# 1 Introduction

In credit risk management it is of critical importance to accurately, or as accurately as possible, estimate the default probability of credit instruments. The reason for this is simple: default risk is the primary determinant in valuing debt-related financial assets. With this in mind, let us briefly describe some of the background of default probability estimation. Typically, when a debt instrument or debt-derived instrument is issued in a public market, such as the New York Stock Exchange, a rating will be assigned to the instrument. The purpose of the rating is to signify to the marketplace the deemed creditworthiness of the instrument, i.e. the likelihood that the debt underlying the instrument will be repaid rather than defaulted. These ratings are assigned by private companies, so-called Credit Rating Agencies (CRAs), whose business it is to research and evaluate the viability of the obligated parties involved in the debt-issuance. Two of the major CRAs in the United States are Moody's Investor Services and Standard and Poor's (S&P's). Each CRA has their own rating schema, but all are based on hierarchical categories which range from the most creditworthy to the least creditworthy, or, alternatively, least risk to most risk. For example, S&P's highest rating is AAA, signifying their opinon that the obligor has the highest likelihood, in S&P's rating schema, of meeting their legally-bound financial commitments. After AAA, S&P's has the following rating categories, progressing in order of increased default risk: AA, A, BBB, BB, B, and C. Beyond C, a rating of D would reflect debt already in default payment status. There are two key assumptions inherent in any such rating system. One is the obvious fact that there is a simple order restriction on default probability among the categories. An entity rated AAA must have a default probability no greater than an entity rated AA, which must have a default probability no greater than an entity rated A, and so on. The second assumption is that all entities within a given rating category have the same default risk. This is clearly not theoretically true, but practically demanded. It is thus assumed that the default probabilities associated with these rating categories follow simple order restrictions and are homogenous with respect to default probability within a rating category.

In what follows, we concern ourselves first with the task of estimating default probabilities under simple order restrictions. Since the probability of default within a given rating category is homogenous, we can use a binomial model for estimation. Second, we discuss testing the hypothesis that the parameters do in fact follow a simple order restriction. A test based on the asymptotic distribution of the likelihood ratio test and a permutation test are reviewed. Next, we seek to express the reliability of these estimates with confidence intervals. This leads to a review of the recent literature regarding methods of forming binomial confidence intervals. Finally, we compare some of these methods in the context of order-restricted estimation.

# 2 Estimation of default probabilities for multiple rating classes with constraints

Given a rating system, let $N$ be the number of distinct rating classes. For each rating class $i$, $i = 1, \ldots, N$, let $p_i$ be the probability of default for a credit instrument within the $i$th class, $d_i$ the observed number of defaults in the class, and $n_i$ the total number of firms in the class. Let $\mathbf{p} = (p_1, \ldots, p_N)$. We assume the default probabilites are ordered as $p_1 \leq p_2 \leq \ldots \leq p_{N-1} \leq p_N$. We further assume that the default of any credit instrument is independent of the default of any other credit instrument, both within rating categories and without. We can then write the likelihood function as:

$$L(\mathbf{p}) = \prod_{i=1}^{N} {p_i}^{d_i} (1 - p_i)^{(n_i - d_i)}$$

and the log-likelihood as:

$$l(\mathbf{p}) = \sum_{i=1}^{N} d_i \ln(p_i) + (n_i - d_i) \ln(1 - p_i)$$

To find the MLE for $\mathbf{p}$ we need to solve the following maximization problem:

$$
\begin{aligned}
\max_{\mathbf{p}} \quad & \sum_{i=1}^{N} d_i \ln(p_i) + (n_i - d_i)\ln(1 - p_i) \\
\text{subject to} \quad & 0 \le p_1 \\
& p_1 \le p_2 \\
& \vdots \\
& p_{N-1} \le p_N \\
& p_N \le 1
\end{aligned}
$$

Since the objective function is a concave function, any local maximum will be a global maximum and the Karush-Kuhn-Tucker(KKT) conditions will give necessary and sufficient conditions for the maximum solution. Let $\lambda_1, \ldots, \lambda_{N+1}$ be the KKT multipliers associated with the constraints of the problem above. Then the solution to the maximization problem above must satisfy the following set of equations:

$$
\begin{aligned}
\lambda_1 p_1 &= 0 \\
\frac{d_i}{p_i} + \frac{n_i - d_i}{p_i - 1} - \lambda_i + \lambda_{i+1} &= 0, \quad i = 1, \ldots, N \\
\lambda_i (p_i - p_{i-1}) &= 0, \quad i = 1, \ldots, N \\
\lambda_{N+1}(1 - p_N) &= 0 \\
\lambda_i &\ge 0, \quad i = 1, \ldots, N+1
\end{aligned} \tag{1}
$$

Note the equations related to the constraints, $\lambda_i(p_i - p_{i-1}) = 0$. We must have either $\lambda_i = 0$ or $p_i = p_{i-1}$. If we set $\lambda_1 = \lambda_2 = 0$, then we can solve for $\hat{p}_1 = d_1/n_1$. Proceeding to solve for $\hat{p}_2$, we set $\lambda_3 = 0$ and get $\hat{p}_2 = d_2/n_2$. Now we must check that $\hat{p}_1 \le \hat{p}_2$. If yes, we can continue on to solve for $\hat{p}_3$ in a similar fashion. If not, we have a constraint violation. We can add the following two equations:

$$
\begin{aligned}
\frac{d_1}{p_1} + \frac{n_1 - d_1}{p_1 - 1} - \lambda_1 + \lambda_2 &= 0 \\
\frac{d_2}{p_2} + \frac{n_2 - d_2}{p_2 - 1} - \lambda_2 + \lambda_3 &= 0
\end{aligned} \tag{2}
$$

and set $\hat{p}_1 = \hat{p}_2$. Solving, we get $\hat{p}_1 = \hat{p}_2 = (d_1 + d_2)/(n_1 + n_2)$. We can continue in this fashion solving for each $\hat{p}_i$ and then checking for constraint violations. Note that when we set two estimates equal to each other we must go back and check that the previously computed estimates still satisfy the constraints.

The above considerations can be extended to the following heuristic algorithm that solves for $\hat{\mathbf{p}}$ sequentially:

1. Set $i = 1$
2. If $i = N + 1$, quit. Else set $\lambda_i = \lambda_{i+1} = 0$ and solve for $\hat{p}_i = d_i/n_i$. Check $\hat{p}_{i-1} \leq \hat{p}_i$. If yes, set $i = i + 1$ and repeat step 2. If not, go to step 3.
3. Set $\hat{p}_{i-1} = \hat{p}_i$ and add the KKT equations so that $\lambda_i$ is eliminated. Solve for $\hat{p}_{i-1} = \hat{p}_i = (d_{i-1} + d_i)/(n_{i-1} + n_i)$. Check that $\hat{p}_{i-2} \leq \hat{p}_{i-1}$. If yes, set $i = i + 1$ and go to step 2. If no, set $\hat{p}_{i-2} = \hat{p}_{i-1} = \hat{p}_i$ and add the corresponding KKT equations, solving for $\hat{p}_{i-2} = \hat{p}_{i-1} = \hat{p}_i$. Repeat until the simple order is restored. Set $i = i + 1$ and go to step 2.

The foregoing development of default estimates is an example of maximum likelihood estimation under simple order restrictions, a special case of the more general problem of isotonic regression. This was first considered by Ayer et al.(1955), where an algorithm later known as Pool Adjacent Violators Algorithm, or PAVA, was introduced. PAVA is equivalent to the above algorithm and simpler to describe. It also has the advantage of being implemented in software packages. See Robertson et al.(1988) or Silvapulle et al.(2005) for technical details. The following description of PAVA is based on the latter.

**PAVA algorithm:** Let $\hat{p}_i = d_i/n_i$ be the intial estimates.

1. If $\hat{p}_1 \leq \ldots \leq \hat{p}_N$, then stop with this solution
2. Otherwise, let $i$ be the smallest index such that $\hat{p}_i > \hat{p}_{i+1}$. Pool these two numbers to get a single estimate $\hat{p}_i = \hat{p}_{i+1} = (d_i + d_{i+1})/(n_i + n_{i+1})$
3. Repeat with the new set of estimates until all estimates are non-decreasing. Set these as the final estimates

# 3 Hypothesis Test of Simple Order Restrictions

Continuing with the above setup, we may wish to test the hypothesis that the simple order restriction holds. Clearly if our maximum likelihood estimates satisfy the constraints, before any algorithm is applied to force constraint satisfaction, we would have no evidence to reject the hypothesis that the order restrictions hold. However if at least one constraint fails to hold among the MLEs, we must judge whether or not we have sufficient evidence to reject our supposed order restriction on the parameters. Below, we consider two testing methods for dealing with this problem. The first is based on the asymptotic distribution of the likelihood ratio test (LRT). The other is a simple permutation test.

## 3.1 Likelihood Ratio Test

We again let $p_i$ be the default probability of the $i$th creditor. Therefore we take as our null hypothesis that $p_1 \leq p_2 \leq \ldots \leq p_{N-1} \leq p_N$. We can write the null and alternative hypotheses as:

$$H_0 : \mathbf{Ap} \geq \mathbf{0} \text{ against } H_1 : \mathbf{Ap} \ngeq \mathbf{0}$$

where $\mathbf{A}$ is a matrix with -1 on the diagonal, 1 on the superdiagonal, and all other entries equal to 0. The first row of $\mathbf{A}$ is then $(-1, 1, 0, 0, \ldots, 0)$, the second row is $(0, -1, 1, 0, 0, \ldots, 0)$ etc., so that every row of $\mathbf{A}$ is a pairwise contrast. The null parameter space is $\{\mathbf{p} \in \mathbb{R}^N : \mathbf{Ap} \geq \mathbf{0}\}$. It has been shown (see, for example, Silvapulle and Sen (1995)) that the likelihood ratio test, Wald-type tests, and score tests for problems of this type are all equivalent and follow a chi-bar-squared ($\bar{\chi}^2$) distribution for every point in the null parameter space. (The $\bar{\chi}^2$ distribution can be defined as follows. Let $\mathcal{C} \in \mathbb{R}^p$ be a closed convex cone and let $\mathbf{Z}_{px1} \sim N(\mathbf{0}, \mathbf{V})$, where $\mathbf{V}$ is a positive definite matrix. Then $\bar{\chi}^2(\mathbf{V}, \mathcal{C})$ is the random variable that has the same distribution as:

$$\mathbf{Z}^T\mathbf{V}^{-1}\mathbf{Z} - \min_{\theta \in \mathcal{C}} (\mathbf{Z} - \theta)^T \mathbf{V}^{-1} (\mathbf{Z} - \theta)$$

See Silvapulle and Sen (1995) or Shapiro (1985) for more details).

Although this distribution depends on the null parameter value, Silvapulle and Sen (2005) show that the least favorable null value is a point in $\{\mathbf{p} \in \mathbb{R}^N : \mathbf{Ap} = \mathbf{0}\}$ and the least favorable null distribution is equal to that of:

$$\inf\{\mathbf{Z} - \mathbf{p}\}^T \mathbf{W}\{\mathbf{Z} - \mathbf{p}\} : \mathbf{Ap} \geq \mathbf{0}\}$$

where $\mathbf{Z} \sim N(\mathbf{0}, \mathbf{W}^{-1})$ and $\mathbf{W}$ is the diagonal matrix whose $i$th entry is equal to $n_i$, $i = 1, \ldots, N$. Further, this distribution is equal to $\bar{\chi}^2(\mathbf{W}^{-1}, \mathcal{C}^0)$, where $\mathcal{C}^0 = \{\mathbf{y} : \mathbf{y}^{\mathbf{T}}\mathbf{W}\mathbf{x} \leq 0 \ \forall x$ satisfying $\mathbf{Ax} \geq \mathbf{0}\}$.

This leaves the question of how to compute p-values by using the $\bar{\chi}^2(\mathbf{W}^{-1}, \mathcal{C}^0)$ distribution. In general, calculating the probabilities of a $\bar{\chi}^2$ distribution is computationally difficult. The distribution is equal to a weighted average of $\chi^2$ distributions (hence the name) and the weights are generally difficult to calculate. However, since we are not interested in the weights here, we can resort to a simulation to approximate the exact tail probability we need. Silvapulle and Sen (1995) note the computational difficulty of computing the exact distribution and suggest the following simulation routine:

1. Generate a multivariate normal random variable, $\mathbf{Z} \sim N(\mathbf{0}, \mathbf{W})$
2. Compute $\bar{\chi}^2(\mathbf{W}^{-1}, \mathcal{C}^0)$
3. Repeat steps 1 and 2 some large number of times $N$
4. Estimate $\bar{\chi}^2(\mathbf{W}^{-1}, \mathcal{C}^0) \geq c$ by $(n/N)$, where $n$ is the number of times out of $N$ where the second step calculation is greater than or equal to $c$, $\bar{\chi}^2(\mathbf{W}^{-1}, \mathcal{C}^0) \geq c$

This can be achieved by any software that allows generating multivariate normal random variables, $\mathbf{Z} \sim N(\mathbf{0}, \mathbf{W})$ and solving a a quadratic program to compute the minimization problem defining the $\bar{\chi}^2$ distribution. Then, if we desire a hypothesis test at the 95% confidence level, we would set $c = .05$ in the above routine.

## 3.2 Permutation Test

Suppose we have two classes, $A$ and $B$, containing $n_A$ and $n_B$ creditors, and observe $d_A$ and $d_B$ defaults for each respective class. If we are interested in testing the hypothesis that an order restriction exists, say $p_A \leq p_B$, where $p_A$ and $p_B$ are the respective class default probabilities, then we could use the statistics $\hat{p}_A = \frac{d_A}{n_A}$ and $\hat{p}_B = \frac{d_B}{n_B}$ to form a natural test statistic, $T = \hat{p}_B - \hat{p}_A$. Clearly, if $p_A \leq p_B$, we would expect $T \geq 0$. If we take as our null hypothesis, $H_0 : p_A = p_B$ vs. $H_A : p_A < p_B$, then we can perform a permutation test in the following way:

First note that to perform an exact permutation test, i.e. one involving all permutations of the data, would require $(n_A + n_B)!$ permutations. Since this is very likely to be prohibitively large, we will need to sample randomly from all permutations to approximate the exact p-value. If we let $t_{obs}$ be the observed value of the test statistic $T$ and $T_i$ be the computed test statistic for the $i$th random permutation, then we can calculate the approximate p-value by:

$$\frac{1}{N} \sum_{i=1}^{N} I(T_i > t_{obs})$$

where $I$ is the indicator function. Then, for example, if the value calculated above is less than 0.05, we can reject $H_0$ at the 95% level. We could use this test on any adjacent pair of rating classes to check that the order restrictions hold with reasonable confidence. Many statistical textbooks contain more information on permutation tests, for example Lehmann and Romano (2005) and Wasserman (2004).
Lastly, we note that since the number of observations in each class, i.e. the number of companies in each rating class, will most likely be large, then the asymptotic results from the LRT should be sufficient to constitute a reliable hypothesis test. It also has the advantage of testing all parameters simultaneously, as opposed to pairwise testing.

