MTH210

Psuedo-Random number generation :

Multiplicative congruential method :

Set seed x_0, and positive integers a, m. Obtain $x_t = ax_{t-1} \; mod \; m$, Return sequence $\frac{x_t}{m}$ for $t = 1, \ldots, n$. a and m are chosen to be large, so as to ensure large jumps in the case of a and m to avoid repetition.

Mixed Congruential Generator :

set $x_t = (ax_{t-1} + c) \; mod \; m$, return sequence x_t/m for $t = 1, \ldots, n$.

> 💡 To generate U(a, b) : $(b - a)U + a \equiv U(a, b)$

Generating Discrete RV

Inverse Transform Method :

Generating uniform Discrete RV :

Suppose we want to generate the value of X which is likely to take any value between {1, 2, . . . , n} i.e. $P(X = j) = 1/n$, for all j. thus $X = j$ if $\frac{j-1}{n} \leq U < \frac{j}{n}$, or in other words, $X = [nU] + 1$

Calculating averages $\bar{a} = \sum_{i=1}^{n} a(i)/i$

Note : n is very large. . We want to approximate $\bar{a}$, and the values $a(i)$ are not easily calculated. We can generate k discrete uniform random variables $X_i, i = 1, ..., k$- by using the above approach, by SLLN(strong law) $\bar{a} \approx \sum_{i=1}^{k} \frac{a(X_i)}{k}$.
Proof can be given as follows, ifi X is a discrete uniform RV over the integers 1 to n, we can state the mean of the RV $a(X)$ is $E[a(X)] = \sum_{i=1}^{n} a(i)P(X = i) = \sum_{i=1}^{n} \frac{a(i)}{n} = \bar{a}$.

Poisson(λ) :

Method 1 is to Generate a random number U. $p = p_0$, i = 0, while(U ≥ p){i++; $p+ = p_i$} return i. But this process is very costly, as the average number of searches is 1 + λ. So we start at i = [λ], if U ≤ f(i), we then decrease i else we increase i. In this case the average number of searches ≈ 1 + $E[|X - \lambda|]$, where X is basically the random variable we are estimating which is nothing but the value of i, as for large lambda the poisson is approximately Normal with mean and variance λ

Generating Binomial Random Variables Binomial (n, p) :

we use the recursive approach to progressively increase p_i by using the equation $P(X = i + 1) = \frac{n-i}{i+1} \frac{p}{1-p} P(X = i)$.

Accept-Reject Technique

Suppose we have an efficient method for simulating a random variable having pmf q_j , and we want to find out the distribution with pmf p_j. First simulate a RV Y with mass function { q_j }, and then accepting this simulated value with a prob proportional to p_Y/q_Y.

💡 $\frac{p_j}{q_j} \leq c$ for all j such that p_j > 0

This simulates a RV X with pmf $p_j = P(X=j)$

💡 Step 1 : Simulate the value of Y, having probability mass function q_j

Step 2: Generate a random number U. If U < $\frac{p_y}{cq_y}$, set X = Y & stop, else return to step 1.

Theorem : The acceptance-rejection algorithm generates a RV X such that $P(X=j) = p_j$, $j = 0, ..$ .The number of iterations of the algoritm needed to obtain X is a geometric RV with mean c.

Proof To begin, let us determine the probability that a single iteration produces the accepted value j. First note that

$$\begin{aligned} P\{Y=j, \text{it is accepted}\} &= P\{Y=j\}P\{\text{accept}|Y=j\} \\ &= q_j \frac{p_j}{cq_j} \\ &= \frac{p_j}{c} \end{aligned}$$

Summing over j yields the probability that a generated random variable is accepted:

$$P\{\text{accepted}\} = \sum_j \frac{p_j}{c} = \frac{1}{c}$$

As each iteration independently results in an accepted value with probability $1/c$, we see that the number of iterations needed is geometric with mean c. Also,

$$\begin{aligned} P\{X=j\} &= \sum_n P\{j \text{ accepted on iteration } n\} \\ &= \sum_n (1-1/c)^{n-1} \frac{p_j}{c} \\ &= p_j \end{aligned}$$

