〈Research Paper〉

Journal of the Korean
Statistical Society
Vol. 17, No. 2, 1988

On Detecting the Best Treatment+

Woo-Chul Kim*

ABSTRACT

We observe independent random variables $Y_i \sim N(\theta_i, 1)$, $i = 1,2,\cdots,k$, and we are interested in detecting the treatment with the largest θ_i. We consider a procedure which infers $\theta_{(k)} \geq \max_{i \neq (k)} \theta_i$ whenever $Y_{(k)} \geq \max_{i \neq (k)} Y_i + C$. The maximum probability of a false inference is found, and it is shown that the inference can be made with the two-sample *one-sided* critical value for the usual error levels. The result also holds in the case of common unknown variance.

1. Introduction

Suppose we observe independent Y_i, $i=1,\cdots,k$, where Y_i is normally distributed with mean θ_i and variance σ^2. Often, θ_i denotes the average treatment effect, and the treatment with the largest θ_i is called the best treatment. The treatment yielding the largest Y_i can be called the sample best. The problem investigated in this article is the following; When can we infer the sample best as the true best?

The common variance σ^2 may be either known or unknown. In the case of unknown variance, we assume that an estimator $\hat{\sigma}^2$ of σ^2 is available so that $\nu\hat{\sigma}^2 / \sigma^2$ has a chi-square distribution with ν degrees of freedom, independently of $Y_1, \cdots, Y_k$.

Suppose we infer $\theta_{(k)} \geq \max_{i \neq (k)} \theta_i$ whenever $Y_{(k)} \geq \max_{i \neq (k)} Y_i + C$, where $C = c\sigma$ or $c\hat{\sigma}(c>0)$ according as σ is known or unknown. Then the probability of a false inference is

$$P_{\underline{\theta}}(\theta_{(k)} < \max_{i \neq (k)} \theta_i,\ Y_{(k)} \geq \max_{i \neq (k)} Y_i + C) \tag{1.1}$$

+ This research is supported by Korea Research Foundation, 1987-1989.

* Department of Computer Science and Statistics, Seoul National University, Seoul, 151-742, Korea

where $\underline{\theta}=(\theta_1, \cdots, \theta_k)$, and the probability of a correct inference is

$$P_{\underline{\theta}}\left(\theta_{(k)} \geqq \max_{i \neq (k)} \theta_i,\ Y_{(k)} \geqq \max_{i \neq (k)} Y_i + C\right). \tag{1.2}$$

Subject to ensuring that $(1.1) \leqq \alpha$ for all $\underline{\theta}$, we want to make (1.2) large.

The above procedure has been considered by Zinger and St-Pierre(1958), Zinger (1961), Gutmann(1985) and Bofinger(1986). All these works were intended to find the maximum error probability in (1.1), but have not been successful in generality even for the case of known variance.

The maximum error probability is found in Section 2. It is also shown that the inference can be made with the two-sample *one-sided* critical value. The sample size aspect is considered in Section 3. An illustrative example and comparisons with previous results are provided in Section 4.

Finally, it should be mentioned that the inference $\theta_{(k)} \geqq \max_{i \neq (k)} \theta_i$ differs from an inference $\theta_{(k)} > \max_{i \neq (k)} \theta_i$. The latter type of inference was considered in Gutmann and Maymin(1987), and Stefansson, Kim and Hsu(1988). Gutmann and Maymin(1987) has shown that the stronger inference $\theta_{(k)} > \max_{i \neq (k)} \theta_i$ is possible with the two-sample *two-sided* cricial value.

2. The Error Probability

In the case of known variance, we may assume $\sigma=1$ without loss of generality. Due to the symmetry of the problem, we may assume $\theta_1 \leqq \theta_2 \cdots \leqq \theta_k$ in finding the maximum error probability in (1.1).

The following lemma is needed to find the maximum error probability. The proof is given in the Appendix A.

