Interval estimation *via* tail functions

Borek Puza[1,2] and Terence O'Neill[1], 6 May 2005

[1] School of Finance and Applied Statistics, Australian National University, ACT 0200, Australia. [2] Corresponding author: borek.puza@anu.edu.au

Abstract

In this paper we describe a new methodology for constructing confidence intervals. The idea is to specify the tail cutoff areas in terms of a function of the target parameter rather than as constants. This function, called the *tail function*, can be engineered so as to provide shorter confidence intervals when prior information is available. It can also be used to improve the coverage properties of approximate confidence intervals. We illustrate the methodology by applying it to inference on the normal mean and binomial proportion, and develop measures of the resulting improvements. Guidelines for choosing the optimal tail function in any situation are provided, and the relationship with Bayesian inference is discussed.

Key words

Confidence interval, shorter, prior, tail function, Wilson, Clopper-Pearson, Blaker.

1. Introduction

The classical theory of two-sided confidence intervals (CI's) (Stuart *et al*., 1999) involves inverting a probability statement regarding a pivotal quantity, *Y*, so as to yield two statistics which 'straddle' the quantity of interest, θ. This theory requires the specification of a confidence level, $1-\alpha$ (*e.g.* 0.95), which 'cuts off' an area of $\alpha/2$ from either tail of *Y*'s distribution. In the statistical literature the tail areas have always been taken as constant. The new method involves specifying these areas in terms of a function of θ, called the *tail function*, and then proceeding to invert the probability statement to create a different CI for θ. The main advantage of this approach is that it can be used to produce shorter CI's when prior information is available. It also provides a mechanism for improving the coverage properties of approximate CI's for parameters of discrete distributions.

The classical theory of CI's is reviewed in Section 2 and generalised in Section 3 by way of tail functions. Section 4 shows how these functions can be used in conjunction with prior information to produce shorter CI's for a normal mean when the normal variance is known. We define measures of the improvements which may be achieved, such as the maximum reduction in interval width. We also develop guidelines for selecting the optimal tail function in any situation, and make comparisons with the Bayesian approach. Section 5 then deals with the case of unknown variance, aided by Monte Carlo methods. Section 6 shows how tail functions can be used to produce attractive variants of the 'standard', Wilson and Clopper-Pearson intervals for a binomial proportion. Section 7 concludes with some general discussion and advice.

2. The classical theory of two-sided confidence intervals

Consider a scalar probability distribution which is dependent on a single unknown constant parameter, θ, and suppose that we are given a random sample of n observations, $X_1,...,X_n$, from that distribution, with realised or possible values denoted $x_1,...,x_n$. Let X and x denote the vectors $(X_1,...,X_n)$ and $(x_1,...,x_n)$, respectively. Next let $Y = g(X,\theta)$ be a scalar function of X and θ with realised or possible value denoted $y = g(x,\theta)$, and suppose that the cumulative distribution function (cdf) of Y, denoted $F_Y(y)$, is continuous and does not depend on θ. Then $F_Y(Y)$ has the standard uniform distribution, and we call Y a pivotal quantity. Next choose the confidence level, $1-\alpha$, where $\alpha \in (0,1)$. Then it is true, for any value which θ may take, that

$$1-\alpha = P(\alpha/2 < F_Y(g(X,\theta)) < 1-\alpha/2). \quad (1)$$

The classical theory of CI's involves manipulating (1) so as to produce the statement

$$1-\alpha = P(L(X) < \theta < U(X)), \quad (2)$$

where L and U are two functions of X which do not depend on θ. The $1-\alpha$ confidence interval (CI) for θ is then defined as (l,u), where l and u are the solutions in θ of the following equations, respectively:

$$F_Y(g(x,\theta)) = 1-\alpha/2 \quad (3)$$
$$F_Y(g(x,\theta)) = \alpha/2. \quad (4)$$

In many cases these equations can be solved analytically or at least easily using standard statistical software. However, in general they may require an iterative search procedure such as the Newton-Raphson algorithm. Explicitly, to solve (4) we choose a suitable starting value θ_0 and repeatedly calculate $\theta_{j+1} = \theta_j - \psi(\theta_j)/\psi'(\theta_j)$ until convergence, where:

$$\psi(\theta) = F_Y(g(x,\theta)) - \alpha/2 \tag{5}$$

$$\psi'(\theta) = \frac{\partial}{\partial\theta} F_Y(g(x,\theta)) = f_Y(g(x,\theta)) \frac{\partial}{\partial\theta} g(x,\theta)\,. \tag{6}$$

In (6), $f_Y(y) = F_Y'(y)$ denotes the probability density function (pdf) of *Y*. Equation (3) can be solved in the same way as (4) but with $\alpha/2$ in (5) replaced by $1-\alpha/2$.

3. A generalisation of the classical theory *via* tail functions

Consider $\tau(\theta)$, a function of θ which is nondecreasing and has a range in the interval from 0 to 1. Then just as (1) is true for all θ, so also

$$1-\alpha = P(\alpha\tau(\theta) < F_Y(g(X,\theta)) < 1-\alpha+\alpha\tau(\theta))\,. \tag{7}$$

We now recognise (1) as being a special case of (7), namely where $\tau(\theta)$ = 1/2 (a constant). Suppose that it is possible to invert (7) so as to produce a statement of the form (2). Then a $1-\alpha$ CI's for θ is (*l*,*u*), where *l* and *u* are the solutions in θ of the following equations, respectively:

$$F_Y(g(x,\theta)) = 1-\alpha+\alpha\tau(\theta) \tag{8}$$

$$F_Y(g(x,\theta)) = \alpha\tau(\theta)\,. \tag{9}$$

Equations (8) and (9) are solvable *via* the Newton-Raphson algorithm in the same way as (3) and (4). Explicitly, to solve (9) we choose a suitable starting value θ_0 and repeatedly calculate $\theta_{j+1} = \theta_j - \psi(\theta_j)/\psi'(\theta_j)$ until convergence, where:

$$\psi(\theta) = F_Y(g(x,\theta)) - \alpha\tau(\theta) \tag{10}$$

$$\psi'(\theta) = f_Y(g(x,\theta)) \frac{\partial}{\partial\theta} g(x,\theta) - \alpha\tau'(\theta), \quad \tau'(\theta) = \frac{\partial}{\partial\theta}\tau(\theta)\,. \tag{11}$$

Equation (8) can be solved in the same way as (9) but with $\alpha\tau(\theta)$ in (10) replaced by $1-\alpha+\alpha\tau(\theta)$, and with (11) unchanged. We call $\tau(\theta)$ the *tail function*.

4. Inference on a normal mean with known variance

To illustrate the new technique, suppose that we have a random sample $X_1,...,X_n$ from the normal distribution with unknown mean μ and known variance σ^2. We wish to construct a $1-\alpha$ CI for μ. The classical approach involves using the result

$$Y = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1), \quad \text{where } \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ is the sample mean,}$$

and taking the tail function as $\tau(\mu) = 1/2$, $-\infty < \mu < \infty$ (constant). The resulting CI is $(\bar{x} \pm z_{\alpha/2}\sigma/\sqrt{n})$, where $\bar{x} = (x_1 + ... + x_n)/n$ and z_p denotes the upper *p*-quantile of the standard normal distribution.

Let us now consider a different tail function, one of the form

$$\tau(\mu) = \delta + (1-2\delta)\Phi\left(\frac{\mu-\eta}{\lambda}\right), \tag{12}$$

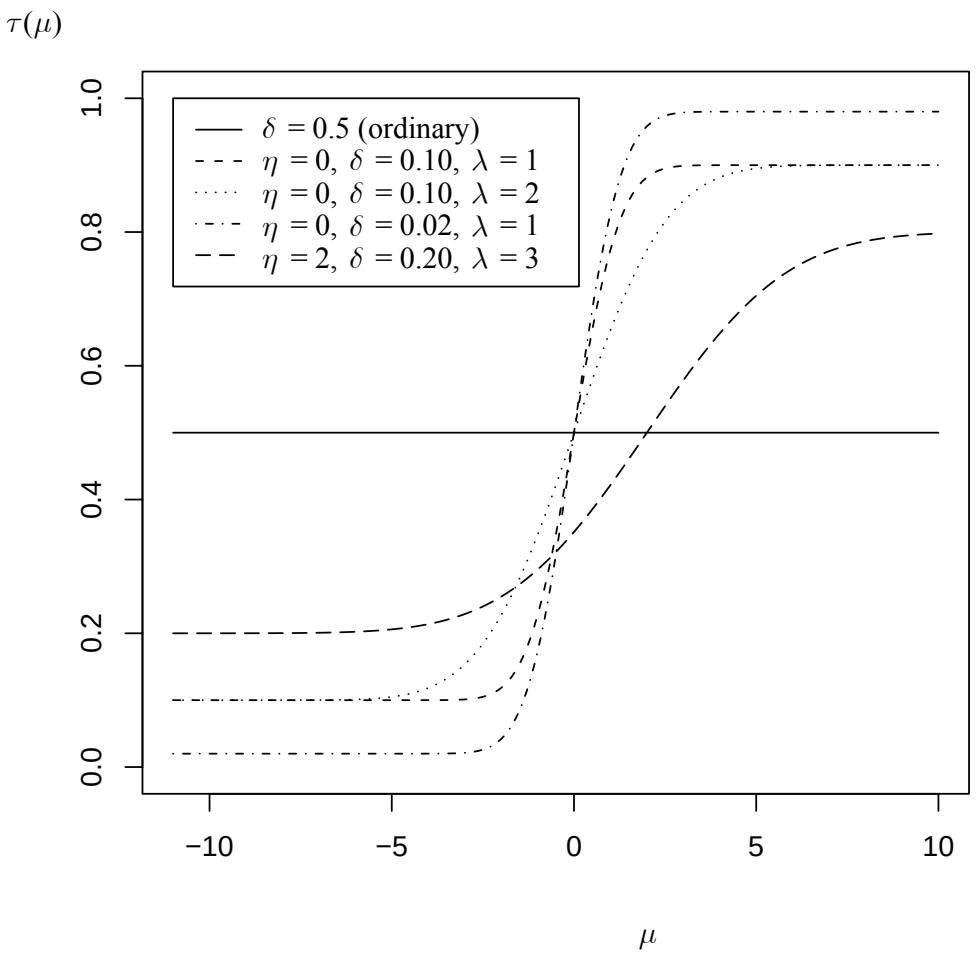
where $\eta \in \Re$, $\lambda \geq 0$, $0 \leq \delta \leq 1/2$, and $\Phi(z)$ denotes the standard normal cdf. This tail function is nondecreasing, with values δ at $-\infty$, 1/2 at η, and $1-\delta$ at $+\infty$. It is also continuous if $\lambda > 0$, in which case its slope is given by

$$\tau'(\mu) = \frac{\partial}{\partial\mu}\tau(\mu) = \left(\frac{1-2\delta}{\lambda}\right)\phi\left(\frac{\mu-\eta}{\lambda}\right), \tag{13}$$

where $\phi(t)$ denotes the standard normal pdf. Several examples of $\tau(\mu)$ are shown in Figure 1. Note that δ determines the minimum deviation of $\tau(\mu)$ from 0 or 1, η is the value of μ at which $\tau(\mu)$ equals 1/2, and λ is inversely proportionate to the slope of $\tau(\mu)$, whose maximum value is $\tau'(\eta) = (1-2\delta)/(\lambda\sqrt{2\pi})$.

We will refer to the tail function (12) and associated CI's as *Gaussian* with parameters η, δ and λ, although the terms *modified*, *alternative* and *new* will often be used instead. Observe that $\tau(\mu)$ reduces to 1/2 if δ = 1/2 or $\lambda = \infty$. In that case we refer to $\tau(\mu)$ and associated CI's as *ordinary* or *classical*.

Figure 1 Several Gaussian tail functions



To find the lower bound of a Gaussian $1-\alpha$ CI, we choose a suitable starting value μ_0 and repeatedly calculate $\mu_{j+1} = \mu_j - \psi(\mu_j)/\psi'(\mu_j)$ until convergence, where:

$$\psi(\mu) = \Phi\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right) - \alpha\tau(\mu) \tag{14}$$

$$\psi'(\mu) = \phi\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right)\left(\frac{-1}{\sigma/\sqrt{n}}\right) - \alpha\tau'(\mu). \tag{15}$$

The upper bound can be found in the same way but with $\alpha\tau(\mu)$ in (14) replaced by $1-\alpha+\alpha\tau(\mu)$, and with (15) unchanged. *S-PLUS* code (Venables and Ripley, 1999) and *R* code (Venables and Smith, 2004) for calculating these bounds and those of all other CI's in this paper can be obtained from the authors upon request.

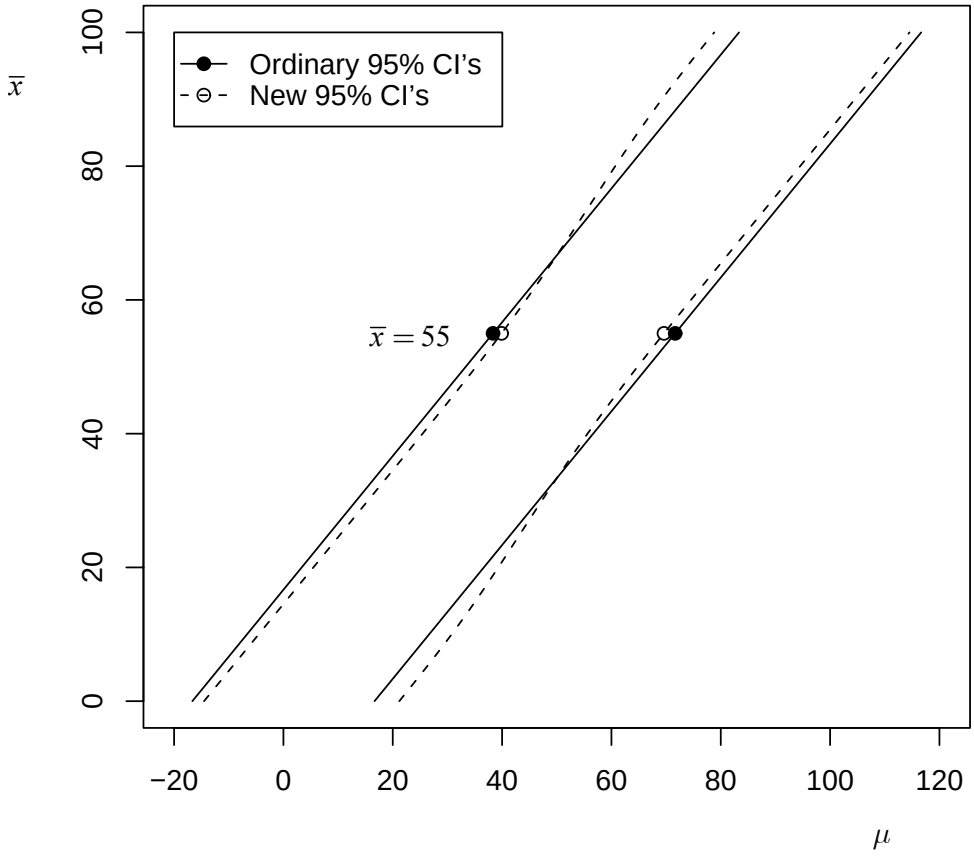
Example 1

Suppose that $n = 4$, $\bar{x} = 55$ and $\sigma = 17$. Then the classical 95% CI for μ is $(55 \pm 1.96(17)/\sqrt{4}) = (38.3, 71.7)$. Implementing the above Newton-Raphson algorithm with starting values equal to the classical bounds, and with $\eta = 50$, $\delta = 1/8$ and $\lambda = 10$, we find that the new CI is (39.9, 69.6). Observe that the lengths of these two intervals are 33.3 and 29.7, respectively. Thus the alternative tail function has resulted in a width reduction of about 11%.

