Distribution Estimation for Fishing Time

Jun Yu and Kjell Leonardsson

Research Report
Centre of Biostochastics

Swedish University of Agricultural Sciences

Report 2011:02
ISSN 1651-8543

Distribution Estimation for Fishing Time

Jun Yu[1]

Centre of Biostochastics
Swedish University of Agricultural Sciences, Umeå, Sweden

Kjell Leonardsson

Department of Wildlife, Fish, and Environmental Studies
Swedish University of Agricultural Sciences, Umeå, Sweden

Abstract

Information on visitation frequencies in recreational fishing is important when dealing with fishing tourism in order to make prognoses for the future upon changes in the management of the aquatic environment (flow regime or habitat restoration), fishing regulations, or to estimate the total harvest of fish. Therefore interviews have been performed in a number of streams and sections of streams, throughout the last twenty years in the Jämtland-Härjedalen region in Sweden. In this work the probability distribution of total fishing hours a day (possibly on different periods) is considered. We found that both Gamma and Weibull distributions can be considered as approximate distributions that generate the data. Gamma distribution fits very well for summer season while Weibull distribution is more appropriate for the other periods. In general, the gamma model is easier to interpret and better fits the mode of the distribution, and therefore, is preferred. Having parameters estimated, we are able to calculate probabilities of different fishing times. It is also suggested to use two periods: mid-summer – the end of August and other dates. The modelling at section level seems to be successful. Both the Gamma and Weibull distributions fits well the data for all periods, providing that the number of observations are not less than 20. The mean fishing hours, however, varies from section to section, even within the same watercourse.

Keywords: Fishing time, density estimation, goodness-of-fit test, randomisation.

[1]E-mail address to the correspondence author: jun.yu@slu.se

1 Introduction

Information on visitation frequencies in recreational fishing is important when dealing with fishing tourism in order to make prognoses for the future upon changes in the management of the aquatic environment (flow regime or habitat restoration), fishing regulations, or to estimate the total harvest of fish. The number of fishing licenses sold does generally not have the resolution needed for analyses of management regulation effects because there might be alternative stream sections with different fishing rules available for the same licence. Therefore interviews have been performed in a number of streams and sections of streams, throughout the last twenty years in the Jämtland-Härjedalen region in Sweden. The interviews have generally been performed once, twice or three times a week throughout the fishing (summer) season. The two most important questions asked during the field interviews with relevance for this report are one question that gives the time spent fishing so far "today" and the other about the expected total time spent fishing "today" (start-stop). Each interview event covers the entire stream sections, which means that the anglers being active will be asked these questions. There are also additional questions, but these are left out here since the focus in this report is on the expected total time per angler. When aggregating these data to seasonal measures, it has not been clear what is the best method to use, and how to best estimate the uncertainty in the measures.

2 Weather independent interview events

The statistics on the number of anglers and their fishing time was gathered at one, two or three days per week throughout the fishing seasons. One concern in the analysis has been whether the interviews are representative in terms of weather. To investigate if the interviewers preferred non-rainy days, data on precipitation and interview events were compared. If there was preference for non-rainy days, the results from the interviews might be biased, such that also fewer anglers might be active during the rainy days. Lack of observations from such days would therefore lead to overestimation of the total number of anglers and their fishing time per season. For weather independent interviews to be the case, the cumulative density functions of precipitation and interviews in relation to the amount of precipitation per day should coincide. The check was performed as follows. The empirical cumulative density functions (ecdf) were compiled for number of days in relation to the precipitation those days. Similarly an ecdf was compiled for the number of interview days in relation

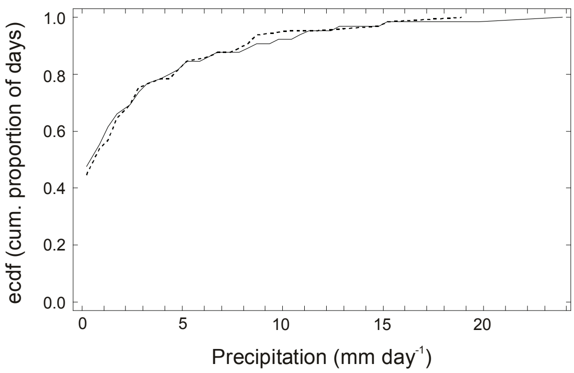

Figure 2.1: Nedre Ammerån, 1995-1999 (excluding 1998)

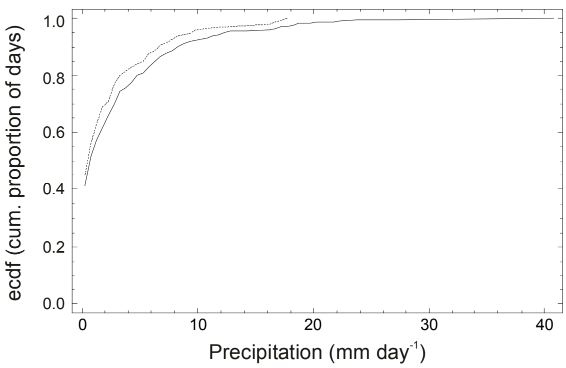

Figure 2.2: Råndan, 1997-2003 (excluding 2001)

to the precipitation those days. This comparison was made for two streams: Nedre Ammerån (Figure **??**) and Råndan (Figure **??**), including all sections for each stream and excluding years with few interviews and when only a subset of the sections were visited by the interviewers. For Nedre Ammerån the two ecdf's were almost identical indicating that the interview events were applied independent on the weather (amount of rain per day). In contrast, there was a tendency for interviews during less rainy days in Råndan than would be expected from the frequency of rainy days.

