On the Null Distribution of Bayes Factors in Linear Regression (Supplementary)

Quan Zhou and Yongtao Guan
Baylor College of Medicine

April 28, 2017

1 Proofs of theorems

Under the null model with the NIG prior, the conditional distribution of $\boldsymbol{y}$ given τ is a multivariate normal with mean $\boldsymbol{0}$ and covariance matrix $\tau^{-1}(\boldsymbol{W}\boldsymbol{V}_a\boldsymbol{W}^t + \boldsymbol{I}_n)$. We integrate out τ to obtain Bayes factor

$$\mathrm{BF} = |\boldsymbol{I}_n + (\boldsymbol{W}\boldsymbol{V}_a\boldsymbol{W}^t + \boldsymbol{I}_n)^{-1}\boldsymbol{L}\boldsymbol{V}_b\boldsymbol{L}^t|^{-1/2} \left\{ \frac{\kappa_2 + \boldsymbol{y}^t(\boldsymbol{W}\boldsymbol{V}_a\boldsymbol{W}^t + \boldsymbol{L}\boldsymbol{V}_b\boldsymbol{L}^t + \boldsymbol{I}_n)^{-1}\boldsymbol{y})}{\kappa_2 + \boldsymbol{y}^t(\boldsymbol{W}\boldsymbol{V}_a\boldsymbol{W}^t + \boldsymbol{I}_n)^{-1}\boldsymbol{y}} \right\}^{-(n+\kappa_1)/2}.$$

Define the projection matrix $\boldsymbol{P} = \boldsymbol{I}_n - \boldsymbol{W}(\boldsymbol{W}^t\boldsymbol{W})^{-1}\boldsymbol{W}^t$. By Woodbury identity and the idempotence of $\boldsymbol{P}$, as $\boldsymbol{V}_a^{-1}$ vanishes,

$$\begin{aligned} \lim_{\boldsymbol{V}_a^{-1}\to\boldsymbol{0}} (\boldsymbol{W}\boldsymbol{V}_a\boldsymbol{W}^t + \boldsymbol{I}_n)^{-1} &= \boldsymbol{P}, \\ \lim_{\boldsymbol{V}_a^{-1}\to\boldsymbol{0}} (\boldsymbol{W}\boldsymbol{V}_a\boldsymbol{W}^t + \boldsymbol{L}\boldsymbol{V}_b\boldsymbol{L}^t + \boldsymbol{I}_n)^{-1} &= \boldsymbol{P} - \boldsymbol{X}(\boldsymbol{X}^t\boldsymbol{X} + \boldsymbol{V}_b^{-1})^{-1}\boldsymbol{X}^t, \end{aligned} \tag{A1}$$

where $\boldsymbol{X} = \boldsymbol{P}\boldsymbol{L}$. By Sylvester's determinant identity and the idempotence of $\boldsymbol{P}$, $|\boldsymbol{I}_n + \boldsymbol{P}\boldsymbol{L}\boldsymbol{V}_b\boldsymbol{L}^t|^{-1/2} = |\boldsymbol{I}_p + \boldsymbol{X}^t\boldsymbol{X}\boldsymbol{V}_b|^{-1/2}$. Now letting $\kappa_1, \kappa_2 \to 0$, we obtain (3) in the main text. The easier case of known error variance can be solved by direct computation using (A1).

Let $\boldsymbol{H}_F$ be the hat matrix for the ordinary least square estimate of $\boldsymbol{b}$, and $\boldsymbol{H}_B$ be the hat matrix for the posterior mean for $\boldsymbol{b}$,

$$\boldsymbol{H}_F \stackrel{\text{def}}{=} \boldsymbol{X}(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t, \quad \boldsymbol{H}_B \stackrel{\text{def}}{=} \boldsymbol{X}(\boldsymbol{X}^t\boldsymbol{X} + \boldsymbol{V}_b^{-1})^{-1}\boldsymbol{X}^t. \tag{A2}$$

