Chapitre II : Generations d'echantillon suivant les différentes lois de probabilité

1- Introduction :

Il y a 3 méthodes pour générer des variables aleatoires qui obeissent à une distribution donnée.

- Méthode d'inversion (cas continue est discret)
- Méthode de Rejet.
- Méthode de composition

2- Quelques résultats de probabilité

Soit U une V.a uniforme sur $[0,1]$.

i)- si X une v.a de fonction de rep F_X alors $X = F_X^{-1}(U)$

ii)- $1-U = U'$ est une V.a uniforme sur $[0,1]$
$\Rightarrow Z_X = F_X^{-1}(1-U)$ a pour fonction de Rep F_X

Démonstration :

i) Soit Y une V.a tellque $Y = F_X^{-1}(U)$ et on cherche la loi de Y?

on suppose que X est continue, donc F_X est strictement croissante donc F_X^{-1} existe.

$\forall y \in [0,1]$

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(F_X^{-1}(u) \leq y) = \mathbb{P}(u \leq F_X(y))$$
$$= F_U(F_X(y)) = F_X(y) = Y = X = F_X^{-1}(u)$$
$$= F_X(y) \text{ car } u \rightsquigarrow U(0,1)$$

Donc la v.a Y a pour fonction de répartition F_X (même fonction que X).

$$\Rightarrow X = Y = F_X^{-1}(u)$$

ii) - $F_T(x) = \mathbb{P}(T \leq t)$ avec $T = F_X^{-1}(1-u)$

$$F_T(x) = \mathbb{P}(T \leq t) = \mathbb{P}(F_X^{-1}(1-u) < t) = \mathbb{P}(1-u \leq F_X(t))$$
$$= F_{1-u}(F_X(t)) = F_X(t) \Rightarrow X = T = F_X^{-1}(1-u)$$

2-2 Application à la génération d'échantillon

Soient $(u_1, \dots, u_n)$ n réalisations $\perp\!\!\!\perp$ issues d'une v.a $U[0,1]$ alors :

$x_1 = F_X^{-1}(u_1)$, $x_2 = F_X^{-1}(u_2)$, ..., $x_n = F_X^{-1}(u_n)$ sont considérés comme n realisation independante de la v.a X de fonction de répartition F_X.

2-2-1 Application aux lois de probabilité :

Méthode d'inversion : (cas continu)

La méthode d'inversion est aussi appelée la méthode de transformation, elle n'est utilisé que si la fonction de densité est connu. Analytiquement est peut être integré facilement.

- Cas de la loi Uniforme :

Soit $X \rightsquigarrow \mathcal{U}[a,b]$ sa densité est donnée

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{si } x \in [a,b] \\ 0 & \text{sinon} \end{cases}$$

$$F_X(x) = \begin{cases} 0 & \text{si } x \leq a \\ \frac{x-a}{b-a} & \text{si } x \in [a,b] \\ 1 & \text{si } x \geq b. \end{cases}$$

Comme $X = F_X^{-1}(U) \iff F_X(u) = u \quad \forall x \in [a,b]$

$$\Rightarrow \frac{x-a}{b-a} = u \Rightarrow x = (b-a)u + a.$$

$$x = (b-a)u + a \qquad u \rightsquigarrow \mathcal{U}[0,1].$$

Conclusion :

Pour simuler une v.a uniforme sur $[a,b]$ il suffit de générer des nombres aléatoires U_i de v.a uniforme sur $[0,1]$ et puis on deduira les realisation x_i

$$x_i = (b-a)U_i + a$$

- Cas de la loi exponentielle :

Soit X une v.a qui suit la loi exponentielle de paramètre λ, Sa fonction de densité est :

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si} \quad x \geq 0 \\ 0 & \text{sinon.} \end{cases}$$

Calcul de la fonction de répartition.

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & \text{si} \quad x \geq 0 \\ 0 & \text{sinon.} \end{cases}$$

Comme $\quad X = F_X^{-1}(U) \iff F_X(x) = U$

$$= 1 - e^{-\lambda x} \quad \forall x \geq 0.$$

$$\Rightarrow e^{-\lambda x} = 1 - U \Rightarrow x = \frac{-1}{\lambda} \log(1 - U)$$

Comme $U \rightsquigarrow U[0,1]$ alors $1 - U \rightsquigarrow U[0,1]$

$$\Rightarrow x = \frac{-1}{\lambda} \log(U) \quad \text{où} \quad U \rightsquigarrow U_{[0,1]}$$

Conclusion.

