Introduction to Computational Physics

University of Maribor,
Summer Term 2003

Franz J. Vesely
University of Vienna

www.ap.univie.ac.at/users/Franz.Vesely/

3. Stochastics

John von Neumann, playing randomly

- Statistics turned upside down
- Production of *Random Numbers* and *Random Sequences* with desired properties
- Random paths through real space (diffusion) or phase space (Monte Carlo Simulation)
- Application of MC to optimization and minimization problems

Equidistributed Random Variates:
- Linear Congruential Generators
- Shift Register Generators

Other Distributions:
- Transformation Method
- Box-Muller Method for the Normal Distribution
- Rejection Method
- Multivariate Gaussian Distribution
- Equidistribution in Orientation Space

Random Sequences:
- Markov Chains and the Monte Carlo method
- Stochastic Optimization
- Simulated Annealing
- Genetic Algorithms

Linear Congruential Generators:

$$\boxed{I_{n+1} = [\, a\, I_n + b\,] \bmod m}$$

where a is some (odd) multiplicative factor, m is the largest integer (hardware-dependent, e.g. $m = 2^{32}$), and b is relatively prime with respect to m.

To obtain random numbers x_n of type *real*, equidistributed over the interval $(0, 1)$, divide I_n by m.

$\Longrightarrow$Library or internal routines RAND, RND, RAN etc.

To minimize serial correlations:

"Erasing tracks:"

1. Produce a list $RLIST(i)$ of Z equidistributed random numbers $x_i \in (0, 1)$; $i = 1 \ldots Z$. (e.g., $Z = 97$.)
2. Sample an additional random number y in $(0, 1)$.
3. Determine a pointer index $j \in [1, Z]$ according to

$$j = 1 + \mathrm{int}(y \cdot Z)$$

   (int(r) ... largest integer smaller than the real number r.)
4. Use the element $RLIST(j)$ corresponding to j as the output random number.
5. Put $y = RLIST(j)$ and replace $RLIST(j)$ by a new random number $\in (0, 1)$; return to (3).

Shift Register Generators:

(Also "Tausworthe" or "XOR" generators) Originally for the production of *random bits*, but one may always generate 16, 32, etc. bits at a time and combine them to a computer word.

Let bits $b_1, b_2, ..b_n$ be already given; then

$$b_{n+1} \;=\; b_k \oplus b_m \oplus \ldots \oplus b_n \, ,$$

with $k < m < .. < n$, and $\oplus$ ... "exclusive or" (XOR)

To find optimal indices $(k, m, \ldots, n)$: see the theory of "primitive polynomials modulo 2".

"Exhaustive" property: Starting such a recursion with an arbitrary combination of n bits (except $0 \ldots 0$), *all possible* configurations of n bits will be realized just once before a new cycle begins.

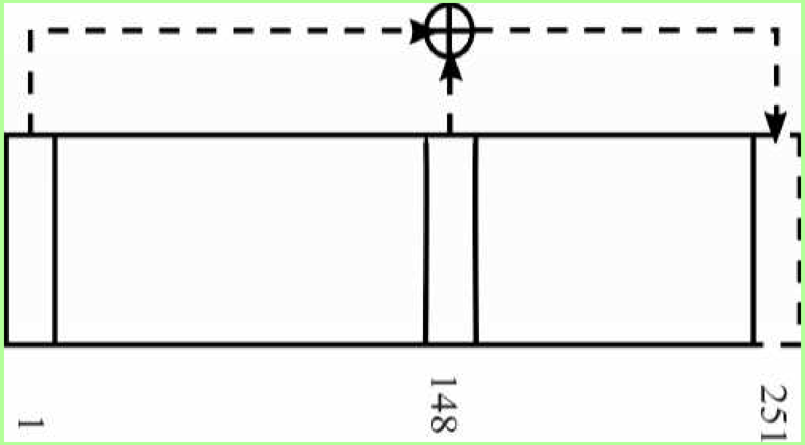
Example: $(1,3)$ is one of the optimal combinations. Starting with the sequence $\{b_1, b_2, b_3\} = \{101\}$ and applying $b_4 = b_3 \oplus b_1$ etc., we find the sequence, reading from left to right,

$$101\,001\,110\,100\,111\,010\,011\,101\,\ldots$$

It is evident that indeed all possible 3-bit groups (except 000) occur before the sequence repeats.

A very popular prescription is the "R250" algorithm of Kirkpatrick-Stoll, based on $m = 103$ and $n = 250$:

$$\boxed{I_s = I_{s-103} \oplus I_{s-250}}$$



For the first 250 random integers, use a linear congruential generator.