Then $\mathrm{BF} = |\boldsymbol{V}_b|^{-1/2}|\boldsymbol{X}^t\boldsymbol{X} + \boldsymbol{V}_b^{-1}|^{-1/2}\{1 + \boldsymbol{y}^t\boldsymbol{H}_B\boldsymbol{y}/\boldsymbol{y}^t(\boldsymbol{P} - \boldsymbol{H}_B)\boldsymbol{y}\}^{n/2}$. Define two components in the expression so that $BF = T \cdot R$, with

$$2\log T \stackrel{\text{def}}{=} -\log|\boldsymbol{V}_b\boldsymbol{X}^t\boldsymbol{X} + \boldsymbol{I}_p|, \quad 2\log R \stackrel{\text{def}}{=} n\log\left\{1 + \frac{\boldsymbol{y}^t\boldsymbol{H}_B\boldsymbol{y}}{\boldsymbol{y}^t(\boldsymbol{P} - \boldsymbol{H}_B)\boldsymbol{y}}\right\}. \tag{A3}$$

Since only R depends on $\boldsymbol{y}$, it will be our focus in the proof of the theorems. For comparison, assuming $\boldsymbol{X}$ has full rank, the likelihood ratio test statistic is

$$2\log L = 2\log\frac{\sup_{\boldsymbol{a},\boldsymbol{b},\tau} p(\boldsymbol{y}|H_1,\boldsymbol{a},\boldsymbol{b},\tau)}{\sup_{\boldsymbol{a},\tau} p(\boldsymbol{y}|H_0,\boldsymbol{a},\tau)} = n\log\left\{1+\frac{\boldsymbol{y}^t\boldsymbol{H}_F\boldsymbol{y}}{\boldsymbol{y}^t(\boldsymbol{P}-\boldsymbol{H}_F)\boldsymbol{y}}\right\}. \tag{A4}$$

Lemma 1. *Recall p is the number of columns of $\boldsymbol{L}$ and $\boldsymbol{X}$, q the number of columns of $\boldsymbol{W}$, and n the number of samples. Without loss of generality, assume $\boldsymbol{H}_B$ defined in* (A2) *has rank p. $\boldsymbol{H}_B$ is positive semi-definite and its spectral decomposition is*

$$\boldsymbol{H}_B = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^t = \sum_{i=1}^{p}\lambda_i\boldsymbol{u}_i\boldsymbol{u}_i^t$$

where $\boldsymbol{\Sigma} = Diag(\lambda_1,\ldots,\lambda_p,0,\ldots,0)$ with $\lambda_1\geq\cdots\geq\lambda_p$, and $\boldsymbol{U}=(\boldsymbol{u}_1,\ldots,\boldsymbol{u}_n)$ where $\boldsymbol{u}_i$ is the ith eigenvector. Then, for some orthonormal vectors $\boldsymbol{k}_1,\ldots,\boldsymbol{k}_{n-p-q}$,

(a) (spectral decomposition) $\boldsymbol{P}-\boldsymbol{H}_B = \sum_{i=1}^{p}(1-\lambda_i)\boldsymbol{u}_i\boldsymbol{u}_i^t + \sum_{j=1}^{n-p-q}\boldsymbol{k}_j\boldsymbol{k}_j^t$;

(b) (spectral decompositions) $\boldsymbol{H}_F = \sum_{i=1}^{p}\boldsymbol{u}_i\boldsymbol{u}_i^t$ and $\boldsymbol{P}-\boldsymbol{H}_F = \sum_{j=1}^{n-p-q}\boldsymbol{k}_j\boldsymbol{k}_j^t$;

(c) for T defined in (A3), *we have $2\log T = \sum_{i=1}^{p}\log(1-\lambda_i)$.*

