Evaluation of the Environmental Caucus's Binomial Method Critique

Steven G. Saiz
Division of Water Quality
SWRCB

March 3, 2004

Introduction

Craig Wilson asked me to evaluate comments sent to the Monitoring and TMDL Listing Unit in response to the recent release of the SWRCB Water Quality Control Policy for Developing California's Clean Water Act Section 303(d) List. Several environmental groups joined efforts and commented on the statistical decision-making errors expected when using the binomial statistical model proposed in the Policy.

My evaluation of the Environmental Caucus's critique reflects my personal observations and is my attempt to explain and expand upon the ideas presented by the Caucus.

Specifically, the Environmental Caucus has argued that the Policy's binomial model is biased in favor of not listing water bodies. Appendix I of the Environmental Caucus's comment document presented an expression to calculate the probability of failing to list a water body when the true exceedance rate is greater than 0.1. This probability is known as a beta statistical error, β. The Environmental Caucus presented the expression below (their equation A1) as an analytical solution to calculate β:

$$P(k < klist, r > 0.1 \mid N) = \sum_{k=0}^{klist-1} \int_{r=0.1}^{1} \left(\frac{N!}{k!(n-k)!} \right) r^k (1-r)^{(N-k)} dr,$$

where N = sample size, klist = number of exceedances in N samples needed in order to list the water body with a significance level of 10% (i.e., with 90% confidence), and r = the true unknown exceedance rate.

Similarly, the Environmental Caucus presented an expression for the probability of incorrectly listing a clean water body when the true exceedance rate is ≤ 0.1. This is known as an alpha statistical error, α. The Environmental Caucus presented the expression below (their equation A4) as an analytical solution to calculate α:

$$P(k \geq klist, r \leq 0.1 \mid N) = \sum_{k=klist}^{N} \int_{r=0}^{0.1} \left(\frac{N!}{k!(n-k)!} \right) r^k (1-r)^{(N-k)} dr.$$

Numeric solutions to these equations for various sample sizes were presented in Table 1 of the Environmental Caucus document and are reproduced in Table 1 of this document.

Observations

<u>Observation 1.</u> The statistical probabilities presented in Table 1 are the sum of all possible statistical errors over all possible alternate exceedance rates.

The given values are an attempt to summarize α and β errors over a continuous range of alternate exceedance rates. The definite integral portion of both equations effectively sums up the total statistical error for infinitely small changes in alternate exceedance rates. The resulting Table 1 "α" values are therefore summations of all αs for $0 \leq r \leq 0.1$ and should more properly be expressed symbolically as $\Sigma\alpha$. Similarly, the Table 1 "β" values are summations of all βs for $0.1 < r \leq 1$ and should more properly be expressed as $\Sigma\beta$. The units of these summations are (*probability of error* x *alternate exceedance rate*) and not *probability of error*, per se.

To help illustrate what the summations are measuring, I created a power curve graph (Fig 1) for a one-sided binomial hypothesis test with N=10. Values in this graph were calculated (Gibbons 1964) by finding the probability of rejecting the null hypothesis (Ho: $r \leq 0.1$) under many alternate exceedance rates when the nominal significance level remains at 10%,

$$P(rejectHo) = P(k \geq klist \mid r, N) = \sum_{k=klist}^{N} \left(\frac{N!}{k!(n-k)!} \right) r^k (1-r)^{(N-k)}.$$

The right side of this expression is the standard cumulative binomial probability distribution. When $r \leq 0.1$ the above expression is an estimate of α. When $r > 0.1$ the above expression is an estimate of statistical power $(1 - \beta)$, the probability of correctly rejecting a false null hypothesis.

Taking unity minus these probabilities produces a graph of the statistical confidence coefficient and β, respectively, (Fig. 2) over a range of alternate exceedance rates. Finally, both statistical errors, α and β, are shown in Figure 3 along with the summations, $\Sigma\alpha$ and $\Sigma\beta$, produced by the definite integrals in equations A4 and A1.

These summations are areas under the statistical error curves. We do not expect $\Sigma\alpha + \Sigma\beta$ to add to one, as is required by a probability density function. Rather we expect $\Sigma\alpha + (0.1-\Sigma\alpha) + \Sigma\beta + (0.9 - \Sigma\beta)$ to equal unity. The expressions in parentheses are the areas above the error curves in Figure 3.

