A Classroom Approach to Illustrate Transformation and Bootstrap Confidence Interval Techniques Using the Poisson Distribution

Per Gösta Andersson[1]

[1] Department of Statistics, Stockholm University, Stockholm, Sweden

Correspondence: Per Gösta Andersson, Department of Statistics, Stockholm University, S-106 91, Stockholm. E-mail: per.gosta.andersson@stat.su.se

Abstract

The Poisson distribution is here used to illustrate transformation and bootstrap techniques in order to construct a confidence interval for a mean. A comparison is made between the derived intervals and the Wald and score confidence intervals. The discussion takes place in a classroom, where the teacher and the students have previously discussed and evaluated the Wald and score confidence intervals. While step by step interactively getting acquainted with new techniques, the students will learn about the effects of e.g. bias and asymmetry and ways of dealing with such phenomena. The primary purpose of this teacher-student communication is therefore not to find the best possible interval estimator for this particular case, but rather to provide a study displaying a teacher and her/his students interacting with each other in an efficient and rewarding way. The teacher has a strategy of encouraging the students to take initiatives. This is accomplished by providing the necessary background of the problem and some underlying theory after which the students are confronted with questions and problem solving. From this the learning process starts. The teacher has to be flexible according to how the students react. The students are supposed to have studied mathematical statistics for at least two semesters.

Keywords: bootstrap/resampling, delta method, score statistic, symmetrizing transformation, variance-stabilizing transformation, Wald statistic

1. Introduction

In this paper we will illustrate two different techniques for confidence interval construction from a teaching perspective, using the Poisson mean as the target parameter. We start with variance-stabilizing and symmetrizing transformations, where the Delta method is the main tool. Nonparametric bootstrap will be the next topic, where we concentrate on the basic and percentile intervals and some modifications thereof.

Papers involving transformation and bootstrap intervals for a Poisson mean include (Barker, 2002, Swift, 2009, Patil & Rajarshi, 2010). Chapter 4 in (DasGupta, 2008) is recommended, especially on the topic of transformation methods in general.

The primary aim is to show how you can teach these two types of techniques using a very simple setup. The students learn about quite complex reasoning in a straightforward way, due to the simplicity of the Poisson distribution. Furthermore, we will be able to compare performances of various interval estimators. From his own experience the author believes that using two seemingly disparate methods such as transformations and the bootstrap on the same problem have great advantages. Usually theory and methods are first presented and the results are applied on data (real or simulated). Then one moves on to additional methods which are applied to other types of data and so on. Students are often having trouble choosing which method to apply in a given situation, since they have been presented to many combinations of methods and structures of data. Keeping the problem fixed, as in this presentation, encourages students to concentrate on the methods at hand, including comparisons of their performances.

While introducing the theory step by step, students will be given tasks in groups to work on and the results are discussed in the classroom with immediate feedback from the teacher. On the one hand the teacher must at all times be able to slow down or even go back if the students seem to get into trouble, while on the other hand he/she should allow for and encourage unexpected initiatives and thoughts, which may lead to activities not planned beforehand. The latter situation symbolizes the strong connection between teaching and research: from what you have learned and are learning thoughts can evolve, leading to new ideas.

We will follow the teacher and the students during five days of lectures in a Master's level statistical inference course with emphasis on asymptotically motivated methods. The students are supposed to be familiar with statistical inference at the level of e.g. (Casella & Berger 2002, Hogg, McKean & Craig, 2005). It is further assumed that previously the

students have investigated properties of confidence intervals for a Poisson mean based on the Wald and score statistics, see (Andersson, 2015). They found that for those values of the mean that we studied, the score interval outperformed the Wald interval. Since then the students have spent yet another semester on studies in probability and statistical inference, so it seems a good idea to explore some other methods in order to construct confidence intervals.

Remark: Whenever it says *Student*, it means a student representing one of the groups the students are divided into.

2. The First Day

Teacher addressing the students: We will start by considering the first three moments of the Wald and score statistics. Ideally, for a pivotal statistic Z in this situation: $E(Z) = 0$, $V(Z) = E(Z^2) = 1$ and $E(Z^3) = 0$. In our previous investigation we saw that for X Poisson distributed with the mean μ, the Wald

$$\frac{X - \mu}{\sqrt{X}} \tag{1}$$

statistic had a negative bias and a variance larger than one and furthermore possessed a substantial negative skewness. The score statistic

$$\frac{X - \mu}{\sqrt{\mu}}, \tag{2}$$

on the other hand, is unbiased with a unit variance and a positive skewness, which in absolute value turned out to be smaller than the skewness for the Wald statistic.

Generally, for a statistic to be really useful for inferential purposes using e.g. the normal distribution, we have at our disposal methods which deal with the following three problems: bias, non-constant variance and skewness. The second problem can be handled by variance-stabilizing transformations and the third by symmetrizing transformations and on top of this we can also adjust for bias, as it turns out.

Now, before we look into these particular methods, and others, we will increase the level of information in the previous hypothetical example. Before we had 28 observed impulses (passing vehicles on a road) during one hour. Suppose that we partition the one-hour-period into ten six-minute intervals and further suppose that in each of these intervals we will get to know the number of impulses. (This will actually not be necessary for the subsequent transformation based intervals.) So, again we may assume, at least as an approximation, that we have a homogenous Poisson process in each interval with unknown intensity.

