This formula is equivalent to assuming that every observation in a group is equal to the group's midpoint. For a limit of 125,000 the expected cost is

$$\frac{1}{227}[3{,}750(99)+12{,}500(42)+25{,}000(29)+50{,}000(28)+96{,}250(17)+125{,}000(12)]=27{,}125.55.$$

For a limit of 300,000, the expected cost is

$$\frac{1}{227}[3{,}750(99)+12{,}500(42)+25{,}000(29)+50{,}000(28)+96{,}250(17)+212{,}500(9)+300{,}000(3)]$$
$$=32{,}907.49.$$

The ratio is 32,907.49/27,125.55 = 1.213, or a 21.3% increase. Note that if the last group has a boundary of infinity, the final integral will be undefined, but the contribution to the sum in the last two lines is still correct. ☐

Exercise 16 *Estimate the variance of the number of accidents using Data Set A assuming that all seven drivers with five or more accidents had exactly five accidents.*

Exercise 17 *Estimate the expected cost per loss and per payment for a deductible of 25,000 and a maximum payment of 275,000 using Data Set C.*

Exercise 18 *(*) You are studying the length of time attorneys are involved in settling bodily injury lawsuits. T represents the number of months from the time an attorney is assigned such a case to the time the case is settled. Nine cases were observed during the study period, two of which were not settled at the conclusion of the study. For those two cases, the time spent up to the conclusion of the study, 4 months and 6 months, was recorded instead. The observed values of T for the other seven cases are as follows – 1, 3, 3, 5, 8, 8, 9. Estimate* $\Pr(3 \le T \le 5)$ *using the Kaplan-Meier product-limit estimate.*

2.3 Estimation for parametric models

2.3.1 Introduction

If a phenomenon is to be modeled using a parametric model, it is necessary to assign values to the parameters. This could be done arbitrarily, but it would seem to be more reasonable to base the assignment on observations from that phenomenon. In particular, we will assume that n independent observations have been collected. For some of the techniques used in this Section it will be further assumed that all the observations are from the same random variable. For others, that restriction will be relaxed.

The methods introduced in the next Subsection are relatively easy to implement, but tend to give poor results. The following Subsection covers maximum likelihood estimation. This method is more difficult to use, but has superior statistical properties and is considerably more flexible.

2.3.2 Method of moments and percentile matching

For these methods we assume that all n observations are from the same parametric distribution. In particular, let the distribution function be given by

$$F(x|\boldsymbol{\theta}),\ \boldsymbol{\theta}^T=(\theta_1,\theta_2,\ldots,\theta_p).$$

That is, $\boldsymbol{\theta}$ is a vector containing the p parameters to be estimated. Furthermore, let

$$\begin{aligned}\mu_k'(\boldsymbol{\theta}) &= E(X^k|\boldsymbol{\theta}) \text{ be the } k\text{th raw moment, and let} \\ &\quad \pi_g(\boldsymbol{\theta}) \text{ be the } 100g\text{th percentile}\end{aligned}$$

of the random variable. That is, $F[\pi_g(\boldsymbol{\theta})|\boldsymbol{\theta}] = g$. For a sample of n independent observations from this random variable, let

$$\begin{aligned}\hat{\mu}_k' &= \frac{1}{n}\sum_{j=1}^{n} x_j^k \text{ be the empirical estimate of the } k\text{th moment, and let} \\ &\quad \hat{\pi}_g \text{ be the empirical estimate of the } 100g\text{th percentile.}\end{aligned}$$

Definition 2.25 *A* ***method of moments estimate*** *of $\boldsymbol{\theta}$ is any solution of the p equations*

$$\mu_k'(\boldsymbol{\theta}) = \hat{\mu}_k',\ k = 1, 2, \ldots, p.$$

The motivation for this estimator is that it produces a model that has the same first p raw moments as the data (as represented by the empirical distribution). The traditional definition of the method of moments uses positive integers for the moments. Arbitrary negative or fractional moments could also be used. In particular, when estimating parameters for inverse distributions, matching negative moments may be a superior approach.[4]

Example 2.26 *Use the method of moments to estimate parameters for the exponential, gamma, and Pareto distributions for Data Set B.*

The first two sample moments are

$$\begin{aligned}\hat{\mu}_1' &= \frac{1}{20}(27 + \cdots + 15,743) = 1,424.4 \\ \hat{\mu}_2' &= \frac{1}{20}(27^2 + \cdots + 15,743^2) = 13,238,441.9.\end{aligned}$$

For the exponential distribution the equation is

$$\theta = 1,424.4$$

with the obvious solution, $\hat{\theta} = 1,424.4$.

