Centre Universitaire Abdelhafid-Boussouf –Mila-
Institut de science et technologie
Département de science et technologie

Cours analyse et modélisation hydrologique

Hydrologie statistique (partie 02)

Master 01 : hydraulique urbaine

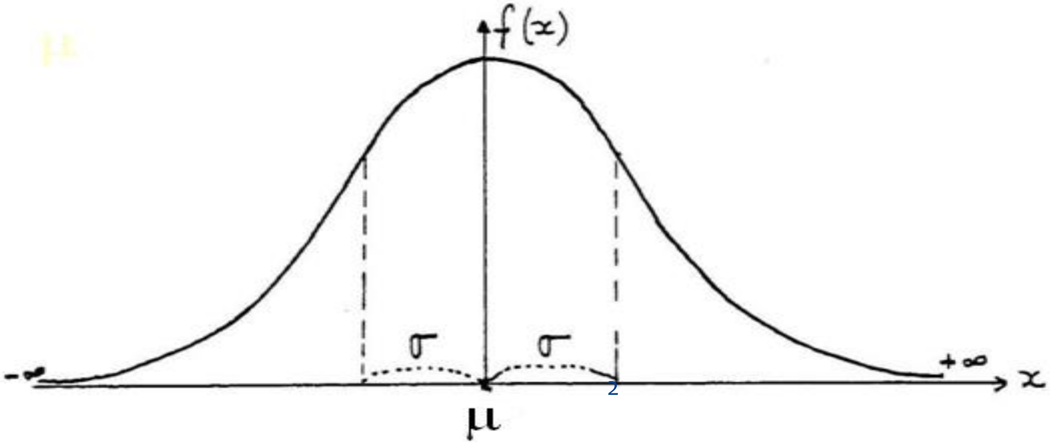
Loi Normale ou loi de Gauss

Fonction de densité:

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

Fonction de distribution:

$$\mathrm{F}(x)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx$$



Loi Normale ou loi de Gauss

La moyenne : $\overline{X} = \frac{\sum x}{n}$

L'écart-type: $S = \sqrt{\frac{\sum(x-\overline{x})^2}{N-1}}$

La droite de Henry : $x = \bar{x}$+z*s

Intervalles de confiance

- **Intervalles de confiance**
- On peut, théoriquement, tirer plusieurs échantillons à partir d'une population donnée.
- On pourrait donc théoriquement tirer plusieurs échantillons de pluies moyennes annuelles ayant chacun une longueur de 20 ans, à partir d'une population de 1000 valeurs de pluies moyennes annuelles, si celle dernière existait.

Intervalles de confiance

- Chaque échantillon aura sa propre moyenne et son propre écart-type, presque tous différents les uns des autres $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_{20}$ et $s_1, s_2, \ldots, s_{20}$ mais oscillant autour de la moyenne de la population μ et son écart-type σ.
- En hydrologie, on ne dispose, en général et si on a de la chance, que d'un échantillon dont on calcule la moyenne et l'écart-type et l'on ignore si les valeurs calculées $\bar{x}, s$ sont égales à celles de la population μ et σ.

Intervalles de confiance

- Il devient nécessaire, devant cette incertitude, de compléter notre information en déterminant autour de la valeur estimée (moyenne, écart-type ou quantile), un intervalle dont on a de bonnes raisons de croire qu'il contient la vraie valeur du paramètre.
- Ceci nous amène à la notion d'intervalle de confiance.

Intervalles de confiance

- Supposons que l'on s'intéresse à un paramètre quelconque d'une population, par exemple la moyenne μ.
- On dispose d'une estimation $\bar{x}$, déterminée à partir d'un échantillon.
- On se propose de déterminer de part et d'autre de les limites x_1 et x_2 de l'intervalle qui a une forte probabilité de contenir la vraie valeur de μ.

