ON CONFIDENT ESTIMATES OF DATASET SIZES

PATRICK BARDSLEY

Abstract. We provide estimates for the number of observations needed from underlying datasets datasets to confidently measure various criteria. One approach uses the asymptotic normality of sample means (i.e., the central limit theorem), which allows us to leverage the χ^2 distribution and obtain an estimate for the number of i.i.d. samples needed to estimate a true population mean within a desired precision ϵ. Our second determines the number of observations needed to estimate an $\epsilon-$level difference between two distinct datasets at a prescribed confidence level. This estimate relies on Hotelling's two-sample T^2 test, and gives the observation number in terms of precision ϵ, size α, and power $1-\beta$ of the T^2 test. We also provide some preliminary dataset analysis procedures and examples.

Contents

1. Notational Conventions and Normality Assumptions

We denote a single observation of an underlying vector-valued feature distribution as $\mathbf{x}_i \in \mathbb{R}^M$, where M is the ambient dimension and $i = 1, \dots, N$. The sample mean and sample covariance of $\{\mathbf{x}_i\}$ shall be denoted as

$$\overline{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_i,$$

$$\widehat{\mathbf{\Sigma}}_x = \frac{1}{N-1}\sum_{i=1}^{N} (\mathbf{x}_i - \overline{\mathbf{x}})^2,$$

respectively. Discussions of distinct datasets should be clear in context, but we will often indicate different datasets with different letters (e.g., $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^{N_x}$, $\mathbf{Y} = \{\mathbf{y}_i\}_{i=1}^{N_y}$). Moreover, we will often omit the indices in the $\{\cdot\}$ set notation for brevity unless required for clarity.

Though all of our sample size estimates ultimately rely on the asymptotic normality of maximum likelihood estimators, which holds by the central limit theorem

regardless the distribution of data $\{\mathbf{x}_i\}$ (see, e.g., [1, 4]), standard practice is to first perform preliminary assessments of any available data. These assessments range from simple familiarization with data shapes, trends, outliers, etc. (e.g., simple scatter plots of data), to validating modeling assumptions of the data distributions. In particular, one may be interested in determining the degree of normality that a given set of data exhibits, if any.

For multivariate data, a simple procedure to assess normality assumptions is a multidimensional QQPlot. In this procedure, we compute the Mahalanobis-deviations

$$d_i = (\mathbf{x}_i - \overline{\mathbf{x}})\widehat{\mathbf{\Sigma}}_x^{-1}(\mathbf{x}_i - \overline{\mathbf{x}}) \quad \text{for } i = 1, \ldots, N,$$

and plot the quantiles of these points against their respective χ^2 quantiles. That is, from $\{d_i\}$ we form the ordered statistics $\{d_{(i)}\}$ (i.e., order the deviation statistics in ascending order), and generate a scatter plot consisting of the points

$$\{(\Phi_{\chi^2}^{-1}(i/N), d_{(i)}) : i = 1, \ldots, N\},$$

where $\Phi_{\chi^2}^{-1}(p)$ denotes the p-th quantile of the $\chi^2(M)$ distribution. If the data $\{\mathbf{x}_i\}$ is in fact multivariate normal, the Mahalanobis-deviation statistics $\{d_i\}$ will be approximately distributed as χ^2 random variables, thus the scatter plot should trace out a straight line. Conversely, deviations from a straight line then indicate deviations (and the degree of deviation) from a multivariate normal distribution.

