Univariate Normal distribution

In general, it has two parameters, μ and σ (mean and standard deviation). A special case is `standardized` Normal distribution, with the mean of 0 and standard deviation equal to 1. Any general X can be converted to standardized Z by

$$Z = \frac{X - \mu}{\sigma}$$

and reverse

$$X = \sigma Z + \mu$$

It is usually a lot easier to deal with Z, and then convert the results to X.

We should recall that in general, if $X \in \mathcal{N}(\mu, \sigma)$, then

$$aX + b \in \mathcal{N}(a\mu + b, |a|\sigma) \tag{1}$$

where a and b are constants.

The probability density function (PDF from now on) of Z and X is

$$f_Z(z) = \frac{\exp(-\frac{z^2}{2})}{\sqrt{2\pi}}$$
$$f_X(x) = \frac{\exp(-\frac{(x-\mu)^2}{2\sigma^2})}{\sqrt{2\pi}\sigma}$$

respectively.

Similarly, the moment generating function (MGF) is

$$\begin{aligned} M_z(t) &= \exp(\frac{t^2}{2}) \\ M_x(t) &= e^{\mu t} \cdot M_z(\sigma t) = \exp(\frac{\sigma^2 t^2}{2} + \mu t) \end{aligned}$$

Bivariate Normal distribution

Again, we consider two versions, the general (X and Y) and standardized (Z_1 and Z_2). The general distribution is defined by 5 parameters (the individual means and variances, plus the correlation coefficient ρ), the standardized version has only one, namely ρ.

The two joint (bivariate) PDF's are

$$f_{zz}(z_1, z_2) = \frac{\exp(-\frac{z_1^2 + z_2^2 - 2\rho z_1 z_2}{2(1-\rho^2)})}{2\pi\sqrt{1-\rho^2}}$$
$$f_{xy}(x, y) = \frac{\exp\left(-\frac{(\frac{x-\mu_1}{\sigma_1})^2 + (\frac{y-\mu_2}{\sigma_2})^2 - 2\rho(\frac{x-\mu_1}{\sigma_1})(\frac{y-\mu_2}{\sigma_2})}{2(1-\rho^2)}\right)}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$$

for the standardized and general case, respectively.

Similarly, the joint MGFs are

$$
\begin{aligned}
&M_{zz}(t_1,t_2) = \exp(\frac{t_1^2+t_2^2+2\rho t_1 t_2}{2}) \\
&M_{xy}(t_1,t_2) = e^{\mu_1 t_1+\mu_2 t_2} \cdot M_{zz}(\sigma_1 t_1, \sigma_2 t_2) = \\
&\exp\left(\frac{\sigma_1^2 t_1^2+\sigma_2^2 t_2^2+2\rho\sigma_1\sigma_2 t_1 t_2}{2}+\mu_1 t_1+\mu_2 t_2\right)
\end{aligned}
$$

We should remember that a joint MGF enables us to find joint simple moments of the distribution by

$$
\mathbb{E}\left(X^n Y^m\right) = \left.\frac{\partial^{(n+m)} M_{xy}(t_1,t_2)}{\partial t_1^n \partial t_2^m}\right|_{t_1=t_2=0}
$$

Also, we can easily find the MGF of a `marginal` distribution of X by setting $t_2 = 0$. This tells us immediately that both Z_1 and Z_2 are standardized Normal.

But now, there is one extra issue to investigate:

Conditional distribution of $Z_1 \mid Z_2 = \mathbf{z}_2$ (using boldface implies that $\mathbf{z}_2$ is no longer variable, but is assumed to have one specific 'observed' value).

To find the corresponding (univariate!) PDF, we have to do this:

$$
\frac{fzz(z_1,\mathbf{z}_2)}{f_z(\mathbf{z}_2)} = \frac{\exp(-\frac{z_1^2+\mathbf{z}_2^2-2\rho z_1\mathbf{z}_2}{2(1-\rho^2)})}{2\pi\sqrt{1-\rho^2}} \div \frac{\exp(-\frac{\mathbf{z}_2^2}{2})}{\sqrt{2\pi}} = \frac{\exp\left(-\frac{(z_1-\rho\mathbf{z}_2)^2}{2(1-\rho^2)}\right)}{\sqrt{2\pi}\sqrt{1-\rho^2}}
$$

by simple algebra. The result can be identified as $\mathcal{N}(\rho\mathbf{z}_2, \sqrt{1-\rho^2})$, i.e. Normal, with mean of $\rho\mathbf{z}_2$ and standard deviation equal to $\sqrt{1-\rho^2}$ (smaller than what it was marginally, i.e. before we observed Z_2). Note that many textbooks use this notation, but put variance in place of standard deviation.

