Analyzing categorical time series in the presence of missing observations

Christian H. Weiß

Department of Mathematics and Statistics, Helmut Schmidt University, Hamburg, Germany

Correspondence
Christian H. Weiß, Department of Mathematics and Statistics, Helmut Schmidt University, Holstenhofweg 85, 22043 Hamburg, Germany.
Email: weissc@hsu-hh.de

In real applications, time series often exhibit missing observations such that standard analytical tools cannot be applied. While there are approaches of how to handle missing data in quantitative time series, the case of categorical time series seems not to have been treated so far. Both for the case of ordinal and nominal time series, solutions are developed that allow to analyze their marginal and serial properties in the presence of missing observations. This is achieved by adapting the concept of amplitude modulation, which allows to obtain closed-form asymptotic expressions for the derived statistics’ distribution (assuming that missingness happens independently of the actual process). The proposed methods are investigated with simulations, and they are applied in a project on migraine patients, where the monitored qualitative time series are often incomplete.

KEYWORDS
dispersion, incomplete data, nominal time series, ordinal time series, serial dependence

1 INTRODUCTION

There is a huge toolbox of models and methods for time series analysis, see, for example, the textbooks by Box et al,1 Weiß,2 These methods usually assume that the given time series has been fully observed. In real applications, however, one is often confronted with missing data, which makes it impossible to apply the standard analytical tools. Such missing data might be caused by different reasons, thus leading to different patterns of the missing observations.3 “Missingness” in time series might be caused by recurrent shutdowns of the measuring device or changes in the measuring period.4 Observations might be lost because of outlier removal in statistical process monitoring (surveillance),5 or observations are missed randomly due to sporadic failures.6,7 To be able to deal with real-valued time series affected by missing observations, there have been several proposals in the literature of how to adapt the standard analytical tools to this scenario, such as the sample autocorrelation function (ACF) or spectral estimators, see Parzen,8 Scheinok,6 Bloomfield,7 Neave,4 Dunsmuir and Robinson,9 Yajima and Nishino,10 and others. These approaches use the idea of Parzen8 to understand real-valued time series with missing observations as a special case of amplitude-modulated time series, where the amplitude-modulating process is binary and independent of the actual process. Then, the considered time series tools are applied to the amplitude-modulated time series, and their asymptotics change according to the type of missingness. Scheinok6 assumed the amplitude-modulating process to be independent and identically distributed (i.i.d.), whereas Bloomfield7 allowed for serial dependence. Neave4 considered a deterministic missingness mechanism, and comprehensive analyses are provided by Dunsmuir and Robinson,9 Yajima and Nishino.10

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2021 The Authors. Statistics in Medicine published by John Wiley & Sons Ltd.

Statistics in Medicine. 2021;1–16. wileyonlinelibrary.com/journal/sim
The aforementioned approaches to deal with missing data refer to real-valued time series. But missing observations may also happen to qualitative time series, which consist of categorical values ordered in time. For such time series, completely different analytical tools have to be used, and, consequently, the aforementioned solutions for missing data are not applicable. The main practical motivation for the present article is a collaborative project involving real-world qualitative time series from migraine patients with missing data.

**Example 1.** N1-Headache is a mobile app for migraine patients that was developed by Curelator Inc. During initial registration, participants give agreement for their data to be used for research and publication purposes by accepting the Curelator Inc terms and conditions and privacy policy. The protocol for analysis and publication of participant data was reviewed and approved by the Biomedical Research Alliance of New York (BRANY) Institutional Review Board (IRB) on October 10, 2019. BRANY granted a full waiver from informed consent. The app requires the patients “to log information about their headaches, migraine symptoms and medication use on days they have an attack, and track a range of factors (moods, weather, diet, etc.) on a daily basis that may influence their risk of attack” (https://n1-headache.com/patients/faq; retrieved April 16, 2021). These data are then used to identify the lifestyle factors and medications that affect the migraine attacks, also see Vives-Mestres et al., Vives-Mestres and Casanova for further background. Many of the considered features are categorical, such as binary symptom features (“true” or “false”), dietary features regarding the consumption of certain foods (“no,” “some,” “a lot”), or emotional features being expressed on a 0 to 10 Likert scale (ranging from “not at all” to “a lot”). Also the variable of main interest, pain peak severity, is measured on an ordinal scale with levels “none,” “mild,” “moderate,” and “severe.”

Thus, for each patient, a lot of categorical time series (measured on a daily basis) are available. These time series, however, often suffer from missing data, because patients did not enter any value into the app at some days, or they stopped before having replied to all questions. Two illustrative examples are shown in Figures 1 and 2. Figure 1 plots the ordinal peak-severity time series (“How bad was your headache pain at its worst today?”) obtained for, say, patient A, which is of length \(n = 225\) but including 19 missing observations (highlighted by dotted lines). Figure 2 refers to another patient, B, and shows one of the emotional features, the variable stress (“How stressed have you felt today?”) being measured on a 0–10 Likert scale. This ordinal time series is of length \(n = 136\) and contains three missing observations. Both time series are further discussed later in Section 4.2.

There are further fields of application where one is commonly confronted with categorical time series exhibiting missing observations. As an example, the measurement of cloud coverage at weather stations leads to ordinal time series, where the measurement at a certain time may fail for several reasons: besides a failure of the measuring device, also weather or environmental conditions (such as heavy precipitation, fog, or light pollution) may prevent the precise determination of the current coverage state. But in what follows, we focus on the migraine data of Example 1.

---

"Personal communication with Walter Koelschtzky from the Deutscher Wetterdienst (German Weather Service)."
In this article, we develop solutions to handle missing data in stationary categorical time series, $X_1, \ldots, X_n$ with $n \in \mathbb{N} = \{1, 2, \ldots \}$. That is, the outcomes $x_i$ of $X_i$ are assumed to stem from a qualitative range $S$ consisting of a finite number of categories. We distinguish between the case, where $S$ exhibits a natural order among the categories (ordinal time series), and the case where the categories are unordered (nominal time series), because this distinction is also associated with different analytic tools. Section 2 briefly reviews the basic approaches for analyzing both types of categorical time series (if no data are missing). Then, we derive solutions of how to handle missing data in categorical time series. We present a unique approach of incorporating missingness and of deriving the asymptotic distributions of the proposed statistics. But as the specific analytic tools differ between ordinal and nominal time series, we present the respective concepts for dealing with the missing data; otherwise, they are postponed to the supplement. Section 4 investigates the finite-sample performance of the obtained asymptotic approximations with some simulations. In particular, we pick up again Example 1 and apply the novel missing-data approaches to these ordinal time series. Due to the unified framework, the ordinal solutions of Section 3 are easily adapted to the case of nominal time series. A summary of the results for the nominal case is presented in Section 5, whereas the necessary derivations are again postponed to the supplement. The article is concluded in Section 6. The companion GitHub repository at https://github.com/cnweiss/CaTS_Missing contains all relevant implementation codes for this article.

## 2 BASICS OF CATEGORICAL TIME SERIES ANALYSIS

Let $(X_t)$ be a stationary categorical process, with either nominal or ordinal range $S$. To simplify notations, let the possible outcomes be arranged in a certain order (either lexicographical or natural order), that is, we denote the range as $S = \{s_0, s_1, \ldots, s_m\}$ with some $m \in \mathbb{N}$. Although the range is denoted in a unique way, ordinal time series have to be analyzed in a different way than nominal ones. For nominal time series, analytic tools are based on the probability mass function (PMF). We denote the marginal PMF by $p = (p_0, \ldots, p_m) \in [0; 1]^{m+1}$ with $p_i = P(X_t = s_i)$, and bivariate probabilities with time lag $h \in \mathbb{N}$ by $p_{ij}(h) = P(X_{t+h} = s_i, X_t = s_j)$. For ordinal time series, we make use of the natural ordering by considering the CDF instead, with the notations $f = (f_0, \ldots, f_{m-1}) \in [0; 1]^m$ where $f_i = P(X_t \leq s_i)$, and $f_{ij}(h) = P(X_{t+h} \leq s_i, X_t \leq s_j)$ for time lag $h \in \mathbb{N}$.

