Classroom Use of R: Coverage Probabilities of Poisson Interval Estimates

Lioudmila Belan, Eric Suess, Bruce Trumbo
Department of Statistics and Biostatistics; California State University, East Bay; Hayward, CA 94542
mila.belan@gmail.com, eric.suess@csueastbay.edu, bruce.trumbo@csueastbay.edu

Abstract
True coverage probabilities of several nominal 95% and 99% interval estimates of the Poisson mean are computed using elementary programs in R. Among the intervals considered are frequentist confidence intervals based on normal approximation and Bayesian posterior probability intervals resulting from a non-informative prior.

Key Words: Confidence interval, Bayesian probability interval, Poisson mean, R/S-Plus, Pedagogy

1. Introduction

As a programming language, R is becoming increasingly popular. It is freely available and it provides broad variety of methods for statistical computations and representations. In addition, its libraries are expanding with contributions from researchers from all around the globe with new modules implementing most contemporary statistical theories and techniques. As a result it is crucial for students of Statistics to be familiar with programming practice in R and able to manipulate data using R.

The goal of the project described in the article is to demonstrate:

- Application of elementary data manipulation tools in R such as vector operations, random number generation and statistical simulation;
- Construction of user-defined functions to apply repeatedly a designated set of operators and functions while preserving readability of programming code;
- Illustrate the effectiveness and practicality of collaborative programming to upper-division and first-year graduate students.

Among the topics included in this paper are:

- Large sample theory for maximum likelihood estimation,
- Bayesian inference,
- Relationship between Poisson and χ^2 distributions.
- Samples of different sizes were examined.

1.1 Poisson Distribution

The Poisson frequency function with parameter λ ($\lambda > 0$) is $P(X = x) = e^{-\lambda} \lambda^x / x!$, where $x = 0, 1, 2, ...$ The Poisson distribution is often used to model the occurrence of rare events in a given period of time or distance. The events may include: calls coming to a telephone system; radioactive particle emitted from radioactive source; cases of a rare non-infectious disease, used to estimate its prevalence; birth defects and genetic mutations; car accidents and claims made to an insurance company; vehicles passing through a street, used for analysing traffic and setting switching time of the traffic lights [Rice, p.45].

1.2 Coverage of Interval Estimates

After an appropriate distribution is chosen, its parameter needs to be estimated to make the model practicable; in the case of the Poisson distribution the parameter is its mean. Therefore, a sample from the population is necessary to suggest a range of values for the estimated parameter. Multiple methods were devised for deriving interval estimates with declared coverage for the parameter λ of Poisson distribution. With repeated execution of a specific procedure to obtain a confidence interval, the researcher's expectation is to find the proportion of intervals that include the true value of λ, to be equal to the declared coverage. In Figure 1, the proportion of simulated 95% confidence intervals that contain true value of λ happens to be exactly 95/100.

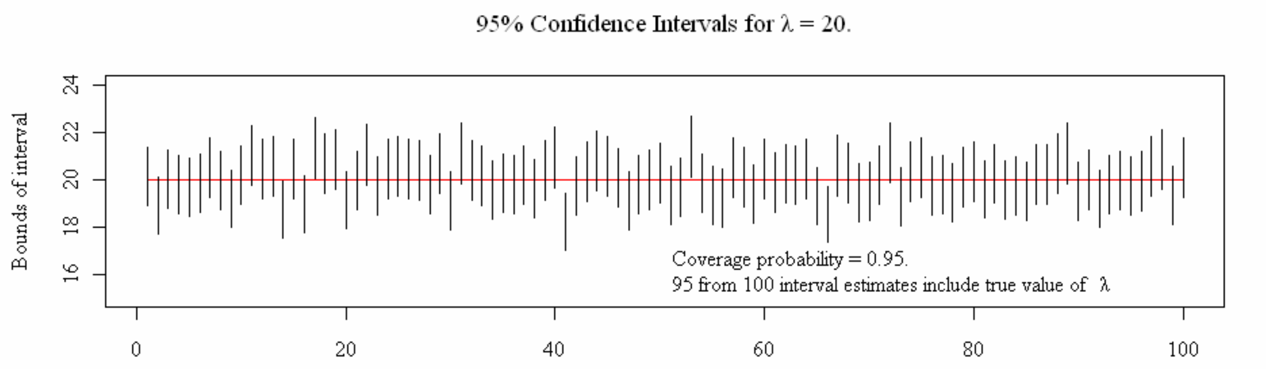

Figure 1. One hundred 95% confidence interval estimates of the parameter; 95 of which include true value.