□

💡 Note

- Note: Since the probability of acceptance in any loop is 1/c, the expected number of loops for one acceptance is c. The larger c is, the more expensive the algorithm.
- Within the support $\{a_j\}$ of $\{\,p_j\}$, the proposal distribution must always be positive, i.e. for all a_j in support of $\{\,p_j\}$, $P(Y=a_j) = q_j > 0$.
- $c = \max_{x=0,1,...} \frac{p_x}{q_x}$
- We want pmf of the proposal and target to match each other as much as possible, so that c is close to 1.

Miscellaneous

chapter_three.pdf (nrbook.com) - Luc Devroye (Reading Material) for accept-reject method

Notes :

- lets say we want to simulate a random variable that has pmf q_i = $\frac{1}{i(i+1)}$ $(i >= 1)$. Observe that $q_i = \frac{1}{i} - \frac{1}{i+1}$, Let the draw be U, if $U > 1 - \frac{1}{i+1}$, then X = i ⇒ $P(X=i)$ = $P(U < \frac{1}{i+1})$ = $\frac{1}{i+1}$. so we can just set $X \Leftarrow [1/U]$.

Generating Continuous RV

Inverse-Transform Function

Proposition : ***Let U be a Uniform(0,1) RV. For any continuous distribution function F the random variable X defined by $X = F^{-1}(U)$ has distribution F.***

$$\begin{aligned} F_X(x) &= \Pr(X \leq x) \\ &= \Pr(F^{-1}(U) \leq x) \\ &= \Pr(F(F^{-1}(U)) \leq F(x)) \\ &= \Pr(U \leq F(x)) \\ &= F(x)\,. \end{aligned}$$

💡 ***U and (1 - U) both have same Uniform(0, 1) distribution.***

> Recall that a poisson distribution with parameter λ results when the time interval between the successive intervalls are independent exponential(λ) distribution. Let N(1) denote the number of events by time 1 and X_i denote the time interval between i^{th} and $(i+1)^{th}$ event, then $N = max(n : \sum_{i=1}^{n} X_i \leq 1)$, thus N(1) can be simulated by -
>
> $$\begin{aligned} N &= \text{Max}\left\{n: \sum_{i=1}^{n} -\frac{1}{\lambda} \log U_i \leq 1\right\} \\ &= \text{Max}\left\{n: \sum_{i=1}^{n} \log U_i \geq -\lambda\right\} \\ &= \text{Max}\{n: \log(U_1 \cdots U_n) \geq -\lambda\} \\ &= \text{Max}\{n: U_1 \cdots U_n \geq e^{-\lambda}\} \end{aligned}$$

Accept - Reject Method :

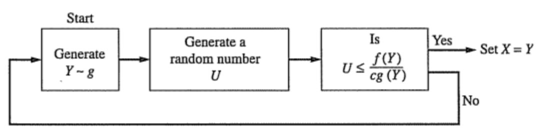

Figure 5.1. The rejection method for simulating a random variable X having density function f.

> In order to choose a good proposal distribution (that yields a finite c), it is important to choose a g so that it has "fatter tails" than f. This ensures that as x → ∞ or x → ∞, g dominates f, so that c → 0 in the extremes, rather than blow up.

Finding proposals for Beta(α, β):

$f_{Beta(\alpha,\beta)}(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$, x ∈ $(0, 1)$, depends on value of α and β.

- if α < 1 && β < 1 : any proposal distribution with a bounded density will not work as beta becomes unbounded.
- Exactly one parameter ≥ 1 (α ≥ 1) $f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} <= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}(1-x)^{\beta-1}$, we can use a function $g(x) = mx^{m-1}$, which is a proper density function on $(0, 1)$ and is easier to simulate.
- else we can just use a uniform(0, 1) to simulate.

