Advances in Microbiology, **2012, 2, 146-161**
http://dx.doi.org/10.4236/aim.2012.22020 Published Online June 2012 (http://www.SciRP.org/journal/aim)

The Poisson Distribution Is Applied to Improve the Estimation of Individual Cell and Micropopulation Lag Phases

Juan Aguirre[1], Carmen Bravo[2], Juan A. Ordóñez[1], Gonzalo García de Fernando[1]
[1]Departamento de Nutrición, Bromatología y Tecnología de los Alimentos, Facultad de Veterinaria, Universidad Complutense, Madrid, Spain
[2]Servicio Informático de Apoyo a Docencia e Investigación, Universidad Complutense, Madrid, Spain
Email: juaguirr@vet.ucm.es

Received January 24, 2012; revised February 14, 2012; accepted March 16, 2012

ABSTRACT

Many articles dealing with individual cell lag phase determination assume that growth, when observed, comes from one cell. This assumption is not in agreement with the Poisson distribution, which uses the probability of growth in a sample to predict how many samples contain one, two, or some other number of cells. This article analyses and compares different approaches to improve the accuracy of lag phase estimation of individual cells and micropopulations. It argues that if the highest initial load, as predicted by the Poisson distribution, is assigned to the sample with the shortest lag phase, the second highest to the sample with the second shortest lag phase and so on, the resulting lag phase distributions would be more accurate. This study also proposes the use of a robust test, permutation test, to compare lag phase distributions obtained in different situations.

Keywords: Individual Cells; Micropopulations; Lag Phase; Poisson Distribution

1. Introduction

The measuring of any parameter characterizing the microbial growth is essential for any quantitative microbial risk assessment. Then, to know the microbial lag phase length of viable cells is critical, especially in RTE products, which nature and storing conditions may allow the growth of viable, pathogen or not, bacteria. In the case of populations of thousands or hundreds of viable cells, the lag phase, is quite reproducible if the pre-inoculation and growth conditions are constant. However, the lag phase of populations composed by few cells, or even by only individual cells, is inherently variable. Therefore, it is understandable that researchers [1-9] have paid attention to the distribution of single cells lag times and to the techniques that can measure them. Measuring the lag time of individual cells requires direct microscopic observation [4,8] or techniques to isolate single cells [10]. Cell isolation can be achieved by diluting [2], sorting by flow cytometry [11] or inactivating all organisms except one [9]. When growth is detected in some samples and not in others, it is commonly assumed that growth comes from one cell [12]. The number of samples must always be high for reliable mathematical treatment. It is recommended that approximately 100 samples show growth [13], and this figure must not be a high percentage of the samples. Guillier *et al.* [5] stated that if 35% of samples show growth, this should not significantly affect individual cell lag phase distributions because at least 80% of samples contain one cell, according to the Poisson distribution function.

"Growth/no growth sampling" has been widely applied to foods and opaque liquids; in the special case of translucent liquids, an apparatus called the Bioscreen C can be used to construct 200 growth curves simultaneously for the same temperature, on the basis of the turbidity resulting from microbial growth. If the specific growth rate (μ) under the experimental conditions is known, the lag phase is determined using the following equations. In the case of translucent liquids analyzed using Bioscreen, the equation is [14]:

$$\text{Lag} = T_d - \left[\text{Ln}\left(N_d\right) - \text{Ln}\left(N_0\right)\right] \quad (1)$$

where T_d is the detection time, *i.e.* the time needed to reach an arbitrary absorbance (turbidity), Ln(N_d) is the natural logarithm of the number of cells generating such absorbance, Ln(N_0) is the natural logarithm of the number of organisms in the inoculum, and μ is the specific growth rate. In the case of opaque samples [6], the equation is:

$$\text{Lag} = T_{\text{count}} - \left[\text{Ln}\left(x_{\text{count}}\right) - \text{Ln}\left(x_{\text{initial}}\right)\right] \quad (2)$$

Copyright © 2012 SciRes.

where T_{count} is the time between inoculation and plating of the sample, $Ln(x_{count})$ is the natural logarithm of the cell number detected at T_{count}, $Ln(x_{initial})$ is the natural logarithm of the initial number of bacteria and μ is the specific growth rate.

When a certain percentage of samples does not show growth, the assumption that growth in the other samples is due to one cell contravenes the predictions of the Poisson distribution. Several researchers have used the Poisson distribution to calculate the proportion of growth-positive samples initially containing more than one cell [5,11,15,16]. McKellar and Hawke [17] recognised that one of the limitations of the Bioscreen as a tool to study single cell behaviour is that it is difficult to ensure that the growth in any given positive well arose from a single cell. Earlier, some authors [2,15] performed a series of binary dilutions to have one cell per sample. Francois *et al.* [2] observed that single cells should be found in wells of Bioscreen microtitre plates where the mean cell number added to each well was less than one. These authors advocate pooling data from the last five binary dilution series to maximise the number of replicate wells; these series contained 0.7812, 0.3906, 0.1951, 0.0977, and 0.0977 cells per well, from a theoretical mean dilution range.

According to the Poisson probability function, if a determined number of occurrences (ρ is expected, then the probability that there are exactly k occurrences (k being a non-negative integer number, $k = 0, 1, 2, \cdots$) is:

$$\phi(\kappa,\rho) = \left(e^{-\rho} * \rho^{\kappa}\right) \Big/ \kappa! \quad (3)$$

where e is the base of the natural logarithm; k is, in our case, the number of organisms in a sample, and the probability of k is given by the function; ρ is a positive real number, which expresses the average number of cells per sample; and $k!$ is the factorial of k. To highlight the relevance of the data that Equation (3) offers, the **Table 1** has been built up. This table shows the percentages of samples predicted by Equation (3) that would contain a determined number of viable cells as a function of the percentage of samples, in which growth was detected. The average number of cell per sample (m) is also shown in **Table 1**. This average is calculated by assuming that the number of cell per sample follows a Poisson distribution. Hence, the following equation was used:

$$m = Ln(P) \quad (4)$$

where P is the probability of there is not any viable cell in a sample. Applying Equation (3) to the data of Francois *et al.* [2], with an average number of cells per sample of 0.78, indicates that 65% of the positive samples contain one cell, 25.3% contain two, 6.6% three and 1.3% four. These figures suggest that the estimated lag phase determinations for individual cells will have a certain error. Indeed, Baranyi *et al.* [18] affirmed that the greater the Poisson parameter (ρ, average number of cells per sample), the less accurately Equations (1) and (2) estimate the distribution of the single cell lag time. If samples are considered to contain only one cell, the value of $Ln(N_0)$ in Equation (1) and $Ln(X_{initial})$ in Equation (2) is zero. However, if the predictions of the Poisson function are applied, we have to assume that some samples contain two, three or more cells, which is an undisputed fact in most real samples. In this case, $Ln(N_0)$ and $Ln(X_{initial})$ are positive numbers that lengthen the lag phase of such samples.

The aim of this study is to compare the individual cell and/or micropopulation lag phase distributions obtained by assuming that all samples with growth contain one

Table 1. Percentage of samples with a determined number of cells, as predicted by Poisson function (Equation (3)).

% of samples with growth	Average No. of cell/sample	% of samples with a determined No. of cells						
		1	2	3	4	5	6	7
90	2.303	23	27	20	12	5	2	1
80	1.609	32	26	14	6	2	0	0
70	1.204	36	22	9	3	1	0	0
60	0.916	37	17	5	1	0	0	0
50	0.693	35	12	3	1	0	0	0
40	0.511	31	8	1	0	0	0	0
30	0.357	25	5	1	0	0	0	0
20	0.223	18	2	0	0	0	0	0
10	0.105	9	1	0	0	0	0	0

Copyright © 2012 SciRes.

cell, with the distributions obtained by assuming a different number of cells per sample according to the Poisson distribution function.

