Metrika (2002) 55: 177–182

Metrika

A note on the maximum deviation of the scale-contaminated normal to the best normal distribution

Dankmar Böhning, Roumporn Ruangroj

Department of Epidemiology, Joint Center for Humanities and Health Sciences, Free University Berlin, Humboldt University at Berlin, Haus 562, Fabeckstr. 60-62, 14195 Berlin, Germany (E-Mail: boehning@zedat.fu-berlin.de)
Department of Mathematics and Statistics, Faculty of Science and Technology, Thammassast University, Pathumthani, Thailand (E-Mail: roumporn@alpha.tu.ac.th)

Received May 2001

Abstracts. In this paper we consider the case of the scale-contaminated normal (mixture of two normals with equal mean components but different component variances: $(1-p)N(\mu, \sigma^2) + pN(\mu, \tau^2)$ with σ and τ being non-negative and $0 \le p \le 1$). Here $c = \frac{\tau}{\sigma}$ is the scale error and p denotes the amount with which this error occurs. It's maximum deviation to the best normal distribution is studied and shown to be montone increasing with increasing scale error. A closed-form expression is derived for the proportion which maximizes the maximum deviation of the mixture of normals to the best normal distribution. Implications to power studies of tests for normality are pointed out.

Key words: mixture of normals with different component variances, scale-contaminated normal, maximum deviation, suprenum's norm

1 Introduction

We are interested in the scale-contaminated normal $(1-p)N(\mu, \sigma^2)+pN(\mu, \tau^2)$ with σ and τ being non-negative and $0 \le p \le 1$. Here $c = \frac{\tau}{\sigma}$ is the scale error and p denotes the amount with which this error occurs. The scale-contaminated normal is a special mixture distribution with mean μ and variance $(1-p)\sigma^2 + p\tau^2$. It is symmetric and bell-shaped, but it is *not* a normal distribution itself. Therefore, it is a suitable candidate for robustness studies as suggested in Barnett and Lewis (1983) or for power studies in tests for departures from normality as has been done by Chen (1971) and some follow-up studies by Chotiwittayatarakorn (1987), Chuchao (1992), and Jantako (1993). In the latter case, the distribution under the null hypothesis is assumed to be normal and, when sampling is done from the specific mixture described above, it's

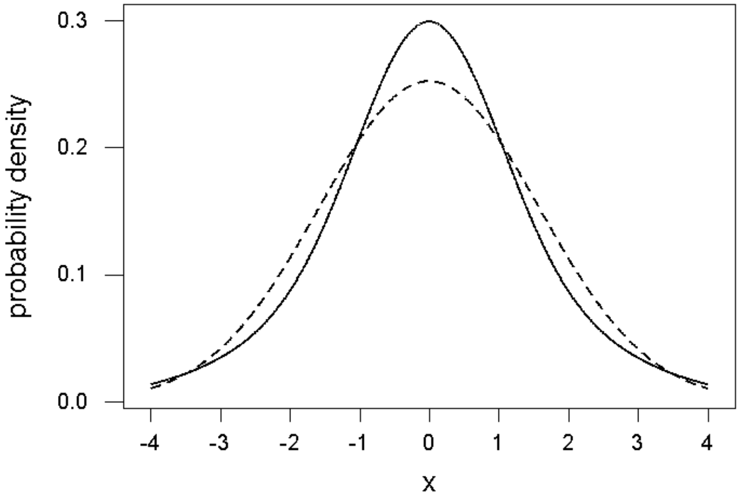

Fig. 1. Scale-contaminated normal $\frac{1}{2}N(0,1)+\frac{1}{2}N(0,2^2)$ (solid line) and best normal approximation $N\left(0,\frac{1}{2}+\frac{1}{2}2^2\right))$ (dashed line)

parameters are already determined as μ for the mean and $(1-p)\sigma^2+p\tau^2$ for it's variance. The situation is illustrated in Figure 1 with $\sigma=1$, $\tau=2$ and $p=\frac{1}{2}$.

