CDF/MEMO/STATISTICS/PUBLIC/6438
Version 1

Coverage of Error Bars for Poisson Data

Joel G. Heinrich—University of Pennsylvania

May 2, 2003

Abstract

The frequentist concept of coverage is explained, and illustrated by calculating the coverage properties of eight error bar schemes for Poisson data. The primary goal is to aid physicists in doing their own coverage calculations, but some conclusions are also drawn concerning the relative coverage performance of the eight schemes. While mostly intended for beginners, some advanced concepts are also introduced.

1 Introduction

When physicists determine an unknown parameter from experimental data, they also provide error bars. The central value and the error bars, often written $V^{+\sigma_2}_{-\sigma_1}$, determine an interval $[V-\sigma_1, V+\sigma_2]$, which is the region within the error bars. One quantity of interest associated with such an interval is the coverage probability (usually just called the coverage) of the interval. We define coverage in the following single-parameter example:

We will assume that there is a single unknown parameter μ that is estimated from the data $\vec{x}$. The experimenters have functions that determine the central value and errors from the data—$V(\vec{x})$, $\sigma_1(\vec{x})$, and $\sigma_2(\vec{x})$ respectively—for all possible data $\vec{x}$. (These functions often are defined by, and implemented through, some fitting procedure.)

The data follow a probability distribution $p(\vec{x},\mu)$ that depends on μ, and is completely known once a specific numerical value for μ is picked. The coverage $C(\mu)$ is a function of μ, defined as the probability that

$$V(\vec{x}) - \sigma_1(\vec{x}) \leq \mu \leq V(\vec{x}) + \sigma_2(\vec{x})$$

for random $\vec{x}$ generated from $p(\vec{x},\mu)$. It is important to note that in this equation, μ is regarded as fixed, and the probability statement applies to the "variables" (really functions) V, σ_1, and σ_2, which in turn depend on $\vec{x}$.

So the coverage is a function $C(\mu)$ of the unknown parameter. It is, for any given μ, the probability that an experiment, with data following the same distribution $p(\vec{x}, \mu)$, and employing the same analysis functions $V(\vec{x})$, $\sigma_1(\vec{x})$, and $\sigma_2(\vec{x})$, will obtain an interval that includes (i.e. "covers") μ. The Holy Grail of Frequentist Statistics is to define $V(\vec{x})$, $\sigma_1(\vec{x})$, and $\sigma_2(\vec{x})$ so that $C(\mu) = C_0$, C_0 being some predefined constant. This optimum case (constant $C(\mu)$) is referred to as exact coverage. By default, physicists use, and will normally assume, $C_0 = 0.682689492137\ldots$ (the area of the Gaussian distribution contained within $\pm 1\sigma$) as the coverage value for error bars[1]: if any other coverage value is used (for error bars), it must be explicitly stated to avoid misunderstandings.

Exact coverage, like the Grail of legend, if approached by any but a perfectly pure and holy frequentist, is borne away and vanishes from sight. So, the following examples mainly demonstrate how badly the coverage deviates from exact when applying common methods to Poisson distributed data.

In the sections that follow we will investigate the coverage $C(\mu)$ achieved by eight different error-bar schemes for Poisson data with no background. In fact, exact coverage cannot be achieved through normal means for the Poisson distribution, or discrete distributions in general—the reasons for this will be manifest upon understanding the examples.

2 Pearson's χ^2 Intervals

In this section we examine the coverage of intervals derived from Pearson's χ^2.[1, 2] Specifically, we choose a Poisson process characterized by parameter $\mu \geq 0$ from which we observe n events. The probability of observing n events is

$$p(n, \mu) = \frac{e^{-\mu} \mu^n}{n!}$$

and Pearson's χ^2 is given by

$$\chi^2(\mu, n) = \frac{(n - \mu)^2}{\mu}$$

(We give the arguments of the probability as (n, μ) to indicate that n is variable and μ is fixed. For the χ^2, it's μ that is variable and n that is fixed, so the arguments are swapped.)

[1] There are other physics conventions for limits, where $C(\mu) = 95\%$ is a common choice.

Having observed n events, we obtain our central estimate V of the unknown parameter μ by minimizing the χ^2 with respect to μ, obtaining $V = n$. For the error bars, we adopt the interval defined as the set of all μ such that $\chi^2(\mu, n) \leq \Delta$. In the language of Minuit[3], these are MINOS errors, and Δ is the Minuit `ERRDEF` parameter. The MINOS error bars are defined as the change in parameter value (μ) required to increase the function value (χ^2) by `ERRDEF` (Δ). In this simple example, the errors are

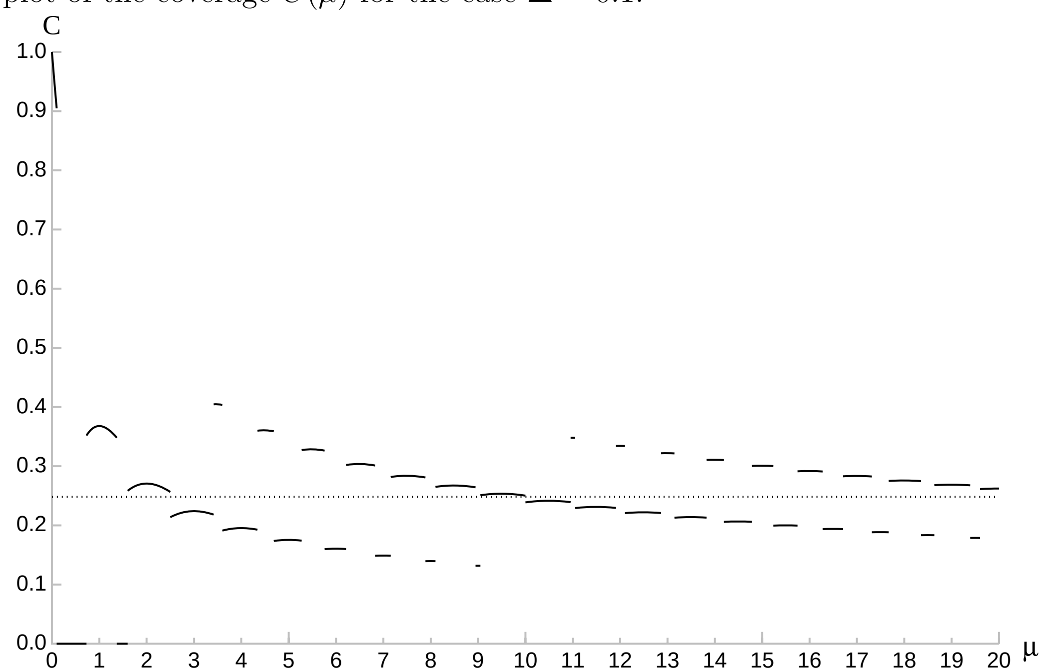
$$\sigma_1 = \sqrt{n\Delta + \Delta^2/4} - \Delta/2 \qquad\qquad \sigma_2 = \sqrt{n\Delta + \Delta^2/4} + \Delta/2$$

and we also have the useful relations $\sigma_1\sigma_2 = n\Delta$ and $\sigma_2 - \sigma_1 = \Delta$. For no observed events ($n = 0$), the above formulas still are valid, and we have $\sigma_1 = 0$ and $\sigma_2 = \Delta$.

The interval $[\mu_1, \mu_2]$ that defines these error bars satisfies

$$\mu_1 = n + \Delta/2 - \sqrt{n\Delta + \Delta^2/4} \qquad\qquad \mu_2 = n + \Delta/2 + \sqrt{n\Delta + \Delta^2/4}$$

and $\mu_1\mu_2 = n^2$, and we wish to calculate the coverage of this interval. Note that the size of the interval, $\sqrt{4n\Delta + \Delta^2}$, grows steadily with the observed number of events n. At this point in the discussion, it is useful to show a plot of the coverage $C(\mu)$ for the case $\Delta = 0.1$:



Coverage (C) vs μ: $\chi^2 < 0.1$ (C → 0.2482 as μ → ∞)

$\Delta = 0.1$ is not the usual choice; this plot is shown first because it is easier to explain than the plot to follow with $\Delta = 1$. There are two shocks to recover from: $C(\mu)$ is discontinuous at ~ 40 points in the region $\mu < 20$, and the value of $C(\mu)$ is all over the map. The first time one sees a plot like this, one assumes that some bug must exist, but, of course, the plot turns out to be correct.

The following explanation should help: At $\mu = 0$, the coverage must be 100%—zero events are aways observed, and the $n = 0$ interval [0,0.1] contains $\mu = 0$ every time. As μ increases slightly, but remains less than 0.1, occasionally $n \geq 1$ is observed, but the interval for $n = 1$ is [0.7298,1.3702], which does not cover $\mu < 0.1$. So $C(\mu) = p(0, \mu) = e^{-\mu}$ for $\mu < 0.1$. For μ greater than 0.1 and less than 0.7298, no possible n has an error interval that covers μ. So $C(\mu) = 0$ for $0.1 < \mu < 0.7298$. When μ is within the region [0.7298,1.3702], only $n = 1$ covers, since the $n = 2$ interval is [1.6,2.5] (exactly). So $C(\mu) = p(1, \mu) = \mu e^{-\mu}$ for $\mu \in [0.7298, 1.3702]$.

Similarly, after another zero coverage region, we have $C(\mu) = p(2, \mu) = \mu^2 e^{-\mu}/2$ for $\mu \in [1.6, 2.5]$. The interval when $n = 3$ is observed is [2.5,3.6] (exactly), and the interval for $n = 4$ is [3.4156,4.6844], which overlaps the $n = 3$ interval. So, for $\mu \in [2.5, 3.4156]$, $C(\mu) = p(3, \mu) = \mu^3 e^{-\mu}/6$. For $\mu \in [3.4156, 3.6]$, both the $n = 3$ and $n = 4$ cases cover, and $C(\mu) = p(3, \mu) + p(4, \mu) = \mu^3 e^{-\mu}/6 + \mu^4 e^{-\mu}/24$.

After that, the form $C(\mu) = (k, \mu)$ alternates with $C(\mu) = p(k, \mu) + p(k+1, \mu)$ for a while. For $\mu > 10$, the error intervals are wide enough so that there are regions where three of them overlap, and $C(\mu) = p(k, \mu) + p(k + 1, \mu)$ alternates with $C(\mu) = p(k, \mu) + p(k + 1, \mu) + p(k + 2, \mu)$.

