JIRSS (2025)
Vol. 24, No. 2, pp 91-106
DOI: 10.22034/jirss.2026.2073280.1143

Markov SIRD Epidemic Model with Semi-Markov Sojourn-Time Analysis of COVID-19

Faihatuz Zuhairoh[1], **M. Rais Ridwan**[1]

[1]Department of Mathematics Education, STKIP YPUP Makassar, Makassar, South Sulawesi, Indonesia.

Received: 01/10/2025, Accepted: 27/01/2026, Published online: xx/xx/2026

Abstract. The SIRD (Susceptible–Infected–Recovered–Deceased) model is a standard framework for analyzing infectious disease dynamics. Classical continuous-time Markov formulations assume constant transition rates and memoryless (exponential) sojourn-times, which may oversimplify empirical epidemic processes. In this study, the Markov SIRD model is employed as a baseline, with transition parameters estimated analytically via maximum likelihood. To assess the validity of the exponential sojourn-time assumption, a semi-Markov framework is introduced exclusively for duration analysis of the infected state. Specifically, the sojourn-times associated with recovery ($I \to R$) and death ($I \to D$) transitions are modeled using exponential and Weibull distributions and compared using likelihood-based criteria. Using COVID-19 data from the Special Region of Yogyakarta, Indonesia, the results show that Weibull distributions provide a substantially better fit than the exponential assumption for both recovery and mortality durations. These findings indicate significant deviations from the memoryless assumption underlying the Markov model. This study does not construct a full dynamic semi-Markov epidemic simulator; instead, the semi-Markov framework is used to statistically characterize and evaluate the temporal structure of infected-state durations. The results highlight the importance of realistic sojourn-time modeling for understanding epidemic progression and for assessing the limitations of classical Markov-based epidemic models.

Keywords. COVID-19; maximum likelihood method; semi-Markov; SIRD model; sojourn-time.

MSC: 60J28, 60J20, 60K15.

Corresponding Author: Faihatuz Zuhairoh (fzuhairoh@gmail.com) .
M. Rais Ridwan (mraisridw4n@gmail.com).

1 Introduction

The COVID-19 pandemic has significantly disrupted global health systems and highlighted the need for accurate epidemiological models to support public health interventions and resource allocation. Early epidemic models, such as the classical SIR (Susceptible, Infected, Recovered) framework, have provided foundational insights into disease dynamics. However, the increasing complexity of real-world epidemics has motivated the development of extended compartmental models—including SIRD, SEIR, SVIRS, and others—that incorporate additional states such as deceased, exposed, and vaccinated individuals. Our previous works (Zuhairoh et al., 2022, 2023b,c,d, 2024) have employed Markov-based formulations for these expanded models, illustrating their usefulness in discrete-time analysis. Other studies have similarly adopted Markov assumptions for modeling infectious disease transmission (Hubbard and Zhou, 2011; Zhang et al., 2019).

Despite their popularity, traditional Markov models rely on constant transition rates and memoryless sojourn-times, where the future evolution of the system depends only on the current state and not on the time already spent in that state (Hsieh et al., 2002; Cohen and Reza, 2011). This simplifying assumption limits their ability to realistically capture the variability of infection, recovery, and mortality durations observed in empirical datasets. As a result, Markov models may underestimate or misrepresent temporal patterns critical for accurate epidemic prediction.

To overcome these limitations, recent studies have proposed semi-Markov extensions that allow the transition intensities to depend on the elapsed time within each state. Semi-Markov models provide greater flexibility by incorporating general sojourn-time distributions, enabling a more realistic representation of disease progression (Foucher et al., 2005; Asanjarani et al., 2021). Applications of semi-Markov modeling in epidemiology include analyses of COVID-19 transmission dynamics (Sun et al., 2023) and broader infectious disease systems (Wang and Mustafa, 2023), demonstrating the potential benefits of relaxing the memoryless assumption.

Although semi-Markov SIRD frameworks have been studied previously, such as in Zuhairoh et al. (2023a), the methodological focus of the present work differs substantially. The earlier study considered a multi-state generalization of SIRD, emphasizing theoretical properties of semi-Markov transitions. In contrast, the current study adopts the classical four-compartment SIRD framework and introduces a two-tier modeling strategy. First, we derive closed-form maximum likelihood estimators for the Markov baseline model (Theorem 2.1), providing analytical expressions for the infection, recovery, and mortality rates. These estimators form the foundation for evaluating model fit and predictive behavior. Second, we apply a semi-Markov extension only to the sojourn-times of infected individuals, fitting Exponential and Weibull distributions to empirically observed durations. This allows us to assess whether relaxing the exponential sojourn-time assumption improves model adequacy.

This distinction is essential: the semi-Markov implementation in this paper is not used to redefine the epidemic transition structure but rather to model the empirical distribution of recovery and mortality times. Consequently, Theorem 2.1 plays a central role in estimating the Markov parameters used in the numerical analysis, whereas the semi-Markov framework provides a complementary evaluation of temporal variability through sojourn-time modeling.

Additional developments combining semi-Markov processes with machine learning and modern computational methods have further enhanced predictive accuracy in infectious disease research (Hale and Aarts, 2023; Getz et al., 2019). These technologies illustrate the growing relevance of flexible duration-based modeling in understanding epidemic trends and designing control strategies.

