

A Bayesian approach to analyse overdispersed longitudinal count data

Fernanda B. Rizzato^a Roseli A. Leandro^b Clarice G.B. Demétrio^b
Geert Molenberghs^c

^a Universidade Federal do Paraná, Caixa Postal 19081, Curitiba, Paraná, Brazil

^b ESALQ, Piracicaba, São Paulo, Brazil

^c I-BioStat, Hasselt University & KU Leuven, Belgium

Supplementary Materials

Appendix A. Review of Basic Concepts

Generalized linear models [20] have unified the regression methodology for a wide variety of discrete and continuous responses that can be assumed to be independent. Considering the random variables Y_1, \dots, Y_n i.i.d., with vector of means $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)$, the three components of a GLM are: (i) the response variable Y_i stemming from the exponential family of distributions with a density of the form:

$$f(y_i; \theta_i, \phi) = \exp\{\phi^{-1}[y_i\theta_i - b(\theta_i)] + c(y_i, \phi)\},$$

with mean $E(Y_i) = \mu_i = b'(\theta_i)$ and variance $\text{Var}(Y_i) = \phi b''(\theta_i)$, where θ_i is the canonical parameter and $\phi > 0$ is the dispersion parameter; (ii) a linear predictor vector $\boldsymbol{\eta}$ given by $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$, where $\boldsymbol{\beta}$ is a vector of p unknown parameters and $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^T$ is an $n \times p$ design matrix; (iii) a link function $g(\cdot)$ relating the mean to the linear predictor, i.e., $\eta_i = g(\mu_i)$ and $\mu_i = g^{-1}(\mathbf{x}_i^T \boldsymbol{\beta})$.

Our focus in this paper is on the particular GLM with Poisson distribution and log link function. Assuming that the random variables Y_i represent counts with means λ_i , the standard Poisson model assumes that $Y_i \sim \text{Poisson}(\lambda_i)$ with probability distribution

$$f_{Y_i}(y_i) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!}, \quad y_i = 0, 1, 2, \dots$$

where $\lambda_i = \exp(\mathbf{x}_i^T \boldsymbol{\beta})$. Parameter estimation conventionally proceeds by maximum likelihood using the iteratively re-weighted least square algorithm.

The Poisson model is based on two strong assumptions: i) independence among measurements and ii) the same mean λ for all individuals treated equally. The failure of one or both of these basic assumptions will induce the observed variance to differ from the mean, the theoretical variance of the Poisson distribution i.e., the aforementioned *overdispersion*. Hinde and Demétrio [8] provide general treatments of overdispersion for univariate data. Underdispersion is possible as well in some settings (e.g., when there is competition for scarce resources among cluster members).

Considering a Poisson model, if the residual deviance is much larger than its associated degrees of freedom, there is evidence of lack-of-fit of the model and probably a problem of overdispersion. An initial exploratory analysis of the epilepsy data suggested two options.

The first option is, for each individual, to consider a Poisson model with a linear predictor $\eta = \log(\mu) = \beta_0 + \beta_1 t$ fitted to the counts, resulting in 45 analyses for

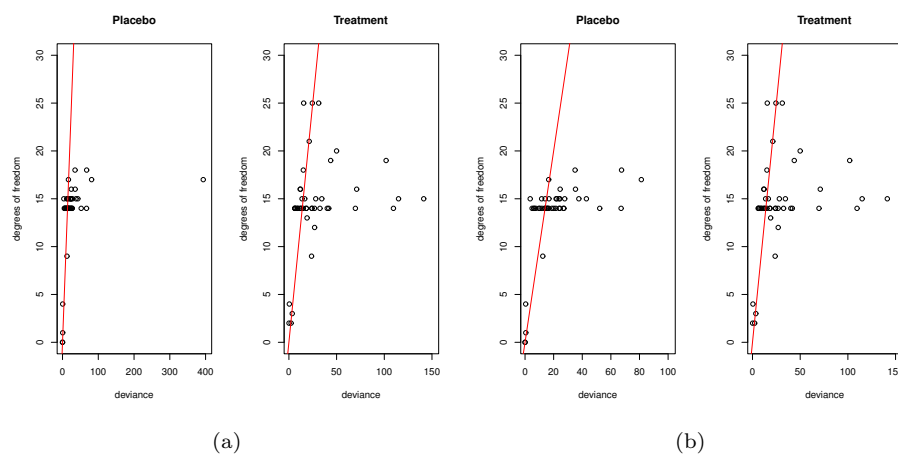


Figure A.1. (a) Values of deviances against its degrees of freedom considering a log-linear Poisson model for each individual. (b) Values of deviances against its degrees of freedom (Δ) considering a log-linear Poisson model for each individual, with some individuals removed.

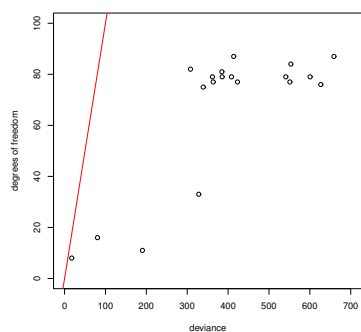


Figure A.2. Values of deviances against its degrees of freedom considering a log-linear Poisson model for each time point.

placebo and 44 analyses for treatment. The values for the deviances are plotted against their degrees of freedom in Figure A.1a, with a close look eliminating some individuals in Figure A.1b. As expected, because of the possible correlation between measurements, we can see that the log-linear Poisson model does not fit to the data. If the models fitted well, the points would nicely scatter around a straight line going through the origin with a slope equal to 1 (confirming that the deviance would equal the degrees-of-freedom).

The second option is, for each time point, to consider a Poisson model with a linear predictor $\eta = \log(\mu) = \beta_0 + \alpha_i$, $i = 1$ (placebo) and $i = 2$ (treatment). The values for the deviances are plotted against their degrees-of-freedom in Figure A.2. The same comment as regarding the previous figure applies. These analysis show the need for two different types of random effects for the analysis of the complete data set: (i) one to accommodate in the model the correlation among measurements in the same individual and (ii) to incorporate in the model the variability among individuals.

Appendix B. OpenBugs Code

B.1 Model 1: Poisson

```

model {
  for (i in 1 : N) {
    for(j in 1 : M){
      y[i,j] ~ dpois(lambda[i,j])
      lambda[i,j] < -exp((int0+slope0*t[i,j])*(1-trt[i])+
        (int1+slope1*t[i,j])*(trt[i]))
    }
  }
  int0 ~ dnorm(0.0,0.000001)
  slope0 ~ dnorm(0.0,0.000001)
  int1 ~ dnorm(0.0,0.000001)
  slope1 ~ dnorm(0.0,0.000001)
  ds < - slope1-slope0
  rs < - slope1/slope0
}

```

B.2 Model 2: Poisson-Gamma

```

model {
  for (i in 1 : N) {
    for(j in 1 : M){
      y[i,j] ~ dpois(lambda[i,j])
      lambda[i,j] < -teta[i,j]*exp((int0+slope0*t[i,j])*(1-trt[i])+
        (int1+slope1*t[i,j])*(trt[i]))
      teta[i,j] ~ dgamma(alpha,alpha)
    }
  }
  int0 ~ dnorm(0.0,0.000001)
  slope0 ~ dnorm(0.0,0.000001)
  int1 ~ dnorm(0.0,0.000001)
  slope1 ~ dnorm(0.0,0.000001)
  alpha < -1/beta
  beta ~ dgamma(0.001,0.001)
  ds < - slope1-slope0
  rs < - slope1/slope0
}

```

B.3 Model 3: Poisson-Normal

```

model {
  for (i in 1 : N) {
    for(j in 1 : M){
      y[i,j] ~ dpois(lambda[i,j])
      lambda[i,j] < -exp((int0+b[i]+slope0*t[i,j])*(1-trt[i])+
        (int1+b[i]+slope1*t[i,j])*(trt[i]))
    }
  }
}

```

S.4

Taylor & Francis and I.T. Consultant

```

        b[i] ~ dnorm(0,tau)
    }
D < - 1 / tau
tau ~ dgamma(0.0001,0.0001)
int0 ~ dnorm(0.0,0.000001)
slope0 ~ dnorm(0.0,0.000001)
int1 ~ dnorm(0.0,0.000001)
slope1 ~ dnorm(0.0,0.000001)
ds < - slope1-slope0
rs < - slope1/slope0
}

