Stat 446
Advanced Short-Term Actuarial Modeling

Chapter 3: Parametric Model: Estimation & Selection

Outline

1. Maximum Likelihood Estimator
2. Variance and Interval Estimation
   - Information Matrix and Delta Method
3. Graphic Comparison
4. Hypothesis Tests
   - Kolmogorov-Smirnov Test
   - Chi-Square Goodness-of-Fit
   - Likelihood Ratio Test
5. Selecting a Model
   - Score-Based Approaches

Mathematical Statistics

- In this chapter we investigate the statistical methods to estimate and select parametric models.
- Brief review of mathematical statistics
  - Point estimation: estimation procedure to produce point estimates for parameters.
    - **Unbiased estimator** $\hat{\theta}$: $\mathbb{E}\left[\hat{\theta}\,\middle|\,\theta\right] = \theta$ for all θ.
    - **Mean-squared error (MSE)**: $\mathrm{MSE}_{\hat{\theta}} = \mathbb{E}\left[\left(\hat{\theta} - \theta\right)^2\,\middle|\,\theta\right]$
  - Interval estimation: construct a $100(1-\alpha)\%$ confidence interval for θ, i.e., $\Pr(L \leq \theta \leq U) \geq 1 - \alpha$.
    - Normal CI: $1 - \alpha \approx \Pr\left(-Z_{1-\alpha/2} \leq \frac{\theta - \hat{\theta}}{\sqrt{v(\hat{\theta})}} \leq Z_{1-\alpha/2}\right)$
  - Hypothesis test: *null* vs. *alternative*
    - Type I error and Type II error
    - Significance level and p value

Parametric Estimation

- How to estimate the parameters in a parametric model?
  - Moments matching
  - Percentile matching
  - Maximum likelihood estimation
  - Bayesian estimation
- The first two methods are easy to implement, but typically provide poor fitting results.
- **Maximum likelihood estimation**: maximize the probability/likelihood of observing the given data.
- **Bayesian estimation**: a *prior* distribution is assumed for the parameter, and further updated to a *posterior* distribution conditioning on the observed data.

Maximum Likelihood Estimator

Maximum Likelihood Estimator

Let the data set consist of n events/observations, $\{A_j\}_{j=1}^n$. The **likelihood function** and the **log-likelihood function** are defined as

$$L(\theta) = \prod_{j=1}^{n} \Pr(X_j \in A_j \,|\, \theta) \quad \& \quad l(\theta) = \ln L(\theta).$$

The **maximum likelihood estimator** (MLE) of θ maximizes $L(\theta)$ or $l(\theta)$.

- A_j can be a single point or an interval (e.g., for grouped or censored data).
- If A_j only contains a single point x_j, the probability mass (or density) function is used for $\Pr(X_j \in A_j \,|\, \theta)$.

Complete, Individual Data

- For **complete, individual data**, there is no truncation or censoring, and the value of each observation is recorded.
- The log-likelihood function can be written as

$$L(\theta) = \prod_{j=1}^{n} f_{X_j}(x_j \mid \theta) \quad \& \quad l(\theta) = \sum_{j=1}^{n} \ln f_{X_j}(x_j \mid \theta)$$

  where f_{X_j} is a pdf or pmf depending on the r.v.
- It is *not necessary* for each observation to come from the same distribution.
- The MLE is obtained by setting the first order partial derivatives to zero.
- Exercises: FAM-S Sample Qs 4, 88, 13, 37

Grouped or Censored Data

- For **complete, grouped data**, let n_j be the number of observations in $(c_{j-1}, c_j]$ for $j = 1, 2, \ldots, k$.
- The likelihood functions become

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

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

- If the data is **censored** by a limit u, i.e., n_c observations in (u, ∞), multiply the likelihood function by $[1 - F(u \,|\, \theta)]^{n_c}$.
- Exercises: FAM-S Sample Qs 2, 9, 49, 52, 56, & 89

Truncated Data

- For **left-truncated data**, any values below the deductible d are excluded.
- The probability of observing x_j, given the values below d are omitted, is a conditional probability/density $\dfrac{f(x_j \mid \theta)}{1 - F(d \mid \theta)}$.
- The likelihood functions become

$$L(\theta) = \prod_{j=1}^{n} \frac{f(x_j \mid \theta)}{1 - F(d \mid \theta)}$$

$$l(\theta) = \sum_{j=1}^{n} \ln\left[f(x_j \mid \theta)\right] - n \ln\left[1 - F(d \mid \theta)\right]$$

- Note that a data set may include *all* of the following cases: individual, grouped/censored, and truncated data.
- Exercises: FAM-S Sample Qs 11, 39, 48, 54, & ASTAM Sample Q17

