The Age Distribution of Firms in Ireland, 1961–2009

Stephen Kinsella*

October 28, 2009

Abstract

This paper studies the age distribution of firms in Ireland from 1961 to 2009 using data on over 159,000 firms. I build a model which mixes a lognormal distribution of firm size with an exponential distribution of firm age to explain the evolution of these data. The model is fitted using maximum likelihood estimation, and tested using Kolmogorov-Smirnov statistics. The model replicates observed empirical regularities.

Keywords: Firm Age, Firm Size, Firm Growth.
JEL Codes: L20; L25; L26.

*Department of Economics, Kemmy Business School, University of Limerick, Ireland. Email: stephen.kinsella@ul.ie, www.stephenkinsella.net. I thank Alex Coad, Helena Lenihan, Donal Palcic, David Ramsey, and Jacco Thijssen for very helpful comments. All errors are, unfortunately, my own.

1. Introduction

Few studies concentrate directly on the age distribution of firms. To my knowledge, no studies of the Irish case exist. Until recently, high resolution data was hard to obtain, difficult to study using costly computational resources, and even then, most data sets only covered larger firms. More recent studies have observed an exponential distribution of firm age[1].

This paper makes two contributions. First, the paper provides a comprehensive description of firm age in Ireland. Second, the paper exposes the previously unexplored data to a new model, using a novel estimation method. The main result is to find that the distribution found in the data can be obtained by mixing a Gibrat-type process to generate firm sizes[2] with an exponential distribution for firm ages.

Further work will describe the implications of the findings of this paper with regards to Ireland's industrial policy over the period.

1.1. Motivation

Beyond simple curiosity, and recognizing the tautological importance of firm age in the success or failure of Irish firms in particular sectors, there are several reasons for studying the age distribution of Irish firms separately from the size distribution.

1. Firm age is a major determinant of firm growth for individual firms. In an influential article on firm age, Thornhill and Amit [2003, p. 498] write that

   > Firms are at the greatest risk of failure when they are young and small. Beyond an early peak in mortality rates, often described as the liability of adolescence, exit rates monotonically decline to a positive asymptote.

2. The age distribution of firms provides indirect data on entry and exit rates in sectors, barriers to entry, survival rates, and technologies used in production in Irish and international firms located in Ireland.

3. Firm age is a major determinant of firm growth in and across sectors. The relationship between firm size and firm age is well studied[3]. Though the likelihood of

[1] Coad [2008] surveys the literature. The literature on firm *size* is vast, with representative papers being Gabaix et al. [2003], Ijiri and Simon [1977], Crosato and Ganugi [2007], Cabral and Mata [2003], Coad and Tamvada [2008]. Firm *age*, however, has not received as much attention until recently.

[2] Details on generating Gibrat processes may be found in Sutton [1997], Aitchison and Brown [1957], de Wit [2005].

[3] Representative studies include Greiner [1998], Evans [1987], Variyam and Kraybill [1992], and Yasuda [2005].

firm growth is essentially stochastic, there seems to be a negative relationship between firm growth and firm age after a certain threshold, which seems to depend on the data being studied, and is as yet inconclusive[4]. The literature in this area is currently emerging, so there are few fixed points of reference for this study.

1.2. Objective

This paper studies the age distribution of firms in Ireland using a database of over 159,000 firms across every sector[5] of the Irish economy from 1961 to 2009[6].

The density of the distribution across all Irish firms appears to follow an approximately straight line, with a negative slope over most of the support. Plotted in rank-frequency space, this straight line suggests that an exponential distribution would be a valid approximation of the empirical age distribution of firms in Ireland.

In this paper, I build a model to generate this exponential distribution which agrees with economic intuition, and I fit the model to real world data using maximum likelihood estimation. I use Kolmogorov-Smirnov (KS) tests of goodness of fit for these distributions, following Clauset et al. [2007], and the KS statistics report a close fit with the exponential with $D \leq 0.1$.

The rest of the paper is laid out as follows. Section 2. explores the data set. Section 3. describes the model used to generate similar distributions to the data. Section 4. fits the empirical distribution to the derived distribution from the model. Section 5. concludes, and appendix A describes the fitting procedure.

