

## 1.1. Introduction

- The starting point for stochastic simulation is the construction of a **Random Number generation**
- Usually, this mechanism generates an integer uniformly on  $[0, n]$  for a large  $n$
- The most common way to do this is using the conventional generation of **leftovers** (§4.5.3)
- The quantity generated/trained samples will be called  $u$
- Note that most methods currently available for generating random numbers are based on deterministic procedures, and therefore are actually **pseudo-random** samples

## 1.2. Generation of discrete random quantities

- Assume we can generate a sample  $u \sim U[0, 1]$
- Let  $x_1, \dots, x_k$ , and for each  $x_i$  there is a associated probability  $p_1, \dots, p_k$  s.t.  $\sum p_i = 1$
- Divide  $[0, 1]$  in  $k$  intervals  $I_1, \dots, I_k$ , where  $I_i = (x_{i-1}, x_i]$  and  $I_0 = [0, x_1]$ ,  $F_i = p_1 + \dots + p_i$ ,  $i=1, \dots, k$   
 $\Rightarrow P(x=x_i) = P(u \in I_i) = F_i - F_{i-1} = p_i$

## Bernoulli distribution

- If  $X \sim \text{Ber}(p) \Rightarrow P(X=x) = \prod_{i=1}^x p^{x_i} (1-p)^{1-x_i}$
- If  $x_1=1, x_2=0$  and  $k=2$ , then:  
 $F_1 = P(X=x_1) = P(u \in I_1) = p \Rightarrow$  if  $0 < u \leq p \Rightarrow x=1$   
otherwise  $\Rightarrow x=0$

## Binomial distribution

- If  $X \sim \text{Bin}(n, p) \Rightarrow P(X=x) = \binom{n}{x} p^x (1-p)^{n-x}$
- Note that, if  $X_1, \dots, X_m$  are R.V. s.t.  $X_i \sim \text{Ber}(p) \Rightarrow X = \sum X_i \sim \text{Bin}(mp)$
- Therefore, for each  $u_1, \dots, u_m$  drawn from  $U[0, 1]$ , we say that:  
if  $0 < u_i \leq p \Rightarrow x_i=1 \quad \Rightarrow \quad x = \sum x_i \sim \text{Bin}(mp)$   
otherwise  $\Rightarrow x_i=0$

## Poisson distribution

- $X \sim \text{Po}( \lambda ) \rightarrow P(X=x) = e^{-\lambda} \frac{\lambda^x}{x!} \prod_{i=1}^x$
- The most efficient way to sample from a poisson dist. is to sample from an exponential dist.

## 1.3. Generation of continuous random quantities

- The basic result for generating continuous random quantities is the **probability integral transform**  
if  $x$  is a random sample from a continuous C.D.F. given by  $F$ , then  
 $u = F(x) \sim U[0, 1] \Leftrightarrow F^{-1}(u) = x \sim F$

## Exponential distribution

- $X \sim \text{exp}(\lambda) \Rightarrow f(x) = \lambda e^{-\lambda x} \prod_{(0, \infty)} \quad \text{and} \quad F(x) = 1 - e^{-\lambda x}$
- $\Rightarrow u = 1 - e^{-\lambda x} \Leftrightarrow -\ln(1-u) = -\lambda x \Leftrightarrow x = -\frac{\ln(1-u)}{\lambda}$
- Therefore, the one procedure to generate random samples could be to draw  $u_i \sim U[0, 1]$  and then apply the transformation  $x_i = -\frac{\ln(1-u_i)}{\lambda}$

## Gamma distribution

- We know that, if  $x_1, \dots, x_n$  are samples from  $\text{exp}(\lambda)$ , then  $x = \sum x_i \sim \text{Gamma}(n, \lambda)$

• Therefore, to sample from the gamma distribution we first sample  $u_1, \dots, u_m \sim U[0, 1]$ , apply the inverse transformation, and finally sum over all inverse transformations

