Proceedings of the 5th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/116/2025.20591

Metropolis-Hastings Algorithms based on a Zero-inflated Poisson Model

Yaoyu Chen[1a,*]

[1]The University of Manchester, Manchester, United Kingdom
a. yaoyu180@gmail.com
**corresponding author*

Abstract: The Metropolis-Hastings algorithm is a well-proven Markov Chain Monte Carlo (MCMC) technique for generating sequences of random samples from probability distributions difficult to sample directly. In this article, we develop two versions of the Metropolis-Hastings algorithm — Independent Metropolis and Random-Walk Metropolis — and integrate them with a zero-inflated Poisson model to calculate the percentage of children among widowed women in Manchester, 2021. The excess of zeros in the data are estimated using a zero-inflated Poisson distribution. So we calculate both Metropolis-Hastings algorithms for estimation of the parameters of the model such as the Poisson rate parameter and the zero-inflation parameter. Using these algorithms we have shown that the zero-inflated Poisson model estimates are quite consistent with observed data and hence useful for modelling count data with excess zeros. These observations indicate that Independent Metropolis and Random-Walk Metropolis methods both fit the model parameters correctly, with Random-Walk Metropolis resulting in stable convergence.

Keywords: Metropolis-Hastings algorithm, Independent Metropolis, Random-walk Metropolis, Markov Chain Monte Carlo, zero-inflated Poisson model

1. Introduction

The Markov Chain Monte Carlo (MCMC) method named by American physicist Metropolis et al [1, 2] was developed and proposed during the "Manhattan Project" to create the atomic bomb. Canadian statistician Hastings [3] and his Ph.D. student Peskun [4] overcame the dimensional bottleneck problem encountered by conventional Monte Carlo methods. They generalized the Metropolis algorithm and extended it as a statistical simulation tool, forming the Metropolis-Hastings algorithm.

Compared with the Metropolis method, the Metropolis-Hastings algorithm s a statistical simulation tool looks more professional. What is important is that the symmetrical proposal distribution function in Metropolis-Hastings algorithm is not necessary, so it is more flexible and convenient to apply [5, 6, 7].

DOI: 10.54254/2755-2721/116/2025.20591

2. Ease

2.1. Metropolis-Hastings algorithm

The Metropolis-Hastings algorithm is one of the most popular MCMC methods and is utilized to simulate a sequence of random samples, a Markov chain converging to a stationary target distribution $\pi(\cdot)$, which is difficult to sample directly. Let $q(x_{t+1}|x_t)$ denote the proposal probability that moves from x_t at time t to x_{t+1} at time t+1. The proposed distribution can facilitate the simulation of a Markov chain corresponding to the target distribution. The algorithm is implemented in the following steps:

Step 1: Initialize the current state x_t;

Step 2: Generate a candidate x_{t+1}^* value from the proposal distribution $q(x_{t+1}^*|x_t)$;

Step 3: Calculate the acceptance probability $\alpha_t(x_t, x_{t+1}^*)$:

$$\alpha_t(x_t, x_{t+1}^*) = min\left\{\frac{\pi(x_{t+1}^*)q(x_t|x_{t+1}^*)}{\pi(x_t)q(x_{t+1}^*|x_t)}, 1\right\} \tag{1}$$

Step 4: Generate u_t from a uniform distribution on [0, 1];

Step 5: If $u_t \leq \alpha_t$, accept the candidate value x_{t+1}^* and set $x_{t+1} = x_{t+1}^*$; otherwise, reject the candidate value x_{t+1}^* and set $x_{t+1} = x_t$;

Step 6: Repeat Step 2 to Step 5 until a sample of the desired size *N* is obtained.

Based on the choice of the proposal distribution, the Independent Metropolis algorithm and Random-walk Metropolis algorithm (RWM) are two typical Metropolis-Hastings algorithms [8].

2.1.1. Independent Metropolis algorithm

In this case, the candidate value x_{t+1}^* is independent of the current state x_t, which is formed as

$q(x_{t+1}^*|x_t) = q(x_t)$,

The acceptance probability $\alpha_t(x_t, x_{t+1}^*)$ is then modified to

$$\alpha_t(x_t, x_{t+1}^*) = min\left\{\frac{\pi(x_{t+1}^*)q(x_t)}{\pi(x_t)q(x_{t+1}^*)}, 1\right\} \tag{2}$$

2.1.2. Random-walk Metropolis algorithm

The critical feature of RWM is that the candidate value x_{t+1}^* is centered distributedly on the current state x_t, i.e.