For the gamma distribution, the two equations are

$$\begin{aligned}E(X) &= \alpha\theta = 1,424.4 \\ E(X^2) &= \alpha(\alpha+1)\theta^2 = 13,238,441.9.\end{aligned}$$

Dividing the second equation by the square of the first equation yields

$$\begin{aligned}\frac{\alpha+1}{\alpha} &= 6.52489 \\ 1 &= 5.52489\alpha\end{aligned}$$

---

[4] One advantage is that with appropriate moments selected, the equations must have a solution within the range of allowable parameter values.

and so $\hat{\alpha} = 1/5.52489 = 0.18100$ and $\hat{\theta} = 1,424.4/0.18100 = 7,869.61$.

For the Pareto distribution, the two equations are

$$\begin{aligned} E(X) &= \frac{\theta}{\alpha - 1} = 1,424.4 \\ E(X^2) &= \frac{2\theta^2}{(\alpha-1)(\alpha-2)} = 13,238,441.9. \end{aligned}$$

Dividing the second equation by the square of the first equation yields

$$\frac{2(\alpha-1)}{(\alpha-2)} = 6.52489$$

with a solution of $\hat{\alpha} = 2.442$ and then $\hat{\theta} = 1,424.4(1.442) = 2,053.985$. □

There is no guarantee that the equations will have a solution, or if there is a solution, that it will be unique.

Exercise 19 *Determine the method of moments estimate for an exponential model for Data Set B with observations censored at 250.*

Exercise 20 *Determine the method of moments estimate for a lognormal model for Data Set B.*

Definition 2.27 *A* ***percentile matching estimate*** *of* $\boldsymbol{\theta}$ *is any solution of the* p *equations*

$$\pi_{g_k}(\boldsymbol{\theta}) = \hat{\pi}_{g_k},\ k = 1, 2, \ldots, p$$

where $g_1, g_2, \ldots, g_p$ *are* p *arbitrarily chosen percentiles. From the definition of percentile, the equations can also be written*

$$F(\hat{\pi}_{g_k}|\boldsymbol{\theta}) = g_k,\ k = 1, 2, \ldots, p.$$

The motivation for this estimator is that it produces a model with p percentiles that match the data (as represented by the empirical distribution). As with the method of moments, there is no guarantee that the equations will have a solution, or if there is a solution, that it will be unique. One problem with this definition is that percentiles for discrete random variables are not always well-defined. For example, with Data Set B and using the definition of percentile from the Part 3 Study Note, the 50th percentile is any number between 384 and 457 (because half the sample is below and above any of these numbers). The convention is to use the midpoint. However, for other percentiles, there is no “official” interpolation scheme. The following definition will be used in this Note.

Definition 2.28 *The* ***smoothed empirical estimate*** *of a percentile is found by*

$$\begin{aligned} \hat{\pi}_g &= (1-h)x_{(j)} + hx_{(j+1)}, \text{ where} \\ j &= [(n+1)g] \text{ and } h = (n+1)g - j. \end{aligned}$$

Here $[\cdot]$ *indicates the greatest integer function and* $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$ *are the order statistics from the sample.*

Unless there are two or more data points with the same value, no two percentiles will have the same value. One feature of this definition is that $\hat{\pi}_g$ cannot be obtained for $g < 1/(n+1)$ or $g > n/(n+1)$. This seems reasonable as we should not expect to be able to infer the value of large or small percentiles from small samples. We will use the smoothed version whenever an empirical percentile estimate is called for.

Example 2.29 *Use percentile matching to estimate parameters for the exponential and Pareto distributions for Data Set B.*

For the exponential distribution, select the 50th percentile. The empirical estimate is the traditional median of $\hat{\pi}_{0.5} = (384 + 457)/2 = 420.5$ and the equation to solve is

$$\begin{aligned} 0.5 &= F(420.5|\theta) = 1 - e^{-420.5/\theta} \\ \ln 0.5 &= -420.5/\theta \\ \hat{\theta} &= -420.5/\ln 0.5 = 606.65. \end{aligned}$$

For the Pareto distribution, select the 30th and 80th percentiles. The smoothed empirical estimates are found as follows:

$$\begin{aligned} \text{30th:} &\quad j = [21(0.3)] = [6.3] = 6,\ h = 6.3 - 6 = 0.3 \\ \hat{\pi}_{0.3} &= 0.7(161) + 0.3(243) = 185.6 \\ \text{80th:} &\quad j = [21(0.8)] = [16.8] = 16,\ h = 16.8 - 16 = 0.8 \\ \hat{\pi}_{0.8} &= 0.2(1,193) + 0.8(1,340) = 1,310.6. \end{aligned}$$

The equations to solve are

$$\begin{aligned} 0.3 &= F(185.6) = 1 - \left(\frac{\theta}{185.6 + \theta}\right)^{\alpha},\ 0.8 = F(1,310.6) = 1 - \left(\frac{\theta}{1310.6 + \theta}\right)^{\alpha} \\ \ln 0.7 &= -0.356675 = \alpha \ln\left(\frac{\theta}{185.6 + \theta}\right),\ \ln 0.2 = -1.609438 = \alpha \ln\left(\frac{\theta}{1310.6 + \theta}\right) \\ \frac{-1.609438}{-0.356675} &= 4.512338 = \frac{\ln\left(\frac{\theta}{1310.6+\theta}\right)}{\ln\left(\frac{\theta}{185.6+\theta}\right)}. \end{aligned}$$

