Assessing the Value of a Second Opinion: the role and structure of exchangeability

Frank Lad
University of Canterbury
F.Lad@math.canterbury.ac.nz

Mario Di Bacco
Universita di Bologna
antropos@unifi.it

Key Words: Conditional exchangeability, risk reduction, forecasting

For presentation at the Conference on Partial Knowledge and Uncertainty: Independence, Conditioning, and Inference, to be held at the University of Rome, 4-6 May, 2000.

Abstract

This paper displays the productive role of the judgment of exchangeability and even conditional exchangeability that should replace the misleading assertion of independence of various experts' opinions regarding uncertain situations. The application is technically rather complicated, and a basic understanding of exchangeability in its simple applications is presumed. One component of the analysis is novel. It identifies how we might learn about one exchangeable sequence of events from the outcomes of another exchangeable sequence. The substantive content of the the paper concerns the use of personal probabilities by experts in assessing the sex of human skulls found in anthropological investigations. Although we initially value the two experts' assertions exchangeably, we learn to value one of the expert's assertions more than the other's. Moreover we identify precisely how much to value the elicitation from a second expert after we have already learned the assertion of the first. Valuation is based on a decision theoretic procedure assessing reduction in risk.

1 Introduction

In every field of expertise, even the most knowledgeable acclaimed experts are uncertain! Whether in medicine, engineering, finance, physical, human, plant or animal sciences, practicing experts have become accustomed to asserting their uncertain judgments in the form of probabilities. In this context which has become generally recognised in the past fifty years, the problem of formalising a precise value for obtaining a second or third expert's opinion has become paramount. Somewhat of a vestige from the lost aura of "certain" scientific knowledge, a second opinion is commonly referred to as an "independent" opinion if the second expert is unknown to the first, or at least if the results of the first's assessment are unknown to the second. Indeed, the statistical results of the original paper that will be extended and improved in this present article rely on this "assumption" of independence in its formulations.

However in any field in which expertise is formally accredited, either by a license or a degree whose award is based on success in an examination by already qualified examiners, the probabilistic concept of "independence" of experts' assessments is not at all relevant to a statistical analysis in which "you," the statistician, wish to make use of the experts' judgments as data to inform yourself. Even when the analyses of two experts are unknown to one another, a statistician or other decision maker would surely adjust his/her conditional probability for the second expert's probability assessment given the value of the first ... precisely because both are recognised as accredited experts! To expand this idea, you are uncertain as to what probabilities the two experts will assert that a specified event in the realm of their expertise will occur. So you can assert your own probability distribution over the possible values of their pair of assertions. In doing so, you would recognise that your assertion for the value of the second expert's probability would be different if you knew what the first expert had asserted than if you did not. This is precisely what it means to say that you do not assess their judgments independently.

Happily, the extensive analysis of the judgment of exchangeability, derived mathematically from the subjective understanding of probability, is directly relevant to this problem. It allows for substantive developments of statistical analytic work that has been begun using the "assumption of independence" merely as a condition for achieving a result, even when the concept of independence should be recognised as irrelevant to the situation. If two assessors really are both expert in their substantive field of knowledge, you would want to regard the pair of their probability assertions regarding any particular event exchangeably, and as strongly interdependent. Moreover, if one of their skills is much greater than the other's, you would surely want to learn about this. Nonetheless in this case too, you would typically still regard their assertions dependently.

The present article portrays the role of exchangeability in illuminating this issue by reformulating the analysis of a study that has already generated results based on the supposed "independence" of experts' assessments. The empirical study by Cencetti et al. (1995) examined the probabilistic assertions of two anthropologists who have been asked to judge a sequence of skulls as to whether each is from a male or a female. The goal of the work was to assess the amount of information gain to be expected from obtaining a second expert's opinion. The framework for the analysis we propose here is far more complex, and will allow a whole range of interesting questions to be examined.

The relevance of exchangeability to the statistical analysis is fairly complicated, and in one aspect it is quite new. The reader will need to pay close attention to which quantities are being judged exchangeably, and also will need to recognise the distinction between exchangeability and partial exchangeability. In the application here it is partial exchangeability in the form of conditional exchangeability that is asserted. A readable technical introduction to the topic can be found in Lad (1996).

The analytic framework we develop will be recognised as relevant to a whole host of diagnosis problems under uncertainty in medicine, engineering, business and science. We defer this discussion until our concluding comments. Without further ado, let us turn to our problem. We shall use the same notation as do the original authors, and we refer the reader to their paper for more ornate details of context and substantive background.

2 Identifying the Sex of a Skull: the experimental data and notation

The article of Cencetti et al. (1995) attempts to assess the value of a second opinion by studying the probability assertions of two expert anthropologists who examine 200 skulls selected from the *Museo Nazionale di Antropologia ed Etnologia* in Florence, Italy. Unbeknown to these two experts, 104 of the skulls were those of males, and 96 were those of females. In fact, the sex for each skull was known to the study designers from records held by the museum. The order in which the skulls were presented to the experts for analysis was chosen by a lottery without replacement. Each of the two anthropologists was asked to assert a probability that each skull was from a male, using only the nearest probability values of $0, .05, .10, ..., .90, .95, 1$. The analysis by Cencetti et al. is based on the assertion that the two probability assessments "are independent." In what follows, we shall provide a reasonable form of analysis based on a judgment to regard the pairs of their assertion values exchangeably across male skulls and across female skulls. The extent of dependence of opinions between experts for any given skull can be learned then from the data. Interestingly, although the pair of probability assertions for the first skull examined are also regarded exchangeably, the data of the probability pairs provides useable information to determine the extent to which this judgment should persist.

The two experts' names are identified by the letters S and G. For each skull we define a vector of quantities (M, P_S, P_G) where M denotes the event identifying whether the skull is that of a male, P_S denotes the probability value for male asserted by S, and P_G denotes the probability value for male asserted by G. The complementary event of a female is denoted by $F \equiv (1 - M)$. Notice that there are 441 possible values for the pair of probability components of any quantity vector (M, P_S, P_G) since each of the probability assertion values has 21 possibilities, and the two are logically independent. (This means that none of the pair possibilities is impossible in principle, according to their definition.) The boldface vector $\boldsymbol{P_M}$ shall denote the vector of all the paired probability assertion values for the male skulls, and $\boldsymbol{P_F}$ will do so for the female skulls. When we refer to any specific possible probability pair vector, as opposed to the unknown quantity, we shall use the lower case notations $\boldsymbol{p_M}$ and $\boldsymbol{p_F}$, respectively. To denote the probability assertion pair for the 201st skull (a new skull to be assessed after we make inferences from the data on the 200 skulls) whose sex is unknown, we shall use the symbol $\boldsymbol{P}_{201}$, with a specific possibility for this pair denoted by $\boldsymbol{p}_{201}$.

Of course the value of M for any particular skull is unknown to anyone who is not one of the experiment's designers, but is known to the designers. We shall describe here the designers' attitudes toward the sequence of (P_S, P_G) pairs (hereafter referred to as "your" attitudes) in such a way that we can specify the coherent inferences conditioned on the experimental data of both the experts' probability assertions. Our primary goal is to compute your coherent inference regarding the sex of a skull in the case that it is not known for sure from records (say, a newly found skull) but that upon examining the new skull, the two assessors specify their own personal probability assertion values. We would like to distinguish this inference based on both experts' assertions from the inferences made if probabilities are elicited exclusively from each one of the two. Moreover, distinct inferences can be determined for the experts' probabilities for any future skull whose sex is already known to you, and also for any future skull whose sex is unknown to anyone.

3 Coherent exchangeable opinions regarding the experts' assertions

We shall presume you regard the *vector pairs* (P_S, P_G) of assertion values exchangeably across the sequence of male skulls, and also across the sequence of female skulls, for you regard them symmetrically across permutations of the orders of the pairs within the sequences. (Virtually everyone would regard these pairs exchangeably within sequences, because the order in which the skulls are presented to the experts has been chosen by lottery.) However, this exchangeability of opinion allows that you are willing to learn from the data how sharply the experts can identify the sex of the skulls. Moreover, we shall presume that the assertions are infinitely exchangeably extendible, at least as an approximation. (See Diaconis and Freedman, 1980, or Lad, 1996.) This aspect of opinion is based on the fact that the two anthropologists are already recognised experts on skulls. You might expect that if they were to learn something new and really substantial regarding skull identification in the course of their subsequent examination career, this might occur only after seeing many, many more skulls. Nonetheless you might also recognise that the process of formalising uncertain judgments numerically via probability assertions is a relatively new prospect for an anthropologist. Thus, you might make a distinction between how you would feel about how the experts' sequence of probability judgments might continue if they were not to learn the results of this assessment exercise and how you would feel about their subsequent assessments if they did! In the former case you would reasonably extend your extendible judgments. In the latter case you might extend them or you might not, depending on what was learned. This distinction is discussed in Lad (1996).

