To compare $P$ and $P_a$, we can equivalently characterize these distributions as follows:

- Draw $X \sim P_X$.
- Conditional on $X$, draw $Z \mid X \in \mathcal{X}_m \sim \text{Bernoulli}(0.5)$ (for the distribution $P$, or for the distribution $P_a$ if $m = 1$), or $Z \mid X \in \mathcal{X}_m \sim \text{Bernoulli}(0.5 + a_m \epsilon)$ (for the distribution $P_a$ if $m \geq 2$).
- Conditional on $X, Z$ draw $Y$ as

$$Y \mid X = x, Z = z \sim P^{\epsilon}_Y \mid X = x.$$  

Define $\tilde{P}$ as the distribution over $(X, Y, Z)$ induced by $P$, and $\tilde{P}_a$ as the distribution over $(X, Y, Z)$ induced by $P_a$. Then the marginal distribution of $(X, Y)$ under $\tilde{P}$ and under $\tilde{P}_a$ is given by $P$ and by $P_a$, respectively.

Now consider comparing two distributions on triples $(X_1, Z_1, Y_1), \ldots, (X_n, Z_n, Y_n)$. We will compare $\tilde{P}_n$ versus the mixture distribution $\tilde{P}_\text{mix}$ defined as follows:

- Draw $A_1, A_2, \ldots \sim \text{Unif}\{\pm 1\}$.
- Conditional on $A_1, A_2, \ldots$, draw $(X_1, Y_1, Z_1), \ldots, (X_n, Y_n, Z_n) \sim \tilde{P}_A$.

Since in our characterization above, the distribution of $Y_1, \ldots, Y_n$ conditional on $X_1, \ldots, X_n$ and on $Z_1, \ldots, Z_n$ is the same for both, the only difference lies in the conditional distribution of $Z_1, \ldots, Z_n$ given $X_1, \ldots, X_n$. Therefore, we can apply Lemma 2 with $\epsilon_1 = 0$ and $\epsilon_2 = \epsilon_3 = \cdots = \epsilon$ to obtain

$$d_{TV}(\tilde{P}_\text{mix}, \tilde{P}^n) \leq 2n \sqrt{\sum_{m \geq 2} \epsilon^4 p_m^2}.$$  

Now let $P_\text{mix}$ be the marginal distribution of $(X_1, Y_1), \ldots, (X_n, Y_n)$ under $\tilde{P}_\text{mix}$. Noting that $P^n$ is the marginal distribution of $(X_1, Y_1), \ldots, (X_n, Y_n)$ under $\tilde{P}^n$, we therefore have

$$d_{TV}(P_\text{mix}, P^n) \leq d_{TV}(\tilde{P}_\text{mix}, \tilde{P}^n) \leq 2n \sqrt{\sum_{m \geq 2} \epsilon^4 p_m^2}.$$  

35th Conference on Neural Information Processing Systems (NeurIPS 2021).
C Proof of Theorem 2

First, define $p_m = P_{X^m} \{ X = x^{(m)} \}$. The following lemma establishes some results on its support, expected value, and concentration properties of $Z$:

Lemma C.1. For $Z$ and $N \geq 2$ defined as in (4) and (5), the following holds:

$$E[Z] = \sum_{m=1}^{\infty} (\mu(x^{(m)}) - \mu_P(x^{(m)}))^2 \cdot (np_m - 1 + (1 - p_m)^n),$$

$$E[Z | X_1, \ldots, X_n] = \sum_{m=1}^{\infty} (np_m - 1 + (1 - p_m)^n) \cdot (\mu(x^{(m)}) - \mu_P(x^{(m)}))^2,$$

$$\text{Var}(E[Z | X_1, \ldots, X_n]) \leq 2E[Z],$$

$$\text{Var}(Z | X_1, \ldots, X_n) \leq N_{\geq 2} + 2E[Z | X_1, \ldots, X_n].$$

In particular, the first part of the lemma will allow us to use $E[Z]$ to bound the error in $\mu$—here the calculations are similar to those in Chan et al. [2014] for the setting of testing discrete distributions. Recalling the definition of $M_{\gamma}^*(P_X)$ given in (6), define

$$\Delta = \sqrt{\frac{2M_{\gamma}^*(P_X) + n}{n(n-1)}},$$

and

$$\sum_{m=1}^{M_{\gamma}^*(P_X)} p_m |\mu(x^{(m)}) - \mu_P(x^{(m)})| \leq \frac{M_{\gamma}^*(P_X)}{\sqrt{2 + np_m}} \cdot \sqrt{2 + np_m},$$

$$\frac{M_{\gamma}^*(P_X)}{n(n-1)} \cdot \sqrt{2M_{\gamma}^*(P_X) + n}$$

where the next-to-last step holds by the following identity:

Lemma C.2. For all $n \geq 1$ and $p \in [0, 1]$, $np - 1 + (1 - p)^n \geq \frac{n(n-1)p^2}{2 + np}$. [2]

Next, we will use Lemma C.1 to relate $\Delta$ and $\hat{\Delta}$. By Chebyshev’s inequality, conditional on $X_1, \ldots, X_n$, with probability at least $1 - \delta/4$ we have

$$Z \geq E[Z | X_1, \ldots, X_n] - \sqrt{\frac{\text{Var}(Z | X_1, \ldots, X_n)}{\delta/4}},$$

which can be relaxed to

$$E[Z | X_1, \ldots, X_n] \leq 2Z + 4 \sqrt{\frac{N_{\geq 2}}{\delta} + 8\delta}.$$

