Transactions on Ecology and the Environment vol 6, © 1994 WIT Press, www.witpress.com, ISSN 1743-3541

The Halphen family of distributions for modeling maximum annual flood series: New developments

L. Perreault, V. Fortin & B. Bobée
INRS-Eau, University of Québec, NSERC - Hydro-Québec Chair on Statistical Hydrology, 2800 rue Einstein, Sainte-Foy, Québec, Canada, G1V 4C7

Abstract

Design of hydraulic structures, optimal operation of reservoirs, and flood forecasting involve estimation of flood flows x_T, with given specified return period T. This estimation is generally achieved by fitting a probability distribution to observed data. In this paper, we briefly present some recent theoretical results of a study on Halphen distributions which were developed in the early forties but have been little used due to the computational complexity. This family of three distributions was motivated by Halphen's desire to construct probability models which met specific requirements of fitting hydrological variables. Halphen probability models have great potential applicability in statistical flood analysis but are not yet familiar to hydrologists.

1 Introduction

Engineering activities such as the design of hydraulic structures, the management of water resources systems, and the prevention of flood damage, all require an estimation of flood characteristics. The objective of flood frequency analysis FFA is to estimate the flood x_T exceeded in average once every T year, which may be used to design, for example, spillways or as prior information in real-time flood forcasting models. At gauged locations, x_T may be estimated by fitting a parametric distribution $F(x;\underline{\theta})$ to observed maximum annual floods, where x refer to an observation and $\underline{\theta}$ to a vector of parameters. If a series of n annual floods has been observed, then, by making use of the i.i.d.-assumption (independently and identically distributed), one attemps to fit a distribution $F(x;\underline{\theta})$ to the n data. It can be shown that $\Pr(X \geq x_T) = 1/T$, so if $F(x;\hat{\underline{\theta}})$ denotes the cumulative distribution with estimated parameters $\hat{\underline{\theta}}$, then the design flood with return period T can be estimated as

$$\hat{x}_T = F^{-1}\left(1 - 1/T\,;\underline{\hat{\theta}}\right) \tag{1}$$

Since it is not possible to derive theoretically the form of the underlying distribution of floods, distributions in FFA are approximate empirical models of the true population of floods and therefore justified primarily by their ability to model flood data. While many distributions have been proposed for use in FFA, very few have been specifically designed to model flood flows. In 1941, Étienne Halphen constructed a family of 3-parameter distributions to model floods. These distributions have appealing statistical properties. However, because of numerical difficulties encountered when estimating the parameters, they have long been neglected. Now that computational power is available, the Halphen distributions have been revisited theoretically and C++ computer routines for estimation of the parameters $\underline{\theta}$, the design flood x_T, and its asymptotic standard error have been developed.

2 Halphen distributions and historical remarks

Despite the number of distributions already in use in hydrology in the middle of the century, the French hydrologist E. Halphen felt the need to develop new density functions (Halphen[4]). Past experience had convinced him that none of the traditional distributions used at that time had shapes that were globally appropriate for modeling flood series in France. Consequently, Halphen developed new distributions, and, although they are largely based on empirical justifications, they possess appealing theoretical properties as it will be shown.

Halphen distributions were motivated by the desire to meet some specific requirements pertaining to the fitting of hydrological variables. These conditions concerned the tail behavior of the probability density functions (pdf), and the statistical properties of the estimators:

- the distribution family should be able to model both exponentially and algebraically decreasing data densities
- the densities should have a lower bound at zero
- the estimation of the parameters should be done using sufficient statistics.

Halphen first considered the 2-parameter harmonic distribution with both lower and upper tails decreasing exponentially (Morlat[6]). However, the application of the harmonic distribution to many types of hydrological data revealed lack-of-fit problems. Halphen generalized the harmonic distribution by adding a third parameter, thereby introducing his Type A probability distribution. Its pdf is given by

$$f_A(x) = \frac{1}{2m^{\upsilon}K_{\upsilon}(2\alpha)}\, x^{\upsilon-1} \exp\left[-\alpha\left(\frac{x}{m} + \frac{m}{x}\right)\right] \quad , \qquad x > 0 \tag{2}$$

where υ and $\alpha > 0$ are shape parameters, and $m > 0$ a scale parameter. The constant $K_\upsilon(2\alpha)$ is a modified Bessel function of the second kind (Watson[8]), defined as

$$K_\upsilon(2\alpha) = \frac{1}{2m^\upsilon} \int_0^{+\infty} x^{\upsilon-1} \exp\left[-\alpha\left(\frac{x}{m} + \frac{m}{x}\right)\right] dx \qquad (3)$$

$f_A(x)$ is unimodal with positive skewness, and its decrease at the lower and upper tails is exponential. It can be shown that the gamma (G) and the inverted gamma (IG) ($1/X \approx$ G) distributions constitute limiting forms of the Type A distribution for specific values of the parameters: if $\alpha \to 0$, $X \approx G$ for $\upsilon > 0$, and $X \approx IG$ for $\upsilon < 0$. Between these limiting cases, the Type A distribution can take a variety of forms which are all of interest for hydrological modeling.

