ESTIMATING THE MASS OF THE HIGGS PARTICLE USING DEMPSTER-SHAFER ANALYSIS

BY PAUL T. EDLEFSEN, CHUANHAI LIU AND ARTHUR P. DEMPSTER

Harvard University and Purdue University

We present a Dempster-Shafer (DS) approach to finding confidence bounds on the mass of the Higgs boson. Dempster-Shafer is a statistical framework that generalizes Bayesian statistics. DS calculus augments traditional probability by allowing mass to be distributed over power sets of the event space. This eliminates the Bayesian dependence on prior distributions while allowing the incorporation of prior information when it is available. We use the Poisson Dempster-Shafer model (DSM) to derive a posterior DSM for the Banff three-Poisson model, from which we make inferences about the unknown mass of the Higgs particle. The results compare favorably with other approaches, demonstrating the utility of the approach. We argue that the reduced dependence on priors afforded by the Dempster-Shafer framework is both practically and theoretically desirable.

1. Introduction.

1.1. The Higgs Particle. The Higgs boson is the only Standard Model (SM) subatomic particle not yet observed. The mass of the particle, if the particle exists, has profound implications for particle physics and for cosmology. Theoretical considerations place the mass somewhere between about 130 and 190 GeV. Previous experimental results suggest that the mass is somewhere between 65 GeV to 186 GeV (Igo-Kemenes, 2006). If the mass is below 130 GeV, new physics would be required to explain the phenomenon. If the boson does not exist, then the fundamental source of mass in the Universe would not be explained by the Standard Model.

Experiments to determine the mass of the Higgs boson involve expensive equipment that is relatively short-lived. At present the most promising apparatus is the Large Hadron Collider (LHC) at CERN, a collaboration of over two thousand physicists from 34 countries, which is expected to become operational in 2008 and to operate for about a decade. Its total cost will be around 8 billion US dollars. When in operation, about 7000 physicists from 80 countries will have access to the LHC, and the data analysis project is expected to involve many more scientists and hobbyists.

Keywords and phrases: Bayesian, belief function, evidence theory, Poisson, high-energy physics, Higgs boson
The experiments do not directly yield the mass of the Higgs boson. Instead, they assess the presence of the measurable electroweak particles into which the Higgs decays. Each measured electroweak particle is called a channel, and the fraction of the Higgs particle that decays into each channel is called the branching ratio. Physical theory states that the branching ratio is a function of the mass of the decaying particle (See Figure 1).

By measuring each channel in the presence of a particle collision with sufficient energy to produce a Higgs particle, and again in the absence of such a collision, the physicists obtain data from which the mass of the Higgs particle might be inferred. Combining these signal plus background ($s+b$) and background ($b$) measurements across channels using a statistical model, confidence intervals may be obtained for the true mass $m_{H^0}$ of the SM Higgs boson.

Unfortunately the choice of the correct model is not entirely clear, nor is the correct statistical methodology clear, given a particular model. In fact the data are processed and filtered substantially during the detection process, further complicating the analysis.

Given the importance of the science and the cost of the data acquisition process, any determination of the mass of the Higgs particle is sure to be debated and its analysis methods scrutinized. Groups of scientists have already
begun to explore, through simulation studies and theoretical arguments, the
certainty that might justifiably be attributed to any future conclusion.

1.2. The BIRS A1 Limits Project. In July 2006, a workshop on statisti-
cal inference problems in high energy physics and astronomy was held at
the Banff International Research Station (BIRS), bringing together physi-
cists, astronomers, and statisticians to “bring the latest methods to the
attention of the scientific community, and to develop statistical theory fur-
ther by considering special aspects that arise in these scientific contexts”
(Linneman et al, 2006). One of the primary objectives of the workshop was
to address issues of statistical significance in the presence of nuisance pa-
rameters such as those that arise in the determination of the mass of the
Higgs particle. By the end of the workshop the participants decided that the
various methods for determining confidence limits on the Higgs mass should
be allowed to compete in a simulated experiment (Linneman et al, 2007).
This open challenge is called the BIRS A1 Limits Project (Heinrich, 2006a).

The challenge is somewhat abstracted from the underlying physics. Data
are provided as counts from three related Poisson processes: one count from
a process that includes only the background $b$; one count from a process of
the “efficiency” $\epsilon$, which is some function of the branching ratio and other
factors such as the running time and the accelerator beam intensity; and
a third count from a process of the combined signal $s$, efficiency $\epsilon$, and
background $b$. As further described below, single-channel and 10-channel
data are provided in a total of 3 tasks. Respondents to the challenge provide
90% and 99% upper (one-sided) confidence bounds on $s$, given the 3 counts
per channel.

Previous work on this model include a Bayesian approach for one chan-
nel in Heinrich et al (2004); Demortier (2005) and for multiple channels
in Heinrich (2005), a Profile Likelihood approach in Rolke et al (2005),
a frequentist-Bayesian hybrid approach in Conrad and Tegenfeldt (2006),
and fully frequentist approaches in Punzi (2005); Cranmer (2003). Prior to
this challenge, these approaches have never been compared using the same
datasets or criteria. No attempt has been made previously to use a hierar-
chical Bayesian approach, or to use Dempster-Shafer analysis. As a result
of the present challenge, all of these approaches may now be compared on
common ground.

1.3. Dempster-Shafer. Our submission uses Dempster-Shafer (DS) the-
tory (Dempster, 1968a; Shafer, 1976; Dempster, 2007) to construct a Bayesian-
style posterior using no priors. Dempster-Shafer calculus assigns probability
mass to elements of the powerset of a state space rather than to elements of
the state space itself. In the context of real-valued parameters, DS models are maps from *ranges* (and sets of ranges) of the parameter space to real numbers, rather than maps from single parameter values to real numbers as is the typical case in frequentist and Bayesian calculi. As such these standard approaches are subsets of the DS calculus.

The principal benefit of the DS approach over the Bayesian is that in DS analysis, information may be transmitted from margins of the space (in the present example, from the nuisance parameters $c$ and $b$) to the joint space without requiring the use of a prior. The Bayesian calculus requires a prior because the map from the marginal space to the joint space must be one-to-one. In DS calculus it may be multivalued, so that the distribution of the smaller space over the larger space need not be specified.

The principal drawback, from a frequentist or classical Bayesian perspective, of the DS calculus is that the interpretation of a distribution over ranges (or more generally over sets of the state space) is awkward for those familiar with a single-valued framework. In this paper we map the distribution over ranges into a distribution over single values using the so-called plausibility transform (Cobb and Shenoy, 2006). This allows us to return a single value for use in comparing the method to other Bayesian and frequentist approaches.