2. Material and Methods

2.1. Simulation

A simulation was generated considering a different average number of cells per sample (0.2 - 2.0). To create the simulation, 100 values of lag phases were randomly generated by assigning them values from 40 to 180 arbitrary time units, following a gamma distribution with the following parameters: shape = 5.5, scale = 16.5, mean = 91.7 and standard deviation = 29.4. A specific growth rate (μ) of 0.0693 h^{-1} was also considered. The resulting distribution data are those of Scenario I (see next section).

2.2. Scenarios

Four scenarios were used to calculate the lag phase distributions: Scenario I assumes that all samples contain one cell. Scenarios II-IV use the Poisson distribution function to assign a number of cells to each sample. In Scenario II, the sample with the shortest lag phase contains the highest number of cells, according to the average number of cells per sample and the Poisson table [19], the sample with the second shortest lag contains the second highest number of cells, and so on. In Scenario III, the number of cells is randomly distributed among samples, regardless of the lag phase length. Scenario IV is calculated like Scenario II, except that all samples with more than one cell are not considered. From the data of Scenario I, lag phases were recalculated according to the assumptions of Scenarios II, III and IV and the corresponding distributions were obtained.

2.3. Statistical Analysis

Pairwise comparisons of the variances of lag phase distributions were carried out using a permutation test to analyse homogeneity of the two variances; this bilateral test assumes that the variances ratio is one. Permutation tests are non-parametric significance tests based on permutation resampling without replacement, with observed lag times drawn at random from the original data and reassigned to the two groups being compared. The distribution of possible variance ratios is calculated for all samples assuming the null hypothesis of homogeneity, and the observed ratio is positioned along this distribution. Values falling outside the main distribution rarely occur by chance and therefore give evidence of heterogeneity of variances [20]. Since our study involves multiple comparisons of several groups, a *p*-value correction must be applied in order to minimise the probability of rejecting a true hypothesis. The Holm-Bonferroni *p*-value correction [21] was applied. This correction is less conservative than those of Bonferroni and Sidak [22], which are also applied in the permutation test program described in appendix A.

A permutation test routine including a multiple comparison test was programmed using R language [23], which is described in Annex 1.

2.4. Application of Scenarios to Experimental Data

To check how well the simulations mimic the reality, the four scenarios were applied to the lag phases of *Enterococcus faecalis*, *Pseudomonas fluorescens*, *Salmonella enterica* serovar Enteritidis and *Listeria innocua* subjected to different irradiation treatments in tryptic soy broth (TSB) and cooked ham and subsequently incubated at different temperatures (experimental data from Aguirre *et al.* [9]). Lag phases were estimated according to Equations (1) and (2) after determining the percentage of samples without growth and considering the Poisson function predictions and the scenarios above described.

3. Results

3.1. Simulation

Figure 1(a) shows the increase in the mean lag and **Figures 1(b)** and **(c)** show, respectively, the decrease in the standard deviation and the coefficient of variation as a function of the average number of cells per sample. The dashed line in **Figure 1** shows the mean and standard deviations of the data in Scenario I. As expected, the greater the number of cells per sample is, the larger the difference between the mean and standard deviation of Scenario I data and those of the others. The distributions of Scenario III were not considered further because the averages were identical to those of Scenario II and their standard deviations were very close to those of Scenario I (data not shown).

Comparison of the distributions obtained in Scenarios I and II, I and IV and II and IV (**Figure 2**) shows that the higher the more probable number per sample is, the smaller the *p* value for comparisons of the mean and standard deviation, according to the permutation test. Significant differences ($\alpha < 0.05$) were found among the three comparisons for the mean and between Scenarios I and II and I and IV for the standard deviation.

3.2. Application of Scenarios to Experimental Data

Table 2 summarises the experimental data, including the expected inactivation according to the irradiation applied, the average number of surviving cells per sample and the

Copyright © 2012 SciRes.

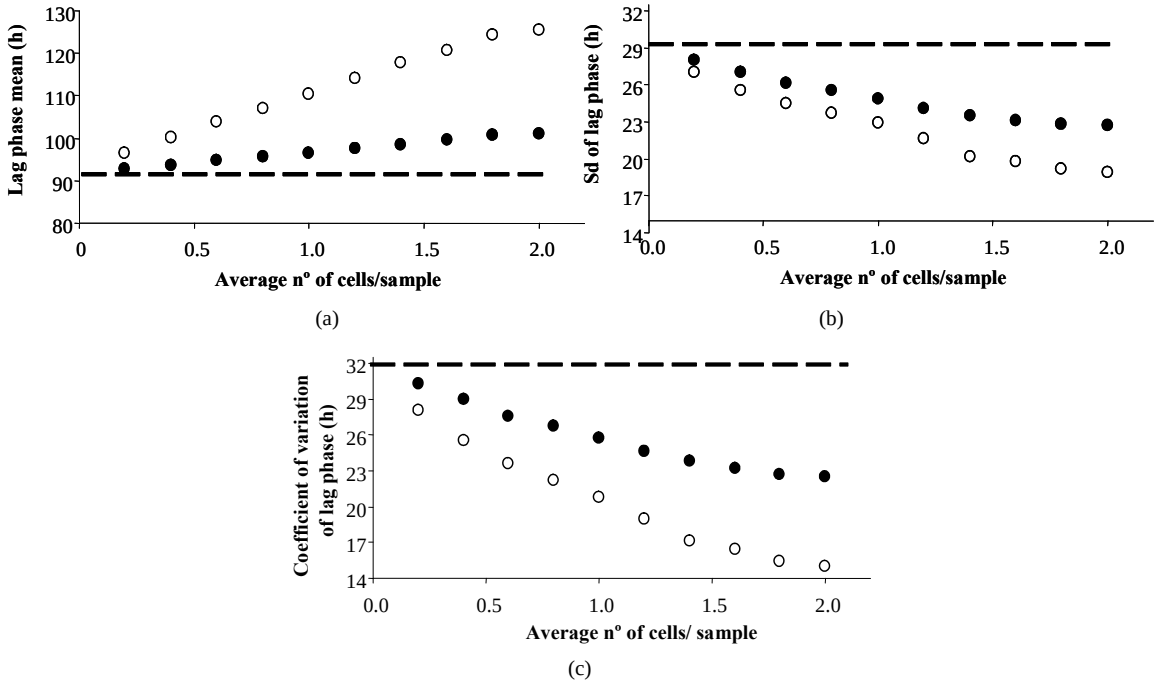

Figure 1. Effect of the average number of cells per sample on the mean (a), standard deviation (b) and coefficient of variation (Sd*100/mean, C) of lag phase distribution in a model system assuming that lag phases follow a gamma distribution and that $\mu = 0.0693\ h^{-1}$. Dashed lines represent the mean, standard deviation and coefficient of variation assuming that all samples contain one cell (Scenario I). Solid symbols show the results assuming a variable number of cells per sample, and assuming that the sample with the shortest lag phase contains the highest number of cells, the sample with the second shortest lag contains the second highest number, and so on, according to the Poisson distribution function (Scenario II). Empty symbols represent the results following the assumptions in (Scenario II) but including only samples with one cell (Scenario IV).

