

Nonlife Actuarial Models

Chapter 13

Model Evaluation and Selection

Learning Objectives

1. Graphical presentation and comparison
2. Misspecification tests and diagnostic checks
3. Kolmogorov-Smirnov test and Anderson-Darling test
4. Likelihood ratio test and chi-square goodness-of-fit test
5. Model selection and information criteria

13.1 Graphical Methods

- We denote $\hat{F}(\cdot)$ as an estimated df using nonparametric method and $F^*(\cdot)$ as a hypothesized df or parametrically estimated df.
- One way to assess if the assumption concerning the distribution is correct is to plot the estimated parametric df against the empirical df.
- If the distributional assumption is incorrect, we would expect the two plotted graphs to differ.

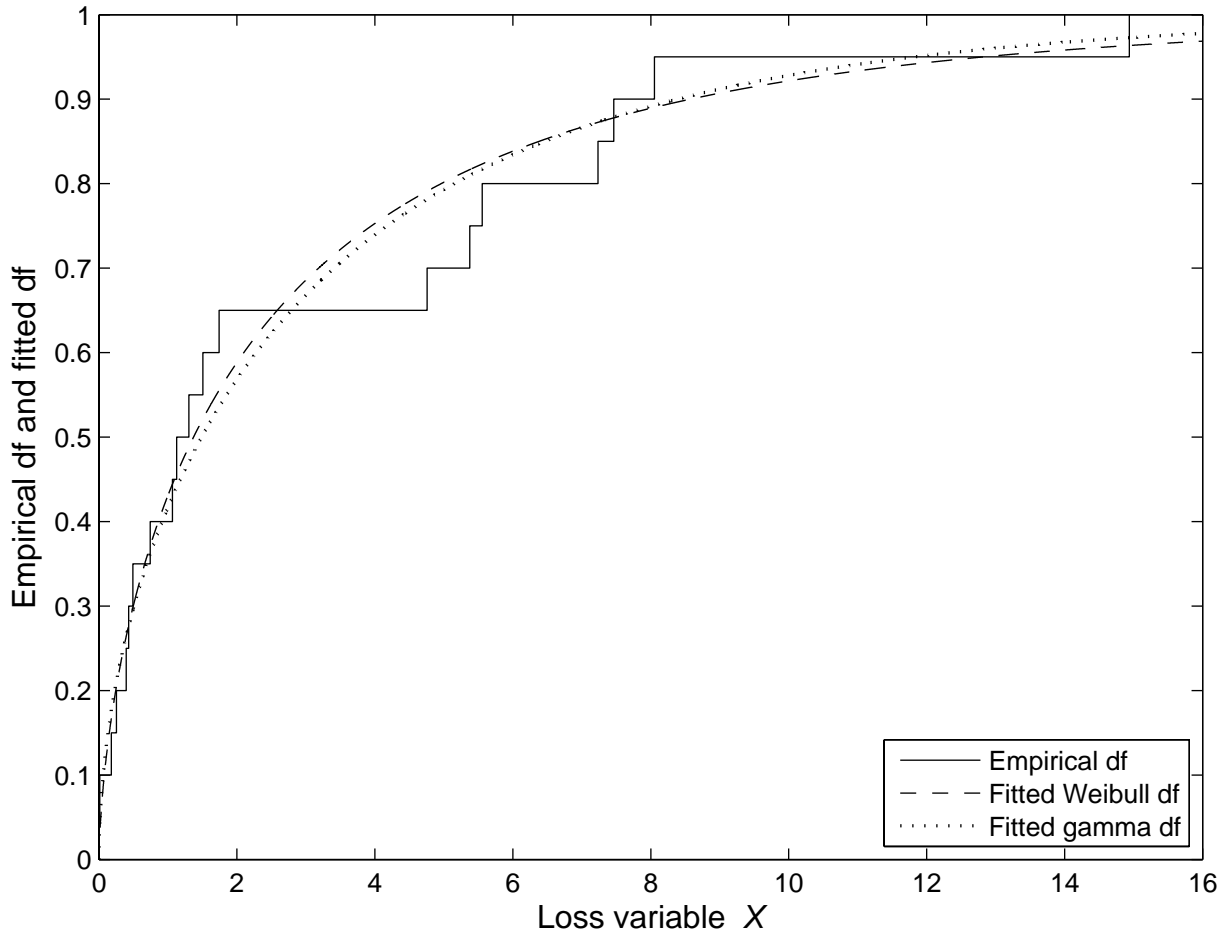
Example 13.1: A sample of 20 loss observations are as follows

0.003, 0.012, 0.180, 0.253, 0.394, 0.430, 0.491, 0.743, 1.066, 1.126,
1.303, 1.508, 1.740, 4.757, 5.376, 5.557, 7.236, 7.465, 8.054, 14.938.

