STOCHASTIC METHODS IN WATER RESOURCES

Unit 2: Hydrological statistics and extremes
Lecture 5: Definitions and analysis of extreme values

Luis Alejandro Morales, Ph.D.

Universidad Nacional de Colombia
Department of Civil and Agriculture Engineering

August 13, 2025

Generalities

- Engineers must deal with different types of risks derived from natural and anthropogenic actions.
- Extremes events, such as floods, droughts and landslides can trigger human, economical and environmental losses.
- This means that civil and environmental engineering design must consider the occurrence of extremes events, even these seldom arise.
- The analysis of extreme events is made based on historical data, usually recorded at daily time scales, of streamflow, rainfall, water levels, temperature, wind, etc. from where probability distribution of extremes are obtained.
- However, because of the length of records, usually the number of extreme values is low and thus the precision of the estimation of the probability distribution decrease.

Hazard, vulnerability and risk

Extreme hydrological event triggers negative effect on economy, environment and population. To understand the these effects is necessary to define the following concepts in the study of extremes.

Hazart (A)

Hazard is the probability that either a natural or anthropogenic event is potentially harmful. Hazards are related to atmospheric, hydrologic, geologic, geothecnical and fires phenomenons that due to their location, severity and frequency are potentially dangerous. For instance, high streamflow is a natural hazard that threat surrounding population. Hazard can be estimated as:

$$A = P[H \geq h]$$

for a period of time t where H is the variable that describe the phenomena, h is the magnitude of H and $P[]$ is the probability of excendance during t.

Vulnerability (V)

Vulnerability is the susceptibility of an exposed element to being affected by a hazard. It depends on the degree of exposure (E) of the element to the hazard and on its resistance or resilience (R) to withstand and absorb the impacts of that hazard.

Hazard, vulnerability and risk

Risk

Risk is defined as the probability of damage of an exposed element upon a hazard. Risk is thus equivalent to the union of a hazard and its vulnerability as shown in the following equation:

$$R = A \otimes V = A \otimes \frac{E}{R}$$

where the $\otimes$ represent a tensor product. This equation shows that to decrease risk, one need to reduce hazard or exposure and increase resilience. There are some hazard that are difficult to reduce such as earthquakes but it is possible to increase the resistance (e.g. anti-earthquake constructions). If the negative effects are represented by the disadvantage scenario S, and if this scenario leads to severe consequences C due to the hazard, then the risk can be estimated as the probability of $S \cap C$, which is a measure of both the probability and the severity of the negative effects. That is:

$$R = P[S \cap C] = P[S]P[C|S]$$

where R is the probability of the negative scenario and its severity, $P[S]$ is the probability of the current scenario S and $P[C|S]$ is the conditional probability of the severity given the occurrence of S. Note that in hydrological systems $P[S] = A$.

Hazard, vulnerability and risk

Risk

Risk can also be estimated as the magnitude of the failure condition as the expected value of the losses (L), the measure of the adverse consequences, as:

$$R = E[L] = \sum_{i=1}^{n} L_i P[L_i]$$

where $i = 1, \cdots, n$ are the loss scenarios, L_i is the loss associated to the ith scenario, and $P[L_i]$ is the probability associated to the occurrence or L_i. For instance, in the case of flood, L_1 can represent minor losses, L_2 moderate losses and L_3 severe losses. Each level of loss is associated to an scenario e.g. the return period of a flood, and scenarios are mutually exclusive and collectively exhaustive. The probability of each level of loss (L_i) is given by:

$$P[L_i] = \sum_{j=1}^{m} P[L_i|S_j]P[S_j]$$

where m is the number of scenarios, and $P[L_i|S_j]$ is the probability of a level of loss L_i given the occurrence of scenario S_j. Risk can thus be estimated as:

$$R = E[L] \sum_{i=1}^{n} \left(\sum_{j=1}^{m} P[L_i|S_j]P[S_j] \right) L_i$$

Note that the scenarios S_j must be defined following the study case (e.g. floods, droughts), and can describe different magnitudes of the same study case.

