# Supplemental Data

# Heterogeneity of Large Macromolecular Complexes Revealed by 3-D Cryo-EM Variance Analysis

**Wei Zhang, Marek Kimmel, Christian M. T. Spahn, and Pawel A. Penczek**

## SUPPLEMENTAL EXPERIMENTAL PROCEDURES

### The Expected Value of the Sample Correlation Coefficient $\rho_B$ between Two Independent Bootstrap Variance Correlations and Correlation Coefficient Between Bootstrap Variance and Distribution Variance

#### Introduction

In this section, we derive the equations for the sample correlation coefficient of bootstrap variance ($\rho_B$) and correlation coefficient between bootstrap variance and the variance of the distribution ($\rho_\gamma$). In order to have a workable paradigm, we will use a model case of a set of one dimensional random vectors filled with independent random variables, and we will consider averaging of the vectors instead of performing the three dimensional reconstruction. We will relate these results to the case of cryo-EM structure reconstruction.

#### Notation

For a one-dimensional vector $\mathbf{y}$, $y_k$ is its *k*-th element. The average of elements of $\mathbf{y}$ is denoted: $\bar{y} = \frac{1}{K}\sum_{k}^{K} y_k$ (if elements of $\mathbf{y}$ were to represent pixels of an image, we would refer to $\bar{y}$ as *between-pixels average*). The sample variance of elements of $\mathbf{y}$ is denoted $S^2(\mathbf{y}) = \frac{1}{K-1}\left(\sum_{k=1}^{K} y_k^2 - K\cdot\bar{y}^2\right)$ (again, it can be referred to as *between-pixels variance*). By $\mathrm{E}(\cdot)$ we denote the expected value of a random variable; for random vectors, the operation is understood component-wise and results in a vector. By $\mathrm{V}(\cdot)$ we denote the variance of a random variable and by $\mathrm{Cov}(\cdot,\cdot)$ the co-variance of two random variables.

#### Definition of the problem for the model case

Our sample set comprises *N* one-dimensional vectors $\mathbf{\Psi} = \{\mathbf{\Psi}_1, \mathbf{\Psi}_2, \ldots, \mathbf{\Psi}_N\}$, each one-dimensional vector has *K* elements. By $\Psi_{ik}$ we denote the *k*-th element of the *i*-th vector in $\mathbf{\Psi}$. In our model, the elements in $\mathbf{\Psi}$ are independent random variables with zero expected values. Also, we assume that for each *K*, the elements $\Psi_{1k}, \Psi_{2k}, \cdots, \Psi_{Nk}$ are identically distributed. We define $m_{2k}$ and $m_{4k}$ as the 2$^{nd}$ and 4$^{th}$ moment of the distribution of the elements $\Psi_{1k}, \Psi_{2k}, \cdots, \Psi_{Nk}$:

$$
\begin{aligned}
&\mathrm{E}(\Psi_{ik}) = \mathrm{E}(\Psi_{jk}) = 0,\\
&\mathrm{E}(\Psi_{ik}^2) = \mathrm{E}(\Psi_{jk}^2) = m_{2k},\\
&\mathrm{E}(\Psi_{ik}^4) = \mathrm{E}(\Psi_{jk}^4) = m_{4k},\\
&\mathrm{E}(\Psi_{ik}\Psi_{jl}) = m_{2k}\delta_{ij}\delta_{kl}.
\end{aligned}
\tag{S.1}
$$

It should be noted here that we do not need assumption for the 3rd moments since in our derivation they are always multiplied by the 1st moments that are equal to zero.

Therefore, the distribution variances in our model are equal to:

$$\mathrm{V}\left(\Psi_{jk}\right)=\sigma_k^2=m_{2k},\left(j=1,\cdots,N\right). \tag{S.2}$$

We will use a convention here by which a vector of variances (or sample variances) is denoted by corresponding bold face symbol, such as $\boldsymbol{\sigma}^2=\left(\sigma_1^2,\cdots,\sigma_K^2\right)$ or $\mathbf{S}^2=\left(S_1^2,\cdots,S_K^2\right)$.

During the bootstrap calculation, the sample set $\boldsymbol{\Psi}$ is resampled with replacement to generate a new set called a bootstrap sample {Shao, 1995 #9989}. For each bootstrap sample, its sample average is calculated component-wise. The procedure is applied repeatedly and yields *B* bootstrap sample averages, which in turn are used to calculate the *bootstrap sample variance* ($S_k^{*2}$ for each pixel *k*, where $k=1,2,...,K$). We prove in Section B of this Supplement that the expected value, variance and co-variance of $S_k^{*2}$ are equal to:

$$\begin{aligned}
&\mathrm{E}\left(S_k^{*2}\right)=\frac{N-1}{N^2}m_{2k}\cong\frac{1}{N}m_{2k},\\
&\mathrm{V}\left(S_k^{*2}\right)=\frac{m_{4k}-m_{2k}^2}{N^3}+\frac{2m_{2k}^2}{N^2B},\\
&\mathrm{Cov}\left(S_{k1}^{*2},S_{k2}^{*2}\right)=\frac{m_{4k}-m_{2k}^2}{N^3}.
\end{aligned} \tag{S.3}$$

$S_{k1}^{*2}$ and $S_{k2}^{*2}$ are bootstrap variances obtained from two independent bootstrap calculations using the same sample set.

