CONFLATIONS OF PROBABILITY DISTRIBUTIONS

THEODORE P. HILL

ABSTRACT. The *conflation* of a finite number of probability distributions $P_1, \ldots, P_n$ is a consolidation of those distributions into a single probability distribution $Q = Q(P_1, \ldots, P_n)$, where intuitively Q is the conditional distribution of independent random variables $X_1, \ldots, X_n$ with distributions $P_1, \ldots, P_n$, respectively, given that $X_1 = \cdots = X_n$. Thus, in large classes of distributions the conflation is the distribution determined by the normalized product of the probability density or probability mass functions. Q is shown to be the unique probability distribution that minimizes the loss of Shannon information in consolidating the combined information from $P_1, \ldots, P_n$ into a single distribution Q, and also to be the optimal consolidation of the distributions with respect to two minimax likelihood-ratio criteria. In that sense, conflation may be viewed as an optimal method for combining the results from several different independent experiments. When $P_1, \ldots, P_n$ are Gaussian, Q is Gaussian with mean the classical weighted-mean-squares reciprocal of variances. A version of the classical convolution theorem holds for conflations of a large class of a.c. measures.

1. INTRODUCTION

Conflation is a method for consolidating a finite number of probability distributions $P_1, \ldots, P_n$ into a single probability distribution $Q = Q(P_1, \ldots, P_n)$. The study of this method was motivated by a basic problem in science, namely, how best to consolidate the information from several independent experiments, all designed to measure the same unknown quantity. The experiments may differ in time, geographical location, methodology and even in underlying theory. Ideally, of course, all experimental data, past as well as present, should be incorporated into the scientific record, but the result would be of limited practical application. For many purposes, a concise consolidation of those distributions is more useful.

For example, to obtain the current internationally-recognized values of each of the fundamental physical constants (Planck's constant, Avogadro's number, etc.),

the U.S. National Institute of Standards and Technology (NIST) collects independent distributional data, often assumed to be Gaussian (see Section 6), from various laboratories. Then, for each fundamental physical constant, NIST combines the relevant input distributions to arrive at a recommended value and estimated standard deviation for the constant. Since these recommended values are usually interpreted as being Gaussian, NIST has effectively combined the several input distributions into a single probability distribution.

The problem of combining probability distributions has been well studied; e.g., [7] describes a "plethora of methods" for finding a summary $T(P_1, \ldots, P_n)$ of n given (subjective) probability measures $P_1, \ldots, P_n$ that represent different expert opinions. Essentially all those methods, however, including the classical convex combination or weighted average ($T(P_1, \ldots, P_n) = \sum_{i=1}^{n} w_i P_i$, with nonnegative weights $\{w_i\}$ satisfying $\sum_{i=1}^{n} w_i = 1$) and its various nonlinear generalizations, are *idempotent*, i.e., $T(P, \ldots, P) = P$. For the purpose of combining probability distributions that represent expert opinions, idempotency is a natural requirement, since if all the opinions $P_1, \ldots, P_n$ agree, the best summary is that distribution.

But for other objectives for combining distributions, such as consolidating the results of independent experiments, idempotency is not always a desirable property. Replications of the same underlying distribution by independent laboratories, for example, should perhaps best be summarized by a distribution with a smaller variance. In addition to the problem of assigning and justifying the unequal weights, another problem with the weighted averages consolidation is that even with normally-distributed input data, this method generally produces a multimodal distribution, whereas one might desire the consolidated output distribution to be of the same general form as that of the input data – normal, or at least unimodal.

Another natural method of consolidating distributional data – one that does preserve normality, and is not idempotent – is to average the underlying input data. In this case, the consolidation $T(P_1, \ldots, P_n)$ is the distribution of $(X_1 + \cdots + X_n)/n$ (or a weighted average), where $\{X_i\}$ are independent with distributions $\{P_i\}$, respectively. With this consolidation method, the variance of $T(P_1, \ldots, P_n)$ is strictly smaller (unless $\{X_i\}$ are all constant) than the maximum variance of the $\{P_i\}$, since $\text{var}(P) = (\text{var}(P_1) + \cdots + \text{var}(P_n))/n^2$. Input data distributions that differ significantly, however, may sometimes reflect a higher uncertainty or variance. More fundamentally, in general this method requires averaging of completely dissimilar data, such as results from completely different experimental methods (see Section 6).

The method for consolidating distributional data presented below, called the *conflation* of distributions, and designated with the symbol "&" to suggest consolidation of P_1 *and* P_2, does not require ad hoc weights, and the mean and/or variance of the conflation may be larger or smaller than the means or variances of the input distributions. In general, conflation automatically gives more weight to input distributions arising from more accurate experiments, i.e. distributions with smaller standard deviations. The conflation of several distributions has several other properties that may be desirable for certain consolidation objectives – conflation minimizes the loss of Shannon information in consolidating the combined information from $P_1, \ldots, P_n$ into a single distribution Q, and is both the unique minimax likelihood ratio consolidation and the unique proportional likelihood ratio consolidation of the given input distributions.

In addition, conflations of normal distributions are always normal, and coincide with the classical weighted least squares method, hence yielding best linear unbiased and maximum likelihood estimators. Many of the other important classical families of distributions, including gamma, beta, uniform, exponential, Pareto, LaPlace, Bernoulli, Zeta and geometric families, are also preserved under conflation. The conflation of distributions has a natural heuristic and practical interpretation – gather data (e.g., from independent laboratories) sequentially and simultaneously, and record the values only when the results (nearly) agree.

In some contexts the basic assumptions, both of the independence of the underlying probability distributions (experiments) and of the equal validity of the distributions (uniform prior), may be quite natural. In other contexts, however, where different experiments may partially share apparatus or methodology, or where there is reason to believe that some experiments may be intrinsically more likely to be valid, a generalization of the notion of conflation introduced below will perhaps be more suitable. Such generalizations, both to distributions that are dependent and to situations with nonuniform priors, are not developed here, but are left for future research.

2. Basic definition and properties of conflations

Throughout this article, $\mathbb{N}$ will denote the natural numbers, $\mathbb{Z}$ the integers, $\mathbb{R}$ the real numbers, $(a, b]$ the half-open interval $\{x \in \mathbb{R} : a < x \leq b\}$, $\mathbb{B}$ the Borel subsets of $\mathbb{R}$, $\mathcal{P}$ the set of all real Borel probability measures, δ_x the Dirac delta measure in $\mathcal{P}$ at the point x (i.e., $\delta_x(B) = 1$ if $x \in B$, and $= 0$ if $x \notin B$), $\|\mu\|$ the total mass of the Borel sub-probability μ, $o(\ )$ the standard "little oh" notation $o(a_n) = b_n$ if and only if $\lim_{n \to \infty} \frac{a_n}{b_n} = 0$, a.c. means absolutely continuous, the p.m.f. of P is the probability mass function $(p(k) = P(\{k\}))$ if P is discrete and p.d.f. is the probability density function (Radon-Nikodyn derivative) of P if P is a.c., $E(X)$ denotes the expected value of the random variable X, ψ_P the characteristic function of $P \in \mathcal{P}$ (i.e., $\psi_P(t) = \int_{-\infty}^{\infty} e^{itx} dP(x)$), I_A is the indicator function of the set A (i.e. $I_A(x) = 1$ if $x \in A$ and $= 0$ if $x \notin A$), $g \otimes h$ is the convolution $(g \otimes h)(t) = \int_{-\infty}^{\infty} g(t - s)h(s)ds$ of g and h, and A^c is the complement $\mathbb{R} \backslash A$ of the set A. For brevity, $\mu((a, b])$ will be written $\mu(a, b]$, $\mu(\{x\})$ as $\mu(x)$, etc.

Definition 2.1. For $P_1, \ldots, P_n \in \mathcal{P}$ and $j \in \mathbb{N}$, $\mu_j(P_1, \ldots, P_n)$ is the purely-atomic j-dyadic sub-probability measure

$$\mu_j(P_1, \ldots, P_n) = \sum_{k \in \mathbb{Z}} \prod_{i=1}^{n} P_i((k-1)2^{-j}, k2^{-j}] \delta_{k2^{-j}}.$$

Remark. The choice of using half-open dyadic intervals closed on the right and of placing the mass in every dyadic interval at the right end point is not at all important — the results which follow also hold if other conventions are used, such as decimal or ternary half-open intervals closed on the left, with masses placed at the center.

Example 2.2. If P_1 is a Bernoulli distribution with parameter $p = \frac{1}{3}$ (i.e. $P = \frac{(2\delta_0 + \delta_1)}{3}$) and P_2 is Bernoulli with parameter $\frac{1}{4}$, then $\mu_j(P_1, P_2) = \frac{(6\delta_{1/2} + \delta_1)}{12}$ for all $j \in \mathbb{N}$.

The next proposition is the basis for the definition of conflation of general distributions below. Recall (e.g. [4, Theorem 4.4.1]) that for real Borel sub-probability measures $\{\nu_j\}$ and ν, the following are equivalent:

(2.1a) $\qquad\qquad\qquad\qquad \nu_j \to \nu$ vaguely as $j \to \infty$;

(2.1b) $\qquad\qquad\qquad\qquad \nu_j(a,b] \to \nu(a,b]$ for all $a < b$ in a dense set $D \subset \mathbb{R}$;

(2.1c) $\qquad \lim_{j\to\infty} \int f(x)d\nu_j(x) = \int f(x)d\nu(x)$

$\qquad\qquad\qquad\qquad\qquad$ for all continuous f that vanish at infinity.

Theorem 2.3. *For all $P, P_1, \ldots, P_n \in \mathcal{P}$,*

(i) *$\mu_{j+1}\left(\frac{a}{2^m}, \frac{b}{2^m}\right] \leq \mu_j\left(\frac{a}{2^m}, \frac{b}{2^m}\right]$ for all $j, m \in \mathbb{N}$, $j > m$; and all $a \leq b$, $a, b \in \mathbb{Z}$;*

(ii) *$\mu_j(P_1, \ldots, P_n)$ converges vaguely to a sub-probability measure $\mu_\infty(P_1, \ldots, P_n)$;*

(iii) *$\lim_{j\to\infty} \|\mu_j(P_1, \ldots, P_n)\| = \|\mu_\infty(P_1, \ldots, P_n)\|$; and*

(iv) *$\mu_\infty(P) = P$, and $\mu_j(P)$ converges vaguely to P as $j \to \infty$.*

The following simple observation — that the square of the sums of nonnegative numbers is always at least as large as the sum of the squares — will be used in the proof of the theorem and several times in the sequel, and is recorded here for ease of reference.

