Characterizing the Distribution of Lysis Time and Burst Size in Lytic Phage

Paul Joyce[1], Craig Miller[1], Dan Weinreich[2]

University of Idaho [1], Brown University [2]

June 14, 2012

Outline

- Briefly explain cycle of a Virus (phage)
- Briefly explain the experiment that monitor a small number of phage through their life cycle
- Use a three step statistical procedure to understand lysis time and burst size that
  - estimates the number of phage in the well
  - estimates the lysis time
  - estimates the burst size
- Conclusions

Outline

- Briefly explain cycle of a Virus (phage)
- Briefly explain the experiment that monitor a small number of phage through their life cycle
- Use a three step statistical procedure to understand lysis time and burst size that
  - estimates the number of phage in the well
  - estimates the lysis time
  - estimates the burst size
- Conclusions

Outline

- Briefly explain cycle of a Virus (phage)
- Briefly explain the experiment that monitor a small number of phage through their life cycle
- Use a three step statistical procedure to understand lysis time and burst size that
  - estimates the number of phage in the well
  - estimates the lysis time
  - estimates the burst size
- Conclusions

Outline

- Briefly explain cycle of a Virus (phage)
- Briefly explain the experiment that monitor a small number of phage through their life cycle
- Use a three step statistical procedure to understand lysis time and burst size that
  - estimates the number of phage in the well
  - estimates the lysis time
  - estimates the burst size
- Conclusions

Outline

- Briefly explain cycle of a Virus (phage)
- Briefly explain the experiment that monitor a small number of phage through their life cycle
- Use a three step statistical procedure to understand lysis time and burst size that
  - estimates the number of phage in the well
  - estimates the lysis time
  - estimates the burst size
- Conclusions

Outline

- Briefly explain cycle of a Virus (phage)
- Briefly explain the experiment that monitor a small number of phage through their life cycle
- Use a three step statistical procedure to understand lysis time and burst size that
  - estimates the number of phage in the well
  - estimates the lysis time
  - estimates the burst size
- Conclusions

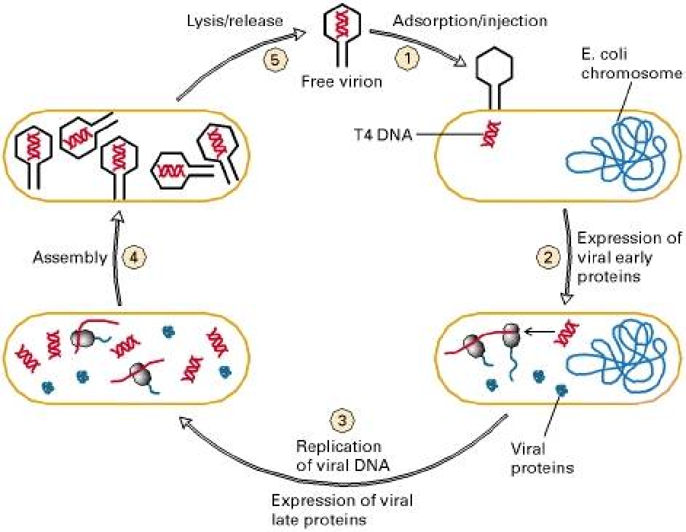
Outline

- Briefly explain cycle of a Virus (phage)
- Briefly explain the experiment that monitor a small number of phage through their life cycle
- Use a three step statistical procedure to understand lysis time and burst size that
  - estimates the number of phage in the well
  - estimates the lysis time
  - estimates the burst size
- Conclusions

Outline

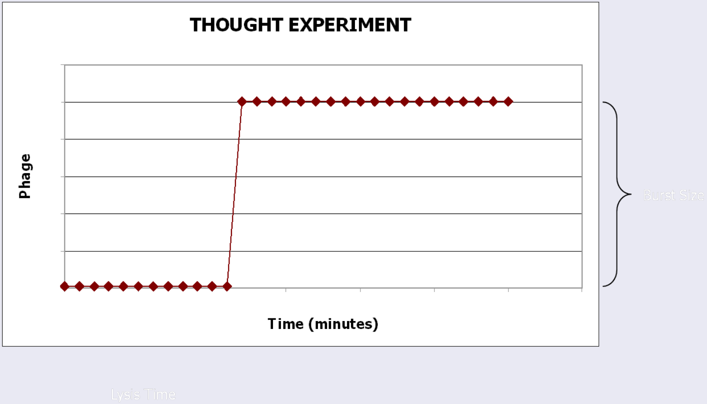
- Briefly explain cycle of a Virus (phage)
- Briefly explain the experiment that monitor a small number of phage through their life cycle
- Use a three step statistical procedure to understand lysis time and burst size that
  - estimates the number of phage in the well
  - estimates the lysis time
  - estimates the burst size
- Conclusions

Lytic Phage Life Cycle

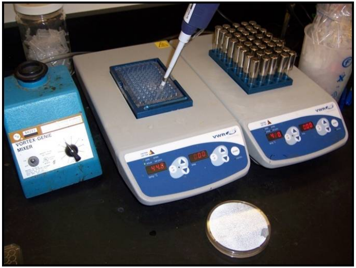


Can We Characterize Burst Size and Lysis Time for an Individual Phage?

Can We Characterize Burst Size and Lysis Time for an Individual Phage?



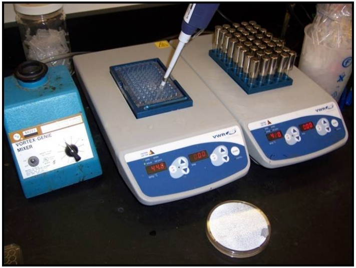
Method



- Phage: $\phi X174$
- In 60 wells add 100 μl host cells.

- Target of $\frac{1}{2}$ phage particle.
- Titrate 20 wells at 5 minutes prior to burst to hone estimates of number of phage per 5 well. We call this the early time data.
- Titrate another well every 30 seconds. We call this the late time point data.

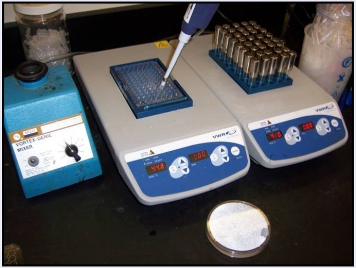
Method



- Phage: $\phi X174$
- In 60 wells add 100 μl host cells.

- Target of $\frac{1}{2}$ phage particle.
- Titrate 20 wells at 5 minutes prior to burst to hone estimates of number of phage per 5 well. We call this the early time data.
- Titrate another well every 30 seconds. We call this the late time point data.

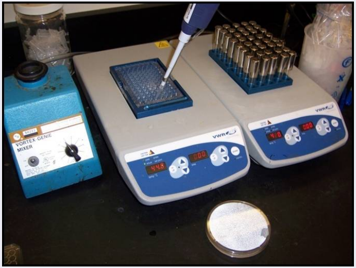
Method



- Phage: $\phi X174$
- In 60 wells add 100 μl host cells.

