START
Selected Topics in Assurance Related Technologies

Kolmogorov-Simirnov: A Goodness of Fit Test for Small Samples

Table of Contents

Introduction

Most statistical methods assume an underlying distribution in the derivation of their results. However, when we assume that our data follow a specific distribution we are taking a serious risk. If such distribution does not hold, then the results obtained may be invalid. For example, the confidence levels of the confidence intervals (CI) or hypotheses tests implemented [1, 2, and 5] may be completely off. Consequences of miss specifying the distribution may prove very costly. One way to deal with this problem is checking the distribution assumptions carefully.

There are two main approaches to checking distribution assumptions [3, 4]. One is via empirical procedures, which are easy to understand and implement and are based on intuitive and graphical properties of the distribution that we want to assess. Empirical procedures can be used to check and validate distribution assumptions. Several of them have been discussed at length in other RAC START sheets [6, 7, and 8].

There are also other, more formal, procedures to assess the underlying distribution of a data set. These are the Goodness of Fit (GoF) tests, based on statistical theory. They are numerically convoluted and usually require specific software to aid the user through their lengthy calculations. But their results are quantifiable and more reliable than those from the empirical procedure [9] are. Here, we are interested in those theoretical GoF procedures, specialized for small samples. Among them, the Anderson-Darling (AD) and the

Kolmogorov-Smirnov (KS) tests stand out. This START sheet discusses the latter of the two; the former (AD) is discussed in [10].

In what follows we overview some issues associated with the implementation of the KS GoF test, especially when assessing the Exponential, Weibull, Normal and Lognormal distribution assumptions. For, these distributions are widely used in quality and reliability (RMQA) work. We first overview some theoretical considerations that will help us better understand (and use) the underlying statistical theory behind these GoF tests. Then, we develop several numerical and graphical examples that illustrate how to implement and interpret the GoF tests for fitting several distributions.

Some Statistical Background

Establishing the underlying distribution of a data set (or random variable) is crucial for the correct implementation of some statistical procedures. For example both, the small sample t test and CI, for the population mean, require that the distribution of the underlying population is Normal. Therefore, we first need to establish (via GoF tests) such Normality before we can correctly implement these statistical procedures.

GoF tests are essentially based on either of two distribution elements: the cumulative distribution function (CDF) or the probability density function (PDF). The Chi-Square test is based on the PDF. Both, the AD and KS GoF tests, use the cumulative distribution function (CDF) approach and therefore belong to the class of "distance tests".

We have selected the AD and KS, among the several distance tests, for two reasons. First, they are among the best distance tests for small samples (and they can also be used for large samples). Secondly, both AD and KS have computerized versions in various statistical packages and, because of this, they are widely used in practice. In this START sheet, we will demonstrate how to use the KS test with the Minitab software.

To implement distance tests, we follow a well-defined series of steps. First, we assume a pre-specified distribution (e.g., Normal). Then, we estimate the distribution parameters (e.g.,

mean and variance) from the data or obtain them from prior experience. Such a process yields a distribution hypothesis, also called the null hypothesis (or H_0), with several parts that must be jointly true. The negation of the assumed distribution (or its parameters) is the alternative hypothesis (also called H_1). We test the assumed (hypothesized) distribution using the data set. Finally, H_0 is rejected whenever any one of the several elements composing the null hypothesis H_0 is not supported by the data.

In the distance tests, when the assumed distribution is correct, the theoretical (assumed) CDF (denoted F_0) closely follows the empirical, step function CDF (denoted F_n). This is conceptually illustrated in Figure 1. There, the data are given as an ordered sample $\{X_1 \leq X_2 \leq ... \leq X_n\}$ and the assumed (H_0) theoretical distribution has a CDF, $F_0(x)$. Then we obtain the corresponding GoF test statistic values. Finally, we compare the theoretical and empirical results. If they agree (probabilistically) then the data supports the assumed distribution. If they do not, the distribution assumption is rejected.
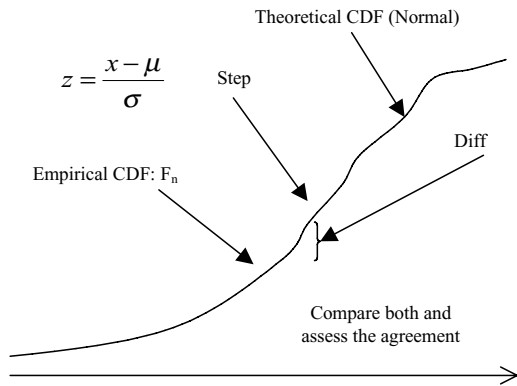

