# Estimate of the error

We want to compute the error on the sample mean, which is again defined as

$$\sigma^2 = \left\langle \left( \frac{1}{N+1} \sum_{n=0}^{N} f(X_n) - \sum_x \pi_x f(x) \right)^2 \right\rangle_{MC,0}$$

where, the average is taken over all possible histories (MC repetitions) that start at $X_0$.

Note: This expression is correct only if there is no bias on the sample mean or if the bias is negligible. For generic $X_0$ this occurs for large values of $N$. The formula is valid for any $N$ if we start in equilibrium (the only case we will consider, see slide 3).

**THEOREM**: The limit

$$\lim_{N\to\infty} N\sigma^2$$

is finite and independent of the initial configuration.

Thus, also in the case of dynamic Monte Carlo fluctuations (errors) decrease as $1/\sqrt{N}$.

We wish now to estimate $\sigma^2$ from the data. If

$$F = \sum_x \pi_x f(x)$$

we have

$$\sigma^2 = \frac{1}{(N+1)^2} \sum_{n=0}^{N} \sum_{m=0}^{N} \langle (f(X_n) - F)(f(X_m) - F) \rangle_{MC,0}$$

Now we define the autocorrelation function $C_f(n,m,X_0)$ :

$$C_f(n,m,X_0) = \langle (f(X_n) - F)(f(X_m) - F) \rangle_{MC,0}$$

so that

$$\sigma^2 = \frac{1}{(N+1)^2} \sum_{n=0}^{N} \sum_{m=0}^{N} C_f(n,m;X_0)$$

The autocorrelation function depends on $X_0$, while the large-$N$ limit of $\sigma$ does not.

**We now assume we start in equilibrium**. From a mathematical point, starting in equilibrium means averaging over the starting point with probability $\pi$: we average the behavior of the system with respect to the initial state using the equilibrium distribution.

In practice, starting in equilibrium means considering a typical configuration as starting conf (if we choose a configuration with probability $\pi_x$, we will obviously select a typical configuration).

In equilibrium, configurations at time $t = n$ are distributed according to $\pi$.

$$\text{prob of } y \text{ at time } n = \sum_{x_0} \pi_{x_0} P^n_{x_0 y}$$

But now we use repeatedly the stationarity relation $\sum_x \pi_x P_{xy} = \pi_y$ so that

$$\sum_{x_0} \pi_{x_0} P^n_{x_0 y} = \sum_{x_0} \pi_{x_0} \sum_z P_{x_0 z} P^{n-1}_{zy} = \sum_z [\sum_{x_0} \pi_{x_0} P_{x_0 z}] P^{n-1}_{zy} = \sum_z \pi_z P^{n-1}_{zy} = \ldots = \pi_y$$

Configurations have always the same distribution, independently of time: the system is time-translation invariant.

We define the equilibrium autocorrelation function:

$$C_f(n,m;\mathrm{eq}) = \sum_x \pi_x C_f(n,m,x)$$

The equilibrium autocorrelation function is time-translation invariant, hence depends only on $|n-m|$: $C_f(n,m;\mathrm{eq}) = C_f(|n-m|)$. Therefore,

$$\sigma^2 = \frac{1}{(N+1)^2}\sum_{n=0}^{N}\sum_{m=0}^{N} C_f(|n-m|)$$

Now, we wish to replace the two sums with a single sum over $k = n-m$.

To understand this point, let us imagine $N = 2$ and compute ($g(n)$ is a generic function)

$$\sum_{n=0}^{N}\sum_{m=0}^{N} g(n-m) = \sum_{n=0}^{2}[g(n)+g(n-1)+g(n-2)]$$

Now we make the summation over $n$ explicit. We obtain

$$g(0)+g(-1)+g(-2) \qquad \text{here } n=0$$

$$g(1)+g(0)+g(-1) \qquad \text{here } n=1$$

$$g(2)+g(1)+g(0) \qquad \text{here } n=2$$

Summing all terms, we get

$$3g(0)+2g(1)+2g(-1)+g(2)+g(-2)$$

For generic values of $N$ we have

$$\sum_{n=0}^{N}\sum_{m=0}^{N} g(n-m) = (N+1)g(0) + N[g(1)+g(-1)] + (N-1)[g(2)+g(-2)] + \ldots g(N) + g(-N)$$

that is

$$\sum_{n=0}^{N}\sum_{m=0}^{N} g(n-m) = (N+1)g(0) + \sum_{k=1}^{N}(N+1-k)[g(k)+g(-k)]$$