One benefit of the DS approach is that, although priors are not required, prior information can be incorporated as easily as in Bayesian analyses. In the DS approach, however, as many or as few priors may be incorporated as there is prior information to incorporate. For instance, it may be the case that strong prior information exists for one of the nuisance parameters, but not for the other. This would easily be accommodated using the DS approach. If Bayesian priors are provided for all three of the parameters, this approach is exactly the same as the corresponding Bayesian approach. Priors may also be provided as DS models (typically referred to as “belief functions” in the literature, following Shafer (1976)), if the prior knowledge is better represented this way.

According to Banff Challenge organizer Joel Heinrich in Heinrich (2006c), “Subjective informative priors for the parameter of interest are very unpopular. We have no confidence in our own opinion, for example, of the mass of the Higgs particle, nor in anyone else’s opinion. Therefore, even subjective priors are invariably uninformative for the parameter of interest. But priors for some nuisance parameters are, in some cases, both subjective and informative – a problem for frequentists.” We argue that the DS approach is ideally suited for situations in which prior information exists for some, but not all, parameters of a model.
2. Problem Statement. In this section we introduce the model and the tasks of the challenge. The official problem statement may be found in Heinrich (2006b) and on the BIRS A1 Limits Project website (Heinrich, 2006a).

2.1. The Model. The model is, for channel $i \in 1 \ldots N$,

$$
n_i \sim \mathcal{P}ois(\epsilon_is + b_i)$$
$$y_i \sim \mathcal{P}ois(t_ib_i)$$
$$z_i \sim \mathcal{P}ois(u_i\epsilon_i).$$

The constants $t_i$ and $u_i$ are given, as well as the counts $n_i$, $y_i$, and $z_i$. The parameter of interest is $s$, with $\epsilon_i$ and $b_i$ considered nuisance parameters. The challenge is to find one-sided 90% and 99% confidence intervals (upper bounds) for $s$.

Three datasets were generated by Joel Heinrich, each corresponding to a separate task of the challenge (Heinrich, 2006a). The data provided for task 1a consist of 60229 independent sets of $(n, y, z)$ counts drawn from the model with a single channel ($N = 1$). In all cases, $t = 33$ and $u = 100$. The data provided for task 1b consist of 39700 permutations of single-channel counts for $n = 0, \ldots, 49$, $z = 0, \ldots, 30$, and $y = y_1, \ldots, y_u$, for $y_i$ varying from 0 to 7, and $y_u$ varying from 13 to 24, depending on $z$. In all cases, $t = 3.3$ and $u = 10$. For this task, the values of $t$ and $u$ are sufficiently small that the 39700 given permutations include every combination of $(n, y, z)$ for which the summand in equation (2.1) is non-negligible for the 101 evaluated values of $s$. The data provided for task 2 consist of 70000 independent sets of $(n_1, x_1, y_1, \ldots, n_{10}, x_{10}, y_{10})$ counts drawn from the model with ten channels ($N = 10$). $t_i = 15 + 2i$ and $u_i = 53 + 2i \forall i \in 1 \ldots 10$.

2.2. Evaluation. The question of the appropriate evaluation procedure is discussed at length in Heinrich et al (2004). The primary metric used for this challenge is the actual coverage, discussed there among several other possibilities.

All three tasks were evaluated by calculating the actual coverage $C(s)$ of the submitted 90% and 99% intervals. Specifically, the actual coverage was calculated for each of 101 different values of $s$ (known only to the evaluator until the all submissions were received), using the formula

$$C(s) = \sum_{(n, y, z) \in \mathcal{R}(n, y, z)} \frac{e^{-\mu} \mu^n e^{-\nu} \nu^y e^{-\rho} \rho^z}{n! y! z!},$$

where

$$(2.1) \quad \mathcal{R}(n, y, z) = \{ (n, y, z) : n, y, z = 0, \ldots, N, n + y + z = \text{constant} \}.$$
where

\[ \mu = \epsilon s + b \]
\[ \nu = tb \]
\[ \rho = \alpha \epsilon, \]

and \( R(n, y, z) \) is the upper bound result submitted by the contender. The values of the nuisance parameters for this calculation were only revealed after the challenge; they were \( \epsilon = 1 \), \( b = 3 \) for tasks 1a and 2, and \( \epsilon = .1 \) and \( b = .31 \) for task 1b.

In theory the sum should be taken over all possible counts \( (n, y, z) \) for which the corresponding results \( R(n, y, z) \) satisfy \( s < R(n, y, z) \). Since in task 1a the 60229 provided counts \( (n, y, z) \) are only a subset of these, the evaluator actually reported only an estimate of \( C(s) \), along with the standard error of that estimate. Task 1b was evaluated in the same manner as was task 1a, except that the coverages were computed exactly (to the reported precision of 4 decimal places), rather than estimated. Task 2 was evaluated in the same manner as task 1a.

3. Dempster-Shafer Theory. Dempster-Shafer theory extends traditional Bayesian/frequentist statistics by appending a third category “don’t know” to the familiar dichotomy “it’s true”/“it’s false”. The theory assigns to any assertion a probability \( p \) for that assertion, a probability \( q \) against that assertion, and a third probability \( r = 1 - p - q \) that remains effectively unassigned. Although the actual state of the described phenomenon is understood to be constrained such that the assertion is either true or false, DS theory allows you, the observer, to describe your evidence for and against the assertion without the traditional constraint that all such evidence be construed unambiguously. The remaining probability, \( r \), represents your residual uncertainty after assessing the available evidence.

The mathematical framework of DS theory is that of random sets. From this perspective, if \( p \) is the probability of the event \( T \) (that the assertion is true), and \( q \) is the probability of the event \( F \) (the assertion is false), then \( r \) is the probability of the set \( \{T, F\} \) (“don’t know”). In general, the DS mass function \( m(A) : 2^S \to [0, 1] \) is mathematically indistinct from a probability measure over an extended state space (the power set of the event space \( S \)). DS theory combines the logic of set theory with this random sets framework to yield a powerful calculus for reasoning about uncertainty.

Defining an assertion \( A \subset S \) as a set of events (in words, “the true state of the described phenomenon is in the set \( A \)”), then the accumulated evidence for that assertion is given by \( p(A) = \sum_{B \subseteq A} m(B) \). That is, it includes
evidence that is unambiguous ("the true state is \( e \), an element of \( A \)) and evidence that is ambiguous ("the true state is in the set \( A' \), a subset of \( A \) with cardinality greater than 1"). The total evidence against the assertion \( A \) is given by \( q(A) = p(A^c) = \sum_{C \subseteq A : C \neq \emptyset} m(C) \). Any evidence that is ambiguous with respect to the assertion \( A \) is accumulated in the residual uncertainty \( r(A) = 1 - p(A) - q(A) \), which is the sum of the evidence on sets that overlap both \( A \) and \( A^c \).