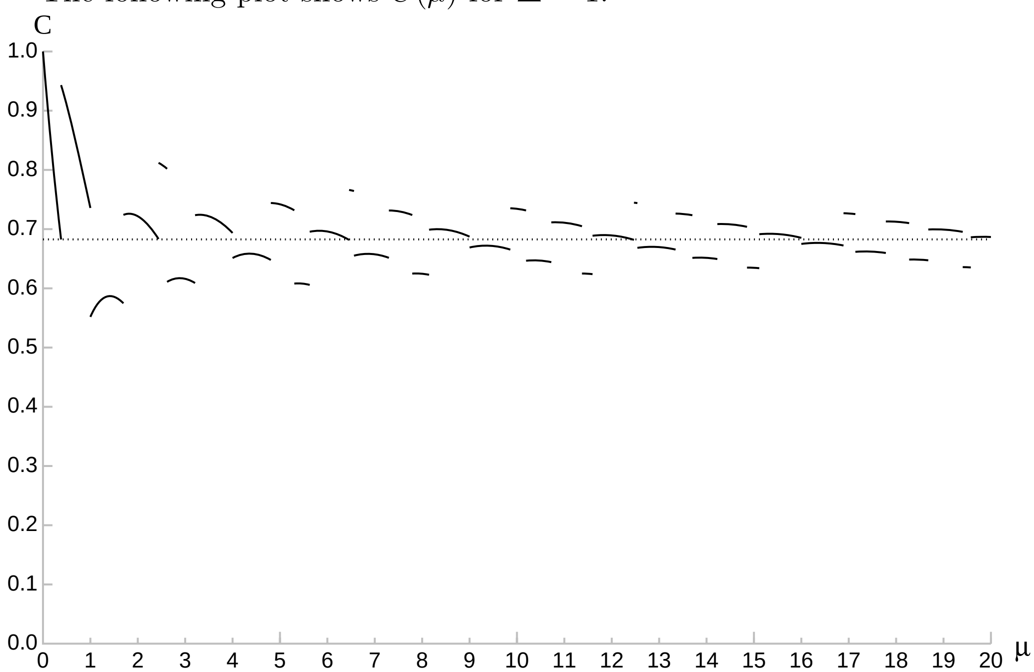
Since discontinuities in $C(\mu)$ occur at the beginning and the end of each interval, within the region $\mu < 20$ there are about 40 discontinuities. This property—average continuous segment width of 0.5—must hold quite accurately over large regions. Strictly speaking, there are two segments of zero width not mentioned above: for $\mu = 2.5$ exactly, under our definition, both the $n = 2$ and $n = 3$ intervals cover, so $C(2.5) = p(2, 2.5)+p(3, 2.5) = 0.4703$. But line segments of zero size simply don't show up on the plot. The other orphan coverage point is at $\mu = 10$.

The following trivial C program suffices to calculate the coverage of Pearson's χ^2 intervals for any μ small enough that $e^{-\mu}$ does not underflow. It expects two command-line arguments: the first is μ, and the second is Δ. As the entire calculation only takes a half a dozen lines or so of C-code, a careful examination of the logic should reward the reader with an improved understanding.

```
#include <stdio.h>
#include <stdlib.h>
#include <math.h>
double chi2(double mu,int n) { return (n) ? (n-mu)*(n-mu)/mu : mu; }

int main(int argc, char* argv[]) {
  const double    mu = (argc>1) ? strtod(argv[1],NULL) : 0.0;
  const double delta = (argc>2) ? strtod(argv[2],NULL) : 1.0;
  double sum=0.0, p=exp(-mu);
  int n;
  for(n=0 ; p>0 ; p *= mu/(++n)) {
    if( chi2(mu,n) <= delta )
      sum += p;
    else if(n>mu)
      break;
  }
  printf("mu=%g delta=%g coverage=%g\n",mu,delta,sum);
  return 0;
}
```

The following plot shows $C(\mu)$ for $\Delta = 1$:



Coverage (C) vs μ: $\chi^2 < 1$ (C → 0.6827 as μ → ∞)

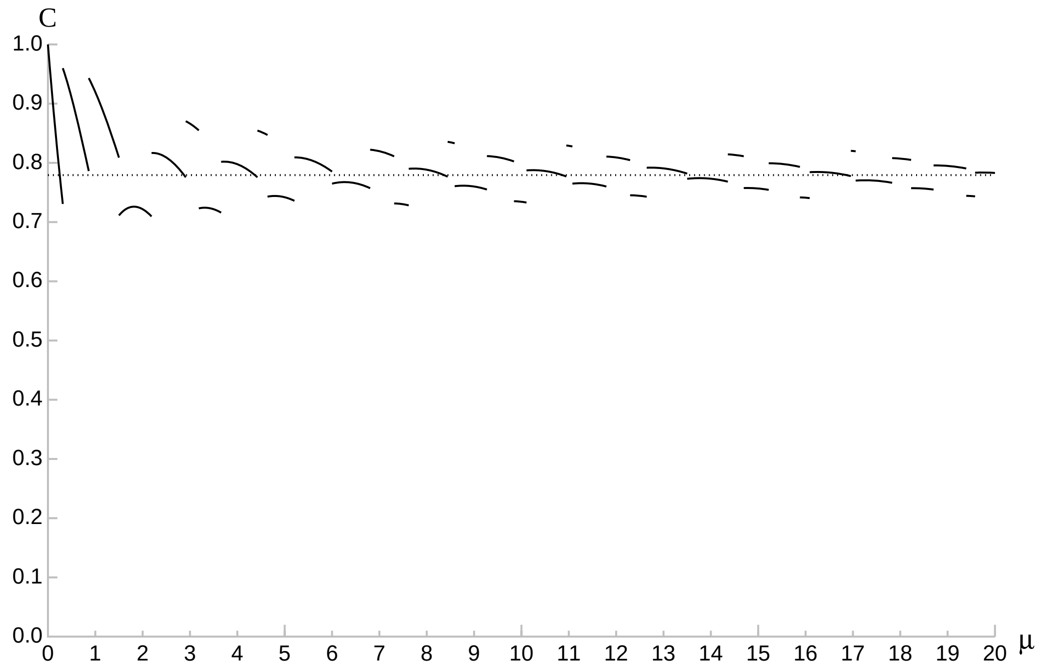
$\Delta = 1$, of course, is the physicist's standard choice for 1σ error bars. The minimum value for $C(\mu)$ on this plot is $1.5e^{-1} = 0.5518$, which is attained in

the limit as μ approaches 1 from above. There is an orphan point at exactly $\mu = 1$, and $C(1) = 2.5e^{-1} = 0.9197$. (To the attentive reader, it will be obvious from looking at the plot that there are also orphan points at $n = 4$, $n = 9$, and $n = 16$.) It seems amazing that one obtains such a complicated structure from such a simple rule. The program given above is sufficient to calculate the coverage for a given value of μ, but it won't indicate where the discontinuities are located; the discontinuities occur at the interval boundaries. The plot is actually produced by locating all the discontinuities first, sorting them in ascending order, and then plotting $C(\mu)$ as continuous curves between the discontinuities.

Suppose we do an experiment and we observe 6 events. Under the above rule, we report $\mu = 6^{+3}_{-2}$. From the frequentist point of view, μ is still an unknown parameter: it could be in the neighborhood of 1 (where the minimum coverage occurs). All we can say about the coverage with absolute certainty is that it is greater than or equal to 55.18% (and less than or equal to 1). Average coverage is a Bayesian concept; for the frequentist, μ, although unknown, has a definite and fixed value[2]. In fact, even if 100 events were observed, the $C \geq 55.18\%$ conclusion is still the only strictly valid statement, although in practice most frequentists would grant that $C \simeq 68.27\%$ is a reasonable approximation for 100 observed events.

But frequentists will generally demand (if the number of observed events is not large) that the minimum coverage be at least 68.27%. One simple strategy is just to boost Δ until this is achieved. In fact, this is first achieved when $\Delta = 1.5$ (minimum coverage as a function of Δ is also discontinuous). This case is shown in the following plot:

[2]The Bayesian and the frequentist both agree that μ has a definite and fixed value; the Bayesian assigns a probability distribution to μ to represent the prior state of his knowledge of what that true value might be.



Coverage (C) vs μ: $\chi^2 < 1.5$ (C → 0.7793 as μ → ∞)

Here the minimum coverage is 0.7095, achieved in the neighborhood of $\mu = 2.1883$. Unfortunately this leads to serious overcoverage[3]: the average coverage is $\sim 78\%$. When physicists see a plot that overcovers significantly (i.e., the error bars cover the theoretical curve at greater than 68%), they tend to accuse the authors of overstating their errors—or of being biased by the theory[4].

3 Neyman's Modified χ'^2 Intervals

Instead of Pearson's χ^2, often Neyman's modified[5] χ'^2 is used:

$$\chi'^2(\mu, n) = \frac{(n - \mu)^2}{n}$$

where the number of observed events n replaces μ in the denominator. Although asymptotically approaching Pearson's χ^2 for large n, χ'^2 is generally thought to be inferior to Pearson's χ^2 at small n. One must also make some choice about what to do when $n = 0$. Once again, we take the error interval

[3] It also overcovers for all μ, since $C(\mu)$ is greater than 68.27% everywhere.

to be the set of all μ such that $\chi'^2(\mu, n) \leq \Delta$. The errors, symmetric for $n \geq \Delta$, are then given by

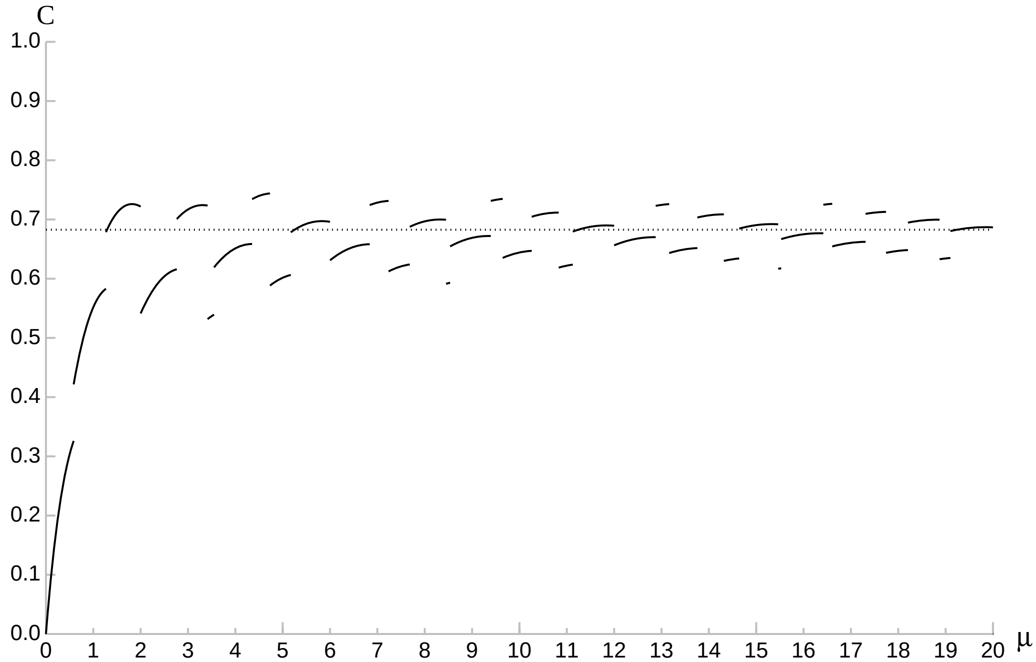
$$\sigma_1 = \min(n, \sqrt{n\Delta}) \qquad\qquad \sigma_2 = \sqrt{n\Delta}$$

the interval is defined by

$$\mu_1 = \max(n - \sqrt{n\Delta}, 0) \qquad\qquad \mu_2 = n + \sqrt{n\Delta}$$

and we have $\mu_1\mu_2 = n \max(n - \Delta, 0)$. For $n = 0$, we take $\sigma_1 = \sigma_2 = 0$, which is simply the limit of the above expressions.