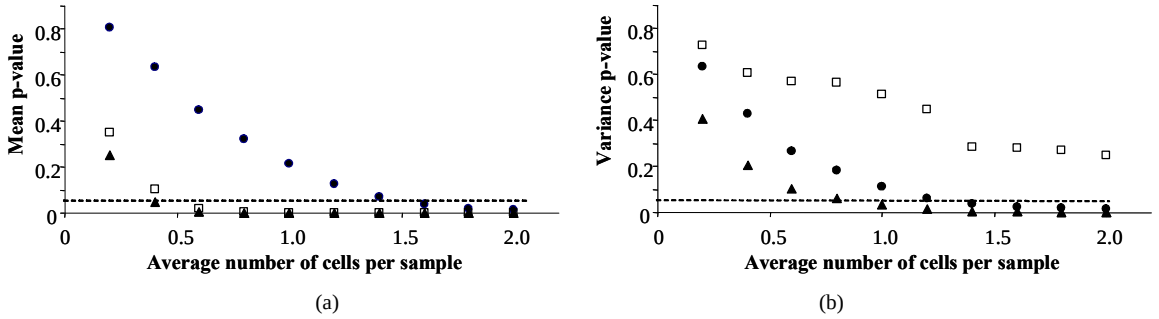

Figure 2. Comparison by permutation test of means (a) and variances (b) of lag phase distributions obtained when all samples with growth are assumed to contain one cell (Scenario I) with those obtained under other assumptions (Scenarios II and IV). The solid circles show the comparison between (Scenarios I and II), in which samples with growth are assumed to contain one or more cells according to the truncated Poisson distribution, and assuming that the sample with the shortest lag phase is supposed to have the highest inocula. The solid triangles show the comparison between (Scenarios I and IV), which makes the same assumptions as (Scenario II) but excludes samples containing more than one cell. Empty squares show the comparison between (Scenarios II and IV). Dashed lines represent the α-value = 0.05. Data below the line indicate significant differences.

characteristics of the distributions (mean, standard deviation and coefficient of variation) for each substrate, treatment (Aguirre *et al.* [9]) and scenario. Scenario III was not considered for the reasons mentioned above. As in the simulations, analysis of the experimental data showed that the mean lag phase when growth is assumed to be due to a variable number of cells (Scenario II) was equal to, or higher than, the mean when growth is assumed to be due to a single cell (Scenario I). In contrast, the standard deviations in Scenario II were always lower than those in Scenario I, except in the case of *Salmonella* Enteritidis growing in ham at 7˚C after no treatment.

The permutation test was used to compare the distributions between Scenarios I and II, I and IV and II and IV.

Copyright © 2012 SciRes.

Table 2. Lag phases of organisms surviving treatment on ham and TSB, as calculated using four scenarios. Explanation on scenarios is in legend of Figure 1.

Organism	Susbstrate and growth temperature	Expected inactivation (log cycles)	Total n° of samples	Nº of samples with growth	Average n° of cells per sample	Mean value of lag phase (h)	Sd (h)	Coefficient of variation (100 Sd/mean)	Mean value of lag phase (h)	Sd (h)	Coefficient of variation (100 Sd/mean)	Mean value of lag phase (h)	Sd (h)	Coefficient of variation (100 Sd/mean)	Mean value of lag phase (h)	Sd (h)	Coefficient of variation (%) (100 Sd/mean)
Entero-coccus faecalis	TSB, 18°C	0.00	100	79	1.56	6.09	2.16	35.45	7.38	1.13	15.31	7.38	2.47	33.52	8.27	1.01	12.19
		1.05	100	85	1.90	13.30	2.67	20.05	14.85	1.55	10.41	14.85	3.17	21.32	16.38	1.50	9.19
		2.10	100	76	1.43	30.84	3.53	11.44	32.00	2.58	8.06	32.00	3.37	10.52	33.57	2.15	6.39
		4.20	100	86	1.97	69.92	7.04	10.07	71.53	5.89	8.23	71.53	6.93	9.69	78.01	3.91	5.01
	Ham, 7°C	0.00	106	79	1.39	59.79	7.92	13.24	71.19	7.07	9.93	71.19	12.86	18.06	65.79	6.48	9.86
		0.90	109	66	0.94	109.59	9.62	8.78	117.27	5.27	4.49	117.27	12.79	10.90	115.80	5.95	5.14
		1.80	114	70	0.94	226.91	19.76	8.71	234.40	14.17	6.04	234.85	21.13	9.00	237.33	17.08	7.20
		3.59	121	68	0.78	458.30	29.58	6.45	464.51	23.11	4.98	464.51	29.08	6.26	473.94	22.26	4.70
Salmonella Enteritidis	TSB, 16°C	0.00	100	49	0.67	34.50	3.90	11.30	35.42	3.11	8.77	35.42	3.90	11.01	36.17	3.39	9.38
		2.40	100	50	0.69	89.79	6.06	6.75	90.69	5.21	5.75	90.69	6.07	6.70	90.94	6.79	7.47
		4.79	100	75	1.39	150.10	8.92	5.94	151.96	7.66	5.04	151.96	9.86	6.49	156.46	8.91	5.69
		5.99	100	52	0.73	187.33	13.44	7.17	188.36	12.39	6.58	188.36	12.95	6.87	193.51	11.47	5.93
	Ham, 7°C	0.00	99	43	0.56	265.38	4.82	1.82	271.82	10.90	4.01	271.82	11.50	4.23	266.16	5.24	1.97
		1.10	105	46	0.58	306.28	8.27	2.70	312.81	8.25	2.64	312.81	13.69	4.38	310.07	6.62	2.13
		2.19	115	47	0.53	433.17	22.75	5.25	439.56	17.96	4.09	439.56	26.09	5.93	441.49	20.45	4.63
		4.39	135	60	0.58	585.64	36.33	6.20	591.80	31.76	5.37	591.80	35.94	6.07	596.55	35.07	5.88
Listeria innocua	TSB, 16°C	0.00	100	54	0.78	24.53	4.23	17.25	25.44	3.27	12.86	25.44	4.14	16.28	27.70	3.49	12.59
		1.08	100	39	0.49	46.44	13.72	29.54	46.93	13.32	28.38	46.93	13.71	29.22	49.48	13.82	27.94
		2.16	100	80	1.61	56.96	15.51	27.24	58.72	14.59	24.84	58.72	15.33	26.10	69.12	18.84	27.25
		4.31	100	59	0.89	130.73	22.19	16.98	131.69	21.39	16.24	131.69	21.94	16.66	140.11	20.85	14.88

Copyright © 2012 SciRes.

Copyright © 2012 SciRes.

Continued

		0.00	106	71	1.11	229.70	5.35	2.33	235.08	4.49	1.91	235.08	8.34	3.55	233.00	4.82	2.07
		1.02	123	81	1.08	285.27	6.56	2.30	290.56	3.94	1.35	290.56	10.02	3.45	289.75	4.91	1.69
	Ham, 7˚C	2.03	119	71	0.92	371.53	10.40	2.80	376.04	6.94	1.85	375.60	11.32	3.01	377.78	8.26	2.19
		4.06	109	85	1.51	512.78	18.10	3.53	520.74	13.83	2.66	520.74	20.07	3.85	528.98	18.52	3.50
		0.00	100	86	1.97	4.30	2.27	52.84	5.76	1.24	21.53	5.76	2.51	43.63	7.00	1.14	16.22
		1.32	100	46	0.62	35.13	5.13	14.60	35.56	4.70	13.22	35.56	5.09	14.32	36.96	4.69	12.70
	TSB, 16˚C	2.63	100	76	1.43	45.06	9.31	20.66	46.12	8.72	18.91	46.12	9.31	20.19	50.11	10.14	20.24
		3.95	100	68	1.14	61.20	15.15	24.76	62.09	14.33	23.08	62.09	15.16	24.41	71.69	11.65	16.25
		6.58	100	53	0.76	98.45	29.17	29.63	99.04	28.51	28.79	99.04	29.20	29.49	113.38	20.59	18.16
		0.00	110	76	1.17	89.27	11.88	13.31	97.77	4.96	5.07	97.77	13.49	13.80	98.54	5.39	5.47
		1.21	138	101	1.31	114.14	17.57	15.40	123.18	9.80	7.95	123.18	20.09	16.31	130.21	9.45	7.26
Pseudo-monas fluorescens	Ham, 2˚C	2.43	149	111	1.35	210.75	22.04	10.46	220.02	14.84	6.74	220.02	24.41	11.10	227.93	16.39	7.19
		4.86	116	75	1.05	382.67	30.97	8.09	389.99	23.92	6.13	389.99	31.89	8.18	404.16	20.75	5.13
		7.28	101	65	1.02	533.26	50.24	9.42	540.70	42.83	7.92	540.70	51.58	9.54	566.11	32.25	5.70
		0.00	101	72	1.24	48.60	3.47	7.13	52.85	2.73	5.16	52.85	4.80	9.08	50.60	0.62	1.22
		1.21	119	83	1.20	59.87	8.12	13.57	63.92	4.84	7.57	63.92	9.05	14.16	66.72	5.00	7.49
	Ham, 7˚C	2.43	136	101	1.35	127.61	14.99	11.75	132.02	11.30	8.56	132.02	17.10	12.95	138.66	9.76	7.04
		4.86	105	67	1.02	271.84	22.06	8.12	275.36	18.34	6.66	275.36	23.62	8.58	286.95	6.59	2.30
		7.28	99	65	1.08	407.68	27.16	6.66	411.31	23.65	5.75	411.31	27.28	6.63	425.39	16.35	3.84

Several significant differences ($\alpha < 0.05$) were found and in order to clarify the reasons of such differences, the average number of cells per sample and the specific growth rate were plotted (**Figure 3**).

