ZERO CROSSINGS OF A GAUSSIAN PROCESS OBSERVED AT DISCRETE RANDOM TIMES AND SOME PECULIAR CONNECTIONS TO THE SIMPLE RANDOM WALK

by

Anirban DasGupta* and Herman Rubin
Purdue University — Purdue University
University of California

Technical Report #97-23

Department of Statistics
Purdue University
West Lafayette, IN USA

September, 1998
Replaces Technical Report #96-6

*Research supported by NSA Grant MDA 904-97-1-0031

ZERO CROSSINGS OF A GAUSSIAN PROCESS OBSERVED AT DISCRETE RANDOM TIMES AND SOME PECULIAR CONNECTIONS TO THE SIMPLE RANDOM WALK

by

Anirban DasGupta* and Herman Rubin
Purdue University
University of California

Herman Rubin
Purdue University

Abstract

Let $X(t)$ be a real valued Gaussian process and F a given absolutely continuous CDF on the time interval $[0, 1]$. Suppose the process is observed at n random times $t_1 < t_2 \ldots < t_n$, which are the order statistics of n samples from the CDF F. We give a formula for the expected number of sign-changes among the values $X(t_i)$ for every fixed n. For the case when $X(t)$ is the standard Brownian motion starting at zero and F is the uniform distribution, the expected number of sign-changes reduces to a neat expression giving a mysterious exact connection to the simple random walk. The expected number of sign-changes. We also consider the random variable T, the epoch of the first sign-changes. A second peculiar phenomenon arises for the Brownian motion case if F is again taken to be uniform. For any given integer $i, P(T > i)$ only depends on i, and not on n, as long as $n > i$. We indicate how to derive a more general formula for $P(T > i)$ for the general Markov case, not just the Brownian motion. The article closes with a formula for the predictor of the end-value X_1 given the earlier observations $X_{t_1}, X_{t_2}, \ldots, X_{t_n}$.

*Research supported by NSA Grant MDA 904-97-1-0031

1 Introduction

Properties of sample paths of the Brownian motion and more general Gaussian processes have been studied in extensive detail by probabilists and mathematicians - see Itô and McKean (1965), Resnick (1992), Ross (1996), among numerous sources. Relatively less seems to have been explored if a Gaussian process is observed at discrete random times, say for instance at the times of realization of an independent Poisson process. In this article, we consider a general mean zero Gaussian process $X(t)$ which is then observed at times $t_1 < t_2 < \ldots < t_n$, the order statistics of a sample of size n from an absolutely continuous CDF F on the interval $[0, 1]$. We consider C_n, the number of zero crossings among $X(t_1), X(t_2), \ldots, X(t_n)$ and T_n, the epoch of the first crossing. Among various results, we establish a rather peculiar phenomenon which is not otherwise obvious. We show that for every $n \geq 2$, the expectation of C_n, when $X(t)$ is the standard Brownian motion and F is $u[0, 1]$, equals $\frac{1}{2}$ times the expected number of returns to the origin of the simple symmetric random walk till time $2n - 2$.

Section 2 gives a general formula for $E(C_n)$ for a general $X(t)$ and a general absolutely continuous F. The standard Brownian motion, the Brownian bridge, and the integrated standard Brownian motion are used as examples to illustrate the general formula. Asymptotics of $E(C_n)$ are considered in Section 3. In Section 4, we consider T_n, the epoch of the first crossing. Under the added assumption of $X(t)$ being a Markov process, we show how to give a formula for $P(T > i)$ and then present another peculiar phenomenon: if $X(t)$ is the standard Brownian motion and F is $u[0, 1]$, then $P(T_n > i)$ is a fixed number depending only on i, but not on n, as long as $n > i$. We then give the exact values of $P(T_n > i)$ for certain values of i. In particular, $P(T_n > 2) = \frac{3}{4}$ for all $n > 2$ and $P(T_n > 3) = \frac{5}{8}$ for all $n > 3$. For larger $i, P(T_n > i)$ can be accurately approximated by Monte Carlo approximation of our exact formula. Section 5 closes the article by giving a formula for the Best Linear Unbiased Predictor of the end-value X_1, given the earlier observations $X_{t_1}, \ldots, X_{t_n}$. Here $X(t)$ is the Standard Brownian motion but F is arbitrary. The effect of F on the predictor is quite interesting.

