BIAS AND SPREAD IN EVT PERFORMANCE TESTS

JOEL G. SMITH
Jet Propulsion Laboratory
California institute of Technology
Pasadena, Calif.

Summary  Performance tests (measurements of probability of error) of communication systems characterized by low bit rates and high reliability requirements frequently utilize classical extreme value theory (EVT) to avoid the excessive test times required by bit error rate (BER) tests. If the underlying noise is gaussian or perturbed-gaussian, the EVT error estimates have either excessive bias or excessive variance if an insufficient number of test samples is used. EVT is examined to explain the cause of this bias and spread; experimental verification is made by testing a known gaussian source, and procedures that minimize these effects are described. Even under these conditions, EVT test results are not particularly better than those of BER.

Introduction  Classically, performance testing of a digital communication system consists of counting the numbers of wrong outputs of the system in response to a sequence of known inputs. This so-called bit error rate (BER) test procedure is known to produce an estimator of the probability of error which converges to the right answer eventually, regardless of the type of underlying noise; i.e., it is asymptotically unbiased, efficient, and distribution-free. But the lower the probability of error of the system, the longer the sequence length must be to generate a sufficient number of errors to construct meaningful confidence intervals; and the lower the data rate of the system, the longer is the time required to generate this sufficient sequence.

As a time-saving alternate to BER testing of communication systems characterized by low bit rate and low probability of error, classical extreme value theory (EVT) [1-9] has been utilized. Procedures based on classical EVT do not record the number of errors generated by the system, but rather examine the noisy data input to the decision element of the system and attempt to measure probability of error from the statistics of the noisiest data - the extremes of the data. Classical EVT is also well-known [1] to be asymptotically unbiased, efficient, and distribution-free within a broad class of noise. Unfortunately, the rate of convergence of the EVT estimate to its asymptote is sensitive

* This work was supported by the National Aeronautics and Space Administration under Contract No. NAS 7-100.
to the statistics of the underlying noise, and particularly slow when that noise is gaussian [6-9]. Further, for a large class of noise, including gaussian noise, the estimator suffers either excessive bias or excessive spread for finite sequence length. These two features combine to render classical EVT of questionable value in a gaussian-noise type of environment for these communication systems.

This paper documents the bias and spread of EVT in the given test environment and indicates how to minimize these effects. Specifically, for a given number of test samples, the effects of the number of extremes and the number of samples from which each extreme is drawn are examined.

Classical EVT Suppose we have a set of n independent samples (data points), $x_1, ..., x_n$, of a random variable X with unknown distribution function $F_X$, and we want to estimate $\Pr(X \geq x_0) = 1 - F_X(x_0)$ for fixed threshold $x_0$. We know that the largest (the extreme) value, X, of this sample has a distribution function $\Phi_n(x)$ equal to $F_X^n(x)$, and that if $F_X$ is sufficiently well-behaved (its derivative may have a tail of the form $A \exp(-x^p)$, for example, and be considered well-behaved), then asymptotically (for large enough n), $\Phi_n(x)$ will be of the form

$$\Phi_n(x) = \exp\left\{-\exp\left[-\alpha_n(x - u_n)\right]\right\}$$

where $\alpha_n$ and $u_n$ are unknown parameters. Thus the procedure followed in classical EVT is this: We take several (k) independent groups of data, each group of size n, and from each group we record the largest (the extreme) sample value. Then, using this block of extremes, we estimate $\alpha_n$ and $u_n$ (using, for example, a maximum-likelihood criterion, a minimum distance criterion, or possibly others [1]). Then our estimate of probability of error becomes

$$p = 1 - F_X(x_0)$$
$$= 1 - \Phi_n^{1/n}(x_0)$$
$$= 1 - \left(\exp\left\{-\exp\left[-\alpha_n(x_0 - u_n)\right]\right\}\right)^{1/n}.$$ 

Note that the total number (m) of samples required for the test is $m = n \cdot k$. Unfortunately, if n is too small, the estimate will be biased. We consider why this is so in the case for gaussian noise.
The Bias of EVT  We define a new random variable $V_n$ (the so-called reduced variate) as a function of the underlying random variable $X$ by (letting $\log$ denote natural logarithm)

$$V_n \triangleq -\log \left[ n(1 - F_X) \right].$$

That is, if the random variable $X$ takes on the extreme value $x$, the reduced variate $V_n$ takes on the value

$$v_n(x) = -\log \left\{ n \left[ 1 - F_X(x) \right] \right\}.$$

This can be rewritten to obtain

$$F_X(x) = 1 - \exp \left[ -\frac{v_n(x)}{n} \right]$$

$$\phi_n(x) = F_X^n(x)$$

$$= \left\{ 1 - \exp \left[ -\frac{v_n(x)}{n} \right] \right\}^n$$

$$= \bar{n} \exp \left\{ -\exp \left[ -v_n(x) \right] \right\}.$$

Thus the extremes of $X$ have a distribution function $\Phi_n(x)$, which asymptotically in $n$ has the characteristic double negative exponential form of classical EVT. We also see that convergence to this asymptotic form should be fairly rapid, in terms of increasing $n$. Further, the development so far has left $F_X$ unspecified (within the broad well-behaved class mentioned earlier).

