Selection Adjusted Confidence Intervals with More Power to Determine the Sign

September 15, 2011

Yoav Benjamini[1] and Asaf Weinstein[2]

Department of Statistics and Operations Research, The Sackler Faculty of Exact Sciences, Tel Aviv University

Email: [1] ybenja@tau.ac.il [2]asafw@wharton.upenn.edu

Abstract

Statistical inference on multiple parameters often involves a preliminary stage of selection. Such situations may arise when θ is of interest only if its estimator X is large enough in either direction, or when there is a physical limitation for observing X within $\pm c$. Since the selection procedure is based on the primary data, it introduces a conditional distribution of the data at hand, which should be accounted for when making subsequent inferences. In this paper we suggest three different methods to construct confidence intervals for a location parameter while conditioning on its estimator being greater in absolute value than some constant threshold. By way of definition they all offer False Coverage-statement Rate control for parameters selected from a larger pool in the above way.

Two of these methods are the primary focus of this paper; Relying on the principle set forth by Benjamini, Hochberg and Stark (1998), they offer not only valid confidence intervals in face of selection, but are also short and at the same time make early sign determinations - they stop including parameter values of opposite signs for relatively small values of $|X|$. Following a simulation study we find that one of the two, the Conditional Quasi-Conventional confidence interval, offers a good balance between length and sign determination while protecting from the effect of selection.

Keywords: Selective Inference, False Coverage Rate, Conditional Confidence Intervals

1 Introduction

Throughout this paper, let $Y = \theta + Z$, where the density of the random variable Z is unimodal and symmetric about 0, and assume that we are interested in the value of the parameter only if $|Y|$ is big enough say bigger than c. Alternatively, we only observe $X = (Y||Y| \geq c)$. This conditional distribution depends on $\theta = E(Y)$, and a confidence interval should be constructed for θ upon observing X. Although the results of this paper are applicable to any such distribution, we will assume that $Z \sim N(0, 1)$ and will denote its density, as usual, by $\phi(z)$, and its cumulative distribution function by $\Phi(z)$. Then,

$$f_X(x;\theta) = \begin{cases} (1-\Phi(c-\theta)+1-\Phi(c+\theta))^{-1}\phi(x-\theta), & |x| \geq c \\ 0, & otherwise \end{cases} . \qquad (1)$$

In other words, the density of the observed random variable X is zero on $(-c, c)$, and is proportional to that of Y elsewhere. It is important to notice that the conditioning on passing a constant threshold materially alters the role of the parameter θ (at least for small values it takes): while it is a mere location parameter for Y (ie. $Y-\theta$ has a distribution invariant of θ), it reflects both the location and shape for X.

In this paper we suggest three different procedures for constructing two-sided confidence intervals for $\theta = E(Y)$. All three utilize the general duality between a family of α level tests of the hypothesis $E(U) = \omega$ and a $(1-\alpha)$ level confidence procedure (Lehmann, 1986), by first constructing a family of $1-\alpha$ acceptance regions $\{A(\omega) : \omega \in \Omega\}$, and then inverting them to obtain a $1-\alpha$ confidence set $S(U) := \{\omega : U \in A(\omega)\}$.

The first procedure is rather straightforward, and in some sense might be considered the equivalent of the conventional confidence interval in the unconditional normal case. The other two follow the principle set up by Benjamini, Hochberg and Stark (1998, hereafter BH&S) whereby a confidence interval can serve the dual goal of (i) bounding the parameter within a short interval while (ii) avoiding parameters of opposite signs.

Property (ii) is referred to as offering weak sign determination. Both opponents and even some proponents of hypothesis testing agree that a null hypothesis such as $\theta = 0$ is an ideal never to be found in practice (Pratt, 1961, Tukey, 1991, BH&S, 1998). Therefore, a confidence interval that includes 0 but no negative values has weakly determined the sign to be nonnegative, and it is an error only if $\theta < 0$. We too adopt this point of view.
In the conditional approach taken here, the strategy remains the same as the one applied by BH&S in the symmetric, unimodal case. However, the loss of symmetry and the different role of θ in the conditional case, in that it is no longer a mere shift parameter, introduce complications in adopting the original constructions.

The proposed confidence intervals answer another rarely addressed yet important concern inherent in many current large problems. In these problems, confidence intervals are made on a limited set of parameters of interest which are selected from the much larger pool of potential parameters based on the value of their estimators. Some examples are: confidence intervals constructed for associations of genetic markers with disease, only for the markers with p-values below a specified threshold; genes are chosen based on whether their expression level differ between two groups by more than 3-fold; risk factors for disease are chosen based on significance, and then their effect is estimated; regions in the brain that are highly correlated with a response are chosen and an estimator of the response is given only for the selected ones. In these situations, and in many others, a parameter is selected only if its associated estimator exceeds (in its absolute value) some constant threshold or, similarly, if the corresponding p-value is small enough.

The ongoing practice is to construct the usual (marginally correct) 95% confidence intervals. Such a practice is defended by arguing that these marginal confidence intervals offer "coverage on the average" over all parameters estimated, and since there is no interest in simultaneous inference their level of coverage suffices: we are accustomed to the 5% level of error, so here too we can bear the fact that 5% of the confidence intervals will not cover their respective parameters (see e.g. Gelman and Hill, 2009).

Benjamini and Yekutieli (2005) addressed this issue in generality. They argue that usually in large multiple inference problems a selection process is taking place, either formally via testing or informally via highlighting in the abstract, and interest lies only in few selected parameters out of the many estimated. They demonstrate that when the coverage of the regular 95% confidence intervals is considered only over the selected parameters, even if on the average and not simultaneously, the non-coverage can deteriorate to be much more than 5%. So even if there is no interest in the strong protection from non-coverage error offered by simultaneous intervals, one should still worry about "coverage on the average over the selected ones" - a coverage property which is not guaranteed with the ongoing practice. Formally, they introduced the False Coverage-statement Rate (FCR),

$$\mathrm{FCR} = E\left(\frac{\#\text{ non-covering intervals}}{\#\text{ intervals selected}}\right),$$

where the ratio is 0 if no interval is selected. They then address the problem of how to construct a set of intervals that offer $FCR \leq q$.

Using any of the three methods we offer here to construct $1-q$ confidence intervals for each selected parameter indeed assures $FCR \leq q$. By constructing each interval based on the conditional distribution, the effect of selection is incorporated into the marginal coverage probability, and conditional on the number of selected parameters $M' = m'$, we have

$$E\left(\frac{\#\text{of non-covering intervals of the } m'\text{constructed}}{m'} \middle| M' = m'\right) \leq \frac{1}{m'} m' q = q,$$

and hence holds when expectation is further taken over M'. Interstingly, the FCR confidence intervals offered by Benjamini and Yekutieli (2005) were the initial motivation for the current work. For $Y_i \sim N(\theta_i, 1),\ i = 1, ..., m$, consider the selection rule picking only those θ_i for which $|y_i| > c$, and denote by m' the size of this subset. In their procedure, a $(1 - \frac{m'\alpha}{m})$ interval is constructed for each of the selected parameters, regardless of the size of $|y_i|$.

In contrast, from a conditional point of view, when $|y_i| >> c$ even the $y_i - z_{1-\alpha/2} < \theta_i < y_i + z_{1-\alpha/2}$ has approximately $1-\alpha$ coverage (because the conditioning effect is negligible), so an interval at a level close to $1-\alpha$ should be sufficient. By the constructions offered in this paper we attempt to avoid the uniform inflation of CQC their confidence intervals and this is indeed achieved by constructing conditional intervals. Once having the conditional confidence intervals at hand, we wanted the constructed intervals to often make a sign determination, that is, to avoid including parameter values of opposite signs. Combining the approach of BH&S with the conditional approach led to the rest of the developments offered here.

The work by Finner (1994) addresses a similar concern to ours, that of confidence interval following the rejection of a two sided hypothesis, and the approach is similar in that acceptance regions are constructed and inverted. Alas Finner (1994) has constructed one-sided intervals, giving up entirely on their length (being always infinite). Interestingly, for large values of the observable X, his confidence interval reverts to the usual unconditional one-sided interval, just as the confidence intervals constructed here revert in the same situation to the regular two-sided interval.

A construction of confidence intervals of bounded intervals for $\theta = E(Y)$ based on a conditional distribution is given in Zhong and Prentice (2008). The methods proposed here are different in a number of essential ways: first, our methods yield exact confidence intervals, while Zhong and Prentice provide "asymptotic" confidence intervals, in that they assume asymptotic distributions for certain terms used to obtain their intervals. Second, while it is not clear what the properties of these asymptotic intervals are, the two main procedures we propose inherently possess favorable sign determination properties, i.e, when using the confidence interval to infer about the sign, they are more powerful.

2 Shortest Acceptance Region

As a first attempt at constructing a family of acceptance regions, we associate each value of $\theta = E(X)$ with the shortest possible region of the observation space that captures a probability of $1 - \alpha$. In general, for each value assumed by θ, this is the set $A(\theta) = \{x : f_\theta(x) > \mathcal{K}_\theta\}$, where $\mathcal{K}_\theta$ is such that $\mathrm{P}\left(A(\theta)\right) = 1 - \alpha$.

In the usual normal case, where $X = Y$, these shortest regions are symmetric around θ, $A(\theta) = \{x : \ z_{1-\alpha/2} < x - \theta < z_{1-\alpha/2}\}$, and when inverted, yield the conventional, symmetric confidence interval, $x \pm z_{1-\alpha/2}$. In our case, θ is no longer a simple location parameter, and the form of these retention regions is not as trivial. In particular, we lose the symmetry which characterized the former situation, where once we found the shortest region for a particular value of θ ($\neq 0$), we have essentially found them all (because, except for $\theta = 0$, they are translations of each other). Still, the fact that the original distribution (that of Y) is symmetric and unimodal makes it easier to obtain these regions even in the 'truncated' case. Indeed, using the fact that on the support of f_θ, the density of X is proportional to that of Y (for fixed c and θ) is key to constructing $\{A(\theta)\}$.

