Hypothesis Testing of Process Capability Indices

5.1 Statement of the Problem

Statistics is all about inferring every thing possible about an entire population based on a few random samples taken from that population. This inference is usually about a few parameters of a particular probability distribution which, we suspect, would adequately model the population of interest. The accuracy of the estimates of the parameters is heavily dependent on the sample size, sampling technique, method of estimation etc. One way of qualifying a single estimate of a statistic is through confidence intervals. Instead of a single estimate of a parameter, the accuracy of which is not sure about, confidence intervals provide a range of values within which we can be reasonably sure that the true parameter value lies. But the uncertainty regarding the accuracy of the estimate, still persists to some extent. In hypothesis testing we determine whether a hypothesized value of the parameter is admissible or not, based on random samples taken from the process and the parameter estimate derived from it. A hypothesis test, this way settles doubt regarding the true value of the unknown parameter, to the more
satisfaction of the investigator. At least in few cases, we have to give a clear answer in the form of either accepting or rejecting a hypothesized value of the parameter.

In quality control situations like process capability analysis, we may be required to give a definite decision in the form of either accepting or rejecting the capability of the process. That is, we have to test hypothesis of the form

\[ H_0 : \text{the process is not capable} \]

against

\[ H_1 : \text{the process is capable}. \]

If \( C \) is the capability index used to assess the capability of the process such that the threshold value of \( C \), that makes the process capable is \( C_0 \), then we have the one-sided statistical testing problem

\[ H_0 : C \leq C_0 \quad \text{(not capable)} \]

against

\[ H_1 : C > C_0 \quad \text{(capable)} \] (5.1.1)

The value of \( C_0 \) is never found smaller than one, but in many cases equal to 1.33.

Kane (1986) had investigated the testing of (5.1) in connection with \( C_p \) index, and succeeded in preparing a table giving critical values of \( C_0 \), which can be used to accept or reject the hypothesis corresponding to various values of \( n \). He had drawn two Operating Characteristic (OC) curves for two situations: (i) \( n=30, C_0=1.33 \); (ii) \( n=70, C_0=1.46 \), by a two-point design fixing two points on the OC curve. These tables can be used to implement the method of testing of hypothesis to carry out process capability analysis through the index \( C_p \). These points were corresponding to one low and one high values of \( C_p \). Hoffman (1993) had used another approach for testing the values of \( C_p \) for many \( n \) and significance levels. Kirmani et al. (1991) considered the testing of \( C_p \) when the data are collected as \( m \) subgroups of size \( n \) each. Kocherlakota (1992) also made a study and arrived at critical values for \( C_p \), which decides whether the process is capable or not using the index \( C_p \).

### 5.2 Definition and Role of Generalized P-values

When the underlying family of distributions contain two or more unknown parameters, conventional tests are available only for special functions of the parameters. This is because in many situations it is not possible or not easy to find test statistics having distributions free of nuisance parameters. The need for extending the definition of con-
conventional test statistics and p-value so that the resulting methods do not depend on nuisance parameters, had been felt by many. For example, the Behrens-Fisher problem has been considered from many angles by Dempster (1967), Kempthorne and Folks (1971), Fraser (1979) and Barnard (1984).

The origin of extended p-values or genealized p-values is the resultant of thinking about an extension of the conventional p-value, which answers the case of nuisance parameters also. In fact, Weerahandi (1987) used a generalized p-value for comparing parameters of two regressions with unequal variances and established that it is the exact probability of a well defined extreme region. Motivated by that application, the idea of generalized p-value had been introduced formally by Tsui and Weerahandi (1989) to deal with statistical testing problems in which nuisance parameters are present and it is impossible or difficult to obtain nontrivial tests with a fixed level of significance. The idea of generalized p-values may be given as follows:

Let \( S = (S_1, S_2, \ldots, S_k) \in R^k \) denote a statistic based on \( X = (X_1, X_2, \ldots, X_n) \).

