Simultaneous Tests for Homogeneity of Two Zero-inflated (Beta) Populations

Supplemental file

Luna Sun and Alix I. Gitelman

Department of Statistics, Oregon State University

Result 1 (Partition of multivariate normal distribution (Seely’s Notes, unpublished)). Suppose $Y \sim N_p(\mu,V)$ where $V$ is nonsingular. Let $Y$, $\mu$ and $V$ be similarly partitioned in the form $Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$, $\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$, and $V = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}$. Assume $Y_1$ is $q \times 1$ and that $s = p - q$ so that $Y_2$ is $s \times 1$. Set $V_{11} = V_{11} - V_{12}V_{22}^{-1}V_{21}$ and let $T = Y_1 - V_{12}V_{22}^{-1}Y_2$. Then

1. $T \sim N_q(\mu_1 - V_{12}V_{22}^{-1}\mu_2, V_{11})$ and $V_{11}$ is nonsingular.

2. $T$ and $Y_2$ are independent random vectors.

Theorem 1 (Asymptotic Independence among Score Tests). $\hat{S}_1$, $\hat{S}_2$ and $\hat{S}_3$ (and also $\tilde{S}_1$, $\tilde{S}_2$ and $\tilde{S}_3$) are asymptotically independent.

Proof. Consider $\hat{S}_1$, $\hat{S}_2$ and $\hat{S}_3$ first: Denote

$$\omega = (p, \mu, \phi)\, \zeta = (\delta, p, \mu, \phi) = (\delta, \omega)' \quad \eta = (\gamma, \delta, p, \mu, \phi) = (\gamma, \zeta)' = (\gamma, \delta, \omega)'$$

$$\theta = (\beta, \gamma, \delta, p, \mu, \phi)' = (\beta, \eta)' = (\beta, \gamma, \zeta)' = (\beta, \gamma, \delta, \omega)'$$

Recall that

$$\hat{S}_1 = \hat{s}_1'(\hat{A}_1 - \hat{C}_1\hat{D}_1^{-1}\hat{C}_1')^{-1}\hat{s}_1,$$

$$\hat{S}_2 = \hat{s}_2'(\hat{A}_2 - \hat{C}_2\hat{D}_2^{-1}\hat{C}_2')^{-1}\hat{s}_2$$

and

$$\hat{S}_3 = \hat{s}_3'(\hat{A}_3 - \hat{C}_3\hat{D}_3^{-1}\hat{C}_3')^{-1}\hat{s}_3.$$

