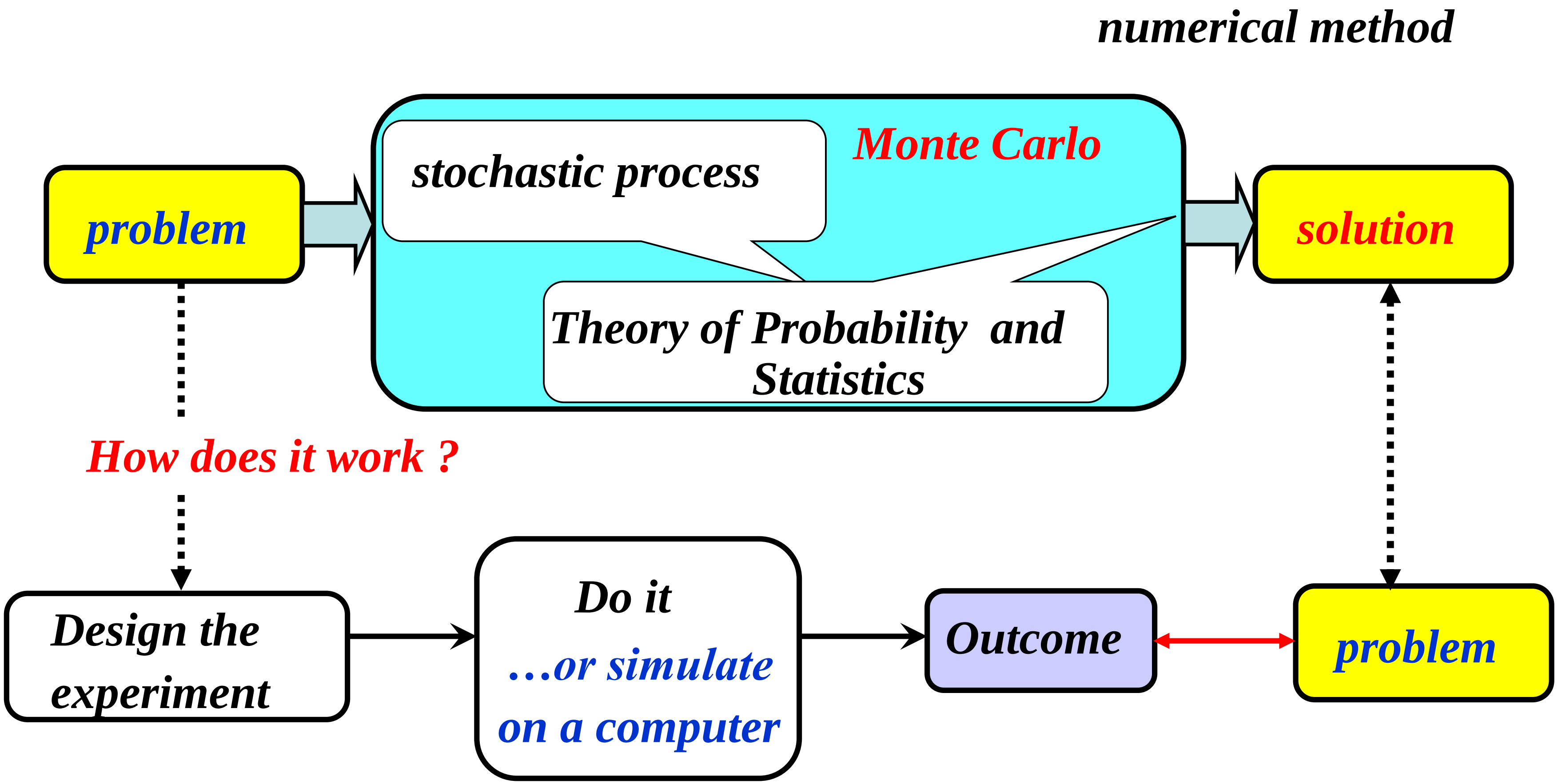
Introduction to Monte Carlo

Carlos Mana
Benasque-2014
Astrofísica y Física de Partículas
CIEMAT

What is the Monte Carlo method ?



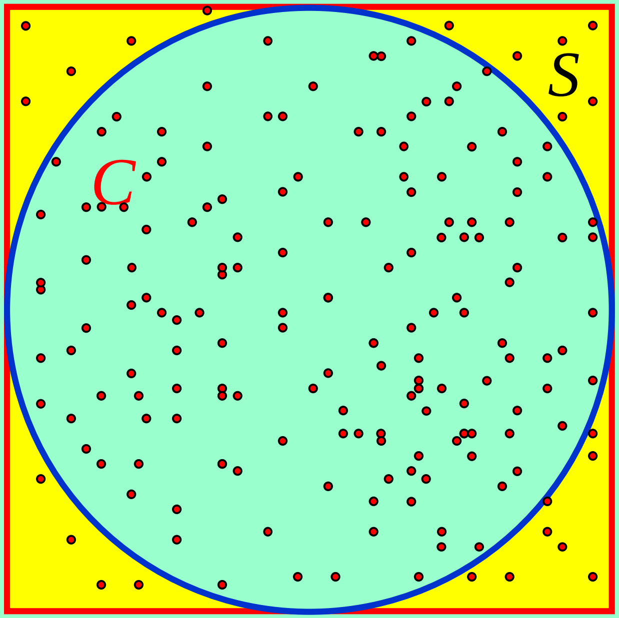
What kind of problems can we tackle ?

Why Monte Carlo ? incapability? Particle physics, Bayesian inference,…

What do we get ? Magic Box?, …, more information?...

Important and General Feature of Monte Carlo estimations:

Example: Estimation of π (… a la Laplace)



Probability that a draw falls inside the circumference

$$\theta = \frac{\mu(C)}{\mu(S)} = \frac{\pi}{4}$$

N ***trials***

event $X =$ ***{Point falls inside the circumference}***

$$X \sim Bi(x \mid N, \theta)$$

Throws (N)	***Accepted (n)***	$\tilde{\pi} \approx 4\frac{n}{N}$	$\sigma_{\tilde{\pi}} \approx \left[\tilde{\pi}(4-\tilde{\pi})/N\right]^{1/2}$
100	83	3.32	0.1503
1000	770	3.08	0.0532
10000	7789	3.1156	0.0166
100000	78408	3.1363	0.0052
1000000	785241	3.141	0.0016

$$\theta \sim Be(\theta \mid n + 1/2, N - n + 1/2)$$

$$\pi = 4\theta$$

… for large N:

$$\tilde{\pi} = \frac{4n}{N}$$

$$\sigma(\tilde{\pi}) \approx \frac{1.6}{\sqrt{N}}$$

The dependence of the uncertainty with $1/\sqrt{N}$ ***is a general feature of the Monte Carlo estimations regardless the number of dimensions of the problem***

Basis of Monte Carlo simulations

Design the experiment → **Do it** *...or simulate on a computer*

$$\left\{ \begin{array}{c} \text{Get a sequence} \\ \{z_1, z_2, \ldots z_n\} \\ \text{of random numbers} \end{array} \right\} \equiv \left\{ \begin{array}{c} \text{Get a sampling of size } n \\ \text{from a stochastic} \\ \text{process} \end{array} \right\}$$

Need sequences of random numbers

Do not do the experiment but….

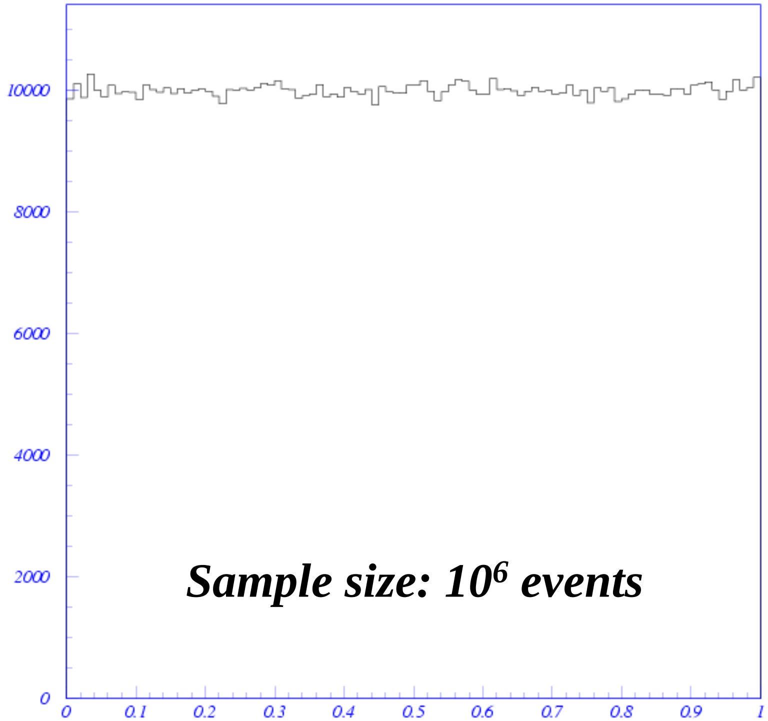
generate sequences of (pseudo-)random numbers on a computer…

"Anyone who considers arithmetical methods of producing random digits is, of course, in a state of sin"

J. von Neumann

(∞ **Refs.**)

$$X \sim Un^D(x \mid 0,1)$$



$$Un(x|0,1)$$

$$\mu = 0.5000$$
$$\sigma^2 = 0.0833 \left(= 1/12\right)$$
$$(\gamma_1) = 0.0$$
$$(\gamma_2) = -1.2 \left(= -6/5\right)$$

Sampling moments:

$$m_k = \frac{1}{n}\sum_{j=1}^{n} x_j^k$$

$$c_k = \frac{1}{n-k}\sum_{j=1}^{n-k} x_j x_{j+k}$$

$$m_1 = 0.5002$$
$$\sigma^2_{sam} = 0.0833 \qquad \rho(x_i, x_{i+1}) = 0.00076$$
$$(\gamma_1)_{sam} = 0.0025 \qquad \rho(x_i, x_{i+2}) = -0.00096$$
$$(\gamma_2)_{sam} = -1.2029 \qquad \rho(x_i, x_{i+3}) = -0.00010$$
$$\delta \approx 0.0003$$

We can generate pseudo-random sequences $\{z_1, z_2, \ldots z_k\}$ ***with*** $Z \sim Un(z|0,1)$

Great!
But life is not uniform…

In Physics, Statistics,… we want to generate sequences from

$$\boldsymbol{X} = \{x_1, x_2, \ldots, x_n\} \sim p(\boldsymbol{x} \,|\, \boldsymbol{\theta})$$

How?

- ***Inverse Transform***
- ***Acceptance-Rejection ("Hit-Miss")***
- ***"Importance Sampling",…***

Usually a combination of them

+

- ***Tricks***
- ***Normal Distribution***
- ***M(RT)²   (Markov Chains)***
- ***Gibbs Sampling,….***

Method 1: Inverse Transform

- *We want to generate a sample of* $X \sim p(x|\theta)$

$$P(X \in (-\infty, x] | \theta) = \int_{-\infty}^{x} p(x|\theta)dx = \int_{-\infty}^{x} dF(x|\theta) = F(x|\theta)$$

- $Y = F(X|\theta) : \Omega_X \to [0,1]$ *How is* Y *distributed?*

$$F_Y(y) \equiv P(Y \le y) = P(F(X|\theta) \le y) = P\left(X \le F^{-1}(y|\theta)\right) = \int_{-\infty}^{F^{-1}(y|\theta)} dF(x|\theta) = y$$

$$F_Y(y) = y \quad \longrightarrow \quad \boxed{Y = F(X|\theta) \sim Un(x|0,1)}$$

Algorithm

n)

i) Sample	$Y \sim Un(y|0,1)$	→	$y_i \sim Un(y|0,1)$
ii) Transform	$X = F^{-1}(Y|\theta)$	→	$x_i = F^{-1}(y_i|\theta)$

$$\{x_1, x_2, \ldots x_n\} \sim p(x|\theta)$$

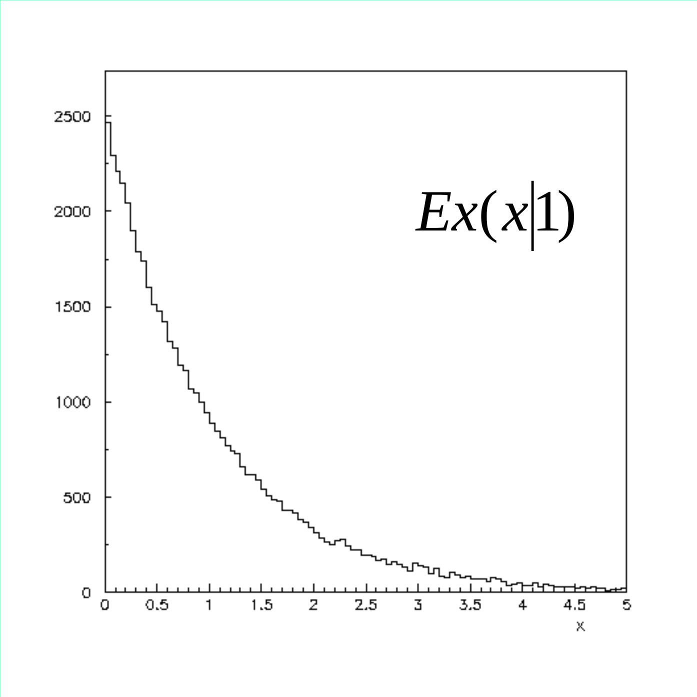
Examples:

Sampling of $X \sim Ex(x|\theta)$

$$p(x|\theta) = \theta\, e^{-\theta x} \qquad F(x|\theta) = 1 - e^{-\theta x}$$

$i)\quad u_i \sim Un(u|0,1)$

$ii)\quad x_i = F^{-1}(u_i|\theta) = -\dfrac{\ln u_i}{\theta}$



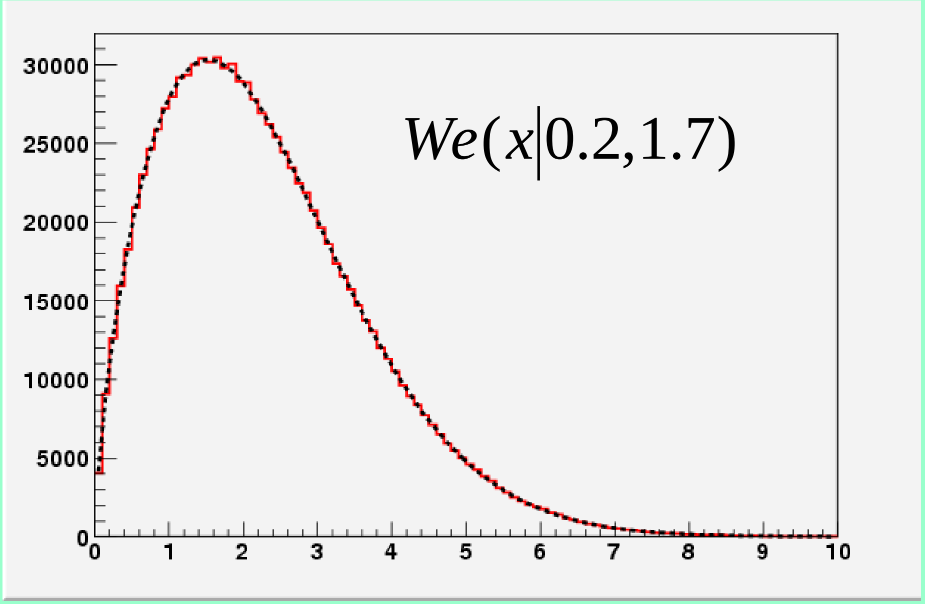
Sampling of $X \sim We(x|\theta_1,\theta_2)$

$$p(x|\boldsymbol{\theta}) = \theta_1\theta_2 x^{\theta_2-1} e^{-\theta_1 x^{\theta_2}}\ \mathbf{1}_{[0,\infty)}(x)$$

$$F(x|\boldsymbol{\theta}) = 1 - e^{-\theta_1 x^{\theta_2}}\ \mathbf{1}_{[0,\infty)}(x)$$

$i)\quad u_i \sim Un(u|0,1)$

$ii)\quad x_i = F^{-1}(u_i\,|\,\boldsymbol{\theta}) = \left[-\dfrac{1}{\theta_1}\ln u_i\right]^{1/\theta_2}$



…some fun…

Problem:

1) Generate a sample : $\{x_1, x_2, \ldots x_n\}$ $\quad n = 10^6$

$X \sim Un(x|0,1)$ $\quad X \sim p(x) = \frac{3}{2} x^2 \, \boldsymbol{I}_{[-1,1]}(x)$ $\quad X \sim Ca(x|0,1) = \frac{1}{\pi} \frac{1}{1+x^2}$

2) Get for each case the sampling distribution of $\quad Y_m = \frac{1}{m} \sum_{k=1}^{m} X_k$

$m = \{2, 5, 10, 20, 50\}$

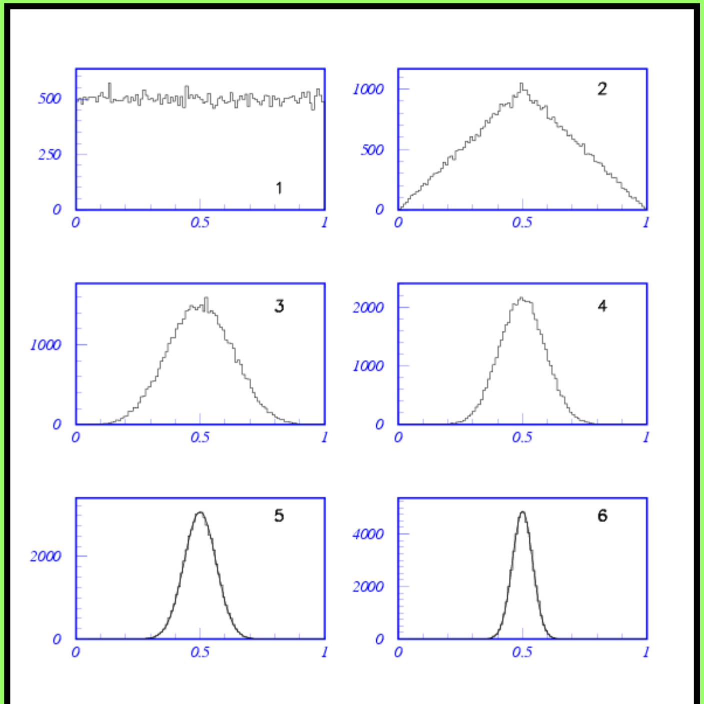
3) Discuss the sampling distribution of Y_m ***in connection with the Law of Large Numbers and the Central Limit Theorem***

4) If $X \sim Un(x|0,1)$ ***and*** $W_n = \prod_{k=1}^{n} U_k \in [0,1]$ ***How is*** $Z_n = -\log W_n$ ***distributed?***

5) If $X_k \sim Ga(x|0,k)$ ***How is*** $Z_n = \frac{X_n}{X_n + X_m}$ ***distributed?***

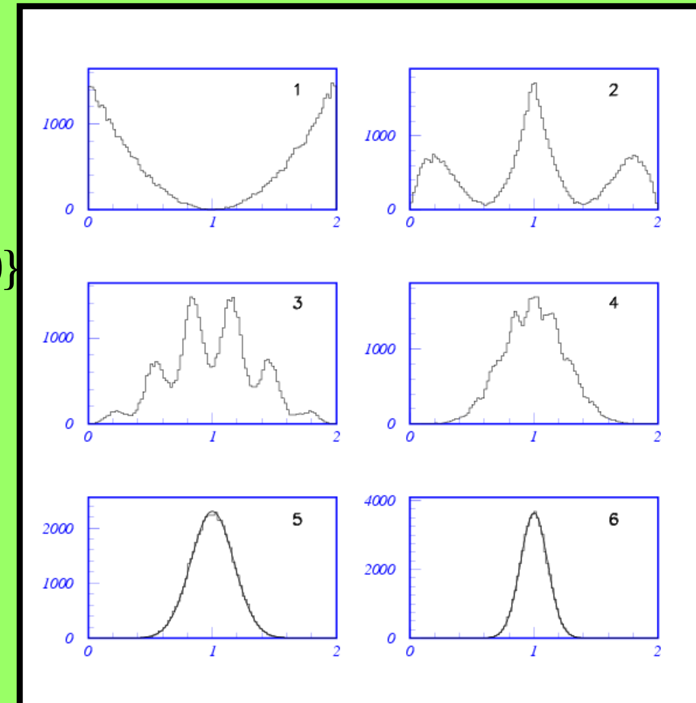
6) If $X_i \sim Ga(x \mid \alpha, \beta_i)$ ***How is*** $Y = \sum_i X_i$ ***distributed?***

(assumed to be independent random quantities)

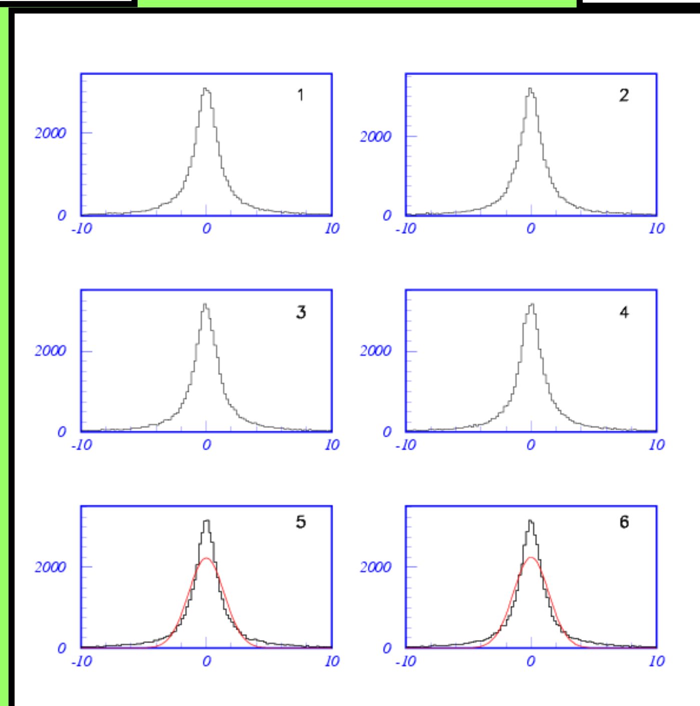


$$Y_m = \frac{1}{m}\sum_{k=1}^{m} X_k$$

$$m = \{1, 2, 5, 10, 20, 50\}$$



$$X \sim Un(x|0,1)$$



$$X \sim p(x) = \frac{3}{2}x^2 \, \boldsymbol{I}_{[-1,1]}(x)$$

$$X \sim Ca(x|0,1) = \frac{1}{\pi}\frac{1}{1+x^2}$$

For discrete random quantities

$$X \sim p(X=k|\theta) = p_k$$

$$k = 0, 1, 2, \ldots$$

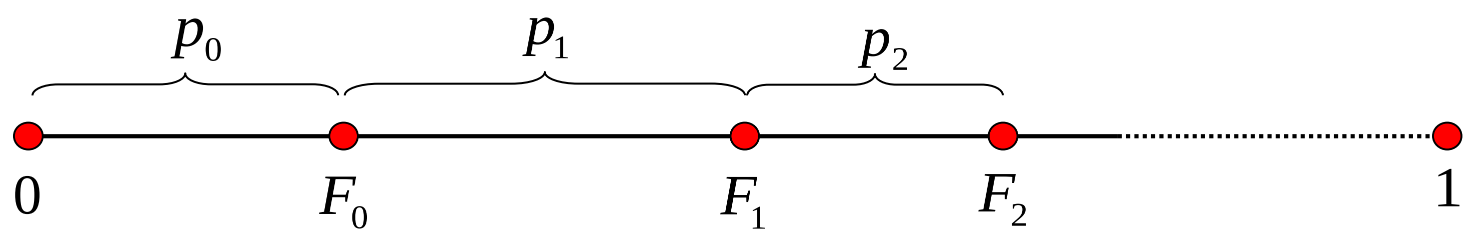
$$F_0 = P(X \le 0|\theta) = p_0$$

$$F_1 = P(X \le 1|\theta) = p_0 + p_1$$

$$\ldots$$

$$F_n = P(X \le n|\theta) = \sum_{k=1}^{n} p_k$$

$$\ldots$$



Algorithm

n)

