LANL LA-UR-93-833, SFI TR-93-07-047

ESTIMATING FUNCTIONS OF PROBABILITY DISTRIBUTIONS FROM A FINITE SET OF SAMPLES

Part II: Bayes Estimators for Mutual Information, Chi-Squared, Covariance, and other Statistics.

by

David R. Wolf[1] and David H. Wolpert[2]

1 - Los Alamos National Laboratory, MS P940, Los Alamos, NM 87545.
(wolf@lanl.gov)

2 - The Santa Fe Institute, 1660 Old Pecos Trail, Suite A, Santa Fe, NM 87501.
(dhw@santafe.edu)

Abstract: This paper is the second in a series of two on the problem of estimating a function of a probability distribution from a finite set of samples of that distribution. In the first paper[1], the Bayes estimator for a function of a probability distribution was introduced, the optimal properties of the Bayes estimator were discussed, and the Bayes and frequency-counts estimators for the Shannon entropy were derived and graphically contrasted. In the current paper the analysis of the first paper is extended by the derivation of Bayes estimators for several other functions of interest in statistics and information theory. These functions are (powers of) the mutual information, chi-squared for tests of independence, variance, covariance, and average. Finding Bayes estimators for several of these functions requires extensions to the analytical techniques developed in the first paper, and these extensions form the main body of this paper. This paper extends the analysis in other ways as well, for example by enlarging the class of potential priors beyond the uniform prior assumed in the first paper. In particular, the use of the entropic and Dirichlet priors is considered.

PACS numbers: 02.50.+s, 05.20.-y

1. BACKGROUND

Consider a system with m possible states and an associated m-vector of probabilities of those states, $\mathbf{p} = (p_i)$, $1 \leq i \leq m$, $(\Sigma_{i=1}^{m} p_i = 1)$. The system is repeatedly and independently sampled according to the distribution $\mathbf{p}$. Let the total number of samples be N and denote the associated vector of counts of states by $\mathbf{n} = (n_i)$, $1 \leq i \leq m$, $(\Sigma_{i=1}^{m} n_i = N)$. By definition, $\mathbf{n}$ is multinomially distributed. In some cases in this paper the states will be indexed by two integers. For these cases $\mathbf{p}$ and $\mathbf{n}$ are matrices.

In many cases what we are interested in is not $\mathbf{p}$, but rather some function of $\mathbf{p}$, $F(\mathbf{p})$. The problem at hand is to estimate such a function $F(\mathbf{p})$ from the data $\mathbf{n}$. More precisely, the problem is to investigate the posterior density of $F(\mathbf{p})$, i.e. the probability density function (pdf)

$$P(F(\mathbf{p}) = f \mid \mathbf{n}) \equiv \int d\mathbf{p}\ \delta(F(\mathbf{p}) - f)\ P(\mathbf{p} \mid \mathbf{n}). \tag{1}$$

Usually it is difficult to analytically compute $P(F(\mathbf{p}) = f \mid \mathbf{n})$. Accordingly, here we instead compute posterior moments of $F(\mathbf{p})$, i.e., the moments of $F(\mathbf{p})$ according to the probability density in Eqn. (1).

These posterior moments do more than simply give us a characterization of $P(F(\mathbf{p}) = f \mid \mathbf{n})$. For example, in [1] we show that the estimator $G(\mathbf{n})$ that minimizes the mean-squared error from the true $F(\mathbf{p})$ is given by the first such posterior moment, the posterior average:

$$G(\mathbf{n}) = E[F(\mathbf{p}) \mid \mathbf{n}] \equiv \int d\mathbf{p}\ F(\mathbf{p})\ P(\mathbf{p} \mid \mathbf{n}) = \int df\ f\ P(f \mid \mathbf{n}). \tag{2}$$

In other words, this $G(\mathbf{n})$ minimizes

$$\int d\mathbf{p}\ P(\mathbf{p})\ \Sigma_{\mathbf{n}} P(\mathbf{n} \mid \mathbf{p}) \times (G(\mathbf{n}) - F(\mathbf{p}))^2, \tag{3}$$

where $P(\mathbf{p})$ is the so-called "prior" pdf of $\mathbf{p}$. In this sense, this $G(\mathbf{n})$ is the optimal estimator for $F(\mathbf{p})$. The estimator $E[F(\mathbf{p}) \mid \mathbf{n}]$ is known as the Bayes estimator for $F(\mathbf{p})$ with prior $P(\mathbf{p})$.

The second posterior moment is also useful. As was mentioned in [1], by using Chebyshev's

inequality, the second posterior moment can be used to bound the probability that $F(\mathbf{p})$ deviates substantially from the Bayes estimator for $F(\mathbf{p})$.

We now introduce the functions $F(\mathbf{p})$ for which we derive Bayes estimators in this paper. The functions $F(\mathbf{p})$ considered here are (the first two) powers of the mutual information, chi-squared for tests of independence, covariance, variance, and average. (In some cases the methods of this paper will allow consideration of all powers of these functions.) This choice of functions is not meant to be exhaustive. Rather it is meant to both exemplify some of the mathematical techniques involved in calculating Bayes estimators of function $F(\mathbf{p})$, and to provide some useful results. The techniques of this paper should be applicable to many other functions of interest as well.

The mutual information is defined in terms of a matrix $\mathbf{p}$ by

$$M(\mathbf{p}) = S((p_{i\cdot})) + S((p_{\cdot j})) - S(\mathbf{p}) \tag{4}$$

Here $(p_{i\cdot})$ and $(p_{\cdot j})$ are the vectors of column and row sums of $\mathbf{p} = (p_{ij})$ respectively, i.e. $p_{i\cdot} \equiv \Sigma_j p_{ij}$ and similarly for $(p_{\cdot j})$. $S(\mathbf{p})$ is the usual Shannon entropy: $S(\mathbf{p}) = -\Sigma_{ij} p_{ij} \log(p_{ij})$, while $S((p_{i\cdot})) = -\Sigma_i p_{i\cdot} \log(p_{i\cdot})$, and similarly for $S((p_{\cdot j}))$. Mutual information is a measure of the amount of information shared between two symbol streams (symbolic dynamical systems) with joint probability p_{ij} [2]. It may also be seen as a measure of the correlation between two symbolic systems with joint probability p_{ij} [3]. The mutual information function has applications in areas such as communication theory [2] (e.g. the measurement of channel capacity), pattern recognition [4], and natural languages [5], to name but a few.

The chi-squared statistic for independence is also given by a function of a matrix $\mathbf{p}$,

$$\chi^2(\mathbf{p}) \equiv \Sigma_{ij} \left((p_{ij} - p_{i\cdot} p_{\cdot j})^2 / (p_{i\cdot} p_{\cdot j}) \right), \tag{5}$$

Chi-squared is commonly used in statistical tests of independence [6], where it appears in a form with the maximum-likelihood estimator of $\mathbf{p}$ substituted for $\mathbf{p}$ in Eq. (5). The form in Eq. (5) is proportional to the asymptotic (large data set) statistic used in these tests, and it is easily shown to be a first order approximation to the mutual information under certain conditions [7-8].

The covariance function of a matrix p is given by

$$\mathrm{Cov}_{xy}(\mathbf{p}) \equiv \Sigma_{ij} p_{ij} \, (X_i - \mu_x) \, (Y_j - \mu_y) \, , \tag{6}$$

where each of the m possible states is associated with some ordered pair (X_i, Y_j) of numbers (there are m index pairs (i, j) altogether). The ij'th state occurs with probability p_{ij}. The means are μ_x and μ_y; $\mu_x \equiv \Sigma_i p_{i\cdot} X_i$ and similarly for μ_y [6].

The variance function of a vector **p** is given by

$$\mathrm{Var}(\mathbf{p}) \equiv \Sigma_i p_i \, (X_i - \mu_x)^2, \tag{7}$$

where the i'th state is associated with the number X_i and occurs with probability p_i.

Finally, the average is a function of a vector **p** given by

$$\mathrm{Avg}_x(\mathbf{p}) \equiv \Sigma_i p_i X_i . \tag{8}$$

In Sec. 2 we derive Bayes estimators with uniform prior for these functions of probability distributions. The notation used in these results is summarized in subsections 2a and 2b. The results themselves, the Bayes estimators for the various functions being considered, are summarized in subsections 2d and 2g. In Sec. 3 we show how these Bayes estimator results along with those of [1], all derived under the assumption of a uniform prior, are modified when various different priors are assumed. In particular, we discuss the entropic prior $P(\mathbf{p}) \propto e^{\alpha S(\mathbf{p})}$ and a broad class of priors which includes the Dirichlet prior, $P(\mathbf{p}) \propto \Pi_{i=1}^{m} p_i^r$.

Throughout this paper it is assumed that the reader is at least passingly familiar with the analysis in [1], the first paper in this two-paper series.

