33. MODEL SELECTION/GOODNESS OF FIT

33.1 Review/Background

In the last chapter, we discussed various type of distributions, we talked about Physical distribution (based on the physics of the problem) and some empirical distributions. We talked that Weibull belongs to both these categories, for example the original Weibull that came from Airplane wing breaking down problem, was empirical, and the Weibull that came from gate dielectric breakdown is physical one. Now once we have a candidate distribution function for a reliability mechanism and we also have experimental data, we should be able to extract the parameters for our model. Many times once we have empirical data we try to fit it with different models like parabolic fit or linear fit etc. This lecture is about how to pick a model and how to determine whether the model is right one for our problem (goodness of fit). Correct fitting of the experimental data with empirical probability distribution functions is important so that we can predict the failure rates for very large population. Equally important is the requirement of “goodness-of-fit”– since many empirical distributions can fit the limited dataset, nevertheless, some fit the data better than others and we will see how to differentiate between them. At first, we will see the problem of matching data with theoretical distribution. After then, this lecture will describe parameter extraction by fitting the probability distribution function with the experimental data. Finally, the later part is selection of most apposite probability distribution functions for a physical process.

33.2 The problem of matching data with theoretical distribution

Having a set of data points from experimental data is always discrete and not enough to fit clearly in some case. Thus, fitting the data points with the theoretical distribution which is called as moment based approach should be carefully done. For example, if there is a few data points having extreme value distribution as shown in Figure. 33.0, some parts of the fitting can be perfect but others can not. In detail, several
data points in the right top of the figure is fitted well and then log-normal distribution and 
Weibull distribution are applied on the points; however, how can we determine the left 
side of the figure is fitting well or not. Therefore, the moment based fitting approach is a 
bad idea in reliability physics.

![Figure. 33.0 Comparison of fitting between Log-normal and Weibull distribution.](image)

33.3 **Parameter extractions: Moments, Linear regression, Maximum likelihood**

33.3.1 **Method of Moments**

This method can be employed to determine parameter estimates for a 
distribution. The method of matching moments sets the distribution moments (theory) 
equal to the data moments (experiment) and solves to obtain estimates for the distribution 
parameters. For example, for a distribution with two parameters, the first two moments of 
the distribution (the mean μ and variance σ of the distribution, respectively) would be set 
equal to the first two moments of the data (the sample mean μ and variance σ, 
respectively) and solved for the parameter estimates.
33.3.2 Method of Least squares

Least square is a mathematical optimization technique which, when given a series of measured data, attempts to find a function which closely approximates the data (the best fit). It attempts to minimize the sum of the squares of the ordinate differences (called residuals) between points generated by the function and corresponding points in the data.

Suppose that the data set $F_{i,\text{exp}}$ consists of the points $(x_i, y_i)$ with $i = 1, 2, ..., n$. We want to find a function $F_{i,\text{theory}}$ such that $F_{i,\text{exp}} \approx F_{i,\text{theory}}$. This is done by minimizing the error between observed and theoretical distributions, i.e.

Error Function:

$$E(\alpha, \beta) = \sum_i (F_{i,\text{exp}}(t_i) - F_{i,\text{theory}}(t_i, \alpha, \beta))^2 \quad 33.1$$

For example, the Weibull parameters $\alpha$ and $\beta$ can be obtained by minimizing the error with respect to those parameters: i.e. $dE/d\alpha$ and $dE/d\beta$ and solving the two equations in the process.

33.3.3 Method of Correlation Coefficient

In method of correlation coefficient, the data is first plotted again $y = a + bx$, then the invert the axis to plot the data as $x = a^* + b^*y$. Therefore, if it was an exactly straight line, the values of coefficients ($a$, $b$, $a^*$, $b^*$) would be the same, however, if the line was not straight in one case we get the error from $(y_i - y)$, where as in the later case we will get an error from $(x_i - x)$, as shown below

$$b = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum y_i^2 - (\sum y_i)^2} \quad 33.2$$

$$b^* = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} \quad 33.3$$
The error estimate is usually provided by the table as shown below.