i) Sample $U \sim Un(x|0,1)$ $\Longrightarrow$ $u_i \sim Un(u|0,1)$

ii) Find $k \mid F_{k-1} < u_i \le F_k$ $\Longrightarrow$ $x_i = k$

$(F_{-1} = 0)$

$$\{x_1, x_2, \ldots x_n\} \sim p(x|\theta)$$

Example: Sampling of $X \sim Po(X \mid \mu)$

$$P(X=k|\mu)=\frac{e^{-\mu}\mu^k}{k!}=\frac{\mu}{k}P(X=k-1|\mu)$$

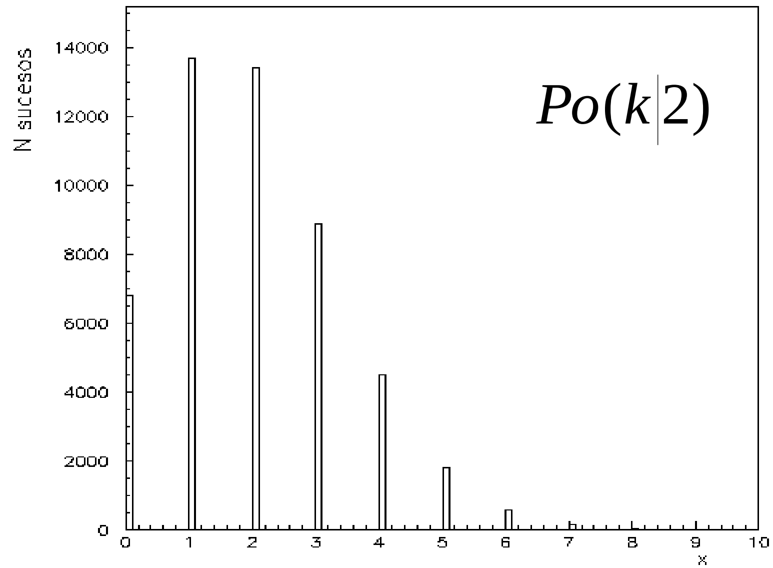
i) $u_i \sim Un(u|0,1)$

ii) a) $p_k=\frac{\mu}{k}p_{k-1}$ $\qquad p_0=e^{-\mu}$

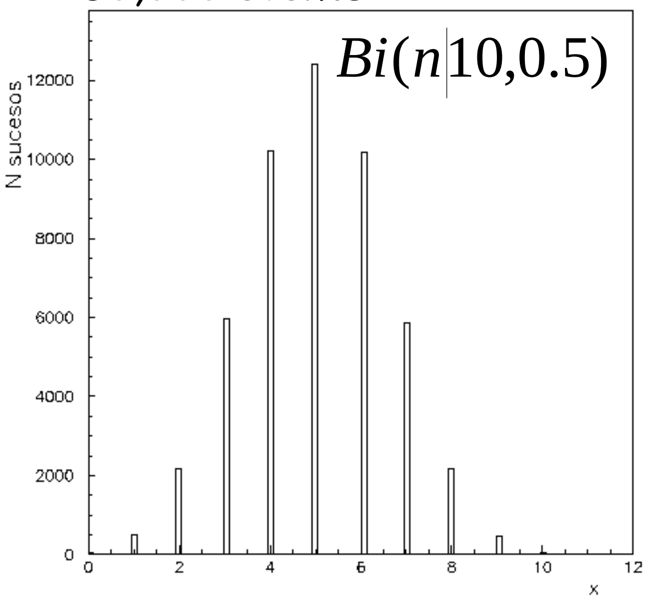
b) $F_k=F_{k-1}+p_k$

from $k=0$ until $F_{k-1}<u_i\le F_k$ → k

50,000 events



50,000 events



Sampling of $X \sim Bi(X \mid N,\theta)=\binom{N}{n}\theta^n(1-\theta)^{N-n}$

$$P(X=n\mid N,\theta)=\frac{N+n-1}{n}\frac{\theta}{1-\theta}P(X=n-1\mid N,\theta)$$

i) $u_i \sim Un(u|0,1)$

ii) a) $p_n=\frac{N+n-1}{n}\frac{\theta}{1-\theta}p_{n-1}$ $\qquad p_0=(1-\theta)^N$

b) $F_n=F_{n-1}+p_n$

from $n=0$ until $F_{n-1}<u_i\le F_n$ → n

Examples:

$$X \sim Bi(X \mid N,\theta) = \binom{N}{n} \theta^n (1-\theta)^{N-n}$$

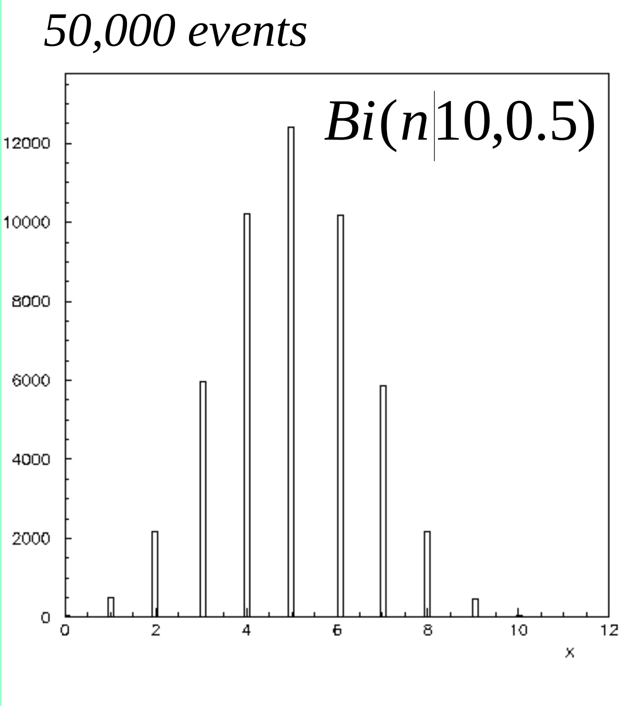
$$P(X = n \mid N,\theta) = \frac{N+n-1}{n} \frac{\theta}{1-\theta} P(X = n-1 \mid N,\theta)$$

i) $u_i \sim Un(u \mid 0,1)$

ii) a) $p_n = \frac{N+n-1}{n} \frac{\theta}{1-\theta} p_{n-1}$ $p_0 = (1-\theta)^N$

b) $F_n = F_{n-1} + p_n$

from $n = 0$ **until** $F_{n-1} < u_i \le F_n$ → n



Sampling of $X \sim Po(X \mid \mu)$

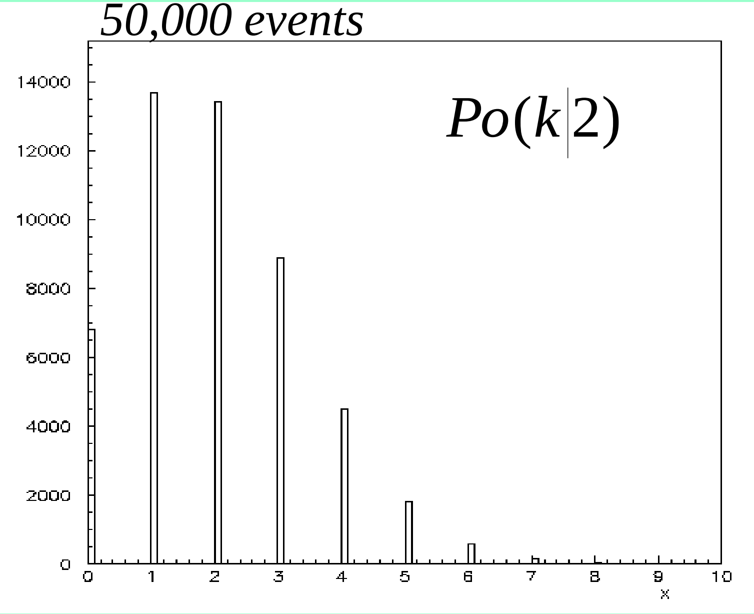
$$P(X = k \mid \mu) = \frac{e^{-\mu} \mu^k}{k!} = \frac{\mu}{k} P(X = k-1 \mid \mu)$$

i) $u_i \sim Un(u \mid 0,1)$

ii) a) $p_k = \frac{\mu}{k} p_{k-1}$ $p_0 = e^{-\mu}$

b) $F_k = F_{k-1} + p_k$

from $k = 0$ **until** $F_{k-1} < u_i \le F_k$ → k



Even though Discrete Distributions can be sampled this way,
usually one can find more efficient algorithms…

Example: $X \sim Po(X \mid \mu)$

$$\left.\begin{array}{l} U_i \sim Un(u \mid 0,1) \\ W_n = \prod_{k=1}^{n} U_k \quad \in [0,1] \end{array}\right\} \quad W_n \sim p(w_n \mid n) = \frac{1}{\Gamma(n)} (-\log w_n)^{n-1}$$

$$\boldsymbol{P}(W_n \le a) = \frac{1}{\Gamma(n)} \int_{-\log a}^{+\infty} e^{-x} x^{n-1} dx \xrightarrow{a = e^{-\mu}} \frac{e^{-\mu}}{\Gamma(n)} \sum_{k=1}^{n} \frac{\mu^{n-k}\Gamma(n)}{\Gamma(n-k+1)} = e^{-\mu} \sum_{m=0}^{n-1} \frac{\mu^m}{\Gamma(m+1)}$$

$$\boldsymbol{P}(W_n \le e^{-\mu}) = e^{-\mu} \sum_{m=0}^{n-1} \frac{\mu^m}{\Gamma(m+1)} = \boldsymbol{Po}(X \le n-1)$$

i) generate $u_i \sim Un(u \mid 0,1)$

ii) multiply them $w_n = u_1 \cdots u_n$

and go to i) while $w_n \le e^{-\mu}$

iii) deliver $x = n-1 \sim Po(x \mid \mu)$

PROBLEMS:

Gamma Distribution:

1) Show that if $X \sim Ga(x\,|\,a,b)$ then $Y = aX \sim Ga(y\,|\,1,b)$

2) Show that if $b = n \in N$ $\quad Y = \sum_{i=1}^{n} Z_i \sim Ga(y\,|\,1,n)$

$$Z_i \sim Ga(z\,|\,1,1) = Ex(z\,|\,1)$$

(→ generate exponentials)

Beta Distribution:

1) Show that if $\begin{cases} X_1 \sim Ga(x_1\,|\,a,b_1) \\ X_2 \sim Ga(x_2\,|\,a,b_2) \end{cases}$ then $Y = \dfrac{X_1}{X_1 + X_2} \sim Be(y\,|\,b_1,b_2)$

Generalization to n dimensions

… trivial but …

$$F(x_1, x_2, \ldots, x_n) = F_n(x_n | x_{n-1}, \ldots, x_1)\, F_{n-1}(x_{n-1} | x_{n-2}, \ldots, x_1) \cdots F_2(x_2 | x_1) F_1(x_1)$$

and

$$p(x_1, x_2, \ldots, x_n) = p_n(x_n | x_{n-1}, \ldots, x_1)\, p_{n-1}(x_{n-1} | x_{n-2}, \ldots, x_1) \cdots p_2(x_2 | x_1)\, p_1(x_1)$$

if X_k ***are independent***

$$p(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} p_i(x_i)$$

$$F(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} F_i(x_i)$$

… but if not … ***there are*** $n!$ ***ways to decompose the pdf and some are easier to sample than others***

Example: $$(X,Y) \sim p(x,y) = 2\exp\left(-\frac{x}{y}\right) \quad ; \quad x \in (0,\infty);\ y \in (0,1]$$

Marginal Densities:

$$p_y(y) = \int_0^\infty p(x,y)\,dx = 2y \qquad p_x(x) = \int_0^1 p(x,y)\,dy = 2x\int_x^\infty \frac{e^{-u}}{u^2}\,du$$

$$p(x,y) \neq p_x(x)\,p_y(y) \longrightarrow \textit{not independent}$$

Conditional Densities:

$$p(x,y) = p(x\,|\,y)\,p_y(y)$$

$$F_y(y) = y^2$$

$$p(x|y) = \frac{p(x,y)}{p_y(y)} = \frac{\exp\left(-\frac{x}{y}\right)}{y}$$

$$F_x(x|y) = 1 - \exp\left(-\frac{x}{y}\right)$$

easy

$$p(x,y) = p(y\,|\,x)\,p_x(x)$$

difficult

Properties:

Direct and efficient in the sense that

one $u_i \sim Un(u|0,1)$ ⇒ *one* $x_i \sim p(x|\theta)$

Useful for many distributions of interest

(Exponential, Cauchy, Weibull, Logistic,…)

…but in general, $F(x|\theta)$ *difficult to invert*

numeric approximations are slow,…

Method 2: Acceptance-Rejection

"Hit-Miss"
J. Von Neuman, 1951

Sample $X \sim p(x|\theta)$

$$X \in \Omega_X = [a,b]$$

$$0 \le p(x|\theta) \le \max_x \{p(x|\theta)\}$$

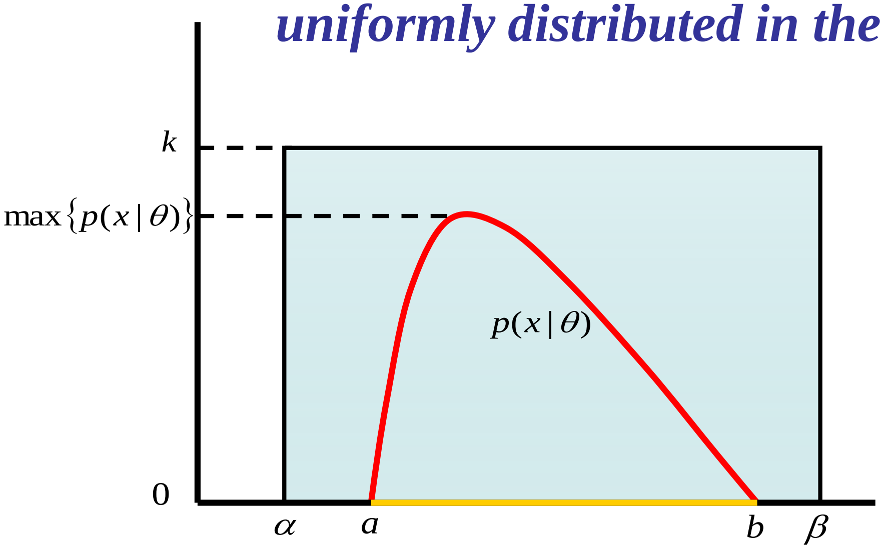
1) Enclose the pairs $(x, y = p(x|\theta))$ ***in a domain*** $\Lambda = [\alpha,\beta] \otimes [0,k]$

(not necessarily hypercube)

such that: $[a,b] \subseteq [\alpha,\beta]$ ***and*** $[0, \max\{p(x|\theta)\}] \subseteq [0,k]$

2) Consider a two-dimensional random quantity $\boldsymbol{Z} = (Z_1, Z_2)$

uniformly distributed in the domain $\Lambda = [\alpha,\beta] \otimes [0,k]$ ***; that is:***



$$g(z_1, z_2 \mid \phi)\, dz_1 dz_2 = \frac{dz_1}{\beta - \alpha} \frac{dz_2}{k}$$

3) Which is the conditional distribution

$P(Z_1 \le x | Z_2 \le p(x|\theta))$ ***?***

$$P(Z_1 \leq x | Z_2 \leq p(x|\theta)) = \frac{P(Z_1 \leq x, Z_2 \leq p(x|\theta))}{P_2(Z_2 \leq p(x|\theta))} =$$

$$= \frac{\int_\alpha^x dz_1 \int_0^{p(z_1|\theta)} g(z_1, z_2 | \phi)\, dz_2}{\int_\alpha^\beta dz_1 \int_0^{p(z_1|\theta)} g(z_1, z_2 | \phi)\, dz_2} = \frac{\int_\alpha^x p(z_1|\theta)\, \mathrm{I}_{[a,b]}(z_1)\, dz_1}{\int_\alpha^\beta p(z_1|\theta)\, \mathrm{I}_{[a,b]}(z_1)\, dz_1} = F(x)$$

Algorithm

n)

i) Sample $Z_1 \rightarrow z_{1i} \sim Un(z_1|\alpha, \beta)$

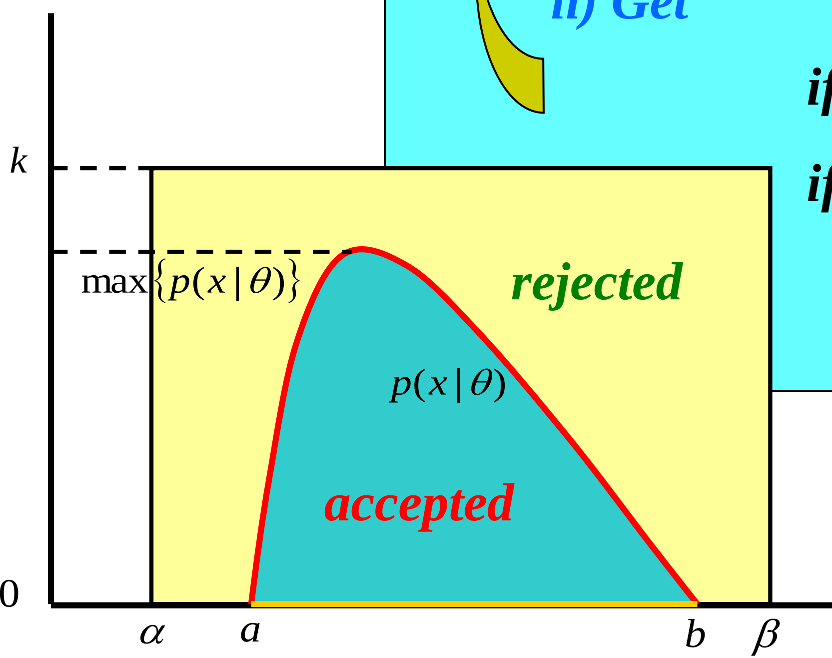
and $Z_2 \rightarrow z_{2i} \sim Un(z_2|0, k)$

ii) Get $X \sim p(x|\theta)$

if $z_{2i} \leq p(z_{1i}|\theta) \rightarrow$ ***accept*** $x_i = z_{1i}$

if $z_{2i} > p(z_{1i}|\theta) \rightarrow$ ***reject*** $x_i = z_{1i}$ ***and go back to i)***

$$\{x_1, x_2, \ldots x_n\} \sim p(x|\theta)$$



Example: $X \sim Be(x|a,b)$ *(pedagogical; low efficiency)*

density: $$p(x \mid a,b) \propto x^{a-1}(1-x)^{b-1} \quad ; \quad x \in [0,1]$$

(normalisation: $\int_0^1 p(x|a,b)\,dx = Be(a,b)$ ***…not needed)***

Covering adjusted for maximum efficiency:

$$\max\{p(x|a,b)\} = p\left(x_{\max} = \frac{a-1}{a+b-2} \,\middle|\, a,b\right) = \frac{(a-1)^{a-1}(b-1)^{b-1}}{(a+b-2)^{a+b-2}}$$

(x,y) ***generated in the domain***

$$\Lambda = [\alpha,\beta]\otimes[0,k] = [0,1]\otimes[0,\max\{p(x|a,b)\}]$$

Algorithm

n)

i) Sample $x_i \sim Un(x|0,1)$

$y_i \sim Un(y|0, max\{p(x|a,b)\})$

ii) Accept-Reject

if $y_i \leq p(x_i|a,b)$ → ***accept*** x_i

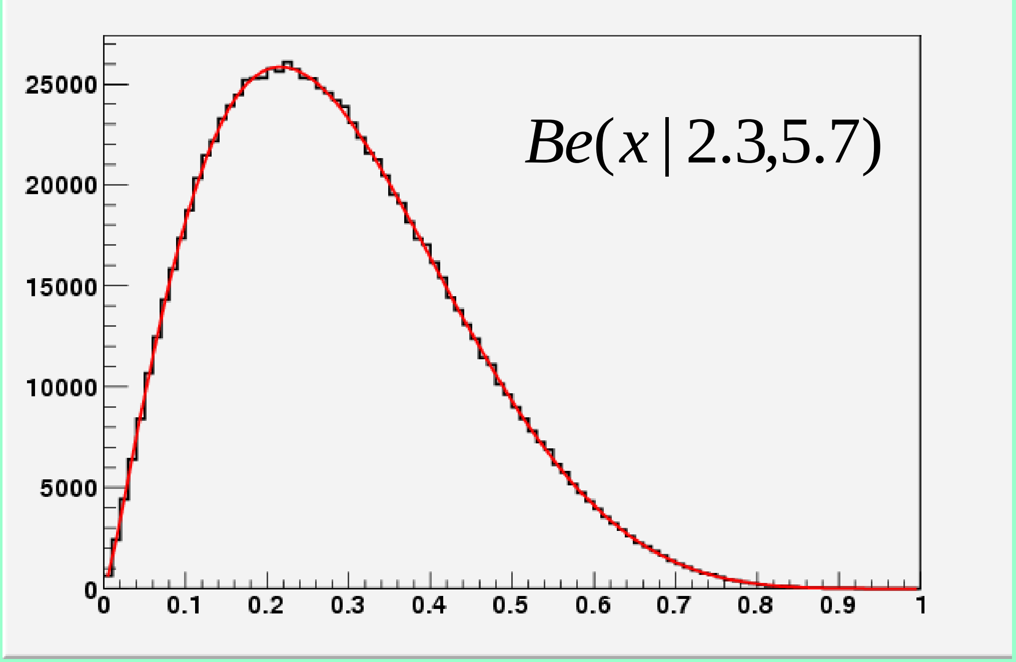
if $y_i > p(x_i|a,b)$ → ***reject*** x_i ***and go to i)***

$a = 2.3\,;\, b = 5.7$ $\Lambda = [0,1] \times [0, \max\{p(x|a,b)\}]$ $\mu(\Lambda) = \max\{p(x|a,b)\}$

$$\left.\begin{array}{l} n_{gen} = 2588191 \\ n_{acc} = 1000000 \end{array}\right\} e_{ff} = \frac{n_{acc}}{n_{gen}} = 0.3864$$

$$\int_{\Omega_X} p(x|\theta)dx = \mu(\Lambda)\left(\frac{n_{acc}}{n_{gen}}\right) = 0.016791 \; (\pm 0.000013)$$

$$Be(2.3,5.7) = 0.01678944$$



… weighted events ???

We can do the rejection as follows:

i) Assign a weight to each generated pair (x_i, y_i) $\qquad w_i = \dfrac{p(x_i|\theta)}{k}$

$$[0, \max\{p(x|\theta)\}] \subseteq [0,k] \;\longrightarrow\; 0 \le w_i \le 1$$

(if we know $\max\{p(x|\theta)\}$ $k = \max\{p(x|\theta)\}$ **)**

ii) Acceptance-rejection

$$u_i \sim Un(u|0,1) \begin{cases} \text{accept} \quad x_i \quad \text{if } u_i \le w_i \\ \text{reject} \quad x_i \quad \text{if } u_i > w_i \end{cases}$$

- **Obviously, events with a larger weight are more likely to be accepted**
- **After step ii), all weights are:** $w_i = \begin{cases} 0 & \text{if rejected} \\ 1 & \text{if accepted} \end{cases}$
- **Sometimes, it is interesting not to apply step ii) and keep the weights**

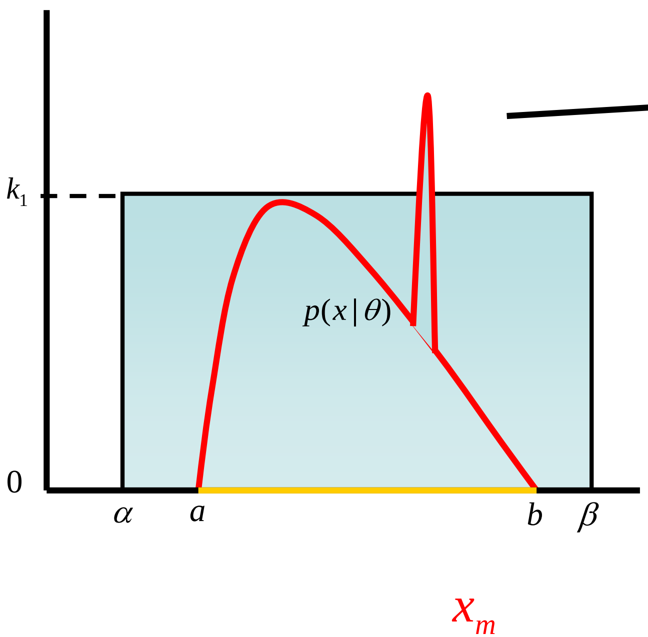
$$\int_{\Omega_x} g(x)\, p(x|\theta)\, dx = \langle g(x) \rangle_p = \frac{1}{w} \sum_{i=1}^{n} g(x_i)\, w_i \qquad ; \qquad w = \sum_{i=1}^{n} w_i$$

And suddenly…

➢ ***Many times we do not know*** $\max\{p(x|\theta)\}$ ***and start with an initial guess*** k_1

➢ ***After having generated*** N_t ***events in the domain*** $\Lambda = [\alpha, \beta] \otimes [0, k_1]$

we find a value $x_m \mid p(x_m \mid \theta) > k_1$



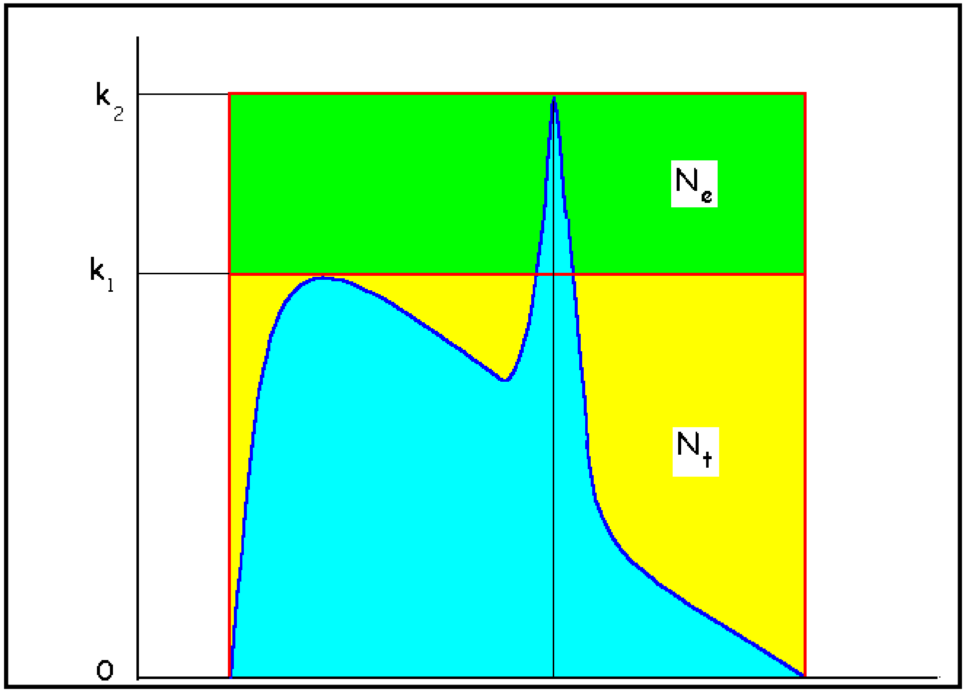
incorrect estimation of

$$k = \max\{p(x|\theta)\}$$

Don't throw away all generated events…

The pairs (x, y) ***have been generated in*** $\Lambda = [\alpha, \beta] \otimes [0, k_1]$

with constant density so $\frac{N_t}{(\alpha - \beta)\, k_1} = \frac{N_t + N_e}{(\alpha - \beta)\, k_2}$ ***with*** $k_2 = p(x_m \,|\, \theta) > k_1$



we have to generate:

$$N_e = N_t \left(\frac{k_2}{k_1} - 1 \right)$$

additional pairs (x, y)

