Fast score test with global null estimation
regardless of missing genotypes

Shuntaro Sato* Masao Ueki†

Details of the method

Hjort and Claeskens framework

We use the asymptotic results of Hjort and Claeskens to study our statistical methods [1]. Hjort and Claeskens derived an asymptotic distribution of the maximum likelihood estimator (MLE) of the parameters \( \theta \) under the following
distribution for \( i \)th subject \( (i = 1, \ldots, n) \),

\[
f_i(Y) = f_i(Y, \theta) = f_i(Y, \theta_0 + \delta/\sqrt{n}),
\]

(1)
in which our \( \theta_0 \) corresponds to \( \gamma_0 \) in equation (2.2) of Hjort and Claeskens and we omit their \( \theta_0 \). The parameter \( \delta \) represents degrees of departures from the null model \( f_i(Y, \theta_0) \). Thus, \( \delta/\sqrt{n} \) represents a small perturbation from the null model, i.e. local alternative. It is suitable to study asymptotic distribution of test statistics.

*shuntarosato@nagasaki-u.ac.jp, Clinical Research Center, Nagasaki University Hospital, 1-7-1 Sakamoto, Nagasaki, Nagasaki 852-8501, Japan. Biostatistics, Graduate School of Medicine, Kurume University, 67 Asahi-machi, Kurume, Fukuoka, 830-0011, Japan.
†Statistical Genetics Team, RIKEN Center for Advanced Intelligence Project, 1-4-1 Nihonbashhi, Chuo-ku, Tokyo, 103-0027, Japan.
Under the assumptions in Hjort and Claeskens,

\[ f_i(Y_i, \theta_0 + \frac{\delta}{\sqrt{n}}) = f_{0i}(Y_i) \left\{ 1 + u_i(Y_i)\frac{\delta}{\sqrt{n}} + o(\frac{1}{\sqrt{n}}) \right\}, \]

where \( f_{0i}(Y_i) = f_i(Y_i, \theta_0) \) and \( u_i(Y_i) = \partial \log f_i(Y_i, \theta_0)/\partial \theta \). Then, the mean of \( u_i(Y_i) \) is given by

\[
E\{u_i(Y_i)\} = \int u_i(Y_i)f_i(Y_i, \theta_0 + \frac{\delta}{\sqrt{n}})dY_i = \int u_i(Y_i)f_{0i}(Y_i) + \int u_i(Y_i)u_i(Y_i)^Tf_{0i}(Y_i)dY_i \frac{\delta}{\sqrt{n}} + o(\frac{1}{\sqrt{n}})
\]

\[ = \frac{1}{\sqrt{n}}J_{11,i} + o(\frac{1}{\sqrt{n}}). \tag{2} \]

where

\[
J_{11,i} = -\int \frac{\partial^2 \log f_i(Y_i, \theta_0)}{\partial \theta \partial \theta^T} f_{0i}(Y_i)dY_i
\]

\[ = \int u_i(Y_i)u_i(Y_i)^T f_{0i}(Y_i)dY_i \]

Similarly, the variance of \( u_i(Y_i) \) is

\[
\text{var}\{u_i(Y_i)\} = \int u_i(Y_i)u_i(Y_i)^T f_{0i}(Y_i) + \frac{\delta}{\sqrt{n}})dY_i - E\{u_i(Y_i)\}E\{u_i(Y_i)\}^T
\]

\[ = \int u_i(Y_i)u_i(Y_i)^T f_{0i}(Y_i)\left\{ 1 + u_i(Y_i)^T \frac{\delta}{\sqrt{n}} + o(\frac{1}{\sqrt{n}}) \right\} dY_i + o(1)
\]

\[ = \int u_i(Y_i)u_i(Y_i)^T f_{0i}(Y_i)dY_i + \int u_i(Y_i)u_i(Y_i)^T u_i(Y_i)f_{0i}(Y_i)dY_i \frac{\delta}{\sqrt{n}} + o(1)
\]

\[ = J_{11,i} + \int u_i(Y_i)u_i(Y_i)^T u_i(Y_i)f_{0i}(Y_i)dY_i \frac{\delta}{\sqrt{n}} + o(1)
\]

\[ = J_{11,i} + o(1). \tag{3} \]

Let \( f(Y, \theta_0) = \prod_{i=1}^n f_i(Y_i, \theta_0) \). Suppose that \( \theta \) is partitioned into two parts,
\((\theta_1^T, \theta_2^T)^T\), in which \(\theta_1\) and \(\theta_2\) are \(s\)-dimensional and \(r\)-dimensional vectors, respectively. For example, a logistic regression model for a SNP is written as

\[
\text{logit}[Pr(Y_i = 1)] = \text{logit}[\pi_i(\beta_0, \beta_e, \beta_g)] = \beta_0 + \beta_e E_i + \beta_g G_i
\]

where \(\pi_i\) is a probability of being case, \(G_i\) is some genotype coding, such as an addition coding \(\{0, 1, 2\}\) and \(E_i\) is a covariate or an environment factor. Letting \(\theta_1 = (\beta_0, \beta_e)^T, \theta_2 = \beta_g, X_1 = (1, E), \) and \(X_2 = G\), the probability of being case at a full model is

\[
\pi_i(\theta) = \pi_i(\theta_1, \theta_2) = \frac{\exp(\beta_0 + \beta_e E_i + \beta_g G_i)}{1 + \exp(\beta_0 + \beta_e E_i + \beta_g G_i)} = \frac{\exp(X_1 \theta_1 + X_2 \theta_2)}{1 + \exp(X_1 \theta_1 + X_2 \theta_2)}.
\]