In an ideal world, the proportion would be always equal the nominal coverage, but due to approximation, assumptions made, and variations in the conditions in which the experiment or sampling was done, the nominal coverage may not always be reached. In an attempt to explore the reliability of a method in terms of coverage, we performed repeated random sampling, and used R to construct confidence intervals for various values of λ.

2. True Coverage of Confidence Intervals Constructed by Different Methods

2.1 Large Sample Theory for Maximum Likelihood Estimation

A confidence interval constructed using large sample theory has a distinct structure: the center of the interval is a point estimator, specifically maximum likelihood estimator (MLE), and limits of the interval are obtained using the asymptotic variance (AV) of the MLE. Therefore, the interval is presented as (MLE-AV; MLE+AV).

2.1.1 Maximum Likelihood Estimator and its Asymptotic Variance

When sampling from a Poisson distributed population, the number that makes the observed sample most likely to appear is the maximum likelihood estimator (MLE) of λ. The theoretical derivation of maximum likelihood estimator is described in [Rice, p.267], where the MLE of the parameter for Poisson distribution is simply the sample mean [p.268]: $\hat{\lambda} = \overline{X} = (1/n)\sum_{i=1}^{n} X_i$, where observations X_i are iid Poisson(λ). Since $\hat{\lambda}$ is calculated from a random sample, it is also randomly distributed; its probability distribution is called the sampling distribution. Large sample theory asserts that, for large sample sizes, the sampling distribution of the maximum likelihood estimator is normal with the MLE as the mean, and AV as the variance. Accordingly, another name for large sample theory is *normal approximation*.

After standardizing the normal sampling distribution, the interval appears as $\hat{\lambda} \pm z_{\alpha/2} AV$, where $z_{\alpha/2}$ denotes the $\alpha/2$ quantile of the standard normal distribution, and AV is the asymptotic variance of the MLE of the Poisson parameter found as $1/nI(\hat{\lambda}) = \hat{\lambda}/n = \overline{X}/n$ [Rice, p.282]. Hence, an approximate 100(1 – α)% confidence interval for λ is $\overline{X} \pm z_{\alpha/2}\sqrt{\overline{X}/n}$. A similar derivation was presented by [Brown] for the parameter of a Binomial distribution.

2.1.2 Realization in R-environment

To see how much coverage the interval estimate can attain, the calculation of its bounds was programmed and samples from Poisson population with specific parameter were simulated with an elementary program in R. A fragment of the programming code is displayed in Figure 2. In the program, repeated sampling for a specific value of λ is presented in the matrix **X**. Each row of this matrix contains one individual sample, and the number of columns is equal to the sample size. With the aid of matrix and vector operations available in R it is possible to avoid using loops, which often run very slowly in large-scale simulations.

```
X = matrix(rpois(m*n, lambda), nrow = m, ncol = n, byrow = TRUE)        # m samples of size n from Poisson(lambda)
x.bar = apply(X, 1, mean)                                  # m sample means
m.err = qnorm(1-a1/2)*sqrt(x.bar/n)                        # m margins of error
lcl = x.bar - m.err; lcl[lcl<0]=0                          # m Lower Confidence Limits
ucl = x.bar + m.err                                        # m Upper Confidence Limits
cover = (lambda >= lcl) & (lambda <= ucl)                  # Vector of 0s and 1s

p.cov = mean(cover)                                # Simulated coverage for specific lambda
width = mean(ucl - lcl)                            # Simulated interval width for specific lambda
```

Figure 2. R code for calculating true coverage of interval estimate based on large sample theory.

Large sample theory is expected to give reliable estimates on samples equal to or larger than 30. In Figure 3, simulated coverage of 95% confidence intervals for samples of 10, 30 and 50 observations are compared. As the plot illustrates, even a sample size larger than 30 does not guarantee the nominal coverage. Moreover, there were not any obvious distinctions in coverage found for different sample sizes. However, in general, larger samples produce shorter intervals for the normal approximation method. Another advantage of a large sample is that the true coverage of confidence intervals substantially becomes closer to the nominal coverage for small values of λ. Specifically, for values of λ in (0, 5] interval estimates based on the normal approximation performed noticeably better for sample size 50. Similar performance was noted for 99% intervals, for which it is worth mentioning that all provided at least 95% coverage.