The additional information is the following sequence of observed impulses for the six-minute intervals: 2, 4, 0, 3, 5, 3, 1, 3, 4, 3. Our primary target is the expected number of impulses in one such interval, which we denote by θ. If we further let $X_1, \ldots, X_{10}$ denote the number of impulses in the intervals, then, of course, $\sum_{i=1}^{10} X_i$ has the expected value $\mu = 10\theta$. Thus a natural estimator of θ is $\hat{\theta} = \bar{X}$ and $\hat{\mu} = 10\hat{\theta}$.

In general, for an observed sequence $x_1, \ldots, x_n$ the $100(1 - \alpha)\%$ Wald interval for θ with observed $\bar{x}$ is

$$\bar{x} \pm \frac{z_{\alpha/2}}{\sqrt{n}} \sqrt{\bar{x}} \tag{3}$$

and the corresponding score interval

$$\bar{x} + \frac{z_{\alpha/2}^2}{2n} \pm \frac{z_{\alpha/2}}{\sqrt{n}} \sqrt{\bar{x} + \frac{z_{\alpha/2}^2}{4n}} \tag{4}$$

We derived these intervals by inverting the statistics (1) and (2), respectively.

In the search for alternative intervals, let us first look at a variance-stabilizing transformation (VST). As it turns out, we will have good use for the Delta Method, which you encountered in a previous probability course. We start assuming that for an arbitrary sequence, say, T_n, we have the following convergence in distribution:

$$\sqrt{n}(T_n - \theta) \overset{D}{\Rightarrow} N(0, \sigma^2(\theta)), \quad n \to \infty$$

and if we have a function g, where the derivative $g'(\theta) \neq 0$, then

$$\sqrt{n}(g(T_n) - g(\theta)) \overset{D}{\Rightarrow} N(0, g'(\theta)^2 \sigma^2(\theta)), \quad n \to \infty$$

We are now going to use this result in a kind of "backward" way. As you can see, the asymptotic variance of T_n is here allowed to depend on θ and if T_n is to be used for the inference of θ, we would like to find a function of T_n which instead has a (known) constant asymptotic variance.

Until the break you will work in groups trying to find a suitable function of T_n.

2.1 After the Break

Student: We think you mean that $g'(\theta)^2 \sigma^2(\theta) = c^2$ (a constant) and a primitive function solving this differential equation is $c \int \frac{1}{\sigma(\theta)} d\theta$. $T_n = \bar{X}$ here?

Teacher: Yes, and formally we consider the iid sequence $X_1, X_2, \ldots X_n$, where $X_i \sim Poi(\theta)$, $i = 1, 2, \ldots n$ and from the Central Limit Theorem:

$$\sqrt{n}(\bar{X} - \theta) \overset{D}{\Rightarrow} N(0, \theta), \quad n \to \infty$$

Student: So here the standard deviation $\sigma(\theta) = \sqrt{\theta}$, which means that $g(\theta)$ should be $c \int \frac{1}{\sqrt{\theta}} d\theta = c\, 2 \sqrt{\theta}$. We are not sure though what is to be done with c.

Teacher: c is arbitrary, but we can put $c = 1/2$ to get the "clean" expression $g(\theta) = \sqrt{\theta}$.

Student: The asymptotic variance of $\sqrt{\bar{X}}$ is then $g'(\theta)^2 \sigma^2(\theta) = (\frac{1}{2} \frac{1}{\sqrt{\theta}})^2 \theta = 1/4$ and we have:

$$\sqrt{n}(\sqrt{\bar{X}} - \sqrt{\theta}) \overset{D}{\Rightarrow} N(0, 1/4), \quad n \to \infty$$

Teacher: Correct and this particular result can be tracked back to (Bartlett, 1936).

Actually, now you have done the hard part of the job and from here we can construct an approximate confidence interval for $\sqrt{\theta}$ and then it is quite straightforward to derive an interval for θ.

First we observe that

$$2 \sqrt{n}(\sqrt{\bar{X}} - \sqrt{\theta}) \overset{appr}{\approx} N(0, 1),$$

so a $100(1 - \alpha)\%$ two-sided symmetric confidence interval for $\sqrt{\theta}$ using the observed $\bar{x}$ is

$$\sqrt{\bar{x}} \pm \frac{z_{\alpha/2}}{2 \sqrt{n}}$$

and the corresponding VST interval for θ is

$$\left(\sqrt{\bar{x}} \pm \frac{z_{\alpha/2}}{2 \sqrt{n}} \right)^2$$

Comment from a student: This looks quite different from the Wald and score intervals!

Teacher: Well, not if we expand the squares! Then we arrive at

$$\bar{x} + \frac{z_{\alpha/2}^2}{4n} \pm \frac{z_{\alpha/2}}{\sqrt{n}} \sqrt{\bar{x}}, \tag{5}$$

a kind of mixture between the Wald and score intervals. Anyway, with $\alpha = 0.05$, $n = 10$ and $\bar{x} = 2.8$, the interval (5) is (1.86, 3.93), so here we get the interval (18.6, 39.3) for μ. This can be compared with the Wald (17.6, 38.4) and score (19.4, 40.5).