Under these setting, the log-likelihood function at the full model is

\[
f(\theta) = \prod_{i=1}^{n} \{\pi_i Y_i (1 - \pi_i)^{(1-Y_i)}\}
\]

\[
\log f(\theta) = \sum_{i=1}^{n} [Y_i \log \pi_i(\theta) + (1 - Y_i) \log \{1 - \pi_i(\theta)\}].
\]

In the same way, a logistic regression model for a joint test is written as

\[
\text{logit}[Pr(Y_i = 1)] = \beta_0 + \beta_e E_i + \beta_g G_i + \beta_{ge} G_i E_i
\]

where \(G_i E_i\) is a gene-environment interaction. Letting \(\theta_1 = (\beta_0, \beta_e)^T, \theta_2 = (\beta_g, \beta_{ge})^T, X_1 = (1, E)\), and \(X_2 = (G, GE)\). We consider the distribution of \(u(Y)\) in the complete data where

\[
u(Y) = \begin{pmatrix} u_1(Y) \\ u_2(Y) \end{pmatrix} = \begin{pmatrix} \partial \log f(Y, \theta_0)/\partial \theta_1 \\ \partial \log f(Y, \theta_0)/\partial \theta_2 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} u_{1i}(Y_i) \\ \sum_{i=1}^{n} u_{2i}(Y_i) \end{pmatrix}.
\]
From (2), the mean of $u(Y)$ is
\[
E\left(\frac{\sum_{i=1}^{n} u_{1i}(Y_i)}{\sum_{i=1}^{n} u_{2i}(Y_i)}\right) = \frac{1}{\sqrt{n}} \left(\frac{\sum_{i=1}^{n} J_{11,i}[1]}{\sum_{i=1}^{n} J_{11,i}[2]}\right) + o(\sqrt{n}).
\]

From (3), the variance of $u(Y)$ is
\[
\text{var}\left(\frac{\sum_{i=1}^{n} u_{1i}(Y_i)}{\sum_{i=1}^{n} u_{2i}(Y_i)}\right) = \frac{n}{\sum_{i=1}^{n} J_{11,i}[1]} \sum_{i=1}^{n} \text{var}\left\{u_{1i}(Y_i)\right\} + \sum_{i=1}^{n} \text{cov}\left\{u_{1i}(Y_i), u_{2i}(Y_i)\right\} + o(n).
\]

By letting
\[
J_{11} = \begin{pmatrix} J_{11}[1,1] & J_{11}[1,2] \\ J_{11}[2,1] & J_{11}[2,2] \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} J_{11,i}[1,1] & \sum_{i=1}^{n} J_{11,i}[1,2] \\ \sum_{i=1}^{n} J_{11,i}[2,1] & \sum_{i=1}^{n} J_{11,i}[2,2] \end{pmatrix},
\]
the distribution of $u(Y)/\sqrt{n}$ is asymptotically
\[
\frac{1}{\sqrt{n}} u(Y) \sim N_{s+r}\left(\frac{1}{n} \begin{pmatrix} J_{11}[1,1] \\ J_{11}[2,1] \end{pmatrix}, \frac{1}{n} \begin{pmatrix} J_{11}[1,1] & J_{11}[1,2] \\ J_{11}[2,1] & J_{11}[2,2] \end{pmatrix}\right) = N_{s+r}\left(\frac{1}{n} J_{11} \delta, \frac{1}{n} J_{11}\right),
\]
which is exactly the statement in Lemma 3.1 of Hjort and Claeskens.

Similarly, we consider the distribution of $u^m(Y)$ in the presence of missing.
data, in which

\[ u^m(Y) = \left( \frac{\sum_{i=1}^{n} u_{1i}(Y_i)I_i}{\sum_{i=1}^{n} u_{2i}(Y_i)I_i} \right). \]

The mean of \( u^m(Y) \) is

\[ E \left( \frac{\sum_{i=1}^{n} u_{1i}(Y_i)I_i}{\sum_{i=1}^{n} u_{2i}(Y_i)I_i} \right) = \frac{1}{\sqrt{n}} \left( \frac{\sum_{i=1}^{n} J_{11,i}[1,1]\delta I_i}{\sum_{i=1}^{n} J_{11,i}[2,1]\delta I_i} \right) + o(\sqrt{n}). \]

The variance of \( u^m(Y) \) is

\[ \text{var} \left( \frac{\sum_{i=1}^{n} u_{1i}(Y_i)I_i}{\sum_{i=1}^{n} u_{2i}(Y_i)I_i} \right) = \sum_{i=1}^{n} \left[ \text{var} \{ u_{1i}(Y_i)I_i \} \right. \]

\[ + \text{cov} \{ u_{1i}(Y_i), u_{2i}(Y_i)I_i \} \]

\[ - \text{cov} \{ u_{2i}(Y_i), u_{1i}(Y_i)I_i \} \] \[ + o(n) \]

\[ = \left( \sum_{i=1}^{n} J_{11,i}[1,1]I_i - \sum_{i=1}^{n} J_{11,i}[1,2]I_i \right) \]

\[ + \left( \sum_{i=1}^{n} J_{11,i}[2,1]I_i - \sum_{i=1}^{n} J_{11,i}[2,2]I_i \right) \] \[ + o(n). \]