The stochastic properties of a nominal process are measured by different approaches than those of an ordinal process. While the location of a nominal process is expressed by computing the mode from $p$, one commonly computes the median from $f$ to express the location of an ordinal process. Also the concepts of dispersion are different. While dispersion is uniquely minimal in the case of a one-point distribution, expressed by the PMF vectors

$$p \in \begin{bmatrix} 1 & 0 & \ldots & 0 \ 0 & 1 & \ldots & 0 \ \vdots & \vdots & \ddots & \vdots \ 0 & 0 & \ldots & 1 \end{bmatrix} =: \{e_0, \ldots, e_m\} \subset [0; 1]^{m+1} \tag{1}$$

in the nominal case, or by the CDF vectors

$$f \in \begin{bmatrix} 1 & 0 & \ldots & 0 \ 1 & 1 & \ldots & 0 \ \vdots & \vdots & \ddots & \vdots \ 1 & 0 & \ldots & 1 \end{bmatrix} =: \{e_0, \ldots, e_m\} \subset [0; 1]^m \tag{2}$$

in the ordinal case, respectively, the concepts of maximal dispersion differ. While a nominal random variable exhibits maximal dispersion if $p$ equals the uniform distribution $(\frac{1}{m+1}, \ldots, \frac{1}{m+1})^\top$ (all outcomes are equally probable), an ordinal
random variable has maximal dispersion for an extreme two-point distribution (the outermost categories have probability 1/2), that is, if the CDF vector \( \hat{f} \) equals \( (\frac{1}{2}, \ldots, \frac{1}{2}) \). Several dispersion measures have been proposed in the literature. A popular choice is the index of qualitative (ordinal) variation, abbreviated as IQV (IOV), which is given by

\[
IQV = \frac{m + 1}{m} \left( 1 - \sum_{i=0}^{m} p_i^2 \right) \quad \text{and} \quad IOV = \frac{4}{m} \sum_{i=0}^{m-1} f_i(1 - f_i),
\]

respectively.\(^{15}\) Both are normalized to values in [0; 1], where 0 (1) is attained exactly in the case of minimal (maximal) dispersion. For ordinal data, also measures of asymmetry or skewness are derived from \( f \), see Klein and Doll,\(^ {16} \) Weiß\(^ {14} \) as well as Section 3.3. Measures of serial dependence, in turn, make use of the bivariate PMF or CDF, respectively. As qualitative counterparts to the ACF, one may use the nominal or ordinal Cohen’s \( \kappa \),\(^ {14,17} \) given by

\[
\kappa_{\text{nom}}(h) = \frac{\sum_{i=0}^{m} (p_i(h) - p_i^2)}{1 - \sum_{i=0}^{m} p_i^2} \quad \text{and} \quad \kappa_{\text{ord}}(h) = \frac{\sum_{i=0}^{m-1} (f_i(h) - f_i^2)}{1 - \sum_{i=0}^{m-1} f_i(1 - f_i)}\]

respectively. A \( \kappa \)-value of zero implies serial independence at lag \( h \), whereas positive (negative) \( \kappa \)-values indicate positive (negative) serial dependence.

The sample counterparts of the measures in (3) and (4) are obtained by replacing the (cumulative) relative frequencies by (cumulative) probabilities. The latter are most easily expressed if switching from \( (X_t) \) to its binarization \( (Y_t) \) or \( (Z_t) \), respectively, defined by \( Y_{t,i} = 1_{\{X_t = s_i\}} \) for \( i \in \{0, \ldots, m\} \) in the nominal case, and by \( Z_{t,i} = 1_{\{X_t \leq A\}} \) for \( i \in \{0, \ldots, m-1\} \) in the ordinal case. Here, \( 1_A \) denotes the indicator function, which takes the value 1 (0) if \( A \) is true (false). Note that the nominal binarization \( (Y_t) \) has the range \( \{e_0, \ldots, e_m\} \) according to (1), whereas the ordinal binarization \( (Z_t) \) has the range \( \{e_0, \ldots, e_m\} \) according to (2). Now, the marginal (cumulative) relative frequencies computed from \( X_1, \ldots, X_n \) are equal to means about their binarizations, ie.

\[
\hat{p} = \frac{1}{n} \sum_{i=1}^{n} Y_i, \quad \hat{f} = \frac{1}{n} \sum_{i=1}^{n} Z_i.
\]

For the bivariate (cumulative) relative frequencies, we compute

\[
\hat{p}_{ij}(h) = \frac{1}{n-h} \sum_{i=h+1}^{n} Y_{i,i} Y_{i-h,j}, \quad \hat{f}_{ij}(h) = \frac{1}{n-h} \sum_{i=h+1}^{n} Z_{i,i} Z_{i-h,j}.
\]

For deriving the asymptotics of the sample measures \( \hat{I}_{\text{IQV}}, \hat{I}_{\text{IOV}} \) in (3) and \( \hat{\kappa}_{\text{nom}}(h), \hat{\kappa}_{\text{ord}}(h) \) in (4), or those of alternative measures, one first requires a CLT for the relative frequencies in (5) to (6). Afterward, one derives the asymptotics of the measures based on Taylor approximations (“Delta method”), see Weiß,\(^ {17,14} \) for details. It should be noted, however, that all of the aforementioned approaches assume the time series to be fully observed.

In the sequel, however, we focus on categorical time series with missing observations, and we present a unified framework for handling missingness in both ordinal and nominal time series. We start with adapting the definition of the above sample measures (3) to (6) to deal with missing data. Then, we follow a similar path as the one sketched before for deriving their asymptotics: First, CLTs are proven for the missingness-corrected types of (cumulative) relative frequencies (5) to (6). Second, these CLTs are applied together with Taylor expansions to achieve the asymptotic distributions of the considered sample measures of dispersion (3), serial dependence (4), and others. Although the general approach for handling missingness is the same for both ordinal and nominal time series, the relevant methods differ in detail (as discussed before, we have different concepts of dispersion, for example, and skewness is well defined only for ordinal data). Therefore, we present the results for both cases separately and start, motivated by Example 1, with ordinal time series in Section 3. But thanks to the unified framework, the presentation and derivation...
of the corresponding nominal results, later in Section 5, is much simplified, also see the discussion in Remark S.4.1 of Supplement S.4.

3 | ORDINAL TIME SERIES WITH MISSING DATA

3.1 | Amplitude modulation

For handling missing data in categorical time series, we pick up the concept of amplitude modulation proposed by Parzen.\(^8\) The amplitude modulation is not applied to the original qualitative process \((X_t)\), but to the binarization derived thereof. Our approach is first explained in detail for the case of ordinal time series and the corresponding binarization \((Z_t)\). Later in Section 5, we show how to adapt it to nominal time series and \((Y_t)\). Let the amplitude-modulating process \((O_t)\) be defined as \(O_t = 1\) if \(X_t\) is observed, and \(O_t = 0\) otherwise. Depending on the considered application context, \((O_t)\) might be deterministic\(^4\) or i.i.d. with \(\pi := P(O_t = 1)\)\(^6\) or, more generally, stationary with some additional dependence structure.\(^7\) We define the amplitude modulation of \((Z_t)\) as \((O_t \cdot Z_t)\), and we set \(\hat{f} := \frac{1}{n} \sum_{t=1}^{n} O_t Z_t\). Then, the mean of \(\hat{f}\) is given by

\[
E[\hat{f}] = \frac{1}{n} \sum_{t=1}^{n} E[O_t Z_t] = \frac{1}{n} \sum_{t=1}^{n} E[Z_t E[O_t | Z_t]].
\]

Thus, the conditional mean \(E[O_t | Z_t]\) is crucial for computing \(E[\hat{f}]\). If \((O_t)\) is assumed to be independent of \((X_t)\) (including the case of \((O_t)\) being deterministic), which is commonly assumed for quantitative time series as well,\(^4,6,10\) then \(E[\hat{f}]\) simplifies to

\[
E[\hat{f}] = \frac{1}{n} \sum_{t=1}^{n} E[O_t] \ E[Z_t] = \left( \frac{1}{n} \sum_{t=1}^{n} E[O_t] \right) \ f_t.
\] \hspace{1cm} (7)

which implies to estimate \(f\) by

\[
\hat{f} := \frac{1}{n} \sum_{t=1}^{n} O_t Z_t = \frac{1}{n} \sum_{t=1}^{n} O_t \]

\hspace{1cm} (8)

If, however, \((O_t)\) would not be independent of \((X_t)\), then the estimation of \(f\) would be more challenging. This is illustrated by the following example.