## 1.3.2. Bivariate Techniques

- If  $(x_1, x_2)$  has joint density  $f_{x_1, x_2}(x_1, x_2)$  and  $g(x_1, x_2) = (y_1, y_2)$  is one-to-one differentiable transformation with inverse  $g^{-1}(y_1, y_2) = (x_1, x_2)$ , then

$$f_g(y_1, y_2) = f_{x_1, x_2}(g^{-1}(y_1, y_2)) J, \quad J = \begin{vmatrix} \frac{\partial g_1}{\partial y_1} & \frac{\partial g_1}{\partial y_2} \\ \frac{\partial g_2}{\partial y_1} & \frac{\partial g_2}{\partial y_2} \end{vmatrix}$$

## Normal Distribution

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

• Let  $\phi = \frac{1}{\sigma^2}$  for convenience, then:

$$f(x) = \frac{\phi^{\frac{1}{2}}}{\sqrt{2\pi}} e^{-\frac{\phi(x-\mu)^2}{2}} \prod_{(-\infty, \infty)}$$

- All available methods to sample from the normal distribution requires the sample of two uniforms

• Box and Muller (1958) showed that, if  $u_1, u_2 \stackrel{i.i.d.}{\sim} U[0, 1]$ , then  $x_1 = \sqrt{-2 \ln(u_1)} \cos(2\pi u_2)$  and  $x_2 = \sqrt{-2 \ln(u_1)} \sin(2\pi u_2) \stackrel{i.i.d.}{\sim} N(0, 1)$

• Therefore, one way to sample  $x \sim N(\mu, \sigma^2)$  is to sample  $u_1, u_2 \stackrel{i.i.d.}{\sim} U[0, 1]$ , then to apply the transformation  $x = \phi^{\frac{1}{2}} \sqrt{-2 \ln(u_1)} \cos(2\pi u_2) + \mu$

• Another more direct way would be to apply the CLT. Sample  $u_1, \dots, u_m$  from  $U[0, 1]$ , then, for  $m \rightarrow \infty$ , we have:

$$x = \phi^{\frac{1}{2}} \left( \frac{u_1 - \frac{1}{2}}{\sqrt{m}} \right) T_m + \mu \sim N(\mu, \sigma^2)$$

## 1.4. Generation of random vectors and matrices

- If  $x = (x_1, \dots, x_d)^T$  has joint distribution given by  $f(x_1, \dots, x_d)$  and  $g(x_1, \dots, x_d) = (y_1, \dots, y_d)^T$  is differentiable one-to-one transformation with inverse  $g^{-1}(y_1, \dots, y_d) = (x_1, \dots, x_d)^T$ , then

$$f_g(y_1, \dots, y_d) = f_x(g^{-1}(y_1, \dots, y_d)) J, \quad J = \frac{\partial x_1}{\partial y_1} \dots \frac{\partial x_d}{\partial y_d}$$

## Multivariate Normal

$$\underline{x} = (x_1, \dots, x_d)^T \sim N_d(\underline{\mu}, \Sigma)$$

$$\Rightarrow f(\underline{x}) = (2\pi)^{-\frac{d}{2}} |\Sigma|^{\frac{1}{2}} \exp\left(-\frac{1}{2} (\underline{x} - \underline{\mu})^T \Sigma^{-1} (\underline{x} - \underline{\mu})\right)$$

where  $\Sigma$  must have full rank  $d$  to exist.

- Recall that, if  $A$  is a  $d \times d$  matrix of constants, and  $b$  is a  $n$ -dimensional vector of constants, then

$$\text{if } \underline{x} \sim N_d(\underline{\mu}, \Sigma) \Rightarrow \underline{y} = A\underline{x} + b \sim N_d(A\underline{\mu} + b, A\Sigma A^T)$$

- A procedure to generate samples from a  $d$ -variate normal would be to first sample  $u_1, \dots, u_m \sim U[0, 1]$ , for  $i=1, \dots, d$ , then apply the CLT to get  $\underline{x} = (x_1, \dots, x_d)^T \sim N(0, 1)$ . Finally, we can apply the transformation

such that:  $\underline{y} = \Sigma^{\frac{1}{2}} \underline{x} + \underline{\mu} \sim N_d(\underline{\mu}, \Sigma)$

- In this situation we just need to specify an envelope constant  $A' = KA$  to ensure a complete envelope