In Figure 1 we show simple scatter plots for the datasets $\mathbf{X} = \{\mathbf{x}_i\}$ and $\mathbf{Y} = \{\mathbf{y}_i\}$ where

$$\begin{aligned} \mathbf{x}_i &\sim \mathcal{N}\left(\sum_i 0.5\mathbf{e}_i, \mathcal{I}\right) \quad \textit{i.i.d.} \text{ for } i = 1, \ldots, N, \\ \mathbf{y}_i &\sim \mathcal{U}\left((0,1)^M\right), \quad \textit{i.i.d.} \text{ for } i = 1, \ldots, N, \end{aligned} \tag{1}$$

where $\{\mathbf{e}_i\}$ denote the canonical basis vectors of $\mathbb{R}^M$ (i.e., 1 in the ith component and 0 elsewhere). Here we have used $\mathcal{N}(\boldsymbol{\mu}, \mathbf{\Sigma})$ to indicate a multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma}$, and similarly $\mathcal{U}((a,b)^M)$ to indicate a multivariate uniform distribution whose samples live in the hypercube defined by the Cartesian product $(a,b)^M$.

Next, in Figure 2 we show the multivariate QQPlot for each of the datasets in (1). Notice the plot for the normal data is linear, whereas the plot for the uniform data is not, which is expected and indicative of the respective distributions.

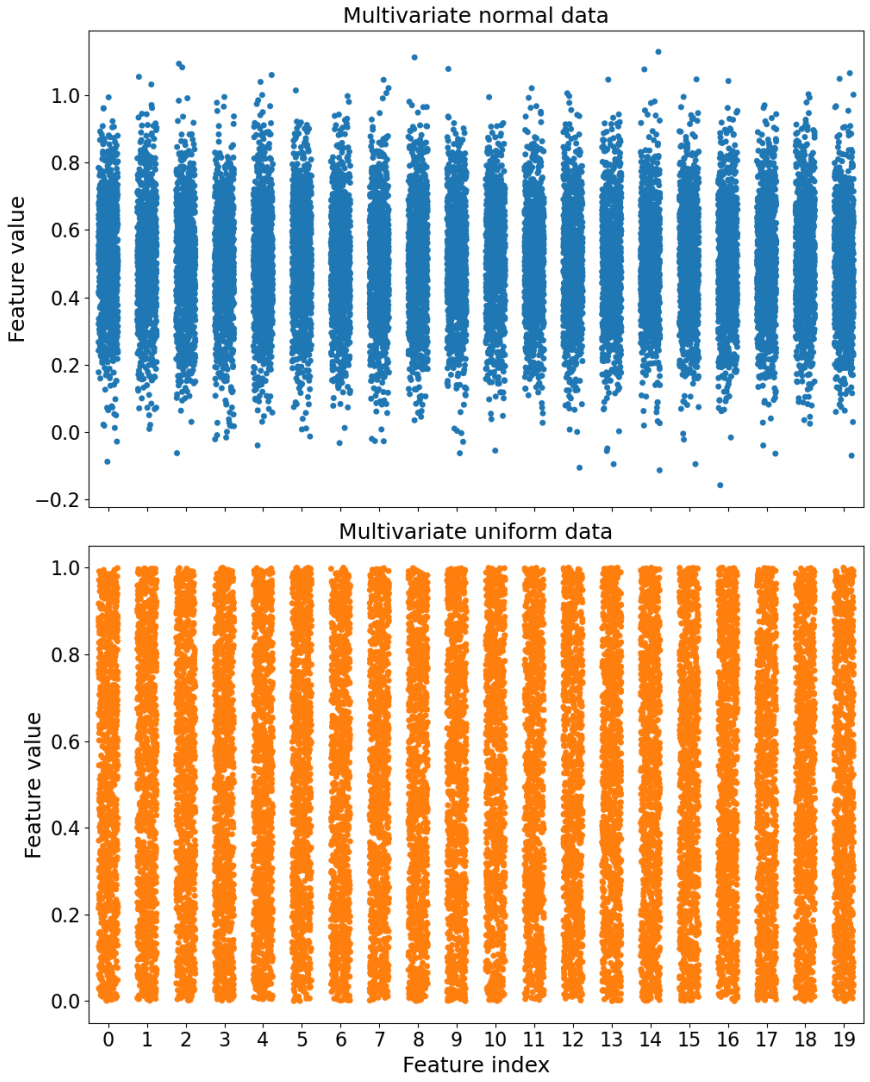

FIGURE 1. Scatter/strip plots of (top) multivariate normal data and (bottom) multivariate uniform data for dimension $M = 20$. Each dimension $m = 1, \ldots, M$, is essentially plotted as an independent scatter plot along a single y-dimension, with some spread to more clearly depict the various values each feature dimension assumes.