$gamma(\alpha, \beta)$:

$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$, $x \in (0, \infty)$ which has mean $\frac{\alpha}{\beta}$, For the proposal you can choose Exponential(λ), which also should have mean the same as the former distribution, so $\frac{1}{\lambda} = \frac{\alpha}{\beta}$. Note that $\alpha > 1$.

$Normal(0, 1)$:

$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, To estimate Normal we can use T-distribution with a fat tail, the fattest T-distribution is cauchy s.t. $g(x) = \frac{1}{\pi}\frac{1}{1+x^2}$, f(x)/g(x) tends to zero as x tends to ∞, thus it converges and maximum value is observed at x = $+1, -1$.

Box-Muller transformation for $N(0, 1)$:

To generate samples for Normal, we will draw random variables (R^2, ϕ). Let X and Y be iid from $N(0, 1)$. $N(0, 1) Rcos\phi$ and Y = $Rsin\phi$, upon solving ϕ is from $U[0, 2\pi]$ and R^2 from $Exp(2)$

💡 ***Algorithm*** :
1. Generate
U_1 and U_2 from U(0, 1) independently
2. Set
$R^2 = -2logU_1$ and $\phi = 2\pi U_2$
3. Set
$X = Rcos\phi$ and $Y = Rsin\phi$

Ratio-of-Uniforms :

Let $f(x)$ be a target density with support $\mathscr{X}$ and distribution F. Define the set $D = \{(u, v) : 0 \leq u \leq \sqrt{f(\frac{v}{u})}\}$. If D is bounded, let (U, V) be unirofmrly distributed over the set D; then V/U ~ $F \Rightarrow$ if we can draw (U, V) ~ Unif(D) then V/U~ F.

Now the question is how to draw uniformly from D : As D is bounded, we can enclose it in a rectangle and apply accept-reject. So consider $(u, v) \in [0, a] \times [b, c]$, clearly $a = sup_{x \in \mathscr{X}} f(x)$. Now $\frac{v}{x} = u \leq f^{\frac{1}{2}}(x)$, thus

$$if\ x < 0 : v >= inf_{x \in \mathscr{X}}(xf^{\frac{1}{2}}(x))$$

$$if\ x > 0 : v <= sup_{x \in \mathscr{X}}(xf^{\frac{1}{2}}(x))$$

Note that if $\sqrt{f(x)}$ or $x^2 f(x)$ is unbounded then D is unbounded and the algorithm cannot work.

💡 ***Algorithm :***
1. Generate (U, V) from
$U[0, a] \times U[b, c]$
2. If U ≤
$f^{\frac{1}{2}}(\frac{V}{U})$, then set $X = V/U$
3. else return to step 1.

Composition method :

Suppose we have an efficient way of simulating $p_j^{(1)}$ and $p_j^{(2)}$, and we want to simulate $Pr(X = j) = \alpha p_j^1 + (1 - \alpha)p_j^2$, j≥0 where $0 < \alpha < 1$. Now we can proof that simulating X is the same as choosing X_1 wp α and X_2 wp $1 - \alpha$.

💡 ***Algorithm*** :
1. Draw U ~ U(0, 1)
2. if U ≤
α, then simulate X_1 ~ $P^{(1)}$ else X_2 and stop.

Similar can be the case for k distributions i.e $F(x) = \sum_{i=1}^{k} \alpha_i F_i(X)$, and upon disfferentiating we get the density mixture. To simulate from composition F choose X_i wp α_i.

Zero Inflated Poisson - $ZIP(\delta, \lambda)$

if X ~ ZIP(δ, λ) : Pr(X = k) = $\delta + (1 - \delta)e^{-\lambda}$ for k = 0 and $(1 - \delta)e^{-\lambda}\frac{\lambda^k}{k!}$ for k = 1, 2, ...
We will use composition method to sample from zip :
$p_j^{(1)}$: Pr(X = 0) = 1 and Pr(X $\neq$ 0) = 0 and $p_j^{(2)}$ be Poisson(λ) : theN $Pr(X = k) = \delta p_k^{(1)} + (1 - \delta)p_k^{(2)}$

