# *Issues of Estimation in the Monitoring of Constant Flow Continuous Streams*

# ISSUES OF ESTIMATION IN THE MONITORING OF CONSTANT FLOW CONTINUOUS STREAMS

N. S. BARNETT AND S. S. DRAGOMIR

Abstract. This paper deals with some fundamental matters pertaining to estimation of critical quantities associated with continuous processes which are frequently related to the quality rating of the product. Specifically, it examines bounds on estimation and bounds on the estimation error variance. It draws on recent results from the theory of mathematical inequalities and their applications.

## 1. Introduction

In the application of statistical techniques to the monitoring and control of industrial processes it is possible to identify two classes of processes - discrete and continuous. A large body of statistical work has been accumulated and disseminated dealing with discrete processes and, frequently, continuous processes are treated as if, in fact, they were discrete. However, this latter approach is not always advisable.

In collecting data from a continuous process, scrutiny will frequently show it to be correlated, which, in general, necessitates a different approach being taken to analysis and interpretation than is generally taken for discrete processes. Discrete processes are characterized by the availability of data which is often times uncorrelated. Discrete processes are generally controlled on the basis that they can be maintained in a state of control. Techniques used and actions taken to correct perceived problems are heavily dependent on this assumption which implies the pre-eminence of the Gaussian (Normal) distribution.

Continuous processes are frequently unable to be controlled in the sense that discrete ones can and sensible approaches to assessment and control draw on more sophisticated statistical techniques. Frequent re-course is made to pre-programmed automated controllers that continually adjust the process to meet stipulated requirements. Such controllers require certain assumptions to be made and need to be properly tuned if they are to obtain appropriate outcomes.

In any process, whether discrete or continuous, data needs to be collected carefully and assumptions drawn with caution so that appropriate techniques of analysis are engaged. Only then is there a likelihood that sound judgements will be made in relation to monitoring, control and quality assessment of the product.

Industrial processes that are classified as continuous are prevalent in the chemical industry. A continuous process deals with products that are not identifiable as discrete entities. Typically, product is liquid, gaseous or fine granular in nature and an item of product only exists in relation to containers in which it is stored

*Date*: June, 1999.

1991 *Mathematics Subject Classification.* Primary 62Xxx; Secondary 26D15.

*Key words and phrases.* Continuous Process, Ostrowski's Inequality .

or despatched. Chemical processes are infamous for having aspects that defy explanation and for being sensitive to apparently insignificant changes in parameters. These all present their own peculiar set of problems and difficulties.

This paper deals with some fundamental matters pertaining to estimation of critical quantities associated with continuous processes and which are frequently related to the quality rating of the product. Specifically, it examines bounds on estimation and bounds on the estimation error variance. It draws on recent results from the theory of mathematical inequalities and their applications.

## 2. Practical Considerations

In sampling a continuous stream with a view to quality assessment, it is generally necessary to estimate the average flow quality, $\bar{X}$, of the stream over a particular time by taking a number of 'grab' samples that return values $x_1, x_2, x_3, ..., x_n$ and which are collected within the same time frame. These are assembled into an average $\bar{x}$ which is then used to estimate $\bar{X}$. Subsequent quantities that are of some interest include $\left|\bar{X} - \bar{x}_n\right|$, $E\left|\bar{X} - \bar{x}_n\right|$ and $E\left[\bar{X} - \bar{x}_n\right]^2$ which are respectively the estimation error, the expectation of the estimation error and the estimation error variance. The first is a purely mathematical quantity but the latter two contain information related to the stochastic nature of the process.

'Grab' samples are frequently a small container of product assumed collected in an instant during or at the conclusion of manufacture. Sometimes, however, a single sample may take an appreciable time to collect. Under these latter circumstances, a single sample can be reasonably considered, itself, a direct measure of average flow so that we are, in effect, estimating the average flow in say $(0, T)$, $\bar{X}(T)$ by the average flow in say $(s, s+p)$, $0 < s + p < T$, $\bar{X}(p)$.

