Kumamoto J. Math.
Vol. 12, 73-79 March (1999)

On a formula for correlation coefficients between normally scattered functions

Shin-ichi Ohwaki and Koh-ichi Matsuda

(Received November 30, 1998)

1. Introduction

We defined in [2, 3] real valued functions $N^{\theta}_{\alpha,\beta}(t)$ with one real variable t, and three parameters $0<\alpha, \beta<1$ and $-\infty<\theta<\infty$, whose values scatter around the whole plane like in Figure 3. For two such functions with the same $\alpha\neq\beta$, we can prove the following simple formula for their correlation coefficients in any interval I:

$$\rho_I(N^{\theta}_{\alpha,\beta}, N^{\zeta}_{\alpha,\beta})=\cos \pi(\theta-\zeta).$$

This formula was found by computer experiments, such as Figure 5.

To prove this formula, we use stochastic Riemannian integral formulae for the compositions of continuous functions f and scattered functions, such as $f(N^{\theta}_{\alpha,\beta}(t))$.

Using this formula, we can define functions with the 2-dimensional normal distributions having any given correlation coefficient.

2. Uniformly scattered functions $U_\alpha(t)$

We defined in [1, 2, 3] real valued functions with one real variable, whose values distribute in [0, 1) with the uniform density, as in Figure 1.

Definition 1. Let $0<\alpha<1$. Using the binary method, a real number t can be written as $t=0.1\ldots\times 2^k$ with an integer k. If its mantissa has the finite length l such as $0.1\ldots1$ with l bits between 1, let m be the smallest integer which is greater than $l\times\alpha$. Then the value of the uniformly scattered function $U_\alpha(t)$ is defined as "0." followed by the lowest m bits of the mantissa.

If the mantissa cannot have finite length, then $U_\alpha(t)$ is defined to be 0.

For example, if $t=1101.001$ by binary method, $t=0.1101001\times 2^4$ and then $U_{0.5}(t)=0.1001$. For irrational t, its mantissa cannot be of finite length, and then $U_\alpha(t)=0$. Therefore, due to the theory of Lebesgue integral [4], these functions are almost everywhere equal to 0, and their integrals over any interval are 0. But by computer experiments, Riemannian sums of these functions behave in another way.

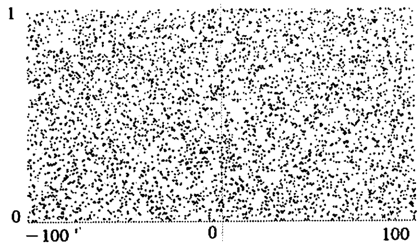

Figure 1 : The graph of $U_\alpha(t)$

As in Figure 1, the values of $U_\alpha(t)$ distributes uniformly in [0, 1), and their density function is equal to 1 over this interval [1, 5].

If $\alpha \neq \beta$, $U_\alpha(t)$ and $U_\beta(t)$ do not influence each other, and are independent.

The product of their differences with their mean 0.5 is

$$f(t)=(U_\alpha(t)-0.5)(U_\beta(t)-0.5).$$

For a fixed number of division n, we calculate their Riemann sum

$$\sum_{k=1}^{n} f(\xi_k)(t_k-t_{k-1}), \quad \xi_k \in [t_k, t_{k-1}]$$

for 10,000 samples, and let their mean be the center of a vertical segment, and let their standard deviation be half of the length of the segment. Figure 2 is the graph of these segments for the number of division from 10 to 10, 000. As the number of division n increases, the distribution of the Riemann sum approach to 0, which is defined in [2] as the stochastic Riemann integral of $f(t)$. Moreover the central limit theorem is valid in general. In this sense, we can write in every interval I,

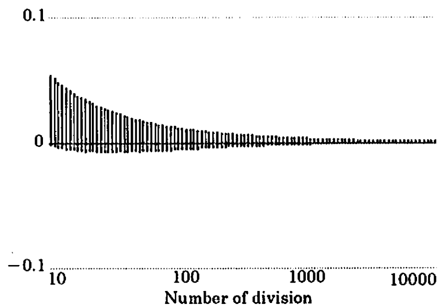

Figure 2 : Distribution of Riemann sum

$$\int_I (U_\alpha(t)-0.5)(U_\beta(t)-0.5)dt=0$$

This means that $U_\alpha(t)-0.5$ and $U_\beta(t)-0.5$ are locally orthogonal.