The Excel Solver or Goal Seek routine[5] can be used to solve this equation for $\hat{\theta} = 715.03$. Then, from the first equation,

$$0.3 = 1 - \left(\frac{715.03}{185.6 + 715.03}\right)^{\alpha}$$

which yields $\hat{\alpha} = 1.54559$. □

Exercise 21 *(*) The 20th and 80th percentiles from a sample are 5 and 12 respectively. Using the percentile matching method, estimate $S(8)$ assuming the population has a Weibull distribution.*

---

[5] Instructions on using these routines is provided in Appendix B

Exercise 22 *(*) From a sample you are given that the mean is 35,000, the standard deviation is 75,000, the median is 10,000, and the 90th percentile is 100,000. Using the percentile matching method, estimate the parameters of a Weibull distribution.*

Exercise 23 *(*) A sample of size five produced the values 4, 5, 21, 99, and 421. You fit a Pareto distribution using the method of moments. Determine the 95th percentile of the fitted distribution.*

Exercise 24 *(*) In year 1 there are 100 claims with an average size of 10,000 and in year 2 there are 200 claims with an average size of 12,500. Inflation increases the size of all claims by 10% per year. A Pareto distribution with* $\alpha = 3$ *and* θ *unknown is used to model the claim size distribution. Estimate* θ *for year 3 using the method of moments.*

Exercise 25 *(*) From a random sample the 20th percentile is 18.25 and the 80th percentile is 35.8. Estimate the parameters of a lognormal distribution using percentile matching and then use these estimates to estimate the probability of observing a value in excess of 30.*

Exercise 26 *(*) A claim process is a mixture of two random variables* A *and* B*.* A *has an exponential distribution with a mean of 1 and* B *has an exponential distribution with a mean of 10. A weight of* p *is assigned to distribution* A *and* $1-p$ *to distribution* B*. The standard deviation of the mixture is 2. Estimate* p *by the method of moments.*

2.3.3 Maximum likelihood

2.3.3.1 Introduction

Estimation by the method of moments and percentile matching is often easy to do, but these estimators tend to perform poorly. The main reason for this is that they use a few features of the data, rather than the entire set of observations. It is particularly important to use as much information as possible when the population has a heavy right tail. For example, when estimating parameters for the normal distribution, the sample mean and variance are sufficient.[6] However, when estimating parameters for a Pareto distribution, it is important to know all the extreme observations in order to successfully estimate α. Another drawback of these methods is that they require that all the observations are from the same random variable. Otherwise, it is not clear what to use for the population moments or percentiles. For example, if half the observations have a deductible of 50 and half have a deductible of 100, it is not clear to what the sample mean should be equated.[7] Finally, these methods allow the analyst to make arbitrary decisions regarding the moments or percentiles to use.

There are a variety of estimators that use the individual data points. All of them are implemented by setting an objective function and then determining the parameter values that optimize that function. For example, we could estimate parameters by minimizing the maximum difference between the distribution function for the parametric model and the distribution function for the Nelson-Åalen estimate. Of the many possibilities, the only one used in this Note is the maximum

---

[6] This applies both in the formal statistical definition of sufficiency (not covered in this Note) and in the conventional sense. If the population has a normal distribution, the sample mean and variance convey as much information as the original observations.

[7] One way to rectify that drawback is to first determine a data-dependent model such as the Kaplan-Meier estimate. Then use percentiles or moments from that model.

likelihood estimator. The general form of this estimator is presented in this introduction. The following Subsections present useful special cases.

To define the maximum likelihood estimator, let the data set consist of n events $A_1, \ldots, A_n$ where A_j is whatever was observed for the jth observation. For example, A_j may consist of a single point or an interval. The latter arises with grouped data or when there is censoring. For example, when there is censoring at u and a censored observation is observed, the event is the interval from u to infinity. Further assume that the event A_j results from observing the random variable X_j. The random variables $X_1, \ldots, X_n$ need not have the same probability distribution, but their distributions must depend on the same parameter vector, $\boldsymbol{\theta}$. In addition, the random variables are assumed to be independent.

Definition 2.30 *The* ***likelihood function*** *is*

$$L(\boldsymbol{\theta}) = \prod_{j=1}^{n} \Pr(X_j \in A_j|\boldsymbol{\theta})$$

and the ***maximum likelihood estimate*** *of $\boldsymbol{\theta}$ is the vector that maximizes the likelihood function.*[8]

Note that there is no guarantee that the function has a maximum at eligible parameter values. It is possible that as various parameters become zero or infinite the likelihood function will continue to increase. Care must be taken when maximizing this function because there may be local maxima in addition to the global maximum. Often, it is not possible to analytically maximize the likelihood function (by setting partial derivatives equal to zero). Numerical approaches, such as those outlined in Appendix B, will usually be needed.

