

Statistical properties of time series (continued)

Fundamental question: how to get reliable results from simulations?

Result: estimate + error bars

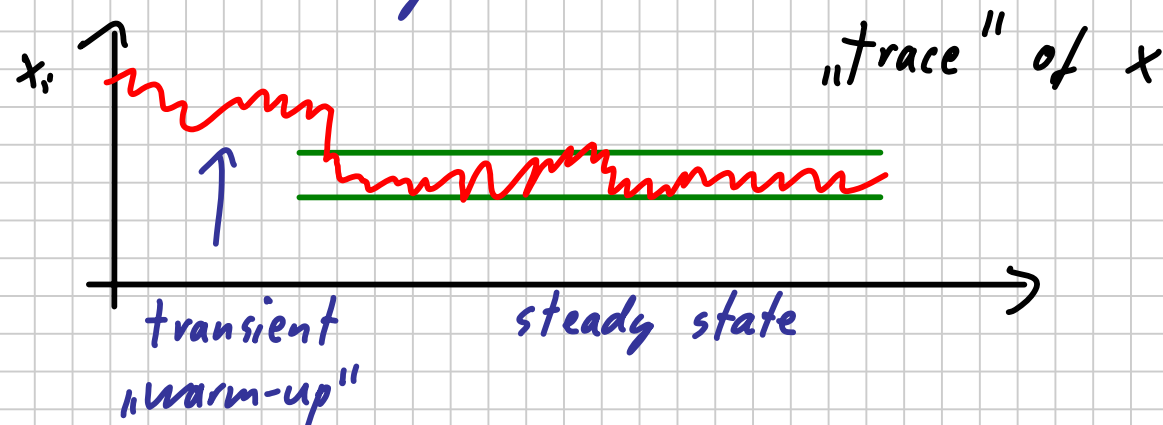
↑ decide whether results are correct or incorrect

- essential for extrapolations (e.g. finite size)

Typical case in computer simulations:

simulations start with non-representative configurations and gradually approach steady state (e.g. thermal equilibrium)

→ both mean and variance of initial measurements can differ strongly from later ones and should be thrown away



Let us now assume that transient parts have been discarded. How to compute best estimate + error bars?

Mathematical excursion

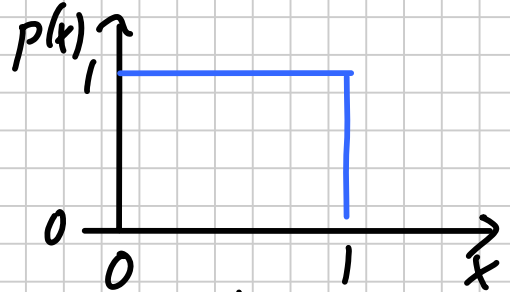
let X be a random variable with probability distribution p :

$$P(x \leq X \leq x+dx) = p(x) dx \quad (p(x) \equiv P_x(x))$$

↑ probability of the statement in brackets to be fulfilled

example: uniform distribution

$$p(x) = \begin{cases} 1 & \text{for } 0 \leq x \leq 1 \\ 0 & \text{else} \end{cases}$$



probability distribution is normalized: $\int_{-\infty}^{\infty} p(x) dx = 1$

mean (1st moment): $\langle X \rangle = \int_{-\infty}^{\infty} X p(x) dx$

2nd moment: $\langle X^2 \rangle = \int_{-\infty}^{\infty} X^2 p(x) dx$

variance: $\sigma_x^2 = \langle (X - \langle X \rangle)^2 \rangle = \langle X^2 \rangle - \langle X \rangle^2$
 $\sigma \equiv$ width of distribution

for uniform dist:

$$\begin{aligned} \langle X \rangle &= \frac{1}{2} \\ \langle X^2 \rangle &= \frac{1}{3} \\ \sigma_x^2 &= \frac{1}{12} \\ \sigma &= \frac{1}{\sqrt{12}} \approx 0.29 \end{aligned}$$

The mean is always additive; the variance only for uncorrelated random variables:

$$\begin{aligned} \langle X+Y \rangle &= \int dx \int dy p(x,y) (x+y) \\ &= \int dx x \int dy p(x,y) + \int dy y \int dx p(x,y) \\ &= \langle X \rangle + \langle Y \rangle \end{aligned}$$

$$\begin{aligned} \sigma_{X+Y}^2 &= \langle [(X+Y) - \langle X+Y \rangle]^2 \rangle = \langle [(X - \langle X \rangle) + (Y - \langle Y \rangle)]^2 \rangle \\ &= \sigma_x^2 + \sigma_y^2 + \underbrace{2 \langle (X - \langle X \rangle)(Y - \langle Y \rangle) \rangle}_{= 0 \text{ for uncorrelated variables}} \end{aligned}$$