**Derivation of $\mathrm{E}\left(\rho_B\right)$ and $\mathrm{E}\left(\rho_r\right)$ for the model case**

The sample correlation coefficient of bootstrap variance ($\rho_B$) is defined based on the following setup. After we obtained the *B* bootstrap sample averages, we randomly split them into two equally sized groups, and for each group we calculate the bootstrap variance. The two obtained bootstrap variances are denoted as $\mathbf{S}_o^{*2}$ and $\mathbf{S}_e^{*2}$, and the sample correlation coefficient of bootstrap variance ($\rho_B$) is calculated as the correlation coefficient between vectors $\mathbf{S}_o^{*2}$ and $\mathbf{S}_e^{*2}$:

$$\rho_B=\frac{\sum_k^K S_{ko}^{*2}S_{ke}^{*2}-K\bar{S}_o^{*2}\bar{S}_e^{*2}}{(K-1)\sqrt{S^2\left(\mathbf{S}_o^{*2}\right)S^2\left(\mathbf{S}_e^{*2}\right)}}, \tag{S.4}$$

while $S^2\left(\mathbf{S}_o^{*2}\right)$ is the sample variance of the elements of $\mathbf{S}_o^{*2}$, and $S^2\left(\mathbf{S}_e^{*2}\right)$ is the sample variance of the elements of $\mathbf{S}_e^{*2}$.

The expected value of $\rho_B$ is approximated by:

$$\mathrm{E}\left(\rho_B\right)\cong\frac{E\left(\sum_k^K S_{ko}^{*2}S_{ke}^{*2}-K\bar{S}_o^{*2}\bar{S}_e^{*2}\right)}{(K-1)\sqrt{E\left[S^2\left(\mathbf{S}_o^{*2}\right)S^2\left(\mathbf{S}_e^{*2}\right)\right]}}. \tag{S.5}$$

The approximation has later been verified by a numerical simulation, see Section D, and proved to be valid for $B\geq N/4$.

The derivation of (S.5) is as follows. From the definition, we have:

$$\mathrm{E}\left(\sum_k^K S_{ko}^{*2} S_{ke}^{*2}\right) = \sum_k^K \mathrm{Cov}\left(S_{ko}^{*2}, S_{ke}^{*2}\right) + \sum_k^K \mathrm{E}\left(S_{ko}^{*2}\right)\mathrm{E}\left(S_{ke}^{*2}\right),$$
$$\mathrm{E}\left(\bar{S}_o^{*2}\bar{S}_e^{*2}\right) = \frac{1}{K^2}\sum_{k,l}^K \mathrm{E}\left(S_{ko}^{*2} S_{le}^{*2}\right) = \frac{1}{K^2}\sum_k^K \mathrm{Cov}\left(S_{ko}^{*2}, S_{ke}^{*2}\right) + \frac{1}{K^2}\sum_k^K \mathrm{E}\left(S_{ko}^{*2}\right)\sum_k^K \mathrm{E}\left(S_{ke}^{*2}\right). \tag{S.6}$$

Combining (S.3) and (S.6), we have:

$$\mathrm{E}\left(\sum_k^K S_{ko}^{*2} S_{ke}^{*2}\right) = \sum_k^K \frac{m_{4k} - m_{2k}^2}{N^3} + \sum_k^K \frac{m_{2k}^2}{N^2},$$
$$\mathrm{E}\left(\bar{S}_o^{*2}\bar{S}_e^{*2}\right) = \frac{1}{K^2}\sum_{k,l}^K \mathrm{E}\left(S_{ko}^{*2} S_{le}^{*2}\right) = \frac{1}{K^2}\sum_k^K \frac{m_{4k} - m_{2k}^2}{N^3} + \frac{1}{K^2}\left(\sum_k^K \frac{m_{2k}}{N}\right)^2, \tag{S.7}$$

which implies:

$$\mathrm{E}\left(\sum_k^K S_{ko}^{*2} S_{ke}^{*2} - K\bar{S}_o^{*2}\bar{S}_e^{*2}\right) = (K-1)\left[\frac{1}{K}\sum_k^K \frac{m_{4k} - m_{2k}^2}{N^3} + \frac{1}{N^2} S^2\left(\boldsymbol{\sigma}^2\right)\right]. \tag{S.8}$$

while $S^2\left(\boldsymbol{\sigma}^2\right)$ is the sample variance of the elements of $\boldsymbol{\sigma}^2$.

We define $\bar{m}_4$ and $\bar{m}_2^2$ as:

$$\bar{m}_4 = \frac{1}{K}\sum_{k=1}^K m_{4k},$$
$$\bar{m}_2^2 = \frac{1}{K}\sum_{k=1}^K m_{2k}^2. \tag{S.9}$$

Then (S.8) can be rewritten as:

$$\mathrm{E}\left(\sum_k^K S_{ko}^{*2} S_{ke}^{*2} - K\bar{S}_o^{*2}\bar{S}_e^{*2}\right) = \frac{(K-1)}{N^2}\left[\frac{\bar{m}_4 - \bar{m}_2^2}{N} + S^2\left(\boldsymbol{\sigma}^2\right)\right]. \tag{S.10}$$

