Contents

A Proofs

The section collects the proofs for the technical results in the main text. Section A.1 and A.2 are devoted to the proof of Lemma 3.1 and Theorem 3.2, respectively. In the end, Section A.3 provides proofs for the coverage gap bounds in the two examples in Section 3.3.

A.1 Proof of Lemma 3.1

Fix a set of locations $\mathcal{B} := \{(i_1, j_1), \dots, (i_{n_{\text{cal}}+1}, j_{n_{\text{cal}}+1})\}$, and an index $1 \le m \le n_{\text{cal}} + 1$. By the definition of conditional probability, we have

$$
\begin{aligned}
&\mathbb{P}\left((i_*, j_*) = (i_m, j_m) \mid \mathcal{S}_{\text{cal}} \cup \{(i_*, j_*)\} = \mathcal{B},\ \mathcal{S}_{\text{tr}} = \mathcal{S}_0\right) \\
=&\frac{\mathbb{P}\left((i_*, j_*) = (i_m, j_m), \mathcal{S}_{\text{cal}} = \mathcal{B}\backslash\{(i_m, j_m)\} \mid \mathcal{S}_{\text{tr}} = \mathcal{S}_0,\ |\mathcal{S}| = n\right)}{\mathbb{P}\left(\mathcal{S}_{\text{cal}} \cup \{(i_*, j_*)\} = \mathcal{B} \mid \mathcal{S}_{\text{tr}} = \mathcal{S}_0,\ |\mathcal{S}| = n\right)} \\
=&\frac{\mathbb{P}\left((i_*, j_*) = (i_m, j_m), \mathcal{S}_{\text{cal}} = \mathcal{B}\backslash\{(i_m, j_m)\} \mid \mathcal{S}_{\text{tr}} = \mathcal{S}_0,\ |\mathcal{S}| = n\right)}{\sum_{l=1}^{n_{\text{cal}}+1} \mathbb{P}\left((i_*, j_*) = (i_l, j_l), \mathcal{S}_{\text{cal}} = \mathcal{B}\backslash\{(i_l, j_l)\} \mid \mathcal{S}_{\text{tr}} = \mathcal{S}_0,\ |\mathcal{S}| = n\right)}.
\end{aligned}
$$

It then boils down to computing $\mathbb{P}\left(\mathcal{S}_{\text{cal}} = \mathcal{S}_1, (i_*, j_*) = (i_l, j_l) \mid \mathcal{S}_{\text{tr}} = \mathcal{S}_0, |\mathcal{S}| = n\right)$ for any fixed set $\mathcal{S}_1$ and fixed location $(i_l, j_l) \notin \mathcal{S}_0 \cup \mathcal{S}_1$. To this end, we have the following claim (we defer its proof to the end of this section)

$$
\begin{aligned}
&\mathbb{P}\left(\mathcal{S}_{\text{cal}} = \mathcal{S}_1, (i_*, j_*) = (i_l, j_l) \mid \mathcal{S}_{\text{tr}} = \mathcal{S}_0, |\mathcal{S}| = n\right) \\
=\ &\frac{1}{d_1 d_2 - n} \cdot \frac{\prod_{(i,j)\in\mathcal{S}_1} \frac{p_{ij}}{1-p_{ij}}}{\sum_{\mathcal{A}\in\Omega_{\mathcal{S}_{\text{tr}},n}} \prod_{(i',j')\in\mathcal{A}} \frac{p_{i'j'}}{1-p_{i'j'}}},
\end{aligned}
\tag{8}
$$

where $\Omega_{\mathcal{S}_{\rm tr},n} = \{\mathcal{A} \subseteq [d_1]\times[d_2] : |\mathcal{A}| = n - |\mathcal{S}_{\rm tr}|,\ \mathcal{A}\cap\mathcal{S}_{\rm tr} = \varnothing\}$. As a result, we obtain

$$\begin{aligned}
&\mathbb{P}\left((i_*, j_*) = (i_m, j_m) \mid \mathcal{S}_{\rm cal}\cup\{(i_*, j_*)\} = \mathcal{B},\ \mathcal{S}_{\rm tr} = \mathcal{S}_0,\ |\mathcal{S}| = n\right) \\
=&\frac{\prod_{(i,j)\in\mathcal{B}\setminus\{(i_m,j_m)\}} \frac{p_{ij}}{1-p_{ij}}}{\sum_{l=1}^{n_{\rm cal}+1}\prod_{(i,j)\in\mathcal{B}\setminus\{(i_l,j_l)\}} \frac{p_{ij}}{1-p_{ij}}} \\
=&\frac{h_{i_m,j_m}}{\sum_{l=1}^{n_{\rm cal}+1} h_{i_l,j_l}},
\end{aligned} \tag{9}$$

where $h_{ij} = (1-p_{ij})/p_{ij}$. This finishes the proof.

Proof of Equation (8). Fix any two disjoint subsets $\mathcal{S}_0, \mathcal{S}_1 \subseteq [d_1]\times[d_2]$ with $|\mathcal{S}_0| + |\mathcal{S}_1| = n$. We have

$$\begin{aligned}
&\mathbb{P}\left(\mathcal{S}_{\rm cal} = \mathcal{S}_1, \mathcal{S}_{\rm tr} = \mathcal{S}_0 \mid |\mathcal{S}| = n\right) \\
=&\mathbb{P}\left(\mathrm{supp}(\mathbf{Z}) = \mathcal{S}_0\cup\mathcal{S}_1, \mathcal{S}_0 \subseteq \mathrm{supp}(\mathbf{W}), \mathcal{S}_1 \subseteq \mathrm{supp}(\mathbf{W})^c \mid |\mathcal{S}| = n\right).
\end{aligned}$$

Since $\mathbf{W}$ and $\mathbf{Z}$ are independent, one further has

$$\begin{aligned}
&\mathbb{P}\left(\mathrm{supp}(\mathbf{Z}) = \mathcal{S}_0\cup\mathcal{S}_1, \mathcal{S}_0 \subseteq \mathrm{supp}(\mathbf{W}), \mathcal{S}_1 \subseteq \mathrm{supp}(\mathbf{W})^c \mid |\mathcal{S}| = n\right) \\
=&\mathbb{P}\left(\mathcal{S}_0 \subseteq \mathrm{supp}(\mathbf{W}), \mathcal{S}_1 \subseteq \mathrm{supp}(\mathbf{W})^c\right)\cdot\mathbb{P}\left(\mathrm{supp}(\mathbf{Z}) = \mathcal{S}_0\cup\mathcal{S}_1 \mid |\mathcal{S}| = n\right) \\
=&q^{|\mathcal{S}_0|}(1-q)^{|\mathcal{S}_1|}\cdot\mathbb{P}\left(\mathrm{supp}(\mathbf{Z}) = \mathcal{S}_0\cup\mathcal{S}_1 \mid |\mathcal{S}| = n\right) \\
=&\frac{q^{|\mathcal{S}_0|}(1-q)^{|\mathcal{S}_1|}}{\mathbb{P}(|\mathcal{S}| = n)}\prod_{(i',j')\in[d_1]\times[d_2]}(1-p_{i'j'})\prod_{(i,j)\in\mathcal{S}_0\cup\mathcal{S}_1}\frac{p_{ij}}{1-p_{ij}},
\end{aligned} \tag{10}$$

where the last two identities are based on direct computations.