The coverage $C(\mu)$ for the standard case $\Delta = 1$ is shown here:



Coverage (C) vs μ: $\chi'^2 < 1$ (C → 0.6827 as μ → ∞)

For $0 < \mu < 2 - \sqrt{2}$, only the interval corresponding to $n = 1$ includes μ, so $C(\mu) = \mu e^{-\mu}$ in this region, and the coverage approaches zero in the limit as $\mu \to 0$. There is an orphan coverage point at $\mu = 0$: $C(0) = 1$ by our definition, since only in that special case is μ included in the $n = 0$ interval. There are also orphan coverage points at $\mu = 2$, 6, 12, and 20.

From the frequentist point of view, the fact that the minimum coverage is zero (for any choice of Δ) is the worst possible outcome. This fate could have been avoided by picking some other (ad hoc) interval for $n = 0$, but we thought it best instead to illustrate what goes wrong in the given case.

Comparing with the corresponding Pearson's χ^2 plot on page 5, it is interesting that there the continuous coverage segments have a general negative slope, while here the segments, although arranged in a similar pattern, have a predominantly positive slope. Neither trend seems present in the $-2\ln\lambda$ plot of page 10 in the next section, where the continuous segments tend to center more closely about their peak location.

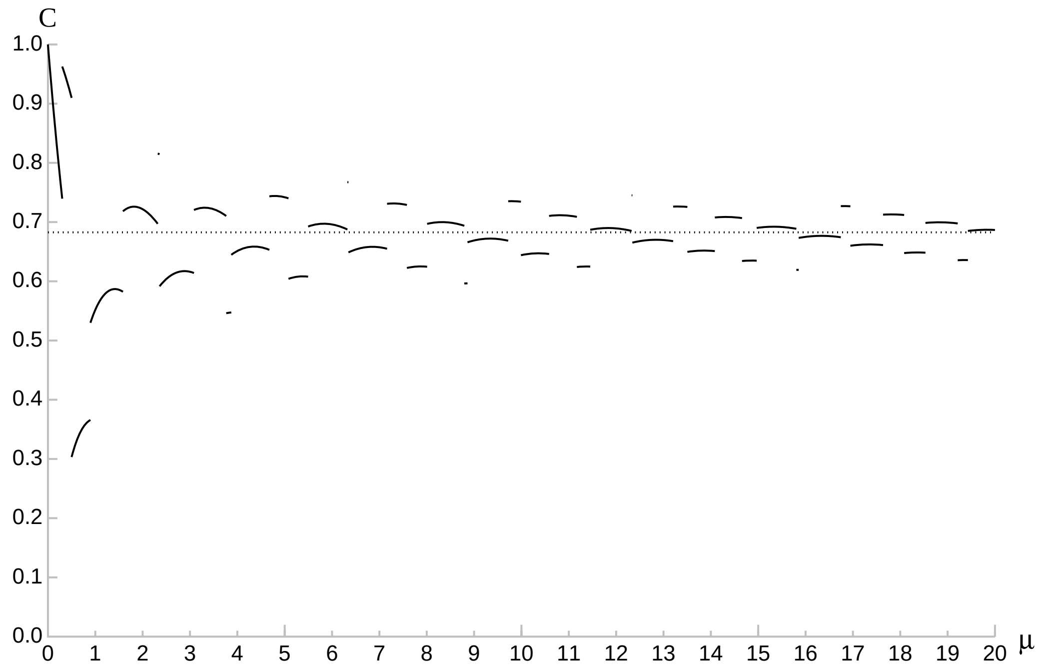
4 Likelihood Intervals

Instead of Pearson's χ^2, or Neyman's modified χ'^2, we can also try error intervals based on the value of the likelihood. Specifically, having observed n events, we can use the error interval defined as the set of all μ such that $-2\ln\lambda(\mu,n) \leq \Delta$, where

$$-2\ln\lambda(\mu,n) = 2[(\mu - n) + n\ln(n/\mu)]$$

is -2 times the log likelihood ratio[4] of the Poisson distribution[6, 7]. This is the quantity that is minimized when one does a maximum-likelihood fit to the Poisson distribution. Once again, standard error bars correspond to $\Delta = 1$. The next plot shows the coverage for this case:

[4] The likelihood ratio $\lambda = p(n,\mu)/p(n,\mu_{\text{best}})$, where μ_{best} is the value of μ that maximizes $p(n,\mu)$ (n being treated as constant). Considered as a function of μ, λ is simply the likelihood renormalized so that the maximum value it can take is 1. In the Poisson case, $\lambda = p(n,\mu)/p(n,n)$. By definition, maximizing the likelihood (with respect to μ) is equivalent to maximizing λ or minimizing $-2\ln\lambda$.

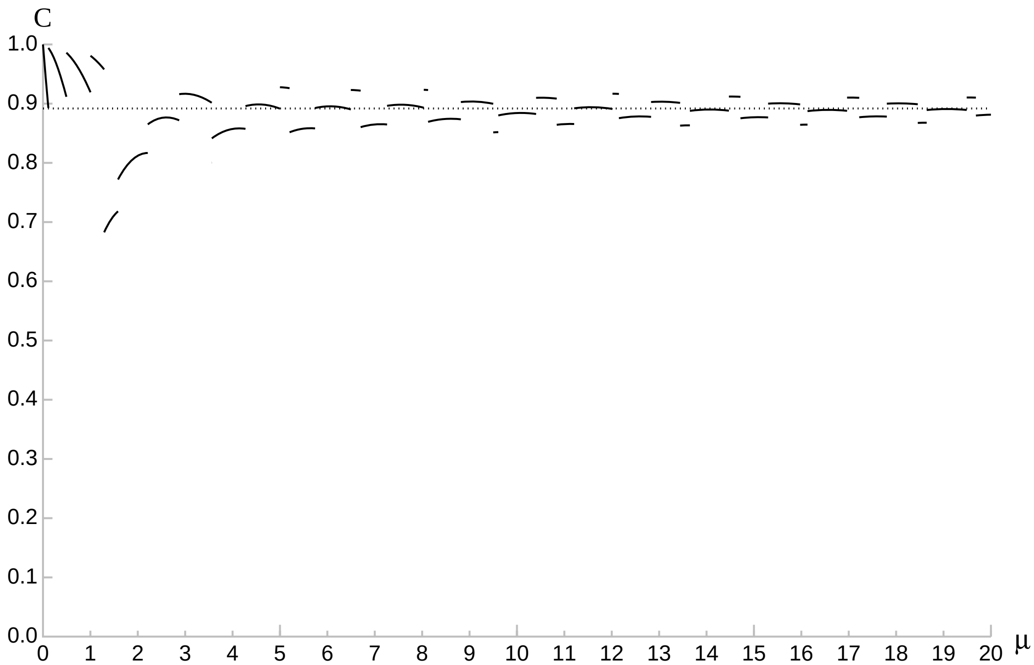


Coverage (C) vs μ: −2lnλ < 1 (C → 0.6827 as μ → ∞)

The minimum coverage, 0.3033, occurs in the neighborhood of $\mu = 0.5$. Surprisingly[5], this is worse than for $\Delta = 1$ Pearson's χ^2 intervals. There are no orphan coverage points—those present in the corresponding Pearson's χ^2 case (compare with the figure on page 5) have "gained weight", and are visible here as short segments.

If we ask again to what value we must increase Δ to obtain a minimum coverage of 68.27%, this time the answer is $\Delta = 2.581$. This is also clearly worse than the corresponding Pearson's χ^2 case. The $\Delta = 2.581$ plot is shown here:

[5]Reference [7] shows that, when the variance is the comparison criterion, $-2\ln\lambda$ is superior to Pearson's χ^2 in the Poisson case.



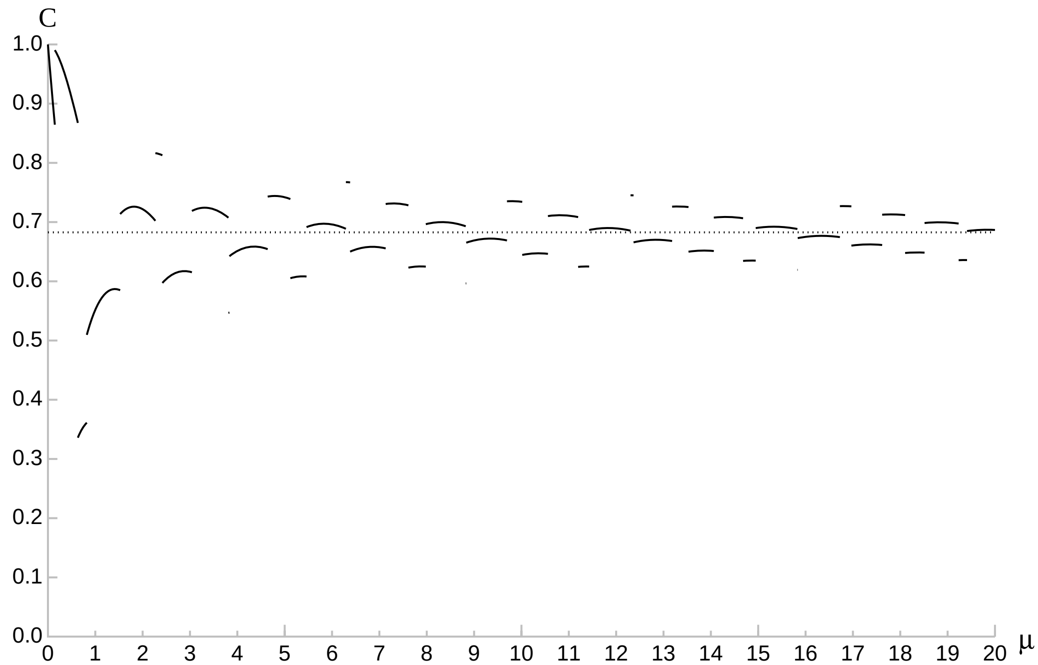
Coverage (C) vs μ: −2lnλ < 2.581 (C → 0.8918 as μ → ∞)

5 "Improved Likelihood Ratio" Intervals

The statistic

$$w = \frac{-2\ln\lambda}{1+\frac{1}{6}\mu^{-1}}$$

is the "improved likelihood ratio"[8, 9, 10] statistic for the Poisson case. As shown in [7], the mean and variance of the Poisson $-2\ln\lambda$ are asymptotically $1+\frac{1}{6}\mu^{-1}+O(\mu^{-2})$ and $2+\frac{2}{3}\mu^{-1}+O(\mu^{-2})$ respectively, when expanded in powers of μ^{-1}. The rationale of the improved likelihood ratio is that w as defined above then has mean $1+O(\mu^{-2})$ and variance $2+O(\mu^{-2})$; i.e., closer to the moments of the χ^2 distribution for 1 degree of freedom (for large μ). The coverage of the resulting intervals is shown here:



Coverage (C) vs μ: w < 1 (C → 0.6827 as μ → ∞)

The coverage looks qualitatively similar to that of the $-2 \ln \lambda$ case shown on page 10.