In order to evaluate the term $\mathrm{E}\left[S^2\left(\mathbf{S}_o^{*2}\right)S^2\left(\mathbf{S}_e^{*2}\right)\right]$ we note that $S^2\left(\mathbf{S}_o^{*2}\right)$ and $S^2\left(\mathbf{S}_e^{*2}\right)$ are correlated; however, the co-variance between them is of the order $N^{-5}$ and is negligible (not shown). Therefore,

$$\mathrm{E}\left[S^2\left(\mathbf{S}_o^{*2}\right)S^2\left(\mathbf{S}_e^{*2}\right)\right] \cong \mathrm{E}\left[S^2\left(\mathbf{S}_o^{*2}\right)\right]\mathrm{E}\left[S^2\left(\mathbf{S}_e^{*2}\right)\right]. \tag{S.11}$$

The relation has later been verified by a numerical simulation, see Section D for details.

From the definition, we have:

$$\mathrm{E}\left[S^2\left(\mathbf{S}_e^{*2}\right)\right] = \frac{1}{K-1}\mathrm{E}\left[\sum_k^K S_{ko}^{*4} - K\bar{S}_o^{*4}\right], \tag{S.12}$$

while:

$$\mathrm{E}\left(S_{ko}^{*4}\right) = \mathrm{E}\left(S_{ko}^{*2}\right)^2 + \mathrm{V}\left(S_{ko}^{*2}\right),$$
$$\mathrm{E}\left(\bar{S}_o^{*4}\right) = \frac{1}{K^2}\sum_k^K \mathrm{E}\left(S_{ko}^{*4}\right) + \frac{1}{K^2}\sum_{k\neq l}^K \mathrm{E}\left(S_{ko}^{*2} S_{lo}^{*2}\right) = \frac{1}{K^2}\sum_k^K \mathrm{V}\left(S_{ko}^{*4}\right) + \frac{1}{K^2}\left[\sum_{k\neq l}^K \mathrm{E}\left(S_{ko}^{*2}\right)\right]^2. \tag{S.13}$$

Combining (S.3), (S.9), (S.12) and (S.13), we obtain:

$$\mathrm{E}\left[S^2\left(\mathbf{S}_o^{*2}\right)\right] = \frac{(K-1)}{N^2}\left[\frac{\bar{m}_4 - \bar{m}_2^2}{N} + \frac{4\bar{m}_2^2}{B} + S^2\left(\boldsymbol{\sigma}^2\right)\right], \tag{S.14}$$

Combining (S.5), (S.10), (S.12) and (S.14), we obtain:

$$\mathrm{E}\left[\rho_B\right] \cong \left[1 + \frac{\frac{4}{B}\bar{m}_2^2}{S^2\left(\boldsymbol{\sigma}_T^2\right) + \frac{\bar{m}_4 - \bar{m}_2^2}{N}}\right]^{-1}, \quad \text{(S.15)}$$

which concludes the derivation of $\mathrm{E}\left(\rho_B\right)$. It should be also noted here that $\bar{m}_4 > \bar{m}_2^2$ as $m_{4k} \geq m_{2k}^2$ holds for any distribution, as a special case of Jensen's inequality.

Using the same procedure, we can derive the equation for the expected value of sample correlation coefficient ($\rho_\gamma$) between bootstrap variance and the distribution variance as:

$$\mathrm{E}\left[\rho_\gamma\right] \cong \sqrt{\frac{S^2\left(\boldsymbol{\sigma}_T^2\right)}{S^2\left(\boldsymbol{\sigma}_T^2\right) + \frac{\bar{m}_4 - \bar{m}_2^2}{N} + \frac{2\bar{m}_2^2}{B}}}. \quad \text{(S.16)}$$

In the case of the bootstrap technique applied to the evaluation of variance in three-dimensional reconstruction from projections, the distribution variance $\boldsymbol{\sigma}^2$ is called the structure variance and we approximate $S^2\left(\boldsymbol{\sigma}^2\right)$ by $S^2\left(\boldsymbol{\sigma}_{Struct}^2\right)$. The latter is calculated as a among-voxels variance of the structure variance map.

## The Expected Value, Variance and Covariance of the Bootstrap Sample Variance

### The definition of the bootstrap variance

For the purpose of this section we assume that the sample set $\boldsymbol{\Psi}$ comprises $N$ independent random variables $\left(\Psi_1, \Psi_2, \cdots\cdots \Psi_N\right)$ with zero expectation originated from the same distribution. It should be noted here that we omitted the subscript $k$ which stands for the pixel number (c.f. (S.1) and (S.17)). The results derived in this section hold for any single pixel. We define $m_2$ and $m_4$ as the second and the fourth moments of the distribution, i.e.:

$$\begin{aligned} &\mathrm{E}\left(\Psi_i\right) = 0, \\ &\mathrm{E}\left(\Psi_i^2\right) = m_2, \\ &\mathrm{E}\left(\Psi_i^4\right) = m_4, \\ &\mathrm{E}\left(\Psi_i \Psi_j\right) = 0. \end{aligned} \quad \text{(S.17)}$$

It should also be noted that we do not have assumptions on the 3$^{rd}$ moment because in our derivation it is always multiplied by the 1$^{st}$ moment that is equal to zero.

