AMDP: An Adaptive Detection Procedure for False Discovery Rate Control in High-Dimensional Mediation Analysis

Appendix

A Selection of an appropriate λ

In Section 2.2, we estimate $\hat{\pi}_{00}(\lambda)$, $\hat{\pi}_{10}(\lambda)$, and $\hat{\pi}_{01}(\lambda)$ using the method proposed by Storey et al. [17] with a fixed parameter λ. The estimators are calculated in (10)-(11). Theoretical considerations suggest that as λ approaches 1, the estimators of the composite null hypothesis become more accurate asymptotically. However, in finite samples scenarios, with a larger value of λ, the chance of these null p-values falling within $(\lambda, 1]$ gets smaller, resulting in less accurate estimates. Conversely, when λ becomes smaller, the bias of the null estimators increases while the variance decreases [17]. Consequently, there exists an inherent bias-variance trade-off in the selection of λ.

To strike a reasonable balance between bias and variance, we aim to determine λ by minimizing the mean-squared error (MSE) of the estimators. The MSE is defined as $E[\{\hat{\pi}_{00}(\lambda) - \pi_{00}\}^2 + \{\hat{\pi}_{10}(\lambda) - \pi_{10}\}^2 + \{\hat{\pi}_{01}(\lambda) - \pi_{01}\}^2]$. For achieving this goal, we consider a range of cutpoints for λ (e.g., $\lambda = 0.1, 0.2, \dots, 0.9$) and calculate the MSE for each value of λ. As highlighted by Barfield et al. [1], a substantial proportion of null hypotheses may exhibit both $\alpha = 0$ and $\beta = 0$ in a genome-wide study involving high-dimensional mediation hypotheses. To investigate the choice of λ in such scenarios, we consider the following settings, as shown in Table A1.

Table A1: The composite null proportions under different scenarios.

Hypothesis Configuration	π_{00}	π_{10}	π_{01}	π_{11}
Scenario 1	0.2	0.3	0.3	0.2
Scenario 2	0.4	0.2	0.2	0.2
Scenario 3	0.5	0.2	0.2	0.1
Scenario 4	0.6	0.15	0.15	0.1
Scenario 5	0.75	0.1	0.1	0.05
Scenario 6	0.85	0.05	0.05	0.05

Table A2 shows means and MSE of the estimated null proportions under the six scenarios with n=1000, J=10000, α_j= 0.2, and β_j=0.3. In each scenario, we identified the top three smallest MSEs among the estimated null proportions. Notably, we observed that the optimal value of λ varied across the different scenarios. However, it is noteworthy that $\lambda = 0.5$ consistently appeared among the top three MSE values in all the simulated scenarios. As mentioned earlier, smaller values of λ tend to result in larger biases of the null estimate, while excessively large values of λ may yield inaccurate estimates in finite sample scenarios. Therefore, the consistent appearance of $\lambda = 0.5$ among the top-performing MSE values suggests that it provides a reasonable trade-off between bias and variance, leading to accurate estimates across a wide range of scenarios. Considering its stability and computational efficiency, we believe that $\lambda = 0.5$ is a suitable choice for estimating the null proportions in high-dimensional mediation analysis.

B Comparison with existing methods

In this section, we demonstrate the loss of information during the ranking step can result in decreased statistical power, despite controlling the FDR at the desired level.

Table A2: The performance of the estimated proportions of the composite null hypothesis (mean and MSE) under the six scenarios. The tuning parameter λ varies from 0.1 to 0.9.

Scenario 1	$\hat{\pi}_{00}$	$\hat{\pi}_{10}$	$\hat{\pi}_{01}$	MSE	Scenario 2	$\hat{\pi}_{00}$	$\hat{\pi}_{10}$	$\hat{\pi}_{01}$	MSE
λ=0.1	0.222	0.278	0.314	1.17e-3	λ=0.1	0.415	0.185	0.215	6.41e-4
λ=0.2	0.211	0.289	0.308	3.10e-4	λ=0.2	0.408	0.192	0.208	1.88e-4
λ=0.3	0.207	0.293	0.306	1.20e-4	λ=0.3	0.406	0.194	0.205	9.14e-5
λ=0.4	0.205	0.295	0.305	6.69e-5	λ=0.4	0.404	0.196	0.204	**4.61e-5**
λ=0.5	0.203	0.296	0.303	**3.03e-5**	λ=0.5	0.402	0.196	0.203	**2.56e-5**
λ=0.6	0.203	0.296	0.303	3.25e-5	λ=0.6	0.403	0.197	0.202	**2.46e-5**
λ=0.7	0.202	0.297	0.302	**1.96e-5**	λ=0.7	0.406	0.194	0.198	7.62e-5
λ=0.8	0.201	0.299	0.302	**5.94e-6**	λ=0.8	0.405	0.195	0.199	5.96e-5
λ=0.9	0.206	0.295	0.296	8.54e-5	λ=0.9	0.405	0.192	0.198	8.69e-5
Scenario 3	$\hat{\pi}_{00}$	$\hat{\pi}_{10}$	$\hat{\pi}_{01}$	MSE	Scenario 4	$\hat{\pi}_{00}$	$\hat{\pi}_{10}$	$\hat{\pi}_{01}$	MSE
λ=0.1	0.515	0.185	0.207	4.84e-4	λ=0.1	0.611	0.139	0.157	2.94e-4
λ=0.2	0.507	0.192	0.204	1.38e-4	λ=0.2	0.606	0.144	0.154	8.07e-5
λ=0.3	0.505	0.194	0.203	6.82e-5	λ=0.3	0.604	0.146	0.152	3.79e-5
λ=0.4	0.504	0.196	0.201	3.81e-5	λ=0.4	0.602	0.148	0.152	**1.18e-5**
λ=0.5	0.503	0.197	0.201	**1.91e-5**	λ=0.5	0.602	0.148	0.151	**7.87e-6**
λ=0.6	0.502	0.196	0.202	2.24e-5	λ=0.6	0.603	0.147	0.150	1.79e-5
λ=0.7	0.503	0.195	0.199	3.15e-5	λ=0.7	0.604	0.148	0.148	2.12e-5
λ=0.8	0.501	0.197	0.202	**1.36e-5**	λ=0.8	0.604	0.147	0.147	2.92e-5
λ=0.9	0.501	0.198	0.201	**6.69e-6**	λ=0.9	0.602	0.150	0.150	**3.19e-6**
Scenario 5	$\hat{\pi}_{00}$	$\hat{\pi}_{10}$	$\hat{\pi}_{01}$	MSE	Scenario 6	$\hat{\pi}_{00}$	$\hat{\pi}_{10}$	$\hat{\pi}_{01}$	MSE
λ=0.1	0.757	0.093	0.104	1.22e-4	λ=0.1	0.854	0.046	0.053	4.12e-5
λ=0.2	0.754	0.096	0.103	3.69e-5	λ=0.2	0.853	0.048	0.052	1.25e-5
λ=0.3	0.753	0.097	0.102	2.22e-5	λ=0.3	0.853	0.048	0.050	1.19e-5
λ=0.4	0.751	0.099	0.102	**6.94e-6**	λ=0.4	0.852	0.049	0.050	**3.79e-6**
λ=0.5	0.752	0.099	0.101	**3.38e-6**	λ=0.5	0.850	0.049	0.051	**9.69e-7**
λ=0.6	0.754	0.097	0.099	2.35e-5	λ=0.6	0.851	0.050	0.051	**1.40e-6**
λ=0.7	0.755	0.096	0.098	4.53e-5	λ=0.7	0.849	0.052	0.052	1.16e-5
λ=0.8	0.751	0.098	0.101	**5.44e-6**	λ=0.8	0.848	0.054	0.056	6.07e-5
λ=0.9	0.765	0.091	0.090	4.09e-4	λ=0.9	0.857	0.056	0.055	1.11e-4

To conduct this investigation, we consider three different approaches for comparison: our proposed AMDP, along with two existing methods, the JS-mixture [5], and the DACT [14]. During the selection step, it is assumed that the information about the proportions of the composite null hypothesis and the distributions of p-values under alternatives are known. This provided knowledge allows for effectively controlling the FDR of all three procedures at the predefined level of α in the selection step. With the FDR under control, we then proceed to investigate how different ranking strategies impact the power performance of the three methods. To achieve this, we compare the rejection regions of each method under various scenarios. Formally, the ranking statistic for each method is as follows:

$$\begin{aligned} \delta^{JS-mixture} &= p_{max}, \\ \delta^{DACT} &= \omega_1 p^{(1)} + \omega_2 p^{(2)} + \omega_3 p_{max}, \\ \delta^{AMDP} &= \text{fdr}(p), \end{aligned}$$

where $p_{max} = p^{(1)} \vee p^{(2)}$, $\vee$ denotes the maximum of the two p-values. ω_1, ω_2 and ω_3 are normalized relative proportions of the composite null. We consider two scenarios to compare the ranking statistic of the three methods:

Scenario 1 Balanced null proportions of H_{01} and H_{10}:

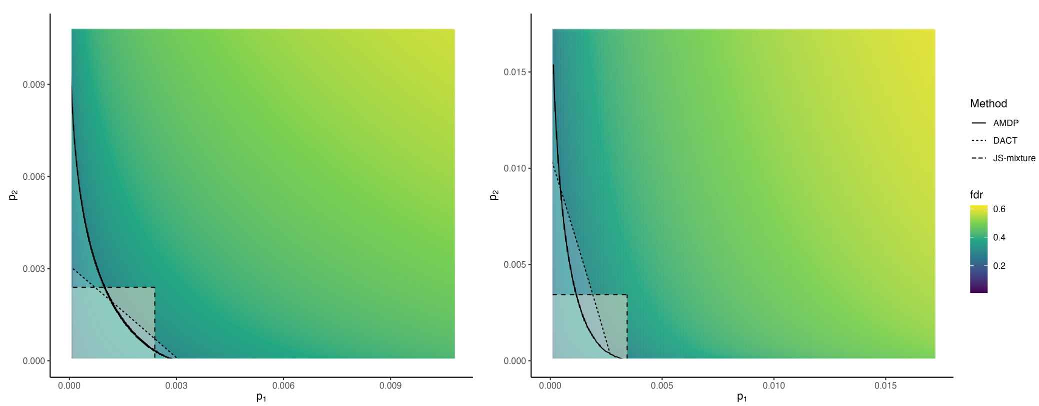
$$f(p^{(1)}, p^{(2)}) = 0.49 + 0.21 \times 0.6 {p^{(1)}}^{-0.4} + 0.21 \times 0.3 {p^{(2)}}^{-0.7} + 0.09 \times 0.18 {p^{(1)}}^{-0.4} {p^{(2)}}^{-0.7},$$

where the density functions of p-values under alternatives are $f(p^{(1)} \mid H_{10}) \sim Beta(0.6, 1)$ and $f(p^{(2)} \mid H_{01}) \sim Beta(0.3, 1)$. The proportions of composite hypothesis are $\pi_{00} = 0.49$, $\pi_{01} = \pi_{10} = 0.21$, and $\pi_{11} = 0.09$.