MLE for $(a, b, 0)$ Class

- For complete data of sample size n, let n_k denote the number of observations with exactly k claims.
  - The sample mean $\overline{x}$ is

$$\overline{x} = \frac{\sum_{k=0}^{\infty} k\, n_k}{\sum_{k=0}^{\infty} n_k} = \frac{\sum_{k=0}^{\infty} k\, n_k}{n}$$

- The likelihood functions are

$$L(\theta) = \prod_{k=0}^{\infty} p_k^{n_k} \quad \& \quad l(\theta) = \sum_{k=0}^{\infty} n_k \ln p_k.$$

- The MLE of the mean of the $(a, b, 0)$ class is the sample mean:
  - Poisson: $\hat{\lambda} = \overline{x}$
  - Negative binomial: $\hat{r}\hat{\beta} = \overline{x}$ (i.e., $\frac{\partial}{\partial \beta} l(r, \beta) = 0$)
  - Binomial: $\hat{m}\hat{q} = \overline{x}$ (i.e., $\frac{\partial}{\partial q} l(m, q) = 0$)
- Exercises: FAM-S Sample Qs 7, 40, 53 & 90

Variance and Confidence Interval

- We are further interested in estimating the variance and confidence intervals of an MLE estimator.
- In a few cases, the variance of the maximum likelihood estimator is easy to obtain.
  - E.g., for an exponential r.v. with mean θ, the variance of the MLE $\hat{\theta}$ is

$$\hat{\theta} = \sum_{i=1}^{n} \frac{X_i}{n}, \quad \mathrm{Var}\left(\hat{\theta}\right) = \frac{\mathrm{Var}(X_i)}{n} = \frac{\sigma^2}{n}$$

- In general, we would have to approximate the variance using the *inverse* of the **Fisher's information** matrix.
- We will also investigate a quantity that is a function of the parameters, where the MLEs of these parameters are used. The **delta method** is used to estimate the mean and variance of this quantity.

Variance of Maximum Likelihood Estimator

- With some mild regularity conditions, the maximum likelihood estimators are *asymptotically multivariate normal* with the variance being the inverse of the information matrix, $[\mathbf{I}(\boldsymbol{\theta})]^{-1}$.
- The **Fisher's information** matrix:

$$\mathbf{I}(\boldsymbol{\theta})_{rs} = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta_r\partial\theta_s} l(\boldsymbol{\theta})\right], \qquad \mathbf{I}(\theta) = -\mathbb{E}\left[\frac{d^2}{d\theta^2} l(\theta)\right]$$

- If the sample has n iid observations, the information matrix is further simplified as

$$\mathbf{I}(\boldsymbol{\theta})_{rs} = -n\mathbb{E}\left[\frac{\partial^2}{\partial\theta_r\partial\theta_s} \ln f(X;\boldsymbol{\theta})\right],$$

$$\mathbf{I}(\theta) = -n\mathbb{E}\left[\frac{d^2}{d\theta^2} \ln f(X;\theta)\right]$$

Example: STAM Sample Q229

Example

A random sample of size n is drawn from a distribution with probability density function:

$$f(x) = \frac{\theta}{(\theta + x)^2}, \quad 0 < x < \infty, \quad \theta > 0$$

Calculate the asymptotic variance of the maximum likelihood estimator of θ. $[3\theta^2/n]$

(Exercise: STAM Sample Q14)

Example: ASTAM Sample Q10(a)-(c)

Example

For a portfolio of short-term insurance contracts, claim amounts are modeled with a density function

$$f(x;\lambda) = \lambda^2 x e^{-\lambda x}, \quad x > 0, \quad \lambda > 0$$

A random sample of n claim amounts, $x_1, x_2, \ldots, x_n$ is used to estimate the parameter λ.

(a) Derive and simplify as far as possible a formula for the maximum likelihood estimate of λ. [$\hat{\lambda} = 2/\overline{x}$]

(b) Use the moment generating function to show that $Y = 2n\lambda\overline{X}$ has a chi-square distribution, and state the degrees of freedom. [$4n$ degrees of freedom]

(c) Determine the Fisher information function, $I(\lambda)$, and use it to derive a formula for the asymptotic variance of $\hat{\lambda}$. [$I(\lambda) = 2n/\lambda^2$, $\mathrm{Var}(\hat{\lambda}) \approx \lambda^2/(2n)$]

Example: ASTAM Sample Q10(d)(e)

Example

(d) A random sample of 10 claim amounts resulted in a sample mean of $\overline{x} = 358$.