During the bootstrap calculation, the sample set $\boldsymbol{\Psi}$ is resampled with replacement and yields a new set called a bootstrap sample. The average of this bootstrap sample is called bootstrap sample average $\left(\Psi^*\right)$. We introduce auxiliary random variables $\left(e_1, e_2, ..., e_N\right)$, such that $\Psi^*$ can be expressed as:

$$\Psi^* = \frac{1}{N}\sum_i^N e_i \Psi_i, \quad \text{(S.18)}$$

The bootstrap step is repeated $B$ times and yields $B$ bootstrap sample averages $\left(\Psi_1^*, \Psi_2^*, \cdots\cdots, \Psi_B^*\right)$, and the bootstrap variance $S^{*2}$ is calculated among these bootstrap sample averages as:

$$S^{*2} = \frac{1}{B-1}\left[\sum_i^B \Psi_i^{*2} - \frac{1}{B}\left(\sum_i^B \Psi_i^*\right)^2\right] = \frac{1}{B}\sum_i^B \Psi_i^{*2} - \frac{1}{B(B-1)}\sum_{i \neq j}^B \Psi_i^* \Psi_j^*. \quad \text{(S.19)}$$

**The expected value of the bootstrap variance**

The expected value of the bootstrap variance $S^{*2}$ (S.19) is evaluated as:

$$\mathrm{E}\left(S^{*2}\right)=\frac{1}{B}\sum_{i}^{B}\mathrm{E}\left(\Psi_i^{*2}\right)-\frac{1}{B(B-1)}\sum_{i\neq j}^{B}\mathrm{E}\left(\Psi_i^{*}\Psi_j^{*}\right). \tag{S.20}$$

Considering that all the bootstrap sample averages are identically distributed, i.e., $\mathrm{E}\left(\Psi_i^{*2}\right)=\mathrm{E}\left(\Psi_j^{*2}\right)$ and $\mathrm{E}\left(\Psi_i^{*}\Psi_j^{*}\right)=\mathrm{E}\left(\Psi_k^{*}\Psi_l^{*}\right)$, (S.20) can be rewritten as:

$$\mathrm{E}\left(S^{*2}\right)=\mathrm{E}\left(\Psi_i^{*2}\right)-\mathrm{E}\left(\Psi_i^{*}\Psi_j^{*}\right). \tag{S.21}$$

is evaluated as:

$$\begin{aligned}\mathrm{E}\left(\Psi_i^{*2}\right)&=\frac{1}{N^2}\mathrm{E}\left(\sum_{j}^{N}e_{ij}^2\Psi_j^2+\sum_{j\neq k}^{N}e_{ij}e_{ik}\Psi_j\Psi_k\right)\\&=\frac{1}{N^2}\sum_{j}^{N}\mathrm{E}\left(e_{ij}^2\right)\mathrm{E}\left(\Psi_j^2\right)+\frac{1}{N^2}\sum_{j\neq k}^{N}\mathrm{E}\left(e_{ij}e_{ik}\right)\mathrm{E}\left(\Psi_j\right)\mathrm{E}\left(\Psi_k\right).\end{aligned} \tag{S.22}$$

$E\left(\Psi_i^{*}\Psi_j^{*}\right)$ is evaluated as:

$$\begin{aligned}\mathrm{E}\left(\Psi_i^{*}\Psi_j^{*}\right)&=\mathrm{E}\left(\sum_{k}^{N}e_{ik}e_{jk}\Psi_k^2+\sum_{k\neq l}^{N}e_{ik}e_{jl}\Psi_k\Psi_l\right)\\&=\frac{1}{N^2}\sum_{j}^{N}\mathrm{E}\left(e_{ik}\right)\mathrm{E}\left(e_{jk}\right)\mathrm{E}\left(\Psi_k^2\right)+\frac{1}{N^2}\sum_{j}^{N}\mathrm{E}\left(e_{ik}\right)\mathrm{E}\left(e_{jl}\right)\mathrm{E}\left(\Psi_k\right)\mathrm{E}\left(\Psi_l\right).\end{aligned} \tag{S.23}$$

We prove in section C that $E\left(e_{ij}^2\right)$, $E\left(e_{ik}\right)$ and $E\left(e_{jl}\right)$ are equal to:

$$\begin{aligned}\mathrm{E}\left(e_{ij}^2\right)&=\frac{2N-1}{N},\\\mathrm{E}\left(e_{ij}e_{ik}\right)&=\frac{N-1}{N},\\\mathrm{E}\left(e_{ij}\right)&=\mathrm{E}\left(e_{kl}\right)=1.\end{aligned} \tag{S.24}$$

Combining (S.17), (S.22), and (S.24), we arrive at:

$$\mathrm{E}\left(\Psi_i^{*2}\right)=\frac{2N-1}{N^2}m_2. \tag{S.25}$$

Combining (S.17), (S.23) and (S.24), we arrive at:

$$\mathrm{E}\left(\Psi_i^{*}\Psi_j^{*}\right)=\frac{1}{N}m_2. \tag{S.26}$$

Combining (S.21), (S.25) and (S.26), we arrive at the final equation of the expected value of bootstrap variance:

$$\mathrm{E}\left(S^{*2}\right)=\frac{N-1}{N^2}m_2, \tag{S.27}$$

which concludes the derivation of the expected value of the bootstrap variance.

