On the use of the imprecise Dirichlet model with fault trees

Matthias C. M. Troffaes
Department of Mathematical Sciences
Durham University
Science Laboratories, South Road
Durham, DH1 3LE, United Kingdom
matthias.troffaes@gmail.com

Frank Coolen
Department of Mathematical Sciences
Durham University
Science Laboratories, South Road
Durham, DH1 3LE, United Kingdom
Frank.Coolen@dur.ac.uk

Abstract

We demonstrate how the imprecise Dirichlet model can be used in modelling fault trees. As a simple example, we consider a system consisting of two parallel subsystems A and B, and assume that the whole system C fails if and only if both components A and B fail. Given test data about the failure of A or B in a sequence of experiments, what can we say about the (interval-valued) posterior predictive probability of a particular component of the system failing upon a further single test or use of the system, taking differing assumptions about the data and the dependence of components A and B into account?

We will use the standard Bayesian framework with a Dirichlet multinomial model, and then generalise this to allow classes of priors, as in Walley's imprecise Dirichlet model (Walley 1996). Restricted to the standard Bayesian approach, this is just a simple special case of system reliability inference with multilevel failure information, which was theoretically developed in the 80's (Vesely, Goldberg, Roberts, and Haasl 1981, p. XI-24).

We extend various features of the imprecise Dirichlet model to accommodate the particular problem at hand, and we are led to study various statistical assumptions about how the sample was generated. This simple example allows us to pin-point a number of interesting effects on the precision of the posterior probabilities under varying assumptions, and also admits an analytical analysis. Scaling these extensions to more complex systems presents a major research challenge.

1 Introduction: The Imprecise Dirichlet Model

In studying the reliability of a system, often, the main sources of information are frequencies of particular primary events, and expert opinions in the form of subjective probabilities of events. A fault tree can then be considered as a tool to combine these frequencies and expert opinions into a measure of reliability of the whole system, for instance, a probability of complete failure.

A frequent problem in fault trees is that sometimes insufficient data is available to arrive at a reliable probability estimate. This is particularly problematic for rare events, and such events often arise in reliability. Another problem is that experts sometimes find it hard to formulate a precise probability, and are more comfortable with providing lower and upper probabilities rather than being forced to pinpoint an accurate probability. This is particularly true in fault tree analysis where failure probabilities are often only known up to a factor 10 or more. The imprecise Dirichlet model (Walley 1996) addresses both of these problems in an elegant way.

The imprecise Dirichlet model begins with a set of k mutually exclusive and exhaustive events $\Omega = \{1, \dots, k\}$. These events could for example be failures, faults, or particular conditions under which certain faults occur. Let θ_i be the uncertain probability of each event i, and denote by $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$. Since we are concerned with the case where the probability vector $\boldsymbol{\theta}$ cannot be specified precisely, a standard Bayesian approach is to put a distribution over $\boldsymbol{\theta}$ to model our prior knowledge about $\boldsymbol{\theta}$, and then apply Bayes theorem using the proper likelihood, as specified by the sampling model. In this way we arrive at a posterior distribution over $\boldsymbol{\theta}$, which can be used for further inferences about $\boldsymbol{\theta}$.

Imagine that our data consists of a particular i.i.d. sequence of N of such events, and say that each event i is observed n_i times in this sequence; let $\boldsymbol{n} = (n_1, \dots, n_k)$. Then the probability of observing such a sequence, if we knew the probability vector $\boldsymbol{\theta}$, is

$$P(\boldsymbol{n}|\boldsymbol{\theta}) = \prod_{i=1}^{k} \theta_i^{n_i} \tag{1}$$

In the Bayesian framework this probability, as a function of $\boldsymbol{\theta}$, is called the *likelihood function.*

