Journal of the American Statistical Association

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/uasa20

A Scale-Free Approach for False Discovery Rate Control in Generalized Linear Models

Chenguang Dai, Buyu Lin, Xin Xing & Jun S. Liu

To cite this article: Chenguang Dai, Buyu Lin, Xin Xing & Jun S. Liu (2023) A Scale-Free Approach for False Discovery Rate Control in Generalized Linear Models, Journal of the American Statistical Association, 118:543, 1551-1565, DOI: 10.1080/01621459.2023.2165930

To link to this article: https://doi.org/10.1080/01621459.2023.2165930

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2023, VOL. 118, NO. 543, 1551–1565: Theory and Methods Discussion
https://doi.org/10.1080/01621459.2023.2165930

A Scale-Free Approach for False Discovery Rate Control in Generalized Linear Models

Chenguang Dai*[a], Buyu Lin*[a], Xin Xing[b], and Jun S. Liu[a]

[a]Department of Statistics, Harvard University, Cambridge, MA; [b]Department of Statistics, Virginia Tech, Blacksburg, VA

ABSTRACT

The Generalized Linear Model (GLM) has been widely used in practice to model counts or other types of non-Gaussian data. This article introduces a framework for feature selection in the GLM that can achieve robust False Discovery Rate (FDR) control. The main idea is to construct a *mirror statistic* based on data perturbation to measure the importance of each feature. FDR control is achieved by taking advantage of the mirror statistic's property that its sampling distribution is (asymptotically) symmetric about zero for any null feature. In the moderate-dimensional setting, that is, $p/n \to \kappa \in (0, 1)$, we construct the mirror statistic based on the maximum likelihood estimation. In the high-dimensional setting, that is, $p \gg n$, we use the debiased Lasso to build the mirror statistic. The proposed methodology is scale-free as it only hinges on the symmetry of the mirror statistic, thus, can be more robust in finite-sample cases compared to existing methods. Both simulation results and a real data application show that the proposed methods are capable of controlling the FDR and are often more powerful than existing methods including the Benjamini-Hochberg procedure and the knockoff filter. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received March 2021
Accepted December 2022

KEYWORDS
Data perturbation; FDR control; Generalized linear model; Feature selection

1. Introduction

The Generalized Linear Model (GLM) is a powerful tool for building a linear relationship between certain characteristic of a non-Gaussian response variable y (e.g., categorical and count data) and p explanatory features $X_1, \ldots, X_p$ through a link function. Although p is often large in the current big data era, the response variable y most likely depends only on a small subset of features. Thus, it is of significant interest to identify those relevant features in order to both sharpen the analysis and better understand the results. A desirable feature selection procedure is expected to control the False Discovery Rate (FDR) (Benjamini and Hochberg 1995) defined as

$$\text{FDR} = \mathbb{E}[\text{FDP}], \quad \text{with FDP} = \frac{|S_0 \cap \widehat{S}|}{|\widehat{S}| \vee 1},$$

in which FDP stands for the "false discovery proportion," and S_0, $\widehat{S}$ denote the index sets of the null and the selected features, respectively. The expectation is taken with respect to the randomness in both the data and the selection procedure. Existing FDR control methods that can be applied to GLMs include the Benjamini-Hochberg (BHq) procedure and the model-X knockoff filter (Candès et al. 2018; Huang and Janson 2020).

There are notable limitations, however, when applying either BHq or Knockoff in practice. BHq requires p-values, which are difficult to construct in high dimensions. Javanmard and Javadi (2019) and Ma, Tony Cai, and Li (2020) considered applying BHq to high-dimensional linear and logistic regression models, respectively, with p-values obtained via the debaised Lasso (Javanmard and Montanari 2014; Van de Geer et al. 2014; Zhang and Zhang 2014). However, the debiased Lasso only provides asymptotically valid p-values, which often appear highly nonuniform under the null in finite-sample cases (e.g., see Figure 1 and also the discussions and empirical evidences in Candès et al. 2018).

Knockoff requires nearly exact knowledge of the joint distribution of all features, potentially limiting its applicability in high dimensions. If this distribution is unknown, Barber, Candès, and Samworth (2020) showed that the inflation of the FDR is proportional to the estimation error in the conditional distribution $\mathbb{P}(X_j \mid X_{-j})$, where $X_{-j} = \{X_1, \ldots, X_p\} \backslash \{X_j\}$. Simulation results in Dai et al. (2022) also suggest that misspecifying the joint distribution of the features may result in an FDR inflation and power loss. Recent developments in generating good knockoff features include Romano, Sesia, and Candès (2019), Jordon, Yoon, and Schaar (2019) (using deep generative models) and Bates et al. (2020) (using sequential MCMC algorithms). Furthermore, Huang and Janson (2020) generalized the model-X knockoff filter using conditioning to allow features following an exponential family distribution with unknown parameters. Bates et al. (2021), Yang et al. (2021), and Marandon et al. (2022) apply the similar "knockoff-type" idea to the novel detection problem. Power analysis of Knockoff and related methods has been carried out in Weinstein, Barber, and Candes (2017), Weinstein et al. (2020) and Ke, Liu, and Ma (2020).

In this article, we propose a new FDR control framework for the GLM, which requires neither p-values nor the joint distribution of features. Two asymptotic regimes for the GLM are considered. The moderate-dimensional setting refers to the regime

CONTACT Jun S. Liu jliu@stat.harvard.edu Department of Statistics, Harvard University, Cambridge, MA.
*These authors contribute equally to this work.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

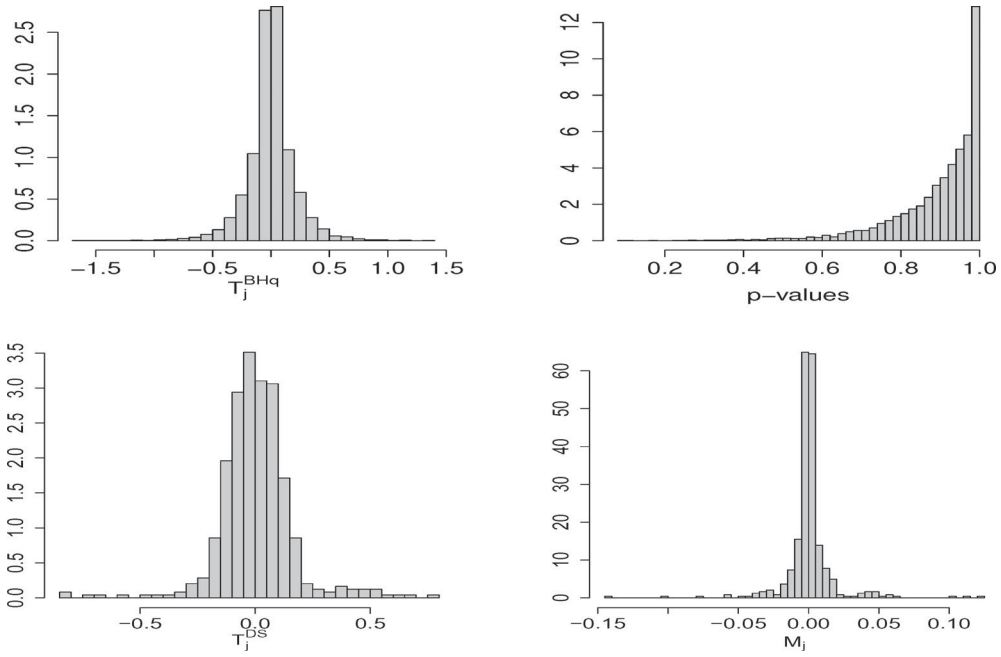

Figure 1. A logistic regression example with $n = 250, p = 500, p_1 = 10$, and $X_i \sim N(0, I_p)$. The true coefficient for a nonnull feature $j \in S_1$ is set to be $\beta_j^\star = \pm 4$ with equal probability. Top left: the normalized debiased Lasso estimates T_j^{BHq} for the null features. Top right: the p-values of the null features. Both histograms are generated based on 20 independent runs (with regeneration of the response vector y) of the algorithm in Ma, Tony Cai, and Li (2020). Bottom left: the normalized debiased Lasso estimates T_j^{DS} (see (21)) for the null features. Bottom right: the mirror statistics of the null features. Both histograms are generated based on a single run of Algorithm 6.

where $p/n \to \kappa \in (0, 1)$. In this case, the classical asymptotic result for the Maximum Likelihood Estimator (MLE) breaks down in the sense that the asymptotic normal distribution of the MLE involves additional bias and variance scaling factors (Sur and Candès 2019). As a result, BHq faces the challenge of estimating these two scaling factors in order to obtain valid p-values, which remains an open problem for GLMs other than logistic/probit regressions. In contrast, our proposed method is scale-free and does not require estimating the aforementioned scaling factors, thus, can be easily and validly applied to all GLMs. The high-dimensional setting refers to the regime where $p \gg n$, in which we use debiased Lasso estimates and prove that the proposed method achieves FDR control under certain regularity conditions. A main advantage of our method over BHq is that we do not need the variances of debiased Lasso estimates and only rely on the symmetry of the estimates under the null to control FDR, which is much easier to achieve in finite-sample cases.

The rest of the article is structured as follows. Section 2 introduces our FDR control framework and two basic methods for constructing the mirror statistic: Gaussian mirror and data splitting. Sections 3 and 4 concern FDR control for the GLM in the moderate-dimensional and the high-dimensional settings, respectively. Sections 5.1 and 5.2 demonstrate the competitive performances of our proposed methods through simulation studies on popular GLMs including logistic, Poisson and negative binomial regressions. Section 5.3 considers selecting relevant genes associated with the glucocorticoid response based on a single-cell RNA sequencing data. Section 6 concludes with final remarks. Proofs and additional numerical results are given in supplementary materials.

2. FDR Control via Mirror Statistics

For a given response variable y, we consider p candidate features $X_1, \ldots, X_p$. Let $X_{n\times p}$ be the design matrix, in which each row x_i^T for $i \in [n]$[1] is an independent realization of these features. Let $y = (y_1, \ldots, y_n)^\mathsf{T}$ be the associated response vector. We assume that the conditional distribution $\mathbb{P}(y \mid X)$ depends only on a subset of features with the corresponding index set denoted as S_1. Let $p_1 = |S_1|$ and $p_0 = p - p_1$. We call X_j relevant (nonnull) if $j \in S_1$; otherwise we call it a null feature. The index set of the null features is denoted as S_0. The goal is to identify as many relevant features as possible with the FDR under control. Throughout, we denote the power of a selection procedure as

$$\text{Power} = \mathbb{E}|S_1 \cap \widehat{S}|/p_1,$$

in which $\widehat{S}$ denotes the index set of the selected features.

Similar to Knockoff, our FDR control framework requires constructing a mirror statistic M_j for each feature X_j with the following two properties:

(A1) A feature with a larger mirror statistic is more likely to be a relevant feature.

(A2) The mirror statistic's distribution under the null is (asymptotically) symmetric about zero.

By Property (A1), we can rank the features by their mirror statistics and select those with their mirror statistics greater than a cutoff. For any cutoff $t > 0$, Property (A2) suggests an approximate upper bound on the number of false positives,

$$\mathrm{FDP}(t) = \frac{\#\{j : j \in S_0, M_j > t\}}{\#\{j : M_j > t\} \vee 1} \lesssim \frac{\#\{j : M_j < -t\}}{\#\{j : M_j > t\} \vee 1}, \qquad (1)$$

[1] $[n]$ denotes the set $\{1, \ldots, n\}$.

which leads to the following FDR control framework as outlined in Algorithm 1.

Algorithm 1 The FDR control framework.

1. Construct the mirror statistic M_j for $j \in [p]$.
2. Given a designated FDR level $q \in (0, 1)$, set the cutoff τ_q as

$$\tau_q = \inf\left\{t > 0 : \widehat{\mathrm{FDP}}(t) = \frac{\#\{j : M_j < -t\} + 1}{\#\{j : M_j > t\} \vee 1} \le q\right\}.$$

3. Select the features $\widehat{S} = \{j : M_j > \tau_q\}$.

Remark 2.1. The data-driven cutoff τ_q and the selection set $\widehat{S}$ are motivated by Knockoff (Barber and Candès 2015). The "+1" in the numerator of $\widehat{\mathrm{FDP}}(t)$ is theoretically redundant for asymptotic FDR control, but is critical for Knockoff to achieve finite-sample FDR control. We recommend to keep the "+1" in practice to make $\widehat{S}$ slightly more conservative, especially when the number of relevant features p_1 is small.

