<table>
<thead>
<tr>
<th>Title</th>
<th>Analytical derivation of sampling distribution of intraclass correlation coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Ogawa, Junjiro</td>
</tr>
<tr>
<td>Citation</td>
<td>Osaka Mathematical Journal. 4(1) P.69–P.76</td>
</tr>
<tr>
<td>Issue Date</td>
<td>1952</td>
</tr>
<tr>
<td>Text Version</td>
<td>publisher</td>
</tr>
<tr>
<td>URL</td>
<td><a href="https://doi.org/10.18910/7306">https://doi.org/10.18910/7306</a></td>
</tr>
<tr>
<td>DOI</td>
<td>10.18910/7306</td>
</tr>
<tr>
<td>Note</td>
<td></td>
</tr>
</tbody>
</table>

Osaka University Knowledge Archive: OUKA
https://ir.library.osaka-u.ac.jp/repo/ouka/all/

Osaka University
Analytical Derivation of Sampling Distribution of Intraclass Correlation Coefficient

By Junjiro Ogawa

The exact sampling distribution of the intraclass correlation coefficient was obtained by R. A. Fisher in 1921. However, his method of derivation is geometrical and too intuitive, so that it is not always easy for every one to comprehend the proof, at least so far as the writer knows. For this reason, it seems to me that, at least from the pedagogical point of view, the analytical derivation of sampling distribution of intraclass correlation coefficient is not at all valueless.

We shall first consider the case when the number of individuals belonging to the family is two. The probability element of the population distribution is given as follows:

\[
df = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)} \left\{ (x-m)^2 - 2\rho(x-m)(x'-m) + (x'-m)^2 \right\}} dx \, dx'.
\]

where \(m, \sigma,\) and \(\rho\) are the mean, standard deviation and intraclass correlation coefficient of the population respectively. The necessary statistics will be defined as follows:

\[
\begin{align*}
nm_2 &= \sum_{i=1}^{n} (x_i - \bar{x})^2, \quad nm_{12} = \sum_{i=1}^{n} (x_i' - \bar{x}')^2, \quad nm_{11} = \sum_{i=1}^{n} (x_i - \bar{x})(x_i' - \bar{x}'), \\
n\bar{x} &= \sum_{i=1}^{n} x_i, \quad n\bar{x}' = \sum_{i=1}^{n} x_i', \\
2n\bar{x} &= \sum_{i=1}^{n} (x_i + x_i'), \\
2n\bar{x}' &= \sum_{i=1}^{n} (x_i + x_i'), \\
nm_2 &= \sum_{i=1}^{n} \left\{ (x_i - \bar{x})^2 + (x_i' - \bar{x})^2 \right\},
\end{align*}
\]

and

\[
nm_2^2 r = \sum_{i=1}^{n} (x_i - \bar{x})(x_i' - \bar{x})
\]

where \(r\) is the intraclass correlation coefficient of the sample and our purpose is to obtain the exact sampling distribution of \(r\).

It is readily seen that
Further, as are well known, the vector variables \((x, x')\) and \((m_{20}, m_{11}, m_{01})\) are mutually independent, and their probability elements are

\[
\frac{n}{2\pi \sigma^2 \sqrt{1 - \rho^2}} e^{-\frac{1}{2\sigma^2 (1 - \rho^2)} \left\{ (x-m)^2 - 2\rho(x-m)(x'-m) + (x'-m)^2 \right\}} \, dx \, dx' \]

and

\[
\frac{n^{n-1}}{4\pi (n-2) \sigma^2 (n-1) (1 - \rho^2)^{n-1}} e^{-\frac{1}{2\sigma^2 (1 - \rho^2)} \left( m_{20} + m_{02} - 2\rho m_{11} \right)} \, dm_{20} \, dm_{02} \, dm_{11}\]

respectively. The probability element of the joint distribution of \(\bar{x}, \bar{x}', m_{20}, m_{02}, m_{11}\) is proportional to

\[
(m_{20} m_{02} - m_{11}^2)^{n-1} e^{-\frac{1}{2\sigma^2 (1 - \rho^2)} \left\{ (x-m)^2 - 2\rho(x-m)(x'-m) + (x'-m)^2 + m_{20} + m_{02} - 2\rho m_{11} \right\}} \, d\bar{x} \, d\bar{x}' \, dm_{20} \, dm_{02} \, dm_{11} .
\]

