FINDING THE MOMENTS OF GENERAL QUADRATIC FORM WITH APPLICATION TO DATA CLASSIFICATION

William W.S. Chen
Department of Statistics
The George Washington University
Washington D.C. 20013

ABSTRACT

We assume a random sample of size k of general quadratic form has been drawn. We wish to find the first four moments of their more general formula. Later, we use these moments to classify a randomly observed vector to one of the two multivariate normal distributions. We also give out the probability of our decision correctly or incorrectly in this classification. As the concluding remark, we give two real life examples that have been published in literature. Kendall and Stuart [1] discussed the case when k=2. Anderson T.W.[2] has a whole chapter 6 discussing the classification of the column vector problem.

Keywords: Characteristic Function, Cumulant Generating Function, Data Classification, First Four Moments, General Quadratic Form, Multivariate Normal Distribution, Real Life Examples.

INTRODUCTION

We assume a random sample of size k of quadratic form has been drawn from multivariate normal distribution. We are interested in finding the first four moments of these quadratic forms. However, this may involve many matrix operations. For example, we cannot operate two column vectors of different sizes or two matrices that are not conformable. We need to carefully define our notation or symbols. In this paper, we define our column vector with bar under-score, such as $\underline{x}$ and corresponding row vector with a prime on the top-right corner, such as $\underline{x}'$. We should be aware of the fact that if an expression starts with a row vector and ends with a column vector, then it always represents a scalar. With respect to the scalar, we may differentiate or integrate as many times as we wish, when it exists. The problem of classification arises when a researcher makes a number of measurements on an individual and wishes to classify the individual into one of several categories on the basis of these measurements. A researcher cannot identify the individual with a category directly, but must use these measurements. We usually assume that there are a finite number of populations from which the individual may have come from. We may also assume that each population has been characterized by a probability distribution of the measurements. Thus, an individual is considered as a random observation from this population. Then the question turns out to be: "given an individual with certain measurements, how do we classify this person?" In this paper, we only consider the case where two populations are admitted; hence we may test one hypothesis of a specified distribution against another. If there is a quota between the populations, it is also possible to create such a test. As the concluding remark, we have given two real-life examples to demonstrate this problem.

Distribution of the Quadratic Forms

Let Q_j is an arbitrary symmetric matrix. Let column vector $\underline{X}$ has multivariate normal distribution with mean vector $\underline{\mu}$ and variance-covariance matrix Σ. Then $q_j = \underline{X}'Q\underline{X}$ has quadratic form. We will find the characteristics function of the joint distribution of $q_1, q_2, q_3, \ldots q_k$. In general, we may assume $q_j = (\underline{x} - \underline{\mu_j})' Q_j (\underline{x} - \underline{\mu_j});$ and hence;

$$\Psi_{q_1, q_2, \ldots q_k}(t_1, t_2, \ldots t_k) = E[e^{it_1 q_1 + it_2 q_2 + \ldots + it_k q_k}] = E[e^{i \sum_{j=1}^{k} t_j q_j}]$$

$$= \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \int \ldots \int e^{i \sum_{j=1}^{k} t_j q_j - \frac{1}{2}(\underline{x} - \underline{\mu})' \Sigma^{-1} (\underline{x} - \underline{\mu})} d\underline{x}$$

Consider the exponential part alone;

$$i \sum_{j=1}^{k} t_j q_j - \frac{1}{2}(\underline{x} - \underline{\mu})' \Sigma^{-1} (\underline{x} - \underline{\mu})$$

$$= i \sum_{j=1}^{k} t_j (\underline{x} - \underline{\mu_j})' Q_j (\underline{x} - \underline{\mu_j}) - \frac{1}{2}(\underline{x} - \underline{\mu})' \Sigma^{-1} (\underline{x} - \underline{\mu})$$

$$= -\frac{1}{2} \underline{x}' [\Sigma^{-1} - 2i \sum_{j=1}^{k} t_j Q_j] \underline{x} - \frac{1}{2} \underline{x}' [-\Sigma^{-1} \underline{\mu} + 2i \sum_{j=1}^{k} t_j Q_j \underline{\mu_j}]$$