**Example 2.** Assume that missing data only happen for “large” observations. This might be the case if, for example, people are asked for their level of alcohol consumption, and if they are ashamed to admit heavy consumption. More precisely, assume that for some \(a \in \{0, \ldots, m\}\), we have

\[
P(O_t = 1 | X_t) = \begin{cases} 1 & \text{if } X_t \leq s_a \Leftrightarrow Z_{t,a} = 1 \\ \pi^{(a)} & \text{if } X_t > s_a \Leftrightarrow Z_{t,a} = 0 \end{cases} = Z_{t,a} + (1 - Z_{t,a}) \pi^{(a)}.
\]

Then, \(E[O_t | Z_t] = \pi^{(a)} + (1 - \pi^{(a)}) Z_{t,a}\), and \(E[O_t Z_t] = \pi^{(a)} f_t + (1 - \pi^{(a)}) \ E[Z_{t,a} Z_t].\) Here, the \(i\)th component of \(E[Z_{t,a} Z_t]\) equals \(E[Z_{t,a} Z_t] = f_{\min(i,a)}\). Thus, we obtain

\[
E[\hat{f}] = \frac{1}{n} \sum_{t=1}^{n} E[O_t Z_t] = \pi^{(a)} f_t + (1 - \pi^{(a)}) f_{\min(i,a)},
\]

which is unbiased for \(i \leq a\), but biased for \(i > a\).

In view of the difficulties illustrated by Example 2 and in accordance with common practice,\(^4,6,10\) for the rest of the article, we assume that \((O_t)\) is independent of \((X_t)\).
3.2 Central limit theorems

Under the assumption of \((O_t)\) being independent of \((X_t)\), let us study the stochastic properties of the corrected CDF estimator \(\hat{f}^* = \hat{f}/\hat{O}\) from (8). If missingness does not happen randomly but according to some deterministic principle, that is, if \(O_t = o_t\) for given binary numbers \(o_t\), then \(E \hat{f} = \hat{O} f\) according to (7). Hence, \(\hat{f}^* = \hat{f}/\hat{O}\) is an exactly unbiased estimator of \(f\). More precisely, \(\hat{f}^* = \sum_{i=1}^n o_i Z_i\) is just a weighted mean of the \(Z_i\), where \(n_{\text{obs}} = \sum_{i=1}^n o_i\) is the number of available observations. Therefore, second-order moments are computed exactly as

\[
\text{Cov}[\hat{f}^*, \hat{f}^*] = \frac{1}{n_{\text{obs}}} \sum_{i=1}^n o_i \text{Cov}[Z_{i,i}, Z_{i,j}] = \frac{1}{n_{\text{obs}}} (f_{\text{min}(i,j)} - f_{ij}) + \frac{1}{n_{\text{obs}}} \sum_{s<t} o_{s} (f_{ij}(t-s) + f_{ij}(t-s) - 2f_{ij}),
\]

see Lemma A.5.1 in Weiß. As an example, if \((X_t)\) would even be i.i.d. (i.e., if \(f_{ij}(h) = f_{ij}\) for \(h > 0\)), then \(\text{Cov}[\hat{f}^*, \hat{f}^*] = \frac{1}{n_{\text{obs}}} (f_{\text{min}(i,j)} - f_{ij})\). Thus, the exact position of the missing data is not relevant in this case, just the “effective sample size” \(n_{\text{obs}}\).

If \((O_t)\) is stochastic, by contrast, and stationary with \(E[O_t] = \pi\), then \(E \hat{f} = \pi f\) such that \(\hat{f}^*\) from (8) is generally biased. Let us study this bias (and the whole asymptotic distribution) of \(\hat{f}^*\) in some more detail. For the rest of Section 3.2, assume that \((O_t)\) is stationary with \(E[O_t] = \pi\), and denote its autocovariance function by \(\gamma_0(h) := \text{Cov}[O_{s}, O_{s+h}]\), so \(\pi(h) := E[O_h O_0] = \gamma_0(h) + h^2\). Define the extended binary vectors

\[
Z^*_i := (O_t, Z_{i,0}, …, Z_{i,m-1})^\top = O_t (1, Z^*_i)^\top,
\]

which form a stationary process \((Z^*_i)\) by the stationarity of \((X_t)\) and \((O_t)\). We number the components of \(Z^*_i\) as \(Z^*_{i-1}, Z^*_{i}, …, Z^*_{i,m-1}\). The mean of \(Z^*_i\) is

\[
E[Z^*_i] = \pi (1, f^*)^\top =: \mu_{Z^*}.
\]

Finally, let us assume that an appropriate mixing assumption on \((X_t)\) and \((O_t)\) is satisfied such as \(\alpha\)-mixing with exponentially decreasing weights, which then passes to \((Z^*_i)\). Then, we have the CLT

\[
\sqrt{n} \left( \bar{Z}_i - \mu_{Z^*} \right) \xrightarrow{d} N(0, \Sigma) \quad \text{with} \quad \Sigma = \sigma_{ij} \delta_{ij}, \quad \text{where}
\]

\[
\sigma_{ij} = \text{Cov}[Z^*_{0,i}, Z^*_{0,j}] + \sum_{h=1}^\infty \left( \text{Cov}[Z^*_{0,i}, Z^*_{h,j}] + \text{Cov}[Z^*_{h,i}, Z^*_{0,j}] \right).
\]

Note that the components of the sample mean \(\bar{Z}_i = (\bar{O}_i, \hat{f}^*)^\top\) are required for calculating \(\hat{f}^*\) according to (8). The covariances \(\sigma_{ij}\) in (11) compute as

\[
\sigma_{ij} = \begin{cases} 
\pi(1 - \pi) + 2 \sum_{h=1}^\infty \gamma_0(h) & \text{if } i = j = -1, \\
\pi(1 - \pi) + 2 \sum_{h=1}^\infty \gamma_0(h) & \text{if } i > j = -1, \\
f_{ij} \sigma_{-1,-1} + \pi (f_{\text{min}(i,j)} - f_{ij}) + \sum_{h=1}^\infty \gamma_0(h) (f_{ij}(h) + f_{ij}(h) - 2f_{ij}) & \text{if } i, j > -1, 
\end{cases}
\]

see Supplement S.1 for the proof.

Let us define the function \(g : [0; 1]^{m+1} \rightarrow [0; 1]^m\) by \(g(x_{-1}, x_0, …, x_{m-1}) = x_j/x_{-1}\) for \(j = 0, …, m - 1\), then \(\hat{f}^* = g(\bar{Z}_i)\) and \(f = g(\mu_{Z^*})\). Let \(D\) denote the Jacobian of \(g\) evaluated in \(\mu_{Z^*} = \pi (1, f^*)^\top\). Then, the linear Taylor approximation \(\hat{f}^* \approx f + D (\bar{Z}_i - \mu_{Z^*})\) together with the CLT (11) implies the following result (Delta method).
Theorem 1. Let the amplitude-modulating binary process \( O_t \) be stationary with moments \( E[O_0] = \pi > 0 \) and \( E[O_h O_0] = \pi(h) > 0 \). Let the stationary ordinal process \( X_t \) be independent of \( O_t \), where \( X_t \) is observed if \( O_t = 1 \). Let \( X_t \) and \( O_t \) satisfy the mixing assumption required for (11). Then, the corrected CDF-estimator \( \hat{f}^* \) according to (8) satisfies

\[
\sqrt{n} \left( \hat{f}^* - f \right) \xrightarrow{d} N(0, \Sigma^*), \quad \text{with} \quad \Sigma^* = (\sigma^*_{ij})_{i,j=0,\ldots,m-1}, \quad \text{where} \quad \sigma^*_{ij} = \frac{1}{\pi} \left( f_{\text{min}(ij)} - f_{ij} \right) + \frac{1}{\pi^2} \sum_{k=1}^{\infty} \pi(h) \left( f_{ij} + f_{i} - 2 f_{ij} \right).
\]

Furthermore, the bias \( E[\hat{f}^*_j] - f_j \) is of order \( o(n^{-1}) \).