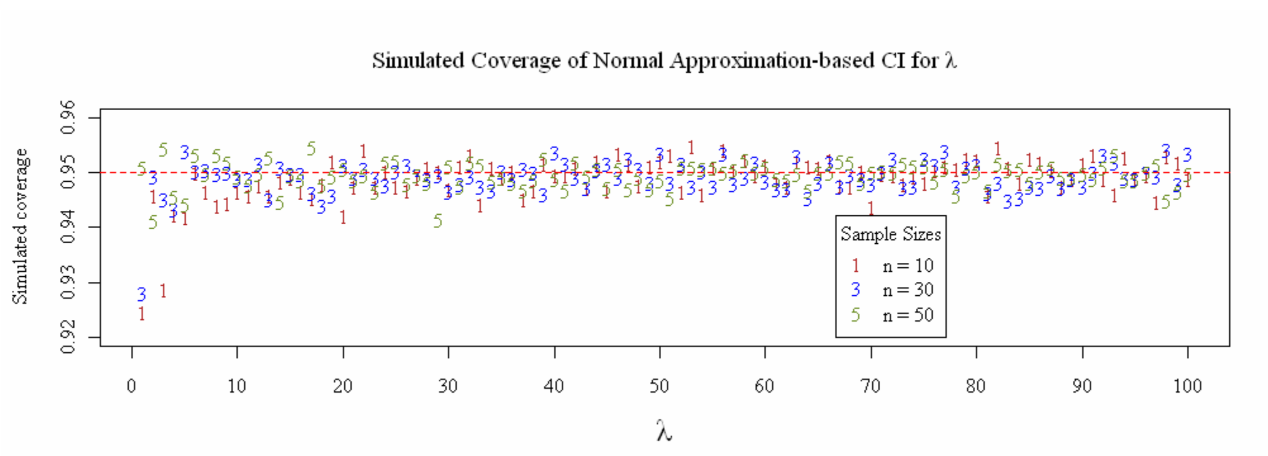

Figure 3. Simulated coverage of 95% confidence intervals based on large sample theory for different sample sizes.

2.2 Bayesian Estimation

In Bayesian Estimation the researcher attempts to incorporate previous knowledge or expert opinion about a situation into the investigation. Technically, the unknown parameter, λ in the case, considered to be as a random variable having a distribution, which usually is called prior distribution. To decide on a distribution family for the prior, it is recommended to consider properties of the parameter of interest. For example, as a parameter of the Poisson distribution, λ can be any number on the positive axis. Hence, the prior distribution should be continuous and defined for positive values. In order to adequately incorporate expert knowledge and to achieve elegant mathematical manipulations, the prior distribution for λ is taken as gamma with parameters (α, β). Then, according to [Lee], the posterior distribution for λ is again gamma, except its parameters are $a^* = \Sigma_i X_i + \alpha$, and $\beta^* = n + \beta$, where the sum is taken from 1 to n. After the posterior distribution is found, bounds of 95% the interval estimate are defined as .025- and

.975-quantiles of gamma with parameters (α^*, β^*). More detailed information on Bayesian inference generally, and on choosing prior distribution for Poisson population in particular can be found in [Lee, p.87].

Regarding the prior distribution: by experimenting with various values for α and β, it is possible to put the desired probability on a specific range. Assume, that the expert considers that the most plausible values of λ are in the range (0, 20], then, our task is to fit a gamma distribution so that most of the probability is in interval (0, 20]. To investigate the effects of using various priors, several different ones were used in this project:

- Almost flat prior Gamma($\alpha = 1$, $\beta = .01$) that puts about 60% of probability into the interval (0, 100),
- Informative prior Gamma ($\alpha = 6.77$, $\beta = .58$) for small values of λ with about 95% of probability in (0,20),
- Informative prior Gamma($\alpha = 84.21$, $\beta = 1.66$) for larger values of λ with about 95% of probability in (40, 60).