However, this distribution function is in terms of the reduced variate values $v_n(x)$, not the extreme values $x$ themselves. That’s the rub! We measure or record extreme values $x$, but we must convert these extreme values to values $v_n(x)$. And clearly this functional relationship depends on the nature of $F_X$.

If $X$ is exponential, then

$$F_X(x) = 1 - \exp (-\alpha x),$$

and

$$v_n(x) = \alpha(x - u_n),$$
where

\[ u_n \triangleq \left( \frac{\log n}{\alpha} \right). \]

Thus if the noise is exponential, there is a nice linear relationship between the extremes \( x \) and the values \( v_n(x) \) of the reduced variate, and clearly

\[ \Phi_n(x) = \exp \left\{ -\exp \left[ -\alpha(x - u_n) \right] \right\} \]

holds for any reasonable value of \( n \).

But suppose the noise is gaussian (it suffices to consider gaussian noise of mean zero and unit variance). Then

\[
1 - F_X(x) = \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} \exp \left( -\frac{x^2}{2} \right) \, dx
\]

which can be expanded asymptotically as [2]

\[
1 - F_X(x) = \frac{\exp \left( -\frac{x^2}{2} \right)}{x \sqrt{2\pi}} \left[ 1 - R(x) \right]
\]

where \( R(x) \) is positive, small, and falls off as \( 1/x^2 \). Defining

\[ \epsilon \triangleq -\log [1 - R(x)] \]

\[ \rho_n \triangleq \sqrt{2 \log n} \]

\[ \ell_n \triangleq \log \sqrt{2\pi \rho_n^2}, \]

and manipulating, we obtain

\[ v_n(x) = \rho_n(x - \rho_n) + \frac{(x - \rho_n)^2}{2} + \log \left( \frac{x}{\rho_n} \right) + \ell_n + \epsilon. \]

This is clearly not a simple linear relationship. However, for values of \( x \) near \( \rho_n \) i.e., where \( |x - \rho_n| \ll \rho_n \), recalling that \( \log y \sim y - 1 \) for \( y \sim 1 \), we have, neglecting \( \epsilon \) and the quadratic term

\[ v_n(x) \sim \alpha_n(x - u_n), \]
where

\[
\alpha_n \triangleq \rho_n + \frac{1}{\rho_n}
\]

\[
u_n \triangleq \rho_n - \frac{\ell_n}{\alpha_n}.
\]

Thus we have for all \(x\) near \(\rho_n\),

\[
\Phi_n(x) \sim \exp \left\{ -\exp \left[ -\alpha_n (x - u_n) \right] \right\}.
\]

Classical EVT draws a “best” linear fit to the collection of extremes. If we assume we have a sufficient number of extremes that the data “clusters” well around the empirical distribution function, and at least a moderate sample size per extreme, then classical EVT draws a “best” linear fit through \(v_n(x)\). The classical EVT estimate of probability of error is a function only of the ordinate of the linear curve at threshold. Thus for classical EVT to yield accurate results, we require a large enough \(n\) that \(v_n(x)\) is reasonably linear near threshold; i.e., we require \(n\) large enough that \(|x - \rho_n| \ll \rho_n\) for all \(x\) near threshold \(x_0\), or, in particular, we might require \(|x_0 - \rho_n| \ll \rho_n\). For example, if \(x_0 = 4\) (i.e., \(x_0 = 4\sigma\) with \((\sigma = 1)\), and from the above considerations we require \(x_0 = \rho_n\), this necessitates an \(n\) of about 3000. Thus we see the need for some minimal sample size per extreme to guarantee dominance of the linear portion of \(v_n(x)\). For gaussian noise of mean \(m\) and variance \(\sigma^2\), \(\alpha_n\) is scaled by \(1/\sigma\), \(u_n\) is scaled by \(\sigma\) and shifted up by \(m\), and our above criterion for dominance of the linear term near threshold becomes

\[
\left| \frac{x_0 - m}{\sigma} - \rho_n \right| \ll \rho_n.
\]

Furthermore, differentiating \(v_n(x)\) twice, we see that \(v_n(x)\) is convex cup for all \(x\) greater than 1. Thus classical EVT places a “best” linear curve through a convex cup function, assuming we have enough extremes that the data clusters around the curve \(v_n(x)\). We point out that if threshold \(x_0\) exceeds the expected largest value \(u_n\), the linear curve must cross threshold at a lower value than \(v_n(x)\). This is shown pictorially in Fig. 1. This is because most of the recorded extremes used in making the best linear fit lie near \(u_n\), and probably none exceed \(x_0\). (If we have a large number of extremes exceeding threshold, our value of \(n\) is sufficiently large to permit classical bit error testing; thus we assume \(u_n < x_0\)).