Because the observations are assumed to be independent, the product in the definition represents the joint probability $\Pr(X_1 \in A_1, \ldots, X_n \in A_n|\boldsymbol{\theta})$, that is, the likelihood function is the probability of obtaining the sample results that were obtained, given a particular parameter value. The estimate is then the parameter value that produces the model under which the actual observations are most likely to be observed. One of the major attractions of this estimator is that it is almost always available. That is, if you can write an expression for the desired probabilities, you can execute this method. If you cannot write and evaluate an expression for probabilities using your model, there is no point in postulating that model in the first place because you will not be able to use it to solve your problem.

Example 2.31 *Suppose the data in Data Set B was censored at 250. Determine the maximum likelihood estimate of θ for an exponential distribution.*

The first 7 data points are uncensored. For them, the set A_j contains the single point equal to the observation x_j. When calculating the likelihood function for a single point for a continuous model it is necessary to interpret $\Pr(X_j = x_j) = f(x_j)$. That is, the density function should be used. Thus the first seven terms of the product are

$$f(27)f(82)\cdots f(243) = \theta^{-1}e^{-27/\theta}\theta^{-1}e^{-82/\theta}\cdots\theta^{-1}e^{-243/\theta} = \theta^{-7}e^{-909/\theta}.$$

---

[8] Some authors write the likelihood function as $L(\boldsymbol{\theta}|\mathbf{x})$ where the vector $\mathbf{x}$ represents the observed data. Because observed data can take many forms, in this Note the dependence of the likelihood function on the data is suppressed in the notation.

For the final 13 terms, the set A_j is the interval from 250 to infinity and therefore $\Pr(X_j \in A_j) = \Pr(X_j > 250) = e^{-250/\theta}$. There are thirteen such factors making the likelihood function

$$L(\theta) = \theta^{-7}e^{-909/\theta}(e^{-250/\theta})^{13} = \theta^{-7}e^{-4159/\theta}.$$

It is easier to maximize the logarithm of the likelihood function. Because it occurs so often, we denote the **loglikelihood function** as $l(\boldsymbol{\theta}) = \ln L(\boldsymbol{\theta})$. Then

$$\begin{aligned} l(\theta) &= -7\ln\theta - 4159\theta^{-1} \\ l'(\theta) &= -7\theta^{-1} + 4159\theta^{-2} = 0 \\ \hat{\theta} &= 4159/7 = 594.14. \end{aligned}$$

In this case, the calculus technique of setting the first derivative equal to zero is easy to do. Also, evaluating the second derivative at this solution produces a negative number, verifying that this solution is a maximum. □

2.3.3.2 Complete, individual data

When there is no truncation, no censoring, and the value of each observation is recorded, it is easy to write the loglikelihood function.

$$L(\boldsymbol{\theta}) = \prod_{j=1}^{n} f_{X_j}(x_j|\boldsymbol{\theta}),\ l(\boldsymbol{\theta}) = \sum_{j=1}^{n} \ln f_{X_j}(x_j|\boldsymbol{\theta}).$$

The notation indicates that it is not necessary for each observation to come from the same distribution.

Example 2.32 *Using Data Set B determine the maximum likelihood estimates for an exponential distribution, for a gamma distribution where α is known to equal 2, and for a gamma distribution where both parameters are unknown.*

For the exponential distribution, the general solution is

$$\begin{aligned} l(\theta) &= \sum_{j=1}^{n} \left(-\ln\theta - x_j\theta^{-1}\right) = -n\ln\theta - n\bar{x}\theta^{-1} \\ l'(\theta) &= -n\theta^{-1} + n\bar{x}\theta^{-2} = 0 \\ n\theta &= n\bar{x},\ \hat{\theta} = \bar{x}. \end{aligned}$$

For Data Set B, $\hat{\theta} = \bar{x} = 1,424.4$. The value of the loglikelihood function is -165.23. For this situation the method of moments and maximum likelihood estimates are identical.

For the gamma distribution with $\alpha = 2$,

$$\begin{aligned} f(x|\theta) &= \frac{x^{2-1}e^{-x/\theta}}{\Gamma(2)\theta^2} = x\theta^{-2}e^{-x/\theta},\ \ln f(x|\theta) = \ln x - 2\ln\theta - x\theta^{-1} \\ l(\theta) &= \sum_{j=1}^{n} \ln x_j - 2n\ln\theta - n\bar{x}\theta^{-1} \\ l'(\theta) &= -2n\theta^{-1} + n\bar{x}\theta^{-2} = 0 \\ \hat{\theta} &= \bar{x}/2. \end{aligned}$$

For Data Set B, $\hat{\theta} = 1,424.4/2 = 712.2$ and the value of the loglikelihood function is -179.98. Again, this estimate is the same as the method of moments estimate.