Marginalizing over $X_1, \ldots, X_n$, this bound holds with probability at least $1 - \delta/4$. Moreover, again applying Chebyshev’s inequality, with probability at least $1 - \delta/4$ we have

$$E[Z | X_1, \ldots, X_n] \geq E[Z] - \sqrt{\frac{\text{Var}(E[Z | X_1, \ldots, X_n])}{\delta/4}},$$

which can be relaxed to

$$E[Z] \leq 2E[Z | X_1, \ldots, X_n] + 8\delta.$$
Combining our bounds, then, we have $\mathbb{E}[Z] \leq 4Z + 8\sqrt{N_{\geq2}/\delta} + 24/\delta$ with probability at least $1 - \delta/2$. Since $\mathbb{P}\{\hat{M}_\gamma \geq M_\gamma^*(P_X)\} \geq 1 - \delta/2$ by Hoeffding’s inequality, this implies that

$$\mathbb{P}\{\hat{\Delta} \geq \Delta\} \geq 1 - \delta.$$

Now we verify the coverage properties of $\hat{C}_n$. We have

$$\mathbb{P}\{\mu_P(X_{n+1}) \notin \hat{C}_n(X_{n+1})\} = \mathbb{P}\{\mu_P(X_{n+1}) - \mu(X_{n+1}) > (\alpha - \delta - \gamma)^{-1}\hat{\Delta}\}$$

$$\leq \mathbb{P}\{\hat{\Delta} < \Delta\} + \mathbb{P}\{\|\mu_P(X_{n+1}) - \mu(X_{n+1})\| > (\alpha - \delta - \gamma)^{-1}\Delta\}$$

$$\leq \mathbb{P}\{\hat{\Delta} < \Delta\} + \mathbb{P}\{X_{n+1} \notin \{x(1), \ldots, x(M_\gamma^*(P_X))\}\}$$

$$\quad + \sum_{m=1}^{M_\gamma^*(P_X)} \mathbb{P}\{X_{n+1} = x(m), \|\mu_P(x_{n+1}) - \mu(x_{n+1})\| > (\alpha - \delta - \gamma)^{-1}\Delta\}$$

$$\leq \delta + \gamma + \sum_{m=1}^{M_\gamma^*(P_X)} \mathbb{P}\{X_{n+1} = x(m), \|\mu_P(x_{n+1}) - \mu(x_{n+1})\| > (\alpha - \delta - \gamma)^{-1}\Delta\}$$

$$\leq \delta + \gamma + \sum_{m=1}^{M_\gamma^*(P_X)} p_m \mathbb{P}\{\mu_P(x(m)) - \mu(x(m))\| > (\alpha - \delta - \gamma)^{-1}\Delta\}$$

$$\leq \delta + \gamma + \frac{\Delta}{(\alpha - \delta - \gamma)^{-1}\Delta} = \alpha,$$

which verifies the desired coverage guarantee.

### D Proof of Theorem

First, we have $\hat{M}_\gamma \leq M$ almost surely by our assumption on $P_X$. Next we need to bound $\mathbb{E}[Z_]$.

We have

$$\mathbb{E}[Z_] \leq \mathbb{E}[(Z - \mathbb{E}[Z | X_1, \ldots, X_n])_+]$$

since this conditional expectation is nonnegative

$$\leq \sqrt{\mathbb{E}[(Z - \mathbb{E}[Z | X_1, \ldots, X_n])^2]}$$

$$= \sqrt{\mathbb{E}[(Z - \mathbb{E}[Z | X_1, \ldots, X_n])^2 | X_1, \ldots, X_n]}$$

$$= \sqrt{\mathbb{E}[(Z | X_1, \ldots, X_n) - \mathbb{E}[Z | X_1, \ldots, X_n])^2]}$$

$$\leq \sqrt{\mathbb{E}[N_{\geq2} + 2\mathbb{E}[Z | X_1, \ldots, X_n)]}$$

$$= \sqrt{\mathbb{E}[N_{\geq2} + 2\mathbb{E}[Z]}. $$

We then have

$$\mathbb{E}[Z] = \mathbb{E}[Z] + \mathbb{E}[Z_] \leq \mathbb{E}[Z] + \sqrt{2\mathbb{E}[Z] + \mathbb{E}[N_{\geq2}]} \leq 1.5\mathbb{E}[Z] + 1 + \sqrt{\mathbb{E}[N_{\geq2}].$$

Next we need a lemma:

**Lemma D.1.** For all $n \geq 1$ and $p \in [0, 1]$, $np - 1 + (1 - p)^n \leq \frac{n^2p^2}{1+np}.$
Combined with the calculation of \( \mathbb{E} [Z] \) in Lemma \( \text{C.1} \) we have

\[
\mathbb{E} [Z] \leq \sum_{m=1}^{M} (\mu(x^{(m)}) - \mu_P(x^{(m)}))^2 \cdot \frac{n^2 \rho_m^2}{1 + n \rho_m} \\
\leq \sum_{m=1}^{M} p_m \cdot (\mu(x^{(m)}) - \mu_P(x^{(m)}))^2 \cdot \frac{n^2 \cdot \eta/M}{1 + n \cdot \eta/M} \\
= \frac{\eta n^2}{M + \eta n} \cdot \mathbb{E}_{\mathbb{P}_X} \left[ (\mu_P(X) - \mu(X))^2 \right] \\
\leq (\text{err}_\mu)^2 \cdot \frac{\eta n^2}{M + \eta n},
\]

since we have assumed that \( \mathbb{P}_X \) is supported on \( \{x^{(1)}, \ldots, x^{(M)}\} \) and that \( \mathbb{P}_{\mathbb{P}_X} \{X = x^{(m)}\} \leq \eta/M \) for all \( m \), where we must have \( \eta \geq 1 \). Furthermore, we have

\[
\mathbb{E} [N_{\geq 2}] = \sum_{m=1}^{M} \mathbb{P} \{n_m \geq 2\} \leq \sum_{m=1}^{M} \mathbb{E} [(n_m - 1)_{+}] \\
= \sum_{m=1}^{M} n \cdot \mathbb{P}_{\mathbb{P}_X} \{X = x^{(m)}\} - 1 + \left( 1 - \mathbb{P}_{\mathbb{P}_X} \{X = x^{(m)}\} \right)^n \text{ as calculated as in the proof of Lemma \( \text{C.1} \)} \\
\leq \sum_{m=1}^{M} n \cdot \eta/M - 1 + (1 - \eta/M)^n \\
\leq \sum_{m=1}^{M} n^2 (\eta/M)^2 \frac{1}{1 + n \eta/M} \text{ by Lemma \( \text{D.1} \)} \\
= \frac{\eta^2 n^2}{M + \eta n}.
\]