Let $Q_c(\theta) = \mathrm{P}_\theta(|X| > c) = 2 - \Phi\left(c - \theta\right) - \Phi\left(c + \theta\right)$. As proved in the Appendix, the shortest acceptance regions for a given cutoff c are formally given by

$$A_{Srt}(\theta) = \begin{cases} \left\{x : x \in \theta \pm \Phi^{-1}\left(1 - \frac{\alpha}{2} Q_c(\theta)\right)\right\} \setminus (-c, c), & 0 \leq \theta < \theta_1 \\ (c, \theta + \Phi^{-1}[\,\Phi(c-\theta) + (1-\alpha) Q_c(\theta)\,])\,, & \theta_1 \leq \theta < \theta_2 \\ \left\{x : \ x \in \theta \pm \Phi^{-1}\left[\frac{1}{2}\left(1 + (1-\alpha) Q_c(\theta)\right)\right]\right\}, & \theta_2 \leq \theta, \end{cases} \tag{2}$$

with $A_{Srt}\left(\theta\right) = -A_{Srt}\left(-\theta\right)$ for $\theta < 0$. The parameters θ_1 and θ_2 are the solutions to

$$\Phi(c+\theta_1) - \Phi(c-\theta_1) = (1-\alpha)Q_c(\theta_1) \tag{3}$$

and

$$2\Phi(\theta_2 - c) - 1 = (1-\alpha)Q_c(\theta_2). \tag{4}$$

Upon inverting these acceptance regions, the marginal confidence set may consist of disjoint intervals. Realizing that in most situations interest will be in confidence *intervals*, we take the convex hull of the confidence region to get the interval $S_{Srt}(X) = (l(X), u(X))$. The symmetry of the acceptance regions about 0 implies the symmetry of the confidence intervals, so that $S_{Srt}(X) = -S_{Srt}(-X)$ for $X < -c$. For $X > c$, the upper end $u(X)$ is the value of θ which solves

$$2\Phi(X-\theta) = \alpha Q_c(\theta),$$

and the lower end is

$$l(X) = \begin{cases} \theta \text{ s.t. } 2(1-\Phi(X-\theta)) = \alpha Q_c(\theta), & c < X < x_1 \\ \max\{\theta: \ \Phi(X-\theta) - \Phi(c-\theta) = (1-\alpha)Q_c(\theta)\}, & x_1 < X < x_2 \\ \theta \text{ s.t. } 2\Phi(X-\theta) = (1-\alpha)Q_c(\theta), & x_2 < X \end{cases}, \tag{5}$$

where

$$x_1 = \sup\{x : x \in A_{Srt}(\theta_1)\} = c + 2\theta_1 \tag{6}$$

and

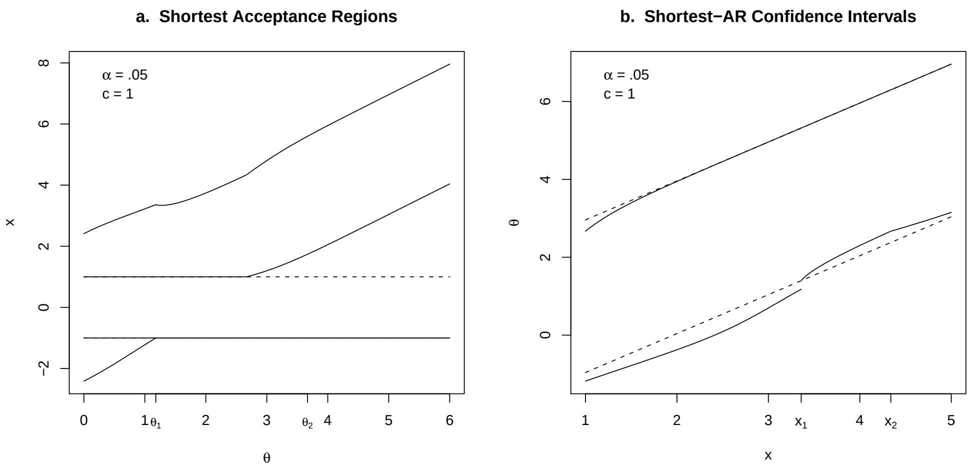

Figure 1: *(a) Acceptance regions vs. θ for the Shortest-Acceptance-Region method ($\theta \geq 0$). Critical values for θ_1 and θ_2 are* 1.18 *and* 2.67*, respectively. (b) Confidence intervals vs. observed x for the Shortest-Acceptance-Region method ($x > c$). Critical points x_1 and x_2 equal* 3.36 *and* 4.34*, respectively. The boundaries of the standard (unadjusted) confidence interval are plotted with dashed lines.*

$$x_2 = \sup \{x : x \in A_{Srt}(\theta_2)\} = 2\theta_2 - c \ . \tag{7}$$

3 A Conditional Modified Pratt (CMP) Procedure

Even though the above procedure yields a valid confidence interval for θ, for a testing procedure acceptance regions that avoid crossing zero to the side opposite to the sign of θ, have more power in determining the sign of the parameter. For a unimodal, symmetric, random variable, Pratt (1961) observed that a confidence interval designed to have the shortest expected length at

$\theta = 0$ enjoys a further favorable sign determination property. He shows that, in general, $E_\theta\,|S(X)| = \int_\nu \mathrm{P}_\theta\{\nu \in S(X)\}d\nu$. Reasoning in acceptance regions, the right term is equivalent to $\int_\nu \mathrm{P}_\theta\{x \in A(\nu)\}d\nu = \int 1 - \mathrm{P}_\theta\{x \notin A(\nu)\}d\nu$, which is minimized when $A(\nu)$ is a maximum power test for testing the hypothesis that the true value is ν aginst the alternative θ. Taking $\theta = 0$, $A(\nu)$ is a one-sided $(1-\alpha)$ ray that is "flushed" to the right for positive values of ν, and to the left for negative ones. The corresponding confidence interval (obtained by inversion) is $(0, X + c_\alpha)$ for $x > 0$ and $(X - c_\alpha)$ for $x < 0$, and is unbounded in length when $|x|$ grows big.

A modification of Pratt's procedure is suggested in BH&S in order to have a bounded length in the unimodal, symmetric case. Namely, they were looking for a confidence procedure corresponding to the most powerful test of $E(Y) = \theta$ against the alternative $E(Y) = 0$, subject to the restriction that the length of the confidence interval never exceed r times the length of the conventional interval, $2c_{\alpha/2}$. In the unimodal, symmetric case, $|A(\theta)| \leq C$ for every θ is a necessary condition to guarantee $|S(x)| \leq C$ for every x. Moreover, the family of CMP tests against $EX = 0$ under the constraint $|A(\theta)| \leq C$ (for every θ), is easily shown to yield confidence intervals with maximum length C. Thus, this family is optimal.

In our case, it is more difficult to see how the restriction on the CI length constraints the structure of the family of acceptance regions. The facts that θ is no longer a mere shift parameter and that the support of $f_\theta(y)$ is not the whole line anymore, pose serious difficulties in solving the restricted optimization problem. Therefore, we directly restrict the length of the acceptance region.

While in the unimodal, symmetric case, the shortest $(1-\alpha)$ region for every θ is of length $2c_{\alpha/2}$, in the truncated case there is no such common quantity, and the length of the shortest region for θ depends on θ, as shown in Figure (2) for $Y \sim N(\theta, 1)$ and $c = 1$. Note that the lengths of the conditional acceptance regions are shorter than $2z_{1-\alpha/2}$, because the decay of the conditional density is faster than that of the unconditional one. A natural thing to do, then, is allow each $A(\theta)$ to extend as much as r times the length of the shortest valid region for that specific θ, and among these seek for a retention region $A_{\text{CMP}}(\theta)$ which corresponds to a CMP test for $EX = \theta$ agains the alternative $EX = 0$.

Since f_θ is proportional to g_θ on $[-c, c]^c$ (for any fixed θ), the ordering of x values in $[-c, c]^c$ by the likelihood ratio under the conditional densities $f_\theta(x)$ is the same as the ordering by the likelihood ratio of the unconditional densities $g_\theta(x)$. Thus, if there were no restriction on the length of the retention region, for any $\theta > 0$ the desired region would be $\{x \in [-c, c]^c : x > t_\theta\}$, where t_θ is such that this set holds exactly $1-\alpha$ probability (under f_θ). As the length of the retention region is constrained, the best we can do (in terms of power) is take, for any $\theta > 0$, $A_{\text{CMP}}(\theta) = \left\{x \in [-c, c]^c : \bar{t}_{\theta,r} < x < \tilde{t}_{\theta,r}\right\}$, where $\left(\bar{t}_{\theta,r}, \tilde{t}_{\theta,r}\right)$ is the pair with biggest value of the first component among all pairs which satisfy $\text{P}\left(A_{\text{CMP}}(\theta)\right) = 1-\alpha$ and $|A_{\text{CMP}}(\theta)| = r\,|A_{\text{S}}(\theta)|$. For $\theta < 0$, we clearly have $A_{\text{CMP}}(\theta) = -A_{\text{CMP}}(-\theta)$, and $A_{\text{CMP}}(0)$ is chosen to be symmetric around zero. Let $\tilde{\theta}_1$ be the value of $\theta \in (0, \theta_1)$ which solves

$$\Phi(c + r\,|A_{Srt}(\theta)| - \theta) - \Phi(c - \theta) = (1-\alpha)Q_c(\theta), \tag{8}$$