Lemma 2.4. *For all $n \in \mathbb{N}$, all $a_{i,k} \geq 0$, and all $J \subset \mathbb{N}$, $\prod_{i=1}^n \sum_{k \in J} a_{i,k} \geq \sum_{k \in J} \prod_{i=1}^n a_{i,k}$.*

Proof of Theorem 2.3. For (i), note that for $j > m$

$$(2.2a) \qquad \mu_j\left(\frac{a}{2^m}, \frac{b}{2^m}\right] = \sum_{k=a2^{j-m-1}}^{b2^{j-m-1}-1} \mu_j\left(\frac{k}{2^{j-1}}, \frac{k+1}{2^{j-1}}\right]$$

$$= \sum_{k=a2^{j-m}}^{b2^{j-m}-1} \mu_j\left(\frac{k}{2^j}, \frac{k+1}{2^j}\right]$$

and

$$(2.2b) \qquad \mu_{j+1}\left(\frac{a}{2^m}, \frac{b}{2^m}\right] = \sum_{k=a2^{j-m}}^{b2^{j-m}-1} \mu_{j+1}\left(\frac{k}{2^j}, \frac{k+1}{2^j}\right].$$

By the definition of μ_j,

$$(2.3a) \qquad \mu_j\left(\frac{k}{2^j}, \frac{k+1}{2^j}\right] = \prod_{i=1}^n P_i\left(\frac{k}{2^j}, \frac{k+1}{2^j}\right]$$

$$= \prod_{i=1}^n \left(P_i\left(\frac{2k}{2^{j+1}}, \frac{2k+1}{2^{j+1}}\right] + P_i\left(\frac{2k+1}{2^{j+1}}, \frac{2k+2}{2^{j+1}}\right]\right)$$

and

$$(2.3b) \qquad \mu_{j+1}\left(\frac{k}{2^j}, \frac{k+1}{2^j}\right] = \prod_{i=1}^n P_i\left(\frac{2k}{2^{j+1}}, \frac{2k+1}{2^{j+1}}\right]$$

$$+ \prod_{i=1}^n P_i\left(\frac{2k+1}{2^{j+1}}, \frac{2k+2}{2^{j+1}}\right].$$

By Lemma 2.4, (2.3a) and (2.3b) imply that

$$(2.4) \qquad \mu_{j+1}\left(\frac{k}{2^j}, \frac{k+1}{2^j}\right] \le \mu_j\left(\frac{k}{2^j}, \frac{k+1}{2^j}\right] \quad \text{for all } j > m,\ j, m \in \mathbb{N},\ k \in \mathbb{Z}.$$

By (2.2a) and (2.2b), this implies (i).

For (ii), note that since every sequence of sub-probability measures contains a subsequence that converges vaguely to a sub-probability measure (e.g. [4, Theorem 4.3.3]), there exists a subsequence $\{\mu_{j_k}(P_1, \ldots, P_n)\}$ of $\{\mu_j(P_1, \ldots, P_j)\}$ and a sub-probability measure $\mu_\infty(P_1, \ldots, P_n)$ so that $\mu_{j_k}(P_1, \ldots, P_n)$ converges vaguely to $\mu_\infty(P_1, \ldots, P_n)$ as $k \to \infty$. Hence by the uniqueness of vague limits (i.e. convergence on intervals from different dense sets results in the same limit measure [4, corollary to Theorem 4.3.1, p. 86]), (i) implies that

$$\lim_{j \to \infty} \mu_j\left(\frac{a}{2^m}, \frac{b}{2^m}\right] = \mu_\infty\left(\frac{a}{2^m}, \frac{b}{2^m}\right],$$

which proves that $\mu_j(P_1, \ldots, P_n)$ converges vaguely to $\mu_\infty(P_1, \ldots, P_n)$.

For (iii), note that

$$\lim_{j \to \infty} \|\mu_j\| = \lim_{j \to \infty} \sum_{k=-\infty}^{\infty} \mu_j(k, k+1]$$

$$= \sum_{k=-\infty}^{\infty} \lim_{j \to \infty} \mu_j(k, k+1] = \sum_{k=-\infty}^{\infty} \mu_\infty(k, k+1] = \|\mu_\infty\|,$$

where the second equality follows by the dominated convergence theorem, and the third by the definition of μ_∞. The special case $n = 1$ of (iv) is immediate. $\qquad \square$

Definition 2.5. $P_1, \ldots, P_n \in \mathcal{P}$ are (*mutually*) *compatible* if $\|\mu_j\| > 0$ for all $j \in \mathbb{N}$.

Clearly every normal distribution is compatible with every probability distribution, every exponential distribution is compatible with every distribution with support in the positive reals, and every geometric distribution is compatible with every discrete distribution having any atoms in $\mathbb{N}$. Even though Theorem 2.3 guarantees that $\mu_j(P_1, \ldots, P_n)$ converges vaguely to a sub-probability measure $\mu_\infty(P_1, \ldots, P_n)$ and that $\lim_{j \to \infty} \|\mu_j(P_1, \ldots, P_n)\| = \|\mu_\infty(P_1, \ldots, P_n)\|$, and compatibility implies that $\frac{\mu_j(P_1, \ldots, P_n)}{\|\mu_j(P_1, \ldots, P_n)\|}$ is a probability measure for all $j \in \mathbb{N}$, $\lim_{j \to \infty} \frac{\mu_j(P_1, \ldots, P_n)}{\|\mu_j(P_1, \ldots, P_n)\|}$ may not be a probability measure, as the next example shows.

Example 2.6. Let $P_1 = \sum_{k \in N} 2^{-k} \delta_k$ and $P_2 = \sum_{k \in \mathbb{N}} 2^{-k} \delta_{k+2^{-k}}$. Then P_1 and P_2 are easily seen to be compatible, but $\lim_{j \to \infty} \frac{\mu_j(P_1, \ldots, P_n)}{\|\mu_j(P_1, \ldots, P_n)\|}$ is the zero measure, since for each $j \in \mathbb{N}$, the support of the probability measure $\frac{\mu_j(P_1, \ldots, P_n)}{\|\mu_j(P_1, \ldots, P_n)\|}$ is contained in $[j, \infty)$.

The next definition is the main definition in this paper.

Definition 2.7. If $\frac{\mu_j(P_1, \ldots, P_n)}{\|\mu_j(P_1, \ldots, P_n)\|}$ converges vaguely to a Borel probability measure Q as $j \to \infty$, this limit Q is called the *conflation* of $P_1, \ldots, P_n$, written $\&(P_1, \ldots, P_n)$.

Theorem 2.8. *The operation* $\&$ *is commutative and associative, that is,* $\&(P_1, P_2) = \&(P_2, P_1)$ *and* $\&(P_1, \&(P_2, P_3)) = \&(\&(P_1, P_2), P_3) = \&(P_1, P_2, P_3)$.

Proof. Immediate from the definition of μ_∞ since multiplication of real numbers is commutative and associative. $\qquad \square$

Example 2.9. Let P_1 be a Bernoulli distribution with parameter $p = \frac{1}{3}$ and P_2 be Bernoulli with parameter $\frac{1}{4}$, as in Example 2.2. Then $\&(P_1, P_2) = \frac{(6\delta_0 + \delta_1)}{7}$.

Example 2.10. Let P_1 be $N(0,1)$ and P_2 be Bernoulli with parameter $p = \frac{1}{3}$. Then it can easily be seen that $\frac{\mu_j(P_1, P_2)}{\|\mu_j(P_1, P_2)\|}$ converges vaguely to $\&(P_1, P_2) = \delta_0 \left(\frac{2}{(2+e^{-1/2})} \right) + \delta_1 \left(\frac{e^{-1/2}}{(2+e^{-1/2})} \right)$, that is, to the probability measure having the same atoms as the discrete measure, weighted according to the product of the atom masses of P_2 and the magnitude of the density of P_1 at 0 and 1.

3. Conflations of discrete and of absolutely continuous distributions

In general, explicit representations of conflations are not known in closed form. For large natural classes of distributions, however, such as collections of discrete distributions with common atoms and collections of a.c. distributions with overlapping densities, explicit forms of the conflations are easy to obtain. The next two theorems give simple and powerful characterizations of conflations in those two cases. Since in practice input data can easily be approximated extremely closely by discrete distributions with common atoms (e.g., by replacing each P_i by the dyadic approximation $\mu_j(P_i)$ above) or can be smoothed (e.g., by convolution with a $U(-\epsilon, \epsilon)$ or an $N(0, \epsilon^2)$ variable), these two cases are of practical interest. The third conclusion in the next two theorems also yields the heuristic and useful interpretation of conflation described in the introduction.

Theorem 3.1. *Let $P_1, \ldots, P_n$ be discrete with p.m.f.'s $p_1, \ldots, p_n$, respectively, and common atoms A, where $\emptyset \neq A \subset \mathbb{R}$. Then $\&(P_1, \ldots, P_n)$ exists, and the following are equivalent:*

  (i) $Q = \&(P_1, \ldots, P_n)$,
  (ii) $Q = \frac{\sum_{x \in A} \delta_x \prod_{i=1}^{n} p_i(x)}{\sum_{y \in A} \prod_{i=1}^{n} p_i(y)}$,
  (iii) *Q is the conditional distribution of X_1 given that $X_1 = X_2 = \cdots = X_n$, where $X_1, \ldots, X_n$ are independent r.v.'s with distributions $P_1, \ldots, P_n$, respectively.*

Proof. Fix $P_1, \ldots, P_n$ and note that by definition of atom, $p_i(x) > 0$ for all $i = 1, \ldots, n$ and all $x \in A$. Fix $k_0 \in \mathbb{Z}$ and $j_0 \in \mathbb{N}$, and let $D = \left(\frac{k_0}{2^{j_0}}, \frac{k_0+1}{2^{j_0}} \right]$. First it will be shown that

$$(3.1) \qquad \mu_\infty(D) = \sum_{x \in A \cap D} \prod_{i=1}^{n} p_i(x).$$

For all $x \in \mathbb{R}$, $j \in \mathbb{N}$, let $D_{x,j}$ denote the unique dyadic interval $\left(\frac{k}{2^j}, \frac{k+1}{2^j} \right]$ containing x. Note that $D_{x,j} \searrow \{x\}$ as $j \to \infty$, so $P_i(D_{x,j}) \searrow p_i(x)$ as $j \to \infty$ for all i and all $x \in \mathbb{R}$.

