FYS4411 MARCH 4

Statistical analysis:
Bootstrap & Blocking;
estimates of covariance.
Our stochastic events are
in general correlated, =>
$\mathrm{cov}(x) \neq 0$ => is that
variance

$\mathrm{var}(x)$ gives an over-
optimistic error (standard-
deviation)

iid = independent and
identically distributed
=> $\mathrm{cov}(x) = 0$
$\mathrm{cov}(x) \geq 0$

iid events:

$X : \{x_1 x_2 \dots x_N\}$

PDF of X

$$P(x_1, x_2, \ldots x_N) = p(x_1)\, p(x_2) \cdots p(x_N)$$

$$\mathrm{cov}(x_i', x_j') =$$

$$\int dx_1 \ldots dx_N \, (x_i' - \mu_i')(x_j - \mu_j') \times P(x_1 x_2 \ldots x_N)$$

$$\mu_i' = \frac{1}{n_i'} \sum_{j=1}^{n_i'} x_j'$$

(iid) : $\mu_i' = \mu_j' = \mu$

$$\mathrm{cov}(x_i' x_j') = \int dx_1 \ldots dx_N \, (x_i' - \mu)(x_j - \mu)\, p(x_1)\, p(x_2) \ldots p(x_N)$$

$$\int p(x_i') \, dx_i' = 1 \qquad \int p(x_i') x_i' \, dx_i' = \mu$$

$$= \int dx_i' \, dx_j' \, (x_i' - \mu)(x_j' - \mu)\, p(x_i) \times p(x_j')$$

$$\int dx_i' (\underline{x_i'} - \mu)\, p(x_i') \int dx_j' (x_j - \mu)\, p(x_j)$$

$$= \left(\mu - \mu \int dx_i\, p(x_i)\right)\left(\mu - \mu \int dx_j\, p(x_j)\right)$$

$$= 0$$

$$\mathrm{cov}(x_i, x_j) = \mathbb{E}\left[(x_i - \mu_i)(x_j - \mu_j)\right]$$

$$\mathrm{cov}(X, Y) = \mathbb{E}\left[(x - \mu_x)(y - \mu_y)\right]$$

if $\mathrm{cov}(x_i, x_j) = 0$

$$\mathrm{var}(x_i) = \mathrm{var}(x) = \mathbb{E}[X^2] - \mu_X^2$$

In our case

we don't have the exact μ nor the exact $\mathrm{var}(x)$.

- sample μ

$$\tilde{\mu} = \frac{1}{N}\sum_{i=1}^{N} x_i \quad \left(\frac{1}{N}\sum_{i=1}^{N} F_L(\vec{R}_i)\right)$$

$\neq$ exact μ

- sample variance

$$\tilde{\sigma}^2 = \frac{1}{\ } \sum^{N} (x_i - \tilde{\mu}_i)^2$$

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

$$\neq \mathrm{var}(x)$$

in our codes

Expected value (sample) of $\mathbb{E}[E_L]$ and $\mathbb{E}[E_L^2]$

How do we estimate the covariance?

Central limit theorem

set of stochastic variables

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

we assume that x_i are iid, $p(x_i) = p(x_j) \ldots = p(x_N)$

what is the distribution (PDF) of the mean value

$$z = \frac{x_1 + x_2 + x_3 + \ldots x_N}{N}$$

$\downarrow$

The PDF is a Gaussian

with $\mu = z$ and variance

$$\sigma_N^2 = \frac{\sigma^2}{N} \qquad \sigma^2 = \int (x-\mu)^2 p(x)\,dx$$

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

$\Rightarrow$ STD (iid)

$$= \sigma/\sqrt{N}$$

PDF for z

$$\tilde{p}(z) = \int dx_1\, p(x_1) \int dx_2\, p(x_2)$$

$$\cdots \int dx_N\, p(x_N)\, \delta\left(z - \frac{x_1+x_2+\dots x_N}{N}\right)$$

$$\delta\left(z - \frac{x_1+x_2+\dots+x_N}{N}\right)$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} dq\, \exp\left[iq\left(z - \frac{x_1+x_2+\dots x_N}{N}\right)\right]$$