**The variance of the bootstrap variance**

Now we begin derivation of the variance of the bootstrap variance. From the definition, it holds:

$$\mathrm{V}\left(S^{*2}\right)=\mathrm{E}\left(S^{*4}\right)-\mathrm{E}\left(S^{*2}\right)^2. \tag{S.28}$$

In equation (S.28), $\mathrm{E}\left(S^{*2}\right)$ is given by (S.27) and thus we only need to derive $\mathrm{E}\left(S^{*4}\right)$. The latter can be split into three parts:

$$\mathrm{E}\left(S^{*4}\right)=\frac{1}{(B-1)^2}\mathrm{E}\left[\left(\sum_i^B \Psi_i^{*2}\right)^2-2B\left(\bar{\Psi}^*\right)^2\sum_i^B \Psi_i^{*2}+B^2\left(\bar{\Psi}^*\right)^4\right], \tag{S.29}$$

The first part in (S.29) is evaluated as:

$$\begin{aligned}\mathrm{E}\left[\left(\sum_i^B \Psi_i^{*2}\right)^2\right]&=\sum_i^B \mathrm{E}\left(\Psi_i^{*4}\right)+\sum_{i\neq j}^B \mathrm{E}\left(\Psi_i^{*2}\Psi_j^{*2}\right)\\&=BE\left(\Psi_i^{*4}\right)+B(B-1)\mathrm{E}\left(\Psi_i^{*2}\Psi_j^{*2}\right).\end{aligned} \tag{S.30}$$

The second part in (S.29) is equal to:

$$\begin{aligned}&2B\,\mathrm{E}\left[\left(\bar{\Psi}^*\right)^2\sum_i^B \Psi_i^{*2}\right]\\&=\frac{2}{B}\mathrm{E}\left(\sum_i^B \Psi_i^{*4}+\sum_{i\neq j}^B \Psi_i^{*2}\Psi_j^{*2}+2\sum_{i\neq j}^B \Psi_i^{*3}\Psi_j^{*}+\sum_{i\neq j\neq k}^B \Psi_i^{*2}\Psi_j^{*}\Psi_j^{*}\right)\\&=2\mathrm{E}\left(\Psi_i^{*4}\right)+2(B-1)\mathrm{E}\left(\Psi_i^{*2}\Psi_j^{*2}\right)+4(B-1)\mathrm{E}\left(\Psi_i^{*3}\Psi_j^{*}\right)\\&+(B-1)(B-2)\mathrm{E}\left(\Psi_i^{*2}\Psi_j^{*}\Psi_k^{*}\right).\end{aligned} \tag{S.31}$$

The third part in (S.29) is equal to:

$$\begin{aligned}&B^2\,\mathrm{E}\left[\left(\bar{\Psi}^*\right)^4\right]\\&=\frac{1}{B^2}\mathrm{E}\left(\sum_i^B \Psi_i^{*4}+4\sum_{i\neq j}^B \Psi_i^{*3}\Psi_j^{*}+3\sum_{i\neq j}^B \Psi_i^{*2}\Psi_j^{*2}+6\sum_{i\neq j\neq k}^B \Psi_i^{*2}\Psi_j^{*}\Psi_j^{*}+\sum_{i\neq j\neq k\neq l}^B \Psi_i^{*}\Psi_j^{*}\Psi_k^{*}\Psi_l^{*}\right)\\&=\frac{1}{B}\mathrm{E}\left(\Psi_i^{*4}\right)+\frac{3(B-1)}{B}\mathrm{E}\left(\Psi_i^{*2}\Psi_j^{*2}\right)+\frac{4(B-1)}{B}\mathrm{E}\left(\Psi_i^{*3}\Psi_j^{*}\right)\\&+\frac{6(B-1)(B-2)}{B}\mathrm{E}\left(\Psi_i^{*2}\Psi_j^{*}\Psi_k^{*}\right)+\frac{(B-1)(B-2)(B-3)}{B}\mathrm{E}\left(\Psi_i^{*}\Psi_j^{*}\Psi_k^{*}\Psi_l^{*}\right).\end{aligned} \tag{S.32}$$

We define intermediate variables $L_1$ through $L_5$ as:

$$\begin{aligned}L_1&=\mathrm{E}\left(\Psi_i^{*4}\right),\\L_2&=\mathrm{E}\left(\Psi_i^{*2}\Psi_j^{*2}\right),\\L_3&=\mathrm{E}\left(\Psi_i^{*3}\Psi_j^{*}\right),\\L_4&=\mathrm{E}\left(\Psi_i^{*2}\Psi_j^{*}\Psi_k^{*}\right),\\L_5&=\mathrm{E}\left(\Psi_i^{*}\Psi_j^{*}\Psi_k^{*}\Psi_l^{*}\right).\end{aligned} \tag{S.33}$$

Combining (S.29), (S.30), (S.31), (S.32) and (S.33), we obtain:

$$\mathrm{E}\left(S^{*4}\right)=L_2-2L_4+L_5+\frac{L_1-L_2-4L_3+8L_4-4L_5}{B}+\frac{2L_2-4L_4+2L_5}{B(B-1)}. \tag{S.34}$$

