Correcting for Serial Dependence in Studies of Respiratory Dynamics

Jen J. Gong, Kin Foon Kevin Wong, Joseph F. Cotten, Ken Solt, Emery N. Brown

Abstract— Understanding the physiological impact of drug treatments on patients is important in assessing their performance and determining possible side effects. While this effect might be best determined in individual subjects, conventional methods assess treatment performance by averaging a physiological measure of interest before and after drug administration for n subjects. Summarizing large numbers of time-series observations in two means for each subject in this way results in significant information loss. Treatment effect can instead be analyzed in individual subjects. Because serial dependence of observations from the same animal must then be considered, methods that assume independence of observations, such as the t-test and z-test, cannot be used. We address this issue in the case of respiratory data collected from anesthetized rats that were injected with a dopamine agonist. In order to accurately assess treatment effect in time-series data, we begin by formulating a method of conditional likelihood maximization to estimate the parameters of a first-order autoregressive (AR) process. We show that treatment effect of a dopamine agonist can be determined while incorporating serial effect into the analysis. In addition, while maximum likelihood estimators of a large sample with independent observations may converge to an asymptotically normal distribution, this result of large sample theory may not hold when observations are serially dependent. In this case, a parametric bootstrap comparison can be used to approximate an appropriate measure of uncertainty.

I. INTRODUCTION

In many respiratory experiments, treatment effect is quantified after taking extensive measurements of variables such as respiratory rate and tidal volume before and after the administration of a drug. The difference between the average value of the respiratory variable before and after treatment for a sample of experimental subjects or animals is used to estimate the average drug effect. Using this approach, a sample with n animals will result in only n data points in the analysis. Because these n data points are the averages of detailed time-series measurements made before and after the drug administration in each animal, this method results in substantial information loss. This suggests that it should be possible to preserve the information content of the original

J.J. Gong is a student at Harvard College jgong@fas.harvard.edu

K.F.K. Wong is with the Neuroscience Statistics Research Laboratory, and Massachusetts General Hospital, Boston, MA 02114 USA wong@nmr.mgh.harvard.edu

J.F. Cotten is with Harvard Medical School and Massachusetts General Hospital, Boston, MA 02114 USA jcotten@partners.org

K. Solt is with Harvard Medical School and Massachusetts General Hospital, Boston, MA 02114 USA ksolt@partners.org

E.N. Brown is with the Neuroscience Statistics Research Laboratory, Harvard Medical School, Massachusetts General Hospital, Boston, MA 02114 USA, and also with the Harvard-Massachusetts Institute of Technology (MIT) Division of Health Science and Technology and the Department of Brain and Cognitive Sciences, MIT, Cambridge, MA 02139 USA enb@neurostat.mit.edu

data by making an inference about the effects of the drug in each animal separately, provided the serial dependence in the recordings of the respiratory variable can be taken into account. This paper presents one approach to addressing serial dependence between observations in constructing a test for treatment effect.

Previous work has developed methods such as three-dimensional phase-space reconstruction to incorporate non-linearity, which is often observed in heart rate variability and respiratory movements, into time-series analysis [1]. However, while these methods can provide insight into the data, they do not necessarily lend themselves well to assessing the statistical significance of the difference in a physiological measure between two particular time intervals. To address this, this paper sets up the framework for a first-order autoregressive process to demonstrate the use of conditional maximum likelihood and bootstrap methods, although the same method can be used for higher order models that may characterize the data more accurately.

The standard method for assessing treatment effect in n animals is to conduct a z-test or t-test for the difference between means. These tests, however, assume independence between observations. Because cardiac and breathing rate observations in the same subject are serially correlated, this assumption of independence is violated (Figs. 1 and 2). Therefore, conventional means of hypothesis testing cannot be used. In addition, construction of 95% confidence intervals for the mean difference using the standard error of the considered data will not accurately represent the uncertainty in the estimate because of the assumption of independent observations.