Question from a student: When you earlier mentioned that we can correct for bias, I was not sure of what you meant. Bias of what?

Teacher: I was alluding to $\sqrt{\bar{X}}$ and the fact that $E(\sqrt{\bar{X}}) \approx \sqrt{\theta}$. But before we start looking for an estimator of θ, which is, at least, closer to being unbiased, we should also pay attention to the asymptotic variance 1/4. (Anscombe, 1948) found that for a Poisson distributed X the variance of $\sqrt{X + \frac{3}{8}}$ is closer to 1/4 than the variance of $\sqrt{X}$ is. We therefore say that $\sqrt{X + \frac{3}{8}}$ is a variance corrected transformation (an approximation of a higher order).

In our case $X = \sum_{i=1}^{n} X_i$ and $\mu = n\theta$ and from this we construct the pivotal

$$2\left(\sqrt{X + \frac{3}{8}} - \sqrt{n\theta + \frac{3}{8}}\right) = 2\sqrt{n}\left(\sqrt{\bar{X} + \frac{3}{8n}} - \sqrt{\theta + \frac{3}{8n}}\right)$$

The variance corrected VST confidence interval for θ is then

$$\left(\sqrt{\bar{x} + \frac{3}{8n}} \pm \frac{z_{\alpha/2}}{2\sqrt{n}}\right)^2 - \frac{3}{8n} = \bar{x} + \frac{z_{\alpha/2}^2}{4n} \pm \frac{z_{\alpha/2}}{\sqrt{n}}\sqrt{\bar{x} + \frac{3}{8n}} \tag{6}$$

It can further be shown that a better estimator of $\sqrt{\theta}$ with respect to bias is $\sqrt{\bar{X} + \frac{1}{4n}}$. Asymptotically both $\sqrt{n}\sqrt{\bar{X} + \frac{3}{8n}}$ and $\sqrt{n}\sqrt{\bar{X} + \frac{1}{4n}}$ have variance $1/4$, so the resulting bias corrected interval for θ using $\sqrt{\bar{X} + \frac{1}{4n}}$ is?

Until next break you will work in groups again to come up with a suggested interval.

2.2 After the Break

Student: We started with the pivotal

$$2\sqrt{n}\left(\sqrt{\bar{x} + \frac{1}{4n}} - \sqrt{\theta}\right)$$

From this we got the bias corrected VST interval

$$\left(\sqrt{\bar{x} + \frac{1}{4n}} \pm \frac{z_{\alpha/2}}{2\sqrt{n}}\right)^2 = \bar{x} + \frac{z_{\alpha/2}^2 + 1}{4n} \pm \frac{z_{\alpha/2}}{\sqrt{n}}\sqrt{\bar{x} + \frac{1}{4n}} \tag{7}$$

Teacher: Yes, I agree with you. Others actually advocate using the pivotal

$$2\sqrt{n}\left(\sqrt{\bar{x} + \frac{1}{4n}} - \sqrt{\theta + \frac{1}{4n}}\right)$$

Anyway, these last three intervals should numerically be close to each other and the 95% intervals for μ for our data using (5), (6) and (7) are (18.6, 39.3), (18.5, 39.4) and (18.8, 39.6), respectively.

Theoretically we arrive at an important conclusion here, namely that we cannot find an "optimal" choice of an estimator of $\sqrt{\theta}$, in terms of reducing bias and at the same time attaining an "accurate" variance.

Now let us look at the other transformation method, that of symmetrizing. Here we also have use for the Delta Method and the aim is to find a function $g(\bar{X})$ with a distribution which is as symmetric as possible. Without explanation of the technical details involved, the result is that we should use $g(\bar{X}) = \bar{X}^{2/3}$ and this is brought up in (DasGupta, 2008). What is then the resulting confidence interval of θ?

I suggest that you go back to your respective groups again and work on this for a few minutes.

2.3 After 10 Minutes

Student: Our group did not arrive at a resulting interval, but I will present what we have agreed upon.

$g(\theta) = \theta^{2/3}$ and $g'(\theta) = (2/3)\theta^{-1/3}$, so

$$\sqrt{n}(\bar{X}^{2/3} - \theta^{2/3}) \overset{D}{\Rightarrow} N(0, (4/9)\theta^{1/3}), \quad n \to \infty$$

which means that

$$P\left(\bar{X}^{2/3} - \frac{2z_{\alpha/2}}{3\sqrt{n}}\theta^{1/6} \le \theta^{2/3} \le \bar{X}^{2/3} + \frac{2z_{\alpha/2}}{3\sqrt{n}}\theta^{1/6}\right) \approx 1 - \alpha$$

$$\Leftrightarrow$$

$$P\left(\left(\bar{X}^{2/3} - \frac{z_{\alpha/2}}{\sqrt{n}}\frac{2}{3}\theta^{1/6}\right)^{3/2} \le \theta \le \left(\bar{X}^{2/3} + \frac{z_{\alpha/2}}{\sqrt{n}}\frac{2}{3}\theta^{1/6}\right)^{3/2}\right) \approx 1 - \alpha \tag{8}$$

But this is not a confidence interval!