Pour simuler une v.a $\exp(\lambda)$, il suffit de générer des nombres aléatoires U_i de v.a. uniforme sur $[0,1]$ et puis on déduira la réalisation x_i tq

$$x_i = \frac{-1}{\lambda} \log(U_i)$$

Méthode d'inversion "Cas discret" methode "tophat"

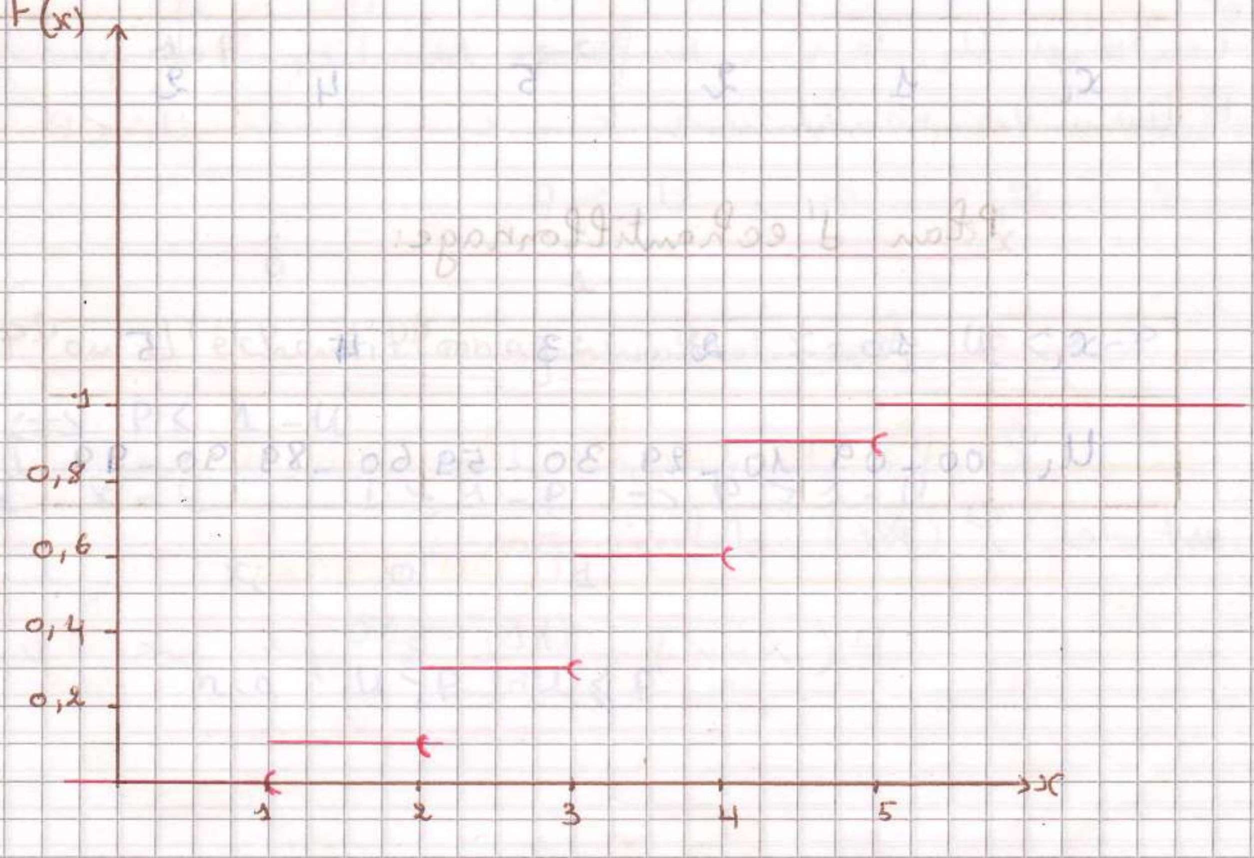
Que l'on ait une distribution de probabilité theorique ou une distribution empirique la méthode de génération est la même supposons que l'on veut généré un echantillon de la loi de probabilité suivante :

x_i	1	2	3	4	5	total
n_i	10	20	30	30	10	$n = 100$
$f_i = \frac{n_i}{n}$	0,1	0,2	0,3	0,3	0,1	1

Calculer la fonction de Répartition de Cette distribution

x	$x < 1$	$1 \leq x < 2$	$2 \leq x < 3$	$3 \leq x < 4$	$4 \leq x < 5$	$x \geq 5$
$F(x)$	0	0,1	0,3	0,6	0,9	1

on trace le graphe de la fonction de Répartition



Plan d'echantillonnage.

