Monte Carlo Methods

Why "Monte Carlo"?

Computation of number π

$\pi/4 = \Sigma_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} \ldots$

$\pi = 37/15 = 2.466666...$

Buffon's Problem

Probability $p = \frac{2}{\pi} \longmapsto \pi = \frac{2}{p}$

Throw needle N times. N_A hits. $\hat{p} = \frac{N_A}{N} \qquad p \approx \hat{p}$

Probabilistic method to "solve" a deterministic problem.

Generalizes the "simulation" idea.

Brief summary of probability results

Set E of results of an "experiment".

Random variable: to assign a number to each result

$$\begin{array}{rccc} \hat{\mathbf{x}}: & E & \longmapsto & R \\ & \xi & \rightarrow & \hat{\mathbf{x}}(\xi) \end{array}$$

We assign probabilities to each possible value of the r.v.

$\hat{\mathbf{x}} \in \{x_1, x_2, ...\}$

Value x_i with probability $p_i = P(\hat{\mathbf{x}} = x_i)$:

$$p_i \geq 0, \qquad \sum_i p_i = 1$$

$\hat{\mathbf{x}}$ continuum variable:

$$P(\hat{\mathbf{x}} \in [a,b]) = \int_a^b f_{\hat{\mathbf{x}}}(x)\, dx$$

$f_{\hat{\mathbf{x}}}(x)$: probability density function of the random variable $\hat{\mathbf{x}}$.

$$f_{\hat{\mathbf{x}}}(x) \geq 0$$

$$\int_{-\infty}^{\infty} f_{\hat{\mathbf{x}}}(x)\, dx = 1$$

Probability distribution function $F_{\hat{\mathbf{x}}}(x)$:

$$F_{\hat{\mathbf{x}}}(x) = \int_{-\infty}^{x} f_{\hat{\mathbf{x}}}(x)\, dx$$

Interpretation:

$$P(x \leq \hat{\mathbf{x}} \leq x + dx) = f_{\hat{\mathbf{x}}}(x)\, dx$$

$$P(\hat{\mathbf{x}} \in \Omega) = \int_\Omega f_{\hat{\mathbf{x}}}(x)\, dx$$

$$P(\hat{\mathbf{x}} \leq x) = F_{\hat{\mathbf{x}}}(x) \qquad P(x_1 \leq \hat{\mathbf{x}} \leq x_2) = F_{\hat{\mathbf{x}}}(x_2) - F_{\hat{\mathbf{x}}}(x_1)$$

Discrete case: sum of Dirac-delta functions:

$$f_{\hat{\mathbf{x}}}(x) = \sum_{\forall i} p_i \delta(x - x_i)$$

Average value of $g(x)$:

$$\langle g\rangle_{\hat{\mathrm{x}}} = E_{\hat{\mathrm{x}}}[g] = \int_{-\infty}^{\infty} f_{\hat{\mathrm{x}}}(x)g(x)\,dx$$

$$\rangle g\rangle = \sum_{\forall i} p_i g(x_i)$$

n-order moments: $\langle \hat{\mathbf{x}}^n\rangle$

Mean: $\mu = \langle \hat{\mathbf{x}}\rangle$

Variance: $\sigma^2[\hat{\mathbf{x}}] = \langle (\hat{\mathbf{x}} - \mu)^2\rangle = \langle \hat{\mathbf{x}}^2\rangle - \langle \hat{\mathbf{x}}\rangle^2$.

$\sigma[\hat{\mathbf{x}}]$:root mean square (rms) of the r.v. $\hat{\mathbf{x}}$.

Bernouilli Distribution: A , $\bar{A}$.