2. Data

Data are taken from the `AMADEUS` and `KOMPASS` databases of firms registered in Ireland, both active and inactive, across all available NACE codes[7]. I take the 'birth date' of any firm to be the date of its registration as recorded on the `AMADEUS` dataset. Appropriately

---

[4]In a rare study, Das [1995] studies rapidly growing Indian technology firms, and finds the opposite result, confirming the economic intuition that explosive growth is more likely to occur at the earliest stages of firm age in certain industries.

[5]The NACE code system is the European standard for industry classifications and was introduced in 1970. In 1990 a revised version became applicable. NACE stands for '*Nomenclature Generale des Activites Economiques dans I'Union Europeenne*' (General Name for Economic Activities in the European Union). The current version from 2008 is based on the 'International Standard Industrial Classification of all economic activities' (ISIC) of the United Nations.

[6]Data was gathered using the `AMADEUS` and `KOMPASS` databases. The author gratefully acknowledges the help of the librarians of Trinity College, Dublin, in gathering these data.

[7]Following Simon (1962) and Velupillai [2000].

filtered, the total number of firms is 159,902. 'Age' is here defined as the date of registration of the company with the Irish Companies Registration Office (CRO). The firm is thus 'born' in this dataset when it registers with the CRO. The legal definition of firm birth has not changed appreciably over the last 49 years in Ireland, although more reporting strictures have been imposed upon newer firms in recent years. There may well be measurement error in the firm age data compiled here—for example, when a firm was registered in 2009, its age was given as 1—but this measurement error is not large enough to discourage an analysis of the age distribution using these data. There are also no data on firms' exits from particular sectors, which would allow a 'mortality' rate to be developed.

There were no upper and lower thresholds placed upon the data. The mean age of the firms is 9.49 years, median age is 7 years, with a standard deviation of 8.46. Firms are rather young relative to international studies[8]. The firms this paper studies range in age from 1 year to 47 years of age. Table 1 shows the relevant data descriptors, including data on firm age by percentile. The percentiles show us a highly skewed distribution, which squares well with the finding that over 97% of firms 'alive' in Ireland are both very young (that is, less than three years' old) and very small.

		Percentile	Year
n	159,902	1%	1
Mean	9.49	10%	1
Std Deviation	8.47	25%	3
Minimum	1	50%	7
Maximum	47	75%	13
		90%	21
		99%	37
		100%	47

Table 1: Descriptive statistics for aggregated AMADEUS data.

Figure 1 shows, in sequence, different characteristics of the data, and provides an empirical justification for modeling these data using lognormal and exponential distri-

[8]For example, Coad and Tamvada [2008] found a median age of 12 years for a sample of 700,000 Indian firms.

butions. Moving left to right, we see simple unranked tally of firms by age (panel 2(a)). Clearly many firms currently alive are very young, with more than 40% of the sample being less than 3 years' old. The data ranked by age, the probability of observing an individual firm x which has been alive for t years is $P(x_t)$. Plotted in log-log space, the frequency of the age distribution is suggestive of an exponential in the support. Panel 2(b) generates 150,000 ages for firms between 1 and 47 from a lognormal distribution, and compares the simulated firm ages (in red) to the actual firm ages (in blue, dotted). Simple OLS confirms there is a high correspondence between the two, and KS tests confirm this at the 99% confidence interval. Panel 2(c) generates the probability distribution function (PDF), for an exponential distribution, and fits to a randomnly sampled empirical distribution function. Panel 2(d) plots the empirical distribution function for the entire sample. Simple fitting by ordinary least squares (OLS) to the tail of the distribution does support the hypothesis that the distribution is exponentially distributed above 100,000 firms[9].

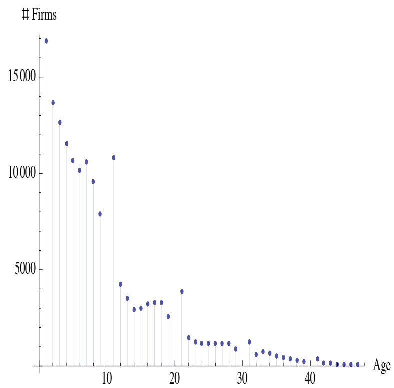
3. Model

Figure 1 plots the same graph for different years, overlaying the results for one year over the results for other years. The general shape of the distribution rather than the 'pooled' version in figure 2(d) could be discerned this way, and also it might be possible to observe long-term changes over time.