Since

\[
(x-m)^2 - 2\rho(x-m)(x'-m) + (x'-m)^2 + m_{20} + m_{02} - 2\rho m_{11} = 2(1 - \rho)(\bar{x} - m)^2 + 2\rho^2 (1 - \rho^2),
\]

the exponential factor in (1) becomes

\[
e^{-\frac{n}{\sigma^2 (1 + \rho)} (x-m)^2 - \frac{n(1 - \rho^2)}{\sigma^2 (1 - \rho^2)} \mu^2} .
\]

Now the following transformation of variables is effected.

\[
\begin{align*}
\xi &= \bar{x} \\
\bar{x} &= \frac{1}{2} \bar{x} + \frac{1}{2} \bar{x}' \\
\eta &= m_{20} \\
2\mu^2 &= \frac{1}{2} (\bar{x} - \bar{x}')^2 + m_{20} + m_{02} \\
\mu^2 r &= -\frac{1}{4} (\bar{x} - \bar{x}')^2 + m_{11}
\end{align*}
\]
Analytical Derivation of Sampling Distribution of Intraclass Correlation Coefficient

\[
\frac{\partial (\tilde{x}, \tilde{x}', m_{20}, m_{02}, m_{11})}{\partial (\xi, \bar{x}, \eta, \mu, r)} = 8\mu^4, \quad (4)
\]

the probability element of the joint distribution of \(\xi, \tilde{x}, \eta, \mu, r\) is proportional to

\[
(m_{20}m_{02}-m_{11}^2)^{\frac{n-4}{2}} e^{-\frac{n}{2\sigma^2(1-\rho)}(\tilde{x}-m)^2 - \frac{n(1-\rho)}{\sigma^2(1-\mu^2)}\mu^2} \mu^3 \text{d}x \text{d}\eta \text{d}\mu \text{d}r. \quad (5)
\]

Now, since

\[
m_{20}m_{02}-m_{11}^2 = \eta \left\{ 2\mu^2 - \eta - \frac{1}{2} (\bar{x} - \tilde{x}')^2 \right\} - \left\{ \mu^2 + \frac{1}{4}(\bar{x} - \tilde{x}')^2 \right\}^2
\]

\[
= \mu^4(1-r^2) - \frac{1}{2} \mu^2(1+r)(\bar{x} - \tilde{x}')^2 - \left\{ \eta - \mu^2 + \frac{1}{4}(\bar{x} - \tilde{x}')^2 \right\}^2
\]

\[
= \mu^4(1-r^2) \left[ 1 - 2\left( \frac{\xi - \tilde{x}}{\mu^2(1-r)} \right)^2 \right] \left[ 1 - \left\{ \frac{\eta - \mu^2 + (\xi - \tilde{x})^2}{\mu^2(1-r)} \right\}^2 \right]
\]

it follows at once that

\[
(m_{20}m_{02}-m_{11}^2)^{\frac{n-4}{2}} = \mu^{2n-8}(1-r^2)^{\frac{n-4}{2}} \left( 1 - 2\left( \frac{\xi - \tilde{x}}{\mu^2(1-r)} \right)^2 \right)^{n-4}
\]

\[
\times \left[ 1 - \left\{ \frac{\eta - \mu^2 + (\xi - \tilde{x})^2}{\mu^2(1-r)} \right\}^2 \right]^{n-4}
\]

Putting

\[
u = \frac{\eta - \mu^2 + (\xi - \tilde{x})^2}{\mu^2(1-r)}
\]

the contribution of the integral with respect to \(\eta\) is

\[
\int \left[ 1 - \left\{ \frac{\eta - \mu^2 + (\xi - \tilde{x})^2}{\mu^2(1-r)} \right\}^2 \right]^{n-1} \text{d}\eta = \mu^2 \sqrt{(1-r^2)} \left( 1 - 2\left( \frac{\xi - \tilde{x}}{\mu^2(1-r)} \right)^2 \right)
\]

\[
\text{hence, integrating out } \eta \text{ from (5), we get}
\]

\[
e^{-\frac{n}{\sigma^2(1+\rho)}(\tilde{x}-m)^2 - \frac{n(1-\rho)}{\sigma^2(1-\mu^2)}\mu^2} \cdot \mu^{2n-3}(1-r^2)^{\frac{n-3}{2}} \left( 1 - 2\left( \frac{\xi - \tilde{x}}{\mu^2(1-r)} \right)^2 \right)^{n-3} \text{d}x \text{d}\xi \text{d}\mu \text{d}r. \quad (7)
\]