# 4 Confidence Intervals

Once we have our estimate of default, it is natural to seek a measure of confidence with which we can gauge the accuracy the estimate, and to this end we move to the construction of confidence intervals. To form confidence intervals, we will consider our probability of default estimates as standard binomial proportion estimates. For many years the Wald confidence interval (given below) was the standard interval of choice, given that some well-known rules of thumb with regard to the parameters were met. However, there have been several papers challenging this conventional wisdom, dating back to at least Ghosh (1979), and including a thorough study by Brown et al.(2001,2002). This literature sheds light on the deficiencies of the standard Wald interval relative to several alternatives and concludes that there are no scenarios in which the Wald interval should be recommended. Cai (2005) reaches a similar conclusion regarding the inadequancy of the Wald interval in the case of one-sided intervals. Here we summarize this literature as well as describe the various alternative intervals.

In what follows, let $\alpha$ be the confidence level and let $\kappa = \Phi^{-1}(1-\alpha/2)$ for whichever $\alpha$ we are using. Note that in the one-sided intervals, $\alpha$ should be used instead of $\alpha/2$ where appropriate. Also, $\hat{p} = X/n$.

Brown et al.(2001,2002) primarily consider two criteria as a basis for comparing these confidence intervals. (They also use simplicity of presentation as an auxiliary consideration.) First, the coverage probability of the interval, i.e. the probability that, for given $n$ and $p$, the interval actually contains $p$. For a given confidence interval $C.I.$ and fixed $n$ and $p$, this is defined as:

$$P(p \in C.I.) \equiv C_n(p) = \sum_{k=0}^{n} I(k,p) \binom{n}{k} p^k (1-p)^{n-k}$$

where $I(k,p) = 1$ if the interval contains $p$ when $X = k$ and 0 if it does not contain $p$.

The coverage probability, $C_n(p)$, is approximated by means of Edgeworth expansions. These are asymptotic expansions often used to approximate the CDF of certain sample statistics, such as standardized means. Approximations based on the Central Limit Theorem can fail to capture skewness in the approximated distribution since the normal distribution is symmetric. An Edgeworth expansion uses the cumulants of the distribution to adjust for skewness and kurtosis. Although any absolutely continuous distribution has an Edgeworth expansion of fairly simple form, since the expansion itself is a smooth function, the expansion for a lattice distribution is slighly more complicated due to the necessary inclusion of extra terms for jumps in the CDF. See Bhattacharya and Rao (1976) for the Edgeworth expansion of the standardized binomial proportion.

The second criterion used is the expected length (also referred to as width) of the interval. This is defined as:

$$L(n,p) \equiv E_{n,p}(\text{length of } C.I.) = \sum_{k=0}^{n} (U(x,n) - L(x,n)) \binom{n}{k} p^k (1-p)^{n-k}$$

where $U$ and $L$ are the upper and lower limits of the confidence interval $C.I.$, respectively. Brown et al.(2002) use standard Taylor expansions to compare expected lengths of the compared intervals.

The Wald interval is based on the asymptotic normality of the sample mean, given by the central limit theorem. As mentioned above, this approximation can be inadequate due to the lattice structure (i.e. discrete nature) and skewness of the binomial distribution. For two-sided confidence intervals, the discreteness of the binomial distribution is the dominant source of error, but for one-sided intervals the skewness can be dominant. The effect of discreteness is an oscillation in the coverage probability for any interval and the effect of skewness is a bias away from the nominal confidence level. The oscillation phenomenon can be understood heuristically by noting that the upper and lower endpoints of a confidence interval will be integer valued and depend on $n$ and $p$. A small change in $n$ or $p$ can thus cause the the endpoints to "jump" to the next higher or lower integer value. The use of Edgeworth expansions of the coverage probabilities allows us to capture the differences in bias and oscillation. It turns out that one-term expansions capture the oscillations well, but are systematically biased from the true coverage probabilities. Hence Brown et al. (2002) make use of the two-term expansions, which are in fact very accurate.

Below we describe several intervals:

## The Standard Wald Interval

The standard Wald Interval is based on the normal approximation of the binomial distribution. It takes the form

$$\hat{p} \pm \kappa n^{-1/2}(\hat{p}(1-\hat{p}))^{(1/2)}$$

The interval is based on inverting the Wald large sample normal test. As Brown et al.(2001), as well as many others, show, this interval has highly erratic coverage probability, even for $n$ large. For this reason it is strongly agreed that this interval should not be used in practice.

The $100(1-\alpha)\%$ upper Wald interval is then defined to be:

$$[0, \hat{p} + \kappa n^{-1/2}(\hat{p}(1-\hat{p}))^{(1/2)}]$$

and the $100(1-\alpha)\%$ lower Wald interval is:

$$[\hat{p} + \kappa n^{-1/2}(\hat{p}(1-\hat{p}))^{(1/2)}, 1]$$

## Wilson Interval

The Wilson interval is based on inverting the standard Wald test statistic using the null standard error, $(pq)^{\frac{1}{2}}n^{-\frac{1}{2}}$ instead of the estimated standard error, $(\hat{p}\hat{q})^{\frac{1}{2}}n^{-\frac{1}{2}}$. Thus we would need to specify a null value. The Wilson interval is then given by:

$$CI_W = \frac{X + \kappa^2/2}{n + \kappa^2} \pm \frac{\kappa n^{\frac{1}{2}}}{n + \kappa^2}(\hat{p}\hat{q} + \kappa^2/(4n))^{\frac{1}{2}}$$

Similar to the one-sided Wald intervals, the upper one-sided Wilson interval is given by 0 to the upper limit above and the lower interval by the lower limit above to 1.

### Modified Wilson Interval

A possible significant problem with the Wilson interval is that its coverage probability flucuates quite a bit for $p$ near 0 or 1. Brown et al.(2001) introduce a modified version of this interval. The idea is to use a Poisson approximation for $x$ close to 0 or $n$. For example, for certain smaller values of $x$, the lower bound of $CI_W$ would be replaced by a lower bound of a Poisson approximation to $x$. Similarly for $x$ near $n$. The modified Wilson interval, $CI_{M-W}$ does indeed perform much better for $p$ values near the boundaries and performs the same for $p$ values in the mid-range.

## Jefferys Interval

The Jefferys interval is a Bayesian method, and so assumes a prior distribution on the default probability $p$. In this case, the standard conjugate to the binomial, the beta distribution, is used. The Jefferys prior is Beta(1/2,1/2). The 100(1-$\alpha$)% equal tailed Jefferys prior interval is defined as:

$$CI_J = [L_J(x), U_J(x)],$$

where $L_J(0) = 0, U_J(n) = 1$ and otherwise:

$$L_J(x) = B(\alpha/2; X + 1/2, n - X + 1/2),$$
$$U_J(x) = B(1 - \alpha/2; X + 1/2, n - X + 1/2)$$

Brown et al. mention that these endpoints must be computed numerically, but also give an approximation formula for easier computation.

For the Jefferys one-sided intervals, if we replace $\alpha/2$ with $\alpha$ in the formulas for $L_J(x)$ and $U_J(x)$, then the upper and lower intervals are $[0, U_J(x)]$ and $[L_J(x), 1]$, respectively.

### Modified Jefferys Interval

Similarly to the Wilson interval, the Jefferys interval has spikes near the boundaries. The spikes result from $U_J(0)$ is too small and $L_J(n)$ is too large. Brown et al. propose a modification:

$$U_{M-J}(0) = p_l \text{ and } L_{M-J}(n) = 1 - p_l$$

where $p_l$ satisfies $(1 - p_l)^n = \alpha/2$. They also make the following adjustment:

$$L_{M-J}(1) = 0 \text{ and } U_{M-J}(n-1) = 1$$

Elsewhere the Jefferys and modified Jefferys are equal. These modifications remove the downard spikes and thus improve the coverage probability near the boundaries.

## Agresti-Coull Interval

The Agresti-Coull interval is the simplest in form and is recommended for use when $n \geq 40$. Let $\tilde{X} = X + \kappa^2/2$ and $\tilde{n} = n + \kappa^2$. Then let $\tilde{p} = \tilde{X}/\tilde{n}$ and $\tilde{q} = 1 - \tilde{p}$. Then the Agresti-Coull interval is given by:

$$CI_{AC} = \tilde{p} \pm \kappa(\tilde{p}\tilde{q})^{1/2}\tilde{n}^{-1/2}$$

Again, similar to the one-sided Wald intervals, the upper one-sided Agresti-Coull interval is given by 0 to the upper limit above and the lower interval by the lower limit above to 1, using $\alpha$ instead of $\alpha/2$.

## Clopper-Pearson Interval

The Clopper-Pearson interval is not based on an approximation, but on the binomial distribution itself. For this reason it is sometimes referred to as an exact method. The lower and upper bounds for the $CI_{CP}$ interval, $L_{CP}(x)$ and $U_{CP}(x)$, respectively, are found by solving for $p$ in the two equations:

$$P_p(X \geq x) = \alpha/2 \text{ and } P_p(X \leq x) = \alpha/2$$

It is clearly conservative since its form guarantees that the coverage probability is always greater than or equal to the nominal confidence level. Brown et al.(2001) show, however, that it can be "wastefully" conservative, i.e. its actual coverage probability can be much higher than the nominal level. They therefore do not recommend this for practical use and refer to other exact methods that perform better if that is what is required.
For the one-sided upper Clopper-Pearson interval, let $p_u$ be the solution in $p$ to $P_p(X \geq x) = \alpha$. Then the upper interval is $[0, p_u]$. Similarly, the lower Clopper-Pearson interval is $[p_l, 0]$, where $p_l$ is the solution in $p$ to $P_p(X \leq x) = \alpha$.

## Other Interval Estimation Methods

There are many other interval estimation methods mentioned in the literature. None have been recommended widely.

### Arcsine interval

The arcsine method is based on a variance stabilizing transformation for the binomial distribution. The transformation leads to the following interval"

$$CI_{Arc} = \left[sin^2(arcsin(\breve{p}^{1/2} - \frac{1}{2}\kappa n^{-1/2})), sin^2(arcsin(\breve{p}^{1/2} + \frac{1}{2}\kappa n^{-1/2}))\right]$$

The interval performs reasonably well except on the boundaries, where it has large spikes.

### Logit interval

The logit interval is based on inverting an interval based on the MLE of the log-odds, $\lambda = \ln(\frac{p}{1-p}))$. The MLE is $\hat{\lambda} = \ln\big(\frac{X}{n-X}\big)$. The variance of $\hat{\lambda}$ can be estimated by:

$$\hat{V} = \frac{n}{X(n-x)}$$

leading to the interval:

$$CI(\lambda) = [\lambda_l, \lambda_u] = [\hat{\lambda} - \kappa \hat{V}^{1/2}, \hat{\lambda} + \kappa \hat{V}^{1/2}]$$

this then is inverted to get the logit interval:

$$CI_{Logit} = [\frac{e^{\lambda_l}}{1+e^{\lambda_l}}, \frac{e^{\lambda_u}}{1+e^{\lambda_u}}]$$

The interval performs well, except, again, near the boundaries. Brown et al.(2001) also point out that its expected length is actually longer than Clopper-Pearson.

### Likelihood ratio interval

As its name indicates, this method is based on the likelihood ratio test which accepts the null hypothesis $p = p_0$ if $-2\ln(\Lambda_n) \leq \kappa^2$, where $\Lambda_n$ is the likelihood ratio. Brown et al.(2001) note that this interval has nice properties but is difficult to compute, and thus do not consider it further.

### Zhou-Li interval

Zhou et al.(2008) suggest yet another alternative confidence interval for binomial proportions, which they refer to as the Zhou-Li interval, or ZL interval. This interval was arrived at through the consideration that the logit transform of $\hat{p}$, $\log(\hat{p}/\hat{q})$, is closer to the normal distribution than $\hat{p}$. Thus they use the following standardized logit transform as a statistic:

$$T = \sqrt{n\hat{p}\hat{q}}\big(\log\big(\frac{\hat{p}}{\hat{q}}\big) - \log\big(\frac{p}{q}\big)\big)$$

The authors then derive the Edgeworth expansion for the CDF of $T$, noting that the skewness term, of order $n^{-1/2}$, is a source of bias that can be eliminated. They thus use an idea of Hall (1992) to transform $T$ monotonically by a function $g$ that will remove the skewness and hence make the distribution of $g(T)$ virtually symmetric. Having done this, they proceed to form a two-sided interval for $g(T)$ using the normal approximation, then use the inverse of $g$ to obtain the interval for $T$, and thus the logit of $\hat{p}$. Finally they take the anti-logit transform to get the ZL interval:

$CI_{ZL} =$

$$\left[\frac{\exp\left(\log\left(\frac{\hat{p}}{\hat{q}} - n^{-1/2}(\hat{p}\hat{q})^{-1/2}g^{-1}(z_{1-\alpha/2})\right)\right.}{1+\exp\left(\log\left(\frac{\hat{p}}{\hat{q}} - n^{-1/2}(\hat{p}\hat{q})^{-1/2}g^{-1}(z_{1-\alpha/2})\right)\right.}, \frac{\exp\left(\log\left(\frac{\hat{p}}{\hat{q}} - n^{-1/2}(\hat{p}\hat{q})^{-1/2}g^{-1}(z_{\alpha/2})\right)\right.}{1+\exp\left(\log\left(\frac{\hat{p}}{\hat{q}} - n^{-1/2}(\hat{p}\hat{q})^{-1/2}g^{-1}(z_{\alpha/2})\right)\right.}\right]$$

where $g^{-1}$ is the inverse of the monotone transformation. Note this interval is always between 0 and 1. (See Zhou et al.(2008) for technical details).

Through simulation the authors show that this interval is comparable to the Agresti-Coull interval in conservativeness, but it's expected width is shorter, especially so when the sample size is small. The authors also give evidence to suggest that the ZL interval has better coverage accuracy than the Wilson or Jefferys intervals, particularly for $p$ values near 0 or 1, and is otherwise comparable with these two.

## Note on One-Sided Confidence Intervals

Cai(2005) studies in a very similar fashion the problem of one-sided confidence intervals for binomial models. Edgeworth expansions are used again to show that not only the Wald, but the Wilson interval performs poorly in one-sided confidence interval estimation. Cai recommends the the Jefferys interval and an interval he calls the "second-order corrected interval", and notes these provide significant improvement over the formerly mentioned. The second-order corrected interval makes use of Edgeworth expansions to eliminate systematic bias in coverage. The performance of these two intervals is very similar.