Lemma 2.1. The error probability is maximized when (i) there are (k-1) non-best treatments, (ii) the second best treatment θ_{k-1} approaches the best θ_k, and (iii) some of $\theta_1, \cdots, \theta_{k-2}$ approach θ_k and the others approach $-\infty$.

The maximum error probability in the case of known variance is given in the next result.

Theorem 2.1. In the case of known variance, the maximum error probability is

$$\max_{1 \leq r \leq k-1} \left\{ r \int_{-\infty}^{\infty} \Phi^r(y-c)\, \phi(y)\, dy \right\}, \tag{2.1}$$

where Φ and ϕ are the cdf and pdf of the standard normal distribution, respectively.

Proof. When $\theta_k > \theta_{k-1} \geqq \cdots \geqq \theta_1$, the error probability is given by

$$P_{\underline{\theta}}(\theta_{(k)} < \max_{i \neq (k)} \theta_i,\ Y_{(k)} \geqq \max_{i \neq (k)} Y_i + C)$$

$$= \sum_{j=1}^{k-1} P_{\underline{\theta}}(Y_j \geqq \max_{i \neq j} Y_i + C)$$

$$= \sum_{j=1}^{k-1} \int_{-\infty}^{\infty} \prod_{\substack{i=1 \\ i \neq j}}^{k} \Phi(y-\theta_i-c)\ \phi(y-\theta_j)\,dy$$

It is clear that this error probability approaches

$$r\int_{-\infty}^{\infty} \Phi^r(y-c)\,\phi(y)\,dy$$

as $\theta_{k-1}, \theta_{k-2}, \cdots, \theta_{k-r} \to \theta_k$ and $\theta_{k-r-1}, \cdots, \theta_1 \to -\infty$. Thus, the result follows from Lemma 2.1. ■

In computing (2.1), the following lemma is useful. Its proof is given in the Appendix A.

Lemma 2.2. Let $A_r = r\int_{-\infty}^{\infty} \Phi^r(y-c)\,\phi(y)dy$, then $A_{r+1} - A_r$ changes sign at most once from $+$ to $-$ as r increases from 1 to (k-1).

The next corollary follows from Theorem 2.1 and Lemma 2.2.

Corollary 2.1. In the case of known variance, the maximum error probability is given by

$$\int_{-\infty}^{\infty} \Phi(y-c)\,\phi(y)\,dy = \Phi(-c/\sqrt{2}) \tag{2.2}$$

provided

$$\int_{-\infty}^{\infty} \Phi(y-c)\,\phi(y)\,dy \geqq 2\int_{-\infty}^{\infty} \Phi^2(y-c)\,\phi(y)\,dy. \tag{2.3}$$

It was found numerically that (2.3) holds for $c \geqq 0.8568$ or equivalently for $\Phi(-c/\sqrt{2}) \leqq 0.2723$. Thus, Corollary 2.1 implies that the inference can be made with the two-sample *one-sided* critical value for any pre-specified level $\alpha \leqq 0.2723$.

When the common variance σ^2 is unknown, the maximum of the conditional error probability, given $\hat{\sigma}^2$, can be found in exactly the same way. Thus the next result follows.

Theorem 2.2. In the case of common unknown variance, the maximum error probability is bounded above by

$$\int_0^{\infty} \max_{1 \le r \le k-1} \{ r \int_{-\infty}^{\infty} \Phi^r(y-cu)\phi(y)\,dy \}\, dF_\nu(u), \tag{2.4}$$

where F_ν is the cdf of $\sqrt{\chi^2(\nu)/\nu}$.

