

Multiple hypotheses testing on dependent count data with covariate effects

WEIZHE SU AND XIA WANG*

Dynamics in the sequence of count data is usually not only affected by the underlying hidden states to be detected, but also quite likely associated with other static or dynamically changing covariates. The multiple hypotheses testing procedure developed here takes these covariates into consideration by the Poisson regression model. Also, a hidden Markov process is applied to model the switches between the null and non-null states as well as the dependence across counts. All model parameters are estimated through Bayesian computation. While a simple distribution is assumed on the null state, the observation distribution under the non-null state usually requires more flexibility. Here a mixture of parametric distributions is assumed. The number of mixture components is decided by model selection criteria, including the Bayesian Information Criterion as well as marginal likelihood methods. Simulation studies are carried out to evaluate the performance of the proposed model and that of the model selection methods. The real data example shows the application of the proposed model and its inference goal differs from the previous testing procedures with no covariate effects considered.

AMS 2000 SUBJECT CLASSIFICATIONS: Primary 62M05, 62F03; secondary 62J12.

KEYWORDS AND PHRASES: Bayesian hierarchical model, Markov switch regression model, Multiple hypotheses testing, Poisson regression, Sampling-based marginal likelihood.

1. INTRODUCTION

We often need to detect “abnormalities” based on count data collected sequentially over time, such as the number of emergency room visits with influenza like symptoms (ILS) per week or the number of log-ins on a website per hour. These observed counts contain information on underlying unobservable states of interest, such as the onset of a flu season (Sun and Cai, 2009). The dynamics of count data usually is not only influenced by the underlying hidden states of interest, but also associated with some common covariate factors. Depending on our goal in testing, it is sometimes necessary to model those covariate effects. For example, checking the number of emergency room visits alone can help detect

the onset of a flu season. If the task is to get alarmed for an “abnormal” season (possible concurrence of multiple respiratory diseases) and get prepared for higher demands on public health resources compared to a normal year, the goal is then to detect any abnormality beyond the dynamics of a “normal” flu season.

To accomplish this goal, we develop a multiple hypotheses testing procedure based on a Poisson Markov switching regression model (PReg-HMM), which can account for covariates related to the “normal” dynamics and extract abnormalities beyond those factors. The Markov switching regression model has been abounding in modeling time series data with hidden “regimes” (Goldfeld and Quandt, 1973). The primary goal in most of the previous studies is to identify the correct clusters of the hidden states. For example, Sebastian et al (2019) identified three states of the infectious disease by a Poisson hidden Markov model and studied its association with climate risk factors after clustering. More recently, Xu et al. (2020) studied how baseline subject characteristics and dynamic environmental factors are related to the disease dynamics using data collected from wearable device.

The purpose of the proposed PReg-HMM is more like that of Ihler et al. (2006), where they used the Poisson hidden Markov model with covariate effects to detect the presence of “unusual” events with a binary Markov process. A similar modeling framework can be applied in testing multiple dependent hypotheses. Previous studies have developed optimal and valid multiple testing procedures based on the hidden Markov model (Sun and Cai, 2009; Wang et al., 2019). Su and Wang (2020) proposed a Bayesian multiple hypotheses testing procedure for sequentially dependent count data (PHMM). Here we extended these studies by incorporating the covariate effects.

All the model parameters and the values of the test statistic are estimated through Bayesian computation. A threshold rejection procedure helps to control the false discovery rate (FDR) and ensure the test is valid. To allow for flexibility in the observation distribution under the non-null state, it is assumed that the distribution can be either a simple standard distribution or a mixture distribution. An important modeling decision to make is then the number of mixture components. A few sampling-based marginal likelihood methods and the Bayesian Information Criterion (BIC) are used in selecting the optimal number of mixture components.

*Corresponding author.

The paper is organized as follows. Section 2 describes the multiple hypotheses testing procedure based on the Poisson Markov switching regression models. Model estimation and selection details are described in Section 3. Results from the simulation studies are shown in Section 4 to demonstrate the performance of the proposed PReg-HMM with different types of covariates and to compare the various model selection methods. Section 5 shows the application of PReg-HMM to the people flow count data and compares the results with those from the PHMM procedure (Su and Wang, 2020). Discussions and conclusions are in Section 6.

2. POISSON MARKOV SWITCHING REGRESSION MODEL FOR MULTIPLE HYPOTHESES TESTING

Assume that there are a total of m hypotheses to be tested based on the sequentially observed data (\mathbf{X}, \mathbf{Z}) , where $\mathbf{X} = (X_1, X_2, \dots, X_m)'$ represents the m observed count data. Denote X_t as the t^{th} data point, where t labels the sequence of the data points, such as the time or the ordered position in the sequence, $t = 1, \dots, m$. At each t , the observed r covariate factors recorded together with X_t is denoted as an $(r+1)$ -dimensional vector $\mathbf{z}_t = \{1, z_{t,1}, \dots, z_{t,r}\}$ with the first element for the intercept term in the regression model. Then $\mathbf{Z} = (\mathbf{z}'_1, \mathbf{z}'_2, \dots, \mathbf{z}'_m)'$ is the $m \times (r+1)$ covariate matrix with m as the number of the observations and r as the number of covariates considered. The first column in \mathbf{Z} , $\mathbf{1}_m$, is an $m \times 1$ vector of ones. The goal of hypotheses testing is to detect “normal” and “abnormal” states based on the observed data.

2.1 Multiple hypotheses testing and hidden Markov model

Given the nature of the hypothesis testing problem, it is natural to assume that the regression model of $\{X_t\}_{t=1}^m$ on $\{\mathbf{z}_t\}_{t=1}^m$ depends on the realizations of a hidden discrete stochastic process $\mathbf{S} = \{S_t\}_{t=0}^m$, where S_t indicates the true state of the regression model at time t and takes discrete value in the set $\{0, 1\}$. The value 0 indicates the null state and 1 for the non-null state. Assume this hidden process $\{S_t\}_{t=0}^m$ satisfies the Markov property, which means that the current state S_t only depends on the one-period ahead hidden state S_{t-1} and is independent of all states prior to time $t-1$. The Markov chain starts from its stationary distribution $\boldsymbol{\pi} = (\pi_0, \pi_1)$, which specifies the proportions for null and non-null states. The probability of $(S_0 = k)$ is $P(S_0 = k) = \pi_k$, $\pi_k \geq 0$, $k = 0, 1$. Let $\mathcal{A} = \{a_{jk}\}$ be the transition matrix of the hidden state process \mathbf{S} , where $a_{jk} \geq 0$ and $\sum_{k=0}^1 a_{jk} = 1$, $j, k = \{0, 1\}$. The transition matrix \mathcal{A} represents the transition probability between null and non-null states and the one-dimensional dependence structure is written out as $P(S_t = k | S_{t-1} = j, \mathcal{A}) = a_{jk}$, $j, k = 0, 1$. The diagonal element a_{jj} is the persistence probability and indicates the expected duration of state

j . The observations X_t 's given S_t 's are independently distributed. Denote the conditional distribution of the observations as $p(X_t | S_t = k, \mathbf{z}_t) = f_k(\mathbf{z}_t)$, $k = 0, 1$. Then $\mathcal{F} = \{f_0(\mathbf{z}_t), f_1(\mathbf{z}_t)\}$ represents the observation distributions under the null and non-null states derived from the regression model, where the non-null distribution f_1 can be either a simple standard distribution or a mixture distribution. We denote the parameters in this two-state hidden Markov model (HMM) as $\vartheta = (\mathcal{A}, \boldsymbol{\pi}, \mathcal{F})$.

