Confidence intervals for a bounded mean in exponential families

Hsiuying Wang[a]

[a] Institute of Statistics, National Chiao Tung University, Hsinchu,
Taiwan
Published online: 11 Sep 2012.

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis
Taylor & Francis Group

Confidence intervals for a bounded mean in exponential families

Hsiuying Wang*

Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan

Setting confidence bounds or intervals for a parameter in a restricted parameter space is an important issue in applications and is widely discussed in the recent literature. In this article, we focus on the distributions in the exponential families, and propose general forms of the truncated Pratt interval and rp interval for the means. We take the Poisson distribution as an example to illustrate the method and compare it with the other existing intervals. Besides possessing the merits from the theoretical inferences, the proposed intervals are also shown to be competitive approaches from simulation and real-data application studies.

Keywords: confidence interval; coverage probability; exponential family; Poisson distribution; minimax; rp interval

1. Introduction

Setting confidence bounds or intervals for a restricted parameter space is widely discussed in recent literature, see Lyons [1], Fraser *et al.* [2], Mandelkern [3], Roe and Woodroofe [4], Woodroofe and Wang [5] and Zhang and Woodroofe [6]. Lyons [1] and Mandelkern [3] pointed out the importance of statistical inference for bounded parameter spaces in physics, and gave examples that the classical Neyman procedure yields unsatisfactory results when the parameter is known to be bounded.

It is worth noting that this problem is particularly important for discrete distributions: the Poisson distribution for physics analyses [1,3,5], the binomial distribution for control charts [7] and the binomial and Poisson distributions for tolerance intervals [8,9]. The confidence interval is directly related to control chart construction, and also to tolerance interval construction because a conventional approach to derive tolerance intervals for discrete distributions is based on the confidence intervals or bounds [10].

Besides the Neyman procedure, there are several traditional methods of constructing the confidence intervals. Most of the approaches focus on the natural parameter space, but not restricted parameter spaces. Although for a confidence interval, we can simply construct the confidence interval for a restricted parameter by considering the intersection of the confidence interval and the restricted parameter space, it may be lack of the advantage of sufficiently using the restriction information because it was not considered to be constructed based on the restriction originally.

---

*Email: wang@stat.nctu.edu.tw

In this article, we focus on the distributions in the exponential families, and propose general forms of the truncated Pratt interval and rp interval for the mean when it is known to be bounded. We consider the one-parameter exponential families. Assume that X follows a one-parameter exponential family with a density or frequency function

$$f(x, \eta) = h(x) \exp\{\eta x - A(\eta)\}, \quad x \in R, \tag{1}$$

where $A(\eta) = \log \int h(x) \exp\{\eta x\} \mathrm{d}x$. The families include the normal, binomial and Poisson distributions. Note that the result in this article can be generalized to the exponential family with a density or frequency function of the form

$$f(x, \eta) = h(x) \exp\{\eta T(x) - A(\eta)\}, \tag{2}$$

where $T(x)$ is a function of x. The mean of the distribution is a function of η. The expected length of a confidence interval is a function of the parameter.

Pratt [11] constructed a minimax expected length interval for the parameter at a fixed point. A minimax expected length interval is a confidence interval with the smallest maximal expected length. Using this fact, Evans *et al.* [12] based on Pratt [11] to construct the 'truncated Pratt interval', which is shown to be a minimax expected interval for the location family when the parameter space is restricted to some region. In this article, we extend the result to the exponential families.

The construction of the rp interval is based on a modified *p*-value which is called a rp-value for hypothesis testing when the parameter space is considered to be restricted. The rp-value is proposed in Wang [13,14] for the normal distribution. In this article, we extend the rp-value to the exponential families and propose a confidence interval of the mean based on the rp-value when the parameter space is known to be restricted.

We take the Poisson distribution as an example and conduct a simulation study to compare the performance of the proposed intervals with the existing methods such as the Wald, score, likelihood ratio (LR) and Jeffreys intervals, in terms of their coverage probabilities and expected lengths. The study shows that the proposed methods have better performance than the existing methods.

The article is organized as follows. The truncated Pratt interval is constructed in Section 2. Section 3 gives the motivation of constructing the rp interval and introduces a terminology to evaluate the test derived from the rp interval. Section 4 describes four existing Poisson confidence intervals in the restricted parameter space, and the explicit forms of the intervals are given. Simulation studies for comparing the six intervals are presented in Section 5. The performances of these intervals are based on their coverage probabilities and expected lengths. Section 6 provides a real-data example to illustrate the method and compare them with the existing methods. A conclusion about the performances of the intervals is given in Section 7.

2. The truncated Pratt interval

Let X be a random variable with the density or frequency function (1), which is the form of a canonical exponential family. By Bickel and Doksum [15], the mean θ of X is $\dot{A}(\eta) = \partial A(\eta)/\partial \eta$.

We will consider the confidence interval for $\theta = \dot{A}(\eta)$. We assume that $\dot{A}(\eta)$ is an strictly increasing function of η. It is reasonable to make this assumption because it can apply to the normal, Poisson and binomial distributions. Note that since θ is a strictly increasing function of η, the subset of $\{\theta : \theta \leq \theta_0\}$ is equivalent to $\{\eta : \eta \leq \eta_0\}$, where θ_0 and η_0 are some constants. Since the rejection region of a most power test for testing $H_0 : \eta \leq \eta_0$ is

$$\{x : x \geq c\}, \tag{3}$$

the rejection region of a most power test for testing $H_0 : \theta \leq \theta_0$ is also (3), where c is a critical value depending on the significant level of the test.

An approach of constructing the minimax interval is referred to Pratt [11], which constructed a minimax expected length interval by inverting from a most powerful test. With an observation x, for the hypothesis $H_0 : \theta \neq \theta_0$ versus $H_1 : \theta = \theta_0$, a most powerful test with type one error α is to accept θ for

$$
\begin{aligned}
\theta \geq \theta_x^1 \quad &\text{if } \theta < \theta_0; \\
\theta \leq \theta_x^2 \quad &\text{if } \theta > \theta_0,
\end{aligned}
\tag{4}
$$

where

$$
\theta_x^1 = \max\{\theta : P_\theta(X \geq x) \leq \alpha\} = \{\theta : P_\theta(X \geq x) = \alpha\}
$$

and

$$
\theta_x^2 = \min\{\theta : P_\theta(X \leq x) \leq \alpha\} = \{\theta : P_\theta(X \leq x) = \alpha\}.
\tag{5}
$$

At $\theta = \theta_0$, any test may be used. For convenience, let all x be in the acceptance region for the null hypothesis $\theta = \theta_0$, see Figure 1. Also from Figure 1, the $1 - \alpha$ confidence interval derived from Equation (4) is

$$
\text{CI} = \begin{cases}
(\theta_0, \theta_x^2) & \text{if } \theta_0 \leq \theta_x^1, \\
(\theta_x^1, \theta_x^2) & \text{if } \theta_x^1 \leq \theta_0 \leq \theta_x^2, \\
(\theta_x^1, \theta_0) & \text{if } \theta_x^2 \leq \theta_0.
\end{cases}
\tag{6}
$$

This confidence interval can be rewritten as $\min\{\theta_0, \theta_x^1\} \leq \theta \leq \max\{\theta_0, \theta_x^2\}$. From Figure 1, the interval may be arbitrarily long or have finite length. The details are referred to Pratt [11].

By Pratt [11], the interval (6) minimizes the expected length when the mean is θ_0. Based on the Pratt interval, for estimating bounded location parameters of one-dimensional shift families, Evans *et al.* [12] constructed a minimax expected length interval for a restricted parameter space by showing that the expected length is a unimodal function of the parameter which is known to be bounded. Here, minimax means minimizing the maximum expected lengths.