Despite these advancements, applications of semi-Markov approaches to COVID-19, specif-

ically using regional datasets from Indonesia, remain scarce. This study aims to address this gap by analyzing COVID-19 data from the Special Region of Yogyakarta (DIY), Indonesia. DIY is a densely populated and highly dynamic region, making it an informative case study for evaluating both Markov and semi-Markov model components.

The contributions of this study are threefold. First, we derive closed-form maximum likelihood estimators for the continuous-time Markov SIRD model, providing an analytically tractable baseline for epidemic dynamics. Second, we conduct a semi-Markov sojourn-time analysis of the infected state by fitting exponential and Weibull distributions to empirical recovery and mortality durations. Third, we evaluate the limitations of the Markov assumption through likelihood-based criteria and Monte Carlo trajectory simulations of the Markov CTMC model. Importantly, the semi-Markov framework in this study is employed for duration modeling rather than for constructing a full semi-Markov epidemic trajectory model.

Section 2 introduces the SIRD model and formulates the Markov transition structure. Section 3 presents the semi-Markov sojourn-time framework. Section 4 details the estimation procedures and model evaluations. Finally, Section 5 concludes the study and outlines future research directions.

2 Deterministic and Stochastic Formulation of the SIRD Model

The SIRD (Susceptible–Infected–Recovered–Deceased) model is a classical compartmental epidemic framework widely used to study the dynamics of infectious diseases. This model originates from the seminal work of Kermack and McKendrick (1927) and has since been developed extensively in modern epidemiological literature (Hethcote, 2008; Andersson and Britton, 2000). It is particularly useful for analyzing the spread of diseases such as COVID-19, providing insight into transmission dynamics and enabling policymakers to assess potential intervention strategies.

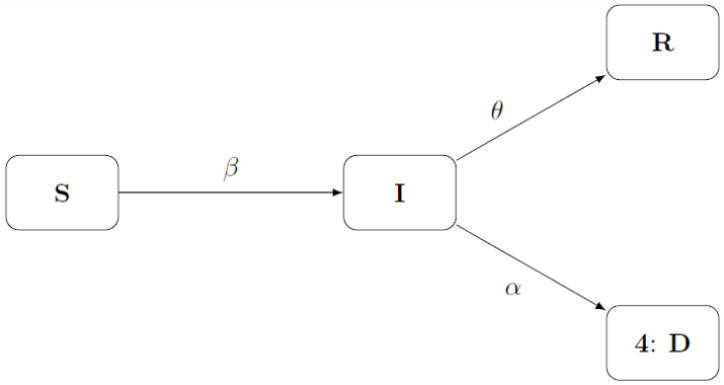

Figure 1: Transition structure of the SIRD epidemic model, illustrating the flow from susceptible (S) to infected (I), recovery (R), and death (D).

The model divides the population into four compartments: $S(w)$: number of susceptible individuals at time,$I(w)$: number of infected individuals,$R(w)$: number of recovered individuals, and $D(w)$: cumulative deaths. The relationships among compartments are illustrated in Figure 1, where individuals move from susceptibility to infection at rate β, infected individuals recover at rate θ, and deaths occur at rate α.

2.1 Deterministic SIRD Model

Under the classical deterministic formulation, the dynamics are described by the system of differential equations:

$$\frac{dS(w)}{dw} = -\frac{\beta S(w)I(w)}{N} + o(\Delta w), \tag{2.1}$$

$$\frac{dI(w)}{dw} = \frac{\beta S(w)I(w)}{N} - \theta I(w) - \alpha I(w) + o(\Delta w), \tag{2.2}$$

$$\frac{dR(w)}{dw} = \theta I(w) + o(\Delta w), \tag{2.3}$$

$$\frac{dD(w)}{dw} = \alpha I(w) + o(\Delta w). \tag{2.4}$$

Equation (2.1) decreases the susceptible population through infection; the term $\frac{\beta S(w)I(w)}{N}$ reflects the mass-action assumption, where contacts are proportional to the product $S\dot{I}$. Equation (2.2) describes the net change in infections: infections increase via the same term from Equation (2.1), and decrease through recoveries $\theta I(w)$ and deaths $\alpha I(w)$. Equation (2.3) represents the flow of individuals recovering from the disease at rate θ. Equation (2.4) captures disease-induced mortality at rate α.

Together, these equations define the deterministic transmission–recovery–mortality dynamics commonly used in COVID-19 modeling (Chatterjee et al., 2021).

2.2 Discrete-Time Transition Probabilities

To link the deterministic system to a stochastic process, we approximate the transitions over a small interval Δw. Let: $P_{S\to I}(w, \Delta w)$ is probability a susceptible becomes infected, $P_{I\to R}(w, \Delta w)$ is probability an infected individual recovers, $P_{I\to D}(w, \Delta w)$ is probability an infected individual dies, $P_{k\to k}(w, \Delta w)$ is probability an individual remains in the same state $k \in \{S, I, R\}$.