This implies

$$(3.2) \qquad \lim_{j \to \infty} \prod_{i=1}^{n} P_i(D_{x,j}) = \prod_{i=1}^{n} p_i(x) \quad \text{for all } x \in \mathbb{R}.$$

Fix $\epsilon > 0$. Since $\{P_i\}$ are discrete, there exists a finite set $A_0 \subset \mathbb{R}$ such that

$$(3.3) \qquad P_i(D \cap A_0^c) < \epsilon \quad \text{for all } i \in \{1, \ldots, n\}.$$

Since $\prod_{i=1}^{n} p_i(x) = 0$ for all $x \in A^c$, (3.3) implies

$$(3.4) \qquad \left| \sum_{x \in A \cap D} \prod_{i=1}^{n} p_i(x) - \sum_{x \in A_0 \cap D} \prod_{i=1}^{n} p_i(x) \right| = \sum_{x \in A \cap A_0^c \cap D} \prod_{i=1}^{n} p_i(x)$$

$$\leq \sum_{x \in A \cap A_0^c \cap D} p_1(x) \leq P_1(D \cap A_0^c) < \epsilon.$$

For each $j \in \mathbb{N}$, let $S_j = \bigcup_{x \in A_0} D_{x,j}$. Then since $x \in D_{x,j}$ for all x and j, (3.3) implies $P_i(D \cap S_j^c) < \epsilon$ for all $i \in \{1, \ldots, n\}$. Thus by definition of $\{\mu_j\}$ and Lemma 2.4,

$$(3.5) \qquad \mu_j(D \cap S_j^c) \leq \prod_{i=1}^{n} P_i(D \cap S_j^c) < \epsilon^n \qquad \text{for all } j \in \mathbb{N}.$$

This implies that

$$(3.6) \qquad \mu_j(D) = \mu_j(D \cap S_j) + \mu_j(D \cap S_j^c)$$

$$= \sum_{x \in D \cap A_0} \mu_j(D_{x,j}) + \mu_j(D \cap S_j^c)$$

$$= \sum_{x \in D \cap A_0} \prod_{i=1}^{n} P_i(D_{x,j}) + \mu_j(D \cap S_j^c),$$

where the second equality follows from the definitions of S_j and $D_{x,j}$. Since $x \in D_{x,j}$, (3.6) implies

$$(3.7) \qquad \mu_j(D) \geq \sum_{x \in D \cap A_0} \prod_{i=1}^{n} P_i(D_{x,j}) \geq \sum_{x \in D \cap A_0} \prod_{i=1}^{n} p_i(x).$$

By (3.6), (3.2) and (3.5),

$$(3.8) \qquad \mu_j(D) \leq \sum_{x \in A_0 \cap D} \prod_{i=1}^{n} p_i(x) + \epsilon^n + \epsilon \quad \text{for sufficiently large } j.$$

By (3.7) and (3.8), $|\mu_j(D) = \sum_{x \in A_0 \cap D} \prod_{i=1}^{n} p_i(x)| \leq \epsilon + \epsilon^n$, so by (3.4),

$$\left| \mu_j(D) - \sum_{x \in A \cap D} \prod_{i=1}^{n} p_i(x) \right| < 2\epsilon + \epsilon^n.$$

Since $\epsilon > 0$ was arbitrary and since $\mu_j \to \mu_\infty$, this implies (3.1). Since D was arbitrary, (3.1) implies that $\|\mu_\infty(P_1, \ldots, P_n)\| = \sum_{x \in A} \prod_{i=1}^{n} p_i(x)$, which proves that $\&(P_1, \ldots, P_n)$ exists. The equivalence of (i) and (ii) follows since $\&(P_1, \ldots, P_n) = \frac{\mu_\infty}{\|\mu_\infty\|}$ and since the measures of dyadic intervals D determine μ_∞. The equivalence of (ii) and (iii) follows immediately from the definition of conditional probability. $\qquad \square$

Example 3.2. If P_1 is binomial with parameters $n = 2$ and $p = \frac{1}{3}$ and P_2 is Poisson with parameter $\lambda = 5$, then $\&(P_1, P_2)$ is discrete with atoms only at 0, 1 and 2 — specifically, $\&(P_1, P_2) = \frac{8\delta_0}{73} + \frac{40\delta_1}{73} + \frac{25\delta_2}{73}$.

Remark. It should be noted that if the input distributions are discrete and have no common atoms, then the conflation does not exist. This could happen if, for example, the underlying experiments were designed to estimate Avogadro's number

(theoretically a 24-digit integer), and the results were given as exact integers. In practice, however, Avogadro's number is known only to seven decimal places, and if the results of the experiments were reported or recorded to eight or nine decimal places of accuracy, then there would almost certainly be common values, and the conflation would be well defined and meaningful. (Restriction to the desired decimal accuracy could be done by the experimenter, or afterwards, e.g., converting each input P_i to $\mu_{20}(P_i)$ as mentioned above.)

The analog of Theorem 3.1 for probability distributions with densities requires an additional hypothesis on the density functions, for the simple reason that the product of a finite number of p.m.f.'s is always the mass function of a discrete sub-probability measure (i.e., is always summable), but the product of a finite number of p.d.f.'s may not be the density function of a finite a.c. measure (i.e., may not be integrable), as will be seen in Example 3.6 below.

The algebraic and Hilbert space properties of normalized products of density functions have been studied for special classes of a.c. distributions with p.d.f.'s with compact support that are bounded from above and bounded from below away from zero [1], [5]; products of p.m.f.'s and p.d.f.'s have been used in certain pattern-recognition problems [8]; and the "log opinion poll" method for combining probability distributions [7] is an a.c. distribution with normalized density $\prod f_i^{w_i}$, which is similar in structure, but is idempotent since the weights sum to one.

Theorem 3.3. *Let $P_1, P_2, \ldots, P_n$ be absolutely continuous with densities $f_1, \ldots, f_n$ satisfying $0 < \int_{-\infty}^{\infty} \prod_{i=1}^{n} f_i(x)dx < \infty$. Then $\&(P_1, \ldots, P_n)$ exists and the following are equivalent:*

  (i) $Q = \&(P_1, \ldots, P_n)$;
  (ii) *Q is absolutely continuous with density $f(x) = \frac{\prod_{i=1}^{n} f_i(x)dx}{\int_{-\infty}^{\infty} \prod_{i=1}^{n} f_i(y)dy}$;*
  (iii) *Q is the (vague) limit, as $\epsilon \searrow 0$, of the conditional distribution of X_1 given that $|X_i - X_j| < \epsilon$ for all $i, j \in \{1, \ldots, n\}$, where $X_1, \ldots, X_n$ are independent r.v.'s with distributions $P_1, \ldots, P_n$, respectively.*

Proof. First suppose that the densities $\{f_i\}$ are nonnegative simple functions on half-open dyadic intervals $(a, b]$, $a, b \in \{\frac{k}{2^j} : k \in \mathbb{Z}, j \in \mathbb{N}\}$. Without loss of generality (splitting the intervals if necessary), there exists $j_0 \in \mathbb{N}$ and a finite set $K \subset \mathbb{N}$ such that

$$(3.9) \qquad f_i = \sum_{k \in K} c_{j,k} I_{D_k} \quad \text{for all } i = 1, \ldots, n,$$

where $c_{i,k} \geq 0$ for all i, k; and D_k are disjoint intervals $\left(a_k, a_k + \frac{1}{2^{j_0}}\right]$, $a_k = \frac{k}{2^{j_0}}$, $k \in K$. Let $\pi_k = \prod_{i=1}^{n} c_{i,k}$ for all $k \in K$, and note that the compatibility of $P_1, \ldots, P_n$ implies that $\sum_{k \in K} \pi_k > 0$. It will now be shown that $\&(P_1, \ldots, P_n)$ is absolutely continuous with density f, where

$$(3.10) \qquad f(x) = \frac{\prod_{i=1}^{n} f_i(x)}{\int_{-\infty}^{\infty} \prod_{i=1}^{n} f_i(s)ds} = \frac{2^{j_0} \sum_{k \in K} \pi_k I_{D_k}}{\sum_{k \in K} \pi_k} \qquad \text{a.s.}$$

Fix $m \in \mathbb{N}$, and let $a_{k,s} = a_k + \frac{s}{2^{j_0+m}}$. First note that since $f_i = c_{i,k}$ a.s. on D_k for each i and k,

$$(3.11) \qquad \pi_k = (2^{j_0+m})^n \prod_{i=1}^{n} P_i(a_{k,s-1}, a_{k,s}]$$
$$\text{for all } s = 1, \ldots, 2^m, \ m \in \mathbb{N}, \ \text{and } k \in K.$$

By (3.11) and the definitions of $\{D_k\}$ and $\{\mu_j\}$,

$$(3.12) \qquad \mu_{j_0+m} = \sum_{k \in K} \sum_{s=1}^{2^m} \prod_{i=1}^{n} P_i(a_{k,s-1}, a_{k,s}] \delta_{a_{k,s}}$$
$$= \sum_{k \in K} \sum_{s=1}^{2^m} \frac{\pi_k}{2^{(j_0+m)n}} \delta_{a_{k,s}} = \frac{1}{2^{(j_0+m)n}} \sum_{k \in K} \pi_k \sum_{s=1}^{2^m} \delta_{a_{k,s}}.$$

Since m, j_0 and n are fixed, and since $\|\sum_{s=1}^{2^m} \delta_{a_{k,s}}\| = 2^m$, (3.12) implies that

$$(3.13) \qquad \frac{\mu_{j_0+m}}{\|\mu_{j_0+m}\|} = \frac{\sum_{k \in K} \pi_k \sum_{s=1}^{2^m} \delta_{a_{k,s}}}{2^m \sum_{k \in K} \pi_k} = \frac{\sum_{k \in K} \pi_k \frac{1}{2^m} \sum_{s=1}^{2^m} \delta_{a_{k,s}}}{\sum_{k \in K} \pi_k}.$$

But since $\frac{1}{2^m} \sum_{s=1}^{2^m} \delta_{a_{k,s}}$ converges vaguely to the probability measure uniformly distributed on D_k for each $k \in K$, and $\frac{\mu_{j_0+m}}{\|\mu_{j_0+m}\|}$ converges vaguely to $\&(P_1, \ldots, P_n)$ as $m \to \infty$, (3.13) implies (3.10). This completes the proof that $\&(P_1, \ldots, P_n)$ exists and (i) and (ii) are equivalent when the densities are simple functions on dyadic intervals. For the general case, use the standard method to extend this result to general simple functions, and then, since densities are a.s. nonnegative, extend this to finite collections of densities whose product is integrable, via the standard argument of approximating below by simple functions, and using monotone convergence.