Two parametric models are fitted to the data using the MLE, assuming that the underlying distribution is (a) $\mathcal{W}(\alpha, \lambda)$, and (b) $\mathcal{G}(\alpha, \beta)$. The fitted models are $\mathcal{W}(0.6548, 2.3989)$ and $\mathcal{G}(0.5257, 5.9569)$. Compare the empirical df against the df of the two estimated parametric models.

Solution: The plots of the empirical df and the estimated parametric df are given in Figure 13.1. It can be seen that both estimated parametric models fit the data quite well. Thus, from the df plots it is difficult to ascertain which is the preferred model. \square

- Another useful graphical device is the p - p plot.

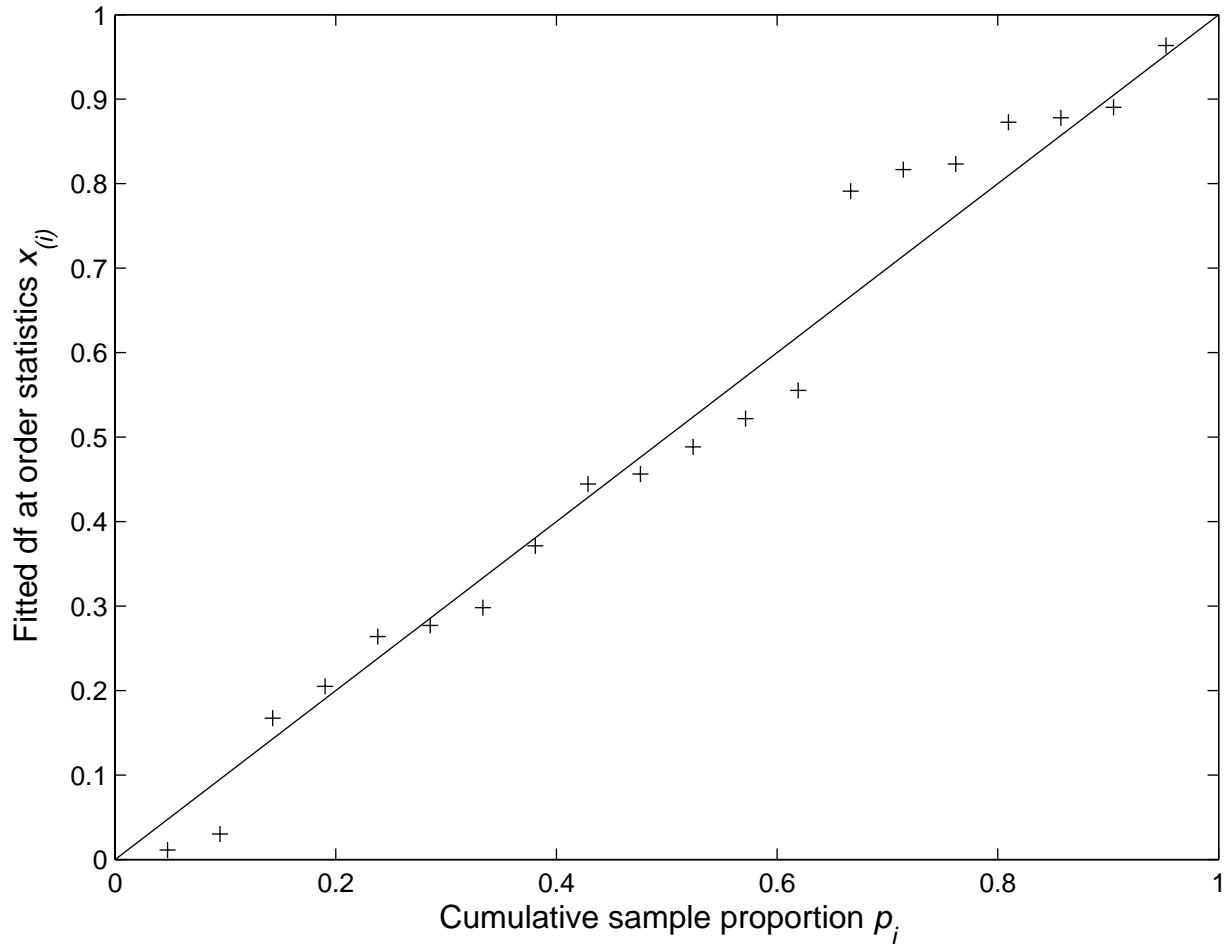