4. Discussion

Obviously, to consign to oblivion the Poisson function of distribution, when growth is observed in a determined percentage of homogeneous samples, and consider that this growth comes from one viable cell, is erroneous because the probability of that a significant number of samples contains more than one cell is very high (see **Table 1**); obviously, the higher the average number of cells per sample is, the higher the probability of finding samples with more than one cell.

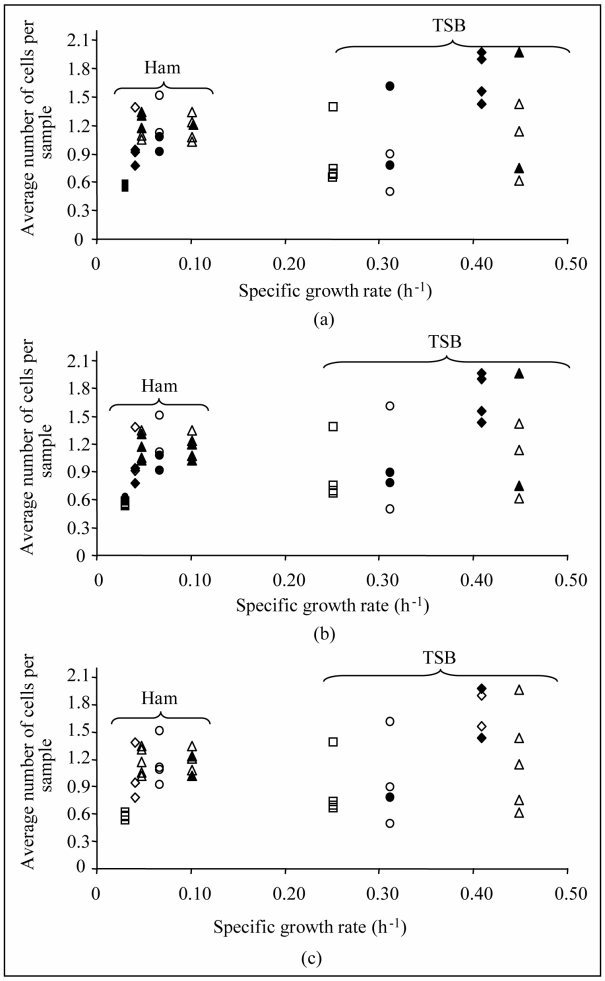

Figure 3. Comparison of lag phase distributions between Scenarios I and II (a), I and IV (b), and II and IV (c) using the permutation test. Empty symbols indicate that there are no significant differences ($\alpha > 0.05$) between distributions. Solid symbols indicate significant differences ($\alpha < 0.05$). Data were taken from Aguirre *et al.* (2011) and correspond to *Listeria innocua* (circles), *Pseudomonas fluorescens* (triangles), *Salmonella* Enteritidis (squares) and *Enterococcus faecalis* (diamonds).

In order to determine more accurately the lag phase, we propose to assume that the sample with the shortest lag phase contains the highest number of cells, according to the Poisson distribution, the sample with the second shortest contains the second highest number of cells, and so on (Scenario II). This assumption is based on the fact that higher inocula need less number of duplications (*i.e.* less time) to reach a determined number of cells (T_d or T_{count} in Equations (1) and (2), respectively) and then, lag is actually longer than if the growth had been generated by only one cell. Furthermore, it is known that the lag phase of individual cells is variable [1,7,9] and, logically, it is more likely to find a fast cell in starting the growth in samples that contain higher number of cells and, obviously, lag phase of these samples must be shortened because of the "fast" cells. Pin and Baranyi [7], working with single cells and micropopulations, stated that samples with low initial number of cells, showed longer lag times—in average—than those initiated with more cells. All these reasonings support the starting hypothesis.

The greater the number of cells per sample is, the larger the difference between the mean and standard deviation of lag phase distributions (**Figure 1**). Then, when individual cell lag phases are determined, larger errors are expected as the number of cells per sample increases, confirming the Baranyi *et al.* [18] statement. In other words, in practical situations, the greater the number of samples with microbial growth is, the greater the expected errors in individual cell lag phase estimation. **Figure 2** corroborates this finding since the higher the average number of cells per sample is, the more probably is to find significant differences ($\alpha < 0.05$) between the distributions of all scenarios analyzed.

Remarkable differences between Scenarios I and IV were observed in both the mean and the standard deviation (**Figure 1**). It is important to remember that in this case, lag phase data, when one cell per sample is assumed (Scenario I), are compared with the lag phase data of cells with longer lags (IV), which are more likely to contain, actually, one cell. Samples with shorter lag phases are ignored in Scenario IV and, therefore, data from fast cells growing alone are not included. Then, the lag phase average is always biased to long times in this scenario.

Scenarios II and III should model the experimental data more closely than Scenario I because Equations (1) and (2), used to calculate the lag phase, reduce the time to reach a given microbial concentration (T_d or T_{count}) by the time that the initial cell number takes to reach the given concentration $\{[(\mathrm{Ln}(N_d) - \mathrm{Ln}(N_0)]/\mu]$ or $[(\mathrm{Ln}(X_{count}) - \mathrm{Ln}(X_{initial})]/\mu]\}$. This is true regardless of the variability in growth rate observed in the first divisions of a cell [4], because this variability seems to be randomly distributed

Copyright © 2012 SciRes.

[24]. Recalculating the lag phases of samples more likely to contain more than one cell is easily accomplished by substituting N_0 or $X_{initial}$ in Equations (1) and (2), respectively, by the number of cells predicted by the Poisson function. When the lag phase is determined by Bioscreen or in food, assuming that all samples contain one cell, the sample with the shortest lag phase is most likely to contain the highest inoculum, and samples with the longest lag phases are most likely to contain only one cell. Based on this reasoning, if more accurate data on initial microbial concentration can be obtained, the quality of lag phase estimation will improve, although these estimations are not absolutely accurate due to the intrinsic cell variability. Actually, one cell may have a short lag phase, even shorter than that of a micropopulation of two medium or long lag phase cells, even considering that one cell have to duplicate once more than two cells to reach the same number of daughter cells and this duplication time is, of course, considered by Equations (1) and (2) when calculating lag phases.