By letting

\[ J_{11}^m = \left( \begin{array}{cc} J_{11}^m[1,1] & J_{11}^m[1,2] \\ J_{11}^m[2,1] & J_{11}^m[2,2] \end{array} \right) = \left( \begin{array}{cc} \sum_{i=1}^{n} J_{11,i}[1,1]I_i & \sum_{i=1}^{n} J_{11,i}[1,2]I_i \\ \sum_{i=1}^{n} J_{11,i}[2,1]I_i & \sum_{i=1}^{n} J_{11,i}[2,2]I_i \end{array} \right), \]

Hence, the distribution of \( u^m(Y)/\sqrt{n} \) is asymptotically

\[ \frac{1}{\sqrt{n}} u^m(Y) \sim N_{s+r} \left( \frac{1}{n} J_{11}^m[1,1]I, \frac{1}{n} J_{11}^m[1,2]I \right) = N_{s+r} \left( \frac{1}{n} J_{11}^m \delta, \frac{1}{n} J_{11}^m \right). \]

(6)
Analogously, we consider the distribution of $u^{m'}(Y)$ in the presence of missing data, where $u^{m'}(Y)$ is

$$ u^{m'}(Y) = \begin{pmatrix} \sum_{i=1}^{n} u_{1i}(Y_i) \\ \sum_{i=1}^{n} u_{2i}(Y_i) I_i \end{pmatrix}. $$

The mean of $u^{m'}(Y)$ is

$$ E \left( \frac{\sum_{i=1}^{n} u_{1i}(Y_i)}{\sum_{i=1}^{n} u_{2i}(Y_i) I_i} \right) = \frac{1}{\sqrt{n}} \left( \sum_{i=1}^{n} J_{11,i}[1,] \delta \right) + o(\sqrt{n}). $$

The variance of $u^{m'}(Y)$ is

$$ \text{var} \left( \frac{\sum_{i=1}^{n} u_{1i}(Y_i)}{\sum_{i=1}^{n} u_{2i}(Y_i) I_i} \right) = \sum_{i=1}^{n} \text{var} \left( \frac{u_{1i}(Y_i)}{u_{2i}(Y_i) I_i} \right) + o(n) $$

$$ = \sum_{i=1}^{n} \begin{pmatrix} \text{var}(u_{1i}(Y_i)) & \text{cov}(u_{1i}(Y_i), u_{2i}(Y_i) I_i) \\ \text{cov}(u_{2i}(Y_i) I_i, u_{1i}(Y_i)) & \text{var}(u_{2i}(Y_i) I_i) \end{pmatrix} + o(n) $$

$$ = \begin{pmatrix} \sum_{i=1}^{n} J_{11,i}[1,1] & \sum_{i=1}^{n} J_{11,i}[1,2] I_i \\ \sum_{i=1}^{n} J_{11,i}[2,1] I_i & \sum_{i=1}^{n} J_{11,i}[2,2] I_i \end{pmatrix} + o(n). $$

Hence, the distribution of $u^{m'}(Y)/\sqrt{n}$ is asymptotically

$$ \frac{1}{\sqrt{n}} u^{m'}(Y) \sim N_{s+r} \left( \frac{1}{n} \begin{pmatrix} J_{11}[1,] \\ J_{11}[2,] \delta \end{pmatrix}, \frac{1}{n} \begin{pmatrix} J_{11}[1,1] & J_{11}[1,2] \\ J_{11}[2,1] & J_{11}[2,2] \end{pmatrix} \right) = N_{s+r} \left( \frac{1}{n} J_{11}^r \delta, \frac{1}{n} J_{11} \right), $$

(7)
\[ J_{11}^f = \begin{pmatrix} J_{11}[1, 1] & J_{11}^m[1, 2] \\ J_{11}^m[2, 1] & J_{11}^m[2, 2] \end{pmatrix} \quad \text{and} \quad J_{11}^{f'} = \begin{pmatrix} J_{11}[1, 1] & J_{11}[1, 2] \\ J_{11}[2, 1] & J_{11}[2, 2] \end{pmatrix}. \]

**Conventional Score Test (CST)**

Analysis using only individuals whose genotype data is observed is referred to as the complete case analysis. We call the score test in the complete case analysis as the conventional score test (CST). Let

\[ \left( \begin{array}{c} u_m^1(\theta_1) \\ u_m^2(\theta_1) \end{array} \right) = \left( \begin{array}{c} u_1^m(\theta) \\ u_2^m(\theta) \end{array} \right) \bigg|_{\theta = (\theta_1^T, 0)^T} = \frac{1}{n} \sum_{i=1}^{n} \frac{u_{1i}(Y_i, \theta)I_i}{\sum_{i=1}^{n} u_{2i}(Y_i, \theta)I_i}, \]

\[ \bigg|_{\theta = (\theta_1^T, 0)^T}. \]

Suppose that the null model correspond to the parameter \( \theta_0 = (\theta_{10}^T, 0)^T \), then

\[ u_m^m(Y) = \left( \begin{array}{c} u_1^m(Y) \\ u_2^m(Y) \end{array} \right) = \left( \begin{array}{c} u_1^m(\theta_{10}) \\ u_2^m(\theta_{10}) \end{array} \right). \]

At the null model on CST, i.e. \( \theta_2 = 0 \), the MLE of \( \theta_1 \) is denoted by \( \hat{\theta}_m^1 \).