How do we utilize this result to find the conditional distribution of X given that Y has been observed to have a value of $\mathbf{y}$. Well, we could use the same procedure, but the algebra would get a lot more messier, or we can do this:

We already know that the conditional distribution of $\frac{X-\mu_1}{\sigma_1} \mid \frac{Y-\mu_2}{\sigma_2} = \frac{\mathbf{y}-\mu_2}{\sigma_2}$ is $\mathcal{N}(\rho\frac{\mathbf{y}-\mu_2}{\sigma_2}, \sqrt{1-\rho^2})$. So we have the conditional distribution of $\frac{X-\mu_1}{\sigma_1} \mid Y = \mathbf{y}$ (which is clearly the same thing). Now, using (1), which holds conditionally as well, we find that the conditional distribution of $X \mid Y = \mathbf{y}$ is

$$
\mathcal{N}\left(\mu_1+\sigma_1\rho\frac{\mathbf{y}-\mu_2}{\sigma_2}, \sigma_1\sqrt{1-\rho^2}\right)
$$

Multivariate Normal Distribution

Consider N independent, standardized, Normally distributed random variables. Their joint PDF is

$$
\begin{aligned}
f(z_1, z_2, ..., z_N) &= (2\pi)^{-N/2} \cdot \exp\left(-\frac{\sum\limits_{i=1}^{N} z_i^2}{2}\right) \\
&\equiv (2\pi)^{-N/2} \cdot \exp\left(-\frac{\mathbf{z}^T\mathbf{z}}{2}\right)
\end{aligned}
$$

Corresponding MGF

$$
\exp\left(\frac{\sum\limits_{i=1}^{N} t_i^2}{2}\right) \equiv \exp\left(\frac{\mathbf{t}^T\mathbf{t}}{2}\right)
$$

The following linear transformation

$$
\mathbf{X} = \mathbb{B}\,\mathbf{Z} + \boldsymbol{\mu}
$$

where $\mathbb{B}$ is an arbitrary (regular) N by N matrix, defines a new set of N random variables having a *general* Normal distribution.

The corresponding PDF is

$$
\begin{gathered}
\frac{\left|\det(\mathbb{B}^{-1})\right|}{\sqrt{(2\pi)^N}} \exp\left(-\frac{(\mathbf{x}-\boldsymbol{\mu})^T(\mathbb{B}^{-1})^T\mathbb{B}^{-1}(\mathbf{x}-\boldsymbol{\mu})}{2}\right) \\
\frac{1}{\sqrt{(2\pi)^N \det(\mathbb{V})}} \exp\left(-\frac{(\mathbf{x}-\boldsymbol{\mu})^T\mathbb{V}^{-1}(\mathbf{x}-\boldsymbol{\mu})}{2}\right)
\end{gathered}
$$

since $\mathbf{Z} = \mathbb{B}^{-1}(\mathbf{X}-\boldsymbol{\mu})$, which further implies that

$$
\begin{aligned}
& (\mathbf{X}-\boldsymbol{\mu})^T(\mathbb{B}^{-1})^T\mathbb{B}^{-1}(\mathbf{X}-\boldsymbol{\mu}) \\
= \;& (\mathbf{X}-\boldsymbol{\mu})^T(\mathbb{B}\mathbb{B}^T)^{-1}(\mathbf{X}-\boldsymbol{\mu}) \\
= \;& (\mathbf{X}-\boldsymbol{\mu})^T\mathbb{V}^{-1}(\mathbf{X}-\boldsymbol{\mu})
\end{aligned}
$$

which must thus have the χ^2_N distribution.

The corresponding MGF is

$$
\begin{aligned}
& \mathbb{E}\left\{\exp\left[\mathbf{t}^T(\mathbb{B}\,\mathbf{Z}+\boldsymbol{\mu})\right]\right\} \\
& = \exp\left(\mathbf{t}^T\boldsymbol{\mu}\right) \cdot \exp\left(\frac{\mathbf{t}^T\mathbb{B}\mathbb{B}^T\mathbf{t}}{2}\right) \\
= \;& \exp\left(\mathbf{t}^T\boldsymbol{\mu}\right) \cdot \exp\left(\frac{\mathbf{t}^T\mathbb{V}\,\mathbf{t}}{2}\right)
\end{aligned}
$$

where $\mathbb{V} \equiv \mathbb{B}\mathbb{B}^T$ is the corresponding variance-covariance matrix (must be symmetric and positive definite). This shows each marginal distribution remains Normal, without a change in the corresponding $\boldsymbol{\mu}$ and $\mathbb{V}$ elements.

Note that there are many different $\mathbb{B}$'s resulting in the same $\mathbb{V}$.

To generate a set of normally distributed random variables having a given variance-covariance matrix $\mathbb{V}$ requires us to solve for the corresponding $\mathbb{B}$ (Maple provides us with $\mathbf{Z}$ only, when typing: **stats[random,normald]**(20)). There is infinitely many such $\mathbb{B}$ matrices, one of them (easy to construct) is lower triangular.

Partial correlation coefficient

The variance-covariance matrix can be converted into the correlation matrix:

$$\mathbb{C}_{ij} \equiv \frac{\mathbb{V}_{ij}}{\sqrt{\mathbb{V}_{ii} \cdot \mathbb{V}_{jj}}}$$

The main diagonal elements of $\mathbb{C}$ are all equal to 1 (the correlation of X_i with itself).

Suppose we have three normally distributed random variables with a given variance-covariance matrix. The conditional distribution of X_2 and X_3 given that $X_1 = \underline{x}_1$ has a correlation coefficient independent of the value of $\underline{x}_1$. It is called the partial correlation coefficient, and denoted $\rho_{23\,|1}$. Let us find its value in terms of the ordinary correlation coefficients..