$\hat{\mathbf{x}} = 1$ if results A, $\hat{\mathbf{x}} = 0$ if results $\bar{A}$

$$\begin{aligned} p(\hat{\mathbf{x}} = 0) &= p \\ p(\hat{\mathbf{x}} = 1) &= 1 - p \equiv q \\ \langle \hat{\mathbf{x}}\rangle &= p \\ \sigma^2[\hat{\mathbf{x}}] &= p(1-p) \end{aligned}$$

Binomial Distribution :

Repetition of binary experiment N times

$$\begin{aligned}
\hat{\mathbf{x}}_i &= 1, 0 \\
\hat{\mathbf{N}}_A &= \Sigma_{i=1}^{N} \hat{\mathbf{x}}_i \\
\hat{p} &= \tfrac{\hat{\mathbf{N}}_A}{N} \\
p(\hat{\mathbf{N}}_A = n) &= \tbinom{N}{n} p^n (1-p)^{N-n} \\
\langle \hat{\mathbf{N}}_A \rangle &= Np \\
\sigma^2[\hat{\mathbf{N}}_A] &= Np(1-p) \\
\langle \hat{\mathbf{p}} \rangle &= p \\
\sigma^2[\hat{\mathbf{p}}] &= \tfrac{p(1-p)}{N}
\end{aligned}$$

Geometric Distribution: To repeat binary experiment until one gets A

$\hat{\mathbf{x}}$: number of trials.

$$p(\hat{\mathbf{x}} = n) = (1-p)^{n-1} p \qquad\qquad n = 1, 2, \ldots, \infty$$

$$\begin{aligned}
\langle \hat{\mathbf{x}} \rangle &= \tfrac{1}{p} \\
\sigma^2[\hat{\mathbf{x}}] &= \tfrac{1}{p^2} - \tfrac{1}{p}
\end{aligned}$$

Poisson Distribution:

Independent repetition with frequency λ

$\hat{\mathbf{x}}$: number of occurrences per unit time

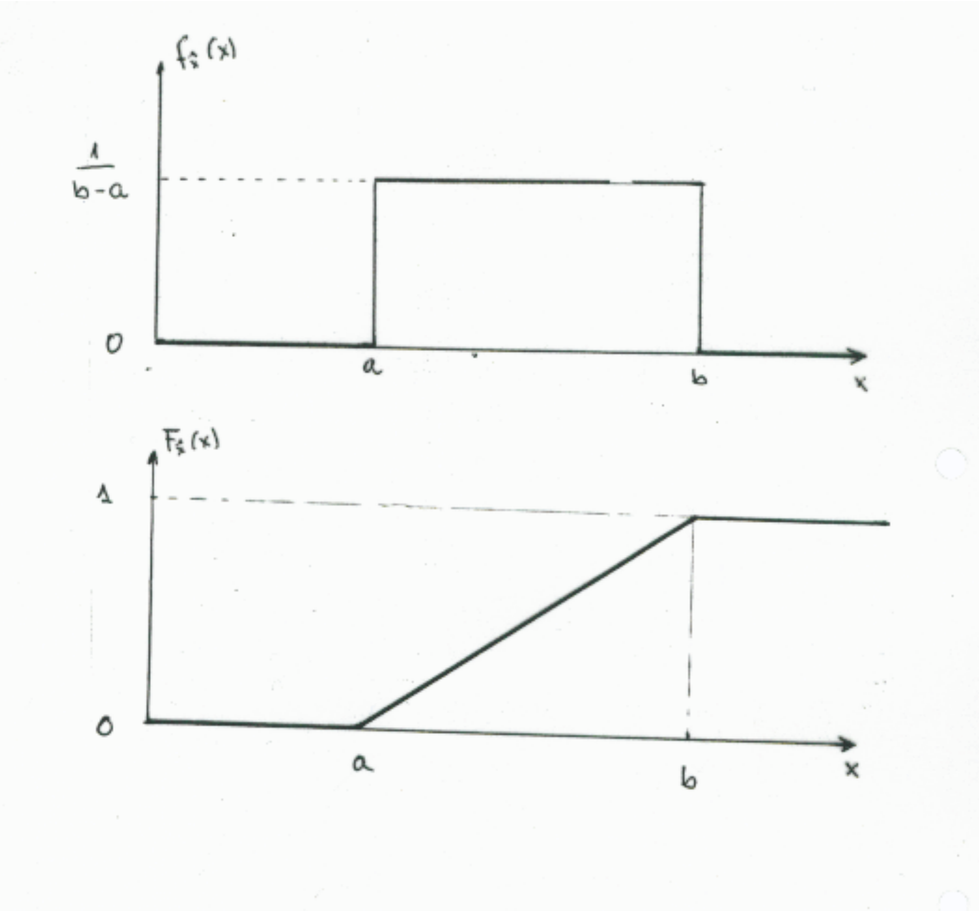
$$p(\hat{\mathbf{x}} = n) = \frac{\lambda^n}{n!} e^{-\lambda}$$
$$\langle \hat{\mathbf{x}} \rangle = \lambda$$
$$\sigma^2[\hat{\mathbf{x}}] = \lambda$$

