Bruno Šubert

Error in generating a normal distribution

Aplikace matematiky, Vol. 12 (1967), No. 2, 130–135

Persistent URL: http://dml.cz/dmlcz/103079

ERROR IN GENERATING A NORMAL DISTRIBUTION

BRUNO ŠUBERT

(Received July 13, 1966.)

INTRODUCTION

For the purposes of Monte Carlo methods it is sometimes required to generate a stationary Gaussian stochastic process. However, there are no physically realizable means for generating such a process in a strict mathematical sense. Any physical method can only yield processes that more or less approximate a Gaussian process.

One of these methods consists in a filtration of a sequence of random pulses by a linear filter [2]. The sequence is formed by pulses of identical shape and width and random polarity. The pulses are supposed to be stochastically independent with both polarities equally probable. Such a sequence can be realized with considerable accuracy [1]. A linear time-invariant filter can likewise be realized, e.g. by analogue techniques. It is obvious that the probability distribution of the output signal at any finite time instant will be discrete. Nevertheless, it has been proved ([4]) that, as the time since the input sequence was applied tends to infinity and the pulse width tends to zero, the output signal approaches a stationary Gaussian process. Hence, the output signal can well approximate the desired process.

For practical purposes, however, it is necessary to know the error of this approximation. In this paper an estimate is presented for the maximum difference of one-dimensional distribution function of the output process and the normal distribution. Other estimates were derived in [5] but they were inexact as they were obtained by neglecting higher terms of infinite series.

GENERAL FORMULA

Let I denote the set of all integers, let $[x]$ denote the integral part of the real number x,

$$[x] = \max \{n \in I : n \leqq x\} .$$

The time variable will be denoted by t. We shall assume that the time scale is normed so that the pulse width is one.

The input process $u_1(t)$, $t \in (-\infty, +\infty)$ is given by

$$u_1(t) = \xi_{[t]} f(t - [t]) , \tag{1}$$

where $\{\xi_n\}_{n \in I}$ is a sequence of independent random variables with values ± 1, identically distributed with

$$\Pr(\xi_n = +1) = \Pr(\xi_n = -1) = \tfrac{1}{2} \tag{2}$$

and $f(t)$, $0 \leqq t < 1$ is the shape of pulses. For simplicity we shall assume that

$$\int_0^1 f(t)\, \mathrm{d}t = 1 . \tag{3}$$

The process $u_2(t)$, $t \in (-\infty, +\infty)$ on the output of a linear time-invariant filter is given by the integral transformation

$$u_2(t) = \int_{-\infty}^{t} u_1(\tau)\, w(t - \tau)\, \mathrm{d}\tau , \tag{4}$$

where the impulse-response function w is supposed to satisfy the condition of physical realizability

$$w(t) = 0 \quad \text{for} \quad t \leqq 0 , \tag{5a}$$

and stability

$$\int_0^{+\infty} |w(t)|\, \mathrm{d}t < +\infty . \tag{5b}$$

We shall investigate the distribution of the output process u_2 at the time $t = 0$. Assuming that the input sequence was applied at $t = -N$, where $N > 0$, $N \in I$, we obtain for the random variable $\eta_N = u_2(0)$ the expression

$$\eta_N = \int_{-N}^{0} \xi_{[\tau]} f(\tau - [\tau])\, w(-\tau)\, \mathrm{d}\tau = \sum_{n=1}^{N} \xi_{-n} \int_0^1 f(\tau)\, w(n - \tau)\, \mathrm{d}\tau . \tag{6}$$

Let us denote $\zeta_n = \xi_{-n} \int_0^1 f(\tau)\, w(n - \tau)\, \mathrm{d}\tau$, $n \in I$. Here ζ_n is the response of a single pulse applied at $t = -n$. From eq. (6) it follows that the random variable η_N is a sum of independent uniformly bounded random variables ζ_n,

$$\eta_N = \sum_{n=1}^{N} \zeta_n . \tag{7}$$

This is a consequence of linearity of the filter.