  (i) Use the results in (b) to calculate an exact 95% confidence interval for λ. [(0.00341, 0.00829)]

  (ii) Use the results in (c) to calculate an approximate (asymptotic) 95% confidence interval for λ. [(0.00314, 0.00804)]

(e) Comment briefly on the comparison of these confidence intervals. [the exact interval is asymmetric and better for $n = 10$]

(Exercise: ASTAM Sample Q6)

Delta Method

- Often, we need the variance of a function of parameters estimated by maximum likelihood.
- Let $\hat{\theta}$ be an estimator of θ (e.g., the MLE) and g a differentiable function. The **delta method** states that

$$\mathrm{Var}\left(g(\hat{\theta})\right) \approx \left(g'(\theta)\right)^2 \mathrm{Var}(\hat{\theta}),$$

  where in practice the derivative is evaluated at $\hat{\theta}$.
- With two estimated parameters $(\hat{\theta}_1, \hat{\theta}_2)$,

$$\mathrm{Var}\left(g(\hat{\theta}_1, \hat{\theta}_2)\right) \approx \mathrm{Var}(\hat{\theta}_1)\left(\frac{\partial g}{\partial \theta_1}\right)^2 + 2\,\mathrm{Cov}(\hat{\theta}_1, \hat{\theta}_2)\frac{\partial g}{\partial \theta_1}\frac{\partial g}{\partial \theta_2} + \mathrm{Var}(\hat{\theta}_2)\left(\frac{\partial g}{\partial \theta_2}\right)^2$$

- In general, $\mathrm{Var}\left(g(\hat{\boldsymbol{\theta}})\right) \approx (\nabla g)^\top \boldsymbol{\Sigma}\,(\nabla g)$, where $\boldsymbol{\Sigma}$ is the covariance matrix of $\hat{\boldsymbol{\theta}}$.

Example: ASTAM Spring 2023 Q2(a)

Example

You are given the following claim frequency data for a line of business with 100 one-year policies.

# of Claims (k)	0	1	2	3	4	5+
Policies (n_k)	46	34	13	4	3	0

You are particularly interested in the probability that a policy will have no claims in a year, denoted p_0. You first fit a Poisson(λ) distribution to this data using Maximum Likelihood Estimation (MLE).

Example: ASTAM Spring 2023 Q2(a) (cont.)

Example

(a) You are given that the MLE of λ is $\hat{\lambda} = 0.84$.

(i) Show that the MLE of p_0 is $\hat{p}_0 = 0.43$ to the nearest 0.01. You should calculate the value to the nearest 0.001. [0.4317]

(ii) Use the delta method to construct an approximate 95% linear confidence interval for p_0 under this model. [(0.3542, 0.5093)]

(Exercises: ASTAM S23 Q2(b)(c), F23 Q4(a), S25 Q4(a)-(c), & Sample Q1(a), Q25)

Example: STAM Sample Qs 179 & 180

Example

The time to an accident follows an exponential distribution. A random sample of size two has a mean time of 6.
Let Y denote the mean of a new sample of size two.

(a) Calculate the maximum likelihood estimate of $\Pr(Y > 10)$. [0.1546]

(b) Calculate the delta method approximation of the variance of the maximum likelihood estimator of $F_Y(10)$. [0.0786]

(Exercises: ASTAM Sample Q13; S24 Q4 & S25 Q1(c))

Example: STAM Sample Q225

Example

You are given:

(i) Fifty claims have been observed from a lognormal distribution with unknown parameters μ and σ.

(ii) The maximum likelihood estimates are $\hat{\mu} = 6.84$ and $\hat{\sigma} = 1.49$.

(iii) The covariance matrix of $\hat{\mu}$ and $\hat{\sigma}$ is $\begin{bmatrix} 0.0444 & 0 \\ 0 & 0.0222 \end{bmatrix}$

(iv) The partial derivatives of the lognormal cumulative distribution function are: $\frac{\partial F}{\partial \mu} = -\frac{\phi(z)}{\sigma}$, $\frac{\partial F}{\partial \sigma} = -\frac{z\phi(z)}{\sigma}$

(v) An approximate 95% confidence interval for the probability that the next claim will be less than or equal to 5000 is $[L, U]$.

Calculate L. [0.795]

(Exercises: STAM Sample Q69 & ASTAM Sample Q19)

Non-normal Confidence Intervals

- The normal interval $\hat{\theta} \pm z_{1-\alpha/2}\sqrt{\widehat{\mathrm{Var}}(\hat{\theta})}$ is symmetric and can include impossible values (e.g., $\theta < 0$), especially for small samples.
- **Likelihood-based interval**: since $2\left[l(\hat{\theta}) - l(\theta)\right]$ is approximately chi-square with k degrees of freedom ($k =$ number of parameters), an approximate $100(1-\alpha)\%$ confidence region is

$$\left\{\theta : \; l(\theta) \geq l(\hat{\theta}) - c\right\}, \qquad c = \tfrac{1}{2}\,\chi^2_{k,\,1-\alpha}$$

For one parameter and 95% confidence, $c = 3.841/2 = 1.921$.

Non-normal Confidence Intervals (cont.)

