Some Statistical Properties of the Heaviside Step Function Based on Sign Data

Louis M. Houston

Abstract—Sign data are the signs of signal plus random noise. Signal amplitudes can be recovered from sign data and this phenomenon is incorporated into seismic data exploration. We show that when sign data are transformed into the Heaviside step function, the expectation value becomes the cumulative distribution function for zero-mean noise. In this paper, we confirm this result and compare related sign data and Heaviside step function quantities for a sample signal with both uniform and Gaussian noise.

Index Terms—Cumulative distribution function, Gaussian noise, Heaviside step function, sign data.

I. INTRODUCTION

Seismic data acquisition generally consists of noisy data that are collected in a redundant manner. Coherent signals are subsequently recovered by summing over the data to reduce the noise amplitude and increase the signal amplitude. Interestingly, for relatively large noise amplitudes, only the signs of the data are needed to recover coherent signal. In contrast, the signal is clipped if the noise amplitude is too low. We point out that signal recovery does not require true amplitude. Sign data recovered signal is generally modulated by a scaling factor related to the noise. Early seismic recording systems had limited memory. Therefore, sign data was often acquired because it presented a way to compress data for storage in the field [1].

Historically, the amplitude recovery of sign data has been exploited to reduce the required dynamic range of seismic recordings. Sign data digital recording means that only the sign of the true amplitude sample is recorded with one bit. In conventional seismic recording, 16 to 20 binary bits per sample point [2] are recorded. The economic advantages of sign data acquisition are immediately obvious. Although it was more widely used during earlier years of the seismic industry, sign data recording remains as a viable tool today [3]. However, sign data recording is a tool that is suitable only for a large number of channels.

Houston [4] has reported on various mathematical properties of sign data. In this paper we discuss some properties of sign data after it is transformed into a binary format. That is, instead of consisting primarily of ones and minus ones, the transformed data consists primarily of ones and zeroes. This binary format is essentially the Heaviside step function. We also compare related sign data and Heaviside step function [5] quantities for a sample signal with both uniform and Gaussian noise.

II. BINARY DATA AMPLITUDE RECOVERY

Sign data amplitude recovery occurs when the average of the signs of signal plus noise reproduces the signal. Signal recovery can occur if rather than averaging sign data, the associated binary data are averaged.

The sign of data, \(d\), is the function,

\[
\text{sgn}(d) = \begin{cases} 
+1, & d > 0 \\
0, & d = 0 \\
-1, & d < 0 
\end{cases}
\]

(1)

The expectation value for sign data amplitude recovery divides into positive and negative components:

\[
E(\text{sgn}(f + X)) = \int_{-\infty}^{-f} \rho(x) \, dx - \int_{-\infty}^{f} \rho(x) \, dx.
\]

(2)
where \( f \) is the signal, \( X \) is the random noise, and \( \rho(x) \) is the noise density.

The transformation
\[
\text{sgn}(f + X) \rightarrow \frac{1 + \text{sgn}(f + X)}{2}
\]
(3)
transforms the data primarily into non-negative binary values:
\[
\{1,-1\} \rightarrow \{1,0\}.
\]
(4)
Consequently, if we retain only the non-negative component, we have
\[
E\left(\frac{1 + \text{sgn}(f + X)}{2}\right) = \int_{-f}^{f} \rho(x) \, dx.
\]
(5)
For a uniform probability distribution,
\[
\rho(x) = \begin{cases} 
1/2a, & \text{for } -a \leq x \leq a \\
0, & \text{else}
\end{cases}
\]
(6)
equation (5) becomes
\[
E\left(\frac{1 + \text{sgn}(f + X)}{2}\right) = \int_{-f}^{f} \frac{1}{2a} \, dx = \frac{a + f}{2a}
\]
(7)
for \( |f| \leq a \).
It follows from (7) that
\[
E(\text{sgn}(f + X)) = \frac{f}{a},
\]
(8)
which is the traditional sign data amplitude recovery.
From (7) we see that amplitude recovery for the binary data is given as
\[
f = 2aE\left(\frac{1 + \text{sgn}(f + X)}{2}\right) - a.
\]
(9)
We point out that if \( \rho(x) \) is symmetric about zero or has zero mean,
\[
\int_{-f}^{f} \rho(x) \, dx = \int_{-\infty}^{\infty} \rho(x) \, dx.
\]
(10)
Thus, for Gaussian noise, \( E((1 + \text{sgn}(f + X))/2) \) is the cumulative distribution function [6], (CDF), \( F(f;0,\sigma^2) \). For the zero-mean (i.e., \( \mu = 0 \) ) Gaussian distribution,
\[
\rho(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2}
\]
(11)
where \( \sigma \) is the standard deviation. Therefore, we can write
\[
E\left(\frac{1 + \text{sgn}(f + X)}{2}\right) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{f} e^{-t^2/2\sigma^2} \, dt
\]
\[
= \frac{1}{2} \left(1 + \text{erf}\left(\frac{f}{\sigma\sqrt{2}}\right)\right).
\]
(12)
where \( \text{erf} \) is the error function.