2.2 Multiple hypotheses testing with Poisson Markov switching regression models

For dependent count data, we assume a Poisson distribution on f_0 and a mixture of Poisson distributions on f_1 with L mixture components ($L = 1$ if f_1 is a single Poisson distribution). Let the ℓ^{th} state specific coefficients $\boldsymbol{\beta}_\ell = (\beta_{0\ell}, \beta_1, \dots, \beta_r)$ be an $(r+1)$ -dimensional row vector, where $\beta_{0\ell}$ is the intercept corresponding to the null state if $\ell = 0$ and the ℓ^{th} mixture component in the non-null state if $\ell \neq 0$. The slope coefficients β_1, \dots, β_r are the covariate effects, which are assumed homogeneous across states. Given the hidden state S_t , the Poisson mean is expressed by the covariates \mathbf{z}_t and the corresponding regression coefficients through a log-linear regression model.

Specifically, the conditional distribution of X_t under the null state is $[X_t | S_t = 0, \mathbf{z}_t] \sim \text{Poisson}(\mu_{0t})$, where $\log(\mu_{0t}) = \beta_{00} + \beta_1 z_{t,1} + \dots + \beta_r z_{t,r}$. The conditional distribution of X_t under the non-null state, when the non-null distribution is assumed as an L -component mixture distribution, is

$$[X_t | S_t = 1, \mathbf{z}_t] \sim \sum_{\ell=1}^L p_\ell \text{Poisson}(\mu_{\ell t}),$$

where p_ℓ is the mixture proportion of the ℓ^{th} component, $0 < p_\ell < 1$, $\sum_{\ell=1}^L p_\ell = 1$, and $\log(\mu_{\ell t}) = \beta_{0\ell} + \beta_1 z_{t,1} + \dots + \beta_r z_{t,r}$, $\forall \ell \in 1, \dots, L$.

Let H_{0t} and H_{1t} represent the t^{th} pair of the null and alternative hypotheses on testing the true state S_t under which the t^{th} observation X_t was collected. Let β_0 be the intercept term in the above log-linear regression model, whose value is decided by S_t . Thus, the hypotheses can be written out as $H_{0t} : \beta_0 = \beta_{00}$ vs $H_{1t} : \beta_0 \neq \beta_{00}$ or $\beta_0 > \beta_{00}$ or $\beta_0 < \beta_{00}$, for $t = 1, \dots, m$. The goal of testing is to detect the underlying true state under which each of the m observations is generated. The focus is then on the hidden state indicators $\{S_t\}_{t=0}^m$. Sun and Cai (2009) proposed a local index of significance (LIS) as the test statistic in multiple testing based on HMM. Wang et al. (2019) showed that the Bayesian hierarchical framework allows easy Bayesian estimates of LIS as the posterior probabilities. Su and Wang (2020) extended the LIS test procedure to Poisson count data.

By the Bayesian framework, the LIS test statistics under PReg-HMM can still be computed as the posterior probabilities of the null states and are defined as $\text{LIS}_t = P_\vartheta(S_t =$

$0|\mathbf{X}, \mathbf{Z}$) by including the covariate effects in estimation. Let $\text{LIS}_{(t)}$ be the ordered t^{th} test statistic in an ascending order and $\text{H}_{(0t)}$ be the corresponding null hypothesis. With the pre-specified FDR at the level α , the PReg-HMM test procedure has the threshold constructed by rejecting hypotheses $\text{H}_{(01)}, \dots, \text{H}_{(0k)}$, where

$$k = \max \left\{ i : \frac{1}{i} \sum_{t=1}^i \text{LIS}_{(t)} \leq \alpha \right\}.$$

3. MODEL ESTIMATION AND SELECTION

The Bayesian scheme makes it straightforward to estimate the model parameters $\vartheta = (\mathcal{A}, \boldsymbol{\pi}, \mathcal{F})$ as well as the hidden state indicators and the LIS test statistics. To decide the number of mixture components in f_1 , the marginal likelihood method and BIC are used in model selection. As there is no closed form for the likelihood function, it has to be evaluated through various sampling methods. We described five different methods to obtain the marginal likelihood estimate. The R package `HMMtesting` has been developed to include algorithms for PReg-HMM (Wang, 2021). The `HMMtesting` package is available on [ResearchGate](#). R codes used in simulation studies and the data example are available on [Github](#). A sample R code is also included as the supplementary material.

3.1 Estimation

To facilitate parameter estimation, two layers of latent indicators are constructed. Besides the underlying null or non-null state indicators $\mathbf{S} = \{S_t\}_{t=0}^m$, the second-layer indicators $\mathbf{S}^* = \{S_t^*\}_{t=0}^m$ label the structure of the non-null mixture distribution and take discrete values in the set $\{1, 2, \dots, L\}$ given $S_t = 1$. Let $P(S_t^* = \ell | S_t = 1) = p_\ell$ for $\ell = 1, \dots, L$, where p_ℓ is the corresponding proportion for the ℓ^{th} component, $0 < p_\ell < 1$ and $\sum_{\ell=1}^L p_\ell = 1$. The mixture indicator $S_t^* = \ell$ indicates that X_t belongs to the ℓ^{th} mixture component in the non-null state. We set $P(S_t^* = 0 | S_t = 0) = 1$.

