

# How to elicit a cost function? Lessons of hope and disappointment from a diced bacon case-study

**Natalie Commeau**

INRA, UMR 518 Math. Info. Appli., F-75005 Paris, France

AgroParisTech, UMR 518 Math. Info. Appli., F-75005 Paris, France

**Eric Parent**

INRA, UMR 518 Math. Info. Appli., F-75005 Paris, France

AgroParisTech, UMR 518 Math. Info. Appli., F-75005 Paris, France

*Statistical decision theory provides an attractive framework to help choose decisions under uncertainty. Unfortunately, it does not seem to be often implemented for specific applications. In this paper, we rely on this theory to determine the optimal sampling plan for a plant producing diced bacon. Sampling plans are widely used in the food industry to assess the quality of products. After presenting the most common sampling plan in use, we develop a Bayesian reanalysis to interpret the common practice for sampling by attribute. Then, we turn to a more elaborate problem and propose a way to get the best plan by minimizing the expected cost a food plant could face. Although the cost function was designed to be easily understandable by manufacturers, we encountered difficulties in determining the correct costs through discussion with an expert. After correction, our alternative approach gives applicable results. We finally discuss what we learnt from this practical experience and give our thoughts on how cost elicitation could be improved and extended by discussing with more manufacturers.*

Keywords : *Optimal design, Cost function, Statistical decision theory, Bayesian statistics, Sampling plans*

## 1. Introduction

Statistical decision theory attempts to formalize many probabilistic engineering design questions into a unified theory of decision-making under uncertainty. This field of statistical decision initiated by [Bernoulli \(1738\)](#) was developed by the work of [Wald \(1950\)](#), searching for a common presentation of hypothesis testing, model choice and estimation techniques within the frequentist framework with a loss function. The field was further enriched by Bayesian statisticians such as [Savage \(1954\)](#); [Raiffa \(1968\)](#) and [Raiffa and Schlaifer](#)

(1961). Following these authors, decision-makers should transform the stakes of the competing decisions (including costs) into a so-called (multiattribute) *utility function* (the modern counterpart of Bernoulli's moral expectancy of the St. Petersburg paradox) and behave in risky situations as optimizers of the expected utility with their state of knowledge being quantified by means of a random variable. The rigorous formalization of the theory is built upon the five mathematical axioms given in [Pratt et al. \(1964\)](#). During the last 60 years, experiments on behavior under risk have exhibited a series of "paradoxes" ([Allais, 1953](#),

1979), mostly linked to discrepancies between observable rationality of a decision maker under risky situations (Piattelli-Palmarini, 1995) and the expected utility optimization principle (Machina, 1987). New models of behavior under uncertainty have been developed (Machina, 1982; Quiggin, 1964; Tvesky and Kahneman, 1992), trying to propose and explore other settings of decision under risk (Munier, 2008). However, many reasons strongly advocate keeping the old rationale of expected utility (Lindley, 1991, 2006), since:

1. It agrees with common actuarial practice, at least for private or public investments when the utility function is the mere addition of costs and benefits, eventually time discounted; (it makes sense for big companies to work with such a loss function as a practical first order simplification that neglects possible risk aversion for extreme stakes)
2. It guarantees an always positive value of information ( *i.e.*, on average, a better state of information can never worsen the decision in the sequential setting of statistical learning);
3. It works coherently within the Bayesian statistical framework (Berger, 1985), which allows for a predictive probabilistic understanding of what the future holds in store (Jordaan, 2005; Kadane, 2011; Pasanisi et al., 2012).

However coherent and simple this statistical decision setting might seem, at least two difficulties occur when trying to go from theory to practice: (i) how does one get the prior distribution describing the current knowledge for the unknown of the problem and (ii) which loss function should one implement? In the applied statistical literature, one can find several papers about the way to encode human expertise into prior distributions (Chaloner, 1996; Jaynes, 1968; Kass and Wasserman, 1996; Kadane et al., 1998; O'Hagan et al., 2006; Albert et al., 2012) but there are fewer articles about how to quantify the consequences of competing decisions (see for instance Krzysztofowicz and Duckstein, 1980; Farquhar, 1984; Abdellaoui et al., 2005). This paper focuses on the latter point and relates the authors' personal

experience of hope with a Bayesian reanalysis of the sampling plan by attribute but also some despair when turning to more elaborate designs and attempting to obtain a loss function. As a very simple but real case study, we study a diced bacon producing plant. Following the Bayesian guidelines, the cost function should reflect the consequences of the decisions based on the sampling plan under the range of possible sanitary conditions in the plant. The main issues about pathogens in such food and the current practice of sampling to monitor the production and to control the microbial conditions in the diced bacon process are explained in Section 2. In Section 3, we review the necessary ingredients for statistical decision theory with special emphasis on the construction of the decision rule, which offers a nice Bayesian reinterpretation of the classical sampling plan by attribute. In Section 4, we undertake the task of eliciting a cost function following this theory and illustrate the many difficulties encountered with the case study depicted in Section 2, however simple it may look at first sight. Section 5 illustrates the tricks to which we had recourse as a remedy to the previous silly results obtained when going bluntly from theory to practice. We obtain more sensible results that are compared to the common ordinary industrial practice. Section 6 contains some conclusions and a discussion on the perspectives opened by this work.

## 2. Context

### 2.1. Diced bacon process and the need for sampling plans

Diced bacon is a typical French product. It is made of pork breast which is tumbled with brine containing salt and different organic acids, then steamed at about 50°C, chilled at -15°C and diced before being packed under a modified atmosphere. The average consumption of a French family reaches 3 kg of diced bacon per year (see Table 5 of Legendre (2008)) thus placing it in the category of common foods. It is usually consumed cooked but one study showed that 14 % of consumers eat diced bacon raw (AFSSA, 2009); as a consequence, it can also be classified as a ready-to-eat product (RTE).

*L. monocytogenes* is a foodborne pathogen commonly found in many food-processing and agricultural environments. It is frequent in raw foods and can also be present in RTE foods due to post-processing contamination. It is not unusual to find *L. monocytogenes* in diced bacon, and temperature and pH conditions allow this pathogen to grow. However, the growth does not seem to be fast (Cornu et al., 2011). Consumption of *L. monocytogenes*-contaminated food may cause listeriosis. The invasive form of this infection particularly affects pregnant women, the elderly, and people with diseases such as cancer, diabetes, AIDS, etc. According to the Center for Disease Control and Prevention (CDC), *L. monocytogenes* is the third pathogen contributing to domestically acquired foodborne illnesses resulting in death (CDC, 2011). In France, the incidence is around 4.5 cases per million inhabitants and per year (InVS, 2010). For the food industry, the costs due to *L. monocytogenes* are mainly due to batch recalls (Jemmi and Stephan, 2006). A recall is an action in which the food business operator (FBO) asks the consumers to return the products, resulting in costs from the circulation of news and alarm calls, transport and destruction of the batch.

