ON THE DETERMINATION OF DOWNSIDE FINANCIAL LOSS OF INSTRUMENTATION NETWORKS IN THE PRESENCE OF GROSS ERRORS

Duy Quang NguyenThanh.[a], Kitipat Siemanond[a], and M. Bagajewicz[b,c]
[a] The Petroleum and Petrochemical College, Chulalongkorn University,
Bangkok 10330, Thailand.
[b] University of Oklahoma, 100 E. Boyd St., T-335, Norman, OK 73019, USA.
E-mail:bagajewicz@ou.edu
[c] Corresponding Author

Expressions for assessing the expected financial loss associated to the accuracy of the instrumentation and the associated probability in the presence of two and more gross errors have been developed in previous work. However, these expressions were given in a form of integrals without closed form solution. This paper presents a generalized method to calculate the expected financial loss and the associated probability when two or more gross errors present in a system. An example is provided.

Keywords: Instrumentation Network Design, Data Reconciliation, Plant Monitoring.

1. Introduction

Errors in measured data undoubtedly lead to deterioration in plant performance. Data reconciliation used together with gross error detection technique can be used to reduce variability and increase accuracy. Recently, accuracy of estimators was defined as the sum of the precision and the maximum undetectable induced bias in a stream (Bagajewicz, 2004a). In addition, a new approach to assess the economic value of precision was presented (Bagajewicz et al., 2003; 2004). In a follow up paper, the economic value of accuracy was discussed (Bagajewicz, 2004b). Specifically, these papers developed expressions to obtain the expected downside financial loss associated to the precision and accuracy of the instrumentation and associate differences of financial loss to the value of added instrumentation. While Bagajewicz (2004b) was able to derive analytical expressions for the financial loss under the presence of one gross error, he did not develop analytical forms of the integral expressions in the presence of more than one gross error, which consist of integrals with discontinuous integrands that do not reduce to a closed form solution.

This paper presents a method that allows the approximate calculation of these integral expressions in the presence of more than one gross error.

2. Background

Bagajewicz et al. (2003, 2004) were able to obtain expressions for assessing the economic value of precision. A formula was developed for such value based on the downside expected loss that occurs when an operator adjusts the throughput of a plant when the measurements or estimators suggest that the targeted production is met or surpassed. However, there is a finite probability that the measurement or estimator is above the target when in fact the real flow is below it. Bagajewicz et al. (2003, 2004) showed that the probability of not meeting the targeted production is 0.25. They also developed an expression for expected downside financial loss associated to this probability. The general expression obtained is the following:

$$DEFL(\hat{\sigma}_p,\sigma_p)=\int_{-\infty}^{m_p^*} g_p(m_p,m_p^*,\sigma_p)\left\{K_s T\int_{-\infty}^{m_p^*}(m_p^*-\hat{m}_p)g_M(\hat{m}_p,m_p,\hat{\sigma}_p)d\hat{m}_p\right\}dm_p \qquad (1)$$

where g_M is the probability distribution of the measurements ($\hat{m}_p$) of the flowrate of the pth product stream around their mean m_p and g_p the distribution of the true value m_p around its mean m_p^* (targeted value); K_s is the cost of the product (or the cost of inventory) and T is the time window of analysis. Under simplified assumptions of negligible process variations (i.e., $\sigma_p/\hat{\sigma}_p << 1$) and normal distributions, the equation (1) reduces to $DEFL = 0.19947\ K_s T\hat{\sigma}_p$ (Bagajewicz et al. 2004).