6 Classical-Frequentist Central-Intervals

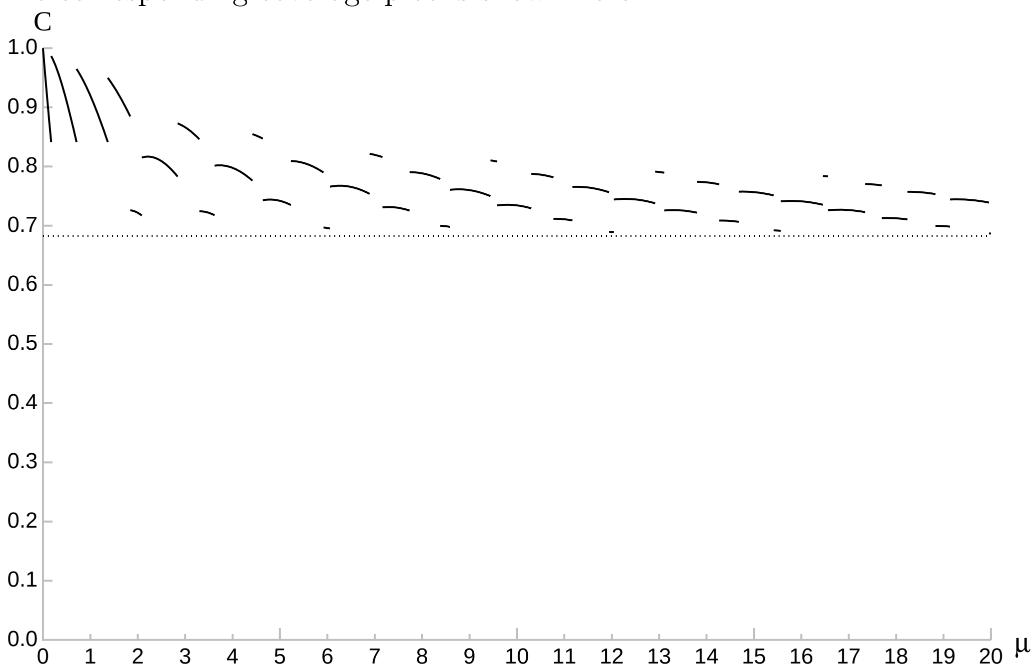
Since, from the frequentist point of view, none of the previous interval schemes have adequate coverage at small μ, we next investigate the coverage achieved by the "68.27%" (central) intervals of the classical frequentist approach. The classical approach to Poisson frequentist intervals is described in Ref. [11]. In this case, the (central) error interval for n observed events is given by the set of all μ such that:

$$\sum_{k=0}^{n} \frac{e^{-\mu}\mu^k}{k!} \geq \frac{1-C_0}{2} \qquad \text{and} \qquad \sum_{k=n}^{\infty} \frac{e^{-\mu}\mu^k}{k!} \geq \frac{1-C_0}{2}$$

The interval for $n = 0$ is defined completely by

$$\mu \leq \ln \frac{2}{1-C_0}$$

since, when $n = 0$, the 2nd inequality, becoming $1 \geq \frac{1-C_0}{2}$, is true for all μ. The corresponding coverage plot is shown here:



Coverage (C) vs μ: Classical Central Intervals (C → 0.6827 as μ → ∞)

The definition of these intervals is tailored so that the minimum coverage is guaranteed to be $\geq C_0$. The average overcoverage is worse here than for the case of the unified intervals (see the figure on page 19) considered next. This seems to be because conservatism is applied twice—there are two inequalities that both need to be satisfied—while the unified approach leads to only a single inequality.

7 Unified Intervals

We next investigate the coverage achieved by the "68.27%" intervals (zero background) of the unified approach[12]. As in the classical frequentist approach, the unified intervals will guarantee that the minimum coverage is $\geq C_0$. The error interval for n observed events is given by the set of all μ satisfying

$$U(\mu, n) = \sum_{k \in \mathcal{A}(\mu,n)} \frac{e^{-\mu}\mu^k}{k!} \; < \; C_0$$

where the set (of zero or more non negative integers) $\mathcal{A}(\mu, n)$ is defined as

$$\mathcal{A}(\mu, n) = \{\, k \mid \; k \in \mathbf{Z},\, k \geq 0, \text{ and } -2\ln\lambda(\mu, k) < -2\ln\lambda(\mu, n) \;\}$$

and $\mathbf{Z}$ denotes the set of integers. The fact that the set $\mathcal{A}(\mu, n)$ is selected using the likelihood ratio is the hallmark of the unified approach. We might (crudely) describe $\mathcal{A}(\mu, n)$ as the set of all integers that give a "better fit" to μ than n does, where "better fit" is defined in terms of the likelihood ratio. Note that $n \notin \mathcal{A}(\mu, n)$.

> Although the terse definition given above is mathematically equivalent to that of Ref. [12], this fact may not be obvious at first glance to readers familiar with that reference. For those readers, the brief explanation in the next paragraph will help to make the connection. (The rest, especially readers unfamiliar with the Neyman construction, may wish to skip the next paragraph entirely.)
>
> In performing the Neyman construction for the Poisson case using likelihood ratio ordering, when constructing the band at any given fixed value of μ, one includes integers k in the band—starting with the one that gives the smallest $-2\ln\lambda(\mu, k)$, and continuing, one by one, in order of smallest $-2\ln\lambda(\mu, k)$ not yet included—only stopping when the probability (given the specified μ) of observing a k within the band finally becomes $\geq C_0$. When one reaches the stage of the construction at which the band is equal to the set $\mathcal{A}(\mu, n)$, the next integer to be considered for inclusion in the band is n, since, by our definition of $\mathcal{A}(\mu, n)$, all k with $-2\ln\lambda(\mu, k) < -2\ln\lambda(\mu, n)$ have already been included in the band. Then, if $\sum_{k\in\mathcal{A}(\mu,n)} e^{-\mu}\mu^k/k! \geq C_0$, the construction of the band (at this μ) has already terminated, meaning n will not be included in the band. On the other hand, if $\sum_{k\in\mathcal{A}(\mu,n)} e^{-\mu}\mu^k/k! < C_0$, the construction of the band must continue, meaning n will be included in the band—independent of what the value of $e^{-\mu}\mu^n/n!$ actually is, or how many additional integers k must be included in the band subsequent to the inclusion of n.

For example, suppose we observe n events, and we want to know if $\mu = n$ is within the error interval for that case. Since $\mathcal{A}(n, n) = \emptyset$ (empty set), $U(n, n) = 0$, which being less than C_0, means that $\mu = n$ is included within the error interval for any $C_0 > 0$ and any n.

Another example: We observe $n = 1$ and we want to know if $\mu = 0.35$ is within the interval for $C_0 = 0.6827$. We have $\mathcal{A}(0.35, 1) = \{0\}$, and

therefore $U(0.35, 1) = e^{-0.35} = 0.7047$. Since $0.7047 \nless 0.6827$, $\mu = 0.35$ is not within the $n = 1$ interval. Based on this, one might falsely suspect that the $n = 1$ interval extends no further down than $-\ln(C_0) = 0.3817$. However, checking $\mu = 0.375$ for coverage when $n = 1$, we have $-2\ln\lambda(0.375, 0) \nless -2\ln\lambda(0.375, 1)$, $\mathcal{A}(0.375, 1) = \emptyset$, $U(0.375, 1) = 0$, and thus $\mu = 0.375$ is contained within the interval. Since $\mathcal{A}(\mu, 1)$ changes from $\{0\}$ to $\emptyset$ at $\mu = e^{-1}$, the low end of the interval for $n = 1$ is given by $\mu_1 = e^{-1} = 0.3679$.

Interestingly, in the unified approach, the width of the error intervals no longer increases at a steady rate with n. This behavior is seen in the following table, which shows the endpoints of the intervals, and their widths, for $n = 0$ to 30:

n	μ_1	μ_2	$\mu_2 - \mu_1$	$-2\ln\lambda(\mu_1, n)$	$-2\ln\lambda(\mu_2, n)$
0	0.0000	1.2904	1.2904	0.0000	2.5807
1	0.3679	2.7505	2.3827	0.7358	1.4775
2	0.7358	4.2504	3.5147	1.4715	1.4853
3	1.1036	5.3012	4.1976	2.2073	1.1865
4	2.3359	6.7764	4.4405	0.9750	1.3356
5	2.7505	7.8064	5.0559	1.4775	1.1577
6	3.8231	9.2783	5.4552	1.0546	1.3256
7	4.2504	10.3006	6.0502	1.4853	1.1931
8	5.3012	11.3187	6.0175	1.1865	1.0852
9	6.3342	12.7905	6.4562	0.9911	1.2544
10	6.7764	13.8060	7.0296	1.3356	1.1617
11	7.8064	14.8194	7.0130	1.1577	1.0819
12	8.8291	16.2920	7.4629	1.0227	1.2456
13	9.2783	17.3043	8.0259	1.3256	1.1724
14	10.3006	18.3152	8.0145	1.1931	1.1075
15	11.3187	19.3249	8.0061	1.0852	1.0495
16	12.3338	20.7991	8.4653	0.9955	1.2039
17	12.7905	21.8084	9.0180	1.2544	1.1480
18	13.8060	22.8169	9.0109	1.1617	1.0971
19	14.8194	23.8247	9.0053	1.0819	1.0506
20	15.8310	25.3003	9.4692	1.0125	1.1972
21	16.2920	26.3079	10.0159	1.2456	1.1512
22	17.3043	27.3150	10.0108	1.1724	1.1087
23	18.3152	28.3216	10.0065	1.1075	1.0692
24	19.3249	29.3278	10.0029	1.0495	1.0325
25	20.3336	30.8049	10.4712	0.9973	1.1699
26	20.7991	31.8110	11.0120	1.2039	1.1328
27	21.8084	32.8169	11.0084	1.1480	1.0980
28	22.8169	33.8223	11.0054	1.0971	1.0653
29	23.8247	34.8275	11.0027	1.0506	1.0345
30	24.8319	36.3057	11.4739	1.0079	1.1648