We also have \( N_{\geq 2} \leq M \) almost surely, and so combining these two bounds, \( \mathbb{E} [N_{\geq 2}] \leq \min \{ \frac{n^2 n^2}{M}, M \} \). Combining everything, then,

\[
\mathbb{E} [Z_{\ast}] \leq 1.5(\text{err}_\mu)^2 \cdot \frac{\eta n^2}{M + \eta n} + 1 + \sqrt{\min \left\{ \frac{\eta^2 n^2}{M}, M \right\}}.
\]

Plugging these calculations into the definition of \( \bar{\Delta} \), we obtain

\[
\mathbb{E} [\bar{\Delta}] = \mathbb{E} \left[ \frac{2M_n + n}{n(n-1)} \cdot \sqrt{4Z_{\ast} + 8 \sqrt{N_{\geq 2} / \delta} + 24 / \delta} \right] \\
\leq \mathbb{E} \left[ \frac{2M_n + n}{n(n-1)} \cdot \sqrt{4\mathbb{E} [Z_{\ast}] + 8 \sqrt{\mathbb{E} [N_{\geq 2} / \delta] + 24 / \delta} \right] \\
\leq \frac{2M_n + n}{n(n-1)} \cdot \sqrt{4 \mathbb{E} [Z_{\ast}] + 8 \sqrt{\mathbb{E} [N_{\geq 2} / \delta] + 24 / \delta}} \\
\leq \sqrt{\frac{2M_n + n}{n(n-1)}} \cdot \sqrt{4 \left( 1.5(\text{err}_\mu)^2 \cdot \frac{\eta n^2}{M + \eta n} + 1 + \sqrt{\min \left\{ \frac{\eta^2 n^2}{M}, M \right\}} \right) + 8 \sqrt{\min \left\{ \frac{n^2}{M}, M \right\}} \cdot 1/\delta + 24 / \delta} \\
\leq \sqrt{\frac{2M_n + n}{n(n-1)}} \cdot \sqrt{6(\text{err}_\mu)^2 \cdot \frac{\eta n^2}{M + \eta n} + \sqrt{4(1 + 2/\sqrt{\delta}) \sqrt{\min \left\{ \frac{\eta^2 n^2}{M}, M \right\}} + \sqrt{4 + 24 / \delta}}.}
\]

We can assume that \( M \leq n^2 \) and \( n \geq 2 \) (as otherwise, the upper bound would be trivial, since we must have \( \text{Leb}(\bar{C}_n(X_{n+1})) \leq 1 \) by construction). If \( M \geq n \), then \( \frac{2M + n}{n(n-1)} \leq \frac{6M}{n^2} \) and the above
simplifies to
\[
\mathbb{E} \left[ \hat{\Delta} \right] \leq 6\sqrt{\eta} \cdot \text{err}_\mu + \sqrt{\frac{6(4 + 24/\delta)M}{n^2}} + \sqrt{24\eta(1 + 2/\sqrt{\delta})}\sqrt{\frac{M}{n^2}},
\]
and since we assume $M \leq n^2$, we therefore have
\[
\mathbb{E} \left[ \hat{\Delta} \right] \leq 6\sqrt{\eta} \cdot \text{err}_\mu + \left( \sqrt{6(4 + 24/\delta)} + \sqrt{24\eta(1 + 2/\sqrt{\delta})} \right) \cdot \sqrt{\frac{M}{n^2}}. \tag{D.2}
\]
If instead $M < n$, then $\frac{2M + n}{n(n-1)} \leq \frac{6}{n}$ and the above bound on $\mathbb{E} \left[ \hat{\Delta} \right]$ simplifies to
\[
\mathbb{E} \left[ \hat{\Delta} \right] \leq 6\sqrt{\eta} \cdot \text{err}_\mu + \left( \frac{6}{n} \right) \cdot \left( \sqrt{4(1 + 2/\sqrt{\delta})M} + \sqrt{4 + 24/\delta} \right),
\]
which again yields the same bound (D.2) since $M \geq 1$ and $\eta \geq 1$. Finally, by definition of $\hat{C}_n(X_{n+1})$, we have
\[
\mathbb{E} \left[ \text{Leb}(\hat{C}_n(X_{n+1})) \right] \leq \mathbb{E} \left[ \hat{\Delta} \right] \cdot \frac{2}{\alpha - \delta - \gamma},
\]
which completes the proof for $c$ chosen appropriately as a function of $\alpha$, $\delta$, $\gamma$, $\eta$.