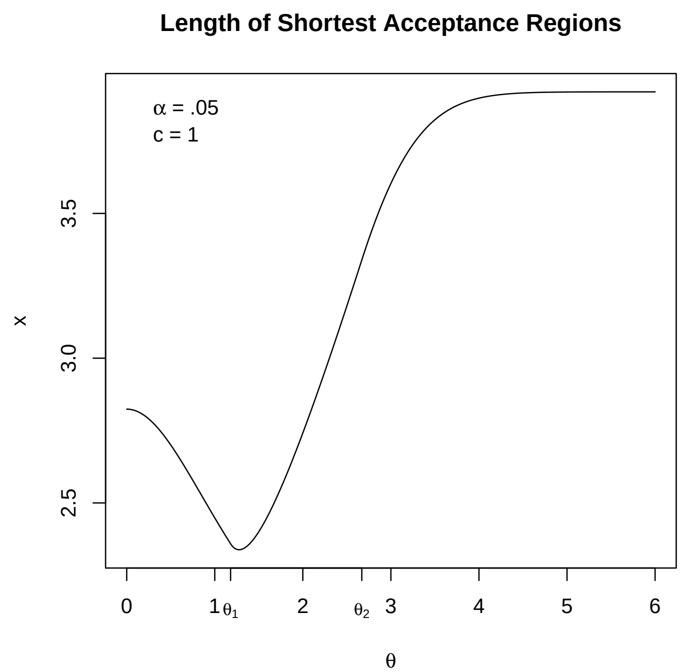

Figure 2: *Length of the shortest acceptance region vs θ for $\alpha = .05$ and $c = 1$. When θ grows big, the size of the approaches $2z_{1-\alpha/2} \approx 3.92$, the length of the standard acceptance region for the usual normal random variable (Note: when the acceptance region consists of disjoint intervals, the length is taken to be the sum of lengths of these intervals)*

and for $0 < \theta < \tilde{\theta}_1$ denote by $g_1(\theta)$ the value of $x \in (\inf A_{Srt}(\theta), -c) = (\theta - \Phi^{-1}(1 - \frac{\alpha}{2}Q_c(\theta)), -c)$ for which

$$1 - \Phi(\theta - x) + 1 - \Phi(c + r\left|A_{Srt}(\theta)\right| - (-c - x) - \theta) = \alpha Q_c(\theta), \tag{9}$$

and by $g_2(\theta)$ the biggest value of $x \in (c, \infty)$ for which

$$\Phi\left(x + r\left|A_{Srt}(\theta)\right| - \theta\right) - \Phi(x - \theta) = (1 - \alpha)Q_c(\theta). \tag{10}$$

Then, as shown in the Appendix,

$$A_{\text{CMP}}(\theta) = \begin{cases} \left[-\Phi^{-1}\left(1 - \frac{\alpha}{2}Q_c(\theta)\right), \Phi^{-1}\left(1 - \frac{\alpha}{2}Q_c(\theta)\right)\right] \setminus (-c, c), & \theta = 0 \\ \left(g_1(\theta), c + r\left|A_{Srt}(\theta)\right| - (-c - g_1(\theta))\right) \setminus (-c, c), & 0 < \theta < \tilde{\theta}_1 \\ \left(g_2(\theta),\ g_2(\theta) + r\left|A_{Srt}(\theta)\right|\right), & \theta > \tilde{\theta}_1 \end{cases}, \tag{11}$$

with $A(\theta) = -A(-\theta)$ for $\theta < 0$.

The confidence set obtained by inverting this family of tests and taking its convex hull, is

$$S_{\text{CMP}}(X) = \begin{cases} (l_1(X), u(X)), & X < \bar{c}_r(0) \\ [0, u(X)), & \bar{c}_r(0) < X < c_{\alpha/2} \\ (0, u(X)), & c_{\alpha/2} < X < \tilde{c}_r(0) \\ (l_2(X), u(X)), & \tilde{c}_r(0) < X < c + r\left|A_{Srt}(\tilde{\theta}_1)\right| \\ (l_3(X), u(X)), & X > c + r\left|A_{Srt}(\tilde{\theta}_1)\right| \end{cases}. \tag{12}$$

In the above, $u(x)$ is the root (w.r.t. θ) such that

$$\Phi(x + r\,|A_{Srt}(\theta)| - \theta) - \Phi(x - \theta) - (1-\alpha)p(c, \theta) = 0; \tag{13}$$

$l_1(x)$ is the value of $x \in (-\tilde{\theta}_1, 0)$ for which

$$1 - \Phi(x - \theta) + 1 - \Phi(\theta - x + 2c + r\,|A_{Srt}(\theta)|) - \alpha p(c, \theta) = 0; \tag{14}$$

$l_2(x)$ is the value of $x \in (0, \tilde{\theta}_1)$ such that

$$1 - \Phi(x - \theta) + 1 - \Phi(\theta - x + 2c + r\,|A_{Srt}(\theta)|) - \alpha p(c, \theta) = 0; \tag{15}$$

$l_3(x)$ is the smaller root of

$$\Phi(x - \theta) - \Phi(x - r\,|A_{Srt}(\theta)| - \theta) - (1-\alpha)p(c, \theta) = 0; \tag{16}$$

$\bar{c}_r(0)$ is the *smaller* root (w.r.t. x), and $\tilde{c}_r(0)$ the *bigger* root, of

$$1 - \Phi(x) + 1 - \Phi(2c + r\,|A_{Srt}(0)| - x) - 2\alpha(1 - \Phi(c)) = 0 \tag{17}$$

and

$$c_{\alpha/2} = \Phi^{-1}(1 - \frac{\alpha}{2}(1 - \Phi(c) + 1 - \Phi(c))). \tag{18}$$

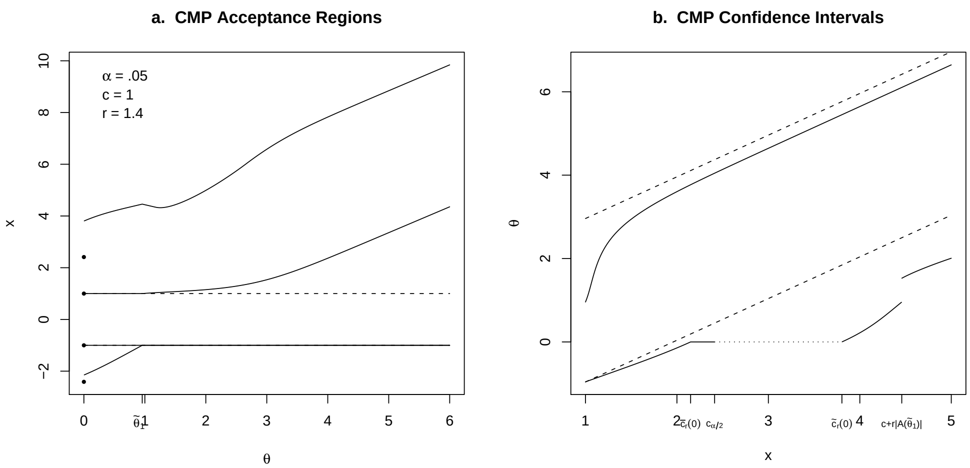

Figure 3: *(a) Conditional Modified Pratt acceptance regions for* $r = 1.4$. *Critical value* $\tilde{\theta}_1$ *is* .96. *The boundaries of* $A_{CMP}(0)$ *are marked with filled circles. (b) Conditional Modified Pratt confidence intervals for* $r = 1.4$. *Critical points* $\bar{c}_{1.4}(0), c_{\alpha/2}, \tilde{c}_{1.4}(0)$ *and* $c + 1.4\left|A_{Srt}(\tilde{\theta}_1)\right|$ *equal* 2.15, 2.41, 3.8, 4.46, *respectively. Dotted line indicates that* 0 *is not included in the interval (but only in its closure), that is, a strict sign determination occurs. The boundaries of the standard (unadjusted) confidence interval are plotted with dashed lines. Note that while a strict sign determination occurs starting at the same point* ($c_{\alpha/2}$) *for the CMP and the Shortest-Acceptance-Region methods, the CMP interval enjoys an earlier weak sign determination. CMP interval length approaches* $2rz_{1-\alpha/2} \approx 5.49$ *as* $x \to \infty$.

4 Conditional Quasi-Conventional (CQC) Confidence Intervals

A different approach to compromise between CI length and more power to determine the sign of the parameter, further suggested by BH&S, is minimizing a weighted sum of the length of the acceptance region $|A(\theta)|$, and the extent to which the acceptance region crosses the origin subject to a size constraint. As in the usual symmetric, unimodal case, this approach yields conditional confidence intervals which revert to the conventional symmetric interval when $|x|$ is large.

4.1 Deriving the Family of Acceptance Regions

Formally, we wish to associate each value of $\theta \neq 0$ with a region

$$A_{\mathrm{CQC}}(\theta) = \arg\min_{A}\{\lambda\,|A(\theta)| + \sup_{x\in A(\theta):\mathrm{sgn}(x)\neq\mathrm{sgn}(\theta)} |x|\} \tag{19}$$

where A is any region which satisfies $\mathrm{P}_\theta(X \in A) \geq 1-\alpha$.

In the usual symmetric, unimodal case, the optimal set of acceptance regions corresponding to a CI with length that never exceeds r times the length of the conventional, symmetric one, has a relatively simple structure. In fact, for a given α, it is completely characterized by the quantity $\bar{c} = \inf_{\theta>0} \inf\{y : y \in A_{\mathrm{CQC}}(\theta)\}$, which is determined by r (or, equivalently, by λ). We use the exact same optimization setup (19) to obtain the family of acceptance regions in our conditional setting. However, now there is considerable complexity in determining the value of λ corresponding to a resulting CI with a certain maximum length. Therefore, we first show how to obtain the acceptance regions for a prescribed λ, and later discuss the relation between λ and the corresponding maximum CI length.