To see what is going on, let us plot the classical and new confidence limits for all values of $\bar{x}$ from 0 to 100. Figure 2 illustrates, with the bounds of the two CI's in Example 1 shown as points (see Note 1). We see that the alternative tail function creates a 'distortion' of the classical bounds, so that when $\bar{x}$ is near $\eta = 50$, the new CI is *narrower*. This effect is compensated for by the new CI being somewhat *wider* when $\bar{x}$ is *far* from η. We find that this pattern is symmetric about η. Thus the alternative CI has a minimum width at $\bar{x} = \eta$, and at $\bar{x} = 45$ its width is the same as at $\bar{x} = 55$, *i.e.* 29.7.

With these observations we have established the significance of the parameter η. It should be specified as a value of μ which is deemed very likely *a priori*, for then $\bar{X}$ will likely be close to η and hence the resulting CI will likely be narrower than the ordinary CI. This idea will be made more precise later on.

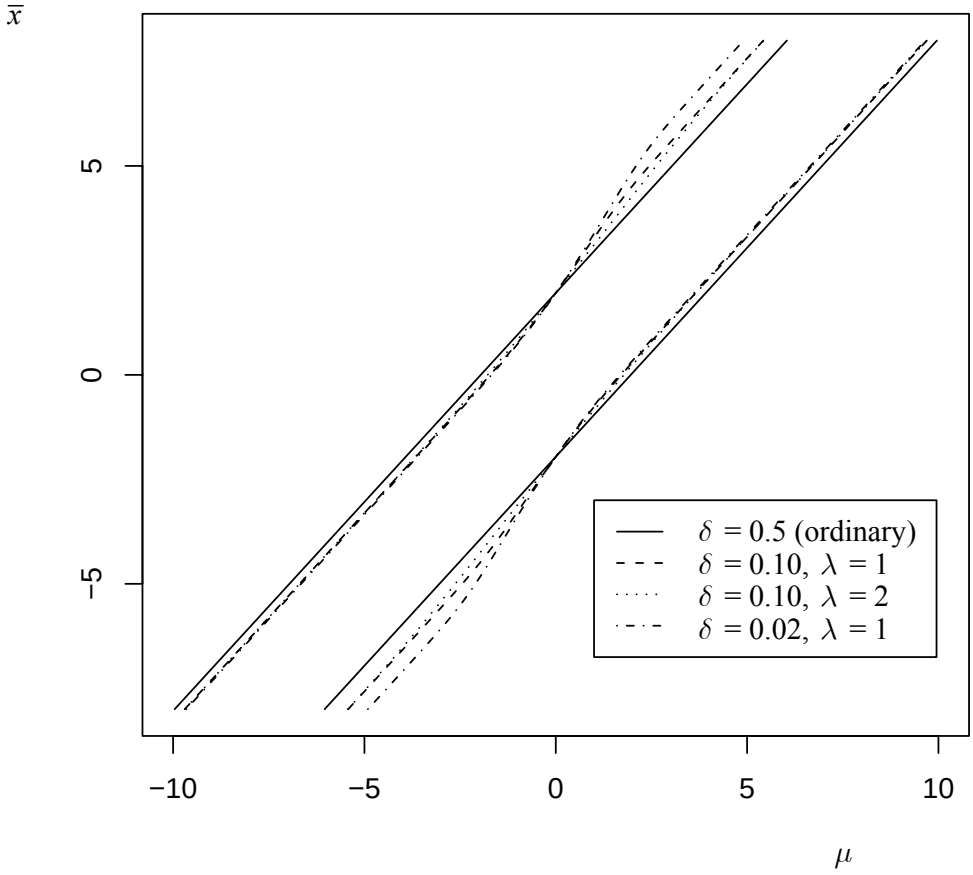
Figure 2 Two sets of 95% confidence bounds when $n = 4$, $\sigma = 17$, $\eta = 50$, $\delta = 1/8$ and $\lambda = 10$



The significance of δ and λ

To gain an understanding of the parameters δ and λ, it is useful to examine some more examples. Figure 3 shows, for $\eta = 0$ and $n = \sigma = 1$, the 95% confidence bounds implied by the first four tail functions in Figure 1 (including the 'ordinary' one). We see that δ determines the minimum and maximum possible lengths of the CI, and λ influences the *rate* at which those extrema are approached as $\bar{x}$ tends to η or $\pm\infty$. Specifying a small value of δ results in a relatively short interval if $\bar{x}$ is close to η and a wide one otherwise. Specifying a large value of λ has the effect of evening out the disturbance due to δ, whilst not altering the crossover points, given generally by $(\mu, \bar{x}) = (\eta, \eta \pm z_{\alpha/2}\sigma/\sqrt{n})$. Examples of these points are $(0, \pm 1.96)$ in Figure 3, and $(50, 50 \pm 1.96(17)/\sqrt{4})$ = (50,33.3) and (50,66.7) in Figure 2.

Figure 3 Several sets of 95% confidence bounds when $n = \sigma = 1$ and $\eta = 0$



The PDL and EPDL

A convenient way to quantify the length (or width) of a modified CI is in terms of the proportion by which it is shorter than the corresponding ordinary CI. This leads us to define the *proportional decrease in length (PDL)* as

$$D_{\delta,\lambda}(\bar{x}) = \frac{W - W_{\delta,\lambda}(\bar{x})}{W}, \tag{16}$$

where $W_{\delta,\lambda}(\bar{x}) = u - l$ is the length of the modified CI and $W = 2z_{\alpha/2}\sigma/\sqrt{n}$ is the length of the ordinary CI.

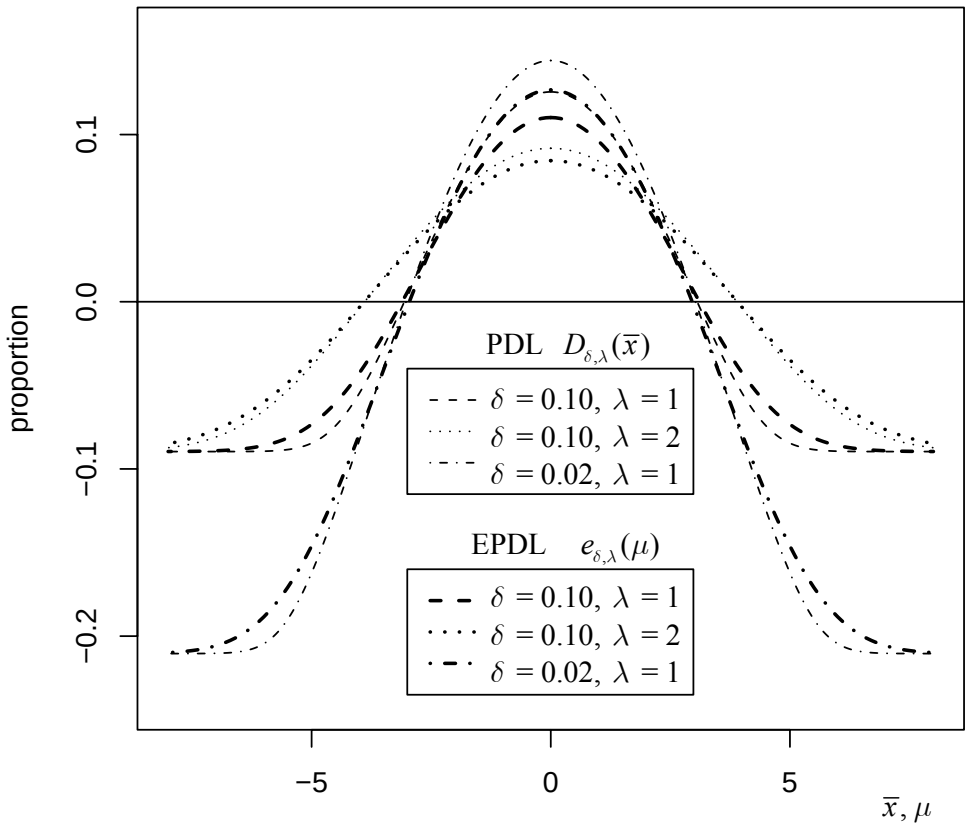
Another useful quantity is the *expected proportional decrease in length (EPDL)*, which may be defined as

$$e_{\delta,\lambda}(\mu) = E\{D_{\delta,\lambda}(\bar{X}) \mid \mu\} = \int_{-\infty}^{\infty} D_{\delta,\lambda}(\bar{x}) f(\bar{x} \mid \mu) d\bar{x}\,, \tag{17}$$

where $f(\bar{x} \mid \mu) = \phi((\bar{x} - \mu)\sqrt{n}/\sigma)\sqrt{n}/\sigma$, since $(\bar{X} \mid \mu) \sim N(\mu, \sigma^2/n)$.

Figure 4 shows both the PDL and EPDL for each of the three new CI's in Figure 3. The first of these two functions shows clearly the values of $\bar{x}$ for which the CI is narrower than the ordinary CI (a neighbourhood around $\eta = 0$). In contrast, the latter function shows the values of μ for which the CI is *expected* to be narrower (again near η). Both the PDL and EPDL are symmetric about and have a maximum at η.

Figure 4 Examples of the PDL and EPDL when *n* = σ = 1 and η = 0



The MPIL

It is also useful to be able to quantify the extent by which a CI in the new class may be *worse* than the ordinary CI. It can be shown that the maximum width of a modified CI is $\tilde{W} = W_{\delta,\lambda}(\infty) = (z_{\alpha\delta} + z_{\alpha(1-\delta)})\sigma/\sqrt{n}$. Accordingly, we define the *maximum proportional increase in length (MPIL)* as

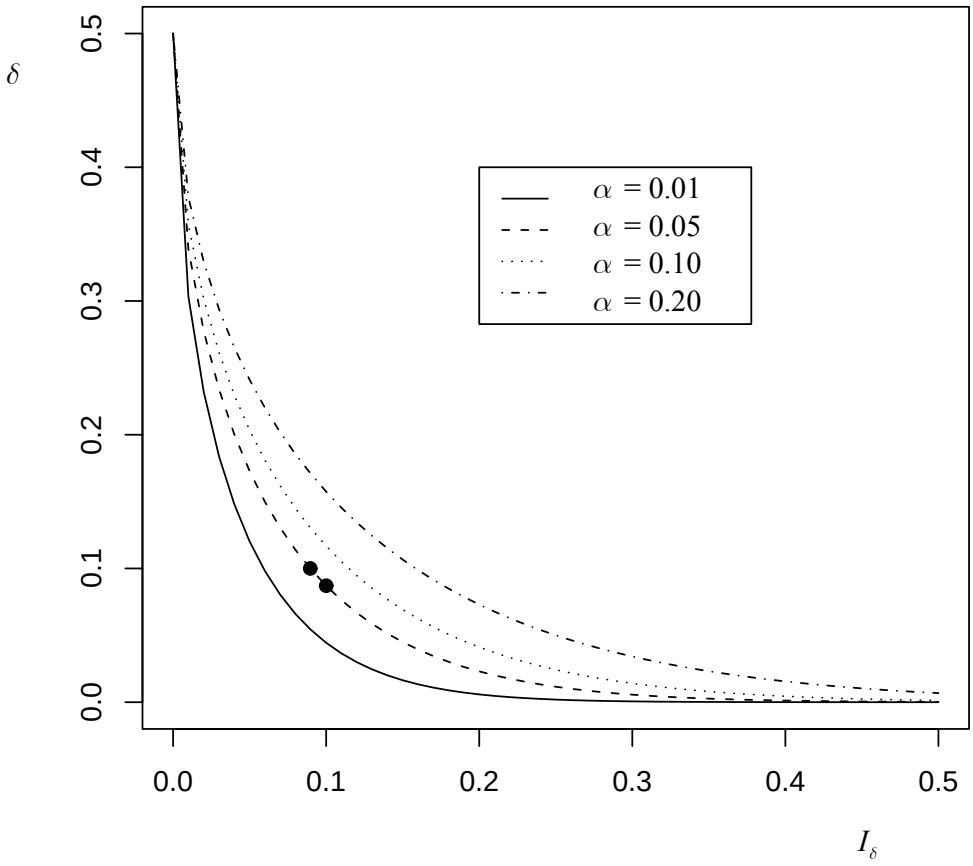
$$I_\delta = \frac{\tilde{W} - W}{W} = \frac{z_{\alpha\delta} + z_{\alpha(1-\delta)}}{2z_{\alpha/2}} - 1. \qquad (18)$$

For example if α = 0.05 then $I_{0.1}$ = 0.0896. Observe that the MPIL does not depend on λ. Also, it is strictly increasing and converges to infinity as δ tends to zero.

It is a fact that decreasing δ always has the effect of increasing the maximum possible decrease in length, namely at $\bar{x} = \eta$ (see Figures 3 and 4). This suggests a useful rule. Suppose we want to be *sure* that the new CI does not exceed the old CI in length by more than 100*I*%. Then we should specify δ as the value for which $I_\delta = I$.

For example, if I = 0.1 and α = 0.05, then we find using the Newton-Raphson algorithm that δ = 0.0871 (see Note 2). Figure 5 illustrates the relationship between δ and the MPIL for various values of α, with the two points being at (I_δ, δ) = (0.0896,0.1) and (0.1,0.0871).