A general recipe for constructing the mirror statistic in regression settings is as follows. For $j \in [p]$, we obtain two estimates, $\widehat{\beta}_j^{(1)}$ and $\widehat{\beta}_j^{(2)}$, of the true coefficient $\beta_j^\star$. We standardize the two estimates (more details in Section 3.2) so that the resulting mirror statistics to have comparable variances. The standardized estimates $T_j^{(1)}$ and $T_j^{(2)}$ should satisfy the following conditions:

Condition 2.1.

- (Independence) The two regression coefficients are (asymptotically) independent.
- (Symmetry) The distribution of either of the two regression coefficients is (asymptotically) symmetric about zero under the null.

The mirror statistic M_j proposed in Dai et al. (2022) takes a general form:

$$M_j = \mathrm{sign}\big(T_j^{(1)} T_j^{(2)}\big) f\big(|T_j^{(1)}|, |T_j^{(2)}|\big), \tag{2}$$

where $f(u, v)$ is a user-specified bivariate function defined on $\mathbb{R}^+ \times \mathbb{R}^+$ that is nonnegative, exchangeable in u and v, that is, $f(u, v) = f(v, u)$, and monotonically increasing in both u and v.

Note that for a relevant feature $j \in S_1$, the two regression coefficients $T_j^{(1)}$ and $T_j^{(2)}$ tend to be large in magnitude and have the same sign if the estimation procedures are reasonably accurate. Since $f(u, v)$ is monotonically increasing in both $|u|$ and $|v|$, the mirror statistic M_j is likely to be positive and relatively large, which implies Property (A1). In addition, for a null feature $j \in S_0$, Property (A2) holds given Condition 2.1 since $T_j^{(1)}$ and $T_j^{(2)}$ are (asymptotically) independent and one of them is (asymptotically) symmetric about zero.

Three convenient choices of $f(u, v)$ are

$$f(u, v) = 2\min(u, v), \quad f(u, v) = uv, \quad f(u, v) = u + v. \tag{3}$$

The first choice leads to the mirror statistic proposed in Xing, Zhao, and Liu (2019), while the third choice corresponds to the "sign-maximum" between $|T_j^{(1)} + T_j^{(2)}|$ and $|T_j^{(1)} - T_j^{(2)}|$, and is optimal in a simplified setting as shown by Dai et al. (2022). The optimality of the sign-max mirror statistic for Knockoff has also been empirically observed by Barber and Candès (2015) and recently proved by Ke, Liu, and Ma (2020) under the weak-and-rare signal setting. The following sections review two recently proposed methods, Gaussian mirror (Xing, Zhao, and Liu 2019) and data splitting (Dai et al. 2022), for constructing the two regression coefficients $T_j^{(1)}$ and $T_j^{(2)}$ that satisfy Condition 2.1.

2.1. Gaussian Mirror

For an easy illustration, we restrict ourselves to low-dimensional ($n > p$) linear models. The idea of Gaussian mirror is to create a pair of perturbed mirror features,

$$X_j^+ = X_j + c_j Z_j, \quad X_j^- = X_j - c_j Z_j, \tag{4}$$

in which c_j is an adjustable scalar and Z_j follows $N(0, 1)$ independently across $j \in [p]$. The linear model with β^* as the true parameter vector can then be equivalently recasted as

$$y = \frac{\beta_j^\star}{2} X_j^+ + \frac{\beta_j^\star}{2} X_j^- + X_{-j}\beta_{-j}^\star + \epsilon. \tag{5}$$

In low dimensions, we obtain $\widehat{\beta}^+$ and $\widehat{\beta}^-$, as well as the normalized estimates T_j^+ and T_j^-, via the ordinary least squares (OLS). For any null feature $j \in S_0$, both T_j^+ and T_j^- follow a t-distribution centered at zero. Thus, the resulting M_j satisfies the symmetry requirement in Condition 2.1. Furthermore, since (T_j^+, T_j^-) asymptotically follows a bivariate normal distribution, we can choose a proper c_j as below so that T_j^+ and T_j^- are asymptotically independent:[2]

$$c_j = \|P_{-j}^\perp X_j\| / \|P_{-j}^\perp Z_j\|, \tag{6}$$

where $P_{-j}^\perp$ is the projection matrix onto the orthogonal complement of the column space spanned by X_{-j}.

It is possible to generate (X_j^+, X_j^-) simultaneously for all $j \in [p]$ and fit the GLM once. However, despite the increasing computational demand, the one-feature-per-fit procedure introduces the least noise, and thus gives a better ranking of the features. In simulation studies, we also consistently observe superior performances of the one-feature-per-fit procedure.

2.2. Data Splitting

The simplest way to obtain two independent regression coefficients is via data splitting. Specifically, we randomly split the data into two halves, $(y^{(1)}, X^{(1)})$ and $(y^{(2)}, X^{(2)})$, and estimate $\widehat{\beta}_j^{(1)}$ and $\widehat{\beta}_j^{(2)}$, as well as their normalized versions $T_j^{(1)}$ and $T_j^{(2)}$, using each part of the data. The independence between the two estimates is naturally implied by data splitting. The symmetry requirement in Condition 2.1 can be satisfied if, for any null feature, either of the estimates is (asymptotically) normal and

[2] Slightly different from Xing, Zhao, and Liu (2019), because we standardize the OLS estimates, we cannot achieve the finite-sample independence between T_j^+ and T_j^- by varying c_j.

centered at zero. As we will show later, desirable estimates can be constructed for the GLM under certain conditions. We assume a half-half data splitting throughout, which generally leads to the highest power based on our empirical observations.

The potential power loss is a major concern of using data splitting. Dai et al. (2022) proposed to remedy this issue by aggregating results from repeated sample splits, which also helps to stabilize the selection result. We use DS and MDS to refer to the single data-splitting and multiple data-splitting methods, respectively. The idea of MDS is to obtain multiple selection results via repeated sample splits, and determine the importance of a feature based on its inclusion rate I_j defined as

$$I_j = \mathbb{E}\left[\frac{\mathbb{1}(j \in \widehat{S})}{|\widehat{S}| \vee 1} \,\middle|\, X, y\right], \quad \widehat{I}_j = \frac{1}{m}\sum_{k=1}^{m} \frac{\mathbb{1}(j \in \widehat{S}^{(k)})}{|\widehat{S}^{(k)}| \vee 1}, \tag{7}$$

where $\widehat{I}_j$ is a natural estimate of I_j, m is the total number of sample splits, and $\widehat{S}^{(k)}$ is the index set of the selected features in the kth sample split. Note that $\sum_j I_j = \Pr(|\widehat{S}| \geq 1)$, and, if all $|\widehat{S}^{(k)}| \geq 1$, then $\sum_j \widehat{I}_j = 1$. We then select those features with their empirical inclusion rates $\widehat{I}_j$ larger than a properly chosen cutoff. The MDS procedure as summarized in Algorithm 2 can be applied on top of the DS procedure designed for the GLM (see Sections 3 and 4).

Algorithm 2 Aggregating selection results from multiple sample splits.

1. Sort the estimated inclusion rates: $0 \leq \widehat{I}_{(1)} \leq \widehat{I}_{(2)} \leq \cdots \leq \widehat{I}_{(p)}$.
2. Given a designated FDR level $q \in (0, 1)$, find the largest $\ell \in [p]$ such that $\widehat{I}_{(1)} + \cdots + \widehat{I}_{(\ell)} \leq q$.
3. Select the features $\widehat{S} = \{j : \widehat{I}_j > \widehat{I}_{(\ell)}\}$.

Ideally, we would like to conduct as many sample splits as possible in order to estimate the inclusion rates accurately. In practice, however, we find that the power of MDS no longer improves much after a small number of independent sample splits (e.g., $m \geq 50$). In addition, Dai et al. (2022) showed that, for the normal means problem, MDS can retrieve almost the full information regarding the feature selection task, in the sense that with high probability the inclusion rates give the same ranking of features as the p-values calculated using the full data.

3. Generalized Linear Models in Moderate Dimensions

For $y = (y_1, \ldots, y_n)^\top$, we consider the following GLM with a canonical link function ρ:

$$p(y \mid X, \beta^\star) = \prod_{i=1}^{n} c(y_i) \exp\left(y_i x_i^\mathsf{T} \beta^\star - \rho(x_i^\mathsf{T} \beta^\star)\right), \tag{8}$$

in which $\beta^\star$ denotes the p dimensional true coefficient vector. In the moderate-dimensional setting, we assume that $p/n \to \kappa \in (0, 1)$. Note that the classical setting with fixed p corresponds to the case $\kappa = 0$. We impose the following assumption on the link function ρ as in Abbasi (2020).

Assumption 3.1 (Assumption 1 in Abbasi (2020)). Define the Moreau envelop of the loss function that is, the negative log-likelihood, of the GLM in (8) as

$$G(x, y, t) = \min_{v \in R} \left\{\frac{1}{2t}(v - x)^2 + \rho(x) - xy\right\}.$$

For all $c_1, c_2 \in \mathcal{R}$ and $\tau > 0$, there exists a continuous function $g : \mathbb{R} \times \mathbb{R} \times \mathbb{R}_{>0} \to \mathbb{R}$ such that

$$\frac{1}{n}\sum_{i=1}^{n} G(c_1 h_i + c_2 x_i^\mathsf{T} \beta^\star, y_i, \tau) \xrightarrow{p} g(c_1, c_2, \tau),$$

where the convergence is in probability over the distribution of y, the random matrix $X \in \mathbb{R}^{n \times p}$ with iid standard Gaussian entries, and the random vector h with iid standard Gaussian entries.

Assumption 3.1 holds for a wide range of GLMs and machine learning models including logistic regression, Poisson regression, and support vector machines (Abbasi 2020). We further assume that x_i's are iid observations for a distribution with mean 0 and covariance matrix Σ, and that the signal strength converges to a constant, that is, $\gamma_n := \mathrm{var}(x_i^\mathsf{T} \beta^\star) \to \gamma^2$.

3.1. Properties of the MLE

Let $\rho(X\beta) = (\rho(x_1^\mathsf{T}\beta), \ldots, \rho(x_n^\mathsf{T}\beta))^\mathsf{T}$. The MLE of $\beta^\star$ can be written as

$$\widehat{\beta} = \arg\min_{\beta \in \mathbb{R}^p} \left\{\frac{1}{n} 1^\mathsf{T} \rho(X\beta) - \frac{1}{n} y^\mathsf{T} X\beta\right\}, \tag{9}$$

which can behave very differently when $\kappa > 0$ compared with the classical setting. First, the GLM may not be identifiable, that is, the MLE does not exist uniquely. For instance, for logistic regression, the MLE does not exist if the two classes are well separated, and the corresponding phase transition curve for the existence of a unique MLE is recently derived by Candès and Sur (2020). Throughout Section 3, we assume the existence of a unique MLE with probability approaching 1 as $n, p \to \infty$ whenever necessary. Second, the MLE is asymptotically biased and its asymptotic variance also differs from the classical result. By generalizing the results of Zhao, Sur, and Candès (2020) and Salehi, Abbasi, and Hassibi (2019), we have the following asymptotic characterization of the MLE.

Proposition 3.1. Consider the GLM defined in (8) in which $x_i \overset{\text{iid}}{\sim} N(0, \Sigma)$. Let $\Theta = \Sigma^{-1}$ and $\tau_j^2 = \Theta_{jj}^{-1}$. Assuming that $\sqrt{n}\tau_j \beta_j^\star = O(1)$, we have

$$\mathbb{P}\left(\frac{\sqrt{n}(\widehat{\beta}_j - \alpha_\star \beta_j^\star)}{\sigma_\star / \tau_j} \leq x \,\middle|\, \text{MLE exists}\right) \longrightarrow \Phi(x), \tag{10}$$

where $\Phi(x)$ is the CDF of the standard Gaussian, and $\alpha_\star, \sigma_\star$ are two universal constants depending on the link function ρ, the true coefficient vector $\beta^\star$, the signal strength γ, and the ratio κ.

