If one or more eigenvalues are negative, this implies negative variances in the diagonalized (uncorrelated) system of variables, and complex numbers are needed to explain them. According to Eq. (9) however, it is still possible for the variance in the mean of a set of random variables to be positive. This raises the question: can the negative eigenvalue solutions (in the uncorrelated system) be tossed and only the positive ones retained to define a new vector space of plausible variables? This subset can then be combined to compute real-valued means and higher-order moments. This sounds like an artificial procedure, but all it means is that joint-normality is now only possible for a reduced set of the variables. These new variables are linear combinations of the initial variables (or physical observables) that had a "bad" (implausible) covariance matrix, but they can still be used for analyzing the principal components of variation.

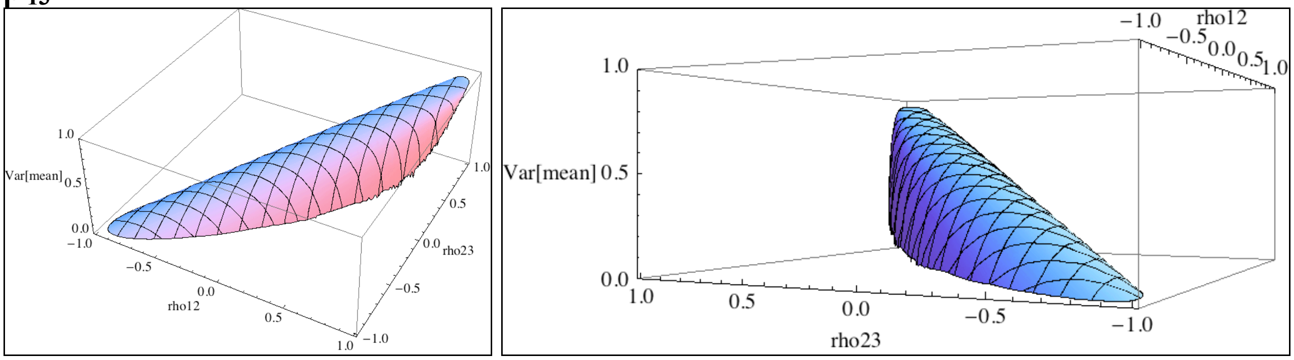
$\rho_{13} = 0$



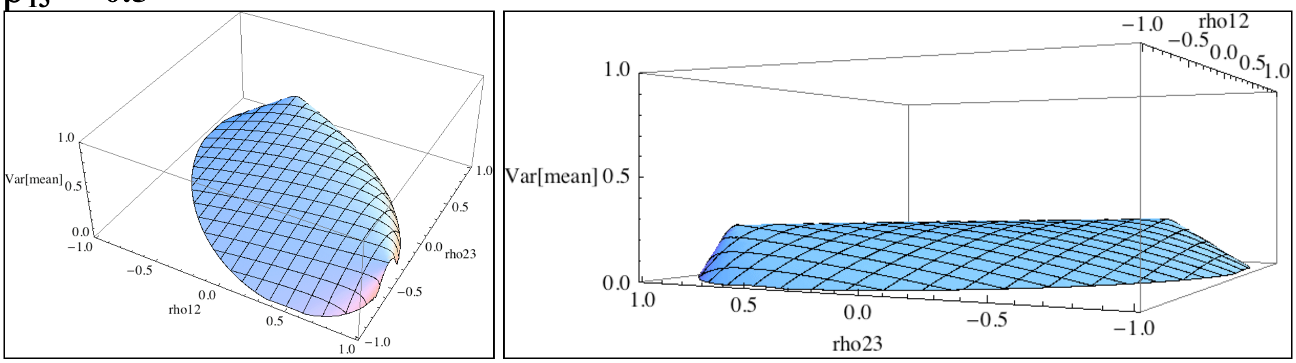
$\rho_{13} = 0.5$



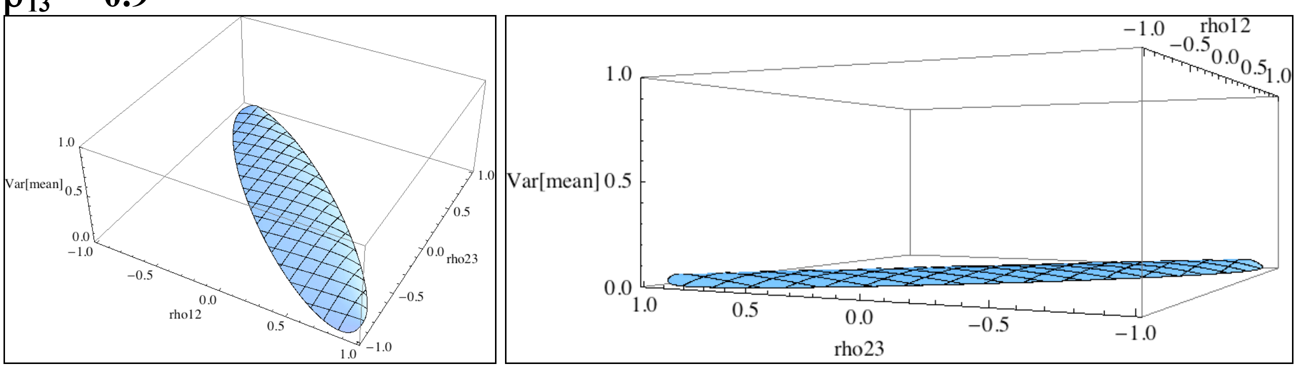
ρ_{13} = 0.9



ρ_{13} = -0.5



ρ_{13} = -0.9



Appendix I: the optimal measure of location and its noise-variance for a correlated sample of Gaussian-distributed variables