In this section, we will extend the result to the exponential families with θ in a restricted parameter space. Without loss of generality, we consider the case of sample size 1 in the following
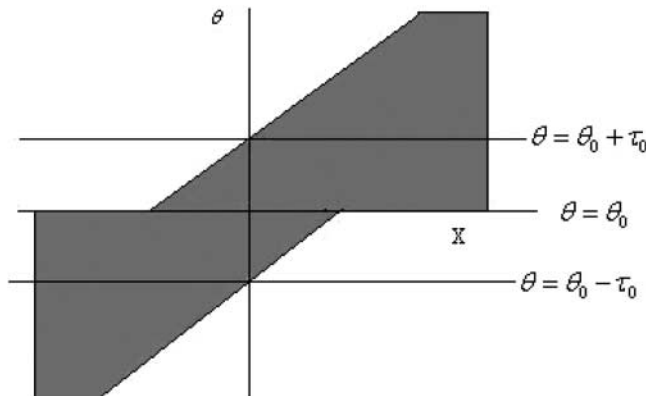

Figure 1. The confidence procedure of Equation (6) for a symmetric distribution.

description. Let

$$d^{\theta_0}(\theta, x) = \begin{cases} 1_{\theta \geq \theta_x^1} & \theta < \theta_0, \\ 1_{\theta \leq \theta_x^2} & \theta > \theta_0, \end{cases} \tag{7}$$

which denotes the indicator function of θ with values 1 when θ is accepted by the test (4).

Under a restricted parameter space Θ, the $1 - \alpha$ confidence interval based on Equation (6) is modified to

$$\text{CI} = \begin{cases} (\theta_0, \theta_x^2) \cap \Theta & \text{if } \theta_0 \leq \theta_x^1, \\ (\theta_x^1, \theta_x^2) \cap \Theta & \text{if } \theta_x^1 \leq \theta_0 \leq \theta_x^2, \\ (\theta_x^1, \theta_0) \cap \Theta & \text{if } \theta_x^2 \leq \theta_0, \end{cases} \tag{8}$$

where θ_0 is the middle point of the space Θ. The interval is called the truncated Pratt interval.

The expected length of the confidence set (8) is

$$L_\theta(d^{\theta_0}) = E_\theta \int_a^b d^{\theta_0}(\zeta, X) \mathrm{d}\zeta.$$

We will show that the interval (8) is a minimax expected length interval for $\theta \in (\theta_0 - \tau, \theta_0 + \tau) \subset (a, b)$ for some τ. The condition of τ is given in the proof of Theorem 1.

THEOREM 1 *Let X be a random variable with density function (1) and $\Theta = [\theta_0 - \tau, \theta_0 + \tau]$ be the parameter space of $\theta = \dot{A}(\eta)$. If (1) is a symmetric function, there exists a constant h such that (8) is the minimax expected length interval for $\theta \in \Theta$, where $\tau < h$. If (1) is not a symmetric function and let $X_1, \ldots, X_n$ be a sample from the density function (1), when the sample size n is large enough, the confidence interval (8) based on the sample mean is the minimax expected length interval for $\theta \in \Theta$.*

Note that h cannot be derived analytically, but it can be derived by numerical calculation. The discussion of choosing h is given in Section 5. Theorem 1 shows that the truncated Pratt interval is the minimax expected length interval for the restricted parameter space $[\theta_0 - \tau, \theta_0 + \tau], \tau < h$. When the parameter space is wider, the truncated Pratt interval may not be the minimax expected length interval. Schafer and Stark [16] has provided free, open-source software with which to approximate minimax expected length confidence interval in the case that the parameter space is too wide for the truncated Pratt interval. The software is also available from http://www.stat.cmu.edu/~cschafer/.

3. The rp interval

In this section, it is assumed that the bounded parameter space is (a, b). We propose the rp interval based on a different criterion from Section 2, which is given by

$$\text{CI}_{\text{rp}} = (\theta_l^{\text{rp}}, \theta_u^{\text{rp}}) \cap (a, b), \tag{9}$$

where

$$\theta_l^{\text{rp}} = \min \left\{ \theta : P_\theta(\bar{X} \geq \bar{x}) \geq \left(1 - \frac{\alpha}{2}\right) P_b(\bar{X} \geq \bar{x}) + \frac{\alpha}{2} P_a(\bar{X} \geq \bar{x}) \right\}$$

and

$$\theta_u^{\text{rp}} = \max \left\{ \theta : P_\theta(\bar{X} \leq \bar{x}) \geq \left(1 - \frac{\alpha}{2}\right) P_a(\bar{X} \leq \bar{x}) + \frac{\alpha}{2} P_b(\bar{X} \leq \bar{x}) \right\}. \tag{10}$$

Without loss of generality, we consider the case of sample size 1 in the following description.

This interval is constructed by converting a modified *p*-value for hypothesis testing in a restricted parameter space. The modified *p*-value is introduced as follows. For the normal distribution, it is referred to Wang [13]. When the sample size is one, for testing $H_0 : \theta < \theta_0$ versus $H_1 : \theta \geq \theta_0$, and $\theta \in (a, b)$,

$$r_{\theta_0}(x) = \frac{P_{\theta_0}(X \geq x) - \min_{a < \theta < b} P_\theta(X \geq x)}{\max_{a < \theta < b} P_\theta(X \geq x) - \min_{a < \theta < b} P_\theta(X \geq x)}, \tag{11}$$

which is called a rp-value, is a modified measure of evidence against the null hypothesis instead of the usual *p*-value $P_{\theta_0}(X \geq x)$. The 'rp' means the *p*-value for restricted parameter spaces. The rp interval for the normal distribution has been discussed in Wang [17]. The motivation for proposing Equation (11) is as follows.

The range of $P_{\theta_0}(X \geq x)$ for $a \leq \theta_0 \leq b$ is $[\min_{\theta_0 \in (a,b)} P_{\theta_0}(X \geq x), \max_{\theta_0 \in (a,b)} P_{\theta_0}(X \geq x)]$. Since no matter what value of $\theta_0 \in (a, b)$ is chosen to be the null hypothesis $\theta = \theta_0$, the *p*-value $p_{\theta_0}(X \geq x)$ is always greater than $\min_{\theta_0 \in (a,b)} P_{\theta_0}(X \geq x)$. Thus, the magnitude $\min_{\theta_0 \in (a,b)} P_{\theta_0}(X \geq x)$ should not be included in a measure of evidence against the null hypothesis. Hence, first we suggest that a reasonable measure of evidence should be the usual *p*-value minus this magnitude, which is

$$p_{\theta_0}(x) - \min_{\theta_0 \in (a,b)} P_{\theta_0}(X \geq x). \tag{12}$$

Moreover, usually we would decide to reject the null hypothesis or accept the null hypothesis by comparing the *p*-value with a value α, where $0 < \alpha < 1$. If we can transform Equation (12) such that the range is $(0, 1)$, then it is more reasonable to compare it with a value α between 0 and 1. Thus, for fixed *x*, we divide (12) by $\max_{\theta_0 \in (a,b)} P_{\theta_0}(X \geq x) - \min_{\theta_0 \in (a,b)} P_{\theta_0}(X \geq x)$ such that its range is $(0, 1)$ which leads to Equation (11). Note that $r_{\theta_0}(x)$ is a $x - dependent$ transformation of $p_{\theta_0}(x)$.

Under this transformation, the range of Equation (11) is $(0, 1)$. Note that when the parameter space is the natural parameter space, Equation (11) is exactly equal to the usual *p*-value because $\max_{\theta_0 \in (a,b)} P_{\theta_0}(X \geq x) = 1$ and $\min_{\theta_0 \in (a,b)} P_{\theta_0}(X \geq x) = 0$.

The $(1 - \alpha/2)$ lower bound constructed by the rp-value is

$$\theta_l^{\rm rp} = \min \left\{ s : \frac{P_s(X \geq x) - \min_{a < \theta < b} P_\theta(X \geq x)}{\max_{a < \theta < b} P_\theta(X \geq x) - \min_{a < \theta < b} P_\theta(X \geq x)} > \frac{\alpha}{2} \right\},$$

which is equal to

$$\min \left\{ s : P_s(X \geq x) > \frac{\alpha}{2} P_b(X \geq x) + \left(1 - \frac{\alpha}{2} \right) P_a(X \geq x) \right\}$$

because $\max_{a < \theta < b} P_\theta(X \geq x) = P_b(X \geq x)$ and $\min_{a < \theta < b} P_\theta(X \geq x) = P_a(X \geq x)$.