Figure 5 Relationship between δ and the MPIL



The MPDL, GMPDL and AGMPDL

Recall that a modified CI is always narrowest at $\bar{x} = \eta$. Accordingly we define the *maximum proportional decrease in length (MPDL)* as $M_{\delta,\lambda} = D_{\delta,\lambda}(\eta)$. We find that for a given δ, this quantity is a nonincreasing function of λ and so has a maximum at $\lambda = 0$. Accordingly, we define the *greatest maximum proportional decrease in length (GMPDL)* as $G_\delta = M_{\delta,0}$. We then find that G_δ is a decreasing function of δ. Accordingly, we define the *absolute greatest maximum proportional decrease in length (AGMPDL)* as $A = G_0$. The following are formulae for all these quantities:

$$
\begin{aligned}
&D_{\delta,\lambda}(\bar{x}) = 1 - W_{\delta,\lambda}(\bar{x})/W && \text{(PDL)} \\
&M_{\delta,\lambda} = D_{\delta,\lambda}(\eta) = 1 - W_{\delta,\lambda}(\eta)/W && \text{(MPDL)} \\
&G_\delta = M_{\delta,0} = 1 - z_{\alpha(1-\delta)}/z_{\alpha/2} && \text{(GMPDL)} \\
&A = G_0 = 1 - z_\alpha / z_{\alpha/2} && \text{(AGMPDL).}
\end{aligned}
\qquad (19)
$$

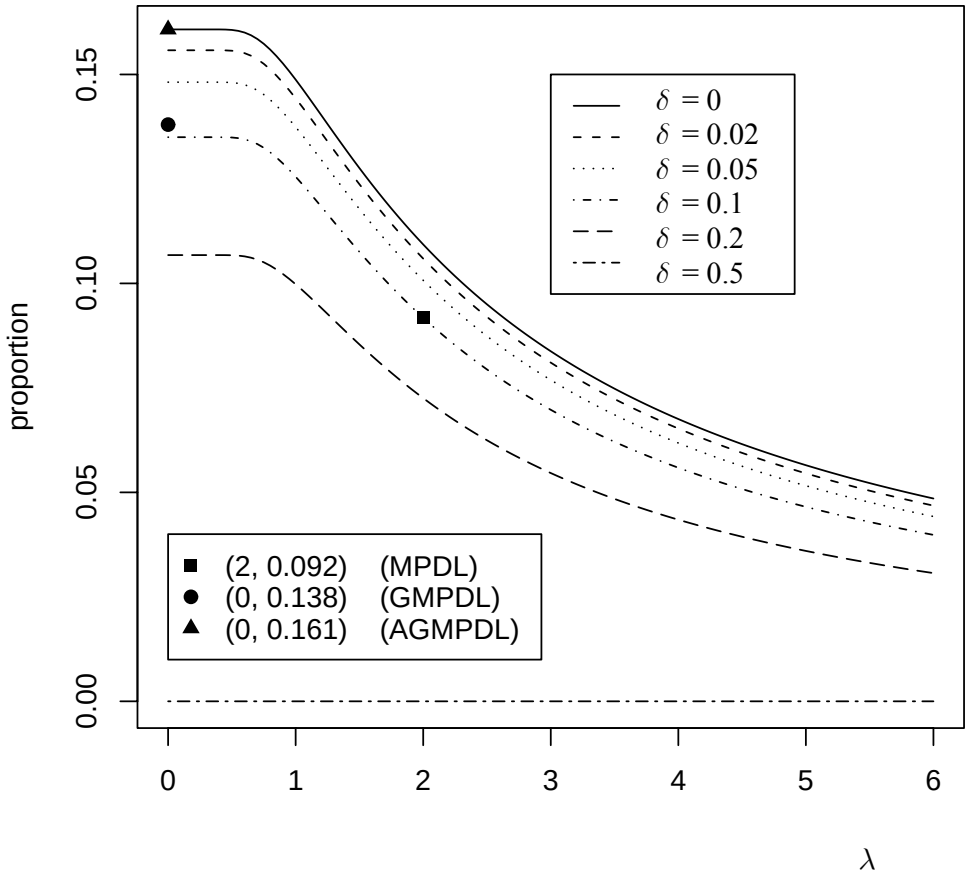
Figure 4 shows three values of the MPDL when α = 0.05 and $\sigma = n = 1$, namely the heights of the three thin lines at $\bar{x}$ = η = 0. These values are $M_{0.1,1}$ = 0.1255, $M_{0.1,2}$ = 0.0918 and $M_{0.02,1}$ = 0.1443. Thus, for example, if δ = 0.1 and λ = 2, the improvement in interval length is limited to 9.2%.

To see the significance of the GMPDL, recall that a MPIL of 10% implies δ = 0.0871. The associated GMPDL is $G_{0.0871} = 1 - z_{0.05(1-0.0871)} / z_{0.05/2}$ = 0.138. This means that if we want to be sure that the new 95% CI will be no more than 10% wider than the ordinary 95% CI, we must also accept that it will be no more than 13.8% *narrower*, with that limit achievable only if λ = 0.

The AGMPDL provides an absolute upper bound on the improvement to a $1-\alpha$ CI which can be achieved. For example, if α = 0.05 then $A = 1 - 1.645/1.96$ = 0.161. Thus a modified 95% CI can be no more than 16.1% narrower than an ordinary 95% CI, regardless of the values of $\bar{x}$, σ, n, η, δ and λ. Some other examples of the AGMPDL are 0.097, 0.221, 0.343 and 1.000 when α = 0.01, 0.10, 0.20 and 0.50, respectively. From this we see that modification *via* tail functions is potentially most effective if an interval with *small* confidence level is required.

Figure 6 provides further illustrations of the MPDL, GMPDL and AGMPDL, with some of the above numbers shown as points. The height of each line at the origin represents a GMPDL, and the height of the highest line at the origin is the AGMPDL for the case α = 0.05. For further discussion of Figure 6 see Note 3.

Figure 6 Examples of the MPDL, GMPDL and AGMPDL when n = σ = 1 and α = 0.05



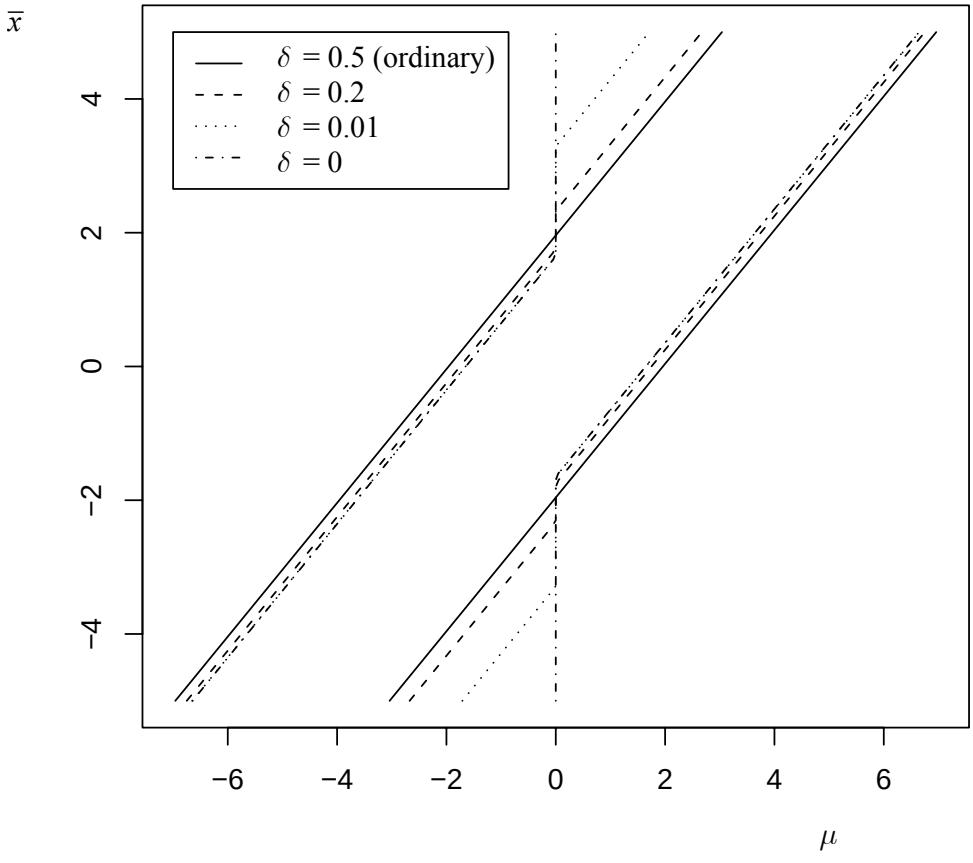
The case of λ **= 0**

In the limiting case λ = 0, the new interval can be written tractably, with no need for the Newton-Raphson algorithm. With $\kappa = \sigma/\sqrt{n}$, its bounds are then:

$$
\begin{array}{lll}
(\overline{x} - z_{\alpha\delta}\kappa, & \overline{x} + z_{\alpha(1-\delta)}\kappa), & \overline{x} > \eta + \kappa z_{\alpha\delta} \\
(\eta, & \overline{x} + z_{\alpha(1-\delta)}\kappa), & \eta + z_{\alpha(1-\delta)}\kappa \le \overline{x} < \eta + z_{\alpha\delta}\kappa \\
(\overline{x} - z_{\alpha(1-\delta)}\kappa, & \overline{x} + z_{\alpha(1-\delta)}\kappa), & \eta - z_{\alpha(1-\delta)}\kappa \le \overline{x} \le \eta + z_{\alpha(1-\delta)}\kappa \\
(\overline{x} - z_{\alpha(1-\delta)}\kappa, & \eta), & \eta - z_{\alpha\delta}\kappa \le \overline{x} < \eta - z_{\alpha(1-\delta)}\kappa \\
(\overline{x} - z_{\alpha(1-\delta)}\kappa, & \overline{x} + z_{\alpha\delta}\kappa), & \overline{x} < \eta - z_{\alpha\delta}\kappa
\end{array}
\qquad (20)
$$

Figure 7 shows 95% CI's corresponding to various values of δ when $\eta = \lambda = 0$ and $\sigma = n = 1$. The case $\lambda = 0$ is important because it implies the maximum possible improvement at $\bar{x} = \eta$ for a given δ. It thereby provides a convenient benchmark against which other values of λ may be assessed. Note that when $\lambda = 0$ the tail function (12) is discontinuous and consists of a single step up of $1-2\delta$ (from δ to $1-\delta$) at $\mu = \eta$. We will refer to this tail function and associated CI's such as those in Figure 7 as *maximal* (or *maximal Gaussian*).

Figure 7 Several maximal 95% CI's ($\lambda = 0$) when $n = \sigma = 1$ and $\eta = 0$



The PEPDL

In order to choose useful values of η, δ and λ, we must have in mind some prior distribution for μ. Suppose that this prior is normal with mean μ_0 and standard deviation σ_0 (see Note 4). In that case we should take $\eta = \mu_0$ and may set δ at the value implied by the desired MPIL. That leaves only λ to be specified. One possibility is to let $\lambda = 0$, because then the PDL at (16) will be maximised. However, it may be preferable to take λ as the value λ_0 which maximizes the *prior expected proportional decrease in length (PEPDL)*, defined as

$$p_\delta(\lambda) = ED_{\delta,\lambda}(\bar{X}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} D_{\delta,\lambda}(\bar{x}) f(\bar{x} \mid \mu) f(\mu)\, d\bar{x}\, d\mu, \qquad (21)$$

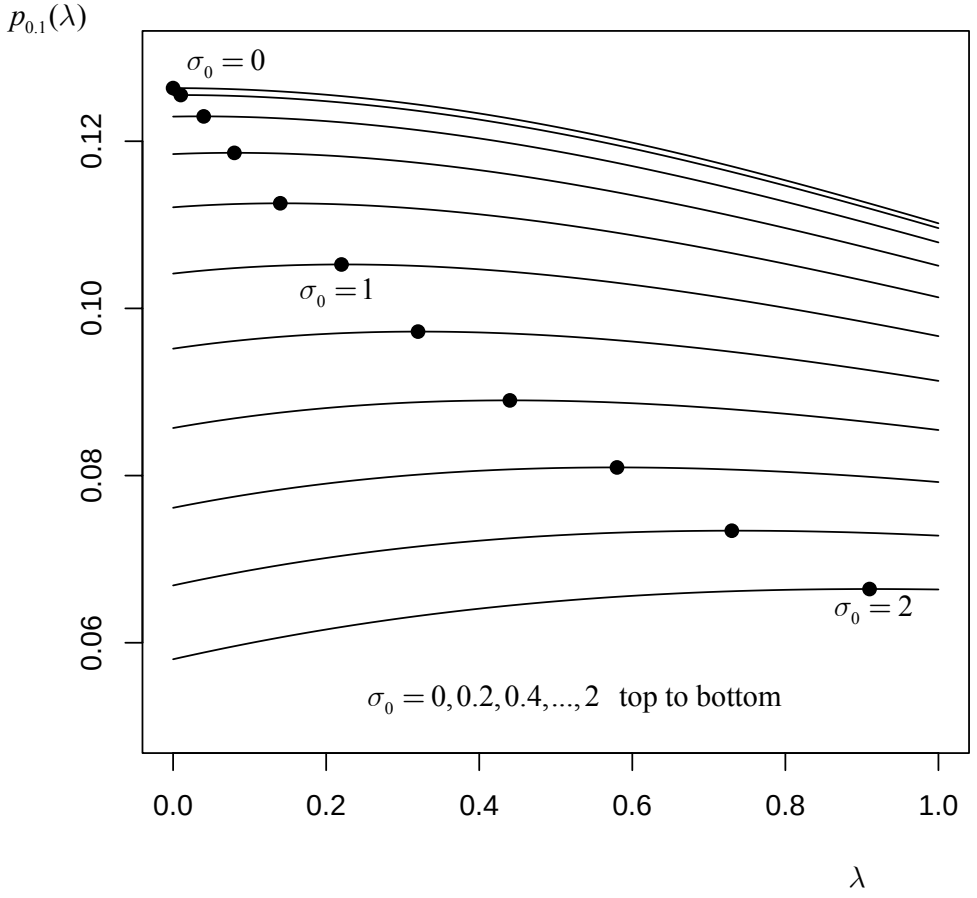
where $f(\mu) = \phi((\mu - \mu_0)/\sigma_0)/\sigma_0$ and $f(\bar{x} \mid \mu) = \phi((\bar{x} - \mu)\sqrt{n}/\sigma)\sqrt{n}/\sigma$. For other ways to write the PEPDL see Note 5.