For the equivalence of (ii) and (iii), for every $\epsilon > 0$ let $P_{1,\epsilon}$ denote the conditional distribution of X_1 given $\{|X_i - X_j| < \epsilon$ for all $i, j \in \{1, \ldots, n\}\}$. That is, for all Borel sets A,

$$P_{1,\epsilon}(A) = \frac{P(X_1 \in A \text{ and } |X_i - X_j| < \epsilon \text{ for all } i, j \in \{1, \ldots, n\})}{P(|X_i - X_j| < \epsilon \text{ for all } i, j \in \{1, \ldots, n\})},$$

where the denominator is always strictly positive since by hypothesis $\int \prod_{i=1}^{n} f_i(x) dx > 0$. Clearly, $P_{1,\epsilon}$ is absolutely continuous with conditional density $f_{1,\epsilon}$, where the independence of the $\{X_i\}$ implies that

$$(3.14) \qquad f_{1,\epsilon}(x) = \frac{f_1(x) \left(\prod_{i=2}^{n} \int_{x-\epsilon}^{x+\epsilon} f_i(z) dz \right)}{\int_{-\infty}^{\infty} f_1(y) \left(\prod_{i=2}^{n} \int_{y-\epsilon}^{y+\epsilon} f_i(z) dz \right) dy}.$$

Next note that by the definition of derivative and integral,

$$(3.15) \qquad \lim_{\epsilon \to 0} f_1(x) \prod_{i=2}^{n} (2\epsilon)^{-1} \int_{x-\epsilon}^{x+\epsilon} f_i(z) dz = \prod_{i=1}^{n} f_i(x).$$

Letting $f_i^M = \min\{f_i, M\}$ for all $M \in \mathbb{N}$, and all $i \in \{1, \ldots, n\}$, calculate

$$(3.16) \qquad \lim_{\epsilon \to 0} \int f_1(y) \left(\prod_{i=2}^n (2\epsilon)^{-1} \int_{y-\epsilon}^{y+\epsilon} f_i(z)dz \right) dy$$

$$= \lim_{\epsilon \to 0} \lim_{M \to \infty} \int f_1^M(y) \left((2\epsilon)^{-1} \prod_{i=2}^n \int_{y-\epsilon}^{y+\epsilon} f_i^M(z)dz \right) dy$$

$$= \lim_{M \to \infty} \lim_{\epsilon \to 0} \int f_1^M(y) \left(\prod_{i=2}^n (2\epsilon)^{-1} \int_{y-\epsilon}^{y+\epsilon} f_i^M(z)dz \right) dy$$

$$= \lim_{M \to \infty} \int \prod_{i=1}^n f_i^M(y)dy = \int \prod_{i=1}^n f_i(y)dy,$$

where the first equality follows from the monotone convergence theorem, the second since the convergence of $\lim_{\epsilon \to 0} \int f_1^M(y) \left(\prod_{i=2}^n (2\epsilon)^{-1} \int_{y-\epsilon}^{y+\epsilon} f_i^M(z)dz \right) dy$ is uniform in M, the third by (3.15) and the bounded convergence theorem since the integrand is bounded by M^n, and the last by the dominated convergence theorem since by hypothesis,

$$\int_{-\infty}^{\infty} \prod_{i=1}^n f_i(x)dx < \infty.$$

Thus by (3.14), (3.15), and (3.16),

$$\lim_{\epsilon \to 0} f_{1,\epsilon}(x) = \frac{\prod_{i=1}^n f_i(x)}{\int \prod_{i=1}^n f_i(y)dy},$$

proving the equivalence of (ii) and (iii). $\qquad\qquad\qquad\qquad\qquad \square$

Example 3.4. Suppose P_1 is $N(0,1)$ and P_2 is exponentially distributed with mean 1. Then $\&(P_1, P_2)$ is a.c. with p.d.f. $f(x)$ proportional to $e^{-x^2/2}e^{-x} = e^{1/2}e^{-(x+1)^2/2}$ for $x > 0$, which is simply the standard normal shifted to the left one unit, and conditioned to be nonnegative.

Example 3.5. Suppose P_1 and P_2 are both standard Cauchy distributions. Then neither P_1 nor P_2 have finite means, but by Theorem 3.3, $\&(P_1, P_2)$ is a.c. with density $f(x) = c(1 + x^2)^{-2}$ for some $c > 0$, and since $\int_{-\infty}^{\infty} x^2(1 + x^2)^{-2}dx < \infty$, $\&(P_1, P_2)$ has both finite mean and variance. In particular, the conflation of Cauchy distributions is not Cauchy, in contrast to the closure of many classical families under conflation (Theorem 7.1 below). This example also shows that the classes of stable and infinitely divisible distributions are not closed under conflation.

In general, the conflation of a.c. distributions, even an a.c. distribution with itself, may not be a.c., let alone have a density proportional to the product of the densities.

Example 3.6. Let $P_1 = P_2$ be a.c. with p.d.f. $f(x) = (4x)^{-1/2}$ for $0 < x < 1$ (and zero elsewhere). Then $f_1(x)f_1(x) = \frac{1}{4x}$ is not integrable, and no scalar multiple is a p.d.f. However, the conflation $\&(P_1, P_2)$ does exist, and by showing that the normalized mass of μ_j is moving to the left as $j \to \infty$, it can be seen that $\&(P_1, P_2) = \delta_0$, the Dirac delta measure at zero (in particular, the conflation is not even a.c.).

The characterization of the conflation of a.c. distributions as the normalized product of the density functions yields another characterization of conflations of a.c. distributions, an analog of the classical convolution theorem in Fourier analysis [3].

Recall that $g \otimes h$ is the convolution of g and h.

Theorem 3.7 (Convolution theorem for conflations). *Let $P_1, P_2, \ldots, P_n$ be compatible and a.c. with densities $\{f_i\}$ and characteristic functions $\{\psi_i\}$. If $0 < \int_{-\infty}^{\infty} \prod_{i=1}^{n} f_i(x)dx < \infty$ and $\{\psi_i\}$ are L^1, then $\&(P_1, \ldots, P_n)$ exists and is the unique a.c. probability distribution with characteristic function $\psi_{\&(P_1,\ldots,P_n)} = \frac{\psi_1 \otimes \psi_2 \otimes \cdots \otimes \psi_n}{(2\pi)^{n-1} \int_{-\infty}^{\infty} \prod_{i=1}^{n} f_i(x)dx}$.*

Proof. The proof will be given only for the case $n = 2$; the general case follows easily by induction and Theorem 2.8. Suppose ψ_1 and ψ_2 are L^1 and $0 < \int_{-\infty}^{\infty} f_1(x)f_2(x)dx < \infty$. Then

$$
\begin{aligned}
(\psi_1 \otimes \psi_2)(t) &= \int_{-\infty}^{\infty} \psi_2(s)\psi_1(t-s)ds = \int_{-\infty}^{\infty} \psi_2(s)\left[\int_{-\infty}^{\infty} e^{i(t-s)x} f_1(x)dx\right]ds \\
&= \int_{-\infty}^{\infty} f_1(x)e^{itx}\left[\int_{-\infty}^{\infty} \psi_2(s)e^{-isx}ds\right]dx \\
&= \int_{-\infty}^{\infty} 2\pi f_1(x)f_2(x)e^{itx}dx = 2\pi\psi_{\&(P_1,P_2)}(t)\int_{-\infty}^{\infty} f_1(x)f_2(x)dx,
\end{aligned}
$$

where the first equality follows from the definition of convolution; the second by definition of ψ_1; the third by Fubini's theorem since ψ_1 and ψ_2 are absolutely integrable; the fourth by the inverse characteristic function theorem (e.g. [4, Theorem 6.2.3]) since ψ_2 is L^1; and the last equality by Theorem 3.3 since $0 < \int_{-\infty}^{\infty} f_1(x)f_2(x)dx < \infty$. $\qquad\square$

The next example is an application of Theorem 3.7, and shows that the conflation of two standard normal distributions is mean-zero normal with half the variance of the standard normal. An intuitive interpretation of this fact is that if the two standard normals reflect the results of two independent experiments, then combining these results effectively doubles the number of trials, thereby halving the variance of the (sample) means. Normality is always preserved under conflation, as will be seen in Theorem 7.1 below.

Example 3.8. Let $P_1 = P_2$ be $N(0,1)$, so $\psi_1(t) = \psi_2(t) = e^{-t^2/2}$. Then $(\psi_1 \otimes \psi_2)(t) = \int_{-\infty}^{\infty} e^{-(t-s)^2/2}e^{-s^2/2}ds = e^{-t^2/4}\int_{-\infty}^{\infty} e^{-\left(s-\frac{t}{2}\right)^2}ds = e^{-t^2/4}\sqrt{\pi}$, so since $\int f_1(x)f_2(x)dx = \int_{-\infty}^{\infty} \frac{e^{-x^2/2}}{\sqrt{2\pi}}\frac{e^{-x^2/2}}{\sqrt{2\pi}}dx = \frac{1}{2\sqrt{\pi}}$, Theorem 3.7 implies that $\&(P_1, P_2)$ is a.c. with characteristic function $\psi(t) = \frac{\sqrt{\pi}e^{-t^2/4}}{2\pi/2\sqrt{\pi}} = e^{-t^2/4}$, so $\&(P_1, P_2)$ is $N(0, \frac{1}{2})$.

In general, the convolution of characteristic functions of discrete measures may not even exist.

Example 3.9. Let $P = P_1 = P_2 = \delta_0$. Then it is easy to see that $\&(P_1, P_2) = \delta_0$, and $\psi_P(t) \equiv 1$, so $\psi_{P_1} \otimes \psi_{P_2}$ does not even exist.

4. Minimal loss of Shannon information

Replacing several distributions by a single distribution will clearly result in some loss of information, however that is defined. A classical measure of information in a stochastic setting is the Shannon information.

Recall that the *Shannon information* $S_P(A)$ (also called the *surprisal*, or *self-information*) of a probability P for the event $A \in \mathbb{B}$ is $S_P(A) = -\log_2 P(A)$ (so the smaller the value of $P(A)$, the greater the information or surprise). The *information entropy*, which will not be addressed here, is simply the expected value of the Shannon information.

Example 4.1. If P is uniformly distributed on $(0,1)$, and $A = (0, \frac{1}{4}) \cup (\frac{1}{2}, \frac{3}{4})$, then $P(A) = \frac{1}{2}$, so $S_P(A) = -\log_2(P(A)) = 1$. Thus exactly one bit of information is obtained by observing A, namely, that the value of the second binary digit is 0.

Definition 4.2. The (*joint*) *Shannon information of $P_1, P_2, \ldots, P_n$ for the event $A \in \mathbb{B}$* is

$$S_{\{P_1,\ldots,P_n\}}(A) = S_P(X_1 \in A, \ldots, X_n \in A) = \sum_{i=1}^{n} S_{P_i}(A) = -\log_2 \prod_{i=1}^{n} P_i(A),$$

where $\{X_i\}$ are independent random variables with distributions $\{P_i\}$, respectively, and the *loss between the Shannon information of $Q \in \mathcal{P}$ and $P_1, \ldots, P_n$ for the event $A \in \mathbb{B}$* is $S_{\{P_1,\ldots,P_n\}}(A) - S_Q(A)$ if $\prod_{i=1}^{n} P_i(A) > 0$, and is 0 if $Q(A) = \prod_{i=1}^{n} P_i(A) = 0$.

Note that the maximum loss is always nonnegative (taking $A = \Omega$).

The next theorem characterizes conflation as the minimizer of loss of Shannon information.

Theorem 4.3. *If $P_1, \ldots, P_n \in \mathcal{P}$ satisfies $\|\mu_\infty(P_1, P_2, \ldots, P_n)\| > 0$, then*

   (i) *the conflation $\&(P_1, P_2, \ldots, P_n)$ exists;*
  (ii) *for every $Q \in \mathcal{P}$, the maximum loss between the Shannon information of Q and $P_1, \ldots, P_n$ is at least $\log_2(\|\mu_\infty(P_1, P_2, \ldots, P_n)\|^{-1})$; and*
 (iii) *the bound in* (ii) *is attained if and only if $Q = \&(P_1, P_2, \ldots, P_n)$.*

Proof. Fix $P_1, \ldots, P_n \in \mathcal{P}$, and for brevity, let $\mu_j = \mu_j(P_1, P_2, \ldots, P_n)$ for all $j \in \mathbb{N}$, and $\mu_\infty = \mu_\infty(P_1, P_2, \ldots, P_n)$. For (i), note that by Theorem 2.3, μ_j converges vaguely to μ_∞, and $\lim_{j\to\infty} \|\mu_j\| = \|\mu_\infty\| > 0$, so $\mu_j\|\mu_j\|^{-1}$ converges vaguely to the probability measure $\mu_\infty\|\mu_\infty\|^{-1}$, which implies that $\&(P_1, P_2, \ldots, P_n)$ exists.