In Figure 4, all of these priors are plotted to illustrate the various probability allocations. All of them were found using the R code.

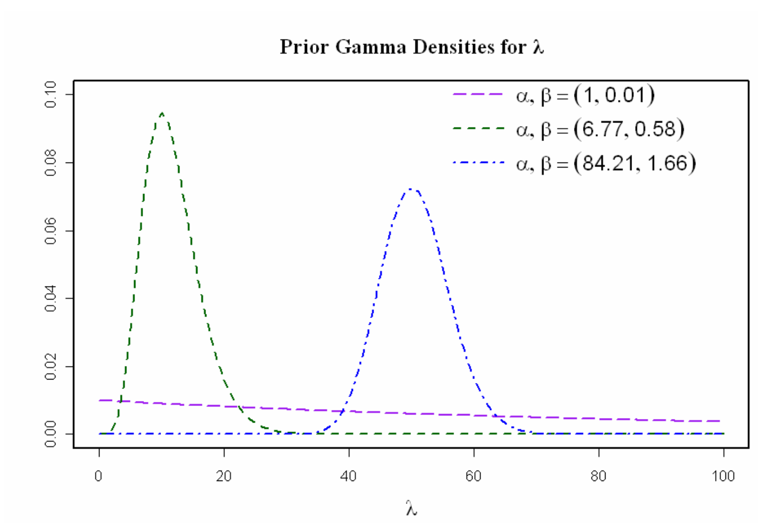

Figure 4. Prior distributions used in Bayesian estimation of Poisson parameter.

```
alpha.prior = 1; beta.prior = .01                          # suggested parameters for prior Gamma distribution
XMAT = matrix(rpois(m*n, lambda), nrow = m, ncol = n, byrow = TRUE)
x.sum = apply(X, 1, sum)
alpha.post = alpha.prior + x.sum; beta.post = beta.prior + n         # calculation of the parameters of posterior Gamma
lambda.lo = qgamma(a1/2, shape = alpha.post, rate = beta.post)   # lower bound of the confidence interval
lambda.up = qgamma(1-a1/2, shape = alpha.post, rate = beta.post)       # upper bound of the confidence interval
p.cov.bay = mean((lambda >= lambda.lo) & (lambda <= lambda.up))       # check if lambda is in the interval
```

Figure 5. R-code for computing Bayesian credible intervals for the Poisson parameter.

During the iteration for each value of λ, repeated sampling of n observations from Poisson(λ) is simulated, parameters of the posterior gamma distribution are calculated, and the bounds of the interval estimate are evaluated. In Figure 5, R code for obtaining Bayesian interval estimates is displayed.

Coverage of a Bayesian credible interval for each prior distribution was simulated as well as coverage of Normal approximation-based intervals. Surprisingly, true coverage that was the closest to nominal coverage was obtained with the almost non-informative prior Gamma($\alpha = 1$, $\beta = .01$). Moreover, it is worth mentioning that the informative priors made it possible to reach the declared coverage and even surpass it, but only for the middle part of the targeted range and for small sample sizes as shown in Figure 6. Perhaps the performance of the informative prior for small values of λ can be improved by changing to a non-modal Gamma distribution with shape parameter 1. For the current analysis, a Gamma distribution with shape parameter 6.77 (and thus mode 10) was used as a prior distribution.