In a similar manner, integrating out \(\xi\) from (7) we get
Thence we get the probability element of the joint distribution of \( \mu \) and \( r \) which is proportional to

\[
e^{-\frac{n}{\sigma^2(1-r)}} e^{-\frac{n(1-r)}{\sigma^2(1-r^2)}} \cdot \mu^{n-2}(1+r)^{n-3} \frac{1}{2} \frac{1}{2} \frac{d\mu}{dr}. \tag{8}
\]

Integrating out \( \mu \) from (9), we finally get the probability element of \( r \) which is proportional to

\[
(1-\rho r)^{-\left(n-\frac{1}{2}\right)} (1-r)^{n-2} \frac{1}{2} (1+r)^{n-3} \frac{dr}. \tag{10}
\]

When the number \( s \) of the individuals belonging to the family is more than two, such an elementary method as above is useless. In this case the probability element of the population distribution is given by

\[
\exp \left\{ -\frac{1}{2\sigma^2(1-r)(1+s-1\rho)} \left\{ (1+s-2\rho) \cdot \sum_{a=1}^{s} (x^{(a)2} - m^2) \right\} \right\} d\bar{x}^{(1)} \ldots d\bar{x}^{(s)}, \tag{11}
\]

omitting the normalizing constant.

If we make use of the following statistics

\[
n\bar{x}^{(a)} = \sum_{i=1}^{n} x^{(a)2},
\]

\[
n\bar{u}_{\alpha\beta} = \sum_{i=1}^{n} \left( x^{(a)} - \bar{x}^{(a)} \right) \left( x^{(\beta)} - \bar{x}^{(\beta)} \right),
\]

the necessary statistics are given as follows:

\[
s \bar{\bar{x}} = \sum_{a=1}^{s} \bar{x}^{(a)},
\]

\[
s \mu^2 = \frac{1}{n} \sum_{i=1}^{n} \sum_{a=1}^{s} (x^{(a)} - \bar{x})^2 = \sum_{a=1}^{s} l_{aa} + \sum_{a=1}^{s} (\bar{x}^{(a)} - \bar{x})^2,
\]

\[
\left( \frac{s}{2} \right) \mu^2 r = \frac{1}{n} \sum_{a<\beta}^{n} \sum_{i=1}^{n} \left( x^{(a)2} - \bar{x} \right) \left( x^{(\beta)} - \bar{x} \right) = \sum_{a<\beta} l_{a\beta} + \sum_{a<\beta} (\bar{x}^{(a)} - \bar{x}) (\bar{x}^{(\beta)} - \bar{x}). \tag{12}
\]

The vector variables \( (\bar{x}^{(1)}, \ldots, \bar{x}^{(s)}) \) and \( (l_{11}, l_{22}, \ldots, l_{ss}, l_{12}, \ldots, l_{ss}) \) are mutually independent and their probability elements are respectively proportional to
Analytical Derivation of Sampling Distribution of Intraclass Correlation Coefficient

\[
\exp \left[-\frac{\sigma^2(1-\rho)}{(1+s-1-k)} \left\{ \sum_{a=1}^{s}(\bar{x}_{(a)} - \bar{x})^2 \right\} \right] d\bar{x}^{(1)} \ldots d\bar{x}^{(k)},
\]

and to

\[
|l_{a\beta}|^{-1} \cdot \exp \left[-\frac{2\sigma^2(1-\rho)}{(1+s-1-k)} \left\{ \sum_{a=1}^{s}l_{aa} - 2\rho \sum_{a \neq \beta} l_{a\beta} \right\} \right] \prod_{a} dl_{aa} \prod_{a \neq \beta} dl_{a\beta}.
\]

