電力系統·狀態推定에서의 不良情報 檢出技法에 관하여

文 永 鉉
연세대학교 전기공학과

1. 緖 言

電力系統 狀態推定은 대규모 電力系統 運轉을 위한 기본적인 기술로서 세계 각국의 電力會社에서널리 채용하고 있다. 狀態推定은 電力系統 全般으로부터 입수된 測定値를 기본으로 하여 系統狀態를推定해 내는 것이므로 計器故障 또는 데이타傳送 중 測定値 流失 등에 의하여 발생되는 잘못된 測定値 즉 不良情報가 사용된다면 정확한 狀態推定을 기할수 없다. 따라서 이러한 경우에 不良情報를 檢出하여 除去한 후 狀態推定을 행하는 것이 바람직하다.

狀態推定 時에는 그 정확성과 신뢰성을 높이기위하여 不良情報 檢出技法을 狀態推定技法과 동시에 사용하여야 한다. 그러나 狀態推定技法은 널리 알려져 있으며 이에 관한 논문이 국내 학술지에 다수 발표되어 있으므로 이에 관한 언급은 피하고 本稿에서는 不良情報檢出技法에 관하여 기초이론과 현재의 연구방향 등에 관하여 기술하였다. 특히 假說檢定에 의한 不良情報檢出技法은 현재 가장 우수한 技法으로 평가되고 있으나 標本抽出理論을 要하는 등 理論이 복잡하며 체계적 이론을 기술한 문헌조차 구하기 힘든 실정이다.

이에 따라 본 稿에서는 假說檢定관련의 標本抽出에 관한 기초이론을 체계적으로 정리하고 假說檢定技法에 의한 不良情報檢出技法의 원리를 소개하고자 한다.

2. 不良情報檢出技法의 소개

測定데이타 중에는 不良情報가 섞여 있을 확률이 항상 존재하므로 不良情報檢出에서는 不良情報의 存在有無를 판정하는 것이 첫번째 과제이며 不良情報가 존재한다고 판단되면 어느 測定値가 不良情報인지를 판단하여 不良情報를 抽出해야 한다. 따라서 본 절에는서 먼저 불량정보유무판정에 관하여 기술하고 다음으로 不良情報判定技法을 소개한다.

2.1 不良情報有無判定

不良情報有無判定을 위해서는 먼저 測定시스템으로 부터 입수된 모든 測定데이타를 사용하여 狀態推定을 행해야 하며 그 결과 推定된 狀態値 $\underline{\hat{x}}$을 얻을 수 있다. 그러면 다음 式에 의거 狀態推定에 의한 計算値와 실제 測定値와의 誤差自乘의 和를 계산한다.

$$J(\underline{\hat{x}}) = [\underline{z} - \underline{h}(\underline{\hat{x}})]^T R^{-1} [\underline{z} - \underline{h}(\underline{\hat{x}})] \tag{1}$$

단 $\underline{z}$: 측정치 벡터
$\underline{h}(\hat{\underline{x}})$: 상태추정치 $\hat{\underline{x}}$에 의한 계산치

만약 測定値 중에 不良情報가 존재하면 추정된狀態値 $\hat{\underline{x}}$는 정확한 추정이 될 수 없으므로 $J(\hat{\underline{x}})$는 큰 값을 갖는다. 따라서 $J(\hat{\underline{x}})$가 不良情報有無를 判定하는 척도가 될 수 있으며, 이 값을 항상 감시함으로써 不良情報有無를 판단할 수 있다. 즉 $J(\hat{\underline{x}})$의 값이 정상적인 데이타에 의해서 생기는 測定誤差한계를 훨씬 벗어 난다면 不良情報가 존재한다는 것을 의미하므로 $J(\hat{\underline{x}})$값이 이 임계치보다 크면 불량정보가 포함되었다고 판정할 수 있다. $J(\hat{\underline{x}})$의 임계치결정은 상당히 어려운 문제에 해당하며 假說檢定에 의한 不良情報有無判定에서 다시 검토될 것이다. 그림 1은 실계통운전에 있어서 不良情報 有無判定에 관한 例를 보인것이다. 自乘推定誤差$J(\hat{\underline{x}})$가 갑자기 증가하는 것은 그 순간에 不良情報가 발생하였다는 것을 의미하며 不良情報를 제거하지 않는다면 推定誤差가 큰 값으로 남아 狀態推定正確度가 크게 떨어지게 되나 다음에 기술되는 不良情報判定法에 의하여 不良情報를 제거하면 정상적인 推定誤差를 유지할 수 있음을 보인 것이다.

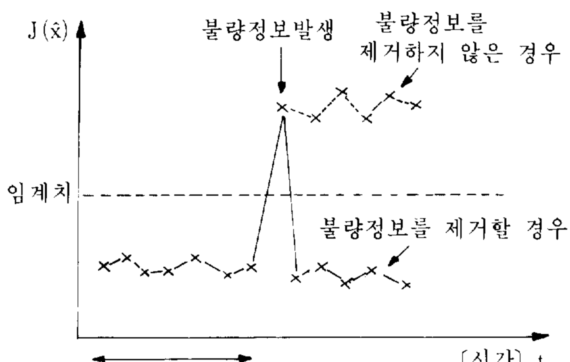

그림 1. 불량정보발생 및 제거에 따른 평가함수 변화

2.2 不良情報判定法

不良情報의 存在가 확인되면 입수된 데이타 중에서 不良데이타를 찾아내어 제거해야 한다. 많은 測定値 중에서 어느 데이타가 不良데이타인지를 판별하는 방법으로 殘留偏差判定法(Residual Error Method)과 傾斜法(Gradient Method)이 보편적으로많이 사용되고 있다.