$x_{t+1}^* = x_t + ò$,

where ò is distributed symmetrically with mean zero. For example, the proposal value x_{t+1}^* given x_t is simulated from a normal distribution $x_{t+1}^*|x_t \sim N(x_t, \sigma^2)$. Therefore, the acceptance probability $\alpha_t(x_t, x_{t+1}^*)$ is simplified to

$$\alpha_t(x_t, x_{t+1}^*) = min\left\{\frac{\pi(x_{t+1}^*)}{\pi(x_t)}, 1\right\} \tag{3}$$

2.2. Application

2.2.1. Data and Model

In this paper, we study the number of widowed women with different number of children in Manchester, 2021, the data of which is shown in Table 1 and Figure 1.

Table 1: The numbers of widows with different number of children

Children	Widows
0	52
1	10
2	25
3	8
4	7
5	3
6	0

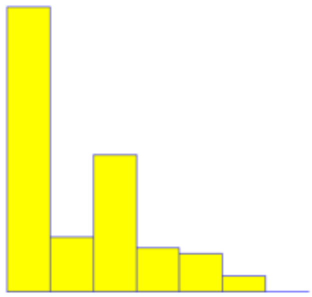


Figure 1: Histogram for the numbers of widows

We use a zero-inflated Poisson model based on the distribution in Figure 1. Let Y be the number of children per widowed woman. The data $y_1, y_2, \ldots, y_n$ are the observations corresponding to Table 1 [9], where n is the total number of widows and assumed independent and identically distributed from the following probability mass function:

$$f(y|\lambda,\varepsilon) = \begin{cases} \varepsilon + (1-\varepsilon)e^{-\lambda} & y = 0 \\ (1-\varepsilon)\frac{\lambda^y e^{-\lambda}}{y!} & y = 1,2,\ldots,n \end{cases} \tag{4}$$

To predict the number of widows with different numbers of children using the zero-inflated Poisson distribution, we first estimate the values of λ and ε through the Metropolis-Hastings algorithms.

For simplicity, assume that both λ and ε have uniform priors. Thus, the posterior distribution for (λ, ε) is proportional to the following form:

$$f(\lambda,\varepsilon|y) \propto \prod_{i=1}^{n} f(y_i|\lambda,\varepsilon) \tag{5}$$

Therefore, we obtain the target distribution of (λ, ε):

$$\pi(\lambda,\varepsilon) = \prod_{i=1}^{n} f(y_i|\lambda,\varepsilon) \tag{6}$$

Based on the properties of zero-inflated Poisson distribution and the observed data, we initialize the value of λ and ε as follows:

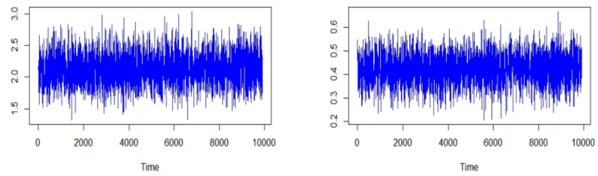
$$\hat{\varepsilon} = f(0) = \frac{52}{n} = 0.4952;$$

$$\hat{\lambda} = y/(1-\hat{\varepsilon}) = 2.3962, as\ EY = \lambda(1-\varepsilon) \tag{7}$$

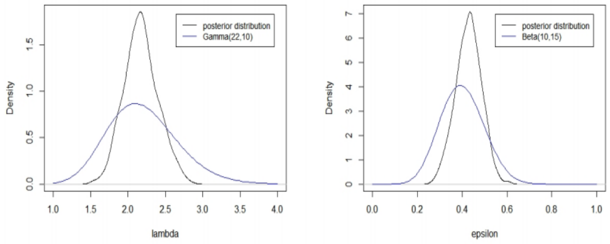
Subsequently, we shall produce sequences of λ and ε simulation with a sample size N=10000 by two different algorithms.

DOI: 10.54254/2755-2721/116/2025.20591

2.2.2. Independent Metropolis



(a) Traceplot for λ (b) Traceplot for ε



(c) Comparison between proposal and posterior for λ (d) Comparison between proposal and posterior for ε

Figure 2: Independent Metropolis

According to the posterior distribution, we assume $\lambda \sim Gamma(\alpha_1, \beta_1)$ and $\varepsilon \sim Beta(\alpha_2, \beta_2)$ to be proposal distributions for λ and ε, respectively. Thus, we have proposal probability $q_1(\lambda)$ and $q_2(\varepsilon)$:

$$q_1(\lambda) = \frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)} \lambda^{\alpha_1 - 1} e^{-\beta_1 \lambda} \tag{8}$$

$$q_2(\varepsilon) = \frac{1}{B} \varepsilon^{\alpha_2 - 1} (1 - \varepsilon)^{\beta_2 - 1} \tag{9}$$