For (ii) and (iii), fix $Q \in \mathcal{P}$, and let $\& = \&(P_1, P_2, \ldots, P_n)$. It must be shown that

(4.1a) $S_{\{P_1,\ldots,P_n\}}(A) - S_Q(A) \geq \log_2(\|\mu_\infty\|^{-1})$ for some Borel A,

(4.1b) $S_{\{P_1,\ldots,P_n\}}(A) - S_Q(A) > \log_2(\|\mu_\infty\|^{-1})$ for some Borel A if $Q \neq \&$,

and

(4.1c) $S_{\{P_1,\ldots,P_n\}}(A) - S_Q(A) \leq \log_2(\|\mu_\infty\|^{-1})$ for all Borel A if $Q = \&$.

By definition of Shannon information and since $\log_2(x)$ is increasing, (4.1a)–(4.1c) are equivalent to

(4.2a) $$\frac{Q(A)}{\prod_{i=1}^{n} P_i(A)} \geq \|\mu_\infty\|^{-1} \text{ for some Borel } A,$$

(4.2b) $$\frac{Q(A)}{\prod_{i=1}^{n} P_i(A)} > \|\mu_\infty\|^{-1} \text{ for some Borel } A \text{ if } Q \neq \&,$$

(4.2c) $$\frac{Q(A)}{\prod_{i=1}^{n} P_i(A)} \leq \|\mu_\infty\|^{-1} \text{ for all Borel } A \text{ if } Q = \&.$$

To establish (4.2a), fix ϵ, $\|\mu_\infty\|^{-1} > \epsilon > 0$. By Theorem 2.3, $\|\mu_j\| \to \|\mu_\infty\|$ as $j \to \infty$, so there exists $j^* \in \mathbb{N}$ such that

(4.3) $$\|\mu_{j^*}\|^{-1} > \|\mu_\infty\|^{-1} - \epsilon > 0.$$

For each $k \in \mathbb{Z}$, let $q_k = Q\left(\frac{k}{2^{j^*}}, \frac{k+1}{2^{j^*}}\right]$, and $p_k = \prod_{i=1}^{n} P_i\left(\frac{k}{2^{j^*}}, \frac{k+1}{2^{j^*}}\right]$. Note that by the definition of $\{\mu_j\}$,

(4.4) $$\|\mu_{j^*}\| = \sum_{k \in \mathbb{Z}} p_k.$$

By (4.3), since Q is a probability, (4.4) implies that $1 = \sum_{k \in \mathbb{Z}} q_k = \sum_{k \in \mathbb{Z}} p_k \|\mu_{j^*}\|^{-1}$, so there exists $k^* \in \mathbb{Z}$ such that

(4.5) $$q_{k^*} \geq p_{k^*} \|\mu_{j^*}\|^{-1} > 0.$$

Hence, by (4.3) and (4.5) and the definition of $\{p_k\}$ and $\{q_k\}$,

(4.6) $$\frac{Q\left(\frac{k^*}{2^{j^*}}, \frac{k^*+1}{2^{j^*}}\right]}{\prod_{i=1}^{n} P_i\left(\frac{k^*}{2^{j^*}}, \frac{k^*+1}{2^{j^*}}\right]} \geq \|\mu_\infty\|^{-1} - \epsilon.$$

By Lyapounov's theorem, the range of a finite-dimensional vector measure is closed (e.g. [9] or [6, Theorem 1.1]), so since ϵ was arbitrarily small, this proves (4.2a).

To prove (4.2b), suppose $Q \neq \&$. Then there exists a $c > 0$, $k^* \in \mathbb{Z}$ and $j^* \in \mathbb{N}$, such that for $D = \left(\frac{k^*}{2^{j^*}}, \frac{k^*+1}{2^{j^*}}\right]$, $\&(D) > 0$ and $Q(D) > \&(D) + c\mu_\infty(D)$. Since $\& = \frac{\mu_\infty}{\|\mu_\infty\|}$, this implies that

(4.7) $$\frac{Q(D)}{\mu_\infty(D)} > \|\mu_\infty\|^{-1} + c.$$

Since $\mu_j(D) \to \mu_\infty(D)$ as $j \to \infty$ by Theorem 2.3(ii), (4.7) implies there exists an $m \in \mathbb{N}$ so that

(4.8) $$\frac{Q(D)}{\mu_{j^*+m}(D)} > \|\mu_\infty\|^{-1} + \frac{c}{2}.$$

Note that $D = \bigcup_{k \in J} D_k$, where $D_k = \left(\frac{k}{2^{j^*+m}}, \frac{k+1}{2^{j^*+m}}\right]$ and $J = \{k^*2^m, k^*2^m + 1, \dots, k^*2^m + 2^m - 1\}$. Next, note that since $\frac{\sum a_k}{\sum b_k} \leq \max_k \left\{\frac{a_k}{b_k}\right\}$ for nonnegative $\{a_k, b_k\}$, there exists $M \in J$ such that

(4.9) $$\frac{\sum_{k \in J} Q(D_k)}{\sum_{k \in J} \prod_{i=1}^{n} P_i(D_k)} \leq \max_{k \in J} \frac{Q(D_k)}{\prod_{i=1}^{n} P_i(D_k)} = \frac{Q(D_M)}{\prod_{i=1}^{n} P_i(D_M)}.$$

Then

(4.10) $$\frac{Q(D_M)}{\prod_{i=1}^{n} P_i(D_M)} \geq \frac{Q(D)}{\mu_{j^*+m}(D)} > \|\mu_\infty\|^{-1},$$

where the first inequality in (4.10) follows by (4.9) and since

$$\mu_{j^*+m}(D) = \sum_{k \in J} \prod_{i=1}^{n} P_i(D_k),$$

and the second follows by (4.8). This proves (4.2b). Finally, suppose that $\& = Q$. Since the class of sets $\left\{ \left(\frac{k}{2^j}, \frac{k+1}{2^j} \right] : j \in \mathbb{N}, k \in \mathbb{Z} \right\}$ generates the Borel sigma algebra on $\mathbb{R}$, and since $Q = \& = \mu_\infty \|\mu_\infty\|^{-1}$, to prove (4.2c) it is enough to show that for all $j \in \mathbb{N}$, all finite sets $J \subset \mathbb{N}$ and all $D = \bigcup_{k \in J} \left(\frac{k}{2^j}, \frac{k+1}{2^j} \right]$,

(4.11)
$$\mu_\infty(D) \le \prod_{i=1}^{n} P_i(D).$$

However, since $\lim_{j \to \infty} \mu_j(D) = \mu_\infty(D)$ and $\mu_{j^*}(D) = \prod_{i=1}^{n} P_i(D)$, (4.11) follows by Theorem 2.3(i).   □

Corollary 4.4. *If $P_1, \ldots, P_n \in \mathcal{P}$ are discrete with common atoms $A \ne \emptyset$, then $\&(P_1, \ldots, P_n)$ is the unique Borel probability distribution that minimizes the maximum loss of Shannon information between single Borel probability distributions and $P_1, P_2, \ldots, P_n$.*

Proof. It is easy to check that for discrete distributions $P_1, \ldots, P_n$ with common atoms A, $\|\mu_\infty(P_1, \ldots, P_n)\| = \sum_{x \in A} \prod_{i=1}^{n} P_i(x)$, which by the definition of A is strictly positive. The conclusion then follows immediately from Theorems 3.1 and 4.3.   □

Theorem 4.5. *If $P_1, P_2, \ldots, P_n$ are a.c. with densities $f_1, \ldots, f_n$, satisfying*

$$0 < \int_{-\infty}^{\infty} \prod_{i=1}^{n} f_i(x)dx < \infty,$$

then there are Borel probability distributions $\{P_{i,j} : i \in \{1, \ldots, n\}, j \in \mathbb{N}\}$ such that

   (i) *for all i, $P_{i,j}$ converges vaguely to P_i as $j \to \infty$,*
   (ii) *$\&(P_{1,j}, \ldots, P_{n,j})$ is the unique minimizer of loss of Shannon information from $P_{1,j}, \ldots, P_{n,j}$, and*
   (iii) *$\&(P_1, \ldots, P_n)$ is the vague limit of $\&(P_{1,j}, \ldots, P_{n,j})$ as $j \to \infty$.*

Proof. For each $i \in \{1, \ldots, n\}$ and $j \in \mathbb{N}$, let $P_{i,j} = \mu_j(P_i)$, and note that $\mu_j(P_i)$ is a discrete p.m. for all i and j, and by Theorem 2.3(iv), $\mu_j(P_i) \to P_i$ vaguely as $j \to \infty$, which proves (i). Since $\{P_{i,j} : i \in \{1, \ldots, n\}\}$ are compatible for all $j \in \mathbb{N}$, $\mu_j(P_1), \ldots, \mu_j(P_n)$ are discrete with at least one common atom, so by Theorem 3.1 and Corollary 4.4, $\&(P_{1,j}, \ldots, P_{n,j}) = \sum_{k \in Z} \prod_{i=1}^{n} P_i((k-1)2^{-j}, k2^{-j}]$ is the unique minimizer of the maximum loss of Shannon information between single Borel p.m.'s and $\{P_{i,j} : i \in \{1, \ldots, n\}\}$, which proves (ii). Finally, note that for all $j \in \mathbb{N}$, $\prod_{i=1}^{n} P_i((k-1)2^{-j}, k2^{-j}] = \prod_{i=1}^{n} \mu_j(k2^{-j})$, so by the definition of $\{\mu_j\}$, $\mu_j(P_1, \ldots, P_n) = \sum_{k \in Z} \prod_{i=1}^{n} \mu_j(k2^{-j})\delta_{k2^{-j}}$, and $\|\mu_j(P_1, \ldots, P_n)\| = \sum_{k \in \mathbb{Z}} \prod_{i=1}^{n} \mu_j(k2^{-j}) > 0$. Hence, by Theorem 3.3,

$$\&(P_{1,j}, \ldots, P_{n,j}) = \frac{\sum_{k \in \mathbb{Z}} \prod_{i=1}^{n} \mu_j(k2^{-j})\delta_{k2^{-j}}}{\sum_{k \in \mathbb{Z}} \prod_{i=1}^{n} \mu_j(k2^{-j})} = \frac{\mu_j(P_1, \ldots, P_n)}{\|\mu_j(P_1, \ldots, P_n)\|}$$

converges vaguely to $\&(P_1, \ldots, P_n)$, proving (iii).   □

5. Minimax likelihood ratio consolidations and proportional consolidations

In classical hypotheses testing, a standard technique to decide from which of n known distributions the given data actually came is to maximize the likelihood ratios, that is, the ratios of the p.m.f.'s or p.d.f.'s. Analogously, when the objective is not to decide from which of n known distributions $P_1, \ldots, P_n$ the data came, but rather to decide how best to consolidate data from those input distributions into a single (output) distribution P, one natural criterion is to choose P so as to make the ratios of the likelihood of observing x under P to the likelihood of observing x under *all* of the (independent) distributions $\{P_i\}$ as close as possible. This motivates the notion of a minimax likelihood ratio.