Halphen investigated the goodness-of-fit of the Type A distribution for various hydrological series. He noted that the distribution was appropriate in many cases, but realized that a new distribution with different asymptotic behavior of the lower tail was needed. As a result of his extensive data study, he introduced the Type B distribution with pdf given by

$$f_B(x) = \frac{2}{m^{2\upsilon} ef_\upsilon(\alpha)} x^{2\upsilon-1} \exp\left[-\left(\frac{x}{m}\right)^2 + \alpha\left(\frac{x}{m}\right)\right] , \qquad x > 0 \qquad (4)$$

where $\upsilon > 0$ and α are shape parameters, and $m > 0$ a scale parameter. The constant $ef_\upsilon(\alpha)$, described in detail by Halphen[5], was named *exponential factorial function,* and is mathematically defined as

$$ef_\upsilon(\alpha) = \int_0^{+\infty} x^{2\upsilon-1} e^{(-x^2+\alpha x)} dx \qquad (5)$$

It can be shown that $ef_\upsilon(\alpha)$ is related to the confluent hypergeometric function (Perreault & Bobée[7]) and for integer values of α to the Hermite polynomials (Morlat[6]). The Halphen Type B distribution can take a large variety of forms. Depending on the values of the parameters, $f_B(x)$ can be unimodal with positive or negative skewness and with various kinds of algebraic decrease near the origin, J-shaped and S-shaped. The Type B distribution also has the gamma distribution as limiting form, obtained when $\alpha \to -\infty$.

The family of distributions formed by Type A and B covered virtually all of Halphen's needs for modeling hydrological data series. Nevertheless, a third distribution type, denoted Type B^{-1}, was introduced by Larcher, one of Morlat's co-workers (Morlat[6]) to complete the Halphen system. Its pdf is given by

$$f_{B^{-1}}(x) = \frac{2}{m^{-2\upsilon} ef_\upsilon(\alpha)} x^{-2\upsilon-1} \exp\left[-\left(\frac{m}{x}\right)^2 + \alpha\left(\frac{m}{x}\right)\right] , \qquad x > 0 \qquad (6)$$

where $\upsilon > 0$ and α are shape parameters, and $m > 0$ a scale parameter. The Type B^{-1} can be deduced from the Type B distribution by substituting x/m by m/x in eqn (4). Therefore, if X is Type B distributed, then $1/X$ is Type B^{-1}. We

Transactions on Ecology and the Environment vol 6, © 1994 WIT Press, www.witpress.com, ISSN 1743-3541

deduce from this result that the inverted gamma distribution is a limiting form of Type B^{-1}, obtained as $\alpha \to -\infty$.

As an illustration of the variety of forms of Halphen distributions, Figure 1 displays the Type B density for some values of the shape parameters α and υ. The scale parameter m is chosen in such a way that the variance is unity.