The analysis of the experimental data of **Table 2** shows that the mean lag phase, when growth is assumed to be due to a variable number of cells (Scenario II), was higher than the mean when growth is assumed to be due to a single cell (Scenario I). Furthermore, the lower the lag phases of Scenario I were, the greater the percentage of increase (**Table 2**). This means that considering that all samples contain one cell in optimum growth conditions or in healthy cells, which imply short lag phases, generates relevant inaccuracies, while the lag phase determination of injured cells and at suboptimal growth conditions, the error may even become negligible (**Table 2**). These statements are only pertinent for the lag phase average because if we consider the transformation from Scenario I to Scenario II, data to data, it is evident that short lag phases estimated according to Scenario I are transformed in longer ones, probably in a non realistic way. An example is shown in **Tables 3** and **4**, which show the lag phases of non-radiated and irradiated (2 kGy) *Enterococcus faecalis* in TSB at 18˚C, respectively (data of distributions are shown in **Table 2**). The analysis of the **Table 3** data allows to affirm that the recalculation of lag phases according to Scenario II, ascribing the highest number of cells to the sample with the shortest lag and the second highest number of cells to the second shortest, etc., result in a too long lag phases for such samples, which hardly correspond to a samples with five cells instead one (the first two data of the left columns of the **Table 3**). The analysis of **Table 4** data drives to the same reasoning, although the lengthening of lag phases due to the irradiation minimizes the differences. In contrast, the standard deviations in Scenario II were almost always lower than those in Scenario I (**Table 2**), which means a lower dispersion of data and, presumably, a

Table 3. Lag phases of non-radiated *Enterococcus faecalis* in TSB at 18˚C calculated by Equation (1) considering a constant μ of 0.409 and assuming: Scenario I: All samples contain 1 cell; Scenario II: The sample in column "Scenario I" with the shortest lag phase contains the highest inoculum, the second shortest, the second highest, and so on. Number of cells per sample was estimated according to the Poisson distribution considering that the average number of cells per sample was 1.56 (Table 2). Then, according to the Poisson function of distribution (Equation (3)), the first two samples must contain five cells, the next five samples four cells, the next thirteen samples three cells, the next twenty six samples two cells and the last thirty three samples one cell. n. Number of sample, decreasingly ordered according to the Scenario I lag phase.

n	Scenario I	Scenario II	n	I	II
1	2.7	6.5	40	5.9	7.6
2	2.7	6.5	41	6.2	7.9
3	2.7	6.1	42	6.2	7.9
4	2.9	6.3	43	6.2	7.9
5	3.0	6.4	44	6.7	8.4
6	3.1	6.5	45	6.7	8.4
7	3.1	6.5	46	6.7	6.7
8	3.2	5.8	47	6.7	6.7
9	3.2	5.8	48	7.0	7.0
10	3.2	5.8	49	7.2	7.2
11	3.2	5.8	50	7.2	7.2
12	3.2	5.8	51	7.2	7.2
13	3.6	6.1	52	7.2	7.2
14	3.7	6.3	53	7.2	7.2
15	3.7	6.3	54	7.7	7.7
16	3.7	6.3	55	7.7	7.7
17	3.7	6.3	56	7.7	7.7
18	4.2	6.8	57	7.7	7.7
19	4.2	6.8	58	7.7	7.7
20	4.2	6.8	59	7.8	7.8
21	4.2	5.9	60	8.0	8.0
22	4.7	6.4	61	8.2	8.2
23	4.7	6.4	62	8.2	8.2
24	4.7	6.4	63	8.2	8.2
25	4.7	6.4	64	8.2	8.2
26	4.7	6.4	65	8.3	8.3
27	4.7	6.4	66	8.4	8.4
28	4.8	6.5	67	8.5	8.5
29	5.0	6.7	68	8.7	8.7
30	5.0	6.7	69	8.7	8.7
31	5.2	6.9	70	9.2	9.2
32	5.2	6.9	71	9.2	9.2
33	5.2	6.9	72	9.2	9.2
34	5.5	7.2	73	9.2	9.2
35	5.7	7.4	74	9.7	9.7
36	5.7	7.4	75	9.7	9.7
37	5.7	7.4	76	9.7	9.7
38	5.7	7.4	77	9.8	9.8
39	5.7	7.4	78	10.7	10.7
		Average (h)		6.09	7.38
		Sd (h)		2.16	1.13

Copyright © 2012 SciRes.

Table 4. Lag phases of *Enterococcus faecalis* irradiated with 2 kGy in TSB at 18˚C calculated by Equation (1) considering a constant μ of 0.409 and assuming: Scenario I: All samples contain 1 cell; Scenario II: The sample in column "Scenario I" with the shortest lag phase contains the highest inoculum, the second shortest, the second highest, and so on. Number of cells per sample was estimated according to the Poisson distribution considering that the average number of cells per sample was 1.97 (Table 2). Then, according to the Poisson function of distribution (Equation (3)), the first sample must contain six cells, the next three samples five cells, the next nine samples four cells, the next eighteen samples three cells, the next twenty seven samples two cells and the last twenty seven samples one cell. n. Number of sample, decreasingly ordered according to the Scenario I lag phase.

n	Scenario I	Scenario II	n	I	II
1	58.7	62.9	44	68.6	70.3
2	58.7	62.5	45	69.1	70.8
3	58.7	62.5	46	69.2	70.9
4	59.2	63.0	47	69.6	71.3
5	59.4	62.8	48	69.6	71.3
6	59.7	63.1	49	69.6	71.3
7	59.9	63.3	50	69.7	71.4
8	60.2	63.6	51	70.7	72.4
9	60.2	63.6	52	70.9	72.6
10	60.6	64.0	53	71.7	73.4
11	61.2	64.6	54	73.9	75.6
12	61.3	64.7	55	74.2	75.9
13	61.6	65.0	56	74.2	75.9
14	61.9	64.4	57	74.7	76.4
15	62.2	64.8	58	74.9	76.6
16	63.9	66.5	59	75.1	75.1
17	63.9	66.5	60	75.2	75.2
18	63.9	66.5	61	75.6	75.6
19	64.5	67.0	62	75.6	75.6
20	64.6	67.2	63	75.6	75.6
21	64.7	67.3	64	75.6	75.6
22	64.9	67.5	65	75.6	75.6
23	65.4	67.9	66	76.1	76.1
24	65.6	68.1	67	76.1	76.1
25	65.7	68.3	68	76.2	76.2
26	65.7	68.3	69	76.2	76.2
27	65.7	68.3	70	76.6	76.6
28	65.7	68.3	71	76.6	76.6
29	65.7	68.3	72	76.6	76.6
30	65.8	68.3	73	76.6	76.6
31	66.3	68.8	74	76.6	76.6
32	66.4	68.1	75	76.6	76.6
33	67.0	68.7	76	76.7	76.7
34	67.2	68.9	77	76.7	76.7
35	67.6	69.3	78	76.7	76.7
36	67.7	69.4	79	77.5	77.5
37	67.9	69.6	80	77.7	77.7
38	68.0	69.7	81	79.2	79.2
39	68.0	69.7	82	80.7	80.7
40	68.1	69.8	83	84.7	84.7
41	68.2	69.9	84	85.2	85.2
42	68.6	70.3	85	86.9	86.9
43	68.6	70.3			
		Average (h)		69.92	71.53
		Sd (h)		7.04	5.89

more accurate lag phase determinations when applying the Scenario II, which means that, in spite of the above discussed arguments, the average of the recalculated distribution (Scenario II) must be more realistic than that of Scenario I.

The robust permutation test (see Annex 1) is proposed to compare lag phase distributions. When this test was applied to the experimental data of Aguirre *et al.* [9], several significant differences ($\alpha < 0.05$) were found between Scenarios I and II, I and IV and II and IV. From **Figure 3**, it may be deduced that the higher the average number of cells per sample was, the greater the probability that there would be significant differences in distributions between Scenario I and II or between I and IV. This is consistent with the predictions of the simulation generated in this study. The influence of the specific growth rate on the distributions is less clear.

5. Conclusion

To estimate lag phase, the assumption that growth comes from one cell in all samples when a certain percentage of them does not show microbial growth contradicts the Poisson distribution function. Taking into account the percentage of samples showing microbial development, the Poisson function allows ascribing higher inocula to the samples with shorter lag phases. Considering Poisson-based predictions of the number of cells per sample, instead of considering that all samples contain one cell, the accuracy of the average lag phase determinations of micropopulations will be improved. In fact, the more samples there are that contain more than one cell, the greater the improvement will be. This improvement is likely to be statistically significant mainly in cases where the average number of cells per sample is relatively high.

6. Acknowledgements

The authors gratefully acknowledge the support of the Ministerio de Educación y Ciencia (Spain), Program Consolider CARNISENUSA CSD2007-0016 and AGL-2010-16598. We thank Armando Chapin Rodríguez for his assistance in the manuscript writing.