```

B.4 Model 4: Poisson-Normal-Gamma

```

model {
  for (i in 1 : N) {
    for(j in 1 : M){
      y[i,j] ~ dpois(lambda[i,j])
      lambda[i,j] < -teta[i,j]*exp((int0+b[i]+slope0*t[i,j])*(1-trt[i])+
      (int1+b[i]+slope1*t[i,j])*(trt[i]))
      teta[i,j] ~ dgamma(alpha,alpha)
    }
    b[i] ~ dnorm(0,tau)
  }
alpha < -1/beta
beta ~ dgamma(0.001,0.001)
D < - 1 / tau
tau ~ dgamma(0.0001,0.0001)
int0 ~ dnorm(0.0,0.000001)
slope0 ~ dnorm(0.0,0.000001)
int1 ~ dnorm(0.0,0.000001)
slope1 ~ dnorm(0.0,0.000001)
ds < - slope1-slope0
rs < - slope1/slope0
}

```

Appendix C. Simulation Results

a. Simulations from a Poisson-gamma model

Simulated data set 1–20: a) Histogram estimate of the posterior distribution of measures S under Poisson model and the corresponding marginal χ^2 distributions. b) Histograms estimate of the posterior distribution of measures S and S_θ under Poisson-gamma model and the corresponding marginal χ^2 distributions. c) Histograms estimate of the posterior distribution of measures S , S_θ and S_b under combined model and the corresponding marginal χ^2 distributions.

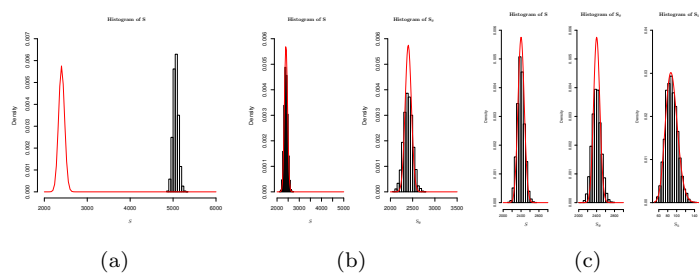


Figure C.1. Simulated data set 1.

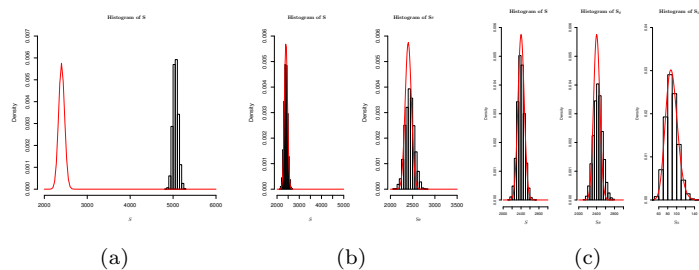


Figure C.2. Simulated data set 2.

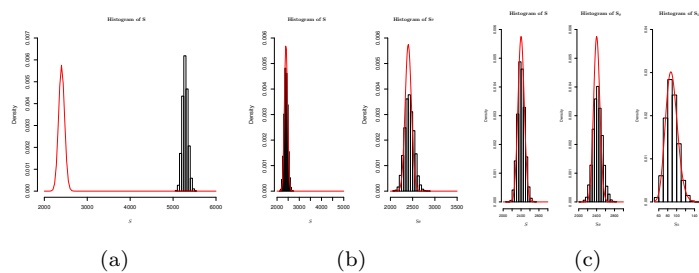


Figure C.3. Simulated data set 3.

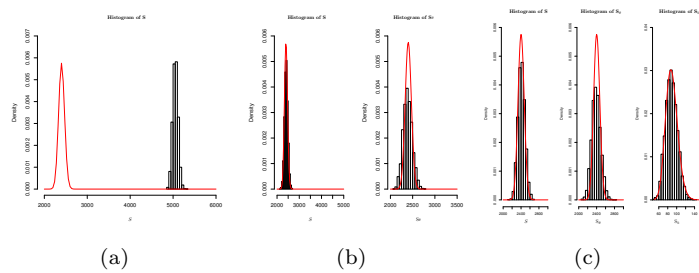


Figure C.4. Simulated data set 4.

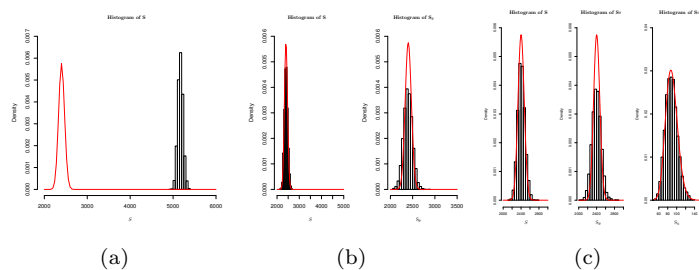


Figure C.5. Simulated data set 5.

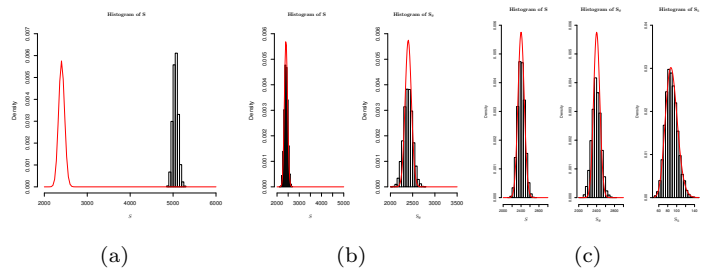


Figure C.6. Simulated data set 6.

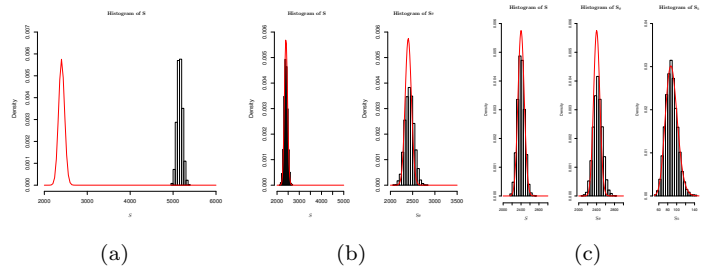


Figure C.7. Simulated data set 7.

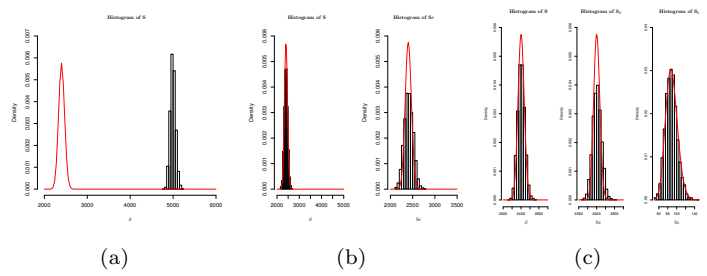


Figure C.8. Simulated data set 8.

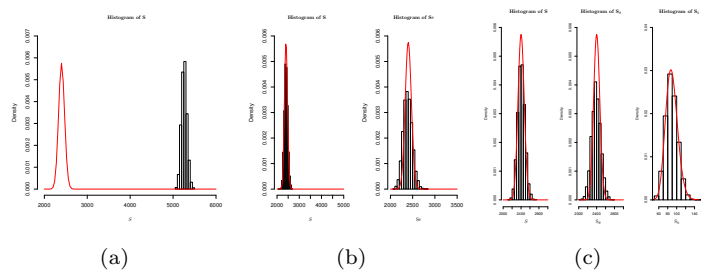


Figure C.9. Simulated data set 9.

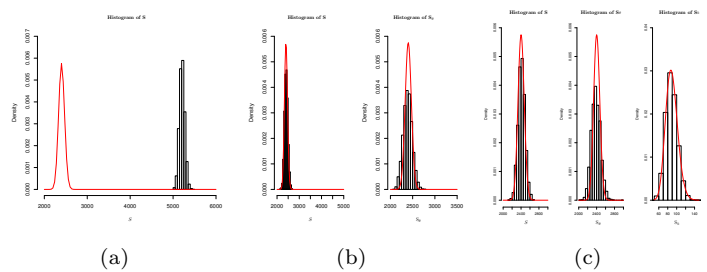


Figure C.10. Simulated data set 10.

