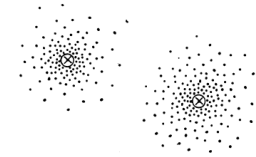
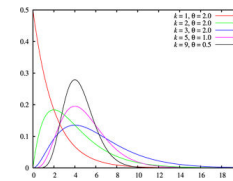


The Relevance of Within-Batch Distributions to Microbiological Criteria in Foods



Dr. Ursula Gonzales Barron
Prof. Francis Butler

Biosystems Engineering, UCD School of Agriculture, Food Science
and Veterinary Medicine
University College Dublin, Ireland

Contents

- Theoretical concepts
 - Within-batch variability
 - “True” and “measurement” distributions
 - Log-normal and gamma distributions
- Applications
 - Fitting to plate counts data
 - Fitting to MPN triplet data
 - Effect on microbiological criteria
- Conclusions and further work

Within-batch variability of microbial load

■ Process

Colony forming units/g at different stages of a model meat process in which frozen, boxed, boneless beef was thawed, minced and bowl chopped. Random 20 g subsamples of meat were taken at each stage

Process stage	Mean log (\bar{x})	Variance (s^2)	Log average ($\log \bar{A}$)	Calculated log average*	Subsamples (n)
Thawed beef	5.934	0.334	6.514	6.599	20
Minced beef	6.389	0.108	6.512	6.521	20
Bowl chopped beef	6.475	0.075	6.563	6.570	20

■ Sampling unit size

Reduction in variance in c.f.u./g with increasing subsample size for sausage meat (Brown 1977)

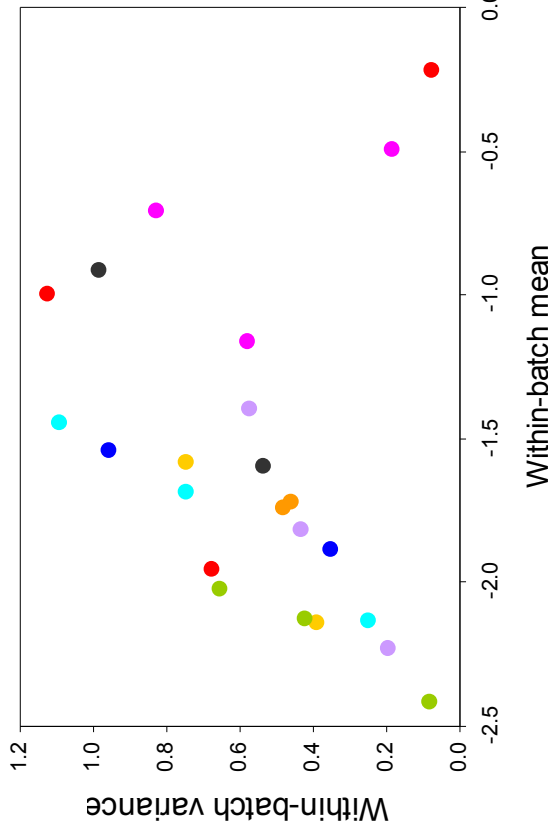
Subsample weight (g)	Mean-log (\bar{x})	Variance (s^2)	Calculated log average*	Subsamples (n)
1	6.58	0.56	7.22	35
2	6.90	0.52	7.56	35
5	7.28	0.23	7.55	35
10	7.30	0.25	7.59	35
15	7.39	0.16	7.57	35
20	7.43	0.11	7.58	20

...Within-batch variability of microbial load

- Level of contamination
 - “Variability of the quality of the end product is assumed more or less constant” (Smelt and Quadt, 1990)
 - “Another assumption is that the variance of the samples is the same for a little or a highly contaminated lot” (Whiting et al., 2006)
 - “A logarithmic transformation is used to remove the correlation between means and variances that have been observed often for plate count data” (FDA, 2009)

...Within-batch variability of microbial load

■ Level of contamination



$n=30$ per batch sampling

□ Should we consider within-batch variance constant?