Uniform distribution: continuum r.v. :$\hat{\mathbf{x}}$, $\hat{U}(a,b)$

$$f_{\hat{\mathbf{x}}}(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a,b] \\ 0 & \text{otherwise} \end{cases}$$

$$\langle \hat{\mathbf{x}} \rangle = \frac{a+b}{2}$$

$$\sigma^2[\hat{\mathbf{x}}] = \frac{(b-a)^2}{12}$$



Theorem: $\hat{\mathbf{x}}$, $F_{\hat{\mathbf{x}}}(x)$, let $\hat{\mathbf{y}} = F_{\hat{\mathbf{x}}}(\hat{\mathbf{x}}) \longmapsto \hat{\mathbf{y}}$ is $\hat{U}(0,1)$.

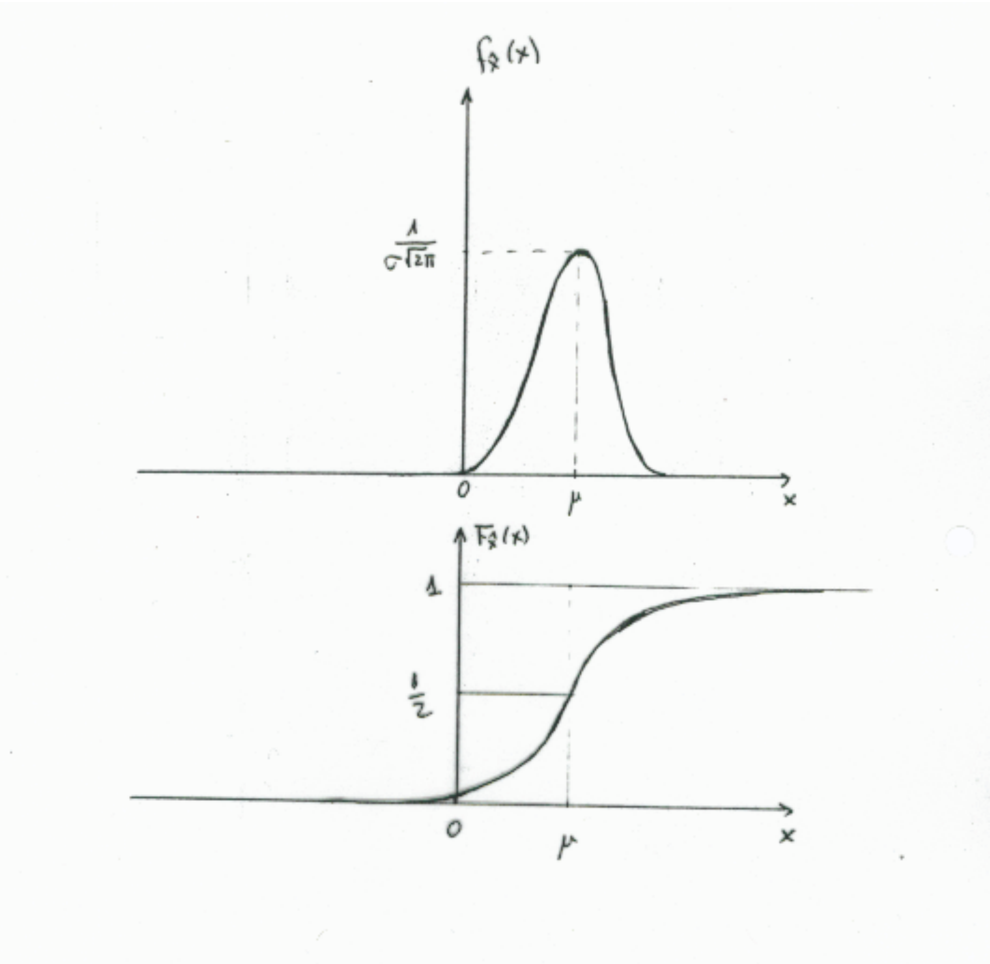
Gaussian Distribution: $\hat{G}(\mu,\sigma)$

