8 Appendix – Lemmas and Derivations

8.1 Lemmas

Lemma 8.1. Suppose a probability distribution $P(\mathbf{x}_i, \mathbf{x}_y) = f(\mathbf{x}_i, \mathbf{x}_j)$ is defined in terms of $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ik}) \in \mathbb{R}^k$ and $\mathbf{x}_j = (x_{j1}, x_{j2}, \dots, x_{jk}) \in \mathbb{R}^k$. Let $\mathbf{m} = \frac{\mathbf{x}_i + \mathbf{x}_j}{2}$ be the midpoint of \mathbf{x}_i and \mathbf{x}_j , and let $\mathbf{d} = \frac{\mathbf{x}_i - \mathbf{x}_j}{2}$ be the points' (symmetric) displacement from the midpoint. Then $P(\mathbf{m}, \mathbf{d}) = 2^k f(\mathbf{x}_i, \mathbf{x}_j)$.

Likewise, the factor is $\frac{1}{2^k}$ for the inverse transformation: $P(\mathbf{x}_i, \mathbf{x}_j) = \frac{1}{2^k} P(\mathbf{m}, \mathbf{d})$.

Proof. The transformation is one-to-one. For each vector component l, the transforms and their inverses are:

$$m_{l} = \frac{x_{il} + x_{jl}}{2} \qquad x_{il} = m_{l} + d_{l}$$
$$d_{l} = \frac{x_{il} - x_{jl}}{2} \qquad x_{jl} = m_{l} - d_{l}$$

We can show by induction that in k dimensions, the determinant of the Jacobian, $|A_k|$, is -2^k . Then, its absolute value 2^k is the factor used in the transformation.

For a single dimension, $|A_1|$ is

$$\frac{\frac{\partial x_{il}}{\partial m_l}}{\frac{\partial x_{jl}}{\partial m_l}} = 1 \quad \frac{\frac{\partial x_{il}}{\partial d_l}}{\frac{\partial x_{jl}}{\partial m_l}} = 1 \quad \frac{\partial x_{jl}}{\frac{\partial d_l}{\partial d_l}} = -1 = -2 = -2^1.$$

In general,

$$A_{k} = \begin{bmatrix} \frac{\partial x_{i1}}{\partial m_{1}} = 1 & \frac{\partial x_{i1}}{\partial d_{1}} = 1 & \frac{\partial x_{i1}}{\partial m_{2}} = 0 & \frac{\partial x_{i1}}{\partial d_{2}} = 0 & \cdots & \frac{\partial x_{i1}}{\partial m_{k}} = 0 & \frac{\partial x_{i1}}{\partial d_{k}} = 0 \\ \frac{\partial x_{j1}}{\partial m_{1}} = 1 & \frac{\partial x_{j1}}{\partial d_{1}} = -1 & \frac{\partial x_{j2}}{\partial m_{2}} = 0 & \frac{\partial x_{j2}}{\partial d_{2}} = 0 & \cdots & \frac{\partial x_{j1}}{\partial m_{k}} = 0 & \frac{\partial x_{i2}}{\partial d_{k}} = 0 \\ \frac{\partial x_{i2}}{\partial m_{1}} = 0 & \frac{\partial x_{i2}}{\partial d_{1}} = 0 & \frac{\partial x_{i2}}{\partial m_{2}} = 1 & \frac{\partial x_{i2}}{\partial d_{2}} = 1 & \cdots & \frac{\partial x_{i2}}{\partial m_{k}} = 0 & \frac{\partial x_{i2}}{\partial d_{k}} = 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{\partial x_{ik}}{\partial m_{1}} = 0 & \frac{\partial x_{ik}}{\partial d_{1}} = 0 & \frac{\partial x_{ik}}{\partial m_{2}} = 0 & \frac{\partial x_{i2}}{\partial d_{2}} = -1 & \cdots & \frac{\partial x_{ik}}{\partial m_{k}} = 0 & \frac{\partial x_{i2}}{\partial d_{k}} = 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{\partial x_{ik}}{\partial m_{1}} = 0 & \frac{\partial x_{ik}}{\partial d_{1}} = 0 & \frac{\partial x_{ik}}{\partial m_{2}} = 0 & \frac{\partial x_{ik}}{\partial d_{2}} = 0 & \cdots & \frac{\partial x_{ik}}{\partial m_{k}} = 1 & \frac{\partial x_{ik}}{\partial d_{k}} = 1 \\ \frac{\partial x_{ik}}{\partial m_{1}} = 0 & \frac{\partial x_{jk}}{\partial d_{1}} = 0 & \frac{\partial x_{jk}}{\partial m_{2}} = 0 & \frac{\partial x_{jk}}{\partial d_{2}} = 0 & \cdots & \frac{\partial x_{ik}}{\partial m_{k}} = 1 & \frac{\partial x_{ik}}{\partial d_{k}} = -1 \end{bmatrix}$$