Here we present a maximum-likelihood estimate for the location and its variance for a set of random variables that may be correlated with each other and drawn from different Gaussian populations. By 'maximum-likelihood', we mean finding the location of some global probability density function that maximizes the chance of obtaining the observed sample (or realization of variables). In the absence of correlations and with variables drawn from the same population, the optimal location measure is simply the arithmetic mean of the sample ($\Sigma_i x_i/N$). Its uncertainty (square root of the variance) is then the constant population σ (or an estimate of it using the sample standard deviation) divided by $\sqrt{N}$. A sample containing mutual correlations and with each variable drawn from different prior populations presents the most general case.

We start with the general matrix form of the χ^2 as defined by Equation (3) with $\boldsymbol{X}$ replaced by $\boldsymbol{X} - \boldsymbol{\mu_X}$ for some real column vector $\boldsymbol{X}$ of variables $x_i = x_1, x_2, x_3, \ldots x_N$:

$$\chi^2 = \left(X - \mu_X\right)^T \Omega_N^{-1} \left(X - \mu_X\right) \qquad \text{(A1)}$$

We make the vector substitution:

$$Z = X - \mu_X$$
$$\Rightarrow \chi^2 = Z^T \Omega_N^{-1} Z \qquad \text{(A2)}$$

The value of $\boldsymbol{\mu_X}$ ($= \hat{\mu}$) that maximizes the Gaussian likelihood (Eq. 5) or equivalently minimizes χ^2, is that where the derivative of χ^2 with respect to $\boldsymbol{\mu_X}$ vanishes:

$$\frac{\partial \chi^2}{\partial \mu_X} = \frac{\partial \chi^2}{\partial Z} \frac{\partial Z}{\partial \mu_X} = 0. \qquad \text{(A3)}$$

We make use of a powerful identity from matrix calculus to obtain:

$$\begin{aligned} \frac{\partial \chi^2}{\partial Z} &= \frac{\partial \left(Z^T \Omega_N^{-1} Z\right)}{\partial Z} \\ &= Z^T \Omega_N^{-1} + Z^T \left(\Omega_N^{-1}\right)^T \\ &= 2 Z^T \Omega_N^{-1}, \end{aligned} \qquad \text{(A4)}$$

since if Ω_N is symmetric ($\Omega_N = \Omega_N^T$), its inverse is also symmetric: $\Omega_N^{-1} = \left(\Omega_N^T\right)^{-1} = \left(\Omega_N^{-1}\right)^T$.

The $\partial Z / \partial \mu_X$ derivative in Eq. (A3) is simply an $N \times 1$ column vector of -1's:

$$\frac{\partial Z}{\partial \mu_X} = \begin{pmatrix} -1 \\ -1 \\ -1 \\ \vdots \\ -1 \end{pmatrix} = -\vec{1}, \tag{A5}$$

where the new vector $\vec{1}$ represents an $N \times 1$ column vector with all elements equal to 1.

Using Equations (A4) and (A5) in (A3), the condition for a vanishing derivative becomes:

$$\begin{aligned} &-2Z^T \Omega_N^{-1} \vec{1} = 0 \\ &\Rightarrow \left(X - \hat{\mu}\right)^T \Omega_N^{-1} \vec{1} = 0 \\ &\Rightarrow X^T \Omega_N^{-1} \vec{1} = \hat{\mu} \vec{1}^T \Omega_N^{-1} \vec{1} \\ &\Rightarrow \hat{\mu} = \frac{X^T \Omega_N^{-1} \vec{1}}{\vec{1}^T \Omega_N^{-1} \vec{1}}. \end{aligned} \tag{A6}$$

To simplify the matrix equation in (A6), we define the inverse of the covariance matrix as W_N:

$$W_N = \Omega_N^{-1} = \begin{pmatrix} w_{11} & w_{12} & w_{13} & \cdots & w_{1N} \\ w_{21} & w_{22} & w_{23} & \cdots & w_{2N} \\ w_{31} & w_{32} & w_{33} & \cdots & w_{3N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ w_{N1} & w_{N2} & w_{N3} & \cdots & w_{NN} \end{pmatrix} \tag{A7}$$

If one uses W_N from (A7) in (A6) and attempts to multiply-out the matrices, one finds that:

$$\hat{\mu} = \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij} x_i}{\sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}}. \tag{A8}$$