Under $H_0$,

$$s_1 = \frac{\partial l_f}{\partial \delta} = \frac{\partial l_f}{\partial \delta}, \quad s_2 = \frac{\partial l_f}{\partial \gamma} = \frac{\partial l_f}{\partial \gamma}, \quad \text{and} \quad s_3 = \frac{\partial l_f}{\partial \beta} = \frac{\partial l_f}{\partial \beta}.$$
And
\[ \hat{s}_1 = s_1(\omega), \hat{s}_2 = s_2(\zeta), \hat{s}_3 = s_3(\eta), \]
where \( \hat{\omega}, \hat{\zeta} \) and \( \hat{\eta} \) are maximum likelihood estimates of \( \omega \) under \( H'_0 \), \( \zeta \) under \( H'_0 \) and \( \eta \) under \( H''_0 \), respectively (i.e., \( \hat{s}_1 = \frac{\partial l}{\partial \delta} \bigg|_{\omega=\hat{\omega}}, \hat{s}_2 = \frac{\partial l}{\partial \gamma} \bigg|_{\zeta=\hat{\zeta}} \) and \( \hat{s}_3 = \frac{\partial l}{\partial \beta} \bigg|_{\eta=\hat{\eta}} \)). Notice that when \( H_0 \) is true, \( H'_0, H''_0 \) and \( H'''_0 \) are all true. Expanding \( \hat{s}_1, \hat{s}_2 \) and \( \hat{s}_3 \) around the true parameters \( \theta_0 \) via Taylor expansion with \( l_f = l \):
\[
\hat{s}_1 = \frac{\partial l}{\partial \delta} - I_{\delta \omega_0} I_{\omega_0 \omega_0}^{-1} \frac{\partial l}{\partial \omega_0} + O_p(1),
\]
\[
\hat{s}_2 = \frac{\partial l}{\partial \gamma} - I_{\gamma \zeta_0} I_{\zeta_0 \zeta_0}^{-1} \frac{\partial l}{\partial \zeta_0} + O_p(1),
\]
and
\[
\hat{s}_3 = \frac{\partial l}{\partial \beta} - I_{\beta \eta_0} I_{\eta_0 \eta_0}^{-1} \frac{\partial l}{\partial \eta_0} + O_p(1),
\]
where
\[
I_{\delta \omega_0} = E \left( -\frac{\partial^2 l}{\partial \delta \partial \omega^T} \bigg| H_0 \right), \quad I_{\omega_0 \omega_0} = E \left( -\frac{\partial^2 l}{\partial \omega \partial \omega^T} \bigg| H_0 \right), \quad I_{\gamma \zeta_0} = E \left( -\frac{\partial^2 l}{\partial \gamma \partial \zeta^T} \bigg| H_0 \right), \quad I_{\zeta_0 \zeta_0} = E \left( -\frac{\partial^2 l}{\partial \zeta \partial \zeta^T} \bigg| H_0 \right),
\]
\[
I_{\gamma \eta_0} = E \left( -\frac{\partial^2 l}{\partial \gamma \partial \eta^T} \bigg| H_0 \right), \quad I_{\eta_0 \eta_0} = E \left( -\frac{\partial^2 l}{\partial \eta \partial \eta^T} \bigg| H_0 \right),
\]
and
\[
\left. \frac{\partial l}{\partial \omega_0} \right|_{\omega=\hat{\omega}}, \quad \left. \frac{\partial l}{\partial \zeta_0} \right|_{\zeta=\hat{\zeta}}, \quad \left. \frac{\partial l}{\partial \eta_0} \right|_{\eta=\hat{\eta}}.
\]
Let
\[
s_{01} = \frac{\partial l}{\partial \delta} - I_{\delta \omega_0} I_{\omega_0 \omega_0}^{-1} \frac{\partial l}{\partial \omega_0}, \quad s_{02} = \frac{\partial l}{\partial \gamma} - I_{\gamma \zeta_0} I_{\zeta_0 \zeta_0}^{-1} \frac{\partial l}{\partial \zeta_0} \quad \text{and} \quad s_{03} = \frac{\partial l}{\partial \beta} - I_{\beta \eta_0} I_{\eta_0 \eta_0}^{-1} \frac{\partial l}{\partial \eta_0}.
\]
Since the score function can be written as
\[
\frac{\partial l}{\partial \theta} = \left( \begin{array}{c}
\frac{\partial l}{\partial \beta} \\
\frac{\partial l}{\partial \gamma} \\
\frac{\partial l}{\partial \delta} \\
\frac{\partial l}{\partial \eta} \\
\frac{\partial l}{\partial \mu} \\
\frac{\partial l}{\partial \phi} \\
\end{array} \right) = \left( \begin{array}{c}
\frac{\partial s_{01}}{\partial \beta} \\
\frac{\partial s_{02}}{\partial \gamma} \\
\frac{\partial s_{03}}{\partial \delta} \\
\frac{\partial s_{04}}{\partial \eta} \\
\frac{\partial s_{05}}{\partial \mu} \\
\frac{\partial s_{06}}{\partial \phi} \\
\end{array} \right) = \left( \begin{array}{c}
\frac{\partial l}{\partial \beta} \\
\frac{\partial l}{\partial \gamma} \\
\frac{\partial l}{\partial \delta} \\
\frac{\partial l}{\partial \eta} \\
\frac{\partial l}{\partial \mu} \\
\frac{\partial l}{\partial \phi} \\
\end{array} \right),
\]
we have $\frac{\partial l}{\partial \theta} \xrightarrow{d} N(0, I(\theta_0))$. Or
\[
\left( \begin{array}{c}
\frac{\partial l}{\partial \beta} \\
\frac{\partial l}{\partial \eta}
\end{array} \right) \xrightarrow{d} N\left( \begin{array}{c}0 \\
0
\end{array} \right), -E \left( \begin{array}{cc}
\frac{\partial^2 l}{\partial \beta^2} & \frac{\partial^2 l}{\partial \beta \eta} \\
\frac{\partial^2 l}{\partial \eta \beta} & \frac{\partial^2 l}{\partial \eta^2}
\end{array} \right) \left| H_0 \right|.
\]

As the sample size going to $+\infty$, using the result from partition of multivariate normal distribution, under $H_0$, $s_{03}$ is independent of $\frac{\partial l}{\partial \eta} = \left( \begin{array}{c}
\frac{\partial l}{\partial \gamma} \\
\frac{\partial l}{\partial \delta}
\end{array} \right) = \left( \begin{array}{c}
\frac{\partial l}{\partial \gamma} \\
\frac{\partial l}{\partial \delta}
\end{array} \right)$. Since $s_{01}$ and $s_{02}$ are only functions of $(\frac{\partial l}{\partial \gamma}, \frac{\partial l}{\partial \delta})$ and $(\frac{\partial l}{\partial \gamma}, \frac{\partial l}{\partial \delta})$, $s_{03}$ is independent of $s_{01}$ and $s_{02}$. Similarly, it is easy to show $s_{01}$ and $s_{02}$ are independent as well.