Recall now that the true probability of error \(p\) is

\[
p = 1 - \exp \left\{ -\exp \left[ -v_n(x_0) \right] \right\}^{1/n}
\]
while the classical EVT probability of error \( p' \) is

\[
p' = 1 - \left( \exp \{-\exp \left[-\alpha_n(x_0 - u_n)\right]\} \right)^{1/n}
\]

Thus since \( \alpha_n(x_0 - u_n) < v_n(x_0) \), then \( p' > p \); that is, classical EVT estimates an excess probability of error, and thus the bias \( b \triangleq p - p' \) is negative. Clearly, the magnitude of the bias will decrease as \( n \) is increased, since \( v_n(x) \) becomes increasingly linear as \( n \) is increased.

**The Spread of EVT**  Given a set of independent data \((x_1,\ldots, x_k)\), ordered so that \( x_1 \leq x_2 \leq \ldots \leq x_k \), with some distribution function \( \Phi_n(x) \), we note that the fraction of points with value no greater than \( x_i \) (for any \( i - 1, \ldots, k \)) is \( i/k \). Thus we define the empirical distribution function \( e_{nk}(x) \), using the step function \( \mathcal{U} \), by

\[
e_{nk}(x) \triangleq \sum_{i=1}^{k} \frac{i}{k} \mathcal{U}(x - x_i).
\]

Clearly, in the limit, as \( k \) gets large,

\[
e_{nk}(x) \xrightarrow{k} \Phi_n(x).
\]

In classical EVT, we are concerned with log log functions of \( e_{nk}(x) \) and hence, to avoid problems, we cheat a little and define

\[
e_{nk}(x) \triangleq \sum_{i=1}^{k} \frac{i}{1 + k} \mathcal{U}(x - x_i).
\]

Again, clearly,

\[
e_{nk}(x) \xrightarrow{k} \Phi_n(x).
\]

We define the reduced empirical variate \( u_{nk}(x) \)

\[
u_{nk}(x) \triangleq -\log \left[-\log e_{nk}(x)\right]
\]

so that if

\[
\Phi_n(x) \sim \exp \{-\exp \left[-v_n(x)\right]\}
\]

which it does for at least moderate \( n \), then

\[
u_{nk}(x) \xrightarrow{k} v_n(x).
\]
This provides a qualitative measure of the number of extremes needed. We require a sufficient number so that the reduced empirical variate \( u_{nk}(x) \) clusters well around the reduced variate \( v_n(x) \). If \( k \) is too low, then the “best” linear fit of the data will not necessarily “fit” well, no matter how linear \( v_n(x) \) may be. That is, an insufficient number of extremes will cause excessive spread of the parameter estimates in any sequence of tests.

**Experimental Evidence**  We have pointed out that the number \( k \) of extremes used must be large enough so that the reduced empirical variate clusters near the reduced variate, and that the sample size \( n \) from which each extreme is drawn must be large enough to minimize the bias (effect of nonlinearity). For a given number \( m \) of total test size, we must ask what values of \( n \) and \( k \) give a best tradeoff between bias (\( n \) too low) and spread (\( k \) too low). Figures 2 and 3 show the effects of various values of \( n \) and \( k \) as a function of test size \( m \) using independent data from a gaussian data source with known statistics, at a threshold \( x_0 \) of 4 sigma points. Figure 3 summarizes the results at \( m = 3, 12, 30, \) and 60 thousand, for \( k = 10, 20, 30, 50, \) and 100. As \( k \) increases, the spread decreases but moves upward, reflecting the increased bias due to decreasing \( n \).

As a reasonable compromise between the effects of bias and spread, a \( k \) of 30 (30 extremes/total cost) was selected for this threshold of 4\( \sigma \), and EVT was compared with BER using the same gaussian source. The performance curves are shown in Fig. 4. The five solid EVT performance curves are the 90% quantiles, the first sigma points, and the average. The dotted BER performance curves are the 90% confidence curve and the actual bit error count. Clearly, EVT holds no distinct edge over BER for this data source. Comparable tests at other threshold values bear out these conclusions.

For noise with density function of the form

\[
f(x) = \alpha \cdot \exp\left(- \frac{x^r}{\gamma}\right)
\]

for \( r > 1 \), analysis reveals again a nonlinear form of \( v_n(x) \) and a negative bias. Experiments have shown the same tradeoffs between spread and bias and comparable performance.

**Conclusions**  We have shown that bias and spread impose conflicting requirements in the use of EVT for testing digital communication systems, have indicated how to minimize the effects of these factors, and have compared EVT with BER testing of a gaussian source. It seems apparent that under the best of conditions EVT does not perform especially better than BER.
References


![Fig. 1 - Typical Reduced Variables at k=30, n = 100, x₀ = 4](image-url)
Fig. 2 - EVT Average Performance as a Function of $k$ for Gaussian Noise and $x_0 = 4\sigma$

Fig. 3 - EVT 90% Spread as a Function of $k$ for Gaussian Noise and $x_0 = 4\sigma$

Fig. 4 - Comparison of BER and EVT for Gaussian Noise at $k = 30$ and $x_0 = 4\sigma$