$$f_{\hat{\mathbf{x}}}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

$$F_{\hat{\mathbf{x}}}(x) = \frac{1}{2} + \frac{1}{2}\text{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)$$

where erf(z) is the error function defined as:

$$\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-y^2}\, dy$$



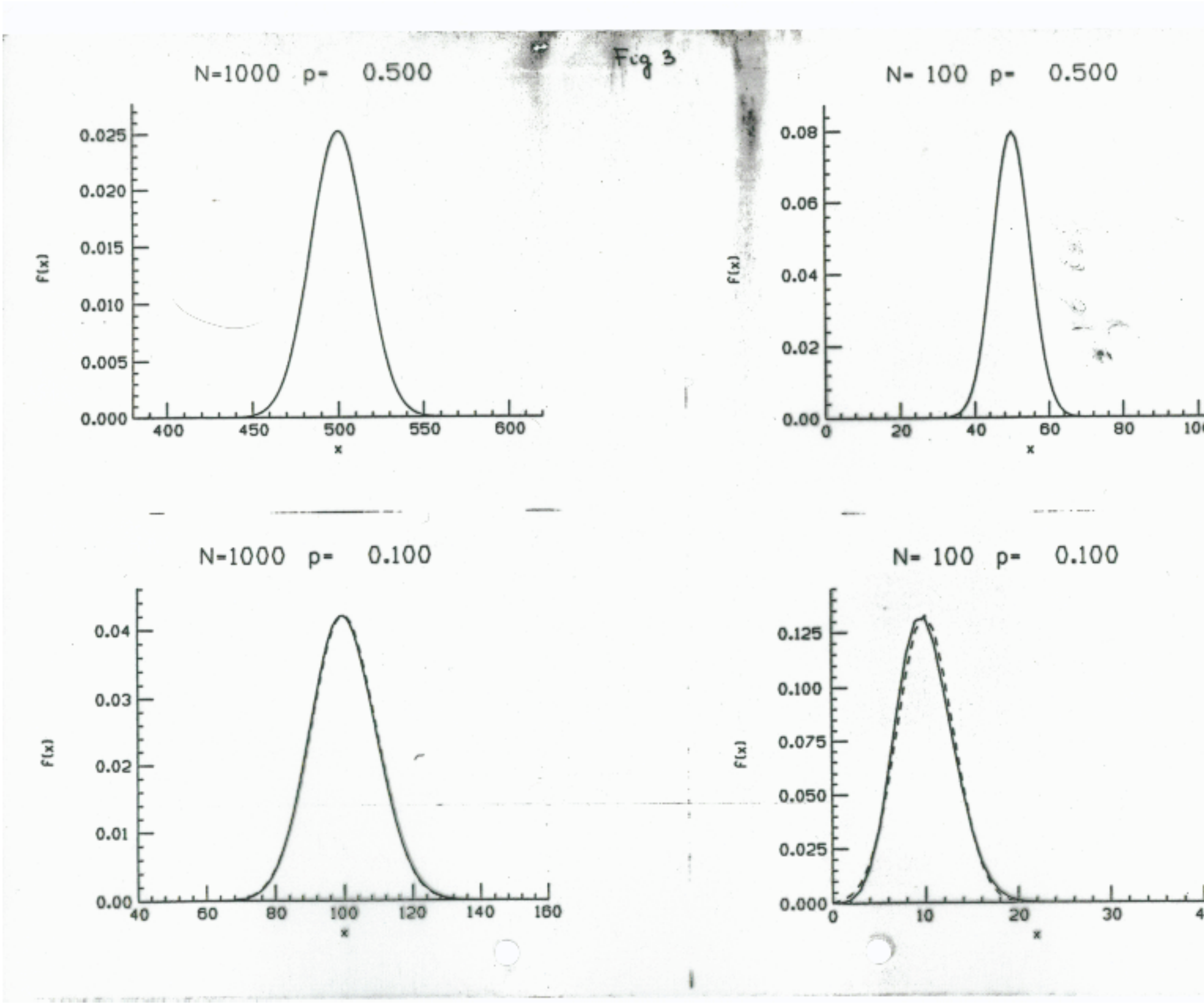
Moivre–Laplace Theorem:

$$p(\hat{\mathbf{N}}_A = n) = \binom{N}{n} p^n (1-p)^{N-n}$$

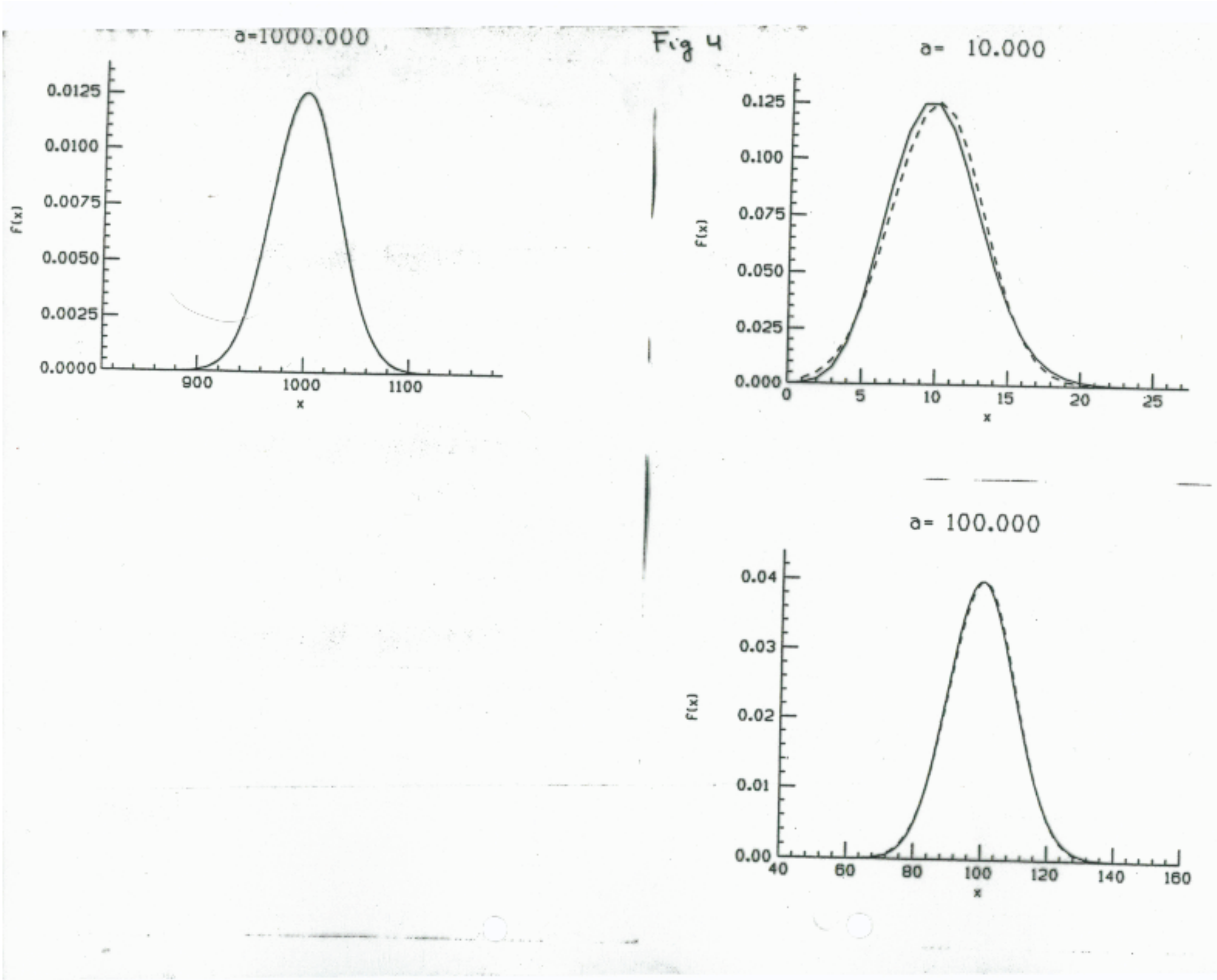
$$\approx \frac{\exp\left[-(n-Np)^2/2Np(1-p)\right]}{\sqrt{2\pi Np(1-p)}}$$

if $N \to \infty$ with $\mid n - Np \mid / \sqrt{Np(1-p)}$ finite.