Teacher: I think we will call it a day and meet again tomorrow, but I encourage you to continue to work on this problem until then.

3. The Second Day

Student: We have been working on the confidence interval based on $\bar{X}^{2/3}$ and we found that the values of $\beta = \theta^{1/3}$, for which the following function $f(\beta)$ is negative, constitute the interval for $\theta^{1/3}$ (and correspondingly, the values of θ for which the function is negative, yield the interval for θ):

$$f(\beta) = \beta^4 - 2\bar{X}^{2/3}\beta^2 - \frac{4z_{\alpha/2}}{9n}\beta + \bar{X}^{4/3}$$

Teacher: What has your group been doing? Let me guess, you have worked on (8) in the same way as when you derived the score interval? Actually this was not what I had in mind. But please go on.

Student: We started with the assumption that the pivotal

$$\sqrt{n}\left(\frac{\bar{X}^{2/3} - \theta^{2/3}}{\frac{2}{3}\theta^{1/6}}\right) \overset{\text{appr}}{\sim} N(0, 1),$$

so those values of θ for which

$$\left|\frac{\bar{X}^{2/3} - \theta^{2/3}}{\frac{2}{3}\theta^{1/6}}\right| \le \frac{z_{\alpha/2}}{\sqrt{n}}$$

constitute a $100(1 - \alpha)\%$ approximate confidence interval of θ, which we may call "score" symm T? Elaborating on this inequality, we found that it is equivalent to

$$\bar{X}^{4/3} - 2\bar{X}^{2/3}\theta^{2/3} + \theta^{4/3} - \frac{4z_{\alpha/2}^2}{9n}\theta^{1/3} \le 0 \tag{9}$$

and letting $\beta = \theta^{1/3}$, We got $f(\beta)$. We thought it was kind of cool, putting $f(\beta) = 0$, to end up with a situation where you have to solve a fourth-order equation: a quartic!

Teacher: I agree that it is quite stimulating. In your case, are you sure that the resulting confidence set for θ is really an interval?

Student: No, we have actually not checked if we generally get an interval for θ out of (9). We were not sure if it is worthwhile to derive the solutions explicitly of the fourth-order equation generally, but we thought we could look at $f(\beta)$ for $n = 10$, $\bar{X} = 2.8$ and $\alpha = 0.05$ and see how it behaves.

Teacher: Absolutely, and I suspect that you have already done this!

Student: But of course! Here is a plot (Figure 1) and we have done it for values of β ranging from 0 to 2.

Teacher: Apparently you get an interval here for $\beta = \theta^{1/3}$ and I guess that from the equation $f(\beta) = 0$, two of the roots will be complex.

Student: Yes, and the two real-valued roots are 1.2358 and 1.5832 (using MATLAB). So, in terms of θ we obtain the interval (1.89, 3.97) and for μ: (18.9, 39.7). This is somewhat closer to the score interval (19.4, 40.7) than the interval obtained from symmetrization.

*Teacher*: Quite right. Do we have an alternative solution from some other group?

Student: Maybe we have been taking an easier way out of this, arguing that $\bar{X}$ is a consistent estimator of θ, replacing $\theta^{1/6}$ by $\bar{X}^{1/6}$, finally arriving at the following Wald "symm T" interval for θ:

$$\left(\bar{X}^{2/3} \pm \frac{2z_{\alpha/2}}{3\sqrt{n}}\bar{X}^{1/6}\right)^{3/2} \tag{10}$$

This interval (10) yields that the corresponding 95% interval for μ is (18.3, 39.0).

Teacher: The result is somewhat closer to the obtained Wald interval (17.6, 38.4), than the intervals using the variance stabilizing transformation. The "Wald solution" as you call it, can be found in (Brown, Cai & DasGupta, 2014).

Good work from both groups and from what I have seen, the rest of the groups have arrived at (10), which was what I had expected from you. Though the "score solution" was a bonus! Now I think it is about time to sum up our findings and make a table of the computed intervals.
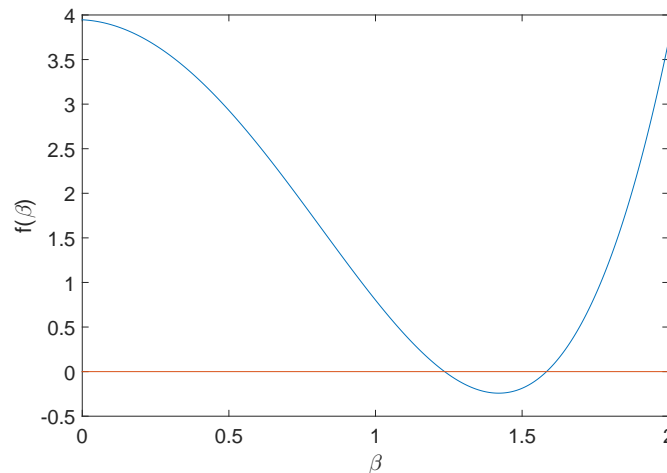
Figure 1. Plot of the function $f(\beta)$

Table 1. Wald, Score and Transformation Intervals for μ with $\bar{x} = 2.8$.