Scenario 2 Unbalanced null proportions of H_{01} and H_{10}:

$$f(p^{(1)}, p^{(2)}) = 0.4 + 0.1 \times 0.4 {p^{(1)}}^{-0.6} + 0.4 \times 0.6 {p^{(2)}}^{-0.4} + 0.1 \times 0.24 {p^{(1)}}^{-0.6} {p^{(2)}}^{-0.4},$$

where the density functions of p-values under alternatives are $f(p^{(1)} \mid H_{10}) \sim Beta(0.4, 1)$ and $f(p^{(2)} \mid H_{01}) \sim Beta(0.6, 1)$. The proportions of composite hypothesis are $\pi_{00} = 0.4$, $\pi_{01} = 0.4$, $\pi_{10} = 0.1$, and $\pi_{11} = 0.1$.



(a) (b)

Figure A1: (a) The rejection regions of the three methods under Scenario 1; (b) The rejection regions of the three methods under Scenario 2.

Theorem 1 has proved that the ranking statistic of AMDP is optimal among those methods that effectively control the FDR at the nominal level. Therefore, we refer to the rejection region of AMDP as the oracle rejection region in the sense that it achieves the highest power under FDR control. In Scenarios 1-2, we compare the rejection regions of JS-mixture and DACT with the oracle rejection region when the FDR level can be precisely controlled to the specified level. This comparison provides deep insights into the impact of information loss during the ranking step on power.

Figure A1 visually presents the local FDR for the four-group model (5) in both Scenarios 1-2, as well as the rejection region for the three procedures. The color intensity in the figure represents the level of local FDR, with darker colors indicating lower local FDR, thus the corresponding hypothesis is more likely to be rejected. While the rejection regions of all three methods are all located in areas with lower local FDR, we emphasize that the AMDP is superior since it simultaneously considers information about proportions of the composite null hypothesis and the distributions of p-values under alternatives, leading to more accurate and reliable selection of mediators. In contrast, the other two methods only take into account partial information during the ranking step, which results in decreased power. The insensitivity of the JS-mixture method to changes in proportions and distributions is noteworthy. As described in panels (a)-(b), the shape of its rejection region remains square regardless of the scenarios. On the other hand, the DACT method only considers partial information about proportions of the null, since its rejection domain remains symmetrical under Scenario 1 (Balanced proportion of H_{01} and H_{10}), and shifts towards the larger distribution side under Scenario 2 (Unbalanced proportion of H_{01} and H_{10}). In contrast, the AMDP method fully captures all relevant information above, as the oracle rejection region is sensitive to the change in both proportions and distributions. This allows the AMDP method to adapt and adjust its rejection region accordingly, making it more effective in identifying significant signals.

C Additional results of Section 4

In this section, we demonstrate additional results related to the data analysis of the prostate cancer dataset in Section 4, including Table A3 and Figures A2-A3.

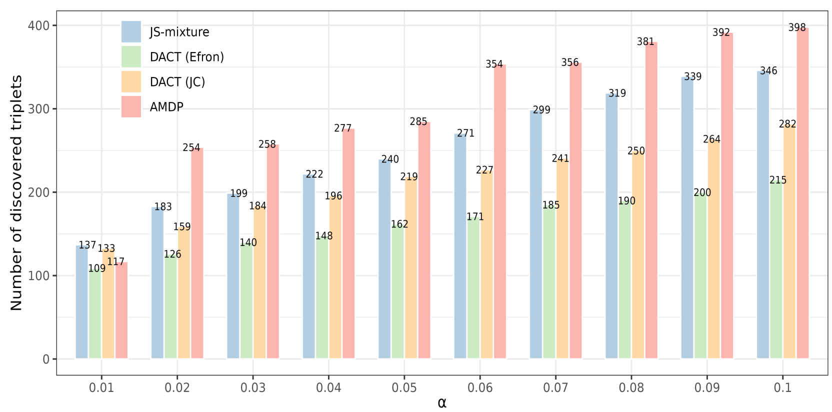


Figure A2: The number of triplets discovered by four methods. The nominal FDR level α varies from 0.01 to 0.1. The blue, green, orange, and pink bars represent the numbers of triplets identified by JS-mixture, DACT (Efron), DACT (JC), and AMDP, respectively.

Figure A3 provides an overview observation of the prostate cancer dataset as well as the rejection regions of JS-mixture, DACT (Efron), DACT (JC), and AMDP. Panel (a) shows the dispersion of p-values. However, the high density of the p-values makes it difficult to observe carefully. Thus, we depict the details of the TCGA dataset in different aspects in panels (b), (c), and (d), respectively, for providing a clearer insight. In panels (e)-(h), and (i)-(l), we compare the rejection regions of four methods: JS-mixture, DACT (Efron), DACT (JC), and AMDP at FDR levels of 0.05 and 0.1, respectively.

From panels (b)-(d), it can be seen that the distribution of p-values is influenced by information related to the composite null hypothesis. In panel (b), there is a slightly denser concentration of p-values near the $p^{(1)} = 0$ axis. This occurrence can be attributed to the presence of $\hat{\pi}_{10} = 0.03$, as mentioned earlier. On the other hand, panel (c) exhibits a notable concentration of p-values near the $p^{(2)} = 0$ axis. This pattern is influenced by the presence of a significant number of cases falling under H_{01}, which affects the distribution of p-values. As a result, we observe an accumulation of p-values near the $p^{(2)} = 0$ axis in the plot. Panel (d) demonstrates a seemingly uniform distribution of p-values. This uniformity can be attributed to the theoretical expectation that only p-values under

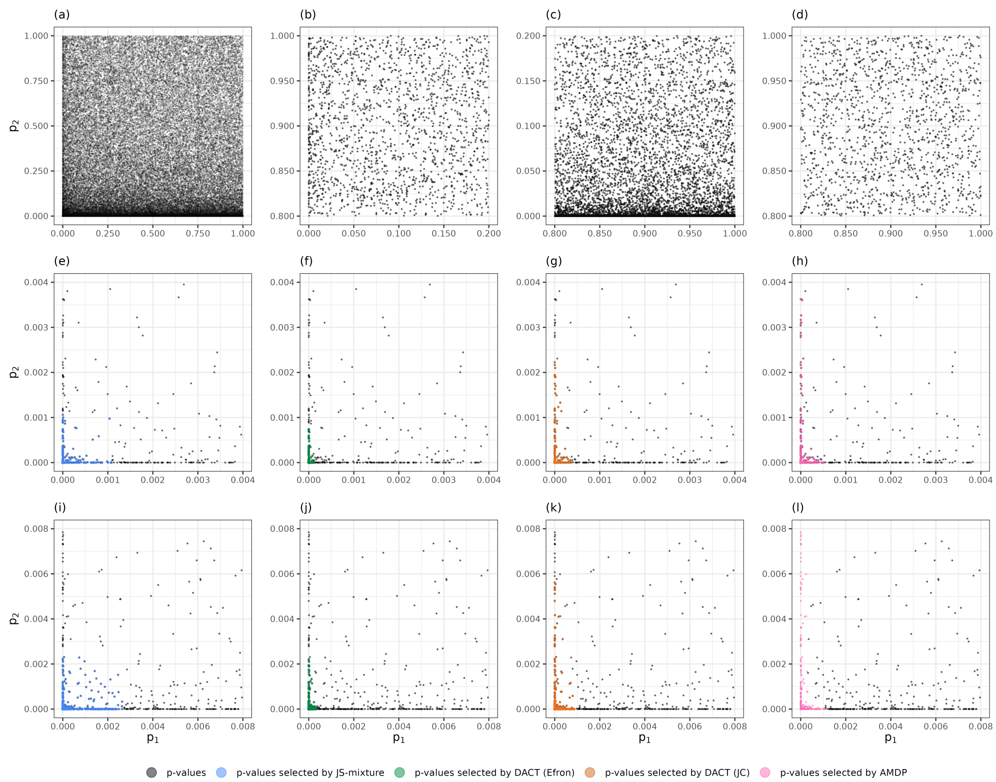


Figure A3: (a) An overview of p-values obtained from TCGA prostate cancer dataset; (b) The dispersion of p-values in the region $[0, 0.2] \times [0.8, 1]$; (c) The dispersion of p-values in the region $[0.8, 1] \times [0, 0.2]$;(d) The dispersion of p-values in the region $[0.8, 1] \times [0.8, 1]$; (e)-(h) The rejection domains of JS-mixture, DACT (Efron), DACT (JC), and AMDP at the targeted FDR level 0.05, respectively; (i)-(l) The rejection domains of JS-mixture, DACT (Efron), DACT (JC), and AMDP at the targeted FDR level 0.1, respectively; The black dots represent the p-values of all triplets. The blue, green, orange, and pink dots represent p-values of triplets identified by JS-mixture, DACT (Efron), DACT (JC), and AMDP, respectively.

H_{00} exist in the region $[0.8, 1] \times [0.8, 1]$. At the FDR level of 0.05, the rejection region of JS-mixture in panel (e) corresponds to a square shape. However, this symmetric shape does not reflect any information related to the distribution of p-values or the proportions of the composite null. In contrast, DACT (Efron) considers the proportion of null hypotheses and demonstrates a preference for rejecting fewer hypotheses with p-values close to $p^{(2)} = 0$ to minimize false discoveries, as shown in panel (f). However, the number of triplets identified by DACT (Efron) is the least among all methods, resulting an overly conservative behavior. The conservatism observed in DACT (Efron) is alleviated by DACT (JC), as panel (g) reveals that DACT (JC) identifies more significant triplets compared to DACT (Efron). DACT (JC) offers a more efficient approach by adjusting the threshold of the rejection region to achieve a higher sensitivity. In panel (h), we observe that the rejection region of AMDP is adaptive. AMDP estimates the number of false discoveries based on symmetric regions of the rejection region, allowing for more effective and accurate control of false discoveries, and well-calibrated adjustments to the rejection region. AMDP strikes a better performance on detecting significant triplets among all procedures. Next, we turn to investigate the rejection regions of these four methods at the FDR level of 0.1. In panel (i), it is observed that the rejection region of JS-mixture remains insensitive to changes in FDR levels, maintaining its square shape. As shown in panels (j)-(k), both DACT (Efron) and DACT (JC) exhibit increased identification of triplets compared to the FDR level of 0.05. Nevertheless, they still appear to be somewhat underpowered in efficiently detecting significant triplets. In contrast, AMDP outperforms all the other methods at the same FDR level, as demonstrated in panel (l). By leveraging information on the proportions of null and calibrating its rejection region dynamically, AMDP achieves better power to identify significant triplets.

