

Comparison of normal distribution based and nonparametric decision limits on the GH-2000 score for detecting growth hormone misuse (doping) in sport

W. Liu^{a*}, F. Bretz^b, D. Böhning^a, R. Holt^c,
W. Böhning^c, N. Guha^d, P. Sönksen^c, D. Cowan^e

^aS3RI and School of Mathematical Sciences
University of Southampton, SO17 1BJ, UK

^bNovartis Pharma AG
Basel, 4002, Switzerland

^cHuman Development and Health Academic Unit
Faculty of Medicine, Southampton General Hospital
University of Southampton, SO16 6YD, UK

^dChemical Pathology and Metabolic Medicine Department of
Clinical Biochemistry, John Radcliffe Hospital
Oxford, OX3 9DU, UK

^e Drug Control Centre, Department of Pharmacy & Forensic Science
King's College London, London SE1 9NH, UK

* w.liu@maths.soton.ac.uk

Abstract

This paper is motivated by the GH-2000 biomarker test, though the discussion is applicable to other diagnostic tests. The GH-2000 biomarker test has been developed as a powerful technique to detect growth hormone misuse by athletes, based on the GH-2000 score. Decision limits on the GH-2000 score have been developed and incorporated into the guidelines of the World Anti-Doping Agency (WADA). These decision limits are constructed, however, under the assumption that the GH-2000 score follows a normal distribution. As it is difficult to affirm the normality of a distribution based on a finite sample, nonparametric decision limits, readily available in the statistical literature, are viable alternatives. In this paper, we compare the normal distribution based and nonparametric decision limits. We show that the decision limit based on the normal distribution may deviate significantly from the nominal confidence level $1 - \alpha$ or nominal false positive rate γ when the distribution of the GH-2000 score departs only slightly from the normal distribution. While a nonparametric decision limit does not assume any specific distribution of the GH-2000 score and always guarantees the nominal confidence level and false positive rate, it requires a much larger sample size than the normal distribution based decision limit. Due to the stringent false positive rate of the GH-2000 biomarker test used by WADA, the sample sizes currently available are much too small, and it will take many years of testing to have the minimum sample size required, in order to use the nonparametric decision limits. Large sample theory about the normal distribution based and nonparametric decision limits is also developed in this paper to help understanding their behaviours when the sample size is large.

Keywords: Asymptotic distribution; Decision Limits; GH-2000 score; Growth hormone misuse detection; Nonparametric methods; Tolerance intervals; Tolerance limits.

1 Introduction

Growth hormone is an endogenous anabolic agent of considerable therapeutic value but also misused in sport for its anabolic and lipolytic properties (cf. Holt, 2009). In order to preserve the fairness of competition, its use is prohibited by the World Anti-Doping Agency (WADA) (WADA, 2016). Two methods are presently available and approved by WADA to detect its misuse: the isoform test developed by Bidlingmaier *et al.* (2000) (see also WADA, 2014, 2016) and the GH-2000 biomarker test developed by the GH-2000 and GH-2004 projects (cf. Holt *et al.*, 2015). The latter method depends on the measurements of two growth hormone (GH) sensitive biomarkers, insulin-like growth factor-I (IGF-I) and the amino-terminal pro-peptide of type III collagen (P-III-NP), both of which rise in response to exogenous GH administration (cf. Longobardi *et al.*, 2000, and Dall *et al.* 2000). The measurements are used, adjusted for sex and age of an athlete, to calculate a score, the GH-2000 score, based on which the compliance of an athlete's sample is determined. The biomarkers are measured by either immunoassay or mass spectrometry. WADA has approved three IGF-I assays and two P-III-NP assays. Each assay generates slightly different results and so the decision limits for the GH-2000 score are assay-specific. The decision limits of both the isoform and GH-2000 biomarker test have been developed using parametric statistics with the aim of controlling the false positive rate (FPR) to less than 1 in 10,000. In this paper, we will use data collected by the GH-2004 team to discuss and illustrate the use of decision limits.

Let X denote an assay-specific GH-2000 score of an athlete without GH misuse and assume that X is a continuous random variable with cumulative distribution function (cdf) $F(\cdot)$ and probability density function (pdf) $f(\cdot)$. WADA (2016) requires that the FPR is controlled at a pre-specified small value γ , leading to a critical value $c = F^{-1}(1 - \gamma)$ such that

$$P\{X > c\} = 1 - F(c) = \gamma, \quad (1)$$

where F^{-1} denotes the inverse or quantile function of F . Currently γ is set at $1/10,000 = 0.0001$ by WADA (2016).

When $F(\cdot)$ is given by the normal distribution $N(\mu, \sigma^2)$, we have $c = \mu + \Phi^{-1}(1 - \gamma)\sigma$, which is equal to $\mu + 3.72\sigma$ for $\gamma = 0.0001$; $\Phi(\cdot)$ and $\phi(\cdot)$ denote, respectively, the cdf and pdf of the standard normal distribution $N(0, 1)$. Clearly, had μ and σ been known, $c = \mu + \Phi^{-1}(1 - \gamma)\sigma$ would have been used as the decision limit as it would give the required FPR γ exactly. However, μ and σ are unknown and need to be estimated by their sample estimators \bar{X} and S , the sample mean and sample standard deviation, respectively. Furthermore the estimated critical value $\hat{c} = \bar{X} + \Phi^{-1}(1 - \gamma)S$ is random and it is not guaranteed that $\bar{X} + \Phi^{-1}(1 - \gamma)S \geq \mu + \Phi^{-1}(1 - \gamma)\sigma$, which is desirable in order to guarantee the FPR γ . Hence it is necessary to incorporate this uncertainty into the construction of the decision limit. This can be accomplished by using an additional uncertainty term rS , where r is a suitably chosen constant, such that

$$P\{\bar{X} + \Phi^{-1}(1 - \gamma)S + rS \geq \mu + \Phi^{-1}(1 - \gamma)\sigma\} = 1 - \alpha. \quad (2)$$

That is, we aim to construct a decision limit $\bar{X} + \Phi^{-1}(1 - \gamma)S + rS$ which has a $1 - \alpha$ (large) confidence level of being greater than the critical value $c = \mu + \Phi^{-1}(1 - \gamma)\sigma$ associated with FPR γ . Usually $1 - \alpha$ is set at 0.95, which is also used currently by WADA.

Holt *et al.* (2015) apply the Δ -method to provide a formula for the constant r so that the probability in (2) is approximately equal to $1 - \alpha$ when the sample size of the available sample (for estimating μ and σ) is large. Realizing from (2) that $\bar{X} + \Phi^{-1}(1 - \gamma)S + rS$ is a $1 - \alpha$ upper confidence bound for $c = \mu + \Phi^{-1}(1 - \gamma)\sigma$ or a $(1 - \gamma)$ -content and $(1 - \alpha)$ -confidence upper tolerance limit for the normal distribution $N(\mu, \sigma^2)$, Böhning *et al.* (2019) point out, by appealing to the available results in the statistical literature (cf. Krishnamoorthy and Mathew, 2009; Liu *et al.*, 2013; Meeker *et al.*, 2017), that the decision limit $\bar{X} + \Phi^{-1}(1 - \gamma)S + rS$ satisfying (2) exactly is given by $\bar{X} - T_{n-1; -\sqrt{n}z_\gamma}^{-1}(\alpha)S/\sqrt{n}$, where n is the sample size, $z_\gamma = \Phi^{-1}(1 - \gamma)$, and $T_{\nu; \delta}^{-1}(q)$ is the q -quantile of the non-central t-distribution with ν degrees of freedom and non-centrality parameter δ .