The $(1 - \alpha/2)$ upper bound constructed by the rp-value is

$$\theta_u^{\rm rp} = \max \left\{ s : \frac{P_s(X \leq x) - \min_{a < \theta < b} P_\theta(X \leq x)}{\max_{a < \theta < b} P_\theta(X \leq x) - \min_{a < \theta < b} P_\theta(X \leq x)} > \frac{\alpha}{2} \right\},$$

which is equal to

$$\max \left\{ s : P_s(X \leq x) > \frac{\alpha}{2} P_a(X \leq x) + \left(1 - \frac{\alpha}{2} \right) P_b(X \leq x) \right\}$$

because $\max_{a < \theta < b} P_\theta(X \leq x) = P_a(X \leq x)$ and $\min_{a < \theta < b} P_\theta(X \leq x) = P_b(X \leq x)$.

Combining the $(1 - \alpha/2)$ upper and lower bounds for θ, Equation (9) is the $(1 - \alpha)$ level rp confidence interval. Wang [17] shows that the rp interval is shown to have higher coverage and shorter expected length than several other intervals for the normal distribution.

The rp-value is evaluated from a testing point of view. We show that the test derived from rp-value is a uniformly most powerful (UMP) test as follows.

A test derived from $r_\theta(x)$ is of the form

$$\text{reject } H_0 \quad \text{if } r_\theta(x) < k$$
$$\text{do not reject } H_0 \quad \text{otherwise,}$$
(13)

where k, between 0 and 1, is a positive constant such that $\max_{\theta \in \Theta_0} P(r(x) \leq k) = \alpha$. Consequently, by Equation (13), the one-sided testing derived from the usual p-value of Neyman–Pearson test is a UMP test. In addition, the test derived from the modified p-value $r_\theta(x)$ with respect to prior $\pi(\theta) = I_{(\Theta_0 \cup \Theta_1)}(\theta)$ is also a UMP test, as demonstrated in Theorem 2.

THEOREM 2   *Let X be a random variable with density or probability function (1) with mean θ. Assume that the parameter space for θ is (a, b). For testing $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta \geq \theta_0$, a level α test derived from the* rp-*value, $r_{\theta_0}(x)$, is*

$$\phi(x) = \begin{cases} 1 & \text{when } r_{\theta_0}(x) < k, \\ e & \text{when } r_{\theta_0}(x) = k, \\ 0 & \text{when } r_{\theta_0}(x) > k, \end{cases}$$

where k and e are determined by $\max_{a \leq v(\theta) \leq v_0} E_\theta \phi(x) = \alpha$. The rejection region of the test is $C_x = \{x : \phi(x) = 1\}$. Then, $\phi(x)$ is a UMP test.

4.   Exact minimum coverage probability for discrete distributions

Using the procedure in Wang [18], we can calculate the minimum and average coverage probabilities of the minimax and rp confidence intervals for discrete distributions if the intervals satisfy that lower bounds and upper bounds of the intervals are increasing functions of the observation and any point in the parameter space is covered in the intervals [18]. These two properties are called the monotone boundary property and the full coverage property, respectively, in Wang [19]. A brief version of the procedure adjusted to fit the restricted parameter space is presented below.

Procedure [18]: Computing the exact minimum coverage probability of a confidence interval $(L(X), U(X))$ for discrete distributions.

Step 1. For a confidence interval, check if the full coverage property is satisfied and the monotone boundary property is satisfied. If the full coverage property is not satisfied, the confidence coefficient is zero, and we do not need to go to step 2.

Step 2. If the monotone boundary property and the full coverage property are satisfied, list the endpoints belonging to the restricted parameter space.

Step 3. Calculate the coverage probabilities at each endpoint of step 2 and at the lower boundary point and the upper boundary point of the parameter space. Then, calculate the minimum value of these coverage probabilities, which is the exact minimum coverage probability of this confidence interval when the parameter is in the restricted parameter space.

To apply this procedure to calculate the minimum coverage probabilities of the minimax expected length and rp intervals, we first need to check the monotone boundary property and the full coverage property. In Example 4.1, the rp intervals and the truncated Pratt intervals for a Poisson mean are presented. It is shown that the two properties hold for the truncated Pratt intervals and the monotone boundary property holds for the rp interval based on the observations from $0, \ldots, 10$.

Example 4.1 A random variable X is assumed to follow a Poisson distribution with parameter λ and a restricted parameter space $(2,4)$ is considered. The range of a Poisson observation is from zero to infinity. Here, we only consider confidence intervals for the observations $X = 0, \ldots, 10$ because the probability of the event $X > 10$ is small for $\lambda \in (2,4)$. The level 0.95 truncated Pratt intervals corresponding to observations $X = 0, \ldots, 10$ are $(2,3)$, $(2,3)$, $(2,3)$, $(2,3)$, $(2,4)$, $(2.613,4)$, $(3,4)$, $(3,4)$, $(3,4)$, $(3,4)$ and $(3,4)$. The level 0.95 rp intervals corresponding to observations $X = 0, \ldots, 10$ are $(2,3.852)$, $(2.092,3.897)$, $(2.029,3.927)$, $(2.041,3.946)$, $(2.058,3.959)$, $(2.085,3.968)$, $(2.125,3.974)$, $(2.184,3.979)$, $(2.263,3.982)$, $(2.361,3.984)$ and $(2.467,3.987)$.

For the observations $X = 0, \ldots, 10$, the truncated Pratt and rp interval have the monotone boundary property. In addition, the rp intervals have the full coverage property for $\lambda \in (2,4)$. However, the rp intervals only have the full coverage property for $\lambda \in (2,3.987)$ based on these observations. We can adopt the above procedure to calculate the approximate minimum coverage probabilities of the truncated Pratt intervals for $\lambda \in (2,4)$ and the rp intervals for $\lambda \in (2,3.987)$.

Beside calculating the minimum coverage probabilities, it may be more objective to consider the overall performance among the restricted parameter space instead of only considering the minimum coverage probability performance. A procedure to calculate the average coverage probability is also provided in Wang [18], which only requires the assumption of the monotone boundary property of a confidence interval. For the Poisson distribution case, adopting Wang's procedure, the average coverage probabilities of the rp and truncated Pratt intervals can be derived.

5. Example of the Poisson distribution

We take the Poisson distribution as an example to compare the proposed methods with the existing methods. Poisson distribution is of particular interest for many applications. An important application is the signal plus noise model in a physics application:

$$X = O + S,$$

where O and S, representing respectively the background and signal, are Poisson random variables with means o and μ. Here, μ is unknown, but o is known. Let $\lambda = o + \mu$. Then, $X \sim \text{Poi}(\lambda)$ with the restriction $\lambda \geq o$. Inference problems concerning this signal plus noise model have been considered, for example, in Feldman and Cousins [20], Roe and Woodroofe [21], Woodroofe and Wang [5] and Fraser *et al.* [2].

Woodroofe and Wang [5] considered one-sided testing for the problem and gave a modified *p*-value for testing $H_0 : \mu = \mu_0$ against $H_1 : \mu > \mu_0$. The modified *p*-value is shown to be admissible under a criterion given in Hwang *et al.* [22], and the usual *p*-value derived from the UMP test is shown to be inadmissible under the same criterion. The modified *p*-value is constructed by using conditional probability given that O is at most the observed value of X.

For this problem, Fraser *et al.* [2] argue that a fixed level confidence approach provides a much more limited statement than the parameter is or is not contained in a given interval. They suggest an improvement to the confidence approach that would be the reporting of confidence limits at a continuum of confidence levels, which is mathematically close to the *p*-value function approach. The proposed interval based on the rp-value can be seen to coincide with this viewpoint.

5.1. *Existing confidence intervals*

Four existing confidence intervals for the Poisson mean are introduced as follows [23]. We consider the forms restricted to the parameter space (a, b). Here, we consider the general case with sample size n instead of only one observation. Assume that $X_1, \ldots, X_n$ is a sample following a Poisson distribution Poi(λ), $\lambda \in (a, b)$. Let $Y = \sum X_i$ and $\bar{X} = \sum X_i / n$.