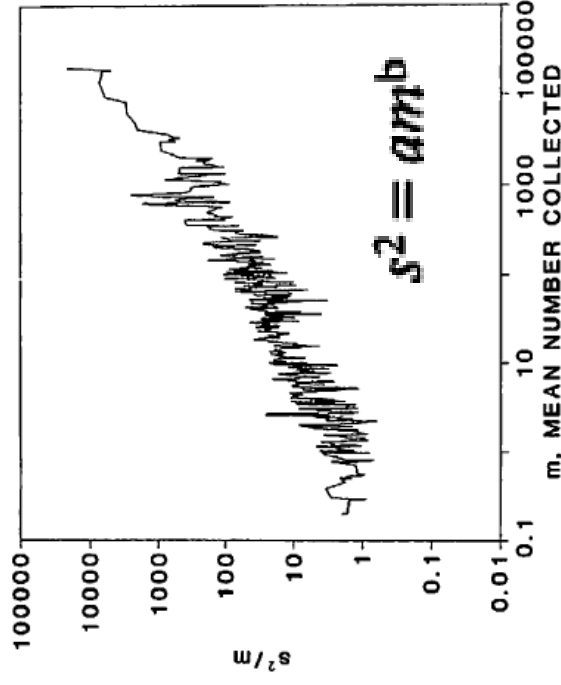
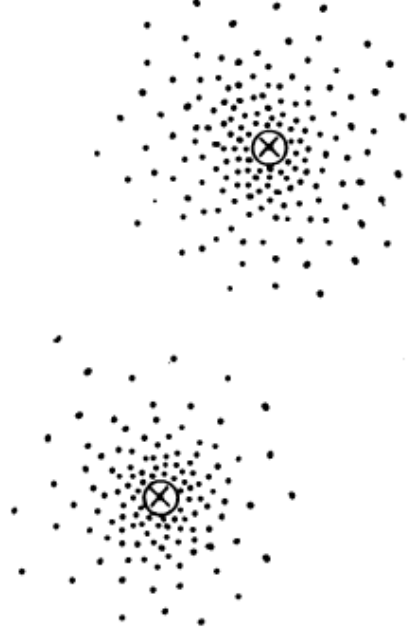


Figure 9.1 Variation in the variance mean ratio (s^2/m) with mean density (no. sample⁻¹) in 1200 sets of replicate zooplankton samples. Data are from marine and freshwater systems (Downing et al., 1987). Means and variances were calculated on

...Within-batch variability of microbial load

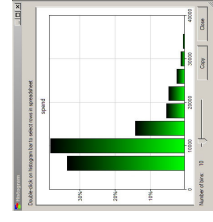
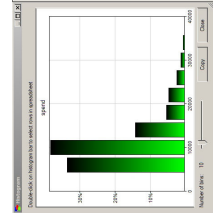
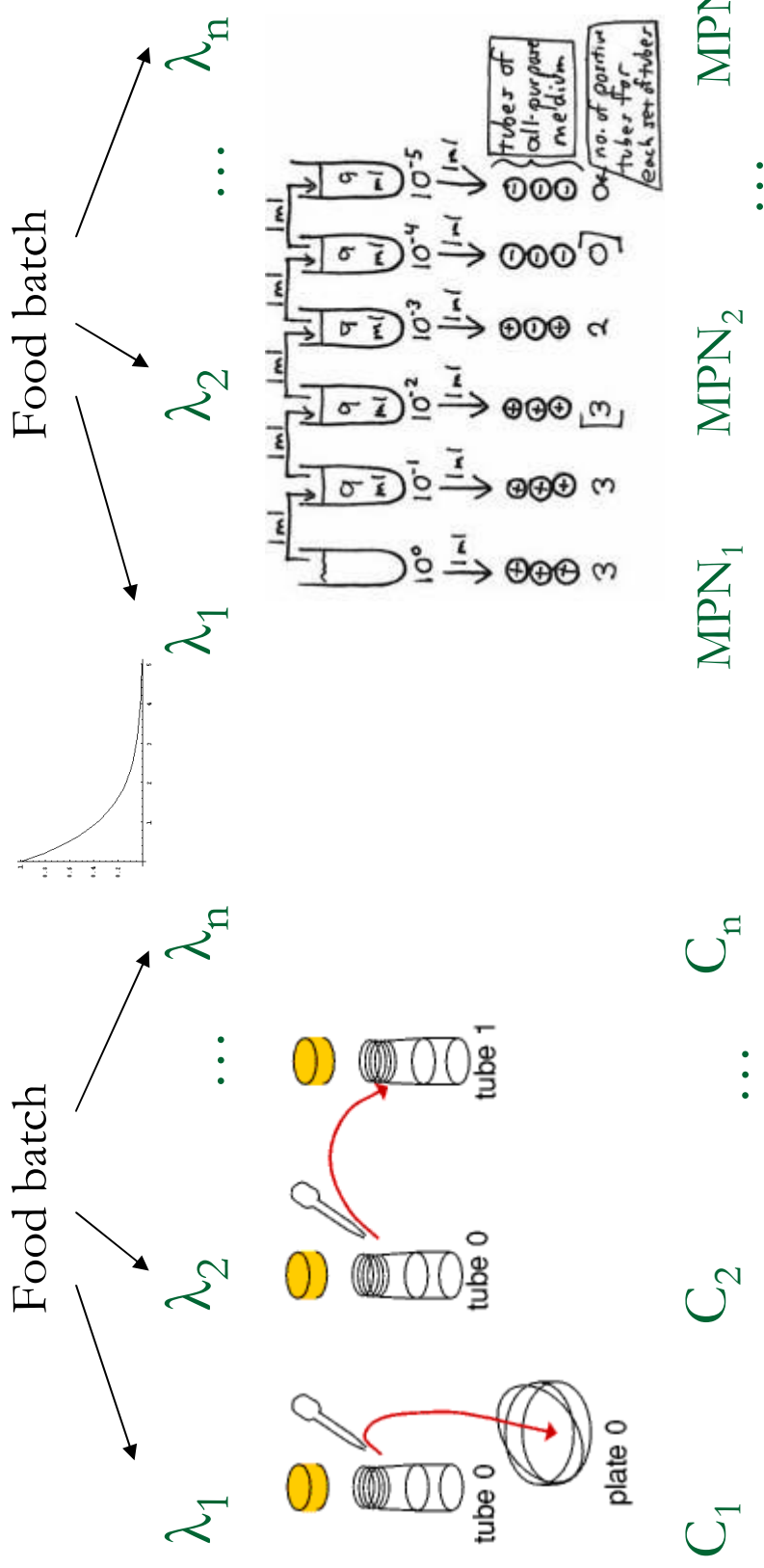
- Clustering or aggregation of bacterial cells
 - Cells adhere to one another after duplication → occurrence of cell clumps rather than individual cells
 - Contagious process