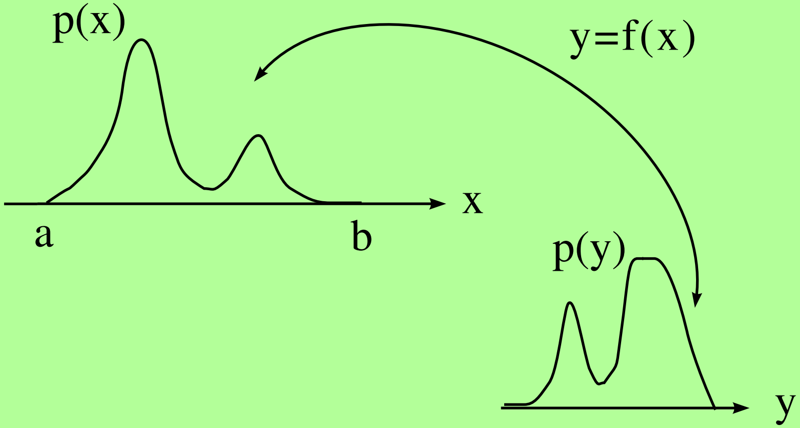
Other Distributions:

Transformation of probability densities:
Given $p(x)$ and a bijective mapping $y = f(x)$; $x = f^{-1}(y)$; then

$$|p(y)\,dy| \quad = \quad |p(x)\,dx|$$

or

$$p(y) \quad = \quad p(x)\left|\frac{dx}{dy}\right| = p[f^{-1}(y)]\left|\frac{df^{-1}(y)}{dy}\right|$$



This relation holds for *any* kind of density.

Example: The spectral density of black body radiation is usually written in terms of the angular frequency ω:

$$I(\omega) = \frac{\hbar\omega^3}{\pi c^3}\frac{1}{e^{\hbar\omega/kT}-1}$$

If we prefer to give the spectral density in terms of the wave length $\lambda \equiv 2\pi c/\omega$, we have

$$I(\lambda) = I[\omega(\lambda)]\left|\frac{d\omega}{d\lambda}\right| = \frac{\hbar}{\pi c^3}\left(\frac{2\pi c}{\lambda}\right)^3\frac{1}{e^{(hc/\lambda)/kT}-1}\left(\frac{2\pi c}{\lambda^2}\right)$$

Exercise: A powder of approximately spherical metallic grains is used for sintering. The diameters of the grains obey a normal distribution with $\langle d\rangle = 2\mu m$ and $\sigma = 0.25\mu m$. Determine the distribution of the grain volumes.

Transformation Method:

Given a probability density $p(x)$:
Find a bijective mapping $y = f(x)$ such that the distribution of y is $p(y) = c$:

$$p(x) \quad = \quad c \left|\frac{dy}{dx}\right| = c \left|\frac{df(x)}{dx}\right| \quad \text{or} \quad \left|\frac{df(x)}{dx}\right| = \frac{1}{c}\, p(x)$$

It is easy to see that

$$f(x) \quad = \quad P(x) \equiv \int_a^x p(x')dx'$$

fulfills this condition, with $c = 1$.

Transformation method:
Let $p(x)$ be a desired density, with $y = P(x) = \int p(x')dx'$. Assume that $P^{-1}(y)$ be known.

- Sample y from an equidistribution in the interval $(0, 1)$.
- Compute $x = P^{-1}(y)$.

The variable x then has the desired probability density $p(x)$.

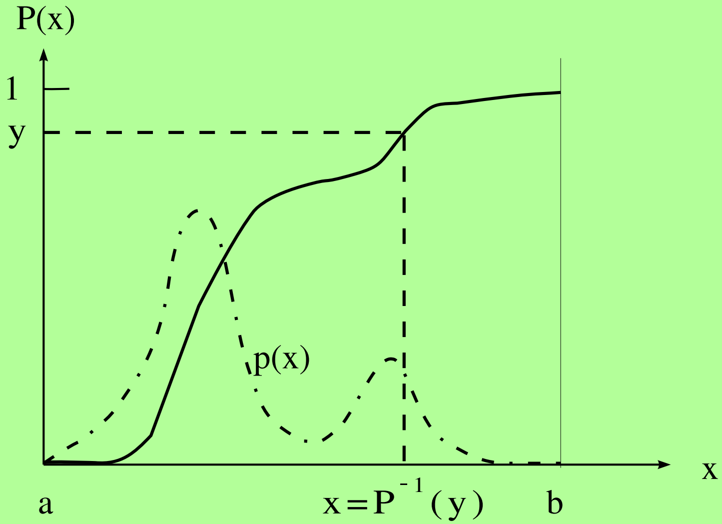
Example: Let

$$p(x) \;=\; \frac{1}{\pi}\frac{1}{1+x^2} \quad \text{(Lorentzian)}, \quad x \epsilon (\pm\infty)$$

Then $y = P(x) = 1/2 + (1/\pi)$ arctan x, with the inverse $P^{-1}(y) = \tan[\pi(y - 1/2)]$. Therefore:

- Sample y equidistributed in $(0, 1)$.
- Compute $x = \tan[\pi(y - \frac{1}{2})]$.

Geometrical interpretation:



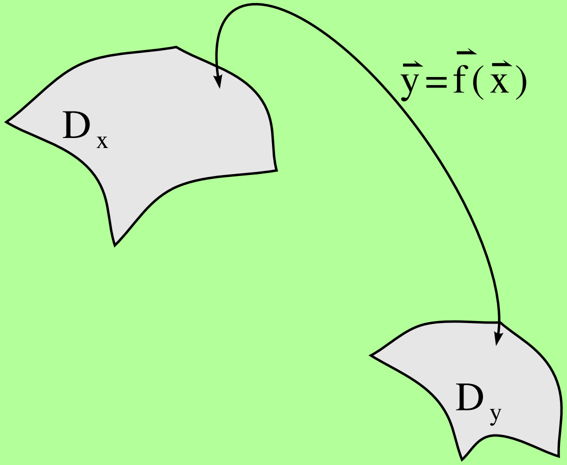
y is sampled from an equidistribution $\in (0,1)$ and $x = P^{-1}(y)$.
$\Longrightarrow$ The regions where $P(x)$ is steeper (i.e. $p(x)$ is large) are hit more frequently.

