

Short Communication

Application of a statistical bootstrapping technique  
to calculate growth rate variance for modelling  
psychrotrophic pathogen growth

Donald W. Schaffner

*Food Science Department, Rutgers State University of New Jersey, New Brunswick, P.O. Box 231,  
NJ 08903, USA*

---

**Abstract**

The inherent variability or ‘variance’ of growth rate measurements is critical to the development of accurate predictive models in food microbiology. A large number of measurements are typically needed to estimate variance. To make these measurements requires a significant investment of time and effort. If a single growth rate determination is based on a series of independent measurements, then a statistical bootstrapping technique can be used to simulate multiple growth rate measurements from a single set of experiments. Growth rate variances were calculated for three large datasets (*Listeria monocytogenes*, *Listeria innocua*, and *Yersinia enterocolitica*) from our laboratory using this technique. This analysis revealed that the population of growth rate measurements at any given condition are not normally distributed, but instead follow a distribution that is between normal and Poisson. The relationship between growth rate and temperature was modeled by response surface models using generalized linear regression. It was found that the assumed distribution (i.e. normal, Poisson, gamma or inverse normal) of the growth rates influenced the prediction of each of the models used. This research demonstrates the importance of variance and assumptions about the statistical distribution of growth rates on the results of predictive microbiological models.

*Keywords:* Predictive microbiology; Bootstrap technique; Growth rate variance; *Listeria monocytogenes*; *Yersinia enterocolitica*

---

**1. Introduction**

The objective of predictive food microbiology is to improve food safety and quality through the application of mathematics and statistics. Proper regression

analysis requires either knowledge of the distribution of the population the data are taken from or that population distribution estimates be made (Ratkowsky et al., 1991; Alber and Schaffner, 1992). Linear regression assumes the data come from a normally distributed population, but most microbial growth data is not normally distributed. Generalized Linear Models (GLM) can be used to model populations that are not normally distributed (Ratkowsky, 1992); however, the error distribution must be either known or estimated. The nature of the relationship between the mean response and the variance can be used to determine the error distribution of the data. In order to get an accurate estimate of the variance for a given mean response, a large number of samples must be collected. Data collection in predictive food microbiology can be very time-consuming and labor-intensive. A new method for generating multiple estimates of growth rate from a single growth curve has been developed in our laboratory. This approach uses a statistical technique known as bootstrapping (Efron and Tibshirani, 1991). A bootstrapped estimates of growth rate variances permit the use of GLM in situations where only a single growth curve is available. Response surface models were fit (using GLM) to growth rate data for three psychrotrophic bacteria previously collected in our laboratory.

## 2. Methods

*Microbiological methods.* *Y. enterocolitica*, *L. monocytogenes*, and *L. innocua* datasets previously collected in our laboratory were used for all analyses. The microbiological methods have been described in the published literature (Alber and Schaffner, 1992; Duh and Schaffner, 1993). Only data collected at temperatures of 37°C or less were used in this study.

*Statistical methods.* The software Genstat (Numerical Algorithms Group) was used for all statistical analyses. The Gompertz equation (Eq. 1) was fit to each growth curve, for each organism, at each temperature.

$$L(t) = A + C \exp[-\exp[-B(t - M)]] \quad (1)$$

The minimum and maximum of the second derivative of the fitted Gompertz function are the points where the slope of the growth curve shows the point of greatest increase, and the point of greatest decrease. These two points defined the beginning and end of the exponential phase of growth (Buchanan and Cygnarowicz, 1990). Statistical bootstrapping was used to resample the data in the exponential phase. Five thousand bootstrapped estimates of the growth rate were made for each organism at each temperature. The variances of the 5000 bootstrapped means for each organism and temperature combination were calculated using Eq. 2:

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1} \quad (2)$$

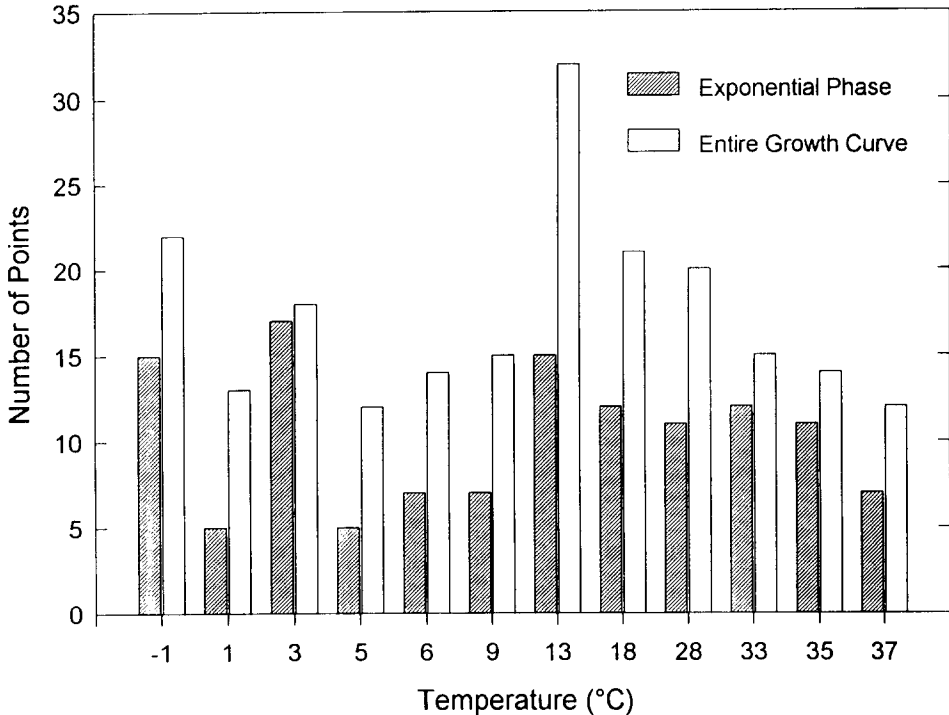


Fig. 1. Total number of data points in the growth curve and the number of data points in the exponential phase used for bootstrapped estimates of growth rate for the *Yersinia enterocolitica* dataset.

The Box–Cox method (Box et al., 1978) was used to find the best relationship between the mean response and the variance. The response surface model was fit to each dataset using four different assumptions (normal, Poisson, gamma or inverse normal) concerning the distribution of the growth rate means.

The mean of the bootstrapped growth rates was predicted as a function of temperature using response surface model after the method of Ratkowsky (1992).

### 3. Results and discussion

Fig. 1 shows the total number of data points at each temperature in the *Y. enterocolitica* dataset, and the number of points that were selected for bootstrapping using the second derivative method described above. At least five points were used, and at one temperature 17 points from the exponential phase were used for bootstrapping.

Fig. 2 is a plot of a typical distribution of growth rates, using the bootstrapped data for *L. monocytogenes* growth rates at 8°C. The number of growth rates for each of 50 different ranges from the lowest (0.026) through the mean (0.030) to the highest (0.034) bootstrapped growth rate are shown in this figure.

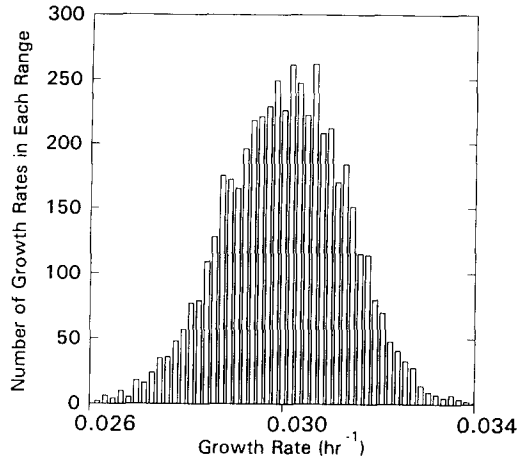


Fig. 2. Distribution of bootstrapped growth rates for *Listeria monocytogenes* grown at 8°C.

The reason for the bootstrapping process, as mentioned above, is that it allows one to calculate the variance of the new population of growth rates. The relationship between the number of bootstrapped estimates of the growth rate and the variance of those estimates is shown in Fig. 3. The variance rises quickly from zero for one estimate to about  $5 \times 10^{-6}$  for about 40 estimates, and is quite stable at that same level after 1000 estimates. The data shown in Fig. 3 are for *L. innocua* at 15°C where the exponential phase contained eight data points.

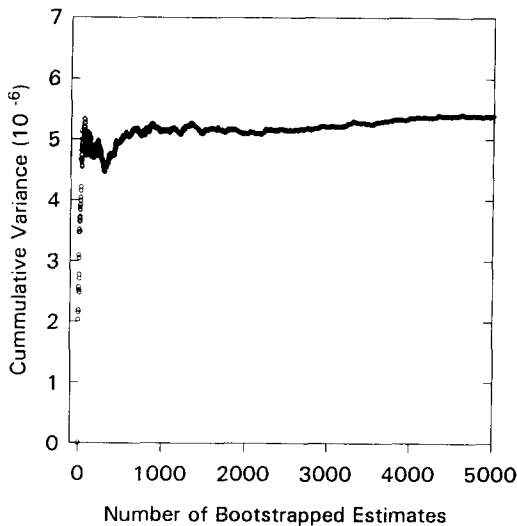


Fig. 3. Relationship between the number of bootstrapped estimates of growth rate for *Listeria innocua* at 15°C and the cumulative estimated growth rate variance.