A particularly interesting form for the prior distribution on $\boldsymbol{\theta}$ is one which is conjugate to the above likelihood:

$$\pi(\boldsymbol{\theta}) \propto \prod_{i=1}^{k} \theta_i^{st_i - 1}$$

This prior is called the *Dirichlet distribution* with hyperparameters $s > 0$ and $\boldsymbol{t} = (t_1, \ldots, t_k)$. The vector $\boldsymbol{t}$ is a probability vector representing the prior expectation of $\boldsymbol{\theta}$ (with each $t_i > 0$ for technical reasons). To understand the meaning of s, consider the posterior distribution

$$\pi(\boldsymbol{\theta}|\boldsymbol{n}) \propto P(\boldsymbol{n}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) = \prod_{i=1}^{k} \theta_i^{n_i + st_i - 1}$$

The posterior expectation of $\boldsymbol{\theta}$ is hence $\frac{\boldsymbol{n}+s\boldsymbol{t}}{N+s}$: the parameter s controls the relative impact of prior versus data on the posterior. For large values of s, more data are required to move the posterior expectation of $\boldsymbol{\theta}$ away from its prior expectation $\boldsymbol{t}$.

We have combined both expert information, through the prior, and frequency data, through the likelihood. The lack of a large amount of data is, so to speak, compensated by expert information, modelled by the hyperparameters s and $\boldsymbol{t}$. As more data become available, the impact of the expert information will decrease. In principle, we could now go on studying fault trees using this model. Indeed, this is just a simple special case of system reliability inference with multilevel failure information, which was theoretically developed in the 80's (Vesely, Goldberg, Roberts, and Haasl 1981, p. XI-24). For computational methods in fault tree analysis, such as Markov chain Monte Carlo, see (Hamada, Martz, Reese, Graves, Johnson, and Wilson 2004), who also provide details of the earlier literature.

However, what if no expert information is available at all? Or what if experts find it hard to assess precise values for the hyperparameters s and $\boldsymbol{t}$? This is where the *imprecise* Dirichlet model comes into play. Instead of considering just a single Dirichlet distribution as prior, the imprecise Dirichlet model starts with the *set* of all Dirichlet distributions for a fixed value of s. That is, for each value of the hyperparameter $\boldsymbol{t}$, the corresponding Dirichlet distribution is updated using the likelihood in Eq. (1), inducing a *set* of posterior Dirichlet distributions, which can then be used for further inferences about $\boldsymbol{\theta}$.

Considering all hypervectors $\boldsymbol{t}$ corresponds (or at least, is meant to correspond) to the extreme case where no expert information is available at all. In case expert information is available, one could ask the expert to provide a set of prior expectations for $\boldsymbol{\theta}$, and this set of prior expectations can be used in the same way as above.

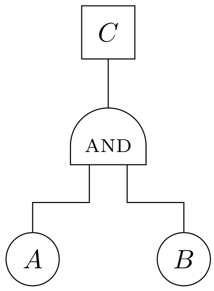
For example, we arrive at the following interval for the probability θ_i that we observe event i:

$$\left[\frac{n_i}{N+s}, \frac{n_i+s}{N+s}\right]$$

For the purpose of quantitative fault tree analysis, it often makes sense to focus on just one end of this interval: as reliability is often concerned with worst case scenarios, we might mostly be interested in the upper probability of total failure. If decisions regarding the system need to be made, then matters are slightly more complicated; we refer to the literature for further discussion (Troffaes (in press)).

2 A Toy Example

In the remainder of this paper, we attempt to apply the imprecise Dirichlet model on fault trees. In doing so, we consider a simple two-component system, with components A and B, where total failure occurs only if both A and B fail. We are both interested in the failure of the whole system, say C, and in the failure of one of the components, say A. So, we shall be studying the following fault tree:



where we identify the event of failure of A with the same letter A, and similar for B and C.

This simple example allows us to pin-point a number of interesting effects on the precision of the posterior probabilities under varying assumptions, and also admits an analytical analysis. Upscaling these extensions to more complex systems presents a major challenge.

We are given test data as follows: in a sequence of $N = N_A + N_B + N_C$ experiments, A failed n_A out of N_A times, B failed n_B out of N_B times, and C failed n_C out of N_C times. What do these data tell us about the probability of failure of the components of the system? More precisely, what can we say about the (interval-valued) posterior predictive probability of the whole system C failing, or a single component A failing, upon a further single test or use of the system, taking differing assumptions about the data and the dependence of components A and B into account?