Wald	(17.6, 38.4)
score	(19.4, 40.5)
VST	(18.6, 39.3)
var corr VST	(18.5, 39.4)
bias corr VST	(18.8, 39.6)
"Wald" symm T	(18.3, 39.0)
"score" symm T	(18.9, 39.7)

The bias and variance corrections make the VST intervals come closer to the score interval and the same holds for the score-type symmetrized interval.

Now I would like to move on to another approach of estimation: the bootstrap.

We will consider both the parametric and non-parametric bootstrap. In this situation the former could be used as a check to whether the assumption of a Poisson process is valid or not.

Comment from a student: Everything we have done so far seems to depend on that assumption! From what I understand, non-parametric bootstrap does not rely on distributional assumptions, but we have not learnt how to construct confidence intervals using the bootstrap.

Teacher: We will get to that soon, but first we have the issue of checking the validity of the (homogenous) Poisson process assumption.

Let us look at an example from (Casella & Berger, 2002). We will compare $\bar{x}$ with s^2 and see if they differ much. They should, at least, be close to one another, if the Poisson assumption is true, since they are both estimators of μ. In our case $\bar{x} = 2.8$ and $s^2 \simeq 2.2$. We can then mimic what they have done in that example and check if $s^2 = 2.2$ is an "unlikely" value when $\theta = 2.8$. In other words, we find the probability distribution of S^2 when $\theta = 2.8$.

So, for tomorrow you will (group-wise) generate 1000 Poisson random samples of size 10, where each value is the outcome of a Poisson distributed random variable with $\theta = 2.8$. Then you will compute, for each sample, S^2 and draw a histogram to find the approximate distribution of S^2.

4. The Third Day

Student: This is our histogram of S^2.

The Poisson process assumption seems valid here since the value 2.2 is far from being an outlier.

Teacher: Yes, your observed sequence of impulses certainly does not contradict this assumption. As you already know, in
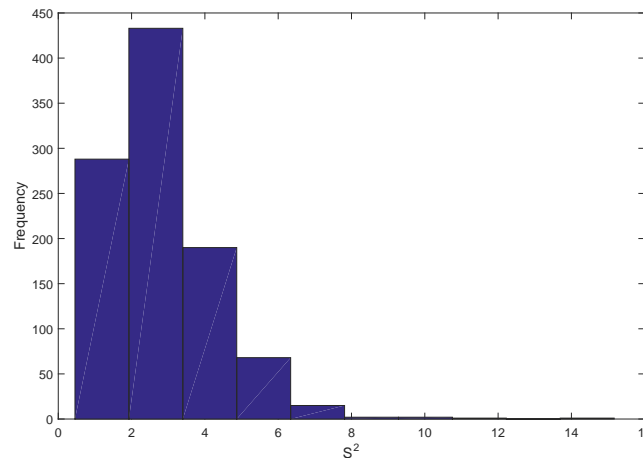
Figure 2. Histogram of S^2 for 1000 random samples from a Poisson distribution with $\theta = 2.8$.

the non-parametric bootstrap we take resamples from the original sample $X_1, \ldots, X_n$ under simple random sampling with replacement. For each such bootstrap sample $X_1^*, \ldots, X_n^*$, we compute the value of some statistic. Here $\theta^* = \bar{X}^*$ and we use this to "obtain" the distribution of $\hat{\theta} = \bar{X}$. One way of constructing a confidence interval for θ is simply by finding the corresponding percentiles of the empirical distribution of θ^*. This is described in e.g. (Hogg, McKean & Craig, 2005).

If we let the percentiles of this distribution be denoted by $a_{\alpha/2}^*$ and $a_{1-\alpha/2}^*$, the interval for θ is simply

$$(a_{\alpha/2}^*, a_{1-\alpha/2}^*) \tag{11}$$

This is the so called percentile bootstrap confidence interval. Actually, this could also be achieved based on your generated Poisson samples from a histogram of $\bar{X}$, as an example of the parametric bootstrap. The other approach, yielding the "basic" bootstrap confidence interval, is to start from the assumption that the distribution of $\hat{\theta} - \theta$ is the same as that of $\theta^* - \hat{\theta}$, i.e. θ^* relates to $\hat{\theta}$ in the same way as $\hat{\theta}$ relates to the true value θ. How should we proceed from this?

Work in groups until the break!

4.1 After the Break

Student: It seemed a bit tough to have three components to keep track of: θ, $\hat{\theta}$ and θ^*, but what you said seems to mean that

$$P(a \le \hat{\theta} - \theta \le b) = P(a \le \theta^* - \hat{\theta} \le b)? \tag{12}$$

Teacher: Yes!

The percentiles $a_{\alpha/2}^*$ and $a_{1-\alpha/2}^*$ from the histogram of θ^* are such that

$$P(a_{\alpha/2}^* \le \theta^* \le a_{1-\alpha/2}^*) = 1 - \alpha \,,$$

but that got us nowhere!