The analysis of inferences regarding a 201^{st} skull in this paper presupposes that your exchangeable opinions do extend to judgments of that skull. That is, your predictive probability is derived on the supposition that the experts have not been informed by the results of this study. The analysis in the case that your exchangeable judgments are not extended because of what the experts will have learned from the results of the exercise would be better left until after the results are aired. This distinction is akin to the general problems in which the result of a measurement of a process is recognised to effect the continuing conditions of the process.

As to the probability pairs asserted for any particular skull, presumably you would initially regard any particular pair (the components of the first pair you observe) symmetrically. (As a statistician, you are only informed that both of the assessors are recognised experts.) However you may well be willing to learn from the data that one of the experts tends to give higher (or lower) probabilities for male to the male skulls than does the other expert, and tends to give higher (or lower) probabilities for female to female skulls. To the contrary on the other hand, you may learn to continue to regard the pair of experts' assessments virtually exchangeably for any given skull! In contradistinction to the question as to whether to extend the exchangeable judgments to assessments of future skulls, the decision as to whether to continue to regard the assertions of the two probability assessors exchangeably can be embedded into the structure of the joint distributions for the assessments as they are regarded by you, the experiment designer.

To see precisely how this works, we need to consider carefully the coherent logic of the relevant probability assertions, to arrive at a useful representation. In the first place, the coherency condition for conditional probabilities tells us directly that

$$P[M_{201}|(\boldsymbol{P}_{201}=\boldsymbol{p}_{201})(\boldsymbol{P_M}=\boldsymbol{p_M})(\boldsymbol{P_F}=\boldsymbol{p_F})]$$
$$= \frac{P[(\boldsymbol{P_M}=\boldsymbol{p_M})(\boldsymbol{P_F}=\boldsymbol{p_F})M_{201}(\boldsymbol{P}_{201}=\boldsymbol{p}_{201})]}{Numerator \ + \ P[(\boldsymbol{P_M}=\boldsymbol{p_M})(\boldsymbol{P_F}=\boldsymbol{p_F})F_{201}(\boldsymbol{P}_{201}=\boldsymbol{p}_{201})]} \quad . \tag{1}$$

On the left-hand-side of equation (1) is the predictive probability that a 201^{st} skull is male, given the pair of probabilities that it is male asserted by the experts and given also the record of the experts' probability assertions for the 104 known male skulls and for the 96 known female skulls. Suppressed in the notation is the fact that the designers (you) know the correct sexes of the skulls for which the experts have asserted these probabilities. The analysis would be different if the correct sexes of the 200 experimental skulls were not known.

To clarify our bearings in this analysis, notice that the data of the $P(M)$ assertion pairs from experts S and G have already been recorded for the known male and female skulls. Thus, what we should like to compute is the value of equation (1) conditional on each of the 441 possible assertion probability pairs for the 201^{st} skull, $\boldsymbol{p}_{201}$, and conditioned as well on the specific data values of $\boldsymbol{p_M}$ and $\boldsymbol{p_F}$. These data of 200 assertion pairs will provide the basis for the several likelihood functions (information transfer functions) used in the inference.

Now a simple factoring of the compound probability appearing in the numerator of (1) yields an equivalent useful representation as the product of three conditional probabilities:

$$P[(\boldsymbol{P_F}=\boldsymbol{p_F})|(\boldsymbol{P_M}=\boldsymbol{p_M})(\boldsymbol{P}_{201}=\boldsymbol{p}_{201})M_{201}]\ P[(\boldsymbol{P_M}=\boldsymbol{p_M})(\boldsymbol{P}_{201}=\boldsymbol{p}_{201})|M_{201}]\ P(M_{201}) \quad . \tag{2}$$

Similarly, the second summand in the denominator of (1) factors into

$$P[(\boldsymbol{P_M}=\boldsymbol{p_M})|(\boldsymbol{P_F}=\boldsymbol{p_F})(\boldsymbol{P}_{201}=\boldsymbol{p}_{201})F_{201}]\ P[(\boldsymbol{P_F}=\boldsymbol{p_F})(\boldsymbol{P}_{201}=\boldsymbol{p}_{201})|F_{201}]\ P(F_{201}) \quad . \tag{3}$$

In the next two Sections we shall elucidate details of the component probabilities in expressions (2) and (3) respectively, particularly the first two multiplicand probabilities in each product.

4 Analysing the assertions $\boldsymbol{P_M}$, $\boldsymbol{P_F}$ and $\boldsymbol{P}_{201}$ conditioning on M_{201}

It is evident from the product expressions (2) and (3) that we need to analyse our understanding of the experts' probability assertion sequences $\boldsymbol{P_M}=\boldsymbol{p_M}$ and $\boldsymbol{P_F}=\boldsymbol{p_F}$ along with the assertions for the new skull, $\boldsymbol{P}_{201}=\boldsymbol{p}_{201}$, as we would understand them conditioned on the 201^{st}

skull being male, M_{201}, and again conditioned on it being female, F_{201}. We shall begin our analysis by conditioning on M_{201} in three subsections with extensive discussion. The analysis conditioning on F_{201} will be equally extensive, but will be similar in structure. Thus, its results can be reported more efficiently in Section 5, capitalising on the developments that we now describe.

4.1 Exchangeability of the assertions regarding male skulls, given M_{201}

We shall begin our analysis of the three multiplicand probabilities in (2) with the term $P[(\boldsymbol{P_M} = \boldsymbol{p_M})(\boldsymbol{P}_{201} = \boldsymbol{p}_{201})|M_{201}]$, for it succumbs most easily to familiar arguments based on exchangeability. Remember and beware, that this part of the analysis of (1) *conditions* on the new skull (of unknown sex) being assessed is male. Moreover, this analysis presumes that the assertion of exchangeability regarding probability assessments of the known male skulls does extend to the 201^{st} skull, conditioning of course on it being a male skull.

In the context of your exchangeable assertions regarding the sequence of probability pairs asserted for known male skulls, and additionally for the new skull conditional on it being male, the sufficient statistics for any particular vector of observations and the conditioning assertion pair is the histogram showing how many of the probability pairs fall into each of the 441 categories that are logically possible. Of the 104 male skulls plus the new skull (conditional on it being male), we shall denote by $\boldsymbol{H_M}(\boldsymbol{p_M}, \boldsymbol{p}_{201})$ the vector of numbers of the probability pairs that fall into each of the categories. The number of observations in any histogram category is denoted merely by $H_{Mi}(\boldsymbol{p_M}, \boldsymbol{p}_{201})$.

Given this opinion structure, de Finetti's representation theorem specifies that a designer's joint probability mass function for the histogram of the assessed probability pairs must be mixture multinomial, and the mass function for any particular sequence of pairs given the histogram must be uniform over all permutations of any sequence that gives rise to it. In extensive notation, a designer's joint probability mass function for any particular vector of the experts' probability assertion values, $(\boldsymbol{p_M}, \boldsymbol{p}_{201})$, has the form

$$f(\boldsymbol{p_M}, \boldsymbol{p}_{201}|M_{201}) \;=\; \int_{\theta_1}\int_{\theta_2} \; ... \; \int_{\theta_{441}} \prod_{i=1}^{441} \theta_i^{\;H_{Mi}(\boldsymbol{p_M}, \boldsymbol{p}_{201})} \; dF(\theta_1, \theta_2, ..., \theta_{441}) \qquad (4)$$

for each possible vector of probability assertions by the experts. The mass function for the histogram vector implied by $(\boldsymbol{p_M}, \boldsymbol{p}_{201})$ would be this same value multiplied by ${}^{N}C_{\boldsymbol{H_M}(\boldsymbol{p_M}, \boldsymbol{p}_{201})}$, the combinatorial multinomial coefficient.

To complete this opinion structure we need only identify some reasonable form for the initial mixing distribution $F(\theta_1, \theta_2, ..., \theta_{441})$ that appears in the integral of (4). The easiest resolution here, and also quite reasonable through the interpretation of of prior opinion as information based on a number of equivalent observations (see Good, 1965), is through use of a natural conjugate Dirichlet form. We can locate the initial Dirichlet mixing function parameters by constructing the initial expectation for the parametric theta values through a conditioning argument, and then select an appropriate scale that might represent the strength of opinion of the experimental designers.

4.1.1 Locating the initial Dirichlet mixture for males, and choosing the scale

Of course there are as many initial mixing functions as there are people to assess them. At this point in these investigations, we shall limit our analysis to a single considered opinion regarding the probability assertions of the two experts, focusing as much on the procedure as on the detail of the content. Variations on the opinions expressed can be studied subsequently in a Bayesian style robustness exercise.