Remark 3.1. The proof of Proposition 3.1 uses the stochastic representation of the MLE in Zhao, Sur, and Candès (2020) and the Convex Gaussian Min-max Theorem (CGMT). The

Gaussian assumption on the x_i's is not required for the DS procedures per se, but allows us to use the CGMT directly to prove theoretical results. The constant pair $(\alpha_\star, \kappa\sigma_\star^2)$ is the limit of

$$\alpha_n = \frac{\langle \widehat{\beta}, \beta^\star \rangle}{\|\widehat{\beta}\|^2}, \quad \sigma_n^2 = \|P_{\widehat{\beta}^\star}^{\perp} \widehat{\beta}\|, \tag{11}$$

in which $P_{\widehat{\beta}^\star}^{\perp}$ is the projection matrix onto the orthogonal complement of $\beta^\star$. The convergence of (α_n, σ_n^2) follows from a routine application of CGMT as in Salehi, Abbasi, and Hassibi (2019). More details can be found in Lemma A.1 in supplementary materials.

Remark 3.2. Note that $\tau_j^2 = \mathrm{var}(X_j \mid X_{-j})$. We can thus estimate it via a a node-wise regression approach. More precisely, we regress X_j against X_{-j} and obtain the residual sum of squares RSS_j. An unbiased estimator of τ_j^2 is then $\widehat{\tau}_j^2 = \mathrm{RSS}_j/(n-p+1)$.

By Proposition 3.1, we need to estimate the bias and variance scaling factors $(\alpha_\star, \sigma_\star)$ in order to obtain p-values. This is in general very challenging as it requires one to first estimate the signal strength γ. Sur and Candès (2019) proposed the *ProbFrontier* method to estimate γ using the phase transition curve that calibrates the existence of a unique MLE. However, to the best of our knowledge, existing results only cover logistic/probit regressions and there is no unified approach to derive the curve for a general GLM. Therefore, it remains unknown how one should apply BHq going beyond logistic/probit regressions.

3.2. FDR Control via Data Splitting

In comparison with BHq, the DS procedure outlined in Algorithm 3 does not require estimating the scaling factors $(\alpha_\star, \sigma_\star)$, thus, is applicable to all GLMs in the moderate-dimensional setting. By (10), we normalize the two independent MLEs $\widehat{\beta}^{(1)}$ and $\widehat{\beta}^{(2)}$ as below,

$$T_j^{(1)} = \widehat{\tau}_j^{(1)} \widehat{\beta}_j^{(1)}, \quad T_j^{(2)} = \widehat{\tau}_j^{(2)} \widehat{\beta}_j^{(2)}, \tag{12}$$

where $\widehat{\tau}_j^{(1)}$ and $\widehat{\tau}_j^{(2)}$ are two independent estimates of τ_j (see Remark 3.2). Although the asymptotic standard deviation of $\widehat{\beta}_j$ is $\sigma_\star/\tau_j$, we can safely drop the constant $\sigma_\star$ because our FDR control framework (Algorithm 1) is scale-invariant with respect to the mirror statistics. DS also does not require estimating the bias scaling factor $\alpha_\star$ since $\alpha_\star \beta_j^\star = 0$ under the null. Thus, the symmetry requirement in Condition 2.1 is asymptotically satisfied according to Proposition 3.1.

To theoretically justify DS, we define $S_{1,\mathrm{strong}}$ to be the largest subset of S_1 such that

$$\sqrt{n} \min_{j \in S_{1,\mathrm{strong}}} |\beta_j^\star| \to \infty. \tag{13}$$

Note such a set might be empty. Let $p_{1,\mathrm{strong}} = |S_{1,\mathrm{strong}}|$. We require the following assumptions.

Assumption 3.2.

1. $1/C \le \sigma_{\min}(\Sigma) \le \sigma_{\max}(\Sigma) \le C$ for some constant $C > 0$.
2. $p_0 \to \infty$, $\liminf p_{1,\mathrm{strong}}/p_0 > 0$ as $n, p \to \infty$.

Algorithm 3 The data-splitting method for GLMs in the moderate-dimensional setting.

1. Split the data into two equal-sized halves $(y^{(1)}, X^{(1)})$ and $(y^{(2)}, X^{(2)})$.
2. For $j \in [p]$, regress $X_j^{(1)}$ onto $X_{-j}^{(1)}$, and regress $X_j^{(2)}$ onto $X_{-j}^{(2)}$. Let
$$\widehat{\tau}_j^{2(1)} = \frac{\mathrm{RSS}_j^{(1)}}{n/2-p+1}, \quad \widehat{\tau}_j^{2(2)} = \frac{\mathrm{RSS}_j^{(2)}}{n/2-p+1},$$
in which RSS_j is the residual sum of squares.
3. Find the MLEs $\widehat{\beta}^{(1)}$ and $\widehat{\beta}^{(2)}$ using each part of the data. For $j \in [p]$, calculate the mirror statistic M_j following (2) based on $T_j^{(1)}$ and $T_j^{(2)}$ defined in (12).
4. Select features using Algorithm 1.

Remark 3.3. Assumption 3.2 (1) also appears in Zhao, Sur, and Candès (2020), in which $\sigma_{\min}(\Sigma)$ and $\sigma_{\max}(\Sigma)$ refer to the smallest and the largest eigenvalues of the covariance matrix Σ. Empirically we observed that DS may still achieve FDR control even if Assumption 3.2 (1) is violated, for example, when features have constant pairwise correlation (see Sections B.1.1 and B.1.2 in supplementary materials). The condition $\liminf p_{1,\mathrm{strong}}/p_0 > 0$ can be possibly relaxed to accommodate the cases of very sparse signals and the global null by imposing some additional weak-correlated assumption among the mirror statistics (e.g., see Assumption 4.1 (2)(a)).

Proposition 3.2. Consider a GLM defined in (8), in which $x_i \overset{\mathrm{iid}}{\sim} N(0, \Sigma)$. For any FDR control level $q \in (0,1)$, under Assumption 3.2, we have

$$\mathrm{FDP} \le q + o_p(1) \quad \text{and} \quad \limsup_{n,p \to \infty} \mathrm{FDR} \le q$$

for the DS procedure outlined in Algorithm 3.

Remark 3.4. We give a sketch of proof for Proposition 3.2. Without loss of generality, we assume $qp_1 < (1-q)p_0$, otherwise we can simply select all features and the FDR is under control. Let

$$H(t) = \mathbb{P}\big(\mathrm{sign}(Z_1 Z_2) f(|Z_1|, |Z_2|) > t\big),$$

in which Z_1, Z_2 are two independent standard Gaussian variables, and f is the bivariate function used in constructing the mirror statistic. The main proof arguments are

$$\sup_{0 \le t \le t^\star} \left| \frac{\sum_{j \in S_0} \mathbb{1}(M_j > t)}{\sum_{j \in S_0} \mathbb{1}(M_j < -t)} - 1 \right| \xrightarrow{p} 0 \quad \text{and} \quad \mathbb{P}(\tau_q \le t^\star) \to 1, \tag{14}$$

in which $t^\star$ satisfies $H(t^\star) = \frac{qp_{1,\mathrm{strong}}}{2(1-q)p_0}$. Assumption 3.2 (2) ensures that $t^\star$ is bounded as $n, p \to \infty$. We show the first part of (14) via Markov's inequality, that is, bounding $\mathrm{cov}(\mathbb{1}(M_i > t), \mathbb{1}(M_j > t))$ for $i, j \in S_0$. For the second part, using the signal strength condition (13), we prove that

$$\mathbb{P}\Big(\min_{j \in S_{1,\mathrm{strong}}} M_j > t^\star \Big) \to 1, \quad n, p \to \infty.$$

By the definitions of $t^\star$ and τ_q, this implies $\widehat{\mathrm{FDP}}(t^\star) \le q$, and thus, $\mathbb{P}(\tau_q \le t^\star) \to 1$.

As discussed in Section 2.2, we can further enhance the power and the stability of the selection result via MDS. The following proposition establishes FDR control for MDS.

Proposition 3.3. Consider the GLM defined in (8), in which $x_i \overset{\text{iid}}{\sim} N(0, \Sigma)$. For any FDR control level $q \in (0, 1)$, if $p_{1,\text{strong}}/p_1 \to 1$ as $n, p \to \infty$, under Assumption 3.2, we have

$$\text{FDP} \le q + o_p(1) \quad \text{and} \quad \limsup_{n,p\to\infty} \text{FDR} \le q$$

for the MDS procedure based on Algorithm 3.

Remark 3.5. We impose an additional signal strength condition $p_{1,\text{strong}}/p_1 \to 1$ when proving FDR control for MDS. Note that DS is free of this assumption (see Proposition 3.2), and numerical studies (e.g., see Section 5 and Table B.2 in supplementary materials) suggest that MDS achieves FDR control whenever DS does, even if $p_{1,\text{strong}}/p_1 \ll 1$. However, it remains unclear how this assumption can be possibly relaxed.

3.3. FDR Control via Gaussian Mirror

Cover (Cover 1964, 1965) showed that the MLE exists uniquely only if $\kappa \in (0, 1/2]$, hence, DS is only applicable when $\kappa \in (0, 1/4]$, that is, $n \ge 4p$. The same issue also occurs to Knockoff as it doubles the number of features. Besides, even if $\kappa \in (0, 1/4]$ and the MLE exists uniquely as $n, p \to \infty$, in the finite-sample case, there is still a chance that the MLE does not exist uniquely for some sample splits used in MDS. To overcome this issue, we consider the Gaussian mirror method, which extends the applicability to $\kappa \in (0, 1/2]$ as long as the MLE exists uniquely on the full data.

As discussed in Section 2.1, we fit a GLM using the response vector y and the augmented set of features (X_{-j}, X_j^+, X_j^-) to find the MLEs, $\widehat{\beta}_j^+$ and $\widehat{\beta}_j^-$, associated with the pair of perturbed mirror features (X_j^+, X_j^-) defined in (4). Let Σ_{aug} be the covariance matrix of (X_{-j}, X_j^+, X_j^-), and let $\Theta_{\text{aug}} = \Sigma_{\text{aug}}^{-1}$. We have the following asymptotic characterization.

Proposition 3.4. Consider fitting a GLM using the response vector y and the augmented set of features (X_{-j}, X_j^+, X_j^-) defined in (4). Then, the asymptotic distribution of the MLE $(\widehat{\beta}_j^+, \widehat{\beta}_j^-)$ is

$$\frac{\sqrt{n}}{\sigma_\star}\left(\begin{pmatrix}\widehat{\beta}_j^+\\ \widehat{\beta}_j^-\end{pmatrix} - \frac{\alpha_\star}{2}\begin{pmatrix}\beta_j^\star\\ \beta_j^\star\end{pmatrix}\right) \xrightarrow{d} N\big(0, \Theta^*\big), \tag{15}$$

in which Θ^* is the 2×2 submatrix at the right bottom of Θ_{aug} corresponding to (X_j^+, X_j^-), $\alpha_\star, \sigma_\star$ are defined as in Proposition 3.1, and the design matrix is Gaussian.

We can choose a proper scalar c_j so that the off-diagonal entry of Θ^* is zero. This implies that the MLEs $\widehat{\beta}_j^+$ and $\widehat{\beta}_j^-$ are asymptotically independent by Proposition 3.4. In practice, we can plug in the sample version of Σ_{aug} and the resulting scalar c_j is in the form of (6). Besides,

$$\begin{aligned} 1/\Theta_{11}^* = 1/\Theta_{22}^* &= \text{Var}(X_j^+ \mid X_j^-, X_{-j}) \\ &\approx \text{var}(X_j^+ \mid X_{-j}) = \tau_j^2 + c_j^2, \end{aligned}$$

where the approximately-equal sign follows from the fact that the asymptotic independence between $\widehat{\beta}_j^+$ and $\widehat{\beta}_j^-$ implies the asymptotic independence between X_j^+ and X_j^- conditioning on X_{-j}. We thus normalize $\widehat{\beta}_j^+$ and $\widehat{\beta}_j^-$ as

$$T_j^+ = \sqrt{\widehat{\tau}_j^2 + c_j^2}\,\widehat{\beta}_j^+, \quad T_j^- = \sqrt{\widehat{\tau}_j^2 + c_j^2}\,\widehat{\beta}_j^-, \tag{16}$$

with $\widehat{\tau}_j^2$ defined in Remark 3.2. We summarize the Gaussian mirror method in Algorithm 4.

Algorithm 4 The Gaussian mirror method for GLMs in the moderate-dimensional setting.

1. For $j \in [p]$, calculate the mirror statistic M_j as follows.
   (a) Simulate Z_j from $N(0, I_n)$.
   (b) Calculate the scaling factor c_j according to (6).
   (c) Fit a GLM using y and (X_{-j}, X_j^+, X_j^-) to find the MLEs $\widehat{\beta}_j^+$ and $\widehat{\beta}_j^-$.
   (d) Estimate $\widehat{\tau}_j^2$ following Remark 3.2.
   (e) Calculate the mirror statistic M_j following (2) based on T_j^+ and T_j^- defined in (16).
2. Select features using Algorithm 1.