Proposition 4.1. *For any $\lambda > 0$, a solution to the optimization problem* (19) *is given by*

$$A_{CQC}(\theta) = \begin{cases} (-\Phi^{-1}\{1-\frac{\alpha}{2}p(c,0)\}, \Phi^{-1}\{1-\frac{\alpha}{2}p(c,0)\}) \setminus (-c,c), & \theta = 0 \\ (-c-d^\star(\theta), c+h_1(\theta)) \setminus (-c,c), & 0 < \theta < \theta_1' \\ (c, c+h_2(\theta)), & \theta_1' \le \theta < \theta_1 \\ A_{Srt}(\theta), & \theta_1 \le \theta, \end{cases} \tag{20}$$

with $A_{CQC}(\theta) = -A_{CQC}(-\theta)$ for $\theta < 0$, and where

(i) θ_1' is the value of θ satisfying

$$1 + \lambda\left(1 - \frac{\phi(c+\theta)}{\phi(\Phi^{-1}(2-\Phi(c+\theta)-\alpha Q_c(\theta)))}\right) = 0, \tag{21}$$

and θ_1 is given in section 2

(ii) $d^\star(\theta)$ is the value of d which solves

$$1 + \lambda\left(1 - \frac{\phi(c+d+\theta)}{\phi(\Phi^{-1}(2-\Phi(c+d+\theta)-\alpha Q_c(\theta)))}\right) = 0 \tag{22}$$

(iii) $h_1(\theta)$ *is the value of* h *which solves*

$$1-\Phi(c+h-\theta)+1-\Phi(d^{\star}(\theta)+c+\theta)=\alpha Q_c(\theta) \tag{23}$$

(iv) $h_2(\theta)$ *is the value of* h *which solves*

$$\Phi(c+h-\theta)-\Phi(c-\theta)=(1-\alpha)Q_c(\theta) \tag{24}$$

(v) $A_{Srt}(\theta)$ *is the shortest acceptance region for* θ*, given in section 2.*

Proof. For $\theta = 0$, the accpetance region can be chosen arbitrarily, and we take it, as usual, to be of the form $(-x,x)\setminus[-c,c]$, where x is such that $\mathrm{P}_{\theta=0}(A(0)) = 1-\alpha$. When $\theta \geq \theta_1$, it is obvious (from the discussion in section 2) that the shortest region possible, $A_{Srt}(\theta)$, is optimal.

For $0<\theta<\theta_1$, denote by $d = d(\theta)$ the amount of extension of any candidate $A(\theta)$ to the left of $-c$, and by $l = l(\theta)$ the total length, $|A(\theta)|$. A solution to (19) will be immediate after proving the following

Claim 1. *Let* $\psi(d,\theta,\lambda) = c+d+\lambda(d-c+\theta+\Phi^{-1}(2-\Phi(d+c+\theta)-\alpha Q_c(\theta)))$*. Furthermore, let* $\underline{d}(\theta) = \max\left(-c-\theta+\Phi^{-1}\left\{1-\alpha Q_c(\theta)\right\},0\right)$*,* $\bar{d}(\theta) = -c-\inf_x A_{Srt}(\theta)$ *and* $\theta^* = \inf\{\theta>0 : \underline{d}(\theta)=0\}$*. Then:*

a. For $0<\theta<\theta_1$*, the region* $A(\theta) = (-c-d^{\star}, c+l^{\star}-d^{\star})\setminus(-c,c)$ *where*

$$d^{\star} = arg \min_{\underline{d}(\theta)\leq d\leq \bar{d}(\theta)} \psi(d,\theta,\lambda) \tag{25}$$

and where $l^\star$ is determined through

$$1 - \Phi(d^\star + c + \theta) + 1 - \Phi(l^\star - d^\star + c - \theta) = \alpha Q_c(\theta),$$

is a solution to the original restricted optimization problem (19)*.*

b. *On $G = \{(d, \theta) : \underline{d}(\theta) < d < \bar{d}(\theta),\ \theta > 0\} \cup \{(0, \theta) : \theta^* < \theta < \theta_1\}$ and for any fixed $\lambda > 0$, the derivative of ψ with respect to d exists, is continuous in d and θ, and is a strictly increasing function in d for any fixed θ and in θ for any fixed d.*

c. *For any fixed $\lambda > 0$, there exists $\theta \in (\theta^*, \theta_1)$ for which*

$$\frac{\partial}{\partial d}\psi(d, \theta, \lambda)\Big|_{d=0} = 0.$$

d. *For fixed $0 < \theta < \theta^*$ and for fixed $\lambda > 0$, there exists $d \in \big(\underline{d}(\theta), \bar{d}(\theta)\big)$ such that $\frac{\partial}{\partial d}\psi(d, \theta, \lambda) = 0$.*

Proof of Claim. a. Note that no region with a given extension d to the right of $-c$ is better (in terms of (19)) than $(x_1, x_2) \setminus [-c, c]$, where $-c - x_1 = d$ and x_2 is then set to make $P_\theta(A(\theta)) = 1 - \alpha$. That the part left of $-c$ should be an interval with a right end at $-c$ is obvious. As for the part which lies to the right of c, first note that we may consider only $d < -c - \inf_S xA(\theta)$, since larger d would increase both terms in (19). Now, from section 2, as long as $0 < \theta < \theta_1$, $A_{Srt}(\theta) = \{x : f_\theta(x) > \mathcal{K}_\theta\}$, where $\mathcal{K}_\theta < f_\theta(-c)$. Since for any $d < -c - \inf_x A_{Srt}(\theta)$,

$$(-c - d, -c) \cup \{x : f_\theta(x) > f_\theta(c)\} \subsetneq A_{Srt}(\theta),$$

we conclude that given any $d < -c - \inf_x A_{Srt}(\theta)$, the shortest region to the right of c we can possibly choose to hold the rest of the probability is an interval with a left end at $-c$.

Let there now be $0 < \theta < \theta_1$. By the discussion above, we may consider only candidates $A(\theta)$ of the form $(x_1, x_2) \setminus [-c, c]$, each of which is characterized by the extension d to the right of $-c$. The total length $l = l(\theta)$ of $A(\theta)$ is determined by d through

$$1 - \Phi(d + c + \theta) + 1 - \Phi(l - d + c - \theta) = \alpha Q_c(\theta). \tag{26}$$

Solving for l and plugging the expression in for $|A(\theta)|$, (19) can be rewritten in terms of d as

$$d^\star(\theta) = \arg\min_d \{c + d + \lambda \left(d - c + \theta + \Phi^{-1}(2 - \Phi(c + \theta + d) - \alpha p(c, \theta))\right), \tag{27}$$

where

$$\max\left(-c - \theta + \Phi^{-1}\left\{1 - \alpha Q_c(\theta)\right\}, 0\right) \leq d \leq -c - \inf_x A_{Srt}(\theta). \tag{28}$$

Here the lower bound is the minimal value d has to take in order for $A(\theta)$ to satisfy the coverage requirement.

b. Let $G_1 = \{(d,\theta) : \underline{\mathrm{d}}(\theta) < d < \bar{d}(\theta),\ \theta > 0\}$ and $G_2 = \{(0,\theta) : \theta^* < \theta < \theta_1\}$. Then ψ is obviously defined on G_1, and since $\Phi(c+\theta) + \alpha Q_c(\theta) > 1$ for $\theta > \theta^*$, it is also defined on G_2. Hence ψ is defined on $G = G_1 \cup G_2$. Now, for any $(d,\theta) \in G$, we have

$$\frac{\partial}{\partial d}\psi(d,\theta,\lambda) = 1 + \lambda\left(1 - \frac{\phi(c+d+\theta)}{\phi(\Phi^{-1}(2 - \Phi(c+d+\theta) - \alpha Q_c(\theta)))}\right), \qquad (29)$$

which is a continuous function of d and θ. Moreover, one can easily verify that on G, the numerator of the quotient in (29) is strictly decreasing, while the denominator is strictly increasing, in d for any fixed θ and in θ for any fixed d.

c. $\{\theta : (0,\theta) \in G\} = (\theta^*, \theta_1)$, and on that set

$$\frac{\partial}{\partial d}\psi(0,\theta,\lambda) = 1 + \lambda\left(1 - \frac{\phi(c+\theta)}{\phi(\Phi^{-1}(2 - \Phi(c+\theta) - \alpha Q_c(\theta)))}\right). \qquad (30)$$

Using the fact that θ^* and θ_1 are the values of θ which solve, respectively,

$$1 - \Phi(c+\theta) = \alpha Q_c(\theta)$$

and

$$2(1 - \Phi(c+\theta)) = \alpha Q_c(\theta),$$

it is easy to verify that for any $\lambda > 0$, the expression on the right hand side of (30) tends to $-\infty$ as $\theta \longrightarrow \theta^*$ from the right, and tends to 1 as $\theta \longrightarrow \theta_1$ from the left. By continuity, it must vanish for some intermediate value.

d. Using the fact that $\underline{\mathrm{d}}(\theta)$ is the value of d such that

$$1 - \Phi(c + d + \theta) = \alpha Q_c(\theta)$$

and $\bar{d}(\theta)$ is the value of d such that

$$2\left(1 - \Phi(d + c + \theta)\right) = \alpha Q_c(\theta),$$

one can easily verify that $\frac{\partial}{\partial d}\psi(d) \longrightarrow -\infty$ as $d \longrightarrow \underline{\mathrm{d}}(\theta)$ from the right, and $\frac{\partial}{\partial d}\psi(d) \longrightarrow 1$ as $d \longrightarrow \bar{d}(\theta)$ from the left. Therefore, there must be some $d \in \left(\underline{\mathrm{d}}(\theta), \bar{d}(\theta)\right)$ such that $\frac{\partial}{\partial d}\psi(d) = 0$.

