I. INTRODUCTION

In our study of mathematical statistics, we have learnt to use data to infer the values of some parameter specifying the law of a random variable modelling the outcomes of the experiment. A set of data:

$$(\underline{x}_1, \ldots, \underline{x}_n) \tag{1}$$

are interpreted as realizations of a sample $(X_1, \ldots, X_n)$, that is:

$$(\underline{x}_1, \ldots, \underline{x}_n) = (X_1(\omega), \ldots, X_n(\omega)) \tag{2}$$

for a ω in some abstract probability space $(\Omega, \mathcal{F}, P)$ where the random variables are defined. This interpretation allows then to define estimators and confidence intervals, to test hypothesis and so on.

In this chapter we will take the opposite point of view: given the law of a random variable X, or, equivalently, its density $p(\underline{x})$ (if it exists), is it possible to *generate* possible realizations of X? This is a central topic in the realm of simulations, and is usually called the **sampling** of random variables, or, equivalently, the **sampling** of probability densities.

II. MONTE CARLO INTEGRATION

One very important application of the sampling of probability densities is **Monte Carlo integration**. It is an extremely useful tool to evaluate integrals arising, for example, from statistical physics and quantum mechanics. It becomes quite the unique way to face multi-dimensional integrals, when typical strategies would require an extremely huge number of operations, beyond the possibility of any computer.

To state the problem, let's consider an integral of the very general form:

$$I = \int_D d\underline{x}\, f(\underline{x})\, p(\underline{x}) \tag{3}$$

where $D \subset \mathbb{R}^d$, f is any (measurable) function and p is a probability density on D:

$$p(\underline{x}) \geq 0, \quad \int_D d\underline{x}\, p(\underline{x}) = 1 \tag{4}$$

The key observation is that, if X is a random variable having $p(\underline{x})$ as its probability density, the following equalities holds:

$$I = E\left[f(X)\right] \tag{5}$$

$$\int_D d\underline{x}\, \left(f(\underline{x}) - I\right)^2 p(\underline{x}) = Var\left(f(X)\right) \tag{6}$$

The law of large numbers and the central limit theorem, provided that $E\left[f(X)\right]$ and $Var\left(f(X)\right)$ are finite, guarantee that, if $\{X_i\}_{i=1}^{+\infty}$ is a sequence of **independent and identically distributed** random variables with density $p(\underline{x})$, we have:

$$I = \lim_{n \to +\infty} \frac{1}{n} \sum_{i=1}^{n} f(X_i) \tag{7}$$

and:

$$\frac{\frac{1}{n} \sum_{i=1}^{n} f(X_i) - I}{\sqrt{\frac{Var(f(X))}{n}}} \tag{8}$$

converges in distribution to a standard normal random variable $N(0,1)$. Thus, if we are able to sample $p(\underline{x})$, i.e., in practice, to generate n points in D:

$$(\underline{x}_1, \ldots, \underline{x}_n) \tag{9}$$

realizations of $\{X_i\}_{i=1}^n$:

$$(\underline{x}_1, \ldots, \underline{x}_n) = (X_1(\omega), \ldots, X_n(\omega)) \tag{10}$$

for a ω in some abstract probability space $(\Omega, \mathcal{F}, P)$ where the random variables are defined, then we can evaluate:

$$I \simeq \frac{1}{n} \sum_{i=1}^{n} f(\underline{x}_i) \tag{11}$$

and use mathematical statistics to estimate confidence intervals and so on.

Precisely in the same way, finite or infinite summations:

$$I = \sum_{\underline{x}} f(\underline{x})\, p(\underline{x}) \tag{12}$$

can be dealt with whenever $p(\underline{x})$ is a discrete probability density.

To summarize, the problem of evaluating an integral is transferred into the problem of building up a (possibly large) number of point $(\underline{x}_1, \ldots, \underline{x}_n)$ starting from the knowledge of a probability density $p(\underline{x})$. Such problem is in general not trivial and requires the formalism of stochastic processes. We start from the basic *generation of random numbers*.

III. RANDOM NUMBER GENERATORS

Our starting point is the existence of the **random number generators**, which are algorithms able to sample a sequence of **independent uniform in** $(0,1)$ random variables. The output of such an algorithm is a sequence:

$$(u_1, \ldots, u_n) \tag{13}$$

$0 < u_i < 1$, realizations of n **independent uniform in** $(0,1)$ random variables:

$$(U_1, \ldots, U_n) \tag{14}$$

We are not going now to enter the details of the theory of random number generation, requiring complex notions of numbers theory beyond the scope of this lectures. We simply mention the simplest algorithm, the **linear congruential generator** (LCG), introduced by D.H. Lehmer in 1949, which builds up the sequence $(u_1, \ldots, u_n)$ using the integers:

$$i_{j+1} = (ai_j + c) \pmod{m}, \quad j = 0, \ldots, n \tag{15}$$

where $m, a, c \in \mathbb{N}$ are positive integer numbers, called **modulus, multiplicator and increment**, while the starting term, i_0, is a non-negative integer called the **seed** of the generator; finally the u_j are obatined as $u_j = i_j/(m-1)$. In the following table we report typical values for the parameters m, a, c:

Table I: parameters m, a, c of the LCG

Source	m	a	c
gclib	2^{31}	1103515245	12345
Numerical Recipes	2^{32}	1664525	1013904223
java.util.Random	2^{48}	25214903917	11

The reader could feel a bit confused now, since we have claimed independence while actually obtaining the sequence applying a deterministic (and very simple!) function to a given number to obtain the following one. This is the reason for the choices of parameter in the given table, providing the conditions for the data $(u_1, \ldots, u_n)$ to be modelled by independent random variables. Statistical and numerical studies have shown that such choices of parameters make the model very accurate, in the sense discussed in the chapter about statistics.