Estimation of $\left|\bar{X} - \bar{x}_n\right|$, $E\left|\bar{X} - \bar{x}_n\right|$ and $E\left[\bar{X} - \bar{x}_n\right]^2$, and for this latter case, $E\left[\bar{X}(T) - \bar{X}(p)\right]$, relate to the veracity of making judgements on the basis of $\bar{x}_n$ and so, in this respect, have some importance. Factors which have an impact on them, include the magnitude of $n$, the times at which the samples are taken and, for the latter two, the intrinsic nature of the continuous stream itself. This latter includes its stochastic behaviour and the flow rate of the stream, if this, in fact, varies [1].

One approach to characterizing the stochastic behaviour of the stream is to describe it by its variogram. For a process with a stationary variogram, this is defined as:-

$$V(u) = \frac{1}{2} E\left[(X(t) - X(t+u))^2\right],$$

where, for current purposes, $X(t)$ is assumed to be a stochastic process continuous over time, with

$$V(0) = 0 \text{ and } V(-u) = V(u).$$

## 3. Previous Work

Whilst, in general, the average flow 'quality', $\bar{X}(t)$ over a time period $(0, T)$ will be estimated by a sample average, $\bar{X}_n$, when this procedure is examined more closely it is apparent that each sample point can be considered to estimate the average flow in its immediate neighborhood. This being the case, in seeking to establish the most appropriate points at which to sample, it is reasonable to consider $E\left[\left(\bar{X} - X(t)\right)^2\right]$

where $t$ is a single sampling time (assumed instantaneous). Barnett et al [1] did this and obtained optimal times at which to sample in order to minimize the estimation error variance for particular flow rates and variograms. In [2], Barnett and Dragomir obtained bounds for the same quantity for a class of variograms and constant flow rate. In so doing, they used a recent development of Ostrowski's integral inequality [3]. The class of variograms for which there exists such an inequality was extended to the Hölder type by Barnett, Dragomir and Gomm [4].

This current paper obtains a bound for the estimation error when sampling is not instantaneous, gives a bound for the estimation error variance when sampling is instantaneous and illustrates how the sample size can be determined by stipulations on this error variance.

## 4. Estimation Error

In [5], the authors prove the following inequality for a differentiable function

$$\begin{aligned}\left|\int_a^b f(x)\,dx-\sum_{i=0}^{n-1} f(\xi_i)\,h_i\right| &\le \|f'\|_\infty \sum_{i=0}^{n-1}\left[\frac{h_i^2}{4}+\left(\xi_i-\frac{x_i+x_{i+1}}{2}\right)^2\right]\\ &\le \frac{\|f'\|_\infty}{2}\sum_{i=0}^{n-1} h_i^2,\end{aligned}$$

where $\|f'\|_\infty := \sup_{t\in(a,b)}|f'(t)| < \infty,\ a = x_0 < x_1 < ... < x_{n-1} < x_n = b$ is an arbitrary partition of $[a,b]$ and $h_i = x_{i+1}-x_i, \xi_i \in [x_i, x_{i+1}],\ i = 0,1,2,...,n-1$.

If, for current purposes, $f(t)$ is chosen to be the stochastic process $X(t)$, then clearly $\bar{X} = \frac{1}{T}\int_0^T X(t)\,dt$ and $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Thus, taking $b = T,\ a = 0,\ \xi_i = \frac{x_i+x_{i+1}}{2}$ and $x_i = i-\frac{T}{n}$, it is possible to obtain

$$\begin{aligned}\left|\bar{X}(T)-\bar{X}_n\right| &= \left|\frac{1}{T}\int_0^T X(t)\,dt-\frac{1}{n}\sum_{i=0}^{n-1} X_{t_i}\right|\\ &\le \frac{\|X'(t)\|_\infty T}{2n},\end{aligned}$$

which provides an upper bound for the estimation error for a particular class of variograms ($\bar{X}(T)$ is being used equivalently for $\bar{X}$ when the time duration is being emphasized).