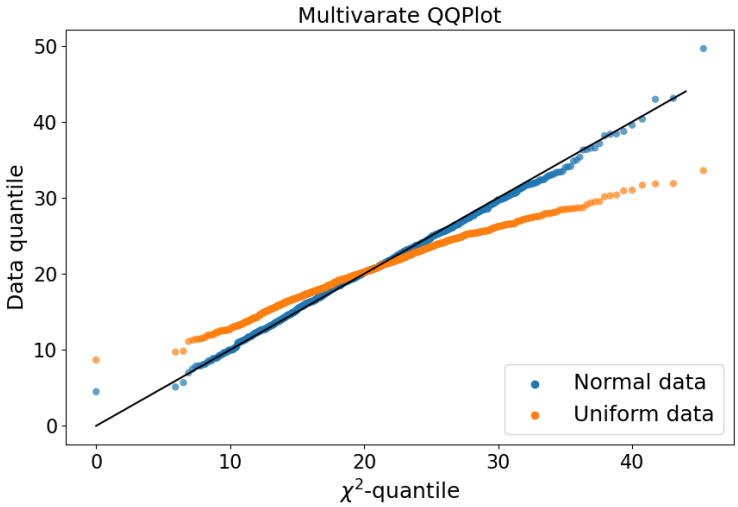

FIGURE 2. Multivariate QQPlot shown for (blue) multivariate normal data and (orange) multivariate uniform data. Straight lines (resp. deviations from straight lines) indicate the feature distribution is well-modeled (resp. not well-modeled) by a multivariate normal distribution.

2. Asymptotic Estimates

In this section we provide our sample size estimates needed to ensure various precision measurements can be obtained with specified confidence. In §2.1 we give sample size estimates required to estimate a dataset's mean within a desired precision ϵ and $100(1-\alpha)\%$ confidence. Then, in §2.2 we give estimates required to measure ϵ-sized differences between two datasets' respective means with $100(1-\alpha)\%$ confidence.

2.1. **Observations required for confident mean estimates.**

For any collection of i.i.d. variables $\{\mathbf{x}_i\}_{i=1}^N$ drawn from *any* distribution with mean and covariance given by $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, respectively, we are guaranteed by the central limit theorem (see e.g., [1, 4]) that

$$\sqrt{N}\boldsymbol{\Sigma}^{-1/2}(\overline{\mathbf{x}}-\boldsymbol{\mu}) \xrightarrow{d} \mathcal{N}(\mathbf{0},\mathbf{I}). \tag{2}$$

In words, in the large sample limit $N\to\infty$, the sample mean is distributed as a multivariate *standard* normal random variable. It follows from (2) and the definition of a χ^2 variable with M degrees of freedom that

$$N(\overline{\mathbf{x}}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{x}}-\boldsymbol{\mu}) \sim \chi^2(M), \tag{3}$$

in the large sample limit $N\to\infty$.

