1　分散の推定

n 個の実数値からなる標本 $s=\left\{x_1, x_2, \ldots, x_n\right\}$ に対して，その分散を求めます。また，これらが N 個の実数値からなる母集団 $S=\left\{X_1, X_2, \ldots, X_N\right\}$ から抽出されたものであるとき，母集団の分散を推定します。

1.1　定義

標本の分散

標準分散　$v_p=\dfrac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2$

不偏分散　$v_s=\dfrac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2 \quad \left(=\dfrac{n}{n-1}v_p\right)$

低減分散　$v_h=\dfrac{1}{n+1}\sum_{i=1}^{n}(x_i-\overline{x})^2 \quad \left(=\dfrac{n}{n+1}v_p\right)$

母集団の分散

標準母分散　$V_p=\dfrac{1}{N}\sum_{i=1}^{N}\left(X_i-\overline{X}\right)^2$

不偏母分散　$V_s=\dfrac{1}{N-1}\sum_{i=1}^{N}\left(X_i-\overline{X}\right)^2 \quad \left(=\dfrac{N}{N-1}V_p\right)$

低減母分散　$V_h=\dfrac{1}{N+1}\sum_{i=1}^{N}\left(X_i-\overline{X}\right)^2 \quad \left(=\dfrac{N}{N+1}v_p\right)$

部品の定義

標本について，　$a=\sum_{i=1}^{n}x_i \quad b=\sum_{i=1}^{n}x_i^2$

母集団について，　$A=\sum_{i=1}^{N}X_i \quad B=\sum_{i=1}^{N}X_i^2$

1.2　v_p, V_p を a, b, A, B で表す

v_p の方だけ示します。V_p も全く同じです。

$$
\begin{aligned}
v_p &= \frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\left\{{x_i}^2-2x_i\overline{x}+(\overline{x})^2\right\} \\
&= \frac{1}{n}\sum_{i=1}^{n}{x_i}^2-2\cdot\frac{1}{n}\sum_{i=1}^{n}x_i\overline{x}+\frac{1}{n}\sum_{i=1}^{n}(\overline{x})^2 \\
&= \overline{x^2}-2(\overline{x})^2+(\overline{x})^2 \\
&= \overline{x^2}-(\overline{x})^2 \qquad (\leftarrow \text{これは有名}) \\
&= \frac{b}{n}-\left(\frac{a}{n}\right)^2 \\
&= \frac{b}{n}-\frac{a^2}{n^2}
\end{aligned}
$$

結論

標本について，$v_p=\dfrac{b}{n}-\dfrac{a^2}{n^2}$　母集団について，$V_p=\dfrac{B}{N}-\dfrac{A^2}{N^2}$

1.3　標本分散の期待値（抽出を組合せで考える場合）

母集団 S から相異なる要素を選んで標本 s とする場合，標本の分散の期待値を求めます。$|S|=N$，$|s|=n$ より，標本の取り出し方は全部で ${}_N\mathrm{C}_n$ 通りあります。

$$
\begin{aligned}
E(v_p) &= \frac{1}{{}_N\mathrm{C}_n}\sum v_p \qquad (\leftarrow \textstyle\sum \text{は全標本に対する合計を表します}) \\
&= \frac{1}{{}_N\mathrm{C}_n}\sum\left(\frac{b}{n}-\frac{a^2}{n^2}\right) \\
&= \frac{1}{{}_N\mathrm{C}_n}\left(\frac{\sum b}{n}-\frac{\sum a^2}{n^2}\right) \\
&= \frac{1}{{}_N\mathrm{C}_n}\cdot\frac{1}{n^2}\left(n\sum b-\sum a^2\right)
\end{aligned}
$$