Generalized Transformation Method:

Same as before, but with *n-tuples* of random variates:
Let $\mathbf{x} = (x_1, \ldots, x_n)$, $\mathbf{x} \epsilon D_x$, and $\mathbf{y} = \mathbf{f}(\mathbf{x})$ with $\mathbf{y} \epsilon D_y$. Then

$$p(\mathbf{y}) \;\; = \;\; p(\mathbf{x}) \left| \frac{\partial \mathbf{x}}{\partial \mathbf{y}} \right|$$

($|\partial \mathbf{x} / \partial \mathbf{y}|$ ... Jacobi determinant of the transformation $\mathbf{x} = \mathbf{f}^{-1}(\mathbf{y})$.)



The following procedure for the production of Gaussian random variates may be understood as an application of this. $\Longrightarrow$

Normal distribution: *

Box-Muller technique:

- Sample $(y_1, y_2) \in (0,1)^2$
- Construct

$$\begin{aligned} x_1 &= \sqrt{-2 \ln y_1} \cos 2\pi y_2 \\ x_2 &= \sqrt{-2 \ln y_1} \sin 2\pi y_2 \end{aligned}$$

x_1, x_2 are then normal-distributed and statistically independent. Gaussian variates with given variances σ_1^2, σ_2^2 are obtained by multiplying x_1 and x_2 by their respective σ_i.

*A "quick and dirty" method to produce *almost* normal variates goes as follows: if $y = x_1 + \ldots + x_n$ is the sum of $n = 10 - 15$ equidistributed random numbers in $(-0.5, 0.5)$, then the distribution of $z \equiv y\sqrt{12/n}$ is almost normal.

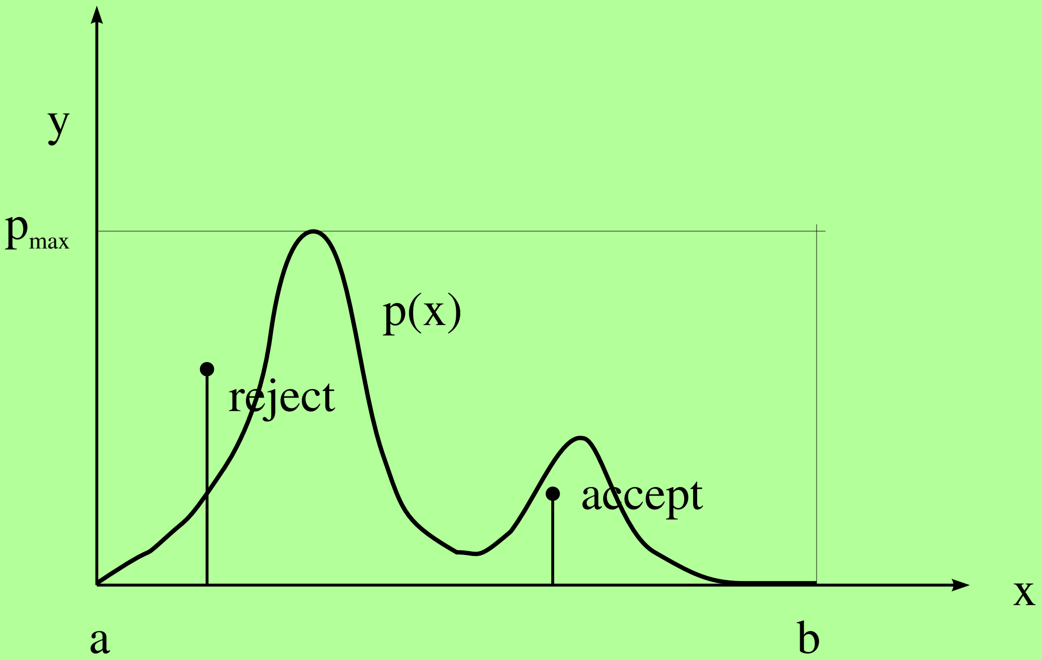
Rejection Method:

A classic: created by John von Neumann, applicable to almost any $p(x)$.

Rejection method:

Let $[a, b]$ be the allowed range of values of the variate x, and p_m the maximum of the density $p(x)$.

1. Sample a pair of equidistributed random numbers, $x \in [a, b]$ and $y \in [0, p_m]$.
2. If $y \leq p(x)$, accept x as the next random number, otherwise return to step 1.



The method is simple and fast, but it becomes inefficient whenever the area of the rectangle $[a, b] \otimes [0, p_m]$ is large compared to the area below the graph of $p(x)$. Otherwise, the "Improved Rejection Method" may be applicable: $\Longrightarrow$

Improved rejection method:

Let $f(x)$ be a test function similar to $p(x)$, with

$$f(x) \geq p(x)\,; \quad x \in [a,b]$$

$F(x) \equiv \int f(x)\,dx$ is assumed to be known and invertible

1. Pick a random number $x \in [a,b]$ from a distribution with density

$$\bar{p}(x) = \frac{f(x)}{F(b) - F(a)}$$

   by using the transformation method. Pick an additional random number y equidistributed in the interval $[0, f(x)]$.
2. If $y \leq p(x)$ accept x as the next random number, else return to Step 1.