The Dirichlet distribution over $\boldsymbol{\theta}_{441}$ is characterised by 441 parameters, $\boldsymbol{\alpha}_{441}$. The relative sizes of these alpha specifications can be located by specifying your initial probability for each

(p_S, p_G) pair as the experts' assessments for the first skull they examine. The scale is then chosen as a proportionality constant to determine the *sum* of the alphas.

An initial distribution over the $\boldsymbol{\theta}_{441}$ parameters will be motivated by two considerations. First is the fact that the anthropologist experts may well *not* be very expert at recognising and assessing their own uncertainties via numerical probability assertions. In this regard they may also not be experienced at paying a monetary price for overexhuberant probabilities as are business practitioners, for example. The practice of assessing their uncertainties as probabilities would be fairly new to them. As a result, one considered feature of their assertions is that they would be regarded as too willing to assert virtual certainty in their judgments, even when their knowledge is infact mistaken! As a result, the initial mixture will be weighted in terms of .3 times a distribution that masses .91 on $(1,1)$, .04 on $(1,0)$, .04 on $(0,1)$, and and masses .01 on $(0,0)$. The remaining .7 weight will be massed across a digital exponential joint mass function that will now be explained.

We shall locate this aspect of the Dirichlet function by describing an appropriate joint p.m.f. for values of a (P_S, P_G) assertion pair within the unit square, massed on the 441 possibility points. This aspect of the p.m.f. will be based on an idea of how precisely the skulls would exhibit readable clues to their sex. Moreover, we would like the initial mixing density to be symmetric with respect to P_G and P_S, being aware that this feature of the mixing distribution may change as data accumulates if that is what the data suggests.

We shall construct this joint p.m.f. by first considering a p.m.f. for the sum of the assertions, $\Sigma = P_S + P_G$, and then considering a conditional p.m.f. for either summand, P_S or P_G, given the value of their sum. Knowing that the skull in question is from a male, and that the assessors S and G are both experts, you might well specify your mass function for the sum of their probabilities by a digitised increasing exponential function that masses only on the prescribed 41 possibilities from 0 to 2 in increments of .05. Algebraically, this function has the form

$$f(\Sigma) = K(\Sigma)e^{-\lambda(2-\Sigma)} \tag{5}$$

over the 41 possibilities of the sum, Σ, just mentioned. The proportionality constant $K(\Sigma)$ merely assures that $f(\Sigma)$ sums to 1 over the 41 possibilities. Remembering that each expert's assertion values are accurate only to within .025, this p.m.f. represents only that higher probability assertions are more likely than lower assertions, and that they rise exponentially by category. The growth rate (negative decay rate to the left from 2) is parameterised in this form by λ.

Now given the sum of their assertion values, you would initially regard either individual's assertion symmetrically about 1/2 of the sum value, as you have no reason initially to suspect that one of the experts asserts higher probabilities for male skulls than does the other. Again, an exponential rate of decay in either direction, parameterised here by λ_C, would yield the conditional p.m.f.

$$f(p_S|\Sigma) = K(\lambda_C, \Sigma)e^{-\lambda_C|p_S - \Sigma|} \tag{6}$$

massed on p_S units withinin the interval of allowable mass points bounded by $[max(0, 1 - \Sigma), min(1, \Sigma)]$.

On the basis of this argument, a construction for the joint p.m.f. for (P_S, P_G), deriving merely from a translation of the product of equations (5) and (6), would have the form

$$f(p_S, p_G) = K(\lambda, \lambda_C)e^{-\lambda(2-p_S-p_G))}e^{-\lambda_C|p_S/2-p_G/2|} \tag{7}$$

over the lattice of (p_S, p_G) possibilities set in the unit-square. This form is obviously symmetric with respect to permutations of its arguments. The parameter λ is an accuracy parameter, increasing with your attitude toward the success of the experts in assessing the skulls as male. The parameter λ_C is a mutuality parameter, increasing with the extent to which you believe the assessors will agree with one another in their assessment.

Finally, merging the two components of the prior specification of the Dirichlet parameters, we can identify the location of the parameters via a mixed p.m.f. derived from a weight of .7 on equation (7) and a weight of .3 on a p.m.f. that masses .91 on $((P_S, P_G) = (1,1))$, .04 on

$((P_S, P_G) = (1,0))$, .04 on $((P_S, P_G) = (0,1))$ and .01 on $((P_S, P_G) = (0,0))$. To cut a long story short at this point, the scale of the Dirichlet parameters has been chosen so that the sum of the location parameters equals 10. This is the value of the "equivalent number of observations" that identifies the strength of the prior information. The sum of 10 identifies a rather mild opinion.

The location of the initial Dirichlet mixture shown in Figure 1 is based on equation (7) with a choice of $\lambda = 2.5$ and $\lambda_C = 5$. To get your bearings, know that with these specifications of λ and λ_C, the doubly exponential p.m.f. component based on equation (7) places probability .84 that P_S is at least .5, .47 that it is at least .8, and .25 that it is at least .9. You, the reader, may regard these probabilities as too pessimistic regarding the experts' abilities, but notice that it is being mixed with the pmf with all weight on the four corner points of the unit-square, .94 being on $(1,1)$. Nonetheless as mentioned, a robustness study is surely possible relative to other specifications. The sharpness of λ_C exhibits double the strength of expected mutuality of the two experts' assessments.

As it turns out, the weight of .3 on the mass function only on the four corner points of the unit-square visually dominates the .7 weight on the exponentially distributed component. Thus, Figure 1 displays the exponential p.m.f. of equation (7) on the left-hand-side separately from the resulting choice of alpha parameters for the Dirichlet prior on the right-hand-side.

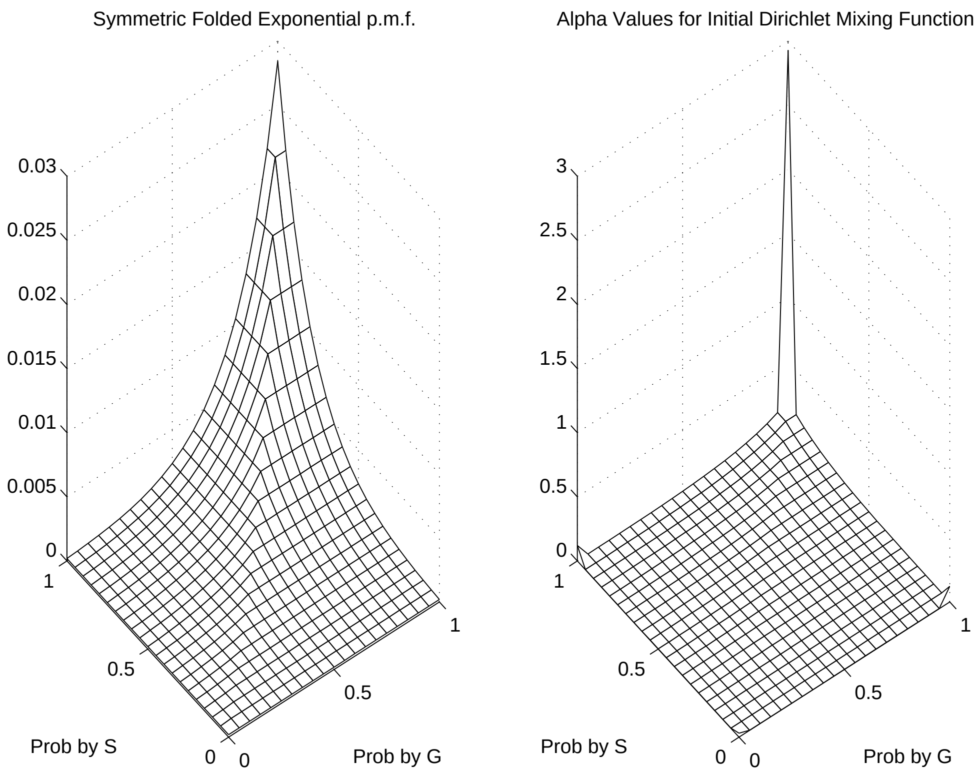

Figure 1: The graph at left displays the joint folded exponential p.m.f. of equation (7) that determines .7 of the weight on the initial Dirichlet location parameters, while the graph at right displays the complete array of Dirichlet location parameters derived from the weighted mixture of the p.m.f. at left and .3 of the weight on the one massing .91 on (1,1), .04 on (0,1), .04 on (1,0), and .01 on (0,0), and using a total scale on the parameters to sum to 10.

4.1.2 The histogram of assertions regarding male skulls

It is merely a technical matter at this point to present the histogram of the probability assertion pairs for the 104 male skulls. This appears at the left-hand-side of Figure 2. At the right-hand-side appears a visual presentation of this histogram augmented by the alpha parameters of the initial Dirichlet mixing distribution, which is relevant to the Polya probabilities for the observed assertion data that we now discuss. Of course to complete this computation, notice that the histogram will need to be supplemented sequentially by a single observation component in each category to represent the possibility of the 201^{st} assertion pair fitting into that category. Figure 2 is displayed merely to present a visual idea of what the experimental data are saying to this point.

