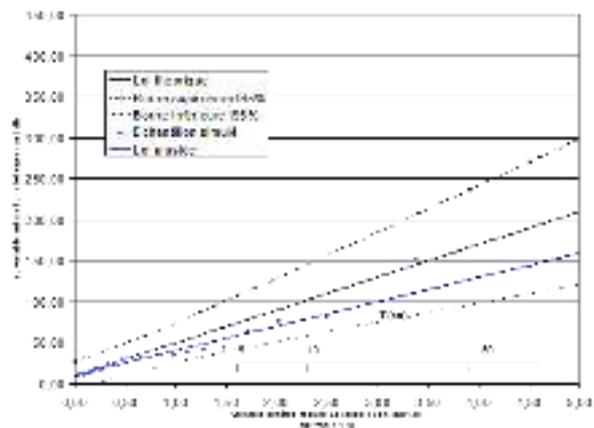
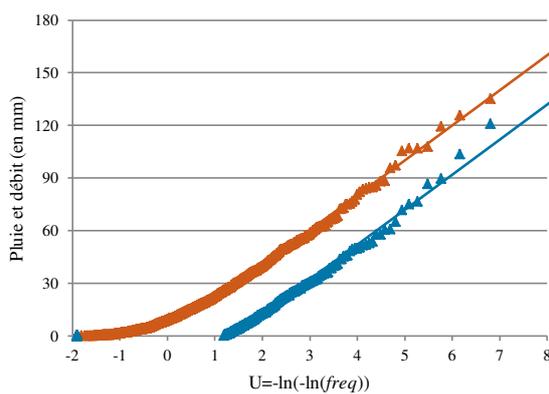


Estimation probabiliste d'évènements extrêmes



Auteur :

GETTVERT Florian

Intervenant :

LANG Michel

EXERCICE 1 : GENERATION D'UN ECHANTILLON DANS UNE LOI EXPONENTIELLE ET MISE A JOUR AUTOMATISEE SOUS EXCEL	3
1. Génération d'un échantillon de N=20 valeurs	3
2. Report graphique	3
3. Vérification du taux de recouvrement de l'intervalle de confiance $I_{95\%}$	4
EXERCICE 2 : ESTIMATION DE LA CRUE CENTENNALE SUR L'ARDECHE A SAINT-MARTIN-D'ARDECHE	7
1. Calcul du débit QJX(10) de période de retour 10 ans	7
2. Estimation du débit instantané QIX(10) de période de retour 10 ans	7
3. Comparaison et limites de l'approche	7
4. Extrapolation du débit centennal de pointe QIX(100)	7
EXERCICE 3 : SIMULATION DU COMPORTEMENT ASYMPTOTIQUE DE LA DISTRIBUTION DES CRUES	9
1. Distribution des pluies et des crues	9
2. Pluies et débits simulés	10
3. Démonstration de la distribution F_Q du débit	11

EXERCICE 1 : Génération d'un échantillon dans une loi exponentielle et mise à jour automatisée sous Excel

Le but est de générer à l'aide d'une macro Excel un échantillon de 20 valeurs issues d'une loi exponentielle : $x_0 = 10$; $a = 40$, et d'avoir immédiatement la fonction de répartition correspondante.

1. Génération d'un échantillon de N=20 valeurs¹

La première étape consiste donc à générer des probabilités p_i pour $i = 1$ à 20 dans une loi uniforme $[0;1]$ et à attribuer à chaque probabilité p_i une valeur $x_p(i)$ telle que $x_p(i) = x_0 + a \cdot [-\ln(1 - p_i)] = x_0 + a \cdot y_p(i)$ avec le changement de variable $y_p(i) = -\ln(1 - p_i)$.

La distribution $(y_p(i) ; x_p(i))$ représente donc la distribution théorique.