REFERENCES

[1] P. J. Stephens, J. A. Joynson, K. W. Davies, R. Holbrook, H. M. Lappin-Scott and T. J. Humphrey, "The Use of an Automated Growth Analyser to Measure Recovery Times of Single Heat-Injured *Salmonella* Cells," *Journal of Applied Microbiology*, Vol. 83, No. 4, 1997, pp. 445-455. doi:10.1046/j.1365-2672.1997.00255.x

[2] K. Francois, F. Devlieghere, A. R. Standaert, A. H. Geeraerd, J. F. Van Impe and J. Debevere, "Modelling the Individual Cell Lag Phase. Isolating Single Cells: Protocol Development," *Letter of Applied Microbiology*, Vol. 37,

Copyright © 2012 SciRes.

No. 1, 2003, pp. 26-30. doi:10.1046/j.1472-765X.2003.01340.x

[3] A. Metris, S. M. George, M. W. Peck and J. Baranyi, "Distribution of Turbidity Detection Times Produced by Single Cell-Generated Bacterial Populations," *Journal of Microbiological Methods*, Vol. 55, No. 3, 2003, pp. 821-827. doi:10.1016/j.mimet.2003.08.006

[4] A. Elfwing, Y. LeMarc, J. Baranyi and A. Ballagi, "Observing Growth and Division of Large Numbers of Individual Bacteria by Image Analysis," *Applied and Environmental Microbiology*, Vol. 70, No. 2, 2004, pp. 675-678. doi:10.1128/AEM.70.2.675-678.2004

[5] L. Guillier, P. Pardon and J. C. Augustin, "Influence of Stress on Individual Lag Time Distributions of *Listeria monocytogenes*," *Applied and Environmental Microbiology*, Vol. 71, No. 6, 2005, pp. 2940-2948. doi:10.1128/AEM.71.6.2940-2948.2005

[6] M. D'Arrigo, G. D. Garcia de Fernando, R. Velasco de Diego, J. A. Ordonez, S. M. George and C. Pin, "Indirect Measurement of the Lag Time Distribution of Single Cells of *Listeria innocua* in Food," *Applied and Environmental Microbiology*, Vol. 72, No. 4, 2006, pp. 2533-2538. doi:10.1128/AEM.72.4.2533-2538.2006

[7] C. Pin and J. Baranyi, "Kinetics of Single Cells: Observation and Modeling of a Stochastic Process," *Applied and Environmental Microbiology*, Vol. 72, No. 3, 2006, pp. 2163-2169. doi:10.1128/AEM.72.3.2163-2169.2006

[8] G. W. Niven, J. S. Morton, T. Fuks and B. M. Mackey, "Influence of Environmental Stress on Distributions of Times to First Division in *Escherichia coli* Populations, as Determined by Digital-Image Analysis of Individual Cells," *Applied and Environmental Microbiology*, Vol. 74, No. 12, 2008, pp. 3757-3763. doi:10.1128/AEM.02551-07

[9] J. S. Aguirre, M. R. Rodriguez and G. D. Garcia de Fernando, "Effects of Electron Beam Irradiation on the Variability in Survivor Number and Duration of Lag Phase of Four Food-Borne Organisms," *International Journal of Food Microbiology*. Vol. 149, No. 3, 2011, pp. 236-246. doi:10.1016/j.ijfoodmicro.2011.07.003

[10] I. A. Swinnen, K. Bernaerts, E. J. Dens, A. H. Geeraerd and J. F. Van Impe, "Predictive Modelling of the Microbial Lag Phase: A Review," *International Journal of Food Microbiology*. Vol. 94, No. 2, 2004, pp. 137-159. doi:10.1016/j.ijfoodmicro.2004.01.006

[11] J. P. Smelt, G. D. Otten and A. P. Bos, "Modelling the Effect Of Sublethal Injury on the Distribution of the Lag Times of Individual Cells of *Lactobacillus plantarum*," *International Journal of Food Microbiology*, Vol. 73, No. 2-3, 2002, pp. 207-212.

[12] A. Metris, S. M. George and J. Baranyi, "Use of Optical Density Detection Times to Assess the Effect of Acetic Acid on Single-Cell Kinetics," *Applied and Environmental Microbiology*, Vol. 72, No. 10, 2006, pp. 6674-6679. doi:10.1128/AEM.00914-06

[13] BACANOVA, "Final Report on the European Project QLRT-2000-01145: Optimisation of Safe Food Processing Methods Based on Accurate Characterisation of Bacterial Lag Time Using Analysis of Variance Techniquesm," BACANOVA Consortium, European Commission, Brussels, 2005.

[14] J. Baranyi and C. Pin, "Estimating Bacterial Growth Parameters by Means of Detection Times," *Applied and Environmental Microbiology*, Vol. 65, No. 2, 1999, pp. 732-736.

[15] R. C. McKellar and K. Knight, "A Combined Discrete-Continuous Model Describing the Lag Phase of *Listeria monocytogenes*," *International Journal of Food Microbiology*, Vol. 54, No. 3, 2000, pp. 171-180. doi:10.1016/S0168-1605(99)00204-4

[16] T. P. Robinson, O. O. Aboaba, A. Kaloti, M. J. Ocio, J. Baranyi and B. M. Mackey, "The Effect of Inoculum Size on the Lag Phase of *Listeria monocytogenes*," *International Journal of Food Microbiology*, Vol. 70, No. 1-2, 2001, pp. 163-173. doi:10.1016/S0168-1605(01)00541-4

[17] R. C. McKellar and A. Hawke, "Assessment of Distributions for Fitting Lag Times of Individual Cells in Bacterial Populations," *International Journal of Food Microbiology*, Vol. 106, No. 2, 2006, pp. 169-175. doi:10.1016/j.ijfoodmicro.2005.06.018

[18] J. Baranyi, S. M. George and Z. Kutalik, "Parameter Estimation for the Distribution of Single Cell Lag Times," *Journal of Theoretical Biology*, Vol. 259, No. 1, 2009, pp. 24-30. doi:10.1016/j.jtbi.2009.03.023

[19] I. Espejo, F. Fernández, M. A. López, M. Muñoz, A. M. Rodríguez, A. Sánchez and C. Valero, "Estadística Descriptiva y Probabilidad," 3rd Edition, U.d.C., Cadiz, 2006. http://knuth.uca.es/repos/l_edyp/pdf/-febrero06/lib_edyp.apendices.pdf

[20] T. Hesterberg, D. S. Moore, S. Monaghan, A. Clipson, R. Epstein and G. P. McCabe, "Bootstrap Methods and Permutation Tests," In: W. H. McCabe, Ed., *Introduction to the Practice of Statistics* (5th Edition), Freeman & Co., New York, 2005, pp. 14.1-14.70.

[21] S. Holm, "A Simple Sequentially Rejective Multiple Test Procedure," *Scandinavian Journal of Statistics*, Vol. 6, 1979, pp. 65-70.

[22] H. Abdi, "Bonferroni and Sidák Corrections for Multiple Comparisons," In: C. S. Neil Salkind, Ed., *Encyclopedia of Measurement and Statistics*, Dallas, Thousand Oaks, 2007, pp. 103-107.

[23] R Version 2.72, "Language and Environment for Statistical Computing," The R Foundation for Statistical Computing, 2008.

[24] C. Pin and J. Baranyi, "Single-Cell and Population Lag Times as a Function of Cell Age," *Applied and Environmental Microbiology*, Vol. 74, No. 8, 2008, pp. 2534-2536. doi:10.1128/AEM.02402-07

Copyright © 2012 SciRes.