For the independence among $\tilde{S}_1, \tilde{S}_2$ and $\tilde{S}_3$, similar arguments can be made.

**Theorem 2 (Size of Hybrid test).** The hypothesis in (8) is

$\text{H}_0 : p_1 = p_2$ and $h_{Y_1} = h_{Y_2}$ vs $\text{H}_1 : \text{Either equality fails.}$

Here, $p_1$ and $p_2$ are the population zero proportions for populations 1 and 2, and $h_{Y_1}$ and $h_{Y_2}$ are the pdf for the non-zero components of populations 1 and 2. To test these hypotheses at the level $\alpha$, the Hybrid procedure involves the following steps:

1. **Test** $H_{01} : p_1 = p_2$ versus $H_{11} : p_1 \neq p_2$.

2. $H_{01}$ is rejected if the p-value is $p < \alpha_1$. In this case, reject $H_0$ in (8); otherwise, test $H_{02} : h_{Y_1} = h_{Y_2}$ versus $H_{12} : h_{Y_1} \neq h_{Y_2}$.

3. $H_{02}$ is rejected if the p-value is $p < \alpha_2$. In this case, reject $H_0$ in (8); otherwise, do not reject $H_0$.

The size of Hybrid is less than or equal to $\alpha_1 + \alpha_2$.

**Proof.** Define
\[
A = \{(p_1, p_2) \in (0, 1)^2 : p_1 = p_2\} \quad \text{and} \quad B = \{h_{Y_1}, h_{Y_2} \in \mathcal{F} : h_{Y_1} = h_{Y_2}\}
\]
where $\mathcal{F}$ is the space of a particular family of pdf. Under the set notation, $H_0, H_1, H_{01}, H_{11}, H_{02}$ and $H_{12}$ can be expressed as $H_0 : A \cap B, H_1 : A^c \cup B^c, H_{01} : A, H_{11} : A^c, H_{02} : A \cap B$ and $H_{12} : A \cap B^c$.

Further define $\mathcal{R}_1, \mathcal{R}_2$ and $\mathcal{R}$ as rejection region for $H_{01}, H_{02}$ and $H_0$, then
\[
C = \{x : x \in \mathcal{R}_1\} \quad \text{and} \quad D = \{x : x \in \mathcal{R}_2\},
\]

3
then we have
\[ C \cup (C^c \cap D) = \{ x : x \in \mathbb{R} \}. \]

It is reasonable to assume \( A \) and \( B \) are independent because the form of \( h_y \) does not depend on \( p \), neither the other way around. Also, it is reasonable to assume \( B \) and \( C \) are independent because \( C \) only concerns the \( p \)'s.

Let \( \alpha_0, \alpha_1 \) and \( \alpha_2 \) denote the type I error rates of \( H_0 \) vs \( H_1 \), \( H_{01} \) vs \( H_{11} \) and \( H_{02} \) vs \( H_{12} \), respectively.

Then
\[ \alpha_0 = P(\text{Reject } H_0 | H_0 \text{ is true}) = P(C \cup (C^c \cap D) | A \cap B), \]
\[ \alpha_1 = P(C | A) \text{ and } \alpha_2 = P(D | A \cap B). \]

Since \( C \cap (C^c \cap D) = \emptyset \), we have \( (C | A \cap B) \cap (C^c \cap D | A \cap B) = \emptyset \). Therefore,
\[ P(C \cup (C^c \cap D) | A \cap B) = P(C | A \cap B) + P((C^c \cap D) | A \cap B). \]

Since \( C^c \cap D \subset D \), there is \( C^c \cap D | A \cap B \subset D | A \cap B \). Therefore,
\[ P((C^c \cap D) | A \cap B) \leq P(D | A \cap B) = \alpha_2. \]

\[ P(C | A \cap B) = \frac{P(C \cap A \cap B)}{P(A \cap B)} = \frac{P(B | A \cap C) \cdot P(A \cap C)}{P(B | A) \cdot P(A)} = P(C | A) \cdot \frac{P(B | A \cap C)}{P(B | A)}. \]

Because we assume \( A \) and \( B \) are independent and \( B \) and \( C \) are independent, \( P(B | A) = P(B) \) and \( P(B | A \cap C) = P(B) \). Hence,
\[ P(C | A \cap B) = P(C | A) = \alpha_1 \]

As a result,
\[ \alpha_0 = P(C \cup (C^c \cap D) | A \cap B) = P(C | A \cap B) + P((C^c \cap D) | A \cap B) \leq \alpha_1 + \alpha_2. \]

Under the Bonferroni correction, \( \alpha_1 = \alpha_2 = \alpha / 2. \)

\[ \square \]

**Reference**

Justus Seely. Linear Model Theory Notes, unpublished. *Department of Statistics, Oregon State University.*