

## BAYESIAN ESTIMATION METHOD FOR PARAMETER OF EPIDEMIC SIR REED-FROST MODEL

Puji Kurniawan

M0104047

ABSTRACT. Infectious diseases is an important health problem in the most of countries, belonging to Indonesia. Some of infectious diseases can spread in the population and cause epidemic. Someone can be infected if appear adequate contact with infected. One of probabilistic model to describe spread of diseases is epidemic SIR ( Susceptible-Infected-Recovery ) Reed-Frost model. Rate of infection in the SIR Reed-Frost model be defined as probability contact an infected and susceptible ( $p$ ). In this research, the Bayesian method is applied to estimate probability rate of infection. The estimation with Bayesian method need prior distribution and likelihood function. Prior distribution and likelihood function are used to determine posterior distribution. The posterior distribution is used to determine probability rate of infection.

As the result, this research gives conclusion that probability rate of infection is  $p=1-q$ , with

$$E[q]=\hat{q} = \frac{(2n_0 + 2n_1 + n_{21} + \alpha)}{(2n_0 + 3n_1 + 2n_2 + n_{21} + \alpha + \beta)}$$

$\hat{q}$  is expected values of the posterior distribution.

**Key words** : SIR Reed-Frost, Bayesian, Prior, Posterior, Likelihood.

### 1. INTRODUCTION

Avian influence, SARS, HIV/AIDS, dengue, malaria, measles are some infectious diseases which still become important issues with high prevalence and cause epidemic. One of the effort is to overcome those diseases i.e. study the way how the spread diseases happens, therefore it can be conducted its prevent and treatment.

According to Hethcote (2000), one of mathematical model, which describes the spread of diseases i.e. SIR (Susceptible-Infected-Recovery) model. *Susceptible* ( $S$ ) is an individual has a potential to be infected the disease. *Infected* ( $I$ ) is an individual which is infected. *Recovery* ( $R$ ) is an individual which is infected the diseases then becoming recovered. Rate of spread in the SIR Reed-Frost model is defined as probability contact between suspected and infected. *SIR* Reed-Frost model assume the spread of diseases happened in closed and

homogeneous population. The rate of spread depends on contact probability ( $p$ ). The pattern of prediction diseases spread, when maximum infection and the process of diseases infection ended can be known based on parameter value which is applied in the model.

The estimation of parameter can use Bayesian method. Estimation parameter in the Bayesian method used data sample and prior information. Prior information can be used to improve prediction quality or parameter prediction.

## 2. SIR REED-FROST MODEL

To formulate the model, we must determine the variables. Notation  $t$  is period, measured in incubation time. The variables are  $S_t$  is number of susceptible in the population,  $I_t$  is number of infected in the population, and  $R_t$  is number of recovery in the population. According to Term (2007) and Fine (2004), there are some assumptions of the SIR Reed-Frost model

1. infection is spread directly from infected to susceptible only by adequate contact, then infected will become recovery,
2. each individual has a equal contact probability,
3. the population is homogeneous and closed,
4. infections occur independet.

The number of infected at time  $t$  will become recovery in the time  $t+1$ . According to Term (2007), if the  $q$  probability an individual not appear adequate contact then each susceptible has a probability  $q^{I_t}$  because infections occur independently. In the next period some of susceptibles have adequate contact with infected will become infected. The others susceptible will remain in susceptible group. This event be binomial distribution, so the susceptible has probability of avoiding infection given by

$$P(S_{t+1}|S_t, I_t) = \binom{S_t}{S_{t+1}} (1 - q^{I_t})^{I_{t+1}} q^{I_t S_{t+1}}.$$

If the infectious occur during  $m$  period then the number of infected each period will given as a pidemic vector  $(I_0, I_1, \dots, I_m)$ , Term (2007). The infectious each

period occur independently, so the joint probability become infected for  $m$  period is

$$P(I_0, I_1, \dots, I_m) = P(I_m | I_{m-1}, S_{m-1}) \dots P(I_1 | I_0, S_0) P(I_0) \quad (1)$$

with  $P(I_0) = 1$ .