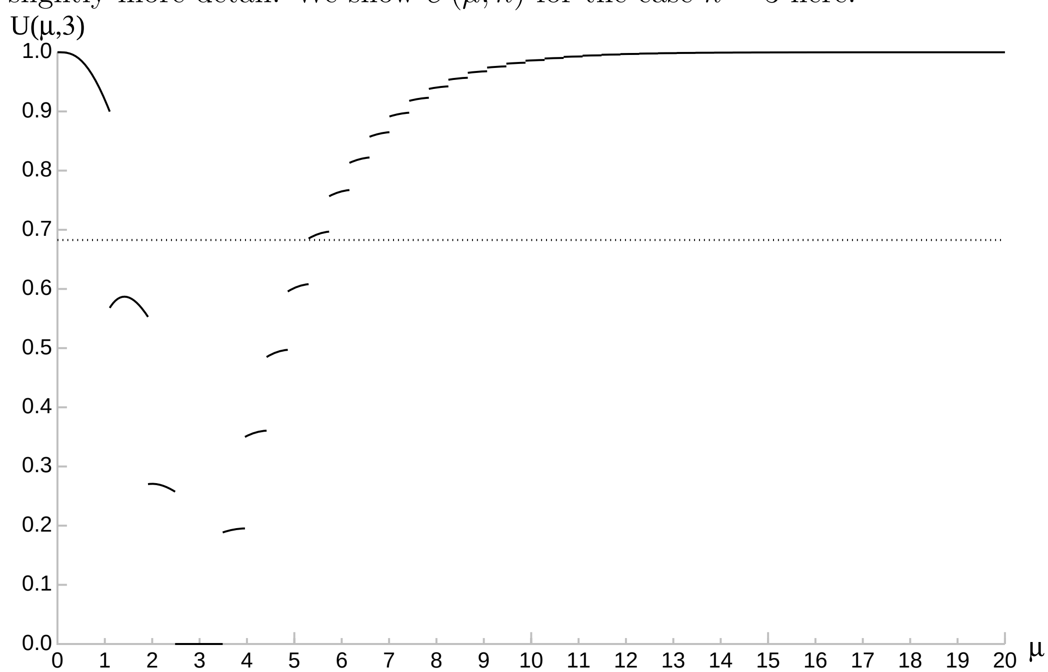
Unified 68.27% intervals

The width $\mu_2 - \mu_1$ of the intervals decreases slightly for the runs $n = 7$ to 8, $n = 10$ to 11, $n = 13$ to 15, $n = 17$ to 19, etc. It may seem strange that the error interval for $n = 8$ is slightly smaller than the error interval for

$n = 7$, for example. This, however, should not be taken too seriously. The interpretation of the size of an interval as proportional to the precision of the measurement is known to suffer when considerations of coverage become dominant. Some formulations of frequentist confidence intervals can occasionally even give empty confidence intervals, or intervals that just contain a single point. These are not to be interpreted as representing infinitely precise measurements: the coverage (in general) at any particular value of μ is not completely determined by the size and location of the interval for a single n, but has significant contributions from several overlapping intervals.

Comparing values from the μ_1 and μ_2 columns of the table, one notices that every value of μ_2 from 2.7505 to 23.8247 also appears somewhere in the μ_1 column. For $C_0 = 68.27\%$, this synchronization continues indefinitely when the table is extended. As observed in a previous section, this means that there are many orphan coverage points that will not show up in the coverage plot.

It is interesting to examine the calculation of the unified intervals in slightly more detail. We show $U(\mu, n)$ for the case $n = 3$ here:



U(μ,3) vs μ: U(μ,3)<C_0= 0.6827 defines Unified n=3 1σ interval

As one would expect from its definition, $U(\mu, 3)$ as a function of μ has discontinuities at discrete points. As μ increases, the first discontinuity is lo-

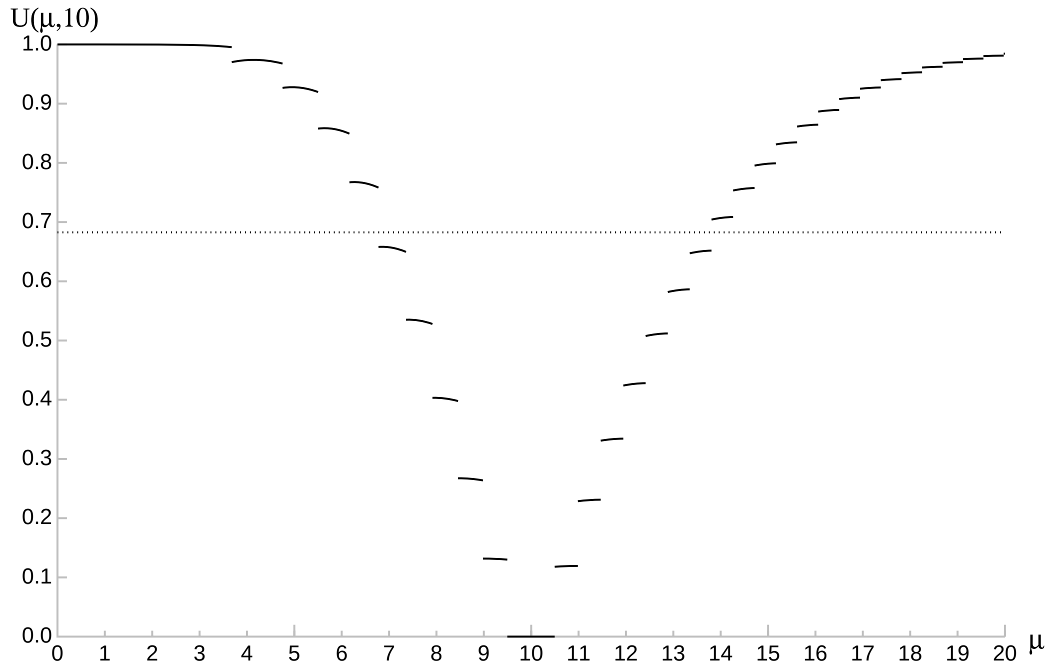
cated at $\mu = 3e^{-1} = 1.1036$, where $U(\mu, 3)$ drops below 68.27%, so we have $\mu_1 = 1.1036$. At $\mu = 5.3012$, $U(\mu, 3)$ jumps discontinuously from 0.6081 to 0.6852, so $\mu_2 = 5.3012$. Thus, the $n = 3$ interval's endpoints are defined by the location of the discontinuities in $U(\mu, 3)$.

Note that, if we tried to calculate the $n = 3$ interval for the case $C_0 =$ 57.5%, we would find that the equation $U(\mu, 3) = 0.575$ has two roots for $\mu <$ 3, leading to an error-interval with a gap (or hole) in its interior, according to the definition given above. Error intervals with holes are generally considered unacceptable. The possibility of obtaining intervals with holes via the unified approach is noted in [12], which therefore adds an additional clause to the definition: any holes in the interior of an interval are to be added to the interval, so that the final interval can always be described simply as $[\mu_1, \mu_2]$.

It turns out that the hole-filling step never needs to be performed for the 68.27% (i.e. 1σ) intervals, but in general, one does need to check for this possibility. Surprisingly, for 1σ intervals, it seems that the equation $U(\mu, n) = C_0$ has no roots at all for $n > 0$. We conjecture[6] that the endpoints of the 1σ intervals for $n > 0$ are always located at discontinuities in $U(\mu, n)$. This seems to be a property specific to the single point $C_0 = 0.682689492137\ldots$ that is not shared by other values for C_0 in that neighborhood.

For general C_0, what does happen is that the equation $U(\mu, n) = C_0$ cannot be satisfied for most (but not all) large n. This seems to be because the continuous segments of $U(\mu, n)$ flatten out as n becomes larger—illustrated here in the case $n = 10$:

[6]This conjecture is based on a numerical search covering $n = 1$ to $n = 10^7$. It would be nice to have a mathematical proof—or a counterexample.



U(μ,10) vs μ: U(μ,10)<C_0= 0.6827 defines Unified n=10 1σ interval

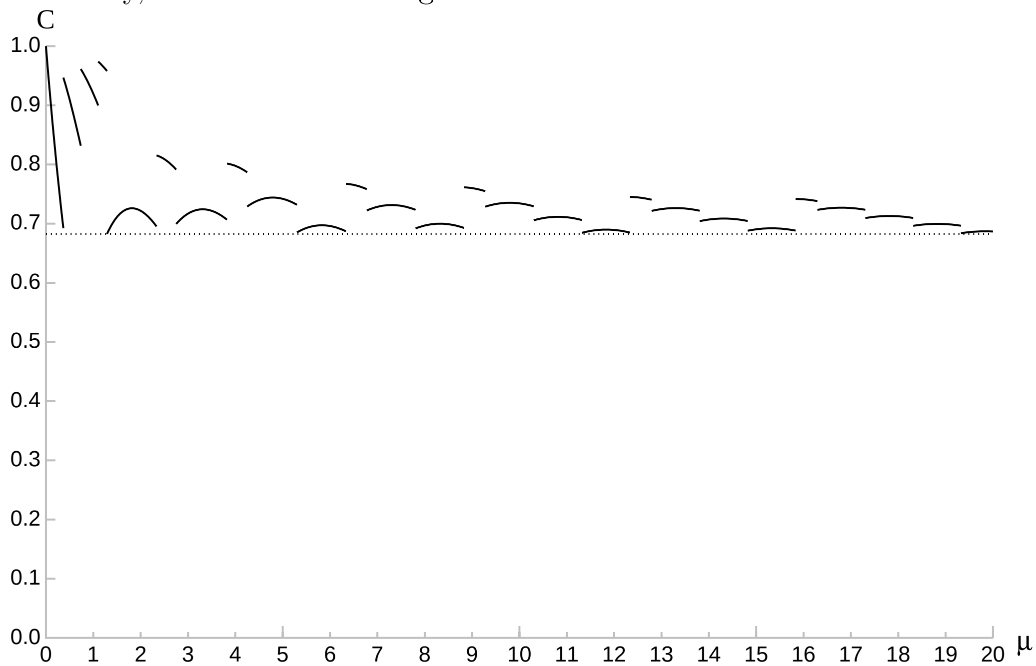
A random horizontal line drawn on this graph already has a relatively high probability of passing through a gap, rather than actually intersecting a segment.

The location of the discontinuities in $U(\mu, n)$ can be described analytically as follows: Solving $-2\ln\lambda(\mu, k) = -2\ln\lambda(\mu, n)$ for μ yields

$$\mu = n \exp\left(\frac{k \ln(n/k)}{n-k} - 1\right) \qquad (k \neq n)$$

Evaluating this expression for $k = 0, 1, 2, \ldots, n-1, n+1, n+2, \ldots$ therefore produces the location of the discontinuities of $U(\mu, n)$ in ascending order.

