Multivariate Statistical Analysis, Exercise 3, Fall 2011, Prof. S.K. Jeng

October 14, 2011 TA: H.C. Cheng

(χ^2 Distribution)

The population in *Exercise 1* is indeed a multivariate Gaussian distribution with the mean vector $\mu = [-0.8\ 0.4]'$, and the covariance matrix $\Sigma = \begin{pmatrix} 0.8000 & 0.7505 \\ 0.7505 & 1.1000 \end{pmatrix}$. Now according to the random sample from *Exercise 1*, please answer the following questions.

(a) The random sample in *Exercise 1* contains 10000 data, *i.e.* $\mathbf{x}_n$, $n = 1, \ldots, 10000$. Now transform the data to a new data $\mathbf{Y} = [y_1, \ldots, y_{10000}]'$; that is

$$y_n = (x - \mu)'\Sigma^{-1}(x - \mu),\ n = 1, \ldots, 10000.$$

Make the histogram of the transformed data $\mathbf{Y}$ and also plot the curve of the χ^2 distribution with degrees of freedom 2 on the same figure. The probability density function (pdf) of χ^2_ν is

$$f_\nu(\chi^2) = \begin{cases} \frac{1}{2^{\nu/2}\Gamma(\nu/2)} (\chi^2)^{\nu/2-1} e^{-\chi^2/2}, & \chi^2 \geq 0; \\ 0, & \text{otherwise.} \end{cases},$$

where the *gamma function* is defined by

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt,\ x \in \mathbb{R}$$

(in our case of degrees of freedom 2, $\Gamma(1) = 1$).

(b) From the histogram derived in (a), we approximate the pdf of $\mathbf{Y}$. Now use the approximated pdf to calculate

$$\Pr\Big[Y \leq \chi^2_2(\alpha)\Big] = \Pr\Big[(\mathbf{X} - \mu)'\Sigma^{-1}(\mathbf{X} - \mu) \leq \chi^2_2(\alpha)\Big],$$

where $\alpha = 0.05$ in this case and $\chi^2_2(\alpha)$ denotes the upper (100α)-th percentile of the $\chi^2_2(\alpha)$ distribution (it is equal to 5.99 from the table). Is it near $1 - \alpha$?