UCLouvain

Analysis and mitigation of floods

Assignment - Statistical methods

Sandra Soares-Frazão

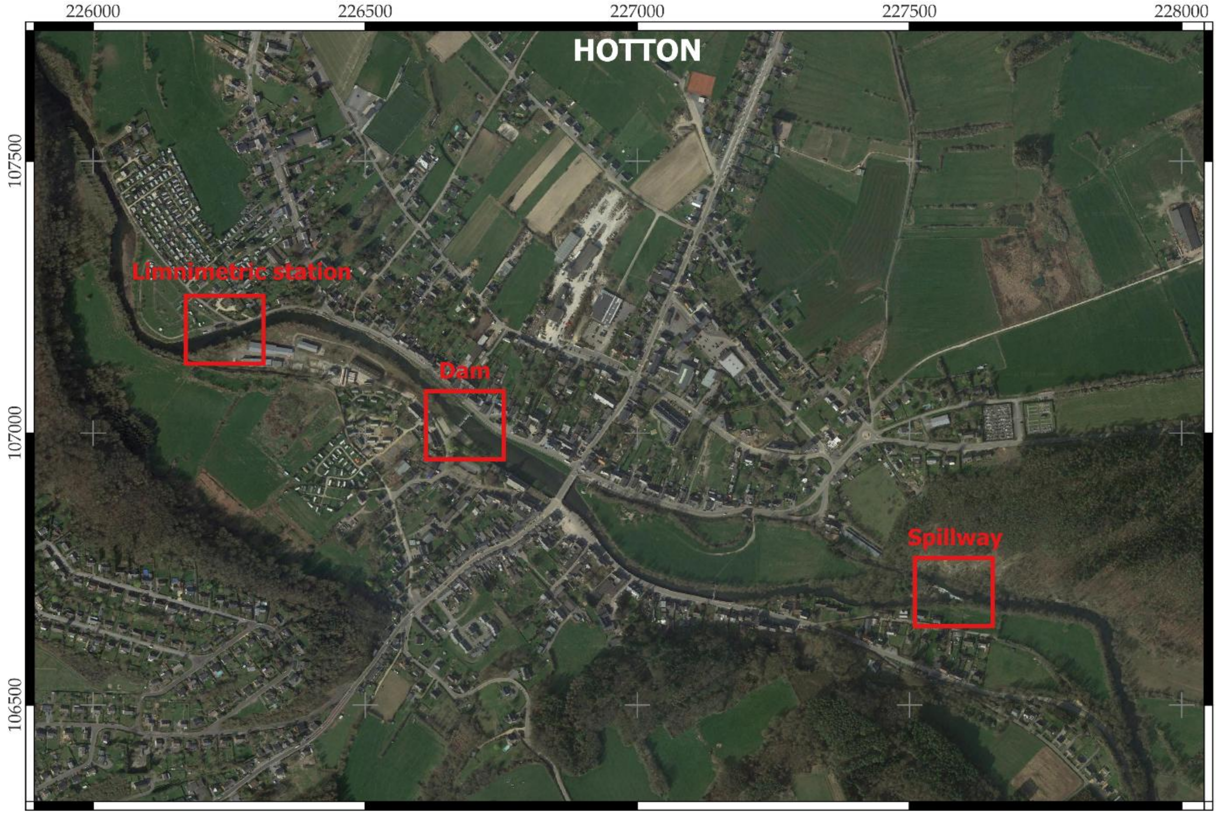
Nathan Delpierre

Martin Petitjean

Contents

Introduction

- Calendar year = 01/01 to 31/12
- Hydrological year = 01/09 to 31/08



Standard normal distribution

- Standard normal distribution $\mathrm{U} \sim \mathcal{N}(0,1)$
  - Probability density function

$$f_U(u) = \frac{1}{\sqrt{2\pi}} e^{\frac{-u^2}{2}}$$

  - Cumulative distribution function

$$F_U(u) = \int_{-\infty}^{u} f_U(v)dv$$

  - Standardization of a random variable X with a mean m_X and standard deviation σ_X

$$u = \frac{X - m_X}{\sigma_X} \qquad F_X(x) = F_U\left(\frac{x - m_X}{\sigma_X}\right)$$