Intervalles de confiance

- On détermine les limites de confiance x_1 et x_2, de telle sorte que $$\Pr ob\{x_1 \leq \mu \leq x_2\} = 1 - \alpha$$
- Ou α est le coefficient de confiance ou coefficient de sécurité.
- On a aussi $$\Pr ob\{\mu < x_1 \text{ ou } \mu > x_2\} = \alpha = \text{coefficient de risque}$$ $$= \text{seuil de signification}$$
- et $$\Pr ob\{\mu < x_1\} = \Pr ob\{\mu > x_2\} = \alpha/2$$

$$\Pr ob\{\bar{x}_1 \leq \mu \leq \bar{x}_2\} = 1 - \alpha$$

Intervalles de confiance

- Ou bien

$$\Pr ob\{\mu < \bar{x}_1\} = \Pr ob\{\mu > \bar{x}_2\} = \alpha/2$$

- En calculant les limites de confiance par rapport a la moyenne observée.

$$\left.\begin{array}{l}\bar{x}_1 = \bar{x} - d_1 \\ \bar{x}_2 = \bar{x} + d_2\end{array}\right\} \Rightarrow \Pr ob(\mu < \bar{x} - d_1) = \Pr ob(\mu > \bar{x} + d_2) = \frac{\alpha}{2}$$

- Ou bien

$$\Pr ob(\bar{x} - \mu > d_1) = \Pr ob(\mu - \bar{x} > d_2) = \frac{\alpha}{2}$$

Intervalles de confiance

- Si on réduit notre variable étudiée, on obtient

$$\Pr ob\left(\frac{\bar{x}-\mu}{s_{\bar{x}}} > \frac{d_1}{s_{\bar{x}}}\right) = \Pr ob\left(\frac{\mu-\bar{x}}{s_{\bar{x}}} > \frac{d_2}{s_{\bar{x}}}\right) = \frac{\alpha}{2}$$

- Les variables centrées réduites

$$\begin{cases} z_1 = \dfrac{x-\mu}{s_{\bar{x}}} \\ z_2 = \dfrac{\mu-\bar{x}}{s_{\bar{x}}} \end{cases}$$

- Donc

$$\Pr ob\left(z_1 > \frac{d_1}{s_{\bar{x}}}\right) = \Pr ob\left(z_2 > \frac{d_2}{s_{\bar{x}}}\right) = \frac{\alpha}{2}$$

Intervalles de confiance

$$\mathrm{Prob}\left(z_1 \leq \frac{d_1}{s_{\bar{x}}}\right) = \mathrm{Prob}\left(z_2 \leq \frac{d_2}{s_{\bar{x}}}\right) = 1 - \frac{\alpha}{2}$$

$$\phi\left(\frac{d_1}{s_{\bar{x}}}\right) = \phi\left(\frac{d_2}{s_{\bar{x}}}\right) = 1 - \frac{\alpha}{2}$$

$$\frac{d_1}{s_{\bar{x}}} = \frac{d_2}{s_{\bar{x}}} = z_{1-\frac{\alpha}{2}} \quad \Longrightarrow \quad d_1 = d_2 = z_{1-\frac{\alpha}{2}} * s_{\bar{x}}$$

- D'ou

$$\left.\begin{aligned} \bar{x}_1 &= \bar{x} - d_1 \\ \bar{x}_2 &= \bar{x} + d_2 \end{aligned}\right\} \Rightarrow \left\{\begin{aligned} \bar{x}_1 &= \bar{x} - z_{1-\frac{\alpha}{2}} * s_{\bar{x}} \\ \bar{x}_2 &= \bar{x} + z_{1-\frac{\alpha}{2}} * s_{\bar{x}} \end{aligned}\right.$$

Intervalles de confiance

- L'intervalle de confiance est donc $\bar{x} \pm z_{1-\frac{\alpha}{2}} * s_{\bar{x}}$

$$\bar{x} - z_{1-\frac{\alpha}{2}} * s_{\bar{x}} \leq \mu \leq \bar{x} + z_{1-\frac{\alpha}{2}} * s_{\bar{x}}$$

- Détérmination de l'écart-type de la moyenne

$$E[\bar{x}] = E\left[\frac{1}{N}\sum x_i\right] = \frac{1}{N}\sum E[x] = \frac{1}{N}.N.E[x] = \bar{x}$$

$$Var[\bar{x}] = Var\left[\frac{1}{N}\sum x_i\right] = \frac{1}{N^2}\sum Var[x] = \frac{1}{N^2}.N.Var[x] = \frac{1}{N}.s^2$$

Intervalles de confiance

- L'écart-type de la moyenne

$$s_{\bar{x}} = \frac{s_x}{\sqrt{N}}$$

- L'intervalle de confiance de la moyenne s'écrit donc

$$\bar{x} - \frac{s_x}{\sqrt{N}} * z_{1-\frac{\alpha}{2}} \leq \mu \leq \bar{x} + \frac{s_x}{\sqrt{N}} * z_{1-\frac{\alpha}{2}}$$

