



Evaluating variability and uncertainty separately in microbial quantitative risk assessment using two R packages

Régis Pouillot^a, Marie Laure Delignette-Muller^{b,c,*}

^a 7403 Wyndale lane, Chevy Chase, MD 20815, United States

^b Université de Lyon, F-69000, Lyon; Université Lyon 1, CNRS, UMR5558, Laboratoire de Biométrie et Biologie Evolutive, F-69622, Villeurbanne, France

^c Université de Lyon, F-69000, Lyon; VetAgro Sup Campus Vétérinaire de Lyon, F-69280 Marcy l'Etoile, France

ARTICLE INFO

Article history:

Received 31 March 2010

Received in revised form 9 June 2010

Accepted 10 July 2010

Keywords:

Freeware tools

Microbial quantitative risk assessment

Second-order Monte Carlo simulations

Distribution fitting

ABSTRACT

Quantitative risk assessment has emerged as a valuable tool to enhance the scientific basis of regulatory decisions in the food safety domain. This article introduces the use of two new computing resources (R packages) specifically developed to help risk assessors in their projects. The first package, “fitdistrplus”, gathers tools for choosing and fitting a parametric univariate distribution to a given dataset. The data may be continuous or discrete. Continuous data may be right-, left- or interval-censored as is frequently obtained with analytical methods, with the possibility of various censoring thresholds within the dataset. Bootstrap procedures then allow the assessor to evaluate and model the uncertainty around the parameters and to transfer this information into a quantitative risk assessment model. The second package, “mc2d”, helps to build and study two dimensional (or second-order) Monte-Carlo simulations in which the estimation of variability and uncertainty in the risk estimates is separated. This package easily allows the transfer of separated variability and uncertainty along a chain of conditional mathematical and probabilistic models. The usefulness of these packages is illustrated through a risk assessment of hemolytic and uremic syndrome in children linked to the presence of *Escherichia coli* O157:H7 in ground beef. These R packages are freely available at the Comprehensive R Archive Network (cran.r-project.org).

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Over the past decade, risk assessment has emerged as a valuable tool that can be used in the management of risks posed by foodborne pathogens and in the determination of standards for food in international trade (FAO/WHO, 2004). According to experts and international organization recommendations (Codex alimentarius Commission, 1999; European Commission, 2003; FAO/WHO, 2006; Vose, 2000), a quantitative risk assessment (QRA) should reflect the variability in the risk and evaluate separately the uncertainty associated with the risk estimates. Variability represents temporal, geographical and/or individual heterogeneity of the risk for a given population. Uncertainty is understood as stemming from a lack of perfect knowledge about the adequacy of the QRA model to reflect the situation and the lack of perfect knowledge about associated parameters. Variability and uncertainty are recommended to be treated separately because each has a different implication for risk management. Risk managers are generally interested in an estimate of the mean population risk and its associated uncertainty. Separation of uncertainty and variability is needed to provide such a result, by first

integrating only variability to calculate a mean population risk, and then evaluating the uncertainty on this estimate.

In order to consider the natural variability of a risk, the distribution of the risk within the population (and/or according to time and/or location) may be estimated from the mathematical combination of distributions reflecting the variability of exposure and susceptibility across the population (and/or time and/or location). As an example, the risk may be mathematically linked to the level of exposure to a given hazard through a dose–response model; this dose–response relationship may be variable according to identified sub-population. Moreover, this exposure may be variable in the population, as a result of the combination of the serving-to-serving variability of contamination and the serving size variability. Distributions should be derived from data (empirical distribution or fitting of a parametric distribution), from statistical theory (e.g., law of large numbers), from knowledge of the underlying process (e.g. binomial process or Poisson process), or from expert elicitation (Vose, 2000). A Monte-Carlo simulation framework is a useful approach to reduce inextricable calculations of distribution densities, by deriving an empirical distribution that is asymptotically similar to the distribution of the risk across the population (and/or time and/or location). Risk managers may not be interested in the variability of the risk in the population, but only in some population statistics such as the mean risk or the expected number of cases. It is sometime possible to

* Corresponding author. VetAgro Sup Campus Vétérinaire de Lyon, 1 avenue Bourgelat, F-69280 Marcy l'Etoile, France.

E-mail address: ml.delignette@vetagro-sup.fr (M.L. Delignette-Muller).

Nomenclature

<i>C</i>	bacterial concentration in <i>E. coli</i> O157:H7 (CFU/g);
<i>CP</i>	cooking preference (1: “rare”, 2: “medium”, 3: “well cooked”);
<i>D</i>	dose per serving (CFU/serving);
<i>HUSRisk</i>	risk of HUS per serving;
$\log_{10}C$	bacterial concentration in <i>E. coli</i> O157:H7 (\log_{10} CFU/g);
$mean.\log_{10}C$	mean parameter of the Normal distribution describing the distribution of $\log_{10}(C)$ (\log_{10} (CFU/g));
$sd.\log_{10}C$	standard deviation parameter of the Normal distribution describing the distribution of $\log_{10}(C)$ (\log_{10} (CFU/g));
<i>p</i>	vector of probabilities for the multinomial distribution characterizing <i>CP</i> ;
<i>r</i>	probability of infection following the ingestion of one cell of <i>E. coli</i> O157:H7 for a less than 5 year old children;
<i>R</i>	\log_{10} reduction of the bacterial concentration during the cooking step (\log_{10} (CFU/g));
<i>S</i>	serving size (g);
<i>Scale.S</i>	scale parameter of the Weibull distribution describing the Serving size distribution;
<i>Shape.S</i>	shape parameter of the Weibull distribution describing the Serving size distribution.

evaluate the mean risk from mathematical combination of the mean of the distributions, notably when the underlying models are linear. Nevertheless, the evaluation of the whole variability of the risk is usually needed to evaluate the mean risk, specifically in the microbiological QRA where non linear model is the rule.

A two-dimensional (or second-order) Monte-Carlo simulation (2D-MC) was proposed to estimate the uncertainty in the risk estimates stemming from parameter uncertainty (Cullen and Frey, 1999; Frey, 1992). A 2D-MC is a Monte-Carlo simulation where the distributions reflecting variability and uncertainty are sampled separately in the Monte-Carlo simulation framework, so that variability and uncertainty in the output may be estimated separately. This framework has been used in various domains of risk assessment (e.g. Jang et al., 2009; Jones et al., 2009b; Özkaynak et al., 2009; Vicari et al., 2007), and is now used in the food safety domain (e.g. FDA/USDA/CDC, 2003; Pouillot et al., 2007; Rimbaud et al., 2010). Nevertheless, variability and uncertainty are gathered in some risk assessment studies through a one dimensional Monte-Carlo simulation (e.g. Danyluk et al., 2006; Ross et al., 2009), most likely because the implementation of two-dimensional simulations remains difficult.

In this article, we describe two R (© The R Foundation for Statistical Computing) packages specifically written to help risk assessors build and study their models within a one dimensional or 2D-MC framework. The package “fitdistrplus” is a set of functions for distribution fitting and the package “mc2d” is a set of functions to build and evaluate 2D-MCs. R is an open-source integrated suite of software facilities for data manipulation, calculation, and graphical display, which is extended by a large collection of packages in which up-to-date statistical methods are implemented (R Development Core Team, 2009).

This article introduces these two R packages through an example from the food safety domain. This example is adapted from a risk assessment for *Escherichia coli* O157:H7 in frozen ground beef patties consumed by children in French households (Delignette-Muller and Cornu, 2008). The aim of the QRA was to evaluate the risk of hemolytic and uremic syndrome (HUS) for less than 5 year old children, following the consumption of ground beef patties accidentally contaminated with *E. coli* O157:H7. These patties were produced

from a contaminated production for which enumeration data was assumed available.¹ The QRA model is as follows: given *S* the serving size (g), *C* the bacterial concentration (CFU/g) in uncooked patties and *R* the \log_{10} -reduction of the bacterial concentration during the cooking step (\log_{10} (CFU/g)), the number of bacteria *D* in a random serving is assumed to follow a Poisson distribution with parameter $S \times C \times 10^{-R}$ CFU per serving. *R* is a function of the cooking preference (*CP*) categorized as “rare”, “medium” or “well-done”. For a given dose *D*, the probability to develop an HUS is assumed to be $\Pr(HUS|D) = 1 - (1-r)^D$, with *r* the constant probability of infection following the ingestion of 1 cell of *E. coli* O157:H7. The directed acyclic graph of the model is provided in Fig. 1.