Multivariate Gaussian Distribution:

$$\boxed{p(x_1,\ldots,x_n) = \frac{1}{\sqrt{(2\pi)^n\, S}}\, e^{-\frac{1}{2}\sum\sum g_{ij}\, x_i\, x_j}}$$

or

$$p(\mathbf{x}) \quad = \quad \frac{1}{\sqrt{(2\pi)^n\, S}}\, e^{-\frac{1}{2}\mathbf{x}^T\cdot\mathbf{G}\cdot\mathbf{x}} \equiv \frac{1}{\sqrt{(2\pi)^n\, S}}\, e^{-\frac{1}{2}\,Q}$$

with the *covariance matrix* of the x_i

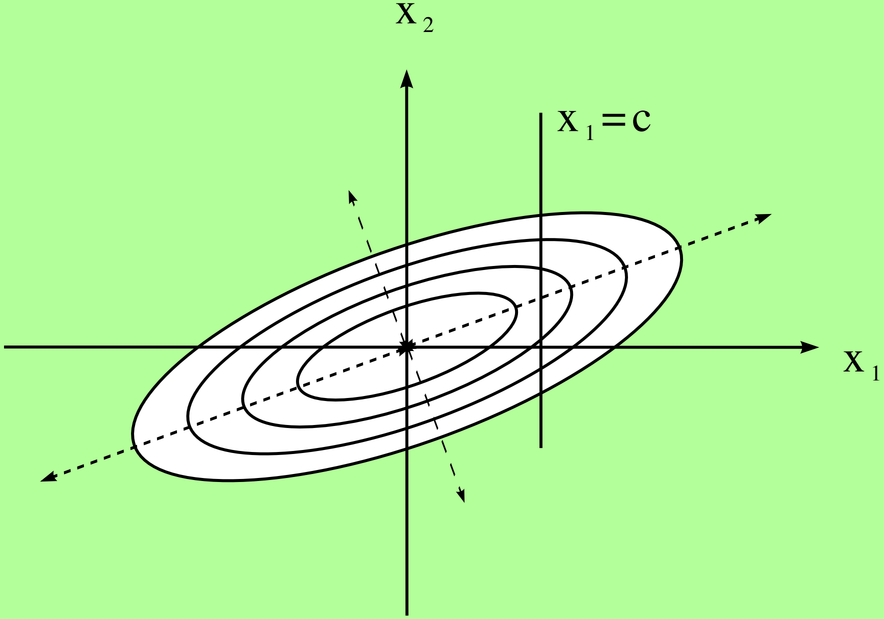
$$\mathbf{S} \quad \equiv \quad \mathbf{G}^{-1} \equiv \begin{pmatrix} \langle x_1^2\rangle & \langle x_1\, x_2\rangle & \ldots \\ \vdots & \langle x_2^2\rangle & \ldots \\ & & \ddots \end{pmatrix}$$

$S \equiv |\mathbf{S}|$ is the determinant of this matrix. $\mathbf{S}$ and $\mathbf{G}$ are symmetric, their eigenvalues are called σ_i^2 and γ_i (sorry!).

Example: Assume that two Gaussian variates have the variances $s_{11} \equiv \langle x_1^2 \rangle = 3$, $s_{22} \equiv \langle x_2^2 \rangle = 4$, and the covariance $s_{12} \equiv \langle x_1 \, x_2 \rangle = 2$:

$$\mathbf{S} = \begin{pmatrix} 3 & 2 \\ 2 & 4 \end{pmatrix}; \quad \mathbf{G} \equiv \mathbf{S}^{-1} = \begin{pmatrix} \frac{1}{2} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{3}{8} \end{pmatrix}$$

The quadratic form Q in the exponent is then $Q = (1/2)\, x_1^2 - (1/2)\, x_1 \, x_2 + (3/8)\, x_2^2$., and the lines of equal density (that is, of equal Q) are ellipses which are inclined with respect to the $x_{1,2}$ coordinate axes:



Rotate the axes of the ellipsoids $Q = const$ to coincide with the coordinate axes: $\Longrightarrow$ cross correlations vanish!

Principal axis transformation:

- Determine eigenvalues γ_j and eigenvectors $\mathbf{g}_j$ of $\mathbf{G}$. (Use NAG-F02AMF, ESSL-SSYGV, or your own code.)
- Combine the n column vectors $\mathbf{g}_j$ into a matrix $\mathbf{T}$. This matrix diagonalizes $\mathbf{G}$ (and consequently Q.)

Since $\mathbf{T}$ is orthogonal ($\mathbf{T}^T = \mathbf{T}^{-1}$) it diagonalizes not only $\mathbf{G} \equiv \mathbf{S}^{-1}$ but also $\mathbf{S}$ itself. $\Longrightarrow$ $\mathbf{S}^{-1}$ need never be computed!

Having found $\mathbf{T}$, we arrive at the following prescription for the production of correlated Gaussian variables: $\Longrightarrow$

Multivariate Gaussian distribution:

Let the covariance matrix $\mathbf{S}$ be given.

- Determine, by principal axis transformation, the diagonalization matrix $\mathbf{T}$ for $\mathbf{S}$ (This step is performed only once.)
- Generate n mutually independent Gaussian random variates y_i with the variances σ_i^2.
- Transform the vector $\mathbf{y} \equiv (y_1 \ldots y_n)^T$ according to

$$\mathbf{x} \;=\; \mathbf{T} \cdot \mathbf{y}$$

The n elements of the vector $\mathbf{x}$ are then random numbers obeying the desired distribution.