Table 3.1: Distribution of observations over month and year

Month	June			July			August			September			Total
Obs.	197			387			264			88			936
Year	94	95	96	97	98	99	00	02	03	04	05	06	07	Total
Obs.	41	68	43	147	62	137	65	53	77	65	80	51	47	936

Table 3.2: Distribution of observations over period

Period	Definition	Obs.
1	Midsummer Day – July 15	278
2	July 16 – August 31	450
3	Weekends outside of Periods 1 & 2	112
4	Weekdays outside of Periods 1 & 2	96
Total		936

3 Data description

The whole data is of size 2611, from year 1994 to 2007 (without observation in 2001); of which 2595 are in months June to September. Of the 2595, there are 936 observations on unique days and sections of watercourses. The total fishing hours, the greater number of "Spent Fishing Time" and "Estimated Fishing Time" for that day, are summed up for each day by different sections. Those data are used as the basis for estimation. In addition, data of the number of anglers every day are also available. The distribution of data over month and year is presented in Table **??**.

Empirical experience suggests that the data may be divided into four periods, each of which represents the seasonality of fishing activities. The first two periods contain the majority of fishing time, due to the summer vacation, where Period 2 shall indicate the peak season. While Period 3 and 4 represent the slack season (first two third part of June and whole September), there should still be some difference between weekends and weekdays. We will show that these four periods have different distribution patterns. The distribution over period is shown in Table **??**.

The primary goal of this work is to estimate the distribution of total fishing hours for one day by sections (of different watercourses in Jämtland, Sweden), years, and/or periods. Since the data length is very limited if data are divided to such detail categories (see Table **??**), it is not appropriate to estimate the densities. Therefore, we will only consider the distribution of total fishing hours per day (possibly on different periods or years). Note that since the average number of anglers per day can be estimated, so can the fishing hours per angler per day.

4 Density estimation

To investigate the probability distributions of total fishing hours, we use parametric statistical inference method, which implies in this case to use Kolmogorov-Smirnov goodness-of-fit test (KS-test) for testing how good the data fits a certain probability distribution. The corresponding parameters in the distribution will be estimated by using the maximum likelihood (ML) method. The nonparametric density estimation will also be applied for illustrating the distributional curve (more smooth than the histogram) and for judgement in comparison of different parametric estimates.

Based on the characteristics of the random variable, "Total fishing hours", such as continuity, nonnegativity and type of lifetime, we choose the following four probability distributions to be fitted for our data:

4.1 The exponential distribution

When a random variable X measures the duration of time until the occurrence of a given phenomenon, it can be described by the *exponential distribution*, whose density is

$$f(x;\beta) = \frac{1}{\beta} e^{-x/\beta}, \qquad \text{if } x \geq 0,$$

where $\beta > 0$ is the scale parameter. This scale parameter is a survival parameter in the sense that if a random variable X is the duration of time that a given biological or mechanical system manages to survive. The expected duration of survival of the system is β units of time, and the sample mean $\bar{X}$ is the ML estimator of the scale parameter.

Moreover, the exponential distribution is the only continuous distribution with a "lack of memory" property. That is, if the lifetime of a part follows the exponential distribution, then the distribution of the time until failure is the same as the distribution of the time until failure given that the part has survived to time t:

$$Pr(X > s + t | X > s) = Pr(X > t), \qquad \text{for } s > 0, t > 0.$$

4.2 The gamma distribution

If $X_1, X_2, ..., X_\alpha$ are α independent exponentially distributed random variables, each of which has a mean of β, then the sum $Y = \sum_{i=1}^{\alpha} X_i$ has the *gamma distribution* with density function

$$f(y;\alpha,\beta) = \frac{y^{\alpha-1}}{\beta^\alpha \Gamma(\alpha)} e^{-y/\beta}, \qquad \text{if } y \geq 0,$$

where $\alpha > 0$ is the shape parameter, $\beta > 0$ is the scale parameter, and Γ is the gamma function. Note that α need not always be an integer. The mean of Y is $\mu = \alpha\beta$. Obviously, the exponential distribution is a special case of the gamma distribution when $\alpha = 1$. A gamma distribution starts to resemble a normal distribution as the shape parameter α tends to infinity. The ML estimators of α and β are the solutions of the following simultaneous equations:

$$\begin{aligned} \log(\hat{\alpha}) - \psi(\hat{\alpha}) &= \log\left[\frac{\bar{Y}}{(\prod_{i=1}^{n} Y_i)^{1/n}}\right], \\ \hat{\beta} &= \frac{\bar{Y}}{\hat{\alpha}}, \end{aligned}$$

where ψ is the digamma function: $\psi(\alpha) = \Gamma'(\alpha)/\Gamma(\alpha)$.

The gamma distribution is frequently a probability model for waiting times; for instance, in life testing, the waiting time until death is a random variable that is frequently modeled with a gamma distribution. Applications of the gamma include life testing, statistical ecology, queuing theory, inventory control and precipitation processes.

4.3 The Weibull distribution

Important examples of nonnegative random variables occurring in applications are lifetimes, waiting times, learning times, durations of epidemics, and travelling times. Nontemporal examples of nonnegative random variables include material strengths, particle dimensions, radioactive intensities, rainfall amounts, and costs of industrial accidents. Although exponential or gamma distributions provide reasonable fits to the frequency distributions of some of these random variables, in some cases the fit is not as close as is desired, and in other cases the fit is unsatisfactory. Hence, other classes of distributions have been introduced to explain the variability of some of these phenomena. One such family of distributions is *Weibull distributions.* The experience of many investigators has shown that the Weibull distributions provide good probability models for describing "length of life" and other endurance data.