By the Taylor expansion, the score function is written as

\[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_{m1}^m(\hat{\theta}_1^m) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ u_{m1}^m(\theta_{10}) + \frac{\partial u_{m1}^m(\theta_{10})}{\partial \theta_1} (\hat{\theta}_m^1 - \theta_{10}) \right\} + o_p(1) \]

\[ = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \frac{1}{n} \sum_{i=1}^{n} u_{2i}^m(\theta_{10}) + \frac{1}{n} \sum_{i=1}^{n} \frac{\partial u_{2i}^m(\theta_{10})}{\partial \theta_1} (\hat{\theta}_m^1 - \theta_{10}) \right\} + o_p(1) \]

\[ = \frac{1}{\sqrt{n}} u_{21}^m(\theta_{10}) - \frac{1}{n} J_{11}[2, 1] \sqrt{n} (\hat{\theta}_m^1 - \theta_{10}) + o_p(1). \]

Using an analogous argument in the proof of Lemma 3.2 of Hjort and Claeskens,
we have
\[ \sqrt{n}(\hat{\theta}_1^m - \theta_{10}) = J_{11}^m[1,1]^{-1}\sqrt{n}u_1^m(\theta_{10}) + o_p(1). \]

Substituting the above equation,
\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_2^m(\hat{\theta}_1^m) = \frac{1}{\sqrt{n}} u_2^m(\theta_{10}) - \frac{1}{n} J_{11}^m[2,1] \left\{ J_{11}^m[1,1]^{-1}\sqrt{n}u_1^m(\theta_{10}) + o_p(1) \right\} + o_p(1)
\]
\[ = \frac{1}{\sqrt{n}} u_2^m(\theta_{10}) - J_{11}^m[2,1]J_{11}^m[1,1]^{-1}\frac{1}{\sqrt{n}} u_1^m(\theta_{10}) + o_p(1) \]  
(8)
\[ = \left( -J_{11}^m[2,1]J_{11}^m[1,1]^{-1} \right) \frac{1}{\sqrt{n}} \begin{pmatrix} u_1^m(\theta_{10}) \\ u_2^m(\theta_{10}) \end{pmatrix} + o_p(1)
\]
\[ = L^m \frac{1}{\sqrt{n}} u^m(Y) + o_p(1), \]
in which
\[ L^m = \begin{pmatrix} -J_{11}^m[2,1]J_{11}^m[1,1]^{-1} & 1 \end{pmatrix}. \]

Applying equation (6), under an alternative hypothesis (1), the distribution of the CST score function is asymptotically
\[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_2^m(\hat{\theta}_1^m) \sim N_r \left( \frac{1}{n} L^m J_{11}^m \delta, \frac{1}{n} L^m J_{11}^m J_{11}^m L^m \right) \]
where
\[ L^m J_{11}^m \delta = \begin{pmatrix} -J_{11}^m[2,1]J_{11}^m[1,1]^{-1} & 1 \end{pmatrix} \begin{pmatrix} J_{11}^m[1,1] & J_{11}^m[1,2] \\ J_{11}^m[2,1] & J_{11}^m[2,2] \end{pmatrix} \delta
\]
\[ = \begin{pmatrix} 0 & -J_{11}^m[2,1]J_{11}^m[1,1]^{-1}J_{11}^m[1,2] + J_{11}^m[2,2] \end{pmatrix} \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix}
\]
\[ = (-J_{11}^m[2,1]J_{11}^m[1,1]^{-1}J_{11}^m[1,2] + J_{11}^m[2,2])\delta_2 \]
and

\[
L^m J_{11}^m L^{mT} = \begin{pmatrix}
-J_{11}^m[2,1]J_{11}^m[1,1]^{-1} & 1 \\
-J_{11}^m[1,1]^{-1}J_{11}^m[1,2] & I
\end{pmatrix}
\begin{pmatrix}
-J_{11}^m[1,1]^{-1}J_{11}^m[1,2] & \\
-J_{11}^m[1,1]^{-1}J_{11}^m[1,2] & I
\end{pmatrix}
\]

\[
= \begin{pmatrix}
0 & -J_{11}^m[2,1]J_{11}^m[1,1]^{-1}J_{11}^m[1,2] + J_{11}^m[2,2] & -J_{11}^m[1,1]^{-1}J_{11}^m[1,2] & \\
-J_{11}^m[2,1]J_{11}^m[1,1]^{-1}J_{11}^m[1,2] + J_{11}^m[2,2] & I
\end{pmatrix}
\]

\[
= -J_{11}^m[2,1]J_{11}^m[1,1]^{-1}J_{11}^m[1,2] + J_{11}^m[2,2].
\]

Let \( u_m = (1/\sqrt{n}) \sum_{i=1}^n u(m)^{2}_{2i}(\hat{\theta}_1^m) \), \( \mu_m = \frac{1}{n} L^m J_{11}^m \delta \) and \( V_m = \frac{1}{n} L^m J_{11}^m L^{mT} \). Then, \( u_m^TV_m^{-1}u_m \) is the score statistic for testing \( H_0 : \theta_2 = 0 \). Under the alternative hypothesis, its asymptotic distribution is non-central chi-squared with \( rdf \) and non-centrality parameter

\[
\lambda_m = \mu_m^TV_m^{-1}\mu_m = \frac{1}{n}(L^m J_{11}^m \delta)^T(L^m J_{11}^m L^{mT})^{-1}(L^m J_{11}^m \delta).
\]