An alternative approach to estimating the $\Sigma\alpha$ and $\Sigma\beta$ values is through the use of Riemann sums (Ellis & Gulick 1994, p.293). For example, let f represent the beta error function in Figure 3. We may partition the interval [0.1, 1] into n subdivisions of size Δr. The area under f is then approximated as,

$$\sum \beta = \int_{0.1}^{1} f(r)dr \approx \sum_{k=1}^{n} f(r_k) \Delta r_k,$$

where r_k is any arbitrary alternate exceedance rate within the kth subdivision and $f(r_k)$ is the beta error associated with the alternate exceedance rate, r_k. Increasing the number of subdivisions will produce increasingly accurate estimates of $\Sigma\beta$ (Table 3) and these approximations will converge on the $\Sigma\beta$ values given in Table 1. Again, note that the elements being summed in the above summation have units of (*probability of error* x *alternate exceedance rate*).

<u>Observation 2.</u> The $\Sigma\alpha$ and $\Sigma\beta$ values should not be used as a measure of overall statistical error or test performance.

[handwritten: or 0.9298 ? why get smaller approaches 0.929]

Note from Table 1 and the above example using N=10 that α ranges from 0 to 0.0 , whereas $\Sigma\alpha$ = 0.002. Similarly, β ranges from 0 to 0.91 , whereas $\Sigma\beta$ = 0.175. These single summation values are being used by the Environmental Caucus to represent the entire range of statistical errors over the alternative exceedance rates. However, these summation values are not measures of central tendency for the statistical error functions.

Expressions A1 and A4 can easily be modified, however, to produce the average value of the statistical error over the limits of integration a and b (Ellis & Gulick 1994, p.308). An expression for the average beta error is,

$$\overline{\beta} = \frac{1}{(1-0.1)} \int_{r=0.1}^{1} \sum_{k=0}^{klist-1}\left(\frac{N!}{k!(n-k)!}\right)r^{k}(1-r)^{(N-k)}dr,$$

which is expression A1 multiplied by the ratio $1/(b-a) = 1/(1-0.1) = 10/9 = 1.111$.

Similarly, an expression for the average alpha error is,

$$\overline{\alpha} = \frac{1}{(0.1-0)} \int_{r=0}^{0.1} \sum_{k=klist}^{N}\left(\frac{N!}{k!(n-k)!}\right)r^{k}(1-r)^{(N-k)}dr,$$

which is expression A4 multiplied by the ratio $1/(b-a) = 1/(0.1-0) = 10$. Note also that these average statistical error values now have the correct units of *probability of error*. I have recalculated Table 1 using the average alpha and beta equations (Table 4). The "willingness to err" ratio is 9 times lower than originally given in Table 1.

Moreover, the average values in Table 4 compare favorably with average values obtained in a typical power analysis. Returning again to the N = 10 example, $\overline{\alpha}$ = 0.0196 and $\overline{\beta}$ = 0.1941 when using the above average value equations. These values compare favorably to the average values obtained from the power analysis results in Table 2 of average alpha = 0.0212 and average beta = 0.1911

Using the average statistical error is one approach for quantifying the overall decision-making error associated with the binomial model. However, Dixon (1953) has cautioned against using an averaging process when comparing power curves because averaging disguises differences in power curve shapes.

3

Observation 3. The "willingness to err" ratio is of theoretical importance only.

In practice, only one type of decision-making error can be made for a given tested water body. Ideally, we will make a correct decision. If we do make a decision error then it will be either an α error or a β error, never both. Therefore, to express the ratio of β to α may not make sense from a practical standpoint.

A better expression might be the total probability of making an α error when making all decisions possible under a true null hypothesis. This can be calculated using the summations in Table 1,

$$P(\alpha \text{ error} \mid Ho \text{ true}) = \Sigma\alpha / (\Sigma\alpha + (0.1 - \Sigma\alpha)) = 10 \Sigma\alpha.$$

In a similar fashion, the total probability of making a β error when making all decisions possible under a false hypothesis is,

$$P(\beta \text{ error} \mid Ho \text{ false}) = \Sigma\beta / (\Sigma\beta + (0.9 - \Sigma\beta)) = (10/9) \Sigma\beta.$$

Notice that these expressions are equivalent to using the average values of alpha and beta.