$\square$

Returning to the proof of our proposition, for any $0 < \theta < \theta_1$ we first need, by (a), to obtain the minimizer in (19). From (c) and (b), there exists a unique value of θ, denoted by θ_1', which satisfies (21). It also follows from (c) and (b) that for $\theta_1' < \theta < \theta_1$, $\frac{\partial}{\partial d}\psi(d, \theta, \lambda) > 0$ for any $d > 0$. Therefore, in that case the minimizer is necessarily $d = 0$. As for $0 < \theta < \theta_1'$, we distinguish between two cases. If $0 < \theta < \theta^*$, by (b) and (d), the value of d for which $\frac{\partial}{\partial d}\psi(d, \theta, \lambda) = 0$ is the minimum. If $\theta^* < \theta < \theta_1'$, then by (b) we have that $\frac{\partial}{\partial d}\psi(0, \theta, \lambda) < \frac{\partial}{\partial d}\psi(0, \theta_1', \lambda) = 0$ and $\frac{\partial}{\partial d}\psi(d, \theta, \lambda) \longrightarrow 1$ as $d \longrightarrow \bar{d}(\theta)$ from the left (as shown in the proof of (d)), thus, by (b), the value of d for which the derivative vanishes is again the unique minimum.

The expressions in (20) for $0 < \theta < \theta_1$ and those in (23) and (24) are an immediate consequence of the description in (a) of an optimal acceptance region once the minimizer d^* is known.

$\square$

Having obtained the above expressions for $A_{CQC}(\theta)$, we now derive the corresponding confidence intervals by inversion.

4.2 Inverting the Acceptance Regions

The convex hull of the set $\{\theta : x \in A_{\mathrm{CQC}}(\theta)\}$, where $A_{\mathrm{CQC}}(\theta)$ is given in (20), is

$$S_{\mathrm{CQC}}(X) = \begin{cases} (-l_1(X), u(X)), & 0 < X < \bar{c}_\lambda(0) \\ [0, u(X), & \bar{c}_\lambda(0) < X < c_{\alpha/2} \\ (0, u(X)), & c_{\alpha/2} < X < \tilde{c}_\lambda(0) \\ (l_2(X), u(X)), & \tilde{c}_\lambda(0) < X < \tilde{c}_\lambda(\theta_1') \\ S_{Srt}(X), & X > \tilde{c}_\lambda(\theta_1'), \end{cases} \tag{31}$$

with $S_{\mathrm{CQC}}(X) = -S_{\mathrm{CQC}}(-X)$ for $X < 0$ and where

(i) $\bar{c}_\lambda(0) = \sup_{\theta<0} \sup A(\theta)$ and is the value of x for which

$$1 + \lambda \left(1 - \frac{\phi(c+d)}{\phi(\Phi^{-1}(2 - \Phi(c+d) - \alpha Q_c(0)))}\right)\Bigg|_{d=x-c} = 0$$

(ii) $c_{\alpha/2} = \sup_x A(0)$

(iii) $\tilde{c}_\lambda(0) = \inf_{\theta>0} \sup_x A(\theta) = \Phi^{-1}\{2 - \Phi(\bar{c}_\lambda(0)) - \alpha Q_c(0)\}$

(iv) $\tilde{c}_\lambda(\theta_1') = \sup_x A(\theta_1') = \theta_1' + \Phi^{-1}\left\{\Phi(c-\theta_1') + (1-\alpha)Q_c(\theta_1')\right\}$

(v) $l_1(x)$ is the value of θ such that

$$1 + \lambda\left(1 - \frac{\phi(c+d+\theta)}{\phi(\Phi^{-1}(2-\Phi(c+d+\theta)-\alpha Q_c(\theta)))}\right)\Big|_{d=x-c} = 0$$

(vi) $l_2(x)$ is the value of θ for which

$$1+\lambda\left(1 - \frac{\phi(c+d+\theta)}{\phi(\Phi^{-1}(2-\Phi(c+d+\theta)-\alpha Q_c(\theta)))}\right)\Big|_{d=-c-\theta+\Phi^{-1}\{2-\Phi(x-\theta)-\alpha Q_c(\theta)\}} = 0$$

(vii) $u(x)$ is the value of θ for which

$$2\left(1-\Phi(\theta-x)\right) = \alpha Q_c(\theta).$$

4.3 Relationship between λ and the maximum length of the confidence interval

We assumed a given value for λ, and derived a family of acceptance regions which yield a confidence interval per that particular value of λ. Ideally, though, we would like to constraint the maximal length of the interval, which is conceptually easier to quantify, and let it dictate a corresponding value for λ, which in turn determines the acceptance region.

In the unimodal, symmetric case, because of the relatively simple structure of the optimal acceptance regions, it was fairly easy to show that the QC inter-

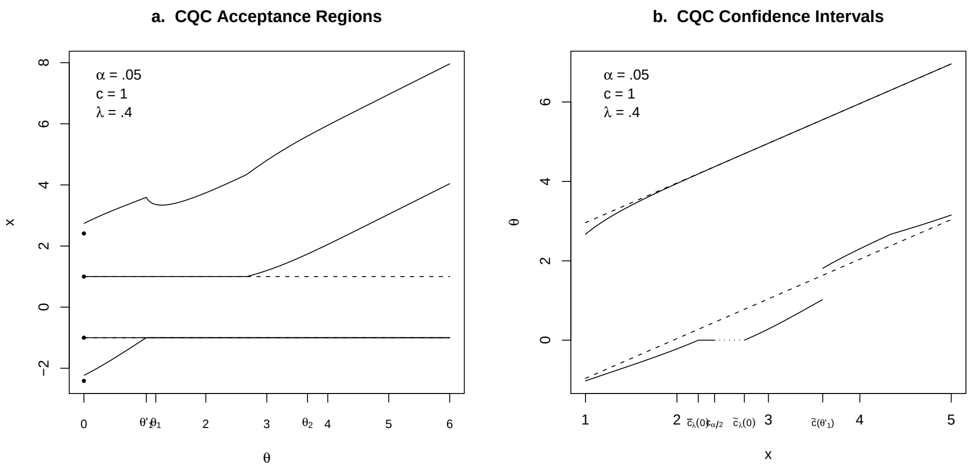

Figure 4: *(a) Conditional Quasi-Conventional acceptance regions for* $\lambda = .4$. *Critical values* θ_1', θ_1 *and* θ_2 *are* 1.02, 1.18 *and* 2.67, *respectively. (b) Conditional Quasi-Conventional confidence intervals for* $\lambda = .4$. *Critical points* $\bar{c}_{.4}(0), c_{\alpha/2}, \tilde{c}_{.4}(0)$ *and* $\tilde{c}_{.4}(\theta_1')$ *are* 2.23, 2.41, 2.74 *and* 3.59, *respectively. The boundaries of the standard (unadjusted) confidence interval are plotted with dashed lines. Compared to the CMP method with* $r = .4$, *the CQC interval (weakly) determines the sign starting at* $\bar{c}_{.4}(0) = 2.23$, *while the CMP interval does it slightly earlier, starting at* $\bar{c}_{1.4}(0) = 2.15$. *While the length of the CMP interval is approximately* 5.49 *for big* x, *the CQC interval approaches the length of the standard, unadjusted interval,* $2z_{1-\alpha/2} = 3.92$. *as* $x \to \infty$.

val always has maximal length greater than $2z_{1-\alpha/2}$ and to specify the family of QC acceptance regions corresponding to a confidence interval with maximal length (across all x values) of r times the length of the conventional interval, for $r > 1$. In our conditional case θ is no longer a mere location parameter, and the family of optimal acceptance regions no longer admit the simple sturcture as before. This makes the relationship between the maximal length of the CQC interval and λ much more complicated and difficult to analyze analytically. Nevertheless, since it is reasonable to expect the CQC interval to be longer than the Shortest-AR interval "for most x", it is interesting to numerically investigate how much worse it really does. Since our reference, the Shortest-AR interval, now has different lengths for different values of x, we coose as a measure of comparison the quantity

$$g_c(\lambda) = \sup_x \{|S^{\lambda}_{CQC}(x)|/|S_{Srt}(x)|\} \tag{32}$$

and evaluate it over a range of λ values. Figure 5 shows $g_c(\lambda)$ as a function of λ for different values of c. In practice, $|S^{\lambda}_{CQC}(x^*)|/|S_{Srt}(x^*)|$ for $x^* = \tilde{c}_{\lambda}(\theta_1') - \epsilon$ (with $\epsilon = 10^{-5}$) was taken as $g_c(\lambda)$, since simulations imply that $g_c(\theta) = \lim_{t \to x^*} |S^{\lambda}_{CQC}(t)|/|S_{Srt}(t)|$ from the left.

Remark For each of the three methods discussed above, the endpoints $L(x)$ and $U(x)$ of the confidence interval are monotone nondecreasing in x for $x > c$ (nonincreasing x for $x < -c$), as can be observed in the above figures. Apart from the fact that the convex hull is taken upon inversion of the acceptance regions, this property is ensured by the facts that (i) $A(\theta) \cap \{x > c\}$

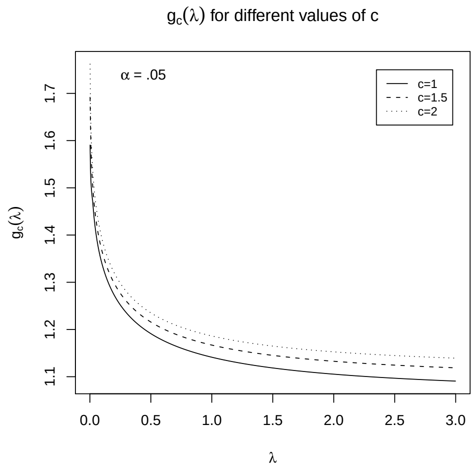

Figure 5: $g_c(\lambda)$ *as a function of* λ *for* $c = 1, 1.5$ *and* 2. *The quantity* $|S^{\lambda}_{CQC}(x^*)|/|S_{Srt}(x^*)|$ *for* $x^* = \tilde{c}_\lambda(\theta_1') - \epsilon$ *(with* $\epsilon = 10^{-5}$*) was taken as* $g_c(\lambda)$*.*

is always an interval and (ii) $l(\theta) = \inf\{x > c : x \in A(\theta)\}$ is nondecreasing with $\lim_{\theta\to-\infty} l(\theta) = c$. That $U(x)$ is nondecreasing is trivial from (ii); This is also the case for $L(x)$, since, letting $\bar{u}(\theta) = \sup_{\theta' \le \theta} \sup_x A(\theta')$ (the upper envelope of $u(\theta)$), we have from (i) and (ii) that $L(x) := \inf\{\theta : x \in A(\theta)\} = \inf\{\theta : \bar{u}(\theta) > x\}$, which is a nondecreasing function of x since $\bar{u}$ is nondecreasing.