가. 殘留偏差判定法(Residual Error Method)[1-3]

測定데이타에 의한 狀態推定結果가 $\hat{x}$이라고 한다면 測定値 z_i에 대한 殘留偏差 Δz_i는 다음과 같이 정의된다.

$$\Delta z_i = z_i - h_i(\hat{x}) \quad (i=1, 2, \cdots, m) \tag{2}$$

그러나 殘留偏差 Δz_i의 크기로서 測定値 z_i가 不良情報인지를 판별하기는 곤란하다. 왜냐하면 z_i에 대한 測定計器가 원래 부정확한 측정밖에 할 수 없다면 이 計器로 부터 평상시 정상적으로 취한 데이타 역시 큰 잔류편차를 가질 것이기 때문이다. 따라서, 각 測定値의 測定雜音狀態를 반영시키기 위하여 다음과 같은 定規化 殘留偏差 Δz_i^N를 정의한다.

$$\Delta z_i^N = [z_i - h_i(\hat{\underline{x}})] \,/\, \sigma_i \quad (i=1, 2, \cdots, m) \tag{3}$$

단 σ_i : 측정치 z_i의 표준편차

그러면, 定規化殘留偏差가 가장 큰 測定値가 통상적인 測定値에서 가장 많이 벗어난 測定値임을 알 수 있으며 이 定規化誤差가 일정치이상이면 不良情報로 간주한다.

나. 傾斜法(Gradient Method)

이 방법은 自乘測定誤差 $J(\hat{x})$의 오차에 대한 傾斜(Gradient)를 관찰함으로써 불량정보를 판정하는 방법이며 傾斜(Gradient)函數는 다음과 같이 계산된다.

$$\nabla J = \frac{\partial J}{\partial(\triangle z)} = \frac{\partial}{\partial \triangle z}\left\{[z - \underline{h}(\hat{\underline{x}})]^T R^{-1}[\underline{z} - \underline{h}(\hat{\underline{x}})]\right\} \tag{4}$$

단 $\hat{\underline{x}} = \underline{x}_0 + [HR^{-1}H]^{-1}HR^{-1}\triangle z$

위식에서 ∇J의 i번째 성분 $[\nabla J]_i$는 측정치z_i에 대한 傾斜를 나타내며 측정치 z_i의 오차 $\triangle z_i$가 평가함수 $J(\hat{x})$에 미치는 영향은 $[\nabla J]_i \triangle z_i$로 주어진다. 따라서 측정치에 불량정보가 존재하는 것을 알고 있다면 평가함수에 가장 큰 영향을 미치는 측정치를 불량정보로 간주할 수 있다. 이것은 $[\nabla J]_i \triangle$

z_i가 가장 큰 측정치를 제거함으로써 평가함수 J $(\hat{x})$의 값을 가장 많이 줄일 수 있기 때문이다.

3. 假說檢定에 의한 不良情報檢出

假說檢定에 의한 不良情報檢出技法은 標本抽出理論에 근거를 두고 있다. 測定시스템에서 무한히 많은 측정을 행할 경우에 관측될 수 있는 모든 測定値를 測定値 母集團으로 하고, 실제 측정으로 부터 입수된 測定데이타를 이 母集團으로부터 취한 標本(sample)데이타로 생각한다. 여기서 測定母集團은 測定시스템의 計器測定誤差 및 通信雜音에 관한 통계적 성질을 알고 있기 때문에 母集團의 統計的 性質을 알고 있다고 할 수 있다. 따라서 假說檢定에 의한 不良情報檢出技法은 입수된 測定値가 母集團과 관계없는 不良測定値를 포함하고 있는지를 母集團과 標本의 관계로부터 판정하는 것이다.

본 절에서는 假說檢定의 원리를 설명하기 위해먼저 標本抽出에 관한 統計理論과 χ^2-分布라는 統計分布函數의 性質을 기술하고 마지막으로 假說檢定原理를 기술하는 절차를 취하였다.

3.1 標本抽出理論

구성요소가 N개이며 평균치 μ, 표준편차 σ인 母集團으로 부터 M개의 標本을 抽出하는 경우를 고려해 보자.

그러면 M개의 標本을 선택하는 가능한 모든 방법수는 ${}_NC_M$가지이며 표본 집단마다 각각의 평균과 표준편차를 갖는다. 그리고 그림 2에서와 같이 각 표본집단의 평균치를 모아 또 다른 하나의 표본집단을 만들면 이 집단은 표본 평균집단이 되며 그 평균과 표준편차는 다음과 같이 계산된다.