Table A3: Top ten triplets identified by AMDP at the FDR level 0.1.

SNP ID	CpG Name	Gene	Chromosome	p_1	p_2
rs12653946	cg00626856	IRX4	5	6.60e-56	2.66e-20
rs12653946	cg03587843	IRX4	5	1.95e-51	1.03e-19
rs12653946	cg06161964	IRX4	5	1.99e-53	2.02e-22
rs12653946	cg09672187	IRX4	5	4.01e-65	2.13e-33
rs12653946	cg11279838	IRX4	5	3.97e-64	3.61e-27
rs12653946	cg14051264	IRX4	5	7.62e-67	8.86e-26
rs12653946	cg26195178	IRX4	5	2.52e-61	2.43e-26
rs5945619	cg16065628	NUDT11	X	6.75e-32	2.79e-42
rs1933488	cg23651356	RGS17	6	7.25e-20	2.81e-16
rs12653946	cg14823763	IRX4	5	8.13e-47	1.27e-15

The top ten triplets identified by AMDP are summaried in Table A3. These ten triplets consist of ten CpG sites and three genes. The CpG sites involved in these triplets are located in close proximity to the transcription starting sites, and their DNA methylation level are closely related to the expression of the corresponding genes [5]. Among the identified triplets, the three genes, IRX4, NUDT11, and RGS17, have been shown to be associated with altered CpG methylation. IRX4 is a causative gene of the prostate cancer susceptibility locus [18]. The corresponding SNP rs12653946, a variant previously confirmed to be associated with prostate cancer, is significantly associated with IRX4 expression [2]. The increased expression of NUDT11 has been confirmed to be associated with the risk variant rs5945619 [9, 13]. RGS17 is a commonly induced gene in prostate tumors, and has been found crucial for the maintenance of the proliferative potential of tumor cells [4].

D Discussions on the parameter choice and assumptions

D.1 Parameter choice

In Figures 1-2 in Section 3, we assess how the four methods (JS-mixture, DACT (Efron), DACT (JC), and AMDP) are influenced by effect size, the large mediator size and sample size. To ensure the realism of our experiments, we carefully selected our simulation parameters. Motivated from several real-world datasets including the TCGA lung cancer cohort dataset [19], the Multi-Ethnic Study of Atherosclerosis [7], and the TCGA prostate cancer dataset [5], we adopt similar parameter settings as those used in [5] to construct the simulation examples in Section 3.

Regarding the choice of nominal FDR level, we initially used an FDR level of 0.1, which is a widely accepted standard in the field [5, 15]. Another common FDR level is 0.05 [10, 16]. To provide a comprehensive analysis, we conducted experiments at the FDR threshold of 0.05 across a wide range of sample sizes (200, 500, 1000, and 5000). We present the experimental results under sparse alternatives scenario and dense alternatives scenario in Tables A4-A5. It's noteworthy that the results are similar with those obtained using the FDR level of 0.1 in Section 3.

D.2 Discussion on Assumptions 1-2

Our method extracts a pair $(p^{(1)}, p^{(2)})$ for each exposure-mediator-outcome relationship and employs these pairs to estimate the FDP on a two-dimensional plane $[0, 1] \times [0, 1]$. The theoretical basis supporting FDP estimation is the assumption that p-values are uniformly distributed under the null hypothesis, which is a widely recognized principle [3, 11]. Due to the presence of a composite null hypothesis in the mediation effect, we elaborate on Assumptions 1-2 to illustrate the properties of the p-value distribution under composite null hypothesis.

For Assumption 1, under H_{00}, both p_{1j} and p_{2j} obey the uniform distribution, resulting in (p_{1j}, p_{2j}) also following the uniform distribution on the two-dimensional plane $[0, 1] \times [0, 1]$. Consequently, the

Table A4: The FDR and power performance of the four methods with effect size α_j=0.2, β_j=0.3 under sparse alternatives scenario. The nominal FDR level is 0.05, and the number of mediators is 15000.

	Method	FDR	Power
n=200	AMDP	0.0530	0.0474
	JS-mixture	0.0284	0.0187
	DACT (Efron)	0.0091	0.0055
	DACT (JC)	0.0859	0.0960
n=500	AMDP	0.0454	0.4665
	JS-mixture	0.0438	0.3714
	DACT (Efron)	0.0140	0.2254
	DACT (JC)	0.0793	0.4960
n=1000	AMDP	0.0488	0.8698
	JS-mixture	0.0500	0.7931
	DACT (Efron)	0.0114	0.6101
	DACT (JC)	0.0299	0.7460
n=5000	AMDP	0.0498	0.9999
	JS-mixture	0.0529	1
	DACT (Efron)	8.00e-05	0.9986
	DACT (JC)	0.0935	1

Table A5: The FDR and power performance of the four methods with effect size α_j=0.2, $\beta_j = 0.3$ under dense alternatives scenario. The nominal FDR level is 0.05, and the number of mediators is 15000.

	Method	FDR	Power
n=200	AMDP	0.0471	0.0331
	JS-mixture	0.0348	0.0206
	DACT (Efron)	0.0449	0.0379
	DACT (JC)	0.1156	0.1673
n=500	AMDP	0.0460	0.4768
	JS-mixture	0.0504	0.4168
	DACT (Efron)	0.0333	0.3488
	DACT (JC)	0.1640	0.6839
n=1000	AMDP	0.0487	0.8734
	JS-mixture	0.0541	0.8208
	DACT (Efron)	0.0315	0.7636
	DACT (JC)	0.0818	0.8674
n=5000	AMDP	0.0501	1
	JS-mixture	0.0544	1
	DACT (Efron)	0.0116	1
	DACT (JC)	0.1085	1

sampling distribution of (p_{1j}, p_{2j}) is symmetrical around $p^{(1)} = 0.5$ and $p^{(2)} = 0.5$. Under H_{01}, p_{1j} still obeys the uniform distribution, but p_{2j} does not, leading to (p_{1j}, p_{2j}) being only symmetrical about $p^{(1)} = 0.5$ on $[0, 1] \times [0, 1]$. Similarly, under H_{10}, p_{2j} obeys the uniform distribution, but p_{1j} does not, resulting in (p_{1j}, p_{2j}) being only symmetrical about $p^{(2)} = 0.5$ on $[0, 1] \times [0, 1]$. It is essential to emphasize that Assumption 1 specifically applies to the null mediators.

For Assumption 2, a non-null p-value theoretically lies within $[0, 0.5)$. Therefore, as the sample size n tends to infinity, the probability of p-values under alternative hypotheses falling within $[0.5, 1]$ approaches zero. For example, as n goes to infinity, p-values under H_{11} and H_{10} are not expected to fall within the region $\tilde{D}_{01} = [0.5, 1] \times [0, 0.5)$ because non-null p_{1j} not lies within $[0.5, 1]$ theoretically, therefore the region $\tilde{D}_{01}$ only contains p-values under H_{00} and H_{01}. Similarly, the region $\tilde{D}_{10}$ theoretically only includes p-values under H_{00} and H_{10}.

E Proofs

E.1 Proof of Theorem 1

For any rejection region $S \in [0,1]^2$, the global FDR in mediation analysis is defined as follows

$$
\begin{aligned}
\mathrm{gFDR}(S) &= \mathbb{P}(H_{00} \cup H_{01} \cup H_{10} = 1 \mid p_j \in S) \\
&= \frac{\pi_{00}\mathbb{P}(p_j \in S \mid H_{00}=1)+\pi_{01}\mathbb{P}(p_j \in S \mid H_{01}=1)+\pi_{10}\mathbb{P}(p_j \in S \mid H_{10}=1)}{\pi_{00}\mathbb{P}(p_j \in S \mid H_{00}=1)+\pi_{01}\mathbb{P}(p_j \in S \mid H_{01}=1)+\pi_{10}\mathbb{P}(p_j \in S \mid H_{10}=1)+\pi_{11}\mathbb{P}(p_j \in S \mid H_{11}=1)} \\
&= \frac{\pi_{00}\int_S f_{00}(p)dp+\pi_{01}\int_S f_{01}(p)dp+\pi_{10}\int_S f_{10}(p)dp}{\pi_{00}\int_S f_{00}(p)dp+\pi_{01}\int_S f_{01}(p)dp+\pi_{10}\int_S f_{10}(p)dp+\pi_{11}\int_S f_{11}(p)dp}.
\end{aligned} \tag{A.1}
$$

We introduce some notations

$$
D_{00}(S)=\int_S f_{00}(p)dp,\ D_{01}(S)=\int_S f_{01}(p)dp,\ D_{10}(S)=\int_S f_{10}(p)dp,\ D_{11}(S)=\int_S f_{11}(p)dp.
$$

Thus, $\mathrm{gFDR}(S)$ is transformed into

$$
\begin{aligned}
\mathrm{gFDR}(S) &= \frac{\pi_{00}D_{00}(S)+\pi_{01}D_{01}(S)+\pi_{10}D_{10}(S)}{\pi_{00}D_{00}(S)+\pi_{01}D_{01}(S)+\pi_{10}D_{10}(S)+\pi_{11}D_{11}(S)} \\
&= \frac{1}{1+\{D_{11}(S)/(\gamma_{00}D_{00}(S)+\gamma_{01}D_{01}(S)+\gamma_{10}D_{10}(S))\}},
\end{aligned} \tag{A.2}
$$

where $\gamma_{00}=\dfrac{\pi_{00}}{\pi_{11}}$, $\gamma_{01}=\dfrac{\pi_{01}}{\pi_{11}}$, $\gamma_{10}=\dfrac{\pi_{10}}{\pi_{11}}$. For any threshold $\zeta \in (0,1]$, define the rejection region $S(\zeta)$ as

$$
\begin{aligned}
S(\zeta) &= \left\{p : \frac{\pi_{00}f_{00}(p)+\pi_{01}f_{01}(p)+\pi_{10}f_{10}(p)}{\pi_{00}f_{00}(p)+\pi_{01}f_{01}(p)+\pi_{10}f_{10}(p)+\pi_{11}f_{11}(p)} \le \zeta\right\} \\
&= \left\{p : \frac{1}{1+\{f_{11}(p)/(\gamma_{00}f_{00}(p)+\gamma_{01}f_{01}(p)+\gamma_{10}f_{10}(p))\}} \le \zeta\right\}.
\end{aligned} \tag{A.3}
$$

Here we prove that $\mathrm{gFDR}(S(\zeta))$ is a non-decreasing function of ζ. Suppose $\zeta_2 > \zeta_1$, considering two cases:

Case 1 $\nu(S(\zeta_2)-S(\zeta_1))=0$. We derive that $\mathrm{gFDR}(S(\zeta_1))=\mathrm{gFDR}(S(\zeta_2))$.

