JACKKNIFE METHOD

Suppose we wish to compute $\langle f \rangle_\pi$ and $J(\langle f \rangle_\pi)$

Generate N number with probability $\pi(x)$

$$X_1 \dots\dots X_N$$

Define the <u>Jackknife</u> averages

$$\bar{f}_i = \frac{1}{N-1} \sum_{\substack{j=1 \\ j \neq i}}^{N} f(X_j) \qquad J_i = J(\bar{f}_i)$$

$\bar{f}_i$ is the sample mean in which we consider all data, except X_i

Example $N = 4$

We compute

$$\bar{f}_1 = \frac{1}{3}\left(f(X_2) + f(X_3) + f(X_4)\right) \qquad J_1 = J(\bar{f}_1)$$

$$\bar{f}_2 = \frac{1}{3}\left(f(X_1) + f(X_3) + f(X_4)\right) \qquad J_2 = J(\bar{f}_2)$$

$$\bar{f}_3 = \frac{1}{3}\left(f(X_1) + f(X_2) + f(X_4)\right) \qquad J_3 = J(\bar{f}_3)$$

$$\bar{f}_4 = \frac{1}{3}\left(f(X_1) + f(X_2) + f(X_3)\right) \qquad J_4 = J(\bar{f}_4)$$

Note
$$\frac{1}{4}\sum_{i=1}^{4} \bar{f}_i = \frac{1}{3}\cdot\frac{1}{4}\left(3f(X_1) + 3f(X_2) + 3f(X_3) + 3f(X_4)\right)$$

$$= \frac{1}{4}\left(f(X_1) + f(X_2) + f(X_3) + f(X_4)\right)$$

$$= \bar{f}$$

This result is true in general

$$\mu_{JK} = \frac{1}{N}\sum_i \bar{f}_i = \frac{1}{N}\sum_i \left[\frac{1}{N-1}\sum_{\substack{j=1\\ j\neq i}}^{N} f(x_j)\right]$$

$$= \frac{1}{N(N-1)} \sum_{i=1}^{N} \left(\sum_{j=1}^{N} f(x_j) - f(x_i) \right)$$

[we add $f(x_i)$ to the sum]

$$= \frac{1}{N(N-1)} \sum_{i=1}^{N} \left[N\bar{f} - f(x_i) \right]$$

$$= \frac{1}{N(N-1)} \left(\sum_{i=1}^{N} N\bar{f} - \sum_{i=1}^{N} f(x_i) \right)$$

$$= \frac{1}{N(N-1)} \left(N^2\bar{f} - N\bar{f} \right) = \bar{f}$$

For the sample mean the average of the jackknife averages is equivalent to $\bar{f}$.

$$\sigma^2_{JK} = \frac{1}{N}\sum_i \bar{f}_i^{\,2} - \mu_{JK}^2 \qquad \text{(jackknife variance)}$$

Now

$$\frac{1}{N}\sum_{i=1}^{N} \bar{f}_i^{\,2} = \frac{1}{N}\sum_{i=1}^{N} \left(\frac{1}{N-1} \sum_{\substack{j=1\\ j\neq i}}^{N} f(x_i) \right)^2$$

$$= \frac{1}{N(N-1)^2} \sum_{i=1}^{N} \left(\sum_{j=1}^{N} f(x_j) - f(x_i) \right)^2$$

$$= \frac{1}{N(N-1)^2} \sum_{i=1}^{N} \left(N\bar{f} - f(x_i) \right)^2$$

$$= \frac{1}{N(N-1)^2} \sum_{i=1}^{N} \left(N^2 \bar{f}^2 - 2N\bar{f}\, f(x_i) + f(x_i)^2 \right)$$

$$= \frac{1}{N(N-1)} \left[N^2 \bar{f}^2 \underbrace{\sum_{i=1}^{N} 1}_{N} - 2N\bar{f} \underbrace{\sum_{i=1}^{N} f(x_i)}_{N\bar{f}} + \underbrace{\sum_{i=1}^{N} f(x_i)^2}_{N\overline{f^2}} \right]$$

$$= \frac{1}{(N-1)^2}\, \overline{f^2} + \frac{N(N-2)}{(N-1)^2}\, \bar{f}^2$$

We plug in the result in σ^2_{JK} $\quad (\mu_{JK} = \bar{f})$

$$\sigma^2 = \frac{1}{(N-1)^2}\, \overline{f^2} + \frac{N(N-2)}{(N-1)^2}\, \bar{f}^2 - \bar{f}^2 =$$

