

Jackknife法による誤差の見積もりについて

Etsuko Itou

April 7, 2023

独立な統計データ

$$A_1, A_2, \dots, A_n$$

があったとする。この時、標準分散を $\tilde{\sigma}$ 、統計誤差を σ と書くと

$$\begin{aligned}\tilde{\sigma}^2 &= \frac{1}{n-1} \sum_{i=1}^n (A_i^2 - \langle A_i \rangle^2) \\ \sigma^2 &= \frac{1}{n} \tilde{\sigma}^2\end{aligned}$$

と書ける。ここで、 $\langle A_i \rangle$ は、 A_i の平均値で

$$\langle A_i \rangle = \frac{1}{n} \sum_{i=1}^n A_i$$

次に以下のように jackknife サンプルを定義する。

$$A_{jack}(i) = \frac{\sum_{j \neq i} A_j}{n-1}$$

つまり、

$$\begin{aligned}A_{jack}(1) &= \frac{1}{n-1} [A_2 + A_3 + \dots + A_{n-1} + A_n] \\ A_{jack}(2) &= \frac{1}{n-1} [A_1 + A_3 + \dots + A_{n-1} + A_n] \\ &\dots \\ A_{jack}(n) &= \frac{1}{n-1} [A_1 + A_2 + A_3 + \dots + A_{n-1}] \end{aligned}$$

とする。すると、

$$\sum_i^n A_{jack}(i) = \sum_i^n A_i \quad (1)$$

$$\langle A_{jack}(i) \rangle = \frac{1}{n} \sum_i^n A_{jack}(i) = \langle A_i \rangle \quad (2)$$

が成り立つ。

一方で、jackknife サンプルの分散を求めると、定義から

$$\begin{aligned} \tilde{\sigma}_{jack}^2 &= \frac{1}{n} \sum_{i=1}^n (A_{jack}^2(i) - \langle A_{jack}(i) \rangle^2) \\ &= \frac{1}{n} \sum_{i=1}^n \left(\left(\frac{\sum_{j \neq i} A_j}{n-1} \right)^2 - \langle A_{jack}(i) \rangle^2 \right) \end{aligned} \quad (3)$$

となる。ここで、

$$\begin{aligned} \left(\sum_{j \neq i} A_j \right)^2 &= ((A_1 + A_2 + \dots + A_n) - A_i)^2 \\ &= \left(\sum_{i=1}^n A_i \right)^2 - 2A_i \left(\sum_{i=1}^n A_i \right) + A_i^2 \end{aligned}$$

なので、

$$\sum_{i=1}^n \left(\sum_{j \neq i} A_j \right)^2 = n \left(\sum_{i=1}^n A_i \right)^2 - 2 \left(\sum_{i=1}^n A_i \right)^2 + \sum_{i=1}^n A_i^2$$

となる。これを式 (3) に代入すると、

$$\begin{aligned} \tilde{\sigma}_{jack}^2 &= \frac{1}{(n-1)^2} \left(\frac{\sum_i^n A_i^2}{n} - \frac{1}{n^2} \left(\sum_{i=1}^n A_i \right)^2 \right) \\ &= \frac{1}{(n-1)^2} \tilde{\sigma}^2 \end{aligned}$$

とかける。以上より、データの統計誤差は、jackknife サンプルの分散を用いて

$$\sigma = \sqrt{n-1} \tilde{\sigma}_{jack} \quad (4)$$

と計算できる。