Note, however, the decision limit $\bar{X} - T_{n-1; -\sqrt{n}z_\gamma}^{-1}(\alpha)S/\sqrt{n}$ hinges on the assumption that the GH-2000 score X has a normal distribution $N(\mu, \sigma^2)$. While there are various statistical tests for detecting the departure of the distribution $F(\cdot)$ from the normal distribution (see, e.g., Chantarangsi *et al.*, 2015, and the references therein), a non-rejection of the null hypothesis that $F(\cdot)$ is a normal distribution does not imply that $F(\cdot)$ is a normal distribution. When $F(\cdot)$ is not normal, the confidence level that the decision limit $\bar{X} - T_{n-1; -\sqrt{n}z_\gamma}^{-1}(\alpha)S/\sqrt{n}$ is larger than $F^{-1}(1 - \gamma)$ could be much smaller than the nominal level $1 - \alpha$, that is, equivalently (see Section 2 below), the probability of the FPR being larger than γ could be much larger than α . For example, shown in Section 2, when $F(\cdot)$ is the t distribution with 30 degrees of freedom (df), which differs little from $N(0, 1)$, the confidence level is 0.010 for $\gamma = 0.001$, $\alpha = 0.05$ and $n = 3100$, which is much smaller than $1 - \alpha = 0.95$. Hence, in this case, the probability of the FPR being larger than $\gamma = 0.001$ is $1 - 0.010 = 0.990$, much larger than $\alpha = 0.05$. It is therefore attractive to use a nonparametric decision limit, which guarantees the nominal confidence level or FPR irrespective of the specific form of $F(\cdot)$.

A nonparametric decision limit can be constructed from the order statistics. Let X_1, \dots, X_n denote the available random sample from the population with cdf $F(\cdot)$, and $X_{[1]} < \dots < X_{[n]}$ the order statistics of X_1, \dots, X_n . It is well-known (cf. Krishnamoorthy and Mathew, 2009) that $F(X_1), \dots, F(X_n)$ are independent, each having a uniform distribution on the interval $(0, 1)$, and that $F(X_{[k]})$ has a beta distribution with parameters k and $n - k + 1$ under the only assumption that X_i is a continuous random variable.

Similar to the requirement in (2), a nonparametric decision limit $D(X_1, \dots, X_n)$ satisfies

$$P \left\{ D(X_1, \dots, X_n) \geq F^{-1}(1 - \gamma) \right\} \geq 1 - \alpha. \quad (3)$$

From this, one realizes that $D(X_1, \dots, X_n)$ is just a $(1 - \gamma)$ -content and $(1 - \alpha)$ -confidence (nonparametric) upper tolerance limit for the distribution $F(\cdot)$. Nonparametric tolerance limits/regions are studied first in Wilks (1941) and extended in Wald (1943) and Tukey (1947) among many others. Excellent overviews are provided in Guttman (1970, 2006), Krishnamoorthy and Mathew (2009) and Meeker *et al.* (2017). The R package `tolerance` of Young (2010) allows the computation of many tolerance intervals/regions. It is known (cf. Krishnamoorthy and Mathew, 2009) that one can set $D(X_1, \dots, X_n) = X_{[k]}$, where the natural number k is chosen such that

$$P \left\{ X_{[k]} \geq F^{-1}(1 - \gamma) \right\}$$

$$= P \{F(X_{[k]}) \geq 1 - \gamma\} \tag{4}$$

$$= 1 - B_{k,n-k+1}(1 - \gamma) \geq 1 - \alpha \tag{5}$$

where $B_{k,n-k+1}(\cdot)$ denotes the cdf of the beta distribution with parameters k and $n - k + 1$.

Note that there may not exist a natural number k such that $1 - B_{k,n-k+1}(1 - \gamma) \geq 1 - \alpha$ for given α, γ and n . The existence of such a natural number k (that satisfies the inequality in (5)) is guaranteed if and only if the inequality in (5) is satisfied with the largest possible value of k given by $k = n$, that is, $1 - B_{n,1}(1 - \gamma) \geq 1 - \alpha$, which is equivalent to $n \geq \ln(\alpha)/\ln(1 - \gamma)$ (cf. Krishnamoorthy and Mathew, 2009).

For $\gamma = 0.0001$ and $\alpha = 0.05$ used currently by WADA, we have $\ln(\alpha)/\ln(1 - \gamma) = 29955.8$ and so the smallest sample size required is $n = 29956$. Hence if $n = 29956$ then the largest order statistic $X_{[n]}$ is a 0.9999-content and 0.95-confidence upper tolerance limit for $F(\cdot)$, and if $n < 29956$ then the largest order statistic $X_{[n]}$ is no longer a 0.9999-content and 0.95-confidence upper tolerance limit for $F(\cdot)$. While reflecting the stringent requirement of the very small FPR $\gamma = 0.0001$ used currently by WADA, this sample size is extremely large considering the sample sizes of the currently available samples in GH-2000 study (Holt *et al.*, 2015) are less than 1,000. For $n = 1000$, for example, $X_{[n]}$ is a 0.9970-content and 0.95-confidence upper tolerance limit, and a 0.9999-content and 0.0952-confidence upper tolerance limit. Even with the latest available data, the sample size is $n = 5053$, still much smaller than the minimum sample size $n = 29956$ required. For $n = 5053$, $X_{[n]}$ is a 0.9994-content and 0.95-confidence upper tolerance limit, and a 0.9999-content and 0.3967-confidence upper tolerance limit.

At the current speed of accruing observations by WADA, it will take many years to reach the required minimum sample size $n = 29956$ for $\gamma = 0.0001$ and $\alpha = 0.05$. Hence, for the purpose of illustration, we will use $n = 5053$ based on the latest available data with the less stringent values $\gamma = 0.001$ and $\alpha = 0.05$.

It should be emphasized that $n \geq \ln(\alpha)/\ln(1 - \gamma)$ is a requirement on the sample size so that a $(1 - \gamma)$ -content and $(1 - \alpha)$ -confidence upper tolerance limit of the form $X_{[k]}$ can be constructed. When $n \geq \ln(\alpha)/\ln(1 - \gamma)$, there may exist more than one natural number k that satisfies the inequality in (5). In this case, the smallest natural number k that satisfies the inequality in (5) should be used, and the corresponding $X_{[k]}$ is the nonparametric decision limit used in the rest of the paper. This smallest k can be computed straightforwardly from the inequality in (5) since $B_{k,n-k+1}(\cdot)$ can be routinely computed in most statistical packages such as R by using the built-in incomplete beta function, or by using the R package *tolerance* of Young (2010).

For example, with $\gamma = 0.001$, $\alpha = 0.05$ and $n = 5053$ (which is larger than $\ln(\alpha)/\ln(1 - \gamma) = 2994.2$), the smallest k that satisfies the inequality in (5) is given by $k = 5052$, and so $X_{[5052]} = 10.18$ is the required nonparametric decision limit. The normal distribution based decision limit in this case is $\bar{X} - T_{n-1; -\sqrt{n}z_\gamma}^{-1}(\alpha)S/\sqrt{n} = 8.426$. Hence the nonparametric decision limit is larger, and so less sensitive, than the normal distribution based decision limit; but the normal distribution based decision limit requires the assumption of normal distribution.