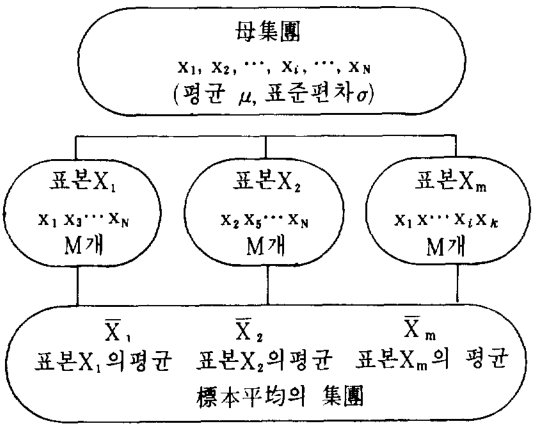

그림 2. 標本抽出 및 標本平均集團의 構成

표본평균집단의 평균 :

$$\mu_{\bar{X}} = \frac{\sum_{i=1}^{m}\bar{X}_i}{m} = E[\bar{X}] = \mu \tag{5}$$

표본평균집단의 표준편차 :

$$\sigma_{\bar{X}} = \sqrt{\frac{1}{m}\left[\sum_{i=1}^{m}(\bar{X}_i-\mu)^2\right]} = E[(\bar{X}-\mu)^2]$$
$$= \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} \tag{6}$$

(단 $\sqrt{\frac{N-n}{N-1}}$ 은 보정계수 임)

무한모집단일 경우(즉 N→∞일 때)는 다음과 같이 간단한 공식을 쓸 수 있다.

$$\mu_{\bar{X}} = \mu \tag{7}$$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \tag{8}$$

이상의 標本의 統計는 다량의 물품매매에 있어서 標本(sample) 檢査에 의하여 제품의 합격·불합격을 결정하는 경우 危險負擔率計算에 직접 적용될수 있다. 狀態推定을 위한 불량정보추출에서는測定데이타를 납품하는 제품으로 생각하면 불량제품을 받아들일 확율이 불량정보를 받아 들일 확율과 같게 된다. 샘플검사에 의한 물품매매에 있어서는 소비자 위험부담율(일명 α-risk)과 생산자 위험부담율(일명 β-risk)을 생각할 수 있으며 이의 계산은다음 예를 통하여 설명하고자 한다.

〔例 1〕 소비자위험율계산

어느 소비자가 생산자로부터 공급받는 제품500개에 대해서 표본 100개를 추출하여 이 표본의 불량율이 8%이하이면, 소비자는 받아 들인다. 표본의 결과 불량율이 p=0.10이다. 이 제품을 소비자가 받아 들일 확율은 얼마인가?

〔풀이〕 제품 500개의 모집단의 불량율은 p=0.10이므로 모집단의 불량율은 σ는 $\sigma=\sqrt{pq}=\sqrt{(0.10)(1-0.10)}$이며 샘플의 불량율 $\hat{p}$의 평균 및 표준편차는 다음 같이 계산된다.

$$\mu_p = E[\hat{p}] = p = 0.10 \quad (9)$$

$$\sigma_p = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{(0.10)(1-0.10)}{100}} \cdot \sqrt{\frac{500-100}{500-1}} = 0.027 \quad (10)$$

따라서 이를 받아 들일 확율은 다음과 같다.

$$P\{\hat{p} \leq 0.08\} = P\left[\frac{\hat{p}-p}{\sigma_p} \leq \frac{0.08-0.10}{0.027}\right]$$
$$= P\{z \leq -0.740\} = 0.5 - \mathrm{erf}(0.740)$$
$$= 0.2296$$

〔例 2〕 생산자 위험의 경우

위와 같은 문제에서 생산자는 불량율이 $p=0.05$인 제품을 생산하여 소비자에게 납품할 경우 불합격이 될 확율은 얼마인가?

〔풀이〕 $\hat{p}$의 표준편차는

$$\sigma_p = \sqrt{\frac{(0.05)(1-0.05)}{100}} \sqrt{\frac{500-100}{500-1}} = 0.0195 \quad (12)$$

이며 불합격이 될 확율은

$$P\{0.08 \leq \hat{p}\} = P\left(\frac{0.08-0.05}{0.0195} \leq \frac{\hat{p}-p}{\sigma_p}\right)$$
$$= 0.5000 - \mathrm{erf}(1.54)$$
$$= 0.0618 \quad (13)$$

이다.

3.2 카이自乘(χ^2)確率分布

서로 獨立인 m개의 確率變數 $X_1, X_2 \cdots X_m$이 정규분포 $N(\mu, \sigma^2)$에 따라 분포하는 확율변수라 하자. 그러면 確率變數 χ^2은 다음과 같이 정의된다.

$$\chi^2 = \sum_{i=1}^{m}\left(\frac{X_i-\mu}{\sigma}\right)^2 \quad (14)$$

이 때의 χ^2의 確率分布를 自由度 m의 χ^2- 分布라 하며 確率密度函數는 다음과 같다.

$$f(\chi^2) = \frac{1}{2^{m/2}\Gamma\left(\frac{m}{2}\right)}(\chi^2)^{m/2-1}e^{-\frac{1}{2}\chi^2} \quad (15)$$

그리고 χ^2分布는 다음과 같은 성질을 갖는다.

1. 平均 : $E[\chi^2] = m$ (16)

2. 共分散 : $V[\chi^2] = E\{[\chi^2-m]^2\} = 2m$ (17)

한편, 確率計算時 수식의 표기를 간단히 하기 위하여 다음과 같은 χ_α^2의 정의를 사용한다.(定規分布에서 $\mathrm{erf}(\alpha)$와 비슷한 규약임)

〔정의〕 χ_α^2는 다음식을 만족하는 χ_α^2값을 의미한다.

$$P\{\chi_\alpha^2 \leq \chi^2 < \infty\} = \int_{\chi_\alpha^2}^{\infty} f(\chi^2)\,d\chi^2 = \alpha \quad (18)$$

예를 들면 自由度 5 에서 $\chi_{0.05}^2$의 값은 11.0705이며 이것은 아래 그림에서 빗금친 부분의 면적이 0.05가 되는 점에서의 χ^2의 수치를 나타낸다.