## 2.2. Sampling plans in food microbiology

Sampling plans consist of taking a sample composed of  $n$  sample units randomly drawn from a population of food items, usually at the end of the process. These units are analyzed and a decision is taken depending on the results. The food population must be homogeneous so that the results on the analyzed sample units give reliable information about the population from which they are taken. The food population is called a *batch*. Homogeneity of batches as defined by the FBO does not always occur in practice, when considering the microbial distributions of the various pieces of production (ICMSF, 2002). Yet in what follows, we make the assumption of a good match between microbial batches and production batches as a simplifying hypothesis.

A very widespread sampling plan is the two-class attribute sampling plan. In this kind of plan, the result given by the analysis is binary: the unit does or does not have a certain prop-

erty (here, the unit may or may not be contaminated with *L. monocytogenes*). The number of positive sample units  $y$  is counted. The two-class acceptance sampling plan is governed by two numbers:

- $n$ , the number of sample units taken for analysis and
- $c$  the maximum allowable number of sample units with positive results for the batch to be accepted (*i.e.* the batch is accepted if  $y \leq c$ ).

Of course, it is always possible to reject a good batch and to accept a poor one. To avoid such risks of misclassification, the FBO should analyze every unit of the whole batch, which does not make sense because the analyzed units are destroyed during the analysis. Would one know the prevalence, *i.e.* some *true* level  $\theta$  for the probability of contamination for a sampled unit, one could evaluate the probability of acceptance  $P_\theta(Y \leq c)$  of the decision rule. Figure 1 shows the curve  $(\theta, P_\theta(Y \leq c))$  known as the operating curve. Indeed, two well-chosen points  $(\theta_0, \alpha = P_{\theta_0}(Y \leq c))$  and  $(\theta_1, 1 - \beta = P_{\theta_1}(Y \leq c))$  are enough to uniquely define the operating curve. The four quantities  $\theta_0, \alpha, \theta_1, \beta$  are to be interpreted as:

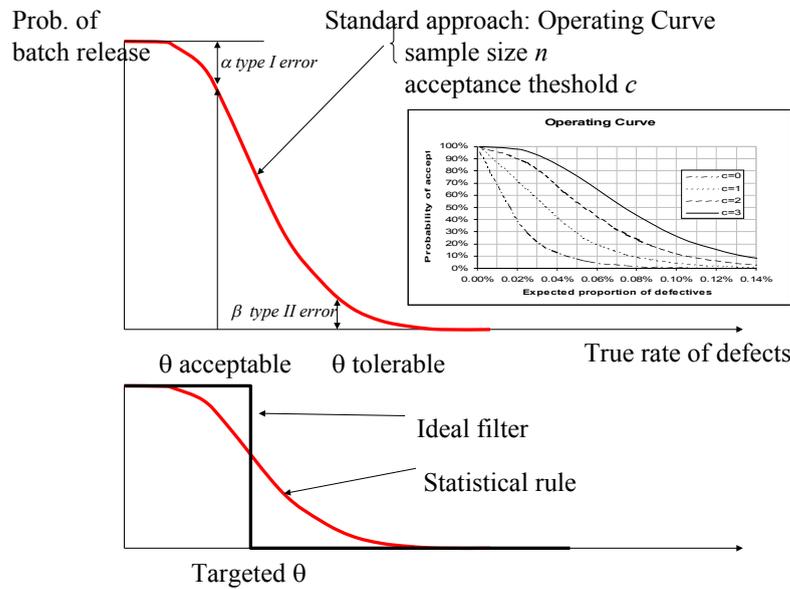
- $\alpha$  is the producer's risk, also known as type I error: the chance of rejecting a good lot that contains defectives equal or less than a threshold  $c$  corresponding to some acceptable quality level  $\theta_0$ , named AQL in quality managers' jargon.
- The type II error or customer's risk  $\beta$  is the chance of accepting a bad lot that contains more defectives than the largest proportion of defectives  $\theta_1$ , the rejecting quality level (RQL) that a consumer is willing to accept.

In theory,  $\theta_0, \alpha, \theta_1, \beta$  should be product- and manufacture-specific. In practice, they also derive from the quest of a compromise between a small value of  $n$  (*i.e.*, small sampling costs) with high risks of errors and a large value of  $n$  but a low risk of errors.

The sampling plan of the diced bacon plant under study should monitor the production so as

to make sure that the product is correct for sale. The diced bacon is sold to a client (here, the retailer), which in turn resells it to consumers. In this research, the standard two-class attribute sampling plan cannot be applied because of the delay between bacteriological controls and production releases. The FBO takes samples several times a month but he could hardly re-

ject a month's production, all the more so as the products are already in the supermarket or in the consumers' stomachs. The FBO analyzes sampled units but does not know how many units per sample to take. In this article, we rely on statistical decision theory to find the most appropriate value for  $n$ .



**Figure 1:** A two-class attribute sampling plan is defined by  $n$ , the sample size and  $c$ , the maximum allowable number of defective units. Equivalently, the errors of type I and II (resp  $\alpha$  and  $\beta$ ) corresponding to the Acceptable Quality Level  $\theta_0$  and the Rejecting Quality Limit  $\theta_1$  provide two points that suffice to plot the probability of accepting the lot for any two-class attribute sampling plan. Such an operating curve is to be compared with the steep one-zero function corresponding to the ideal filter (bottom panel).

We further define a batch as the period of production (e.g. a month) during which a sample of  $n$  units is taken and analyzed. Once this production period is over and the bacteriological controls known, the plant manager takes a decision with respect to the plant working conditions, and the client may renegotiate his contract with the FBO. We take the point of view of the FBO and make some additional simplification to depict the retailer's behavior.

### 3. Statistical decision theory

Decision theory under uncertainty (Jordaan, 2005) is a branch of Bayesian statistics (Berger, 1985; Kadane, 2011) dealing with decision-making (here, choosing a sampling plan) un-

der imperfect knowledge and partial information. In this section we give the general setting and some illustration from the two-class sampling plan by attribute.