insert $e^{i\mu q - i\mu q}$

$$\tilde{p}(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dq\, e^{i(z-\mu)q} \times$$

$$\left[\int_{-\infty}^{\infty} dx\, p(x)\, e^{iq(\mu - x)/N}\right]^N$$

$-\infty$

$$\int_{-\infty}^{\infty} dx\, p(x)\, e^{iq(\mu-x)/N}$$

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

$$= \int_{-\infty}^{\infty} dx\, p(x)\left[1 + \frac{iq(\mu-x)}{N} - \frac{q^2(\mu-x)^2}{2N^2} + \dots\right]$$

$$1 + 0 - \frac{q^2}{2N^2}\sigma^2 + \dots$$

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

$\Rightarrow$

$$\left[\int_{-\infty}^{\infty} dx\, p(x)\, e^{iq(\mu-x)/N}\right]^N$$

$$\simeq \left[1 - \frac{q^2\sigma^2}{2N^2}\right]^N$$

$\Rightarrow$

$$\tilde{p}(z) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dq\, e^{iq(z-\mu)}$$

$$\times \left[1 - \frac{q^2\sigma^2}{2N^2}\right]^N$$

$$\tilde{p}(z) = \frac{1}{\sqrt{2\pi}\,\sigma/\sqrt{N}} e^{-\frac{(z-\mu)^2}{2\sigma^2/N}}$$

$$\Rightarrow \quad z \sim N(\mu, \sigma^2/N)$$

each x_i in $\{x_1, x_2, \ldots x_N\}$

has $\mathbb{E}[x] = \mu$ and

$\text{var}(x) = \sigma^2$

$\tilde{p}(z)$ has a STD $\sim \frac{1}{\sqrt{N}}$

$\text{cov}(z) = 0$

sample covariance — we want to estimate it.

$X_\alpha = \{x_{\alpha,1}, x_{\alpha,2}, \ldots x_{\alpha,N}\}$

sample mean

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

sample variance

$$\sigma^2 = \frac{1}{N}\sum_{k=1}^{N} (x_{\alpha,k} - \mu_\alpha)^2$$

Repeat this experiment -M- times.

$$\mu_M = \frac{1}{M} \sum_{\alpha=1}^{M} \mu_\alpha = \frac{1}{MN} \sum_{\alpha k} x_{\alpha,k}$$

<u>Total variance</u>

$$\sigma_M^2 = \frac{1}{M} \sum_{\alpha=1}^{M} \left(\underline{\mu_\alpha} - \mu_M \right)^2$$

$$\mu_\alpha - \mu_M = \frac{1}{MN} \sum_{\alpha k} \left(x_{\alpha k} - \mu_M \right)$$

$$\sigma_M^2 = \frac{1}{M} \sum_{\alpha=1}^{M} \left\{ \mu_\alpha^2 - \mu_\alpha \mu_M - \mu_\alpha \mu_M + \mu_M^2 \right\}$$

$$= \left(\frac{1}{M} \sum_{\alpha=1}^{M} \underline{\mu_\alpha^2} \right) - \mu_M^2$$

$$= \frac{1}{M} \sum_{\alpha=1}^{M} \frac{1}{N} \sum_{k=1}^{N} \frac{1}{N} \sum_{\ell=1}^{N} x_{\alpha,k} x_{\alpha\ell} - \mu_M^2$$

$$= \frac{1}{MN^2} \sum_{\alpha} \sum_{k\ell} \left(x_{\alpha k} - \mu_M \right) \left(x_{\alpha\ell} - \mu_M \right)$$

$$= \left(\frac{1}{MN^2} \sum_{\alpha k} (x_{\alpha k} - \mu_M)^2 \right.$$

$$\left. + \underbrace{\frac{2}{MN^2} \sum_{\alpha=1}^{M} \sum_{k<\ell}^{N} (x_{\alpha k} - \mu_M)(x_{\alpha \ell} - \mu_M)}_{\text{cov}} \right)$$

sample variance of all MN experiments

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

$$\boxed{\sigma_M^2 = \frac{\sigma^2}{N} + \boxed{\text{cov}}}$$

use resampling techniques