Based on (10), one can further compute

$$\begin{aligned}
&\mathbb{P}\left(\mathcal{S}_{\rm tr} = \mathcal{S}_0 \mid |\mathcal{S}| = n\right) \\
=&\sum_{\mathcal{A}\in\Omega_{\mathcal{S}_{\rm tr},n}}\mathbb{P}\left(\mathcal{S}_{\rm cal} = \mathcal{A}, \mathcal{S}_{\rm tr} = \mathcal{S}_0 \mid |\mathcal{S}| = n\right) \\
=&\frac{q^{|\mathcal{S}_0|}(1-q)^{n-|\mathcal{S}_0|}}{\mathbb{P}(|\mathcal{S}| = n)}\sum_{\mathcal{A}\in\Omega_{\mathcal{S}_{\rm tr},n}}\prod_{(i',j')\in[d_1]\times[d_2]}(1-p_{i'j'})\prod_{(i,j)\in\mathcal{S}_0\cup\mathcal{A}}\frac{p_{ij}}{1-p_{ij}},
\end{aligned} \tag{11}$$

where the last identity uses Equation (10).

Now we are ready to prove (8). Recall that the new data point $(i_*, j_*) \mid \mathcal{S}$ is drawn uniformly from $\mathrm{Unif}(\mathcal{S}^c)$. Therefore one has

$$\begin{aligned}
&\mathbb{P}\left(\mathcal{S}_{\rm cal} = \mathcal{S}_1, (i_*, j_*) = (i_{n+1}, j_{n+1}) \mid \mathcal{S}_{\rm tr} = \mathcal{S}_0, |\mathcal{S}| = n\right) \\
=&\frac{\mathbb{P}\left(\mathcal{S}_{\rm cal} = \mathcal{S}_1, \mathcal{S}_{\rm tr} = \mathcal{S}_0, (i_*, j_*) = (i_{n+1}, j_{n+1}) \mid |\mathcal{S}| = n\right)}{\mathbb{P}\left(\mathcal{S}_{\rm tr} = \mathcal{S}_0 \mid |\mathcal{S}| = n\right)} \\
=&\frac{\mathbb{P}\left((i_*, j_*) = (i_{n+1}, j_{n+1}) \mid \mathcal{S} = \mathcal{S}_1\cup\mathcal{S}_0\right)\mathbb{P}\left(\mathcal{S}_{\rm cal} = \mathcal{S}_1, \mathcal{S}_{\rm tr} = \mathcal{S}_0 \mid |\mathcal{S}| = n\right)}{\mathbb{P}\left(\mathcal{S}_{\rm tr} = \mathcal{S}_0 \mid |\mathcal{S}| = n\right)} \\
=&\frac{1}{d_1 d_2 - n}\cdot\frac{\prod_{(i,j)\in\mathcal{S}_1}\frac{p_{ij}}{1-p_{ij}}}{\sum_{\mathcal{A}\in\Omega_{\mathcal{S}_{\rm tr},n}}\prod_{(i',j')\in\mathcal{A}}\frac{p_{i'j'}}{1-p_{i'j'}}},
\end{aligned} \tag{12}$$

where the last line follows from (10) and (11).

A.2 Proof of Theorem 3.2

First, fix any $a \in [0,1]$. Lemma 3.1, together with the weighted conformal prediction framework of Tibshirani et al. (2019), implies that

$$\mathbb{P}\left(M_{i_*j_*} \in \widehat{M}_{i_*j_*} \pm q^*_{i_*j_*}(a)\cdot\widehat{s}_{i_*j_*} \,\middle|\, \mathcal{S}_{\rm tr}, \mathcal{S}_{\rm cal}\cup\{(i_*, j_*)\}\right) \geq 1 - a,$$

where

$$q^*_{i_*j_*}(a) = \text{Quantile}_{1-a}\left(\sum_{(i,j)\in\mathcal{S}_{\mathrm{cal}}\cup\{(i_*,j_*)\}} w_{ij}\cdot\delta_{R_{ij}}\right),$$

and

$$w_{ij} = \frac{h_{ij}}{\sum_{(i',j')\in\mathcal{S}_{\mathrm{cal}}\cup\{(i_*,j_*)\}} h_{i'j'}}.$$

Indeed, here a can be any function of the random variables we are conditioning on—that is, a may depend on n, on $\mathcal{S}_{\mathrm{tr}}$, and on $\mathcal{S}_{\mathrm{cal}}\cup\{(i_*,j_*)\}$.

Next define $a = \alpha + \Delta$ where Δ is defined as in (6). We observe that, since each $\widehat{h}_{ij}$ is a function of $\mathcal{S}_{\mathrm{tr}}$, then Δ (and thus also a) can therefore be expressed as a function of n, $\mathcal{S}_{\mathrm{tr}}$, and $\mathcal{S}_{\mathrm{cal}}\cup\{(i_*,j_*)\}$. Therefore, applying the work above, we have

$$\mathbb{P}\left(M_{i_*j_*}\in\widehat{M}_{i_*j_*}\pm q^*_{i_*j_*}(\alpha+\Delta)\cdot\widehat{s}_{i_*j_*}\;\middle|\;\mathcal{S}_{\mathrm{tr}},\mathcal{S}_{\mathrm{cal}}\cup\{(i_*,j_*)\}\right)\geq 1-\alpha-\Delta$$

and thus, after marginalizing,

$$\mathbb{P}\left(M_{i_*j_*}\in\widehat{M}_{i_*j_*}\pm q^*_{i_*j_*}(\alpha+\Delta)\cdot\widehat{s}_{i_*j_*}\right)\geq 1-\alpha-\mathbb{E}[\Delta].$$

Next, we verify that

$$\widehat{q}\geq q^*_{i_*j_*}(\alpha+\Delta)$$

holds almost surely—if this is indeed the case, then we have shown that

$$\mathbb{P}\left(M_{i_*j_*}\in\widehat{M}_{i_*j_*}\pm\widehat{q}\cdot\widehat{s}_{i_*j_*}\right)\geq 1-\alpha-\mathbb{E}[\Delta],$$

which establishes the desired result. Thus we only need to show that $\widehat{q}\geq q^*_{i_*j_*}(\alpha+\Delta)$, or equivalently,

$$\text{Quantile}_{1-\alpha}\left(\sum_{(i,j)\in\mathcal{S}_{\mathrm{cal}}}\widehat{w}_{ij}\cdot\delta_{R_{ij}}+\widehat{w}_{\text{test}}\cdot\delta_{+\infty}\right)\geq\text{Quantile}_{1-\alpha-\Delta}\left(\sum_{(i,j)\in\mathcal{S}_{\mathrm{cal}}\cup\{(i_*,j_*)\}}w_{ij}\cdot\delta_{R_{ij}}\right).$$