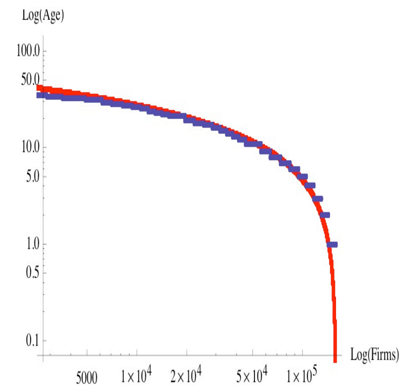
Here the modeling strategy is to model the size of the firm— relative to its age—in two parts. I posit a standard, lognormal-distributed generating function for the firm size, following Simon [1955], Coad [2008], Cabral and Mata [2003], initially assuming firms are of identical ages, and then to let t flow freely by mixing an exponential distribution of firm age with a lognormal distribution of firm size, following Coad [2008], Coad and Tamvada [2008].

Let x_t be the size of firm x at time t. 'Size' here can be measured by a number of proxies, for example: market capitalization, number of employees, turnover, operating profit, and market share. Let ϵ_t be a random, *iid* variable representing a multiplicative growth shock between each period. The following relation then holds:

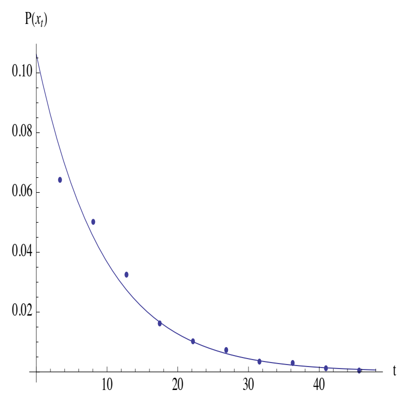
[9] OLS results computed in *Mathematica* regressing the data on the assumed distributional shape age $P(\text{AGE}_t) = \frac{C}{\text{AGE}_t^{-k}}$, where a is a constant and α the scaling parameter, show an R^2 of 0.847, an estimated variance of 24.7, with both C and k strongly significant at the 5% confidence interval. The best fitted distributional function, according to OLS, is $P(\text{AGE}_t) = 336/\text{AGE}_t^{0.32}$



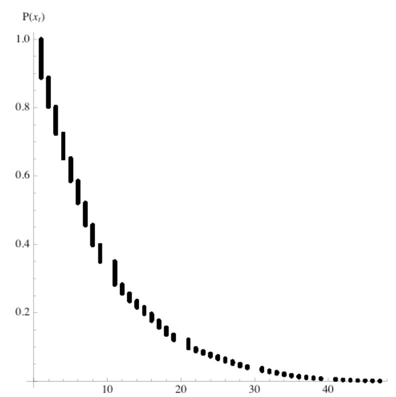
(a) Unranked tally of number of firms by age.



(b) Ranked plot of ages of firms on log-log scales. Red line shows 150,000 simulated lognormally distributed firms. Blue (dotted) line is actual firms.



(c) Fitting of exponential probability density function to actual data.



(d) Empirical distribution function of $P(x_{it} > x_{jt})$.

Figure 1: Graphical representations of the actual and simulated data. An exponential and/or lognormal distribution of firm age is clearly evident.

$$x_t - x_{t-1} = \epsilon x_{t-1}. \quad (1)$$

Expanding equation 1, we can be obtain:

$$x_t = (1 + \epsilon_t)x_{t-1} = x_0(1 + \epsilon_1) + (1 + \epsilon_2) + \ldots \bar{\epsilon} + (1 + \epsilon_t). \quad (2)$$

Taking logs either side of equation 2 to approximate $\log(1+\epsilon_t)$ by ϵ_t, we get

$$\log x_t \approx \log x_0 + \epsilon_1 + \epsilon_2 + \ldots + \epsilon_t = \log(x_0) + \sum_{s=1}^{t} \epsilon_s \tag{3}$$

As t become large, the logged initial size of the firm will become very small, and may be dropped.

$$\log(x_t) \approx \sum_{i=1}^{t} \epsilon_i. \tag{4}$$

Applying the central limit theorem, we can say that $\log(x_t)$ is normally distributed for large t, and therefore firm size is lognormally distributed.