E Proofs of lemmas

E.1 Proof of Lemma 1

Let $x_{\text{med}}$ be the median of $Q$. Define
\[
q_\lt = \mathbb{P}_Q \{ X < x_{\text{med}} \}, \quad q_\gt = \mathbb{P}_Q \{ X > x_{\text{med}} \},
\]
and note that $q_\lt$, $q_\gt \in [0, 0.5]$. For $X \sim Q$, let $Q_\lt$ be the distribution of $X$ conditional on $X < x_{\text{med}}$ and let $Q_\gt$ be the distribution of $X$ conditional on $X > x_{\text{med}}$. Then we can write
\[
Q = q_\lt \cdot Q_\lt + (1 - q_\lt - q_\gt) \cdot \delta_{x_{\text{med}}} + q_\gt \cdot Q_\gt,
\]
where $\delta_t$ denotes the point mass distribution at $t$. Now define
\[
Q_0 = 2q_\lt \cdot Q_\lt + (1 - 2q_\lt) \cdot \delta_{x_{\text{med}}}
\]
and
\[
Q_1 = 2q_\gt \cdot Q_\gt + (1 - 2q_\gt) \cdot \delta_{x_{\text{med}}}.
\]
Then clearly $Q = 0.5Q_0 + 0.5Q_1$. Next let $\mu_0, \mu_1$ be the means of these two distributions, satisfying $\frac{\mu_0 + \mu_1}{2} = \mu$ where $\mu$ is the mean of $Q$, and let $\sigma_0^2, \sigma_1^2$ be the variances of these two distributions. By the law of total variance, we have
\[
\sigma^2 = \text{Var}(0.5\delta_{\mu_0} + 0.5\delta_{\mu_1}) + \mathbb{E} \left[ 0.5\delta_{\sigma_0^2} + 0.5\delta_{\sigma_1^2} \right]
\]
\[
= \left( \mu_1 - \mu_0 \right)^2 + 0.5\sigma_0^2 + 0.5\sigma_1^2.
\]
Next, $Q_0$ is a distribution supported on $[0, x_{\text{med}}]$ with mean $\mu_0$, so its variance is bounded as
\[
\sigma_0^2 \leq \mu_0(x_{\text{med}} - \mu_0),
\]
where the maximum is attained if all the mass is placed on the endpoints 0 or $x_{\text{med}}$. Similarly, $Q_1$ is a distribution supported on $[x_{\text{med}}, 1]$ with mean $\mu_1$, so its variance is bounded as
\[
\sigma_1^2 \leq (1 - \mu)(\mu_1 - x_{\text{med}}).
\]
Using the fact that $\frac{\mu_0 + \mu_1}{2} = \mu$, we can simplify to
\[
\sigma_0^2 + \sigma_1^2 \leq \mu_0(x_{\text{med}} - \mu_0) + (1 - \mu)(\mu_1 - x_{\text{med}})
\]
\[
= \mu(x_{\text{med}} - \mu_0) + (1 - \mu)(\mu_1 - x_{\text{med}}) - 0.5(\mu_1 - \mu_0)^2.
\]
Therefore, we have
\[
\sigma^2 = \frac{(\mu_1 - \mu_0)^2}{4} + 0.5\sigma_0^2 + 0.5\sigma_1^2 \leq 0.5\mu(x_{\text{med}} - \mu_0) + 0.5(1 - \mu)(\mu_1 - x_{\text{med}})
\]
\[
= 0.5(2\mu - 1)x_{\text{med}} - 0.5\mu_0 + 0.5(1 - \mu)\mu_1 = 0.5(2\mu - 1)(x_{\text{med}} - \mu) + 0.25(\mu_1 - \mu_0).
\]
Next, $|2\mu - 1| \leq 1$ since $\mu \in [0, 1]$, and $|x_{\text{med}} - \mu| \leq 0.5|\mu_1 - \mu_0|$ since $\mu_0 \leq x_{\text{med}} \leq \mu_1$ and $\frac{\mu_0 + \mu_1}{2} = \mu$. Therefore, $\sigma^2 \leq 0.5(\mu_1 - \mu_0)$, proving the lemma.
E.2 Proof of Lemma 2

First we need a supporting lemma.

**Lemma E.1.** For any $N \geq 1$ and any $\epsilon \in [0, 0.5]$, 
\[ d_{KL}(0.5 \cdot \text{Binom}(N, 0.5 + \epsilon) + 0.5 \cdot \text{Binom}(N, 0.5 - \epsilon) \parallel \text{Binom}(N, 0.5)) \leq 8N(N-1)\epsilon^4. \]

**Proof of Lemma E.1.** Let $f_0$ be the probability mass function of the Binom$(N, 0.5)$ distribution, and let $f_1$ be the probability mass function of the mixture $0.5 \cdot \text{Binom}(N, 0.5+\epsilon) + 0.5 \cdot \text{Binom}(N, 0.5-\epsilon)$. Then we would like to bound $d_{KL}(f_1 \parallel f_0)$. We calculate the ratio

\[
\frac{f_1(k)}{f_0(k)} = \frac{0.5 \cdot \binom{N}{k}(0.5 + \epsilon)^k(0.5 - \epsilon)^{N-k} + 0.5 \cdot \binom{N}{k}(0.5 - \epsilon)^k(0.5 + \epsilon)^{N-k}}{\binom{N}{k}(0.5)^N} \\
= \frac{(1 + 2\epsilon)^k(1 - 2\epsilon)^{N-k} + (1 - 2\epsilon)^k(1 + 2\epsilon)^{N-k}}{2}.
\]

Therefore, it holds that

\[
E_{\text{Binom}(N,0.5)} \left( \frac{f_1(X)}{f_0(X)} \right)^2 \\
= E_{\text{Binom}(N,0.5)} \left( \frac{(1 + 2\epsilon)^X(1 - 2\epsilon)^{N-X} + (1 - 2\epsilon)^X(1 + 2\epsilon)^{N-X}}{2} \right)^2 \\
= E_{\text{Binom}(N,0.5)} \left( \frac{(1 + 2\epsilon)^2X - 2N\cdot 2\epsilon(1 - 2\epsilon)^{N-X} + (1 - 2\epsilon)^2X(1 + 2\epsilon)^{N-X} + 2(1 - 4\epsilon^2)^N}{4} \right) \\
= \frac{(1 - 2\epsilon)^2N E_{\text{Binom}(N,0.5)} \left( \left( \frac{1 + 2\epsilon}{1 - 2\epsilon} \right)^2 \right)^N}{4} + (1 + 2\epsilon)^2N E_{\text{Binom}(N,0.5)} \left( \left( \frac{1 - 2\epsilon}{1 + 2\epsilon} \right)^2 \right)^N + 2(1 - 4\epsilon^2)^N \\
= \frac{0.5(1 + 2\epsilon)^2 + 0.5(1 - 2\epsilon)^2}{4} + 0.5(1 - 2\epsilon)^2 + 0.5(1 + 2\epsilon)^2 + 2(1 - 4\epsilon^2)^N \\
= \frac{(1 + 4\epsilon^2)^N + (1 - 4\epsilon^2)^N}{2} \\
= 1 + \sum_{k \geq 1} \frac{N(N-1)\ldots(N-2k+2)(N-2k+1)}{(2k)!}(4\epsilon^2)^{2k} \\
\leq 1 + \sum_{k \geq 1} \frac{(N(N-1))^k}{2^k k!}(4\epsilon^2)^{2k} \\
\leq e^{8\epsilon^4N(N-1)}.
\]