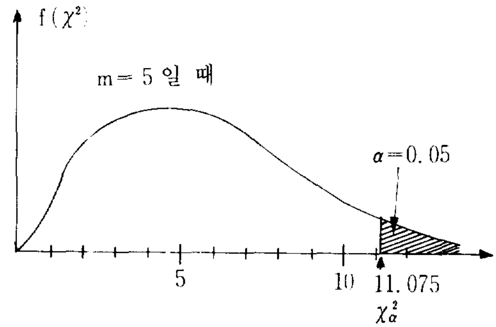

그림 3. 自由度 5 의 χ^2分布 및 χ_α^2

3.3 標本分散의 分布 및 母集團의 分散推定

平均μ, 標準偏差 σ인 특정한 母集團에서 n개의 원소를 취하여 표본 χ를 만든다고 하자. 그러면 표본의 원소는 모집단의 원소중에서 임의로 선택할 수 있으므로 하나의 확율변수로 취급할 수 있으며 이것을 확율변수 X_i로 표시한다. 그리고 표본 χ는 다음과 같이 나타낼 수 있다.

$$\chi = \{X_1, X_2, \cdots, X_n\} \quad (19)$$

위의 표본 χ에 대한 평균치를 $\bar{X}$로 표시하면 $\bar{X}$ 역시 하나의 확율변수가 되며 다음과 같이 나타낼 수 있다.

$$\bar{X} = \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right) \quad (20)$$

그리고 표본집단 χ의 분산을 標本分散이라 정의하며 s^2으로 표기한다. 標本分散은 다음과 같이 계산된다.

$$\hat{s}^2 = \frac{1}{n}\left[\sum_{i=1}^{n}(X_i-\overline{X})^2\right] \tag{21}$$

위 식에 의하여 계산된 標本分散 $\hat{s}^2$은 표본크기가 1일 경우(n=1) $\hat{s}^2=0$이 된다. 그러나 이것은 모집단의 분산과 큰 차이가 있으며 실제 분산값은 不定(indefinite)이라고 할 수 있다. 따라서 式(21)의 표본분산을 직접사용하지 않고 다음과 같이 표본에 대한 不偏分散 s^2을 정의하여 사용한다.

$$s^2 = \frac{1}{(n-1)}\left[\sum_{i=1}^{n}(X_i-\overline{X})^2\right] = \frac{n}{(n-1)}\hat{s}^2 \tag{22}$$

위 식의 불편분산의 기대치는 모집단의 분산 σ^2과 같다.

$$E[s^2] = \sigma^2 \tag{23}$$

그리고 표본분산 $\hat{s}^2$의 기대치는 모집단의 분산 σ^2과 다음 관계를 갖는다.

$$E[\hat{s}^2] = \left(\frac{n-1}{n}\right)\sigma^2 \tag{24}$$

한편 式(21)로 정의되는 표본분산 $\hat{s}^2$의 확율분포는 前節에서 기술한 χ^2-分布를 사용하여 나타낼 수 있다.

즉 평균 μ 분산 σ^2의 母集團으로부터 n개의 要素를 취하여 만든 표본집단에 대한 標本分散$\hat{s}^2$은 다음과 같은 確率分布를 갖는다.

① $E[\hat{s}^2] = \frac{n-1}{n}\sigma^2$ (25)

② χ^2은 다음과 같이 나타낼 수 있으며

$$\chi^2 = \frac{n\hat{s}^2}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{\sigma}\right)^2 \tag{26}$$

위 식의 χ^2의 확율분포는 自由度(n-1)의 χ^2-分布를 따른다.

$\left(E[\chi^2] = E\left[\frac{n\hat{s}^2}{\sigma^2}\right] = (n-1)\right.$ 임에 유의할 것$\left.\right)$

위에서 標本分散 $\hat{s}^2$에 대한 χ^2-分布의 自由度가 (n-1)이 된 이유는 모든 χ_i가 독립적으로 결정되는 것이 아니고 다음관계를 만족해야 하기 때문이다.

$$\sum_{i=1}^{n}(X_i-\overline{X}) = 0 \tag{27}$$

이상의 標本分散 $\hat{s}^2$의 統計的 성질을 이용하면 標本 χ로부터 母集團의 分散 σ^2을 추정할 수 있다.

(가) 小標本에 대한 母集團分散 σ^2의 信賴區間推定

母集團의 分散 σ^2을 危險率 α로서 추정한다면 σ^2의 信賴區間은 다음 식으로 부터 계산될 수 있다.

$$P\{^2 < \chi_1^2\} = \int_0^{\chi_1^2} f(\chi_2)\,d\chi^2 = \frac{\alpha}{2} \tag{28}$$

$$P\{\chi_2^2\} = \int_{\chi_2^2}^{\infty} f(\chi^2)\,d\chi^2 = \frac{\alpha}{2} \tag{29}$$

위 두식에서의 χ_1^2과 χ_2^2은 χ_α^2의 정의로부터 다음과 같이 주어진다.

$$\chi_1^2 = \chi_{1-\frac{\alpha}{2}}^2 \tag{30}$$

$$\chi_2^2 = \chi_{\frac{\alpha}{2}}^2 \tag{31}$$

式 (28), (29)를 합치면

$P\{\chi_1^2 > \chi^2 \text{ or } \chi_2^2 < \chi^2\} = \alpha$

이므로 다음 식을 얻을 수 있다.