Calculation of λ_0 requires a search by trial and error. Figure 8 shows the PEPDL at each $\lambda = 0, 0.01, 0.02, ..., 1$ for various values of σ_0 when $n = \sigma = 1$, $\delta = 0.1$, $\mu_0 = 0$ and $\alpha = 0.05$. For example, if $\sigma_0 = 1$ then the PEPDL is maximized at $\lambda = 0.22$, to the nearest 0.01. Refining our search we find that $\lambda_0 = 0.2233$ and $p_{0.1}(\lambda_0) = 0.1053$. We also find that $p_{0.1}(0) = 0.1042$. Thus in this case there was not much gain from finding λ_0 (only a 1.0% decrease in the PEPDL). For inference we may as well take λ as zero and use the simple formulae at (20).

For the case $\sigma_0 = 2$, we find that the optimal value of λ is 0.9123, with associated PEPDL 0.0664. This amounts to an improvement of 14.5% over 0.0580, the PEPDL when $\lambda = 0$. From this we see that when the prior information is more diffuse, there are greater benefits to be had from finding λ_0. On the other hand, less prior information implies a smaller possible improvement. This is illustrated in Figure 8 where the maxima uniformly decrease as the prior standard deviation σ_0 increases.

Observe that $0.9123 \approx 0.2233 \times 4$. Thus λ_0 at $\sigma_0 = 2$ is roughly 2^2 times λ_0 at $\sigma_0 = 1$. Also, $\lambda_0 = 0$ when $\sigma_0 = 0$. These facts are part of a general pattern which may be written $\lambda_0 \approx k_\delta \sigma_0^2 \sqrt{n}/\sigma$. Three examples of the constant k_δ here are $k_{0.1} = 0.22$, $k_{0.05} = 0.27$ and $k_{0.20} = 0.16$. Although this formula is only approximate it does provide some guidance to conducting the search for λ_0 in any given situation.

Figure 8 Examples of the PEPDL and its maximization when α = 0.05, n = σ = 1, $\eta = \mu_0$ = 0 and δ = 0.1



Comparison with Bayesian inference

It may appear that the tail function methodology is *Bayesian* (see Lee, 1989) since it requires the specification of a prior distribution. However, this is not so, because nowhere does that methodology involve a posterior calculation. The resulting CI's are all *frequentist* in the usual sense. It may then be asked why Bayesian posterior intervals should not be used instead, considering that with informative priors they can result in far greater width reductions than those which have been mentioned (*e.g.* an AGMPDL of 16.1% when α = 0.05).

The answer is that a Bayesian posterior interval may fail to contain the target parameter with the required probability if that parameter happens to be outside 'the most probable region'. In contrast, frequentist CI's have the *correct* coverage probabilities for *all* possible values of the target parameter. Tail functions provide a tool whereby prior information can be usefully incorporated into the inferential process *without* sacrificing this very important feature of the frequentist approach.

Example 2

Suppose that μ's prior is $N(\mu_0, \sigma_0^2)$ (see Note 4). Then it is a standard result that μ's posterior is $N(\mu_*, \sigma_*^2)$, where $\mu_* = (1-k)\mu_0 + k\bar{x}$, $\sigma_*^2 = k\sigma^2/n$ and $k = 1/\{1+\sigma^2/(n\sigma_0^2)\}$ (the 'credibility factor'). It follows that μ's $1-\alpha$ highest posterior density region (HPDR) is $H(\bar{x}) = (\mu_* \pm z_{\alpha/2}\sigma_*)$. The length of this interval is $2z_{\alpha/2}\sigma_*$ and so the associated PDL is $1 - 2z_{\alpha/2}\sigma_* / \{2z_{\alpha/2}\sigma/\sqrt{n}\} = 1-\sqrt{k}$.

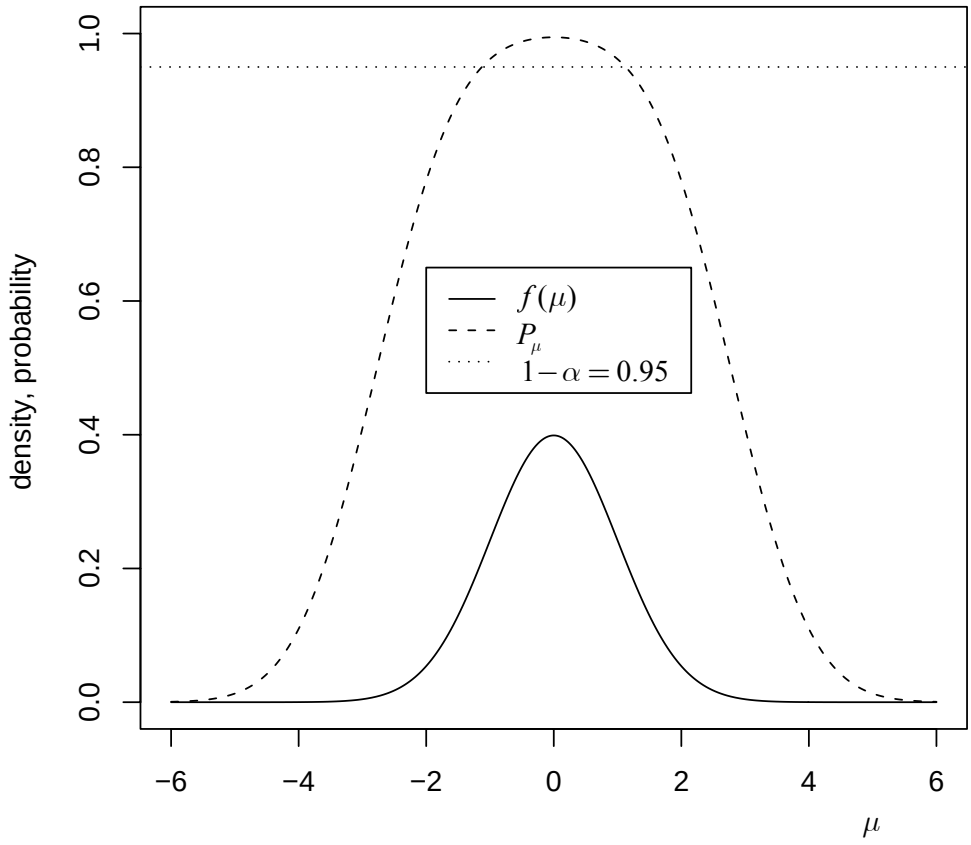
For example, suppose that $\mu_0 = 0$ and $n = \sigma = 1$. Then for $\sigma_0 = 0$, 1, 2, respectively, we find that k = 0, 0.5, 0.8; and the associated PDL's are 1, 0.293 and 0.106. These results may be compared to the corresponding PEPDL's in Figure 8, whose maxima are 0.126, 0.105 and 0.066. Thus the Bayesian 95% HPDR is in every case *shorter* in expectation than the optimal modified 95% CI defined by tail function (12), $\eta = \mu_0$ and $\delta = 0.1$.

Let us now examine the frequentist coverage properties of μ's Bayesian HPDR. It can be shown that the conditional probability of that HPDR containing μ is

$$P_\mu = P(\mu \in H(\bar{X}) \mid \mu) = \Phi\left(\frac{(\mu-\mu_0)(1-k)}{k\sigma/\sqrt{n}} + \frac{z_{\alpha/2}}{\sqrt{k}}\right) - \Phi\left(\frac{(\mu-\mu_0)(1-k)}{k\sigma/\sqrt{n}} - \frac{z_{\alpha/2}}{\sqrt{k}}\right) \quad (22)$$

(see Note 6). Figure 9 shows this probability when $\alpha = 0.05$, $\mu_0 = 0$ and $n = \sigma = \sigma_0 = 1$, and also μ's prior pdf, $f(\mu) = \phi((\mu-\mu_0)/\sigma_0)/\sigma_0$. We see that $P_\mu \longrightarrow 0$ as $\mu \longrightarrow \pm\infty$. Also, $P_\mu > 0.95$ if and only if $-1.12 < \mu < 1.12$; and therefore $P(P_\mu > 0.95) = 0.737$. Thus the frequentist coverage probability of the Bayesian 95% HPDR falls severely short of the required 95% level if μ happens to be far from its prior mean, μ_0; and the prior probability of that coverage probability being at least 95% is only 74%. In contrast, any 95% CI in the class defined by tail function (12) contains μ with probability *exactly* 95% for *all* possible values of μ.

Figure 9 Prior density of μ and coverage probabilities of μ's HPDR



5. Inference on a normal mean with unknown variance

Consider the scenario in Section 4 but with the normal variance σ^2 unknown rather than known. In that case the standard classical approach is to make use of the result

$$Y = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)\text{, where } S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \text{ is the sample variance,}$$

and thereby construct a $1-\alpha$ CI for μ of the form $(\bar{x} \pm t_{\alpha/2}(n-1)s/\sqrt{n})$, where $t_p(v)$ denotes the upper *p*-quantile of the *t*-distribution with *v* degrees of freedom.

To find the upper bound of a modified $1-\alpha$ CI for μ in this case, we choose a suitable starting value μ_0 and repeatedly calculate $\mu_{j+1} = \mu_j - \psi(\mu_j)/\psi'(\mu_j)$ until convergence, where now:

$$\psi(\mu) = F_{t(n-1)}\left(\frac{\bar{x}-\mu}{s/\sqrt{n}}\right) - \alpha\tau(\mu) \tag{23}$$

$$\psi'(\mu) = f_{t(n-1)}\left(\frac{\bar{x}-\mu}{s/\sqrt{n}}\right)\left(\frac{-1}{s/\sqrt{n}}\right) - \alpha\tau'(\mu)\,. \tag{24}$$

Here, $F_{t(n-1)}(.)$ and $f_{t(n-1)}(.)$ denote the cdf and pdf of the *t* distribution with $n-1$ degrees of freedom (*cf.* (14) and (15)). The lower bound can be found in the same way but with $\alpha\tau(\mu)$ in (23) replaced by $1-\alpha+\alpha\tau(\mu)$, and with (24) unchanged.

When λ = 0, the confidence bounds are given by (20) with each z_p replaced by $t_p(n-1)$, and with σ in each κ replaced by *s*. Similar modifications apply to all the other equations in Section 4, such as that for the MPIL at (18). As in Section 4, it is a good practise to specify η as μ's prior mean, μ_0, and to set δ equal to the value implied by a desired MPIL.

If an optimal value of λ is desired, we first define the PEPDL in this context as

$$p_\delta(\lambda) = ED_{\delta,\lambda}(\bar{X},S) = \int_{-\infty}^{\infty}\int_{0}^{\infty}\int_{-\infty}^{\infty}\int_{0}^{\infty} D_{\delta,\lambda}(\bar{x},s)\,f(\bar{x},s\,|\,\mu,\sigma)\,f(\mu,\sigma)\,d\bar{x}\,ds\,d\mu\,d\sigma \tag{25}$$

(*cf.* (21)), where $D_{\delta,\lambda}(\bar{x},s) = 1 - W_{\delta,\lambda}(\bar{x},s)/W(s)$ is the PDL (*cf.* (16)), $W_{\delta,\lambda}(\bar{x},s)$ is the width of the modified CI, $W(s) = 2t_{\alpha/2}(n-1)s/\sqrt{n}$ is the width of the ordinary CI, $f(\bar{x},s\,|\,\mu,\sigma)$ is the joint conditional pdf of $\bar{X}$ and *S*, and $f(\mu,\sigma)$ is the joint prior pdf of μ and σ. For more details see Note 7.

As in Section 4 our task is to find the value of λ which maximizes $p_\delta(\lambda)$. Fortunately, (25) can easily be approximated *via* Monte Carlo for any given value of λ. One way is to draw values of μ and σ from their joint prior distribution, generate a random sample of size *n* from the associated $N(\mu,\sigma^2)$ distribution, calculate the corresponding standard and new CI's, repeat many times, and each time record the PDL. The result will be a random sample, $d_1,...,d_J$, from the prior distribution of the PDL. An unbiased point estimate of and (ordinary) 95% CI for the PEPDL, $p_\delta(\lambda)$, are then $\bar{d} = J^{-1}\sum_{j=1}^{J} d_j$ and $(\bar{d} \pm 1.96 s_d/\sqrt{J})$, where $s_d^2 = (J-1)^{-1}\sum_{j=1}^{J}(d_j - \bar{d})^2$. Repeating this whole process using many different values of λ will lead to an estimate of the value λ_0 which maximises $p_\delta(\lambda)$.

Example 3

Suppose that we are about to sample n = 4 observations from the $N(\mu,\sigma^2)$ distribution where σ^2 is unknown and then construct a 95% CI for μ. We want the CI to be shorter than the standard 95% CI, but don't want to risk it being more than 10% longer. Solving $\{t_{\alpha\delta}(3) + t_{\alpha(1-\delta)}(3)\}/\{2t_{\alpha/2}(3)\} - 1 = 0.1$ with $\alpha = 0.05$ (see (18)), we find that the appropriate value of δ is 0.2227.

Suppose that our prior distributions for μ and σ are independently $N(\mu_0,\sigma_0^2)$ and $Gamma(a,b)$, where μ_0 = 50, σ_0 = 5, a = 100 and b = 10 (so that $E\sigma = a/b = 10$). These priors are shown in Figure 10. Applying the above Monte Carlo procedure with η = 50, δ = 0.2227, J = 1000 and each λ = 0,1,2,...,20, we obtain the estimates and (ordinary) 95% CI's for the PEPDL shown in Figure 11. We see that there is little gain to be had from refining the search, and decide to take λ = 0. Figure 11 reveals that if our priors are 'correct' then we can expect the maximal CI defined by (η,δ,λ) = (50,0.2227,0) to be about 14.5% narrower than the standard 95% CI (see Note 8).

We now observe the data and find that $\bar{x}$ = 55 and s = 17. Hence by (20) the maximal 95% CI for μ is $(55 \pm t_{0.05(1-0.2227)}(3)17/\sqrt{4})$ = (32.6,77.4). In contrast, the ordinary 95% CI is $(55 \pm t_{0.025}(3)17/\sqrt{4})$ = (27.9, 82.1). The lengths of these CI's are 44.9 and 54.1. Therefore the new methodology has resulted in an improvement of 17.1%.

Note that this is the maximum improvement which was possible, since by (19) the GMPDL is $G_{0.2227} = 1 - t_{\alpha(1-0.2227)}(3)/t_{\alpha/2}(3)$ = 0.171. Also note that exactly the same improvement would have been achieved had the sample mean, $\bar{x}$ = 55, been any other number in the interval $(50 \pm t_{0.05(1-0.2227)}(3)17/\sqrt{4})$ = (27.6, 72.4).