Figure 1. Distance Goodness of Fit Test Conceptual Approach

There is, however, an important caveat to make. Theoretically, distance tests require the knowledge of the assumed distribution parameters. These are seldom known in practice. Therefore, adaptive procedures are used to circumvent this problem, when implementing GoF tests in the real world (e.g., see [4], Chapter 7). This is a major drawback of the KS GoF test, which otherwise is very powerful and general. In the examples we develop here some of these adaptive procedures will be used.

Under a Normal Assumption, F_0 is normal (mu, sigma).

Fitting a Normal Distribution Using the Kolmogorov-Simirnov GoF Test

The KS GoF test is widely used in practice. It is very versatile (any continuous distribution can be fit with it), works well with small samples and is included in statistical software packages. In this section we develop an example using KS for testing Normality. In the next section we develop another one for test-

ing the Weibull assumption. Finally, we develop a counter example, where the assumption fails the test.

First, we use real data from Section 8.3.7 of MIL-HDBK-17 (1E), Composite Materials. The data set (Table 1) contains six tensile strengths, drawn from the same population.

Table 1. Data for the KS GoF Test

338.7	308.5	317.7	313.1	322.7	294.2

We obtain the descriptive statistics (Table 2) yielding point estimations of the assumed Normal distribution unknown parameters: sample mean and standard deviation.

Table 2. Descriptive Statistics of the Data in Table 1

Variable	N	Mean	Median
Data Set	6	315.82	315.40

There are several computational methods for the KS [3, 4]. Here, we follow the KS test statistic version in [4] and a well-defined scheme. First, sort the data. Then establish the assumed distribution (null hypothesis) and estimate its parameters. Then, obtain both the theoretical (assumed CDF) distribution (F_0) as well as the empirical (F_n) at each data point. Since KS is a distance test, we need to find the maximum distance $|F_0 - F_n|$ between the theoretical and empirical distributions. Its two basic functions are defined as:

$$F_0(X_i) = P_0(X \leq X_i) = CDF(X_i);$$

That is $F_0(X_i)$ is the assumed cumulative distribution function evaluated at Xi and $F_n(X_i)$ is the empirical distribution function obtained by the proportion of the data smaller than X_i in the data set size n.

$$F_n(X_i) = \frac{\text{\# of X's} \leq X_i}{n} = \frac{i}{n}; i = 1, ..., n$$

Then, define: $D+ = F_n - F_0$ and $D- = F_0 - F_{n-1}$ for every data point X_i. The KS statistic is:

$$D = \text{Maximum of all } D+ \text{ and } D- (\geq 0); \text{ for } i = 1, ..., n \quad (1)$$

The KS logic is as follows: if the maximum departure between the assumed (CDF) and empirical distributions is small, then the assumed CDF will likely be correct. But if this discrepancy is "large" then the assumed F_0 is likely not the underlying data distribution.

As an example, we calculate the values for the smallest data point (294.2) in Table 1.

$$F_0(294.2) = P_{\mu=315.82, \sigma=14.85}(294.2) =$$

$$\text{Normal}\left(\frac{294.2 - 315.82}{14.88}\right) = 0.0727$$

Which is the assumed normal probability evaluated at 294.2.

$$F_n(294.2) = \frac{1}{n} = \frac{1}{6} = 0.167; \quad F_{n-1}(294.2) = \frac{1-1}{n} = \frac{0}{6} = 0$$

Which is the empirical probability at 294.2.