The prior distributions for parameters $\vartheta = (\mathcal{A}, \boldsymbol{\pi}, \mathcal{F})$ as well as the hidden indicators are

$$\begin{aligned} a_{j0} &\sim \text{Beta}(e_{j0}, e_{j1}), j = 0, 1, \\ p(\mathbf{S}|\mathcal{A}) &= p(S_0|\mathcal{A}) \prod_{t=1}^m P(S_t|S_{t-1}, \mathcal{A}), \\ S_t^*|\mathbf{p}, S_t = 1 &\sim \text{Multinomial}(\mathbf{p}), t = 1, \dots, m, \\ \mathbf{p} &\sim \text{Dirichlet}(e_1, \dots, e_L), \\ \boldsymbol{\beta}_\ell &\sim N_{(r+1)}(b_0, B_0), \forall \ell \in 0, 1, \dots, L, \end{aligned}$$

where the prior hyper-parameters are set as $e_{jj} = 2$, $e_{jj'} = 1$, $j, j' = 0, 1$ and $j \neq j'$, $e_l = 4$, $l = 1, \dots, L$, $b_0 = 0$ and $B_0 = 100I_{r+1}$ as suggested by Frühwirth-Schnatter (2006). More discussion on the priors of a_{j0} 's can be found in Frühwirth-Schnatter (2001).

There are two main steps in the Markov chain Monte Carlo (MCMC) implementation. The first step is to sample the parameters ϑ conditional on the hidden state indicator \mathbf{S} and the mixture indicator \mathbf{S}^* . The second step is to sample \mathbf{S} and \mathbf{S}^* conditional on parameters ϑ . The inference for the hidden state indicator \mathbf{S} is carried out by Forward-Filtering-Backward-Sampling (FFBS) (Frühwirth-Schnatter, 2006). The test statistic $P_\vartheta(S_t = 0|\mathbf{X}, \mathbf{Z})$ is the backward smoothed probability. The posterior density of the stationary probability π_0 given \mathbf{S} and \mathbf{S}^* is straightforwardly sampled with the Gibbs's sampling. The conjugate posterior distributions of the transition matrix elements a_{jj} 's and the proportions \mathbf{p} of the components in the mixture distribution under the non-null state are all available in closed forms.

The posterior distributions for the ℓ^{th} state specific regression coefficients are

$$\begin{aligned} p(\boldsymbol{\beta}_0|\mathbf{X}_{\mathbf{S}=0}, \mathbf{Z}, \mathbf{S}) &\propto p(\mathbf{X}_{\mathbf{S}=0}|\mathbf{Z}, \mathbf{S}, \boldsymbol{\beta}_0)p(\boldsymbol{\beta}_0), \\ p(\boldsymbol{\beta}_\ell|\mathbf{X}_{\mathbf{S}^*=\ell|\mathbf{S}=1}, \mathbf{Z}, \mathbf{S}, \mathbf{S}^*) &\propto p(\mathbf{X}_{\mathbf{S}^*=\ell|\mathbf{S}=1}|\mathbf{Z}, \mathbf{S}, \mathbf{S}^*, \boldsymbol{\beta}_\ell)p(\boldsymbol{\beta}_\ell), \end{aligned}$$

for $\ell = 0, 1, \dots, L$, which do not have closed-forms. A normal approximation of the posterior distribution was discussed by El-Sayyad (1973) and Chan and Vasconcelos (2009). However, it requires large Poisson means and does not provide a good approximation under count sparsity. Thus, we employ the Metropolis-Hastings algorithm as well as the adaptive rejection sampling in $\boldsymbol{\beta}_\ell$'s sampling.

3.2 Model selection

Below are five sampling-based estimation methods to obtain marginal likelihood estimates. The performance of each method is examined and compared in simulation studies (Section 4).

The importance sampling method (IS) is obtained through

$$\hat{p}_{IS}(\mathbf{X}|\mathbf{Z}, \mathcal{M}_L) = \frac{1}{J} \sum_{j=1}^J \frac{p(\mathbf{X}|\mathbf{Z}, \vartheta_L^{(j)})p(\vartheta_L^{(j)})}{q(\vartheta_L^{(j)})},$$

where $\vartheta_L^{(j)}$ is the j^{th} sample from the prior $p(\vartheta_L)$, $j = 1, \dots, J$. The importance density is

$$\begin{aligned} (1) \quad q(\vartheta_L^{(j)}) &= \frac{1}{\mathcal{I}} \sum_{i=1}^{\mathcal{I}} p(\mathcal{A}^{(j)}|\mathbf{S}^{(j,i)})p(\mathbf{p}|\mathbf{S}^{*(j,i)})p(\boldsymbol{\beta}_0^{(j)}|\mathbf{S}^{(j,i)}, \mathbf{X}, \mathbf{Z}) \\ &\quad \cdot \sum_{\ell=1}^L p(\boldsymbol{\beta}_\ell^{(j)}|\mathbf{S}^{*(j,i)}, \mathbf{X}, \mathbf{Z}), \end{aligned}$$

where j stands for the j^{th} sample from a total of J MCMC samplings and i stands for the i^{th} sample from a total of \mathcal{I} samplings. The \mathcal{I} samplings is a subset of randomly selected samples from the J MCMC samplings. The choice of \mathcal{I} is based on the number of the total MCMC samplings J and

the number of the hidden Markov states. We choose $\mathcal{I} = 200$ for a two-state PReg-HMM with $J = 2000$ after 3000 burn-ins.

The reciprocal importance sampling (RI) is

$$\hat{p}_{RI}(\mathbf{X}|\mathbf{Z}, \mathcal{M}_L) = \left\{ \frac{1}{D} \sum_{d=1}^D \frac{q(\vartheta_L^{(d)})}{p(\mathbf{X}|\mathbf{Z}, \vartheta_L^{(d)})p(\vartheta_L^{(d)})} \right\}^{-1},$$

where $\vartheta_L^{(d)}$ is the d^{th} posterior sample from a total of D MCMC samplings and $q(\vartheta_L^{(d)})$ is the importance density for the model as in (1).

The harmonic mean estimator (HM) by Newton and Raftery (1994) is

$$\hat{p}_{HM}(\mathbf{X}|\mathbf{Z}, \mathcal{M}_L) = \left\{ \frac{1}{D} \sum_{d=1}^D \frac{1}{p(\mathbf{X}|\mathbf{Z}, \vartheta_L^{(d)})} \right\}^{-1},$$

where again $\vartheta_L^{(d)}$ is the d^{th} posterior sample from the D MCMC samplings and $p(\mathbf{X}|\mathbf{Z}, \vartheta_L^{(d)})$ is the mixture likelihood evaluated at the d^{th} sampling. While the computation for the HM estimator is convenient and faster, its performance is shown to be less accurate compared with IS and RI methods.

The modified harmonic mean estimator (HM₂) is obtained through

$$\hat{p}_{HM2}(\mathbf{X}|\mathbf{Z}, \mathcal{M}_L) = \left\{ \frac{1}{D} \sum_{d=1}^D \frac{1}{p(\mathbf{X}|\mathbf{Z}, \tilde{\mathbf{S}}^{(d)})} \right\}^{-1},$$

where $\tilde{\mathbf{S}}^{(d)}$ is the d^{th} posterior sample of the indicators (\mathbf{S}, \mathbf{S}^*) from a total of D posterior samplings.