While precision is important, most instruments present biases. To understand how biases corrupt measurement accuracy, Bagajewicz (2004a) has introduced the concept of software accuracy, which is based on the notion that data reconciliation with some test statistics is used to detect biases. If biases are too small to be detected, they smear all the estimators, including those of the variables for which the corresponding instruments have no bias, called induced bias. Thus, the accuracy of an estimator is defined as the sum of precision of the estimator plus the maximum possible undetected induced bias in that variable. Using the above concept, Bagajewicz (2004b) developed the theory of economic value of accuracy. He assumed that, when an instrument fails, which happens with a certain probability $f_i(t)$ (a function of time), and that the size of the bias follows a certain distribution $\mathrm{h_i}(\theta_\mathrm{i},\bar{\delta}_\mathrm{i},\rho_\mathrm{i})$ with mean $\bar{\delta}_i$ and variance ρ_i^2. Thus, one needs to integrate over all possible values of the gross errors and multiply by the probability of such bias to develop. If it is assumed that one instrument fails at a time, then the probability of n_b instruments failing and the others not is given by: $\Phi_{i1,i2,\ldots,in_b}^{n_T}=f_{i1}(t)\ldots f_{in_b}(t)\prod_{s\neq i1,\ldots s\neq in_b}[1-f_s(t)]$. With all these assumptions, Bagajewicz (2004b) obtained general expressions for the probability of missing the target, and the associated downside financial loss. He was also able to derive analytical form for the expressions in the presence of one bias, but did not provide analytical forms for the expressions for the presence of more than one bias because the integrals involved require integrating a discontinuous function that changes form in different regions.

3. Approximate Generalized Method

Under simplified assumptions (negligible process variation and normal distributions), the general expressions given by Bagajewicz (2004b) for the probability and the associated financial loss when two gross errors being present in the system reduce to:

$$P = \Phi_{i1,i2}^{2} \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \frac{1}{4} \left\{ \begin{array}{l} h_1(\theta_1,\bar{\delta}_1,\rho_1) h_2(\theta_2,\bar{\delta}_2,\rho_2) \ \text{(when no bias is detected)} \\ +[1+erf(A_1)] h_1(\theta_1,\bar{\delta}_1,\rho_1) h_2(\theta_2,\bar{\delta}_2,\rho_2) \ \text{(both biases are detected)} \\ +[1+erf(A_2)] h_1(\theta_1,\bar{\delta}_1,\rho_1) h_2(\theta_2,\bar{\delta}_2,\rho_2) \ \text{(only bias 1 is detected)} \\ +[1+erf(A_3)] h_1(\theta_1,\bar{\delta}_1,\rho_1) h_2(\theta_2,\bar{\delta}_2,\rho_2) \ \text{(only bias 2 is detected)} \end{array} \right\} d\theta_1 d\theta_2 \quad (2)$$

$$DEFL = \frac{K_s T}{2} \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \left\{ \begin{array}{l} \frac{\hat{\sigma}_p}{\sqrt{2\pi}} h_1(\theta_1,\bar{\delta}_1,\rho_1) h_2(\theta_2,\bar{\delta}_2,\rho_2) \ \text{(when no bias is detected)} \\ +\left(\frac{\hat{\sigma}_p e^{-A_1^2}}{\sqrt{2\pi}} - \frac{\sqrt{2}\hat{\sigma}_p A_1 \{1-erf(A_1)\}}{2} \right) \left(\begin{array}{c} h_1(\theta_1,\bar{\delta}_1,\rho_1) h_2(\theta_2,\bar{\delta}_2,\rho_2) \\ \text{both biases are detected} \end{array} \right) \\ +\left(\frac{\hat{\sigma}_p e^{-A_2^2}}{\sqrt{2\pi}} - \frac{\sqrt{2}\hat{\sigma}_p A_2 \{1-erf(A_2)\}}{2} \right) \left(\begin{array}{c} h_1(\theta_1,\bar{\delta}_1,\rho_1) h_2(\theta_2,\bar{\delta}_2,\rho_2) \\ \text{only bias 1 is detected} \end{array} \right) \\ +\left(\frac{\hat{\sigma}_p e^{-A_3^2}}{\sqrt{2\pi}} - \frac{\sqrt{2}\hat{\sigma}_p A_3 \{1-erf(A_3)\}}{2} \right) \left(\begin{array}{c} h_1(\theta_1,\bar{\delta}_1,\rho_1) h_2(\theta_2,\bar{\delta}_2,\rho_2) \\ \text{only bias 2 is detected} \end{array} \right) \end{array} \right\} d\theta_1 d\theta_2 \quad (3)$$