5 Example

In an ongoing experiment on the response to stress as reflected in brain activity and connectivity as measured by functional magnetic resonance imaging, subjects were exposed to two segments of movies that differed in the level of stress they project. Both the activity at voxels in the brain and the level of Cortisol in their blood were recorded while being exposed to the differing segments. The Cortisol levels, which are known to reflect the level of stress exposure, were log-transformed before taking the difference per subject. The resulting distribution of the estimates is quite Gaussian. The difference in activity level was estimated from a generalized linear model. The results per voxel, as inspected for a sample of voxels, are again quite symmetric and close to Gaussian. One of the questions that interests us is the correlation between the difference in activity and the difference in Cortisol levels in the promising voxels. The results for 16 subjects are available at this stage, as the study is still ongoing (more subjects' data will become available). For the same reason the study has not been reported yet, so we shall avoid giving further details about the experiment and the analysis leading to the correlations.

For our purpose it is enough to start from the 14756 correlations we have - one for each voxel. Interest lies only with voxels for which the correlation is high. In this case we looked at correlations larger in absolute value than 0.6, and there were 15 positive and 21 negative such correlations. The correlations were Fisher transformed into Gaussian variates, then CQC intervals were calculated using the values $\sigma = 1/\sqrt{16-3}$ and $\lambda = .4$. The standard unadjusted 95% marginal confidence intervals were also calculated on this scale. Both sets of intervals were back transformed into the correlation scale and are presented in Figure 6. The CQC intervals are given by the lines and the standard intervals are marked by the tick marks. Note that all the 36 selected standard intervals do not cover 0, excluding parameter values of the opposite sign, and offering evidence for both high positive correlations and high negative ones. In contrast, only the two largest correlations and the two smallest correlations have conditional intervals that exclude correlations of opposite signs, where the two largest ones include 0 and the two smallest ones exclude 0. The other 32 conditional intervals extend to beyond 0, indicating there is no evidence to exclude correlations of the other sign, based on the results at hand. Considering 36 intervals at 0.95 level we expect non-coverage by 1.8 intervals, so we do not have yet an evidence of either positive or negative correlation from the conditional confidence intervals. This point of view is the more realistic one in our case.

Note also that the upper side of the confidence intervals are almost the same as those of the standard ones. In fact, those of the CQC intervals are just a bit shorter than those of the standard ones very close to the conditioned upon value.

We are thankful to Prof. Talma Hendler from Tel Aviv University who is leading this research, with whom the first author is cooperating. We are thankful to Sharon Vaisvaser and to Yonatan Weinetraub who conducted the analysis leading to the correlations.

6 Comparison with Other Methods

In their work concerning selection bias of estimators in genome-wide association studies (GWAS), Zhong and Prentice (2008) suggest methods to obtain confidence intervals for $\theta = E(Y)$, where $Y \sim N(\theta, \sigma^2)$, upon observing $X = Y|(|Y| \geq c\sigma)$. Because σ is assumed to be known, there is no loss of generality in setting $\sigma = 1$.

First, they suggest using the asymptotic distribution of the log-likelihood

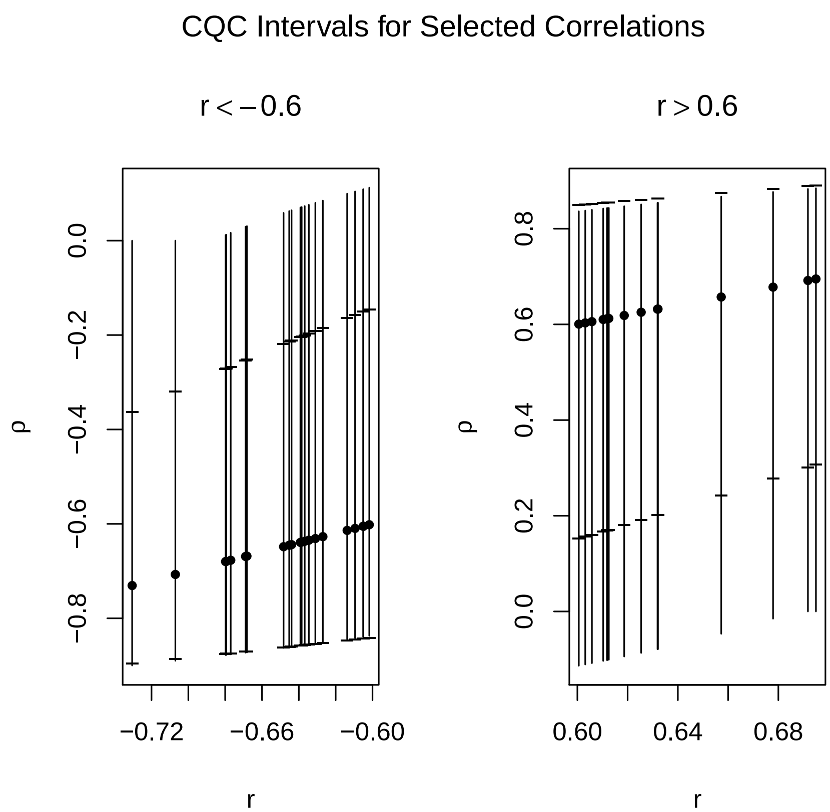

Figure 6: *CQC and standard (unadjusted) intervals for correlations between Cortisol level changes in the blood and activity changes in the brain at various locations, when viewing two movie segments. Only correlations that are greater in absolute value than* 0.6 *are displayed. CQC intervals are given by the vertical lines and the endpoints of the unadjusted standard are indicated by tick marks. While all of the standard intervals are separated from zero, only the two largest correlations and the two smallest ones have conditional intervals that exclude values of opposite signs, the intervals for the two largest ones including 0 and the ones for the two smallest ones excluding 0.*

ratio, namely

$$2ln(L_x(\hat{\theta}_{MLE})/L_x(\theta_0)) \sim \chi^2_1,$$

where $L_x(\theta)$ is the likelihood of θ with respect to the conditional distribution in 1.

To obtain acceptance regions we now take, for each $\theta = \theta_0$,

$$A_{LR}(\theta_0) = \left\{x : 2ln(L_x(\hat{\theta}_{MLE})/L_x(\theta_0)) \leq \chi^2_{1;1-\alpha}\right\} \tag{33}$$

Inverting this family yields their confidence sets

$$S_{LR}(X) = \left\{\theta : lnL_X(\theta) \geq lnL_X(\hat{\theta}_{MLE}) - \chi^2_{1;1-\alpha}/2\right\}. \tag{34}$$

This approximation is obviously not supposed to hold for a small sample sizes. Even when many observations are combined into a single estimator $\hat{\theta}$, but then θ is estimated conditional on $\left|\hat{\theta}\right| \geq \sigma c$, we are practically at a situation where we attempt to construct the confidence set from just a single observation.

A Second, quantile-based, confidence interval is proposed by having each acceptance region leave an $\alpha/2$ probability on each tail :

$$A_{QB}(\theta_0) = (t_{\alpha/2;\theta_0}, t_{1-\alpha/2;\theta_0}) \setminus (-c, c), \tag{35}$$

where $t_{\xi;\theta_0} = t : \int_{-\infty}^{t} f(x;\theta_0)dx = \xi$ (it is the ξ percentile of X under $\theta = \theta_0$). Inverting those regions, X being the observed value, we have

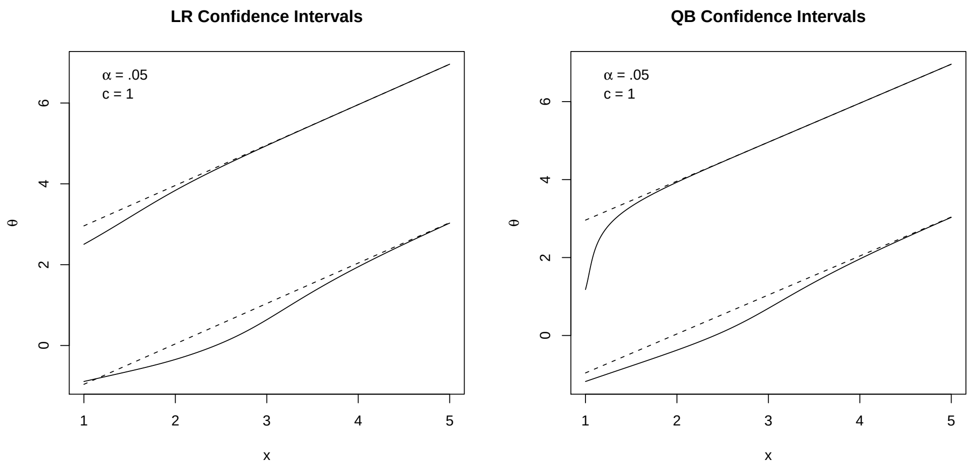

Figure 7: *Confidence intervals for the Likelihood Ratio and Quantile-based methods*

$$S_{QB}(X) = [\theta_l, \theta_u],$$

where

$$\theta_l = \gamma : \int_{-\infty}^{X} f(x;\gamma)dx = 1 - \alpha/2$$

and

$$\theta_u = \gamma : \int_{-\infty}^{X} f(x;\gamma)dx = \alpha/2$$

Examining the acceptance regions of each of the above two methods, we can notice that both yield confidence intervals which are roughly symmetric around x for large values of x, as displayed by figure (7). Notice also that the quantile-based acceptance regions are the same as those of the Shortest-