Define

$$w'_{ij}=\frac{\widehat{h}_{ij}}{\sum_{(i',j')\in\mathcal{S}_{\mathrm{cal}}\cup\{(i_*,j_*)\}}\widehat{h}_{i'j'}} \tag{13}$$

for all $(i,j)\in\mathcal{S}\cup\{(i_*,j_*)\}$. Then by definition of $\widehat{\boldsymbol{w}}$, we see that $w'_{ij}\geq\widehat{w}_{ij}$ for $(i,j)\in\mathcal{S}$, and therefore,

$$\begin{aligned}&\text{Quantile}_{1-\alpha}\left(\sum_{(i,j)\in\mathcal{S}_{\mathrm{cal}}}\widehat{w}_{ij}\cdot\delta_{R_{ij}}+\widehat{w}_{\text{test}}\cdot\delta_{+\infty}\right)\\ &\geq\text{Quantile}_{1-\alpha}\left(\sum_{(i,j)\in\mathcal{S}_{\mathrm{cal}}}w'_{ij}\cdot\delta_{R_{ij}}+w'_{(i_*,j_*)}\cdot\delta_{+\infty}\right)\geq\text{Quantile}_{1-\alpha}\left(\sum_{(i,j)\in\mathcal{S}_{\mathrm{cal}}\cup\{(i_*,j_*)\}}w'_{ij}\cdot\delta_{R_{ij}}\right)\end{aligned}$$

holds almost surely. Therefore it suffices to show that

$$\text{Quantile}_{1-\alpha}\left(\sum_{(i,j)\in\mathcal{S}_{\mathrm{cal}}\cup\{(i_*,j_*)\}}w'_{ij}\cdot\delta_{R_{ij}}\right)\geq\text{Quantile}_{1-\alpha-\Delta}\left(\sum_{(i,j)\in\mathcal{S}_{\mathrm{cal}}\cup\{(i_*,j_*)\}}w_{ij}\cdot\delta_{R_{ij}}\right)$$

holds almost surely. Indeed, we have

$$\mathsf{d}_{\mathsf{TV}}\left(\sum_{(i,j)\in\mathcal{S}_{\mathrm{cal}}\cup\{(i_*,j_*)\}}w'_{ij}\cdot\delta_{R_{ij}},\sum_{(i,j)\in\mathcal{S}_{\mathrm{cal}}\cup\{(i_*,j_*)\}}w_{ij}\cdot\delta_{R_{ij}}\right)\leq\frac{1}{2}\sum_{(i,j)\in\mathcal{S}_{\mathrm{cal}}\cup\{(i_*,j_*)\}}|w'_{ij}-w_{ij}|=\Delta,$$

where d_{TV} denotes the total variation distance. This completes the proof.

A.3 Proofs for examples in Section 3.3

Recall the definition of Δ:

$$\Delta = \frac{1}{2} \sum_{(i,j)\in\mathcal{S}_{\mathrm{cal}}\cup\{(i_*,j_*)\}} \left| \frac{\widehat{h}_{ij}}{\sum_{(i',j')\in\mathcal{S}_{\mathrm{cal}}\cup\{(i_*,j_*)\}} \widehat{h}_{i'j'}} - \frac{h_{ij}}{\sum_{(i',j')\in\mathcal{S}_{\mathrm{cal}}\cup\{(i_*,j_*)\}} h_{i'j'}} \right|.$$

Note that $\Delta \le 1$ by definition. We start with stating a useful lemma to bound the coverage gap Δ using the estimation error of $\widehat{h}_{ij}$.

Lemma A.1. *By the definition of the coverage gap Δ, we have*

$$\Delta \le \frac{\sum_{(i,j)\in\mathcal{S}_{\mathrm{cal}}\cup\{(i_*,j_*)\}} |\widehat{h}_{ij} - h_{ij}|}{\sum_{(i',j')\in\mathcal{S}_{\mathrm{cal}}} \widehat{h}_{i'j'}}. \tag{14}$$

A.3.1 Proof for Example 3.3.1

Under the logistic model, one has for any (i,j)

$$\mathbb{P}((i,j)\in\mathcal{S}_{\mathrm{tr}}) = q\cdot p_{ij} = \frac{q\cdot\exp(u_i+v_j)}{1+\exp(u_i+v_j)}.$$

In this case, if $\widehat{\boldsymbol{u}}$ and $\widehat{\boldsymbol{v}}$ are the constrained maximum likelihood estimators in Example 3.3.1, Theorem 6 in Chen et al. (2023) implies that

$$\|\widehat{\mathbf{u}}-\mathbf{u}\|_\infty = O_{\mathbb{P}}\left(\sqrt{\frac{\log d_1}{d_2}}\right), \qquad \|\widehat{\mathbf{v}}-\mathbf{v}\|_\infty = O_{\mathbb{P}}\left(\sqrt{\frac{\log d_2}{d_1}}\right),$$

with the proviso that $\|\boldsymbol{u}\|_\infty + \|\boldsymbol{v}\|_\infty \le \tau < \infty$, $d_2 \gg \sqrt{d_1}\log d_1$ and $d_1 \gg (\log d_2)^2$. Then for $h_{ij} = \exp(-u_i - v_j)$ and $\widehat{h}_{ij} = \exp(-\widehat{u}_i - \widehat{v}_j)$, we have

$$\max_{i,j} |\widehat{h}_{ij} - h_{ij}| = \max_{i,j} e^{-u_i-v_j}\left(e^{-(\widehat{u}_i-u_i)-(\widehat{v}_j-v_j)} - 1\right) = O_{\mathbb{P}}\left(\sqrt{\frac{\log d_1}{d_2}} + \sqrt{\frac{\log d_2}{d_1}}\right).$$

Further, as $\min_{i,j} h_{i,j} = \min_{i,j}\exp(-u_i-v_j) \ge e^{-\tau}$, then with probability approaching one, for every $(i,j)\in[d_1]\times[d_2]$, one has $\widehat{h}_{i,j} \ge h_{i,j} - |h_{ij} - \widehat{h}_{ij}| \ge e^{-\tau}/2 =: h_0$. By the upper bound (14), we have

$$\Delta \lesssim \sqrt{\frac{\log d_1}{d_2}} + \sqrt{\frac{\log d_2}{d_1}} \asymp \sqrt{\frac{\log\max\{d_1,d_2\}}{\min\{d_1,d_2\}}}.$$

Further, as $\Delta \le 1$, we have $\mathbb{E}[\Delta] \lesssim \sqrt{\frac{\log\max\{d_1,d_2\}}{\min\{d_1,d_2\}}}$.