where $A_1 = \frac{\alpha_1\theta_1 + \alpha_2\theta_2}{\sqrt{2}\hat{\sigma}_p}$, $A_2 = \frac{\alpha_2\theta_2}{\sqrt{2}\hat{\sigma}_p}$ and $A_3 = \frac{\alpha_1\theta_1}{\sqrt{2}\hat{\sigma}_p}$. In addition α_1 and α_2 are parameters in the expressions for the induced bias: $\hat{\delta}_p^{i1,i2}(\theta_1,\theta_2) = \alpha_1\theta_1 + \alpha_2\theta_2$; $\hat{\delta}_p^{i1}(\theta_1) = \alpha_1\theta_1$; $\hat{\delta}_p^{i2}(\theta_2) = \alpha_2\theta_2$.

In order to evaluate the above integrals, one needs to identify the different regions according to the existence of undetected gross errors in the system. If a serial elimination strategy is used to detect gross errors based on the maximum power MT test, the test statistics for the two measurements i1, i2 are given by: $Z_1^{MP} = \frac{|W_{i1i1}\theta_1 + W_{i1i2}\theta_2|}{\sqrt{W_{i1i1}}}; Z_2^{MP} = \frac{|W_{i2i1}\theta_1 + W_{i2i2}\theta_2|}{\sqrt{W_{i2i2}}}$; where $W = A^T(ASA^T)^{-1}A$. The threshold value for the MT test statistics is ξ (which usually is 1.96). In the serial elimination strategy, if one gross error has been detected, the corresponding measurement is eliminated. Then, at the next stage, the new test statistics for the two measurements reduce to $Z_1' = \frac{|W'_{i1i1}\theta_1|}{\sqrt{W'_{i1i1}}}; Z_2' = \frac{|W''_{i2i2}\theta_2|}{\sqrt{W''_{i2i2}}}$, respectively ($W'$, W'' are the updated matrix W after one measurement has been eliminated). Thus, these expression define the boundary lines for the different regions, which are shown in Figure 1.

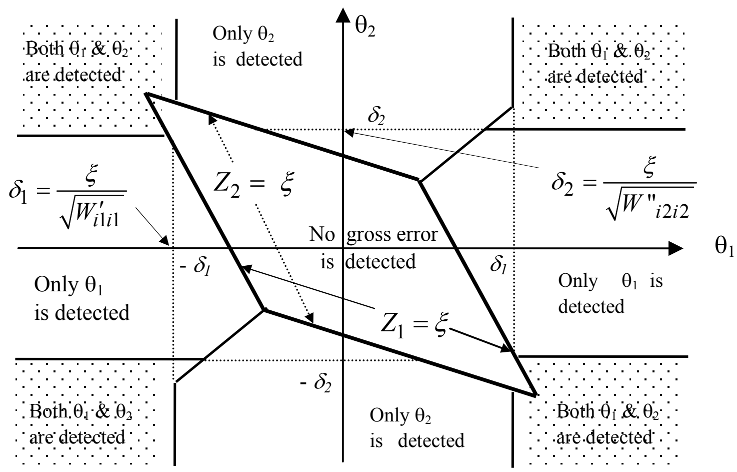

Figure 1. *Different regions when two gross errors are present in the system*

3.1 Approximation method

To calculate P and DEFL, one needs to evaluate integrals at the form

$$J=\int_{a}^{b} e^{-z^2}\ erf\left\{z\right\}dz \tag{4}$$

The essence of the approximation method is to replace one of the integrand functions by a piece wise linear function. Thus if the error function is replaced by a linear expression, J becomes