Teacher: If we combine this with what you have in (12) and also put those probabilities to $1 - \alpha$, we get

$$P(a + \hat{\theta} \le \theta^* \le b + \hat{\theta}) = P(a_{\alpha/2}^* \le \theta^* \le a_{1-\alpha/2}^*) = 1 - \alpha,$$

so $a + \hat{\theta} = a_{\alpha/2}^*$ and $b + \hat{\theta} = a_{1-\alpha/2}^*$. Now you are almost done if you go back to (12).

Student: I think I understand, this implies that:

$$P(\hat{\theta} - a_{1-\alpha/2}^* + \hat{\theta} \le \theta \le \hat{\theta} - a_{\alpha/2}^* + \hat{\theta}) = P(2\hat{\theta} - a_{1-\alpha/2}^* \le \theta \le 2\hat{\theta} - a_{\alpha/2}^*)$$

and the resulting confidence interval is

$$(2\hat{\theta} - a_{1-\alpha/2}^*, 2\hat{\theta} - a_{\alpha/2}^*), \tag{13}$$

which looks quite different from (11)!

Teacher: Yes, but there are situations when they produce similar endpoints. If we put $a^*_{\alpha/2} = 2\hat{\theta} - a^*_{1-\alpha/2}$, we arrive at

$$\hat{\theta} = \frac{a^*_{\alpha/2} + a^*_{1-\alpha/2}}{2} \tag{14}$$

and we get the same result starting with the upper bound. What does this mean?

Student: Symmetry!

Teacher: Yes, the empirical distribution of θ^* tries to capture the distribution of $\hat{\theta}$ and (14) implies that this distribution is symmetric. In this case we know that $\bar{X}$ is skewed, so these intervals should produce different results. However, as (Navidi, 2006) has put it: "...which is better is a matter of some controversy".

There is a command in MATLAB yielding bootstrap resamples. For tomorrow, use that to compute the two bootstrap intervals.

5. The Fourth Day

Student: We simulated 10 000 bootstrap resamples. So, the approximate 95% percentile confidence interval for μ was (19.0, 36.0) and the corresponding basic interval (20.0, 37.0). They have the same length and are shorter than all the previous intervals. We became curious enough to carry out a simulation study in order to find out the coverage properties of these intervals.

Teacher: Good, I am curious too!

Student: We wanted to be able to compare the results with the earlier simulation study for the Wald and score intervals. Therefore, we chose $\theta = 4$ and $n = 10$ (as in our hypothetical example), so $\mu = 40$. The results were not good for the bootstrap intervals with coverage rates 91.3% (percentile) and 90.4% (basic). 10 000 Poisson samples were drawn and for each sample 1000 bootstrap resamples were generated. For reference, the coverage rate for the Wald interval was 94.1%.

We guessed that $n = 10$ is not enough for the bootstrap method?

Teacher: No, but you could try $\theta = 2$ and $n = 20$ and $\theta = 1$ combined with $n = 40$. This will improve the results for the bootstrap method, while not affecting the Wald interval. But before you do this I would like to throw in yet another bootstrap interval! We could start by considering the pivotal

$$t^* = \frac{\theta^* - \hat{\theta}}{se(\theta^*)},$$

for each bootstrap resample. The empirical distribution of t^* then gives us percentiles $t^*_{\alpha/2}$ and $t^*_{1-\alpha/2}$ (t^* not necessarily symmetric around 0). From this, what will the resulting confidence interval be?

I will give you before the next break to think about this in your groups.

5.1 After the Break

Student: In this situation we assume that the distribution of

$$t^* = \sqrt{n}\frac{\bar{X}^* - \bar{X}}{\sqrt{\bar{X}^*}}$$

should "imitate" that of

$$\sqrt{n}\frac{\bar{X} - \theta}{\sqrt{\bar{X}}},$$

so the resulting interval is

$$\left(\bar{X} - \frac{t^*_{1-\alpha/2}}{\sqrt{n}}\sqrt{\bar{X}}, \bar{X} - \frac{t^*_{\alpha/2}}{\sqrt{n}}\sqrt{\bar{X}}\right) \tag{15}$$

Teacher: Quite right, but you have actually assumed that the underlying distribution is the Poisson, since you let $se(\hat{\theta}) = \bar{X}/\sqrt{n}$, so you get a kind of mixture between the non-parametric and parametric bootstrap. You might call this a "plug-in" estimator and the result is called a bootstrap t-interval, This is brought up in e.g. (Efron & Tibshirani, 1993), Efron being the inventor of the bootstrap. For a pure non-parametric approach, you should have used $se(\hat{\theta}) = S/\sqrt{n}$.

Student: You are right of course, but we also thought about trying a "score" bootstrap interval. Since we know that the score interval is at least an improvement of the Wald interval, we could instead start with

$$t_{sc}^* = \sqrt{n}\frac{\bar{X}^* - \bar{X}}{\sqrt{\bar{X}}}$$

and arrive at

$$P(t_{sc,\alpha/2}^* < \sqrt{n}\frac{\bar{X} - \theta}{\sqrt{\theta}} < t_{sc,1-\alpha/2}^*) \approx 1 - \alpha$$

Then it is just about finding those values of θ for which

$$t_{sc,\alpha/2}^* < \sqrt{n}\frac{\bar{x} - \theta}{\sqrt{\theta}} < t_{sc,1-\alpha/2}$$

Teacher: Good idea. All groups can work on this for a while.