Case 2 $\nu(S(\zeta_2)-S(\zeta_1)) > 0$. We can prove that $\mathrm{gFDR}(S(\zeta))$ is a non-decreasing function of ζ if

$$
\begin{aligned}
&\frac{D_{11}(S(\zeta_2)-S(\zeta_1))}{\gamma_{00}D_{00}(S(\zeta_2)-S(\zeta_1))+\gamma_{01}D_{01}(S(\zeta_2)-S(\zeta_1))+\gamma_{10}D_{10}(S(\zeta_2)-S(\zeta_1))} \\
&< \frac{D_{11}(S(\zeta_1))}{\gamma_{00}D_{00}(S(\zeta_1))+\gamma_{01}D_{01}(S(\zeta_1))+\gamma_{10}D_{10}(S(\zeta_1))}
\end{aligned} \tag{A.4}
$$

holds, the reason is as follows. Let

$$
m_1 = \sup\left\{\frac{f_{11}(p)}{\gamma_{00}f_{00}(p)+\gamma_{01}f_{01}(p)+\gamma_{10}f_{10}(p)} : p \in S(\zeta_2)-S(\zeta_1)\right\},
$$

$$
m_2 = \inf\left\{\frac{f_{11}(p)}{\gamma_{00}f_{00}(p)+\gamma_{01}f_{01}(p)+\gamma_{10}f_{10}(p)} : p \in S(\zeta_1)\right\}.
$$

By the definition of region $S(\zeta)$, we have $m_2 > m_1$ obviously. Therefore, we have

$$
\begin{aligned}
&\frac{D_{11}(S(\zeta_1))}{\gamma_{00}D_{00}(S(\zeta_1))+\gamma_{01}D_{01}(S(\zeta_1))+\gamma_{10}D_{10}(S(\zeta_1))} \\
&\ge m_2\frac{\gamma_{00}D_{00}(S(\zeta_1))+\gamma_{01}D_{01}(S(\zeta_1))+\gamma_{10}D_{10}(S(\zeta_1))}{\gamma_{00}D_{00}(S(\zeta_1))+\gamma_{01}D_{01}(S(\zeta_1))+\gamma_{10}D_{10}(S(\zeta_1))} \\
&> m_1\frac{\gamma_{00}D_{00}(S(\zeta_2)-S(\zeta_1))+\gamma_{01}D_{01}(S(\zeta_2)-S(\zeta_1))+\gamma_{10}D_{10}(S(\zeta_2)-S(\zeta_1))}{\gamma_{00}D_{00}(S(\zeta_2)-S(\zeta_1))+\gamma_{01}D_{01}(S(\zeta_2)-S(\zeta_1))+\gamma_{10}D_{10}(S(\zeta_2)-S(\zeta_1))} \\
&\ge \frac{D_{11}(S(\zeta_2)-S(\zeta_1))}{\gamma_{00}D_{00}(S(\zeta_2)-S(\zeta_1))+\gamma_{01}D_{01}(S(\zeta_2)-S(\zeta_1))+\gamma_{10}D_{10}(S(\zeta_2)-S(\zeta_1))}.
\end{aligned} \tag{A.5}
$$

Furthermore, we decompose the region $S(\zeta_2)$ as follows

$$\frac{D_{11}(S(\zeta_2))}{\gamma_{00}D_{00}(S(\zeta_2))+\gamma_{01}D_{01}(S(\zeta_2))+\gamma_{10}D_{10}(S(\zeta_2))}$$
$$= \Big\{D_{11}(S(\zeta_2)-S(\zeta_1))+D_{11}(S(\zeta_1))\Big\}\Big/\left\{\begin{matrix}\gamma_{00}D_{00}(S(\zeta_2)-S(\zeta_1))+\gamma_{01}D_{01}(S(\zeta_2)-S(\zeta_1)) \\ +\gamma_{10}D_{10}(S(\zeta_2)-S(\zeta_1))+\gamma_{00}D_{00}(S(\zeta_1)) \\ +\gamma_{01}D_{01}(S(\zeta_1))+\gamma_{10}D_{10}(S(\zeta_1))\end{matrix}\right\}. \tag{A.6}$$

Combined with (A.4), we obtain

$$\frac{D_{11}(S(\zeta_1))}{\gamma_{00}D_{00}(S(\zeta_1))+\gamma_{01}D_{01}(S(\zeta_1))+\gamma_{10}D_{10}(S(\zeta_1))} > \frac{D_{11}(S(\zeta_2))}{\gamma_{00}D_{00}(S(\zeta_2))+\gamma_{01}D_{01}(S(\zeta_2))+\gamma_{10}D_{10}(S(\zeta_2))}. \tag{A.7}$$

Moreover, by the definition of $\mathrm{gFDR}(S)$, it holds that

$$\mathrm{gFDR}(S(\zeta_1)) < \mathrm{gFDR}(S(\zeta_2)).$$

Under the Assumption (ii) in Theorem 1, for a given $\alpha \in (0,1)$, there exists a threshold $\zeta^\star > 0$, s.t. $\mathrm{gFDR}(S(\zeta^\star)) = \alpha$. For the ease of presentation, we denote $S(\zeta^\star)$ as $S^\star$. In the following, we will prove that $S^\star$ is the optimal rejection region.

Considering any set T that satisfies $D_{11}(T) > D_{11}(S^\star)$. Let $R_T = T - S^\star$ and $R_S = S^\star - T$. We can derive that

$$\begin{aligned} D_{11}(T) &= D_{11}(T\cap S^\star) + D_{11}(R_T), \\ D_{11}(S) &= D_{11}(T\cap S^\star) + D_{11}(R_S). \end{aligned} \tag{A.8}$$

Then, we have $D_{11}(R_T) > D_{11}(R_S)$. By the definition of $S^\star$, we have

$$\inf\left\{\frac{\gamma_{00}f_{00}(p)+\gamma_{01}f_{01}(p)+\gamma_{10}f_{10}(p)}{f_{11}(p)} : p\in R_T\right\} > \sup\left\{\frac{\gamma_{00}f_{00}(p)+\gamma_{01}f_{01}(p)+\gamma_{10}f_{10}(p)}{f_{11}(p)} : p\in R_S\right\}. \tag{A.9}$$

Therefore,

$$\frac{\gamma_{00}D_{00}(R_T)+\gamma_{01}D_{01}(R_T)+\gamma_{10}D_{10}(R_T)}{D_{11}(R_T)} > \frac{\gamma_{00}D_{00}(R_S))+\gamma_{01}D_{01}(R_S)+\gamma_{10}D_{10}(R_S)}{D_{11}(R_S)}. \tag{A.10}$$

In a similar way, we can derive that

$$\frac{\gamma_{00}D_{00}(R_T)+\gamma_{01}D_{01}(R_T)+\gamma_{10}D_{10}(R_T)}{D_{11}(R_T)} > \frac{\gamma_{00}D_{00}(T\cap S^\star))+\gamma_{01}D_{01}((T\cap S^\star)+\gamma_{10}D_{10}((T\cap S^\star)}{D_{11}((T\cap S^\star)}. \tag{A.11}$$

Finally, we have

$$\begin{aligned} &\frac{\gamma_{00}D_{00}(T)+\gamma_{01}D_{01}(T)+\gamma_{10}D_{10}(T)}{D_{11}(T)} \\ &= \frac{\gamma_{00}D_{00}(T\cap S^\star)+\gamma_{01}D_{01}(T\cap S^\star)+\gamma_{10}D_{10}(T\cap S^\star)+\gamma_{00}D_{00}(R_T)+\gamma_{01}D_{01}(R_T)+\gamma_{10}D_{10}(R_T)}{D_{11}(T\cap S^\star)+D_{11}(R_T)} \\ &> \frac{\gamma_{00}D_{00}(T\cap S^\star)+\gamma_{01}D_{01}(T\cap S^\star)+\gamma_{10}D_{10}(T\cap S^\star)+\gamma_{00}D_{00}(R_S)+\gamma_{01}D_{01}(R_S)+\gamma_{10}D_{10}(R_S)}{D_{11}(T\cap S^\star)+D_{11}(R_S)} \\ &= \frac{\gamma_{00}D_{00}(S^\star)+\gamma_{01}D_{01}(S^\star)+\gamma_{10}D_{10}(S^\star)}{D_{11}(S^\star)}. \end{aligned} \tag{A.12}$$

The second inequality holds because $D_{11}(R_T) > D_{11}(R_S)$, implying $\mathrm{gFDR}(T) > \mathrm{gFDR}(S(\zeta^\star)) = \alpha$. Therefore, we can conclude that the rejection region $S(\zeta^\star)$ is optimal.

E.2 Proof of Theorem 2

E.2.1 The consistent estimator of local FDR

To justify Assumption (i) for the corresponding local FDR estimator in Theorem 2, we first prove the consistency of local FDR estimator under L_∞ norm in Proposition 1.

Proposition 1. *Assume that the smoothing parameter b satisfies*

$$\lim_{J\to\infty} b = 0 \textit{ and } \lim_{J\to\infty} Jb^2 = +\infty.$$

Then, we have

$$\sup_{p_j\in[0,1]^2} \left|\widehat{\mathrm{fdr}}(p_j) - \mathrm{fdr}(p_j)\right| \xrightarrow{P} 0 \textit{ as } n, J \to \infty.$$

Let g be a probability density on $[0,1]$, and $\hat{g}$ be the beta kernel estimator:

$$\hat{g}(p^{(i)}) = J^{-1}\sum_{j=1}^{J} K^{\star}_{p^{(i)},b}(p_{ij}), \quad i = 1, 2.$$

To prove the consistency of the beta kernel estimator $\hat{g}$, i.e

$$\sup_{p^{(i)}\in[0,1]} \left|\hat{g}(p^{(i)}) - g(p^{(i)})\right| \xrightarrow{P} 0 \text{ as } J \to \infty, \tag{A.13}$$

we first need to establish the uniform convergence of its bias on the interval $[0,1]$.