An asymptotic $100(1-\alpha)\%$ confidence interval for the quantity $(\overline{\mathbf{x}}-\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{x}}-\boldsymbol{\mu})$ can be found by requiring

$$\begin{aligned} 1-\alpha &\approx P(N(\overline{\mathbf{x}}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{x}}-\boldsymbol{\mu}) \le \Phi^{-1}_{\chi^2}(1-\alpha)) \\ &= P((\overline{\mathbf{x}}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{x}}-\boldsymbol{\mu}) \le N^{-1}\Phi^{-1}_{\chi^2}(1-\alpha)), \end{aligned}$$

where $\Phi^{-1}_{\chi^2}(1-\alpha)$ denotes the $(1-\alpha)$-quantile of the $\chi^2(M)$ distribution. Furthermore, the approximation in the first line becomes equality in the large sample limit $N\to\infty$. Requiring our estimate of the true population mean $\boldsymbol{\mu}$ to be accurate within ϵ, in terms of the Mahalanobis distance, we require

$$(\overline{\mathbf{x}}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{x}}-\boldsymbol{\mu}) \le \epsilon^2.$$

Thus, to obtain a $100(1-\alpha)\%$ confident estimate of $\boldsymbol{\mu}$ with specified precision ϵ, we require

$$N^{-1}\Phi^{-1}_{\chi^2}(1-\alpha) < \epsilon^2.$$

This final inequality can be solved for N to determine the number of i.i.d. observations required to measure a dataset's (e.g., language-specific dataset) feature vector mean with precision ϵ and confidence $100(1-\alpha)\%$:

$$N(\epsilon,\alpha) \ge \frac{\Phi^{-1}_{\chi^2}(1-\alpha)}{\epsilon^2}. \tag{4}$$

In Table 1, we give minimal values required for various precision and confidence levels, respectively ϵ and $(1-\alpha)$, and for feature dimension $M=20$.

ϵ	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
1.00	29	32	38
0.50	114	126	151
0.25	455	503	602
0.10	2842	3142	3757

TABLE 1. Number of i.i.d. M-variate observations required to achieve a $100(1-\alpha)\%$ estimate of a dataset's population mean $\boldsymbol{\mu}$ to within a precision level ϵ, where precision is measured in terms of the Mahalanobis distance: $(\overline{\mathbf{x}} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\overline{\mathbf{x}} - \boldsymbol{\mu}) < \epsilon^2$. Here the feature dimension $M = 20$.

2.2. **Observations required for confident estimates of differences in dataset means.** We now derive an estimate on the number of observations for two datasets, that provides confident detection of differences between the datasets' means. Denote two independent datasets as $\mathbf{X} = \{\mathbf{x}_i\}$ and $\mathbf{Y} = \{\mathbf{y}_i\}$, where

$$\begin{aligned} \mathbf{x}_i &\sim N(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}) \quad \textit{i.i.d.} \text{ for } i = 1 \ldots, N_x, \\ \mathbf{y}_i &\sim N(\boldsymbol{\mu}_y, \boldsymbol{\Sigma}) \quad \textit{i.i.d.} \text{ for } i = 1 \ldots, N_y. \end{aligned}$$

We consider Hotelling's two-sample T^2 test ([2, 3, 5, 6]) for testing the equality of $\boldsymbol{\mu}_x$ and $\boldsymbol{\mu}_y$, whose null-hypothesis, alternative hypothesis, and test statistic are given by

$$\begin{aligned} H_0 &: \boldsymbol{\mu}_x = \boldsymbol{\mu}_y, \\ H_A &: \boldsymbol{\mu}_x \neq \boldsymbol{\mu}_y, \\ T^2 &= \frac{N_x N_y}{N_x + N_y} (\overline{\mathbf{x}} - \overline{\mathbf{y}})^T \widehat{\boldsymbol{\Sigma}}^{-1} (\overline{\mathbf{x}} - \overline{\mathbf{y}}), \end{aligned}$$

where

$$\begin{aligned} \overline{\mathbf{x}} &= N_x^{-1} \sum_{i=1}^{N_x} \mathbf{x}_i, \\ \overline{\mathbf{y}} &= N_y^{-1} \sum_{i=1}^{N_y} \mathbf{y}_i, \\ \widehat{\boldsymbol{\Sigma}} &= (N_x + N_y - 2)^{-1} \left(\sum_{i=1}^{N_x} (\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^T + \sum_{i=1}^{N_y} (\mathbf{y}_i - \overline{\mathbf{y}})(\mathbf{y}_i - \overline{\mathbf{y}})^T \right). \end{aligned}$$