$$D+(294.2) = F_n(294.2) - F_0(294.2) = 0.167 - 0.0727 = 0.0939$$

$$D-(294.2) = F_0(294.2) - F_{n-1}(294.2) = 0.0727 - 0 = 0.0727$$

The resulting values are shown in Table 3. The two columns of KSINDEX denote F_n and F_{n-1}, respectively. Column NormProb denotes the evaluation of the data using the assumed (Normal) distribution F_0 and the last two columns are statistics D+ and D. The result of the KS GoF test statistic (1) for this data set and assumption is:

$$D = \text{Maximum } (D+, D-) = 0.1549$$

KS tables of critical values (CV) can be found in several texts [3, 4]. These CV are valid when the distribution parameters are known. When the parameters are estimated from the data, these CV are only approximate. One adaptive procedure uses a CV for

Table 3. Intermediate Values for the KS GoF Test for Normality

Row	DataSet	F_0	F_n	F_{n-1}	D+=Fn-Fo	D-=Fo-Fn-1
1	294.2	0.072711	0.16667	0.000000	0.093955	0.072711
2	308.5	0.311031	0.33333	0.166667	0.022302	0.144365
3	313.1	0.427334	0.50000	0.333333	0.072666	0.094001
4	317.7	0.550371	0.66667	0.500000	0.116295	0.050371
5	322.7	0.678425	0.83333	0.666667	0.154908	0.011759
6	338.7	0.938310	1.00000	0.833333	0.061690	0.104977
				Max:	0.154908	0.144365

an error α', four times the a for which we are testing. Critical values are given in Table 4.

Table 4. KS CV for α = 0.05 and α' = 0.2 and n = 4(1)10

Size "n"	CV (0.20)	CV (0.05)
4	0.494	0.624
5	0.446	0.564
6	0.411	0.521
7	0.381	0.486
8	0.358	0.457
9	0.339	0.432
10	0.322	0.411
For n>10, use the approximation C.V. (0.05) = 1.36/√n C.V. (0.20) = 1.07/√n Approximation Error < 0.02		

The KS statistic value (D = 0.155) is less than the corresponding table value (0.521) for n = 6 and an error α = 0.05. If the true Normal parameters were used, the results for a would allow us to assume the null hypothesis (H_0) that the distribution of the population is Normal (315.8; 14.9). Since the two Normal parameters were estimated from the data, we must use an adaptive test procedure. We test at level α = 0.05 but compare the KS statistic with the CV at level α' = 4*α = 4*0.05 = 0.2. Hence, the adaptive KS CV for this case would be 0.411, which is still larger than the KS statistic value of 0.155.

There are software versions of the KS GoF test available. We present, in Figure 2, the Minitab graphical version. Verify how D+, D- and D (KS Statistics) coincide with the hand calculations

in Table 3. In addition, the Minitab version of the KS test provides the approximate p-value = 0.15 > α = 0.05 (also leading us to assume Normality of the data). Finally, the KS graph from the Minitab program shows a linear trend. This fact lends additional support to the Normality assumption.
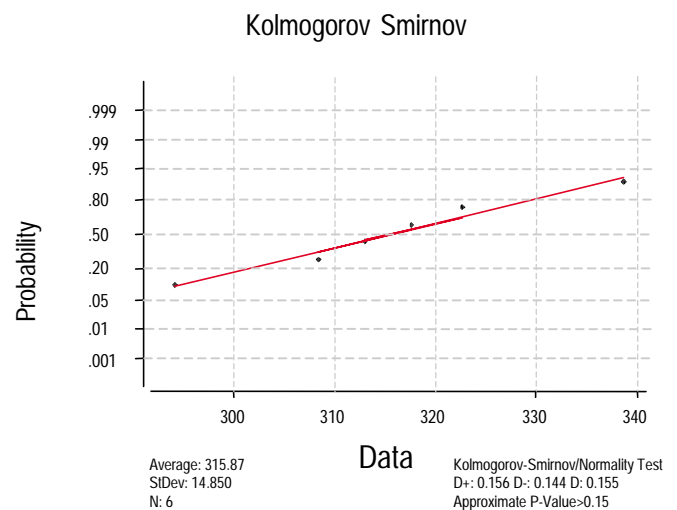

Kolmogorov Smirnov

Average: 315.87
StDev: 14.850
N: 6

Kolmogorov-Smirnov/Normality Test
D+: 0.156 D-: 0.144 D: 0.155
Approximate P-Value>0.15

Figure 2. Computer (Minitab) Version of the KS GoF Test for Normality

Hence, since all results coincide, we can assume that the data comes from a Normal (μ = 315.8; σ = 14.9) population. The stepwise KS GoF test procedure is in Table 5.