The first-order approximations are:

$$p_{S\to I}(w, \Delta w) = \frac{\beta S(w)I(w)}{N}\Delta w, \tag{2.5}$$

$$p_{I\to R}(w, \Delta w) = \theta I(w)\Delta w, \tag{2.6}$$

$$p_{I\to D}(w, \Delta w) = \alpha I(w)\Delta w, \tag{2.7}$$

$$p_{k\to k}(w, \Delta w) = 1 - \left(\frac{\beta S(w)I(w)}{N} + \theta I(w) + \alpha I(w)\right)\Delta w. \tag{2.8}$$

In these expressions, the instantaneous compartment counts are denoted by $s = S(w), i = I(w), r = R(w)$, and $d = D(w)$. The use of alphabetic labels avoids potential confusion that may arise when employing numerical indices (e.g., 1, 2, 3, 4) to represent compartmental states. Together, these transition probabilities constitute a discrete-time Markov approximation of the continuous-time SIRD process.

2.3 Continuous-Time Transition Rates

The corresponding continuous-time transition probability for observing a change from state (s, i, r) to $(s + x, i + y, r + z)$ over interval Δw is:

$$p(\Delta w) = \begin{cases} \dfrac{\beta}{N} si\Delta w + o(\Delta w), & (x,y,z) = (-1,1,0) \\ \theta i\Delta w + o(\Delta w), & (x,y,z) = (0,-1,1) \\ \alpha i\Delta w + o(\Delta w), & (x,y,z) = (0,-1,0) \\ 1 - \left(\dfrac{\beta si}{N} + \theta i + \alpha i\right)\Delta w + o(\Delta w), & (x,y,z) = (0,0,0). \\ o(\Delta w), & \text{elsewhere} \end{cases} \tag{2.9}$$

These transitions depend on the current state and the model parameters β, θ, and α.Such a formulation treats the epidemic as a continuous-time Markov chain (CTMC), in which events occur one at a time with exponentially distributed sojourn-times. This forms the foundation for the maximum likelihood estimation developed in the next section.

Theorem 2.1 (Maximum likelihood estimators for the continuous-time Markov SIRD model)**.** *Consider a continuous-time Markov jump process $X(t)$ on the finite state space $\mathcal{S} = \{S, I, R, D\}$ representing the susceptible, infected, recovered and deceased compartments. Assume the following time-homogeneous transition intensities (per unit time):*

$$\lambda_{S\to I}(t) = \beta\frac{S(t)I(t)}{N}, \qquad \lambda_{I\to R}(t) = \theta\, I(t), \qquad \lambda_{I\to D}(t) = \alpha\, I(t),$$

and no other transitions are possible. Let the observation window be $[0, T]$. Let

1. *a denote the total number of observed $S \to I$ transition events in $[0, T]$,*
2. *b denote the total number of observed $I \to R$ transition events in $[0, T]$,*
3. *c denote the total number of observed $I \to D$ transition events in $[0, T]$,*

and define the integrated exposures

$$T_{SI} := \int_0^T \frac{S(t)I(t)}{N}\, dt, \qquad T_I := \int_0^T I(t)\, dt.$$

Assume standard regularity conditions for likelihood-based inference (identifiability of parameters, differentiability of the log-likelihood and interchangeability of differentiation and integration). Then the maximum likelihood estimators of (β, θ, α) are

$$\hat{\beta} = \frac{a}{T_{SI}}, \qquad \hat{\theta} = \frac{b}{T_I}, \qquad \hat{\alpha} = \frac{c}{T_I}.$$

Proof. We provide a complete derivation. The proof follows the standard construction of the likelihood for a continuous-time Markov jump process and direct maximization.

1. Intensity (hazard) representation. By assumption the total instantaneous rate (intensity) of events at calendar time t is

$$\lambda_{\text{tot}}(t) = \lambda_{S\to I}(t) + \lambda_{I\to R}(t) + \lambda_{I\to D}(t) = \beta\frac{S(t)I(t)}{N} + (\theta + \alpha)I(t).$$

At any time t when an event occurs, the probability that the observed event is of a particular type is proportional to its intensity. This standard representation underlies the likelihood factorization below.

2. Likelihood factorization for observed jump times. Let $\{t_1, t_2, \ldots, t_m\}$ be the ordered jump times in $[0, T]$ (all jump types combined), and for each jump time t_ℓ let the observed jump type be denoted by $J_\ell \in \{S \to I,\ I \to R,\ I \to D\}$. The likelihood of the observed sample path (jump times and jump types) for a Markov jump process has the form (see e.g. Andersen et al., counting process theory; or standard CTMC likelihood derivations):

$$L(\beta, \theta, \alpha) = \prod_{\ell=1}^{m} \lambda_{J_\ell}(t_\ell)\, \exp\Big(- \int_0^T \lambda_{\text{tot}}(t)\, dt \Big), \tag{2.10}$$

where $\lambda_{J_\ell}(t_\ell)$ is the intensity of the observed transition at time t_ℓ, and the exponential term is the survival term (no-event probability) accounting for the integrated intensity over $[0, T]$.

