REMARK. g^* is called the Bayes decision and $L^* = \mathbf{P}\{g^*(X) \neq Y\}$ is referred to as the Bayes probability of error, Bayes error, or Bayes risk. The proof given above reveals that

$$L(g) = 1 - \mathbf{E}\left\{ I_{\{g(X)=1\}}\eta(X) + I_{\{g(X)=0\}}(1 - \eta(X)) \right\},$$

and in particular,

$$L^* = 1 - \mathbf{E}\left\{ I_{\{\eta(X)>1/2\}}\eta(X) + I_{\{\eta(X)\leq 1/2\}}(1 - \eta(X)) \right\}. \quad \square$$

We observe that the a posteriori probability

$$\eta(x) = \mathbf{P}\{Y = 1 | X = x\} = \mathbf{E}\{Y | X = x\}$$

minimizes the squared error when Y is to be predicted by $f(X)$ for some function $f : \mathcal{R}^d \to \mathcal{R}$:

$$\mathbf{E}\left\{(\eta(X) - Y)^2\right\} \leq \mathbf{E}\left\{(f(X) - Y)^2\right\}.$$

To see why the above inequality is true, observe that for each $x \in \mathcal{R}^d$,

$$\begin{aligned}
&\mathbf{E}\left\{(f(X) - Y)^2 | X = x\right\} \\
&\quad = \mathbf{E}\left\{(f(x) - \eta(x) + \eta(x) - Y)^2 | X = x\right\} \\
&\quad = (f(x) - \eta(x))^2 + 2(f(x) - \eta(x))\mathbf{E}\{\eta(x) - Y | X = x\} \\
&\qquad + \mathbf{E}\left\{(\eta(X) - Y)^2 | X = x\right\} \\
&\quad = (f(x) - \eta(x))^2 + \mathbf{E}\left\{(\eta(X) - Y)^2 | X = x\right\}.
\end{aligned}$$

The conditional median, i.e., the function minimizing the absolute error $\mathbf{E}\{|f(X) - Y|\}$ is even more closely related to the Bayes rule (see Problem 2.12).

2.2 A Simple Example

Let us consider the prediction of a student's performance in a course (pass/fail) when given a number of important factors. First, let $Y = 1$ denote a pass and let $Y = 0$ stand for failure. The sole observation X is the number of hours of study per week. This, in itself, is not a foolproof predictor of a student's performance, because for that we would need more information about the student's quickness of mind, health, and social habits. The regression function $\eta(x) = \mathbf{P}\{Y = 1 | X = x\}$ is probably monotonically increasing in x. If it were known to be $\eta(x) = x/(c + x)$, $c > 0$, say, our problem would be solved because the Bayes decision is

$$g^*(x) = \begin{cases} 1 & \text{if } \eta(x) > 1/2 \text{ (i.e., } x > c) \\ 0 & \text{otherwise.} \end{cases}$$