In words, this test rejects the null hypothesis of equal means in favor of the alternative hypothesis that the means differ, if the test statistic value T^2 is overly large. It can be shown (see, e.g., [5] or Appendix A), under the null hypothesis H_0, T^2 is distributed according to a standard (i.e., *central*) $F-$distribution:

$$\frac{N_x + N_y - M - 1}{(N_x + N_y - 2)M} T^2 \sim F(M, N_x + N_y - 1 - M, 0).$$

Conversely, under the alternative hypothesis H_A, T^2 is distributed according to a *noncentral* $F-$distribution (see Appendix A):

$$\frac{N_x+N_y-M-1}{(N_x+N_y-2)M}T^2 \sim F(M, N_x+N_y-1-M, \lambda)$$

with noncentrality parameter λ given by the *population* Mahalanobis distance

$$\lambda = \frac{N_x N_y}{N_x+N_y}(\boldsymbol{\mu}_x-\boldsymbol{\mu}_y)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_x-\boldsymbol{\mu}_y).$$

We leverage this hypothesis test to determine a sample size required to guarantee rejection of H_0 when H_A is indeed true (i.e., correct detection of dataset mean differences) with probability $1-\beta$, while keeping the probability of false rejection of H_0, namely α, low. In statistical terminology, we select our sample sizes for a two-sample T^2 test of size α to guarantee a specified detection power of $1-\beta$. We first compute our $\alpha-$level critical value via

$$\begin{aligned} 1-\alpha &= P\left(\frac{N_x+N_y-M-1}{(N_x+N_y-2)M}T^2 \leq C_*|H_0\right) \\ &= \Phi_F(C_*; M, N_x+N_y-1-M, 0), \end{aligned}$$

where $\Phi_F(\cdot; M, N_x+N_y-1-M, \lambda)$ denotes the cumulative distribution function for the (possibly noncentral) $F(M, N_x+N_y-1-M, \lambda)$ distribution. Thus,

$$C_* = \Phi_F^{-1}(1-\alpha; M, N_x+N_y-1-M, 0),$$

where Φ_F^{-1} denotes the associated quantile function. We then require power

$$\begin{aligned} 1-\beta &\leq P\left(\frac{N_x+N_y-M-1}{(N_x+N_y-2)M}T^2 > C_*|H_A\right) \\ &= 1-\Phi_F(C_*; M, N_x+N_y-1-M, \lambda). \end{aligned}$$

Therefore, given a specified value of α, β, and detection precision $\epsilon = (\boldsymbol{\mu}_x-\boldsymbol{\mu}_y)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_x-\boldsymbol{\mu}_y)$, we can select N_x and N_y such that

$$0 \leq \beta - \Phi(C_*(N_x, N_y); M, N_x+N_y-1-M, \lambda(N_x, N_y, \epsilon)).$$

For simplicity in estimates, we will require $N \equiv N_x \equiv N_y$.

We depict this procedure in Figure 3. Here we show the power of the T^2 test (i.e., the probability of true rejection of H_0) for fixed values of $\alpha = 0.05$ and $\epsilon = 0.1$, and varying values of the sample size N. The dashed lines here indicate different values of $1-\beta$. In Table 2, we give the estimates of the required sample size $N \equiv N_x \equiv N_y$ for various values of α, β and ϵ. As an example, from Figure 3 and/or Table 2, we see that we require $N \geq 5235$ for 90% detection (i.e., $\beta = 0.1$) of an $\epsilon = 0.1$ difference between the means with 95% confidence (i.e., $\alpha = 0.05$).