The Bridge Sampling (BS) (Meng and Wong (1996)) is specified as

$$\hat{p}_{BS,t}(\mathbf{X}|\mathbf{Z}, \mathcal{M}_L) = \frac{D_{IS}^{-1} \sum_{d_{IS}=1}^{D_{IS}} \alpha(\tilde{\vartheta}_L^{(d_{IS})}) p^*(\tilde{\vartheta}_L^{(d_{IS})}|\mathbf{X}, \mathbf{Z}, \mathcal{M}_L)}{D_{RI}^{-1} \sum_{d_{RI}=1}^{D_{RI}} \alpha(\vartheta_L^{(d_{RI})}) q(\vartheta_L^{(d_{RI})})}$$

with

$$\alpha(\tilde{\vartheta}_L^{(d_{IS})}) = \left[q(\tilde{\vartheta}_L^{(d_{IS})}) D_{IS} + \frac{p^*(\tilde{\vartheta}_L^{(d_{IS})}|\mathbf{X}, \mathbf{Z}, \mathcal{M}_L) D_{RI}}{\hat{p}_{BS,t-1}} \right]^{-1},$$

$$\alpha(\vartheta_L^{(d_{RI})}) = \left[q(\vartheta_L^{(d_{RI})}) D_{IS} + \frac{p^*(\vartheta_L^{(d_{RI})}|\mathbf{X}, \mathbf{Z}, \mathcal{M}_L) D_{RI}}{\hat{p}_{BS,t-1}} \right]^{-1},$$

and

$$p^*(\vartheta_L|\mathbf{X}, \mathbf{Z}, \mathcal{M}_L) = p(\mathbf{X}|\mathbf{Z}, \vartheta_L)p(\vartheta_L),$$

where $\tilde{\vartheta}_L^{(d_{IS})}$ is the d_{IS}^{th} sample from the prior $p(\vartheta_L)$ with $d_{IS} = 1, \dots, D_{IS}$ and $\vartheta_L^{(d_{RI})}$ is the d_{RI}^{th} sample from the posterior $p(\vartheta_L|\mathbf{X}, \mathbf{Z}, \mathcal{M}_L)$ with $d_{RI} = 1, \dots, D_{RI}$. The IS

estimator $\hat{p}_{IS}(\mathbf{X}|\mathbf{Z}, \mathcal{M}_L)$ is set as the start value of BS sampling.

Besides the marginal likelihood methods, we also implement BIC. It is defined as $\text{BIC}_L = -2 \log p(\mathbf{X}|\mathbf{Z}, \hat{\vartheta}_L, \mathcal{M}_L) + d_L \log(m)$, where $\hat{\vartheta}_L$ represents the estimation on the model parameter $(\mathcal{A}, \boldsymbol{\pi}, \mathcal{F})$ in model M_L and d_L is the number of distinct parameters in ϑ_L . The smaller the BIC_L , the better the model fits.

4. SIMULATION STUDY

Here the proposed Poisson Markov switching regression model (PReg-HMM) is compared with the Poisson hidden Markov model (PHMM) (Su and Wang, 2020) for their performance in multiple testing on dependent count data under different scenarios. In each scenario, a total of 100 data sets with $m = 3000$ observations are generated. In all the simulation studies, the pre-specified FDR is set at $\alpha = 0.10$. The initial stationary distribution $\boldsymbol{\pi}^0 = (0, 1)$. The transition matrix \mathcal{A} is with $a_{00} = 0.95$, $a_{01} = 0.05$, $a_{10} = 0.2$ and $a_{11} = 0.8$. The values of state specific regression coefficients $\boldsymbol{\beta}_\ell$ are described later in each case. For each data set, both PReg-HMM and PHMM are fitted and the results are compared.

To compare the performance in hypotheses testing, three metrics are examined, including the false discovery rate (FDR), the false negative rate (FNR), and the average true positives (ATP). As defined in Sun and Cai (2009), an FDR procedure is valid if it controls FDR at a pre-specified level α (here $\alpha = 0.1$), and optimal if it has the smallest FNR among all the FDR procedures at a level α . ATP is an additional metric for efficiency among the valid FDR procedures. The higher ATP, the more efficient the procedure.

Estimation is carried out using R software version 4.0.2. on Owens of the Ohio Supercomputer Center (Ohio Supercomputer Center, 1987) and the cluster through the Cincinnati Children's Hospital Medical Center. To ensure convergence, a total of 5000 samplings are run with 3000 burn-ins for each of the PReg-HMM and PHMM models. The Geweke statistics are examined for all the parameters in $\vartheta = (\mathcal{A}, \boldsymbol{\pi}, \mathcal{F})$.

4.1 Case 1: PReg-HMM $L = 2$

In this case, observations are generated from a PReg-HMM with $L = 2$ and two covariates (z_{1t} and z_{2t}), where z_{1t} is sampled from a standard normal distribution and z_{2t} is from a Bernoulli distribution with the probability of success equal to 0.5. Observations under the null state are sampled from f_{0t} as $\text{Poisson}(\exp(0.5 - 0.25z_{1t} + z_{2t}))$ and those under the non-null state are sampled from f_{1t} as a two-component Poisson mixture $0.6 \cdot \text{Poisson}(\exp(1 - 0.25z_{1t} + z_{2t})) + 0.4 \cdot \text{Poisson}(\exp(\beta_{02} - 0.25z_{1t} + z_{2t}))$. To examine the performance under varying levels of signal strength, the intercept term β_{02} increases from 1.5 to 3 by 0.25 in the simulation. A total of 700 data sets are generated.