Figure 10 Prior densities of the normal mean and standard deviation

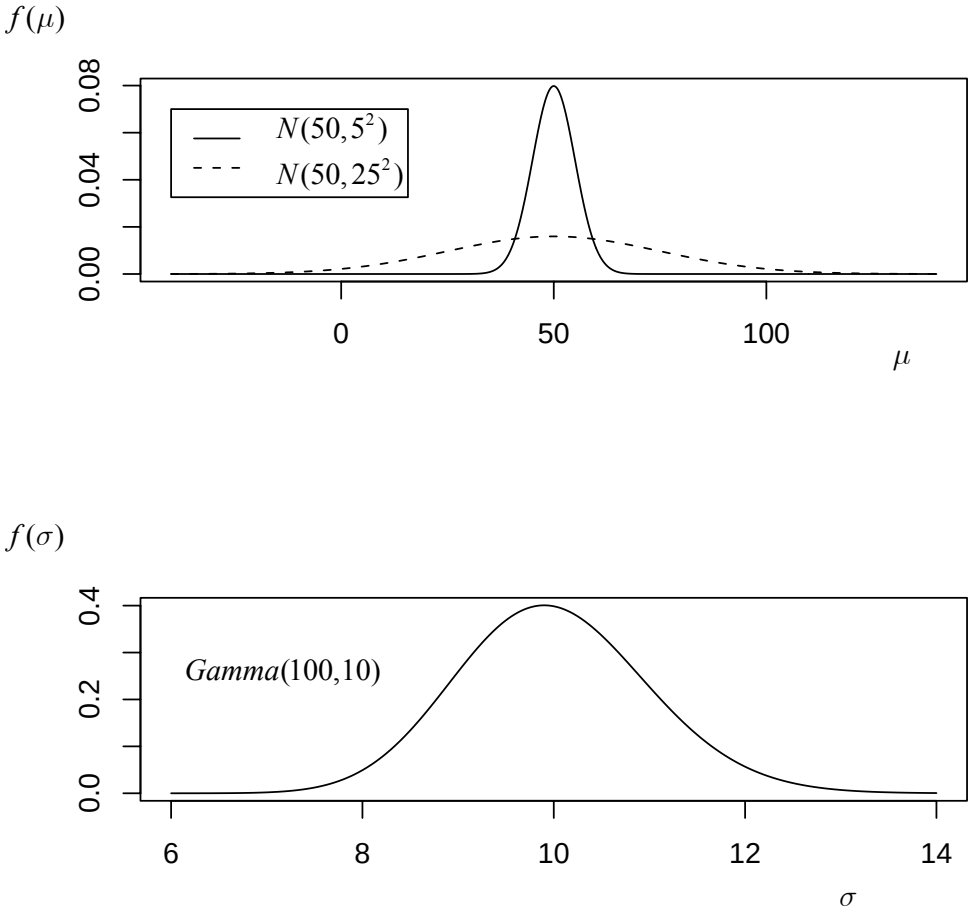
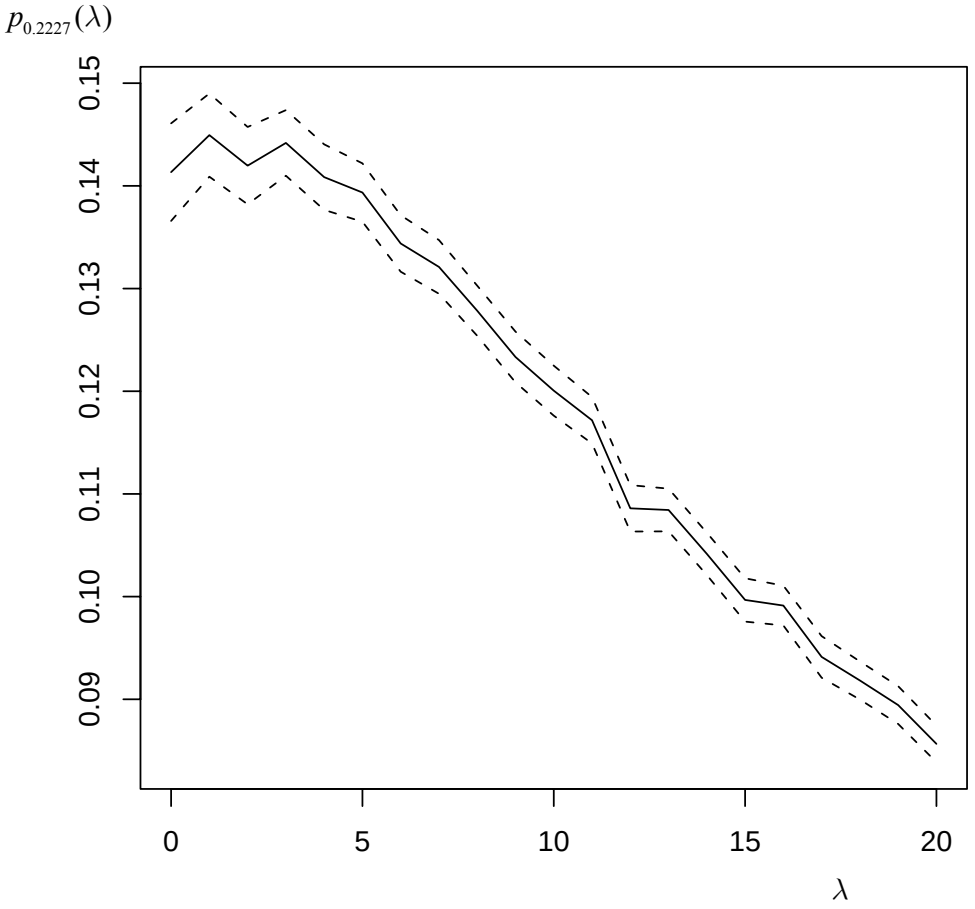


Figure 11 Monte Carlo estimates of the PEPDL when σ_0 = 5



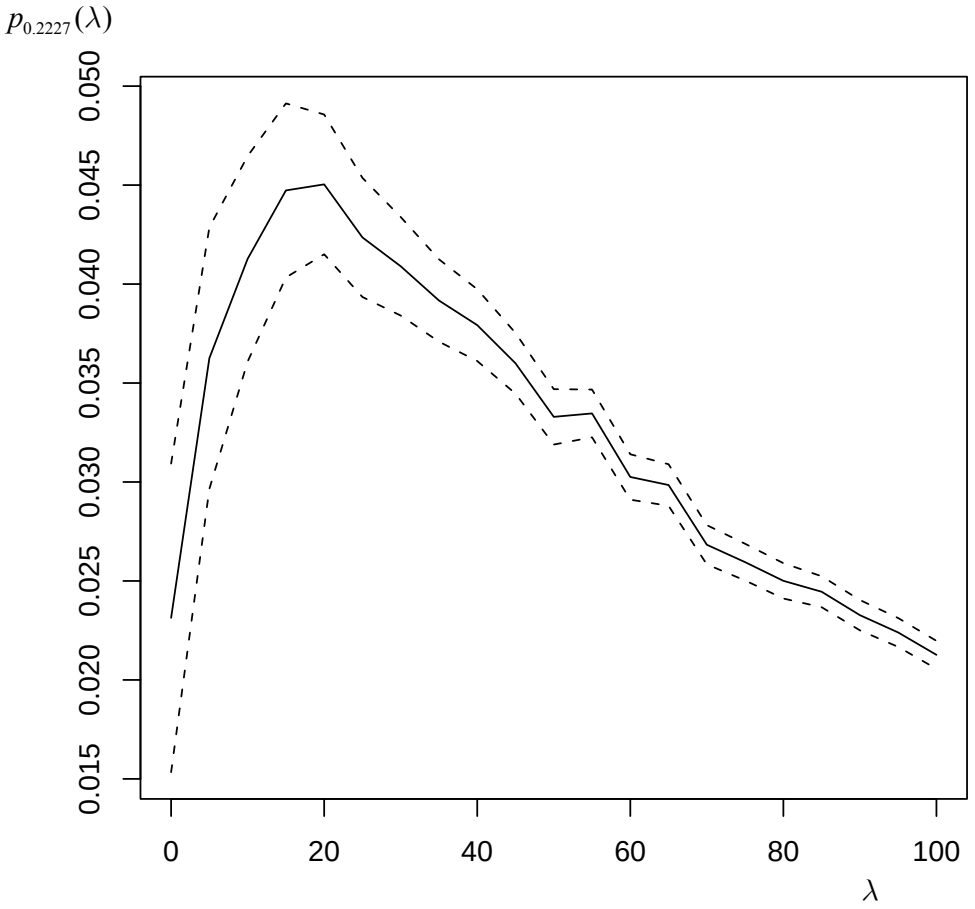
Example 4

Consider the situation in Example 3, but with σ_0 = 25 (rather than 5). The corresponding prior on μ is shown in Figure 10. We repeat the Monte Carlo search procedure, this time over λ = 0,5,10,...,100, and conclude on the basis of the results shown in Figure 12 that the optimal value of λ is about 20 (see Note 9). Applying the Newton Raphson algorithm defined by (23) and (24), with η = 50, δ = 0.2227, λ = 20, n = 4, $\bar{x}$ = 55 and s = 17, we find that the new 95% CI for μ is (31.2,78.0).

Observe that this interval is slightly wider than the maximal one, (32.6,77.4), and only 13.5% narrower than the ordinary one, (27.9, 82.1). This makes sense because a poorer performance is to be expected with a decrease in prior information. Actually, 13.5% is well above the 4.5% improvement which could on the basis of Figure 12 be expected when σ_0 = 25 and λ = 20. Note that with σ_0 = 25 and λ = 0, the expected decrease in interval length is only about 2.5% (see Note 9).

The fact that the *actual* improvement using $\lambda = 0$ is 17.1% (see Example 3) may be attributed to $\bar{x}$ happening *by chance* to fall very near μ's prior mean, $\mu_0 = \eta = 50$. When σ_0 is 25 (rather than 5), specifying λ as 20 (rather than 0) may be thought of as taking out *insurance* — at the cost of a slightly wider interval — against the possibility of a much wider interval resulting due to $\bar{x}$ being far from μ_0 (then a more likely outcome).

Figure 12 Monte Carlo estimates of the PEPDL when σ_0 = 25



6. Inference on the binomial proportion

We will now focus on the problem of estimating a binomial proportion p based on X successes in n Bernoulli trials. It will be shown how tail functions can be used to modify three favourite CI's for p, namely the 'standard', Wilson, and Clopper-Pearson intervals (see Brown *et al*., 2001), so as to reduce their prior expected lengths. In some cases the modified interval also exhibits better coverage properties.

The 'standard' interval

We first consider the most commonly used $1-\alpha$ CI for p, namely the *'standard' interval*. This may be written

$$(\overline{x} \pm z_{\alpha/2}\sigma/\sqrt{n}), \tag{26}$$

where $\overline{x} = x/n$, x is the observed value of X, $\sigma^2 = \overline{x}(1-\overline{x})$, and where bounds less than 0 or greater than 1 are taken as 0 or 1, respectively. This approximate CI is a consequence of the central limit theorem which, with $\overline{X} = X/n$, implies that

$$\frac{\overline{X}-p}{\sqrt{\overline{X}(1-\overline{X})/n}} \xrightarrow{d} N(0,1) \quad \text{as } n \longrightarrow \infty. \tag{27}$$

The lower and upper bounds of (26) are obtained as the values of p for which $\Phi((\overline{x}-p)\sqrt{n}/\sigma)$ equals $1-\alpha/2$ and $\alpha/2$, respectively.

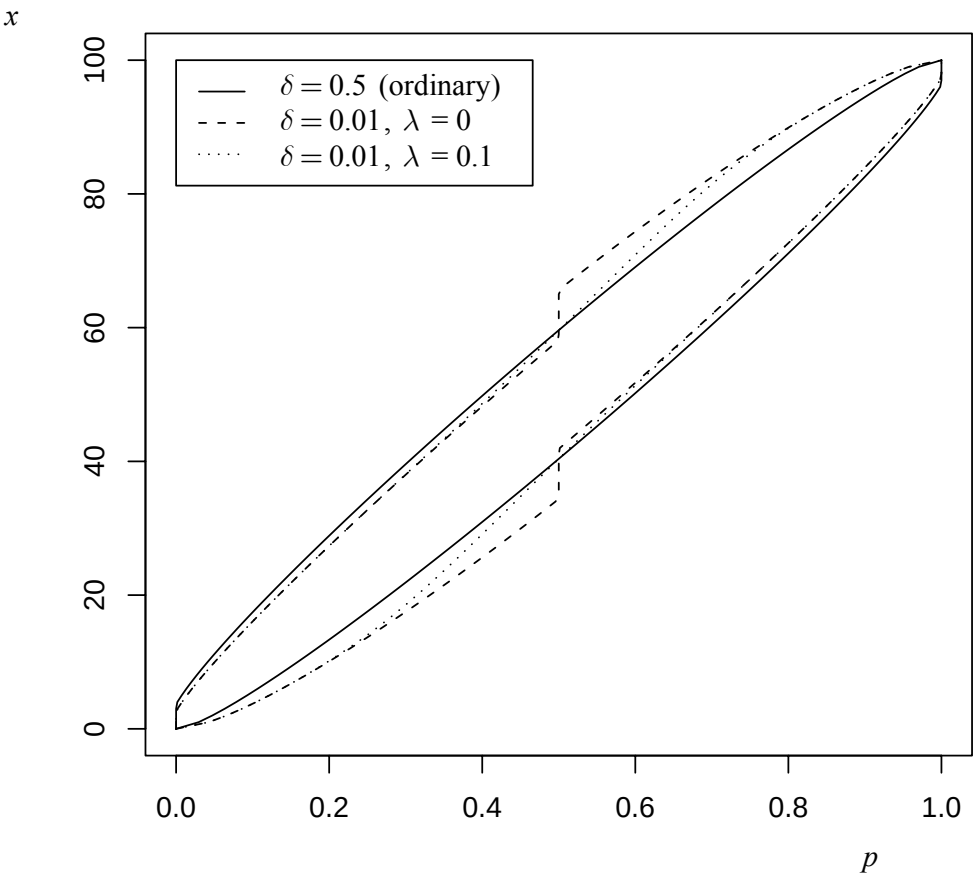
We see that this scenario is *identical* to the one in Section 4, with $\mu = p$. Thus if prior information regarding p is available, we may replace $1-\alpha/2$ and $\alpha/2$ in the last paragraph with $1-\alpha+\alpha\tau(p)$ and $\alpha\tau(p)$, where $\tau(p)$ is given by (12) and suitable choices of η, δ and λ. Note that $\sigma^2 = 0$ if $x = 0$ or n, in which case we define the modified standard CI as the single point 0 or 1, respectively (*i.e.* we make no modification to the ordinary standard CI).