The purpose of this paper is to compare this nonparametric decision limit, denoted as NonpL hereafter, and the normal distribution based decision limit $\bar{X} - T_{n-1; -\sqrt{n}z_\gamma}^{-1}(\alpha) S/\sqrt{n}$, denoted as NormL henceforth. In Section 2, NonpL and NormL are compared for finite sample size behaviour. In particular, we demonstrate by simulation that the true confidence level of NormL can be much smaller than the nominal level $1 - \alpha$ (correspondingly, the probability of the FPR being larger than γ can be much larger than α) even if the true distribution $F(\cdot)$ departs only slightly from the normal distribution. We also assess by simulation the efficiency loss of NonpL relative to NormL when the true distribution $F(\cdot)$ is normal. To have a deep understanding of NormL and NonpL, their large sample asymptotic behaviours are studied in Section 3. Finally Section 4 contains a brief discussion.

The dataset containing 5053 GH-2000 scores and the R codes for reproducing the results in this paper are available on the journal's web page <http://onlinelibrary.wiley.com/doi/xxx/supinfo> as Supporting Information.

2 Finite sample comparison of the decision limits

In this section, NormL and NonpL are compared numerically. Note firstly that NormL can be applied only if the score X distribution $F(\cdot)$ is normal but NonpL is applicable as long as X is a continuous random variable. Secondly, NormL only needs the sample size n to be at least two (so that sample variance is available for estimating σ^2) but NonpL requires a much large sample size n , at least $\ln(\alpha)/\ln(1 - \gamma)$, as pointed out in Section 1.

Although it is difficult to affirm the exact distribution of $F(\cdot)$ based on a finite sample, NonpL always guarantees that the FPR is no more than γ with $1 - \alpha$ confidence about the randomness in the sample X_1, \dots, X_n , in contrast to NormL. In Subsection 2.1, the performance of NormL is assessed when $F(\cdot)$ departs from the normal distribution in several scenarios. On the other hand, when $F(\cdot)$ does have a normal distribution, NonpL is less efficient than NormL, which is tailor-made for the normal distribution. In Subsection 2.2, the efficiency loss of NonpL relative to NormL is assessed when the distribution $F(\cdot)$ is normal.

2.1 Distribution $F(\cdot)$ is not normal

In this subsection, we assess the probability

$$P \left\{ \bar{X} - T_{n-1; -\sqrt{n}z_\gamma}^{-1}(\alpha) S/\sqrt{n} > F^{-1}(1 - \gamma) \right\} \quad (6)$$

when $F(\cdot)$ is not a normal distribution. This probability is the confidence level that NormL is larger than $F^{-1}(1 - \gamma)$. We show that it can be substantially smaller than the nominal level $1 - \alpha$ when $F(\cdot)$ is not a normal distribution.

After the sample of observations being collected, NormL $\bar{X} - T_{n-1; -\sqrt{n}z_\gamma}^{-1}(\alpha) S/\sqrt{n}$ can be

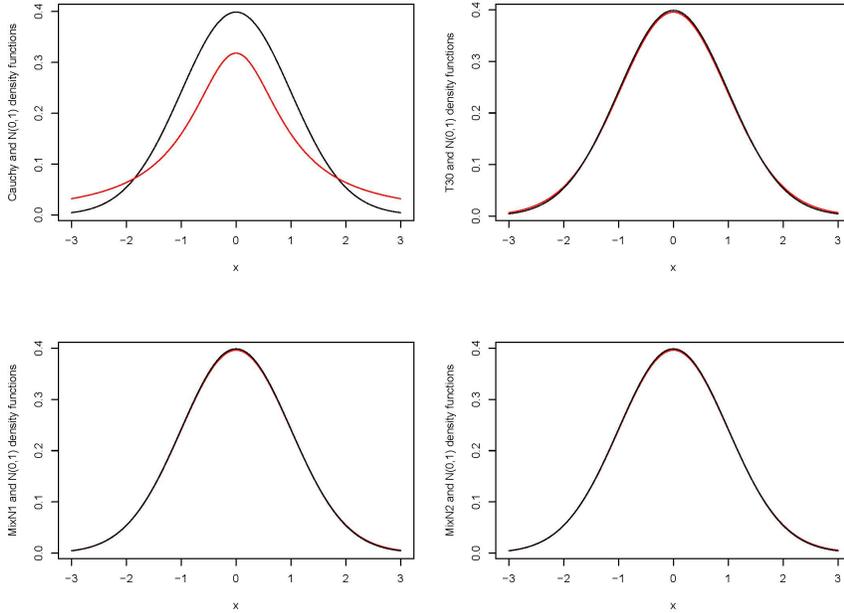


Figure 1: Plots of the four pdfs overlaying the $N(0, 1)$ pdf

calculated. When NormL is applied to future observations, the FPR is given by

$$P \left\{ X > \bar{X} - T_{n-1; -\sqrt{n}z_\gamma}^{-1}(\alpha) S/\sqrt{n} \right\} = 1 - F \left(\bar{X} - T_{n-1; -\sqrt{n}z_\gamma}^{-1}(\alpha) S/\sqrt{n} \right) \quad (7)$$

where X denotes a future observation from $F(\cdot)$ independent of the sample, and the probability is calculated with respect to X since the same NormL is used to all future observations. Clearly the FPR in (7) depends on \bar{X} and S , and so is random. Furthermore, the FPR is smaller than γ if and only if $\bar{X} - T_{n-1; -\sqrt{n}z_\gamma}^{-1}(\alpha) S/\sqrt{n} > F^{-1}(1 - \gamma)$. Hence the probability in (6) is also the probability that the FPR is smaller than γ , which can be substantially smaller than the nominal level $1 - \alpha$ when $F(\cdot)$ is not a normal distribution as shown in Table 1 below.

The probability in (6) can be computed easily by using statistical simulation. For a given distribution $F(\cdot)$ and γ , compute $F^{-1}(1 - \gamma)$ first. Then simulate a simple random sample of size n from $F(\cdot)$, compute the corresponding statistic $\bar{X} - T_{n-1; -\sqrt{n}z_\gamma}^{-1}(\alpha) S/\sqrt{n}$, and check whether the statistic is larger than $F^{-1}(1 - \gamma)$. Repeat this process R times, and use the proportion of times out of the R replications that the statistic is larger than $F^{-1}(1 - \gamma)$ as the probability in (6). The value $R = 1,000,000$ is used in all our simulations in this paper so that this proportion from simulation is within $2.576\sqrt{p(1-p)/R} \leq 0.0013$ of the true probability in (6) with 99% confidence level about the randomness in the simulation.