$$P(x_t) = \frac{1}{x_t\sqrt{2\pi\sigma^2}} e^{-\frac{(\log(x_t)-\bar{\epsilon_t})^2}{2\sigma_t^2}}. \tag{5}$$

Now remove the assumption that t has an identical value for all firms, and allow t's value to float freely. If we assume the distribution of the firm's age to be exponentially distributed, we can posit the following exponential distribution for t:

$$P(t) = \lambda e^{-\lambda t}. \tag{6}$$

It is important to discover a rule to mix the two distributions: the lognormal distribution for firm size, and the exponential distribution for firm age. Huberman and Adamic [1999] mix these distributions according to this rule: if a variable α's distribution $P(\alpha|\beta)$ depends on a parameter β, and β itself is distributed according to another probability distribution $P(\beta)$, then the distribution of α is given by $P(\alpha) = \int P(\beta) \cdot P(\alpha|\beta) d\beta$. This rule is, admittedly, arbitrary, but other rules can be posited without difficulty, and further work will concentrate on discovering appropriate mixings of these distributions.

Mixing the lognormal distribution of firm size and the exponential distribution of firm age, we can obtain

$$P(x_t) = \int_0^\infty \left(-\lambda e^{\lambda t} \cdot \frac{1}{x_t\sqrt{2\pi\sigma^2}} e^{-\frac{(\log(x_t)-\bar{\epsilon_t})^2}{2\sigma_t^2}} \right) dt, \tag{7}$$

which can be developed to give

$$P(x_t) = C \cdot x_t^{-k}. \tag{8}$$

C is a constant whose value is given by $C = \lambda/\sigma(\sqrt{(\bar{\epsilon})^2 + 2\lambda})$, with $1 < k < \infty$. k's value is given by $k = 1 - \frac{\bar{\epsilon}}{\sigma^2} + \frac{\sqrt{\bar{\epsilon}^2 + 2\lambda\sigma^2}}{\sigma^2}$.

There is a simple economic intuition behind equation 8. If λ is large, then the exponential decay is strong, so it will be unusual to find firms with an age above the median age. When the mean growth rate is close to zero, then $\bar{\epsilon}$ will be close to 1, and if λ is small, then the likelihood of finding firms with an age greater than the average will be low.

We can then study the evolution of both λ and $\bar{\epsilon}$ over the 47 years of Irish firm age data we have available.

The next section exposes our simple model to the data described in section 2..

4. Fitting

Details of the fitting procedure are given in the appendix. We fit 159,902 ages of Irish firms to equation 8 via maximum likelihood[10]. Figure **??** plots the fit of the firms ages to equation (8). The best fitted values for λ and ϵ by the maximum likelihood method are 26.2 and 3.1, respectively, across the entire period. The observed functional form for the model is

$$P(x_t) = \frac{24.3}{x^{0.09}}. \tag{9}$$

R^2 and adjusted R^2 for this model are 0.54 and 0.54, respectively, with $p < 0.05$ in all cases studied.

It is possible to break the age distribution down further by NACE code, and I will do so for four representative sectors: firms with more than 500 employees, firms in export sectors, and firms in high technology sectors, and light manufacturing. The influence of Irish industrial policy over the time period can clearly be seen, even up to the present day.

5. Conclusion and Further Work

This paper studies the age distribution of Irish firms from 1961 to 2009. The paper builds a model which mixes lognormal (firm size) and exponential (firm age) distributions to obtain an empirically valid and tractable approximation to the data on over 159,000 firms available. The simple distributional model is simulated, and tested for several parameter ranges and does replicate the observed findings on firm age tolerably.

Further work will concentrate on expanding the description of firm growth and decay

[10] *Mathematica* code for this fitting is available from the author.

by including firm death, when certain firms go out of business. Then a true birth and death rate of firms in Ireland can be computed exactly.

References

Aitchison, J. and J. A. C. Brown, *The Lognormal Distribution, with Special Reference to Its Use in Economics*, New York: Cambridge University Press, 1957.

Cabral, Luís M. B. and José Mata, "On the Evolution of the Firm Size Distribution: Facts and Theory," *American Economic Review*, 2003, *94* (4), 1075–1090.

