10.3 Parameter Estimation in the Binomial SIR Epidemic Model

Consider the system of difference equations in the binomial SIR model:

$$\begin{cases} S(j+1) = S(j) - X, \\ \\ I(j+1) = I(j) + X - Y, \quad , \quad j = 0, 1, 2, \dots, \\ \\ R(j+1) = R(j) + Y. \end{cases}$$

where $X \sim Bi(S(j), \beta\, I(j)/N)$ and $Y \sim Bi(I(j), \gamma)$. Taking expected values, we arrive at the same system of difference equations as in the deterministic SIR model:

$$\begin{cases} S(j+1) = S(j) - \beta\, \dfrac{S(j)\, I(j)}{N}, \\ \\ I(j+1) = I(j) + \beta\, \dfrac{S(j)\, I(j)}{N} - \gamma I(j), \quad , \quad j = 0, 1, 2, \dots. \\ \\ R(j+1) = R(j) + \gamma I(j). \end{cases}$$

Denoting the increments by $\Delta S(j) = S(j+1) - S(j), \Delta I(j) = I(j+1) - I(j)$, and $\Delta R(j) = R(j+1) - R(j)$, we can rewrite the equations as

$$\begin{cases} \dfrac{\Delta S(j)}{S(j)I(j)} \;=\; -\,\dfrac{\beta}{N}, \\ \\ \dfrac{\Delta I(j)}{I(j)} \;=\; \dfrac{\beta}{N}\, S(j) - \gamma, \qquad j = 0, 1, 2, \dots. \\ \\ \dfrac{\Delta R(j)}{I(j)} \;=\; \gamma. \end{cases}$$

From these equations, it follows that we have two ways to estimate the parameters β and γ. The first one is the **method of moments**, according to which

$$\widehat{\beta}_{MM} \;=\; -\,N \left(\text{sample mean of } \frac{\Delta S(j)}{S(j)I(j)}\right),$$

and

$$\widehat{\gamma}_{MM} \;=\; \text{sample mean of } \frac{\Delta R(j)}{I(j)}.$$

The second method of estimation is the **regression method** where we regress $\dfrac{\Delta I(j)}{I(j)}$ on $S(j)$. The estimators of the parameters can be obtained as

$$\widehat{\beta}_R = N\,\widehat{\text{slope}}, \quad \text{and} \quad \widehat{\gamma}_R = -\,\widehat{\text{intercept}}.$$

Remark. In the deterministic SIR model, these estimators are always exactly equal to the true values of the parameters. □

Below we show how to estimate the parameters in SAS and R.

In SAS: Here we use the data simulated above (with seed=809197, N=1000, beta=0.7, gamma=0.2, and time=30).

```
data SIR;
set SIR;
lag_S=lag(S);
beta_est=(S-lag(S))/(lag(S)*lag(I));
gamma_est=(R-lag(R))/lag(I);
response_var=(I-lag(I))/lag(I);
if beta_est ne .;
run;

proc sql;
 select -1000*mean(beta_est) as beta_hatMM,
 mean(gamma_est) as gamma_hatMM
  from SIR;
quit;
```

beta_hatMM	gamma_hatMM
0.685137	0.185146

```
proc glm data=SIR;
model response_var=lag_S;
run;
```

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	-.2151709787	0.05347189	-4.02	0.0004
lag_S	0.0007667129	0.00007969	9.62	<.0001

From this output, $\beta_R = 1000 \cdot \widehat{\text{slope}} = 0.76506$ and $\gamma_R = -\widehat{\text{intercept}} = 0.21517$.

In R: Here we use the data simulated earlier (with seed=191962, N=1000, beta=0.7, gamma=0.2, and time=30).

```
simdata<- data.frame(time, S, I, R)

simdata$S.delta<- c(simdata$S[-1],0)-simdata$S
simdata$I.delta<- c(simdata$I[-1],0)-simdata$I
simdata$R.delta<- c(simdata$R[-1],0)-simdata$R
simdata<- simdata[-nrow(simdata),]

for (j in 1:nrow(simdata)) {
simdata$beta.est[j]<- simdata$S.delta[j]/(simdata$S[j]*simdata$I[j])
    simdata$gamma.est[j]<- simdata$R.delta[j]/simdata$I[j]
      simdata$response.var[j]<- simdata$I.delta[j]/simdata$I[j]
}

#estimating using the method of moments
print(beta.hatMM<- -1000*mean(simdata$beta.est))
```

```
0.6978012
```

```
print(gamma.hatMM<- mean(simdata$gamma.est))
```

```
0.1924247
```

```
#estimating using the regression method
reg<- glm(response.var ~ S, data=simdata)
reg.summary <- summary(reg)
```

print(beta.hatR<- N*reg.summary$coefficients[2,1])

```
0.7587301
```

print(gamma.hatR<- -reg.summary$coefficients[1,1])

```
0.2220625
```

□