Modeling time-series data with an autoregressive (AR) process, however, allows us to use conditional likelihood methods to estimate the parameters and account for the serial dependence between observations. Thus, a standard hypothesis test can be modified to test for treatment effect in serially correlated data by replacing the typically used sample mean and standard error with the analogous maximum likelihood estimates and standard deviations of these estimates, as approximated by the inverse of the observed Fisher information matrix [3]. Similar methods have been used in constructing state-space models based on Kalman filtering to analyze optical imaging data, although the full likelihood function is used instead of the conditional one [4]. The observed Fisher information gives a reliable estimate for the variances of the parameters under the assumption that when the sample size is large, the conditional likelihood estimates behave according to the distributions they asymptotically approach as the sample size goes to infinity. These asymptotic distributions
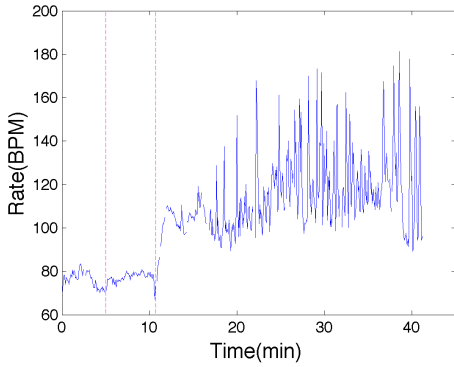
Fig. 1. Breathing rate (BPM) of rat 1. Injections of saline and dopamine agonist marked by left and right dotted lines, respectively.
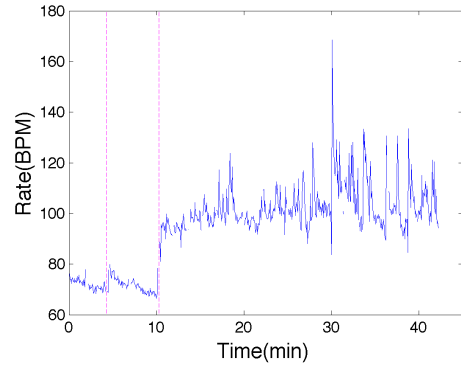

Fig. 2. Breathing rate (BPM) of rat 2. Injections of saline and dopamine agonist marked by left and right dotted lines, respectively.

are Gaussian with variances given by the inverse Fisher information matrix [3]. This theory assumes that the data have a large number of independent observations. In the case of serially correlated data, however, there may be a substantial amount of information loss because the observations depend on each other. As a consequence, the approximate variances of the maximum likelihood estimates from the observed Fisher information matrix may not accurately describe the actual variance in the parameters. To better approximate this uncertainty, parametric bootstrap methods can be used to generate a sample of values for the maximum likelihood estimate of the parameter μ [5].

By addressing these two issues in the context of assessing the difference in mean breathing rate before and after the injection of a dopamine agonist in two rats, we propose a correction for serial dependence on a standard hypothesis test that allows us to maintain the information content of the time-series data for each animal.

II. METHODS

A. Experimental Procedure

Two rats were maintained for 40 minutes under stable general anesthesia using 1.4% isoflurane. Respiratory rate (BPM) and tidal volume were continuously measured. After 40 minutes, a dopamine agonist was injected to assess its effect on respiratory rate.

B. Experimental Data

To investigate whether or not injection of a dopamine agonist resulted in a significant change in mean breathing rate, the BPM data from before the administration of the drug was compared to BPM data from after the administration of the drug for two rats. The raw breathing rate data, as shown in Figures 1 and 2, can be seen to start at a baseline level and transition to a higher level, where it eventually flattens out. Missing data values were removed from all data sets. The conditional log-likelihood is

$$\log\left(L(\theta|x)\right) = -(T-1)\log(2\pi\sigma_\epsilon^2)/2$$
$$-\sum_{t=2}^{T}\left(x_t - \mu - \phi(x_{t-1} - \mu)\right)^2/2\sigma_\epsilon^2. \quad (1)$$

C. Conditional Likelihood Estimation

A stationary AR(1) model with mean μ, $|\phi| < 1$, and white noise ϵ_t follows

$$x_t - \mu = \phi(x_{t-1} - \mu) + \epsilon_t. \quad (2)$$

The white noise ϵ_t is assumed to be independently and identically Gaussian distributed for all t, with mean 0 and variance σ_ϵ^2. We are interested in finding values for the parameters $\theta = (\mu, \phi, \sigma_\epsilon^2)$ that maximize the likelihood of the data.