In theory we can consider an independent copy \( X^*_1, X^*_2, \ldots, X^*_n \) of \( X_1, X_2, \ldots, X_n \) and denote the statistic based on \( X^*_i \)'s by \( S^* \). Thus \( S^* \) is an independent copy of the observable random vector \( S \) whose distribution is indexed by \( \xi \in R^p \). We will use \( s \) and \( s^* \) to denote realized values of \( S \) and \( S^* \) respectively.

Suppose \( \theta = \pi(\xi) \) be the parametric function of interest about which we want to test a hypothesis of the form

\[
H_0 : \theta \leq \theta_0 \text{ against } H_1 : \theta > \theta_0 \tag{5.2.1}
\]

Similar to the case of conventional p-value, to define the idea of generalized p-value, we need the definition of what is known as a Generalized Test Variable (GTV) \( T_\theta(S, S^*, \xi) \) which is a function of \( (S, S^*, \xi) \) satisfying the following conditions.

\[
\begin{align*}
\text{(GTV1)} & : \quad \text{The conditional distribution of } T_{\theta_0}(S, S^*, \xi), \text{ conditional on } \quad S = s, \text{ is free of } \xi \\
\text{(GTV2)} & : \quad \text{For every allowable } s \in R^k, \quad T_{\theta_0}(s, s, \xi) = \theta_0 \\
\text{(GTV3)} & : \quad T_{\theta}(S, S^*, \xi) \text{ is stochastically increasing or decreasing in } \theta, \quad \text{for every allowable } s \in R^k
\end{align*}
\]

As in the case of test statistics, the search for generalized test variables can also be confined to functions of sufficient statistics if exist, for ensuring no loss of information about \( \theta \). Then for every \( t = T_{\theta}(s, s, \xi) \) corresponding to \( s \in R^k \), the generalized p-value
for testing (5.2.1) is given by
\[ p = \begin{cases} 
    P[T_\theta(S, S^*, \xi) \geq t|\theta = \theta_0], & \text{if } T_\theta(S, S^*, \xi) \text{ is increasing in } \theta \\
    P[T_\theta(S, S^*, \xi) \leq t|\theta = \theta_0], & \text{if } T_\theta(S, S^*, \xi) \text{ is decreasing in } \theta 
\end{cases} \]

It may be noted from the definition of a GPQ defined in section 1.6 that a GPQ is also satisfying (GTV1) and (GTV2) of GTVs. So a GTV \( \tilde{T}_\theta(S, S^*, \xi) \) for \( \theta \) can always be constructed by using the corresponding GPQ \( R_\theta(S, S^*, \xi) \) for \( \theta \) by the relation,
\[ \tilde{T}_\theta(S, S^*, \xi) = R_\theta(S, S^*, \xi) - \theta. \]

Then \( \tilde{T}_\theta(S, S^*, \xi) \) will be a decreasing function in \( \theta \), and hence the generalized p-value for testing (5.2.1) is given by
\[ p = P[\tilde{T}_\theta(S, S^*, \xi) \leq 0|\theta_0] \] (5.2.2)
as the observed value becomes zero. The generalized p-value is the exact probability of an extreme region. Smaller the p-value, the greater is the evidence against the null hypothesis.

The concept of generalized p-value has been used in a large number of problems to test hypothesis regarding “nonstandard” quantities involving nuisance parameters. Since the class of test statistics is a particular case of test variables, this is indeed a generalization of conventional fixed-level test. Generalized fixed-level tests may be of interest to decision theorists who insist on conventional fixed-level tests as well as to others.

As in the case of GCIs, there were apprehensions regarding the holding of repeated sampling properties by the generalized hypothesis tests using the idea of generalized p-values. A comparison between the two-testing schemes is possible by comparing the power and size of the conventional fixed level and generalized fixed level tests. A number of simulation studies have been conducted to understand the power and size of the generalized fixed level tests in various contexts. According to their findings, not only did the generalized tests provide excellent approximate tests with nominal sizes, but they are also as good as, or often substantially out perform, the other approximate tests. Moreover, with typical values of nuisance parameters, they found the size of fixed level tests based on generalized p-values to be less than the nominal size, suggesting the holding of the repeated sampling property in a more desirable way. A few works in this direction are: Griffths and Judge (1992), Thursby (1992), Weerahandi and Johnson.