A.3.2 Proof of Example 3.3.2

Define the link function $\psi(t) = q(1+e^{-\phi(t)})$, where ϕ is monotonic. Applying Theorem 1 in the paper (Davenport et al., 2014), we obtain that with probability at least $1 - C_1/(d_1+d_2)$

$$\frac{1}{d_1 d_2}\|\widehat{\mathbf{A}} - \mathbf{A}\|_{\mathrm{F}}^2 \le \sqrt{2}\widetilde{C}_\tau\sqrt{\frac{k(d_1+d_2)}{d_1 d_2}}, \tag{15}$$

with the proviso that $d_1 d_2 \ge (d_1+d_2)\log(d_1 d_2)$. Here $\widetilde{C}_\tau = 2^{39/4}e^{9/4}(1+\sqrt{6})\tau L_\tau \beta_\tau$ with

$$L_\tau = \sup_{-\tau\le t\le\tau} \frac{|\psi'(t)|}{\psi(t)(1-\psi(t))}, \quad \text{and} \quad \beta_\tau = \sup_{-\tau\le t\le\tau} \frac{\psi(t)(1-\psi(t))}{|\psi'(t)|^2}. \tag{16}$$

Denote this high probability event to be $\mathcal{E}_0$. Since $\|\mathbf{A}\|_1 \le \sqrt{d_1 d_2}\|\mathbf{A}\|_{\mathrm{F}}$, on this event $\mathcal{E}_0$, we further have

$$\frac{1}{d_1 d_2}\|\widehat{\mathbf{A}} - \mathbf{A}\|_1 \le C_\tau\left(\frac{k(d_1+d_2)}{d_1 d_2}\right)^{1/4},$$

where $C_\tau = \zeta\sqrt{L_\tau \beta_\tau}$ and ζ is a universal constant.

Recall that $h_{ij} = \exp(-\phi(A_{ij}))$ and $\|\mathbf{A}\|_\infty \le \tau$. On the same event $\mathcal{E}_0$, we further have

$$\frac{1}{d_1 d_2}\|\widehat{\mathbf{H}} - \mathbf{H}\|_1 \le C'_\tau \left(\frac{k(d_1+d_2)}{d_1 d_2}\right)^{1/4}, \tag{17}$$

where $C'_\tau = 2e^{\phi(\tau)}\sqrt{\widetilde{C}_\tau}$.

By the feasibility of the minimizer $\widehat{\mathbf{A}}$ and the fact that $\widehat{h}_{ij} = \exp(-\phi(\widehat{A}_{ij}))$, we have $\widehat{h}_{ij} \ge h_0 = e^{-\phi(\tau)}$ for all (i,j). This together with the upper bound (14) implies that

$$\Delta \le \frac{1}{h_0 n_{\text{cal}}} \sum_{(i,j)\in\mathcal{S}_{\text{cal}}\cup\{(i_*,j_*)\}} |h_{ij} - \widehat{h}_{ij}| \le \frac{1}{h_0 n_{\text{cal}}}\|\widehat{\mathbf{H}} - \mathbf{H}\|_1.$$

Define a second high probability event $\mathcal{E}_1 = \{n_{\text{cal}} \ge (1-c)(1-q)\|\mathbf{P}\|_1\}$. Using the Chernoff bound, we have $\mathbb{P}(\mathcal{E}_1) \ge 1 - C_0(d_1d_2)^{-5}$. Therefore, on the event $\mathcal{E}_0 \cap \mathcal{E}_1$, we have

$$\Delta \le \frac{d_1 d_2}{ch_0(1-q)\|\mathbf{P}\|_1} C'_\tau \left(\frac{k(d_1+d_2)}{d_1 d_2}\right)^{1/4}.$$

Under the assumptions that $p_{ij} = 1/(1+e^{-\phi(A_{ij})})$, and that $\|\mathbf{A}\|_\infty \le \tau$, we know that $p_{ij} \ge C_2$ for some constant that only depends on τ. As a result, we have $\|\mathbf{P}\|_1 \ge C_2 d_1 d_2$, which further leads to the conclusion that

$$\Delta \le \frac{1}{cC_2 h_0(1-q)} C'_\tau \left(\frac{k(d_1+d_2)}{d_1 d_2}\right)^{1/4}.$$

on the event $\mathcal{E}_0 \cap \mathcal{E}_1$. In addition, on the small probability event $(\mathcal{E}_0 \cap \mathcal{E}_1)$, one trivially has $\Delta \le 1$. Therefore simple combinations of the cases yields the desired bound on $\mathbb{E}[\Delta]$.

A.3.3 Proof of Lemma A.1

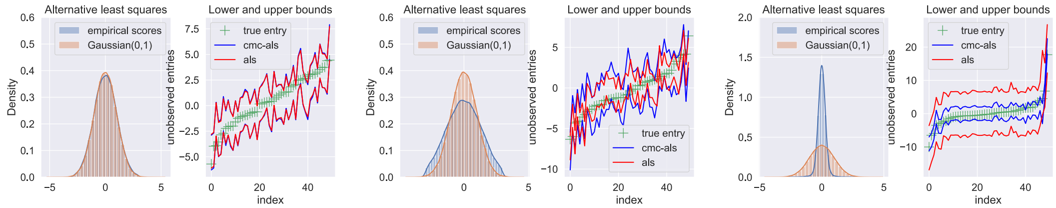
Reusing the definition of w' (13), one has