$$- \frac{1}{2} [-\underline{\mu}' \Sigma^{-1} + 2i \sum_{j=1}^{k} t_j \underline{\mu_j}' Q_j] \underline{x} - \frac{1}{2} [\underline{\mu}' \Sigma^{-1} \underline{\mu} - 2i \sum_{j=1}^{k} t_j \underline{\mu_j}' Q_j \underline{\mu_j}]$$

$$Assume\ v_t^{-1} = \Sigma^{-1} - 2i \sum_{j=1}^{k} t_j Q_j \quad and\ \underline{\beta_t} = v_t^{-1} \underline{\mu} = \Sigma^{-1} \underline{\mu} - 2i \sum_{j=1}^{k} t_j Q_j \underline{\mu_j}$$

$$c = \underline{\mu}' \Sigma^{-1} \underline{\mu} - 2i \sum_{j=1}^{k} t_j \underline{\mu_j}' Q_j \underline{\mu_j} \quad aware\ that\ v_0^{-1} = \Sigma^{-1} \quad and\ \underline{\beta_0} = \Sigma^{-1} \underline{\mu}$$

Exponential can be written as; $i \sum_{j=1}^{k} t_j (\underline{x} - \underline{\mu_j})' Q_j (\underline{x} - \underline{\mu_j}) - \frac{1}{2}(\underline{x} - \underline{\mu})' \Sigma^{-1} (\underline{x} - \underline{\mu})$

$$= -\frac{1}{2} \underline{x}' v^{-1} \underline{x} + \frac{1}{2} \underline{x' \beta} + \frac{1}{2} \underline{\beta' x} - \frac{1}{2} c$$

$$= -\frac{1}{2} (\underline{x} - \underline{v\beta})' v^{-1} (\underline{x} - \underline{v\beta}) + \frac{1}{2} \underline{\beta' v \beta} - \frac{1}{2} c$$

Finally, the joint characteristic function is:

$$\Psi_{q_1 q_2 \ldots q_k}(t_1, t_2, \ldots t_k) = \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} e^{\frac{1}{2} \underline{\beta' v \beta} - \frac{1}{2} c} \int \ldots \int exp[-\frac{1}{2} (\underline{x} - \underline{v\beta})' v^{-1} (\underline{x} - \underline{v\beta})] dx$$

$$= \frac{(2\pi)^{\frac{p}{2}} |v|^{\frac{1}{2}}}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} exp[\frac{1}{2} \underline{\beta' v \beta} - \frac{1}{2} c]$$

$$K_q(t') = ln\Psi(t_1....t_k) = \frac{1}{2}ln|v| - \frac{1}{2}ln|\Sigma| + \frac{1}{2}\underline{\beta}' v\underline{\beta} - \frac{1}{2}\underline{\mu}'\Sigma^{-1}\underline{\mu} + i\,\underset{j=1}{\overset{k}{Sum}} t_j \underline{\mu}_j' Q_j \underline{\mu}_j$$

$$where\ \ v_t = [\Sigma^{-1} - 2i\underset{j=1}{\overset{k}{Sum}} t_j Q_j]^{-1},\ \ \underline{\beta_t} = [\Sigma^{-1}\underline{\mu} - 2i\underset{j=1}{\overset{k}{Sum}} t_j Q_j \underline{\mu_j}]$$

$$v_0^{-1} = \Sigma^{-1} \quad v_0 = \Sigma \ \ at\ \ t = 0\ \ and\ \underline{\beta_0} = \Sigma^{-1}\underline{\mu}$$

$$K(0') = \frac{1}{2}ln|\Sigma| - \frac{1}{2}ln|\Sigma| + \frac{1}{2}\underline{\mu}'\Sigma^{-1}\underline{\mu} - \frac{1}{2}\underline{\mu}'\Sigma^{-1}\underline{\mu} = 0$$

$$\frac{\partial v}{\partial t_j} = -[\Sigma^{-1} - 2i\underset{j=1}{\overset{k}{Sum}} t_j Q_j]^{-2}(-2iQ_j) = 2iQ_j v^2 = 2ivQ_j v$$

$$\frac{\partial \underline{\beta}}{\partial t_j} = -2iQ_j\underline{\mu_j}; \qquad \frac{\partial v^{-1}}{\partial t_j} = -2iQ_j \quad ;$$