Combining (S.27), (S.28) and (S.34), we arrive at:

$$\begin{aligned}&\mathrm{V}\left(S^{*2}\right)=\\&L_2-2L_4+L_5-\frac{(N-1)^2}{N^4}m_2^2+\frac{L_1-L_2-4L_3+8L_4-4L_5}{B}+\frac{2L_2-4L_4+2L_5}{B(B-1)}.\end{aligned} \tag{S.35}$$

The $L_1, L_2, \cdots, L_5$ are derived as:

$$\begin{aligned} L_1 &= \frac{E\left(e_{ij}^4\right)}{N^3} m_4 + E\left(e_{ij}^2 e_{ik}^2\right) \frac{3(N-1)}{N^3} m_2^2, \\ L_2 &= \frac{(2N-1)^2}{N^5} m_4 + \frac{(N-1)\left(6N^2-8N+3\right)}{N^5} m_2^2, \\ L_3 &= \frac{E\left(e_{ij}^3\right)}{N^3} m_4 + \frac{3(N-1)E\left(e_{ij}^2 e_{ik}\right)}{N^3} m_2^2, \\ L_4 &= \frac{(2N-1)}{N^4} m_4 + \frac{(N-1)(4N-3)}{N^4} m_2^2, \\ L_5 &= \frac{1}{N^3} m_4 + \frac{3(N-1)}{N^3} m_2^2. \end{aligned} \tag{S.36}$$

We proved in Section C that $\mathrm{E}\left(e_{ij}^4\right)$, $\mathrm{E}\left(e_{ij}^2 e_{ik}^2\right)$, $\mathrm{E}\left(e_{ij}^3\right)$, and $\mathrm{E}\left(e_{ij}^2 e_{ik}\right)$ are equal to:

$$\begin{aligned} \mathrm{E}\left(e_{ij}^4\right) &= \frac{(N-1)(N-2)(N-3)}{N^3} + \frac{6(N-1)(N-2)}{N^2} + \frac{7(N-1)}{N} + 1, \\ \mathrm{E}\left(e_{ij}^2 e_{ik}^2\right) &= \frac{(N-1)(N-2)(N-3)}{N^3} + \frac{2(N-1)(N-2)}{N^2} + \frac{(N-1)}{N}, \\ \mathrm{E}\left(e_{ij}^3\right) &= \frac{(N-1)(N-2)}{N^2} + \frac{3(N-1)}{N} + 1, \\ \mathrm{E}\left(e_{ij}^2 e_{ik}\right) &= \frac{(N-1)(N-2)}{N^2} + \frac{(N-1)}{N}. \end{aligned} \tag{S.37}$$

Combining (S.35), (S.36), and (S.37), we arrive at the final equation for the variance of bootstrap variance:

$$\begin{aligned} \mathrm{V}\left(S^{*2}\right) &= \left(\frac{m_4 - m_2^2}{N^3} - \frac{2m_4 - 4m_2^2}{N^4} + \frac{m_4 - 3m_2^2}{N^5}\right)\left[1 + \frac{2}{B(B-1)}\right] \\ &+ \frac{1}{B}\left(\frac{2m_2^2}{N^2} + \frac{5m_4 - 9m_2^2}{N^3} - \frac{11m_4 - 25m_2^2}{N^4} + \frac{14m_4 - 18m_2^2}{N^5}\right). \end{aligned} \tag{S.38}$$

The higher order terms in (S.38) can be neglected in practice, and therefore we have:

$$\mathrm{V}\left(S^{*2}\right) = \frac{m_4 - 2m_2^2}{N^3} + \frac{2m_2^2}{N^2 B}, \tag{S.39}$$

which concludes the derivation of the variance of the bootstrap variance.

**Covariance of bootstrap variance**

The co-variance of bootstrap variances is defined as:

$$\mathrm{Cov}\left(S_i^{*2}, S_j^{*2}\right) = \mathrm{E}\left(S_i^{*2} S_j^{*2}\right) - \mathrm{E}\left(S_i^{*2}\right)\mathrm{E}\left(S_j^{*2}\right), \tag{S.40}$$

while $S_i^{*2}$ and $S_j^{*2}$ are bootstrap variances obtained from two independent bootstrap calculations using the same sample set.

We know that:

$$\mathrm{E}\left(S_i^{*2}\right) = \mathrm{E}\left(S_j^{*2}\right) = \frac{N-1}{N^2} m_2. \tag{S.41}$$

Thus, the only unknown part in (S.41) is $E\left(S_i^{*2} S_j^{*2}\right)$, which is evaluated as:

$$\mathrm{E}\left(S_i^{*2}S_j^{*2}\right)=\frac{1}{(B-1)^2}\mathrm{E}\left[\left(\sum_i^B\Psi_i^{*2}-B\bar{\Psi}_i^{*2}\right)\left(\sum_i^B\Psi_j^{*2}-B\bar{\Psi}_j^{*2}\right)\right]$$
$$=\frac{1}{(B-1)^2}\mathrm{E}\left[\sum_i^B\Psi_i^{*2}\sum_i^B\Psi_j^{*2}-2B\bar{\Psi}_i^{*2}\sum_i^B\Psi_j^{*2}+B^2\bar{\Psi}_i^{*2}\bar{\Psi}_j^{*2}\right]. \tag{S.42}$$

For the first part of $E\left(S_i^{*2}S_j^{*2}\right)$, we have:

$$\mathrm{E}\left(\sum_i^B\Psi_i^{*2}\sum_i^B\Psi_j^{*2}\right)=B^2\,\mathrm{E}\left(\Psi_i^{*2}\Psi_j^{*2}\right)=B^2L_2. \tag{S.43}$$