This article follows the different steps used to derive a QRA model that separates variability and uncertainty: the first section depicts how to obtain a parametric distribution for parameters from quantitative or semi-quantitative data, the second section illustrates how to model uncertainty of the parameters, the third section illustrates how to integrate uncertainty and variability in the 2D-MC model separately, and the fourth section describes how to study the model and the results. The code, along with commentary lines used to derive the results, is provided in the Appendix A.

2. Fitting distributions to data

Within a risk assessment, one often has to fit a parametric distribution to a set of observed data (Haas et al., 1999; Vose, 2000). Fitting distributions to data consists of selecting the best fitting probability distribution from a predefined family of distributions. It requires judgment and expertise and generally needs an iterative process of distribution choice, parameter estimation, and quality of fit evaluation. The package “fitdistrplus” provides a set of functions dedicated to help the overall process of fitting a univariate parametric distribution to various types of data, including censored data.

2.1. Fitting distribution to quantitative data

In Delignette-Muller and Cornu (2008), a consumer survey provided a set of 254 serving sizes of ground beef consumed by children less than 5 years of age. The risk assessment required the characterization of the inter-individual variability distribution of the serving size *S* (g). The package “fitdistrplus” offers different ways to characterize and visualize a dataset in order to help the assessor choose of one or more parametric distribution(s). One function (“plotdist”) provides the plot of the empirical distribution as a histogram and as an empirical cumulative distribution function (ecdf) (results not shown). The skewness and kurtosis are often useful to help the selection of appropriate distribution candidates to fit a dataset. For this purpose, Cullen and Frey (1999) proposed to display on a graph, the relationship between the square of skewness and the kurtosis of standard parametric distributions, superimposed with skewness and kurtosis values estimated from the dataset. A function (“descdist”) is proposed in the package, which provides values of various descriptive parameters describing an empirical distribution, and a skewness–kurtosis plot as proposed by Cullen and Frey (1999). Fig. 2 shows this graph for the serving size dataset *S* (see the code in Appendix A.1). It suggests that three distributions are potential candidates to fit these data: the lognormal, the gamma and the Weibull distributions.

The next step in selecting a distribution consists of estimating parameters for one or more chosen distributions. For a given distribution, a function (“fitdist”) may be used to estimate the

¹ The data and the model were adapted to better illustrate the use of the R packages: the results provided in this article will not and should not be interpreted as an estimation of the actual risk from consumption of ground beef patties. Interested readers should refer to the original article (Delignette-Muller and Cornu, 2008) for a full description of the model, the data, and a comprehensive discussion of the results.

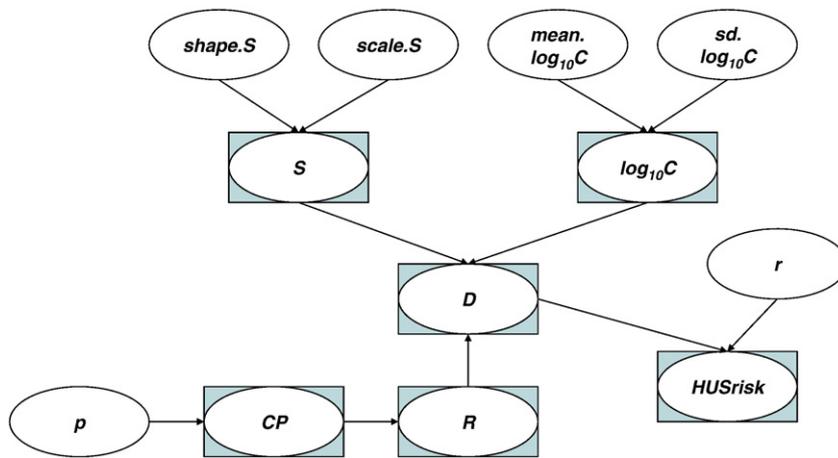


Fig. 1. Directed acyclic graph of the example: *Escherichia coli* O157:H7 in ground beef. See text and nomenclature for a description of the parameters and their relationships. Ovals stand for uncertain parameters and ovals-in-rectangles stand for variable and uncertain parameters.

parameters using the maximum likelihood (ML) method or, alternatively, the method of matching moments, if available. The output of this function, for the fitting of a Weibull distribution to the serving size dataset, is shown in Fig. 3. The best estimates are 2.19 for the shape parameter (standard error (s.e.): 0.105) and 83.35 for the scale parameter (s.e.: 2.527). The standard errors of the ML estimates are derived classically from the Hessian matrix. The inclusion of the function “fitdist” within the R environment allows evaluation of the fit of a large number of univariate distributions. The quality of the fit is assessed using classical goodness-of-fit statistics (Chi-squared, Kolmogorov–Smirnov and Anderson–Darling statistics (Anderson and Darling, 1952), Fig. 3) and graphs (empirical and theoretical

distributions plot in density and in cumulative density function, P–P plot and Q–Q plot, Fig. 4). Note that for continuous data, the Anderson–Darling statistic and the Q–Q plot are especially relevant, as they emphasize deviations at distribution tails that may have a strong impact in quantitative risk assessment.

Similar adjustments may be done with other distributions. In our example (code and results not shown), a gamma distribution and a Weibull distributions fit the data equally well, with the same value of the Anderson Darling statistic ($AD=3.57$). A slightly higher value ($AD=4.54$) is obtained with a lognormal distribution. The Weibull distribution may be preferred to the gamma distribution due to a far lower correlation between the parameters of this latter distribution ($r=0.32$ for the Weibull distribution and $r=0.94$ for the gamma distribution).

In addition to the fitting of distributions for continuous data as just described, the “fitdistrplus” package can also accommodate discrete variables, such as count numbers. Specific distributions such as the Poisson distribution or the negative binomial distribution can be fitted with the same functions. They will provide somewhat different statistics and graphs taking into account the discrete type of the modeled variable.

2.2. Fitting distribution to semi-quantitative or censored data

Analytical methods often lead to semi-quantitative results which are referred to as “censored values”. Some results may be under a limit of detection: these are said to be “left censored” data. Others lie between two values, for example above the limit of detection but under the limit of quantification: these are termed “interval censored” data. Some “right censored” data could also be observed if a maximum level of quantification is reached or exceeded (Busschaert et al., 2010; Lorimer and Kiermeier, 2007). Thus, the final dataset may contain right-, left-, or interval-censored data, or it may be a mixture of these three categories, possibly with different upper and lower bounds if data were obtained with several analytical methods. Fitting distributions to these types of datasets without considering the censoring would lead to biased results (Busschaert et al., 2010). A recommended approach to correctly model such data for risk assessment is based upon maximum likelihood estimation (Busschaert et al., 2010; Helsel, 2005). This may be done using the “fitdistrplus” package using the function “fitdistscens”.