On souhaite alors simuler un échantillon de 20 valeurs. Pour cela, on attribue à chaque valeur $x_p(i)$ calculée précédemment une fréquence empirique $freq = \frac{i+0,3}{20+0,4}$. A partir de cet échantillon simulé, on estime alors une nouvelle distribution dite distribution théorique ajustée sur l'échantillon $Expo(\hat{x}_0; \hat{a})$ telle que :

$$\hat{x}_0 = \min(x_p(1); \dots; x_p(20)) \quad \text{valeur seuil}$$

$$\hat{a} = \text{moyenne}(x_p(1); \dots; x_p(20)) - \hat{x}_0$$

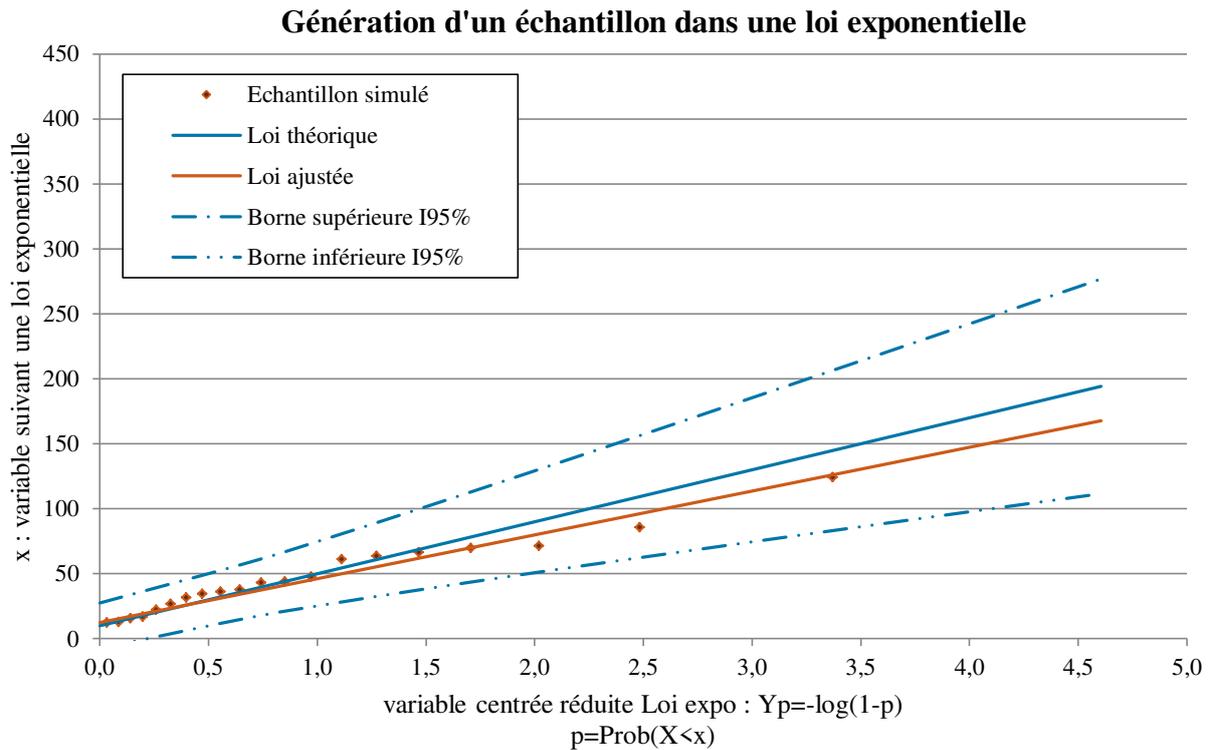
On calcule également les bornes de l'intervalle de confiance $I_{95\%}$ qui est défini par « la probabilité que x_p appartienne à $I_{95\%}$ vaut 0,95 » avec $I_{95\%} = [x_p(0,025) ; x_p(0,975)]$.

↳ Cf. l'onglet *Feuille de calculs* du fichier Excel "ENTPE_3A_Exercice_1_GETTVERT_Florian" pour vérifier le bon fonctionnement de la Macro 1 qui permet de créer la colonne tampon dans le but de figer les valeurs de probabilités et de les trier dans un même temps par ordre croissant. Sur ce même onglet, sont également renseignés tous les détails des calculs.

2. Report graphique

Le report graphique des différentes distributions susmentionnées permet de représenter le graphique fréquentiel ci-dessous avec en ordonnée une valeur x et en abscisse la variable centrée réduite de Gumbel ($y_p = -\ln(1 - p)$) associée telle que $p = \text{Prob}(X < x)$.

¹ Cf. le fichier Excel « ENTPE_3A_Exercice_1_GETTVERT_Florian » pour visionner les feuilles de calculs et les macros.



