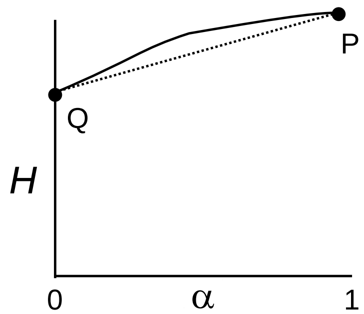
Multivariate Methods

Homework #3 (2020-2021), Answers

Q1. Here we show that the entropy of a mixture is no less than the mixture of the entropies. Given two distributions P *and* Q*, with entropies* $H(P) = -\sum_i p_i \log p_i$ *and* $H(Q) = -\sum_i q_i \log q_i$*, a mixture distribution* $M_\alpha = \alpha P + (1-\alpha)Q$ *is defined by the probabilities* $m_{\alpha,i} = \alpha p_i + (1-\alpha) q_i$*, for* $0 \le \alpha \le 1$*. Show* $H(M_\alpha) \ge \alpha H(P) + (1-\alpha) H(Q)$*. Note that, since* $H(M_0) = H(Q)$ *and* $H(M_1) = H(P)$*, it suffices to show that* $\frac{d^2}{d\alpha^2} H(M_\alpha) \le 0$*, as this means that* $H(M_\alpha)$ *(solid line) is concave downward, and therefore above the line (dashed) of mixtures of entropies.*



This is a straightforward calculation:
$H(M_\alpha) = -\sum_i m_{\alpha,i} \log m_{\alpha,i} = -\sum_i \left(\alpha p_i + (1-\alpha) q_i\right) \log\left(\alpha p_i + (1-\alpha) q_i\right)$.

So

$$\frac{d}{d\alpha} H(M_\alpha) = \frac{d}{d\alpha}\left(-\sum_i \left(\alpha p_i + (1-\alpha) q_i\right) \log\left(\alpha p_i + (1-\alpha) q_i\right)\right) = -\sum_i \left(p_i - q_i\right) \log\left(\alpha p_i + (1-\alpha) q_i\right) - \sum_i \left(p_i - q_i\right) = -\sum_i \left(p_i - q_i\right) \log\left(\alpha p_i + (1-\alpha) q_i\right)$$

, with the last equality because $\sum_i p_i = \sum_i q_i = 1$.

Then
$\frac{d^2}{d\alpha^2} H(M_\alpha) = \frac{d}{d\alpha}\left(-\sum_i \left(p_i - q_i\right) \log\left(\alpha p_i + (1-\alpha) q_i\right)\right) = -\sum_i \frac{\left(p_i - q_i\right)^2}{\left(\alpha p_i + (1-\alpha) q_i\right)}$. Numerators and denominators are both non-negative, so $\frac{d^2}{d\alpha^2} H(M_\alpha) \le 0$, as required. Note that if the denominator of the i th term is zero, then both $p_i = q_i = 0$, and this term can be omitted, as it doesn't contribute to $H(P)$, $H(Q)$, or $H(M_\alpha)$.

Q2: Here we show that the entropy of a joint distribution is maximized when the variables are independently distributed. Let P *be a discrete probability distribution on a set of* M *values* $\{x_i\}$*, i.e.,* P_i *is the probability that a random draw chooses the value* x_i*. Similarly, let* Q *be a discrete probability distribution on a set of* N

values $\{y_j\}$, i.e., Q_j is the probability that a random draw chooses the value y_j. Let R be a discrete distribution on a set of $M \times N$ values $\{(x_i, y_j)\}$, i.e., $R_{i,j}$ is the probability that a random draw chooses the pair of values (x_i, y_j). Find the joint distribution R that maximizes entropy, subject to the constraints that its marginals are compatible with P and Q, i.e., that $P_i = \sum_j R_{i,j}$ and that $Q_j = \sum_i R_{i,j}$. Lagrange multipliers will work nicely.