Let's try it out: $\Longrightarrow$

Example: Once more, let

$$\mathbf{S} = \begin{pmatrix} 3 & 2 \\ 2 & 4 \end{pmatrix}, \text{ with the inverse } \mathbf{G} = \begin{pmatrix} \frac{1}{2} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{3}{8} \end{pmatrix}$$

Principal axis transformation: The eigenvalues of $\mathbf{S}$ are $\sigma^2_{1,2} = (7 \pm \sqrt{17})/2 = 5.562|1.438$, and the corresponding eigenvectors are

$$\mathbf{s}_1 = \begin{pmatrix} 0.615 \\ 0.788 \end{pmatrix} \quad \mathbf{s}_2 = \begin{pmatrix} 0.788 \\ -0.615 \end{pmatrix} \quad \text{Thus} \quad \mathbf{T} = \begin{pmatrix} 0.615 & 0.788 \\ 0.788 & -0.615 \end{pmatrix}$$

Generator: To produce pairs (x_1, x_2) of Gaussian random numbers with the given covariance matrix:

- Draw y_1 and y_2 Gaussian, uncorrelated, with variances 5.562 and 1.438.
- Compute x_1 and x_2 according to

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0.615 & 0.788 \\ 0.788 & -0.615 \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

Exercise: Write a program that generates a sequence of bivariate Gaussian random numbers with the statistical properties as assumed in the foregoing example. Determine $\langle x_1^2 \rangle$, $\langle x_2^2 \rangle$, and $\langle x_1 x_2 \rangle$ to see if they indeed approach the given values of 3, 4, and 2.

Homogeneous distributions in Orientation Space:

Equidistribution on the unit circle:

- Draw a pair of equidistributed random numbers $(y_1, y_2) \in (-1, 1)^2$; compute $r^2 = y_1^2 + y_2^2$; if necessary, repeat until $r^2 \leq 1$.
- $x_1 \equiv y_1/r$ and $x_2 \equiv y_2/r$ are the cartesian coordinates of points that are homogeneously distributed on the circumference of the unit circle.

Equidistribution on a spherical surface:

- Draw pairs of random numbers $(y_1, y_2) \in (-1, 1)^2$ until $r^2 \equiv y_1^2 + y_2^2 \leq 1$.
- The quantities

$$
\begin{aligned}
x_1 &= 2y_1\sqrt{1-r^2} \\
x_2 &= 2y_2\sqrt{1-r^2} \\
x_3 &= 1-2r^2
\end{aligned}
$$

  are then the cartesian coordinates of points out of a homogeneous distribution on the surface of the unit sphere.

(Generalization to *hyperspherical* surfaces: see Vesely, Comp. Phys.)

Random Sequences:

So far: random numbers, preferably no serial correlations $\langle x_n \, x_{n+k} \rangle$.
Now: *sequences* of r. n. with *given* serial correlations.

Let $\{x(t)\}$ be an ensemble of functions of time t. Then

$$P_1(x;t) \quad \equiv \quad \mathcal{P}\,\{x(t) \le x\} \quad \text{and} \quad p_1(x;t) \equiv \frac{dP_1(x;t)}{dx}$$

are the probability distribution and the respective density.

Example: Let $x_0(t)$ be a deterministic function of time, and assume that the quantity $x(t)$ at any time t be Gauss distributed about the value $x_0(t)$:

$$p_1(x;t) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{1}{2}\,[x - x_0(t)]^2\,/\sigma^2}$$

A random process is called a *random sequence* if the variable t may assume only discrete values $\{t_k\,;\; k = 0, 1, \ldots\}$. In this case one often writes $x(k)$ for $x(t_k)$.

The foregoing definitions may be generalized in the following manner:

$$
\begin{aligned}
P_2(x_1, x_2; t_1, t_2) &\equiv \mathcal{P}\{x(t_1) \le x_1, x(t_2) \le x_2\} \\
&\vdots \\
P_n(x_1, \ldots, x_n; t_1, \ldots, t_n) &\equiv \mathcal{P}\{x(t_1) \le x_1, \ldots, x(t_n) \le x_n\}
\end{aligned}
$$

Thus $P_2(..)$ is the compound probability for the events $x(t_1) \le x_1$ *and* $x(t_2) \le x_2$. These higher order distribution functions and the corresponding densities

$$
p_n(x_1, \ldots, x_n; t_1, \ldots, t_n) = \frac{d^n P_n(x_1, \ldots, x_n; t_1, \ldots, t_n)}{dx_1 \ldots dx_n}
$$

describe the random process in ever more – statistical – detail.

Stationarity: A random process is stationary if

$$P_n(x_1, \ldots, x_n; t_1, \ldots, t_n) \;=\; P_n(x_1, \ldots, x_n; t_1 + t, \ldots, t_n + t)\,,$$

This means that the origin of time is of no importance:

$$p_1(x;t) = p_1(x) \qquad \text{and} \qquad p_2(x_1, x_2; t_1, t_2) = p_2(x_1, x_2; t_2 - t_1)$$

Autocorrelation:

$$\langle x(0)\, x(\tau)\rangle \;\equiv\; \int_a^b \int_a^b x_1 x_2\, p_2(x_1, x_2; \tau)\, dx_1\, dx_2\,,$$

For $\tau \to 0$ the autocorrelation function (acf) approaches the variance $\langle x^2\rangle$. For finite τ it tells us how rapidly a particular value of $x(t)$ will be "forgotten".