💡 ZIP :
1. Draw U ~ U(0, 1)
2. if
$U \leq \delta$ then $X = 0$ else simulate X ~ $Poisson(\lambda)$

- Similarly a mixture of normals can also be simulated and also zero-inflated gamma distribution

Relationships b/w distributions :

1. ***Binomial Distribution :*** sum of iid Bern(p) = bin(n, p)
2. ***Negative Binomial Distribution :*** sum of iid Geom(p) = NB(r, p) .
3. If X ~ Gamma($\alpha, 1$) and Y ~ Gamma($\beta, 1$), then $\frac{X}{X+Y}$ ~ $Beta(\alpha, \beta)$.
4. ***Dirichlet distribution :*** $f(x_1, x_2, \ldots, x_k) = \frac{\Gamma(\alpha_1 + \ldots + \alpha_k)}{\prod_{i=1}^{k}\Gamma(\alpha_i)} \prod_{i=1}^{k} x_i^{\alpha_i - 1},\ 0 \leq x_i \leq 1$, $\sum_{i=1}^{k} x_i = 1$. This is a dirichlet function of a general beta distribution
   - if Y_i ~ $Gamma(\alpha_i, 1)$, then $X_i = \frac{Y_i}{\sum_{i=1}^{k} Y_i}$ ⇒ $(X_1, \ldots X_k)$~ $Dir(\alpha_1, \ldots, \alpha_k)$
5. Chi- Squared : sum of K iid Normal(0, 1) $\Rightarrow \mathscr{X}_k^2$

6. ***T-Distribution*** : Let Z ~ N(0 ,1) and Y ~ $\mathscr{X}_k^2$, then X = $\frac{Z}{\sqrt{\frac{Y}{k}}}$ ~ t_k
7. $Y = \mu + \sigma Z$ ⇒ $F_Y(y) = F_Z(\frac{y-\mu}{\sigma})$ ⇒ upon differentiating ⇒ $f_Y(y) = \sigma^{-1} f_Z(\frac{y-\mu}{\sigma})$

Multi-dimensional :

Consider a RV $\mathbb{X} = (X_1, X_2, \dots, X_k)$ with a joint pdf f(x), to simulate f(x) we use conditional properties i.e. $f(x) = f_{X_1}(x_1) f_{X_2|X1}(x_2) \dots f_{X_k|X_1, \dots X_{k-1}}(x_k)$

Multivariate Normal :

Consider sampling from a $N_K(\mu, \sum)$, wher $\sum$ is positive definite(⇒ eigenvalue decomposition possble). To simulate this, look at $\mathbb{Z} = \sum^{-1/2}(\mathbb{X} - \mu)$, this $\mathbb{Z}$ ~ $N_k(0, I_k)$, as in case of normal covariance = 0 ⇒ independence thus, we can simulate $Z_1, Z_2, \dots Z_k$ ~iid from $N(0,1)$ and set $\mathbb{Z} = (Z_1, \dots, Z_k)$. Then $X := \mu + \sum^{1/2} \mathbb{Z}$ ~ $N_k(\mu, \sum)$

Importance Sampling

Simple Monte carlo

Suppose F is a distribution with density f. We wish to estimate the expectation of a function $h : \mathscr{X} \to \mathbb{R}$ w.r.t F i.e. $\theta := E_F[h(X)] = \int_{\mathscr{X}} h(x) f(x) dx < \infty$. We also assume that $\sigma^2 = Var_F(h(X)) < \infty$. To achieve this we can draw iid samples X_i ~ F(iid), then by WLLN $\hat{\theta} = 1/N \sum_{t=1}^{N} h(X_t) \to \theta$ in probability as $N \to \infty$

Variance of the estimator : $Var(\hat{\theta}) = Var(\frac{1}{N} \sum h(X_t)) = \frac{1}{N^2} \sum Var_F(h(X_t)) = \frac{Var_F(h(X_1))}{N} = \frac{\sigma^2}{N}$ by independence and identical distribution. As CLT holds if $\sigma^2 < \infty$ ⇒ as $N \to \infty$, $\sqrt{N}(\hat{\theta} - \theta) \to N(0, \sigma^2)$

