Simultaneous statistical inference for epidemic trends

Marina Khismatullina Michael Vogt

1 Introduction

There are many questions surrounding the current COVID-19 pandemic that are not well understood yet. One important question is the following: How does the outbreak pattern of COVID-19 compare across countries? Are the time trends of daily new infections (and daily deaths) comparable across countries, or is the virus spreading differently in different regions of the world? The main aim of this paper is to develop new statistical methods that help to shed light on this issue.

Let X_{it} be the number of new infections on day t in country i and suppose we observe a sample of data $\mathcal{D}_i = \{X_{it} : 1 \le 1 \le T\}$ for n different countries i. A simple way to model the count data X_{it} is to use a Poisson distribution. Specifically, we may assume that the random variables X_{it} are Poisson distributed with time-varying intensity parameter $\lambda_i(t/T)$, that is, $X_{it} \sim \mathcal{P}_{\lambda_i(t/T)}$. Since $\lambda_i(t/T) = \mathbb{E}[X_{it}] = \text{Var}(X_{it})$, we can model the observations X_{it} by the nonparametric regression equation

$$X_{it} = \lambda_i\Big(\frac{t}{T}\Big) + u_{it}, \tag{1.1}$$

where $u_{it} = X_{it} - \mathbb{E}[X_{it}]$ with $\mathbb{E}[u_{it}] = 0$ and $\text{Var}(u_{it}) = \lambda_i(t/T)$. As usual in nonparametric regression, we let the regression function λ_i in model (1.1) depend on rescaled time t/T rather than on real time t; cp. [cite].

In model (1.1), the outbreak pattern of COVID-19 in country i is determined by the intensity function λ_i. Hence, the question whether the outbreak patterns are comparable across countries amounts to the question whether the intensity functions λ_i have the same shape across countries i. In this paper, we construct a multiscale test which allows to *identify* and *locate* the differences between the intensity functions λ_i. More specifically, let $\mathcal{F} = \{\mathcal{I}_k \subseteq [0,1] : k = 1,\dots,K\}$ be a family of (rescaled) time intervals and let $H_0^{(ijk)}$ be the hypothesis that the intensity functions λ_i and λ_j are the same on the interval $\mathcal{I}_k$, that is,

$$H_0^{(ijk)} : \lambda_i(w) = \lambda_j(w) \text{ for all } w \in \mathcal{I}_k.$$

We design a method to test the hypothesis $H_0^{(ijk)}$ *simultaneously* for all pairs of countries i and j and for all intervals $\mathcal{I}_k$ in the family $\mathcal{F}$. The main theoretical result of the paper shows that the method controls the family-wise error rate, that is, the probability of wrongly rejecting at least one null hypothesis $H_0^{(ijk)}$. As a consequence, we can make simultaneous confidence statements of the following form for a given significance level $\alpha \in (0,1)$:

> *With probability at least $1-\alpha$, the intensity functions λ_i and λ_j differ on the interval $\mathcal{I}_k$ for each (i,j,k) for which the test rejects $H_0^{(ijk)}$.*

Hence, our method allows us to make simultaneous confidence statements (i) about which intensity functions differ from each other and (ii) about where, that is, in which time intervals $\mathcal{I}_k$ they differ.

2 Model setting

As already discussed in the Introduction, the assumption that $X_{it} \sim \mathcal{P}_{\lambda_i(t/T)}$ leads to a nonparametric regression model of the form

$$X_{it} = \lambda_i\Big(\frac{t}{T}\Big) + u_{it} \quad \text{with} \quad u_{it} = \sqrt{\lambda_i\Big(\frac{t}{T}\Big)}\eta_{it}, \tag{2.1}$$

where η_{it} has zero mean and unit variance. In this model, both the mean and the noise variance are described by the same function λ_i. In empirical applications, however, the noise variance often tends to be much larger than that implied by the Poisson distribution; cp. [cite]. To deal with this issue, so-called quasi-Poisson models are frequently used. In our context, a quasi-Poisson model of X_{it} has the form

$$X_{it} = \lambda_i\Big(\frac{t}{T}\Big) + \varepsilon_{it} \quad \text{with} \quad \varepsilon_{it} = \sigma\sqrt{\lambda_i\Big(\frac{t}{T}\Big)}\eta_{it}, \tag{2.2}$$

where σ is a scaling factor that allows the noise variance to be a multiple of the mean function λ_i. In what follows, we assume that the observed data X_{it} are produced by model (2.2), where the noise residuals η_{it} have zero mean and unit variance but we do not impose any further distributional assumptions on them. In order to derive our theoretical results, we impose the following regularity conditions on model (2.2):