Gaussian process: The random variables $x(t_1), \ldots, x(t_n)$ obey a multivariate Gaussian distribution. The covariance matrix elements are $\langle x(0)\, x(t_j - t_i)\rangle$, i.e. the values of the autocorrelation function at the specific time displacement:

$$p_2(x_1\,, x_2\,; \tau) \;\; = \;\; \frac{1}{\sqrt{(2\pi)^2 S_2(\tau)}}\, e^{-\frac{1}{2}Q}$$

with

$$Q \;\; \equiv \;\; \frac{\langle x^2\rangle x_1^2 - 2\langle x(0)x(\tau)\rangle x_1\, x_2 + \langle x^2\rangle x_2^2}{S_2(\tau)}$$

and

$$S_2(\tau) \;\; \equiv \;\; |\mathbf{S}_2(\tau)| = \langle x^2\rangle^2 - \langle x(0)\, x(\tau)\rangle^2$$

Markov Process: A stationary random sequence $\{x_n; n = 0, 1 \ldots\}$ has the *Markov property* if its "memory" goes back only one time step:

$$p(x_n | x_{n-1} \ldots x_1) \quad = \quad p(x_n | x_{n-1})$$

where the *conditional density*

$$p(x_n | x_{n-1}; \tau) \quad = \quad \frac{p_2(x_{n-1}, x_n)}{p_1(x_{n-1})}$$

is the density of x_n *under the condition* that $x(n-1) = x_{n-1}$.
Thus all statistical properties of the process are contained in $p_2(x_{n-1}, x_n)$.

An even shorter memory would mean that successive elements of the sequence were not correlated at all.

Gaussian Markov processes: To describe them uniquely not even $p_2(\ldots)$ is needed. If the autocorrelation function $\langle x(n)\, x(n+l)\rangle$ is known, $p_2(..)$ and consequently all statistical properties of the process follow.
Note: The acf of a stationary Gaussian Markov process is always an exponential:

$$\langle x(0)\, x(\tau)\rangle \;=\; \langle x^2\rangle e^{-\beta\tau}$$

or

$$\langle x(n)\, x(n+k)\rangle \;=\; \langle x^2\rangle e^{-\beta\Delta t\, k}$$

How to produce a Markov sequence? $\Longrightarrow$

Generating a stationary Gaussian Markov sequence:

Solve the stochastic differential equation

$$\dot{x}(t) \;\;=\;\; -\beta\, x(t) + s(t)$$

with a stochastic "driving" process $s(t)$, assumed to be uncorrelated Gaussian noise, i.e. Gauss distributed about $\langle s \rangle = 0$, with $\langle s(0)\, s(t) \rangle = A\,\delta(t)$.

The general solution to this equation reads

$$x(t) \;\;=\;\; x(0)\, e^{-\beta\, t} + \int_0^t e^{-\beta(t-t')}\, s(t')\, dt'$$

Inserting $t = t_n$ and $t = t_{n+1} \equiv t_n + \Delta t$ one finds that

$$x(t_{n+1}) \;\;=\;\; x(t_n)\, e^{-\beta\,\Delta t} + \int_0^{\Delta t} e^{-\beta(\Delta t - t')}\, s(t_n + t')\, dt'$$

At any time t, the values of $x(t)$ belong to a stationary Gauss distribution with $\langle x^2 \rangle = A/2\beta$, and the process $\{x(t_n)\}$ has the Markov property.

The integrals

$$z(t_n) \equiv \int_0^{\Delta t} e^{-\beta(\Delta t - t')}\, s(t_n + t')\, dt'$$

are elements of a random sequence, with z Gauss distributed with zero mean and $\langle z(t_n)\, z(t_{n+k}) \rangle = 0$ for $k \neq 0$. Their variance is

$$\langle z^2 \rangle = \frac{A}{2\beta}\,(1 - e^{-2\beta\,\Delta t})$$

Here is the resulting recipe for generating a stationary, Gaussian Markov sequence: $\Longrightarrow$

"Langevin Shuffle":

Let the desired stationary Gaussian Markov sequence $\{x(n)\,;\; n = 0, \ldots\}$ be defined by the autocorrelation function

$$\langle x(n)\, x(n+k)\rangle \;=\; \frac{A}{2\beta}\, e^{-\beta\, k\, \Delta t}$$

with given parameters A, β and Δt. Choose a starting value $x(0)$, either as $x(0) = 0$ or from a Gauss distribution with $\langle x\rangle = 0$ and $\langle x^2\rangle = A/2\beta$.

- Draw $z(n)$ from a Gaussian distribution with $\langle z\rangle = 0$ and

$$\langle z^2\rangle \;=\; \frac{A}{2\beta}\,(1 - e^{-2\beta\, \Delta t})$$

- Construct

$$x(n+1) \;=\; x(n)\, e^{-\beta\, \Delta t} + z(n)$$

The random sequence thus produced has the desired properties.

If $\beta\, \Delta t \ll 1$, replace the exponential by its linear Taylor approximation. The iteration prescription then reads $x(n+1) = x(n)\,(1 - \beta\, \Delta t) + z'(n)$, where $z'(n)$ is Gaussian with $\langle z'^2\rangle = A\, \Delta t\, (1 - \beta\, \Delta t)$.