Lemma 1. *Let g be the probability density on $[0,1]$, and $\hat{g}$ be the beta kernel estimator. We have*

$$\sup_{p^{(i)}\in[0,1]} \left|\mathrm{E}\left\{\hat{g}(p^{(i)})\right\} - g(p^{(i)})\right| \to 0 \textit{ as } b \to 0, \quad i = 1, 2.$$

Proof of Lemma 1. Without loss of generality, we replace $p^{(i)}, i = 1, 2$ with p for simplifying the proof steps, and discuss three cases in the following.

Case 1 $p \in (2b, 1-2b)$ Denote μ_1 and σ_1^2 are mean and variance of P, a variable following Beta$(p/b, (1-p)/b)$. According to Johnson et al. [12], there exists a constant C such that

$$\mu_1 = p, \tag{A.14}$$

$$\sigma_1^2 = bp(1-p) + R_2(p), \tag{A.15}$$

where $R_2(p) \le Cb^2$. Because f is a probability density on $[0,1]$, for $\varepsilon > 0$, there exists a $\delta > 0$ such that

$$|g(t) - g(p)| < \varepsilon \quad \text{for } |p - t| < \delta \tag{A.16}$$

for all $p \in (2b, 1-2b)$; According to (A.14), we have

$$|\mu_1 - p| < \delta/2 \text{ for all } p \in (2b, 1-2b). \tag{A.17}$$

Therefore, we can derive that

$$\begin{aligned}
|E\{\hat{g}(p)\} - g(p)| =& \left|\int_{2b}^{1-2b} \{g(t) - g(p)\} K\left(t, \frac{p}{b}, \frac{1-p}{b}\right) dt\right| \\
\le& \int_{|t-\mu_1|<\delta/2} |g(t) - g(p)| K\left(t, \frac{p}{b}, \frac{1-p}{b}\right) dt \\
&+ \int_{|t-\mu_1|>\delta/2} |g(t) - g(p)| K\left(t, \frac{p}{b}, \frac{1-p}{b}\right) dt \\
\le& \int_{|t-\mu_1|<\delta/2} |g(t) - g(p)| K\left(t, \frac{p}{b}, \frac{1-p}{b}\right) dt \\
&+ 2 \sup_{t\in(2b,1-2b)} |g(t)| \int_{|t-\mu_1|>\delta/2} K\left(t, \frac{p}{b}, \frac{1-p}{b}\right) dt \\
\equiv& \mathcal{M}_1 + \mathcal{M}_2.
\end{aligned}$$

According (A.16) and (A.17), we obtain

$$\mathcal{M}_1 \le \varepsilon. \tag{A.18}$$

Combining the Chebyshev's inequality and (A.15), and there also exists b_ε such that

$$\mathcal{M}_2 \le \left\{8 \sup_{t\in(2b,1-2b)} |g(t)|\sigma_1^2\right\}/\delta^2 \le \left\{2 \sup_{t\in(2b,1-2b)} |g(t)|\left(b+4Cb^2\right)\right\}/\delta^2 \le \varepsilon \text{ for all } b \le b_\varepsilon. \tag{A.19}$$

Thus, from (A.18) and (A.19), we conclude that

$$\sup_{p\in(2b,1-2b)} |\mathrm{E}\{\hat{g}(p)\} - g(p)| < 2\varepsilon \quad \text{for all } b \le b_\varepsilon.$$

Case 2 $p \in [0, 2b]$ Based on the notations of Case 1, we have

$$\mu_2 = p + \xi(p, b), \tag{A.20}$$

$$\sigma_2^2 = R_2(p), \tag{A.21}$$

where μ_2 and σ_2^2 are mean and variance of of P, a variable following Beta $(\rho(p,b),(1-p)/b)$, $\xi(p,b) = (1-p)\{\rho(p,b) - p/b\}/\{1 + b\rho(p,b) - p\}$, and $R_2(p) \le Cb^2$. For $\varepsilon > 0$, there exists a $\delta > 0$ such that

$$|g(t) - g(p)| < \varepsilon \quad \text{for } |p-t| < \delta \tag{A.22}$$

for all $p \in [0, 2b]$; According to (A.20), since $\xi(p, b)$ is a bounded function for $p \in [0, 2b]$, there also exists b_δ such that

$$|\mu_2 - p| < \delta/2 \text{ for } b \le b_\delta \quad \text{for all } p \in [0, 2b]. \tag{A.23}$$

Therefore, we can derive that

$$\begin{aligned}
|E\{\hat{g}(p)\} - g(p)| =& \left|\int_0^{2b} \{g(t) - g(p)\}K\left(t, \rho(p,b), \frac{1-p}{b}\right) dt\right| \\
\le& \int_{|t-\mu_2|<\delta/2} |g(t) - g(p)|K\left(t, \rho(p,b), \frac{1-p}{b}\right) dt \\
&+ \int_{|t-\mu_2|>\delta/2} |g(t) - g(p)|K\left(t, \rho(p,b), \frac{1-p}{b}\right) dt \\
\le& \int_{|t-\mu_2|<\delta/2} |g(t) - g(p)|K\left(t, \rho(p,b), \frac{1-p}{b}\right) dt \\
&+ 2 \sup_{t\in[0,2b]} |g(t)| \int_{|t-\mu_2|>\delta/2} K\left(t, \rho(p,b), \frac{1-p}{b}\right) dt \\
\equiv& \mathcal{M}_1 + \mathcal{M}_2.
\end{aligned}$$

According to (A.22) and (A.23), there exists $b_\varepsilon^{(1)}$ such that

$$\mathcal{M}_1 \le \varepsilon \text{ for all } b \le b_\varepsilon^{(1)}. \tag{A.24}$$

Combining the Chebyshev's inequality and (A.21), and there also exists $b_\varepsilon^{(2)}$ such that

$$\mathcal{M}_2 \le \left\{8 \sup_{t\in[0,2b]} |g(t)|\sigma_2^2\right\}/\delta^2 \le \left\{8 \sup_{t\in[0,2b]} |g(t)|Cb^2\right\}/\delta^2 \le \varepsilon \text{ for all } b \le b_\varepsilon^{(2)}. \tag{A.25}$$

Thus, from (A.24) and (A.25), we conclude that

$$\sup_{p\in[0,2b]} |\mathrm{E}\{\hat{g}(p)\} - g(p)| < 2\varepsilon \quad \text{for all } b \le \min\left(b_\varepsilon^{(1)}, b_\varepsilon^{(2)}\right).$$

Case 3 $p \in [1-2b, 1]$ Case 3 can be proven a similar procedure. We note that

$$\mu_3 = p - b \cdot \xi(1-p, b), \tag{A.26}$$

$$\sigma_3^2 = R_2(p), \tag{A.27}$$

where μ_3 and σ_3^2 are mean and variance of P, a variable following Beta$(p/b, \rho(1-p,b))$, and $R_2(p) \le Cb^2$.

This completes the proof of Lemma 1. □

Proof of Proposition 1. To prove the consistency of the beta kernel estimator, we use the inequality:

$$\sup_{p\in[0,1]} |\hat{g}(p) - g(p)| \leq \sup_{p\in[0,1]} |\hat{g}(p) - \mathrm{E}\{\hat{g}(p)\}| + \sup_{p\in[0,1]} |\mathrm{E}\{\hat{g}(p)\} - g(p)|\,. \tag{A.28}$$

From Lemma 1, the second term converges to zero. In the following, we prove that

$$\sup_{p\in[0,1]} |\hat{g}(p) - \mathrm{E}\{\hat{g}(p)\}| \xrightarrow{P} 0 \text{ as } J \to \infty. \tag{A.29}$$

We also consider three cases:

Case 1 $p \in (2b, 1-2b)$ The beta kernel estimator $\hat{g}(p)$ is expressed as

$$\hat{g}(p) = \int_{2b}^{1-2b} K\left(t, \frac{p}{b}, \frac{1-p}{b}\right) dF_n(t), \tag{A.30}$$

where F_n is the empirical distribution. The expectation of the beta kernel estimator is

$$E\{\hat{g}(p)\} = \int_{2b}^{1-2b} K\left(t, \frac{p}{b}, \frac{1-p}{b}\right) dF(t). \tag{A.31}$$

Thus, for $p \in (2b, 1-2b)$, we can derive that

$$\begin{aligned} |\hat{g}(p) - \mathrm{E}\{\hat{g}(p)\}| &= \left|\int_{2b}^{1-2b} K\left(t, \frac{p}{b}, \frac{1-p}{b}\right) d\{F_n(t) - F(t)\}\right| \\ &\leq \sup_{t\in(b,1-2b)} |F_n(t) - F(t)| \int_{2b}^{1-2b} \left|dK\left(t, \frac{p}{b}, \frac{1-p}{b}\right)\right|. \end{aligned} \tag{A.32}$$

Note that the integral in (A.32) is bounded above by

$$\frac{1-b}{b}\int_{2b}^{1-2b} \left|K\left(t, \frac{p}{b}-1, \frac{1-p}{b}\right) - K\left(t, \frac{p}{b}, \frac{1-p}{b}-1\right)\right| dt \leq 2\frac{1-b}{b}. \tag{A.33}$$

Therefore,

$$|\hat{g}(p) - E\{\hat{g}(p)\}| \leq 2\frac{1-b}{b} \sup_{t\in(2b,1-2b)} |F_n(t) - F(t)|\,. \tag{A.34}$$

From Dvoretzky et al. [8], we can obtain

$$\begin{aligned} \mathrm{P}\left[\sup_{p\in(2b,1-2b)} |\hat{g}(p) - \mathrm{E}\{\hat{g}(p)\}| \geq \varepsilon\right] &\leq \mathrm{P}\left\{\sup_{t\in(2b,1-2b)} |F_n(t) - F(t)| \geq \frac{\varepsilon}{2}\cdot\frac{b}{1-b}\right\} \\ &\leq 2\exp\left\{-J\frac{\varepsilon^2}{2}\frac{b^2}{(1-b)^2}\right\}. \end{aligned} \tag{A.35}$$

By utilizing the Borel-Cantelli Lemma, it is shown that under the beta kernel estimator is consistent.