Definition 5.1. A discrete probability distribution $P^* \in \mathcal{P}$ (with p.m.f. p^*) is the *minimax likelihood ratio (MLR) consolidation of discrete distributions* $P_1, \ldots, P_n$ (with p.m.f.'s $\{p_i\}$) if

$$\min_{\text{p.m.f.'s } p} \left\{ \max_{x \in \mathbb{R}} \frac{p(x)}{\prod_{i=1}^n p_i(x)} - \min_{x \in \mathbb{R}} \frac{p(x)}{\prod_{i=1}^n p_i(x)} \right\}$$

is attained by $p = p^*$ (where $0/0 := 1$). Similarly, an a.c. distribution $P^* \in \mathcal{P}$ (with p.d.f. f^*) is the *MLR consolidation of a.c. distributions* $P_1, \ldots, P_n$ (with p.d.f.'s $f_1, \ldots, f_n$) if

$$\min_{\text{p.d.f.'s } f} \left\{ \operatorname{ess\,sup}_{x \in \mathbb{R}} \frac{f(x)}{\prod_{i=1}^n f_i(x)} - \operatorname{ess\,inf}_{x \in \mathbb{R}} \frac{f(x)}{\prod_{i=1}^n f_i(x)} \right\}$$

is attained by f^*.

The min-max terms in (5.1) and (5.2) are similar to the min-max criterion for loss of Shannon information (Theorem 4.3), whereas the others are dual max-min criteria. Just as conflation was shown to minimize the loss of Shannon information, conflation will now be shown to also be the MLR consolidation of the given input distributions.

Theorem 5.2. *If $P_1, \ldots, P_n \in \mathcal{P}$ are discrete with at least one common atom, or are a.c. with p.d.f.'s $\{f_i\}$ satisfying $0 < \int \prod_{i=1}^n f_i(x)dx < \infty$, then $\&(P_1, \ldots, P_n)$ is the unique MLR consolidation of $P_1, \ldots, P_n$.*

Proof. First consider the discrete case, let $\{p_i\}$ denote the p.m.f.'s of $\{P_i\}$, respectively, and let $\emptyset \neq A \subset \mathbb{R}$ denote the common atoms of $\{P_i\}$, i.e. $A = \{x \in \mathbb{R} : \prod_{i=1}^n p_i(x)\} > 0$. By Theorem 3.1, $\&(P_1, \ldots, P_n)$ is discrete with p.m.f. $p^*(x) = \frac{\prod_{i=1}^n p_i(x)}{\sum_{y \in A} \prod_{i=1}^n p_i(y)}$. For each p.m.f. p, let

$$\Delta(p) = \sup_{x \in \mathbb{R}} \frac{p(x)}{\prod_{i=1}^n p_i(x)} - \inf_{x \in \mathbb{R}} \frac{p(x)}{\prod_{i=1}^n p_i(x)} \, .$$

Then, since $p^*(x) = 0$ for every $x \in A^c$, it follows from the definition of p^* (and the convention $0/0 := 1$) that $\Delta(p^*) = \left(\sum_{y \in A} \prod_{i=1}^n p_i(y) \right)^{-1} - 1 \geq 0$. Thus, to establish the theorem for $P_1, \ldots, P_n$ discrete, it suffices to show that for all p.m.f.'s p

$$(5.1) \qquad \Delta(p) \geq \left(\sum_{y \in A} \prod_{i=1}^n p_i(y) \right)^{-1} - 1, \text{ with equality if and only if } p = p^*.$$

If $\sum_{y \in A} p(y) < 1$, then there exists an $x_0 \in A^c$ with $p(x_0) > 0$, so $\frac{p(x_0)}{\prod_{i=1}^n p_i(x_0)} = \infty$ and $\Delta(p) = \infty$, so (5.1) is trivial. On the other hand if $\sum_{y \in A} p(y) = 1$, then $\min_{x \in \mathbb{R}} \frac{p(x)}{\prod_{i=1}^n p_i(x)} \leq 1$, which implies that $\Delta(p) \geq \max_{x \in \mathbb{R}} \frac{p(x)}{\prod_{i=1}^n p_i(x)} - 1$ for all p, and the argument in the proof of Theorem 4.3 shows equality holds if and only if $\frac{p(x)}{\prod_{i=1}^n p_i(x)}$ is constant, i.e. if and only if $p = p^*$. This proves (5.1) and completes the argument when $\{P_i\}$ are discrete.

For the a.c. conclusion, fix $\{P_i\}$ a.c. with p.d.f.'s satisfying $0 < \int \prod_{i=1}^n f_i(x) dx < \infty$. By Theorem 3.3 $\&(P_1, \ldots, P_n)$ is a.c. with p.d.f. $f^*(x) = \frac{\prod_{i=1}^n f_i(x)}{\int \prod_{i=1}^n f_i(y) dy}$. For each p.d.f. f, let

$$\Delta(f) = \text{ess sup}_{x \in \mathbb{R}} \frac{f(x)}{\prod_{i=1}^n f_i(x)} - \text{ess inf}_{x \in \mathbb{R}} \frac{f(x)}{\prod_{i=1}^n f_i(x)}.$$

Case 1. $\int \prod_{i=1}^n f_i(x) dx \in (0, 1]$, $\prod_{i=1}^n f_i(x) > 0$ a.s. (e.g., $\{P_i\}$ arbitrary normal distributions). Then since $\prod_{i=1}^n f_i(x) > 0$, $\frac{f^*(x)}{\prod_{i=1}^n f_i(x)} = \frac{1}{\int \prod_{i=1}^n f_i(y) dy}$, a.s., which is constant, so $\Delta(f^*) = 0$. Thus it suffices to show that for all f as in Case 1,

(5.2) $$\Delta f(x) \geq 0 \quad \text{with equality if and only if } f = f^*.$$

If f is not positive a.s., then $\text{ess inf} \frac{f}{\prod_{i=1}^n f_i} = 0$ since $\prod_{i=1}^n f_i(x) > 0$ a.s., so $\Delta(f) = \text{ess sup} \frac{f}{\prod_{i=1}^n f_i} > 0$, and the inequality in (5.2) is satisfied. On the other hand, if $f > 0$ a.s., then $\Delta(f) = \text{ess sup}_{x \in \mathbb{R}} \frac{f(x)}{\prod_{i=1}^n f_i(x)} - \text{ess inf}_{x \in \mathbb{R}} \frac{f(x)}{\prod_{i=1}^n f_i(x)} \geq 0$, with equality if and only if $\frac{f(x)}{\prod_{i=1}^n f_i(x)}$ is constant a.s.; i.e. if and only if $f = f^*$ a.s., which completes the argument for Case 1.

The three other cases

$$\left\{ \int \prod_{i=1}^n f_i(x) dx \in (0, 1], \prod_{i=1}^n f_i(x) \text{ not } > 0 \text{ a.s.} \right\},$$

$$\left\{ \int \prod_{i=1}^n f_i(x) dx \in (1, \infty), \prod_{i=1}^n f_i(x) > 0 \text{ a.s.} \right\},$$

$$\left\{ \int \prod_{i=1}^n f_i(x) dx \in (1, \infty), \prod_{i=1}^n f_i(x) \text{ not } > 0 \text{ a.s.} \right\}$$

follow similarly. $\square$

If the $\{P_i\}$ are a.c. but do not satisfy the integrability condition in the hypotheses of Theorem 5.2, both parts of the conclusion of Theorem 5.2 may fail: the conflation may not be MLR; and MLR distributions may not be unique.

Example 5.3. Let $n = 2$, and $P_1 = P_2$ be as in Example 3.6, so the conflation $\&(P_1, P_2)$ exists and is δ_0, which is not MLR for P_1, P_2 since it is not even a.c. However, every a.c. distribution with p.d.f. $f_\alpha(x) = \alpha x^{\alpha-1}$ for $x \in (0, 1)$ (and $= 0$ otherwise), $0 < \alpha \leq \frac{1}{4}$, is MLR for P_1, P_2. To see this, recall that $\prod_{i=1}^n f_i(x) = (4x)^{-1}$ for $x \in (0, 1)$, and $= 0$ otherwise. Thus $\frac{f_\alpha(x)}{\prod_{i=1}^n f_i(x)} = 4x f_\alpha(x) = 4\alpha x^\alpha$ for $x \in (0, 1)$, so $\text{ess sup}_{x \in \mathbb{R}} \frac{f_\alpha(x)}{\prod_{i=1}^n f_i(x)} = 1$, since off $(0, 1)$, $\frac{f_\alpha(x)}{\prod_{i=1}^n f_i(x)} = 1$, and on $(0, 1)$, $\text{ess sup}_{x \in \mathbb{R}} \frac{f_\alpha(x)}{\prod_{i=1}^n f_i(x)} = 4\alpha \leq 1$. Next, $\text{ess inf}_{x \in \mathbb{R}} \frac{f_\alpha(x)}{\prod_{i=1}^n f_i(x)} = 0$ since $\frac{f_\alpha(x)}{\prod_{i=1}^n f_i(x)} = 4\alpha x^\alpha$ for $x \in (0, 1)$. Thus $\Delta(f_\alpha) = 1$, so to show f_α is MLR, requires showing

that $\Delta(f) \geq 1$ for all p.d.f.'s f. Fix f, and note that if $\operatorname{ess\,inf}_{x \in \mathbb{R}} \frac{f(x)}{\prod_{i=1}^{n} f_i(x)} = \delta > 0$, then on $(0, 1)$, $\frac{f(x)}{\prod_{i=1}^{n} f_i(x)} = 4xf(x) \geq \delta$ a.s., so $f(x) \geq \frac{\delta}{4x}$ a.s., which cannot be a density since it is not integrable. Hence, $\operatorname{ess\,inf}_{x \in \mathbb{R}} \frac{f(x)}{\prod_{i=1}^{n} f_i(x)} = 0$. But $\operatorname{ess\,sup}_{x \in \mathbb{R}} \frac{f(x)}{\prod_{i=1}^{n} f_i(x)} \geq 1$, since f is a.s. nonnegative and $\prod_{i=1}^{n} f_i(x) = 0$ for all x not in $(0, 1)$. Thus $\Delta(f) \geq 1$, so f_α is MLR.

In the underlying problem of consolidating the independent distributions $P_1, \ldots, P_n$ into a single distribution Q, a criterion similar to MLR is to require that Q reflect the relative likelihoods of identical individual outcomes under the $\{P_i\}$. For example, if the likelihood of all the experiments $\{P_i\}$ observing the identical outcome x is twice that of the likelihood of all the experiments $\{P_i\}$ observing y, then $Q(x)$ should also be twice as large as $Q(y)$. This motivates the notion of proportional consolidation.