$$|A_{k}| = \begin{vmatrix} 1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & -1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & & & \\ \vdots & \vdots & & \begin{bmatrix} A_{k-1} \end{bmatrix} \end{vmatrix} = (-1)|A_{k-1}| - (1)|A_{k-1}| = -2(-2^{k-1}) = -2^{k}$$

Lemma 8.2. Given points \mathbf{x}_i and $\mathbf{x}_j \in \mathbb{R}^k$, with midpoint \mathbf{m}_{ij} and displacement vector \mathbf{d}_{ij} from the midpoint. Define $m = \|\mathbf{m}_{ij} - \boldsymbol{\mu}\|$ and $d = \|\mathbf{d}_{ij}\|$. Then $\|\mathbf{x}_i - \boldsymbol{\mu}\|^2 + \|\mathbf{x}_j - \boldsymbol{\mu}\|^2 = 2(m^2 + d^2)$.

It then follows (see main text, Eqs. (7) and (8)) that if $\phi \sim \text{Normal}(\mu, \sigma^2 I)$, then $P(\mathbf{x}_i \mid \phi) P(\mathbf{x}_i \mid \phi)$ depends on \mathbf{x}_i and \mathbf{x}_j only through m and d.

Proof. The vectors $(\boldsymbol{\mu}, \mathbf{x}_i)$ and $(\boldsymbol{\mu}, \mathbf{x}_j)$ define a plane, which contains \mathbf{m}_{ij} . In that plane, Figure 4 is as shown. The law of cosines tells us that $\|\mathbf{x}_i - \boldsymbol{\mu}\|^2 = m^2 + d^2 - 2md\cos f$, and that $\|\mathbf{x}_j - \boldsymbol{\mu}\|^2 = m^2 + d^2 - 2md\cos(\pi - f)$. Since $\cos(\pi - f) = -\cos f$, we can rewrite $\|\mathbf{x}_j - \boldsymbol{\mu}\|^2 = m^2 + d^2 + 2md\cos f$. This yields $\|\mathbf{x}_i - \boldsymbol{\mu}\|^2 + \|\mathbf{x}_j - \boldsymbol{\mu}\|^2 = 2(m^2 + d^2)$.



Figure 4: Triangle formed by \mathbf{x}_i , \mathbf{x}_j , and $\boldsymbol{\mu}$.

Lemma 8.3. Given $\mathbf{z} \in \mathbb{R}^k$ with magnitude $z = \|\mathbf{z}\|$, and with a probability density that depends only on that magnitude: $f_{\mathbf{z}}(\mathbf{z}) = g(z)$. Then, changing variables to write the density as a function of z gives $f_z(z) = g(z) \frac{2z^{k-1}\pi^{\frac{k}{2}}}{\Gamma(\frac{k}{2})}$.