a-

x	1	2	3	4	5
n. a	$0 \leq U_i \leq 0{,}1$	$0{,}1 \leq U_i < 0{,}3$	$0{,}3 \leq U_i < 0{,}6$	$0{,}6 \leq U_i < 0{,}9$	$0{,}9 \leq U_i < 1$

On utilisant les n. a données :

Soient U_i $i = \overline{1,5}$ $U_1 = 0{,}02$; $U_2 = 0{,}05$

$U_3 = 0{,}98$; $U_4 = 0{,}50$; $U_5 = 0{,}41$

l'échantillon de la v. a X est donnée par :

U_i	0,02	0,05	0,98	0,50	0,41
x_i	1	1	5	3	3

Supposons que j'ai la suite de n. a suivante :

$100 \times U_i$	0,9	11	98	68	20	....
x_i	1	2	5	4	2	....

Plan d'échantillonnage :

x_i	1	2	3	4	5
U_i	00 - 09	10 - 29	30 - 59	60 - 89	90 - 99

4- Génération d'echantillon de V.a discrets

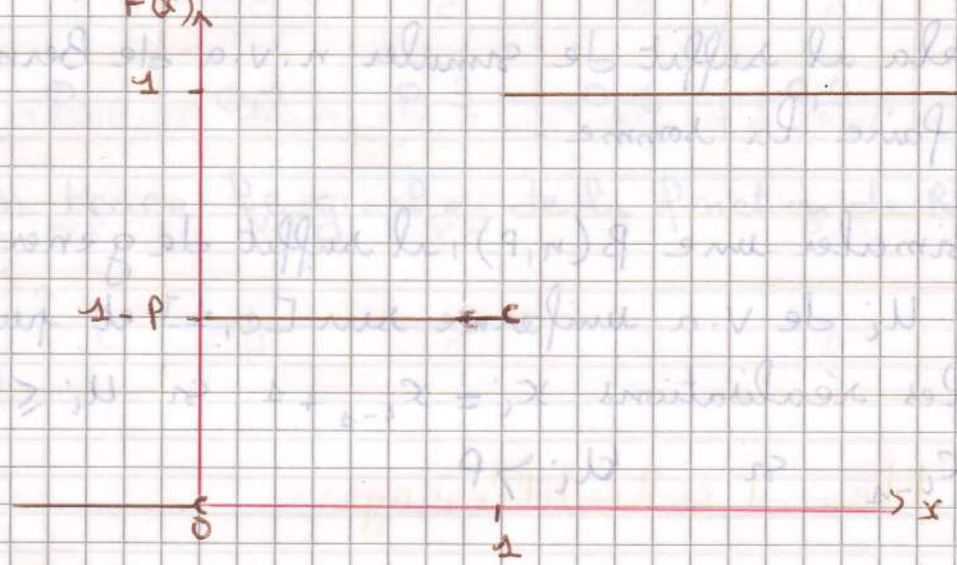
4-1- Loi de Bernoulli de paramètre 'P'

Soit X une V.a ~> Bernoulli de paramètre P.

La loi de probabilité de X est :

$$\begin{cases} \mathbb{P}(X=1) = P \\ \mathbb{P}(X=0) = 1-P \end{cases}$$

x	$]-\infty, 0[$	$[0, 1[$	$[1, +\infty[$
$F_X(x)$	0	$1-P$	1.



Plan d'échantillonnage : • si $x=0$, $U_i < 1-P$

$\iff P < 1-U$

• si $x=1$ $\quad U \geq 1-P \Rightarrow P > 1-U$

x_i	0	1
n.a	$U > P$	$U \leq P$

Pour simuler les v.a de la loi de Bernoulli, il suffit de générer des v.a U_i de v.a uniforme sur $[0,1]$ puis on déduit les réalisations de x_i telleque :

$$\begin{cases} x_i = 0 & \text{si } U_i > P \\ x_i = 1 & \text{si } U_i \leqslant P \end{cases}$$

4-2- Loi Binomiale :

On utilise le faite que la somme de n v.a de loi de Bernoulli de paramètre P suit loi Binomiale de paramètre n P

$$(X_i)_{i=\overline{1,n}} \rightsquigarrow \beta(P) \quad \text{alors} \quad X = \sum_{i=1}^{n} X_i \rightsquigarrow \beta(n,P)$$

Pour cela il suffit de simuler n.v.a de Bernoulli et dont faire la somme.