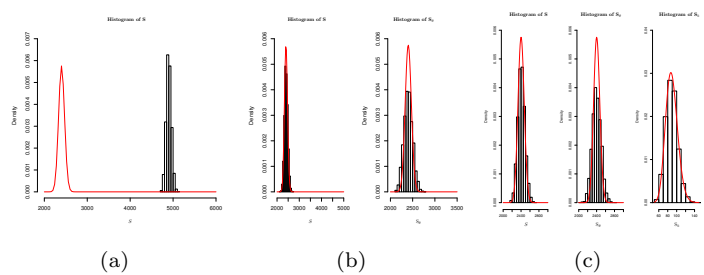


Figure C.11. Simulated data set 11.

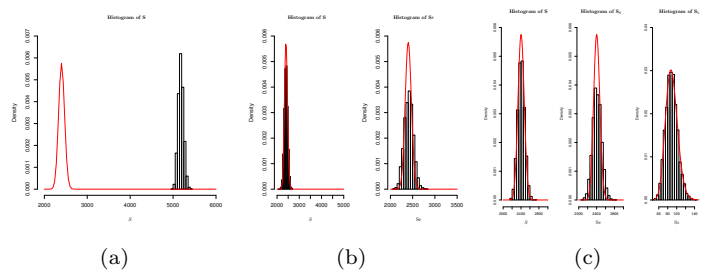


Figure C.12. Simulated data set 12.

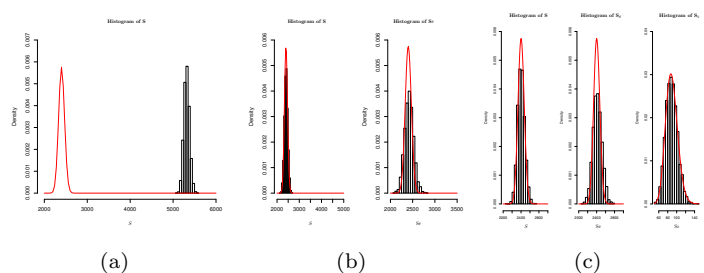


Figure C.13. Simulated data set 13.

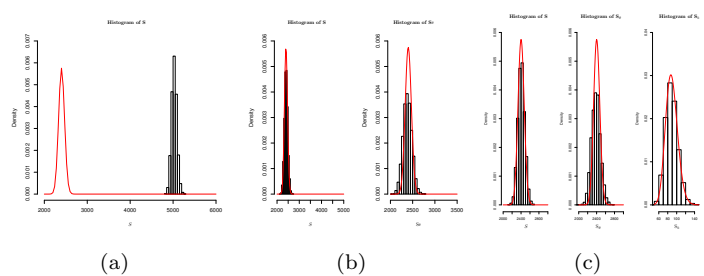


Figure C.14. Simulated data set 14.

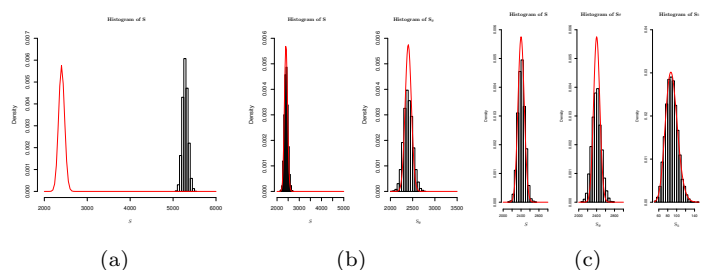


Figure C.15. Simulated data set 15.

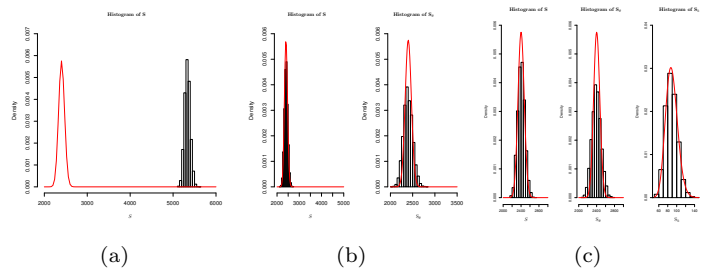


Figure C.16. Simulated data set 16.

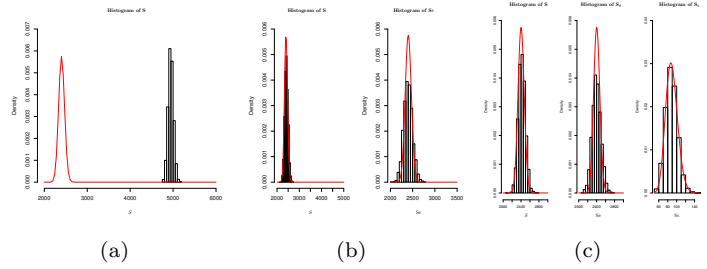


Figure C.17. Simulated data set 17.

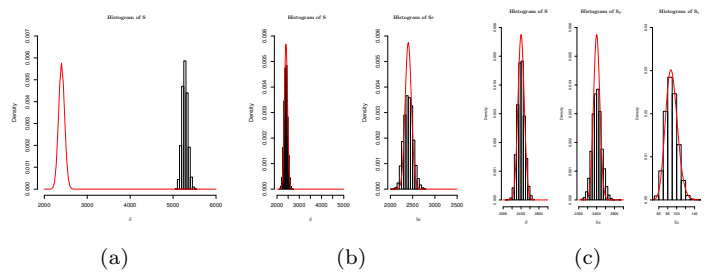


Figure C.18. Simulated data set 18.

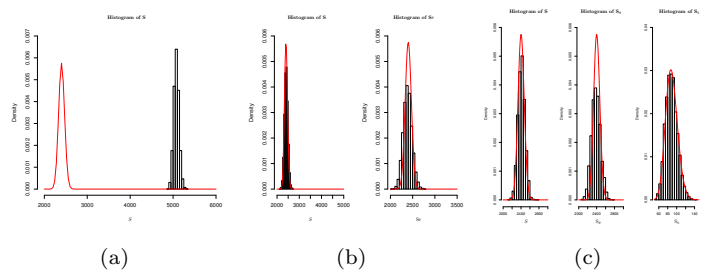


Figure C.19. Simulated data set 19.

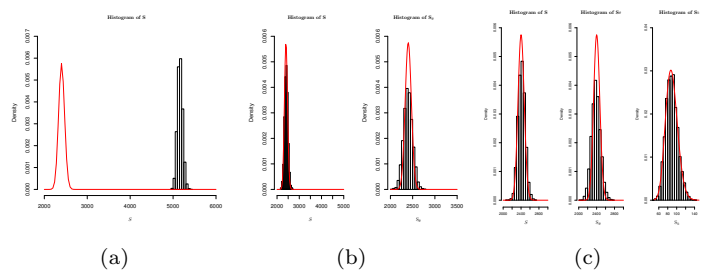


Figure C.20. Simulated data set 20.

Table C.1. DIC criterion for Poisson, Poisson-gamma and Poisson-normal-gamma models obtained using the OpenBugs software for each simulated data set (simulated from Poisson-gamma model).