The Weibull distribution has a density function as follows:

$$f(x;\alpha,\beta) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{(\alpha-1)} e^{-(x/\beta)^{\alpha}}, \qquad \text{if } x \geq 0,$$

where $\alpha > 0$ is the shape parameter and $\beta > 0$ is the scale parameter. When $\alpha = 1$, the Weibull distribution becomes the exponential distribution with the same scale parameter.

The ML estimators of α and β are the solutions of the following simultaneous equations:

$$\begin{aligned}
\hat{\alpha} &= \frac{n}{\left\{(1/\hat{\beta})^{\hat{\alpha}} \sum_{i=1}^{n} \left[X_i^{\hat{\alpha}} \log(X_i)\right]\right\} - \sum_{i=1}^{n} \log(X_i)} \\
\hat{\beta} &= \left[\frac{1}{n}\sum_{i=1}^{n} X_i^{\hat{\alpha}}\right]^{1/\hat{\alpha}}
\end{aligned}$$

The Weibull distribution is often used in the field of life data analysis due to its flexibility – it can mimic the behaviour of other statistical distributions such as the normal and the exponential. If the failure rate decreases over time, then $\alpha < 1$. If the failure rate is constant over time, then $\alpha = 1$. If the failure rate increases over time, then $\alpha > 1$.

An understanding of the failure rate may provide insight as to what is causing the failures:

- A decreasing failure rate would suggest ”infant mortality”. That is, defective items fail early and the failure rate decreases over time as they fall out of the population.
- A constant failure rate suggests that items are failing from random events.
- An increasing failure rate suggests ”wear out” - parts are more likely to fail as time goes on.

4.4 The lognormal distribution

A lognormal distribution is a probability distribution of a random variable whose logarithm is normally distributed. If Y is a random variable with a normal distribution, then $X = \exp(Y)$ has a lognormal distribution; likewise, if X is lognormally distributed, then $\log(X)$ is normally distributed.

Weibull distribution, with density

$$f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left\{-\frac{1}{2\sigma^2}[\log(x) - \mu]^2\right\}, \qquad \text{if } x \geq 0,$$

where μ and $\sigma > 0$ are parameters of the distribution. They are mean and standard deviation of the distribution of $\log(X)$. Thus the ML estimators of μ and σ^2 will be sample mean and sample variance, respectively, applied to log-transformed data.

A variable might be modelled as log-normal if it can be thought of as the multiplicative product of many independent random variables each of which is positive. Because the empirical distribution of many variables are inherently positive and skewed to the right (e.g., size of organisms, amount of rainfall, size of income, etc.), the lognormal distribution has been widely applied in several fields, including economics, business, industry, biology, ecology, atmospheric science, and geology

The lognormal distributions are important competitors to the exponential, gamma, or Weibull distributions as models for nonnegative phenomena.

It is worth noting that all these four distributions belong to a more general distribution class, the generalised Gamma distribution, where three parameters determine determine the distribution family.

5 Goodness-of-fit test

5.1 The whole data set

Assume that all observations are independent and from same distribution (population). From Figure **??**, the candidate densities are fitted and compared with the histogram and nonparametric density estimate using Gaussian kernel of all the data, where the parameters in Exponential, Weibull, Gamma and Lognormal distributions are estimated by the maximum likelihood method. The estimated parameters and P-values using Kolmogorov-Smirnov goodness-of-fit test (KS-test) are presented in the legend of Figure **??**. Note that there are only 129 unique values among the 936 observations; the KS-test cannot give an accurate P-value for data with ties. Thus, the P-values reported are used only as a relative reference.

5.2 Data by periods

5.2.1 Four periods

To begin with looking at the data from each period separately, the summary statistics for these four periods is presented in Table **??**. It indicates similarities between Period 1 and 2 as well as between Period 3 and 4. Due to some extremely large observations, the mean and standard deviation are bigger than the median and IQR.

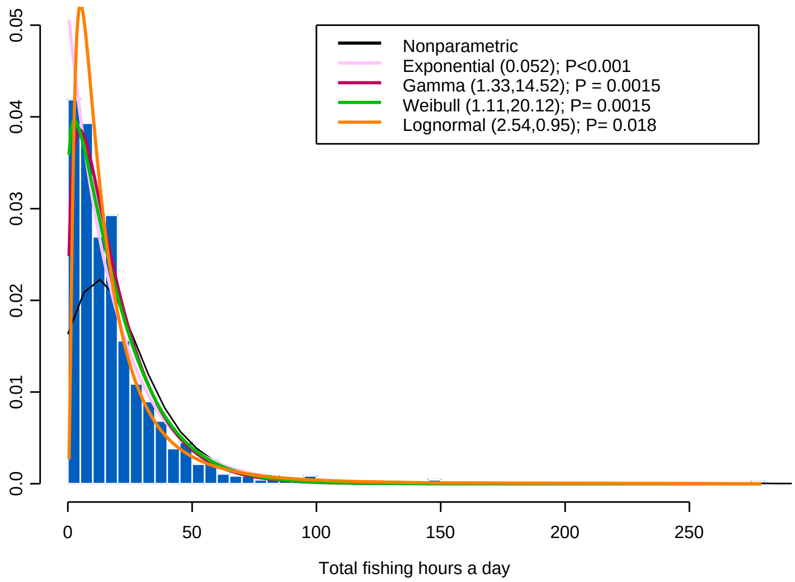

Figure 5.1: Histogram and density estimates for whole data

Table 5.1: Summary Statistics for fishing time in four different periods

Period	1	2	3	4
Mean	22.1	20.0	14.3	13.4
Std	21.1	20.9	17.9	16.2
Median	15.0	16.0	8.0	6.8
IQR	20.0	18.0	11.5	12.6