Return period

Return period

The concept of **return period** is important to the analysis of extreme events. To define it is important remember the concept of independent events where two events A and B are statistically independent when the ocurrence of B does not affect the ocurrence of A and viceversa. This means that $P[A|B] = P[A]$ and $P[B|A] = P[B]$. Note that statistically independent events can be analysed independently of the order of occurrence. To understand this concept, the following example is analysed.

Return period in reservoir

Consider a reservoir designed to control floods, where the main outflow structure is sized based on the allowable downstream flood discharge. Under normal conditions, water flows through the structure and the reservoir does not store any volume. When the incoming streamflow exceeds the structure's design capacity, only a fraction of the flow passes through the structure while the remaining portion is stored, attenuating the flood peak. The reservoir is also equipped with a spillway that evacuates flow when the water level exceeds an admissible limit. For this reason, the design flow for the spillway corresponds to a relatively infrequent event. Suppose that the reservoir's performance is analyzed over a 50-year period, which is the useful life of the system. For the analysis, the highest annual inflows are considered to determine whether they exceed the spillway capacity. This problem can be analysed after supposing the occurrence of Bernoully trials, so that, the data can be described using a Binomial distribution considering that each flow event is independent and with probability p that the maximum annual flow exceed the spillway capacity.

Return period

Return period

Return period in reservoir

Suppose that $p = 0.01$, so that, What is the probability that inflows exceed the spillway capacity in exactly five of the 50 years of the system's useful life? Using the binomial pdf we have:

$$\begin{pmatrix} 50 \\ 5 \end{pmatrix} 0.01^5 (1 - 0.01)^{45} = 0.000135$$

This show that the probability is quite low. However, when the number of year is reduced from five to two, the probability increases to 0.0756. So that, the probability of the system failure during its useful life increases as the number of years of possible occurrence decrease. Also, while the probability of no occurrence of floods during the system's useful like that surpasses the spillway capacity is 0.6050, the complement $1 - 0.6050 = 0.3950$ indicate the probability of hydrological risk of the system.

Return period

Return period

Return period in reservoir

Another interesting question would be to compute the number of years N until the first occurrence of a flood that surpasses the spillway capacity. Considering N as a random variable that follow a geometric distribution. If Q_i represent the maximum flow in the year i and the maximum flows are independent events, the probability that the time interval between exceeds T of a flood of magnitude q being equal to n is:

$$P[T = n] = P[Q_1 < q]P[Q_2 < q]P[Q_3 < q]P[Q_1 < q] \cdots P[Q_{n-1} < q]P[Q_n > q]$$

If Q are equally distributed:

$$P[T = n] = (P[Q < q])^{n-1} P[Q > q] = (1 - p)^{n-1}p$$

which is the geometric pdf. The geometric cdf is thus:

$$P[T > n] = (1 - p)^n p$$

If one want to estimate the probability that $T > 10$, applying this equation, one has $P[T > 10] = 0.9044$, for $P[T > 25] = 0.7778$ and $P[T > 100] = 0.366$. Note that the probability of the spillway capacity being surpassed for the first time decreases progressively over time.

Return period

Return period

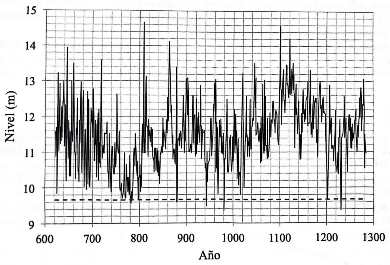
The **return period** is thus defined as the expected value of the time (usually years) between exceedances of a specific streamflow value, that is:

$$E[T] = T_R = \frac{1}{P[Q > q]} = \frac{1}{p}$$

For the reservoir example, this was designated for $T_R = 1/p = 1/0.01 = 100$ years. In the case of minimum flows, The return period is the expected value of the time between magnitudes less than or equal to an specific value.