For the second part of $E\left(S_i^{*2}S_j^{*2}\right)$, we have:

$$2B\,\mathrm{E}\left(\bar{\Psi}_i^{*2}\sum_i^B\Psi_j^{*2}\right)=2B\,\mathrm{E}\left(\Psi_i^{*2}\Psi_j^{*2}\right)+2B(B-1)\mathrm{E}\left(\Psi_i^{*}\Psi_j^{*}\Psi_k^{*2}\right)=2BL_2+2B(B-1)L_4. \tag{S.44}$$

For the third part in $E\left(S_i^{*2}S_j^{*2}\right)$, we have:

$$B^2\,\mathrm{E}\left(\bar{\Psi}_i^{*2}\bar{\Psi}_j^{*2}\right)=L_2+2(B-1)L_4+(B-1)^2L_5. \tag{S.45}$$

Combining (S.42), (S.43), (S.44) and (S.45), we arrive at:

$$\mathrm{E}\left(S_i^{*2}S_j^{*2}\right)=L_2-2L_4+L_5. \tag{S.46}$$

Finally, combining (S.36), (S.40) and (S.46), we arrive at the equation for the co-variance of the bootstrap variance:

$$\mathrm{Cov}\left(S_i^{*2},S_j^{*2}\right)=\frac{m_4-m_2^2}{N^3}-\frac{2m_4-4m_2^2}{N^4}+\frac{m_4-3m_2^2}{N^5}. \tag{S.47}$$

The higher order terms in (S.47) are negligible, and thus we have:

$$\mathrm{Cov}\left(S_i^{*2},S_j^{*2}\right)=\frac{m_4-m_2^2}{N^3}, \tag{S.48}$$

which concludes the derivation of the co-variance of the bootstrap variance.

Finally we note that the final results of this section (S.27), (S.39) and (S.48) have been verified by a numerical simulation, see Section D for details.

**Additional Relations for Auxiliary Random Variables** $(e_1,e_2,...,e_N)$

The set of auxiliary random variables $(e_1,e_2,...,e_N)$ are introduced such that the bootstrap sample average $\Psi^*$ can be expressed as:

$$\Psi^*=\frac{1}{N}\sum_i^N e_i\Psi_i. \tag{S.49}$$

The random variables $(e_1,e_2,...,e_{N-1})$ are distributed multinomially with parameters $p_1=p_2=...=p_{N-1}=\frac{1}{N}$, i.e., $(e_1,e_2,...,e_{N-1})\sim M\left(\frac{1}{N},\cdots,\frac{1}{N};N\right)$, so that:

$$\mathrm{P}\left(e_1=n_1,\cdots,e_{N-1}=n_{N-1}\right)=\frac{N!}{n_1!\cdots n_{N-1}!}\left(\frac{1}{N}\right)^{n_1+\cdots+n_{N-1}} \tag{S.50}$$

while $n_1\ge 0,\cdots,n_{N-1}\ge 0$ and $n_1+\cdots+n_{N-1}\le N$, whereas $e_N=N-(e_1+\cdots+e_{N-1})$.

Based on the known properties of the multinomial distribution in equation (S.50), we can determine the probability of $e_i = k,\ \left(0 \le k \le N\right)$ to be:

$$\mathrm{P}\left(e_i = k\right) = \frac{N!}{k!\left(N-k\right)!}\frac{\left(N-1\right)^{N-k}}{N^N}. \quad \text{(S.51)}$$

The expected value of $e_i$ is equal to:

$$\begin{aligned} &\mathrm{E}\left(e_i\right) = p_i N = 1, (i = 1, \cdots, N-1) \\ &\mathrm{E}\left(e_N\right) = N - \left[\mathrm{E}\left(e_i\right) + \cdots + \mathrm{E}\left(e_{N-1}\right)\right] = 1 \end{aligned} \quad \text{(S.52)}$$

$\mathrm{E}\left(e_i^2\right)$, $\mathrm{E}\left(e_i^3\right)$, and $\mathrm{E}\left(e_i^4\right)$ are evaluated using the similar method as follows:

$$\begin{aligned} &\mathrm{E}\left(e_i^2\right) = \frac{2N-1}{N}, \\ &\mathrm{E}\left(e_i^3\right) = \frac{\left(N-1\right)\left(N-2\right)}{N^2} + \frac{3\left(N-1\right)}{N} + 1, \\ &\mathrm{E}\left(e_i^4\right) = \frac{\left(N-1\right)\left(N-2\right)\left(N-3\right)}{N^3} + \frac{6\left(N-1\right)\left(N-2\right)}{N^2} + \frac{7\left(N-1\right)}{N} + 1. \end{aligned} \quad \text{(S.53)}$$

It should be noted here that $e_i$ and $e_j$ are not independent, and the combined probability of $e_i = k$ and $e_j = l$ while $\left(0 \le k, l \le N\right)$ is equal to:

$$\mathrm{P}\left(e_i = k, e_j = l\right) = \frac{N!}{k!l!\left(N-k-l\right)!}\frac{\left(N-2\right)^{N-k}}{N^N}. \quad \text{(S.54)}$$

In addition,

$$\begin{aligned} &\mathrm{E}\left(e_i e_j\right) = \frac{\left(N-1\right)}{N}, \\ &\mathrm{E}\left(e_i^2 e_j\right) = \frac{\left(N-1\right)\left(N-2\right)}{N^2} + \frac{\left(N-1\right)}{N}, \\ &\mathrm{E}\left(e_i^2 e_j^2\right) = \frac{\left(N-1\right)\left(N-2\right)\left(N-3\right)}{N^3} + \frac{2\left(N-1\right)\left(N-2\right)}{N^2} + \frac{\left(N-1\right)}{N}. \end{aligned} \quad \text{(S.55)}$$

The relations (S.49)-(S.55) were used in the Section B.