The proof of Theorem 1 is given in Supplement S.1. For \( \pi = 1 = \pi(h) \) (no missing data), (13) agrees with the CLT (20) in Weiß.14 Also note the analogous structure of (9) and (13). Generally, (13) can be used to derive the asymptotics of statistics based on \( \hat{f}^* \), which is illustrated in Section 3.3. Note that (12) and (13) are related to each other by \( \sigma_{ij} = f_{ij} \sigma_{-1,-1} + \pi^2 \sigma^*_{ij} \) for \( i, j \geq 0 \).

Example 3. Consider the case where the ordinal data \( (X_t) \) are i.i.d. Then, \( f_j(h) = f_{ij} \) for \( h > 0 \) such that (13) simplifies to \( \sigma^*_{ij} = \frac{1}{\pi} \left( f_{\text{min}(ij)} - f_{ij} \right) \). Thus, the dependence structure of \( O_t \) is without effect in this case, only the expected rate of observations, \( \pi \), matters. The variances of \( \hat{f} \) are proportional to \( 1/(n \pi) \) (instead of solely \( 1/n \) if no data were missing), where \( n \pi \) is the expected number of available observations.

By contrast, if \( (O_t) \) is i.i.d. but \( (X_t) \) exhibits serial dependence, then \( \pi(h) = \gamma_0(h) + \pi^2 = \pi^2 \) and \( \frac{1}{\pi^2} \pi(h) = 1 \) for \( h \geq 1 \). Thus, \( \pi \) cancels out of the second term in (13):

\[
\sigma^*_{ij} = \frac{1}{\pi} \left( f_{\text{min}(ij)} - f_{ij} \right) + \sum_{h=1}^{\infty} \left( f_{ij} + f_{i} - 2 f_{ij} \right).
\]

As a result, the effect of \( (X_t) \)'s dependence structure on \( \sigma^*_{ij} \) does not change in the presence of missing data.

Finally, let us return to the bias result provided by Theorem 1. The linear Taylor approximation \( \hat{f}^*_j \approx f_j + D^{(j)} \left( \overline{Z_t} - \mu_Z \right) \), where \( D^{(j)} \) denotes the \( j \)th row of \( D \), implies that the expansion for the bias \( E[\hat{f}^*_j] - f_j \) is of order \( o(n^{-1/2}) \). To compute the \( n^{-1/2} \)-term, we use the quadratic Taylor approximation \( \hat{f}^*_j \approx f_j + D^{(j)} \left( \overline{Z_t} - \mu_Z \right) + \frac{1}{2} \left( \overline{Z_t} - \mu_Z \right)^\top H^{(j)} \left( \overline{Z_t} - \mu_Z \right) \), where \( H^{(j)} \) is the Hessian of \( g_j \) evaluated in \( \mu_Z \). It turns out, see Supplement S.1, that the only nonzero entries of \( H^{(j)} \) are \( h_{-1,-1}^{(j)} = \frac{2 f_j}{\pi^2} \) and \( h_{-1,j}^{(j)} = \frac{1}{\pi^2} \). This implies the bias approximation

\[
n E \left[ \hat{f}^*_j - f_j \right] \approx \frac{1}{2} h_{-1,-1}^{(j)} \sigma_{-1,-1} + h_{-1,j}^{(j)} \sigma_{-1,j} = 0.
\]

Hence, the bias \( E[\hat{f}^*_j] - f_j \) is of order \( o(n^{-1}) \), as stated in Theorem 1, and thus negligible for practice.

3.3 Marginal properties of ordinal time series

The CLTs (11) and (13) can now be used to derive the asymptotics of statistics computed from \( \hat{f}^* \). This is illustrated by considering the asymptotics of the dispersion measure IOV defined in Section 2, but an analogous derivation would be possible for other ordinal dispersion measures as well. For the case of complete data, the IOV asymptotics are provided by Theorem 7.1.1 in Weiß.14 We now adapt his derivations to the case of incomplete data. If applying the Delta method directly to (13), we use the linear approximation of \( \text{IOV} \) in Theorem 7.1.1 of Weiß14 together with \( \hat{f}^* \), namely

\[
\text{IOV} \approx \text{IOV} + \frac{4}{m} \sum_{i=0}^{m-1} (1 - 2 f_i) (\hat{f}^*_i - f_i).
\]
This linear approximation exists if \( \mathbf{f} \) is not equal to the extreme two-point distribution \( \left( \frac{1}{2}, \ldots, \frac{1}{2} \right)^\top \) (recall Section 2; for \( \mathbf{f} = \left( \frac{1}{2}, \ldots, \frac{1}{2} \right)^\top \), the linear term in the expansion of \( \hat{\text{IOV}} \) would vanish). Then, we conclude that

\[
V \left[ \hat{\text{IOV}} \right] \approx \frac{1}{n} \frac{16}{m^2} \sum_{i,j=0}^{m-1} (1 - 2f_i)(1 - 2f_j) \sigma_{ij}^*.
\]

Similarly, the mean of \( \hat{\text{IOV}} \) is computed approximately as

\[
E \left[ \hat{\text{IOV}} \right] = \frac{4}{m} \sum_{i=0}^{m-1} \left( E \left[ \hat{f}_i \right] - E \left[ \hat{f}_i^* \right] \right) - V \left[ \hat{f}_i \right] \approx \text{IOV} - \frac{1}{n} \frac{4}{m} \sum_{i=0}^{m-1} \sigma_{ii}^* \approx \text{IOV} \left( 1 - \frac{1}{n} \left( \frac{1}{\pi} + \frac{2}{\pi^2} \sum_{h=1}^\infty \pi(h) \kappa_{\text{ord}}(h) \right) \right).
\]

Thus, we have a negative bias in estimating IOV, also see Weiß.\(^{14}\) Note that exactly the same results are obtained if using (11) as the starting point, which is demonstrated in Supplement S.2. If being concerned with i.i.d. ordinal data, that is, if \( \kappa_{\text{ord}}(h) = 0 \), then we simply obtain \( -\frac{1}{n} \text{IOV} \) for the bias.

In the same way, the asymptotics of further statistics regarding the marginal distribution of ordinal time series could be obtained, for example, that of alternative measures of dispersion, or of measures about its shape. As an example, a simple measure of ordinal skewness is given by skew = \( \frac{1}{m} \sum_{i=0}^{m-1} f_i - 1 \), see Klein and Doll.\(^{16}\) Weiß.\(^{14}\) It is estimated by replacing the \( f_i \) by \( \hat{f}_i^* \), thus leading to the mean

\[
E \left[ \hat{\text{skew}} \right] = \frac{2}{m} \sum_{i=0}^{m-1} E \left[ \hat{f}_i^* \right] - 1 \approx \text{skew} + o(n^{-1}),
\]

and to the asymptotic variance

\[
V \left[ \hat{\text{skew}} \right] = \frac{4}{m^2} \sum_{i,j=0}^{m-1} \text{Cov} \left[ \hat{f}_i^*, \hat{f}_j^* \right] \approx \frac{1}{n} \frac{4}{m^2} \sum_{i,j=0}^{m-1} \sigma_{ij}^*.
\]

Let us summarize these results in the following theorem.