… in the domain $\Delta = [\alpha, \beta] \otimes [k_1, k_2]$

with the pseudo-distribution

$$g(x|\theta) = [p(x|\theta) - k_1]\, \mathbf{1}_{p(x|\theta) > k_1}(x)$$

… and proceed with $\Lambda_2 = [\alpha, \beta] \otimes [0, k_2]$

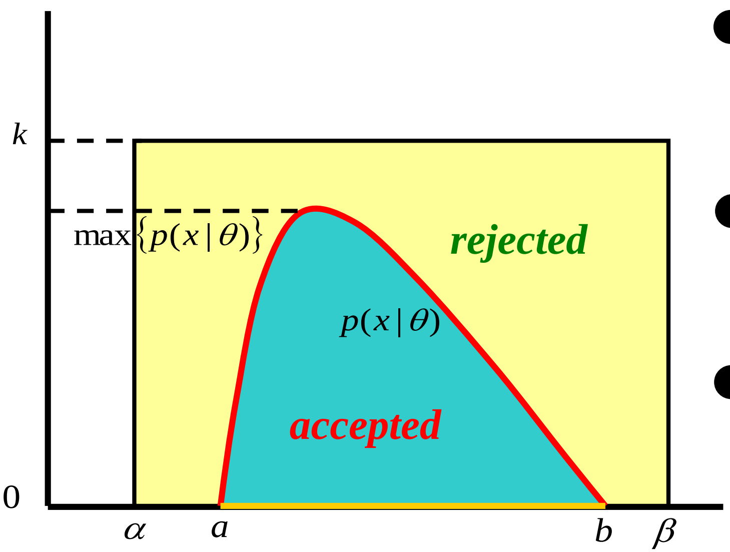
Properties:

Easy to implement and generalize to n dimensions

Efficiency *depends on the covering domain* $\Lambda = [\alpha, \beta] \otimes [0, k]$

$$e_f = \frac{\text{area under } p(x|\theta)}{\text{area covering domain}} = \frac{\#\text{ accepted events}}{\#\text{ generated events}} \leq 1$$

- $$\int_{\Omega_X} p(x|\theta)dx = \mu(\Lambda)\, e_f = \mu(\Lambda)\frac{N_{acc}}{N_{gen}}$$
- $e_f = 1$ *is equivalent to the Inverse Transform*
- *The better adjusted the covering is, the higher the efficiency*



Straight forward generalization to ***n*** *dimensions*

Algorithm:

i) Generate n+1 $Un(x|\Omega)$ ***distributed random quantities***

$$(x_i^{1)}, x_i^{2)}, \ldots, x_i^{n)}; y_i) \sim Un(x^{1)}|\Omega_1) \otimes Un(x^{2)}|\Omega_2) \otimes \ldots$$
$$\ldots \otimes Un(x^{n)}|\Omega_n) \otimes Un(y|0,k)$$

k)

ii) Accept or reject

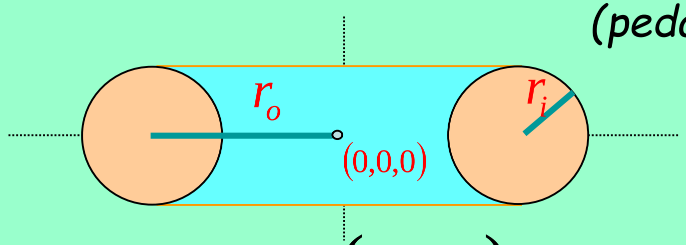
accept $(x_i^{1)}, x_i^{2)}, \ldots, x_i^{n)})$ ***if***

$$y_i \leq p(x_i^{1)}, x_i^{2)}, \ldots, x_i^{n)} \mid \theta)$$

reject if not and go back to i)

(pedagogical; low efficiency)

3D Example:



Points inside a torus of radius (r_i, r_o) centred at $(0,0,0)$

Cover the torus by a parallelepiped:

$$[-(r_i+r_o),(r_i+r_o)]\otimes[-(r_i+r_o),(r_i+r_o)]\otimes[-r_i,r_i]$$

Algorithm:

i) Generate

$$x_i \sim Un(x \,|\, -(r_i+r_o),(r_i+r_o))$$
$$y_i \sim Un(y \,|\, -(r_i+r_o),(r_i+r_o))$$
$$z_i \sim Un(y \,|\, -r_i,r_i)$$

n)

ii) Reject if

$$x_i^2 + y_i^2 > (r_i + r_o)^2$$
or $$x_i^2 + y_i^2 < (r_i - r_o)^2$$
or $$z_i^2 + (r_0 - (x_i^2 + y_i^2)^{1/2})^2 > r_i^2$$

And go back to i)

otherwise accept (x_i, y_i, z_i)

$$\{(x_1,y_1,z_1),(x_2,y_2,z_2),\ldots,(x_n,y_n,z_n)\}\in T(r_o,r_i,0,0,0)$$

$r_i = 1; \quad r_o = 3$

$$e_f = \frac{\text{volume toroid}}{\text{volume parallelepiped}} = \frac{2\pi^2 r_0 r_i^2}{8 r_i (r_0 + r_i)^2}$$

$N_{generated} = 10{,}786$

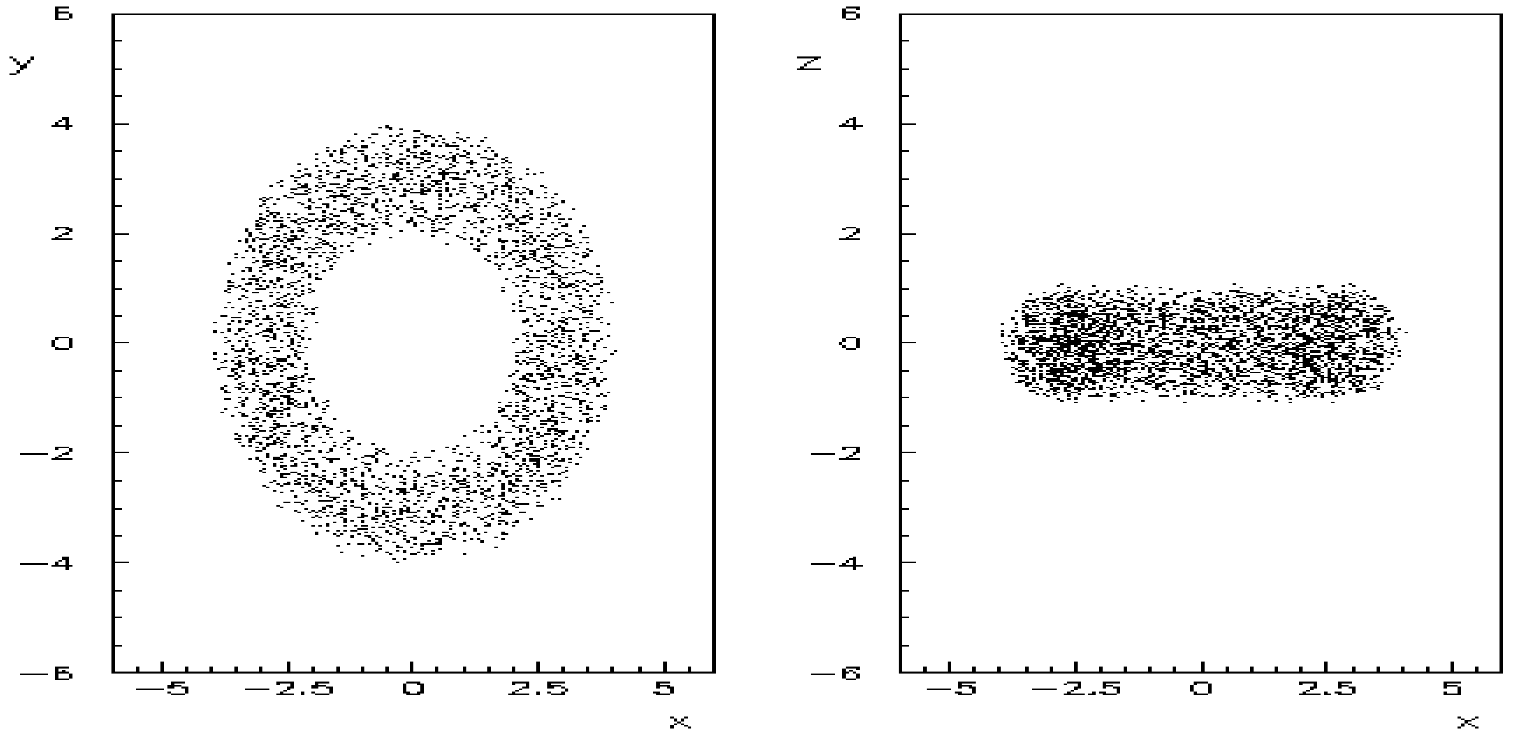
$N_{accepted} = 5{,}000$

$$\frac{N_{accepted}}{N_{generated}} = 0.4636$$

knowing that: $V_{parallelepiped} = 128$

we can estimate: $V_{toroid} = e_f \, V_{parallelepipedo} \approx 59{,}34 \pm 0{,}61$

$V_{toroid} = 59{,}218$

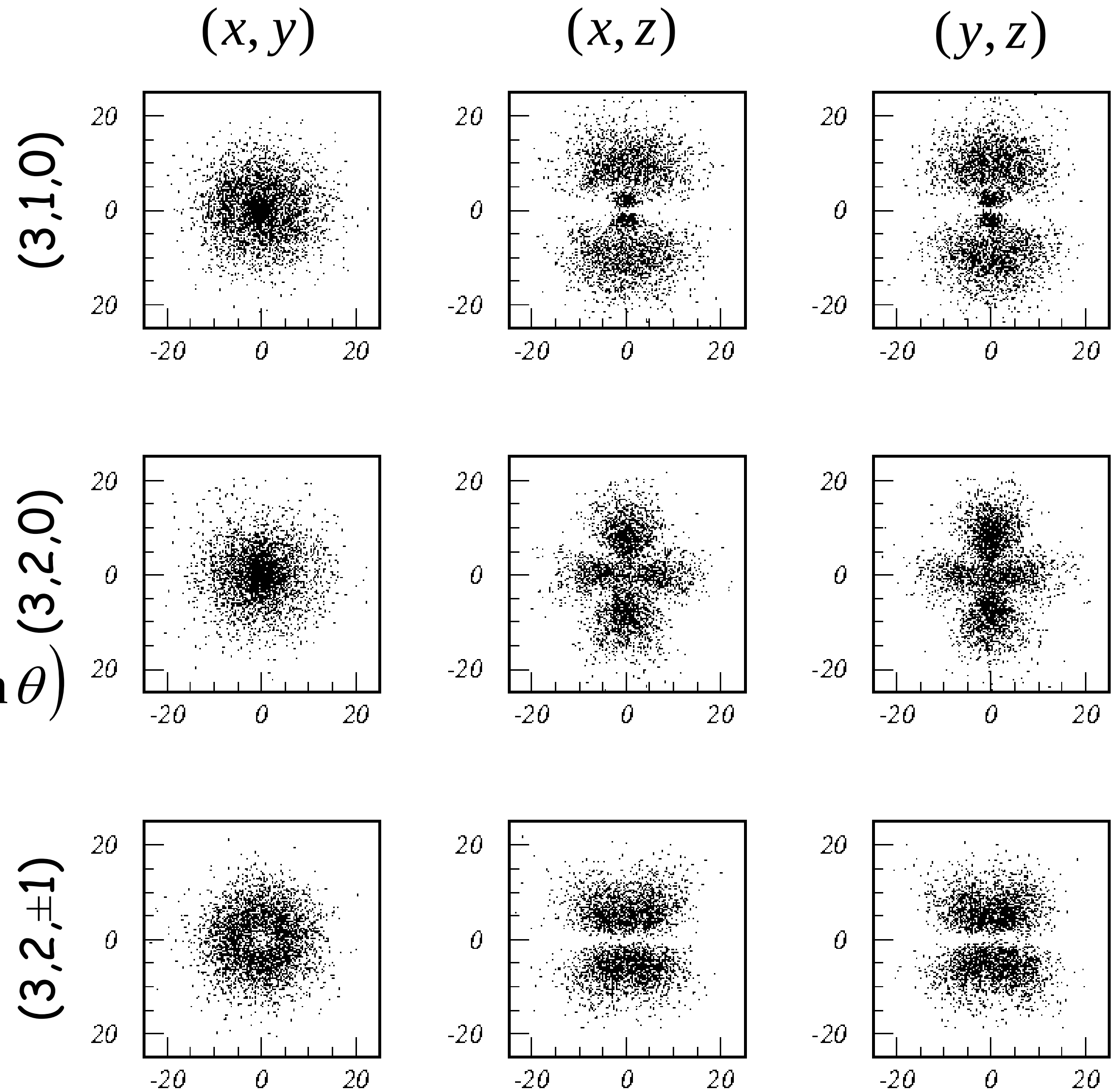


Problem 3D:

Sampling of Hidrogen atom wave functions

$$\Psi_{n,l,m}(r,\theta,\phi) = R_{n,l}(r) Y_{l,m}(\theta,\phi)$$

$$P(r,\theta,\phi | n,l,m) = R_{n,l}(r)^2 \left|Y_{l,m}(\theta,\phi)\right|^2 \left(r^2 \sin\theta\right)$$



Evaluate the energy using Virial Theorem

$$\langle T \rangle = -\tfrac{1}{2}\langle V \rangle$$

$$\langle E \rangle_n = \tfrac{1}{2}\langle V \rangle_n$$

$$V(r) = -\frac{e^2}{4\pi\varepsilon_0}\frac{1}{r}$$

Problems with low efficiency

(problem for convergence of estimations,…)

Example in n-dimensions:

n-dimensional sphere of radius $r = 1$ centred at $(0,0,\ldots,0)$

i) Sample $\left(x_i^{1)},\ldots,x_i^{n)}\right) \sim Un\left(x^{1)}\middle|-1,1\right)\otimes\ldots\oplus Un\left(x^{n)}\middle|-1,1\right)$

ii) Acceptance-rejection: $y^2 = \left(x_i^{1)}\right)^2 + \left(x_i^{2)}\right)^2 + \ldots + \left(x_i^{n)}\right)^2$

if $y^2 \le 1$ accept $\left(x_i^{1)},\ldots,x_i^{n)}\right)$ as inner point of the sphere

or $\left(\frac{x_i^{1)}}{y},\ldots,\frac{x_i^{n)}}{y}\right)$ over the surface

if $y^2 > 1$ reject the n-tuple and go back to i)

$$e_f = \frac{V(S_n)}{V(H_n)} = \frac{2\pi^{n/2}/n\Gamma(n/2)}{2^n} \xrightarrow{n\to\infty} 0$$

$e_f(n=4) \approx 31\%$

$e_f(n=5) \approx 16\%$

Why do we have low efficiency?

Most of the samplings are done in regions of the sample space that have low probability

Example: ***Sampling of***

$$X \sim p(x|\theta) \sim e^{-\theta x} \quad ; x \in [0,1]$$

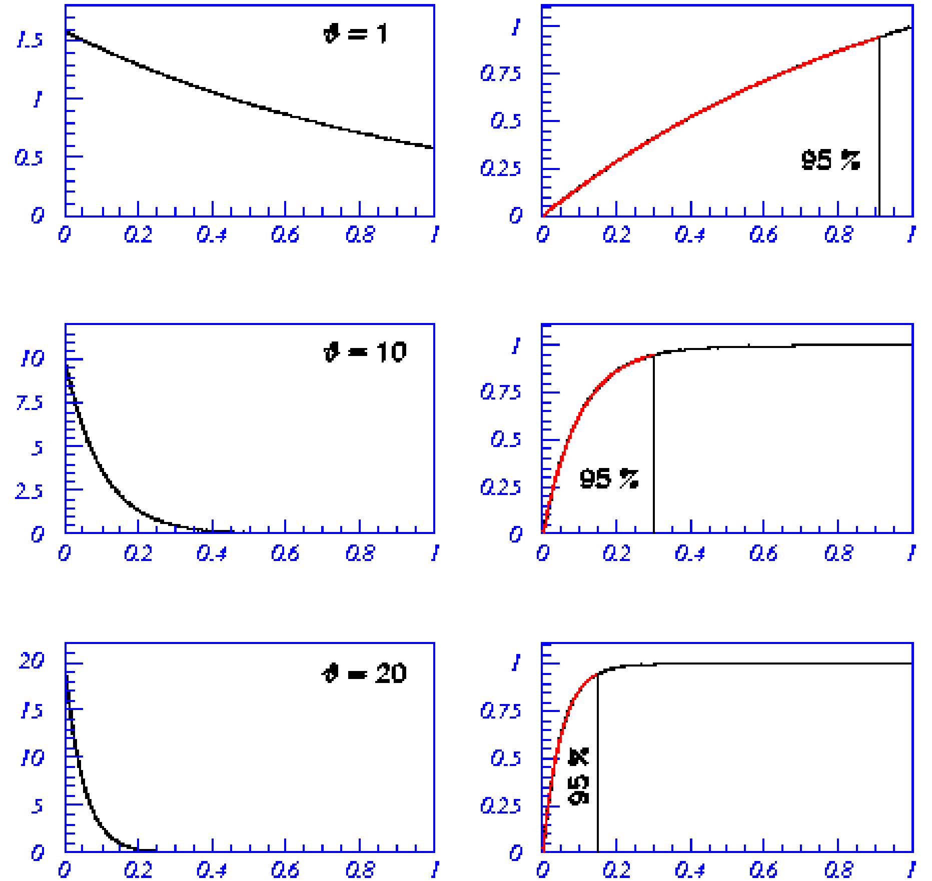
Generate values of X *uniformly in* $\Omega_x = [0,1]$

… do a more clever sampling…

usually used in combination with *"Importance Sampling"*

$$p(x|\theta) = \frac{\theta}{1-e^{-\theta}} e^{-\vartheta x}$$

$$F(x|\theta) = \frac{1-e^{-\theta x}}{1-e^{-\theta}}$$



Example of inefficiencies: Usual problem in Particle Physics…

$$d\sigma = \frac{(2\pi)^4}{F} \, | iM_{fi} |^2 \, d\Phi_n(p_1, \dots p_n \mid p)$$

i) Sample phase-space variables $d\Phi_n(p_1, \dots p_n \mid p)$

ii) Acceptance-rejection on $| iM_{fi} |^2$ $\longrightarrow$ *sample of events with dynamics of the process*

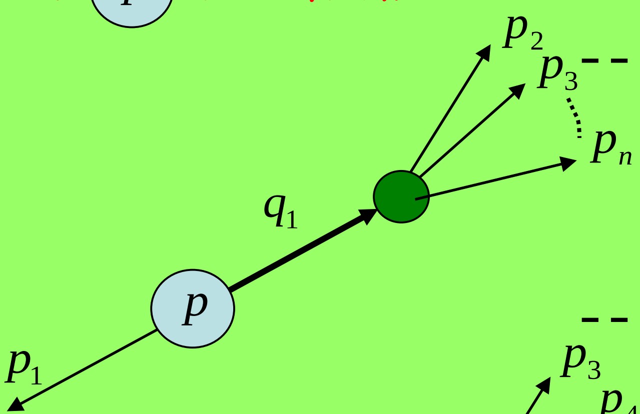
iii) Estimate cross-section $E[| iM_{fi} |^2]$

"Naïve" and simple generation of n-body phase-space

$$d\Phi_n(p_1, \dots, p_n \mid p) = d\Phi_j(p_1, \dots, p_j \mid q) \times d\Phi_{n+1-j}(q, p_{j+1}, \dots, p_n \mid p) (2\pi)^3 dm_q^{\,2}$$

In (p) *rest-frame:*

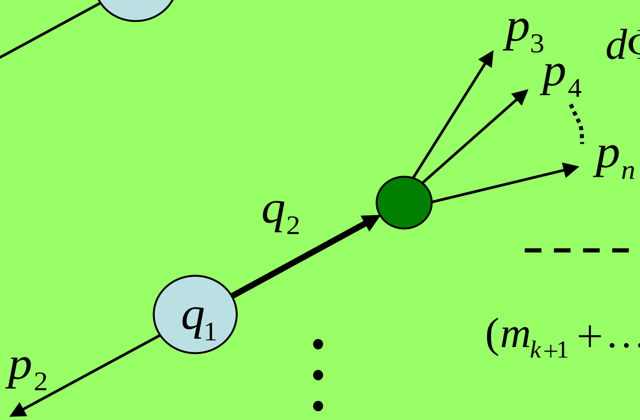
$$S_k = \sum_{i=k}^{n} m_i \qquad m_{q_0} = M_p; \quad m_{q_{n-1}} = m_n$$



$$d\Phi_n(p_1,\dots,p_n \mid p) \propto d\Phi_{n-1}(p_2,\dots,p_n \mid q_1) \times \underbrace{d\Phi_2(q_1, p_1 \mid p)}\, dm_{q_1}^{\;2}$$

$$(m_2 + \dots + m_n) \le m_{q_1} \le M_p - m_1$$

$$u_1 \sim Un(0,1) \longrightarrow m_{q_1} = S_2 + u_1(m_{q_0} - S_1)$$



$$d\Phi_{n-1}(p_2,\dots,p_n \mid q_1) \propto d\Phi_{n-2}(p_3,\dots,p_n \mid q_2) \times \underbrace{d\Phi_2(q_2, p_2 \mid q_1)}\, dm_{q_2}^{\;2}$$

$$(m_3 + \dots + m_n) \le m_{q_2} \le m_{q_1} - m_2$$

$$u_2 \sim Un(0,1) \longrightarrow m_{q_2} = S_3 + u_2(m_{q_1} - S_2)$$

⋮

$$(m_{k+1} + \dots + m_n) \le m_{q_k} \le m_{q_{k-1}} - m_k$$

$$u_k \sim Un(0,1) \longrightarrow m_{q_k} = S_{k+1} + u_k(m_{q_{k-1}} - S_k)$$

for the $k = 1,\dots,n-2$ *intermediate states*

and then weight each event…

$$d\Phi_2(p_1, p_2 \mid p) = \frac{1}{4\,(2\pi)^6} \frac{|\vec{p}_1|}{m_p} d\Omega_1$$

$$W_T = \prod_{k=0}^{n-2} \frac{\lambda^{1/2}(m_{q_k}, m_{n-k}, m_{q_{k+1}})}{m_{q_k}}$$

$$\lambda(x,y,z) = [x^2 - (y-z)^2][x^2 - (y+z)^2]$$

+ Lorentz boosts to overall centre of mass system

…but $|iM_{fi}|^2$ ***, cuts,…!***

usually inefficient ("very")

Method 3: Importance Sampling

Sample $X \sim p(x\,|\,\theta) \geq 0 \quad ; \forall x \in \Omega_X$

1) Express $p(x\,|\,\theta) = g(x\,|\,\theta_1)\, h(x\,|\,\theta_2) \geq 0 \quad ; \forall x \in \Omega_X$

i) $g(x|\theta_1) \geq 0 ; \quad h(x|\theta_2) \geq 0 \quad ; \forall x \in \Omega_X$

ii) $h(x|\theta_2)$ ***probability density***

$$p(x\,|\,\theta)dx = g(x\,|\,\theta_1)\, h(x\,|\,\theta_2)\, dx = g(x\,|\,\theta_1)\, dH(x\,|\,\theta_2)$$

In particular, take a convenient $h(x\,|\,\theta_2) > 0$ ***(easiness) and define*** $g(x\,|\,\theta_1) = \dfrac{p(x\,|\,\theta)}{h(x\,|\,\theta_2)}$

2) Consider a sampling of $X \sim h(x\,|\,\theta_2)$

and apply the acceptance-rejection algorithm to $g(x|\theta_1) \geq 0$

How are the accepted values distributed?

$$P\left(X \le x \mid Y \le g(x\mid\theta_1)\right) = \frac{P\left(X \le x, Y \le g(x\mid\theta_1)\right)}{P\left(Y \le g(x\mid\theta_1)\right)} = \frac{\frac{1}{\Lambda}\int_{-\infty}^{x} h(x'\mid\theta_2)dx' \int_{-\infty}^{g(x'\mid\theta_1)} dy}{\frac{1}{\Lambda}\int_{-\infty}^{\infty} h(x'\mid\theta_2)dx' \int_{-\infty}^{g(x'\mid\theta_1)} dy}$$

$$\Lambda = \Omega_X \times \Omega_Y$$

$$\mu(\Lambda) = \mu(\Omega_X)\mu(\Omega_Y)$$

$$= \frac{\int_{-\infty}^{x} h(x'|\theta_2)g(x'|\theta_1)dx'}{\int_{-\infty}^{\infty} h(x'|\theta_2)g(x'|\theta_1)dx'} = \frac{\int_{-\infty}^{x} p(x'|\theta)dx'}{\int_{-\infty}^{\infty} p(x'|\theta)dx'} = F(x\mid\theta)$$