## Numerical Simulations to Verify the Results

The following numerical simulations have been performed to verify some of the obtained equations.

## Simulations to verify (S.5): The approximation of sample correlation coefficient of bootstrap variances

The procedure to calculate the sample correlation coefficient of bootstrap variance ($\rho_B$) has been described in Section A. Here we set $K$=20, tried different combination of $N$ and $B$. The procedure has been repeated 10,000 times. By $\langle \cdot \rangle$ we denote the average value over the 10,000 trails.

| N | B | $\langle \rho_B \rangle$ | $\left\langle \sum_k^K S_{ko}^{*2} S_{ke}^{*2} - K \bar{S}_o^{*2} \bar{S}_e^{*2} \right\rangle$ | $\left\langle S^2\left(\mathbf{S}_o^{*2}\right)\right\rangle \left\langle S^2\left(\mathbf{S}_e^{*2}\right)\right\rangle$ | $\dfrac{\left\langle \sum_k^K S_{ko}^{*2} S_{ke}^{*2} - K \bar{S}_o^{*2} \bar{S}_e^{*2} \right\rangle}{(K-1)\sqrt{\left\langle S^2\left(\mathbf{S}_o^{*2}\right)\right\rangle \left\langle S^2\left(\mathbf{S}_e^{*2}\right)\right\rangle}}$ |
|---|---|---|---|---|---|
| 100 | 20 | 0.7044 | 31.4328 | 6.1700 | 0.6660 |

| | | | | | |
|---|---|---|---|---|---|
| 150 | 20 | 0.7087 | 13.8413 | 1.1830 | 0.6697 |
| 200 | 20 | 0.7088 | 7.8984 | 0.3795 | 0.6747 |
| 100 | 50 | 0.8646 | 31.7627 | 3.8388 | 0.8468 |
| 150 | 50 | 0.8649 | 13.9927 | 0.7554 | 0.8573 |
| 200 | 50 | 0.8618 | 7.8786 | 0.2414 | 0.8439 |
| 100 | 100 | 0.9291 | 32.0549 | 3.3619 | 0.9201 |
| 150 | 100 | 0.9286 | 14.0818 | 0.6501 | 0.9192 |
| 200 | 100 | 0.9270 | 7.8619 | 0.2031 | 0.9182 |
| 100 | 10000 | 0.9992 | 32.2604 | 2.8879 | 0.9991 |

The simulations indicate that a bootstrap sample of *B*=50 is sufficient for the approximate $\rho_B$ to be satisfactory.

**Simulations to verify the correlation between $\mathbf{S}_o^{*2}$ and $\mathbf{S}_e^{*2}$**

Here we use a sample set with K=20, B=10,000 and N=100, 150, 200. The values reported are average values over the 1000 trails.

| N | $\left\langle S^2\left(\mathbf{S}_o^{*2}\right)S^2\left(\mathbf{S}_e^{*2}\right)\right\rangle$ | $\left\langle S^2\left(\mathbf{S}_o^{*2}\right)\right\rangle\left\langle S^2\left(\mathbf{S}_e^{*2}\right)\right\rangle$ | $\operatorname{cov}\left(\mathbf{S}_o^{*2},\mathbf{S}_e^{*2}\right)$ | $N^5\operatorname{cov}\left(\mathbf{S}_o^{*2},\mathbf{S}_e^{*2}\right)/5.44e8$ |
|---|---|---|---|---|
| 100 | 2.887954 | 2.833545 | 0.054408 | 1.0000 |
| 150 | 0.551595 | 0.544690 | 0.006804 | 0.9496 |
| 200 | 0.172754 | 0.171038 | 0.001716 | 1.0091 |

**Simulations to verify the expected value and variance of bootstrap sample variance of a single pixel**

The procedure to calculate the bootstrap variance of a single pixel has been explained in Section B. We use the setting B=100 and N=100, 150, 200. The procedure we repeated 10,000 times.

| N | $\left\langle S^{*2}\right\rangle$ | $\left\langle S^2\left(S^{*2}\right)\right\rangle$ | $E\left(S^{*2}\right)$ | $V\left(S^{*2}\right)$ |
|---|---|---|---|---|
| 100 | 0.99e-2 | 3.93e-4 | 1.01e-2 | 3.93e-4 |
| 150 | 6.61e-3 | 1.47e-6 | 6.71e-3 | 1.46e-6 |
| 200 | 4.97e-3 | 7.58e-7 | 5.02e-3 | 7.44e-7 |

There is a very good agreement between the simulated and analytically calculated value of the expectation and the variance of the bootstrap variance.