For the case where sampling is not instantaneous, the estimation error can be considered to be

$$\left|\bar{X}(T)-\bar{X}(p)\right|,$$

where

$$\bar{X} = \frac{1}{T}\int_0^T X(t)\,dt,$$

as previously, and

$$\bar{X}(p) = \frac{1}{p}\int_s^{s+p} X(t)\,dt,$$

where $s$ is the time at which sampling commences and $s+p$ is the time at which it is complete. To obtain a bound for the error, we require a further extension of Ostrowski's inequality given by the following lemma.

**Lemma 1.** *If $f:[a,b]\to\mathbb{R}$ is an absolutely continuous mapping on $[a,b]$, $[c,d]\subseteq[a,b]$ and*

$$\|f'\|_\infty := \sup_{t\in(a,b)}|f'(t)| < \infty,$$

*then*

$$\begin{aligned}
&\left|\frac{1}{b-a}\int_a^b f(t)\,dt - \frac{1}{d-c}\int_c^d f(s)\,ds\right| \\
\leq\ &\left\{\frac{1}{4}(b-a)+\frac{(d-c)}{2}+\frac{1}{b-a}\left[\left|\frac{c+d}{2}-\frac{a+b}{2}\right|-\frac{d-c}{2}\right]^2\right\}\|f'\|_\infty .
\end{aligned}$$

*Proof.* Ostrowski's inequality for absolutely continuous mappings is

$$\left|f(x)-\frac{1}{b-a}\int_a^b f(t)\,dt\right| \leq \|f'\|_\infty\left[\frac{1}{4}+\left(\frac{x-\frac{a+b}{2}}{b-a}\right)^2\right](b-a)$$

for all $x\in[a,b]$.
By the triangle inequality we have:

$$\begin{aligned}
&\left|\frac{1}{b-a}\int_a^b f(t)\,dt - \frac{1}{d-c}\int_c^d f(s)\,ds\right| \\
\leq\ &\left|f(x)-\frac{1}{b-a}\int_a^b f(t)\,dt\right| + \left|f(x)-\frac{1}{d-c}\int_c^d f(s)\,ds\right| \\
\leq\ &\left[\frac{1}{4}+\frac{\left(x-\frac{a+b}{2}\right)^2}{(b-a)^2}\right](b-a)\|f'\|_\infty + \left[\frac{1}{4}+\frac{\left(x-\frac{c+d}{2}\right)^2}{(d-c)^2}\right](d-c)\|f'\|_\infty
\end{aligned}$$

for all $x\in[c,d]\subseteq[a,b]$.
Right hand side is

$$\frac{1}{4}\|f'\|_\infty(b-a+d-c)+\|f'\|_\infty\left[\frac{\left(x-\frac{a+b}{2}\right)^2}{(b-a)}+\frac{\left(x-\frac{c+d}{2}\right)^2}{(d-c)}\right].$$

If this is denoted by $h(x)$, it is apparent that

$$\inf_{x\in[c,d]} h(x) = \min\{h(c),h(d),h(u)\},$$

where $u$ is the turning point of

$$\begin{aligned}
y\ =\ &\frac{1}{4}(b-a+d-c)+\left(\frac{1}{b-a}+\frac{1}{d-c}\right)x^2 \\
&-\left(\frac{a+b}{b-a}+\frac{c+d}{d-c}\right)x+\frac{1}{b-a}\left(\frac{a+b}{2}\right)^2+\frac{1}{d-c}\left(\frac{c+d}{2}\right)^2 .
\end{aligned}$$

So

$$\left|\frac{1}{b-a}\int_a^b f(t)\,dt - \frac{1}{d-c}\int_c^d f(s)\,ds\right| \leq \min\{h(c),h(d),h(u)\}.$$