- The endpoints of the likelihood-based interval are usually found numerically. The interval need not be symmetric, stays inside the parameter space, and transforms directly: the interval for $g(\theta)$ is g applied to the endpoints.
- **Log-transformed interval** for $\theta > 0$: apply the normal approximation to $\ln \hat{\theta}$, whose variance is about $\widehat{\mathrm{Var}}(\hat{\theta})/\hat{\theta}^2$ (delta method):

$$\left(\hat{\theta}/U,\ \hat{\theta}\,U\right), \qquad U = \exp\left(z_{1-\alpha/2}\,\frac{\sqrt{\widehat{\mathrm{Var}}(\hat{\theta})}}{\hat{\theta}}\right)$$

Example: Non-normal Confidence Intervals

Example

A random sample of 10 losses from an exponential distribution with mean θ has $\sum x_i = 5000$. Construct an approximate 95% confidence interval for θ using

(a) the normal approximation, [(190.1, 809.9)]

(b) the log-transformed interval, and [(269.0, 929.3)]

(c) the likelihood-based interval. [(285.1, 997.9)]

Model Selection

- We next introduce a variety of ways to evaluate a model and compare competing models.
  - Graphic comparison: parametric model vs. empirical distribution
  - Hypothesis tests
  - Score-based approach
- Note that whatever model we select, it is only an approximation of reality.
- Our goal is to determine a model that is "good enough".

Graphic Comparison

Empirical Distribution Function

Given complete, individual data, the **empirical distribution function** is defined as

$$F_n(x) = \frac{\text{number of observations} \leq x}{n}$$

- The most direct way to see how well the model and data match is to plot the respective density/distribution function $F^*(x)$ with the data $F_n(x)$.

Graphic Comparison: Example

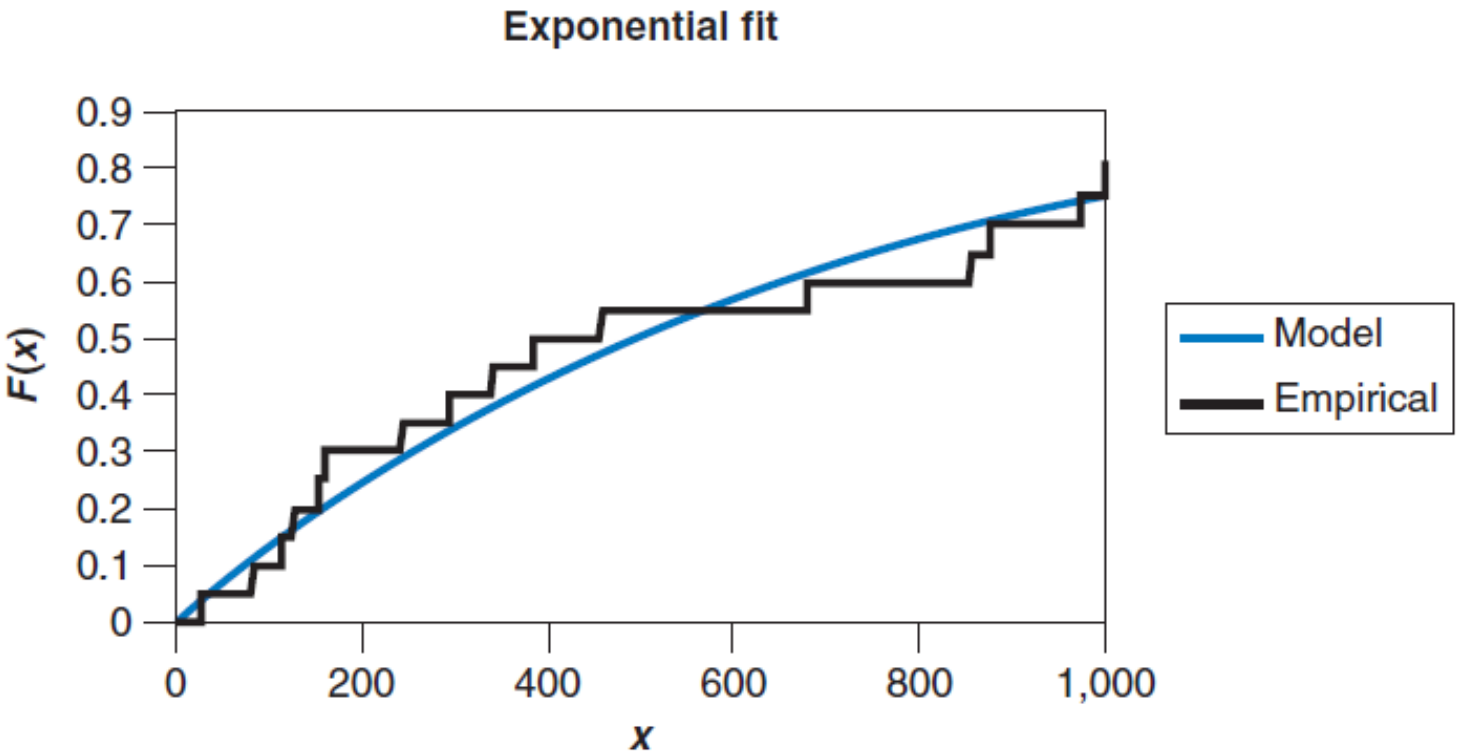


Figure 15.3 Model versus data cdf plot for Data Set B censored at 1,000.

