PROBLEM SET 1
STAT 221

WON I. LEE

1. QUESTION 1

1.1. We write $x = \log\left(\frac{u}{1-u}\right)$ elementwise, where $u \in \mathbb{H}^d$. We then have:

$$f_U(u) = f_X(g^{-1}(u))\left\|\frac{dx}{du}\right\|$$

as given by the transformation of variables formula. We know f_X by the multivariate Gaussian, so all that remains is to compute the Jacobian. Since we have the function g^{-1}, we can compute:

$$\frac{\partial x_i}{\partial u_j} = \frac{\partial}{\partial u_j}\log\left(\frac{u_i}{1-u_i}\right) = \begin{cases} i = j: & \frac{\partial}{\partial u_i}[\log u_i - \log(1-u_i)] = \frac{1}{u_i} + \frac{1}{1-u_i} = \frac{1}{u_i(1-u_i)} \\ i \neq j: & 0 \end{cases}$$

Thus, we have that the Jacobian is simply a diagonal matrix with entries $\frac{1}{u_i(1-u_i)}$ on the diagonal, so that:

$$\left\|\frac{dx}{du}\right\| = \frac{1}{\prod_{i=1}^d u_i(1-u_i)}$$

We can now plug this into the original transformation of variables formula and immediately obtain the desired expression:

$$f_U(u) = \frac{1}{|2\pi\Sigma|^{1/2}\prod_i u_i(1-u_i)}\exp\left[-\frac{1}{2}\left(\log\left(\frac{u}{1-u}\right) - \mu\right)^T \Sigma^{-1}\left(\log\left(\frac{u}{1-u}\right) - \mu\right)\right]$$

1.2. We take the Taylor series of $u_i = g_i(x_i)$ componentwise:

$$g_i(x_i) = g_i(\mu_i) + g_i'(\mu_i)(x_i - \mu_i) + \frac{1}{2}g_i''(\mu_i)(x_i - \mu_i)^2 + \cdots$$

but we note that the third-order term $(x - \mu)^3$ has expectation zero, since the Normal distribution has no higher-order odd moments. Thus, we can write (as an approximation):

$$E[u_i] = E[g_i(x_i)] = g_i(\mu_i) + g_i'(\mu_i)E[x_i - \mu_i] + \frac{1}{2}g_i''(\mu_i)E[(x_i - \mu_i)^2]$$

but clearly $E[x_i - \mu_i] = 0$ and $E[(x_i - \mu_i)^2] = \text{Var}(x_i) = \alpha$, so that we simply need to evaluate $g_i''(\mu_i)$. We note that in fact:

$$\frac{du_i}{dx_i} = u_i(1-u_i)$$

1

and so:

$$g_i'' = \frac{d^2 u_i}{dx_i^2} = \frac{d}{dx_i} u_i(1 - u_i) = u_i(1 - u_i)(1 - 2u_i)$$

after simplification. Define:

$$\tilde{\mu}_i \equiv g_i(\mu_i)$$

Plugging in $u_i = g_i(x_i)$ and setting $x_i = \mu_i$, we have:

$$g_i''(\mu_i) = \tilde{\mu}_i(1 - \tilde{\mu}_i)(1 - 2\tilde{\mu}_i)$$

Thus, we have:

$$E[u_i] = \tilde{\mu}_i + \frac{\alpha}{2}\tilde{\mu}_i(1 - \tilde{\mu}_i)(1 - 2\tilde{\mu}_i)$$

1.3. Assume we have n observations $u^i, \ldots, u^n$. Note that the transformation $u_= g(x)$ is smooth and does not depend on the parameters. Thus, by the invariance of the MLE (or more simply noting that the PDF of the distribution of U has no additional dependence on the parameters relative to the standard multivariate Normal), we see that the MLE of μ, α, β is identical to estimating the MLE using the multivariate Normal, with the x values evaluated by $g^{-1}(u)$. Thus, immediately:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x^i = \frac{1}{n}\sum_{i=1}^{n} \log(u^i/(1 - u^i))$$

We also know that the MLE of the covariance matrix Σ for a multivariate Normal is given by:

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(x^i - \hat{\mu})(x^i - \hat{\mu})^T$$

By the transformation of parameters $(\mu, \Sigma) \to (\mu, \alpha, \beta)$, we can see that:

$$\alpha = \frac{1}{n}\sum_{i=1}^{n}\Sigma_{ii}$$

$$\beta = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j \neq i}\Sigma_{ij}$$

and so we can use invariance to transform the MLE of Σ to the MLE of α, β:

$$\hat{\alpha} = \frac{1}{n}\sum_{i=1}^{n}\hat{\Sigma}_{ii}$$

$$\hat{\beta} = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j \neq i}\hat{\Sigma}_{ij}$$