Exercise: Employ the above procedure to generate a Markov sequence $\{x_n\}$ with a given β. Check if the sequence shows the expected autocorrelation.

Wiener-Lévy Process (Unbiased Random Walk)

With $\beta = 0$ in the above differential equation, we find

$$x(n+1) \;=\; x(n) + z(n)$$

where $z(n)$ is Gaussian with

$$z(n) \;\equiv\; \int_0^{\Delta t} s(t_n + t')\, dt' \quad \langle z \rangle = 0 \quad \langle z^2 \rangle = A\Delta t$$

Since z and x are uncorrelated, we have

$$\langle [x(n)]^2 \rangle \;=\; n\, A\, \Delta t$$

<u>Example:</u> Let x be one cartesian coordinate of a diffusing particle. Then $\langle [x(n)]^2 \rangle$ is the mean squared displacement after n time steps. In this case we may relate the coefficient A to the diffusion constant according to $A = 2D$.

Wiener-Lévy process:

Let $A\Delta t$ be given. Choose $x(0) = 0$.

- Pick $z(n)$ from a Gauss distribution with variance $A\Delta t$.
- Compute

$$x(n+1) \;=\; x(n) + z(n)$$

The random sequence thus produced is a nonstationary Gaussian process with variance $[x(n)]^2 = n\,A\,\Delta t$.

Exercise: 500 *random walkers* set out from positions $x(0)$ homogeneously distributed in the interval $[-1, 1]$. The initial particle density is thus rectangular. Each of the random walkers is now set on its course to perform its own one-dimensional trajectory, with $A\,\Delta t = 0.01$. Sketch the particle density after 100, 200, ... steps.

It is not really necessary to draw $z(n)$ from a Gaussian distribution. If $z(n)$ comes from an equidistribution in $[-\Delta x/2,\ \Delta x/2]$, the "compound" x-increment after every $10 - 15$ steps will again be Gauss distributed (central limit theorem).

We may even discretize the x-axis: $z = 0, \pm\Delta x$ with equal probability $1/3$: after many steps, and on a scale which makes Δx appear small, the results will again be the same.

To simulate 2- or 3-dimensional diffusion, apply the above procedure independently to 2 or 3 coordinates.

Markov Chains (Biased Random Walk)

A Markov sequence of discrete x_α is called a *Markov chain*.

We generalize the discussion to "state vectors" $\{\mathbf{x}_\alpha, \alpha = 1, \ldots M\}$. The conditional probability

$$p_{\alpha\beta} \quad \equiv \quad \mathcal{P}\left\{\mathbf{x}(n) = \mathbf{x}_\beta \,|\, \mathbf{x}(n-1) = \mathbf{x}_\alpha\right\}$$

is called *transition probability* between the states α and β.

Let M be the total number of possible states. The $M \times M$-matrix $\mathbf{P} \equiv \{p_{\alpha\beta}\}$ and the M-vector $\mathbf{p}$ consisting of the individual probabilities $p_\alpha \equiv \mathcal{P}\{\mathbf{x} = \mathbf{x}_\alpha\}$ determine the statistical properties of the Markov chain uniquely.

A Markov chain is *reversible* if

$$p_\alpha \, p_{\alpha\beta} \quad = \quad p_\beta \, p_{\beta\alpha} \tag{1}$$

– Meaning?

The M^2 elements of the matrix $\mathbf{P}$ are not uniquely defined by the $M(M-1)/2$ reversibility conditions. $\Longrightarrow$For a given distribution density $\mathbf{p}$ there are many reversible transition matrices. $\Longrightarrow$"Asymmetrical rule" (N. Metropolis):

N. Metropolis' asymmetric rule:

Let Z be the number of states $\mathbf{x}_\beta$ accessible from $\mathbf{x}_\alpha$, and let the a priori access probability be $\pi_{\alpha\beta} = 1/Z$. Then

$$p_{\alpha\beta} = \pi_{\alpha\beta} \qquad \text{if } \; p_\beta \geq p_\alpha$$

$$p_{\alpha\beta} = \pi_{\alpha\beta} \frac{p_\beta}{p_\alpha} \qquad \text{if } \; p_\beta < p_\alpha$$

$\Longrightarrow p_{\alpha\beta}$ is reversible!

Monte Carlo Method

Central theorem:

> If the stationary Markov chain characterized by $\mathbf{p} \equiv \{p_\alpha\}$ and $\mathbf{P} \equiv \{p_{\alpha\beta}\}$ is reversible, then each state $\mathbf{x}_\alpha$ will be visited, in the course of a sufficiently long chain, with the relative frequency p_α.

$\Longrightarrow$Here is yet another recipe for generating random numbers with a given probability density $\mathbf{p}$:

Random numbers à la Metropolis:

Let $\mathbf{p} \equiv \{p_\alpha;\ \alpha = 1, 2, \ldots\}$ be the vector of probabilities of the events $x = x_\alpha$. To generate a random sequence $\{x(n)\}$ in which the relative frequency of the event $x(n) = x_\alpha$ approaches p_α:

- After the n-th step, let $x(n) = x_\alpha$. Draw a value x_β from a region around x_α, e.g. according to

  $$x_\beta = x_\alpha + (\xi - 0.5)\,\Delta x$$

  where $\xi \in (0, 1)$.
- If for $p_\beta \equiv p(x_\beta)$ we have $p_\beta \geq p_\alpha$, then let $x(n+1) = x_\beta$.
- If $p_\beta < p_\alpha$, then pick a random number $\xi \in (0, 1)$; if $\xi < p_\beta / p_\alpha$, let $x(n+1) = x_\beta$; else put $x(n+1) = x_\alpha$.