Graphic Comparison: $D(x)$ Plot

- To amplify the difference between $F^*(x)$ and $F_n(x)$, we may plot

$$D(x) = F_n(x) - F^*(x)$$

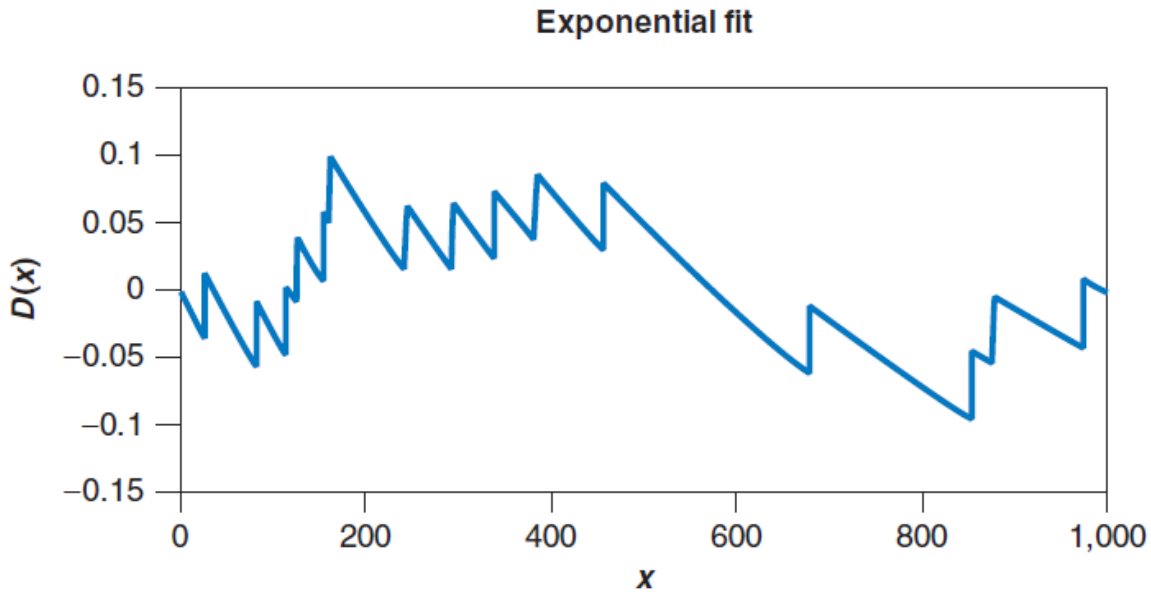


Figure 15.5 Model versus data $D(x)$ plot for Data Set B censored at 1,000.

Graphic Comparison: p–p Plot

- Alternatively, we highlight the differences with a p–p **(probability–probability) plot**: plot the points $(F_n(x_j), F^*(x_j))$ for the **ordered observations** $x_1 \leq \cdots \leq x_n$, where $F_n(x_j) = \frac{j}{n+1}$.
- If the model fits well, the points lie close to the 45-degree line. (A q–q plot is different: it compares quantiles rather than probabilities.)

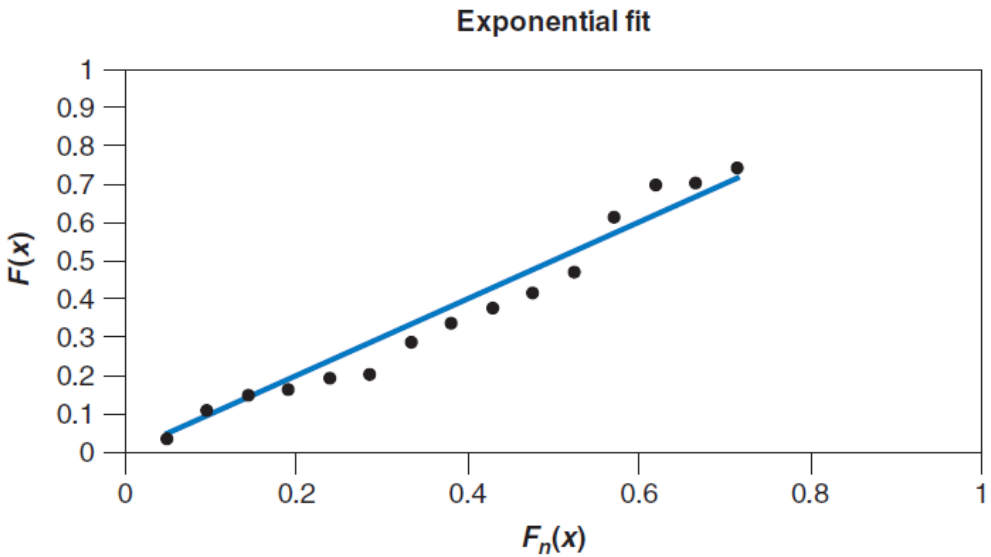
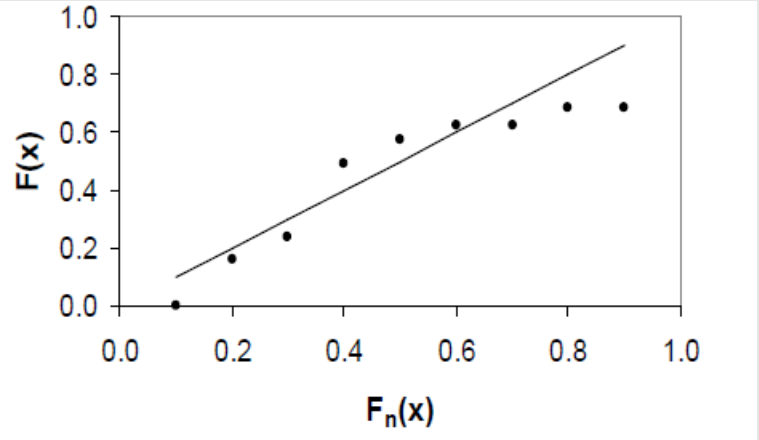