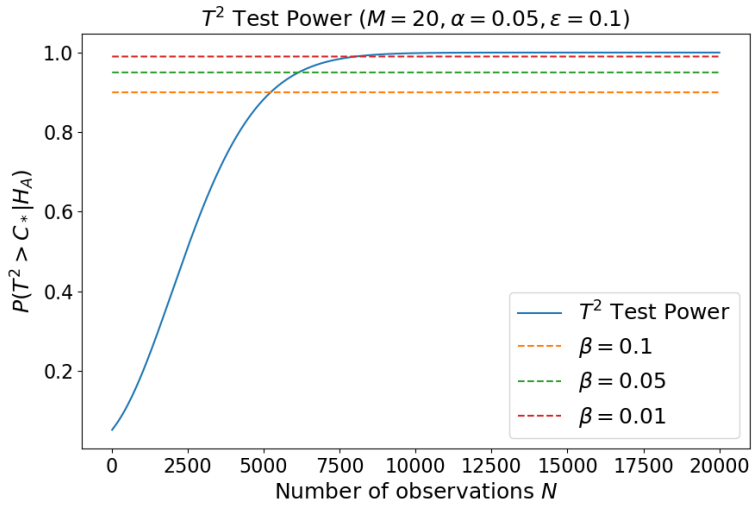

FIGURE 3. Power of Hotelling's T^2 test for $M = 20, \alpha = 0.05$ and $\epsilon = 0.1$ for various values of $N \equiv N_x = N_y$. The dashed lines indicate different values of $1-\beta$ which can be used to determine the minimum number of observations N required to achieve a desired power-level for the given α and ϵ.

$\epsilon = 1.00$			
α	$\beta = 0.1$	$\beta = 0.05$	$\beta = 0.01$
0.10	53	61	79
0.05	61	70	89
0.01	78	88	108

$\epsilon = 0.50$			
α	$\beta = 0.1$	$\beta = 0.05$	$\beta = 0.01$
0.10	186	221	293
0.05	218	254	329
0.01	281	321	402

$\epsilon = 0.25$			
α	$\beta = 0.1$	$\beta = 0.05$	$\beta = 0.01$
0.10	721	861	1148
0.05	845	991	1292
0.01	1093	1254	1579

$\epsilon = 0.10$			
α	$\beta = 0.1$	$\beta = 0.05$	$\beta = 0.01$
0.10	4467	5340	7132
0.05	5235	6153	8028
0.01	6780	7783	9815

TABLE 2. The number $N \equiv N_x = N_y$ of i.i.d. M-variate observations required for a two-sample T^2 test to detect an $\epsilon-$level difference between datasets with a true (resp. false) detection rate of $1-\beta$ (resp. α). Here the feature dimension $M = 20$ and ϵ is measured in terms of the Mahalanobis distance: $(\overline{\mathbf{x}}-\overline{\mathbf{y}})^T\mathbf{\Sigma}^{-1}(\overline{\mathbf{x}}-\overline{\mathbf{y}}) \geq \epsilon^2$.

Appendix A. Distribution of Hotelling's two-sample T^2 statistic

Here we derive the distributional form of the two-sample T^2 test statistic:

$$T^2 = (\overline{\mathbf{x}} - \overline{\mathbf{y}})^T \widehat{\mathbf{\Sigma}}^{-1} (\overline{\mathbf{x}} - \overline{\mathbf{y}}), \tag{5}$$

where

$$\begin{aligned} \mathbf{x}_i &\sim \mathcal{N}(\boldsymbol{\mu}_x, \mathbf{\Sigma}) \quad i.i.d. \quad \text{for } i = 1, \ldots, N_x \\ \mathbf{y}_i &\sim \mathcal{N}(\boldsymbol{\mu}_y, \mathbf{\Sigma}) \quad i.i.d. \quad \text{for } i = 1, \ldots, N_y. \end{aligned}$$

