U.S. DEPARTMENT OF COMMERCE
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
NATIONAL WEATHER SERVICE
SYSTEMS DEVELOPMENT OFFICE
TECHNIQUES DEVELOPMENT LABORATORY

TDL OFFICE NOTE 81-8

A BETA CLASSIFICATION MODEL

Robert G. Miller and Donald L. Best

November 1981

A BETA CLASSIFICATION MODEL

Robert G. Miller and Donald L. Best

1. INTRODUCTION

This paper introduces a new classification procedure using beta probability density functions (pdf) to compute threshold probability values. The classification problem is this: given a probability distribution for the occurrence of an event, how does one make a categorical decision? In decision theory, such classifications are made under the control of some underlying utility function. The decisionmaker may then choose categorical selections that either maximize some gain or minimize some loss. In weather forecasting, utility is usually some verification statistic which is to be optimized (e.g., percent correct, hits, threat score, or skill score). This paper departs from the decision-theoretic approach by using a much simpler, albeit approximate, procedure incorporating threshold probabilities and a successive pair-wise comparison test. Using threshold probability values is not new; however, what has yet to be achieved is a threshold model that would provide a wide range of desired categorical responses that in turn control the verification statistic. The Beta classification model presented here accomplishes this objective. This procedure can maximize threat score, and can produce a marginal distribution balance (i.e., the number of forecast events equals the number of events observed).

2. REGRESSION PROBABILITY MODEL

The first step in the classification problem is to establish a function which can provide event probabilities. Linear regression of a selected dependent variable onto the desired independent variables accomplishes this. Here we define the independent variables, or predictors, as X_1, X_2, X_3, $...X_K$. We represent the dependent variable, the predictand, as Y; its estimate is $\hat{Y}$. The desired probability model is then:

$$\hat{Y} = d_0 + d_1X_1 + d_2X_2 + ... + d_KX_K \tag{1}$$

The solution of the coefficients (d_i's) is obtained through regular multiple regression techniques with or without screening. The definition of the predictand values is absolutely necessary. The event must be exhaustive and mutually exclusive of all other possible events. If the event over the developmental data sample is observed to fall within this preselected definition of occurrence, the Y-value is assigned a "1"; otherwise it is assigned a "0." The Y-data are, therefore, binary variables representing whether the event occurred or not. The predictor variables may be either scalar, binary, or some combination of either.

Introduction of a binary predictand Y into a least-squares linear regression program produces a model which then will estimate probabilities of future events. Since there are many possible combinations of the predictors, the probability model produces a range of probability values. These values can be grouped according to verification and examined through their frequency distributions as illustrated in figure 1. This figure also shows several features that are important to the understanding of the following discussion.

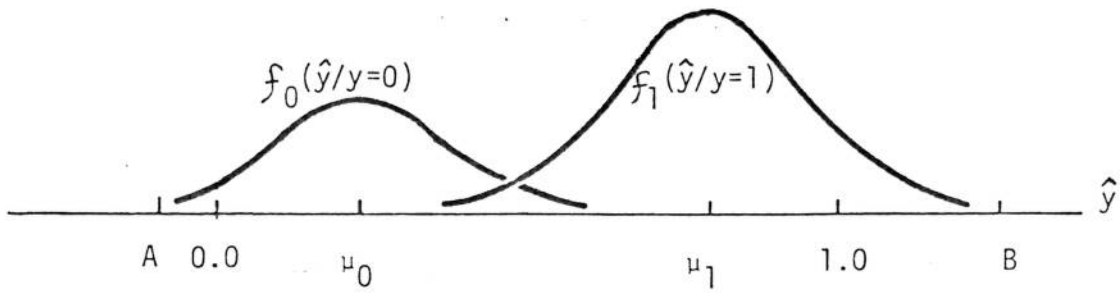

Figure 1.--Schematic depiction of the probability-value ($\hat{y}$) distributions when Y=1 and Y=0. The μ values represent distribution means.