1.4. Please see attached R code.

1.5. We report the results given the parameters, using 250 points for every example, in Table 1.

1.6. The MLE we compute are in Table 2.

μ	α	β	$\hat{\mu}$	$\hat{\alpha}$	$\hat{\beta}$
(5,5)	10	1	(5.07, 5.18)	9.597	0.804
(1,1,10)	1	0.1	(0.92, 1.03, 9.96)	0.924	0.064
(5,-5,10,-10)	50	5	(4.52, -5.34, 9.92, -9.54)	48.568	3.163

TABLE 1. Problem 1.5

$\hat{\mu}$	$\hat{\alpha}$	$\hat{\beta}$
(0.379, 0.366)	2.915	0.75

TABLE 2. Problem 1.6

1.7. We can employ the bootstrap method to obtain an approximate variance estimate of the MLE. That is, we subsample from our original data with replacement, and obtain the MLE for the subsampled data. A histogram of these subsampled MLE values can provide an idea of the variability of the original MLE. We conduct 1000 such bootstrap simulations, and provide the histograms for this data in `wonlee_fig_1.pdf`.

1.8. In this model, we simply have $d \cdot n$ independent Beta variates, where d is the dimension and n is the number of data points. Thus, assuming that the data are given by:

$$u_j^i \sim \mathcal{B}(a, b)$$

the likelihood is simply:

$$L(a, b) = \prod_{i=1}^{n} \prod_{j=1}^{d} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} (u_j^i)^{a-1} (1 - u_j^i)^{b-1}$$

Because MLE for the parameters of the Beta is generally intractable, we employ a method of moments estimator for a, b:

$$\hat{a} = \bar{x} \left(\frac{\bar{x}(1 - \bar{x})}{s^2} - 1 \right)$$

$$\hat{b} = (1 - \bar{x}) \left(\frac{\bar{x}(1 - \bar{x})}{s^2} - 1 \right)$$

where $\bar{x}$ is the sample mean and s^2 is the sample variance.

The histograms of the data alongside the PDF of the estimated Beta density are given in `wonlee_fig_2.pdf`. The histograms do not reveal a substantial error of goodness-of-fit using the Beta density. We explore further by using the Kolmogorov-Smirnov test. Using each dimension as a univariate distribution with the same parameters, we obtain the following test statistics (using the method of moments estimated parameters for the Beta and MLE parameters for the hypercube), in Table 3.

We see that none of the distributions fail the goodness-of-fit test. Interestingly, the first dimension has better fit with the estimated Beta distribution, whereas the second dimension shows better fit with the hypercube distribution.

Variable	Distribution	T	p-value
u_1	Beta	0.06	0.97
u_2	Beta	0.075	0.8475
u_1	Hypercube	0.085	0.721
u_2	Hypercube	0.055	0.988

TABLE 3. Problem 1.8

2. QUESTION 2

2.1. The relevant variables are G, Θ, where:

$$G = \begin{pmatrix} g_{L,1} & g_{H,1} \\ \vdots & \vdots \\ g_{L,N} & g_{H,N} \end{pmatrix}$$

$$\Theta = \begin{pmatrix} \theta_{L,1} & \cdots & \theta_{L,P} \\ \theta_{H,1} & \cdots & \theta_{H,P} \end{pmatrix}$$

Then the log-likelihood is:

$$l(G,\theta) = \log P(X|G,\Theta) = \sum_{n=1}^{N}\sum_{p=1}^{P} \log P(X_{np}|G_{n,\cdot}, \Theta_{\cdot,p}) = \sum_{n=1}^{N}\sum_{p=1}^{P} \log \left[g_{L,n}\theta_{L,p,X_{np}} + g_{H,n}\theta_{H,p,X_{np}} \right]$$

where each $\theta_{L,p}, \theta_{H,p}$ is assumed to be a vector of multinomial probabilities $\theta_{L,p} = (\theta_{L,p,1}, \ldots, \theta_{L,p,m_p})$ and thus $\theta_{L,p,X_{np}}$ selects the vector component according to the observed X_{np}, and similarly for H.

2.3. We can write the derivatives as follows:

$$\frac{\partial l}{\partial g_{ij}} = \sum_{n=1}^{N}\sum_{p=1}^{P} \frac{\theta_{i,p,X_{np}}\mathbb{1}_{n=j}}{g_{L,n}\theta_{L,p,X_{np}} + g_{H,n}\theta_{H,p,X_{np}}} = \sum_{p=1}^{P} \frac{\theta_{i,p,X_{jp}}}{g_{L,j}\theta_{L,p,X_{jp}} + g_{H,j}\theta_{H,p,X_{jp}}}$$

$$\frac{\partial l}{\partial \theta_{ijk}} = \sum_{n=1}^{N}\sum_{p=1}^{P} \frac{g_{i,n}\mathbb{1}_{X_{np}=k}}{g_{L,n}\theta_{L,p,X_{np}} + g_{H,n}\theta_{H,p,X_{np}}}$$

and perform coordinate ascent.