Kott and Liu(2007) also study one-sided confidence intervals for binomial proportions. They conclude that the Cai second-order corrected interval above, as well as an interval they propose (Kott-Liu interval) perform best for one-sided binomial intervals. This interval is not decribed here.

# 5 Comparison of Confidence Intervals under Constrained Inference

Based on the preceding discussion, and in the context of default estimation, we will compare the Agresti-Coull (AC), Wilson (or Score), Jefferys, and Zhou-Li (ZL) two-sided 95% confidence intervals under two different constrained estimation strategies. Here we will consider just two proportions. The first estimation strategy is simply the PAVA method described above. The second strategy, called Strategy 2 here, is similar to PAVA in that the sample proportions are the estimates as long as they satisfy the constraints. The second strategy differs however by making both estimates equal to the higher estimated proportion when the constraints are violated. This could be considered a more conservative approach than PAVA in the case of default estimation, since we are cautious not to underestimate a default probability. Both strategies are weighted averages of the two estimates, with PAVA giving weights proportional to the number of observations for each estimate, and Strategy 2 giving all weight to the higher estimate.

For comparison purposes, we use the coverage probability (CP) and expected length (EL) of the different intervals. Thus for our two proportion scenario, we would have the coverage probability of the first interval given as:

$$CP_1 = \sum_{x_2=0}^{n_2} \sum_{x_1=0}^{n_1} I(p_1 \in CI_1) \binom{n_2}{x_2} p_2^k (1-p_2)^{n-k} \binom{n_1}{x_1} p_1^k (1-p_1)^{n-k}$$

where $I(p_1 \in CI_1)$ is the indicator function indicating whether $p_1$ is contained in confidence interval 1, $CI_1$. A similar definition applies for $CP_2$. Also, the EL for both intervals is calculated in a similar way by using the interval length in the above formula instead of the indicator function. Thus the EL of interval 1 would be:

$$EL_1 = \sum_{x_2=0}^{n_2} \sum_{x_1=0}^{n_1} (U_1 - L_1) \binom{n_2}{x_2} p_2^k (1-p_2)^{n-k} \binom{n_1}{x_1} p_1^k (1-p_1)^{n-k}$$

where $U_1$ and $L_1$ are the upper and lower endpoints, respectively, of interval 1.

If the estimates satisfy the constraints, then the confidence intervals are formed separately as described in the section above. If not, the estimates are formed according to either PAVA or Strategy 2. The intervals are then formed with these modified estimates, but with their original number of observations. For example, under PAVA, if $p_1 \leq p_2$, but $\hat{p}_1 = \frac{3}{20}$ and $\hat{p}_2 = \frac{1}{10}$, then both intervals are formed with the estimate $\frac{1+3}{10+20} = \frac{4}{30}$, but the interval for $p_1$ is formed with $n = 20$ and the interval for $p_2$ with $n = 10$. Similarly for Strategy 2.

For comparison purposes, we used the following default probability estimates from Standard & Poors (available at www.standardandpoors.com):

| Rating | Default Prob. |
|---|---|
| AAA | 0.0 |
| AA | 0.00004 |
| A | 0.00008 |
| BBB | 0.0027 |
| BB | 0.0096 |
| B | 0.0459 |
| C | 0.2758 |

The estimates were based on data from 1981 to 2011. The lowest number of issuers among classes was 93, which was AAA. The rest were well over 100. We chose the number 100 for both classes in each comparison for computational reasons. (A small sample of larger numbers showed the same comparative effects.)

Since the default estimate for class AAA was 0, we chose to start the comparisons with AA and A. For both strategies, and for each two adjacent classes beginning with AA and A, we use the table above to first fix the lower probability, $p_1$, and then vary $p_2$, plotting and looking at the coverage probability and expected length for both class intervals using the AC, Score, Jefferys, and ZL methods. We then do the same for fixed $p_2$ and varying $p_1$. (When $p_1$ is fixed, $p_2$ varies from slightly more than p1 to slightly more than the default probability for the corresponding class in the table. When $p_2$ is fixed, $p_1$ varies from slightly less than it's value in the table to slighly less than $p_2$). Refer to the graphs at the end of this section.

Over all the comparisons, the expected length maintained a consistent ranking, with the Jefferys interval producing the lowest EL, followed by the Score, ZL, and AC respectively. We thus keep this in mind and focus on the coverage probabilities. One other general observation from the comparisons is that the CP for the fixed probability class is nearly constant or slowly changing as the other probability varies. It is very close to it's independent value.

When looking at the first two classes, AA and A, we first fix $p_1$ and notice the Score and Jefferys interval both underestimate the coverage probability for $p_2$, for both PAVA and Strategy 2. Although overly conservative, the ZL interval is slighltly less conservative than the AC. The coverage probability of the ZL interval is virtually identical here between PAVA and Strategy 2. Now fixing $p_2$, we note that none of the intervals CP falls below .95 for $p_1$, but the Jefferys and Score are consistently below .95 for $p_2$. This is consistent with the work cited in the previous section and the probability values considered. Strategy 2 is virtualy identical to PAVA on the AC and ZL intervals and is only slightly more conservative for some probabilities with the Score and Jefferys. There is no strong evidence to recommend one strategy over the other here. It should also be noted that although the Jefferys interval falls below .95 for some probability values, it never falls below .90. This might be considered along with the Jefferys smaller EL.

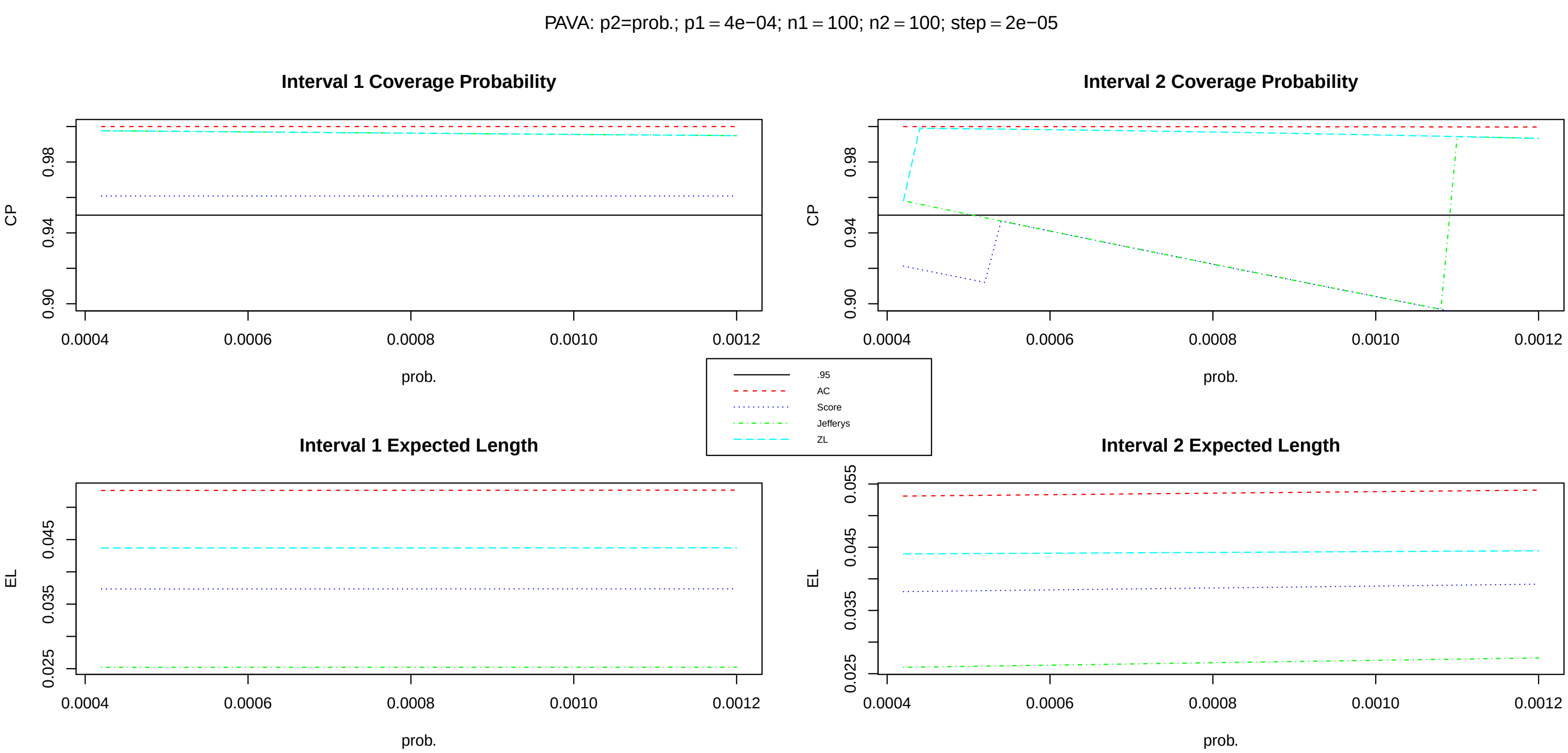
The case for the next two classes, A and BBB, is the same. There is virtually no difference between PAVA and Strategy 2, with PAVA CPs being slightly closer to the nominal .95 for the Jefferys and Score intervals with $p_2$ fixed. With fixed $p_1$, the Jefferys and Score intervals underestimate the coverage probability of $p_2$, although again the Jefferys does not fall below .90.

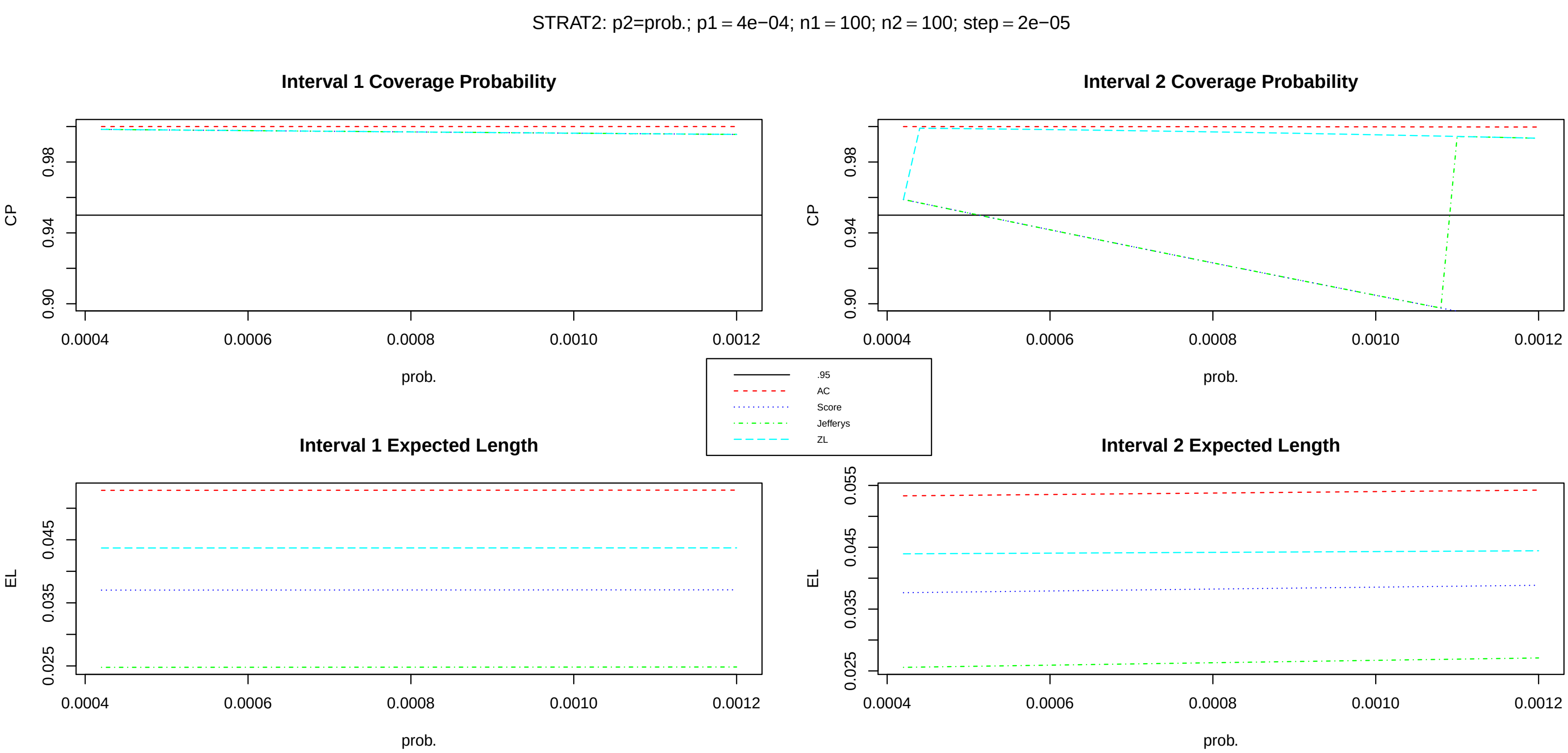
For the next two classes, BBB and BB, the PAVA and Strategy 2 again give similar results. In this case, the Score interval again underestimates the CP for some probabilities, but never falls below .90. The Jefferys interval only slighly underestimates CP and given it's low EL should be considered along with the ZL, which maintains CP very close to Jefferys throughout, but never falls below nominal.