Correction: here sigma^2 is sigma^2_JK

$$= \frac{1}{(N-1)^2} \left(\overline{f^2} - \bar{f}^2 \right)$$

Now note that the error σ on $\bar{f}$ is

$$\sigma^2 = \frac{1}{(N-1)} \left(\overline{f^2} - \bar{f}^2 \right) \longrightarrow \sigma^2 = (N-1)\sigma^2_{JK}$$

Thus, we can estimate

$$\sigma = \sqrt{(N-1)\, \sigma_{JK}}$$

- To compute the error on $\bar{f}$ there is <u>NO ADVANTAGE</u> in using the jackknife method
- Results are equivalent

The jackknife method is useful for estimating $J(\langle f \rangle_\pi)$

AVERAGE

jackknife estimate

$$\text{est}\left[J(\langle f\rangle_\pi)\right] = N J(\bar{f}) - \frac{(N-1)}{N} \sum_{i=1}^{N} J(\bar{f}_i)$$

ERROR

$$\sigma^2 = (N-1)\left[\frac{1}{N} \sum_{i=1}^{N} J(\bar{f}_i)^2 - \left(\frac{1}{N} \sum_{i=1}^{N} J(f_i) \right)^2 \right]$$

CORRECTION: Add a bar on top of the last term. The error depends on J_i = J(bar(f)_i)

Justification of the error:

ⓐ It is equivalent to the error propagation formula for $N \to \infty$ (N large)

ⓑ It provides a stable method to estimate the error even if N is not large

Proof of ⓐ (exercise):

note that $\bar{f}_i = \frac{1}{N-1} \sum_{\substack{j=1 \\ j \neq i}}^{N} f(x_j) = \bar{f} + \underbrace{\frac{1}{N-1}(\bar{f} - f(x_i))}_{\Delta}$

Δ is small so that

$$J(\bar{f}_i) = J(\bar{f}) + J'(\bar{f})\,\Delta_i + \text{negligible terms.}$$

DEFINITION OF THE AVERAGE: MOTIVATIONS

$$\text{est.}\left[J(\langle f\rangle_\pi)\right] = N J(\bar{f}) - \frac{N-1}{N}\sum_{i=1}^{N} J(\bar{f}_i)$$

- $$\langle J(\bar{f})\rangle_{MC} = J(\langle f\rangle_\pi) + \frac{1}{2N} J''(\langle f\rangle_\pi)\, Var_\pi f + O(N^{-2})$$
$$= J(\langle f\rangle_\pi) + \frac{a}{N} + O(N^{-2})$$

- $$\langle J(\bar{f}_i)\rangle_{MC} = J(\langle f\rangle_\pi) + \frac{a}{N-1} + O(N^{-2})$$ [of course, independent of i]

[$\bar{f}_i$ is indeed a sample mean over $(N-1)$ measures]

It follows

$$\left\langle N J(\bar{f}) - \frac{N-1}{N}\sum_{i=1}^{N} J(\bar{f}_i)\right\rangle_{MC}$$

$$= N\left(J(\langle f\rangle_\pi + \frac{a}{N}\right) - \frac{N-1}{N}\cdot \underset{\substack{\uparrow\\ \text{there are } N \text{ terms in the sum}}}{N}\left(J(\langle f\rangle_\pi + \frac{a}{N-1}\right)$$

$$= N J(\langle f\rangle_\pi) + a - (N-1)\left[J(\langle f\rangle_\pi) + \frac{a}{N-1}\right]$$

$$= J(\langle f\rangle_\pi) + O(N^{-2})$$

The average is defined so as to eliminate the bias of order N