# 1

# Bayesian statistics and the agro-food production chain

*Tom Fearn*[#]

## Abstract

The essence of the Bayesian approach to inference is that all uncertainly is expressed through the specification of probability distributions. Its use is particularly advantageous in situations where uncertainties of various kinds have to be combined, in situations where the desired end is a probability statement about a hypothesis, in problems that involve making a decision, and in the modelling of complex systems. Opportunities of all these types exist in the agro-food production chain.

## Introduction

The brief for this contribution was to identify opportunities for the Bayesian approach in the agro-food production chain. The paper begins with a discussion in Section 2 of the essential features of the approach and some of its strengths and weaknesses. Section 3 then tries to highlight some of the opportunities.

## The Bayesian approach

### Essentials

The essence of the Bayesian approach to inference is that all uncertainly is expressed through the specification of probability distributions. This includes uncertainty about quantities that would be treated as unknown parameters in a classical statistical approach. When data are observed, these probability distributions are updated by conditioning on the observations.

Suppose, for example, that we are interested in the unknown amount $x$ of pesticide present in a food sample, and that we are about to make a relevant measurement $y$ on the sample. What we need to do is specify a joint probability distribution $p(x,y)$ for the two quantities $x$ and $y$. This will usually, but not necessarily, be done by specifying a conditional distribution $p(y|x)$ for the measurement given the amount present and expressing our beliefs about the possible values of $x$ by specifying a probability distribution $p(x)$ for $x$. Then the joint distribution is $p(x,y) = p(y|x)p(x)$. Once $y$ is known we condition this joint distribution on the observed value, calculating $p(x|y) = p(x,y)/p(y)$.

What we now believe about $x$ is captured in the conditional distribution $p(x|y)$. We might, for example, use the mean of this distribution as an estimate of $x$, or quote an interval that includes 95% of the probability as a confidence interval for $x$. In some

[#] Department of Statistical Science, University College London, Gower Street, London, WC1E 6BT, UK. E-mail: tom@stats.ucl.ac.uk

situations, when $y$ is a precise measurement and we do not know much about $x$ before observing $y$, the statements we make may be numerically very similar to those resulting from a classical statistical approach to the same problem. However the interpretation is very different. The Bayes 95% confidence interval really does have the interpretation that we believe that $x$ lies between these two numbers with probability 0.95. A classical one does not, even though many people think it does. In other situations, where we are able to use $p(x)$ to incorporate extra information about $x$, the Bayesian inference may be quite different to the classical one, typically being rather more precise.

There are many textbooks on the subject for those wanting to learn more. Of the introductory level texts, (Berry 1996) is one of the most accessible, whilst (Lee 1997) is rather more mathematical. I particularly like (Gelman et al. 1995), which would be a suitable introduction for someone with some experience of statistics but little knowledge of Bayes. Although (Lindley 1971) is about making decisions, it would provide a non-statistician with a very readable introduction to the fundamental ideas in Bayesian statistics.

**Advantages**

The main advantage of the Bayesian approach is that it provides a logical and coherent framework for dealing with uncertainty.

In a classical framework we cannot talk about the probability that a hypothesis is true, or the probability that a parameter, such as $x$ in the pesticide example above, lies in some interval. Avoiding these, by making probability statements that appear to be about hypotheses or parameters but are actually about data we might have observed but did not, requires considerable ingenuity and causes students of statistics many problems of understanding. If we adopt the Bayesian approach we can calculate the probability of anything we like and it will have a straightforward interpretation. For example it might be of considerable interest whether $x$ is greater than some allowable threshold $T$. We can easily find prob($x > T$) from the distribution $p(x|y)$. We are not even allowed to consider this probability in a classical framework. Instead the inference is typically based on the probability of observing a $y$ less than or equal to the one actually observed given that $x = T$. This is not at all the same thing and not obviously relevant.

When the motivation for the analysis is to make a decision in the face of uncertainty, the argument for using the Bayesian approach becomes even stronger. If you want to build a theory that can tell you whether to accept or reject a batch it seems highly desirable to start with a framework in which it is possible to discuss the probability that the batch is defective.

Bayesian decision theory (Lindley 1971; Berger 1993) combines the probability calculus described above with so-called utilities that quantify the desirability of outcomes to provide a rational framework for optimizing decisions.