N 個中 n 個を選ぶとき，特定の 1 個が含まれる組合せが ${}_{N-1}\mathrm{C}_{n-1}$ 通り，特定の 2 個が含まれる組合せが ${}_{N-2}\mathrm{C}_{n-2}$ 通りであることを考えると，$\sum b = {}_{N-1}\mathrm{C}_{n-1}B$，$\sum a^2 = {}_{N-2}\mathrm{C}_{n-2}(A^2-B)+{}_{N-1}\mathrm{C}_{n-1}B$ となります。詳しいことは説明しにくいので省きます。これを代入して計算を進めると，

$$
\begin{aligned}
E(v_p) &= \frac{1}{{}_N\mathrm{C}_n}\cdot\frac{1}{n^2}\left[n\left\{{}_{N-1}\mathrm{C}_{n-1}B\right\}-\left\{{}_{N-2}\mathrm{C}_{n-2}(A^2-B)+{}_{N-1}\mathrm{C}_{n-1}B\right\}\right] \\
&= \frac{1}{{}_N\mathrm{C}_n}\cdot\frac{1}{n^2}\left\{n\,{}_{N-1}\mathrm{C}_{n-1}B-{}_{N-2}\mathrm{C}_{n-2}(A^2-B)-{}_{N-1}\mathrm{C}_{n-1}B\right\} \\
&= \frac{1}{{}_N\mathrm{C}_n}\cdot\frac{1}{n^2}\left\{(n-1)\,{}_{N-1}\mathrm{C}_{n-1}B-{}_{N-2}\mathrm{C}_{n-2}(A^2-B)\right\} \\
&= \frac{1}{{}_N\mathrm{C}_n}\cdot\frac{1}{n^2}\left\{(n-1)\frac{n}{N}\,{}_N\mathrm{C}_nB-\frac{n(n-1)}{N(N-1)}\,{}_N\mathrm{C}_n(A^2-B)\right\} \\
&= \frac{1}{n^2}\left\{\frac{n(n-1)}{N}B-\frac{n(n-1)}{N(N-1)}(A^2-B)\right\} \\
&= \frac{n-1}{n}\left\{\frac{1}{N}B-\frac{1}{N(N-1)}(A^2-B)\right\} \\
&= \frac{n-1}{n}\left\{\frac{B}{N-1}-\frac{A^2}{N(N-1)}\right\} \\
&= \frac{n-1}{n}\cdot\frac{N}{N-1}\left(\frac{B}{N}-\frac{A^2}{N^2}\right) \\
&= \frac{n-1}{n}\cdot\frac{N}{N-1}V_p
\end{aligned}
$$

したがって，

$$
\begin{aligned}
\frac{n}{n-1}E(v_p) &= \frac{N}{N-1}V_p \\
E\left(\frac{n}{n-1}v_p\right) &= \frac{N}{N-1}V_p \\
E(v_s) &= V_s
\end{aligned}
$$

標本の不偏分散の期待値は，母集団の不偏母分散であることが分かりました。

結論

抽出を組合せで考える場合，$E(v_s)=V_s$

1.4　標本分散の期待値（抽出を重複順列で考える場合）

今度は母集団 S から標本 s を抽出するとき，要素を重複して選ぶことを許すことにします。重複順列で考えると，標本の取り出し方は全部で N^n 通りあります。

$$
\begin{aligned}
E(v_p) &= \frac{1}{N^n}\sum v_p \qquad (\leftarrow\textstyle\sum\text{は全標本に対する合計を表します}) \\
&= \frac{1}{N^n}\sum\left(\frac{b}{n}-\frac{a^2}{n^2}\right) \\
&= \frac{1}{N^n}\left(\frac{\sum b}{n}-\frac{\sum a^2}{n^2}\right) \\
&= \frac{1}{N^n}\cdot\frac{1}{n^2}(n\textstyle\sum b-\sum a^2)
\end{aligned}
$$