2 Expected Number of Sign-Changes

In this section, we present a general formula for the number of zero crossings of a general mean zero Gaussian process observed at discrete random times. Thus, let F be a CDF on $[0,1]$ and let for $n \geq 2, t_1 < t_2 <, \ldots, < t_n$ be the order statistics of a random sample of size n from F. $X(t)$ is a mean zero Gaussian process on $[0,1]$ with covariance kernel $C(s,t)$ and let C_n denote the number of sign changes among $X(t_1), X(t_2), \ldots, X(t_n)$.

Theorem 1

$$E(C_n) = \frac{n!}{\pi} \sum_{i=1}^{n-1} \frac{1}{(i-1)!(n-i-1)!}$$
$$\int_0^1 \int_0^t \cos^{-1} \frac{C(s,t)}{\sqrt{C(s,s)C(t,t)}} \overset{i-1}{F(s)} \,(1-F(t)) \overset{n-i-1}{f(s)} \, f(t) ds\, dt \qquad (1)$$

Proof: Note that $C_n = \sum_{i=1}^{n-1} (X(t_i)X(t_{i+1}) < 0)$

$$\begin{aligned} \text{Therefore E(C}_\text{n}) &= \sum_{i=1}^{n-1} P(X(t_i)X(t_{i+1}) < 0) \\ &= n - 1 - \sum_{i=1}^{n-1} P(X(t_i)X(t_{i+1}) \geq 0) \\ &= n - 1 - 2\sum_{i=1}^{n-1} P(X(t_i) \geq 0, X(t_{i+1}) \geq 0) \qquad (2) \end{aligned}$$

Now note that if (X, Y) has a bivariate normal distribution with means zero and correlation ρ, then $P(X \geq 0, Y \geq 0) = \frac{1}{2} - \frac{1}{2\pi} \cos^{-1} \rho$ (see Tong (1990). Since t_i, t_{i+1} are the order statistics of a sample of size n from F and therefore have the joint density $\frac{n!}{(i-1)!(n-i-1)!} \overset{i-1}{F(t_i)} \,(1 - F(t_{i+1}))^{n-i-1} f(t_i) f(t_{i+1}), 0 \leq t_i \leq t_{i+1} \leq 1$, (1) now follows from (2).

Example 1. Suppose $X(t)$ is the Standard Brownian motion (SBM) on $[0,1]$ and F is the $u[0,1]$ CDF. Then, by Theorem 1,

$$
\begin{aligned}
E(C_n) &= \frac{n!}{\pi} \sum_{i=1}^{n-1} \frac{1}{(i-1)!(n-i-1)!} \\
&\quad \int_0^1 \int_0^t \cos^{-1} \sqrt{\frac{s}{t}} s^{i-1}(1-t)^{n-i-1} ds\ dt \\
&= \frac{n!}{\pi} \sum_{i=1}^{n-1} \frac{2}{(i-1)!(n-i-1)!} \\
&\quad \int_0^1 \int_0^1 \cos^{-1} x \cdot x^{2i-1} t^i (1-t)^{n-i-1} dx\ dt \qquad (3)
\end{aligned}
$$

Since $\int_0^1 \cos^{-1} x \cdot x^{2i-1} dx = \frac{\pi(2i-1)!}{(i!)^2 2^{2i+1}}$, (Gradshteyn and Ryzhik (1980), pp 607) it follows on simplification from (3) that

$$
E(C_n) = \frac{1}{2} \sum_{i=1}^{n-1} \frac{\binom{2i}{i}}{2^{2i}} \qquad (4)
$$

One therefore has the following mysterious corollary:

Corollary 1 *For every $n \geq 2, E(C_n) = \frac{1}{2} \cdot E$ (Number of returns to the origin of the simple symmetric random walk in $2n-2$ steps).*

The fact that Corollary 1 holds as an identify for every n is an interesting and quite remarkable fact and we find it intriguing.