$P\{\chi_1^2 \leqq \chi^2 \leqq \chi_2^2\} = 1-\alpha$

여기에 式 (26)을 代入하면

$$P\left\{\chi_1^2 \leqq \frac{n\hat{s}^2}{\sigma^2} \leqq \chi_2^2\right\} = 1-\alpha \tag{32}$$

따라서, 위험율 α에 대한 σ^2의 신뢰구간은 다음과 같이 주어진다.

$$\frac{n\hat{s}^2}{\chi_2^2} \leqq \sigma^2 \leqq \frac{n\hat{s}^2}{\chi_1^2} \tag{33}$$

〔例 3〕 크기 20의 確率標本을 수집하였다. 이의 標本平均 $\overline{X}$는 12.0 그리고 標準偏差 $\hat{s}$는 2.0 이었다. 이 경우에 대하여 母集團의 分散 σ^2의 95% 信賴區間을 推定하라.

〔풀이〕 표본의 크기는 n=20, 표본분산은 $\hat{s}^2=4.0$, χ^2-분포의 자유도는 m=n-1=19이다. 위험율 0.05를 좌우 양측에 각각0.025 씩 나누면 χ^2-분포표로 부터 χ_1^2 및 χ_2^2의 값을 찾을 수 있다. 즉 $\chi_1^2=8.90652$, $\chi_2^2=32.8523$ 얻는다. 그러면 式 (32)로 부터 σ^2의 95% 신뢰구간은 다음과 같이 표시되며

$$P\left\{8.90652 \leqq \frac{20\times4.0}{\sigma^2} \leqq 32.8523\right\} = 0.95$$

위 식에서 σ^2의 구간을 계산하면 다음과 같다.

$2.44 \leqq \sigma^2 \leqq 8.98$

(나) 大標本($n \geqq 30$)일 경우 母集團分散推定

母集團이 分散 σ^2의 정규분포를 이루고 이 母集團으로 부터 취한 標本의 크기 n이 충분히 클 경우 표본의 分散 $\hat{s}^2$은 그 確率分布가 근사적인 定規分布를 이루게 되며 標本分散의 平均과 標本分散$\hat{s}^2$의 分散은 다음과 같이 주어진다.

$$\text{평균} : E[\hat{s}^2] = \frac{n-1}{n}\sigma^2 \tag{33}$$

$$\text{分散} : V[\hat{s}^2] = E[\{\hat{s}^2 - E[\hat{s}^2]\}^2] = \frac{2(n-1)}{n^2}\sigma^4 \tag{34}$$

(증명은 Papoulis[4] 제 8 장 참조)

즉 n이 큰 경우 分散 $\hat{s}^2$의 平均 μ_{s^2}과 分散$\hat{s}^2$ 의 分散 $\sigma^2_{s^2}$은 다음과 같이 근사적으로 나타낼 수 있다.

$$\mu_{s^2} = \sigma^2 \tag{35}$$

$$\sigma^2_{s^2} = \frac{2}{n}\sigma^4 \tag{36}$$

따라서 위험율 α(또는 신뢰도$(1-\alpha)$)에 대한 σ^2의 信賴區間은 定規分布를 사용하여 다음과 같이근사적으로 推定될 수 있다.

$$P\left\{\hat{s}^2 - z_{\alpha/2}\left(\sqrt{\frac{2\sigma^4}{n}}\right) \leqq \sigma^2 \leqq \hat{s}^2 + z_{\alpha/2}\sqrt{\frac{2\sigma^2}{n}}\right\} = 1 - \alpha \tag{37}$$

(단 $z_{\alpha/2} = \mathrm{erf}(\alpha/2)$)

위 식의 부등식을 각각 풀어서 정리하면 다음 식을 얻는다.

$$P\left\{\frac{\hat{s}^2}{1 + z_{\alpha/2}\sqrt{\frac{2}{n}}} \leqq \sigma^2 \leqq \frac{\hat{s}^2}{1 - z_{\alpha/2}\sqrt{\frac{2}{n}}}\right\} = 1 - \alpha \tag{38}$$

따라서 위험율 α에 대한 信賴區間은

$$\frac{\hat{s}^2}{1 + z_{\alpha/2}\sqrt{\frac{2}{n}}} \leqq \sigma^2 \leqq \frac{\hat{s}^2}{1 - z_{\alpha/2}\sqrt{\frac{2}{n}}} \tag{39}$$

으로써 주어진다.

〔例 4〕 크기 100개의 확율분포을 수집하였다. 이표본의 평균은 $\overline{X} = 120$, 표준편차는 $\hat{s} = 3.70$이었다. 이 경우에 대해 모집단의 분산 σ^2의 95% 신뢰구간을 추정하라.

〔풀이〕 χ^2-분포는 n이 커지면 정규분포에 가깝기 때문에 n이 30이상이면 일반적으로 정규분포를사용하여 근사해를 구하는 것이 편리하다. 95% 신뢰구간추정이므로 위험율은 $\alpha = 0.05$가 되며 $z_{\alpha/2} = z_{0.025} = 1.96$이다. 이것을 式(39)에 대입하면 다음과 같다.

$$\frac{3.70^2}{1 + \sqrt{\frac{2}{100}} \times 1.96} \leqq \sigma^2 \leqq \frac{3.70^2}{1 - \sqrt{\frac{2}{100}} \times 1.96} \tag{40}$$

즉 95%신뢰구간은 근사적으로 다음과 같다.

