Example 1.1 Assume that we have a large, randomly mixed batch of n "good" and "bad" observations x_i of the same quantity μ. Each single observation with probability $1-\varepsilon$ is a "good" one, with probability ε a "bad" one, where ε is a small number. In the former case x_i is $\mathcal{N}(\mu,\sigma^2)$, in the latter $\mathcal{N}(\mu,9\sigma^2)$. In other words all observations have the same mean, but the errors of some are increased by a factor of 3.

Equivalently, we could say that the x_i are independent, identically distributed with the common underlying distribution

$$F(x)=(1-\varepsilon)\Phi\left(\frac{x-\mu}{\sigma}\right)+\varepsilon\Phi\left(\frac{x-\mu}{3\sigma}\right), \tag{1.1}$$

where

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-y^2/2}\,dy \tag{1.2}$$

is the standard normal cumulative.

Two time-honored measures of scatter are the mean absolute deviation

$$d_n=\frac{1}{n}\sum|x_i-\bar{x}| \tag{1.3}$$

and the mean square deviation

$$s_n=\left[\frac{1}{n}\sum(x_i-\bar{x})^2\right]^{1/2}. \tag{1.4}$$

There was a dispute between Eddington (1914, p. 147) and Fisher (1920, footnote on p. 762) about the relative merits of d_n and s_n. Eddington advocated the use of the former: "This is contrary to the advice of most textbooks; but it can be shown to be true." Fisher seemingly settled the matter by pointing out that for normal observations s_n is about 12% more efficient than d_n.

Of course, the two statistics measure different characteristics of the error distribution. For instance, if the errors are exactly normal, s_n converges to σ, while d_n converges to $\sqrt{2/\pi}\ \sigma\approx0.80\sigma$. So we must be precise about how their performances are to be compared; we use the asymptotic relative

efficiency (ARE) of d_n relative to s_n, defined as follows:

$$\mathrm{ARE}(\varepsilon)=\lim_{n\to\infty}\frac{\operatorname{var}(s_n)/(Es_n)^2}{\operatorname{var}(d_n)/(Ed_n)^2}$$

$$=\frac{\left[\dfrac{3(1+80\varepsilon)}{(1+8\varepsilon)^2}-1\right]\Big/4}{\dfrac{\pi(1+8\varepsilon)}{2(1+2\varepsilon)^2}-1}.$$

The results are summarized in the Exhibit 1.1.1.

The result is disquieting: just 2 bad observations in 1000 suffice to offset the 12% advantage of the mean square error, and ARE(ε) reaches a maximum value greater than 2 at about $\varepsilon=0.05$.

This is particularly unfortunate since in the physical sciences typical "good data" samples appear to be well modeled by an error law of the form (1.1) with ε in the range between 0.01 and 0.1. (This does not imply that these samples contain between 1% and 10% gross errors, although this is very often true; the above law (1.1) may just be a convenient description of a slightly longer-tailed than normal distribution.) Thus it becomes painfully clear that the naturally occurring deviations from the idealized model are large enough to render meaningless the traditional asymptotic optimality theory: in practice we should certainly prefer d_n to s_n, since it is better for all ε between 0.002 and 0.5.

ε	ARE(ε)
0	0.876
0.001	0.948
0.002	1.016
0.005	1.198
0.01	1.439
0.02	1.752
0.05	2.035
0.10	1.903
0.15	1.689
0.25	1.371
0.5	1.017
1.0	0.876

ARE + gross error model
B+D p 357
[ch 5]
[§5.4]
+ §3.5.3

Exhibit 1.1.1 Asymptotic efficiency of mean absolute relative to mean square deviation.
From Huber (1977b), with permission of the publisher.