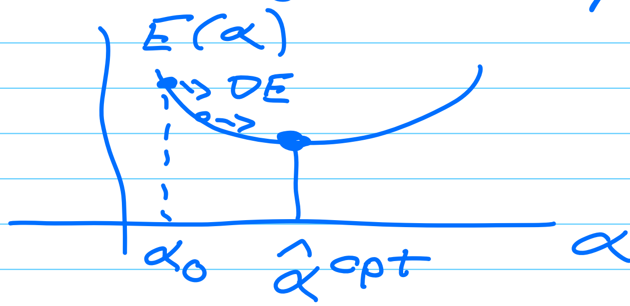
FYS 4411/9411, MARCH 16, 2023

- VMC machinery

  - ✓ Metropolis–Hastings (importance sampling)

  - ✓ Optimization of variational parameters



find $\hat{\alpha}^{opt}$ with few cycles in order to launch the final VMC calculations with $\hat{\alpha}^{opt}$ in order to have a reliable statistical determination of various expectation values.

- Resampling techniques used to provide a "reliable" estimation of expectation values with statistical uncertainties.

$$\mathbb{E}[f(x)] = \int_{x \in D} p(x) f(x) dx$$

$$= \mu_f$$

$$\mathbb{E}[f(x)^n] = \int_{x \in D} p(x) f^n(x)\, dx$$

$$\mathrm{var}[f(x)] = \sigma_f^2 =$$

$$\mathbb{E}[f^2(x)] - (\mathbb{E}[f(x)])^2$$

$$= \int_{x \in D} p(x) (f(x) - \mu_f)^2 dx$$

$$\simeq \frac{1}{M} \sum_{i=1}^{M} (f(x_i) - \bar{\mu}_f)^2$$

$$(\bar{\mu}_f \neq \mu_f) = \bar{\sigma}_f^2$$

$$\parallel$$

sample variance

STD (standard deviation)

$$= \sqrt{\bar{\sigma}_f^2}$$

$$\bar{\mu}_f \neq \mu_f \wedge \bar{\sigma}_f^2 \neq \sigma_f^2$$

- Parallelization

<u>Statistical elements</u>

i.i.d = independent

and identically distributed stochastic events

$$\mathrm{cov}(x_i\, x_j) = \int dx_i\, dx_j\, P(x_i\, x_j)\,(x_i - \bar{x}_i)(x_j - \bar{x}_j)$$

iid : $P(x_i\, x_j) = \underbrace{p(x_i)\, p(x_j)}$

$$\bar{x}_i = \bar{x}_j = \bar{x} = \int x\, p(x)\, dx$$

$$\mathrm{cov}(x_i\, x_j) = \int dx_i\, p(x_i)\,(x_i - \bar{x}) \int dx_j\, p(x_j)\,(x_j - \bar{x})$$

$$\langle x_i\, x_j \rangle - \bar{x}^2 = 0$$

$$\underbrace{\int dx_i\, p(x_i)\, x_i}_{\bar{x}}\ \underbrace{\int dx_j\, p(x_j)\, x_j}_{\bar{x}}$$

<u>Central limit theorem</u>

$x \in$ iid $p(x)$

series of measurements

$$X = \{x_0, x_1, \ldots, x_{n-1}\}$$

$$\bar{x}_0 = \frac{1}{n}\sum_{i=0}^{n-1} x_{0i}$$

sloppy notation $\bar{x}_0 \Rightarrow x_0$

$$\bar{z} = \frac{x_0 + x_1 + \cdots x_{n-1}}{n} = z$$

which $p(z)dz$?