(1) The Wald interval. The $1 - \alpha$ Wald interval has the form

$$\mathrm{CI}_s = \left(\bar{X} - z_{\alpha/2} \sqrt{\frac{\bar{X}}{n}}, \bar{X} - z_{\alpha/2} \sqrt{\frac{\bar{X}}{n}} \right) \cap (a, b).$$

(2) The score interval. Denote $\tau(\bar{X}) = (\bar{X} + z_{\alpha/2}^2/2)/n$. The $1 - \alpha$ score interval is defined as

$$\mathrm{CI}_R(\bar{X}) = (\tau(\bar{X}) - z_{\alpha/2} W(\bar{X}), \tau(\bar{X}) + z_{\alpha/2} W(\bar{X})) \cap (a, b),$$

where

$$W(\bar{X}) = \sqrt{\frac{\bar{X}}{n} + \frac{z_{\alpha/2}^2}{(4n^2)}}$$

(3) The LR interval. The $1 - \alpha$ LR interval is

$$\mathrm{CI}_{\theta_n}(\bar{X}) = \left\{ \theta : -2 \log \frac{\mathrm{e}^{-n\theta} \theta^Y}{\mathrm{e}^{-Y} \bar{X}^Y} < \chi_{1,\alpha}^2 \right\} \cap (a, b),$$

where $\chi_{1,\alpha}^2$ is the upper α cutoff point of chi-square distribution with degree of freedom 1.
(4) The equal-tailed Jeffreys interval. The $1 - \alpha$ equal-tailed Jeffreys prior interval is given by

$$\mathrm{CI}_J(Y) = (G_{\alpha/2, Y+1/2, 1/n}, G_{1-\alpha/2, Y+1/2, 1/n}) \cap (a, b),$$

where $G(\alpha, s_1, s_2)$ denotes the αth quantile of a Gamma(s_1, s_2) distribution.

Including the two intervals proposed in the previous two sections, there are six intervals introduced in this article. The constructions of these intervals are basically based on pre-data or post-data inferences. Note that the usual confidence interval approach is constructed by pre-data inference approach, which means that the confidence intervals are constructed by inverting a pivotal less than a cutoff point only depending on the confidence level, but not depending on the data, such as the Wald, the score and the LR intervals. For this kind of confidence intervals from pre-data inference, usually the coverage probability of a confidence interval is an increasing function of its expected length. If we want to obtain a confidence interval with higher coverage, then their expected length will be relatively larger. It is impossible to obtain a confidence interval with higher coverage and shorter expected length compared with the existing confidence intervals from pre-data inference unless we use post-data inference. Here, the post-data inference means the cutoff point depending on both confidence level and the observation data.

Different from the three intervals, the rp and the minimax expected length intervals are constructed by post-data inference because $P_b(\bar{X} \geq \bar{x})$ and $P_a(\bar{X} \geq \bar{x})$ in Equation (10) depend on the observation x and the form of Equation (8) depends on the observation x.

The Jeffreys interval is a credible interval in which the endpoints depend on the posterior distribution. The Jeffreys interval is most likely from post-data inference because the rp-value can be viewed as a Bayes estimator [13]. The post-data confidence interval inference is widely used in constructing better confidence interval for a normal variance when the mean is unknown, which is referred to Cohen [24], Goutis and Casella [25] and Shorrock [26].

5.2. *Simulation results*

The performances of the six intervals are evaluated in terms of coverage probability and expected length in this section. Since the minimax property of the truncated Pratt interval does not hold when the bounded parameter spaces is too wide, the parameter space we considered here is the restricted parameter space such that the minimax property for the truncated Pratt interval is valid in this parameter space. When the sample size is large, we can have more information for the unknown parameter. Thus, we can selected the bounds such that the restricted parameter space is shorter and the minimax property holds in this restricted parameter space.

The Wald, score and Jeffreys intervals have similar performances in terms of coverage probability and expected length. Their coverage probabilities are around 0.95, but their expected lengths are longer than those of the other intervals. The LR interval has the similar expected length performance such as the Wald, score and Jeffreys intervals, but its coverage probability is slightly larger than the coverage probabilities of the other intervals. The simulation study is based on 10,000 replicates.

The performances of the truncated Pratt interval and the rp interval are different from the other four intervals. The restricted parameter spaces in Figures 2–5 are chosen such that the expected length of the truncated Pratt interval has a unimodal shape. The expected length of the truncated Pratt interval is a unimodal shape if $(a, b) \in (\theta_0 - h, \theta_0 - h)$, where the condition of choosing h is given in the proof of Theorem 1. When a is less than $\theta_0 - h$ and b is large than $\theta_0 + h$, the expected length may become a convex function with a minimum at θ_0 or near θ_0. Since h has not been specifically derived, we just try to find a and b such that the expected length is a unimodal function. Then, it can guarantee that it is a minimax expected length interval in the restricted parameter space. The restricted parameter space considered here is reasonable when the sample size n is large. As in Figures 2 and 3, assume that we know that the true parameter θ is between 2 and 4. When $n = 20$, the variance of the sample mean is between $\frac{2}{20} = 0.1$ and $\frac{4}{20} = 0.2$. Since the true parameter almost falls within two standard deviations of the mean, the interval $(2, 4)$ is not too short in this case unless the true parameter is near 2 or 4. The expected length of the truncated Pratt interval in Figure 3 is the shortest one among the six intervals.

The rp interval has higher coverage probability than the other intervals when the true θ is not near the boundary. Its expected length is shorter than the other intervals except the truncated Pratt interval when the true parameter is not near the boundary. Although we may expect a confidence interval to have a coverage probability close to the nominal level, a confidence interval with a higher coverage probability and a smaller expected length is preferable because we can shorten the interval such that it has a coverage probability close to the nominal level. A way to shorten the rp interval is to adjust the α value in the interval. For example, when we increase the α value, the length of the rp interval decreases. Thus, we can adjust the α value such that the rp interval has coverage probability closer to the nominal level for the mean in a region. Therefore, it shows that the rp interval is a potential competitor to the other intervals.