To guarantee a pre-specified level α, We find $c=c_\alpha$ by equating (2.4) to be α. We have computed c_α values for α=0.01, 0.05, 0.10, k=3(1)12, 15, 20, 30, 50 and for ν=5(1)20, 24, 30, 40, 60, 120, 250. It turned out that, except for α=0.10 and ν=5,6,7, the values of c_α coincide, up to sixth decimal places, with the two-sample *one-sided* critical values, i.e.,

$$c_\alpha/\sqrt{2} = t_\alpha(\nu)$$

where $t_\alpha(\nu)$ is the upper α-quantile of the t-distribution with ν degrees offreedom.

For α=0.10 and ν=5,6,7, the values of $c_\alpha/\sqrt{2}$ are given in Table 1. It can be observed that, even in these cases, these values are very close to $t_\alpha(\nu)$. Details of the computation are given in Appendix B.

These observations indicate that, even in the case of unknown variance, the two-sample *one-sided* critical value $c_\alpha=\sqrt{2}\,t_\alpha(\nu)$ can be used at most of practical levels.

Table 1. Values of $c_\alpha/\sqrt{2}$ and $t_\alpha(\nu)$ for $\alpha=0.10$

ν \ k	3	4	5	6	7
5	1.481473	1.481897	1.481970	1.481993	1.482000
6	1.442696	1.442864	1.442885	1.442890	1.442891
7	1.416484	1.416555	1.416563	1.416564	1.416564

ν \ k	8	9	10	$t_{0.10}(\nu)$
5	1.482002	1.482003	1.482004	1.476
6	1.442891	1.442891	1.442891	1.440
7	1.416564	1.416564	1.416564	1.415

Finally, it should be mentioned that, in the case of known variance, Bofinger(1986) obtained an upper bound on the maximum error probability. Then, she could obtain a result similar to Corollary 2.1 through some numerical computations.

3. The Sample Size Aspect

The probability of a correct inference in(1.2) is analogous to the power in a standard hypothesis testing. In this section, we consider a method to determine the sample size to control the probability of a correct inference.

We consider the following one-way balanced model with normal errors e_{ij};

$$X_{ij} = \theta_i + e_{ij},\ j = 1, \cdots, n\ ;\ i = 1, \cdots, k. \tag{3.1}$$

Let $\hat{\sigma}^2$ denote the mean squared error with $\nu = k(n-1)$ degrees of freedom.

Obviously, the results in Section 2 can be applied to this model by taking $Y_i = \sqrt{n}\,\bar{X}_i$, $i = 1, \cdots, k$. We control the probability of a correct inference when the best treatment is significantly better than the others, i.e. $\theta_{[k]} - \theta_{[k-1]} \geqq \delta\sigma$ where $\delta > 0$ is specified before the experiment and $\theta_{[1]} \leqq \cdots \leqq \theta_{[1]}$ denote the ordered $\theta_1, \cdots, \theta_k$.

For $\theta_{[k]} - \theta_{[k-1]} \geqq \delta\sigma$, the probability of a correct inference is given by

$$P_\theta(X_{[k]} = \max_{1 \leq i \leq k} X_i,\ X_{[k]} \geqq \max_{i \neq [k]} X_i + c\sigma/\sqrt{n})$$

$$= \int_0^\infty \int_{-\infty}^\infty \prod_{i=1}^{k-1} \Phi(y + \sqrt{n}\,(\theta_{[k]} - \theta_{[i]})/\sigma - cu\}\ \phi(y)\,dy\ \ dF_\nu(u)\ .$$

It is easy to observe that the minimum probability of a correct inference for $\theta_{[k]} - \theta_{[k-1]} \geqq \delta\sigma$ is given by

$$\int_0^\infty \int_{-\infty}^\infty \Phi^{k-1}(y + \sqrt{n}\,\delta - cu)\ \phi(u)\,dy\ \ dF_\nu(u).$$

For given k, α and $\delta > 0$, the minimum probability of a correct inference(3.1) can be computed as a function of n. The results for $\alpha = 0.05$, $k = 3, 6, 10$ and $\delta = 0.5, 1.0, 1.5$ are given in Figure 1 for illustrative purposes. The experimenter specifies the separation between the best and the others as $\theta_{[k]} - \theta_{[k-1]} \geqq \delta\sigma$ with a plausible guess for σ. Then, he or she can find the necessary sample size n for desired probability of a correct inference.