Example 2. Suppose $X(t)$ is the Standard Brownian Bridge (SBB) on $[0,1]$ and F is the $u[0,1]$ CDF. Then,

$$
\begin{aligned}
E(C_n) &= \frac{n!}{\pi} \sum_{i=1}^{n-1} \frac{1}{(i-1)!(n-i-1)!} \int_0^1 \int_0^t \\
&\quad \cos^{-1} \sqrt{\frac{s(1-t)}{(1-s)t}} s^{i-1}(1-t)^{n-i-1} ds\ dt \qquad (5)
\end{aligned}
$$

Making the substitution $x = \sqrt{\frac{s(1-t)}{(1-s)t}}$ and on using the representation $\int_0^1 t^i(1-t)^{n-i}(1-\alpha t)^{-i-1} dt = B(i+1, n-i+1) F(i+1, i+1; n+2, \alpha)$, where$F(\cdot,\cdot,\cdot,\cdot)$

denotes the ${}_2F_1$ hypergeometric function, it follows from (5) on using Fubini's theorem that

$$E(C_n) = \frac{2}{\pi(n+1)} \sum_{i=1}^{n-1} i(n-i)$$
$$\int_0^1 x^{2i-1}(\cos^{-1} x) F(i+1, i+1; n+2; 1-x^2)dx \tag{6}$$

It does not seem possible to simplify (6) further to a closed form.

Example 3. Suppose $X(t)$ is the once integrated Brownian motion (ISBM), $X(t) = \int_0^t \mathcal{E}(u)du$, where $\mathcal{E}(u)$ is the SBM on [0, 1]. Then, on application of Theorem 1, one gets

$$E(C_n) = 1 + \frac{2}{\pi} \sum_{i=1}^{n-1} ia_i, \text{ where } a_i = \int_0^1 x^{2i-1} \cos^{-1}\left(\frac{3}{2}x - \frac{1}{2}x^3\right) dx. \tag{7}$$

Again, although, as in (6), simplification to a closed form does not seem possible, expressions (6) and (7) are useful in investigating the asymptotics of $E(C_n)$ as $n \to \infty$. The following table gives values of $E(C_n)$ for some selected values of n and the three processes discussed in the above examples.

Table 1

n		$E(C_n)$	
	SBM	SBB	ISBM
2	.25	.33333	1.17301
3	.4375	.6	1.27853
4	.59375	.82857	1.35461
5	.73047	1.03174	1.41414
10	1.26197	1.83772	1.60111
20	2.00741	2.98814	1.79003

3 Asymptotics of $E(C_n)$

There is some intrinsic interest in knowing the rate of growth of $E(C_n)$ as the number of points $n \to \infty$. Since the SBB returns to the origin at $t = 1$, intuition might suggest that $E(C_n)$ grows faster for the SBB than the SBM. Also, since the sample paths of the ISBM are more smooth than those of the SBM, one would expect that in that case $E(C_n)$ might grow slower than for the SBM. This is apparent in the numbers presented in Table 1 as well. We have the following result.

Theorem 2 *$E(C_n) \sim \frac{\sqrt{n}}{\sqrt{\pi}}$ for the SBM and the SBB*

$$\sim 1 + \frac{\sqrt{3}}{2\pi} \log n \quad \text{for the ISBM}$$

Proof: For the SBM, we have the closed form expression (4): $E(C_n) = \frac{1}{2} \cdot \sum_{i=1}^{n-1} \frac{\binom{2i}{i}}{2^{2i}}$ from which the result follows immediately. The rate for the SBB can be derived either from the SBM, or by following the steps of the ISBM case given below. For the ISBM, from (7),

$$E(C_n) = 1 + \frac{2}{\pi} \sum_{i=1}^{n-1} i a_i,$$

where

$$a_i = \int_0^1 x^{2i-1} \cos^{-1}\left(\frac{3}{2}x - \frac{x^3}{2}\right) dx.$$

Now note that $\cos z \sim \sqrt{1-z^2}$ as $z \to 0$, i.e., $\cos^{-1} y \sim \sqrt{1-y^2}$ as $y \to 1$. Thus, as $x \to 1$,

$$\begin{aligned} \cos^{-1}\left(\frac{3}{2}x - \frac{x^3}{2}\right) &\sim \sqrt{1 - \left(\frac{3}{2}x - \frac{x^3}{2}\right)^2} \\ &= \sqrt{(1-x^2)^2\left(1 - \frac{x^2}{4}\right)} \\ &\sim \frac{\sqrt{3}}{2}(1-x^2) \end{aligned} \tag{8}$$

Since $\int_0^1 x^{2i-1}(1-x^2)dx = \frac{1}{2i(i+1)}$, it follows that $\sum_{i=1}^{n} i a_i \sim \frac{\sqrt{3}}{4} \log n$ and the stated result of the Theorem follows.