3. CLASSIFICATION BY THRESHOLDING

There are two well defined clusters of probability values grouped into occurrence $f_1(Y/Y=1)$ and non-occurrence $f_0(Y/Y=0)$ of the event. The respective means of these distributions are μ_1 and μ_0. Some values fall outside the (0,1) range. The (A,B) interval represents the lower and upper bounds of possible probability values. The property that the "probability" estimate can fall outside the (0,1) range is more a nuisance to the classification problem than a mystical fact.

This property is actually of little concern, because the two distributions' overlapping values are of greater concern to us than the out-of-range values. Figure 2 portrays the overlapping problem with a given threshold value, p*.

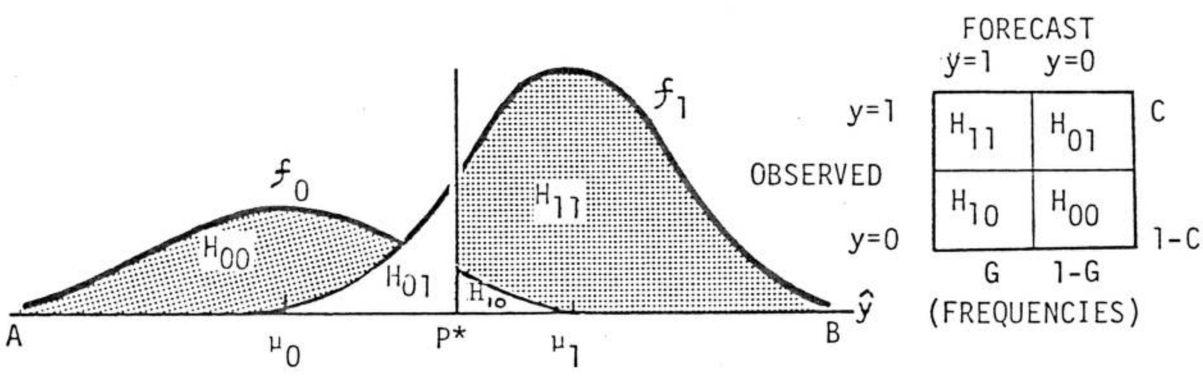

Figure 2.--Illustration of how a chosen p* (threshold probability) would control the frequency of positive classifications. A verification table is also shown. Subscripts on densities H_{ij} represent forecast category i and verified category j.

Since these two distributions describe the forecast model's response in an expected sense, we can construct an expected verification table upon which various statistical scores can be computed. The verification table's entries (H_{ij}) are estimated from the two distributions and the selected p* by these relationships:

$$H_{11} = C \int_{p*}^{B} f_1 d\hat{Y}$$

$$H_{10} = (1-C) \int_{p*}^{B} f_0 d\hat{Y}$$

$$H_{01} = C \int_{A}^{p*} f_1 \, d\hat{Y} = C - H_{11} \tag{2}$$

$$H_{00} = (1-C) \int_{A}^{p*} f_0 \, d\hat{Y} = (1-C) - H_{10}$$

To control the frequency of positive classifications (the G measure in figure 2), simply solve for the p* that gives the desired frequency result:

$$G = H_{11} + H_{10} \tag{3}$$

For example, classification control to balance the classification table's margins can be accomplished by finding the p* which yields G = C. Other scores can likewise be maximized by stepping p* through the (A,B) interval, deriving the expected verification table (the H_{ij} values will change), computing the desired statistical score, and stopping where the desired maximum or minimum score is found. For example, to maximize the threat score find the p* which yields $T_{max} = H_{11}/ (H_{11} + H_{10} + H_{01})$, or to maximize the Heidke skill score find *p such that

$$S_{MAX} = \frac{H_{11} + H_{00} - CG - (1-C)(1-G)}{1 - CG - (1-C)(1-G)} \tag{4}$$

A decision-theory application is also available. If a user has a known utility or value-assessment to apply against the expected verification table, one merely varies the p* until an expected maximum gain or minimum loss value results.

4. STATISTICS OF THE PROBABILITY VALUE DISTRIBUTIONS

Specifying the analytic form of the underlying distributions is a vital component of a threshold model because the H_{ij} values defined previously require some analytic function to integrate. The properties of the distributions in question are examined:

Definitions:

- C Relative frequency of the predictand event when Y=1.
- R The correlation between the Y and $\hat{Y}$ over the dependent sample (also known as the multiple correlation coefficient).
- f_i Shorthand notation for the distributions $f_i\ (\hat{Y}/Y=i)$, i=0,1.
- μ_i Mean value of the distribution f_i, i=0,1.
- σ_i^2 Variance of $\hat{Y}$ about μ_i when Y=i, i=0,1.
- σ^2 Total predictand variance.
- σ_w^2 Pooled predictand variance.