According Fine(2004), Reed-Frost model following the chain of binomial, so the number of susceptible in the next period is expected value for random variable  $S_{t+1}$ , and given by

$$E[S_{t+1}] = \hat{S}_{t+1} = S_t q^{I_t}. \quad (2)$$

The new infected in the next period is some susceptible who have adequate contact with infected. the number of infected in the next period given by

$$I_{t+1} = S_t - S_{t+1} \quad (3)$$

Based on assumption that probability infected will become recovery is one, so the number of recovery in the next period is given by

$$R_{t+1} = R_t + I_t. \quad (4)$$

From equations (2.2), (2.3), and (2.4), Reed-Frost epidemic model is obtained

$$\begin{aligned} S_{t+1} &= S_t q^{I_t} \\ I_{t+1} &= S_t - S_{t+1} \\ R_{t+1} &= R_t + I_t \end{aligned} \quad (5)$$

where  $S_0 > 0, I_0 > 0, R_0 = 0, 0 \leq q \leq 1$ .

### 3. PRIOR DISTRIBUTION

The spread of diseases pattern in the model of SIR Reed-Frost epidemic is a case which Bernoulli distributed, Fine(1977). The determination of prior distribution is based on conjugate prior. Because of the spread of diseases is Bernoulli distributed, it is discrete distribution, while  $q$  is parameter which is distributed to continue ( $q$  runs all dots in the 0 and 1 interval). it needs conjugate from Bernoulli distribution in the continue form. It is beta distribution, Soejoeti and Soebanar(1988). If the data sample follows Bernoulli process and its likelihood function are binominal distribution, therefore the parameter estimation uses Bayesian method needs beta distribution, conjugate from Bernoulli

distribution. Based on the information, prior distribution from SIR Reed-Frost epidemic model is beta distribution. Random variable of  $q$  is distributed beta with  $\alpha$  and  $\beta$ , notated  $q \sim \text{Beta}(\alpha, \beta)$  i.e.

$$p(q, \alpha, \beta) = \frac{1}{B(\alpha, \beta)} q^{\alpha-1} (1-q)^{\beta-1}$$

where  $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ , constant.  $\alpha > 0, \beta > 0$ , and  $0 \leq q \leq 1$ .

Expected value of beta distribution is  $E[q] = \frac{\alpha}{\alpha + \beta}$

#### 4. LIKELIHOOD FUNCTION FOR HOUSEHOLD OF SIZE 3

Let a household of size 3, then the possible paths in epidemic vector are (1), (1, 1), (1, 1, 1) and (1, 2). Using the equation (2.1) with  $p = 1-q$ , we get probability each vector are given on Table.1.

Table 1. Probability each vector

Vector	probability
(1)	$q^2$
(1,1)	$2pq^2$
(1,1,1)	$2p^2q$
(1,2)	$p^2$

If we had a larger population, let observed number of household each vector notated by  $m, n, o, p$  with  $m$  observed number of chains of form (1),  $n$  observed number of chains of form (1,1),  $o$  observed number of chains of form (1,1,1), and  $p$  observed number of chains of form (1,2).

Based on probability each vector, if we have  $n$  household in population then the expected number of household in the population given by

$$E(i) = n * \text{probability of vector } i \text{ occurring}$$

with  $i$  is epidemic vector, then the likelihood function can be constructed by

$$L = E[(1)]^m E[(1,1)]^n E[(1,1,1)]^o E[(1,2)]^p \quad (6)$$

Let  $n_j$  be the number of households where an epidemic of final size  $j$  occurs where  $j=0,1,2$ , and let  $(1,1,1)$  be expressed as  $n_{21}$ . Nothing that  $n_2 = (1,1,1) + (1,2)$  so  $n_2 = (1,1,1) + (1,2)$  and so hence  $n_2 = n_{21} + (1,2)$ , using Equation (3.1), this gives the full condition likelihood function for  $q$  be

$$L(q; n_0, n_1, n_2, n_{21}) = n^{n_0+n_1+n_2} 2^{n_1+n_{21}} q^{2n_1+2n_2+n_{21}} (1-q)^{n_1+2n_2}$$