Case 2 $p \in [0, 2b]$ Case 2 can be proven a similar procedure of Case 1. Note that, for all $p \in [0, 2b]$,

$$\begin{aligned} |\hat{g}(p) - \mathrm{E}\{\hat{g}(p)\}| &= \left|\int_{0}^{2b} K\left(t, \rho(p,b), \frac{1-p}{b}\right) d\{F_n(t) - F(t)\}\right| \\ &\leq \sup_{t\in[0,2b]} |F_n(t) - F(t)| \int_{0}^{2b} \left|dK\left(t, \rho(p,b), \frac{1-p}{b}\right)\right|. \end{aligned} \tag{A.36}$$

Since $\rho(p,b)$ is monotonic increasing in $[0,2b], \rho(0,b) = 1, \rho(2b,b) = 2$. For $p \in (0, 2b]$, the integral in (A.36) is bounded above by $2\frac{1+b}{b}$. For $p = 0$, it is bounded above by

$$(\rho(p,b) + \frac{1-p}{b} - 1)\int_{0}^{2b} \left|K\left(t, \rho(p,b), \frac{1-p}{b}-1\right)\right| dt = \frac{1+b}{b}. \tag{A.37}$$

Thus, we can obtain

$$\begin{aligned}\mathrm{P}\left[\sup_{p\in[0,2b]}|\hat{g}(p)-\mathrm{E}\left\{\hat{g}(p)\right\}|\geq\varepsilon\right]&\leq\mathrm{P}\left\{\sup_{t\in[0,2b]}|F_n(t)-F(t)|\geq\frac{\varepsilon}{2}\cdot\frac{b}{1+b}\right\}\\&\leq 2\exp\left\{-J\frac{\varepsilon^2}{2}\frac{b^2}{(1+b)^2}\right\},\end{aligned}\tag{A.38}$$

which concludes the proof of the consistency of beta kernel estimator in Case 2.

Case 3 $p\in[1-2b,1]$ Case 3 can be proven a similar procedure of Case 1. Note that for all $p\in[1-2b,1]$,

$$\begin{aligned}|\hat{g}(p)-\mathrm{E}\left\{\hat{g}(p)\right\}|&=\left|\int_{1-2b}^{1}K\left(t,\frac{p}{b},\rho(1-p,b)\right)d\left\{F_n(t)-F(t)\right\}\right|\\&\leq\sup_{t\in[1-2b,1]}|F_n(t)-F(t)|\int_{1-2b}^{1}\left|dK\left(t,\frac{p}{b},\rho(1-p,b)\right)\right|.\end{aligned}\tag{A.39}$$

For $p\in[1-2b,1)$, the integral in (A.39) is bounded above by $2\frac{1+b}{b}$. And for $p=1$, it is bounded above by

$$(\rho(1-p,b)+\frac{p}{b}-1)\int_{1-2b}^{1}\left|K\left(t,\frac{p}{b}-1,\rho(1-p,b)\right)\right|dt=\frac{1+b}{b}.\tag{A.40}$$

Thus for all $p\in[1-2b,1]$,

$$|\hat{g}(p)-E\left\{\hat{g}(p)\right\}|\leq 2\frac{1+b}{b}\sup_{t\in[1-2b,1]}|F_n(t)-F(t)|.\tag{A.41}$$

Similarly, we obtain

$$\begin{aligned}\mathrm{P}\left[\sup_{p\in[1-2b,1]}|\hat{g}(p)-\mathrm{E}\left\{\hat{g}(p)\right\}|\geq\varepsilon\right]&\leq\mathrm{P}\left\{\sup_{t\in[1-2b,1]}|F_n(t)-F(t)|\geq\frac{\varepsilon}{2}\cdot\frac{b}{1+b}\right\}\\&\leq 2\exp\left\{-J\frac{\varepsilon^2}{2}\frac{b^2}{(1+b)^2}\right\},\end{aligned}\tag{A.42}$$

which concludes the proof of the consistency of beta kernel estimator in Case 3.

From Dai et al. [5], for a fixed J and λ, the biases of $\hat{\pi}_{00}$, $\hat{\pi}_{10}$, and $\hat{\pi}_{01}$ go to zero as $n\to\infty$:

$$\lim_{n\to\infty}\hat{\pi}_{00}=\pi_{00},\quad\lim_{n\to\infty}\hat{\pi}_{01}=\pi_{01},\quad\lim_{n\to\infty}\hat{\pi}_{10}=\pi_{10}.\tag{A.43}$$

And we can derive

$$\hat{f}(p)=\hat{f}(p^{(1)},p^{(2)})=\hat{g}(p^{(1)})\cdot\hat{g}(p^{(2)}).\tag{A.44}$$

By combining equations (A.13), (A.43), and (A.44), according to continuous mapping theorem [6], we have

$$\sup_{p\in[0,1]}\left|\widehat{\mathrm{fdr}}(p)-\mathrm{fdr}(p)\right|\xrightarrow{P}0\text{ as }n,J\to\infty.\tag{A.45}$$

□

According to equation (A.45), we can verify the rationality of Assumption (i) in Theorem 2:

$$\frac{1}{J}\sum_{j=1}^{J}\left|\widehat{\mathrm{fdr}}(p_j)-\mathrm{fdr}(p_j)\right|\xrightarrow{P}0\text{ as }n,J\to\infty.\tag{A.46}$$

E.2.2 Proof of Theorem 2

To begin with, we introduce some notations. For $\zeta \in (0,1]$, denote

$$\widehat{G}_J^{00}(\zeta) = \frac{1}{J_{00}} \sum_{j\in\Delta_{00}} \mathbb{I}\left(p_j \in \widehat{S}(\zeta)\right), \quad \widehat{G}_J^{01}(\zeta) = \frac{1}{J_{01}} \sum_{j\in\Delta_{01}} \mathbb{I}\left(p_j \in \widehat{S}(\zeta)\right),$$

$$\widehat{G}_J^{10}(\zeta) = \frac{1}{J_{10}} \sum_{j\in\Delta_{10}} \mathbb{I}\left(p_j \in \widehat{S}(\zeta)\right), \quad \widehat{G}_J^{11}(\zeta) = \frac{1}{J_{11}} \sum_{j\in\Delta_{11}} \mathbb{I}\left(p_j \in \widehat{S}\zeta)\right),$$

$$G_J^{00}(\zeta) = \frac{1}{J_{00}} \sum_{j\in\Delta_{00}} \mathbb{P}\left(p_j \in \widehat{S}(\zeta)\right), \quad G_J^{01}(\zeta) = \frac{1}{J_{01}} \sum_{j\in\Delta_{01}} \mathbb{P}\left(p_j \in \widehat{S}(\zeta)\right),$$

$$G_J^{10}(\zeta) = \frac{1}{J_{10}} \sum_{j\in\Delta_{10}} \mathbb{P}\left(p_j \in \widehat{S}(\zeta)\right),$$

$$\widehat{V}_J^{00}(\zeta) = \frac{1}{J_{00}} \sum_{j\in\Delta_{00}} \mathbb{I}\left(p_j \in \tilde{S}_{00}(\zeta)\right), \quad \widehat{V}_J^{01}(\zeta) = \frac{1}{J_{01}+J_{00}} \sum_{j\in\Delta_{01}\cup\Delta_{00}} \mathbb{I}\left(p_j \in \tilde{S}_{01}(\zeta)\right),$$

$$\widehat{V}_J^{10}(\zeta) = \frac{1}{J_{10}+J_{00}} \sum_{j\in\Delta_{10}\cup\Delta_{00}} \mathbb{I}\left(p_j \in \tilde{S}_{10}(\zeta)\right),$$

where $J_{00} = |\Delta_{00}|$, $J_{01} = |\Delta_{01}|$, $J_{10} = |\Delta_{10}|$, $J_{11} = |\Delta_{11}|$. Denote $r_J^{00} = J_{00}/J_{11}, r_J^{01} = J_{01}/J_{11}, r_J^{10} = J_{10}/J_{11}, v_J = \dfrac{J_{11}}{J_{00}+J_{01}+J_{10}}$. And

$$\begin{aligned}
\overline{K}_J^0(\zeta) &= v_J\{r_J^{00}G_J^{00}(\zeta) + r_J^{01}G_J^{01}(\zeta) + r_J^{10}G_J^{10}(\zeta)\},\\
K_J^0(\zeta) &= v_J\{r_J^{00}\widehat{G}_J^{00}(\zeta) + r_J^{01}\widehat{G}_J^{01}(\zeta) + r_J^{10}\widehat{G}_J^{10}(\zeta)\},\\
\widehat{K}_J^0(\zeta) &= v_J\{(r_J^{01}+r_J^{00})\widehat{V}_J^{01}(\zeta) + (r_J^{10}+r_J^{00})\widehat{V}_J^{10}(\zeta) - r_J^{00}\widehat{V}_J^{00}(\zeta)\},\\
\mathrm{FDP}_J(\zeta) &= \frac{r_J^{00}\widehat{G}_J^{00}(\zeta) + r_J^{01}\widehat{G}_J^{01}(\zeta) + r_J^{10}\widehat{G}_J^{10}(\zeta)}{r_J^{00}\widehat{G}_J^{00}(\zeta) + r_J^{01}\widehat{G}_J^{01}(\zeta) + r_J^{10}\widehat{G}_J^{10}(\zeta) + \widehat{G}_J^{11}(\zeta)},\\
\mathrm{FDP}_J^{\dagger}(\zeta) &= \frac{\widehat{K}_J^0(\zeta)/v_J}{r_J^{00}\widehat{G}_J^{00}(\zeta) + r_J^{01}\widehat{G}_J^{01}(\zeta) + r_J^{10}\widehat{G}_J^{10}(\zeta) + \widehat{G}_J^{11}(\zeta)},\\
\overline{\mathrm{FDP}}_J(\zeta) &= \frac{\overline{K}_J^0(\zeta)/v_J}{r_J^{00}G_J^{00}(\zeta) + r_J^{01}G_J^{01}(\zeta) + r_J^{10}G_J^{10}(\zeta) + \widehat{G}_J^{11}(\zeta)}.
\end{aligned}$$

Before proceeding with the proof of Theorem 2, we prove Lemma 2 first.

Lemma 2. *Under Assumption (i)-(ii) in Theorem 2, if $J_{00} \to \infty$, $J_{01} \to \infty$, $J_{10} \to \infty$ as $J \to \infty$, and $n \to \infty$, we have in probability,*

$$\sup_{\zeta\in(0,1]} \left|\widehat{G}_J^{00}(\zeta) - G_J^{00}(\zeta)\right| \longrightarrow 0, \quad \sup_{\zeta\in(0,1]} \left|\widehat{G}_J^{01}(\zeta) - G_J^{01}(\zeta)\right| \longrightarrow 0,$$
$$\sup_{\zeta\in(0,1]} \left|\widehat{G}_J^{10}(\zeta) - G_J^{10}(\zeta)\right| \longrightarrow 0, \quad \sup_{\zeta\in(0,1]} \left|\widehat{K}_J^{0}(\zeta) - \overline{K}_J^{0}(\zeta)\right| \longrightarrow 0.$$

Proof of Lemma 2. We consider three cases under composite null hypothesis.