2. CALCULATIONS FOR THE BAYES ESTIMATORS

In this section we present the calculations needed to derive the first two moments of the functions discussed in Sec. 1. The subsections are organized as follows. In Sec. 2a we discuss the form

of the integrations to be done. Sec. 2b contains a presentation of notation motivated by considering the special case where $F(\mathbf{p})$ is the mutual information raised to some power. Section 2c contains intermediate results, Thms. 9-15. In Sec. 2d we use the results of Sec. 2c to derive the Bayes estimators for the first two powers of most of the functions described in Sec. 1. The results appear as Thms. 16-21. Sections 2e-g parallel Secs. 2b-d (notation, intermediate results (Thms. 22-25), and Bayes estimators (Thms. 26 and 27)), but for integrals more complicated than those considered in Secs. 2b-d.

The reader interested only in the results for the Bayes estimators for the $F(\mathbf{p})$ described in Sec. 1 should see Secs. 2d and 2g.

2a. THE FORM OF THE INTEGRALS.

Recall from Sec. 1 that the Bayes estimator for $F(\mathbf{p})$ is the posterior average $E[F(\mathbf{p})|\mathbf{n}] = \int d\mathbf{p}\ P(\mathbf{p}|\mathbf{n})\, F(\mathbf{p})$. Let $\Delta(\mathbf{p}) \equiv \delta(\Sigma_i p_i - 1)$ and $\Theta(\mathbf{p}) \equiv \Pi_i \theta(p_i)$, where $\theta(x) = 1$ for $x \geq 0$, 0 otherwise. Define $I[F(\mathbf{p}), \mathbf{n}]$ by

$$I[F(\mathbf{p}), \mathbf{n}] \equiv \int d\mathbf{p} F(\mathbf{p}) \Delta(\mathbf{p}) \Theta(\mathbf{p}) \Pi_{i=1}^{m} p_i^{n_i}. \tag{9}$$

Using this notation, when the prior $P(\mathbf{p})$ is uniform, i.e. when $P(\mathbf{p}) \propto \Delta(\mathbf{p})\Theta(\mathbf{p})$, it is easily shown (see [1]) that

$$E[F(\mathbf{p})|\mathbf{n}] = I[F(\mathbf{p}), \mathbf{n}] / I[1, \mathbf{n}]. \tag{10}$$

The result for $I[1, \mathbf{n}]$ appears in Thm. 3 of [1]. (In general, references to Thms. 1-8 are to the theorems so numbered in [1]; this paper, being a continuation of [1], starts numbering its theorems at 9.) Therefore, when the prior is uniform, finding the Bayes estimator $E[F(\mathbf{p})|\mathbf{n}]$ reduces to evaluating the integral $I[F(\mathbf{p}), \mathbf{n}]$. In the rest of this section, this integral is calculated for each of the $F(\mathbf{p})$ mentioned in Sec. 1.

2b. NOTATION.

To understand the types of calculations that must be performed, consider the case where $F(\mathbf{p}) = M^k(\mathbf{p})$, $M(\mathbf{p})$ being the mutual information. In this case the integral of interest is

$$I[M^k(\mathbf{p}), \mathbf{n}] = \int_{\Delta_{\mathbf{P}}} d\mathbf{p}\ \Pi_{i=1}^{m} p_i^{n_i} \times [-\Sigma_i p_{i\cdot} \log(p_{i\cdot}) - \Sigma_j p_{\cdot j} \log(p_{\cdot j}) + \Sigma_{ij} p_{ij} \log(p_{ij})]^k \quad (11)$$

(Here and in the rest of this paper we do not explicitly write the $\Delta(\mathbf{p})\Theta(\mathbf{p})$ factor in the integrand. Rather, it is indicated by subscripting the integrals with $\Delta_{\mathbf{P}}$.) The right-hand side of Eq. (11) expands to a sum of integrals of the form $I[\rho_1^{q_1}\log^{r_1}(\rho_1)\ldots\rho_k^{q_k}\log^{r_k}(\rho_k), \mathbf{n}]$, with $q_i = r_i$, and with each ρ_i a sum of a subset of the p_i's.

Since such sums of subsets of p_i's will often arise, we introduce some special notation. Indicate a subset of indices of the p_i's by σ, and the sum of the p_i's with indices $i \in \sigma$ by $\rho \equiv \Sigma_{i\in\sigma} p_i$. If there are k such subsets, these will be represented by σ_u, $u = 1, \ldots, k$, and the corresponding subset sums will be given by $\rho_u \equiv \Sigma_{i\in\sigma_u} p_i$, $u = 1, \ldots, k$. In the case where $\sigma_u \cap \sigma_v = \varnothing$ for $u \neq v$, the subsets will be called *non-overlapping*. If on the other hand $\sigma_u \cap \sigma_v \neq \varnothing$ and yet $\sigma_u \neq \sigma_v$, the subsets will be called *overlapping*. Any expression involving non-overlapping (overlapping) subset sums will be called a non-overlapping (overlapping) term. *Pair-wise overlapping* will be used to indicate a term involving two subsets which are overlapping (as opposed to multiple such subsets).

Since we will often be dealing with overlapping subsets, the notation σ_{uv} will be used for $\sigma_u \cap \sigma_v$. (No confusion arises since we never explicitly refer to σ_u with u a double-digit number.) Similarly, we use the notation $\sigma_{u-uv} \equiv \sigma_u - \sigma_{uv}$ to refer to those indices in σ_u but not in σ_{uv}, $\sigma_{u+v} \equiv \sigma_u \cup \sigma_v$ to refer to those indices in either σ_u and/or in σ_v, and obvious extensions when more than two sets are being considered.

Note that in Eq. (11), for $k = 1$ the integral reduces to a sum of integrals of non-overlapping terms. For $k = 2$ some of the integrands are pair-wise overlapping terms. However for $k = 2$ no term occurs that involves more than two overlapping subsets.

We will use the following conventions for various other quantities which arise. For generic numbers appearing as exponents in convolution expressions the variable α will be used. Just as we previously defined ρ's to refer to sums of those p_i picked out by index subsets σ, we also need notation for the sums of n_i's picked out by such σ's. The variable β will be used to indicate these sums, with the convention discussed above for subscripts on ρ holding for β's subscripts; $\beta_u \equiv \Sigma_{i \in \sigma_u} (n_i + 1)$. It will also sometimes be useful to use the notational convention $\nu_i \equiv n_i + 1$, $i = 1, \ldots, m$. In the case where **n** and **p** are matrices instead of vectors we use $\nu_{ij} \equiv n_{ij} + 1$. Similarly, to denote row or column sums of n_{ij}'s we use $\nu_{i\cdot} \equiv \Sigma_j \nu_{ij}$ and $\nu_{\cdot j} \equiv \Sigma_i \nu_{ij}$.

We will also have need of the following notation: $\beta \equiv \nu \equiv \Sigma_{i=1}^{m} \nu_i$; $\gamma_{\mathbf{n}} \equiv \Pi_{i=1}^{m} \Gamma(\nu_i)$; $\eta \equiv \Sigma_{u=1}^{k} \eta_u$ (η_u is associated with subset σ_u, $u = 1, \ldots, k$); $\Phi^{(n)}(z) \equiv \Psi^{(n-1)}(z)$, where $\Psi^{(n)}(z)$ is given by $\Psi^{(n)}(z) = \partial_z^{n+1} \log(\Gamma(z))$ (see [9], Eq. 6.3.1); and the "delta-phi" function is given by $\Delta\Phi^{(n)}(z_1, z_2) \equiv \Phi^{(n)}(z_1) - \Phi^{(n)}(z_2)$.

Notation for the hypergeometric functions used here appears in App. A. In particular, the special use of subscripts on parentheses (e.g., $(a)_b$) is defined there, as are functions of the form ${}_pF_q$ and functions of the form ${}_{p_1, p_2, p_{12}}F_{q_1, q_2, q_{12}}$.

As it stands, the integrand in Eq. (11) is not "factorable" as defined in Sec. 4b of [1]. Having an integrand in factorable form is desirable because it allows us to apply the procedure of Sec. 4b of [1] to evaluate the integral. In what follows we utilize the T transform (see App. B.1) to convert the integrand of Eq. (11) (and similar integrands to appear) into factorable form. Once the integrand is factorable, we can perform the integration. After the integration we inverse T-transform to arrive at the final result for the integral of interest.

2c. NON-OVERLAP AND PAIR-WISE OVERLAP CONVOLUTION INTEGRALS.

In this subsection we derive a number of intermediate results. We begin this subsection with the fundamental convolution integrals needed to derive non-overlap and pair-wise overlap convolutions (Thms. 9-11). We then present specific integrals $I[\cdot,\cdot]$ of these two types (Thms. 12-15). Theorems 12 and 13a,b apply to non-overlap terms; Theorems 14a,b and 15a,b apply to pair-wise overlap terms. In Sec. 2d we use these intermediate results to evaluate Bayes estimators for several of the functions $F(\mathbf{p})$ being considered.