Example 5

Suppose that we are about to conduct a binomial experiment with $n = 100$ trials and feel confident that p lies near 0.5. We consider using a *Gaussian standard* 95% CI with parameters $\eta = 0.5$ and $\delta = 0.01$. Figure 13 shows the bounds of this CI for all

possible outcomes, x = 0,1...,100, both when $\lambda = 0$ and when $\lambda = 0.1$. Also shown are the bounds of the *ordinary standard* 95% CI ($\delta = 0.5$).

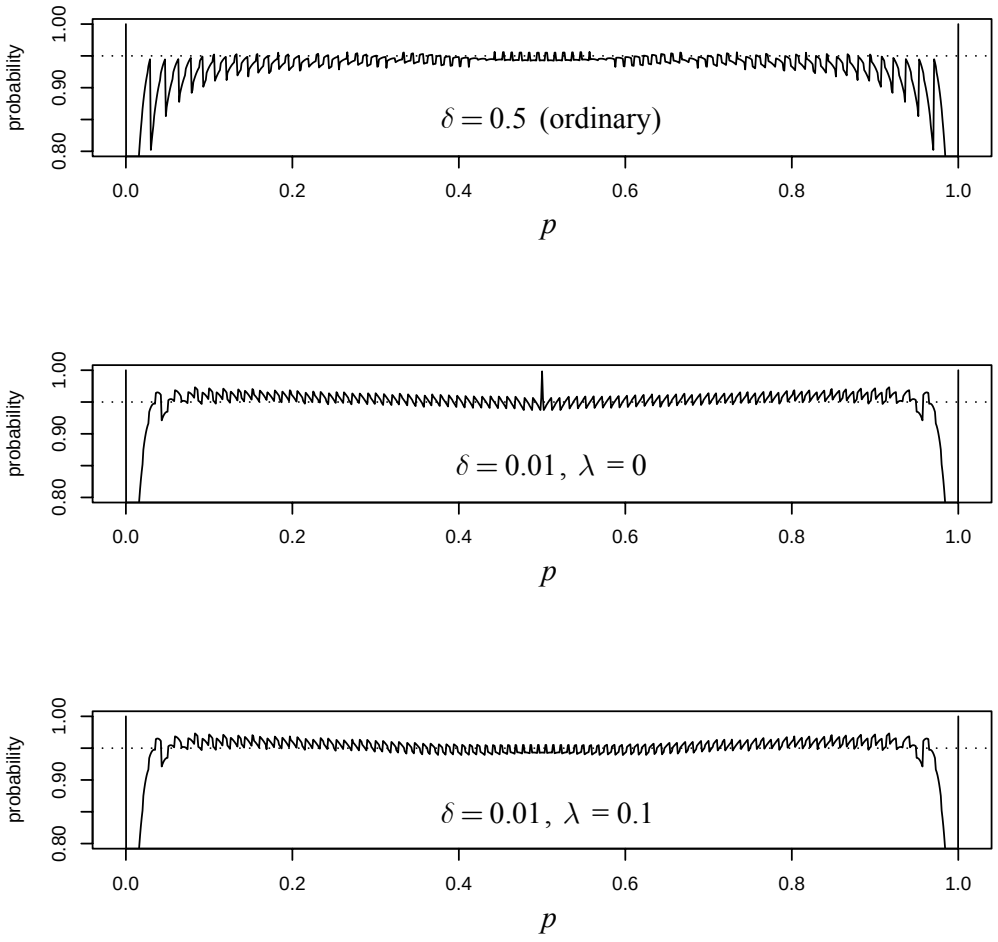
Figure 13 Bounds of three standard 95% CI's when $n = 100$ and $\eta = 0.5$



Observe in Figure 13 that the two modified CI's are shorter than the ordinary CI when x is close to 50, corresponding to p being close to $\eta = 0.5$. This effect is 'paid for' by the modified CI's being longer for other values of x.

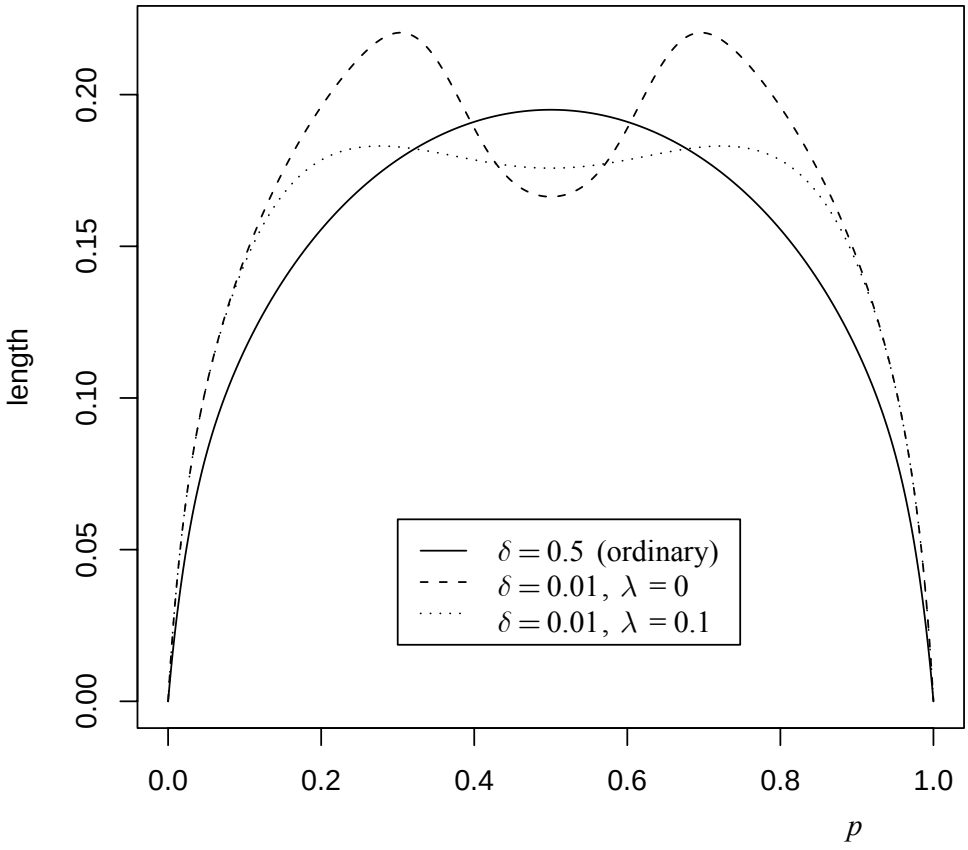
Let us now compare the coverage probabilities of the three CI's. These are illustrated in Figure 14, and it is interesting that both of the modified 95% CI's have coverage properties which are distinctly better than those of the standard 95% CI.

Figure 14 Coverage probabilities of three standard 95% CI's when *n* = 100 and η = 0.5



Let us now compare expected lengths. In Figure 15 we see that the maximal CI (defined by λ = 0) has lengths which are *smaller* on average than those of the ordinary CI for values of *p* between about 0.4 and 0.6, and *larger* for values outside that range. The CI defined by λ = 0.1 provides a greater range for which there is improvement over the ordinary interval (roughly 0.3-0.7), and also a smaller maximum expected length (0.18 rather than 0.22). However, these improvements come at the expense of it being slightly longer than the maximal CI for *p* very near 0.5.

Figure 15 Expected lengths of three standard 95% CI's when n = 100 and η = 0.5



Our final decision regarding which interval to use will depend on the strength of our prior beliefs. For example, if we believe very strongly that p is between 0.45 and 0.55 then a good choice is the maximal CI defined by (η, δ, λ) = (0.5,0.01,0). If we believe that p might be anywhere between 0.3 and 0.7 then a 'safer' choice is (η, δ, λ) = (0.5,0.01,0.1). If we believe that p could be anywhere from 0.2 to 0.8, then perhaps it is best to stick with the ordinary standard interval (δ = 0.5), or else investigate other values of δ and λ.

The Wilson interval

An interval with better coverage properties than the standard interval, especially when *n* is small, is the *Wilson interval* (Wilson, 1927), whose bounds are given by:

$$(l_a, u_a) = \left(\frac{\overline{x} + \frac{z_a^2}{2n} - z_a\sqrt{\frac{\overline{x}(1-\overline{x})}{n} + \frac{z_a^2}{4n^2}}}{1+\frac{z_a^2}{n}}, \frac{\overline{x} + \frac{z_a^2}{2n} + z_a\sqrt{\frac{\overline{x}(1-\overline{x})}{n} + \frac{z_a^2}{4n^2}}}{1+\frac{z_a^2}{n}} \right), \qquad (28)$$

where $a = \alpha/2$. This interval is based on the result

$$\frac{\overline{X} - p}{\sqrt{p(1-p)/n}} \xrightarrow{d} N(0,1) \quad \text{as } n \longrightarrow \infty, \qquad (29)$$

where the rate of convergence is faster than in (27). It can be shown that if we apply the tail function (12) with $\lambda = 0$, then the resulting modified $1-\alpha$ Wilson CI is:

$$\begin{array}{lll} (l_{\alpha\delta}, & u_{\alpha(1-\delta)}), & \overline{x} > \eta + d \\ (\eta, & u_{\alpha(1-\delta)}), & \eta + c < \overline{x} \le \eta + d \\ (l_{\alpha(1-\delta)}, & u_{\alpha(1-\delta)}), & \eta - c \le \overline{x} \le \eta + c \\ (l_{\alpha(1-\delta)}, & \eta), & \eta - d \le \overline{x} < \eta + c \\ (l_{\alpha(1-\delta)}, & u_{\alpha\delta}), & \overline{x} < \eta - d, \end{array} \qquad (30)$$

where $c = z_{\alpha(1-\delta)}\sqrt{\eta(1-\eta)/n}$ and $d = z_{\alpha\delta}\sqrt{\eta(1-\eta)/n}$. Note that if δ = 1/2 then the *maximal Wilson* CI (30) reduces to the *ordinary Wilson* CI (28), which in turn reduces to the *ordinary standard* CI (26) in the limit as *n* tends to infinity.

If $\lambda > 0$ then the modified Wilson CI can be obtained *via* the Newton-Raphson algorithm as follows. The upper bound may be found by choosing a suitable starting value p_0 and iterating $p_{j+1} = p_j - \psi(p_j)/\psi'(p_j)$ until convergence, where:

$$\psi(p) = \Phi(g(x,p)) - \alpha\tau(p) \qquad (31)$$

$$\psi'(p) = \phi(g(x,p))\frac{\partial}{\partial p}g(x,p) - \alpha\tau'(p) \qquad (32)$$

$$g(x,p) = \frac{\overline{x} - p}{\sqrt{p(1-p)/n}} \quad \text{and} \quad \frac{\partial}{\partial p}g(x,p) = \frac{\overline{x}(p-1/2) - p/2}{\sqrt{p^3(1-p)^3/n}}. \qquad (33)$$

The lower bound can be found in the same way but with $\alpha\tau(p)$ in (31) replaced by $1-\alpha+\alpha\tau(p)$, and with (32) unchanged. Note that this Newton-Raphson algorithm can also be used to find the modified standard CI after replacing (33) with $g(x,p)=(\overline{x}-p)/\sqrt{\overline{x}(1-\overline{x})/n}$ and $\partial g(x,p)/\partial p=-1/\sqrt{\overline{x}(1-\overline{x})/n}$.

Example 6

Suppose that we are about to conduct a binomial experiment with n = 25 trials and feel confident that p lies near 0.3. Since n is small we consider using a modified Wilson 95% CI with parameters η = 0.3 and δ = 0.03, where λ is either 0 or 0.15.

Figures 16-18 provide the same kinds of information as Figures 13-15. In Figure 18 we see that when p is near 0.3, the maximal Wilson CI (defined by λ = 0) provides greater gains in expected length over the standard Wilson CI than the one defined by λ = 0.15, and that this comes at the expense of it being more 'risky' overall. Note that the coverage probabilities of the two modified intervals in Figure 17 are asymmetric, unlike in Figure 14. This is because η equals 0.3 and not 0.5.

Figure 16 Bounds of three Wilson 95% CI's when $n = 25$ and $\eta = 0.3$

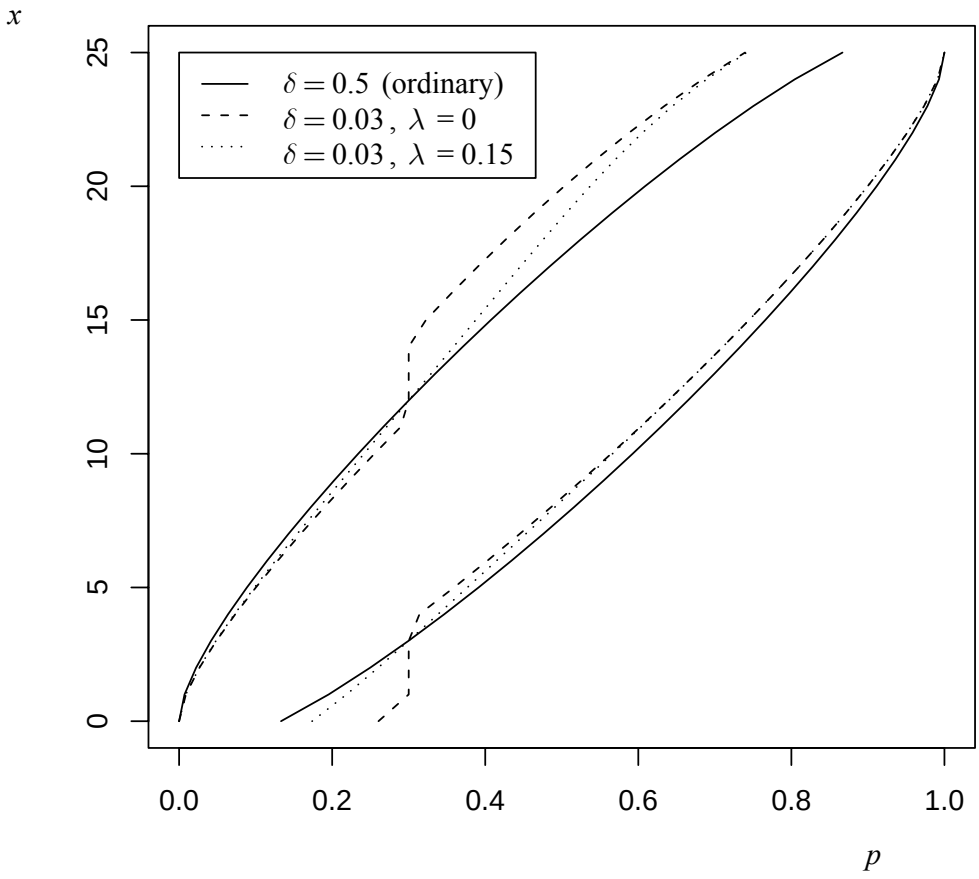


Figure 17 Coverage probabilities of three Wilson 95% CI's when $n = 25$ and $\eta = 0.3$