(C1) The random variables η_{it} are independent both across i and t. Moreover, for any i and t, $\mathbb{E}[\eta_{it}] = 0$, $\mathbb{E}[\eta_{it}^2] = 1$ and $\mathbb{E}[\eta_{it}^\theta] \le C_\theta < \infty$ for some $\theta >$??.

(C2) The intensity functions λ_i are uniformly Lipschitz continuous, that is, $|\lambda_i(u) - \lambda_i(v)| \le L|u-v|$ for all $u,v \in [0,1]$, where the constant L does not depend on i.

[Discuss why independence of η_{it} across i and t is justified.]

3 The multiscale test

Let $\mathcal{S} \subseteq \{(i,j) : 1 \le i < j \le n\}$ be the set of all pairs of countries (i,j) whose intensity functions λ_i and λ_j we want to compare. Moreover, as already introduced above, let $\mathcal{F} = \{\mathcal{I}_k : 1 \le k \le K\}$ be the family of (rescaled) time intervals under consideration. Finally, write $\mathcal{M} := \mathcal{S} \times \{1,\dots,K\}$ and let $p := |\mathcal{M}|$ be the cardinality of $\mathcal{M}$. In this section, we devise a method to test the null hypothesis $H_0^{(ijk)}$ simultaneously for all pairs of countries $(i,j) \in \mathcal{S}$ and all time intervals $\mathcal{I}_k \in \mathcal{F}$, that is, for all $(i,j,k) \in \mathcal{M}$. As shown by our theoretical results in Section [number], the method works under very weak conditions on the dimension p. In particular, p may be much larger than the time series length T, which allows us to deal with potentially very high-dimensional test problems.

3.1 Construction of the test statistics

A statistic to test the hypothesis $H_0^{(ijk)}$ for a given triple (i,j,k) can be constructed as follows. To start with, we introduce the expression

$$\hat{s}_{ijk,T} = \sum_{t=1}^{T} w_k\Big(\frac{t}{T}\Big)(X_{it} - X_{jt}),$$

where $w_k(t/T)$ is a (rectangular) kernel weight defined by

$$w_k\Big(\frac{t}{T}\Big) = \frac{\mathbf{1}(\frac{t}{T} \in \mathcal{I}_k)}{\{\sum_{s=1}^{T} \mathbf{1}(\frac{s}{T} \in \mathcal{I}_k)\}^{1/2}}.$$

Under (C1), it holds that

$$\nu_{ijk,T}^2 := \mathrm{Var}(\hat{s}_{ijk,T}) = \sigma^2 \sum_{t=1}^{T} w_k^2\Big(\frac{t}{T}\Big)\Big\{\lambda_i\Big(\frac{t}{T}\Big) + \lambda_j\Big(\frac{t}{T}\Big)\Big\}.$$

In order to normalize the variance of the statistic $\hat{s}_{ijk,T}$, we scale it by an estimator of $\nu_{ijk,T}$. In particular, we estimate $\nu_{ijk,T}^2$ by

$$\hat{\nu}_{ijk,T}^2 = \hat{\sigma}^2 \sum_{t=1}^{T} w_k^2\Big(\frac{t}{T}\Big)\{X_{it} + X_{jt}\},$$

where $\hat{\sigma}^2 = n^{-1}\sum_{i=1}^{n} \hat{\sigma}_i^2$ and

$$\hat{\sigma}_i^2 = \frac{\sum_{t=2}^{T}(X_{it} - X_{it-1})^2}{2\sum_{t=1}^{T} X_{it}}.$$