Clauset, Aaron, Cosma Rohilla Shalizi, and M. E. J. Newman, "Power-law distributions in empirical data," arXiv:0706.1062 2007.

Coad, Alex, "The exponential age distribution and the Pareto firm size distribution," Jena Economic Research Papers in Economics 2008-072, Friedrich-Schiller-University Jena, Max-Planck-Institute of Economics, Thueringer Universitaets- und Landesbibliothek September 2008.

Coad, Alexander and Jagannadha Pawan Tamvada, "The Growth and Decline of Small firms In Developing Countries," Papers on Econonmics and Evolution 2008-08, Max Planck Institute of Economics, Evolutionary Economics Group October 2008.

Crosato, L. and P. Ganugi, "Statistical regularity of firm size distribution: the Pareto IV and truncated Yule for Italian SCI manufacturing," *Statistical Methods and Applications*, 2007, *16* (1), 85–115.

Das, S., "Size, Age, and Firm Growth in an Infant Industry: The Computer Hardware industry in India," *International Journal of Industrial Organization*, 1995, *13* (1), 111–126.

de Wit, G., "Firm size distributions an overview of steady-state distributions resulting from firm dynamics models," *International Journal of Industrial Organization*, 2005, *23* (5), 423–450.

Evans, David S., "The Relationship between Firm Growth, Size and Age: Estimates for 100 Manufacturing Industries," *Journal of Industrial Economics*, 1987, *35* (3), 567–581.

Gabaix, X., Gopikrishnan P., Plerou V., and Stanley H.E., "A theory of power-law distributions in financial market fluctuations," *Nature*, May 2003, *15* (423(6937)), 267–70.

Greiner, L., "Evolution and Revolution as Organisations Grow," *Family Business Review*, 1998, *10* (4), 397–409.

Huberman, B.A. and L.A. Adamic, "Growth Dynamics of the World Wide Web," *Nature*, 1999, *401* (6749), 131–131.

Ijiri, Y. and H. A. Simon, *Skew Distributions and the Sizes of Business Firm*, North-Holland: Amsterdam, 1977.

Jeynes, E. T., *The Maximum Entropy Formalism*, MIT Press, 1978.

Simon, H. A, "On a Class of skew distribution functions," *Biometrika*, 1955, *42*, 425–440.

Simon, Herbert A., "The Architecture of Complexity," *Proceedings of the American Philosophical Society*, December 1962, *106* (6), 467–482.

Sutton, John, "Gibrat's Legacy," *Journal of Economic Literature*, 1997, *XXXV* (1), 40–59.

Thornhill, S. and R. Amit, "Learning about Failure: Bankruptcy, Firm Age, and the Resource-Based View," *Organization Science*, 2003, *14* (5), 497–509.

Variyam, J. N. and D. S. Kraybill, "Empirical Evidence on Determinants of Firm Growth," *Economics Letters*, 1992, *38* (4), 31–36.

Velupillai, K. V., *Computable economics: the Arne Ryde memorial lectures*, Oxford, UK: Oxford University Press, 2000.

Yasuda, T., "Firm Growth, Size, Age and Behaviour in Japanese Manufacturing," *Small Business Economics*, 2005, *24* (1), 1–15.

A Maximum Likelihood Estimation

Let X denote the variable(s) of interest that has (or is assumed to have) a probability density function (pdf): $f(x;\theta)$, dependent upon a $(k \times 1)$ parameter $\theta \in \Theta \subset \mathbb{R}^k$, whose true value θ_0 is unknown.

Assume the functional form of f is known.

Let $(X_1, \ldots, X_n)$ denote a random sample of size n drawn on X, in practice a time series of arbitrary length. It is assumed that the pdf of the sample $f_1, \ldots, n|(x_1, \ldots, x_n; \theta)$ can be derived from the knowledge we have about f, so the joint density we derive herein will depend upon the unknown parameter θ. A key point we need to stress is the likelihood function is mathematically equivalent to the joint distribution of the sample for the data types under study.

The real benefit of MLE in the age of vast computational resources is that instead of regarding the likelihood as a function of the X_i's, the likelihood can be interpreted as a function of θ defined over the parameter space Θ for fixed values of each $X_i = x_i$. The advantage of this approach is the recursive structure each x_i inherently has.