Define the Laplace convolution operator "$\otimes$" acting on functions f and g by

$$(f \otimes g)(\tau) \equiv \int_0^\tau dx\ f(x)\, g(\tau - x)\,. \tag{12}$$

In the following, the convolution operator is (usually) implicitly assumed to refer to functions of the variable p. As described in [1], this convolution operator is fundamental to evaluating $I[\cdot,\cdot]$ integrals.

Theorem 9. If $\mathrm{Re}(\alpha_i) > 0$, $i = 1, 2$. Then

$$(9.1)\quad (p^{\alpha_1 - 1} \otimes p^{\alpha_2 - 1})(\tau) = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1 + \alpha_2)} \times \tau^{\alpha_1 + \alpha_2 - 1},$$

$$(9.2)\quad (p^{\alpha_1 - 1} e^{-pt} \otimes p^{\alpha_2 - 1} e^{-pt})(\tau) = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1 + \alpha_2)} \times \tau^{\alpha_1 + \alpha_2 - 1} \times e^{-\tau t}.$$

Proof: To Prove (9.1) note that by Thm. 2 (the Laplace convolution theorem) and the fact that $L[p^{\alpha - 1}](s) = \dfrac{\Gamma(\alpha)}{s^\alpha}$ for $\mathrm{Re}(\alpha) > 0$ we have

$$(p^{\alpha_1 - 1} \otimes p^{\alpha_2 - 1})(\tau) = L^{-1}[L[p^{\alpha_1 - 1}]\, L[p^{\alpha_2 - 1}]]\,(\tau)$$

$$= L^{-1}\left[\frac{\Gamma(\alpha_1)}{s^{\alpha_1}}\frac{\Gamma(\alpha_2)}{s^{\alpha_2}}\right](\tau) = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1+\alpha_2)}\times\tau^{\alpha_1+\alpha_2-1}.$$

To prove (9.2) note that $L[p^{\alpha-1}e^{-pt}](s) = \frac{\Gamma(\alpha)}{(s+t)^{\alpha}}$. The remainder of the proof parallels the proof of (9.1) above: Substitute $L[p^{\alpha-1}e^{-pt}]$ for $L[p^{\alpha}]$. QED.

If desired, Thm. 3 may be thought of as corollary to Thm. 9.1, by induction.

In order to discuss the next results succinctly, we need to use the confluent hypergeometric function, discussed in App. A.

Theorem 10. If $\mathrm{Re}(\alpha_1) > 0$, $\mathrm{Re}(\alpha_2) > 0$, and $\bar{\alpha} = \alpha_1 + \alpha_2$ then

(10.1)

$$(p^{\alpha_1-1}e^{-pt_1}\otimes p^{\alpha_2-1}e^{-p(t_1+t_2)})(\tau) = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\bar{\alpha})}\times\tau^{\bar{\alpha}-1}\times e^{-\tau(t_1+t_2)}$$

$$\times {}_1F_1(\alpha_1;\ \bar{\alpha};\ t_2\tau)$$

(10.2)

$$(p^{\alpha_1-1}\otimes p^{\alpha_2-1}e^{-pt})(\tau) = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\bar{\alpha})}\times\tau^{\bar{\alpha}-1}\times e^{-\tau t}\times {}_1F_1(\alpha_1;\ \bar{\alpha};\ t\tau)$$

Proof: To prove (10.1) write the convolution $(p^{\alpha_1-1}e^{-pt_1}\otimes p^{\alpha_2-1}e^{-p(t_1+t_2)})(\tau)$ in its integral form as $e^{-\tau(t_1+t_2)}\int_0^{\tau} e^{pt_2}\, p^{\alpha_1-1}(\tau-p)^{\alpha_2-1}dp$. Make the change of variables $\tau u = p$, which introduces the factor $\tau^{\bar{\alpha}-1}$. Substitute $\alpha = \alpha_1$, $\beta = \bar{\alpha}$, $z = t_2\tau$ and compare with the form in App. A to find the result. The factor $\frac{\Gamma(\beta)}{\Gamma(\alpha)\Gamma(\beta-\alpha)}$ in the hypergeometric function in

App. A cancels the factor $\frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\bar{\alpha})}$. To prove (10.2) substitute zero for t_1 in result (10.1).

QED.

Theorem 11 gives the primary equality needed to derive pair-wise overlap results. See App. A for the definition of the generalized hypergeometric function ${}_{1,1,0}F_{0,0,1}$.

Theorem 11. If $\mathrm{Re}(\alpha_1) > 0$, $\mathrm{Re}(\alpha_{12}) > 0$, $\mathrm{Re}(\alpha_2) > 0$ and $\alpha \equiv \alpha_1 + \alpha_{12} + \alpha_2$, then

$$(p^{\alpha_1 - 1} e^{-pt_1} \otimes p^{\alpha_{12} - 1} e^{-p(t_1 + t_2)} \otimes p^{\alpha_2 - 1} e^{-pt_2})(\tau) =$$

$$\frac{\Gamma(\alpha_1)\Gamma(\alpha_{12})\Gamma(\alpha_2)}{\Gamma(\alpha)} \times \tau^{\alpha - 1} \times e^{-\tau(t_1 + t_2)} \times {}_{1,1,0}F_{0,0,1}[\alpha_1, \alpha_2;\ \alpha;\ t_2\tau, t_1\tau]$$

Proof: The convolution $p^{\alpha_1 - 1} e^{-pt_1} \otimes p^{\alpha_{12} - 1} e^{-p(t_1 + t_2)}$ was done in Thm. 10.1. Expand ${}_1F_1$ in that result in its series representation (see App. A)

$${}_1F_1(\alpha_1;\ \alpha_1 + \alpha_{12};\ t_2\tau) = \Sigma_{k=0}^{\infty} \frac{(\alpha_1)_k}{(\alpha_1 + \alpha_{12})_k} \frac{(t_2\tau)^k}{\Gamma(k+1)}$$

and convolve it with $p^{\alpha_2 - 1} e^{-pt_2}$ term-by-term (apply Thm. 10.2) to find the desired result. (This is valid since the series is uniformly convergent on $[0, \tau]$.) QED.

Thms. 9-11 are now used to derive integrals $I[\cdot,\cdot]$ for some non-overlap and pair-wise overlap terms. Theorems 12 and 13 discuss the non-overlap case, while Thms. 14 and 15 discuss the pair-wise overlap case.

Theorem 12. If the subsets σ_u, defined for $u = 1, \ldots, k$, satisfy $\sigma_{uv} = \varnothing$ for all $u \neq v$, and if $\mathrm{Re}(\beta_u + \eta_u) > 0$ for all $u = 1, \ldots, k$, and if $\mathrm{Re}(\nu_i) > 0$ for all $i = 1, \ldots, m$ then

$$I[\rho_1^{\eta_1}\ldots\rho_k^{\eta_k}, \mathbf{n}] = \frac{\gamma_{\mathbf{n}}}{\Gamma(\beta+\eta)}\Pi_{u=1}^{k}\frac{\Gamma(\beta_u+\eta_u)}{\Gamma(\beta_u)}.$$

Proof: Assume $k = 1$ and $\mathrm{Re}(\eta_1) < 0$ to begin. Apply $T^{-1}T$ with respect to η_1 to the integral $I[\rho_1^{\eta_1}, \mathbf{n}]$ (see App. B.1 for the definition of the T transform) and evaluate the inner T transform. Noting $e^{-\rho_1 t} = \Pi_{i\in\sigma_1} e^{-p_i t}$ and that $T[z^{\eta}](t) = e^{-zt}$ for $\eta_1 < 0$, find

$$I[\rho_1^{\eta_1}, \mathbf{n}] = T^{-1}\left[\int_{\Delta_{\mathbf{p}}} d\mathbf{p}\,(\Pi_{i\in\sigma_1} p_i^{n_i} e^{-p_i t}) \times (\Pi_{i\notin\sigma_1} p_i^{n_i})\right].$$

(See App. C.1 for the justification of the interchange the integral over $\mathbf{p}$ and the T transform.) Now, write the transformed integral above as the convolution

$$I[\rho_1^{\eta_1}, \mathbf{n}] = T^{-1}((\otimes_{i\in\sigma_1} p_i^{n_i} e^{-p_i t}) \otimes (\otimes_{i\notin\sigma_1} p_i^{n_i}))(1).$$

Use Thm. 9.1 and induction to find (with $\bar{\beta} \equiv \beta - \beta_1$)

$$(\otimes_{i\notin\sigma_1} p_i^{n_i})(\tau) = \frac{\Pi_{i\notin\sigma_1}\Gamma(\nu_i)}{\Gamma(\bar{\beta})} \times \tau^{\bar{\beta}-1}.$$