Computations and relationships:

$$C = \frac{1}{N} \sum_{j=1}^{N} Y_j \quad \text{(N=sample size)}$$

$$R^2 = (SST\text{-}SSR)/SST; \; SST\text{=sum of squares of total, } \sum_{j=1}^{N} (Y_j - C)^2$$

$$SSR\text{=sum of squares of residuals, } \sum_{j=1}^{N} (\hat{Y}_j - Y_j)^2$$

SST-SSR=SSEX or sum of squares explained.

$\mu_0 = C \ (1-R^2)$ (see proof #1)

$\mu_1 = R^2 + C \ (1-R^2)$ (see proof #1) (Notice that: $\mu_1 - \mu_0 = R^2$)

$\sigma^2 = C \ (1-C)$ (see proof #2)

$\sigma_w^2 = C \ (1-C) \ R^2 \ (1-R^2)$ (see proof #3)

We have reason to suspect the distributions f_0 and f_1 to be beta pdf's, but to prove this is quite another matter. We postulate, therefore, that if we could parameterize the constants (also known as shape parameters) of the beta pdf using only the basic statistics described and defined above, we could compute likelihoods and use the Bayes theorem to test whether the input probability value ($\hat{Y}$) is unaltered after being transformed through a beta pdf. We surmise that, if an input value is transformed into a form which accomplishes desired results, then the transformation function is appropriate. In this case the input is the probability $\hat{Y}$, and the transformation function is the Bayes theorem using likelihoods (β_i) generated from the beta pdf's. That is, we want to show that

$$\hat{Y} = \frac{C \ \beta_1 \ (\hat{Y}|Y=1)}{C \ \beta_1(\hat{Y}|Y=1) + (1-C) \ \beta_0(\hat{Y}|Y=0)}, \tag{5}$$

with

$$\beta_i \ (\hat{Y}|Y=i) = \frac{\Gamma(\alpha_i+\nu_i)}{\Gamma(\alpha_i) \cdot \Gamma(\nu_i)} \ \hat{Y}^{\alpha_i-1} \ (1-\hat{Y})^{\nu_i-1}, \ (i=0,1) \tag{6}$$

Several empirical results substantiated that the beta pdf was the required distribution, but with the relationships given above we can also demonstrate it mathematically. (See proof #4.)

5. HANDLING THE OUT-OF-RANGE PROBLEM

The beta pdf is defined over the (0,1) interval, but figure 1 illustrates the true situation where some probability values can fall outside these bounds. One could argue, therefore, that any model which produces probabilities outside of the permissible range of the beta pdf must in fact not be replicating a beta pdf. Wadsworth and Bryan (1960) show, however, that a beta pdf can be "stretched" to other bounds such as (A,B). Stretching is performed by a transformation $U = (\hat{Y}-A)/(B-A)$ from the $\hat{Y}$-scale to a U-scale. The range of (0,1) thereby expands to (A,B). Wadsworth and Bryan also state that the solution of the stretched beta pdf uses the same shape parameters α_i and ν_i. The proper beta pdf for integration to solve the H_{ij} terms becomes:

$$\beta_i\ (\hat{Y}|Y=i) = \frac{\Gamma(\alpha_i+\nu_i)}{\Gamma(\alpha_i)\cdot\Gamma(\nu_i)}\quad U^{\alpha_i-1}\,(1-U)^{\nu_i-1}\ ,\ (i=0,1) \tag{7}$$

where proof #4 shows that:

$$\alpha_i = \mu_i(\mu_i(1-\mu_i) - S_i^2)/S_i^2\ ,\qquad i=0,1$$

$$\nu_i = \alpha_i(1-\mu_i)/\mu_i\ ,\qquad i=0,1 \tag{8}$$

if

$$S_i^2 = \frac{R^2}{(1+R^2)}\ \mu_i(1-\mu_i)\ ,\qquad i=0,1 \tag{9}$$

This information allows us to solve the H_{ij} verification values from the standard beta pdf.