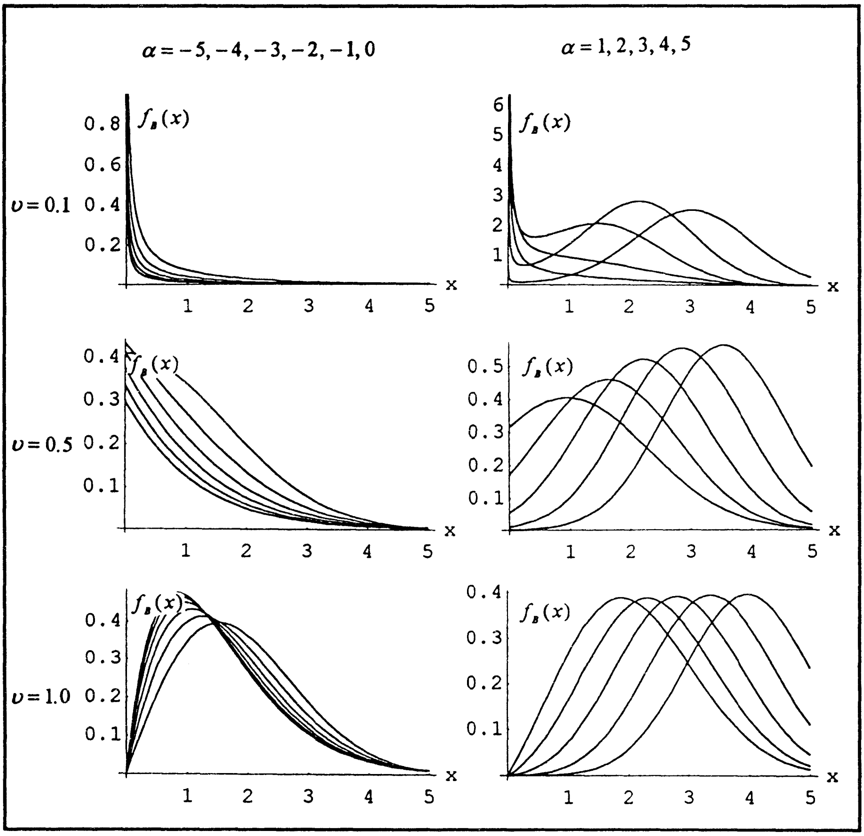

Figure 1: Plots of probability density function of Halphen's Type B.

As it is the case with the Pearson system, we can say that the three Halphen distributions form a complete system. Morlat[6] showed that the scale parameter m can be eliminated by considering the moment ratios $\delta_1 = \ln(\mu_1/\mu_0)$ and $\delta_2 = \ln(\mu_0/\mu_{-1})$ where $(\mu_{-1})^{-1}$, μ_0 and μ_1 are respectively the harmonic, geometric and arithmetic means of the population. They correspond to the moments of order -1, quasi-zero and 1, respectively, (Bobée and Ashkar[2]). As the moment ratios β_1 and β_2 (functions of the coefficients of skewness and of kurtosis) of the traditional diagram introduced by Karl Pearson provide a

Transactions on Ecology and the Environment vol 6, © 1994 WIT Press, www.witpress.com, ISSN 1743-3541

complete taxonomy of the Pearson system, so do δ_1 and δ_2 for the Halphen system. The two diagrams were studied and compared by Bobée *et al.*[3]. Figure 2 illustrates how the three types of Halphen distributions harmoniously form a complete system. In particular, the moment ratio diagram clearly shows the symmetry between Type A distributions with positive and negative values of the shape parameter υ, and between the Type B and the Type B^{-1}. It also shows the limiting cases G and IG.

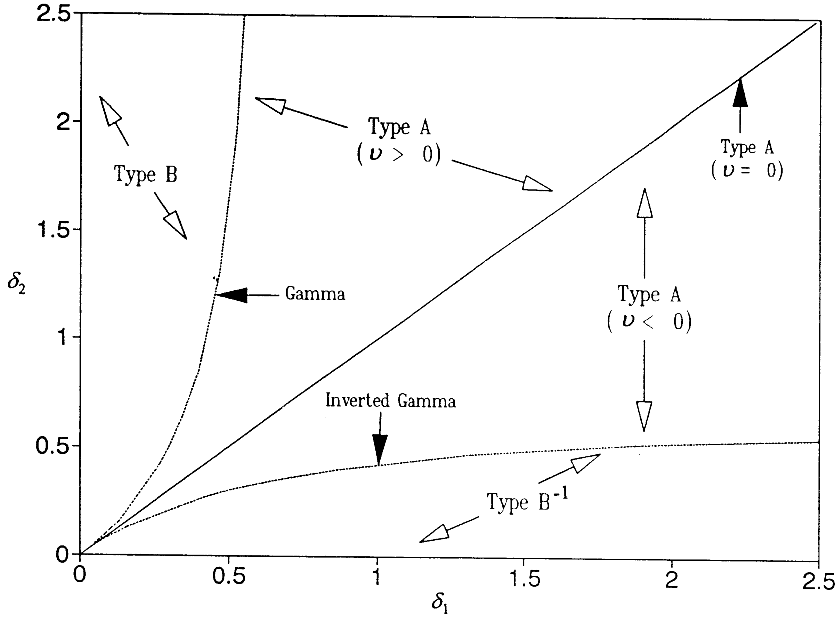

Figure 2: Halphen's distributions in the (δ_1, δ_2) moment ratio diagram.