Simple importance sampling

For h : $\mathscr{X}$ → R, we want to estimate $\theta = E_F[h(X)]$. Let G be a distribution with density g defined on $\mathscr{X}$ so that $E_F[h(X)] = E_G[\frac{h(Z)f(Z)}{g(Z)}]$, Z ~ G. The estimator $\hat{\theta}_g = \frac{1}{N} \sum \frac{h(Z_t)f(Z_t)}{g(Z_t)}$ is the **importance sampling estimator**, the method is called **importance sampling** and G is the **importance distribution**. Observe that this estimator is a weighted average of $h(Z_t)$. i.e. a weight $w(Z_t) = f(Z_t)/g(Z_t)$ is attached to each point Z_t.

Properties :

- *Unbiasedness* : The importance samppling estimaor $\hat{\theta}_g$ is unbiased for θ
- *The importance sampling estimator is consistent for* θ i.e. as N → ∞, $\hat{\theta}_g \to^p \theta$. Now we are interested in quanityfing the variability in our estimator. Thus CLT holds iff Var($\hat{\theta}_g) = Var_g(\frac{1}{N} \sum_t \frac{h(Z_t)f(Z_t)}{g(Z_t)}) = \frac{1}{N} Var_g(\frac{h(Z_1)f(Z_1)}{g(Z_1)}) =: \frac{\sigma_g^2}{N} < \infty$. By CLT : as $N \to \infty$, $\sqrt{N}(\hat{\theta}_g - \theta) \to N(0, \sigma_g^2)$ in distribution. Further we also have the estimator $\hat{\sigma}_g^2 = \frac{1}{N-1} \sum_{t=1}^{N} (\frac{h(Z_t)f(Z_t)}{g(Z_t)} - \hat{\theta}_g^2)^2$ as we already have N samples of $s(Z_t)$(the fraction)

> 💡 A sufficient condistion for the finiteness of above : Suppose $\sigma^2 = Var_F(h(X)) < \infty$. If g is chosen s.t $sup_{z \in \mathscr{X}} \frac{f(z)}{g(z)} \le M < \infty$ ⇒ $\sigma_g^2 < \infty$

Optimal Proposals :

The proposal g should be chosen so that sampling from G is relatively easy and the $Var_g(\hat{\theta}_g) = \frac{\sigma_g^2}{N}$ is smaller than the regular monte carlo variance estimator

Theorem : If $E_F[|h(x)|] \neq 0$, the importance density g^* that minimizes variance is $g^*(z) = \frac{|h(z)|f(z)}{E_F[|h(z)|]}$ for proof we show that $\theta^2 + \sigma_{g^*}^2 \le \theta^2 + \sigma_g^2$ for any g defined on $\mathscr{X}$.

with this choice of g^* we can also show that $\sigma_{g^*}^2 = E_F[|h(z)|]^2 Var_{G*}(\frac{h(Z)}{|h(Z)|})$. ⇒ if on support h only takes positive values then the variance of the importance sampling is zero !

Examples -

- For a gamma(α, β), and h(x) = x^k, the optimum importance distribution is $Gamma(\alpha + k, \beta)$. The variance in this case of estimator is zero as h is non negative in $\mathscr{X}$.
- ***Mean of a standard normal :*** i.e. h(x) = x, we get $g^* = \frac{|x|e^{-x^2/2}}{\int |x|e^{-x^2/2}}$, it is challenging to draw samples from the it, so will search for some other proposal which is more efficient than sampling from the target.