3. Separate contributions by event type. Group the product in (2.10) according to event types. Using the notation that a is the number of $S \to I$ events occurring at times $u_1, \ldots, u_a$, b is the number of $I \to R$ events at times $v_1, \ldots, v_b$, and c is the number of $I \to D$ events at times $w_1, \ldots, w_c$, we can rewrite the product as

$$\prod_{\ell=1}^{m} \lambda_{J_\ell}(t_\ell) = \Big(\prod_{j=1}^{a} \lambda_{S\to I}(u_j) \Big) \Big(\prod_{j=1}^{b} \lambda_{I\to R}(v_j) \Big) \Big(\prod_{j=1}^{c} \lambda_{I\to D}(w_j) \Big).$$

Substituting the intensity forms gives

$$\prod_{\ell=1}^{m} \lambda_{J_\ell}(t_\ell) = \beta^a \prod_{j=1}^{a} \frac{S(u_j) I(u_j)}{N} \ \times\ \theta^b \prod_{j=1}^{b} I(v_j) \ \times\ \alpha^c \prod_{j=1}^{c} I(w_j).$$

4. Exponential survival term (integrated intensity). The integrated intensity term in (2.10) is

$$\int_0^T \lambda_{\text{tot}}(t)\, dt = \beta \int_0^T \frac{S(t) I(t)}{N}\, dt + (\theta + \alpha) \int_0^T I(t)\, dt = \beta\, T_{SI} + (\theta + \alpha) T_I.$$

Therefore the full likelihood is

$$L(\beta, \theta, \alpha) = \beta^a \theta^b \alpha^c \Big(\prod_{j=1}^{a} \frac{S(u_j) I(u_j)}{N} \Big) \Big(\prod_{j=1}^{b} I(v_j) \Big) \Big(\prod_{j=1}^{c} I(w_j) \Big) \exp\Big[- \beta T_{SI} - (\theta + \alpha) T_I \Big].$$

5. Log-likelihood. Taking logarithms and collecting terms that depend on the parameters yields

$$\ell(\beta, \theta, \alpha) := \log L(\beta, \theta, \alpha) = a \log \beta + b \log \theta + c \log \alpha - \beta T_{SI} - (\theta + \alpha) T_I + C,$$

where C is a term that does not depend on β, θ, α (it collects the log of the observed state-dependent factors $\prod S(u_j) I(u_j)/N$ and the $\log I(\cdot)$ terms).

6. First-order conditions (score equations). Differentiate the log-likelihood with respect to each parameter and set derivatives equal to zero.

$$\frac{\partial \ell}{\partial \beta} = \frac{a}{\beta} - T_{SI} = 0 \quad \Longrightarrow \quad \hat{\beta} = \frac{a}{T_{SI}}.$$

$$\frac{\partial \ell}{\partial \theta} = \frac{b}{\theta} - T_I = 0 \quad \Longrightarrow \quad \hat{\theta} = \frac{b}{T_I}.$$

$$\frac{\partial \ell}{\partial \alpha} = \frac{c}{\alpha} - T_I = 0 \quad \Longrightarrow \quad \hat{\alpha} = \frac{c}{T_I}.$$

These are critical points of the log-likelihood.

7. Second-order conditions (concavity). The second derivatives (observed information diagonal entries) are

$$\frac{\partial^2 \ell}{\partial \beta^2} = -\frac{a}{\beta^2} < 0, \qquad \frac{\partial^2 \ell}{\partial \theta^2} = -\frac{b}{\theta^2} < 0, \qquad \frac{\partial^2 \ell}{\partial \alpha^2} = -\frac{c}{\alpha^2} < 0,$$

so the log-likelihood is strictly concave in each parameter on the positive parameter domain. Hence the critical points above are global maxima for $\beta, \theta, \alpha > 0$.

8. Regularity and practical computation remark. The derivation assumes that $T_{SI} > 0$ and $T_I > 0$ so that the denominators are nonzero; these conditions are naturally satisfied in any dataset with at least one observed infection and at least one observed recovery or death. The standard regularity conditions (identifiability of parameters, differentiability and integrability allowing interchange of integration and differentiation) ensure asymptotic normality and consistency of the MLEs; these conditions are standard in counting-process based inference for epidemic models (see Andersen *et al.*, 1993, for a general treatment).

Combining the results of steps 6 and 7 yields the stated expressions for $\hat{\beta}, \hat{\theta}, \hat{\alpha}$, completing the proof. □

3 The Sojourn-time Approach in a Semi-Markov Process

A semi-Markov process is a generalization of a continuous-time Markov chain in which the transition mechanism depends not only on the current state but also on the time already spent in that state (the sojourn-time). Let $Z_n : n \geq 0$ denote the embedded Markov chain that governs state-to-state transitions, and let $T_n : n \geq 0$ be the sequence of associated sojourn-times. The continuous-time process is then defined as

$$X(t) = Z_n \quad \text{for} \quad \sum_{i=0}^{n} T_i \leq t < \sum_{i=0}^{n+1} T_i. \tag{3.1}$$

For each possible transition $k \to l$, let H_{kl} denote the sojourn-time spent in state k before transitioning to state l. Following standard semi-Markov theory, the sojourn-time distribution is given by

$$F_{kl}(w) = Pr(H_{kl} \leq w), \qquad w \geq 0, \tag{3.2}$$

with corresponding density

$$f_{kl}(w) = F'_{kl}(w)$$

whenever the derivative exists.

It is important to note that, unlike the classical Markov framework, the semi-Markov model allows different sojourn-time distributions for different transitions $k \to l$, permitting flexible modeling of epidemiological pathways.

