The Institute for Perception

For more technical reports from The Institute for Perception or for information about our short courses, please visit www.ifpress.com or email us at mail@ifpress.com

Reprinted from IFPress (1998) 1(2)2,3

Replicated Difference and Preference Testing with Applications to Claims Support

Jian Bi and Daniel M. Ennis

Background:

As part of an effort to compare a new deodorant with a competitor's product, 10 experienced judges evaluate the two products on the left and right arms of 30 subjects. In a counterbalanced design in which the two products are alternately placed on the left or right arms within a subject, each judge reports the least malodorous arm.

In this example, there is the possibility that there may be trial-to-trial differences in malodor. The chemical reactions of subjects to the deodorants may differ so that one product may be less effective on some subjects than others. This type of result leads to the need to account for inter-trial variability either because of the need to provide defensible claims in the face of inter-trial variation or because this type of variation is of fundamental interest itself, as may occur when one is identifying preference segments.

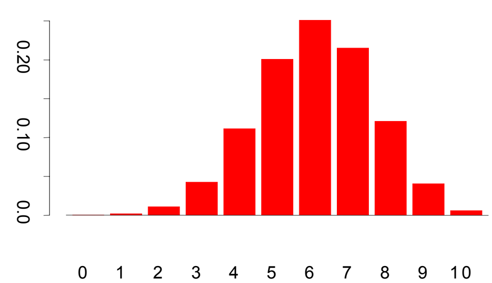
In a recent paper[1], we discussed the use of the Beta-Binomial (BB) model for replicated difference and preference tests to deal with exactly the type of problem just discussed. The BB model is a natural extension and generalization of the binomial model. In the BB model, two sources of variation (inter-trial and intra-trial variation) are accounted for and two parameters are used to fit the data collected from a replicated difference or preference test. The BB model provides a much better fit to many data sets than the binomial model. This model should improve the validity of sensory difference and preference tests where inter-trial variation occurs.

Inapplicability of the Binomial Model - Sensory difference and consumer preference tests are widely used in sensory and consumer studies. The traditional model for these experiments is the binomial distribution. The binomial model is valid under the assumption that there is only one source of variation in the data because the choice probability (preference or probability of a correct response, for instance) is assumed to be constant from trial-to-trial. However, this assumption is almost always violated in practice. If variation due to inter-trial differences cannot be ignored, then there are two sources of variation: variation due to judgments and variation due to trials. Returning to the earlier example about the treatment of malodor, if the judges are homogeneous, but subject reactions to deodorants are not, then variation within a trial (the assessment of one subject) may be very different from variation across subjects. This extra variation is called *overdispersion* and the BB model is designed to account for this extra variation.

False Claims of Product Superiority - When there is more than one source of variance in a difference or preference test, the variability in the data may exceed binomial variability. If we still use the binomial model, an underestimate of the standard error can be obtained and so a seriously misleading conclusion may be drawn about the superiority of one product over another. A simulation study based on actual product testing experience was reported by Ennis and Bi[1]. The results of the study showed that, for overdispersed binomial data, the Type I error may be 0.44 when the experimenter thinks it is 0.05. This means that in a test to demonstrate product superiority, a product may be declared significantly better at the 5% level, when its superiority can only be demonstrated at the 44% level! It is clear that the traditional binomial model should not be applied to overdispersed binomial data and the current approach to analyzing difference and preference tests needs to be revised.

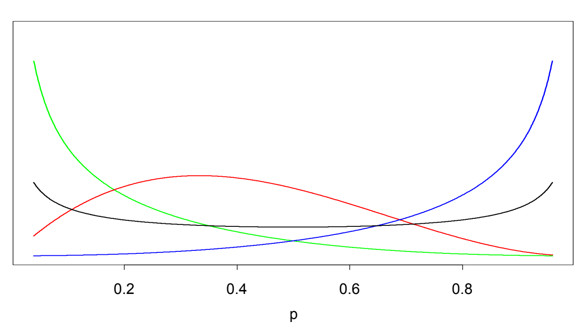
Binomial, Beta and Beta-Binomial Models - The binomial distribution is a discrete distribution and when applied to choice data models the probability that a particular choice outcome will occur. For instance in the malodor example, the binomial distribution models the probability that of the 10 judges evaluating a particular subject, 0, 1, 2,..., 10 of them will choose your competitor's brand as least malodorous. If we pool data across subjects, we assume that each subject reacts to both products in the same way. But suppose that when your competitor's product is used on subject 1 there is a 60% chance of being less malodorous but that on subject 2, there is only a 20% chance. By combining data we are mixing data binomially distributed in different ways and cannot assume that the combined data follows a binomial distribution. Figure 1 shows what the binomial distribution looks like for $n = 10$ and $p = 0.6$ and shows the probability of 0 out of 10 to 10 out of 10 correct responses.

Figure 1. Binomial: n = 10, p = .6



As we consider the possibility that p may change from trial (subject) to trial, how is p distributed? One very general possibility is to consider that p follows a beta distribution. The beta distribution allows a broad variety of shapes for the distribution of p, four of which are shown in Figure 2.