The estimator $\hat{\sigma}^2$ is motivated by the fact that under appropriate smoothness conditions on the functions λ_i (in particular under the Lipschitz condition of (C2)),

$$\begin{aligned}
X_{it} - X_{it-1} &= \lambda_i\Big(\frac{t}{T}\Big) - \lambda_i\Big(\frac{t-1}{T}\Big) + \sigma\sqrt{\lambda_i\Big(\frac{t}{T}\Big)}(\eta_{it} - \eta_{it-1}) \\
&\quad + \sigma\Big(\sqrt{\lambda_i\Big(\frac{t}{T}\Big)} - \sqrt{\lambda_i\Big(\frac{t-1}{T}\Big)}\Big)\eta_{it-1} \\
&= \sigma\sqrt{\lambda_i\Big(\frac{t}{T}\Big)}(\eta_{it} - \eta_{it-1}) + (1 + |\eta_{it-1}|)O(T^{-1}).
\end{aligned}$$

This suggests that $T^{-1}\sum_{t=2}^T (X_{it} - X_{it-1})^2 = 2\sigma^2 \int_0^1 \lambda_i(u)du + o_p(1)$. Moreover, since $T^{-1}\sum_{t=1}^T X_{it} = \int_0^1 \lambda_i(u)du + o_p(1)$, we expect that $\hat{\sigma}_i^2 = \sigma^2 + o_p(1)$ for any i and thus $\hat{\sigma}^2 = \sigma^2 + o_p(1)$. Normalizing the statistic $\hat{s}_{ijk,T}$ by the estimator $\hat{\nu}_{ijk,T}$ yields the expression

$$\hat{\psi}_{ijk,T} := \frac{\hat{s}_{ijk,T}}{\hat{\nu}_{ijk,T}} = \frac{\sum_{t=1}^T \mathbf{1}(\frac{t}{T} \in \mathcal{I}_k)(X_{it} - X_{jt})}{\hat{\sigma}\{\sum_{t=1}^T \mathbf{1}(\frac{t}{T} \in \mathcal{I}_k)(X_{it} + X_{jt})\}^{1/2}}, \tag{3.1}$$

which serves as our test statistic of the hypothesis $H_0^{(ijk)}$. For later reference, we additionally introduce the statistic

$$\hat{\psi}^0_{ijk,T} = \frac{\sum_{t=1}^T \mathbf{1}(\frac{t}{T} \in \mathcal{I}_k)\,\sigma\overline{\lambda}_{ij}^{1/2}(\frac{t}{T})(\eta_{it} - \eta_{jt})}{\hat{\sigma}\{\sum_{t=1}^T \mathbf{1}(\frac{t}{T} \in \mathcal{I}_k)(X_{it} + X_{jt})\}^{1/2}} \tag{3.2}$$

with $\overline{\lambda}_{ij}(u) = \{\lambda_i(u) + \lambda_j(u)\}/2$, which is identical to the test statistic $\hat{\psi}_{ijk,T}$ under $H_0^{(ijk)}$.

3.2 Construction of the test procedure

Our multiscale test is carried out as follows: For a given significance level $\alpha \in (0,1)$ and each $(i,j,k) \in \mathcal{M}$, we reject $H_0^{(ijk)}$ if

$$|\hat{\psi}_{ijk,T}| > c_{ijk,T}(\alpha),$$

where $c_{ijk,T}(\alpha)$ is the critical value for the (i,j,k)-th test problem. The critical values $c_{ijk,T}(\alpha)$ are chosen such that the familywise error rate (FWER) is controlled at the level α, which is defined as the probability of wrongly rejecting $H_0^{(ijk)}$ for at least one (i,j,k). More formally speaking, for a given significance level $\alpha \in (0,1)$,

$$\begin{aligned}
\text{FWER}(\alpha) &= \mathbb{P}\Big(\exists (i,j,k) \in \mathcal{M}_0 : |\hat{\psi}_{ijk,T}| > c_{ijk,T}(\alpha)\Big) \\
&= 1 - \mathbb{P}\Big(\forall (i,j,k) \in \mathcal{M}_0 : |\hat{\psi}_{ijk,T}| \le c_{ijk,T}(\alpha)\Big)
\end{aligned}$$

$$= 1 - \mathbb{P}\Big(\max_{(i,j,k) \in \mathcal{M}_0} |\hat{\psi}_{ijk,T}| \le c_{ijk,T}(\alpha) \Big),$$

where $\mathcal{M}_0 \subseteq \mathcal{M}$ is the set of triples (i,j,k) for which $H_0^{(ijk)}$ holds true.