Observation 4. Alpha and beta errors can only be balanced asymptotically, i.e., with large sample sizes, when using a test of medians (i.e, Ho: $r \leq 0.50$) at the 50% confidence level.

The Environmental Caucus suggested raising the significance level to 50% in order to balance the alpha and beta errors. They further suggested lowering the exceedance rate of the binomial test to 5% (Ho: $r \leq 0.05$). Table 5 illustrates that binomial testing under these conditions will not balance the errors, but reduce the willingness to err (using averages) ratio to below unity. This means that we will list in error more often than we will incorrectly fail to list a water body.

Table 6 shows the effect of binomial testing at the 50% confidence level with Ho: $r \leq 0.50$. The willingness to err ratio now approaches unity with moderate sample sizes (say N>30). Because of the symmetry of power function graphs, this is the only test design that will balance both type of statistical errors.

The Policy's default one-sided null hypothesis attempts to put statistical control on alpha errors by testing at the 10% significance level. Beta errors are inversely related to the significance level. Therefore, if we lower the significance level of the test we will invariably increase beta errors. Conversely, if we increase the significance level we will decrease beta errors. The only way to simultaneously reduce alpha and beta errors is to increase the sample size.

4

References

Dixon, W. J. (1953). *Power functions of the sign test and power efficiency for normal alternatives*. Annals of Mathematical Statistics 24:467-473.

Ellis, R. and D. Gulick. 1994. *Calculus with analytic geometry*. Saunders College Publishing, NY.

Gibbons, J. D. 1964. *Effect of non-normality on the power function of the sign test*. American Statistical Assoc. Journal 59:142-148.

Fig 1. Statistical alpha error and power associated with the binomial model for N=10, klist = 3, and a nominal significance level of 10%. Alpha error is the blue line function to the left of the 0.1 exceedance rate. Statistical power is the blue line function to the right of the 0.1 alternate exceedance rate.
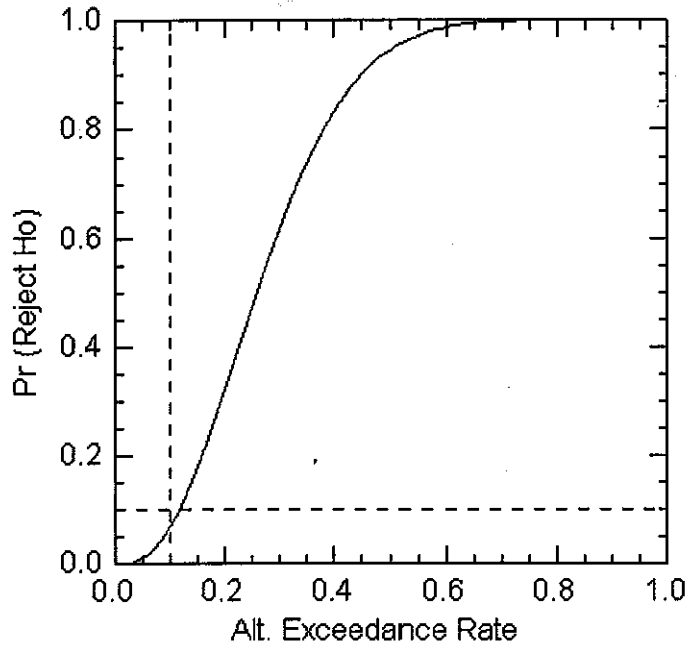
Fig 2. Statistical confidence and beta error associated with the binomial model for N=10, klist = 3, and a nominal significance level of 0.1. Statistical confidence coefficient is the blue line function to the left of the 0.1 exceedance rate. Beta error is the blue line function to the right of the 0.1 alternate exceedance rate.
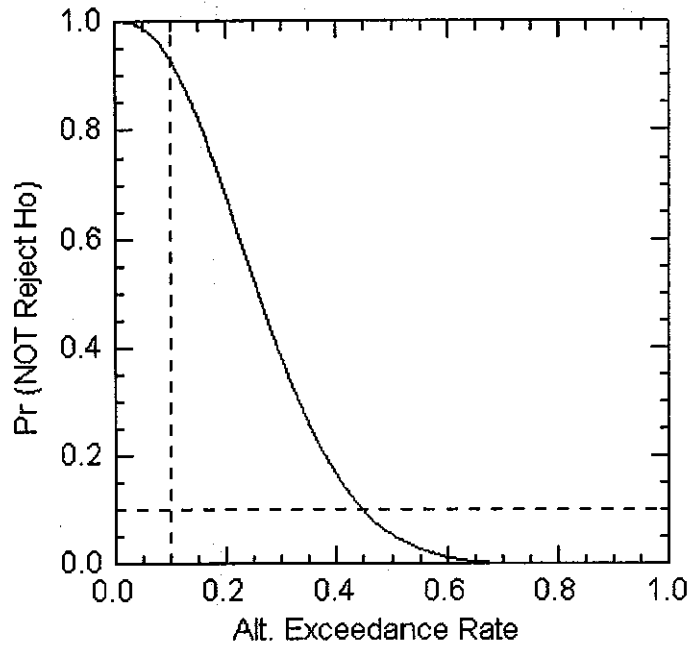
Fig 3. Statistical errors associated with the binomial model when N=10, klist=3, and a nominal significance level of 0.1. The green shaded area is the sum of all alpha errors, $\Sigma\alpha = 0.002$ and the blue shaded area is the sum of all beta errors, $\Sigma\beta = 0.17$.