Pour simuler une $\beta(n,P)$, il suffit de générer desnbres aleatoires U_i de v.a uniforme sur $[0,1]$ et puis on déduira les réalisations $x_i = x_{i-1} + 1$ si $U_i \leqslant P$ ou $x_i = x_{i-1}$ si $U_i > P$

4-3- Loi de Poisson :

La loi du nombre d'occurences pendant $[0, t[$ est de $\mathcal{P}(\lambda t)$, definie par :

$$P(X=n) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}, \text{ on pose } \mu = \lambda t$$

on a

$$\prod_{i=1}^{n+1} U_i < e^{-\mu} < \prod_{i=1}^{n} U_i$$

Pour simuler une v.a de $P(\mu)$ il suffit de générer de nombre aleatoire de v.a Uniforme $[0,1]$ et puis on multiplie entre elles, les observations obtenus. on arrête le processus dès que le produit obtenue devient inferieure à $\exp(-\mu)$. et on deduit alors une réalisation n et on recommence pour une autre réalisation.

4-3- La loi de Erlang :

Soit X une v.a qui suit une loi $Erg(k,\lambda)$ ($Gamma(k, \frac{1}{k})$) tq :

$$F_X(x) = \begin{cases} 1 - \sum_{m=0}^{k-1} \frac{(\lambda x)^m}{m!} e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{sinon.} \end{cases}$$

$X_i \rightsquigarrow Exp(\lambda)$

$$X = \sum_{i=1}^{k} x_i \left(-\frac{1}{\lambda}\right) Ln(U_i)$$

$$= -\frac{1}{\lambda} \sum_{i=1}^{k} Ln(U_i)$$

$$x = -\frac{1}{\lambda} Ln\left(\prod_{i=1}^{k} U_i\right)$$

Pour simuler une v.a $Erg(k, \lambda)$, il suffit de générer k nombres aléatoires et calculer :

$$x = \frac{-1}{\lambda} Ln \prod_{i=1}^{k} U_i$$

5- Génération d'échantillons de la loi normal :

Dans le cas de la loi normale, la méthode d'inversion n'est pas utilisable à cause de l'expression de la fonction de répartition.

5.1- Méthode approximation.

Cette méthode est basé sur le théorème central limite.

Théorème central limite :

Soient $Z_1, Z_2, \dots, Z_k$, k v.a $\perp\!\!\!\perp$ ($\perp\!\!\!\perp$ indépendante) et de même loi tq :

$$\begin{cases} E(Z_i) = \theta \\ Var(Z_i) = \sigma^2 \end{cases} \quad i = \overline{1,k}$$

Lorsque $k \to +\infty$, la v.a $y = \frac{\sum_{i=1}^{k} z_i - k\theta}{\sigma\sqrt{k}} \to N(0,1)$

*) Alors y_i sont des réalisations aléatoires générées approximativement suivant une loi $N(0,1)$

$$y_i = \frac{\sum_{i=1}^{k} z_i \; k\theta}{\sigma\sqrt{k}}$$

**) Supposons que les k v.a z_i ⫫ sont uniforme sur $[a,b]$

$z_i \rightsquigarrow U[a,b]$

$$E(z_i) = \theta = \frac{a+b}{2} \quad , \quad Var(z_i) = \sigma_z^2 = \frac{(b-a)^2}{12}$$

$\Rightarrow \sigma_z = \frac{b-a}{\sqrt{12}}$ on remplace dans (*) on obtient

$$y_i = \frac{\left(\sum_{i=1}^{k} z_i - k\left(\frac{a+b}{2}\right)\right)\sqrt{12}}{(b-a)\sqrt{k}}$$ sont des n.a générés approximativement de la loi $N(0,1)$

***) Supposons que les k v.a z_i sont ⫫ et $\rightsquigarrow U[0,1]$ qu'on note $u_1, \dots, u_k$ on a :

$\theta = \frac{1}{2}$, $\sigma^2 = \frac{1}{12}$ et en remplaçant dans (**)

On obtient :

$$y_i = \frac{\sum_{i=1}^{k} z_i - \frac{k}{2}}{\sqrt{\frac{k}{12}}}$$ qui sont des n.a générés approx de la loi $N(0,1)$

Remarque :

À partir de $k=12$, on obtient une bonne approximation

$y_i = \sum_{i=1}^{12} u_i - 6$ sont des n.a générés approx de $N(0,1)$.