Applying Jensen’s inequality, we then have

\[
d_{KL}(f_1 \parallel f_0) = \sum_{k=0}^{n} f_1(k) \log \left( \frac{f_1(k)}{f_0(k)} \right) = E_{f_1} \left[ \log \left( \frac{f_1(X)}{f_0(X)} \right) \right] \leq \log \left( E_{f_1} \left[ \frac{f_1(X)}{f_0(X)} \right] \right) \\
= \log \left( E_{\text{Binom}(N,0.5)} \left[ \left( \frac{f_1(X)}{f_0(X)} \right)^2 \right] \right) \leq \log \left( e^{8\epsilon^4N(N-1)} \right) = 8\epsilon^4N(N-1).
\]
Now we turn to the proof of Lemma 2. Let \( p_m = \mathbb{P} \{ X \in \mathcal{X}_m \} \) for each \( m = 1, 2, \ldots \). Define a distribution \( P_0' \) on \((W, Z) \in \mathbb{N} \times \{0, 1\}\) as:

\[
\text{Draw } W \sim \sum_{m=1}^{\infty} p_m \delta_m, \text{ and draw } Z \sim \text{Bernoulli}(0.5), \text{ independently from } W.
\]

and for any signs \( a_1, a_2, \ldots \in \{\pm 1\} \), define a distribution \( P_a' \) on \((W, Z) \in \mathbb{N} \times \{0, 1\}\) as:

\[
\text{Draw } W \sim \sum_{m=1}^{\infty} p_m \delta_m, \text{ and conditional on } W, \text{ draw } Z|W = m \sim \text{Bernoulli}(0.5 + a_m \cdot \epsilon_m).
\]

Then define \( \tilde{P}_0' = (P_0')^n \) and define \( \tilde{P}_1' \) as the following mixture distribution.

- Draw \( A_1, A_2, \ldots \iiid \text{Unif}\{\pm 1\} \).
- Conditional on \( A_1, A_2, \ldots \), draw \((W_1, Z_1), \ldots, (W_n, Z_n) \iid P_A' \).

Note that \((X_1, Z_1), \ldots, (X_n, Z_n) \sim \tilde{P}_0 \) can be drawn by first drawing \((W_1, Z_1), \ldots, (W_n, Z_n) \sim \tilde{P}_0' \) and then drawing \( X_i|W_i \sim P_{X|X \in X_{W_i}} \) for each \( i \). Similarly, \((X_1, Z_1), \ldots, (X_n, Z_n) \sim \tilde{P}_1 \) is equivalent to first drawing \((W_1, Z_1), \ldots, (W_n, Z_n) \sim \tilde{P}_1' \) and then drawing \( X_i|W_i \sim P_{X|X \in X_{W_i}} \) for each \( i \). This implies \( \text{d} \text{TV}(\tilde{P}_1||\tilde{P}_0) \leq \text{d} \text{TV}(\tilde{P}_1'||\tilde{P}_0') \).

Now we can calculate the probability mass function of \( \tilde{P}_0' \) as

\[
\tilde{P}_0'((w_1, z_1), \ldots, (w_n, z_n)) = \prod_{i=1}^{n} (p_{w_i} \cdot 0.5),
\]

and for \( \tilde{P}_1' \) as

\[
\tilde{P}_1'((w_1, z_1), \ldots, (w_n, z_n)) = \mathbb{E}_{A \iiid \text{Unif}\{\pm 1\}} \left[ \prod_{i=1}^{n} (p_{w_i} \cdot (0.5 + A_{w_i} \epsilon_m)^{z_i} \cdot (0.5 - A_{w_i} \epsilon_m)^{1-z_i}) \right].
\]

Defining summary statistics

\[
n_m = \sum_{i=1}^{n} \mathbb{1} \{ w_i = m \} \text{ and } k_m = \sum_{i=1}^{n} \mathbb{1} \{ w_i = m, z_i = 1 \},
\]

we can rewrite the above as

\[
\tilde{P}_0'((w_1, z_1), \ldots, (w_n, z_n)) = \prod_{m=1}^{\infty} p_m^{n_m} \cdot 0.5^{n_m},
\]

and

\[
\tilde{P}_1'((w_1, z_1), \ldots, (w_n, z_n)) = \mathbb{E}_{A \iiid \text{Unif}\{\pm 1\}} \left[ \prod_{m=1}^{\infty} p_m^{n_m} \cdot (0.5 + A_{m} \epsilon_m)^{k_m} \cdot (0.5 - A_{m} \epsilon_m)^{n_m-k_m} \right]
\]

\[
= \prod_{m=1}^{\infty} p_m^{n_m} \cdot \frac{1}{2} \sum_{a_m \in \{\pm 1\}} (0.5 + a_m \epsilon_m)^{k_m} \cdot (0.5 - a_m \epsilon_m)^{n_m-k_m}
\]

7
We then calculate
\[
d_{\text{KL}}(P_1 || P_0) = \mathbb{E}_{\hat{P}_1} \left[ \log \left( \frac{P_1((W_1, Z_1), \ldots, (W_n, Z_n))}{P_0((W_1, Z_1), \ldots, (W_n, Z_n))} \right) \right]
\]
\[
= \mathbb{E}_{\hat{P}_1} \left[ \log \left( \frac{\prod_{m=1}^{\infty} P_m^{N_m} \cdot \frac{1}{2} \sum_{a_m \in \{\pm 1\}} (0.5 + a_m \epsilon_m) K_m \cdot (0.5 - a_m \epsilon_m) N_m - K_m}{\prod_{m=1}^{\infty} P_m^{N_m} \cdot (0.5) N_m} \right) \right]
\]
\[
= \sum_{m=1}^{\infty} \mathbb{E}_{\hat{P}_1} \left[ \log \left( \frac{\frac{1}{2} \sum_{a_m \in \{\pm 1\}} (0.5 + a_m \epsilon_m) K_m \cdot (0.5 - a_m \epsilon_m) N_m - K_m}{(0.5) N_m} \right) \right] | N_m \right] \] ,