There are different ways to construct critical values $c_{ijk,T}(\alpha)$ that ensure control of the FWER. In the traditional approach, the same critical value $c_T(\alpha) = c_{ijk,T}(\alpha)$ is used for all (i,j,k). In this case, controlling the FWER at the level α requires to determine the critical value $c_T(\alpha)$ such that

$$1 - \mathbb{P}\Big(\max_{(i,j,k) \in \mathcal{M}_0} |\hat{\psi}_{ijk,T}| \le c_T(\alpha) \Big) \le \alpha. \tag{3.3}$$

This can be achieved by choosing $c_T(\alpha)$ as the $(1-\alpha)$-quantile of the statistic

$$\tilde{\Psi}_T = \max_{(i,j,k) \in \mathcal{M}} |\hat{\psi}^0_{ijk,T}|,$$

where $\hat{\psi}^0_{ijk,T}$ has been introduced in (3.2). (Note that both the statistic $\tilde{\Psi}_T$ and the quantile $c_T(\alpha)$ depend on p in general. To keep the notation simple, we however suppress this dependence throughout the paper. We use the same convention for all other quantities that are defined in the sequel.)

In more modern approaches, different critical values $c_{ijk,T}(\alpha)$ are assigned to the test problems (i,j,k). In particular, the critical values are allowed to depend on the length h_k of the time interval $\mathcal{I}_k$, that is, on the scale of the test problem. A general approach to construct scale-dependent critical values was pioneered by [cite] and has been used in many other studies since then; cp. [cite]. In our context, the approach of [cite] leads to the critical values

$$c_{ijk,T}(\alpha) = c_T(\alpha, h_k) := b_k + q_T(\alpha)/a_k,$$

where $a_k = \{\log(e/h_k)\}^{1/2}/\log\log(e^e/h_k)$ and $b_k = \sqrt{2\log(1/h_k)}$ are scale-dependent constants and the quantity $q_T(\alpha)$ is determined by the following consideration: Since

$$\begin{aligned}
\mathrm{FWER}(\alpha) &= \mathbb{P}\Big(\exists (i,j,k) \in \mathcal{M}_0 : |\hat{\psi}_{ijk,T}| > c_T(\alpha, h_k) \Big) \\
&= 1 - \mathbb{P}\Big(\forall (i,j,k) \in \mathcal{M}_0 : |\hat{\psi}_{ijk,T}| \le c_T(\alpha, h_k) \Big) \\
&= 1 - \mathbb{P}\Big(\forall (i,j,k) \in \mathcal{M}_0 : a_k\big(|\hat{\psi}_{ijk,T}| - b_k\big) \le q_T(\alpha) \Big) \\
&= 1 - \mathbb{P}\Big(\max_{(i,j,k) \in \mathcal{M}_0} a_k\big(|\hat{\psi}_{ijk,T}| - b_k\big) \le q_T(\alpha) \Big), \qquad (3.4)
\end{aligned}$$

we need to choose the quantity $q_T(\alpha)$ as the $(1-\alpha)$-quantile of the statistic

$$\hat{\Psi}_T = \max_{(i,j,k)\in\mathcal{M}} a_k\big(|\hat{\psi}^0_{ijk,T}| - b_k\big)$$

in order to ensure control of the FWER at level α. Comparing (3.4) with (3.3), the current approach can be seen to differ from the traditional one in the following respect: the maximum statistic $\tilde{\Psi}_T$ is replaced by the rescaled version $\hat{\Psi}_T$ which re-weights the individual statistics $\hat{\psi}_{ijk,T}$ by the scale-dependent constants a_k and b_k. As demonstrated above, this translates into scale-dependent critical values $c_{ijk,T}(\alpha) = c_T(\alpha, h_k)$.