- Target of $\frac{1}{2}$ phage particle.
- Titrate 20 wells at 5 minutes prior to burst to hone estimates of number of phage per 5 well. We call this the early time data.
- Titrate another well every 30 seconds. We call this the late time point data.

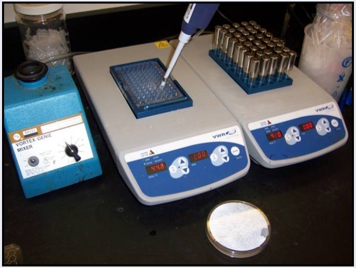
Method



- Phage: $\phi X174$
- In 60 wells add 100 μl host cells.

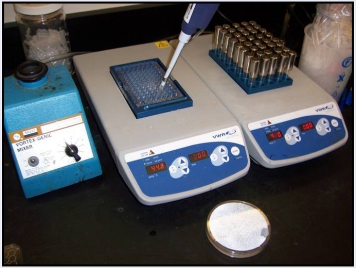
- Target of $\frac{1}{2}$ phage particle.
- Titrate 20 wells at 5 minutes prior to burst to hone estimates of number of phage per 5 well. We call this the early time data.
- Titrate another well every 30 seconds. We call this the late time point data.

Method



- Phage: $\phi X174$
- In 60 wells add 100 μl host cells.

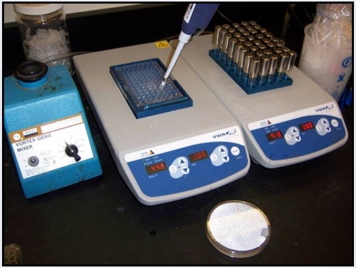
Method



- Phage: $\phi X174$
- In 60 wells add 100 μl host cells.

- Target of $\frac{1}{2}$ phage particle.
- Titrate 20 wells at 5 minutes prior to burst to hone estimates of number of phage per 5 well. We call this the early time data.
- Titrate another well every 30 seconds. We call this the late time point data.

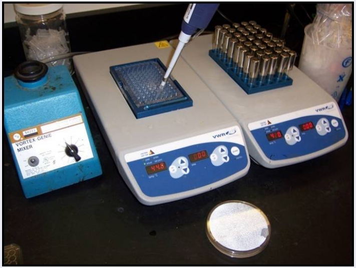
Method



- Phage: $\phi X174$
- In 60 wells add 100 μl host cells.

- Target of $\frac{1}{2}$ phage particle.
- Titrate 20 wells at 5 minutes prior to burst to hone estimates of number of phage per 5 well. We call this the early time data.
- Titrate another well every 30 seconds. We call this the late time point data.

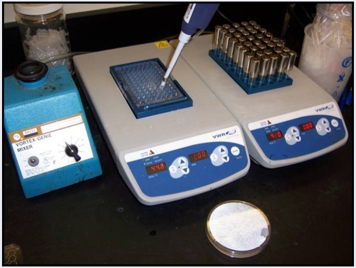
Method



- Phage: $\phi X174$
- In 60 wells add 100 μl host cells.

- Target of $\frac{1}{2}$ phage particle.
- Titrate 20 wells at 5 minutes prior to burst to hone estimates of number of phage per 5 well. We call this the early time data.
- Titrate another well every 30 seconds. We call this the late time point data.

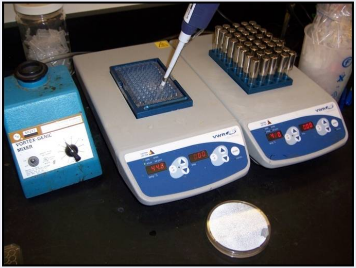
Method



- Phage: $\phi X174$
- In 60 wells add 100 μl host cells.

- Target of $\frac{1}{2}$ phage particle.
- Titrate 20 wells at 5 minutes prior to burst to hone estimates of number of phage per 5 well. We call this the early time data.
- Titrate another well every 30 seconds. We call this the late time point data.

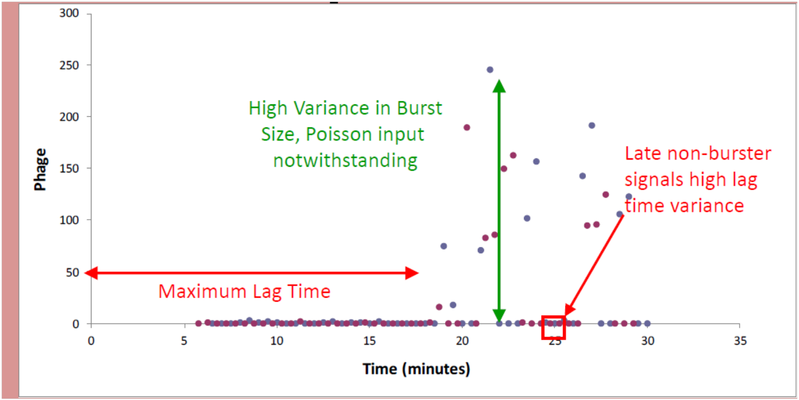
Method



- Phage: $\phi X174$
- In 60 wells add 100 μl host cells.

- Target of $\frac{1}{2}$ phage particle.
- Titrate 20 wells at 5 minutes prior to burst to hone estimates of number of phage per 5 well. We call this the early time data.
- Titrate another well every 30 seconds. We call this the late time point data.

Example Sample Data



Notation

Notation

- T_i= time to burst of phage i.
- t = time when well is sampled
- C_T = burst size at time T for phage i.
- C_t = observed count in well sampled at time t.
- B = the number of phage in the well.
- N_t = the number of phage in a well that have burst by time t, when well is sampled.
- X_t = sum of burst times, T_i, for all phage that have burst
- α = slope of the linear function relating burst size to time.
- μ = intercept at time when burst is first possible $t > t_0$.
- σ=standard deviation or average error of predicted versus observed burst size.

Notation

Notation

- T_i= time to burst of phage i.
- t = time when well is sampled
- C_T = burst size at time T for phage i.
- C_t = observed count in well sampled at time t.
- B = the number of phage in the well.
- N_t = the number of phage in a well that have burst by time t, when well is sampled.
- X_t = sum of burst times, T_i, for all phage that have burst
- α = slope of the linear function relating burst size to time.
- μ = intercept at time when burst is first possible $t > t_0$.
- σ=standard deviation or average error of predicted versus observed burst size.

Notation

Notation

- T_i= time to burst of phage i.
- t = time when well is sampled
- C_T = burst size at time T for phage i.
- C_t = observed count in well sampled at time t.
- B = the number of phage in the well.
- N_t = the number of phage in a well that have burst by time t, when well is sampled.
- X_t = sum of burst times, T_i, for all phage that have burst
- α = slope of the linear function relating burst size to time.
- μ = intercept at time when burst is first possible $t > t_0$.
- σ=standard deviation or average error of predicted versus observed burst size.