Sur l'exemple ci-dessus, la représentation linéarisée (variable centrée réduite Y_p) permet de zoomer sur les valeurs les plus fortes pour mieux juger de l'adéquation de la loi ajustée avec les valeurs observées et la loi théorique.

On constate peu d'éloignement des valeurs simulées avec la droite de la loi ajustée. De même, une faible différence entre la loi théorique et la loi ajustée est notée. Cependant, toutes les valeurs de l'échantillon simulé ainsi que la loi ajustée sont comprises dans l'intervalle de confiance $I_{95\%}$. Ainsi, la distribution empirique (loi ajustée) décrit la variable x de manière satisfaisante.

↳ Cf. l'onglet *Graphique* du fichier Excel "ENTPE_3A_Exercice_1_GETTVERT_Florian" pour vérifier le bon fonctionnement de la Macro 2 qui permet de faire apparaître le graphique ci-dessus sur un onglet à part entière tout en générant un nouvel échantillon et l'ajustement de la distribution théorique $Exp(\hat{x}_0; \hat{\alpha})$ correspondant à ce nouvel échantillon.

3. Vérification du taux de recouvrement de l'intervalle de confiance $I_{95\%}$

Pour pouvoir comparer le taux théorique de sortie (5%) de l'intervalle $I_{95\%}$ avec le taux empirique, on simule 100 échantillons sur lesquels on regarde le fait que :

- $\text{Prob}(x_i \notin I_{95\%})$ pour $i = 1, \dots, 20$ c'est-à-dire la probabilité que la valeur x_i de rang i de l'échantillon est située hors des limites de l'intervalle $I_{95\%}$;
- $\text{Prob}(\hat{x}_{0,90} \notin I_{95\%})$: la probabilité que l'estimation de la crue décennale $\hat{x}_{0,90}$ est située hors des limites de l'intervalle $I_{95\%}$;
- $\text{Prob}(\exists i / x_i \notin I_{95\%} \text{ pour } i = 1, \dots, 20)$: la probabilité qu'une des 20 valeurs de l'échantillon est située hors des limites de l'intervalle $I_{95\%}$.

EXERCICE 1 : GENERATION D'UN ECHANTILLON DANS UNE
LOI EXPONENTIELLE ET MISE A JOUR AUTOMATISEE SOUS EXCEL

Voici le tableau récapitulatif comportant les résultats sur 100 échantillons simulés :

$p = \text{Prob}(x_i \notin I_{95\%})$ pour $i = 1, \dots, 20$:

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
p	0	0	0	0	0	0	0	0	0,01	0,02	0,04	0,05	0,05	0,08	0,10	0,07	0,07	0,09	0,12	0,10

$\text{Prob}(\hat{x}_{0,90} \notin I_{95\%}) = 0,04$ et $\text{Prob}(\exists i / x_i \notin I_{95\%}) = 0,23$

On constate que le taux théorique de sortie de 5% n'est pas vérifié pour l'ensemble des valeurs x_i de rang i de l'échantillon. Plus la valeur x_i est élevée, plus sa probabilité au non-dépassement est élevée et plus son incertitude est grande. On a donc une bonne adéquation de la loi ajustée avec la loi théorique pour des valeurs « faibles à moyennes », mais il n'est pas rare qu'une ou deux valeurs fortes se trouvent en dehors de l'intervalle de confiance $I_{95\%}$. Expérimentalement, ces valeurs fortes ont un taux empirique de sortie supérieure à 5%.

D'ailleurs, on observe que l'on a quasiment une chance sur quatre pour qu'une des 20 valeurs de l'échantillon est située hors des limites de l'intervalle $I_{95\%}$. Ainsi, 1, 2 voire 3 valeurs (sur un échantillon de 20 valeurs) en dehors de l'intervalle de confiance $I_{95\%}$ ne sont pas préjudiciables pour commenter l'adéquation de la loi ajustée avec la loi théorique et les valeurs d'échantillon simulées.