Return period of river water levels

The time series of the figure represent the minimum annual water levels in the Nile River at the El Cairo station between 622 and 1284. According to the definition of T_R, p is the probability for any year the river water level goes below or equal to a specific value and its inverse value is T_R. Suppose that specific values is 9.66 m. Analysing the time series, we can get that the average time between events where the river goes below 9.66 m is 101.5 years, which is a expected value or the T_R. Note that the probability of occurrence of water level under 9.66 m is thus nearly 0.01.



Return period

Hydrological risk

According to the definition of risk, this involves the hazard of an certain elements. The **hydrological risk** (R_h) is defined as the probability that the exposed element is affected by the hazard at least once during an exposure period of n years, that is:

$$R_h = P[T \leq n] = 1 - (1-p)^n$$

Note that the hazard is defined by the probability $p = \frac{1}{T_R}$ and that R_h is the complement of the probability that the exposed element is not affected by the hazard during the exposure period of n years, this means, it is the complement of $(1-p)^n$. The hydrological risk is also known in the literature as the **failure risk**. The hydrological risk is important in the design of hydraulic structures because these are designed for a design period (T_D) and a service life of n years. Accordingly, R_h can be written as:

$$R_h = 1 - \left(1 - \frac{1}{T_D}\right)^n$$

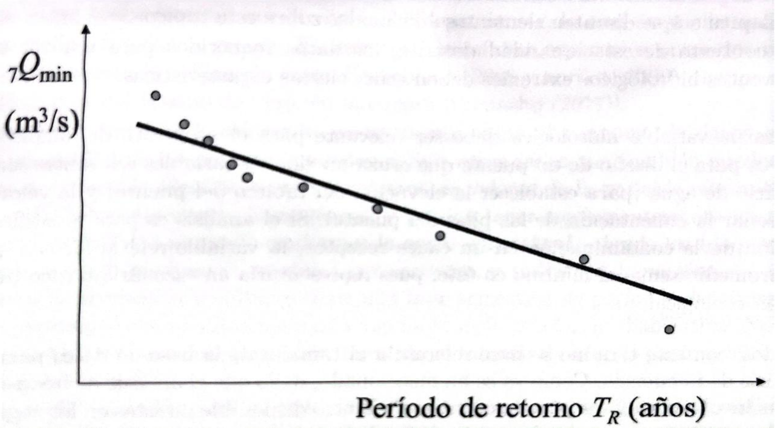
For instance, an structure with a service life of 25 years and an acceptable hydrological risk of 15%, the design period can be estimate from this equation and it is 154 years. It is expected that with a probability of 0.15, this could be surpassed at least once in the 25 years of service life. Note that is bad practice in design to make $T_D = n$ because $R_h = 2/3$, which is quite high.

The purpose of the frequency analysis of extreme hydrological events

The frequency analysis of extreme hydrological events seeks to estimate hazards and in particular to answer the following questions:

- What is the return period of, for instance, a flood occurred in the XXXX year at the YY location where the water lever rise to ZZ m?
- What the water level in the river XX at the YY location for a NN return period? This question can be considered as the inverse of the aforementioned question.

The frequency analysis can be shown graphically through the **frequency curve**. An example of the frequency curve for the weekly average minimum streamflows are shown in the figure.



The data needed for the analysis

The characteristics needed to the analysis of extreme hydrological events are described below:

- Use historical time series of records assuming that the data is stationary, which means that the data pdf does not change in time. Note that this assumption is partially denied because climate change introduce non-stationary traits to the data (e.g. long term trends).
- The hydrological variable must be relevant for the porpoise of the analysis. For instance, if an engineers is designing a bridge over a river, water level and flow speed are data needed for the design.
- The amount of data must be sufficient to the frequency analysis. As in the analysis of extreme events the number of data extracted from the data set is usually low, it is recommended to implement probability distribution with few parameters. These data must be independent and homogeneous (from the same probability distribution). The minimum number of extreme values to perform an frequency analysis must be between 15 and 30. The larger the number the lesser the uncertainty in the estimation of the frequency curve.
- The data must come from the same gage in order avoid a mixing of errors.
- The data must represent or come from a homogeneous hydrological regime. This means that is not correct to mix within the same period, for instance, natural flows with regulated ones.