- We approximate the probability of having an observation less than or equal to $x_{(i)}$ using the sample data by the sample proportion $p_i = i/(n + 1)$.
- A plot of $F^*(x_{(i)})$ against p_i is called the p - p plot. If $F^*(\cdot)$ fits the data well, the p - p plot should approximately follow the 45 degree line.

Example 13.2: For the data and the fitted Weibull model in Example 13.1, assess the model using the p - p plot.

Solution: The p - p plot is presented in Figure 13.2. It can be seen that most points lie closely to the 45 degree line, apart from some deviations around $p_i = 0.7$. □

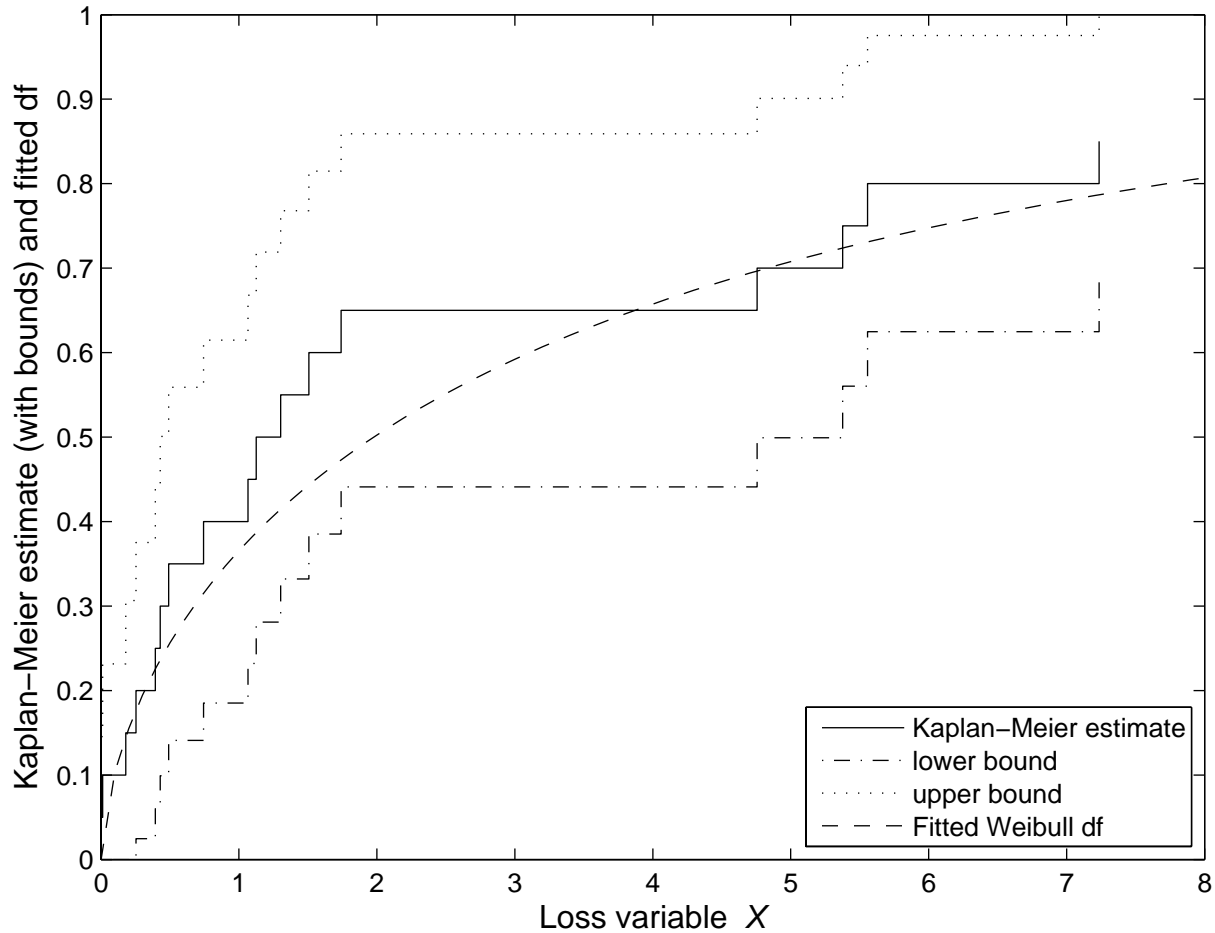
- Another graphical method equivalent to the p - p plot is the q - q plot.