In (5) we utilize the unbiased parameter estimates

$$\begin{aligned} \overline{\mathbf{x}} &= \frac{1}{N_x} \sum_{i=1}^{N_x} \mathbf{x}_i, \quad \widehat{\mathbf{\Sigma}}_x = \frac{1}{N_x - 1} \sum_{i=1}^{N_x} (\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^T \\ \overline{\mathbf{y}} &= \frac{1}{N_y} \sum_{i=1}^{N_y} \mathbf{y}_i, \quad \widehat{\mathbf{\Sigma}}_y = \frac{1}{N_y - 1} \sum_{i=1}^{N_y} (\mathbf{y}_i - \overline{\mathbf{y}})(\mathbf{y}_i - \overline{\mathbf{y}})^T, \end{aligned}$$

and the pooled covariance matrix (which is also an unbiased estimate of $\mathbf{\Sigma}$) as

$$\widehat{\mathbf{\Sigma}} = \frac{1}{N-2}((N_x - 1)\widehat{\mathbf{\Sigma}}_x + (N_y - 1)\widehat{\mathbf{\Sigma}}_y) \quad \text{with } N = N_x + N_y.$$

Theorem 1 (Hotelling's T^2 statistic $\iff$ Noncentral F-distribution)**.** *Hotelling's two-sample T^2 test statistic* (5) *is a noncentral F-distributed random variable, i.e.,*

$$\left(\frac{N - M - 1}{M} \right) \frac{N_x N_y}{N(N-2)} T^2 \sim F(M, N - M - 1, \lambda),$$

with noncentrality parameter

$$\lambda = N_x N_y N^{-1} (\boldsymbol{\mu}_x - \boldsymbol{\mu}_y)^T \mathbf{\Sigma}^{-1} (\boldsymbol{\mu}_x - \boldsymbol{\mu}_y).$$

Proof. We begin by rewriting the T^2 statistic as

$$\begin{aligned} T^2 &= \mathbf{d}^T \widehat{\mathbf{\Sigma}}^{-1} \mathbf{d} \\ &= \mathbf{d}^T \mathbf{\Sigma}^{-1} \mathbf{d} \left(\frac{\mathbf{d}^T \mathbf{\Sigma}^{-1} \mathbf{d}}{\mathbf{d}^T \widehat{\mathbf{\Sigma}}^{-1} \mathbf{d}} \right)^{-1}, \end{aligned} \tag{6}$$

where, for notational convenience, we have used

$$\mathbf{d} := \overline{\mathbf{x}} - \overline{\mathbf{y}}.$$

We proceed by determining the distributions of the numerator and denominator, sequentially.

Observing that $\mathbf{x}_i \sim \mathcal{N}(\boldsymbol{\mu}_x, \mathbf{\Sigma})$ for all i, and similarly for $\mathbf{y}_i$, we have

$$\overline{\mathbf{x}} \sim \mathcal{N}(\boldsymbol{\mu}_x, N_x^{-1} \mathbf{\Sigma}) \quad \text{and} \quad \overline{\mathbf{y}} \sim \mathcal{N}(\boldsymbol{\mu}_y, N_y^{-1} \mathbf{\Sigma}).$$

Thus, their difference $\mathbf{d} = \overline{\mathbf{x}} - \overline{\mathbf{y}}$ is also normally distributed as

$$\mathbf{d} \sim \mathcal{N}(\boldsymbol{\delta}, N(N_x N_y)^{-1} \mathbf{\Sigma}),$$

where $\boldsymbol{\delta} := \boldsymbol{\mu}_x - \boldsymbol{\mu}_y$. It follows then that

$$(N_x N_y N^{-1})^{1/2} \mathbf{\Sigma}^{-1/2} \mathbf{d} \sim \mathcal{N}((N_x N_y N^{-1})^{1/2} \mathbf{\Sigma}^{-1/2} \boldsymbol{\delta}, \mathcal{I}).$$