Figure 15.7 p–p plot for Data Set B censored at 1,000.

Example: STAM Sample Q176

Example

You are given the following p–p plot:



The plot is based on the sample: 1 2 3 15 30 50 51 99 100.
Determine the fitted model underlying the p–p plot. [A]

A. $F(x) = 1 - x^{-0.25}$, $x \geq 1$ B. $F(x) = x/(1+x)$, $x \geq 0$
C. Uniform on $[1, 100]$ D. $\mathrm{Exp}(10)$ E. $N(\mu = 40, \sigma = 40)$

(Exercise: STAM Sample Q59)

Hypothesis Tests

- While pictures provide an intuitive representation, sometimes it is best to replace them with mathematical demonstrations.
- We may evaluate a model through the following test:

  H_0: The data came from a population with the stated model.
  H_1: The data did not come from such a population.

  - Reject H_0 if Test Statistic $>$ Critical Value (based on α)
  - Reject H_0 if p value (based on Test Statistic) $< \alpha$

Critical Value and p Value

The **critical value** is the boundary of a rejection region, while the p **value** $(0 < p < 1)$ is the probability of the test statistic being as extreme as the observed one given that H_0 is true.

Kolmogorov-Smirnov Test

- For individual data, the **Kolmogorov-Smirnov statistic** D is defined as

$$D = \max_{t \le x \le u} |F_n(x) - F^*(x)|$$

  where t is the left truncation point and u is the right censoring point.
- For $x_1 \le x_2 \le \cdots \le x_n$, we only need to compare the values at x_j and x_j^-.
  - Step I: For each x_j, take the maximum of $|F^*(x_j) - F_n(x_j)|$ and $\left|F^*(x_j) - F_n(x_j^-)\right|$.
  - Step II: Find the maximum considering all j.
- The critical values will be given in exams.

Example: STAM Sample Q172

Example

You are given:

1. A random sample of five observations from a population is:
   0.2 0.7 0.9 1.1 1.3
2. You use the Kolmogorov-Smirnov test for testing the null hypothesis, H_0, that the probability density function for the population is:
$$f(x) = \frac{4}{(1+x)^5}, \quad x > 0$$

Example: STAM Sample Q172 (cont.)

Example

3. Critical values for the Kolmogorov-Smirnov test are:

Level of Significance	0.10	0.05	0.025	0.01
Critical Value	$\frac{1.22}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.48}{\sqrt{n}}$	$\frac{1.63}{\sqrt{n}}$

Determine the result of the test. [$D = 0.680$: reject at 0.025, not at 0.01]

(Exercises: STAM Sample Qs 40 & 160)

Example: ASTAM Fall 2023 Q3(a)

Example

An actuary is analyzing a sample of 1000 claims, as part of a review of pricing assumptions. Claims are assumed to be independent and identically distributed. Let x_i denote the amount of the ith claim in the sample. You are given that the sample mean is $\overline{x} = 1015.92$.