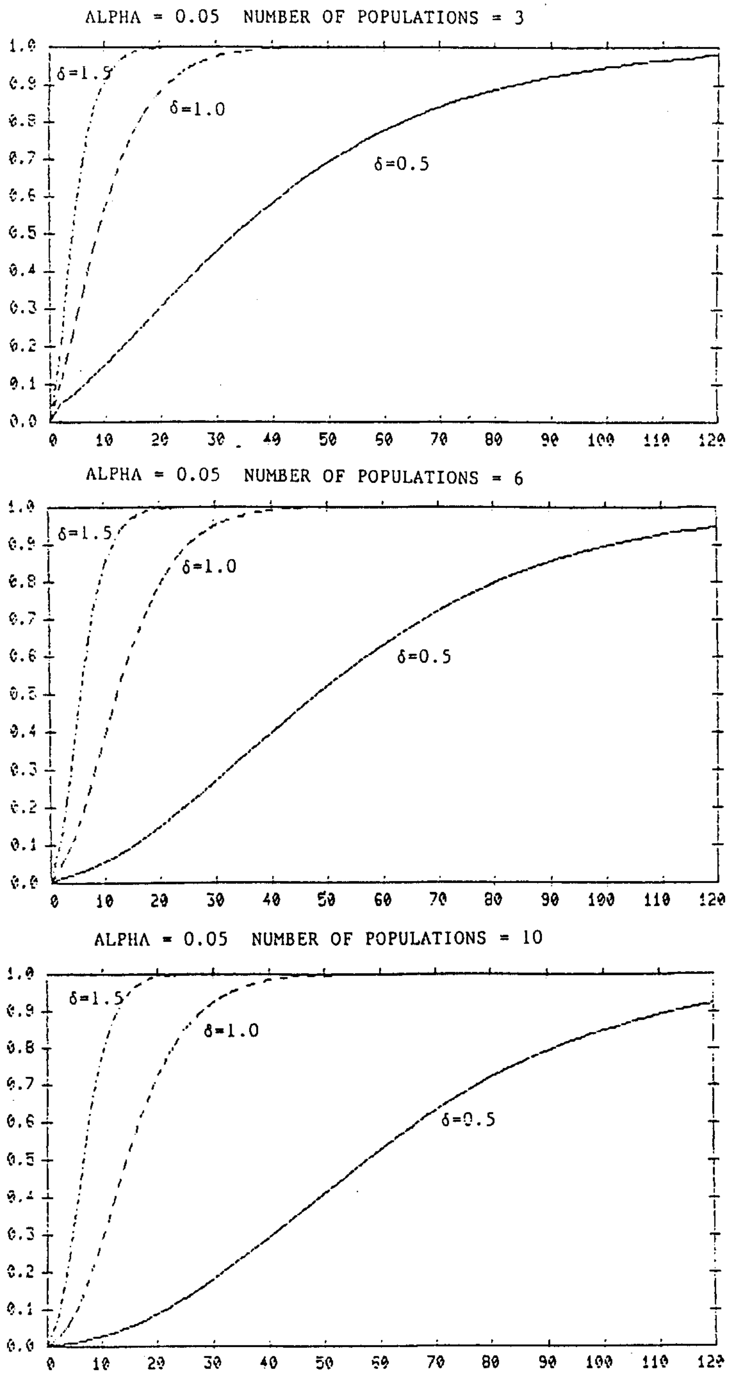

Figure 1. Sample size and the probability of a correct inference

4. An Example

As an illustration, consider an example cited in Steel and Torrie(1960), in which the effects of k=6 inoculation methods for red clover plants were compared in terms of nitrogen contents. The data are given in Table 2.