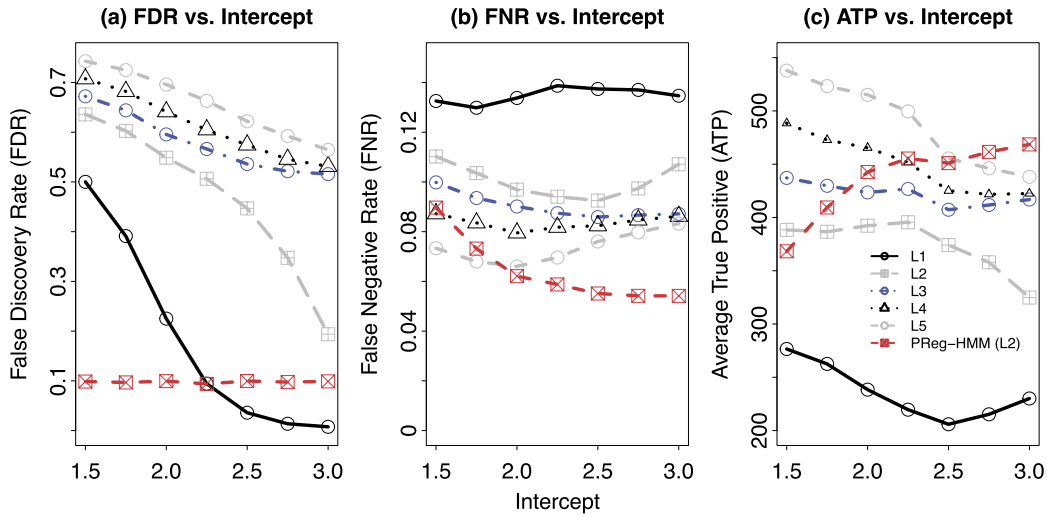


Figure 1. Comparison of FDR, FNR and ATP under PHMM and PReg-HMM. The distribution under the null state is $f_{0t} \sim \text{Poisson}(\exp(0.5 - 0.25z_{1t} + z_{2t}))$ and the distribution under the non-null state is $f_{1t} \sim 0.6 \cdot \text{Poisson}(\exp(1 - 0.25z_{1t} + z_{2t})) + 0.4 \cdot \text{Poisson}(\exp(\beta_{02} - 0.25z_{1t} + z_{2t}))$ with β_{02} varying from 1.5 to 3 by 0.25.

Table 1. Model Comparison (Case 1: PReg-HMM $L = 2$): the count (percentage) of a given model being selected as the best in the 700 simulated data sets

| Method | PHMM | | | | | PReg-HMM $L = 2$ |
|---------------------------------|----------|----------|----------|----------|----------|------------------|
| | $L = 1$ | $L = 2$ | $L = 3$ | $L = 4$ | $L = 5$ | |
| $\log p_{BS}(\mathbf{X} M_L)$ | — | 156(22%) | 249(36%) | 209(30%) | 86(12%) | 700(100%) |
| $\log p_{IS}(\mathbf{X} M_L)$ | — | 142(20%) | 246(35%) | 214(31%) | 98(14%) | 700(100%) |
| $\log p_{RI}(\mathbf{X} M_L)$ | — | 174(25%) | 242(35%) | 216(31%) | 68(10%) | 697(99.6%) |
| $\log p_{HM}(\mathbf{X} M_L)$ | — | 46(7%) | 243(35%) | 249(36%) | 162(23%) | 697(99.6%) |
| $\log p_{HM_2}(\mathbf{X} M_L)$ | — | 10(1%) | 64(9%) | 197(28%) | 429(61%) | 697(99.6%) |
| BIC_L | 235(34%) | 351(50%) | 36(5%) | 62(9%) | 16(2%) | 700(100%) |

Figure 1 shows the proposed PReg-HMM controls FDR reasonably around the given level 0.1 for all β_{02} values, while all the PHMM models do not perform well. The PHMMs with $L \geq 2$ can not control FDR and thus are not valid FDR procedures. The PHMM with $L = 1$ has the highest FNR and the lowest ATP, while it either has FDR out of control under weak signals or be conservative in FDR under stronger signals.

Table 1 shows the counts and percentages of a given model being selected as the best by a specific model selection criterion. There are a total of 700 simulated data sets for the 7 values of the intercept (β_{02}). Both the marginal likelihood method (all five sampling-based estimates) and BIC choose PReg-HMM with $L = 2$ as the best model almost 100% of the time, when comparing between PReg-HMM and PHMMs ($L = 1$ to $L = 5$). For comparison among PHMMs with different L 's, the HM_2 method chooses PHMM with $L = 5$ as the best model with a probability of 61%. The BS, IS, RI and HM methods choose randomly from PHMMs

with $L = 2$ to $L = 5$. The BIC criterion chooses PHMM with $L = 2$ as the best model with a probability of 50%. However, as shown in Figure 1, neither PHMM $L = 5$ nor $L = 2$ provides a valid testing procedure when ignoring the covariate effect. PReg-HMM with $L = 2$ is the best model with the lowest FNR, the highest ATP, and the FDR under control. Extensive simulations have been carried out on cases with data sets generated from PReg-HMMs $L = 1$ and $L = 3$ and different types of covariates. The results are similar and details are not reported here.

4.2 Case 2: PHMM $L = 1$

PHMM (Su and Wang, 2020) is a special case of the proposed PReg-HMM. Here we show that the proposed PReg-HMM model can provide similar inference as the PHMM model when data are actually generated under the simpler PHMM model.

In the simulation, observations under the null state are

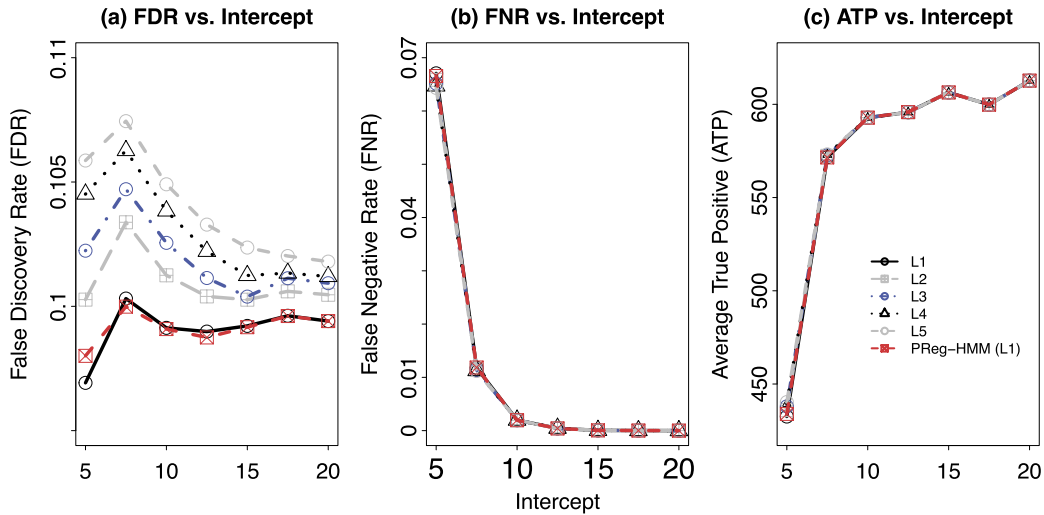


Figure 2. Comparison of FDR, FNR and ATP under PHMM and PReg-HMM test procedures. The distribution under the null state f_0 is Poisson(2) and the distribution under the non-null state f_1 is Poisson(μ_1) with μ_1 varying from 5 to 20 by 2.5.