$$\frac{\partial K}{\partial t_j} = \frac{1}{2}tr[v^{-1}(2ivQ_j v)] + \frac{1}{2}(-2i\underline{\mu_j'}Q_j v\underline{\beta}) + \frac{1}{2}\underline{\beta}'(2ivQ_j v)\underline{\beta}$$

$$+\frac{1}{2}\underline{\beta}'v(-2iQ_j\mu_j) + i\underline{\mu_j'}Q_j\underline{\mu_j}$$

$$= itrQ_j v - i\underline{\mu_j'}Q_j v\underline{\beta} + i\underline{\beta}'vQ_j v\underline{\beta} - i\underline{\beta}'vQ_j\underline{\mu_j} + i\underline{\mu_j'}Q_j\underline{\mu_j}$$

$$= itr(Q_j v) + i(\underline{v\beta - \mu_j})'Q_j(\underline{v\beta - \mu_j})$$

$$Assume\ \ \underline{\gamma_j} = -v\underline{\beta} + \underline{\mu_j}\ \ then\ \ \underline{\gamma_0} = -\Sigma\Sigma^{-1}\underline{\mu} + \underline{\mu_j} = \underline{\mu_j - \mu}$$

$$\frac{\partial K}{\partial t_j} = itrQ_j v + i\underline{\gamma_j'}Q_j\underline{\gamma_j}\ \ at\ t = 0\ we\ have$$

$$\frac{\partial K}{\partial t_j}\Big|_{t=0} = itrQ_j\Sigma + i\underline{\gamma_0'}Q_j\underline{\gamma_0} = i[trQ_j\Sigma + (\underline{\mu_j - \mu})'Q_j(\underline{\mu_j - \mu})]$$

$$i.e\ \ first\ cumulant\ of\ \ q_j$$

$$\frac{\partial K}{\partial t_j}\Big|_{t=0} = E(q_j) = trQ_j\Sigma + (\underline{\mu_j - \mu})'Q_j(\underline{\mu_j - \mu})$$

$$\frac{\partial \gamma_j}{\partial t_k} = -\frac{\partial v}{\partial t_k}\underline{\beta} - v\frac{\partial \beta}{\partial t_k} = -2ivQ_k v\underline{\beta} + 2ivQ_k\underline{\mu_k} = 2ivQ_k\underline{\gamma_k}$$

$$\frac{\partial \gamma_j}{\partial t_k} = 2ivQ_k\underline{\gamma_k}$$

$$\frac{\partial^2 K}{\partial t_j \partial t_k} = itrQ_j(2ivQ_k v) + i(2i\underline{\gamma_k'}Q_k vQ_j\underline{\gamma_j}) + i\underline{\gamma_j'}Q_j(2ivQ_k\underline{\gamma_k})$$

$$= -2trQ_j vQ_k v - 2\underline{\gamma_k'}Q_k vQ_j\underline{\gamma_j} - 2\underline{\gamma_j'}Q_j vQ_k\underline{\gamma_k}$$

$$Cov(q_j, q_k) = (coefficient\ of\ -i^2\ in\ \frac{\partial^2 K}{\partial t_j \partial t_k}\ at\ t = 0)$$

$$= 2trQ_j\Sigma Q_k\Sigma + 4(\underline{\mu_j - \mu})'Q_j\Sigma Q_k(\underline{\mu_k - \mu})$$

$$\frac{\partial^2 K}{\partial t_j \partial t_k} = -2trQ_j vQ_k v - 4\underline{\gamma_j'}Q_j vQ_k \underline{\gamma_k}$$

$$\frac{\partial^3 K}{\partial t_j \partial t_k \partial t_l} = -4i\,trQ_j vQ_l vQ_k v - 4i\,trQ_j vQ_k vQ_l v - 8i\,\underline{\gamma_j'}vQ_k v\underline{\gamma_l}$$

$$-8i\,\underline{\gamma_j'}vQ_l v\underline{\gamma_k} - 8i\,\underline{\gamma_k'}vQ_j v\underline{\gamma_l}$$

$$Suppose\ k = 1 \quad q = (\underline{x - \mu_1})' Q_1(\underline{x - \mu_1}) \ where\ Q_1 = \Sigma^{-1} = (var(x))^{-1}$$