Table 2. The effects of inoculation methods

Treatments	Nitrogen contents in milligrams				
1	19.4	32.6	27.0	32.1	33.0
2	17.7	24.8	27.9	25.2	24.3
3	17.0	19.4	9.1	11.9	15.8
4	20.7	21.0	20.5	18.8	18.6
5	14.3	14.4	11.8	11.6	14.2
6	17.3	19.4	19.1	16.9	20.8

From Table 2, we find sample means and the pooled sample variance as follows;

$\bar{X}_1$	$\bar{X}_2$	$\bar{X}_3$	$\bar{X}_4$	$\bar{X}_5$	$\bar{X}_6$	$\hat{\sigma}^2$
28.82	23.98	14.64	19.92	13.26	18.70	11.79

Under the usual one-way model with normal errors, the homogeneity hypothesis is rejected even at 1%. Then a natural question is whether the treatment 1 yielding the largest sample mean can be inferred as the best.

The p-value can be found by evaluating the integral(2.4) with $c=\sqrt{n}(\bar{X}_1-\max_{i\neq 1}\bar{X}_i)/\hat{\sigma}=3.15$. The p-value in this example is found to be $\hat{\alpha}=0.018$. Thus at level $\hat{\alpha}=0.018$, we can infer that $\theta_1\geqq\max_{i\neq 1}\theta_i$.

By comparison, it can be verified that Hsu's(1984) method for simultaneous comparisons with the best can not make this last assertion at $\alpha=0.018$. The reason is that Hsu's method is not specifically designed for this type of inference.

It should also be noted that the methods of Stefansson, Kim and Hsu(1988), and of Gutmann and Maymin(1987) can be used to make an inference $\theta_1>\max_{i\neq 1}\theta_i$ at $\alpha=0.05$. It can be, however, verified that none of these methods can make such an assertion at $\alpha=0.018$.

References

(1) Bofinger, E.(1986), The Least Significant Difference for 'one versus the rest' Normal Populations , Unpublished manuscript, University of New England, Australia.

(2) Gutmann, S.(1985), Is the Selected Population Unsurpassed? (Three or Four Normal Populations) , Unpublished manuscript, Northeastern University, U.S.A.

(3) Gutmann, S., and Maymin, Z.(1987), Is the Selected Population the Best? , *Annals of Statistics,* 15, 456-461.

(4) Hsu, J.C.(1984), Constrained Two-sided Simultaneous Confidence Intervals for Multiple Comparisons with the Best , *Annals of Statistics,* 12, 1135-1144.

(5) Lehmann, E.L.(1986), *Testing Statistical Hypotheses,* second edition, New york: John Wiley.

(6) Steel, R.G.D. and Torrie, J.H.(1960), *principles and Procedures of Statistics,* New York: McGraw-Hill.

(7) Stefansson, G., Kim, W.C., and Hsu, J.C.(1988), On Confidence Sets In Multiple Comparisons , *Statistical Decision Theory and Related Topics* IV (ed. S.S. Gupta and J.O. Berger), Vol. 2, 89-104.

(8) Zinger, A.(1961), Detection of Best and Outlying Normal Populations With Known Variances , *Biometrika,* 48, 457-461.

(9) Zinger, A., and St-Pierre, J.(1958), On the Choice of the Best Among Three Normal Populations With Known Variances , *Biometrika,* 45, 436-446.

Appendix A: Proofs of Lemmas

By the symmetry, we assume $\theta_1 \leqq \theta_2 \leqq \cdots \leqq \theta_k$ in the sequel. Let $\Omega_r (r=1,2,\cdots.k-1)$ denote the set of parameters where there are exactly r non-best treatments, i.e., $\Omega_r = \{\theta : \theta_1 \leqq \cdots \leqq \theta_r < \theta_{r+1} = \cdots = \theta_k\}$.