Proof. To prove (a), first since $\boldsymbol{P}$ is an orthogonal projection matrix, it has $n-q$ eigenvalues equal to 1 and q zero eigenvalues. Therefore, we only need to show that $\boldsymbol{u}_i,\ldots,\boldsymbol{u}_p$ and $\boldsymbol{k}_1,\ldots,\boldsymbol{k}_{n-p-q}$ are eigenvectors for both $\boldsymbol{P}$ and $\boldsymbol{H}_B$. $\boldsymbol{H}_B$ starts with $\boldsymbol{X}=\boldsymbol{P}\boldsymbol{L}$, which combines $\boldsymbol{H}_B\boldsymbol{u}_i=\lambda_i\boldsymbol{u}_i$ to give $\boldsymbol{P}\boldsymbol{u}_i=\boldsymbol{u}_i$. Now let $\boldsymbol{k}_1,\ldots,\boldsymbol{k}_{n-p-q}$ be the rest $n-p-q$ eigenvectors of $\boldsymbol{P}$ with unit eigenvalues. They are orthogonal to every $\boldsymbol{u}_i$ and thus must lie in the null space of $\boldsymbol{H}_B$, i.e., $\boldsymbol{H}_B\boldsymbol{k}_i=0$. To prove (b), notice that $\boldsymbol{H}_F$ is an orthogonal projection matrix, and $\boldsymbol{H}_F$ and $\boldsymbol{H}_B$ have the same null space. Hence $(\boldsymbol{u}_i,\ldots,\boldsymbol{u}_p,\boldsymbol{k}_1,\ldots,\boldsymbol{k}_{n-p-q})$ is also a set of eigenvectors for $\boldsymbol{H}_F$. To prove (c), notice that λ_i is also an eigenvalue of the matrix $(\boldsymbol{X}^t\boldsymbol{X}+\boldsymbol{V}_b^{-1})^{-1}\boldsymbol{X}^t\boldsymbol{X}$ with the corresponding eigenvector $(\boldsymbol{X}^t\boldsymbol{X}+\boldsymbol{V}_b^{-1})^{-1}\boldsymbol{X}^t\boldsymbol{u}_i$. With simple matrix computation one can show that the eigenvalues of $\boldsymbol{V}_b(\boldsymbol{X}^t\boldsymbol{X})$ are $\lambda_i/(1-\lambda_i)$ ($i=1,\ldots,p$), which proves (c). Lastly, these eigenvalues must be non-negative since they are also the eigenvalues of $(\boldsymbol{X}^t\boldsymbol{X})^{1/2}\boldsymbol{V}_b(\boldsymbol{X}^t\boldsymbol{X})^{1/2}$, a positive semi-definite matrix. So for $i=1,\ldots,p$, $\lambda_i\in[0,1)$ and $\lambda_i\in(0,1)$ if $\boldsymbol{X}$ has full rank. □

1.1 Proof of Theorem 1

Proof. We use $o_P(1)$ to denote a term converging to 0 in probability and $O_P(1)$ to denote a term that is stochastically bounded. Let $\boldsymbol{z}=\tau^{1/2}(\boldsymbol{y}-\boldsymbol{W}\boldsymbol{a})$ as defined in the main text. Here $\boldsymbol{a}$ and τ can be interpreted as either the true values (from a Frequentist's perspective) or the values that generate $\boldsymbol{y}$ (from a Bayesian perspective). Lemma 1(c) gives the expression for $2\log T$. So we only need to deal with $2\log R$. Since $\boldsymbol{P}\boldsymbol{W}=\boldsymbol{0}$,

$2\log R$ can always be expressed using $\boldsymbol{z}$,

$$2\log R = n\log\left\{1 + \frac{\boldsymbol{z}^t\boldsymbol{H}_B\boldsymbol{z}}{\boldsymbol{z}^t(\boldsymbol{P}-\boldsymbol{H}_B)\boldsymbol{z}}\right\}. \tag{A5}$$

We shall prove a more general result without assuming "under the null":

If $\boldsymbol{z}^t\boldsymbol{H}_F\boldsymbol{z} = O_P(1)$, *then* $2\log R = \boldsymbol{z}^t\boldsymbol{H}_B\boldsymbol{z} + o_P(1)$.