Similarly, use Thm. 9.2 and induction to find

$$(\otimes_{i\in\sigma_1} p_i^{n_i} e^{-p_i t})(\tau) = \frac{\Pi_{i\in\sigma_1}\Gamma(\nu_i)}{\Gamma(\beta_1)} \times \tau^{\beta_1-1} \times e^{-\tau t}.$$

Substituting the last two expressions into $I[\rho_1^{\eta_1}, \mathbf{n}]$ yields

$$I[\rho_1^{\eta_1}, \mathbf{n}] = \frac{\gamma_{\mathbf{n}}}{\Gamma(\beta_1)\Gamma(\bar{\beta})} T^{-1}(\tau^{\beta_1-1} e^{-\tau t} \otimes \tau^{\bar{\beta}-1})(1).$$

The T^{-1} transform may now be taken to find (see Apps. A, C)

$$I[\rho_1^{\eta_1}, \mathbf{n}] = \frac{\gamma_{\mathbf{n}}}{\Gamma(\beta_1)\Gamma(\bar{\beta})}(\tau^{\beta_1+\eta_1-1} \otimes \tau^{\bar{\beta}-1})(1).$$

Now apply Thm. 9.1 in this expression to find for $\eta_1 < 0$

$$I[\rho_1^{\eta_1}, \mathbf{n}] = \frac{\gamma_{\mathbf{n}}\Gamma(\beta_1 + \eta_1)}{\Gamma(\beta_1)\Gamma(\beta + \eta)}.$$

Refer to App. D for the continuation to $\mathrm{Re}(\eta_1) \geq 0$. Refer to App. E for the existence conditions. Now, for $k > 1$ apply the identity operator $T^{-1}T$ k times (with respect to $\eta_1 \ldots \eta_k$ respectively) and evaluate only the T transforms initially. Since $\sigma_{uv} = \varnothing$ for $u \neq v$, the convolution form of the transformed $I[\rho_1^{\eta_1}, \mathbf{n}]$ becomes

$$I[\rho^{\eta_1} \ldots \rho^{\eta_k}, \mathbf{n}] = T_1^{-1} \ldots T_k^{-1}((\otimes_{u=1}^{k} (\otimes_{i \in \sigma_u} p_i^{n_i} e^{-p_i t_u})) \otimes (\otimes_{i \notin \cup_{u=1}^{k} \sigma_u} p_i^{n_i})) (1),$$

while $\bar{\beta}$ becomes $\bar{\beta} \equiv \beta - \beta_{1 + \ldots + k}$. Extend the application of Thm. 9.2 to the k convolution products $\otimes_{i \in \sigma_u} p_i^{n_i} e^{-p_i t}$ for $u = 1, \ldots, k$. Do the substitutions and take the inverse transforms to find the result. QED.

Footnote 1 contains a derivation of an interesting identity based on an alternate form of the result in Thm. 12. Since observations being integer counts in practice, we'll be interested in non-negative integer n_i; in this regard, Thm. 12 is more general than we need. (However, below we will want to differentiate with respect to η's, to get logarithms into the integrand. So a result only applicable to integer η's would not suffice.)

Theorem 13 applies Thm. 12 to find non-overlap results needed specifically for the expression of Bayes estimators for the first two moments of the entropy, mutual information, and various other functions.

Theorem 13. If the subsets σ_u, defined for $u = 1, \ldots, k$, satisfy $\sigma_{uv} = \varnothing$ for all $u \neq v$, and if $\mathrm{Re}(\beta_u + \eta_u) > 0$ for all $u = 1, \ldots, k$, and if $\mathrm{Re}(\nu_i) > 0$ for all $i = 1, \ldots, m$ then the following hold.

(13.1) One logarithm of a subset sum.

$$I\left[\left(\rho_1^{\eta_1}\ldots\rho_k^{\eta_k}\right)\times\log\left(\rho_u\right),\mathbf{n}\right] = \frac{\gamma_{\mathbf{n}}}{\Gamma(\beta+\eta)}\times\Pi_{w=1}^{k}\frac{\Gamma(\beta_w+\eta_w)}{\Gamma(\beta_w)}\times\Delta\Phi^{(1)}(\beta_u+\eta_u,\beta+\eta).$$

(13.2) Two logarithms of subset sums, $u \neq v$.

$$I\left[\left(\rho_1^{\eta_1}\ldots\rho_k^{\eta_k}\right)\times\log\left(\rho_u\right)\times\log\left(\rho_v\right),\mathbf{n}\right] = \frac{\gamma_{\mathbf{n}}}{\Gamma(\beta+\eta)}\times\Pi_{w=1}^{k}\frac{\Gamma(\beta_w+\eta_w)}{\Gamma(\beta_w)}$$

$$\times\left\{\Delta\Phi^{(1)}(\beta_u+\eta_u,\beta+\eta)\Delta\Phi^{(1)}(\beta_v+\eta_v,\beta+\eta)-\Phi^{(2)}(\beta+\eta)\right\}.$$

(13.3) Squared logarithm of a subset sum.

$$I\left[\left(\rho_1^{\eta_1}\ldots\rho_k^{\eta_k}\right)\times\log\left(\rho_u\right)^2,\mathbf{n}\right] = \frac{\gamma_{\mathbf{n}}}{\Gamma(\beta+\eta)}\times\Pi_{w=1}^{k}\frac{\Gamma(\beta_w+\eta_w)}{\Gamma(\beta_w)}$$

$$\times\left\{\Delta\Phi^{(1)}(\beta_u+\eta_u,\beta+\eta)^2+\Delta\Phi^{(2)}(\beta_u+\eta_u,\beta+\eta)\right\}.$$

Proof: The proof is done for result (13.1); results (13.2) and (13.3) follow in a similar manner. Differentiate both sides of the formula for $I\left[\rho^{\eta_1}\ldots\rho^{\eta_k},\mathbf{n}\right]$ given in Thm. 12 with respect to η_u using the fact that $\partial_\eta\rho^\eta = \rho^\eta\log(\rho)$. (See App. C.2 for justification of the interchange the integral and derivative.) Doing this gives the desired result. QED.

Theorems 12 and 13 dealt with non-overlap sums. Theorems 14 and 15 below discuss pair-wise overlap sums. In Thm. 14a the non-contained overlap case is discussed. In Thm. 14b the contained overlap case is discussed. See App. A for the definition of the hypergeometric function ${}_{2,2,0}F_{0,0,1}$.

Theorem 14a. If σ_1 and σ_2 satisfy $\sigma_{12}\neq\varnothing$, $\sigma_1\neq\sigma_{12}\neq\sigma_2$, $\mathrm{Re}\left(\beta_1+\eta_1\right)>0$, $\mathrm{Re}\left(\beta_2+\eta_2\right)>0$, and $\mathrm{Re}\left(\nu_i\right)>0, i = 1, \ldots, m$, then

$$I\left[\rho_1^{\eta_1}\rho_2^{\eta_2},\mathbf{n}\right] = \frac{\gamma_{\mathbf{n}}}{\Gamma(\beta+\eta)}\times\frac{\Gamma(\beta_{1+2}+\eta)}{\Gamma(\beta_{1+2})}$$

$$\times{}_{2,2,0}F_{0,0,1}\left[\left(\beta_{1-12},-\eta_2\right),\left(\beta_{2-12},-\eta_1\right);\beta_{1+2};1,1\right].$$

Proof: To begin, assume that $\mathrm{Re}(\eta_i) < 0$, $i = 1, 2$ and that the η_i are not integers. Apply $T_1^{-1}T_2^{-1}T_1T_2$ (T_i is with respect to η_i, see App. B.1) to the integral $I[\rho_1^{\eta_1}\rho_2^{\eta_2}, \mathbf{n}]$. Evaluating the (non-inverse) T transforms yields the factorable form

$$I[\rho_1^{\eta_1}\rho_2^{\eta_2}, \mathbf{n}] = T_1^{-1}T_2^{-1}\int_{\Delta_{\mathbf{P}}} d\mathbf{p}\,(\Pi_{i\in(\sigma_{1-12})} p_i^{n_i}e^{-p_i t_1}) \times (\Pi_{i\in\sigma_{12}} p_i^{n_i}e^{-p_i(t_1+t_2)}) \times (\Pi_{i\in(\sigma_{2-12})} p_i^{n_i}e^{-p_i t_2}) \times (\Pi_{i\notin\sigma_{1+2}} p_i^{n_i}).$$

(See App. C.1 for justification of the interchange of the integral over $\mathbf{p}$ and the T transform.) Now, write the transformed integral as the convolution (see Thm. 1)

$$I[\rho_1^{\eta_1}\rho_2^{\eta_2}, \mathbf{n}] = T_1^{-1}T_2^{-1}((\otimes_{i\in\sigma_{1-12}} p_i^{n_i}e^{-p_i t_1})\otimes(\otimes_{i\in\sigma_{12}} p_i^{n_i}e^{-p_i(t_1+t_2)}) \otimes(\otimes_{i\in\sigma_{2-12}} p_i^{n_i}e^{-p_i t_2})\otimes(\otimes_{i\notin\sigma_{1+2}} p_i^{n_i}))(1).$$