where
\[
N_m = \sum_{i=1}^{n} \mathbb{1} \{ W_i = m \} \text{ and } K_m = \sum_{i=1}^{n} \mathbb{1} \{ W_i = m, Z_i = 1 \} ,
\]
Next, we calculate the conditional expectation in the last expression above. If \( N_m = 0 \) then trivially it is equal to \( \log(1) = 0 \). If \( N_m \geq 1 \), then under \( \hat{P}_1 \), we can see that
\[
K_m \mid N_m \sim 0.5 \cdot \text{Binom}(N_m, 0.5 + \epsilon_m) + 0.5 \cdot \text{Binom}(N_m, 0.5 - \epsilon_m),
\]
and therefore,
\[
\mathbb{E}_{\hat{P}_1} \left[ \log \left( \frac{\frac{1}{2} \sum_{a_m \in \{\pm 1\}} (0.5 + a_m \epsilon_m) K_m \cdot (0.5 - a_m \epsilon_m) N_m - K_m}{(0.5) N_m} \right) \right] | N_m \right] \] = \( d_{\text{KL}} \left( 0.5 \cdot \text{Binom}(N_m, 0.5 + \epsilon_m) + 0.5 \cdot \text{Binom}(N_m, 0.5 - \epsilon_m) \mid \text{Binom}(N_m, 0.5) \right) \leq 8N_m(N_m - 1)\epsilon_m^4,
\]
where the last step applies Lemma [E.1] Therefore,
\[
d_{\text{KL}}(\hat{P}_1 || \hat{P}_0) \leq \sum_{m=1}^{\infty} \mathbb{E}_{\hat{P}_1} \left[ 8N_m(N_m - 1)\epsilon_m^4 \right]
\]
\[
= 8 \sum_{m=1}^{\infty} \epsilon_m^4 \mathbb{E}_{\hat{P}_1} \left[ N_m^2 - N_m \right]
\]
\[
= 8 \sum_{m=1}^{\infty} \epsilon_m^4 \left( (np_m(1 - p_m) + n^2 p_m^2) - np_m \right)
\]
\[
= 8 \cdot n(n - 1) \sum_{m=1}^{\infty} \epsilon_m^2 p_m^2,
\]
since \( N_m \sim \text{Binom}(n, p_m) \) by definition. Applying Pinsker’s inequality and \( d_{\text{TV}}(\hat{P}_1 || \hat{P}_0) \leq d_{\text{TV}}(\hat{P}_1 || \hat{P}_0) \) completes the proof.

E.3 Proof of Lemma[C.1]

Define
\[
Z_m = \begin{cases} 
(n_m - 1) \cdot ((\bar{y}_m - \mu(x^{(m)}))^2 - n_m^{-1} s_m^2), & n_m \geq 2, \\
0, & n_m = 0 \text{ or } 1.
\end{cases}
\]
Then \( Z = \sum_{m=1}^{\infty} Z_m \). Now we calculate the conditional mean and variance. Conditional on \( X_1, \ldots, X_n, \bar{y}_m \) and \( s_m^2 \), are the sample mean and sample variance of \( n_m \) i.i.d. draws from a distribution with mean \( \mu_P(x^{(m)}) \) and variance \( \sigma_P^2(x^{(m)}) \), supported on \([0, 1] \), where we let \( \sigma_P^2(x^{(m)}) \) be the variance of the distribution of \( Y | X = x^{(m)} \), under the joint distribution \( P \). For any \( m \) with \( n_m \geq 2 \), we therefore have
\[
\mathbb{E} [\bar{y}_m \mid X_1, \ldots, X_n] = \mu_P(x^{(m)}), \quad \text{Var}(\bar{y}_m \mid X_1, \ldots, X_n) = n_m^{-1} \sigma_P^2(x^{(m)}) = \mathbb{E} [n_m^{-1} s_m^2 \mid X_1, \ldots, X_n] ,
\]
and so
\[
\mathbb{E} \left[ (\bar{y}_m - \mu(x^{(m)}))^2 - n_m^{-1} s_m^2 \middle| X_1, \ldots, X_n \right] = n_m^{-1} \sigma_P^2(x^{(m)}) + \mu_P(x^{(m)}) - \mu(x^{(m)})^2 - n_m^{-1} \sigma_P^2(x^{(m)}) = (\mu_P(x^{(m)}) - \mu(x^{(m)}))^2.
\]

Next, we have \( (n_1, \ldots, n_M) \sim \text{Multinom}(n, p) \), which implies that marginally \( n_m \sim \text{Binom}(n, p_m) \) and so
\[
\mathbb{E} \left[ (n_m - 1)_+ \right] = \mathbb{E} [n_m - 1 + \mathbb{I} \{n_m = 0\}] = np_m - 1 + (1 - p_m)^n.
\]
Combining these calculations completes the proof for the expected value \( \mathbb{E} [Z] \) and conditional expected value \( \mathbb{E} [Z \mid X_1, \ldots, X_n] \).