Table 5.2: Distributions selected for different periods with the greatest P-value and the corresponding parameter estimation

Period	1	2	3	4
Distribution	Gamma	Gamma	Weibull	Weibull
P-value	0.05	0.24	0.07	0.05
Parameter(s)	(1.40,15.79)	(1.49,13.46)	(1.03, 14.48)	(0.97, 13.17)

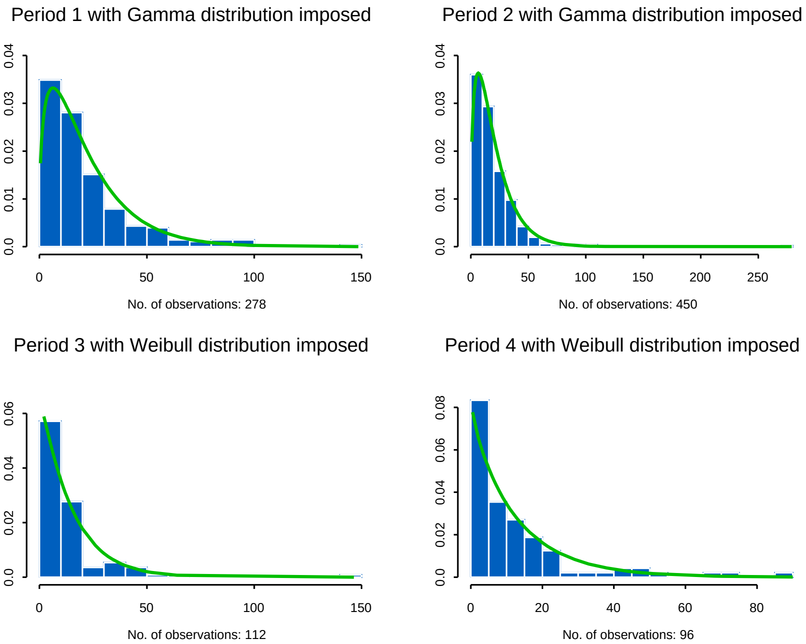

Figure 5.2: The histograms and densities for data from each period

Table **??** shows the distributions with greatest P-values from KS-test for data from four periods and the estimated parameters for Gamma or Weibull distributions (from ML estimation). Both Periods 1 and 2 entail gamma distributions. While Weibull distribution has greatest P-value in Periods 3 and 4, the closeness of the shape parameter to 1 implies that the exponential distribution is also a good approximation. Figure **??** shows the histograms and the estimated Gamma or Weibull densities for these four periods. Summarising the analysis above, it seems that the distributions for Periods 1 and 2 differ only slightly from each other, seen both from the first two moments and the distributions estimated from the data. So do Periods 3 and 4.

In addition, the P-values of KS-test for testing the distributional equivalences between Periods 1 and 2, and Periods 3 and 4 are 0.49 and 0.41 respectively. All other two sample KS-tests showed significant distributional differences. Therefore, it will be worth to merge the data into two periods and to look further their distributional behaviours.

5.2.2 Two periods

In this subsection we make the two combinations of data over periods: Period I = Periods 1 + 2, Period II = Periods 3 + 4. Figure **??** shows histograms and density estimates for these combined periods. Note that the ranges in x-axis are different. P-values from the KS-test indicate that none of four distributions

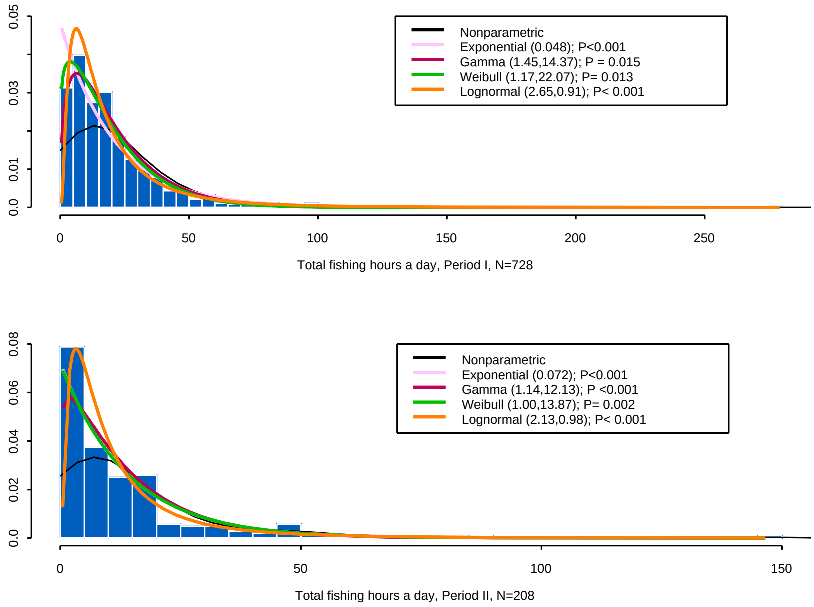

Figure 5.3: Histogram and densities for Period I (upper) and Period II (Lower)

is fitted well.