**Proposed Method 1 (PM1)**

Our first proposed method is to use MLE of \( \theta_1 \) using all individuals at the null model for the score function \( u(m)^{2}_{2i}(\hat{\theta}_1) \) instead of the MLE \( \hat{\theta}_1^m \) ignoring individuals whose genotype is missing. The MLE is denoted by \( \tilde{\theta}_1^m \). We call this method the proposed method 1 (PM1). Let

\[
\begin{align*}
\text{Proposed Method 1 (PM1)} & \\
\quad & \\
\text{Let } u_1(\theta_1) = u_1(\theta) \bigg|_{\theta=(\theta_1^T,0^T)^T} = \sum_{i=1}^n u_{1i}(Y_i, \theta) \bigg|_{\theta=(\theta_1^T,0^T)^T} \\
& = \sum_{i=1}^n \partial \log f_i(Y_i, \theta)/\partial \theta_2 \bigg|_{\theta=(\theta_1^T,0^T)^T}.
\end{align*}
\]
As in the analysis for CST, suppose that the null model correspond to the parameter \( \theta_0 = (\theta^T_{10}, 0^T)^T \), then

\[
u^m(Y) = \left( \begin{array}{c} u_1(Y) \\ u_2^m(Y) \end{array} \right) = \left( \begin{array}{c} u_1(\theta_{10}) \\ u_2^m(\theta_{10}) \end{array} \right).
\]

Using an analogous argument in the proof of Lemma 3.2 of Hjort and Claeskens, we have

\[
\sqrt{n}(\hat{\theta}^f_1 - \theta_{10}) = J_{11}[1, 1]^{-1}\sqrt{n}u_1(\theta_{10}) + o_p(1).
\]

Substituting the above equation, the score function is written as

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_{2i}^m(\hat{\theta}^f_1) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ u_{2i}^m(\theta_{10}) + \frac{\partial u_{2i}^m(\theta_{10})}{\partial \theta_1} (\hat{\theta}^f_1 - \theta_{10}) \right\} + o_p(1)
\]

\[
= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ u_{2i}^m(\theta_{10}) + \frac{1}{n} \sum_{i=1}^{n} \frac{\partial u_{2i}^m(\theta_{10})}{\partial \theta_1} (\hat{\theta}^f_1 - \theta_{10}) \right\} + o_p(1)
\]

\[
= \sqrt{n} \left\{ \frac{1}{n} u_{2i}^m(\theta_{10}) - \frac{1}{n} J_{11}^m[2, 1](\hat{\theta}^f_1 - \theta_{10}) \right\} + o_p(1)
\]

\[
= \frac{1}{\sqrt{n}} u_{2i}^m(\theta_{10}) - \frac{1}{n} J_{11}^m[2, 1] \sqrt{n}(\hat{\theta}^f_1 - \theta_{10}) + o_p(1)
\]

\[
= \frac{1}{\sqrt{n}} u_{2i}^m(\theta_{10}) - \frac{1}{n} J_{11}^m[2, 1] \{ J_{11}[1, 1]^{-1} \sqrt{n}u_1(\theta_{10}) + o_p(1) \} + o_p(1)
\]

\[
= \frac{1}{\sqrt{n}} u_{2i}^m(\theta_{10}) - J_{11}^m[2, 1] J_{11}[1, 1]^{-1} \frac{1}{\sqrt{n}} u_1(\theta_{10}) + o_p(1)
\]

\[
= \left( -J_{11}^m[2, 1] J_{11}[1, 1]^{-1} I \right) \frac{1}{\sqrt{n}} \left( \begin{array}{c} u_1(\theta_{10}) \\ u_2^m(\theta_{10}) \end{array} \right) + o_p(1)
\]

\[
= L^f \frac{1}{\sqrt{n}} u^m(Y) + o_p(1).
\]
where
\[
L^f = \left( -J_{11}^m[2,1]J_{11}[1,1]^{-1} \ 1 \right).
\]

Applying equation (7), under an alternative hypothesis (1), the distribution of the PM1 score function is
\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_{f}^m(\tilde{\theta}_{1}^f) \sim N_r(\frac{1}{n}L^fJ_{11}^f\delta, \frac{1}{n}L^fJ_{11}^fL^f)^T),
\]
where
\[
L^fJ_{11}^f\delta = \left( -J_{11}^m[2,1]J_{11}[1,1]^{-1} \ 1 \right) \left( J_{11}[1,1] \ J_{11}[1,2] \ J_{11}^m[2,1] \ J_{11}^m[2,2] \right) \delta
\]
\[
= \left( 0 \ -J_{11}^m[2,1]J_{11}[1,1]^{-1}J_{11}[1,2] + J_{11}^m[2,2] \right) \delta
\]
\[
= (-J_{11}^m[2,1]J_{11}[1,1]^{-1}J_{11}[1,2] + J_{11}^m[2,2])\delta_2,
\]
and
\[
L^fJ_{11}^fL^f = \left( -J_{11}^m[2,1]J_{11}[1,1]^{-1} \ 1 \right) \left( J_{11}[1,1] \ J_{11}^m[1,2] \ J_{11}[2,1] \ J_{11}^m[2,2] \right) \left( -J_{11}[1,1]^{-1}J_{11}^m[1,2] \ 1 \right)
\]
\[
= \left( 0 \ -J_{11}^m[2,1]J_{11}[1,1]^{-1}J_{11}^m[1,2] + J_{11}^m[2,2] \right) \left( -J_{11}[1,1]^{-1}J_{11}^m[1,2] \ 1 \right)
\]
\[
= -J_{11}^m[2,1]J_{11}[1,1]^{-1}J_{11}^m[1,2] + J_{11}^m[2,2].
\]