A particular strength of modern Bayesian methodology is the ability to deal with large and complex problems. This is partly because it provides a clearly defined method for dealing with inference about many unknown parameters. Our beliefs about the parameters, either before or after observing data, are represented in a joint probability distribution. To make inference about any one parameter, we have to integrate this joint distribution to find the marginal probability distribution for the quantity of interest. In contrast, classical statistical methodology tends to resort to a variety of ad-hoc methods when faced with problems involving many parameters. The other reason for the strength, and the reason for the qualification ‘modern’ in the opening sentence of this paragraph, is that recent advances in computational methods for Bayesian statistics have made it possible to evaluate the desired probability distributions in a relatively straightforward manner.

**Difficulties**

The distribution $p(x)$ that describes our belief about $x$ before we observe $y$ is an essential ingredient of the procedure. It is called the prior distribution of $x$. We may specify it explicitly, or implicitly through the joint distribution $p(x,y)$, but we must specify it. The Bayesian paradigm then tells us how to update this belief about $x$ in the light of observing $y$ to give the posterior distribution $p(x|y)$.

Criticisms of the approach tend to focus on the role of this prior distribution: who should specify this, whether personal beliefs have any role in science, and so on. Similar criticisms can be made over the specification of utilities. The force of these criticisms depends very much on the context of the analysis. When the inference or decision is an in-house matter, for example a decision to pass a batch of material on to the next stage in a process or to reject and send it for reworking, it is clearly appropriate for the manufacturer concerned to use whatever historical data and expert opinion he can bring to bear on the problem. As we get closer to the public domain, and especially when legislation and external regulation are involved, the arguments get more difficult. Sometimes the best one can do is to use a range of plausible assumptions and examine the sensitivity of the inference.

## Opportunities

What follows is an attempt to identify some of the opportunities for Bayesian methods in the agro-food production chain. It is possible to argue, with some force, that the Bayesian approach is the only correct one in any situation involving inference or decision in the face of uncertainty. However, there are many situations in which it makes very little practical difference which approach is adopted. The areas highlighted below are ones where the choice of methodology does matter, and where there is likely to be a distinct advantage in using a Bayesian approach. The list makes no claims for completeness. Similarly, the examples cited are drawn from the author's own work and do not pretend to be a bibliography of agro-food applications. The other papers in this volume make a good introduction to some of these applications.

**Probability**

The fact that all uncertainty may be, indeed must be, described by probability is a major advantage in situations where uncertainties of various kinds need to be combined, and in situations where the natural inference is a probability statement about a hypothesis.

As an example of the first situation, any attempt to carry out a risk analysis for, say, some particular type of microbiological hazard in a foodstuff will necessarily involve a blend of information from many sources, ranging from hard experimental data to expert opinion.

As an example of the second, "what is the probability that this food sample contains more than 1.5% GMO?" is a question which can be answered directly only in a Bayesian framework.

**Decision theory**

In the agro-food industry, as in others, many measurements are made in order to inform decisions. Does this crop need more fertilizer? Should this batch of raw material be accepted or rejected? Then a Bayesian decision-theory approach may lead to significant economic gains.

It is not always easy to apply. Not only do we have to quantify all uncertainly by assigning probabilities, as for any Bayesian analysis, but we also have to quantify the desirability of all the possible outcomes subsequent to the decision. This desirability may be expressed in simple monetary terms for some problems. For others, particularly when very large potential losses such as a possible product recall are involved, the concept of utility (Lindley 1971) may be needed. In essence this allows a nonlinear transformation of the monetary scale so that, for example, a loss of one million euro with probability one in one million is not regarded as equivalent to a certain loss of one euro. It is because their consumers have a nonlinear utility for money that both lotteries and insurance companies are able to prosper.

In some industrial situations, including many in the agro-food industry, the availability of historical production data may make the specification of appropriate probabilities relatively easy. Quantifying the costs or utilities of potential outcomes may be harder. An argument commonly met is that these costs are not quantifiable, and that therefore decision theory cannot be applied. The obvious counter argument is that each decision taken implies some judgement about possible costs. By making these judgements explicit we expose them to scrutiny and we may get better decisions as a result. There is enormous scope for improving decision making by putting it on a more rational basis, and very little evidence that anyone is tackling the practical problem.