The following table illustrates the usefulness ofthe asymptotic rates presented in Theorem 2.

Table 2

n	$E(C_n)$			Asymptotic expression		
	SBM	SBB	ISBM	SBM	SBB	ISBM
20	2.007	2.988	1.79003	2.52313	2.52313	1.826
30	2.57735	3.75332	1.90106	3.09019	3.09019	1.93759
50	3.47946	*	2.041	3.98942	3.98942	2.078
100	5.13485	*	2.232	5.6419	5.6419	2.269

For $n \geq 50$, exact evaluation of $E(C_n)$ for the SBB became numerically difficult, thereby making the asymptotic expression even more valuable.

4 Epoch of First Crossing

In this section, we present a second quite remarkable phenomenon. Let T_n denote the time at which the first zero crossing happens, i.e.,

$$T_n > i \;\text{ if }\; X(t_1), \ldots, X(t_i) > 0(<0). \tag{9}$$

There is some interest in knowing the distribution of T_n. We assume the same structure as before, i.e., $t_1 < t_2 < \ldots < t_n$ are the order statistics of a random sample from an absolutely continuous CDF F on $[0,1]$ and $\{X(t)\}$ is a zero mean Gaussian process on $[0,1]$. With the added condition that $\{X(t)\}$ is Markov, one can give a general formula for $P(T > i)$ for $i = 2, \ldots, n$. We present the case when F is the $u[0,1]$ CDF and $\{X(t)\}$ is the SBM and present a nice phenomenon. In the following theorem, a product $\prod_{i=m}^{n} a_j$ is defined to be 1 if $m > n$.

Theorem 3 *Let $\{X(t)\}$ be the SBM on $[0,1]$ and let $t_1 < t_2 < \ldots < t_n$ be the order statistics of a random sample from the uniform distribution on $[0,1]$. Let $T = T_n$ denote the epoch of the first sign-change among $X(t_1), X(t_2), \ldots, X(t_n)$. Then for any $i < n$,*

$$P(T > i) = \frac{1}{\prod_{j=1}^{i-1} B(j, i-j)} \int_{[0,1]^{i-1}} g_i\left(\underset{\sim}{u}\right) \left\{ \prod_{j=1}^{i-1} u_j^{j-1}(1-u_j)^{i-j-1} \right\} d\underset{\sim}{u} \tag{10}$$

where

$$g_i\left(\underset{\sim}{u}\right) = \frac{\prod_{j=1}^{i-2}\left\{1-\frac{1}{2\pi}\left(\cos^{-1}\sqrt{u_j}+\cos^{-1}\sqrt{u_{j+1}}+\cos^{-1}\sqrt{u_j u_{j+1}}\right)\right\}}{\prod_{j=2}^{i-2}\left\{1-\frac{1}{\pi}\cos^{-1}\sqrt{u_j}\right\}} \tag{11}$$

In particular, for any given i, $P(T > i)$ depends only on i and not on n.

Proof: We use the notation $Y_i = I(X(t_i)X(t_{i+1}) \leq 0)$. Then,

$$\begin{aligned} P(T > i) &= P(Y_1 = 0, \ldots, Y_{i-1} = 0) \\ &= EP(Y_1 = 0, \ldots, Y_{i-1} = 0 | t_1, \ldots, t_i), \end{aligned} \tag{12}$$

where $E(\cdot)$ means expectation with respect to the joint distribution of $(t_1, \ldots, t_i)$. Now, given the times $t_1, \ldots, t_i$, due to the Markov property of SBM, one has the identity

$$\begin{aligned} &P(Y_1 = 0, \ldots, Y_{i-1} = 0|\underset{\sim}{t}) \\ &= \frac{P(Y_1 = 0, Y_2 = 0|\underset{\sim}{t}) \cdot P(Y_2 = 0, Y_3 = 0|\underset{\sim}{t} \cdots P(Y_{i-2} = 0, Y_{i-1} = 0|\underset{\sim}{t})}{P(Y_2 = 0|\underset{\sim}{t}) \cdots P(Y_{i-2} = 0|\underset{\sim}{t})}. \end{aligned} \tag{13}$$

(13) is obtained by induction on i and by using the fact that for a Markov process, given the present, the future and the past are independent.