3.1 Survival and hazard functions

The survival function associated with the sojourn-time H_{kl} is defined as Asanjarani et al. (2021)

$$S_{kl}(w) = \Pr(H_{kl} > w) = 1 - F_{kl}(w). \tag{3.3}$$

The hazard function represents the instantaneous transition risk at age w. Formally, it is defined as

$$\mu_{kl}(w) = \lim_{\Delta w \to 0} \frac{\Pr(w < H_{kl} \leq w + \Delta w | H_{kl} > w)}{\Delta w} = \frac{f_{kl}(w)}{S_{kl}(w)}. \tag{3.4}$$

This definition emphasizes that the transition intensity depends explicitly on the time already spent in state k, distinguishing semi-Markov models from their Markov counterparts, which assume constant hazards.

3.2 Exponential Distribution of Sojourn-time

A commonly used model for sojourn-times is the Exponential distribution with mean λ by Ross (2007)

$$f(w) = \frac{1}{\lambda} e^{-\frac{w}{\lambda}}, \qquad S(w) = e^{-\frac{w}{\lambda}}, \qquad \mu(w) = \frac{1}{\lambda} \tag{3.5}$$

The constant hazard rate implies that the probability of transitioning out of a state is memoryless—consistent with the Markov assumption. Within a Markov SIRD model, the parameters θ and α correspond to the constant hazards for the transitions $I \to R$ and $I \to D$, respectively.

3.3 Weibull Distribution of Sojourn-time

To accommodate more realistic, non-memoryless transitions, we adopt the Weibull distribution with shape parameter $k > 0$ and scale parameter $\lambda > 0$. The density, survival, and hazard functions are:

$$f_w = \frac{k}{\lambda} \left(\frac{w}{\lambda}\right)^{k-1} \exp\left[-\left(\frac{w}{\lambda}\right)^k\right], \tag{3.6}$$

$$S_w = \exp\left[-\left(\frac{w}{\lambda}\right)^k\right], \tag{3.7}$$

$$\mu_w = \frac{k}{\lambda} \left(\frac{w}{\lambda}\right)^{k-1}. \tag{3.8}$$

When $k = 1$, the Weibull distribution reduces to the Exponential distribution.

For $k < 1$, the hazard decreases over time, representing early transitions followed by stabilization; for $k > 1$, the hazard increases, capturing transitions that become more likely as the sojourn-time grows. This flexibility allows the semi-Markov model to capture the empirical variability of recovery and mortality durations in infectious disease dynamics.

3.4 Application to the SIRD epidemic model

In the SIRD context, the semi-Markov formulation is applied to the infected state I, where transitions may occur to recovery R or death D. Let H_{IR} and H_{ID} denote the respective sojourn-times. The competing hazards for leaving state I at age w are given by:

$$\mu_{IR}(w), \qquad \mu_{ID}(w),$$

with total hazard

$$\mu_I(w) = \mu_{IR}(w) + \mu_{ID}(w).$$

This structure allows recovery and death to follow different distributions, making the model more realistic than the Markov formulation, which assumes a single constant exit rate.

4 Data, Estimation, and Evaluation

4.1 Data Description

This study uses the official COVID-19 time-series published by the Provincial Government of the Special Region of Yogyakarta (DIY), Indonesia. The dataset contains 139 consecutive daily observations from March 15, 2020 to July 31, 2020. The available series include cumulative confirmed cases, cumulative recoveries, cumulative deaths, and the number of active cases.

The initial susceptible population was set to

$$S_0 \approx 3.6 \; \textit{million}$$

based on the 2020 population statistics published by BPS (2020). This follows the standard SIRD modeling assumption that all individuals are initially susceptible before community transmission accelerates. The initial compartment values were therefore:

$$S - 0 = 3,600,000, I_0 = 1, R_0 = 0, D_0 = 0$$

Daily transition counts were reconstructed from first differences of the cumulative series. From these we obtained:

$$n_I = 673, n_R = 399, n_D = 20$$

These values determine the event contributions in the Markov CTMC likelihood.

4.2 Parameter Estimation for the Markov SIRD Model

The closed-form estimators derived in Theorem 2.1 are directly used to obtain the Markov CTMC parameter estimates reported below, which serve as the baseline for both trajectory simulations and subsequent duration-based evaluation. Under the continuous-time Markov (CTMC) representation described in Section 2, the log-likelihood of the model is:

$$\ell_M = n_I \log \beta + n_R \log \theta + n_D \log \alpha - \beta T_{SI} - (\theta + \alpha) T_I \tag{4.1}$$

where

$$T_{SI} = \sum_w \frac{S(w)I(w)}{N}, \qquad T_I = \sum_w I(w)$$

represent the infection exposure integral and infected person-time.

Using the closed-form MLEs (Theorem 2.1):

$$\hat{\beta} = \frac{n_I}{T_{SI}}, \qquad \hat{\theta} = \frac{n_R}{T_I}, \qquad \hat{\alpha} = \frac{n_D}{T_I} \tag{4.2}$$

The exposure integrals computed from the observed series are:

$$T_{SI} = 8615.285, \qquad T_I = 8616$$

we obtained:

$$\hat{\beta} = 0.0781, \qquad \hat{\theta} = 0.0463, \qquad \hat{\alpha} = 0.0023.$$

The estimated parameters indicate that approximately 7.8% of effective susceptible–infected contacts result in new infections per unit time. Furthermore, about 4.63% of infected individuals recover each day, while approximately 0.23% die per day. These transition rates provide a quantitative description of the transmission, recovery, and mortality dynamics of COVID-19 in Yogyakarta during the study period. The corresponding event-based likelihood results are shown in Table 1.

Table 1: Likelihood Summary for the Markov CTMC SIRD Model.