Given this background, it might be of some considerable interest to understand for which values of $c=\dfrac{\tau}{\sigma}$ and p the departure of $(1-p)N(\mu,\sigma^2)+pN(\mu,\tau^2)$ from $N(\mu,(1-p)\sigma^2+p\tau^2)$ is the largest. These values might serve as standard of comparison for power studies of tests for normality. Any test for departure from nomality should have its maximum power for this value of p. In fact, the issue of this contribution did arise while the second author was undertaking a comparative power study of various tests for normality (Ruangroj 2000) and maximizing powers occurred at values of p quite different from 0.5. Therefore, the departure of the scale contaminated normal $(1-p)N(\mu,\sigma^2)$ $+pN(\mu,\tau^2)$ from the best normal distribution is of some interest. Here, "best" means that normal distribution having first two moments identical to the those of the contaminated normal. This topic will be addressed in the next section. Without limitation of generality we let $\mu=0$, $\sigma=1$, and $c=\tau>1$.

2 Results

Let us denote the absolute deviation of the mixture of normals $\phi_{\text{mix}}(x)=(1-p)\phi(x)+p\dfrac{1}{c}\phi\left(\dfrac{x}{c}\right)$ to the best normal $\phi_{\text{best}}(x)=\dfrac{1}{\sqrt{1-p+pc^2}}\cdot\phi(x/\sqrt{1-p+pc^2})$

$$f(x,p,c)=\sqrt{2\pi}|\phi_{\text{mix}}(x)-\phi_{\text{best}}(x)| \tag{1}$$

where $\phi(x)=\dfrac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$.

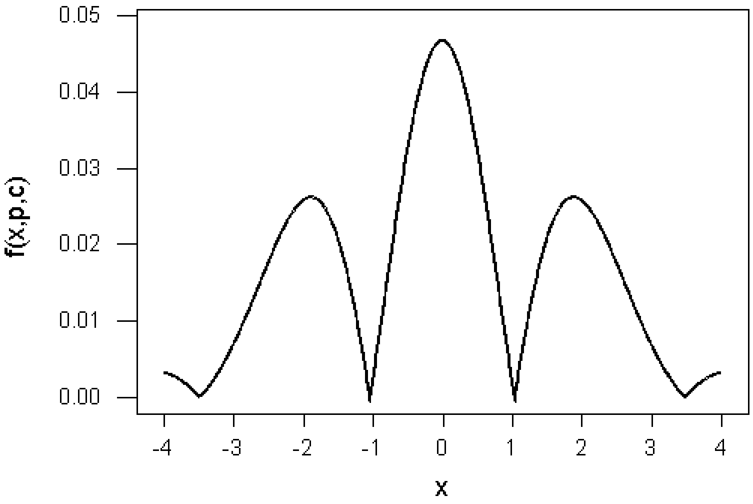

Fig. 2. Example of $f(x, p, c)$ for $p = \frac{1}{2}$ and $c = 2$

From empirical evidence (see also Figure 2) it can be conjectured that $f(x) \equiv f(x, p, c)$ is maximized at $x = 0$.

2.1 *Conjecture*

Let $p \in (0, 1)$ and $c > 1$ be given and fixed. Then

$$\sup_{-\infty<x<\infty} f(x, p, c) = f(0, p, c) = 1 - p + \frac{p}{c} - \frac{1}{\sqrt{1 - p + pc^2}} \tag{2}$$

We are able to provide the following (weaker) result.

2.2 *Theorem*

Let $p \in (0, 1)$ and $c > 1$ be given and fixed. Then:

a) $f(0, p, c) = 1 - p + \frac{p}{c} - \frac{1}{\sqrt{1 - p + pc^2}} > 0$ and,
b) $f(x)$ has a *local maximum* at $x = 0$.