Our theory allows us to work with both the traditional choice $c_{ijk,T}(\alpha) = c_T(\alpha)$ and the more modern, scale-dependent choice $c_{ijk,T}(\alpha) = c_T(\alpha, h_k)$. Since the latter choice produces a test with better theoretical properties (cp. [cite]), we restrict attention to the critical values $c_T(\alpha, h_k)$ in the sequel. There is, however, one complication we need to deal with: As the quantiles $q_T(\alpha)$ are not known in practice, we can not compute the critical values $c_T(\alpha, h_k)$ exactly in practice but need to approximate them. This can be achieved as follows: Under appropriate regularity conditions, it can be shown that

$$\begin{aligned}\hat{\psi}^0_{ijk,T} &= \frac{\sum_{t=1}^T \mathbf{1}(\frac{t}{T} \in \mathcal{I}_k)\, \sigma \overline{\lambda}_{ij}^{1/2}(\frac{t}{T})(\eta_{it} - \eta_{jt})}{\hat{\sigma}\{\sum_{t=1}^T \mathbf{1}(\frac{t}{T} \in \mathcal{I}_k)(X_{it} + X_{jt})\}^{1/2}} \\ &\approx \frac{1}{\sqrt{2Th_k}} \sum_{t=1}^T \mathbf{1}\Big(\frac{t}{T} \in \mathcal{I}_k\Big)\{\eta_{it} - \eta_{jt}\}.\end{aligned}$$

A Gaussian version of the statistic displayed in the final line above is given by

$$\phi_{ijk,T} = \frac{1}{\sqrt{2Th_k}} \sum_{t=1}^T \mathbf{1}\Big(\frac{t}{T} \in \mathcal{I}_k\Big)\big\{Z_{it} - Z_{jt}\big\},$$

where Z_{it} are independent standard normal random variables for $1 \le t \le T$ and $1 \le i \le n$. Hence, the statistic

$$\Phi_T = \max_{(i,j,k)\in\mathcal{M}} a_k\big(|\phi_{ijk,T}| - b_k\big)$$

can be regarded as a Gaussian version of the statistic $\hat{\Psi}_T$. We approximate the unknown quantile $q_T(\alpha)$ by the $(1-\alpha)$-quantile $q_{T,\text{Gauss}}(\alpha)$ of Φ_T, which can be computed (approximately) by Monte Carlo simulations and can thus be treated as known.

To summarize, we propose the following procedure to simultaneously test the hypotheses $H_0^{(ijk)}$ for all $(i,j,k) \in \mathcal{M}$ at the significance level $\alpha \in (0,1)$:

$$\text{For each } (i,j,k) \in \mathcal{M}, \text{ reject } H_0^{(ijk)} \text{ if } |\hat{\psi}_{ijk,T}| > c_{T,\text{Gauss}}(\alpha, h_k), \tag{3.5}$$

where $c_{T,\text{Gauss}}(\alpha, h_k) = b_k + q_{T,\text{Gauss}}(\alpha)/a_k$ with $a_k = \{\log(e/h_k)\}^{1/2}/\log\log(e^e/h_k)$ and $b_k = \sqrt{2\log(1/h_k)}$.

4 Theoretical results

Theorem 4.1. *Let (C1) and (C2) be satisfied and assume that* $p \leq ??$, $\min_k h_k \geq ??$ *and* $\max_k h_k \leq ??$. *Then it holds that*

$$\text{FWER}(\alpha) := \mathbb{P}\Big(\exists (i,j,k) \in \mathcal{M}_0 : |\hat{\psi}_{ijk,T}| > c_{T,\text{Gauss}}(\alpha, h_k)\Big) \leq \alpha + o(1)$$

for any given $\alpha \in (0,1)$. *Hence, the simultaneous test defined in* (3.5) *asymptotically controls the FWER at level* α *for any given significance level* $\alpha \in (0,1)$.