For the gamma distribution with unknown parameters the equation is not as simple.

$$f(x|\alpha,\theta) = \frac{x^{\alpha-1}e^{-x/\theta}}{\Gamma(\alpha)\theta^{\alpha}},\ \ln f(x|\alpha,\theta) = (\alpha-1)\ln x - x\theta^{-1} - \ln\Gamma(\alpha) - \alpha\ln\theta.$$

The partial derivative with respect to α requires the derivative of the gamma function. The resulting equation cannot be solved analytically. Using the Excel Solver, the estimates are quickly identified as $\hat{\alpha} = 0.55616$ and $\hat{\theta} = 2,561.1$ and the value of the loglikelihood function is -162.29. These do not match the method of moments estimates. □

Exercise 27 *Repeat the previous example using the inverse exponential, inverse gamma with $\alpha = 2$, and inverse gamma distributions. Record the value of the loglikelihood function at the maximum (it will be used later) and then compare your estimates with the method of moments estimates.*

2.3.3.3 Complete, grouped data

When data are complete and grouped, the observations may be summarized as follows. Begin with a set of numbers $c_0 < c_1 < \cdots < c_k$ where c_0 is the smallest possible observation (often zero) and c_k is the largest possible observation (often infinity). From the sample, let n_j be the number of observations in the interval $(c_{j-1}, c_j]$. For such data, the likelihood function is

$$L(\boldsymbol{\theta}) = \prod_{j=1}^{k}[F(c_j|\boldsymbol{\theta}) - F(c_{j-1}|\boldsymbol{\theta})]^{n_j},\ l(\boldsymbol{\theta}) = \sum_{j=1}^{k} n_j \ln[F(c_j|\boldsymbol{\theta}) - F(c_{j-1}|\boldsymbol{\theta})].$$

Example 2.33 *From Data Set C, determine the maximum likelihood estimate for an exponential distribution.*

The loglikelihood function is

$$\begin{aligned} l(\theta) &= 99\ln[F(7,500) - F(0)] + 42\ln[F(17,500) - F(7,500)] + \cdots + 3\ln[1 - F(300,000)] \\ &= 99\ln(1 - e^{-7,500/\theta}) + 42\ln(e^{-7,500/\theta} - e^{-17,500/\theta}) + \cdots + 3\ln e^{-300,000/\theta}. \end{aligned}$$

A numerical routine is needed to produce $\hat{\theta} = 29,721$ and the value of the loglikelihood function is -406.03. □

Exercise 28 *From Data Set C, determine the maximum likelihood estimates for gamma, inverse exponential, and inverse gamma distributions.*

2.3.3.4 Truncated or censored data

When data are censored, there is no additional complication. As noted in Example 2.31, right censoring simply creates an interval running from the censoring point to infinity. In that example, data below the censoring point was individual data, and so the likelihood function contains both density and distribution function terms.

Exercise 29 *Determine maximum likelihood estimates for Data Set B using the inverse exponential, gamma, and inverse gamma distributions. Assume the data have been censored at 250 and then compare your answers to those obtained in Example 2.32 and Exercise 27.*

Truncated data presents more of a challenge. There are two ways to proceed. One is to shift the data by subtracting the truncation point from each observation. The other is to accept the fact that there is no information about values below the truncation point, but then attempt to fit a model for the original population.

Example 2.34 *Assume the values in Data Set B had been truncated from below at 200. Using both methods, estimate the value of* α *for a Pareto distribution with* $\theta = 800$ *known. Then use the model to estimate the cost per payment with deductibles of 0, 200, and 400.*

Using the shifting approach, the data becomes 43, 94, 140, 184, 257, 480, 655, 677, 774, 993, 1,140, 1,684, 2,358, and 15,543. The likelihood function is

$$\begin{aligned}
L(\alpha) &= \prod_{j=1}^{14} \frac{\alpha(800^{\alpha})}{(800+x_j)^{\alpha+1}} \\
l(\alpha) &= \sum_{j=1}^{14}[\ln\alpha + \alpha\ln 800 - (\alpha+1)\ln(x_j+800)] \\
&= 14\ln\alpha + 93.5846\alpha - 103.969(\alpha+1) \\
&= 14\ln\alpha - 103.969 - 10.384\alpha \\
l'(\alpha) &= 14\alpha^{-1} - 10.384,\ \hat{\alpha} = 14/10.384 = 1.3482.
\end{aligned}$$

Because the data have been shifted, it is not possible to estimate the cost with no deductible. With a deductible of 200, the expected cost is the expected value of the Pareto variable, $800/0.3482 = 2,298$. Raising the deductible to 400 is equivalent to imposing a deductible of 200 on the modeled distribution. From the Part 3 Note, the expected cost per payment is

$$\frac{E(X) - E(X \wedge 200)}{1 - F(200)} = \frac{\frac{800}{0.3482}\left(\frac{800}{200+800}\right)^{0.3482}}{\left(\frac{800}{200+800}\right)^{1.3482}} = \frac{1000}{0.3482} = 2,872.$$