#### 3.1. General framework with illustration

Statistical decision theory requires the following ingredients (Ulmo and Bernier, 1973):

- The definition of the set  $\Theta$  of all unknown states of Nature. Here,  $\Theta$  stands for the set of all possible values  $\theta$  for the "true" microbial level, for instance a prevalence value or some contamination concentration, which will remain unknown throughout the batch production. It is assumed that some prior knowledge

is available about the unknown  $\theta$  under the form of a probabilistic bet, encoded by the probability distribution function  $[\theta]$ . Following [Gelfand and Smith \(1990\)](#), we use the bracket notation for pdfs in what follows. In practice, the prior pdf  $[\theta]$  may be obtained through an expert's interview as in [O'Hagan et al. \(2006\)](#) or from some historical records of estimations of  $\theta$  in similar working conditions as in [Berger \(1985\)](#). In the acceptance sampling plan by attribute,  $\theta$  is defined as the proportion of defective items in the lot. A similar definition is taken in the diced bacon case study (*i.e.*, the prevalence). Because the support of  $\theta$  is the unit interval, it is most convenient to opt for a beta distribution with tuning coefficients  $a$  and  $b$  to be understood as some virtual number of a priori successes and failures:

$$[\theta|a, b] = \theta^{a-1}(1-\theta)^{b-1} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}. \quad (1)$$

- In a sequential setting, potential actions are composed of two pieces under the form  $(e, d)$  with  $e \in \mathcal{E}$  being the experimental device to be first selected before decision  $d \in \mathcal{D}$  is chosen at some terminal stage. Such terminal decision can be often taken in the light of new information, *i.e.*, collected after the experimental condition  $e$  has been set. For instance in the case of the two-class sampling plan by attribute, the experimental condition  $e$  can be specified as “a sample of size  $n$  is drawn at random without replacement from the batch”, while the second component  $d$  takes either the value 1 for “recall” or 0 for “let go”.
- The definition of the set  $\mathcal{Y}_e$  of observable events under experimental conditions  $e$ . A likelihood function of  $\theta$  is available under the pdf  $[y|\theta, e]$  for the event  $Y = y$ . In our case, a typical event will be the observation  $y$  of a random sample  $Y$  from a batch with contamination  $\theta$  under experiment  $e$  (referring to the sample size  $n$ ). In the two-class attribute sampling plan, one typically assumes that  $y$ , the observed number of defects, is sampled from a sufficiently large production of items so that a binomial distribution of

order  $n$  and probability  $\theta$  can be taken for  $[y|\theta, n]$ :

$$[y|\theta, n] = \theta^y(1-\theta)^{n-y} \frac{\Gamma(n+1)}{\Gamma(y+1)\Gamma(n-y+1)}. \quad (2)$$

- The definition of an evaluation criterion, depending on both the decision made (*i.e.* the chosen design plan) and the (unknown) state of nature. This criterion is called *loss function*, noted  $L$ , defined on  $\mathcal{E} \times \mathcal{D} \times \Theta \times \mathcal{Y}$  and taking values in  $\mathbb{R}$ . Heuristically, this loss function evaluates the consequence of picking the composite decision  $(e, d)$  while the state of nature is  $\theta$  and while data  $y$  is recorded. For instance in the acceptance sampling by attribute, a simple and prototypal loss function would be independent of the observed results  $y$  under the form:

$$L(n, d, \theta, y) = k \times n + C\theta \times (1-d) + d \quad (3)$$

with  $k$  the cost of controlling one item,  $C \times \theta$  the cost of releasing ( $d = 0$ ) a fraction  $\theta$  of the lot with defective items, while expressing all costs relatively to the value of the wasted production when  $d = 1$  (*i.e.*, the loss that corresponds to not delivering the goods is 1). More generally, assuming sufficient regularity conditions so that the function  $(e, d) \mapsto L(e, d, \theta, y)$  exhibits a single minimum  $(e^*(\theta, y, d), d^*(\theta, y))$  for all  $(\theta, y) \in \Theta \times \mathcal{Y}$ , it is often easier to work with the opportunity loss:

$$L(e, d, \theta, y) - L(e^*(\theta, y, d^*), d^*(\theta, y), \theta, y).$$

In more interpretable words, the opportunity loss function evaluates the cost of knowing neither  $\theta$  nor  $y$  when deciding to enforce the composite action  $(e, d)$ .

The optimal choice according to Bayesian decision theory is the composite action  $(\widehat{e}, \widehat{d})$  minimizing the expected loss (or expected opportunity loss):

$$(\widehat{e}, \widehat{d}) = \arg \min_{(e, d)} \int \int_{\mathcal{Y}} L(e, d, \theta, y) [y|\theta, e] [\theta] dy d\theta.$$

The decision tree ([Raiffa, 1968](#)) of Figure 2 better puts forward the sequential nature of the

previous mathematical setting: decision  $d$  is to be taken once information  $y$  has been observed. The role of the statistician might thus be understood as:

1. Consider an experimental design  $e$ , (for instance in the acceptance sampling plan, take a given sample size  $n$ )
2. Given such experiment  $e$ , evaluate the decision rule  $\hat{\delta}_e$  which makes the best map between the random output  $Y$  that will occur as an observation of experiment  $e$  and the decision to be taken  $d$ . When the sets  $\mathcal{D}$  and  $\Theta$  are identical, such a correspondance is known as an *estimator*.

$$\hat{\delta}_e = \arg \min_{\delta} W(e, \delta)$$

$$W(e, \delta) = \int_y \int_{\theta} L(e, \delta(y), \theta, y) [y|\theta, e] [\theta] dy d\theta$$

3. Find (upon eventual requirement) the *optimal* design  $\hat{e}$  as a solution minimizing

the Bayes risk  $W(e, \hat{\delta}_e)$ . For the acceptance sampling plan, one finds:

$$W(n, \hat{\delta}_n) = \sum_{y=0}^n \text{Min}(1, C \frac{a+y}{n+a+b}) [y]$$

$$\hat{n} = \arg \min_n \{k \times n + W(n, \hat{\delta}_n)\}. \quad (4)$$

Note that, since  $[y|\theta, e] \times [\theta] = [\theta|y, e] \times [y|e]$ , once the experimental setting  $e$  is picked, the construction of the optimal decision rule  $\hat{\delta}_e$  for the Bayesian risk  $W(e, \delta)$  is a simpler one-step problem. Given observation  $y$ , one simply seeks the minimum of Bayesian posterior expected loss. As a consequence,  $\hat{\delta}_e$  is such that:

$$y \mapsto \hat{\delta}_e(y) = \hat{d} = \arg \min_d \int_{\theta} L(e, d, \theta, y) [\theta|y, e] d\theta.$$

Coming back to the acceptance sampling plan example, conjugate properties of the beta prior with respect to the binomial likelihood yields a beta posterior and a Polya predictive:

$$[\theta|y, n, a, b] = \theta^{y+a-1} (1-\theta)^{n-y+b-1} \frac{\Gamma(n+a+b)}{\Gamma(a+y)\Gamma(n-y+b)}$$

$$[y] = \frac{\Gamma(n+1)\Gamma(a+b)\Gamma(a+y)\Gamma(b+n-y)}{\Gamma(y+1)\Gamma(n-y+1)\Gamma(a)\Gamma(b)\Gamma(a+b+n)}. \quad (5)$$

Consequently,  $\hat{\delta}_n$  will in this case be such that:

$$\hat{\delta}_n(y) = \arg \min_d \int_{\theta=0}^1 (C\theta \times (1-d) + (d)) [\theta|y, n, a, b] d\theta$$

$$\hat{\delta}_n(y) = \arg \min_d \{(1-d) \times C \times \mathbb{E}_y(\theta) + d\}$$