When all evidence is unambiguous, DS theory coincides completely with Bayesian (and frequentist) statistics, with \( p(A) = \sum_{e \in A} \mathbb{P}(e) = 1 - q(A) \). What the DS framework adds is the ability to tolerate ambiguous evidence, which is particularly useful when describing the joint distribution of non-independent margins: with the Bayesian constraint that \( r(\cdot) = 0 \), marginal evidence must be combined with evidence or assumptions about conditional distributions when extending that evidence to a joint state space. The DS framework allows the observer to remain agnostic when extending evidence on margins into a joint space. For instance if two non-independent Bernoullis (with respective marginal probabilities \( p_1 \) and \( p_2 \) of states \( T_1 \) and \( T_2 \)) are described by a joint Dempster-Shafer model (DSM), the evidence \( p_1 \) on \( T_1 \) in the first margin is, in the joint model, construed as evidence on the set \( \{(T_1, F_2), (T_1, T_2)\} \). The Bayesian constraint would require the ambiguity between the constituent states \( (T_1, F_2) \) and \( (T_1, T_2) \) to be resolved before analysis could proceed.

This feature of the DS theory can be exploited by Bayesians wishing to minimize dependence on convenience priors. Historically, post-analysis depictions of uncertainty about model parameters of interest have been restricted to point estimates and confidence regions. Bayesian posterior distributions are richer depictions that can be summarized with point estimates and credibility regions as needed. DS models are richer still, and just as full Bayesian posteriors have gained acceptance with the passage of time and the improvement of mathematical and computational tools to store them and compute with them, we predict that full DSMs will ultimately gain acceptance as intermediate and final products of statistical analysis. In the interim it is nevertheless convenient for Bayesians who wish to report standard posterior distributions, or for frequentists who wish to report confidence regions, to combine the prior-free analysis summarized by a DSM with a non-ambiguity constraint, as we do for the Banff challenge. This approach provides all of the benefits of the Bayesian paradigm with a great deal of additional flexibility and a reduced dependence on priors.

The principal operations of the Dempster-Shafer calculus (DSC) involve extending marginal evidence to a joint space, combining the evidence in
the joint space, and projecting from a joint space to a margin. Evidence is represented using DSMs, which as previously discussed are essentially probability measures over power sets of state space models (SSMs). In our previous example of two Bernoulli SSMs, the joint DSM was created by first extending the Bernoulli distributions on the two margins to the joint state space (yielding two DSMs with mass functions mapping from the power set of the 4-element joint SSM). Assuming that the evidence yielding the values \( p_1 \) and \( p_2 \) can be considered mutually non-comprimising (that is, our evidence that the probability of \( T_1 \) is \( p_1 \) is independent of our evidence that the probability of \( T_2 \) is \( p_2 \)), combination is a straightforward multiplication operation, generalizing the Bayesian operation of multiplying likelihoods.

The combination operation is always performed with two DSMs over the same SSM (in our Bernoulli case, we first extend the marginal DSMs to the joint SSM, then combine). If \( C \) is the DSM that is the result of combining DSMs \( D_1 \) and \( D_2 \) (written \( C = D_1 \oplus D_2 \)), then the combined evidence \( m_C(A) \) on any set \( A \) is given by accumulating relevant portions of evidence from the two constituent DSMs:

\[
m_C(A) = \sum_{A_1,A_2:A_1 \cap A_2 = A} m_{D_1}(A_1)m_{D_2}(A_2).
\]

This operation is most conveniently performed using the “commonality” set function \( c(A) = \sum_{B \subseteq A} m(B) \), since \( c_{D_1 \oplus D_2}(A) = c_{D_1}(A)c_{D_2}(A) \). Thus the DS combination operation is simply computed by multiplying commonalities of like sets, generalizing the Bayesian combination operation of multiplication of probability masses (or mass densities) on like elements of the SSM. As with the corresponding Bayesian computation, the resulting DSM will usually be normalized to remove mass on the empty set, although normalization may be postponed to the end of the analysis, or avoided altogether if proportional values are sufficient.

The three alternate representations of a DSM (the mass function \( m(\cdot) \), the “belief” function \( p(\cdot) \), and the commonality function \( c(\cdot) \)) are interchangeable and have a unique correspondence. Glenn Shafer, who coined many of the terms of DS theory, including “belief” and “commonality”, described in his foundational book (Shafer, 1976) elegant Möbius transforms for passing back and forth among the alternate representations. Subsequent work by Thoma (1989, 1991) described a fast Möbius transform analogous to the famous FFT (see Kennes, 1992, for a thorough proof and discussion).

A useful conceptualization of Dempster-Shafer Models is that of the multivalued map. Dempster’s earliest work on the subject (1967a; 1967b; 1968a; 1968b) described the theory in this way, and more recent work by Kohlas,

imsart-aoas ver. 2007/12/10 file: DSBanff.tex date: December 31, 2007
Monney, and others (see Kohlas and Monney, 1994, for a review) has emphasized this perspective, in which a DSM’s mass function \(m(\cdot)\) is derived from an associated probability distribution \(\mathbb{P}\) over an auxiliary state space \(\Omega\) and a function \(\Lambda : \Omega \mapsto 2^S\). Then \(m(A) = \sum_{\omega \in \Omega: \Lambda(\omega) = A} \mathbb{P}(A)\). In what follows we will use the prefix “\(\Lambda\)” when referring to the associated probability model, or to any auxiliary probability distribution that is used to characterize a DSM.

A simple example that illustrates DS analysis is the case of repeated flips of a single coin. The Binomial DSM for \(n\) trials is a joint DS model over the state space of \(k \in \{0, \ldots, n\}\) and \(p \in [0,1]\). We may summarize our evidence about \(p\) as the projection of this joint DSM onto the \(p\) margin. We can condition on the observation of \(k\) heads by first combining the joint DSM with a deterministic one, effectively placing 0 mass on any set in the joint space that contradicts the observed number of heads. An obvious associated probability model is the Uniform distribution: if \(X \sim U(0,1)\), then the indicator that \(X \leq p\) has a Bernoulli distribution, so if we define \(\Lambda_h(x) = \{p' : p' \geq x\}\), then the tuple \((\Omega = [0,1], \mathbb{P} = f_U(\cdot), \Lambda = \Lambda_h, S = [0,1])\) defines a DSM over the state space \(S\) of possible values of \(p\) when we have observed a coin flip and it is heads (if it were tails, we would use \(\Lambda_t(x) = \{p' : p' < x\}\)). Such a tuple, called a “hint” by Kohlas and Monney (1995), summarizes our evidence about the unknown parameters of interest implied by an observation. Combining \(n\) such hints yields the projection of the conditioned Binomial DSM onto the \(p\) margin. This could be equivalently expressed as a single hint with an \(n\)-dimensional Uniform associated probability model \((x_i \sim U(0,1)\) for \(i \in 1..n\)) and a multivalued map \(\Lambda(x_1,..x_n) = \{p' : p' \geq x_i \Rightarrow i \in H\}\) (where \(H\) is the index set of coins that came up heads). In words, if we observe that \(k\) of \(n\) coin flips are heads, then \(p\) is somewhere between the \(k\)th and \((k+1)\)th ordered Uniforms. That is, for any interval \(A := (l,u)\), the mass function \(m(A)\) is proportional to the joint density \(f(l,u)\) of the \(k\) and \(k+1\) order statistics of \(n\) independent Uniforms.