Simulation	DIC		
	Poisson	Poisson-gamma	Poisson-normal-gamma
1	11070	9622	9622
2	11000	9634	9635
3	11170	9633	9634
4	11020	9575	9576
5	11130	9609	9610
6	11060	9615	9615
7	11050	9613	9614
8	10940	9577	9579
9	11180	9569	9570
10	11160	9589	9590
11	10900	9588	9589
12	11140	9642	9644
13	11180	9603	9603
14	11000	9556	9558
15	11250	9653	9654
16	11140	9501	9502
17	10930	9595	9596
18	11050	9461	9463
19	11080	9600	9601
20	11190	9675	9676

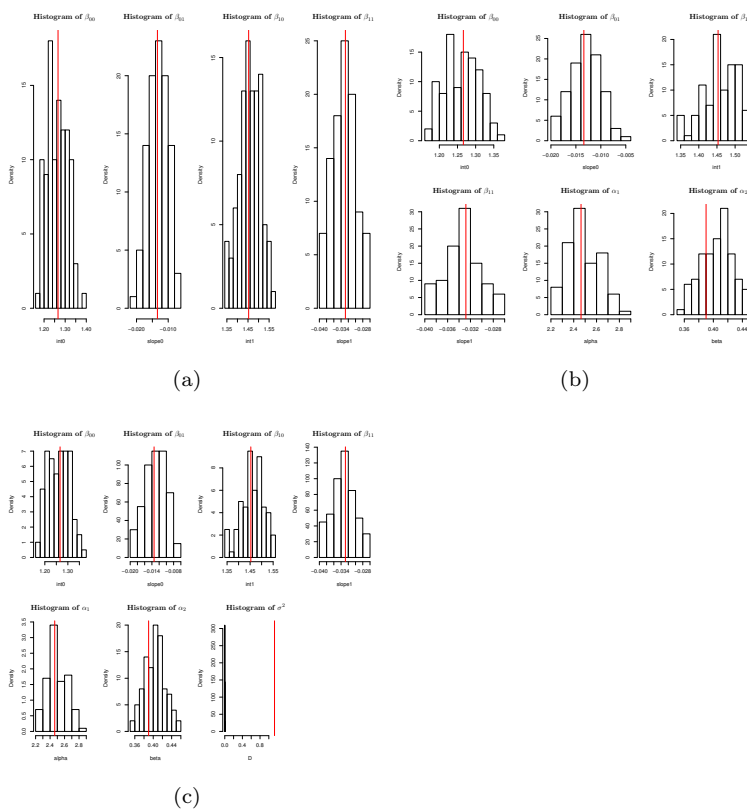


Figure C.21. a) Histograms for each parameter of Poisson model of posterior summaries for the 100 simulated data set and the respective simulated value for each parameter (red line). b) Histograms for each parameter of Poisson-gamma model of posterior summaries for the 100 simulated data set and the respective simulated value for each parameter (red line). c) Histograms for each parameter of Poisson-normal-gamma model of posterior summaries for the 100 simulated data set and the respective simulated value for each parameter (red line).

Table C.2. DIC criterion for Poisson, Poisson-normal and Poisson-normal-gamma models obtained using the OpenBugs software for each simulated data set (simulated from Poisson-normal model).

Simulation	DIC		
	Poisson	Poisson-gamma	Poisson-normal-gamma
1	35510	9068	9068
2	24660	9507	9508
3	23920	8869	8870
4	23520	9288	9290
5	33810	9182	9183
6	23990	8827	8828
7	18440	9557	9558
8	19270	9006	9006
9	19720	8882	8883
10	19880	8980	8979
11	19780	9173	9175
12	24630	9007	9002
13	20980	9134	9135
14	25050	9171	9171
15	26810	9020	9021
16	17790	9273	9275
17	16110	8914	8915
18	28100	9191	9190
19	23090	9126	9127
20	24410	9033	9034

b. Simulations from a Poisson-normal model

Simulated data set 1–20: a) Histogram estimate of the posterior distribution of measures S under Poisson model and the corresponding marginal χ^2 distributions. b) Histograms estimate of the posterior distribution of measures S and S_b under Poisson-normal model and the corresponding marginal χ^2 distributions. c) Histograms estimate of the posterior distribution of measures S , S_θ and S_b under combined model and the corresponding marginal χ^2 distributions.

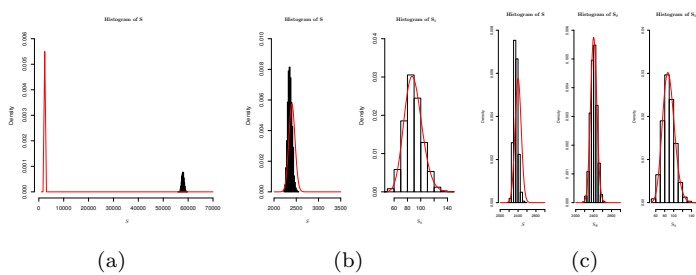


Figure C.22. Simulated data set 1.

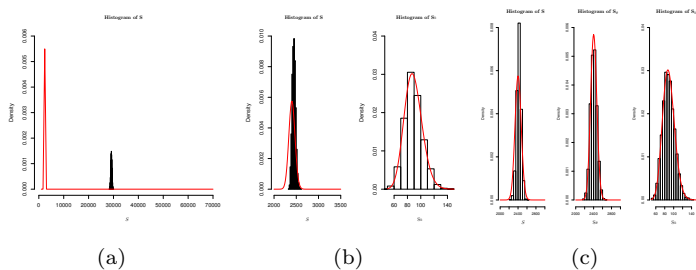


Figure C.23. Simulated data set 2.

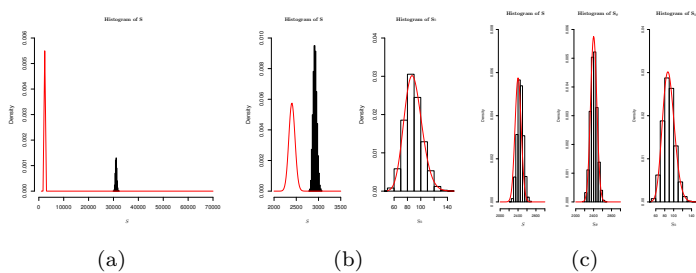


Figure C.24. Simulated data set 3.

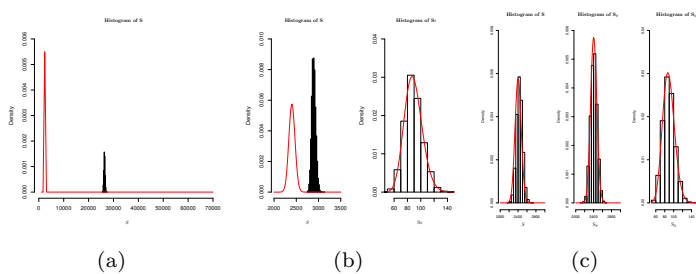


Figure C.25. Simulated data set 4.

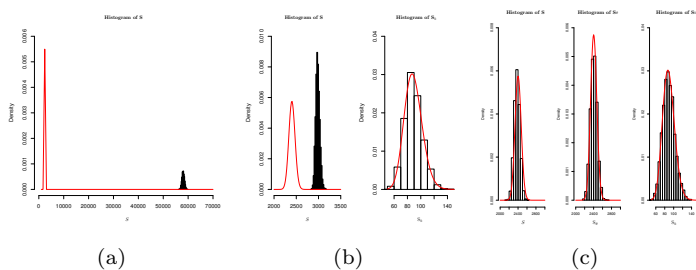


Figure C.26. Simulated data set 5.

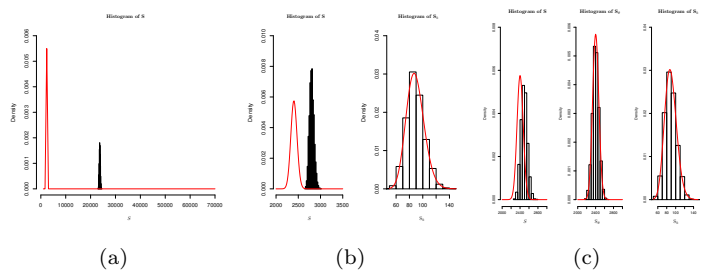


Figure C.27. Simulated data set 6.

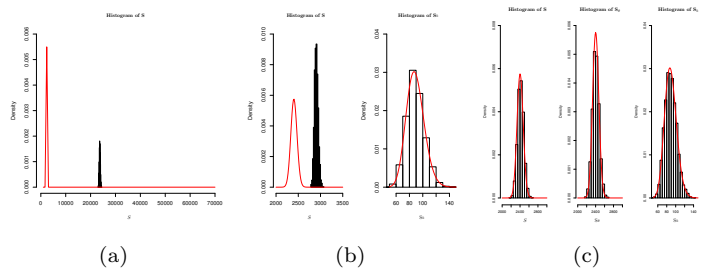


Figure C.28. Simulated data set 7.