The other important point is the seed i_0: it can be chosen to be equal to any non-negative integer number. If a program is used twice with the same seed, it gives exactly the same output. Actually the seed can be thought as the point $\omega \in \Omega$ in some abstract probability space $(\Omega, \mathcal{F}, P)$ determining the output of the "experiment".

$$(u_1, \ldots, u_n) = (U_1(\omega), \ldots, U_n(\omega)) \tag{16}$$

IV. SIMULATION OF NORMAL RANDOM VARIABLES

We have thus learned that, with a very simple alogorithm, we can **sample** the uniform distribution in $(0, 1)$. What about other probability densities? In general, this problem is highly non trivial and requires the formalism of stochastic processes which we will discuss in the following chapters. Nevertheless, there are some situations allowing to solve the problem in a simple and elegant way, relying on transformations between random variables. Due to the outstanding importance of normal random variables, our first example is the sampling the density $N(0, 1)$:

$$p(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \tag{17}$$

We will exploit the transformation law for densities:

$$p_Y(\underline{y}) = p_X(g^{-1}(\underline{y}))\left|\det(J_{g^{-1}}(\underline{y}))\right| \tag{18}$$

valid whenever $Y = g(X)$, g being a diffeomorphism between open subsets of $\mathbb{R}^d$. $J_{g^{-1}}(\underline{y})$ is the Jacobian matrix of the inverse g^{-1}.

We specialize the transformation law to the special case in two dimensions $X = (U_1, U_2)$, U_1, U_2 being **independent uniform in** $(0, 1)$ and:

$$g(u_1, u_2) = \left(\sqrt{-2\log(u_1)}\cos(2\pi u_2), \sqrt{-2\log(u_1)}\sin(2\pi u_2)\right) \tag{19}$$

We let $Y = (Y_1, Y_2)$ and evaluate its density. The inverse if g is simply checked to be:

$$g^{-1}(y_1, y_2) = \left(\exp\left(-\frac{y_1^2 + y_2^2}{2}\right), \frac{1}{2\pi} artg\left(\frac{y_2}{y_1}\right)\right) \tag{20}$$

while its Jacobian is given by:

$$J_{g^{-1}}(y_1, y_2) = \begin{pmatrix} -y_1 \exp\left(-\frac{y_1^2+y_2^2}{2}\right) & -y_2 \exp\left(-\frac{y_1^2+y_2^2}{2}\right) \\ -\frac{y_2/y_1^2}{2\pi\left(1+y_2^2/y_1^2\right)} & \frac{1/y_1}{2\pi\left(1+y_2^2/y_1^2\right)} \end{pmatrix} \tag{21}$$

The determinant is:

$$\det\left(J_{g^{-1}}(y_1, y_2)\right) = -\frac{(1 + y_2^2/y_1^2)}{2\pi\,(1 + y_2^2/y_1^2)} \exp\left(-\frac{y_1^2 + y_2^2}{2}\right) \tag{22}$$

implying that $Y = (Y_1, Y_2) = g(U_1, U_2)$ has density:

$$p_Y(y_1, y_2) = \frac{1}{2\pi} \exp\left(-\frac{y_1^2 + y_2^2}{2}\right) \tag{23}$$

which means that Y_1 and Y_2 are **independent standard normal random variables**.

Thus, in practice, in order to sample one standard normal random variable, it is possible to use a random number generator twice, obtaining two numbers (u_1, u_2), and to apply the following **Box-Muller formula**:

$$y = \sqrt{-2\log(u_1)}\cos(2\pi u_2) \tag{24}$$

V. THE INVERSE CUMULATIVE DISTRIBUTION FUNCTION

We present now another very important example of the possibility of sampling one-dimensional random variables given a random number generator. Let's consider a given probability density $p(x)$ on $\mathbb{R}$ and let $F(x) = \int_{-\infty}^{x} dy\, p(y)$, the cumulative distribution function of a random variable having $p(x)$ as probability density. We work under the hypothesis that there exists an interval (α, β), $-\infty \leq \alpha < \beta \leq +\infty$ such that