• The method consists of sampling  $\underline{x}$  from  $u \sim U[0, 1]^d$ , and accepting  $\underline{x}$  as a sample of  $\Pi$  if  $A \underline{u} g(\underline{x}) \leq \Pi(\underline{x})$ , otherwise  $\underline{x}$  is rejected

• Samples here can have any size, which is rarely a problem to have small samples for the rejection method

• Samples can be drawn with or without replacement

## 1.5. Resampling Methods

- Resampling Methods can basically be described as generation techniques that require sampling of random variables

- Typically, they consist of two steps with the first one providing a value sampled from an approximating distribution, and the second would be some sort of correction mechanism

- Let the density of interest be  $\Pi$  and the auxiliary density be  $g$

## 1.5.1. Rejection Method

- This method uses an auxiliary distribution  $g$  for random generation of distributions that are not amenable to analytical treatment.

- The idea is to use of to draw samples from  $\Pi$

• The only restriction on  $g$  is that there must exists a constant  $A < \infty$  s.t.  $\Pi(x) \leq A g(x), \forall x$ .

↳ envelope density  
↳ envelope constant

- To generate samples using this method, we don't even have to know the complete expression for  $\Pi$

- This will be particularly useful for Bayesian posterior, where the constant  $K = \int \Pi(x) dx$ ,  $\Pi'(x) = K \Pi(x)$ , is usually not completely known

• In this situation we just need to specify an envelope constant  $A' = KA$  to ensure a complete envelope

- The method consists of sampling  $\underline{x}$  from  $u \sim U[0, 1]^d$ , and accepting  $\underline{x}$  as a sample of  $\Pi$  if  $A \underline{u} g(\underline{x}) \leq \Pi(\underline{x})$ , otherwise  $\underline{x}$  is rejected

• Samples here can have any size, which is rarely a problem to have small samples for the rejection method

• Samples can be drawn with or without replacement

## 1.5.2. Weighted Resampling Methods - Rubin (1981) - SIR method

- Determination of the constant  $A$  is central to the rejection method, and is not a trivial task

- Weighted Resampling techniques use the same idea of drawing samples from an auxiliary density  $g$  but without relying on the constant  $A$

- One disadvantage of the method is that it only provides an approximation to  $\Pi$

- Assuming a sample  $x_1, \dots, x_m$  from  $g$ , the weights of the method are given by:

$$w_i = \frac{\Pi(x_i)/g(x_i)}{\sum_{j=1}^m \Pi(x_j)/g(x_j)}, \quad i=1, \dots, m$$

- A second sample of size  $m$  is drawn from the discrete distribution on  $\{x_1, \dots, x_m\}$  with probabilities  $w_1, \dots, w_m$

• Note that here again we don't need to know the normalizing constant, that is, if we need to sample from  $\Pi'(x) = K \Pi(x)$ , and we don't know  $K$ , the weights  $w_i$  will still be the same

- For both the methods presented previously  $g$  must reflect as much as possible the characteristics of  $\Pi$

• Samples here can have any size, which is rarely a problem to have small samples for the rejection method

• Samples can be drawn with or without replacement

## 1.5.3. Weighted Resampling Methods - Rubin (1981) - SIR method

- Determination of the constant  $A$  is central to the rejection method, and is not a trivial task

- Weighted Resampling techniques use the same idea of drawing samples from an auxiliary density  $g$  but without relying on the constant  $A$

- One disadvantage of the method is that it only provides an approximation to  $\Pi$

- Assuming a sample  $x_1, \dots, x_m$  from  $g$ , the weights of the method are given by:

$$w_i = \frac{\Pi(x_i)/g(x_i)}{\sum_{j=1}^m \Pi(x_j)/g(x_j)}, \quad i=1, \dots, m$$

- A second sample of size  $m$  is drawn from the discrete distribution on  $\{x_1, \dots, x_m\}$  with probabilities  $w_1, \dots, w_m$

• Note that here again we don't need to know the normalizing constant, that is, if we need to sample from  $\Pi'(x) = K \Pi(x)$ , and we don't know  $K$ , the weights  $w_i$  will still be the same

- For both the methods presented previously  $g$  must reflect as much as possible the characteristics of  $\Pi$

• Samples here can have any size, which is rarely a problem to have small samples for the rejection method

• Samples can be drawn with or without replacement