Pour obtenir un échantillon de v.a $X \rightsquigarrow N(\mu, \sigma^2)$ on obtient le changement de variable suivant.

$$Y = \frac{X + \mu}{\sigma} \Rightarrow X = \sigma Y + \mu .$$

Donc $X_i = \sigma \left(\sum_{i=1}^{12} U_i - 6 \right) + \mu$ sont des v.a générés approximativement de la loi $N(\mu, \sigma^2)$.

5-2- Methode de Box muller:

Cette méthode est plus rapide que les precedentes et produit des couples de réalisations aléatoires exactes,

Soient X et Y deux variables aléatoire ⫫ qui suivent la même loi $N(0,1)$ et soit $f(x,y)$ la densité conjointe du couple (X,Y) définie par:

$$f(x,y) = \frac{1}{2\pi} e^{-\left(\frac{x^2+y^2}{2}\right)} \quad \text{où } (x,y) \in \mathbb{R}^2$$

On transforme X et Y en coordonnée polaire et on obtient:

$$\begin{cases} x = r\cos\theta = \varphi_1(r,\theta) \\ y = r\sin\theta = \varphi_2(r,\theta) \end{cases} \Rightarrow \begin{cases} r^2 = x^2 + y^2 \\ \theta = \operatorname{artg}\left(y/x\right) \end{cases}$$

Montrer que les v.a θ et $\frac{R^2}{2}$ sont ⫫ ?

La densité conjointe du couple (R, θ)

$$f(r,\theta)\,dr\,d\theta = \frac{1}{2\pi} e^{-\frac{r^2}{2}} r\,dr\,d\theta$$

Comme $\frac{d\left(\frac{r^2}{2}\right)}{dr} = r \Rightarrow r\,dr = d\left(\frac{r^2}{2}\right)$

$$\Rightarrow f(r,\theta)\,dr\,d\theta = \frac{1}{2\pi} e^{-\frac{r^2}{2}} d\left(\frac{r^2}{2}\right) d\theta = f\left(\frac{r^2}{2}, \theta\right) d\left(\frac{r^2}{2}\right) d\theta$$

on pose : $f\left(\frac{r^2}{2}\right) = e^{-\frac{r^2}{2}}$ donc $\frac{R^2}{2} \rightsquigarrow Exp(1)$

et $f(\theta) = \frac{1}{2\pi}$ donc $\theta \rightsquigarrow \mathcal{U}[0, 2\pi[$

on déduit $f\left(\frac{r^2}{2}, \theta\right) = f\left(\frac{r^2}{2}\right) \times f(\theta) \Rightarrow \frac{R^2}{2}$ et θ sont $\perp\!\!\!\perp$.

Supposons que l'on génère 2 lois uniforme U_1 et U_2 et $\perp\!\!\!\perp$ sur $[0,1[$ en posant :

$$U_1 = e^{-\frac{r^2}{2}}, \quad U_2 = \frac{\theta}{2\pi}.$$

$$\begin{cases} r = \sqrt{-2\ln U_1} \\ \theta = 2\pi U_2 \end{cases}$$

on produira un couple de v.a (x, y) $\perp\!\!\!\perp$ de loi $N(0,1)$ tq :

$$\begin{cases} x = r\left(\sqrt{-2\ln(U_1)}\cos(2\pi U_2)\right) + m \\ y = r\left(\sqrt{-2\ln(U_1)}\sin(2\pi U_2)\right) + m. \end{cases}$$

Montrons que U_1 et U_2 sont $\perp\!\!\!\perp$ et suivent la loi unif$(0,1)$

* $\perp\!\!\!\perp$ de U_1 et U_2 : car $\frac{R^2}{2}$ et θ sont $\perp\!\!\!\perp$.
* $U_1 \rightsquigarrow \mathcal{U}[0,1[$?

$t > 0$

$$F_{U_1}(t) = \mathbb{P}(U_1 < t) = \mathbb{P}\left(e^{-\frac{R^2}{2}} < t\right) = \mathbb{P}\left(\frac{-R^2}{2} < \log t\right)$$

$$= \mathbb{P}\left(\frac{R^2}{2} \geq -\log t\right) = 1 - \left(1 - e^{-(-\log t)}\right) = t$$

* $U_2 \rightsquigarrow \mathcal{U}[0,1[$

$$F_{U_2}(u_2) = P(U_2 < u_2) = \mathbb{P}\left(\frac{\theta}{2\pi} < u_2\right) = P(\theta < 2\pi u_2)$$

$$= F_\theta(2\pi u_2) = u_2$$

5-3- Methode de BoxMuller (Méthode Polaire)