ここで，$\sum b = nN^{n-1}B$，$\sum a^2 = n(n-1)N^{n-2}A^2+nN^{n-1}B$ が成り立ちます。これも詳しい説明は省きます。面積図などを利用して確かめて下さい。これを代入して計算を進めると，

$$
\begin{aligned}
E(v_p) &= \frac{1}{N^n}\cdot\frac{1}{n^2}\Big[\,n(nN^{n-1}B)-\big\{n(n-1)N^{n-2}A^2+nN^{n-1}B\big\}\,\Big] \\
&= \frac{1}{N^n}\cdot\frac{1}{n^2}\big\{n(n-1)N^{n-1}B-n(n-1)N^{n-2}A^2\big\} \\
&= \frac{1}{N^n}\cdot\frac{n-1}{n}(N^{n-1}B-N^{n-2}A^2) \\
&= \frac{n-1}{n}\left(\frac{B}{N}-\frac{A^2}{N^2}\right) \\
&= \frac{n-1}{n}V_p
\end{aligned}
$$

したがって，

$$
\begin{aligned}
\frac{n}{n-1}E(v_p) &= V_p \\
E\left(\frac{n}{n-1}v_p\right) &= V_p \\
E(v_s) &= V_p
\end{aligned}
$$

標本の不偏分散の期待値は，母集団の標準母分散であることが分かりました。

結論

抽出を重複順列で考える場合，$E(v_s) = V_p$

1.5　標本分散の期待値（抽出を重複組合せで考える場合）

今度は標本の取り出し方を重複組合せで考えます。標本の取り出し方は全部で ${}_N\mathrm{H}_n$ 通りあります。

$$
\begin{aligned}
E(v_p) &= \frac{1}{{}_N\mathrm{H}_n}\Sigma v_p \qquad (\leftarrow \Sigma \text{は全標本に対する合計を表します}) \\
&= \frac{1}{{}_N\mathrm{H}_n}\Sigma\left(\frac{b}{n} - \frac{a^2}{n^2}\right) \\
&= \frac{1}{{}_N\mathrm{H}_n}\left(\frac{\Sigma b}{n} - \frac{\Sigma a^2}{n^2}\right) \\
&= \frac{1}{{}_N\mathrm{H}_n}\cdot\frac{1}{n^2}(n\Sigma b - \Sigma a^2)
\end{aligned}
$$

今回も説明は省きますが，$\Sigma b = {}_{N+1}\mathrm{H}_{n-1}B$，$\Sigma a^2 = {}_{N+2}\mathrm{H}_{n-2}A^2 + {}_{N+2}\mathrm{H}_{n-1}B$ が成り立ちます。これを代入して計算を進めると，

$$
\begin{aligned}
E(v_p) &= \frac{1}{{}_N\mathrm{H}_n}\cdot\frac{1}{n^2}\left\{n({}_{N+1}\mathrm{H}_{n-1}B) - ({}_{N+2}\mathrm{H}_{n-2}A^2 + {}_{N+2}\mathrm{H}_{n-1}B)\right\} \\
&= \frac{1}{{}_N\mathrm{H}_n}\cdot\frac{1}{n^2}\left\{n\frac{n}{N}{}_N\mathrm{H}_N B - \frac{n(n-1)}{N(N+1)}{}_N\mathrm{H}_n A^2 - \frac{n(N+n)}{N(N+1)}{}_N\mathrm{H}_n B\right\} \\
&= \frac{1}{n^2}\left\{\frac{n^2}{N}B - \frac{n(n-1)}{N(N+1)}A^2 - \frac{n(N+n)}{N(N+1)}B\right\} \\
&= \frac{1}{n}\left\{\frac{n}{N}B - \frac{n-1}{N(N+1)}A^2 - \frac{N+n}{N(N+1)}B\right\} \\
&= \frac{1}{n}\left\{\frac{n-1}{N+1}B - \frac{n-1}{N(N+1)}A^2\right\} \\
&= \frac{n-1}{n}\left\{\frac{1}{N+1}B - \frac{1}{N(N+1)}A^2\right\} \\
&= \frac{n-1}{n}\cdot\frac{N}{N+1}\left(\frac{B}{N} - \frac{A^2}{N^2}\right) \\
&= \frac{n-1}{n}\cdot\frac{N}{N+1}V_p
\end{aligned}
$$