Model	Distribution	$\log L$	k	n	AIC	BIC
Markov	Exponential	−4155.05	3	139	8316.10	8324.91

Here:

1. $k = 3$ parameters (β, θ, α),
2. $n = 139$ observation days,
3. AIC and BIC computed as

$$AIC = -2\ell_M + 2k, \qquad BIC = -2\ell_M + k\ln(n) \tag{4.3}$$

4.3 Sojourn-time Analysis and Semi-Markov Estimation

To assess the validity of the exponential assumption in the Markov model, the empirical sojourn durations for the transitions $I \to R$ and $I \to D$ were fitted using exponential and Weibull distributions. Tables 2 and 3 summarize the fitted log-likelihoods, AIC, BIC, and the likelihood ratio tests (LRT) for each transition.

Table 2: Model comparison for the $I \to R$ durations (recovery).

Model	$\log L$	k	AIC	BIC
Exponential	−1530.10	1	3062.20	3066.19
Weibull	−1477.50	2	2959.01	2966.98

Table 3: Model comparison for the $I \to D$ durations (death).

Model	$\log L$	k	AIC	BIC
Exponential	−37.2563	1	76.51	77.08
Weibull	−35.1779	2	74.36	75.49

The LRT strongly rejects the exponential distribution for the I→R transition ($\chi^2 = 105.19, p < 10^{-24}$) and significantly rejects it for the I→D transition ($\chi^2 = 4.157, p = 0.041$). These results confirm that the Markov (memoryless) assumption is unsuitable for the infected compartment.

4.4 Model Evaluation Through Trajectory Simulation

To evaluate the trajectory behavior of the baseline Markov SIRD model, we performed Monte Carlo simulations of the continuous-time Markov chain using the estimated parameters. These simulations aim to illustrate the dynamical implications of the exponential sojourn-time assumption, rather than to provide a generative Monte Carlo validation of a semi-Markov epidemic model.

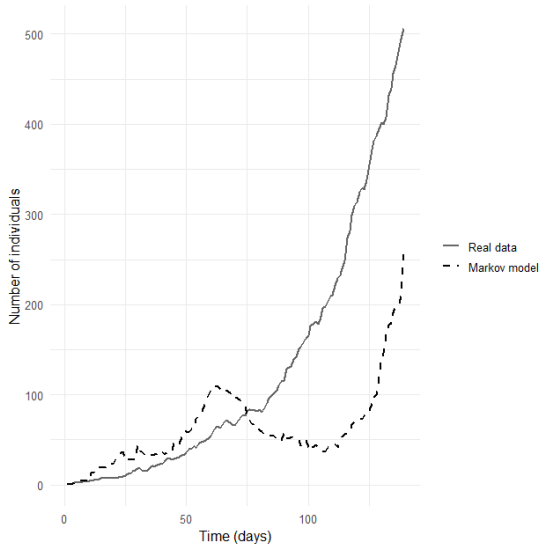

Figure 2: Infected cases: real data vs Markov model

The Markov model captures the general trend of infection but produces a smoother trajectory than the observed data. This occurs because exponentially distributed holding times impose constant transition hazards, which dampen short-term fluctuations and lead to a more gradual decline in active cases.

The model underestimates the rapid early increase in recoveries and produces a more uniform growth pattern. This discrepancy reflects the limitations of assuming a constant recovery rate θ, which does not account for the temporal heterogeneity in real recovery times.

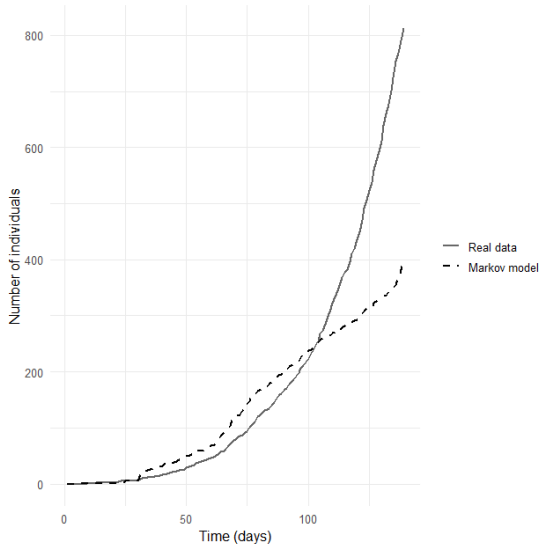

Figure 3: Recovered cases: real data vs Markov model

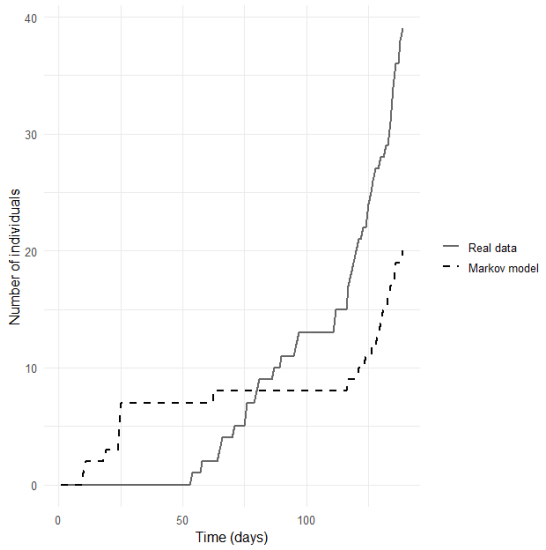

Figure 4: Death cases: real data vs Markov model

To illustrate the out-of-sample trajectory behavior of the baseline Markov SIRD model, we performed a 20% temporal hold-out experiment, withholding the final portion of the data and generating out-of-sample forecasts using the Markov SIRD model. A total of 500 Monte Carlo forward simulations were computed from the state at the end of the training window, yielding a point forecast and a 95% pointwise predictive uncertainty band.