Proof.

a) The function $g(x) = x^{-1/2}$ is *strictly* convex on the interval $[0, 1]$, since $g'(x) = -\frac{1}{2}x^{-3/2}$ and $g''(x) = \frac{3}{4}x^{-5/2} > 0$. Therefore, $g((1 - p)1 + pc^2) < (1-p)g(1)+pg(c^2)$, or $1-p+\frac{p}{c} - \frac{1}{\sqrt{1-p+pc^2}} = f(0, p, c) > 0$.
b) Consider $\frac{df}{dx}$ which equals

$$
\begin{aligned}
&(1-p)e^{-x^2/2}(-x)+\frac{p}{\tau}e^{-\frac{1}{2}x^2/\tau^2}(-x/\tau^2)\\
&\qquad -\frac{1}{\sqrt{1-p+p\tau^2}}e^{-\frac{1}{2}x^2/(1-p+p\tau^2)}\frac{-x}{1-p+p\tau^2}\\
&\quad =\left[\frac{1}{(1-p+p\tau^2)^{3/2}}e^{-\frac{1}{2}x^2/(1-p+p\tau^2)}-(1-p)e^{-x^2/2}-\frac{p}{\tau^3}e^{-\frac{1}{2}x^2/\tau^2}\right]x
\end{aligned}
$$

in a neighborhood of $x = 0$. Clearly, $\frac{df}{dx} = 0$ at $x = 0$. Note in addition that

$$
\left.\frac{d^2f}{(dx)^2}\right|_{x=0} = \frac{1}{(1-p+p\tau^2)^{3/2}} - (1-p+p/\tau^3).
$$

Since $g(x) = x^{-3/2}$ is *strictly* convex on the interval $[0, 1]$, it follows that

$$
\begin{aligned}
&0 > g(1-p+p\tau^2) - (1-p)g(1) + pg(\tau^2)\\
&\quad = \frac{1}{(1-p+p\tau^2)^{3/2}} - (1-p+p/\tau^3).
\end{aligned}
$$

This is part b) of the Theorem and ends the proof.

It is quite intuitive to assume that the statement "*the larger the contanimation e.g. the constant c, the larger the difference in the distance between the two densities*" is correct. This, however, needs confirmation as the following result provides.

2.3 Theorem

Let p as in Lemma 1. Let further be $f_p(c) = \sup_{-\infty<x<\infty} f(x, p, c)$ defined for $c > 1$. Then, $f_p(c)$ is *monotone increasing* in c.

Proof. It is sufficient to show that $f_p(c) = 1 - p + \frac{p}{c} - \frac{1}{\sqrt{1-p+pc^2}}$ is monotone increasing in c. We show that $f_p'(c) = -p\frac{1}{c^2} + (1-p+pc^2)^{-3/2}pc$ is positive. This can be accomplished by showing that $(1-p+pc^2)^{-3/2} - \frac{1}{c^3}$ is positive, as the factor pc is positive. Or, if $(1-p+pc^2)^{-3/2} > \frac{1}{c^3}$, or if $1-p+pc^2 < \left(\frac{1}{c^3}\right)^{-2/3} = c^2$. The latter inequaltiy is holding, since $p < 1$. This ends the proof.

Similarly, one might assume that the statement "*the larger the amount of contanimation e.g. the proportion constant p, the larger the difference in the distance between the two densities*" is true. Typically, one could have the impression that a maximum amount of contamination is provided with $p = 0.5$. In a simulation study investigating the power of the Shapiro-Wilks-test by means of the scale-contaminated normal, Chen (1971) writes as follows: "*$P(n, \lambda, p)$ is an increasing function of n, λ, and p*"[1]. This statement is needs to be read with care with regard to p as the following Theorem is illuminating.