Table 2. Model Comparison (Case 2: PHMM $L = 1$): the count (percentage) of a given model being selected as the best in the 700 simulated data sets

| Method | PHMM | | | | | PReg-HMM $L = 1$ |
|---------------------------------|-----------|-----------|----------|-----------|-----------|------------------|
| | $L = 1$ | $L = 2$ | $L = 3$ | $L = 4$ | $L = 5$ | |
| $\log p_{BS}(\mathbf{X} M_L)$ | 682 (97%) | - | - | - | - | 700 (100%) |
| $\log p_{IS}(\mathbf{X} M_L)$ | 504 (72%) | 1 (0.15%) | 10 (1%) | 47 (7%) | 131 (19%) | 511 (73%) |
| $\log p_{RI}(\mathbf{X} M_L)$ | 663 (95%) | 2 (0.29%) | - | - | - | 698 (99.7%) |
| $\log p_{HM}(\mathbf{X} M_L)$ | 373 (53%) | 271 (39%) | 86 (12%) | 15 (2%) | 6 (1%) | 322 (46%) |
| $\log p_{HM_2}(\mathbf{X} M_L)$ | 454 (65%) | 233 (33%) | 7 (2%) | 4 (0.57%) | 1 (0.15%) | 414 (59%) |
| BIC_L | 404 (58%) | 164 (23%) | 70 (10%) | 49 (7%) | 47 (7%) | 370 (53%) |

sampled from f_0 as Poisson(2) and the non-null state observations are sampled from a single Poisson distribution f_1 as Poisson(μ_1) with μ_1 varying from 5 to 20 by 2.5. Under this setup, the true parameter values in a PReg-HMM model are approximately 0.7 for β_{00} under the null state and in the range of 1.6 \sim 3 under the non-null state for β_{01} , respectively. The true underlying model is PHMM $L = 1$ and thus there is no covariate effect. When fitting the PReg-HMM, a randomly generated binary covariate $z_{1t} \sim$ Bernoulli (0.5) is included in the regression model part. The true values for the regression coefficient β_1 is 0.

Figure 2 compares results in multiple testing. Among PHMMs ($L = 1$ to 5), PHMM with $L = 1$ controls FDR around the given level 0.1. There is no benefit over-fitting PHMMs with $L \geq 2$, which leads to higher FDR and no improvement in FNR or ATP. The proposed PReg-HMM with $L = 1$ has almost identical performance as the true model PHMM with $L = 1$ in terms of FDR, FNR and ATP.

The model comparison results are shown in Table 2. When comparing among all 6 models (PHMMs with $L =$

1, \dots , 5 and PReg-HMM with $L = 1$), all sampling-based marginal likelihood estimates and BIC choose the true model PHMM with $L = 1$ as the best model with the highest probabilities (53 \sim 97%). Between PHMMs ($L = 2$ to 5) and PReg-HMM ($L = 1$), PReg-HMM has been selected as the best model with the highest probabilities (46 – 100%).

5. PEOPLE FLOW DATA EXAMPLE

The people flow data set contains 3-month records of people flowing in and out of the Calit2 institute building at the University of California, Irvine (Dua (2017), <https://archive.ics.uci.edu/ml/index.php>). The data were collected between July 25th, 2005 and November 5th, 2005 with half hour count aggregates. The sensor at the main entrance of the building was used to collect the data. There are a total of 5,040 recorded counts with the minimum 0 and the maximum 114. The PHMM model was applied to detect the active periods of the building (Su and Wang, 2020). The analysis, which provided good results when comparing

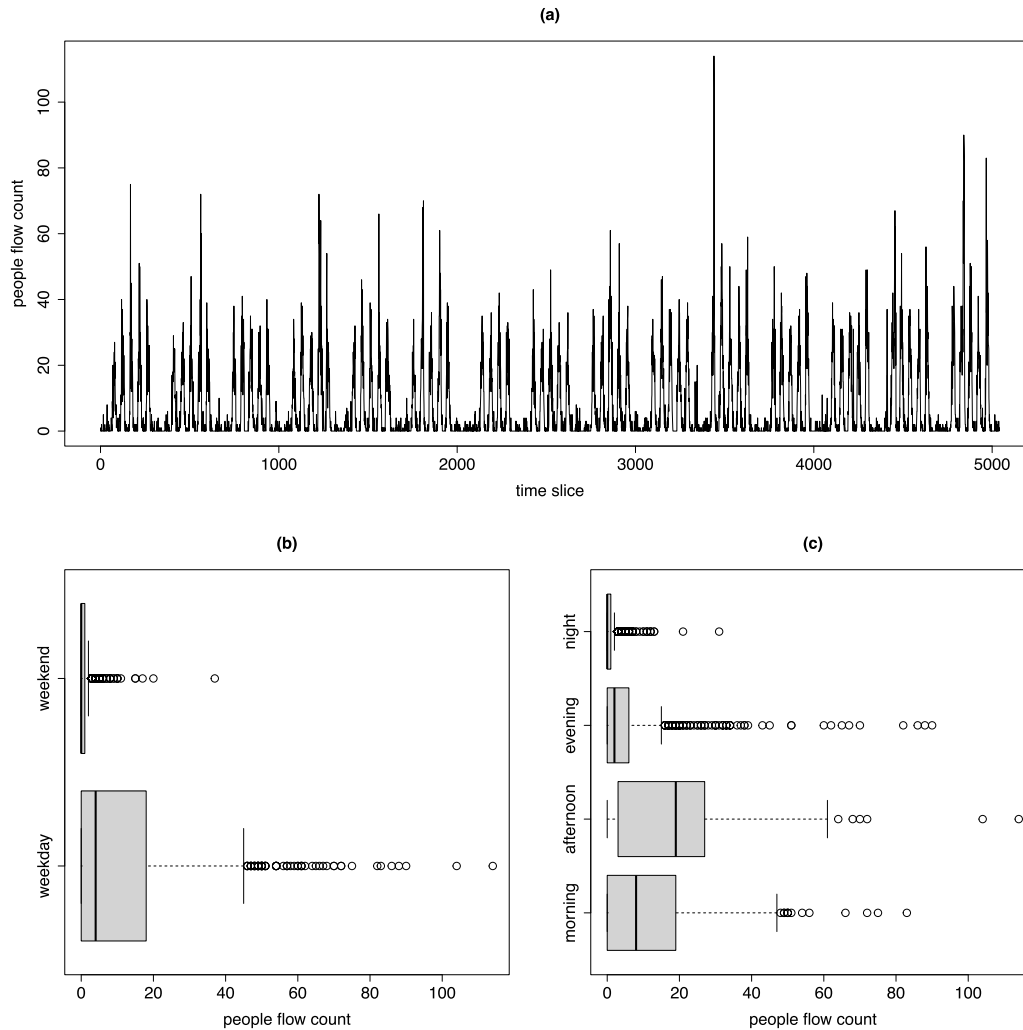


Figure 3. (a) Raw counts of people flowing in and out every 30-minute interval (b) the day of the week and the count (c) the time of the day and the count. There are a total of 5,040 observations.

with an event log file, was able to detect the period with high people flow. However, the people flow naturally fluctuates due to factors such as the time of the day and the day of the week. The PHMM model cannot distinguish natural dynamics of the day from the “abnormally” high flow periods due to factors other than the natural fluctuations. Here we showcase that the regression model structure in PReg-HMM can achieve the inference goal to detect abnormalities beyond the natural dynamics.