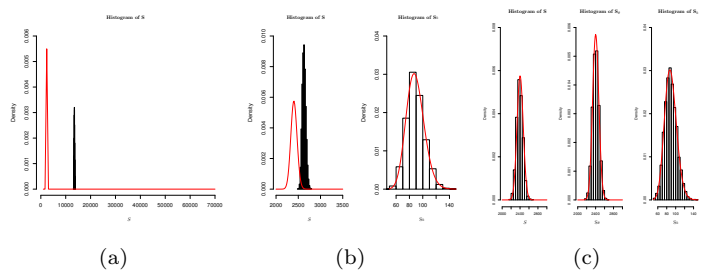


Figure C.29. Simulated data set 8.

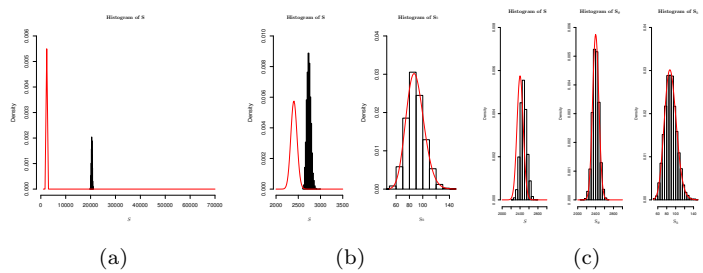


Figure C.30. Simulated data set 9.

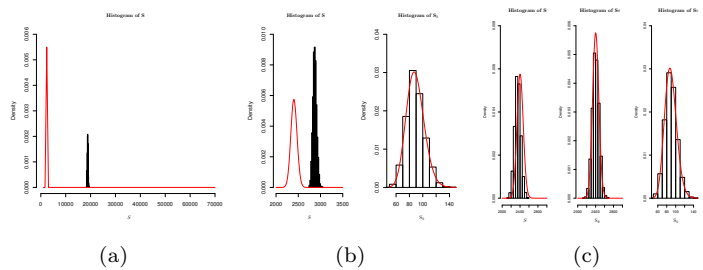


Figure C.31. Simulated data set 10.

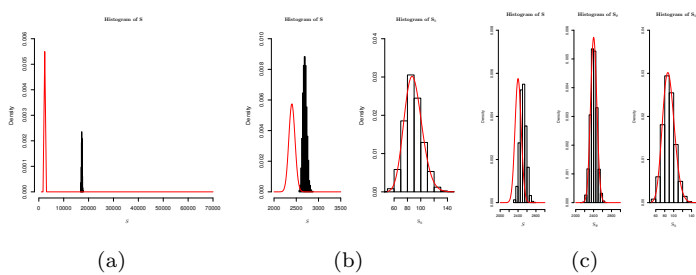


Figure C.32. Simulated data set 11.

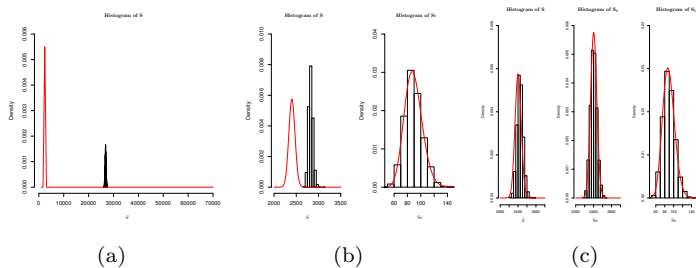


Figure C.33. Simulated data set 12.

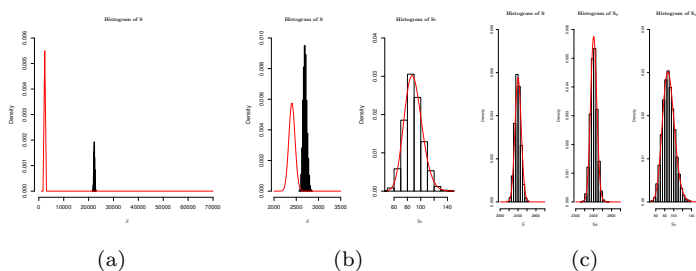


Figure C.34. Simulated data set 13.

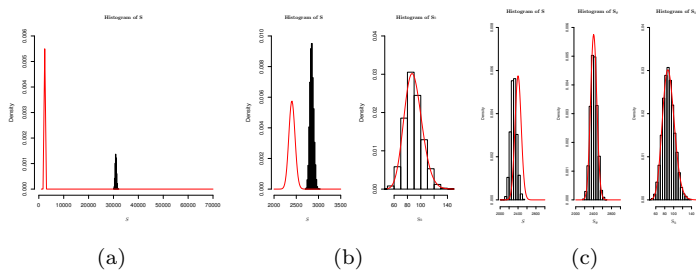


Figure C.35. Simulated data set 14.

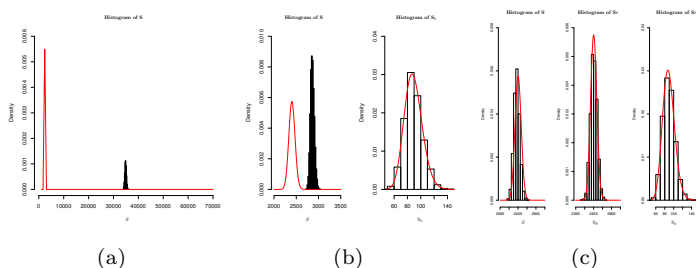


Figure C.36. Simulated data set 15.

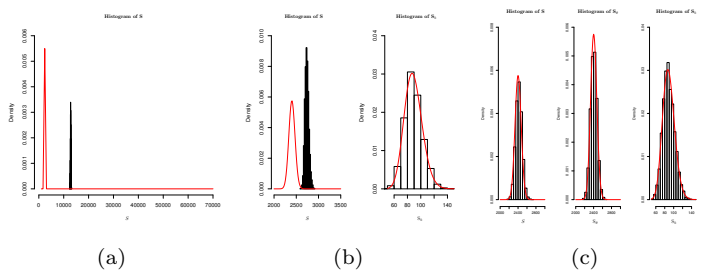


Figure C.37. Simulated data set 16.

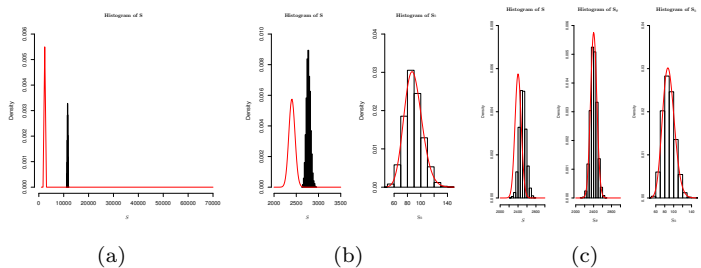


Figure C.38. Simulated data set 17.

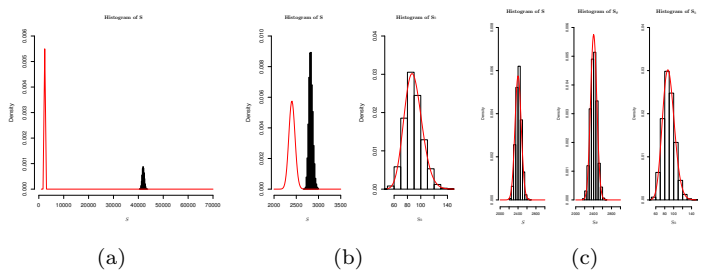


Figure C.39. Simulated data set 18.

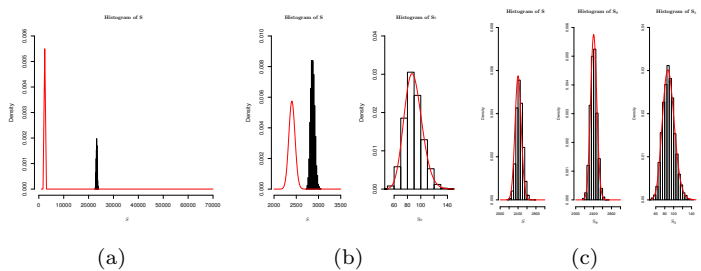


Figure C.40. Simulated data set 19.