3. Normally scattered functions $N^{\theta}_{\alpha,\beta}(t)$

From uniformly scattered functions, we can define normally (Gaussian) scattered functions in the following way [2, 3].

Defenition 2. Let $0<\alpha, \beta<1$, and θ be a real number. We define normally scattered functions in the following way:

$$N^{\theta}_{\alpha,\beta}(t)=\sqrt{-2\cdot\log U_{\alpha}(t)}\cdot\sin\pi(2U_{\beta}(t)+\theta)$$

If $\alpha\neq\beta$, their graph is like as Figure 3, and the density function of $N^{\theta}_{\alpha,\beta}(t)$ is the following standard normal (Gaussian) density function

$$\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}.$$

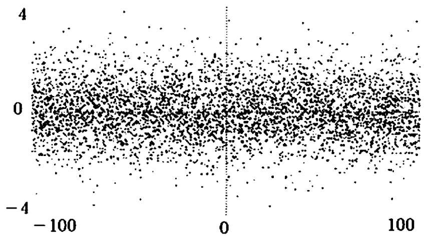

Figure 3: The graph of $N^{0}_{0.4,0.6}(t)$

If $\alpha, \beta, \gamma, \delta$ are different each other, then in any interval I,

$$\int_I N^{\theta}_{\alpha,\beta}(t)N^{\xi}_{\gamma,\delta}(t)dt=0$$

and these two functions $N^{\theta}_{\alpha,\beta}(t)$ and $N^{\xi}_{\gamma,\delta}(t)$ are locally orthogonal. This fact cannot be understood from their graphs like Figure 3.

But by plotting

$$(N^{\theta}_{\alpha,\beta}(t), N^{\xi}_{\gamma,\delta}(t), t)$$

in (x, y, t)-space as in Figure 4, we can visualize this orthogonality relation. Its density becomes the 2-dimensional normal distribution with the correlation coefficient 0.

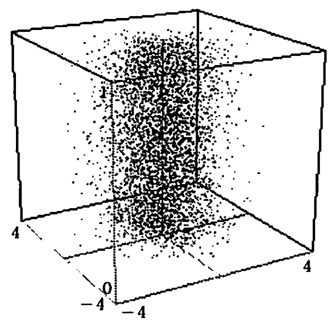

Figure 4: $(N^{0}_{0.3,0.5}(t), N^{0}_{0.4,0.6}(t), t)$

4. Correlation coefficients

To define correlation coefficients between two scattered functions, we consider the usual correlation coefficients for finite samples.

In an interval $I=[a, b]$, take $n+1$ points $a=t_0<t_1<\cdots<t_{n-1}<t_n=b$ with the equal spacings, and let

$$\begin{aligned}f(t_k)&=(f_1(t_k), f_2(t_k))=(x_k, y_k)\\&=(N^{\theta}_{\alpha,\beta}(t_k), N^{\xi}_{\gamma,\delta}(t_k)),\quad k=1,2,\cdots,n\end{aligned}$$

then, the mean of the values of the first coordinate is

$$m_1=\frac{x_1+x_2+\cdots+x_n}{n}=\frac{1}{b-a}\frac{b-a}{n}\sum_{k=1}^{n}f_1(t_k)$$
$$\approx\frac{1}{b-a}\int_a^b N_{\alpha,\beta}^{\theta}(t)dt=0.$$