An important corollary to the transformation of $\hat{Y}$ to a standard beta variate U is that any value of $\hat{Y}$ lying between A and B can be transformed to lie between 0 and 1 through the formula

$$U = \frac{\hat{Y} - A}{B - A} \tag{10}$$

Since A and B are not normally precisely known, a set of reasonable values has been found:

$$A = \mu_0 - 2\sigma_w \quad \text{for } \mu_0<2\sigma_w$$

$$A = 0 \text{ elsewhere}$$

$$B = \mu_1 + 2\sigma_w \text{ for } (1-\mu_1)<2\sigma_w$$

$$B = 1 \text{ elsewhere} \tag{11}$$

also, set

$$U = 0 \text{ when } \hat{Y} < A$$

$$U = 1 \text{ when } \hat{Y} > B \tag{12}$$

Proof #5 demonstrates some relationships which pertain to estimating the beta distribution parameters from known sample estimates.

6. SUMMARY

In problems such as weather forecasting it is often important to make a categorical decision about a future event. Given that we have a probability estimate of the future state of the atmosphere, we are left with the challenge of deciding whether the probability value is sufficiently large to warrant a categorical "yes it will occur" forecast. To do this we need something to compare the probability forecast against, hence the need for a critical value called the threshold probability.

When there are various users of weather-forecast information, the same probability of occurrence can evoke different categorical responses because each will most likely have different "thresholds of pain," so to speak. For example, if a 20% chance of a severe thunderstorm is forecast, one customer with a threshold probability of 30% will pick a "no it will not happen" category while another with a 15% threshold will definitely make plans for its occurrence. The simplicity of this classification procedure is to answer the question: does the probability forecast exceed the threshold probability? If it does, forecast an occurrence; otherwise do not. The beta pdf threshold model allows us to specify the threshold probability value needed by the user through the control of the expected frequency of positive classification (or "yes" forecasts).

APPENDIX

Proof #1: Prove that

$$\mu_0 = C(1-R^2) \tag{1}$$

and that

$$\mu_1 = R^2 + C(1-R^2). \tag{2}$$

Given that

$$R^2 = \frac{SSEX}{SST} , \tag{3}$$

where the sum of squares explained can be obtained from

$$SSEX = \sum_{k=1}^{K} d_k \sum_{j=1}^{N} X_{jk} Y_j - NC^2 \tag{4}$$

and (see proof #2)

$$SST = NC(1-C). \tag{5}$$

In addition, the mean of $\hat{Y}$ when the event occurs can be obtained from

$$\mu_1 = \sum_{k=1}^{K} d_k \sum_{j=1}^{N} X_{jk} Y_j / NC \tag{6}$$

Then, using (3), (4), and (5) we get

$$R^2 = (NC\mu_1 - NC^2)/NC(1-C). \tag{7}$$

Combining (7) with (6) will yield

$$\mu_1 = R^2 + C(1-R^2) \tag{8}$$

and since, with (8),

$$C = C\mu_1 + (1-C)\mu_0 , \tag{9}$$

then,

$$\mu_0 = C(1-R^2). \qquad \text{QED} \tag{10}$$

Proof #2

$$\sigma^2 = C(1-C). \tag{1}$$

Given that Y is a binary variable (0 or 1)

$$\sigma^2 = \frac{1}{N} \cdot SST$$

$$\sigma^2 = \frac{1}{N} \sum_{j=1}^{N} (Y_j - \overline{Y})^2 \tag{2}$$

$$\sigma^2 = \frac{1}{N} \sum_{j=1}^{N} (Y_j - 2Y_j\overline{Y} + \overline{Y}^2)$$

$$\sigma^2 = \frac{1}{N} \sum_{j=1}^{N} Y_j - \frac{2Y}{N} \sum_{j=1}^{N} Y_j + \overline{Y}^2$$

Since $Y^2 = Y$ then $\sum_{j=1}^{N} Y_j^2 = \sum_{j=1}^{N} Y_j$ and $\overline{Y} = C$.