- In a q - q plot, $F^{*-1}(p_i)$ is plotted against $x_{(i)}$. If $F^*(\cdot)$ fits the data well, the q - q plot should approximately follow a straight line.

Example 13.3: For the data in Example 13.1, assume that observations larger than 7.4 are censored. Compare the estimated df based on the MLE under the Weibull assumption against the Kaplan-Meier estimate.

Solution: In the data set three observations are larger than 7.4 and are censored. The plots of the Kaplan-Meier estimate and the estimated df using the MLE of the Weibull model are given in Figure 13.3. For the Kaplan-Meier estimate we also plot the lower and upper bounds of the 95% confidence interval estimates of the df. It can be seen that the estimated parametric df falls inside the band of the estimated Kaplan-Meier estimate. □



13.2 Misspecification Tests and Diagnostic Checks

13.2.1 Kolmogorov-Smirnov Test

- We specify a null hypothesis about the df of a continuous loss variable, and denote it by $F^*(\cdot)$. To examine if the data support the null hypothesis, we compare $F^*(\cdot)$ against the empirical df $\hat{F}(\cdot)$ and consider the statistic

$$\max_{x_{(1)} \leq x \leq x_{(n)}} | \hat{F}(x) - F^*(x) |, \quad (13.1)$$

where $x_{(1)}$ and $x_{(n)}$ are the minimum and maximum of the observations, respectively.

- As $\hat{F}(\cdot)$ is a right-continuous increasing step function and $F^*(\cdot)$ is also increasing we only need to compare the differences at the ob-

served data points, namely, at the order statistics $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$.

- The maximum may only occur at a jump point $x_{(i)}$ or immediately to the left of it. We now denote the statistic in expression (13.1) by D , which is called the **Kolmogorov-Smirnov statistic** and can be written as

$$D = \max_{i \in \{1, \dots, n\}} \left\{ \max \left\{ \left| \hat{F}(x_{(i-1)}) - F^*(x_{(i)}) \right|, \left| \hat{F}(x_{(i)}) - F^*(x_{(i)}) \right| \right\} \right\}, \quad (13.2)$$

where $\hat{F}(x_{(0)}) \equiv 0$.

- When we have complete individual observations,

$$\hat{F}(x_{(i)}) = \frac{i}{n}, \quad (13.3)$$

and D can be written as

$$D = \max_{i \in \{1, \dots, n\}} \left\{ \max \left\{ \left| \frac{i-1}{n} - F^*(x_{(i)}) \right|, \left| \frac{i}{n} - F^*(x_{(i)}) \right| \right\} \right\}. \quad (13.4)$$

- When $F^*(\cdot)$ is completely specified the critical values of D for some selected values of α are given as follows

Level of significance α	0.10	0.05	0.01
Critical value	$\frac{1.22}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.63}{\sqrt{n}}$

- The critical values above apply to all df, as long as they are completely specified.
- Any unknown parameters in the df have to be estimated for the computation of $F^*(x_{(i)})$. Then the critical values above will not apply.