3 Independent Components

Let θ_A and θ_B be the (possibly uncertain) probabilities of the events A and B. Assuming that failure of A is statistically independent from failure of B, we have $P(C|\theta_A, \theta_B) = \theta_A\theta_B$.

We denote the data on A by $\boldsymbol{n}_A = (N_A, n_A)$, and similarly for the events B and C. All data together are denoted by $\boldsymbol{n}$. Then the likelihood function is

$$
\begin{aligned}
P(\theta_A, \theta_B|\boldsymbol{n}) \propto\ & \theta_A^{n_A}(1-\theta_A)^{N_A-n_A} \\
& \times \theta_B^{n_B}(1-\theta_B)^{N_B-n_B} \\
& \times (\theta_A\theta_B)^{n_C}(1-\theta_A\theta_B)^{N_C-n_C}
\end{aligned}
$$

A convenient (and standard) choice of prior $p(\theta_A, \theta_B)$ is derived by assuming prior independence of these two parameters, and choosing conjugate priors on each parameter:

$$p(\theta_A) \propto \theta_A^{st_A-1}(1-\theta_A)^{s(1-t_A)-1}$$

and

$$p(\theta_B) \propto \theta_B^{st_B-1}(1-\theta_B)^{s(1-t_B)-1}$$

with $s > 0$, and t_A and $t_B \in (0, 1)$. The parameter t_A is the prior expectation of θ_A, and similarly for t_B. Again, the parameter s controls the impact of the prior on the posterior. We use the same value s for the prior both on θ_A and θ_B: this is not essential, but it simplifies the formulas; generalization is straightforward.

The posterior distribution follows directly

$$
\begin{aligned}
p(\theta_A, \theta_B|\boldsymbol{n}) \propto\ & \theta_A^{n_A+st_A-1}(1-\theta_A)^{N_A-n_A+s(1-t_A)-1} \\
& \times \theta_B^{n_B+st_B-1}(1-\theta_B)^{N_B-n_B+s(1-t_B)-1} \\
& \times (\theta_A\theta_B)^{n_C}(1-\theta_A\theta_B)^{N_C-n_C}
\end{aligned}
\tag{2}
$$

To calculate the posterior predictive probability of A and C (i.e. the posterior expectation of θ_A and θ_C) we need to integrate over the above posterior distribution. This calculation is not straightforward, because the posterior is not a product of a Dirichlet distribution in θ_A and a Dirichlet distribution in θ_B.

Elsewhere (Troffaes and Coolen (submitted)), we have shown that these posterior predictive probabilities can be calculated analytically. Unfortunately, the analytical expressions are quite large. For the purpose of this paper, let us simply give bounds on the posterior predictive probabilities of A and of C, also derived in (Troffaes and Coolen (submitted)):

$$P(C|\boldsymbol{n}) \in \left[\frac{n_A+st_A+n_C}{N_A+s+n_C}\times\frac{n_B+st_B+n_C}{N_B+s+N_C}, \frac{n_A+st_A+n_C}{N_A+s+N_C}\times\frac{n_B+st_B+N_C}{N_B+s+N_C}\right] \cap \left[\frac{n_B+st_B+n_C}{N_B+s+n_C}\times\frac{n_A+st_A+n_C}{N_A+s+N_C}, \frac{n_B+st_B+n_C}{N_B+s+N_C}\times\frac{n_A+st_A+N_C}{N_A+s+N_C}\right]$$

(where $[a,b]=\{x\colon \min\{a,b\}\leq x\leq \max\{a,b\}\}$) and

$$P(A|\boldsymbol{n}) \in \left[\frac{n_A+st_A+n_C}{N_A+s+N_C}, \frac{n_A+st_A+n_C}{N_A+s+n_C}\right]$$

These expressions give reasonably accurate estimates of the posterior predictive probabilities of failure of A, and of C, based on the data $\boldsymbol{n}$, and the initial expert estimates t_A and t_B.

