HW2 Solution, ECE7252, February 6, 2008

1. Let $\mathbf{u}$ be a D-dimension random vector, which has a multivariate normal density with a mean vector μ and an invertible and positive definite covariance matrix Σ, then the probability density function (pdf), $f(\mathbf{u}) = \frac{1}{\sqrt{(2\pi)^D|\Sigma|}} \exp[-\frac{1}{2}(\mathbf{u}-\mu)\Sigma^{-1}(\mathbf{u}-\mu)^t]$. For a bivariate normal random vector, $\mathbf{u} = [x, y]^t$, with $\mu = [\mu_x, \mu_y]^t$ and $\Sigma = [\mathbf{v}_1, \mathbf{v}_2]$, with $\mathbf{v}_1 = [\sigma_x^2, \rho\sigma_x\sigma_y]^t$ and $\mathbf{v}_2 = [\rho\sigma_x\sigma_y, \sigma_y^2]^t$. With the constraints that σ_x and σ_y being positive, and $|\rho| < 1$, then the pdf is simply:

$$f(\mathbf{u}) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp[-\frac{1}{2(1-\rho^2)}\{(\frac{x-\mu_x}{\sigma_x})^2 + 2\rho(\frac{x-\mu_x}{\sigma_x})(\frac{y-\mu_y}{\sigma_y}) + (\frac{y-\mu_y}{\sigma_y})^2\}] \tag{1}$$

Given an iid set of samples, $\mathbf{U} = \{(x_i, y_i), i = 1, ..., N\}$, the log likelihood function can be expressed as: $\mathbf{L}(\mathbf{U}) = \log f(\mathbf{U}) = \mathbf{L}_1 + \mathbf{L}_2(\mathbf{U})$, with

$$\mathbf{L}_1 = -N\log(2\pi) - N\log(\sigma_x) - N\log(\sigma_y) - \frac{N}{2}\log(1-\rho^2), \tag{2}$$

and

$$\mathbf{L}_2(\mathbf{U}) = -\frac{1}{2(1-\rho^2)}\sum_{i=1}^{N}[(\frac{x_i-\mu_x}{\sigma_x})^2 + 2\rho(\frac{x_i-\mu_x}{\sigma_x})(\frac{y_i-\mu_y}{\sigma_y}) + (\frac{y_i-\mu_y}{\sigma_y})^2]. \tag{3}$$

The maximum likelihood solution for μ_x and μ_y can be solved separately by taking derivatives of the log likelihood function over the two mean parameters to get:

$$-\frac{2}{\sigma_x}\sum_{i=1}^{N}[(\frac{x_i-\mu_x}{\sigma_x}) + \rho(\frac{y_i-\mu_y}{\sigma_y})] = 0, \tag{4}$$

and

$$-\frac{2}{\sigma_y}\sum_{i=1}^{N}[\rho(\frac{x_i-\mu_x}{\sigma_x}) + (\frac{y_i-\mu_y}{\sigma_y})] = 0. \tag{5}$$

It is easy to show that the two individual sum over the two random variables are zero, or:

$$\hat{\mu}_x = \frac{1}{N}\sum_{i=1}^{N} x_i = \bar{x}, \tag{6}$$

and

$$\hat{\mu}_y = \frac{1}{N}\sum_{i=1}^{N} y_i = \bar{y}. \tag{7}$$

The next three parameters have to be solved together by taking derivatives and set them to zero, we have the following three equations after some simplifications:

$$-N + \frac{1}{(1-\rho^2)} \sum_{i=1}^{N} [(\frac{x_i - \mu_x}{\sigma_x})^2 + \rho(\frac{x_i - \mu_x}{\sigma_x})(\frac{y_i - \mu_y}{\sigma_y})] = 0, \tag{8}$$

$$-N + \frac{1}{(1-\rho^2)} \sum_{i=1}^{N} [\rho(\frac{x_i - \mu_x}{\sigma_x})(\frac{y_i - \mu_y}{\sigma_y}) + (\frac{y_i - \mu_y}{\sigma_y})^2] = 0, \tag{9}$$

and

$$-\frac{N\rho}{(1-\rho^2)} - \frac{\rho}{(1-\rho^2)}\mathbf{L}_2(\mathbf{U}) - \frac{1}{(1-\rho^2)} \sum_{i=1}^{N} [(\frac{x_i - \mu_x}{\sigma_x})(\frac{y_i - \mu_y}{\sigma_y})] = 0. \tag{10}$$

By summing Equations (8) and (9), it is clear that $\mathbf{L}_2(\mathbf{U}) = -2N$. Plugging it into Equation (10), we have the maximum likelihood estimate of ρ as:

$$\hat{\rho} = \frac{1}{N} \sum_{i=1}^{N} [(\frac{x_i - \bar{x}}{\sigma_x})(\frac{y_i - \bar{y}}{\sigma_y})]. \tag{11}$$