Finally, we show the coverage of the 1σ unified intervals:



Coverage (C) vs μ: Unified Intervals (C → 0.6827 as μ → ∞)

As is desired, $C(\mu) \geq 68.27\%$ for all μ. If the previously mentioned conjecture is correct, $C(\mu) = C_0$ at only the single point $\mu \simeq 1.2904$, which is μ_2 for the $n = 0$ interval. At all other values of μ, $C(\mu) > C_0$. Because of the "synchronization" of the μ_1 and μ_2 mention above, this coverage plot looks qualitatively different from the previously shown coverage plots.

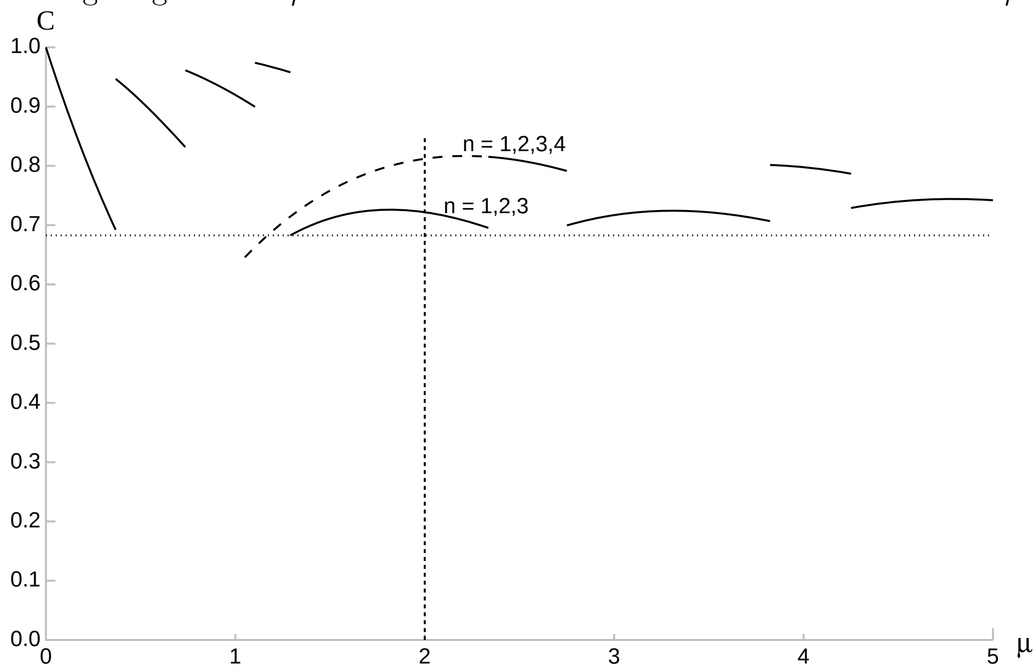
8 Interval Bias

There is an interesting concept that involves coverage considerations which can be introduced at this point: interval bias[7]. Quoting from reference [13]:

> Further, it seems highly desirable that a good confidence interval should cover a value of θ with higher probability when it is the true value than when it is not, so that the confidence coefficient will exceed the probability of covering any false value. Such a confidence interval is called *unbiased*—this use of the term is unconnected with estimation bias.

[7]The concept of interval bias is directly related to test bias.

We can illustrate this concept using any of the coverage plots shown so far—they all demonstrate interval bias. Coverage is usually calculated for a given μ, assuming the "data" is generated with that μ. To investigate interval bias, we assume the data is generated at μ_{gen}, and the coverage is calculated for a different value μ_{cov}. For example, here we expand the μ-axis of the unified coverage plot from page 19, and extend the original coverage segment at $\mu \simeq 2.5$ as a dashed line down to lower values of μ:



Coverage (C) vs μ: Unified Intervals (C → 0.6827 as μ → ∞)

As with the other coverage plots, the solid curves above give the coverage at the generated value (i.e. $\mu_{\text{gen}} = \mu_{\text{cov}}$). To determine the coverage for $\mu_{\text{cov}} = 2.5$ when the data is generated at $\mu_{\text{gen}} = 2.0$, the solid segment at $\mu = 2.5$ needs to be extended to the left, shown as dashed. Interval bias will be demonstrated by the fact that the dashed curve is higher than the solid curve at $\mu = 2$:

Suppose the true value of μ is 2.0. Then the coverage of that point under the unified scheme is

$$e^{-2}\left(2 + 2^2/2 + 2^3/6\right) = \frac{16}{3}e^{-2} \simeq 0.7218$$

because the μ intervals for $n = 1, 2, 3$ include $\mu = 2$ (see Table on page 15). However, given that the true value of μ is 2.0, the coverage of the false value

$\mu = 2.5$ (or any false value in that segment) is

$$e^{-2}\left(2 + 2^2/2 + 2^3/6 + 2^4/24\right) = 6e^{-2} \simeq 0.8120$$

because the μ intervals for $n = 1, 2, 3, 4$ include $\mu = 2.5$. That is, although the Poisson probabilities are calculated using the true value ($\mu = 2$) in both cases, the coverage probability of the false point $\mu = 2.5$ (or any false point in that segment) gains the $n = 4$ interval when compared to the coverage probability of the true value. Since this false value of the parameter μ is more likely to lie within the error bars than the true value, we have an example of interval bias.

It should be obvious both that interval bias is undesirable, and that is present in all the cases we have examined so far. The size of the bias is proportional to the size of the discontinuities in the $C(\mu)$ function, so we have another reason to want to keep the size of the "jumps" as small as possible. On the whole, because of the "synchronization" of the intervals, the unified approach does better than the others examined so far in keeping the size of the jumps small.

9 Pearson's χ^2 Ordering

In analogy with the unified approach, we next consider intervals constructed by imposing an ordering based on Pearson's χ^2 (instead of $-2\ln\lambda$). Explicitly, the error interval for n observed events is given by the set of all μ satisfying

$$U(\mu, n) = \sum_{k \in \mathcal{A}(\mu,n)} \frac{e^{-\mu}\mu^k}{k!} \;<\; C_0$$

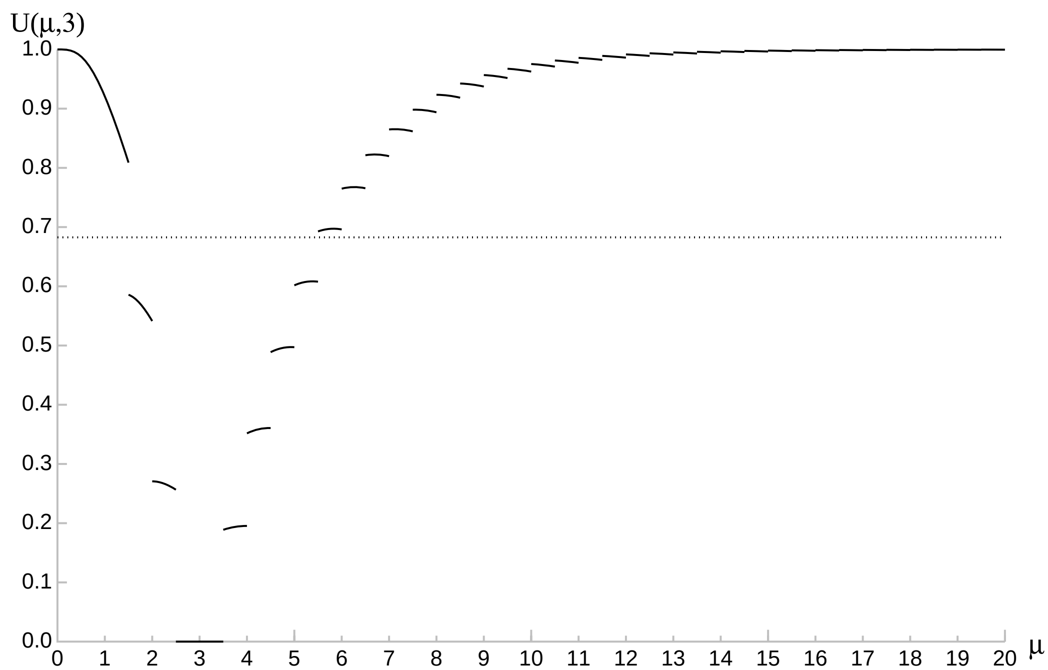
where now the set $\mathcal{A}(\mu, n)$ is defined as

$$\mathcal{A}(\mu, n) = \{\, k \mid\; k \in \mathbf{Z},\, k \geq 0, \text{ and } \tfrac{(k-\mu)^2}{\mu} < \tfrac{(n-\mu)^2}{\mu} \;\}$$

The locations of the discontinuities of $U(\mu, n)$, which occur at values of μ that satisfy $(k-\mu)^2/\mu = (n-\mu)^2/\mu$, are then simply given by

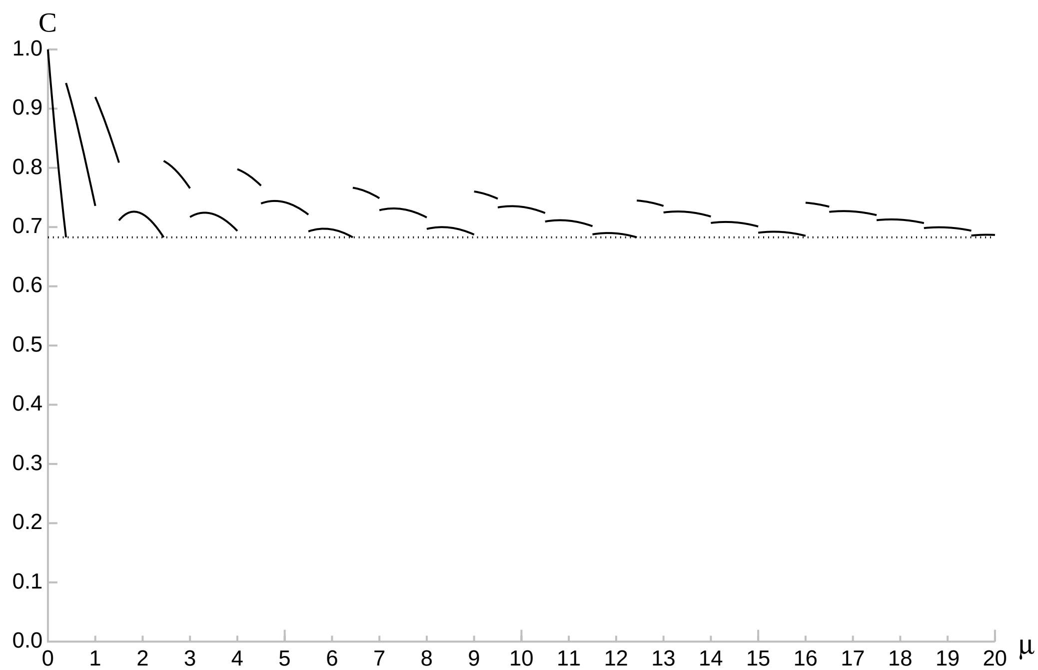
$$\mu = \frac{k+n}{2} \qquad\qquad (k \neq n)$$

for $k = 0, 1, 2, \ldots, n-1, n+1, n+2, \ldots$. The function $U(\mu, n)$ for the case $n = 3$ is shown here:



$U(\mu,3)$ vs μ: $U(\mu,3)<C_0= 0.6827$ defines χ^2 ordered n=3 1σ interval

The resulting coverage function $C(\mu)$ is shown here:



Coverage (C) vs μ: χ^2 Ordering Intervals (C → 0.6827 as μ → ∞)

It is qualitatively similar to the unified coverage plot on page 19.

