Probabilistic modeling in the inference of fungi viability

Jorge Fernandez-de-Cossio[1*], Ariadna Mendoza-Cuevas[2], Amanda Paz Torres[2,3]

[1] Bioinformatic Department, Center for Genetic Engineering and Biotechnology, Cuba.

[2] Archaeometry Laboratory, Oficina del Historiador de la Habana. Cuba

[3] Biology Faculty, Havana University, Cuba.

* jorge.cossio@cigb.edu.cu to who correspondence should be addressed.

Abstract

Preservation of large volumes of documentary materials requires massive treatments to control fungal biodeterioration. Multi-inoculum in limited-dilution experiments is an economical format to provide informative data for fungal viability. Data gathering can be simpler and more precise, when the model for the assessment of viability depends on the number of colonized wells instead of the number of colonies, since the later vary in shape and definition. Probabilistic modeling for this kind of data led us to a truncated form of the Beta-exponential function, a variant from a family of generalized distributions, recently derived. We discuss performance from actual experimental data, and evaluate antifungal effect of biocide treatments in terms of viability.

Introduction

The quality and performance of statistical analysis depends on the chosen probability models [1–4]. Over-reliance on standard error, or the uncritical adherence to a limited set of well-known probability distributions, does not provide sufficient flexibility to account for all common datasets [5–8]. A renewed interest in developing generalized probability distributions are conferring more flexibility in fitting real-life data, and has become an area of rapid advances in the last years [6,9,10]. The extensions are derived by mathematically compounding two or more distributions, incorporating one or more parameters to the baseline model, transforming variables, or by elaborating the distribution of a function of their parameters (typically sum, ratio, or convolutions) [1,6,9,11–25].

The Beta generalized family involve nearly all of the well-known models [20,23,26–32]. This generalized class of distributions can be defined by

$$F(x) = I_{G(x)}(a, b), \qquad x > 0, a > 0, b > 0 \tag{1}$$

where, $I_z(a, b) = \mathrm{B}(z; a, b)/\mathrm{B}(a, b)$ is the regularized incomplete beta function and G is a cumulative distribution function (cdf). Recently, by taking G as the cdf of an exponential distribution, Nadarajah & Kotz [29] obtained the beta-exponential density function $f(x) = \mathrm{d}F(x)/\mathrm{d}x$, expressed in the form

$$f(x) = \frac{\lambda}{B(a,b)} \exp(-b\lambda x) \{1 - \exp(-\lambda x)\}^{a-1}, \qquad (2)$$

with the associated hazard rate function $\lambda(x)$, and provided a comprehensive treatment of their properties.

Mathematical generalizations like this family of functions, are often derived without a particular real-world question or data in mind. The resulting distributions compete *a posteriori* for the possibility of fitting into some application scenario.[1,9,11–14,17–20,24,25,33–35]. Despite being theoretically well supported, the pace of the practical introduction of these extensions are far behind the rate at which they are mathematically derived. In consequence, its use in life-sciences applications remains scarce and underestimated.

On another modeling avenue, insights about the nature of the data guide from the beginning the selection or construction of the distribution. The parameters enter by design, with a world-domain interpretation, and so do the logical structures involved in the construction of the probability models. But often, the model starts to look intractable, and too early approximations prevent later arriving at well-studied extended distributions. The issue also arises with data of apparent simple nature.

Species of fungi are well known for their cellulolytic and endoglucanase activities, and have been described as spoilers of paper and archival materials, impairing their quality and usefulness [36–38]. Evaluation of biocidal effect is of relevance for the conservation of such materials. We undertake here the later modeling path to assess the viability of fungi (i.e., their ability to form progeny), with data from multiple inoculums in limited diluted experiments. The model is not selected here a priori from a tool-set of well-studied mathematical functions with built-in tuning parameters for shaping and data fitting. The viability model acquires the final structure after successive application of the rules of probability theory. Step by step, guided by the nature of the data, led us to a truncated form of the Beta-exponential function (2). The truncated variant, here derived, need re-normalization and impose border considerations not explicitly accounted in the comprehensive mathematical treatment of the original derivation [29]. We delve into the logical structures of the model to account for the evidence and inferences expected from the experimental setup. Finally, we discuss performance in the estimation of viability of fungi spores, from actual data gathered by the experimental format from which the model was derived, and elaborate on the assessment of the impact and magnitude of the treatment effect.

Materials and Methods

Data

Data were taken from a study to evaluate the effect of irradiation in the viability of fungi spores, executed in the frame of an IAEA technical project for the development of treatments for paper conservation against biodeterioration [39–41].

Definition of viability

Criteria for determining microbial viability remain ambiguous [42], but there is agreement in define viability by the capacity of a microbial unit to proliferate [43,44]. By viability we mean here the expected frequency of an average cell to form progeny or to multiply. This definition of viability, for example, can serve as a quantitative measure to assess the effect of a biocidal treatments.

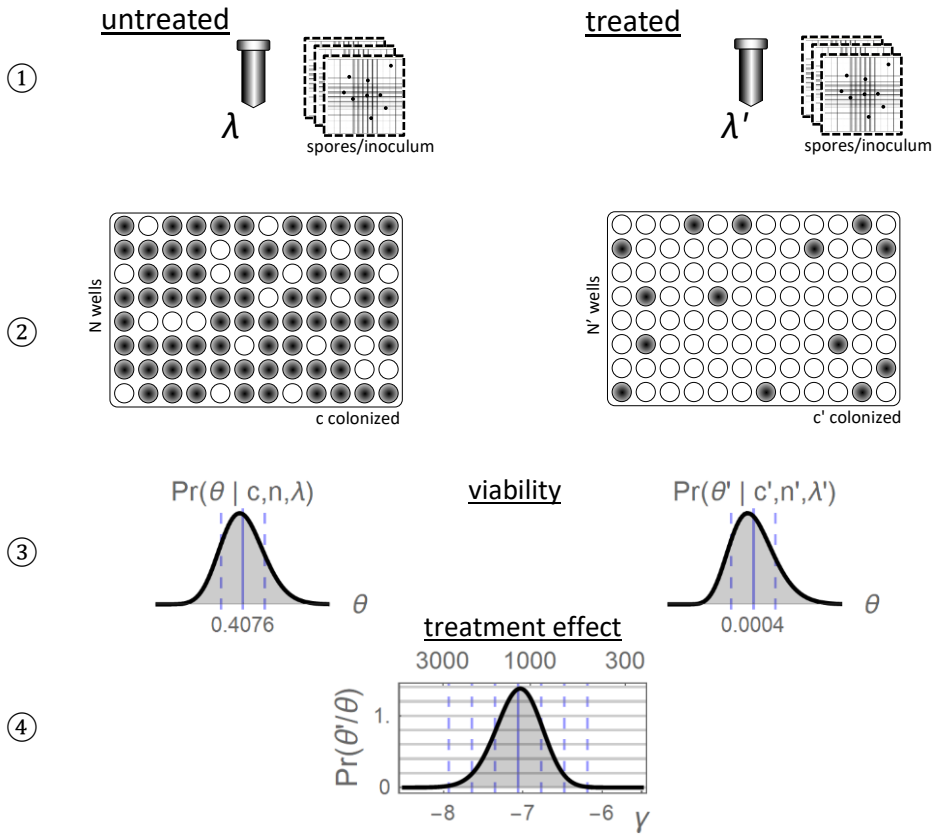


Figure 1: Experiment format delivering data for the model of viability. Left and right workflows are identical except that the sample of spores of the right is previously treated. 1) A suspension is prepared with an appropriate concentration of spores. The mean number of spores per inoculation λ is accurately estimated by counting spores in replicates samples. 2) Multiple wells are inoculated with replicate samples drawn from the solution (ex. 96 wells plate). 3) Inferences of viability is preformed from the collected data (c, N, and λ). 4) The effect-size of the treatment is estimated from the untreated and treated data.