$$\begin{aligned}
\Delta = &\ \frac{1}{2}\sum_{(i,j)\in\mathcal{S}_{\text{cal}}\cup\{(i_*,j_*)\}} |w_{ij} - w'_{ij}| \\
= &\ \frac{1}{2}\sum_{(i,j)\in\mathcal{S}_{\text{cal}}\cup\{(i_*,j_*)\}} \frac{\left|h_{ij}\sum_{(i',j')\in\mathcal{S}_{\text{cal}}\cup\{(i_*,j_*)\}}\widehat{h}_{i'j'} - \widehat{h}_{ij}\sum_{(i',j')\in\mathcal{S}_{\text{cal}}\cup\{(i_*,j_*)\}} h_{i'j'}\right|}{\left(\sum_{(i',j')\in\mathcal{S}_{\text{cal}}\cup\{(i_*,j_*)\}}\widehat{h}_{i'j'}\right)\left(\sum_{(i',j')\in\mathcal{S}_{\text{cal}}\cup\{(i_*,j_*)\}} h_{i'j'}\right)} \\
\le &\ \frac{1}{2}\sum_{(i,j)\in\mathcal{S}_{\text{cal}}\cup\{(i_*,j_*)\}} \left\{\frac{h_{ij}\left|\sum_{(i',j')\in\mathcal{S}_{\text{cal}}\cup\{(i_*,j_*)\}}\widehat{h}_{i'j'} - \sum_{(i',j')\in\mathcal{S}_{\text{cal}}\cup\{(i_*,j_*)\}} h_{i'j'}\right|}{\left(\sum_{(i',j')\in\mathcal{S}_{\text{cal}}\cup\{(i_*,j_*)\}}\widehat{h}_{i'j'}\right)\left(\sum_{(i',j')\in\mathcal{S}_{\text{cal}}\cup\{(i_*,j_*)\}} h_{i'j'}\right)}\right. \\
&\qquad \left. + \frac{|\widehat{h}_{ij} - h_{ij}|\sum_{(i',j')\in\mathcal{S}_{\text{cal}}\cup\{(i_*,j_*)\}} h_{i'j'}}{\left(\sum_{(i',j')\in\mathcal{S}_{\text{cal}}\cup\{(i_*,j_*)\}}\widehat{h}_{i'j'}\right)\left(\sum_{(i',j')\in\mathcal{S}_{\text{cal}}\cup\{(i_*,j_*)\}} h_{i'j'}\right)}\right\} \\
= &\ \frac{1}{2}\frac{\left|\sum_{(i',j')\in\mathcal{S}_{\text{cal}}\cup\{(i_*,j_*)\}}\widehat{h}_{i'j'} - \sum_{(i',j')\in\mathcal{S}_{\text{cal}}\cup\{(i_*,j_*)\}} h_{i'j'}\right|}{\left(\sum_{(i',j')\in\mathcal{S}_{\text{cal}}\cup\{(i_*,j_*)\}}\widehat{h}_{i'j'}\right)} + \frac{1}{2}\frac{\sum_{(i,j)\in\mathcal{S}_{\text{cal}}\cup\{(i_*,j_*)\}}|\widehat{h}_{ij} - h_{ij}|}{\left(\sum_{(i',j')\in\mathcal{S}_{\text{cal}}\cup\{(i_*,j_*)\}}\widehat{h}_{i'j'}\right)} \\
\le &\ \frac{\sum_{(i,j)\in\mathcal{S}_{\text{cal}}\cup\{(i_*,j_*)\}}|\widehat{h}_{ij} - h_{ij}|}{\sum_{(i',j')\in\mathcal{S}_{\text{cal}}\cup\{(i_*,j_*)\}}\widehat{h}_{i'j'}}.
\end{aligned} \tag{18}$$

This completes the proof.

B Additional numerical experiments

In this section, we provide additional simulation results. In Section B.1, the confidence intervals are visualized. We also present the simulation results for the convex relaxation (`cvx`) and the conformalized convex method (`cmc-cvx`) in both settings with homogeneous and heterogeneous missingness in Section B.2 and B.4.

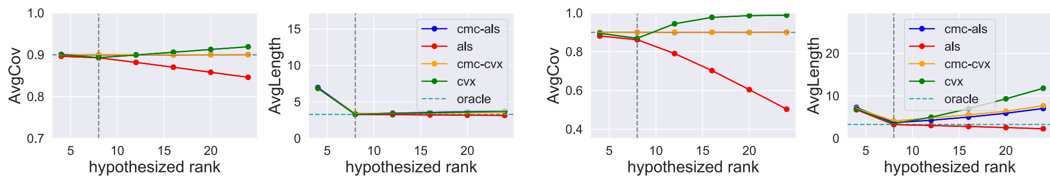


(a) Large sample size + Gaussian noise: $r = r^* = 8$

(b) Large sample size + Gaussian noise: $r^* = 8 < 50 = r$

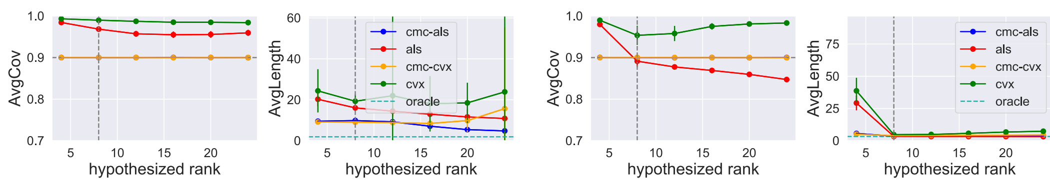
(c) Coherent matrix: $r = 6 < 8 = r^*$

Figure 4: Histogram of standardized scores for `als` and prediction lower and upper bounds for 50 distinct unobserved entries.



(a) Setting 1: Large sample size + Gaussian noise

(b) Setting 2: Small sample size+Gaussian noise



(c) Setting 3: Large sample size + heavy-tailed noise

(d) Setting 4: Violation of incoherence

Figure 5: Comparison between conformalized and model-based matrix completion approaches.

B.1 Illustration of confidence intervals and normalized scores

In Figure 4, we present the histogram of standardized scores $(\widehat{M}_{ij} - M_{ij})/\sqrt{\widehat{\theta}_{ij}^2 + \widehat{\sigma}^2}$ and the plot of the upper and lower bounds for three settings. In Figure 4a, when the model assumptions are met and $r = r^*$, the scores match well with the standard Gaussian and the prediction bounds produced by `als` and `cmc-als` are similar. With the same data generating process, when the rank is overparametrized, the distribution of scores cannot be captured by the standard Gaussian, thus the quantiles are misspecified. As we can see from the confidence intervals, `als` tends to have smaller intervals which lead to the undercoverage. In the last setting, the underlying matrix is no longer incoherent. When the rank is underestimated, the $r^* - r$ factors will be captured by the noise term and the high heterogeneity in the entries will further lead to overestimated noise level. As a result, the intervals by `als` are much larger while the conformalized intervals are more adaptive to the magnitude of entries.

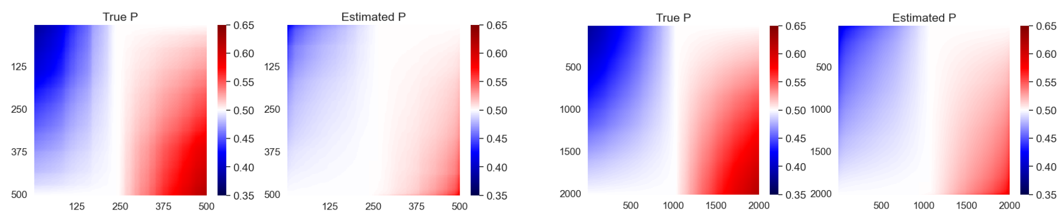
B.2 Additional results for homogeneous misingness

In this section, we present the results for synthetic simulation with the convex relaxation `cvx` and the conformalized convex matrix completion method `cmc-cvx`. Setting 1, 2, 3, 4 are the same as introduced in Section 4.1. The true rank is $r^* = 8$ and the hypothesized rank varies from 4 to 24 with the stepsize 4.