10 Probability Ordering

Instead of $-2\ln\lambda$ or Pearson's χ^2, we can also order on the Poisson probability itself. In this case, the error interval for n observed events is given by the set of all μ satisfying

$$U(\mu, n) = \sum_{k \in \mathcal{A}(\mu,n)} \frac{e^{-\mu}\mu^k}{k!} \;<\; C_0$$

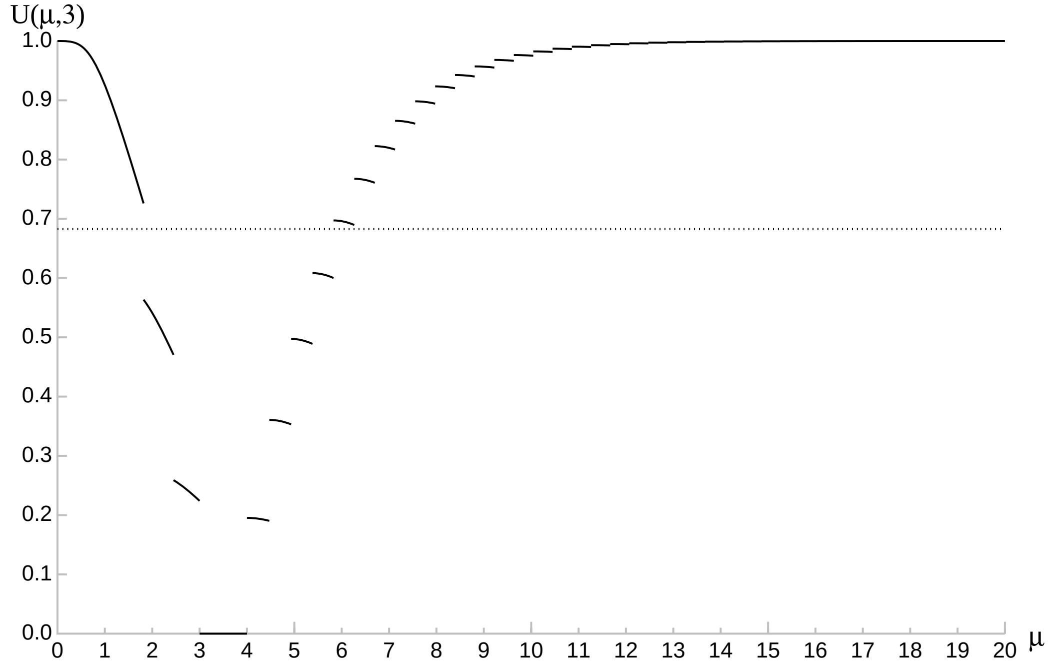
with the set $\mathcal{A}(\mu, n)$ defined as

$$\mathcal{A}(\mu, n) = \{\, k \mid\; k \in \mathbf{Z},\, k \geq 0, \text{ and } \tfrac{e^{-\mu}\mu^k}{k!} > \tfrac{e^{-\mu}\mu^n}{n!} \;\}$$

The locations of the discontinuities of $U(\mu, n)$ are then given by

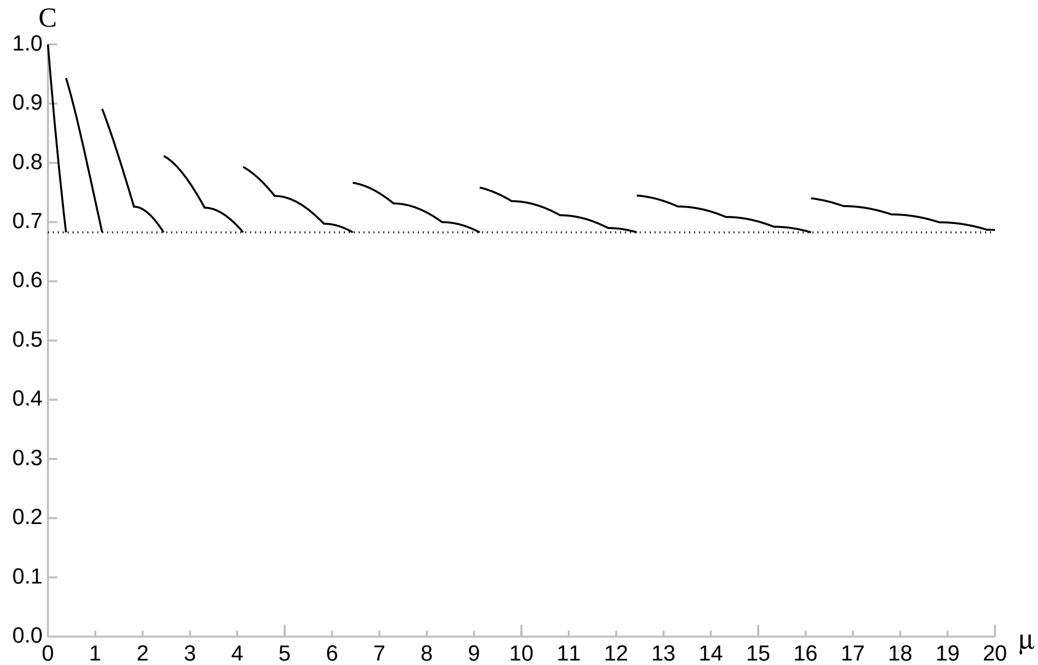
$$\mu = \left(\frac{n!}{k!}\right)^{\frac{1}{n-k}} \qquad\qquad (k \neq n)$$

for $k = 0, 1, 2, \ldots, n-1, n+1, n+2, \ldots$. The function $U(\mu, n)$ for the case $n = 3$ is shown here:



U(μ,3) vs μ: U(μ,3)<C_0= 0.6827 defines probability ordered n=3 1σ interval

The resulting coverage function $C(\mu)$ is shown here:



Coverage (C) vs μ: Probability Ordering Intervals (C → 0.6827 as μ → ∞)

Although quantitatively it is not so different from the unified coverage plot on page 19 or the Pearson's χ^2 ordering coverage plot on page 23, it seems at first glance quite different because the smaller discontinuities in $C(\mu)$ are "closed up", leaving only a discontinuity in the first derivative[8]. At these points of discontinuity in the derivative, $C'(\mu)$ actually changes from a negative value to exactly zero. This behavior represents an improvement, since it eliminates the interval bias that was present in the neighborhood of the small discontinuities, but the main discontinuities still remain about the same.

11 Summary and Conclusions

- The intervals (or error bars) based on the change in the value of Pearson's χ^2, Neyman's modified χ'^2, the likelihood ratio $(-2\ln\lambda)$, and the "improved likelihood ratio" all produce both overcoverage and undercoverage (at different values of μ). Of these four schemes, the Pearson's χ^2 approach seems to give the best results. However, (any) undercoverage is deemed unacceptable by many frequentists—all of these methods

[8]We ignore the orphan coverage points present at these locations.

fail that test.

- Classical central intervals, unified intervals, Pearson's χ^2 ordered intervals, and probability ordered intervals all effectively eliminate undercoverage. The overcoverage of the classical central intervals is clearly worse than in the other three cases. The coverage functions $C(\mu)$ of the other three are quite similar, with probability ordered intervals arguably giving slightly better properties. However, all these conclusions only apply to the specific case investigated here: Poisson data with no background.

- Overcoverage is undesirable, so one is justified in trying to minimize it. Physicists intuitively expect exact coverage for error bars, but unfortunately, exact coverage is not attainable through normal means for discrete distributions—the Poisson case being a prime example. Interval bias (a wrong value of μ covering at a higher rate than the correct value of μ), like non exact coverage, is both undesirable and unavoidable in the Poisson case (and in the general discrete case).