- L'intervalle de confiance de l'ecart-type s'écrit donc

$$s_x - \frac{s_x}{\sqrt{2N}} * z_{1-\frac{\alpha}{2}} \leq \sigma \leq s_x + \frac{s_x}{\sqrt{2N}} * z_{1-\frac{\alpha}{2}}$$

Intervalles de confiance

- L'intervalle de confiance du quantile

$$x_F = \bar{x} + s_x * z_F$$

$$s_{x_F}^2 = s_{\bar{x}}^2 + s_{s_x}^2 * z_F^2$$

$$s_{x_F}^2 = \frac{s_x^2}{N} + \frac{s_x^2}{2N} * z_F^2 = \frac{s_x^2}{2N}\left(2 + z_F^2\right)$$

$$s_{x_F} = \frac{s_x}{\sqrt{2N}}\sqrt{\left(2 + z_F^2\right)}$$

- L'intervalle de confiance du quantile s'écrit donc

$$x_F - \frac{s_x}{\sqrt{2N}}\sqrt{2 + z_F^2} * z_{1-\frac{\alpha}{2}} \le x_F \le x_F + \frac{s_x}{\sqrt{2N}}\sqrt{2 + z_F^2} * z_{1-\frac{\alpha}{2}}$$

Intervalles de confiance

- **a**- pour la moyenne

$$\bar{x} - z_{\frac{1-\alpha}{2}} \frac{s}{\sqrt{N}} < \mu < \bar{x} + z_{\frac{1-\alpha}{2}} \frac{s}{\sqrt{N}}$$

- **b**- pour l'écart-type

$$s - z_{\frac{1-\alpha}{2}} \frac{s}{\sqrt{2N}} < \sigma < s + z_{\frac{1-\alpha}{2}} \frac{s}{\sqrt{2N}}$$

- **c**- et pour un quantile x_p de probabilité p

$$x_p - z_{\frac{1-\alpha}{2}} \frac{s}{\sqrt{2N}} \sqrt{2 + z_p^2} < x_p < x_p + z_{\frac{1-\alpha}{2}} \frac{s}{\sqrt{2N}} \sqrt{2 + z_p^2}$$

Intervalles de confiance

L'intervalle de confiance de la loi Normale

La moyenne :

$$\bar{x} - z_{\frac{1-\alpha}{2}} \frac{s}{\sqrt{N}} < \mu < \bar{x} + z_{\frac{1-\alpha}{2}} \frac{s}{\sqrt{N}}$$

Intervalles de confiance

L'écart-type :

$$s - z_{\frac{1-\alpha}{2}} \frac{s}{\sqrt{2N}} < \sigma < s + z_{\frac{1-\alpha}{2}} \frac{s}{\sqrt{2N}}$$

un quantile xp de probabilité p :

$$x_p - z_{\frac{1-\alpha}{2}} \frac{s}{\sqrt{2N}} \sqrt{2 + z_p^2} < x_p < x_p + z_{\frac{1-\alpha}{2}} \frac{s}{\sqrt{2N}} \sqrt{2 + z_p^2}$$