Q?

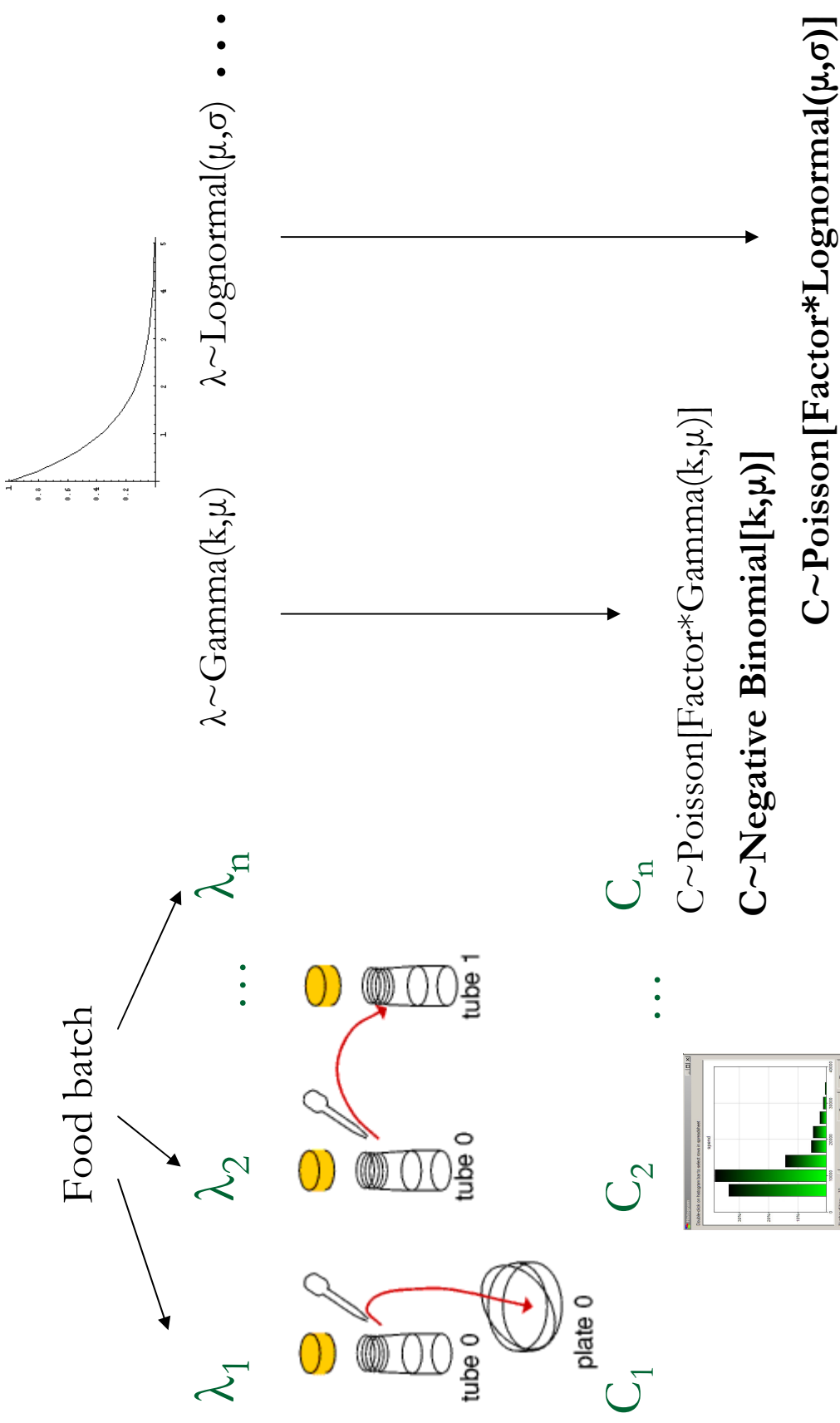
Q?