The simple likelihood for θ is

$$L\left(\theta | x_1, \ldots, x_n\right) \equiv f_{1,\ldots,n}\left(x_1, \ldots, x_n; \theta\right). \tag{10}$$

It makes most sense to shorten the notation for the likelihood to just $L(\theta)$, because the finer points of the definition are not used below.

Construction of the joint pdf may at first sight seem a daunting task. However, if the variables in $(X_1, \ldots, X_n)$ are mutually independent, then the joint pdf is given by the product of the marginals shown in equation 11:

$$f_{1,\ldots,n}\left(x_1, \ldots, x_n; \theta\right) = \prod_{i=1}^{n} f\left(x_i; \theta\right), \tag{11}$$

which usually makes it easy to construct the joint pdf and hence the likelihood for θ. The literature on the necessary independence of explanatory variables and the controversy this assumption has created is summarised and criticized in Jeynes [1978], though his contribution has spawned a large literature which is beyond the scope of this paper.

It is important usually to distinguish between two forms of the likelihood for θ, namely, the likelihood function, and the *observed* likelihood function.

The likelihood function is defined as the likelihood for θ, given the random sample prior to observation; it is given by $L(\theta|X_1, \ldots, X_n)$, and is a random variable.

The observed likelihood is defined as the likelihood for θ evaluated for a given sample of observed data, and it is *not* random. The examples developed below illustrate the construction of the likelihood, and its observed counterpart.

In simple cases, the likelihood function is easily constructed, since due to mutual independence, the joint distribution of the random sample is simply the product of the marginal distributions. In practical financial situations, however, mutual independence amongst the sampling variables does not occur, and so the derivation of the likelihood function requires more effort.

Maximum likelihood parameter estimation is based on choosing values for θ so as to maximise the likelihood function. That is, the MLE of θ, denoted $\hat{\theta}$, is the solution to the maximized problem:

$$\hat{\theta} = \arg\max_{\theta \in \Theta} L\left(\theta | X_1 = x_1, \ldots, X_n = x_n\right). \tag{12}$$

$\hat{\theta}$ is the value of the argument of the likelihood, selected from anywhere in the parameter space that maximises the value of the likelihood after we have been given the sample. In other words, we seek the particular value of θ, namely, $\hat{\theta}$, which makes it most likely to have observed the sample that we actually have. We may view the solution to 12 in two ways depending on whether the objective function is the *likelihood function* or the *observed likelihood function*. If the objective is the likelihood, then 12 defines the ML estimator,$\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$; since this is a function of the random sample, $\hat{\theta}$ is a random variable. If the objective is the observed likelihood, then 12 defines the ML estimate,$\hat{\theta} = \hat{\theta}(x_1, \ldots, x_n)$, where $x_1, \ldots, x_n$ denotes observed data; in this case $\hat{\theta}$ is a point estimate.

The solution to 12 is invariant to any monotonic increasing transformation of the objective. Since the natural logarithm is a monotonic transformation, it follows that

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \log L(\theta), \tag{13}$$

which we will use from now on as the definition of the estimator (estimate). The natural logarithm of the likelihood, log L(θ), is called the log-likelihood function. A weaker definition of the MLE, but one that, in practice, is often equivalent to 12 is

$$\hat{\theta} = \arg\max_{\tilde{\theta} \in \tilde{\Theta}} \log L\left(\tilde{\theta}\right) \tag{14}$$

where $\tilde{\Theta}$ denotes a finite, non-null set whose elements $\tilde{\theta}$ satisfy the conditions

$$\frac{\partial}{\partial \theta} \log L(\hat{\theta}) = 0 \tag{15}$$

and

$$\frac{\partial^2}{\partial \theta^2} \log(\hat{\theta}) \leq 0. \tag{16}$$

The two parts of 16 express, respectively, the first- and second-order conditions familiar from basic calculus for determining local maxima of a function. Generally speaking, we determine MLE through computation of the 16, although the example below, finding MLE for the exponential parameter, relies on 13 alone. One further piece of notation is the so-called score (or 'efficient score' in some texts), defined as the gradient of the log-likelihood,

$$S(\theta) = \frac{\partial}{\partial \theta} \log L(\theta). \tag{17}$$

For example, the first-order condition is simply $S(\tilde{\theta}) = 0$.