Note that the Binomial DSM describes \(p\) as contained in an a-random interval \((L,U)\) \((L \leq p \leq U)\), where \(L \sim \text{Beta}(k,n)\) and \((U - L) \sim \text{Beta}(1,n)\). The “a-random” quantities \(L\) and \(U\) do not describe the distribution of \(p\) in the usual sense. In order to obtain a distribution (a “precise” DSM) for \(p\), all of the mass in the DSM must be restricted to the singleton sets. Combining any DSM with any precise DSM will yield a precise DSM. The Bayesian practice of combining a likelihood with a prior distribution, for instance, ensures that the resulting posterior is precise. Combining the Binomial DSM for \(p\) with a Bayesian prior leads to the same result as the corresponding Bayesian analysis. A Uniform prior, for instance, yields a \(\text{Beta}(k+1, n-k+1)\)
posterior.

Combination with an uninformative (uniform) prior is an example of the “plausibility transform” (Cobb and Shenoy, 2006), a method for transforming between a DSM and a probability distribution, which takes \( P(e) \propto c(\{e\}) \) for all elements \( e \in S \). Glenn Shafer coined the term “plausibility” to refer to the total evidence not in contradiction with an assertion: \( \text{Plaus}(A) = 1 - q(A) = 1 - p(A^c) = p(A) + r(A) \). For singleton sets, \( \text{Plaus}(\{e\}) = c(\{e\}) \). In the Binomial DSM example, the plausibility transform yields \( P(e) \propto P_{L,U}(e \in [L,U]) \), which turns out to be the \( \text{Beta}(k + 1, n - k + 1) \) density.

3.1. The Poisson DSM. Our solution to the Banff challenge uses the Poisson DSM, which is the Poisson analogue to the Binomial DSM. The full state space model of the Poisson DSM is the cross of the natural numbers \( N \) (for the count, \( X \)) with the non-negative Reals \( \mathbb{R}^{0+} \) (for the rate, \( L \)). Figure 2 depicts the full SSM. Conditioning on \( L = \lambda \), the \( X \) margin has a \( \text{Pois}(\lambda) \) distribution (a precise DSM). For inference about \( L \) we are concerned with the DSM on the \( L \) margin after conditioning on an observed count \( X = k \).

The Poisson DSM is defined mathematically by assigning a mass distribution over a-random subsets of its \((L,X)\) state space. These subsets are dete-
Fig 3. A typical a-random subset is the union of intervals at levels \( X=0,1,2,\ldots \) (from Dempster, 2007)

mined by an auxiliary sequence of a-random points \( 0 \leq V_1 \leq V_2 \leq V_3 \leq \ldots \) on the \( L \) axis. As illustrated in Figure 3, the auxiliary sequence \( V_1, V_2, V_3, \ldots \) defines a corresponding sequence of intervals \( 0 \leq V_1, V_1 \leq V_2, V_2 \leq V_3, \ldots \) at respective levels \( X = 0, 1, 2, \ldots \). The union of these intervals becomes an a-random set in the state space \((L, X)\) when the lengths of the intervals are independently and identically distributed with the unit scale exponential density \( \exp(-u) \) for \( u \geq 0 \).

Note that when you condition on a fixed \( L = \lambda \) (by combining the Poisson DSM with a deterministic DSM with mass 1 on the set \( \{(\lambda, x) : x \in \mathcal{N}\} \)), the resulting DSM does indeed yield a precise Pois(\( \lambda \)) margin for \( X \), since the number of unit exponential intervals that elapse in time \( \lambda \) is described by a Poisson process with rate 1. Conditioning instead on a fixed count \( k \) restricts the mass on the resulting DSM to the line corresponding to \( X = k \). From Figure 3 we see that this places \( L \) in the range \((V_k, V_{k+1})\) (letting \( V_0 := 0 \) for notational convenience). This characterization of inference about \( L \) was first given in Almond (1989, 1995).

The left end \( V_k \) of this a-random interval is defined by the sum of \( k \) independent a-random unit-scale exponentials, and hence has a unit-scale Gamma distribution with shape \( k \). The length of the interval \((V_{k+1} - V_k)\) is
independently exponentially distributed, and the \((V_k, V_{k+1})\) pair are jointly distributed as the \(k\) and \(k+1\) transition times of a unit-rate Poisson process. In words, observing that a unit-rate Poisson process has transitioned \(k\) states in an unknown amount of time \(\lambda\) provides evidence about \(\lambda\) in the form of bounds: since in \(\lambda\) time, \(k\) exponentials elapsed, \(\lambda > V_k\), where \(V_k \sim \text{Gamma}(k)\), and since the next exponential has not yet elapsed, \(\lambda < V_{k+1}\). The joint distribution of \((V_k, V_{k+1})\) is characterized by the formula

\[
P(V_k \leq u, V_{k+1} \geq v) = \frac{1}{k!} u^k \exp(-v), \forall v \geq u \geq 0.
\]

In standard probability terms, this is a form of the bivariate cumulative distribution of the ends of the a-random interval \((V_k, V_{k+1})\). In DS terms, however, it is the commonality function \(c(u, v)\) of the interval \((u, v)\) for the posterior DSM of \(L\) given the observation \(X = k\).

Note that \(\lambda \in (V_k, V_{k+1})\) only when \(V_k \leq \lambda\) and \(V_{k+1} \geq \lambda\), so \(\lambda \notin (V_k, V_{k+1})\) whenever \(V_{k+1} < \lambda\) or \(V_k > \lambda\). These are mutually exclusive events, so

\[
\text{Plaus}((\lambda)) = c(\lambda, \lambda) = \mathbb{P}(V_k, V_{k+1}) (\lambda \in (V_k, V_{k+1})))
= 1 - \mathbb{P}(\lambda \notin (V_k, V_{k+1}))
= 1 - (F_{V_{k+1}}(\lambda) + (1 - F_{V_k}(\lambda)))
= F_{V_k}(\lambda) - F_{V_{k+1}}(\lambda).
\]

3.2. Join Trees. In the Banff challenge, the counts \(y_i\) and \(z_i\) are each from scaled Poisson distributions with known, constant scale factors \(t_i\) and \(u_i\), respectively. There are at least three ways to extend the unscaled Poisson DSM as described above for use with scaled Poissons. Perhaps the simplest is to argue via representation that if an observation of \(Y' = y'\) from a \(\text{Po}(b)\) yields a posterior DSM on the \(B\) margin with the commonality function described above, which bounds \(b\) by \(V_{y'}\) and \(V'_{y'+1}\), then an observation \(Y = y\) from a \(\text{Pois}(tb)\) should yield a posterior DSM bounding \(b\) by \(V_y\) and \(V_{y+1}\).