Four specific non-normal distributions of $F(\cdot)$ are investigated in our simulation study. Figure 1 plots each pdf of the four distributions, overlaying the pdf of $N(0, 1)$. The first is the standard Cauchy distribution (top left panel of Figure 1), whose pdf has a bell-shape too but much larger tails than $N(0, 1)$. The second is the central t distribution with thirty df (top right panel of Figure 1), T_{30} , whose pdf has only very slightly larger tails than $N(0, 1)$. Indeed, the results of Chantarangsi *et al.* (2015, Table 3) indicate that even the T_6 distribution is not

Table 1: The confidence level in expression (6) of NormL when $F(\cdot)$ is not a normal distribution and $\alpha = 0.05$

γ	n	k	NonpCL	Cauchy	T_{30}	$MixN_1$	$MixN_2$
0.01	900	897	0.979	0.933	0.813	0.749	0.983
	1000	996	0.971	0.946	0.802	0.739	0.985
	1100	1095	0.963	0.956	0.793	0.728	0.986
	1500	1492	0.963	0.981	0.752	0.688	0.990
0.001	3100	3100	0.955	0.341	0.010	0.000	0.905
	3500	3500	0.970	0.361	0.005	0.000	0.905
	5053	5052	0.961	0.425	0.000	0.000	0.901
	10000	9996	0.971	0.567	0.000	0.000	0.893
0.0001	30000	30000	0.950	0.129	0.000	0.000	0.000
	50000	49999	0.960	0.166	0.000	0.000	0.000

easily distinguishable from $N(0, 1)$ by the usual tests of normality when the sample size is not very large. The third distribution is the mixed normal distribution $0.99\Phi(x) + 0.01\Phi((x - 1.5)/2.5)$ (bottom left panel of Figure 1), denoted as $MixN_1$, which is only very slightly skewed to the right. Again, there are only very small differences between this distribution and $N(0, 1)$. The fourth distribution is the mixed normal distribution $0.99\Phi(x) + 0.01\Phi((x + 1.5)/2.5)$ (bottom right panel of Figure 1), denoted as $MixN_2$, whose pdf is the reflection about the vertical line $x = 0$ of the pdf of the $MixN_1$ above, and so only very slightly skewed to the left. Overall the Cauchy pdf has much larger tails than $N(0, 1)$ pdf, and there is little difference between the other three distributions from $N(0, 1)$.

Table 1 presents the simulation results on the confidence level in (6) for $\alpha = 0.05$ and various combinations of FPR γ and sample size n . The last four columns gives the true confidence levels of NormL, whilst the column NonpCL gives the true confidence levels of NonpL. The k value for given α, γ and n is for NonpL $X_{[k]}$. The true confidence level of $X_{[k]}$ is given by $1 - B_{k, n-k+1}(1 - \gamma)$ in (5) which is also computed and agrees to three decimal places with the corresponding simulated value given in the column NonpCL for all the cases in Table 1. This indicates that the simulation results on the confidence level of NormL can also be expected to be accurate to three decimal places.

It is clear from Table 1 that the true confidence level of NormL is drastically smaller than the nominal level $1 - \alpha = 0.95$ for most cases, and close to zero (to three decimal places) in several cases. This is surprising especially for the T_{30} distribution which seems very close to $N(0, 1)$ as shown in Figure 1. As expected, the true confidence level of NonpL is at least $1 - \alpha$ and has nothing to do with the specific distribution of $F(\cdot)$. The key message from Table 1 is that if the true distribution $F(\cdot)$ departs only very slightly from the normal distribution then the nominal confidence level of NormL can be seriously breached. Hence NormL should be used only if one is comfortable with the assumption that $F(\cdot)$ is a normal distribution.

If one is directly interested in the FPR in (7), which is a random variable taking values in the interval $(0, 1)$, one can simulate its value for a large number R times as above to study

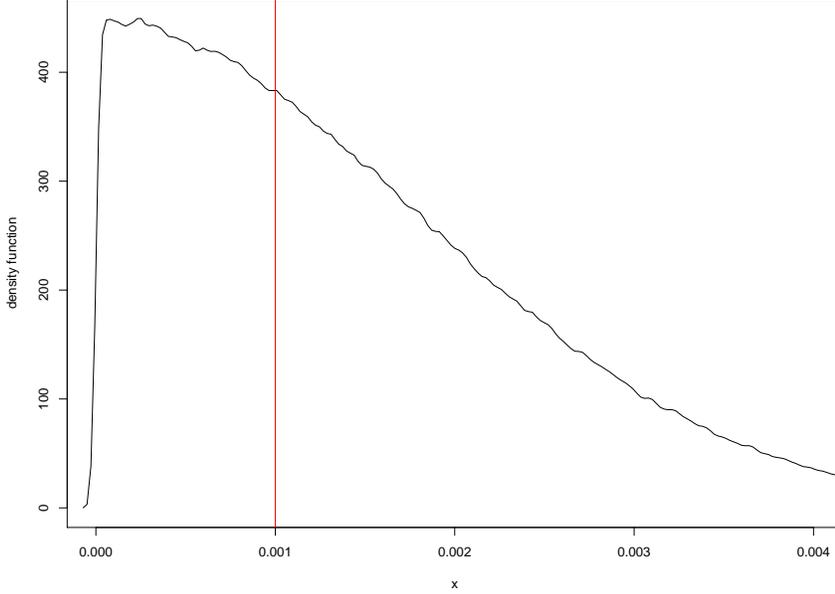


Figure 2: Estimate of the pdf of the FPR in (7)

its properties. When $n = 5053$, $\gamma = 0.001$, $\alpha = 0.05$ and $F(\cdot)$ is the standard Cauchy distribution, the kernel density estimate (cf. Wand and Jones, 1995, and the companion R package `KernSmooth`) of the pdf of the FPR, based on the $R = 1,000,000$, is given in Figure 2, where the vertical line indicates the nominal FPR $\gamma = 0.001$. In Figure 2, the probability that the FPR is smaller than γ is given by the area under the curve to the left of the vertical line, which is given by the corresponding entry 0.425 in Table 1 due to relationship pointed out under expression (7). From Figure 2, it is clear that the FPR in (7) can be substantially larger than $\gamma = 0.001$. Of course, a decision limit that has inflated FPR will also increase its chance of detecting true positives.

Subsections 3.1 and 3.2 below investigate the asymptotic distributions of $\bar{X} - T_{n-1; -\sqrt{n}z_\gamma}^{-1}(\alpha) S/\sqrt{n}$ and $X_{[k]}$, which provide insights on NormL and NonpL from a different perspective.

2.2 Distribution $F(\cdot)$ is normal

In this subsection, the efficiency loss of NonpL relative to NormL is assessed when $F(\cdot)$ is a normal distribution $N(\mu, \sigma^2)$. Specifically, the difference between the two decision limits $X_{[k]} - \left[\bar{X} - T_{n-1; -\sqrt{n}z_\gamma}^{-1}(\alpha) S/\sqrt{n} \right] = \sigma D_0$ is of interest, where

$$D_0 = \frac{X_{[k]} - \mu}{\sigma} - \left[\frac{\bar{X} - \mu}{\sigma} - T_{n-1; -\sqrt{n}z_\gamma}^{-1}(\alpha) \frac{S/\sigma}{\sqrt{n}} \right]. \quad (8)$$

Note that D_0 does not depend on μ or σ and hence, without loss of generality, we assume $\mu = 0$ and $\sigma^2 = 1$ when studying D_0 . We will focus on the distribution of D_0 and, in particular, $P\{D_0 > 0\}$ which gives the probability that NonpL is larger than NormL. These can be assessed again by simulation. To simulate one D_0 , one generates a simple random

Table 2: The values of $P\{D_0 > 0\}$ (top), $E(D_0)$ (middle) and $sd(D_0)$ (bottom) when $F(\cdot)$ is a normal distribution and $\alpha = 0.05$

γ	$n = 900$	1,100	1,500	3,000	5053	10,000	30,000	50,000	∞
0.01	0.924	0.873	0.872	0.877	0.859	0.841	0.834	0.828	0.824
	0.220	0.147	0.120	0.080	0.055	0.035	0.019	0.014	0
	0.163	0.132	0.107	0.070	0.051	0.035	0.019	0.015	0
0.001	NA	NA	NA	0.908	0.925	0.940	0.927	0.905	0.897
				0.381	0.276	0.187	0.088	0.059	0
				0.323	0.211	0.128	0.062	0.045	0
0.0001	NA	NA	NA	NA	NA	NA	0.933	0.944	0.929
							0.368	0.265	0
							0.287	0.187	0

sample of size n from $N(0, 1)$ and compute D_0 using (8) with $\mu = 0$ and $\sigma = 1$. Repeat this $R = 1, 000, 000$ times to generate a simple random sample of R observations of D_0 , based on which one can approximate accurately the distribution of D_0 and $P\{D_0 > 0\}$. Based on the R simulated observations of D_0 , we use the kernel density estimate as before to assess the pdf of D_0 . We use the proportion of times $D_0 > 0$ in the R replications as $P\{D_0 > 0\}$, and use the sample mean and standard deviation as $E(D_0)$ and $sd(D_0)$ respectively.