Two factors are added to the Poisson regression model for the observed counts. One is the day of the week. If the observation is recorded on a weekend day (Sunday or Saturday), the covariate “weekend” is set to 1, and it is set to 0 otherwise. The other factor is the time of day. There are four levels in the time of day, including “morning” (8:00 AM to 12:00 PM), “afternoon” (12:00 PM to 5:00 PM), “evening” (5:00 PM to 10:00 PM) and “night” (10:00 PM to 8:00 AM). The night is set as the reference level, and three binary co-

variates are included for the other three levels.

Figure 3 Panel (a) shows the raw counts of people flowing in and out every 30-minute interval during the recorded period. Panels (b) and (c) shows the boxplots of the counts versus the the day of the week and the time of the day levels. Clearly, the weekdays usually have higher people flow than weekends. During the day, the afternoon has the highest flow on average, followed by the morning. The evening usually has a lower flow but with quite some outlying periods with high counts. The quietest time is the night. The goal is to examine “abnormally” active periods in the building, given the natural fluctuations due to the time of the day and the day of the week effects.

Model comparison results are shown in Table 3. Comparing among PReg-HMMs with $L = 1$ to 7, all five sampling-based marginal likelihood estimates and BIC choose PReg-HMM $L = 6$ as the best model. The last column in Table 3 is for PHMM with $L = 5$, which is the best model selected

Table 3. Model Comparison: People Flow Data Example

| Method | PReg-HMM | | | | | | | PHMM $L = 5$ |
|---------------------------------|----------|---------|---------|---------|---------|---------|---------|--------------|
| | $L = 1$ | $L = 2$ | $L = 3$ | $L = 4$ | $L = 5$ | $L = 6$ | $L = 7$ | |
| $\log p_{BS}(\mathbf{X} M_L)$ | -13715 | -12668 | -11367 | -11208 | -11185 | -11159 | -11160 | -12109 |
| $\log p_{IS}(\mathbf{X} M_L)$ | -13715 | -12667 | -11366 | -11206 | -11185 | -11158 | -11159 | -12109 |
| $\log p_{RI}(\mathbf{X} M_L)$ | -13714 | -12669 | -11365 | -11206 | -11184 | -11158 | -11158 | -12115 |
| $\log p_{HM}(\mathbf{X} M_L)$ | -13678 | -12625 | -11311 | -11145 | -11119 | -11077 | -11070 | -12023 |
| $\log p_{HM_2}(\mathbf{X} M_L)$ | -12417 | -10881 | -9300 | -8884 | -8779 | -8704 | -8680 | -9553 |
| BIC_L | 27992 | 26027 | 23368 | 23080 | 23070 | 23021 | 23026 | 24881 |

among PHMMs (Su and Wang, 2020). PReg-HMMs with $L \geq 3$ are considered better than PHMM with $L = 5$, as the marginal likelihood estimates are larger and the BIC values are smaller.

Under the selected PReg-HMM $L = 6$, the estimated null state intercept is around -1.21 with the posterior standard deviation (SD) as 0.06 . The six estimated non-null state intercepts are 0.29 (SD = 0.07), 0.79 (SD = 0.07), 1.21 (SD = 0.07), 1.64 (SD = 0.07), 2.40 (SD = 0.06), and 3.73 (SD = 0.12), respectively. The corresponding proportions in the non-null state are 0.30 (SD = 0.03), 0.42 (SD = 0.03), 0.20 (SD = 0.03), 0.06 (SD = 0.01), 0.02 (SD = 0.003), 0.005 (SD = 0.002), respectively. The estimate on the week-end effect shows that the log of expected counts for week-ends decreases by 2.29 (SD = 0.06) compared to weekdays. Comparing the time of day effect with night as the reference level, the log of expected counts for morning, afternoon and evening increases by 2.11 (SD = 0.06), 2.37 (SD = 0.06) and 1.81 (SD = 0.06), respectively. The estimated null state proportion π_0 is round 0.48 (SD = 0.02) with the transition matrix as $a_{00} = 0.88$ (SD = 0.008) and $a_{11} = 0.89$ (SD = 0.007).

PReg-HMM with $L = 6$ detects 2,740 events, which is more than 2,088 detected by PHMM with $L = 5$. A closer look into the testing results from PReg-HMM and PHMM helps show different perspectives in the questions being addressed by these two procedures. Table 4 shows the covariate information for those observations where testing decisions from PReg-HMM and PHMM differ. As shown in Figure 3, even an “abnormally” high people flow on a weekend may be far lower than that on an average weekday. Furthermore, the posterior estimates on the regression coefficients confirm that people flow is low on average during weekends and at the night. Since PHMM does not account for the covariate effect, these periods are highly likely be claimed as inactive periods. On the other hand, PReg-HMM aims at detecting active periods while accounting for usually higher people flow during weekdays and mornings and afternoons. As shown in Table 4, all the time intervals (291) which are claimed as non-null under PHMM but null under PReg-HMM are from weekdays and mostly from mornings and

Table 4. People flow data: difference in hypotheses testing results on null ($S_t^{model} = 0$) and non-null ($S_t^{model} = 1$) states by PHMM ($L = 5$) and PReg-HMM ($L = 6$)

| | $S_t^{PHMM} = 1$ & $S_t^{PReg-HMM} = 0$ | | $S_t^{PHMM} = 0$ & $S_t^{PReg-HMM} = 1$ | |
|-----------|--|---------|--|---------|
| | weekday | weekend | weekday | weekend |
| morning | 102 | 0 | 0 | 144 |
| afternoon | 2 | 0 | 0 | 184 |
| evening | 186 | 0 | 0 | 41 |
| night | 1 | 0 | 252 | 324 |

evenings. None is from the weekends and only 1 is from the nights during the weekdays. On the other hand, among the 954 time intervals which are claimed as non-null by PReg-HMM but null by PHMM, 702 are from weekends with almost a half from the night time of the weekends. The remaining 252 intervals are all from the night period on weekdays.