The conformalized methods, regardless of the based algorithm adopted, have nearly exact coverage around $1 - \alpha$. But we can observe different behaviors between `als` and `cvx` since the convex relaxation is free of the choice of r until the projection of $\widehat{\mathbf{M}}_{\texttt{cvx}}$ onto the rank-r subspace (Chen et al., 2020). As a result, when $r > r^*$, `cvx` tends to overestimate the strength of the noise. In Figure 5a, 5b and 5d, when $r > r^*$, `cvx` has coverage rate higher than the target level and the confidence interval is more conservative than conformalized methods. Since the accuracy of `cvx` is based on the large sample size, in Figure 5b, when the effective sample size is insufficient with small p_{ij}, the residuals from `cvx` have a large deviation from the standard distribution and the intervals are much larger than the oracle ones. Besides, in Figure 5c, when the noise has heavy tails than Gaussian variables, `cvx` overestimates the noise strength similar to `als` and is conservative in coverage. When the incoherence condition is violated in Figure 5d, if $r < r^*$, both `cvx` and `als` fit the missed factor by overestimating the noise strength and produce extremely large intervals.

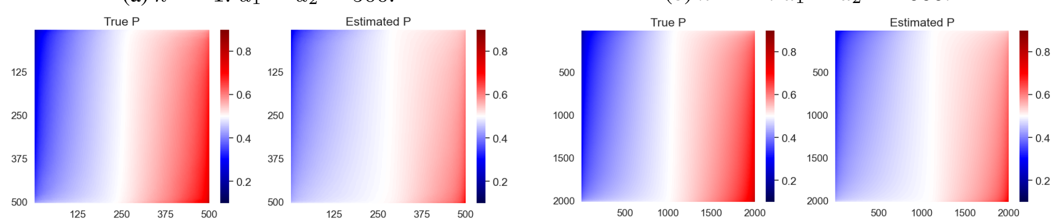
B.3 Estimation error for one-bit matrix estimation

The estimation error in $\widehat{p}_{ij}$ can be visualized from the following heatmaps comparing $\mathbf{P}$ and $\widehat{\mathbf{P}}$. Here the entries are sorted by the order of p_{ij} for each row and each column.



(a) $k^* = 1$: $d_1 = d_2 = 500$.

(b) $k^* = 1$: $d_1 = d_2 = 2000$.



(c) $k^* = 5$: $d_1 = d_2 = 500$.

(d) $k^* = 5$: $d_1 = d_2 = 2000$.

Figure 6: Heatmaps for $\mathbf{P}$ and $\widehat{\mathbf{P}}$.

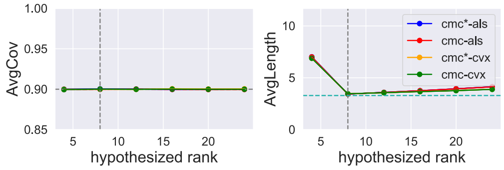
B.4 Additional results for heterogeneous missingness

In Figure 7, we present the results for synthetic simulation with the convex relaxation `cvx` as the base algorithm, where we denote `cmc-cvx` and `cmc`*`-cvx` as the conformalized matrix completion method with estimated weights and true weights, respectively. Three settings with heterogeneous missingness are the same as Figure 2.

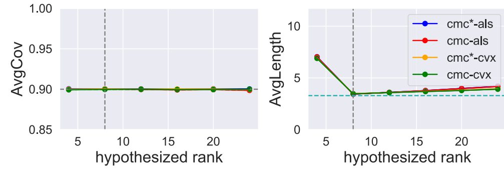
B.5 Additional plots for the sales dataset

Denote $\mathbf{M}$ the underlying matrix in the sales dataset. In Figure 8a, we plot singular values of $\mathbf{M}$ and top-5 singular values contain a large proportion of the information. In Figure 8b, we plot the histogram of entries M_{ij}'s of the underlying matrix, and the sales dataset has the range from 0 to over 20 thousand with a heavy tail.

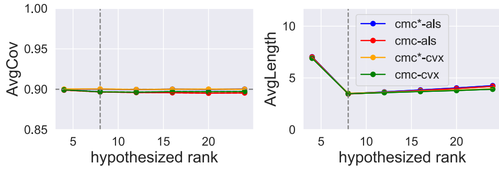
In Figure 9, `cmc-cvx` has nearly exact coverage at $1 - \alpha$, but `cvx` tends to have higher coverage than the target level. Besides, the convex approach has much larger intervals when r is large, which can be caused by the overfitting of the observed entries. As conformalized approach leaves out a proportion of observed entries as the training set, intervals produced by `cmc-cvx` are less accurate than `cvx` due to the poorly behaved base algorithm.



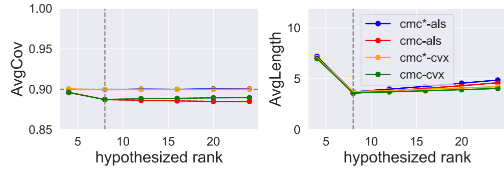
(a) Gaussian noise: $k^* = 1$.



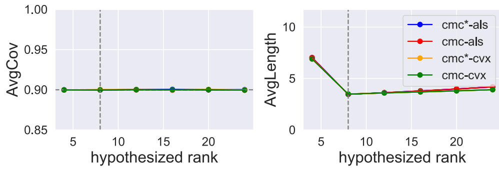
(b) Gaussian noise: $k^* = 5$.



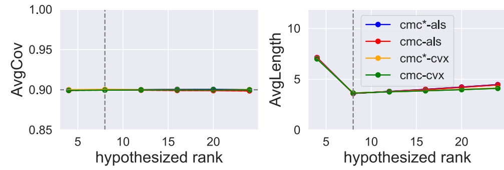
(c) Adversarial heterogeneous noise: $k^* = 1$.



(d) Adversarial heterogeneous noise: $k^* = 5$.

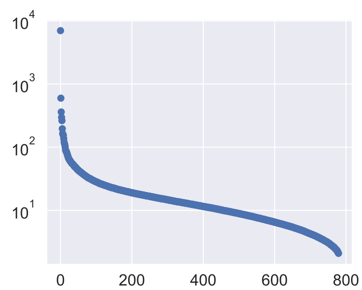


(e) Random heterogeneous noise: $k^* = 1$.

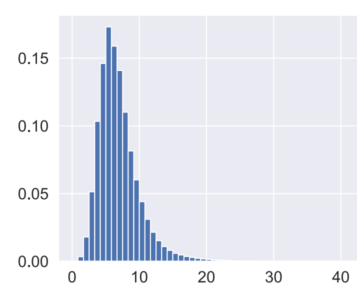


(f) Random heterogeneous noise: $k^* = 5$.

Figure 7: Comparison under heterogenous missingness.