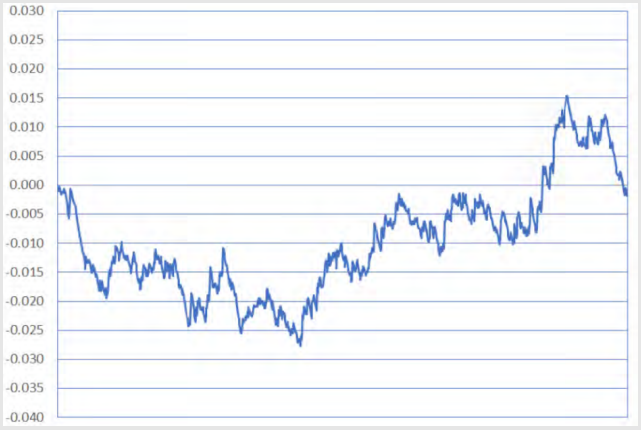
(a) The actuary first fits an exponential distribution to the claims data using maximum likelihood estimation. You are given that the maximum likelihood estimate of the exponential distribution parameter θ is $\hat{\theta} = \overline{x}$.

(i) Show that the maximum value of the log-likelihood function under the exponential distribution is -7925 to the nearest 5. You should calculate the value to the nearest 0.1. [-7923.55]

(ii) The actuary creates the following plot of the difference between the empirical distribution function and the fitted exponential distribution function for each data point, where the observations are sorted from low to high.

Example: ASTAM Fall 2023 Q3(a) (cont.)

Example



(a) (iii) You are given that the critical value for the Kolmogorov-Smirnov test at the 5% significance level is $1.36/\sqrt{n}$.
Conduct a Kolmogorov-Smirnov goodness of fit test at the 5% significance level. You should state clearly your conclusion.
[$D \approx 0.0275 < 0.043$: do not reject]

(Exercises: ASTAM F23 Q3(b), S25 Q1(b) & Sample Q1(c))

Chi-Square Test Statistic

- Y follows a **chi-square** (χ^2) distribution with ν degrees of freedom if $Y = \sum_{i=1}^{\nu} Z_i^2$, where Z_i is standard normal. Also, $Y \sim \mathrm{Gamma}(\nu/2, 2)$.
- The chi-square test allows grouping data into k intervals (by arbitrarily choosing $k-1$ breaking points c_j).
- The chi-square test statistic is

$$\chi^2 = \sum_{j=1}^{k} \frac{n(p_{nj} - \hat{p}_j)^2}{\hat{p}_j} = \sum_{j=1}^{k} \frac{(O_j - E_j)^2}{E_j} = \sum_{j=1}^{k} \frac{O_j^2}{E_j} - n$$

  - $\hat{p}_j = F^*(c_j) - F^*(c_{j-1})$ and $p_{nj} = F_n(c_j) - F_n(c_{j-1})$.
  - E_j: the *expected* number of observations in the interval
  - O_j: the *actual* number of observations in the interval

Degrees of Freedom

- The critical value of the chi-square statistic depends on the degrees of freedom.
- The degrees of freedom is related to several factors:
  1. If n observations are divided into k groups, the degrees of freedom is
  $$\mu = k - 1 - r$$
     - Subtract 1 since the counts in the k groups must add up to n.
     - Subtract r if r parameters are **estimated from the data** (note that $r = 0$ if the parameters are given in advance or estimated from another data source).
  2. If the data is naturally broken down by period, the degrees of freedom is
  $$\mu = k - r$$

Chi-Square Test: Grouping the Data

- The chi-square approximation needs enough expected observations in each group. A common rule (used on SOA exams) is that every group should have $E_j \geq 5$.
- If a group has $E_j < 5$, combine it with a neighboring group (usually in the tail) and recompute O_j and E_j for the combined group. Use the largest number of groups that satisfies the rule.
- The degrees of freedom use the number of groups *after* combining.
- Example: Poisson counts grouped as 0, 1, 2, and 3+ with $E_{3+} = 3.8$. Combine the last two groups into "2 or more," so $k = 3$; with λ estimated from the data, $\mu = 3 - 1 - 1 = 1$.

Example: STAM Sample Q201

Example

You test the hypothesis that a given set of data comes from a known distribution with distribution function $F(x)$. The following data were collected:

Interval	$F(x_i)$	Number of Observations
$x < 2$	0.035	5
$2 \leq x < 5$	0.130	42
$5 \leq x < 7$	0.630	137
$7 \leq x < 8$	0.830	66
$8 \leq x$	1.000	50
Total		300

where x_i is the upper endpoint of each interval. You test the hypothesis using the chi-square goodness-of-fit test.
Determine the result of the test. [$\chi^2 = 11.02$, 4 df: reject at 0.05, not at 0.025]

(Exercises: STAM Sample Qs 13, 71, 140)

Example: STAM Sample Q222

Example

1000 workers insured under a workers compensation policy were observed for one year. The number of work days missed is given below:

Number of Days of Work Missed	Number of Workers
0	818
1	153
2	25
3 or more	4
Total	1000
Total Number of Days Missed	230

Example: STAM Sample Q222 (cont.)

Example

The chi-square goodness-of-fit test is used to test the hypothesis that the number of work days missed follows a Poisson distribution where:

(i) The Poisson parameter is estimated by the average number of work days missed.