Le risque α

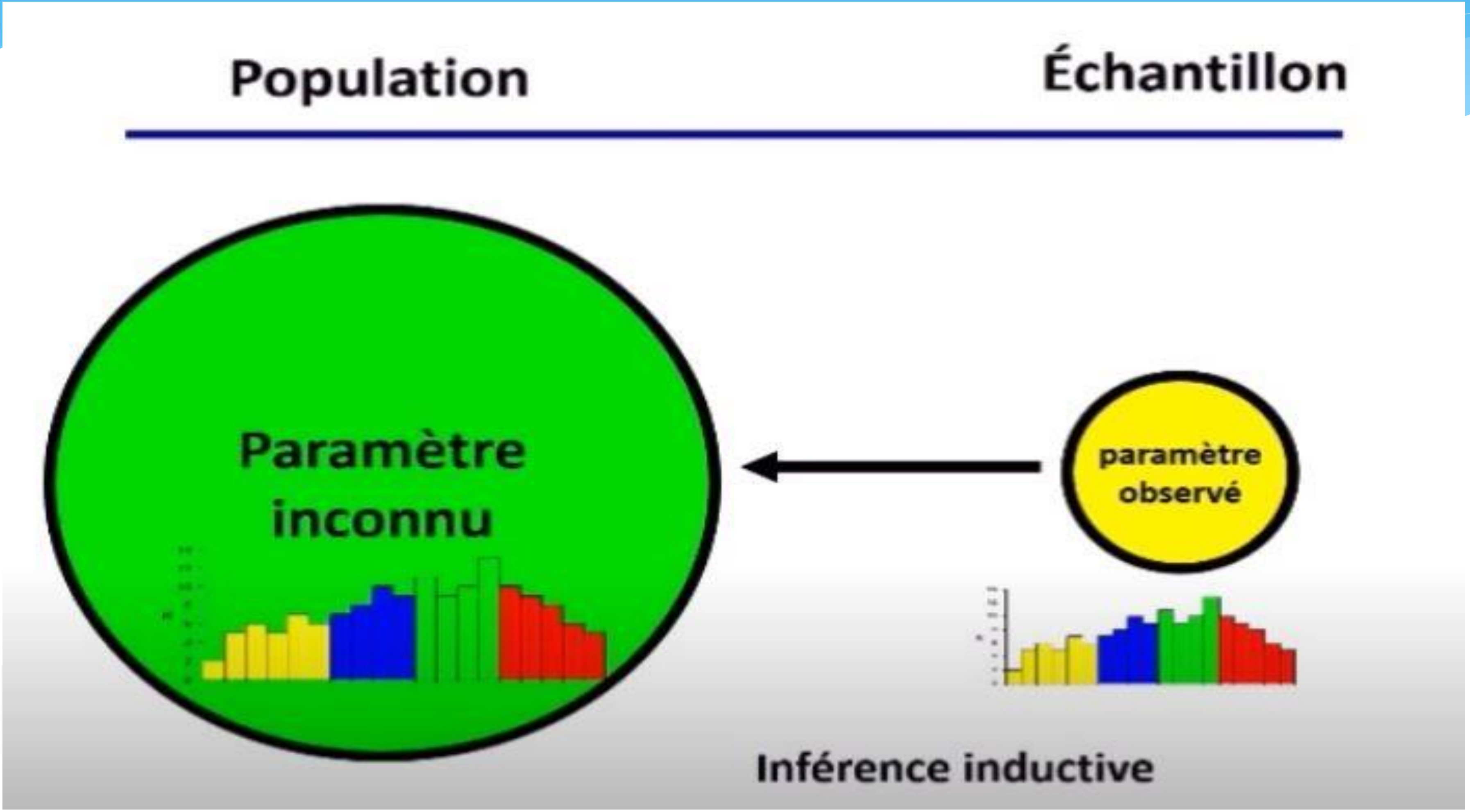
Le risque α est une probabilité

Le risque α est un risque d'erreur

A>B **α= 0.03**

Il y a 3 chances sur cent que A>B soit faux

Le risque α



Population
Échantillon
Paramètre
inconnu
paramètre
observé
Inférence inductive

Domaine du risque α

- Estimation d'un paramètre par intervalle de confiance
- Résultat d'un test statistique
- Adéquation d'un modèle avec situation observée