Next, we calculate conditional and marginal variance. We have
\[
\text{Var} \left( (\bar{y}_m - \mu(x^{(m)}))^2 - n_m^{-1} s_m^2 \middle| X_1, \ldots, X_n \right) = \text{Var} \left( (\bar{y}_m - \mu(x^{(m)}))^2 - n_m^{-1} s_m^2 - (\mu_P(x^{(m)}) - \mu(x^{(m)}))^2 \right) \middle| X_1, \ldots, X_n)
\]
\[
\leq \mathbb{E} \left[ (\bar{y}_m - \mu(x^{(m)}))^2 - n_m^{-1} s_m^2 - (\mu_P(x^{(m)}) - \mu(x^{(m)}))^2 \right]^2 \middle| X_1, \ldots, X_n)
\]
\[
= \mathbb{E} \left[ (\bar{y}_m - \mu_P(x^{(m)})) (\mu_P(x^{(m)}) - \mu(x^{(m)})) \right)^2 \middle| X_1, \ldots, X_n)
\]
\[
\leq 4 \mathbb{E} \left[ (\bar{y}_m - \mu_P(x^{(m)}))^4 \right] + 2 \mathbb{E} \left[ (2(\bar{y}_m - \mu_P(x^{(m)}))(\mu_P(x^{(m)}) - \mu(x^{(m)})) \right] \middle| X_1, \ldots, X_n
\]
\[
+ 4 \mathbb{E} \left[ (n_m^{-1} s_m^2)^2 \right] \middle| X_1, \ldots, X_n
\]
where the last step holds since \((a + b + c)^2 \leq 4a^2 + 2b^2 + 4c^2\) for any \(a, b, c\). Now we bound each term separately. First, we have
\[
\mathbb{E} \left[ (\bar{y}_m - \mu_P(x^{(m)}))^4 \right] = 4 \sum_{i_1, i_2, i_3, i_4 \text{ s.t. } X_{i_1} = X_{i_2} = X_{i_3} = X_{i_4} = x^{(m)}} \mathbb{E} \left[ \prod_{k=1}^{4} (Y_{i_k} - \mu_P(x^{(m)})) \right] \middle| X_1, \ldots, X_n
\]
\[
= \frac{1}{n_m^4} \left[ n_m \cdot \mathbb{E} \left[ (Y - \mu_P(x^{(m)}))^4 \right] \middle| X = x^{(m)} \right] + 3n_m(n_m - 1) \cdot \mathbb{E} \left[ (Y - \mu_P(x^{(m)}))^2 \right] \middle| X = x^{(m)} \right] \left[ 3n_m + 1 \right]
\]
\[
\leq \frac{1}{n_m^4} \left[ n_m \cdot \sigma_P^4(x^{(m)}) + 3n_m(n_m - 1) \cdot (\sigma_P^2(x^{(m)}))^2 \right]
\]
\[
\leq \frac{1}{n_m^4} \left[ n_m \cdot \frac{1}{4} + 3n_m(n_m - 1) \cdot (\frac{1}{4})^2 \right] = \frac{3n_m + 1}{16n_m^3},
\]
where the second step holds by counting tuples \((i_1, i_2, i_3, i_4)\) where either all four indices are equal, or there are two pairs of equal indices (since otherwise, the expected value of the product is zero). Next,
\[
\mathbb{E} \left[ (2(\bar{y}_m - \mu_P(x^{(m)}))(\mu_P(x^{(m)}) - \mu(x^{(m)})) \right] \middle| X_1, \ldots, X_n
\]
\[
= 4(\mu_P(x^{(m)}) - \mu(x^{(m)}))^2 \mathbb{E} \left[ (\bar{y}_m - \mu_P(x^{(m)}))^2 \right] \middle| X_1, \ldots, X_n
\]
\[
= 4(\mu_P(x^{(m)}) - \mu(x^{(m)}))^2 \cdot n_m^{-1}(\sigma_P^2(x^{(m)})) \leq n_m^{-1}(\mu_P(x^{(m)}) - \mu(x^{(m)}))^2.
\]
Finally, since $s_m^2 \leq \frac{n_m}{4(n_m - 1)}$ holds deterministically,

$$
\mathbb{E} \left[ (n_m s_m^{-1})^2 \mid X_1, \ldots, X_n \right] \leq n_m^{-2} \cdot \frac{n_m}{4(n_m - 1)} \cdot \mathbb{E} \left[ s_m^2 \mid X_1, \ldots, X_n \right] = n_m^{-2} \cdot \frac{n_m}{4(n_m - 1)} \cdot \sigma_m^2(x^{(m)}) \leq \frac{1}{16n_m(n_m - 1)}.
$$

Combining everything, then,

$$
\text{Var} \left( (\bar{y}_m - \mu(x^{(m)}))^2 - n_m s_m \right) \mid X_1, \ldots, X_n \right) \leq 4 \cdot \frac{3n_m + 1}{16n_m^3} + 2 \cdot n_m^{-1} (\mu_P(x^{(m)}) - \mu(x^{(m)}))^2 + 4 \cdot \frac{1}{16n_m(n_m - 1)},
$$

and so for $n_m \geq 2$,

$$
\text{Var} (Z_m \mid X_1, \ldots, X_n) \leq \left( n_m - 1 \right)^2 \left[ 4 \cdot \frac{3n_m + 1}{16n_m^3} + 2 \cdot n_m^{-1} (\mu_P(x^{(m)}) - \mu(x^{(m)}))^2 + 4 \cdot \frac{1}{16n_m(n_m - 1)} \right] \leq 1 + 2(n_m - 1) \cdot (\mu_P(x^{(m)}) - \mu(x^{(m)}))^2 = 0.5 + 2\mathbb{E} [Z_m \mid X_1, \ldots, X_n].
$$

If instead $n_m = 0$ or $n_m = 1$ then $Z_m = 0$ by definition, and so $\text{Var} (Z_m \mid X_1, \ldots, X_n) = 0$. Therefore, in all cases, we have

$$
\text{Var} (Z_m \mid X_1, \ldots, X_n) \leq \mathbb{I} \{n_m \geq 2\} + 2\mathbb{E} [Z_m \mid X_1, \ldots, X_n].
$$

It is also clear that, conditional on $X_1, \ldots, X_n$, the $Z_m$’s are independent, and so

$$
\text{Var} (Z \mid X_1, \ldots, X_n) = \sum_{m=1}^{\infty} \text{Var} (Z_m \mid X_1, \ldots, X_n) \leq N_{\geq 2} + 2\mathbb{E} [Z \mid X_1, \ldots, X_n].
$$

Finally, we need to bound $\text{Var} (\mathbb{E} [Z \mid X_1, \ldots, X_n])$. First, we have

$$
\text{Var} (\mathbb{E} [Z_m \mid X_1, \ldots, X_n]) = \text{Var} ((n_m - 1)_+ \cdot (\mu_P(x^{(m)}) - \mu(x^{(m)}))^4 \leq \text{Var} ((n_m - 1)_+ \cdot (\mu_P(x^{(m)}) - \mu(x^{(m)}))^2),
$$

and we can calculate

$$
\text{Var} ((n_m - 1)_+) = \text{Var} (n_m + \mathbb{I} \{n_m = 0\}) = \text{Var} (n_m) + \text{Var} (\mathbb{I} \{n_m = 0\}) + 2\text{Cov} (n_m, \mathbb{I} \{n_m = 0\}) = \text{Var} (n_m) + \text{Var} (\mathbb{I} \{n_m = 0\}) - 2\mathbb{E} [n_m] \mathbb{E} [\mathbb{I} \{n_m = 0\}] \text{ since } n_m \cdot \mathbb{I} \{n_m = 0\} = 0 \text{ almost surely}
$$