In practice, $N \geq 100$ if $p = 0.5$, $N \geq 1000$ si $p = 0.1$.



Poisson $\lambda \to \hat{G}(\lambda, \sqrt{\lambda})$, if $\lambda \to \infty$ ($\lambda \geq 100$).



Sequence of Random Variables

Joint density probability function $f_{\hat{\mathbf{x}}_1,\ldots,\hat{\mathbf{x}}_N}(x_1, ..., x_N)$

$$P((\hat{\mathbf{x}}_1, ..., \hat{\mathbf{x}}_N) \in \Omega) = \int_\Omega dx_1 .... dx_N f_{\hat{\mathbf{x}}_1,\ldots,\hat{\mathbf{x}}_N}(x_1, ..., x_N)$$

$\hat{\mathbf{x}}_1, ...\hat{\mathbf{x}}_N$ independent r.v. :

$$f_{\hat{\mathbf{x}}_1,\ldots,\hat{\mathbf{x}}_N}(x_1, ..., x_N) = f_{\hat{\mathbf{x}}_1}(x_1) \ldots f_{\hat{\mathbf{x}}_N}(x_N)$$

$$\langle g(x_1, ..., x_N) \rangle =$$

$$\int_{-\infty}^{\infty} dx_1 \ldots \int_{-\infty}^{\infty} dx_N g(x_1, ...x_N) f_{\hat{\mathbf{x}}_1,...,\hat{\mathbf{x}}_N}(x_1, ..., x_N)$$

If $\hat{\mathbf{x}}_1, ...\hat{\mathbf{x}}_N$ are independent:

$$\langle \lambda_1 g_1(x_1) + ... + \lambda_N g_N(x_N) \rangle = \lambda_1 \langle g_1(x_1) \rangle + ...\lambda_N \langle g_N(x_N) \rangle$$

$$\sigma^2[\lambda_1 g_1 + ... + \lambda_N g_N] = \lambda_1^2 \sigma^2[g_1] + ... + \lambda_N^2 \sigma^2[g_N]$$

Cros-correlation (covariance) between $\hat{\mathbf{x}}_i,\ \hat{\mathbf{x}}_j$:

$$C[\hat{\mathbf{x}}_i, \hat{\mathbf{x}}_j] \equiv C_{ij} \equiv \langle (\hat{\mathbf{x}}_i - \mu_i)(\hat{\mathbf{x}}_j - \mu_j) \rangle$$

If $\hat{\mathbf{x}}_i,\ \hat{\mathbf{x}}_j$ are independent:

$$C_{ij} = \sigma^2[\hat{\mathbf{x}}_i] \delta_{ij}$$

Varianza of sum of two functions $g_1(x),\ g_2(x)$:

$$\sigma^2[g_1 + g_2] = \langle (g_1 + g_2)^2 \rangle - \langle g_1 + g_2 \rangle^2$$

expanding and reordering:

$$\sigma^2[g_1 + g_2] = \sigma^2[g_1] + \sigma^2[g_2] + 2C[g_1, g_2]$$

Correlation coefficient $\rho[\hat{\mathbf{x}}_i, \hat{\mathbf{x}}_j]$, between $\hat{\mathbf{x}}_i,\ \hat{\mathbf{x}}_j$:

$$\rho[\hat{\mathbf{x}}_i, \hat{\mathbf{x}}_j] = \frac{C[\hat{\mathbf{x}}_i, \hat{\mathbf{x}}_j]}{\sigma[\hat{\mathbf{x}}_i] \sigma[x_j]}$$

$$|\rho[\hat{\mathbf{x}}_i, \hat{\mathbf{x}}_j]| \leq 1$$

Marginal probability density functions:

$$f_{\hat{\mathbf{x}}_1}(x_1) = \int_{-\infty}^{\infty} f_{\hat{\mathbf{x}}_1\hat{\mathbf{x}}_2}(x_1, x_2)\, dx_2$$

$$f_{\hat{\mathbf{x}}_2\hat{\mathbf{x}}_4} = \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_3 \; f_{\hat{\mathbf{x}}_1\hat{\mathbf{x}}_2\hat{\mathbf{x}}_3\hat{\mathbf{x}}_4}(x_1, x_2, x_3, x_4)$$