Case 1 Under H_{00}: We can derive that

$$\begin{aligned}
\widehat{G}_J^{00}(\zeta) &= \frac{1}{J_{00}} \sum_{j\in\Delta_{00}} \left\{\mathbb{I}\left(p_j \in \widehat{S}(\zeta)\right) - \mathbb{I}\Big(p_j \in S(\zeta)\Big) + \mathbb{I}\Big(p_j \in S(\zeta)\Big)\right\},\\
G_J^{00}(\zeta) &= \frac{1}{J_{00}} \sum_{j\in\Delta_{00}} \left\{\mathbb{P}\left(p_j \in \widehat{S}(\zeta)\right) - \mathbb{P}\Big(p_j \in S(\zeta)\Big) + \mathbb{P}\Big(p_j \in S(\zeta)\Big)\right\}.
\end{aligned} \tag{A.47}$$

Thus, we have

$$\begin{aligned}\sup_{\zeta\in(0,1]}\left|\widehat{G}_J^{00}(\zeta)-G_J^{00}(\zeta)\right| \le & \sup_{\zeta\in(0,1]}\frac{1}{J_{00}}\sum_{j\in\Delta_{00}}\left|\mathbb{I}\left(p_j\in\widehat{S}(\zeta)\right)-\mathbb{I}\left(p_j\in S(\zeta)\right)\right| \\ & + \sup_{\zeta\in(0,1]}\frac{1}{J_{00}}\sum_{j\in\Delta_{00}}\left|\mathbb{P}\left(p_j\in\widehat{S}(\zeta)\right)-\mathbb{P}\left(p_j\in S(\zeta)\right)\right| \\ & + \sup_{\zeta\in(0,1]}\frac{1}{J_{00}}\sum_{j\in\Delta_{00}}\left|\mathbb{I}\left(p_j\in S(\zeta)\right)-\mathbb{P}\left(p_j\in S(\zeta)\right)\right|.\end{aligned} \tag{A.48}$$

To deal with the first term, we have

$$\begin{aligned}&\frac{1}{J_{00}}\sum_{j\in\Delta_{00}}\left|\mathbb{I}\left(p_j\in\widehat{S}(\zeta)\right)-\mathbb{I}\left(p_j\in S(\zeta)\right)\right| \\ &\quad=\frac{1}{J_{00}}\sum_{j\in\Delta_{00}}\left|\mathbb{I}\left\{\widehat{\mathrm{fdr}}\left(p_j\right)\le\zeta\right\}-\mathbb{I}\left\{\mathrm{fdr}\left(p_j\right)\le\zeta\right\}\right| \\ &\quad=\frac{1}{J_{00}}\sum_{j\in\Delta_{00}}\left[\mathbb{I}\left\{\widehat{\mathrm{fdr}}\left(p_j\right)\le\zeta,\mathrm{fdr}\left(p_j\right)>\zeta\right\}+\mathbb{I}\left\{\mathrm{fdr}\left(p_j\right)\le\zeta,\widehat{\mathrm{fdr}}\left(p_j\right)>\zeta\right\}\right] \\ &\quad=\frac{1}{J_{00}}\sum_{j\in\Delta_{00}}\left[\mathbb{I}\left\{\widehat{\mathrm{fdr}}\left(p_j\right)\le\zeta,\zeta+\epsilon\ge\mathrm{fdr}\left(p_j\right)>\zeta\right\}+\mathbb{I}\left\{\zeta-\epsilon<\mathrm{fdr}\left(p_j\right)\le\zeta,\widehat{\mathrm{fdr}}\left(p_j\right)>\zeta\right\}\right] \\ &\qquad+\frac{1}{J_{00}}\sum_{j\in\Delta_{00}}\left[\mathbb{I}\left\{\widehat{\mathrm{fdr}}\left(p_j\right)\le\zeta,\mathrm{fdr}\left(p_j\right)>\zeta+\epsilon\right\}+\mathbb{I}\left\{\mathrm{fdr}\left(p_j\right)\le\zeta-\epsilon,\widehat{\mathrm{fdr}}\left(p_j\right)>\zeta\right\}\right] \\ &\qquad\le\frac{1}{J_{00}}\sum_{j\in\Delta_{00}}\mathbb{I}\left\{\zeta-\epsilon<\mathrm{fdr}\left(p_j\right)\le\zeta+\epsilon\right\}+\frac{1}{J_{00}\epsilon}\sum_{j\in\Delta_{00}}\left|\widehat{\mathrm{fdr}}\left(p_j\right)-\mathrm{fdr}\left(p_j\right)\right|.\end{aligned} \tag{A.49}$$

Combine with the Glivenko-Cantelli theorem and Assumption (i), we can derive

$$\begin{aligned}Q:=&\sup_{\zeta\in(0,1]}\frac{1}{J_{00}}\sum_{j\in\Delta_{00}}\left|\mathbb{I}\left\{\widehat{\mathrm{fdr}}\left(p_j\right)\le\zeta\right\}-\mathbb{I}\left\{\mathrm{fdr}\left(p_j\right)\le\zeta\right\}\right| \\ \le&\sup_{\zeta\in(0,1]}\frac{1}{J_{00}}\sum_{j\in\Delta_{00}}\mathbb{I}\left\{\zeta-\epsilon<\mathrm{fdr}\left(p_j\right)\le\zeta+\epsilon\right\}+\frac{1}{J_{00}\epsilon}\sum_{j\in\Delta_{00}}\left|\widehat{\mathrm{fdr}}\left(p_j\right)-\mathrm{fdr}\left(p_j\right)\right| \\ \le&\sup_{\zeta\in(0,1]}\frac{1}{J_{00}}\sum_{j\in\Delta_{00}}\left|\mathbb{P}(\zeta-\epsilon<\mathrm{fdr}\left(p_j\right)\le\zeta+\epsilon)\right| \\ &+2\sup_{\zeta\in(0,1]}\left|\frac{1}{J_{00}}\sum_{j\in\Delta_{00}}\mathbb{I}\left\{\mathrm{fdr}\left(p_j\right)\le\zeta\right\}-\frac{1}{J_{00}}\sum_{j\in\Delta_{00}}\mathbb{P}\left\{\mathrm{fdr}\left(p_j\right)\le\zeta\right\}\right| \\ &+\frac{1}{J_{00}\epsilon}\sum_{j\in\Delta_{00}}\left|\widehat{\mathrm{fdr}}\left(p_j\right)-\mathrm{fdr}\left(p_j\right)\right| \\ \le&\sup_{\zeta\in(0,1]}\frac{1}{J_{00}}\sum_{j\in\Delta_{00}}\left|\mathbb{P}(\zeta-\epsilon<\mathrm{fdr}\left(p_j\right)\le\zeta+\epsilon)\right|+o_p(1).\end{aligned}$$

Since ϵ can be arbitrarily small, $\sup_{\zeta\in(0,1]}\frac{1}{J_{00}}\sum_{j\in\Delta_{00}}|\mathbb{P}(\zeta-\epsilon<\mathrm{fdr}\left(p_j\right)\le\zeta+\epsilon)|$ can be small due to Assumption (ii). Consequently, we have $Q=o_p(1)$ and thus the first term holds.

Before addressing the second term, we obtain that

$$\begin{aligned}&\mathbb{P}\left(\widehat{\mathrm{fdr}}\left(p_j\right)\le\zeta\right) \\ \le&\mathbb{P}\left(\widehat{\mathrm{fdr}}\left(p_j\right)\le\zeta,\mathrm{fdr}\left(p_j\right)\le\zeta+\epsilon\right)+\mathbb{P}\left(\widehat{\mathrm{fdr}}\left(p_j\right)\le\zeta,\mathrm{fdr}\left(p_j\right)>\zeta+\epsilon\right) \\ \le&\mathbb{P}\left(\mathrm{fdr}\left(p_j\right)\le\zeta+\epsilon\right)+\mathbb{P}\left(\left|\widehat{\mathrm{fdr}}\left(p_j\right)-\mathrm{fdr}\left(p_j\right)\right|>\epsilon\right).\end{aligned} \tag{A.50}$$

Combine with Assumption (i), we can derive that $\mathbb{P}\left(\left|\widehat{\mathrm{fdr}}\left(p_j\right)-\mathrm{fdr}\left(p_j\right)\right|>\epsilon\right) \rightarrow 0$.

Then, we have

$$
\begin{aligned}
& \sup_{\zeta \in(0,1]} \sum_{j \in \Delta_{00}}\left|\mathbb{P}\left(p_j \in \widehat{S}(\zeta)\right)-\mathbb{P}\left(p_j \in S(\zeta)\right)\right| \\
& =\sup_{\zeta \in(0,1]} \frac{1}{J_{00}} \sum_{j \in \Delta_{00}}\left|\mathbb{P}\left(\widehat{\mathrm{fdr}}\left(p_j\right) \leq \zeta\right)-\mathbb{P}\left(\mathrm{fdr}\left(p_j\right) \leq \zeta\right)\right| \\
& \leq \sup_{\zeta \in(0,1]} \frac{1}{J_{00}} \sum_{j \in \Delta_{00}}\left|\mathbb{P}\left(\mathrm{fdr}\left(p_j\right) \leq \zeta+\epsilon\right)-\mathbb{P}\left(\mathrm{fdr}\left(p_j\right) \leq \zeta\right)\right| \\
& =\sup_{\zeta \in(0,1]} \frac{1}{J_{00}} \sum_{j \in \Delta_{00}}\left|\mathbb{P}\left(\zeta<\mathrm{fdr}\left(p_j\right) \leq \zeta+\epsilon\right)\right| .
\end{aligned} \tag{A.51}
$$

As ϵ can be arbitrarily small, $\sup_{\zeta \in(0,1]} \frac{1}{J_{00}} \sum_{j \in \Delta_{00}}\left|\mathbb{P}\left(\zeta<\mathrm{fdr}\left(p_j\right) \leq \zeta+\epsilon\right)\right| \rightarrow 0$.

The third term can be proved using the Glivenko-Cantelli theorem. Thus, we have shown the proof of the first claim in Lemma 2.

