

Bayesian Decision Theory

Median Decision Rule

Jonas Petersen
Ander Runge Walther

October 28, 2024

Abstract

This study investigates the derivation of the median as the optimal decision rule from a specific cost function within the framework of Bayesian decision theory. By exploring the implications of a linear asymmetric cost function, we illustrate the skewness in penalty structures associated with underestimation and overestimation. The analysis leads to the formulation of an expected cost that highlights the importance of the median in decision-making scenarios.

This study considers how the median can be derived as the optimal decision rule from a particular cost function using Bayesian decision theory. Let 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 expressed viz

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

where the function $\text{swish}(z, \beta)$ is defined as

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

For brevity, we set $z \equiv U(x, D) - s$. When $\beta \gg 1$, equation (1) approximates two linear segments around the origin, as illustrated in Figure 1.

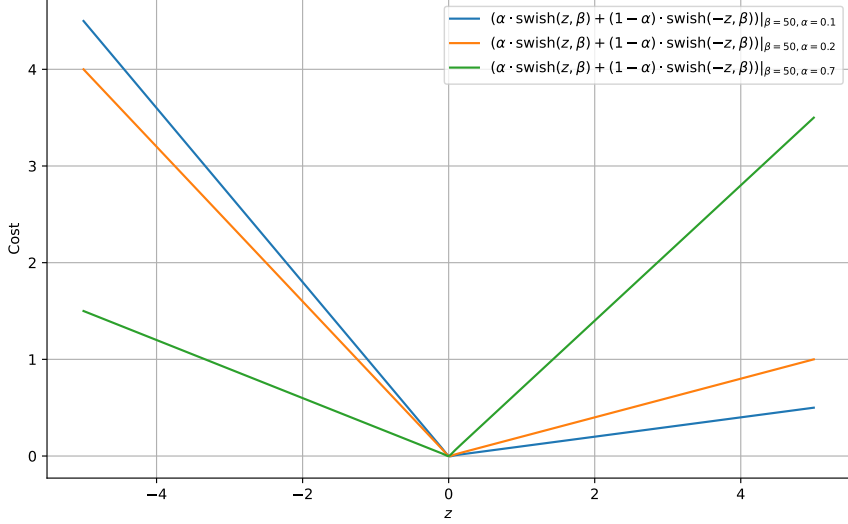


Figure 1: Illustration of equation (1) with different parameter values α and β .

By taking $\alpha \ll 1$, we observe that when $z < 0$ (indicating underestimation), the penalty incurred is relatively greater than for $z > 0$ (indicating overestimation). This skewed penalty structure emphasizes the greater cost associated with underestimating the true value.

We can now compute the expected cost given the input and data as follows

$$\begin{aligned} \mathbb{E}[C|x, D, I] = \int ds p(s|x, D, I) & \left(\alpha \cdot \text{swish}(U(x, D) - s, \beta) \right. \\ & \left. + (1 - \alpha) \cdot \text{swish}(s - U(x, D), \beta) \right), \end{aligned} \quad (3)$$

where I denotes 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

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

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

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

This means

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

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.

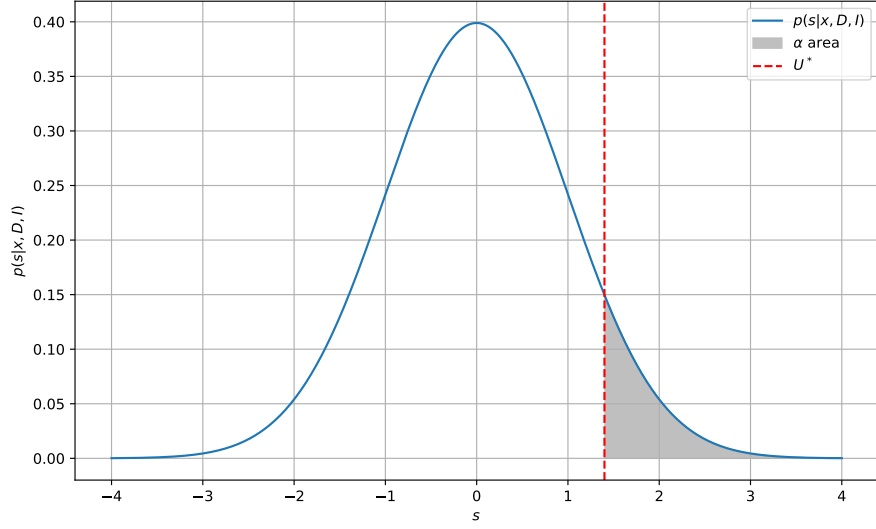


Figure 2: Illustration of equation (7) with $\alpha = 0.05$. A simple normal distribution has been used for illustrative purposes while it is understood equation (7) generalize beyond this simple setting.

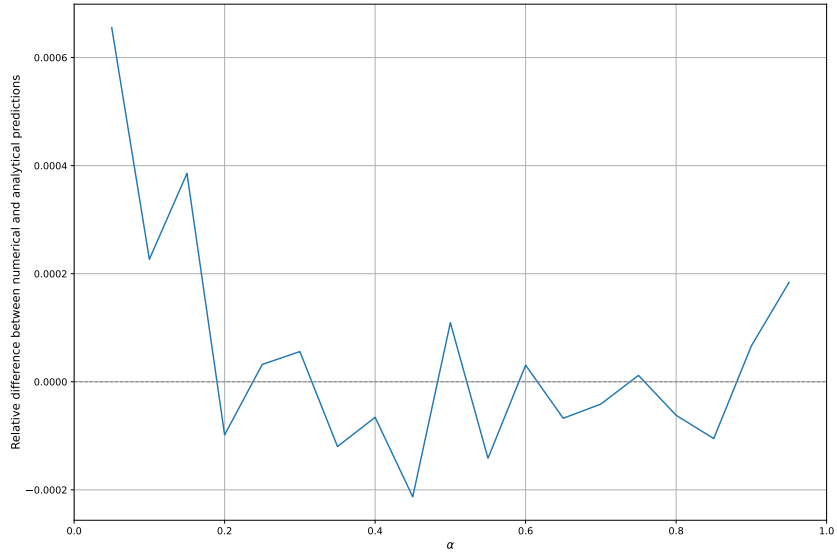


Figure 3: The relative difference between numerically optimizing the expected cost and the percentile, as a function of α .

References

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