It follows from (12) that

\[
E(\text{sgn}(f + X)) = \text{erf}\left(\frac{f}{\sigma \sqrt{2}}\right).
\]

From (12) we see that amplitude recovery is given as

\[
f = \sigma \sqrt{2} \text{erf}^{-1}\left(2E\left(\frac{1 + \text{sgn}(f + X)}{2}\right) - 1\right).
\]

Since the cumulative distribution function is the probability given as

\[
F(x; \mu, \sigma^2) = P(X \leq x),
\]

We can write

\[
E\left(\frac{1 + \text{sgn}(f + X)}{2}\right) = P(X \leq f).
\]

(16)

Since the probability density is the derivative of the cumulative distribution function, we have

\[
\frac{d}{df} E\left(\frac{1 + \text{sgn}(f + X)}{2}\right) = \frac{1}{\sqrt{2\pi\sigma}} e^{-f^2 / 2\sigma^2}.
\]

(17)

Since

\[
\frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} e^{-f^2 / 2\sigma^2} df = 1,
\]

(18)

it follows that

\[
\int_{-\infty}^{\infty} \frac{d}{df} E\left(\frac{1 + \text{sgn}(f + X)}{2}\right) df = 1.
\]

(19)

We can confirm equation (19). Use the fact that

\[
\frac{d}{dx} \text{sgn}(x) = 2\delta(x),
\]

(20)

[7], where \( \delta(x) \) is the delta function and (19) becomes

\[
\int_{-\infty}^{\infty} E(\delta(f + X)) df = 1
\]

(21)

or

\[
\int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} e^{-x^2 / 2\sigma^2} \delta(f + x) dx\right] df = 1.
\]

(22)

Based on the property of the delta function, (22) becomes (18).

We can simplify our results by introducing the Heaviside step function, [8]:

\[
H(x) = \begin{cases} 
0, & x < 0 \\
\frac{1}{2}, & x = 0 \\
1, & x > 0 
\end{cases}
\]

(23)

Therefore, we can write

\[
\text{sgn}(f + X) = 2H(f + X) - 1.
\]

(24)

Based on (24), we have
\[
E\left(\frac{1 + \text{sgn}(f + X)}{2}\right) = E(H(f + X)).
\]
(25)

\[
E(H(f + X)) = \frac{1}{2}(1 + \text{erf}\left(\frac{f}{\sigma\sqrt{2}}\right)).
\]
(26)

Obviously, we must have

\[
\int_{-\infty}^{\infty} \frac{d}{df} [E(H(f + X))] = 1.
\]
(27)

In order to confirm (27) use

\[
\frac{d}{df} H(f + X) = \delta(f + X).
\]
(28)

Thus, (27) becomes

\[
\int_{-\infty}^{\infty} E(\delta(f + X)) = 1,
\]
which is confirmed by equation (22). Amplitude recovery is thus

\[
f = \sigma\sqrt{2}\text{erf}^{-1}(2E(H(f + X)) - 1).
\]
(29)

The variance is given as

\[
\text{Var}(H(f + X)) = E(H^2(f + X)) - (E(H(f + X)))^2.
\]
(30)

We can use the fact that

\[
H^2(f + X) = H(f + X).
\]

This follows since \(\text{sgn}^2(f + X) = 1\):

\[
H^2(f + X) = \left[\frac{1}{2} + \frac{1}{2}\text{sgn}(f + X)\right]^2
\]
(31)

\[
= \frac{1}{4} + \frac{1}{2}\text{sgn}(f + X) + \frac{1}{4}
\]
\[
= H(f + X).
\]
(32)

Substituting (32) into (30) yields

\[
\text{Var}(H(f + X)) = E(H(f + X))(1 - E(H(f + X))).
\]
(33)

Consequently, for uniform noise, \(\text{Var}(H(f + X))\) is minimal when

\[
\frac{a + f}{2a} = 1
\]
or
\[
f = a.
\]
(34)

For Gaussian noise, \(\text{Var}(H(f + X))\) is minimal when

\[
\frac{1}{2}(1 + \text{erf}\left(\frac{f}{\sigma\sqrt{2}}\right)) = 1
\]
(36)

or
\[ \text{erf}(\frac{f}{\sqrt{2}\sigma}) = 1. \]  
(37)

Equation (37) implies that
\[ f = \sqrt{2}\sigma \text{erf}^{-1}(1) \]

or
\[ f \to \infty. \]  
(38)

It is interesting to contrast this to
\[ \text{Var}(\text{sgn}(f + X)) = E(\text{sgn}^2(f + X)) - (E(\text{sgn}(f + X))^2. \]  
(39)

Equation (39) reduces to
\[ \text{Var}(\text{sgn}(f + X)) = 1 - (E(\text{sgn}(f + X))^2. \]  
(40)

For uniform noise, \( \text{Var}(\text{sgn}(f + X)) \) is minimal when
\[ \frac{f^2}{a^2} = 1 \]

or
\[ f = \pm a. \]  
(41)

For Gaussian noise, \( \text{Var}(\text{sgn}(f + X)) \) is minimal when
\[ \text{erf}^2(\frac{f}{\sqrt{2}\sigma}) = 1 \]

or
\[ f \to \pm \infty. \]  
(42)

III. COMPUTATIONAL TESTS

We compared various quantities for sign data by using the signal function [9]:
\[ f(t) = -\frac{\sin(t/3)}{t/3} e^{-(t/10)^2}. \]  
(43)

We computed expectation values for sign data by replacing the integral with a summation over a range from one to \( N \). In Figure 1, we show \( f, f + X, \text{sgn}(f + X), E(\text{sgn}(f + X)), \) and \( E(H(f + X)) \) for uniform noise. In Figure 2, we show the same quantities for Gaussian noise. The noise has unit amplitude and \( N = 1000 \).

Fig. 1. A plot of \( f, f + X, \text{sgn}(f + X), E(\text{sgn}(f + X)), \) and \( E(H(f + X)) \).
IV. CONCLUSION

We have shown that the expectation value of the Heaviside step function based on sign data is the cumulative distribution function for zero-mean random noise and we have examined some of its statistical properties. In addition, we have compared quantities related to sign data and the Heaviside step function for a sample signal with both uniform and Gaussian noise.

ACKNOWLEDGMENT

Discussions with Maxwell Lueckenhoff are appreciated.

REFERENCES