The data needed for the analysis

Series

Series are usually a time dependent sequence of data. From the series, the extreme data extracted can be classified as: 1) data blocks of maximum/minimum and 2) maximum/minimum data over/under a certain threshold. While the first type is extreme data extracted from a block (subset) of data, the second type is data that is over or under a predefined threshold. The most common first type data set is the annual series. Annual series are formed after extracting the maximum/minimum values of the raw time series for each year, where one year is one block. It is also needed to prove that the values extracted for each year are independent. This means that they must be separated enough in time so that no physical relationship (catchment time concentration) between values is possible. An example of annual series is the maximum annual instantaneous discharge in a river gauge station. Semesterly series can also be formed when independence and homogeneity are guaranteed. In this case, the return period is estimated in semesters. Note that the smaller the block the larger the difficulty to guarantee independence and homogeneity. In the case of the series formed after extracting maximum/minimum data over/under a certain threshold, the number of data per block can be greater than 1. For instance there must be more than one value of an specific year. Because is difficult to establish a return period from these series, this one serves to define a series of excess, that must be constituted by n_y (number of years) values. In this series, the largest n_y values are selected. In the case of bimodal streamflow regimes, the semesterly series make sense, so the chosen of the block must obey the physical functioning of the natural system.

The data needed for the analysis

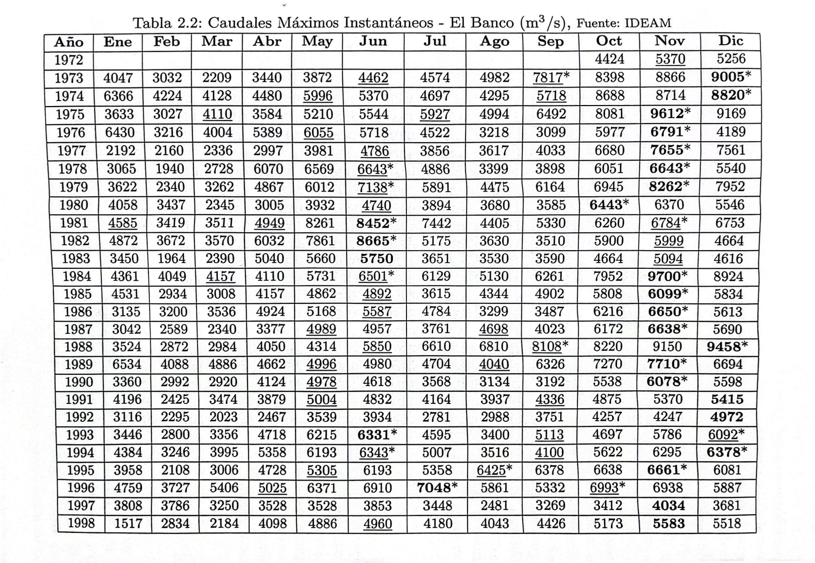


Tabla 2.2: Caudales Máximos Instantáneos - El Banco (m^3/s), Fuente: IDEAM