したがって，

$$
\begin{aligned}
\frac{n}{n-1}E(v_p) &= \frac{N}{N+1}V_p \\
E\left(\frac{n}{n-1}v_p\right) &= \frac{N}{N+1}V_p \\
E(v_s) &= V_h
\end{aligned}
$$

標本の不偏分散の期待値は，母集団の低減母分散であることが分かりました。

結論

抽出を重複組合せで考える場合，$E(v_s) = V_h$

2　要素間の差を使う

分散を計算するとき，各要素の平均からの偏差ではなく，2 つずつの要素の差を使うことにします。2 つの要素をとり出すとき，相異なる要素をとるか，同じ要素を重複してとることを許すかによって，2 通りの分散を考えます。相異なる要素をとる場合を**組合せ分散**，重複を許す場合を**重複順列分散**および**重複組合せ分散**と呼ぶことにします。

2.1　組合せ分散

組合せ分散　$d_s = \dfrac{1}{{}_n\mathrm{C}_2} \displaystyle\sum_{i=1}^{n} \sum_{j=i+1}^{n} (x_i - x_j)^2$

d_s を a, b で表し，他の分散との関係を考えます。

$$
\begin{aligned}
d_s &= \frac{1}{{}_n\mathrm{C}_2} \sum_{i=1}^{n} \sum_{j=i+1}^{n} (x_i - x_j)^2 \\
&= \frac{1}{{}_n\mathrm{C}_2} \sum_{i=1}^{n} \sum_{j=1}^{n} (x_i - x_j)^2 \cdot \frac{1}{2} \quad \text{（← } \textstyle\sum \text{ で足すものを 2 倍に増やして 2 で割っておいた）} \\
&= \frac{1}{{}_n\mathrm{C}_2} \left\{ \begin{array}{|c|c|c|c|c|} \hline (x_1 - x_1)^2 & (x_1 - x_2)^2 & (x_1 - x_3)^2 & \cdots & (x_1 - x_n)^2 \\ \hline (x_2 - x_1)^2 & (x_2 - x_2)^2 & (x_2 - x_3)^2 & \cdots & (x_2 - x_n)^2 \\ \hline (x_3 - x_1)^2 & (x_3 - x_2)^2 & (x_3 - x_3)^2 & \cdots & (x_3 - x_n)^2 \\ \hline \cdots & \cdots & \cdots & \cdots & \cdots \\ \hline (x_n - x_1)^2 & (x_n - x_2)^2 & (x_n - x_3)^2 & \cdots & (x_n - x_n)^2 \\ \hline \end{array} \text{ の和} \right\} \frac{1}{2} \\
&= \frac{1}{{}_n\mathrm{C}_2} \left\{ n({x_1}^2 + \cdots + {x_n}^2) + n({x_1}^2 + \cdots + {x_n}^2) - 2(x_1 + \cdots + x_n)(x_1 + \cdots + x_n) \right\} \frac{1}{2} \\
&= \frac{1}{{}_n\mathrm{C}_2} \left\{ n({x_1}^2 + \cdots + {x_n}^2) - (x_1 + \cdots + x_n)^2 \right\} \\
&= \frac{1}{{}_n\mathrm{C}_2} (nb - a^2) \\
&= \frac{2}{n(n-1)} (nb - a^2) \\
&= 2 \cdot \frac{n}{n-1} \left(\frac{b}{n} - \frac{a^2}{n^2} \right) \\
&= 2 \cdot \frac{n}{n-1} v_p \\
&= 2 v_s
\end{aligned}
$$