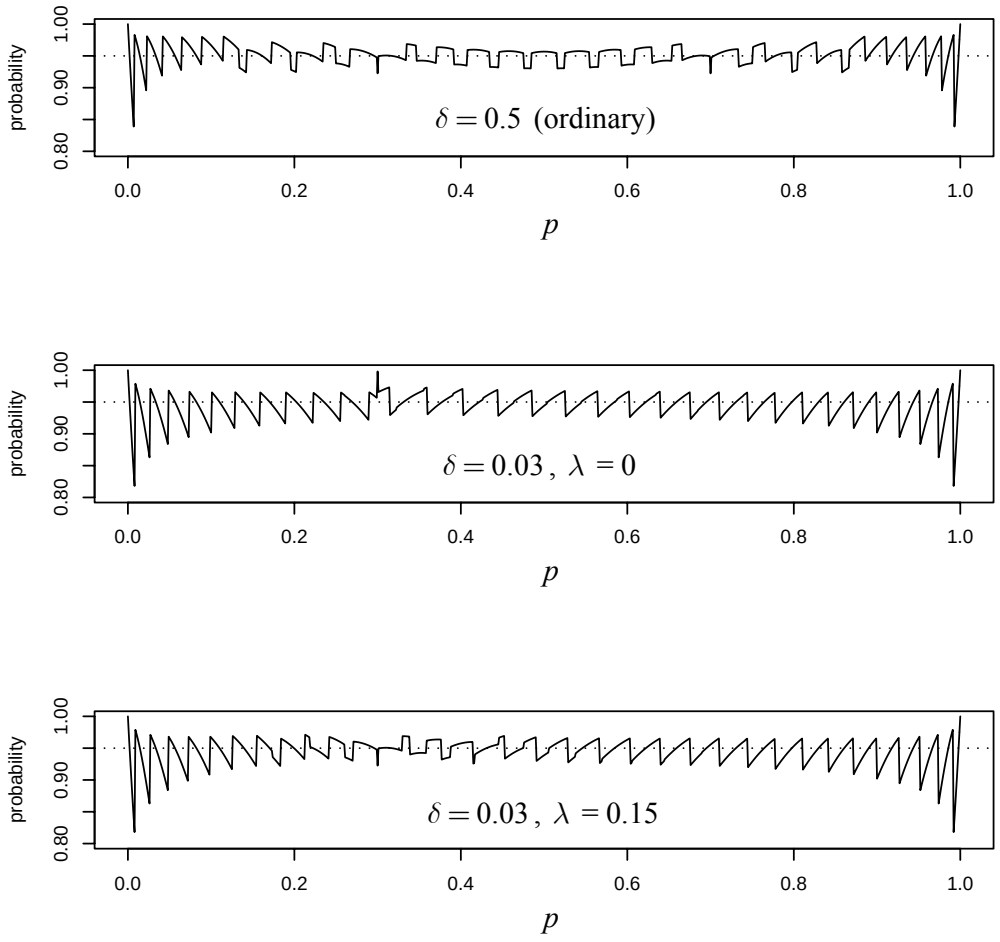
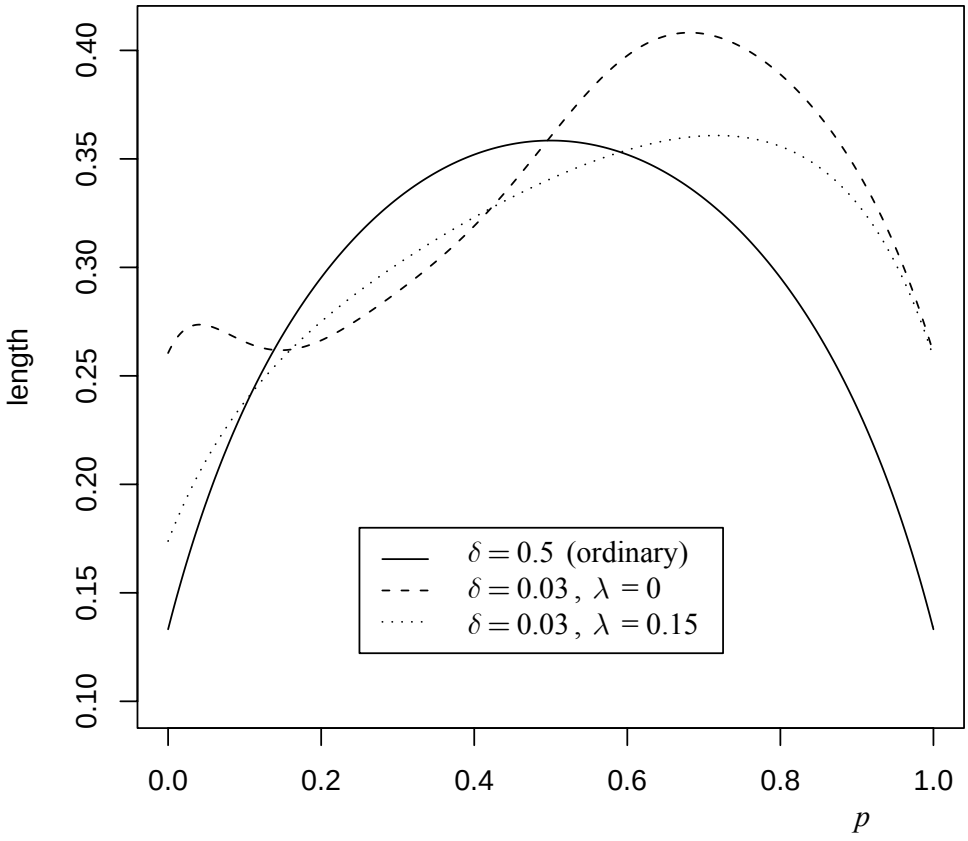


Figure 18 Expected lengths of three Wilson 95% CI's when $n = 25$ and $\eta = 0.3$



The Clopper-Pearson interval

As is evident from Figures 14 and 17, the coverage probabilities of the standard and Wilson intervals are only *approximately* equal to the desired level, $1-\alpha$. One CI which contains p with probability *at least* $1-\alpha$ is the so-called 'exact' *Clopper-Pearson* (CP) CI, after Clopper and Pearson (1934). This interval may be written $(L(a),U(a))$ where, with $a=\alpha/2$, $L(a)$ and $U(a)$ are the solutions in p of the equations:

$$P(X \geq x; p) = a$$
$$P(X \leq x; p) = a,$$

respectively, except that $L(a) = 0$ if $x = 0$ and $U(a) = 1$ if $x = n$. It can be shown that $L(a) = Beta(a; x, n-x+1)$ and $U(a) = Beta(1-a; x+1, n-x)$, where $Beta(q; r, s)$ denotes the (lower) q quantile of the beta distribution with parameters r and s.

A problem with the CP CI is that it tends to be wider than necessary, and recently Blaker (2000) has modified it so as to produce a new 'gold standard' that is uniformly shorter on average whilst still being 'exact'. Moreover, Blaker's interval has an attractive nesting condition which is absent from some earlier 'exact' CI's, for example those developed by Blyth and Still (1983) and Casella (1986).

The tail function methodology can also be applied in this context. Given a tail function $\tau(p)$, the *modified Clopper-Pearson* CI is defined as (l,u), where l and u are the solutions in p of the equations:

$$P(X \geq x; p) = \alpha(1-\tau(p))$$
$$P(X \leq x; p) = \alpha\tau(p), \qquad (34)$$

respectively, except that once again $l = 0$ if $x = 0$ and $u = 1$ if $x = n$. These equations can be solved *via* the Newton-Raphson algorithm in the usual manner, after noting that $P(X \leq x; p) = \sum_{t=0}^{x} f_X(t; p)$ where $f_X(t; p) = C_t^n p^t (1-p)^{n-t}$. The calculations can also be facilitated by noting certain relationships between the binomial and F distributions, and by interpreting CP CI's, both ordinary and modified, as a natural byproduct of randomised confidence interval theory. For more details see Puza and O'Neill (2004).

An important special case is where the tail function is a *step function*, in which case the bounds of the modified CP CI can be written in terms of the ordinary CP CI. For example, suppose that the tail function is Gaussian (12) with $\lambda = 0$. Then the bounds of the associated 'maximal' CI are:

$$l = \begin{cases} L(\alpha(1-\delta)) & \text{if } L(\alpha(1-\delta)) \le \eta \\ L(\alpha\delta) & \text{if } L(\alpha\delta) \ge \eta \\ \eta & \text{otherwise} \end{cases}$$

$$u = \begin{cases} U(\alpha(1-\delta)) & \text{if } U(\alpha(1-\delta)) \ge \eta \\ U(\alpha\delta) & \text{if } U(\alpha\delta) \le \eta \\ \eta & \text{otherwise} \end{cases} \qquad (35)$$

These equations can easily be generalized to accommodate any number of steps in the tail function. For an example, see Note 10.

Example 7

Suppose that we are about to conduct a binomial experiment with only $n = 10$ trials, where the goal is an 'exact' 95% CI for p, whose value is believed to be small and certainly less than 0.4. We consider five candidates: the standard CP CI, the Blaker CI, the one-sided CP CI $(0, U_\alpha)$, and the modified CP CI's defined by (η, δ, λ) = (0.2,0,0) and (0.25,0,0). Note that the one-sided CI can also be thought of as a modified CP CI defined by (η, δ, λ) = (0,0,0), or equivalently, by $\tau(p) = 1$.

Figures 19-21 show the same diagnostics as in previous examples, but for the above five CI's. We see in Figure 21 that in terms of expected length, the modified CI defined by η = 0.25 is clearly the best of the five intervals for all p less than 0.4. In particular, it is everywhere narrower on average than the one-sided CI, except at p = 0, where the two intervals have exactly the same expected length. For further discussion see Note 11.

Figure 19 Bounds of five 'exact' 95% CI's when $n = 10$ and $\delta = \lambda = 0$

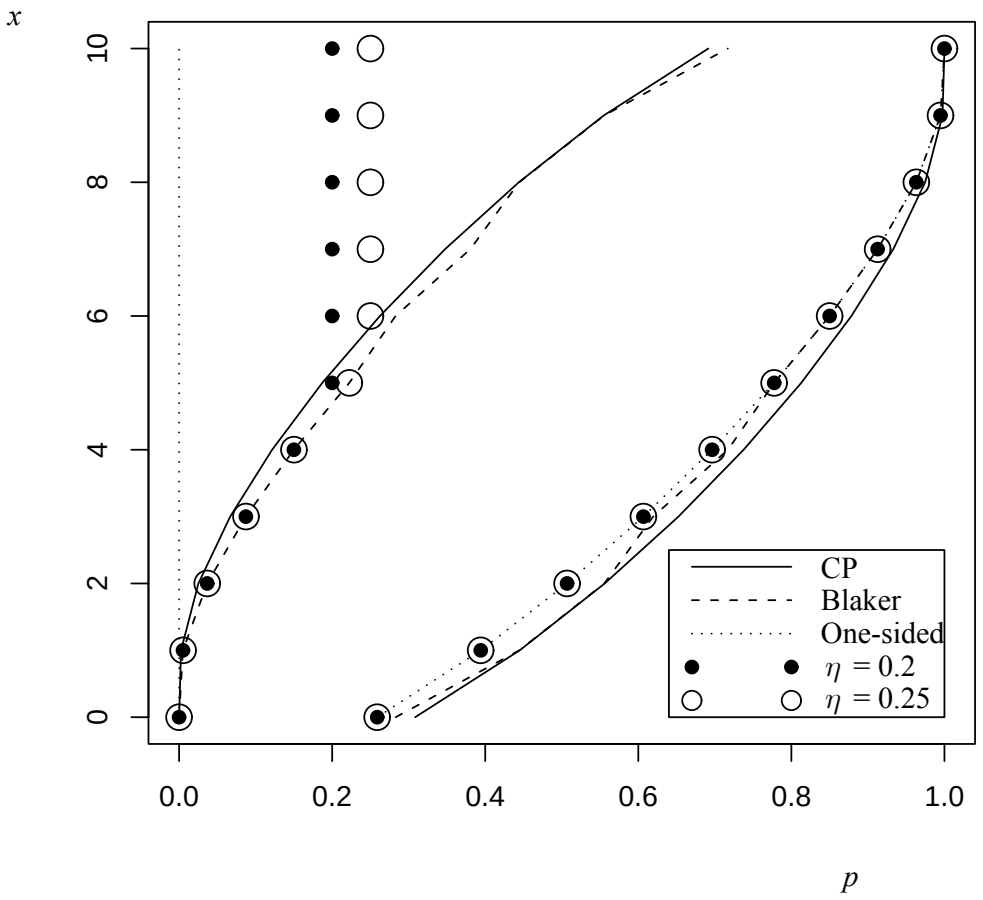
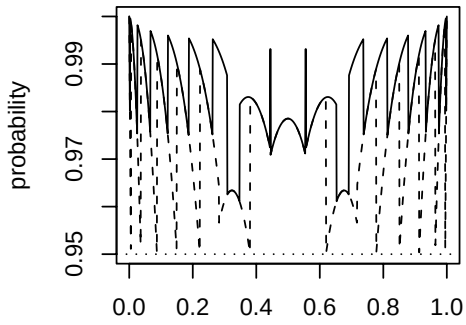
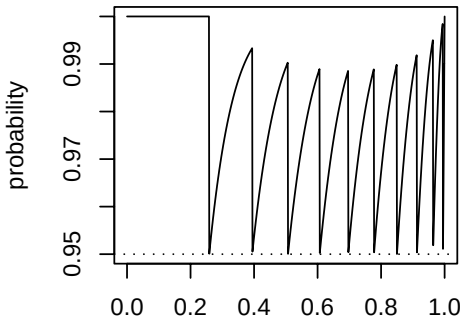


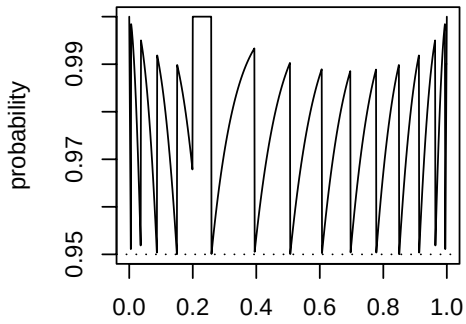
Figure 20 Coverage probabilities of five 'exact' 95% CI's when $n = 10$ and $\delta = \lambda = 0$