Proposition 3.5. Consider a GLM defined in (8), in which $x_i \overset{\text{iid}}{\sim} N(0, \Sigma)$. For any FDR control level $q \in (0, 1)$, under Assumption 3.2, we have

$$\text{FDP} \le q + o_p(1) \quad \text{and} \quad \limsup_{n,p\to\infty} \text{FDR} \le q$$

for the Gaussian mirror method outlined in Algorithm 4.

The normality assumption on the design matrix in this section is mainly for technical purposes to apply the Convex Gaussian Min-max Theorem (CGMT). We expect that all propositions in Section 3 hold more generally with the proviso that the joint distribution of features has a sufficiently light tail (see the discussion in Sur and Candès 2019). Empirically, the proposed methods work well for non-Gaussian designs (e.g., see Figure B.6 in supplementary materials).

4. Generalized Linear Models in High Dimensions

We consider the high-dimensional setting ($p \gg n$), in which we base the mirror statistic on the regularized estimator instead of the MLE. Considering computational feasibility, we focus on the data-splitting methods. For the GLM defined in (8), we define its *loss function* as

$$\ell(u, v) = -uv + \rho(v),$$

which is just the negative log-likelihood up to an additive constant. Denote

$$\begin{aligned} \dot{\ell}(u, v) &= \frac{\partial \ell(u, v)}{\partial v}, \quad \ddot{\ell}(u, v) = \frac{\partial^2 \ell(u, v)}{\partial v^2}, \quad \dot{\ell}_\beta(y, x) \\ &= \frac{\partial \ell(y, x^\intercal \beta)}{\partial \beta}, \quad \ddot{\ell}_\beta(y, x) = \frac{\partial^2 \ell(y, x^\intercal \beta)}{\partial \beta \partial \beta^\intercal}. \end{aligned}$$

Note that both $\dot{\ell}(u,v)$ and $\ddot{\ell}(u,v)$ are scalars, whereas $\dot{\ell}_\beta(y,x)$ is a $p \times 1$ vector and $\ddot{\ell}_\beta(y,x)$ is a $p \times p$ matrix. For a general mapping g defined on the space (y,x), we define

$$P_n g = \frac{1}{n}\sum_{i=1}^{n} g(y_i, x_i) \text{ and } Pg = \mathbb{E}\left[P_n g\right].$$

Let W_β be an $n \times n$ diagonal matrix with $W_{ii}^2 = \ddot{\rho}(x_i^\intercal \beta)$. Then the sample version of the Hessian matrix can be written as $P_n \ddot{\ell}_\beta = X_\beta^\intercal X_\beta / n$, in which $X_\beta = W_\beta X$ is the weighted design matrix.

4.1. Construction of the Mirror Statistic via Debiased Lasso

The mirror statistic is built upon the Lasso estimator:

$$\widehat{\beta}(y, X; \lambda) = \arg\min_{\beta \in \mathbb{R}^p} \left\{ \frac{1}{n}\sum_{i=1}^{n} \ell(y_i, x_i^\intercal \beta) + \lambda \|\beta\|_1 \right\}. \quad (17)$$

To symmetrize $\widehat{\beta}$, we consider the following debiasing adjustment (Van de Geer et al. 2014):

$$\widehat{\beta}^d = \widehat{\beta} - \widehat{\Theta} P_n \dot{\ell}_{\widehat{\beta}}. \quad (18)$$

Here $\widehat{\Theta}$ is referred to as the decorrelating matrix.

Various proposals of $\widehat{\Theta}$ have been documented in the literature. For example, Javanmard and Montanari (2014) proposed an optimization approach to simultaneously minimize the bias and the variance of $\widehat{\beta}^d$. We follow the approach in Javanmard and Montanari (2013) and Zhang and Zhang (2014), and set $\widehat{\Theta}$ as an estimator of $\Theta = \Sigma^{-1}$, where $\Sigma = \mathbb{E}[X_{\beta^\star}^\intercal X_{\beta^\star}]/n$ is the Hessian matrix evaluated at the true coefficient vector $\beta^\star$.

One natural way to construct $\widehat{\Theta}$ in high dimensions is via regularized node-wise regression as detailed in Algorithm 5, based on the fact that $\Theta_{j,-j}$ corresponds to the coefficients of the best linear predictor of $X_{\beta^\star,j}$ using $X_{\beta^\star,-j}$. Note that estimating Θ only involves the second moment and does not require any distributional assumption of X (e.g., normality).

Algorithm 5 Construction of the decorrelating matrix $\widehat{\Theta}$.

1. Node-wise Lasso regression. For $j \in [p]$, let

$$\widehat{\gamma}_j = \arg\min_{\gamma \in \mathbb{R}^{p-1}} \left\{ \frac{1}{2n}\|X_{\widehat{\beta},j} - X_{\widehat{\beta},-j}\gamma\|_2^2 + \lambda_j \|\gamma\|_1 \right\}.$$

2. Define matrix $\widehat{C}$ with $\widehat{C}_{jj} = 1$ and $\widehat{C}_{jk} = -\widehat{\gamma}_{j,k}$ for $k \neq j$, where $\widehat{\gamma}_{j,k}$ is the kth entry of $\widehat{\gamma}_j$.
3. Let $\widehat{\Theta} = \widehat{G}^{-2}\widehat{C}$, in which $\widehat{G}^2 = \text{diag}(\widehat{\tau}_1^2, \ldots, \widehat{\tau}_p^2)$ with

$$\widehat{\tau}_j^2 = (X_{\widehat{\beta},j} - X_{\widehat{\beta},-j}\widehat{\gamma}_j)^\intercal X_{\widehat{\beta},j}/n.$$

As an example, for linear models, the canonical link function is $\rho(v) = v^2/2$, and the loss function and its derivative simplify to

$$\ell(u,v) = -uv + v^2/2, \quad \dot{\ell}_\beta(y,x) = -x^\intercal(y - x^\intercal\beta).$$

In addition, since $\ddot{\rho}(v) \equiv 1$, W_β and X_β simplify to I_n and X, respectively. Thus, $\Theta = \Sigma^{-1}$ is simply the population precision matrix of features. As a result, the debiased Lasso estimator for linear models takes the form of

$$\widehat{\beta}^d = \widehat{\beta} + \frac{1}{n}\widehat{\Theta} X^\intercal (y - X\widehat{\beta}).$$

Plugging in $y = X\beta^\star + \epsilon$, we have the following decomposition,

$$\sqrt{n}(\widehat{\beta}^d - \beta^\star) = Z + \Delta, \quad Z \mid X \sim N(0, \sigma^2 \widehat{\Theta}\widehat{\Sigma}\widehat{\Theta}^\intercal), \quad \Delta = \sqrt{n}(\widehat{\Theta}\widehat{\Sigma} - I)(\beta^\star - \widehat{\beta}), \quad (19)$$

where $\widehat{\Sigma} = (X^\intercal X)/n$ is the sample covariance matrix.

For GLMs, we have a similar decomposition as (19),

$$\sqrt{n}(\widehat{\beta}_j^d - \beta_j^\star) = Z_j + \Delta_j \text{ for } j \in [p], \quad (20)$$

in which Z_j, defined as

$$Z_j = -\sqrt{n}\Theta_{j,\cdot} P_n \dot{\ell}_{\beta^\star} = -\sqrt{n}\sum_{i=1}^{n} \Theta_{j,\cdot} x_i [-y_i + \dot{\rho}(x_i^\intercal \beta^\star)],$$

asymptotically follows the normal distribution by the central limit theorem. Here $\Theta_{j,\cdot}$ denotes the jth row of Θ. Under certain conditions, we show that the bias term Δ vanishes asymptotically (see Proposition 4.1), thus, the symmetry requirement in Condition 2.1 is satisfied. Note that the asymptotic variance of Z_j is

$$\sigma_j^2 = (\Theta \mathbb{E}[P_n \dot{\ell}_{\beta^\star} \dot{\ell}_{\beta^\star}^\top]\Theta)_{jj} = (\Theta\Sigma\Theta)_{jj} = \Theta_{jj},$$

and we further normalize $\widehat{\beta}_j^d$ as

$$T_j = \widehat{\beta}_j^d / \widehat{\sigma}_j \text{ with } \widehat{\sigma}_j^2 = (\widehat{\Theta} P_n \dot{\ell}_{\widehat{\beta}} \dot{\ell}_{\widehat{\beta}}^\top \widehat{\Theta}^\top)_{jj}, \quad (21)$$

in which $\widehat{\sigma}_j^2$ is a consistent estimator of σ_j^2.

DS proceeds by first randomly splitting the data into two halves, $(y^{(1)}, X^{(1)})$ and $(y^{(2)}, X^{(2)})$, and then calculating the two independent debiased Lasso estimates $\widehat{\beta}^{(1,d)}$ and $\widehat{\beta}^{(2,d)}$ following (18), in which $\widehat{\beta}^{(1)}$, $\widehat{\beta}^{(2)}$ and $\widehat{\Theta}^{(1)}$, $\widehat{\Theta}^{(2)}$ are computed via (17) and Algorithm 5, respectively. The mirror statistic M_j is constructed following (2) based on the normalized estimators $T^{(1)}$ and $T^{(2)}$ defined in (21). A summary of the DS procedure for high-dimensional GLMs is given in Algorithm 6.

For FDR control, DS relies only on the symmetry of the debiased Lasso estimator, whereas BHq further requires estimating its variance in order to obtain the Z-score. Our simulations show that the symmetry requirement is much less stringent and kicks in much earlier than the asymptotic normality of the Z-score in finite-sample cases. In the case of linear models, BHq needs to estimate the noise level σ, which is a challenging task in high dimensions (e.g., see the right panel of Figure B.9 in supplementary materials). If the variance is under-estimated, BHq is at the risk of losing FDR control since the resulting p-values for the null features would skew to the left. On the other hand, if the variance is over-estimated, BHq can be overly conservative, leading to a significant power loss. In contrast, DS is scale-free, that is, any rescaling of all the mirror statistics does not materially change the selection result, thus, is expected to perform more robustly than BHq. Numerical comparisons between DS and BHq can be found in Figures B.8 and B.10 in supplementary materials.

Algorithm 6 The data-splitting method for GLMs in the high-dimensional setting.

1. Split the data into two equal-sized halves $(y^{(1)}, X^{(1)})$ and $(y^{(2)}, X^{(2)})$.
2. Construct the normalized debiased Lasso estimator on each part of the data.
   (a) Calculate the Lasso estimators $\widehat{\beta}^{(1)}$ and $\widehat{\beta}^{(2)}$ via (17).
   (b) Obtain $\widehat{\Theta}^{(1)}$ and $\widehat{\Theta}^{(2)}$ following Algorithm 5. For $j \in [p]$, let
   $$\widehat{\sigma}_j^{2(1)} = \big(\widehat{\Theta}^{(1)} P_{n/2} \dot{\ell}_{\widehat{\beta}^{(1)}} \dot{\ell}_{\widehat{\beta}^{(1)}}^\top \widehat{\Theta}^{(1)\top}\big)_{jj},$$
   $$\widehat{\sigma}_j^{2(2)} = \big(\widehat{\Theta}^{(2)} P_{n/2} \dot{\ell}_{\widehat{\beta}^{(2)}} \dot{\ell}_{\widehat{\beta}^{(2)}}^\top \widehat{\Theta}^{(2)\top}\big)_{jj}.$$
   (c) Normalize the debiased Lasso estimators $\widehat{\beta}^{(1,d)}$ and $\widehat{\beta}^{(2,d)}$ from (18), that is, for $j \in [p]$,
   $$T_j^{(1)} = \widehat{\beta}_j^{(1,d)}/\widehat{\sigma}_j^{(1)}, \quad T_j^{(2)} = \widehat{\beta}_j^{(2,d)}/\widehat{\sigma}_j^{(2)}. \tag{22}$$
3. For $j \in [p]$, calculate the mirror statistic M_j following (2) based on $T_j^{(1)}$ and $T_j^{(2)}$.
4. Select features using Algorithm 1.