Now, in (13), use the following probability expressions:

$$\begin{aligned} P\left(Y_j = 0|\underset{\sim}{t}\right) &= 2P\left(X(t_j) > 0, X(t_{j+1}) > 0|\underset{\sim}{t}\right) \\ &= 1 - \frac{1}{\pi}\cos^{-1}\sqrt{\frac{t_j}{t_{j+1}}} \end{aligned} \tag{14}$$

and

$$\begin{aligned} &P\left(Y_j = 0, Y_{j+1} = 0|\underset{\sim}{t}\right) \\ &= 2P\left(X\left(t_j\right) > 0, X\left(t_{j+1}\right) > 0, X\left(t_{j+2}\right) > 0|\underset{\sim}{t}\right) \end{aligned}$$

$$
\begin{aligned}
&= 2\left\{\frac{1}{8}+\frac{1}{4\pi}\left(\sin^{-1}\sqrt{\frac{t_j}{t_{j+1}}}+\sin^{-1}\sqrt{\frac{t_{j+1}}{t_{j+2}}}+\sin^{-1}\sqrt{\frac{t_j}{t_{j+2}}}\right)\right\} \\
&\quad \text{(see Tong(1990))} \\
&= 1-\frac{1}{2\pi}\left(\cos^{-1}\sqrt{\frac{t_j}{t_{j+1}}}+\cos^{-1}\sqrt{\frac{t_{j+1}}{t_{j+2}}}+\cos^{-1}\sqrt{\frac{t_j}{t_{j+2}}}\right) \qquad (15)
\end{aligned}
$$

If we now write $\frac{t_j}{t_{j+1}} = u_j$, then substitution of (14) and (15) into (13), leads to

$$P(T > i) = Eg_i(\underset{\sim}{u}) \qquad (16)$$

where $g_i(\cdot)$ is as it is defined in (11). We now use the fact that the ratios $u_j, 1 \leq j \leq i-1$, of the successive order statistics of the uniform distribution have the property that they are independent with u_j having the marginal $\beta(j, i-j)$ density (see, e.g., Reiss(1989)). Formula (10) follows immediately.

Remark. Formula (10) can be evaluated exactly to give $P(T > 2) = \frac{3}{4}$ and $P(T > 3) = \frac{5}{8}$. For larger $i, P(T > i)$ can be approximated from formula (10) by Monte Carlo simulation; i.e., for a specified simulation size N, one may simulate N uniform vectors from the $(i-1)$-dimensional unit cube and form an average of the entire integrand in (10) and divide by the constant $\prod_{j=1}^{i-1} B(j, i-j)$. We report some values ($N =$ the simulation size $= 7500$).

Table 3

i	2	3	4	5	6
$P(T > i)$	.750	.625	.546	.350	.003

5 Predicting the Final Location

Finally, we close the article by deriving a predictor of the end-value X_1 given the earlier observations $X_{t_1}, X_{t_2}, \ldots, X_{t_n}$. In this section, $\{X(t)\}$ is assumed to be the standard Brownian motion and $t_1 < \ldots < t_n$, as before, the order statistics of a sample from a CDF F on $[0, 1]$. We will derive the Best Linear Unbiased Predictor of X_1.

Definition. The Best Linear Unbiased Predictor of X_1 is one that minimizes $\mathrm{Var}\left(X_1 - \sum_{i=1}^{n} a_i X_{t_i}\right)$ among all linear unbiased predictors $\sum_{i=1}^{n} a_i X_{t_i}$.

Remark. Without further constraints, the coefficients $\{a_i\}$ can be arbitrary if $\{X(t)\}$ is the SBM. However, we will use the restriction $\sum_{i=1}^{n} a_i E_F(t_i) = 1$. This is the condition for unbiasedness of $\sum_{i=1}^{n} a_i X_{t_i}$ if there is a linear drift in the Brownian motion. Thus, the predictors are unbiased even if there is a linear drift, but the best predictor given below is under the assumption of no drift.