Table 2 presents some results on $P\{D_0 > 0\}$, $E(D_0)$ and $sd(D_0)$, given by the top, middle and bottom entries respectively. The entry ‘NA’ represents the case that, for the given values of α, γ and n , there is no k that satisfies the inequality in (5) and so NonpL cannot be constructed. As expected, $P\{D_0 > 0\}$ is larger than 0.5 for each case and so NonpL is larger than normL most times. However, $P\{D_0 > 0\} < 1$ and so NonpL is smaller than NormL with a positive probability, that is, for some samples of observations X_1, \dots, X_n , NonpL is smaller than NormL. For example, for the case of $\gamma = 0.01$ and $n = 10, 000$ in Table 2, NonpL is actually smaller than NormL for $100 \times (1 - 0.84)\% = 16\%$ of the samples.

More detailed information about the magnitude of D_0 is provided by the kernel density estimate. Figure 3 provides the kernel density estimate for the case of $\alpha = 0.05, \gamma = 0.01$ and $n = 1500$. Based on the R simulated observations, one can easily compute $E(D_0) = 0.120$ and $sd(D_0) = 0.107$ using the sample mean and standard deviation, for example, which indicates that the average amount (scaled by σ) by which NonpL is larger than NormL is $E(D_0) = 0.120$. The density estimate in Figure 3 also indicates that the likely values of D_0 are in the interval $[-0.2, 0.5]$.

Subsection 3.3 provides the asymptotic distribution of D_0 to provide insight on D_0 from a different viewpoint. The entries in the last column corresponding to $n = \infty$ of Table 2 are the asymptotic values of $P\{D_0 > 0\}$, $E(D_0)$ and $sd(D_0)$ as $n \rightarrow \infty$, using the results in Subsection 3.3.

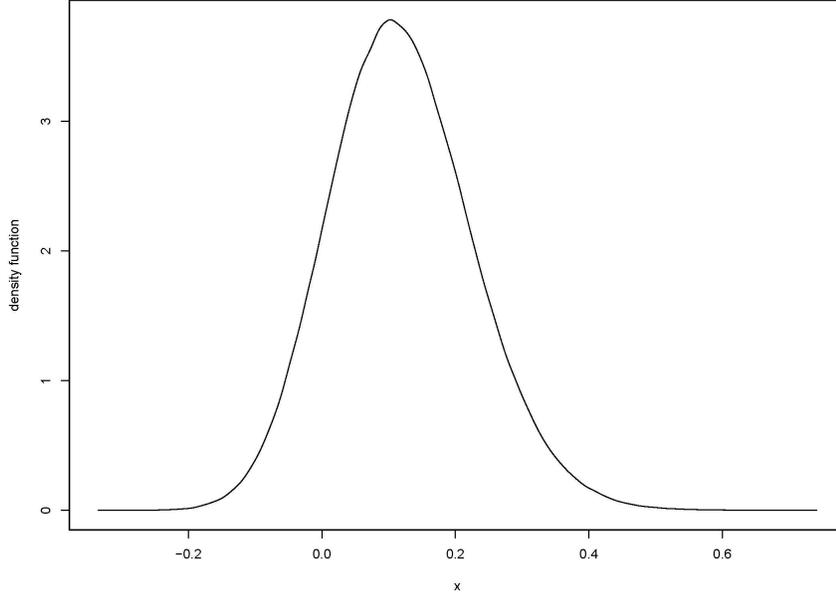


Figure 3: Estimate of the pdf of D_0

3 Large sample theory of the decision limits

3.1 NormL

To understand better why the probability in (6) may be different from the nominal confidence level $1 - \alpha$ when $F(\cdot)$ is not normal, we study the large sample (as $n \rightarrow \infty$ throughout this section) behaviour of the probability in (6). Let μ and σ^2 denote the mean and variance, respectively, of the distribution $F(\cdot)$. The readers are referred to van der Vaart (1998) for the meanings of the standard notations such as $o(a_n)$ and $o_p(a_n)$, and for basic asymptotic results about the sample mean and variance.

First we show that the constant $\lambda = T_{n-1; -\sqrt{n}z_\gamma}^{-1}(\alpha)$ used in NormL has the asymptotic expansion

$$\lambda = -z_\gamma \sqrt{n} - z_\alpha \sqrt{1 + z_\gamma^2/2} + o(1). \quad (9)$$

Note that, from its definition (via a non-central t distribution), λ satisfies

$$P \left\{ Z - \sqrt{n}z_\gamma \leq \lambda \sqrt{\chi_{n-1}^2/(n-1)} \right\} = \alpha \quad (10)$$

where Z denotes a standard normal $N(0, 1)$ random variable, χ_{n-1}^2 denotes a chi-square random variable with $n-1$ df, and Z and χ_{n-1}^2 are independent. By writing $\sqrt{\chi_{n-1}^2/(n-1)} = \sqrt{1 + (\chi_{n-1}^2 - (n-1))/(n-1)}$ and using Taylor expansion $\sqrt{1+x} = 1+x/2+o(x)$ as $x \rightarrow 0$, we have $\sqrt{\chi_{n-1}^2/(n-1)} = 1 + \frac{1}{2}(\chi_{n-1}^2 - (n-1))/(n-1) + o_p(1/\sqrt{n})$ as $n \rightarrow \infty$, where $o_p(1/\sqrt{n})$ denotes a term that satisfies $o_p(1/\sqrt{n})/(1/\sqrt{n}) \rightarrow 0$ in probability as $n \rightarrow \infty$.

Hence the probability in (10) becomes

$$P \left\{ Z - \frac{\lambda}{\sqrt{2(n-1)}} \frac{\chi_{n-1}^2 - (n-1)}{\sqrt{2(n-1)}} - \lambda \cdot o_p(1/\sqrt{n}) \leq \lambda + \sqrt{n}z_\gamma \right\}. \quad (11)$$

It is therefore clear that, when λ is replaced with the expansion in (9), we have

$$\begin{aligned} \lambda + \sqrt{n}z_\gamma &\rightarrow -z_\alpha \sqrt{1 + z_\gamma^2/2}, \\ \frac{\lambda}{\sqrt{2(n-1)}} \frac{\chi_{n-1}^2 - (n-1)}{\sqrt{2(n-1)}} &\xrightarrow{\mathcal{L}} N(0, z_\gamma^2/2), \\ Z - \frac{\lambda}{\sqrt{2(n-1)}} \frac{\chi_{n-1}^2 - (n-1)}{\sqrt{2(n-1)}} - \lambda \cdot o_p(1/\sqrt{n}) &\xrightarrow{\mathcal{L}} N(0, 1 + z_\gamma^2/2) \end{aligned}$$

due to the independence of Z and χ_{n-1}^2 ; here ‘ $\xrightarrow{\mathcal{L}}$ ’ means convergence in distribution. Hence the limit of the probability in (11) is α as required by the expression in (10).