Any correlation coefficient is independent of scaling. We can thus choose the three X's to be standardized (but *not* independent), having the following 3-D PDF:

$$\frac{1}{\sqrt{(2\pi)^3 \det(\mathbb{C})}} \cdot \exp\left(-\frac{\mathbf{z}^T \mathbb{C}^{-1} \mathbf{z}}{2}\right)$$

where

$$\mathbb{C} = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{bmatrix}$$

Since the marginal PDF of z_1 is

$$\frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{z_1^2}{2}\right)$$

the conditional PDF we need is

$$\frac{1}{\sqrt{(2\pi)^2 \det(\mathbb{C})}} \cdot \exp\left(-\frac{\mathbf{z}^T \mathbb{C}^{-1} \mathbf{z} - z_1^2}{2}\right)$$

The information about the five parameters of the corresponding bi-variate distribution is in

$$\mathbf{z}^T\mathbb{C}^{-1}\mathbf{z}-z_1^2=$$

$$\frac{\left(\dfrac{z_2-\rho_{12}z_1}{\sqrt{1-\rho_{12}^2}}\right)^2+\left(\dfrac{z_3-\rho_{13}z_1}{\sqrt{1-\rho_{13}^2}}\right)^2 -2\,\dfrac{\rho_{23}-\rho_{12}\rho_{13}}{\sqrt{1-\rho_{12}^2}\sqrt{1-\rho_{13}^2}}\left(\dfrac{z_2-\rho_{12}z_1}{\sqrt{1-\rho_{12}^2}}\right)\left(\dfrac{z_3-\rho_{13}z_1}{\sqrt{1-\rho_{13}^2}}\right)}{1-\left(\dfrac{\rho_{23}-\rho_{12}\rho_{13}}{\sqrt{1-\rho_{12}^2}\sqrt{1-\rho_{13}^2}}\right)^2}$$

which, in terms of the two conditional means and standard deviations agrees with what we know from MATH 2F81. The extra parameter is our partial correlation coefficient

$$\rho_{23\,|1}=\frac{\rho_{23}-\rho_{12}\cdot\rho_{13}}{\sqrt{1-\rho_{12}^2}\sqrt{1-\rho_{13}^2}}$$

or

$$\rho_{ij\,|k}=\frac{\rho_{ij}-\rho_{ik}\cdot\rho_{jk}}{\sqrt{1-\rho_{ik}^2}\sqrt{1-\rho_{jk}^2}}$$

in general.

To get the conditional mean, standard deviation and correlation coefficient given more than one X has been observed, one can 'iterate' in the following manner:

$$\begin{aligned}
\mu_{i\ |K\ell} &= \mu_{i\ |K}+\sigma_{i\ |K}\rho_{i\ell\ |K}\frac{x_\ell-\mu_{\ell\ |K}}{\sigma_{\ell\ |K}}\\
\sigma_{i\ |K\ell} &= \sigma_{i\ |K}\sqrt{1-\rho_{i\ell\ |K}^2}\\
\rho_{ij|K\ell} &= \frac{\rho_{ij|K}-\rho_{i\ell|K}\cdot\rho_{j\ell|K}}{\sqrt{1-\rho_{i\ell|K}^2}\sqrt{1-\rho_{j\ell|K}^2}}
\end{aligned}$$

etc., where K now represents any number of indices (corresponding to the already observed X's).

A more direct way to find these is presented in the following section.

General conditional distribution:

When the N variables are partitioned into two subsets, say $\mathbf{X}_{(1)}$ and $\mathbf{X}_{(2)}$, with means $\boldsymbol{\mu}_{(1)}$ and $\boldsymbol{\mu}_{(2)}$, and the variance-covariance matrix

$$\begin{bmatrix}\mathbb{V}_{11} & \mathbb{V}_{12}\\ \mathbb{V}_{21} & \mathbb{V}_{22}\end{bmatrix}$$

whose inverse is

$$\mathbb{A}=\begin{bmatrix}(\mathbb{V}_{11}-\mathbb{V}_{12}\mathbb{V}_{22}^{-1}\mathbb{V}_{21})^{-1} & -(\mathbb{V}_{11}-\mathbb{V}_{12}\mathbb{V}_{22}^{-1}\mathbb{V}_{21})^{-1}\mathbb{V}_{12}\mathbb{V}_{22}^{-1}\\ -(\mathbb{V}_{22}-\mathbb{V}_{21}\mathbb{V}_{11}^{-1}\mathbb{V}_{12})^{-1}\mathbb{V}_{21}\mathbb{V}_{11}^{-1} & (\mathbb{V}_{22}-\mathbb{V}_{21}\mathbb{V}_{11}^{-1}\mathbb{V}_{12})^{-1}\end{bmatrix}$$