Another approach would be to extend the definition of the Poisson DSM such that the auxiliary sequence \(V_1, V_2, V_3, \ldots\) are separated by exponentials with scale \(\frac{1}{\lambda}\). In this section we introduce a third approach that uses join trees, which are a fundamental component of DS analysis (DSA) that greatly simplify the process of computing with DSMs.

If we define \(L_y := tb\), then (when the observed count is \(Y = y\)) the Poisson DSM as described in the previous section provides a posterior inference about \(L_y\) in the form of a DSM with commonality function

\[
c(u, v) = \frac{1}{y!} u^y \exp(-v), \forall v \geq u \geq 0.
\]
If we treat this as a DSM on the \( L_y \) margin of the larger SSM that also includes the \( T \) and \( B \) states, then we can extend this marginal DSM to the full SSM, and likewise extend the constant DSM \( T = t \), combine them there and then project the resulting DSM to the \( B \) margin to yield an inference about \( b \).

The join tree theorem of DS analysis due to Shenoy and Shafer (1986) or Kong (1986) states that we may (as just described) first project the Poisson DSM over the \((L_y, Y)\) SSM to the \( L_y \) margin, then extend from that margin up to the joint \((L_y, T, B)\) SSM, combine there with the DSM about \( T \) and then project to the \( B \) margin (this process is called “propagation”). We do not need to extend all margins up to the full joint state space (which would include the \( Y, L_y, T, \) and \( B \) states) and combine there, since all relevant information on the \( Y \) margin is contained in the projection of \((L_y, Y)\) onto \( L_y \).

Figure 4 depicts the join tree for this example. A join tree (also called a Markov tree) is a hypergraph depicting nodes for each constituent margin of the SSM and hyperedges corresponding to each hint. In this example the hints are

- the Poisson DSM, relating \( L_y \) and \( Y \),
- the observed count \( Y = y \),
- the relationship given by the definition \( L_y := T \times B \), and
• the known scale $T = t$.

Note that if one hint refers to a group of margins $J$, and another refers to a group $K$, and $J \subseteq K$, then the two hints can be combined on the $K$ SSM. For example the observed count $y$ and the Poisson DSM may be combined on the $(L_y,Y)$ SSM. If we define $\mathcal{J}$ as those hyperedges remaining after all of these trivial combinations are performed, then the hyperedges in the “scheme” $\mathcal{J}$ are said to be nodes of a join tree if it is possible to define edges $\mathcal{E}$ (pairs of hyperedges in the scheme) such that

• $(\mathcal{J}, \mathcal{E})$ is a tree, and

• if $J_i$ and $J_k$ are distinct vertices of the tree $(\mathcal{J}, \mathcal{E})$, then $J_i \cap J_k$ is contained in every vertex on the unique path from $J_i$ to $J_k$ in the tree.

The join tree for this example contains only two nodes, one formed from the Poisson DSM and one from the definition relating $L_y$ to $B$ and $T$. The unique path between them includes no other nodes. If it did, then their intersection $\{L_y\}$ would need to be contained in every node on that path.

The join tree theorem states that when such a tree exists, then DSA can be performed by propagating evidence from leaves of the join tree towards the node(s) containing the margin of interest (in this case, $B$). The full joint space of all states need never be constructed. In the present example this is a small convenience, but in larger problems (such as the Banff Challenge), the join tree theorem provides a significant reduction in computational complexity. We refer the interested reader to one of the many available tutorials on DS propagation for a more thorough treatment (eg. Almond (1988) or Kohlas and Monney (1994)).

4. DS Banff Challenge Solution. The join tree for the (single-channel) Banff Challenge is depicted in Figure 5. The components involving $L_y$ and $L_z$ are just as described in the previous section. The additional components relate the Poisson DSM on $(L_N, N)$ to the other components via the definition $L_N := ES + B$. An additional hint constrains $S$ to the non-negative Reals (since the mass of the Higgs particle must be $\geq 0$).

We will ultimately characterize our uncertainty about the quantity of interest $s$ by the distribution of an a-random variable $S$. If $F_S(s) = \int_0^s f_S(x)dx$ is the cumulative distribution function of $S$, then our goal is to find $s_{90}$ and $s_{99}$ such that $F_S(s_{90}) = .90$ and $F_S(s_{99}) = .99$. The function $f_S(\cdot)$ is the result of a “plausibility transformation” from the $S$ margin of the posterior DSM. In the $n$-channel case we get $f_S(x) \propto \prod_{i=1}^n r_i(x)$, where $r_i(x) = c_i(x, x)$ is the DS commonality of the singleton $\{x\}$ on the $S$ margin of the sub-DSM corresponding to channel $i$.
The DS commonality (for channel $i$) of the singleton \{x\},

$$r_i(x) = \left( F_{S^i}(x) - F_{S_u}(x) \right),$$

is the difference between the CDFs of the a-random variables $S^i$ and $S_u$ for the lower end and upper end of the a-random interval. To complete the DS solution we need the distribution functions $F_{S^i}(\cdot)$ and $F_{S_u}(\cdot)$. Equations for these functions are derived later in this section. A simplified result is provided here.

In the cases in which $n > 0$, $y > 0$, and $z > 0$, and ignoring the constraint that $s \geq 0$, the formulas are

$$F_{S^i}(x) = \mathbb{P}\left( \frac{N^i - \frac{1}{t_i} Y^i}{Z^i_u} \leq x \right)$$

and

$$F_{S_u}(x) = \mathbb{P}\left( \frac{N^i - \frac{1}{z_i} Y^i}{Z^i_u} \leq x \right),$$

where $N^i$, $Z^i$, $Y^i$ are the lower ends of the a-random intervals for $L^i_n := (\epsilon t + b_i)$, $L^i_y := (t_i b_i)$, and $L^i_z := (u_i \epsilon_i)$, and $N^i_u$, $Z^i_u$, $Y^i_u$ are the upper ends. The lower ends of these a-random intervals are distributed according
to unit-scale independent gammas:

\[ \mathbf{N}_i \sim \text{Gamma}(n_i), \]
\[ \mathbf{Y}_i \sim \text{Gamma}(y_i), \]
\[ \mathbf{Z}_i \sim \text{Gamma}(z_i), \]

The upper ends are also gamma distributed, such that the differences \( \mathbf{N}_i - \mathbf{N}_i, \mathbf{Y}_i - \mathbf{Y}_i, \text{ and } \mathbf{Z}_i - \mathbf{Z}_i \) are each independently exponentially distributed \( \sim \text{Exp}(1) \).

