Matthias C. M. Troffaes and Frank P. A. Coolen. Applying the imprecise Dirichlet model in cases with partial observations and dependencies in failure data. *International Journal of Approximate Reasoning*, 50(2):257-268, February 2009.

Applying the imprecise Dirichlet model in cases with partial observations and dependencies in failure data

Matthias C. M. Troffaes

Department of Mathematical Sciences, Durham University, Science Laboratories, South Road, Durham, DH1 3LE, England

Frank P. A. Coolen

Department of Mathematical Sciences, Durham University, Science Laboratories, South Road, Durham, DH1 3LE, England

Abstract

Imprecise probabilistic methods in reliability provide exciting opportunities for dealing with partial observations and incomplete knowledge on dependencies in failure data. In this paper, we explore the use of the imprecise Dirichlet model for dealing with such information, and we derive both exact results and bounds which enable analytical investigations. However, we only consider a very basic two-component system, as analytical solutions for larger systems will become very complex. We explain how the results are related to similar analyses under data selection or reporting bias, and we discuss some challenges for future research.

Key words: imprecise Dirichlet model, independence, selection bias, partial observations, Bayesian inference, robustness

Email addresses: `matthias.troffaes@gmail.com` (Matthias C. M. Troffaes), `Frank.Coolen@durham.ac.uk` (Frank P. A. Coolen).

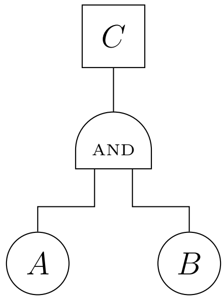

Fig. 1. A simple 3-node fault tree.

1 Introduction

Consider the 3-node fault tree depicted in Fig. 1. It is convenient to think of A and B as events of failure of the two components of a parallel system. Node C then represents the failure of the whole system. The fault tree describes how failures propagate: the whole system fails if and only if both components A and B fail.

In many systems, components are generally tested prior to assembly. For instance, they may be produced, tested, and assembled in separate locations. Hence, often we have test data of each component separately. In this paper, for the system depicted in Figure 1, we shall imagine a series of tests on component A only, a series of tests on component B only, and a series of tests on C only, i.e. observing failure of the whole system C without directly observing the components A and B as well.

So, we are given test data as follows: in a series 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 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 first use the standard Bayesian framework with a binomial model and beta prior (beta-binomial model), and then generalize this to allow classes of priors (as in Walley's book [1], and a special case of his imprecise Dirichlet model [2]). 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 [3, p. XI-24], with computational methods (MCMC) presented by Hamada et al. [4], who also provide details of the earlier literature. Coolen [5] discusses the use of the imprecise Dirichlet model for inference on lifetimes in reliability based on data including right-censored observations. Similar approaches to reliability with imprecise probabilities, in different settings, have been pursued by Utkin and Gurov [6] and Utkin [7]. A general overview of imprecise methods in reliability is presented by Utkin and Coolen [8].

As we go, we extend various features of the imprecise Dirichlet model to accommodate the particular problem at hand, and we study various statistical assumptions about how the sample was generated. The simple example depicted in Fig. 1 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. We also stress that we only look at posterior probabilities on the next observation, and not at the full posterior. Upscaling these extensions to more complex systems and making more involved inferences still presents a major challenge.

The paper is organized as follows. In Section 2 we model the 3-node fault tree through a beta-binomial model, additionally assuming that A and B are statistically independent. Section 3 is concerned with dropping the independence assumption, and with studying the effect of selection bias. We conclude in Section 4: we look at how the various structural assumptions affect the precision of the posterior, and suggest an alternative approach to approximating lower and upper posterior probabilities. The appendix contains relevant formulas for quick reference, and deals with the tedious calculation of one of the posteriors.

2 Independent components

Let θ_x be the Bernoulli parameter related to failure of component $x = A, B$, let $F_x = 1(0)$ denote failure (success) of $x = A, B, C$, so $P(F_x = 1|\theta_x) = \theta_x$ for $x = A, B$, and $P(F_C = 1|\theta_A, \theta_B) = \theta_A\theta_B$, assuming that failure of A is statistically independent from failure of B.

We denote the data on $x = A, B, C$ by $D_x = (N_x, n_x)$ meaning n_x failures out of N_x observations, with $0 \leq n_x \leq N_x$. All data together are denoted by $D = (D_A, D_B, D_C)$. Then the likelihood function is

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

2.1 Precise Bayesian approach

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 per 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 mean of θ_A; similarly for t_B. The parameter s determines the shape of each prior: for high values the prior will peak around its mean, and for low values the distribution will have a more uniform shape. We are using 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 joint posterior distribution follows directly:

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

where the proportionality constant follows from normalization, and will be a function of the data, and the hyper-parameters s, t_A, and t_B. The posterior predictive probabilities of interest are

$$P(F_C = 1|D) = \int_0^1 \int_0^1 \theta_A\theta_B p(\theta_A, \theta_B|D) \mathrm{d}\theta_A \mathrm{d}\theta_B, \tag{2}$$
$$P(F_A = 1|D) = \int_0^1 \int_0^1 \theta_A p(\theta_A, \theta_B|D) \mathrm{d}\theta_A \mathrm{d}\theta_B.$$

Of course, $P(F_B = 1|D)$ follows from $P(F_A = 1|D)$ by swapping A and B.