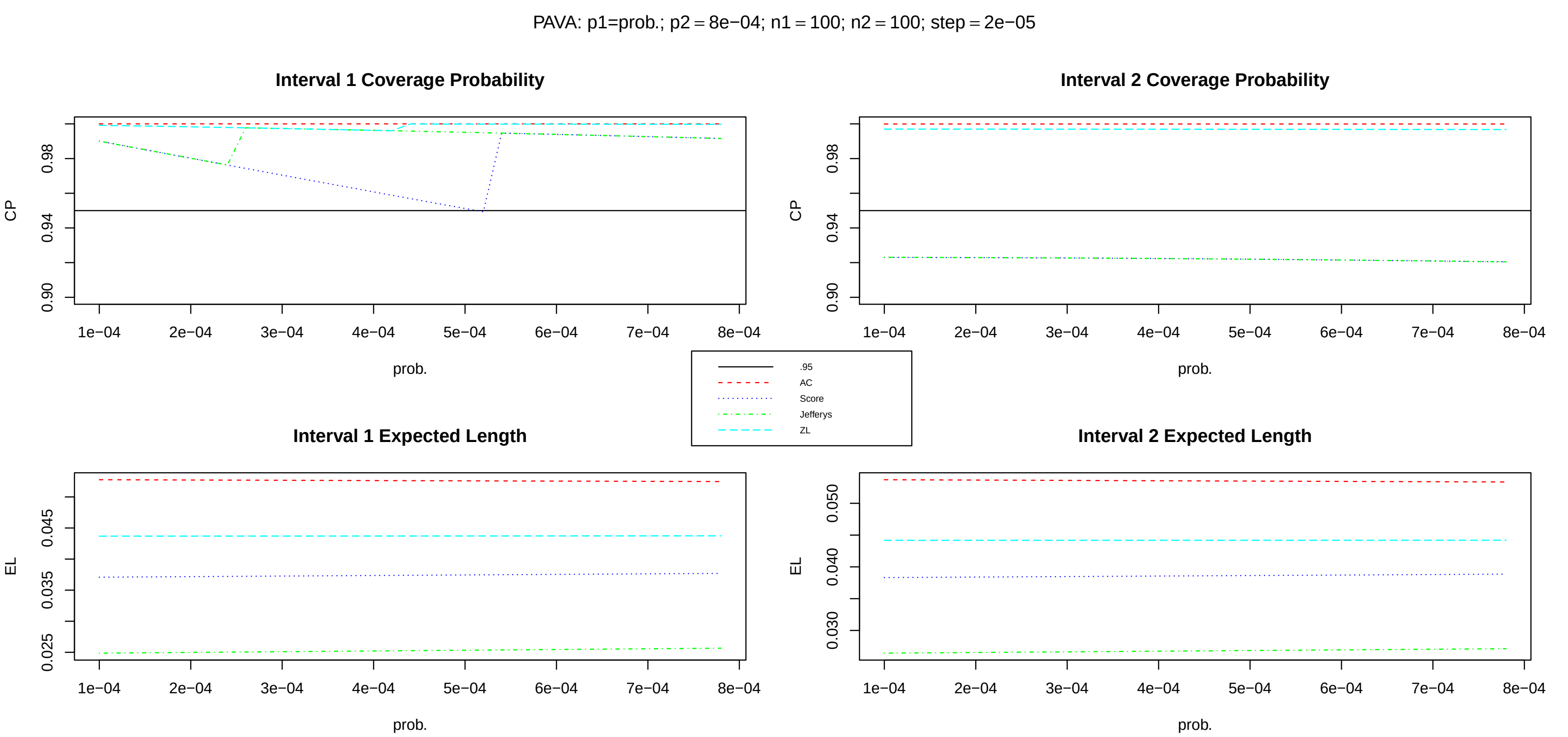
For BB and B, PAVA and Strategy 2 show slight differences. With $p_1$ fixed and $p_2$ varying, the Jefferys interval has better coverage probability for PAVA, but with $p_2$ fixed, Jefferys has undercoverage for both strategies. The ZL interval again performs well, although there are a few instances of undercoverage.

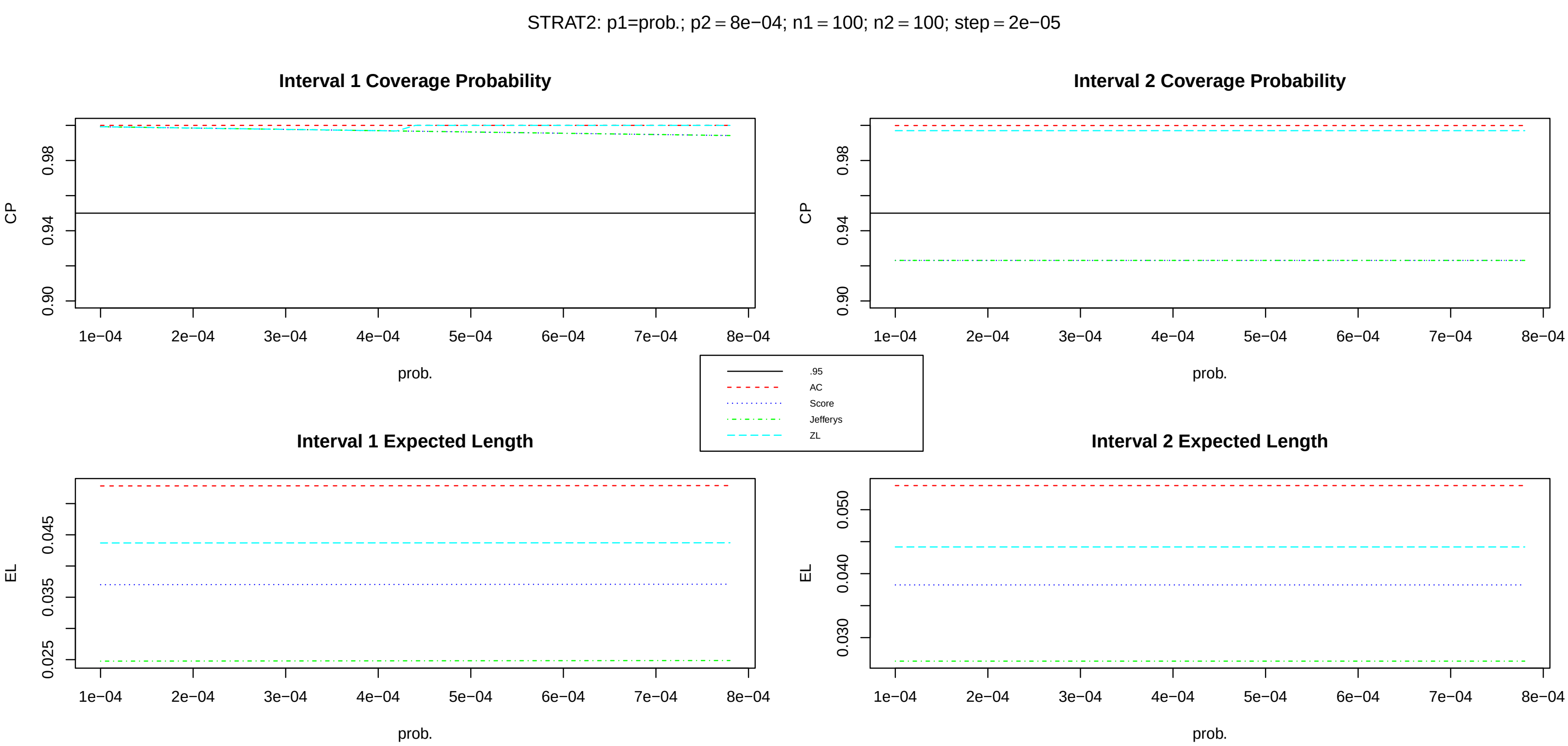
For the last classes, B and C, PAVA and Strategy 2 again perform very similarly for all intervals considered, although PAVA appears slightly less erratic. Again, the Jefferys and Score show some undercoverage, as does the ZL. The AC performs well here, never falling below the nominal .95 and staying very close.

Overall, the results are not surprising considering the analysis in the literature cited above. There is somewhat of a tradeoff between EL and CP. The ZL interval tends to perform better on the small probability scales than the Jefferys and Score, and the AC seems too conservative on these scales. However, given the lower EL of the Jefferys interval and the fact that it does not fall below .90 on coverage probability, it might be considered. As for PAVA and Strategy 2, they perform very similarly, with PAVA being slightly less conservative and less erratic.

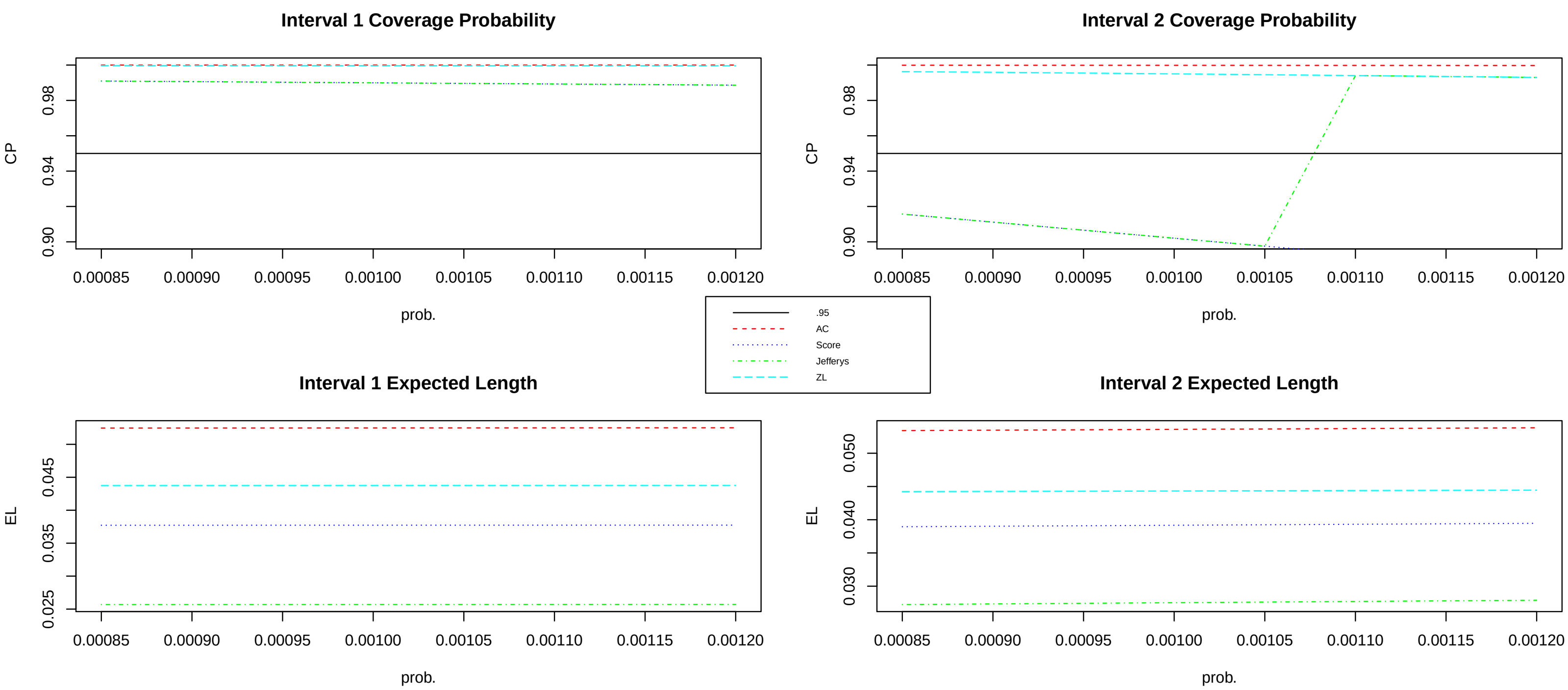

PAVA: p2=prob.; p1 = 4e−04; n1 = 100; n2 = 100; step = 2e−05
Interval 1 Coverage Probability
Interval 2 Coverage Probability
CP
0.90
0.94
0.98
0.0004
0.0006
0.0008
0.0010
0.0012
prob.
.95
AC
Score
Jefferys
ZL
Interval 1 Expected Length
Interval 2 Expected Length
EL
0.025
0.035
0.045
0.055

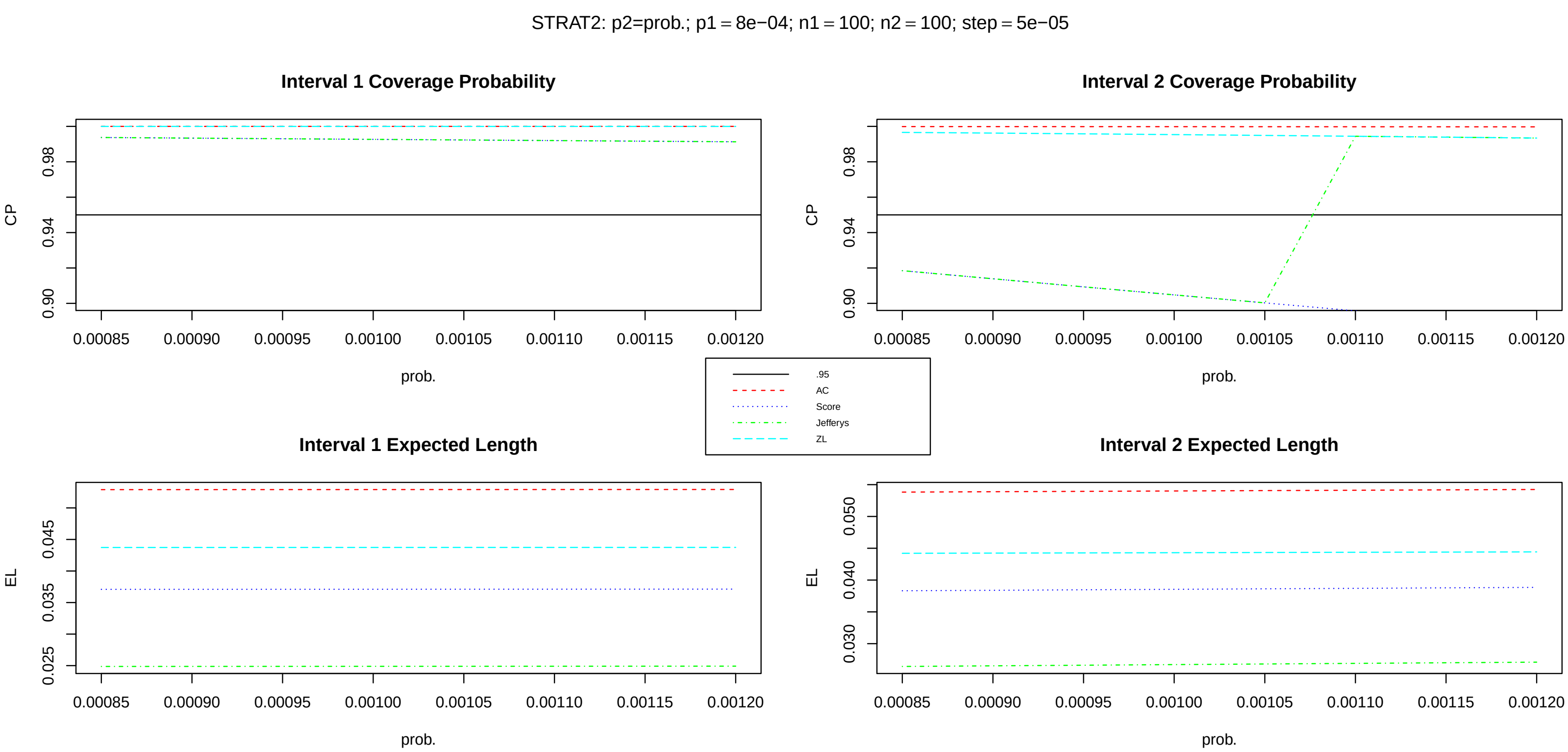

STRAT2: p2=prob.; p1 = 4e−04; n1 = 100; n2 = 100; step = 2e−05
Interval 1 Coverage Probability
Interval 2 Coverage Probability
Interval 1 Expected Length
Interval 2 Expected Length
CP
EL
prob.
.95
AC
Score
Jefferys
ZL