<table>
<thead>
<tr>
<th>Data points (n)</th>
<th>Correlation (r)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>r = 0.5</td>
</tr>
<tr>
<td>3</td>
<td>0.667</td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.141 (~14%)</td>
</tr>
</tbody>
</table>

The table provides an estimate of the goodness of fit, e.g. if correlation coefficient $r = 0.9$ and we have 4 data points ($n = 4$), there is only a 10% chance the fit is poor. Similarly, if $r = 0.5$, with 3 data points, the chances of the line being nonlinear is ~67%; whereas only ~14% with 10 data points. In general, if we increase either the correlation factor or data points the accidental probability comes down.

33.3.4 Fisher’s Maximum Likelihood Method

This method is credited to Fisher (1920) and is widely used for robust fitting of experimental data. Maximum likelihood estimation begins with writing a mathematical
expression known as the Likelihood Function of the sample data. In other words, the likelihood of a set of data is the probability of obtaining that particular set of data, given the selected probability distribution model. This expression contains the unknown model parameters. The values of these parameters that maximize the sample likelihood are known as the Maximum Likelihood Estimates.

To get better understanding of this method, we go through following example. Let us consider Weibull distribution as the theoretical PDF to fit the experimental data. The unknown model parameters are $\alpha$ and $\beta$, respectively. By changing the model parameters ($\alpha$ and $\beta$) assuming normal Gaussian distribution for variables in terms of variance and mean, Maximum Likelihood Estimator (MLE) finds particular parametric values that make the observed data the most probable as shown in Figure 33.2. Hence, the Likelihood Function of the sample data is assumed to be

$$L = \prod_{i=1}^{n} f(t_i, \alpha, \beta)$$

Since product function is difficult to evaluate, one often use the log of the previous expression, i.e.
\[ \ln L = \prod_{i=1}^{n} \ln f(t_i, \alpha, \beta) \]  

33.5

with the reasonable assumption that same parameters would maximize \( \log(L) \) and \( L \).

To find the values of \( \alpha \) and \( \beta \) so that the Likelihood Function reaches its maximum value, we set \( \frac{d \ln L}{d \alpha} = 0 \) and \( \frac{d \ln L}{d \beta} = 0 \).

For Weibull distribution, one can readily show that the above conditions produce two equations with two unknowns (eq 33.6 give value of \( \beta \) and eq 33.7 gives \( \alpha \))

\[ \frac{\sum_{i=1}^{n} t_i^\beta \ln(t_i)^\beta}{\sum_{i=1}^{n} t_i^\beta} - \frac{1}{n} \sum_{i=1}^{n} \ln(t_i)^\beta = 1 \]  

33.6

\[ \alpha = \left[ \frac{1}{n} \sum_{i=1}^{n} \ln(t_i)^\beta \right]^{\frac{1}{\beta}} \]  

33.7

Obviously, one can fit various theoretical distributions to the same set of experimental data and obtain relevant parameters. There is nothing in the mathematical procedure in Method of Moments, Least Square methods, or MLE to suggest if one model is better than the other. However, when \( \alpha \) and \( \beta \) are calculated, these are internally dependent values. Therefore, in order to find appropriate \( \alpha \) and \( \beta \), complicated calculation is needed. For this reason, MLE was not popular before computers was not invented. Furthermore, there is a problem with \( n \)-parameter distribution by the MLE. For example if sample size (\( n \)) for \( \sigma \), \( n \) local maxima one close to each observation. Also, at times, we are interested in the tails of the distribution and various theoretical distributions would predict different values of the tail of the distributions. Therefore the question is: Which one of these distributions fits the data the best such that the tail of the distribution is reliable? Therefore, the remaining part of this lecture is about “Goodness of Fit”.
33.3.5 Goodness of fit: Residual, Pearson, Cox, Akika

In the first half of the lecture we discussed how to extract the different parameters of a probability distribution function (PDF) by curve fitting the experimental data. In most occasions, the available data will be so few that more than one PDF might describe it quite satisfactorily (Figure 33.3). But the tail of these distribution functions will be entirely different and the hence the projections made using them will also differ from each other. Note that we still don’t know about the underlying physical mechanism of the process, which would have definitely helped in deciding the exact probability distribution. So ideally one would like to have some empirical methods that can help us determine which PDF describes the experimental data better and hence should be chosen for projection. Specifically the problem we analyze in this section is how to decide which function fits the data well in the absence of any information regarding the physics behind it.