$10.718 \leqq \sigma^2 \leqq 18.94$

한편 더 정밀한 계산을 위해서는 χ^2분포를 사용한 공식 즉 式(30), (31), (33)을 사용할 수 있으며 자유도 $m = 99$에 대한 $\chi^2_{\alpha/2}$ 및 $\lambda^2_{1-\alpha/2}$의 값을 χ^2-분포표로 부터 보간법칙을 사용하여 계산하면 다음과 같다.

$$\chi^2_2 = \chi^2_{0.025} = 129.561 - (129.561 - 118.136) \times \frac{1}{10} = 128.419$$

$$\chi^2_1 = \chi^2_{0.975} = 74.2219 - (74.2219 - 65.6466) \times \frac{1}{10} = 73.3644$$

이것을 식 (33)에 대입하면 σ^2구간은 다음과 같다.

$10.67 \leqq \sigma^2 \leqq 18.76$

이것을 앞에서 구한 근사해와 비교하면 정규분포에 의한 근사계산을 사용하여도 충분히 정확한 결과를 얻을 수 있음을 알 수 있다.

3.4 統計的 假說檢定

統計的 假說檢定에 앞서 실험자료에 의한 동전의 均質性判定에 관한 문제를 생각해 보자.

問) 어떤 동전을 100회 던진 결과 앞면이 40회나

타났다. 이 동전은 앞면과 뒷면이 나올 확률이 같다고 할 수 있는가?

위의 실험에 있어서 동전의 앞면이 50회에 가까운 회수가 나타났다면 그 동전에 대한 균질성을 의심하지 않았을 것이다. 그러나 50회에 비하여 다소 적은 40회의 앞면이 나타났다면 그 동전의 균질성을 의심하게 될 것이다.

동전의 균질성을 판정하기 위해서 동전이 균질하다는 가정하에서 100회 던졌을 때 앞면이 40회 혹은 그 이하가 나타날 확률을 계산해 보자.

앞면이 나타날 확률 : $P=\frac{1}{2}$: 가설

시행회수 : $n=100$

앞면회수의 기대치 : $\mu=np=50$

앞면의회수 표준편차 : $\sigma=\sqrt{npq}=5$

40회 이하가 될 확률 :

$$P=0.5-\mathrm{erf}\left(\frac{50-40.5}{5}\right)$$
$$=0.5-0.4713=0.0287$$

이상에서 만약 동전이 균질하다고 하면 앞면이40회 이하가 될 확률은 단지 2.87%밖에 되지 않으며 이러한 결과를 얻을 확률이 매우 낮다는 것을 알 수 있다. 따라서 다음과 같은 2가지 결론을 내릴 수 있다.

① 가설은 옳으나 다소 드문 현상이 우연히 나타났다.

② 가설이 옳지 않다.

그러나 統計的 假說檢定에서는 일반적으로 후자의 결론을 채택한다. 왜냐하면 위의 실험결과를 우연한 사실의 탓으로 돌리기 보다는 母數와 假說이 잘못 되었기 때문에 나타난 결과라고 보아야 하기 때문이다.

이상의 가설검증에서는 가설이 옳지 않다고 판정되었지만 실제 동전이 균질한 경우에도 앞면이 40회이하 나타날 확률이 2.87%이며 이것은 현재 내린 판정이 오류를 범할 확률이 2.87%임을 뜻한다. 이 확률을 위험부담율이라고 한다.

다음은 假說檢定에서 자주 사용되는 용어들의 정의이다.

〈정의〉 통계적 가설검증에 있어서 표본으로 부터 계산된 표본통계량이 假說을 부정하는 일정한 棄却域에 들어갈 확율 즉 위험부담율의 한계치를 有意水準α라 한다.

따라서 有意水準α로서 假說檢定을 실시한다면 위험부담확율이 α보다 크면 假說이 잘못 되었다고 할 수 없다. 반면에 위험부담확율이 α보다 적으면 假說이 잘못 되었다고 판정할 수 있으며 假說을 棄却한다. 위의 동전의 균질성검정문제에서 有意水準을 $\alpha=0.05$로 檢定한다면 "동전이 균질하지 않다" (假說棄却)는 결론을 내릴 수 있다. 그러나 有意水準 $\alpha=0.01$에서 檢定한다면 그 결론은 "동전이 균질하지 않다고 단정할 수 없으며 시험결과 앞면이 40회 나타난 것은 다소 드문 현상이 우연히 일어난 것으로 볼 수도 있다"로 내려진다.

〈정의〉 계산된 확율값이 有意水準 α보다 적으면 有意的(significant) 하다고 한다. 즉"유의적"이란 일상적으로 드물게 나타나는 유의할만한 현상이란 뜻이다.

〈정의〉 歸無假說과 對立假說

어떤 假說을 증명하기 위해 두개의 대립되는 假說을 설정하며 그 중 기각할 목적으로 설정된 假說을 假無假說 H_0(null hypothesis)라고 하고 나머지 假說을 對立假說 H_1(alternative hypothesis)라 한다.

假說檢定에서 歸無假說이 기각된다면 對立假說H_1이 채택된다. 그러므로 假說檢定의 절차로서 증명하고자 하는 對立假說과 반대의 歸無假說을 설정한 다음에 歸無假說이 부정되는 것을 보임으로써 간접적으로 對立假說이 옳다는 것을 증명한다.