Theorem 1 is then a special case of the above claim. Later we will also use this claim to handle the alternative. The conditional distribution of $\boldsymbol{z}$ given $\boldsymbol{b}$ is $\mathrm{MVN}(\tau^{1/2}\boldsymbol{L}\boldsymbol{b}, \boldsymbol{I}_n)$. Apply the spectral decomposition results given in Lemma 1(a) to get $\boldsymbol{z}^t\boldsymbol{H}_B\boldsymbol{z} = \sum_{i=1}^p \lambda_i(\boldsymbol{u}_i^t\boldsymbol{z})^2$, and $\boldsymbol{z}^t(\boldsymbol{P}-\boldsymbol{H}_B)\boldsymbol{z} = \sum_{j=1}^{n-p-q}(\boldsymbol{k}_j^t\boldsymbol{z})^2 + \sum_{i=1}^p(1-\lambda_i)(\boldsymbol{u}_i^t\boldsymbol{z})^2$. By Lemma 1(a), we have $\boldsymbol{P}\boldsymbol{k}_j = \boldsymbol{k}_j$ and $\boldsymbol{H}_B\boldsymbol{k}_j = \boldsymbol{0}$, which jointly imply that $\boldsymbol{k}^t\boldsymbol{L} = \boldsymbol{0}$. Thus we can define $Q_0 = \sum_{j=1}^{n-p-q}(\boldsymbol{k}_j^t\boldsymbol{z})^2 \sim \chi^2_{n-p-q}$. Recall by Lemma 1(b), $\boldsymbol{z}^t\boldsymbol{H}_F\boldsymbol{z} = \sum_{i=1}^p(\boldsymbol{u}_i^t\boldsymbol{z})^2$ and $\lambda_i \in [0,1)$. Hence if $\boldsymbol{z}^t\boldsymbol{H}_F\boldsymbol{z} = O_P(1)$, we have $\boldsymbol{z}^t\boldsymbol{H}_B\boldsymbol{z} = O_P(1)$ and $\boldsymbol{z}^t(\boldsymbol{P}-\boldsymbol{H}_B)\boldsymbol{z} = Q_0 + O_P(1)$. Direct calculations show that $n^{-1}\boldsymbol{z}^t(\boldsymbol{P}-\boldsymbol{H}_B)\boldsymbol{z}$ converge to 1 in probability at rate n^{-1}. By the continuous mapping theorem, its reciprocal also converges to 1 in probability and thus $(\boldsymbol{z}^t(\boldsymbol{P}-\boldsymbol{H}_B)\boldsymbol{z}/n)^{-1} = 1 + o_P(1)$. Now apply Slutsky's Theorem and Taylor expansion to rewrite (A5) into $2\log R = n\log\{1 + n^{-1}\boldsymbol{z}^t\boldsymbol{H}_B\boldsymbol{z} + o_P(1)\} = \boldsymbol{z}^t\boldsymbol{H}_B\boldsymbol{z} + o_P(1)$. Lastly, define $Q_i = (\boldsymbol{u}_i^t\boldsymbol{z})^2$ and we have $2\log R = \sum_{i=1}^p \lambda_i Q_i + o_P(1)$. This finishes the proof of our general result (when the error variance is known, this claim holds trivially). Under the null, $\boldsymbol{z}$ is an n-dimensional standardized normal variable. Therefore, $Q_i \sim \chi_1^2$ and $\boldsymbol{z}^t\boldsymbol{H}_F\boldsymbol{z} = \sum_{i=1}^p Q_i \sim \chi_p^2$, which implies $\boldsymbol{z}^t\boldsymbol{H}_F\boldsymbol{z} = O_P(1)$. □

Consider the two special cases of $\boldsymbol{V}_b$ given in (6) and (7) in the main text. For the independent prior, $\boldsymbol{V}_b = \sigma_b^2\boldsymbol{I}_p$, we have $\lambda_i = \delta_i^2/(\delta_i^2 + \sigma_b^{-2})$ where δ_i is the ith (sorted) singular value of $\boldsymbol{X}$. For g-prior, $\boldsymbol{V}_b = g(\boldsymbol{X}^t\boldsymbol{X})^{-1}$, we have $\lambda_i = g/(g+1)$, for $i = 1, \ldots, p$.

1.2 Proof for Corollary 1

Proof. From Theorem 1, the likelihood ratio defined in (A4) becomes $2\log L = \sum_{i=1}^p Q_i + o_P(1)$, which is a special case of Wilks' Theorem [Wilks, 1938]. For $p = 1$, we have $\log \mathrm{BF} = \lambda_1 \log L + (\log(1-\lambda_1))/2 + o_P(1)$, which implies $p_B \approx p_F$. In addition, $2\log\mathrm{BF}$ can be expressed using the F statistic ($F = (n-1-q)Q_1/Q_0$), $2\log\mathrm{BF} = n\log[1 + F\lambda_1/\{F(1-\lambda_1) + (n-1-q)\}] + \log(1-\lambda_1)$. Thus the p-value of F-test exactly equals the p-value for BF. □