Central limit theorem

$p(z)dz$ is a gaussian

$$p(z)dz = \frac{1}{\sqrt{2\pi\sigma^2/n}} \exp\left[-\frac{(z-\mu)}{2(\sigma/\sqrt{n})^2}\right]$$

$$\mu = \int x p(x) dx$$

$$\sigma^2 = \int (x-\mu)^2 p(x) dx$$

new distribution with variance σ^2/n

$$STD = \sigma/\sqrt{n}$$

$$\bar{z} = z = \mu$$

Toward the Blocking method

assume we have m - experiments and each experiment has n - observations each experiment – α –

$$\mu_\alpha = \frac{1}{n}\sum_{k=1}^{n} x_{\alpha,k}$$

$$\sigma_\alpha^2 = \frac{1}{n}\sum_{k=1}^{n}\left(x_{\alpha,k}-\mu_\alpha\right)^2$$

– repeat – m – times –

$$\mu_m = \frac{1}{m}\sum_{\alpha=1}^{m}\mu_\alpha$$

$$= \frac{1}{mn}\sum_{\alpha k} x_{\alpha,k}$$

Total variance

$$\sigma_m^2 = \frac{1}{m}\sum_{\alpha=1}^{m}\left(\mu_\alpha-\mu_m\right)^2$$

$$= \frac{1}{mn}\sum_{\alpha k}\left(x_{\alpha,k}-\mu_m\right)^2$$

$$= \frac{1}{m}\sum_{\alpha=1}^{m}\left(\frac{1}{n}\sum_{k=1}^{n}x_{\alpha,k}\,\frac{1}{n}\sum_{l=1}^{n}x_{\alpha,l}\right)$$

$$-\ \mu_m^2$$

$$= \frac{1}{mn^2} \sum_{\alpha=1}^{m} \sum_{kl=1}^{n} (x_{\alpha,k} - \mu_m) \times (x_{\alpha,l} - \mu_m)$$

$$+ \frac{1}{mn^2} \sum_{\alpha k} (x_{\alpha,k} - \mu_m)^2$$

$$+ \frac{2}{mn^2} \sum_{\alpha=1}^{m} \sum_{k<l}^{n} (x_{\alpha,k} - \mu_m) \times (x_{\alpha l} - \mu_m)$$

sample variance of all mn - experiments

$$\sigma^2 = \frac{1}{mn} \sum_{\alpha k} (x_{\alpha k} - \mu_m)^2$$

$$\sigma_m^2 = \frac{\sigma^2}{n} + \text{cov}$$

$$\text{cov} = \frac{2}{mn^2} \sum_{\alpha=1}^{m} \sum_{k<l}^{n} (x_{\alpha k} - \mu_m) \times (x_{\alpha,l} - \mu_m)$$

if $\text{cov} = 0$

$$\sigma_m^2 = \sigma^2 / M \quad (\text{single loop})$$

cov includes a double loop $\sum_{k<l}^{M}$

Strategy : is to produce an array of observations $X_{\alpha,k}$

in the statistical postanalysis we will use these $X_{\alpha,k}$

1 find the optimal $\hat{\alpha}$

2 Run an VMC calculation parallelize and produce $X_{\alpha,k}$

3 statistical postanalysis to evaluate $\langle X_{\alpha,k} \rangle$ and σ^2

Bootstrap (jackknife) and blocking evaluate the cov without evaluating the double loop.