First, note that the error probability in (1.1) is 0 when all θ_i's are equal. Thus, in finding the maximum error probability, we may assume. $r \geqq 1$.

For $\underline{\theta} \epsilon \Omega_r$, the error probability in (1.1) is given by.

$$P_{\underline{\theta}} \left(\max_{1 \leq j \leq r} Y_j = Y_{(k)} \geqq \max_{i \neq (k)} Y_i + c \right),$$

which is obviously non-increasing in $\theta_{r+1}= \cdots =\theta_k$. Thus, it is maximized as $\theta_{r+1}= \cdots =\theta_k$ decreases to θ_r. Therefore for r=1,2,⋯,k-1, we have

$$\sup_{\underline{\theta}\varepsilon\Omega_r} P_{\underline{\theta}}(\max_{1\le j\le r} Y_j=Y_{(k)} \geqq \max_{i\ne(k)} Y_i + c)$$

$$= \sup_{\underline{\theta}\varepsilon\bar{\Omega}_{r-1}} P_{\underline{\theta}}(\max_{1\le j\le r} Y_j=Y_{(k)} \geqq \max_{i\ne(k)} Y_i + c)$$

$$\geqq \sup_{\underline{\theta}\varepsilon\bar{\Omega}_{r-1}} P_{\underline{\theta}}(\max_{1\le j\le r-1} Y_j=Y_{(k)} \geqq \max_{i\ne(k)} Y_i + c)$$

$$\geqq \sup_{\underline{\theta}\varepsilon\Omega_{r-1}} P_{\underline{\theta}}(\max_{1\le j\le r-1} Y_j=Y_{(k)} \geqq \max_{i\ne(k)} Y_i + c)$$

where $\bar{\Omega}_{r-1}$ denotes the closure of Ω_{r-1}.

Hence the maximum error probability over Ω_r is non-decreasing in r=1,2,⋯k-1. Thus, the result (i) in Lemma 2.1 follows.

It follows from Lemma 2.1(i) and the translation invariance that we may assume $\theta_1 \leqq \cdots \leqq \theta_{k-1} < \theta_k = 0$. Let $G_{k-1}(\theta_1, \cdots, \theta_{k-1})$ denotes the error probability in (1.1) for $\theta_1 \leqq \cdots \leqq \theta_{k-1} < \theta_k = 0$, i.e.,

$$\begin{aligned} &G_{k-1}(\theta_1, \cdots, \theta_{k-1}) \\ &= P_{\underline{\theta}}(\max_{1\le j\le k-1} Y_j=Y_{(k)} \geqq \max_{i\ne(k)} Y_i + c) \\ &= \sum_{j=1}^{k-1} \int_{-\infty}^{\infty} \prod_{\substack{i=1\\ i\ne j}}^{k-1} \Phi(y-\theta_i-c)\,\Phi(y-c)\,\phi(y-\theta_j)\,dy\,. \end{aligned} \tag{A.1}$$

Denoting θ_{k-1} by θ, we have from (A.1)

$$\begin{aligned} &\frac{\partial}{\partial\theta} G_{k-1}(\theta_1, \cdots, \theta_{k-2}, \theta) \\ &= \int_{-\infty}^{\infty} \prod_{i=1}^{k-2} \Phi(y-\theta_i-c)\,\phi(y-c)\,\phi(y-\theta)\,dy \\ &+ \sum_{j=1}^{k-1} \int_{-\infty}^{\infty} \prod_{\substack{i=1\\ i\ne j}}^{k-2} \Phi(y-\theta_i-c)\,\phi(y-c) \\ &\{\phi(y-\theta_j-c)\,\phi(y-\theta)-\phi(y-\theta-c)\,\phi(y-\theta_j)\}\,dy\,, \end{aligned}$$

which is non-negative for $\theta_1 \leqq \cdots \leqq \theta_{k-2} \leqq \theta < 0$. Therefore the result (ii) in Lemma 2.1 follows.