Enfin, l'estimation de la crue décennale $\hat{x}_{0,90}$ est assez bien déterminée et assez fiable puisque la probabilité $\text{Prob}(\hat{x}_{0,90} \notin I_{95\%})$ vaut 4

↳ Cf. l'onglet *Vérif taux de recouvrement* du fichier Excel "ENTPE_3A_Exercice_1_GETTVERT_Florian" pour vérifier le bon fonctionnement de la Macro 4 qui simule 100 échantillons et calcule les probabilités ci-dessus associées à ces 100 échantillons simulés dans le but de les comparer au taux théorique de sortie (5%) de l'intervalle $I_{95\%}$. Le script de cette Macro 4 est présenté et commenté ci-dessous.

Si on réalise le même tableau récapitulatif mais cette fois-ci sur 10 000 échantillons simulés, on obtient alors les valeurs suivantes :

$p = \text{Prob}(x_i \notin I_{95\%})$ pour $i = 1, \dots, 20$:

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
100*p	0,0	0,0	0,1	0,3	0,5	0,8	1,0	1,2	1,6	2,0	2,7	3,3	3,9	4,5	5,4	6,2	6,9	8,2	10,5	17,4

$\text{Prob}(\hat{x}_{0,90} \notin I_{95\%}) = 0,028$ et $\text{Prob}(\exists i / x_i \notin I_{95\%}) = 0,323$

Pour un plus large échantillon (10000 échantillons), les probabilités obtenues sont plus représentatives du comportement global observé. On peut en ressort ainsi les mêmes conclusions (mais avec davantage de précisions sur les chiffres) : une bonne estimation de la crue décennale, 1/3 de chance qu'une des valeurs de notre échantillon soit en dehors de l'intervalle de confiance et ce sont les plus fortes valeurs qui ont la probabilité la plus grande d'être en dehors de cet intervalle de confiance $I_{95\%}$. Plus la valeur est élevée, plus l'incertitude sur cette dernière est grande en hydrologie.



Description de la macro 4 avec commentaires

```
Sub Macro4_tx_de_recouvrement()  
,  
' Macro4_tx_de_recouvrement Macro  
,  
' Touche de raccourci du clavier: Ctrl+Shift+U  
,  
  
Dim numero As Integer 'Variable numero qui fixe le nombre de boucles à simuler  
  
Dim y As Integer 'Initialisation des cellules  
For i = 9 To 29  
Cells(i, 14) = 0 'Remise à 0 des probabilités suivantes :  $\text{Prob}(x_i \notin I_{95\%})$  et  $\text{Prob}(\hat{x}_{0,90} \notin I_{95\%})$   
Next i  
numero = 1 'Numéro de départ  
Cells(30, 14) = 0 ' Remise à 0 de la probabilité :  $\text{Prob}(\exists i / x_i \notin I_{95\%} \text{ pour } i = 1, \dots, 20)$   
  
While numero <= 100 'Début de boucle de simulation pour 100 échantillons  
Dim x As Integer  
For x = 9 To 29  
Cells(x, 14) = Cells(x, 14) + Cells(x, 11) 'Calcul des probabilités suivantes :  $\text{Prob}(x_i \notin I_{95\%})$  et  $\text{Prob}(\hat{x}_{0,90} \notin I_{95\%})$   
Next x 'Fin de la boucle For  
Cells(30, 14) = Cells(30, 14) + Cells(31, 9) 'Calcul de la probabilité  $\text{Prob}(\exists i / x_i \notin I_{95\%} \text{ pour } i = 1, \dots, 20)$   
numero = numero + 1 'La variable numero est augmenté de 1 à chaque boucle  
Sheets("Vérif taux de recouvrement").Select  
Application.Run _  
    "ENTPE_3A_Exercice_1_GETTVERT_Florian.xlsm!Macro3_Copie_generation_p_recouvrement"  
    'Génération d'un nouvel échantillon de 20 valeurs pour la prochaine simulation  
Wend 'Fin de la boucle While  
  
End Sub 'Fin du script
```



EXERCICE 2 : Estimation de la crue centennale sur l'Ardèche à Saint-Martin-d'Ardèche

La station hydrométrique de Saint-Martin-d'Ardèche contrôle un bassin versant de 2240 km². Une étude fréquentielle sur les débits de crue a donné les informations suivantes :

- le débit moyen maximal journalier (QJX) suit une loi de Gumbel de paramètres ($x_0 = 590$; $a = 460$) ;
- le rapport moyen de pointe de crue RX est égal à 2,3 et représente la moyenne des rapports (pour chaque évènement de crue) entre le débit de pointe QIX et le débit QJX ;
- le gradex des pluies journalières a_p (paramètre d'échelle de la distribution des pluies) vaut 32 mm.