$$= np_m(1 - p_m) + (1 - p_m)^n(1 - (1 - p_m)^n) - 2np_m(1 - p_m)^n.
$$

Therefore,

$$
2\mathbb{E} [(n_m - 1)_+] - \text{Var} ((n_m - 1)_+) = 2np_m - 2(1 - p_m)^n - np_m(1 - p_m) - (1 - p_m)^n(1 - (1 - p_m)^n) + 2np_m (1 - p_m)^n
$$

$$= np_m(1 + p_m) + (1 - p_m)^n(1 + 2np_m + (1 - p_m)^n) - 2
$$

$$\geq 0,$$

where the last step holds since, defining $f(t) = nt(1 + t) + (1 - t)^n(1 + 2nt + (1 - t)^n)$, we can see that $f(0) = 2$ and $f'(t) \geq 0$ for all $t \in [0, 1]$. This verifies that

$$
\text{Var} (\mathbb{E} [Z_m \mid X_1, \ldots, X_n]) \leq \text{Var} ((n_m - 1)_+ \cdot (\mu_P(x^{(m)}) - \mu(x^{(m)}))^2
$$

$$\leq 2\mathbb{E} [(n_m - 1)_+] \cdot (\mu_P(x^{(m)}) - \mu(x^{(m)}))^2 = 2\mathbb{E} [Z_m].
$$
Next, for any $m \neq m'$,  
\[
\text{Cov}(\mathbb{E}[Z_m | X_1, \ldots, X_n], \mathbb{E}[Z_{m'} | X_1, \ldots, X_n]) = \text{Cov}((n_m - 1)_+, (n_{m'} - 1)_+) \cdot (\mu_p(x^{(m)}) - \mu(x^{(m)}))^2 \cdot (\mu_p(x^{(m')}) - \mu(x^{(m')}))^2 \leq 0.
\]
For the last step, we use the fact that $\text{Cov}((n_m - 1)_+, (n_{m'} - 1)_+) \leq 0$, which holds since, conditional on $n_m$, we have $n_{m'} \sim \text{Binom}\left(n - n_m, \frac{p_m}{1 - p_m}\right)$, and so the distribution of $n_{m'}$ is stochastically smaller whenever $n_m$ is larger. Therefore, 
\[
\text{Var}(\mathbb{E}[Z | X_1, \ldots, X_n]) \leq \sum_{m=1}^{\infty} \text{Var}(\mathbb{E}[Z_m | X_1, \ldots, X_n]) \leq \sum_{m=1}^{\infty} 2\mathbb{E}[Z_m] = 2\mathbb{E}[Z].
\]

### E.4 Proofs of Lemma C.2 and Lemma D.1

Replacing $p$ with $1 - s$, equivalently, we need to show that, for all $s \in [0, 1]$,  
\[
\frac{n(1 - s) - 1 + s^n}{2 + n(1 - s)} \leq n(1 - s) - 1 + s^n \leq \frac{n^2(1 - s)^2}{1 + n(1 - s)}.
\]

After simplifying, this is equivalent to proving that  
\[
\frac{n(1 - s)^2 + 2n(1 - s)}{2 + n(1 - s)} \geq 1 - s^n \geq \frac{n(1 - s)}{1 + n(1 - s)},
\]

which we can further simplify to  
\[
\frac{n(1 - s) + 2n}{2 + n(1 - s)} \geq 1 + s + \cdots + s^{n-1} \geq \frac{n}{1 + n(1 - s)}\tag{E.2}
\]

by dividing by $1 - s$ (note that this division can be performed whenever $s < 1$, while if $s = 1$, then the desired inequalities hold trivially).

Now we address the two desired inequalities separately. For the left-hand inequality in (E.2), define  
\[
h(s) = (2 + n(1 - s)) \cdot (s + s^2 + \cdots + s^{n-1}) = ns + 2(s + s^2 + \cdots + s^{n-1}) - ns^n.
\]

We calculate $h(1) = 2(n - 1)$, and for any $s \in [0, 1]$,  
\[
h'(s) = n + \sum_{i=1}^{n-1} 2is^{i-1} - n^2s^{n-1} \geq n + \sum_{i=1}^{n-1} 2is^{n-1} - n^2s^{n-1} = n + s^{n-1}\left(\sum_{i=1}^{n-1} 2i - n^2\right) = n - ns^{n-1} \geq 0,
\]

where the first inequality holds since $s^{i-1} \geq s^{n-1}$ for all $i = 1, \ldots, n-1$, and the second inequality holds since $s^{n-1} \leq 1$. Therefore, $h(s) \leq h(1) = 2(n - 1)$ for all $s \in [0, 1]$, and so  
\[
1 + s + \cdots + s^{n-1} = \frac{(1 + s + \cdots + s^{n-1}) \cdot (2 + n(1 - s))}{2 + n(1 - s)} \leq \frac{2 + n(1 - s) + 2(n - 1)}{2 + n(1 - s)} = \frac{n(1 - s) + 2n}{2 + n(1 - s)},
\]
as desired.

To verify the right-hand inequality in (E.2), we have  
\[
1 + s + \cdots + s^{n-1} = \frac{(1 + s + \cdots + s^{n-1}) \cdot (1 + n(1 - s))}{1 + n(1 - s)} = \frac{(n + 1)(1 + s + \cdots + s^{n-1}) - n(s + s^2 + \cdots + s^n)}{1 + n(1 - s)} = \frac{n + (1 + s + \cdots + s^{n-1}) - ns^n}{1 + n(1 - s)} \geq \frac{n}{1 + n(1 - s)},
\]

where the last step holds since, for $s \in [0, 1]$, we have $s^i \geq s^n$ for all $i = 0, 1, \ldots, n - 1.$
References

Siu-On Chan, Ilias Diakonikolas, Paul Valiant, and Gregory Valiant. Optimal algorithms for testing closeness of discrete distributions. In Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms, pages 1193–1203. SIAM, 2014.