Multiple-inoculums in limited diluted experiment

An experimental format providing relevant data for the estimation of viability θ, can be described as follows (Figure 1). *N* wells are inoculated with an average number λ of spores. This

average is accurately estimated from repeated samples drawn from a spore suspension, using a cell counting device. Samples of $10\mu L$ inoculated in a well, with a given average of spores, are drawn from suspensions of spores at corresponding concentrations. Multiple inoculums experiments were performed in 96 and 24 wells plates.

A well is not colonized if none of the inoculated (founder) spores produce progeny. The higher the viability, the more plausible the colonization of each well, and the more wells are expected to be colonized. The number c of colonized wells is collected for viability estimation.

To assess the effect of treatments, the viability of spores from untreated and treated samples are compared. The comparative experiment is sketched in Figure 1. Left and right workflows are identical except for that the sample of the right is previously treated. It is assumed that the experiments are conducted so that any differences in colonization found between untreated and treated inoculums can only be attributed to genuine effects of the treatment and not confounded with other factors. Viability, as here defined, does not consider possible cases of interaction.

Software

The computations of the models and plots have been performed with scripts written in Mathematica [45].

Results and discussion

Probabilistic modeling

Inoculum colonization

In an inoculation of n spores with viability θ, the chance that r of them germinates follows a binomial distribution

$$\Pr(r|n,\theta) = \binom{n}{r}\theta^r(1-\theta)^{n-r} \qquad (3)$$

The inoculated well will be colonized if one or more of the spores germinates. The probability of colonization $\equiv r > 0$ is then

$$\Pr(\text{colonization}|n,\theta) = 1 - \Pr(r = 0|n,\theta) = 1 - (1-\theta)^n \qquad (4)$$

From practical conveniences (time, resources, and labors), counting the number n of spores at each inoculation is eluded in our experiments. Therefore, equation (4) cannot be directly used. All that the model demand to know, as accurately as possible, is the average number λ of spores inoculated per well, on the assumptions that the samples are draw from an inexhaustive suspension with a constant concentration of spore λ. On this fundamental hypothesis, the uncertainty on the actual number n of spores inoculated is ruled by a Poisson rule [46].

$$\Pr(n|\lambda) = e^{-\lambda}\frac{\lambda^n}{n!} \qquad (5)$$

The probability of colonization of a well, given λ, can be obtained by averaging (4) over the possible actual number n of spores distributed by (5), that is

$$\Pr(\text{colonization}|\lambda,\theta)=\sum_{n}\Pr(\text{colonization}|n,\theta)\Pr(n|\lambda)=1-e^{-\lambda\theta} \qquad (6)$$

Multiple inoculums

Let $w_1,\dots,w_c$ be the colonized wells from a total of N wells, all inoculated with an average number λ of spores with viability θ. The sampling probability of $\{w_i\}$ depends only on the number c of wells that turned out colonized, and is ruled by the binomial distribution

$$\Pr(c|\theta,\lambda,N)=\left(1-e^{-\lambda\theta}\right)^{c}\left(e^{-\lambda\theta}\right)^{N-c} \qquad (7)$$

When samples are drawn from various suspension with different concentrations corresponding to the average inoculations $\lambda_1,\dots\lambda_m$, the sampling becomes

$$\begin{aligned}\Pr(c_1,\dots,c_m|\theta,\lambda_1,\dots\lambda_m,N_1,\dots,N_m)&=\prod_{i=1}^{m}\Pr(c_i|\theta,\lambda_i,N_i)\\&=\prod_{i=1}^{m}\left(1-e^{-\lambda_i\theta}\right)^{c_i}\left(e^{-\lambda_i\theta}\right)^{N_i-c_i}\end{aligned} \qquad (8)$$

When $\lambda_i=\lambda$, for $i=1,\dots,m$, equation (8) reduce to (7), where $c=\sum_i c_i$ and $N=\sum_i N_i$.

These expressions (7) and (8) as a function of the unknown θ (the likelihoods), carry all the relevant information contained in the data for the inference of viability.

Inferences of viability

By the Bayes theorem, the probability of θ given these data is

$$\begin{aligned}&\Pr(\theta|c_1,\dots,c_m,\lambda_1,\dots\lambda_m,N_1,\dots,N_m)\\&\qquad=\frac{\Pr(\theta)\,\Pr(c_1,\dots,c_m|\theta,\lambda_1,\dots\lambda_m,N_1,\dots,N_m)}{\Pr(c|\lambda_1,\dots\lambda_m,N_1,\dots,N_m)}\end{aligned} \qquad (9)$$

where $\Pr(\theta)=\Pr(\theta|\lambda_1,\dots\lambda_m,N_1,\dots,N_m)$, since by the experiment design, the number of spores per inoculums λ_i, and the number of inoculated wells N_i does not elicit information about the viability of the spores, without knowing the number of colonized wells.

The denominator of (9) can be computed from

$$\Pr(c_1,\dots,c_m|\lambda_1,\dots\lambda_m,N_1,\dots,N_m)=\int_0^1\prod_{i=1}^{m}\Pr(c_i|\theta,\lambda_i,N_i)\,\mathrm{d}\theta \qquad (10)$$

In the particular case when $\lambda_i=\lambda$, for $i=1,\dots,m$

$$\Pr(c_1,\dots,c_m|\lambda,N_1,\dots,N_m)=\lambda^{-1}\mathrm{B}_{1-e^{-\lambda}}(c+1,N-c) \qquad (11)$$

where $c = \sum_i c_i$ and $N = \sum_i N_i$. Therefore, the exact expression for the probability distribution of viability (9) becomes

$$\Pr(\theta|c_1, \dots, c_m, \lambda, N_1, \dots, N_m) = \frac{\lambda}{\mathrm{B}_{1-e^{-\lambda}}(c+1, N-c)} \left(1-e^{-\lambda\theta}\right)^c \left(e^{-\lambda\theta}\right)^{N-c}, \qquad (12)$$

which is the Beta-exponential density (2), a member of the extended Beta family [29], except for the provision that symbol θ replace x, with the connotation that $0 < \theta < 1$, while $0 < x$ in (2). Consequently, the normalization constant of this truncated form (12) is not the Euler beta but the incomplete beta $B_z(a, b)$ function, with $z = 1 - e^{-\lambda}$. The more general form is however (9), provided (8) and (10), which has more parameters than what can be found in the so far known beta extended families.

Moment generating function

The moment generating function defined by $\mathrm{Mgf}(t) = \langle \exp(\theta t) \rangle$, for the case $\lambda_i = \lambda$ is

$$\mathrm{Mgf}(t) = \frac{\lambda}{\mathrm{B}_{1-e^{-\lambda}}(c+1, N-c)} \int_0^1 \left(1-e^{-\lambda\theta}\right)^c \left(e^{-\lambda\theta}\right)^{(N-c-t/\lambda)} \mathrm{d}\theta$$

Which is an incomplete beta function, thus

$$\mathrm{Mgf}(t) = \frac{\mathrm{B}_{1-e^{-\lambda}}(c+1, N-c-t/\lambda)}{\mathrm{B}_{1-e^{-\lambda}}(c+1, N-c)}, \qquad t < \lambda$$

We find no analytical expression for the first and second moments of this function. Though there are codes to numerically compute the required derivatives [47], we choose to calculate the mean value of any function $g(\theta)$ of the viability by the numerical integration:

$$\langle g(\theta) \rangle = \frac{\lambda}{\mathrm{B}_{1-e^{-\lambda}}(c+1, N-c)} \int_0^1 g(\theta) \left(1-e^{-\lambda\theta}\right)^c \left(e^{-\lambda\theta}\right)^{N-c} \mathrm{d}\theta \qquad (13)$$

Of particular interest here are the first two moments of both, $g(\theta) = \theta$ and $g(\theta) = \log\theta$, and their variance.

