Bayesian tests for correlated binary choices

Daniel Navarro

October 13, 2008

One very basic experimental data analysis situation arises when two decision makers are presented with a collection of N objects, and asked to divide them into two distinct subsets, generally a chosen or "in" group, and a non-chosen or "out" group. Suppose that the first decision maker assigns a total of k_1 of the N objects to one class, and $N - k_1$ of the objects to the other class (or alternatively, chooses k_1 of the objects and leaves the others). The second decision-maker then selects k_2 of the objects, of which s also belong to the set of k_1 objects chosen by the first decision-maker. The result is that we have a data set in which n_{11} objects are chosen by both decision-makers, n_{00} objects are chosen by neither, n_{10} are selected by the first but not the second decision-maker, and n_{01} are chosen by the second but not the first; where $n_1 1 = s$, $n_{10} = k_1 - s$, $n_{01} = k_2 - s$ and $n_{00} = N - k_1 - k_2 + s$. Letting

$$\mathbf{N} = \left[\begin{array}{cc} n_{11} & n_{10} \\ n_{01} & n_{00} \end{array} \right] \qquad (1)$$

denote these data, and (for the sake of simplicity) assuming that a person's choices can be approximated by a set of independent Bernoulli trials (i.e., he or she chooses a particular object with probability θ, independently for each object). In those circumstances, there would seem to be two substantively important questions to answer:

1. Do the two decision-makers have the same choice probability? That is, if we let θ_1 denote the probability with which decision-maker 1 chooses an item, and let θ_2 denote the corresponding rate for decsion-maker 2, then should we conclude that $\theta_1 = \theta_2$, or $\theta_1 \neq \theta_2$ in light of the observed data $\mathbf{N}$?

2. Are the two decision-makers correlated in some way? This is a slightly trickier question to formalize sensibly, but the qualitative idea is to ask whether the number of shared choices $n_{11}+n_{00}$ is too large (or small) to be plausible if the decision-makers are making decisions independently of one another. More precisely, if $\theta_2^{(0)}$ denotes the probability that decision-maker 2 chooses a particular object conditional on the knowledge that decision-maker 1 did not, and $\theta_2^{(1)}$ denotes the corresponding probability if decision-maker did choose that object, then should the data $\mathbf{N}$ lead us to conclude that $\theta_2^{(0)} = \theta_2^{(1)}$ or $\theta_2^{(0)} \neq \theta_2^{(1)}$?

In practice, question #2 is likely to be of deeper interest than question #1, but both seem useful, and in any case it seems sensible to try to disentangle the two from one another. So, if we let C_1 and C_2 denote the two sets of "chosen" objects, we end up with four statistical models of interest:

- $\mathcal{M}_0$: *independent and identically-distributed processes.* In this case, we assume that the two object sets are generated using the same choice probability θ, and moreover that (assuming a known value of θ) the observation that a particular object belongs to C_1 does not influence the likelihood that it also belongs to C_2 (or vice versa).

- $\mathcal{M}_1$: *independent processes with different distributions.* In this case, we assume that the choice probabilities differ between the two sets, (i.e., $\theta_1 \neq \theta_2$), but as before, there are no dependencies between the two.

- $\mathcal{M}_2$: *correlated processes with identical distributions.* In this case, we still assume that the two object sets share the same choice rate θ, but we now introduce a separate parameter ϕ that describes the probability that C_1 and C_2 agree on any particular object (i.e., both in or both out). Note that under the *i.i.d.* model $\mathcal{M}_0$, this agreement parameter would (in effect) be fixed at $\phi = \theta^2 + (1-\theta)^2$.

- $\mathcal{M}_3$: *correlated processes with different distributions.* Finally, we have a model in which the two sets have different rate parameters θ_1 and θ_2, but can *also* agree with each other at a rate ϕ that differs from the independence prediction (in this case $\phi = \theta_1\theta_2 + (1-\theta_1)(1-\theta_2)$).

To state the obvious, some of these models are nested inside one another. Specifically, $\mathcal{M}_0 \subset \mathcal{M}_1 \subset \mathcal{M}_3$ and $\mathcal{M}_0 \subset \mathcal{M}_2 \subset \mathcal{M}_3$, but there is no nesting relationship between $\mathcal{M}_1$ and $\mathcal{M}_2$.