Año	**Ene**	**Feb**	**Mar**	**Abr**	**May**	**Jun**	**Jul**	**Ago**	**Sep**	**Oct**	**Nov**	**Dic**
1972										4424	5370	5256
1973	4047	3032	2209	3440	3872	4462	4574	4982	7817*	8398	8866	**9005***
1974	6366	4224	4128	4480	5996	5370	4697	4295	5718	8688	8714	**8820***
1975	3633	3027	4110	3584	5210	5544	5927	4994	6492	8081	**9612***	9169
1976	6430	3216	4004	5389	6055	5718	4522	3218	3099	5977	**6791***	4189
1977	2192	2160	2336	2997	3981	4786	3856	3617	4033	6680	**7655***	7561
1978	3065	1940	2728	6070	6569	6643*	4886	3399	3898	6051	**6643***	5540
1979	3622	2340	3262	4867	6012	7138*	5891	4475	6164	6945	**8262***	7952
1980	4058	3437	2345	3005	3932	4740	3894	3680	3585	**6443***	6370	5546
1981	4585	3419	3511	4949	8261	**8452***	7442	4405	5330	6260	6784*	6753
1982	4872	3672	3570	6032	7861	**8665***	5175	3630	3510	5900	5999	4664
1983	3450	1964	2390	5040	5660	**5750**	3651	3530	3590	4664	5094	4616
1984	4361	4049	4157	4110	5731	6501*	6129	5130	6261	7952	**9700***	8924
1985	4531	2934	3008	4157	4862	4892	3615	4344	4902	5808	**6099***	5834
1986	3135	3200	3536	4924	5168	5587	4784	3299	3487	6216	**6650***	5613
1987	3042	2589	2340	3377	4989	4957	3761	4698	4023	6172	**6638***	5690
1988	3524	2872	2984	4050	4314	5850	6610	6810	8108*	8220	9150	**9458***
1989	6534	4088	4886	4662	4996	4980	4704	4040	6326	7270	**7710***	6694
1990	3360	2992	2920	4124	4978	4618	3568	3134	3192	5538	**6078***	5598
1991	4196	2425	3474	3879	5004	4832	4164	3937	4336	4875	5370	**5415**
1992	3116	2295	2023	2467	3539	3934	2781	2988	3751	4257	4247	**4972**
1993	3446	2800	3356	4718	6215	**6331***	4595	3400	5113	4697	5786	6092*
1994	4384	3246	3995	5358	6193	6343*	5007	3516	4100	5622	6295	**6378***
1995	3958	2108	3006	4728	5305	6193	5358	6425*	6378	6638	**6661***	6081
1996	4759	3727	5406	5025	6371	6910	**7048***	5861	5332	6993*	6938	5887
1997	3808	3786	3250	3528	3528	3853	3448	2481	3269	3412	**4034**	3681
1998	1517	2834	2184	4098	4886	4960	4180	4043	4426	5173	**5583**	5518