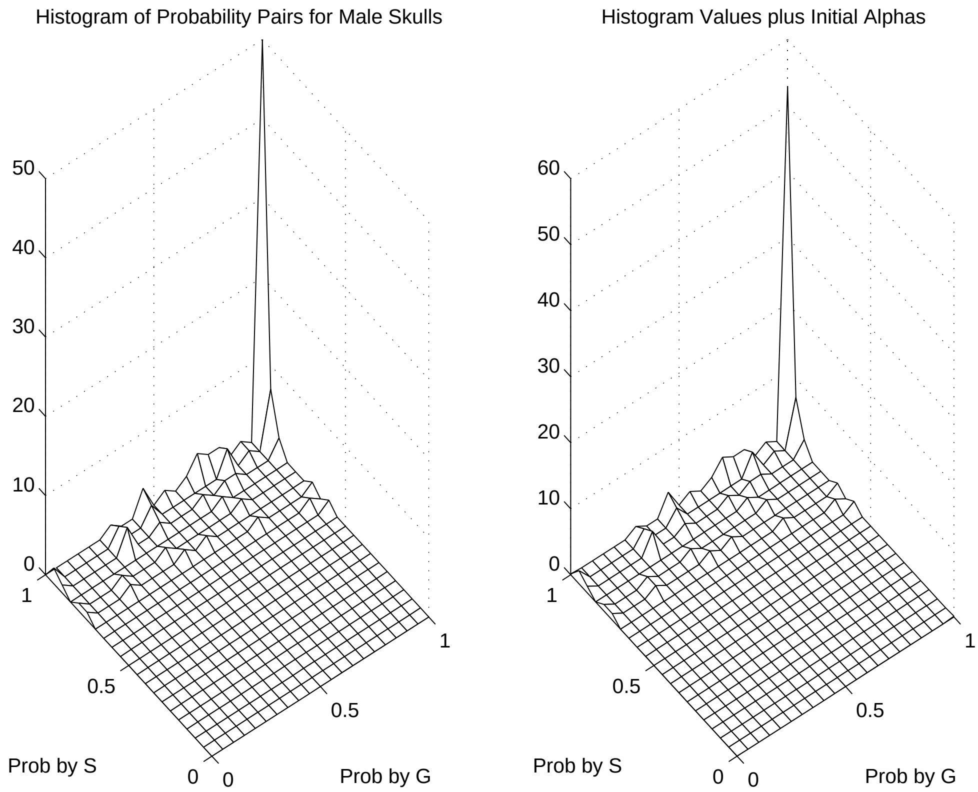

Figure 2: At left is the histogram of probability assertion pairs (P_S, P_G) for the 104 male skulls. At right appear the sum of each histogram category and its corresponding initial alpha specification.

The observant reader may well be surprised, as we were, in observing that histogram. For it appears that expert S is far more proficient than is expert G at identifying male skulls! (For your information, and to aid your musing, the marginal histograms of probability assertions by G and by S for the male skulls are shown in Table 1.) Notice, for example, that all of S's probabilities exeed .5, and 90% of them exceed .9; whereas G asserts probability near to 0 for 5 of the skulls, and rather diffuse probabilities for skulls that S assesses with probabilities exceeding .75, even those as high as .9. Moreover, the empirical cdf of G's probabilities stochastically dominate those of S over the entire range of possibilities short of certainty. Rather than get lost in discussion at this point, however, we leave the reader to mull over this as we continue to develop the full details of the complete structure of the analysis.

Table 1. Marginal Histograms for the P_G and P_S Assertions Regarding the Male Skulls.

p	0	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
$\boldsymbol{H(p_S)}$	0	0	0	0	0	0	0	0	0	0	0
$\boldsymbol{H(p_G)}$	5	0	0	0	2	0	6	0	1	7	0

p	.55	.60	.65	.70	.75	.80	.85	.90	.95	1
$\boldsymbol{H(p_S)}$	0	1	1	2	4	4	1	10	17	64
$\boldsymbol{H(p_G)}$	2	0	2	4	5	6	0	2	1	61

4.1.3 Polya probabilities for the data on male skulls

As is well-known (see Lad, 1996, p. 322) in this Dirichlet-multinomial mixture context, equation (4) integrates to yield an individuated Polya mass function for the experts' probability assertion sequence. The detailed equation to compute these probabilities is

$$f(\boldsymbol{p_M}, \boldsymbol{p}_{201}|M_{201}) \;=\; \frac{,\,(\sum_{i=1}^{441}\alpha_i)\;\prod_{i=1}^{441},\,(\alpha_i + H_{Mi}(\boldsymbol{p_M}, \boldsymbol{p}_{201}))}{,\,(105+\sum_{i=1}^{441}\alpha_i)\;\prod_{i=1}^{441},\,(\alpha_i)} \tag{8}$$

Remembering that $\boldsymbol{p_M}$ is fixed by the experimental data, but that $\boldsymbol{p}_{201}$ is variable, depending on what the two experts might assert about the new skull (the 201^{st} skull), it is useful to recognise that equation (8) reduces to

$$f(\boldsymbol{p_M}, \boldsymbol{p}_{201}|M_{201}) \;=\; K\;(\alpha^*(\boldsymbol{p}_{201}) + H_{M^*}(\boldsymbol{p_M}, \boldsymbol{p}_{201})) \tag{9}$$

presuming that the assertion for the new skull, $\boldsymbol{p}_{201}$, falls into category * of the 441 possibilities. The proportionality constant in equation (9) is

$$K \;=\; \frac{,\,(\sum_{i=1}^{441}\alpha_i)\;\prod_{i=1}^{441},\,(\alpha_i + H_{Mi}(\boldsymbol{p_M}))}{,\,(105+\sum_{i=1}^{441}\alpha_i)\;\prod_{i=1}^{441},\,(\alpha_i)} \tag{10}$$

The results of this computation appear visually by the right-hand-side of Figure 2, modulo the addition of 1 to every point on that grid. The proportionality constant, which will be relevant to the total computation of equation (1), equals $1.1188 \times 10(-167)$.

4.2 Exchangeability of the assertions regarding female skulls, conditioned on $(\boldsymbol{P_M} = \boldsymbol{p_M})$ and M_{201}

We now shift our analysis to the first term of the factored expression (2), $P[(\boldsymbol{P_F} = \boldsymbol{p_F})|(\boldsymbol{P_M} = \boldsymbol{p_M})(\boldsymbol{P}_{201} = \boldsymbol{p}_{201})M_{201}]$. That is, we need to discuss your opinion about the two experts' probability assertions for the female skulls, conditioned on the knowledge you obtain from observing what they had asserted for the skulls that were known to be male, $(\boldsymbol{P_M} = \boldsymbol{p_M})$ and also conditioned on their assertions for the 201^{st} skull presuming it too to be male.

To begin, remember in what follows that the assertions $\boldsymbol{P_F} = \boldsymbol{p_F}$ are the numerical probability values for the event that each skull is male, in the state tht you, the experimenter, know that the skulls are all females! Recognise therefore that since the assessors are anthropological experts you are going to be expecting their assertion values to be small, generally less than .5 and actually closer to zero rather than large. We mention this now just to help you get your bearings in following some symmetric transposition arguments that follow.

The experimental setup continues to merit your assertion of exchangeability of the vector pairs in the sequence $\boldsymbol{P_F} = \boldsymbol{p_F}$. However the fact that we are now conditioning on $(\boldsymbol{P_M} = \boldsymbol{p_M})(\boldsymbol{P}_{201} = \boldsymbol{p}_{201})M_{201}$ means that the exchangeable representation equation (4) translates in this context to

$$f(\boldsymbol{p_F}|(\boldsymbol{P_M} = \boldsymbol{p_M})(\boldsymbol{P}_{201} = \boldsymbol{p}_{201})M_{201}) \;=\;$$

$$\int_{\phi_1}\int_{\phi_2} \cdots \int_{\phi_{441}} \prod_{i=1}^{441} \phi_i^{H_{Fi}(\boldsymbol{p_F})}\; dF[\phi_1, \phi_2, ..., \phi_{441}|(\boldsymbol{P_M} = \boldsymbol{p_M})(\boldsymbol{P}_{201} = \boldsymbol{p}_{201})M_{201}] \tag{11}$$

This equation (11) is quite similar to (4). Interestingly, however, the mixture distribution here requires the initial mixing function of the parameters $\boldsymbol{\phi}_{441}$ to be posterior to the observed probability assertion pairs by the experts for the male skulls! (Additionally, the condition is supplemented distinctly by each of the possible assertion pairs for the 201^{st} skull, $\boldsymbol{P}_{201} = \boldsymbol{p}_{201}$.) Thus, an initial mixture over $\boldsymbol{\phi}_{441}$, the probability pair parameters for female skulls, must be transformed to a posterior conditional on the assessed probabilities for the known male skulls.