Cases 2-3 Under H_{01} and H_{10}: Following the similar procedure in Case 1, we can conclude the proof of the second and third claims in Lemma 2.

Case 4 According to the symmetric property of p-values p_j for $j \in \Delta_{00} \cup \Delta_{01} \cup \Delta_{10}$, we follow the similar steps in Case 1, thus, we have

$$
\sup_{\zeta \in(0,1]}\left|\widehat{K}_J^0(\zeta)-\overline{K}_J^0(\zeta)\right| \longrightarrow 0. \tag{A.52}
$$

This concludes the proof of the fourth claim.

□

Proof of Theorem 2. For any $\epsilon \in(0, \alpha)$, suppose there exists $\zeta_{\alpha-\epsilon}>0$, then

$$
\mathbb{P}\left(\operatorname{FDP}\left(\zeta_{\alpha-\epsilon}\right) \leq \alpha-\epsilon\right) \rightarrow 1 .
$$

By Lemma 2, for any constant $c>0$, we have

$$
\sup_{0<\zeta \leq c}\left|\operatorname{FDP}_J^{\dagger}(\zeta)-\operatorname{FDP}_J(\zeta)\right| \xrightarrow{p} 0. \tag{A.53}
$$

By the definition of $\zeta^{\star}$, i.e., $\zeta^{\star}=\sup \left\{\zeta \in(0,1]: \operatorname{FDP}_J^{\dagger}(\zeta) \leq \alpha\right\}$, we have

$$
\begin{aligned}
\mathbb{P}\left(\zeta^{\star} \geq \zeta_{\alpha-\epsilon}\right) \geq & \mathbb{P}\left(\operatorname{FDP}_J^{\dagger}\left(\zeta_{\alpha-\epsilon}\right) \leq \alpha\right) \\
& \geq \mathbb{P}\left(\left|\operatorname{FDP}_J^{\dagger}\left(\zeta_{\alpha-\epsilon}\right)-\operatorname{FDP}_J\left(\zeta_{\alpha-\epsilon}\right)\right| \leq \epsilon, \operatorname{FDP}\left(\zeta_{\alpha-\epsilon}\right) \leq \alpha-\epsilon\right) \\
& =\mathbb{P}\left(\operatorname{FDP}\left(\zeta_{\alpha-\epsilon}\right) \leq \alpha-\epsilon\right)-\mathbb{P}\left(\left|\operatorname{FDP}_J^{\dagger}\left(\zeta_{\alpha-\epsilon}\right)-\operatorname{FDP}_J\left(\zeta_{\alpha-\epsilon}\right)\right|>\epsilon, \operatorname{FDP}\left(\zeta_{\alpha-\epsilon}\right) \leq \alpha-\epsilon\right) \\
& \geq \mathbb{P}\left(\operatorname{FDP}\left(\zeta_{\alpha-\epsilon}\right) \leq \alpha-\epsilon\right)-\mathbb{P}\left(\left|\operatorname{FDP}_J^{\dagger}\left(\zeta_{\alpha-\epsilon}\right)-\operatorname{FDP}_J\left(\zeta_{\alpha-\epsilon}\right)\right|>\epsilon\right) \\
& \geq 1-\epsilon,
\end{aligned} \tag{A.54}
$$

for J large enough. Thus, we have

$$
\mathbb{P}\left(\zeta^{\star} \geq \zeta_{\alpha-\epsilon}\right) \geq 1-\epsilon. \tag{A.55}
$$

Conditioning on the event $\zeta^{\star} \geq \zeta_{\alpha-\epsilon}$, we have

$$
\begin{aligned}
\limsup_{n,J\to\infty} \mathbb{E}\left[\mathrm{FDP}_J\left(\zeta^\star\right)\right] &\leq \limsup_{n,J\to\infty} \mathbb{E}\left[\mathrm{FDP}_J\left(\zeta^\star\right) \mid \zeta^\star \geq \zeta_{\alpha-\epsilon}\right] \mathbb{P}\left(\zeta^\star \geq \zeta_{\alpha-\epsilon}\right) + \epsilon \\
&\leq \limsup_{n,J\to\infty} \mathbb{E}\left[\left|\mathrm{FDP}_J\left(\zeta^\star\right) - \overline{\mathrm{FDP}}_J\left(\zeta^\star\right)\right| \mid \zeta^\star \geq \zeta_{\alpha-\epsilon}\right] \mathbb{P}\left(\zeta^\star \geq \zeta_{\alpha-\epsilon}\right) \\
&+ \limsup_{n,J\to\infty} \mathbb{E}\left[\left|\mathrm{FDP}^\dagger_J\left(\zeta^\star\right) - \overline{\mathrm{FDP}}_J\left(\zeta^\star\right)\right| \mid \zeta^\star \geq \zeta_{\alpha-\epsilon}\right] \mathbb{P}\left(\zeta^\star \geq \zeta_{\alpha-\epsilon}\right) \\
&+ \limsup_{n,J\to\infty} \mathbb{E}\left[\mathrm{FDP}^\dagger_J\left(\zeta^\star\right) \mid \zeta^\star \geq \zeta_{\alpha-\epsilon}\right] \mathbb{P}\left(\zeta^\star \geq \zeta_{\alpha-\epsilon}\right) + \epsilon \\
&\leq \limsup_{n,J\to\infty} \mathbb{E}\left[\sup_{\zeta\in[\zeta_{\alpha-\epsilon},1]} \left|\mathrm{FDP}_J(\zeta) - \overline{\mathrm{FDP}}_J(\zeta)\right|\right] \\
&+ \limsup_{n,J\to\infty} \mathbb{E}\left[\sup_{\zeta\in[\zeta_{\alpha-\epsilon},1]} \left|\mathrm{FDP}^\dagger_J(\zeta) - \overline{\mathrm{FDP}}_J(\zeta)\right|\right] \\
&+ \limsup_{n,J\to\infty} \mathbb{E}\left[\mathrm{FDP}^\dagger_J\left(\zeta^\star\right)\right] + \epsilon.
\end{aligned} \tag{A.56}
$$

The first two terms are 0 based on Lemma 2 and the dominated convergence theorem. For the third term, we have $\mathrm{FDP}^\dagger_J\left(\zeta^\star\right) \leq \alpha$ by the definition of $\zeta^\star$. This concludes the proof of Theorem 2. □

References for Appendix

[1] R. Barfield, J. Shen, A. C. Just, P. S. Vokonas, J. Schwartz, A. A. Baccarelli, T. J. VanderWeele, and X. Lin. Testing for the indirect effect under the null for genome-wide mediation analyses. *Genetic epidemiology*, 41(8):824–833, 2017.

[2] S. Benafif, Z. Kote-Jarai, and R. A. Eeles. A review of prostate cancer genome-wide association studies (gwas). *Cancer Epidemiology, Biomarkers & Prevention*, 27(8):845–857, 2018.

[3] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300, 1995.

[4] C. R. Bodle, D. I. Mackie, and D. L. Roman. Rgs17: an emerging therapeutic target for lung and prostate cancers. *Future Medicinal Chemistry*, 5(9):995–1007, 2013.

[5] J. Y. Dai, J. L. Stanford, and M. LeBlanc. A multiple-testing procedure for high-dimensional mediation hypotheses. *Journal of the American Statistical Association*, 117(537):198–213, 2022.

[6] A. W. Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge University Press, 2000.

[7] J. Du, X. Zhou, D. Clark-Boucher, W. Hao, Y. Liu, J. A. Smith, and B. Mukherjee. Methods for large-scale single mediator hypothesis testing: Possible choices and comparisons. *Genetic Epidemiology*, 47(2):167–184, 2023.

[8] A. Dvoretzky, J. Kiefer, and J. Wolfowitz. Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *The Annals of Mathematical Statistics*, 27(3):642–669, 1956.

[9] C. Grisanzio, L. Werner, D. Takeda, B. C. Awoyemi, M. M. Pomerantz, H. Yamada, P. Sooriakumaran, B. D. Robinson, R. Leung, A. C. Schinzel, et al. Genetic and functional analyses implicate the nudt11, hnf1b, and slc22a3 genes in prostate cancer pathogenesis. *Proceedings of the National Academy of Sciences*, 109(28):11252–11257, 2012.

[10] X. Guo, R. Li, J. Liu, and M. Zeng. Estimations and tests for generalized mediation models with high-dimensional potential mediators. *Journal of Business & Economic Statistics*, pages 1–14, 2023.

[11] H. M. J. Hung, R. T. O'Neill, P. Bauer, and K. Kohne. The behavior of the p-value when the alternative hypothesis is true. *Biometrics*, 53(1):11–22, 1997.

[12] N. L. Johnson, S. Kotz, and N. Balakrishnan. *Continuous univariate distributions*. Wiley, New York., 1995.

[13] H. Liu, B. Wang, and C. Han. Meta-analysis of genome-wide and replication association studies on prostate cancer. *The Prostate*, 71(2):209–224, 2011.

[14] Z. Liu, J. Shen, R. Barfield, J. Schwartz, A. A. Baccarelli, and X. Lin. Large-scale hypothesis testing for causal mediation effects with applications in genome-wide epigenetic studies. *Journal of the American Statistical Association*, 117(537):67–81, 2022.

[15] M. O. Mosig, E. Lipkin, G. Khutoreskaya, E. Tchourzyna, M. Soller, and A. Friedmann. A whole genome scan for quantitative trait loci affecting milk protein percentage in israeli-holstein cattle, by means of selective milk dna pooling in a daughter design, using an adjusted false discovery rate criterion. *Genetics*, 157(4):1683–1698, 2001.

[16] Y. Song, X. Zhou, M. Zhang, W. Zhao, Y. Liu, S. L. R. Kardia, A. V. D. Roux, B. L. Needham, J. A. Smith, and B. Mukherjee. Bayesian shrinkage estimation of high dimensional causal mediation effects in omics studies. *Biometrics*, 76(3):700–710, 2020.

[17] J. D. Storey, J. E. Taylor, and D. Siegmund. Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(1):187–205, 2004.

[18] X. Xu, W. M. Hussain, J. Vijai, K. Offit, M. A. Rubin, F. Demichelis, and R. J. Klein. Variants at irx4 as prostate cancer expression quantitative trait loci. *European Journal of Human Genetics*, 22(4):558–563, 2014.

[19] H. Zhang, Y. Zheng, L. Hou, C. Zheng, and L. Liu. Mediation analysis for survival data with high-dimensional mediators. *Bioinformatics*, 37(21):3815–3821, 2021.