Adjust the parameter Δx such that approximately one out of two trial moves leads to a new state, $x(n+1) = x_\beta$.

Warning: The random numbers thus produced are serially correlated: $\langle x(n)\, x(n+k) \rangle \neq 0$.

<u>Exercise:</u> Let $p(x) = A\, exp[-x^2]$ be the desired probability density. Apply the Metropolis' prescription to generate random numbers with this density. Confirm that $\langle x(n)\, x(n{+}k)\rangle \neq 0$.

Advantage of Metropolis' method: only p_β/p_α is needed, not p_α.

$\Longrightarrow$Statistical-mechanical Monte Carlo simulation: only *relative* thermodynamic probabilities needed!

Stochastic Optimization

Finding the global extremum of a function of many variables:
– Nonlinear fit to a set of table values
– improvement of complex electronic circuits (“travelling salesman problem”)
– find the most stable (i. e. lowest energy) configuration of microclusters or biopolymers.
– . . .

Two methods:

– **Simulated Annealing**

– **Genetic Algorithms**

Simulated Annealing

Consider a Metropolis walk through the space of "states" $\mathbf{x}_\alpha$ with

$$p_\alpha \quad = \quad A \exp - \beta U(\mathbf{x})$$

where $U(x_1, \ldots x_M)$ is a "cost function" to be minimized, and β a tunable parameter (a reciprocal "temperature".)

– Low $\beta \Longrightarrow$ smaller variation of p_α; higher $U(\mathbf{x})$ are accessible

– High $\beta \Longrightarrow \mathbf{x}$ will tend to go "downhill"

Simulated Annealing:

Draw a starting vector $\mathbf{x}^0 \equiv \{x_1^0, \ldots x_M^0\}$ at random, and choose a high initial "temperature" kT.

Carefully lower the temperature: $\Longrightarrow$regions with lower $U(\mathbf{x})$ will be visited more frequently than the higher ranges.

Finally, for $kT \rightarrow 0$ the system point will come to rest in a minimum that very probably (not with certainty!) will be the global minimum.

<u>Exercise:</u> Create (fake!) a table of “measured values with errors” according to

$$y_i = f(x_i; c_1, \ldots c_6) + \xi_i\,, \quad i = 1, 20 \tag{2}$$

with ξ_i coming from a Gauss distribution with suitable variance, and with the function f defined by

$$f(x; \mathbf{c}) \equiv c_1 e^{-c_2 (x - c_3)^2} + c_4 e^{-c_5 (x - c_6)^2} \tag{3}$$

($c_1 \ldots c_6$ being a set of arbitrary coefficients).

Using these data, try to reconstruct the parameters $c_1 \ldots c_6$ by fitting the theoretical function f to the table points (x_i, y_i). The cost function is

$$U(\mathbf{c}) \equiv \sum_i \left[y_i - f(x_i; \mathbf{c}) \right]^2 \tag{4}$$

Choose an initial vector $\mathbf{c}^0$ and perform an MC random walk through $\mathbf{c}$-space, slowly lowering the temperature.

Genetic Algorithms

Evolution of biological systems:
– Adaptation of species to external conditions: optimization

– Adaptation strategy itself has evolved over time: sexual reproduction

$\Longrightarrow$More sex:

Consider some oscillatory function $f(x)$ of a single variable, having one global minimum within the range of definition, $x \epsilon [a,b]$. Find x^* with $f(x^*) = min\{f(x),\, x\epsilon[a,b]\}$.

1. Start with a *population* of randomly chosen numbers (*individuals*), $\{x_i^0 \epsilon [a,b],$ $i = 1, \ldots N\}$. (N is kept constant.)
   – "Gene": *bit string* of x_i^0
   – "Fitness": low $f_i \equiv f(x_i^0)$ = high fitness and vice versa
   – Relative fitness (probability of reproduction): $p_i \equiv f_i / \sum_{i=1}^N f_i$. This is a probability density, and $P(x_i) \equiv P_i \equiv \sum_{j=1}^i p_j$ is its cumulative distribution function.

2. Draw N individuals in accordance with their reproduction probability (Transformation method!).

   The new population $\{x_i',\ i = 1, \ldots N\}$ is fitter than the original one. However, thus far we have remained at the level of primitive selective reproduction without mutation or sexual crossover.

3. Pick pairs of individuals at random and submit their genetic strings are to crossover:
   () Draw a position m within the bit strings; () swap the bits following m between the two strings. The number of such pairings, the "crossover rate",

should be around $0.6\,N$. The resulting set $\{x_i'',\ i = 1, \ldots N\}$ is called the *offspring* population.

4. Finally, *mutation* comes into play: within each string x_i'' every single bit is reversed with a probability $p_{mut} \approx 0.01$.

   The resulting population is regarded as the next generation, $\{x_i^1,\ i = 1, \ldots N\}$, and we are back at step 2.

<u>Exercise:</u> Apply the simple genetic algorithm to find the minimum of the function $[2\,sin(10\,x-1)]^2+10\,(x-1)^2$ within the interval $[0, 2]$.