Statistical Inferences and Fundamental Approaches

Bhaveshkumar C. Dharmani
Lovely Professional University,
Jalandhar-Delhi G. T. Road, Punjab – 144411
Bhavesh.23409@lpu.co.in

Abstract

There are various problems based on statistical inferences, such as, parameter estimation, detection, classification, prediction, pattern recognition and others. There are also various techniques to derive some statistical inferences. The article first reviews various problems based on statistical inferences with their differences. At the next level, the article compares the fundamental approaches towards the solutions of these problems. In the third part of the article, it explains the way various techniques for statistical may differ.

Keywords: Statistical inference, estimation, prediction, state estimation, Fisher estimation approach, Bayesian estimation

1. Introduction

We as a human race, want a machine to assist us in decision making either to emphasize our own decisions or to take a co-operative decision or even sometimes completely to rely on the decision by machines. For that the machines also have to be enabled with all the mental faculties to think and to decide, such as mental abilities of analysis, differentiation, classification, prediction, recognition, imagination, memorize, learning form experiences and others. Machines have computational abilities and these abilities have to be utilized to induce human-like intelligence. The complexities involved due to indeterministic and random variabilities in natural phenomena requires machines to use Statistical Inferences to reach to a decision based on a quantitative assessment of the real scenario. As ‘Artificial Intelligence’ and ‘Machine Learning’ are becoming two buzz-words for everyone -- from a laymen using mobile phone for day-to-day shopping or gaming or other works to an expert in other field but who is novice these area, it is important to understand the basic philosophy and approaches behind various types of statistical inferences. Also, there is a trend that the technological users are switching over to become a techno-developers and inventors.

To support such transformation, the article is addressed for anyone who is new to the area of statistical inferences. The approach is to provide comparative and intuitive understanding of the topic. Also, a guidance that in a given scenario which type of statistical estimation technique be used. More general treatment of the topic can be found from the books [1-3].

The next Section 2 discusses various types of statistical. Section 3 explains the basic philosophy to solve these inference problems. Section 4 then discusses various approaches applicable towards the available philosophy. Finally, the last section concludes the article.

2. Statistical Inferences

Statistical inferences are based on the data. Accordingly, a branch of Statistics, identified as Design or Applied Statistics, aims designing experiments or sample surveys for data collection to match with certain statistical models or a real scenario. The other branch of Statistics, Mathematical Statistics or Analysis, deals with analyzing the experimental outcomes or measurements with or without the prior knowledge about the experiment to bring inferences with some assured certainties. The measurements or the data is random in nature due to either there is natural variability in the source or noise from another source/communication channel is added in the measurements or there are measurement errors due to inaccurate measuring instrument or method.

There are variety of problems covered under the larger umbrella of statistical inferences.

Estimation: It is a process of selecting a value as the “best estimate” from a given continuous set of possible values. A *point estimate* is the function $\hat{x} = g(Z)$ of the observation vector $Z = \{z_1, z_2, \dots, z_n\}$ that minimizes

$\hat{x} - x$ in some sense, where x is the true value. An *Interval Estimate* of a parameter x is an interval $x_1 - x_2$ with $x_1 = g_1(Z)$ and $x_2 = g_2(Z)$ as the functions of the observation vector Z. Also, (x_1, x_2) is a γ confidence interval of x if the probability of x lying between x_1 and x_2 is γ. γ is a given constant called the *confidence coefficient* of the estimate and the difference $\delta = 1 - \gamma$ is called the *confidence level*.

Decision: It can be viewed as the selection of one "the best choice" out of a set of discrete alternatives or a discrete space. So, **estimation in discrete space** can be viewed as the possibility of not making a choice but obtaining some conditional probabilities of the various alternatives. This information can be used without making "hard decisions".

Classification: It is a task of assigning a class label to an object, a physical process or an event. This is same as taking a decision from a discrete space.

Detection: It is a special case of classification with only two class labels e.g. 'yes' and 'no'.