(ii) Any interval in which the expected number is less than one is combined with the previous interval.

Determine the results of the test. [$\chi^2 = 9.36$, 2 df: reject at 0.01, not at 0.005]

(Exercises: ASTAM Sample Qs 1(b), 11 & S25 Q1(a))

Likelihood Ratio Test

- An alternative test to the previous "yes or no" tests is

  H_0: The data came from a population with distribution A.
  H_1: The data came from a population with distribution B.

  Model A must be nested in model B, i.e., A is a special case of B (e.g., A: Weibull with $\tau = 2$; B: Weibull with τ unrestricted). The test compares the fit of the two models.
- The **likelihood ratio test** is performed as follows:
  - Let $L_0(\theta_0)$ be the maximum likelihood based on only values of the parameters within the null hypothesis.
  - Let $L_1(\theta_1)$ be the maximum likelihood based on all possible values of the parameters from the alternative hypothesis.
  - The test statistic is

$$T = 2 \ln (L_1/L_0) = 2 (\ln L_1 - \ln L_0)$$

  - T has a chi-square distribution with degrees of freedom μ equal to the *difference* of the number of free parameters from H_1 and H_0.

Example: STAM Sample Q235

Example

You are given:

(i) A random sample of losses from a Weibull distribution is:

595 700 789 799 1109

(ii) At the maximum likelihood estimates of θ and τ, $\sum \ln [f(x_i)] = -33.05$.

(iii) When $\tau = 2$, the maximum likelihood estimate of θ is 816.7.

(iv) You use the likelihood ratio test to test the hypothesis

$$H_0 : \tau = 2 \quad \text{vs.} \quad H_1 : \tau \neq 2$$

Determine the result of the test. [$T = 4.46$, 1 df: reject at 0.05, not at 0.025]

(Exercise: ASTAM F24 Q1)

Selecting a Model

- To select models, we may use **judgement-based approaches** and **score-based approaches**.
- Judgement-based approaches
  - Graphs, such as p–p plots.
  - Success of similar models in similar situations.
  - Situations may completely determine the distribution.
  - Judgement is required when algorithmic approaches do not agree.
- Score-based approaches
  1. Lowest Kolmogorov-Smirnov test statistic
  2. Lowest chi-square test statistic
  3. Highest p-value for chi-square test
  4. Highest value of the likelihood function at its maximum

Score-Based: Penalty for Parameters

- Potential issue: no penalty for adding additional parameters
  - All but Approach 3 tend to favor a more complex model (with more parameters).
  - The *p*-value for the chi-square test is immune from this issue due to the auto-adjustment of the degrees of freedom.
- Compare models with different numbers of parameters:
  - Use the likelihood ratio test when one model is a special case of another.
  - In general, we may penalize the loglikelihood for parameters.
    - **Schwarz Bayesian Criterion (SBC)**: for r parameters and sample size n, $\mathrm{SBC} = l - \frac{r}{2} \ln n$
    - **Akaike Information Criterion (AIC)**: $\mathrm{AIC} = l - r$
    - Choose the model with the **largest** SBC or AIC. (Some texts use -2 times these values; on that scale, the smallest is best.)

Example: STAM Sample Qs 306 & 307

Example

Five models are fitted to a sample of $n = 260$ observations with the following results:

Model	Number of Parameters	Loglikelihood
I	1	−414
II	2	−412
III	3	−411
IV	4	−409
V	6	−409

(a) Determine the model favored by the Schwarz Bayesian criterion. [Model I]

(b) Determine the model favored by the Akaike Information criterion. [Model IV]

Example: ASTAM Fall 2023 Q3(c)-(e)

Example

In part (a), the exponential fit to the $n = 1000$ claims had maximized loglikelihood -7923.55; in part (b), a gamma fit gives $\hat{\alpha} = 0.92463$, $\hat{\theta} = 1098.73$, and maximized loglikelihood -7921.5.

(c) State with reasons which of the two distributions is preferred based on each of the following criteria:

- (i) The Akaike Information Criterion (AIC). [gamma]
- (ii) The Schwarz Bayes Criterion (SBC/BIC). [exponential]
- (iii) The likelihood ratio test at the 5% significance level. [$T = 4.10 > 3.84$: gamma]

(d) The actuary is concerned about the risk of claims over 6000.

- (i) Calculate the probability that a randomly selected claim exceeds 6000 using the fitted exponential distribution. [0.00272]
- (ii) Calculate the probability that a randomly selected claim exceeds 6000 using the fitted gamma distribution. [0.00352]

Example: ASTAM Fall 2023 Q3(c)-(e) (cont.)

Example

(e) For the task of determining margins for solvency in reserve calculations, it is important to adequately reflect the tail risk of losses. State with reasons which of the two distributions you would recommend for this task. [gamma]

(Exercises: ASTAM S23 Q2(d), F25 Q2; Sample Qs 1(d), 12, 18 & 24)