In our example, we used a fictitious dataset of 18 counts of *E. coli* O157:H7 from the contaminated production of ground beef. This dataset included two left-censored data points, under the 0.04 CFU/g limit of detection, and four interval-censored data points with different thresholds. Those data are needed to characterize the inter-batch variability of *C*, the concentration of *E. coli* O157:H7 in

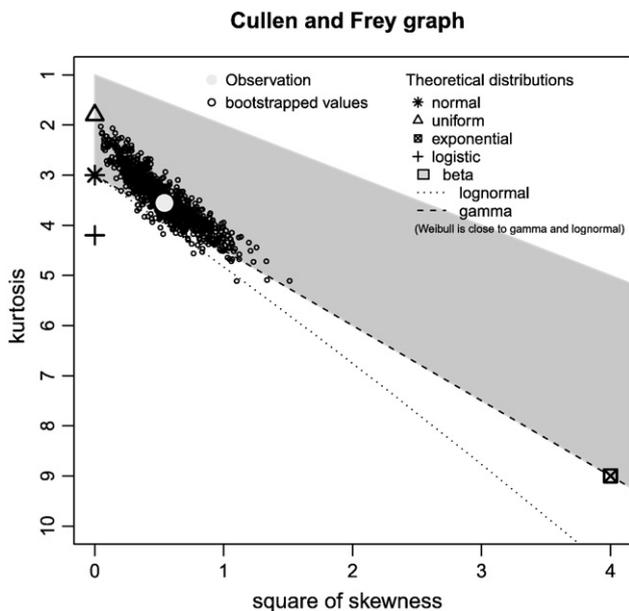


Fig. 2. Output of the “descdist” function: skewness–kurtosis plot (Cullen and Frey, 1999) for the serving size dataset. The large circular point is located at the square of the skewness (x-axis) and the kurtosis (y-axis) evaluated from the dataset. This location is compared to theoretical locations for various standard distributions. For some distributions (normal, uniform, logistic, exponential), there is only one possible value for the skewness and the kurtosis (for a normal distribution for example, skewness = 0 and kurtosis = 3), and the distribution is thus represented by a point on the plot. For other distributions, areas of possible values are represented, consisting in lines (gamma and lognormal distributions), or larger areas (beta distribution). In this example, the lognormal, the gamma, and the Weibull distributions appears as potential candidates to fit the serving size dataset. One hundred non-parametric bootstrap samples allow the evaluation of the robustness of the proposals.

```
> summary(fit.S)
Fitting of the distribution 'weibull' by maximum likelihood
Parameters :
      estimate Std. Error
shape      2.19      0.105
scale     83.35     2.527
Loglikelihood: -1255   AIC: 2514   BIC: 2522
Correlation matrix:
      shape scale
shape 1.000 0.322
scale 0.322 1.000

> gofstat(fit.S)
Kolmogorov-Smirnov statistic: 0.140
Anderson-Darling statistic: 3.57
```

Fig. 3. Summary of the fitting procedure of a Weibull distribution to the serving size dataset and goodness-of-fit values and tests results.

uncooked ground beef. Once coded in a relevant way, so as to precisely define each censored value (see Code in Appendix A.2), such data may be fitted to any parametric distribution in a similar way as was shown for non-censored data. Here, a normal distribution was fitted to $\log_{10}(C)$ data. The best estimates are 0.232 for the mean parameter (S.E.: 0.323) and 1.346 for the standard deviation parameter (S.E.: 0.247). A goodness-of-fit graph is provided to evaluate the fit, with interval and left censored values represented by segments on the cumulative distribution plot (Fig. 5).

3. Modeling uncertainty on each variable of a risk assessment model

In order to enable two-dimensional Monte Carlo simulations such as proposed by Cullen and Frey (1999), one needs to quantitatively describe

the uncertainty associated with the parameters in the risk assessment model. Bootstrap resampling method, Bayesian derivation, or expert elicitation are some of the available methods (Vose, 2000). This section illustrates how the two packages can be used to derive such uncertainty distributions. The uncertainty and variability of the model inputs will then be transferred to the two-dimensional Monte Carlo framework.

3.1. Simulating uncertainty from quantitative or semi-quantitative data

Parametric or non-parametric bootstrap resampling (Efron and Tibshirani, 1993) may be used to evaluate and simulate the uncertainty in parameters estimated from a distribution (Busschaert et al., 2010; Cullen and Frey, 1999). The “mc2d” package can derive non-parametric bootstrap resampling from empirical distributions. The “fitdistrplus” package can derive parametric or non-parametric bootstrap procedures

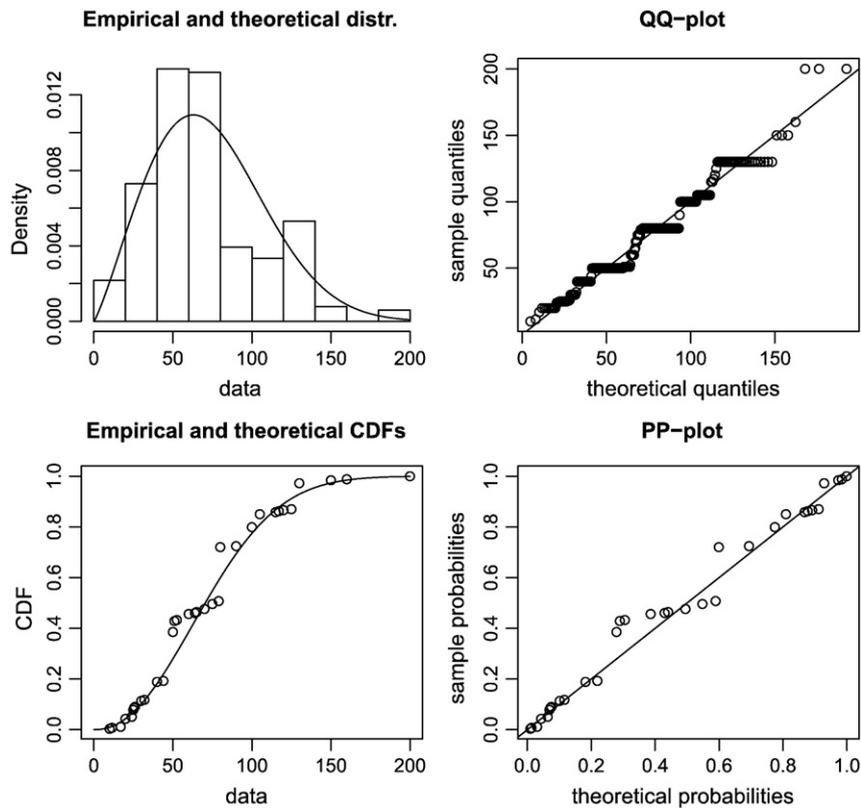


Fig. 4. Results of the fitting of a Weibull distribution to the serving size dataset. Clockwise: i) histogram of the empirical distribution (data) superimposed with the density function of the theoretical fitted distribution; ii) Q-Q plot, i.e. plot of the quantiles of the theoretical fitted distribution (x-axis) against the empirical quantiles of the data (y-axis); iii) P-P plot, i.e. for each value of the data set, the cumulative density function of the fitted distribution (x-axis) is plotted against the empirical cumulative density function (y-axis); iv) empirical density function of the data superimposed with the cumulative density function of the theoretical fitted distribution.

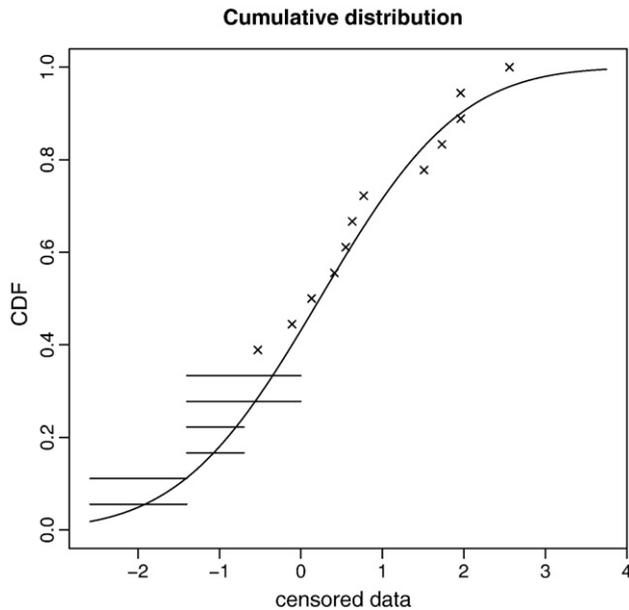


Fig. 5. Goodness-of-fit graph for the fit of a lognormal distribution censored dataset of the decimal logarithm microbial concentration from the example. Crosses stand for non censored data. Segments lie between the limits of the censoring interval.

for any parametric distribution preliminarily fitted to quantitative data, and non-parametric bootstrap for any parametric distribution preliminarily fitted to semi-quantitative data. In this example, a parametric bootstrap is used to model the uncertainty around the scale and shape parameters of the fitted Weibull distribution for S and a non-parametric bootstrap is used for the uncertainty around the mean and standard deviation of the fitted lognormal distribution for C . The joint probability distribution obtained in the bootstrap sample of parameters may then be used to characterize uncertainty about these parameters (Cullen and Frey, 1999) and be directly input into a 2D-MC model (see Code in Appendix A.3). Note that measurement uncertainties (imprecision and bias), which are particularly relevant in microbiological risk assessment, could be similarly transferred in the model, provided that the risk assessor is able to measure it.