Appendix A

```
##########################################################################
# PROGRAM PERMUTATION_VARIANCE_RATIO_TEST v2.0
##########################################################################
# Authors: María del Carmen Bravo and Juan Aguirre - November 2011
# This R program (http://cran.es.r-project.org/) applies the permutation
# test to the null hypothesis of homogeneity of two variances.
# It applies a multiple test procedure to comparisons of several variance
# pairs. To minimize the probability of rejecting a true hypothesis,
# Bonferroni and Sidak p-value corrections, as well as the less conservative
# Holm-Bonferroni corrections, are applied to the p-values.
# The lag-time distribution of single cells of different microorganisms is
# analysed. For each microorganism, the analysis is carried out using 3, 4, or
# 5 dose or scenario groups. By default, the number of variance pair
# comparisons in these cases are 3, 6 and 10, respectively.
# All tests are bilateral (alternative hypothesis is non-homogeneity).
# Extension to unilateral tests (alternative hypothesis is that one variance
# is larger), and to different numbers of doses or scenarios, is possible.
# Also the number of comparisons may be reduced.
# Permutation test:
# -----------------
# The homogeneity of two variances test is equivalent to the variance
# ratio test, with the null hypothesis stating that the ratio of variances
# is equal to 1.
# For two dose or scenario groups, the variances of the lag-time of both
# groups and their ratio (variance_ratio) are computed.
# For a number of times, n_permutations:
#     - Lag-time observations are randomly distributed into the two dose
#       or scenario groups. Original dose or scenario sizes are maintained.
# - Variances of both groups and their ratio are computed.
# A distribution of the variance ratios is obtained.
# If the variance_ratio value is an anomalous value of this distribution,
# then the hypothesis of homogeneity of variances is rejected.
# A histogram of the variance ratio distribution is generated.
# Bonferroni and Sidak p-value corrections:
# -----------------------------------------
# The output of each homogeneity of variances test is its p-value and
# the corrected Sidak and Bonferroni p-values that take into account the
# number of variance pair comparisons. The Bonferroni correction is the
# test p-value multiplied by the number of comparisons. This value
# must be compared to a multiple significance level alpha.
# Holm-Bonferroni p-value corrections:
# ------------------------------------
# p-values of homogeneity of variance tests are ordered from lower to higher.
# These values are compared to alpha/n_comparisons, alpha/(n_comparisons-1),
# alpha/(n_comparisons-2),..., alpha, respectively. The p-values must be lower # than this second set of values to be significant with a multiple
# significance level alpha.
# References:
# -----------
# Abdi, H. 2007. Bonferroni and Sidák corrections for multiple comparisons.
# In: N.J. Salkind (ed.) Encyclopedia of Measurement and Statistics, Thousand # Oaks, CA: Sage. pp. 103-107
# Hesterberg, T., Moore, D.S., Monaghan, S., Clipson, A., Epstein, R.. McCabe, # G.P., 2005, Bootstrap Methods and Permutation Tests. In: Moore, D.S. &
# McCabe, G.P.: Introduction to the Practice of Statistics, Fifth Ed., W.H.
# Freeman & Co. pp. 14.1-14.70.
# Holm, S., 1979. A simple sequentially rejective multiple test procedure.
# Scandinavian Journal of Statistics, 6, 65-70.
# R version 2.7.2, 2008. R: A Language and Environment for Statistical
# Computing. Reference Index. The R Foundation for Statistical Computing.
# ISBN 3-900051-07-0.
# To run the program in R: Mouse-selection of several sentences + F5

##########################################################################
```

Copyright © 2012 SciRes.

```
#                                   Function definitions
##############################################################################

# The function definitions part of the program is executed once.
# permutation_test function definition.
# g1, g2: ordinal numbers identifying dose or scenario groups.

permutation_test <- function(g1,g2) {

      assign("k",k+1,envir=.GlobalEnv)

      lag2g<-LAG[Key==keys[g1] | Key==keys[g2]]

      ng1<-n_groups[g1]
      ng2<-n_groups[g2]

      n1<-ng1
      n11<-ng1+1
      n2<-ng1+ng2

variance_ratio<-var_groups[g1]/var_groups[g2]
      s_variance_ratio<-numeric(n2)
      s_variance_ratio<-variance_ratio

      for (i in 1:(n_permutations-1)) {
      s_lag2g<-sample(lag2g)

s_variance_ratio<-c(s_variance_ratio,var(s_lag2g[1:n1])/var(s_lag2g[n11:n2]))
      }

      position<-length(s_variance_ratio[s_variance_ratio<=variance_ratio])

      # p_inf: left tail estimated probability for variance_ratio value
      p_inf<-position/n_permutations
      p_sup<-1-p_inf

      p_value<-print_results(g1,g2,variance_ratio,p_inf,p_sup)

# If the user does not want to see the histograms, add the
# symbol as the
# first character in the following two sentences.
win.graph()
      histogram(s_variance_ratio,g1,g2)

      p_value
      }

# print_results function definition

print_results <- function (g1,g2,coc,p_inf,p_sup) {

      cat ('---------- Permutation test ----------','\n')
      cat('Groups:', keys[g1],keys[g2],'\n')
      cat('Variance ratio:',coc,'\n')

      # p_value: p-value for the two-tailed test
      if (1/n_permutations < p_inf & p_inf <= 0.5) p_value <- p_inf*2
      if (0 < p_sup & p_sup < 0.5) p_value <- p_sup*2
      # p_bon, p_sidak: corrected Bonferronni and Sidak p_values for
# n_comparisons tests

if ((1/n_permutations<p_inf & p_inf<=0.5) | (0<p_sup & p_sup<0.5)) {
p_bon <- p_value*n_comparisons
            p_sidak <- 1-(1-p_value)^n_comparisons }

      if (p_inf==1/n_permutations | p_sup==0) {
          p_value <- 1/n_permutations
          p_bon <- 1/n_permutations
          p_sidak <- 1/n_permutations
          cat('Two-tailed p_value lower than',1/n_permutations,'\n')
          cat('Bonferroni corrected p_value lower than',1/n_permutations)
          cat('\n','Sidak corrected p_value lower than',1/n_permutations,'\n')
     }

      if (p_value > 1/n_permutations) cat('Two-tailed p_value:',p_value,'\n')

      if (1/n_permutations < p_value & p_value < alpha) {
            cat('Bonferroni corrected p_value:',p_bon,'\n')
            cat('Sidak corrected p_value:',p_sidak,'\n') }

      if (min(p_bon,p_sidak)>alpha    )
       { cat('** Non significant differences between variances **','\n')}
      if (min(p_bon,p_sidak)<alpha)
          {cat ('** Variances of doses groups are different **','\n')}

      cat ('-----------------------------------------------------------','\n')
      p_value
```

Copyright © 2012 SciRes.

```
}

# histogram function definition

histogram <-function (x,g1,g2) {

graph_name<-paste(file_name,sheet_name,'Groups:',keys[g1],'and',keys[g2])
        hist(x, main = graph_name)    }

##############################################################################
#                                End of function definitions
##############################################################################

##############################################################################
#                                Main program
##############################################################################
# Program parameters:
# -------------------
#    alpha: Multiple significance level for n_comparisons of permutation
#                test of the homogeneity of two variances. Usual value: 0.05
#    n_permutations: Number of permutations. Usual value: 10000
#    file_name: Name of the Excel file with the input data
# sheet_name: Name of the Excel sheet with the input data
# Name of columns: Key LAG
# Key refers to treatment names (doses or scenarios)
#    LAG refers to LAG-time
# groups_number: Number of dose or scenario groups.
# Possible values: 3, 4, 5
# n_comparisons: Number of pairwise comparisons. Values: 3, 6, 10
# for respective group_number values of 3, 4, 5
# Its value is used for p_value corrections
# keys: group_number dimension vector with dose or scenario group names
#    digits: Number of digits for output of numerical values
# output_file_name: Name of output file when user wants program output
# diverted to a file instead of to the R terminal

# This main program must be executed once for all variance comparisons of
# lag-time distribution for each microorganism

# IMPORTANT NOTE: The user must assign values to these program parameters
# ---------------
alpha <- 0.05
n_permutations <- 10000
# file_name and sheet_name are only necessary for the output, to identify
# results
file_name <- 'file-name.xls'
sheet_name <- 'sheet-name'
groups_number <- 3

# Give the dose or scenario names in the case of groups_number equal
# to 3, 4 or 5. Names are case-sensitive.
if (groups_number == 3) keys <- c('A','B','C')
if (groups_number == 4) keys <- c('A','B','C','D')
if (groups_number == 5) keys <- c('A','B','C','D','E')

# Output:
# -------
# Program output can be diverted to the R terminal (default) or to a file.
# If output is written to a file, specify the name of the file
# (for example: output_file_name="c:/output_files/Lag_HAM.txt")
# and run the following two sentences without the # symbol.

# output_file_name="complete-file-name"
# sink(file = output_file_name, append = TRUE, type = "output",split = FALSE)

# the append = TRUE option means that the output will be appended
# to the file; otherwise, it will overwrite the contents of the file.
# If during an R session you want to divert output to the R terminal,
# run the next sentence without the # symbol.

# sink(file=NULL)

###################                Data entry ###################
#    1. Open the Excel file file_name, sheet sheet_name
#    2. Select the columns Key and LAG (names are case-sensitive)
#    3. Click menu option Edit --> Copy (or CTRL + C )
#    4. Execute the next two R sentences
# If the decimal point is written as a comma in your input data, read
# the note at the end of the program
```