**Theorem 2.** Let the assumptions of Theorem 1 hold. Then, the sample IOV and skew, computed based on \( \hat{\mathbf{f}}^* \), are asymptotically normally distributed with

\[
E \left[ \hat{\text{IOV}} \right] \approx \text{IOV} \left( 1 - \frac{1}{n} \left( \frac{1}{\pi} + \frac{2}{\pi^2} \sum_{h=1}^\infty \pi(h) \kappa_{\text{ord}}(h) \right) \right),
\]

\[
V \left[ \hat{\text{IOV}} \right] \approx \frac{1}{n} \frac{16}{m^2} \sum_{i,j=0}^{m-1} (1 - 2f_i)(1 - 2f_j) \sigma_{ij}^*,
\]

and

\[
E \left[ \hat{\text{skew}} \right] \approx \text{skew} + o(n^{-1}),
\]

\[
V \left[ \hat{\text{skew}} \right] \approx \frac{1}{n} \frac{4}{m^2} \sum_{i,j=0}^{m-1} \sigma_{ij}^*,
\]

respectively.

### 3.4 Extension: Measuring serial dependence

The CLTs (11) and (13) for \( \hat{\mathbf{f}}^* \) are used for statistics referring to the marginal distribution of \( (X_t) \), recall Section 3.3. If, however, we want to compute a measure of serial dependence in the presence of missing data, such as \( \kappa_{\text{ord}}(h) \) defined in
Section 2, then the concept of amplitude modulation and the corresponding CLTs have to be extended. Like in (8), we use \( \hat{f}^* = \hat{f} / \hat{O} \) as an estimator for the marginal distribution \( f \). Then, we redefine the ordinal Cohen’s \( \kappa \) as

\[
\hat{\kappa}_{\text{ord}}(h) = \frac{\sum_{j=0}^{m-1} (\hat{f}^*_j(h) - \hat{f}^*_j)^2}{\sum_{j=0}^{m-1} \hat{f}^*_j(1 - \hat{f}^*_j)} \quad \text{for } h \in \mathbb{N}.
\] (15)

Here, \( \hat{f}^*_j(h) \) denotes an estimator of \( f^*_j(h) \) that is defined as follows. In analogy to Parzen,\(^8\) Dunsmuir and Robinson,\(^9\) we first consider the bivariate cumulative relative frequency \( \hat{f}_{ij}(h) \) of the amplitude-modulated binarization \( (O_t, Z_t) \), defined by

\[
\hat{f}_{ij}(h) = \frac{1}{n - h} \sum_{t=h+1}^{n} O_t O_{t-h} Z_{t,i} Z_{t-h,j}.
\]

By the independence of \( (O_t) \) to \( (X_t) \), its mean equals

\[
E \left[ \hat{f}_{ij}(h) \right] = \frac{1}{n - h} \sum_{t=h+1}^{n} E[O_t O_{t-h}] E[Z_{t,i} Z_{t-h,j}] = \left( \frac{1}{n - h} \sum_{t=h+1}^{n} E[O_t O_{t-h}] \right) f_{ij}(h),
\]

which implies to define the estimator

\[
\hat{f}^*_j(h) = \frac{1}{n - h} \sum_{t=h+1}^{n} O_t O_{t-h} Z_{t,i} Z_{t-h,j} =: \hat{f}_{ij}(h) / O_t O_{t-h},
\] (16)

in analogy to (8). If \( (O_t) \) is deterministic, this leads again to an exactly unbiased estimator of \( f_{ij}(h) \) (again a type of weighted mean). By contrast, if \( (O_t) \) is stochastic, namely stationary with autocovariance function \( \gamma_0(h) \) and \( \pi(h) = \gamma_0(h) + \pi^2 \), then \( E \left[ \hat{f}_{ij}(h) \right] = \pi(h) f_{ij}(h) \) such that \( \hat{f}^*_j(h) \) from (16) is generally biased.

To derive the asymptotics of the \( \hat{f}^*_j(h) \) and, in particular, of \( \hat{\kappa}_{\text{ord}}(h) \) from (15), we have to extend our derivations from Section 3.2 in analogy to the approach in Supplements A.5 and A.7 in Weiβ.\(^14\) Therefore, we define the \( 2(m+1) \)-dimensional binary vectors

\[
Z^{(h)*}_t = \left( \begin{array}{c}
Z^{(h)*}_{t,-2,}\ldots, Z^{(h)*}_{t,-1,}, Z^{(h)*}_{t,i,j}, \ldots, Z^{(h)*}_{t,m+j,}\ldots,
\end{array} \right)^T,
\]

where the components numbered by \(-1, 0, \ldots, m - 1\) agree with those of \( Z^*_t \) in (10), and where the additional components at positions \( -2 \) and \( m, \ldots, 2m - 1 \) refer to the relevant interactions at times \( t \) and \( t - h \). Note that we only included the “\( jk \)” pairs \( Z_{i,j} Z_{t-h,j} \) in \( Z^{(h)*}_t \), because \( \hat{\kappa}_{\text{ord}}(h) \) in (15) only makes use of the “agreement frequencies” \( \hat{f}^*_j(h) \). If also the asymptotics of \( \hat{f}^*_j(h) \) with \( i \neq j \) would be required, then \( Z^{(h)*}_t \) would have to be extended by further components \( O_t O_{t-h} Z_{t,i} Z_{t-h,j} \).

To keep it simple, we continue with the vectors \( Z^{(h)*}_t \) defined by (17). These allow us to conclude on the joint asymptotics of \( O_t O_{t-h}, \hat{O}, \hat{f}_{ij}(h) \) in analogy to Section 3.2. Then, again like in Section 3.2, we derive the joint asymptotic distribution of \( \hat{f}^*_j(h), \hat{f}^*_j(h), \) and finally that of \( \hat{\kappa}_{\text{ord}}(h) \). The details are postponed to Supplement S.3. As the main result from a practical perspective, we observe the following theorem, which shall later be applied to test for significant serial dependence in \( (X_t) \).

**Theorem 3.** Let the assumptions of Theorem 1 hold, and denote
\[ \pi(h_1, \ldots, h_r) := E \left[ O_0 \cdot O_{h_1} \cdots O_{h_r} \right] > 0 \quad \text{with} \quad r \geq 1 \quad \text{and} \quad 0 < h_1 < \ldots < h_r. \]

Under the null hypothesis of \((X_t)\) being i.i.d., the distribution of \(\hat{c}_{\text{ord}}(h)\) at lag \(h \in \mathbb{N}\), defined by (15), is approximately normal with mean and variance given by

\[
\frac{-1}{n \pi} \quad \text{and} \quad \frac{1}{n} \cdot \frac{\sum_{i,j=0}^{m-1} (f_{\text{min}(i,j)} - f_{j}) (f_{\text{min}(i,j)} - f_{j}) + 2 (1 + h \pi^2) \frac{\pi^2}{\pi(h)} - 2 \frac{\pi^2}{\pi^3} f_{j})}{\pi(h) \left( \sum_{k=0}^{m-1} f_k(1-f_k) \right)^2}, \quad \text{respectively.}
\]

Theorem 3 shows that the (negative) bias of \(\hat{c}_{\text{ord}}(h)\) is not influenced by the dependence structure of \((O_t)\), it is just inversely proportional to the expected number of available observations, \(n \pi\). The approximate variance, by contrast, depends on joint second- and third-order moments of \((O_t)\), namely on \(\pi(h)\) and \(\pi(h, 2h)\). If the data are fully observed, then \(\pi = \pi(h) = \pi(h, 2h) = 1\), and Theorem 3 simplifies to Theorem 7.2.1 in Weiß.\(^{14}\) If, by contrast, data are missing with \((O_t)\) being i.i.d., so \(\pi(h) = \pi^2\) and \(\pi(h, 2h) = \pi^3\), then the variance in Theorem 3 simplifies to

\[
V \left[ \hat{c}_{\text{ord}}(h) \right] \approx \frac{1}{n} \cdot \frac{1}{\pi^2} \cdot \frac{\sum_{i,j=0}^{m-1} (f_{\text{min}(i,j)} - f_{j})^2}{\left( \sum_{k=0}^{m-1} f_k(1-f_k) \right)^2} + \frac{2}{n} \frac{1-\pi}{\pi^2} \cdot \frac{\sum_{i,j=0}^{m-1} f_{j} \cdot (f_{\text{min}(i,j)} - f_{j})}{\left( \sum_{k=0}^{m-1} f_k(1-f_k) \right)^2}.
\]

But still, the variance depends in a more complex way on the missing data than the mean: the “no-missings variance” (first summand) is inflated by the factor \(1/\pi^2\), and we have an additional summand being proportional to \((1-\pi)/\pi^2\). This kind of dependence on \(\pi\) is more complex than the weaker \(1/\pi\)-dependence in the case of the marginal statistics discussed in Example 3.