Appendix: Interval Tables

	Pearson's χ^2		Neyman's χ'^2		Likelihood		Improved L	
n	μ_1	μ_2	μ_1	μ_2	μ_1	μ_2	μ_1	μ_2
0	0.0000	1.0000	0.0000	0.0000	0.0000	0.5000	0.0000	0.6319
1	0.3820	2.6180	0.0000	2.0000	0.3017	2.3577	0.1457	2.4170
2	1.0000	4.0000	0.5858	3.4142	0.8976	3.7654	0.8212	3.8117
3	1.6972	5.3028	1.2679	4.7321	1.5840	5.0802	1.5252	5.1198
4	2.4384	6.5616	2.0000	6.0000	2.3185	6.3463	2.2691	6.3815
5	3.2087	7.7913	2.7639	7.2361	3.0841	7.5811	3.0408	7.6131
6	4.0000	9.0000	3.5505	8.4495	3.8719	8.7936	3.8329	8.8232
7	4.8074	10.1926	4.3542	9.6458	4.6765	9.9891	4.6408	10.0168
8	5.6277	11.3723	5.1716	10.8284	5.4946	11.1711	5.4615	11.1973
9	6.4586	12.5414	6.0000	12.0000	6.3237	12.3422	6.2926	12.3670
10	7.2984	13.7016	6.8377	13.1623	7.1619	13.5040	7.1326	13.5277
11	8.1459	14.8541	7.6834	14.3166	8.0080	14.6580	7.9802	14.6807
12	9.0000	16.0000	8.5359	15.4641	8.8609	15.8052	8.8344	15.8270
13	9.8599	17.1401	9.3944	16.6056	9.7198	16.9463	9.6944	16.9674
14	10.7251	18.2749	10.2583	17.7417	10.5840	18.0822	10.5596	18.1025
15	11.5949	19.4051	11.1270	18.8730	11.4529	19.2132	11.4295	19.2330
16	12.4689	20.5311	12.0000	20.0000	12.3262	20.3401	12.3035	20.3592
17	13.3467	21.6533	12.8769	21.1231	13.2033	21.4630	13.1814	21.4816
18	14.2280	22.7720	13.7574	22.2426	14.0839	22.5823	14.0627	22.6005
19	15.1125	23.8875	14.6411	23.3589	14.9679	23.6984	14.9472	23.7161
20	16.0000	25.0000	15.5279	24.4721	15.8548	24.8115	15.8347	24.8288
21	16.8902	26.1098	16.4174	25.5826	16.7445	25.9218	16.7250	25.9387
22	17.7830	27.2170	17.3096	26.6904	17.6368	27.0295	17.6178	27.0460
23	18.6782	28.3218	18.2042	27.7958	18.5315	28.1348	18.5129	28.1510
24	19.5756	29.4244	19.1010	28.8990	19.4285	29.2378	19.4103	29.2537
25	20.4751	30.5249	20.0000	30.0000	20.3276	30.3387	20.3098	30.3543
26	21.3765	31.6235	20.9010	31.0990	21.2287	31.4377	21.2113	31.4530
27	22.2798	32.7202	21.8038	32.1962	22.1317	32.5347	22.1146	32.5497
28	23.1849	33.8151	22.7085	33.2915	23.0364	33.6300	23.0197	33.6447
29	24.0917	34.9083	23.6148	34.3852	23.9429	34.7235	23.9264	34.7381
30	25.0000	36.0000	24.5228	35.4772	24.8509	35.8155	24.8347	35.8298
31	25.9098	37.0902	25.4322	36.5678	25.7605	36.9060	25.7446	36.9201
32	26.8211	38.1789	26.3431	37.6569	26.6715	37.9950	26.6558	38.0089
33	27.7337	39.2663	27.2554	38.7446	27.5838	39.0826	27.5685	39.0963
34	28.6477	40.3523	28.1690	39.8310	28.4975	40.1689	28.4824	40.1824
35	29.5628	41.4372	29.0839	40.9161	29.4125	41.2540	29.3976	41.2673
36	30.4792	42.5208	30.0000	42.0000	30.3286	42.3379	30.3139	42.3510
37	31.3967	43.6033	30.9172	43.0828	31.2459	43.4206	31.2314	43.4335
38	32.3153	44.6847	31.8356	44.1644	32.1643	44.5022	32.1500	44.5150
39	33.2350	45.7650	32.7550	45.2450	33.0838	45.5827	33.0697	45.5953
40	34.1557	46.8443	33.6754	46.3246	34.0043	46.6622	33.9904	46.6747
41	35.0774	47.9226	34.5969	47.4031	34.9258	47.7407	34.9121	47.7531
42	36.0000	49.0000	35.5193	48.4807	35.8482	48.8183	35.8347	48.8305
43	36.9235	50.0765	36.4426	49.5574	36.7716	49.8949	36.7582	49.9070
44	37.8479	51.1521	37.3668	50.6332	37.6958	50.9707	37.6826	50.9826
45	38.7732	52.2268	38.2918	51.7082	38.6209	52.0456	38.6079	52.0574
46	39.6993	53.3007	39.2177	52.7823	39.5468	53.1197	39.5339	53.1314
47	40.6261	54.3739	40.1443	53.8557	40.4735	54.1930	40.4608	54.2045
48	41.5538	55.4462	41.0718	54.9282	41.4010	55.2655	41.3884	55.2769
49	42.4822	56.5178	42.0000	56.0000	42.3293	56.3372	42.3168	56.3486
50	43.4113	57.5887	42.9289	57.0711	43.2583	57.4083	43.2459	57.4195

Lower and upper limits of "68.27%" intervals.

	Classical		Unified		χ^2 Order		Probability Order	
n	μ_1	μ_2	μ_1	μ_2	μ_1	μ_2	μ_1	μ_2
0	0.0000	1.8410	0.0000	1.2904	0.0000	1.5000	0.0000	1.8171
1	0.1728	3.2995	0.3679	2.7505	0.3817	3.0000	0.3817	3.3098
2	0.7082	4.6379	0.7358	4.2504	1.0000	4.5000	1.1447	4.7894
3	1.3673	5.9182	1.1036	5.3012	1.5000	5.5000	1.8171	5.8274
4	2.0857	7.1628	2.3359	6.7764	2.4438	7.0000	2.4438	7.2989
5	2.8403	8.3825	2.7505	7.8064	3.0000	8.0000	3.3098	8.3239
6	3.6201	9.5836	3.8231	9.2783	4.0000	9.5000	4.1226	9.7947
7	4.4185	10.7703	4.2504	10.3006	4.5000	10.5000	4.7894	10.8143
8	5.2316	11.9451	5.3012	11.3187	5.5000	11.5000	5.8274	11.8303
9	6.0565	13.1102	6.3342	12.7905	6.4382	13.0000	6.4382	13.3019
10	6.8913	14.2669	6.7764	13.8060	7.0000	14.0000	7.2989	14.3160
11	7.7344	15.4165	7.8064	14.8194	8.0000	15.0000	8.3239	15.3282
12	8.5847	16.5598	8.8291	16.2920	9.0000	16.5000	9.1183	16.8010
13	9.4413	17.6976	9.2783	17.3043	9.5000	17.5000	9.7947	17.8123
14	10.3035	18.8304	10.3006	18.3152	10.5000	18.5000	10.8143	18.8225
15	11.1706	19.9587	11.3187	19.3249	11.5000	19.5000	11.8303	19.8315
16	12.0422	21.0831	12.3338	20.7991	12.4363	21.0000	12.4363	21.3059
17	12.9178	22.2037	12.7905	21.8084	13.0000	22.0000	13.3019	22.3147
18	13.7971	23.3210	13.8060	22.8169	14.0000	23.0000	14.3160	23.3228
19	14.6798	24.4352	14.8194	23.8247	15.0000	24.0000	15.3282	24.3301
20	15.5656	25.5465	15.8310	25.3003	16.0000	25.5000	16.1169	25.8058
21	16.4542	26.6552	16.2920	26.3079	16.5000	26.5000	16.8010	26.8132
22	17.3455	27.7614	17.3043	27.3150	17.5000	27.5000	17.8123	27.8199
23	18.2393	28.8652	18.3152	28.3216	18.5000	28.5000	18.8225	28.8263
24	19.1354	29.9669	19.3249	29.3278	19.5000	29.5000	19.8315	29.8321
25	20.0337	31.0666	20.3336	30.8049	20.4355	31.0000	20.4355	31.3094
26	20.9340	32.1643	20.7991	31.8110	21.0000	32.0000	21.3059	32.3153
27	21.8362	33.2602	21.8084	32.8169	22.0000	33.0000	22.3147	33.3209
28	22.7403	34.3544	22.8169	33.8223	23.0000	34.0000	23.3228	34.3262
29	23.6461	35.4470	23.8247	34.8275	24.0000	35.0000	24.3301	35.3311
30	24.5535	36.5380	24.8319	36.3057	25.0000	36.5000	25.1162	36.8095
31	25.4624	37.6276	25.3003	37.3110	25.5000	37.5000	25.8058	37.8146
32	26.3729	38.7158	26.3079	38.3160	26.5000	38.5000	26.8132	38.8194
33	27.2847	39.8026	27.3150	39.3207	27.5000	39.5000	27.8199	39.8240
34	28.1979	40.8881	28.3216	40.3251	28.5000	40.5000	28.8263	40.8283
35	29.1123	41.9724	29.3278	41.3294	29.5000	41.5000	29.8321	41.8325
36	30.0280	43.0555	30.3335	42.8090	30.4351	43.0000	30.4351	43.3122
37	30.9449	44.1376	30.8049	43.8134	31.0000	44.0000	31.3094	44.3164
38	31.8628	45.2185	31.8110	44.8176	32.0000	45.0000	32.3153	45.3205
39	32.7819	46.2984	32.8169	45.8216	33.0000	46.0000	33.3209	46.3244
40	33.7020	47.3773	33.8223	46.8254	34.0000	47.0000	34.3262	47.3281
41	34.6231	48.4552	34.8275	47.8291	35.0000	48.0000	35.3311	48.3317
42	35.5452	49.5322	35.8323	49.3097	36.0000	49.5000	36.1159	49.8124
43	36.4682	50.6083	36.3057	50.3135	36.5000	50.5000	36.8095	50.8161
44	37.3921	51.6835	37.3110	51.3171	37.5000	51.5000	37.8146	51.8197
45	38.3168	52.7579	38.3160	52.3206	38.5000	52.5000	38.8194	52.8231
46	39.2424	53.8315	39.3207	53.3240	39.5000	53.5000	39.8240	53.8264
47	40.1688	54.9043	40.3251	54.3273	40.5000	54.5000	40.8283	54.8296
48	41.0960	55.9763	41.3294	55.3304	41.5000	55.5000	41.8325	55.8327
49	42.0240	57.0476	42.3335	56.8121	42.4349	57.0000	42.4349	57.3144
50	42.9527	58.1182	42.8090	57.8153	43.0000	58.0000	43.3122	58.3176

Lower and upper limits of “68.27%” intervals.

References

[1] W.T. Eadie, D. Drijard, F.E. James, M. Roos, and B. Sadoulet "Statistical Methods in Experimental Physics", (North-Holland Publishing Co, Amsterdam, 1971), §11.2, p 257.

[2] Glen Cowan, "Statistical Data Analysis", (Oxford University Press, Oxford, 1998), §4.7 p 61.

[3] F. James, "Minuit, Function Minimization and Error Analysis", CERN long writeup D506, `wwwinfo.cern.ch/asdoc/minuit/minmain.html`

[4] Paul Harrison, "Blind Analyses" in Proceedings of the Conference on Advanced Techniques in Particle Physics, M. Whalley and L. Lyons (ed.), IPPP/02/39 (July 2002), page 278. `www.ippp.dur.ac.uk/Workshops/02/statistics/proceedings/harrison.ps`

[5] Harold Jeffreys, "Theory of Probability", 3rd ed., (Oxford University Press, Oxford, 1961), §4.1, p 197; W.T. Eadie, *et al.*, *ibid.*, §8.4.5, p 171; Glen Cowan, *ibid.*, §7.4 p 101.

[6] K. Hagiwara et al., Physical Review D **66**, 010001-229 (2002), Statistics Review, revised October 2001 by G. Cowan, `pdg.lbl.gov/2002/contents_sports.html#mathtoolsetc`

[7] J.G. Heinrich, "The Log Likelihood Ratio of the Poisson Distribution for Small μ", CDF note 5718, version 2, `www-cdf.fnal.gov/physics/statistics/notes/cdf5718_loglikeratv2.ps`

[8] W.T. Eadie, et al., *ibid.*, §10.5.5, p 236;

[9] Alan Stuart, J. Keith Ord, and Steven Arnold, "Kendall's Advanced Theory of Statistics", Volume 2A 6th ed., (Oxford University Press, Oxford, 1999), §22.9, p 249.

[10] Morten Frydenberg and Jens Ledet Jensen, "Is the 'improved likelihood ratio statistic' really improved in the discrete case?", Biometrika **76** p 655 (1989).

[11] Glen Cowan, *ibid.*, §9.4 p 126.

[12] G.J. Feldman and R.D. Cousins, "Unified approach to the classical statistical analysis of small signals", Physical Review D **57**, 3873 (1998).
`link.aps.org/abstract/PRD/v57/p3873`
`www.hepl.harvard.edu/~feldman/Unified_Approach.ps`

[13] Alan Stuart et al., *ibid.*, §19.14 p 128.