For the unshifted approach we need to ask the key question required when constructing the likelihood function. That is, what is the probability of observing each value, knowing that values under 200 are omitted from the data set? This becomes a conditional probability and therefore the likelihood function is (where the x_j values are now the original values)

$$\begin{aligned}
L(\alpha) &= \prod_{j=1}^{14} \frac{f(x_j|\alpha)}{1-F(200|\alpha)} = \prod_{j=1}^{14} \left[\frac{\alpha(800^\alpha)}{(800+x_j)^{\alpha+1}} \Big/ \left(\frac{800}{800+200} \right)^\alpha \right] \\
&= \prod_{j=1}^{14} \frac{\alpha(1000^\alpha)}{(800+x_j)^{\alpha+1}} \\
l(\alpha) &= 14 \ln \alpha + 14\alpha \ln 1000 - (\alpha+1) \sum_{j=1}^{14} \ln(800+x_j) \\
&= 14 \ln \alpha + 96.709\alpha - (\alpha+1)105.810 \\
l'(\alpha) &= 14\alpha^{-1} - 9.101,\ \hat{\alpha} = 1.5383.
\end{aligned}$$

This model is for losses with no deductible, and therefore the expected cost without a deductible is $800/0.5383 = 1,486$. Imposing deductibles of 200 and 400 produces the following results:

$$\begin{aligned}
\frac{E(X) - E(X \wedge 200)}{1 - F(200)} &= \frac{1000}{0.5383} = 1,858 \\
\frac{E(X) - E(X \wedge 400)}{1 - F(400)} &= \frac{1200}{0.5383} = 2,229.
\end{aligned}$$

□

Exercise 30 *Repeat the above example using a Pareto distribution with both parameters unknown.*

It should now be clear that the contribution to the likelihood function can be written for most any observation. The following two steps summarize the process:

1. For the numerator, use $f(x)$ if the exact value, x, of the observation is known. If it is only known that the observation is between y and z, use $F(z) - F(y)$.
2. For the denominator, let d be the truncation point (use zero if there is no truncation). The denominator is then $1 - F(d)$.

Example 2.35 *Determine Pareto and gamma models for the time to death for Data Set D2.*

The following table shows how the likelihood function is constructed for these values. For deaths, the time is known and so the exact value of x is available. For surrenders or those reaching time 5, the observation is censored and therefore death is known to be some time in the interval from the surrender time, y, to infinity. In the table, $z = \infty$ is not noted because all interval observations end at infinity.

Obs.	x, y	d	L	Obs.	x, y	d	L
1	$y = 0.1$	0	$1 - F(0.1)$	16	$x = 4.8$	0	$f(4.8)$
2	$y = 0.5$	0	$1 - F(0.5)$	17	$y = 4.8$	0	$1 - F(4.8)$
3	$y = 0.8$	0	$1 - F(0.8)$	18	$y = 4.8$	0	$1 - F(4.8)$
4	$x = 0.8$	0	$f(0.8)$	19-30	$y = 5.0$	0	$1 - F(5.0)$
5	$y = 1.8$	0	$1 - F(1.8)$	31	$y = 5.0$	0.3	$\frac{1-F(5.0)}{1-F(0.3)}$
6	$y = 1.8$	0	$1 - F(1.8)$	32	$y = 5.0$	0.7	$\frac{1-F(5.0)}{1-F(0.7)}$
7	$y = 2.1$	0	$1 - F(2.1)$	33	$x = 4.1$	1.0	$\frac{f(4.1)}{1-F(1.0)}$
8	$y = 2.5$	0	$1 - F(2.5)$	34	$x = 3.1$	1.8	$\frac{f(3.1)}{1-F(1.8)}$
9	$y = 2.8$	0	$1 - F(2.8)$	35	$y = 3.9$	2.1	$\frac{1-F(3.9)}{1-F(2.1)}$
10	$x = 2.9$	0	$f(2.9)$	36	$y = 5.0$	2.9	$\frac{1-F(5.0)}{1-F(2.9)}$
11	$x = 2.9$	0	$f(2.9)$	37	$y = 4.8$	2.9	$\frac{1-F(4.8)}{1-F(2.9)}$
12	$y = 3.9$	0	$1 - F(3.9)$	38	$x = 4.0$	3.2	$\frac{f(4.0)}{1-F(3.2)}$
13	$x = 4.0$	0	$f(4.0)$	39	$y = 5.0$	3.4	$\frac{1-F(5.0)}{1-F(3.4)}$
14	$y = 4.0$	0	$1 - F(4.0)$	40	$y = 5.0$	3.9	$\frac{1-F(5.0)}{1-F(3.9)}$
15	$y = 4.1$	0	$1 - F(4.1)$				

The likelihood function must be maximized numerically. For the Pareto distribution there is no solution. The likelihood function keeps getting larger as α and θ get larger.[9] For the gamma distribution the maximum is at $\hat{\alpha} = 2.617$ and $\hat{\theta} = 3.311$. □

Exercise 31 *Repeat the above example, this time finding the distribution of the time to surrender.*

Discrete data presents no additional problems.