Corollary 4.1. *Under the conditions of Theorem 4.1,*

$$\mathbb{P}\Big(\forall (i,j,k) \in \mathcal{M} : \textit{ If } |\hat{\psi}_{ijk,T}| > c_{T,\text{Gauss}}(\alpha, h_k), \textit{ then } (i,j,k) \notin \mathcal{M}_0\Big) \geq 1 - \alpha + o(1).$$

Hence, with asymptotic probability at least $1-\alpha$, *the two intensity functions* λ_i *and* λ_j *are different on the interval* $\mathcal{I}_k$ *for all* $(i,j,k) \in \mathcal{M}$ *for which the test rejects* $H_0^{(ijk)}$.

Proof of Theorem 4.1. The strategy to prove Theorem 4.1 is as follows:

(1) Let

$$\hat{\psi}^0_{ijk,T} = \frac{\sum_{t=1}^T \mathbf{1}(\frac{t}{T} \in \mathcal{I}_k)\, \sigma \overline{\lambda}_{ij}^{1/2}(\frac{t}{T})(\eta_{it} - \eta_{jt})}{\hat{\sigma}\{\sum_{t=1}^T \mathbf{1}(\frac{t}{T} \in \mathcal{I}_k)(X_{it} + X_{jt})\}^{1/2}}$$
$$\psi^0_{ijk,T} = \frac{1}{\sqrt{2Th_k}} \sum_{t=1}^T \mathbf{1}\Big(\frac{t}{T} \in \mathcal{I}_k\Big)(\eta_{it} - \eta_{jt}).$$

Prove that

$$\max_{(i,j,k) \in \mathcal{M}} \big|\hat{\psi}^0_{ijk,T} - \psi^0_{ijk,T}\big| = o_p(r_T),$$

where $r_T \to 0$ sufficiently fast as $T \to \infty$. This implies that

$$\big|\hat{\Psi}_T - \Psi_T\big| = o_p(r_T), \tag{4.1}$$

where we define $\Psi_T = \max_{(i,j,k)\in\mathcal{M}} a_k(|\psi^0_{ijk,T}| - b_k)$.

(2) Define

$$V_t^{(ijk)} = V_{t,T}^{(ijk)} := \sqrt{\frac{T}{2Th_k}} \mathbf{1}\Big(\frac{t}{T} \in \mathcal{I}_k\Big)(\eta_{it} - \eta_{jt})$$

for $(i,j,k) \in \mathcal{M}$ and let $\boldsymbol{V}_t = (V_t^{(ijk)} : (i,j,k) \in \mathcal{M})$ be the p-dimensional random vector with the entries $V_t^{(ijk)}$. With this notation, we get that $\psi^0_{ijk,T} = T^{-1/2} \sum_{t=1}^T V_t^{(ijk)}$ and thus

$$\begin{aligned}
\Psi_T &= \max_{(i,j,k) \in \mathcal{M}} a_k \big(|\psi_{ijk,T}| - b_k\big) \\
&= \max_{(i,j,k) \in \mathcal{M}} a_k \Big\{ \Big| \frac{1}{\sqrt{T}} \sum_{t=1}^T V_t^{(ijk)} \Big| - b_k \Big\}.
\end{aligned}$$

Analogously, we define

$$W_t^{(ijk)} = W_{t,T}^{(ijk)} := \sqrt{\frac{T}{2Th_k}} \mathbf{1}\Big(\frac{t}{T} \in \mathcal{I}_k\Big)(Z_{it} - Z_{jt})$$

with Z_{it} i.i.d. standard normal and let $\boldsymbol{W}_t = (W_t^{(ijk)} : (i,j,k) \in \mathcal{M})$. The vector $\boldsymbol{W}_t$ is a Gaussian version of $\boldsymbol{V}_t$ with the same mean and variance. In particular, $\mathbb{E}[\boldsymbol{W}_t] = \mathbb{E}[\boldsymbol{V}_t] = 0$ and $\mathbb{E}[\boldsymbol{W}_t \boldsymbol{W}_t^\top] = \mathbb{E}[\boldsymbol{V}_t \boldsymbol{V}_t^\top]$. Similarly as before, we can write $\phi_{ijk,T} = T^{-1/2} \sum_{t=1}^T W_t^{(ijk)}$ and

$$\begin{aligned}
\Phi_T &= \max_{(i,j,k) \in \mathcal{M}} a_k \big(|\phi_{ijk,T}| - b_k\big) \\
&= \max_{(i,j,k) \in \mathcal{M}} a_k \Big\{ \Big| \frac{1}{\sqrt{T}} \sum_{t=1}^T W_t^{(ijk)} \Big| - b_k \Big\}.
\end{aligned}$$