The conditional PDF of $\mathbf{X}_{(1)}$ given $\mathbf{X}_{(2)} = \underline{\mathbf{x}}_{(2)}$ is obviously

$$\frac{1}{\sqrt{(2\pi)^N \det(\mathbb{V})}} \exp\left(-\frac{(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\mathbb{V}^{-1}(\mathbf{x}-\boldsymbol{\mu})}{2}\right) \div$$
$$\frac{1}{\sqrt{(2\pi)^{N_2} \det(\mathbb{V}_{22})}} \exp\left(-\frac{(\underline{\mathbf{x}}_{(2)}-\boldsymbol{\mu}_{(2)})^{\mathrm{T}}\mathbb{V}_{22}^{-1}(\underline{\mathbf{x}}_{(2)}-\boldsymbol{\mu}_{(2)})}{2}\right)$$

i.e. still Normal. To get the resulting (conditional) variance-covariance matrix, all we need to do is to invert the corresponding block of $\mathbb{A}$, getting

$$\mathbb{V}_{(1,2)} \equiv \mathbb{V}_{11} - \mathbb{V}_{12}\mathbb{V}_{22}^{-1}\mathbb{V}_{21}$$

Similarly, the conditional mean (say $\boldsymbol{\mu}_{(1|2)}$) is found based on

$$-\mathbf{x}_{(1)}^{\mathrm{T}}\mathbb{V}_{(1|2)}^{-1}\boldsymbol{\mu}_{(1|2)} = -\mathbf{x}_{(1)}^{\mathrm{T}}(\mathbb{V}_{11} - \mathbb{V}_{12}\mathbb{V}_{22}^{-1}\mathbb{V}_{21})^{-1}\boldsymbol{\mu}_{(1)} -$$
$$\mathbf{x}_{(1)}^{\mathrm{T}}(\mathbb{V}_{11} - \mathbb{V}_{12}\mathbb{V}_{22}^{-1}\mathbb{V}_{21})^{-1}\mathbb{V}_{12}\mathbb{V}_{22}^{-1}(\underline{\mathbf{x}}_{(2)} - \boldsymbol{\mu}_{(2)})$$

It equals

$$\boldsymbol{\mu}_{(1|2)} = \boldsymbol{\mu}_{(1)} + \mathbb{V}_{12}\mathbb{V}_{22}^{-1}(\underline{\mathbf{x}}_{(2)} - \boldsymbol{\mu}_{(2)})$$

Proof:

$$[\mathbf{x}_{(1)}-\boldsymbol{\mu}_{(1)}-\mathbb{V}_{12}\mathbb{V}_{22}^{-1}(\underline{\mathbf{x}}_{(2)}-\boldsymbol{\mu}_{(2)})]^{\mathrm{T}}(\mathbb{V}_{11}-\mathbb{V}_{12}\mathbb{V}_{22}^{-1}\mathbb{V}_{21})^{-1}[\mathbf{x}_{(1)}-\boldsymbol{\mu}_{(1)}-\mathbb{V}_{12}\mathbb{V}_{22}^{-1}(\underline{\mathbf{x}}_{(2)}-\boldsymbol{\mu}_{(2)})]$$

$$\begin{aligned} = \quad & (\mathbf{x}_{(1)} - \boldsymbol{\mu}_{(1)})^{\mathrm{T}}(\mathbb{V}_{11} - \mathbb{V}_{12}\mathbb{V}_{22}^{-1}\mathbb{V}_{21})^{-1}(\mathbf{x}_{(1)} - \boldsymbol{\mu}_{(1)}) \\ & -(\underline{\mathbf{x}}_{(2)} - \boldsymbol{\mu}_{(2)})^{\mathrm{T}}\mathbb{V}_{22}^{-1}\mathbb{V}_{21}(\mathbb{V}_{11} - \mathbb{V}_{12}\mathbb{V}_{22}^{-1}\mathbb{V}_{21})^{-1}(\mathbf{x}_{(1)} - \boldsymbol{\mu}_{(1)}) \\ & -(\mathbf{x}_{(1)} - \boldsymbol{\mu}_{(1)})^{\mathrm{T}}(\mathbb{V}_{11} - \mathbb{V}_{12}\mathbb{V}_{22}^{-1}\mathbb{V}_{21})^{-1}\mathbb{V}_{12}\mathbb{V}_{22}^{-1}(\underline{\mathbf{x}}_{(2)} - \boldsymbol{\mu}_{(2)}) \\ & +(\underline{\mathbf{x}}_{(2)} - \boldsymbol{\mu}_{(2)})^{\mathrm{T}}\mathbb{V}_{22}^{-1}\mathbb{V}_{21}(\mathbb{V}_{11} - \mathbb{V}_{12}\mathbb{V}_{22}^{-1}\mathbb{V}_{21})^{-1}\mathbb{V}_{12}\mathbb{V}_{22}^{-1}(\underline{\mathbf{x}}_{(2)} - \boldsymbol{\mu}_{(2)}) \end{aligned}$$