However, what if no initial expert estimates are available? In that case, let us consider an imprecise Dirichlet model on both θ_A and θ_B, i.e., we consider the set of all possible values of $t_A \in (0,1)$ and $t_B \in (0,1)$. As the above bounds are monotone in t_A and t_B, we can immediately infer lower and upper bounds on the lower and upper predictive probabilities as well,

For example, the lower and upper probabilities $\underline{P}(C|\boldsymbol{n})$ and $\overline{P}(C|\boldsymbol{n})$ are the infimum and supremum, respectively, of the set of corresponding predictive posterior probabilities $P(C|\boldsymbol{n})$ for all possible values of the prior parameters t_A and t_B, and are therefore bounded by

$$\max\left\{\min\left\{\begin{matrix}\frac{n_A+n_C}{N_A+s+n_C}\times\frac{n_B+n_C}{N_B+s+N_C}, \\ \frac{n_A+n_C}{N_A+s+N_C}\times\frac{n_B+N_C}{N_B+s+N_C}\end{matrix}\right\}, \min\left\{\begin{matrix}\frac{n_B+n_C}{N_B+s+n_C}\times\frac{n_A+n_C}{N_A+s+N_C}, \\ \frac{n_B+n_C}{N_B+s+N_C}\times\frac{n_A+N_C}{N_A+s+N_C}\end{matrix}\right\}\right\},$$
$$\leq \underline{P}(C|\boldsymbol{n}) \leq \overline{P}(C|\boldsymbol{n}) \leq$$
$$\min\left\{\max\left\{\begin{matrix}\frac{n_A+s+n_C}{N_A+s+n_C}\times\frac{n_B+s+n_C}{N_B+s+N_C}, \\ \frac{n_A+s+n_C}{N_A+s+N_C}\times\frac{n_B+s+N_C}{N_B+s+N_C}\end{matrix}\right\}, \max\left\{\begin{matrix}\frac{n_B+s+n_C}{N_B+s+n_C}\times\frac{n_A+s+n_C}{N_A+s+N_C}, \\ \frac{n_B+s+n_C}{N_B+s+N_C}\times\frac{n_A+s+N_C}{N_A+s+N_C}\end{matrix}\right\}\right\}$$

Similarly, we have

$$\frac{n_A+n_C}{N_A+s+N_C} \leq \underline{P}(A|\boldsymbol{n}) \leq \overline{P}(A|\boldsymbol{n}) \leq \frac{n_A+s+n_C}{N_A+s+n_C}$$

Next we discuss a few special cases.

3.1 No Observations of C

If $N_C=0$, then

$$\frac{n_A}{N_A+s}\times\frac{n_B}{N_B+s} \leq \underline{P}(C|\boldsymbol{n}) \leq \overline{P}(C|\boldsymbol{n}) \leq \frac{n_A+s}{N_A+s}\times\frac{n_B+s}{N_B+s}$$

and

$$\frac{n_A}{N_A+s} \leq \underline{P}(A|\boldsymbol{n}) \leq \overline{P}(A|\boldsymbol{n}) \leq \frac{n_A+s}{N_A+s}$$

Recall that in this case, the posterior is an independent product of two Dirichlet distributions. Therefore, it is not so surprising that the posterior predictive probability intervals are a product of the two posterior predictive probability intervals induced by each imprecise Dirichlet model on A and on B separately.

3.2 Only observations of C

If $N_A = N_B = 0$ then

$$\frac{n_C}{n_C + s} \times \frac{n_C}{N_C + s} \le \underline{P}(C|\boldsymbol{n}) \le \overline{P}(C|\boldsymbol{n}) \le \frac{n_C + s}{N_C + s} \tag{3}$$

which is slightly more conservative than the Dirichlet model without the independence assumption—which would yield $\frac{n_C}{N_C+s}$ as a lower bound.