Cette méthode n'utilise pas les fonctions trigonometrique elle est donc plus rapide que la précedente.
on considère deux v.a v_1 et v_2 ⫫ de loi uniforme sur $[-1,1]$ et $W \rightsquigarrow \mathcal{U}[0,1]$ tel que :

$$W = v_1^2 + v_2^2 \quad et \quad W \leq 1.$$

on deroule l'algorithme suivant :

1- Tirer deux nombres aléatoire v_1 et v_2 suivant $U_{[-1,1]}$

2- Si $w = (v_1^2 + v_2^2) \geq 1$, Alors retourner ver ①

3- Sinon

$$x = \sigma\left(v_1 \sqrt{-2\ln(w)/w}\right) + m$$
$$y = \sigma\left(v_2 \sqrt{-2\ln(w)/w}\right) + m.$$

6- Methode de Rejet:

Soit X une v.a de fonction de densité $f(x)$, si $f(x)$ est bornée et x appartient aux domaine borné aussi $a \leq x \leq b$ alors la méthode de rejet peut être utilisé, elle se resume en 4 étapes :

1- normaliser le domaine de f à l'aide d'une échelle "C" de telle sorte que :

$$g(x) = C\,[f(x)] \leq 1 \quad si \quad a \leq x \leq b$$

2- Défini x comme fonction lineaire de U.

$$x = a + (b-a)U$$

3- engendrer (génerer) une paire de nbres aleatoire (U_1, U_2)

4- Chaque fois que l'on rencontre une paire de n.a satisfaisants $x_2 \leq C[f(x_1)] = C[f(a+(b-a)U_1)]$ on accepte $x = a+(b-a)U_1$ comme valeur de la V.a X de densité $f(x)$

Exemple :

Soit X une V.a $\rightsquigarrow$ Beta$(\alpha=0, \beta=1)$

$$f(x) = 2x \quad \text{si} \quad 0 \leq x \leq 1$$

on est dans le cas où $a=0$ et $b=1$.

1- on $C = \frac{1}{2}$, on a $g(x) = \frac{1}{2} f(x) = x \leq 1$.

2- $x = a+(b-a)U \Rightarrow g(U) = \frac{1}{2} f(U)$

3- on génère un n.a U_1 et on calcul $g(U_1) = \frac{1}{2} f(U_1)$ on génère un autre n.a U_2 et on le compare à $g(U_1)$

4- Si $U_2 \leq g(U_1)$ on accepte U_1 comme réalisation x de la V.a X possédant comme fonction densité $f(x)$.

si $U_2 > g(U_1)$ on rejette U_1 et on retourne à l'étape3

Soient $0{,}09 - 0{,}65$

$g(U_1) = U_1 = 0{,}09$

$g(U_2) = U_2 = 0{,}65$

Comme $U_2 \geq 0{,}09$ on rejette $U_1 = 0{,}09$

$U_1 = 0{,}69$, $U_2 = 0{,}66$

$g(U_1) = U_1 = 0{,}69$

comme $U_2 \leq 0{,}69$, on accepte $U_1 = 0{,}69$

Donc $x = U_1 = 0{,}69$ est une realisation de X de fonction de densité Béta.

7- Méthode de decomposition.

Cette methode a remplacer $f(x)$ par un melange probabiliste de fct de densité $g_n(x)$ judicieusement chaisis. autrement dit elle explaite une relation du type $\sum_n g_n(x) \times P_n$.

Exemple:

soit la v.a X de densité de probabilité :

$$f(x) = \begin{cases} \frac{\alpha}{2} e^{-x} + \frac{1-\alpha}{2} e^{-\frac{x}{2}} + \frac{\alpha}{6} e^{-\frac{x}{3}} & \text{si } x \geq 0 \\ 0 & \text{sinon.} \end{cases}$$

On a: $f(x) = \frac{\alpha}{2} e^{-x} + (1-\alpha)\left(\frac{1}{2} e^{-x/2}\right) + \frac{\alpha}{2}\left(\frac{1}{3} e^{-x/3}\right)$

Alors on definit la v.a. discret Y tq:

$Y = y$	1	2	3
$P(Y=y)$	$\frac{\alpha}{2}$	$1-\alpha$	$\frac{\alpha}{2}$

Et on definit aussi les variables aleatoires:

$X_1 \rightsquigarrow Exp(1)$; $X_2 \rightsquigarrow Exp(1/2)$

$X_3 \rightsquigarrow Exp(1/3)$