Lognormal distribution

- Y can be represented as a product of random variables W_i

$$Y = W_0 \cdot W_1 \cdot W_2 \ldots W_n$$

- Central limit theorem

If X_i is a series of i.i.d (independent and identically distributed) random variable, then $Y = X_1 + X_2 + \ldots + X_n$ tend to follows a normal distribution for $n \to \infty$.

- We need a sum !

$$\mathrm{X} = \ln(Y) = \ln(W_0 \cdot W_1 \cdot W_2 \ldots W_n) = \ln(W_0) + \ldots + \ln(W_n)$$

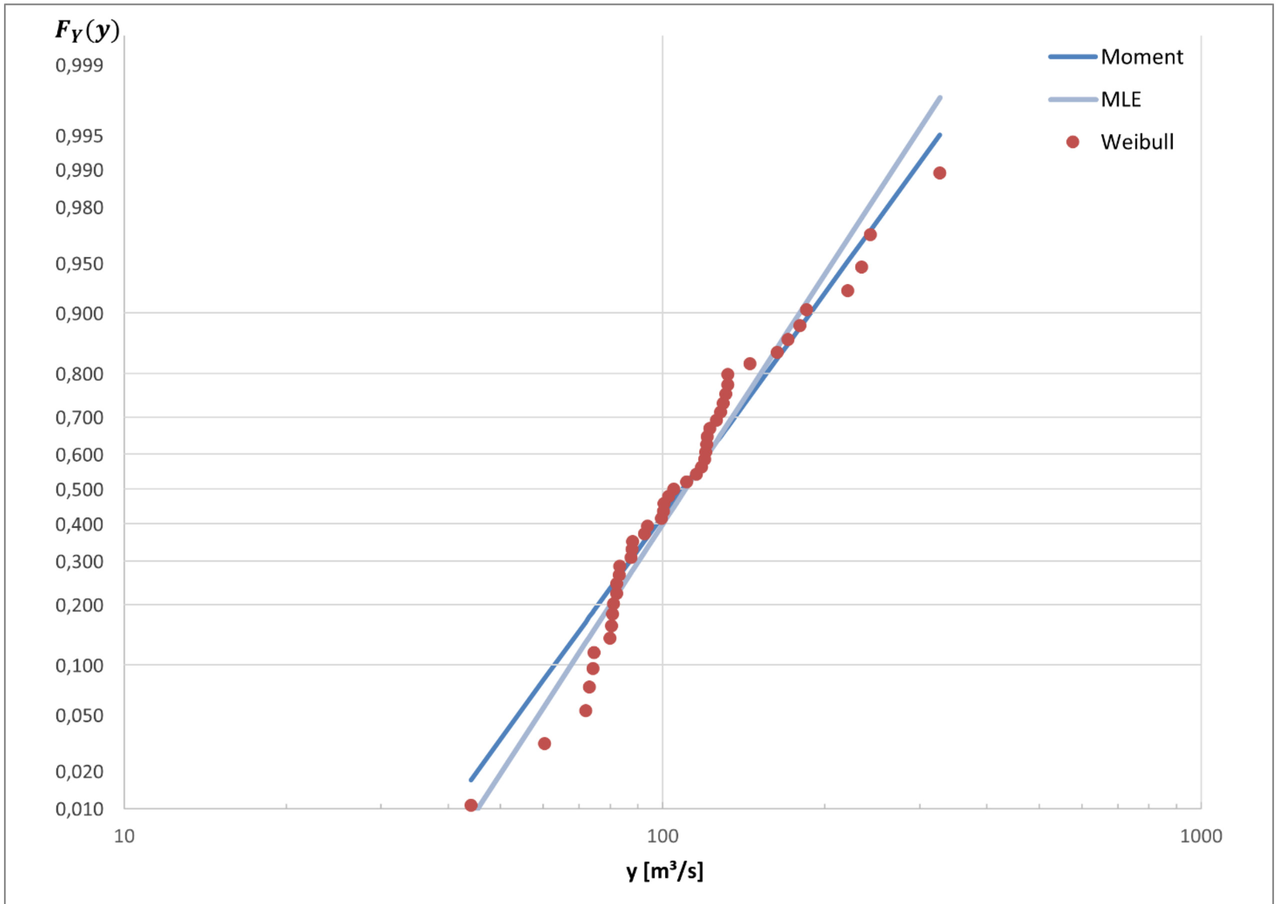
$$f_Y(y) = \frac{1}{y} \frac{1}{\sqrt{2\pi}\, \sigma_{\ln Y}} e^{\frac{-1}{2}\left(\frac{\ln y - m_{\ln Y}}{\sigma_{\ln Y}}\right)^2}$$