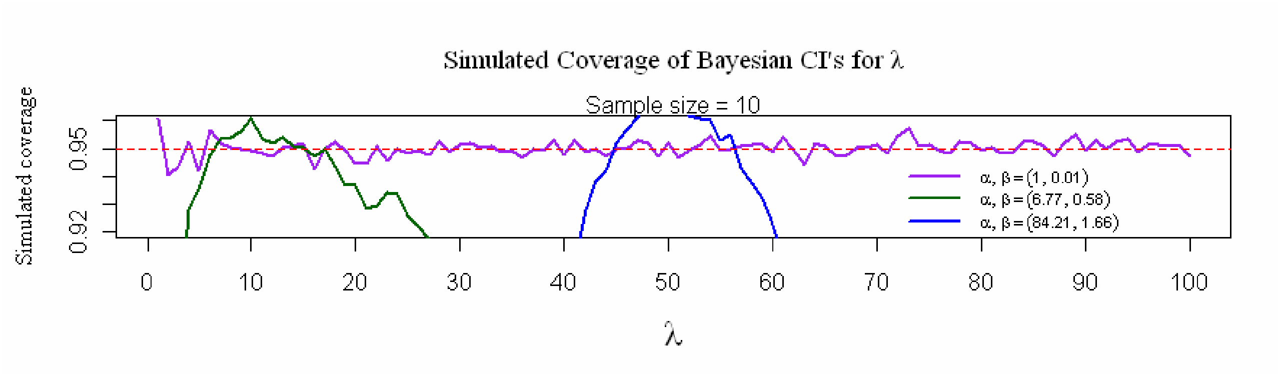

Figure 6. Simulated coverage of Bayesian credible intervals for different prior distributions. The colors of the three curves correspond to colors of the density curves for the three prior distributions in Figure 5.

Because the width of the interval is also an important attribute, expected widths of the intervals were also calculated. It is clear that the wider the interval the more chances for lambda to be inside the interval, but at the same time a wider interval provides less precise inference. Unlike the *simulated* expected width of normal confidence intervals, expected widths of Bayesian credible intervals was directly *calculated*. It is sufficient to know parameters of prior distribution, sample size, and sum of the observations to define the posterior distribution of λ. Then for each specific λ, specific prior and specific sample size, the problem of computing the expected width reduces to finding what would be the width of the interval for each possible sum of the observations, and probability to observe each sum given that population has Poisson distribution with specific λ. Luckily, sum of identical independently distributed Poisson random variables still has a Poisson distribution, except the parameter now equals λ times sample size. In addition, it is acceptable to set an upper bound for the *sum* variable since at certain values of *sum*, the probability of its appearing is so small that it does not affect the expected width by much. For example, the probability that a *sum* equalling 150 occurs with a distribution having parameter $\lambda = 5$ is 8.262944e–161, and thus, the range for calculating the expected width can be truncated to extend from 0 to 150 in this particular instance. R code for computing the expected width of a Bayesian credible interval for a Poisson parameter is shown in Figure 7.

```
h = lambda*n*3+1                    # h diff values of possible observations; "+1" stands for all x's are zero
sumx = 0:(h-1)                      # sumx is sum of n observations from Poisson
alpha = alpha.prior + sumx          # shape parameter for posterior Gamma
lo = qgamma(a1/2, shape = alpha, rate = beta.post)
up = qgamma(1-a1/2, shape = alpha, rate = beta.post)
width.x = up - lo                   # width.x = width of the interval for each possible sum of observations
width.bay = sum(width.x*dpois(sumx[1:length(sumx)], lambda*n))          # width = expected width of the interval
```

Figure 7. Fragment of the R-code for calculating expected width of Bayesian credible interval estimate.

The expected widths of the intervals were obtained for estimates based on different priors, and widths are compared in Figure 8. Here we can see that, for small values of λ, an almost non-informative prior allowed us to get intervals that not only have better coverage, but also have shorter intervals. However, with a larger sample size the difference in width became negligible. For values of λ in a range from 40 to 60, an informative prior showed consistently narrower intervals than did our non-informative priors.

2.3 Using the χ^2 Distribution

This section describes the calculation of coverage probabilities for interval estimates of the parameter of a Poisson distribution that are based on the χ^2-distribution, as was first suggested by [Garwood]. Using the derivation in his paper and the fact that sum of independent identically distributed Poisson variables is still Poisson-distributed, but with parameter equal to the original parameter times sample size [Rice, p.159], bounds of a $(1 - \alpha)100\%$ confidence interval can be calculated as $(\chi^2_{2T,\alpha/2}/2n,\ \chi^2_{2T+2,1-\alpha/2}/2n)$, where $T = \sum_{i=1}^{n} x_i$ is the sum of observations in a sample of size n from a Poisson distribution with unknown mean λ, $\chi^2_{df,q}$ is the q-quantile of the χ^2-distribution with df degrees of freedom, and α is chosen to designate a $(1-\alpha)100\%$ confidence interval.