After 20 minutes:

Student: We started with the "observed" pivotal

$$\sqrt{n}\frac{\bar{x} - \theta}{\sqrt{\theta}},$$

which is a decreasing function of θ, so we should get the lower limit from letting

$$\sqrt{n}\frac{\bar{x} - \theta}{\sqrt{\theta}} = t_{sc,1-\alpha/2}^*$$

and the upper limit from

$$\sqrt{n}\frac{\bar{x} - \theta}{\sqrt{\theta}} = t_{sc,\alpha/2}^*$$

Teacher: Sound reasoning!

Student: We put $\beta = \sqrt{\theta}$ in both cases and in the first we got

$$\beta^2 + \frac{t_{sc,1-\alpha/2}^*}{\sqrt{n}}\beta - \bar{x} = 0$$

with roots

$$-\frac{t_{sc,1-\alpha/2}^*}{2\sqrt{n}} \pm \sqrt{\bar{x} + \frac{(t_{sc,1-\alpha/2}^*)^2}{4n}}$$

Only the positive root is possible here and

$$\left(-\frac{t_{sc,1-\alpha/2}^*}{2\sqrt{n}} + \sqrt{\bar{x} + \frac{(t_{sc,1-\alpha/2}^*)^2}{4n}}\right)^2 = \bar{x} + \frac{(t_{sc,1-\alpha/2}^*)^2}{2n} + \frac{t_{sc,1-\alpha/2}^*}{\sqrt{n}}\sqrt{\bar{x} + \frac{(t_{sc,1-\alpha/2}^*)^2}{4n}}$$

From this the upper limit must be

$$\bar{x} + \frac{(t_{sc,\alpha/2}^*)^2}{2n} + \frac{t_{sc,\alpha/2}^*}{\sqrt{n}}\sqrt{\bar{x} + \frac{(t_{sc,\alpha/2}^*)^2}{4n}}$$

Teacher: Correct! And now having all these intervals we should summarize our findings. I want you to perform simulations to study coverage and non-coverage rates for the seven types of intervals included in Table 1 and the four bootstrap intervals: percentile, basic, *t* and the last proposal, which we can name "score". Draw 10 000 Poisson samples, all of size $n = 10$. For the bootstrap intervals 1000 resamples should be drawn. Put the nominal confidence level to 95%. Let further $\theta = 1, 1.5, 4$ and 10.

Good luck!

6. The Fifth Day

Student: We have put our results in a table.

Teacher: This will be interesting to study. What are your comments on the results in your table?

Table 2. Coverage (CR) and Non-Coverage (lower: LNCR, upper: UNCR) rates for the Wald, Score, Transformation and Bootstrap Intervals with Nominal Confidence Level 95%

	$\theta = 1$			$\theta = 1.5$		
	CR	LNCR	UNCR	CR	LNCR	UNCR
Wald	.926	.067	.007	.915	.073	.012
score	.964	.009	.027	.953	.019	.028
VST	.956	.029	.015	.943	.039	.018
VST var corr	.956	.029	.015	.943	.039	.018
VST bias corr	.944	.029	.027	.943	.039	.018
symm T	.918	.067	.015	.943	.039	.018
"score" symm T	.944	.029	.027	.943	.039	.018
percentile bs	.915	.066	.019	.918	.060	.022
basic bs	.904	.079	.017	.905	.072	.023
t bs	.944	.012	.044	.928	.031	.041
"score" bs	.901	.045	.054	.902	.052	.046

	$\theta = 4$			$\theta = 10$		
	CR	LNCR	UNCR	CR	LNCR	UNCR
Wald	.944	.042	.014	.943	.038	.019
score	.953	.018	.029	.948	.023	.029
VST	.951	.029	.020	.947	.029	.024
VST var corr	.951	.029	.020	.947	.029	.024
VST bias corr	.951	.029	.020	.947	.029	.024
symm T	.951	.029	.020	.947	.029	.024
"score" symm T	.951	.029	.020	.947	.029	.024
percentile bs	.913	.054	.033	.904	.056	.040
basic bs	.904	.064	.032	.898	.061	.041
t bs	.905	.044	.051	.897	.051	.052
"score" bs	.901	.048	.051	.896	.052	.052

Student: Compared with the transformation and bootstrap based intervals, the "old" score interval is still very competitive. The transformation based intervals also perform well and actually yield identical results except for the smallest value of θ.

As we saw earlier, the percentile and basic bootstrap intervals are not accurate with respect to either coverage or non-coverage. This goes for the other two bootstrap intervals as well, but the "score" interval is, at least, well balanced for all values of θ. Furthermore we have, as you suggested, tested other combinations of n and θ corresponding to $\mu = 40$. The results improve for increasing n, but the coverage rates are still below the nominal level (Table 3).

Table 3. Coverage (CR) and Non-Coverage (lower: LNCR, upper: UNCR) rates for Bootstrap Intervals with Nominal Confidence Level 95%.