Intuitively, it is clear that each of the four models provides different constraints on the probabilities of the four possible events whose frequencies are counted in the data matrix $\mathbf{N}$ (i.e., 11, 01, 10 and 00). We denote these four probabilities by λ_{11}, λ_{10}, λ_{01} and λ_{00} and note that (as always) each such probability lies between 0 and 1, and all four must sum to 1. In all four cases we may describe the joint probability of C_1 and C_2 as follows:

$$P(C_1, C_2|\Lambda, \mathcal{M}) = {\lambda_{11}}^{n_{11}} {\lambda_{10}}^{n_{10}} {\lambda_{01}}^{n_{01}} {\lambda_{00}}^{n_{00}}, \tag{2}$$

where the matrix Λ is defined such that the first row corresponds to the events in which a particular object belongs to C_1 and the first column corresponds to events in which the object belongs to C_2. That is,

$$\Lambda = \begin{bmatrix} \lambda_{11} & \lambda_{10} \\ \lambda_{01} & \lambda_{00} \end{bmatrix}. \tag{3}$$

Having done so, we can express the four models in terms of a set of expressions for Λ. For $\mathcal{M}_0$ and $\mathcal{M}_1$ these are straightforward, but for $\mathcal{M}_2$ and $\mathcal{M}_3$ they are a little less obvious and are explained later. The expressions in question are:

$$\Lambda(\mathcal{M}_0, \theta) = \begin{bmatrix} \theta^2 & \theta(1-\theta) \\ \theta(1-\theta) & (1-\theta)^2 \end{bmatrix} \tag{4}$$

$$\Lambda(\mathcal{M}_1, \theta_1, \theta_2) = \begin{bmatrix} \theta_1\theta_2 & \theta_1(1-\theta_2) \\ (1-\theta_1)\theta_2 & (1-\theta_1)(1-\theta_2) \end{bmatrix} \tag{5}$$

$$\Lambda(\mathcal{M}_2, \theta, \phi) = \frac{1}{2}\begin{bmatrix} \phi - 1 + 2\theta & 1-\phi \\ 1-\phi & \phi + 1 - 2\theta \end{bmatrix} \tag{6}$$

$$\Lambda(\mathcal{M}_3, \theta_1, \theta_2, \phi) = \frac{1}{2}\begin{bmatrix} \theta_1 + \theta_2 + \phi - 1 & -\theta_1 + \theta_2 - \phi + 1 \\ \theta_1 - \theta_2 - \phi + 1 & -\theta_1 - \theta_2 + \phi + 1 \end{bmatrix}, \tag{7}$$

where the constraints on the rate parameters are $0 \leq \theta, \theta_1, \theta_2 \leq 1$ in all cases. For the agreement parameter ϕ, in order to ensure that the appropriate constraints on λ are met, we have $|1-2\theta| \leq \phi \leq 1$ for model $\mathcal{M}_2$ and $|1-\theta_1-\theta_2| \leq \phi \leq 1$ for model $\mathcal{M}_3$. Note that in the two models that assume equal rates (i.e., $\mathcal{M}_0$ and $\mathcal{M}_2$), the matrix Λ is constrained such that

$$\lambda_{11} + \lambda_{10} = \theta \quad (8)$$
$$\lambda_{11} + \lambda_{01} = \theta \quad (9)$$
$$\lambda_{00} + \lambda_{10} = 1-\theta \quad (10)$$
$$\lambda_{00} + \lambda_{01} = 1-\theta, \quad (11)$$

and hence θ can indeed be interpreted as a choice probability in both cases. For the models that assume different rates (i.e., $\mathcal{M}_1$ and $\mathcal{M}_3$) we have the equivalent constraint:

$$\lambda_{11} + \lambda_{10} = \theta_1 \quad (12)$$
$$\lambda_{11} + \lambda_{01} = \theta_2 \quad (13)$$
$$\lambda_{00} + \lambda_{10} = 1-\theta_2 \quad (14)$$
$$\lambda_{00} + \lambda_{01} = 1-\theta_1, \quad (15)$$

and once again, θ_1 and θ_2 can be treated as choice probabilities for the two object sets. Finally, for models in which the agreement rate ϕ is treated as a free parameter (i.e., $\mathcal{M}_2$ and $\mathcal{M}_3$), note that

$$\lambda_{11} + \lambda_{00} = \phi \quad (16)$$
$$\lambda_{10} + \lambda_{01} = 1-\phi, \quad (17)$$

which is indeed the relationship that we would expect if ϕ is to act as an agreement-rate. Moreover, we can now be explicit about the nature of the nesting relationships by stating that:

$$\mathcal{M}_0 \equiv \mathcal{M}_1 : \theta_1 = \theta_2 \quad (18)$$
$$\mathcal{M}_0 \equiv \mathcal{M}_2 : \phi = \theta^2 + (1-\theta)^2 \quad (19)$$
$$\mathcal{M}_1 \equiv \mathcal{M}_3 : \phi = \theta_1\theta_2 + (1-\theta_1)(1-\theta_2) \quad (20)$$
$$\mathcal{M}_2 \equiv \mathcal{M}_3 : \theta_1 = \theta_2. \quad (21)$$