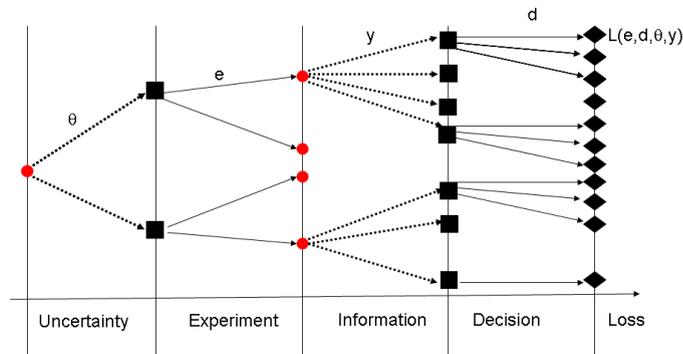
$$\hat{\delta}_n(y) = \begin{cases} 1 & \text{if } y > c \\ 0 & \text{if } y \leq c \end{cases}$$

with, because  $\mathbb{E}_y(\theta) = \frac{a+y}{a+b+n}$ ,

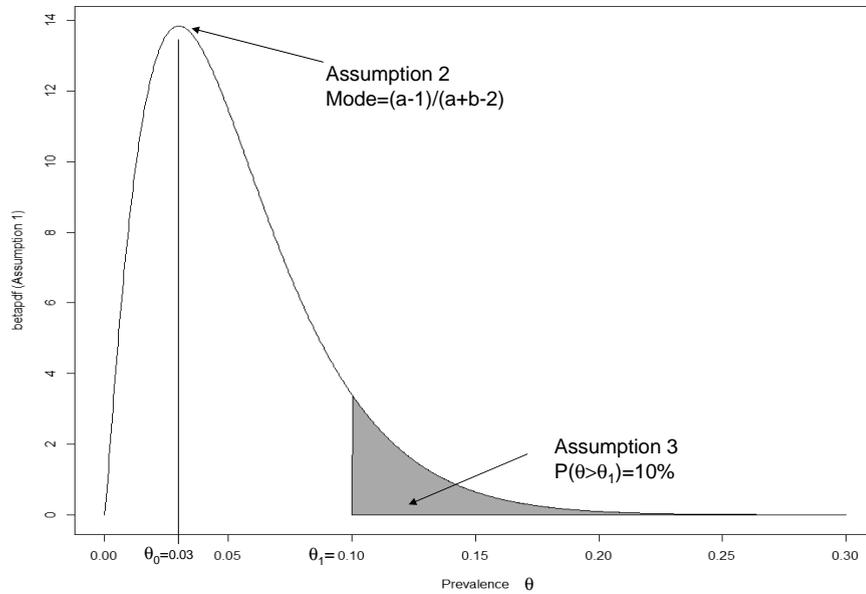
$$c = \lfloor \frac{n+a+b}{C} - a \rfloor. \quad (6)$$

Equation (6) offers a coherent interpretation: the control threshold  $c$  decreases with coefficient  $a$  (a small  $a$  corresponds to a low prior level of prevalence). The severity is also re-

inforced for higher values of  $C$  i.e., when important consequences stem from uncontrolled release of defective items.



**Figure 2:** Decision tree for the sequential setting of the optimization program: the stochastic nodes are indicated by circles while deterministic nodes are squares.



**Figure 3:** Relying on a beta distribution (with  $a = 2.15$  and  $b = 38.18$ ) to interpret AQL and RQL. With such values, Assumptions 2 and 3 are fulfilled.

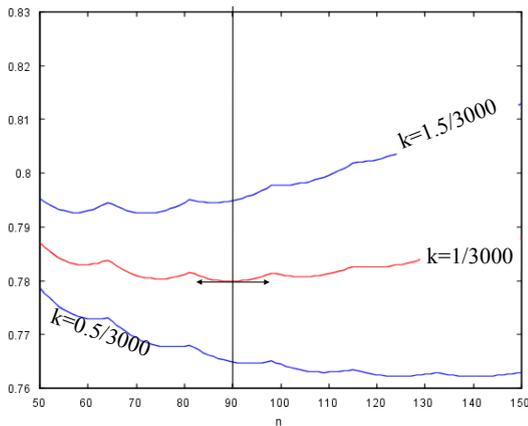
3.2. A Bayesian reanalysis of the two-class sampling plan

FBOs may have conceptual difficulties interpreting type I and II errors so as to set these errors to meaningful values: indeed, they must imagine conceptual repetitions of situations when prevalence  $\theta$  is known to be either the AQL  $\theta_0$  or the RQL  $\theta_1$ , whereas in operational situations only the data  $y$  are given. Therefore in the spirit of p-values, *ad hoc* values such as 1, 5 or 10% are often recommended with no deep understanding of the role of such errors. Consider for instance the sampling plan  $\theta_0 = 0.03$  for  $\alpha = 5\%$  and  $\theta_1 = 0.1$  for  $\beta = 10\%$ , which

corresponds to a standard scheme of inspection by attribute with  $n = 90$  and  $c = 5$ . A Bayesian reanalysis may help to interpret such a plan. We need five assumptions.

1. As in Figure 3, a beta distribution with coefficients  $a$  and  $b$  to be tuned as in Equation (1) is used to encode a prior probabilistic judgement on the level of prevalence  $\theta$ .
2. The AQL  $\theta_0$  is interpreted as the mode of this distribution, i.e., the best guess of prevalence in a routine operation.  $\theta_0 = \frac{a-1}{a+b-2}$ .

3. The RQL  $\theta_1$  could be, in light of misinterpretations of (frequentist) confidence intervals pointed out by [Lecoutre \(2006\)](#), the 90% quantile of this distribution of (Bayesian) credible values.  $a = 2.15$  and  $b = 38.18$  met this requirement and the previous one. One can easily check that the mode of a  $beta(2.15, 38.18)$  distribution is  $\theta_0 = 0.03$  and that such a distribution exceeds the value  $\theta_1 = 0.1$  with 10% of its mass.
4. Adopting the loss function given by Equation (3) and the binomial likelihood, the sampling threshold derived from a Bayesian decision analysis is given by Equation (6). Because of the rounding to the nearest larger integer, this equation is verified with  $C = 17$ , to be taken as the cost linked to unstopped defective items (indeed, with  $n = 90$ ,  $c = 5$ ,  $a = 2.15$  and  $b = 38.18$ , any value of  $C$  between 16 and 18 would still be appropriate).
5. The analysis is further developed to find which sampling cost  $k$  would match  $n = 90$  as an optimal solution for Equation (4). Figure 4 illustrates that  $k = 1/3000$  provides a rounded quasi-optimum solution.



**Figure 4:** Trials and errors when searching for the sampling cost  $k$  that minimizes  $k \times n + W(n, \hat{\delta}_n)$ , where  $n = 90$ , and  $W(90, \hat{\delta}_{90}) = \sum_{y=0}^{90} \min(1, 17 \frac{2.15+y}{90+2.15+38.18} [y])$ .