4.2.1 Specifying the Dirichlet mixture for females, and informing it with the assertion data on male skulls

The required mixture can be computed simply enough using Bayesian inference. The natural form for the initial mixture on $\boldsymbol{\phi}_{441}$ would again be Dirichlet, but with $\boldsymbol{\alpha}_{441}$ parameters generated by a 180 degree rotation of the male parameter matrix about its center point, $(.5, .5)$. Algebraically, this means that $\alpha(p_S, p_G)_F = \alpha(1-p_S, 1-p_G)_M$. Computationally, it is achieved by flipping the rows of the alpha matrix for males up and down to reverse their order, and again flipping the columns back and forth to reverse their order. Conceptually, this means that your expectations regarding the incidence of an (p_S, p_G) pair for a female skull are identical to your expectations regarding the incidence of the pair $(1 - p_S, 1 - p_G)$ for a male skull. There is no need to display the results of this generating procedure because you can imagine the results from examining Figure 1 and merely inverting the scales on the P_S and P_G axes to read from 1 to 0 rather than from 0 to 1.

The second stage of transforming this prior form to a posterior form given the results of the experts' assertions requires an explanatory remark. The observable clues to the anthropologist that a skull is from a male are precisely complementary to the clues that the skull is from a female. Thus, you might *expect* that if you learned that one expert was much more precise in asserting high probabilities for male to the male skulls, you would imagine that person also to tend to be more precise in asserting low probabilities for male to female skulls. However the making of these two judgments regarding any two skulls are not precisely mirroring activities. Thus, rather than formulating the likelihood function for the female $\boldsymbol{\phi}_{441}$ components based on the probability pairs asserted for the male skulls as a Dirichlet form from a pure Binomial mass function, we suppose you would want to reduce the informative power of the data on male skulls to a fraction of its power in informing us about assessments for male skulls. For now, we will specify the fraction as .3, yielding the likelihood function as

$$L(\boldsymbol{\phi}_{441}; \boldsymbol{P_M} = \boldsymbol{p_M}) \;=\; \prod_{i=1}^{441} \phi_i^{.3H_{Mi}(\mathbf{1} - \boldsymbol{p_M}, \mathbf{1} - \boldsymbol{p}_{201})} \quad . \tag{12}$$

The expressions $\mathbf{1} - \boldsymbol{p_M}$ and $\mathbf{1} - \boldsymbol{p}_{201}$ in that likelihood function denote that each of the probability components of the vector pairs $\boldsymbol{p_M}$ and $\boldsymbol{p}_{201}$ is arithmetically inverted to its additive reciprocal of $1 - p$.

The left-hand side of Figure 3 displays the numerical results of the weakened transformed histogram of equation (12), while the right-hand side displays the sum of these histogram values with the initial alpha parameters of the Dirichlet form, to yield the alpha parameters representing the mixing function $F[\phi_1, \phi_2, ..., \phi_{441} | (\boldsymbol{P_M} = \boldsymbol{p_M})(\boldsymbol{P}_{201} = \boldsymbol{p}_{201})M_{201}]$ that pertain to the integral mixture equation (11).

4.2.2 The histogram of assertions regarding female skulls

Parallel to the presentation of Figure 2, the left-hand-side of Figure 4 portrays the histogram of the 96 $\boldsymbol{p_F}$ assertion pairs for the female skulls that fall into each of the 441 possible categories. On the right-hand-side appear these histogram values augmented by the alpha parameters of the mixing Dirichlet function for the female skulls, which had been displayed on the right-hand-side of Figure 3.

For the moment once again, we defer detailed comment on the content of the histogram, leaving you to mull over the fact that again expert S appears to be more knowledgable in skull identification than is G, although this is not quite so clearcut as it seemed for male skulls. Whereas G asserts probability near 0 for a higher number of skulls than does S, the remainder of G's assertions are spread widely across the domain of possibilities, so that the empirical cdf for S's assertions dominate those of G for probability assertions above .05. G even asserts probabilities (of male) near 1 for five of those female skulls whereas S does so only for one of them. Moreover, both experts appear to be somewhat more uncertain in their assertions than

Figure 3: At left is the transposed histogram of assertions for male skulls, diminished by a factor of .3. At right is the sum of this diminished histogram with the initial alpha specifications for the female skull assertions.

they were for male skulls. The marginal histograms for the assertions of S and G are shown in Table 2.

Table 2. Marginal Histograms for the P_G and P_S Assertions Regarding the Female Skulls.

p	0	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
$\boldsymbol{H}(\boldsymbol{p_S})$	27	13	18	7	2	5	3	5	3	5	0
$\boldsymbol{H}(\boldsymbol{p_G})$	41	0	11	3	4	1	4	2	6	4	0

p	.55	.60	.65	.70	.75	.80	.85	.90	.95	1
$\boldsymbol{H}(\boldsymbol{p_S})$	0	0	1	1	0	1	0	2	2	1
$\boldsymbol{H}(\boldsymbol{p_G})$	2	3	3	0	3	2	0	2	0	5

4.2.3 Polya probabilities for the data on female skulls

Polya probabilities for the observed assertion pairs of the experts on female skulls that result from the Dirichlet-Binomial mixture are displayed visually on the right-hand-side of Figure 4. Algebraically, they derive from equations similar to (8), (9) and (10). However the computational detail is a bit different because of the different way that the variable vector $\boldsymbol{p}_{201}$ appears in the equation as part of the "initial α_i." The computational equations are

$$f(\boldsymbol{p_F}|\boldsymbol{p_M}, \boldsymbol{p}_{201}, M_{201}) \; = \; K \; \frac{,\, (\alpha^*(\boldsymbol{p}_{201}) + .3(H_{M^*}(\boldsymbol{p_M}) + 1) + H_{F^*}(\boldsymbol{p_F}))\ ,\, (\alpha^*(\boldsymbol{p}_{201}) + .3(H_{M^*}(\boldsymbol{p_M})))}{,\, (\alpha^*(\boldsymbol{p}_{201}) + .3(H_{M^*}(\boldsymbol{p_M})) + H_{F^*}(\boldsymbol{p_F}))\ ,\, (\alpha^*(\boldsymbol{p}_{201}) + .3(H_{M^*}(\boldsymbol{p_M}) + 1))} \qquad (13)$$

presuming that the assertion for the new skull, $\boldsymbol{p}_{201}$, falls into category * of the 441 possibilities. The proportionality constant in equation (13) is

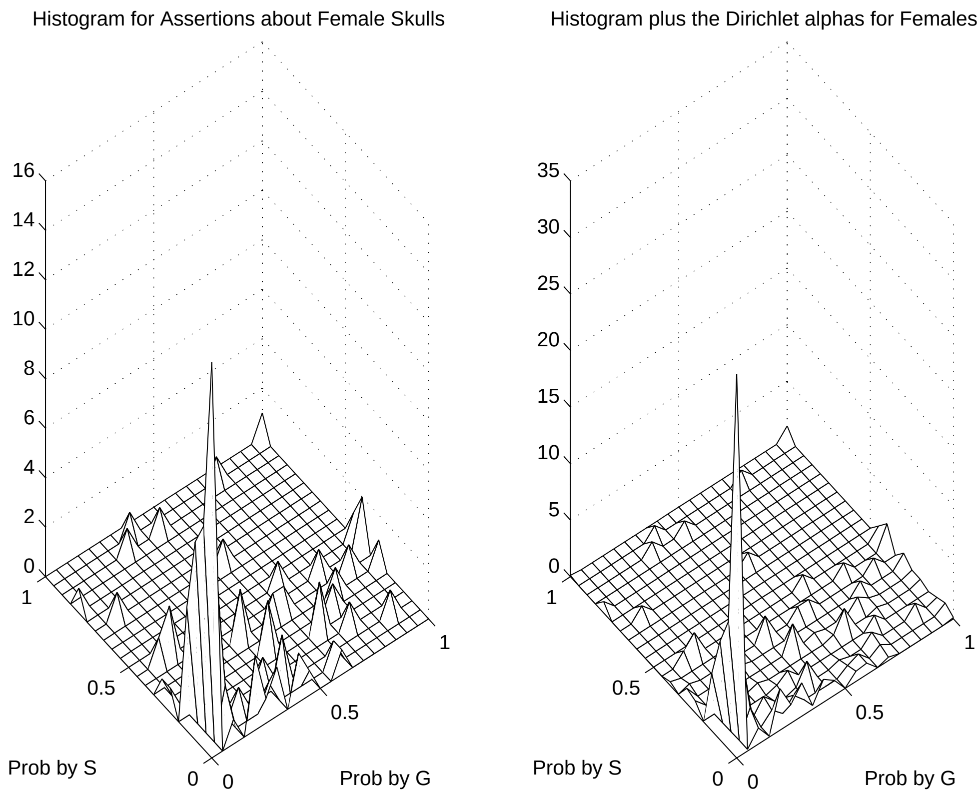

Figure 4: At left is the histogram of assertions for the 96 female skulls. At right appears the sum of this histogram with the corresponding Dirichlet alphas that are posterior to the observations on the male skulls.

$$K \;=\; \frac{,\,(10+.3(105))\;\prod_{i=1}^{441}\,,\,(\alpha_i+.3H_{Mi}(\boldsymbol{p_M})+H_{Fi}(\boldsymbol{p_F}))}{,\,(96+10+.3(105))\;\prod_{i=1}^{441}\,,\,(\alpha_i+.3\;H_{Mi}(\boldsymbol{p_M}))} \tag{14}$$

4.3 The matrix of numerator probabilities

The final computational step in producing the numerical results for the numerator of equation (1) factored in expression (2), requires merely that we multiply equations (9) and (13).