Treated and untreated

Parameters and data related to the treatment experiment will be denoted by an accent, i.e. $\theta', \lambda', c', N'$. The setup experimental parameters are denoted $D = \{\lambda, N\}$ and $D' = \{\lambda', N'\}$. After the observed numbers of colonized wells c and c' of untreated and treated inoculums, the joint distribution of viabilities θ and θ' can be obtained from Bayes theorem.

$$\Pr(\theta, \theta'|c, c', D, D') = \Pr(\theta, \theta'|D, D') \frac{\Pr(c, c'|\theta, \theta', D, D')}{\Pr(c, c'|D, D')} \qquad (14)$$

Conditional on wildtype and treated data, the observations can be regarded as independent. Given the viabilities θ and θ', the colonized counts c and c' can be regarded as identically distributed and independent.

$$\Pr(c, c'|\theta, \theta', D, D') = \Pr(c|\theta, D)\Pr(c'|\theta', D') \quad (15)$$

There is no reason to a priori consider logical dependency between θ and θ', then

$$\Pr(\theta, \theta'|D, D') = \Pr(\theta|D)\Pr(\theta'|D') \quad (16)$$

In consequence, the denominator also factors into $\Pr(c|D)\Pr(c'|D')$, and by (11) yields

$$\Pr(c, c'|D, D') = \frac{1}{\lambda}\mathrm{B}_{1-e^{-\lambda}}(c+1, N-c) \times \frac{1}{\lambda'}\mathrm{B}_{1-e^{-\lambda'}}(c'+1, N'-c') \quad (17)$$

Substituting all these expressions in the joint distribution of viabilities (14), and using (11) and (12) yields

$$\begin{aligned}\Pr(\theta, \theta'|c, c', D, D') &= \Pr(\theta|c, D)\Pr(\theta'|c', D') \\ &= \frac{\left(1-e^{-\lambda\theta}\right)^{c}\left(e^{-\lambda\theta}\right)^{N-c}}{\lambda^{-1}\mathrm{B}_{1-e^{-\lambda}}(c+1, N-c)} \times \frac{\left(1-e^{-\lambda'\theta'}\right)^{c'}\left(e^{-\lambda'\theta'}\right)^{N'-c'}}{\lambda'^{-1}\mathrm{B}_{1-e^{-\lambda'}}(c'+1, N'-c')}\end{aligned} \quad (18)$$

The modelling can further complicate, for example, when anticipating in advance that the treatment deteriorate viability. Being that the case, the prior join probability of untreated and treated viabilities cannot factorize like in (16). The less compromised way to introduce this prior knowledge is by keeping in $\Pr(\theta'|\theta, D, D')$ the structure of $\Pr(\theta'|D')$ in the range $\theta' \leq \theta$, and cero otherwise, that is

$$\Pr(\theta'|\theta, D, D') = \frac{\left(1-e^{-\lambda'\theta'}\right)^{c'}\left(e^{-\lambda'\theta'}\right)^{N'-c'}}{\lambda'^{-1}\mathrm{B}_{1-e^{-\lambda'\theta}}(c'+1, N'-c')}[\theta' \leq \theta] \quad (19)$$

Notice the dependency on the viability θ of the untreated sample, introduced after normalization in the range $\theta' \leq \theta$ by the incomplete beta function in the denominator. This later case was not further elaborated in the present work.

Treatment effect size

The viability of the treated sample can be expressed in term of the wildtype sample as $\theta' = e^{\gamma}\theta$. Thus, the ratio of viability in a log scale $\gamma = \log(\theta'/\theta)$, is a symmetric measure of the effect of the treatment.

The expected effect size $\langle\gamma\rangle$ can be computed from (13), since $\langle\gamma\rangle = \langle\log\theta\rangle - \langle\log\theta'\rangle$, and the variance is equal to the sum of the variance of $\log\theta$ and $\log\theta'$, which can be also computed from (13). The probability density distribution of the effect γ, given the data, is obtained by the double integral

$$\Pr(\gamma|c,c',D,D') = \int_0^1 \int_0^1 \Pr(\gamma|\theta,\theta') \Pr(\theta,\theta'|c,c',D,D') \mathrm{d}\theta' \, \mathrm{d}\theta$$

By the definition of the effect size γ, the a priori density is $\Pr(\gamma|\theta,\theta') = \delta(\gamma - \log(\theta'/\theta))$, and from (16) the posterior density becomes

$$\int_0^1 \Pr(\theta|c,D) \int_{-\infty}^0 \delta(\gamma - \log\theta' + \log\theta) e^{\log\theta'} \Pr\left(e^{\log\theta'}\middle|c',D'\right) \mathrm{d}(\log\theta') \mathrm{d}\theta,$$

which reduce to the single integral

$$\Pr(\gamma|c,c',D,D') = e^{\gamma} \int_0^1 \theta \Pr(\theta|c,D) \Pr(e^{\gamma}\theta|c',D') \, [0 \le \theta e^{\gamma} \le 1] \, \mathrm{d}\theta \qquad (20)$$

The cumulative distribution of γ is computed by

$$\Pr(\gamma < \gamma_o|c,c',D,D') = \int_0^{\gamma_o} \Pr(\gamma|c,c',D,D') \, \mathrm{d}\gamma \qquad (21)$$

Is there any effect?

We proceed to assess if the inferences of the effect size are significant. The hypotheses $\gamma < 0$, $\gamma = 0$ and $0 < \gamma$, corresponds to deterioration (biocide), no effect, or improvement (invigoration) of viability. The Bayes factor test is adopted to compare the null hypotheses $\gamma = 0$ versus the alternative hypothesis $\gamma \neq 0$. We are not biased for either in advance, so we assign them equal prior probabilities. Any other choice lessens or strengthen the tolerance of decision on either hypothesis.

$$\mathrm{BF}(\gamma = 0 \text{ vs. } \gamma \neq 0) = \frac{\Pr(c,c'|\gamma = 0,D,D')}{\Pr(c,c'|\gamma \neq 0,D,D')} \qquad (22)$$

The numerator of (22) was already stablished by (10), since $\gamma = 0$ imply $\theta = \theta'$. For the particular case $\lambda = \lambda'$, it becomes $\lambda^{-1}\mathrm{B}_{1-e^{-\lambda}}\left(c + c' + 1, N + N' - (c + c')\right)$ by (11). The denominator of (22) is the product of incomplete beta functions, already stablished by (17).

In its general case, the Bayes factor is expressed by

$$\mathrm{BF}(\gamma = 0 \text{ vs. } \gamma \neq 0|\lambda \neq \lambda') = \lambda \frac{\int_0^1 \Pr(c|\theta,D) \Pr(c'|\theta,D') \, \mathrm{d}\theta}{\mathrm{B}_{1-e^{-\lambda}}(c+1,N-c)\mathrm{B}_{1-e^{-\lambda'}}(c'+1,N'-c')} \qquad (23)$$

The particular case $\lambda = \lambda'$, it can be expressed by the closer or more treatable form

$$\mathrm{BF}(\gamma = 0 \text{ vs.} \gamma \neq 0|\lambda) = \lambda \frac{\mathrm{B}_{1-e^{-\lambda}}(c + c' + 1, N + N' - c - c')}{\mathrm{B}_{1-e^{-\lambda}}(c + 1, N - c)\mathrm{B}_{1-e^{-\lambda}}(c' + 1, N' - c')} \quad (24)$$

Minimum (maximum) effect

The previous epigraph is suitable, for example, to ask if some environmental parameter is affecting in any degree a material. We can be interested in the hypothesis that a treatment produces an effect that is stronger than a given magnitude, with the addition of the uncertainty. For example, the viability of spores is reduced by 90% or more with probability $\Pr(\gamma < \log \Delta \,|\, \cdots)$, computed by (21), where $\Delta = \theta'/\theta = 0.1$. The hypothesis against lower viability can be expressed in terms of Bayes factor by

$$\mathrm{BF}(\mathrm{LD}_{90}) = \frac{\Pr(\gamma < \log \Delta \,|\, \cdots)}{1 - \Pr(\gamma < \log \Delta \,|\, \cdots)} \quad (25)$$

Theoretical anticipations

Accuracy and uncertainty

The evidence provided by the number c of colonized wells, out of $N = 11$, is profiled in Figure 2. The probability distribution of the viability θ, i.e. $\Pr(\theta|c, \lambda, N)$, is plotted for various parameters and data, in four hypothetical setups arranged by columns with distinct average number of spores per inoculation ($\lambda = 0.8$, 1, 2 and 20).