$$= (\underline{x - \mu_1})' \Sigma^{-1}(\underline{x - \mu_1})$$

$$Let\ \underline{x} \sim N(\underline{\mu_1}, \Sigma)\ also\ suppose\ \underline{x} \sim N(\underline{\mu_2}, \Sigma)$$

$$q\ has\ noncentral\ \chi^2_{(p)}\ distribution\ with\ noncentral\ parameter$$

$$(\underline{\mu_1 - \mu_2})' \Sigma^{-1}(\underline{\mu_1 - \mu_2}) = \delta^2_{\pi_1\pi_2}$$

$$E(q) = tr\Sigma Q + (\underline{\mu_1 - \mu})' Q(\underline{\mu_1 - \mu})\ in\ this\ case\ \mu = \mu_2\ and\ Q = \Sigma^{-1}$$

$$E(q) = p + (\underline{\mu_1 - \mu_2})' \Sigma^{-1}(\underline{\mu_1 - \mu_2}) = p + \delta^2$$

$$Var(q) = 2tr(Q\Sigma)^2 + 4(\underline{\mu_1 - \mu_2})' Q\Sigma Q(\underline{\mu_1 - \mu_2}) = 2p + 4\delta^2$$

$$K_3(q) = 8p + 24\delta^2; \qquad K_4(q) = 48p + 192\delta^2;$$

$$in\ general;\ \ K_r(q) = 2^{r-1}(r-1)!\,[\,p + r\delta^2\,].$$

The above result due to Lancaster.

APPLICATION IN CLASSIFICATION

Assume column vector $\underline{X}$ comes from population $,\pi_1,$ or from $,\pi_2,$ then If $\underline{X}$ from population 1 then $ln\,L(\underline{X}/\pi_1)$ would be the likelihood . If $\underline{X}$ from population 2 then $ln\,L(\underline{X}/\pi_2)$ would be the likelihood . Our objective is to classify $\underline{X}$ as belonging to which population for which $ln\,L(\underline{X})$ is larger.

$$R_1 = \{x : lnL(x|\pi_1) > ln\,L(x|\pi_2)\}\ then\ we\ make\ decision\ D_1\ and\ assign\ to\ \pi_1.$$

$$R_2 = \{x : lnL(x|\pi_1) < ln\,L(x|\pi_2)\}\ then\ we\ make\ decision\ D_2\ and\ assign\ to\ \pi_2.$$

$$\pi_1 \sim N(\mu_1, \Sigma); \qquad \pi_2 \sim N(\mu_2, \Sigma);$$

$$lnL(\underline{x}|\pi_1) = -\frac{p}{2}ln(2\pi) - \frac{1}{2}ln|\Sigma| - \frac{1}{2}(\underline{x - \mu_1})' \Sigma^{-1}(\underline{x - \mu_1})$$

$$lnL(\underline{x}|\pi_2) = -\frac{p}{2}ln(2\pi) - \frac{1}{2}ln|\Sigma| - \frac{1}{2}(\underline{x - \mu_2})' \Sigma^{-1}(\underline{x - \mu_2})$$

$$u = lnL(x|\pi_1) - ln\,L(x|\pi_2)$$

$$= -\frac{1}{2}(\underline{x - \mu_1})' \Sigma^{-1}(\underline{x - \mu_1}) + \frac{1}{2}(\underline{x - \mu_2})' \Sigma^{-1}(\underline{x - \mu_2})$$

$$where\ classification\ statistics\ u\ make\ decision\ D_1\ if\ u > 0$$

$$u = -\frac{1}{2}\underline{x'}\Sigma^{-1}\underline{x} + \underline{\mu_1'}\Sigma^{-1}\underline{x} - \frac{1}{2}\underline{\mu_1'}\Sigma^{-1}\underline{\mu_1} + \frac{1}{2}\underline{x'}\Sigma^{-1}\underline{x} - \underline{\mu_2'}\Sigma^{-1}\underline{x} + \frac{1}{2}\underline{\mu_2'}\Sigma^{-1}\underline{\mu_2}$$

$$=(\underline{\mu_1'-\mu_2'})\Sigma^{-1}\underline{x}-\frac{1}{2}\underline{\mu_1'}\Sigma^{-1}\underline{\mu_1}+\frac{1}{2}\underline{\mu_2'}\Sigma^{-1}\underline{\mu_2}$$

$$=(\underline{\mu_1'-\mu_2'})\Sigma^{-1}(\underline{x}-\frac{1}{2}(\underline{\mu_1+\mu_2}))$$