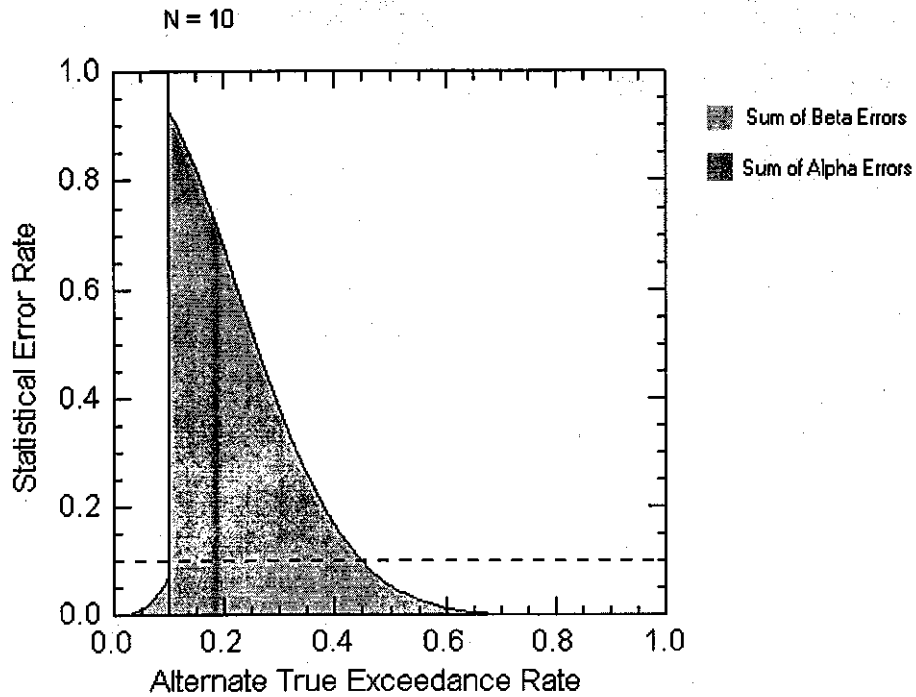*or 0.1941   or 0.0196*

8

Table 1. Reproduction of the Environmental Caucus's Table 1. Probabilities of making listing errors under the Draft Policy

Sample Size, N	Place on 303(d) list if at least this number of exceedances, klist	Probability of listing a clean water body, α	Probability of failing to list an impaired water body, β	Willingness to Err Ratio, (β/α)
10	3	0.002	0.175	89
12	4	0.001	0.208	362
19	5	0.001	0.151	213
26	6	0.001	0.123	169
33	7	0.001	0.107	153
41	8	0.001	0.091	122
48	9	0.001	0.084	124
56	10	0.001	0.076	111
64	11	0.001	0.070	102
72	12	0.001	0.065	97
80	13	0.001	0.061	93
89	14	0.001	0.056	81
97	15	0.001	0.054	81
105	16	0.001	0.052	81

Table 2. Probability of rejecting the null hypothesis (Ho: $r \leq 0.1$) for N = 10 and klist = 3. Average alpha = 0.0212, average power = 0.8089, average beta = 1 - 0.8089 = 0.1911.