Algorithm

i) Sample $x_i \sim h(x|\theta_2)$

ii) Apply Acceptance-Rejection to $g(x|\theta_1)$

$\{x_1, x_2, \ldots x_n\}$ ***sample drawn from*** $p(x|\theta) = g(x|\theta_1)\, h(x|\theta_2)$

…∞ manners to choose $h(x|\theta_2)$

- ***Easy to invert*** ***so we can apply the Inverse Transform Method***

Easiest choice: $h(x|\theta_2) = \frac{1}{\mu(\Delta)} \mathbf{1}_{\Delta}(x)$

… but this is just the acceptance – rejection procedure

- ***We would like to choose*** $h(x|\theta_2)$ ***such that,***
  instead of sampling uniformly the whole domain Ω_X ***,***
  we sample with higher probability from those regions where
  $p(x|\theta)$ ***is "more important"***

➡ ***Take*** $h(x|\theta_2)$ ***as "close" as possible to*** $p(x|\theta)$

➡ $g(x|\theta_1) = \frac{p(x|\theta)}{h(x|\theta_2)}$ ***will be "as flat as possible" ("flatter" than*** $p(x|\theta)$ ***)***
and we shall have a higher efficiency when applying the acceptance-rejection algorithm

Example: $X \sim Be(x|a,b)$ $\quad a = 2.3\,;\quad b = 5.7$

density: $$p(x\,|\,a,b) \propto x^{a-1}(1-x)^{b-1}\,\mathbf{1}_{[0,1]}(x)$$

1) We did "standard" acceptance-rejection

2) "Importance Sampling"

$$m = [a] = 2 \rightarrow a = m + \delta_1\,;\quad \delta_1 = 0.3$$

$$n = [b] = 5 \rightarrow b = n + \delta_2\,;\quad \delta_2 = 0.7$$

$$p(x\,|\,a,b) \propto \frac{x^{a-1}(1-x)^{b-1}}{x^{m-1}(1-x)^{n-1}}\,x^{m-1}(1-x)^{n-1}$$

$$\propto [x^{\delta_1}(1-x)^{\delta_2}][x^{m-1}(1-x)^{n-1}]$$

$$dP(x\,|\,a,b) \sim [x^{\delta_1}(1-x)^{\delta_2}]\,dBe(x\,|\,m,n)$$

$$p(x \mid a,b) \propto [x^{\delta_1}(1-x)^{\delta_2}][x^{m-1}(1-x)^{n-1}]$$

2.1) Generate $X \sim Be(x|m,n)$

$$\left.\begin{aligned} U_k &\sim Un(x|0,1) \\ Y_m &= -\log\left(\prod_{k=1}^{m} U_k\right) \sim Ga(x|1,m) \end{aligned}\right\} Z = \frac{Y_m}{Y_m + Y_n} \sim Be(x|m,n)$$

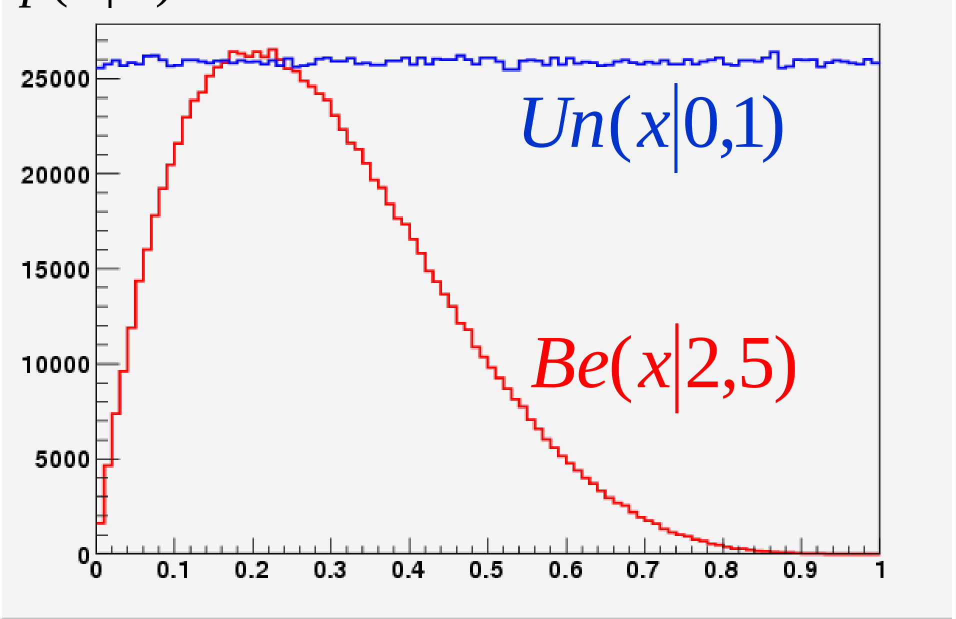
2.2) Acceptance-Rejection on

$$g(x) = x^{\delta_1}(1-x)^{\delta_2} \qquad 0<\delta_1,\delta_2<1$$

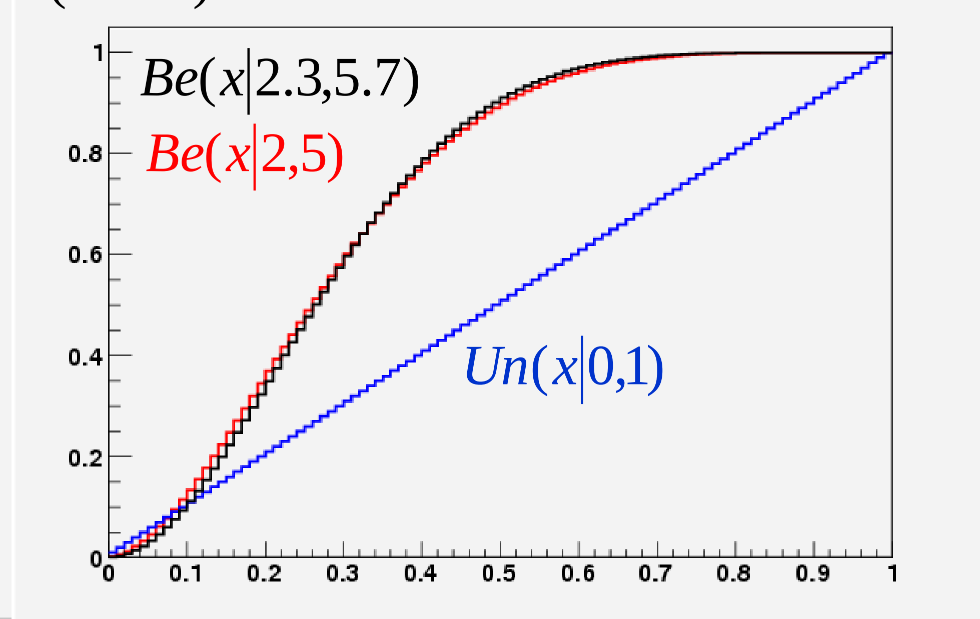
$g(x)$ **concave on** $x \in [0,1]$

$$\max_x\{g(x)\} = \frac{\delta_1^{\delta_1}\delta_2^{\delta_2}}{(\delta_1+\delta_2)^{\delta_1+\delta_2}}$$

$p(x \mid \theta)$



$P(X \leq x)$



acceptance-rejection

$$n_{gen} = 2588191$$

$$n_{acc} = 1000000$$

$$eff = 0.3864$$

$$Be^{*}(2.3,5.7) = eff \times f_{\max} = $$
$$= 0.016791$$
$$\pm 0.000013$$

$$Be(2.3,5.7) = 0.01678944$$

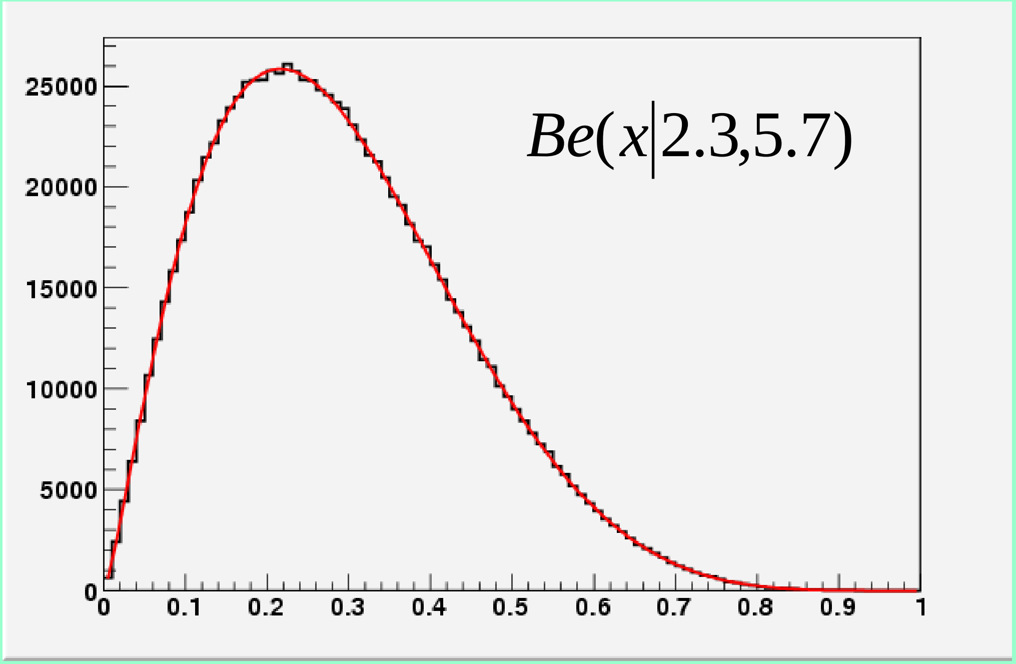
Importance Sampling

$$n_{gen} = 1077712$$

$$n_{acc} = 1000000$$

$$eff = 0.9279$$

$$Be^{*}(2.3,5.7) = eff \times f_{\max} \times Be(2,5) =$$
$$= 0.0167912$$
$$\pm 0.0000045$$



Method 4: Decomposition (… trick)

Idea: Decompose the pdf as a sum of simpler densities

$$p(x\,|\,\theta) = \sum_{i=1}^{m} a_i \; p_i(x\,|\,\theta)$$

$$p_i(x\,|\,\theta) \geq 0 \qquad a_i > 0\,; \quad \sum_i a_i = 1 \qquad \begin{cases} \forall i = 1, m \\ \forall x \in \Omega_x \end{cases}$$

Normalisation: $\displaystyle\int_{-\infty}^{+\infty} p(x\,|\,\theta)\,dx = \sum_{i=1}^{m} a_i \int_{-\infty}^{+\infty} p_i(x\,|\,\theta)\,dx = \sum_{i=1}^{m} a_i = 1$

Each density $p_i(x\,|\,\theta)$ has a relative weight a_i

⇒ Sample $X \sim p_i(x\,|\,\theta)$ with probability $p_i = a_i$

… thus, more often from those with larger weight

Algorithm

n)

i) Generate $u_i \sim Un(u|0,1)$ to select $p_i(x\,|\,\theta)$ with probability a_i

ii) Sample $X \sim p_i(x|\theta)$

Note:

Sometimes the pdf $p_i(x\,|\,\theta)$ can not be easily integrated

… normalisation unknown → generate from $f_i(x|\theta) \propto p_i(x|\theta)$

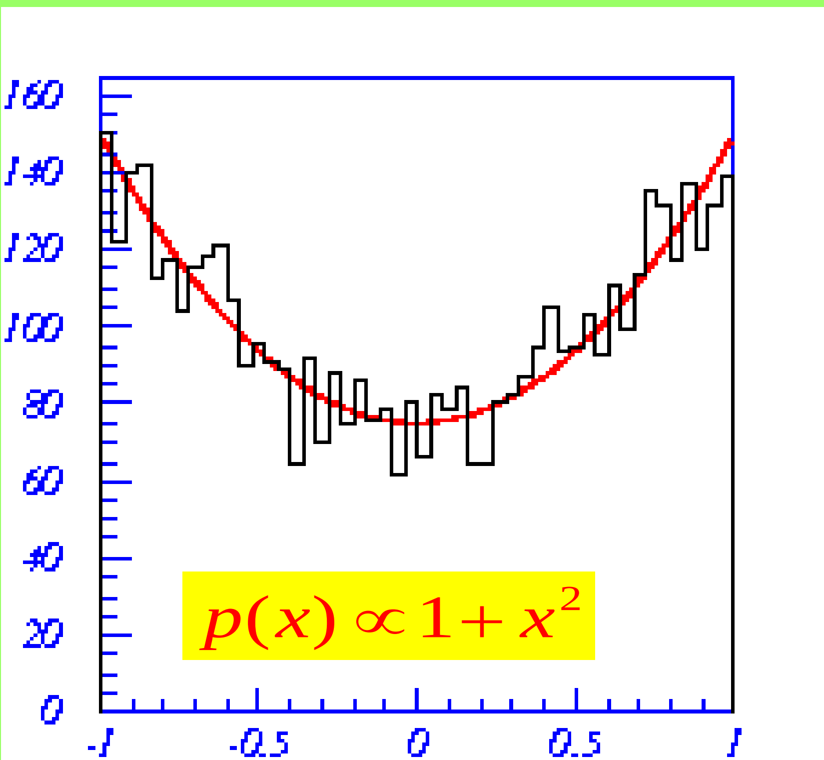
Evaluate the normalisation integrals I_i during the generation process (numeric integration) and assign eventually a weight $w_i = a_i I_i$ to each event

$$p(x\,|\,\theta) = \sum_{i=1}^{m} (a_i I_i) \frac{f_i(x|\theta)}{I_i} = \sum_{i=1}^{m} (a_i I_i) p_i(x|\theta)$$

Example: $X \sim p(x) = \frac{3}{8}\left(1+x^2\right) \quad ; x \in [-1,1]$

$$g_1(x) = 1, \quad g_2(x) = x^2 \quad \xrightarrow{\text{normalisation}} \quad p_1(x) = \tfrac{1}{2}, \quad p_2(x) = \tfrac{3x^2}{2}$$

$$p(x) = \frac{3}{4}\, p_1(x) + \frac{1}{4}\, p_2(x)$$



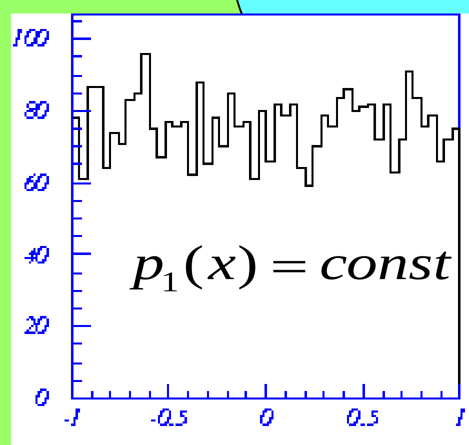
Algorithm:

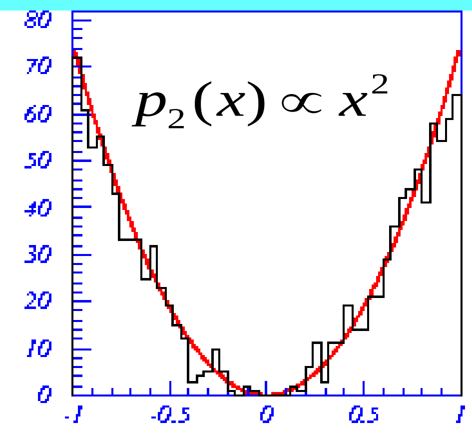
i) Generate $u_i \sim Un(u|0,1)$

$u_i \le \frac{3}{4}$	$\frac{3}{4} < u_i \le 1$

ii) Sample from:

$p_1(x) = const$	$p_2(x) \propto x^2$
75% of the times	*15% of the times*





Generalisation:

Extend $\sum_{i=1}^{m} g_i(x)$ to a continuous family $\int g(y)\, dy$ and consider

$p(x \mid \theta)$ as a marginal density:

$$p(x \mid \theta) = \int_{\Omega_y} p(x, y \mid \phi)\, dy = \int_{\Omega_y} p(x \mid y, \xi)\; p(y \mid \varphi)\, dy$$

Algorithm:

n) i) Sample $y_i \sim p(y \mid \varphi)$

ii) Sample $x_i \sim p(x \mid y_i, \xi)$

Structure in bayesian analysis $\quad p(\theta \mid x) \propto p(x \mid \theta)\, \pi(\theta)$

Experimental resolution $\quad p_{obs}(y \mid \theta, \lambda) = \int_{\Omega_Y} R(y \mid x, \theta)\; p_t(x \mid \lambda) dx$

$$p(x, y \mid \theta, \lambda) = R(y \mid x, \theta)\; p_t(x \mid \lambda)$$

The Normal Distribution

$$X \sim N(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$$

Standardisation $Z = \dfrac{X-\mu}{\sigma}$

$$Z \sim N(z|0,1) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \Longrightarrow X = \mu + Z\sigma$$

but...

i) Inverse Transform: $F(z)$ ***is not an elementary function***

entire function → ***series expansion convergent but inversion slow,…***

ii) Central Limit Theorem: $U_i \sim Un(u|0,1) \Longrightarrow Z = \sum_{i=1}^{12} U_i - 6$

$Z \in [-6,6]\,;\quad E[Z]=0\,;\quad V(Z)=1 \Longrightarrow Z \overset{approx}{\sim} N(z|0,1)$

iii) Acceptance - Rejection: ***not efficient although…***

iv) Importance Sampling: ***easy to find good approximations***

Sometimes, going to higher dimensions helps: in two dimensions…

Inverse Transform in two dimensions

(G.E.P.Box, M.E. Muller; 1958)

i) Consider two independent random quantities

$$X_1, X_2 \sim N(x|0,1) \qquad p(x_1, x_2) = \frac{1}{2\pi}\, e^{-\frac{1}{2}\left(x_1^2 + x_2^2\right)}$$

ii) Polar Transformation: $(X_1, X_2) \longrightarrow (R, \Theta) \in [0, \infty) \otimes [0, 2\pi)$

$$\left.\begin{aligned} X_1 &= R\cos\Theta \\ X_2 &= R\sin\Theta \end{aligned}\right\} \qquad p(r,\theta) = \frac{1}{2\pi}\, e^{-\frac{r^2}{2}}\, r = p(r)\, p(\theta)$$

(R, Θ) ***independent***

iii) Marginal Distributions:

$$F_r(r) = \int_0^{2\pi} p(r,\theta)\, d\theta = 1 - e^{-\frac{r^2}{2}}$$

$$F_\theta(\theta) = \int_0^{\infty} p(r,\theta)\, dr = \frac{\theta}{2\pi}$$

…trivial inversion

Algorithm

n)

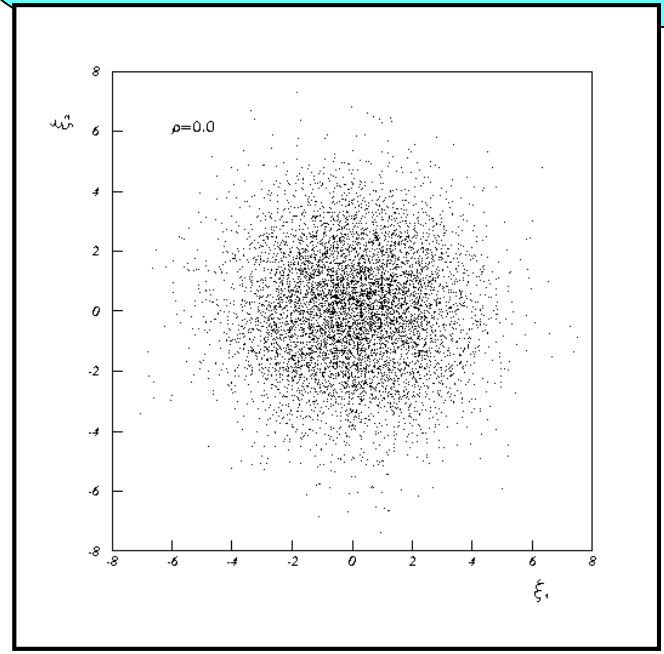
i) Generate $u_1, u_2 \sim Un(u|0,1)$

ii) Invert $F_r(r)$ and $F_\theta(\theta)$ to get $\begin{cases} r = \sqrt{-2\ln u_1} \\ \theta = 2\pi u_2 \end{cases}$

iii) Obtain $x_1, x_2 \sim N(x|0,1)$ as

$$x_1 = \sqrt{-2\ln(u_1)}\ \cos(2\pi u_1) \qquad x_2 = \sqrt{-2\ln(u_1)}\ \text{sen}(2\pi u_1)$$

(independent)



Example: 2-dimensional Normal random quantity

Sampling from Conditional Densities

$$(X_1, X_2) \sim N(x_1, x_2 | \mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$$

Standardisation

$$W_i = \frac{X_i - \mu_i}{\sigma_i}$$

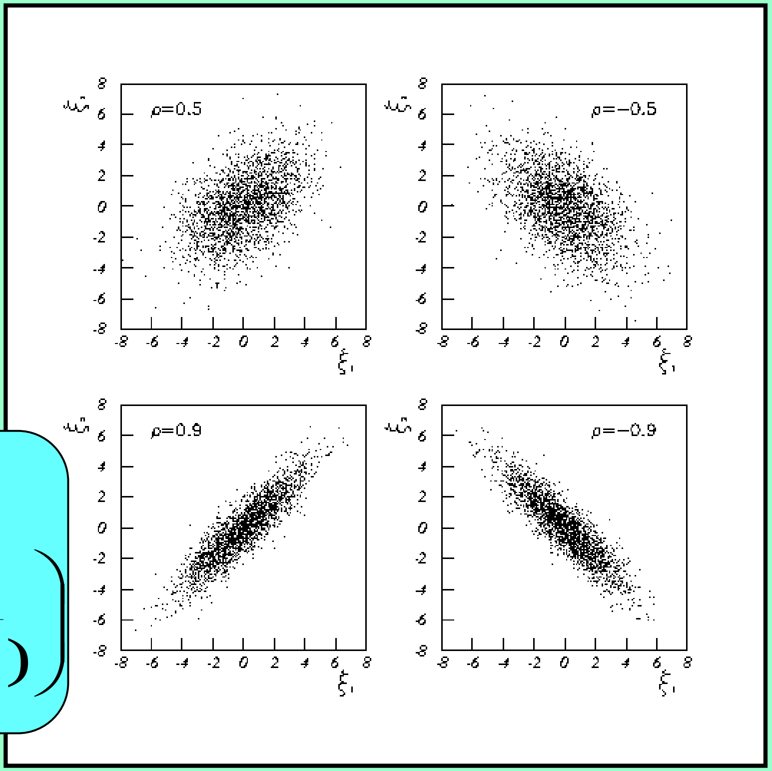
$$p(w_1, w_2 | \theta) = p(w_2 | w_1, \rho)\, p(w_1)$$

$$= N(w_2 | \rho\, w_1, (1-\rho^2)^{1/2})\; N(w_1 | 0,1)$$

$$Z_2 = \frac{W_2 - \rho W_1}{\sqrt{1-\rho^2}} \sim N(z_2 | 0,1) \qquad Z_1 (= W_1) \sim N(z_1 | 0,1)$$

i) Generate $z_1, z_2 \sim N(z | 0,1)$

ii) $$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} z_1 \sigma_1 \\ \sigma_2 (z_1 \rho + z_2 \sqrt{1-\rho^2}) \end{pmatrix}$$



n-dimensional Normal Density

… different ways

Factorisation Theorem: *(Cholesky)*

if $\mathbf{V} \in \Re^{n \times n}$ *symmetric and positive defined*

$\exists\, \mathbf{C}$ *unique, lower triangular and with positive diagonal elements* *such that* $\mathbf{V} = \mathbf{C}\,\mathbf{C}^T$

$$\mathbf{V}^{-1} = \left[\mathbf{C}^{-1}\right]^T \mathbf{C}^{-1} \Longrightarrow \left[\mathbf{x}-\boldsymbol{\mu}\right]^T \mathbf{V}^{-1}\left[\mathbf{x}-\boldsymbol{\mu}\right] = \left[\mathbf{C}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]^T\left[\mathbf{C}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]$$

if $\boldsymbol{X} \sim N(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{V})$ *take*

$\mathbf{C}$ *such that* $\boldsymbol{X} = \boldsymbol{\mu} + \boldsymbol{C}\boldsymbol{Y}$

and $\boldsymbol{Y} \sim N(\boldsymbol{y}|\boldsymbol{0},\boldsymbol{I})$

$$C_{i1} = \frac{V_{i1}}{\sqrt{V_{11}}} \quad ;1 \le i \le n$$

$$C_{ij} = \frac{V_{ij} - \sum_{k=1}^{j-1} C_{ik}C_{jk}}{\sqrt{C_{jj}}} \quad ;1 < j < i \le n$$

$$C_{i1} = \left(V_{ii} - \sum_{k=1}^{i-1} C_{ik}^2\right)^{1/2} \quad ;1 < i \le n$$

$$C_{ij} = 0 \quad ;\forall j > i$$

since it is triangular inferior matrix

Algorithm:

0) Determine matrix $\mathbf{C}$ *from* $\mathbf{V}$

i) Generate n independent random quantities $\{z_1, z_2, \ldots, z_n\}$

each as $z_i \sim N(z|0,1)$

$$\mathbf{z} = \mathbf{C}^{-1}(\mathbf{x} - \boldsymbol{\mu})$$

$$\boldsymbol{x} = \boldsymbol{\mu} + \boldsymbol{C}\,\boldsymbol{z}$$

n)

ii) Get $x_i = \mu_i + \sum_{j=1}^{n} C_{ij}\, z_j$

$\Rightarrow$ $\mathbf{x} \sim N(\mathbf{x}|\boldsymbol{\mu}, \mathbf{V})$

Example: 2-dimensional Normal random quantity

$$\mathbf{V} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

$$\mathbf{C} = \begin{pmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sigma_2\sqrt{1-\rho^2} \end{pmatrix}$$