$$J^{*}=\sum\int_{a_i}^{b_i}(y_{1i}+y_{2i}z)e^{-z^2}\ dz \tag{5}$$

which can be calculated analytically. Certain choices of piece wise linear functions guarantee that the answer is an underestimate J^{*}_{L} or an overestimate J^{*}_{U} of J (figure 2). Thus, the integral J can be approximated as the average of J^{*}_{L} and J^{*}_{U}, that is:
$J = (J_L^{*} + J_U^{*})/2$. Because J_U^{*} and J_L^{*} are overestimator and underestimator of J, then, the maximum error is given by $(J_U^{*} - J_L^{*})$.
Finally, the expression for the probability is given by:

$$P=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} Fd\theta_1 d\theta_2=\int_{-K_1}^{K_1}\int_{-K_2}^{K_2} Fd\theta_1 d\theta_2(P_4)+\left\{\int_{-\infty}^{-K_1}+\int_{K_1}^{\infty}\right\}\left\{\int_{-\infty}^{-K_2}+\int_{K_2}^{\infty}\right\}Fd\theta_1 d\theta_2(P_1)$$
$$+\left\{\int_{-\infty}^{-K_1}+\int_{K_1}^{\infty}\right\}\int_{-K_2}^{K_2}Fd\theta_1 d\theta_2(P_2)\ +\int_{-K_1}^{K_1}\left\{\int_{-\infty}^{-K_2}+\int_{K_2}^{\infty}\right\}Fd\theta_1 d\theta_2(P_3) \tag{6}$$

Where P_1, P_2, P_3, and P_4 are the regions indicated in figure 3.

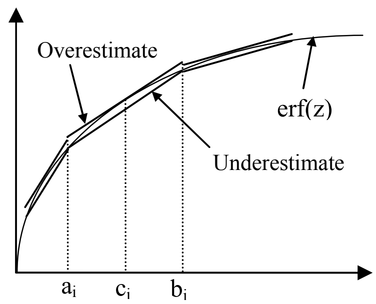

Figure 2. *Approximation method*

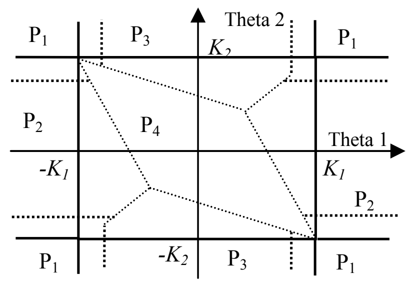

Figure 3. *Regions used in calculation*

Clearly, each region has a different integrand, and therefore some, like P_1, can be calculated analytically, but others need to be calculated approximately. The above procedure to calculate integral expressions can be extended to calculate P & DEFL when more than two gross errors are present in the system. The principle is to partition the space of variables into regions in the same way, but in more dimensions.

4. Example

Consider the following process which was used by Bagajewicz (2004a):