Table 5. Step-by-Step Summary of the KS GoF Test for

•   Establish the (H_0) assumed distribution: Normal (μ; σ)
•   Estimate the Normal parameters: $\mu = 315.8$; $\sigma = 14.9$
•   Sort the data in ascending order (Col. 1, Table 3)
•   Obtain the Theoretical distribution (Col. 2)
•   Obtain the Empirical distributions (Cols. 3 & 4)
•   Obtain D+ and D- (Cols. 5 & 6)
•   Obtain the KS Statistic: D = Maximum (D+, D-) = 0.155
•   Obtain the KS table critical value CV(α') = 0.411 (Adaptive)
•   Since KS < CV assume the data come from a Normal ($\mu = 315.8$; $\sigma = 14.9$)
•   When available, use the computer software and the p-value

Normality

Finally, if we want to fit a Lognormal distribution, we first take the logarithm of the data and then implement the above KS GoF procedure on these transformed data. For, if the original data is Lognormal then its logarithm is Normally distributed and we can use KS to test for Lognormality.

Fitting a Weibull Distribution Using the Kolmogorov-Simirnov GoF Test

We now develop an example of testing for the Weibull assumption, using the data in Table 6. These data consist of six measurements, drawn from the same Weibull ($\alpha = 10$; $\beta = 2$) population. In our example, however, parameters are unknown and estimated.

Table 6. Data for Testing the Weibull Assumption

11.7216	10.4286	8.0204	7.5778	1.4298	4.1154

We obtain the descriptive statistics (Table 7) and, using graphical methods in [11], also the point estimations of the assumed Weibull parameters: shape $\beta = 1.3$ and scale $\alpha = 8.7$.

Using the KS statistic (1) and following the procedures already discussed in the previous section, we obtain the intermediate KS results shown in Table 8.

Table 7. Descriptive Statistics

Variable	N	Mean	Median	StDev	Min	Max	Q1	Q2
DataSet	6	7.22	7.80	3.86	1.43	11.72	3.44	10.75

Table 8. Intermediate Values for the KS GoF Test for Weibull

Row	DataSet	F_0	F_n	F_{n-1}	D+=Fn-Fo	D-=Fo-Fn-1
1	1.430	0.091176	0.16667	0.000000	0.075491	0.091176
2	4.115	0.314692	0.33333	0.166667	0.018641	0.148025
3	7.578	0.566413	0.50000	0.333333	-0.066413	0.233079
4	8.020	0.593296	0.66667	0.500000	0.073371	0.093296
5	10.429	0.717949	0.83333	0.666667	0.115384	0.051282
6	11.722	0.770846	1.00000	0.833333	0.229154	-0.062488
				Max:	0.229154	0.233079

The step-by-step procedure (Table 9) shows how the KS GoF test statistic (1) value (0.23) is smaller than the KS table critical values (0.52 or 0.411) for $\alpha = 0.05$ and a sample of size n = 6. Based on these results, we do not reject the hypothesis that the population from where these data were obtained, is distributed Weibull ($\alpha = 8.7$; $\beta = 1.3$). Again, recall that the KS test is theoretically used with the true (but usually unknown) distribution parameters and not their estimations. Therefore, when implementing KS with parameters estimated from the data, use an adaptive test procedure (e.g., $\alpha' = 4$ times Error α), as we have done in the present case.

Finally, assume we want to assess the data for the Exponential (or any other continuous distribution) assumption. In such case, we obtain the Exponential mean (or any other appropriate distribution parameter) and apply the KS procedure, described in this section for the Weibull distribution. But now, we use instead the Exponential (or other pertinent probabilities) instead of the Weibull, in Column 2 of Table 8.