6. Real-data example

The comparison of the six intervals is illustrated by a real-data example, about testing the average number of passengers to Taiwan from other countries each month from January 1991 to February 2007. We have a government database of 194 numbers of passengers in the unit of 10,000 from January 1991 to February 2007, which is assumed to follow a Poisson distribution. In this example, the population is the whole data set, that is, 194 numbers of each month passengers. The true mean of the population is 20.9 in the unit of 10,000 and the variance of the population is 23.1. Note that in this case, although the mean and variance are not exactly the same, the deviation is
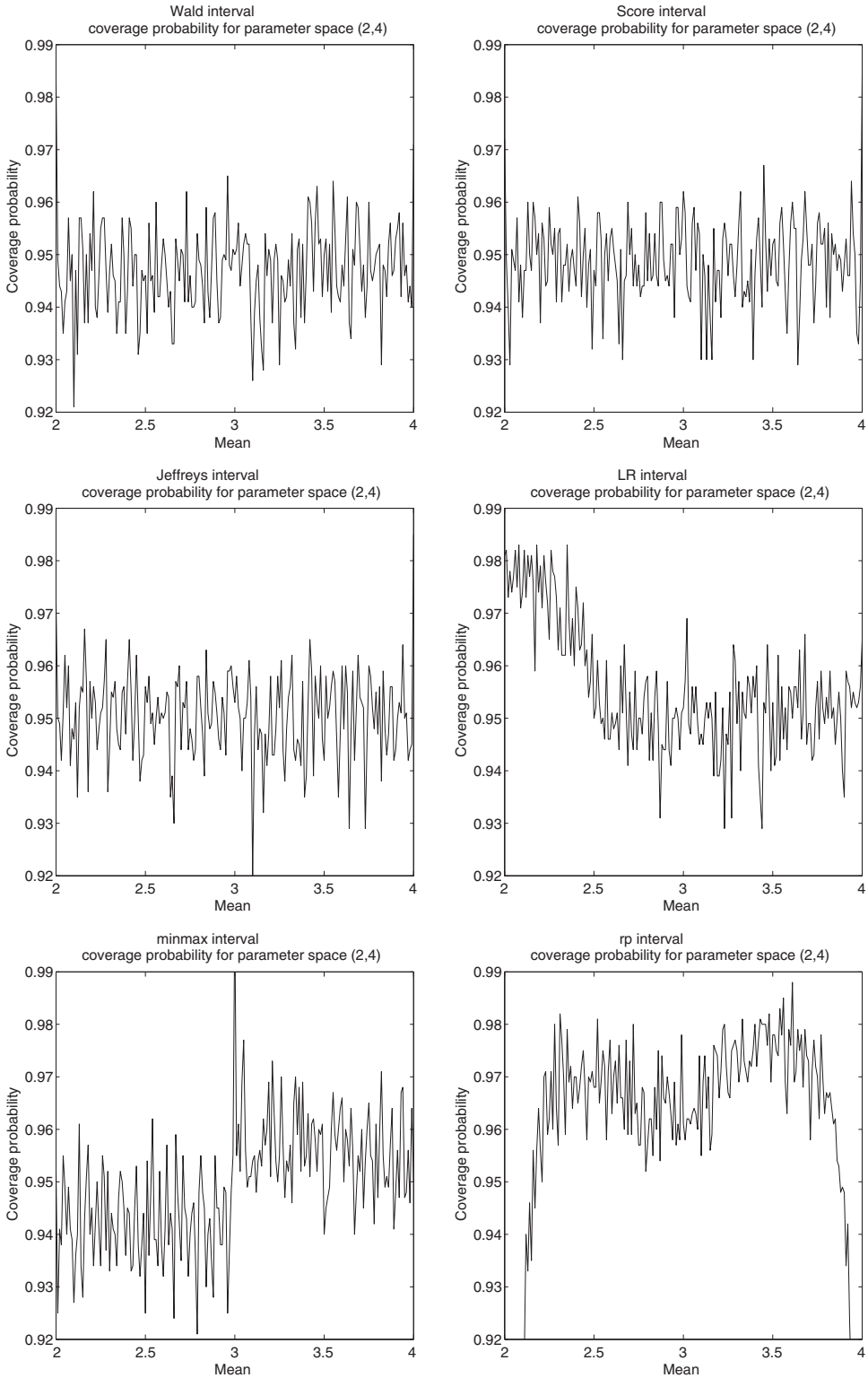
Figure 2.   The coverage probabilities of the six intervals when $n = 20$ with parameter space $(2, 4)$.
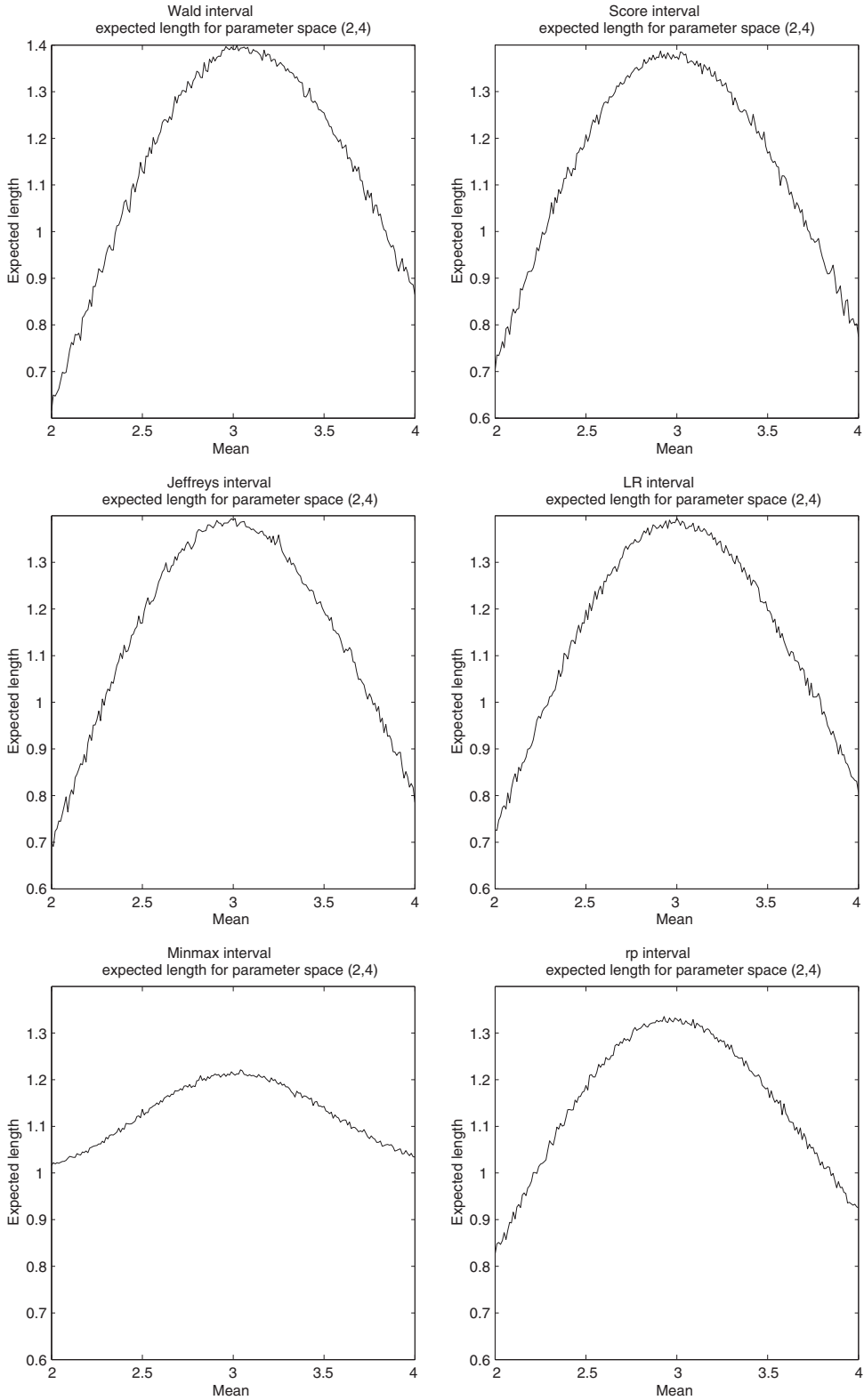
Figure 3.   The expected length of the six intervals when $n = 20$ with parameter space $(2, 4)$.
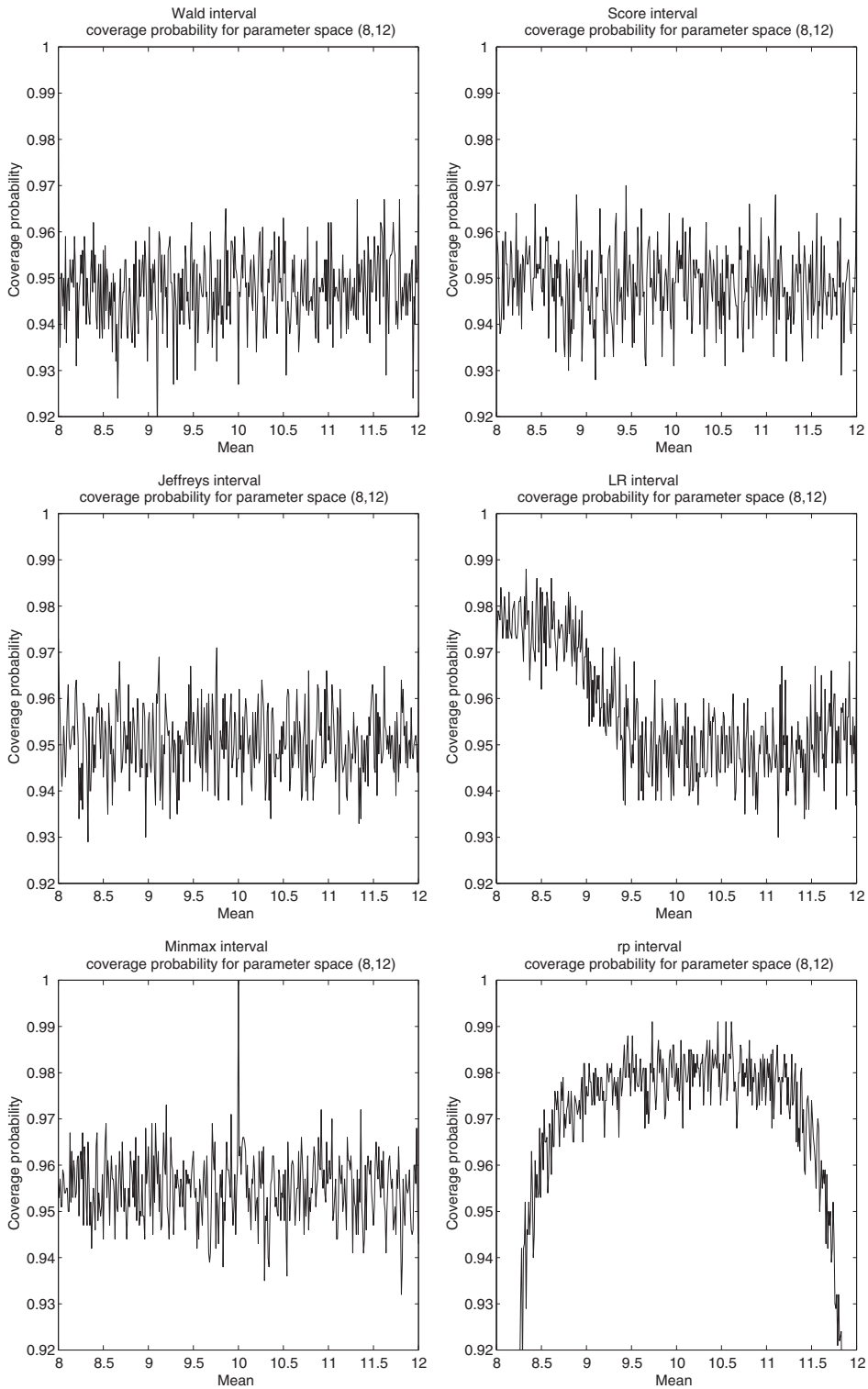
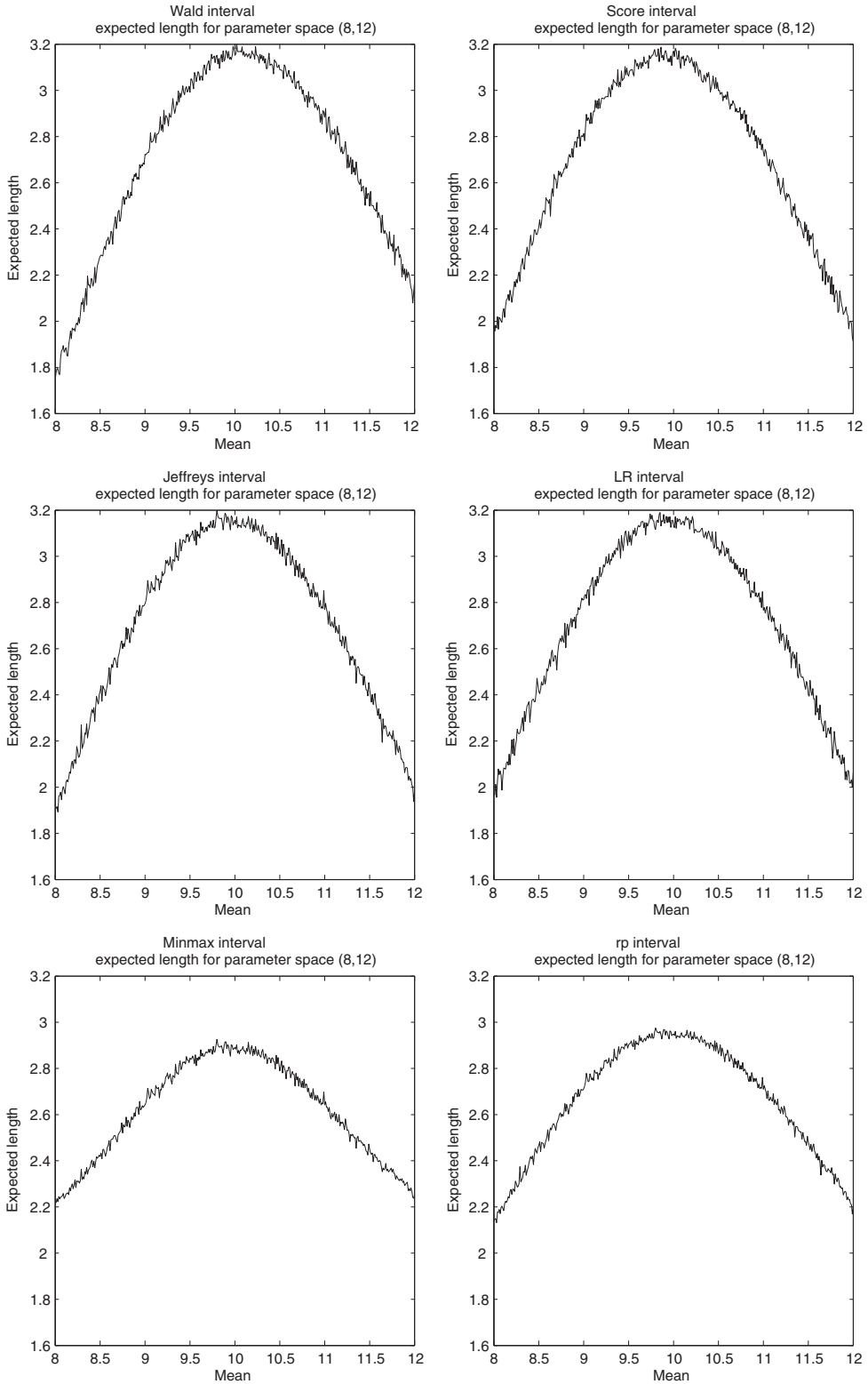Figure 4.   The coverage probabilities of the six intervals when $n = 10$ with parameter space $(8, 12)$.