The case $\lambda = 1$ might be easier to grasp, since if exactly one spore is inoculated in a well, the chance of colonization (4) will be θ, the parameter of interest, and the inference of θ would be ruled by a Beta distribution. However, λ is not the exact number but the average of spores inoculated per well, and by this experimental format it is quite difficult to inoculate exactly the same number of spores per well. Having inoculated one spore on average imply that some wells cannot colonize, not because the stochastic nature of the spore to yield progeny, but because the well contains no spore at all. The Beta distribution, having no provision for the chance of no-spore well, underestimate θ, amounting all colonization failures to the viability of the spore. Further, since the uncertainty on the actual number of spores inoculated is not accounted in the Beta distribution model, the accurracy of their predictions are expected to be overoptimistic. The provision for the uncertainty on the actual number of spores inoculated is introduced by equation (6) into the model, leading by the rules of probability theory to the wider Beta-exponential distribution (12). These predicted issues are clearly exhibited in the column corresponding to $\lambda = 1$ of Figure 2, comparing the posterior distribution of θ of both distributions. The Beta distribution (dotted red profile) lag behind the Beta-exponential distribution (blue profile), and is also narrower. There is also the chance that a well be inoculated with more than one spore, increasing the chance of colonization. But the probability of no spore is 1.42 times the probability of more than one spore according to (5), supporting again the observed in the plots.

Weigh of the evidence

Whether or not we are aware of the viability of the spores, the number of colonized wells will depend on the actual viability of the inoculated spores. Let then suppose that the unknown but actual viabilities of the spores in the four experiments are $\theta = 0.5, 0.4, 0.2, 0.02$, respectively.

These values have been chosen so that $\lambda \times \theta = 0.4$ in each experiment, to simplify comparison (see columns heading in Figure 2). Consequently, the probability $1 - e^{-\lambda\theta}$ of colonizing a well is also the same, 0.33, and the sampling probability of colonizing c of $N = 11$ wells is ruled by the same binomial distribution (7) in the four setup (green bargraph on the left of Figure 2). Therefore, the expected number of colonized wells is $N\left(1 - e^{-\lambda\theta}\right) = 3.63$, and the mode of (7) is $c = 3$ (Figure 2).

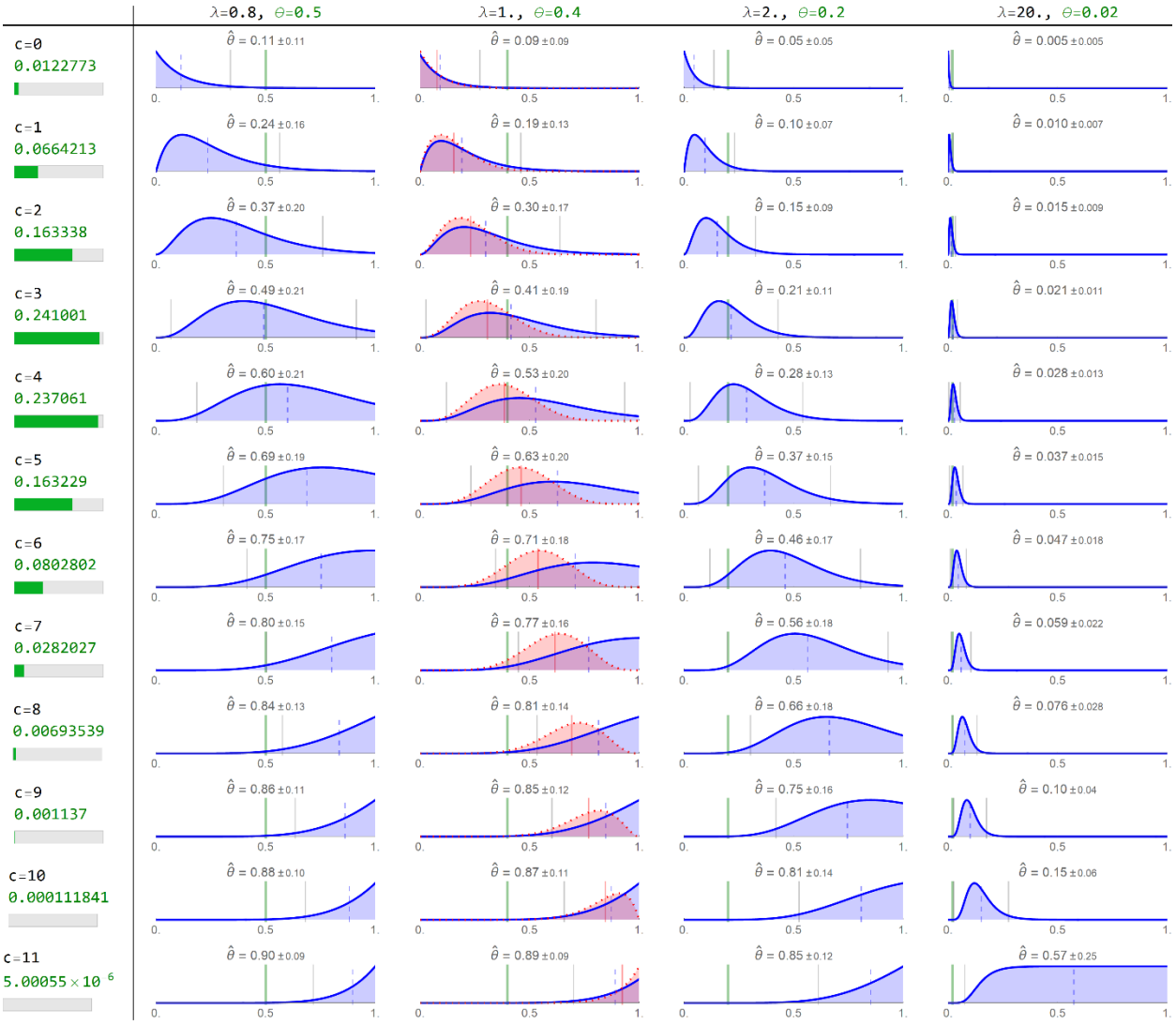


Figure 2: The extended beta exponential distribution of the viability θ, equation (12), is profiled in blue for various parameters and data, in an experiment with $N = 11$ wells. Dotted red profiles corresponds to the Beta distribution of θ, with parameters $c + 1$ and $N - c + 1$. Rows corresponds to the number c of colonized wells. The left column shows in green the sampling probability of gathering c colonized wells, were the unknown actual viability θ is indicated in green at the top of each column. At each plot, the mean ± standard deviation of θ is displayed at the top, according to the extended truncated beta exponential.

Vertical bars (gray) indicate 2 standard deviation from the mean (dashed line) of the distribution, and the putative viability at the top of the column is indicated with a vertical green bar in the plot. Red vertical line locates the mean of the beta distribution. Columns corresponds to distinct average number λ of spores per inoculation (indicated in the column head).

Thus, in experiments of $N = 11$ wells inoculated with $\lambda = 1$ spores per well on average, the most frequent outcome expected is $c = 3$ colonized wells, when the (unknown) viability of the spores is $\theta = 0.4$. Looking at the row panel (Figure 2) corresponding to the outcome $c = 3$, the extended beta exponential distribution of θ, conditioning on this outcome has mean $\hat{\theta} = 0.41 \pm 0.19$ (blue dashed line), pretty close to the actual one, while the mean 0.3 (red line) of the beta distribution is markedly separated. So arise with the respetive mean $\hat{\theta}$ of the other four θ in the row $c = 3$ of Figure 2.

Although all outcomes from $c = 0$ to $c = 11$ are stochastically plausible, the extreme one estimate mean $\hat{\theta}$ farther than two standandard deviation from the actual θ. However, these extreme cases are practically impossible, in the four setup. For exampe, the outcome $c = 8$ are pretty unplausible, occurring with probability 0.007. The outcomes $c = 2,3,4,5$ accounts for more than 80% chance, and all yields extended beta exponential distribution with mean of θ within one standard deviation of the actual viability.

