Competence Development

Bayesian Decision Theory Asymmetric Cost

Jonas Petersen
Ander Runge Walther

Contents

1 Introduction

This note considers how quantile regression can be derived, from a particular cost function, as the optimal decision rule using Bayesian decision theory. Consider a regression scenario where s denote the true value and $U(x, D)$ the estimate based on input x and data D. A linear asymmetric cost function can be written viz

$$C(U(x,D), s) = \alpha \cdot \text{swish}(U(x,D) - s, \beta) + (1 - \alpha) \cdot \text{swish}(s - U(x,D), \beta) \tag{1}$$

where

$$\text{swish}(z, \beta) = \frac{z}{1 + e^{-z\beta}} \tag{2}$$

and $z \equiv U(x,D) - s$ for brevity. For $\beta \gg 1$ equation (1) approximate two linear segments around Origin as shown in figure 1.
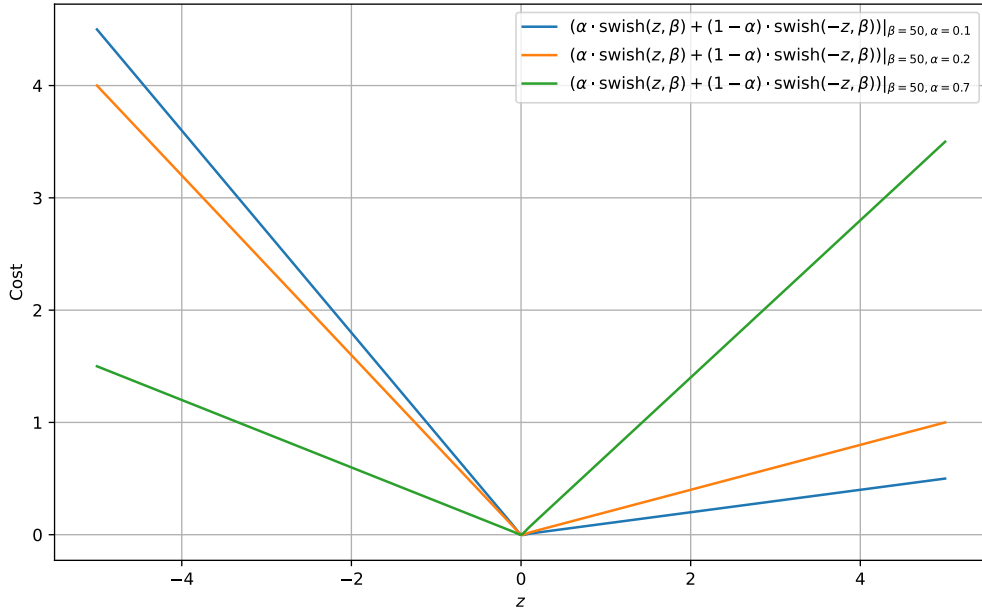

Figure 1: Illustration of equation (1) with different parameter values α, β.

Taking $\alpha \ll 1$, then $z < 0$ will be penalized relatively more than $z > 0$. $z < 0$ corresponds to underestimation, so this is penalized greater relative to overestimation. Now

$$\mathbb{E}[C|x, D, I] = \int ds\, p(s|x, D, I) \Big(\alpha \cdot \text{swish}(U(x,D) - s, \beta) + (1 - \alpha) \cdot \text{swish}(s - U(x,D), \beta) \Big) \tag{3}$$

where I denote the background information (Sivia and Skilling, 2006). The optimal decision rule $U^*(x, D)$ can be found from the derivative

$$
\begin{aligned}
\frac{dC}{dU} &= \frac{dC}{dz}\frac{dz}{dU} \\
&= \left(\frac{\alpha}{1 + e^{-\beta z}} - \frac{1 - \alpha}{1 + e^{\beta z}} + \frac{\alpha\beta e^{-\beta z}z}{(1 + e^{-\beta z})^2} + \frac{(1 - \alpha)\beta e^{\beta z}z}{(1 + e^{\beta z})^2} \right) \frac{dz}{dU} \\
&= \frac{\beta z e^{\beta z} - e^{\beta z} - 1}{(1 + e^{\beta z})^2} + \alpha + \mathcal{O}(\alpha^2) \\
&\approx \alpha - \frac{1}{(1 + e^{\beta z})^2}
\end{aligned}
\tag{4}
$$

where it has been used that $\alpha \ll 1$. Combining equations (3) and (4) alongside imposing the derivative vanish for the optimal decision rule

$$
\begin{aligned}
\frac{d\mathbb{E}[C|x, D, I]}{dU}\bigg|_{U=U^*} &\approx \int ds\, p(s|x, D, I)\left(\alpha - \frac{1}{(1 + e^{\beta z})^2} \right) \\
&= \alpha - \int ds\, p(s|x, D, I)\frac{1}{(1 + e^{\beta z})^2} \\
&= 0
\end{aligned}
\tag{5}
$$

$\frac{1}{(1+e^{\beta z})^2}$ approximate a unit step which is 1 for $z < 0$ and 0 otherwise. $z < 0 \Rightarrow s > U(x)$. This means

$$
\int_{-\infty}^{\infty} ds\, p(s|x, D, I)\frac{1}{(1 + e^{\beta z})^2} \approx \int_{U^*}^{\infty} ds\, p(s|x, D, I)
\tag{6}
$$