The results are summarized in Figure 5. The Markov model systematically underestimates

the steep upward trend in late-stage infections, and although the uncertainty band widens over time, it remains substantially below the observed trajectory. Quantitative predictive accuracy metrics further support this conclusion:

$$RMSE = 63.05, \qquad MAE = 45.61$$

These results highlight the limited predictive flexibility of the exponential sojourn-time assumption, particularly during periods of rapidly changing transmission dynamics.

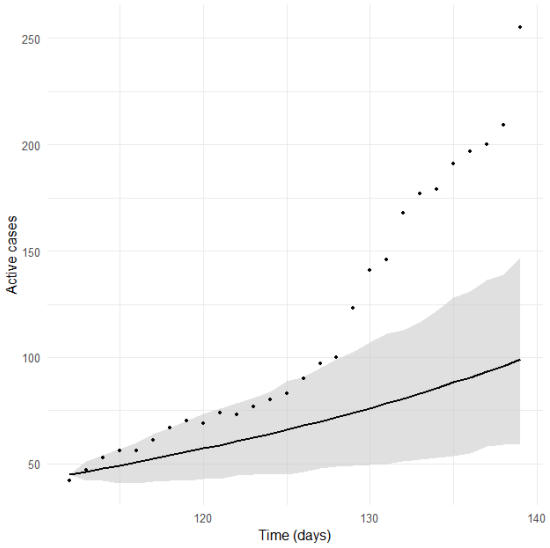

Figure 5: Observed and Markov CTMC–simulated active cases with 95% pointwise uncertainty band.

Figure 5 hold-out Monte Carlo trajectory assessment of the Markov SIRD model. Black points represent observed active cases. The red line is the mean Markov prediction derived from 500 Monte Carlo simulations, and the grey region denotes the 95% predictive uncertainty band.

The Markov model slightly overpredicts cumulative deaths, particularly in the later phase of the epidemic. While the constant mortality hazard α captures the overall monotonic trend, it does not fully reflect the deceleration observed in real death counts.

Overall, these results show that although the Markov model provides a reasonable approximation of population-level epidemic dynamics, its reliance on exponential holding times limits its flexibility. This finding aligns with the duration analysis in Section 4.3, which statistically rejects exponential sojourn-times in favor of the Weibull distribution.

4.5 Summary of Findings

The Markov CTMC model provides parameter estimates that are consistent with the observed epidemic dynamics but relies on exponential sojourn-times, which are statistically rejected for both recovery and death transitions ($p < 10^{-24}$ and $p \approx 0.041$, respectively). The Weibull-based semi-Markov analysis demonstrates a significantly better fit to individual-level transition

durations, indicating substantial deviations from the memoryless assumption. Since this study focuses on sojourn-time estimation rather than constructing a full semi-Markov SIRD trajectory model, epidemic curves are evaluated only under the Markov specification. Overall, the results highlight the importance of accommodating non-exponential sojourn-time distributions to more accurately characterize the temporal structure of infectious disease progression.

The trajectory-based evaluation further supports this conclusion. Although the Markov model provides reasonable trajectory estimates, its hold-out prediction errors (RMSE = 63.05, MAE = 45.61) and wide uncertainty bands highlight the limitations of exponential holding times. In contrast, the Weibull model achieves substantially better likelihood-based fit for sojourn-time distributions and is statistically preferred under the likelihood ratio test, reinforcing the conclusion that non-exponential sojourn-times more accurately characterize the epidemic process.

5 Conclusion

This study analyzed the SIRD epidemic model under both Markov and semi-Markov frameworks using COVID-19 data from the Special Region of Yogyakarta. The Markov continuous-time formulation provided closed-form maximum-likelihood estimators for the infection, recovery, and mortality rates through Theorem 2.1 and offered a baseline trajectory for comparison. However, likelihood ratio tests strongly rejected the exponential sojourn-time assumption underlying the Markov model for both recovery and death durations. The Weibull distribution provided a significantly better empirical fit, revealing substantial heterogeneity in individual sojourn-times that cannot be captured by the memoryless Markov specification.

Because the present study focuses on duration modeling rather than constructing a full semi-Markov SIRD trajectory model, epidemic curves were evaluated only under the Markov formulation, while the semi-Markov component contributed through improved estimation of state-dependent sojourn distributions. These findings underscore that incorporating flexible, non-exponential sojourn-time distributions is essential for accurately representing the temporal structure of individual-level disease progression, rather than for improving epidemic trajectory prediction. Future research may extend this work by developing full semi-Markov epidemic simulators or hybrid models that integrate duration-based and trajectory-based components within a unified framework.

Acknowledgements

The author would like to express sincere gratitude to the authors of the cited works for providing valuable insights that have greatly contributed to the development of this paper. Appreciation is also extended to the editorial board and anonymous reviewers journal of the Iranian Statistical Society Statistics for their constructive feedback and support throughout the review process.