From the expression in (9) it is clear that NormL satisfies: $\bar{X} - \lambda S/\sqrt{n} \xrightarrow{a.s.} \mu + z_\gamma \sigma$. Next we establish the asymptotic distribution of

$$\begin{aligned} &(\sqrt{n}/\sigma) \left(\bar{X} - \lambda S/\sqrt{n} - (\mu + z_\gamma \sigma) \right) \\ &= \sqrt{n}(\bar{X} - \mu)/\sigma + z_\gamma \sqrt{n}(S/\sigma - 1) + \left(z_\alpha \sqrt{1 + z_\gamma^2/2} + o(1) \right) S/\sigma \\ &= \sqrt{n}\bar{Y} + z_\gamma \sqrt{n}(S_Y - 1) + \left(z_\alpha \sqrt{1 + z_\gamma^2/2} + o(1) \right) S_Y \end{aligned} \quad (12)$$

where \bar{Y} and S_Y^2 denote, respectively, the sample mean and variance of $Y_1 = (X_1 - \mu)/\sigma, \dots, Y_n = (X_n - \mu)/\sigma$. Applying the same Taylor expansion for $\sqrt{1+x}$ as in the last paragraph to $S_Y = \sqrt{1 + (S_Y^2 - 1)}$ gives

$$\begin{aligned} \sqrt{n}(S_Y - 1) &= \frac{1}{2} \sqrt{n}(S_Y^2 - 1) + o_p(1) \\ &= \frac{1}{2} \sqrt{n} \left(\frac{1}{n} \left(\sum_{i=1}^n Y_i^2 - n(\bar{Y})^2 \right) - 1 \right) + o_p(1) \\ &= \frac{1}{2} \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n Y_i^2 - 1 + o_p(1/\sqrt{n}) \right) + o_p(1) \\ &= \frac{1}{2} \sum_{i=1}^n (Y_i^2 - 1)/\sqrt{n} + o_p(1). \end{aligned}$$

Substituting this into the expression in (12) gives

$$\begin{aligned} &(\sqrt{n}/\sigma) \left(\bar{X} - \lambda S/\sqrt{n} - (\mu + z_\gamma \sigma) \right) \\ &= \sum_{i=1}^n \left(Y_i + z_\gamma (Y_i^2 - 1)/2 \right) / \sqrt{n} + z_\alpha \sqrt{1 + z_\gamma^2/2} + o_p(1) \end{aligned} \quad (13)$$

$$\xrightarrow{\mathcal{L}} N(0, v^2) + z_\alpha \sqrt{1 + z_\gamma^2/2} \quad (14)$$

where $v^2 = \text{Var}(Y_i + z_\gamma (Y_i^2 - 1)/2) = 1 + z_\gamma E(Y_1^3) + z_\gamma^2 (E(Y_1^4) - 1)/4$.

The asymptotic distribution in (14) describes how $\bar{X} - \lambda S/\sqrt{n}$ behaves around $\mu + z_\gamma\sigma$ when the sample size n is large. From (14) we have for a large n

$$\begin{aligned}
& P \left\{ \bar{X} - \lambda S/\sqrt{n} > F^{-1}(1 - \gamma) \right\} \\
&= P \left\{ (\sqrt{n}/\sigma) \left(\bar{X} - \lambda S/\sqrt{n} - (\mu + z_\gamma\sigma) \right) > (\sqrt{n}/\sigma) \left(F^{-1}(1 - \gamma) - (\mu + z_\gamma\sigma) \right) \right\} \\
&\rightarrow P \left\{ N(0, v^2) > (\sqrt{n}/\sigma) \left(F^{-1}(1 - \gamma) - (\mu + z_\gamma\sigma) \right) - z_\alpha \sqrt{1 + z_\gamma^2/2} \right\} \tag{15}
\end{aligned}$$

It is clear that, when $F(\cdot)$ is normal, $v^2 = 1 + z_\gamma^2/2$, $F^{-1}(1 - \gamma) = \mu + z_\gamma\sigma$ and so the probability in (15) is equal to $1 - \alpha$ as expected. On the other hand, when $F(\cdot)$ is not normal, $F^{-1}(1 - \gamma)$ is unlikely equal to $\mu + z_\gamma\sigma$ and hence the probability in (15) could be very different from the nominal level $1 - \alpha$. Especially when $F(\cdot)$ has a larger right tail such that $F^{-1}(1 - \gamma) - (\mu + z_\gamma\sigma) > 0$ the probability in (15) can be really small when the sample size n is large, which is also observed from the simulation results in Table 1 of Section 2.

3.2 NonpL

In this section we study the large sample behaviour of NonpL $X_{[k]}$. For this, we first derive a large sample expansion for k which satisfies

$$P \left\{ 1 - \gamma \leq F(X_{[k]}) \right\} \rightarrow 1 - \alpha \text{ as } n \rightarrow \infty. \tag{16}$$

Recall that $F(X_1), \dots, F(X_n)$ are independent random variables each having a uniform distribution on the interval $(0, 1)$. Let M denote the number of random variables among $F(X_1), \dots, F(X_n)$ that fall in the interval $(0, 1 - \gamma)$ and so $M \sim \text{Binomial}(n, 1 - \gamma)$. It is clear that

$$\begin{aligned}
& P \left\{ 1 - \gamma \leq F(X_{[k]}) \right\} = P \{ M < k \} \\
&= P \left\{ \frac{M - n(1 - \gamma)}{\sqrt{n(1 - \gamma)\gamma}} < \frac{k - n(1 - \gamma)}{\sqrt{n(1 - \gamma)\gamma}} \right\} \\
&\rightarrow P \left\{ N(0, 1) < \frac{k - n(1 - \gamma)}{\sqrt{n(1 - \gamma)\gamma}} \right\}.
\end{aligned}$$

In order that the last probability has the limit $1 - \alpha$ as required in (16), it is clear that k must be of the form

$$k = n(1 - \gamma) + z_\alpha \sqrt{n(1 - \gamma)\gamma} + o(\sqrt{n}).$$

From this expansion of k we have

$$\begin{aligned}
& \sqrt{n} \left(X_{[k]} - F^{-1}(1 - \gamma) \right) \\
&= \sqrt{n} \left(X_{[k]} - F_n^{-1}(1 - \gamma) \right) + \sqrt{n} \left(F_n^{-1}(1 - \gamma) - F^{-1}(1 - \gamma) \right) \\
&= \frac{z_\alpha \sqrt{(1 - \gamma)\gamma}}{f(F^{-1}(1 - \gamma))} - \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{I_{\{X_i \leq F^{-1}(1 - \gamma)\}} - (1 - \gamma)}{f(F^{-1}(1 - \gamma))} + o_p(1) \tag{17}
\end{aligned}$$

$$\xrightarrow{\mathcal{L}} \frac{\sqrt{\gamma(1-\gamma)}}{f(F^{-1}(1-\gamma))} (z_\alpha + N(0,1)), \quad (18)$$

where $F_n^{-1}(\cdot)$ denotes the empirical quantile function based the sample X_1, \dots, X_n (van der Vaart, 1998, p.305), I_A denotes the index function of set A , and the equality in (17) follows directly from the lemma of van der Vaart (1998, p.308) and the corollary of van der Vaart (1998, p.307).