Lognormal distribution

- Moments
  - $\hat{m}_{ln\,Y} = \ln m_Y - \frac{1}{2}\sigma^2_{\ln Y}$
  - $\hat{\sigma}^2_{\ln Y} = \ln\left(\frac{\sigma_Y^2}{m_Y^2} + 1\right)$
- Maximum Likelihood Expectation (MLE)
  - $\hat{m}_{ln\,Y} = \frac{1}{n}\sum_{i=1}^{n} \ln Y_i$
  - $\hat{\sigma}^2_{\ln Y} = \frac{1}{n}\sum_{i=1}^{n}(\ln Y_i - m_{\ln Y})^2$
- Standardization of Y and PDF / CDF using $\mathrm{U} \sim \mathcal{N}(0,1)$
  - $U_i = \frac{\ln(Y_i) - m_{\ln Y}}{\sigma_{\ln Y}}$

Graphical representation

- $U_i = \frac{X_i - m_{\ln Y}}{\sigma_{\ln Y}} = \frac{\ln(Y_i) - m_{\ln Y}}{\sigma_{\ln Y}}$



x-axis : Q (discharge in m³/s) which corresponds to Y_i and are plotted on a log scale

y-axis : U_i (standardized variable) to get a linear relationship BUT labels corresponds to the cumulative probability, whereas their position on the y-axis corresponds to the inverse of the CDF of a normal distribution (since the y-axis represents the standardized variables)

Gumbel distribution

- Y can be represented as maximum of random variables W_i

$$Y = \max(W_1, W_2, \dots, W_n)$$

- CDF of Y : $F_Y(y) = P[y \le Y] = P[(W_1 \le y) \cap \cdots \cap (W_n \le y)]$

$$= F_1(y) \cdot F_{W_2}(y) \cdot \dots \cdot F_{W_n}(y)$$

$$= (F_W(y))^n \ if\ all\ W_i \sim same\ distribution$$

- Assumptions : $F_W(w) = 1 - e^{-g(w)}$

$$\left.\begin{aligned} &F_W(w) = 1 - e^{-g(w)} \\ &u_n \to \infty \text{ so that } F_W(u_n) = 1 - \frac{1}{n} \end{aligned}\right\} \frac{e^{g(u_n)}}{n} = 1$$