Estimation d'un paramètre par intervalle de confiance

Exemple : étude de l'écart-type des pluies

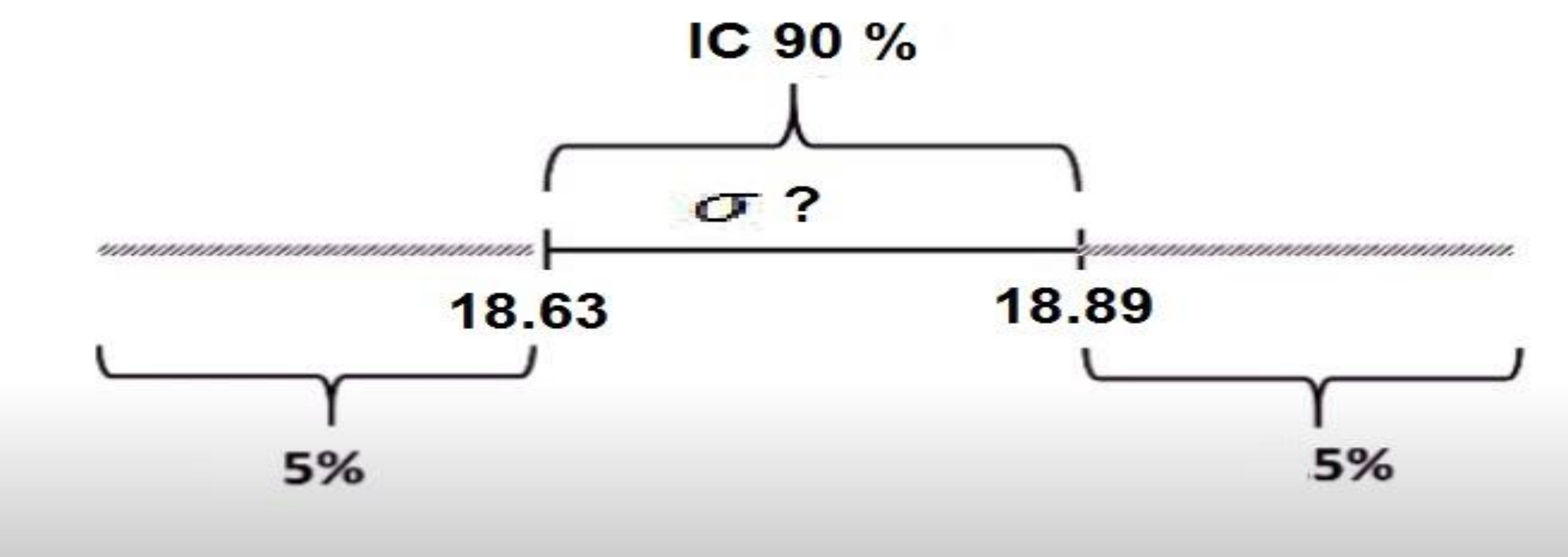
L'écart-type de la pluie :

L'écart-type de l'Echantillon = 18.76 mm ; N = 42

L'écart-type de la population : µ ?

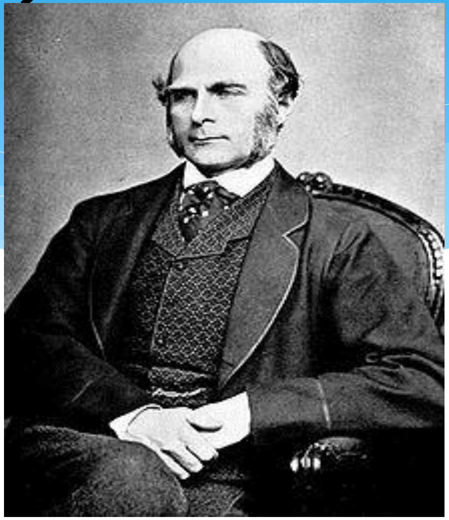
IC 90 % [18.63, 18.89]

Estimation d'un paramètre par intervalle de confiance



Loi log Normale (Galton)

Si le logarithme d'une variable quelconque a une distribution normale, alors cette variable possède une distribution log-normale.



Francis Galton

Si par exemple : y = Ln x suit une loi normale alors x suit une loi log-normale.

Loi log Normale (Galton)

La fonction de densité de probabilité s'écrit alors :

$$\begin{cases} f(x) = \dfrac{1}{x\sigma_n\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(Lnx-\mu_n)^2}{\sigma_n^2}} \quad ; \quad x > 0 \\ f(x) = 0 \qquad x \leq 0 \end{cases}$$

Loi log Normale (Galton)

Avec :

$$\begin{cases} \mu_n = \dfrac{1}{N}\displaystyle\sum_{1}^{N} Lnx_i \\ \sigma_n^2 = \dfrac{1}{N}\displaystyle\sum_{1}^{N}\left(Lnx_i - \mu_n\right)^2 \end{cases}$$

Loi log Normale (Galton)

La moyenne et la variance de la distribution log-normale peuvent aussi s'écrire :

$$\begin{cases} \mu_n = \dfrac{1}{2} Ln\left(\dfrac{\mu_x^2}{C_v^2+1}\right) \\ \sigma_n^2 = Ln\left(C_v^2+1\right) \end{cases} \longrightarrow \begin{cases} \mu_x = e^{\mu_n + \frac{\sigma_n^2}{2}} \\ \sigma_x^2 = \mu_x^2\left(e^{\sigma_n^2}-1\right) \end{cases}$$

Où μ_x est la moyenne de la variable étudiée (sans logarithme) et C_v est le coefficient de variation : $C_v = \sigma/\mu$.