Theorem 3 and Equation (18) can be applied to test for significant serial dependence (on some level \(\alpha\)) in the presence of missing data. For this purpose, we have to plug-in estimated probabilities into the formula for the approximate variance, leading to the value \(\hat{\sigma}_\alpha^2\). Then, the critical values are given by \(-1/(n \hat{\pi}) \pm z_{\alpha/2} \hat{\sigma}_\alpha\), where \(z_{\alpha/2}\) denotes the \((1-\alpha/2)\)-quantile of the standard normal distribution.

**Example 4.** An important special case for practice is that of a binary process \((X_t)\), i.e., where \(m = 1\). For simplicity, one usually assumes the range to be coded as \(S = \{0, 1\}\), with marginal distribution given by \(p_1 = p\) and \(p_0 = f_0 = 1-p\). Then, \(c_{\text{ord}}(h)\) and \(c_{\text{nom}}(h)\) are identical, and they also agree with the ACF (the same holds for the respective sample counterparts). The variance in Theorem 3 simplifies to

\[
V \left[ \hat{c}_{\text{ord}}(h) \right] \approx \frac{1}{n} \cdot \frac{p + 2 \left( 1 + \frac{\pi(h, 2h)}{\pi(h)} - 2 \frac{\pi^2}{\pi^3} \right) (1-p)}{\pi(h) \cdot p}.
\]

It is equal to \(1/n\) if no data are missing, which are the well-known ACF asymptotics. For \((O_t)\) being i.i.d., we get

\[
V \left[ \hat{c}_{\text{ord}}(h) \right] \approx \frac{1}{n} \cdot \frac{1}{\pi^2} \cdot \frac{2}{n} \cdot \frac{1-\pi}{\pi^2} \cdot \frac{1-p}{p}.
\]

These results can be applied to test for significant dependence (autocorrelation) in binary time series with missing data.

### 4 | EMPIRICAL ILLUSTRATIONS

#### 4.1 | Results from a simulation study

To check the finite-sample performance of the asymptotic results derived in Section 3, a simulation study with 10,000 replications per scenario was done. The ordinal time series were generated according to the rank count approach described in Weiß,\(^{14}\) that is, a process \((I_t)\) of bounded counts with range \(\{0, \ldots, m\}\) was generated, and the ordinal data \((X_t)\) with range \(S = \{s_0, \ldots, s_m\}\) was obtained by setting \(X_t = s_{I_t}\). For the rank counts, a first-order binomial autoregressive (BinAR(1)) model is assumed, see Section 3.3 in Weiß\(^2\) for details. The BinAR(1) model with mean parameter \(p \in (0; 1)\) and dependence parameter \(r \in [0; 1]\) has a binomial marginal distribution, namely \(\text{Bin}(m, p)\), and the lag-\(h\) ACF \(r^h\),...
where \( r = 0 \) leads to i.i.d. data. Inspired by the migraine data (recall Example 1), we consider the parametrizations \((m, p, r) = (3, 0.20, 0.35)\) and \((10, 0.45, 0.50)\). In addition, we also use \( r = 0 \) within the aforementioned parametrizations as examples of i.i.d.-simulations. The sample sizes are \( n \in \{50, 100, 250, 500, 1000\} \), and the probability \( \pi \) of the i.i.d. amplitude-modulating process \( (O_t) \) is chosen as \( \pi \in \{0.95, 0.9, 0.85, 0.8, 0.75, 0.5\} \). For each of these scenarios, the CDF, IOV, skew, and \( \hat{\kappa}_{\text{ord}}(h) \) are computed, and we compare the sample properties obtained from the simulated data with the asymptotic properties computed according to Section 3. The full simulation and asymptotic results are offered as supplementary material. In what follows, we discuss a few illustrative graphs created thereof.

Figure 3 shows diverse error plots, where the central dot corresponds to the mean, and where the error bars are drawn as the standard deviation (SD). The first row of Figure 3 shows the (asymptotic vs simulated) mean \( \pm \) SD of the dispersion measure \( \hat{\text{IOV}} \), recall Section 3.3. To save some space, only two levels of missingness are considered, namely \( \pi = 1 \) (complete data) and \( \pi = 0.75 \) (25\% missing data), and these are represented by different dots as explained in the legend. If we compare the filled dots (\( \pi = 1 \)) to the incomplete dots (\( \pi = 0.75 \)), we recognize a slight effect of the missing data. The negative bias of \( \hat{\text{IOV}} \) becomes stronger with decreasing \( \pi \) (the true IOV values are 0.468 and 0.351, respectively), and the SDs increase in any scenario. But the crucial point is that the asymptotics well capture these changes. Only for \( n = 50 \), we see a modest deviation between asymptotic and simulated SD for \( \hat{\text{IOV}} \). Otherwise, we have a rather close agreement. The conclusions for the skewness measure \( \hat{\text{skew}} \) in the second row of Figure 3 are essentially the same, except that there is virtually no bias. The true skew values (=asymptotic means) are 0.6 and 0.1, respectively, and also for the simulated means of \( \hat{\text{skew}} \), there is hardly any effect of \( n \) and \( \pi \). Altogether, the derived asymptotics are well suited to approximate the true finite-sample distribution of \( \hat{\text{IOV}} \) and \( \hat{\text{skew}} \).

Next, let us turn to the dependence measure \( \hat{\kappa}_{\text{ord}}(h) \), recall Section 3.4 for the asymptotics. The first row of Figure 4 shows the (asymptotic vs simulated) mean \( \pm \) SD of \( \hat{\kappa}_{\text{ord}}(1) \) for i.i.d. ordinal data (tabulated results are provided by the supplementary material). We see a rather good agreement between the asymptotic approximation and the actual true sampling distribution except for the small-sample scenarios \( n = 50, 100 \). We also recognize that decreasing \( \pi \) again intensifies the negative bias and, in particular, the SDs severely increase (more than twice the SD if 25\% missing data). Next, we apply the dependence test (5\%-level) described before Example 4. For this purpose, we estimate \( \pi \) by \( \hat{\pi} = \bar{O} \) and \( f \) by...
and these are plugged-in into the asymptotics (18). Then, the dependence test is executed by comparing $\hat{\kappa}_{ord}(1)$ to the critical values $-1/(n\hat{\sigma}) \pm z_{0.975} \hat{\sigma}_\kappa$. Counting across all replications, we get the simulated rejection rates plotted in the second row of Figure 4. These rejection rates express the size in the i.i.d.-case $r = 0$, and the power in the dependence case $r > 0$. It becomes clear that the simulated sizes are reasonably close to the nominal level 0.05 for $n \geq 100$, irrespective of the actual extent of missingness. However, we note a strong effect of missing data on the power of the test. Especially for the small sample sizes $n = 50, 100$, we get a severe deterioration in power if 25% of the data are missing. For example, if $n = 100$ and 25% are randomly missing (so the “effective sample size” is 75), we have a worse power than if $n = 50$ without missing data. This is caused by the strong increase of the SD for decreasing $\pi$, and is also confirmed by asymptotic computations using (18): For $(m, p, r) = (3, 0.20, 0)$, we have the asymptotic SD $\approx 0.112$ if $(n, \pi) = (50, 1)$, but the increased value $\approx 0.163$ if $(n, \pi) = (100, 0.75)$; for $(m, p, r) = (10, 0.45, 0)$, the asymptotic SDs are $\approx 0.085$ and $\approx 0.122$, respectively. From an intuitive point of view, it is better to estimate a dependence measure such as $\hat{\kappa}_{ord}(h)$ from a short but connected categorical sequence than from a long but disconnected one. Thus, missing observations have a much stronger effect on serial than on marginal properties.