Apply Thm. 9.1 and induction to find (where $\bar{\beta} \equiv \beta - \beta_{1+2}$)

$$(\otimes_{i\notin\sigma_{1+2}} p_i^{n_i})(\tau) = \frac{\Pi_{i\notin\sigma_{1+2}}\Gamma(\nu_i)}{\Gamma(\bar{\beta})} \times \tau^{\bar{\beta}-1}.$$

Similarly, use Thm. 9.2 and induction to find

$$((\otimes_{i\in\sigma_{1-12}} p_i^{n_i}e^{-p_i t_1}) \otimes (\otimes_{i\in\sigma_{12}} p_i^{n_i}e^{-p_i(t_1+t_2)}) \otimes (\otimes_{i\in\sigma_{2-12}} p_i^{n_i}e^{-p_i t_2}))(\tau) =$$

$$\frac{\Pi_{i\in\sigma_{1+2}}\Gamma(\nu_i)}{\Gamma(\beta_{1-12})\Gamma(\beta_{12})\Gamma(\beta_{2-12})} (p^{\beta_{1-12}-1}e^{-pt_1}\otimes p^{\beta_{12}-1}e^{-p(t_1+t_2)}\otimes p^{\beta_{2-12}-1}e^{-pt_2})(\tau).$$

Substitute the result for Thm. 11 into the triple convolution above, and substitute the last two expressions into the convolution form of the transformed integral to find

$$I[\rho_1^{\eta_1}\rho_2^{\eta_2}, \mathbf{n}] = T_1^{-1}T_2^{-2}\Big[\ \frac{\gamma_{\mathbf{n}}}{\Gamma(\beta_{1+2})\Gamma(\bar{\beta})}$$

$$\times\ ((p^{\beta_{1+2}-1}\times e^{-p(t_1+t_2)}\times {}_{1,1,0}F_{0,0,1}[\beta_{1-12}, \beta_{2-12}; \beta_{1+2}; t_2 p, t_1 p])\otimes p^{\bar{\beta}-1})\ \Big](1).$$

Now, take inverse T transforms and apply Thm. 9.1 to find the desired result. Refer to App. E to

determine the conditions for the existence of the identity. Refer to App. D for the continuation of the result to $\mathrm{Re}\,(\eta_i) \geq 0$. Finally, for values of $\eta \geq 0$, refer to App. F. QED.

See footnote 2 for a derivation of two interesting identities resulting from alternate forms of this proof. When $\sigma_2 \subset \sigma_1$ the above result simplifies as in Thm. 14b below.

Theorem 14b. If σ_1, σ_2 satisfy $\sigma_{12} \neq \varnothing$, $\sigma_2 \subset \sigma_1$, $\mathrm{Re}\,(\beta_1 + \eta) > 0$, $\mathrm{Re}\,(\beta_{12} + \eta_2) > 0$, and $\mathrm{Re}\,(\nu_i) > 0, i = 1, \ldots, m$, then

$$(14b.1)\quad I\,[\rho_1^{\eta_1}\rho_2^{\eta_2}, \mathbf{n}] = \frac{\gamma_{\mathbf{n}}}{\Gamma(\beta + \eta)} \times \frac{\Gamma(\beta_1 + \eta)}{\Gamma(\beta_1)} \times {}_2F_1\,[\,(\beta_{1-12}, -\eta_2)\,;\beta_1\,;1\,]\,,$$

$$(14b.2)\quad = \frac{\gamma_{\mathbf{n}}}{\Gamma(\beta_{12})} \times \frac{\Gamma(\beta_{12} + \eta_2)}{\Gamma(\beta_1 + \eta_2)} \times \frac{\Gamma(\beta_1 + \eta)}{\Gamma(\beta + \eta)}\,.$$

Proof: Similar to proof of Thm. 14a, but apply Thm. 10.1 instead of Thm. 11. The second form (14b.2) of the result is derived by applying Gauss's identity (see Footnote 1) to the first form of the result above. QED.

Theorems 15a and 15b build upon the results of Thms. 14a and 14b respectively and state results needed to express specific terms of the various Bayes estimators. Theorem 15a contains results for the non-contained overlap case. ("Non-contained" means that neither subset of indices is properly contained within the other.) Thm. 15b contains results for the contained overlap case. Since we are most directly interested in non-negative integer η's and because simplification occurs at those η's, Thm. 15a is stated only for non-negative integer η's.

Theorem 15a. If $\eta_1 \geq 0$ and $\eta_2 \geq 0$ are integers and the conditions for Thm. 14a hold, then

$$(15a.1)\quad I\,[\rho_1^{\eta_1}\rho_2^{\eta_2}, \mathbf{n}] = C^{(0)}\,F^{(00)}\,,$$

(15a.2) $$I[\rho_1^{\eta_1}\log(\rho_1)\,\rho_2^{\eta_2}, \mathbf{n}] = C^{(1)}F^{(00)} + C^{(0)}F^{(10)},$$

(15a.3) $$I[\rho_1^{\eta_1}\log(\rho_1)\,\rho_2^{\eta_2}\log(\rho_2), \mathbf{n}] =$$

$$C^{(2)}F^{(00)} + C^{(1)}[F^{(10)} + F^{(01)}] + C^{(0)}F^{(11)},$$

(15a.4) $$I[\rho_1^{\eta_1}\log^2(\rho_1)\,\rho_2^{\eta_2}, \mathbf{n}] = C^{(2)}F^{(00)} + 2C^{(1)}F^{(10)} + C^{(0)}F^{(20)},$$

where:

$$C^{(0)} \equiv \frac{\gamma_{\mathbf{n}}}{\Gamma(\beta_{1+2})} \times \frac{\Gamma(\beta_{1+2}+\eta)}{\Gamma(\beta+\eta)},\ C^{(1)} \equiv C^{(0)} \times \Delta\Phi^{(1)}(\beta_{1+2}+\eta, \beta+\eta),$$

$$C^{(2)} \equiv C^{(0)} \times \{\Delta\Phi^{(1)}(\beta_{1+2}+\eta, \beta+\eta)^2 + \Delta\Phi^{(2)}(\beta_{1+2}+\eta, \beta+\eta)\},$$

and

(a) $$F^{(00)} \equiv \eta_1!\eta_2!\Sigma_{i=0}^{\eta_2}\Sigma_{j=0}^{\eta_1}\frac{(\beta_{1-12})_i\,(\beta_{2-12})_j}{(\beta_{1+2})_{i+j}}\frac{1}{(\eta_1-j)!\,(\eta_2-i)!}\frac{(-1)^{i+j}}{i!j!},$$

(b) $$F^{(10)} \equiv \eta_1!\eta_2!\Sigma_{i=0}^{\eta_2}\Sigma_{j=0}^{\infty}\frac{(\beta_{1-12})_i\,(\beta_{2-12})_j}{(\beta_{1+2})_{i+j}}\frac{1}{(\eta_2-i)!}\frac{(-1)^{i}}{i!j!}Q_1(j, \eta_1),$$

with Q_1 given by

$$Q_1(j, \eta_1) \equiv (1-\theta(j-\eta_1-1))\frac{(-1)^j}{(\eta_1-j)!}\Sigma_{r=0}^{j-1}\frac{1}{\eta_1-r}$$

$$+\theta(j-\eta_1-1)(-1)^{\eta_1+1}\Gamma(j-\eta_1).$$

(c) $F^{(01)}$ is the same as (b) with $i \leftrightarrow j$ and $\eta_1 \leftrightarrow \eta_2$.