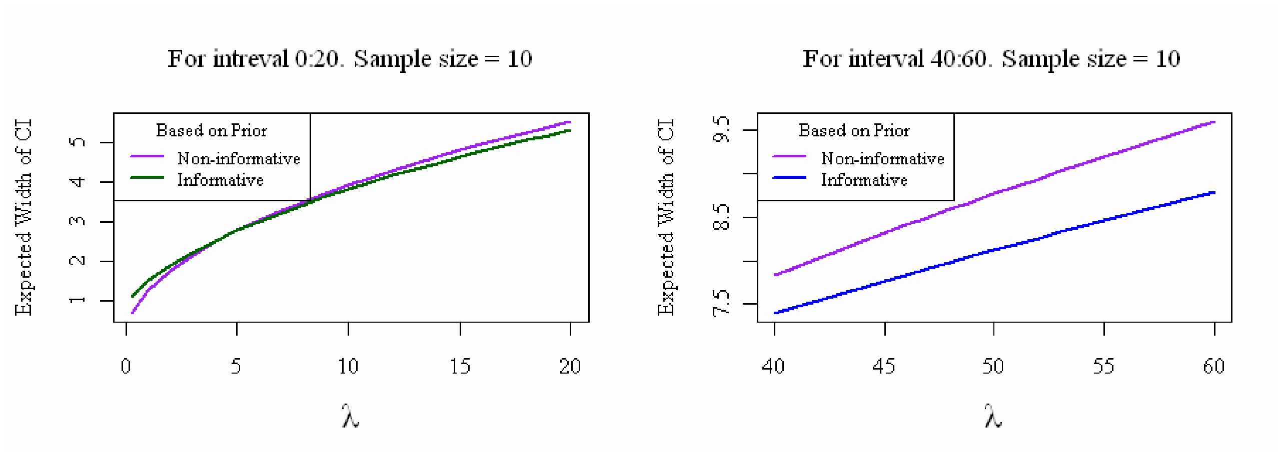

Figure 8. Plots demonstrating widths of the interval estimates based on informative and non-informative priors.

Therefore, once the sample from the population is given, the bounds can be computed. Table 1 presents calculated confidence interval bounds for sample of just 1 observation and sample of 10 observations. Interestingly, width of the interval based on a sample of size 10 is 10 times shorter than based on sample of size 1, or in general, width of the confidence interval is inversely proportional to sample size, *n*, as well as for large sample-based intervals.

Table 1: Calculated 95% confidence interval bounds and width for single observation and for sample of size 10

Sum of observations	*n=1*			*n=10*		
	Lower bound	*Upper bound*	*Width*	*Lower bound*	*Upper bound*	*Width*
0	0.0000	3.6889	3.6889	0.0000	0.3689	0.3689
1	0.0253	5.5716	5.5463	0.0025	0.5572	0.5546
2	0.2422	7.2247	6.9825	0.0242	0.7225	0.6982
3	0.6187	8.7673	8.1486	0.0619	0.8767	0.8149
4	1.0899	10.2416	9.1517	0.1090	1.0242	0.9152
5	1.6235	11.6683	10.0448	0.1623	1.1668	1.0048
6	2.2019	13.0595	10.8576	0.2202	1.3059	1.0858
7	2.8144	14.4227	11.6083	0.2814	1.4423	1.1608
8	3.4538	15.7632	12.3094	0.3454	1.5763	1.2309
9	4.1154	17.0848	12.9694	0.4115	1.7085	1.2969
10	4.7954	18.3904	13.5950	0.4795	1.8390	1.3595

2.3.1 Probability that λ lies in the confidence interval

Let us examine the 95% CIs for sample of size 10 in Table 1. As we see, values of both bounds are increasing as the observed sum increases. Consider a range (0, 0.3689] (where 0.3689 is the 1st upper bound). Then for any λ from this range there is no chance that the upper limit of a confidence interval will miss λ, because the lowest possible upper bound (0.3689) is still greater than (or equal to) λ. Now assume that $0.3689 < \lambda \le 0.5572$; that is, the true value of λ is located between the 1st and 2nd upper bounds. Then the upper limit of an interval will miss λ only if all 10 observations in the obtained sample equal 0. Hence, the probability that λ is larger than upper bound of the interval equals probability of obtaining 10 zeroes from Poisson distribution where λ has a particular value from the interval (0.3689, 0.5572]. With analogous reasoning it is possible to get the probability of missing λ by upper bound for arbitrary range of λ.