It follows

$$\sigma^2 = \frac{1}{(N+1)^2}\left[(N+1)C_f(0) + 2\sum_{k=1}^{N}(N+1-k)C_f(k)\right]$$

$$\sigma^2 = \frac{1}{(N+1)}\left[C_f(0) + 2\sum_{k=1}^{N}C_f(k)\right] - \frac{2}{(N+1)^2}\sum_{k=1}^{N}kC_f(k)$$

The function $C_f(|k|)$ decays exponentially (not at continuous transitions but we are not discussing this issue here). This guarantees that for $N \to \infty$ we have
(a) $\sum_{k=1}^{N} C_f(k) \approx \sum_{k=1}^{\infty} C_f(k)$, with (exponentially small) corrections that vanish as $N \to \infty$.
(b) $\sum_{k=1}^{N} k C_f(k)$ is finite for $N \to \infty$, so that the last term in the expression of $\sigma^2$ is of order $N^{-2}$.
Hence, we have proved that the error is of order $1/N$ and obtained the relation

$$\sigma^2 = \frac{1}{N+1}\left[C_f(0) + 2\sum_{k=1}^{\infty} C_f(k)\right].$$

**Observation:** $C_f(0)$ is the variance with respect to $\pi$
We use the definition

$$C_f(n,n;X_0) = \langle (f(X_n) - F)^2 \rangle_{MC,0} = \sum_x P^n_{X_0 x}(f(x) - F)^2$$

Now we assume that we are in equilibrium:

$$C_f(0) = \sum_{x_0} C_f(n,n;x_0)\pi_{x_0} = \sum_{x_0}\sum_x \pi_{x_0} P^n_{x_0 x}(f(x) - F)^2 = \sum_x (f(x) - F)^2 \sum_{x_0} \pi_{x_0} P^n_{x_0 x}$$

$$C_f(0) = \sum_x (f(x) - F)^2 \pi_x = \langle (f(x) - F)^2 \rangle_\pi = \mathrm{Var}_\pi f$$

**Observation.** For static algorithms we obtain the usual formula (do not get confused by the $N+1$ in the normalization: here we sum from $0$ to $N$ and therefore there are $N+1$ data).
Indeed, if $n \neq m$ we have $\langle f(X_n) f(X_m) \rangle_{MC} = \langle f(X_n) \rangle_{MC} \langle f(X_m) \rangle_{MC}$. It follows $C_f(n,m;X_0) = 0$ if $n \neq m$.
For uncorrelated data the autocorrelation function $C_f(n)$ vanishes except for $n = 0$.

It follows

$$\sigma^2 = \frac{1}{N+1} C_f(0) = \frac{1}{N+1} \mathrm{Var}_\pi f$$

# Integrated autocorrelation time

We define an **integrated autocorrelation time** $\tau_{\text{int},f}$ as

$$\tau_{\text{int},f} = \frac{1}{2} + \sum_{k=1}^{\infty} \frac{C_f(k)}{C_f(0)}.$$

The error becomes

$$\sigma^2 = \frac{C_f(0)}{N+1} 2\tau_{\text{int},f}.$$

The definition of $\tau_{\text{int},f}$ can be understood by considering the simple case in which $C_f(k) = C_f(0)\exp(-k/\tau)$. We use the result for the geometric series $(1-x)^{-1} = \sum_n x^n$, so that

$$\sum_{k=1}^{\infty} e^{-k/\tau} = -1 + \sum_{k=0}^{\infty} e^{-k/\tau} = -1 + \frac{1}{1-e^{-1/\tau}}$$

It follows

$$\tau_{\text{int},f} = -\frac{1}{2} + \frac{1}{1-e^{-1/\tau}} \approx \tau$$

where the last equality holds for $\tau \gg 1$ ($1/\tau$ is small so that we can expand the expression in powers of $1/\tau$, $e^{-1/\tau} \approx 1 - 1/\tau$).
For uncorrelated data $\tau_{\text{int,f}} = 1/2$.