Regarding A, we have

$$\frac{n_C}{N_C + s} \le \underline{P}(A|\boldsymbol{n}) \le \overline{P}(A|\boldsymbol{n}) \le 1$$

The upper bound of 1, which is completely uninformative, can be explained as follows: if we only have observations regarding C then, in each case where C does not fail, there is no way of telling whether A failed or whether B failed. In fact, we only learn about the product of θ_A and θ_B. For example, imagine we have many observations of C, so $f_C = \frac{n_C}{N_C}$ can be used as a good estimate for θ_C. Because A and B are statistically independent, also approximately $\theta_A\theta_B = f_C$. But, the only fact about θ_A we can infer from $\theta_A\theta_B = f_C$, is that θ_A must belong to $[f_C, 1]$. This interval is the best we can do without additional prior information.

4 Dropping the Independence Assumption

Let us now analyse the fault trees again, without assuming independence of A and B.

4.1 Partial observations

Because the fault tree involves observations not just from mutually exclusive events—for example, if we observe failure of A, then we do not know whether B failed or not—we cannot apply the usual imprecise Dirichlet model directly. Let us investigate this problem in more detail.

The simplest sample space which fully models all possible outcomes of the fault tree consists of four elements:

	A	A^c
B	1	2
B^c	3	4

Category 1 obtains when both A and B fail, 2 when B fails but A does not, 3 when A fails but B does not, and 4 if both do not fail. C corresponds to category 1.

Recall, we are given $N = N_A + N_B + N_C$ observations, which can be summarized in the following contingency table:

event	count
A	n_A
A^c	$N_A - n_A$
B	n_B
B^c	$N_B - n_B$
C	n_C
C^c	$N_C - n_C$

But, not all of our observations correspond to the observation of a single category. We are dealing with *partial observations*: for instance, during the sequence of N_A experiments where A was monitored, we have not been told whether B failed or not. During that sequence, we only learn that the true state of the system belongs to either $\{1, 3\}$ or $\{2, 4\}$, but nothing more.

4.2 An Imprecise Dirichlet Model for Partial Observations

Partial observations can be dealt with by a straightforward extension of the imprecise Dirichlet model. Assume we have k categories, and let $\Omega = \{1, \ldots, k\}$ be the sample space. A multinomial sampling model generates a sequence of N outcomes $(\omega_1, \ldots, \omega_N)$ where each ω_i is independently chosen from Ω with an identical probability distribution $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$.

The likelihood of observing the sequence of events (rather than single categories, as in the traditional imprecise Dirichlet model) $(O_1, \ldots, O_N)$ is given by

$$\prod_{i=1}^{N} \left(\sum_{j \in O_i} \theta_j \right) = \prod_{O \subseteq \Omega} \left(\sum_{j \in O} \theta_j \right)^{n_O}$$

where n_O is the number of times event O occurs in the sequence $(O_1, \ldots, O_N)$. This can also be written as

$$= \sum_{\nu_{Oj}} \left(\prod_{O \subseteq \Omega} \binom{n_O}{\nu_{O1}, \ldots, \nu_{Ok}} \right) \left(\prod_{j=1}^{k} \theta_j^{\sum_{O \subseteq \Omega} \nu_{Oj}} \right)$$

if we take the convention that $\nu_{Oj} = 0$ whenever $j \notin O$, and where it is understood that the sum over ν_{Oj} runs over all counts ν_{Oj} for all $O \subseteq \Omega$ and all $j = 1, \ldots, k$ such that $\sum_{j \in O} \nu_{Oj} = n_O$. Again, the likelihood depends on the observations only through $\boldsymbol{n} = (n_O; O \subseteq \Omega)$.

If we are unsure about the true value of $\boldsymbol{\theta}$, the standard approach is to model our knowledge about $\boldsymbol{\theta}$ by a Dirichlet prior with parameters s and $\boldsymbol{t}$. After observation we have by Bayes theorem a posterior distribution

$$\pi(\boldsymbol{\theta}|\boldsymbol{n}) \propto \sum_{\nu_{Oj}} \prod_{O \subseteq \Omega} \binom{n_O}{\nu_{O1}, \ldots, \nu_{Ok}} \prod_{j=1}^{k} \theta_j^{st_j + \sum_{O \subseteq \Omega} \nu_{Oj} - 1} \tag{4}$$

where the proportionality constant follows from normalization. Note that the posterior is now a convex combination of Dirichlet distributions. It is possible to arrive at an analytic expression for the posterior predicitive of an event O in Ω, however the expression is rather large (Troffaes and Coolen (submitted)).