3 Some properties of Halphen distributions

Morlat[6] presented the noncentral moments of the three Halphen distributions. The moments are functions of the Bessel function $K_\upsilon(2\alpha)$ for Type A, and of the exponential factorial function $ef_\upsilon(\alpha)$ for Type B and B^{-1}. He also gave some recommendations on parameter estimation, but in a rather intuitive way. As mentioned above, the Halphen probability density functions were partly motivated by the desire to construct distributions which allow efficient parameter estimation, i.e. estimates which are functions of sufficient statistics. A simple rearrangement of the pdf.'s (eqns (2), (4) and (6)) shows that the Halphen distributions belong to the exponential class of probability density functions of order 3. We therefore deduce that a triplet of minimal sufficient and complete statistics exist for each distribution. These are given in Table 1. The statistics are seen to be sample moments of various order.

Transactions on Ecology and the Environment vol 6, © 1994 WIT Press, www.witpress.com, ISSN 1743-3541

The existence of sufficient and complete statistics implies that the ML-estimators of the parameters exist and are unique. Moreover, from Lehman-Sheffe's theorem (Bickel & Doksum[1]) we can deduce that unbiased minimum variance estimators of the parameters are functions of the ML-estimators. Hence, no investigation to determine the optimal estimation method is necessary, as it is for distributions like the Pearson Type 3, the log-Pearson Type 3, the 3-parameter log-normal, etc. Among the various 3-parameter distributions used in hydrology, the Halphen distributions are the only ones which possess this property.

Table 1. Triplets of minimal sufficient statistics for each distributions.

Type	Sufficient statistics		
A	$A = \frac{1}{n}\sum_{i=1}^{n} x_i$	$H = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i}\right)^{-1}$	$G = \exp\left(\frac{1}{n}\sum_{i=1}^{n}\log x_i\right)$
B	$Q = \frac{1}{n}\sum_{i=1}^{n} x_i^2$	$A = \frac{1}{n}\sum_{i=1}^{n} x_i$	$G = \exp\left(\frac{1}{n}\sum_{i=1}^{n}\log x_i\right)$
B^{-1}	$IQ = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i^2}$	$H = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i}\right)^{-1}$	$G = \exp\left(\frac{1}{n}\sum_{i=1}^{n}\log x_i\right)$

4 Estimation of the parameters

The reason why only little attention has been paid to Halphen distribution since their introduction fifty years ago has to do with the practical problems pertaining to the estimation of the parameters. The systems of ML-equations are non-linear and must be solved using iterative procedures. They involve evaluation of Bessel functions, exponential factorial functions, and of their derivatives, which are all difficult to handle. The numerical solutions turn out to be rather tedious. To illustrate the problems, we discuss briefly the estimation of parameters of the Halphen Type A distribution. Detailed calculations are presented in Perreault & Bobée[7].

To find the ML-estimators we considered a two step approach. Initially, the parameters α and m are estimated numerically with υ fixed. Then, the partially maximized log-likelihood function $\log L(\upsilon; \hat{\alpha}, \hat{m})$ is used to determine an estimate of the parameter υ.

The shape parameter υ being fixed, the ML-equations can be rearranged to get the alternative system of equations :

$$D_\upsilon(\alpha) = \frac{K_{\upsilon+1}(2\alpha)\,K_{\upsilon-1}(2\alpha)}{K_\upsilon^2(2\alpha)} = \frac{A}{H} \tag{7}$$

$$m = A\frac{K_\upsilon(2\alpha)}{K_{\upsilon+1}(2\alpha)} \tag{8}$$

Thus, if the likelihood equations have a solution, it may be found by solving eqn (7) for α and then inserting the result $\hat{\alpha}$ in eqn (8) to get $\hat{m}$. It follows that the properties of the function $D_\upsilon(\alpha)$ are central in the discussion of the estimation. $D_\upsilon(\alpha)$ is a positive and strictly decreasing function of α for any given υ. By using the first order approximation of $K_\upsilon(2\alpha)$ as $\alpha \to 0$, we can show that:

$$\mathrm{Max}\{D_\upsilon(\alpha)\} = \lim_{\alpha\to 0} D_\upsilon(\alpha) = \begin{cases} +\infty \ , \text{ if } |\upsilon| \le 1 \\ \dfrac{|\upsilon|}{|\upsilon|-1}, \text{ if } |\upsilon| > 1 \end{cases} \tag{9}$$

Eqn (7) can therefore only be solved if $|\upsilon| \le 1$ or if $AH^{-1} < |\upsilon|/(|\upsilon|-1)$. If we define $U = AH^{-1}/(AH^{-1}-1)$, we have that $U > 1$ because $H \le A$, and hence, from the properties of $D_\upsilon(\alpha)$, we see that the likelihood equations (eqns (7) and (8)) have a solution if and only if $|\upsilon| > U$. In the opposite case ($|\upsilon| \le U$), the estimates correspond to the ML-estimators of the gamma ($\upsilon > 1$) or the inverted gamma ($\upsilon < -1$) distribution. Therefore, the estimators point at one of these two limiting cases (depending on the sign of υ) as being the most likely distribution.