Since

$$\mathbb{V}_{21}(\mathbb{V}_{11} - \mathbb{V}_{12}\mathbb{V}_{22}^{-1}\mathbb{V}_{21})^{-1} \equiv \mathbb{V}_{22}(\mathbb{V}_{22} - \mathbb{V}_{21}\mathbb{V}_{11}^{-1}\mathbb{V}_{12})^{-1}\mathbb{V}_{21}\mathbb{V}_{11}^{-1}$$

the last matrix equals

$$\begin{aligned} & (\mathbb{V}_{22} - \mathbb{V}_{21}\mathbb{V}_{11}^{-1}\mathbb{V}_{12})^{-1}\mathbb{V}_{21}\mathbb{V}_{11}^{-1}\mathbb{V}_{12}\mathbb{V}_{22}^{-1} \\ = \quad & (\mathbb{V}_{22} - \mathbb{V}_{21}\mathbb{V}_{11}^{-1}\mathbb{V}_{12})^{-1}(\mathbb{V}_{21}\mathbb{V}_{11}^{-1}\mathbb{V}_{12} - \mathbb{V}_{22} + \mathbb{V}_{22})\mathbb{V}_{22}^{-1} \\ = \quad & (\mathbb{V}_{22} - \mathbb{V}_{21}\mathbb{V}_{11}^{-1}\mathbb{V}_{12})^{-1} - \mathbb{V}_{22}^{-1} \end{aligned}$$

in full agreement with the conditional PDF quoted above.

Finally, to show that $\det(\mathbb{V}) \div \det(\mathbb{V}_{22}) = \det(\mathbb{V}_{11} - \mathbb{V}_{12}\mathbb{V}_{22}^{-1}\mathbb{V}_{21})$, take the determinant of each side of

$$\begin{bmatrix} \mathbb{I} & -\mathbb{V}_{12}\mathbb{V}_{22}^{-1} \\ \mathbb{O} & \mathbb{V}_{22}^{-1} \end{bmatrix} \begin{bmatrix} \mathbb{V}_{11} & \mathbb{V}_{12} \\ \mathbb{V}_{21} & \mathbb{V}_{22} \end{bmatrix} = \begin{bmatrix} \mathbb{V}_{11} - \mathbb{V}_{12}\mathbb{V}_{22}^{-1}\mathbb{V}_{21} & \mathbb{O} \\ \mathbb{V}_{22}^{-1}\mathbb{V}_{21} & \mathbb{I} \end{bmatrix}$$

Estimating μ, ρ and ρ

The standard method of finding good `estimators` of distribution parameters is called `Maximum Likelihood` (ML) technique. We will demonstrate it on the general Normal bivariate case.

First, take the natural logarithm of the PDF of a random independent sample of n pairs of X and Y observations (product of individual PDFs), namely

$$\ln \prod_{i=1}^{n} \frac{\exp\left(-\frac{(\frac{x_i-\mu_1}{\sigma_1})^2+(\frac{y_i-\mu_2}{\sigma_2})^2-2\rho(\frac{x_i-\mu_1}{\sigma_1})(\frac{y_i-\mu_2}{\sigma_2})}{2(1-\rho^2)}\right)}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$$

$$= -\frac{\sum_{i=1}^{n}\left[(\frac{x_i-\mu_1}{\sigma_1})^2+(\frac{y_i-\mu_2}{\sigma_2})^2-2\rho(\frac{x_i-\mu_1}{\sigma_1})(\frac{y_i-\mu_2}{\sigma_2})\right]}{2(1-\rho^2)}$$

$$-n\ln(2\pi)-n\ln\sigma_1-n\ln\sigma_2-\frac{n}{2}\ln(1-\rho^2)$$

Then, replace the x_i and y_i variables by the actual sample values (switch to boldface), and maximize this expression with respect to the μ_1, μ_2, σ_1, σ_2 and ρ parameters (by setting the corresponding derivatives equal to zero):

$$\sum_{i=1}^{n}(\frac{\mathbf{x}_i-\mu_1}{\sigma_1})-\rho\sum_{i=1}^{n}(\frac{\mathbf{y}_i-\mu_2}{\sigma_2})=0$$

$$\sum_{i=1}^{n}(\frac{\mathbf{y}_i-\mu_2}{\sigma_2})-\rho\sum_{i=1}^{n}(\frac{\mathbf{x}_i-\mu_1}{\sigma_1})=0$$

$$\frac{1}{\sigma_1^3}\sum_{i=1}^{n}(\mathbf{x}_i-\mu_1)^2-\frac{\rho}{\sigma_1^2}\sum_{i=1}^{n}(\mathbf{x}_i-\mu_1)(\frac{\mathbf{y}_i-\mu_2}{\sigma_2})=\frac{n}{\sigma_1}(1-\rho^2)$$

$$\frac{1}{\sigma_2^3}\sum_{i=1}^{n}(\mathbf{y}_i-\mu_2)^2-\frac{\rho}{\sigma_2^2}\sum_{i=1}^{n}(\mathbf{y}_i-\mu_2)(\frac{\mathbf{x}_i-\mu_1}{\sigma_1})=\frac{n}{\sigma_2}(1-\rho^2)$$

$$\frac{1}{1-\rho^2}\sum_{i=1}^{n}(\frac{\mathbf{x}_i-\mu_1}{\sigma_1})(\frac{\mathbf{y}_i-\mu_2}{\sigma_2})-\frac{\rho}{(1-\rho^2)^2}\cdot$$

$$\cdot\sum_{i=1}^{n}\left[(\frac{\mathbf{x}_i-\mu_1}{\sigma_1})^2+(\frac{\mathbf{y}_i-\mu_2}{\sigma_2})^2-2\rho(\frac{\mathbf{x}_i-\mu_1}{\sigma_1})(\frac{\mathbf{y}_i-\mu_2}{\sigma_2})\right]$$