5.3 Data by year

If we consider the distribution of fishing time each year separately, the results are much more promising. Table **??** summarises the results from KS-tests with the P-values and the estimated parameters for each four distributions. It can be seen that both Gamma and Weibull distributions fit well the data for all years except 2003 and 2005. Note that year 2005 is also fitted good by these two distributions if the extreme value would be omitted.

5.4 Data by section

If we look at the data based on each watercourse section, then to our delight, it is shown in Table **??** that both Gamma and Weibull distributions fits very well, for all kind of periods. Note that only those sections having more than 20 observations were tested. As examples, Figure **??** shows the fitted models for data from Section 7 and Section 15 in Period I and Period II, respectively. The major difference between two parametric models is their performance around the mode. In general, the gamma distribution fits the peak better than the Weibull distribution.

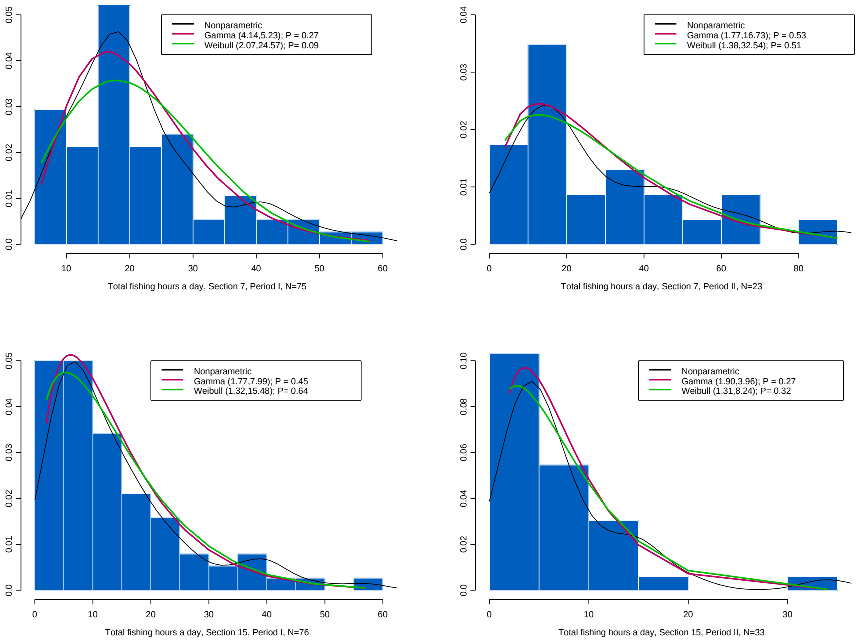

Figure 5.4: Histogram and densities for Section 7 (upper) and Section 15 (Lower)

5.5 Ties and Randomisation

As mentioned earlier, due to the problem with ties, the P-values obtained from the KS-test are not accurate. In this section we consider some issues for solving the problem.

5.5.1 "‘Thinning”

Firstly, we simply eliminate ties by “thinning” the data. That is, we consider only those 129 unique values in our data. Under this circumstance, we would be very satisfied with the goodness of fit by either Gamma or Weibull distribution, as shown in Figure **??**.

Next, we will investigate the effect of randomisation. Ties in the observations will be untied by resampling. That is, we resample the tied data by using randomisation within the interval of plus-minus half hour, in principle. The randomisation is conducted according to three different distributions: triangular, uniform, and mixture of uniform distributions. The number of replicates is 100 each.

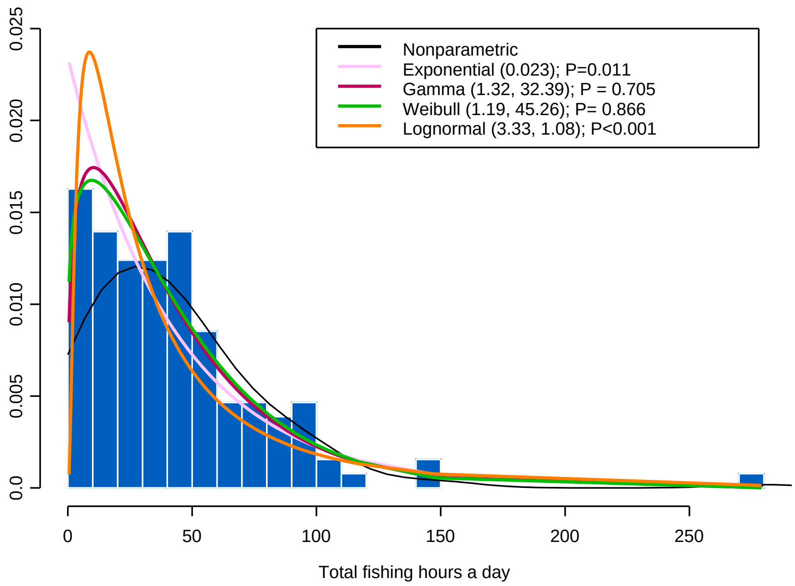

Figure 5.5: Histogram and density estimates for the unique values in the whole data

5.5.2 Triangular distribution

Each observed fishing time t_i is resampled according to the triangular distribution on the interval $[(t_i - 0.5) \vee 0, t_i + 0.5]$ with mode $= t_i$:

$$\mathrm{Tri}(a, b, c) = \mathrm{Tri}\left((t_i - \frac{1}{2}) \vee 0, t_i + \frac{1}{2}, t_i\right).$$

