NPTEL course offered by IIT Madras

Risk and Reliability of Offshore structures

<u>Tutorial 3: Random variables</u>

Answer all questions **Total marks: 25**

1. State and explain qualitative rules of plausible reasoning

The deductive logic is based on the fact that If A is true, the B is true A. Then,

p (B|AC)=p(AB|C)/p(A|C)=p(A|C)=1

If A is true then B is true ≡ $A \Rightarrow B$. If B is false, then A is false. Then, we get

$p(A|BC) = \left[p(B\acute{A}|C)\acute{/}P(B|\acute{C})\right] = 0$

If A is true then B is true. $C \equiv A \Rightarrow B$. Also, if B is true, A becomes more plausible. Hence, we get:

$p(A|BC) = p(A|C)[P(B|AC)/p(B|C]$

But since p(B|AC)=1 and p(B|C)≤ 1, we get:

p (A|BC)≥ $p(A|C)$

Also, following statements of qualitative rules hold good:

p (B|AC)≥p(B|C)

p(A|BC)=p(A|C)$[p(B|AC/p(B|C)]$

P (A|BC)≥p(A|C)

2. What do you understand by Random variable?

Any variable is considered as random variable not because the value of the variable assumes any random number. Value assigned to the random variable is not random but unique; randomness is due to the fact that the assigned value is not known. As it is being guessed in the wide range of possible values, these variables are termed as random variables.

3. List various methods of estimating parameters of probability distribution

There are various methods, both numerical and graphical, for estimating the parameters of a probability distribution.

1. Method of moments
2. Maximum likelihood
3. Least squares
4. PPCC and probability plots

Method of Moments

The method of moments equates sample moments to parameter estimates. When moment methods are available, they have the advantage of simplicity. The disadvantage is that they are often not available and they do not have the desirable optimality properties of maximum likelihood and least squares estimators.

The primary use of moment estimates is as starting values for the more precise maximum likelihood and least squares estimates.

Maximum likelihood

Maximum likelihood estimation begins with the mathematical expression known as a likelihood function of the sample data. Loosely speaking, the likelihood of a set of data is the probability of obtaining that particular set of data given the chosen probability model. This expression contains the unknown parameters. Those values of the parameter that maximize the sample likelihood are known as the maximum likelihood estimates.

Maximum likelihood provides a consistent approach to parameter estimation problems. This means that maximum likelihood estimates can be developed for a large variety of estimation situations. For example, they can be applied in reliability analysis to censored data under various censoring models. Maximum likelihood methods have desirable mathematical and optimality properties. Specifically, they become minimum variance unbiased estimators as the sample size increases. By unbiased, we mean that if we take (a very large number of) random samples with replacement from a population, the average value of the parameter estimates will be theoretically exactly equal to the population value. By minimum variance, we mean that the estimator has the smallest variance, and thus the narrowest confidence interval, of all estimators of that type. They have approximate normal distributions and approximate sample variances that can be used to generate confidence bounds and hypothesis tests for the parameters. Several popular statistical software packages provide excellent algorithms for maximum likelihood estimates for many of the commonly used distributions. This helps mitigate the computational complexity of maximum likelihood estimation.

The disadvantages of this method are as follows:

1) The likelihood equations need to be specifically worked out for a given distribution and estimation problem. The mathematics is often non-trivial, particularly if confidence intervals for the parameters are desired.

2) The numerical estimation is usually non-trivial. Except for a few cases where the maximum likelihood formulas are in fact simple, it is generally best to rely on high quality statistical software to obtain maximum likelihood estimates. Fortunately, high quality maximum likelihood software is becoming increasingly common.

3) Maximum likelihood estimates can be heavily biased for small samples. The optimality properties may not apply for small samples. Maximum likelihood can be sensitive to the choice of starting values.

Least squares

Non-linear least squares provide an alternative to maximum likelihood.

1) Advantages

Non-linear least squares software may be available in many statistical software packages that do not support maximum likelihood estimates.