Loi log Normale (Galton)

La moyenne : $\overline{Ln\ p} = \frac{\sum Ln\ pi}{N}$

L'écart-type: $S_{Lnp} = \sqrt{\frac{\sum(lnp_i)^2 - N\overline{Lnp^2}}{N-1}}$

La droite de Henry : $\ln p_i = \overline{Ln\ p} + Z_i * S_{Lnp}$

Intervalles de confiance

L'intervalle de confiance de la loi log Normale (GALTON)

La moyenne :

$$\ln \bar{P} - Z_{\frac{1-\alpha}{2}} * \frac{S_{\ln p}}{\sqrt{2N}} * \sqrt{2 + Z^2_{Ln\bar{P}}} < Ln\,\bar{P} < \ln \bar{P} + Z_{\frac{1-\alpha}{2}} * \frac{S_{\ln p}}{\sqrt{2N}} * \sqrt{2 + Z^2_{Ln\bar{P}}}$$

$$Z_{Lnp} = \frac{ln\bar{P} - \overline{LNP}}{Slnp}$$

L'écart-type :

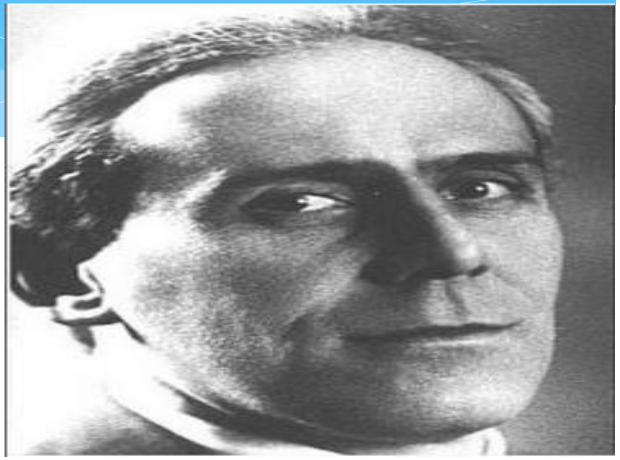
$$S_{lnp} - Z_{\frac{1-\alpha}{2}} * \frac{S_{\ln p}}{\sqrt{2N}} < S_{lnp} < S_{lnp} + Z_{\frac{1-\alpha}{2}} * \frac{S_{\ln p}}{\sqrt{2N}}$$

Un quantile xp de probabilité p :

$$\ln P_{10} - Z_{\frac{1-\alpha}{2}} * \frac{S_{\ln p}}{\sqrt{2N}} * \sqrt{2 + Z^2_{Ln\,p_{10}}} < \ln p_{10} < \ln P_{10} + Z_{\frac{1-\alpha}{2}} * \frac{S_{\ln p}}{\sqrt{2N}} * \sqrt{2 + Z^2_{Ln\,p_{10}}}$$

Loi de Gumbel

C'est une loi très importante, qui sert dans l'analyse fréquentielle des valeurs extrêmes, et sera notamment l'ingrédient essentiel, en hydrologie opérationnelle, de la méthode du Gradex pour le calcul des crues de projet.



Émil Julius Gumbel

Loi de Gumbel

Pour l’étude des pluies extrêmes (ou n'importe quel autre événement d'une rare fréquence), on constitue un échantillon de N valeurs, chacune d'elles représentant la précipitation journaliere la plus forte d'une des N années.

Loi de Gumbel

On parvient généralement à ajuster à cet échantillon la loi de Gumbel ou la loi Log Normale.

La fonction de répartition de la loi de Gumbel est :

$$\boxed{F(x) = e^{-e^{-\alpha(x-\beta)}}}$$

Loi de Gumbel

Ou F(x) = fréquence au non dépassement =FND =F.

et α, β = coefficients d'ajustement.

Par un changement de variable y = α (x-β), la loi de Gumbel est :

$$\boxed{F(x) = F(y) = e^{-e^{-y}}}$$

Ou y est la variable réduite de Gumbel, liée à la probabilité attachée à la valeur x,

et F(y) = fréquence au non dépassement de la variable réduite y.

Loi de Gumbel

L'équation $y = \alpha\,(x-\beta)$ présentée sous forme :

$$\boxed{x = \frac{1}{\alpha}y + \beta}$$

est l'équation d'une droite qui représente la loi de Gumbel sur du papier à l'échelle de probabilité Gumbel.