Copyright © 2012 SciRes.

```
data<- read.table("clipboard",header=T)
attach(data)

# digits controls the number of digits to print # when printing numeric values.
# options(digits=5)

################## Running the analysis ##############
# User has two options:
# 1) Select all sentences through to the end of the program and execute them
# 2) Execute the following command blocks as appropriate

if (groups_number == 3) n_comparisons <- 3
if (groups_number == 4) n_comparisons <- 6
if (groups_number == 5) n_comparisons <- 10

cat('Input data: File ', file_name ,', sheet ', sheet_name,"\n")
cat('Number of doses or scenario groups:',groups_number,"\n")
cat('Number of comparisons:',n_comparisons,"\n")
cat('Multiple significance level:', alpha,"\n")

g1<-Key[Key==keys[1]]
ng1<-length(g1)
g2<-Key[Key==keys[2]]
ng2<-length(g2)
g3<-Key[Key==keys[3]]
ng3<-length(g3)

n_groups<-c(ng1,ng2,ng3)

var1<-var(LAG[Key==keys[1]])
var2<-var(LAG[Key==keys[2]])
var3<-var(LAG[Key==keys[3]])

var_groups<-c(var1,var2,var3)

# Running this block is required for groups_number > 3, though it can be run
# for any value of groups_number.
if (groups_number>3) {
      g4<-Key[Key==keys[4]]
      ng4<-length(g4)
      n_groups<-c(n_groups,ng4)
      var4<-var(LAG[Key==keys[4]])
      var_groups<-c(var_groups,var4)
      }

# Running this block is required for groups_number = 5, though it can be run
# for any value of groups_number.
if (groups_number==5) {
      g5<-Key[Key==keys[5]]
      ng5<-length(g5)
      n_groups<-c(n_groups,ng5)
      var5<-var(LAG[Key==keys[5]])
      var_groups<-c(var_groups,var5)
      }

# Continue running the program
std_groups <- sqrt(var_groups)
cat('Names of doses or scenario groups:', keys,"\n")
cat('Variances:', var_groups,"\n")
cat('Standard deviations:', std_groups,"\n")

# k: Initialization of the number of comparisons # performed. User should not
# change this value
# k increments its value by one each time #permutation_test function is run
k<-0
c1<-numeric(n_comparisons)
c2<-numeric(n_comparisons)
p_values<-numeric(n_comparisons)

# Execution of the permutation_test function as permutetion_test(g1,g2),
# where g1, g2:  ordinal numbers identifying dose or scenario groups.
# n_comparisons permutation_test definition functions must be run.

p<-permutation_test(1,2);      c1[k]<-1;      c2[k]<-2;
p_values[k]<-p

p<-permutation_test(1,3);      c1[k]<-1;      c2[k]<-3;
p_values[k]<-p

p<-permutation_test(2,3);      c1[k]<-2;      c2[k]<-3;
p_values[k]<-p

# Running this block is required for groups_number > 3, though it can be run
# for any value of groups_number.
if (groups_number>3) {
      p<-permutation_test(1,4);    c1[k]<-1;    c2[k]<-4;
p_values[k]<-p
p<-permutation_test(2,4);      c1[k]<-2;      c2[k]<-4;
p_values[k]<-p
      p<-permutation_test(3,4);    c1[k]<-3;    c2[k]<-4;
p_values[k]<-p
      }

# Running this block is required for groups_number = 5,
```

Copyright © 2012 SciRes.

```
though it can be run
# for any value of groups_number.
if (groups_number==5) {
      p<-permutation_test(1,5);     c1[k]<-1;     c2[k]<-5;
p_values[k]<-p
      p<-permutation_test(2,5);     c1[k]<-2;     c2[k]<-5;
p_values[k]<-p
      p<-permutation_test(3,5);     c1[k]<-3;     c2[k]<-5;
p_values[k]<-p
      p<-permutation_test(4,5);     c1[k]<-4;     c2[k]<-5;
p_values[k]<-p
      }

# Continue running the program:
# -to control possible errors in k value and
# -to perform tests that apply Holm-Bonferroni p-value
corrections

if (k>n_comparisons) {
      cat ('******* ERROR *******',"\n")
      cat ('The number of permutation tests run is greater
than', n_comparisons)
      cat ("\n",'Start the running of the program from
sentence k<-0') }

# ord_p: test p-values in order of increasing value
ord_p <- order(p_values)

# alpha_i: Holm-Bonferroni significance level for the ith
test for a multiple
# significance level alpha
alpha_i<-c(alpha/n_comparisons:1)

C<-matrix(c(c1[ord_p],c2[ord_p],p_values[ord_p],alpha
_i),nrow=n_comparisons)

aux<-numeric(n_comparisons)
for (i in 1:n_comparisons) {
      aux[i]<-0
      if (C[i,3]< C[i,4]) aux[i]<-1 }

CC<-matrix(c(C,aux),nrow=n_comparisons)

# L: number of significance tests
L <-sum(CC[,5])

# CCC: matrix of significance tests.
#     columns 1 and 2: Identification numbers of
# dose or scenario groups
#     column 3: p-value for the bilateral test
#     column 4: alpha_i value to be compared with
p-value
CCC<-CC[,1:4][CC[,5]==1]
dim(CCC)<-c(L,4)

cat ("\n")
cat ('** Application of Holm-Bonferroni p-value correc-
tions **',"\n")
cat ('-------------------------------------------------------',"\n")
if (L==0) {
cat ('** Non significant differences between any pair of
variances **',"\n")}
if (L > 0) {
cat ('** Pairs of variances that are different **',"\n")
      cat  ('Doses          Doses          p-value
H-B correction',"\n")
      cat  ('-----          -----          -------
--------------',"\n")
      for (i in 1:L) {
cat(format(keys[CCC[i,1]],width=12),format(keys[CCC[
i,2]],width=12),
      format(CCC[i,3],width=8,digits=5),
format(CCC[i,4],width=9,digits=5),"\n")}
cat ('Significant because p-value < Holm-Bonferroni
correction',"\n")
      }
cat ('-------------------------------------------------------',"\n")
cat
('*********************************************
***********',"\n")

# End of the program

# Notes:
# ------
# When running attach(data), the message
#       The following object(s) are masked from data #
(position 3):
#   Key LAG
# should be ignored
# For data in which the decimal is marked with a comma
instead of a period,
# use the following sentence without the # symbol
#data<-read.table("clipboard",dec=",",header=T#)
# Data can be read from CSV files exported from Excel
# For data in which the decimal is marked with a # pe-
riod, use the following
# sentence without the # symbol
#data<-read.csv2("complete-file-name",header=T#)
# For data in which the decimal is marked with a comma,
use the following
# sentence without the # symbol
#data<- read.csv("complete-file-name",header=T)
# Example of complete-file-name:
# c:/data/VT2kgy.csv
# To avoid the following error
# Error en win.graph() : too many devices open
```

Copyright © 2012 SciRes.

```
# close graphing devices when many are open
# To avoid the following error
# Error en var(LAG[Key == keys[1]]) : 'x' is empty
# ensure that the names in the column Key of the # Excel file are
# the same as the names given in the sentence
# if (groups_number ...) keys <- c('A',...)
```

Copyright © 2012 SciRes.