It can be applied more generally than maximum likelihood. That is, if your software provides non-linear fitting and it has the ability to specify the probability function you are interested in, then you can generate least squares estimates for that distribution. This will allow you to obtain reasonable estimates for distributions even if the software does not provide maximum likelihood estimates.

2) Disadvantages

It is not readily applicable to censored data.

It is generally considered to have less desirable optimality properties than maximum likelihood.

It can be quite sensitive to the choice of starting values.

Probability Plots

The PPCC plot can be used to estimate the shape parameter of a distribution with a single shape parameter. After finding the best value of the shape parameter, the probability plot can be used to estimate the location and scale parameters of a probability distribution.

Advantages :

It is based on two well-understood concepts.

The linearity (i.e., straightness) of the probability plot is a good measure of the adequacy of the distributional fit.

The correlation coefficient between the points on the probability plot is a good measure of the linearity of the probability plot.

1) It is an easy technique to implement for a wide variety of distributions with a single shape parameter. The basic requirement is to be able to compute the percent point function, which is needed in the computation of both the probability plot and the PPCC plot.

2) The PPCC plot provides insight into the sensitivity of the shape parameter. That is, if the PPCC plot is relatively flat in the neighbourhood of the optimal value of the shape parameter, this is a strong indication that the fitted model will not be sensitive to small deviations, or even large deviations in some cases, in the value of the shape parameter.

3) The maximum correlation value provides a method for comparing across distributions as well as identifying the best value of the shape parameter for a given distribution. For example, we could use the PPCC and probability fits for the Weibull, lognormal, and possibly several other distributions. Comparing the maximum correlation coefficient achieved for each distribution can help in selecting which is the best distribution to use.

Disadvantages:

1) It is limited to distributions with a single shape parameter.

2) PPCC plots are not widely available in statistical software packages other than Data plot (Data plot provides PPCC plots for 40+ distributions). Probability plots are generally available. However, many statistical software packages only provide them for a limited number of distributions.

3) Significance levels for the correlation coefficient (i.e., if the maximum correlation value is above a given value, then the distribution provides an adequate fit for the data

with a given confidence level) have only been worked out for a limited number of distributions.

4. **If $P(A) = \frac{1}{2}, P(A \cap B) = \frac{3}{5}$ and $P(B) = p$**

Find p if given that the events A and B are such that they are (i) mutually exclusive (ii) independent

Solution:

$$P(A) = \frac{1}{2}, P(A \cap B) = \frac{3}{5}, and\ P(B) = p$$

It is given that

(i)When A and B are mutually exclusive $A \cap B = \emptyset$

$P(A \cap B) = 0$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$\Rightarrow \frac{3}{5} = \frac{1}{2} + p - 0$$

$$\Rightarrow p = \frac{3}{5} - \frac{1}{2} = \frac{1}{10}$$

(ii) when A and B are independent $P(A \cap B) = P(A).P(B) = \frac{1}{2}p$

It is know that,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$\Rightarrow \frac{3}{5} = \frac{3}{2} + p - \frac{1}{2}p$$

$$\Rightarrow \frac{3}{5} = \frac{1}{2} + \frac{p}{2}$$

$$\Rightarrow \frac{p}{2} = \frac{3}{5} - \frac{1}{2} = \frac{1}{10}$$

$$\Rightarrow p = \frac{2}{10} = \frac{1}{5}$$

5. **If** A and B be independent events with P(A)=0.3 and P(B)=0.4 find the following: (i)P(A∩B); (ii)P(A∪B); (iii)P(A|B); (iv)P(B|A)

Solution:

It is given that $P(A)=0.3$ and $P(B)=0.4$

If A and B are independent events, then

$P(A\cap B)=P(A).P(B)=0.3*0.4=0.12$

It is known that,

$P(A\cup B)=P(A)+P(B)-P(A\cap B)$

$\Rightarrow P(A\cup B)=0.3+0.4-0.12=0.58$

(iii)it is known that $P(A|B)=(P(A\cap B))/(P(B))$

$\Rightarrow P(A|B)=0.12/0.4=0.3$

It is known that,

$P(B|A)=(P(A\cap B))/(P(A))$

$\Rightarrow P(B|A)=0.12/0.3=0.4$