Noting that each of these four models is parametrized in a manner that incorporates all the relevant assumptions that we want to make (e.g., note that $\mathcal{M}_3$ is actually equivalent to a saturated model, but written in terms of the three interesting marginal probabilities θ_1, θ_2 and ϕ), it seems reasonable for each model to place a uniform prior across the (valid) parameter space, and thus define the marginal probability assigned to the pair of object sets as follows:

$$P(C_1, C_2|\mathcal{M}) = \frac{1}{V(\mathcal{M})} \int_{\Lambda \in \mathcal{M}} P(C_1, C_2|\Lambda)\, d\Lambda \quad (22)$$

where $V(\mathcal{M})$ denotes the volume of the valid parameter space for the model. With that in mind, it is straightforward to use the parameter constraints to show that

$$V(\mathcal{M}_0) = 1 \quad (23)$$

$$V(\mathcal{M}_1) = 1 \quad (24)$$
$$V(\mathcal{M}_2) = 1/2 \quad (25)$$
$$V(\mathcal{M}_3) = 2/3, \quad (26)$$

though in the case of the saturated model $\mathcal{M}_3$ it will turn out that for the purposes of deriving marginal probabilities it is easier to operate in the λ parameterization in which the volume is 1/3.

Not surprisingly, the models without correlated decision-makers are equivalent to standard beta-binomial models, and are hence very tractable. For $\mathcal{M}_0$ we have:

$$\begin{aligned} P(C_1, C_2|\mathcal{M}_0) &= \int_0^1 \theta^{2n_{11}+n_{10}+n_{01}}(1-\theta)^{n_{10}+n_{01}+2n_{00}}\ d\theta & (27) \\ &= B(2n_{11}+n_{10}+n_{01}+1, n_{10}+n_{01}+2n_{00}+1) & (28) \\ &= \frac{(2n_{11}+n_{10}+n_{01})!\ (n_{10}+n_{01}+2n_{00})!}{(2N+1)!}, & (29) \end{aligned}$$

where $B(\cdot,\cdot)$ is the beta function; and since the arguments are always positive integers this reduces to simple ratios of factorial functions, as illustrated in the third line. Similarly, $\mathcal{M}_1$ gives an integral that factorizes into two beta functions in the the way one might expect:

$$\begin{aligned} P(C_1, C_2|\mathcal{M}_1) &= \int_0^1\int_0^1 \theta_1^{n_{11}+n_{10}}(1-\theta_1)^{n_{00}+n_{01}}\theta_2^{n_{11}+n_{01}}(1-\theta_2)^{n_{00}+n_{10}}\ d\theta_1\ d\theta_2 & (30) \\ &= \int_0^1 \theta_1^{n_{11}+n_{10}}(1-\theta_1)^{n_{00}+n_{01}}d\theta_1 \int_0^1 \theta_2^{n_{11}+n_{01}}(1-\theta_2)^{n_{00}+n_{10}}\ d\theta_2 & (31) \\ &= \frac{(n_{11}+n_{10})!\ (n_{00}+n_{01})!}{(N+1)!}\ \frac{(n_{11}+n_{01})!\ (n_{00}+n_{10})!}{(N+1)!}. & (32) \end{aligned}$$

For the correlated but equiprobable choices model, $\mathcal{M}_2$, the marginals probabilities are a little more complex, and so need to be handled with more care. Noting that we can rewrite the parameter constraints as $\phi \in [0,1]$ and $\theta|\phi \in [\frac{1-\phi}{2}, \frac{1+\phi}{2}]$, the probability in which we are interested is the solution to

$$P(C_1, C_2|\mathcal{M}_2) = 2\int_0^1\int_{\frac{1-\phi}{2}}^{\frac{1+\phi}{2}} \left(\theta - \frac{1-\phi}{2}\right)^{n_{11}} \left(\frac{1+\phi}{2} - \theta\right)^{n_{00}} \left(\frac{1-\phi}{2}\right)^{n_{10}+n_{01}} d\theta d\phi. \quad (33)$$