The above integrals are not straightforward to calculate analytically. If we have observed any successful tests of the whole system ($N_C - n_C > 0$) then the posterior is not a product of a function of θ_A times a function of θ_B. However, perhaps surprisingly, we can still calculate Eq. (2) analytically. In Appendix C it is shown that

$$P(F_C = 1|D) = \sum_{m=0}^{N_C - n_C} \tilde{\gamma}_m^{A,B} \frac{n_A + st_A + n_C}{N_A + s + n_C + m} \times \frac{n_B + st_B + n_C + m}{N_B + s + N_C}, \tag{3}$$

$$= \sum_{m=0}^{N_C - n_C} \tilde{\gamma}_m^{B,A} \frac{n_B + st_B + n_C}{N_B + s + n_C + m} \times \frac{n_A + st_A + n_C + m}{N_A + s + N_C}, \tag{4}$$

and

$$P(F_A = 1|D) = \sum_{m=0}^{N_C - n_C} \tilde{\gamma}_m^{A,B} \frac{n_A + st_A + n_C}{N_A + s + n_C + m}, \tag{5}$$

$$= \sum_{m=0}^{N_C - n_C} \tilde{\gamma}_m^{B,A} \frac{n_A + st_A + n_C + m}{N_A + s + N_C}, \tag{6}$$

where $\tilde{\gamma}_m^{A,B}$ and $\tilde{\gamma}_m^{B,A}$ are defined in the Appendix, Eqs. (C.5)–(C.6). They are non-negative real numbers depending on N_A, n_A, N_B, n_B, N_C, n_C, s, t_A, and t_B, satisfying

$$\sum_{m=0}^{N_C - n_C} \tilde{\gamma}_m^{A,B} = \sum_{m=0}^{N_C - n_C} \tilde{\gamma}_m^{B,A} = 1. \tag{7}$$

The analytical properties of the coefficients $\tilde{\gamma}_m^{A,B}$ and $\tilde{\gamma}_m^{B,A}$ are not easily captured, although they can be easily calculated numerically, as is apparent from Eq. (C.5). To allow analytical investigation, we shall bound the predictive probabilities.

First, observe that the factors that follow $\tilde{\gamma}_m^{A,B}$ and $\tilde{\gamma}_m^{B,A}$ in Eqs. (3)–(6) are monotone functions of m. For example, in Eq. (3), $\frac{n_A+st_A+n_C}{N_A+s+n_C+m} \times \frac{n_B+st_B+n_C+m}{N_B+s+N_C}$ is an increasing function of m if $N_A + s \geq n_B + st_B$, and decreasing otherwise. Hence, we immediately infer useful bounds for the probabilities (set $m = 0$ or $m = N_C - n_C$):

$$\begin{aligned} &P(F_C = 1|D) \\ &\in \left[\frac{n_A + st_A + n_C}{N_A + s + n_C} \times \frac{n_B + st_B + n_C}{N_B + s + N_C}, \frac{n_A + st_A + n_C}{N_A + s + N_C} \times \frac{n_B + st_B + N_C}{N_B + s + N_C}\right] \\ &\quad \cap \left[\frac{n_B + st_B + n_C}{N_B + s + n_C} \times \frac{n_A + st_A + n_C}{N_A + s + N_C}, \frac{n_B + st_B + n_C}{N_B + s + N_C} \times \frac{n_A + st_A + N_C}{N_A + s + N_C}\right] \end{aligned} \tag{8}$$

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

$$P(F_A = 1|D) \in \left[\frac{n_A + st_A + n_C}{N_A + s + N_C}, \frac{n_A + st_A + n_C}{N_A + s + n_C}\right]. \tag{9}$$

These bounds allow us to state a number of interesting results. Further on, they will allow us to bound the imprecise posterior probabilities analytically as well.

2.1.1 *No observations of C*

If $N_C = 0$, and hence $n_C = 0$, then the intervals in Eq. (8) reduce to a singleton,

$$P(F_C = 1|D) = \frac{n_A + st_A}{N_A + s} \times \frac{n_B + st_B}{N_B + s},$$

and similarly, from Eq. (9),

$$P(F_A = 1|D) = \frac{n_A + st_A}{N_A + s}, \qquad P(F_B = 1|D) = \frac{n_B + st_B}{N_B + s}.$$

In this case, the posterior probability of $F_C = 1$ is simply a product of the posterior probabilities of $F_A = 1$ and $F_B = 1$, because the joint posterior probability on (θ_A, θ_B) is a product of Dirichlet posteriors.

2.1.2 *Only observations of C*

If $N_A = N_B = 0$, we expect inferences about A and B still to depend largely on the parameters of the prior, because in each case where C does not fail, there is no way of telling whether A or B have failed.

Regarding C, Eq. (8) reduces to

$$\begin{aligned} P(F_C = 1|D) \in & \left[\frac{st_A + n_C}{s + n_C} \times \frac{st_B + n_C}{s + N_C}, \frac{st_A + n_C}{s + N_C} \times \frac{st_B + N_C}{s + N_C}\right] \\ & \cap \left[\frac{st_B + n_C}{s + n_C} \times \frac{st_A + n_C}{s + N_C}, \frac{st_B + n_C}{s + N_C} \times \frac{st_A + N_C}{s + N_C}\right] \end{aligned}$$

and regarding A, Eq. (9) becomes

$$P(F_A = 1|D) \in \left[\frac{st_A + n_C}{s + N_C}, \frac{st_A + n_C}{s + n_C}\right].$$

It is instructive to investigate the limit of Eqs. (3) and (5) for $s \to 0$ in case $N_C > n_C > 0$. From Eq. (1), one can see that the limit $s \to 0$ maximizes the effect of the data on the posterior. In Appendix C.3 it is shown that

$$\lim_{s\to 0} P(F_C = 1|D) = \frac{n_C}{N_C}, \qquad \lim_{s\to 0} P(F_A = 1|D) = \frac{(1 - t_B) + (1 - t_A)\frac{n_C}{N_C}}{2 - t_A - t_B}.$$

So, $P(F_A = 1|D)$ is a weighted average of 1 and the relative frequency of C, with t_A and t_B determining the weights. In conclusion, regardless of the size of N_C, the posterior probabilities of A and B will still depend largely on the parameters t_A and t_B of the prior, whereas the posterior probability of C does not depend on the prior at all, in case we observe just C. For larger values of s, the impact of the prior on the posterior increases. Therefore, in those cases, the posterior will vary even more as a function of t_A and t_B.

We shall come back to this issue and provide a more in-depth explanation at the end of Section 2.2.

2.2 Imprecise beta model

In many applications of interest, especially in reliability theory (for instance, see [3, p. XI-9]), it is hard to assess prior probabilities. Along the lines of Walley's imprecise beta model [1, §5.3, pp. 217–222] (which is of course the 2-categories special case of his imprecise Dirichlet model [2]), we take all beta priors into account, that is for all values t_A and t_B in the open interval $(0, 1)$ and for a fixed value of s. We then arrive at a set of posteriors instead of just a single one by application of Eq. (1) on each prior.