Example 13.5: Compute the Kolmogorov-Smirnov statistics for the data in Example 13.1, with the estimated Weibull and gamma models as the hypothesized distributions.

Solution: We denote $F_1^*(x_{(j)})$ as the df of $\mathcal{W}(0.6548, 2.3989)$ evaluated at $x_{(j)}$, and $F_2^*(x_{(j)})$ as the df of $\mathcal{G}(0.5257, 5.9569)$ evaluated at $x_{(j)}$, which are the df of the estimated Weibull and gamma distributions in Example 13.1. We further denote

$$D_{ij} = \max \left\{ \left| \hat{F}(x_{(j-1)}) - F_i^*(x_{(j)}) \right|, \left| \hat{F}(x_{(j)}) - F_i^*(x_{(j)}) \right| \right\},$$

for $i = 1, 2$, and $j = 1, \dots, 20$. Note that

$$\hat{F}(x_{(j)}) = \frac{j}{20},$$

as we have complete individual data. The results are summarized in Table 13.1.

Table 13.1: Results for Example 13.5

$\hat{F}(x_{(j)})$	$F_1^*(x_{(j)})$	D_{1j}	$F_2^*(x_{(j)})$	D_{2j}
0.0500	0.0112	0.0388	0.0191	0.0309
0.1000	0.0301	0.0699	0.0425	0.0575
0.1500	0.1673	0.0673	0.1770	0.0770
0.2000	0.2050	0.0550	0.2112	0.0612
0.2500	0.2639	0.0639	0.2643	0.0643
0.3000	0.2771	0.0271	0.2761	0.0261
0.3500	0.2981	0.0519	0.2951	0.0549
0.4000	0.3714	0.0286	0.3617	0.0383
0.4500	0.4445	0.0445	0.4294	0.0294
0.5000	0.4563	0.0437	0.4405	0.0595
0.5500	0.4885	0.0615	0.4710	0.0790
0.6000	0.5218	0.0782	0.5030	0.0970
0.6500	0.5553	0.0947	0.5357	0.1143
0.7000	0.7910	0.1410	0.7813	0.1313
0.7500	0.8166	0.1166	0.8097	0.1097
0.8000	0.8233	0.0733	0.8172	0.0672
0.8500	0.8726	0.0726	0.8728	0.0728
0.9000	0.8779	0.0279	0.8789	0.0289
0.9500	0.8903	0.0597	0.8929	0.0571
1.0000	0.9636	0.0364	0.9728	0.0272

The Kolmogorov-Smirnov statistics D for the Weibull and gamma distributions are, 0.1410 and 0.1313, respectively, both occurring at $x_{(14)}$. The critical value of D at the level of significance of 10% is

$$\frac{1.22}{\sqrt{20}} = 0.2728,$$

which is larger than the computed D for both models. However, as the hypothesized df are estimated, the critical value has to be adjusted. Monte Carlo methods can be used to estimate the p -value of the tests, which will be discussed in Chapter 15. □

13.2.2 Anderson-Darling Test

- The **Anderson-Darling test** can be used to test for the null hypothesis that the variable of interest has the df $F^*(\cdot)$.
- Assuming we have complete and individual observations arranged in the order: $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$, the Anderson-Darling statistic, denoted by A^2 , is defined as

$$A^2 = -\frac{1}{n} \left[\sum_{j=1}^n (2j - 1) \left\{ \log [F^*(x_{(j)})] + \log [1 - F^*(x_{(n+1-j)})] \right\} \right] - n. \quad (13.6)$$

- If $F^*(\cdot)$ is fully specified with no unknown parameters, the critical values of A^2 are 1.933, 2.492 and 3.857 for level of significance $\alpha = 0.10, 0.05$ and 0.01 , respectively.

- Critical values are available for certain distributions with unknown parameters. Otherwise, they may be estimated using Monte Carlo methods.