Loi de Gumbel

Le papier de probabilité Gumbel porte en graduation d'abscisse deux (2) échelles :

- Une échelle de fréquence au non dépassement FND;
- Une échelle arithmétique de la variable réduite y.

A chaque valeur de y de la seconde échelle correspond, sur le papier échelle, la valeur de la fréquence au non dépassement calculée par l'expression :

$$\boxed{F(x) = e^{-e^{-y}}}$$

Loi de Gumbel

L'ordonnée, sur le papier de probabilité Gumbel, représente, sur une échelle arithmétique, la variable étudiée x.

La représentation graphique de l'échantillon des valeurs extrêmes sur papier Gumbel est obtenue en portant en ordonnées les valeurs de x et en abscisses les fréquences expérimentales au non dépassement :

$$F_i(x) = {(n_i - 0,5)}/{N}$$

Loi de Gumbel

Les valeurs de $1/\alpha$ et β sont déterminées en utilisant :

- Soit la méthode des moindres carrés qui minimise la somme des carrés des écarts entre les valeurs observées et les valeurs estimées par le modèle considéré;

- Soit la résolution d'un système d'équations formé avec les moments des deux premiers ordres.

Loi de Gumbel

Ces méthodes ne sont pas détaillées ici.

L'estimation des paramètres α et β par la méthode des moments donne :

$$\begin{cases} \alpha = \dfrac{\pi}{\sigma\sqrt{6}} \\ et \\ \beta = \mu - \dfrac{0{,}5772}{\alpha} \end{cases}$$

Ou μ et σ sont respectivement la moyenne et l'écart-type de la serie étudiée.

Loi de Gumbel

Une fois les paramètres de la droite de Gumbel déterminés, on trace la droite en calculant trois valeurs de x à partir de la valeur de y en utilisant l'équation de la variable réduite de Gumbel.

L'estimation de la valeur que prendrait la variable étudiée pour une probabilité donnée peut se faire soit par la lecture directe du graphe, soit en la calculant grâce à la formule :

$$\boxed{x_T = \frac{1}{\alpha} y + \beta}$$

Avec : $\boxed{y = -Ln(-Ln(F(x)))}$

Loi de Gumbel

En remplaçant α et β par leurs expressions respectives on obtient :

$$\boxed{x_T = \frac{\sigma\sqrt{6}}{\pi} y + \mu - \frac{0{,}5772}{\alpha} = \frac{\sigma\sqrt{6}}{\pi} y + \mu - 0{,}5772 * \frac{\sigma\sqrt{6}}{\pi}}$$

$$\boxed{x_T = \mu - \frac{\sigma\sqrt{6}}{\pi}\left(0{,}5772 - y\right)}$$

Or :

$$\boxed{y = -Ln(-Ln(F_{ND}(x)))}$$ et : $$\boxed{F_{ND} = \frac{T-1}{T}}$$

Loi de Gumbel

$$y = -Ln(-Ln(\frac{T-1}{T})) = -Ln\left(Ln\frac{T}{T-1} \right)$$

D'où :

$$x_T = \mu - \frac{\sqrt{6}}{\pi}\left\{0{,}5772 + Ln\left[\frac{T}{T-1}\right]\right\} * \sigma$$

$$x_T = \mu + K_T * \sigma$$

Avec :

$$K_T = -\frac{\sqrt{6}}{\pi}\left\{0{,}5772 + Ln\left[\frac{T}{T-1}\right]\right\}$$

$$K = -0{,}7797\left[0{,}5572 + \ln\left\{\ln\left(\frac{T}{T-1}\right)\right\}\right]$$

K_T est appelé **facteur de fréquence**.

Loi de Gumbel

La moyenne : $$\overline{X} = \frac{\sum x}{n}$$

L'écart-type: $S = \sqrt{\frac{\sum(x-\overline{x})^2}{N}}$

La droite de Henry : $x = \frac{1}{\alpha}\,y + X_0$

$\frac{1}{\alpha}$ = 0.78 *S

X_0 = $\overline{X}$ - $\frac{0.577}{\alpha}$

Loi de Gumbel

L'intervalle de confiance de la loi de GUMBEL

$$h_1, h_2 = \frac{\left(t_\alpha / N^{0,5}\right) * \left(1 + 1{,}13 * t_F + 1{,}1 * t_F^2\right)^{0,5} \pm t_\alpha^2 / N \left(1{,}1 * t_F + 0{,}57\right)}{1 - 1{,}1\, t_\alpha^2 / N}$$