Table 9. Step-by-Step Summary of the KS GoF Test for the Weibull

•   Establish the (Null Hypothesis) assumed distribution: Weibull (α; β).
•   Estimate the Weibull parameters: $\alpha = 8.7$; $\beta = 1.3$.
•   Sort the data in ascending order (Col. 1, Table 8).
•   Obtain the Theoretical distribution (Col. 2).
•   Obtain the Empirical distributions (Cols. 3 and 4).
•   Obtain D+ and D- (Cols. 5 and 6).
•   Obtain the KS statistic: D = Maximum (D+, D-) = 0.2331.
•   Obtain the KS table critical value(s) = 0.521 (or 0.411; Adaptive).
•   Since D < CV do not reject the Weibull ($\alpha = 8.7$; $\beta = 1.3$).
•   Software for Weibull version of KS is not commonly available.

A Counter Example

As an example of the use of KS where the assumed distribution is not true, we now refit the data (recall that these are Normal data) in Table 1 for the Weibull. We again estimate, using the graphical methods given in [11], the two unknown Weibull parameters: shape = 8 and scale = 350. Again, we use formula (1) and the KS procedures of the previous section and obtain the intermediate results shown in Table 10.

Table 10. Intermediate Values for the KS GoF Test for the Weibull

Row	DataSet	F_0	F_n	F_{n-1}	D+=Fn-Fo	D-=Fo-Fn-1
1	294.2	0.220598	0.16667	0.000000	-0.053931	0.220598
2	308.5	0.305339	0.33333	0.166667	0.027994	0.138673
3	313.1	0.336435	0.50000	0.333333	0.163565	0.003102
4	317.7	0.369275	0.66667	0.500000	0.297391	-0.130725
5	322.7	0.406793	0.83333	0.666667	0.426540	-0.259873
6	338.7	0.536565	1.00000	0.833333	0.463435	-0.296768
				Max:	0.463435	0.220598

The KS GoF test statistic (1) value (0.463) is now three times higher than before. It is larger than the KS test CV for $\alpha' = 0.2$ (0.411) and smaller than the CV for $\alpha = 0.05$ (0.52), both for n = 6. Given that both Weibull parameters were estimated from the sample, the test error α is different from the one stated and we have to use an adaptive procedure (i.e., compare with 0.411). Based on these adaptive KS results, we reject the hypothesis that the population from where these data were obtained, is distributed Weibull ($\alpha = 350$; $\beta = 8$).

The entire KS GoF test procedure is summarized in Table 11.

Table 11. Step-by-Step Summary of the KS GoF Test for the Weibull

- Establish the (Null Hypothesis) assumed distribution: Weibull (α; β).
- Estimate the Weibull parameters: $\alpha = 350$; $\beta = 8$.
- Sort the data in ascending order (Col. 1, Table 10).
- Obtain the Theoretical distribution (Col. 2).
- Obtain the Empirical distributions (Cols. 3 and 4).
- Obtain D+ and D- (Cols. 5 and 6).
- Obtain the KS statistic: D = Maximum (D+, D-) = 0.463.
- Obtain the KS table critical value CV(α') = 0.411 (Adaptive).
- Since KS > CV reject that data comes from Weibull ($\alpha = 350$; $\beta = 8$).
- Software for Weibull version of KS is not commonly available.

Summary

In this START sheet we have discussed the important problem of (GoF) assessment of statistical distributions, especially for small samples, via the Kolmogorov Smirnov GoF test. Alternatively, one can also implement the Anderson Darling test [10]. These tests can also be used with large samples. We have provided several numerical and graphical examples for testing the Normal, Lognormal, Exponential and Weibull distributions, relevant in reliability and maintainability studies (Lognormal is a special case of the Normal). We have also discussed some related theo-retical and practical issues, providing several references to background information and further readings.

The large sample GoF problem is often better dealt with via the Chi-Square test [9]. For, it does not require knowledge of the distribution parameters -something that both, AD and KS tests do require and affects their power. On the other hand, the Chi Square GoF test requires that the number of data points be large enough for the test statistic to converge to its underlying Chi-Square distribution -something that neither AD nor KS require. Due to their complexity, the Chi Square and the Anderson Darling GoF tests are treated in more detail in separate papers [9, 10].