(a) Singular values of the underlying matrix $\mathbf{M}$.



(b) Histogram of entries in $\mathbf{M}$.

Figure 8: Descriptive plots for the underlying matrix $\mathbf{M}$.

C Additional details of algorithms and extensions

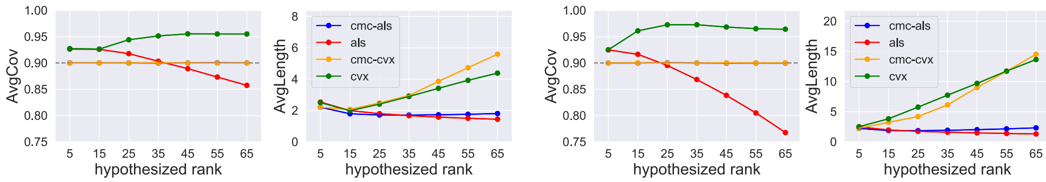
C.1 Extension to likelihood-based scores and categorical matrices

We will show in this section that `cmc` can also be applied to categorical matrix completion (Cao and Xie, 2015) or a more general setting in (19), the validity of which is also guaranteed by the presented theorem.

Setup To formulate the problem, consider an underlying parameter matrix $\mathbf{M}^* \in [d_1] \times [d_2]$ and the observations $\{M_{ij} : \ i \in [d_1], j \in [d_2]\}$ are drawn from the distribution

$$M_{ij} \mid M^*_{ij} \sim \mathcal{P}_{M^*_{ij}}, \tag{19}$$

where $\{\mathcal{P}_\theta\}_{\theta \in \Theta}$ can be a family of parametric distributions with probability density p_θ. The categorical matrix completion is a specific example where the support of M_{ij} is finite or countable. For example, a Poisson matrix is generated by $M_{ij} \sim \mathrm{Pois}(M^*_{ij})$, where M^*_{ij} is the Poisson mean. Similar to the previous setup, we treat $\mathbf{M}$ as deterministic, and entries in the subset $\mathcal{S} \subseteq [d_1] \times [d_2]$ are available. Here $\mathcal{S}$ is sampled in the same manner as before with the matrix $\mathbf{P} \in [d_1] \times [d_2]$.



(a) Homogeneous missingness.

(b) Heterogeneous missingness with $k^* = 1$.

Figure 9: Comparison between conformalized and model-based matrix completion with sales dataset.

Split conformal approach Consider the split approach with the partition $\mathcal{S} = \mathcal{S}_{\text{tr}} \cup \mathcal{S}_{\text{cal}}$ as in Algorithm 1. With the training set $\mathbf{M}_{\text{tr}}$, we obtain an estimated likelihood function $\widehat{\pi}(m; i, j)$ such that

$$\widehat{\pi}(m; i, j) = \widehat{p}_{M^*_{ij}}(m),$$

which is an estimate for the true likelihood of M_{ij} at m. The estimation can be feasible given certain low-complexity structures. For example, if a hypothesized distribution family $\{\mathcal{Q}_\theta\}_{\theta \in \Theta}$ with probability density q_θ is given and the underlying mean matrix $\mathbf{M}^*$ is assumed to be low-rank. Then, $\mathbf{M}$ can be viewed as a perturbation of $\mathbf{M}^*$ and we can estimate $\mathbf{M}^*$ via matrix completion algorithms with entries in $\mathbf{M}_{\text{tr}}$. Denote $\widehat{\mathbf{M}}$ as the estimate for $\mathbf{M}^*$, then we have the estimated likelihood

$$\widehat{\pi}(m; i, j) = q_{\widehat{M}_{ij}}(m).$$

The odds ratios are also estimated from the training set, i.e. $\widehat{h}_{ij}$, from which we compute the weights

$$\widehat{w}_{ij} = \frac{\widehat{h}_{ij}}{\sum_{(i',j') \in \mathcal{S}_{\text{cal}}} \widehat{h}_{i'j'} + \max_{(i',j') \in \mathcal{S}^c} \widehat{h}_{i'j'}}, \; (i,j) \in \mathcal{S}_{\text{cal}}, \;\; \widehat{w}_{\text{test}} = \frac{\widehat{h}_{i_* j_*}}{\sum_{(i',j') \in \mathcal{S}_{\text{cal}}} \widehat{h}_{i'j'} + \max_{(i',j') \in \mathcal{S}^c} \widehat{h}_{i'j'}}.$$

For each $(i, j) \in \mathcal{S}_{\text{cal}}$, calculate the likelihood-based nonconformity score

$$R_{ij} = -\widehat{\pi}(M_{ij}; i, j).$$

Then, for any test point $(i_*, j_*) \in \mathcal{S}^c$, we can construct the confidence interval

$$\widehat{C}(i_*, j_*) = \{m \in [K] : \; \widehat{\pi}(m; i_*, j_*) \leq \widehat{q}\},$$

where $\widehat{q}$ is the weighted quantile

$$\widehat{q} = \text{Quantile}_{1-\alpha}\left(\sum_{(i,j) \in \mathcal{S}_{\text{cal}}} \widehat{w}_{ij} \cdot \delta_{R_{ij}} + \widehat{w}_{\text{test}} \cdot \delta_{+\infty} \right).$$

More examples of conformal methods for classification are shown in Romano et al. (2020), Angelopoulos et al. (2020), etc.

C.2 Full conformalized matrix completion

In Algorithm 2, the procedure of the full conformalized matrix completion (`full-cmc`) is presented. This full conformal version of `cmc` offers the same coverage guarantee as given in Theorem 3.2 for the split version of `cmc` (except with the entire observed set $\mathcal{S}$ in place of $\mathcal{S}_{\text{cal}}$, when defining $\Delta(\widehat{\boldsymbol{w}})$; the formal proof of this bound for full conformal is very similar to the proof of Theorem 3.2, using an analogous weighted exchangeability argument as in Lemma 3.1, and so we omit it here.