1. Calcul du débit QJX(10) de période de retour 10 ans

La distribution des débits moyens maximum journaliers (QJX) suit une loi de Gumbel $Gu(x_0; a)$ tel que $x_0 = 590$ m³/s et $a = 460$ m³/s. La valeur de période de retour 10 ans est égale au quantile d'ordre 0,9 de la distribution des QJX. Ce quantile peut être calculé en inversant la fonction de répartition de la loi de Gumbel : $QJX(10) = x_0 - a \cdot \ln(-\ln(0,9))$.

On obtient alors **QJX(10) = 1625,2 m³/s**.

2. Estimation du débit instantané QIX(10) de période de retour 10 ans

On souhaite à présent estimer le débit instantané QIX(10) de période de retour 10 ans. Pour cela, on utilise la méthode Crupedix : $QIX(10) = S^{0,8} \cdot [PJ(10)/80]^2 \cdot K$ avec PJ(10) la pluie journalière décennale et K le coefficient régional ($0,1 < K < 1,75$).

Dans le cas présent, PJ(10) = 170 mm et K = 1,50. On trouve alors un débit de pointe décennal **QIX(10) de 3243,5 m³/s**.

3. Comparaison et limites de l'approche

Si l'on compare l'estimation du débit de pointe décennal QIX(10) ci-dessus de 3243,5 m³/s avec celle issue de la loi de Gumbel $QIX(10) = RX \cdot QJX(10) = 3737,9$ m³/s, on observe une différence d'environ 500 m³/s (soit 13% d'écart) entre les deux débits de pointe estimés.

Dans une optique « sécuritaire », on retiendra la valeur la plus forte pour le débit de pointe QIX(10) qui est celle issue de la loi de Gumbel : $QIX(10) = 3737,9$ m³/s.

4. Extrapolation du débit centennal de pointe QIX(100)

On souhaite à présent extrapoler au-delà de la crue décennal avec la méthode du Gradex et avec le modèle Agregée pour calculer la valeur du débit centennal de pointe QIX(100).



EXERCICE 2 : ESTIMATION DE LA CRUE CENTENNALE
SUR L'ARDECHE A SAINT-MARTIN-D'ARDECHE

	Méthode Gradex	Méthode Agréee
QJX(100) en m³/s	3535,5	3110,3
QIX(100) en m³/s	8131,6	7153,7

On a des estimations de la valeur du débit de pointe centennal assez éloignées : 8131 m³/s et 7153 m³/s (écart de 12%). Idem que précédemment, dans une optique « sécuritaire », on retiendra la valeur la plus forte pour le débit de pointe QIX(100) qui est QIX(100) = 7153,7 m³/s.



EXERCICE 3 : Simulation du comportement asymptotique de la distribution des crues



On considère un modèle hydrologique très simplifié du type :

$$Q = P - \min(P, S_{\max} - S).$$

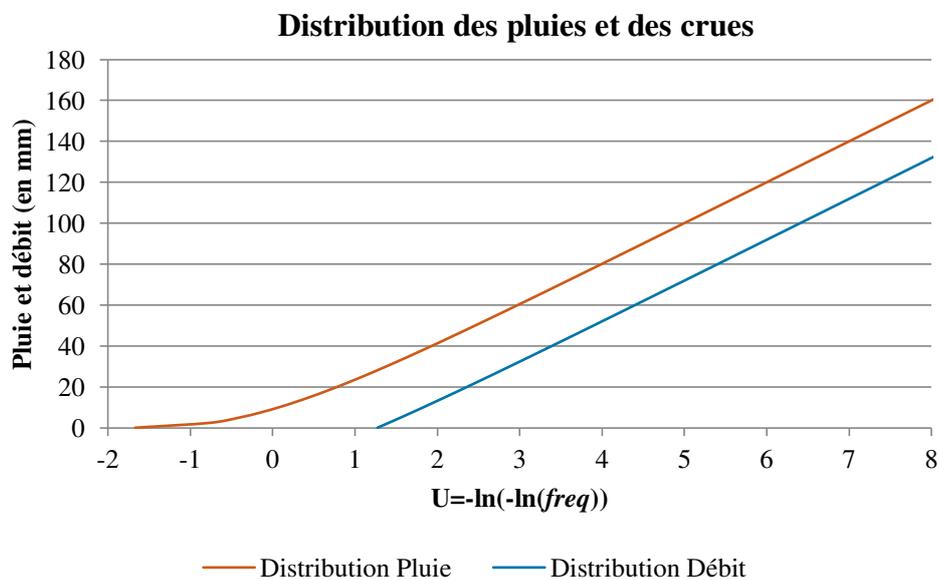
La distribution des pluies P est supposée suivre une loi exponentielle simple : $F_P(x) = \text{Prob}(P < x) = 1 - \exp(-x/a)$. La distribution de l'aléa hydrique (hauteur S du réservoir) est supposée suivre une loi uniforme : $F_S(y) = \text{Prob}(S < y) = y/S_{\max}$.