5.5.3 Uniform distribution

Each observed fishing time t_i is resampled according to the uniform distribution on the interval $[(t_i - 0.5) \vee 0, t_i + 0.5]$:

$$U\left[(t_i - \frac{1}{2}) \vee 0, t_i + \frac{1}{2}\right].$$

5.5.4 Mixture of uniform distributions

Each observed fishing time t_i is resampled according to the mixture of four uniform distributions given as follows:

$$p_1 U\left[t_i - \frac{1}{2}, t_i - \frac{1}{4}\right] + p_2 U\left[t_i - \frac{1}{4}, t_i\right] + p_3 U\left[t_i, t_i + \frac{1}{4}\right] + p_4 U\left[t_i + \frac{1}{4}, t_i + \frac{1}{2}\right],$$

where $(p_1, p_2, p_3, p_4) = (1/6, 1/3, 1/3, 1/6)$. Note that above mixture is applied when the observed fishing time is an integer, which is the most frequent

Table 5.3: Distributions selected for different periods with the greatest P-value, after three methods of randomization. The mean and coefficient of variation (CV) of the P-values from KS-test over 100 replicates

	Gamma				Weibull		
Triangular	Whole	Period I	Period 1	Period 2	Period II	Period 3	Period 4
Mean	0.01	0.08	0.11	0.50	0.01	0.14	0.13
CV	0.33	0.25	0.21	0.13	0.33	0.22	0.28
Uniform							
Mean	0.02	0.09	0.13	0.53	0.02	0.15	0.15
CV	0.38	0.24	0.27	0.12	0.37	0.32	0.29
Mixture							
Mean	0.01	0.08	0.16	0.49	0.02	0.15	0.11
CV	0.29	0.19	0.11	0.10	0.18	0.14	0.11

occurrence. In other cases, we simplify the distribution as follows:

$$\begin{cases} U\left[t_i - \dfrac{1}{4}, t_i + \dfrac{1}{4}\right] & \text{if } 4t_i = \text{integer}; \\ U\left[\dfrac{[4t_i]}{4}, \dfrac{[4t_i]+1}{4}\right], & \text{otherwise.} \end{cases}$$

where $[x]$ denotes the largest integers $\leq x$.

Table **??** presents the selected distributions by the greatest mean P-values of KS-test after randomisation. The tests for the whole data, data from four periods (Periods 1–4), as well as data from two combined periods (Periods I and II) are included. 100 resampled copies are generated and tested for goodness-of-fit for distributions Exponential, Gamma, Weibull, and Log-normal, respectively. It shows that Gamma distribution fits best for the first two periods and and their combination, i.e., the summer season from mid-summer day to the end of August, and among them, Period 2 (16th of July to 31th of August) has the highest P-value. For those periods outside this summer season, i.e., Periods 3 and 4 and their combination Period II, Weibull distribution fits better. However, the whole data set rejects all candidates of distributions, as we have seen for the original tied or unique data. This convinces us that fishing times are distributed differently for different time periods.

If we compare the three methods of randomisation, the resulting mean P-values seems similar, and the uniform gives the highest one in most of cases. It is more interesting to note that the mixture method has the lowest coefficient of variation (CV) in all cases.

Considering the distribution of fishing time each year separately, the results are better than that from the original data. Table **??** presents the mean P-values from KS-test on each year over 100 replicates of randomisation using the three methods. Both Gamma and Weibull distributions fit well the data for all years except 2003. All three methods of randomisation give consistent results, while the mixture provides the least variation.

Furthermore, regarding the most "stable" behaviour from the KS-test on the data by section, it is interesting to know how they behave after randomisation. As one can expect, all the KS-tests based on the randomised data show that both Gamma and Weibull distributions are good candidates to model the fishing time at section level. Since it is easier to interpret the parameters in the Gamma distribution with respect to fishing time, and also better fits the mode of the distribution. we choose to summarise only the testing results for Gamma distribution and the mixture method of randomisation. Table **??** shows the average of estimated parameters from the KS-test and the mean fishing hours μ on each section at different periods, over 100 replicates of randomisation.

We take two sections as examples for interpretation the fitted models. Recall that the Gamma distribution with the shape parameter α and the scale parameter β can be regarded as a sum of α independent exponentially distributed random variables, each of which has a mean of β, and the mathematical expectation $\mu = \alpha\beta$.

Consider first section 15 in the watercourse Nedre Ammerån. It has 109 observations in total and the estimated parameters are $\alpha = 1.6, \beta = 7.5$. Thus, the mean fishing time is about 12 hours at this section, where in average 1.6 anglers showed up, having 7.5 fishing hours each. If we look at different periods, it is found that the mean fishing time is almost twice longer in Period I (14 hours) than in Period II (7.4 hours), and within Period I, the mean fishing time is slightly longer in Period 1 (15.7 hours) than in Period 2 (13.3 hours). Consider now section 7 in the watercourse Råndan. The estimated mean fishing time is about 23.5 hours, where in average 3 anglers showed up, each having 8 fishing hours. At the seasonal level, the mean fishing time is longer in Period II (29.6 hours) compared to Period I (21.7 hours). Within Period I, it has also longer mean fishing time in Period 1 (25.5 hours) compared to Period 2 (19.6 hours).

We observe also that the mean fishing time is in general much longer for the sections within the watercourse Hotagsströmmen than those in within the watercourse Nedre Ammerån.

Finally, it shows that after randomisation, there are no ties and P-values are consequently larger than that for the true data. On the other hand, randomisation would not influence the parameter estimation, and the densities estimated from the true and randomised data are almost same.