3.2. Simulating uncertainty from qualitative data

Qualitative data are also often used in risk assessment and, assuming some independence, the variability distribution of the data may be characterized by a binomial or a multinomial distribution. If the observed dataset is small, one may want to describe and model uncertainty of the parameters of such distributions. A simple Bayesian integration may be obtained if a beta distribution is used as a prior distribution for the probability in a binomial process (Vose, 2000), or if a Dirichlet distribution is used as a prior distribution for the probabilities in a multinomial process (Christensen et al., 2005). This may be illustrated in the *E. coli* model on the cooking preference (CP) for ground beef (Delignette-Muller and Cornu, 2008): let us denote in this example p_i ($i = 1, 2, 3$) the proportion of consumers in the population whose cooking preference (CP) for ground beef is “rare”, “medium” and “well-done”, respectively. An *ad-hoc* survey indicated that $n_1 = 14$, $n_2 = 59$ and $n_3 = 71$ random consumers cooked the ground beef for their children as rare, medium or well-done, respectively. Let us consider the uncertainty around \hat{p}_i inferred from these results. The Dirichlet distribution, multivariate generalization of the beta distribution, and conjugate prior to the multinomial, is used to model such uncertainty. Using a Dirichlet(1, 1, 1) distribution as a prior distribution, the posterior distribution is then a Dirichlet distribution with parameter $\alpha = (n_1 + 1, n_2 + 1, \dots, n_k + 1)$, where (n_1, n_2, \dots, n_k) are the observed

counts (Christensen et al., 2005). In our example, the Dirichlet distribution is then parameterized by $\alpha = (14 + 1, 59 + 1, 71 + 1)$. The “mc2d” package allows the user to easily manipulate multivariate distributions such as the Dirichlet distribution. The corresponding code for this example is provided Appendix A.4.

3.3. Simulating uncertainty from expert knowledge and some more possibilities

In a risk assessment model, uncertainty of inputs may in some cases be defined from expert knowledge. Triangular and PERT distributions are commonly used for this purpose (Vose, 2000). These distributions were also included in the “mc2d” package. In our example, the dose–response model parameter r , could be considered uncertain. This parameter corresponds in the single-hit model to the probability of HUS following the ingestion of a single cell (Haas et al., 1999). In the original paper, Delignette-Muller and Cornu (2008) estimated this parameter, for children under the age of 5 from epidemiological data collected during a French outbreak, to be 1.2×10^{-3} (95% CI: $[5.3 \times 10^{-4}, 2.3 \times 10^{-3}]$). Since this estimation was obtained from only one dataset and without considering uncertainty on consumption data, it seems reasonable to enlarge this interval. The uncertainty on r may be defined using a PERT distribution characterized by three parameters, with the minimum equal to 10^{-4} , the most likely value equal to 1.2×10^{-3} , and the maximum equal to 10^{-2} (see Appendix A.5 for the corresponding R code).

Other functions are available in the package “mc2d” in order to simulate the uncertainty of model inputs. The function “mcprobtree” may be used, for example, to build a probability tree distribution (or mixture of distributions) representing the uncertainty of choice between two or more models for an input. This function may also be used to generate samples from a mixture distribution for variability (see next section and Appendix A.6). This package also provides some distributions that can be useful to the risk assessor, i.e., the Bernoulli, the discrete empirical, the continuous empirical, the generalized beta (beta distribution on any finite domain) and the multi-normal distributions. Other tools are also available to sample from truncated distributions, to sample from univariate distributions using Latin hypercube sampling and to build a rank correlation structure between two or more parameters using the Iman and Conover (1982) method.

4. Integrating uncertainty and variability in a two-dimensional Monte Carlo simulation framework

4.1. General framework

The major appeal of the “mc2d” package is the easy transfer of variability and uncertainty which are separated along the chain of conditional mathematical and probabilistic models. The framework is adapted from the one proposed by Frey (1992) and continued in Cullen and Frey (1999) (Fig. 6). Once all inputs have been disaggregated in uncertain or variable inputs, the 2D-MC simulation is developed as follows:

1. Uncertain inputs are randomly sampled from their respective distributions in row vectors of n_u elements, where n_u is a specified number of iterations in the 2D-MC uncertainty dimension;
2. Variable inputs are randomly sampled from their respective distributions in column vectors of n_v elements, where n_v is a specified number of iterations in the 2D-MC variability dimension;
3. For each uncertain and variable input, n_v values are randomly sampled conditionally to each of the n_u values of its uncertain parameters, leading to an $n_v \times n_u$ matrix of values;
4. The mathematical risk assessment model eventually combines these elements to derive the $n_v \times n_u$ final matrix of outputs, so that each $(v, u)^{th}$ entry corresponds to the combination of the v th

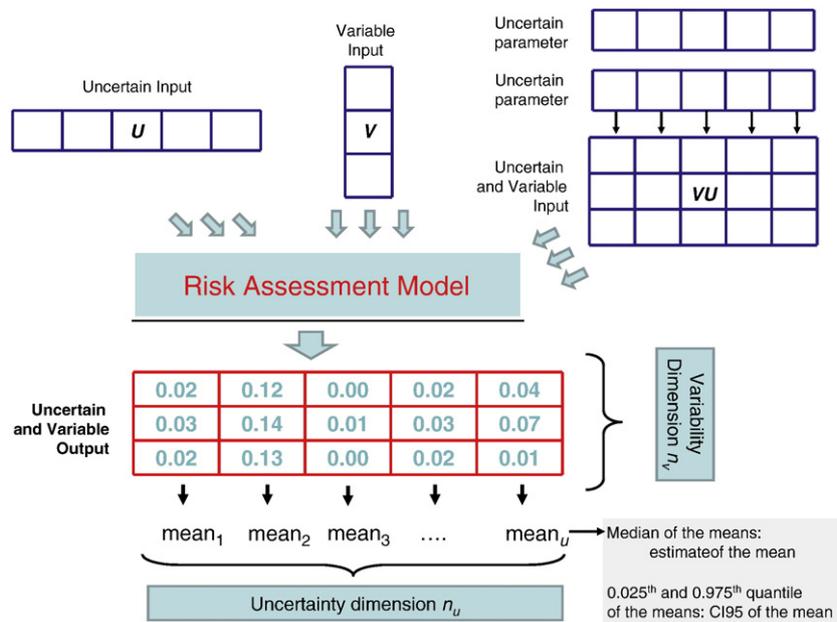


Fig. 6. Illustration of the two dimensional Monte-Carlo simulation.

random assay of the variable inputs, the u th random assay of the uncertain inputs and the $(v, u)^{th}$ random assay of the uncertain and variable parameters.

For a given u , the (\bullet, u) output values may be considered as the result of a one-dimensional Monte-Carlo simulation of size n_v , in which the uncertain parameters are treated as fixed. This QRA takes into account the variability in all variable parameters, and leads to a density function reflecting the variability of risk across the population, conditional upon the uncertain parameters.