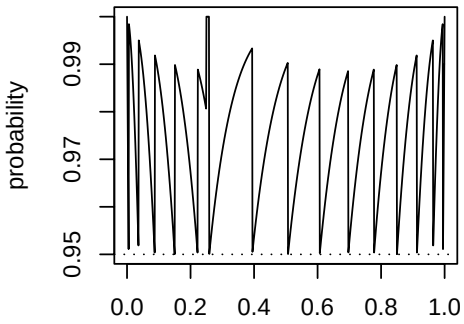
Ordinary CP (solid lines) and Blaker (dashed lines)



One-sided CP ($\eta = 0$)

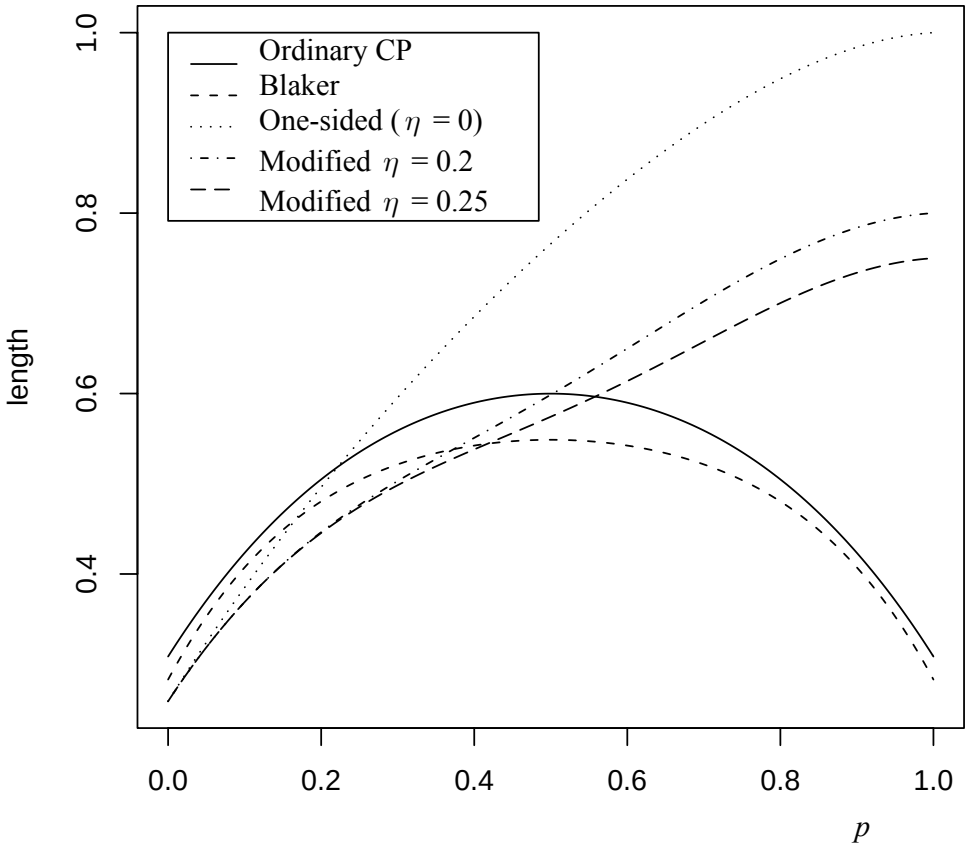


Modified CP $\eta = 0.2$



Modified CP $\eta = 0.25$

Figure 21 Expected lengths of five 'exact' 95% CI's when $n = 10$ and $\delta = \lambda = 0$



7. Summary and discussion

In this paper we have shown how confidence intervals can be constructed using *tail functions*. Classical methods may be thought of as a special case of the new theory, with all tail functions taken as *constant*. Focussing on the normal mean and binomial proportion, we have provided several examples of how a suitable choice of tail function can lead to a CI which is shorter in expectation than its 'ordinary' or classical counterpart. Such improvements rely on prior information being available and are most effective when that information is strong.

The tail function methodology provides an attractive frequentist alternative to the Bayesian approach when it is desired that the coverage probability be correct for all values of the target parameter. Also, it has the advantage of not requiring prior information to be expressed in an exact way. For example, if we are about to conduct a binomial experiment with 10 trials and know only that the binomial proportion is less than 0.4 then a good tail function can be found without any additional information (see Example 7). In contrast, the Bayesian approach requires a specific prior distribution, such as the beta, in order to proceed.

A useful observation is that a modified CI tends to be relatively short on average for values of the target parameter at which the tail function is *steep*. This fact provides a guide to choosing the optimal tail function in any situation. One important special case is where the tail function consists of a single step. The implied CI can often be calculated easily, without the Newton-Raphson algorithm (see (20), (30) and (35)). Moreover, this case allows for the maximum possible reduction in width and is a good choice in many situations. This suggests that we give special consideration to tail functions which consist only of steps (see Note 10)). We leave these and other tail functions as a topic for future research.

It should be kept in mind that there is a price to be paid for a CI with smaller expected width, namely the risk that it will be *longer* if the data happens to be extreme or the prior has been misspecified. Thus the proposed methodology involves an element of gambling. Also, the choice of tail function should not be allowed to depend on the observed data. If it does, then any reduction in interval width may be illusory due to a deflated coverage effect (see Note 12). Ideally, the exact form of the CI should be specified *before* the data have been observed. Note that this advice applies equally in any situation where there are two or more ways to construct an interval estimate.

Notes

1. Each of the two pairs of lines in Figure 2 also satisfy the following equations:

$$\bar{x} = \mu + \Phi^{-1}(1-\alpha+\alpha\tau(\mu))\sigma/\sqrt{n} \quad \text{(upper line)}$$
$$\bar{x} = \mu + \Phi^{-1}(\alpha\tau(\mu))\sigma/\sqrt{n} \quad \text{(lower line).}$$

The continuous lines correspond to $\tau(\mu) = 1/2$, and the dashed ones to (12) with η = 50, δ = 1/8 and λ = 10. This simple way to graph interval bounds follows from (7), (8) and (9) after putting $\theta = \mu$, $g(x,\theta) = y = (\bar{x}-\mu)\sqrt{n}/\sigma$ and $F_Y(y) = \Phi(y)$.

2. The equation $I_\delta = I$ can be solved via Newton-Raphson as follows. Let $h(\delta) = I_\delta - I$ and $h'(\delta) = \alpha(z'_{\alpha\delta} - z'_{\alpha(1-\delta)})/(2z_{\alpha/2})$, where $z'_t = dz_t/dt = -1/\phi(z_t)$. Then choose a starting value δ_0 and repeatedly calculate $\delta_{j+1} = \delta_j - h(\delta_j)/h'(\delta_j)$ until convergence.

3. Except for values of λ near the origin, the MPDL's in Figure 6 are proportionate to $\sigma/\sqrt{n}$. For example, if (n,σ) = (1,1) is changed to either (4,1) or (1,0.5) then $M_{0.1,2}$ = 0.092 changes to 0.056. At the same time, each GMPDL (*e.g.* $G_{0.0871} = 0.138$) stays the same, and the 'plateaus' near the origin become shorter. In general, $M_{\delta,\lambda} \longrightarrow G_\delta I(\lambda = 0)$ as $\sigma \longrightarrow 0$ or $n \longrightarrow \infty$ (for all δ), and $M_{\delta,\lambda} \longrightarrow G_\delta$ as $\sigma \longrightarrow \infty$ (for all δ and λ).

4. One way in which the prior $\mu \sim N(\mu_0, \sigma_0^2)$ might arise is if a random sample of n_0 values from the $N(\mu,\sigma^2)$ distributions were previously observed and their average was $\bar{x}_0$. Using a flat 'pre-prior' for μ and standard Bayesian calculations (see Lee, 1989), we may then specify $\mu_0 = \bar{x}_0$ and $\sigma_0^2 = \sigma^2/n_0$.

5. An equivalent definition for the PEPDL at (21) is $p_\delta(\lambda) = \int_{-\infty}^{\infty} D_{\delta,\lambda}(\bar{x}) f(\bar{x})\, d\bar{x}$, where $f(\bar{x}) = \phi((\bar{x}-\mu_0)/\sigma_\#)/\sigma_\#$ and $\sigma_\#^2 = \sigma_0^2 + \sigma^2/n$. This is because $\bar{X} \sim N(\mu_0, \sigma_\#^2)$. We may also write $p_\delta(\lambda) = Ee_{\delta,\lambda}(\mu) = \int_{-\infty}^{\infty} e_{\delta,\lambda}(\mu) f(\mu) d\mu$, where $e_{\delta,\lambda}(\mu)$ is the EPDL at (17).

6. In contrast to P_μ at (22), the *unconditional* probability of μ's HPDR containing μ is exactly 95%. This can be shown by integrating $P_\mu f(\mu)$ over the whole real line, but is most easily proved as follows:

$$P(\mu \in H(\bar{X})) = EP(\mu \in H(\bar{X}) \mid \bar{X}) = E(1-\alpha) = 1-\alpha.$$

7. In (25), $f(\bar{x},s\,|\,\mu,\sigma)=f(\bar{x}\,|\,\mu,\sigma)f(s\,|\,\sigma)$,
where: $f(\bar{x}\,|\,\mu,\sigma)=\phi((\bar{x}-\mu)\sqrt{n}/\sigma)\sqrt{n}/\sigma$
$f(s\,|\,\sigma)=f(u\,|\,\sigma)du/ds$
$u=(n-1)s^2/\sigma^2,\quad du/ds=2(n-1)s/\sigma^2$
$f(u\,|\,\sigma)=u^{(1/2)(n-1)-1}e^{-u/2}/\{2^{(1/2)(n-1)}\Gamma((n-1)/2)\}$.
This follows because $\bar{X}$ and S^2 are conditionally independent, because $(\bar{X}\,|\,\mu,\sigma)\sim N(\mu,\sigma^2/n)$, and because $((n-1)S^2/\sigma^2\,|\,\mu,\sigma)\sim\chi^2(n-1)$.

8. Using numerical techniques to evaluate the integral at (25) exactly, we find that the PEPDL in Figure 11 equals 0.144 at λ = 0 and has a maximum of 0.145 at λ = 0.9.

9. Using numerical techniques to evaluate the integral at (25) exactly, we find that the PEPDL in Figure 12 equals 0.022 at λ = 0. It also equals 0.043 at λ = 20 and has a maximum of 0.043 at λ = 19.1 (same to three decimals).

10. Suppose that the tail function is

$$\tau(p)=\begin{cases}\delta_1, & 0\le p<\eta_1\\ \delta_2, & \eta_1\le p<\eta_2\\ \delta_3, & \eta_2\le p\le 1\end{cases}$$

where $0\le\eta_1\le\eta_2\le 1$ and $0\le\delta_1\le\delta_2\le\delta_3\le 1$. Then the bounds of the associated $1-\alpha$ Clopper-Pearson CI are:

$$l=\begin{cases}L(\alpha(1-\delta_1)) & \text{if } L(\alpha(1-\delta_1))\le\eta_1\\ \eta_1 & \text{if } L(\alpha(1-\delta_1))>\eta_1 \text{ and } L(\alpha(1-\delta_2))\le\eta_1\\ L(\alpha(1-\delta_2)) & \text{if } \eta_1<L(\alpha(1-\delta_2))\le\eta_2\\ \eta_2 & \text{if } L(\alpha(1-\delta_2))>\eta_2 \text{ and } L(\alpha(1-\delta_3))\le\eta_2\\ L(\alpha(1-\delta_3)) & \text{if } L(\alpha(1-\delta_3))>\eta_2\end{cases}$$

$$u=\begin{cases}U(\alpha\delta_3) & \text{if } U(\alpha\delta_3)\ge\eta_2\\ \eta_2 & \text{if } U(\alpha\delta_3)<\eta_2 \text{ and } U(\alpha\delta_2)\ge\eta_2\\ U(\alpha\delta_2) & \text{if } \eta_1\le U(\alpha\delta_2)<\eta_2\\ \eta_1 & \text{if } U(\alpha\delta_2)<\eta_1 \text{ and } U(\alpha\delta_1)\ge\eta_1\\ U(\alpha\delta_1) & \text{if } U(\alpha\delta_1)<\eta_1\end{cases}$$

11. This relationship also holds between the one-sided CI and *all* modified CP CI's defined by $\delta = \lambda = 0$ and $0 < \eta \leq 0.2589$. For $\eta > 0.2589$ the modified CI is wider in expectation for values of p in a neighbourhood of zero whose size increases as η increases. The reason for this is that the one-sided CI at $x = 0$ is (0,0.2589), which follows from (34), whereby $u = 1 - 0.05^{1/10} = 0.2589$. Thus any increase in η past 0.2589 has the effect of making u at (35) also increase to η when $x = 0$ (see Figure 19). The result is that the modified CI is wider than the one-sided CI when $x = 0$ and so has a larger prior expected width if p is sufficiently small.

12. For example, the Gaussian tail function (12) with $\delta = \lambda = 0$ and $\eta = \bar{x}$ implies an automatic reduction in expected length (relative to the ordinary 95% CI for μ in Section 4) of 16.1% (the AGMPDL at (19) when $\alpha = 0.05$). However, by (20) it also implies an *actual* coverage probability of only 90%.

References

Blaker, H. (2000). Confidence curves and improved exact confidence intervals for discrete distributions. *The Canadian Journal of Statistics*, **28**, 783-798.

Blaker, H. (2001). Corrigenda: Confidence curves and improved exact confidence intervals for discrete distributions. *The Canadian Journal of Statistics*, **29**, 681.

Blyth, C. R. and Still, H. A. (1983). Binomial confidence intervals. *Journal of the American Statistical Association*, **78**, 108-116.

Brown, D. B., Cai, T. T. and DasGupta, A. (2001). Interval estimation for a binomial proportion. *Statistical Science*, **16**, 101-133 (with discussion).

Casella, G. (1986). Refining binomial confidence intervals. *The Canadian Journal of Statistics*, **14**, 113-129.

Clopper, C. J. and Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, **26**, 404-413.

Lee, P.M., 1989. Bayesian Statistics: An Introduction. Oxford University Press, New York.

Puza, B. D, and O'Neill, T. J. (2004). Generalised Clopper-Pearson confidence intervals for the binomial proportion. Under review by the *Journal of Statistical Computation and Simulation.*

Stuart, A., Ord, J. K., and Arnold, S. (1999). *Kendall's Advanced Theory of Statistics, Volume 2A*. London: Arnold.

Venables, W. N., and Ripley, B. D. (1999). *Modern Applied Statistics with S-Plus*, 3rd ed. New York: Springer.

Venables, W. N., Smith, D. M., and the *R* Development Core Team (2004). *An Introduction to R.* ISBN: 3-900051-05-4. Website: http://www.r-project.org

Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, **22**, 209-212.