Bibliography

1. Practical Statistical Tools for Reliability Engineers, Coppola, A., RAC, 1999.
2. A Practical Guide to Statistical Analysis of Material Property Data, Romeu, J.L. and C. Grethlein, AMPTIAC, 2000.
3. An Introduction to Probability Theory and Mathematical Statistics, Rohatgi, V.K., Wiley, NY, 1976.
4. Methods for Statistical Analysis of Reliability and Life Data, Mann, N., R. Schafer, and N. Singpurwalla, John Wiley, NY, 1974.
5. Statistical Confidence, Romeu, J.L., RAC START, Volume 9, Number 4, http://rac.alionscience.com/pdf/STAT_CO.pdf.
6. Statistical Assumptions of an Exponential Distribution, Romeu, J.L., RAC START, Volume 8, Number 2, http://rac.alionscience.com/pdf/E_ASSUME.pdf.
7. Empirical Assessment of Normal and Lognormal Distribution Assumptions, Romeu, J.L., RAC START, Volume 9, Number 6, http://rac.alionscience.com/pdf/NLDIST.pdf.
8. Empirical Assessment of Weibull Distribution, Romeu, J.L., RAC START, Volume 10, Number 3, http://rac.alionscience.com/pdf/WEIBULL.pdf.

9. The Chi-Square: A Large-Sample Goodness of Fit Test, Romeu, J.L., RAC START, Volume 10, Number 4, http://rac.alionscience.com/pdf/Chi_Square.pdf.
10. Anderson-Darling: A Goodness of Fit Test for Small Samples Assumptions, Romeu, J.L., RAC START, Volume 10, Number 5.
11. Weibull Analysis, Dodson, B., ASQ Press, WI, 1994.

About the Author

Dr. Jorge Luis Romeu has over thirty years of statistical and operations research experience in consulting, research, and teaching. He was a consultant for the petrochemical, construction, and agricultural industries. Dr. Romeu has also worked in statistical and simulation modeling and in data analysis of software and hardware reliability, software engineering and ecological problems.

Dr. Romeu has taught undergraduate and graduate statistics, operations research, and computer science in several American and foreign universities. He teaches short, intensive professional training courses. He is currently an Adjunct Professor of Statistics and Operations Research for Syracuse University and a Practicing Faculty of that school's Institute for Manufacturing Enterprises. Dr. Romeu is a Chartered Statistician Fellow of the Royal Statistical Society, Full Member of the Operations Research Society of America, and Fellow of the Institute of Statisticians.

Romeu is a senior technical advisor for reliability and advanced information technology research with Alion Science and Technology. Since joining Alion in 1998, Romeu has provided consulting for several statistical and operations research projects. He has written a State of the Art Report on Statistical Analysis of Materials Data, designed and taught a three-day intensive statistics course for practicing engineers, and written a series of articles on statistics and data analysis for the AMPTIAC Newsletter and RAC Journal.

Other START Sheets Available

Many Selected Topics in Assurance Related Technologies (START) sheets have been published on subjects of interest in reliability, maintainability, quality, and supportability. START sheets are available on-line in their entirety at <http://rac.alionscience.com/rac/jsp/start/startsheet.jsp>.

For further information on RAC START Sheets contact the:

Reliability Analysis Center
201 Mill Street
Rome, NY 13440-6916
Toll Free: (888) RAC-USER
Fax: (315) 337-9932

or visit our web site at:

<http://rac.alionscience.com>

About the Reliability Analysis Center

The Reliability Analysis Center is a Department of Defense Information Analysis Center (IAC). RAC serves as a government and industry focal point for efforts to improve the reliability, maintainability, supportability and quality of manufactured components and systems. To this end, RAC collects, analyzes, archives in computerized databases, and publishes data concerning the quality and reliability of equipments and systems, as well as the microcircuit, discrete semiconductor, and electromechanical and mechanical components that comprise them. RAC also evaluates and publishes information on engineering techniques and methods. Information is distributed through data compilations, application guides, data products and programs on computer media, public and private training courses, and consulting services. Located in Rome, NY, the Reliability Analysis Center is sponsored by the Defense Technical Information Center (DTIC). Alion, and its predecessor company IIT Research Institute, have operated the RAC continuously since its creation in 1968. Technical management of the RAC is provided by the U.S. Air Force's Research Laboratory Information Directorate (formerly Rome Laboratory).