$Suppose\ \underline{x}\sim N(\underline{\mu},\Sigma)$ $Let\ \underline{y}=(\underline{x}-\frac{1}{2}(\underline{\mu_1+\mu_2}))\sim N(\underline{\mu}-\frac{1}{2}(\underline{\mu_1+\mu_2}),\Sigma)$

$Then\ \ u=(\underline{\mu_1'-\mu_2'})\Sigma^{-1}\underline{y}$

$Hence: E(u)=(\underline{\mu_1-\mu_2})'\,\Sigma^{-1}(\underline{\mu}-\frac{1}{2}(\underline{\mu_1+\mu_2}))$

$var(u)=(\underline{\mu_1-\mu_2})'\,\Sigma^{-1}\Sigma\Sigma^{-1}(\underline{\mu_1-\mu_2});$

$u\sim N((\underline{\mu_1-\mu_2})'\Sigma^{-1}(\underline{\mu-\bar{\mu}}),\ \delta^2)$

$where\ \ \underline{\bar{\mu}}=\frac{1}{2}(\underline{\mu_1+\mu_2});\quad \delta^2=(\underline{\mu_1-\mu_2})'\Sigma^{-1}(\underline{\mu_1-\mu_2})$

$Pr\{classify\ \underline{x}\ as\ coming\ from\ \pi_1\ if, in\ fact, it\ does\}$

$=Pr\{u>0\,|\,\underline{x}\in\pi_1\}=Pr\{u>0\,|\,\underline{\mu=\mu_1}\}$

$$=1-\phi(\frac{0-\frac{1}{2}\delta^2}{\delta})=1-\phi(-\frac{1}{2}\delta)=\phi(\frac{1}{2}\delta)$$

$Since\ if\ \underline{\mu=\mu_1}\ \ then\ u\sim N(\frac{1}{2}\delta^2,\delta^2)$ $If\ the\ two\ population\ are\ 1\ standard\ unit$

$away\ then\ Pr\{correct\ classify\ into\ \pi_1\,|\,\pi_1 is\ the\ true\ population)=\phi(\frac{1}{2})\approx 0.65$

$Pr\{u<0\,|\,\underline{\mu=\mu_1}\}=1-Pr\{u>0\,|\,\underline{\mu=\mu_1}\}=1-\phi(\frac{\delta}{2})$

$if\ E(\underline{x})=\underline{\mu_2}\ \ then\ u\sim N(-\frac{\delta^2}{2},\delta^2)$ $\ Pr\{u<0\,|\,\mu=\mu_2\}=\phi(\frac{0+\frac{\delta^2}{2}}{\delta})=\phi(\frac{\delta}{2})$

$Pr\{u>0\,|\,\mu=\mu_2\}=1-Pr\{u<0\,|\,\mu=\mu_2\}=1-\phi(\frac{\delta}{2})$

$we\ summary\ above\ result$

	π_1	π_2
D_1	$\phi(\frac{\delta}{2})$	$1-\phi(\frac{\delta}{2})$
D_2	$1-\phi(\frac{\delta}{2})$	$\phi(\frac{\delta}{2})$

Let q_1 denote the prior probability that $\underline{X}$ comes from π_1,

$q_2=1-q_1$ that $\underline{X}$ come from π_2. Our decision rule is optimal only if $q_1=q_2=\frac{1}{2}$ then

$Pr\{misclassification\}=1-\phi(\frac{\delta}{2})$

$q_1 = Quota\ for\ population\ \pi_1.\ 1 - q_1 = Quota\ for\ population\ \pi_2.$

$R_1 = \{ \underline{x} : q_1 L(\underline{x} \mid \pi_1) > q_2 L(\underline{x} \mid \pi_2) \}$

If $\underline{x}$ is observed in R_1 and make decision D_1 : classify $\underline{x}$ as belonging to π_1.