## 5. POSTERIOR DISTRIBUTION

According to Soejoeti and Soebanar (1988), posterior distribution is equivalent with multiplication likelihood function and prior distribution, so from prior distribution as known and likelihood function obtain, posterior distribution for  $(q | n_0, n_1, n_2, n_{21})$  given by

$$\begin{aligned} p(q|n_0, n_1, n_2, n_{21}) &\approx L(q; n_0, n_1, n_2, n_{21})p(q) \\ &\approx (n^{n_0+n_1+n_2} 2^{n_1+n_{21}} q^{2n_1+2n_2+n_{21}} (1-q)^{n_1+2n_2})x \frac{1}{B(\alpha, \beta)} q^{\alpha-1} (1-q)^{\beta-1} \\ &\approx \frac{n^{n_0+n_1+n_2} 2^{n_1+n_{21}}}{B(\alpha, \beta)} q^{2n_0+2n_1+n_{21}+\alpha-1} (1-q)^{n_1+2n_2+\beta-1} \\ &\approx \text{Beta}(2n_0 + 2n_1 + n_{21} + \alpha, n_1 + 2n_2 + \beta) \end{aligned}$$

with parameter  $a = (2n_0 + 2n_1 + n_{21} + \alpha)$  and  $b = (n_1 + 2n_2 + \beta)$

According to Bain and Engelhardt (1992), estimation of  $q$  constitute expected values from posterior distribution, given by

$$\begin{aligned} E(q) = \hat{q} &= \frac{a}{a+b} \\ &= \frac{(2n_0 + 2n_1 + n_{21} + \alpha)}{(2n_0 + 3n_1 + 2n_2 + n_{21} + \alpha + \beta)} \end{aligned} \quad (7)$$

consider  $n_0, n_1, n_2, n_{21}$  known.

## 6. EXAMPLE

Refer to investigation into measles in Rhode Island, O'Neill and Roberts (1999). There are only using the data on household of size 3. All members of the population are aged between 7 months and 10 years, the data have a population of size  $n = 334$  household. The result shows in Tabel.2.

Table.2. Number household for each vector

Type of chain	Expected number	Household number
( 1 )	$nq^2$	34
( 1,1 )	$2npq^2$	25
( 1,1,1 )	$2np^2q$	239
( 1,2 )	$np^2$	36

Based on Tabel.2,  $n_0 = 34$ ,  $n_1 = 25$ ,  $n_2 = 36 + 239 = 275$ ,  $n_{21} = 36$ . Using equation (7) estimation parameter of  $q$  obtained  $\hat{q} = 0.2125$ , so  $\hat{p} = 0.7875$ . Based on  $\hat{q}$  spread of measeles in Rhode island occur with probability contact  $\hat{p} = 0.7875$ . With equations (5), SIR Reed-Frost epidemic model is obtained  $S_{t+1} = (0.2125)^t S_t$ ,  $I_{t+1} = S_t - S_{t+1}$ , and  $R_{t+1} = R_t + I_t$ .

Each household in this investigation of size 3, so number of individu in the population is 1002. Prediction spread of measeles using SIR Reed Frost model with  $I_0 = 1$  shows in Tabel.3.

Tabel.3. Number spread of diseases

Periode	t	S	I	R
1	0	1001	1	0
2	1	213	788	1
3	2	0	213	789
4	3	0	0	1002

Tabel 4.3 shows that spread prediction measles every period. Maximum infection happens in the second periode, while final epidemic ends after the fourth period. Comparison number of infected each period for different  $I_0$  can be seen in the Figure.1.