Now, let us consider lower limits for the same purpose. Assume that $0 \le \lambda < 0.0025$ (between the 1st and 2nd lower bounds), then the only way to miss λ with the lower bound is to obtain at least one observation not equal to 0. That is, the probability of missing λ with the lower bound equals the probability of getting anything greater than 0 from Poisson distribution with parameter λn from interval [0, 0.0025n). With a similar argument we can obtain the probabilities of missing λ by lower bound for any range of λ.

Finally, if λ is not in the interval, then either bound has missed it. Therefore, the probability of this event is the sum of probabilities of missing λ with the lower bound and missing λ with the upper bound for a particular value of λ; and the coverage probability sought is the compliment of the event of missing λ. Coverage probabilities of CIs constructed with the χ^2-distribution for different sample sizes are demonstrated in Figure 9, where we also see that their coverage always is at least as good as the stated confidence level.

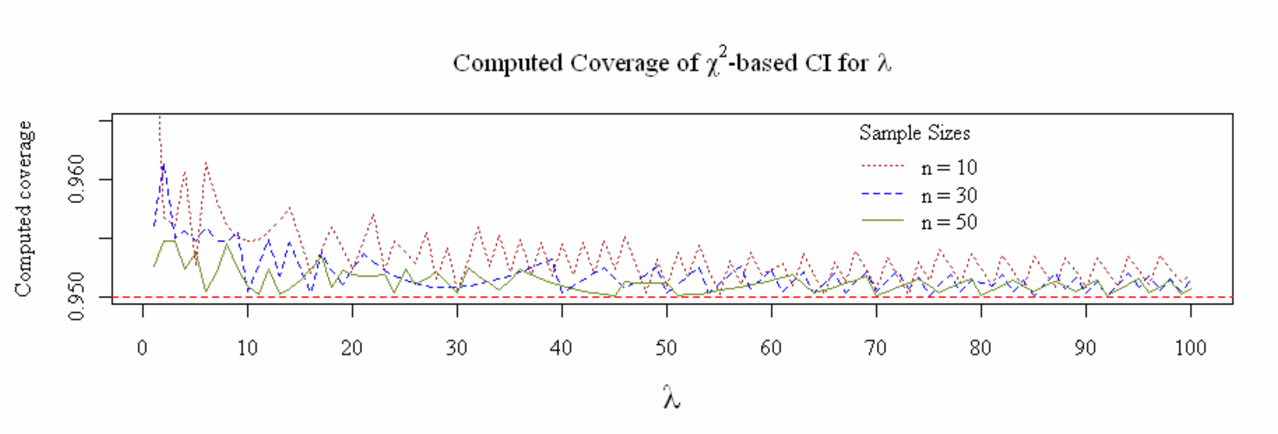

Figure 9. Simulated coverage for confidence intervals based on χ^2-distribution for samples of different sizes.

3. Comparing Interval Estimates Constructed by Different Methods

3.1 Comparing Expected Widths

Now that we are aware of the true coverage of various intervals, it is time to take a look at their expected width. In Figure 10, variability of the widths of the intervals is readily seen to be a deviation from the width of the χ^2-based interval, taken as a reference point. Therefore, if a curve on the plot goes higher then line $y = 0$, then the interval is shorter than the χ^2-distribution-based interval. Moreover, the higher the curve on the plot, the shorter the interval represented by the curve.

Based on the limited scope of our investigations of true coverage probabilities and expected widths of the intervals, the following recommendations seem reasonable:

- For a small sample and small assumed values of λ it seems best to use the method based on the χ^2-distribution, because of guaranteed coverage; fairly good coverage can be attained also by employing the Bayesian method, which gives shorter intervals than χ^2-method;
- For a small sample and large values of λ, the Bayesian method with an appropriate prior gives the narrowest intervals maintaining required coverage;
- For a large sample and small values of λ, use the χ^2-distribution to obtain at least the required coverage or a Bayesian interval estimate with reference to the appropriate informative prior for the narrowest intervals;
- For a large sample and large values of λ, large-sample and Bayesian methods give close to the required coverage and short intervals.

