Course 07/1

Notions of probability. Central limit theorem

- Some simple definitions in probability theory
- Distribution of averages.
- Central limit theorem
- Central limit theorem for generating a Gaussian distribution

Some definitions in probability theory

Random number	Unpredictable numerical quantity, result from an experiment (observable), which can vary discretely or continuously.
Event	Obtaining an observable.
Frequency	$f\,(x) = n_x / n$, where n_x is the number of events with observable x and n is the total number of events.
Distribution of frequencies	$f\,(x)$ vs x.

Probability
(discrete variable)

$$P\,(x) = \lim_{n \to \infty} \frac{n_x}{n}$$

Properties: i. $0 \leq P\,(x) \leq 1$

ii. $\sum_x P\,(x) = 1$

Probability
(continuous variable)

$$\int_{x_{\mathrm{inf}}}^{x_{\mathrm{sup}}} p\,(x)\,dx = 1$$

Some definitions in probability theory

Mean value

$$\langle x \rangle = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{1}{n} \sum_{x} x\, n_x = \sum_{x} x \frac{n_x}{n}$$

$$\underset{\lim n \to \infty}{=} \sum_{x} x\, P(x)$$

$$\underset{\text{continuous}}{=} \int_{-\infty}^{\infty} x \; p(x)\, dx$$

Some definitions in probability theory

Variance σ^2, Var

Standard deviation σ

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

$$= \frac{1}{n} \sum_{i=1}^{n} (x_i - \langle x \rangle)^2$$

$$= \frac{1}{n} \sum_{i=1}^{n} (x_i^{\,2} - 2x_i \langle x \rangle + \langle x \rangle^2)$$

$$= \frac{1}{n} \sum_{i=1}^{n} x_i^{\,2} - \langle x \rangle^2$$

$$\boxed{\sigma^2 = \langle x^2 \rangle - \langle x \rangle^2}$$

Distribution of averages

Think at people ($j = 1,\ldots N$) measuring in sequence the size of nails x_i ($i = 1,\ldots n$).

$\ldots, x_i^{(1)}, x_{i+1}^{(1)}, \ldots$ $\qquad$ $i = 1, \ldots, n$

$\ldots, x_i^{(2)}, \ldots$

$\ldots, x_i^{(j)}, \ldots$

$\vdots$

$\ldots, x_i^{(N)}, \ldots$

The series are fully independent !

Definition of "average" variable A_i

$$A_i = \frac{1}{N} \sum_{j=1}^{N} x_i^{(j)}$$

Properties of distribution of averages

Mean value

$$\langle A \rangle = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{N}\sum_{j=1}^{N} x_i^{(j)} = \frac{1}{N}\sum_{j=1}^{N} \langle x \rangle = \frac{1}{N} N \langle x \rangle = \langle x \rangle$$

Variance

$$\mathrm{Var}\, A = \langle (A - \langle A \rangle)^2 \rangle = \langle \left(\frac{1}{N}\sum_{j=1}^{N} x^{(j)} - \langle x \rangle \right)^2 \rangle = \frac{1}{N^2} \langle \left(\sum_{j=1}^{N} x^{(j)} - N \langle x \rangle \right)^2 \rangle$$

$$= \frac{1}{N^2} \langle \left(\sum_{j=1}^{N} (x^{(j)} - \langle x \rangle) \right)^2 \rangle$$

$$= \frac{1}{N^2} \langle \left(\sum_{j=1}^{N} (x^{(j)} - \langle x \rangle) \right) \left(\sum_{j'=1}^{N} (x^{(j')} - \langle x \rangle) \right) \rangle$$

Properties of distribution of averages

Variance

$$\mathrm{Var}\, A \;=\; \frac{1}{N^2} \left\langle \left(\sum_{j=1}^{N} (x^{(j)} - \langle x \rangle) \right) \left(\sum_{j'=1}^{N} (x^{(j')} - \langle x \rangle) \right) \right\rangle$$

terms with $j \neq j'$ are independent:

$\langle \Delta x^{(j)} \Delta x^{(j')} \rangle = \langle \Delta x^{(j)} \rangle \cdot \langle \Delta x^{(j')} \rangle = 0$

Hence, only terms with $j = j'$ remain.

$$\mathrm{Var}\, A \;=\; \frac{1}{N^2} \left\langle \sum_{j=1}^{N} (x^{(j)} - \langle x \rangle)^2 \right\rangle \;=\; \frac{1}{n} \sum_{i=1}^{n} \frac{1}{N^2} \sum_{j=1}^{N} (x_i^{(j)} - \langle x \rangle)^2$$

All $x^{(j)}$ are independent: mean of sums → sum of means

$$\mathrm{Var}\, A \;=\; \frac{1}{N^2} \sum_{j=1}^{N} \langle (x^{(j)} - \langle x \rangle)^2 \rangle \;=\; \frac{1}{N^2} N \sigma_x^2 \;=\; \frac{\sigma_x^2}{N}$$

Properties of distribution of averages : Summary

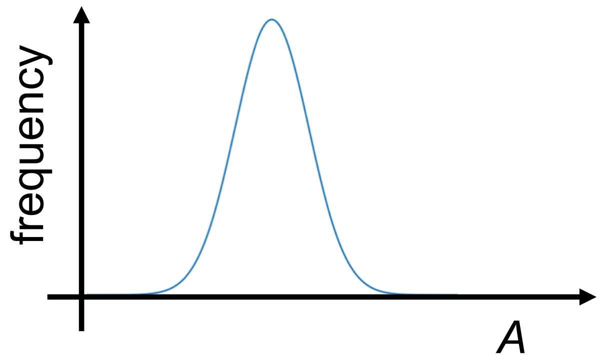
Mean value

$$\mu_A = \langle A \rangle = \langle x \rangle = \mu_x$$

Variance

$$\sigma^2{}_A = \frac{\sigma^2{}_x}{N}$$

What about the full distribution ?



Central limit theorem

- Important theorem in statistics

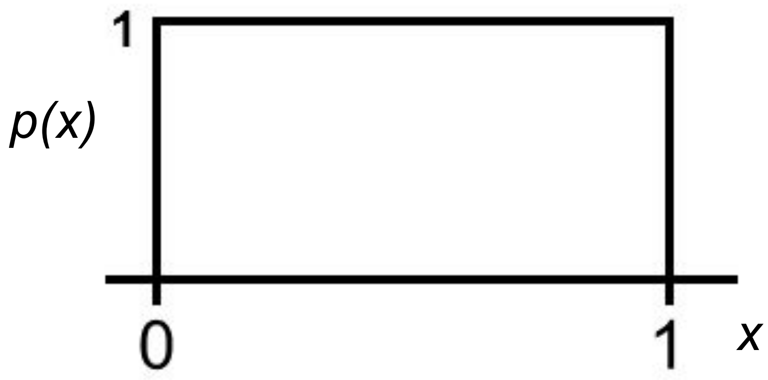
- In the limit $N \to \infty$, a random number that corresponds to the sum of N random numbers is distributed according to a normal distribution (Gaussian), provided the variances of all random numbers in the sum are:

  i. finite
  ii. of the same order

(there are some exceptions)

Central limit theorem for generating a Gaussian distribution

Computer-generated uniform distribution *p(x)* of random numbers *x* between 0 and 1.