Notation

Notation

- T_i= time to burst of phage i.
- t = time when well is sampled
- C_T = burst size at time T for phage i.
- C_t = observed count in well sampled at time t.
- B = the number of phage in the well.
- N_t = the number of phage in a well that have burst by time t, when well is sampled.
- X_t = sum of burst times, T_i, for all phage that have burst
- α = slope of the linear function relating burst size to time.
- μ = intercept at time when burst is first possible $t > t_0$.
- σ=standard deviation or average error of predicted versus observed burst size.

Notation

Notation

- T_i= time to burst of phage i.
- t = time when well is sampled
- C_T = burst size at time T for phage i.
- C_t = observed count in well sampled at time t.
- B = the number of phage in the well.
- N_t = the number of phage in a well that have burst by time t, when well is sampled.
- X_t = sum of burst times, T_i, for all phage that have burst
- α = slope of the linear function relating burst size to time.
- μ = intercept at time when burst is first possible $t > t_0$.
- σ=standard deviation or average error of predicted versus observed burst size.

Notation

Notation

- T_i= time to burst of phage i.
- t = time when well is sampled
- C_T = burst size at time T for phage i.
- C_t = observed count in well sampled at time t.
- B = the number of phage in the well.
- N_t = the number of phage in a well that have burst by time t, when well is sampled.
- X_t = sum of burst times, T_i, for all phage that have burst
- α = slope of the linear function relating burst size to time.
- μ = intercept at time when burst is first possible $t > t_0$.
- σ=standard deviation or average error of predicted versus observed burst size.

Notation

Notation

- T_i= time to burst of phage i.
- t = time when well is sampled
- C_T = burst size at time T for phage i.
- C_t = observed count in well sampled at time t.
- B = the number of phage in the well.
- N_t = the number of phage in a well that have burst by time t, when well is sampled.
- X_t = sum of burst times, T_i, for all phage that have burst
- α = slope of the linear function relating burst size to time.
- μ = intercept at time when burst is first possible $t > t_0$.
- σ=standard deviation or average error of predicted versus observed burst size.

Notation

Notation

- T_i= time to burst of phage i.
- t = time when well is sampled
- C_T = burst size at time T for phage i.
- C_t = observed count in well sampled at time t.
- B = the number of phage in the well.
- N_t = the number of phage in a well that have burst by time t, when well is sampled.
- X_t = sum of burst times, T_i, for all phage that have burst
- α = slope of the linear function relating burst size to time.
- μ = intercept at time when burst is first possible $t > t_0$.
- σ=standard deviation or average error of predicted versus observed burst size.

Notation

Notation

- T_i= time to burst of phage i.
- t = time when well is sampled
- C_T = burst size at time T for phage i.
- C_t = observed count in well sampled at time t.
- B = the number of phage in the well.
- N_t = the number of phage in a well that have burst by time t, when well is sampled.
- X_t = sum of burst times, T_i, for all phage that have burst
- α = slope of the linear function relating burst size to time.
- μ = intercept at time when burst is first possible $t > t_0$.
- σ=standard deviation or average error of predicted versus observed burst size.

Notation

Notation

- T_i= time to burst of phage i.
- t = time when well is sampled
- C_T = burst size at time T for phage i.
- C_t = observed count in well sampled at time t.
- B = the number of phage in the well.
- N_t = the number of phage in a well that have burst by time t, when well is sampled.
- X_t = sum of burst times, T_i, for all phage that have burst
- α = slope of the linear function relating burst size to time.
- μ = intercept at time when burst is first possible $t > t_0$.
- σ=standard deviation or average error of predicted versus observed burst size.

Notation

Notation

- T_i= time to burst of phage i.
- t = time when well is sampled
- C_T = burst size at time T for phage i.
- C_t = observed count in well sampled at time t.
- B = the number of phage in the well.
- N_t = the number of phage in a well that have burst by time t, when well is sampled.
- X_t = sum of burst times, T_i, for all phage that have burst
- α = slope of the linear function relating burst size to time.
- μ = intercept at time when burst is first possible $t > t_0$.
- σ=standard deviation or average error of predicted versus observed burst size.

Three Step Algorithm for Estimating Lysis Time and Burst Size

Step 1: Model and Estimate the mean number of phage per well

- Assume a Poisson number of phage in a well with mean β_d for day d.
- Use the early time point data as direct observations.
- Use late time points as indirect observation.
  - If a burst occurs by time t then there was at least one phage in that well. Assign that event probability $1 - e^{-\beta_d}$
  - If no burst occurs by time t assign that event probability $e^{-\beta_d}$
  - Treat the late burst times as binary data.
- Combine early and late data using maximum likelihood to estimate β_d–the mean number of phage in a well on day d.

Three Step Algorithm for Estimating Lysis Time and Burst Size

Step 1: Model and Estimate the mean number of phage per well

- Assume a Poisson number of phage in a well with mean β_d for day d.
- Use the early time point data as direct observations.
- Use late time points as indirect observation.
  - If a burst occurs by time t then there was at least one phage in that well. Assign that event probability $1 - e^{-\beta_d}$
  - If no burst occurs by time t assign that event probability $e^{-\beta_d}$
  - Treat the late burst times as binary data.
- Combine early and late data using maximum likelihood to estimate β_d–the mean number of phage in a well on day d.

Three Step Algorithm for Estimating Lysis Time and Burst Size

Step 1: Model and Estimate the mean number of phage per well

- Assume a Poisson number of phage in a well with mean β_d for day d.
- Use the early time point data as direct observations.
- Use late time points as indirect observation.
  - If a burst occurs by time t then there was at least one phage in that well. Assign that event probability $1 - e^{-\beta_d}$
  - If no burst occurs by time t assign that event probability $e^{-\beta_d}$
  - Treat the late burst times as binary data.
- Combine early and late data using maximum likelihood to estimate β_d–the mean number of phage in a well on day d.

Three Step Algorithm for Estimating Lysis Time and Burst Size

Step 1: Model and Estimate the mean number of phage per well

- Assume a Poisson number of phage in a well with mean β_d for day d.
- Use the early time point data as direct observations.
- Use late time points as indirect observation.
    - If a burst occurs by time t then there was at least one phage in that well. Assign that event probability $1 - e^{-\beta_d}$
    - If no burst occurs by time t assign that event probability $e^{-\beta_d}$
    - Treat the late burst times as binary data.
- Combine early and late data using maximum likelihood to estimate β_d–the mean number of phage in a well on day d.

Three Step Algorithm for Estimating Lysis Time and Burst Size

Step 1: Model and Estimate the mean number of phage per well

- Assume a Poisson number of phage in a well with mean β_d for day d.
- Use the early time point data as direct observations.
- Use late time points as indirect observation.
  - If a burst occurs by time t then there was at least one phage in that well. Assign that event probability $1 - e^{-\beta_d}$
  - If no burst occurs by time t assign that event probability $e^{-\beta_d}$
  - Treat the late burst times as binary data.
- Combine early and late data using maximum likelihood to estimate β_d–the mean number of phage in a well on day d.