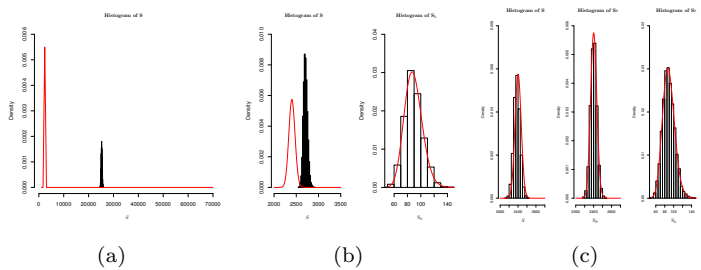


Figure C.41. Simulated data set 20.

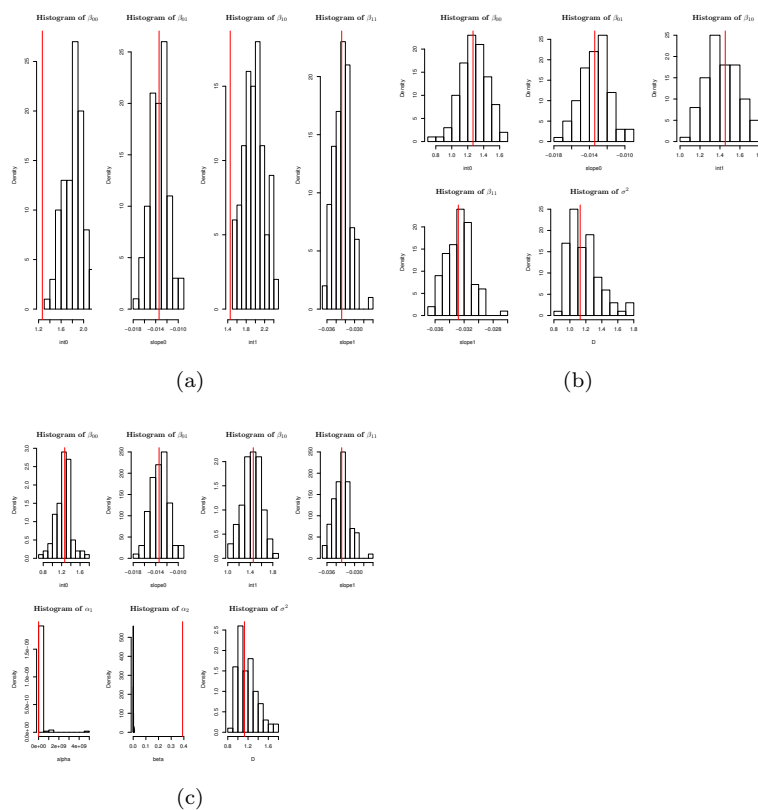


Figure C.42. a) Histograms for each parameter of Poisson model of posterior summaries for the 100 simulated data set and the respective simulated value for each parameter (red line). b) Histograms for each parameter of Poisson-normal model of posterior summaries for the 100 simulated data set and the respective simulated value for each parameter (red line). c) Histograms for each parameter of Poisson-normal-gamma model of posterior summaries for the 100 simulated data set and the respective simulated value for each parameter (red line).

Table C.3. DIC criterion for Poisson, Poisson-gamma, Poisson-normal and Poisson-normal-gamma models obtained using the OpenBugs software for each simulated data set (simulated from Poisson-normal-gamma model).

Simulation	DIC			
	Poisson	Poisson-gamma	Poisson-normal	Poisson-normal-gamma
1	24140	10180	13340	9963
2	30500	9956	13770	9712
3	19650	10060	12570	9785
4	23150	10120	12950	9853
5	25770	10090	13180	9791
6	32530	9888	13820	9695
7	33110	10190	14420	10010
8	24980	9797	12800	9549
9	27380	9704	13000	9486
10	20790	9370	11670	9137
11	21330	9424	11750	9236
12	22370	9905	12710	9687
13	21580	9800	12250	9543
14	23600	9993	13320	9814
15	26790	10380	13680	10130
16	31520	10250	13690	9971
17	26580	10050	13230	9788
18	23790	9473	11990	9252
19	28830	9913	13440	9683
20	21020	9828	12350	9591

c. Simulations from the combined model

Simulated data set 1–20: a) Histogram estimate of the posterior distribution of measures S under Poisson model and the corresponding marginal χ^2 distributions. b) Histograms estimate of the posterior distribution of measures S and S_θ under poisson-gamma model and the corresponding marginal χ^2 distributions. c) Histograms estimate of the posterior distribution of measures S and S_b under Poisson-normal model and the corresponding marginal χ^2 distributions. d) Histograms estimate of the posterior distribution of measures S , S_θ and S_b under combined model and the corresponding marginal χ^2 distributions.

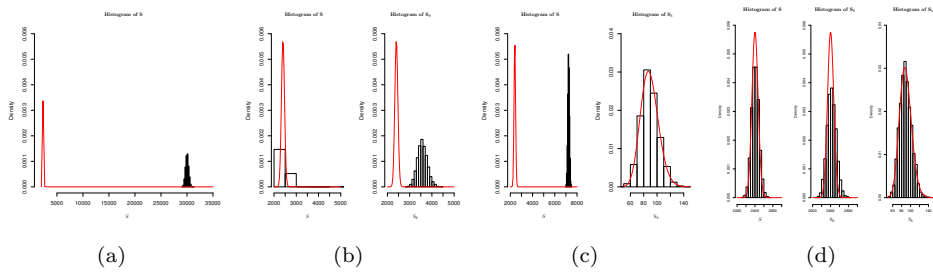


Figure C.43. Simulated data set 1.

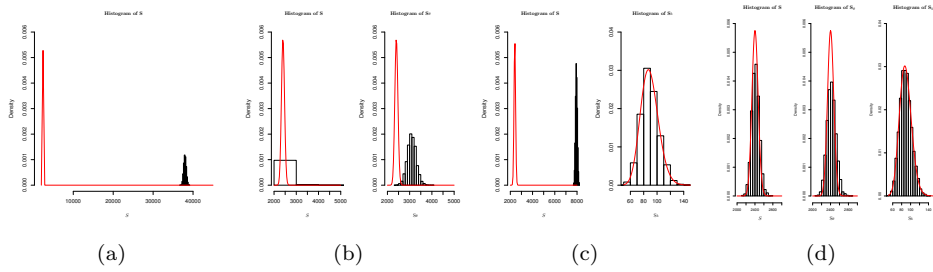


Figure C.44. Simulated data set 2.

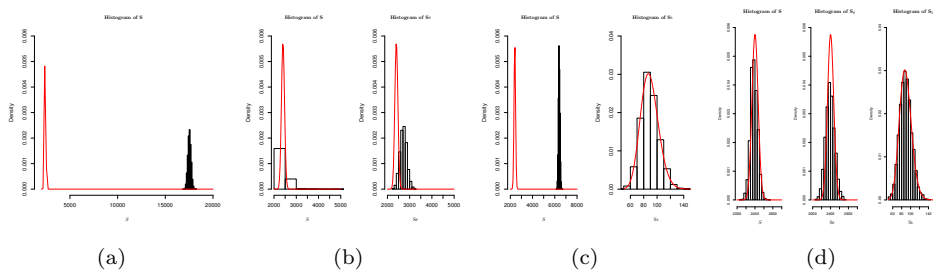


Figure C.45. Simulated data set 3.

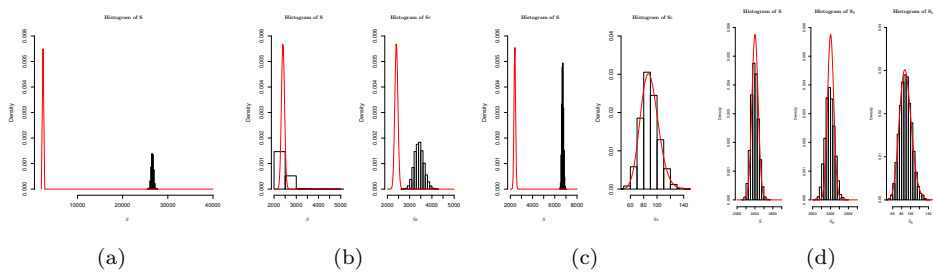


Figure C.46. Simulated data set 4.