Definition 5.4. For discrete $P_1, \ldots, P_n \in \mathcal{P}$ with p.m.f.'s $p_1, \ldots, p_n$, respectively, the discrete distribution $Q \in \mathcal{P}$ is a *proportional consolidation of* $P_1, \ldots, P_n$ if its p.m.f. q satisfies

$$\frac{q(x)}{q(y)} = \frac{\prod_{i=1}^{n} p_i(x)}{\prod_{i=1}^{n} p_i(y)} \qquad \text{for all } x, y \in \mathbb{R}.$$

Similarly, for a.c. $P_1, \ldots, P_n \in \mathcal{P}$ with p.d.f.'s $f_1, \ldots, f_n$, respectively, the a.c. distribution $Q \in \mathcal{P}$ is a *proportional consolidation of* $P_1, \ldots, P_n$ if its p.d.f. g satisfies

$$\frac{g(x)}{g(y)} = \frac{\prod_{i=1}^{n} f_i(x)}{\prod_{i=1}^{n} f_i(y)} \quad \text{for Lebesgue-almost-all } x, y \in \mathbb{R}.$$

Theorem 5.5. *If $P_1, \ldots, P_n \in \mathcal{P}$ are discrete with at least one common atom, or are a.c. with p.d.f.'s $\{f_i\}$ satisfying $0 < \int \prod_{i=1}^{n} f_i(x)dx < \infty$, then the conflation $\&(P_1, \ldots, P_n)$ is the unique proportional consolidation of $P_1, \ldots, P_n$.*

Proof. First consider the case where $\{P_i\}$ are discrete, and let $\{p_i\}$ be the p.m.f.'s for $\{P_i\}$, respectively. By Theorem 3.1 again, $\&(P_1, \ldots, P_n)$ is discrete with p.m.f. $p^*(x) = \frac{\prod_{i=1}^{n} p_i(x)}{\sum_{y \in \mathbb{R}} \prod_{i=1}^{n} p_i(y)}$ for all $x \in \mathbb{R}$. Thus $\frac{p^*(x)}{p^*(y)} = \frac{\prod_{i=1}^{n} p_i(x)}{\prod_{i=1}^{n} p_i(y)}$, so $\&(P_1, \ldots, P_n)$ is a proportional consolidation of $P_1, \ldots, P_n$. To see that $\&(P_1, \ldots, P_n)$ is the unique proportional consolidation, suppose $Q \neq \&(P_1, \ldots, P_n)$, and set $q(x) = Q(x)$ for all $x \in \mathbb{R}$. Since, $Q \neq \&(P_1, \ldots, P_n)$, it follows from Theorem 3.1 that there exist $x, y \in \mathbb{R}$ so that $q(x) > \frac{\prod_{i=1}^{n} p_i(x)}{\sum_{z \in \mathbb{R}} \prod_{i=1}^{n} p_i(z)}$ and $q(y) < \frac{\prod_{i=1}^{n} p_i(y)}{\sum_{z \in \mathbb{R}} \prod_{i=1}^{n} p_i(z)}$, so $\frac{q(x)}{q(y)} > \frac{\prod_{i=1}^{n} p_i(x)}{\prod_{i=1}^{n} p_i(y)}$, and Q is not a proportional consolidation of $P_1, \ldots, P_n$. The case where $P_1, \ldots, P_n$ are a.c. follows similarly, again using Theorem 3.3 in place of Theorem 3.1. $\square$

Here, too, the conclusion for a.c. distributions may fail if the integrability hypothesis condition is not satisfied.

Example 5.6. Let $n = 2$, and $P_1 = P_2$ be as in Example 3.5, so again $\prod_{i=1}^{n} f_i(x) = (4x)^{-1}$ for $x \in (0, 1)$, and $= 0$ otherwise. This implies that $\frac{\prod_{i=1}^{n} f_i(x)}{\prod_{i=1}^{n} f_i(y)} = \frac{y}{x}$ for Lebesgue almost all $x, y \in (0, 1)$. But there are no p.d.f.'s f with support on $(0, 1)$ such that $\frac{f(x)}{f(y)} = \frac{y}{x}$ a.s., since then for fixed y, $f(x) = \frac{yf(y)}{x}$ for almost all $x \in (0, 1)$, and $\int_0^1 cx^{-1}dx = 0$ if $c = 0$ and $= \infty$ if $c > 0$. Thus, there is no proportional consolidation of this P_1, P_2 (in contrast to the conclusion of Example 5.3 for these same distributions, where it was seen that there are many MLR consolidations).

6. Conflations of normal distributions

In describing the method used to obtain values for the fundamental physical constants from the input data, NIST explains that certain data "are the means of tens of individual values, with each value being the average of about ten data points" ([13, p. 679]), and predicates interpretation of some of their conclusions on the condition "If the probability distribution associated with each input datum is assumed to be normal" ([11, p. 483]). After comparing the most recent (2006) results from electrical watt-balance and from silicon-lattice sphere experiments used to estimate Planck's constant, however, NIST determined that the means and standard deviations of several distributions of input data were not sufficiently close, and reported that their "data analysis uncovered two major inconsistencies with the input data", conceding that the resulting official NIST 2006 set of recommended values for the fundamental physical constants "does not rest on as solid a foundation as one might wish" ([12, p. 54]). In order to eliminate this perceived inconsistency, the NIST task group "ultimately decided that ... the a priori assigned uncertainties of the input data involved in the two discrepancies would be weighted by the multiplicative factor 1.5", which "reduced the discrepancies to a level comfortably between two standard deviations" ([12, p. 54]).

But if the various input distributions are all normal, for example, as in the NIST assumption, then every interval centered at the unknown positive true value of Planck's constant has a positive probability of occurring in every one of the independent experiments. If the input data distributions happen to have different means and variances, that does not imply the input is "inconsistent". Thus in consolidating data from several independent sources, special attention should be paid to the normal case.

The conflation of normal distributions has several important properties – it is itself normal (hence unimodal), and in addition to minimizing the loss of Shannon information (Theorem 4.3) and being the unique MLR consolidation (Theorem 5.2) and the unique proportional consolidation (Theorem 5.5), the conflation of normal distributions also yields the classical weighted mean squares and best linear unbiased estimators for general unbiased data, and the maximum likelihood estimators for normally-distributed unbiased input data.

Theorem 6.1. *If P_i is $N(\mu_i, \sigma_i^2)$, $i = 1, \ldots, n$, then*

$$\&(P_1, \ldots, P_n) = N\left(\frac{\sum_{i=1}^n \frac{\mu_i}{\sigma_i^2}}{\sum_{i=1}^n \frac{1}{\sigma_i^2}}, \frac{1}{\sum_{i=1}^n \sigma_i^{-2}}\right).$$

Proof. By Theorem 3.3, $\&(P_1, \ldots, P_n)$ is a.c. with density proportional to the product of the densities for each distribution, and the conclusion then follows immediately from the definition of normal densities and a routine calculation by completing the square. □

Example 6.2. If P_1 is $N(1, 1)$ and P_2 is $N(2, 4)$, then $\&(P_1, P_2)$ is $N(\frac{6}{5}, \frac{4}{5})$.

The mean of the conflations of normals which was given in Theorem 6.1, $\sum_{i=1}^n \mu_i \sigma_i^{-2} \left(\sum_{i=1}^n \sigma_i^{-2}\right)^{-1}$, is precisely the value of the weighted least squares estimate given by Aitken's generalization of the Gauss-Markov Theorem, and this simple observation will next be exploited to obtain several conclusions relating conflation and statistical estimators.

First, however, it must be remarked that the mean of the conflation is not in general the same as the weighted least squares estimate. Conflation disregards outlier or "inconsistent" data values, whereas weighted least squares gives full weight to all values. For instance, if one of the input distributions includes negative entries (e.g., is reported as a true Gaussian), and the others do not, then conflation eliminates the negative values. The following example for the uniform distribution illustrates this, and the same argument can easily be applied to other distributions such as truncated normals (Theorem 7.2 below).

Example 6.3. Let P_1 be $U(0, 1)$ and P_2 be $U(-0.1, 1)$. By Theorem 3.3, the conflation of P_1 and P_2 is $\&(P_1, P_2) = U(0, 1)$, which ignores the negative values of P_2 and has mean $\frac{1}{2}$. The weighted least squares estimate, however, is easily seen to be $\left(\frac{12}{1} + \frac{12}{1.1^2}\right)^{-1} \left(\frac{12}{2} + \left(\frac{9}{20}\right)\left(\frac{12}{1.1^2}\right)\right) < .48$.

To establish the link between conflation and statistical estimators, recall that a random variable X is an unbiased estimator of an unknown parameter θ if $EX = \theta$, and note that if X is a r.v., then $N(X, \sigma^2)$ is a *random normal distribution* with variable mean X and fixed variance σ^2.

Theorem 6.4. *If $X_1, \ldots, X_n$ are independent unbiased estimators of θ with finite variances $\sigma_1^2, \ldots, \sigma_n^2$, respectively, then $\Theta = mean(\&(N_1, \ldots, N_n))$ is the best linear unbiased estimator for θ, where $\{N_i\}$ are the random normal distributions $N_i = N(X_i, \sigma_i^2)$, $i = 1, \ldots, n$.*

Proof. By Theorem 6.1, where $\{\mu_i\}$ and $\{\sigma_i^2\}$ are the means and variances of $\{N_i\}$, $\&(N_1, \ldots, N_n)$ is $N\left(\sum_{i=1}^n \mu_i \sigma_i^{-2} \left(\sum_{i=1}^n \sigma_i^{-2}\right)^{-1}, \left(\sum_{i=1}^n \sigma_i^{-2}\right)^{-1}\right)$, respectively. Since N_i is $N(X_i, \sigma_i^2)$ for each $i = 1, \ldots, n$, where the $\{X_i\}$ are r.v.'s, this implies that $\&(N_1, \ldots, N_n)$ is the random distribution $N\left(\sum_{i=1}^n X_i \sigma_i^{-2} \left(\sum_{i=1}^n \sigma_i^{-2}\right)^{-1}, \left(\sum_{i=1}^n \sigma_i^{-2}\right)^{-1}\right)$, so

$$(6.1) \qquad mean(\&(N_1, \ldots, N_n)) = \left(\sum_{i=1}^n \sigma_i^{-2}\right)^{-1} \sum_{i=1}^n X_i \sigma_i^{-2}.$$

Since the right hand side of (6.1) is the classical weighted least squares estimator for θ, Aitken's generalization of the Gauss-Markov Theorem (e.g. [1], [14, Theorem 7.8a]) implies that it is the best linear unbiased estimator for θ. $\qquad \square$

Note that normality of the distributions is in the *conclusion*, not the hypotheses, of Theorem 6.4. If, in addition, the underlying data distributions are normal, this estimator is even a maximum likelihood estimator.

Theorem 6.5. *If $X_1, \ldots, X_n$ are independent normally-distributed unbiased estimators of θ with finite variances $\sigma_1^2, \ldots, \sigma_n^2$, respectively, then*

$$\Theta = mean(\&(N_1, \ldots, N_n))$$

is a maximum likelihood estimator for θ, where $\{N_i\}$ are the random normal distributions $N_i = N(X_i, \sigma_i^2)$, $i = 1, \ldots, n$.