Weighted Importance Sampling

suppose target density $f(x) = a\tilde{f}(x)$ and the proposal density is $g(x) = b\tilde{g}(x)$, where a and b are unknown⇒ suppose we are interested in calculating $\theta := \int_{\mathscr{X}} h(x) f(x) dx$, and we want to use g as the importance distribution. Consider $Z_1, ..., Z_N \sim G$. The **weighted importance sampling estimator of** θ is

$$\hat{\theta}_w = \frac{\sum_{t=1}^{N} \frac{h(Z_t)\tilde{f}(Z_t)}{\tilde{g}(Z_t)}}{\sum_{t=1}^{N} \frac{\tilde{f}(Z_t)}{\tilde{g}(Z_t)}}$$

proof !!!

Properties :

- The weighted importance sampling estimator is consistent. So as N → ∞, $\hat{\theta}_w \to \theta$.

$w(Z) = \frac{\tilde{f}(Z)}{\tilde{g}(Z)}$ is the **un-normzlized importance sampling weight**.

Likelihood Based Estimation

Likelihood function

Suppose $X_1, \ldots, X_n$ is a random sample from a given sitribution with density $f(x|\theta)$, for $\theta \in \Theta$. After obtained the real data, from F, we want to estimate θ and assess the quality of this estimator. One useful method is the *maximum likelihood estimation(MLE)*. Let $\mathbb{X} = (X_1, \ldots, X_n)$

$$L(\theta|\tilde{X} = \tilde{x}) = f(\tilde{x}|\theta) = f(x_1, ..., x_n|\theta)$$

💡 Note that $L(\theta|\tilde{x})$ is not a distribution over θ, it is just a function, that quantifies how likely a value of θ is.

Maximum Likelihood Estimation

$$\hat{\theta}_{MLE} = arg\ max_{\theta \in \Theta} L(\theta|\tilde{x})$$

It is the"most likely" value of θ habing observed the data. $\hat{\theta}_{MLE}$ is the maximum likelihood estimator of θ

Definitions:

Concave Function(1D) : a function $h(x)$ is concave if $h''(x) \leq 0$ for all x.

Concave Function : a function $h(\tilde{x})$ is concave if the hessian matrix $\nabla^2 h(\tilde{x})$, is **negative semi definite** for all $\tilde{x}$. That is, if all eigenvalues of the Hessian are non-positive or $\tilde{a}^T(\nabla^2 h(\tilde{x}))\tilde{a} < 0,\ \forall \tilde{a}$

$$\nabla_{\tilde{x}} = \begin{bmatrix} d/dx_1 \\ d/dx_2 \\ . \\ . \\ . \\ d/dx_n \end{bmatrix} \And \nabla^2 = \begin{bmatrix} d^2/dx_1^2 & . & . & . & d^2/(dx_1)(dx_n) \\ . & d^2/dx_2^2 & . & . & d^2/(dx_2)(dx_n) \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & \frac{d^2}{dx_n^2} \end{bmatrix}$$

Regression

Let $Y_1, ..., Y_n$ be observations known as response. Let $x_i = (x_{i1}, ..., x_{ip})^T \in \mathbb{R}^p$ be the *i*th corresponding vector of covariates for the *i*th observation. Let $\beta \in \mathbb{R}^p$ be the *regression coefficient* so that for $\sigma^2 > 0$, $Y_i = x_i^T\beta + \epsilon_i$, where ϵ_i ~ $\mathcal{N}(0, \sigma^2)$. Define $\tilde{X} := (x_1^T\ x_2^T\ , \ldots, x_n^T)^T$. Now,

$$\tilde{Y} = \begin{bmatrix} y_1 \\ . \\ y_i \\ . \\ y_n \end{bmatrix} = \begin{bmatrix} x_{11} & . & . & . & x_{1p} \\ . & . & . & . & . \\ x_{i1} & . & . & . & x_{ip} \\ . & . & . & . & . \\ x_{n1} & . & . & . & x_{np} \end{bmatrix} \begin{bmatrix} \beta_1 \\ . \\ . \\ . \\ \beta_p \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ . \\ . \\ . \\ \epsilon_n \end{bmatrix} = \tilde{X}\tilde{\beta} + \epsilon \sim \mathcal{N}_n(X\beta, \sigma^2\mathbb{I}_n)$$

This model is built to estimate β, which measures the linear effect of X on Y