Now,

$$\min\{h(c),h(d)\} = \frac{1}{4}\|f'\|_\infty (b-a+d-c) + \|f'\|_\infty \frac{(d-c)}{4} + \frac{\|f'\|_\infty}{b-a}\min\left\{\left(c-\frac{a+b}{2}\right)^2, \left(d-\frac{a+b}{2}\right)^2\right\}.$$

Simplifying,

$$\min\left\{\left(c-\frac{a+b}{2}\right)^2, \left(d-\frac{a+b}{2}\right)^2\right\},$$

observe that this is:

$$\begin{aligned}
&\frac{1}{2}\left[\left(c-\frac{a+b}{2}\right)^2 + \left(d-\frac{a+b}{2}\right)^2 - \left|\left(c-\frac{a+b}{2}\right)^2 - \left(d-\frac{a+b}{2}\right)^2\right|\right] \\
= &\frac{1}{4}(d-c)^2 + \left(\frac{a+b}{2}-\frac{c+d}{2}\right)^2 - \frac{1}{2}\left|(c-d)(c+d-(a+b))\right| \\
= &\frac{1}{4}(d-c)^2 + \left(\frac{a+b}{2}-\frac{c+d}{2}\right)^2 - (d-c)\left|\frac{c+d}{2}-\frac{a+b}{2}\right| \\
&\left[\left|\frac{c+d}{2}-\frac{a+b}{2}\right| - \frac{d-c}{2}\right]^2,
\end{aligned}$$

and so

$$\begin{aligned}
&\left|\frac{1}{b-a}\int_a^b f(t)\,dt - \frac{1}{d-c}\int_c^d f(s)\,ds\right| \\
\leq\ &\|f'\|_\infty \left\{\frac{1}{4}(b-a) + \left(\frac{d-c}{2}\right) + \frac{1}{b-a}\left[\left|\frac{c+d}{2}-\frac{a+b}{2}\right| - \frac{d-c}{2}\right]^2\right\}
\end{aligned}$$

as required, since

$$\min\{h(c),h(d),h(u)\} \leq \min\{h(c),h(d)\}.$$

∎

For application of this result to estimation of the mean flow quality, take $a = 0$, $b = T$ and $c = s$, $d = s+p$ with $s+p < T$, with respect to the time, $s$, at which sampling commences it is interesting to note that with reference to the mid point of the time period over which it is desired to estimate $\bar{X}$, $(0,T)$, if sampling commences at the midpoint, i.e., $s = \frac{T}{2}$, the bound is

$$\left(\frac{s+p}{2}\right)\|X'(t)\|_\infty .$$

If sampling concludes at the midpoint, i.e., $s+p = \frac{T}{2}$, the bound is

$$\left(\frac{2s+3p}{4}\right)\|X'(t)\|_\infty$$

and the tightest bound is provided when sampling is symmetrical about the mid point of the time period, that is,

$$\frac{T}{2} = \frac{2s+p}{2},$$

in which case the bound is:-

$$\left|\bar{X}(T)-\bar{X}(p)\right|\leq\left(\frac{T+p}{4}\right)\left\|X'(t)\right\|_{\infty}.$$

## 5. Estimation Error Variance- Instantaneous Sampling

By application of the cubature formula in [3], and using the approach given in [4], it can be shown that for instantaneous sampling and estimation of the average of the continuous flow by sample average:

$$E\left[\left(\bar{X}-\bar{X}_n\right)^2\right]\leq\frac{d^2}{4}\left\|V''\right\|_{\infty},$$

provided the variogram is twice differentiable in $(-d,d)$ and where instantaneous samples are taken at $\frac{d}{2},\frac{3d}{2},...,\frac{(2n-1)d}{2},\ T=nd.$