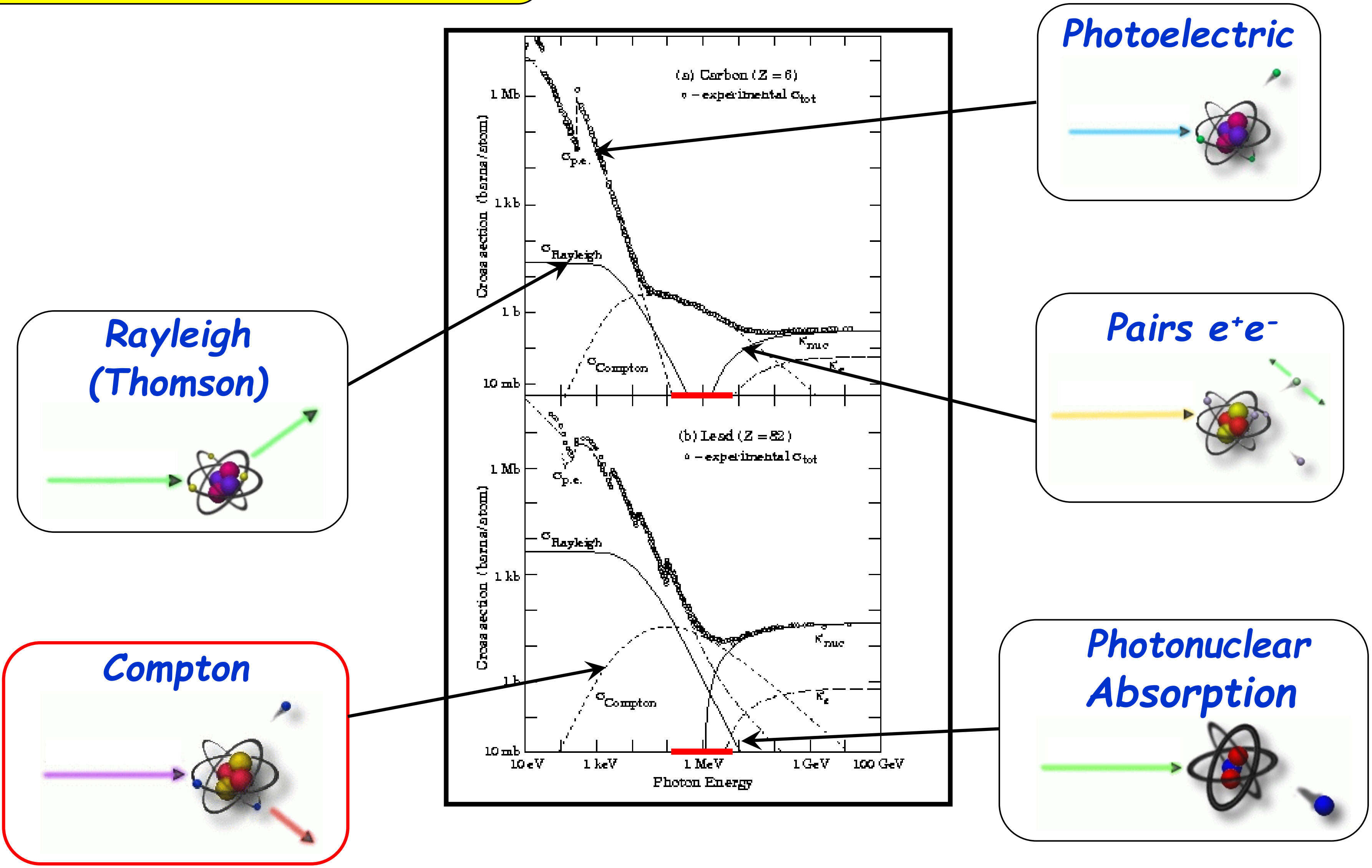
$$\begin{cases} C_{11} = \dfrac{V_{11}}{\sqrt{V_{11}}} = \sigma_1 & C_{12} = 0 \\ C_{21} = \dfrac{V_{21}}{C_{11}} = \rho\sigma_2 & C_{22} = \left(V_{22} - C_{21}^2\right)^{1/2} = \\ & = \sigma_2\sqrt{1-\rho^2} \end{cases}$$

$z_1, z_2 \sim N(z|0,1)$ $\Rightarrow$ $\boldsymbol{x} = \boldsymbol{\mu} + \boldsymbol{C}\,\boldsymbol{z}$

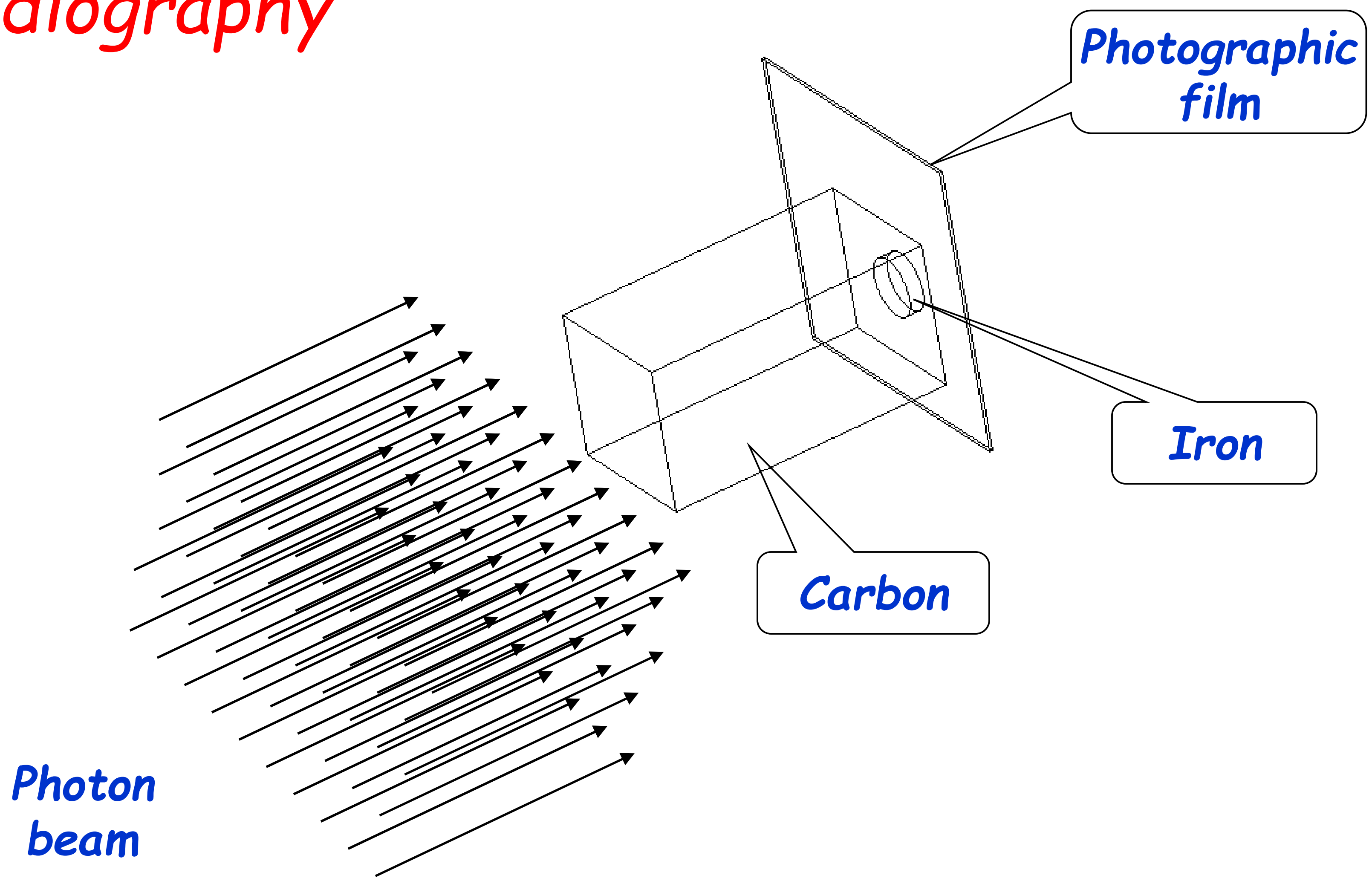
Example: Interaction of photons with matter

(Compton scattering)

Interaction of photons with matter



Radiography



1) What kind of process?

$$\sigma_t = \sigma_{Compton} + \sigma_{pairs} + \cdots$$

$$p_{Compton} = \frac{\sigma_{Compton}}{\sigma_t}$$

$$p_{pairs} = \frac{\sigma_{pairs}}{\sigma_t}$$

$$\cdots$$

σ* = cross-section =** [***"probability of interaction" with atom expressed in cm²]

$$F_1 = p_{Compton}$$

$$F_2 = p_{Compton} + p_{pairs}$$

$$\cdots$$

$$F_i = p_{Compton} + p_{pairs} + \cdots$$

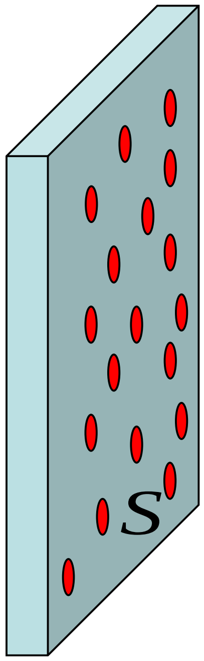
$$\cdots$$

$$F_n = 1$$

$$u \sim Un(u|0,1)$$

Compton	$u \le F_1$
pairs	$F_1 < u \le F_2$
$\cdots$	

2) Where do we have the first/next interaction?



In a thin slab of surface S and thickness ε

Volume $S\varepsilon$ *cm³*

Density ρ *g cm⁻³*

We have $S\varepsilon\rho \frac{N_A}{A}$ *atoms*

σ = *cross-section* = [*"probability of interaction" with atom expressed in cm²*]

"Interaction surface" covered by atoms $S_{eff} = \sigma\left[S\varepsilon\rho \frac{N_A}{A}\right]$ *cm²*

Total surface S → *Probability to interact with an atom in a thin slab*

$$p \equiv P_I(\varepsilon) = \frac{S_{eff}}{S} = \sigma\varepsilon\rho \frac{N_A}{A} \equiv \frac{\varepsilon}{\lambda}$$

$$P_{One\ I}(x+\varepsilon) = P_{No\ I}(x) P_I(\varepsilon)$$

$$\lambda \equiv \frac{A}{\sigma\rho N_A}$$

x *has* $m = x\varepsilon^{-1}$ *slabs of thickness* ε

$$P_{One\ I}(x+\varepsilon) = (1-p)^m p \overset{p<<}{\approx} e^{-mp} p \rightarrow e^{-x/\lambda}\lambda^{-1}dx$$

$$p(x\,|\,\lambda) = e^{-\frac{x}{\lambda}}\lambda^{-1}\ ; \quad x \in [0,\infty)$$

"mean free path"

$$E[x] = \lambda$$

Get the "mean free path"

$$\lambda = \frac{A}{\rho \sigma_t N_A} \; cm$$

$$\sigma_t = \sigma_{Compton} + \sigma_{pairs} + \cdots$$

Generate distance until the next interaction

$$p_{\text{int}}(x \mid \lambda) = e^{-\frac{x}{\lambda}} \lambda^{-1}$$

$$F_{int}(x \mid \lambda) = 1 - e^{-x/\lambda}$$

… Inverse Transform

$$u \sim Un(u \mid 0,1)$$

$$x = -\lambda \ln u \quad [cm]$$

$$(2) \leftrightarrow (1)$$

In this example, we shall simulate only the Compton process so I took the "Mean Free Path" for Compton Interaction only

$$\sigma = \int_{-1}^{1} \frac{d\sigma}{dx}\, dx$$

$$\gamma + e^-_{atomic} \rightarrow \gamma' + e^-_{free}$$

$$\sigma_0(E_\gamma) = \frac{\sigma_{Thomson}}{4}\left(\left(\frac{1+a}{a^2}\right)\left(\frac{2(1+a)}{1+2a} - \frac{1}{a}\ln(1+2a)\right) + \frac{1}{2a}\ln(1+2a) - \frac{1+3a}{(1+2a)^2}\right)$$

$$\gamma + atom \rightarrow \gamma' + atom^+ + e^-_{free}$$

$$\sigma_t(E_\gamma) = Z\sigma_0(E_\gamma)$$

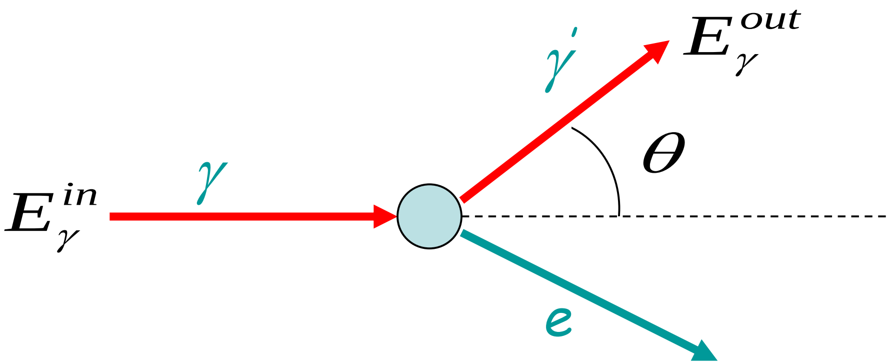
$$a = \frac{E_\gamma}{m_e}$$

$$\lambda = \frac{A}{\rho[Z\sigma_0(E_\gamma)]N_A}\ cm$$

$$(2) \leftrightarrow (1)$$

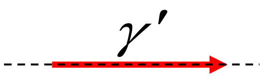
3) Compton Interaction

easy: **2 variables (8-2-4)** θ, ϕ



$$\varepsilon = \frac{E_\gamma^{out}}{E_\gamma^{in}} = \frac{1}{1 + a(1-\cos\theta)}$$

$$a = \frac{E_\gamma^{in}}{m_e} \qquad \theta \in [0, \pi]$$



$\theta = 0 \qquad \varepsilon = 1 \qquad E_\gamma^{out} = max[E_\gamma] = E_\gamma^{in}$

$\theta = \pi \qquad \varepsilon = \dfrac{1}{1+2a} \qquad E_\gamma^{out} = min[E_\gamma] = \dfrac{E_\gamma^{in}}{1+2a}$

Perturbative expansion in Relativistic Quantum Mechanics

$$\frac{d\sigma}{dx} = \frac{3\,\sigma_{Thomson}}{8} f(x) \qquad x = \cos\theta \in [-1,1] \qquad \phi \in [0, 2\pi]$$

(integrated)

$$f(x) = \frac{1}{[1 + a(1-x)]^2}\left(1 + x^2 + \frac{a^2(1-x)^2}{1+a(1-x)}\right)$$

$$\sigma_{Thomson} = 0.665\, barn = 0.665 \cdot 10^{-24}\ cm^2$$

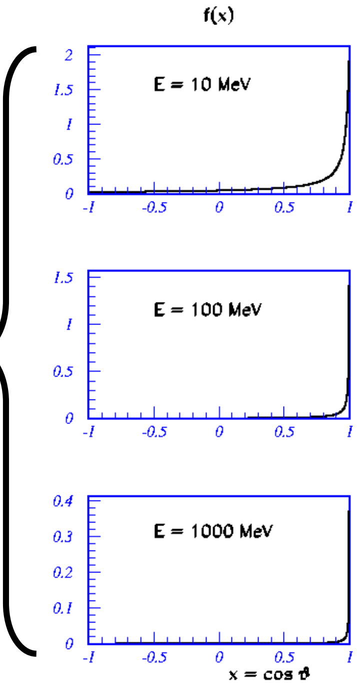
3.1) Generate polar angle for the outgoing photon (θ)

$$x = \cos\theta \in [-1,1]$$

$$f(x) = \frac{1}{[1+a(1-x)]^2}\left(1 + x^2 + \frac{a^2(1-x)^2}{1+a(1-x)}\right)$$

complicated to apply inverse transform…

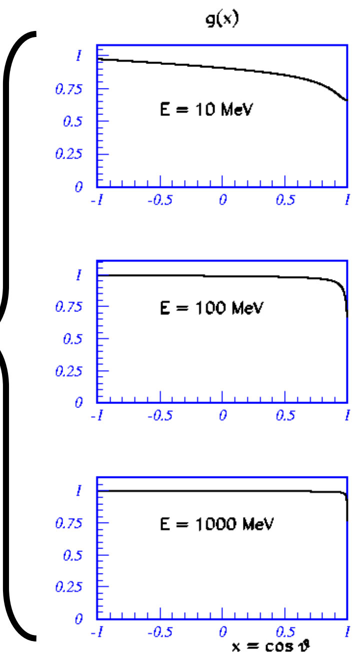
straight-forward acceptance-rejection very inefficient...



Use: Decomposition + inverse transform + acceptance-rejection

$$f(x) = [f_1(x) + f_2(x) + f_3(x)] \cdot g(x)$$

$$f_n(x) = \frac{1}{[1+a(1-x)]^n} \qquad g(x) = 1 - \frac{(2-x^2)\ f_1(x)}{1 + f_1(x) + f_2(x)}$$



$f_n(x) > 0$	*easy to invert… Inverse Transform*
$g(x) > 0$	*"fairly flat"… Acceptance-rejection*

$$f(x) = [f_1(x) + f_3(x) + f_3(x)] \cdot g(x)$$

$$f_n(x) = \frac{1}{[1 + a(1-x)]^n}$$

Probability Densities

$$p_i(x) = \frac{1}{w_i} f_i(x)$$

$$\int_{-1}^{1} p_i(x)\,dx = 1$$

$$w_i = \int_{-1}^{1} f_i(x)\,dx$$

$$w_1 = \frac{1}{a} \ln b$$

$$w_2 = \frac{2}{a^2(b^2-1)}$$

$$w_3 = \frac{2b}{a^3(b^2-1)^2}$$

$$b = 1 + 2a$$

$$f(x) = [f_1(x) + f_3(x) + f_3(x)] \cdot g(x) =$$

$$= [w_1 p_1(x) + w_2 p_2(x) + w_3 p_3(x)] \cdot g(x) =$$

$$= w_T\ [\alpha_1 p_1(x) + \alpha_2 p_2(x) + \alpha_3 p_3(x)] \cdot g(x) =$$

$$w_T = w_1 + w_2 + w_3$$

$$\alpha_i = \frac{w_i}{w_T}$$

$$\alpha_1 + \alpha_2 + \alpha_3 = 1$$

$$= w_T\ h(x)\ \ g(x) = f(x)$$

… we have everything…

1) Sample $X_g \sim w_T \left[\alpha_1 p_1(x) + \alpha_2 p_2(x) + \alpha_3 p_3(x)\right] \cdot g(x)$

Decomposition

$$u \sim Un(u|0,1) \quad \left\{ \begin{array}{c} u < \alpha_1 \\ \alpha_1 < u \le \alpha_1 + \alpha_2 \\ \alpha_1 + \alpha_2 < u \end{array} \right\} \quad \begin{array}{l} x_g \sim p_1(x) \\ x_g \sim p_2(x) \\ x_g \sim p_3(x) \end{array}$$

Inverse transform

$$F(x)_i = \int_{-1}^{x} p_i(s)\, ds \qquad u \sim Un(u|0,1)$$

$$F_1(x) = 1 - \frac{\ln\left[1 + a(1-x)\right]}{\ln b} \qquad x_g = \frac{1 + a - b^u}{a}$$

$$F_2(x) = \frac{b^2 - 1}{2(b-x)} - \frac{1}{2a} \qquad x_g = b - \frac{a(b^2-1)}{1+2au}$$

$$F_3(x) = \frac{1}{4a(1+a)}\left[\frac{(b+1)^2}{(b-x)^2} - 1\right] \qquad x_g = b - \frac{b+1}{\left[1 + 4a(1+a)u\right]^{1/2}}$$

acceptance-rejection $u \sim Un(u|0, g_M)$

$$g_M = max\left[g(x)\right] = $$
$$= g(x=-1) = 1 - \frac{b}{1+b+b^2}$$

$u \le g(x_g)$ accept x_g

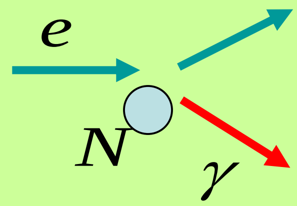
$u > g(x_g)$ reject x_g

2-body kinematics

$$x_g \sim f(x)$$

$$E_\gamma^{out} = \frac{E_\gamma^{in}}{1 + a(1 - x_g)}$$

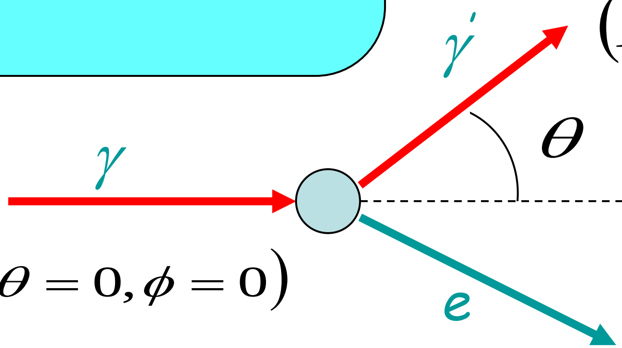
$$\phi_g \sim Un(\phi \mid 0, 2\pi)$$



... for the electron:

$$E_e^{out} = E_\gamma^{in}\left(1 + \frac{1}{a} - \varepsilon\right)$$

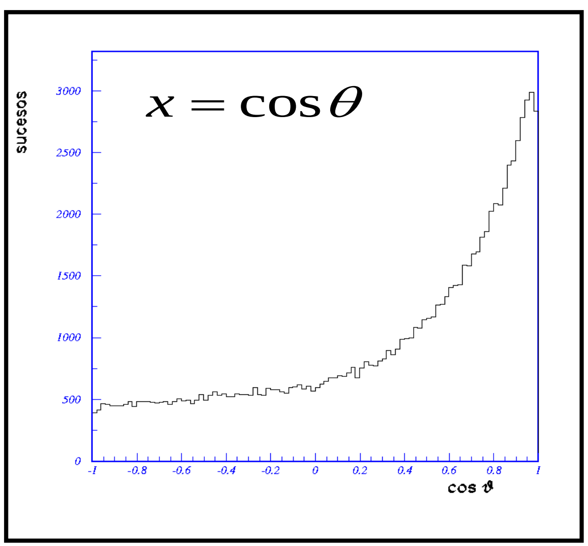
$$\tan\theta_e = \frac{\cot \theta/2}{1 + a}$$