Example 2.36 *For Data Set A, assume that the 7 drivers with 5 or more accidents all had exactly 5 accidents. Determine the maximum likelihood estimate for a Poisson distribution and for a binomial distribution with $m = 8$.*

In general, for a discrete distribution with complete data, the likelihood function is

$$L(\boldsymbol{\theta}) = \prod_{j=1}^{\infty} [p(x_j|\boldsymbol{\theta})]^{n_j}$$

---

[9] For a Pareto distribution, the limit as the parameters α and θ become infinite with the ratio being held constant is an exponential distribution. Thus, for this example, the exponential distribution is a better model (as measured by the likelihood function) than any Pareto model.

where x_j is one of the observed values, $p(x_j|\boldsymbol{\theta})$ is the probability of observing x_j, and n_j is the number of times x_j was observed in the sample. For a Poisson distribution

$$\begin{aligned}
L(\lambda) &= \prod_{x=0}^{\infty}\left(\frac{e^{-\lambda}\lambda^x}{x!}\right)^{n_x} = \prod_{x=0}^{\infty}\frac{e^{-n_x\lambda}\lambda^{xn_x}}{x!^{n_x}} \\
l(\lambda) &= \sum_{x=0}^{\infty}[-n_x\lambda + xn_x\ln\lambda - n_x\ln x!] = -n\lambda + n\bar{x}\ln\lambda - \sum_{x=0}^{\infty}n_x\ln x! \\
l'(\lambda) &= -n + n\bar{x}/\lambda = 0,\ \hat{\lambda} = \bar{x}.
\end{aligned}$$

For a binomial distribution

$$\begin{aligned}
L(q) &= \prod_{x=0}^{m}\left[\binom{m}{x}q^x(1-q)^{m-x}\right]^{n_x} = \prod_{x=0}^{m}\frac{m!^{n_x}q^{xn_x}(1-q)^{(m-x)n_x}}{x!^{n_x}(m-x)!^{n_x}} \\
l(q) &= \sum_{x=0}^{m}[n_x\ln m! + xn_x\ln q + (m-x)n_x\ln(1-q) - n_x\ln x! - n_x\ln(m-x)!] \\
l'(q) &= \sum_{x=0}^{m}\frac{xn_x}{q} - \frac{(m-x)n_x}{1-q} = \frac{n\bar{x}}{q} - \frac{mn-n\bar{x}}{1-q} = 0,\ \hat{q} = \bar{x}/m.
\end{aligned}$$

For this problem, $\bar{x} = [81,714(0)+11,306(1)+1,618(2)+250(3)+40(4)+7(5)]/94,935 = 0.16313$. Therefore $\hat{\lambda} = 0.16313$ and $\hat{q} = 0.16313/8 = 0.02039$. □

Exercise 32 *Repeat the previous example, but this time assume that the actual values for the 7 drivers who have 5 or more accidents are unknown.*

Exercise 33 *(*) Lives are observed in order to estimate* q_{35}*. Ten lives are first observed at age 35.4 and six die prior to age 36 while the other four survive to age 36. An additional twenty lives are first observed at age 35 and eight die prior to age 36 with the other twelve surviving to age 36. Determine the maximum likelihood estimate of* q_{35} *given that the time to death from age 35 has density function* $f(t) = w$, $0 \le t \le 1$ *with* $f(t)$ *unspecified for* $t > 1$.

Exercise 34 *(*) The model has hazard rate function* $h(t) = \lambda_1$, $0 \le t < 2$ *and* $h(t) = \lambda_2$, $t \ge 2$. *Five items are observed from age zero, with the following results:*

Age last observed	*Cause*
1.7	*death*
1.5	*censoring*
2.6	*censoring*
3.3	*death*
3.5	*censoring*

Determine the maximum likelihood estimates of λ_1 *and* λ_2.

Exercise 35 *(*) Your goal is to estimate* q_x*. The time to death for a person age x has a constant density function. In a mortality study, 10 lives were first observed at age x. Of them, one died and one was removed from observation alive at age* $x+0.5$*. Determine the maximum likelihood estimate of* q_x*.*

Exercise 36 *(*) Ten lives are subject to the survival function*

$$S(t) = \left(1 - \frac{t}{k}\right)^{1/2}, \ 0 \le t \le k$$

where t is time since birth. There are ten lives observed from birth. At time 10, two of the lives die and the other eight are withdrawn from observation. Determine the maximum likelihood estimate of k.