In this case, the acceptance probability $\alpha_t\{(\lambda_t, \varepsilon_t), (\lambda_{t+1}^*, \varepsilon_{t+1}^*)\}$ is

$$\alpha_t = min\left\{\frac{\pi(\lambda_{t+1}^*, \varepsilon_{t+1}^*)}{\pi(\lambda_t, \varepsilon_t)} \cdot \frac{q_1(\lambda_t)}{q_1(\lambda_{t+1}^*)} \cdot \frac{q_2(\varepsilon_t)}{q_2(\varepsilon_{t+1}^*)}, 1\right\} \tag{10}$$

To confirm a specific distribution, we need to choose reasonable parameters $\alpha_1, \beta_1, \alpha_2, \beta_2$ in the proposal probabilities $q_1(\lambda)$ and $q_2(\epsilon)$. The key point in this selection is to ensure that the proposal distributions for λ and ϵ are both close to the posterior distribution and exhibit greater dispersion than the posterior distribution [10].

For example, $q_1(\lambda) \sim N(\alpha_1, \beta_1^2)$ and $q_2(\epsilon) \sim N(\alpha_2, \beta_2^2)$ are possible choices. Figure 2(a) and Figure 2(b) show the trace plots for the simulations of λ and ϵ, respectively. Both trace plots in Figure 2 illustrate stationary Markov chains that move quickly, indicating that the algorithm performs effectively. Additionally, the comparison between the proposal and posterior distributions for λ and ϵ in Figures 2(c) and 2(d) demonstrates that the proposals not only have means similar to the posteriors but also exhibit greater dispersion than the posterior distributions. These characteristics confirm that the chosen parameters for the proposal distributions are appropriate and contribute to the effectiveness of the algorithm.

2.2.3. Random-walk Metropolis

Based on the RWM algorithm, Normal proposal distribution is assumed for both λ and ε. Therefore, we propose to sample $\varepsilon^* \sim N(\varepsilon, \sigma_\varepsilon^2)$ and $\lambda^* \sim N(\lambda, \sigma_\lambda^2)$, where σ_ε^2 and σ_λ^2 represent the proposed variances for λ and ε, respectively. In this case, the acceptance probability is

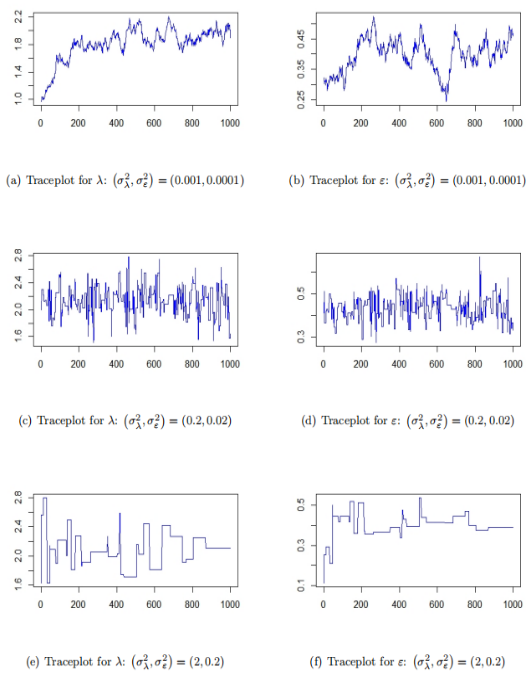
DOI: 10.54254/2755-2721/116/2025.20591

$$\alpha_t = min\left\{\frac{\pi(\lambda_{t+1}^*, \varepsilon_{t+1}^*)}{\pi(\lambda_t, \varepsilon_t)}, 1\right\} \tag{11}$$

Then tuning σ_ε^2 and σ_λ^2 is able to optimize the performance of the algorithm. We choose $(\sigma_\lambda^2, \sigma_\varepsilon^2)$ in different values (0.001, 0.0001), (0.2, 0.02), (2, 0.2), and all the implements of the first 1000 samples are shown in the Figure 3.

For the case $(\sigma_\lambda^2, \sigma_\varepsilon^2) = (0.001, 0.0001)$, most candidate values of λ and ε are accepted. However, the move in each step is very small. Thus it takes a long time for the Markov chains of λ and ε to explore and they do not converge to posterior distributions. $(\sigma_\lambda^2, \sigma_\varepsilon^2) = (0.2, 0.02)$ is the other extreme case, where few candidate values of λ and ε in the moving are accepted. For the case $(\sigma_\lambda^2, \sigma_\varepsilon^2) = (0.2, 0.02)$, the traceplots show that both Markov chains move fast and appear to be converged and therefore, it is the optimal choice for the RWM algorithm.