Now that we can estimate the parameters m and α for any given υ, we have to determine $\hat{\upsilon}$. The principal tool for the estimation of υ is the partially maximized log-likelihood function $\log L(\upsilon; \hat{\alpha}, \hat{m})$ obtained by inserting the estimates $\hat{\alpha}$ and $\hat{m}$ in the log-likelihood function. It can be shown to be:

$$\log L(\upsilon;\hat{\alpha},\hat{m}) \propto \begin{cases} \upsilon \ln[\upsilon/M] - \Gamma(\upsilon) + (\upsilon-1)\ln G - \upsilon, & \text{if } \upsilon \ge U \\ \ln\left[\dfrac{G^{\upsilon-1}}{\hat{m}^\upsilon K_\upsilon(2\hat{\alpha})}\right] - \hat{\alpha}\left[\dfrac{K_{\upsilon+1}(2\hat{\alpha})}{K_\upsilon(2\hat{\alpha})} + \dfrac{K_{\upsilon-1}(2\hat{\alpha})}{K_\upsilon(2\hat{\alpha})}\right], & \text{if } |\upsilon| < U \\ -\upsilon \ln[-\upsilon H] - \Gamma(-\upsilon) + (\upsilon-1)\ln G + \upsilon, & \text{if } \upsilon \le -U \end{cases} \tag{10}$$

In practice, having implemented the estimation of α and m for fixed values of υ (step 1) on the computer, it is straight forward to determine $\hat{\upsilon}$ numerically (step 2), e.g. by a simple tabulation of $\log L(\upsilon;\hat{\alpha},\hat{m})$. Let us mention that the derivatives of $\log L(\upsilon;\hat{\alpha},\hat{m})$ at U and $-U$ can be easily computed and are independent of the unknown parameters. These derivatives are interesting because they can be used before step 1 to determine whether $\hat{\upsilon}$ is inside or outside the interval [$-U$, U] and hence whether the likelihood equations have a solution.

Finally, it is obvious from eqn (2) that the cumulative distribution function for the Halphen Type A does not exist in closed form. Therefore, it is not possible to develop an explicit expression for quantiles, or T-year events. An estimate $\hat{x}_T$ of x_T must be determined by solving numerically the equation:

$$\int_0^{\hat{x}_T} f_A(x;\hat{m},\hat{\alpha},\hat{\upsilon})dx = 1-\frac{1}{T} \tag{11}$$

The approaches used to estimate the parameters and x_T events for the Type B and Type B^{-1} distributions are quite similar and shall not be discussed. The reader is referred to Perreault & Bobée[7].

5 Conclusions

In this paper, we have presented a synthesis of the theoretical aspects of the Halphen family of distributions which are not yet familiar to hydrologists, but have great potential applicability in FFA. Those distributions are complex since they require numerical methods involving in particular the evaluation of the Bessel and exponential factorial functions.

To help engineers perform FFA, we have developed the software ***Ajuste II*** which, in addition to other more classical distributions, include C++ computer routines for estimation of parameters, x_T, and its asymptotic standard error for the Halphen distributions. These distribution have been successfully applied to 179 series of flood data in Eastern Canada (Québec-Ontario area). Simulation studies are, however, still needed to show the advantage of their systematic application over other distributions in current use.

References

1. Bickel, P.J. & Doksum, K.A. *Mathematical Statistics: Basic Ideas and Selected Topics*, Holden-Day, Oakland, 1977.
2. Bobée, B. & Ashkar, F. The Gamma Family and Derived Distributions Applied in Hydrology, Water Resources Publications, Littleton, 1991.
3. Bobée, B., Ashkar, F & Perreault, L. Two kinds of moment ratio diagrams and their applications in hydrology, Stochastic Hydrology and Hydraulics, 1993, **7**, 41-65.
4. Halphen, E. Sur un nouveau type de courbe de fréquence, *Compte rendus de l'Académie des Sciences*, 1941, **213**, 634-635. (published under the name of D. Dugué)
5. Halphen, E. Les fonctions factorielles, *Institut de statistique de l'Université de Paris*, 1955, **4**, 21-39.
6. Morlat, G. Les lois de probabilité de Halphen, *Revue de statistique appliquée*, 1956, **4**, 21-46.
7. Perreault, L. & Bobée, B. Les lois de Halphen, *Research Report*, INRS-Eau, 1994, in preparation.
8. Watson, G.N. *A Treatise on the Theory of Bessel Functions*, Cambridge University Press, New York, 1966.