(d) $$F^{(11)} \equiv \eta_1!\eta_2!\Sigma_{i=0}^{\infty}\Sigma_{j=0}^{\infty}\frac{(\beta_{1-12})_i\,(\beta_{2-12})_j}{(\beta_{1+2})_{i+j}}\frac{1}{i!j!}Q_1(j, \eta_1)Q_1(i, \eta_2).$$

(e) $$F^{(20)} \equiv \eta_1!\eta_2!\Sigma_{i=0}^{\eta_2}\Sigma_{j=0}^{\infty}\frac{(\beta_{1-12})_i\,(\beta_{2-12})_j}{(\beta_{1+2})_{i+j}}\frac{1}{(\eta_2-i)!}\frac{(-1)^{i}}{i!j!}Q_2(j, \eta_1),$$

with Q_2 given by

$$Q_2(j, \eta_1) \equiv \{ \ (1 - \theta(j - \eta_1 - 1)) \ \frac{(-1)^j}{(\eta_1 - j)!} \Sigma^{j-1}_{r,\, s = 0,\, r \neq s} \frac{1}{(\eta_1 - r) \ (\eta_1 - s)}$$

$$+ \theta(j - \eta_1 - 1) \ (-1)^{\eta_1 + 1} 2\Gamma(j - \eta_1) \Sigma^{j-1}_{r = 0,\, r \neq \eta_1} \frac{1}{\eta_1 - r} \} \ .$$

Proof: The proof is done for (15a.2). The other cases have similar proofs. Differentiate both sides of the expression for $I[\rho_1^{\eta_1} \rho_2^{\eta_2}, \mathbf{n}]$ given in Thm. 14a with respect to η_1. After differentiating, the left-hand side is given by $I\left[\rho_1^{\eta_1} \log(\rho_1) \, \rho_2^{\eta_2}, \mathbf{n}\right]$. (The justification of the interchange of the integral and derivative is given in App. C.1.) Write the differentiated right-hand side as

$$\partial_{\eta_1}(C^{(0)} \times {}_{2,2,0}F_{0,0,1}[(\beta_{1-12}, -\eta_2), (\beta_{2-12}, -\eta_1); \beta_{1+2}; 1, 1]).$$

This expands to $\partial_{\eta_1}(C^{(0)})_{2,2,0}F_{0,0,1}[\ldots] + C^{(0)} \partial_{\eta_1}({}_{2,2,0}F_{0,0,1}[\ldots])$. The derivative of $C^{(0)}$ is given by $\partial_{\eta_1} C^{(0)} = C^{(0)} \times \Delta\Phi^{(1)}(\beta_{1+2} + \eta, \beta + \eta) = C^{(1)}$. The undifferentiated hypergeometric is evaluated at η_1 and η_2 using the results in App. F, cases 1 and 2. This evaluates to $F^{(00)}$ defined in (a) above. The derivative of the hypergeometric may be taken term-by-term (this is justified below). Use the results in App. F, cases 1 and 2, Eqs. (F.4) and (F.6), to evaluate this derivative at η_1 and η_2. Doing this gives the expression $F^{(10)}$ defined in (b). With these derivatives and evaluations, the result (15a.2) follows immediately. Now consider the validity of term-by-term differentiation of the hypergeometric. There exists a closed neighborhood N containing the integer η_1 with $\mathrm{Re}(\beta_1 + x) > 0 \ \forall x \in N$. The results of App. F show that any truncation (in j) of the series for ${}_{2,2,0}F_{0,0,1}[(\beta_{1-12}, -\eta_2), (\beta_{2-12}, -x); \beta_{1+2}; 1, 1]$ (see App. A) may be differentiated with respect to x on N. The sequence of derivatives of the increasing order truncations converges uniformly on N. (To see this, note that $S_i(x) \equiv \Sigma^{\infty}_{j = \eta_1 + 1} \frac{\Gamma(\beta_{2-12} + j)}{\Gamma(\beta_{1+2} + i + j)} \frac{\Gamma(-x + j)}{j!}$ is convergent for

each i, and $\mathrm{Re}(\beta_1 + x) > 0$. Now, note that $S_i(x)$ is a series of terms each monotonic on N with the same monotonicity in x holding for each term, and that the summation over i in (b) is finite. These observations and the convergence just established demonstrate the claim of uniform convergence.) Finally, by Thm. 7.17 of [10], the sequence of derivatives of the increasing order truncations converges to the derivative of the limit of the series on N, justifying the term-by-term differentiation of the infinite series. QED.

See footnote 3 for some comments regarding alternate forms for the results given above in Thm. 15a.

Theorem 15b builds on Thm. 14b and states the results for the case in which there are two subset sums, with the indices of one subset completely contained in the other. Here, unlike in Thm. 15a, there is no hypergeometric function to consider, so the presentation of these results is much shorter. Further, unlike Thm. 15a, the expressions given are valid for all η's in the range specified (not just at nonnegative integers as in Thm. 15a) because there are no poles in the expressions being considered at the integers and therefore no further simplification occurs at these points.

Theorem 15b. If the conditions for Thm. 14b hold, then

(15b.1) $$I\left[\rho_1^{\eta_1}\log(\rho_1)\rho_2^{\eta_2}, \mathbf{n}\right] = C^{(00)} \times \Delta\Phi^{(1)}(\beta_1 + \eta, \beta + \eta),$$

(15b.2)

$$I\left[\rho_1^{\eta_1}\rho_2^{\eta_2}\log(\rho_2), \mathbf{n}\right] = C^{(00)} \times \left\{\Delta\Phi^{(1)}(\beta_{12} + \eta_2, \beta_1 + \eta_2) + \Delta\Phi^{(1)}(\beta_1 + \eta, \beta + \eta)\right\},$$

(15b.3)

$$I\left[\rho_1^{\eta_1}\log(\rho_1)\rho_2^{\eta_2}\log(\rho_2), \mathbf{n}\right] = C^{(00)} \times \left\{\Delta\Phi^{(1)}(\beta_1 + \eta, \beta + \eta)^2 + \Delta\Phi^{(1)}(\beta_{12} + \eta_2, \beta_1 + \eta_2)\Delta\Phi^{(1)}(\beta_1 + \eta, \beta + \eta) + \Delta\Phi^{(2)}(\beta_1 + \eta, \beta + \eta)\right\},$$

(15b.4)

$$I\left[\rho_1^{\eta_1}\log^2(\rho_1)\rho_2^{\eta_2}, \mathbf{n}\right] = C^{(00)} \times \left\{\Delta\Phi^{(1)}(\beta_1 + \eta, \beta + \eta)^2 + \Delta\Phi^{(2)}(\beta_1 + \eta, \beta + \eta)\right\}$$

(15b.5)

$$I[\rho_1^{\eta_1}\rho_2^{\eta_2}\log^2(\rho_2), \mathbf{n}] = C^{(00)} \times \{ (\Delta\Phi^{(1)}(\beta_{12}+\eta_2, \beta_1+\eta_2) + \Delta\Phi^{(1)}(\beta_1+\eta, \beta+\eta))^2 + \Delta\Phi^{(2)}(\beta_{12}+\eta_2, \beta_1+\eta_2) + \Delta\Phi^{(2)}(\beta_1+\eta, \beta+\eta) \},$$

where $C^{(00)} = \frac{\gamma_{\mathbf{n}}}{\Gamma(\beta_{12})} \times \frac{\Gamma(\beta_{12}+\eta_2)}{\Gamma(\beta_1+\eta_2)} \times \frac{\Gamma(\beta_1+\eta)}{\Gamma(\beta+\eta)}$.

Proof: The proof is done for (15b.2). The proofs of the other results follow in a similar manner. The result of Thm. 14b is

$$I[\rho_1^{\eta_1}\rho_2^{\eta_2}, \mathbf{n}] = C^{(00)} = \frac{\gamma_{\mathbf{n}}}{\Gamma(\beta_{12})} \times \frac{\Gamma(\beta_{12}+\eta_2)}{\Gamma(\beta_1+\eta_2)} \times \frac{\Gamma(\beta_1+\eta)}{\Gamma(\beta+\eta)}.$$

Differentiate both sides of this with respect to η_2. The left-hand side of the differentiated expression is $I[\rho_1^{\eta_1}\rho_2^{\eta_2}\log(\rho_2), \mathbf{n}]$. (The justification of the interchange of the integral and derivative is given in App. C.2.) The derivative of $\frac{\Gamma(\beta_{12}+\eta_2)}{\Gamma(\beta_1+\eta_2)}$ is given by $\Delta\Phi^{(1)}(\beta_{12}+\eta_2, \bar{\beta}+\eta_2)$. The derivative of $\frac{\Gamma(\beta_1+\eta)}{\Gamma(\beta+\eta)}$ is given by $\Delta\Phi^{(1)}(\beta_1+\eta, \beta+\eta)$. Substituting these expressions for the appropriate derivatives in the overall derivative of the right-hand side of the equality above for $C^{(00)}$ gives the claimed result. QED.

2d. BAYES ESTIMATORS FOR NON-OVERLAP AND PAIR-WISE OVERLAP TERMS

In this subsection we present those Bayes estimators for the functions of Sec. 1 that can be expressed with non-overlapping and pair-wise overlapping terms. These include the first and second posterior moments of the entropy, mutual information, average, and variance, and the first posterior moments of chi-squared and covariance. The second posterior moments of chi-squared and covariance appear in Sec. 2g. All of the results presented in this section assume a uniform prior.

1. Entropy, $S(\mathbf{p}) = -\Sigma_i p_i \log(p_i)$. This result appears in [1] in a different form and is stated here for completeness.