Joint Gaussian random variables

$$f(x_1, ..., x_N) = \sqrt{\frac{|\, A\, |}{(2\pi)^N}} \exp\left[-\frac{1}{2} \sum_{i,j=1}^{N} (x_i - \mu_i) A_{ij} (x_j - \mu_j)\right]$$

$$\langle \hat{\mathbf{x}}_i \rangle = \mu_i$$

$$C_{ij} = (A^{-1})_{ij}$$

Interpretation of the rms. Statistical errors.

$$P(|\, \hat{\mathbf{x}}(\xi) - \mu \,| \leq k\, \sigma) \geq 1 - \frac{1}{k^2}$$

Gaussian random variable:

$$P(|\, \hat{\mathbf{x}}(\xi) - \mu \,| \leq k\, \sigma) = \mathrm{erf}(\frac{k}{\sqrt{2}})$$

which takes the following values:

$$P(|\, \hat{\mathbf{x}}(\xi) - \mu \,| \leq \sigma) = 0.68269...$$

$$P(|\hat{\mathbf{x}}(\xi) - \mu| \leq 2\sigma) = 0.95450...$$

$$P(|\hat{\mathbf{x}}(\xi) - \mu| \leq 3\sigma) = 0.99736...$$

$$\mu = \hat{\mathbf{x}}(\xi) \pm \sigma$$

Buffon'a problem: we measure $\hat{\mathbf{p}} = \hat{\mathbf{N}}_A/N$

Follows the binomial distribution

$$\langle\hat{\mathbf{p}}\rangle = p$$
$$\sigma^2[\hat{\mathbf{p}}] = \frac{p(1-p)}{N}$$

$$p = \hat{\mathbf{p}} \pm \sqrt{\frac{p(1-p)}{N}} \approx \hat{\mathbf{p}} \pm \sqrt{\frac{\hat{\mathbf{p}}(1-\hat{\mathbf{p}})}{N}}$$

Error decreases as $N^{-1/2}$.

$p = 2/\pi = 0.6366$, $N = 100$, relative error $\approx 7.5\%$.

We do not know σ. Take the sample mean:

$$\hat{\mu}_N = \frac{1}{N}\sum_{i=1}^{N}\hat{\mathbf{x}}_i$$

$$\langle\hat{\mu}_N\rangle = \langle\hat{\mathbf{x}}_i\rangle = \mu$$

Sample variance:

$$\hat{\sigma}_N^2 = \frac{1}{N-1}\sum_{i=1}^{N}(\hat{\mathbf{x}}_i - \hat{\mu}_N)^2 = \frac{N}{N-1}\left(\frac{1}{N}\sum_{i=1}^{N}\hat{\mathbf{x}}_i^2 - \left(\frac{1}{N}\sum_{i=1}^{N}\hat{\mathbf{x}}_i\right)^2\right)$$

$$\langle\hat{\sigma}_N^2\rangle = \sigma^2$$

$$\sigma^2[\hat{\mu}_N] = \frac{1}{N}\sigma^2$$

$$\mu = \hat{\mu}_N(\Xi) \pm \sigma[\hat{\mu}_N] = \hat{\mu}_N(\Xi) \pm \frac{\sigma}{\sqrt{N}} \approx \hat{\mu}_N(\Xi) \pm \frac{\hat{\sigma}_N[\Xi]}{\sqrt{N}}$$

$N \to \infty$, **central limit theorem**

$$\hat{\mu}_N \to \hat{G}(\mu, \frac{\sigma}{\sqrt{N}})$$

Asymptotic result:

$$P\left(\mid \hat{\sigma}_N(\Xi) - \sigma \mid \leq \frac{k\hat{\sigma}_N(\Xi)}{\sqrt{2N}} \right) = \mathrm{erf}(\frac{k}{\sqrt{2}})$$

$$\sigma = \hat{\sigma}_N(\Xi) \pm \frac{\hat{\sigma}_N(\Xi)}{\sqrt{2N}}$$