4.1. Derivation of the DS Solution. Recall that our goal is to find \( s_{99}^* \) and \( s_{99}^* \) such that \( F_S(s_{99}^*) = .90 \) and \( F_S(s_{99}^*) = .99 \), where \( F_S(s) = \int_0^s f_S(x) \, dx \).

The evidence given by the data corresponding to channel \( i \), \( (n_i, y_i, z_i, t_i, u_i) \), constrains the unknown \( s \) to the a-random interval \( (S_i^l, S_i^u) \). By characterizing the distribution functions of the a-random variables \( S_i^l \) and \( S_i^u \), we can calculate \( r_i(x) = \left( F_{S_i^l}(x) - F_{S_i^u}(x) \right) \), and from this, \( f_S(x) \propto \prod_{i=1}^n r_i(x) \).

We assume, for now, that \( n_i > 0, y_i > 0, \text{ and } z_i > 0 \); the Special Cases section, below, addresses the cases in which one or more of these counts is 0. By the Poisson DSM, we know that the distributions of the lower ends of the a-random intervals for \( L_i := (\epsilon_i s + b_i), L_i^u := (t_i b_i), \text{ and } L_i^u := (u_i \epsilon_i) \) are independent unit-scale gammas:

\[ \mathbf{N}_i \sim \text{Gamma}(n_i), \]
\[ \mathbf{Y}_i \sim \text{Gamma}(y_i), \]
\[ \mathbf{Z}_i \sim \text{Gamma}(z_i), \]

and that the upper ends are also gamma distributed, such that the differences \( \mathbf{N}_i - \mathbf{N}_i, \mathbf{Y}_i - \mathbf{Y}_i, \) and \( \mathbf{Z}_i - \mathbf{Z}_i \) are each independently exponentially distributed \( \sim \text{Exp}(1) \). From this we get the constraints that

\[ \frac{1}{l_i} \mathbf{Y}_i \leq b_i \leq \frac{1}{l_i} \mathbf{Y}_i, \]
\[ \frac{1}{u_i} \mathbf{Z}_i \leq \epsilon_i \leq \frac{1}{u_i} \mathbf{Z}_i. \]

From these and from the additional constraint that \( s \geq 0 \), we see that

\[ S_i^l = \frac{\max(0, \mathbf{N}_i - \frac{1}{u_i} \mathbf{Y}_i)}{\frac{1}{u_i} \mathbf{Z}_i} \]
\[ S_i^u = \frac{\mathbf{N}_i - \frac{1}{u_i} \mathbf{Y}_i}{\frac{1}{u_i} \mathbf{Z}_i} \]
in the equation $S^i_s \leq s \leq S^i_u$.

Thus, if we ignore (momentarily) the constraint that $s \geq 0$, we may characterize the CDFs of $S^i_s$ and $S^i_u$ as

$$F^*_s(s) = \mathbb{P}\left(\frac{N^i_u - \frac{1}{t^i_u}Y^i_u}{\frac{1}{u_i}Z^i_u} \leq s\right) \quad \text{and}$$

$$F^*_u(s) = \mathbb{P}\left(\frac{N^i_u - \frac{1}{t^i_u}Y^i_u}{\frac{1}{u_i}Z^i_u} \leq s\right).$$

Rearranging, we may write this as

$$F^*_s(s) = \mathbb{P}(N^i_u - \frac{1}{t^i_u}Y^i_u \leq \frac{x}{u_i}Z^i_u) \quad \text{and}$$

$$F^*_u(s) = \mathbb{P}(N^i_u - \frac{1}{t^i_u}Y^i_u \leq \frac{x}{u_i}Z^i_u).$$

We are ultimately interested in the normalized quantities

$$F_{S^i_s}(x) = \frac{F^*_s(x) - \mathbb{P}(S^i_u < 0)}{1 - \mathbb{P}(S^i_u < 0)} \quad \text{and}$$

$$F_{S^i_u}(x) = \frac{F^*_u(x) - \mathbb{P}(S^i_u < 0)}{1 - \mathbb{P}(S^i_u < 0)},$$

where we condition on the upper end of the interval, $S^i_u$, being non-negative. Since this condition is met whenever $N^i_u \geq \frac{1}{t^i_u}Y^i_u$, we have

$$F_{S^i_s}(x) = \frac{\mathbb{P}(N^i_u - \frac{1}{t^i_u}Y^i_u + \frac{x}{u_i}Z^i_u) - \mathbb{P}(N^i_u < \frac{1}{t^i_u}Y^i_u)}{1 - \mathbb{P}(N^i_u < \frac{1}{t^i_u}Y^i_u)} \quad \text{and}$$

$$F_{S^i_u}(x) = \frac{\mathbb{P}(N^i_u - \frac{1}{t^i_u}Y^i_u + \frac{x}{u_i}Z^i_u) - \mathbb{P}(N^i_u < \frac{1}{t^i_u}Y^i_u)}{1 - \mathbb{P}(N^i_u < \frac{1}{t^i_u}Y^i_u)}. \quad \text{(4.1)}$$

As we show in the following sections, these may be expressed in terms of
the Beta CDF as

\[
F_{S_i^l}(x) = 1 - \frac{pB(\alpha_i, z_i + 1, n_i) - \int_0^{\alpha_i} pB(\frac{1}{1+t_i(1-\gamma\alpha_i)}, z_i + 1 + n_i, y_i + 1)d\gamma}{pB(\frac{1}{t_i+1}, y_i, n_i + 1)}
\]

and

\[
F_{S_i^u}(x) = 1 - \frac{pB(\alpha_i, z_i, n_i + 1) - \int_0^{\alpha_i} pB(\frac{1}{1+t_i(1-\gamma\alpha_i)}, z_i + n_i, n_i + 1, y_i + 1)d\gamma}{pB(\frac{1}{t_i+1}, y_i, n_i + 1)}
\]

where \( pB(\cdot, \alpha, \beta) \) is the PDF of a Beta distribution with parameters \( \alpha \) and \( \beta \), \( pB(\cdot, \alpha, \beta) \) is its CDF, and \( \alpha_i := \frac{u_i}{u_i+x} \).