2.2.4. Result and Analysis



(a) Traceplot for λ: $(\sigma_\lambda^2, \sigma_\varepsilon^2) = (0.001, 0.0001)$

(b) Traceplot for ε: $(\sigma_\lambda^2, \sigma_\varepsilon^2) = (0.001, 0.0001)$

(c) Traceplot for λ: $(\sigma_\lambda^2, \sigma_\varepsilon^2) = (0.2, 0.02)$

(d) Traceplot for ε: $(\sigma_\lambda^2, \sigma_\varepsilon^2) = (0.2, 0.02)$

(e) Traceplot for λ: $(\sigma_\lambda^2, \sigma_\varepsilon^2) = (2, 0.2)$

(f) Traceplot for ε: $(\sigma_\lambda^2, \sigma_\varepsilon^2) = (2, 0.2)$

Figure 3: Random-walk Metropolis

Based on the samples of size N=10000 for λ and ε from two algorithms, we estimate the numbers of widows with different children in Manchester from the zero-inflated Poisson model using Monte-Carlo method. The results are summarized in Table 2 and shown in Figure 3 and Figure 4.

Table 2: Result

Children	Widows (Observed)	Widows (Ind)	Widows (RWM)
0	52	51.76	51.8
1	10	15.4	15.38
2	25	16.32	16.31
3	8	11.53	11.52
4	7	6.11	6.11
5	3	2.59	2.59
6	0	0.92	0.91

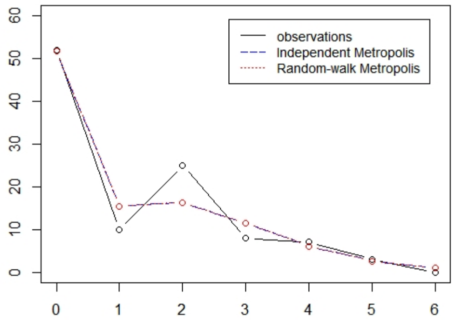


Figure 4: Comparison of result

We observe that both Independent Metropolis and RWM have good agreement on the numbers of widowed women with different numbers of children. Comparing with the observations, both algorithms provide accurate estimations for the number of widows with 0, 4, 5, and 6 children. However, there are large differences between the expected numbers and observations for cases that widowed women have 1, 2, 3 children.

3. Conclusion

In this paper, two variations of the Metropolis-Hastings algorithm—the Independent Metropolis and the Random-Walk Metropolis (RWM)—were applied to a zero-inflated Poisson model to estimate the distribution of widows with different numbers of children in Manchester. By adjusting the parameters within the algorithms, reasonable samples of the zero-inflated Poisson model parameters π and λ were obtained.

When comparing the predicted values (obtained from the model using the estimated parameters) with the observed values (actual data from Manchester), the model demonstrates a generally good fit. However, there is a noticeable gap between the predicted values and the observed values in specific cases, particularly for widows with 1, 2, or 3 children. This suggests that while the model performs adequately overall, the true distribution of the data may not be fully captured by the zero-inflated Poisson model.

To address these discrepancies, future studies could consider exploring alternative models, such as the Negative Binomial Zero-Inflated model, which may better account for overdispersion in the data and provide an improved fit to the observed values.

References

[1] N. Metropolis and S. Ulam, "The monte carlo method," Journal of the American statistical association, vol. 44, no. 247, pp. 335–341, 1949.

[2] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," The journal of chemical physics, vol. 21, no. 6, pp. 1087–1092, 1953.

[3] W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," Biometrika, vol. 57, no. 1, pp. 97–109, 1970.

[4] P. H. Peskun, "Optimum monte-carlo sampling using markov chains," Biometrika, vol. 60, no. 3, pp. 607–612, 1973.

[5] A. Gelman, W. R. Gilks, and G. O. Roberts, "Weak convergence and optimal scaling of random walk metropolis algorithms," The annals of applied probability, vol. 7, no. 1, pp. 110–120, 1997.

[6] C. Andrieu, N. De Freitas, A. Doucet, and M. I. Jordan, "An introduction to mcmc for machine learning,"Machine learning, vol. 50, no. 1, pp. 5–43, 2003.

[7] C. Robert and G. Casella, "A short history of markov chain monte carlo: Subjective recollections from incomplete data," Statistical Science, vol. 26, no. 1, pp. 102–115, 2011.

DOI: 10.54254/2755-2721/116/2025.20591

[8] Madrigal-Cianci, Juan P., Fabio Nobile, and Raul Tempone. "Analysis of a class of multilevel Markov chain Monte Carlo algorithms based on independent Metropolis–Hastings." SIAM/ASA Journal on Uncertainty Quantification 11.1 (2023): 91-138.

[9] Teixeira, J. S., et al. "A new adaptive approach of the Metropolis-Hastings algorithm applied to structural damage identification using time domain data." Applied Mathematical Modelling 82 (2020): 587-606.

[10] Choi, Michael CH, and Lu-Jing Huang. "On hitting time, mixing time and geometric interpretations of metropolis–hastings reversiblizations." Journal of Theoretical Probability 33.2 (2020): 1144-1163.