1.3 Proof of Proposition 1

Proof. By the definition of sBF we have $\mathbb{E}_0[\log \mathrm{sBF}] = \mathbb{E}_0[\,\log\mathrm{BF} - \mathbb{E}_0[\log\mathrm{BF}]\,] = \mathbb{E}_0[\log\mathrm{BF}] - \mathbb{E}_0[\log\mathrm{BF}] = 0$. Also by definition $\mathrm{sBF}/\mathrm{BF} = \prod_{i=1}^p \exp(-\lambda_i/2)(1-\lambda_i)^{-1/2}$. Because $\log(1-\lambda_i) < -\lambda_i$, $\exp(-\lambda_i/2)(1-\lambda_i)^{-1/2} > 1$ and therefore $\log\mathrm{sBF} > \log\mathrm{BF}$. These prove (a). Because the scaling factor (sBF/BF) is independent of $\boldsymbol{y}$, BF and sBF have the same Bayesian p-value and this proves (b). Since permutation simulates the null, $\log\mathrm{sBF} =$

$\mathbb{E}_P[\log \mathrm{BF}(\boldsymbol{y}) - \log \mathrm{BF}(\tilde{\boldsymbol{y}})]$ where the expectation is with respect to the permutation. By Jensen's inequality, $\mathrm{sBF} < \mathbb{E}_P[\mathrm{BF}(\boldsymbol{y})/\mathrm{BF}(\tilde{\boldsymbol{y}})]$, which proves (c). $\square$

1.4 Proof of Proposition 3

Proof. Since $\mathbb{E}_0[\log \mathrm{sBF}_j] = 0$ for $j = 1, 2$ by Proposition 2(a). We have $\mathbb{E}_0[\log \mathrm{sBF}_1 - \log \mathrm{sBF}_2] = 0$. Because Proposition 2(a) applies to arbitrary dimension, $\mathbb{E}_0[\log \mathrm{sBF}_1 - \log \mathrm{sBF}_2] = 0$ in fact holds for arbitrary dimension. Due to lack of consistent way to compare informativeness of covariates of more than one dimension, we focus on one dimension situation of BF. By calculating the first derivative with respect to c_j, it's straightforward to show that $\mathbb{E}_0[2\log \mathrm{BF}_j] = c_j + \log(1 - c_j)$ is monotone decreasing in $c_j \in (0, 1)$ (we use c_1, c_2 to denote the values of λ_1 of the two tests). If $c_1 > c_2$, we have $\mathbb{E}_0[\log \mathrm{BF}_1] < \mathbb{E}_0[\log \mathrm{BF}_2]$. $\square$

1.5 Proof of Theorem 4

Proof. Consider a sequence local alternatives $\boldsymbol{b} = \boldsymbol{\beta}/\sqrt{n\tau}$. Then we have $\boldsymbol{z} = \tau^{1/2}(\boldsymbol{y} - \boldsymbol{W}\boldsymbol{a}) \sim \mathrm{MVN}(\boldsymbol{L}\boldsymbol{\beta}/\sqrt{n}, \boldsymbol{I}_n)$. According to the proof of Theorem 1, we only need to show $\boldsymbol{z}^t\boldsymbol{H}_F\boldsymbol{z} = O_P(1)$. We use $\chi^2_p(\rho)$ to denote a noncentral chi-squared random variable with $d.f. = p$ and the noncentrality parameter ρ. By Lemma 1(a),

$$\boldsymbol{z}^t\boldsymbol{H}_F\boldsymbol{z} = \sum_{i=1}^{p}(\boldsymbol{u}_i^t\boldsymbol{z})^2 \sim \chi^2_p(\frac{1}{n}\sum_{i=1}^{p}(\boldsymbol{u}_i^t\boldsymbol{L}\boldsymbol{\beta})^2) \tag{A6}$$