Let \( u_f = (1/\sqrt{n}) \sum_{i=1}^{n} u_{f}^m(\tilde{\theta}_{i}^f) \), \( \mu_f = \frac{1}{n}L^fJ_{11}^f\delta \), and \( V_f = \frac{1}{n}L^fJ_{11}^fL^f)^T). \)

Then, \( u_f^TV_f^{-1}u_f \) is the score statistics. Its asymptotic distribution is non-central.
chi-squared with \( r_d \) and non-centrality parameter

\[
\lambda_f = \mu_f^T V_f^{-1} \mu_f = \frac{1}{n} (L_f J_{11}^T \delta) ^T (L_f J_{11}^T L_f^T)^{-1} (L_f J_{11}^T \delta).
\]

**Proposed Method 2 (PM2)**

We propose another score test whose power is greater than PM1, as shown in what follows. Exploiting the approximation in (8), we define a modified score function \( u_2^*(\theta_1) \) by

\[
u_2^*(\theta_1) = u_2^{m1}(\theta_1) - J_{11}^m[2,1](\theta_1) J_{11}^m[1,1](\theta_1)^{-1} u_1^{m1}(\theta_1),
\]

where \( J_{11}^m[2,1](\theta) \) and \( J_{11}^m[1,1](\theta) \) are \(-\partial u_2^{m1}(\theta)/\partial \theta_1 \)/n and \(-\partial u_1^{m1}(\theta)/\partial \theta_1 \)/n, respectively.

Then, we propose using \( u_2^*(\hat{\theta}_1) \), which is expanded as

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_2^*(\hat{\theta}_1) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ u_2^{m1}(\theta_{10}) - J_{11}^m[2,1](\hat{\theta}_1) J_{11}^m[1,1](\hat{\theta}_1)^{-1} u_1^{m1}(\theta_{10}) \right\}
\]

\[
= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \frac{\partial u_2^{m1}(\theta_{10})}{\partial \theta_1} - J_{11}^m[2,1](\hat{\theta}_1) J_{11}^m[1,1](\hat{\theta}_1)^{-1} \frac{\partial u_1^{m1}(\theta_{10})}{\partial \theta_1} \right\} (\hat{\theta}_1 - \theta_{10}) + o_p(1)
\]

\[
= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ u_2^{m1}(\theta_{10}) - J_{11}^m[2,1](\hat{\theta}_1) J_{11}^m[1,1](\hat{\theta}_1)^{-1} u_1^{m1}(\theta_{10}) \right\}
\]

\[
+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \frac{\partial u_2^{m1}(\theta_{10})}{\partial \theta_1} - J_{11}^m[2,1](\hat{\theta}_1) J_{11}^m[1,1](\hat{\theta}_1)^{-1} \frac{\partial u_1^{m1}(\theta_{10})}{\partial \theta_1} \right\} (\hat{\theta}_1 - \theta_{10}) + o_p(1)
\]

\[
= K_1 + K_2 (\hat{\theta}_1 - \theta_{10}) + o_p(1).
\]
Hence, 

\[
K_2 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \frac{\partial u_{1i}^m(\theta_{1i})}{\partial \theta_1} - J_{11}^m[2, 1](\hat{\theta}_1^f)J_{11}^m[1, 1](\hat{\theta}_1^f)^{-1} \frac{\partial u_{1i}^m(\theta_{1i})}{\partial \theta_1} \right\}
\]

= \frac{1}{\sqrt{n}} \left[ -J_{11}^m[2, 1](\theta_{10}) - J_{11}^m[2, 1](\hat{\theta}_1^f)J_{11}^m[1, 1](\hat{\theta}_1^f)^{-1} \{ -J_{11}^m[1, 1](\theta_{10}) \} + o_p(n) \right]

= \frac{1}{\sqrt{n}} \left[ -J_{11}^m[2, 1](\theta_{10}) + \{ J_{11}^m[2, 1](\theta_{10}) + o_p(n) \} \{ J_{11}^m[1, 1](\theta_{10})^{-1} + o_p(1/n) \} J_{11}^m[1, 1](\theta_{10}) + o_p(n) \right]

= \frac{1}{\sqrt{n}} \left[ -J_{11}^m[2, 1](\theta_{10}) + J_{11}^m[2, 1](\theta_{10})J_{11}^m[1, 1](\theta_{10})^{-1} J_{11}^m[1, 1](\theta_{10}) + o_p(n) \right]

= \frac{1}{\sqrt{n}} o_p(n)

= o_p(\sqrt{n}).