**4.2 Data application: Migraine patients**

Let us return to the migraine data introduced in Example 1. Using the amplitude modulation approach developed in Section 3, we estimated the marginal CDF via (8) as well as corresponding measures of location, dispersion, and skewness. The obtained results are summarized in Figure 5 for the peak-severity series, and in Figure 6 for the stress series (we plot the PMF instead of the CDF as this is easier to interpret). The stress series in Figure 6 has a nearly symmetric marginal distribution with median 5 (and the outermost categories are never observed), so patient B experiences medium stress levels throughout the observation period. The peak-severity distribution in Figure 5, by contrast, is clearly skewed to the right with median category “none.” Nevertheless, patient A had an at least mild peak on more than 40% of all days.
Before further analyzing the serial dependence structure of both time series, and before applying the asymptotics derived before, the pattern of missingness needs to be analyzed in some more detail. For the migraine data, there is no reason to assume the amplitude-modulating process to be deterministic. The binary amplitude-modulating time series \( o_1, \ldots, o_n \) for Figures 1 and 2 take the value 0 at the dotted lines and 1 otherwise. Thus, the probability \( \pi \) of getting an observation at time \( t \) is estimated as \( \hat{\pi} = 206/225 \approx 0.916 \) (Figure 1) and \( \hat{\pi} = 133/136 \approx 0.978 \) (Figure 2), respectively. We analyze for possible serial dependence within \( (o_t) \) by using the ordinary sample ACF (also recall Example 4). In both cases, the obtained ACF values are not significant at the 5%-level: \( 0.080, -0.036, 0.026, \ldots \) with approximate standard error \( 1/\sqrt{n} \approx 0.067 \) for Figure 1, and \( -0.023, \ldots \) with \( 1/\sqrt{n} \approx 0.086 \) for Figure 2. Thus, there is no contradiction to assume \( (O_t) \) being i.i.d. in both examples. Hence, we can test for the serial dependence of the ordinal peak-severity and stress series by using the simplified expression (18) for the approximate variance of \( \hat{\kappa}_{\text{ord}}(h) \).

The results obtained for the \( \hat{\kappa}_{\text{ord}}(h) \)-test (approximate critical values at 5% level) are also summarized in Figures 5 and 6. In both cases, we have significantly positive but decreasing values at lags 1 and 2 (0.341, 0.213, … for peak severity, 0.392, 0.203, … for stress). This information could be used for model fitting, and it implies that an AR(1)-like model for the respective rank counts (in analogy to the simulations of Section 4.1) might give a feasible solution. Furthermore, the clearly positive lag-1 values indicate some tendency for inertia, in the sense that the peak or stress level, respectively, of the previous day continues in a similar way also at the present day. While these information would be relevant for the considered patients, we conclude by discussing another question: What would have happened if we would have ignored the missing data? As an experiment (not recommended for imitation in practice), we just skip the missing observations, leading to complete time series of reduced lengths \( n = 206 \) (peak severity) and \( n = 133 \) (stress), respectively. If using these reduced time series for computing marginal statistics, we obtain the same estimates as before, which is clear from Equation (8). But both the estimates and the critical values for \( \hat{\kappa}_{\text{ord}}(h) \) are affected. For the peak-severity series, where the number of missing (and now ignored) values was rather large (19 out of 225), the \( \hat{\kappa}_{\text{ord}}(h) \)-values change from 0.341, 0.213, … to 0.258, 0.138, … (and the approximate standard error from 0.072 to 0.055). Thus, ignoring the missing data, we would arrive at much lower dependence values. And even for the stress series, where only three out of 136 observations are missing, there is a visible effect: \( \hat{\kappa}_{\text{ord}}(h) \) changes from 0.392, 0.203, … to 0.370, 0.195, …, and the approximate standard error from 0.060 to 0.057. Hence, it becomes clear that missing observations in ordinal time series should not just be ignored but carefully considered for time series analysis.
5 | NOMINAL TIME SERIES WITH MISSING DATA

Although we focused on ordinal time series up to now (motivated by the real application about migraine patients), the problem of missing data is certainly also relevant for nominal time series. But thanks to the unified framework for handling missingness, the necessary derivations for the nominal case are rather similar to those of the ordinal case in Section 3. In what follows, we briefly discuss the necessary modifications of Section 3 as well as the obtained results, while the detailed derivations are postponed to Supplement S.4.

In this section, \( (X_t) \) is a stationary nominal process (recall the notations from Section 2), and we consider the amplitude modulation \( (O_t \cdot Y_t) \) of its binarization \( (Y_t) \), where \( (O_t) \) is independent of \( (X_t) \). Similar to Section 3.1, we define the estimator \( \hat{\mathbf{p}} := \frac{1}{n} \sum_{t=1}^{n} O_t \mathbf{y}_t \), which has the mean \( E[\mathbf{p}] = (\frac{1}{n} \sum_{t=1}^{n} E[O_t]) \mathbf{p} \) in analogy to (7). Thus, as the nominal counterpart to (8), we estimate \( \mathbf{p} \) by

\[
\hat{\mathbf{p}}^* := \frac{1}{n} \sum_{t=1}^{n} O_t \mathbf{y}_t =: \hat{\mathbf{p}} / \bar{O}.
\]  

(19)

The case of \( (O_t) \) being deterministic is discussed in (S.2) in Supplement S.4. Then, \( \hat{\mathbf{p}}^* = \sum_{t=1}^{n} \frac{O_t}{\bar{O}} \mathbf{y}_t \) is again a weighted mean of the \( \mathbf{y}_t \), and the cross-covariance computes as \( \text{Cov}[\hat{\mathbf{p}}^*, \hat{\mathbf{p}}_j^*] = 1/n_{\text{obs}} p_j (\delta_{ij} - p_i) + \frac{1}{n_{\text{obs}}} \sum_{t=1}^{n} O_t \mathbf{y}_t (p_j(t) + p_j(t) - 2 p_i p_j) \) in analogy to (9).

If \( (O_t) \) is stationary with \( \pi, \gamma_0(h), \) and \( \pi(h) \) as before, we consider the extended binary vectors

\[
\mathbf{y}_t^* = (\mathbf{y}_{t-1}^*, \mathbf{y}_{t-2}^*, \ldots, \mathbf{y}_{t-n+1}^*)^\top := O_t (1, \mathbf{y}_t^*)
\]  

(20)

with mean \( \mathbf{\mu}_y^* := E[\mathbf{y}_t^*] = \pi (1, \mathbf{p}^\top)^\top \). In analogy to Section 3.2, with the same mixing assumptions as for (11), we establish the CLT

\[
\sqrt{n} (\mathbf{y}_t^* - \mathbf{\mu}_y^*) \overset{d}{\to} \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}^*) \quad \text{with} \quad \mathbf{\Sigma}^* = (\sigma^*_y)_{i,j=-1,\ldots,m-1}, \quad \text{where}
\]

\[
\sigma^*_y = \begin{cases} 
\pi (1 - \pi) + 2 \sum_{h=1}^{\infty} \gamma_0(h) & \text{if } i = j = -1, \\
\pi_i \sigma_{i,-1} & \text{if } i > j = -1, \\
\pi_j \sigma_{-1,j} + \pi_p (\delta_{ij} - p_i) + \sum_{h=1}^{\infty} \pi(h) (p_j(h) + p_j(h) - 2 p_i p_j) & \text{if } i, j > -1.
\end{cases}
\]  

(21)

Using the Delta method, this result is transferred to the asymptotics of the corrected estimator \( \hat{\mathbf{p}}^* \).

**Theorem 4.** Let \( (X_t) \) be a stationary nominal process, but apart from that, let the assumptions of Theorem 1 still hold. Then, the corrected PMF-estimator \( \hat{\mathbf{p}}^* \) according to (19) satisfies

\[
\sqrt{n} (\hat{\mathbf{p}}^* - \mathbf{p}) \overset{d}{\to} \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}^*), \quad \text{with} \quad \mathbf{\Sigma}^* = (\sigma^*_y)_{i,j=0,\ldots,m-1}, \quad \text{where}
\]

\[
\sigma^*_y = \frac{1}{\pi} \pi p_j (\delta_{ij} - p_i) + \frac{1}{\pi} \sum_{h=1}^{\infty} \pi(h) (p_j(h) + p_j(h) - 2 p_i p_j).
\]  

(22)

**Furthermore, the bias E[\hat{\mathbf{p}}^*] - \mathbf{p} is of order o(n^{-1}).**

These results can now be applied to derive the asymptotics of nominal dispersion measures such as the IQV, also see Section 3.3.