The asymptotic distribution in (18) describes how NonpL $X_{[k]}$ behaves around $F^{-1}(1-\gamma)$ when n is large. In particular, from (18) it is clear that

$$\begin{aligned} & P \left\{ F^{-1}(1-\gamma) \leq X_{[k]} \right\} \\ &= P \left\{ \sqrt{n} \left(X_{[k]} - F^{-1}(1-\gamma) \right) \geq 0 \right\} \\ &\rightarrow P \left\{ \frac{\sqrt{\gamma(1-\gamma)}}{f(F^{-1}(1-\gamma))} (z_\alpha + N(0,1)) \geq 0 \right\} = 1 - \alpha \end{aligned}$$

irrespective of the form of the distribution $F(\cdot)$ as expected.

3.3 Comparison of NormL and NonpL

Combine the expressions in (13) and (17) to give

$$\begin{aligned} \tilde{D} &:= -\sqrt{n} \left(\bar{X} - \lambda S / \sqrt{n} - X_{[k]} \right) \\ &= -\sqrt{n} \left(\bar{X} - \lambda S / \sqrt{n} - (\mu + z_\gamma \sigma) \right) + \sqrt{n} \left(X_{[k]} - F^{-1}(1-\gamma) \right) \\ &\quad - \sqrt{n} \left((\mu + z_\gamma \sigma) - F^{-1}(1-\gamma) \right) \\ &= -\frac{1}{\sqrt{n}} \sum_{i=1}^n \left((X_i - \mu) + z_\gamma \left((X_i - \mu)^2 - \sigma^2 \right) / 2 \right) - \sigma z_\alpha \sqrt{1 + z_\gamma^2 / 2} \\ &\quad - \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{I_{\{X_i \leq F^{-1}(1-\gamma)\}} - (1-\gamma)}{f(F^{-1}(1-\gamma))} + \frac{z_\alpha \sqrt{(1-\gamma)\gamma}}{f(F^{-1}(1-\gamma))} + o_p(1) \\ &\quad - \sqrt{n} \left((\mu + z_\gamma \sigma) - F^{-1}(1-\gamma) \right) \end{aligned}$$

for any distribution function $F(\cdot)$. We are particularly interested in the situation that $F(\cdot)$ has a normal distribution $N(\mu, \sigma^2)$ in order to understand how NonpL $X_{[k]}$ can be less efficient than NormL $\bar{X} - \lambda S / \sqrt{n}$.

When $F(\cdot)$ has a normal distribution $N(\mu, \sigma^2)$, we have $\tilde{D} = \sigma \tilde{D}_0$ where

$$\begin{aligned} \tilde{D}_0 &= -\frac{1}{\sqrt{n}} \sum_{i=1}^n \left(Y_i + z_\gamma \left(Y_i^2 - 1 \right) / 2 \right) - z_\alpha \sqrt{1 + z_\gamma^2 / 2} \\ &\quad - \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{I_{\{Y_i \leq z_\gamma\}} - (1-\gamma)}{\phi(z_\gamma)} + \frac{z_\alpha \sqrt{(1-\gamma)\gamma}}{\phi(z_\gamma)} + o_p(1) \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n W_i - z_\alpha \sqrt{1 + z_\gamma^2/2} + z_\alpha \sqrt{(1 - \gamma)\gamma}/\phi(z_\gamma) + o_p(1) \\
&\xrightarrow{L} \sqrt{\text{Var}(W_1)} N(0, 1) - z_\alpha \sqrt{1 + z_\gamma^2/2} + z_\alpha \sqrt{(1 - \gamma)\gamma}/\phi(z_\gamma)
\end{aligned} \tag{19}$$

where $Y_i = (X_i - \mu)/\sigma \sim N(0, 1)$, $W_i = -Y_i - z_\gamma (Y_i^2 - 1)/2 - (I_{\{Y_i \leq z_\gamma\}} - (1 - \gamma))/\phi(z_\gamma)$, and straightforward calculation gives

$$\text{Var}(W_1) = 1 + z_\gamma^2/2 + \gamma(1 - \gamma)/\phi^2(z_\gamma) + \left(2 \int_{-\infty}^{z_\gamma} z\phi(z)dz + z_\gamma \left(\int_{-\infty}^{z_\gamma} z^2\phi(z)dz - (1 - \gamma) \right) \right) / \phi(z_\gamma). \tag{20}$$

It is interesting to observe that the term

$$-\sqrt{1 + z_\gamma^2/2} + \sqrt{(1 - \gamma)\gamma}/\phi(z_\gamma)$$

in (19) is always positive for $\gamma \in (0, 1/2)$, which is easy to check by plotting it against $\gamma \in (0, 1/2)$. Hence $P\{\tilde{D}_0 > 0\} > 1/2$ asymptotically at least, that is, NonpL $\bar{X}_{[k]}$ is larger than NormL $\bar{X} - \lambda S/\sqrt{n}$ with a greater than half chance for at least a large n . This agrees with the simulation results in Table 2 by noting that $D_0 = \tilde{D}_0/\sqrt{n}$ and so $P\{D_0 > 0\} = P\{\tilde{D}_0 > 0\}$.

From $D_0 = \tilde{D}_0/\sqrt{n}$ and the asymptotic distribution of \tilde{D}_0 in (19), we have

$$\begin{aligned}
P\{D_0 > 0\} &= P\{\tilde{D}_0 > 0\} \\
&\approx P\left\{ \sqrt{\text{Var}(W_1)} N(0, 1) - z_\alpha \sqrt{1 + z_\gamma^2/2} + z_\alpha \sqrt{(1 - \gamma)\gamma}/\phi(z_\gamma) > 0 \right\} \\
&= \Phi\left\{ z_\alpha \left(-\sqrt{1 + z_\gamma^2/2} + \sqrt{(1 - \gamma)\gamma}/\phi(z_\gamma) \right) / \sqrt{\text{Var}(W_1)} \right\}.
\end{aligned} \tag{21}$$

The expression in (21) can easily be computed numerically by using numerical quadrature to compute $\text{Var}(W_1)$ from the expression in (20); its values for some cases are given in the last column of Table 2 corresponding to $n = \infty$. For example, when $\alpha = 0.05$ and $\gamma = 0.01$, the expression in (21) is computed to be 0.824. Hence the entries (based on simulation) in the first row of Table 2 are expected to converge to 0.824 as n becomes large. The entries in the first row of Table 2 indicate that the asymptotic approximation in (21) becomes reasonably accurate when $n \geq 10,000$. For $\alpha = 0.05$ and $\gamma = 0.01$, we have compared the kernel density estimate of the pdf of $\sqrt{n}(X_{[k]} - F^{-1}(1 - \gamma))$ based on $R = 1,000,000$ simulations with the pdf of the asymptotic normal distribution in (18) for various values of n . Our investigation shows that the kernel density estimate is not close to the asymptotic normal distribution unless $n \geq 10,000$. Similar investigation shows that the kernel density estimate of the pdf of $(\sqrt{n}/\sigma)(\bar{X} - \lambda S/\sqrt{n} - (\mu + z_\gamma\sigma))$ is quite close to the asymptotic normal distribution in (14) when $n \geq 300$. Hence the asymptotic normal distribution in (14) requires a much smaller sample size n than the asymptotic normal distribution in (18) to be a reasonably accurate approximation. It is clear from the kernel density estimate in Figure 3 that, for $n = 1,500$, the distribution of D_0 (or \tilde{D}_0) is skewed to the right and so not close to normal yet.