It is readily seen that

\[
\begin{align*}
\mu_{x}^2 &= \sum_{a=1}^{s} l_{aa} + \sum_{a=1}^{s} (\bar{x}_{(a)} - \bar{x})^2, \\

s(s-1)\mu_{x}^2r &= \sum_{a \neq \beta} l_{a\beta} + \sum_{a \neq \beta} (\bar{x}_{(a)} - \bar{x})(\bar{x}_{(\beta)} - \bar{x}).
\end{align*}
\]

We shall calculate the moment generating function \(\varphi(t, u)\) of the joint distribution of \(\mu_{x}^2\) and \(\mu_{x}^2r\). Since

\[
\varphi(t, u) = E(e^{t\mu_{x}^2 + u\mu_{x}^2r}) = E(e^{t\mu_{x}^2 + u\mu_{x}^2r}),
\]

putting

\[
\xi_{(a)} = \bar{x}_{(a)} - \bar{x}, \quad s\bar{x} = \sum_{a=1}^{s} \xi_{(a)},
\]

and

\[
D = \sigma^2(1-\rho)(1+s-1-k),
\]

we have

\[
\varphi(t, u) \propto \left\{ \begin{array}{c}
\exp \left[ -\frac{n}{2} \left\{ \sum_{a=1}^{s} \xi_{(a)}^2 - 2p \cdot \sum_{a \neq \beta} \xi_{(a)} \xi_{(\beta)} \right\} - \frac{2D}{n} \cdot \sum_{a=1}^{s} l_{aa} - 2p \cdot \sum_{a \neq \beta} l_{a\beta} \right\} \\
\prod_{a} dl_{aa} \prod_{a \neq \beta} dl_{a\beta}
\end{array} \right\} \prod_{a} dl_{aa} \prod_{a \neq \beta} dl_{a\beta}
\]

\[
\times \left\{ \begin{array}{c}
\exp \left[ -\frac{n}{2} \left\{ \sum_{a=1}^{s} \xi_{(a)}^2 - 2p \cdot \sum_{a \neq \beta} \xi_{(a)} \xi_{(\beta)} \right\} - \frac{2D}{n} \cdot \sum_{a=1}^{s} l_{aa} - 2p \cdot \sum_{a \neq \beta} l_{a\beta} \right\} \\
\prod_{a \neq \beta} dl_{a\beta}
\end{array} \right\} \prod_{a} dl_{aa} \prod_{a \neq \beta} dl_{a\beta}
\]

Whence we get
Consequently we have

$$q(t, u) \propto \left(1 + s - 1 \cdot \rho - \frac{2D}{n} \frac{1}{s} \left(t - \frac{u}{s} - 1\right)\right)^{-\frac{m(s-1)}{2}} \cdot \left(1 - \rho - \frac{2D}{n} \frac{1}{s}\left(t + u\right)\right)^{-\frac{n-1}{2}}.$$

(15)

By the famous inversion formula due to P. Lévy, we shall have the density function $f(x, y)$ of the joint distribution of $\mu^2$ and $\mu^2 r$ as

$$f(x, y) \propto \lim_{\substack{n \to \infty \atop t \to x}} \int_{-\infty}^{\infty} e^{-itx - iuy} \cdot q(it, iu) dt du.$$  

(16)

Putting

$$t + u = \xi, \quad t - \frac{u}{s-1} = \eta,$$

(16) becomes

$$f(x, y) \propto \lim_{n \to \infty} \left\{ \int_{-\infty}^{x} \frac{1}{s} e^{-\frac{s}{2} \left(x + s - 1 \cdot y\right) i\xi} \right\} \left\{ \int_{-\infty}^{\infty} e^{-\frac{s}{2} \left(\xi - 1\right)^2} d\xi \right\} \left\{ \int_{-\infty}^{\infty} e^{-\frac{s}{2} \left(x - y\right) i\eta} \right\} \left\{ \int_{-\infty}^{\infty} e^{-\frac{s}{2} \left(\eta + \frac{ns}{2D} (1 - s - 1 \cdot \rho)\right) i\eta} \right\}.$$  