Hence,

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_{2i}^{*}(\hat{\theta}_1^f) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ u_{2i}^m(\theta_{1i}) - J_{11}^m[2, 1](\hat{\theta}_1^f)J_{11}^m[1, 1](\hat{\theta}_1^f)^{-1} u_{1i}^m(\theta_{1i}) \right\} + o_p(\sqrt{n})(\hat{\theta}_1^f - \theta_{10}) + o_p(1)
\]

= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ u_{2i}^m(\theta_{1i}) - J_{11}^m[2, 1](\hat{\theta}_1^f)J_{11}^m[1, 1](\hat{\theta}_1^f)^{-1} u_{1i}^m(\theta_{1i}) \right\} + o_p(\sqrt{n})O_p(1/\sqrt{n}) + o_p(1)

= \frac{1}{\sqrt{n}} \left\{ u_{2i}^m(\theta_{1i}) - J_{11}^m[2, 1](\hat{\theta}_1^f)J_{11}^m[1, 1](\hat{\theta}_1^f)^{-1} u_{1i}^m(\theta_{1i}) \right\} + o_p(1)

= \frac{1}{\sqrt{n}} u_{2i}^m(\theta_{1i}) - J_{11}^m[2, 1](\hat{\theta}_1^f)J_{11}^m[1, 1](\hat{\theta}_1^f)^{-1} \frac{1}{\sqrt{n}} u_{1i}^m(\theta_{1i}) + o_p(1)

= \frac{1}{\sqrt{n}} u_{2i}^m(\theta_{1i}) - \{ J_{11}^m[2, 1](\theta_{1i}) + o_p(n) \} \{ J_{11}^m[1, 1](\theta_{1i})^{-1} + o_p(1/n) \} \frac{1}{\sqrt{n}} u_{1i}^m(\theta_{1i}) + o_p(1)

= \frac{1}{\sqrt{n}} u_{2i}^m(\theta_{1i}) - \{ J_{11}^m[2, 1](\theta_{1i}) J_{11}^m[1, 1](\theta_{1i})^{-1} + o_p(1) \} \frac{1}{\sqrt{n}} u_{1i}^m(\theta_{1i}) + o_p(1)

= \frac{1}{\sqrt{n}} u_{2i}^m(\theta_{1i}) - J_{11}^m[2, 1](\theta_{1i}) J_{11}^m[1, 1](\theta_{1i})^{-1} \frac{1}{\sqrt{n}} u_{1i}^m(\theta_{1i}) - o_p(1) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_{1i}^m(\theta_{1i}) + o_p(1)

= \frac{1}{\sqrt{n}} u_{2i}^m(\theta_{1i}) - J_{11}^m[2, 1](\theta_{1i}) J_{11}^m[1, 1](\theta_{1i})^{-1} \frac{1}{\sqrt{n}} u_{1i}^m(\theta_{1i}) + o_p(1). \]

Comparing (8) and (9), PM2 is asymptotically equivalent to CST.
Comparison of Power between the Conventional Score Test and the Proposed Method 1

Under the alternative hypothesis, the score function and its distribution of CST are

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_{2i}^{m}(\hat{\theta}_{1}^{m}) \sim N_{r}(L^{m} J_{11}^{m} \delta, L^{m} J_{11}^{m} L^{mT}),
\]

\[
L^{m} J_{11}^{m} \delta = (-J_{11}^{m}[2,1]J_{11}^{m}[1,1]^{-1}J_{11}^{m}[1,2] + J_{11}^{m}[2,2])\delta_{2},
\]

\[
L^{m} J_{11}^{m} L^{mT} = -J_{11}^{m}[2,1]J_{11}^{m}[1,1]^{-1}J_{11}^{m}[1,2] + J_{11}^{m}[2,2],
\]

and that of PM1 are

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_{2i}^{m}(\hat{\theta}_{1}^{f}) \sim N_{r}(L^{f} J_{11}^{f} \delta, L^{f} J_{11}^{f} L^{fT}),
\]

\[
L^{f} J_{11}^{f} \delta = (-J_{11}^{m}[2,1]J_{11}^{f}[1,1]^{-1}J_{11}^{m}[1,2] + J_{11}^{m}[2,2])\delta_{2},
\]

\[
L^{f} J_{11}^{f} L^{fT} = -J_{11}^{m}[2,1]J_{11}^{f}[1,1]^{-1}J_{11}^{m}[1,2] + J_{11}^{m}[2,2].
\]

First, we compare the mean of CST and that of PM1. Recall that \(I_{i}\) is a binary indicator of whether the genotype \(G_{i}\) is observed or not. We make the following assumptions: (i) \(I_{i}\) independently and identically follows a binomial distribution \(I_{i} \sim \text{Bin}(1, 1 - R)\) for \(i = 1, \cdots, n\), where \(R\) is a probability of random missing. Then expected value of \(I_{i}\) is \(E(I_{i}) = 1 - R\); (ii) \(I_{i}\) is independent of a differential score function \(\partial u_{i}(\theta)/\partial \theta\); (iii) \((1/n) \sum_{i=1}^{n} \partial u_{i}(\theta)/\partial \theta = O(1)\).