![Figure 33.3. The functions \( f_1(x) \) and \( f_2(x) \) fits the experimental data well in the observed range but have different tails](image)

Tests for goodness of fit are mainly used in statistics for hypothesis testing, to test whether two samples are drawn from identical distribution or whether the PDF obtained experimentally follows a specified distribution (which is our task in this section).
It is important to note that these are just statistical tests with no physical basis. The different methods that we discuss are (a) Residual method, (b) Method of correlation coefficient, (c) Cox-Oakes method, and (d) Kolmogorov-Smirnov Algorithm, (e) Pearson $\chi^2$-test algorithm and (f) number based method (Adjusted Residual method and Akaike Information Criteria).

### 33.3.6 Q-Q plot method

The quantile-quantile (Q-Q) plot is a graphical method for determining whether two different sources of data sets have common distribution. At first, Take the q-quantile values of the original data and plot in the y-axis. And then, take another the q-quantile values of the test-distribution to compare (i.e., calculate $x=F^{-1}(q)$) to define the x-axis. Finally, visually inspect and establish deviation from linearity as shown in Figure 33.4.

![Q-Q plot](image)

Figure 33.4. Quantile-Quantile (Q-Q) plot. Taking method for quantiles in each data (left) and plotting to compare each quantile for checking analogy.
33.3.7 **Residual method**

This is visual estimation method. Here the difference between the observed data and theoretical estimate (using the parameters obtained by fitting the corresponding function to the experimental data) is plotted (i.e., \( \Delta y = y_{fit} - y_{expt} \)). If the difference is random or balanced the fit obtained is regarded as good (Figure. 33.5 (a)). Generally the sample data which is 3\( \sigma \) away from the mean are discarded (data point shown as in Error! Reference source not found.(b)). But if the error shows a systematic behavior then it is almost certain that the fit obtained is a poor one (Figure. 33.5(b)). To decide among different fitting functions, the residuals are plotted together and only those which show a random behavior for residuals are retained.

![Residual distribution for a good fit](image)

Figure. 33.5. Residual distribution for a good fit. The data point denoted as E is 3 away from the mean (shown as dotted line) and is discarded (left). (b) residue shows a systematic behavior which shows that the PDF used doesn’t describe the data very well (right)

33.4 **Cox-Oakes measure**

Here the normalized third moment of the data is computed for each of the distribution functions using the formula
\[ \mu_3^* = \frac{\mu_3}{\sigma^3} \]  

\[ \mu = \frac{1}{n} \sum_{i=1}^{n} t_i \]  

\[ \sigma^2 = \frac{1}{n-1} \sum_{i=1}^{n} (t_i - \mu)^3 \]  

where \( n \) is the number of samples, \( \mu \) and \( \sigma \) are the mean and standard deviation obtained by curve fitting the corresponding PDF with experimental data.

The third moment provides a measure of asymmetry in the PDF. Now the extracted \( \mu_3^* \) is plotted as a function of \( \sigma/\mu \) for different PDFs.

Figure. 33.6. For a given \( \alpha, \beta \) Weibull has a specific \( \mu, \sigma, \mu_3 \) (blue triangle). It is seen that every distribution has different shape.

The normalized third moment of the experimental data is also plotted in the same figure (using the and obtained from raw data using statistical techniques discussed in lecture 31-32).
When the point is close to certain distribution, the distribution is the best fit among different distribution. The PDF which gives a better estimate for the normalized third moment is chosen as the optimum one. For example in Figure. 33.6, the Weibull distribution gives a better estimate for the normalized third moment.

### 33.5 Kolmogorov-Smirnov algorithm

This criteria instead of total sum, it focuses on the maximum error i.e. it computes

\[
D_n = \max |F_{obs}(t_i) - F_{theory}(t_i)|
\]  \hspace{1cm} 33.11

Where, \(D_n\) is sample size. If \(D_n > D_n^{crit}\), then the fit is said to be poor, otherwise good. We can plot distributions \((D_n)\), for various number of data points \(n\), but before we do this, we have to have a \(D_{crit}^{table}\) table for a particular significance level for various data points. The table for 5% significance levels, is shown as inset to the Figure. 33.7. (\(D_{crit}^{table}\) are also marked on the Figure. 33.7).

![Figure. 33.7. \(D_n\) and \(D_{crit}^{table}\) for \(n = 5\) and \(n = 20\)](image)

58
The way we interpret the goodness of fit is, e.g. if we have 10 data points. In order to have 5% confidence level in the fitness, no data point should be $> D_{critical}$ (0.409) away from the theoretical curve. The matlab code for one sample is

$$[h,p]=kstest(x,’CDF’,test_cdf,’Alpha’,0.01),$$

(the hypothesis test result - h, p-value - p, and significance level - Alpha respectively).