### 3.3. Going one step further

The previous reanalysis provides an appealing interpretation of the two class sampling

plan. To sum it up, a decision-maker behaving according to statistical decision theory would opt for  $n = 90$  and  $c = 5$  if:

- his prior knowledge about the prevalence  $\theta$  can be expressed by the probabilistic judgement with betting odds shown by Figure 3,
- when unduly released, defective items induced losses up to 17 times their own values,
- the amount of effort necessary to inspect 3000 items would hamper the entire benefits of the whole production.

Most managers met by the authors found the previous decisional interpretation of the two-class sampling plan by attribute at least interesting, at best exciting. Despite their long practice of ISO and ANSI standards and their knowledge of Codex Alimentarius recommendations, many are more inclined to set their own specific values to  $a, b, C$  and  $k$  rather than answer questions about  $\theta_0, \theta_1, \alpha$  and  $\beta$  for implementation of sampling rules ([ICMSF, 1974, 2002](#)). Is such a hope in statistical decision theory justified for practical purposes? As depicted in what follows, turning to a slightly more sophisticated three-class sampling plan, we went from hope to disappointment.

## 4. How to build an appropriate loss function?

### 4.1. First attempt

Often, standard loss functions are chosen, for example when there is no information available to build a dedicated function. At the top rank of the *ad hoc* loss functions ([Berger, 1985](#)), one finds the unavoidable squared-error loss and the 0-1 loss. In this work, we choose to build an appropriate loss function through elicitation. We do it in order to measure, on a monetary scale, the costs that a plant has to bear. According to statistical decision theory, one needs to find the values of  $L$  for every couple (state of Nature, decision). Let us first define the set of possible decisions and the set of the states of Nature. Because the result of the analysis the FBO carries out on his product is binary, the state of Nature of a batch can be taken to be the prevalence  $\theta$ . We turned to

an expert to define the set of decisions. This expert works in the field of the pork meat industry. We were not able to find any FBO willing to take time to answer our questions. However we kept working on with hope: even though the expert we had knew less about the decisions and the costs of a specific plant, he had nevertheless a broader view of what happens in the diced bacon industry in France than a FBO.

According to the expert, a good way to model all the possible decisions a plant can take is to define three different decisions:

- $d_0$ : contamination is under control, no correction needed;
- $d_1$ : contamination is a little too high, a slight correction is needed;
- $d_2$ : contamination is too high, a big correction is needed.

To make it simple, we divide the prevalence accordingly into three classes. The prevalence is considered:

- low if  $\theta \in \Theta_0 = [0; \theta_0]$  ;
- medium if  $\theta \in \Theta_1 = ]\theta_0; \theta_1]$  ;
- high if  $\theta \in \Theta_2 = ]\theta_1; 1]$ ,

where  $\theta_0 < \theta_1$ ,  $\bigcup_{i=0}^2 \Theta_i = [0; 1]$  and  $\bigcap_{i=0}^2 \Theta_i = \emptyset$ . This can be understood as binning the beta distribution of Equation (1) into a three-category ordinal variable. With the set of decisions and the set of states of Nature in hand, we can now determine losses of  $L(n, d, \theta)$  for all the nine cases (here, the experimental condition  $e$  is equal to the size  $n$  of the sampling plan). First, for all the cases, the FBO has to pay for microbiological analyses. One analysis costs  $K \text{ €}$ . This cost is not mentioned below in this section to simplify the presentation. The way we decided to model the costs a plant faces is represented in Table 1. Each decision has a cost: 0 for  $d_0$ ,  $C_1$  for  $d_1$  and  $C_2$  for  $d_2$ . We made the assumption that the client knows in which class the prevalence of the defaults in the production belongs. When the prevalence  $\theta$  is not in  $\Theta_0$  and if the FBO has taken decision  $d_0$ , then the client imposes a fine on the FBO:  $pb$  if  $\theta \in \Theta_1$  and  $PB$  if  $\theta \in \Theta_2$  (see Table 1).

**Table 1:** Values taken by the loss function  $L$  depending on the decision  $d$  taken by the plant and the prevalence  $\theta$  of the production period.

	Decisions		
$\theta$	$d_0$	$d_1$	$d_2$
$\theta \in \Theta_0$	0	$C_1$	$C_2$
$\theta \in \Theta_1$	$pb$		
$\theta \in \Theta_2$	$PB$		

The expert is asked to define precisely the corrective actions and the fines and to give costs to all of them.

**Slight correction (cost  $C_1$ )** The expert divides this correction into two sub-corrections:

1. a reinforced cleanup during a week;
2. an alert of the suppliers of pork breasts by carrying out analyses on 20 of their batches.

The sum of the costs is assessed at  $C_1 = 4,250 \text{ €}$ .

**Large correction (cost  $C_2$ )** As for the slight correction, the expert divides this correction into two sub-corrections:

1. the plant is closed for 24 h so the FBO loses a day's production;
2. the cleanup is more reinforced than for the slight correction because devices are taken apart and cleaned.

The total cost is assessed at  $C_2 = 14,000 \text{ €}$ .

**Fine if  $\theta \in \Theta_1$**  If the client detects that the prevalence of the production is medium and if the FBO has taken decision  $d_0$ , then the expert describes the fine as follows:

1. a penalty because specifications were not respected;
2. 10 additional analyses on the products of the FBO.

The total cost is assessed at  $pb = 4,200 \text{ €}$ .

**Fine if  $\theta \in \Theta_2$**  If the client detects that the prevalence of the production is high and if the FBO has taken decision  $d_0$ , then the expert describes the fine as follows:

1. a penalty because specifications were not respected;
2. a recall of the batch found positive by the client (for instance, a batch is the production of diced bacon done over a day);
3. additional analyses during a month for each batch (5 sample units per batch);
4. an audit.

The total cost is assessed at  $PB = 90,000 \text{ €}$ .

How did we fill the last four boxes of Table 1? The costs  $C_1$  and  $C_2$  remain the same whatever the prevalence of the production. On the contrary, the fines decrease when the FBO has taken corrections: the bigger the correction, the smaller the fine because we considered that, although the prevalence is not low, the FBO discovered the problem and took corrections. For instance, when the prevalence is medium and the decision taken by the FBO is  $d_1$ , the FBO has to pay the penalty but he only makes 5 additional analyses instead of 10 so the cost of the fine is 4,100 €. Finally, the costs are presented on Table 2. The cost of  $n$  microbiological analyses is added to all 9 cases since the FBO samples and analyzes the production over every period. The cost of one analysis is denoted  $K$  and was assessed at 20 € by the expert (to be added to the costs mentioned in Table 2).

**Table 2:** Values of the loss function  $L$  (without the cost of analysis).

Decisions			
$\theta$	$d_0$	$d_1$	$d_2$
$\theta \in \Theta_0$	0	4,250	14,000
$\theta \in \Theta_1$	4,200	8,350	18,000
$\theta \in \Theta_2$	90,000	93,300	99,000

When looking at Table 2, we notice that the cheapest decision to take whatever the prevalence is decision  $d_0$  (i.e. do nothing). In this situation, there is no need for the FBO to sample as he does not have to choose between several decisions.