$p(x) > 0$ for $x \in (\alpha, \beta)$ and $p(x) = 0$ outside that interval. $F(x)$ is thus strictly increasing on (α, β) and its values lie in $[0, 1]$. We define now $Y = F^{-1}(U)$ where U is uniform in $(0, 1)$. The key point is that the cumulative distribution of Y coincides with $F(x)$, in fact:

$$F_Y(y) = P(Y \leq y) = P(F^{-1}(U) \leq y) = P(U \leq F(y)) = F(y) \tag{25}$$

and thus:

$$p_Y(y) = p(y) \tag{26}$$

This means that we can sample any one-dimensional probability density $p(x)$ using a random number generator if we are able to evaluate F^{-1}: the generator provides a realization of a uniform random variable U, and, if we apply F^{-1}, we obtain a realization of a random variable with density $p(x)$.

$$\omega \rightsquigarrow U(\omega) \rightsquigarrow F^{-1}(U(\omega)) \tag{27}$$

For example, if we wish to sample the **lorentzian** probability density:

$$p(x) = \frac{1}{\pi} \frac{\Gamma}{\Gamma^2 + x^2} \tag{28}$$

we evaluate:

$$F(x) = \int_{-\infty}^{x} dy\, p(y) = \frac{1}{\pi} artg\left(\frac{x}{\Gamma}\right) + \frac{1}{2} \tag{29}$$

We know that, if U is uniform in $(0, 1)$:

$$Y = F^{-1}(U) = \Gamma \tan\left(\pi\left(U - \frac{1}{2}\right)\right) \tag{30}$$

has density $p(x)$.

As another example, if $p(x)$ is the **exponential density** with parameter λ:

$$p(x) = \lambda \exp(-\lambda x)\, 1_{(0,+\infty)}(x), \quad \lambda > 0 \tag{31}$$

we evaluate:

$$F(x) = \int_{-\infty}^{x} dy\, p(y) = (1 - \exp(-\lambda x))\ 1_{(0,+\infty)}(x) \tag{32}$$

so that, if U is uniform in $(0, 1)$:

$$Y = F^{-1}(U) = -\frac{1}{\lambda} \log(1 - U) \tag{33}$$

has density $p(x)$.

VI. DISCRETE RANDOM VARIABLES

We discuss now the typical situation when we wish to sample a discrete probability density $p(x)$, non-zero only in the discrete set $\{x_1, \ldots, x_n\}$, which we assume to be finite. The typical tool we can use is the following: we define $q_0 = 0$, $q_1 = p(x_1)$, $q_2 = p(x_1) + p(x_2)$, $q_{m-1} = p(x_1) + p(x_2) + \cdots + p(x_{m-1})$ and, finally, $q_n = 1$. We have naturally $0 = q_0 < \cdots < q_n = 1$. If U is uniform in $(0, 1)$, we define:

$$Y = x_j, \quad if \quad q_{j-1} \leq U \leq q_j \tag{34}$$

Y is clearly a discrete random variable and has precisely the discrete density $p(x)$, as follows from the following simple calculation:

$$P(Y = x_j) = P(q_{j-1} \leq U \leq q_j) = q_j - q_{j-1} = p(x_j) \tag{35}$$

VII. PERSPECTIVES

In the previous section we have presented some tools to sample random variables, once a random number generator is available. Unfortunately, these tools are, in general, not available in the multidimensional case, which is the most interesting situation for Monte Carlo integration to be used. For example, in classical statistical mechanics, one wishes, for example, to sample the Boltzmann weight of a classical fluid in thermal equilibrium at temperature $T = 1/\beta$:

$$p(\vec{r}_1, \ldots, \vec{r}_N) = \frac{\exp\left(-\beta \sum_{i<j} u\left(|\vec{r}_i - \vec{r}_j|\right)\right)}{\mathcal{Z}} \tag{36}$$

$u(r)$ being the interatomic potential. Another example is the Ising model, desribing a collection of *spins* $(\sigma_1, \ldots, \sigma_N)$, $\sigma_i = \pm 1$ on an hypercubic lattice, whose equilibrium properties, at temperature $T = 1/\beta$ and at the presence of a magnetic field B, are described by the (discrete) probability density:

$$p(\sigma_1, \ldots, \sigma_N) = \frac{\exp\left(-\beta\left(-\sum_{\langle i,j\rangle} J\sigma_i\sigma_j - B\sum_i \sigma_i\right)\right)}{Z} \tag{37}$$

$J > 0$ describing a ferromagnetic coupling between the nearest neighbours spins (the simble $\langle i, j\rangle$ denotes nearest neighbours).

Using the Ising model as a guiding example, we wish for example to estimate the magnetization of the system, as a function of the temperature and the magnetic field:

$$m(\beta, B) = \frac{1}{N}\sum_{i=1}^{N}\langle\sigma_i\rangle \tag{38}$$

where we have used the common notation of statistical physics for the expected value:

$$\langle \sigma_i \rangle = E\left[\sigma_i\right] = \sum_{(\sigma_1,\ldots,\sigma_N)} \sigma_i \, p(\sigma_1,\ldots,\sigma_N) \tag{39}$$

This summation (2^N terms) is not feasible when N becomes large, requiring Monte Carlo integration techniques.

If we are able to sample $p(\sigma_1,\ldots,\sigma_N)$, then we can estimate the magnetization. The reader will notice that none of the techniques presented in this chapter is suitable. We will learn to perform such a sampling, but we will need the formalism of stochastic processes that we will present in the next chapter.