Now by plugging the summation in Equation (10) into Equation (8), and after some simplification, we have

$$\frac{1}{(1-\rho^2)} \sum_{i=1}^{N} (\frac{x_i - \bar{x}}{\sigma_x})^2 = N[1 + \frac{\rho^2}{(1-\rho^2)}], \tag{12}$$

the ML estimate for the variances can therefore be solved as:

$$\hat{\sigma}_x^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2, \hat{\sigma}_y^2 = \frac{1}{N} \sum_{i=1}^{N} (y_i - \bar{y})^2. \tag{13}$$

Finally by plugging Equation (13) into Equation (11), we have

$$\hat{\rho} = \frac{\sum_{i=1}^{N} [(x_i - \bar{x})(y_i - \bar{y})]}{\sqrt{[\sum_{i=1}^{N} (x_i - \bar{x})^2][\sum_{i=1}^{N} (y_i - \bar{y})^2]}}. \tag{14}$$

Therefore the maximum likelihood estimates of the five parameters are shown in Equations (6), (7), (13), and (14), respectively. Since they are are statistics of the observations, you can take mathematical expectation of them. If they are equal to the parameters to be estimated, they are called unbiased estimates. Are our ML estimates unbiased? You can also compute the second order expectation to obtain the variance of the ML estimates. They will give you some idea if the ML estimates are good, and if there are other estimates that may potentially be better than ML because they may offer less variances. The "real" parameters used to generate the data: $\mu_x = 4$, $\mu_y = -3$, $\sigma_x = 1$, $\sigma_y = 2$, and $\rho = -0.7$.

2. The multinomial distribution (probability mass) function of a random vector $\mathbf{R} = [r_1, ..., r_M]$ can be expressed as

$$P(r_1, ..., r_m, ..., r_M, N|p_1, ..., p_m, ..., p_M) = \frac{N!}{\prod_{m=1}^{M} r_m!} \prod_{m=1}^{M} {p_m}^{r_m}, \tag{15}$$

with the constraints that $\sum_{m=1}^{M} r_m = N$, $0 \leq r_m \leq N$, $0 < p_m < 1$, and $\sum_{m=1}^{M} p_m = 1$. Of course we need to satisfy the probability requirement that

$$\sum_{all\mathbf{R}} P(r_1, ..., r_m, ..., r_M, |p_1, ..., p_m, ..., p_M) = \sum_{all\mathbf{R}} \frac{N!}{\prod_{m=1}^{M} r_m!} \prod_{m=1}^{M} {p_m}^{r_m} = 1. \tag{16}$$

To obtain the expectation $E[\mathbf{R}]$, some understanding about the above equality is needed. To evaluate $E[r_m]$, the following equality is essential, by setting another random variable $s_m = (r_m - 1)$ (a trivial modification because $s_m = 0$ will not affect the following summation):

$$E[r_m] = \sum_{all\mathbf{R}} \frac{N!}{\prod_{m=1}^{M} r_m!} r_m \prod_{m=1}^{M} {p_m}^{r_m} = (Np_m) \sum_{all\mathbf{R}} P(r_1, ..., s_m, ..., r_M, (N-1)|p_1, ..., p_m, ..., p_M) = Np_m. \tag{17}$$

The same trick can be used to compute $E[r_m^2] = E[r_m(r_m - 1) + r_m]$, or

$$Var[r_m] = E[r_m^2] - (E[r_m])^2 = N(N-1)p_m^2 + Np_m - (Np_m)^2 = Np_m(1 - p_m). \tag{18}$$

Similarly the cross-correlation can also be computed as follows using the same trick,

$$Cov[r_m, r_l] = E[(r_m - E[r_m])(r_l - E[r_l])] = N(N-1)p_m p_l - (Np_m)(Np_l) = -Np_m p_l. \tag{19}$$

Now to obtain the maximum likelihood estimate of the membership probability, p_m, given a random observation vector, $\mathbf{Q} = [\mathbf{r_1} = \mathbf{q_1}, ..., \mathbf{r_m} = \mathbf{q_m}, ..., \mathbf{r_M} = \mathbf{q_m}]$ with the constraints, $\sum_{m=1}^{M} q_m = N$, and $\sum_{m=1}^{M} p_m = 1$. The log likelihood function can be expressed as:

$$L(\mathbf{Q}|p_1, ..., p_M) = \log[P(q_1, ..., q_M, N|p_1, ..., p_M)] = (Constant) + \sum_{m=1}^{M} [q_m \log p_m]. \tag{20}$$

We now define an objective function to be optimized through Lagrange formulation as:

$$F(p_1, ..., p_M) = L(\mathbf{Q}|p_1, ..., p_M) + \lambda[\sum_{m=1}^{M} p_m - 1]. \tag{21}$$

To maximize the above we can take derivatives with respect to p_m and λ to obtain

$$\frac{q_m}{p_m} + \lambda = 0, m = 1, ...M, \tag{22}$$

and

$$\sum_{m=1}^{M} p_m = 1. \tag{23}$$

Now summing over the M equations in Equation (22) and the apply the constraint in Equation (23), we have $\lambda = -N$, therefore the ML estimate for p_m is simply

$$\hat{p}_m = \frac{q_m}{N}. \tag{24}$$