Alternate exceedance rate, r	Decision If Ho is rejected	P(reject Ho)	P(NOT reject Ho) = 1- P(reject Ho)
0.00	incorrect, alpha error	0.000	1.000
0.01	incorrect, alpha error	0.000	1.000
0.02	incorrect, alpha error	0.001	0.999
0.03	incorrect, alpha error	0.003	0.997
0.04	incorrect, alpha error	0.006	0.994
0.05	incorrect, alpha error	0.012	0.989
0.06	incorrect, alpha error	0.019	0.981
0.07	incorrect, alpha error	0.028	0.972
0.08	incorrect, alpha error	0.040	0.960
0.09	incorrect, alpha error	0.054	0.946
0.10	incorrect, alpha error	0.070	0.930
0.11	correct, power	0.088	0.912
0.12	correct, power	0.109	0.891
0.13	correct, power	0.131	0.869
0.14	correct, power	0.155	0.846
0.15	correct, power	0.180	0.820
0.16	correct, power	0.206	0.794
0.17	correct, power	0.234	0.766
0.18	correct, power	0.263	0.737
0.19	correct, power	0.292	0.708
0.20	correct, power	0.322	0.678
0.21	correct, power	0.353	0.647
0.22	correct, power	0.383	0.617
0.23	correct, power	0.414	0.586
0.24	correct, power	0.444	0.556
0.25	correct, power	0.474	0.526
0.26	correct, power	0.504	0.496
0.27	correct, power	0.534	0.467
0.28	correct, power	0.562	0.438
0.29	correct, power	0.590	0.410
0.30	correct, power	0.617	0.383
0.31	correct, power	0.643	0.357
0.32	correct, power	0.669	0.331
0.33	correct, power	0.693	0.307
0.34	correct, power	0.716	0.284
0.35	correct, power	0.738	0.262
0.36	correct, power	0.760	0.241
0.37	correct, power	0.779	0.221
0.38	correct, power	0.798	0.202
0.39	correct, power	0.816	0.184
0.40	correct, power	0.833	0.167
0.41	correct, power	0.848	0.152
0.42	correct, power	0.863	0.137
0.43	correct, power	0.876	0.124
0.44	correct, power	0.889	0.111
0.45	correct, power	0.900	0.100
0.46	correct, power	0.911	0.089
0.47	correct, power	0.921	0.079

0.48	correct, power	0.930	0.070
0.49	correct, power	0.938	0.062
0.50	correct, power	0.945	0.055
0.51	correct, power	0.952	0.048
0.52	correct, power	0.958	0.042
0.53	correct, power	0.963	0.037
0.54	correct, power	0.968	0.032
0.55	correct, power	0.973	0.027
0.56	correct, power	0.976	0.024
0.57	correct, power	0.980	0.020
0.58	correct, power	0.983	0.017
0.59	correct, power	0.985	0.015
0.60	correct, power	0.988	0.012
0.61	correct, power	0.990	0.010
0.62	correct, power	0.991	0.009
0.63	correct, power	0.993	0.007
0.64	correct, power	0.994	0.006
0.65	correct, power	0.995	0.005
0.66	correct, power	0.996	0.004
0.67	correct, power	0.997	0.003
0.68	correct, power	0.998	0.003
0.69	correct, power	0.998	0.002
0.70	correct, power	0.998	0.002
0.71	correct, power	0.999	0.001
0.72	correct, power	0.999	0.001
0.73	correct, power	0.999	0.001
0.74	correct, power	0.999	0.001
0.75	correct, power	1.000	0.000
0.76	correct, power	1.000	0.000
0.77	correct, power	1.000	0.000
0.78	correct, power	1.000	0.000
0.79	correct, power	1.000	0.000
0.80	correct, power	1.000	0.000
0.81	correct, power	1.000	0.000
0.82	correct, power	1.000	0.000
0.83	correct, power	1.000	0.000
0.84	correct, power	1.000	0.000
0.85	correct, power	1.000	0.000
0.86	correct, power	1.000	0.000
0.87	correct, power	1.000	0.000
0.88	correct, power	1.000	0.000
0.89	correct, power	1.000	0.000
0.90	correct, power	1.000	0.000
0.91	correct, power	1.000	0.000
0.92	correct, power	1.000	0.000
0.93	correct, power	1.000	0.000
0.94	correct, power	1.000	0.000
0.95	correct, power	1.000	0.000
0.96	correct, power	1.000	0.000
0.97	correct, power	1.000	0.000
0.98	correct, power	1.000	0.000
0.99	correct, power	1.000	0.000