Three Step Algorithm for Estimating Lysis Time and Burst Size

Step 1: Model and Estimate the mean number of phage per well

- Assume a Poisson number of phage in a well with mean β_d for day d.
- Use the early time point data as direct observations.
- Use late time points as indirect observation.
  - If a burst occurs by time t then there was at least one phage in that well. Assign that event probability $1 - e^{-\beta_d}$
  - If no burst occurs by time t assign that event probability $e^{-\beta_d}$
  - Treat the late burst times as binary data.
- Combine early and late data using maximum likelihood to estimate β_d–the mean number of phage in a well on day d.

Three Step Algorithm for Estimating Lysis Time and Burst Size

Step 1: Model and Estimate the mean number of phage per well

- Assume a Poisson number of phage in a well with mean β_d for day d.
- Use the early time point data as direct observations.
- Use late time points as indirect observation.
  - If a burst occurs by time t then there was at least one phage in that well. Assign that event probability $1 - e^{-\beta_d}$
  - If no burst occurs by time t assign that event probability $e^{-\beta_d}$
  - Treat the late burst times as binary data.
- Combine early and late data using maximum likelihood to estimate β_d–the mean number of phage in a well on day d.

Three Step Algorithm for Estimating Lysis Time and Burst Size

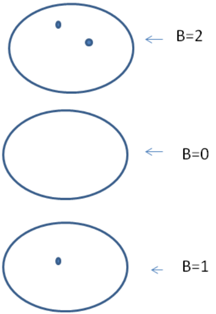
Step 1: Model and Estimate the mean number of phage per well

- Assume a Poisson number of phage in a well with mean β_d for day d.
- Use the early time point data as direct observations.
- Use late time points as indirect observation.
    - If a burst occurs by time t then there was at least one phage in that well. Assign that event probability $1 - e^{-\beta_d}$
    - If no burst occurs by time t assign that event probability $e^{-\beta_d}$
    - Treat the late burst times as binary data.
- Combine early and late data using maximum likelihood to estimate β_d–the mean number of phage in a well on day d.

Step 1: Model and Estimate the mean number of phage per well

Use early time points as direct observations

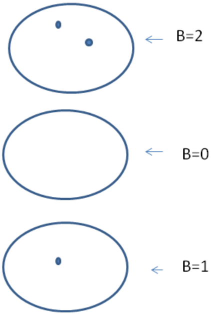
- Assume a Poisson number of phage in a well with mean λ_0 for day 0.
- Use the early time point data as direct observations



Step 1: Model and Estimate the mean number of phage per well

Use early time points as direct observations

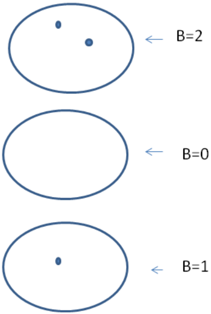
- Assume a Poisson number of phage in a well with mean β_d for day d.
- Use the early time point data as direct observations.



Step 1: Model and Estimate the mean number of phage per well

Use early time points as direct observations

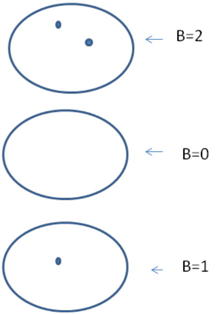
- Assume a Poisson number of phage in a well with mean β_d for day d.
- Use the early time point data as direct observations.



Step 1: Model and Estimate the mean number of phage per well

Use early time points as direct observations

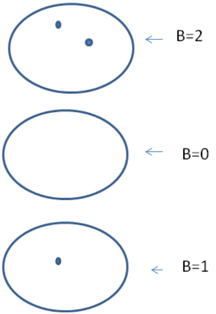
- Assume a Poisson number of phage in a well with mean β_d for day d.
- Use the early time point data as direct observations.



Step 1: Model and Estimate the mean number of phage per well

Use early time points as direct observations

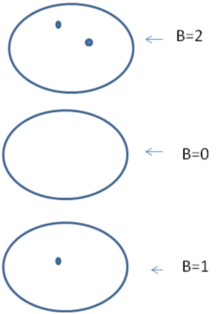
- Assume a Poisson number of phage in a well with mean β_d for day d.
- Use the early time point data as direct observations.



Step 1: Model and Estimate the mean number of phage per well

Use early time points as direct observations

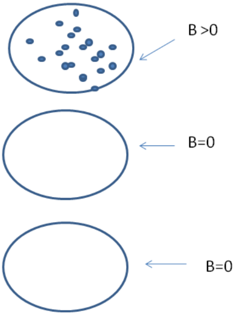
- Assume a Poisson number of phage in a well with mean β_d for day d.
- Use the early time point data as direct observations.



Step 1: Model and Estimate the mean number of phage per well

Use late time points as indirect observation

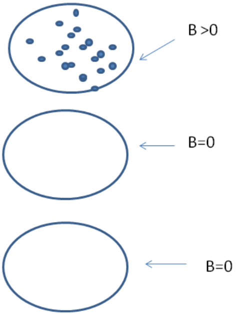
- If a burst occurs by time t then there was at least one phage in that well. Assign that event probability $1 - e^{-\beta_u}$
- If no burst occurs by time t assign that event probability $e^{-\beta_u}$
- Treat the late burst times as binary data



Step 1: Model and Estimate the mean number of phage per well

Use late time points as indirect observation

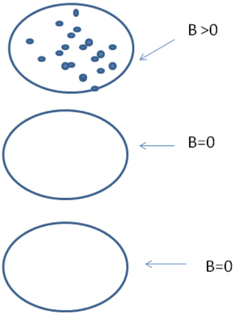
- If a burst occurs by time t then there was at least one phage in that well. Assign that event probability $1 - e^{-\beta_d}$
- If no burst occurs by time t assign that event probability $e^{-\beta_d}$
- Treat the late burst times as binary data.



Step 1: Model and Estimate the mean number of phage per well

Use late time points as indirect observation

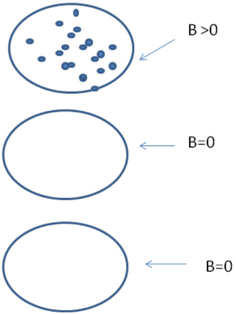
- If a burst occurs by time t then there was at least one phage in that well. Assign that event probability $1 - e^{-\beta_d}$
- If no burst occurs by time t assign that event probability $e^{-\beta_d}$
- Treat the late burst times as binary data.



Step 1: Model and Estimate the mean number of phage per well

Use late time points as indirect observation

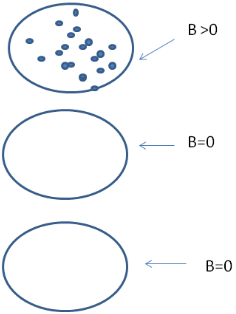
- If a burst occurs by time t then there was at least one phage in that well. Assign that event probability $1 - e^{-\beta_d}$
- If no burst occurs by time t assign that event probability $e^{-\beta_d}$
- Treat the late burst times as binary data.