Let G be the Gaussian distribution function

$$G(x) = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{x} e^{-u^2/2}\, du \,,$$

let F_N be the distribution function of the random variable $\eta_N/\sigma(\eta_N)$ where $\sigma(\eta_N)$ denotes the standart deviation. Applying the Berry - Esseen normal approximation theorem ([3], p. 288) to eq. (7) we obtain

$$\sup_{x \in (-\infty, +\infty)} |F_N(x) - G(x)| \leqq c \frac{\sum_{n=1}^{N} E|\zeta_n|^3}{\sigma^3(\eta_N)} \,, \tag{8}$$

where the symbol E denotes the mathematical expectation. The Berry-Esseen constant $c < 1{,}322$ (cf. [6]). Because of eq. (2),

$$E\,|\zeta_n|^3 = \left| \int_0^1 f(\tau)\, w(n - \tau)\, d\tau \right|^3$$

and since the summands ζ_n are independent random variables

$$\sigma^3(\eta_N) = (E(\sum_{n=1}^{N} \zeta_n)^2)^{3/2} = (\sum_{n=1}^{N} E(\zeta_n)^2)^{3/2} =$$
$$= \left(\sum_{n=1}^{N} \left(\int_0^1 f(\tau)\, w(n - \tau)\, d\tau \right)^2 \right)^{3/2} .$$

Substituing into eq. (8) we obtain the general formula

$$\sup_{x \in (-\infty, +\infty)} |F_N(x) - G(x)| \leqq c \frac{\sum_{n=1}^{N} \left| \int_0^1 f(\tau)\, w(n - \tau)\, d\tau \right|^3}{\left(\sum_{n=1}^{N} \left(\int_0^1 f(\tau)\, w(n - \tau)\, d\tau \right)^2 \right)^{3/2}} . \tag{9}$$

Let $\eta = \lim_{N \to \infty} \eta_N$ and let F denote the distribution function of η. This corresponds to the case where the input signal was applied at minus infinity. The relation (9) becomes

$$\sup_{x \in (-\infty, +\infty)} |F(x) - G(x)| \leqq c \frac{\sum_{n=1}^{\infty} \left| \int_0^1 f(\tau)\, w(n - \tau)\, d\tau \right|^3}{\left(\sum_{n=1}^{\infty} \left(\int_0^1 f(\tau)\, w(n - \tau)\, d\tau \right)^2 \right)^{3/2}} . \tag{10}$$

Because of ineq. (5b), both series on the right-hand side of ineq. (10) converge. Hence, for N sufficiently large, we may use ineq. (10) instead of ineq. (9).

We shall consider two special cases of the impulse-response function w. For simplicity we shall confine ourselves to rectangular pulses, $f(t) = 1$ for $0 \leqq t < 1$.

A. Let $w(t) = Ce^{-\alpha t}$, $t > 0$, where α and C are positive real constants. Then

$$\int_0^1 f(\tau)\, w(n - \tau)\, d\tau = C A_0 e^{-\alpha n}, \quad \text{where}$$

$$A_0 = \int_0^1 e^{\alpha\tau}\, d\tau = \frac{1}{\alpha}(e^\alpha - 1). \tag{11}$$

Denoting the bound in ineq. (10) by $K(\alpha)$,

$$K(\alpha) = \frac{\sum_{n=1}^{\infty} \left| \int_0^1 f(\tau)\, w(n - \tau)\, d\tau \right|^3}{\left(\sum_{n=1}^{\infty} \left(\int_0^1 f(\tau)\, w(n - \tau)\, d\tau \right)^2 \right)^{3/2}} \tag{12}$$

we find that in our case

$$K(\alpha) = \frac{(1 - e^{-2\alpha})^{3/2}}{1 - e^{-3\alpha}}. \tag{13}$$

It is easily seen that

$$\lim_{\alpha \to 0+} K(\alpha) = 0.$$

We are interested in asymptotic behaviour of $K(\alpha)$ for small α. Expanding the exponential functions into power series and neglecting terms of higher order we obtain

$$K(\alpha) \underset{\alpha \to 0+}{\approx} 0{,}94 \sqrt{\alpha}. \tag{14}$$

B. Let $w(t) = Cte^{-\alpha t}$, $t > 0$ with $\alpha > 0$, $C > 0$. In this case

$$\int_0^1 f(\tau)\, w(n - \tau)\, d\tau = Cne^{-\alpha n} \left(A_0 - \frac{1}{n} A_1 \right),$$

where A_0 denotes the integral (11) and

$$A_1 = \int_0^1 \tau e^{\alpha\tau}\, d\tau = \frac{1}{\alpha} e^\alpha - \frac{1}{\alpha^2}(e^\alpha - 1). \tag{15}$$

Since $0 < A_1 < A_0$ we have the following inequalities

$$0 < C(A_0 - A_1)\, n e^{-\alpha n} \leqq \int_0^1 f(\tau)\, w(n - \tau)\, \mathrm{d}\tau \leqq C A_0 n e^{-\alpha n} .$$