Using the “mc2d” package, the user has to specify for each input if the associated randomness represents uncertainty or variability. The type of dispersion is then automatically propagated within the mathematical model. Say that V is a variable parameter and that U is an uncertain parameter. Illustrating a given mathematical function by the addition, $V + V$ leads to a variable parameter, $U + U$ leads to an uncertain parameter and $V + U$ leads to a variable and uncertain parameter, say W . Each of the $(v, u)^{th}$ entries of W corresponds to the addition of the v th random assay of V and the u th random assay of U . Similarly, $V + W$, $U + W$ and $W + W$ lead to a variable and uncertain parameter, with a correct recycling of the data. Variable, uncertain and/or variable and uncertain parameters may also be combined in parametric or empirical distributions, according to the user’s conditional model and to its dispersion characterization. Variability and uncertainty are then naturally transferred into the final outputs.

4.2. Simulating variability and uncertainty along the *E. coli* model

In our example, four inputs are defined at this step, S – the uncertain and variable serving size (g), C – the uncertain and variable bacterial concentration (CFU/g) in uncooked patties, CP – the uncertain and variable cooking preference categorized as “rare”, “medium” or “well-done” and r – the uncertain dose–response parameter. Each of the variable inputs (S , C and CP) may be easily simulated conditionally to the uncertain parameters of their distribution of variability, i.e. *Shape.S* and *Scale.S* for S , *mean.log10C* and *sd.log10C* for C and p for CP (see Appendix A.6 for the corresponding R code).

The \log_{10} -reduction of the bacterial concentration during the cooking step (R in $\log_{10}(\text{CFU/g})$) is a function of the cooking

preference CP . For each of the three consumption preferences (rare, medium, well-done), the number of decimal reductions of *E. coli* O157:H7 due to cooking, R_i , was assumed to be variable, defined by a uniform distribution: a Uniform(0, 0.9) for “rare”, a Uniform(0.2, 1.4) for “medium” and a Uniform(1.2, 2.8) for “well cooked” (Delignette-Muller and Cornu, 2008) where Uniform(a, b) is the uniform distribution between a and b . The “mcprobtree” function may be used to build the corresponding node in the Monte-Carlo model (see Appendix A.6 for the corresponding R code).

From all of the previously defined inputs, the number of bacteria D in a random serving and the corresponding probability to develop an HUS following the ingestion of this dose were computed (see Fig. 1 and Appendix A.6 for the corresponding R code). With the “mc2d” package, uncertainty and variability are automatically transferred to the outputs of the model using standard arithmetic operations to define D and the final HUS risk.

5. Study of a risk assessment model

Once the QRA model is developed, some additional tools are needed to study the model. The “mc2d” package includes functions to study the model through *ad-hoc* summaries, graphics and sensitivity analyses. Various statistics (e.g. the mean, the standard deviation, percentiles) of the resulting density function in the variability dimension are evaluated (Fig. 7). Repeating this procedure in the uncertainty dimension for $u = 1, \dots, n_u$ eventually leads to n_u means, n_u standard deviations and n_u set of percentiles, reflecting the uncertainty in the mean, the standard deviations, or the percentiles of the risk. As final outputs, the median (or, alternatively, the mean) of each statistic may be used as a point estimate of this statistic; the 2.5th and the 97.5th percentiles of each statistic are used to establish a 95% credible interval (CI95). Sensitivity analyses through rank correlations and “tornado charts” and other statistics proposed by Özkaynak et al. (2009) (variability ratio, uncertainty ratio and overall uncertainty ratio) are also available.

The output for the “summary” function applied to the *E. coli* model is provided in Fig. 7 (see Appendix A.7 for the corresponding R code). For each uncertain parameter (r in this example), the median, the mean, the 2.5th and 97.5th percentiles are estimated as a measure of the dispersion of this uncertainty. For variable and uncertain inputs ($\log_{10}C$, S and CP) and the output ($HUSrisk$), the median of the

```

> summary(modelEcoli)

log10C :
      mean   sd  Min 2.5%   25%   50%   75% 97.5%  Max  nsv Na's
median 0.245 1.302 -3.94 -2.30 -0.6434 0.246 1.127 2.81 4.48 1001 0
mean   0.237 1.306 -3.99 -2.31 -0.6425 0.237 1.116 2.79 4.51 1001 0
2.5%   -0.425 0.925 -5.87 -3.50 -1.4433 -0.431 0.399 1.75 2.89 1001 0
97.5%   0.854 1.732 -2.44 -1.32 0.0875 0.864 1.700 3.66 6.25 1001 0

S :
      mean   sd  Min 2.5%   25%   50%   75% 97.5%  Max  nsv Na's
median 73.8 35.3 3.065 15.9 47.3 70.5 96.6 150 205 1001 0
mean   73.8 35.4 3.253 16.0 47.3 70.6 96.6 151 208 1001 0
2.5%   69.4 32.3 0.692 12.0 42.0 65.4 90.7 139 177 1001 0
97.5%   78.6 39.2 6.663 20.2 52.4 75.7 103.2 163 249 1001 0

r :
      NoVar
median 0.002158
mean   0.002444
2.5%   0.000307
97.5%   0.006282

CP :
      mean   sd  Min 2.5%   25%   50%   75% 97.5%  Max  nsv Na's
median 2.39 0.662 1 1 2 2.00 3 3 3 1001 0
mean   2.39 0.662 1 1 2 2.43 3 3 3 1001 0
2.5%   2.28 0.596 1 1 2 2.00 3 3 3 1001 0
97.5%   2.50 0.724 1 1 2 3.00 3 3 3 1001 0

HUSrisk :
      mean   sd  Min 2.5%   25%   50%   75% 97.5%  Max  nsv Na's
median 0.1412 0.277 0 0 0.00000 0.009984 0.1019 1.000 1.000 1001 0
mean   0.1471 0.269 0 0 0.00104 0.016287 0.1493 0.939 0.999 1001 0
2.5%   0.0405 0.134 0 0 0.00000 0.000925 0.0111 0.442 1.000 1001 0
97.5%   0.2803 0.371 0 0 0.00639 0.065314 0.4986 1.000 1.000 1001 0

> tor <- tornado(modelEcoli)
> print(tor)

Spearman's rho statistic
Output: HUSrisk
$HUSrisk
      log10C      S      CP
median 0.8069322 0.14381736 -0.4046471
mean   0.8008802 0.14366550 -0.4041324
2.5%   0.6983308 0.07383674 -0.5261821
97.5%   0.8682931 0.21615094 -0.2978436

```

Fig. 7. Output of the code “summary(modelEcoli)”. For each node, some statistics (mean, sd, Min, ...) are evaluated on the variability dimension (in line). The confidence intervals of these statistics are provided through their 2.5th and 97.5th percentiles evaluated in the uncertainty dimension. Output of the code “print(tor)”: an estimate of the Spearman's rank correlation is provided for each node (in line). The confidence intervals of these rank correlations are provided through their 2.5th and 97.5th percentiles evaluated in the uncertainty dimension.

$n_u = 1,001$ statistics evaluated in the variability dimension are provided, as well as the mean, the 2.5th and the 97.5th percentiles. The mean risk, which is the output of interest for risk managers, is estimated by the median of the n_u means as 0.14 and the uncertainty around this estimate may be described by the 95% credible interval [0.04, 0.28]. A mean number of HUS cases between 4 and 28 for 100 consumptions may then be expected for children under the age of 5 consuming ground beef patties from the fictitious production defined in this example. The “plot” function provides a variability cumulative distribution graph, as proposed by Frey (1992) (Fig. 8). The Spearman rank order correlation coefficient ρ between the output and each input may also be computed using the function “tornado”, with its 2.5th and 97.5th uncertainty percentiles. In this example, the impact of the variability of C has the greatest affect on the risk (95% CI on ρ : [0.70, 0.87]), followed by a negative correlation with CP (95% CI on ρ : [-0.53, -0.30]) and a positive rank correlation with S (95% CI on ρ : [0.07, 0.22]) (Fig. 9).