### 33.6 Pearson $\chi^2$- Test Algorithm

The Pearson $\chi^2$ test is one of the powerful tests for determining the goodness of a fit. Here the available data is represented as a histogram as shown in Figure. 33.8. The entire range is divided into $n^*$ bins such that each bin has at least 5 data points. As before, a probability distribution function is fitted to this experimental data (shown as dotted line).

![Figure. 33.8. Histogram of the experimental data. The dotted line represents the](image)

An estimate of the Pearson coefficient ($\chi^2$) is obtained by the formula

$$\chi^2_s = \sum_{i=1}^{n^*} \frac{(o_i - e_i)^2}{e_i} \quad 33.12$$
where $O_i$ and $E_i$ represents the $i$th observed and estimated value.

In a $\chi^2$-test, the decision regarding the goodness of a fit is made by comparing the estimated $\chi^2$ with the 5% or 1% tail of the theoretical $\chi^2$ distribution. The distribution is given by

$$f(\chi^2) = k(v)(\chi^2)^{\frac{v-2}{2}}e^{-\chi^2/2}$$  \hspace{1cm} (33.13)

$$v = n^* - k - 1$$  \hspace{1cm} (33.14)

where $v$ is the degrees of freedom and $k$ is the number of parameters extracted while fitting the PDF with the experimental data. The $\chi^2$ function is plotted for different $v$ in Figure. 33.9.

![Figure. 33.9. $\chi^2$ distribution function](image)

We proceed as follows: The Pearson coefficient is estimated for all PDFs being considered. Then the $\chi^2$-function with the corresponding degrees of freedom is used to determine the probability of observing the estimated $\chi^2$ for each of the PDFs being
considered. Subsequently, if one plans to determine the goodness-of-fit to \( p \) percent significance level, then one finds \( \chi^2_{\text{critical}} \) by requiring that
\[
p = \int_{\chi^2_{\text{crit}}}^{\infty} f_\nu(\chi^2) d(\chi^2).
\]
A fit is good if \( \chi^2_{\text{expt}} < \chi^2_{\text{crit}} \).

### 33.7 Parameter number vs. goodness of fit

In terms of number of parameters there are three methods using R value blow

\[
R \equiv \sum_{n=1}^{n} (t_i - t_{i,\text{fit}})^2
\]

### 33.8 Method of adjusted residual

This one is different than normal residual, which is \( R^2 \), here we have number of parameter included in the expression as follows

\[
R^2_{\text{adj}} = \frac{(n - 1)R^2 - (M - 1)}{n - M}
\]

where \( n \) is number of samples and \( M \) is number of parameters.

### 33.9 Akaike Information Criterion (AIC)

The AIC criterion is given by:
\[ AIC = n \times \ln R^2 + 2M \]  

where \( f_i \) and \( f_e \) are the estimated and experimental probabilities and \( M \) is the number of parameters extracted while curve fitting the pdf to the experimental data (same \( k \) in section 33.3.5 i.e. Pearson \( \chi^2 \) test). The PDF which gives the least value of AIC is often chosen to represent the data.

### 33.10 Schwarz Information Criterion (BIC)

BIC is based on the likelihood function and it is similar with AIC BIT criterion is given by

\[ BIC = n \times \ln R^2 + M \times \ln(n) \]

The BIC resolves the problem from overfitting of highly increasing the likelihood adding parameters by introducing a penalty term for the number of parameters in the model. The penalty term is larger in BIC than in AIC.

In practice all the above tests are done for a particular set of PDFs (for a given experimental data). Usually the best fit function satisfies most of these criteria very well.

### 33.11 Conclusion

In this lecture first we discussed method of moments, linear regression and maximum likelihood method for extraction of parameters. In addition, at times tail of the distribution is important, which necessitates correct fitting of the experimental data, therefore, in the later part of lecture we considered the various statistical methods used for evaluating the goodness of fit of a particular empirical PDF with the experimental data.
Reference


[33.9] The example of KS analysis was taken from “Introduction to Probability and Mathematical Statistics”, L. J Bain, Ex. 13.8.3.