The data needed for the analysis

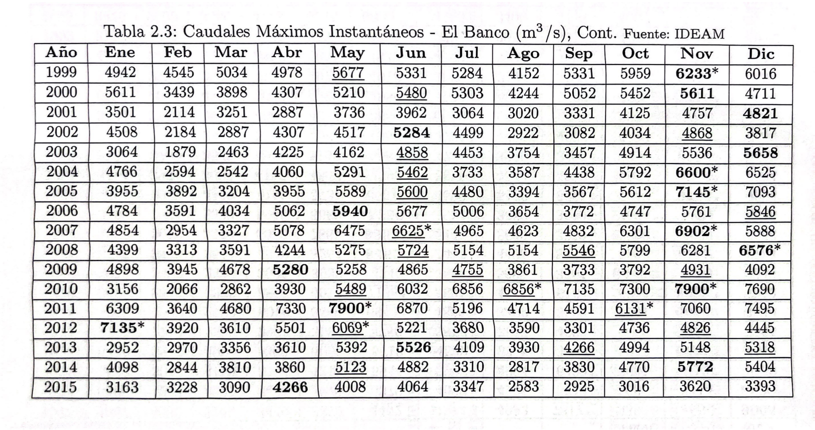


Tabla 2.3: Caudales Máximos Instantáneos - El Banco (m^3/s), Cont. Fuente: IDEAM

Año	Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
1999	4942	4545	5034	4978	5677	5331	5284	4152	5331	5959	**6233***	6016
2000	5611	3439	3898	4307	5210	5480	5303	4244	5052	5452	**5611**	4711
2001	3501	2114	3251	2887	3736	3962	3064	3020	3331	4125	4757	**4821**
2002	4508	2184	2887	4307	4517	**5284**	4499	2922	3082	4034	4868	3817
2003	3064	1879	2463	4225	4162	4858	4453	3754	3457	4914	5536	**5658**
2004	4766	2594	2542	4060	5291	5462	3733	3587	4438	5792	**6600***	6525
2005	3955	3892	3204	3955	5589	5600	4480	3394	3567	5612	**7145***	7093
2006	4784	3591	4034	5062	**5940**	5677	5006	3654	3772	4747	5761	5846
2007	4854	2954	3327	5078	6475	6625*	4965	4623	4832	6301	**6902***	5888
2008	4399	3313	3591	4244	5275	5724	5154	5154	5546	5799	6281	**6576***
2009	4898	3945	4678	**5280**	5258	4865	4755	3861	3733	3792	4931	4092
2010	3156	2066	2862	3930	5489	6032	6856	6856*	7135	7300	**7900***	7690
2011	6309	3640	4680	7330	**7900***	6870	5196	4714	4591	6131*	7060	7495
2012	**7135***	3920	3610	5501	6069*	5221	3680	3590	3301	4736	4826	4445
2013	2952	2970	3356	3610	5392	**5526**	4109	3930	4266	4994	5148	5318
2014	4098	2844	3810	3860	5123	4882	3310	2817	3830	4770	**5772**	5404
2015	3163	3228	3090	**4266**	4008	4064	3347	2583	2925	3016	3620	3393

The data needed for the analysis

Hypothesis test for series

As we have told before data in any type of series must be independent and homogeneous. The Kendall test to verify that the data is stationary or not, check that whether μ is constant (null hypothesis H_0) or not (non null hypothesis H_1). To apply the test is needed to estimate:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} sgn(x_j - x_i)$$

where x represent the hydrological variable, sgn is the sign function where $sgn(a) = 1$ if $a > 0$, $sgn(a) = 0$ if $a = 0$ and $sgn(a) = -1$ if $a < 0$. Once S is calculated, K is calculated as:

$$K = \begin{cases} \frac{S-1}{\sqrt{n(n-1)(2n+5)/18}}, & \text{for } S > 0 \\ 0, & \text{for } S = 0 \\ \frac{S+1}{\sqrt{n(n-1)(2n+5)/18}} & \text{for } S < 0 \end{cases}$$

where K follow a normal distribution with $N(0,1)$. Accordingly, H_0 is accepted if $-z_{\alpha/2} \leq K \leq z_{-\alpha/2}$, where $z_{\alpha/2}$ is the value of K that follow $N(0,1)$ which is exceeded with probability $\alpha/2$, where α is the significance.

Conceptual framework for frequency analysis

Generalities

For a annual series is of interest to estimate the return period (T_R) for a specific value q; for instance the T_R for a extreme discharge occurred in a given year. As T_R is the inverse of the probability p that a value q being equal or exceeded for maximum value series, and being equal or below for minimum value series, this means that p must be estimated for a event of interest (minimum or maximum). There are different equations to estimate p where the results among the equations are similar for short T_R and they differ for long T_R. A general equation to estimate the probability p is:

$$p_i = \frac{i - 0.439}{n + 0.526}$$

where i is the order of data; maximum values are sorted in descending order and minimum values are sorted in ascending order. It is also of interest to estimate the extreme value q for a given T_R. This T_R in many cases (e.g. $T_R = 100$ yrs) is larger than the time series length. For this, the data need to be fit to a probability distribution to extrapolate up to the T_R of interest. To do this, for a series of annual maximum streamflows Q, note that $p = P[Q > q] = 1 - F_Q(q)$, where $F_Q(q)$ is the fitted cdf of Q annual extreme values. For a series of annual minimum streamflow $p = P[Q < q] = F_Q(q)$.