Step 1: Model and Estimate the mean number of phage per well

Use late time points as indirect observation

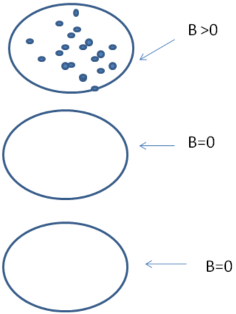
- If a burst occurs by time t then there was at least one phage in that well. Assign that event probability $1 - e^{-\beta_d}$
- If no burst occurs by time t assign that event probability $e^{-\beta_d}$
- Treat the late burst times as binary data.



Step 1: Model and Estimate the mean number of phage per well

Use late time points as indirect observation

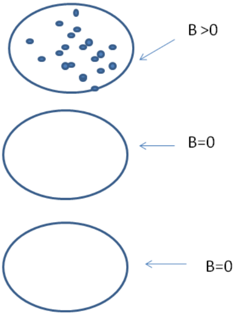
- If a burst occurs by time t then there was at least one phage in that well. Assign that event probability $1 - e^{-\beta_d}$
- If no burst occurs by time t assign that event probability $e^{-\beta_d}$
- Treat the late burst times as binary data.

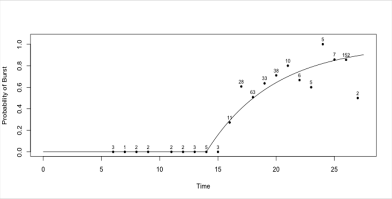


Step 1: Model and Estimate the mean number of phage per well

Use late time points as indirect observation

- If a burst occurs by time t then there was at least one phage in that well. Assign that event probability $1 - e^{-\beta_d}$
- If no burst occurs by time t assign that event probability $e^{-\beta_d}$
- Treat the late burst times as binary data.



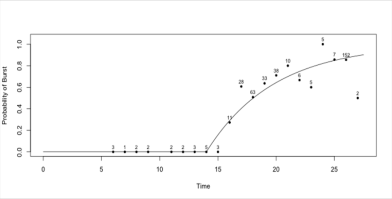


Step 2: Model and Estimate Lysis Time

- Assume the time to burst, T follows a lagged exponential

$$P(T < t) = w_t = 1 - e^{-(t-t_0)/\lambda}$$

- N_t = the number of phage in a well that have burst by time t, when well is sampled. N_t is Poisson with mean βw_t.
- Let Y_t be 1 if $N_t > 0$ and 0 otherwise
- Use binary Y_t to estimate λ and t_0

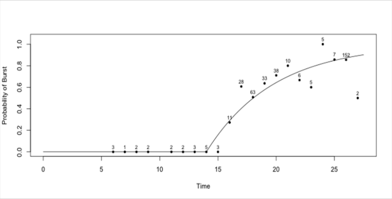


Step 2: Model and Estimate Lysis Time

- Assume the time to burst, T follows a lagged exponential

$$P(T < t) = w_t = 1 - e^{-(t-t_0)/\lambda}$$

- N_t = the number of phage in a well that have burst by time t, when well is sampled. N_t is Poisson with mean βw_t.
- Let Y_t be 1 if $N_t > 0$ and 0 otherwise
- Use binary Y_t to estimate λ and t_0

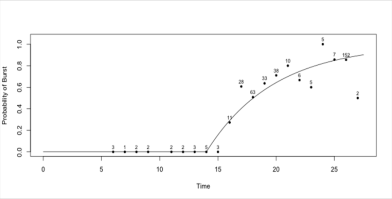


Step 2: Model and Estimate Lysis Time

- Assume the time to burst, T follows a lagged exponential

$$P(T < t) = w_t = 1 - e^{-(t-t_0)/\lambda}$$

- N_t = the number of phage in a well that have burst by time t, when well is sampled. N_t is Poisson with mean βw_t.
- Let Y_t be 1 if $N_t > 0$ and 0 otherwise
- Use binary Y_t to estimate λ and t_0

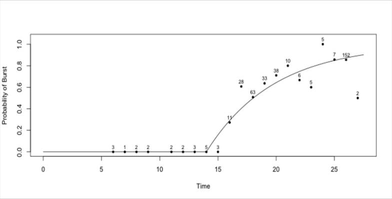


Step 2: Model and Estimate Lysis Time

- Assume the time to burst, T follows a lagged exponential

$$P(T < t) = w_t = 1 - e^{-(t-t_0)/\lambda}$$

- N_t = the number of phage in a well that have burst by time t, when well is sampled. N_t is Poisson with mean βw_t.
- Let Y_t be 1 if $N_t > 0$ and 0 otherwise
- Use binary Y_t to estimate λ and t_0



Step 2: Model and Estimate Lysis Time

- Assume the time to burst, T follows a lagged exponential

$$P(T < t) = w_t = 1 - e^{-(t-t_0)/\lambda}$$

- N_t = the number of phage in a well that have burst by time t, when well is sampled. N_t is Poisson with mean βw_t.
- Let Y_t be 1 if $N_t > 0$ and 0 otherwise
- Use binary Y_t to estimate λ and t_0

Three Step Algorithm for Estimating Lysis Time and Burst Size

Step 3: Model and Estimate Burst Size

$$C_T = \alpha T + \mu + \epsilon$$

where ϵ is normally distributed with mean 0 and variance σ^2. Note that none of the terms in the equation are directly observable

Step 3: Model and Estimate Burst Size

Observable Counts Equation

$$C_t = \alpha \sum_{i=1}^{B} T_i I\{T_i < t\} + \mu \sum_{i=1}^{B} I\{T_i < t\} + \sum_{i=1}^{B} \epsilon_i I\{T_i < t\}$$

$$C_t = \alpha X_t + \mu N_t + \delta$$

where δ is normally distributed with mean zero and variance $N_t \sigma^2$

Step 3: Model and Estimate Burst Size using EM Algorithm

1. Begin with initial guesses for N_t and X_t for each well.
2. Given these values, estimate α and μ using least-squares

$$\min_{\alpha,\mu} \sum_{t,r} \frac{(C_{t,r} - \alpha X_{t,r} - \mu N_{t,r})^2}{N_{t,r}} = \sum_{t,r} \frac{(C_{t,r} - \hat{\alpha} X_{t,r} - \hat{\mu} N_{t,r})^2}{N_{t,r}}$$

   where r indexes the well.
3. Given estimates $\hat{\alpha}$ and $\hat{\mu}$ estimate σ^2 using

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{t,r} \frac{(C_{t,r} - \hat{\alpha} X_{t,r} - \hat{\mu} N_{t,r})^2}{N_{t,r}}$$

4. Impute the values X_t and N_t using their posterior expected values

$$E(X_t | C_{t,r} \hat{\alpha}, \hat{\mu}, \hat{\sigma}^2)$$

$$E(N_t | C_{t,r} \hat{\alpha}, \hat{\mu}, \hat{\sigma}^2)$$

5. return to step 2 until convergence occurs.