The problem of sample size for the assessment of a continuous stream may well be resolved by restricting $E\left[\left(\bar{X}-\bar{X}_n\right)^2\right]$ and obtaining the smallest integer $n$ that satisfies the restriction. This is illustrated for the case where the variogram is linear. Set the condition $E\left[\left(\bar{X}-\bar{X}_n\right)^2\right]\leq M$. It can be shown [6] that

$$\begin{aligned}&E\left[\left(\bar{X}-\bar{X}_n\right)^2\right]\\ =\;&-\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}V\left(t_i-t_j\right)-\frac{1}{T^2}\int_0^T\int_0^T V\left(u-v\right)dudv\\ &+\frac{2}{nT}\int_0^T\sum_{i=1}^{n}V\left(u-t_i\right)du\end{aligned}$$

where $t_i$ are the times at which instantaneous samples are taken.

These are assumed to be equidistant apart since for a constant flowing stream this procedure has been shown to be optimal [1]. Hence $t_i=\frac{(2i-1)}{2}d,\ i=1,...,n.$

It can be shown further that

$$\begin{aligned}&-\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}V\left(t_i-t_j\right)\\ =\;&\frac{-2}{n^2}\left\{(n-1)V(d)+(n-2)V(2d)+...+V((n-1)d)\right\}\\ =\;&\frac{-2}{n^2}\left\{(n-1)V\left(\frac{T}{n}\right)+(n-2)V\left(\frac{2T}{n}\right)+...+V\left((n-1)\frac{T}{n}\right)\right\}.\end{aligned}$$

Now, for the case of a linear variogram we have:-

$$V(u)=A+Bu,$$

where for the current application we would expect both $A,B>0$.

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}V\left(t_i-t_j\right)$$

then simplifies down to

$$\frac{-A(n-1)}{n}-\frac{BT(n-1)(n+1)}{3n^2},$$

$-\frac{1}{T^2}\int_0^T\int_0^T V(u-v)\,dudv$ simplifies to $-A$, and $\frac{2}{nT}\int_0^T\sum_{i=1}^n V(u-t_i)\,du$ simplifies to $2A$.

The sample size sought is then the smallest $n$ such that

$$\frac{-A(n-1)}{n}-\frac{BT(n-1)(n+1)}{3n^2}+A\leq M,$$

which essentially reduces to solving a quadratic for which the solution is:-

$$n=\frac{3A+\sqrt{9A^2+4(3M+BT)BT}}{2(3M+BT)}.$$

## References


[1] N. S. Barnett, I. S. Gomm and L. Armour, Location of the optimal sampling point for the quality assessment of continuous streams, *Australian J. of Statistics,* **37**(2), 1995, 145-152.

[2] N. S. Barnett and S. S. Dragomir, A note on the bounds for the estimation error variance of a continuous stream with stationary variogram, *J. of Korea Society for Industrial and Applied Mathematics,* **2**(2), 1998, 49-56.

[3] N. S. Barnett and S. S. Dragomir, An Ostrowski type inequality for double integrals and applications to cubature formulae, *(to appear).*

[4] N. S. Barnett, S. S. Dragomir and I. S. Gomm, Further Bounds for the estimation error variance of a continuous stream with stationary variogram, *(to appear).*

[5] S. S. Dragomir and S. Wang, Applications of Ostrowski's inequality to the estimation of error bounds for some special means and some numerical quadrature rules, *Appl. Math. Lett.,* **11** (1998), 105-109.

[6] I. W. Saunder, G. K. Robinson, T. Lwin and R. J. Holmes, A simplified variogram method for the estimation error variance in sampling from a continuous stream, *Internat. J. Mineral Processing,* **25** (1989), 175-798.



School of Communications and Informatics, Victoria University of Technology, Melbourne, Australia., On leave at Salford University, Manchester, U.K.

(S. S. Dragomir), School of Communications and Informatics, Victoria University of Technology, PO Box 14428, Melbourne City, MC 8001 Australia

*E-mail address*: sever@matilda.vu.edu.au