'True' and 'Measurement' Distributions



Measurement and sampling errors

...‘True’ and ‘Measurement’ Distributions



Lognormal (true) distribution

$$f_X(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}, \quad x > 0$$

- Known advantages
 - Easy to fit (straight to the transformed plate counts) and can approximate the heterogeneity in a food batch.
 - Traditionally, used for convenience and to induce data normality.
- Known drawback
 - Not sure how to proceed with data with high proportion of zeros: left-censored model (?)

Poisson-lognormal (observed) distribution

$$f_{N_i}(n_i) = \int_{-\infty}^{\infty} \frac{\exp(-\gamma_i) \gamma_i^{n_i}}{n_i!} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2} \left(\frac{\varepsilon_i}{\sigma}\right)^2\right) d\varepsilon_i$$

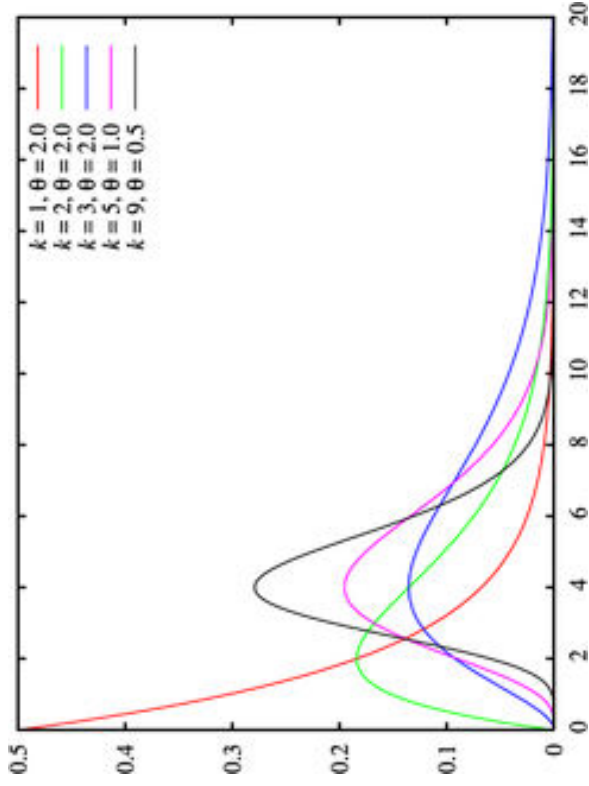
- Advantages
 - Can be fitted (at least in theory) to ‘observed’ data consisting of high proportion of zeros, to infer the ‘unobserved’ lognormal distribution.
- Drawback
 - Difficult to fit! PDF does not have a closed form!

Gamma (true) distribution

$$f(x; k, \theta) = x^{k-1} \frac{e^{-x/\theta}}{\theta^k \Gamma(k)} \text{ for } x > 0 \text{ and } k, \theta > 0.$$

$X \sim \Gamma(k, \theta)$ or $X \sim \text{Gamma}(k, \theta)$.

- Known advantage
 - Can model high proportion of zeros of the observed data (Gonzales-Barron et al., 2010)
- Known drawback (?)
 - Related to the theoretical interpretations of the observed Negative Binomial (k, μ) distribution



Negative binomial (observed) distribution

$$f(k) = \frac{\Gamma(k+r)}{k! \cdot \Gamma(r)} (1-p)^r p^k = \frac{\lambda^k}{k!} \cdot \frac{\Gamma(r+k)}{\Gamma(r) (r+\lambda)^k} \cdot \frac{1}{\left(1 + \frac{\lambda}{r}\right)^r}$$