Conceptual framework for frequency analysis

Probability distributions for the frequency analysis

Suppose a set of independent random variables $X_1, X_2, \cdots, X_n$ with a common cdf $F_X(x)$, where x is an observed value and n is the number of values, usually, equally spaced, for instance, one year. Additionally, $X_{max} = max[X_1, X_2, \cdots, X_n]$, and $F_{X_{max}}(x_{max})$ is given by the joint probability that $X_i \leq x_{max}$, that is:

$$\begin{aligned} P[X_{max} \leq x_{max}] &= P[X_1 \leq x_{max}, X_2 \leq x_{max}, \cdots, X_n \leq x_{max}] \\ &= F_{X_1, X_2, \cdots, X_n}(x_{max}, x_{max}, \cdots, x_{max}) \end{aligned}$$

As X_i are independent variables and follow the same probability distribution:

$$F_{X_{max}}(x_{max}) = \prod_{k=1}^{n} P[X_k \leq x_{max}] = \prod_{k=1}^{n} F_{X_k}(x_{max}) = [F_X(x_{max})]^n$$

If $n \rightarrow \infty$ and if X_{max} is standardized and converted into Y, the probability distribution must be one of the following types:

$$\text{Type I:} \quad F_Y(y) = e^{-e^{-y}}, \text{ for } -\infty < y < \infty$$

$$\text{Type II: } F_Y(y) = \begin{cases} e^{-y^{-\gamma}} \text{ for } y > 0 \\ 0 \text{ for } y \leq 0 \end{cases}$$

$$\text{Type III: } F_Y(y) = \begin{cases} e^{-(-y)^{\gamma}} \text{ for } y < 0 \\ 1 \text{ for } y \geq 0 \end{cases}$$

Conceptual framework for frequency analysis

Probability distributions for the frequency analysis

This types can be the following probability distributions:

- Type I: Exponential, Gamma, Weibull, Normal, Lognormal, Logistic and Gumbel type I.
- Type II: Pareto, t-student, Cauchy, Loggamma, Frechet type II
- Type III: Uniform, Beta, Weibull type III.

Note that the type I distributions are not bounded and do not serve to represent maximum and minimum values, while the type II distribution are bounded in the inferior limit and are suitable to fit maximum values. In the case of type III distributions, these are bounded in the superior limit and are suitable for minimum values. Some of these distribution converge slowly to the limit distribution and are thus unsuitable for the analysis of extreme events. In consequence, there are other distributions or modifications to some of the distributions mentioned that can be implemented for the analysis of extremes.

Conceptual framework for frequency analysis

Frequency equation

The **frequency equation** is given by:

$$x_T = \mu_X + K_T \sigma_X$$

where x_T is the magnitude of the event with a return period T, μ_X and σ_X are the mean and the standard deviation of the distribution, respectively, and K_T are the **frequency factor** that depends of the return period and the specific probability distribution. The equations to estimate the frequency factor are found probability textbooks. Note that μ_X and σ_X need to be estimated based in the pdf parameters estimated based on the annual series.

Conceptual framework for frequency analysis

Estimation of the confidence limits

It is important to point out that the magnitude x_T of the extreme event with a return period T is a random variable that follow a pdf. For an ascending sorted annual series $x_{1:n} \leq x_{2:n} \leq \cdots \leq x_{n:n}$, the pdf of $X_{k:n}$ is:

$$f_{X_{k:n}}(x) = \left(\begin{array}{c} n \\ k \end{array} \right) (k) \left[F_X(x) \right]^{k-1} \left[1 - F_X(x) \right]^{n-k} f_X(x)$$

where $f_X(x)$ and $F_X(x)$ are the fitted pdf and cdf, respectively, based on the annual series. Note that the pdf of $X_{k:n}$ has the same characteristics of the pdf fitted to the annual series and a position of order k. Accordingly, the fdp of X_T is:

$$f_{X_T}(x_T) = \left(\begin{array}{c} n \\ m \end{array} \right) (m) \left[F_X(x_T) \right]^{m-1} \left[1 - F_X(x_T) \right]^{n-m} f_X(x_T)$$

where m is the integer fraction of nr and $r = 1 - \frac{1}{T}$. So if $f_{X_T}(x_T)$ is known, the confidence limits can be estimated. For instance, to the 90% of confidence, the inferior and the superior limits must satisfy that $F_{X_T}(x_{T_inf}) = 0.05$ and $F_{X_T}(x_{T_sup}) = 0.95$.