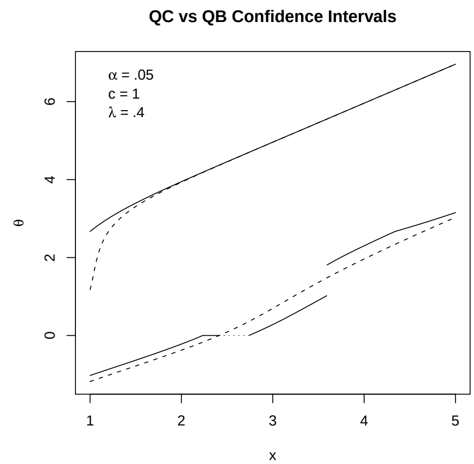

Figure 8: *Quantile-Based (dashedline) vs Conditional Quasi-Conventional (solid line) confidence intervals. Parameter value used for the CQC interval is* $\lambda = .4$. *The CQC interval makes an earlier (weak) sign determination at the expense of a later separation from zero, and is approximately the standard, symmetric interval for large* x

Acceptance-Region ones for small θ $(0 \leq \theta \leq \theta_1,\ \theta_1$ defined in (3)), and in particular both methods make (weak) sign determination starting at the same value of the observed x. We can further see in (7) that, similarly to the CMP confidence interval, the quantile-based interval has a sharp drop towards the origin, which is due to the separation of the acceptance regions from c at $\theta = \theta_1$. In figure (8) we can see the tradeoff between early sign determination and late separation from zero: The CQC interval, as expected, weakly determines the sign earlier than does the Quantile-based interval, but does it at the cost of a later separation from zero.

Table (6) presents estimated values for the length, coverage probability and probability of making a weak sign determination per each of the methods discussed in the paper, through a simulation of $n = 2000$ samples with cutoff

$c = 1, 3$ and $\theta = 0, 1, 3, 5$. We used $\theta = -10^{-6}$ with the notation θ^{-} instead of $\theta = 0$ in order to distinguish between a correct sign determination and a wrong one. In terms of coverage, all methods except LR always offer the right coverage, even if sometimes conservatively so. The LR is unstable in terms of coverage, sometimes lower than needed ($c = 1,\ \theta = 1$), sometimes higher ($c = 3,\ \theta = 1$). As can be seen, the CMP procedure , but has significantly inflated average length when θ is big compared to the other methods (up to 30% longer). It is interesting to note that the CQC interval has a smaller expected length for big θ ($\theta = 5$) than do the LR and QB methods, while for small θ its expected length is very close to that of these two methods.

The three conditional intervals offered here always enjoy, as expected, higher power to determine the sign than do the LR and the QB methods. with the CMP having higher power than the other two. However the power of the CQC approaches that of the CMP as θ increases, and never falls by much.

In summary, it seems that the CQC method, while enjoying exact coverage properties, also reaches a good overall compromise between sign determination and expected length, across values that the conditioning constant c and the penalty term λ may take, and across θ where it matters.

7 Generalizations

It is worth emphasizing that the only properties of the Gaussian distribution that were used in the derivation of the confidence intervals are symmetry and unimodality. Hence confidence intervals can be constructed with the relevant

Table 1: *Simulation estimates of the expected length (Lng), probability for weak sign determination (PSD) and coverage probability (Cvr) for the Shortest-acceptance-region (Srt), CCMP, CQC, Likelihood-ratio (LR) and Quantile-based (QB) confidence intervals, and for different values of* θ*. The first table gives simulated values for* $c = 1$ *and the second for* $c = 3$*. Here sample size* $n = 2000$*,* $\sigma = 1$ *and the parameters of CCMP and CQC methods are* $r = 1.5$ *and* $\lambda = .4$*, respectively. Standard error of simulation based Cvr is bounded by* $.005$*, and that of PSD is bounded by* $.01$ *using the binomial approximation; standard error of Lng was estimated from the simulation data, and depends both on the method used for the CI and on the specific value of* θ*, but in any case did not exceed* $.01$*.*

$c = 1$	$\theta = 0^-$			$\theta = 1$			$\theta = 3$			$\theta = 5$		
	Lng	PSD	Cvr	Lng	PSD	Cvr	Lng	PSD	Cvr	Lng	PSD	Cvr
Srt	4.127	.025	.95	4.181	.146	.955	4.07	.743	.954	3.826	.996	.948
CMP	3.259	.051	.95	3.53	.236	.951	4.615	.813	.948	5.196	.999	.951
CQC	3.998	.04	.952	4.098	.203	.95	4.229	.791	.954	3.872	.998	.948
LR	3.784	.021	.955	3.922	.141	.937	4.143	.736	.945	3.969	.995	.948
QB	3.833	.025	.95	3.997	.146	.955	4.158	.743	.951	3.963	.996	.948

$c = 3$	$\theta = 0^-$			$\theta = 1$			$\theta = 3$			$\theta = 5$		
	Lng	PSD	Cvr	Lng	PSD	Cvr	Lng	PSD	Cvr	Lng	PSD	Cvr
Srt	5.448	.026	.949	5.45	.11	.954	5.242	.414	.94	4.216	.902	.954
CMP	4.186	.056	.95	4.467	.18	.961	5.046	.511	.95	5.182	.934	.955
CQC	5.394	.042	.951	5.422	.154	.975	5.318	.473	.94	4.301	.924	.954
LR	4.703	.018	.962	4.837	.095	.976	5.063	.384	.971	4.574	.891	.943
QB	4.779	.026	.949	4.966	.11	.947	5.146	.414	.945	4.529	.902	.956

constants calculated for any other distribution enjoying these properties, including in particular the t_ν-distribution.

The CMP and the CQC intervals can be constructed for Gaussian distributions with known standard error of (Y) other than 1. The change is trivial: dividing the estimator and the conditioned upon constant c by σ, calculating the CIs and re-inflating them by σ. Yet care as to be taken as to the interpretation of the constant c. If c serves as a condition on the value of the estimator, the above is fine. If c expresses significance, say being $z_{1-\alpha/2}$, it should not be divided by σ (or equivalently should be first multiplied by σ before being divided.)

In spite of these two observations, the Gaussian case of unknown σ with small sample size being estimated from the data using $\widehat{\sigma}$ presents a challenge. Finner (1994) addresses similar issues at length when dealing with the problem of conditional one-sided confidence intervals. He makes the important distinction between the above two cases. When c serves as a condition on significance the problem involves using properties of the non-central t-distribution, and when c serves as a condition on the estimated value his solution is more complicated. In both cases the problem we face is more difficult because we cannot rely on the symmetry of the distribution as we do. We therefore leave this problem for future research.

Nevertheless, when the standard errors are estimated but can be used under the asymptotic Gaussian regime, the proposed confidence interval can be used. This would include one-sample mean, two-sample means difference and regression coefficients, with large enough sample sizes; correlation coefficients after using Fisher's transformation; the coefficients in a logistic regression; coefficients in survival analysis, and such. Other nuisance parameters that do not interfere with the role of the parameter of interest as a location parameter, or with the conditioning process.

8 Discussion

We have presented conditional confidence intervals, where the selection whether to construct a confidence interval or not depends on a fixed threshold c. As explained in the introduction, an important virtue of such confidence intervals is their possible role in large problems of inference. When facing a family of m parameters, it is often the case that the parameters of interest are only the large ones, or those significantly different from 0 at some given level. All three types of conditional intervals can be used then to set confidence intervals for these parameters only, and the set of intervals still assures control over the FCR. The selection rules can follow individual testing (i.e. p-value $\leq \alpha$), multiplicity adjusted testing such as Bonferroni (p-value $\leq \alpha/m$), or any other fixed value.

However, there is some difficulty that arises from the dual interpretation of selection via testing, which is already evident in the single parameter case. When selecting a parameter by (unconditional) testing whether it is significantly different from 0, the conditional confidence interval at the same level falls below zero when $|x|$ is close to the threshold, meaning it includes values on both sides of 0. Benjamini and Yekutieli (2005) discuss and demonstrate the inevitability of this phenomenon, and further study its limiting implications when addressing multiple parameters. This is the motivation behind their introduction of the FCR as a goal in selective inference, rather than the conditional coverage. With this goal in mind, one may use the procedure in Benjamini and Hochberg (1995) for multiple testing at FDR level α to select R parameters, and set $1 - \alpha R/2m$ confidence intervals on the selected ones according to Benjamini and Yekutieli (2005). The confidence intervals on the selected ones will not intrude into the other side of 0, even when the estimator is close to the threshold of significance.

The conditional methods in this paper do not cover the situations where the selection rule for each parameter depends on the value of the estimators of the other parameters as well. Therefore they do not cover the case where the threshold is chosen according to the results of the above FDR controlling testing, or any other step-up or step-down or other adaptive multiple testing procedure controlling whatever criterion. Extension of the current theory to such problems is highly desirable. Still, in large problems, under the asymptotic mixture model where each parameter is either 0 with probability $1 - p_m$ or comes from distribution F_1 with probability p_m and $\lim p_m = p > 0$, the FDR threshold of Benjamini and Hochberg (1995) converges in probability to a constant (e.g. Genovese and Wasserman, 2002). So we can assume the properties of the conditional intervals will not change by much.

In view of that, the challenge that motivated this work is still intriguing if but from a new perspective. We learned that it is possible to select according to a multiple testing procedure that assures FDR control at some level, and avoid undue long confidence intervals when the estimators are very large. But can we achieve both goals and avoid including 0 in the confidence intervals when the estimator is near the significance threshold? It seems that in order to succeed in achieving both goals, some modification of the selection procedure should perhaps be considered.