5.3 Relation between GCIs and Generalized P-Values

As already noted, the construction of the generalized p-value required for performing generalized hypothesis tests, is based on a suitable generalized test variable. This is similar to the construction of a GCI using the definition of GPQ. It is already noted in the previous section that a generalized test variable can be reached from a generalized pivotal quantity. The route from confidence interval estimation to testing of hypothesis may be explained in connection with that of the capability index $C_{pk}$.

It has been noticed in section 2.3.1 that $R_{C_{pk}} = \frac{d-n[R_{\mu}-M]}{3\sqrt{R_\sigma^2}}$ is a GPQ for deriving the GLCLs for $C_{pk}$, where $R_{\mu}$ and $R_{\sigma^2}$ are defined by

$$R_{\mu} = \bar{X} - \sqrt{\frac{n-1}{n}} \frac{((\bar{X}^* - \mu)/(\sigma/\sqrt{n}))}{(n-1)S^2/\sigma^2} S$$

and

$$R_{\sigma^2} = \frac{S^2\sigma^2}{S^*^2} = \frac{(n-1)S^2}{U^*^2}$$

where

$$Z^* = \frac{\bar{X}^* - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

and

$$U^*^2 = \frac{(n-1)S^*^2}{\sigma^2} \sim \chi^2_{n-1}$$

Now, let us consider the following equation

$$T_{C_{pk}} = R_{C_{pk}} - C_{pk}.$$ 

Then, it can be verified that $T_{C_{pk}}$ satisfies the conditions (GTV1), (GTV2), (GTV3) and $T_{C_{pk}}$ is stochastically decreasing in $C_{pk}$. Therefore, the p-value for testing

$$H_0 : C_{pk} \leq C_0 \text{ against } C_{pk} > C_0 \quad (5.3.1)$$

may be given as $P[T_{C_{pk}} \leq 0 | C_{pk} = C_0]$ as the observed value of $T_{C_{pk}}$ is zero. This probability may also be written as

$$P[R_{C_{pk}} - C_{pk} \leq 0 | C_{pk} = C_0] = P[R_{C_{pk}} \leq C_0]$$
We have calculated the 100\(\alpha\)th percentile points of \(R_{C_{pk}}\) as the 100\((1 - \alpha)\)% lower confidence limits for \(C_{pk}\), together with its coverage probability in chapter 2. Actually, what we have obtained is the expected value of the lower confidence limits as \(E(G_{pk})\). Observing the connection between the coverage probability and the generalized p-value, we can construct generalized test procedure based on the idea of generalized p-value by making use of generalized lower confidence limits, and vice versa. Thus, if \(T_{\alpha}\) is the 100\((1 - \alpha)\)% generalized lower confidence limit computed in the case of \(C_{pk}\), then the generalized test procedure for testing \(H_0 : C_{pk} \leq C_0\) versus \(H_1 : C_{pk} > C_0\), is to reject \(H_0\) if \(T_{\alpha} > C_0\), and otherwise accept. We have also computed the power of the corresponding generalized test procedure based on the generalized p-value, for the various situations considered in the simulation studies, described in the previous chapters in connection with the construction of lower confidence limits for various indices. It has been noticed that the power of the generalized test procedure was quite satisfactory. The results of these simulation experiments are reported in Kurian and Sebastian (2008). However, we are not reporting it here as we have presented the coverage probabilities in extensive volume in the previous chapters. The ultimate conclusion is that the idea of generalized p-values is equally efficient in Hypothesis testing as its counterpart the idea of generalized confidence intervals in Interval estimation, when nuisance parameters are involved in the corresponding problems.

Bibliography