Hypothesis Testing: In the first step, there is made a statistical hypothesis or an assumption about the value or parameters of estimation based on a statistical model. In the next step, the validity of the made assumption or hypothesis is tested with some certainty or confidence. This is different from a classification problem where there are no pre assumed hypothesis.

Prediction: Given a set of observations for a random variable X, we wish to predict its value x at a future trial. A *Point Prediction* of X deals with the determination of a constant c, so as to minimize in some sense the error $X - c$.
An Interval Prediction of X is the determination of two constants $c_1$and c_2 such that $P(c_1 < X < c_2) = \gamma = 1 - \delta$, where γ is a given constant called the *confidence coefficient*.
In prediction, the model is completely known. So, from the model we infer the future output. Compare to it, in estimation, from the data, we derive some of the missing model parameter(s).

State Estimation: In state estimation, one tries to describe a state, either assigning a class label to it or deriving a set of parameters, for a process which vary with time or space.

Online estimation is the process of estimation of the present state using all the measurements that are available till now, i.e. future measurements cannot be used as in *offline estimation* but all the present and previous measurements are available.
Prediction is the estimation of future states.
Retrodiction is the offline estimation or smoothing, where there is a estimation of past states.

3. General Philosophy of Statistical Inferences

There are different approaches to Statistical inference, once the measurements are available.

1. **Parametric Methods**: These methods consider the data to follow a mathematical model of the physical process or an event from where the measurements are taken.

Unfortunately, the physical structures responsible for generating the object/ physical process/ event are only partly understood and so partly known, partly unknown. So is the model – partly correct, partly incorrect. That is why there is a scope that the real time data when fitted to a mathematical model may result into randomness of parameter(s).

2. **Non-Parametric Methods** or **Data Analysis**: These methods analyze the data without any assumption of a specific probability distribution function or a mathematical model on it.

The data is analyzed or summarized in a way that brings the main features and clarifies underlying structure. For example, averaging all the measurements for estimation of the real parameter.

Again, the empirical data collected constitutes only a sample set and so possible measurements. So, if too much weight is given to the data at hand, the risk of over fitting occurs. The resulting model will depend too much on the accidental peculiarities (or noise or outliers) of the data. On the other hand, if too little weight is given, nothing will be learned and the model completely relies on the prior knowledge. So, size of the data and the quality i.e. exact representation of all possible cases are important.

According to the statistical significance given to the data at hand, there are two approaches to statistical inference.

1. **Classical** or **Frequentist** or **Non-Bayesian** or **Fisher approach**: This approach assumes that though the measurements are random, the decision taken i.e. the parameter estimated or the class labeled is unknown but deterministic (not random).

Sometimes, it is known that the parameter to be estimated is random, but under given set of measurements with given variability of them we may assume the parameter to be almost constant or deterministic and so classical approach can be still useful.

2. **Bayesian Approach**: This approach assumes that the parameter to be estimated is a random variable with known distribution and so different set of observations may lead to different decisions. Accordingly, instead of completely lying on the data at hand, the prior information regarding the unknowns is also given importance. The prior distribution is modified in the light of the data to determine a posterior distribution (the conditional distribution of a parameter given the data). The approach is based on Bayes' formula: $P(x|Z) = \frac{P(x,Z)}{P(Z)} = \left(\frac{1}{c}\right) P(Z|x)P(x)$, where $P(x|Z)$ is posterior probability or probability of x given the observations Z, $P(Z|x)$ is the likelihood function or the probability of having the observations given the random variable $X = x$, $P(x)$ is the prior probability, c is the normalization constant, which does not depend on x.

3. **Minimum Distance based Estimation Approach:** The approach is based on the estimation of a parameter that minimizes (or maximizes depending upon problem definition) the distance between true distribution and the measurements [3, 4, 5]. Usually, the assumption or availability of the true distribution makes the approach parametric but there are techniques available for minimum distance based non-parametric estimation.

The general philosophies translate into various techniques based on their usefulness and efficiency in a given scenario.

4. Different Techniques under the light of general philosophy

Corresponding to each general philosophical approach, there are variety of techniques available.

1. **Nonparametric approach**: The approach uses data at hand, without super imposing any model.