Theorem 16. If $\mathrm{Re}(\nu_i) > 0$, $i = 1, \ldots, m$ then

(16.1) $$E[S(\mathbf{p})|\mathbf{n}] = -\Sigma_i \frac{\nu_i}{\nu} \Delta\Phi^{(1)}(\nu_i + 1, \nu + 1),$$

(16.2) $$E[S^2(\mathbf{p})|\mathbf{n}] =$$

$$\Sigma_{i \neq j} \frac{\nu_i \nu_j}{\nu(\nu+1)} \times \{\Delta\Phi^{(1)}(\nu_i + 1, \nu + 2)\Delta\Phi^{(1)}(\nu_j, \nu + 2) - \Phi^{(2)}(\nu + 2)\}$$

$$+ \Sigma_i \frac{\nu_i(\nu_i + 1)}{\nu(\nu+1)} \times \{\Delta\Phi^{(1)}(\nu_i + 2, \nu + 2)^2 + \Delta\Phi^{(2)}(\nu_i + 2, \nu + 2)\}.$$

2. Mutual information, $M(\mathbf{p}) = \Sigma_{ij} p_{ij} \log\left(\frac{p_{ij}}{p_{i\cdot}\, p_{\cdot j}}\right)$. In this case the observed counts form a matrix. Define $\bar{\nu}_{ij} \equiv \nu_{i\cdot} + \nu_{\cdot j} - \nu_{ij}$.

Theorem 17. If the ν_{ij} are non-negative integers $\forall\, ij$ (the integer condition is used only in the simplification of the $\overline{IN}$ term in (17.2); for the other terms it may be relaxed) then

(17.1) $E[M(\mathbf{p})|\mathbf{n}] = \overline{IJ} - \bar{I} - \bar{J}$ where

$$\overline{IJ} = E[\Sigma_{ij} p_{ij} \log(p_{ij})|\mathbf{n}] = \Sigma_{ij} \frac{\nu_{ij}}{\nu} \Delta\Phi^{(1)}(\nu_{ij} + 1, \nu + 1),$$

$$\bar{I} = E[\Sigma_i p_{i\cdot} \log(p_{i\cdot})|\mathbf{n}] = \Sigma_i \frac{\nu_{i\cdot}}{\nu} \Delta\Phi^{(1)}(\nu_{i\cdot} + 1, \nu + 1),$$

$$\bar{J} = E[p_{\cdot j} \log(p_{\cdot j})|\mathbf{n}] = \Sigma_j \frac{\nu_{\cdot j}}{\nu} \Delta\Phi^{(1)}(\nu_{\cdot j} + 1, \nu + 1), \text{ and}$$

(17.2) $E[M^2(\mathbf{p})|\mathbf{n}] = \overline{IJMN} + \overline{IM} + \overline{JN} - 2(\overline{IJM} + \overline{IJN} - \overline{IN})$ where

$$\overline{IJMN} = E\left[\Sigma_{ijmn} p_{ij} \log(p_{ij})\, p_{mn} \log(p_{mn}) \,|\, \mathbf{n}\right] =$$

$$\Sigma_{ij}\Sigma_{mn \neq ij} \frac{\nu_{ij}\nu_{mn}}{\nu(\nu+1)} \times \{\Delta\Phi^{(1)}(\nu_{ij}+1, \nu+2)\Delta\Phi^{(1)}(\nu_{mn}+1, \nu+2) - \Phi^{(2)}(\nu+2)\}$$

$$+ \Sigma_{ij} \frac{\nu_{ij}(\nu_{ij}+1)}{\nu(\nu+1)} \times \{\Delta\Phi^{(1)}(\nu_{ij}+2, \nu+2)^2 + \Delta\Phi^{(2)}(\nu_{ij}+2, \nu+2)\}$$

$$\overline{IM} = E\left[\Sigma_{im} p_{i\cdot} \log(p_{i\cdot})\, p_{m\cdot} \log(p_{m\cdot}) \,|\, \mathbf{n}\right] =$$

$$\Sigma_{i}\Sigma_{m \neq i} \frac{\nu_{i\cdot}\nu_{m\cdot}}{\nu(\nu+1)} \times \{\Delta\Phi^{(1)}(\nu_{i\cdot}+1, \nu+2)\Delta\Phi^{(1)}(\nu_{m\cdot}+1, \nu+2) - \Phi^{(2)}(\nu+2)\}$$

$$+ \Sigma_{i} \frac{\nu_{i\cdot}(\nu_{i\cdot}+1)}{\nu(\nu+1)} \times \{\Delta\Phi^{(1)}(\nu_{i\cdot}+2, \nu+2)^2 + \Delta\Phi^{(2)}(\nu_{i\cdot}+2, \nu+2)\}$$

(To find $\overline{JN}$ substitute $\nu_{\cdot i}$ for $\nu_{i\cdot}$ and $\nu_{\cdot m}$ for $\nu_{m\cdot}$ in the expression for $\overline{IM}$.)

$$\overline{IJM} = E\left[\Sigma_{ij}\Sigma_{m} p_{ij} \log(p_{ij})\, p_{m\cdot} \log(p_{m\cdot}) \,|\, \mathbf{n}\right] =$$

$$\Sigma_{ij}\Sigma_{m \neq i} \frac{\nu_{ij}\nu_{m\cdot}}{\nu(\nu+1)} \times \{\Delta\Phi^{(1)}(\nu_{ij}+1, \nu+2)\Delta\Phi^{(1)}(\nu_{m\cdot}+1, \nu+2) - \Phi^{(2)}(\nu+2)\}$$

$$+ \Sigma_{ij} \frac{\nu_{ij}(\nu_{i\cdot}+1)}{\nu(\nu+1)} \times \{\Delta\Phi^{(1)}(\nu_{i\cdot}+2, \nu+2)^2 + \Delta\Phi^{(1)}(\nu_{ij}+1, \nu_{i\cdot}+1)\Delta\Phi^{(1)}(\nu_{i\cdot}+2, \nu+2)$$

$$+ \Delta\Phi^{(2)}(\nu_{i\cdot}+2, \nu+2)\}$$

(To find $\overline{IJN}$ substitute $\nu_{\cdot m}$ for $\nu_{m\cdot}$ in the expression for $\overline{IJM}$.)

$$\overline{IN} = E\left[\Sigma_{i}\Sigma_{n} p_{i\cdot} \log(p_{i\cdot})\, p_{\cdot n} \log(p_{\cdot n}) \,|\, \mathbf{n}\right] =$$

$$\Sigma_{in} \frac{\bar{\nu}_{in}(\bar{\nu}_{in}+1)}{\nu(\nu+1)} \times \{ (\Delta\Phi^{(1)}(\bar{\nu}_{in}+2, \nu+2)^2 + \Delta\Phi^{(2)}(\bar{\nu}_{in}+2, \nu+2)) \times \{1 - \frac{\nu_{i\cdot} + \nu_{\cdot n} - 2\nu_{in}}{\nu} + \frac{(\nu_{i\cdot} - \nu_{in})(\nu_{\cdot n} - \nu_{in})}{\nu(\nu+1)}\}$$

$$+ \Delta\Phi^{(1)}(\bar{\nu}_{in}+2, \nu+2) \times \Sigma_{r=0}^{\infty} \frac{Q_1(r,1)}{r!} \left[\frac{(\nu_{\cdot n} - \nu_{in})_r}{(\bar{\nu}_{in})_r} \left(1 + \frac{\nu_{i\cdot} - \nu_{in}}{\bar{\nu}_{in}+r}\right) + \frac{(\nu_{i\cdot} - \nu_{in})_r}{(\bar{\nu}_{in})_r} \left(1 + \frac{\nu_{\cdot n} - \nu_{in}}{\bar{\nu}_{in}+r}\right) \right]$$

$$+ \Sigma_{r=0}^{\infty} \Sigma_{s=0}^{\infty} \frac{(\nu_{i\cdot} - \nu_{in})_r (\nu_{\cdot n} - \nu_{in})_s}{(\bar{\nu}_{in})_{r+s}} \frac{Q_1(r,1)}{r!} \frac{Q_1(s,1)}{s!} \}$$

where Q_1 is defined in Thm. 15a.

Proof: Write $M(\mathbf{p})$ as $M(\mathbf{p}) = \Sigma_{ij} p_{ij} \log(p_{ij}) - \Sigma_i p_{i\cdot} \log(p_{i\cdot}) - \Sigma_j p_{\cdot j} \log(p_{\cdot j})$.

To prove (17.1) write $E[M(\mathbf{p})|\mathbf{n}] = \bar{IJ} - \bar{I} - \bar{J}$, where $\bar{IJ} \equiv E[\Sigma_{ij} p_{ij} \log(p_{ij}) \mid \mathbf{n}]$ etc. Recall that for the uniform prior $E[F(\mathbf{p})|\mathbf{n}] = I[F(\mathbf{p}), \mathbf{n}] / I[1, \mathbf{n}]$, and apply Thm. 13.1 with $k = 1$, $\eta_u = 1$, $\rho_u = p_{ij}$, $\rho_u = p_{i\cdot}$, $\rho_u = p_{\cdot j}$ to simplify the numerators in the expressions $\bar{IJ}$, $\bar{I}$, $\bar{J}$ respectively. From Thm. 3, $I[1, \mathbf{n}] = \frac{\gamma_{\mathbf{n}}}{\Gamma(\nu)}$. The results follow by substitution.