For any $q \in \mathbb{R}$,

$$\begin{aligned}
\mathbb{P}\big(\Psi_T \leq q\big) &= \mathbb{P}\Big(\max_{(i,j,k) \in \mathcal{M}} a_k \Big\{ \Big| \frac{1}{\sqrt{T}} \sum_{t=1}^T V_t^{(ijk)} \Big| - b_k \Big\} \leq q \Big) \\
&= \mathbb{P}\Big(\Big| \frac{1}{\sqrt{T}} \sum_{t=1}^T V_t^{(ijk)} \Big| \leq q_{ijk} \text{ for all } (i,j,k) \in \mathcal{M} \Big) \\
&= \mathbb{P}\Big(\Big| \frac{1}{\sqrt{T}} \sum_{t=1}^T \boldsymbol{V}_t \Big| \leq \boldsymbol{q} \Big),
\end{aligned}$$

where $\boldsymbol{q}$ is the $\mathbb{R}^p$-vector with the entries $q_{ijk} = q/a_k + b_k$. Analogously,

$$\mathbb{P}\big(\Phi_T \leq q\big) = \mathbb{P}\Big(\Big| \frac{1}{\sqrt{T}} \sum_{t=1}^T \boldsymbol{W}_t \Big| \leq \boldsymbol{q} \Big).$$

We want to prove that

$$\sup_{q\in\mathbb{R}}\Big|\mathbb{P}\big(\Psi_T\le q\big)-\mathbb{P}\big(\Phi_T\le q\big)\Big|=o(1). \tag{4.2}$$

To achieve this, we make use of Proposition 2.1 from [cite]. Applied to the situation at hand, this proposition can be stated as follows:

Proposition 4.1. *Assume that*

(a) $T^{-1}\sum_{t=1}^T\mathbb{E}(V_t^{(ijk)})^2\ge c>0$ *for all* $(i,j,k)\in\mathcal{M}$.

(b) $T^{-1}\sum_{t=1}^T\mathbb{E}[|V_t^{(ijk)}|^{2+r}]\le C_T^r$ *for all* $(i,j,k)\in\mathcal{M}$ *and* $r=1,2$, *where* $C_T\ge 1$ *are constants that may tend to infinity as* $T\to\infty$.

(c) $\mathbb{E}[\{\max_{(i,j,k)\in\mathcal{M}}|V_t^{(ijk)}|/C_T\}^s]\le 2$ *for all* t *and some* $s>2$.

Then

$$\begin{aligned}\sup_{\boldsymbol{q}\in\mathbb{R}^p}\Big|\mathbb{P}\Big(\Big|\frac{1}{\sqrt{T}}\sum_{t=1}^T\boldsymbol{V}_t\Big|\le\boldsymbol{q}\Big)-\mathbb{P}\Big(\Big|\frac{1}{\sqrt{T}}\sum_{t=1}^T\boldsymbol{W}_t\Big|\le\boldsymbol{q}\Big)\Big|&\\ \le C\Big\{\Big(\frac{C_T^2\log^7(pn)}{n}\Big)^{1/6}+\Big(\frac{C_T^2\log^3(pn)}{n^{1-2/s}}\Big)^{1/3}\Big\},&\end{aligned}$$

where C *depends only on* c *and* s.

Proposition 4.1 obviously implies (4.2). We merely need to verify that assumptions (a)–(c) are satisfied under the conditions of Theorem 4.1.