Similarly, $m_2=\frac{y_1+y_2+\cdots+y_n}{n}\approx 0.$

Consider two vectors, $X=\begin{pmatrix} x_1-m_1 \\ x_2-m_1 \\ \vdots \\ x_n-m_1 \end{pmatrix}\approx\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$, $Y=\begin{pmatrix} y_1-m_2 \\ y_2-m_2 \\ \vdots \\ y_n-m_2 \end{pmatrix}\approx\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$. Their usual correlation coefficient is given by

$$\rho(X, Y)=\frac{X\cdot Y}{\|X\|\cdot\|Y\|}=\frac{\sum_{k=1}^{n}x_k y_k}{\sqrt{\sum_{k=1}^{n}x_k^2}\sqrt{\sum_{k=1}^{n}y_k^2}}.$$

As n become large,

$$\rho(X, Y)\approx\frac{\frac{b-a}{n}\sum_{k=1}^{n}f_1(t_k)f_2(t_k)}{\sqrt{\frac{b-a}{n}\sum_{k=1}^{n}f_1(t_k)^2}\sqrt{\frac{b-a}{n}\sum_{k=1}^{n}f_2(t_k)^2}},$$

and each sums approach to fixed values, in the sense of stochastic Riemann integral, hence the following definition is natural.

Definition 3. The correlation coefficient of two functions $f_1(t)$ and $f_2(t)$ in the interval I is defined to be

$$\rho_I(f_1, f_2)=\frac{\int_I f_1(t)f_2(t)dt}{\sqrt{\int_I f_1(t)^2dt}\sqrt{\int_I f_2(t)^2dt}}.$$

5. Numerical experiments of $\rho_I(N_{\alpha,\beta}^{\theta}, N_{\gamma,\delta}^{\zeta})$

In this paper, we examine the case where $\alpha=\gamma$ and $\beta=\delta$. If we fix α and β and the interval I, the correlation coefficient ρ is the function of (θ, ζ). We divide the square $[0,1]\times[0,1]$ in (θ, ζ)-plane into 50×50 equally spaced meshes, and calculate correlation coefficients for 100,000 samples at each mesh point. Figure 5 is the

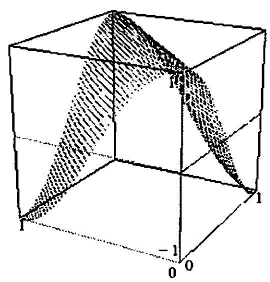

Figure 5 : Graph of (θ, ζ, ρ)

perspective view of 51×51 points (θ, ζ, ρ) in $[0,1]\times[0,1]\times[-1,1]$.

From our careful examination, we conclude that the following formula is valid.

Theorem. If $0<\alpha\neq\beta<1$, then the correlation coefficients ρ between $N^{\theta}_{\alpha,\beta}(t)$ and $N^{\zeta}_{\alpha,\beta}(t)$ is given by.

$$\rho_I(N^{\theta}_{\alpha,\beta}, N^{\zeta}_{\alpha,\beta})=\cos\pi(\theta-\zeta)$$

6. Proof of the theorem

The numerator is calculated as follows.

$$\int_I N^{\theta}_{\alpha,\beta}(t)N^{\zeta}_{\alpha,\beta}(t)dt=$$

$$=\int_I\sqrt{-2\cdot\log U_\alpha(t)}\cdot\sin\pi(2U_\beta(t)+\theta)\sqrt{-2\cdot\log U_\alpha(t)}\cdot\sin\pi(2U_\beta(t)+\zeta)dt$$

$$=\int_I(-2\cdot\log U_\alpha(t))\left(-\frac{1}{2}\right)\{\cos\pi(4U_\beta(t)+\theta+\zeta)-\cos\pi(\theta-\zeta)\}dt$$

$$=\int_I(\log U_\alpha(t))\cos\pi(4U_\beta(t)+\theta+\zeta)dt-\cos\pi(\theta-\zeta)\int_I\log U_\alpha(t)dt$$

Since $U_\beta(t)$ distributes in (0, 1) uniformly, $\pi(4U_\beta(t)+\theta+\zeta)$ distributes in $(\pi(\theta+\zeta), \pi(4+\theta+\zeta))$ uniformly. Therefore $\log U_\alpha(t)\cdot\cos\pi(4U_\beta(t)+\theta+\zeta)$ distributes in $(0,\infty)$ and $(-\infty,0)$ in the same density, and therefore its integral becomes 0. Then the first term is equal to 0.