Step 3: Model and Estimate Burst Size using EM Algorithm

1. Begin with initial guesses for N_t and X_t for each well.
2. Given these values, estimate α and μ using least-squares

$$\min_{\alpha,\mu} \sum_{t,r} \frac{(C_{t,r} - \alpha X_{t,r} - \mu N_{t,r})^2}{N_{t,r}} = \sum_{t,r} \frac{(C_{t,r} - \hat{\alpha} X_{t,r} - \hat{\mu} N_{t,r})^2}{N_{t,r}}$$

   where r indexes the well.
3. Given estimates $\hat{\alpha}$ and $\hat{\mu}$ estimate σ^2 using

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{t,r} \frac{(C_{t,r} - \hat{\alpha} X_{t,r} - \hat{\mu} N_{t,r})^2}{N_{t,r}}$$

4. Impute the values X_t and N_t using their posterior expected values

$$E(X_t | C_{t,r} \hat{\alpha}, \hat{\mu}, \hat{\sigma}^2)$$

$$E(N_t | C_{t,r} \hat{\alpha}, \hat{\mu}, \hat{\sigma}^2)$$

5. return to step 2 until convergence occurs.

Step 3: Model and Estimate Burst Size using EM Algorithm

1. Begin with initial guesses for N_t and X_t for each well.
2. Given these values, estimate α and μ using least-squares

$$\min_{\alpha,\mu} \sum_{t,r} \frac{(C_{t,r} - \alpha X_{t,r} - \mu N_{t,r})^2}{N_{t,r}} = \sum_{t,r} \frac{(C_{t,r} - \hat{\alpha} X_{t,r} - \hat{\mu} N_{t,r})^2}{N_{t,r}}$$

   where r indexes the well.
3. Given estimates $\hat{\alpha}$ and $\hat{\mu}$ estimate σ^2 using

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{t,r} \frac{(C_{t,r} - \hat{\alpha} X_{t,r} - \hat{\mu} N_{t,r})^2}{N_{t,r}}$$

4. Impute the values X_t and N_t using their posterior expected values

$$E(X_t | C_{t,r} \hat{\alpha}, \hat{\mu}, \hat{\sigma}^2)$$

$$E(N_t | C_{t,r} \hat{\alpha}, \hat{\mu}, \hat{\sigma}^2)$$

5. return to step 2 until convergence occurs.

Step 3: Model and Estimate Burst Size using EM Algorithm

1. Begin with initial guesses for N_t and X_t for each well.
2. Given these values, estimate α and μ using least-squares

$$\min_{\alpha,\mu} \sum_{t,r} \frac{(C_{t,r} - \alpha X_{t,r} - \mu N_{t,r})^2}{N_{t,r}} = \sum_{t,r} \frac{(C_{t,r} - \hat{\alpha} X_{t,r} - \hat{\mu} N_{t,r})^2}{N_{t,r}}$$

   where r indexes the well.
3. Given estimates $\hat{\alpha}$ and $\hat{\mu}$ estimate σ^2 using

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{t,r} \frac{(C_{t,r} - \hat{\alpha} X_{t,r} - \hat{\mu} N_{t,r})^2}{N_{t,r}}$$

4. Impute the values X_t and N_t using their posterior expected values

$$E(X_t | C_{t,r} \hat{\alpha}, \hat{\mu}, \hat{\sigma}^2)$$

$$E(N_t | C_{t,r} \hat{\alpha}, \hat{\mu}, \hat{\sigma}^2)$$

5. return to step 2 until convergence occurs.

Step 3: Model and Estimate Burst Size using EM Algorithm

1. Begin with initial guesses for N_t and X_t for each well.
2. Given these values, estimate α and μ using least-squares

$$\min_{\alpha,\mu} \sum_{t,r} \frac{(C_{t,r} - \alpha X_{t,r} - \mu N_{t,r})^2}{N_{t,r}} = \sum_{t,r} \frac{(C_{t,r} - \hat{\alpha} X_{t,r} - \hat{\mu} N_{t,r})^2}{N_{t,r}}$$

   where r indexes the well.
3. Given estimates $\hat{\alpha}$ and $\hat{\mu}$ estimate σ^2 using

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{t,r} \frac{(C_{t,r} - \hat{\alpha} X_{t,r} - \hat{\mu} N_{t,r})^2}{N_{t,r}}$$

4. Impute the values X_t and N_t using their posterior expected values

$$E(X_t | C_{t,r} \hat{\alpha}, \hat{\mu}, \hat{\sigma}^2)$$

$$E(N_t | C_{t,r} \hat{\alpha}, \hat{\mu}, \hat{\sigma}^2)$$

5. return to step 2 until convergence occurs.

Step 3: Model and Estimate Burst Size using EM Algorithm

1. Begin with initial guesses for N_t and X_t for each well.
2. Given these values, estimate α and μ using least-squares

$$\min_{\alpha,\mu} \sum_{t,r} \frac{(C_{t,r} - \alpha X_{t,r} - \mu N_{t,r})^2}{N_{t,r}} = \sum_{t,r} \frac{(C_{t,r} - \hat{\alpha} X_{t,r} - \hat{\mu} N_{t,r})^2}{N_{t,r}}$$

   where r indexes the well.
3. Given estimates $\hat{\alpha}$ and $\hat{\mu}$ estimate σ^2 using

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{t,r} \frac{(C_{t,r} - \hat{\alpha} X_{t,r} - \hat{\mu} N_{t,r})^2}{N_{t,r}}$$

4. Impute the values X_t and N_t using their posterior expected values

$$E(X_t | C_{t,r} \hat{\alpha}, \hat{\mu}, \hat{\sigma}^2)$$

$$E(N_t | C_{t,r} \hat{\alpha}, \hat{\mu}, \hat{\sigma}^2)$$

5. return to step 2 until convergence occurs.

Imputing X_t and N_t

$$E(N_t|C_t) = \sum_{x_t} \sum_{n_t} \frac{n_t f_{N_t,X_t}(n_t, x_t|\beta, t_0, \lambda) f_{C_t}(c_t|n_t, x_t, \alpha, \mu, \sigma^2)}{f_{C_t}(c_t)}$$

$$E(X_t|C_t) = \sum_{x_t} \sum_{n_t} \frac{x_t f_{N_t,X_t}(n_t, x_t|\beta, t_0, \lambda) f_{C_t}(c_t|n_t, x_t, \alpha, \mu, \sigma^2)}{f_{C_t}(c_t)}$$

$$f_{C_t}(C_t) = \sum_{n_t, x_t} f_{N_t,X_t}(n_t, x_t|\beta, t_0, \lambda) f_{C_t}(c_t|n_t, x_t, \alpha, \mu, \sigma^2)$$

- β is estimated in step 1
- λ, t_0 is estimated in step 2
- μ, σ^2 are updated from the "M" part of the EM algorithm in step 3