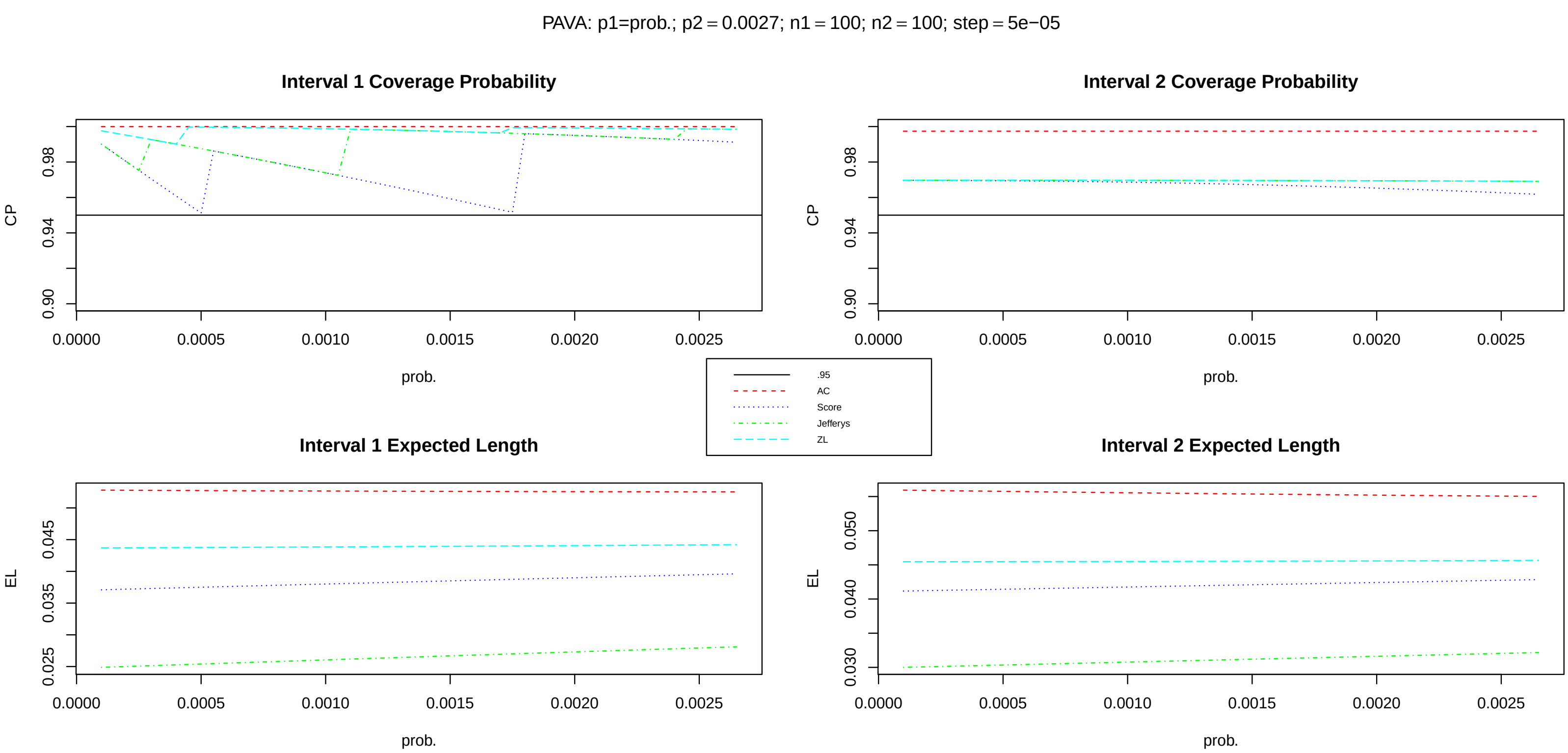

PAVA: p1=prob.; p2 = 8e−04; n1 = 100; n2 = 100; step = 2e−05
Interval 1 Coverage Probability
Interval 2 Coverage Probability
Interval 1 Expected Length
Interval 2 Expected Length
CP
EL
prob.
.95
AC
Score
Jefferys
ZL

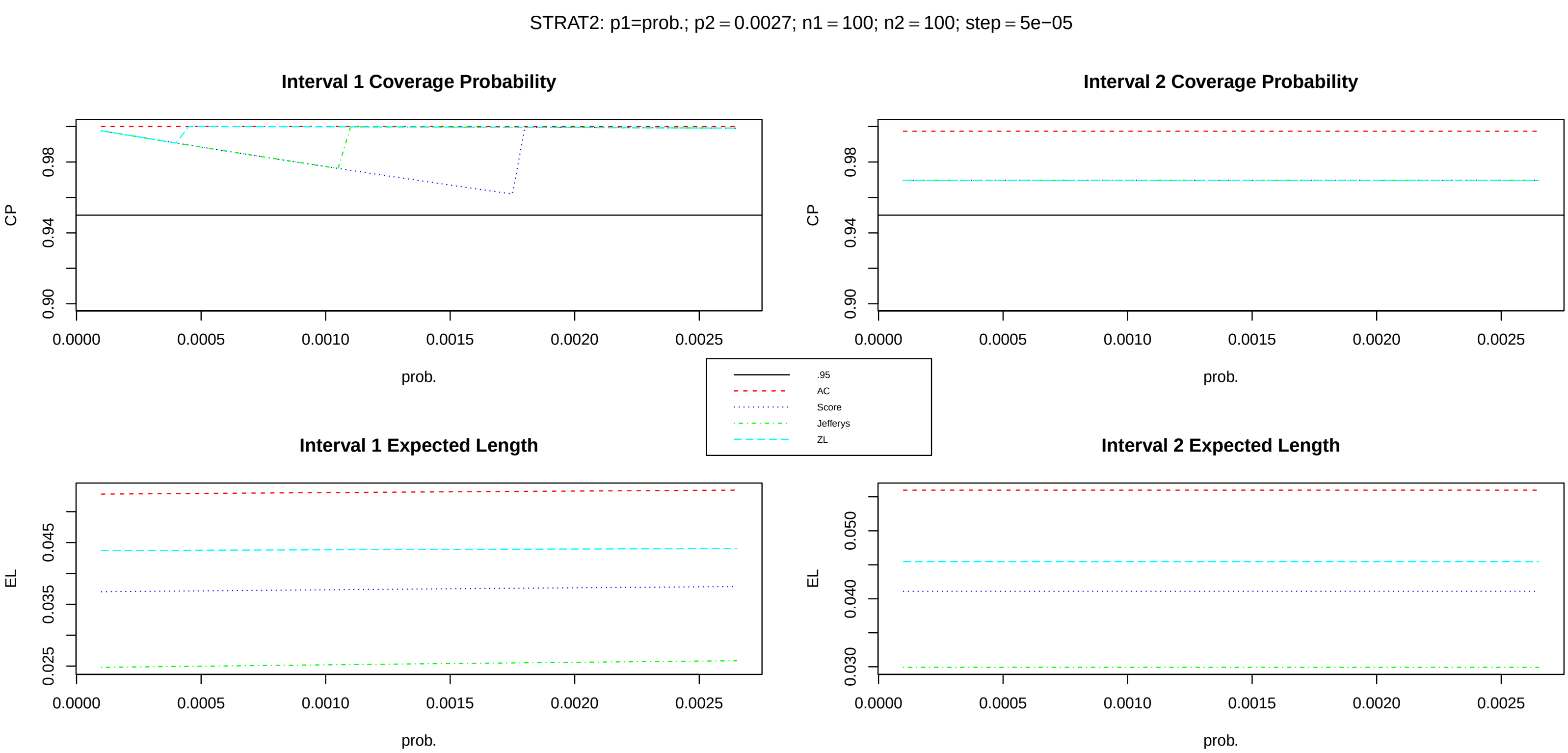

STRAT2: p1=prob.; p2 = 8e−04; n1 = 100; n2 = 100; step = 2e−05
Interval 1 Coverage Probability
Interval 2 Coverage Probability
Interval 1 Expected Length
Interval 2 Expected Length
CP
EL
prob.
.95
AC
Score
Jefferys
ZL

PAVA: p2=prob.; p1 = 8e−04; n1 = 100; n2 = 100; step = 5e−05

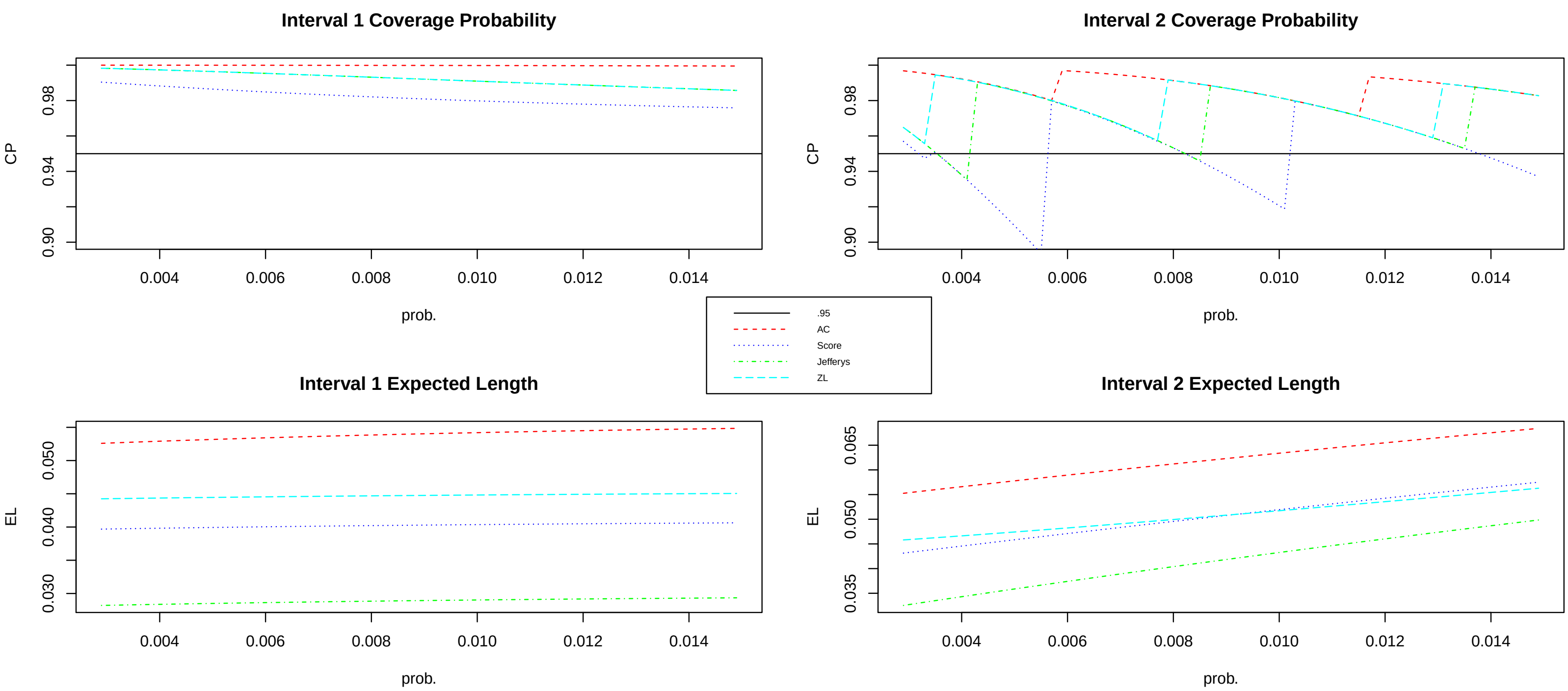



STRAT2: p2=prob.; p1 = 8e−04; n1 = 100; n2 = 100; step = 5e−05
Interval 1 Coverage Probability
Interval 2 Coverage Probability
Interval 1 Expected Length
Interval 2 Expected Length
CP
EL
prob.
.95
AC
Score
Jefferys
ZL

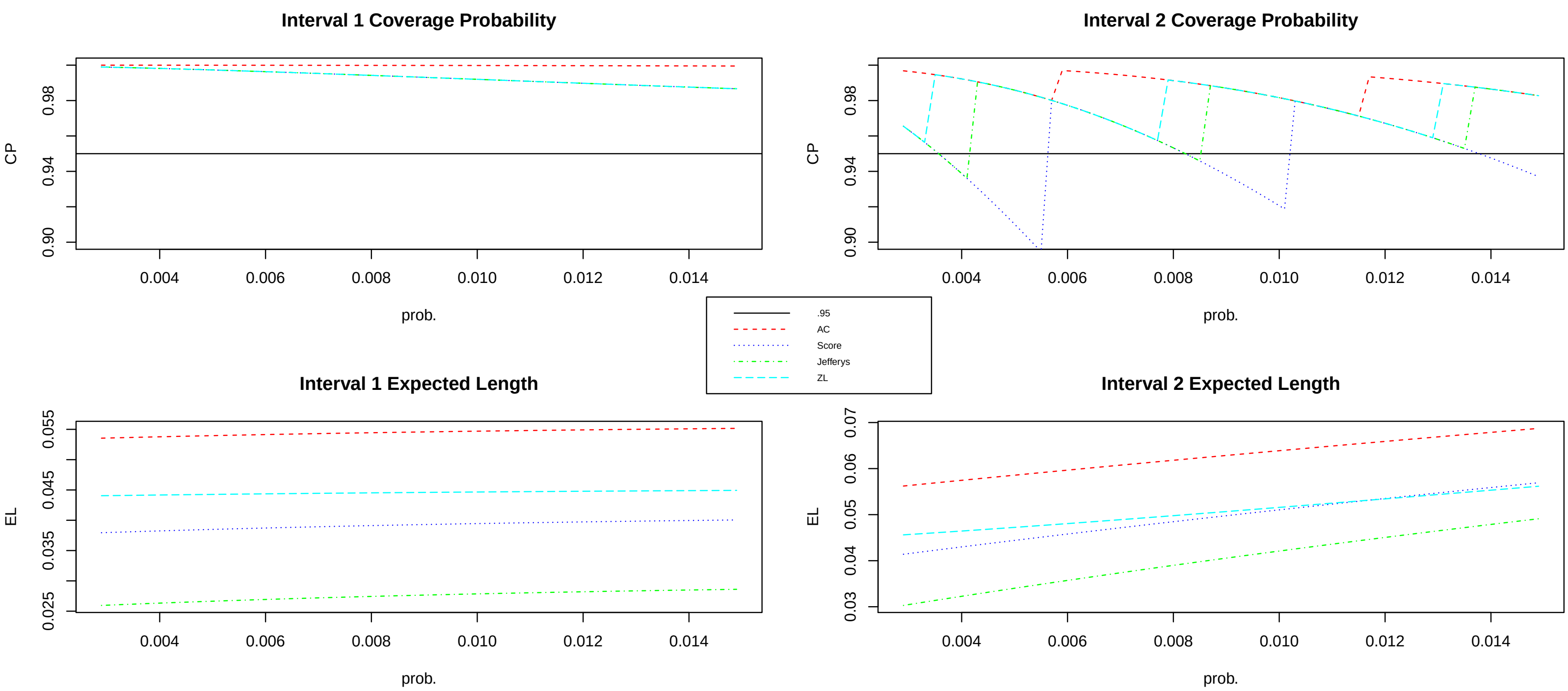

PAVA: p1=prob.; p2 = 0.0027; n1 = 100; n2 = 100; step = 5e−05
Interval 1 Coverage Probability
Interval 2 Coverage Probability
CP
0.90
0.94
0.98
0.0000
0.0005
0.0010
0.0015
0.0020
0.0025
prob.
.95
AC
Score
Jefferys
ZL
Interval 1 Expected Length
Interval 2 Expected Length
EL
0.025
0.035
0.045
0.030
0.040
0.050