4.2. Non-negativity Constraint. The normalization of \( F_{S_i^l}(\cdot) \) and \( F_{S_i^u}(\cdot) \) requires calculating the probability that the constraint \( s \geq 0 \) is violated. The constraint is violated whenever \( N_{i}^u < \frac{1}{t_i} Y_{i}^l \), which occurs with probability

\[
P(N_{i}^u < \frac{1}{t_i} Y_{i}^l).
\]

We can thus express the constraint violation probability as

\[
P(N_{i}^u < \frac{1}{t_i} Y_{i}^l) = P\left(\frac{N_{i}^u}{N_{i}^u + Y_{i}^l} < \frac{1}{t_i}\right) = P\left(\frac{Y_{i}^l}{N_{i}^u + Y_{i}^l} > \frac{t_i}{t_i+1}\right)
\]

Since \( N_{i}^u \) and \( Y_{i}^l \) are gamma-distributed with unit scale, and with shape parameters \( (n_i + 1) \) and \( y_i \), respectively,

\[
\frac{Y_{i}^l}{N_{i}^u + Y_{i}^l} \sim \text{Beta}(y_i, n_i + 1).
\]

So the probability that the non-negativity constraint is violated is

\[
P(N_{i}^u < \frac{1}{t_i} Y_{i}^l) = 1 - pB(\frac{1}{t_i+1}, y_i, n_i + 1).
\]

4.3. Unnormalized CDFs. We now use similar techniques to characterize the unnormalized components \( F_{S_i^u}^*(\cdot) \) and \( F_{S_i^l}^*(\cdot) \). Consider first

\[
F_{S_i^u}^*(x) = P(N_{i}^u \leq \frac{1}{t_i} Y_{i}^l + \frac{x}{u_i} Z_{i}^l).
\]
Noting that gamma random variables are never negative, we may apply the law of total probability to rewrite this as

\[
F^*_{\mathbf{S}_u}(x) = \mathbb{P}(\mathbf{N}_u^i \leq \frac{1}{t_i} \mathbf{Y}_i^i + \frac{x}{u_i} \mathbf{Z}_i^i \text{ and } \mathbf{N}_u^i \leq \frac{x}{u_i} \mathbf{Z}_i^i) \\
+ \mathbb{P}(\mathbf{N}_u^i \leq \frac{1}{t_i} \mathbf{Y}_i^i + \frac{x}{u_i} \mathbf{Z}_i^i \text{ and } \mathbf{N}_u^i > \frac{x}{u_i} \mathbf{Z}_i^i),
\]

the first term of which can be expressed as

\[
\mathbb{P}(\mathbf{N}_u^i \leq \frac{1}{t_i} \mathbf{Y}_i^i + \frac{x}{u_i} \mathbf{Z}_i^i \text{ and } \mathbf{N}_u^i \leq \frac{x}{u_i} \mathbf{Z}_i^i) = \mathbb{P}(\mathbf{N}_u^i \leq \frac{x}{u_i} \mathbf{Z}_i^i) \\
= \mathbb{P}(\frac{\mathbf{Z}_i^i}{\mathbf{Z}_i^i + \mathbf{N}_u^i} > \frac{u_i}{u_i + x}) \\
= 1 - \text{Beta}(\frac{u_i}{u_i + x}, z_i, n_i + 1),
\]

by an argument similar to that used in deriving the Beta CDF representation for the probability of violating the non-negativity constraint.

The latter part may be simplified also. We can rewrite \(\mathbf{N}_u^i - \frac{x}{u_i} \mathbf{Z}_i^i\) as

\[
(1 - \frac{\mathbf{Z}_i^i + \mathbf{N}_u^i}{u_i + x})(\mathbf{Z}_i^i + \mathbf{N}_u^i),
\]

since

\[
(1 - \frac{\mathbf{Z}_i^i + \mathbf{N}_u^i}{u_i + x})(\mathbf{Z}_i^i + \mathbf{N}_u^i) = (\mathbf{Z}_i^i + \mathbf{N}_u^i) - \frac{\mathbf{Z}_i^i}{u_i + x} \\
= (\mathbf{Z}_i^i + \mathbf{N}_u^i) - \frac{\mathbf{Z}_i^i}{1 + \frac{x}{u_i}} \\
= (\mathbf{Z}_i^i + \mathbf{N}_u^i) - \mathbf{Z}_i^i(1 + \frac{x}{u_i}) \\
= \mathbf{N}_u^i - \frac{x}{u_i} \mathbf{Z}_i^i.
\]

This leads to

\[
\mathbb{P}(\mathbf{N}_u^i \leq \frac{1}{t_i} \mathbf{Y}_i^i + \frac{x}{u_i} \mathbf{Z}_i^i \text{ and } \mathbf{N}_u^i > \frac{x}{u_i} \mathbf{Z}_i^i) \\
= \mathbb{P}(\frac{1}{t_i} \mathbf{Y}_i^i \geq (1 - \frac{\mathbf{Z}_i^i + \mathbf{N}_u^i}{u_i + x})(\mathbf{Z}_i^i + \mathbf{N}_u^i) \text{ and } \mathbf{N}_u^i > \frac{x}{u_i} \mathbf{Z}_i^i).
\]

(4.4)

Recognizing again that the event that \(\mathbf{N}_u^i > \frac{x}{u_i} \mathbf{Z}_i^i\) is the same as the event that \(\frac{\mathbf{Z}_i^i}{\mathbf{Z}_i^i + \mathbf{N}_u^i} \leq \frac{u_i}{u_i + x}\), the complicated probability in (4.4) may be simplified
by conditioning on the value of $\frac{Z^i}{Z^i+N^u} = \gamma$:

\[
\mathbb{P}(N^i_u \leq \frac{1}{t_i} Y^i_l + \frac{x}{u_i} Z^i_l \text{ and } N^i_u > \frac{x}{u_i} Z^i_l)
\]

\[
= \mathbb{P}
\left( \frac{1}{t_i} Y^i_l \geq (1 - \frac{Z^i_l}{Z^i_l+N^u})(Z^i_l+N^u) \text{ and } N^i_u > \frac{x}{u_i} Z^i_l \right)
\]

\[
= \mathbb{P}
\left( \frac{1}{t_i} Y^i_l \geq (1 - \frac{Z^i_l}{Z^i_l+N^u})(Z^i_l+N^u) \text{ and } \frac{Z^i_l}{Z^i_l+N^u} \leq \frac{u_i}{u_i+x} \right)
\]

\[
= \int_0^{\alpha_i} \mathbb{P}
\left( \frac{1}{t_i} Y^i_l \geq (1 - \frac{\gamma}{\alpha_i})(Z^i_l+N^u) \right) d\mathbb{P}(\frac{Z^i_l}{Z^i_l+N^u} < \gamma)
\]

\[
= \int_0^{\alpha_i} \mathbb{P}
\left( \frac{Z^i_l+N^u}{Z^i_l+N^u+Y^i_l} \leq \frac{1}{1+t_i(1-\frac{\gamma}{\alpha_i})} \right) d\mathbb{P}(\frac{Z^i_l}{Z^i_l+N^u} < \gamma),
\]

where $\alpha_i := \frac{u_i}{u_i+x}$. Since $(Z^i_l+N^u) \perp \frac{Z^i_l}{Z^i_l+N^u}$, we get

\[
\mathbb{P}(N^i_u \leq \frac{1}{t_i} Y^i_l + \frac{x}{u_i} Z^i_l \text{ and } N^i_u > \frac{x}{u_i} Z^i_l)
\]

\[
= \int_0^{\alpha_i} p\mathcal{B}(\frac{1}{1+t_i(1-\frac{\gamma}{\alpha_i})}, z_i + n_i + 1, y_i) d\mathcal{B}(\gamma, z_i, n_i + 1) d\gamma.
\]

Note that this can be approximated to any desired precision by a straightforward numerical integration.