6 Discussion

A. Parametric models

From the testing results, it is seen that both Gamma and Weibull distributions can be considered as approximate distributions that generate the data. They give close approximation to the data in most of the cases. Gamma distribution fits very well for summer season while Weibull distribution is more appropriate for the other periods.

Recall that exponential distribution can be understood as the data generating mechanism that anglers come and leave randomly. The “memorylessness” property of exponential can be understood in a same manner. In addition, exponential occurs when describing the lengths of the inter-arrival times in a homogeneous Poisson process, where the intensity of the Poisson process indicates the mean number of anglers arrived.

Therefore it could be reasonable to describe the fishing time that each angler spent per day as a random variable with exponential distribution. Consequently, the total fishing time per day can be well characterised by the Gamma distribution, because the sum of independent exponential variables is Gamma-distributed.

In general, the gamma model is easier to interpret and better fits the mode of the distribution, and therefore, is preferred. Having parameters estimated, we are able to calculate probabilities of different fishing times.

B. Periods

The distributions for Periods 1 and 2 differ only slightly from each other, seen both from the first two moments and the distributions estimated from the data. So do Periods 3 and 4. Therefore, it may be suggested to use two periods, i.e. Period I: from mid-summer to the end of August and Period II: other dates.

C. Fishing hours relate to the dates

It is not easy to relate the amount of fishing time to the dates. In that case, the data have to be divided yearly such that the data size is rather short for every year. There is no clear pattern for different years. In addition, using “date” (or similar time variable) as explanatory variable leads to clear conclusion. The explanatory variable is usually not significant (at level 5%).

D. Sections

The modelling at section level seems to be successful. Both the Gamma and Weibull distributions fits well the data for all periods, providing that the number of observations are not less than 20. The mean fishing hours varies from section to section, even within the same watercourse.

E. Others

Although there are more than 4000 total observations, only 936 are on unique days and watercourse sections. The data length is still very limited if they are divided into different years (13 years) or sections (22 sections). This raises problem of density estimation. The temperature, precipitation, and water temperature data have been analysed as explanatory variables to fishing time. Note that those series cover different time period of fishing data and fishing data are discontinuous (in terms of days), therefore it is not clear how to set a regression model. We picked up all fishing time data with temperature and precipitation data of the dates that are available, and tried different combination of explanatory variables, including previous days' temperature and/or precipitation. The explanatory power is limited and the estimated parameters are usually not significant (at level 5%). Other data or approaches are to be pursued in this aspect.

Acknowledgement

We acknowledge support from the research program “Adaptive management of Fish and Wildlife” financed by the Swedish Environmental Protection Agency and from the research program “Hydropower - Environmental impacts, mitigation measures and costs in regulated waters”, an R&D program established and financed by Elforsk, the Swedish Energy Agency, the National Board of Fisheries and the Swedish Environmental Protection Agency.

A Appendix: KS-tests by section and year

Table A.1: P-values and estimated parameters from the KS-test for sections over different periods

Section	Whole data				Period I				Period II				Period 1				Period 2			
ID	n	p	α	β	n	p	α	β	n	p	α	β	n	p	α	β	n	p	α	β
1	36	0.2	1.4	13.9	33	0.2	1.5	13.5												
2	21	0.7	1.3	15.3	21	0.7	1.3	15.3												
5	26	0.6	1.4	15.5																
6	46	0.4	2.7	6.7	32	0.7	3.2	5.5												
7	98	0.1	3.0	7.9	75	0.3	4.1	5.2	23	0.5	1.8	16.7	26	0.5	4.1	6.2	49	0.7	4.6	4.3
8	39	0.9	4.4	4.1	29	0.7	4.0	4.4												
9	85	0.5	2.7	7.0	70	0.5	2.9	6.4					27	0.6	2.8	7.5	43	0.6	3.1	5.6
15	109	0.1	1.6	7.5	76	0.5	1.8	8.0	33	0.3	1.9	4.0	25	0.8	1.4	11.4	51	0.6	2.1	6.4
16	107	0.1	1.3	7.4	76	0.6	1.3	8.7	31	0.2	1.8	3.2	29	0.7	1.4	7.6	47	0.3	1.2	9.3
17	107	0.1	1.8	5.1	73	0.5	2.0	5.2	34	0.1	2.0	2.9	28	0.9	2.2	4.4	45	0.8	1.9	5.7
20	39	0.8	1.8	32.7	35	0.7	2.1	28.0									23	0.4	1.9	28.6
21	27	0.3	1.8	26.5	26	0.3	2.0	24.1												
22	21	0.9	1.9	7.8																
23	46	0.5	1.8	18.1	38	0.2	2.2	15.9									23	0.2	2.6	13.5
24	73	0.9	1.7	10.4	63	0.9	1.6	10.7					24	0.8	1.5	13.2	39	0.7	1.7	8.9
33	21	0.8	1.9	9.6																

Table A.2: P-values from the KS-test for different years and the corresponding parameter estimation