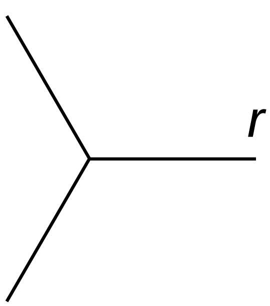
We set up a Lagrange multiplier problem, with M constraints $P_i = \sum_j R_{i,j}$ (one for each i), and N constraints $Q_j = \sum_i R_{i,j}$ (one for each j). Assigning these constraints the multipliers λ_i and μ_j, we need to extremize

$$F(R,\vec{\lambda},\vec{\mu}) = -\sum_{i,j} R_{i,j} \log R_{i,j} - \sum_{i,j} \lambda_i R_{i,j} - \sum_{i,j} \mu_j R_{i,j} .$$

$$\frac{\partial}{\partial R_{k,m}} F(R,\vec{\lambda},\vec{\mu}) = -1 - \log R_{k,m} - \lambda_k - \mu_m .$$

So, $\frac{\partial}{\partial R_{k,m}} F(R,\vec{\lambda},\vec{\mu}) = 0$ implies that $R_{k,m} = e^{-1-\lambda_k-\mu_m}$, which displays $R_{k,m}$ as a product of a function of k and a function of m, but these functions are arbitrary, because the Lagrange multipliers are free to vary. That is, $R_{i,j} = f(i)g(j)$. The constraints are satisfied by choosing $f(i) = P_i$ and $g(j) = Q_j$.

Q3: ICA: toy examples with cubic and quartic surrogates for entropy.



A. Consider the above distribution for bivariate data (centered at the origin), and its projection onto a line whose orientation with respect to the horizontal is given by θ. Determine the angular dependence of the second moment M_2, the third moment M_3, and the fourth moment M_4. Which of these is sensitive to the structure in the data?

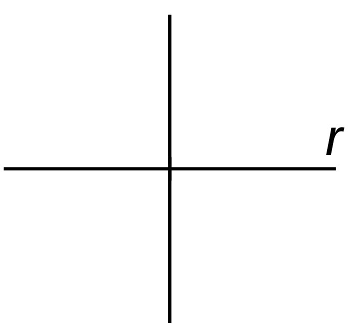
It suffices to consider the angular dependence of points at unit distance from the origin along the spokes, as points at other distances will have the same angular dependence, just contributing more or less to each moment. So $M_k(\theta) = C\sum_{s=0}^{2}\left(\cos\left(\frac{2\pi s}{3} - \theta\right)\right)^k$. We use the complex representation of the cosine, $\cos u = \frac{e^{iu} + e^{-iu}}{2}$, and then the binomial theorem to expand the power.

$M_2(\theta)=\frac{C}{4}\sum_{s=0}^{2}\left(e^{i\left(\frac{2\pi s}{3}-\theta\right)}+e^{-i\left(\frac{2\pi s}{3}-\theta\right)}\right)^2=\frac{C}{4}\sum_{s=0}^{2}\left(e^{2i\left(\frac{2\pi s}{3}-\theta\right)}+2+e^{-2i\left(\frac{2\pi s}{3}-\theta\right)}\right)=\frac{3C}{2}$. The key observation is that in the final step, the expressions with exponentials are the three cube roots of unity – equally-spaced around the unit circle – so they sum to zero.
For the third moment,
$M_3(\theta)=\frac{C}{8}\sum_{s=0}^{2}\left(e^{i\left(\frac{2\pi s}{3}-\theta\right)}+e^{-i\left(\frac{2\pi s}{3}-\theta\right)}\right)^3=\frac{C}{8}\sum_{s=0}^{2}\left(e^{3i\left(\frac{2\pi s}{3}-\theta\right)}+3e^{i\left(\frac{2\pi s}{3}-\theta\right)}+3e^{-i\left(\frac{2\pi s}{3}-\theta\right)}+e^{-3i\left(\frac{2\pi s}{3}-\theta\right)}\right)$. Here, since $3\left(\frac{2\pi s}{3}\right)$ is always a multiple of 2π, the first and last terms are independent of s (so these terms persist), but the middle terms vanish as above, because they are the three cube roots of unity:
$M_3(\theta)=\frac{C}{8}\sum_{s=0}^{2}\left(e^{-3i\theta}+e^{+3i\theta}\right)=\frac{3C}{4}\cos 3\theta$.