For example, the lower and upper probabilities $\underline{P}(F_C = 1|D)$ and $\overline{P}(F_C = 1|D)$ are then the infimum and supremum, respectively, of the set of corresponding predictive posterior probabilities $P(F_C = 1|D)$ for all possible values of the prior parameters t_A and t_B. Using the bounds for the precise posterior probabilities obtained in Eq. (8) and Eq. (9)—which should not be confounded with lower and upper probabilities—and taking into account that the interval bounds can swap, we arrive at the following bounds for the imprecise posterior probabilities:

$$
\begin{aligned}
&\max\left\{\min\begin{pmatrix}\frac{n_A+n_C}{N_A+s+n_C}\times\frac{n_B+n_C}{N_B+s+N_C},\\ \frac{n_A+n_C}{N_A+s+N_C}\times\frac{n_B+N_C}{N_B+s+N_C}\end{pmatrix},\min\begin{pmatrix}\frac{n_B+n_C}{N_B+s+n_C}\times\frac{n_A+n_C}{N_A+s+N_C},\\ \frac{n_B+n_C}{N_B+s+N_C}\times\frac{n_A+N_C}{N_A+s+N_C}\end{pmatrix}\right\}\\
&\leq \underline{P}(F_C = 1|D) \leq \overline{P}(F_C = 1|D)\\
&\leq \min\left\{\max\begin{pmatrix}\frac{n_A+s+n_C}{N_A+s+n_C}\times\frac{n_B+s+n_C}{N_B+s+N_C},\\ \frac{n_A+s+n_C}{N_A+s+N_C}\times\frac{n_B+s+N_C}{N_B+s+N_C}\end{pmatrix},\max\begin{pmatrix}\frac{n_B+s+n_C}{N_B+s+n_C}\times\frac{n_A+s+n_C}{N_A+s+N_C},\\ \frac{n_B+s+n_C}{N_B+s+N_C}\times\frac{n_A+s+N_C}{N_A+s+N_C}\end{pmatrix}\right\}, \qquad (10)\\
&[\underline{P}(F_A = 1|D), \overline{P}(F_A = 1|D)] \subseteq \left[\frac{n_A+n_C}{N_A+s+N_C}, \frac{n_A+s+n_C}{N_A+s+n_C}\right]. \qquad (11)
\end{aligned}
$$

The two special cases discussed in Sec. 2.1 generalize straightforwardly.

2.2.1 *No observations of C*

If $N_C = 0$, and hence $n_C = 0$, then (where we have equality because of similar arguments as in Sec. 2.1.1)

$$[\underline{P}(F_C = 1|D), \overline{P}(F_C = 1|D)] = \left[\frac{n_A}{N_A + s} \times \frac{n_B}{N_B + s}, \frac{n_A + s}{N_A + s} \times \frac{n_B + s}{N_B + s}\right], \quad (12)$$

and

$$[\underline{P}(F_A = 1|D), \overline{P}(F_A = 1|D)] = \left[\frac{n_A}{N_A + s}, \frac{n_A + s}{N_A + s}\right]. \quad (13)$$

2.2.2 *Only observations of C*

If $N_A = N_B = 0$ then from Eq. (10)

$$[\underline{P}(F_C = 1|D), \overline{P}(F_C = 1|D)] \subseteq \left[\frac{n_C}{n_C + s} \times \frac{n_C}{N_C + s}, \frac{n_C + s}{N_C + s}\right]. \quad (14)$$

For $n_C \gg s$, Eq. (14) is approximately equal to $[\frac{n_C}{N_C+s}, \frac{n_C+s}{N_C+s}]$ which is the interval we would obtain from the usual imprecise Dirichlet model with the same hyper-parameter s. Next,

$$[\underline{P}(F_A = 1|D), \overline{P}(F_A = 1|D)] \subseteq \left[\frac{n_C}{N_C + s}, 1\right]. \quad (15)$$

The right hand side of Eq. (15) is a result of only learning about $\theta_A \times \theta_B$ in case we only observe C. Indeed, recall the limit of $P(F_C = 1|D)$ and $P(F_A = 1|D)$ for $s \to 0$, in the precise case:

$$\lim_{s\to 0} P(F_C = 1|D) = \frac{n_C}{N_C}, \quad \lim_{s\to 0} P(F_A = 1|D) = \frac{(1 - t_B) + (1 - t_A)\frac{n_C}{N_C}}{2 - t_A - t_B}$$

Taking the infimum and supremum of these expressions over t_A and t_B, we find

$$\lim_{s\to 0}[\underline{P}(F_C = 1|D), \overline{P}(F_C = 1|D)] = \left\{\frac{n_C}{N_C}\right\},$$
$$\lim_{s\to 0}[\underline{P}(F_A = 1|D), \overline{P}(F_A = 1|D)] = \left[\frac{n_C}{N_C}, 1\right]$$

These results admit a very interesting interpretation: because in the limit we only learn that $\theta_A \times \theta_B$ equals $\frac{n_C}{N_C}$, it must hold that θ_A belongs to $[\frac{n_C}{N_C}, 1]$, and this interval is the best we can do without additional prior information.

3 Dropping the independence assumption

In Section 2, we assumed statistical independence of the components A and B. Although there are scenarios where this assumption is reasonable, it is difficult to justify in many situations. In this section, we explore the inferences without any assumptions about independence between A and B.

3.1 Partial observations

Because the 3-node fault tree involves observations not just from exclusive categories, we cannot apply the 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 table:

event	A	A^c	B	B^c	C	C^c
count	n_A	$N_A - n_A$	n_B	$N_B - n_B$	n_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 N_A experiments where A was monitored, we have not been told whether B failed or not. During these experiments, we only learn that category belongs to either $\{1, 3\}$ or $\{2, 4\}$, but nothing more.