〈정의〉 第Ⅰ種過誤와 第Ⅱ種過誤

檢定된 假說이 실제로는 옳은 것인데도 불구하고, 표본으로부터 얻은 결과만으로 내린 결론에서 假說이 그릇된 것이라고 판정하는 경우를 第Ⅰ種過誤α(type Ⅰ error)라 한다. 반면에실제로는 그릇된 假說임에도 불구하고 이를 옳은 것이라고 판정할 경우 이를 第Ⅱ種過誤β(type Ⅱ error)라 한다. 특히 第Ⅰ種過誤가 일어날 비율을 危險率이라 한다.

第Ⅰ種 및 第Ⅱ種過誤를 범할 確率 α, β는 다음과 같이 표시할 수 있다.

$$\alpha = P\{제\ \mathrm{I}\ 종과오\} = P\{H_0\ 기각 | H_0\} \tag{41}$$

$$\beta = P\{제\ \mathrm{II}\ 종과오\} = P\{H_0\ 채택 | 그릇된 H_0\} \tag{42}$$

母集團의 母數에 관한 假說檢定에서의 棄却域決定은 통계처리자의 자의적 선택에 의하여 이루어진다. 그러나 I, II 種過誤는 한쪽의 確率α를 적게 하려고 하면 다른 쪽의 確率β가 증가하는 배타적 관계가 있으며 적절한 타협점을 찾아 기각역을 결정해야 한다. 일반적인 방법으로는 제 I 종 과오가 제 II 종 과오보다 더 중대한 문제를 야기시키므로 제 I 종과오확률α를 통상 5% 또는 1%정도로 적게 취하되 제 2 종과오확률 β를 충분히 적게 할 수 있도록 기각역을 설정하는 것이 檢定의 관례로 되어 있다.

다음은 표본분산에 관한 χ^2-分布檢定方法을 예시한 것이다.

〔例〕 어느 業種에 대한 제조업체 근로자 평균임금 분포의 母分散 σ_0^2은 26,000〔원〕^{2}이었다. 어느 確率標本 20명에 대한 標本의 不偏分散은 $s^2=31,000$〔원〕^{2}이었다. 이 때 5% 有意水準에서 이 標本이 母分散 $\sigma_0^2=26,000$〔원〕^{2}의 母集團에서 抽出된 標本이라고 할 수 있는가를 檢定하라.

〔풀이〕 標本에서 얻은 不偏分散이 母分散 σ_0^2 보다 크므로 우측 검정을 행하는 것이 바람직하다. 절차는 다음과 같다.

① 假說設定 : $\begin{cases} H_0 : \sigma^2 = \sigma_0^2\ (단\ \sigma^2 은 표본의 모분산) \\ H_1 : \sigma^2 > \sigma_0^2 \end{cases}$

② 有意水準 : $\alpha = 0.05$, 자유도 : 19

③ 檢定統計量 :

$$\chi_0^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{19 \times 31,000}{26,000} = 22.65$$

$$P\{s^2 \geq 31,000\} = P\{\chi^2 \geq \chi_0^2\} \doteqdot 0.27 > 0.5$$

④ 結論 : 불편분산 s^2이 31,000이상이될 확률이 有意水準보다 그므로 假說 H_0를 부정할 수 없다. 즉 이 標本은 분산 $\sigma_0^2=26,000$의 모집단으로부터 취한 표본이 아니라고 말할 수 없다.

3.5 假說檢定에 의한 狀態推定不良情報檢出

狀態推定에서의 線形化된 測定方程式은 다음과같다.

$$\underline{z} = H\underline{x} + \underline{v} \tag{43}$$

단 $H = \dfrac{\partial h(x_0)}{\partial x}$: 운전점 $\underline{x}_0$에서의 Jacobian 행렬

$\underline{z}$: m-차원 측정치벡터

$\underline{x}$: n-차원 상태변수벡터

$\underline{v}$: m-차원 측정잡음벡터

(v_i는 $N(0, \sigma_i^2)$의 정규분포 잡음임)

위의 선형화 측정함수를 정규화하면 다음 방정식을 얻는다.

$$\underline{z}^N = H_N \underline{x} + \underline{v}^N \tag{44}$$

단 $\underline{z}^N = R^{-1/2}\underline{z}$, $H_N = R^{-1/2}H$

$\underline{v}^N = R^{-1/2}\underline{v}$, $R = E[\underline{v}\ \underline{v}^T] = \mathrm{diag}(\sigma_1^2, \sigma_2^2 \cdots, \sigma_m^2)$

그러면 定規化·雜音 v_i^N는 $N(0, 1)$의 定規分布를 갖게 되며 m개의 定規化 測定誤差$\{v_1^N, v_2^N, \cdots, v_m^N\}$는 $N(0, 1)$의 定規分布無限母集團에서 m개의標을 취한 것으로 생각할 수 있다. 따라서 다음과 같이 정의되는 狀態推定評價函數 J는 自由度$(m-n)$의 χ^2-分布를 이루게 된다(여기서는 評價函數J를 하나의 확율변수로 취급한다)

$$J = [\underline{z}^N - H^N\underline{\hat{x}}]^T[\underline{z}^N - H^N\underline{\hat{x}}] = \sum_{i=1}^{n} (v_i^N)^2 \tag{45}$$

(단 $\hat{x}$는 상태 x의 추정치임)

이상으로부터 不良情報 存在與否削定問題는 입수된 標本 $\{z_1^N, z_2^N, \cdots z_m^N\}$이 $N(0, 1)$의 N母集團으로부터 抽出된 標本이라고 볼 수 있는지 여부를 假說檢定(hypothesis test)를 통하여 檢定하는 問題로 定式化되며 檢定節次는 假說設定, 假說檢定 그리고 不良情報의 存在가 확인되면 不良情報除去 등의 3 단계로 나눌 수 있다.