Imputing X_t and N_t

$$E(N_t|C_t) = \sum_{x_t}\sum_{n_t} \frac{n_t f_{N_t,X_t}(n_t, x_t|\beta, t_0, \lambda) f_{C_t}(c_t|n_t, x_t, \alpha, \mu, \sigma^2)}{f_{C_t}(c_t)}$$

$$E(X_t|C_t) = \sum_{x_t}\sum_{n_t} \frac{x_t f_{N_t,X_t}(n_t, x_t|\beta, t_0, \lambda) f_{C_t}(c_t|n_t, x_t, \alpha, \mu, \sigma^2)}{f_{C_t}(c_t)}$$

$$f_{C_t}(C_t) = \sum_{n_t,x_t} f_{N_t,X_t}(n_t, x_t|\beta, t_0, \lambda) f_{C_t}(c_t|n_t, x_t, \alpha, \mu, \sigma^2)$$

- β is estimated in step 1
- λ, t_0 is estimated in step 2
- μ, σ^2 are updated from the "M" part of the EM algorithm in step 3

Imputing X_t and N_t

$$E(N_t|C_t) = \sum_{x_t}\sum_{n_t} \frac{n_t f_{N_t,X_t}(n_t, x_t|\beta, t_0, \lambda) f_{C_t}(c_t|n_t, x_t, \alpha, \mu, \sigma^2)}{f_{C_t}(c_t)}$$

$$E(X_t|C_t) = \sum_{x_t}\sum_{n_t} \frac{x_t f_{N_t,X_t}(n_t, x_t|\beta, t_0, \lambda) f_{C_t}(c_t|n_t, x_t, \alpha, \mu, \sigma^2)}{f_{C_t}(c_t)}$$

$$f_{C_t}(C_t) = \sum_{n_t, x_t} f_{N_t,X_t}(n_t, x_t|\beta, t_0, \lambda) f_{C_t}(c_t|n_t, x_t, \alpha, \mu, \sigma^2)$$

- β is estimated in step 1
- λ, t_0 is estimated in step 2
- μ, σ^2 are updated from the "M" part of the EM algorithm in step 3

Imputing X_t and N_t

$$E(N_t|C_t) = \sum_{x_t}\sum_{n_t} \frac{n_t f_{N_t,X_t}(n_t, x_t|\beta, t_0, \lambda) f_{C_t}(c_t|n_t, x_t, \alpha, \mu, \sigma^2)}{f_{C_t}(c_t)}$$

$$E(X_t|C_t) = \sum_{x_t}\sum_{n_t} \frac{x_t f_{N_t,X_t}(n_t, x_t|\beta, t_0, \lambda) f_{C_t}(c_t|n_t, x_t, \alpha, \mu, \sigma^2)}{f_{C_t}(c_t)}$$

$$f_{C_t}(C_t) = \sum_{n_t,x_t} f_{N_t,X_t}(n_t, x_t|\beta, t_0, \lambda) f_{C_t}(c_t|n_t, x_t, \alpha, \mu, \sigma^2)$$

- β is estimated in step 1
- λ, t_0 is estimated in step 2
- μ, σ^2 are updated from the "M" part of the EM algorithm in step 3

Initial Data on Burst Size

- Burst size was only recorded after 26 minutes. Thus there was no way to detect a time trend for this data.
- This would appear to make data analysis easier. Since we would need to assume $\alpha = 0$, so the observed burst counts depend only on N_t the initial number of phage in the well.

Initial Data on Burst Size

- Burst size was only recorded after 26 minutes. Thus there was no way to detect a time trend for this data.
- This would appear to make data analysis easier. Since we would need to assume $\alpha = 0$, so the observed burst counts depend only on N_t the initial number of phage in the well.

Initial Data on Burst Size

- Burst size was only recorded after 26 minutes. Thus there was no way to detect a time trend for this data.
- This would appear to make data analysis easier. Since we would need to assume $\alpha = 0$, so the observed burst counts depend only on N_t the initial number of phage in the well.

Special case of no time trend $\alpha = 0$

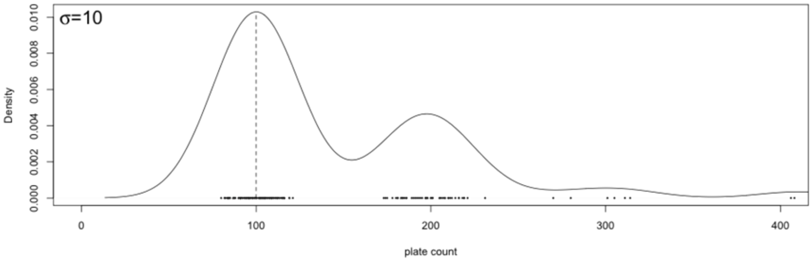
Observable Counts Equation

$$C_t = \mu \sum_{i=1}^{B} I\{T_i < t\} + \sum_{i=1}^{B} \epsilon_i I\{T_i < t\}$$

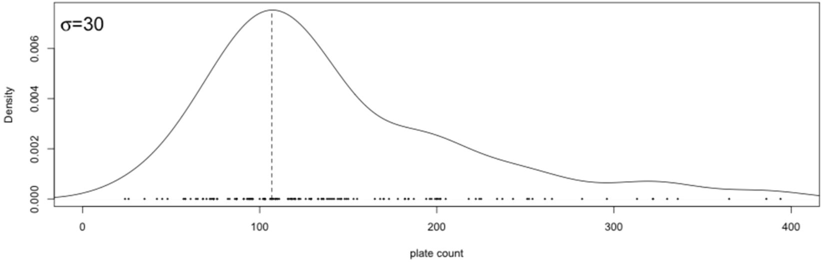
$$C_t = \mu N_t + \delta$$

where δ is normally distributed with mean zero and variance $N_t \sigma^2$

Estimating Mean Burst Size μ with Low Variance

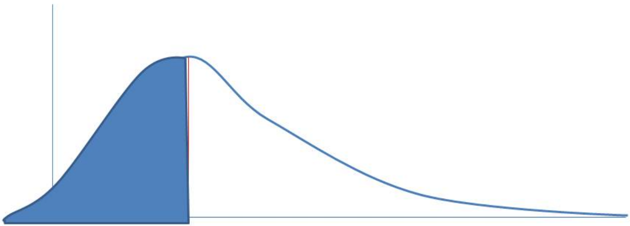


Estimating Mean Burst Size μ with high Variance

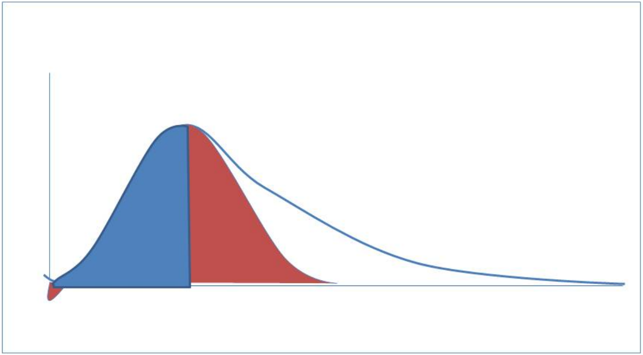


Estimating Mean Burst Size μ with high Variance

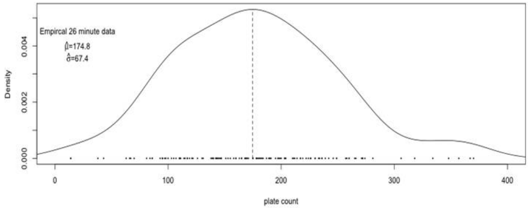
Estimating Mean Burst Size μ with high Variance



Estimating Mean Burst Size μ with high Variance



Estimating Mean Burst Size μ from 26 Minute Data



More Data Came in

New Data Structure

- Process was monitored more extensively so bursts events and multiple time points were recorded.
- This revealed a time trend in burst size, which helped explain the data.
- This produced lower variance, and better imputation.
- The EM algorithm now worked.