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

Consider a series of events $(O_1, \dots, O_N)$ where each of these events can be identified with a subset of Ω. The likelihood of observing this series of events 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 series $(O_1, \dots, O_N)$. This can also be written as (see Appendix B and explanation of $\sum_{\nu_{O,j}}$ further on)

$$= \sum_{\nu_{O,j}} \left(\prod_{O \subseteq \Omega} \binom{n_O}{\nu_{O,j}; j \in O} \prod_{j \in O} \theta_j^{\nu_{O,j}} \right)$$

and if we write $\nu_{O,j}$ briefly as ν_{Oj},

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

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

where it is understood that the sum over ν_{Oj} runs over all tuples $(\nu_{Oj}; O \subseteq \Omega, j \in \Omega)$ which satisfy $\sum_{j \in O} \nu_{Oj} = n_O$ for all $O \subseteq \Omega$ and $\nu_{Oj} = 0$ whenever $j \notin O$. 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 $\boldsymbol{\theta}$, it is convenient to model our knowledge about $\boldsymbol{\theta}$ by a Dirichlet prior with parameters $(s, t_1, \dots, t_k) = (s, \boldsymbol{t})$, where $s > 0$, $0 < t_j < 1$, and $\sum_{j=1}^{k} t_j = 1$:

$$\pi(\boldsymbol{\theta} | s, \boldsymbol{t}) \propto \prod_{j=1}^{k} \theta_j^{s t_j - 1}. \tag{16}$$

The set of all feasible values of $\boldsymbol{t}$ will be denoted by T. Using this Dirichlet prior, after observation of the series of events $(O_1, \dots, O_N)$, by Bayes rule, we have a posterior distribution

$$\pi(\boldsymbol{\theta}|s,\boldsymbol{t},\boldsymbol{n}) = \frac{\displaystyle\sum_{\nu_{Oj}}\prod_{O\subseteq\Omega}\binom{n_O}{\nu_{O1},\ldots,\nu_{Ok}}\prod_{j=1}^{k}\theta_j^{st_j+\sum_{O\subseteq\Omega}\nu_{Oj}-1}}{\displaystyle\sum_{\nu_{Oj}}\prod_{O\subseteq\Omega}\binom{n_O}{\nu_{O1},\ldots,\nu_{Ok}}\frac{\prod_{j=1}^{k}\Gamma(st_j+\sum_{O\subseteq\Omega}\nu_{Oj})}{\Gamma(s+N)}}. \tag{17}$$

where we use that for any real numbers $\alpha_1, \ldots, \alpha_k > 0$ it holds that

$$\int \theta_1^{\alpha_1-1}\cdots\theta_k^{\alpha_k-1}\mathrm{d}\boldsymbol{\theta} = \frac{\prod_{j=1}^{k}\Gamma(\alpha_j)}{\Gamma(\sum_{j=1}^{k}\alpha_j)}, \tag{18}$$

The posterior is now a convex combination of Dirichlet distributions.

In similar spirit to the imprecise Dirichlet model, if we start out with the set $\mathcal{M}_0$ of all Dirichlet priors with parameters $(s,\boldsymbol{t})$, where s is a fixed constant and $\boldsymbol{t}$ varies over all possible values in T, we end up with a set of posteriors $\mathcal{M}_{\boldsymbol{n}}$, each posterior in the set being a convex combination of Dirichlet distributions. As an example, we investigate the lower probability of the next observation to belong to the set $S \subseteq \Omega$. We now have

$$\begin{aligned}
&\underline{P}(\omega_{N+1}\in S|s,\boldsymbol{n}) \\
&= \inf_{\boldsymbol{t}\in T}\int\left(\sum_{\ell\in S}\theta_\ell\right)\frac{\displaystyle\sum_{\nu_{Oj}}\prod_{O\subseteq\Omega}\binom{n_O}{\nu_{O1},\ldots,\nu_{Ok}}\prod_{j=1}^{k}\theta_j^{st_j+\sum_{O\subseteq\Omega}\nu_{Oj}-1}}{\displaystyle\sum_{\nu_{Oj}}\prod_{O\subseteq\Omega}\binom{n_O}{\nu_{O1},\ldots,\nu_{Ok}}\frac{\prod_{j=1}^{k}\Gamma(st_j+\sum_{O\subseteq\Omega}\nu_{Oj})}{\Gamma(s+N)}}\mathrm{d}\boldsymbol{\theta} \\
&= \inf_{\boldsymbol{t}\in T}\sum_{\ell\in S}\frac{\displaystyle\sum_{\nu_{Oj}}\prod_{O\subseteq\Omega}\binom{n_O}{\nu_{O1},\ldots,\nu_{Ok}}\prod_{j=1}^{k}\Gamma\left(st_j+\sum_{O\subseteq\Omega}\nu_{Oj}\right)\frac{st_\ell+\sum_{O\subseteq\Omega}\nu_{O\ell}}{s+N}}{\displaystyle\sum_{\nu_{Oj}}\prod_{O\subseteq\Omega}\binom{n_O}{\nu_{O1},\ldots,\nu_{Ok}}\prod_{j=1}^{k}\Gamma\left(st_j+\sum_{O\subseteq\Omega}\nu_{Oj}\right)}
\end{aligned} \tag{19}$$

An expression for $\overline{P}(\omega_{N+1}\in S|s,\boldsymbol{n})$ follows by replacing inf by sup in Eq. (19). Clearly, we have to rely on numerical methods for calculating the infimum and the supremum.

However, as in Sec. 2.2, we can easily come up with bounds for Eq. (19):

$$[\underline{P}(\omega_{N+1} \in S|s, \boldsymbol{n}), \overline{P}(\omega_{N+1} \in S|s, \boldsymbol{n})] \subseteq \left[\inf_{\nu_{Oj}} \frac{\sum_{O\subseteq\Omega} \nu_{OS}}{N+s}, \sup_{\nu_{Oj}} \frac{\sum_{O\subseteq\Omega} \nu_{OS} + s}{N+s}\right] \quad (20)$$

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

$$\nu_{OS} = \sum_{\ell \in S} \nu_{O\ell}.$$

Eq. (20) can be interpreted in terms of selection bias, which we address in the following section.