13.2.3 Chi-Square Goodness-of-Fit Test

- The chi-square goodness-of-fit test is applicable to grouped data.
- Suppose the sample observations are classified into the intervals $(0, c_1], (c_1, c_2], \dots, (c_{k-1}, \infty)$, with n_j observations in $(c_{j-1}, c_j]$ such that $\sum_{j=1}^k n_j = n$.
- The expected number of observations in $(c_{j-1}, c_j]$ based on $F^*(\cdot)$ is

$$e_j = n [F^*(c_j) - F^*(c_{j-1})]. \quad (13.7)$$

- The chi-square goodness-of-fit statistic X^2 is defined as

$$X^2 = \sum_{j=1}^k \frac{(e_j - n_j)^2}{e_j} = \left(\sum_{j=1}^k \frac{n_j^2}{e_j} \right) - n. \quad (13.8)$$

- If $F^*(\cdot)$ is fully specified with no unknown parameters, X^2 is approximately distributed as a χ_{k-1}^2 .
- If the parameters of $F^*(\cdot)$ are estimated using the **multinomial MLE**, then the asymptotic distribution of X^2 is χ_{k-r-1}^2 , where r is the number of parameters estimated.
- To compute the multinomial MLE, we use the log-likelihood function

$$\log L(\theta; \mathbf{n}) = \sum_{j=1}^k n_j \log [F^*(\theta; c_j) - F^*(\theta; c_{j-1})], \quad (13.9)$$

where θ is the r -element parameter vector.

- The expected frequency e_j in interval $(c_{j-1}, c_j]$ is then given by

$$e_j = n [F^*(\hat{\theta}; c_j) - F^*(\hat{\theta}; c_{j-1})]. \quad (13.10)$$

Example 13.7: For the data in Example 13.4, compute the chi-square goodness-of-fit statistic assuming the loss distribution is (a) $\mathcal{G}(\alpha, \beta)$ and (b) $\mathcal{W}(\alpha, \lambda)$. In each case, estimate the parameters using the multinomial MLE.

Solution: The multinomial MLE for the $\mathcal{G}(\alpha, \beta)$ assumption can be found in Example 13.4. We estimate the Weibull case using the multinomial MLE method to obtain the distribution $\mathcal{W}(1.9176, 17.3222)$. For each of the fitted distributions the expected frequencies are computed and compared alongside the observed frequencies in each interval. The results are given in Table 13.2.

Table 13.2: Results for Example 13.7

$(c_{j-1}, c_j]$	$(0, 6]$	$(6, 12]$	$(12, 18]$	$(18, 24]$	$(24, 30]$	$(30, \infty)$
n_j	3	9	9	4	3	2
gamma	3.06	9.02	8.27	5.10	2.58	1.97
Weibull	3.68	8.02	8.07	5.60	2.92	1.71

Thus, for the gamma distribution, the chi-square statistic is

$$X^2 = \frac{(3)^2}{3.06} + \frac{(9)^2}{9.02} + \frac{(9)^2}{8.27} + \frac{(4)^2}{5.10} + \frac{(3)^2}{2.58} + \frac{(2)^2}{1.97} - 30 = 0.3694.$$

Similarly, the X^2 statistic for the fitted Weibull distribution is 0.8595. The degrees of freedom of the test statistics is $6 - 2 - 1 = 3$ for both fitted distributions, and the critical value of the test statistic at the 5% level of significance is $\chi_{3,0.95}^2 = 7.815$. Thus, both the gamma and Weibull assumptions cannot be rejected for the loss distribution. \square

13.2.4 Likelihood Ratio Test

- We have an unconstrained model defined by the **alternative hypothesis**.
- The model under the parametric constraints is the **null hypothesis**, which is **nested** within the alternative hypothesis.
- The constraints imposed on the parameter vector θ can be zero restrictions (i.e., some of the parameters in θ are zero), linear restrictions or nonlinear restrictions.
- Let $\hat{\theta}_U$ denote the unrestricted MLE under the alternative hypothesis, $\hat{\theta}_R$ denote the restricted MLE under the null hypothesis, and r denote the number of restrictions.