To prove (17.2) first square $M(\mathbf{p})$. In a manner similar to the proof of (17.1), find $E[M^2(\mathbf{p})|\mathbf{n}] = \overline{IJMN} + \overline{IM} + \overline{JN} - 2(\overline{IJM} + \overline{IJN} - \overline{IN})$, where each term may be expressed as a ratio of $I[\cdot, \cdot]$ integrals, with $I[1, \mathbf{n}]$ in the denominator. The rest of this proof outlines the calculation of the numerator terms in these ratios.

Apply Thm. 13.2 with $k = 2$, $\eta_u = 1$, $\eta_v = 1$, $\rho_u = p_{ij}$, $\rho_v = p_{mn}$ to find the $ij \neq mn$ terms of $I[\overline{IJMN}|\mathbf{n}]$, and apply Thm. 13.3 with $k = 1$, $\eta_u = 2$, $\rho_u = p_{ij}^2$ to find the $ij = mn$ terms of $I[\overline{IJMN}|\mathbf{n}]$.

Apply Thm. 13.2 with $k = 2$, $\eta_u = 1, \eta_v = 1$, $\rho_u = p_{i\cdot}, \rho_v = p_{m\cdot}$ to find the $i \neq m$ terms of $I[\overline{IM}|\mathbf{n}]$, and apply Thm. 13.3 with $k = 1$, $\eta_u = 2$, $\rho_u = p_{i\cdot}^2$ to find the $i = m$ terms of $I[\overline{IM}|\mathbf{n}]$. To find $I[\overline{JN}|\mathbf{n}]$ substitute $\nu_{\cdot i}$ for $\nu_{i\cdot}$ and $\nu_{\cdot m}$ for $\nu_{m\cdot}$ in the expression for $I[\overline{IM}|\mathbf{n}]$.

Apply Thm. 13.2 with $k = 2$, $\eta_u = 1, \eta_v = 1$, $\rho_u = p_{ij}, \rho_v = p_{m\cdot}$ to find the $i \neq m$ terms of $I[\overline{IJM}|\mathbf{n}]$, and apply Thm. 15b.3 with $\eta_1 = 1, \eta_2 = 1$ and $\rho_1 = p_{i\cdot}, \rho_2 = p_{ij}$ to find the $i = m$ terms of $I[\overline{IJM}|\mathbf{n}]$. To find $I[\overline{IJN}|\mathbf{n}]$ substitute $\nu_{\cdot m}$ for $\nu_{m\cdot}$ in the expression for $I[\overline{IJM}|\mathbf{n}]$.

Apply Thm. 15a.3 with $\eta_1 = 1, \eta_2 = 1$ and $\rho_1 = p_{i\cdot}, \rho_2 = p_{\cdot n}$ to find $I[\overline{IN}|\mathbf{n}]$. QED.

3) Average $A(\mathbf{p}) = \Sigma_{i=1}^{m} p_i X_i$. (Note that for $X_i = \delta_{ij}$, (1) of Thm. 18 below gives the Laplace Law of Succession estimator for p_j.)

Theorem 18. If $\mathrm{Re}(\nu_i) > 0 \ \forall i$ then

$$(18.1) \quad E[A(\mathbf{p}) \mid \mathbf{n}] = \Sigma_i \frac{\nu_i}{\nu} X_i .$$

$$(18.2) \quad E[A^2(\mathbf{p}) \mid \mathbf{n}] = \Sigma_{i \neq j} \frac{\nu_i \nu_j}{\nu(\nu+1)} X_i X_j + \Sigma_i \frac{\nu_i(\nu_i+1)}{\nu(\nu+1)} X_i^2 .$$

4) Variance $V(\mathbf{p}) = \Sigma_{i=1}^{m} p_i (X_i - \mu_x)^2$.

Note that $E[V(\mathbf{p}) \mid \mathbf{n}] \neq E[(A(\mathbf{p}) - E[A(\mathbf{p}) \mid \mathbf{n}])^2 \mid \mathbf{n}]$; μ_x is the true mean, not the expected mean, and $V(\mathbf{p})$ refers to the true variance, not the variance in the estimator $E[A(\mathbf{p}) \mid \mathbf{n}]$.

Theorem 19. If $\mathrm{Re}(\nu_i) > 0 \ \forall i$ then

$$(19.1) \quad E[V(\mathbf{p}) \mid \mathbf{n}] = \Sigma_i \frac{\nu_i(\nu - \nu_i)}{\nu(\nu+1)} X_i^2 - \Sigma_{i \neq j} \frac{\nu_i \nu_j}{\nu(\nu+1)} X_i X_j .$$

(19.2)

$$E[V^2(\mathbf{p}) \mid \mathbf{n}] = \Sigma_{ij} E[p_i p_j \mid \mathbf{n}] X_i^2 X_j^2 - 2\Sigma_{ijk} E[p_i p_j p_k \mid \mathbf{n}] X_i^2 X_j X_k + \Sigma_{ijkl} E[p_i p_j p_k p_l \mid \mathbf{n}] X_i X_j X_k X_l,$$

where the expectations are found by applying Thm. 12.

5) Covariance $C(\mathbf{p}) = \Sigma_{ij} p_{ij} (X_i - \mu_x) (Y_i - \mu_y)$.

Theorem 20. If $\mathrm{Re}(\nu_{ij}) > 0 \; \forall ij$ then

$$E[C(\mathbf{p}) \mid \mathbf{n}] = \Sigma_{ij} X_i Y_j \frac{1}{\nu(\nu+1)} [\nu\nu_{ij} - \nu_{i\cdot}\nu_{\cdot j}],$$

found by applying Thm. 14a.

The second posterior moment of $C(\mathbf{p})$ is given in Thm. 26.

6) Chi-Squared $\chi^2(\mathbf{p}) = \Sigma_{ij} \frac{(p_{ij} - p_{i\cdot} p_{\cdot j})^2}{p_{i\cdot} p_{\cdot j}}$.

Theorem 21. If $\mathrm{Re}(\nu_{ij}) > 0 \; \forall ij$, $\mathrm{Re}(\nu_{i\cdot}) > -1$, and $\mathrm{Re}(\nu_{\cdot j}) > -1$ then

$$E[\chi^2(\mathbf{p}) \mid \mathbf{n}] = -1 + \Sigma_{ij} \frac{(\nu-1)(\nu-2)}{(\nu_{i\cdot} + \nu_{\cdot j} - \nu_{ij} + 1)(\nu_{i\cdot} + \nu_{\cdot j} - \nu_{ij})} \times \Sigma_{m=0}^{\infty} \Sigma_{n=0}^{\infty} \frac{(\nu_{i\cdot} - \nu_{ij})_m (\nu_{\cdot j} - \nu_{ij})_n}{(\nu_{i\cdot} + \nu_{\cdot j} - \nu_{ij} + 2)_{m+n}},$$

found by applying Thm. 14a.

The second posterior moment of $\chi^2(\mathbf{p})$ is given in Thm. 27.

2e. NOTATION FOR MULTIPLE OVERLAP INTEGRALS.

The calculation of multiple overlap integrals is not straightforward. The process can be summarized as follows. First T transforms are applied and then Thms. 9.1 and 9.2 are used, leaving a convolution product of m terms. Each term in this product has the form $\tau^{\alpha_k - 1}\Pi_{i=1}^{n} e^{-\tau a_i t_i}$, $k = 1, \ldots, m$, with the a_i's taking values in $\{0, 1\}$, the t_i's the T transform variables, τ the convolution variable, and the α_k's constants (see Thm. 14a for an example). The upper index on the product, n, is determined by the overlap structure.

Since the convolution operation is both commutative and associative, the convolutions may be done in any order. However, if the convolutions are done in random order, it is quite easy to arrive at an expression for which the inverse T transforms cannot be evaluated in closed form. Further, given any particular ordering of the convolution operations, much bookkeeping needs to be done to actually find the result. Thus, it is important to have a method for quickly determining the salient aspects of any convolution ordering (whether or not the particular order chosen for the convolutions leads to an invertible expression), without actually having to do the convolutions.

In order to facilitate this, *Convolution Form* (CF) notation is now introduced. This notation captures the relevant aspects of the expressions involved and provides a guide for the rapid calculation of multiple overlap convolution integrals. First, we show how to translate convolution product terms into this notation. Then we use the CF notation to state the relevant algebraic properties of convolutions. It is not expected that the justifications for these properties should be transparent to the reader. In fact, it is not even expected that the reader will have a complete and formal understanding of CF notation; in the interests of simplicity, CF's are presented here simply as a useful algebraic framework for performing calculations. The reader interested in the formal details is directed to App. G.

Write each expression $\tau^{\alpha_k - 1}\Pi_{i=1}^{n} e^{-\tau a_i t_i}$ appearing in the convolution product (the T transformed integral mentioned two paragraphs ago) in CF notation as the CF $(\mathbf{a};\mathbf{0})$, where $\mathbf{a}$ is the n-