Thus,

$$\sigma^2 = C - 2C^2 + C^2 \tag{3}$$

or

$$\sigma^2 = C(1-C). \qquad \text{QED} \tag{4}$$

Proof #3: Prove that

$$\sigma_w^2 = C(1-C)\ R^2\ (1-R^2) \tag{1}$$

given that

$$\sigma_w^2 = \frac{1}{N} SSR. \tag{2}$$

Further, from the Analysis of Variance in regression,

$$SSR = SST-SSEX \tag{3}$$

However, we know that

$$SST = NC(1-C)R^2 \tag{4}$$

and

$$SSEX = n_0\ (\mu_0-C)^2 + n_1(\mu_1-C)^2 \tag{5}$$

where

$$n_0 = N(1-C) \tag{6}$$

$$n_1 = NC$$

Thus,

$$SSR = NC(1-C)R^2-N(1-C)(\mu_0-C)^2-CN(\mu_1-C)^2\ . \tag{7}$$

But, from proof #1

$$\mu_0 = C(1-R^2) \qquad (8)$$

$$\mu_1 = R^2 + C(1-R^2).$$

We then get

$$\sigma_w^2 = C(1-C)R^2-(1-C)(C-CR^2-C)^2-C(R^2+C-CR^2-C)^2$$

$$\sigma_w^2 = C(1-C)R^2-(1-C)C^2R^4-C(1-C)^2R^4$$

$$\sigma_w^2 = C(1-C)[R^2-CR^4-(1-C)R^4]$$

$$\sigma_w^2 = C(1-C)(R^2-CR^4-R^4+CR^4)$$

$$\sigma_w^2 = C(1-C)(1-R^2)R^2$$

<u>Proof #4</u>: Prove that

$$\hat{Y} = \frac{C \cdot \beta_1(\hat{Y}|Y=1)}{C \cdot \beta_1(\hat{Y}|Y=1)+(1-C) \cdot \beta_0(\hat{Y}|Y=0)} \qquad (1)$$

where

$$\beta_i(\hat{Y}|Y=i) = \frac{\Gamma(\alpha_i+\nu_i)}{\Gamma(\alpha_i) \cdot \Gamma(\nu_i)} \hat{Y}^{\alpha_i-1}(1-\hat{Y})^{\nu_i-1}, \quad (i=0,1) \qquad (2)$$

This is tantamount to showing that event probability forecasts, $\hat{Y}$, in the beta distribution produce likelihoods which, when applied to the Bayes theorem, yields itself.

Or, that

$$\hat{Y} = \frac{Cf_1}{Cf_1+(1-C)f_0} \qquad (3)$$

Basic relationships and definitions:

$$f_1 = \frac{\Gamma(\alpha_1+\nu_1)}{\Gamma(\alpha_1)\ \Gamma(\nu_1)} \hat{Y}^{\alpha_1-1}(1-\hat{Y})^{\nu_1-1} \qquad (4)$$

$$f_0 = \frac{\Gamma(\alpha_0+\nu_0)}{\Gamma(\alpha_0)\Gamma(\nu_0)} \hat{Y}^{\alpha_0-1}(1-\hat{Y})^{\nu_0-1} \qquad (5)$$

$$\alpha_i = \mu_i \left(\mu_i(1-\mu_i)-S_i^2\right)/S_i^2 \qquad i=0,1 \qquad (6)$$

$$\nu_i = \left(\frac{1-\mu_i}{\mu_i}\right) \alpha_i \qquad i=0,1 \qquad (7)$$

where

μ_1 = mean of Y when $\hat{Y}=1$

μ_0 = mean of $\hat{Y}$ when $Y=0$

S_1^2 = variance of $\hat{Y}$ about μ_1 when $Y=1$

S_0^2 = variance of $\hat{Y}$ about μ_0 when $Y=0$

with

$$\mu_1 = R^2 + C(1-R^2) = R^2 + \mu_0 \qquad \text{(Proof \#1)} \qquad (8)$$

$$\mu_0 = C(1-R^2) \qquad \text{(Proof \#1)} \qquad (9)$$

$$S_i^2 = \frac{R^2}{1+R^2}\, \mu_i(1-\mu_i), \qquad (i=0,1) \qquad (10)$$

and

R^2 = Reduction of variance of the forecast equation, or the square of the correlation between the forecast probabilities and the dependent variable over the dependent sample.