Figure 1 illustrates the above discussion via logistic regression. Detailed algorithmic settings are in the figure caption. We obtain the normalized debiased Lasso estimators T_j^{BHq} and T_j^{DS} following Ma, Tony Cai, and Li (2020) and Algorithm 6, respectively. We see that for BHq, the histogram of T_j^{BHq} under the null is far from that of the standard Gaussian, and the resulting p-values of the null features are significantly right-skewed. In contrast, for DS, the symmetry requirement in Condition 2.1 is approximately satisfied.

Xing, Zhao, and Liu (2019) and Dai et al. (2022) consider high-dimensional linear models based on the same FDR control framework described in Section 2. However, both methods may not be easily applicable to GLMs. Xing, Zhao, and Liu (2019) hinges on post-selection adjustments to symmetrize the regression coefficients under the null. To the best of our knowledge, post-selection adjustments have only been worked out for specific GLMs including logistic regressions and Cox's proportional hazards model (Taylor and Tibshirani 2018), but remain unknown for general GLMs. The DS procedure proposed in Dai et al. (2022) requires the sure screening property, which is a stronger signal strength assumption than the one in this article.

4.2. Theoretical Justification of the Data-Splitting Methods

Let $\Theta_{ij}^0 = \Theta_{ij}/(\Theta_{ii}\Theta_{jj})^{1/2}$, $\varrho = \max_{i\neq j} |\Theta_{ij}^0|$, $s = \max_{j\in[p]} \#\{i : \Theta_{ij}^0 \neq 0\}$, and define γ_j as

$$\gamma_j = \arg\min_{\gamma\in\mathbb{R}^{p-1}} \mathbb{E}\left[\|X_{\beta^\star,j} - X_{\beta^\star,-j}\gamma\|_2^2\right],$$

that is, the coefficient vector of the best linear predictor of $X_{\beta^\star,j}$ using $X_{\beta^\star,-j}$. Define $S_{1,\mathrm{strong}}$ as the largest subset of S_1 such that

$$\sqrt{n/\log p} \min_{j\in S_{1,\mathrm{strong}}} |\beta_j^\star| \to \infty. \tag{23}$$

Assumption 4.1. There exist some constants $\alpha_1 \in (0, (1-\varrho)/(1+\varrho))$, $\alpha_2 > 0$, $C > 0$, $c > 0$ such that the following conditions are satisfied.

1. Bounded design. $\max_{i\in[n],\, j\in[p]} |X_{ij}| \le C$ almost surely, and
   $$\mathbb{P}\Big(\max_{i\in[n]} |x_i^\intercal \beta^\star| \vee \max_{j\in[p]} \|X_{\beta^\star,-j}\gamma_j\|_\infty \ge C\Big) = O(p^{-c}).$$
2. Regularity conditions.
   (a) $\ddot{\rho}(v)$ is Lipschitz continuous for $|v| \le C$;
   (b) $|\dot{\rho}(v)|$ and $|\ddot{\rho}(v)|$ are upper bounded for $|v| \le C$;
   (c) $1/C \le \sigma_{\min}(\Sigma) \le \sigma_{\max}(\Sigma) \le C$.
3. Sparsity conditions.
   (a) $s = o(\sqrt{n}/\log^{3/2} p)$ and $s = O(p^{\alpha_1})$;
   (b) $p_1 = o(\sqrt{n}/\log^{3/2} p \wedge p^{1/2-\alpha_2})$ and $p_{1,\mathrm{strong}} \ge C(\log p)^{\alpha_2}$.

Remark 4.1. We consider the random-design scenario with a boundedness assumption (Ma, Tony Cai, and Li 2020). Similar conditions for the canonical link function ρ appear in Van de Geer et al. (2014), and hold for popular GLMs including logistic, Poisson, and negative binomial regressions. In contrast to the moderate-dimensional setting, we impose certain sparsity conditions on the true coefficient vector $\beta^\star$ and the asymptotic covariance matrix Θ of the debiased Lasso estimator, so that their respective estimators enjoy a fast convergence rate (see Bickel, Ritov, and Tsybakov 2009, Javanmard and Montanari 2013, and Van de Geer et al. 2014).

Proposition 4.1. Under Assumption 4.1, we have $\|\Delta\|_\infty = O_p(\max\{s, p_1\}\log p/\sqrt{n})$. where Δ is the bias term as defined in (20).

Proposition 4.2. For any FDR control level $q \in (0,1)$, in the asymptotic regime $p = O(n^r)$ for some constant $r > 1$, under Assumption 4.1, we have

$$\mathrm{FDP} \le q + o_p(1) \quad \text{and} \quad \limsup_{n,p\to\infty} \mathrm{FDR} \le q$$

for the DS procedure outlined in Algorithm 6.

Remark 4.2. The proof of Proposition 4.2 still hinges on (14), except that $t^\star$ can no longer be bounded given that features are sparse. We conduct a more delicate analysis of $\mathrm{cov}(\mathbb{1}(M_i > t), \mathbb{1}(M_j > t))$ for $i, j \in S_0$ under Assumption 4.1 to obtain a tighter upper bound. The key technical tools are Lemmas 6.1 and 6.2 in Liu (2013).

Proposition 4.3. Assume that there exists some constant $C > 0$ such that $\max_{j\in S_1\setminus S_{1,\mathrm{strong}}} \sqrt{n}|\beta_j^\star| \le C$. Then, under Assumption 4.1, the MDS procedure based on Algorithm 6 satisfies

$$\mathrm{FDP} \le q + o_p(1) \quad \text{and} \quad \limsup_{n,p\to\infty} \mathrm{FDR} \le q$$

in the asymptotic regime $p = O(n^r)$ for some constant $r > 1$.

Remark 4.3. Similar signal strength conditions on $S_{1,\mathrm{strong}}$ and $S_1\setminus S_{1,\mathrm{strong}}$ also appear in Bühlmann and Mandozzi (2014). Although we do not find a workaround to drop those conditions, we conjecture that they are not essential for MDS to achieve FDR control.

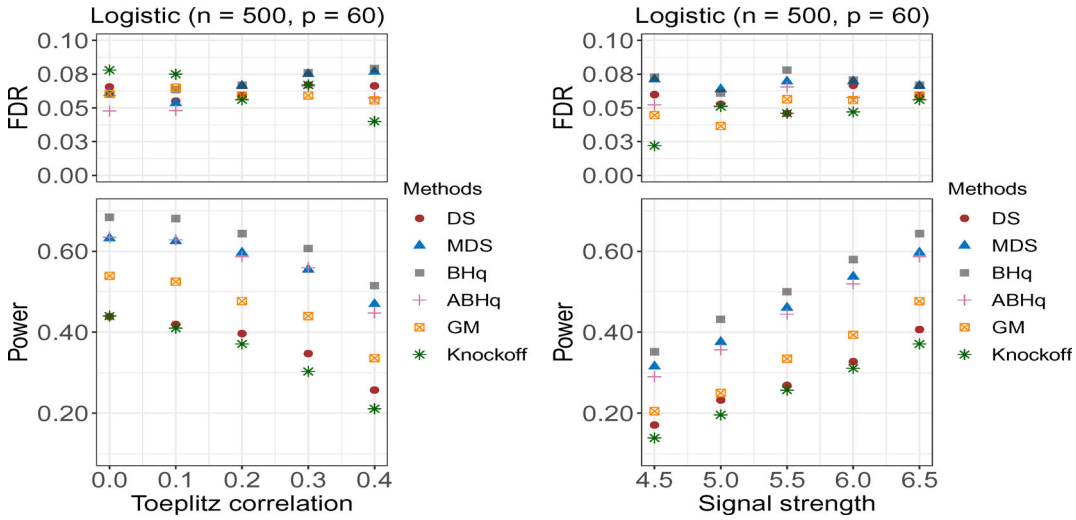

Figure 2. Empirical FDRs and powers for logistic regressions in the small-n-and-p setting. In the left panel, we fix the signal strength at $|\beta_j^\star| = 6.5$ for $j \in S_1$ and vary the correlation r. In the right panel, we fix the correlation at $r = 0.2$ and vary the signal strength. The number of the relevant features is $p_1 = 30$ across all settings. Knockoff features are created using the minimum variance-based reconstructability (MVR) construction.

Remark 4.4. In both moderate and high dimensions, DS and MDS are able to select all relevant features in $S_{1,\text{strong}}$ with probability approaching one. Thus, an immediate lower bound of the power is $p_{1,\text{strong}}/p_1$. See Dai et al. (2022) for more discussions on the power guarantees of DS, MDS (for linear models) and Knockoff.

5. Numerical Illustrations

Recall that the abbreviations DS, MDS, BHq, GM, and Knockoff refer to the single data-splitting method, the multiple data-splitting method, the Benjamini-Hochberg procedure, the Gaussian mirror method, and the model-X knockoff filter, respectively. For DS, we construct the mirror statistic by (2) with $f(u, v) = uv$. For MDS, we replicate DS for 50 times and aggregate the results using Algorithm 2. For Knockoff, we test out the following constructions and report the best results: (a) the second-order construction, including the *equi-correlated* construction and the *asdp*[3] construction; (b) the minimizing reconstructability construction (Spector and Janson 2020), including the *minimum variance-based reconstructability* (MVR) construction and the *maximum entropy* (ME) construction. We assume that the covariance matrix of the features is unknown, and examine two estimators including the Ledoit–Wolf estimator (the python package *knockpy*) and a James-Stein-type shrinkage estimator (the R package *knockoff*). The computational costs of different procedures are summarized in Table B.3 in supplementary materials.

For all the synthetic examples, we set $|\beta_j^\star|$ the same across the relevant features $j \in S_1$ and randomly generate their signs with equal probability. The elements of S_1 are randomly drawn from $\{1, \ldots, p\}$. With a bit abuse of terminology, we refer to $|\beta_j^\star|$ for $j \in S_1$ as the signal strength. The designated FDR control level is set to be $q = 0.1$ henceforth. Each dot in the figures represents the average from 50 independent runs.

[3] *asdp* refers to approximate semidefinite program.

5.1. The Moderate-Dimensional Setting

5.1.1. Logistic Regression

We consider two moderate-dimensional settings for logistic regressions. The first one is the classical small-n-and-p setting, with sample size $n = 500$, dimension $p = 60$, and the dimension-to-sample-size ratio $\kappa = p/n = 0.12$. The second one concerns with the large-n-and-p setting, with $n = 3000$, $p = 500$, and $\kappa = 1/6$. The number of the relevant features p_1 is 30 and 50 in the small-n-and-p and the large-n-and-p settings, respectively. In both settings, we consider six competing methods based on the MLE, including DS, MDS, GM, BHq along with its adjusted version ABHq, and Knockoff. The implementation details of DS and GM are given in Algorithms 3 and 4, respectively. BHq uses the classical p-values calculated via the Fisher information, whereas ABHq is based upon the adjusted p-values derived recently by Sur and Candès (2019).

Each row of the design matrix is independently drawn from $N(0, \Sigma)$ with a Toeplitz correlation structure, that is, $\Sigma_{ij} = r^{|i-j|}$. The variance of each feature is then standardized to be $1/n$. We consider the scenarios with different correlations r and signal strengths. The detailed simulation settings can be found in the captions of Figures 2 and 3. We report additional results for different types of covariance matrix Σ in Section B.1.1 of supplementary materials, including the case where features have constant pairwise (partial) correlation.