4. Conclusion

The subject of this discussion originated from a class project in a graduate seminar on simulation given during summer 2007, where an R program to compute coverage probabilities of confidence intervals for a parameter of a binomial distribution was demonstrated, based on [Brown, et al.]. Hopefully, the project will be helpful to future students for illustrating the concepts of interval estimation and of simulation with R.

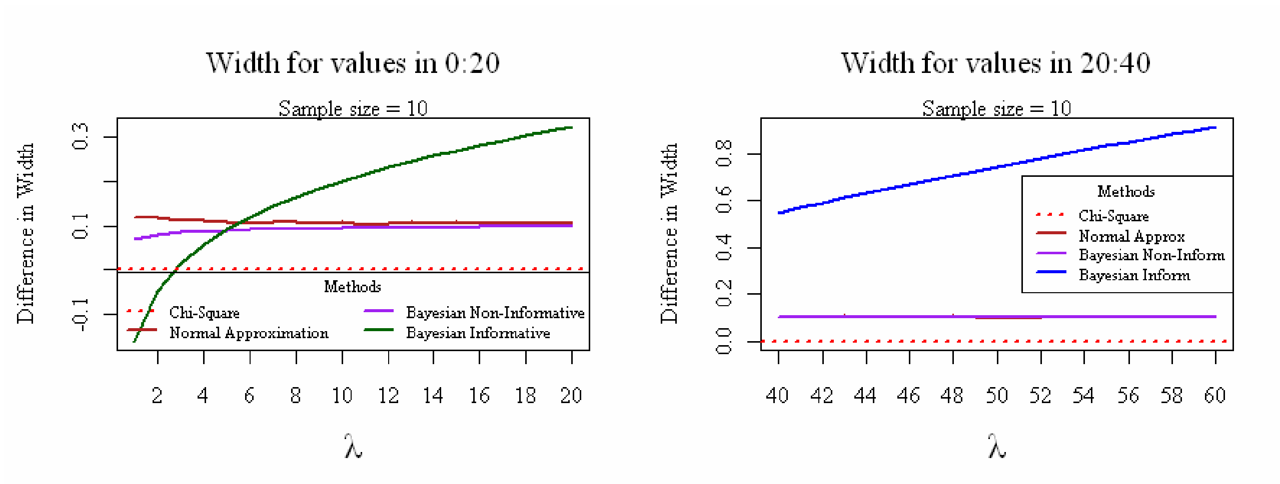

Figure 10. Difference in widths of CI's constructed by different methods in comparison to width of an interval estimate based on the χ^2-distribution.

References

Barker, Lawrence (2002), "A Comparison of Nine Confidence Intervals for a Poisson Parameter When the Expected Number of Events Is ≤ 5," *The American Statistician*, Vol. 56, No. 2, May 2002, pp. 85-89.

Brown, Lawrence D., Cai, T. Tony, and Das Gupta, Anirban (2001), "Interval estimation for a binomial proportion, " *Statistical Science,* 16:2, 101–133.

Casella, George, and Robert, Christian (1989), "Refining Poisson Confidence Intervals," *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, Vol. 17, No.1, March 1989, pp. 45-57.

Crow, Edwin L., and Gardner, Robert S. (1959), "Confidence Intervals for the Expectation of a Poisson Variable," *Biometrika*, Vol. 46, No. 3/4, December 1959, pp. 441-453.

Garwood, F. (1936), "Fiducial Limits for the Poisson Distribution," *Biometrika*, Vol. 28, No. 3/4, December 1936, pp. 437-442.

Kabaila, Paul, and Byrne, John (2001), "Exact Short Poisson Confidence Intervals," *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, Vol. 29, No.1, March 2001, pp. 99-106.

Lee, Peter M. (2004), *Bayesian Statistics: an Introduction*, London: Hodder Arnold

Rice, John A. (2007), *Mathematical Statistics and Data Analysis*, Brooks/Cole: Thomson, pp. 159, 183, 268-269, 282-283.

Sahai, Hardeo, and Khurshid, Anwer (1993), "Confidence Intervals for the Mean of a Poisson Distribution: A Review," *Biometrical journal*, Vol. 35, No. 7, pp. 771-896.

Stapleton, James H. (2008), Models for Probability and Statistical Inference: Theory and Applications, John Wiley & Sons, Inc., pp.319-322.