For the fourth moment,
$M_4(\theta)=\frac{C}{8}\sum_{s=0}^{2}\left(e^{i\left(\frac{2\pi s}{3}-\theta\right)}+e^{-i\left(\frac{2\pi s}{3}-\theta\right)}\right)^4=\frac{C}{16}\sum_{s=0}^{2}\left(e^{4i\left(\frac{2\pi s}{3}-\theta\right)}+4e^{2i\left(\frac{2\pi s}{3}-\theta\right)}+6+4e^{-2i\left(\frac{2\pi s}{3}-\theta\right)}+e^{-4i\left(\frac{2\pi s}{3}-\theta\right)}\right)=\frac{9C}{8}$. As with the second moment, all terms but the constant term sum to zero -- the exponentials are the three cube roots of unity.

So of these choices for ICA, only M_3 is sensitive to the structure in the data.

B. Same as A, but for this distribution.



$M_2(\theta)=\frac{C}{4}\sum_{s=0}^{3}\left(e^{i\left(\frac{2\pi s}{4}-\theta\right)}+e^{-i\left(\frac{2\pi s}{4}-\theta\right)}\right)^2=\frac{C}{4}\sum_{s=0}^{3}\left(e^{2i\left(\frac{2\pi s}{4}-\theta\right)}+2+e^{-2i\left(\frac{2\pi s}{4}-\theta\right)}\right)=\frac{C}{2}$. The expressions with exponentials are take on the values of +1 and -1, each twice, so their sums vanish.

For the third moment,
$M_3(\theta)=\frac{C}{8}\sum_{s=0}^{3}\left(e^{i\left(\frac{2\pi s}{4}-\theta\right)}+e^{-i\left(\frac{2\pi s}{4}-\theta\right)}\right)^3=\frac{C}{8}\sum_{s=0}^{3}\left(e^{3i\left(\frac{2\pi s}{4}-\theta\right)}+3e^{i\left(\frac{2\pi s}{4}-\theta\right)}+e^{-i\left(\frac{2\pi s}{4}-\theta\right)}+e^{-3i\left(\frac{2\pi s}{4}-\theta\right)}\right)=0$. Here, all terms vanish because they are all sums over the fourth roots ofunity $\{+1,i,-1,-i\}$.

For the fourth moment,

$$M_4(\theta)=\frac{C}{8}\sum_{s=0}^{3}\left(e^{i\left(\frac{2\pi s}{4}-\theta\right)}+e^{-i\left(\frac{2\pi s}{4}-\theta\right)}\right)^4=\frac{C}{16}\sum_{s=0}^{3}\left(e^{4i\left(\frac{2\pi s}{4}-\theta\right)}+4e^{2i\left(\frac{2\pi s}{4}-\theta\right)}+6+4e^{-2i\left(\frac{2\pi s}{4}-\theta\right)}+e^{-4i\left(\frac{2\pi s}{4}-\theta\right)}\right).$$

The second and fourth terms vanish because they take on the values of +1 and -1, each twice. The other terms are independent of s, since the portion of the exponent that depends on s is always an integer multiple of 2π. This yields

$$M_4(\theta)=\frac{C}{4}\left(e^{-4i\theta}+6+e^{\theta 4i}\right)=\frac{C}{2}\left(\cos 4\theta+3\right).$$

So of these choices for ICA, only M_4 is sensitive to the structure in the data.