Figure 5.   The expected length of the six intervals when $n = 10$ with parameter space $(8, 12)$.

Table 1. The coverage probabilities and the expected lengths of the six 0.95 intervals when the bounded parameter space is (18, 24).

	Wald	Score	Likelihood	Truncated Pratt	rp
Coverage probability	0.947	0.943	0.945	0.949	0.955
Expected length	3.8159	3.38135	3.8156	3.347	3.6903

Table 2. The coverage probabilities and the expected lengths of the six 0.95 intervals when the bounded parameter space is (16, 25).

	Wald	Score	Likelihood	Truncated Pratt	rp
Coverage probability	0.947	0.943	0.945	0.946	0.948
Expected length	4.003	4.00	4.0021	3.4701	3.9650

not large. Thus, it is still reasonable to use the Poisson model to approximate it. Suppose that we do not know the whole data set. From an empirical knowledge, we may have rough bounds of the mean. In this example, two restricted parameter spaces (18, 24) and (16, 25) are considered.

Since the mean of the whole data set is 20.9, we conduct a simulation to investigate the performances of the six confidence intervals by comparing their coverage probabilities at the true mean and their expected lengths. We choose samples of 20 numbers from the whole data set with 1000 replicates, and calculate the coverage probability when θ is the true mean 20.9. The expected lengths of the six intervals are also calculated. The coverage probabilities and the expected lengths of the six 0.95 confidence intervals with respect to different bounded parameter spaces are given in Tables 1 and 2.

From Tables 1 and 2, it is shown that the truncated Pratt interval (denoted as the minimax interval in figures) has coverage probabilities close to the nominal level and has the shortest expected lengths; the rp interval has higher coverage probabilities and shorter expected lengths than most of the existing intervals. Basically, these two intervals have good performances and the comparison results coincide with the simulation result conducted in Section 5.

