Stochastic Simulation.

Alberto Policriti

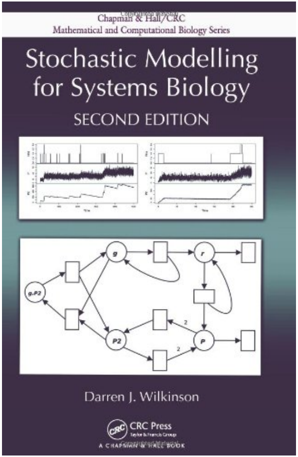
Dipartimento di Matematica e Informatica
Istituto di Genomica Applicata

Summary

1. Continuous probability models

2. Monte-Carlo simulation

Material from



Basic definitions

Continuous random quantities

Random quantities with a continuous sample space

... too *large* to define the probability of a single event.

Definition (Probability density function)

X continuous random quantity, the Probability Density Function (PDF) $f_X(x)$ is defined as a function such that:

1. $f_X(x) \geq 0$ for all x
2. $\int_{-\infty}^{\infty} f_X(x)dx = 1$
3. $P(a \leq x \leq b) = \int_a^b f_X(x)dx$ for any $a \leq b$

... densities can be greater than 1 (as long as they integrate to 1)

Definition (Cumulative distribution function)

X continuous random quantity with PDF $f_X(x)$, the Cumulative Distribution Function (CDF) is defined as a function $F_X(x)$ such that:

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= P(-\infty \leq X \leq x) \\ &= \int_{-\infty}^{x} f_X(z)dz \end{aligned}$$

... as the first derivative of $F_X(x)$ is $f_X(x)$: *the slope of the CDF is the PDF*

Definition (Expectation and variance)

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} x f_X(x) dx \\
Var(X) &= \int_{-\infty}^{\infty} [x - E(X)]^2 f_X(x) dx
\end{aligned}
$$

The Law of large numbers

Suppose we want to approximate X by $X_1, \ldots, X_n$ independent realizations of X (our *runs*). Namely we consider the quantity:

Sample mean

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

... the key is to get n *large enough* to ensure—under appropriate conditions—that $\bar{X}$ is a good estimate of $E(X)$.

Lemma

Assume $E(X) = \mu$ and $Var(X) = \sigma^2$ finite, and $\bar{X}$ defined as above, then

$$E(\bar{X}) = \mu \quad \textit{and} \quad Var(\bar{X}) = \frac{\sigma^2}{n}.$$

Inequalities

Lemma (Markov's inequality)

Assume $X \geq 0$ *and* $E(X) = \mu$ *finite, then* $\forall a \geq 0$

$$P(X \geq a) \quad \leq \quad \frac{\mu}{a}.$$

Lemma (Chebyshev's inequality)

Assume $E(X) = \mu$ *and* $Var(X) = \sigma^2$ *finite, then* $\forall k \geq 0$

$$P(|X - \mu| < k\sigma) \quad \geq \quad 1 - \frac{1}{k^2}.$$

The laws

Lemma (Weak law of large numbers)

Assume $E(X) = \mu$ *and* $Var(X) = \sigma^2$ *finite, and* $\bar{X}$ *defined as above, then* $\forall \epsilon > 0$

$$P(|\bar{X} - \mu| < \epsilon) \quad \geq \quad 1 - \frac{\sigma^2}{n\epsilon^2} \overset{n}{\underset{\infty}{\rightarrow}} 1.$$

Lemma (Strong law of large numbers)

Assume $E(X) = \mu$ *and* $Var(X) = \sigma^2$ *finite, and* $\bar{X}$ *defined as above, then* $\forall \epsilon > 0$

$$P(\bar{X} \overset{n}{\underset{\infty}{\rightarrow}} \mu) \quad = \quad 1.$$

Monte-Carlo integration

Rationale

To understand a statistical model, simulate many realizations of it.

Concrete view

A way of numerically solve a difficult integration problem.

Assumptions

X *continuous random variable*
$f_X(x)$ *probability density function (PDF)*
some function $g(\cdot)$ given

Problem

Evaluate $E(g(X))$.

We know that

$$E(g(X)) \quad = \quad \int_X g(x) f_X(x) dx$$

... in general: *analytically intractable*

Practical solution

1. simulate $x_1, \ldots, x_n$ *realizations* of X
2. produce $g(x_1), \ldots, g(x_n)$ *realizations* of $g(X)$
3. assume the variance of $g(X)$ is finite (to apply the law of large numbers)

Hence

$$E(g(X)) \simeq \frac{1}{n}\sum_{i=1}^{n} g(x_i)$$

What if we cannot simulate realizations of X?

... but we can simulate realizations $y_1, \ldots, y_n$ of Y (r.q. analogous to X) with PDF $h(\cdot)$

$$
\begin{aligned}
E(g(X)) &= \int_X g(x) f_X(x) dx \\
&= \int_X \frac{g(x) f_X(x)}{h(x)} h(x) dx
\end{aligned}
$$

and hence

$$
E(g(X)) \simeq \frac{1}{n} \sum_{i=1}^{n} \frac{g(y_i) f_X(y_I)}{h(y_i)}
$$

Are we done?

Two possibilities:

1. the two PDF (f and h) *almost* agree.
2. come up with a method to simulate a *PDF* with another one

The first strategy (when applicable) is called

Importance Sampling

The second approach is more general (and more challenging)

The uniform case

A non trivial assumption

We assume to have a *(pseudo) random number generator*

This is equivalent to say that we can simulate the uniform distribution.

Notation

$$U \quad \sim \quad U(0,1)$$

Transformation methods

We can now hope to *move* from the uniform distribution to a generic PDF $f(\cdot)$. This can be done IF the corresponding F is invertible.

Theorem (Inverse distribution method)

Let $F(\cdot)$ be an invertible cumulative distribution function (CDF), then

$$X = F^{-1}(U)$$

has CDF $F(\cdot)$

Proof.

$$
\begin{aligned}
P(X \le x) &= P(F^{-1}(U) \le x) \\
&= P(U \le F(x)) \\
&= F_U(F(x)) \\
&= F(x)
\end{aligned}
$$

$\square$

The discrete-event continuous-time case

We must simulate which event and how much time has passed. Let us start with time, using the inverse distribution (inversion) method.

Proposition

If $\lambda > 0$, *then*

$$X = \frac{1}{\lambda} \log(U),$$

has an $Exp(\lambda)$ *distribution.*

Proof

Observe that $X \sim Exp(\lambda)$ has density $f(x) = \lambda e^{-\lambda x}$ and distribution $F(x) = 1 - e^{-\lambda x}$. Then apply the inversion method.

The discrete-event continuous-time case

To simulate $X \sim Exp(\lambda)$ (using U)

$$x = -\frac{1}{\lambda}\log(u).$$