FYS 4411 MARCH 11

Central limit theorem

n - measurements

$$\underline{\mu_\alpha} = \frac{1}{n}\sum_{i=1}^{n} x_{\alpha i}$$

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

sample variance

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$$

These X - measurements are iid

Repeat these measurements m - times

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

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

same PDF for all $x_{\alpha k}$

CLT states

$\mu_\alpha \Rightarrow \mu \Rightarrow$

$$\mu_m = \frac{1}{m}\sum_{i=1}^{m} \mu = \mu$$

$$\sigma_m^2 = \frac{\sigma^2}{m}$$

PDF on μ_m and σ_m^2

is $N\left(\mu_m, \frac{\sigma^2}{m}\right)$

Resampling technique:

Bootstrap:

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

1) compute μ and σ^2

2) reshuffle data randomly by selecting n data points with replacement

$$X' = \{x_3, x_5, x_5, x_1, x_{100} \ldots\}$$

compute σ, μ.

3) Repeat 2) - m times.

4) $$\sigma_m^2 = \frac{1}{m} \sum_{i=1}^{m} \sigma^2(i)$$

time consuming when m is large

Blocking

sample mean

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

$$var(x) = \frac{1}{n} \sum_{k=1}^{n} (x_{\alpha k} - \mu)^2$$

m- times

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

Total variance:

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

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

Double loop

covariance

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

Diag $k = l$

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

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

$$\sigma_m^2 = \frac{\sigma^2}{n} + \frac{2}{mn^2} \sum_{\alpha} \sum_{k<l}^{n}$$

introduce f_d

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

$$d = |k - l|$$

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

autocorrelation

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

$d=0$

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

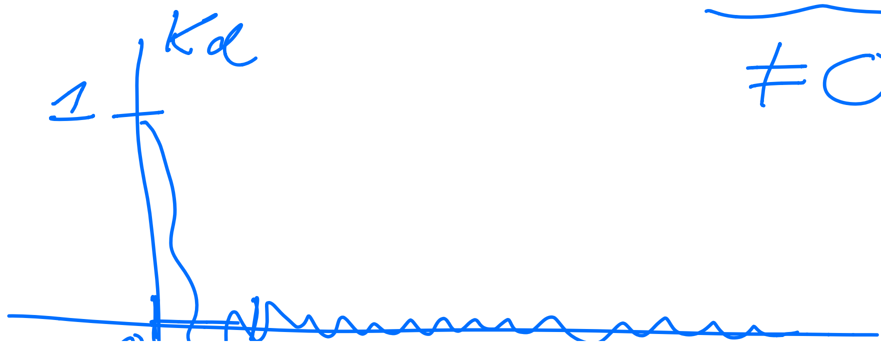
$$= \sigma^2$$

$$K_0 = 1$$

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

$m=1$

$$= \frac{\sigma^2}{n}\left[1 + 2\underbrace{\sum_{d=1}^{n-1} K_d}_{\neq 0}\right]$$



$\underline{\alpha}$

Blocking method

Flyvbjerg - Petersen

J chem Phys 91, 461 (1989)

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\sigma^2(\mu) = \mathbb{E}[\mu^2] - \mathbb{E}[\mu]^2$$

$$\boxed{m = 1}$$

$$\sigma_m^2 = \frac{\sigma^2}{n} + \underbrace{\frac{2}{n^2}\sum_{k<l}(x_l-\mu_l)(x_k-\mu_k)}_{\text{covariance}}$$

$$= \sigma^2(\mu)$$

$$\gamma_{ij} = \mathbb{E}[x_i x_j] - \mathbb{E}[\mu]^2$$

$$= \gamma_t \qquad t = |i-j|$$

$$\boxed{\sigma^2(\mu) = \frac{1}{n^2}\sum_{ij}\gamma_{ij}}$$

$$= \frac{1}{n}\left[\gamma_0 + 2\sum_{t=1}^{n-1}\left(1-\frac{t}{n}\right)\gamma_t\right]$$

σ^- $\qquad C_t$

$$C_t = \frac{1}{\underset{\Lambda}{\underline{n-t}}} \sum_{k=1}^{n-t=\Lambda} (x_k - \mu)(x_{k+t} - \mu)$$

Λ plays a role of a cutoff.
when $C_t(\Lambda)$ smaller than a specific, then stop.