Rather than maximizing the full likelihood, which is less computationally tractable, we greatly simplify the likelihood calculations by conditioning on the first observation, x_1. If we let $x = (x_1, ..., x_T)$, then we have

$$L(\theta|x) = \prod_{t=2}^{T} f(x_t|x_{t-1}, \theta) \quad (3)$$

Maximizing this with respect to μ, ϕ, and σ_ϵ^2, we find that the maximum likelihood estimates of the three parameters are:

$$\hat{\phi} = \sum_{t=2}^{T}(x_t - \hat{\mu})(x_{t-1} - \hat{\mu})\left/\sum_{t=2}^{T}(x_{t-1} - \hat{\mu})^2\right. \quad (4)$$

$$\hat{\mu} = \sum_{t=2}^{T}(x_t - \hat{\phi}x_{t-1})/\left((T-1)(1-\hat{\phi})\right) \quad (5)$$

$$\hat{\sigma}_\epsilon^2 = \sum_{t=2}^{T}\left(x_t - \hat{\mu} - \hat{\phi}(x_{t-1} - \hat{\mu})\right)^2/(T-1) \quad (6)$$

While the conditional maximum likelihood estimates of $\hat{\mu}$ and $\hat{\phi}$ are independent of $\hat{\sigma}_\epsilon^2$, they are not independent of each other. To find these estimates, we apply a cyclic descent algorithm. These estimates convergence after a few iterations, and the final values are used to calculate $\hat{\sigma}_\epsilon^2$ [6].

D. Variance of estimated parameters

The observed Fisher information matrix is computed from the second derivatives of the conditional log-likelihood with respect to μ, ϕ, and σ_ϵ^2. The variance for each parameter is

obtained from the inverse of this matrix and used to construct 95% confidence intervals for the parameters.

The observed Fisher information matrix is

$$I(\hat{\theta}) = - \begin{bmatrix} \frac{\partial^2 l}{\partial \mu^2} & \frac{\partial^2 l}{\partial \mu \partial \phi} & \frac{\partial^2 l}{\partial \mu \partial \sigma_\epsilon^2} \\ \frac{\partial^2 l}{\partial \phi \partial \mu} & \frac{\partial^2 l}{\partial \phi^2} & \frac{\partial^2 l}{\partial \phi \partial \sigma_\epsilon^2} \\ \frac{\partial^2 l}{\partial \sigma_\epsilon^2 \partial \mu} & \frac{\partial^2 l}{\partial \sigma_\epsilon^2 \partial \phi} & \frac{\partial^2 l}{\partial (\sigma_\epsilon^2)^2} \end{bmatrix} \Bigg|_{\hat{\theta}=(\hat{\mu},\hat{\phi},\hat{\sigma}_\epsilon)} . \quad (7)$$

The variances of the parameter estimates are obtained from the elements of the matrix $[I(\hat{\theta})]^{-1}$.

E. Hypothesis Test

Using these conditional maximum likelihood estimates, we can conduct a hypothesis test of the difference in means between two sets of data. From our AR(1) model, we know that $x_t - \mu - \phi(x_{t-1} - \mu) = \epsilon_t$, where the ϵ_t are independently and identically distributed according to a normal distribution with mean 0 and variance σ_ϵ^2 for all t. For each of the two data sets we are comparing, we find the maximum likelihood estimate of μ. In addition, we can approximate the values of $\sigma_{\mu_1}^2$ and $\sigma_{\mu_2}^2$, the variances of these estimated means, using the observed Fisher information matrix (7). Thus, we can construct the test statistic

$$z = (\hat{\mu}_1 - \hat{\mu}_2)/\sqrt{\hat{\sigma}_{\hat{\mu}_1}^2 + \hat{\sigma}_{\hat{\mu}_2}^2}. \quad (8)$$

To analyze the difference in mean breathing rate before and after the injection of the dopamine agonist, as illustrated in Figures 3 and 4, we conduct a test of the null hypothesis $H_0 : \mu_{\text{before}} = \mu_{\text{after}}$ against the alternative $H_1 : \mu_{\text{before}} \neq \mu_{\text{after}}$ for the two animals.