Empirical FDRs and powers of different methods under the small-n-and-p setting are summarized in Figure 2. The FDRs of all the six competing methods are under control across all settings. In terms of the power, BHq performs the best in all cases and MDS performs the second best. We see that ABHq is less powerful than BHq, indicating that the asymptotics for the p-value adjustment has not kicked in when n and p are small. Except for MDS, all the perturbation-based methods, such as DS, GM and Knockoff, are not as powerful as the p-value-based methods. A possible reason is that when n and p are small, perturbations employed in these methods may have diluted the signal too much. More interestingly, however, MDS

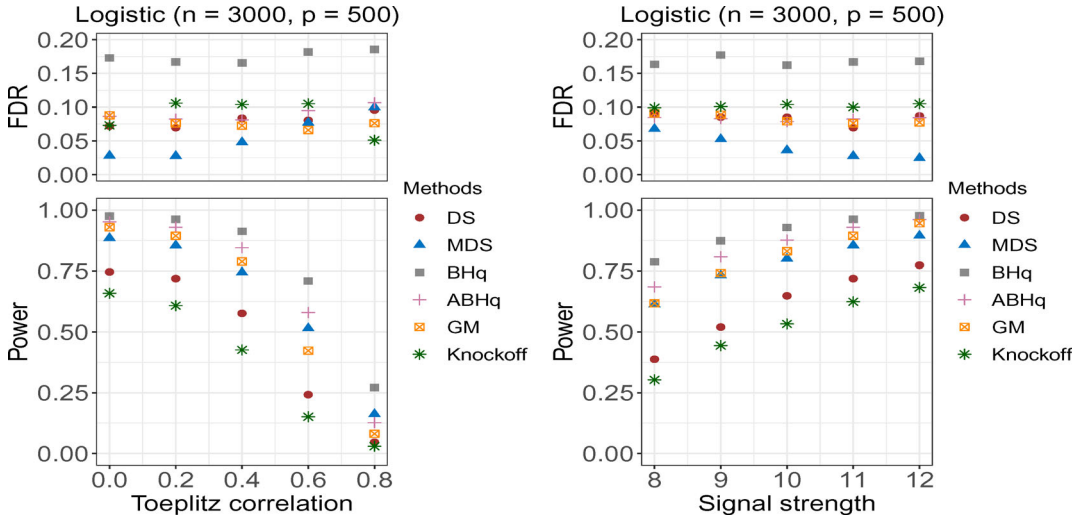

Figure 3. Empirical FDRs and powers for logistic regressions in the large-n-and-p setting. In the left panel, we fix the signal strength at $|\beta_j^\star| = 11$ for $j \in S_1$ and vary the correlation r. In the right panel, we fix the correlation at $r = 0.2$ and vary the signal strength. The number of the relevant features is $p_1 = 50$ across all settings. Knockoff features are created using the minimum variance-based reconstructability (MVR) construction.

gains back almost all the lost power due to sample splitting without sacrificing FDR control.

Results under the large-n-and-p setting are summarized in Figure 3. We see that BHq loses FDR control because the classical p-value under the null, calculated via the Fisher information, is nonuniform and skew to the left. In contrast, ABHq still controls the FDR well and enjoys the highest power, verifying the asymptotics of the MLE derived in Sur and Candès (2019). MDS and GM also perform competitively and have a slightly lower power than ABHq. In particular, GM shows much improved performances compared with the small n-and-p setting.

5.1.2. Negative Binomial Regression

We consider a negative binomial regression model with the dispersion parameter set to be 2, that is, the target number of successful trials is 2. We set sample size $n = 3000$ and dimension $p = 500$, resulting in a dimension-to-sample-size ratio of $\kappa = 1/6$. We simulate the design matrix as described in Section 5.1.1, and test out the scenarios with different correlations r and signal strengths. The detailed simulation settings can be found in the caption of Figure 4. The number of the relevant features is $p_1 = 50$ across all settings. In Section B.1.2 of Supplementary Materials, we report additional results for the case where features have constant pairwise correlation.

We consider five competing methods based on the MLE, including DS, MDS, BHq, GM, and Knockoff. The implementation details of DS and GM are given in Algorithms 3 and 4, respectively. BHq is based upon the classical p-values calculated via the Fisher information, which is known to be incorrect Candès and Sur (2020). However, to the best of our knowledge, the exact asymptotic distribution of the MLE has not been derived and no proper adjustment exists.

The empirical FDRs and powers of different methods are summarized in Figure 4. We see that BHq is the only method losing FDR control because of the nonuniformity (skew to the left) of the p-values under the null. We also tested using the debiased LASSO method to get p-values for BHq and FDRs were still out of control (see Figure B.7 in supplementary materials). Among the methods with FDR control, GM and MDS consistently perform the best over different levels of correlation and signal strength. MDS has a slightly lower power but also a lower FDR compared with GM, and is significantly better than DS in the sense that it simultaneously reduces the FDR and boosts the power. Knockoff has the lowest power among all competing methods for the same reason as discussed in Section 5.1.1.

5.2. The High-Dimensional Setting

5.2.1. Logistic Regression

We consider a case with sample size n=800 and dimension p=2000. Each row of the design matrix is independently drawn from $N(0, \Sigma)$. Following a similar setup as in Ma, Tony Cai, and Li (2020), we let $\Sigma = 0.1 \times \Sigma_B$, where Σ_B is blockwise diagonal consisting of 10 identical unit-diagonal Toeplitz matrices with the correlation factor r=0.3. In Section B.2 of supplementary materials, we give more details about the simulation setup and report additional results for r=0.1. Scenarios with different sparsity levels p_1 and signal strengths are examined, for which details can be found in the caption of Figure 5. We consider four competing methods, including DS, MDS, Knockoff, and the BHq procedure in Ma, Tony Cai, and Li (2020). DS and MDS use the debiased Lasso estimator, with implementation details given in Algorithm 6. For Knockoff, we empirically found that the equi-correlated construction of knockoff features yields the highest power.

The empirical FDRs and powers of different methods are summarized in Figure 5. We see that all methods control the FDR successfully. In terms of the power, MDS is the leading method across different levels of sparsity and signal strength, and has a significantly higher power than DS. Even DS appears to be more powerful than BHq, suggesting that the p-values constructed following Ma, Tony Cai, and Li (2020) can be highly non-informative (skew to the right) in finite-sample cases. Knockoff performs competitively when the signal is sparse, but can potentially suffer when p_1 becomes larger.

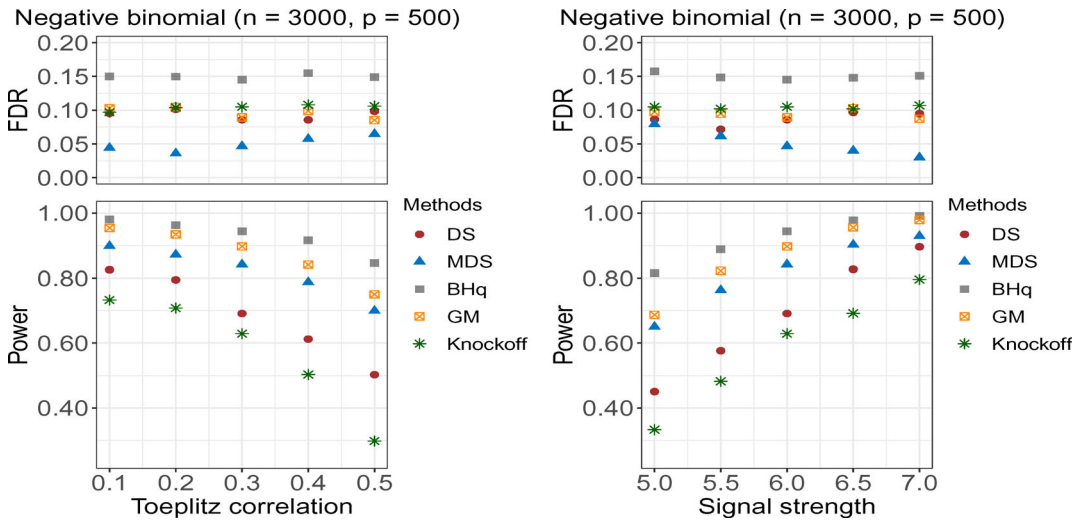

Figure 4. Empirical FDRs and powers for negative binomial regressions. In the left panel, we fix the signal strength at $|\beta_j^\star| = 6$ for $j \in S_1$ and vary the correlation r. In the right panel, we fix the correlation at $r = 0.3$ and vary the signal strength. The number of the relevant features is $p_1 = 50$ across all setting. Knockoff features are created using the minimum variance-based reconstructability (MVR) construction.

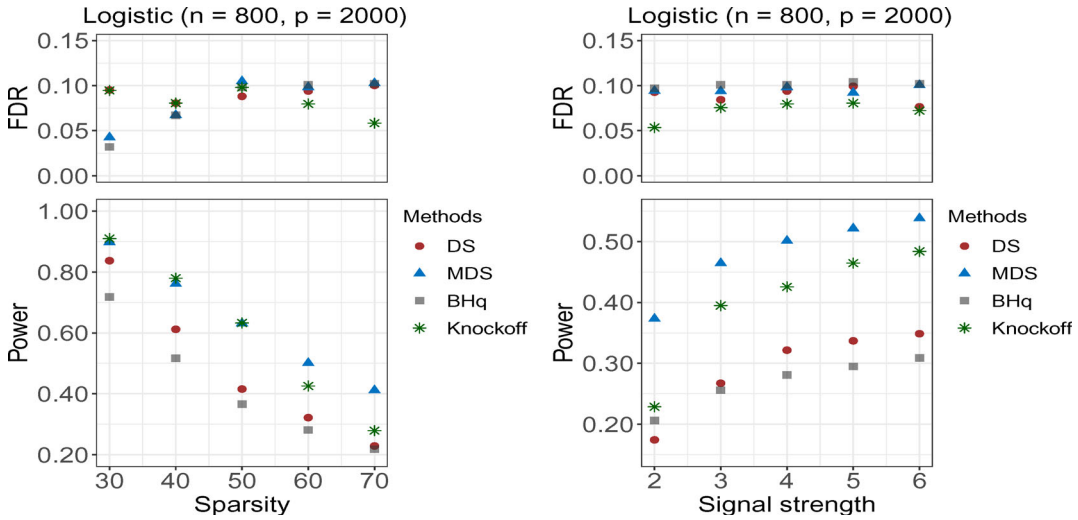

Figure 5. Empirical FDRs and powers for logistic regressions. Left panel: the signal strength is fixed at $|\beta_j^\star|$=4 for $j \in S_1$, and the number of the relevant features p_1 (i.e., sparsity) varies. Right panel: p_1 is fixed at 60 and the signal strength varies. Knockoff features are created using the equi-correlated construction.

5.2.2. Poisson Regression

We consider a case with sample size $n = 800$ and dimension $p = 1600$. In Section B.2.2 of supplementary materials, we report additional results for various values of p. Similar as described in Section 5.2.1, each row of the design matrix is independently drawn from a $(|x_i^\intercal \beta^\star| < 4)$-truncated multivariate normal distribution, with covariance matrix $\Sigma = 0.01 \times \Sigma_B$, where Σ_B is blockwise diagonal consisting of eight identical unit-diagonal matrices, of which all the off-diagonal entries are set to be r. We consider the scenarios with different correlation factors r and signal strengths (see the caption of Figure 6). Methods DS, MDS, and Knockoff are examined. To the best of our knowledge, there is no available BHq procedure for high-dimensional Poisson regression. We follow Algorithm 6 to implement DS and MDS, and use the equi-correlated construction to create knockoff features.

The empirical FDRs and powers of different methods are summarized in Figure 6. All the three competing methods have FDR under control across all settings. In terms of the power, Knockoff achieves the highest power when the correlation among features is small. However, the power of Knockoff decreases rapidly when the correlation becomes larger and can be much lower compared with MDS. Besides, we see that MDS significantly boosts the performance of DS by simultaneously reducing the FDR and increasing the power.

5.3. Real Data Application

Compared with traditional bulk RNA sequencing technologies, single-cell RNA sequencing (scRNAseq) allows researchers to examine the sequence information of each individual cell, thus, promises to advance research in cancer genomics and metagenomics. In this section, we consider selecting the relevant genes with respect to the glucocorticoid response in a human breast cancer cell line, using the scRNAseq data in Hoffman et al. (2020). A total of 400 T47D A1–2 human breast cancer cells were treated with 100 nM synthetic glucocorticoid dexamethasone (Dex) at 1 hr, 2 hr, 4 hr, 8 hr, and 18 hr timestamps. An scRNASeq

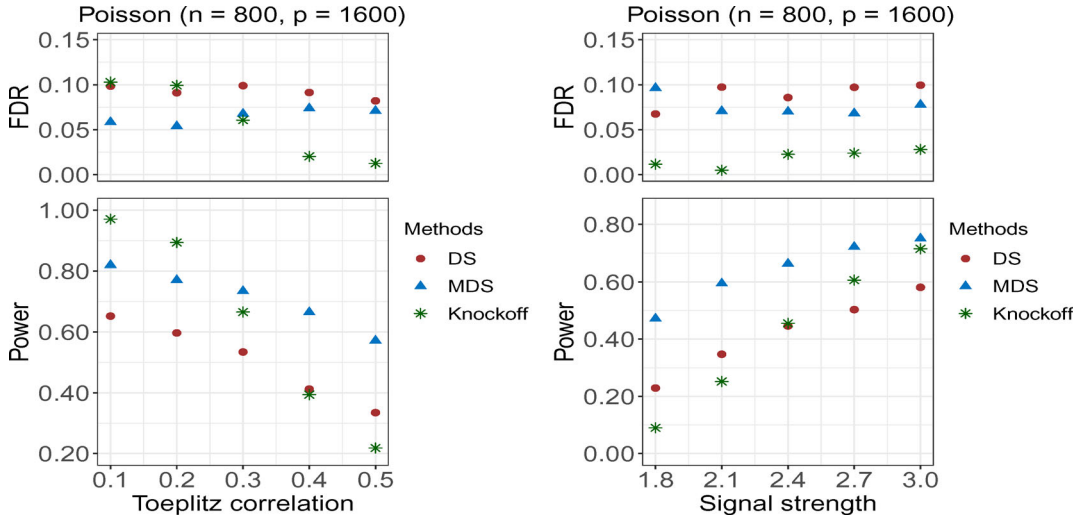

Figure 6. Empirical FDRs and powers for Poisson regressions. Left panel: the signal strength is fixed at $|\beta_j^*|$=2 for $j \in S_1$ and the correlation factor r varies. Right panel: the correlation factor is fixed at r =0.5 and the signal strength varies. The number of the relevant features is p_1=50.

experiment was performed at each timestamp, resulting in a total of 2,000 samples of gene expressions for the treatment group. For the control group, there are 400 vehicle-treated control cells. An scRNAseq experiment was performed at the 18h timestamp to obtain the corresponding profile of gene expressions. After proper normalization, the final scRNAseq data[4] contains 2400 samples, each with 32,049 gene expressions. To further reduce the dimensionality, we first screen out the genes detected in fewer than 10% of cells, and then pick up the top 500 most variable genes following Hoffman et al. (2020).