La distribution du débit est obtenue :

- Par approche analytique : $F_Q(z) = \text{Prob}(Q < z) = 1 - \frac{a}{S_{\max}} \left[1 - \exp\left(-\frac{S_{\max}}{a}\right) \right] \exp\left(-\frac{z}{a}\right)$
- Par simulation :
 - ✓ Tirage aléatoire des pluies selon une loi exponentielle simple et des niveaux de remplissage selon une loi uniforme ;
 - ✓ Transformation des couples P-S en débits grâce au modèle hydrologique ;
 - ✓ Classement et probabilisation empirique des débits.

1. Distribution des pluies et des crues

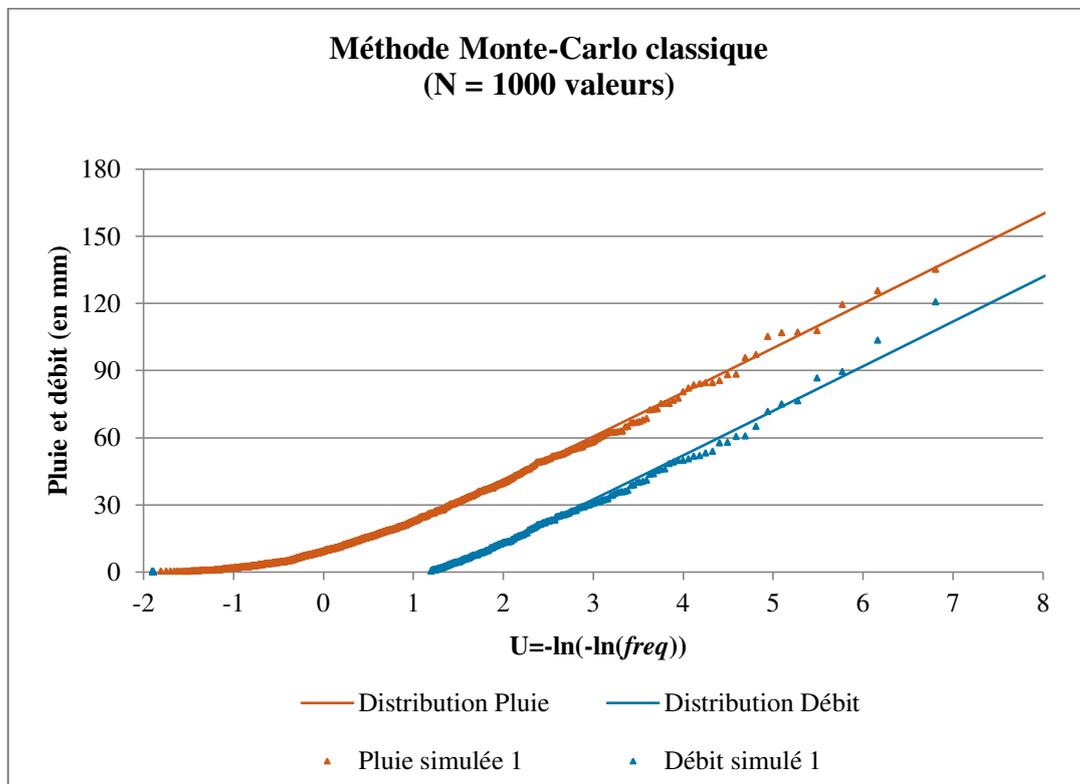
La distribution des pluies et des crues à partir des formulations analytiques susmentionnées (avec $a=20$ et $S_{\max}=80$) est tracée sur le graphique ci-dessous. En abscisse, le changement de variable suivant $U = -\ln(-\ln(freq))$ est utilisé.