6. Discussion

The aim of this article was to introduce the use of the two R packages to help risk assessors building and studying a model, and separate uncertainty and variability through a two-dimensional Monte Carlo simulation framework. A fictitious example adapted from a published

risk assessment (Delignette-Muller and Cornu, 2008) was chosen for sake of simplicity and to show various modeling possibilities offered in these two packages. The implementation of another published example of quantitative risk assessment of waterborne cryptosporidiosis in France (Pouillot et al., 2004) and the adaptation of a published example of quantitative risk assessment of listeriosis from cold-smoked salmon consumption in France (Pouillot et al., 2009, 2007) may be found in the documents² included in the package distributions. For technical details, manuals are included in each package. R and the packages introduced in this paper are freely available at the Comprehensive R Archive Network (CRAN, <http://cran.r-project.org>).

The function “fitdistr” in the R package “MASS” (Venables and Ripley, 2002) is a well known general-purpose maximum-likelihood fitting routine for the parameter estimation step in R. Other steps of the process may be developed using R (Ricci, 2005) but, to our knowledge, no specific package has been implemented previously for distribution fitting purpose. This package will be of help for risk assessors, but also for any other R user in need of a practical and efficient tool to fit parametric distributions. This package is especially helpful to analyze observed data with censored values, as is often

² In R jargon, the “vignettes”.

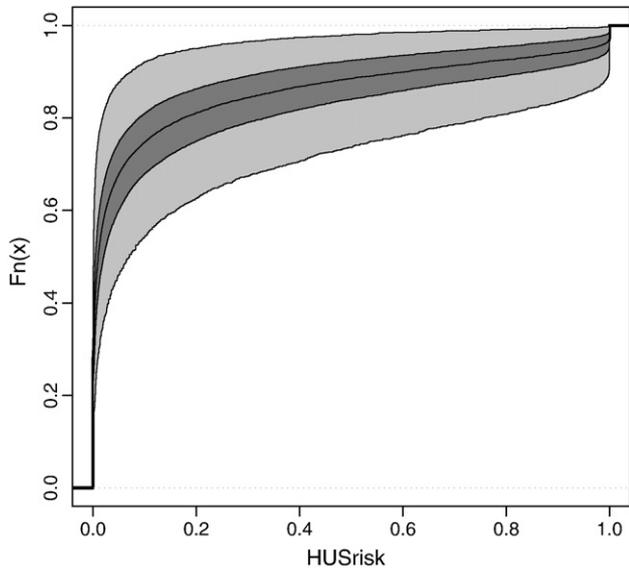


Fig. 8. Variability cumulative distribution plots of the output of the *E. coli* model, with uncertainty indicated: light gray band corresponds to the 95% uncertainty range on each quantile of variability and dark gray band corresponds to the 50% uncertainty range on each quantile of variability.

encountered in risk assessment, as a result of limits of detection or quantification of analytical methods (Busschaert et al., 2010).

The package “mc2d” is more specific to the risk assessment domain. To our knowledge, it is the first attempt to provide users with tools to derive and study their 2D-MC models within the R environment. While the separated estimation of uncertainty and variability is a classical recommendation (Codex alimentarius Commission, 1999; European Commission, 2003; FAO/WHO, 2006; Vose, 2000), it remains challeng-

ing and infrequently implemented. Yet, this strict separation and a careful tracking of uncertainty may outline a great uncertainty in the final estimates that can impair some of the conclusions of the risk assessment (Pouillot et al., 2009) and more generally may avoid serious interpretation errors due to the confounding of uncertainty and variability. This package may help to increase the use of this 2D-MC method, even if the identification and assessment of the overall data and model uncertainty remain a challenge for risk assessors.

The R environment is now largely used for statistical computing in many areas, and is increasingly used by several authors in microbial quantitative risk assessment to perform Monte-Carlo simulations (Afchain et al., 2008). The R environment is also useful for dealing with microbial censored data using maximum likelihood and censored regression (Busschaert et al., 2010; Lorimer and Kiermeier, 2007), to estimate parameters of dose-response models (Jones et al., 2009a), and to fit non linear predictive models (Delignette-Muller, 2009). Both packages use classical R objects: “mc2d” builds classical three dimensional arrays (one dimension for the variability, one for the uncertainty and one for the different variates in multivariate distributions). These objects can be accessed directly for interactions with sophisticated statistical models from the classical R distribution or other packages. The output may directly be used in ANOVA, linear regression models, CART or other recommended sensitivity analysis tools, like Saltelli’s method (Frey et al., 2004), all of which are available in the R environment.

Like all R packages, the “fitdistrplus” and “mc2d” code is open, which is a guarantee of the recommended transparency in quantitative risk assessment models. Moreover, these tools are free, which could help students, resource limited institutes, or resource limited countries, and the development of training to perform 2D-MC (or one dimensional Monte-Carlo) QRAs.

These packages are mainly suitable for users with an intermediate experience with R. Nevertheless, a training session performed in France by the authors demonstrated that students with no prior experience with R are able to develop their own models in a few days. Moreover, many free introductory manuals and commercial books are now available to easily learn the basics of R language (see <http://cran.r-project.org>). On the contrary, experienced users of R may find that these flexible functions are more memory-consuming and time-consuming than the specific code they themselves might write. Indeed, the major objective in the development of these packages was the flexibility of the models that could be implemented. For those experienced users, we speculate that some functions will be of great help, notably some functions from “fitdistrplus”, but also from “mc2d” designed to analyze 2D-MC simulations results, as they may easily be used on independent simulation results converted as an mc object.

Future planned developments for “mc2d” are to develop *ad-hoc* functions for more efficient sensitivity analyses, both in the uncertainty and the variability dimension (Frey et al., 2004) and to provide a larger choice of plots (e.g. spaghetti-looking graphs and box-plots of variability statistics) to help the interpretation of simulation results. With respect to “fitdistrplus”, we plan to provide graphs of likelihood contours in order to detect potential identifiability issues. We also plan to develop other methods to fit a distribution (such as quantiles method) and other goodness-of-fit plots and statistics, especially for censored data. Thus, these packages are constantly in development through an R-Forge project (<https://r-forge.r-project.org/projects/riskassessment/>). We encourage readers to test the current packages and developers to improve current functions, implement additional functions, and/or share their own packages on this platform.

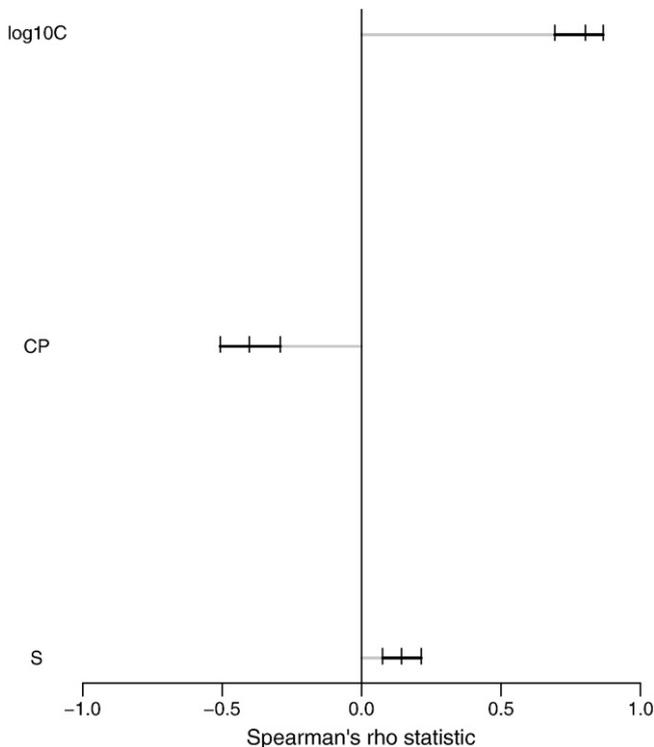


Fig. 9. Tornado chart from the example. The mark is the median estimate of the spearman’s rank correlation between the inputs and the output, bounded by the 95% uncertainty range.

Acknowledgements

We would like to thank Meryl B. Lubran for her help in reviewing this article. We also thank J.-B. Denis, C. Dutang and D.L. Kelly for their help in the development of the packages and their documentation.