Thus we have

\[
(4.5)
\]

\[
F^*_S(x) = 1 - p\mathcal{B}(\frac{u_i}{u_i+x}, z_i, n_i + 1)
\]

\[
+ \int_0^{\alpha_i} p\mathcal{B}(\frac{1}{1+t_i(1-\frac{\gamma}{\alpha_i})}, z_i + n_i + 1, y_i) d\mathcal{B}(\gamma, z_i, n_i + 1) d\gamma,
\]

and, by an analogous derivation,

\[
(4.6)
\]

\[
F^*_S(x) = 1 - p\mathcal{B}(\frac{u_i}{u_i+x}, z_i + 1, n_i)
\]

\[
+ \int_0^{\alpha_i} p\mathcal{B}(\frac{1}{1+t_i(1-\frac{\gamma}{\alpha_i})}, z_i + 1 + n_i, y_i + 1) d\mathcal{B}(\gamma, z_i + 1, n_i) d\gamma.
\]
4.4. Normalized CDFs. Plugging equations (4.5), (4.6), and (4.3) into equation (4.1) concludes the derivation of equation (4.2), the CDFs of the upper and lower bounds on $s$ induced by the evidence from a single channel. Combining evidence across channels, normalizing, and applying the plausibility transform yields

$$f_S(x) = \frac{\prod_{i=1}^{n} \left( F_{S_i^l}(x) - F_{S_i^u}(x) \right)}{\int_{0}^{\infty} \prod_{i=1}^{n} \left( F_{S_i^l}(x) - F_{S_i^u}(x) \right)}.$$  

4.5. Special Cases. The above derivation for $r_i(\cdot)$ assumed that $n_i$, $y_i$, and $z_i$ are all positive. In the event that $n_i = 0$, $S_i^l = 0$ and $F_{S_i^l}(x) = 1 \forall x \geq 0$. When $y_i = 0$, there is no violation of the non-negativity constraint, and $F_{S_i^u}^*(x) = 1 - p\beta\left(\frac{u}{u_i+y_i+z_i}, n_i+1\right)$, unless $z_i = 0$. Whenever $z_i = 0$, $S_i^l = \infty$, so $F_{S_i^l}(x) = 0 \forall x < \infty$.

Note that when $n_i = z_i = 0$, we learn nothing new about $s$.

5. Results. As discussed in the Introduction, we applied the DS method to the three datasets corresponding to tasks 1a, 1b, and 2. We implemented the algorithm in the Perl programming language, using a crude 100-point rectangle integration procedure to numerically compute the integrals in equations (4.5) and (4.6) for each of 100 values of $x$. The 100 values of $x$ were chosen separately for each data channel, with the maximum value chosen to include the non-negligible range of both $F_{S_i^l}(x)$ and $F_{S_i^u}(x)$ for that channel’s values of $(n, y, z, t, u)$. After combining the values of $\left( F_{S_i^l}(x) - F_{S_i^u}(x) \right)$ from each channel $i$, we reported the 90th and 99th percentiles of $f_S(x)$ as the 90% and 99% one-sided upper bounds on $s$.

As discussed earlier, the evaluator computed the coverage value $C(s)$ for each of 101 values of $s$ in the range from 0 to 25. For tasks 1a and 2, he also computed a standard error. The coverage results are plotted below, with a red line indicating the desired coverage. Note that the 90% plots are on a different scale than the 99% plots.

The coverage results for task 1a are shown in Figure 6. The results for task 1b are shown in Figure 7. The results for task 2 are shown in Figure 8.

The coverage results for tasks 1a and 2 were very close to the desired coverage after an initial period of overcoverage. The initial period of over-coverage, also found in Heinrich et al (2004), is inevitable at $s = 0$ and understandable at small values of $s$, due to the truncation at 0. The results for task 1b show actual coverage closer to 95% and 99.5%, slightly above the desired values of 90% and 99%. This is presumably due to the smaller
Figure 6. Task 1a Coverage Plots (estimated coverage and 95% CI)
Fig 7. Task 1b Coverage Plots

ESTIMATING THE MASS OF THE HIGGS USING DSA

imsart-aos ver. 2007/12/10 file: DBanff.tex date: December 31, 2007
Task 2
Dempster–Shafer Method
90% Coverage Plot

(a) 90% Coverage

Task 2
Dempster–Shafer Method
99% Coverage Plot

(b) 99% Coverage

Fig 8. Task 2 Coverage Plots (estimated coverage and 95% CI)
values of \( t \) and \( u \) used in this task. In the project specification, task 1b is referred to as the “large uncertainty” task because the nuisance parameters are more difficult to estimate.

Overall, the results are impressively accurate. For comparison we include a figure from Heinrich et al (2004) showing the result of applying a pure Bayesian approach to a similar problem (but in which the background \( b \) is fixed and known rather than a nuisance parameter). Note that in this figure (Figure 9), \( \kappa \) is what we refer to as \( u \). At \( \kappa = 100 \) it corresponds to the value used in task 1a.

6. Discussion. The DS approach described in this paper is conceptually straightforward, simple to implement, efficient to compute, and performs very well at the given tasks. Unlike pure Bayesian approaches there is no need to specify a prior. If prior information is available, however, this approach can easily accommodate that information, unlike frequentist approaches.

It remains unclear whether the tasks of this challenge are representative of the actual scientific goal of placing confidence intervals on the mass of the Higgs particle. The three-Poisson model for the background, efficiency, and combined (signal times efficiency plus background) counts may or may not be the best representation of the problem. These questions are perhaps best left to the physics and astronomy communities to debate.

What we have shown is that the DS approach, which heretofore has not
been considered for the particle mass estimation problem, is an approach at least on par with the more commonly considered techniques. DSA is a statistical framework that is not well understood by most statisticians, though it has steadily gained practitioners since its quiet inception in the middle of the last century. With the Bayesian-frequentist debate hinging primarily on the power and danger of incorporating prior information into an analysis, the Dempster-Shafer approach deserves consideration.

Acknowledgements. The authors would like to acknowledge Mathis Thoma, Joel Heinrich, and the members of the California/Harvard Astrostatistics Collaboration (CHASC) for their inspiration and helpful feedback during the course of this project.

References.


ESTIMATING THE MASS OF THE HIGGS USING DSA


Harvard University Statistics Department
Science Center
1 Oxford Street
Cambridge, MA 02138-2901 USA
E-mail: edlefsen@fas.harvard.edu
dempster@fas.harvard.edu

Department of Statistics
Purdue University
250 N. University Street
West Lafayette, IN 47907-2066 USA
E-mail: chuanhai@stat.purdue.edu