We have now completed a computable representation of the multiplicand components of the numerator of equation (1). In the next Section we shall perform the similar compuational steps of the second denominator term, expression (3).

5 Analysing the experts' assertions conditioning on F_{201}

All the computational proceedings of this Section will mimic those in Section 4, except that we shall be conditioning on the 201^{st} skull being female. As a result the order in which the probability assessments for the male and the female skulls will be reversed. Thus, the explanations will be rather terse for now, just in order to get the results down. More extensive commentary will appear subsequently in discussion.

We shall be computing the second term in the denominator of equation (1), which is detailed in factored form as expression (3).

5.1 Exchangeability of the assertions regarding the female skulls, given F_{201}

Given that the new skull is female, the pair of assertions by S and by G are regarded exchangeably with their assertions for the other 96 female skulls. Again, a multinomial mixture mass function is required to represent the p.m.f. $f(\boldsymbol{p_F}, \boldsymbol{p}_{201}|F_{201})$, similar to equation (4).

5.1.1 Locating the initial Dirichlet mixture for females, and choosing the scale

The initial mixing function for the female skulls is the 180 degree rotation of the mixing function for males as described in Subsection 4.1.1. Hopefully it requires no further explanation at this time. For the record, it is displayed in two portions in Figure 5.

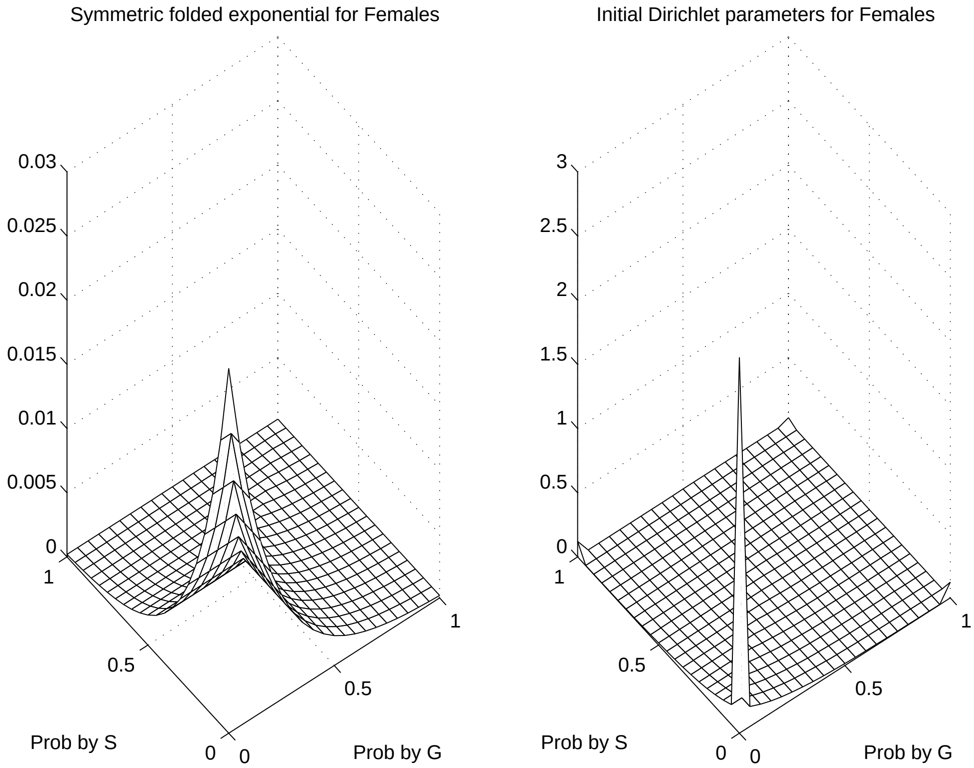

Figure 5: The graph at left displays the joint folded exponential p.m.f. of equation (7) that determines .7 of the weight on the initial Dirichlet location parameters, while the graph at right displays the complete array of Dirichlet location parameters derived from the weighted mixture of the p.m.f. at left and .3 of the weight on the one massing .91 on (0,0), .04 on (0,1), .04 on (1,0), and .01 on (1,1), and using a total scale on the parameters to sum to 10. It should be evident that these graphs are merely rotations of the graphs displayed in Figure 1.

5.1.2 The histogram of assertions regarding female skulls

Figure 6 displays at left the histogram of the experts' assertions regarding the 96 known female skulls. At right it shows the sum of these category counts with the alpha values of the initial Dirichlet mixing density.

The reader may wish to reexamine the marginal histograms of assertions for the female skulls by S and by G, which has appeared in Table 2.

5.1.3 Polya probabilities for the data on female skulls

The Polya probilities that derive from the Dirichlet mixture for the assertions regarding female skulls can be examined visually from the right-hand-side of Figure 6, since they are proportional to those values. The argument supporting this mimics the parallel discussion in Subsection 4.1.3.

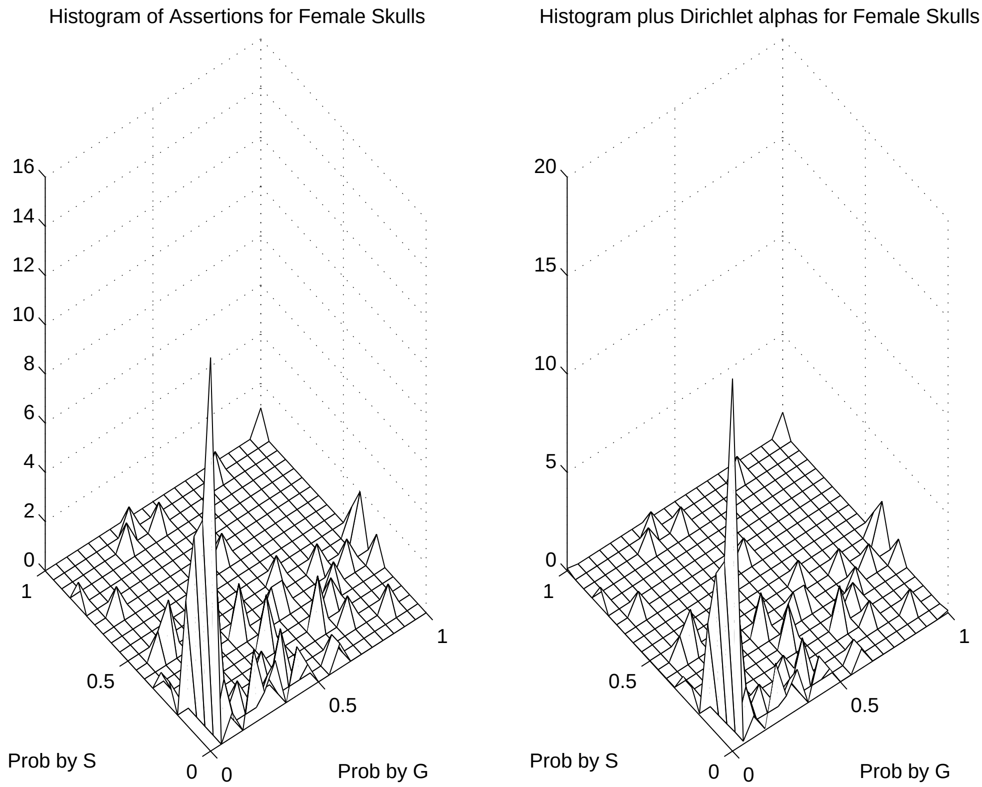

Figure 6: At left is the histogram of probability assertion pairs (P_S, P_G) for the 96 female skulls. At right appear the sum of each histogram category and its corresponding initial alpha specification.

5.2 Exchangeability of the assertions regarding male skulls, conditioned on $(\boldsymbol{P_F} = \boldsymbol{p_F})$ and F_{201}

Again, we regard the assertions about male skulls exchangeably given the results on the female skulls, deriving a mixture distribution parallel to that shown in Section 4.2.