Exercise 37 *(*) 500 losses are observed. Five of the losses are 1100, 3200, 3300, 3500, and 3900. All that is known about the other 495 losses is that they exceed 4000. Determine the maximum likelihood estimate of the mean of an exponential model.*

Exercise 38 *(*) 100 people are observed at age 35. Of them, 15 leave the study at age 35.6, 10 die sometime between ages 35 and 35.6, and 3 die sometime after age 35.6 but before age 36. The remaining 72 people survive to age 36. Determine the product-limit estimate of q_{35} and the maximum likelihood estimate of q_{35}. For the latter, assume the time to death is uniform between ages 35 and 36.*

Exercise 39 *(*) The survival function is $S(t) = 1 - t/w$, $0 \le t \le w$. Five claims were studied in order to estimate the distribution of the time from reporting to settlement. After 5 years, four of the claims were settled, the times being 1, 3, 4, and 4. Actuary X then estimates w using maximum likelihood. Actuary Y prefers to wait until all claims are settled. The fifth claim is settled after 6 years at which time Actuary Y estimates w by maximum likelihood. Determine the two estimates.*

Exercise 40 *(*) Four automobile engines were first observed when they were 3 years old. They were then observed for r additional years. By that time three of the engines had failed, with the failure ages being 4, 5, and 7. The fourth engine was still working at age $3 + r$. The survival function has the uniform distribution on the interval 0 to w. The maximum likelihood estimate of w is 13.67. Determine r.*

Exercise 41 *(*) Ten claims were observed. The values of 7 of them (in thousands) were 3, 7, 8, 12, 12, 13, and 14. The remaining 3 claims were all censored at 15. The proposed model has a hazard rate function given by*

$$h(t) = \begin{cases} \lambda_1, & 0 < t < 5 \\ \lambda_2, & 5 \le t < 10 \\ \lambda_3, & t \ge 10. \end{cases}$$

Determine the maximum likelihood estimates of the three parameters.

Exercise 42 *(*) You are given the five observations 521, 658, 702, 819, and 1217. Your model is the single parameter Pareto distribution with distribution function*

$$F(x) = 1 - \left(\frac{500}{x}\right)^{\alpha}, \ x > 500, \ \alpha > 0.$$

Determine the maximum likelihood estimate of α.

Exercise 43 *(*) You have observed the following five claim severities: 11.0, 15.2, 18.0, 21.0, and 25.8. Determine the maximum likelihood estimate of μ for the following model:*

$$f(x) = \frac{1}{\sqrt{2\pi x}} \exp\left[-\frac{1}{2x}(x-\mu)^2\right],\ x, \mu > 0.$$

Exercise 44 *(*) A random sample of size 5 is taken from a Weibull distribution with $\tau = 2$. Two of the sample observations are known to exceed 50 and the three remaining observations are 20, 30, and 45. Determine the maximum likelihood estimate of θ.*

Exercise 45 *(*) Phil and Sylvia are competitors in the light bulb business. Sylvia advertises that her light bulbs burn twice as long as Phil's. You were able to test 20 of Phil's bulbs and 10 of Sylvia's. You assumed that both of their bulbs have an exponential distribution with time measured in hours. You have separately estimated the parameters as $\hat{\theta}_P = 1000$ and $\hat{\theta}_S = 1500$ for Phil and Sylvia respectively, using maximum likelihood. Using all thirty observations, determine $\hat{\theta}^*$, the maximum likelihood estimate of θ_P restricted by Sylvia's claim that $\theta_S = 2\theta_P$.*

Exercise 46 *(*) A sample of 100 losses revealed that 62 were below 1000 and 38 were above 1000. An exponential distribution with mean θ is considered. Using only the given information, determine the maximum likelihood estimate of θ. Now suppose you are also given that the 62 losses that were below 1000 totalled 28,140 while the total for the 38 above 1000 remains unknown. Using this additional information, determine the maximum likelihood estimate of θ.*

Exercise 47 *(*) The following values were calculated from a random sample of 10 losses:*

$$\begin{aligned}
\sum_{j=1}^{10} x_j^{-2} &= 0.00033674 & \sum_{j=1}^{10} x_j^{-1} &= 0.023999 \\
\sum_{j=1}^{10} x_j^{-0.5} &= 0.34445 & \sum_{j=1}^{10} x_j^{0.5} &= 488.97 \\
\sum_{j=1}^{10} x_j &= 31,939 & \sum_{j=1}^{10} x_j^{2} &= 211,498,983.
\end{aligned}$$

Losses come from a Weibull distribution with $\tau = 0.5$ (so $F(x) = 1 - e^{-(x/\theta)^{0.5}}$). Determine the maximum likelihood estimate of θ.

Exercise 48 *(*) For claims reported in 1997, the number settled in 1997 (year 0) was unknown, the number settled in 1998 (year 1) was 3 and the number settled in 1999 (year 2) was 1. The number settled after 1999 is unknown. For claims reported in 1998 there were 5 settled in year 0, 2 settled in year 1, and the number settled after year 1 is unknown. For claims reported in 1999 there were 4 settled in year 0 and the number settled after year 0 is unknown. Let N be the year in which a randomly selected claim is settled and assume that it has probability function* $\Pr(N = n) = p_n = (1-p)p^n$, $n = 0, 1, 2, \ldots$. *Determine the maximum likelihood estimate of p.*