This means

$$
\alpha \approx \int_{U^*}^{\infty} ds\, p(s|x, D, I)
\tag{7}
$$

which is the definition of a the α-quantile. Figure 2 show how α and U^* are related via equation (7) in graphical terms (the α-quantile). Figure 3 show the relative difference between numerically optimizing the expected cost and the percentile, as a function of α, for an arbitrary gamma distribution. Numerical imperfections aside, the figure shows that the two are identical. The code can be found in the src folder in the project directory.
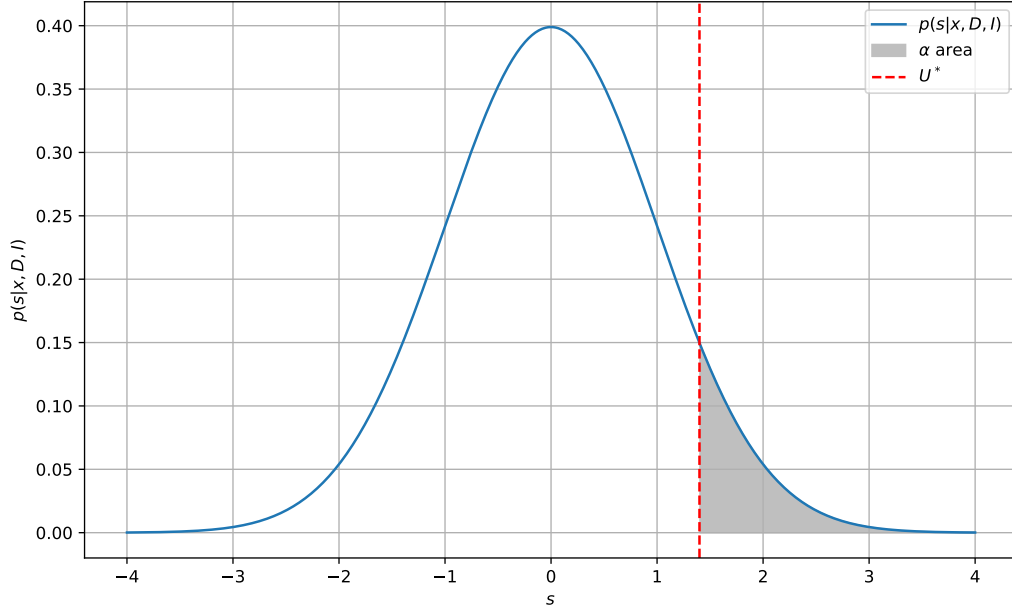
Figure 2: Illustration of equation (7) with $\alpha = 0.05$. A simple normal distribution has been used for illustrative purposes while it is understood equation (7) generalize beyond this simple setting.
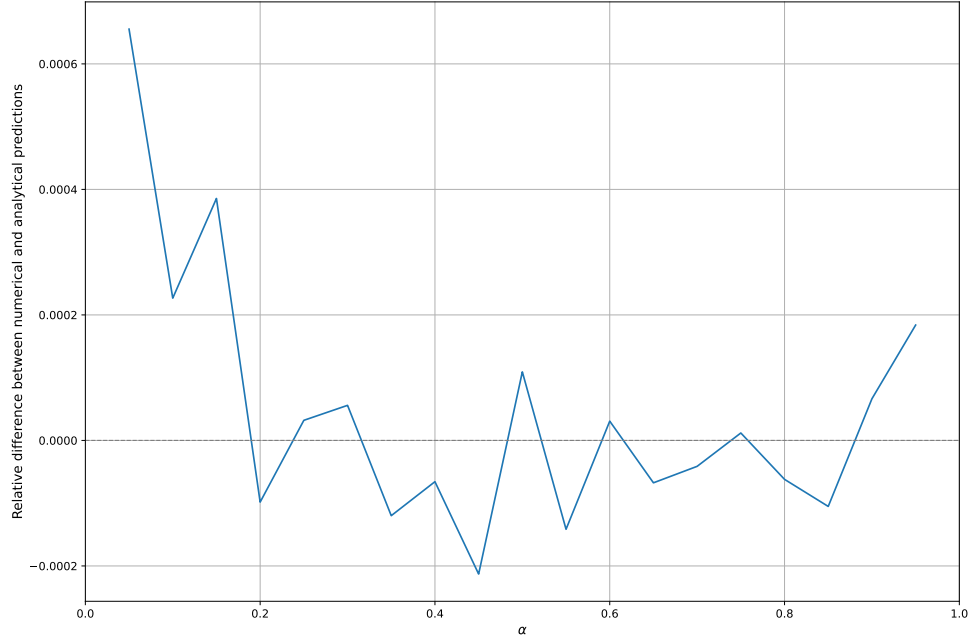

Figure 3: The relative difference between numerically optimizing the expected cost and the percentile, as a function of α.

2 Summary and Discussion

It has been shown how quantile regression can, using Bayesian decision theory, be derived as the optimal decision rule in case of an asymmetric linear cost function. Conversely, it can be deduced that any deviation from such a cost function will yield a different optimal decision rule.

References

D. S. Sivia and J. Skilling. *Data Analysis - A Bayesian Tutorial*. Oxford Science Publications. Oxford University Press, 2nd edition, 2006.