Appendix A. Code

Beginning users of R can find information in one of the multiple free manuals provided on the Comprehensive R Archive Network in rubrics “Manuals” or “Contributed” (www.r-project.org). Lines beginning with a pound sign are commentaries and thus do not need to be typed in R.

A.1. Code to fit a Weibull distribution to serving size data using *fitdistrplus*

Below is the R code used to obtain the results presented in Section 2.1, Figs. 2–4 for serving size data.

```
# load the fitdistrplus package
require(fitdistrplus)
# load the data
data(groundbeef)
serving <- groundbeef$serving
# Cullen and Frey graph to help the choice of candidate
distributions
# (Fig. 2)
descdist(serving, boot = 1000, obs.col = "white", boot.
  col = "black")
# fit of a Weibull distribution
fit.S <- fitdist(serving, "weibull")
# print the estimates (Fig. 3)
summary(fit.S)
# compute and print goodness-of-fit statistics (Fig. 3)
gofstat(fit.S)
# four goodness-of-fit graphs (Fig. 4)
plot(fit.S)
```

A.2. Code to fit a lognormal distribution to *E. coli* O157:H7 concentration in uncooked ground beef data using *fitdistrplus*

Below is the R code used to fit a lognormal distribution to fictitious (but realistic) semi-quantitative data of *E. coli* O157:H7 concentration in uncooked ground beef (Section 2.2 and Fig. 5). It is first necessary to create (or to import) an R data frame (named `data.log10C` in this example) with two columns respectively named “left” and “right” in order to code this dataset. Left and right values correspond to bounds of interval censored values, and are equal for non-censored values. For left (resp. right) censored values, the left (resp. right) value is fixed to “NA”. In this example, we would have:

```
# building data.log10C, a data frame containing the
log10
# of the observed concentrations
data.log10C <- data.frame(
  left = c(1.73, 1.51, 0.77, 1.96, 1.96, -1.4, -1.4, NA,
    -0.11, 0.55, 0.41, 2.56, NA, -0.53, 0.63,
    -1.4, -1.4, 0.13),
  right = c(1.73, 1.51, 0.77, 1.96, 1.96, 0, -0.7, -1.4,
    -0.11, 0.55, 0.41, 2.56, -1.4, -0.53, 0.63,
    0, -0.7, 0.13)
)
```

The first value is not censored (left = right = 1.73), the 6th is interval-censored (left = -1.4, right = 0) and the 8th is left-censored (left = NA, right = -1.4), etc. Once `data.log10C` is built, the function `fitdistcens` may be used to fit the distribution as follows:

```
# fit a normal distribution to the decimal logarithm of C
fit.log10C <- fitdistcens(data.log10C, "norm")
# print the estimates
```

```
summary(fit.log10C)
# goodness-of-fit graphs for censored data (Fig. 5)
plot(fit.log10C)
```

A.3. Code to simulate uncertainty on distribution parameters by bootstrap resampling from quantitative or semi-quantitative data

The functions “bootdist” and “bootdistcens” may be used to simulate a bootstrap sample of estimated parameters of a fitted distribution to describe their uncertainty. Below is the R code used to simulate bootstrap samples for the Weibull distribution fitted to serving size data and for the lognormal distribution fitted to microbial concentration in uncooked ground beef (see Section 3.1):

```
# parametric bootstrap resampling for Weibull S
distribution
boot.S <- bootdist(fit.S, niter = 1001, bootmethod =
  "param")
# nonparametric bootstrap resampling for normal log10C
distribution
boot.log10C <- bootdistcens(fit.log10C, niter = 1001)
# bootstrap medians and 95% percentile confidence inter-
vals for log10C
summary(boot.log10C)
```

To enable the use of these bootstrap samples in a two-dimensional simulation framework using `mc2d`, it is necessary to transform them (using the function `mcdata`) in a `mcnode`. A `mcnode` is the class of objects corresponding to parameters of a risk assessment model within `mc2d`. Each `mcnode` has a `type`, V, U, VU or 0 whether it is considered variable, uncertain, variable and uncertain (in two dimensions) or constant, respectively. Thus each bootstrap sample must be transformed in a `mcnode` of type U as below. Prior to the transformation, the package “`mc2d`” should be loaded and the uncertainty and variability dimensions should be defined using the functions `ndunc` and `ndvar`, respectively.

```
# load mc2d package
require(mc2d)
# specification of uncertainty and variability
dimensions
ndunc(1001); ndvar(1001)
# uncertainty on the distribution parameters of the
serving size
shape.S <- mcdata(boot.S$estim$shape, type = "U")
scale.S <- mcdata(boot.S$estim$scale, type = "U")
# uncertainty on the distribution parameters of the
log10 concentration
mean.log10C <- mcdata(boot.log10C$estim$mean, type = "U")
sd.log10C <- mcdata(boot.log10C$estim$sd, type = "U")
```

A.4. Code to simulate uncertainty on distribution parameters from qualitative data

As `mcdata`, the function `mcstoc` aims at defining a `mcnode` but this time by sampling a set of values from a specified distribution. It can be used here to simulate uncertainty on the multivariate parameter `p` of the distribution of the consumption preference CP.

Note the use of the `nvariables` option to specify the multivariate property of `p`.

```
# uncertainty on the probabilities of CP distribution
p <- mcstoc(rdirichlet, type = "U", alpha = c(14+1, 59+
  1, 71+1), nvariables = 3)
```

A.5. Code to simulate uncertainty from expert knowledge

The function `mcstoc` may be used here to define a `mcnode` of type `U` from a PERT distribution specified from expert knowledge of the dose–response parameter `r`.

```
# uncertainty on the dose-response parameter r
r <- mcstoc(rpert, type = "U", min = 1e-4, mode = 1.2e-3,
            max = 1e-2)
```

A.6. Code to simulate variability conditionally to uncertainty and transfer both along the model

The function `mcstoc` is used below to sample `mcnode` of type `VU` for `S`, `log10C`, `CP` and `R` variability conditionally to uncertainty.

```
# variability of the serving size conditionally to
uncertainty
S <- mcstoc(rweibull, type = "VU", shape = shape.S, scale =
            scale.S)
# variability of the log10 concentration conditionally
to uncertainty
log10C <- mcstoc(rnorm, type = "VU", mean = mean.log10C,
                sd = sd.log10C)
# variability of the consumption preference condi-
tionally to uncertainty
# CP is a node with values 1 for rare, 2 for medium, 3 for
well cooked.
CP <- mcstoc(rempiricalD, type = "VU", prob = p, values =
            1:3)
# log10 reduction according to CP as a mixture of uniform
distributions
R <- mcprobtree(CP,
               list("1" = mcstoc(runif, type = "V", min = 0, max = 0.9),
                    "2" = mcstoc(runif, type = "V", min = 0.2, max =
                                1.4),
                    "3" = mcstoc(runif, type = "V", min = 1.2, max =
                                2.8)),
               type = "VU")
```

The number of bacteria `D` in a random serving is sampled using `mctsoc` from a Poisson distribution in which parameters are calculated from other inputs with automatic transfer of their variability and uncertainty. At last, the HUS risk is calculated by simple arithmetic operations from `r` and `D`.

```
# ingested dose D defined by a Poisson distribution
D <- mcstoc(rpois, type = "VU", lambda = S * 10^(log10C - R))
# HUS risk resulting from r (type U) and D (type VU)
HUSrisk <- 1 - (1 - r)^D
```

A.7. Code to study the results of the model

Various nodes of a Monte Carlo model may be gathered in a `mc` object (function `mc`) and the results may be summarized, plotted, and studied using the following code:

```
modelEcoli <- mc(log10C, S, r, CP, HUSrisk)
#Summary (Fig. 7)
summary(modelEcoli)
#Plot (Fig. 8)
plot(HUSrisk)
```

Note that the results slightly differ from one run to the other, due to the random process involved in Monte Carlo simulations.