To define this algorithm, we need some notation: given the observed data $\mathbf{M}_{\mathcal{S}}$, plus a test point location (i_*, j_*) and a hypothesized value m for the test point $M_{i_* j_*}$, define a matrix $\mathbf{M}^{(m)}$ with entries

$$M^{(m)}_{ij} = \begin{cases} M_{ij}, & (i,j) \in \mathcal{S}, \\ m, & (i,j) = (i_*, j_*), \\ \emptyset, & \text{otherwise}, \end{cases} \tag{20}$$

Algorithm 2 `full-cmc`: full conformalized matrix completion

Input: target level $1-\alpha$; partially observed matrix $\mathbf{M}_{\mathcal{S}}$.
Using the training data $\mathbf{M}_{\mathcal{S}}$, compute an estimate $\widehat{\mathbf{P}}$ of the observation probabilities (with $\widehat{p}_{ij}$ estimating p_{ij}, the probability of entry (i,j) being observed).
for (i_*, j_*) in $\mathcal{S}^c$ **do**
  for $m \in \mathcal{M}$ **do**
    Augment $\mathbf{M}_{\mathcal{S}}$ with one additional hypothesized entry, $\{M_{i_*,j_*} = m\}$, to obtain $\mathbf{M}^{(m)}$ defined as in (20).
    Using the imputed matrix $\mathbf{M}_{\mathcal{S}}^{(m)}$, compute:

- An initial estimate $\widehat{\mathbf{M}}^{(m)}$ using any matrix completion algorithm (with $\widehat{M}_{ij}^{(m)}$ estimating the target M_{ij});
- Optionally, a local uncertainty estimate $\widehat{\mathbf{s}}^{(m)}$ (with $\widehat{s}_{ij}^{(m)}$ estimating our relative uncertainty in the estimate $\widehat{M}_{ij}^{(m)}$), or otherwise set $\widehat{s}_{ij}^{(m)} \equiv 1$;
- An estimate $\widehat{\mathbf{P}}$ of the observation probabilities (with $\widehat{p}_{ij}$ estimating p_{ij}, the probability of entry (i,j) being observed).

    Compute normalized residuals for $(i,j) \in \mathcal{S} \cup \{(i_*, j_*)\}$,

$$R_{ij}^{(m)} = \frac{|M_{ij} - \widehat{M}_{ij}^{(m)}|}{\widehat{s}_{ij}^{(m)}}.$$

    Compute weights

$$\widehat{w}_{ij} = \frac{\widehat{h}_{ij}}{\sum_{(i',j') \in \mathcal{S} \cup \{(i_*, j_*)\}} \widehat{h}_{i'j'}}, \qquad \widehat{h}_{ij} = \frac{1-\widehat{p}_{ij}}{\widehat{p}_{ij}}, \qquad (i,j) \in \mathcal{S} \cup \{(i_*, j_*)\}.$$

    Compute the weighted quantile

$$\widehat{q}^{(m)}(i_*, j_*) = \text{Quantile}_{1-\alpha}\left(\sum_{(i,j)\in\mathcal{S}} \widehat{w}_{ij}\delta_{R_{ij}^{(m)}} + \widehat{w}_{i_*j_*}\delta_{+\infty}\right) \tag{21}$$

  end for
end for
Output: $\left\{\widehat{C}\left(i_*, j_*\right) = \left\{m \in \mathcal{M} : \ R_{i_*j_*}^{(m)} \leq \widehat{q}^{(m)}(i_*, j_*)\right\} : \ (i_*, j_*) \in \mathcal{S}^c\right\}$

where, abusing notation, "$M_{ij} = \emptyset$" denotes that no information is observed in this entry.

We note that, when $\mathcal{M} = \mathbb{R}$ (or an infinite subset of $\mathbb{R}$), the computation of the prediction set is impossible in most of the cases. In that case, our algorithm can be modified via a trimmed or discretized approximation; these extensions are presented for the regression setting in the work of Chen et al. (2016, 2018), and can be extended to the matrix completion setting in a straightforward way.

C.3 Exact split conformalized matrix completion

In Algorithm 3, we present the exact split approach, which is less conservative than our one-shot approach given in Algorithm 1, but may be less computationally efficient. In this version of the algorithm, the quantile $\widehat{q} = \widehat{q}(i_*, j_*)$ needs to be computed for each missing entry since the weight vector $\widehat{\boldsymbol{w}}$ depends on the value of $\widehat{p}_{i_*,j_*}$.

Algorithm 3 `split-cmc`: split conformalized matrix completion

Input: target coverage level $1-\alpha$; data splitting proportion $q \in (0,1)$; observed entries $\mathbf{M}_{\mathcal{S}}$.

Split the data: draw $W_{ij} \overset{\text{i.i.d.}}{\sim} \mathrm{Bern}(q)$, and define training and calibration sets,

$$\mathcal{S}_{\text{tr}} = \{(i,j) \in \mathcal{S} : W_{ij} = 1\}, \quad \mathcal{S}_{\text{cal}} = \{(i,j) \in \mathcal{S} : W_{ij} = 0\}.$$

Using the training data $\mathbf{M}_{\mathcal{S}_{\text{tr}}}$ indexed by $\mathcal{S}_{\text{tr}} \subseteq [d_1] \times [d_2]$, compute:

- An initial estimate $\widehat{\mathbf{M}}$ using any matrix completion algorithm (with $\widehat{M}_{ij}$ estimating the target M_{ij});
- Optionally, a local uncertainty estimate $\widehat{\mathbf{s}}$ (with $\widehat{s}_{ij}$ estimating our relative uncertainty in the estimate $\widehat{M}_{ij}$), or otherwise set $\widehat{s}_{ij} \equiv 1$;
- An estimate $\widehat{\mathbf{P}}$ of the observation probabilities (with $\widehat{p}_{ij}$ estimating p_{ij}, the probability of entry (i,j) being observed).

Compute normalized residuals on the calibration set,

$$R_{ij} = \frac{\left|M_{ij} - \widehat{M}_{ij}\right|}{\widehat{s}_{ij}}, \ (i,j) \in \mathcal{S}_{\text{cal}}.$$

Compute estimated odds ratios for the calibration set and test set,

$$\widehat{h}_{ij} = \frac{\widehat{p}_{ij}}{1-\widehat{p}_{ij}}, \ (i,j) \in \mathcal{S}_{\text{cal}} \cup \mathcal{S}^c,$$

for $(i_*, j_*) \in \mathcal{S}^c$ **do**

Compute weights for the calibration set and test point,

$$\widehat{w}_{ij} = \frac{\widehat{h}_{ij}}{\sum_{(i',j') \in \mathcal{S}_{\text{cal}}} \widehat{h}_{i'j'} + \widehat{h}_{i_*j_*}}, \ (i,j) \in \mathcal{S}_{\text{cal}}, \ \ \widehat{w}_{\text{test}} = \frac{\widehat{h}_{i_*j_*}}{\sum_{(i',j') \in \mathcal{S}_{\text{cal}}} \widehat{h}_{i'j'} + \widehat{h}_{i_*j_*}}.$$

Compute threshold

$$\widehat{q}(i_*, j_*) = \text{Quantile}_{1-\alpha}\left(\sum_{(i,j) \in \mathcal{S}_{\text{cal}}} \widehat{w}_{ij} \cdot \delta_{R_{ij}} + \widehat{w}_{\text{test}} \cdot \delta_{+\infty} \right),$$

where δ_t denotes the point mass at t.

end for

Output: confidence intervals

$$\widehat{C}(i_*, j_*) = \widehat{M}_{i_*j_*} \pm \widehat{q}(i_*, j_*) \cdot \widehat{s}_{i_*j_*}$$

for each unobserved entry $(i_*, j_*) \in \mathcal{S}^c$.