Specifically for $\sigma_x = \sigma_y = \sigma$ in uncorrelated case:

$$\sigma\left[\frac{x+y}{2}\right] = \frac{\sigma}{\sqrt{2}}$$

(and in fully correlated case $X=Y$: $\sigma\left[\frac{x+y}{2}\right] = \sigma$)

and for arithmetic average of N uncorrelated random numbers with equal variance σ :

$$\sigma_{\bar{x}} = \sigma\left[\frac{1}{N} \sum_{i=1}^N x_i\right] = \frac{\sigma}{\sqrt{N}}$$

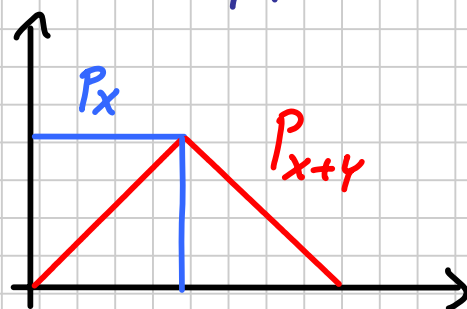
This is one formulation of the **central limit theorem**

Full probability distribution of sum/average?

$$\begin{aligned} p_{x+y}(z) &= \int dx p_x(x) \int dy p_y(y) \delta(z - (x+y)) \\ &= \int dx p_x(x) p_y(z-x) \end{aligned}$$

specifically for uniformly distributed X, Y :

$$p_{x+y}(z) \begin{cases} z & \text{for } 0 \leq z \leq 1 \\ 2-z & \text{" } 1 \leq z \leq 2 \\ 0 & \text{else} \end{cases}$$



Look at higher order cumulants \rightarrow normal distribution

$$p_{\bar{x}}(x) \xrightarrow{N \rightarrow \infty} \frac{1}{\sigma_{\bar{x}} \sqrt{2\pi}} \exp\left[-\frac{1}{2} \left(\frac{\bar{x} - \langle x \rangle}{\sigma_{\bar{x}}}\right)^2\right]; \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}}$$

central limit theorem (uncorrelated case!)

Now: application to simulation / measurements

again: random variables X_i with same distribution,
non necessarily uncorrelated

Problem: probability distribution, mean, variance
unknown. Best unbiased estimates?

easy: $\langle X \rangle \approx \bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$

\bar{X} is (best) unbiased estimator for $\langle X \rangle$ since

$$\langle \bar{X} \rangle = \frac{1}{N} \sum_{i=1}^N \underbrace{\langle X_i \rangle}_{=\langle X \rangle} = \langle X \rangle$$

(also unbiased: $\tilde{X} = \sum_{i=1}^N a_i X_i$ with $\sum_{i=1}^N a_i = 1$)
(also estimator: $\tilde{\tilde{X}} = \frac{1}{N-1} \sum_{i=1}^N X_i$ (since $\langle \tilde{\tilde{X}} \rangle = \frac{N}{N-1} \langle X \rangle \xrightarrow{N \rightarrow \infty} \langle X \rangle$)

difficult: best estimator for error of \bar{X} ?

(i) unbiased estimator for σ_X ?

$$\begin{aligned}
\langle \sum_{i=1}^N (x_i - \bar{x})^2 \rangle &= N \langle (x_i - \bar{x})^2 \rangle \\
&= N \langle (x_i - \frac{1}{N} \sum_{j=1}^N x_j)^2 \rangle \\
&= N \langle [(x_i - \langle x \rangle) - \frac{1}{N} \sum_{j=1}^N (x_j - \langle x \rangle)]^2 \rangle \\
&= N \langle [(1 - \frac{1}{N})(x_i - \langle x \rangle) - \frac{1}{N} \sum_{\substack{j=1 \\ j \neq i}}^N (x_j - \langle x \rangle)]^2 \rangle \\
&\stackrel{\text{no corr.}}{=} N \left[(1 - \frac{1}{N})^2 \langle (x_i - \langle x \rangle)^2 \rangle + \frac{1}{N^2} \sum_{\substack{j=1 \\ j \neq i}}^N \langle (x_j - \langle x \rangle)^2 \rangle \right] \\
&= \frac{(N-1)^2}{N} \sigma^2 + \frac{N-1}{N} \sigma^2 \\
&= \sigma^2 \frac{[(N-1)+1](N-1)}{N} = (N-1) \sigma^2
\end{aligned}$$

Thus, $\sigma_{\text{est}}^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2$ is an unbiased estimator for the variance σ^2 if the data is uncorrelated.

Homework: Is this estimator also unbiased for data with finite autocorrelation?