3.3 Compensating for selection bias

It is an interesting observation that the containing interval obtained in Eq. (20) 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, for example for economic reasons or perhaps even 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 is reorganized to report only particular events if particular categories have 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: data is missing. But, even if we know 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 (see De Cooman and Zaffalon [9], and Utkin [10]). Those completions ν_{Oj} are exactly the counts introduced in Sec. 3.2:

$$L(\boldsymbol{\theta}|\nu_{Oj}) = \prod_{O\subseteq\Omega}\prod_{j\in O}\theta_j^{\nu_{Oj}} = \prod_{j\in O}\theta_j^{\sum_{O\subseteq\Omega}\nu_{Oj}} \tag{21}$$

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 right hand side of Eq. (20).

3.4 Application to the 3-node fault tree

Applying Eq. (20) to our example, we immediately arrive at

$$[\underline{P}(F_C = 1|D), \overline{P}(F_C = 1|D)] \subseteq \left[\frac{n_C}{N+s}, \frac{n_A + n_B + n_C + s}{N+s}\right] \tag{22}$$

$$[\underline{P}(F_A = 1|D), \overline{P}(F_A = 1|D)] \subseteq \left[\frac{n_A + n_C}{N+s}, \frac{n_A + N_B + N_C + s}{N+s}\right] \tag{23}$$

For example, the lower bound for $\underline{P}(F_C = 1|D)$ 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 $\overline{P}(F_C = 1|D)$ corresponds to the case in which, for all failures of A, B failed as well, and vice versa.

The lower bound for $\underline{P}(F_A = 1|D)$ obtains when A never failed in case C^c, B, or B^c was observed. The upper bound 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.

If $N_C = 0$, then

$$[\underline{P}(F_C = 1|D), \overline{P}(F_C = 1|D)] \subseteq \left[0, \frac{n_A + n_B + s}{N_A + N_B + s}\right], \tag{24}$$

$$[\underline{P}(F_A = 1|D), \overline{P}(F_A = 1|D)] \subseteq \left[\frac{n_A}{N_A + N_B + s}, \frac{n_A + N_B + s}{N_A + N_B + s}\right]. \tag{25}$$

If $N_A = N_B = 0$, then

$$[\underline{P}(F_C = 1|D), \overline{P}(F_C = 1|D)] \subseteq \left[\frac{n_C}{N_C + s}, \frac{n_C + s}{N_C + s}\right], \tag{26}$$

$$[\underline{P}(F_A = 1|D), \overline{P}(F_A = 1|D)] \subseteq \left[\frac{n_C}{N_C + s}, 1\right]. \tag{27}$$

4 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 modeling dependencies correctly.

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

Next, we note that there is an alternative, and intuitively more appealing approach to arrive at the bounds given in Sec. 2: simple bounds can be attained by making extreme assumptions on the data, in this case on the numbers of non-failures observed for C, such as to keep the factorization of the joint posterior, which keeps the posterior integral simple to calculate. Such an approach can often be used to obtain conservative bounds quickly. Suppose one needs to make decisions with such posteriors as inputs, then a quick lower bound for the lower probability (and upper for the upper) might provide sufficient information on which to base a decision. For many monotone systems (which are such that reliability of the system never improves if that of a component becomes worse), such bounds can be derived pretty easily, although it is certainly not always as trivial as below.

For example, in our problem, posterior dependence between θ_A and θ_B occurred due to lack of detailed information on the components' failure in case of system observations which are not failures (there are $N_C - n_C$ such observations). Clearly, in this monotone system, the predictive posterior probability for the event $F_C = 1$ is decreasing as function of the numbers of failures observed, n_C, for fixed N_A, N_B, and N_C. Hence, a lower bound for this lower probability corresponds to the case where all $N_C - n_C$ non-failing system tests actually related to both components A and B not failing. Let us denote the data with this additional assumption as D_l, which therefore is equal to data with the information on components A and B being respectively $(N_A + N_C, n_A + n_C)$ and $(N_B + N_C, n_B + n_C)$. Hence, to arrive at a lower bound for the posterior probability of $F_C = 1$, we can use the distribution

$$\begin{aligned} p(\theta_A, \theta_B | D_l) &\propto \theta_A^{n_A + n_C + st_A - 1} (1 - \theta_A)^{N_A - n_A + N_C - n_C + s(1-t_A) - 1} \\ &\quad \times \theta_B^{n_B + n_c + st_B - 1} (1 - \theta_B)^{N_B - n_B + N_C - n_c + s(1-t_B) - 1} \\ &\propto p(\theta_A | D_{l,A}) \times p(\theta_B | D_{l,B}) \end{aligned}$$

where $D_{l,x} = (N_x + N_C, n_x + n_C)$ for $x = A$, B. Now we are back to the simple situation of posterior independence of the parameters θ_A and θ_B. Because the joint posterior distribution is just the product of two beta distributions, the corresponding posterior lower probability of failure of C is the product of the posterior lower probabilities of failure of A and B (attained at $t_A = t_B = 0$), leading to

$$\underline{P}(F_C = 1 | D) \geq P(F_C = 1 | D_l) = \frac{n_A + n_C}{N_A + N_C + s} \times \frac{n_B + n_C}{N_B + N_C + s}$$

Note that the approximation obtained in Eq. (10) is slightly tighter.

For the upper probability, we can do the same, but as the information on $N_C - n_C$ non-failing system tests can only imply, in the most pessimistic scenario (corresponding to upper probability of system failure for monotone systems), that for each of these observations either A or B failed, but not both, we must take all the following cases into account: assume that, for these successful system tests, actually A failed y times, and B failed $N_C - n_C - y$ times, with $y \in \{0, 1, \ldots, N_C - n_C\}$. This, again, results for every single value of y in posterior independence for θ_A and θ_B, and then we can maximize the resulting posterior predictive probability over y. Details are left to the reader.

Another 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. Identical components—when we know a priori that $\theta_A = \theta_B$—is clearly one important case of dependence. This could be studied either analytically along similar lines as in Sec. 2, or by making extreme assumptions as demonstrated above in this section if quick bounds are sufficient. 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 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 worked on most of the results reported in this paper. Both authors thank Scott Ferson for suggesting the fault tree problem studied in this paper, and Jean-Marc Bernard for useful comments that have improved the presentation of the paper.