Mean value

$$\mu_x = \int_0^1 x\, dx = \frac{1}{2} x^2 \Big|_0^1 = \frac{1}{2}$$

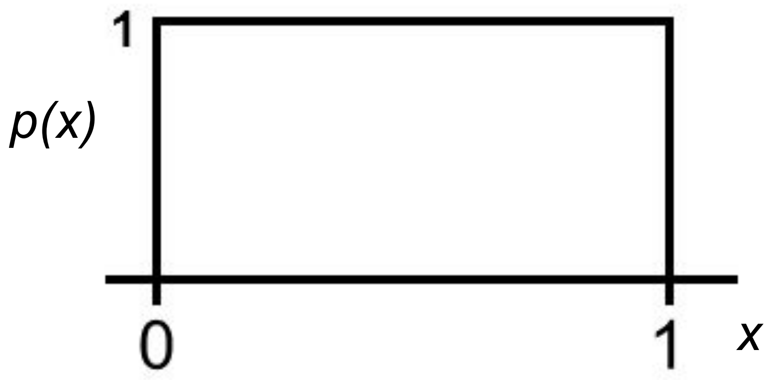
Variance

$$\sigma_x^2 = \int_0^1 x^2\, dx - \frac{1}{4} = \frac{1}{3} x^3 \Big|_0^1 - \frac{1}{4} = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$$

Central limit theorem
for generating a Gaussian distribution

New variable: average over N values

$$A = \frac{1}{N} \sum_{j=1}^{N} x^{(j)}$$



Mean value and variance

$$\mu_A = \langle A \rangle = \frac{1}{N} (N \langle x \rangle) = \frac{1}{2}$$

$$\sigma_A^{\,2} = \frac{\sigma_x^{\,2}}{N} = \frac{1}{12N}$$

Approximate Gaussian variable

$$A' = \frac{1}{\sigma_A} (A - \langle A \rangle)$$

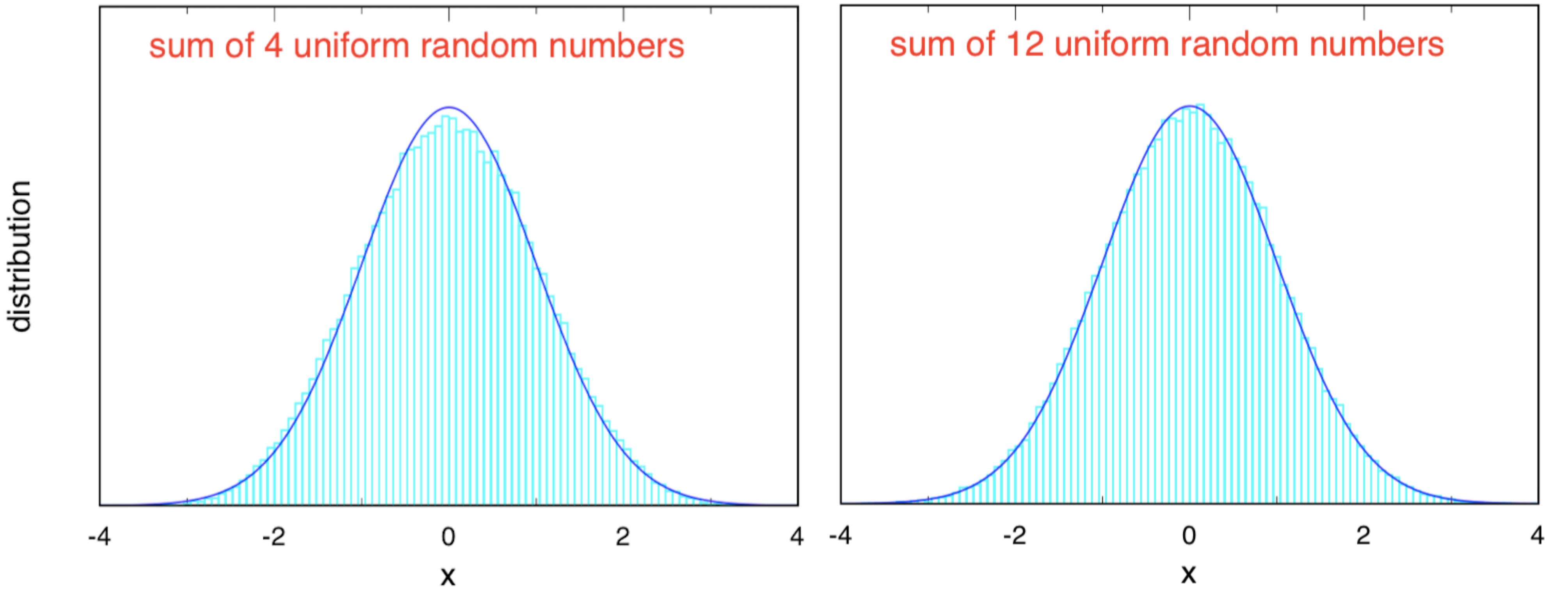
Gaussian with $\mu' = 0$ and $\sigma' = 1$

$$A'' = \mu_{\mathrm{D}} + \sigma_{\mathrm{D}} \left[\frac{1}{\sigma_A} (A - \langle A \rangle) \right]$$