		Exponential		Gamma			Weibull			Lognormal		
Year	n	P-value	β	P-value	α	β	P-value	α	β	P-value	μ	σ
1994	41	0.01	0.06	0.33	1.55	11.08	0.46	1.38	18.78	0.00	2.49	1.00
1995	68	0.00	0.13	0.15	1.47	5.21	0.18	1.16	8.14	0.01	1.66	0.86
1996	43	0.05	0.08	0.59	1.63	7.43	0.66	1.35	13.18	0.50	2.15	0.96
1997	147	0.04	0.06	0.06	1.35	12.18	0.13	1.16	17.47	0.00	2.39	0.96
1998	62	0.03	0.07	0.67	1.75	8.10	0.53	1.31	15.51	0.53	2.34	0.81
1999	137	0.00	0.07	0.78	1.79	7.77	0.85	1.38	15.27	0.00	2.33	0.84
2000	65	0.02	0.05	0.66	1.87	10.81	0.70	1.43	22.32	0.04	2.71	0.82
2002	53	0.00	0.05	0.63	2.57	7.63	0.41	1.54	21.97	0.04	2.77	0.64
2003	77	0.00	0.04	0.02	1.87	13.82	0.02	1.39	28.48	0.00	2.96	0.78
2004	65	0.02	0.03	0.78	1.57	19.79	0.95	1.35	33.84	0.02	3.08	0.97
2005	80	0.07	0.02	0.04	1.38	29.61	0.05	1.13	43.03	0.08	3.31	0.90
2006	51	0.01	0.08	0.66	2.02	6.48	0.62	1.39	14.49	0.50	2.31	0.73
2007	47	0.04	0.06	0.29	1.74	9.27	0.20	1.32	17.67	0.50	2.47	0.82
Total	936	0.00	0.05	0.00	1.33	14.52	0.00	1.11	20.12	0.00	2.54	0.95

Table A.3: Mean P-values from the KS-test by year, over 100 replicates of randomisation, using three methods: Triangular, Uniform, and Mixture

		Exponential			Gamma			Weibull			Lognormal		
Year	n	Tri	Uni	Mix	Tri	Uni	Mix	Tri	Uni	Mix	Tri	Uni	Mix
1994	41	0.02	0.02	0.01	0.43	0.46	0.32	0.53	0.55	0.52	0.00	0.00	0.00
1995	68	0.02	0.03	0.01	0.26	0.28	0.25	0.35	0.38	0.31	0.24	0.37	0.09
1996	43	0.05	0.05	0.04	0.59	0.60	0.52	0.66	0.66	0.59	0.43	0.33	0.49
1997	147	0.05	0.06	0.06	0.07	0.08	0.09	0.14	0.15	0.17	0.01	0.01	0.01
1998	62	0.04	0.04	0.05	0.69	0.68	0.71	0.62	0.65	0.67	0.31	0.34	0.49
1999	137	0.00	0.00	0.00	0.93	0.93	0.95	0.93	0.95	0.92	0.05	0.13	0.02
2000	65	0.03	0.04	0.03	0.77	0.79	0.79	0.80	0.82	0.81	0.21	0.33	0.06
2002	53	0.00	0.00	0.00	0.66	0.64	0.68	0.44	0.42	0.44	0.33	0.41	0.41
2003	77	0.00	0.00	0.00	0.04	0.04	0.05	0.04	0.04	0.05	0.01	0.02	0.03
2004	65	0.03	0.03	0.03	0.84	0.85	0.90	0.95	0.94	0.97	0.02	0.03	0.02
2005	80	0.08	0.08	0.07	0.06	0.07	0.08	0.08	0.09	0.10	0.46	0.49	0.50
2006	51	0.01	0.02	0.01	0.83	0.83	0.82	0.73	0.75	0.61	0.50	0.50	0.50
2007	47	0.09	0.09	0.14	0.32	0.36	0.33	0.23	0.26	0.24	0.50	0.50	0.50

Table A.4: Mean estimated parameters from the KS-test and the mean fishing hours (μ) for sections over different periods, based on 100 replicates of mixture randomisation

Section	Whole data				Period I				Period II				Period 1				Period 2			
ID	n	μ	α	β	n	μ	α	β	n	μ	α	β	n	μ	α	β	n	μ	α	β
1	36	19.6	1.4	13.9	33	20.2	1.5	13.5												
2	21	20.2	1.3	15.4	21	20.2	1.3	15.4												
5	26	21.7	1.4	15.2																
6	46	18.2	2.7	6.7	32	17.4	3.1	5.6												
7	98	23.5	3.0	7.9	75	21.6	4.1	5.2	23	29.7	1.8	16.6	26	25.5	4.1	6.2	49	19.6	4.6	4.3
8	39	18.3	4.4	4.2	29	17.7	4.0	4.5												
9	85	19.0	2.8	6.9	70	18.8	3.0	6.4					27	21.1	2.8	7.6	43	17.5	3.2	5.5
15	109	12.1	1.6	7.5	76	14.1	1.8	7.9	33	7.5	1.9	4.0	25	15.7	1.4	11.2	51	13.3	2.1	6.3
16	107	9.4	1.3	7.4	76	10.9	1.3	8.6	31	5.8	1.8	3.2	29	10.5	1.3	7.8	47	11.2	1.2	9.1
17	107	9.0	1.8	5.1	73	10.6	2.0	5.3	34	5.8	2.0	2.9	28	9.8	2.2	4.5	45	11.0	1.9	5.8
20	39	58.4	1.8	32.9	35	58.4	2.1	28.0									23	55.5	1.9	28.6
21	27	46.7	1.7	26.9	26	48.3	2.0	24.4												
22	21	14.5	1.9	7.5																
23	46	31.8	1.8	18.2	38	35.5	2.2	16.2									23	34.6	2.5	13.8
24	73	17.2	1.7	10.4	63	17.2	1.6	10.7					24	20.1	1.50	13.42	39	15.4	1.8	8.7
33	21	18.3	1.9	9.5																