2. Pluies et débits simulés

On souhaite simuler à présent des pluies et des débits de deux façons² : par tirage Monte-Carlo classique (méthode n°1) et par tirage Monte-Carlo pondéré (méthode n°2).

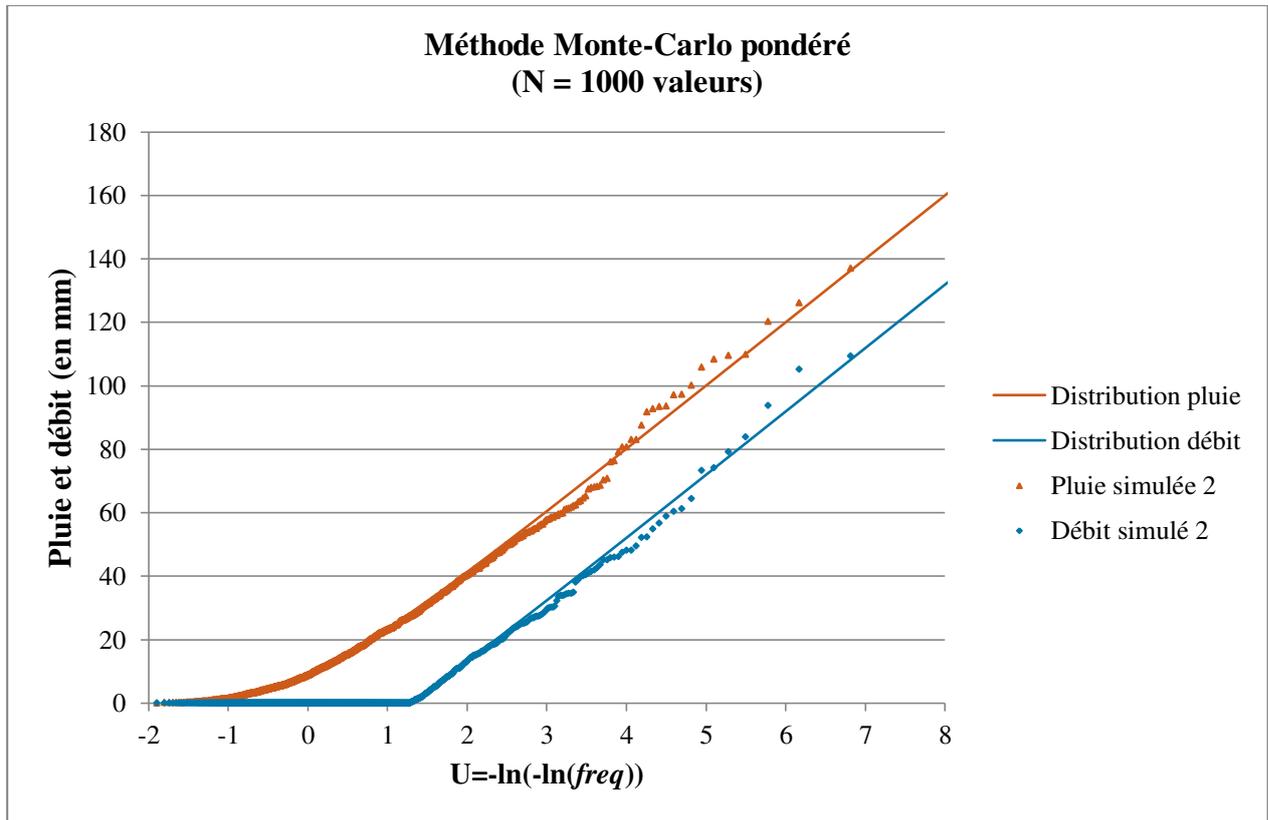
- Méthode n°1 : tirage Monte-Carlo classique



Avec le tirage Monte Carlo classique, on a **une bonne estimation des distributions de pluie et de débit théoriques** puisque l'on constate peu d'éloignement des valeurs simulées avec les courbes théoriques et il n'y a pas de différence de forme notable entre les courbes empiriques (échantillon simulé) et théoriques qui laisserait présager un manque d'adéquation entre la distribution estimée et les observations.