Viability allowance

The Beta distribution allot for high extreme viability, i.e. $\theta \approx 1$, only when all the wells are colonized, i.e. $c = N$ (last row of Figure 2). The conditioning that each well contain exactly one spore precludes incomplete outcomes, i.e., $c < N$, unless $\theta < 1$. But because $\lambda = 1$ is the average spore per well, a single well have 37% chance to receive no spore, and in 11 wells there is more than 62% chance that four or more wells contain no spore. Those “empty” wells will not colonize even if the spores are 100% viable (i.e. $\theta = 1$). Therefore, a realistic model should not rule out the $\theta \approx 1$ possibility.

The allowance for $\theta \approx 1$ is rightly accounted by the Beta-exponential distribution in the truncated form, as can be seen by positive right tails in rows from $c = 5$ to $c = 11$ (Figure 2). This allowance is more acentuated when $\lambda = 0.8$, since more wells are expected to receive no spore. Even for $\lambda = 2$, there is allowance for that. The obvios case is $c = N$, where no distribution can discard $\theta = 1$ (last row of Figure 2).

Shaping the distribution

The parameter N was given by experimental design, λ were measured, and the data c is a discrete outcome, obtainable by counting colonized wells in the multiple inoculums experiment. The commonly used probability models can't accommodate the variety of shapes delivered by this extended function (Figure 2). Viability θ is our inferential target.

Data analysis

Inference of viability

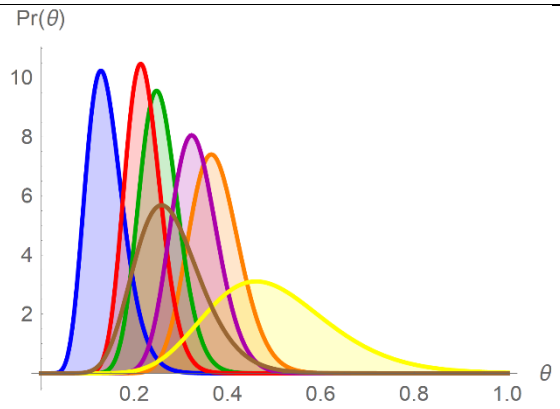
Data collected from the multiple inoculums’ experiments designed in these studies, are used to evaluate the statistical model for the inference of viability and treatment effect against

biodeterioration agents. Samples of $10\mu L$ with an average λ of spores, were drawn from spores' suspensions of Aspergillus niger with the corresponding concentrations, and inoculated in 96 and 24 wells plates. The outcomes and inferences are shown in Figure 3 for both, wildtype and treated samples.

An average of 0.95, 1.9 and 3.8 wildtype spores were inoculated in 96 and 24 wells plates. An average of 420, 470 and 510 treated spores were inoculated in 96 well plates, as indicated in the left panel of Figure 3. Few wildtype spores ($1 < \lambda < 4$) were sufficient to colonize some wells out of 24, while about 500 are required for treated spores. This numbers alone accounts in favor of treatment efficacy, but there are aspects of the model's performance on real data, which merit a bit further discussion.

A) wildtype

	λ	N	c	Θ	σ
	0.95	96	11	0.139779	0.04038
	1.9	96	48	0.373052	0.054415
	1.9	96	36	0.254505	0.042249
	1.9	96	32	0.220262	0.038623
	1.9	96	44	0.3305	0.050081
	1.9	24	14	0.497593	0.131835
	3.8	24	15	0.278448	0.072882



B) treated (2 kGy)

	λ'	N'	c'	Θ'	σ'
	420.	96	28	0.000851	0.000159
	470.	96	34	0.000959	0.000163
	510.	92	42	0.001226	0.00019
	510.	96	29	0.00073	0.000134
	510.	96	30	0.00076	0.000137
	510.	96	32	0.000821	0.000144

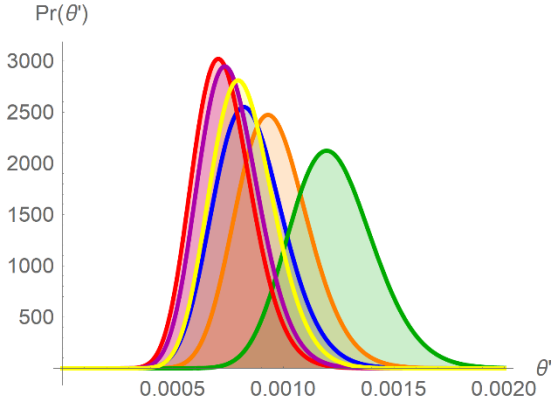


Figure 3: Parameters, outcomes, and inferences of multiple inoculums' experiments performed for wildtype (A) and treated (B) samples of Aspergillus niger cultures. The viability θ and standard deviation σ, as estimated from the model. The truncated beta exponential distribution of θ is profiled for each outcome at the right panel.

The probability distributions of viability (12), profiled in the right panel of Figure 3, tightly overlap in both, wildtype and treated samples. Therefore, the inference are consistent, even for diffent number of inoculated spores. Each plot, however, show a seemingly outlier profile (yellow and green respectively). In these two particular cases, the inoculations exhausted the suspension, in the case of the treated sample, it fall short of 96 wells. This is a departure from the experimental assumptions. The modeling rely on the Poisson rule early in the derivations (5

), but it apply there on the assumption of inexhaustible suspensions. This non-compliance may affect all of the following, in particular, it may not lead to the truncated beta exponential distribution (12). Even though, this “outliers” did not performed so badly. The modeling based on fitting the data to a family of distributions are precluded from this type of analysis.