Since for $i = 1, \ldots, p$, $\boldsymbol{u}_i^t\boldsymbol{P} = \boldsymbol{u}_i^t$ and for $i = p+1, \ldots, n$, $\boldsymbol{u}_i^t\boldsymbol{P} = \boldsymbol{0}$, $\sum_{i=1}^{p}(\boldsymbol{u}_i^t\boldsymbol{L}\boldsymbol{\beta})^2 = \sum_{i=1}^{p}(\boldsymbol{u}_i^t\boldsymbol{P}\boldsymbol{L}\boldsymbol{\beta})^2 = \sum_{i=1}^{n}(\boldsymbol{u}_i^t\boldsymbol{P}\boldsymbol{L}\boldsymbol{\beta})^2$. Since $\boldsymbol{u}_i$ is the singular vector of $\boldsymbol{X}$, $\sum_{i=1}^{n}(\boldsymbol{u}_i^t\boldsymbol{P}\boldsymbol{L}\boldsymbol{\beta})^2 = (\boldsymbol{L}\boldsymbol{\beta})^t\boldsymbol{P}(\boldsymbol{L}\boldsymbol{\beta}) = (\boldsymbol{X}\boldsymbol{\beta})^t(\boldsymbol{X}\boldsymbol{\beta})$. Because a weakly convergent random sequence is stochastically bounded [Shao, 2003, p127], to show $\boldsymbol{z}^t\boldsymbol{H}_F\boldsymbol{z} = O_P(1)$ we only need to show that the noncentrality parameter in (A6), $n^{-1}(\boldsymbol{X}\boldsymbol{\beta})^t(\boldsymbol{X}\boldsymbol{\beta})$, either converges to a constant or is bounded. Due to projection, $(\boldsymbol{X}\boldsymbol{\beta})^t(\boldsymbol{X}\boldsymbol{\beta}) \leq (\boldsymbol{L}\boldsymbol{\beta})^t\boldsymbol{L}\boldsymbol{\beta}$ and thus it is sufficient to show $n^{-1}\boldsymbol{\beta}^t\boldsymbol{L}^t\boldsymbol{L}\boldsymbol{\beta}$ converges or is bounded, which are the two assumptions in the theorem. Thus, we have $2\log R = \boldsymbol{z}^t\boldsymbol{H}_B\boldsymbol{z} + o_P(1)$. The distribution of the quadratic form $\boldsymbol{z}^t\boldsymbol{H}_B\boldsymbol{z}$ is $\boldsymbol{z}^t\boldsymbol{H}_B\boldsymbol{z} = \sum_{i=1}^{p}\lambda_i(\boldsymbol{u}_i^t\boldsymbol{z})^2 = \sum_{i=1}^{p}\lambda_i Q_i$ where $Q_i \overset{ind.}{\sim} \chi^2_1(n^{-1}(\boldsymbol{u}_i^t\boldsymbol{L}\boldsymbol{\beta})^2)$. $\square$

1.6 Proof of Proposition 5

Proof. Part (a) can be directly derived from Theorem 4. For simple linear regression, the noncentrality parameter of Q_1 is $\sigma_b^2\delta_1^2 = 1/(\lambda_1^{-1} - 1)$. We prove part (b) for the multi-linear regression. The expression of Q_i is $Q_i = (\boldsymbol{u}_i^t\boldsymbol{z})^2$ where $\boldsymbol{z} = \tau^{1/2}(\boldsymbol{y} - \boldsymbol{W}\boldsymbol{a})$. Under the alternative $\boldsymbol{b} \sim \mathrm{MVN}(\boldsymbol{0}, \tau^{-1}\boldsymbol{V}_b)$, we have $\boldsymbol{z} \sim \mathrm{MVN}(\boldsymbol{0}, \boldsymbol{L}\boldsymbol{V}_b\boldsymbol{L}^t + \boldsymbol{I})$. Using $\boldsymbol{P}\boldsymbol{u}_i = \boldsymbol{u}_i$ and $\boldsymbol{X} = \boldsymbol{P}\boldsymbol{L}$, we obtain $\boldsymbol{u}_i^t\boldsymbol{z} \sim \mathrm{MVN}(\boldsymbol{0}, \boldsymbol{u}_i^t\boldsymbol{X}\boldsymbol{V}_b\boldsymbol{X}^t\boldsymbol{u}_i + 1)$. So Q_i is a scaled central chi-squared variable with $d.f. = 1$ and $\mathbb{E}[Q_i] = \boldsymbol{u}_i^t\boldsymbol{X}\boldsymbol{V}_b\boldsymbol{X}^t\boldsymbol{u}_i + 1$. Consider the two special cases. First under the independent prior $\boldsymbol{V}_b = \sigma_b^2 \boldsymbol{I}$, $\mathbb{E}[Q_i] = \sigma_b^2 \delta_i^2 + 1 = 1/(1-\lambda_i)$. Second, under the g-prior $\boldsymbol{V}_b = g(\boldsymbol{X}^t \boldsymbol{X})^{-1}$, we have $\mathbb{E}[Q_i] = g + 1 = 1/(1-\lambda_i)$ as well. □