Before we solve (3) simplify some of the above parameters:

Putting (10) into (6) reduces $\alpha_i = \frac{\mu_i}{R^2}$, $\qquad i=0,1 \qquad (11)$

Putting (8) or (9) into (7) reduces $\nu_i = \frac{1-\mu_i}{R^2}$, $\qquad i=0,1 \qquad (12)$

Now, $\alpha_i + \nu_i = \frac{1}{R^2}$ $\qquad i=0,1 \qquad (13)$

Rewriting (3) as $\dfrac{1}{1 + \frac{(1-C)}{C} \cdot \frac{f_0}{f_1}} = \dfrac{1}{1 + D}$

and reducing the term D: Returning to (4) and (5), D becomes:

$$D = \frac{1-C}{C} \cdot \frac{\Gamma(\alpha_0 + \nu_0)}{\Gamma(\alpha_1 + \nu_1)} \cdot \frac{\Gamma(\alpha_1)}{\Gamma(\alpha_0)} \cdot \frac{\Gamma(\nu_1)}{\Gamma(\nu_0)} \cdot \hat{Y}^{\alpha_0-\alpha_1} (1-\hat{Y})^{\nu_0-\nu_1} \qquad (14)$$

From (11) $\alpha_0 - \alpha_1 = \frac{\mu_0 - \mu_1}{R^2}$ (15)

and from (12) $\nu_0 - \nu_i = \frac{\mu_1 - \mu_0}{R^2}$ (16)

But we also see from (8) that $\mu_1 - \mu_0 = R^2$ (17)

Therefore, (15) and (16) become

$$\alpha_0 - \alpha_1 = -1 \qquad (18)$$
$$\nu_0 - \nu_1 = 1$$

From (13) we see that $\Gamma(\alpha_0 + \nu_0) = \Gamma(\alpha_1 + \nu_1) = \Gamma(\frac{1}{R^2})$ (19)

Now (14) becomes, with (15), (16), and (17):

$$D = \frac{1-C}{C} \cdot \frac{\Gamma(\alpha_1}{\Gamma(\alpha_0)} \cdot \frac{\Gamma(\nu_1)}{\Gamma(\nu_0)} \cdot \frac{(1-\hat{Y}}{\hat{Y}} \qquad (20)$$

Next we look at the ratio $\frac{\Gamma(\alpha_1)}{\Gamma(\alpha_0)}$: (21)

from (11) and (8)

$$\Gamma(\alpha_1) = \Gamma(\frac{\mu_1}{R^2}) = \Gamma(1+ \frac{\mu_0}{R^2}) \qquad (21)$$

From (11)

$$\Gamma(\alpha_0) = \Gamma(\frac{\mu_0}{R^2}) \qquad (22)$$

Using the feature of the Gamma function that $\Gamma(1+Z) = Z\,\Gamma(Z)$, $Z>0$

we change (21) to

$$\Gamma(1+ \frac{\mu_0}{R^2}) = \frac{\mu_0}{R^2} \Gamma(\frac{\mu_0}{R^2}) \qquad (23)$$

Now from (22), (23), and (9)

$$\frac{\Gamma(\alpha_1)}{\Gamma(\alpha_0)} = \frac{\mu_0}{R^2} = \frac{C(1-R^2)}{R^2} \qquad (24)$$

Next look at the ratio $\frac{\Gamma(\nu_1)}{\Gamma(\nu_0)}$:

From (12) and (8)

$$\Gamma\ (\nu_1) = \Gamma\ (\frac{1-\mu_1}{R^2}) = \Gamma\ (\ \frac{1-\mu_0-R^2}{R^2}\ ) = \Gamma\ (-[1-\frac{1-\mu_0}{R^2}\ ]). \qquad (25)$$