$$pr\{ \underline{x} : \log q_1 + \log L(\underline{x} \mid \pi_1) - \log q_2 - \log L(\underline{x} \mid \pi_2) > 0 \}$$

$$= Pr\{ \log L(\underline{x} \mid \pi_1) - \log L(\underline{x} \mid \pi_2) > \log(\frac{1-q_1}{q_1}) \}$$

Assume $\pi_1 \sim N(\underline{\mu_1}, \Sigma)$; $\quad \pi_2 \sim N(\underline{\mu_2}, \Sigma)$

$$if\ \ E(\underline{x}) = \underline{\mu_1}\ \ then\ u \sim N(\frac{\delta^2}{2}, \delta^2);\ \ if\ \ E(\underline{x}) = \underline{\mu_2}\ \ then\ u \sim N(-\frac{\delta^2}{2}, \delta^2)$$

$$where\ u = \underline{(\mu_1 - \mu_2)}' \Sigma^{-1} (\underline{x} - \frac{1}{2} \underline{(\mu_1 + \mu_2)});$$

$$\delta^2 = \underline{(\mu_1 - \mu_2)}' \Sigma^{-1} \underline{(\mu_1 - \mu_2)}\ \ and\ \log(\frac{1-q_1}{q_1}) = c$$

$if\ q_1 < q_2\ \ then\ c > 0$

$$Pr\{ x \in R_1 \mid \pi_1 \} = Pr\{ u > c \mid \pi_1 \} = 1 - Pr\{ u < c \mid \pi_1 \} = 1 - \phi(\frac{c - \frac{\delta^2}{2}}{\delta}) = \phi(\frac{\delta}{2} - \frac{c}{\delta})$$

$$Pr\{ x \in R_2 \mid \pi_2 \} = Pr\{ u < c \mid \pi_2 \} = \phi(\frac{c + \frac{\delta^2}{2}}{\delta}) = \phi(\frac{\delta}{2} + \frac{c}{\delta})$$

$$Pr\{ Misclassification \} = \phi(-\frac{\delta}{2} - \frac{c}{\delta}) + \phi(\frac{c}{\delta} - \frac{\delta}{2})$$

$$choose\ c\ such\ that\ \max\{ \phi(-\frac{\delta}{2} - \frac{c}{\delta}), \phi(\frac{c}{\delta} - \frac{\delta}{2}) \}\ is\ minimized$$

Clearly $c = 0$ then $q_1 = q_2$ mini-max decision.

State of nature as follow:

	π_1	π_2
D_1	$\phi(\frac{\delta}{2} - \frac{c}{\delta})$	$\phi(-\frac{\delta}{2} - \frac{c}{\delta})$
D_2	$\phi(\frac{c}{\delta} - \frac{\delta}{2})$	$\phi(\frac{\delta}{2} + \frac{c}{\delta})$

CONCLUDING REMARKS

At the end, we give two real life examples to demonstrate how to apply our results to solve a real-life problem. Example 1. Rao C.R. [3] considered three populations, the Brahmin, Artison, and Korwa castes of India. He assumed that each of these populations could be characterized by four characters – stature (x_1), sitting height (x_2), nasal depth (x_3), and nasal height (x_4)- of each member of the population. On the basis of sample observations on these characters from these three populations, the problem is to classify an individual with observation $\underline{x}' = (x_1, x_2, x_3, x_4)$' into one of the three populations. Rao C.R. used a linear discriminator to obtain this solution.

Example 2. On a patient with a diagnosis of myocardial infarction, observation on his systolic blood pressure (x_1), diastolic blood pressure (x_2), heart rate (x_3), stroke index (x_4), and mean arterial pressure (x5) are taken. On the basis of these observations, it is possible to predict whether or not the patient will survive.

REFERENCES

[1] Kendall, M. and Stuart, A. (1978) The Advanced Theory of Statistics. Vol. 2. 4th edition. Macmillan publishing Company, Inc.
[2] Anderson, T.W.(1984) An Introduction to Multivariate Statistical Analysis. Second Edition. John Wiley & Sons, Inc.
[3] Rao, C.R.(1948) The utilization of multiple measurements in problem of biological classification. J.R. Statistical Society B10 : 159-203.