(17)
Now, we consider the function
\[ \psi(z) = e^{-\frac{1}{s}(x+s-1)yz} \times \left( -z + \frac{ns}{2D} (1-\rho) \right)^{-\frac{n-1}{2}}, \] 
which has branch points at 
\[ z = \frac{ns}{2D} (1-\rho) \text{ and } z = \infty \]
and otherwise regular unless \( \frac{n-1}{2} \) is an integer. Hence by the Cauchy's integral theorem, we have (see Fig. 1)
\[ \int_{-iT}^{iT} \psi(z)\,dz + \int_{c_1} \psi(z)\,dz + \int_{c_2} \psi(z)\,dz = -\int_{c} \psi(z)\,dz. \]
But on the contour \( C_i \), we see easily that
\[ \left| \int_{c_i} \psi(z)\,dz \right| \leq \int_{c_i} \left| \psi(z) \right|\,dz \leq \frac{\pi}{2} T^{\frac{n}{4}} \left( \frac{ns}{2D} (1-\rho) T + \frac{ns^2}{4D^2} (1-\rho)^2 \right)^{-\frac{n-1}{4}} \frac{1}{s} |x+s-1| y, T. \]
It follows that when \( T \to \infty \),
\[ \int_{c_i} \psi(z)\,dz \to 0, \text{ for } i = 1, 2. \]
Consequently we have
\[ \lim_{r \to \infty} \int_{-r}^{r} e^{-\frac{1}{s}(x+s-1)yz} \left( -z + \frac{ns}{2D} (1-\rho) \right)^{-\frac{n-1}{2}} \,dz = i \int_{1/V_{\infty}} \psi(z)\,dz, \] 
which becomes after translating the origin and changing the scale,
\[ = i e^{-\frac{n}{2D} (1-\rho)(x+s-1)y} \times (x+s-1)^{\frac{n-1}{2}} \times \left( \frac{1}{s} \right)^{\frac{n-1}{2}} \times \int_{1/V_{\infty}} e^{-\frac{s}{2}(x+s-1)y} \,dz. \]
Here the path of integration $\Gamma$ is as shown in the following:

\[ \int_{-\infty}^{\infty} \]

Hence we have

\[
\lim_{r \to \infty} \int_{-r}^{r} \frac{e^{-\frac{1}{s}(x+s-1 \cdot y)/s}}{2D} \left( \frac{n}{2} \right)^{n-1} d\xi
\]

\[
= \frac{2\pi}{\Gamma \left( \frac{n-1}{2} \right)} \left( \frac{1}{s} \right)^{n-3} (x+s-1 \cdot y)^{n-3} \cdot e^{-\frac{n}{2D}(1+s-1 \cdot x+s-1 \cdot y)}.
\]

In a similar manner we have

\[
\lim_{r \to \infty} \int_{-r}^{r} \frac{e^{-\frac{s-1}{s}(x-y)/s}}{2D} \left( \frac{n}{2} \right)^{n-1} d\eta
\]

\[
= \frac{2\pi}{\Gamma \left( \frac{n(s-1)}{2} \right)} \left( \frac{s-1}{s} \right)^{n(s-1)-2} (x-y)^{n(s-1)-2} \cdot e^{-\frac{(s-1)n}{2D}(1+s-1 \cdot x+s-1 \cdot y)}.
\]

Consequently the required density $f(x, y)$ is proportional to

\[
e^{-2\pi(1-\rho)(1+s-1 \cdot \mu)^2} (x+s-1 \cdot y)^{n-3} (x-y)^{n(s-1)-2},
\]

hence the probability element of the joint distribution of $\mu$ and $r$ is proportional to

\[
e^{-\frac{n}{2}(1+s-1 \cdot \mu)^2} \cdot \mu^{n-3} \cdot \frac{n(s-1)-2}{2} \frac{n-3}{2} \cdot (1-r)^2 d\mu dr.
\]

Whence it follows that the probability element of $r$ is proportional to

\[
(1+s-2 \cdot \rho-s-1 \cdot \rho r)^{n(s-1)-2}(1+r)^{n-3}(1-r)^{n(s-1)-2} dr,
\]

which was to be proved.

(Received December 14, 1951)

Reference


2) When $\frac{n-1}{2}$ is an integer it will be seen that the calculation of residue is sufficient for our purpose.