**Theorem 5.** Let the assumptions of Theorem 4 hold. Then, the sample IQV, computed based on \( \hat{\mathbf{p}}^* \), is asymptotically normally distributed with

\[
E[1_{\text{IQV}}] \approx \text{IQV} \left( 1 - \frac{1}{n} \left( \frac{1}{\pi} + \frac{2}{\pi^2} \sum_{h=1}^{\infty} \pi(h) \kappa_{\text{nom}}(h) \right) \right), \quad \text{Var}[1_{\text{IQV}}] \approx \frac{n}{n^2} \left( \frac{m+1}{m} \right)^2 \frac{1}{\sum_{i,j=0}^{m} \pi_i p_j \sigma^*_y}.
\]

Finally, it is possible to adapt the derivations of Section 3.4 to obtain the asymptotics of the nominal Cohen’s \( \kappa \).
\[ \hat{k}_{\text{nom}}(h) = \frac{\sum_{j=0}^{m} (\hat{p}_{ij}^*(h) - (\hat{p}_j^*)^2)}{1 - \sum_{i=0}^{m} (\hat{p}_i^*)^2} \quad \text{with time lag } h \in \mathbb{N}. \]  

Here,

\[ \hat{p}_{ij}^*(h) = \frac{\hat{p}_{ij}(h)}{O_t O_{t-h}} : = \frac{1}{n-h} \sum_{t=h+1}^{n} O_t O_{t-h} Y_{t,h} Y_{t-h,j} . \]  

In Supplement S.4, it is first shown that the vector \( \sqrt{n} (\ldots , \hat{p}_{ij}^*(h) - p_{ij}, \ldots , \hat{p}_{ij}^*(h) - p_{ij}(h), \ldots ) \), the entries of which are defined by (19) and (24), is approximately normally distributed with mean 0 and some covariance matrix \( \Sigma^{(h)} \), see the derivations up to (S.7). Then, these asymptotics are used to derive the following counterpart to Theorem 3.

**Theorem 6.** Let the assumptions of Theorem 4 hold, and use the notation \( \pi(h_1, \ldots , h_r) \) of Theorem 3. Under the null hypothesis of \( (X_t) \) being i.i.d., the distribution of \( \hat{k}_{\text{nom}}(h) \) at lag \( h \in \mathbb{N} \), defined by (23), is approximately normal with mean and variance given by

\[ -\frac{1}{n \pi} \quad \text{and} \quad \frac{1}{n} \sum_{j=0}^{m} p_{ij}^2 \left( \delta_{ij} - p_i \right) \left( \delta_{ij} + p_i \left( 1 + 2 \pi(h,2h) \pi(h) - 4 \pi(h) \pi \right) \right), \]

respectively.

To sum up, for nominal time series with missing data, the same battery of results is available as for the ordinal case. The obtained formulae differ only slightly from their ordinal counterparts such that software implementations are very similar for both cases.

**6 | CONCLUSIONS**

Motivated by a project on migraine patients, where the collected time series on qualitative features such as stress or pain peak severity are often incomplete, we considered the task of analyzing categorical time series in the presence of missing data. We proposed a unified framework for handling missingness in categorical time series and for deriving the asymptotics of the relevant statistics. We proved expressions for the asymptotic distribution of the univariate and bivariate sample CDF (PMF) computed from an ordinal (nominal) time series that explicitly account for missing observations. These expressions, in turn, were used to derive the asymptotics of, for example, measures of dispersion or serial dependence in such time series. The finite-sample performance of the resulting asymptotic approximations was investigated by simulations, and the developed approaches were applied to example time series from the migraine project. It became clear that especially the measurement of serial dependence is strongly affected by incomplete data, and simply ignoring the missing observations may cause quite misleading results. Thus, missing observations in categorical time series should be carefully considered when analyzing these time series.

We conclude by briefly outlining a direction for future research that seems relevant for the migraine project. Besides analyzing the time series individually, it would also be relevant to analyze for cross-dependencies between the different features of a patient in time. Thus, one has to deal with missingness again, but now with respect to statistics that are applied to features of different type, for example, ordinal features with different ranges (such as peak severity vs stress) or even features being measured on different scales.

**ACKNOWLEDGEMENTS**

The author thanks the associate editor and the reviewers for carefully reading the article and for their comments, which greatly improved the article. The author is very grateful to Curelator Inc (https://n1-headache.com/) for providing the N1-Headache data that was used for Example 1 and Section 4.2. In particular, he wants to thank Dr Marina Vives-Mestres and Dr Amparo Casanova for their support and for fruitful discussions about these data.

**CONFLICT OF INTEREST**

The author declares no potential conflict of interests.
DATA AVAILABILITY STATEMENT
The relevant R codes for this research can be found in the accompanying GitHub repository at https://github.com/cnweiss/CaTS_Missing.

ORCID
Christian H. Weiß https://orcid.org/0000-0001-8739-6631

REFERENCES
1. Box GEP, Jenkins GM, Reinsel GC, Ljung GM. Time Series Analysis: Forecasting and Control. 5th ed. Hoboken, NJ: John Wiley & Sons, Inc; 2015.
2. Weiß CH. An Introduction to Discrete-Valued Time Series. Chichester, NJ: John Wiley & Sons, Inc; 2018.
3. Rubin DB. Inference and missing data. Biometrika. 1976;63(3):581-592.
4. Neave HR. Spectral analysis of a stationary time series using initially scarce data. Biometrika. 1970;57(1):111-122.
5. Weiß CH, Testik MC. On the Phase I analysis for monitoring time-dependent count processes. IEEE Trans. 2015;47(3):294-306.
6. Scheinok PA. Spectral analysis with randomly missed observations: the binomial case. Ann Math Stat. 1965;36(3):971-977.
7. Bloomfield P. Spectral analysis with randomly missing observations. J Royal Stat Soc B. 1970;32(3):369-380.
8. Parzen E. On spectral analysis with missing observations and amplitude modulation. Sankhyā A. 1963;25(4):383-392.
9. Dunsmuir W, Robinson PM. Asymptotic theory for time series containing missing and amplitude modulated observations. Sankhyā A. 1981;43(3):260-281.
10. Yajima Y, Nishino H. Estimation of the autocorrelation function of a stationary time series with missing observations. Sankhyā A. 1999;61(2):189-207.
11. Vives-Mestres M, Casanova A, Buse DC, et al. Patterns of perceived stress throughout the migraine cycle: a longitudinal cohort study using daily prospective diary data. Headache J Head Face Pain. 2021;61(1):90-102.
12. Vives-Mestres M, Casanova A. Modelling and visualizing two-way contingency tables using compositional data analysis: a case-study on individual self-prediction of migraine days. Stat Med. 2021;40(2):213-225.
13. Weiß CH. Regime-switching discrete ARMA models for categorical time series. Entropy. 2020;22(4):458.
14. Weiß CH. Distance-based analysis of ordinal data and ordinal time series. J Am Stat Assoc. 2020;115(531):1189-1200.
15. Kleiñseth TO. Coefficients of variation for nominal and ordinal categorical data. Perceptual Motor Skills. 1995;80(3):843-847.
16. Klein I, Doll M. Tests on asymmetry for ordered categorical variables. J Appl Stat. 2020.
17. Weiß CH. Serial dependence of NDARMA processes. Comput Stat Data Anal. 2013;68:213-238.
18. Ibragimov I. Some limit theorems for stationary processes. Theory Prob Appl. 1962;7(4):349-382.
19. Billingsley P. Convergence of Probability Measures. 2nd ed. New York, NY: John Wiley & Sons, Inc; 1999.

SUPPORTING INFORMATION
Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Weiß CH. Analyzing categorical time series in the presence of missing observations. Statistics in Medicine. 2021;1–16. https://doi.org/10.1002/sim.9089