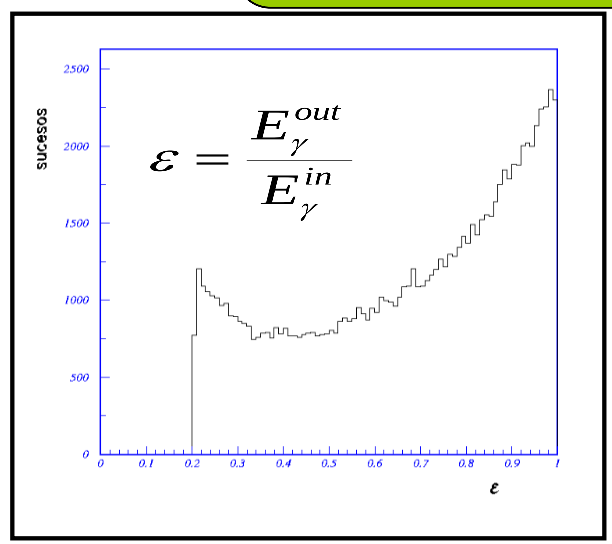


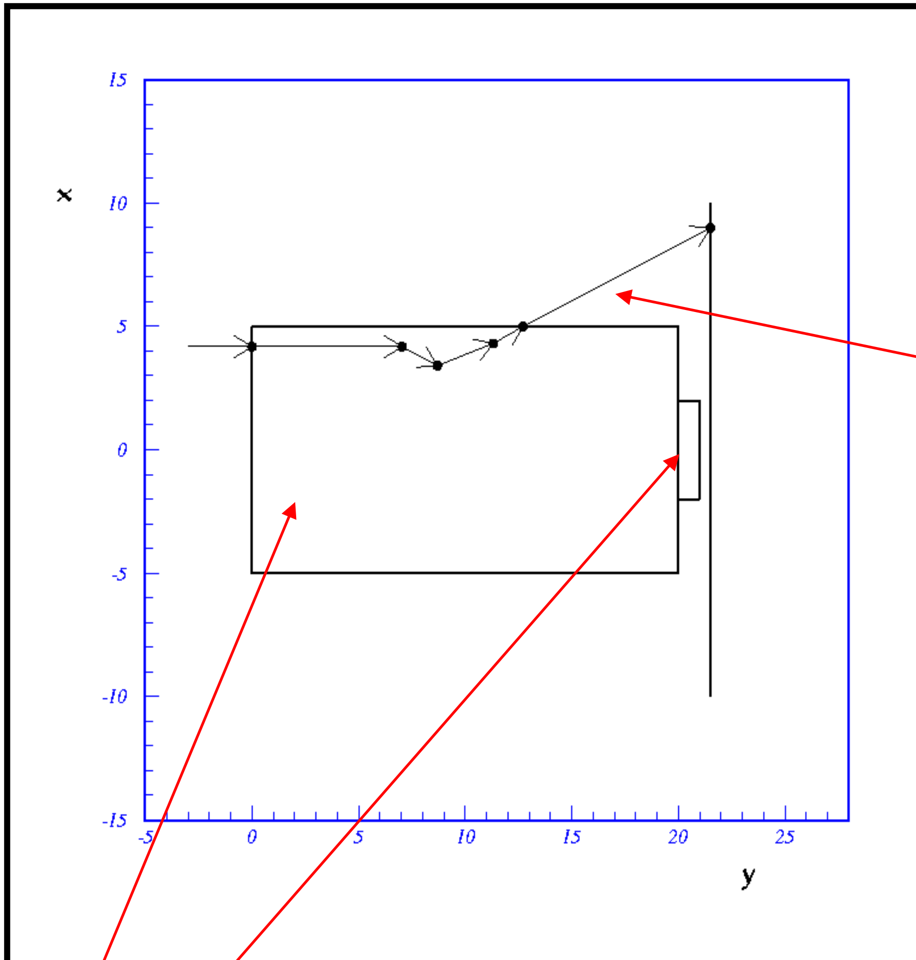
$$\left(E_\gamma^{out}, \theta_g = \mathrm{acos}(x_g), \phi_g\right)$$

$$\left(E_\gamma^{in}, \theta = 0, \phi = 0\right)$$

θ_g *with respect to the direction of incidence of the photon !! ... rotation*



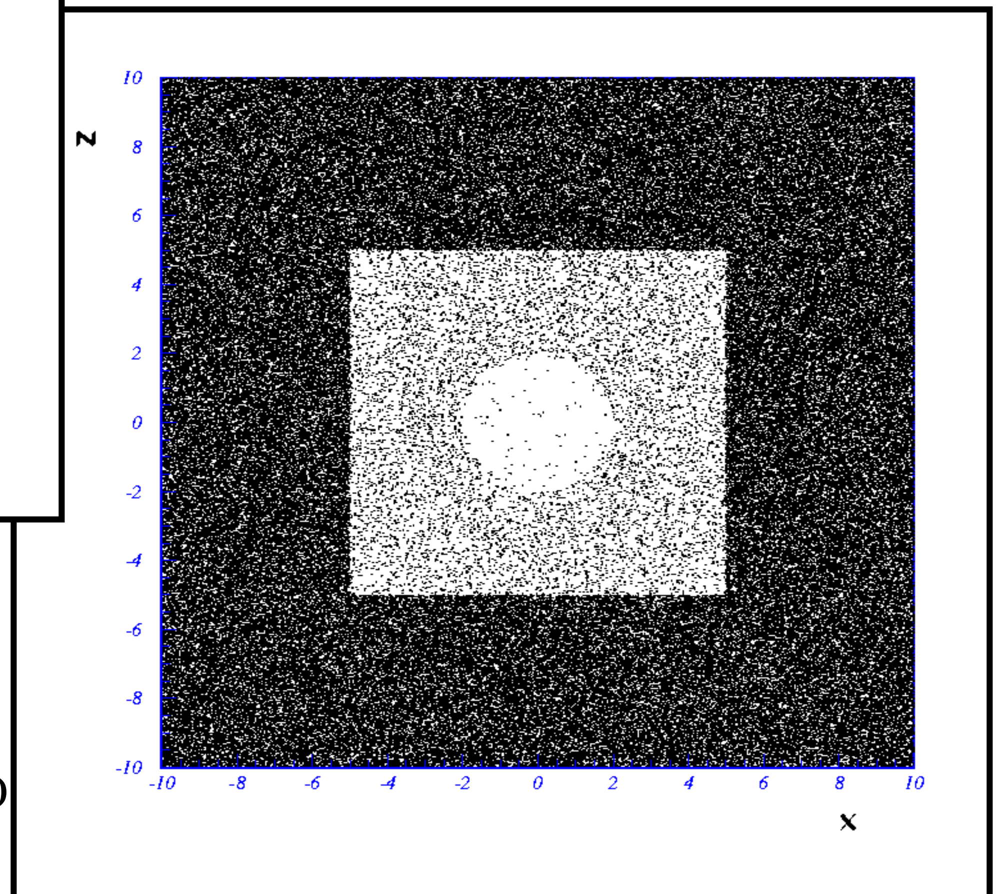




100,000 generated photons

$$E_{\gamma}^{in} = 1\ \mathrm{MeV}$$

trajectory of one photon



C : $(Z = 6, A = 12, \rho = 2.26\ gr \cdot cm^{-3})$

Fe : $(Z = 26, A = 55.85, \rho = 7.87\ gr \cdot cm^{-3})$

END OF FIRST PART

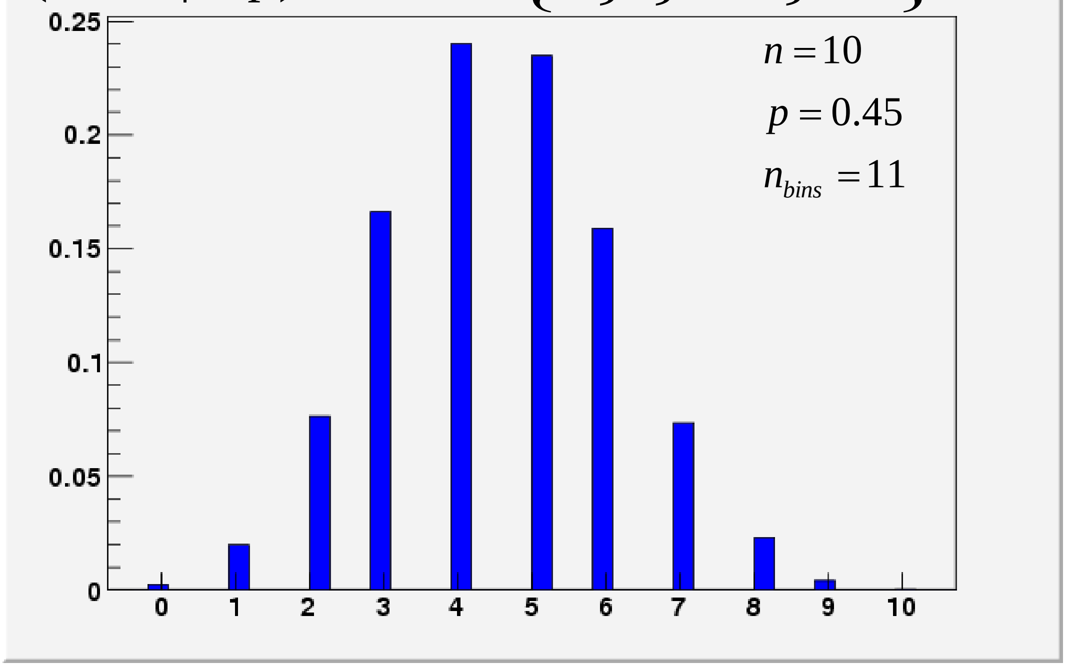
To come:
Markov Chain Monte Carlo (Metropolis, Hastings, Gibbs,…)
Examples: Path Integrals in Quantum Mechanics
Bayesian Inference

Method 5: Markov Chain Monte Carlo

$$X \sim Bi(k \mid n, p) \qquad P(X = k \mid n, p) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$k = 0, 1, 2, \ldots, n$$

$P(X = k \mid n, p) \quad k = \{0, 1, \ldots, 10\}$



$$k = \{0, 1, \ldots, n\}$$

$$\pi_{k+1} = P(X = k)$$

$$\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_{n+1})$$

Probability vector

$$\pi_i \geq 0 ; \quad \sum_{i=1}^{n+1} \pi_i = 1$$

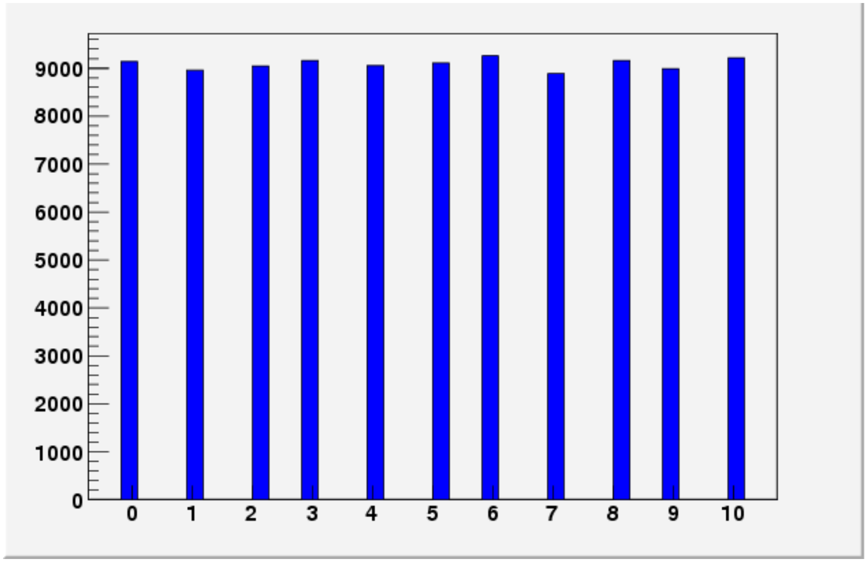
Step 0) **Generate N=100,000**

number of bins = 11

$$Un^{D}(k|1,\ldots,n+1=11)$$

$$P(X=k)=1/_{n+1}=1/_{11}\quad;\quad\forall k$$

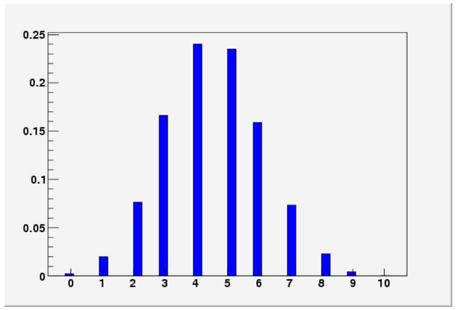
$$\left(d_1,d_2,\ldots,d_{n+1}\right)\qquad\sum_{i=1}^{n+1}d_i=N$$



Sampling probability vector $\boldsymbol{\pi}^{(0)}=\left(\pi_1^{(0)},\pi_2^{(0)},\ldots,\pi_{n+1}^{(0)}\right)=\left(d_1/N,d_2/N,\ldots,d_{n+1}/N\right)$

First step done: ***We have already the N=100,000 events***

But we want this: →



Second step: ***redistribute all N generated events moving each event from its present bin to a different one (or the same) to get eventually a sampling from*** $P(X=k)=\pi_k$

HOW?

In one step: an event in bin *i* goes to bin *j* with probability $P(X = j \mid n, p)$

… But this is equivalent to sample from $Bi(k \mid n, p)$

Sequentially: At every step of the evolution, we move all events from the bin where they are to a new bin (may be the same) with a **migration probability**

$$(\boldsymbol{P})_{ij} = P(i \to j) = a(j \mid i)$$

step	populations	→	probability vector
(0)	$\left(d_1^{(0)}, d_2^{(0)}, \ldots, d_{n+1}^{(0)}\right)$		$\boldsymbol{\pi}^{(0)} = \left(\pi_1^{(0)}, \pi_2^{(0)}, \ldots, \pi_{n+1}^{(0)}\right)$
(1)	$\left(d_1^{(1)}, d_2^{(1)}, \ldots, d_{n+1}^{(1)}\right)$		$\boldsymbol{\pi}^{(1)} = \left(\pi_1^{(1)}, \pi_2^{(1)}, \ldots, \pi_{n+1}^{(1)}\right)$
$\vdots$			
(i)	$\left(d_1^{(i)}, d_2^{(i)}, \ldots, d_{n+1}^{(i)}\right)$		$\boldsymbol{\pi}^{(i)} = \left(\pi_1^{(i)}, \pi_2^{(i)}, \ldots, \pi_{n+1}^{(i)}\right)$
$(i+1)$	$\left(d_1^{(i+1)}, d_2^{(i+1)}, \ldots, d_{n+1}^{(i+1)}\right)$		$\boldsymbol{\pi}^{(i+1)} = \left(\pi_1^{(i+1)}, \pi_2^{(i+1)}, \ldots, \pi_{n+1}^{(i+1)}\right)$
$\vdots$			

Transitions among states of the system

$$\boldsymbol{\pi}^{(0)}$$

$$\boldsymbol{\pi}^{(1)} = \boldsymbol{\pi}^{(0)}\mathbf{P}$$

$$\boldsymbol{\pi}^{(2)} = \boldsymbol{\pi}^{(1)}\mathbf{P} = \boldsymbol{\pi}^{(0)}\mathbf{P}^2$$

$$\boldsymbol{\pi}^{(k)} = \boldsymbol{\pi}^{(k-1)}\boldsymbol{P} = \boldsymbol{\pi}^{(k-2)}\boldsymbol{P}^2 = \cdots = \boldsymbol{\pi}^{(0)}\boldsymbol{P}^k$$

Transition Matrix among states

$$\boldsymbol{P} = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{pmatrix}$$

$$(\boldsymbol{P})_{ij} = P(i \to j) = a(j \mid i)$$

Goal: Find a Transition Matrix $\boldsymbol{P} \in R^{n \times n}$ that allows to go from $\boldsymbol{\pi}^{(0)}$ to desired $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_N)$

$\boldsymbol{\pi}^{(i+1)}$ depends upon $\boldsymbol{\pi}^{(i)}$ and not on $\boldsymbol{\pi}^{(j<i)}$

Markov Chain with transition matrix P

Transition Matrix is a Probability Matrix

$$\sum_{j=1}^{n} p_{ij} = \sum_{j=1}^{n} P(i \to j) = 1$$

the probability to go from state i to whatever some other state is 1

Remainder: *If the Markov Chain is…*

irreducible : ***all the states of the system communicate among themselves***

and *ergodic:* *that is, the states of the system are*

recurrent : being at one state, we shall return to it with $p=1$

positive: we shall go to it in a finite number of steps

aperiodic: ***the system is not trapped cycles***

i) There is a unique stationary distribution $\boldsymbol{\pi}$ *with* $\boldsymbol{\pi}=\boldsymbol{\pi}\mathbf{P}$ *(unique fix vector)*

ii) Starting at any arbitrary state $\boldsymbol{\pi}^{(0)}$ *, the sequence*

$$\boldsymbol{\pi}^{(0)}, \boldsymbol{\pi}^{(1)}=\boldsymbol{\pi}^{(0)}\boldsymbol{P}, \ldots, \boldsymbol{\pi}^{(n)}=\boldsymbol{\pi}^{(n-1)}\boldsymbol{P}=\ldots=\boldsymbol{\pi}^{(0)}\boldsymbol{P}^{n}, \ldots$$

tends asymptotically to the fix vector $\boldsymbol{\pi}=\boldsymbol{\pi}\mathbf{P}$

iii)

$$lim_{n\to\infty}\mathbf{P}=\begin{pmatrix} \pi_1 & \pi_2 & \cdots & \pi_N \\ \pi_1 & \pi_2 & \cdots & \pi_N \\ \vdots & \vdots & \cdots & \vdots \\ \pi_1 & \pi_2 & \cdots & \pi_N \end{pmatrix}$$

∞ *ways to choose the* ***Probability Transition Matrix***

A sufficient condition (not necessary) for $\boldsymbol{\pi}$ ***to be a fix vector of*** $\mathbf{P}$ ***is that the Detailed Balance relation*** $\pi_i \, (\boldsymbol{P})_{ij} = \pi_j \, (\boldsymbol{P})_{ji}$ ***is satisfied***

Why? ***It assures that*** $\boldsymbol{\pi} \, \boldsymbol{P} = \boldsymbol{\pi}$

$$\boldsymbol{\pi} \, \boldsymbol{P} = \left(\sum_{i=1}^{N} \pi_i (\boldsymbol{P})_{i1}, \sum_{i=1}^{N} \pi_i (\boldsymbol{P})_{i2}, \ldots, \sum_{i=1}^{N} \pi_i (\boldsymbol{P})_{iN} \right) = \boldsymbol{\pi}$$

If $\pi_i \, (\boldsymbol{P})_{ij} = \pi_j \, (\boldsymbol{P})_{ji} \longrightarrow \sum_{i=1}^{N} \pi_i (\boldsymbol{P})_{ik} = \sum_{i=1}^{N} \pi_k (\boldsymbol{P})_{ki} = \pi_k$

$$\forall k = 1,2,\ldots, N$$

Procedure to follow

After specifying $(\boldsymbol{P})_{ij} = P(i \to j) = a_{ij}$ *according to the Detailed Balance condition*

step	populations	probability vector (state of the system)
(i)	$\left(d_1^{(i)}, d_2^{(i)}, \ldots, d_{n+1}^{(i)}\right)$	$\boldsymbol{\pi}^{(i)} = \left(\pi_1^{(i)}, \pi_2^{(i)}, \ldots, \pi_{n+1}^{(i)}\right)$

For all the 11 bins:

i) **For each event at bin** $i = 1,\ldots,11$ **choose a candidate bin to go (j) among the 11 possible ones as** $Un^{disc}(k\,|\,1,11)$

ii) **Accept the migration** $i \to j$ **with probability** a_{ij}

Accept the migration $i \to i$ **with probability** $1 - a_{ij}$

$(i+1)$	$\left(d_1^{(i+1)}, d_2^{(i+1)}, \ldots, d_{n+1}^{(i+1)}\right)$	$\boldsymbol{\pi}^{(i+1)} = \left(\pi_1^{(i+1)}, \pi_2^{(i+1)}, \ldots, \pi_{n+1}^{(i+1)}\right)$

How do we take $a_{ij} = a(j \mid i)$ **so that** $\lim_{i\to\infty} \boldsymbol{\pi}^{(i)} = \boldsymbol{\pi} = (p_0, p_1, \ldots, p_{10})$

Detailed Balance condition $\pi_i \, (\boldsymbol{P})_{ij} = \pi_i \, a_{ij} = \pi_j \, a_{ji} = \pi_j \, (\boldsymbol{P})_{ji}$

$$\pi_i > \pi_j \rightarrow a(j \mid i) = \min\left\{1, \frac{\pi_j}{\pi_i}\right\} = \frac{\pi_j}{\pi_i} \qquad a(i \mid j) = \min\left\{1, \frac{\pi_i}{\pi_j}\right\} = 1$$

$$\pi_i < \pi_j \rightarrow a(j \mid i) = \min\left\{1, \frac{\pi_j}{\pi_i}\right\} = 1 \qquad a(i \mid j) = \min\left\{1, \frac{\pi_i}{\pi_j}\right\} = \frac{\pi_i}{\pi_j}$$

$$\Bigg\} \; \pi_i \, a(j \mid i) = \pi_j \, a(i \mid j)$$

$\pi_i \, (\mathbf{P})_{ij} = \pi_j \, (\mathbf{P})_{ji}$ ⇒ $lim_{i\to\infty} \boldsymbol{\pi}^{(i)} = \boldsymbol{\pi}$

For instance, at step t …

For an event in bin $i = 7$

Choose a bin j to go as $J \sim Un^{D}(j|1,11)$

$$j = 2 \qquad a_{72} = a(7 \to 2) = \min\left[1, \frac{\pi_2}{\pi_7} = \frac{P(X=1)}{P(X=6)} = 0.026\right] = 0.026$$

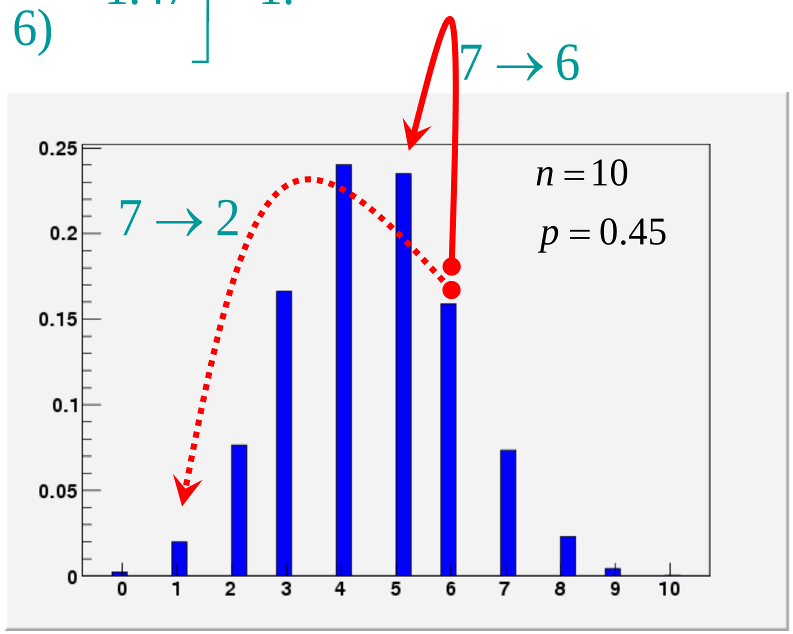
$$u \sim Un(u|0,1)$$

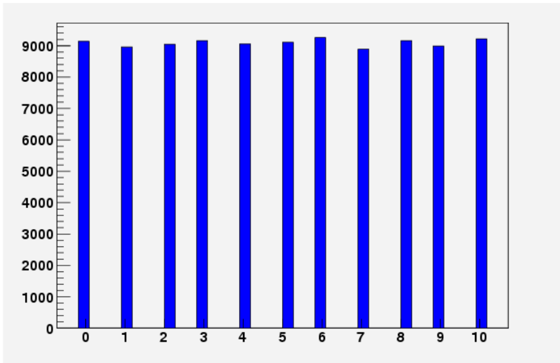
$u \leq 0.026$ Move event to bin 2

$u > 0.026$ Leave event in bin 7

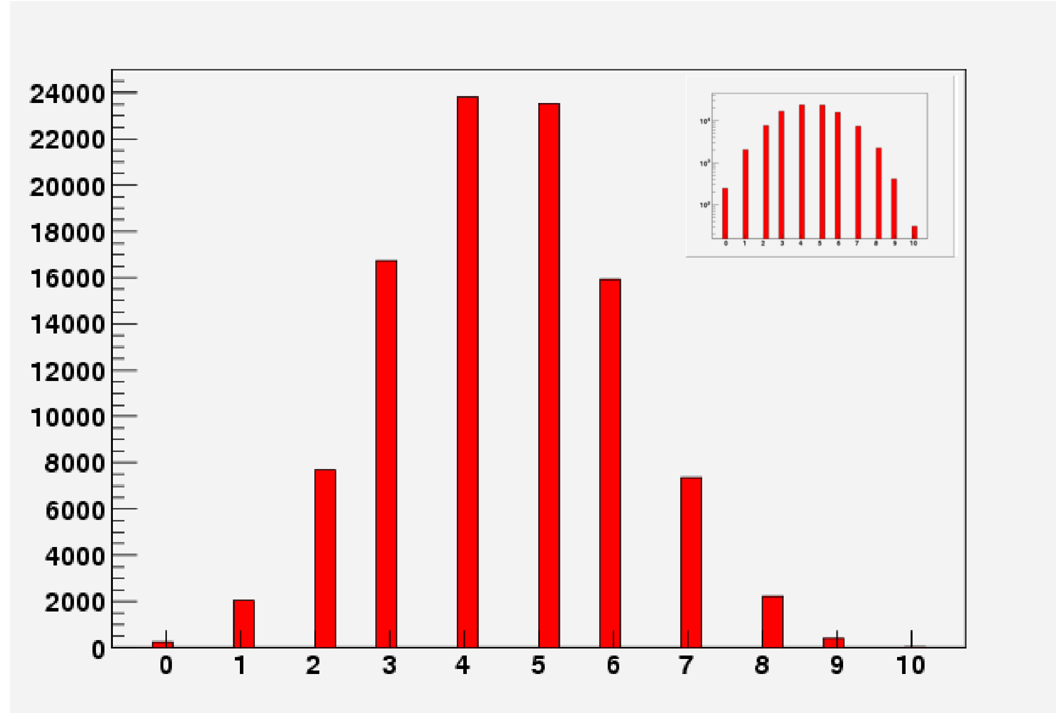
$$j = 6 \qquad a_{76} = a(7 \to 6) = \min\left[1, \frac{\pi_6}{\pi_7} = \frac{P(X=5)}{P(X=6)} = 1.47\right] = 1.$$