Hence

$$\sum_{n=1}^{\infty} \left| \int_0^1 f(\tau)\, w(n - \tau)\, \mathrm{d}\tau \right|^3 \leqq C^3 A_0^3 e^{-3\alpha} \frac{1 + 4e^{-3\alpha} + e^{-6\alpha}}{(1 - e^{-3\alpha})^4}$$

and

$$\sum_{n=1}^{\infty} \left(\int_0^1 f(\tau)\, w(n - \tau)\, \mathrm{d}\tau \right)^2 \geqq C^2 (A_0 - A_1)^2 e^{-2\alpha} \frac{1 + e^{-2\alpha}}{(1 - e^{-3\alpha})^3} .$$

Substituing into eq. (12) we obtain

$$K(\alpha) \leqq \left(\frac{A_0}{A_0 - A_1} \right)^3 \frac{(1 - e^{-2\alpha})^3}{(1 - e^{-3\alpha})^4} \left(\frac{1 - e^{-2\alpha}}{1 + e^{-2\alpha}} \right)^{3/2} (1 + 4e^{-3\alpha} + e^{-6\alpha}) .$$

Substituing further for A_0 and A_1 we have

$$\frac{A_0}{A_0 - A_1} = \frac{1/\alpha\, (e^{\alpha} - 1)}{1/\alpha^2\, (e^{\alpha} - 1 - \alpha)} < \alpha ,$$

since $\alpha > 0$, and hence

$$K(\alpha) < \alpha^3 \frac{(1 - e^{-2\alpha})^3}{(1 - e^{-3\alpha})^4} \left(\frac{1 - e^{-2\alpha}}{1 + e^{-2\alpha}} \right)^{3/2} (1 + 4e^{-3\alpha} + e^{-6\alpha}) . \tag{16}$$

It follows $\lim_{\alpha \to 0+} K(\alpha) = 0$ as in the case **A** but now

$$K(\alpha) \underset{\alpha \to 0+}{\approx} \left(\frac{2}{3} \right)^3 \alpha^{7/2} \left(1 - \frac{15}{2} \alpha \right) . \tag{17}$$

If we compare the expressions (14) and (17) we conclude that, for α small, the Berry-Esseen bound is much smaller for the impulse-response function in case **B** than in case **A**. As corresponding transfer functions differ only by the multiplicity of the pole we conjecture that the multiplicity of poles improves the approximation of normal distribution in general.

Acknowledgement

The author wishes to thank Ing. K. Janáč, C.Sc., from the Institute of Information Theory and Automation, who suggested the problem.

References

[1] *J. Havel*: An Electronic Generator of Random Sequences, Transactions of the Second Prague Conference on Information Theory, Prague 1959.

[2] *K. Janáč*: Determination of Correlation Function on the Output of a Generator of Continuous Random Process (in Czech), Aplikace Matematiky, vol. 6 (1961), No. 1.

[3] *M. Loève*: Probability Theory, 3rd eddition, Van Nostrand, 1963.

[4] *C. Pantelopulos*: Processus Aléatoires Asymptotiquement Stationnaires Laplaciens Produits par Filtrage d'une Suite Périodique d'Impulsions Aléatoires, Transactions of the Second Prague Conference on Information Theory, Prague 1959.

[5] *B. Šubert*: Stochastic Properties of a Random Binary Sequence on the Output of a Linear Filter (in Czech), Research Memorandum No. 145 (1966), Institute of Information Theory and Automation, Czechoslovak Academy of Sciences.

[6] *В. М. Золотарев:* Абсолютная оценка остаточного члена в центральной предельной теореме, Теория вероятностей и ее применения, *11* (1966), 1, 108—119.

Výtah

CHYBA PŘI GENEROVÁNÍ NORMÁLNÍHO ROZLOŽENÍ

BRUNO ŠUBERT

V článku je odvozen odhad pro maximum odchylky jednorozměrné distribuční funkce náhodného procesu vzniklého filtrací posloupnosti nezávislých impulsů náhodné polarity lineárním filtrem od normální distribuční funkce. Na dvou speciálních případech je studován vliv váhové funkce filtru na velikost odchylky.

Резюме

ОШИБКА ПРИ ГЕНЕРИРОВАНИИ ГАУССОВСКОГО РАСПРЕДЕЛЕНИЯ

БРУНО ШУБЕРТ (BRUNO ŠUBERT)

В статье выведена оценка максимального отклонения одномерной функции распределения вероятностного процесса, возникающего фильтрацией последовательности независимых импульсов случайной полярности линейным фильтром и гауссовской функции распределения. В двух частных случаях изучается зависимость отклонения от весовой функции фильтра.

Author's address: Ing. *Bruno Šubert*, C.Sc., Ústav teorie informace a automatizace ČSAV, Praha 2, Vyšehradská 49.