Besides the above simulation study, confidence intervals based on a random sample with size 20 drawn from the data set when the parameter space is assumed to be (18, 24) are presented as follows for readers to compare them. The Wald interval, score interval, LR, truncated Pratt interval and rp interval based on the sample are (18.46, 22.42), (18.56, 22.52), (18.52, 22.49), (18.77, 22.14) and (18.61, 22.47).

7. Conclusion

Two new intervals, the truncated Pratt interval and the rp interval, for a bounded mean in exponential families are proposed in this article. The rp interval is inverted from a rp-value. And the truncated Pratt interval has the property of the minimax expected length when the parameter space is restricted to the region satisfying the condition of Theorem 1. Otherwise, the truncated Pratt interval is not the minimax expected length interval because its expected length increases with the width of the parameter space region.

Both intervals have shorter maximum expected length compared with the four existing methods, the Wald, score, LR and equal-tailed Jeffreys intervals, from the theoretical derivation or the simulation study for the Poisson distribution. The rp interval has higher coverage probability than the other intervals. It is shown from the theoretical and simulation studies that the two proposed intervals are competitive approaches.

Acknowledgements

References

[1] L. Lyons, *Open statistical issues in particle physics*, Ann. Appl. Stat. 2 (2008), pp. 887–915.

[2] D.A.S. Fraser, N. Reid, and A.C.M. Wong, *Inference for bounded parameters*, Phys. Rev. D 69 (2004), p. 033002.

[3] M. Mandelkern, *Setting confidence intervals for bounded parameters*, Stat. Sci. 17 (2002), pp. 149–172.

[4] B.P. Roe and M.B. Woodroofe, *Setting confidence belts*, Phys. Rev. D 60 (2001), pp. 3009–3015.

[5] M. Woodroofe and H. Wang, *The problem of low counts in a signal plus noise model*, Ann. Stat. 28 (2000), pp. 1561–1569.

[6] T. Zhang and M. Woodroofe, *Credible and confidence sets for restricted parameter spaces*, J. Stat. Plann. Infer. 115 (2003), pp. 479–490.

[7] H. Wang, *Comparison of p control charts for low defective rate*, Comput. Stat. Data Anal. 53 (2009), pp. 4210–4220.

[8] H. Wang and F. Tsung, *Tolerance intervals with improved coverage probabilities for binomial and Poisson variables*, Technometrics 51 (2009), pp. 25–33.

[9] T. Cai and H. Wang, *Tolerance intervals for discrete distributions in exponential families*, Statist. Sinica 19 (2009), pp. 905–923.

[10] G.J. Hahn and W.Q. Meeker, *Statistical Intervals: A Guide for Practitioners*, John Wiley & Sons, New York, 1991.

[11] J.W. Pratt, *Length of confidence intervals*, J. Am. Stat. Assoc. 56 (1961), pp. 549–567.

[12] S.N. Evans, B.B. Hansen, and P.B. Stark, *Minimax expected measure confidence sets for restricted location parameters*, Bernoulli 11 (2005), pp. 571–590.

[13] H. Wang, *Modified p-values for one-sided testing in restricted parameter spaces*, Stat. Probab. Lett. 77(15) (2007), pp. 625–631.

[14] H. Wang, *Modified p-value of two-sided test for normal distribution with restricted parameter space*, Commun. Stat. Theory Methods 35 (2006), pp. 1361–1374.

[15] P.J. Bickel and K.A. Doksum, *Mathematical Statistics*, Basic Ideas and Selected Topics Vol. I, Prentice Hall, Upper Saddle River, NJ, 2001.

[16] C. Schafer and P. Stark, *Using what we know: Inference with physical constraints*, Stat. Probl. Part. Phys, Astrophys. Cosmol. (2003), pp. 25–34.

[17] H. Wang, *Confidence intervals for the mean of a normal distribution with restricted parameter space*, J. Stat. Comput. Simul. 78(9) (2008), pp. 829–841.

[18] H. Wang, *Exact average coverage probabilities and confidence coefficients of confidence intervals for discrete distributions*, Stat. Comput. 19 (2009a), pp. 139–148.

[19] H. Wang, *Monotone boundary property and full coverage property of confidence intervals for a binomial proportion*, J. Stat. Plann. Infer. 140 (2010), pp. 495–501.

[20] G.J. Feldman and R.D. Cousins, *Unified approach to the classical statistical analysis of small signals*, Phys. Rev. D 57 (1998), pp. 3873–3889.

[21] B.P. Roe and M. Woodroofe, *Improved probability method for estimating signal in the presence of background*, Phys. Rev. D 60 (1999), p. 053009.

[22] J.T. Hwang, G. Casella, C. Robert, M. Wells, and R. Farrell, *Estimation of accuracy in testing*, Ann. Stat. 20 (1992), pp. 490–509.

[23] L. Brown, T. Cai, and A. DasGupta, *Interval estimation in exponential families*, Statist. Sinica 13 (2003), pp. 19–49.

[24] A. Cohen, *Improved confidence intervals for the variance of a normal distribution*, J. Am. Stat. 67 (1972), pp. 382–387.

[25] C. Goutis and G. Casella, *Improved invariant confidence intervals for a normal variance*, Ann. Stat. 19(4) (1991), pp. 2015–2031.

[26] G. Shorrock, *Improved confidence intervals for a normal variance*, Ann. Stat. 18(2) (1990), pp. 972–980.

Appendix

Proof of Theorem 1 We consider the continuous case here. The discrete case can be replaced the integration by an summation. Note that for a fixed x, Equation (8) is equivalent to $\{\theta : d^{\theta_0}(\theta, x) = 1\} \cap \Theta$. For a $\theta \in (\theta_0 - \tau, \theta_0 + \tau)$, since θ is a strictly increasing function of η, there exists a η such that $\theta = \dot{A}(\eta)$. The expected length of Equation (8) for this θ is

$$L_\eta(d^{\theta_0}) = \int \left(\int 1_{\{\zeta : \theta_x^1 \geq \dot{A}(\zeta)\}} \, d\zeta \right) h(x) \exp\{\eta x - A(\eta)\} \, dx + \int \left(\int 1_{\{\zeta : \theta_x^2 \geq \dot{A}(\zeta)\}} \, d\zeta \right) h(x) \exp\{\eta x - A(\eta)\} dx. \quad (A1)$$

We need to show that $L_\eta(d^{\theta_0})$ is a unimodal function of θ for $\theta \in (\theta_0 - \tau, \theta_0 + \tau)$.

We have

$$\frac{\mathrm{d}}{\mathrm{d}\eta} L_\eta(d^{\theta_0}) = \int \left(\int 1_{\{\zeta : \theta_x^1 \geq \dot{A}(\zeta)\}} \, \mathrm{d}\zeta \right) h(x) \exp\{\eta x - A(\eta)\}(x - \dot{A}(\eta))\mathrm{d}x$$

$$+ \int \left(\int 1_{\{\zeta : \theta_x^2 \geq \dot{A}(\zeta)\}} \mathrm{d}\zeta \right) h(x) \exp\{\eta x - A(\eta)\}(x - \dot{A}(\eta))\mathrm{d}x \tag{A2}$$

If X has a symmetric density function function symmetric about θ_0, Equation (A2) is equal to zero when $\theta = \theta_0$.

Note that for a $\theta_1 < \theta_0$, we have $x - \theta_1 > x - \theta_0$ and for a $\theta_2 > \theta_0$, we have $x - \theta_2 < x - \theta_0$. By this fact, $\theta = \dot{A}(\eta)$ and Equation (A2) is equal to zero when $\theta = \theta_0$, we have Equation (A2) > 0 for $\theta_1 < \theta_0$ near θ_0 and Equation (A2) < 0 for $\theta_2 < \theta_0$ near θ_0. Consequently, Equation (A1) has a local maximum at $\theta = \theta_0$. Therefore, there exist a constant h such that Equation (A1) is a unimodal function for $\theta \in (\theta_0 - \tau, \theta_0 + \tau)$ when $\tau < h$. Since Equation (8) is confidence interval with the shortest expected length at $\theta = \theta_0$, by the continuity of the expected length of a confidence interval and that fact that (A1) is a unimodal function with a maximum at θ_0, Equation (8) is a minimax expected length confidence interval $\theta \in (\theta_0 - \tau, \theta_0 + \tau)$ when $\tau < h$.