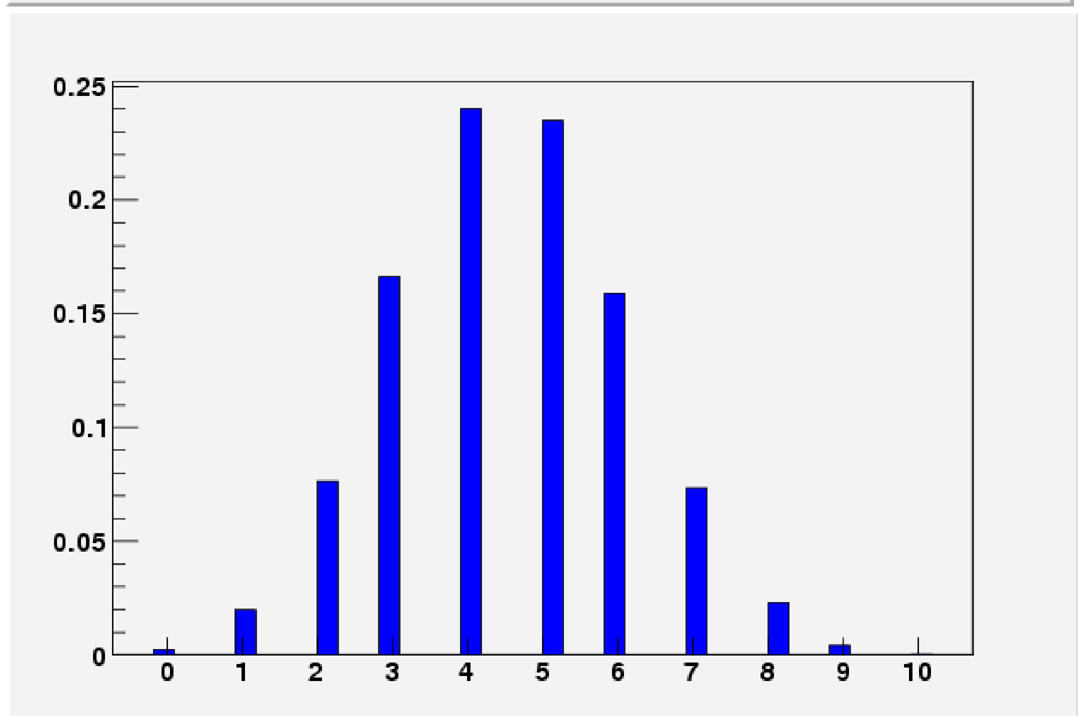
Move event to bin 6





After 20 steps…

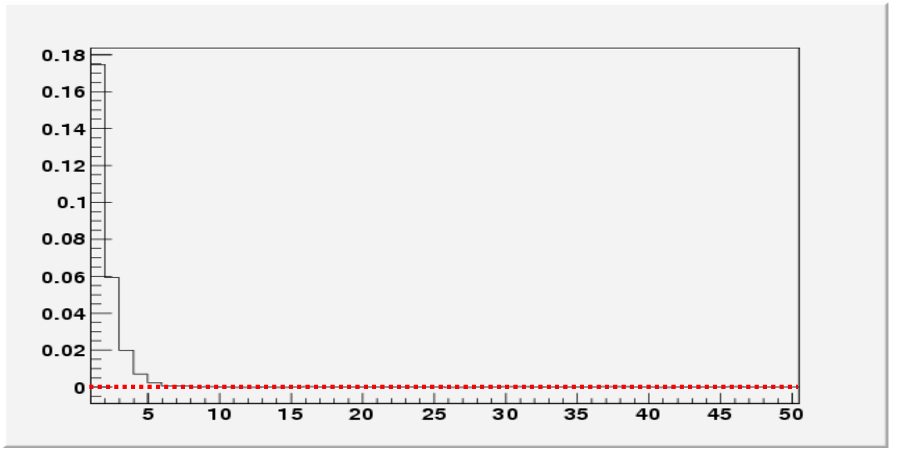




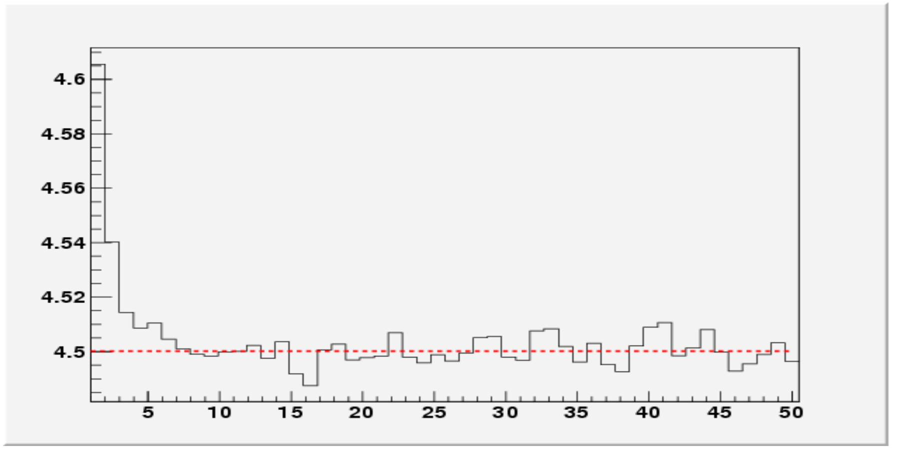
$$X \sim Bi(k \mid n = 10, p = 0.45)$$

Convergence?

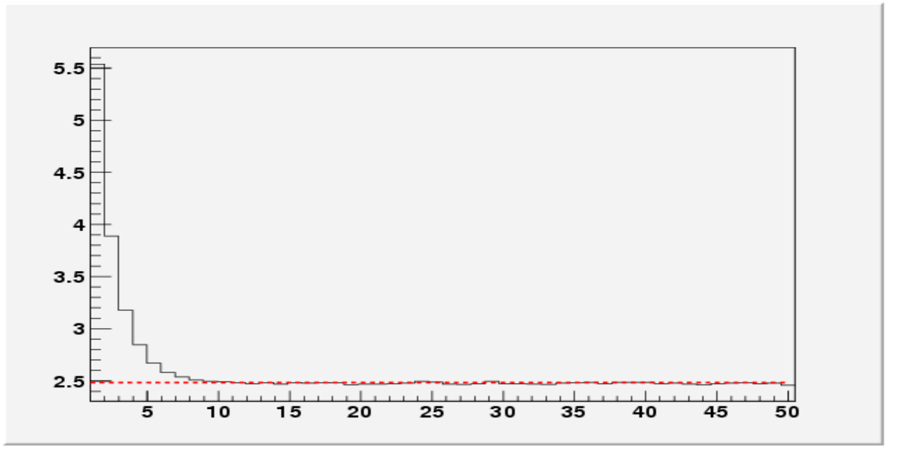
$$D_{KL}[p \mid \tilde{p}] = \sum_k p_k \log \frac{p_k}{\tilde{p}_k}$$



$$E[X] = N\theta = 4.5$$



$$V[X] = N\theta(1-\theta) = 2.475$$



Watch for trends, correlations, "good mixing",...

Detailed Balance Condition → **Still Freedom to choose the Transition Matrix P**

- **Trivial election:** $(\mathbf{P})_{ij} = \pi_j \longrightarrow \pi_i \, \pi_j = \pi_j \, \pi_i$

  Select a new state migrating events from one bin to another with probability $P(X = k) = \pi_k$

- **Basis for Markov Chain Monte Carlo simulation**

$$(\mathbf{P})_{ij} = q(j|i) \cdot a_{ij}$$

***Simple* probability to choose a new possible bin j for an event that is at bin i** (for $q(j|i)$)

probability to accept the proposed new bin j for an event at bin i taken in such a way that the Detailed Balance Condition is satisfied (for a_{ij})

$$\pi_i \, (\mathbf{P})_{ij} = \pi_j \, (\mathbf{P})_{ji}$$

Metropolis-Hastings *algorithm*

$$a_{ij} = min\left[1, \frac{\pi_j\, q(i|j)}{\pi_i\, q(j|i)}\right]$$

$$a_{ij} = min\left[1, \frac{\pi_j\, q(i|j)}{\pi_i\, q(j|i)}\right] = \frac{\pi_j\, q(i|j)}{\pi_i\, q(j|i)}$$

$$a_{ji} = min\left[1, \frac{\pi_i\, q(j|i)}{\pi_{ji}\, q(i|j)}\right] = 1$$

$$\pi_i\,(\mathbf{P})_{ij} = \pi_i\, q(j|i)\, a_{ij} = \pi_i\, q(j|i)\, \frac{\pi_j\, q(i|j)}{\pi_i\, q(j|i)} =$$

$$= \pi_j\, q(i|j) = \pi_j\, q(i|j)\, a_{ji} = \pi_j\,(\mathbf{P})_{ji}$$

Can take: $q(j|i) \neq q(i|j)$ **(not symmetric), even** $q(j|i) = q(j)$

but better as close as possible to desired distribution for high acceptance probability

$$q(i|j) \to \pi_i \quad \Rightarrow \quad a_{ij} \to 1$$

If symmetric: $q(j|i) = q(i|j)$ →

Metropolis algorithm

$$q(j|i) = q(i|j)$$

(symmetric)

$$a_{ij} = min\left[1, \frac{\pi_j}{\pi_i}\right]$$

➢ **For absolute continuous distributions**

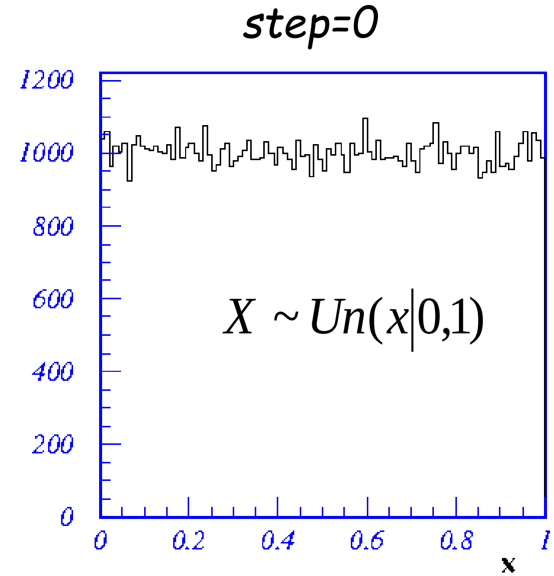
$$X \sim \pi(\boldsymbol{x} \mid \boldsymbol{\theta})$$

Probability density function $\pi(\boldsymbol{x} \mid \boldsymbol{\theta})$

$$p(\boldsymbol{x} \to \boldsymbol{x}') = q(\boldsymbol{x}' \mid \boldsymbol{x})\, a(\boldsymbol{x} \to \boldsymbol{x}') \qquad \boldsymbol{x}, \boldsymbol{x}' \in \Omega_X$$

$$a(\boldsymbol{x} \to \boldsymbol{x}') = \min\left[1, \frac{\pi(\boldsymbol{x}' \mid \boldsymbol{\theta})\, q(\boldsymbol{x} \mid \boldsymbol{x}')}{\pi(\boldsymbol{x} \mid \boldsymbol{\theta})\, q(\boldsymbol{x}' \mid \boldsymbol{x})}\right]$$

Example: $X \sim Be(x|4,2)$

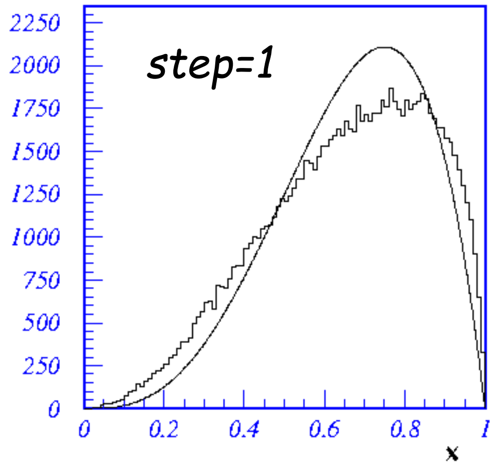


$$\pi(s) \propto s^3\,(1-s) \qquad p(s \to s') = q(s|s') \cdot a(s \to s')$$

Metropolis-Hastings

$$q(s|s') = 2\,s$$

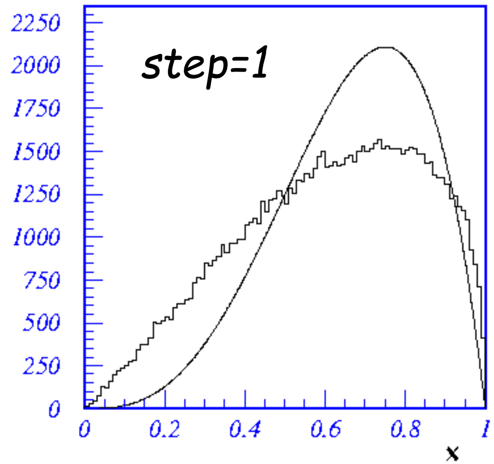
$$a(s \to s') = min\left[1, \frac{s'^2(1-s')}{s^2(1-s)}\right]$$



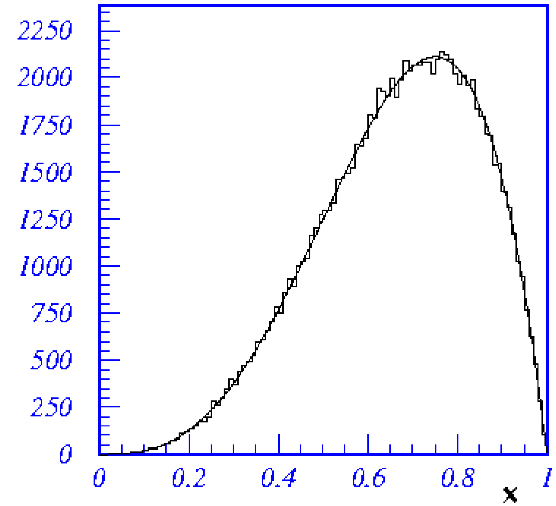
Metropolis

$$q(s|s') = Un(s|0,1)$$

$$a(s \to s') = min\left[1, \frac{s'^3(1-s')}{s^3(1-s)}\right]$$



... and after 20 steps..



In practice, we do not proceed as in the previous examples, (meant for illustration)

Once equilibrium reached…

> ***different steps → different samplings from "same" distribution***

1) At step $t = 0$ **, choose an admissible arbitrary state** $\boldsymbol{x}^{(0)}$ $\{\boldsymbol{x}^{(0)} \in \Omega_x ; \pi(\boldsymbol{x}^{(0)} \mid \boldsymbol{\theta}) > 0\}$

$t = t + 1$

2) Generate a proposed new value $\boldsymbol{x}'$ **from the distribution**

$$q(\boldsymbol{x}' \mid \boldsymbol{x}^{(t-1)}, \phi) = q(\boldsymbol{x}^{(t-1)} \mid \boldsymbol{x}', \phi)$$

(symmetric)

$$q(\boldsymbol{x}' \mid \boldsymbol{x}^{(t-1)}, \phi) \neq q(\boldsymbol{x}^{(t-1)} \mid \boldsymbol{x}', \phi)$$

(not symmetric)

3) Accept the new value $\boldsymbol{x}'$ **with probability**

$$a(\boldsymbol{x}', \boldsymbol{x}^{(t-1)}) = \min\left\{1, \frac{\pi(\boldsymbol{x}' \mid \boldsymbol{\theta})}{\pi(\boldsymbol{x}^{(t-1)} \mid \boldsymbol{\theta})}\right\}$$

$$a(\boldsymbol{x}', \boldsymbol{x}^{(t-1)}) = \min\left\{1, \frac{\pi(\boldsymbol{x}' \mid \boldsymbol{\theta})}{\pi(\boldsymbol{x}^{(t-1)} \mid \boldsymbol{\theta})} \frac{q(\boldsymbol{x}^{(t-1)} \mid \boldsymbol{x}', \phi)}{q(\boldsymbol{x}' \mid \boldsymbol{x}^{(t-1)}, \phi)}\right\}$$

4) If accepted, set $\boldsymbol{x}^{(t)} = \boldsymbol{x}'$ **Otherwise, set** $\boldsymbol{x}^{(t)} = \boldsymbol{x}^{(t-1)}$

Metropolis ***Metropolis-Hastings***

1) Need some initial steps to reach "equilibrium" (stable running conditions)

2) For samplings, take draws every few steps to reduce correlations (if important)

Properties:

- *“Easy” and powerful* — *Virtually any density regardless the number of dimensions and analytic complexity*
- *Changes of configurations depend on*

$$r = \frac{\pi(s')\, q(s|s')}{\pi(s)\, q(s'|s)}$$

Normalization is not relevant

- *Sampling “correct” asymptotically* — *Need previous sweeps (“burn-out”, “thermalization”) to reach asymptotic limit*
- *Correlations among different states* — *If an issue, additional sweeps*

… nevertheless, in many circumstances is the only practical solution...

And Last…

1) Some ("many") times, we do not have the explicit expression of the pdf but know the conditional densities

2) Usually, conditionals densities have simpler expressions

3) Bayesian Structure:

$$p(\theta,\phi \mid x) \propto p(x \mid \theta,\phi)\, p(\phi \mid \theta)\, p(\theta)$$

(and in particular hierarchical models)

…Sampling from Conditional densities

Method 5: Markov Chain Monte Carlo with Gibbs Sampling

EXAMPLE: Sampling from Conditionals

Sampling of $X \sim St(x|\nu) \quad ;\nu \in \Re^{+}$ $\qquad p(x|\nu) = N\left(1+x^2/\nu\right)^{-\left(\frac{\nu+1}{2}\right)}$

$$F(x|\nu) = \frac{1}{2} + \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{\pi\nu}}\left[xF(1/2,(\nu+1)/2,3/2,-x^2/2)\right]$$

0) Consider that $\int_0^{\infty} e^{-au} u^{b-1} du = \Gamma(b) a^{-b}$

1) Introduce a new random quantity $U \sim Ga(u|a,b)$ **(extra dimension) with**

$$a = 1 + \frac{x^2}{\nu} \qquad b = \frac{\nu+1}{2} \qquad \boxed{p(x,u|\nu) = \frac{N}{\Gamma(b)} u^{b-1} e^{-au}} \qquad \int_{-\infty}^{\infty} dx \int_0^{\infty} p(x,u)du = 1$$

Marginals:

$$p(x|\nu) = \int_0^{\infty} p(x,u|\nu)du \propto a^{-b}$$
$$\sim St(x|\nu)$$

$$p(u|\nu) = \int_{-\infty}^{\infty} p(x,u|\nu)dx \propto u^{b-3/2} e^{-u}$$
$$\sim Ga(u|a,b)$$

Conditionals:

$$p(x|u,\nu) = \frac{p(x,u|\nu)}{p(u|\nu)} = N(x|0,\sigma^2 = \nu u^{-1})$$

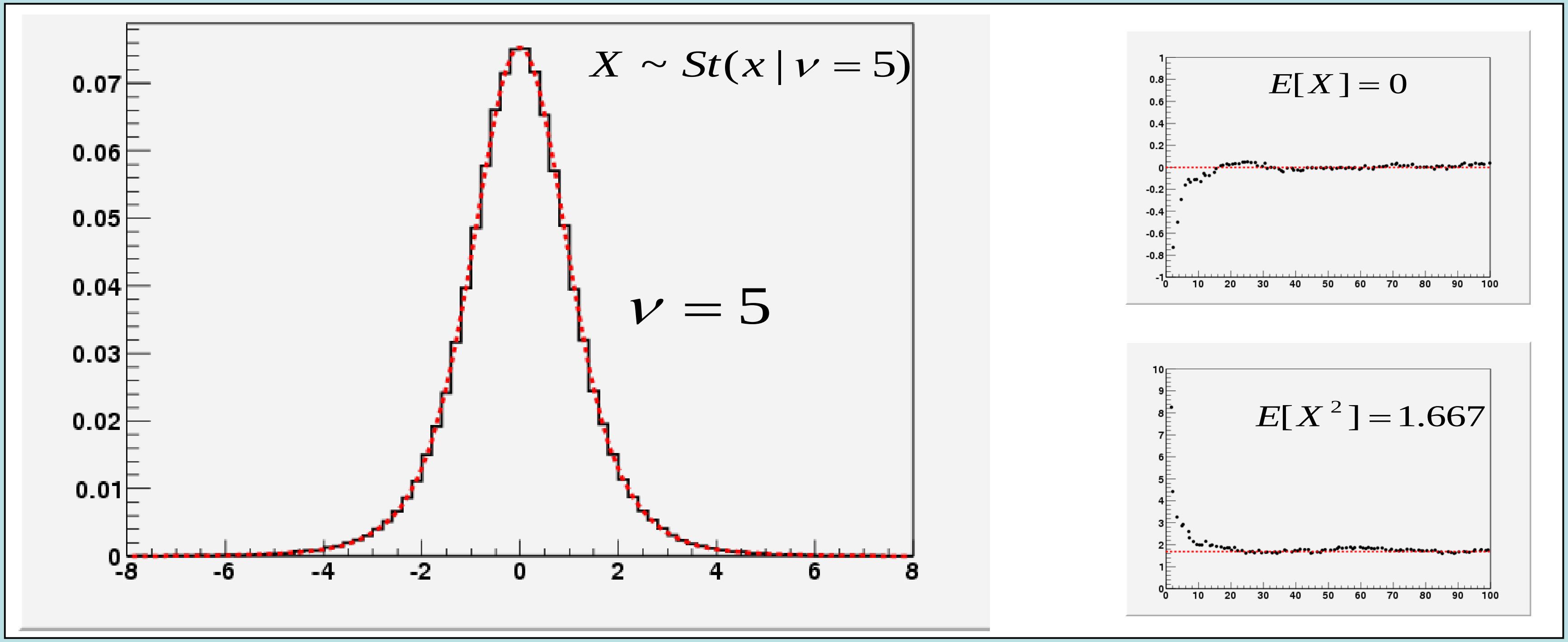
$$p(u|x,\nu) = \frac{p(x,u|\nu)}{p(x|\nu)} = Ga(u|a,b)$$

Sampling:

1) Start at t=0 with $x_0 \in R$ $\qquad b = \frac{\nu+1}{2}$

2) At step t Sample $U \mid X \sim Ga(u \mid a, b)$ $\qquad a = 1 + \nu^{-1} x_{t-1}^2$

$$X \mid U \sim N(x \mid 0, \sigma^2 = \nu u^{-1})$$



Yet another EXAMPLE…

Trivial marginals for the Behrens-Fisher problem (lect. 2)

$$W = \mu_1 - \mu_2 \sim \quad p(w \mid \boldsymbol{x}_1, \boldsymbol{x}_2) \sim \int_{-\infty}^{\infty} \left(s_1{}^2 + (\bar{x}_1 - w - u)^2 \right)^{-(n_1+1/2)} \left(s_2{}^2 + (\bar{x}_2 - u)^2 \right)^{-(n_2+1/2)} du$$

However:

$$p(\mu_1, \mu_2 \mid \boldsymbol{x}_1, \boldsymbol{x}_2) = p(\mu_1 \mid \boldsymbol{x}_1)\, p(\mu_2 \mid \boldsymbol{x}_2)$$

$$T_i = \sqrt{n_i - 1}\left(\frac{\mu_i - \bar{x}_i}{s_i} \right) \sim St(t_i \mid n_i - 1) \qquad i = 1,2$$

so, instead of sampling from $p(w \mid \boldsymbol{x}_1, \boldsymbol{x}_2)$

Algorithm

n)

1) $t_i \sim St(t_i \mid n_i - 1) \qquad \mu_i = \bar{x}_i + t_i s_i (n_i - 1)^{-1/2} \qquad i = 1,2$

2) $w = \mu_1 - \mu_2$

Basic Idea:

We want a sampling of $\boldsymbol{X} = \left(X_1, X_2, \ldots, X_n\right) \sim p\left(x_1, x_2, \ldots, x_n\right)$

marginal densities $p(s_i) = \int_{\Omega} p(x_1, x_2, \ldots, x_n)\, dx_i$

$$s_i = \{x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n\}$$

conditional densities $p(x_i \mid s_i) = \dfrac{p(x_1, x_2, \ldots, x_n)}{p(s_i)}$

2) Sample the n-dimensional random quantity $\mathbf{X}$
from the conditional distributions

$$p\left(x_1 \mid x_2, x_3, \ldots, x_n\right)$$
$$p\left(x_2 \mid x_1, x_3, \ldots, x_n\right)$$
$$\vdots$$
$$p\left(x_n \mid x_1, x_2, \ldots, x_{n-1}\right)$$

First, sampling from conditional densities:

$$p(\mathbf{x} \to \mathbf{x}') = q(\mathbf{x}' \mid \mathbf{x})\, a(\mathbf{x} \to \mathbf{x}')$$

Consider:

1) The probability density function

$$q\left(x_1, x_2, x_3, \ldots, x_n\right)$$

2) The conditional densities

$$q\left(x_1 \mid x_2, x_3, \ldots, x_n\right)$$
$$q\left(x_2 \mid x_1, x_3, \ldots, x_n\right)$$
$$\vdots$$
$$q\left(x_n \mid x_1, x_2, \ldots, x_{n-1}\right)$$

3) An arbitrary initial state

$$(x_1(0), x_2(0), \ldots, x_n(0)) \in \Omega_X$$

(usually less than n are needed)

4) Sampling of the $X_k\ ; k = 1, \ldots, n$ random quantities at step t

step $t-1$ $\quad (x_1(t-1), x_2(t-1), \ldots, x_n(t-1)) \in \Omega_Z$

step t $\quad$ *Generate a possible new value* $x_k(t)$ *of* X_k *from the conditional density*

$$q\left(x_k \mid \underbrace{x_1(t), x_2(t), \ldots, x_{k-1}(t)}_{t}, \underbrace{x_{k+1}(t-1), \ldots, x_n(t-1)}_{t-1}\right)$$

- *Propose a change of the system from*

the state $s = \left(x_1(t), x_2(t), \ldots, x_{k-1}(t), x_k(t-1), x_{k+1}(t-1), \ldots, x_n(t-1)\right)$

to

the state $s' = \left(x_1(t), x_2(t), \ldots, x_{k-1}(t), x_k(t), x_{k+1}(t-1), \ldots, x_n(t-1)\right)$