Gaussian with desired $\mu'' = \mu_{\mathrm{D}}$ and $\sigma'' = \sigma_{\mathrm{D}}$

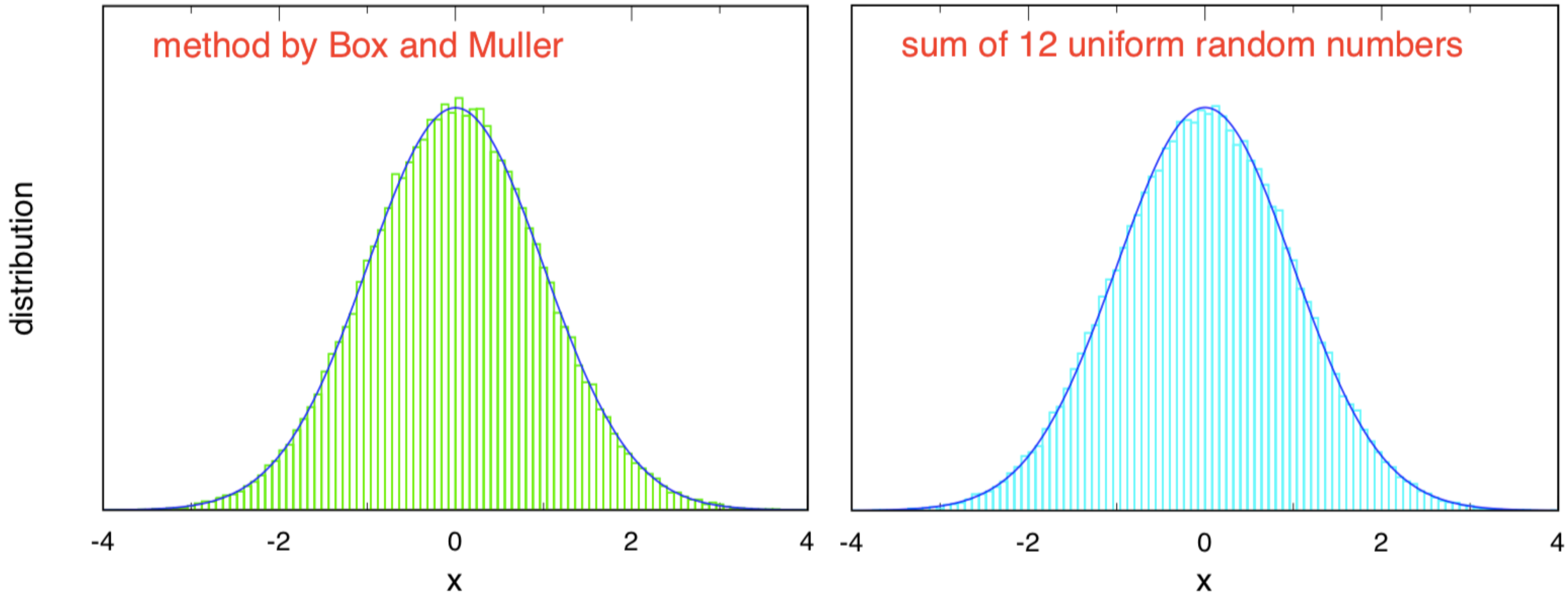
Central limit theorem for generating a Gaussian distribution

10’000 variates



Central limit theorem
for generating a Gaussian distribution

10’000 variates



Course 07/1

Notions of probability. Central limit theorem

- Some simple definitions in probability theory
- Distribution of averages.
- Central limit theorem
- Central limit theorem for generating a Gaussian distribution