In practice, which procedure is appropriate would depend on the inference goal. Particularly, it relies on the definition of “abnormal” events to detect. For the people flow example, PHMM detects active periods with high people flow regardless of the day of the week or the time of the day, while PReg-HMM detects *relatively* active periods, by controlling the day of the week and the time of the day effects.

6. DISCUSSION

When the inference goal is on detecting abnormal events beyond natural fluctuations, hypotheses testing can achieve better performance by taking into consideration of relevant covariate effects. One of our simulation studies with one binary covariate (not shown here) shows that under strong non-null signals, it is possible for the PHMM with a large number of mixture components performs similarly as a PReg-HMM. However, it is challenging for PHMM to achieve similar performance when PReg-HMM is the true model with more complicated covariate structure (such as the case shown in Section 4.1).

The PReg-HMM model presented here assumes that the covariates have homogeneous effects across states. Thus, the regression models under the null and non-null states only differ in the intercept terms. It is likely that both the intercepts and the covariate effects are state-specific. Further flexibility can be achieved by allowing $\beta_0, \beta_1, \dots, \beta_r$ to vary across both between the states and among the mixtures and thus the regression model has heterogeneous covariate effects under different states. In both homogeneous and heterogeneous cases, the detection of the null and non-null states is about testing the hypotheses on a regression coefficient, which is now related either to the intercept terms or to a regression coefficients of a covariate. When the covariate effects are homogeneous across states, they are not of interest in hypotheses testing itself directly. However, if the covariate effects are not considered, the changes in those covariates may be mistakenly included in modeling the hidden states of interest. On the other hand, when the covariate effects are heterogeneous and change across states, the regression coefficients are more directly related to the hypotheses testing problem. In many cases, hypotheses testing is directly on the effect switching across states of a given covariate. Under the Bayesian hierarchical model framework, it is straightforward to extend the current testing model to include heterogeneous covariate effects. A couple of simulation examples are included in R package `HMMtesting` (Wang, 2021). In practice, when assuming heterogeneous covariate effects, the testing procedure requires a pre-specified order on one or several regression coefficients.

ACKNOWLEDGEMENTS

Dr. Xia Wang's work was supported in part by NIH/NIMH Grant R01MH119814.

Received 25 December 2020

REFERENCES

- CHAN AB, VASCONCELOS N (2009) Bayesian Poisson regression for crowd counting. In: 2009 IEEE 12th International Conference on Computer Vision, pp 545–551. [10.1109/ICCV.2009.5459191](https://doi.org/10.1109/ICCV.2009.5459191)
- DUA D, CASEY, G (2017) UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>]. Irvine, CA: University of California, School of Information and Computer Science.
- EL-SAYYAD GM (1973) Bayesian and classical analysis of Poisson regression. *Journal of the Royal Statistical Society, Series B: Methodological* 35(3): 445–451. [MR0341729](https://doi.org/10.2307/23441729)
- FRÜHWIRTH-SCHNATTER S (2001) Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models". *Journal of the American Statistical Association* 96(453): 194–209. [MR1952732](https://doi.org/10.2307/131952732)
- FRÜHWIRTH-SCHNATTER (2006) *Finite Mixture and Markov Switching Models*. Springer, New York. [MR2265601](https://doi.org/10.1007/978-1-4939-9811-1)
- GOLDFELD S, QUANDT RE (1973) The estimation of structural shifts by switching regressions. *Annals of Economic and Social Measurement* 2(4): 475–485.
- IHLER AT, HUTCHINS J, SMYTH P (2006) Adaptive event detection with time-varying Poisson processes. In: KDD'06: Proceedings of the 12th ACM SIGKDD international conference on knowledge discovery and data mining, 207–216.
- MENG XL, WONG WH (1996) Simulating ratios of normalizing constants via a simple identity: a theoretical exploration. *Statistica Sinica* 6(4): 831–860. [MR1422406](https://doi.org/10.1214/153922406)
- NEWTON M, RAFTERY A (1994) Approximate Bayesian inference by the weighted likelihood bootstrap. *Journal of the Royal Statistical Society Series B (Statistical Methodology)* 56(1): 3–26. [MR1257793](https://doi.org/10.2307/23467793)
- OHIO SUPERCOMPUTER CENTER (1987) Ohio Supercomputer Center. Columbus, OH: Ohio Supercomputer Center. <http://osc.edu/ark:/19495/f5s1ph73>
- SEBASTIAN T, JEYASEELAN V, JEYASEELAN L, ANANDAN S, GEORGE S, BANGDIWALA S (2019) Decoding and modeling of time series count data using Poisson hidden Markov model and Markov ordinal logistic regression models. *Statistical Methods in Medical Research* 28(5): 1551–1563. [MR3941093](https://doi.org/10.1093/bjbm/abz019)
- SU WZ, WANG X (2020) Hidden Markov model in multiple testing on dependent count data. *Journal of Statistical Computation and Simulation* 90(5): 889–906. [MR4067352](https://doi.org/10.1080/00949655.2020.18067352)
- SUN W, CAI TT (2009) Large-scale multiple testing under dependence. *Journal of the Royal Statistical Society Series B (Statistical Methodology)* 71(2): 393–424. [MR2649603](https://doi.org/10.1111/j.1467-9868.2009.00660.x)
- WANG X (2021) `HMMtesting` 2.1: Bayesian hidden Markov models for dependent multiple hypotheses testing. [10.13140/RG.2.2.33172.53129/1](https://doi.org/10.13140/RG.2.2.33172.53129/1)
- WANG X, SHOJAIE A, ZOU J (2019) Bayesian hidden Markov models for dependent large-scale multiple testing. *Computational Statistics and Data Analysis* 136: 123–136. [MR3944676](https://doi.org/10.1016/j.csda.2019.04.006)
- XU Z, LABER EB, STAICU AM (2020) Hierarchical continuous time hidden Markov model, with application in zero-inflated accelerometer data. In: Zhao Y, Chen DG (eds) *Statistical Modeling in Biomedical Research. Emerging Topics in Statistics and Biostatistics*. Springer, Cham. [10.1007/978-3-030-33416-1_7](https://doi.org/10.1007/978-3-030-33416-1_7)

Weizhe Su
 Division of Statistics and Data Sciences
 Department of Mathematical Sciences
 University of Cincinnati
 USA
 E-mail address: [suwh@mail.uc.edu](mailto:suw@mail.uc.edu)

Xia Wang
 Division of Statistics and Data Sciences
 Department of Mathematical Sciences
 University of Cincinnati
 USA
 E-mail address: xia.wang@uc.edu