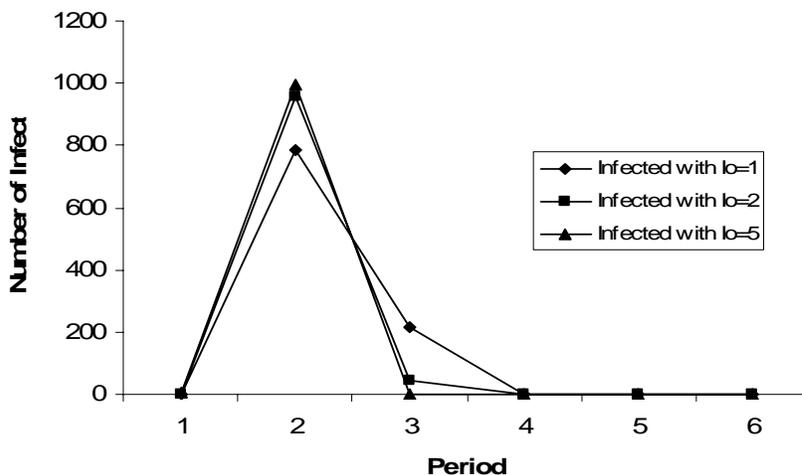


Figure.1. The spread of infected

### 7. PROBABILITY VALUE FOR DIFFERENT $n_{21}$

Spread of diseases in SIR Reed-Frost model depend on first infected and probability value. Probability contact obtained based on equation 6. Used the data investigation into measles in Rhode island with number of husehold  $n=334$ , can be made simulation for number household with vector  $(1,1,1)$ , notated  $n_{21}$ , are diferent. Comparison probability value of contact shown in tabel 4.

Table.4. Probability value for different  $n_{21}$

$n_0$	$n_1$	$n_{21}$	$n_2$	$\hat{q}$	$\hat{p}$
34	25	0	275	0.170274	0.829726
34	25	36	275	0.211248	0.788752
34	25	137	275	0.307229	0.692771
34	25	239	275	0.383047	0.616953
34	25	275	275	0.405992	0.594008

Table 4 show that more and more spread of diseases with vektor (1,1,1) show the smaller the probability contact between infected with susceptible.

## 8. CONCLUSION

1. The estimation of *SIR* Reed-Frost model uses Bayesian method with household of size 3 is obtained its equation i.e.

$$\hat{q} = \frac{(2n_0 + 2n_1 + n_{21} + \alpha)}{(2n_0 + 3n_1 + 2n_2 + n_{21} + \alpha + \beta)}$$

with  $n_0, n_1, n_2, n_{21}$  is known.

2. Based on observation numeric with household of size 3, spread of diseases depend on first infected and probability contact in population. largest probability contact occur if the number of household that epidemic  $n_{21}$  smallest.

## REFERENCES

- Bain, L. J. and M. Engelhardt, *Introduction to probability and mathematical Statistics*, 2 ed., Duxbury Press, California, 1992.
- Fine, P. E. M., *A commentary on the Reed-Frost epidemic model*. American journal of Epidemiology 106 (1977), 87-100.
- Hethcote, H. W., *The mathematics of infectious diseases*, [www.siam.org/journals/sirev/42-4/37190.html](http://www.siam.org/journals/sirev/42-4/37190.html) 42 (2000), 599-653.
- Philip O'Neil, P. and Gareth Roberts. M, *Bayesian Inference of fo Partially observed Stochastic Epidemics*, J.R. Statist. Soc A (1999), 162. part1. pp121-129.
- Soejoeti, Z. dan Soebanar, *Inferensi bayesian*, Universitas Terbuka, Karunika, Jakarta, 1988.
- Term, H., *Using probabilistic models to infer infection rates in viral outbreaks*. [http : //mathgen.stats.ox.ac.uk/bioinformatics/infection\\_rates.Pdf](http://mathgen.stats.ox.ac.uk/bioinformatics/infection_rates.Pdf) (2007).