We consider a logistic regression model with $n = 2400$ and $p = 500$. Since the MLE does not exist uniquely on this data, we cannot use the method in Sur and Candès (2019) to obtain p-values. Instead, we apply DS outlined in Algorithm 6 using the debiased Lasso estimator. As sample size n is larger than dimension p, we directly estimate the precision matrix Θ (see (18) and the discussion therein) by inverting the sample Hessian matrix $\widehat{\Sigma}$. We then replicate DS for 500 times and aggregate the results using the MDS procedure outlined in Algorithm 2. Table 1 summarizes the selected genes by MDS, BHq (Ma, Tony Cai, and Li 2020), Knockoff, and three de-randomized Knockoff procedures. BHq only selects 1 gene, RPL10, which is also selected by MDS.

We run Knockoff for 500 times and report the genes whose selection frequencies are above 0.5. The set of the selected genes appears unstable. Specifically, the highest selection frequency is only 0.75, and the size of the selection set ranges from 0 to 53 with a mean of 16 (see Figure B.12 in supplementary materials). As shown in Table 1, there are 17 genes whose selection frequency by Knockoff is above 0.5, among which 14 are also selected by MDS. We also found that Knockoff does not select any genes for approximately 20% of the 500 runs. Further, FDR control is no longer guaranteed for the aggregated list of the genes selected by multiple Knockoff runs.

To stabilize Knockoff, we test out three de-randomized Knockoff procedures proposed in the following papers.

[4]The data is available at *https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE141834*.

Table 1. Genes selected by MDS, Knockoff and de-randomized Knockoff procedures (Gimenez and Zou 2019; Nguyen et al. 2020; Ren, Wei, and Candès 2020).

Gene	MDS	Knockoff	Nguyen et al.	Gimenez and Zou	Ren et al.		
					$k=1$	$k=2$	$k=3$
NFKBIA	✓	0.75	✓	0.59	✓	✓	✓
HSPB1	✓	0.71	✓		✓	✓	✓
LY6E	✓	0.66	✓	0.41	✓	✓	✓
BLOC1S1	✓	0.65	✓	0.34		✓	✓
IGFBP4	✓	0.64	✓		✓	✓	✓
ATF4	✓	0.64	✓			✓	✓
SERPINA6	✓	0.63	✓			✓	✓
RPL10	✓	0.61	✓		✓	✓	✓
DDIT4	✓	0.60	✓			✓	✓
EEF1A1	✓	0.59	✓		✓	✓	✓
SEMA3C	✓	0.58	✓			✓	✓
EIF4EBP1	✓	0.57	✓		✓	✓	✓
FKBP5	✓	0.54	✓		✓	✓	✓
NUPR1	✓	0.51				✓	✓
DSCAM-AS1	✓		✓			✓	✓
BCL6	✓					✓	✓
HSPA1A	✓						
KRT19		0.52	✓			✓	✓
MSX2		0.51		0.40		✓	✓
S100A11		0.50				✓	✓
RPLP0P6							✓
UHMK1							✓

NOTE: For Knockoff and the method in Gimenez and Zou (2019), we report the genes with selection frequencies above 0.5 and 0.3 among 500 independent runs, respectively. For MDS and the methods in Nguyen et al. (2020) and Ren, Wei, and Candès (2020), we repeat the corresponding base procedure for 50 times and aggregate the selection results. The de-randomized Knockoff procedure in Ren, Wei, and Candès (2020) controls the k family-wise error rate, and we report the selection results for $k \in \{1, 2, 3\}$.

1. Ren, Wei, and Candès (2020). The method repeats the base procedure ν-Knockoff (Janson and Su 2016) and controls the k family-wise error rate, that is, the number of the false discoveries, rather than the FDR. Table 1 lists the selected genes with $k \in \{1, 2, 3\}$. Compared to the selection results of MDS, the set of the selected genes with $k = 2$ are similar, while the selection results with $k = 1$ and $k = 3$ are more and less conservative, respectively.
2. Nguyen et al. (2020). Based on "intermediate p-values," the method summarizes the selection results of multiple Knockoff runs using quantile aggregation (Meinshausen, Meier, and

Table 2. The references in support of the selected genes.

Gene	References
NFKBIA	Auphan et al. (1995); Deroo and Archer (2001)
HSPB1	Barr and Dokas (1999); Tuckermann et al. (1999)
LY6E	–
BLOC1S1	–
IGFBP4	Cheung et al. (1994); Okazaki, Riggs, and Conover (1994); Conover, Clarkson, and Bale (1995)
ATF4	Adams (2007)
SERPINA6	GeneCards-SERPINA6, Zhou et al. (2008)
RPL10	Zorzatto et al. (2015)
DDIT4	Wang et al. (2003); Boldizsár et al. (2006); Wolff, McKay, and Brugarolas (2014)
EEF1A1	NCBI-EEF1A1
SEMA3C	Williamson, Garg, and Wells (2020)
EIF4EBP1	Watson et al. (2012)
FKBP5	AGCOH-FKBP5, Nair et al. (1997)
NUPR1	Wikigenes-NUPR1, Mukaida et al. (1994)
DSCAM-AS1	Zhao et al. (2016); Chen and Cai (2020)
BCL6	Goodman et al. (2016)
HSPA1A	NCBI-Gene, Kirschke et al. (2014)
KRT19	Romanò et al. (2020)
MSX2	Jaskoll, Luo, and Snead (1998)
S100A11	–
RPLP0P6	–
UHMK1	–

NOTE: There are five genes that we do not find direct supporting evidence in the existing literature for their interactions with the glucocorticoid receptor (GR), which might be of interest for further investigations. The red hyperlinks point to the documented information of the corresponding genes in some widely referred databases including GeneCards, Strings, Wikigenes, the National Center for Biotechnology Information (NCBI), and Atlas of Genetics and Cytogenetics in Oncology and Harmatology (AGCOH).

Bühlmann 2009). As shown in Table 1, it selects 15 genes, among which 14 are also selected by MDS.

3. Gimenez and Zou (2019). The method simultaneously samples two knockoff copies in the base Knockoff procedure in order to stabilize the selection result. We independently repeat the procedure for 500 times and report the selection frequency of each gene. As shown in Table 1, there is only 1 gene with selection frequency above 0.5, and only four genes with selection frequency above 0.3. The selection result becomes more conservative when we further increase the number of knockoff copies.

The existing literature confirms interactions between the glucocorticoid receptor (GR) and a majority of the genes selected by MDS. A summary of the references associated with the selected genes are given in Table 2, and we highlight some of the supporting evidences in Section B.3 of supplementary materials. Curiously, we do not find any relevant literature documenting how GR may interact with genes LY6E and BLOC1S1, which have been consistently selected by most of the tested methods and can be of interest for further investigations. Figure B.11 in supplementary materials demonstrates the sharp difference in the gene expression distributions between the treatment group and the control group for four genes: NFKBIA, EEF1A1, FKBP5, and RPL10, all of which are selected by MDS and having a Knockoff selection frequency above 0.5.

6. Conclusion

We have described a general framework for feature selection in GLMs with FDR asymptotically under control. In particular, we detail the constructions of the mirror statistic under two asymptotic regimes, that is, the moderate-dimensional setting ($p/n \to \kappa \in (0, 1)$) and the high-dimensional setting ($p \gg n$). Compared to BHq, the proposed methodology enjoys a wider applicability and improved robustness due to its scale-free property. Compared to Knockoff, it does not require the knowledge of the joint distribution of features and is less affected by the correlations among features.

We conclude by pointing out several directions for future work. First, it is of immediate interest to generalize the proposed methods to handle cases where subsets of explanatory features exhibit group structures. Second, we would like to investigate the applicability of our FDR control framework to dependent observations (e.g., stationary time series data). These two types of data structures appear a lot in practice including genetic studies and financial engineering. Third, moving beyond parametric models, we can consider the FDR control problem in semi-parametric single-index models, in which the link function becomes unknown.

Supplementary Materials

The supplementary materials include detailed proofs of all the main theorems and some additional simulation results.

ORCID

Xin Xing https://orcid.org/0000-0001-9121-0086
Jun S. Liu https://orcid.org/0000-0002-4450-7239

References

Abbasi, E. (2020), "Universality Laws and Performance Analysis of the Generalized Linear Models," Dissertation Ph.D., California Institute of Technology. [1554]

Adams, C. M. (2007), "Role of the Transcription Factor ATF4 in the Anabolic Actions of Insulin and the Anti-anabolic Actions of Glucocorticoids," *Journal of Biological Chemistry*, 282, 16744–16753. [1563]

Auphan, N., DiDonato, J. A., Rosette, C., Helmberg, A., and Karin, M. (1995), "Immunosuppression by Glucocorticoids: Inhibition of NF-kappa B Activity through Induction of I kappa B Synthesis," *Science*, 270, 286–290. [1563]

Barber, R. F., and Candès, E. J. (2015), "Controlling the False Discovery Rate via Knockoffs," *The Annals of Statistics*, 43, 2055–2085. [1553]

Barber, R. F., Candès, E. J., and Samworth, R. J. (2020), "Robust Inference with Knockoffs," *The Annals of Statistics*, 48, 1409–1431. [1551]

Barr, C. S., and Dokas, L. A. (1999), "Glucocorticoids Regulate the Synthesis of HSP27 in Rat Brain Slices," *Brain Research*, 847, 9–17. [1563]

Bates, S., Candés, E. J., Janson, L., and Wang, W. (2020), "Metropolized Knockoff Sampling," *Journal of the American Statistical Association*, 116, 1413–1427. [1551]

Bates, S., Candès, E., Lei, L., Romano, Y., and Sesia, M. (2021), "Testing for Outliers with Conformal p-values," arXiv preprint: 2104.08279. [1551]

Benjamini, Y., and Hochberg, Y. (1995), "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society*, Series B, 57, 289–300. [1551]

Bickel, P. J., Ritov, Y., and Tsybakov, A. B. (2009), "Simultaneous Analysis of Lasso and Dantzig Selector," *The Annals of Statistics*, 37, 1705–1732. [1558]

Boldizsár, F., Pálinkás, L., Czömpöly, T., Bartis, D., Németh, P., and Berki, T. (2006), "Low Glucocorticoid Receptor (GR), high Dig2 and low Bcl-2 Expression in Double Positive Thymocytes of BALB/c mice Indicates their Endogenous Glucocorticoid Hormone Exposure," *Immunobiology*, 211, 785–796. [1563]

Bühlmann, P., and Mandozzi, J. (2014), "High-Dimensional Variable Screening and Bias in Subsequent Inference, with an Empirical Comparison," *Computational Statistics*, 29, 407–430. [1558]

Candès, E., Fan, Y., Janson, L., and Lv, J. (2018), "Panning for Gold: Model-X Knockoffs for High Dimensional Controlled Variable Selection," *Journal of the Royal Statistical Society*, Series B, 80, 551–577. [1551]