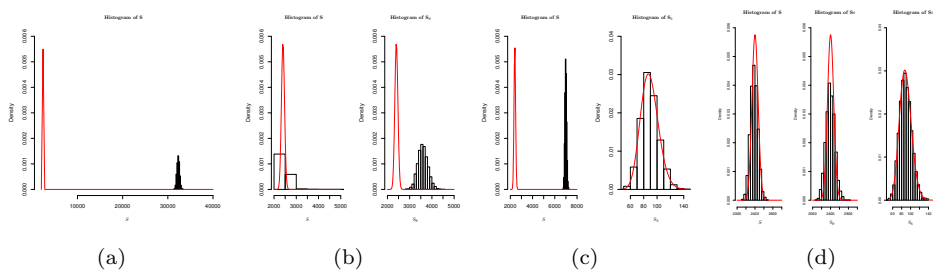


Figure C.47. Simulated data set 5.

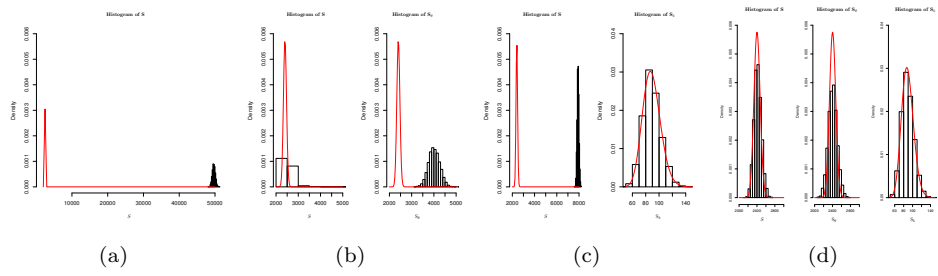


Figure C.48. Simulated data set 6.

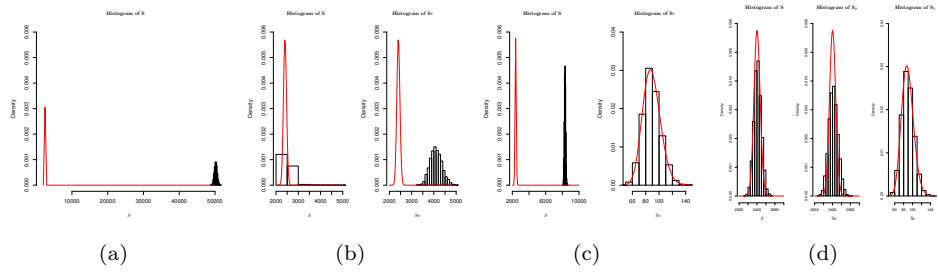


Figure C.49. Simulated data set 7.

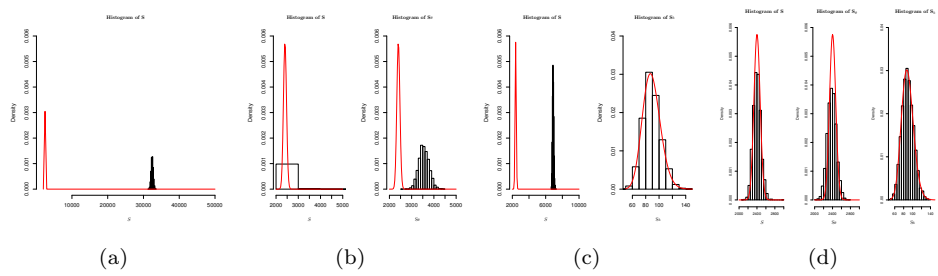


Figure C.50. Simulated data set 8.

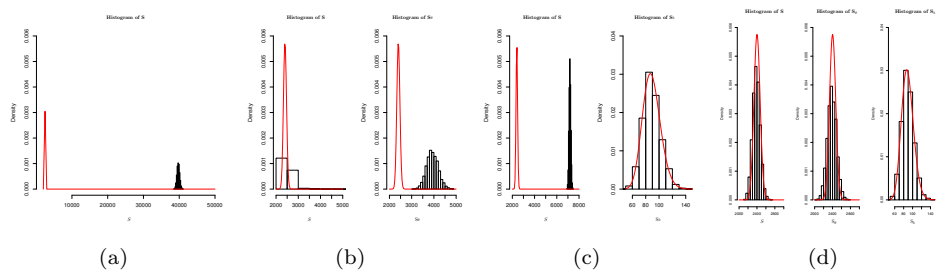


Figure C.51. Simulated data set 9.

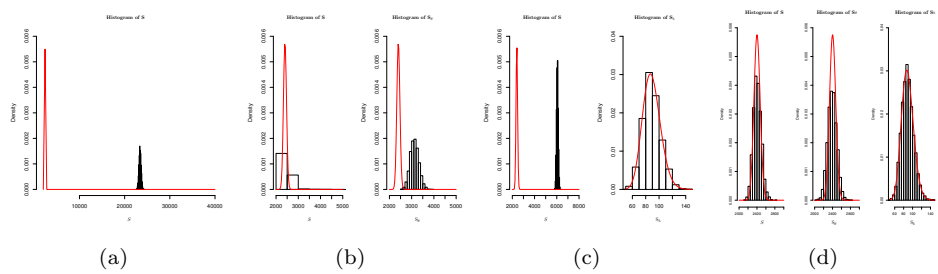


Figure C.52. Simulated data set 10.

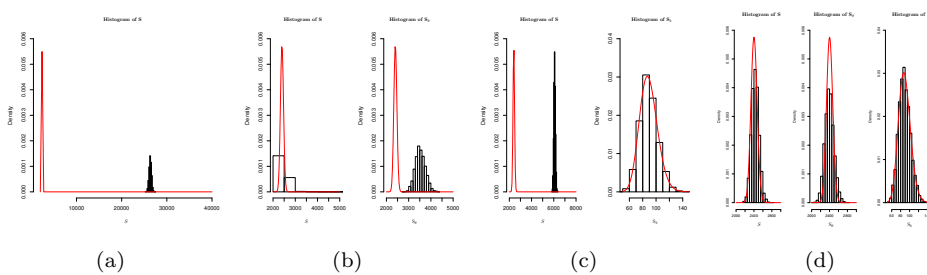


Figure C.53. Simulated data set 11.

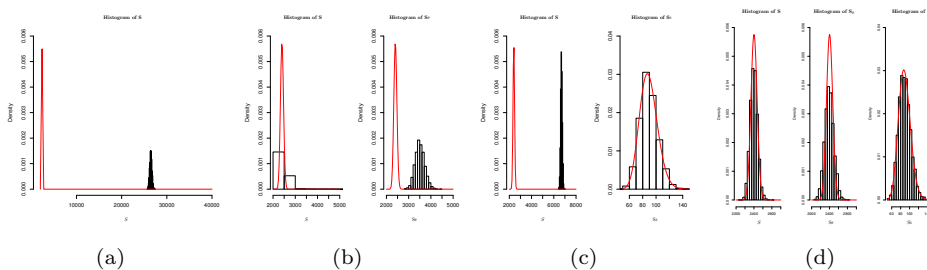


Figure C.54. Simulated data set 12.

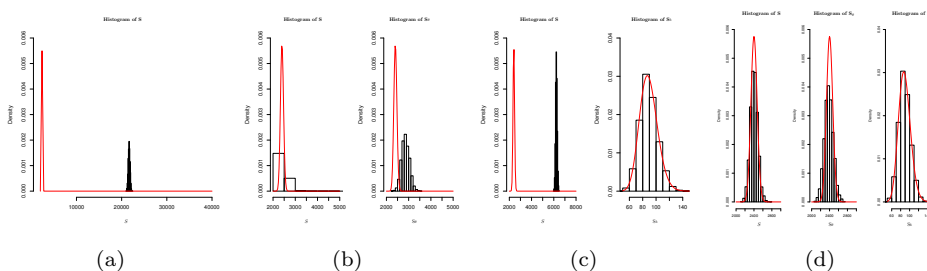


Figure C.55. Simulated data set 13.

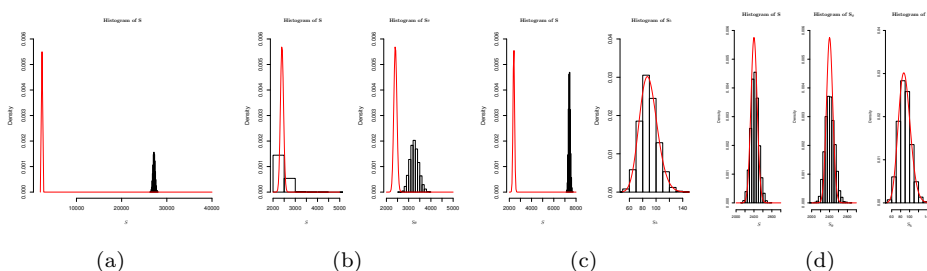


Figure C.56. Simulated data set 14.