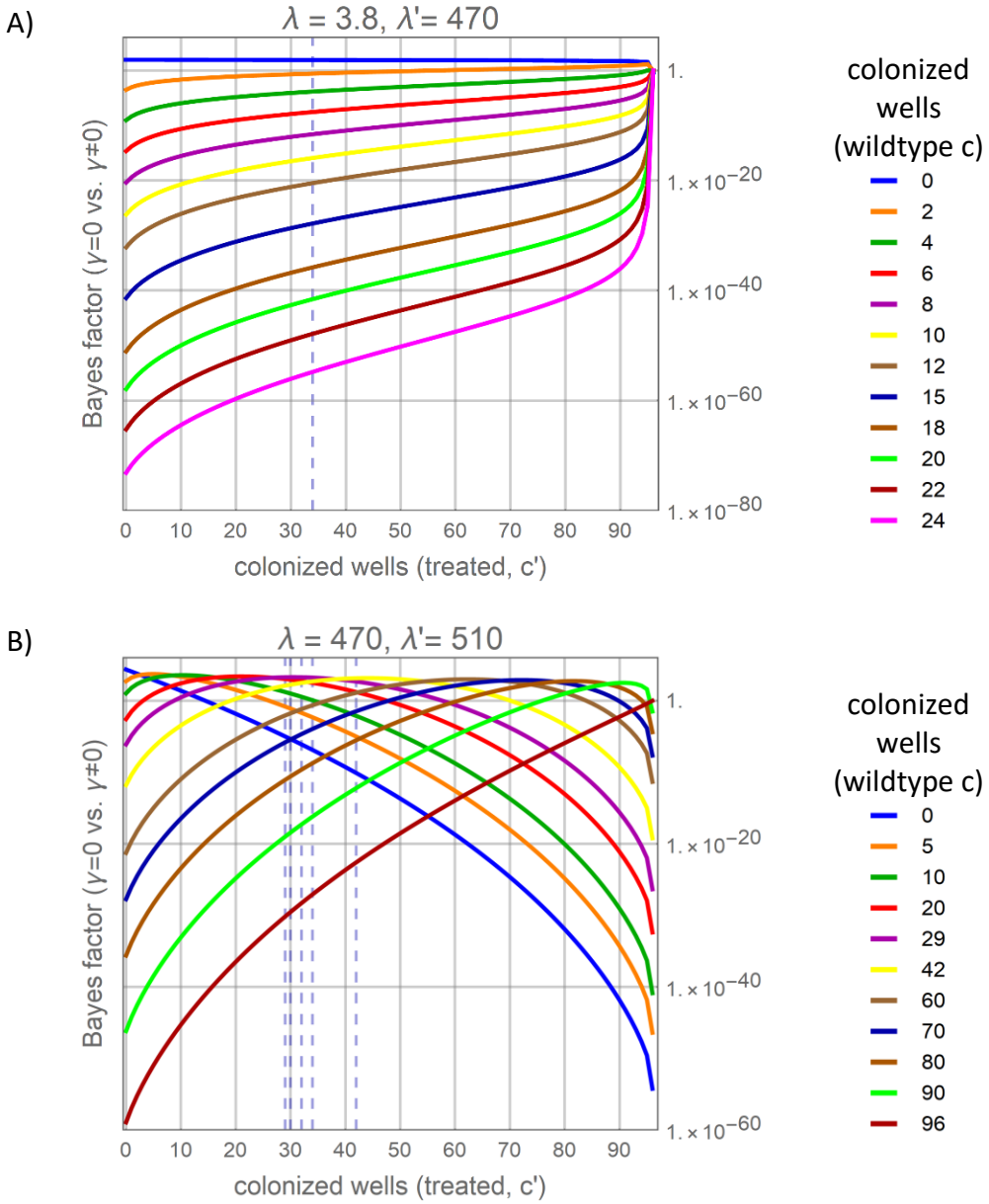


Figure 4: Tests for the hypothesis of no-effective vs effective treatment ($\gamma = 0$ vs. $\gamma \neq 0$). Each curve corresponds to a number c of colonized wells in the wildtype plate (color legend at the right), and the Bayes factor is plotted against the number c' of colonized wells in the treated plate. A) Testing the effectiveness of the treatment by comparing untreated $\lambda = 3.8$ vs. treated $\lambda' = 470$ samples in the experiments with $N = 24$ and $N' = 96$ wells, respectively. A vertical dashed line indicates the actual treatment outcome $c' = 34$. B) Testing the effectiveness of the treatment by comparing treated $\lambda = 470$ vs. treated $\lambda' = 510$ samples in the experiments with 96 inoculated wells. Vertical dashed lines indicate treatment outcomes $c' = 29,30,32,34,42$.

Treatment comparison

The effectiveness of the treatment is established by the model in terms of how much it reduces the viability of the wildtype spore population. Some anticipations can be realized at first glance from the outcomes of our experiments, without requiring a model. But in general, the ratios c/N and c'/N' does not rightly indicate these directions and magnitudes. It depends as well on the average number of spores per inoculation in wildtype λ and treated λ' sample. The treated samples inoculations, Figure 3B, was performed with more than 100 times the number of spores inoculated from the wildtype sample Figure 3A. The biocidal effect of the treatment can be overwhelmingly established from these plots, since the support of θ and θ' do not overlap. We can see, how the model performs in terms of our definition of viability, comparing the hypothesis of no-effect versus effect, in terms of the Bayes factor test (23) and (24). Figure 4 profiles the Bayes Factor, testing the hypothesis of no effect against effect of the treatment. Each curve corresponds to a fixed number c of wildtype-colonized wells. With c fixed, the curve is plotted along the Bayes factor computed for each number c' of treated-colonized wells.

In Figure 4A, $\lambda = 3.8$ wildtype spores were on average inoculated in $N = 24$ wells, while $\lambda' = 470$ treated spores were on average inoculated in $N' = 96$ wells. The Bayes factor for the outcome $c = 15$ and $c' = 34$ colonized wells, respectively, equal 1.6×10^{-28} (see the intersection of the dark blue curve with the vertical dashed line in Figure 4A).

We now compare the treated sample experiments between them, as if the spores in each experiment were subjected to different treatments with similar competing effects. We know they are no different since all these inoculations are actually from spores subjected to the same 2kGy treatment. Hence, we expect in every case a verdict in favor of the hypothesis of no effect. Figure 4B, compare treated $\lambda = 470$ vs. treated $\lambda' = 510$ samples in the experiments with 96 inoculated wells. Vertical lines correspond to the "treated" sample outcomes $c' = 29{,}30{,}32{,}34{,}42$. For the sake of clearness, only the curves $c = 29$ and $c = 42$ (magenta and yellow) are plotted, since the other curves of the actual outcomes are closely bounded between them. From $c' = 20$ to $c' = 50$, all the Bayes factor traced between the curves $c = 29$ and $c = 42$ are well above one, indicating no evidence in support of treatment effect.

Effect direction

Once stablished the hypothesis that the treatment is effective, we need to know in which direction, biocide or invigorating, and quantitatively by how much. Bayes factor alone does not indicate the direction neither the magnitude of the effect, but something can be anticipated from the monotony of the curves in Figure 4. Bayes factor increase when wildtype and treated viability approach each other, i.e. when the treatment effect is smaller. Hence, since c is constant along a curve, and Bayes factor increase with the number c' of colonized wells, that means that by increasing treated viability, it gets closer to wildtype viability, thus the treatment is biocide. On the opposite, if Bayes factor decrease with the number c' of colonized wells, that means that the treatment is invigorating. The curve in Figure 4A are all time increasing, which means that the treatment in this experiment is biocide. For comparable hypothesis, Figure 4B,

the monotony changes with the outcomes, which means that for similar treatments, the same experimental setup can resolve between biocide or invigorating.

Effect magnitude γ

We chose a dataset of wildtype and treated samples with different number of inoculated wells $N = 24$ vs. $N' = 96$, and different average of spores per wells $\lambda = 3.8$ and $\lambda' = 470$, respectively. Figure 5 plot the probability distribution of the effect (20) for the actual outcome $c = 15$ and $c' = 34$ of colonized wells, and for outcome combination of non and all colonized wells. Each plot can be identified by the number of colonized wells pair c, c' as indicated in the top row and left column, respectively.