introduce a shorthand

$$f_d = \frac{1}{nm}\sum_{\alpha=1}^{m}\sum_{k=1}^{n-d}(x_{\alpha,k}-\mu_m)$$

$$\times(x_{\alpha+d}-\mu_m)$$

$$d = |k-l|$$

autocorrelation function

$$R_d = \frac{f_d}{\sigma^2}$$

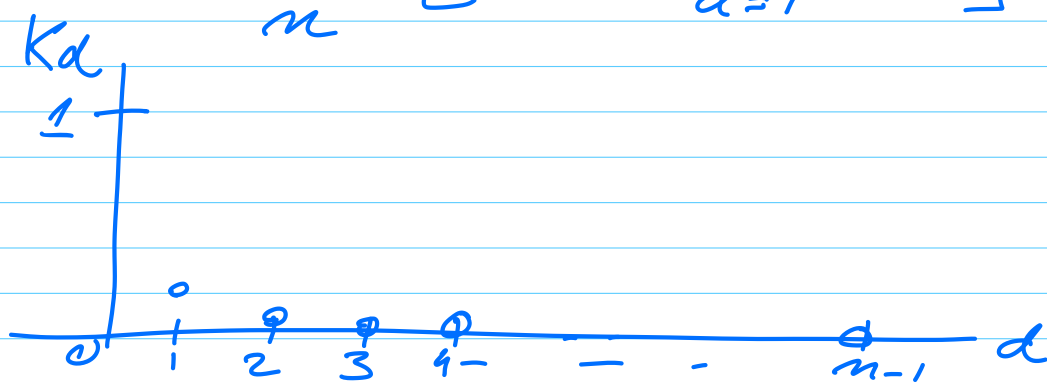
$$\underset{d=0}{f_0} = \frac{1}{m}\frac{1}{n}\sum_{\alpha=1}^{m}\sum_{k=1}^{n}(x_{\alpha k}-\mu_m)^2$$

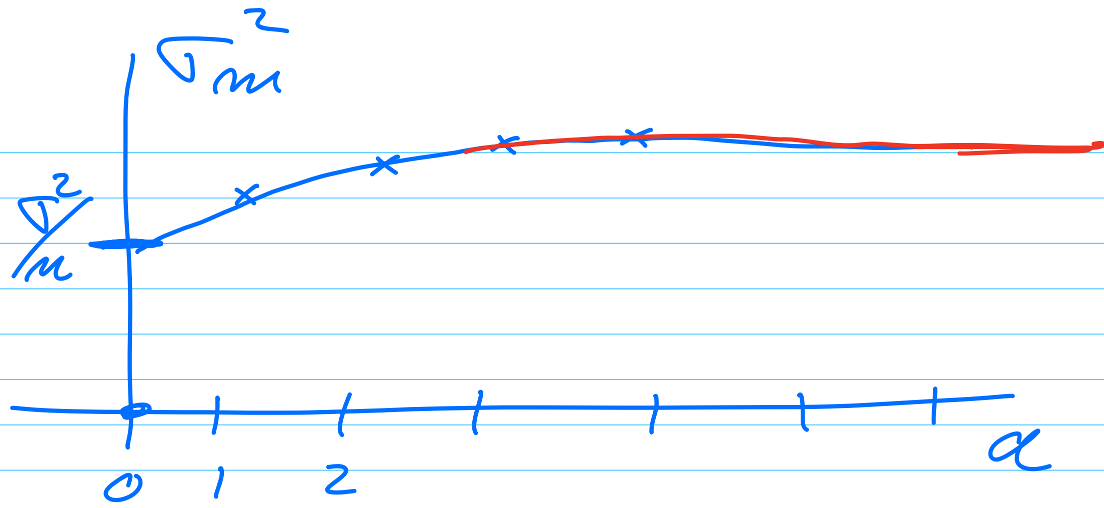
$$= \sigma^2$$

$$R_0 = \sigma^2/\sigma^2 = 1$$

$$\sigma_m^2 = \frac{\sigma^2}{n} + \frac{2}{n}\sum_{d=1}^{n-1} f_d$$

$$= \frac{\sigma^2}{n}\left[1 + 2\sum_{d=1}^{n-1} R_d\right]$$





Resampling: Bootstrap.

$X = \{x_1, x_2, x_3, \ldots x_n\}$

m - bootstrap operations

for $i = 1, m$

- compute μ_i', σ_i^2
- reshuffle data randomly by selecting n points by replacement

$X' = \{x_3, x_5, x_5, x_1, \ldots x_{n-45}\}$

- compute $\mu_{i+1}, \sigma_{i+1}^2$ with X'

- repeat till $i = m$

end DO

Final stage: compute

$$\text{final } \mu = \frac{1}{m}\sum_{i=1}^{m} \mu_i$$

$$\text{final } \sigma^2 = \frac{1}{m}\sum_{i=1}^{m} (\mu_i - \mu)^2$$

For large data sets, requires many FLOPS to perform the statistical post analysis

$\Rightarrow$ Blocking method.