Gumbel distribution

- CDF of Y : $F_Y(y) = (F_W(y))^n = \left(1 - e^{-g(y)} \cdot 1\right)^n$

$$= \left(1 - \frac{e^{g(u_n)-g(y)}}{n}\right)^n$$

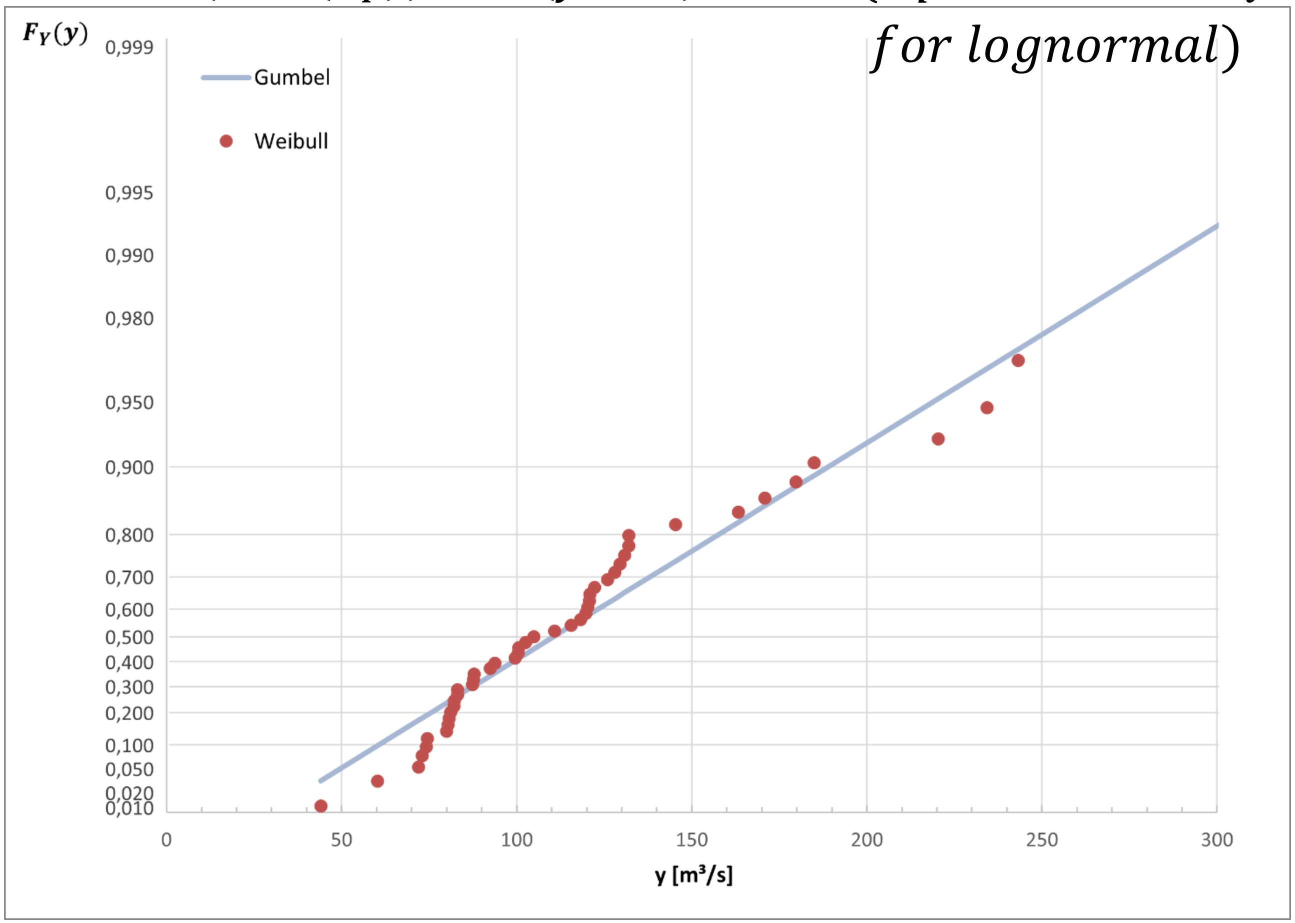
  - Taylor : $g(y) = g(u_n) + \overset{\alpha_n}{\boxed{\left(\frac{dg(y)}{dy}\right)_{u_n}}}(y - u_n) + \ ...$
  - $\lim_{n \to \infty} \left(1 + \frac{\beta}{n}\right)^n = e^\beta$