Explanation. To transform, we would write \mathbf{z} in spherical coordinates ($\mathbf{z} = (z, \alpha_1, \alpha_2, \dots, \alpha_{k-1})$, with $\alpha_1, \dots, \alpha_{k-2} \in [0, \pi]$ and $\alpha_{k-1} \in [0, 2\pi)$), then integrate out the angles. The factor introduced by this process is $S_k = \frac{2z^{k-1}\pi^{\frac{k}{2}}}{\Gamma(\frac{k}{2})}$, the surface area of a hypersphere of radius z in \mathbb{R}^k . (References for the formula: http://mathworld.wolfram.com/Hypersphere.html and http://en.wikipedia.org/wiki/Hypersphere.)

Lemma 8.4. If $f_y(y) = \frac{2^{1-\frac{k}{2}}}{\Gamma(\frac{k}{2})} \frac{y^{k-1}}{a^k} e^{\frac{-y^2}{2a^2}}$, defined on $y \ge 0$ with $a \in \mathbb{R}$ and $k \in \mathbb{Z}, k > 0$, then $z = \frac{y}{a} \sim \chi_k$.

Proof. Perform the change of variables: $z = \frac{y}{a}$, so y = az and $\frac{dy}{dz} = a$.

$$f_y(y) = \frac{2^{1-\frac{k}{2}}}{\Gamma(\frac{k}{2})} \frac{y^{k-1}}{a^k} e^{\frac{-y^2}{2a^2}}$$
(14)

$$f_z(z) = \frac{2^{1-\frac{k}{2}}}{\Gamma(\frac{k}{2})} z^{k-1} \left(\frac{1}{a}\right) e^{\frac{-z^2}{2}} |\frac{dy}{dz}|$$
(15)

$$=\frac{2^{1-\frac{\kappa}{2}}}{\Gamma(\frac{k}{2})}z^{k-1}e^{\frac{-z^2}{2}}$$
(16)

Equation (16) is exactly the density function for the distribution χ_k . We will also use the notation $\chi_k(z)$ to represent that density function, so we would write above that $f_z(z) = \chi_k(z)$ and also that $f_y(y) = \chi_k(\frac{y}{a})(\frac{1}{a})$.

8.2 Distributions for positive and negative pairs

Beginning with the expression from (6) for the density function for positive pairs,

$$P(\mathbf{m}_{ij} \mid \phi) P(\mathbf{d}_{ij} \mid \epsilon) = \left(\frac{1}{\sqrt{2\pi\sigma}}\right)^k e^{-\frac{m^2}{2\sigma^2}} \left(\frac{1}{\sqrt{2\pi\nu}}\right)^k e^{-\frac{d_{ij}^2}{2\nu^2}},\tag{17}$$

we apply Lemma 8.3 to the first and second parts, respectively, to get

$$P(m \mid \phi)P(d \mid \epsilon) = \left(\frac{2m^{k-1}\pi^{\frac{k}{2}}}{\Gamma(\frac{k}{2})}\right) \left(\frac{1}{\sqrt{2\pi\sigma}}\right)^k e^{-\frac{m^2}{2\sigma^2}} \left(\frac{2d^{k-1}\pi^{\frac{k}{2}}}{\Gamma(\frac{k}{2})}\right) \left(\frac{1}{\sqrt{2\pi\nu}}\right)^k e^{-\frac{d_{ij}^2}{2\nu^2}}$$
(18)

$$= \left(\frac{2^{1-\frac{\kappa}{2}}}{\Gamma(\frac{k}{2})}\right) \frac{m^{k-1}}{\sigma^k} e^{-\frac{m^2}{2\sigma^2}} \left(\frac{2^{1-\frac{\kappa}{2}}}{\Gamma(\frac{k}{2})}\right) \frac{d^{k-1}}{\nu^k} e^{-\frac{d^2}{2\nu^2}}.$$
(19)

With the use of Lemma 8.4, we can recognize this as the product of two χ_k distributions: $\frac{m}{\sigma} = m' \sim \chi_k$, and $\frac{d}{\nu} = \frac{d'}{t} \sim \chi_k$. Rewriting in terms of m', d' and t, and using the notation $\chi_k(z)$ to represent the density function for χ_k :

$$P(m' \mid \phi) P(d' \mid \epsilon) = \chi_k(m') \chi_k\left(\frac{d'}{t}\right) \left(\frac{1}{t}\right).$$
⁽²⁰⁾

This is intuitively reasonable because χ_k is known to describe the distance from the origin to points that are distributed as k-dimensional normals, which is exactly where \mathbf{m}_{ij} and \mathbf{d}_{ij} came from.