Re-expressing the numerator term in (6) as

$$\mathbf{d}^T \mathbf{\Sigma}^{-1} \mathbf{d} = \left(\mathbf{\Sigma}^{-1/2} \mathbf{d} \right)^T \left(\mathbf{\Sigma}^{-1/2} \mathbf{d} \right)$$

we readily find

$$N_x N_y N^{-1} \mathbf{d}^T \boldsymbol{\Sigma}^{-1} \mathbf{d} \sim \chi^2(M, \lambda), \tag{7}$$

which denotes the noncentral χ^2 distribution with M degrees of freedom and noncentrality parameter

$$\lambda = N_x N_y N^{-1} \boldsymbol{\delta}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\delta}. \tag{8}$$

Next, because $\mathbf{x}_i \sim \mathcal{N}(\boldsymbol{\mu}_x, \boldsymbol{\Sigma})$ and $\mathbf{y}_i \sim \mathcal{N}(\boldsymbol{\mu}_y, \boldsymbol{\Sigma})$ for all i, it can be shown (see, e.g., [3, Chapter 3]) that

$$(N_x - 1)\widehat{\boldsymbol{\Sigma}}_x \sim \mathcal{W}(\boldsymbol{\Sigma}, N_x - 1) \quad \text{and} \quad (N_y - 1)\widehat{\boldsymbol{\Sigma}}_y \sim \mathcal{W}(\boldsymbol{\Sigma}, N_y - 1),$$

where $\mathcal{W}(\boldsymbol{\Sigma}, P)$ denotes the Wishart distribution with scale matrix $\boldsymbol{\Sigma}$ and P degrees of freedom. It follows then (see, e.g., [3, Theorem 3.4.3]) that

$$(N - 2)\widehat{\boldsymbol{\Sigma}} \sim \mathcal{W}(\boldsymbol{\Sigma}, N - 2).$$

Therefore, as is shown in, e.g., [3, Theorem 3.4.7] we have

$$(N - 2)\frac{\mathbf{d}^T \boldsymbol{\Sigma}^{-1} \mathbf{d}}{\mathbf{d}^T \widehat{\boldsymbol{\Sigma}}^{-1} \mathbf{d}} \sim \chi^2(N - M - 1). \tag{9}$$

Finally, we combine (7), (8), and (9), to find

$$\begin{aligned}
\left(\frac{N - M - 1}{M}\right) \frac{N_x N_y}{N(N-2)} T^2 &= \left(\frac{N - M - 1}{M}\right) N_x N_y N^{-1} \mathbf{d}^T \boldsymbol{\Sigma}^{-1} \mathbf{d} \times \\
&\qquad \left((N-2)\frac{\mathbf{d}^T \boldsymbol{\Sigma}^{-1} \mathbf{d}}{\mathbf{d}^T \widehat{\boldsymbol{\Sigma}} \mathbf{d}}\right)^{-1} \\
&\sim \left(\frac{N - M - 1}{M}\right) \frac{\chi^2(M, \lambda)}{\chi^2(N - M - 1)} \\
&\sim F(M, N - M - 1, \lambda).
\end{aligned}$$

□

References

[1] Patrick Billingsley. *Probability and measure*. John Wiley & Sons, 2008.

[2] Harold Hotelling. The Generalization of Student's Ratio. *The Annals of Mathematical Statistics*, 2(3):360 – 378, 1931. doi: 10.1214/aoms/1177732979. URL https://doi.org/10.1214/aoms/1177732979.

[3] JT Kent, John Bibby, and KV Mardia. *Multivariate analysis*. Academic Press Amsterdam, 1979.

[4] Erich L Lehmann and Joseph P Romano. *Testing statistical hypotheses*. Springer Science & Business Media, 2006.

[5] Alvin C Rencher. *Multivariate statistical inference and applications*. Wiley New York, 1998.

[6] Alvin C Rencher. *Methods of multivariate analysis*, volume 492. John Wiley & Sons, 2003.