Proof. Analogous to proof of Theorem 6.4, using [14, Theorem 7.8b]. $\qquad \square$

7. Closure and truncation properties of conflation

If input data distributions are of a particular form, it is often desirable that consolidation of the input also have that same form. Theorem 6.1 showed that the conflation of normal distributions is always normal, and the next theorem shows that many other classical families of distributions are closed under conflation.

Recall that a discrete probability distribution is *Bernoulli* with parameter $p \in [0,1]$ if its p.m.f. is $p(1) = 1 - p(0) = p$, is *geometric* with parameter $p \in [0,1]$ if its p.m.f. is $p(k) = (1-p)^{k-1}p$ for all $k \in \mathbb{N}$, is *discrete uniform* on $\{1, 2, \ldots, n\}$ if its p.m.f. is $p(k) = n^{-1}$ for all $k \in \{1, 2, \ldots, n\}$, is *Zipf* with parameters $\alpha > 0$ and $n \in \mathbb{N}$ if its p.m.f. is proportional to $k^{-\alpha}$ for all $k \in \{1, 2, \ldots, n\}$, and is *Zeta* with parameter $\alpha > 1$ if its p.m.f. is proportional to $k^{-\alpha}$ for all $k \in \mathbb{N}$. Also recall that an a.c. probability distribution is *gamma* with parameters $\alpha \in \mathbb{N}$ and $\beta > 0$ if its p.d.f. is proportional to $x^{\alpha-1}e^{-x/\beta}$ for $x > 0$, is *beta* with parameters $\alpha > 1$ and $\beta > 1$ if its p.d.f. is proportional to $x^{\alpha-1}(1-x)^{\beta-1}$ for $0 < x < 1$, is *uniform* on (a, b) for $a < b$ if its p.d.f. is constant $(b-a)^{-1}$ for $a < x < b$, is standard *LaPlace* (or double-exponential) with parameter $\alpha > 0$ if its p.d.f. is proportional to $e^{-|x|/\beta}$, $-\infty < x < \infty$, is *Pareto* with parameters $\alpha > 0$ and $\beta > 0$ if its p.d.f. is proportional to $x^{-(\alpha+1)}$ for $\beta < x < \infty$, and is *exponential* with mean $a > 0$ if its p.d.f. is proportional to $e^{-x/\alpha}$ for $x > 0$.

Theorem 7.1. *Let $P_1, P_2, \ldots, P_n$ be compatible.*

(i) *If $\{P_i\}$ are* Bernoulli *with parameters $\{p_i\}$, respectively, then*
$\&(P_1, \ldots, P_n)$ *is* Bernoulli *with parameter* $p = \frac{\prod_{i=1}^{n} p_i}{\left(\prod_{i=1}^{n} p_i + \prod_{i=1}^{n}(1-p_i)\right)}$.

(ii) *If $\{P_i\}$ are* geometric *with parameters $\{p_i\}$, respectively,*
then $\&(P_1, \ldots, P_n)$ is geometric *with parameter* $p = 1 - \prod_{i=1}^{n}(1-p_i)$.

(iii) *If $\{P_i\}$ are* discrete uniform *on $\{1, \ldots, n_i\}$, respectively,*
then $\&(P_1, \ldots, P_n)$ is uniform on $\{1, \ldots, \min_i\{n_i\}\}$.

(iv) *If $\{P_i\}$ are* Zipf *with parameters $\{\alpha_i\}$ and $\{n_i\}$, respectively, then*
$\&(P_1, \ldots, P_n)$ *is* Zipf *with parameters $\alpha = \sum_{i=1}^{n} \alpha_i$ and $n = \min_i\{n_i\}$.*

(v) *If $\{P_i\}$ are* Zeta *with parameters $\{\alpha_i\}$, respectively, then $\&(P_1, \ldots, P_n)$ is*
Zeta *with parameter $\alpha = \sum_{i=1}^{n} \alpha_i$.*

(vi) *If $\{P_i\}$ are* gamma *with parameters $\{\alpha_i, \beta_i\}$, respectively,*
then $\&(P_1, \ldots, P_n)$ is gamma *with parameters $\alpha = \sum_{i=1}^{n} \alpha_i - (n-1)$,*
$\beta = \left(\sum_{i=1}^{n}(\beta_i)^{-1}\right)^{-1}$.

(vii) *If $\{P_i\}$ are* beta *with parameters $\{\alpha_i, \beta_i\}$, respectively, then $\&(P_1, \ldots, P_n)$*
is beta *with parameters $\alpha = \sum_{i=1}^{n} \alpha_i - (n-1)$, $\beta = \sum_{i=1}^{n} \beta_i - (n-1)$.*

(viii) *If $\{P_i\}$ are* continuous uniform *on intervals $\{(a_i, b_i)\}$, respectively, then*
$\&(P_1, \ldots, P_n)$ *is uniform on $(\max_i a_i, \min_i b_i)$.*

(ix) *If $\{P_i\}$ are* LaPlace *with parameters $\{\alpha_i\}$, respectively, then*
$\&(P_1, \ldots, P_n)$ *is* LaPlace *with parameter $\alpha = \left(\sum_{i=1}^{n}(\alpha_i)^{-1}\right)^{-1}$.*

(x) *If $\{P_i\}$ are* Pareto *with parameters $\{\alpha_i, \beta_i\}$, respectively, then*
$\&(P_1, \ldots, P_n)$ *is* Pareto *with parameters $\alpha = \sum_{i=1}^{n} \alpha_i + n - 1$ and $\beta = \max_i \beta_i$.*

(xi) *If $\{P_i\}$ are* exponential *with means $\{\alpha_i\}$, respectively, then $\&(P_1, \ldots, P_n)$*
is exponential *with mean $\alpha = \left(\sum_{i=1}^{n} \alpha_i^{-1}\right)^{-1}$.*

Proof. Conclusions (i)–(v) follow from Theorem 3.1 and routine calculations, and (vi)–(xi) follow from Theorem 3.3 and calculations. □

Note that for smaller values of the parameters of beta distributions, the conflation may not be beta simply because the product of the densities may not be integrable. The families of distributions identified in Theorem 7.1 that are closed under conflation are by no means exhaustive. For example, the conflation of n Poisson distributions is not classical Poisson, but is a discrete Conway-Maxwell-Poisson (CMP) distribution with p.m.f. proportional to $\frac{\lambda^k}{(k!)^n}$, $k = 0, 1, \ldots$, and clearly the CMP family is closed under conflation.

Recall that the conflation of Cauchy distributions is not Cauchy, as was shown in Example 3.5. It is easy to see that the families of binomial distributions and of chi-square distributions are not closed under conflation, but chi-square comes very close in the following sense: if X is a random variable with distribution $\&(P_1, \ldots, P_n)$ where $\{P_i\}$ are chi-square with $\{k_i\}$ degrees of freedom, respectively, then X/n is chi-square with $\sum_{i=1}^{n} k_i - 2n + 2$ degrees of freedom.

In practice, assumptions are often made about the form of the input distributions, such as NIST's essential assumption that underlying data is often normally distributed. But the true and estimated values for Planck's constant clearly are never negative, so the underlying distribution is certainly not truly normally distributed – more likely it is truncated normal. The additional assumption of exact normality, in addition to their use of linearizing the observational equations and then applying generalized least squares ([11, p. 481]), introduces further errors into the NIST estimates.

Using conflations, however, the problem of truncation essentially disappears – it is automatically taken into account. The reason is that another important feature of conflations is that it preserves many classes of truncated distributions, where a distribution of a certain type is called *truncated* if it is the conditional distribution of that type conditioned to be in a (finite or infinite) interval. For example, truncated normal distributions include normal distributions conditioned to be positive (that is, a.c. distributions with density function proportional to $e^{-(x-\mu)^2/2\sigma^2}$, $x > 0$ (and zero elsewhere)), as is often the case in experimental data involving estimates of many of the fundamental physical constants.

Theorem 7.2. *If $P_1, P_2, \ldots, P_n$ are compatible truncated normal (exponential, gamma, LaPlace, Pareto) distributions, then $\&(P_1, P_2, \ldots, P_n)$ is also a truncated normal (exponential, gamma, LaPlace, Pareto, respectively) distribution.*

Proof. Immediate from Theorem 3.3. □

The above example of determination of the values of the fundamental physical constants is only one among many scientific situations where consolidation of dissimilar data is problematic. Some government agencies, such as the Methods and Data Comparability Board of the National Water Quality Monitoring Council [10], have even established special programs to address this issue. Perhaps the method of *conflating* input data will provide a practical and simple, yet optimal and rigorous method to address this problem.

Acknowledgement

the editor and referee, and especially to Professor Ron Fox for his extensive input. The author is also grateful to the Free University of Amsterdam and the California Polytechnic State University in San Luis Obispo, where much of this research was accomplished.

References

[1] Aitchison, J. (1982) The statistical analysis of compositional data, *J. Royal. Statist. Soc. Series B* **44**, 139–177. MR676206 (83m:62104)

[2] Aitken, A. (1934) On least-squares and linear combinations of observations, *Proc. Royal Soc. Edinburgh* **55**, 42–48.

[3] Bracewell, R. (1999) *The Fourier Transform and its Applications*, 3rd Ed., McGraw-Hill.

[4] Chung, K.L. (2001) *A Course in Probability Theory*, 3rd Ed. Academic Press, New York. MR1796326 (2001g:60001)

[5] Egozcue, J., Diaz-Barrero, J. and Pawlowsky-Glahn, V. (2006) Hilbert space of p.d.f.s based on Aitchison geometry, *Acta Math. Sinica* **22**, 1175–1182. MR2245249 (2007k:28006)

[6] Elton, J. and Hill, T. (1987) A generalization of Lyapunov's convexity theorem to measures with atoms, *Proc. Amer. Math. Soc.* **99**, 297–304. MR870789 (88k:28014)

[7] Genest, C. and Zidek, J. (1986) Combining probability distributions: A critique and an annotated bibliography, *Statist. Sci.* **1**, 114–148. MR833278 (87f:62004)

[8] Lubinski, P. (2004) Averaging spectral shapes, *Monthly Notices of the Royal Astronomical Society* **350**, 596–608.

[9] Lyapounov, A. (1940) Sur les fonctions-vecteurs completement additives, *Bull. Acad. Sci URSS Math* **4**, 465–478. MR0004080 (2:315e)

[10] `http://acwi.gov/methods/about/background.html`

[11] Mohr, P. and Taylor, B. (2000) CODATA recommended values of the fundamental physical constants: 1998, *Rev. Mod. Physics* **72**, 351–495.

[12] Mohr, P., Taylor, B. and Newell, D. (2007) The fundamental physical constants, *Physics Today*, 52–55.

[13] Mohr, P., Taylor, B. and Newell, D. (2008) Recommended values of the fundamental physical constants: 2006, *Rev. Mod. Phys.* **80**, 633–730.

[14] Rencher, A. and Schaalje, G. (2008) *Linear Models in Statistics*, Wiley. MR2401650 (2009a:62006)

School of Mathematics, Georgia Institute of Technology, Atlanta, Georgia 30332-0160

E-mail address: `hill@math.gatech.edu`