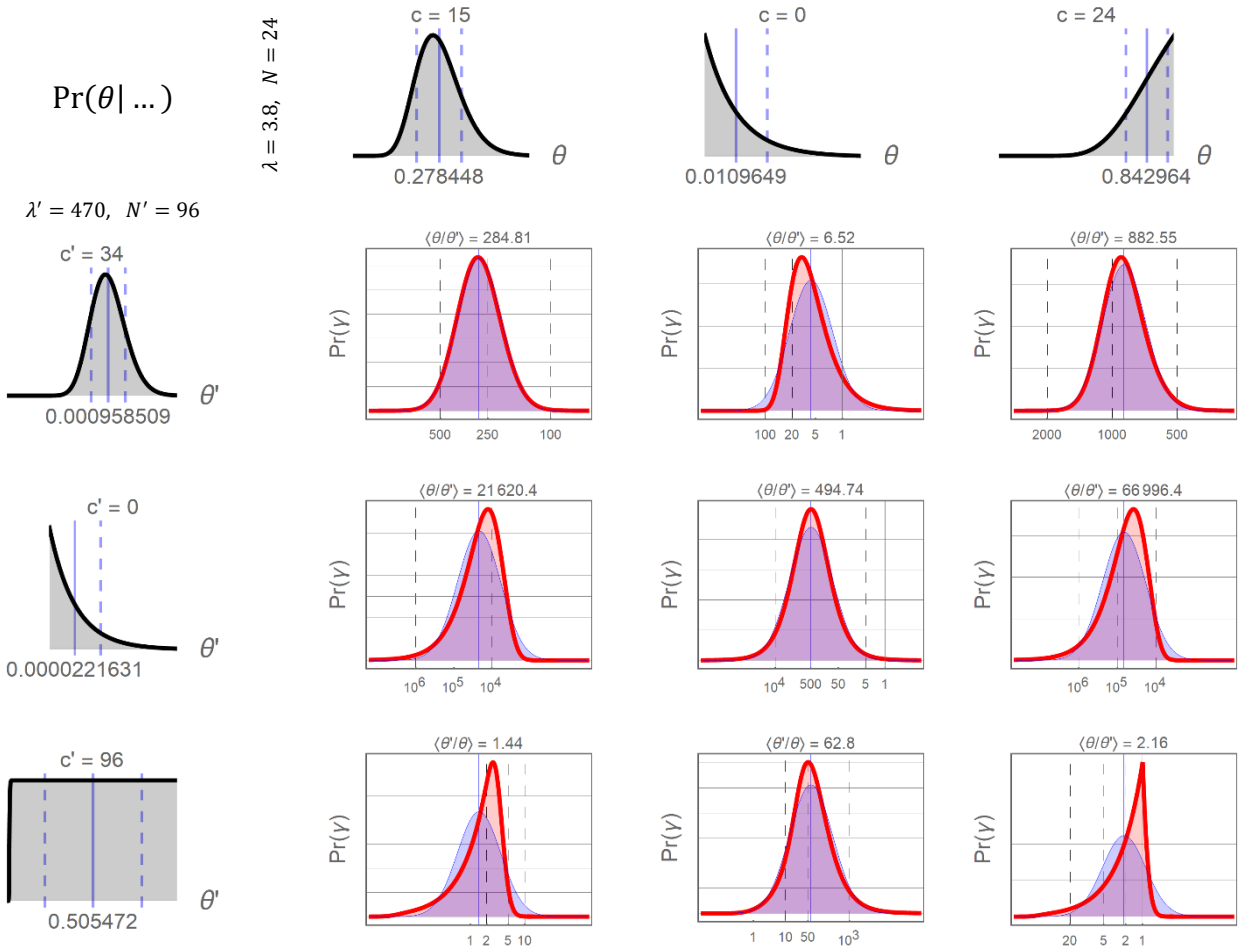


Figure 5: Inference of the magnitude of treatment efficacy. Red curve represents the probability distribution (20) of the efficacy $\gamma = ln\, \theta'/\theta$, obtained from the number c' of wells colonized by the treated samples with respect to the untreated ones c. Blue curve are the normal distributions with the same mean and standard deviation as the red one. In the untreated samples, a total of $N = 24$ wells were inoculated with 3.8 spores on average. In treated samples, $N' = 96$ wells were inoculated with 4.7×10^2 spores on average. The respective probability distributions of viabilities are at the top and left margins coordinates.

Figure 5 explore other hypothetical outcomes c and c', for $N = 24$, $\lambda = 3.8$ and $N' = 96$, $\lambda = 470$, including the extreme full colonization and the cero colonization. Suppose the outcome is $c = N$ and $c' = 0$ colonized wells from each sample respectively. A dramatic treatment effect can be anticipated, since full colonization is obtained with only 3.8 wildtype spores per well,

while 470 treated spores per well were unable to colonize a single well in 96. Indeed, the probability distribution of the effect θ/θ' is located between four and six order of magnitude, concentrated about $66\,996$.

Perhaps, a provocative outcome is the no colonization $c = c' = 0$ for both wildtype and treated samples. Our first thinking did not anticipate relevant information favoring an effect from this outcome. However, the burden of the probability distribution of the effect θ/θ' is located about 500, between 5 and 10^4, indicating a more probable biocide effect.

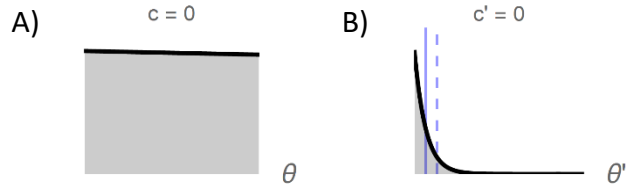


Figure 6: The range $0 < \theta < 0.00033$ of the probability distribution of viability of the plots in Figure 5, given the outcomes $c = c' = 0$.

We find an explanation by looking at the marginal probability distributions of the viability of wildtype and treated spores for these outcomes (Figure 6). The outcome $c' = 0$ of 96 wells with 470 treated spores on average, drastically concentrate the plausible range of θ' to the interval $0 < \theta < 0.00033$, with a negligible probability of $\sim 10^{-8}$ outside this range, practically ruling out this possibility (Figure 6B). However, the outcome $c = 0$ of 24 wells of the wildtype spores does not convey such limited range for the viability (Figure 6A), allowing lot more possible combinations for the ratio θ/θ' favoring the biocide effect. Such kind of explanations were not immediately provided by our unaided intuition, but the model surface it for us. In cases that such explanation did not emerge after an exhaustive search, the primordial assumptions should be back revised, trying to identify cogent prior information not considered in the model.

Table 1: Confidence of lethal dose 2kGy at various percentage 100Δ, where $\Delta^{-1} = \theta/\theta'$, from the data $c = 15$, $N = 24$, $\lambda = 3.8$ and $c' = 34$, $N' = 96$, $\lambda = 470$.

2kGy	Δ^{-1}	$\Pr(\theta/\theta' > \Delta^{-1})$	$\mathrm{Bf}(\theta/\theta' > \Delta^{-1}\ vs.\ \theta/\theta' < \Delta^{-1})$
$D_{\%99}$	100	0.999	1132.81
$D_{\%99.5}$	200	0.869	6.614
$D_{\%99.67}$	300	0.441	0.788
$D_{\%99.75}$	400	0.139	0.162
$D_{\%99.8}$	500	0.034	0.035

Lethal dose

We are now interested in the hypothesis that a treatment produces an effect that is stronger than a given magnitude, say a dose that guarantees a reduction of the viability by 90%. We can take it for the lethal dose LD_{90} [48], by quantifying the uncertainty with a probability, usually neglected. Inference of lethal dose from the wildtype data $c = 15$, $N = 24$, $\lambda = 3.8$ and treated data $c' = 34$, $N' = 96$, $\lambda = 470$, was computed with (21) and shown in Table 1, at various confidences.

Conclusions

The truncated probabilistic model for the assessment of viability from colonized wells counts, here derived, is resolved at the spore level, and is numerically treatable. The resources required for this experimental format are economic and affordable to any labs. Data analysis also leverage information from extreme-boundary cases that are missed with some probability distribution approximations. The performance demonstrated, are suitable for the development of a methodology for antifungal treatments on massive documentary material of cultural heritage relevance.

Funding Information

This research received no external funding

Acknowledgments

Author thanks the support of the International Atomic Energy Agency in the frame of the project RLA1019: Strengthening Capabilities for the Utilization of Nuclear and Radiation Technology to Characterize, Conserve and Preserve the Cultural Heritage (ARCAL CLXVII).

Data availability statement

Data is contained within the article or supplementary material

Conflict of Interest and other Ethics Statements

The authors declare no conflicts of interest.

References

[1] L. Anzagra, S. Sarpong, and S. Nasiru, “Chen-G class of distributions,” Cogent Mathematics & Statistics **7**(1), 1721401 (2020).

[2] E. Check Hayden, “Weak statistical standards implicated in scientific irreproducibility,” Nature, 1 (2013).

[3] L. Hanin, “Why statistical inference from clinical trials is likely to generate false and irreproducible results,” BMC Med Res Methodol **17**(1), 127 (2017).

[4] L. Kennedy-Shaffer, “Before $p < 0.05$ to Beyond $p < 0.05$: Using History to Contextualize p-Values and Significance Testing,” The American Statistician **73**(sup1), 82–90 (2019).

[5] S.K. Khosa, A.Z. Afify, Z. Ahmad, M. Zichuan, S. Hussain, and A. Iftikhar, “A New Extended-F Family: Properties and Applications to Lifetime Data,” Journal of Mathematics **2020**, 1–9 (2020).

[6] S. Kotz, and D. Vicari, “Survey of developments in the theory of continuous skewed distributions,” METRON - International Journal of Statistics **LXIII**(2), 225–261 (2005).

[7] H.J. Motulsky, “Common Misconceptions about Data Analysis and Statistics,” J Pharmacol Exp Ther **351**(1), 200–205 (2014).

[8] M. Shrahili, I. Elbatal, I. Muhammad, and M. Muhammad, “Properties and applications of beta Erlang-truncated exponential distribution,” J. Math. Computer Sci. **22**(01), 16–37 (2020).