Blocking algo

Transform data

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

into half as large a data set

$$\{x_1', x_2', \ldots x_{n'}'\} \qquad n' = \frac{1}{2} n$$

$$x_i' = \frac{1}{2}\left[x_{2i-1} + x_{2i}\right]$$

$$\mu' = \mu \quad \wedge \quad \sigma(\mu') = \sigma(\mu)$$

$$X = \{x_1\ x_2\ x_3\ x_4\ x_5\ x_8\}$$

$$= \{1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6\}$$

$$\mu = \frac{1}{6}\sum_{i=1}^{6} x_i = 21/6 = 7/2$$

$$\sigma^2 = \frac{1}{6}\sum_i x_i^2 - \mu^2 = 8/3$$

$$\sigma'^2 = \frac{1}{3}\sum_{i=1}^{3}\left(\frac{x_{2i-1}+x_{2i}}{2}\right)^2 - \underset{\mu}{\mu'^2}$$

$$= \frac{1}{3}\left(\left[\frac{x_1+x_2}{2}\right]^2 + \left[\frac{x_3+x_4}{2}\right]^2 + \left[\frac{x_5+x_6}{2}\right]^2\right)$$

$$- \mu^2 = 8/3$$

$$x_i' \qquad \gamma_{i'j'} = \gamma_t \rightarrow \gamma_{i'j}' = \gamma_t'$$

$$\sigma^2$$

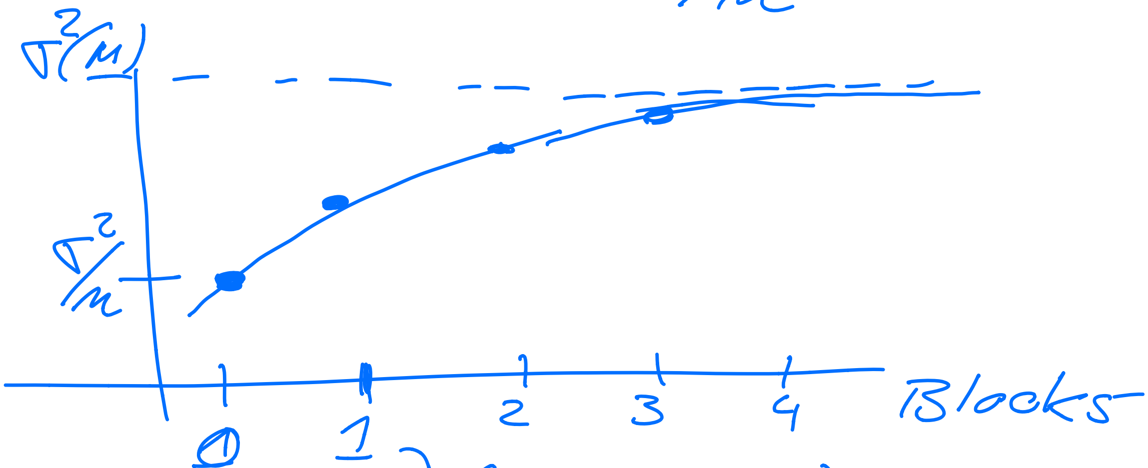
$$\gamma_t' = \begin{cases} \frac{1}{2}\gamma_0 + \frac{1}{2}\gamma_1 & t = 0 \\ \frac{1}{4}\gamma_{2t-1} + \frac{1}{2}\gamma_{2t} + \frac{1}{4}\gamma_{2t+1} & t > 0 \end{cases}$$

$$\underline{\gamma_0'} = \frac{1}{2}\underline{\gamma_0} + \frac{1}{2}\underline{\underline{\gamma_1}}$$

$$\sigma^2(\mu) \geq \frac{\gamma_0}{n} = \frac{\sigma^2}{n}$$

$\gamma_0' > \gamma_0$ one loop

$\sigma^2(m) \approx \gamma_0'/m$



$\{x_1' \, x_2' \, \dots \, x_{m'}'\}$ $m' = \frac{1}{2} m$

$\{x_1 \, x_2 \, \dots \, x_m\}$

$$\sigma^2 = \frac{1}{m} \sum_{k=1}^{m} (x_k - \mu)^2$$

$\{E_1, E_2, \dots \, E_{MCS}\}$

From your MC code

$\rightarrow$ post analysis of data and compute $\mathbb{E}[E_L]$ + $\sigma^2(\mathbb{E}[E_L])$