For the negative density, we take the expression from Equation (8) and transform from coordinates $(\mathbf{x}_i, \mathbf{x}_j)$ to $(\mathbf{m}_{ij}, \mathbf{d}_{ij})$ to (m, d).

$$P(\mathbf{x}_i \mid \phi) P(\mathbf{x}_j \mid \phi) = \left(\frac{1}{\sqrt{2\pi\sigma}}\right)^{2k} e^{-\frac{m^2 + d^2}{\sigma^2}}$$
(21)

$$P(\mathbf{m}_{ij} \mid \phi) P(\mathbf{d}_{jj} \mid \phi) = 2^k \left(\frac{1}{\sqrt{2\pi\sigma}}\right)^k e^{\frac{-m^2}{\sigma^2}} \left(\frac{1}{\sqrt{2\pi\sigma}}\right)^k e^{\frac{-d^2}{\sigma^2}}$$
(22)

$$P(m \mid \phi)P(d \mid \phi) = \left(\frac{2m^{k-1}\pi^{\frac{k}{2}}}{\Gamma(\frac{k}{2})}\right) \left(\frac{1}{\sqrt{\pi}\sigma}\right)^k e^{\frac{-m^2}{\sigma^2}} \left(\frac{2d^{k-1}\pi^{\frac{k}{2}}}{\Gamma(\frac{k}{2})}\right) \left(\frac{1}{\sqrt{\pi}\sigma}\right)^k e^{\frac{-d^2}{\sigma^2}}$$
(23)

$$= \left(\frac{2^{1-\frac{k}{2}}}{\Gamma(\frac{k}{2})}\right) \left(\frac{\sqrt{2}}{\sigma}\right)^{k} m^{k-1} e^{\frac{-m^{2}}{\sigma^{2}}} \left(\frac{2^{1-\frac{k}{2}}}{\Gamma(\frac{k}{2})}\right) \left(\frac{\sqrt{2}}{\sigma}\right)^{k} d^{k-1} e^{\frac{-d^{2}}{\sigma^{2}}}$$
(24)
(25)

Lemma 8.4 now applies, to show this term is also a product of two χ_k distributions: $\frac{m\sqrt{2}}{\sigma} \sim \chi_k$, and $\frac{d\sqrt{2}}{\sigma} \sim \chi_k$, or in terms of m' and d':

$$P(m' \mid \phi) P(d' \mid \phi) = 2\chi_k(m'\sqrt{2})\chi_k(d'\sqrt{2}).$$
(26)

8.3 The effect of increasing k, the number of dimensions

Figure 2 was for 2-dimensional data. In k dimensions the contour lines have the same shape for a given value of t, but as k grows the distributions end up better separated.

The peak of the negatives is always at $(\frac{\sqrt{k-1}}{\sqrt{2}}, \frac{\sqrt{k-1}}{\sqrt{2}})$, and that of the positives is at $(\sqrt{k-1}, t\sqrt{k-1})$. As the dimensionality varies, the relative positions of the peaks are the same apart for the scale factor $\sqrt{k-1}$. However, the variances of the distributions do not scale as fast. So, compared to their positions, the distributions get proportionally narrower and easier to distinguish.



Figure 5: Theoretical distributions of positive and negative pairs, as functions of parameters, and score assigned. From left to right, t takes on values (.02, .3, .7, 1). From top to bottom, number of dimensions k = 1, 2, 10, 100. n = 25, E(r) = 10.