We consider the quantity in the mean of CST,

\[
J_{11}^{m}[1,1] = -\frac{1}{n} \frac{\partial u_{i}^{m}(\theta)}{\partial \theta_{1}} = -\frac{1}{n} \sum_{i=1}^{n} \frac{\partial u_{i}^{m}(\theta)}{\partial \theta_{1}} \mathcal{I}_{i}
\]

\[
= -\frac{1}{n} \sum_{i=1}^{n} \frac{\partial u_{i}(\theta)}{\partial \theta_{1}} I_{i}.
\]
Here, the expected value and variance of \( \frac{1}{n} \sum_{i=1}^{n} \frac{\partial u_{i_1}(\theta)}{\partial \theta_1} I_i \) with respect to \( I_i \) are

\[
E \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{\partial u_{i_1}(\theta)}{\partial \theta_1} I_i \right\} = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial u_{i_1}(\theta)}{\partial \theta_1} E(I_i) = (1 - R) \cdot \frac{1}{n} \sum_{i=1}^{n} \frac{\partial u_{i_1}(\theta_1)}{\partial \theta_1}
\]

and

\[
\text{var} \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{\partial u_{i_1}(\theta)}{\partial \theta_1} I_i \right\} = \sum_{i=1}^{n} \left( \frac{1}{n} \frac{\partial u_{i_1}(\theta)}{\partial \theta_1} \right)^2 \text{var}(I_i) = R(1 - R) \cdot \frac{1}{n^2} \sum_{i=1}^{n} \left( \frac{\partial u_{i_1}(\theta)}{\partial \theta_1} \right)^2.
\]

From the assumption (iii),

\[
\text{var} \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{\partial u_{i_1}(\theta)}{\partial \theta_1} I_i \right\} = O(1/n).
\]

By Markov’s inequality, for any \( \epsilon > 0 \),

\[
P_r \left( \left| \frac{1}{n} \sum_{i=1}^{n} \frac{\partial u_{i_1}(\theta)}{\partial \theta_1} I_i - \frac{1 - R}{n} \sum_{i=1}^{n} \frac{\partial u_{i_1}(\theta)}{\partial \theta_1} \right| > \epsilon \right) = P_r \left( \left| \frac{1}{n} \sum_{i=1}^{n} \frac{\partial u_{i_1}(\theta)}{\partial \theta_1} I_i - \frac{1 - R}{n} \sum_{i=1}^{n} \frac{\partial u_{i_1}(\theta)}{\partial \theta_1} \right|^2 > \epsilon^2 \right) \leq E \left\{ \left| \frac{1}{n} \sum_{i=1}^{n} \frac{\partial u_{i_1}(\theta)}{\partial \theta_1} I_i - \frac{1 - R}{n} \sum_{i=1}^{n} \frac{\partial u_{i_1}(\theta)}{\partial \theta_1} \right|^2 \right\} = \frac{\text{var} \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{\partial u_{i_1}(\theta)}{\partial \theta_1} I_i \right\}}{\epsilon^2} = O(1/n) \rightarrow 0.
\]
Thus, by the law of large numbers,

\[
J^m_{11}[1, 1] = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial u_{11}(\theta)}{\partial \theta_1} I_i
\]

\[
\mathbb{E}_p \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{\partial u_{11}(\theta)}{\partial \theta_1} I_i \right\}
\]

\[
= (1 - R) \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{\partial u_{11}(\theta)}{\partial \theta_1} \right\}
\]

\[
= (1 - R) J_{11}[1, 1].
\]

(10)

Similarly, \(J^m_{11}[1, 2]\) is

\[
J^m_{11}[1, 2] \xrightarrow{P} (1 - R) J_{11}[1, 2].
\]

(11)

From (10) and (11), it follows that

\[
J^m_{11}[1, 1]^{-1} J^m_{11}[1, 2] \xrightarrow{P} \frac{1}{1 - R} J_{11}[1, 1]^{-1} (1 - R) J_{11}[1, 2]
\]

\[
= J_{11}[1, 1]^{-1} J_{11}[1, 2].
\]

(12)

Hence, from (12), the mean of CST, \(L^m J^m_{11} \delta\), is asymptotically equivalent to that of PM1, \(L^f J^f_{11} \delta\).

Second, we compare the variance of CST and that of PM1. To this end, we may only compare \(J^m_{11}[1, 1]\) and \(J_{11}[1, 1]\). The former is

\[
J^m_{11}[1, 1] = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial u_{11}(\theta)}{\partial \theta_1} I_i
\]

and the letter is

\[
J_{11}[1, 1] = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial u_{11}(\theta)}{\partial \theta_1}.
\]
Assuming positive-semidefiniteness of $\partial u_i(\theta)/\partial \theta_1$ for all $i$, we deduce that $J_{11}[1,1] \geq J_{11}^m[1,1]$, in which we denote that, for two matrices $A$ and $B$, $A \succeq B$ if $A - B$ is positive-semidefinite. Thus, $-J_{11}[1,1]^{-1} \geq -J_{11}^m[1,1]^{-1}$. Consequently, the variance of PM1 $L^m J_{11}^m L^m^T$, is bigger than that of CST, $L^T J_{11}^m L^T$.

If a score function $u$ follows the normal distribution with mean $\mu$ and variance $V$, the quadratic form $u^T V^{-1} u$ (test statistic) follows the non-central chi-square distribution, where the non-centrality parameter is $\mu^T V^{-1} \mu$. The power of the test statistic increases as the non-centrality parameter increases. In the above arguments, we have shown that the mean of CST score function is asymptotically equivalent to that of PM1 score function while the variance of PM1 score function is bigger than the variance of CST score function. That is, the magnitude of non-centrality parameter is dominated only by the magnitude of variance, and the non-centrality parameter of PM1 is smaller than that of CST. Therefore, the power of PM1 is smaller than that of CST.

References

[1] Hjort NL, Claeskens G. Frequentist model average estimators. Journal of the American Statistical Association. 2003;98(464):879–899.