25118

Table 3. Estimation of $\sum\beta$ using Riemann sums. The interval [0.1, 1] on the x-axis of Figure 3 is partitioned into n subdivisions of size Δr.

Number of Partitions, n	Subdivision Size, Δr	$\sum\beta$ Approximation
4	0.25	0.1452
9	0.1	0.1297
90	0.01	0.1701
900	0.001	0.1752
9000	0.0001	0.1747
∞	lim Δr$\rightarrow$0	0.175 (from Table 1)

Table 4. Table 1 recalculated using the average statistical error expected over the range of alternate exceedance values. Ho: $r \leq 0.1$ and significance level is 10%.

Sample Size, N	Place on 303(d) list if at least this number of exceedances, klist	Average probability of listing a clean water body, $\overline{\alpha}$	Average probability of failing to list an impaired water body, $\overline{\beta}$	Willingness to Err Ratio, $(\overline{\beta}/\overline{\alpha})$
10	3	0.0196	0.1941	9.9
12	4	0.0058	0.2314	40.2
19	5	0.0071	0.1675	23.7
26	6	0.0073	0.1366	18.8
33	7	0.0070	0.1184	17.0
41	8	0.0075	0.1014	13.5
48	9	0.0068	0.0937	13.8
56	10	0.0069	0.0846	12.3
64	11	0.0069	0.0777	11.3
72	12	0.0067	0.0723	10.7
80	13	0.0066	0.0679	10.4
89	14	0.0069	0.0625	9.0
97	15	0.0067	0.0597	9.0
105	16	0.0064	0.0573	9.0

Table 5. Average statistical error expected over the range of alternate exceedance rates when Ho: $r \leq 0.05$ and significance level is 50%.

Sample Size, N	Place on 303(d) list if at least this number of exceedances, klist	Average probability of listing a clean water body, $\overline{\alpha}$	Average probability of failing to list an impaired water body, $\overline{\beta}$	Willingness to Err Ratio, $(\overline{\beta}/\overline{\alpha})$
10	1	0.2160	0.0544	0.25
12	1	0.2513	0.0416	0.17
19	2	0.0943	0.0576	0.61
26	2	0.1529	0.0334	0.22
33	2	0.2132	0.0205	0.10
41	3	0.1130	0.0285	0.25
48	3	0.1529	0.0199	0.13
56	4	0.0887	0.0259	0.29
64	4	0.1219	0.0186	0.15
72	4	0.1578	0.0134	0.08
80	5	0.1010	0.0177	0.17
89	5	0.1330	0.0128	0.10
97	6	0.0886	0.0165	0.19
105	6	0.1119	0.0128	0.11

Table 6. Average statistical error expected over the range of alternate exceedance rates when Ho: $r \leq 0.5$ and significance level is 50%.

Sample Size, N	Place on 303(d) list if at least this number of exceedances, klist	Average probability of listing a clean water body, $\overline{\alpha}$	Average probability of failing to list an impaired water body, $\overline{\beta}$	Willingness to Err Ratio, $(\overline{\beta}/\overline{\alpha})$
10	6	0.0776	0.1685	2.2
12	7	0.0743	0.1513	2.0
19	10	0.0881	0.0881	1.0
26	14	0.0590	0.0960	1.6
33	17	0.0679	0.0679	1.0
41	21	0.0612	0.0612	1.0
48	25	0.0471	0.0675	1.4
56	29	0.0443	0.0618	1.4
64	33	0.0420	0.0574	1.4
72	37	0.0400	0.0537	1.3
80	41	0.0383	0.0506	1.3
89	45	0.0419	0.0419	1.0
97	49	0.0402	0.0402	1.0
105	53	0.0387	0.0387	1.0