- The unrestricted and restricted maximized likelihoods are denoted by $L(\hat{\theta}_U, \mathbf{x})$ and $L(\hat{\theta}_R, \mathbf{x})$, respectively.
- The **likelihood ratio statistic**, denoted by ℓ , is defined as

$$\ell = 2 \log \left[\frac{L(\hat{\theta}_U, \mathbf{x})}{L(\hat{\theta}_R, \mathbf{x})} \right] = 2 \left[\log L(\hat{\theta}_U, \mathbf{x}) - \log L(\hat{\theta}_R, \mathbf{x}) \right]. \quad (13.11)$$

- When the null is true, $\ell \rightarrow \chi_r^2$ as $n \rightarrow \infty$.
- Reject the null hypothesis (i.e., conclude the restrictions do not hold) if $\ell > \chi_{r, 1-\alpha}^2$ at level of significance α , where $\chi_{r, 1-\alpha}^2$ is the 100(1− α)-percentile of the χ_r^2 distribution.

Example 13.8: For the data in Example 13.1, estimate the loss distribution assuming it is exponential. Test the exponential assumption against the gamma assumption using the likelihood ratio test.

Solution: The exponential distribution is the $\mathcal{G}(\alpha, \beta)$ distribution with $\alpha = 1$. For the alternative hypothesis where α is not restricted, the fitted distribution is $\mathcal{G}(0.5257, 5.9569)$ and the log-likelihood is

$$\log L(\hat{\theta}_U, \mathbf{x}) = -39.2017.$$

The MLE of λ for the $\mathcal{E}(\lambda)$ distribution is $1/\bar{x}$ (or the estimate of β in $\mathcal{G}(\alpha, \beta)$ with $\alpha = 1$ is \bar{x}). Now $\bar{x} = 3.1315$ and the maximized restricted log-likelihood is

$$\log L(\hat{\theta}_R, \mathbf{x}) = -42.8305.$$

Thus, the likelihood ratio statistic is

$$\ell = 2(42.8305 - 39.2017) = 7.2576.$$

As $\chi_{1,0.95}^2 = 3.841 < 7.2576$, the null hypothesis of $\alpha = 1$ is rejected at the 5% level of significance. □

13.3 Information Criteria for Model selection

- When two non-nested models are compared, the larger model with more parameters have the advantage of being able to fit the in-sample data with a more flexible function and thus possibly a larger log-likelihood.
- To compare models on more equal terms, **penalized log-likelihood** may be adopted.
- The **Akaike information criterion**, denoted by AIC, is defined as

$$\text{AIC} = \log L(\hat{\theta}; \mathbf{x}) - p, \quad (13.13)$$

where p is the number of estimated parameters in the model.

- Based on this approach the model with the **largest** AIC is selected.

- Consider two models \mathcal{M}_1 and \mathcal{M}_2 , so that $\mathcal{M}_1 \subset \mathcal{M}_2$. Using AIC, the probability of choosing \mathcal{M}_1 when it is true converges to a number that is strictly less than 1 when the sample size tends to infinity.
- In this sense, we say that the Akaike information criterion is *inconsistent*.
- The above problem can be corrected by imposing a different penalty on the log-likelihood.
- The **Schwarz information criterion**, also called the **Bayesian information criterion**, denoted by BIC, is defined as

$$\text{BIC} = \log L(\hat{\theta}; \mathbf{x}) - \frac{p}{2} \log n. \quad (13.13)$$

- Unlike the AIC, the BIC is consistent.

Example 13.10: For the data in Example 13.1, consider the following models: (a) $\mathcal{W}(\alpha, \lambda)$, (b) $\mathcal{W}(0.5, \lambda)$, (c) $\mathcal{G}(\alpha, \beta)$, and (d) $\mathcal{G}(1, \beta)$. Compare these models using AIC and BIC.

Solution: Table 13.3 summarizes the results.

Table 13.3: Results for Example 13.10

Model	$\log L(\hat{\theta}; \mathbf{x})$	AIC	BIC
$\mathcal{W}(\alpha, \lambda)$	-39.5315	-41.5315	-42.5272
$\mathcal{W}(0.5, \lambda)$	-40.5091	-41.5091	-42.0070
$\mathcal{G}(\alpha, \beta)$	-39.2017	-41.2017	-42.1974
$\mathcal{G}(1, \beta)$	-42.8305	-43.8305	-45.3284

AIC is maximized for the $\mathcal{G}(\alpha, \beta)$ model. BIC is maximized for the $\mathcal{W}(0.5, \lambda)$ model. Based on the BIC, $\mathcal{W}(0.5, \lambda)$ is the preferred model.