The fact that the third term depends on ϕ but not θ is convenient, since we can remove it from the inner integral as a constant term, and hence start with the following:

$$\begin{aligned} g(\phi) &= \int_{\frac{1-\phi}{2}}^{\frac{1+\phi}{2}} \left(\theta - \frac{1-\phi}{2}\right)^{n_{11}} \left(\frac{1+\phi}{2} - \theta\right)^{n_{00}} d\theta & (34) \\ &= \int_0^\phi u^{n_{11}}(\phi - u)^{n_{00}}\ du & (35) \\ &= \phi^{n_{11}+n_{00}+1}\int_0^1 v^{n_{11}}(1-v)^{n_{00}}\ dv & (36) \\ &= \phi^{n_{11}+n_{00}+1}\frac{n_{11}!\ n_{00}!}{(n_{11}+n_{00}+1)!}, & (37) \end{aligned}$$

where the second line is the result of the substitution $u = \theta - \frac{1-\phi}{2}$, and the third line is obtained by substituting $v = u/\phi$ and then moving constant terms outside the integral. Since the resulting integral describes a beta function with integer-valued arguments, the final line follows. Hence we have,

$$\begin{aligned} P(C_1, C_2|\mathcal{M}_2) &= 2\frac{n_{11}!\ n_{00}!}{(n_{11}+n_{00}+1)!}\int_0^1 \left(\frac{1-\phi}{2}\right)^{n_{10}+n_{01}} \phi^{n_{11}+n_{00}+1}\ d\phi & (38)\\ &= \frac{n_{11}!\ n_{00}!}{(n_{11}+n_{00}+1)!\ 2^{n_{10}+n_{01}-1}}\int_0^1 (1-\phi)^{n_{10}+n_{01}}\phi^{n_{11}+n_{00}+1}\ d\phi & (39)\\ &= \frac{n_{11}!\ n_{00}!}{(n_{11}+n_{00}+1)!\ 2^{n_{10}+n_{01}-1}}\ \frac{(n_{10}+n_{01})!\ (n_{11}+n_{00}+1)!}{(n_{10}+n_{01}+n_{11}+n_{00}+2)!} & (40)\\ &= \frac{n_{11}!\ n_{00}!\ (n_{10}+n_{01})!}{(n_{11}+n_{00}+1)!\ (N+2)!\ 2^{n_{10}+n_{01}-1}}. & (41) \end{aligned}$$

For the last model, $\mathcal{M}_3$ in which the decision-makers are correlated and have different choice probabilities, it is worth observing firstly that the model has the same degrees of freedom (i.e., 3) as a saturated model in which Λ is allowed to vary arbitrarily, and by construction satisfies the same constraints. Specifically, noting that $\lambda_{00} = 1 - \lambda_{11} - \lambda_{10} - \lambda_{01}$, it is helpful to observe that we can map from one to the other via the straightforward linear transformation:

$$2\begin{bmatrix} \lambda_{11} \\ \lambda_{10} \\ \lambda_{01} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & -1 \end{bmatrix}\begin{bmatrix} \theta_1 \\ \theta_2 \\ \phi \end{bmatrix} + \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}. \quad (42)$$

In other words, the parameters of $\mathcal{M}_3$ are actually just a rigid rotation, reflection, translation and dilation of those used in the saturated model. Without any nonlinearities involved, a uniform prior on valid choices of $(\theta_1, \theta_2, \phi)$ is also uniform on all valid choices of $(\lambda_{11}, \lambda_{10}, \lambda_{01})$. Hence,

$$\begin{aligned} P(C_1, C_2|\mathcal{M}_3) &= 3\int_\Lambda {\lambda_{11}}^{n_{11}}{\lambda_{10}}^{n_{10}}{\lambda_{01}}^{n_{01}}{\lambda_{00}}^{n_{00}}\ d\Lambda & (43)\\ &= 3\ \frac{n_{11}!\ n_{10}!\ n_{01}!\ n_{00}!}{(N+1)!}. & (44) \end{aligned}$$

At this point, it is straightforward to specify the posterior probability associated with the d-th model via Bayes' theorem:

$$P(\mathcal{M}_d|C_1, C_2) = \frac{P(C_1, C_2|\mathcal{M}_d)\ P(\mathcal{M}_d)}{\sum_{j=0}^3 P(C_1, C_2|\mathcal{M}_j)\ P(\mathcal{M}_j)}, \quad (45)$$

where one would generally choose a uniform prior over models, $P(\mathcal{M}_j) = 1/4$ for all j.