A The Gamma function and Binomial Coefficients

A basic property of the Gamma function is that $\Gamma(x+1) = x\Gamma(x)$ for any real number $x > 0$. For any real number x and any natural number $n \geq 0$ define

$$\left\{ {n \atop x} \right\} = \frac{\Gamma(x+n)}{\Gamma(x)} = \begin{cases} (x+n-1)(x+n-2)\cdots x & \text{if } n \geq 1, \\ 1 & \text{otherwise.} \end{cases} \tag{A.1}$$

The Beta function and binomial function are defined as

$$B(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} \qquad \binom{x+y}{x} = \frac{\Gamma(x+y+1)}{\Gamma(x+1)\Gamma(y+1)}$$

If the right hand side in any of the above expressions has a zero denominator, the analytical extension is assumed. For example, if m and n are natural numbers, then

$$\binom{m-n-1}{m} = \begin{cases} (-1)^m \binom{n}{m} & \text{if } m \leq n, \\ 0 & \text{if } m > n. \end{cases} \tag{A.2}$$

We shall need the following properties:

$$B(x+1,y) = \frac{x}{x+y} B(x,y) \tag{A.3}$$

$$\left\{ {n+m \atop x+y} \right\} B(x+n, y+m) = \left\{ {n \atop x} \right\} \left\{ {m \atop y} \right\} B(x,y) \tag{A.4}$$

$$\left\{ {n \atop x+y} \right\} B(x+n, y) = \left\{ {n \atop x} \right\} B(x,y) \tag{A.5}$$

$$\left\{ {n \atop y-n} \right\} B(x+n, y-n) = \left\{ {n \atop x} \right\} B(x,y) \tag{A.6}$$

B Multinomial Sums

The following equality holds:

$$\left(\sum_{j=1}^{n} a_j\right)^m = \sum_{\mu_1+\mu_2+\cdots+\mu_n=m} \binom{m}{\mu_1,\mu_2,\ldots,\mu_n} \prod_{j=1}^{n} a_j^{\mu_j},$$

where

$$\binom{m}{\mu_1,\mu_2,\ldots,\mu_n} = \frac{m!}{\mu_1!\mu_2!\ldots\mu_n!}.$$

Also, recall that

$$\prod_{i=1}^{n}\left(\sum_{\mu=1}^{m_i} a_{i\mu}\right) = \sum_{\mu_1=1}^{m_1}\cdots\sum_{\mu_n=1}^{m_n}\prod_{i=1}^{n} a_{i\mu_i}$$

Therefore, it holds that

$$\begin{aligned}\prod_{i=1}^{\ell}\left(\sum_{j=1}^{n_i} a_{ij}\right)^{m_i} &= \prod_{i=1}^{\ell}\sum_{\mu_1+\cdots+\mu_{n_i}=m_i}\binom{m_i}{\mu_1,\mu_2,\ldots,\mu_{n_i}}\prod_{j=1}^{n_i} a_{ij}^{\mu_j},\\ &= \sum_{\mu_{11}+\cdots+\mu_{1n_1}=m_1}\cdots\sum_{\mu_{\ell 1}+\cdots+\mu_{\ell n_\ell}=m_\ell}\prod_{i=1}^{\ell}\prod_{j=1}^{n_i}\binom{m_i}{\mu_{i1},\mu_{i2},\ldots,\mu_{in_i}} a_{ij}^{\mu_{ij}}\end{aligned}$$

C Calculating the posterior probabilities

C.1 An integral

First, we find an analytic expression for the following integral:

$$\begin{aligned}&\int_0^1\int_0^1 x^{a-1}(1-x)^{\overline{a}-1}y^{b-1}(1-y)^{\overline{b}-1}(xy)^c(1-xy)^{\overline{c}}\mathrm{d}x\mathrm{d}y\\ &= \int_0^1 y^{b+c-1}(1-y)^{\overline{b}-1}\int_0^1 x^{a+c-1}(1-x)^{\overline{a}-1}(1-xy)^{\overline{c}}\mathrm{d}x\mathrm{d}y\end{aligned}$$

and using [11, p. 558, §15.3.1], with $\tilde{a} = -\overline{c}$, $\tilde{b} = a + c$, and $\tilde{c} = \overline{a} + a + c$,

$$= \int_0^1 y^{b+c-1}(1-y)^{\overline{b}-1}\frac{\Gamma(\tilde{b})\Gamma(\tilde{c}-\tilde{b})}{\Gamma(\tilde{c})}F(\tilde{a},\tilde{b};\tilde{c};y)\mathrm{d}y$$

where F is the hypergeometric function, and by [11, p. 558, §15.3.4]

$$= \int_0^1 y^{b+c-1}(1-y)^{\bar{b}-1}(1-y)^{-\tilde{a}} \frac{\Gamma(\tilde{b})\Gamma(\tilde{c}-\tilde{b})}{\Gamma(\tilde{c})} F(\tilde{a}, \tilde{c}-\tilde{b}; \tilde{c}; \frac{y}{y-1})\mathrm{d}y$$

and by [11, p. 556, §15.1.1],

$$= \frac{\Gamma(\tilde{b})\Gamma(\tilde{c}-\tilde{b})}{\Gamma(\tilde{c})} \frac{\Gamma(\tilde{c})}{\Gamma(\tilde{a})\Gamma(\tilde{c}-\tilde{b})} \sum_{m=0}^{\infty} \frac{\Gamma(\tilde{a}+m)\Gamma(\tilde{c}-\tilde{b}+m)}{\Gamma(\tilde{c}+m)}$$
$$\int_0^1 y^{b+c-1}(1-y)^{\bar{b}+\bar{c}-1} \frac{y^m}{(-1)^m(1-y)^m m!}\mathrm{d}y$$

and, now by [11, p. 258, §6.2.1]