F. Bootstrap Analysis

From the observed Fisher information matrix, it is easy to show that

$$\sigma_{\hat{\mu}}^2 = \sigma_\epsilon^2 / \big((T-1)(1-\phi)^2\big). \quad (9)$$

From this expression for the variance of the estimated mean, it can be seen that it takes a form analogous to that of the variance of the mean in a sample with independent observations
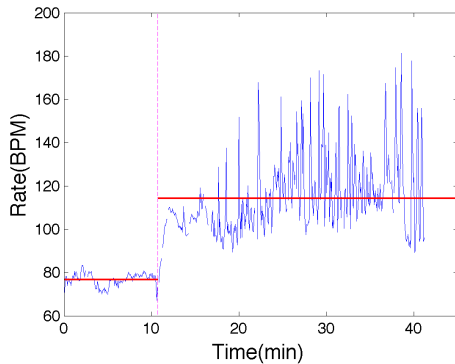
$$\sigma_\mu^2 = \sigma^2/n, \quad (10)$$

where n is the number of observations and σ^2 is the variance of the population. Thus, the denominator of the variance of the mean for serially correlated data has what can be seen to be approximately $(T-1)(1-\phi)^2$ independent observations.

For data for which this denominator is small (when there is large serial correlation between observations), large sample theory may not provide a suitable estimate of the parameter variance. Therefore, this estimate of the variance of $\hat{\mu}$ may not be an accurate measure of the uncertainty and will give an incorrect z-statistic. To find a more accurate estimate of the variance of the estimated mean ($\hat{\sigma}_{\hat{\mu}}$), we use a parametric bootstrap comparison [5]. Using our original data and conditional maximum likelihood estimates derived previously, we simulate multiple bootstrap samples and find the conditional maximum likelihood estimates of the parameters for each of these samples.

The bootstrap algorithm is as follows:

A. For $t = 2, ..., T$
   1. Draw ϵ_t^* from $N(0, \hat{\sigma}_\epsilon^2)$.
   2. Compute $x_t^* = \hat{\mu} + \hat{\phi}(x_{t-1}^* - \hat{\mu}) + \epsilon_t^*$.
B. Given $x^* = (x_1, x_2^*, ..., x_T^*)$, compute $\hat{\theta}^*$ by maximizing $l(\theta|x^*)$.
C. Repeat A and B 200 times.

Using this process, we obtained a set of 200 values of $\hat{\mu}$. Taking the variance of the bootstrap sample gives a bootstrap estimate of $\hat{\sigma}_{\hat{\mu}}$. The values were used to calculate the modified z-statistic for the data from both animals (Table 2).

III. RESULTS

A. Convergence of maximum likelihood estimates

Convergence of conditional maximum likelihood estimation of parameters for all but one of the considered data sets occurred in fewer than 20 iterations. Convergence time for each $(\hat{\mu}, \hat{\phi})$ pair was defined as the number of iterations it took before both parameters reached a value less than 0.00001 from their values in the previous iteration.
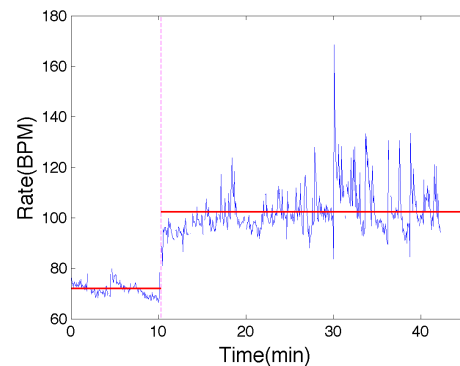

Fig. 3. Comparing mean breathing rate (BPM) before and after injection of dopamine agonist (marked by dotted line) in rat 1

Fig. 4. Comparing mean breathing rate (BPM) before and after injection of dopamine agonist (marked by dotted line) in rat 2

TABLE I
CONDITIONAL MAXIMUM LIKELIHOOD ESTIMATES

Animal		$\hat{\mu}$	$Var(\hat{\mu})$	$\hat{\phi}$	$Var(\hat{\phi})$	$\hat{\sigma}_\epsilon^2$	$Var(\hat{\sigma}_\epsilon^2)$
1	Pre-DA	76.983	0.52007	0.8267	0.00179	2.4394	0.075806
	Post-DA	114.52	3.1179	0.6441	0.0013303	171	135.06
2	Pre-DA	72.231	0.78295	0.7997	0.0044509	4.3426	0.26561
	Post-DA	102.31	0.71939	0.59669	0.0013932	53.826	12.597

TABLE II
COMPARING MEAN BREATHING RATE BEFORE AND AFTER INJECTION OF DOPAMINE AGONIST (DA) WITH

$$H_0 : \mu_{\text{BEFORE}} = \mu_{\text{AFTER}}, H_1 : \mu_{\text{BEFORE}} \neq \mu_{\text{AFTER}}.$$