To prove the result (iii) in lemma 2.1, we need the follwing;
For $l=1,2,\cdots,k\text{-}2$,

$$G_{k-1}(\theta_1,\cdots\theta_l,0,\cdots,0)$$

$$=(k-1-l)\int_{-\infty}^{\infty}\prod_{i=1}^{l}\Phi(y-\theta_i-c)\Phi^{k-1-l}(y-c)\phi(y)\,dy \tag{A.2}$$

$$+\sum_{j=1}^{l}\int_{-\infty}^{\infty}\prod_{\substack{i=1\\ i\neq j}}^{l}\Phi(y-\theta_i-c)\Phi^{k-l}(y-c)\phi(y-\theta_j)\,dy,$$

and

$$\frac{\partial}{\partial\theta_1}G_{k-1}(\theta_1,\cdots;\theta_l,0,\cdots;0)$$

$$=\{(k-l)e^{\theta_1 c}-(k-1-l)\}\int_{-\infty}^{\infty}\prod_{i=2}^{l}\Phi(y-\theta_1-c)\Phi^{k-1-l}(y-c) \tag{A.3}$$

$$\phi(y-\theta_1-c)\phi(y)\,dy$$

$$+\sum_{j=2}^{l}\int_{-\infty}^{\infty}\prod_{\substack{i=2\\ i\neq j}}^{l}\Phi(y-\theta_i-c)\Phi^{k-l}(y-c)$$

$$\{\phi(y-\theta_j-c)\phi(y-\theta_1)-\phi(y-\theta_1-c)\phi(y-\theta_j)\}dy.$$

It can be verified from (A.2) and (A.3) that

$$\sup\{G_{k-1}(\theta_1,\cdots,\theta_l,0,\cdots,0):\theta_1\leqq\cdots\leqq\theta_l\leqq\frac{1}{c}\log(\frac{k-1-l}{k-l})\}$$

$$=\lim_{\theta_1\to-\infty}\cdots\lim_{\theta_l\to-\infty}G_{k-1}(\theta_1,\cdots,\theta_l,0,\cdots,0)$$

and that

$$\sup\{G_{k-1}(\theta_1,\cdots,\theta_l,0,\cdots,0):\theta_1\leqq\cdots\leqq\theta_l,\frac{1}{c}\log(\frac{k-1-l}{k-l})<\theta_l<0\}$$

$$=\sup\{G_{k-1}(\theta_1,\cdots,\theta_{l-1},0,\cdots,0):\theta_1\leqq\cdots\leqq\theta_{l-1}<0\}.$$

Therefore, the result (iii) in Lemma 2.1 follows by applying the same arguments iteratively w.r.t. ℓ. This completes the proof of Lemma 2.1.

To prove Lemma 2.2, we note that the integration by parts can be applied to get the following;

$$A_r = r\int_{-\infty}^{\infty}\Phi^r(y-c)\,\phi(y)\,dy$$

$$= e^{-\frac{1}{2}c^2}\int_{-\infty}^{\infty}\Phi^r(y)\{c\Phi(y)-\phi(y)\}e^{-cy}\,dy.$$

Thus we have, after a little algebra,

$$A_{r+1}-A_r = e^{-\frac{1}{2}c^2}\int_{-\infty}^{\infty}\Phi^r(y)\left(\frac{\phi(y)}{\Phi(y)}-c\right)(1-\Phi(y))\,\Phi(y)\,e^{-cy}\,dy.$$

Hence Lemma 2.2 follows from the log-concavity of $\Phi(y)$ and the TP_2 property of $\Phi^r(y)$ (see, e.g., Lehmann 1986).

Appendix B: Details of Computation

The integration in (2.4) was evaluated by Gauss-Hermite and Gauss-Legendre quadrature, and the critical value c_α was found via the modified regula falsi method with accuracy up to 10^{-7}.