Let $X \sim \text{Exp}(\theta)$, where parameter $\theta \in \mathbb{R}_+$. Here is its pdf, assuming the domain of f runs from x to infinity, and $\theta > 0 : f = \frac{1}{\theta} e^{-x/\theta}$.

For a random sample of size n drawn on X, the log-likelihood function is:

$$\log L\theta = \log \left[\prod_{i=1}^{n} (f/.x \to x_i) \right]$$

$$= -\frac{n\theta \log[\theta] + \sum_{i=1}^{n} x_i}{\theta} \tag{18}$$

The score function is the gradient of the log-likelihood with respect to θ:

$$\text{Score } = \frac{-n\theta + \sum_{i=1}^{n} x_i}{\theta^2}. \tag{19}$$

Setting the score to zero and solving for θ corresponds to the first-order condition given in 16. We find:

$$\theta = \frac{\sum_{i=1}^{n} x_i}{n}. \tag{20}$$

The unique solution corresponds to the sample mean, the first moment of the distribution. The nature of the solution is not yet clear; that is, does the sample mean correspond to a local minimum, local maximum, or saddle point of the log-likelihood? A check of the second-order condition, evaluated at the solution point gives

$$-\frac{n^3}{\left(\sum_{i=1}^{n} x_i\right)^2}, \tag{21}$$

which reveals that the Hessian is strictly negative at the sample mean and therefore the log-likelihood is maximized at the sample mean. Hence, the MLE of θ is

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n}(X_i). \tag{22}$$

Now let $X \sim N(\mu, \sigma^2)$, where $\mu \in \mathbb{R}$ and $\sigma^2 \in \mathbb{R}_+$, with pdf $f(x; \mu, (\sigma^2)$:

$$f = \frac{1}{\sigma\sqrt{2\pi}}\text{Exp}\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]; \tag{23}$$

Assuming domain on f of $-\infty \leftarrow x \rightarrow \infty$ and a real-valued mean, for a random sample of size n drawn on X, the log-likelihood for parameter $\theta = (\mu, \sigma)$ is:

$$\log L\theta = \log\left[\prod_{i=1}^{n}(f/.x \rightarrow x_i)\right] \tag{24}$$

which gives

$$-\frac{1}{2\sigma^2}\left(n\left(\mu^2 + \sigma^2\text{Log}[2\pi] + 2\sigma^2\log[\sigma]\right) - 2\mu\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} x_i^2\right) \tag{25}$$

The score vector $S(\theta) = S(\mu, \sigma)$ is given by:

$$\left(\frac{-n\mu + \sum_{i=1}^{n} x_i}{\sigma^2}, \frac{n\mu^2 - n\sigma^2 - 2\mu \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} x_i^2}{\sigma^3}\right) \tag{26}$$

Setting the score to zero and solving yields:

$$\left(\left(\sigma \to -\frac{\sqrt{-\frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n} + \sum_{i=1}^{n} x_i^2}}{\sqrt{n}}, \mu \to \frac{\sum_{i=1}^{n} x_i}{n}\right), \left(\sigma \to \frac{\sqrt{-\frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n} + \sum_{i=1}^{n} x_i^2}}{\sqrt{n}}, \mu \to \frac{\sum_{i=1}^{n} x_i}{n}\right)\right) \tag{27}$$

Clearly, the negative-valued solution for σ lies outside the parameter space and is therefore invalid; thus, the only permissible solution to the first-order conditions is:

$$\left(\sigma \to \frac{\sqrt{-\frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n} + \sum_{i=1}^{n} x_i^2}}{\sqrt{n}}, \mu \to \frac{\sum_{i=1}^{n} x_i}{n}\right) \tag{28}$$

Then

$$\hat{\theta} = (\hat{\mu}, \hat{\sigma})$$

is the MLE of θ, where $\hat{\mu}$ and $\hat{\sigma}$ are the formulae given as the solution values for θ (second-order conditions must be checked below). The functional form for $\hat{\sigma}$ may appear unfamiliar. However, if we utilize the following identity for the sum of squared deviations about the sample mean,

$$\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2 = \sum_{i=1}^{n} X_i^2 - n\bar{X}^2$$

where $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} (X_i)$, then

$$\hat{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2}.$$

Invoking MLE's Invariance Property the MLE of σ^2 is

$$(\hat{\sigma})^2 = \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2 \tag{29}$$

which is the 2nd sample central moment.