References

J. Shao. *Mathematical Statistics*. Springer Texts in Statistics. Springer, second edition, 2003. ISBN 9780387953823.

S. S. Wilks. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, 9(1):60–62, 1938.

2 R code computing BF and sBF

```
# Function sbf takes 4 arguments and returns a pair of log(BF) and log(sBF).
# W is an n-by-q matrix; L is an n-by-p matrix; y is an n-vector; sigma_b is a scalar.
sbf <- function(W, L, y, sigma_b){
    n = nrow(W)
    q = ncol(W)
    p = ncol(L)
    PW = diag(n) - W %*% solve(t(W) %*% W) %*% t(W)
    X = PW %*% L
    HB = X %*% solve(t(X) %*% X + diag(1/sigma_b/sigma_b,p)) %*% t(X)
    log.R = -0.5*n*log(1 - (t(y) %*% HB %*% y) / (t(y) %*% PW %*% y ))
    delta = svd(X)$d
    lambda = delta^2 / (delta^2 + 1/sigma_b/sigma_b)
    log.T = sum(log(1-lambda))/2
    # one can check:
    # log.T = -0.5*log(det( t(X) %*% X * sigma_b * sigma_b + diag(p)))
    log.bf = log.T + log.R
    log.sbf = log.R - 0.5 * sum(lambda)
    return(c(log.bf,log.sbf))
}

## an example
sigma_b = 0.2
W = matrix(rexp(3000),1000,3)
L = matrix(rnorm(5000),1000,1)
y = rnorm(1000) + L %*% rnorm(1,0,0.1)
sbf(W,L,y,sigma_b)
```

3 Top 20 single SNP associations

SNP	Chr	Pos	MAF	$\log_{10}$ BF	$\log_{10}$ sBF	$-\log_{10} p_B$
rs12120962	1	10.53	0.384	3.549 (6)	4.624 (5)	5.628 (5)
rs12127400	1	10.54	0.384	3.271 (9)	4.346 (9)	5.339 (9)
rs4656461	1	163.95	0.140	5.494 (2)	6.424 (2)	7.507 (2)
rs7411708	1	163.99	0.428	3.360 (8)	4.438 (7)	5.434 (8)
rs10918276	1	163.99	0.427	3.258 (10)	4.336 (10)	5.328 (10)
rs7518099	1	164.00	0.140	5.829 (1)	6.758 (1)	7.852 (1)
rs972237	2	125.89	0.119	2.781 (14)	3.674 (17)	4.649 (18)
rs2728034	3	2.72	0.090	3.584 (5)	4.434 (8)	5.452 (7)
rs7645716	3	46.31	0.254	3.019 (12)	4.045 (11)	5.027 (11)
rs7696626	4	8.73	0.023	3.069 (11)	3.644 (18)	4.698 (16)
rs2025751	6	51.73	0.466	3.443 (7)	4.527 (6)	5.527 (6)
rs1081076	6	132.97	0.022	2.690 (19)	3.260 (44)	4.287 (36)
rs10757601	9	26.18	0.443	2.748 (15)	3.829 (13)	4.797 (13)
rs10778292	12	102.78	0.140	3.738 (4)	4.665 (4)	5.683 (4)
rs2576969	12	102.80	0.271	2.745 (16)	3.777 (14)	4.745 (14)
rs17034938	12	102.85	0.127	2.953 (13)	3.862 (12)	4.845 (12)
rs1955511	14	32.30	0.076	2.684 (20)	3.497 (25)	4.472 (25)
rs12150284	17	9.97	0.353	4.638 (3)	5.707 (3)	6.752 (3)
rs6017819	20	44.48	0.069	2.736 (17)	3.524 (24)	4.505 (23)
rs279728	20	44.51	0.087	2.704 (18)	3.541 (22)	4.515 (22)

Table S1: The "Pos" column gives the genomic position in megabase pair (reference HG18). The rankings by the three statistics are given in the parentheses. SNP IDs are in bold if they are mentioned specifically in the main text. Results were obtained using $\sigma = 0.5$.