Theorem 4 *Let $\{X(t)\}$ be the SBM and F an absolutely continuous CDF on $[0,1]$. Let $t_1 < t_2 < \ldots < t_n$ be the order statistics of a sample of size n from F. The Best Linear Unbiased Predictor of X_1 given $X_{t_1}, X_{t_2}, \ldots, X_{t_n}$ is $\frac{1}{E_F(t_n)} X_{t_n}$.*

Proof: Let $\Sigma = \Sigma_n$ denote the unconditional covariance matrix of $(X_{t_1}, \ldots, X_{t_n})$ and let $\underset{\sim}{\alpha} = \underset{\sim}{\alpha}_n$ denote the vector with coordinates $\alpha_i = E_F(t_i)$.

Thus, $\Sigma = \left((E_F \min(t_i, t_j))\right)$. Also note that $\mathrm{Cov}(X_1, X_{t_i}) = E(X_1 X_{t_i}) = E_F(t_i) = \alpha_i$. Hence, by a direct calculation,

$$\mathrm{Var}\left(X_1 - \sum_{i=1}^{n} a_i X_{t_i}\right) = \underset{\sim}{a}'\Sigma\underset{\sim}{a} - 2\underset{\sim}{a}'\underset{\sim}{\alpha} + 1 \tag{17}$$

and so the Best Linear Unbiased Predictor corresponds to the vector $\underset{\sim}{a}$ that minimizes $\underset{\sim}{a}'\Sigma\underset{\sim}{a}$ subject to $\underset{\sim}{a}'\underset{\sim}{\alpha} = 1$. This is given by

$$\underset{\sim}{a} = \frac{\Sigma^{-1}\underset{\sim}{\alpha}}{\underset{\sim}{\alpha}'\Sigma^{-1}\underset{\sim}{\alpha}} \tag{18}$$

Now Σ is of the form

$$\Sigma = \begin{pmatrix} \alpha_1 & \alpha_1 & \ldots & \alpha_1 \\ \alpha_1 & \alpha_2 & \ldots & \alpha_2 \\ & \cdot & \cdot & \cdot \\ \alpha_1 & \alpha_2 & \ldots & \alpha_n \end{pmatrix} \tag{19}$$

and

$$\Sigma^{-1}\underset{\sim}{\alpha} = (0\ 0\ \ldots\ 0\ 1)' \tag{20}$$

Hence, from (18),

$$\underset{\sim}{q} = \left(0\ 0\ \ldots\ 0\ \frac{1}{\alpha_n}\right) = \left(0\ 0\ \ldots\ 0\ \frac{1}{E_F(t_n)}\right),$$

as needed.

The following table illustrates the effect of the CDF F on the predictor. We will use $f(x)$ to denote the density of F.

Table 4

	$\frac{1}{E_F(t_n)}$			
	$f(x) = 2x$	$f(x) = 1$	$f(x) = 6x(1-x)$	$f(x) = 2(1-x)$
n				
2	1.25	1.5	1.5909	2.1429
3	1.1667	1.3333	1.4433	1.8421
4	1.125	1.25	1.3645	16845
5	1.1	1.2	1.3145	15858
10	1.05	1.1	1.2031	1.3704

Acknowledgement. I have had many stimulating conversations with Burgess Davis, Tom Sellke and Herman Rubin. Tom Sellke, in particular, contributed a lot to my understanding of the invariance result in Theorem 3.

References

[1] Gradshteyn, I.S. and Ryzhik, I.M., *Table of Integrals*, Series, and Products, Academic Press, New York, (1980).

[2] Itô, K. and McKean, J.P., *Diffusion processes and their sample paths*, Springer Verlag, New York, (1965).

[3] Reiss, R.D., Approximate distributions of Order Statistics, Springer-Verlag, New York, (1989).

[4] Resnick, S. I., *Adventures in Stochastic Processes*, Birkhauser, Basel, (1992).

[5] Ross, S. M., *Stochastic Processes*, Wiley, New York, (1996).

[6] Tong, Y-L., *The Multivariate Normal Distribution*, Springer Verlag, New York, (1990).