$$=-\frac{n\rho}{(1-\rho^2)}$$

The first two equations are solved by making both $\sum_{i=1}^{n}(\mathbf{x}_i-\mu_1)$ and $\sum_{i=1}^{n}(\mathbf{y}_i-\mu_2)$ equal to zero, which means

$$\hat{\mu}_1=\frac{\sum_{i=1}^{n}\mathbf{x}_i}{n}\equiv\overline{\mathbf{x}}$$

$$\hat{\mu}_2 = \frac{\sum_{i=1}^{n} \mathbf{y}_i}{n} \equiv \overline{\mathbf{y}}$$

(the usual sample means).

The next three equations can be re-written as

$$\sum_{i=1}^{n} (\frac{\mathbf{x}_i - \mu_1}{\sigma_1})^2 - \rho \sum_{i=1}^{n} (\frac{\mathbf{x}_i - \mu_1}{\sigma_1})(\frac{\mathbf{y}_i - \mu_2}{\sigma_2}) = n(1-\rho^2)$$

$$\sum_{i=1}^{n} (\frac{\mathbf{y}_i - \mu_2}{\sigma_2})^2 - \rho \sum_{i=1}^{n} (\frac{\mathbf{x}_i - \mu_1}{\sigma_1})(\frac{\mathbf{y}_i - \mu_2}{\sigma_2}) = n(1-\rho^2)$$

$$(1-\rho^2) \sum_{i=1}^{n} (\frac{\mathbf{x}_i - \mu_1}{\sigma_1})(\frac{\mathbf{y}_i - \mu_2}{\sigma_2}) -$$

$$-\rho \sum_{i=1}^{n} \left[(\frac{\mathbf{x}_i - \mu_1}{\sigma_1})^2 + (\frac{\mathbf{y}_i - \mu_2}{\sigma_2})^2 - 2\rho (\frac{\mathbf{x}_i - \mu_1}{\sigma_1})(\frac{\mathbf{y}_i - \mu_2}{\sigma_2}) \right]$$

$$= -n\rho(1-\rho^2)$$

Using the first two equations, the second term of the last equation can be simplified to $-2\rho n(1-\rho^2)$. The last equation thus reads

$$(1-\rho^2) \sum_{i=1}^{n} (\frac{\mathbf{x}_i - \mu_1}{\sigma_1})(\frac{\mathbf{y}_i - \mu_2}{\sigma_2}) = n\rho(1-\rho^2)$$

which implies

$$\rho = \frac{\sum_{i=1}^{n} (\mathbf{x}_i - \mu_1)(\mathbf{y}_i - \mu_2)}{n \; \sigma_1 \sigma_2}$$

Substituting this for the second term of the first two equations makes the second term read $-n\rho^2$. Cancelling with the corresponding term on the right hand side yields

$$\sum_{i=1}^{n} (\frac{\mathbf{x}_i - \mu_1}{\sigma_1})^2 = n$$

$$\sum_{i=1}^{n} (\frac{\mathbf{y}_i - \mu_2}{\sigma_2})^2 = n$$

and the following final estimators:

$$\hat{\sigma}_1 = \sqrt{\frac{\sum_{i=1}^{n} (\mathbf{x}_i - \overline{\mathbf{x}})^2}{n}}$$

$$\hat{\sigma}_2 = \sqrt{\frac{\sum_{i=1}^{n} (\mathbf{y}_i - \overline{\mathbf{y}})^2}{n}}$$

and

$$\hat{\rho} = \frac{\sum_{i=1}^{n} (\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{y}_i - \overline{\mathbf{y}})}{\sqrt{\sum_{i=1}^{n} (\mathbf{x}_i - \overline{\mathbf{x}})^2 \cdot \sum_{i=1}^{n} (\mathbf{y}_i - \overline{\mathbf{y}})^2}}$$

(sample correlation coefficient), more commonly denoted r.

We know that $\overline{X}$, $\overline{Y}$, $\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{\sigma_1^2}$ and $\frac{\sum_{i=1}^{n}(Y_i-\overline{Y})^2}{\sigma_2^2}$ have simple distributions, the distribution of r is a lot more complicated; this is its PDF:

$$\frac{(n-2)\cdot\Gamma(n-1)\cdot(1-\rho^2)^{\frac{n-1}{2}}}{\sqrt{2\pi}\cdot\Gamma(n-\frac{1}{2})}\cdot\frac{(1-r^2)^{\frac{n}{2}-2}}{(1-\rho\ r)^{n-\frac{3}{2}}}\cdot F(\tfrac{1}{2},\tfrac{1}{2};n-\tfrac{1}{2};\tfrac{1+\rho\ r}{2})$$

where F is the hypergeometric function defined by

$$F(a,b;c;x)=1+\frac{a\cdot b}{c\cdot 1!}\ x+\frac{a(a+1)\cdot b(b+1)}{c(c+1)\cdot 2!}\ x^2$$

$$+\frac{a(a+1)(a+2)\cdot b(b+1)(b+2)}{c(c+1)(c+2)\cdot 3!}\ x^3+...$$

This result is due to Hotelling (1953).