Estimation of the confidence limits

These pdf equations correspond to the Beta distribution. This is actually the incomplete Beta function standardized by the Beta function. The exact confidence intervals can be estimated numerically solving this equations:

$$I[P; n'P_{inf}, n'(1-P_{inf})] = 1-\alpha/2$$

$$I[P; n'P_{sup}, n'(1-P_{sup})] = \alpha/2$$

where: $$I(z; a, b) = \frac{\Gamma(a+b)}{\Gamma(a+1)\Gamma(b)} z^a (1-z)^b + \frac{\Gamma(a+b+1)}{\Gamma(a+1)\Gamma(b)} \int_0^z t^a (1-t)^{b-1} dt$$

and $n' = (n+1)$. Note that int the equation for $I[]$, $z = P$, $a = n'P_{inf}$ or $a = n'P_{sup}$, and $b = n'(1-P_{inf})$ or $b = n'(1-P_{sup})$. The problems is to find the values of a and b using the two equations for a given value of α P. Suppose that one want to estimate the confidence limits for a 90% confidence for the value from the frequency curve that correspond to a return period of $T = 50$ years. The equations above are solved for $\alpha = 1 - 0.9 = 0.1$ and $P = 1 - \frac{1}{T} = 1 - 1/50 = 0.98$. Note that traditionally, the confidence limits for x_T have estimated assuming that X_T follow a Normal distribution, so the confidence limits can be estimated approximately as:

$$x_{T_{inf}} = x_T - z_{\alpha/2} s_T$$

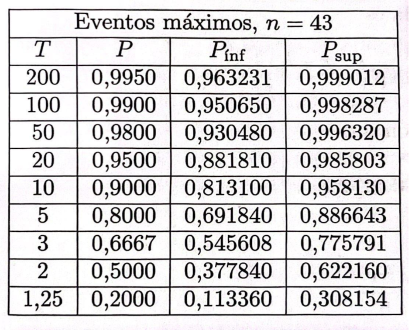
$$x_{T_{sup}} = x_T + z_{\alpha/2} s_T$$

where α is the level of significance, $z_{\alpha/2}$ is the value of the standardized variable which is exceeded with probability $\alpha/2$, and s_T is the estimate of the standard deviation of the event with a return period T. Note that s_T is usually estimated using the Moments or the Maximum-likelihood methods.

Conceptual framework for frequency analysis

Confidence limits of return period

The exact confidence limits for the 90% of confidence of the frequency curve of instantaneous maximum streamflows for the Banco station at the Magdalena river (see the tables with data given before) are shown in the table below:



Eventos máximos, $n = 43$			
T	P	$P_{\text{ínf}}$	P_{sup}
200	0,9950	0,963231	0,999012
100	0,9900	0,950650	0,998287
50	0,9800	0,930480	0,996320
20	0,9500	0,881810	0,985803
10	0,9000	0,813100	0,958130
5	0,8000	0,691840	0,886643
3	0,6667	0,545608	0,775791
2	0,5000	0,377840	0,622160
1,25	0,2000	0,113360	0,308154

Note that the equations given before are solved for: $n = 43$ (# of years), $\alpha = 0.1$, and various values of T. The equations are solved to find the values for P_{inf} and P_{sup} given in the table. The confidence limits can then be estimated using the fitted cdf as $x_{T_{inf}} = F^{-1}(P_{inf})$ and $x_{T_{sup}} = F^{-1}(P_{sup})$.