The second-order conditions may, for example, be checked by examining the eigenvalues of the Hessian matrix evaluated at $\hat{\theta}$:

$$\left(\frac{n^3}{\left(\sum_{i=1}^n x_i\right)^2 - n\sum_{i=1}^n x_i^2}, \frac{2n^3}{\left(\sum_{i=1}^n x_i\right)^2 - n\sum_{i=1}^n x_i^2}\right) \tag{30}$$

Given the identity for the sum of squared deviations, the eigenvalues of the Hessian are $-n\hat{\sigma}^{-2}$ and $-2n\hat{\sigma}^{-2}$, which clearly are negative. Thus, the Hessian is negative definite at $\hat{\theta}$ and therefore the log-likelihood is maximized at $\hat{\theta}$.

B MLE for Pareto-distributed data

Let $X \sim \text{Pareto}(\alpha, \beta)$, where parameters $\alpha \in \mathbb{R}_+$ and $\beta \in \mathbb{R}_+$. Assuming $\alpha > 0, \beta > 0$ and allowing the domain of f to fluctuate from $x \to \infty$, the pdf of X is given by:

$$f = \alpha\beta^{\alpha}x^{-(\alpha+1)}. \tag{31}$$

Since $X \geq \beta$, there exists dependence between the parameter and sample spaces. Given a random sample of size n collected on X, the log-likelihood for $\theta = (\alpha, \beta)$ is:

$$(\log[\alpha] + \alpha\log[\beta]) - (1+\alpha)\sum_{i=1}^n \log[x_i] \tag{32}$$

The score vector is given by:

$$\left(n\left(\frac{1}{\alpha} + \log[\beta]\right) - \sum_{i=1}^n \log[x_i], \frac{n\alpha}{\beta}\right). \tag{33}$$

If we attempt to solve the first-order conditions in the usual way, we cannot find a solution to the equations. However, if we focus on solving just the first of the first-order conditions, we find:

$$\left(\alpha \to -\frac{n}{n\log[\beta] - \sum_{i=1}^n \log[x_i]}\right) \tag{34}$$

This time a solution is provided, albeit in terms of β; that is, $\hat{\alpha} = \hat{\alpha}(\beta)$. We now take this solution and substitute it back into the log-likelihood:

$$n\left(-1 + \log\left[\frac{n}{-n\log[\beta] + \sum_{i=1}^n \log[x_i]}\right]\right) - \sum_{i=1}^n \log[x_i] \tag{35}$$

This function is known as the concentrated log-likelihood. It corresponds to $\log L(\hat{\alpha}(\beta), \beta)$. Since it no longer involves α, we can maximise it with respect to β. Let $\hat{\beta}$ denote the solution to this optimization problem.

This solution can then be substituted back to recover $\hat{\alpha} = \hat{\alpha}(\hat{\beta})$; then $\hat{\theta} = (\hat{\alpha}, \hat{\beta})$ would be the MLE of θ by recursion.

In general, when the first-order conditions can be solved uniquely for some subset of parameters in θ, then those solutions can be substituted back into the log-likelihood to yield the concentrated log-likelihood. The concentrated log-likelihood is then maximized with respect to the remaining parameters, usually using numerical techniques.

For my example, maximizing the concentrated log-likelihood using standard calculus will not work. This is because the parameter space depends on the sample space. However, by inspection, it is apparent that the concentrated log-likelihood is increasing in β. Therefore, we should select β as large as possible. Now, since each $X_i \geq \beta$, we can choose β no larger than the smallest observation. Hence, the MLE for β is

$$\hat{\beta} = \min(X_1, X_2, \ldots, X_n) \tag{36}$$

which is the smallest order statistic. Replacing β in $\hat{\alpha}(\beta)$ with $\hat{\beta}$ yields the MLE for α,

$$\hat{\alpha} = n / \sum_{i=1}^{n} \log\left(\frac{X_i}{\min(X_1, X_2, \ldots, X_n)}\right) \blacksquare \tag{37}$$