² Cf. le fichier Excel « ENTPE_3A_Exercice_3_GETTVERT_Florian » pour visionner les feuilles de calculs.

- **Méthode n°2 : tirage Monte-Carlo pondéré**



De même que précédemment, on a une **bonne estimation des distributions de pluie et de débit théoriques** avec le tirage Monte Carlo pondéré puisque l'on constate une nouvelle fois peu d'éloignement des valeurs simulées avec les courbes théoriques et il n'y a pas non plus de différence de forme notable entre les courbes empiriques (échantillon simulé) et théoriques qui laisserait présager un manque d'adéquation entre la distribution estimée et les observations.

3. Démonstration de la distribution F_Q du débit

Afin de retrouver la formulation analytique de la distribution F_Q du débit, on se propose de raisonner en deux étapes pour traiter l'équation : débit $Q_i = x_i - \min(x_i, S_{\max} - y_j)$.

- **Cas du débit non nul** : $Q > 0$ si $P > S_{\max} - y$

Lorsque la pluie est supérieure à la capacité d'infiltration du sol, le débit vaut $Q = P - (S_{\max} - y)$ et la fonction de répartition F_Q s'exprime ainsi :

$$F_Q(z) = Prob(Q < z) = \int_0^{S_{\max}} Prob(P < S_{\max} - y + z) \cdot f_s(y) \cdot dy \text{ avec } f_s(y) = F'_s(y) = 1/S_{\max}$$

On obtient alors :



$$F_Q(z) = \frac{1}{S_{max}} \int_0^{S_{max}} \left(1 - e^{-\frac{S_{max}-y+z}{a}}\right) \cdot dy = \frac{1}{S_{max}} \left(S_{max} - e^{-\frac{S_{max}+z}{a}} \int_0^{S_{max}} \frac{y}{e^{\frac{y}{a}}} dy\right)$$

$$= \frac{1}{S_{max}} \left[S_{max} - e^{-\frac{S_{max}+z}{a}} \cdot \left(a \cdot e^{S_{max}/a} - a\right)\right] = 1 - \frac{a}{S_{max}} \left(e^{\frac{S_{max}}{a}} - 1\right) e^{-(S_{max}+z)/a}$$

D'où : $F_Q(z) = \text{Prob}(Q < z) = 1 - \frac{a}{S_{max}} \left[1 - \exp\left(-\frac{S_{max}}{a}\right)\right] \exp\left(-\frac{z}{a}\right)$ pour $z > 0$.

- **Cas du débit nul :** $Q = 0$ si $P < S_{max} - y$

$$\text{Prob}(Q = 0) = \int_0^{S_{max}} \text{Prob}(P < S_{max} - y) \cdot f_S(y) \cdot dy = \frac{1}{S_{max}} \int_0^{S_{max}} \left(1 - e^{-\frac{S_{max}-y}{a}}\right) dy$$

$$= \frac{1}{S_{max}} \left[S_{max} - e^{-\frac{S_{max}}{a}} \int_0^{S_{max}} e^{\frac{y}{a}} dy\right] = 1 - \frac{1}{S_{max}} e^{-\frac{S_{max}}{a}} \left(a \cdot e^{\frac{S_{max}}{a}} - a\right)$$

D'où : $\text{Prob}(Q = 0) = 1 - \frac{a}{S_{max}} \left(1 - e^{-\frac{S_{max}}{a}}\right)$

Remarque : on a bien la continuité de la distribution F_Q pour $z = 0$ puisque $F_Q(z) \xrightarrow{z \rightarrow 0} \text{Prob}(Q = 0)$.

Donc, pour tout $z \geq 0$, on a bien $F_Q(z) = \text{Prob}(Q < z) = 1 - \frac{a}{S_{max}} \left[1 - \exp\left(-\frac{S_{max}}{a}\right)\right] \exp\left(-\frac{z}{a}\right)$.