$$\begin{aligned}
&= \frac{\Gamma(\tilde{b})}{\Gamma(\tilde{a})} \sum_{m=0}^{\infty} \frac{(-1)^m}{m!} \frac{\Gamma(\tilde{a}+m)\Gamma(\tilde{c}-\tilde{b}+m)}{\Gamma(\tilde{c}+m)} B(b+c+m, \bar{b}+\bar{c}-m) \\
&= \sum_{m=0}^{\infty} \frac{(-1)^m}{m!} \frac{\Gamma(\tilde{a}+m)}{\Gamma(\tilde{a})} \frac{\Gamma(\tilde{b})\Gamma(\tilde{c}-\tilde{b}+m)}{\Gamma(\tilde{c}+m)} B(b+c+m, \bar{b}+\bar{c}-m) \\
&= \sum_{m=0}^{\infty} (-1)^m \frac{\Gamma(m-\bar{c}-1+1)}{\Gamma(m+1)\Gamma(-\bar{c}-1+1)} B(\tilde{b}, \tilde{c}-\tilde{b}+m) B(b+c+m, \bar{b}+\bar{c}-m) \\
&= \sum_{m=0}^{\infty} (-1)^m \binom{m-\bar{c}-1}{m} B(a+c, \bar{a}+m) B(b+c+m, \bar{b}+\bar{c}-m)
\end{aligned}$$

and by Eq. (A.5), Eq. (A.6) and Eq. (A.2) (assuming $\bar{c}$ is a natural number),

$$= B(a+c, \bar{a}) B(b+c, \bar{b}+\bar{c}) \sum_{m=0}^{\bar{c}} \binom{\bar{c}}{m} \frac{\left\{ {m \atop \bar{a}} \right\}}{\left\{ {m \atop \bar{a}+a+c} \right\}} \frac{\left\{ {m \atop b+c} \right\}}{\left\{ {m \atop \bar{b}+\bar{c}-m} \right\}}$$

C.2 The posterior probabilities of failure of C and A

With $a = n_A + st_A$, $b = n_B + st_B$, $c = n_C$, $\bar{a} = N_A - n_A + s(1-t_A)$, $\bar{b} = N_B - n_B + s(1-t_B)$, and $\bar{c} = N_C - n_C$,

$$\begin{aligned}
P(F_C = 1|D) &= \int_0^1 \int_0^1 \theta_A \theta_B p(\theta_A, \theta_B|D) \mathrm{d}\theta_A \mathrm{d}\theta_B \\
&= \frac{\sum_{m=0}^{\bar{c}} \binom{\bar{c}}{m} B(a+c+1, \bar{a}+m) B(b+c+m+1, \bar{b}+\bar{c}-m)}{\sum_{m=0}^{\bar{c}} \binom{\bar{c}}{m} B(a+c, \bar{a}+m) B(b+c+m, \bar{b}+\bar{c}-m)}
\end{aligned}$$

and by Eq. (A.3)

$$= \frac{\sum_{m=0}^{\bar{c}} \binom{\bar{c}}{m} B(a+c, \bar{a}+m) B(b+c+m, \bar{b}+\bar{c}-m) \frac{a+c}{\bar{a}+a+c+m} \frac{b+c+m}{\bar{b}+b+\bar{c}+c}}{\sum_{m=0}^{\bar{c}} \binom{\bar{c}}{m} B(a+c, \bar{a}+m) B(b+c+m, \bar{b}+\bar{c}-m)}$$

and now applying Eq. (A.5) and Eq. (A.6)

$$= \frac{\displaystyle\sum_{m=0}^{\bar{c}} \binom{\bar{c}}{m} \frac{\left\{ {m \atop \bar{a}} \right\}}{\left\{ {m \atop \bar{a}+a+c} \right\}} \frac{\left\{ {m \atop b+c} \right\}}{\left\{ {m \atop \bar{b}+\bar{c}-m} \right\}} \frac{a+c}{\bar{a}+a+c+m} \frac{b+c+m}{\bar{b}+b+\bar{c}+c}}{\displaystyle\sum_{m=0}^{\bar{c}} \binom{\bar{c}}{m} \frac{\left\{ {m \atop \bar{a}} \right\}}{\left\{ {m \atop \bar{a}+a+c} \right\}} \frac{\left\{ {m \atop b+c} \right\}}{\left\{ {m \atop \bar{b}+\bar{c}-m} \right\}}} \tag{C.1}$$

$$= \frac{\displaystyle\sum_{m=0}^{N_C-n_C} \gamma_m^{A,B} \frac{n_A+st_A+n_C}{N_A+s+n_C+m} \frac{n_B+st_B+n_C+m}{N_B+s+N_C}}{\displaystyle\sum_{m=0}^{N_C-n_C} \gamma_m^{A,B}} \tag{C.2}$$

$$= \sum_{m=0}^{N_C-n_C} \tilde{\gamma}_m^{A,B} \frac{n_A+st_A+n_C}{N_A+s+n_C+m} \frac{n_B+st_B+n_C+m}{N_B+s+N_C} \tag{C.3}$$

and by symmetry also

$$= \sum_{m=0}^{N_C-n_C} \tilde{\gamma}_m^{B,A} \frac{n_B+st_B+n_C}{N_B+s+n_C+m} \frac{n_A+st_A+n_C+m}{N_A+s+N_C} \tag{C.4}$$

where we defined

$$\begin{aligned} \gamma_m^{A,B} &= \binom{\bar{c}}{m} \frac{\left\{ {m \atop \bar{a}} \right\}}{\left\{ {m \atop \bar{a}+a+c} \right\}} \frac{\left\{ {m \atop b+c} \right\}}{\left\{ {m \atop \bar{b}+\bar{c}-m} \right\}}. \\ &= \binom{N_C-n_C}{m} \frac{\left\{ {m \atop N_A-n_A+s(1-t_A)} \right\}}{\left\{ {m \atop N_A+s+n_C} \right\}} \frac{\left\{ {m \atop n_B+st_B+n_C} \right\}}{\left\{ {m \atop N_B-n_B+s(1-t_B)+N_C-n_C-m} \right\}} \end{aligned} \tag{C.5}$$

and

$$\tilde{\gamma}_m^{A,B} = \frac{\gamma_m^{A,B}}{\displaystyle\sum_{m'=0}^{N_C-n_C} \gamma_{m'}^{A,B}} \tag{C.6}$$

and $\gamma_m^{B,A}$ and $\tilde{\gamma}_m^{B,A}$ by swapping A and B in the above expressions.