Candès, E. J., and Sur, P. (2020), "The Phase Transition for the Existence of the Maximum Likelihood Estimate in High-Dimensional Logistic Regression," *The Annals of Statistics*, 48, 27–42. [1554,1560]

Chen, H., and Cai, K. (2020), "DSCAM-AS1 Mediates Pro-hypertrophy Role of GRK2 in Cardiac Hypertrophy Aggravation via Absorbing miR-188-5p," *In Vitro Cellular and Developmental Biology. Animal*, 56, 286–295. [1563]

Cheung, P. T., Wu, J., Banach, W., and Chernausek, S. D. (1994), "Glucocorticoid Regulation of an Insulin-like Growth Factor-Binding Protein-4 Protease Produced by a Rat Neuronal Cell Line," *Endocrinology*, 135, 1328–1335. [1563]

Conover, C. A., Clarkson, J. T., and Bale, L. K. (1995), "Effect of Glucocorticoid on Insulin-like Growth Factor (IGF) Regulation of IGF-Binding Protein Expression in Fibroblasts," *Endocrinology*, 136, 1403–1410. [1563]

Cover, T. M. (1964), "Geometrical and Statistical Properties of Linear Threshold Devices," Ph.D. thesis. [1556]

——— (1965), "Geometrical and Statistical Properties of Systems of Linear Inequalities with Applications in Pattern Recognition," *IEEE Transactions on Electronic Computers*, 14, 326–334. [1556]

Dai, C., Lin, B., Xing, X., and Liu, J. S. (2022), "False Discovery Rate Control via Data Splitting," *Journal of the American Statistical Association*, arXiv preprint 2002.08542. [1551,1553,1554,1558,1559]

Deroo, B. J., and Archer, T. K. (2001), "Glucocorticoid Receptor Activation of the I kappa B alpha Promoter Within Chromatin," *Molecular Biology of the Cell*, 12, 3365–3374. [1563]

Gimenez, J. R., and Zou, J. (2019), "Improving the Stability of the Knockoff Procedure: Multiple Simultaneous Knockoffs and Entropy Maximization," in *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, Volume 89 of Proceedings of Machine Learning Research, eds. K. Chaudhuri and M. Sugiyama, pp. 2184–2192. PMLR. [1562,1563]

Goodman, C. R., Sato, T., Peck, A. R., Girondo, M. A., Yang, N., Liu, C., Yanac, A. F., Kovatich, A. J., Hooke, J. A., and Shriver, C. D. (2016), "Steroid Induction of Therapy-Resistant Cytokeratin-5-Positive Cells in Estrogen Receptor-Positive Breast Cancer through a BCL6-dependent Mechanism," *Oncogene*, 35, 1373–1385. [1563]

Hoffman, J. A., Papas, B. N., Trotter, K. W., and Archer, T. K. (2020), "Single-Cell RNA Sequencing Reveals a Heterogeneous Response to Glucocorticoids in Breast Cancer Cells," *Communications Biology*, 3, 1–11. [1561,1562]

Huang, D., and Janson, L. (2020), "Relaxing the Assumptions of Knockoffs by Conditioning," *The Annals of Statistics*, 48, 3021–3042. [1551]

Janson, L., and Su, W. (2016), "Familywise Error Rate Control via Knockoffs," *Electronic Journal of Statistics*, 10, 960–975. [1562]

Jaskoll, T., Luo, W., and Snead, M. L. (1998), "Msx-2 Expression and Glucocorticoid-Induced Overexpression in Embryonic Mouse Submandibular Glands," *Journal of Craniofacial Genetics and Developmental Biology*, 18, 79–87. [1563]

Javanmard, A., and Javadi, H. (2019), "False Discovery Rate Control via Debiased Lasso," *Electronic Journal of Statistics*, 13, 1212–1253. [1551]

Javanmard, A., and Montanari, A. (2013), "Nearly Optimal Sample Size in Hypothesis Testing for High-Dimensional Regression," in *2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 1427–1434. IEEE. [1557,1558]

Javanmard, A., and Montanari, A. (2014), "Confidence Intervals and Hypothesis Testing for High-Dimensional Regression," *Journal of Machine Learning Research*, 15, 2869–2909. [1551,1557]

Jordon, J., Yoon, J., and Schaar, M. V. D. (2019), "KnockoffGAN: Generating Knockoffs for Feature Selection Using Generative Adversarial Networks," in *International Conference on Learning Representations*. [1551]

Ke, Z. T., Liu, J. S., and Ma, Y. (2020), "Power of FDR Control Methods: The Impact of Ranking Algorithm, Tampered Design, and Symmetric Statistic," arXiv preprint: 2010.08132. [1551,1553]

Kirschke, E., Goswami, D., Southworth, D., Griffin, P. R., and Agard, D. A. (2014), "Glucocorticoid Receptor Function Regulated by Coordinated Action of the Hsp90 and Hsp70 Chaperone Cycles," *Cell*, 157, 1685–1697. [1563]

Liu, W. (2013), "Gaussian Graphical Model Estimation with False Discovery Rate Control," *The Annals of Statistics*, 41, 2948–2978. [1558]

Ma, R., Tony Cai, T., and Li, H. (2020), "Global and Simultaneous Hypothesis Testing for High-Dimensional Logistic Regression Models," *Journal of the American Statistical Association*, 116, 984–998. [1551,1552,1558,1560,1562]

Marandon, A., Lei, L., Mary, D., and Roquain, E. (2022), "Machine Learning Meets False Discovery Rate," arXiv preprint: 2208.06685. [1551]

Meinshausen, N., Meier, L., and Bühlmann, P. (2009), "*p*-values for High-Dimensional Regression," *Journal of the American Statistical Association*, 104, 1671–1681. [1563]

Mukaida, N., Morita, M., Ishikawa, Y., Rice, N., Okamoto, S., Kasahara, T., and Matsushima, K. (1994), "Novel Mechanism of Glucocorticoid-Mediated Gene Repression. Nuclear Factor-kappa B is Target for Glucocorticoid-Mediated Interleukin 8 Gene Repression," *Journal of Biological Chemistry*, 269, 13289–13295. [1563]

Nair, S. C., Rimerman, R. A., Toran, E. J., Chen, S., Prapapanich, V., Butts, R. N., and Smith, D. F. (1997), "Molecular Cloning of Human FKBP51 and Comparisons of Immunophilin Interactions with HSP90 and Progesterone Receptor," *Molecular and Cellular Biology*, 17, 594–603. [1563]

Nguyen, T.-B., Chevalier, J.-A., Thirion, B., and Arlot, S. (2020), "Aggregation of Multiple Knockoffs," in *Proceedings of the 37th International Conference on Machine Learning*, Volume 119 of Proceedings of Machine Learning Research, eds. H. D. III and A. Singh, pp. 7283–7293. PMLR. [1562]

Okazaki, R., Riggs, B. L., and Conover, C. A. (1994), "Glucocorticoid Regulation of Insulin-Like Growth Factor-Binding Protein Expression in Normal Human Osteoblast-Like Cells," *Endocrinology*, 134, 126–132. [1563]

Ren, Z., Wei, Y., and Candès, E. (2020), "Derandomizing Knockoffs," arXiv preprint: 2012.02717. [1562]

Romanò, N., Duncan, P., McClafferty, H., Nolan, O., Ding, Q., Homer, N., Le Tissier, P., Walker, B., Shipston, M., and Chambers, T. (2020), "Dissection of the Corticotroph Transcriptome in a Mouse Model of Glucocorticoid-Induced Suppression of the HPA Axis," bioRxiv. [1563]

Romano, Y., Sesia, M., and Candès, E. J. (2019), "Deep Knockoffs," *Journal of the American Statistical Association*, 115, 1861–1872. [1551]

Salehi, F., Abbasi, E., and Hassibi, B. (2019), "The Impact of Regularization on High-Dimensional Logistic Regression," in *Advances in Neural Information Processing Systems*, pp. 11982–11992. [1554,1555]

Spector, A., and Janson, L. (2020), "Powerful Knockoffs via Minimizing Reconstructability," arXiv preprint: 2011.14625. [1559]

Sur, P., and Candès, E. J. (2019), "A Modern Maximum-Likelihood Theory for High-Dimensional Logistic Regression," *Proceedings of the National Academy of Sciences*, 116, 14516–14525. [1552,1555,1556,1559,1560,1562]

Taylor, J. E., and Tibshirani, R. (2018), "Post-Selection Inference for ℓ_1-penalized Likelihood Models," *Canadian Journal of Statistics*, 46, 41–61. [1558]

Tuckermann, J. P., Reichardt, H. M., Arribas, R., Richter, K. H., Schütz, G., and Angel, P. (1999), "The DNA Binding-Independent Function of the Glucocorticoid Receptor Mediates Repression of AP-1–dependent Genes in Skin," *Journal of Cell Biology*, 147, 1365–1370. [1563]

Van de Geer, S., Bühlmann, P., Ritov, Y., and Dezeure, R. (2014), "On Asymptotically Optimal Confidence Regions and Tests for High-Dimensional Models," *The Annals of Statistics*, 42, 1166–1202. [1551,1557,1558]

Wang, Z., Malone, M. H., Thomenius, M. J., Zhong, F., Xu, F., and Distelhorst, C. W. (2003), "Dexamethasone-Induced Gene 2 (Dig2) is a Novel Pro-survival Stress Gene Induced Rapidly by Diverse Apoptotic Signals," *Journal of Biological Chemistry*, 278, 27053–27058. [1563]

Watson, M. L., Baehr, L. M., Reichardt, H. M., Tuckermann, J. P., Bodine, S. C., and Furlow, J. D. (2012), "A Cell-Autonomous Role for the Glucocorticoid Receptor in Skeletal Muscle Atrophy Induced by Systemic Glucocorticoid Exposure," *American Journal of Physiology-Endocrinology and Metabolism*, 302, E1210–E1220. [1563]

Weinstein, A., Barber, R., and Candes, E. (2017), "A Power and Prediction Analysis for Knockoffs with Lasso Statistics," arXiv preprint: 1712.06465. [1551]

Weinstein, A., Su, W. J., Bogdan, M., Barber, R. F., and Candès, E. J. (2020), "A Power Analysis for Knockoffs with the Lasso Coefficient-Difference Statistic," arXiv preprint arXiv:2007.15346. [1551]

Williamson, M., Garg, R., and Wells, C. M. (2020), "PlexinB1 Promotes Nuclear Translocation of the Glucocorticoid Receptor," *Cells*, 9, 3. [1563]

Wolff, N. C., McKay, R. M., and Brugarolas, J. (2014), "REDD1/DDIT4-Independent mTORC1 Inhibition and Apoptosis by Glucocorticoids in Thymocytes," *Molecular Cancer Research*, 12, 867–877. [1563]

Xing, X., Zhao, Z., and Liu, J. S. (2019), "Controlling False Discovery Rate using Gaussian Mirrors," arXiv preprint: 1911.09761. [1553,1558]

Yang, C.-Y., Lei, L., Ho, N., and Fithian, W. (2021), "Bonus: Multiple Multivariate Testing with a Data-Adaptivetest Statistic." [1551]

Zhang, C. H., and Zhang, S. S. (2014), "Confidence Intervals for Low Dimensional Parameters in High Dimensional Linear Models," *Journal of the Royal Statistical Society*, Series B, 76, 217–242. [1551,1557]

Zhao, Q., Sur, P., and Candès, E. J. (2020), "The Asymptotic Distribution of the MLE in High-Dimensional Logistic Models: Arbitrary Covariance," arXiv preprint: 2001.09351. [1554,1555]

Zhao, W., Wang, D., Liu, C., and Zhao, X. (2016), "G-protein-Coupled Receptor Kinase 2 Terminates G-protein-coupled Receptor Function in Steroid Hormone 20-hydroxyecdysone Signaling," *Scientific Reports*, 6, 1–13. [1563]

Zhou, A., Wei, Z., Stanley, P. L. D., Read, R. J., Stein, P. E., and Carrell, R. W. (2008), "The S-to-R Transition of Corticosteroid-Binding Globulin and the Mechanism of Hormone Release," *Journal of Molecular Biology*, 380, 244–251. [1563]

Zorzatto, C., Machado, J. P. B., Lopes, K. V. G., Nascimento, K. J. T., Pereira, W. A., Brustolini, O. J. B., Reis, P. A. B., Calil, I. P., Deguchi, M., and Sachetto-Martins, G. (2015), "NIK1-Mediated Translation Suppression Functions as a Plant Antiviral Immunity Mechanism," *Nature*, 520, 679–682. [1563]