Figure 2. Shapes of the Beta Distribution

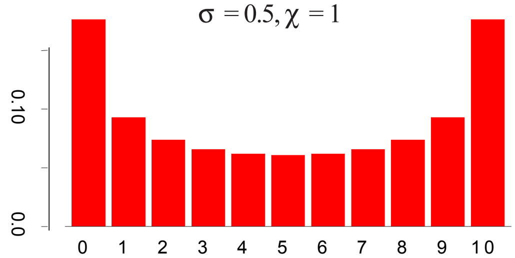


In one case (the black line), the most likely values of p are close to 0 or 1. Applied to the deodorant problem, this means that your product is more effective on some subjects and worse on others than your competitor's product.

Combining the binomial and beta distributions produces the beta-binomial distribution, one example is in Figure 3. In this distribution, we assume that a binomial distribution with $p = 0.6$ applies within a trial and that as we move from trial to trial, the p's follow a beta distribution, like the black line in Figure 2. In order to estimate the p for a particular trial, we need replications. In order to estimate the parameters of the beta distribution, we need more than one replicated trial. Data fit by a beta-binomial are always in the form of replications and trials. The meaning of these two terms varies depending on the problem.

Figure 3. Beta-Binomial: n = 10

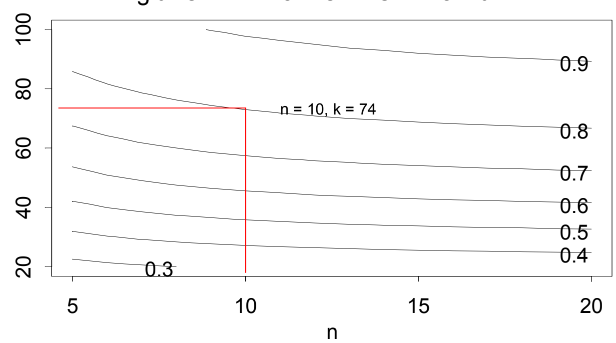
$\sigma = 0.5, \chi = 1$



Fitting the BB Model to Data - Two parameters are needed to fit the BB model. These parameters are σ and χ and they measure the mean and spread of the distribution of the choice probability, p. If χ is zero, the BB reduces to the binomial distribution with a single parameter p. The BB parameters are fitted to replicated difference and preference tests using the method of maximum likelihood. In the deodorant claim example, the estimate of σ was 0.60 and the estimate of χ was 0.47. A comparison of the binomial and BB models showed that the BB model fits the data significantly better. This means that χ cannot be assumed to be zero. When the binomial was used to compare products on the combined data from subjects and judges, your competitor's product was significantly better ($p < 0.001$). However, when we used the BB model and accounted for inter-trial variability we found that the products were not significantly different at ζ # .05. The apparent superiority of your competitor's brand may have been due to overdispersion.

How Many Trials and Replications? - In the malodor example, 30 trials with 10 replications per trial were used. Analysis showed that the two products did not differ significantly at the 5% level using the BB model. How many replications and trials would be needed to be 80% sure that if the real σ # and χ were 0.6 and 0.47 the result would be significant? Figure 4 shows contours of equal power for values of n (replications) and k (trials) when $\sigma = 0.6$ and χ # 0.47. It can be seen that if 10 judges were used, the number of subjects would need to be increased to 74. Other combinations of n and k can be used to achieve 80% power. The choice would depend on the relative cost of n and k.

Figure 4. Power vs. *n* and *k*



Other Applications - In a separate project, you have an interest in determining whether a new deodorant has a broader appeal to young males than your current product or whether there are latent preference segments. These segments may differ in their preferences for the new or current deodorant. A sample of 300 young male consumers provide triplicate forced choice preference responses to pairs of the current and new deodorant.

In this preference test, some consumers may prefer one deodorant over the other and other consumers may have the opposite preference. Some may be indifferent. We can use the BB model to test for the existence of latent preference groups very much the same way that we used it to study subject effects in the malodor example. In the preference test, the trials are consumers and replications occur within a consumer. In this case if the BB is significantly different from the binomial, then we would conclude that there are latent preference groups. Another way of expressing this is that χ would be greater than zero.

There are many sources of inter-trial variance. Differences in experimental material (such as the subjects in the malodor example), individual preferences, individual sensory acuity, manufacturing locations and time are all sources of inter-trial variance. In some cases, we may use the BB to find interesting latent groups. In other cases, we use the BB to deal with extra nuisance variation that we would like to work toward eliminating. If inter-trial variation could be eliminated, we would be justified in using the traditional binomial model.

Conclusion - Product superiority claims in terms of preference or the intensity of some attribute (product A is less malodorous than product B, for instance) are often based on binomial tests of choice proportions. If overdispersion exists, then a superiority claim could be refuted on the basis of an erroneous Type I error. Conversely, a superiority claim based on the BB model can be defended. The BB model may be used as an offensive and defensive weapon in claims support and refutation.

1. Ennis, D. M. and Bi, J. 1998. The Beta-Binomial model: Accounting for inter-trial variation in replicated difference and preference tests. *Journal of Sensory Studies.* *In press.*