If the FBO knew the prevalence of his production (i.e. if he knew the value of  $\theta$ ), the optimal decision rule to take would be decision  $d_i$  when  $\theta \in \Theta_i, i = 0, 1, 2$ , because when the prevalence is not low, the FBO wants to take corrective actions to lower the prevalence. With the costs assessed by the expert, the FBO has negative opportunity loss (Table 3) that is to say that a sub-optimal decision is cheaper than the optimal one!

**Table 3:** Values of the opportunity loss function (without the cost of analysis).

Decisions			
$\theta$	$d_0$	$d_1$	$d_2$
$\theta \in \Theta_0$	0	4,250	14,000
$\theta \in \Theta_1$	-4,150	0	9,650
$\theta \in \Theta_2$	-9,000	-5,750	0

In the assessment made by the expert, the decrease in the fines is smaller than the marginal increase in the decision costs. For example, when the FBO makes a slight correction instead of doing nothing, the cost increases from 0 to  $C_1 = 4,250 \text{ €}$ . If the prevalence is medium, the fine only decreases from 4,200 to 4,100 €.

We also explain this unsatisfactory result by the fact that it was difficult for the expert to assess the costs because he never thought this way. In addition, he may be reluctant to consider disastrous situations of contamination. As we wanted to avoid cases where the sample size is not trivially determined, we looked for ways to constrain the opportunity loss to be positive.

#### 4.2. A more sensible way to build a loss function

Let us keep the first line and the first column of Table 2 that correspond to situations easy to think of. We do not consider as reliable the four remaining left boxes. To maintain coherence, we decide the fine would be lowered by

- $1 - \alpha$  % when the FBO has taken decision  $d_1$  compared to decision  $d_0$ ;
- $1 - \beta$  % when the FBO has taken decision  $d_2$  compared to decision  $d_0$ ,

with marginal effects such that  $0 < \beta < \alpha$  (see Table 4) in order to keep positive opportunity losses.

**Table 4:** Values of the opportunity loss function (without the cost of analysis).

Decisions $\theta$	$d_0$	$d_1$	$d_2$
$\theta \in \Theta_0$	0	4,250	14,000
$\theta \in \Theta_1$	4,200	$4,250 + \alpha 4,200$	$14,000 + \beta 4,200$
$\theta \in \Theta_2$	90,000	$4,250 + \alpha 90,000$	$14,000 + \beta 90,000$

As a consequence, when decision  $d_1$  has been taken, the overall cost when prevalence is in  $\Theta_1$  is lower than when prevalence is in  $\Theta_0$ . However, 4,200 remain lower than  $4,250 + \alpha 4,200$ . To repair this discrepancy in the initial assessment of the fine, we corrected the cost of the fine  $pb$  from 4,200 to 6,200 €. This intends to make sure that the penalty the client charges the FBO is higher when prevalence is high than when the prevalence is medium. In agreement with the expert's reanalysis,  $\alpha$  has been fixed at 0.3 and  $\beta$  at 0.15 so that the optimal decision is always the cheapest one under perfect information about  $\theta$ . Table 5 describes the final costs of the loss function, which were approved after feedback with the expert.

**Table 5:** Values of the loss function  $L$  (the cost of analysis is not included).

Decisions $\theta$	$d_0$	$d_1$	$d_2$
$\theta \in \Theta_0$	0	4,250	14,000
$\theta \in \Theta_1$	6,200	6,100	14,900
$\theta \in \Theta_2$	90,000	31,300	27,500

## 5. The best sample size for a three-way cost table

In order to calculate the Bayesian risk, we need to determine:

- the prior distribution of prevalence  $\theta$ ;
- the distribution of the observations.

As in Section 3, we chose a Beta distribution with parameters  $a$  and  $b$  given by Equation (1) for the prior of the prevalence. For the distribution of the number of positive observations  $y$  in the sample knowing  $\theta$ , we kept the binomial distribution with parameters  $n$  and  $\theta$

(see Equation (2)). In that case, we recall that the posterior distribution of  $\theta$  is also a Beta distribution:  $[\theta|y] = \mathcal{Be}(a + y; b + n - y)$  and the marginal distribution of  $y$  is straightforward to calculate (see Equation (5)). To carry out this study, we merely chose  $a = 2$  and  $b = 3$  for the prior distribution of  $\theta$ . With such a prior, we cautiously assumes a very vague state of knowledge for some expert, equivalent to a mind experiment with  $a = 2$  defective units as outcomes from a virtual sample of size only  $a + b = 5$ . As the FBO does not know the prevalence, he takes a decision depending on the value of  $\int L(n, \delta_n(y), \theta)[\theta|y, n]d\theta$ , where  $\delta_n$  is the decision rule. We added the index  $n$  because the rule depends on the sample size (decision step previously  $e$  in the general framework given in Section 3). Decision  $d_0$  is taken when

$$\int L(n, d_0, \theta)[\theta|y, n]d\theta \leq \int L(n, d_1, \theta)[\theta|y, n]d\theta.$$

In fact, the optimal decision rule is generally as follows:

$$\begin{cases} \hat{\delta}_n(y) = d_0 & \Leftrightarrow y \leq c_1 \\ \hat{\delta}_n(y) = d_1 & \Leftrightarrow c_1 < y \leq c_2 \\ \hat{\delta}_n(y) = d_2 & \Leftrightarrow y > c_2, \end{cases}$$

where  $c_1$  and  $c_2$  are two thresholds with  $c_1 < c_2$ . But there are special cases when neither threshold exists. For instance, if decision  $d_2$  is never taken when  $y$  increases from 0 to  $n$ , threshold  $c_2$  is not defined. The values of  $c_1$  and  $c_2$  are determined for each value of  $n$ . Still, we do not mention the dependence in order to simplify notations. If  $c_1$  exists, it is the value for which  $L(n, \delta(c_1) = d_0, \theta) \leq L(n, \delta(c_1) = d_1, \theta)$  and  $L(n, \delta_n(c_1 + 1) = d_0, \theta) > L(n, \delta_n(c_1 + 1) = d_1, \theta)$ . Finding the value of  $c_2$  is performed similarly.