Animal	1	2
Naive z-statistic	43.667	61.359
Naive p-value	$< 10^{-16}$	$< 10^{-16}$
Modified z-statistic	34.277	45.205
Modified p-value	$< 10^{-16}$	$< 10^{-16}$
Bootstrap z-statistic	19.379	25.116
Bootstrap p-value	$< 10^{-16}$	$< 10^{-16}$

B. Maximum likelihood estimates and standard deviations

Maximum likelihood estimates of the parameters and their variances for the breathing rate data before and after the dopamine agonist injection are shown in Table 1. These estimates were calculated using a cyclic descent algorithm, and the variances were calculated from the observed Fisher information matrix.

C. Hypothesis tests

Hypothesis tests for two animals were conducted to examine the statistical significance of the difference in mean breathing rate before and after the injection of the dopamine agonist. Under the null hypothesis, the means before and after the injection would be the same, and therefore, the difference between the two would equal zero. For each animal, three versions of this hypothesis test were performed. The first, which we call the naive test, was done without accounting for the serial correlation between observations. The second, called the modified test, accounted for serial correlation and used the conditional maximum likelihood estimates and the Fisher information matrix to calculate the z-statistic. The third test also used the conditional maximum likelihood estimates of the parameters, but parametric bootstrap methods were used to estimate the variance of $\hat{\mu}$ instead of the Fisher information matrix. The results of the hypothesis tests are given in Table 1.

Although all of the p-values were found to be significant ($< 10^{-16}$), there is a noticeable difference in the z-statistics for the three methods considered. The lower magnitude of the bootstrap z-statistic compared to the modified one agrees with our stated logic; because there are a large number of correlated observations, large sample theory need not hold. Therefore, the inverse of the Fisher information matrix may not give an accurate measure of the uncertainty in the parameters.

IV. DISCUSSION & CONCLUSIONS

We have provided a two step correction for the problem of conducting a simple hypothesis test of the difference in mean respiratory rate before and after administration of a dopamine agonist. To correct for the serial dependence in each set of measurements, we first fit separate autoregressive models to the BPM data before and after administration of the drug treatment using a cyclic descent procedure to maximize the conditional maximum likelihood of each data set. Second, we used a parametric bootstrap to compute the uncertainty in the parameters before and after the drug treatment to correct for the possible inaccuracies in the parameter variance estimates based on the observed Fisher information. The bootstrap estimates of the parameter variances are used to conduct the hypothesis test comparing the means. The two step procedure provided a correction for the naive hypothesis test which assumes that both sets of data are independent. Our approach makes it possible to perform within subject hypothesis testing instead of having a substantial loss of power because each subject contributes a only a single pair of observations to a group hypothesis test.

Our approach suggests several important extensions. These include use of a higher order autoregressive model to characterize the data better and use of a non-parametric bootstrap procedure to further help compensate for lack of fit. In addition, the non-constant mean respiratory rate can be accounted for. These extensions will be considered when applying these methods to further analyze the respiratory data examined in this study.

V. ACKNOWLEDGMENTS

REFERENCES

[1] K.L. Myers, A.E. Brockwell, W.F. Eddy, "State-space models for optical imaging," *Statistics in Medicine*, vol. 26, no. 21, pp. 3862-3874, 2007.
[2] G.E.P. Box, G.M. Jenkins, and G.C. Reinsel, *Time-Series Analysis: Forecasting and Control*, 4th ed., Wiley and Son, New York, 2008.
[3] Y. Pawitan, *In All Likelihood: Statistical Modelling and Inference Using Likelihoods*, Oxford Press, Oxford, 2001.
[4] D. Hoyer, K. Schmidt, R. Bauer, U. Zweiner, M. Kohler, B. Luthke, and M. Eiselt, "Nonlinear analysis of heart rate and respiratory dynamics," *Engineering in Medicine and Biology Magazine, IEEE*, vol. 16, no. 1, pp. 31-39, 1997.
[5] B. Efron and R. Tibshirani, *An Introduction to the Bootstrap*, Chapman & Hall, 1993.
[6] A.L. Garkavi, "Method of cyclic descent in the problem of best approximation," *Mathematical Notes*, vol. 27, no. 4, pp. 270-274, 1980.