A tornado chart including uncertainty may be calculated and plotted:

```
tor <- tornado(modelEcoli)
# Spearman's rank correlations (Fig. 7)
print(tor)
# Tornado chart (Fig. 9)
plot(tor)
```

References

- Afchain, A.L., Carlin, F., Nguyen-The, C., Albert, I., 2008. Improving quantitative exposure assessment by considering genetic diversity of *B. cereus* in cooked, pasteurised and chilled foods. *International Journal of Food Microbiology* 128, 165–173.
- Anderson, T.W., Darling, D.A., 1952. Asymptotic theory of certain goodness of fit criteria based on stochastic processes. *The Annals of Mathematical Statistics* 23, 193–212.
- Busschaert, P., Geeraerd, A.H., Uyttendaele, M., Van Impe, J.F., 2010. Estimating distributions out of qualitative and (semi)quantitative microbiological contamination data for use in risk assessment. *International Journal of Food Microbiology* 138, 260–269.
- Christensen, B.B., Rosenquist, H., Sommer, H.M., Nielsen, N.L., Fagt, S., Andersen, N.L., Norrung, B., 2005. A model of hygiene practices and consumption patterns in the consumer phase. *Risk Analysis* 25, 49–60.
- Codex alimentarius Commission, 1999. Principles and Guidelines for the Conduct of Microbiological Risk Assessment. FAO Edition, Rome. 6 Available at http://www.codexalimentarius.net/download/standards/357/CXG_030e.pdf. 1999.
- Cullen, A.C., Frey, H.C., 1999. Probabilistic techniques in Exposure assessment. Plenum Press, New York.
- Danyluk, M.D., Harris, L.J., Schaffner, D.W., 2006. Monte Carlo simulations assessing the risk of salmonellosis from consumption of almonds. *Journal of Food Protection* 69, 1594–1599.
- Delignette-Muller, M.L., 2009. Principles of Predictive Modeling. In: Toldra, F. (Ed.), *Safety of meat and processed meat*. Springer, New York, USA, pp. 535–558.
- Delignette-Muller, M.L., Cornu, M., 2008. Quantitative risk assessment for *Escherichia coli* O157:H7 in frozen ground beef patties consumed by young children in French households. *International Journal of Food Microbiology* 128, 158–164.
- Efron, B., Tibshirani, R.J., 1993. *An Introduction to the Bootstrap*. Chapman & Hall.
- European Commission, 2003. Risk Assessment of food Borne Bacterial Pathogens: Quantitative Methodology Relevant for Human Exposure Assessment. European Commission. 89 Available at http://europa.eu.int/comm/food/fs/sc/ssc/out308_en.pdf. 2003.
- FAO/WHO, 2004. Risk assessment of *Listeria monocytogenes* in ready to eat foods—Technical report, Microbiological Risk Assessment Series, no 5. Food and Agriculture Organization of the United Nations and World Health Organization, Rome. 269 Available at <http://www.fao.org/docrep/010/y5394e/y5394e00.htm>. 2004.
- FAO/WHO, 2006. Food safety risk analysis. A guide for national food safety authorities, FAO Food and Nutrition. FAO/WHO, Roma. 37–63 Available at http://www.fao.org/icatalog/search/dett.asp?aries_id=108051. 2006.
- FDA/USDA/CDC, 2003. Quantitative Assessment of Relative Risk to Public Health from Foodborne *Listeria monocytogenes* Among Selected Categories of Ready-to-Eat Foods [Online] Available at <http://www.foodsafety.gov/~dms/lmr2-toc.html>. 2003.
- Frey, H.C., 1992. Quantitative Analysis of Uncertainty and Variability in Environmental Policy Making. AAAS/U.S. EPA, Washington, DC. 78 Available at http://www4.ncsu.edu/~frey/reports/frey_92.pdf. 1992.
- Frey, H.C., Mokhtari, A., Zheng, J., 2004. Recommended Practice Regarding Selection, Application, and Interpretation of Sensitivity Analysis Method Applied to Food Safety Process Risk Models. North Carolina State University for Office of Risk Assessment and Cost-Benefit Analysis, U.S. Department of Agriculture, Washington, D.C. Available at www.ce.ncsu.edu/risk/. 2004.
- Haas, C.N., Rose, J.B., Gerba, C.P., 1999. *Quantitative Microbial Risk Assessment*. Wiley & Sons, New York.
- Helsel, D.R., 2005. *Nondetects and Data Analysis. Statistics for Censored Environmental Data*. Wiley-Interscience.
- Iman, R.L., Conover, W.J., 1982. A distribution-free approach to inducing rank correlation among input variables. *Communication in Statistics* B11, 311–334.
- Jang, H.K., Kim, J.Y., Lee, J.K., 2009. Radiological risk assessment for field radiography based on two dimensional Monte Carlo analysis. *Applied Radiation and Isotopes* 67, 1521–1525.
- Jones, R.M., Masago, Y., Bartrand, T., Haas, C.N., Nicas, M., Rose, J.B., 2009a. Characterizing the risk of infection from *Mycobacterium tuberculosis* in commercial passenger aircraft using quantitative microbial risk assessment. *Risk Analysis* 29, 355–365.
- Jones, S., Lacey, P., Walshe, T., 2009b. A dynamic hydrological Monte Carlo simulation model to inform decision-making at Lake Toolibin, Western Australia. *Journal of Environmental Management* 90, 1761–1769.
- Lorimer, M.F., Kiermeier, A., 2007. Analysing microbiological data: Tobit or not Tobit? *International Journal of Food Microbiology* 116, 313–318.
- Özkaynak, H., Frey, H.C., Burkea, J., Pinder, R.W., 2009. Analysis of coupled model uncertainties in source-to-dose modeling of human exposures to ambient air pollution: A PM_{2.5} case study. *Atmospheric Environment* 43, 1641–1649.
- Pouillot, R., Beaudou, P., Denis, J.-B., Derouin, F., AFSSA Cryptosporidium Study Group, 2004. A quantitative risk assessment of waterborne cryptosporidiosis in France using second-order Monte Carlo simulation. *Risk Analysis* 24, 1–17.

- Pouillot, R., Miconnet, N., Afchain, A.L., Delignette-Muller, M.L., Beaufort, A., Rosso, L., Denis, J.B., Cornu, M., 2007. Quantitative risk assessment of *Listeria monocytogenes* in French cold-smoked salmon: I. Quantitative exposure assessment. *Risk Analysis* 27, 683–700.
- Pouillot, R., Goulet, V., Delignette-Muller, M.L., Mahe, A., Cornu, M., 2009. Quantitative risk assessment of *Listeria monocytogenes* in French cold-smoked salmon: II. Risk Characterization. *Risk Analysis* 29, 806–819.
- R Development Core Team, 2009. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing Available at <http://www.R-project.org>. 2009.
- Ricci, V., 2005. Fitting distributions with R Available at <http://cran.r-project.org/doc/contrib/Ricci-distributions-en.pdf>. 2005.
- Rimbaud, L., Heraud, F., La Vieille, S., Leblanc, J.C., Crepet, A., 2010. Quantitative risk assessment relating to adventitious presence of allergens in food: a probabilistic model applied to peanut in chocolate. *Risk Analysis* 30, 7–19.
- Ross, T., Rasmussen, S., Fazil, A., Paoli, G., Summer, J., 2009. Quantitative risk assessment of *Listeria monocytogenes* in ready-to-eat meats in Australia. *International Journal of Food Microbiology* 131, 128–137.
- Venables, W.N., Ripley, B.D., 2002. *Modern Applied Statistics with S*, Fourth edition. Springer.
- Vicari, A.S., Mokhtari, A., Morales, R.A., Jaykus, L.A., Frey, H.C., Slenning, B.D., Cowen, P., 2007. Second-order modeling of variability and uncertainty in microbial hazard characterization. *Journal of Food Protection* 70, 363–372.
- Vose, D., 2000. *Risk Analysis, A quantitative guide*, 2nd ed. Wiley and Sons, Chichester.