In similar spirit to the imprecise Dirichlet model, if we keep s fixed and let $\boldsymbol{t}$ vary over all possible values, $0 < t_j < 1$, and $\sum_{j=1}^{k} t_j = 1$, we end up with a set of posteriors, which are now convex combinations of Dirichlet distributions. This time, we have to rely on numerical methods for calculating lower and upper probabilities, because the posterior predictive probabilities are strongly non-linear in $\boldsymbol{t}$.

However, again, we can easily come up with bounds for the posterior predictive lower and upper probabilities:

$$\inf_{\nu_{Oj}} \frac{\sum_{O \subseteq \Omega} \nu_{OB}}{N + s} \leq \underline{P}(\omega_{N+1} \in B|\boldsymbol{n}) \leq \overline{P}(\omega_{N+1} \in B|\boldsymbol{n}) \leq \sup_{\nu_{Oj}} \frac{\sum_{O \subseteq \Omega} \nu_{OB} + s}{N + s}$$

where we denote by ν_{OB} the partial sum of $\nu_{O\ell}$ over all $\ell \in B$:

$$\nu_{OB} = \sum_{\ell \in B} \nu_{O\ell}.$$

This expression can be interpreted in terms of selection bias, which we address in the following section.

4.3 Compensating for Selection Bias

It is a very interesting observation that the containing interval just obtained exactly entails taking possible selection bias into account. One could for instance imagine a mechanism which reports specific events O for specific outcomes of the multinomial process.

For example, in case of our fault tree, we could imagine A only to be tested if B did not fail, in an attempt to make component A come out better in the resulting statistics. The statistics will be biased towards component A, but unless such crucial details about the experimental setup are revealed, we have no way to tell in general how much bias there is towards this or that event.

Another instance of selection bias happens when the data are reorganized to report only particular events if particular categories had been observed, effectively selecting part of the data. For example, one could report failure of only B whenever actually both components failed, so all failures of C would be reported as failures of B, and all failures of A would be instances where B did not fail. In this way one explicitly removes information from the contingency table: data are missing. But, even if we know that the data may have been tampered with, we usually do not know what selecting mechanism was used.

The proper way to model such situations where we cannot exclude the possibility of selection bias or missing data, but we wish to account for it, is by considering the set of all likelihood functions induced by all possible selection mechanisms, or equivalently, all possible *completions* ν_{Oj} of the counts n_O (De Cooman and Zaffalon 2004; Utkin 2006). Those completions ν_{Oj} are exactly the counts introduced previously:

$$P(\boldsymbol{\theta}|\nu_{Oj}) = \prod_{O\subseteq\Omega}\prod_{j\in O}\theta_j^{\nu_{Oj}} = \prod_{j\in O}\theta_j^{\sum_{O\subseteq\Omega}\nu_{Oj}}$$

Hence, applying the imprecise Dirichlet model, but now with a set of likelihood functions, and hence, a set of counts of the form $n_j = \sum_{O\subseteq\Omega}\nu_{Oj}$ running over all possible completions ν_{Oj}, we immediately recover the bounding interval mentioned in the previous section.

4.4 Application on Fault Trees

In our example, we arrive at

$$\frac{n_C}{N+s} \leq \underline{P}(C|\boldsymbol{n}) \leq \overline{P}(C|\boldsymbol{n}) \leq \frac{n_A+n_B+n_C+s}{N+s}$$

and

$$\frac{n_A+n_C}{N+s} \leq \underline{P}(A|\boldsymbol{n}) \leq \overline{P}(A|\boldsymbol{n}) \leq \frac{n_A+N_B+N_C+s}{N+s}$$

For example, the lower bound for $\underline{P}(C|\boldsymbol{n})$ obtains exactly when in all n_A failures of A, B did not fail, and in all n_B failures of B, A did not fail (a full compensation effect). The upper bound for $\overline{P}(C|\boldsymbol{n})$ corresponds to the case in which all failures of A, B failed as well, and vice versa.