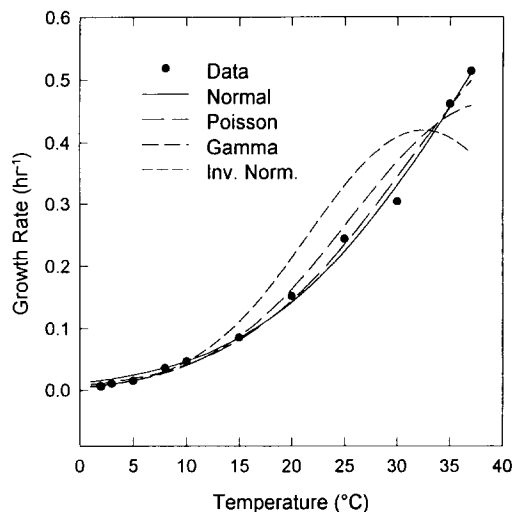


Fig. 4. Growth rate predictions for *Listeria innocua* as a function of temperature using a response surface model with four distribution assumptions.

The  $\alpha$ -values from the Box–Cox plots are as follows: *L. innocua*: 0.31; *L. monocytogenes*: 0.34; and *Y. enterocolitica*: 0.14. The  $\alpha$ -values for the standard distributions are: Normal: 0; Poisson: 0.5; Gamma: 1; and Inverse normal: 1.5. From this we can see that the distribution of all three datasets is between normal and Poisson.

A typical example of the final result of the bootstrapping and GLM regression process is shown in Fig. 4. The data are shown as closed circles, while the regression results using the response surface model are shown as different dashed lines. The normal or Poisson plots (Fig. 4) seem to closely describe the actual data points. Careful inspection of Fig. 4 reveals, however, that the normal distribution may slightly overestimate growth rate at temperatures of less than 7°C. The gamma distribution plot over-estimates growth rates from 20°C to 35°C and under-estimates growth rate from 35°C to 37°C. The inverse normal distribution plot over-estimates growth rate from 15°C to 35°C and also under-estimates growth rate from 35°C to 37°C. The  $\alpha$  value for the *L. innocua* dataset is 0.31, which would indicate a distribution between normal and Poisson, and Fig. 4 supports this result. This finding also seems to concur with the results previously reported in the literature, which indicate that the square root transformation (indicating a Poisson distribution) seems most appropriate for microbial growth rates (Alber and Schaffner, 1992; Ratkowsky, 1992; Zwietering et al., 1994).

#### 4. Conclusions

The growth rate bootstrapping method described here was used to estimate growth rate variances from a single growth curve. At least 40 and ideally 1000

bootstrapped estimates of growth rate are required to obtain a consistent growth rate variance. The variances obtained by bootstrapping can be used to determine the theoretical distribution of the growth rate data using the Box–Cox method. All three psychrotrophic pathogen growth rate datasets had a distribution between normal and Poisson, which is consistent with the literature. The assumed distribution influenced the predictions obtained using generalized linear regression.

### Acknowledgments

This is paper number D-10122-2-93 of the New Jersey Agricultural Experiment Station. This research was partially supported by state appropriations and US Hatch Act Funds.

### References

- Alber, S.A. and Schaffner, D.W. (1992) Evaluation of data transformations used with the square root and Schoolfield models for predicting bacterial growth rate. *Appl. Environ. Microbiol.* 58, 3337–3342.
- Box, G.E.P., Hunter, W.G. and Hunter, J.S. (1978) Statistics for experimenters, In: *An Introduction To Design, Data Analysis, and Model Building. Statistics For Experimenters.* Wiley, New York, pp. 232–235.
- Buchanan, R.L. and Cygnarowicz, M.L. (1990) A mathematical approach toward defining and calculating the duration of the lag phase. *Food Microbiol.* 7, 237–240.
- Duh, Y.-H. and Schaffner, D.W. (1993) Modeling the effect of temperature on the growth rate and lag time of *Listeria innocua* and *Listeria monocytogenes*. *J. Food Prot.* 56, 205–210.
- Efron, B. and Tibshirani, R. (1991) Statistical data analysis in the computer age. *Science* 253, 390–395.
- Ratkowsky, D.A. (1992) Predicting response times in predictive food microbiology. Occasional Paper 1992/1, Fisheries and Energy, Research and Development Unit, Biometrics Section, Department of Primary Industry, Tasmania.
- Ratkowsky, D.A., Ross, T., McMeekin, T.A. and Olley, J. (1991) Comparison of Arrhenius-type and Bělehrádek-type models for prediction of bacterial growth in foods. *J. Appl. Bacteriol.* 71, 452–459.
- Zwietering, M.H., Cuppers, H.G.A.M., De Wit, J.C. and Van 't Riet, K. (1994) Evaluation of data transformations and validation of models for the effect of temperature on bacterial growth. *Appl. Environ. Microbiol.* 60, 195–203.