Some problems that are perhaps less obvious candidates for decision theory may also be clarified by this approach. Designing a sampling scheme and choosing measurement methods that are fit for purpose in some particular context is a problem that has been studied using decision theory (Fearn et al. 2002). The idea is to trade off the sampling and measurement costs against the possible losses due to uncertainty in the resulting measurement. Treating the problem in this way serves to make more precise a concept that, in analytical chemistry at least, has been rather vague until recently.

**Complex problems**

Bayesian methods have been very successful in tackling two types of complex problem. The distinction between the two is somewhat artificial and rather blurred, and probably arises mainly because different computational strategies are involved, rather than because of any deep difference in the type of problem.

The first type of problem may be characterized as an attempt to model a system involving many variables, some of which affect others, possibly in a causal way, and about some or all of which there is uncertainty as to their values. A simple non-food example would be diagnosis of a medical condition given the values of various tests or symptoms. Bayesian belief networks (Cowell et al. 1999) are a powerful tool for modelling such systems. The system is described by graphical network, where links between variables indicate dependencies, and the uncertainty is described by conditional probability distributions, specifying the probability for each variable given the values of the ones linked to it. If we restrict these probability distributions to be discrete ones, or to belong to a particular mixture family, then we can use powerful algorithms to propagate probability calculations through the system in response to, for example, a new piece of evidence. The paper by Barker in this volume describes a food safety example in detail. The technology to handle these networks is relatively new, easily available, and has very many potential applications in food safety and quality.

The second type of problem could be described as one involving many parameters. Here parameter is used in the statistical sense of an unknown quantity, usually denoted by a Greek letter, that is part of a stochastic model describing some system. There may be many parameters because the system is high-dimensional, with inference from spectroscopic data being an example (Brown, Vannucci and Fearn 1998). Alternatively it may be convenient to introduce many parameters in order to model, for example, the growth of many individual animals. Often a hierarchical structure is a natural one. For example we might have two or three parameters that characterize the growth curve for each individual animal, and consider these to be drawn from a common distribution, characterized by further parameters, at the next level of the hierarchy. Although such 'random-effects models' can be treated in a non-Bayesian way, they are particularly amenable to a Bayesian approach. Markov chain Monte Carlo (MCMC) methods (Gilks, Richardson and Spiegelhalter 1996), which efficiently simulate from the required posterior distributions, work particularly well with these hierarchical structures. Applications in agro-food production are most likely to be found at the research end of the spectrum, in areas where parametric models are fitted to data in order to understand and/or predict complex systems.

## Conclusion

It should be clear from the above discussion that this author, at least, is convinced that there are many as yet unexploited opportunities to apply Bayesian methods in the agro-food production chain. It may be relevant that one of the major growth areas for applications of these methods has been in biostatistics (Berry and Stangl 1996). Researchers and producers in the agro-food context face many of the same problems, notably the need to cope with large amounts of natural variation in anything 'bio' and a desire to understand and model large and complex systems. It seems likely that some of the same tools will prove useful.

## References

Berger, J.O., 1993. *Statistical decision theory and Bayesian analysis*. Springer-Verlag, New York. Springer series in statistics.

Berry, D.A., 1996. *Statistics: a Bayesian perspective*. Duxbury Press, Belmont. Alexander Kugushev book.

Berry, D.A. and Stangl, D.K., 1996. *Bayesian biostatistics*. Dekker, New York. Statistics: textbooks and monographs, vol. 151.

Brown, P.J., Vannucci, M. and Fearn, T., 1998. Bayesian wavelength selection in multicomponent analysis. *Journal of Chemometrics,* 12 (3), 173-182.

Cowell, R.G., Dawid, A.P., Lauritzen, S., et al., 1999. *Probabilistic networks and expert systems*. Springer, New York. Statistics for engineering and information science.

Fearn, T., Fisher, S.A., Thompson, M., et al., 2002. A decision theory approach to fitness for purpose in analytical measurement. *Analyst,* 127 (6), 818-824.

Gelman, A., Carlin, J.B., Stern, H.S., et al., 1995. *Bayesian data analysis*. Chapman and Hall, London. Texts in statistical science series.

Gilks, W.R., Richardson, S. and Spiegelhalter, D.J., 1996. *Markov Chain Monte Carlo in practice*. Chapman & Hall, London. Interdisciplinary statistics.

Lee, P.M., 1997. *Bayesian statistics: an introduction*. Arnold, London. [http://www-users.york.ac.uk/~pml1/bayes/book.htm]

Lindley, D.V., 1971. *Making decisions*. John Wiley, London.