(3) We next want to show that

$$\sup_{q\in\mathbb{R}}\Big|\mathbb{P}\big(\hat{\Psi}_T\le q\big)-\mathbb{P}\big(\Phi_T\le q\big)\Big|=o(1). \tag{4.3}$$

It holds that

$$\begin{aligned}&\sup_{q\in\mathbb{R}}\Big|\mathbb{P}\big(\hat{\Psi}_T\le q\big)-\mathbb{P}\big(\Phi_T\le q\big)\Big|\\ &=\sup_{q\in\mathbb{R}}\Big|\mathbb{P}\Big(\Psi_T\le q+\big\{\Psi_T-\hat{\Psi}_T\big\}\Big)-\mathbb{P}\big(\Phi_T\le q\big)\Big|\\ &\le\max\Big\{\sup_{q\in\mathbb{R}}\Big|\mathbb{P}\Big(\Psi_T\le q+\big|\Psi_T-\hat{\Psi}_T\big|\Big)-\mathbb{P}\big(\Phi_T\le q\big)\Big|,\\ &\qquad\quad\sup_{q\in\mathbb{R}}\Big|\mathbb{P}\Big(\Psi_T\le q-\big|\Psi_T-\hat{\Psi}_T\big|\Big)-\mathbb{P}\big(\Phi_T\le q\big)\Big|\Big\}\end{aligned}$$

$$\le \max\Big\{\sup_{q\in\mathbb{R}}\Big|\mathbb{P}\Big(\Psi_T \le q + r_T\Big) - \mathbb{P}(\Phi_T \le q)\Big| + \mathbb{P}\Big(\big|\Psi_T - \hat{\Psi}_T\big| > r_T\Big),$$

$$\sup_{q\in\mathbb{R}}\Big|\mathbb{P}\Big(\Psi_T \le q - r_T\Big) - \mathbb{P}(\Phi_T \le q)\Big| + \mathbb{P}\Big(\big|\Psi_T - \hat{\Psi}_T\big| > r_T\Big)\Big\}.$$

Moreover,

$$\begin{aligned}
&\sup_{q\in\mathbb{R}}\Big|\mathbb{P}\Big(\Psi_T \le q \pm r_T\Big) - \mathbb{P}(\Phi_T \le q)\Big| + \mathbb{P}\Big(\big|\Psi_T - \hat{\Psi}_T\big| > r_T\Big)\\
&\le \sup_{q\in\mathbb{R}}\Big|\mathbb{P}\Big(\Psi_T \le q \pm r_T\Big) - \mathbb{P}\Big(\Phi_T \le q \pm r_T\Big)\Big|\\
&\quad + \sup_{q\in\mathbb{R}}\Big|\mathbb{P}\Big(\Phi_T \le q \pm r_T\Big) - \mathbb{P}(\Phi_T \le q)\Big| + \mathbb{P}\Big(\big|\Psi_T - \hat{\Psi}_T\big| > r_T\Big)\\
&= \sup_{q\in\mathbb{R}}\Big|\mathbb{P}\Big(\Phi_T \le q \pm r_T\Big) - \mathbb{P}(\Phi_T \le q)\Big| + o_p(1)
\end{aligned}$$

by (4.1) and (4.2). Finally, by Nazarov's inequality [cite],

$$\begin{aligned}
&\sup_{q\in\mathbb{R}}\Big|\mathbb{P}\Big(\Phi_T \le q \pm r_T\Big) - \mathbb{P}(\Phi_T \le q)\Big|\\
&\le \sup_{\boldsymbol{q}\in\mathbb{R}^p}\Big|\mathbb{P}\Big(\Big|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{W}_t\Big| \le \boldsymbol{q} \pm r_T\Big) - \mathbb{P}\Big(\Big|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{W}_t\Big| \le \boldsymbol{q}\Big)\Big| \le C r_T\sqrt{\log(2p)},
\end{aligned}$$

where C only depends on the constant c defined in condition (a) of Proposition 4.1.

(4) Since $\mathbb{P}(\Phi_T \le q_{T,\mathrm{Gauss}}(\alpha)) = 1-\alpha$ for any $\alpha \in (0,1)$ [add proof of this claim], (4.3) immediately implies that

$$\mathbb{P}\big(\hat{\Psi}_T \le q_{T,\mathrm{Gauss}}(\alpha)\big) = 1 - \alpha + o(1), \tag{4.4}$$

from which Theorem 4.1 immediately follows. □