	$\theta = 4,\ n = 10$			$\theta = 2,\ n = 20$		
	CR	LNCR	UNCR	CR	LNCR	UNCR
percentile	.913	.054	.033	.934	.045	.021
basic	.904	.064	.032	.927	.053	.020
t	.905	.044	.051	.929	.034	.037
"score"	.901	.048	.051	.924	.038	.038
	$\theta = 1,\ n = 40$			$\theta = .5,\ n = 80$		
	CR	LNCR	UNCR	CR	LNCR	UNCR
percentile	.943	.037	.020	.951	.032	.017
basic	.936	.046	.017	.943	.043	.014
t	.939	.028	.033	.948	.026	.026
"score"	.934	.031	.035	.943	.029	.028

Teacher: Nearly all of these bootstrap intervals do indeed have coverage rates below 95%. Having observed this it should also be pointed out that we will obtain increased coverage rates using the parametric bootstrap, where the resamples are generated from the Poisson distribution, see e.g. (Patil & Rajarshi, 2010). A very short summary of your results is that the "good old" score interval is the most preferable of all the intervals you have studied in terms of maximizing the coverage rate. For some cases it is clearly conservative though.

As you have shown empirically, the differences between the transformation based intervals with respect to coverage and non-coverages are practically nil. If we compare the score interval

$$\bar{x} + \frac{z_{\alpha/2}^2}{2n} \pm \frac{z_{\alpha/2}}{\sqrt{n}} \sqrt{\bar{x} + \frac{z_{\alpha/2}^2}{4n}}$$

with the bias corrected VST interval

$$\left(\sqrt{\bar{x} + \frac{1}{4n}} \pm \frac{z_{\alpha/2}}{2\sqrt{n}}\right)^2 = \bar{x} + \frac{z_{\alpha/2}^2 + 1}{4n} \pm \frac{z_{\alpha/2}}{\sqrt{n}} \sqrt{\bar{x} + \frac{1}{4n}},$$

we find that the mid-point of the score interval is shifted away (to the right) from $\bar{x}$ almost twice as far as for the bias corrected VST interval. From the results this seems to increase the coverage and yield more balanced non-coverages.

Another comment: In our study we have not compared lenghts of intervals, but from the interval expressions we may observe that the score interval is slightly longer than the transformation based intervals, see e.g. (Barker, 2002).

By the way, we have the Bayesian approach left to investigate for this situation. Where are you going?

Student: We do have other classes, you know!

7. Summary

The Poisson process/distribution was used to illustrate various techniques for construction of confidence intervals in an educational context. The students will learn more about the properties of the Poisson distribution and about some of the many options available to obtain confidence intervals of the mean. Most important though is to convey to the students a scientific and analytic way of thinking. We started with an explicit problem and discovered that the standard method did not lead to satisfactory results. We then tried to find out why the method failed and started looking for alternative approaches, which means that we need more advanced tools, such as the Delta Mehod. From a theoretical point of view the discussion presented is not very far from the research front. Generally, as commented in (DasGupta, 2008) on page 59: "No studies seem to have been made that compare bias-corrected VSTs to bias-corrected symmetrizing transforms."

References

Andersson, P. G. (2015). A Classroom Approach to the Construction of an Approximate Confidence Interval of a Poisson Mean Using One Observation. *The American Statistician*, *69*, 159-164. https://doi.org/10.1080/00031305.2015.1056830

Anscombe, F. J. (1948). The Transformation of Poisson, Binomial and Negative-Binomial Data. *Biometrika*, *37*, 358–382. https://doi.org/10.1093/biomet/37.3-4.358

Barker, L. (2002). A Comparison of Nine Confidence Intervals for a Poisson Parameter When the Expected Number of Events is ≤ 5. *The American Statistician, 56*, 85–89. https://doi.org/10.1198/000313002317572736

Bartlett, M. S. (1936). The Square Root Transformation in Analysis of Variance. *Supplement to the Journal of the Royal Statistical Society*, *3*, 68–78. https://doi.org/10.2307/2983678

Brown, L. D., Cai, T. T., & DasGupta, A. (2014). On Selecting a Transformation: With Applications. *Technical report,* Purdue University, USA.

Casella, G., & Berger, R. L. (2002). *Statistical Inference, Second Edition*, Duxbury Advanced Series.

Efron, B., & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*, Chapman & Hall. https://doi.org/10.1007/978-1-4899-4541-9

DasGupta. (2008). *Asymptotic Theory of Statistics and Probability*, Springer.

Hogg, R. V., McKean, J. W., & Craig, A. T. (2005). *Introduction to Mathematical Statistics, sixth edition*, Pearson - Prentice Hall.

Navidi, W. (2006). *Statistics for Engineers and Scientists*, McGraw-Hill.

Patil, D. D., & Rajarshi, M. B. (2010). Comparison of Confidence Intervals for the Poisson Mean. *Model Assisted Statistics and Applications*, *5*, 79–91.

Swift, M. B. (2009). Comparison of Confidence Intervals for a Poisson Mean – Further Considerations. *Communication in Statistics – Theory and Methods*, *38*, 748–759. https://doi.org/10.1080/03610920802255856