For continuous functions $f(x)$, the following general formula is valid [5].

$$\int_a^b f(U_\alpha(t))dt=(b-a)\int_0^1 f(y)dy$$

Applying this formula to $f(x)=\log x$, we have

$$\int_a^b\log U_\alpha(t)dt=(b-a)\int_0^1\log(y)dy=a-b.$$

Then the numerator is

$$\int_a^b N^{\theta}_{\alpha,\beta}(t)N^{\zeta}_{\alpha,\beta}(t)dt=(b-a)\cos\pi(\theta-\zeta).$$

For continuous functions $f(x)$, the following general formula is valid [5].

$$\int_a^b f(N^{\theta}_{\alpha,\beta}(t))dt=(b-a)\int_{-\infty}^{\infty}f(y)\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}dy.$$

Applying this formula to $f(x)=x^2$, we have

$$\int_a^b N^{\theta}_{\alpha,\beta}(t)^2 dt=(b-a)\int_{-\infty}^{\infty} y^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy=b-a.$$

Then the denominator is

$$\sqrt{\int_a^b N^{\theta}_{\alpha,\beta}(t)^2 dt}\sqrt{\int_a^b N^{\zeta}_{\alpha,\beta}(t)^2 dt}=b-a.$$

From these calculation of the numerator and the denominator of the correlation coefficients, the formula of the theorem follows.

7. An application of the theorem

Using the above formula, we can create 2 dimensional normally scattered functions with any given correlation coefficient.

Let $\rho=0.8$ and $\theta=0$, then from

$$0.8=\cos \pi(-\zeta),$$

we obtain

$$\zeta=\frac{1}{\pi}\cos^{-1} 0.8 \approx 0.2048.$$

Figure 6 is the graph of $(N^0_{0.4,0.6}(t), N^{0.2048}_{0.4,0.6}(t), t)$ with 5,000 points in $0 \leqq t \leqq 1$.

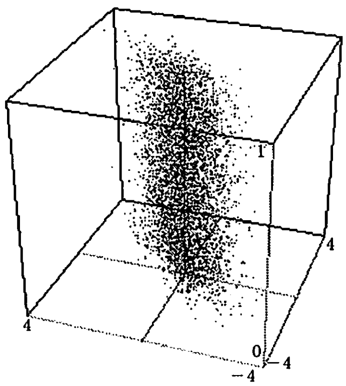

Figure 6 : $(N^0_{0.4,0.6}(t), N^{0.2048}_{0.4,0.6}(t), t)$

References

[1] S. Ohwaki, Everywhere discontinuous functions and differential equations, *Proceedings of JSIAM 1997 annual conference*, (1997), 346-347 (in Japanese).

[2] S. Ohwaki, Methods to solve differential equations with fluctuations using non-Lebesgue integral formulae for everywhere discontinuous functions, *Proceedings of the 56th IPSJ conference* Vol. 1, (1998), 76-77 (in Japanese).

[3] S. Ohwaki and K. Matsuda, On the generation of Brownian motions using high dimensional normally scattered functions, *Proceedings of JSIAM 1998 annual conference*, (1998), 302-303 (in Japanese).

[4] H. Lebesgue, Intégrale, longueur, aire, *Ann. Mat. Pure Appl.*, (3) 7, (1902), 231-359.

[5] http: //www. math. sci. kumamoto-u. ac. jp/~ohwaki/VisualPDE98/SFindex. html (in Japanese)

Shin-ichi Ohwaki
Department of Mathematics
Faculty of Science
Kumamoto University
Kumamoto 860-8555, JAPAN

Koh-ichi Matsuda
Graduate School of Natural Sciences
Kumamoto University
Kumamoto 860-8555, JAPAN