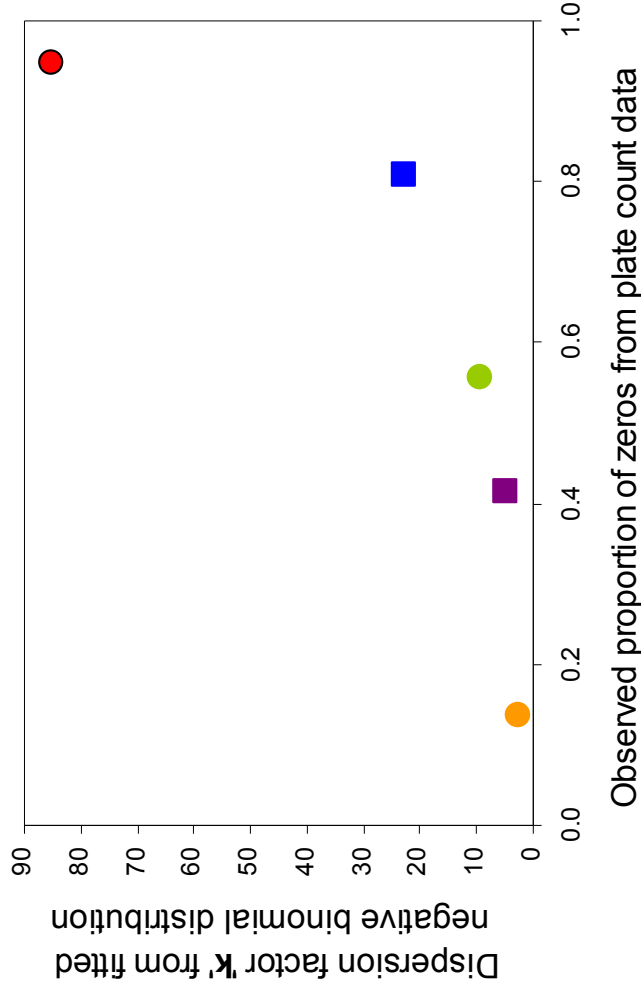
$$X \sim \text{NB}(r, p)$$

■ Theoretical framework

- True contagious process: Each “favourable” event increases the probability of future “favourable” events → **Generalised Poisson processes**: Polyá, Neymann, Double-Poisson → **k=spatial inhomogeneity**.
- Random inhomogeneity: Each event is independent. An inhomogeneity in the populations leads to an apparent contagion due to our method of sampling → **Heterogeneous Poisson processes**: Neymann, Poisson-Gamma, Poisson-Lognormal → **k=sampling inhomogeneity**
- A Poisson mixture of Logarithmic(p)-distributed random variables.
- Sum of geometric distributions...

Dispersion factor 'k'

- Each point is a different data set (bacteria-foodstuff)!
- The higher the “k”, the more skewed the distribution, and the higher the proportion of zero counts it can represent.
- As “k” → 0, NB → Poisson

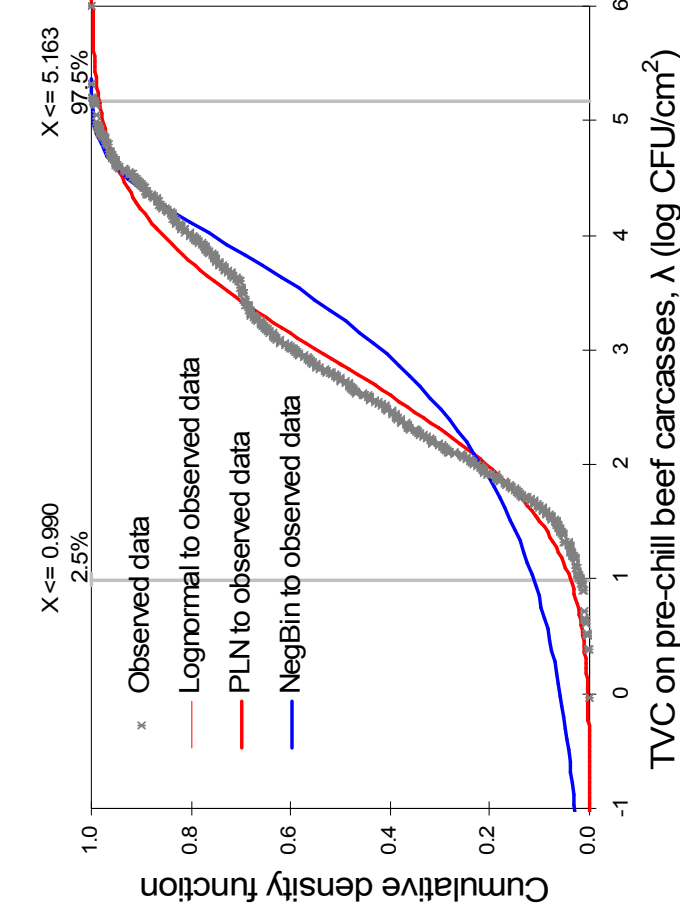


Applications

- Fitting to plate counts data
 - High bacterial populations
 - Low bacterial populations
- Fitting to MPN triplet data
- Effect on microbiological criteria

Fitting to plate count data: High bacterial populations (All positive counts)

Distribution of ‘unobserved’ λ



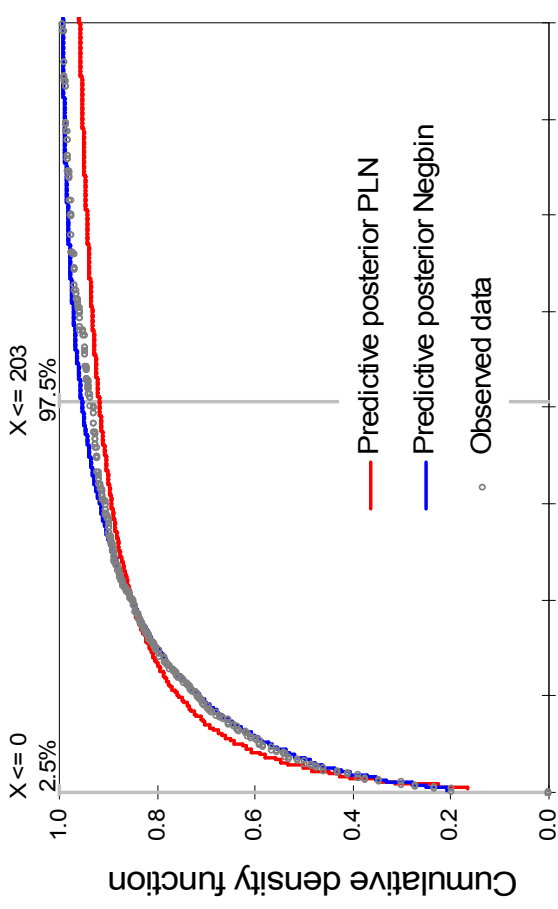
Findings:

- To estimate “lambda” for high populations, fitting plate counts to Poisson-lognormal (observed distribution) does not produce significantly different results to fitting straight to lognormal.
- The lognormal fits better than the gamma for high populations.

Fitting to plate count data:

Low bacterial populations (20% zero counts)

Distribution of 'observed' CFU



Findings:

- ❑ Negative binomial fitted better to the plate count data than the Poisson-lognormal.
- ❑ Poisson-lognormal can represent the proportion of zero counts, but for its nature of being so skewed, overestimates the 'unobserved' lambda.
- ❑ Not a good idea to fit lognormal straight to the observed data when there are many zero counts.

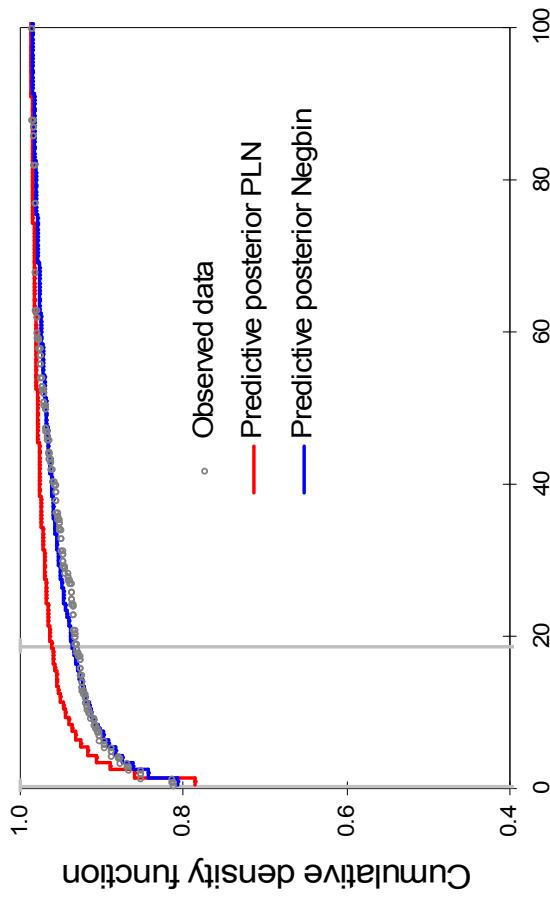
Fitting to plate count data:

Even lower bacterial populations (82% zero counts)

Distribution of 'observed' CFU

Findings:

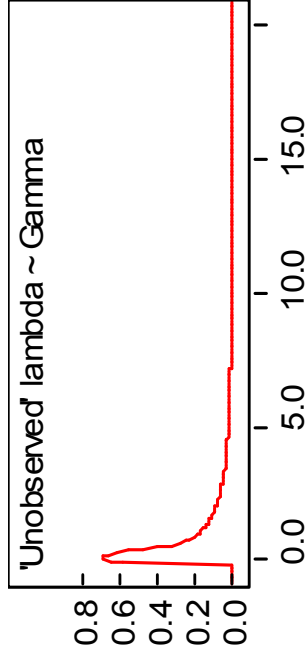
- At this level, Poisson-lognormal under-represents the proportion of zero counts.
- Too skewed that overestimates the 'unobserved' lambda



E. coli plate counts from post-chill beef carcasses (CFU)

Fitting to MPN triplets data

Salmonella MPN in sausages: (3-1-1), (1-0-0), (0-0-0)...



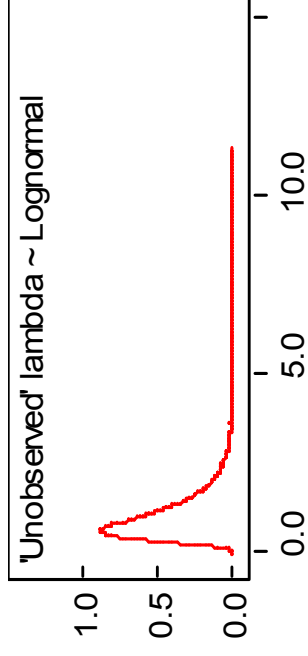
$E(\lambda) = 2.09$ cells/g

95% CI = 0 – 13.90 cells/g

DIC = 24

■ Findings:

- Gamma had a longer tail than lognormal
- Gamma distribution may (?) represent better the within-batch variability for low counts in a MPN framework



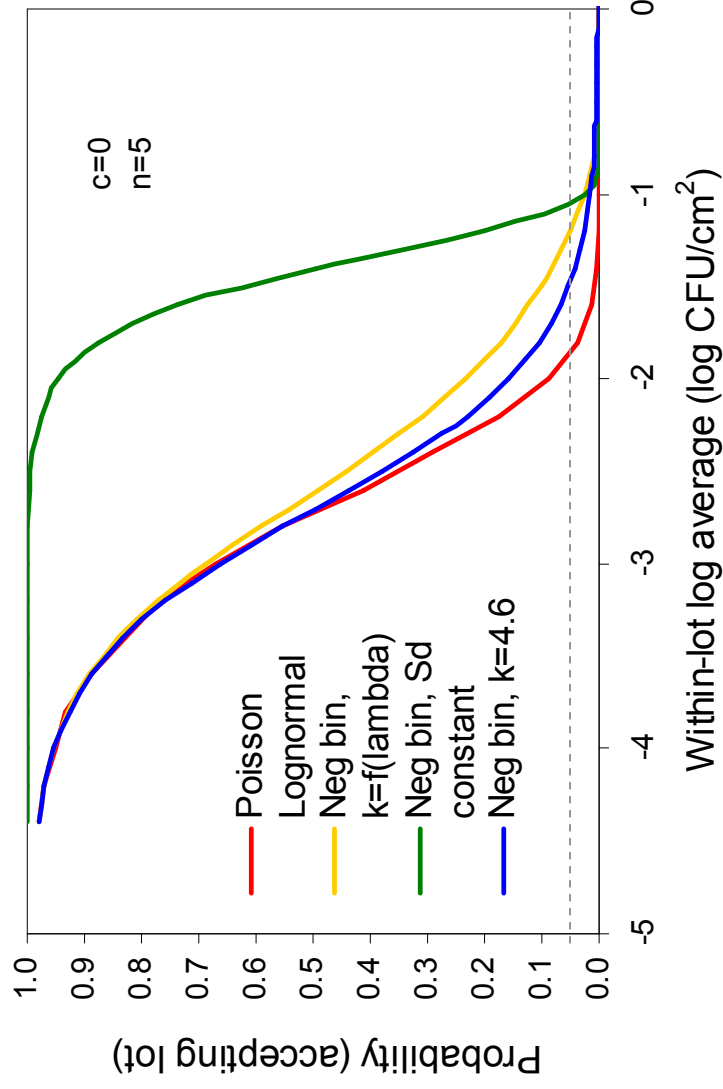
$E(\lambda) = 1.06$ cells/g

95% CI = 0.25 – 2.95 cells/g

DIC = 39

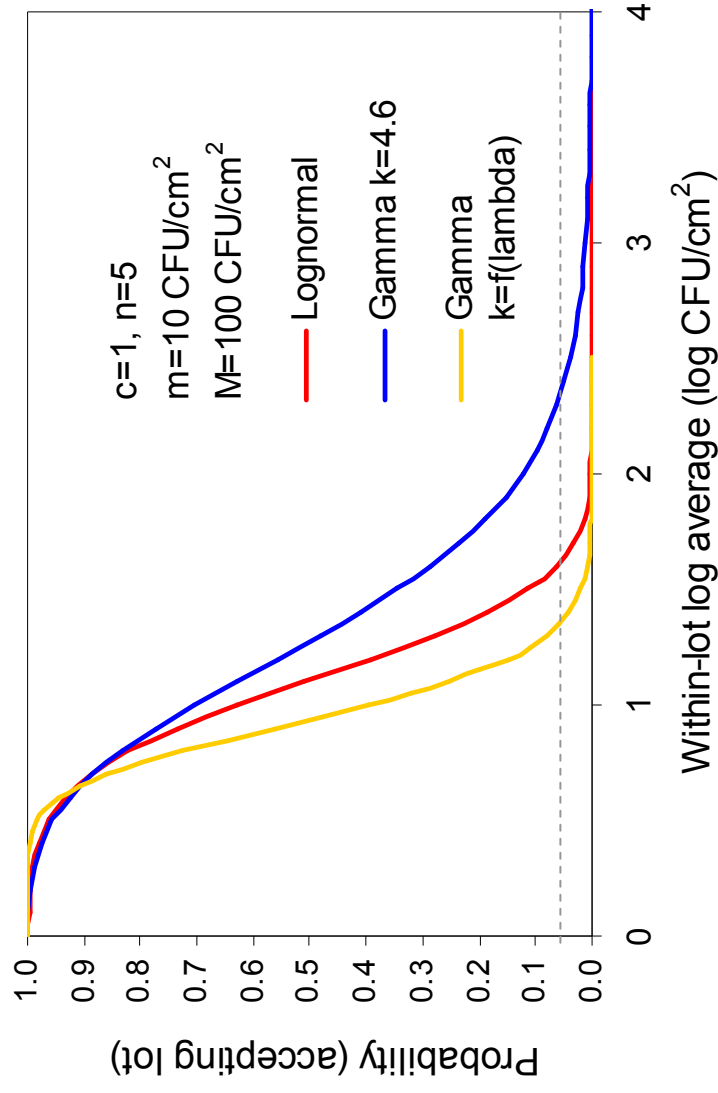
Effect on microbiological criteria

- Attributes sampling plan
- Poisson log-normal offers lower level of safety



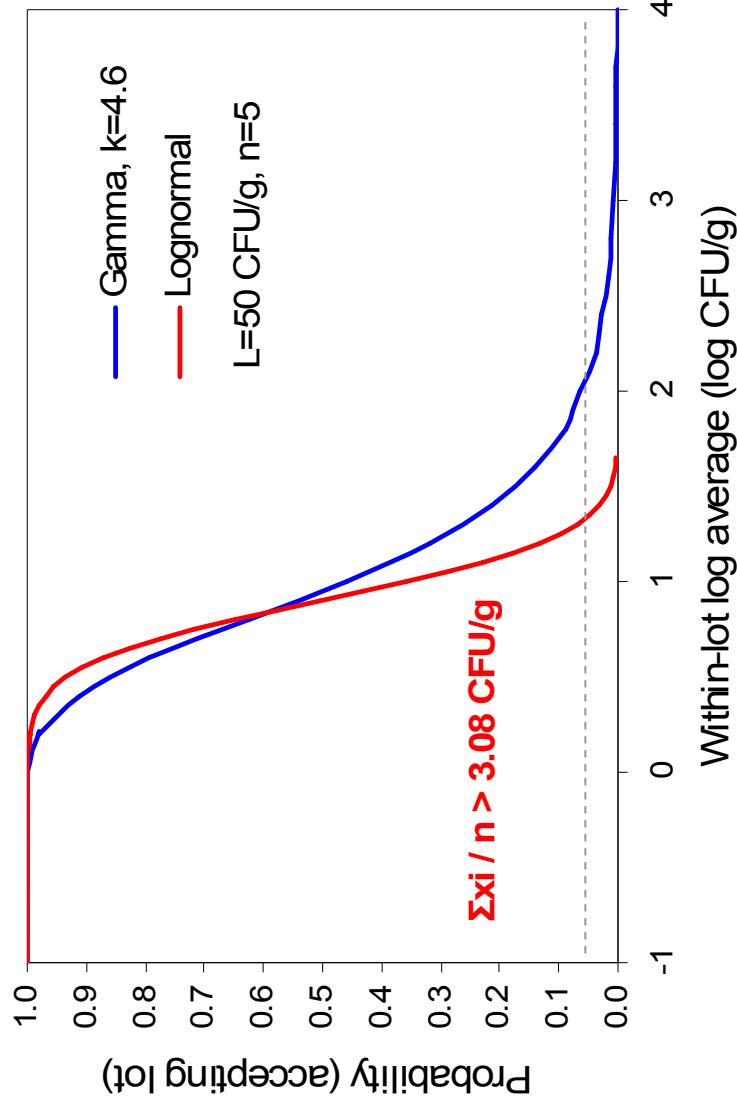
...Effect on microbiological criteria

- Attributes
three-class
sampling plan
(LOD=10
CFU/cm²)



...Effect on microbiological criteria

- Variables sampling plan
- A decision rule was worked out based on the lognormal
- That rule would operate at a lower level of safety if lambda followed a gamma



Conclusions

- Lognormal seems more suitable for high microbial counts, while gamma for low microbial counts. However, other distributions should be assessed.
- For high counts, lognormal can be fitted straight to transformed plate count data.
- For low counts, negative binomial should be fitted to plate count data in order to model the (assumed) true distribution (gamma).
- Setting microbiological criteria is substantially affected by the population distribution, and therefore further research is needed.

...Conclusions

- In setting microbiological criteria, revise:
 - Lognormal?
 - True or measurement distribution?
 - Should variance be assumed constant independently of the batch mean?
 - Should 'k' be assumed constant independently of the batch mean?
 - Sample weight plays a role in our 'observed' variability!!

Further work to answer these questions

- Modelling macro-level: A LOT OF data is needed
 - ❑ Multiple counts from batches (good idea to assess historical data)
 - ❑ Within-lot mean and standard deviation to model relationship
- Modelling micro-level: Understand and model spatial clustering of cells
 - ❑ Single cell behaviour modelling
 - ❑ Behaviour of the population over space and time
 - ❑ Population ecology modelling, spatial aggregation
 - ❑ Scale-dependent theory (rather than scale-free)
- We can move towards “more realistic” assumptions (robust simulation tools, Markov Chain Monte Carlo)

Acknowledgments

- Irish Department of Agriculture, Fisheries and Food

