

Procedure

For a given trial i , let $X^{(c_1, c_2)} = \{(rt_1, rt_2); (c_1, c_2) = \mathbf{c}\}$ be $n \times d$ sample matrix fall into the choice category c .

1. Transform (rt_1, rt_2) into log-space ($Z^{(c_1, c_2)} = \{(\log(rt_1), \log(rt_2))\}$)
 - To make sure the samples in log-space are always greater than 0 (e.g., $rt_1 < 1\text{ms}$, although it might be impossible to happen), we shift these samples by a small constant $\log(rt) = \log(rt + \text{shift})$
 - After the transformation, Jacobian matrix is required to ensure the probability density remains correct in the original space, that is given by:

$$\hat{f}_X(\mathbf{x}) = \hat{f}_Z(\mathbf{z}) \cdot \left| \frac{\partial \mathbf{z}}{\partial \mathbf{x}} \right| = \hat{f}_Z(\log(\mathbf{z})) \cdot \begin{vmatrix} \frac{1}{x_1 + \text{shift}_1} & 0 \\ 0 & \frac{1}{x_2 + \text{shift}_2} \end{vmatrix}. \quad (1)$$

2. Calculate the covariance matrix Σ and bandwidth matrix H
 - Covariance matrix Σ is calculated by using all these samples in Z
 - Applying either Scott's rule (Eq. 2) or Silverman's rule (Eq. 3) to calculate the bandwidth matrix H

$$H = n^{-1/(d+4)} \Sigma, \quad (2)$$

$$H = \left(\frac{4}{d+2} \right)^{1/(d+4)} n^{-1/(d+4)} \Sigma. \quad (3)$$

- If numbers of samples is smaller than dimension $n < d$, we returned a identity matrix I and a small constant ε as the likelihood for this trial
3. We apply the multivariate gaussian kernel to these sample to estimate the probability density of the joint reaction time distribution

$$\phi_H(\mathbf{u}) = \frac{1}{(2\pi)^{d/2} |H|^{1/2}} \exp\left(-\frac{1}{2} \mathbf{u}^\top H^{-1} \mathbf{u}\right). \quad (4)$$

we need to compute the determinant $|H|$ and the inverse matrix \mathbf{H}^{-1} . To do that, we conduct Cholesky decomposition:

$$H = LL^\top \quad (5)$$

Where $L = \begin{bmatrix} \ell_{11} & 0 \\ \ell_{21} & \ell_{22} \end{bmatrix}$. At this step, Cholesky decomposition requires the principal elements on the diagonal to be positive. Only in this case does the matrix can be positive definite. When the matrix doesn't have full rank(i.e., linearly dependent samples), the matrix can be positive semidefinite, which makes the Cholesky decomposition failed.

- An extreme case of matrix H would be $\begin{bmatrix} \sigma & 0 \\ 0 & 0 \end{bmatrix}$, when the variance of one dimension is σ while the other dimension is 0. In this case, the determinant of the matrix will be 0, which makes the density to be infinite, and the inverse matrix is not normal as well.

4. For the human reaction time data point (v_1, v_2) in this trial to be estimated, we first transform into the log-space following the step 1 and then calculate the vector \mathbf{u} in Eq. 4 by calculating the difference between this data point and the samples in matrix \mathbf{Z} :

$$\mathbf{u} = \mathbf{v}^{\log} - \mathbf{x}_i = \begin{bmatrix} v_1^{\log} - z_{i1} \\ v_2^{\log} - z_{i2} \end{bmatrix}. \quad (6)$$

We can implicitly derive \mathbf{H}^{-1} using $\mathbf{u}^\top \mathbf{H}^{-1} \mathbf{u} = \mathbf{u}^\top (\mathbf{L} \mathbf{L}^\top)^{-1} \mathbf{u} = \mathbf{u}^\top (\mathbf{L}^\top)^{-1} \mathbf{L}^{-1} \mathbf{u}$. Let $\mathbf{t} = \mathbf{L}^{-1} \mathbf{u}$ then

$$\mathbf{u}^\top \mathbf{H}^{-1} \mathbf{u} = \mathbf{t}^\top \mathbf{t} = ||\mathbf{t}||^2 \quad (7)$$

Thus, we need to solve the equation:

$$\mathbf{L} \mathbf{t} = \mathbf{u} \quad (8)$$

Because L is a lower triangular matrix, it can be easily solved using [forward substitution](#). We

can derive $t_1 = \frac{u_1}{\ell_{11}}, \quad t_2 = \frac{u_2 - \ell_{21}t_1}{\ell_{22}}, \mathbf{t} = \begin{bmatrix} \frac{u_1}{\ell_{11}} \\ \frac{u_2 - \ell_{21}t_1}{\ell_{22}} \end{bmatrix}$.

For determinant $|H|$, since $|H| = |L| \cdot |L^\top| = |L|^2, |H|^{1/2} = |L|$.

5. To makes the value more stable, let's take the log transformation on the Eq. 4, then it becomes:

$$\log(\phi_H(u)) = -(\frac{d}{2} \log(2\pi) + |L|) - \frac{1}{2} ||\mathbf{t}||^2 \quad (9)$$

We apply the Eq. 9 to every samples in matrix \mathbf{Z} to estimate the likelihood and take the average:

$$\hat{f}_{\mathbf{H}}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n K_{\mathbf{H}}(\mathbf{x} - \mathbf{x}_i) = \frac{1}{n} \sum_{i=1}^n \exp(\log(\phi_H(u))). \quad (10)$$

At last, we return the log likelihood

$$\log \hat{f}(z) = s_{\max} + \log \left(\sum_i \exp(\log(\phi_H(u)) - s_{\max}) \right) - \log n. \quad (11)$$

Here, s_{\max} is the maximal likelihood among all these simulated samples $s_{\max} = \max_i \log(\phi_H(u))$

6. Now we have $\hat{f}_Z(\log(z))$ in Eq. 1, to transform back to the original space, we need to multiply the Jacobian matrix again

$$\log f_V(v) = \log f_Z(z) - \log(v_1 + \text{shift}_1) - \log(v_2 + \text{shift}_2). \quad (12)$$

That gives us the log-likelihood on the original space.