For the case that the density or probability function of X is not symmetric to a point, when the sample size n is large, by the normal approximate, $\sum_{i=1}^n T(X_i)$ tends to be symmetric to $n\theta_0$. Thus, for the sample size being large enough, Equation (8) is a minimax expected length confidence interval $\theta \in (\theta_0 - \tau, \theta_0 + \tau)$ when $\tau < h$. ∎

Proof of Theorem 2   Since $\dot{A}(\eta)$ is a strictly increasing function of η, there exists three points v_0, a' and b' such that $\theta_0 = \dot{A}(v_0), a = \dot{A}(a')$ and $b = \dot{A}(b')$. We consider that X is a continuous random variable here. A similar argument can be applied to discrete random variables.

(i) Assume $b = \infty$. The rejection region in Theorem 2 is $\{x : r_{\theta_0}(x) < k\}$ which is equivalent to

$$\{x : P_{\theta_0}(X \geq x) - (1 - k)P_a(X \geq x) - k < 0\}.$$

Let $d(x) = P_{\theta_0}(X \geq x) - (1 - k)P_a(X \geq x) - k$, which is equal to $\int_x^\infty h(x)\mathrm{e}^{v_0 x - A(v_0)}\mathrm{d}x - (1 - k)\int_x^\infty h(x)\mathrm{e}^{a'x - A(a')}$ $\mathrm{d}x - k$. Note that

$$\frac{\partial d(x)}{\partial x} = -h(x)\mathrm{e}^{v_0 x - A(v_0)} - (1 - k)h(x)\mathrm{e}^{a'x - A(a')}$$

$$= -h(x)\mathrm{e}^{v_0 x}((1 - k)\mathrm{e}^{(a'-v_0)x - A(a')} - \mathrm{e}^{-A(v_0)}) \tag{A3}$$

Set Equation (A3) to zero, then we have $x = x^* = 1/v_0 - a' \log((1 - k)\mathrm{e}^{-A(a')+A(v_0)})$. Since Equation (A3) is positive for $x < x^*$ and negative for $x > x^*$, $d(x)$ is increasing for $x < x^*$ and decreasing for $x > x^*$. Consequently, $d(x)$ has a local maximum at $x = x^*$. Note also that $d(a_0) = 0$ and $d(x) \to -k$ as $x \to \infty$, where a_0 is the lower bound of the sample space, that is, $x \geq a_0 = -\infty$. Using these facts, we have $\{x : d(x) < 0\}$ is equivalent to $\{x : x \geq x_\alpha\}$, where x_α is a positive constant. Thus, this is the rejection region of a Neyman–Pearson test. So, the test is a UMP test.

(ii) Assume b is a finite constant. The rejection region is $\{x : r_{\theta_0}(x) < k\}$ which is now equivalent to

$$\{x : P_{\theta_0}(X \geq x) - (1 - k)P_a(X \geq x) - kP_b(X \geq x) < 0\}.$$

Let

$$d(x) = P_{\theta_0}(X \geq x) - (1 - k)P_a(X \geq x) - kP_b(X \geq x).$$

$$\frac{\partial d(x)}{\partial x} = -(h(x)\mathrm{e}^{v_0 x - A(v_0)} - (1 - k)h(x)\mathrm{e}^{a'x - A(a')} - kh(x)\mathrm{e}^{b'x - A(b')})$$

$$= h(x)\mathrm{e}^{v_0 x}((1 - k)\mathrm{e}^{(a-v_0)x - A(a')} + k\mathrm{e}^{(b'-v_0)x - A(b')} - \mathrm{e}^{-A(v_0)}) \tag{A4}$$

Let $g(x) = (1 - k)\mathrm{e}^{(a'-v_0)x - A(a')} + k\mathrm{e}^{(b'-v_0)x - A(b')} - \mathrm{e}^{-A(v_0)}$. It can be verified that the function g cross the x-axis at most twice.

We have

$$g'(x) = (1 - k)\mathrm{e}^{(a'-v_0)x - A(a')}(a - v_0) + k\mathrm{e}^{(b'-v_0)x - A(b')}(b' - v_0).$$

$$g''(x) = (1 - k)\mathrm{e}^{(a'-v_0)x - A(a')}(a - v_0)^2 + k\mathrm{e}^{(b'-v_0)x - A(b')}(b' - v_0)^2.$$

Note that $g''(t) > 0$, which implies that $g'(x)$ is an increasing function.

We consider two cases for the lower bound of the sample space: (I) $a_0 = -\infty$; (II) $a_0 > -\infty$. For example, for the normal distribution, the lower bound is $-\infty$; for the Poisson distribution, the lower bound of the sample space is 0.

First consider the case (I). Note that $g(x) \to \infty$ as $x \to \infty$ or $x \to -\infty$, and $d(x) \to 0$ as $x \to \infty$ or $x \to -\infty$. Note that $g'(x)$ is an increasing function.

There are two cases for $g'(-\infty)$:

Case 1: $g'(-\infty) < 0$. It leads that (i) $g'(x) < 0$ for all x or (ii) $g'(x)$ is negative first then positive.

Case 2: $g'(-\infty) > 0$. It leads that $g'(x) > 0$ for all x.

Since $g(-\infty) = \infty$, the corresponding situations for $g(x)$ for the above Cases 1 and 2 are:

Case 1: (1) $g(x)$ is always positive (corresponding to (i) of the above Case (i)) (2) $g(x)$ is always positive, or is positive first, then negative and positive again (corresponding to (ii) of the above Case 1)

Case 2: $g(x)$ is always positive.

Since $d(-\infty) = d(\infty) = 0$, the possible situations for $d(x)$ are:

Case 1: (1) There does not exist $d(x)$ satisfying the condition.

(2) There does not exist $d(x)$ satisfying the first part. The $d(x)$ must increases, then decrease to a negative constant, and then increase to zero.

Case 2: There does not exist $d(x)$ satisfying the condition.

Thus in case (I), there is only one situation for $d(x)$ which increases first, then decrease to a negative constant, and then increase to zero. The rejection region $\{x : d(x) < 0\}$ for this situation is equivalent to $\{x > x_*\}$, where x_* is a constant.

Now we consider the case (II). In this case, we do not know the sign for $g'(a_0)$. There are two cases for $g'(a_0)$:

Case 1: $g'(a_0) < 0$. It leads that (i) $g'(x) < 0$ for all x or (ii) $g'(x)$ is negative first then positive.

Case 2: $g'(a_0) > 0$. It leads that $g'(x) > 0$ for all x.

The corresponding situations for $g(x)$ for Cases 1 and 2 are:

Case 1: (1) $g(a_0) > 0$ and $g(x)$ is always positive (corresponding to (i)). Since $g(\infty) = \infty$ and $g'(x) < 0$, it is impossible that $g(a_0) < 0$.

(2) $g(a_0) > 0$ and $g(x)$ is first positive, then negative and positive again (corresponding to (ii))

Case 2: (1) $g(a_0) > 0$ and $g(x)$ is always positive or $g(a_0) < 0$ and $g(x)$ is first negative and then positive.

Since $d(-\infty) = d(\infty) = 0$, the possible situations for $d(x)$ are:

Case 1: (1) There does not exist $d(x)$ satisfying the condition.

(2) $d(x)$ must increases, then decrease to a negative constant, and then increase to 0.

Case 2: There does not exist satisfied $d(x)$ satisfying the first part. $d(x)$ must decrease to a negative constant, and then increase to 0.

Thus in case (II), there are two possible situations for $d(x)$:

The first one is that $d(x)$ increases, then decrease to a negative constant, and then increase. The rejection region $\{x : d(x) < 0\}$ for this situation is equivalent to $\{x > x_*\}$, where x_* is a constant.

The second one is that decreases to a negative constant then increases to 0. The rejection region $\{x : d(x) < 0\}$ for this situation is equivalent to $\{x > a_0\}$.

Combining the above results, the rejection region for each case is $\{x > c\}$ for a constant c, which is a rejection region of a Neyman–Pearson test. Thus, the test is a UMP test. ∎