For example: (i) The simplest possible model to estimate a parameter will be either averaging all the measurements or taking any anyone of them as a representation of the data set. If the n measured values are $z_1, z_2, \dots, z_n$, then an estimate for x is by mean,
$\hat{x} = (z_1 + z_2 + \dots + z_n)/n$. (ii) There may be other measures selected to find mean from the data directly, like least square. Say, the approximated value is $\hat{x}$, for which the sum of the squared difference $\sum(x_i - \hat{x})^2$ is minimum.

These methods are simple in use, but as they do not use any relational model between the measurements and the parameter, are inaccurate most of the time.

2. **Parametric Methods**: The approach assumes data at hand is generated through predefined model.

Let say, the measurements are contaminated with noise. Then, $z = x + w(n)$. We assume that the expectation of the random measurements will give the estimation and w(n) follows some model, i.e., $\hat{x} = E\{Z\} = E\{z|x\} = \int_z z f_z(z|x) dz$, where $E\{\cdot\}$ denotes expectation.

Some Simple Models:

Normal distribution: $z \sim N(\mu, \sigma^2)$; $w(n) \sim N(0, {\sigma_n}^2)$.

Linear Regression: Normal distribution with $E\{z\} = a + bz_i$, where a and b are constants.
Normal Theory, Nonlinear regression: $E\{Z\} = a + be^{cz_i}$ or $E\{Z\} = f(b, z_i)$, where f is a nonlinear function of some unknown b and the measurements z_i.
Exponential Distribution: ρexp(-ρ*zi)
Location and Scale Family: If we assume that the target (tracking) or the measurement location (navigation) is moving, we imagine that the observations are coming from some mean location. In that case, the distribution function $g(z)$ takes the form of $g(z - \mu)$, where we call μ as the location parameter. Sometimes the observations also need to be scaled. The generalized density function then becomes, 1/α(g((z- μ)/α), where α is a positive parameter called a scaled parameter and family of distribution is called location and scale family. Example may be, say if the measurements of temperature are in centigrade and the fitting model requires it into Kelvin.

As said before, according to the statistical significance given to the data at hand, there are two general philosophies. The ***Classical*** or ***Frequentist*** or ***Non-Bayesian*** or ***Fisher approach*** targets to have an inference which has maximum probability of generating the given samples. The unknown $p \times 1$ sized parameter vector is considered to be deterministic. So, there is no reliance on the prior pdf and total reliance on the pdf of the measurements conditioned on the parameter, or the likelihood function of the parameter, $L\big(z(x)\big)\ or\ L(x; z) = f(Z|x)$. The ***Bayesian approach*** uses the Bayes' formula based on the prior and posterior probability distributions.

From the above examples, it is understood that there can be various mechanism or methods to fulfil the general philosophy. The various inference methods also depend upon the used numerical optimization method, as well, focus on achieving maximal performance based on problem dependent various performance measures. For example, given the problem of estimation biasedness, consistency and efficiency are some of the performance measures. In case of classification, the performance of a classifier is measured based on accuracy, precision and recall are some of the performance measures.

5. Conclusion and Discussion

The article presents various types of statistical problems under a same umbrella of statistical inferences. It has explained general philosophy behind and the fundamental approaches towards those philosophy. Understanding the way various methods are derived may be useful to derive some new techniques based on general philosophies.

Reference:

1. Kay, S. M. (1997). *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall.
2. Kay, S. M. (1997). *Fundamentals of Statistical Signal Processing: Detection Theory*. Prentice Hall.
3. A. Basu, H. Shioya and C. Park (2011). *Statistical inference: The minimum distance approach*. Chapman & Hall/CRC Press.
4. García-Dorado, A., & Gallego, A. (2004). Maximum likelihood vs. minimum distance: searching for hills in the plain. *Genetics*, *168*(2), 1085–1086.
5. Woodward, W., Parr, W., Schucany, W., & Lindsey, H. (1984). A Comparison of Minimum Distance and Maximum Likelihood Estimation of a Mixture Proportion. *Journal of the American Statistical Association*, *79*(387), 590-598.