結論

組合せ分散 d_s は不偏分散 v_s と相似であり，$d_s = 2v_s$

2.2 重複順列分散

重複順列分散 $d_p = \dfrac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - x_j)^2$

d_p を a, b で表し，他の分散との関係を考えます。

$$
\begin{aligned}
d_p &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - x_j)^2 \\
&= \frac{1}{n^2}\left\{ \begin{array}{|c|c|c|c|c|} \hline (x_1-x_1)^2 & (x_1-x_2)^2 & (x_1-x_3)^2 & \cdots & (x_1-x_n)^2 \\ \hline (x_2-x_1)^2 & (x_2-x_2)^2 & (x_2-x_3)^2 & \cdots & (x_2-x_n)^2 \\ \hline (x_3-x_1)^2 & (x_3-x_2)^2 & (x_3-x_3)^2 & \cdots & (x_3-x_n)^2 \\ \hline \cdots & \cdots & \cdots & \cdots & \cdots \\ \hline (x_n-x_1)^2 & (x_n-x_2)^2 & (x_n-x_3)^2 & \cdots & (x_n-x_n)^2 \\ \hline \end{array} \text{の和} \right\} \\
&= \frac{1}{n^2}\left\{ n({x_1}^2 + \cdots + {x_n}^2) + n({x_1}^2 + \cdots + {x_n}^2) - 2(x_1 + \cdots + x_n)(x_1 + \cdots + x_n) \right\} \\
&= \frac{2}{n^2}\left\{ n({x_1}^2 + \cdots + {x_n}^2) - (x_1 + \cdots + x_n)^2 \right\} \\
&= \frac{2}{n^2}(nb - a^2) \\
&= 2\left(\frac{b}{n} - \frac{a^2}{n^2}\right) \\
&= 2v_p
\end{aligned}
$$

結論

重複順列分散 d_p は標準分散 v_p と相似であり，$d_p = 2v_p$

2.3 重複組合せ分散

重複組合せ分散 $\quad d_h = \dfrac{1}{{}_n\mathrm{H}_2} \sum\limits_{i=1}^{n} \sum\limits_{j=i}^{n} (x_i - x_j)^2$

d_h を a, b で表し，他の分散との関係を考えます。

$$
\begin{aligned}
d_h &= \frac{1}{{}_n\mathrm{H}_2} \sum_{i=1}^{n} \sum_{j=i}^{n} (x_i - x_j)^2 \\
&= \frac{1}{{}_n\mathrm{H}_2} \sum_{i=1}^{n} \sum_{j=i+1}^{n} (x_i - x_j)^2 \quad （\leftarrow i = j \text{ のときは } x_i - x_j = 0 \text{ なので}） \\
&= \frac{1}{{}_n\mathrm{H}_2}(nb - a^2) \quad （\leftarrow \text{組合せ分散より}） \\
&= \frac{1}{{}_{n+1}\mathrm{C}_2}(nb - a^2) \\
&= \frac{2}{n(n+1)}(nb - a^2) \\
&= 2 \cdot \frac{n}{n+1}\left(\frac{b}{n} - \frac{a^2}{n^2}\right) \\
&= 2 \cdot \frac{n}{n+1} v_p \\
&= 2v_h
\end{aligned}
$$

結論

重複組合せ分散 d_h は低減分散 v_h と相似であり，$d_h = 2v_h$

3　2次元への拡張

n 個の点 $(x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots, (x_n, y_n)$ に対し，これらの分散を考えます。

不偏分散	$v_s = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 + \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2$
標準分散	$v_p = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 + \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2$
低減分散	$v_h = \frac{1}{n+1}\sum_{i=1}^{n}(x_i - \overline{x})^2 + \frac{1}{n+1}\sum_{i=1}^{n}(y_i - \overline{y})^2$
組合せ分散	$d_s = \frac{1}{{}_n\mathrm{C}_2}\sum_{i=1}^{n}\sum_{j=i+1}^{n}\left\{(x_i - x_j)^2 + (y_i - y_j)^2\right\}$
重複順列分散	$d_p = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\{(x_i - x_j)^2 + (y_i - y_j)^2\right\}$
重複組合せ分散	$d_h = \frac{1}{{}_n\mathrm{H}_2}\sum_{i=1}^{n}\sum_{j=i}^{n}\left\{(x_i - x_j)^2 + (y_i - y_j)^2\right\}$

同様に，$d_s = 2v_s$，$d_p = 2v_p$，$d_h = 2v_h$ が成り立ちます。