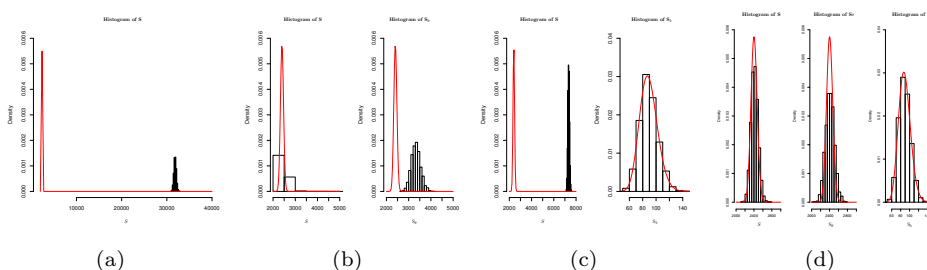


Figure C.57. Simulated data set 15.

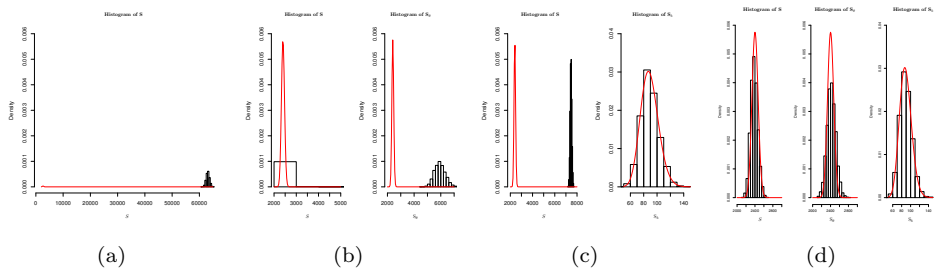


Figure C.58. Simulated data set 16.

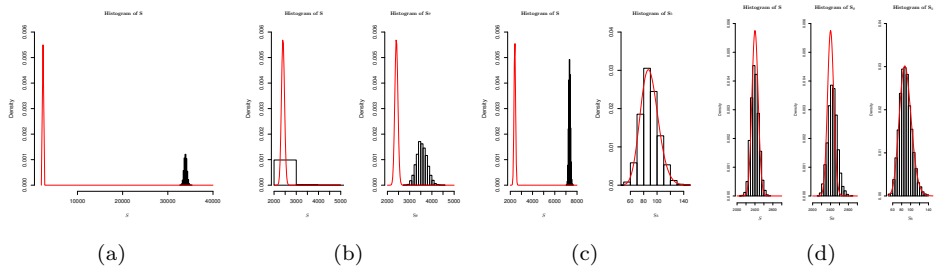


Figure C.59. Simulated data set 17.

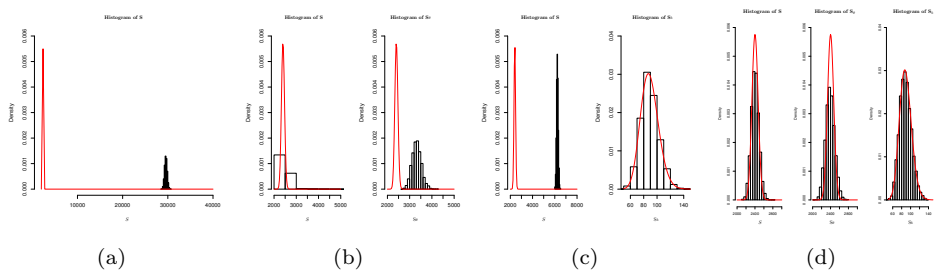


Figure C.60. Simulated data set 18.

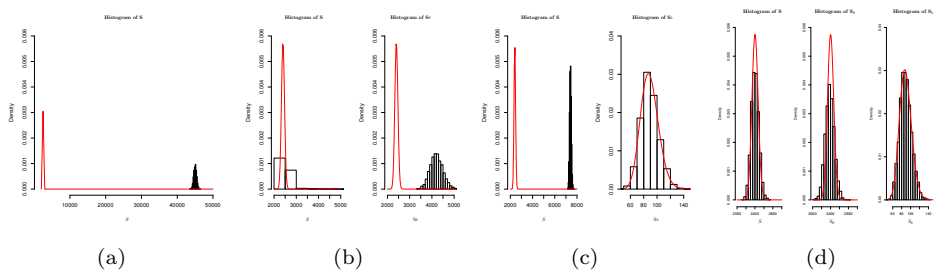


Figure C.61. Simulated data set 19.

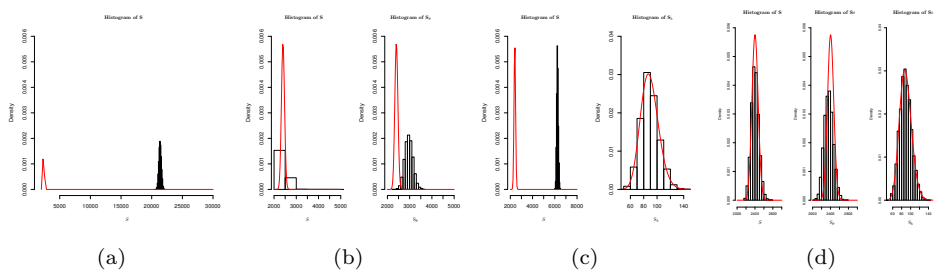


Figure C.62. Simulated data set 20.

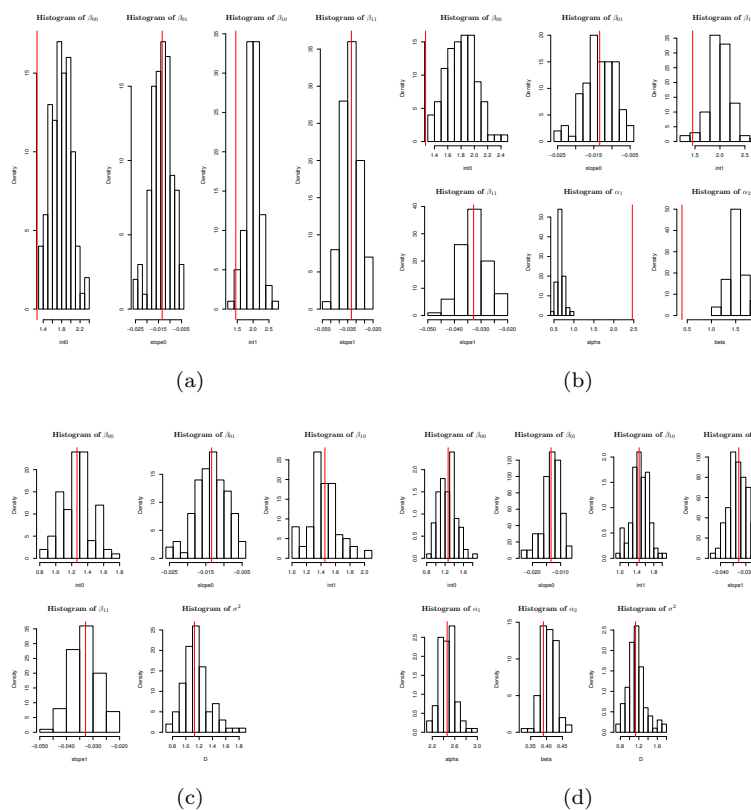


Figure C.63. a) Histograms for each parameter of Poisson model of posterior summaries for the 100 simulated data set and the respective simulated value for each parameter (red line). b) Histograms for each parameter of Poisson-gamma model of posterior summaries for the 100 simulated data set and the respective simulated value for each parameter (red line). c) Histograms for each parameter of Poisson-normal model of posterior summaries for the 100 simulated data set and the respective simulated value for each parameter (red line). d) Histograms for each parameter of Poisson-normal-gamma model of posterior summaries for the 100 simulated data set and the respective simulated value for each parameter (red line).