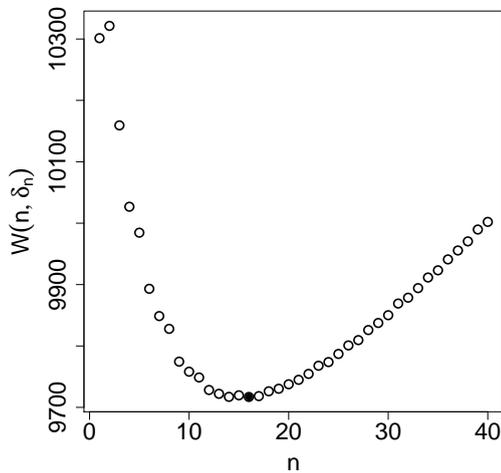
Now, we can calculate the Bayesian risk:

$$\begin{aligned}
 W(n, \delta_n) &= Kn + \int \left( \int L(n, \hat{\delta}_n(y), \theta) [\theta|y, n] d\theta \right) [y] dy \\
 &= Kn + \int \left( 1_{\hat{\delta}_n(y)=d_0} (6,200P_1 + 90,000P_2) + \int 1_{\hat{\delta}_n(y)=d_1} (4,250P_0 + 6,100P_1 + 31,300P_2) \right. \\
 &\quad \left. + \int 1_{\hat{\delta}_n(y)=d_2} (14,000P_0 + 14,900P_1 + 27,500P_2) \right) [y] dy \\
 &= Kn + \sum_{y=0}^{c_1} (6,200P_1 + 90,000P_2) P(Y = y) + \sum_{y=c_1+1}^{c_2} (4,250P_0 + 6,100P_1 + 31,300P_2) P(Y = y) \\
 &\quad + \sum_{y=c_2+1}^n (14,000P_0 + 14,900P_1 + 27,500P_2) P(Y = y), \tag{7}
 \end{aligned}$$

where  $P_i = P(\theta \in \Theta_i|y)$ ,  $i = 0, 1, 2$ ,  $K$  is the cost of a microbiological analysis. Remember that it was set at 20 € by the expert. In agreement with him, we set the classes of  $\theta$  as follows:

- low prevalence:  $\Theta_0=[0;0.2]$ ;
- medium prevalence:  $\Theta_1=[0.2;0.6]$ ;
- high prevalence:  $\Theta_2=[0.6;1]$ .

The formula in Equation (7) only depends on  $n$ . It is possible to determine the optimal sample size  $n$  for which the Bayesian risk stands at the lowest. Figure 5 shows the values of the Bayesian risk  $y$  for different values of  $n$ . The minimum is obtained for  $n = 16$ , with  $c_1 = 4$  and  $c_2 = 11$ .



**Figure 5:** Different Bayesian risk values  $W(n, \delta_n)$  when  $n$  increases from 1 to 50. The minimum is reached for  $n = 16$ , with  $c_1 = 4$  and  $c_2 = 11$ .

## 6. Discussion

Building an appropriate loss function is not an easy task. We chose to build a monetary function so that we can compare different consequences on the same scale: the decisions taken by the FBO and compliance with the client specifications.

Of course, the loss function was simplified in order to be able to obtain it through discussion with the expert. First, we reduced the number of decisions to only three and we only kept the two essential modes of correction that FBOs are used to. For instance, a slight correction is always made in a reinforced cleanup during a week and an alert of the suppliers of pork breasts to carry out analyses on 20 of their batches. Yet in operational practice, the FBO will slightly adjust the correction to the variety of problems he may encounter. Second, we divided the prevalence into three categories (low, medium and high). Had we attempted to build a function with costs depending on the prevalence  $\theta$  in a continuous way, this would have been a waste of time and an inappropriate approach. The client and the FBO cannot assess the prevalence precisely and even experienced experts are not able to consistently give costs for a large collection of prevalence values.

How should the elicitation of loss function be performed? Loss elicitation can be done directly through experts (Gorton et al., 1997), through population surveys and inquiries (Teisl and Roe, 2010) or through specialized magazines or journals (van der Gaag et al., 2004). Maybe we should have spent more time training the expert in probabilistic judgments

even though he was fully educated in statistics. Works about elicitation by experts can be found in O'Hagan et al. (2006); Kuhnert et al. (2010). Here, we only tried to incorporate his successive feedback into the cost function within the limited amount of time devoted to the study.

When figuring out a sampling plan, the brains of the FBO and of the statistical analyst certainly do not activate the same neuronal connections. Understandably, FBOs feel more at ease with a reassuring normative approach than within a prescriptive perspective. The FBO makes plans because it is compulsory, legally speaking (e.g. see Regulation (CE) No 2073/2005 (2005)). Sampling plans can also be a task set in the stone of the client's specifications. To our knowledge, sampling plans have never been designed to lower the expected costs borne by the FBO, at least in this food industry.

In this work, we made the assumption that the client knows the prevalence of each production period. Of course, the client lives in the very same world of uncertainty as the FBO: he can only assess the prevalence through microbiological testing. Additional modeling of the client's uncertainty is not difficult but we do not present it in this paper because we wanted here to describe how we built a loss function keeping things as simple as possible.

Even if the exercise seems difficult, it is worthwhile putting effort into the modeling of a loss function. As an essential component of the statistical decision theory ingredients, there are no *ad hoc* recipes, every case is specific and results differ greatly according to which loss function is considered.

## 7. Conclusion

This article highlights the following concluding points:

- conversely to the standard theory of optimal design working with general *ad hoc* mathematical functions such as Kullback-Leibler divergence or errors of type I, II, etc., Bayesian decision theory can consider plant-specific objectives: obviously the cost function for diced bacon, a rather ordinary food, will be different from the cost function for cold smoked salmon,

quickly alterable and mostly consumed at special events.

- Since monetary costs are an essential concept for decision makers, plant specific cost functions such as the ones proposed in equations (3) or (7) are readily understandable by FBOs. At the same time, writing the objective under a quantitative form favors communication and discussions among the various members of the often multidisciplinary team in charge of quality improvement. Additionally, such formulation may raise questions that FBOs might not have thought about otherwise: the first version of the perceived loss function given in Section 4, Table 3 was definitely not coherent with the decision process and had to be improved.
- However, the function to optimize for the diced bacon sampling plan was rather flat around the minimum. If this situation is a general feature when searching for optimum Bayesian decisions in practice, such a lack of sensitivity may help in understanding why discussion between mathematicians and FBOs is bound to remain long and uneasy.
- The authors cannot deny the many difficulties encountered when eliciting a cost function. Nevertheless, for the case study relying on the coherent improvement detailed in Table 5, we get to a recommended sampling rule with  $n = 16$ ,  $c_1 = 4$  and  $c_2 = 11$ . With regards to the stakes of the production, this change from today's rule ( $n = 1, c = 0$ ) looks quite acceptable and even on the more cautious side.

## Acknowledgements

We thank Pascal Garry who accepted to assess different costs supported by the FBOs in various circumstances. This work was supported by a grant from the Agence Nationale de la Recherche (ANR) (France), as a part of the Quant'HACCP project.