Now we derive the variance in the location estimate given by (A8), or equivalently (A6). We start with (A6) and rewrite it in terms of a *true vector* $\boldsymbol{t} = (t_1\ t_2\ t_3\ \ldots\ t_N)^{\mathrm{T}}$ + *error vector* $\boldsymbol{\varepsilon} = (\varepsilon_1\ \varepsilon_2\ \varepsilon_3\ \ldots\ \varepsilon_N)^{\mathrm{T}}$:

$$t_\mu + \varepsilon_\mu = \frac{1}{D}\left(t + \varepsilon\right)^T \Omega_N^{-1} \vec{1}. \tag{A9}$$

where

$$D = \vec{1}^T \Omega_N^{-1} \vec{1} = \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}$$

Equation (A9) can be written:

$$t_\mu + \varepsilon_\mu = \frac{1}{D}\left(t^T \Omega_N^{-1} \vec{1} + \varepsilon^T \Omega_N^{-1} \vec{1}\right). \tag{A10}$$

squaring both sides of (A10) and taking expectation values:

$$\left\langle \left(t_\mu + \varepsilon_\mu\right)^2 \right\rangle = \frac{1}{D^2}\left\langle \left(t^T \Omega_N^{-1} \vec{1} + \varepsilon^T \Omega_N^{-1} \vec{1}\right)^2 \right\rangle$$

$$\Rightarrow \left\langle t_\mu^2 + 2\varepsilon_\mu t_\mu + \varepsilon_\mu^2 \right\rangle = \frac{1}{D^2}\left\langle \left(t^T \Omega_N^{-1} \vec{1}\right)^2 + 2\left(\varepsilon^T \Omega_N^{-1} \vec{1}\right)\left(t^T \Omega_N^{-1} \vec{1}\right) + \left(\varepsilon^T \Omega_N^{-1} \vec{1}\right)\left(\varepsilon^T \Omega_N^{-1} \vec{1}\right) \right\rangle$$

$$\Rightarrow \left\langle t_\mu^2 \right\rangle + 2\left\langle \varepsilon_\mu t_\mu \right\rangle + \left\langle \varepsilon_\mu^2 \right\rangle = \frac{1}{D^2}\left\langle \left(t^T \Omega_N^{-1} \vec{1}\right)^2 \right\rangle + \frac{2}{D^2} \vec{1}^T \Omega_N^{-1} \left\langle \varepsilon t^T \right\rangle \Omega_N^{-1} \vec{1} + \frac{1}{D^2}\left\langle \varepsilon^T \Omega_N^{-1} \vec{1} \varepsilon^T \Omega_N^{-1} \vec{1} \right\rangle,$$

where for the middle product on the right, we made the replacement: $\varepsilon^T \Omega_N^{-1} \vec{1} = \vec{1}^T \Omega_N^{-1} \varepsilon$.

Given $\left\langle \varepsilon_\mu t_\mu \right\rangle = 0$ and $\left\langle \varepsilon t^T \right\rangle = 0$, we make the following associations:

$$\left\langle t_\mu^2 \right\rangle = \frac{1}{D^2}\left\langle \left(t^T \Omega_N^{-1} \vec{1}\right)^2 \right\rangle,$$

$$\left\langle \varepsilon_\mu^2 \right\rangle = \frac{1}{D^2}\left\langle \varepsilon^T \Omega_N^{-1} \vec{1} \varepsilon^T \Omega_N^{-1} \vec{1} \right\rangle.$$

With the replacement $\varepsilon^T \Omega_N^{-1} \vec{1} = \vec{1}^T \Omega_N^{-1} \varepsilon$,

$$\left\langle \varepsilon_\mu^2 \right\rangle = \frac{1}{D^2}\left\langle \vec{1}^T \Omega_N^{-1} \varepsilon \varepsilon^T \Omega_N^{-1} \vec{1} \right\rangle.$$

Identifying $\left\langle \varepsilon \varepsilon^T \right\rangle = \Omega_N$, we have:

$$\left\langle \varepsilon_\mu^2 \right\rangle = \frac{1}{D^2}\left\langle \vec{1}^T \Omega_N^{-1} \vec{1} \right\rangle = \frac{D}{D^2} = \frac{1}{D} \tag{A11}$$

using the definition of *D* below Equation (A9).

Therefore, the corresponding variance for the optimum location measure is given by:

$$\boxed{v_\mu = \sigma_\mu^2 = \left[\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}\right]^{-1}} \tag{A12}$$

Note: prior to the author discovering the above matrix derivation, he started with Equation (A8), expressed the x_i in terms of 'truth' + ε_i, squared the terms and took expectation values, and arrived at:

$$v_\mu = \frac{1}{D^2}\left\{\sum_{i=1}^{N}\left[\sum_{j=1}^{N} w_{ij}\right]^2 \sigma_i^2 + 2\sum_{m=1}^{N}\sum_{n<m}^{N}\sum_{i}^{N}\sum_{j}^{N} \rho_{mn}\sigma_m\sigma_n w_{mi} w_{nj}\right\}. \tag{A13}$$

Equation (A13) will give results numerically equivalent to Equation (A12). Take your pick! I prefer (A12).