Metropolis-Hastings: ***Desired density*** $p\left(x_1, x_2, \ldots, x_n\right)$

Acceptance factor

$$\alpha = \frac{p\left(x_1(t), x_2(t), \ldots, x_{k-1}(t), x_k(t), x_{k+1}(t-1), \ldots, x_n(t-1)\right)}{p\left(x_1(t), x_2(t), \ldots, x_{k-1}(t), x_k(t-1), x_{k+1}(t-1), \ldots, x_n(t-1)\right)} \times$$

$$\times \frac{q\left(x_k(t-1) \mid x_1(t), x_2(t), \ldots, x_{k-1}(t), x_{k+1}(t-1), \ldots, x_n(t-1)\right)}{q\left(x_k(t) \mid x_1(t), x_2(t), \ldots, x_{k-1}(t), x_{k+1}(t-1), \ldots, x_n(t-1)\right)}$$

so we shall accept the change with probability $a(s \to s') = \min(1, \alpha)$

At the end of step t ➡ $(x_1(t), x_2(t), \ldots, x_n(t)) \in \Omega_X$

the sequences $\left\{x_1, x_2, \ldots, x_n\right\}$

will converge towards the stationary p.d.f $p\left(x_1, x_2, \ldots, x_n\right)$

regardless the starting sampling $(x_1(0), x_2(0), \ldots, x_n(0)) \in \Omega_X$

In particular,… Gibbs algorithm…

Gibbs Algorithm: ***Take as conditional densities to generate a proposed value the desired conditional densities***

$$q\left(x_k \mid x_1(t), x_2(t), \ldots, x_{k-1}(t), x_{k+1}(t-1), \ldots, x_n(t-1)\right) =$$
$$= p\left(x_k \mid x_1(t), x_2(t), \ldots, x_{k-1}(t), x_{k+1}(t-1), \ldots, x_n(t-1)\right)$$

acceptance factor:

$$q\left(x_k(t-1) \mid x_1(t), x_2(t), \ldots, x_{k-1}(t), x_{k+1}(t-1), \ldots, x_n(t-1)\right) =$$
$$= p\left(x_k(t-1) \mid x_1(t), x_2(t), \ldots, x_{k-1}(t), x_{k+1}(t-1), \ldots, x_n(t-1)\right) =$$
$$= \frac{p\left(x_1(t), x_2(t), \ldots, x_{k-1}(t), x_k(t-1), x_{k+1}(t-1), \ldots, x_n(t-1)\right)}{p\left(x_1(t), x_2(t), \ldots, x_{k-1}(t), x_{k+1}(t-1), \ldots, x_n(t-1)\right)}$$

$$q\left(x_k(t) \mid x_1(t), x_2(t), \ldots, x_{k-1}(t), x_{k+1}(t-1), \ldots, x_n(t-1)\right) =$$
$$= p\left(x_k(t) \mid x_1(t), x_2(t), \ldots, x_{k-1}(t), x_{k+1}(t-1), \ldots, x_n(t-1)\right) =$$
$$= \frac{p\left(x_1(t), x_2(t), \ldots, x_{k-1}(t), x_k(t), x_{k+1}(t-1), \ldots, x_n(t-1)\right)}{p\left(x_1(t), x_2(t), \ldots, x_{k-1}(t), x_{k+1}(t-1), \ldots, x_n(t-1)\right)}$$

$$\alpha = 1$$

So we accept the "proposed" change with probability $a(s \to s') = 1$

After enough steps to erase the effect of the initial values in the samplings and to achieve a good approximation to the asymptotic limit, we shall have sequences

$$\left(x_1^{1)}, x_2^{1)}, \ldots, x_n^{1)}\right), \ldots, \left(x_1^{m)}, x_2^{m)}, \ldots, x_n^{m)}\right) \quad \sim p(x_1, x_2, \ldots, x_n)$$

Example: $\boldsymbol{X} = \{X_1, \dots, X_n\} \sim Di(\boldsymbol{x} \mid \boldsymbol{\alpha})$

Conjugated prior for Multinomial:

$$p(\boldsymbol{x} \mid \boldsymbol{\alpha}) = \frac{\Gamma(\alpha_0)}{\prod_{k=1}^{n} \Gamma(\alpha_k)} x_1^{\alpha_1 - 1} x_2^{\alpha_2 - 1} \cdots x_n^{\alpha_n - 1}$$

$$x_k \in [0,1] \qquad \sum_{k=1}^{n} x_k = 1$$

$$\alpha_k > 0 \qquad \sum_{k} \alpha_k = \alpha_0$$

$$p(x_k \mid s_k, \boldsymbol{\alpha}) \propto x_k^{\alpha_k - 1} (1 - S_k - x_k)^{\alpha_n - 1} \quad \longrightarrow \quad z_k = \frac{x_k}{1 - S_k} \sim Be(z \mid \alpha_k, \alpha_n)$$

$$s_k = \{x_1, x_2, \dots, x_{k-1}, x_{k+1}, x_{n-1}\}$$

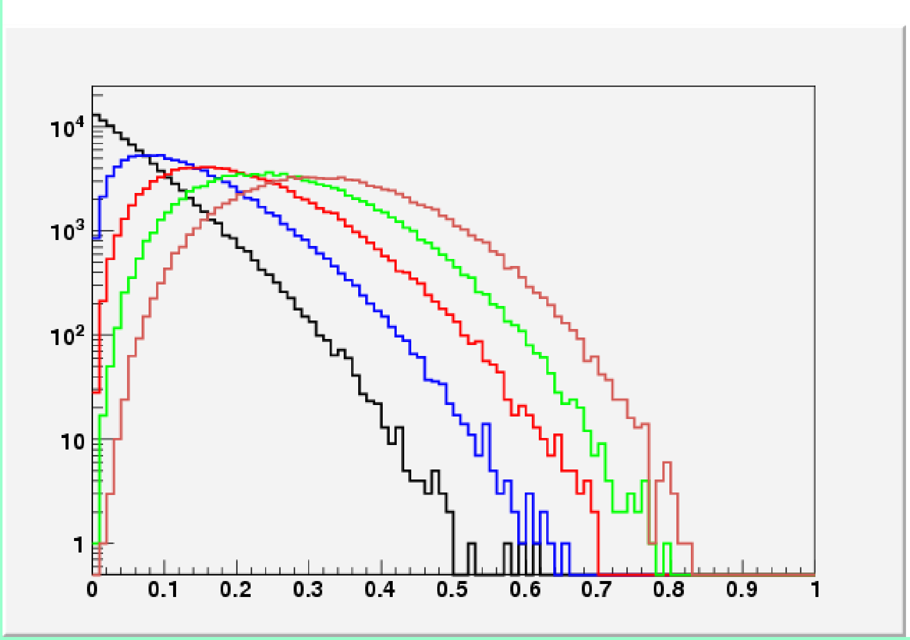
$$S_k = \sum_{\substack{j=1 \\ j \neq k}}^{n} x_j$$

$$n = 5$$

$$\boldsymbol{\alpha} = \{1, 2, 3, 4, 5\}$$

$$\boldsymbol{x}^{(0)} = \{0.211, 0.273, 0.262, 0.101, 0.152\}$$

$$\sum_k x_k^{(0)} = 1$$



Example: Dirichlet and Generalised Dirichlet

Conjugated priors for Multinomial:

$$\boldsymbol{X} = \{X_1,\ldots,X_n\} \sim Di(\mathbf{x}\,|\,\boldsymbol{\alpha}) \qquad p(\mathbf{x}\,|\,\boldsymbol{\alpha}) = D(\boldsymbol{\alpha})\; x_1^{\alpha_1-1} x_2^{\alpha_2-1}\cdots x_n^{\alpha_n-1}$$

$$\alpha_k > 0 \qquad D(\boldsymbol{\alpha}) = \frac{\Gamma(\alpha_0)}{\prod_{k=1}^{n}\Gamma(\alpha_k)} \qquad \alpha_0 = \sum_{k=1}^{n}\alpha_k$$

$$x_k \in [0,1] \qquad \sum_{k=1}^{n} x_k = 1 \;\longrightarrow\; x_n = 1 - \sum_{k=1}^{n-1} x_k$$

Degenerated Distribution:

$$p(\mathbf{x}\,|\,\boldsymbol{\alpha}) = D(\boldsymbol{\alpha})\left(\prod_{i=1}^{n-1} x_i^{\alpha_i-1}\right)\left(1-\sum_{k=1}^{n-1}x_k\right)^{\alpha_n-1}$$

$$E[X_i] = \frac{\alpha_i}{\alpha_0} \qquad V[X_i,X_j] = \frac{\alpha_i\alpha_0\delta_{ij} - \alpha_i\alpha_j}{\alpha_0^2(\alpha_0+1)}$$

$$Z_k \sim Ga(1,\alpha_k) \qquad X_j = \frac{Z_j}{\sum_{k=1}^{n} Z_k} \qquad \boldsymbol{X} \sim \{X_1,\ldots,X_n\} \sim Di(\mathbf{x}\,|\,\boldsymbol{\alpha})$$

$$\boldsymbol{X} \sim \{X_1,\ldots,X_n\} \sim GDi(\boldsymbol{x}\,|\,\boldsymbol{\alpha},\boldsymbol{\beta})$$

$$p(\boldsymbol{x}\,|\,\boldsymbol{\alpha}) = \prod_{i=1}^{n-1} \frac{\Gamma(\alpha_i+\beta_i)}{\Gamma(\alpha_i)\Gamma(\beta_i)} x_i^{\alpha_i-1}\left(1-\sum_{k=1}^{i} x_k\right)^{\gamma_i}$$

$$0 < x_i < 1 \qquad \sum_{k=1}^{n-1} x_k < 1 \qquad x_n = 1-\sum_{k=1}^{n-1} x_k$$

$$\alpha_k > 0 \qquad \beta_k > 0 \qquad \gamma_i = \begin{cases} \beta_i - \alpha_{i+1} - \beta_{i+1}\,; & i = 1,2,\ldots,n-2 \\ \beta_{n-1} - 1\,; & i = n-1 \end{cases}$$

$$E[X_i] = \frac{\alpha_i}{\alpha_i+\beta_i} S_i \qquad S_i = \prod_{k=1}^{i-1} \beta_k(\alpha_k+\beta_k)^{-1}$$

$$V[X_i] = E[X_i]\left(\frac{\alpha_i+\delta_{ij}}{\alpha_i+\beta_i+1} T_i - E[X_i]\right) \qquad T_i = \prod_{k=1}^{i-1} (\beta_k+1)(\alpha_k+\beta_k+1)^{-1}$$

$$X_1 \sim Z_1$$

$$X_2\,|\,X_1 \sim Z_2(1-X_1)$$

$$X_3\,|\,X_1, X_2 \sim Z_3(1-X_1-X_2)$$

$$X_k = Z_k\left(1-\sum_{j=1}^{k-1} X_j\right) \qquad Z_k \sim Be(\alpha_k,\beta_k)$$

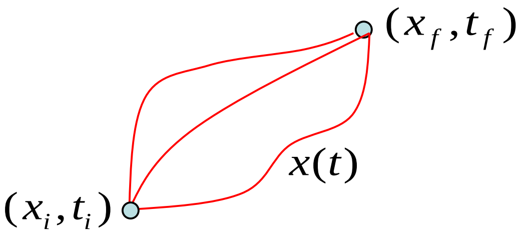
$$k = 1,\ldots,n-1$$

$$X_n = 1-\sum_{j=1}^{n-1} X_j$$

$$\boldsymbol{X} \sim \{X_1,\ldots,X_n\} \sim GDi(\boldsymbol{x}\,|\,\boldsymbol{\alpha},\boldsymbol{\beta})$$

Quantum Mechanics

Path Integral formulation of Q.M. (R.P. Feynman)



Probability Amplitude

$$K(x_f, t_f | x_i, t_i) \propto \int_{paths} e^{i/\hbar \, S[x(t)]} D[x(t)]$$

$$S[x(t)] = \int_{t_i}^{t_f} L(\dot{x}, x; t)\, dt$$

Propagator

$$\Psi(x_f, t_f) = \int K(x_f, t_f | x_f, t_f) \Psi(x_i, t_i) dx_i \quad ; \quad t_f > t_i$$

Local Lagrangians (additive actions)

$$K(x_f, t_f | x_i, t_i) = \int K(x_f, t_f | x, t) K(x, t | x_i, t_i) dx$$

Chapman-Kolmogorov

Feynman-Kac Expansion Theorem

$$K(x_f, t_f | x_i, t_i) = \sum_n e^{-i/\hbar \, E_n (t_f - t_i)} \phi_n(x_f) \phi_n^*(x_i)$$

Expected Value of an operator $A(x)$

$$\langle A \rangle = \frac{\int A[x(t)] e^{i/\hbar \, S[x(t)]} D[x(t)]}{\int e^{i/\hbar \, S[x(t)]} D[x(t)]}$$

One Dimensional Paticle

$$L(\dot{x}, x; t) = \frac{1}{2} m\, \dot{x}(t)^2 - V[x(t)]$$

Wick Rotation

$$\tau = e^{i\pi/2} t = it \qquad t \rightarrow -i\tau$$

$$S[x(t)] \rightarrow i \int_{\tau_i}^{\tau_f} \left(\frac{1}{2} m\, \dot{x}(t)^2 + V[x(t)] \right) dt$$

Feynman-Kac expansion theorem

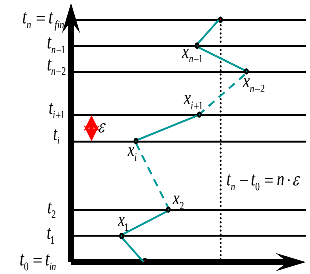
$$K(x_f, \tau_f | x_i, \tau_i) = \sum_n e^{-E_n(\tau_f - \tau_i)} \phi_n(x_f) \phi_n^*(x_i)$$

$$x_f = x_i = 0$$
$$\tau_i = 0$$

$$K(0, \tau_f | 0,0) = \sum_n e^{-E_n \tau_f} \phi_n(0) \phi_n^*(0)$$

$$\approx e^{-E_1 \tau_f} \phi_1(0) \phi_1^*(0) + \dots$$

Discretised time



$$D[x(t)] \rightarrow A_N \prod_{j=1}^{N-1} dx_j$$

$$S_N(x_0, x_1, \ldots, x_N) = \varepsilon \sum_{j=1}^{N} \left(\frac{1}{2} m \left(\frac{x_j - x_{j-1}}{\varepsilon} \right)^2 + V(x_j) \right)$$

$$K(x_f, t_f | x_i, t_i) = A_N \int_{x_1} dx_1 \cdots \int_{x_{N-1}} dx_{N-1} \exp\{- S_N(x_0, x_1, \ldots, x_N)\}$$

$$\langle A \rangle = \frac{\int \prod_{j=1}^{N-1} dx_j \, A(x_0, x_1, \ldots, x_N) \exp\{- S_N(x_0, x_1, \ldots, x_N)\}}{\int \prod_{j=1}^{N-1} dx_j \, \exp\{- S_N(x_0, x_1, \ldots, x_N)\}}$$

$$\langle A \rangle = \frac{\int \prod_{j=1}^{N-1} dx_j \; A(x_0, x_1, \ldots, x_N) \exp\{-S_N(x_0, x_1, \ldots, x_N)\}}{\int \prod_{j=1}^{N-1} dx_j \; \exp\{-S_N(x_0, x_1, \ldots, x_N)\}}$$

Goal: generate N_{tray} as

(importance sampling)

$$p(x_0, x_1, \ldots, x_N) \propto \exp\{-S_N(x_0, x_1, \ldots, x_N)\}$$

and estimate $\langle A \rangle = \sum_{k=1}^{N_{tray}} A(x_0, x_1^{(k)}, \ldots, x_{N-1}^{(k)}, x_N)$

Very complicated…

Markov Chain Monte Carlo

Harmonic Potential

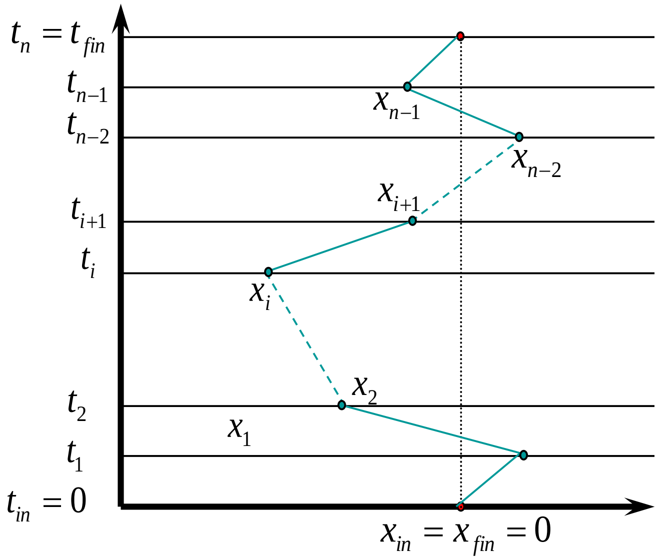
$$V[x(t)] = \frac{1}{2} k \; x(t)^2$$

$$S_N(x_0, x_1, \ldots, x_N) = \frac{\varepsilon}{2} \sum_{j=1}^{N} \left(m \left(\frac{x_j - x_{j-1}}{\varepsilon} \right)^2 + k \; x_j^2 \right)$$

Parameters

$x_0 = x(t_{in}) = 0$ $x_j \in R \quad ; \quad j = 1, \ldots, N-1$

$x_N = x(t_{fin}) = 0$ $\ldots x_j \in [-10, 10]$



$\varepsilon = 0.25 \qquad (\varepsilon << \to \approx t \; continuo)$

$N = 2000 \qquad (N >> \to \underbrace{\tau_{fin} = N\varepsilon >>)}$

To isolate fundamental state

Termalisation and correlations

$N_{term} = 1000$

$N_{tray} = 3000$

$N_{util} = 1000$

Generación de trayectorias

1) Generate initial trajectory $(x_0, x_1^{(0)}, \ldots, x_{N-1}^{(0)}, x_N)$

2) Sweep over $x_1, \ldots, x_{N-1}$

N-1 veces

$$x'_j \sim Un(x|-10,10)$$

$$P(x_j \to x'_j) = \exp\{-S_N(x_0, x_1, \ldots, x'_j, \ldots, x_N)\}$$

$$P(x_j \to x_j) = \exp\{-S_N(x_0, x_1, \ldots, x_j, \ldots, x_N)\}$$

$$a(x_j \to x'_j) = min\left[1, \frac{P(x_j \to x'_j)}{P(x_j \to x_j)}\right]$$

$$= min\left[1, \exp\{S_N(x_{j-1}, x_j x_{j+1}) - S_N(x_{j-1}, x'_j x_{j+1})\}\right]$$

3) Repeat 2) $N_{term} = 1000$ times

4) Repeat 2) $N_{tray} = 3000$ times and take one trajectory out of 3

$N_{util} = 1000$ trajectories $(x_0 = 0, x_1, \ldots, x_{N-1}, x_N = 0)$

Virial Theorem

$$\langle T \rangle_\Psi = \frac{1}{2} \langle \vec{x} \cdot \vec{\nabla} V(\vec{x}) \rangle_\Psi$$

Harmonic Potential

$$V[x(t)] = \frac{1}{2} k \ x(t)^2$$

$$\langle T \rangle_\Psi = \langle V \rangle_\Psi = \frac{1}{2} k \langle x^2 \rangle_\Psi$$

$$\langle E \rangle_\Psi = \langle T \rangle_\Psi + \langle V \rangle_\Psi = k \langle x^2 \rangle_\Psi$$

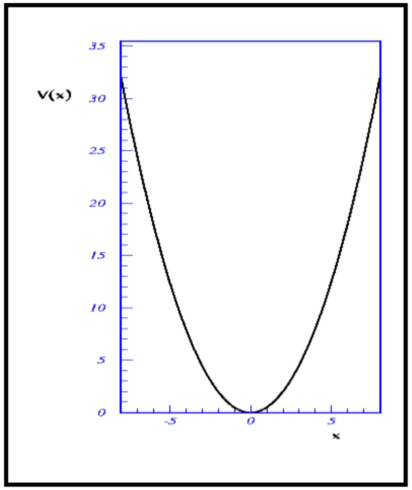
X^4 Potential

$$V[x(t)] = \frac{a^2}{4} \left[\left(\frac{x}{a} \right)^2 - 1 \right]^2$$

$$\langle T \rangle_\Psi = \frac{a^2}{2} \left(\frac{x}{a} \right)^2 \left[\left(\frac{x}{a} \right)^2 - 1 \right]$$

$$\langle E \rangle_\Psi = \frac{3}{4a^2} \langle x^4 \rangle_\Psi - \langle x^2 \rangle_\Psi + \frac{a^2}{4}$$

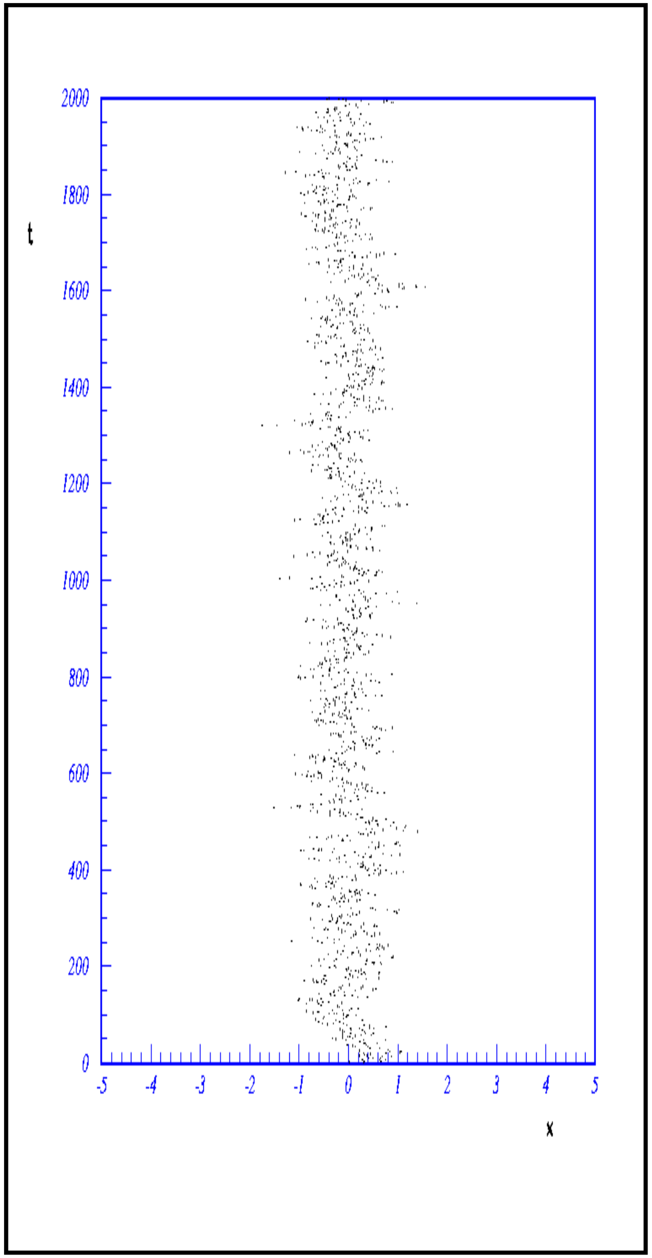
Harmonic Potential

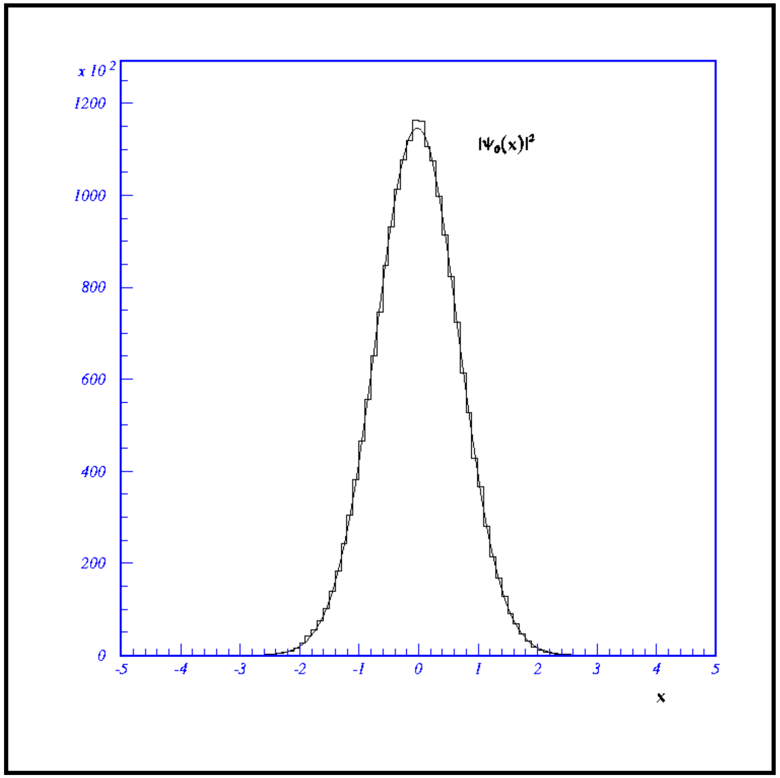


$$k = 1$$

$$\langle E \rangle_0 = \langle x^2 \rangle_0 = 0.486$$

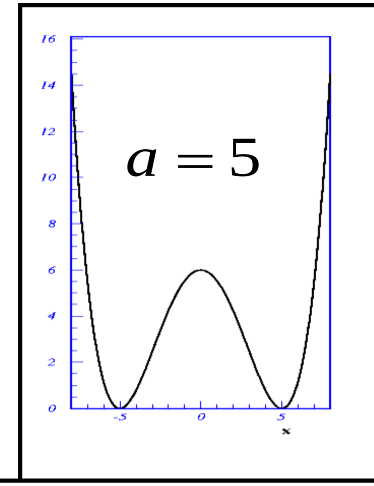
$$\langle E \rangle_0^{exact} = 0.5$$





X^4 Potential

$$V[x(t)] = \frac{a^2}{4}\left[\left(\frac{x(t)}{a}\right)^2 - 1\right]$$



$\varepsilon = 0.25$

$N = 9000$

$$\langle E \rangle_0 = \frac{3}{4a^2}\langle x^4 \rangle_0 - \langle x^2 \rangle_0 + \frac{a^2}{4} = 0.668$$

$$\langle E \rangle_0^{exact} = 0.697$$