[9] C. Lee, F. Famoye, and A.Y. Alzaatreh, “Methods for generating families of univariate continuous distributions in the recent decades: Methods for generating families of univariate continuous distributions,” WIREs Comp Stat **5**(3), 219–238 (2013).
[10] R.O. Momanyi, and J.A.M. Ottieno, “Generating Distribution Functions Based on Burr Differential Equation,” JAMCS, 55–64 (2020).
[11] Z. Ahmad, “The Zubair-G Family of Distributions: Properties and Applications,” Ann. Data. Sci. **7**(2), 195–208 (2020).
[12] M.A. Aldahlan, F. Jamal, C. Chesneau, I. Elbatal, and M. Elgarhy, “Exponentiated power generalized Weibull power series family of distributions: Properties, estimation and applications,” PLoS ONE **15**(3), e0230004 (2020).
[13] A. Alzaatreh, C. Lee, and F. Famoye, “A new method for generating families of continuous distributions,” METRON **71**(1), 63–79 (2013).
[14] F. Jamal, C. Chesneau, D.L. Bouali, and M. Ul Hassan, “Beyond the Sin-G family: The transformed Sin-G family,” PLoS ONE **16**(5), e0250790 (2021).
[15] S. Nadarajah, “The gamma beta ratio distribution,” Braz. J. Probab. Stat. **26**(2), (2012).
[16] S. Nadarajah, and S. Kotz, “On the Linear Combination of Exponential and Gamma Random Variables,” Entropy **7**(2), 161–171 (2005).
[17] V.B.V. Nagarjuna, R.V. Vardhan, and C. Chesneau, “Kumaraswamy Generalized Power Lomax Distributionand Its Applications,” Stats **4**(1), 28–45 (2021).
[18] S. Nasiru, P.N. Mwita, and O. Ngesa, “Exponentiated Generalized Power Series Family of Distributions,” Ann. Data. Sci. **6**(3), 463–489 (2019).
[19] E.M.M. Ortega, F. Prataviera, and G.M. Cordeiro, “Four generalized Weibull distributions: similar properties and applications,” CeN **42**, e10 (2020).
[20] A. Saboor, M.N. Khan, G.M. Cordeiro, I. Elbatal, and R.R. Pescim, “The beta exponentiated Nadarajah-Haghighi distribution: theory, regression model and application,” Mathematica Slovaca **69**(4), 939–952 (2019).
[21] K.I. Santoro, H.J. Gómez, I. Barranco-Chamorro, and H.W. Gómez, “Extended Half-Power Exponential Distribution with Applications to COVID-19 Data,” Mathematics, 16 (2022).
[22] V.L.D. Tomazella, S.R. Jesus, A.B. Gazon, F. Louzada, S. Nadarajah, D.C. Nascimento, F.A. Rodrigues, and P.L. Ramos, “Bayesian Reference Analysis for the Generalized Normal Linear Regression Model,” Symmetry **13**(5), 856 (2021).
[23] M. Wu, and W. Gui, “Estimation and Prediction for Nadarajah-Haghighi Distribution under Progressive Type-II Censoring,” Symmetry **13**(6), 999 (2021).
[24] S.M. Zaidi, M.M.A. Sobhi, M. El-Morshedy, and A.Z. Afify, “A new generalized family of distributions: Properties and applications,” AIMS Mathematics **6**(1), 456–476 (2021).
[25] M. Zichuan, S. Hussain, A. Iftikhar, M. Ilyas, Z. Ahmad, D.M. Khan, and S. Manzoor, “A New Extended- X Family of Distributions: Properties and Applications,” Computational and Mathematical Methods in Medicine **2020**, 1–13 (2020).
[26] G.M. Cordeiro, and A.J. Lemonte, “The beta Laplace distribution,” Statistics & Probability Letters **81**(8), 973–982 (2011).
[27] N. Eugene, C. Lee, and F. Famoye, “Beta-normal distribution and its applications,” Communications in Statistics - Theory and Methods **31**(4), 497–512 (2002).
[28] M.C. Jones, “Families of distributions arising from distributions of order statistics,” Test **13**(1), 1–43 (2004).

[29] S. Nadarajah, and S. Kotz, "The beta exponential distribution," Reliability Engineering & System Safety **91**(6), 689–697 (2006).
[30] V. Nekoukhou, H. Bidram, and R. Roozegar, "The Beta-Weibull distribution on the lattice of integers," CeN **39**(1), 40 (2016).
[31] K.P. Singh, C.M.-S. Lee, and E.O. George, "On Generalized Log-Logistic Model for Censored Survival Data," Biom. J. **30**(7), 843–850 (1988).
[32] N. Singla, K. Jain, and S. Kumar Sharma, "The Beta Generalized Weibull distribution: Properties and applications," Reliability Engineering & System Safety **102**, 5–15 (2012).
[33] Z.A. Al-saiary, R.A. Bakoban, and A.A. Al-zahrani, "Characterizations of the Beta Kumaraswamy Exponential Distribution," Mathematics **8**(1), 23 (2019).
[34] M. Ghitany, E. Gómez-Déniz, and S. Nadarajah, "A New Generalization of the Pareto Distribution and Its Application to Insurance Data," JRFM **11**(1), 10 (2018).
[35] M. Korkmaz, E. Altun, H. Yousof, A. Afify, and S. Nadarajah, "The Burr X Pareto Distribution: Properties, Applications and VaR Estimation," JRFM **11**(1), 1 (2017).
[36] M.K.L. Das, J.S. Prasad, and S.K. Ahmad, "Endoglucanase production by paper-degrading mycoflora," Letters in Applied Microbiology **25**(5), 313–315 (1997).
[37] G. Piñar, M.C. Sclocchi, F. Pinzari, P. Colaizzi, A. Graf, M.L. Sebastiani, and K. Sterflinger, "The Microbiome of Leonardo da Vinci's Drawings: A Bio-Archive of Their History," Front. Microbiol. **11**, 593401 (2020).
[38] T. Sandle, R. Vijayakumar, M. Saleh Al Aboody, and S. Saravanakumar, "*In vitro* fungicidal activity of biocides against pharmaceutical environmental fungal isolates," J Appl Microbiol **117**(5), 1267–1273 (2014).
[39] A. Mendoza-Cuevas, A. Paz, C. Nela, R. Martinez, and G. Lemus, in *XVII Workshop on Nuclear Physics* (Habana, Cuba, 2019).
[40] A. Paz, Metodología para la determinación de la efectividad de tratamientos fungicidas con irradiación en material documental, Thesis Dissertation, Havana University, 2021.
[41] Technical Cooperation Project RLA1019, "Strengthening Capabilities for the Utilization of Nuclear and Radiation Technology to Characterize, Conserve and Preserve the Cultural Heritage (ARCAL CLXVII)," (2020).
[42] S.S. Kumar, and A.R. Ghosh, "Assessment of bacterial viability: a comprehensive review on recent advances and challenges," Microbiology **165**(6), 593–610 (2019).
[43] O. Braissant, M. Astasov-Frauenhoffer, T. Waltimo, and G. Bonkat, "A Review of Methods to Determine Viability, Vitality, and Metabolic Rates in Microbiology," Front. Microbiol. **11**, 547458 (2020).
[44] H. Davey, and S. Guyot, "Estimation of Microbial Viability Using Flow Cytometry," Current Protocols in Cytometry **93**(1), (2020).
[45] I. Wolfram Research, "Mathematica, Version 13.0.0," (2021).
[46] M.J. Faddy, and D.M. Smith, "Extended Poisson Process Modelling of Dilution Series Data," Journal of the Royal Statistical Society. Series C (Applied Statistics) **57**(4), 461–471 (2008).
[47] A. Zwaenepoel, "Derivatives of the incomplete beta function beta(a, b, x)," (2021).
[48] M. Cortesão, A. de Haas, R. Unterbusch, A. Fujimori, T. Schütze, V. Meyer, and R. Moeller, "Aspergillus niger Spores Are Highly Resistant to Space Radiation," Front. Microbiol. **11**, 560 (2020).