9 Appendix

9.1 Obtaining the Shortest Acceptance Regions

For $\theta = 0$, $A(\theta)$ is obviously symmetric around 0, because of symmetry and unimodality of g_θ. To obtain the form of the shortest region for $\theta > 0$, we make the following simple observations (wherever $\mathrm{P}(\cdot)$ appears, it is meant that the probability is taken under f_θ). Denote $g_\theta(y) = \phi(y - \theta)$, and let $S_1(\theta) = \{x \in \mathrm{Supp} f_\theta : g_\theta(x) > g_\theta(-c)\} = (c, c + 2\theta)$ and $S_2(\theta) = \{x \in \mathrm{Supp} f_\theta : g_\theta(x) > g_\theta(c)\}$. Then:

a. Any point in the interval $S_1(\theta)$ has higher f_θ density than any point outside this interval.

b. As a function of θ, $\mathrm{P}\left(\{x \in \mathrm{Supp}(f_\theta) : g_\theta(x) > g_\theta(-c)\}\right) = \mathrm{P}\left((c, c+2\theta)\right)$ is strictly increasing and approaches 1 as $\theta \longrightarrow \infty$ ($\theta > 0$).

c. $S_2(\theta)$ is an empty set for $0 < \theta \leq c$, and is a symmetric interval around θ with its left end at c for $\theta > c$.

d. For all $\theta > 0$, $S_2(\theta) \subsetneq S_1(\theta)$ and $\mathrm{P}\left(S_2(\theta)\right) < \mathrm{P}\left(S_1(\theta)\right)$.

e. As a function of θ, $\mathrm{P}\left(S_2(\cdot)\right)$ is (i) continuous, (ii) equals zero on $0 < \theta \leq c$, (iii) strictly increasing for $\theta > c$ and (iv) approaches 1 as $\theta \longrightarrow \infty$.

f. Whenever $S_2(\theta)$ is not empty, each member of $S_2(\theta)$ has a higher f_θ density than any member of $S_1(\theta) \setminus S_2(\theta)$, for any θ.

In the above, (a) is because on the support of f_θ the density of X is proportional to that of Y (for fixed c and θ). (b) is a consequence of (a) and of the fact that g_θ is symmetric and unimodal. For (c), we have

$$\mathrm{P}((c, c+2\theta)) = \frac{\Phi(c+\theta) - \Phi(c-\theta)}{1 - \Phi(c+\theta) + 1 - \Phi(c-\theta)},$$

and both the numerator and the denominator approach 1 as $\theta \longrightarrow \infty$. By taking the derivative, it is easy to verify that $\mathrm{P}(S_1(\theta))$ is strictly increasing on $\theta > 0$.

Assertions (d) and (e) are a consequence of the symmetry and unimodality of g_θ and of the fact that $c > -c$. In (f), (ii) follows trivially from (d); For $\theta > c$, we have

$$\mathrm{P}(S_2(\theta)) = \frac{2\Phi(\theta - c) - 1}{1 - \Phi(c+\theta) + 1 - \Phi(c-\theta)}.$$

As $\theta \longrightarrow \infty$, both the numerator and the denominator approach 1, establishing (iv). Again taking the derivative, we see that $\left[\mathrm{P}\left(S_2(\theta)\right)\right]'$ is strictly increasing on $\theta > c$). Finally, as $\theta \longrightarrow c$ from the right, the numerator tends to 0, while the denominator approaches $2 - 0.5 - \Phi(2c) > 0.5$, hence $h(\theta) \longrightarrow 0$, and together with the fact that h is continuous on $\theta > c$, we have (i).

It follows from (b), (e) and (d) that there exist θ_1 and θ_2, $\theta_1 < \theta_2$, such that $A(\theta_1) = (c, c + 2\theta_1)$, $A(\theta_2) = (c, 2\theta_2 - c)$. By (a), (b) and (e), for each $0 < \theta < \theta_1$, $A(\theta)$ is the union of the interval $(c, c+2\theta)$ and two equally spanned extensions to the right of $c + 2\theta$ and to the left of $-c$ that make the entire probability captured in these components add up to $1 - \alpha$. From (c), (f) and (a), we conclude that for $\theta > \theta_2$, $A(\theta)$ is a symmetric interval around θ with its left end bigger than c.

Now that the form of the acceptance region for each $\theta > 0$ is known, the exact boundaries of the region are obtained by requiring that $\mathrm{P}(A(\theta)) = 1-\alpha$, and (1) results. (b) and (e) assure that θ_1 and θ_2 are indeed the unique solutions to (3) and (4), respectively.

9.2 Obtaining the CMP Acceptance Regions

The discussion in section 3 which precedes the formal statement of the CMP acceptance regions, gives a qualitative description of $A(\theta)$. To obtain the exact expression in (11) we need to determine for which θ values $A(\theta)$ intersects $(-\infty, -c)$ and when it is entirely contained in (c, ∞), and then, distinguishing between these two cases, obtain the desired boundaries of $A(\theta)$ as the proper roots of the respective equations. We observe that:

a. For $\theta > \theta_1$, $A(\theta)$ is contained in (c, ∞).

b. $\mathrm{P}_\theta\left(c, c + r\left|A_{Srt}(\theta)\right|\right)$ is strictly increasing in θ on $0 < \theta < \theta_1$.

c. For $0 < \theta < \theta_1$, $A(\theta)$ intersects $(-\infty, -c)$ if and only if $\mathrm{P}\left(c, c + r\left|A_{Srt}(\theta)\right|\right) < 1 - \alpha$.

d. $\mathrm{P}_\theta([x, x+r\,|A_{Srt}(\theta)|+2c]^c)$ is strictly increasing in x on $x \in (\inf A_{Srt}(\theta), -c)$.

In the above, (a) follows from the fact that for $\theta > \theta_1$, the shortest retention region is contained in (c, ∞). (c) is because for $\theta < \theta_1$, no region contained in (c, ∞) and of total length $r\,|A_{Srt}(\theta)|$ has higher probability than $(c, c + r\,|A_{Srt}(\theta)|)$ (under θ). (b) and (d) can be verified by taking derivatives.

Now, $\mathrm{P}_\epsilon\,(c, c+r\,|A_{Srt}(\epsilon)|) < 1-\alpha$ for some $0 < \epsilon < \theta_1$ because of continuity and the fact that for $\epsilon = 0$, $P_\epsilon(c, \infty) = 0.5 < 1-\alpha$, while $\mathrm{P}_{\theta_1}\,(c, c+r\,|A_{Srt}(\theta_1)|) > 1-\alpha$ because of continuity and since $P_{\theta_1}(c, c + |A_{Srt}(\theta_1)|) = 1-\alpha$. Together with continuity of $\mathrm{P}_\theta\,(c, c+r\,|A_{Srt}(\theta)|)$ in θ, it follows that there exist $0 < \tilde{\theta}_1 < \theta_1$ such that $\mathrm{P}_{\tilde{\theta}_1}\left(c, c+r\left|A_{Srt}(\tilde{\theta}_1)\right|\right) = 1-\alpha$. By (b) we conclude that

$$\begin{cases} \mathrm{P}_\theta\,(c, c+r\,|A_{Srt}(\theta)|) < 1-\alpha, & \theta < \tilde{\theta}_1 \\ \mathrm{P}_\theta\,(c, c+r\,|A_{Srt}(\theta)|) > 1-\alpha, & \theta > \tilde{\theta}_1. \end{cases}$$

It now follows from (a) and (c) that $A(\theta)$ intersects $(-\infty, -c)$ when $0 < \theta < \tilde{\theta}_1$ and is entirely contained in (c, ∞) for $\theta > \tilde{\theta}_1$. Using the discussion in section 3, (11) is true for $g_1(\theta)$ and $g_2(\theta)$ which satisfy (9) and (10), respectively. Finally, $g_2(\theta)$ is unique by definition, while (d) implies that there is indeed a unique solution to (9) in $(\inf A_{Srt}(\theta), -c)$.

Acknowledgements

We are thankful to Professor Larry Brown who, following the presentation of the FCR controlling intervals at the 2005 IMS/Bernoulli meeting in Barcelona, suggested constructing conditional confidence intervals as an alternative. We would also like to thank Dr. Dani Yekutieli for the enlightening discussions on selective inference.

This work was partially funded by grants from the Israeli Science Foundation and the US Department of Defense.

References

[1] Benjamini and Hochberg (1995), *Controlling the False Discovery Rate in multiple comparisons: a new and powerful approach* Journal of the Royal Statistical Society B, 57,289-300.

[2] Benjamini, Y., Hochberg, Y., and Stark, P.B. (1998), *Confidence Intervals with More Power to Determine the Sign: Two Ends Constrain the Means.* Journal of the American Statistical Association, 93,441,309- 317.

[3] Benjamini, Y., and Yekutieli, D. (2005), *False Discovery Rate–Adjusted Multiple Confidence Intervals for Selected Parameters.* Journal of the American Statistical Association, 100, 469, 71-81.

[4] Finner, H. (1994), *Two-sided tests and one sided confidence bounds* The Annals of Statistics, 22, 3,1502-1516.

[5] Gelman, A. and Hill, J. (2009), *Data Analysis Using Regression and Multilevel/Hierarchical Models (2^{nd}ed.).* Cambridge University Press, Cambridge.

[6] Genovese, C. and Wasserman, L. 2002) *Operating Characteristics and extensions of the false discovery rate* Journal of the Royal Statistical Society B, 64,3,499-517.

[7] Hua Zhong and Ross L. Prentice (2008), *Bias-Reduced Estimators and Confidence Intervals for Odds Ratios in Genome-Wide Association Studies.* Biostatistics, 9, 621- 634.

[8] Pratt, J. W. (1961), *Length of Confidence Intervals*. Journal of the American Statistical Association, 56, 549-567.

[9] Tukey, J. (1991) *The Philosophy of Multiple Comparisons*, Statistical Science, 6, 100-116.