① 假說設定

$\begin{cases} 歸無假說\ H_0 : 不良情報가\ 存在하지\ 않는다. \\ 對立假說\ H_1 : 不良情報가\ 存在한다. \end{cases}$

- 有意水準決定 : 일반적으로 $\alpha=0.05$ 또는 $\alpha=0.01$로 정

② 假說檢定

假說檢定은 歸無假說 H_0가 옳다는 가정하에서 실험결과가 나타날 확율을 계산함으로써 假說의 옳고

그름을 판정한다. 따라서 먼저 H_0가 옳다는 가정하에서 狀態推定을 행하고 式(45)에 의거 J를 계산한다. 이 때 계산된 J의 값을 J_{cal}이라 하면 J_{cal}은 주어진 표본, 즉 입수된 測定値에 대한 χ^2값과 같다. 만약 J_{cal}이 χ^2의 평균치 즉 $(m-n)$보다 적다면 이것은 測定誤差가 충분히 적다는 것을 뜻하기 때문에 H_0를 채택한다. 반면에 J_{cal}이 $(m-n)$보다 크면 J가 J_{cal}보다 크게 될 확율 P를 계산한다. (χ^2 - 分布의 右測檢定)

$$P = P\{J > J_{cal}\} = P\{\chi^2 > J_{cal}\} \tag{46}$$

위에서 계산된 확율 P가 α보다 크면 H_0를 채택하고 그렇지 않으면 H_1을 채택한다.

그러나 式(46)에서의 확율계산이 어려우므로 통상적으로 有意水準 α에 대한 χ^2_α 값〔節3.2 참조〕을 계산해 두고 다음과 같이 가설채택을 결정한다.

(a) $Jcal < \chi^2_\alpha$이면 H_0를 채택

(b) $Jcal \geq \chi^2_\alpha$이면 H_1을 채택

自由度$(m-n)$이 큰 경우에 대한 χ^2_α의 계산은 定規分布에 의한 近似計算法을 사용할 수 있다.(節 3.3 참조)

만약 對立假說H_1이 채택되면 不良情報가 존재한다고 결론을 내릴 수 있으므로 다음의 절차에 따라 不良情報를 推出해야 한다.

③ 不良情報抽出

不良情報抽出은 節2-2에 소개된 바 있는 방법중 어느 방법이나 적용이 가능하지만 보편적으로 定規化殘留偏差에 의한 不良情報抽出法을 사용한다. 이 경우 殘留偏差가 큰 순으로 不良情報를 제거하며 그 때마다 평가함수 J를 다시 계산하여 J가 χ^2_α 보다 적을 때까지 不良情報除去를 계속한다.

이상의 설명에서 假說檢定 및 不良情報抽出에 대해 단지 第1種過誤確率(危險率) α만을 고려하였으나 실제 응용에서는 전술한 바 있는 第2種過誤確率 β를 동시에 결정하고 있음을 부언해 둔다. 本稿에서는 지면관계상 이에 관한 이론을 직접 기술할 수 없음을 유감으로 생각하며 관심있는 독자는 참고문헌 〔1, 2, 5〕를 참고하길 부탁드리는 바이다.

4. 結 言

狀態推定에서 不良情報抽出의 중요성은 널리 인정되어 왔으나, 이 분야의 국내 연구가 활성화되지 못하였으며 그 이유는 관련이론의 복잡성에 기인한다고 생각된다. 특히 假說檢定에 의한 不良情報抽出은 이론이 복잡다양하여 통계학 전문서적을 읽어도 직접 관련성을 이해하기 어려우며 또한 이에 관련된 이론을 정리하여 소개한 책자도 거의 없는 실정이다.

필자는 본 稿를 통하여 假說檢定에 의한 不良情報抽出의 원리과 이론을 일관성있게 소개할 수 있게 된 것을 기쁘게 생각한다. 지면관계상 다소 미흡한 점이 있지만 관심있는 독자들에게 不良情報假說檢定의 원리이해에 가이드(guide)가 될 수 있길 바라며 앞으로 이 분야에 많은 관심이 모아지길 기대해 본다.

참 고 문 헌

1) L. Milli, Th. Van Cutsem, M. Ribbens-Pavella "Hypothesis Testing Identification-A new method for bad data analysis in power system state estimation", IEEE Trans. on PAS, Vol. PAS-103 No. 11, Nov. 1984.

2) L. Milli, Th. Van Cutsem, M. Pibbens-Pavella, "Bad Data Identification Method in Power System State Estimation-A comparative study", IEEE/PES Winter Meeting, Paper #WM 060-9, 1985.

3) 일본 電力中央研究所-總合報告 102 "電力系統의 狀態推定法의 開発", 電力中央研究所, 電力技術研究所, 昭和55년 3월.

4) Papoulis, Probility, Random Variables and Stochastic Process, MCGraw-Hill, 1965.

5) 金正年, 統計學〔增補版〕, 經文社, 1985.