References

Andersson H, Britton T. Stochastic Epidemic Models and Their Statistical Analysis. New York: Springer; 2000.

Asanjarani A, Liquet B, Nazarathy Y. Estimation of semi-Markov multi-state models: A comparison of the sojourn times and transition intensities approaches. International Journal of Biostatistics. 2021;p. 1–20. https://doi.org/10.1515/ijb-2020-0083.

Chatterjee S, Sarkar A, Chatterjee S, Karmakar M, Paul R. Studying the progress of COVID-19 outbreak in India using SIRD model. Indian Journal of Physics. 2021;95:1941–1957. https://doi.org/10.1007/s12648-020-01766-8.

Cohen T, Reza Y. Generalized Markov Models of Infectious Disease Spread: A Novel Framework for Developing Dynamic Health Policies. European Journal of Operational Research. 2011;215(3):679–687. https://doi.org/10.1016/j.ejor.2011.07.016.

Foucher Y, Mathieu E, Saint-Pierre P, Durand JF, Daurès JP. A semi-markov model based on generalized Weibull distribution with an illustration for HIV disease. Biometrical Journal. 2005;47(6):1–9. https://doi.org/10.1002/bimj.200410170.

Getz WM, Carlson C, Dougherty E, Porco TC, Salter R. An agent-based model of school closing in under-vaccinated communities during measles outbreaks. Simulation. 2019;95. https://doi.org/10.1177/0037549717721754.

Hale WW, Aarts E. Hidden Markov model detection of interpersonal interaction dynamics in predicting patient depression improvement in psychotherapy: Proof-of-concept study. Journal of Affective Disorders Reports. 2023 12;14:100635. https://doi.org/10.1016/J.JADR.2023.100635.

Hethcote HW. THE BASIC EPIDEMIOLOGY MODELS: MODELS, EXPRESSIONS FOR R 0 , PARAMETER ESTIMATION, AND APPLICATIONS. In: Mathematical Understanding of Infectious Disease Dynamics.; 2008. p. 1–61. http://www.worldscientific.com/doi/abs/10.1142/9789812834836_0001. http://www.worldscientific.com/doi/abs/10.1142/9789812834836_0001.

Hsieh HJ, Chen THH, Chang SH. Assessing chronic disease progression using non-homogeneous exponential regression Markov models: An illustration using a selective breast cancer screening in Taiwan. Statistics in Medicine. 2002;21(22):3369–3382. https://doi.org/10.1002/sim.1277.

Hubbard RA, Zhou XH. A comparison of non-homogeneous markov regression models with application to Alzheimer's disease progression. Journal of Applied Statistics. 2011;38(10):2313–2326. https://doi.org/10.1080/02664763.2010.547567.

Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. Proceedings of the Royal Society of London Series A. 1927;115(772):700–721. https://doi.org/10.1098/rspa.1927.0118.

Ross SM. Introduction to Probability Models. United States of America: Elsevier Academic Press; 2007.

Sun Q, Tan D, Zhang S. Epidemic threshold of a COVID-19 model with gaussian white noise and semi-Markov switching. Journal of the Franklin Institute. 2023;360. https://doi.org/10.1016/j.jfranklin.2023.04.035.

Wang C, Mustafa S. A data-driven Markov Process for Infectious Disease Transmission. PLoS ONE. 2023;18:1–20. https://doi.org/10.1371/journal.pone.0289897.

Zhang L, Lim CY, Maiti T, Li Y, Choi J, Bozoki A, et al. Analysis of conversion of Alzheimer's disease using a multi-state Markov model. Statistical Methods in Medical Research. 2019;28(9):2801–2819. https://doi.org/10.1177/0962280218786525.

Zuhairoh F, Rosadi D, Effendie AR. Multi-state SIRD Model with Semi-Markov Assumptions on the Spread of COVID-19 Disease Based on Sojourn Time Distribution. AIP Conference Proceedings (***Accepted***). 2023;ICMSE2021.

Zuhairoh F, Rosadi D, Effendie AR. Multi-state Discrete-time Markov Chain SVIRS Model on the Spread of COVID-19. Engineering Letters. 2022;30(2):598–608.

Zuhairoh F, Rosadi D, Effendie AR. Continuous-time Hybrid Markov/semi-Markov Model with Sojourn Time Approach in the Spread of Infectious Diseases. IAENG International Journal of Computer Science. 2023 8;50:1108–1114.

Zuhairoh F, Rosadi D, Effendie AR. Prediction of disease outbreaks with the SIR model and Richards model in multi-wave cases. AIP Conference Proceedings. 2023;080005(19751 (1)). https://doi.org/10.1063/5.0181069.

Zuhairoh F, Rosadi D, Effendie AR. SVIRD Epidemic Model with Discrete-time Hybrid Markov/semi-Markov Assumptions. Communications in Mathematical Biology and Neuroscience. 2023 3;2023:1–19. https://doi.org/10.28919/cmbn/7887.

Zuhairoh F, Rosadi D, Ronnie AR. Multi-state SVIRD Model with Continuous-time Markov Chain Assumption on the Spread of Infectious Diseases. Austrian Journal of Statistics. 2024;53(1):70–94. https://doi.org/10.17713/ajs.v53i1.1633.