STRAT2: p1=prob.; p2 = 0.0027; n1 = 100; n2 = 100; step = 5e−05
Interval 1 Coverage Probability
Interval 2 Coverage Probability
Interval 1 Expected Length
Interval 2 Expected Length
CP
EL
prob.
0.0000
0.0005
0.0010
0.0015
0.0020
0.0025
0.90
0.94
0.98
0.025
0.035
0.045
0.030
0.040
0.050
.95
AC
Score
Jefferys
ZL

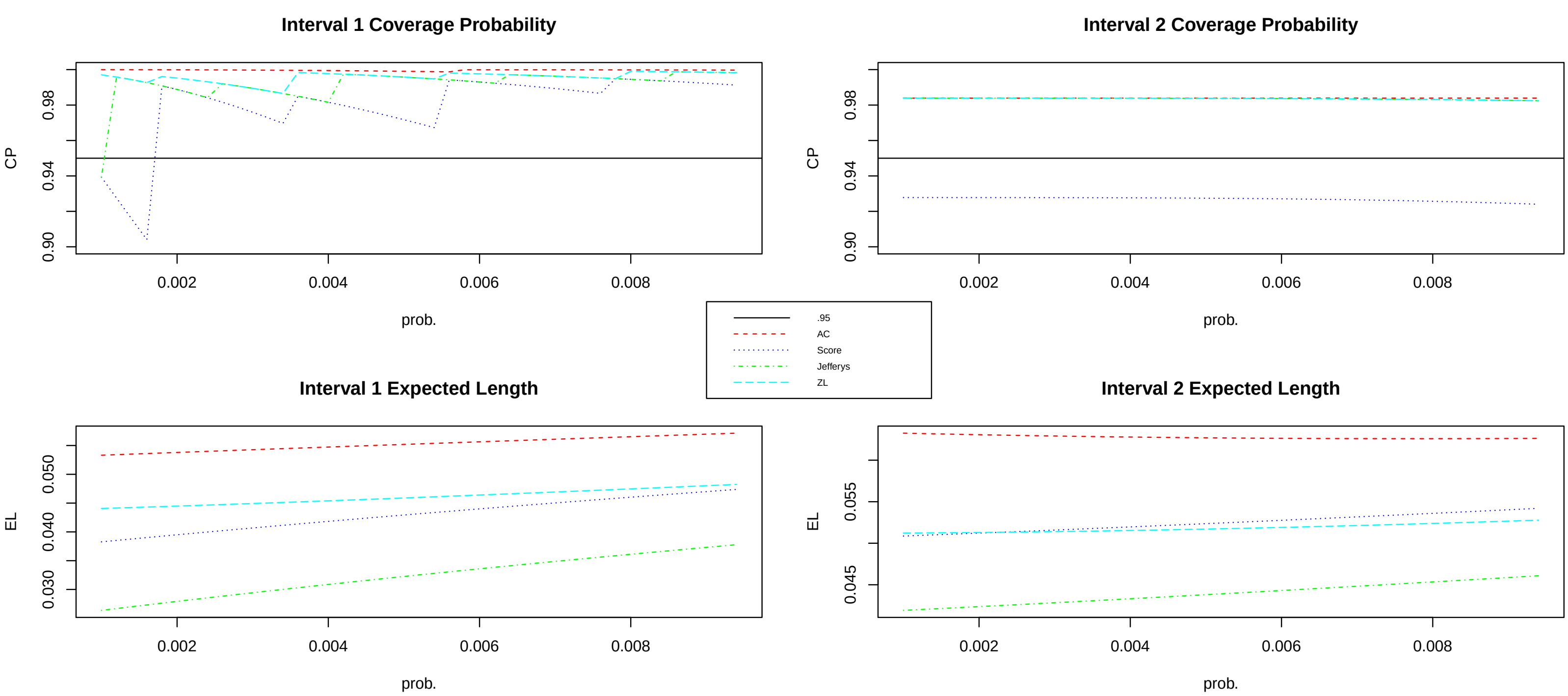
PAVA: p2=prob.; p1 = 0.0027; n1 = 100; n2 = 100; step = 2e−04

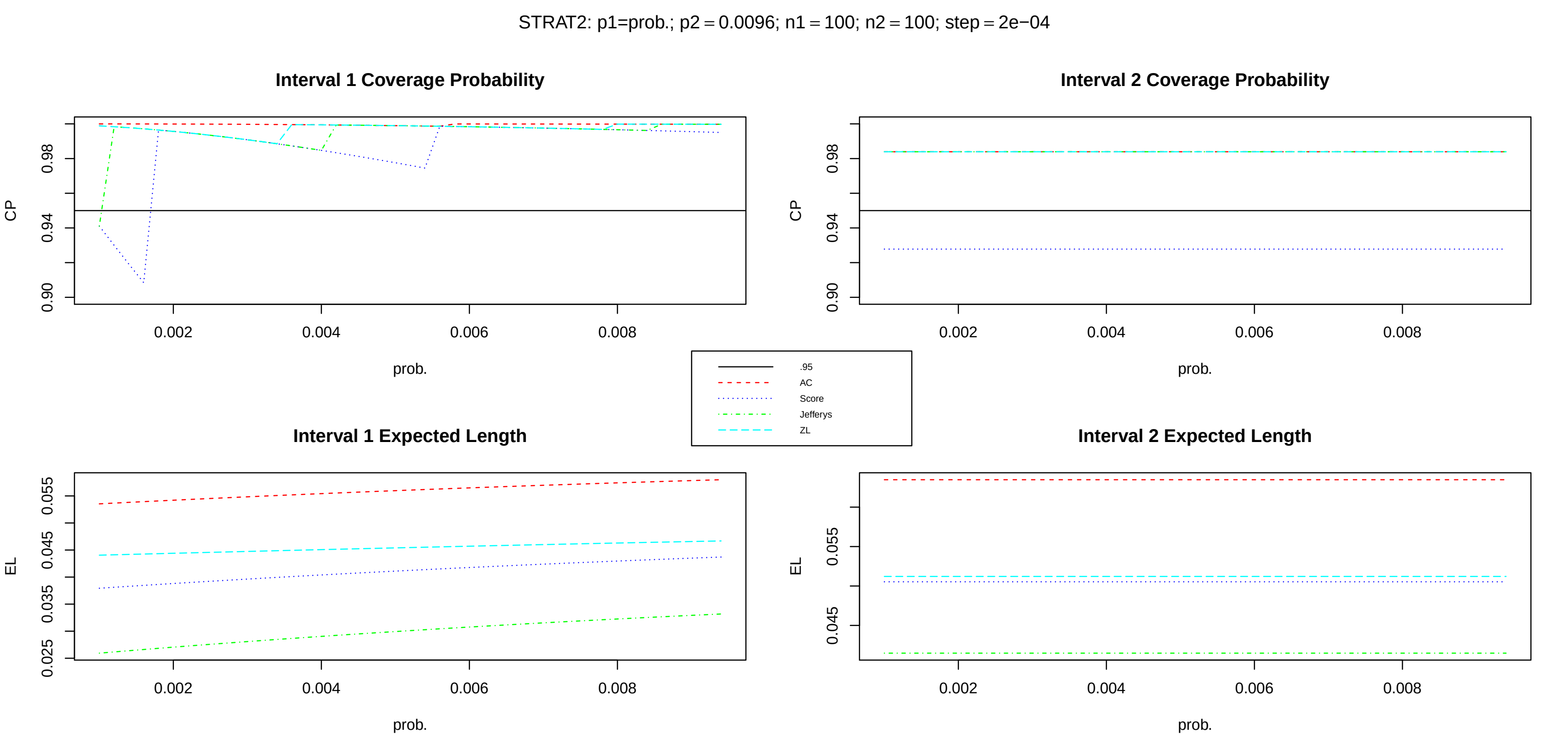


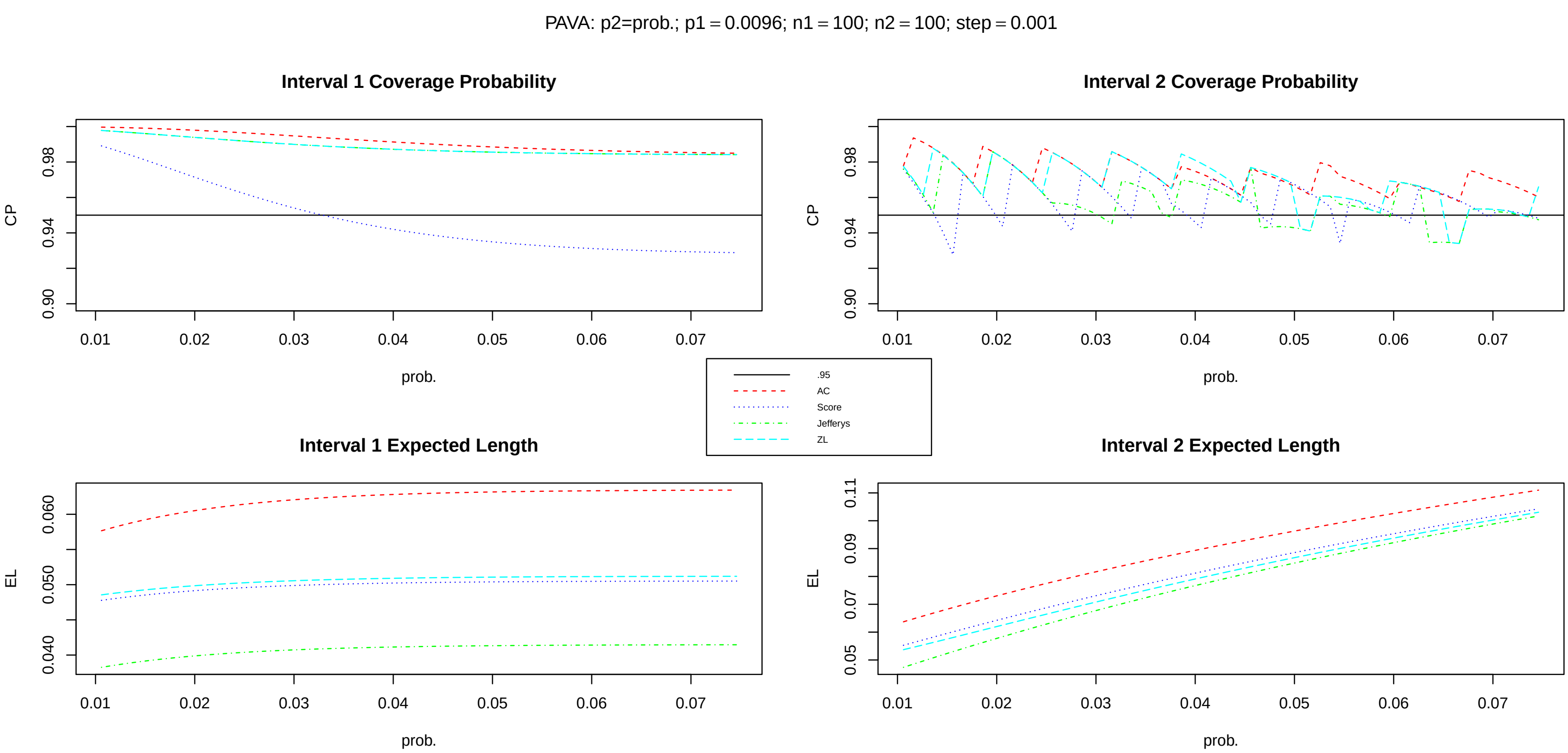
STRAT2: p2=prob.; p1 = 0.0027; n1 = 100; n2 = 100; step = 2e−04