The lower bound for $\underline{P}(A|\boldsymbol{n})$ obtains when A never failed in case C^c, B, or B^c was observed. The upper bound for $\overline{P}(A|\boldsymbol{n})$ obtains if A always failed if C^c, B, or B^c was observed.

Note that in general these bounds are very imprecise, even when the counts are large. If we have no model of the selection mechanism, then additional observations do not necessarily improve precision.

4.5 No Observations of C

If $N_C = 0$, then

$$0 \leq \underline{P}(C|\boldsymbol{n}) \leq \overline{P}(C|\boldsymbol{n}) \leq \frac{n_A+n_B+s}{N_A+N_B+s}$$

and

$$\frac{n_A}{N_A+N_B+s} \leq \underline{P}(A|\boldsymbol{n}) \leq \overline{P}(A|\boldsymbol{n}) \leq \frac{n_A+N_B+s}{N_A+N_B+s}$$

4.6 Only Observations of C

If $N_A = N_B = 0$, then

$$\frac{n_C}{N_C+s} \leq \underline{P}(C|\boldsymbol{n}) \leq \overline{P}(C|\boldsymbol{n}) \leq \frac{n_C+s}{N_C+s}$$

and

$$\frac{n_C}{N_C+s} \leq \underline{P}(A|\boldsymbol{n}) \leq \overline{P}(A|\boldsymbol{n}) \leq 1$$

5 Concluding Remarks

Independence has an obvious effect on the imprecision of the posterior. This effect is most clear in case we have no observations about C, i.e., when $N_C = 0$. In case we make no assumptions regarding the independence of A and B, and even take possible selection bias into account, then the posterior predictive probability intervals for both A and C usually become wider. These intervals will also not converge to points as more data become available. In conclusion, wrongfully assuming independence, we may end up with a too precise posterior and thereby underestimate the true risk of the system. This stresses the need for making good assumptions about data, and in particular the importance of modelling dependencies correctly.

A huge problem is how these calculations can be expanded to larger fault trees used in practice. For example, can we formulate simple rules by which imprecision propagates in a fault tree along particular gates?

An interesting question for future research is how various forms of dependence between the components A and B can be taken into account, and how one can learn about such dependence from the data. There are, of course, many ways to take dependence into account. Exchangeable components—when we know a priori that $\theta_A = \theta_B$—is clearly one important case of dependence. More generally, it may be difficult to learn about the form of dependence from the data. In particular, it is not clear how to arrive at a model which allows updating of dependencies.

Acknowledgments

The first author would like to thank Teddy Seidenfeld for useful comments and references in particular regarding selection bias. The first author is also grateful for the financial support provided by the Belgian American Educational Foundation during his stay at Carnegie Mellon University, where he has worked on most of the results reported in this paper. Both authors would like to thank Scott Ferson for suggesting the fault tree problem studied in this paper.

References

De Cooman, G. and M. Zaffalon (2004). Updating beliefs with incomplete observations. *Artificial Intelligence 159*(1–2), 75–125.

Hamada, M. S., H. Martz, C. S. Reese, T. Graves, V. Johnson, and A. Wilson (2004). A fully bayesian approach for combining multilevel failure information in fault tree quantification and optimal follow-on resource allocation. *Reliability Engineering and System Safety 86*, 297–305.

Troffaes, M. C. M. Decision making under uncertainty using imprecise probabilities. International Journal of Approximate Reasoning. In press.

Troffaes, M. C. M. and F. Coolen. Applying the imprecise Dirichlet model in cases with partial observations and dependencies in failure data. Submitted.

Utkin, L. V. (2006). A method for processing the unreliable expert judgments about parameters of probability distributions. *European Journal of Operational Research 175*(1), 385–398.

Vesely, W. E., F. F. Goldberg, N. H. Roberts, and D. F. Haasl (1981, January). *Fault Tree Handbook*. Wachington D. C.: U. S. Nuclear Regulatory Commision. NUREG-0492.

Walley, P. (1996). Inferences from multinomial data: Learning about a bag of marbles. *Journal of the Royal Statistical Society 58*(1), 3–34.