The corresponding expected value of r is

$$\frac{\rho\ \Gamma(\frac{n}{2})^2}{\Gamma(\frac{n-1}{2})\Gamma(\frac{n+1}{2})}F(\tfrac{1}{2},\tfrac{1}{2};\tfrac{n+1}{2};\rho^2)$$

so the estimator is clearly biased. Expanded in powers of $\frac{1}{n}$, this becomes

$$\rho-\frac{\rho(1-\rho^2)}{2n}-\frac{3\rho(1-\rho^2)(1+3\rho^2)}{8n^2}...$$

One finds that the distribution of arctanh(r) converges to Normal distribution a lot faster (it has a smaller bias, and a lot smaller skewness).

To derive this PDF, we first need some formulas concerning

n dimensional sphere

of radius r is defined as the following set

$$x_1^2+x_2^2+...+x_n^2\le r^2$$

Its volume $V_n(r)=r^nV_n(1)$ and surface area $S_n(r)$ are clearly related by

$$S_n(r)=\frac{dV_n(r)}{dr}=nr^{n-1}V_n(1)=r^{n-1}S_n(1)$$

It is easier to find S_n from

$$\pi^{n/2}=\int\cdots\int\exp(-x_1^2-x_2^2-...-x_n^2)dx_1dx_2...dx_n=$$

$$\begin{aligned}S_n(1)\int_0^\infty r^{n-1}\exp(-r^2)dr &= \\ \frac{S_n(1)}{2}\int_0^\infty y^{n/2-1}\exp(-y)dy &= \frac{S_n(1)}{2}\Gamma\left(\frac{n}{2}\right)\end{aligned}$$

which implies that

$$S_n(1) = \frac{2\pi^{n/2}}{\Gamma\left(\frac{n}{2}\right)}$$

$$S_n(r) = \frac{2\pi^{n/2} r^{n-1}}{\Gamma\left(\frac{n}{2}\right)}$$

and

$$V_n(r) = \frac{2\pi^{n/2} r^n}{n\Gamma\left(\frac{n}{2}\right)}$$

Sample from standardized bi-variate Normal distribution

has a PDF given by

$$(2\pi)^{-n}(1-\rho^2)^{-n/2} \exp\left(-\frac{\sum(x_i^2 + y_i^2 - 2\rho x_i y_i)}{2(1-\rho^2)}\right) dx_1 ... dy_n =$$
$$(2\pi)^{-n}(1-\rho^2)^{-n/2} \cdot$$
$$\exp\left(-\frac{n(s_1^2 + \bar{x}^2 + s_2^2 + \bar{y}^2 - 2\rho r s_1 s_2 - 2\rho\bar{x}\bar{y})}{2(1-\rho^2)}\right) dx_1 ... dy_n$$

where

$$\begin{aligned} \bar{x} &= \frac{\sum x_i}{n} \quad \bar{y} = \frac{\sum y_i}{n} \\ s_1^2 &= \frac{\sum(x_i - \bar{x})^2}{n} \quad s_2^2 = \frac{\sum(y_i - \bar{y})^2}{n} \\ r &= \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n s_1 s_2} \end{aligned}$$

All we have to do now is to figure out the volume of the $2n$ dimensional region filled by taking each of the five quantities s_1, s_2, $\bar{x}$, $\bar{y}$ and r, and increasing them by ds_1, ds_2, $d\bar{x}$, $d\bar{y}$ and dr respectively.

We know that

$$\frac{\sum x_i}{\sqrt{n}} = \sqrt{n}\bar{x}$$

is a plane in the 'first' n dimensions (the x-space), going through ($\bar{x}$, $\bar{x}$, ...$\bar{x}$) and having $\frac{(1,1,...1)}{\sqrt{n}}$ as normal. Similarly

$$\sum(x_i - \bar{x})^2 = n s_1^2$$

is an n dimenstional sphere centered on ($\bar{x}$, $\bar{x}$, ...$\bar{x}$), with the radius of $s_1\sqrt{n}$. The plane cuts the sphere in 'the middle', thus creating (the corresponding cross section) an $n-1$ dimensional sphere, whose surface is

$$\frac{2\pi^{(n-1)/2} s_1^{n-2} n^{(n-2)/2}}{\Gamma\left(\frac{n-1}{2}\right)}$$

This needs to be further multiplied (since the $\sqrt{n}ds_1$ and $\sqrt{n}d\bar{x}$ directions are perpendicular) by $n\ d\bar{x}\ ds_1$.