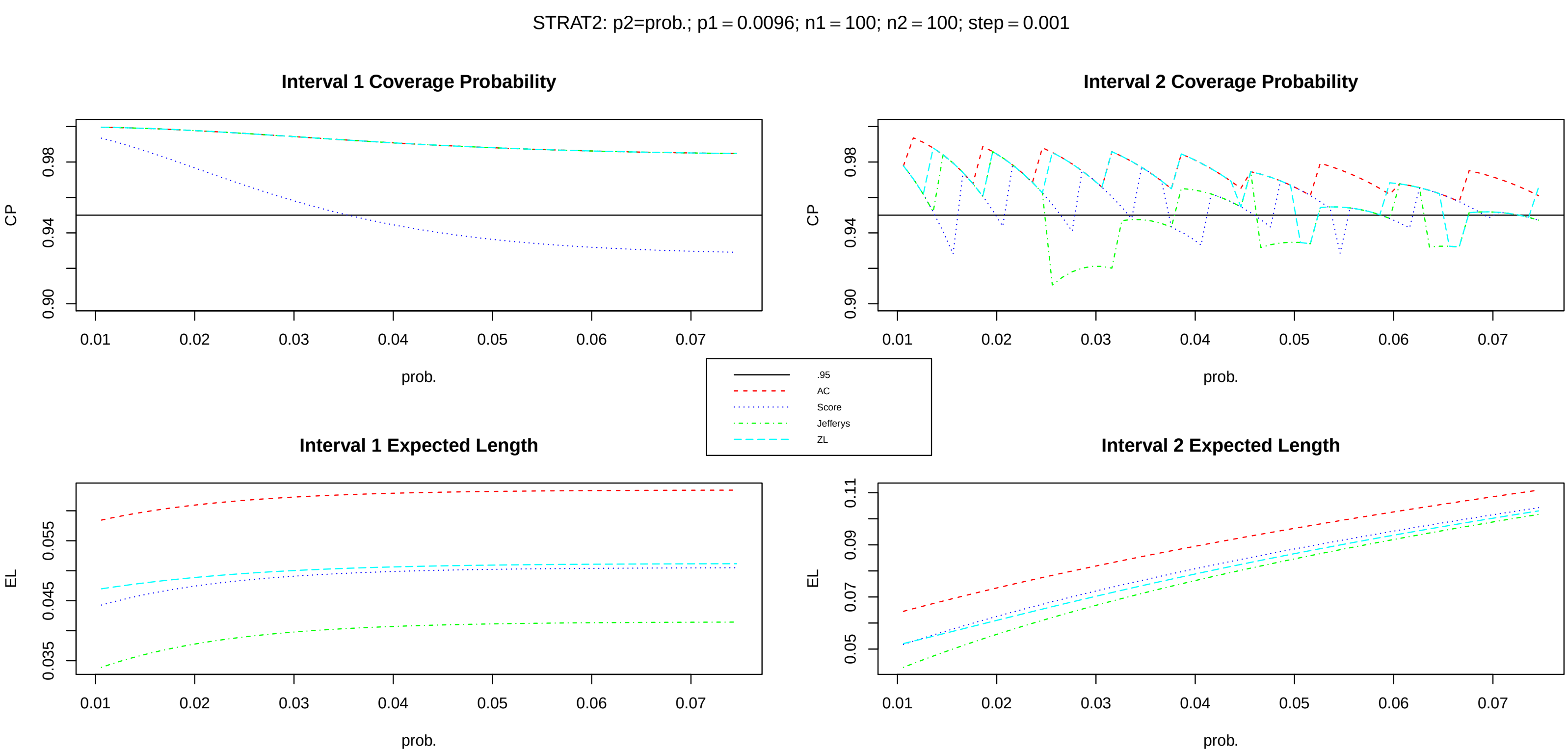


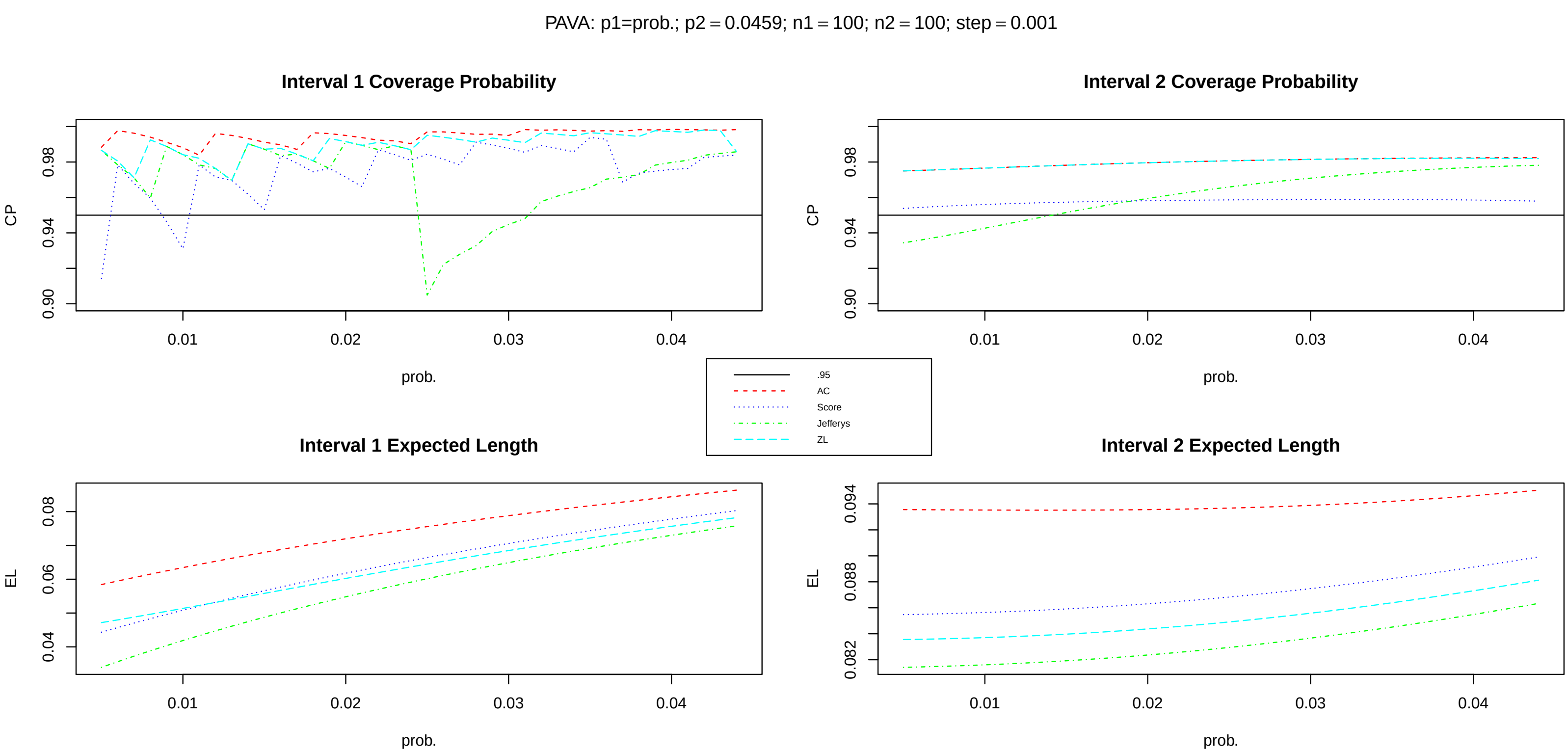
PAVA: p1=prob.; p2 = 0.0096; n1 = 100; n2 = 100; step = 2e−04



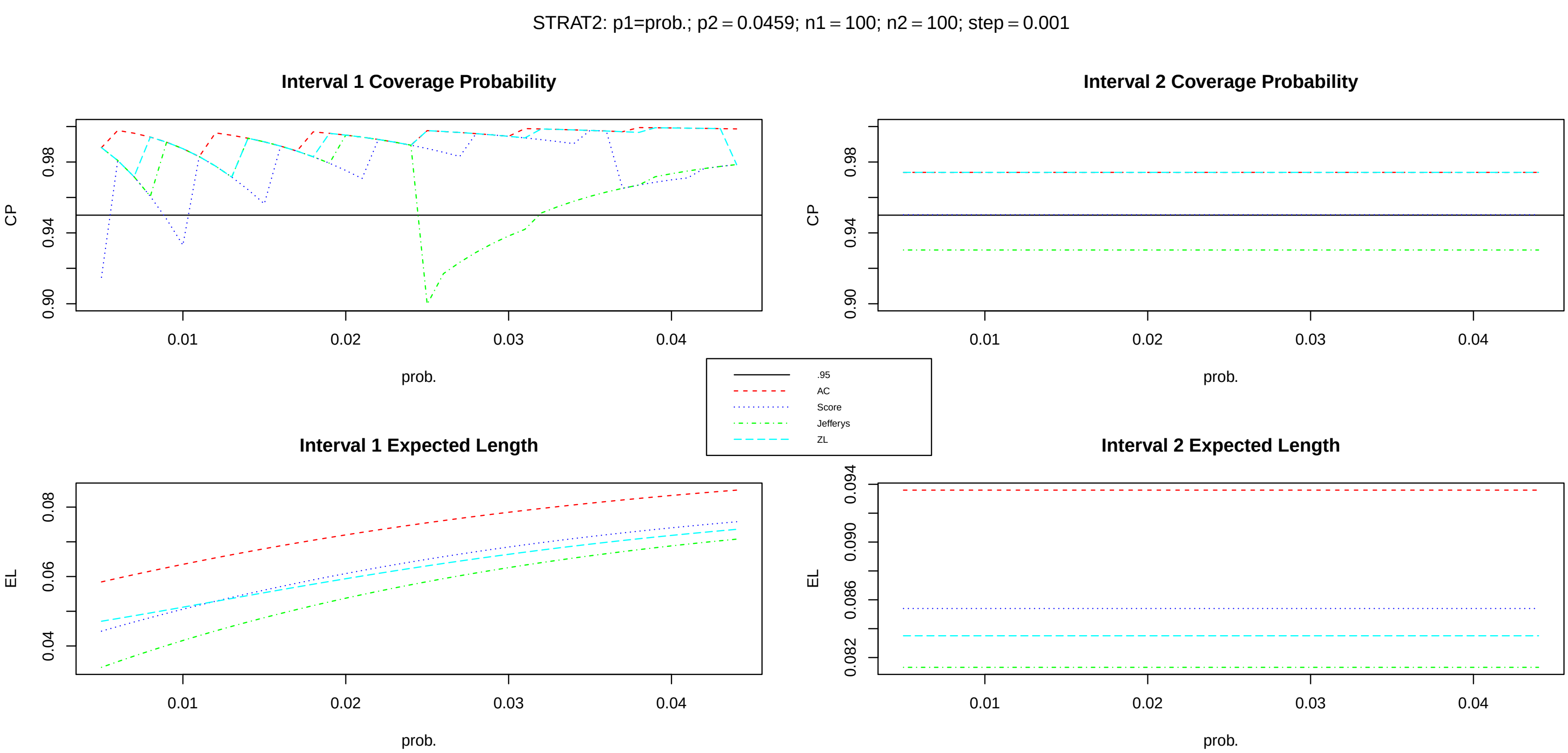

STRAT2: p1=prob.; p2 = 0.0096; n1 = 100; n2 = 100; step = 2e−04
Interval 1 Coverage Probability
Interval 2 Coverage Probability
Interval 1 Expected Length
Interval 2 Expected Length
CP
EL
prob.
0.002
0.004
0.006
0.008
0.90
0.94
0.98
0.025
0.035
0.045
0.055
.95
AC
Score
Jefferys
ZL

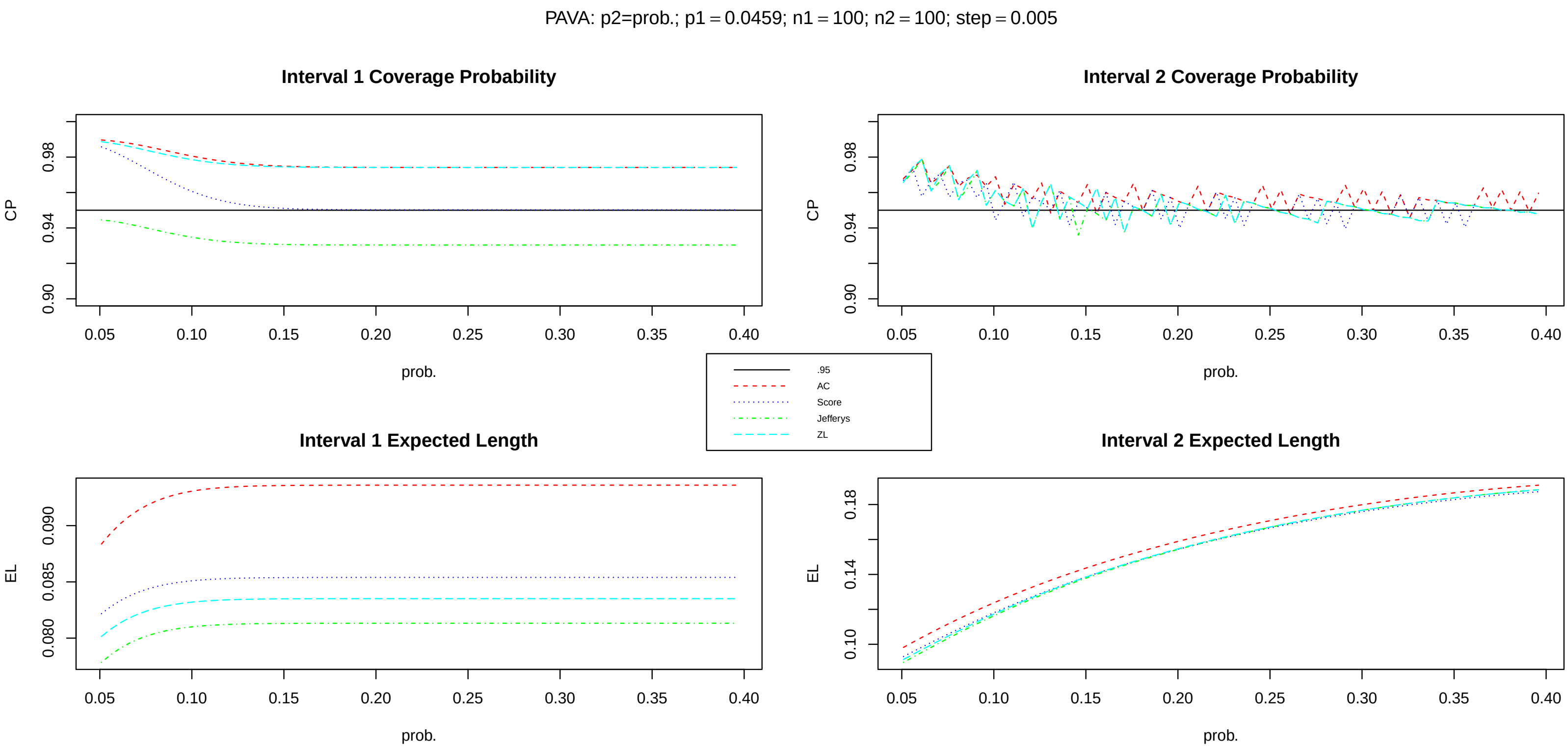

PAVA: p2=prob.; p1 = 0.0096; n1 = 100; n2 = 100; step = 0.001
Interval 1 Coverage Probability
Interval 2 Coverage Probability
Interval 1 Expected Length
Interval 2 Expected Length
CP
EL
prob.
0.90
0.94
0.98
0.040
0.050
0.060
0.05
0.07
0.09
0.11
0.01
0.02
0.03
0.04
0.05
0.06
0.07
.95
AC
Score
Jefferys
ZL


STRAT2: p2=prob.; p1 = 0.0096; n1 = 100; n2 = 100; step = 0.001
Interval 1 Coverage Probability
Interval 2 Coverage Probability
Interval 1 Expected Length
Interval 2 Expected Length
CP
EL
prob.
.95
AC
Score
Jefferys
ZL