It follows directly from the relationship $D_0 = \tilde{D}_0/\sqrt{n}$ and the asymptotic distribution of \tilde{D}_0 in (19) that both $E(D_0)$ and $sd(D_0)$ approach 0 as $n \rightarrow \infty$, which are indicated by the corresponding entries in the last column of Table 2.

4 Discussion

We have shown by both simulation and large sample theory that the exact decision limit based on the normal distribution of the GH-2000 score for growth hormone misuse can seriously violate the nominal confidence level $1 - \alpha$ when the distribution $F(\cdot)$ of the GH-2000 score departs only slightly from the normal distribution. Note also that the non-rejection of normality by a goodness-of-fit test does not mean that the distribution $F(\cdot)$ is exactly normal. Hence one should be aware of the potential risk when using NormL. The nonparametric decision limit NonpL is an attractive alternative since it always guarantees the nominal confidence level $1 - \alpha$ irrespective of the specific form of $F(\cdot)$. One price to pay for using NonpL is that a much larger sample size is required than NormL in order to be able to construct NonpL, especially when γ or α are very small. The other price for using NonpL is that, when $F(\cdot)$ is a normal distribution, NonpL tends to be larger, and so has a smaller chance of detecting doping, than NormL. The large sample theory developed in this paper helps us to understand the behaviours of NonpL and NormL when the sample size is large. The R codes are provided for computing all the results given in this paper.

When the score X is a discrete random variable, $F(X)$ no longer has a uniform distribution on the interval $(0, 1)$ and so the construction of a nonparametric decision limit is expected to be very different from the NonpL considered in this paper. This warrants further research nevertheless.

The direct implication of the results in this paper on the GH-2000 score for growth hormone misuse detection is that the current use of NormL is clearly based on the assumption of normality. NonpL is a simple, easy-to-calculate and distribution-free alternative. However, the sample size currently available is too small to use NonpL for $\gamma = 0.0001$ and $\alpha = 0.05$ required by WADA. On the other hand, using $\gamma = 0.001$ and $\alpha = 0.05$, for example, would lead to a sample size requirement of at least $n = 2995$ which is realistic, given the currently available sample size is $n = 5053$. However, false positive rate $\gamma = 0.001$ does not meet the current WADA legal standard of proof. Since it is likely to take many years to accrue $n = 29956$ observations, the onus is on WADA to decide whether NonpL with a less stringent γ or the distribution sensitive NormL should be used, while this paper shows the advantages and disadvantages of both NormL and NonpL.

The results of this paper have a wider implication to the many distribution-specific tolerance intervals/regions published in the statistical literature (cf. Krishnamoorthy and Mathew, 2009, and Meeker *et al.*, 2017). These distribution-specific procedures are likely to have similar problems as NormL identified in this paper and so should be used with a clear understanding of potential pitfalls. It is our view that one should use the nonparametric procedures if available.

Acknowledgments: We thank the referees and editors for constructive comments.

References

- [1] Bidlingmaier, M., Wu, Z. and Strasburger, C.J. (2000). Test method: GH. *Baillieres Best Pract Res Clin Endocrinol Metab.*, 14(1), 99-109.
- [2] Böhning, D., Liu, W., Holt, R., Böhning, W., Guha, N., Sönksen, P.H., Cowan, D.A. and Liang, T. (2019). Exact statistical calculation of the uncertainty term in the decision limits based on the GH-2000 score for growth hormone misuse detection (doping). *Statistical Methods in Medical Research*, 28(3), 928-936.
- [3] Chantarangsi, W., Liu, W., Bretz, F., Kiatsupaibul, S., Hayter, A.J. and Wan, F. (2015). Normal probability plots with confidence. *Biometrical Journal*, 57(1), 52-63.
- [4] Dall, R., Longobardi, S., Ehrnborg, C., Keay, N., Rosen, T., Jorgensen, J.O., *et al.* (2000). The effect of four weeks of supraphysiological growth hormone administration on the insulin-like growth factor axis in women and men. GH-2000 Study Group. *J. Clin. Endocrinol. Metab.*, 85(11), 4193-200.
- [5] Guttman, I. (1970). *Statistical Tolerance Regions: Classical and Bayesian*. Griffin: London.
- [6] Guttman, I. (2006). Tolerance Regions, in *Encyclopedia of Statistical Sciences, 2nd edition*, edited by Kotz S *et al.*, 8644-8659, Wiley: New York.
- [7] Holt, R.I. (2009). Is human growth hormone an ergogenic aid? *Drug Test Anal.*, 1, 412-418. doi: 10.1002/dta.58.
- [8] Holt, R.I., Böhning, W., Guha, N., Bartlett, C., Cowan, D.A., Giraud, S., Bassett, E.E., Sönksen, P.H. and Böhning, D. (2015). The development of decision limits for the GH-2000 detection methodology using additional insulin-like growth factor-I and amino-terminal pro-peptide of type III collagen assays. *Drug Test Anal.*, 7, 745-755. doi: 10.1002/dta.1772.
- [9] Krishnamoorthy, K. and Mathew, T. (2009). *Statistical Tolerance Regions – Theory, Applications, and Computation*. Wiley: New York.
- [10] Liu, W., Bretz, F., Hayter, A. J. and Glimm, E. (2013). Simultaneous inference for several quantiles of a normal population with applications. *Biometrical Journal*, 55(3), 360-369.
- [11] Longobardi, S., Keay, N., Ehrnborg, C., Cittadini, A., Rosen, T., Dall, R., *et al.* (2000). Growth hormone (GH) effects on bone and collagen turnover in healthy adults and its potential as a marker of GH abuse in sports: a double blind, placebo-controlled study. The GH-2000 Study Group. *J. Clin. Endocrinol. Metab.*, 85(4), 1505-12.
- [12] Meeker, W.Q., Hahn, G.J. and Escobar, L.A. (2017). *Statistical Intervals: A Guide For Practitioners And Researchers, 2nd ed.*. Wiley: New York.

- [13] Tukey, J. (1947). Nonparametric estimation II: statistical equivalence blocks and tolerance regions – the continuous case. *Annals of Mathematical Statistics*, 18, 529-539.
- [14] van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press: Cambridge.
- [15] WADA World Antidoping Agency (2014). World Anti-Doping Program Guidelines for hGH Isoform Differential Immunoassays for anti-doping analyses, <https://wada-main-prod.s3.amazonaws.com/resources/files/WADA-Guidelines-for-hGH-Differential-Immunoassays-v2.1-2014-EN.pdf> .
- [16] WADA (2016). The World Anti-Doping Code International Standard: Prohibited List 2016, <https://wada-main-prod.s3.amazonaws.com/resources/files/wada-2016-prohibited-list-en.pdf>
- [17] Wald, A. (1943). An extension of Wilks' method for setting tolerance limits. *Annals of Mathematical Statistics*, 14, 45-55.
- [18] Wand, M. and Jones, M.C. (1995). *Kernel Smoothing*. Springer: New York.
- [19] Wilks, S. S. (1941). Determination of sample sizes for setting tolerance limits. *Annals of Mathematical Statistics*, 12, 91-96.
- [20] Young, D.S. (2010). *tolerance*: An R Package for Estimating Tolerance Intervals. *Journal of Statistical Software*, 36, 1-39.