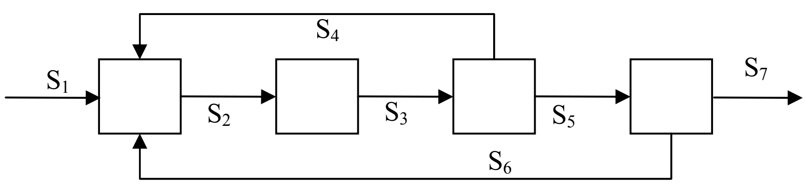

Figure 4. *Example process*

We assume that all variables are measured. The variance-covariance matrix of measurements S = diag(1.0,0.2,1.0,1.0,1.0,1.0,1.0). We also assume that biases have zero means and standard deviations ρ_1 = 3.0, ρ_2 = 4.0, ρ_3 = 4.0, ρ_4 = 5.0, ρ_5 = 6.0, ρ_6 = 5.0. The results (Table 1&2) were obtained using an Intel 2.4 GHz processor and 1024 MB RAM memory. In Table 1 and 2, different interval sizes for the intervals used are indicated and the maximum error $E=(J_U^* - J_L^*)$ is also reported. The probability when two and three gross errors are present in the systems has also been calculated to be about $\Phi^2_{i1,i2}\mathbf{x}0.25$ and $\Phi^3_{i1,i2,i3}\mathbf{x}0.25$ (for all cases); respectively. This probability is viewed as the confidence with which the financial loss as shown below is known. When more gross errors are present, DEFL increases (which is expected). Moreover, DEFL in the presence of biases is larger than DEFL without biases ($DEFL^0$ = $0.19947K_sT$). Clearly we see that $DEFL/K_sT > 0.19947 = DEFL^0/K_sT$. The time is also reported.

Table 1. DEFL/ K_sT for two gross errors present in the system

Locations of biases	Interval size = 0.1			Interval size = 0.2		
	Solution	Time (millisecond)	Error	Solution	Time (millisecond)	Error
1, 2	0.2162	10	$0.57*10^{-5}$	0.2162	3	$2.29*10^{-5}$
1, 3	0.2148	11	$0.47*10^{-5}$	0.2148	2	$1.85*10^{-5}$
2, 3	0.2017	69	$0.01*10^{-5}$	0.2017	17	$0.03*10^{-5}$
2, 5	0.2024	17	$0.03*10^{-5}$	0.2024	3	$0.09*10^{-5}$
3, 6	0.2057	12	$0.08*10^{-5}$	0.2057	4	$0.32*10^{-5}$
4, 5	0.2098	52	$0.05*10^{-5}$	0.2098	13	$0.19*10^{-5}$
5, 6	0.2051	38	$0.03*10^{-5}$	0.2052	9	$0.13*10^{-5}$

Table 2. DEFL/ K_sT for gross errors present in the system (interval size = 0.2)

Locations of biases	Solution	Time (second)	E	Locations of biases	Solution	Time (second)	E
1, 2 , 3	0.2217	10	$3.24*10^{-5}$	1, 3, 6	0.2489	3	$3.93*10^{-5}$
1, 2, 4	0.2192	4	$1.92*10^{-5}$	1, 4, 5	0.2577	19	$4.75*10^{-5}$
1, 2, 5	0.2279	3	$4.8*10^{-5}$	1, 4, 6	0.2517	7	$4.71*10^{-5}$
1, 2, 6	0.2498	3	$4.04*10^{-5}$	2, 3, 5	0.2127	53	$0.51*10^{-5}$

5. Discussion & conclusions

When the interval size decreases, accuracy increases (which is preferred) but computation time also increases. The “optimum” interval size should be chosen as a compromise between accuracy and computation time. When the number of gross errors increases one (from 2 to 3), the computation time increases significantly. On the same interval size basis, computation time increases 50-100 times. Short computation time are important because the financial loss calculation is used in sensor networks design which needs to explore many alternatives combinatorially. Future work will include comparisons with Monte Carlo-based integration methods.

References

Bagajewicz, M., 2004a, On the Definition of Software Accuracy in Redundant Measurement Systems. To appear. AIChE J., (available at http://www.ou.edu/class/che-design/Unpublished-papers.htm).

Bagajewicz, M., 2004b, Value of Accuracy in Linear Systems. Submitted to AIChE Journal, (available at http://www.ou.edu/class/che-design/Unpublished-papers.htm).

Bagajewicz M. and M. Markowski. *Instrumentation Design and Upgrade using an Unconstrained Method with Pure Economical Objectives.* FOCAPO 2003. (Foundations of Computer Aided Process Operations) Coral Springs, FL, USA, January , (2003).

Bagajewicz, M., Markowski, M., and Budek, A., 2004, Economic Value of Precision in the Monitoring of Linear Systems, To appear. AIChE Journal (available at http://www.ou.edu/class/che-design/Unpublished-papers.htm).

Acknowledgements: Funding from US-NSF. Grant CTS-0350501 is acknowledged.