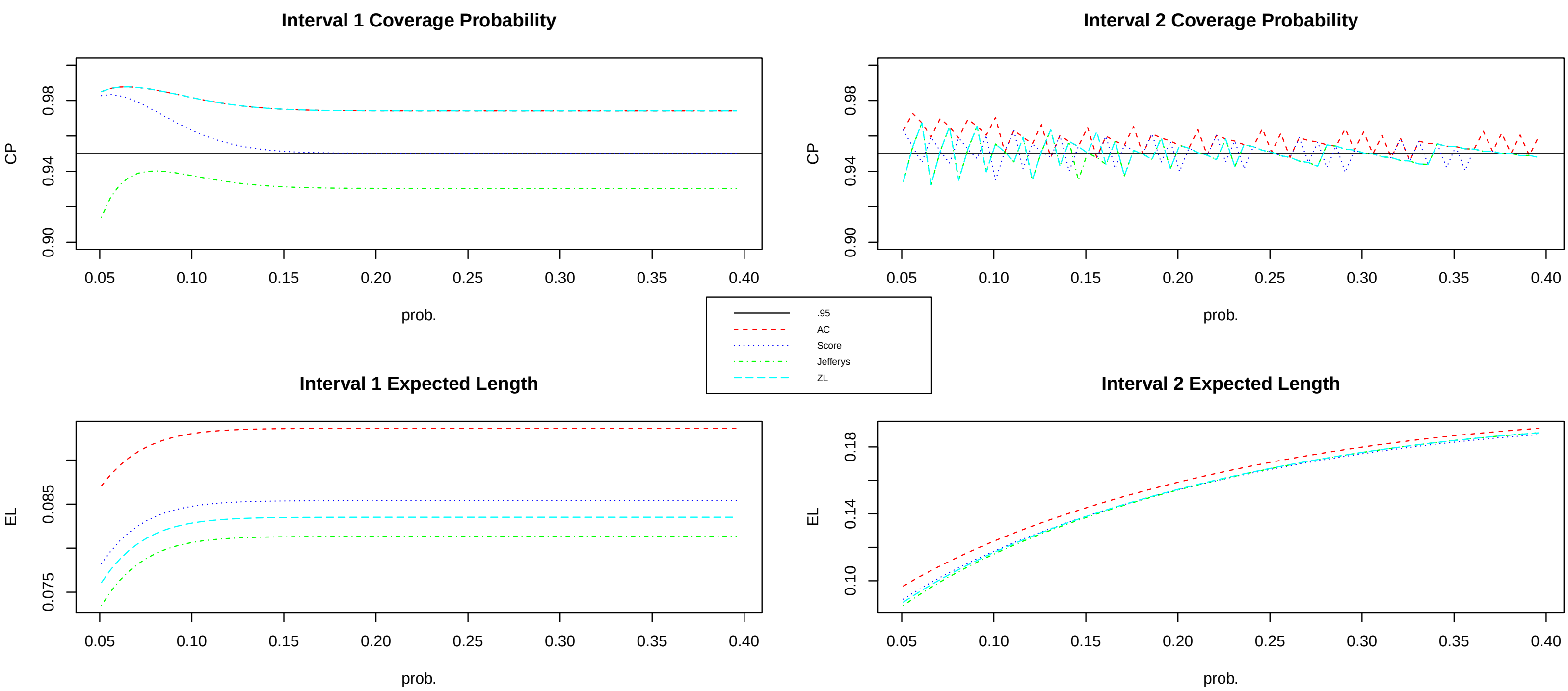


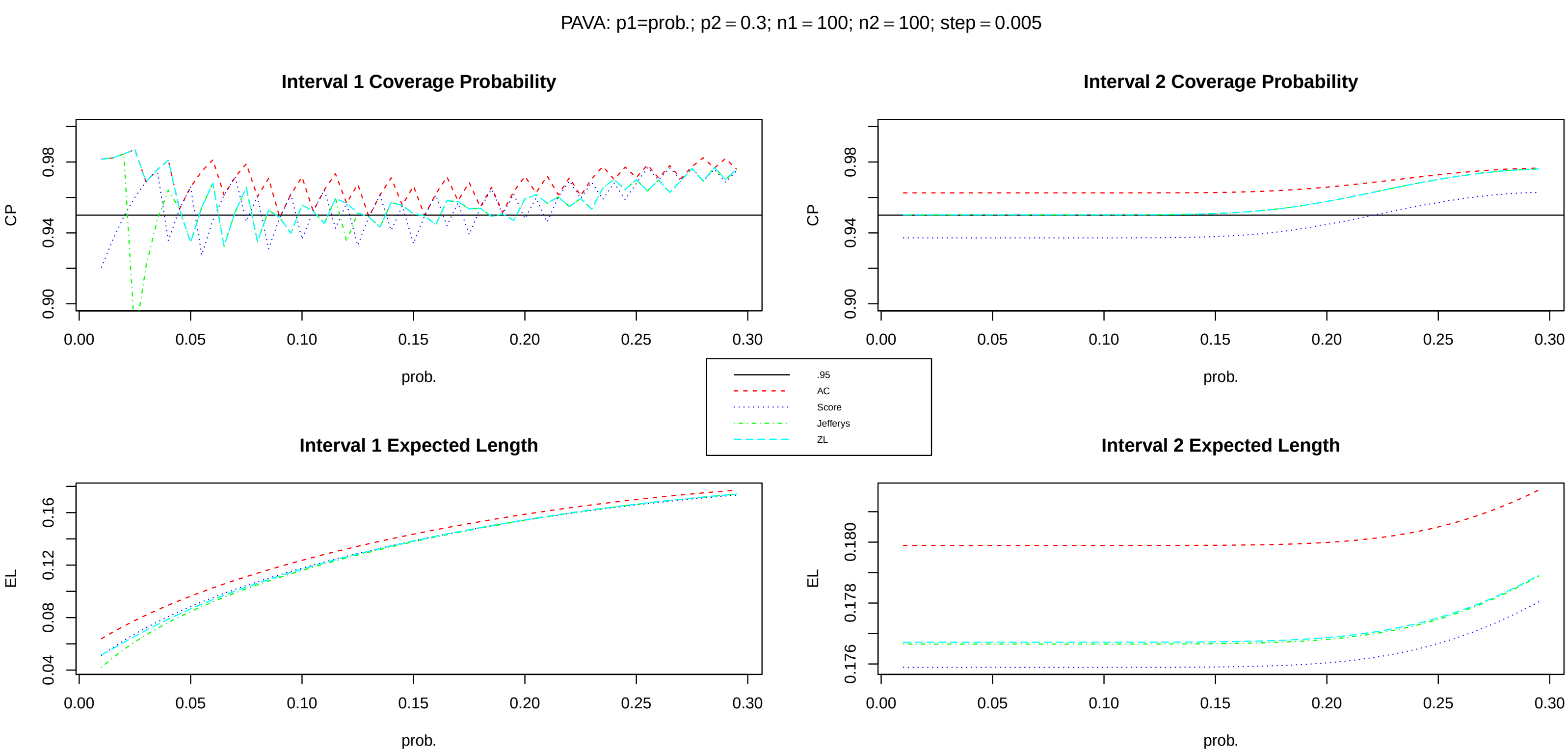
PAVA: p1=prob.; p2 = 0.0459; n1 = 100; n2 = 100; step = 0.001
Interval 1 Coverage Probability
Interval 2 Coverage Probability
CP
0.90
0.94
0.98
prob.
0.01
0.02
0.03
0.04
.95
AC
Score
Jefferys
ZL
Interval 1 Expected Length
Interval 2 Expected Length
EL
0.04
0.06
0.08
0.082
0.088
0.094


STRAT2: p1=prob.; p2 = 0.0459; n1 = 100; n2 = 100; step = 0.001
Interval 1 Coverage Probability
Interval 2 Coverage Probability
CP
0.90
0.94
0.98
0.01
0.02
0.03
0.04
prob.
.95
AC
Score
Jefferys
ZL
Interval 1 Expected Length
Interval 2 Expected Length
EL
0.04
0.06
0.08
0.082
0.086
0.090
0.094

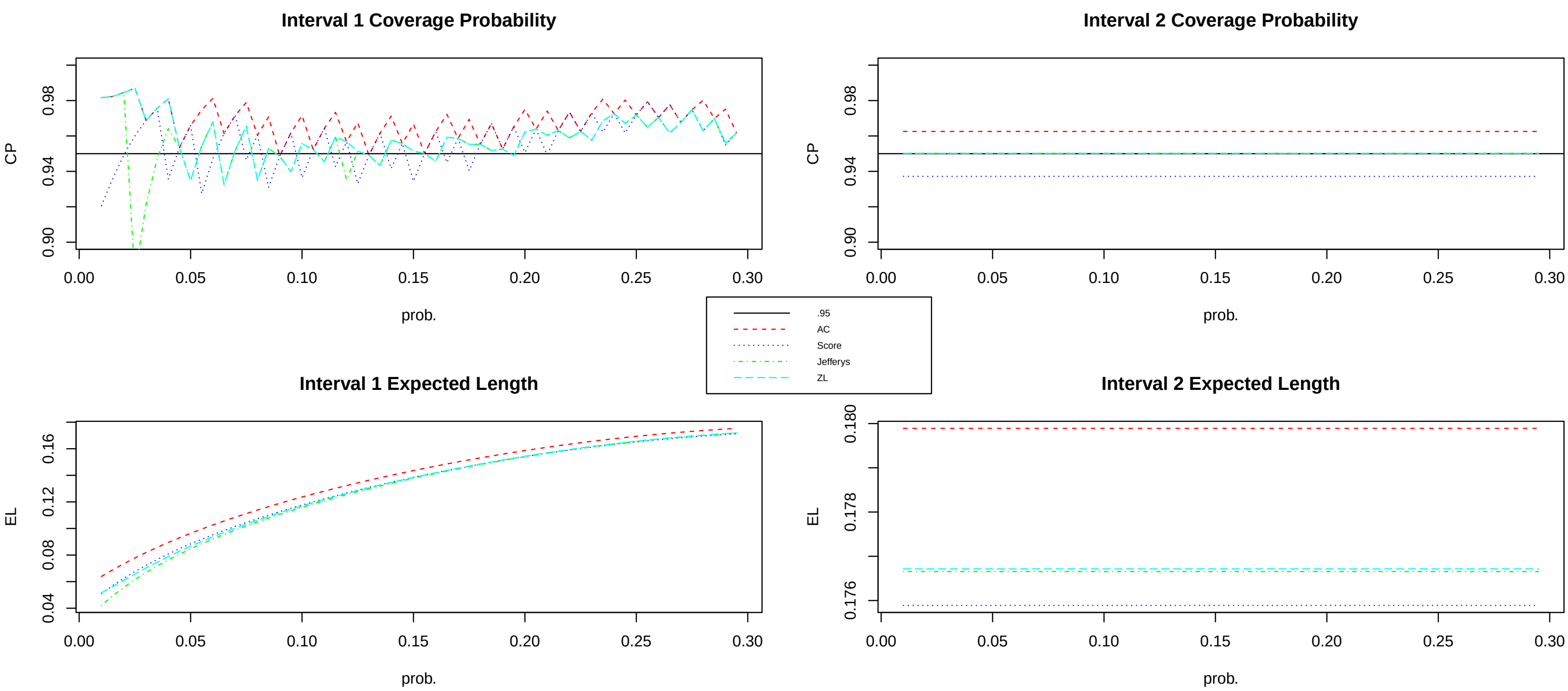

PAVA: p2=prob.; p1 = 0.0459; n1 = 100; n2 = 100; step = 0.005
Interval 1 Coverage Probability
Interval 2 Coverage Probability
Interval 1 Expected Length
Interval 2 Expected Length
CP
EL
prob.
.95
AC
Score
Jefferys
ZL

STRAT2: p2=prob.; p1 = 0.0459; n1 = 100; n2 = 100; step = 0.005




PAVA: p1=prob.; p2 = 0.3; n1 = 100; n2 = 100; step = 0.005
Interval 1 Coverage Probability
Interval 2 Coverage Probability
Interval 1 Expected Length
Interval 2 Expected Length
CP
EL
prob.
.95
AC
Score
Jefferys
ZL

STRAT2: p1=prob.; $p2 = 0.3$; $n1 = 100$; $n2 = 100$; $step = 0.005$



# References

[1] Agresti, A., Coull, B.A., 1998. Approximate is better than "exact" for interval estimation of binomial proportions. American Statistician 52, 119-126.

[2] Ayer, M., Brunk, H.D., Ewing, G.M., Reid, W.T., Silverman, E., 1955. An empirical distribution function for sampling with incomplete information. Ann. Math. Statis., 26, 641-647.

[3] Bhattacharya, R.N., Rao, R.R.,1976. Normal Approximation and Asymptotic Expansions. Wiley, New York.

[4] Brown, L.D., Cai, T., Dasgupta, A., 2001. Interval estimation for a binomial proportion (with discussion). Statistical Science 16, 101-133.

[5] Brown, L.D., Cai, T., Dasgupta, A., 2002. Confidence intervals for a binomial proportion and Edgeworth expansions. Annals of Statistics 30, 160-201.

[6] Cai, T., 2005. One-Sided Confidence Intervals in Discrete Distributions. Journal of Statistical Planning and Inference 131, 63-88.

[7] Casella, G., Berger, R.L., 2001. Statistical Inference. Duxbury Press.

[8] Dasgupta, A., 2008. Asymptotic Theory of Statistics and Probability. Springer.

[9] Ghosh, B.K., 1979. A comparison of some approximate confidence intervals for the binomial parameter. J. Amer. Stat. Assoc. 74, 894-900.

[10] Hall, P. 1992. On the removal of skewness by transformation. J. R. Stat. Soc. B 54, 221-228.

[11] Jiang, J., 2010. Large Sample Techniques for Statistics. Springer.

[12] Lehmann, E., Romano, J. , 2005. Testing Statistical Hypotheses. Springer.

[13] Liu, Kott. 2007. Evaluating alternative one-sided confidence intervals for an extreme binomial proportion. ASA Joint. Stat. Meetings, Section on Survey Research Methods.

[14] Robertson, T., Wright, F.T., Dykstra, R.L., 1988. Order Restricted Statistical Inference. Wiley.

[15] Shapiro, A., 1985. Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints. Biometrika, 72,133-144.

[16] Silvapulle, M., Sen, P., 2005. Constrained Statistical Inference. Wiley-Interscience.

[17] Wasserman, L., 2004. All of Statistics. Springer.

[18] Zhou, X.H., Li, C.M., Yang, Z., 2008. Improving interval estimation of binomial proportions. Phil.Trans. of the Royal Society 366,2405-2418.