- CDF of Y : $F_Y(y) = e^{-e^{-\alpha_n(y-u_n)}} = e^{-e^{-\alpha(y-u)}} \qquad -\infty \leq y \leq \infty$
- PDF of Y : $f_Y(y) = \frac{dF_Y}{dy} = \ \alpha e^{-\alpha(y-u)-e^{-\alpha(y-u)}}$

with $m_Y = u + \frac{0{,}577}{\alpha} \qquad \sigma_Y = \frac{\pi}{\sqrt{6}\,\alpha} = \frac{1{,}282}{\alpha}$

Graphical representation

- CDF of Y : $F_Y(y) = e^{-e^{-\alpha(y-u)}}$ $-\infty \leq y \leq \infty$
- $-\ln(-\ln(F_Y)) = \alpha(y-u)$ $(equivalent\ to\ U_i$ $for\ lognormal)$



x-axis : Q (discharge in m³/s) which corresponds to Y_i and are plotted on a linear scale

y-axis : $\alpha(y-u)$ to get a linear relationship BUT labels corresponds to the cumulative probability F_y, whereas their position on the y-axis corresponds to $-\ln(-\ln(F_Y))$

Kolmogorov-Smirnov test

- $P\left[\max_i \left|\frac{i}{n} - F_X(X_i)\right| \leq C_\alpha \middle| \, adequate \ distribution\right] = 1 - \alpha$

C_α / n	α		
	0.10	0.05	0.01
5	0.51	0.56	0.67
10	0.37	0.41	0.49
15	0.30	0.34	0.40
20	0.26	0.29	0.35
25	0.24	0.26	0.32
30	0.22	0.24	0.29
40	0.19	0.21	0.25
> 40	$1.22/\sqrt{n}$	$1.36/\sqrt{n}$	$1.63/\sqrt{n}$

Chi-squared test

- $P\left[\boxed{\sum_{i=1}^{k}\frac{(N_i-np_i)^2}{np_i}}^{\chi^2} \leq \chi^2_{\alpha,k-p-1} \mid adequate\ distribution\right] = 1-\alpha$
  - N_i = # of elements in class i
  - n = total # of elements
  - p_i = probability of being in class i
  - k = # of classes
  - p = # of parameters

Degrees of freedom	α		
	0.10	0.05	0.01
1	2.71	3.84	6.63
2	4.61	5.99	9.21
3	6.25	7.81	11.34
4	7.78	9.49	13.28
5	9.24	11.07	15.09
6	10.64	12.59	16.81
7	12.02	14.07	18.47
8	13.36	15.51	20.09
9	14.68	16.92	21.67
10	15.99	18.31	23.21
12	18.55	21.03	26.22
14	21.06	23.68	29.14
16	23.54	26.30	32.00
18	25.99	28.87	34.80
20	28.41	31.41	37.57
25	34.38	37.65	44.31
30	40.26	43.77	50.89

Return period

- Probability of having a discharge lower than the flood discharge y_r (return period T_r) during a period of T years
  - $P[(Y \leq y_r)_T] = \left(1 - \frac{1}{T_r}\right)^{\frac{T_r}{T_r} \cdot T} = e^{-\frac{T}{T_r}}$ $\boxed{\lim_{T \to \infty} \left(1 + \frac{\beta}{T}\right)^{T} = e^{\beta}}$
- We know $\left(1 - \frac{1}{T_r}\right) = P[Y \leq y_r] = F_Y(y)$ and we are looking for y :
  - $F_Y(y) = F_U(u) \; with \; u = \frac{\ln(y) - m_{\ln Y}}{\sigma_{\ln Y}}$ for lognormal distribution
  - $F_Y(y) = e^{-e^{-\alpha(y-u)}}$ for Gumbel distribution