## References

- Abdellaoui, M., Vossman, F., and Weber, M. (2005). Choice-based elicitation and decomposition of decision weights for gains and losses under uncertainty. *Management Science*, 51(9):1384–1399.
- AFSSA (2009). Enquête individuelle et nationale sur les consommations alimentaires 2 (INCA 2). <http://www.afssa.fr/Documents/PASER-Ra-INCA2.pdf>, last accessed on September 4, 2014.
- Albert, I., Donnet, S., Guihenneuc-Jouyaux, C., Low-Choy, S., Mengersen, K., and Rousseau, J. (2012). Combining expert opinions in prior elicitation. *Bayesian Analysis*, 7(3):503–532.
- Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école américaine. *Econometrica*, 21(4).
- Allais, M. (1979). *Expected Utility Hypotheses and the Allais Paradox*, chapter 12: The so-called Allais paradox and rational decisions under uncertainty, pages 437–681. Dordrecht/Boston, Reidel.
- Berger, J. (1985). *Statistical Decision Theory and Bayesian Analysis*. In Statistics. Springer-Verlag, New York.
- Bernoulli, D. (1738). Specimen theoriae novae de mensura sortis. *Commentarii Academiae Scientiarum Imperialis Petropolitanae*, 5.
- CDC (2011). CDC estimates of foodborne illness in the United States. Technical report. <http://www.cdc.gov/foodborneburden/2011-foodborne-estimates.html>, last accessed on July 6, 2014.
- Chaloner, K. (1996). Elicitation of prior distributions. In *Case Studies in Bayesian Biostatistics*. Edited by Berry, D. and Stangl, D., Marcel Dekker, New York, USA.
- Cornu, M., Billoir, E., Bergis, H., Beaufort, A., and Zuliani, V. (2011). Modeling microbial competition in food: Application to the behavior of *listeria monocytogenes* and lactic acid flora in pork meat products. *Food Microbiology*, 28:639–647.
- Farquhar, P. H. (1984). State of the art—utility assessment methods. *Management Science*, 30(11):1283–1300.
- Gelfand, A. and Smith, A. (1990). Sampling based approaches to calculating marginal densities. *Journal Amer. Stat. Assoc.*, 85:398–409.
- Gorton, S., Kliebenstein, J., and Beran, G. (1997). Cost of on-farm microbial testing for *Salmonella*: An application by farm size and prevalence level. Technical Report ASL-R1413, Iowa State University.
- ICMSF (1974). *Micro-organisms in foods 2 : sampling for microbiological analysis: principles and specific application*. Univ. of Toronto Press.
- ICMSF (2002). *Micro-organisms in Foods 7: microbiological testing in food safety management*. New-York, Kluwer academic/Plenum Publisher.
- InVS (2010). Nombre de cas de listérioses déclarées par an et taux d'incidence annuel de 1999 à 2010. <http://www.invs.sante.fr/Dossiers-thematiques/Maladies-infectieuses/Maladies-a-declaration-obligatoire/Listeriose/Donnees-epidemiologiques.pdf>, last accessed on September 4, 2014.
- Jaynes, E. T. (1968). Prior probabilities. *IEEE Trans. Systems Sci. Cybernetics*, 4(3).
- Jemmi, T. and Stephan, R. (2006). *Listeria monocytogenes* : food-borne pathogen and hygiene indicator. *Rev. sci. tech. Off. int. Epiz*, 25(2):571–580.
- Jordaan, I. (2005). *Decisions Under Uncertainty: Probabilistic Analysis For Engineering Decisions*. Cambridge University Press.
- Kadane, J. (2011). *Principles of Uncertainty*. Texts in Statistical Science. Chapman & Hall.
- Kadane, J., Wolson, L., O'Hagan, A., and Craig, K. (1998). Papers on elicitation with discussions. *The Statistician*, pages 3–53.
- Kass, R. and Wasserman, L. (1996). The selection of prior distributions by formal rules. *Journal of the American Statistical Association*, 91, Ni½ 435:1343–1370.

- Krzysztofowicz, R. and Duckstein, L. (1980). Assessment errors in multiattribute utility functions. *Organizational Behavior and Human Performance*, 26(3):326 – 348.
- Kuhnert, P., Martin, T., and Griffiths, S. (2010). A guide to eliciting and using expert knowledge in Bayesian ecological models. *Ecology Letters*, 13:900–914.
- Lecoutre, B. (2006). And if you were a Bayesian without knowing it? *AIP Conference Proceedings*, 872(1):15.
- Legendre, V. (2008). Les consommateurs de porc frais et de charcuterie : qui sont-ils ? éclairage sociologique. *TechniPorc, la revue technique de l'Institut Français des Industries du Porc*, 31(4).
- Lindley, D. (1991). *Making Decision*. John Wiley & Sons, INC, New York.
- Lindley, D. (2006). *Understanding Uncertainty*. John Wiley & Sons, INC, New York.
- Machina, M. (1982). Expected utility without the independence axiom. *Econometrica*, 50.
- Machina, M. J. (1987). Choice under uncertainty: Problems solved and unsolved. *The Journal of Economic Perspectives*, 1(1):121–154.
- Munier, B. (2008). *Risk, Decision and Rationality*. Dordrecht, Reidel Publishing Company.
- O'Hagan, A., Buck, C., Daneshkhah, A., Eiser, R., Garthwaite, P., Jenkinson, D., Oakley, J., and Rakow, T. (2006). *Uncertain judgements: eliciting expert probabilities*.
- Pasanisi, A., Keller, M., and Parent, E. (2012). Estimation of a quantity of interest in uncertainty analysis: Some help from Bayesian decision theory. *Reliability Engineering & System Safety*, 100(0):93 – 101.
- Piattelli-Palmarini, M. (1995). *La Réforme du Jugement: Ou Comment Ne Plus Se Tromper*. Odile Jacobs.
- Pratt, J., Raiffa, H., and Schlaifer, R. (1964). The foundations of decision under uncertainty: an elementary exposition. *Journal of the American Statistical Association*, 59(306):353–375.
- Quiggin, A. (1964). A theory of anticipated utility. *Journal of Economic Behavior and Organization*, 3:323–343.
- Raiffa, H. (1968). *Decision Analysis - Introductory lectures on choices under uncertainty*. Addison Wesley, Reading, MA. traduction française: *Analyse de la décision : introduction aux choix en avenir incertain*, Dunod, 1973.
- Raiffa, H. and Schlaifer, R. (1961). *Applied Statistical Decision Theory*. Harvard University Press, Harvard.
- Regulation (CE) No 2073/2005 (2005). On microbiological criteria for foodstuffs. *Official Journal of the European Union*, L338:1–26.
- Savage, L. J. (1954). *The Foundations of Statistics*. Dover Publications, New York.
- Teisl, M. and Roe, B. (2010). Consumer willingness-to-pay to reduce the probability of retail foodborne pathogen contamination. *Food policy*, 35:521–530.
- Tvesky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5:297–321.
- Ulmo, J. and Bernier, J. (1973). *Eléments de Décision Statistique*. PUF.
- van der Gaag, M., Saatkamp, H., Backus, G., van Beek, P., and B.M.H., R. (2004). Cost-effectiveness of controlling *Salmonella* in the pork chain. *Food Control*, 15:173–180.
- Wald, A. (1950). *Statistical Decision Functions*. Wiley.

Correspondence: [eric.parent@agroparistech.fr](mailto:eric.parent@agroparistech.fr).