5.2.1 Specifying the Dirichlet mixture for males, and informing it with the assertion data on female skulls

Figure 7 shows at left the initial Dirichlet parameters for the assertions regarding male skulls, but supplemented by the weakened rotated histogram of results regarding female skulls. Remember that the initial Dirichlet parameters without this supplemental information have already appeared in Figure 1. Since the histogram of assertions regarding male skulls has also already been displayed, the right-hand-side of Figure 7 shows only the "updated" prior parameters of the left-hand-side with the histogram components for the male skulls added to them.

5.2.2 The histogram of assertions regarding male skulls

The results of this subsection already appear as described in Figure 7. That figure merely shows the male skull histogram added onto the augmented male skull Dirichlet parameters.

5.3 Polya probabilities for the data on male skulls

The computational equations for $f(\boldsymbol{p_M}|\boldsymbol{p_F}, \boldsymbol{p}_{201}, F_{201})$ would mimic those of equations (13) and (14).

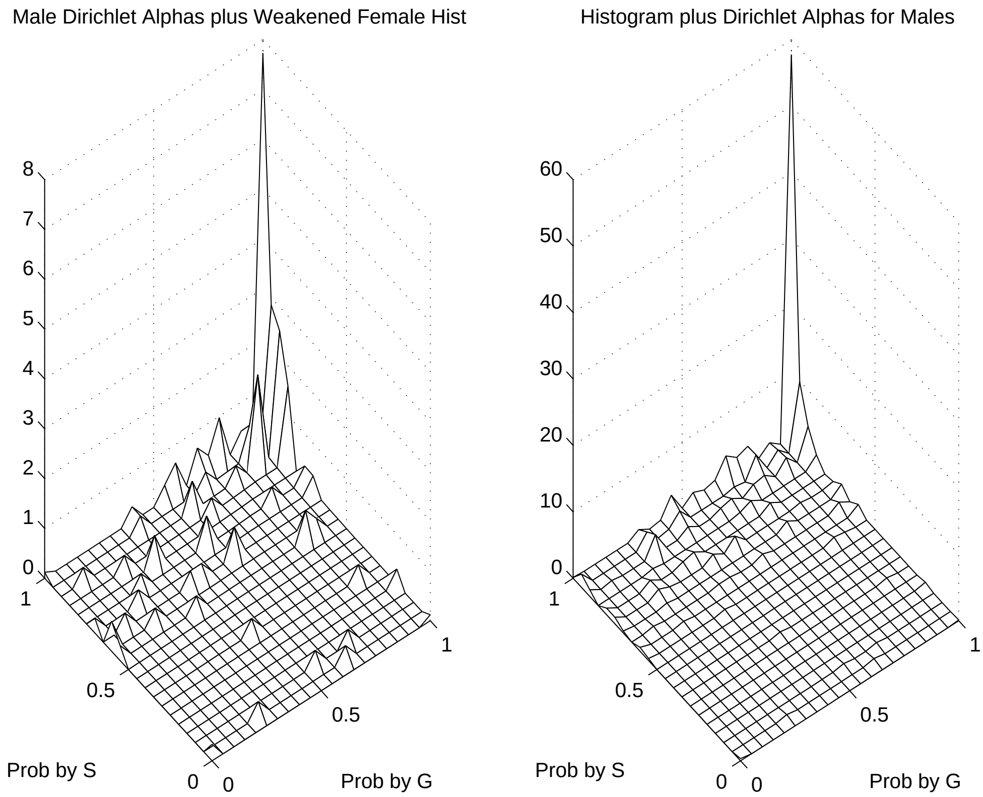

Figure 7: At left are the initial Dirichlet parameters for male skulls augmented by the weakened rotated histogram of assertions regarding female skulls. At right, these augmented prior parameters appear in summation with the male skulls assertion histogram.

5.4 The matrix of denominator probabilities

The computation of the second term in the denominator of equation (1) now requires merely the multiplication of the two matrices of probabilities based on the various possibilities of $\boldsymbol{p}_{201}$ that we described in Sections 5.1.3 and 5.3.

6 Computing the inference from the experts' assertions pertaining to a newly found skull

Our analysis has allowed us now to compute numerical values of equation (1) for various values of $\boldsymbol{p}_{201}$, yielding $P[M_{201}|(\boldsymbol{P}_{201} = \boldsymbol{p}_{201})(\boldsymbol{P_M} = \boldsymbol{p_M})(\boldsymbol{P_F} = \boldsymbol{p_F})]$. This appears graphically on the right-hand-side of Figure 8 for each possible (p_S, p_G) pair of $\boldsymbol{p}_{201}$. Substantively, these results are of interest by comparison with the probabilities $P(M_{201}|(\boldsymbol{P}_{201} = \boldsymbol{p}_{201})$, the probability you assert that the 201^{st} skull is male given only the two experts' assertions, not their assertion records on the known male and female skulls. The matrix of these probabilities appears in the left-hand-side of Figure 8. It is computed merely as the relevant initial Dirichlet α values for the male skulls relative to the sum of these same alpha values for the male skulls plus their corresponding values for the female skulls.

The visual comparison of the two sides of Figure 8 illuminates the inference regarding the usefulness of the two experts' opinions that can be gleaned from the data. The most immediate impression is that the posterior surface on the right-hand-side of Figure 8 is no longer symmetric. Rather, it has shifted higher on the side of S's large probabilities than on G's large probabilities, and it also shifts lower on the section of S's low probabilities than it does for G. These visual impressions will be confirmed and formalised algebraically via the concepts of risk reduction and information gain in the next Section.

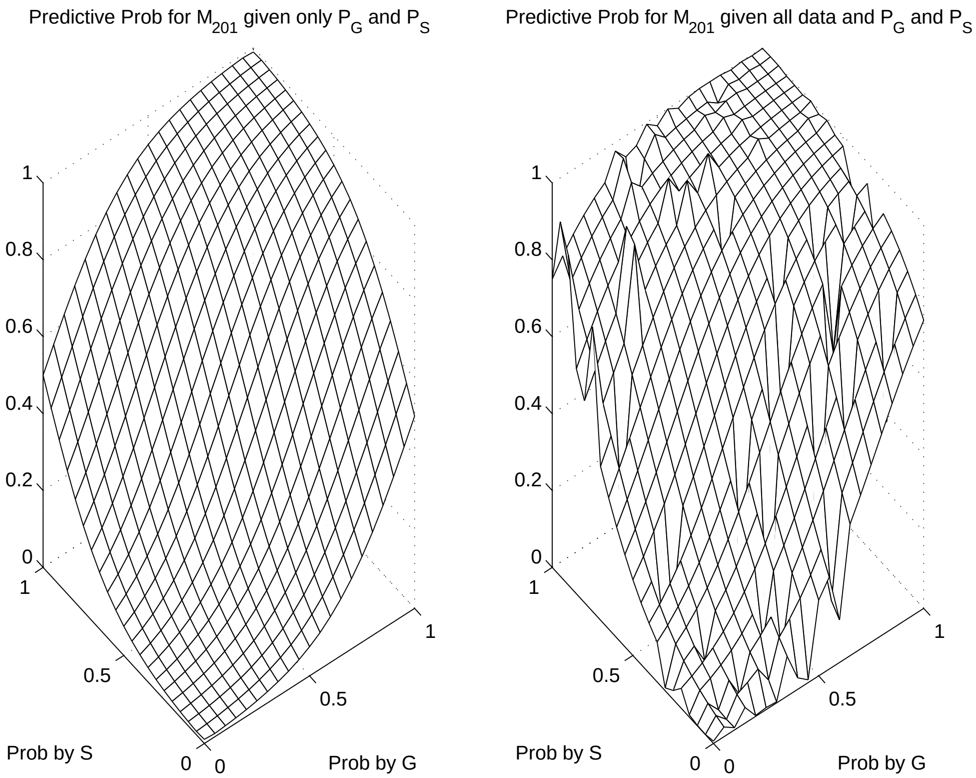

Figure 8: The matrix of conditional predictive probabilities defined by equation (1), given the predictive probabilities that the 201^{st} skull is from a male given the probability assertions by G and by S.

7 Comparing the information contained in each expert's probability assertion to the information contained in both

Our final effort is directed to comparing the inference regarding the 201^{st} skull if we are given only the value of P_S or P_G for that skull, not the both of them, but still conditioning on the evidence of both experts' assertions for the two hundred skulls of known sex.

To begin, Table 3 displays your posterior probability for M_{201} given only the assertion probability by S or by G for that skull, based on the experimental data of the assertions for the 200 known skulls. This Table conveys the inferential use you would make of S's or G's probability assertions regarding M_{201} in formulating your own assertion probability for M_{201}. These probabilities show a fairly regular ascending order based on both experts' assertion levels, but there are some anomalies in both columns as you can see for yourself.

Table 3. Predictive Probability for M_{201} given only the single assertion P_S or P_G for the 201^{st} skull, but given all the experimental data on the assessments for known skulls.