More Data Came in

New Data Structure

- Process was monitored more extensively so bursts events and multiple time points were recorded.
- This revealed a time trend in burst size, which helped explain the data.
- This produced lower variance, and better imputation.
- The EM algorithm now worked.

More Data Came in

New Data Structure

- Process was monitored more extensively so bursts events and multiple time points were recorded.
- This revealed a time trend in burst size, which helped explain the data.
- This produced lower variance, and better imputation.
- The EM algorithm now worked.

More Data Came in

New Data Structure

- Process was monitored more extensively so bursts events and multiple time points were recorded.
- This revealed a time trend in burst size, which helped explain the data.
- This produced lower variance, and better imputation.
- The EM algorithm now worked.

More Data Came in

New Data Structure

- Process was monitored more extensively so bursts events and multiple time points were recorded.
- This revealed a time trend in burst size, which helped explain the data.
- This produced lower variance, and better imputation.
- The EM algorithm now worked.

Estimating Slope α and Intercept μ

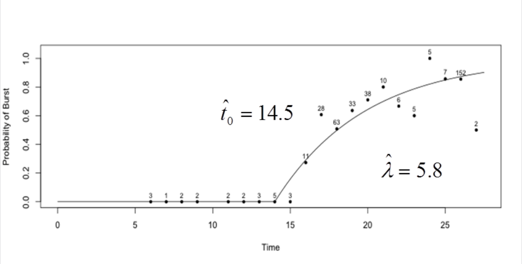
Observable Counts Equation

$$C_t = \alpha \sum_{i=1}^{B} T_i I\{T_i < t\} + \mu \sum_{i=1}^{B} I\{T_i < t\} + \sum_{i=1}^{B} \epsilon_i I\{T_i < t\}$$

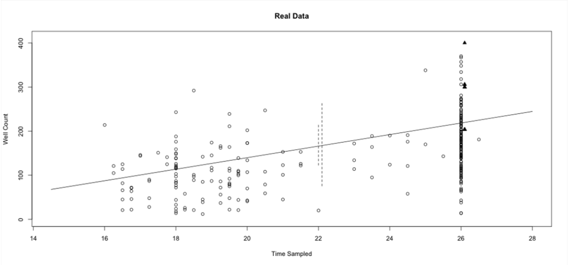
$$C_t = \alpha X_t + \mu N_t + \delta$$

where δ is normally distributed with mean zero and variance $N_t \sigma^2$

The estimated lysis time until burst probability > 0, t_0= 14.5 minutes. Estimated mean lysis time is $\lambda + t_0 = 5.8 + 14.5 = 20.3$ minutes.



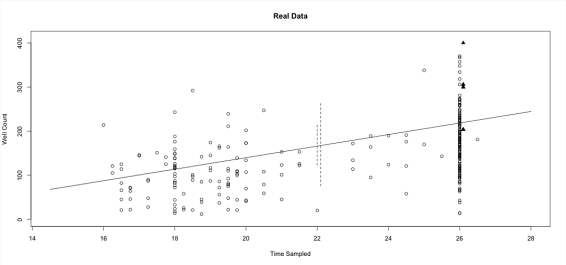
Full Data



Slope intercept

- The estimated burst size at t_0 is $\mu = 67.0$. The estimated increase in burst size per minute is $\alpha = 13.1$.
- The estimated mean burst size is 143 phage: 67 + 13.1(5.8).

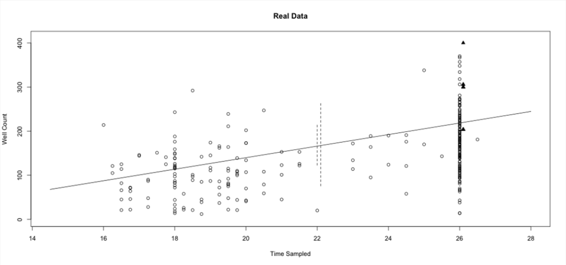
Full Data



Slope intercept

- The estimated burst size at t_0 is $\mu = 67.0$. The estimated increase in burst size per minute is $\alpha = 13.1$.
- The estimated mean burst size is 143 phage: 67 + 13.1(5.8).

Full Data



Slope intercept

- The estimated burst size at t_0 is $\mu = 67.0$. The estimated increase in burst size per minute is $\alpha = 13.1$.
- The estimated mean burst size is 143 phage: 67 + 13.1(5.8).

Parameter estimates

Parameter	Estimate	Confidence Interval
α	13.1	$(6.0, 22.5)$
μ	67.0	$(56.8, 87.0)$
σ	47.7	$(36.3, 57.8)$

Conclusions

- Either 'Missing data' or 'censored data' challenge any statistical modeling effort. This data set has both.
- We address the challenges in 3 steps: (1) estimating the number of phage in each well, (2) estimating the lysis time, and (3) estimate the burst size.
- A rigorously validation process using simulations offers credibility to the lysis time and size estimates for ϕ X 174 and indicates the methods may be extended to other lytic phage.

Support: Paul Joyce and Craig Miller were funded by NIH R01 GM076040

Conclusions

- Either 'Missing data' or 'censored data' challenge any statistical modeling effort. This data set has both.
- We address the challenges in 3 steps: (1) estimating the number of phage in each well, (2) estimating the lysis time, and (3) estimate the burst size.
- A rigorously validation process using simulations offers credibility to the lysis time and size estimates for ϕ X 174 and indicates the methods may be extended to other lytic phage.

Support: Paul Joyce and Craig Miller were funded by NIH R01 GM076040

Conclusions

- Either 'Missing data' or 'censored data' challenge any statistical modeling effort. This data set has both.
- We address the challenges in 3 steps: (1) estimating the number of phage in each well, (2) estimating the lysis time, and (3) estimate the burst size.
- A rigorously validation process using simulations offers credibility to the lysis time and size estimates for ϕ X 174 and indicates the methods may be extended to other lytic phage.

Support: Paul Joyce and Craig Miller were funded by NIH R01 GM076040

Conclusions

- Either ‘Missing data’ or ‘censored data’ challenge any statistical modeling effort. This data set has both.
- We address the challenges in 3 steps: (1) estimating the number of phage in each well, (2) estimating the lysis time, and (3) estimate the burst size.
- A rigorously validation process using simulations offers credibility to the lysis time and size estimates for ϕ X 174 and indicates the methods may be extended to other lytic phage.

Support: Paul Joyce and Craig Miller were funded by NIH R01 GM076040