From (12)

$$\Gamma(\nu_0) = \Gamma\ (\frac{1-\mu_0}{R^2}) \qquad (26)$$

Using the feature of the Gamma function that

$$\Gamma\ (-Z) = -\frac{\Gamma(1-Z)}{Z}\ ,\ Z > 0$$

Change 25 to

$$\Gamma\ (\ -\ [1-\frac{1-\mu_0}{R^2}\ ]\ ) = \frac{\Gamma\left(\frac{1-\mu_0}{R^2}\right)}{\frac{1-\mu_0}{R^2}-1} \qquad (27)$$

and using (26) and (27)

$$\frac{\Gamma(\nu_1)}{\Gamma(\nu_0)} = \frac{1}{\frac{1-\mu_0}{R^2}-1} \qquad (28)$$

Before returning to solve D, (28) can be simplified further:

From (9) $$\frac{\Gamma\ (\nu_1)}{\Gamma\ (\nu_0)} = \frac{1}{\frac{1-C(1-R^2)}{R^2}-1} = \frac{R^2}{1-C+CR^2\ -R^2} \qquad (29)$$

$$= \frac{R^2}{(1-C)-(1-C)R^2}$$

$$= \frac{R^2}{(1-C)(1-R^2)}$$

Returning (24) and (29) to (20) yields:

$$D = \frac{1-\hat{Y}}{\hat{Y}} \qquad (30)$$

Now reordered the form of (4) using (30), we finally prove

$$\hat{Y} = \frac{1}{\frac{1+1-\hat{Y}}{\hat{Y}}} = \frac{\hat{Y}}{\hat{Y}+1-\hat{Y}} = \hat{Y} \qquad \text{QED}$$

Proof #5: Show that

$$\hat{\alpha}_i = \hat{\mu}_i\left(\hat{\mu}_i(1-\hat{\mu}_i)-\sigma_1^2\right)/\sigma_1^2 \qquad i=0,1 \qquad (1)$$

$$\hat{\nu}_i = \hat{\alpha}_i(1-\hat{\mu}_i)/\hat{\mu}_i \qquad i=0,1 \qquad (2)$$

Given, from the Beta distribution (see Feller 1966, p. 49) that

$$\mu_i = \frac{\alpha_i}{\alpha_i+\nu_i} \qquad i=0,1 \qquad (3)$$

and

$$\sigma_i^2 = \frac{\alpha_i\nu_i}{(\alpha_i+\nu_i)^2(\alpha_i+\nu_i+1)} \qquad i=0,1 \qquad (4)$$

From (3) and the estimates $\hat{\mu}_i$ and $\hat{\sigma}_i^2$ of μ_i and σ_i^2, respectively, we satisfy (2) by

$$\hat{\nu}_i = \frac{\hat{\alpha}_i(1-\hat{\mu}_i)}{\hat{\mu}_i} \qquad i=0,1 \qquad (5)$$

Now from (4) with μ_i and σ_i^2 replaced by their estimates $\hat{\mu}_i$ and $\hat{\sigma}_i^2$, respectively,

$$\hat{\sigma}_i^2 = \frac{\hat{\mu}_i^2 - \hat{\mu}_i^3}{\hat{\alpha}_i + \hat{\mu}_i} \qquad i=0,1 \qquad (6)$$

Therefore (1) is satisfied by using (4) and (6) or

$$\hat{\alpha}_i = \hat{\mu}_i\left(\hat{\mu}_i(1-\hat{\mu}_i)-\hat{\sigma}_i^2\right)/\hat{\sigma}_i^2 \qquad i=0,1 \qquad (7)$$

It is practical to employ σ_w^2 in place of σ_1^2 and σ_0^2, since the latter two require reference to the raw data and σ_w^2 does not. In fact,

$$\sigma_w^2 = R^2(1-R^2)\ C(1-C), \qquad (8)$$

from proof #3 QED

Experimental evidence has shown that using σ_w^2 for the individual group beta distributions or using σ^2 for the total beta distribution, with $\hat{Y}$ providing the likelihood ratios, performs equally well on the integration needed to determine P*.

References

Feller, William, An Introduction to Probability Theory and Its Applications, Volume II., John Wiley & Sons, New York, 1966, p. 49.

Wadsworth, G. P., and Joseph G. Bryan, Introduction to Probability and Random Variables, McGraw-Hill Book Co., 1960, pp. 101-104.