3.6 Robust Statistics

This can easily be done by Monte Carlo, but analytic calculations are also not difficult. Define the mixture by the distribution

$$f(x) = \frac{e^{-\frac{x^2}{2}}(1-p)}{\sqrt{2\pi}} + \frac{e^{-\frac{x^2}{18}}p}{3\sqrt{2\pi}}.$$

The mean value

$$\sqrt{\int_{-\infty}^{\infty} x^2 f(x)\, dx} = \sqrt{1+8p}$$

and the mean deviation

$$2\int_0^{\infty} x\, f(x)\, dx = (1+2p)\sqrt{\frac{2}{\pi}}.$$

A robust width estimator (the interquartile mean c) satisfies

$$\int_{-c}^{c} f(x)\, dx = \frac{1}{2}$$

which gives the equation

$$p\mathrm{Erf}\left[\frac{c}{3\sqrt{2}}\right] + \mathrm{Erf}\left[\frac{c}{\sqrt{2}}\right] - p\mathrm{Erf}\left[\frac{c}{\sqrt{2}}\right] = \frac{1}{2}$$

for c. This has to be solved numerically. The answers are graphed in Figure 1 as a function of p. It is clear that the robust measure is less sensitive to outliers.

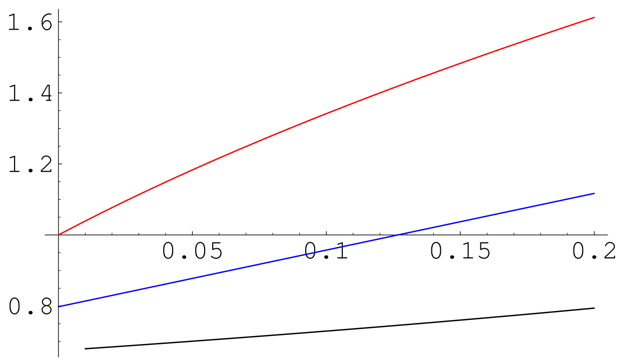

Figure 1: The standard deviation (red), mean deviation (blue) and interquartile range (black) as a function of the outlier parameter p.