Similarly,

$$
\begin{aligned}
P(F_A=1|D) &= \int_0^1\int_0^1 \theta_A p(\theta_A,\theta_B|D)\mathrm{d}\theta_A\mathrm{d}\theta_B \\
&= \frac{\sum_{m=0}^{\bar{c}}\binom{\bar{c}}{m}B(a+c+1,\bar{a}+m)B(b+c+m,\bar{b}+\bar{c}-m)}{\sum_{m=0}^{\bar{c}}\binom{\bar{c}}{m}B(a+c,\bar{a}+m)B(b+c+m,\bar{b}+\bar{c}-m)}
\end{aligned}
$$

and by Eq. (A.3)

$$
= \frac{\sum_{m=0}^{\bar{c}}\binom{\bar{c}}{m}B(a+c,\bar{a}+m)B(b+c+m,\bar{b}+\bar{c}-m)\frac{a+c}{\bar{a}+a+c+m}}{\sum_{m=0}^{\bar{c}}\binom{\bar{c}}{m}B(a+c,\bar{a}+m)B(b+c+m,\bar{b}+\bar{c}-m)}
$$

and hence, along similar lines,

$$
= \sum_{m=0}^{N_C-n_C}\tilde{\gamma}_m^{A,B}\frac{n_A+st_A+n_C}{N_A+s+n_C+m} = \sum_{m=0}^{N_C-n_C}\tilde{\gamma}_m^{B,A}\frac{n_A+st_A+n_C+m}{N_A+s+N_C} \tag{C.7}
$$

The result for $P(F_B=1|D)$ follows by swapping A and B in the above expression.

C.3 Special case

In this section we consider the special case $N_A=N_B=0$, $N_C>0$, and $s\to 0$. If $N_C>n_C>0$, we have

$$
\begin{aligned}
\lim_{s\to 0}\gamma_m^{A,B} &= \lim_{s\to 0}\binom{N_C-n_C}{m}\frac{\left\{{m \atop s(1-t_A)}\right\}}{\left\{{m \atop s+n_C}\right\}}\frac{\left\{{m \atop st_B+n_C}\right\}}{\left\{{m \atop s(1-t_B)+N_C-n_C-m}\right\}} \\
&= \begin{cases} 1 & \text{if } m=0 \\ \frac{1-t_A}{1-t_B} & \text{if } m=N_C-n_C \\ 0 & \text{otherwise,} \end{cases}
\end{aligned}
$$

if $n_C=0$, then

$$
\lim_{s\to 0}\gamma_m^{A,B} = \begin{cases} 1 & \text{if } m=0 \\ \frac{1-t_A}{1-t_B}t_B & \text{if } m=N_C \\ 0 & \text{otherwise,} \end{cases}
$$

and finally, if $n_C=N_C$, then $\gamma_0^{A,B}=1$.

Hence, in the limit $s \to 0$, if $N_C > n_C > 0$, from Eq. (C.2) we deduce that

$$\lim_{s\to 0} P(F_C = 1|D) = \frac{\frac{n_C}{n_C}\frac{n_C}{N_C} + \frac{1-t_A}{1-t_B}\frac{n_C}{N_C}\frac{N_C}{N_C}}{1 + \frac{1-t_A}{1-t_B}} = \frac{n_C}{N_C}, \tag{C.8}$$

and similarly from Eq. (C.7)

$$\lim_{s\to 0} P(F_A = 1|D) = \frac{\frac{n_C}{n_C} + \frac{1-t_A}{1-t_B}\frac{n_C}{N_C}}{1 + \frac{1-t_A}{1-t_B}} = \frac{(1-t_B) + (1-t_A)\frac{n_C}{N_C}}{2 - t_A - t_B} \tag{C.9}$$

If $n_C = 0$, then $\lim_{s\to 0} P(F_C = 1|D) = 0$, and

$$\lim_{s\to 0} P(F_A = 1|D) = \frac{t_A + 0}{1 + \frac{1-t_A}{1-t_B} t_B} = \frac{t_A - t_A t_B}{1 - t_A t_B} \tag{C.10}$$

If $n_C = N_C$, then $\lim_{s\to 0} P(F_C = 1|D) = \lim_{s\to 0} P(F_A = 1|D) = 1$.

References

[1] P. Walley, Statistical Reasoning with Imprecise Probabilities, Chapman and Hall, London, 1991.

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

[3] W. E. Vesely, F. F. Goldberg, N. H. Roberts, D. F. Haasl, Fault Tree Handbook, U. S. Nuclear Regulatory Commision, Wachington D. C., 1981, NUREG-0492.

[4] M. S. Hamada, H. Martz, C. S. Reese, T. Graves, V. Johnson, A. Wilson, 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 (2004) 297–305.

[5] F. P. A. Coolen, An imprecise Dirichlet model for Bayesian analysis of failure data including right-censored observations, Reliability Engineering and System Safety 56 (1997) 61–68.

[6] L. V. Utkin, S. V. Gurov, New reliability models based on imprecise probabilities, in: C. Hsu (Ed.), Advanced Signal Processing Technology, World Scientific, 2001, Ch. 6, pp. 110–139.

[7] L. Utkin, Cautious reliability analysis of multi-state and continuum-state systems based on the imprecise Dirichlet model, International Journal of Reliability, Quality and Safety Engineering 13 (5) (2006) 433–453.

[8] L. V. Utkin, F. P. A. Coolen, Imprecise reliability: an introductory overview, in: Computational Intelligence in Reliability Engineering, Volume 2: New Metaheuristics, Neural and Fuzzy Techniques in Reliability, Springer, 2007, pp. 261–306.

[9] G. de Cooman, M. Zaffalon, Updating beliefs with incomplete observations, Artificial Intelligence 159 (1–2) (2004) 75–125.

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

[11] M. Abramowitz, I. A. Stegun, Handbook of mathematical functions, 10th Edition, National Bureau of Standards, Washington D.C., 1972.