Conditioning p	P_S	P_G
0	0.0199	0.0754
0.0500	0.0710	0.1036
0.1000	0.0835	0.0640
0.1500	0.0765	0.0833
0.2000	0.3669	0.2536
0.2500	0.2052	0.2014
0.3000	0.4319	0.5864
0.3500	0.3641	0.4646
0.4000	0.6914	0.2880
0.4500	0.3667	0.6673
0.5000	0.6830	0.6830
0.5500	0.7541	0.6472
0.6000	0.7232	0.5193
0.6500	0.8271	0.6296
0.7000	0.7493	0.8463
0.7500	0.9133	0.8543
0.8000	0.8083	0.9149
0.8500	0.9752	0.9672
0.9000	0.9659	0.8935
0.9500	0.9794	0.9900
1.0000	0.9973	0.9864

The paper of Cencetti, et al. (1995) presents the numerical results of a procedure designed by Di Bacco (1992) to assess the value of eliciting the probability from a second opinion on the basis of the expected value of the information it would contain. Basically, the procedure presumes that the personal loss incurred by making an incorrect assessment of the sex of a skull is the same, no matter which is the correct sex of the skull. Such a loss would be incurred by asserting probability of male exceeding 1/2 for a female skull, or by asserting probability of male less than 1/2 for a male skull. The numerical computation in the Cencetti report presumes that the assertive assessments by the two experts are independent. The measure of the information gain expected from the elicitation is basically the reduction in risk expected from hearing and using what the expert might say. Applied to expert S for example, the measure reduces to

$$\Delta_S = E(\frac{1/2 - P(M_{201}|(P_S = p)\ and\ data)}{1/2}) \tag{15}$$

for various values of p when $P(M_{201}|(P_S = p)\ and\ data)$ is less than 1/2. If this exceeds 1/2 it is replaced in the formula by $1 - P(M_{201}|(P_S = p)\ and\ data)$. In the case that the probability has not yet been elicited, the computation is based on the expectation of S's probability given the experimental data. You can imagine the modifications to this measure that occur when P_G is elicited instead of or in addition to P_S, and whether the computation is made prior to or posterior to the experimental data on the 200 skulls.

In the present article we have computed the posterior probabilities for the new skull being male, M_{201}, conditioned upon either or both of the experts' elicited probabilities, so we can now apply DiBacco's information measure for each expert to derive the expected gain through reduced risk that would be achieved by asking that expert alone, or asking them both. The computation of the gain expected from asking both experts is made at first based only on your distribution for what the experts might say. These results appear in Table 4.

Table 4. Measures of Information Gain Expected from the Elicitation of One or Both Experts' Probabilities, both prior to observing the experimental data and posterior to the data.

	prior to data	posterior to data
Δ_S	0.7645	0.8456
Δ_G	0.7645	0.7398
$\Delta_{S,G}$	0.8152	0.8900

However, we can also derive the expected information gain from asking a second expert after having heard the response of the first expert. In this case, the information measure (15) is modified to

$$\Delta_{S|(P_G=p*)\ and\ data} = E(\frac{P(M_{201}|(P_G = p*)\ and\ data) - P(M_{201}|(P_S = p)(P_G = p*)\ and\ data)}{P(M_{201}|(P_G = p*)\ and\ data)}) \tag{16}$$

with similar modifications as to equation (15) when the various conditional probabilities in equation (16) exceed 1/2 or are less than 1/2. The expectation is computed with respected to the conditional distribution of S's probability assertion, given the value of G's assertion as $p*$. The results of these computations appear in Table 5.

Table 5. Measures of Information Gain Expected from the additional elicitation of the second expert's probability given that you know the assertion of the first. The conditioning assertion value for the expert is printed in column 1; the second and third columns present the measures of expected gain from asking each of S and G as the second expert, knowing the conditioning value of the first's assertion, but prior to the data on the 200 skulls. Columns four and five present these measures again, but conditioned on the experimental data. Column six presents the difference in the gains expected from asking S in addition to G less the gain expected from asking G in addition to S, based on the same elicited probability response from G and from S in the first elicitation.

Conditioning p	$\Delta_{S\|(P_G=p)}$	$\Delta_{G\|(P_S=p)}$	$\Delta_{S\|(P_G=p)\ and\ data}$	$\Delta_{G\|((P_S=p)\ and\ data}$	Difference
0	0	0	0.5725	0.4513	0.1212
0.0500	0.0106	0.0106	0.0610	0.2590	-0.1979
0.1000	0.0302	0.0302	0.1865	0.1418	0.0447
0.1500	0.0575	0.0575	0.0584	0.0262	0.0322
0.2000	0.0923	0.0923	0.7258	0.4235	0.3023
0.2500	0.1344	0.1344	0.4666	0.3379	0.1287
0.3000	0.1841	0.1841	0.8943	0.5595	0.3348
0.3500	0.2419	0.2419	0.6399	0.2685	0.3714
0.4000	0.3089	0.3089	0.7086	0.2805	0.4281
0.4500	0.3862	0.3862	0.8745	0.5680	0.3064
0.5000	0.4756	0.4756	0.2730	0.2730	0
0.5500	0.3862	0.3862	0.4524	0.3688	0.0836
0.6000	0.3089	0.3089	0.6396	0.4015	0.2381
0.6500	0.2419	0.2419	0.5357	0.0410	0.4947
0.7000	0.1841	0.1841	0.0958	0.5169	-0.4211
0.7500	0.1344	0.1344	0.2701	0	0.2701
0.8000	0.0923	0.0923	0	0	0
0.8500	0.0575	0.0575	0	0	0
0.9000	0.0302	0.0302	0.0617	0	0.0617
0.9500	0.0106	0.0106	0	0	0
1.0000	0	0	0	0	0

It is evident now just how systematically the information provided by S's probability assertions is to be valued more highly than that provided by G's. At only two probability assertion levels (.05 and .70) by expert S is the amount of information expected from the elicitation of G's probability greater than that expected from eliciting S's probability after G had asserted the same level as S.

8 Substantive Conclusions

We began this study presuming that you regard the two experts' probability assessments exchangeably, regardless of whether the skull in question is male or female. Moreover we assessed exchangeably the sequence of probability pairs by the two experts conditionally separately on the skulls being male or being female. The histograms of probability assertions turned out to be far from symmetric, however, both for male skulls and for female skulls. The assertions of expert S appeared to be more informative to you in forecasting than did those of G. Based on the partial exchangeability structure on the assertion pairs, we learned just how much more to value the ellicitation of a probability from S than from G. Assessment of the gain in value from S's assertions relative to G's were formalised in the decision theoretic expectational measure of risk reduction.

9 Methodological Comments

While the substantial results of this analysis are rather conclusive, our analysis should be recognised in its technical limitations which can be investigated further.

In the first place the use of the Dirichlet mixture distributions allows that the computed results on forecasting probabilities $P(M_{201}|data)$ are somewhat bumpy, a feature exacerbated whenever "holes" of probabilty pair categories were not observed occurring in the histogram. It would be worthwhile to compare the results computed here to those based on more integrated forms of prior distributions such the tractable forms of special functions or the logistic normal distribution on the unit-simplex, which allow positive covariance between neighboring categories, rather than uniformly negatively covariances between categories as is required by the Dirichlet.

A second limitation should be recognised in the sharp specification of the weakening parameter value of .3 in formulating the likelihood function (information transfer function) detailing how observations for male skulls inform us about the experts' judgments on female skulls, and vice-versa. (This refers to Sections 4.2 and 5.2.) It would be interesting to soften this parameter specification by mixing it over other reasonable values to study its influence.

Finally, the $0-1$ loss structure on forecasting errors (no loss if the skull is diagnosed appropriately with a probability exceeding $1/2$ for male skulls and less than $1/2$ for female skulls, and an equal sized loss for misdiagnosis of either sexed skull should be recognised as appropriate to the specific problem studied here. However for other problems that are structurally similar, such as medical diagnosis, other forms of loss functions should be admitted, allowing even a different specification of loss for the doctor and for the patient.

Acknowledgments: Joint work on this project was begun by correspondence while Frank Lad was supported as a research visitor to University of Perugia during June, 1998.

References

Cencetti, S., D'Amore, G., Di Bacco, M., and Pacciani, E. (1995) Visual Diagnosis of the Sex of a Human Skull, and the Wagers of de Finetti, *Atti XI Congresso Degli Antropologi Italiani*, 601-608.

Di Bacco, M. (1992) *La diagnosi come regola di decisione*, Bologna: Pittagora.

Good, I.J. (1965) *The Estimation of Probabilities: an essay on modern Bayesian methods*, Cambridge, Massachusetts: MIT Press.

Lad, F. (1996) *Operational Subjective Statistical Methods: a mathematical, philosophical, and historical introduction*, New York: John Wiley.