Similar argument can be made in the y space, except now we need to keep the $(y_1-\bar{y}, y_2-\ \bar{y},\ ...y_n-\bar{y})$ vector at a fixed angle to $(x_1-\bar{x}, x_2-\ \bar{x},\ ...x_n-\bar{x})$, an angle whose cos is r. This reduces the $n-1$ dimensional shperical surface of radius $s_2\sqrt{n}$ to an $n-2$ dimensional spherical surface of radius $s_2\sqrt{n(1-r^2)}$. This then contributes

$$\frac{2\pi^{(n-2)/2} s_1^{n-3} n^{(n-3)/2}(1-r^2)^{(n-3)/2}}{\Gamma\left(\frac{n-2}{2}\right)} n\ d\bar{y}\ ds_2\ \frac{s_2\sqrt{n}}{\sqrt{1-r^2}}dr$$

So now we have

$$\frac{n^n(1-r^2)^{(n-4)/2} s_1^{n-2} s_2^{n-2}}{2^{n-2}\pi^{3/2}(1-\rho^2)^{n/2}\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{n-2}{2}\right)}$$
$$\cdot \exp\left(-\frac{n(s_1^2+\bar{x}^2+s_2^2+\bar{y}^2-2\rho r s_1 s_2-2\rho\bar{x}\bar{y})}{2(1-\rho^2)}\right) d\bar{x}d\bar{y}ds_1 ds_2 dr$$

(R. A. Fisher, 1915).

The $d\bar{x}d\bar{y}$ integration (separable, from $-\infty$ to ∞ each) can be carried out easily, yielding

$$\frac{n^{n-1}(1-r^2)^{(n-4)/2} s_1^{n-2} s_1^{n-2}}{\pi(1-\rho^2)^{(n-1)/2}(n-3)!}\cdot$$
$$\exp\left(-\frac{n(s_1^2+s_2^2-2\rho r s_1 s_2)}{2(1-\rho^2)}\right) ds_1 ds_2 dr$$

(we have also replaced $2^{n-3}\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{n-2}{2}\right)$ by $(n-3)!\sqrt{\pi}$).

To integrate over s_1 and s_2, we realize that the $s_1 > s_2$ and $s_1 < s_2$ regions must contribute the same amount (the function is $s_1 \leftrightarrow s_2$ symmetric), so we multiply the integrand by 2 and integrate over the latter.

Introducing the following one-to-one transformation (ξ goes from 0 to ∞, η from 1 to infinity):

$$\begin{aligned}\xi &= s_1 s_2\\ \eta &= \frac{1}{2}\left(\frac{s_1}{s_2}+\frac{s_2}{s_1}\right) = \frac{s_1^2+s_2^2}{2s_1 s_2}\end{aligned}$$

whose Jacobian is

$$\det\begin{bmatrix} s_2 & s_1 \\ \frac{1}{2}\left(\frac{1}{s_2}-\frac{s_2}{s_1^2}\right) & \frac{1}{2}\left(\frac{1}{s_1}-\frac{s_1}{s_2^2}\right)\end{bmatrix} =$$
$$\left(\frac{s_2}{s_1}-\frac{s_1}{s_2}\right) = 2\sqrt{\eta^2-1}$$

results in

$$\frac{n^{n-1}(1-r^2)^{(n-4)/2}\xi^{n-2}}{\pi(1-\rho^2)^{(n-1)/2}(n-3)!}\exp\left(-\frac{n(\eta-\rho r)\xi}{1-\rho^2}\right)\frac{d\xi d\eta}{\sqrt{\eta^2-1}}dr$$

which can be easily integrated over ξ, getting

$$\frac{(n-2)(1-r^2)^{(n-4)/2}}{\pi(1-\rho^2)^{(n-1)/2}} \cdot \left(\frac{1-\rho^2}{\eta-\rho r}\right)^{n-1} \frac{d\eta}{\sqrt{\eta^2-1}}dr$$

And the last substitution is

$$\eta = \frac{1-\rho r z}{1-z}$$

(z going from 0 to 1) replaces $\eta - \rho r$ by

$$\frac{1-\rho r}{1-z}$$

$d\eta$ by

$$\frac{1-\rho r}{(1-z)^2}dz$$

and $\sqrt{\eta^2-1}$ by

$$\frac{\sqrt{2z(1-\rho r)} \cdot \sqrt{1-\frac{1+\rho r}{2}z}}{1-z}$$

getting

$$\frac{(n-2)(1-r^2)^{(n-4)/2} z^{-1/2}\left(1-\frac{1+\rho r}{2}z\right)^{-1/2}}{\sqrt{2}\pi}$$
$$\cdot \frac{(1-\rho^2)^{(n-1)/2}(1-z)^{n-2}}{(1-\rho r)^{n-3/2}} dz\ dr$$

Maple tells us that

$$\int_0^1 z^k(1-z)^m(1-az)^\ell dz =$$
$$\frac{\Gamma(k+1)\Gamma(m+1)}{\Gamma(k+m+2)} F(-\ell, k+1; m+k+2; a)$$

which in our case means

$$\frac{\Gamma(\frac{1}{2})\Gamma(n-1)}{\Gamma(n-\frac{1}{2})} F(\tfrac{1}{2}, \tfrac{1}{2}; n-\tfrac{1}{2}; \tfrac{1+\rho r}{2})$$

The final answer is thus

$$\frac{(n-2)(n-2)!(1-\rho^2)^{(n-1)/2}}{\sqrt{2\pi}\Gamma(n-\frac{1}{2})} \cdot$$
$$\frac{(1-r^2)^{(n/2-2)}}{(1-\rho r)^{n-3/2}} F(\tfrac{1}{2}, \tfrac{1}{2}; n-\tfrac{1}{2}; \tfrac{1+\rho r}{2})dr$$