2.4 Theorem

Let c as in Lemma 1. Let further be $f_c(p) = \sup_{-\infty<x<\infty} f(x, p, c)$ defined for $c > 1$. Then, $f_c(p)$ is maximized in $(0, 1)$ by

$$p_{\max} = \frac{\left(\frac{c(c+1)}{2}\right)^{2/3} - 1}{c^2 - 1}. \tag{3}$$

Proof. It is sufficient to show that $f_c(p) = 1 - p + \frac{p}{c} - \frac{1}{\sqrt{1 - p + pc^2}}$ is maximized at $p_{\max}$. Now, consider $f_c'(p) = \frac{1}{c} - 1 + \frac{1}{2}(1 - p + pc^2)^{-3/2}(c^2 - 1)$ and equate it to 0. This leads to

$$(1 - p + pc^2)^{-3/2} = \frac{2}{c(c+1)} \tag{4}$$

or,

$$1 - p + pc^2 = \left(\frac{c(c+1)}{2}\right)^{2/3} \tag{5}$$

from which the result $p_{\max} = \frac{C - 1}{c^2 - 1}$ with $C = \left(\frac{c(c+1)}{2}\right)^{3/2}$ follows. Clearly, $p_{\max} > 0$. We show $p_{\max} < 1$. This can be accomplished by schowing that $\left(\frac{c(c+1)}{2}\right)^{2/3} < c^2$, or, equivalently $c(c+1) < 2c^3$. The latter inequality holds for $c > 1$, evidently. Furthermore, $f_c''(p) = -\frac{3}{4}(1 - p + pc^2)^{-5/2}(c^2 - 1)^2 < 0$ for all $p \in [0, 1]$ showing that $f_c(p)$ is strictly concave in p. This ends the proof.

3 Discussion

As Figure 3 illustrates the value of $p_{\max}$ drops down quickly if the contamination constant c increases. For example, for $c = 2$ we find that $p_{\max} = 0.3600$, for $c = 5$ the maximizing value is $p_{\max} = 0.2118$. These results provide some guidance as where to expect the maximum power of tests for detecting departure from normality when using scale-contaminated normals as basis for the power study.

[1] The Power of the test under consideration; $\lambda = c$ in our notation; *The Authors*

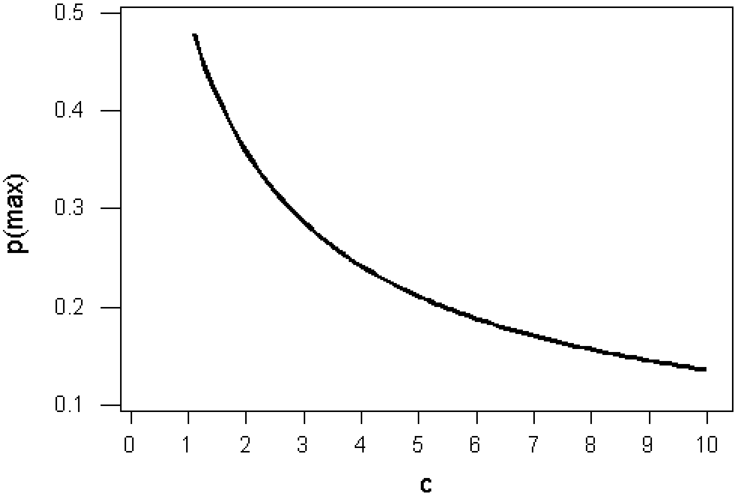

Fig. 3. p_{max} as a function of scale-contanimation constant c

Acknowledgements. This research is done under support of the *German Research Foundation.*

References

Barnett V, Lewis T (1983) Outliers in statistical data. Second Edition, Wiley, London

Chen EH (1971) The power of the Shapiro-Wilk W test for normality in samples from contaminated normal distributions. Journal of the American Statistical Association 66:760–762

Chotiwittayatarakorn S (1987) A comparison of power of some standard goodness-of-fit tests of normality. M.S. Thesis in Statistics. Bangkok: Graduate School, Chulalongkorn University

Chuchao K (1992) A comparison of power of some statistical methods for testing normality. M.S. Thesis in Statistics. Pathumthani: Faculty of Science and Technology, Thammassat University

Jantako K (1993) On the comparison of location and scale parameter estimators of contaminated normal. M.S. Thesis in Statistics. Pathumthani: Faculty of Science and Technology, Thammassat University

Ruangroj R (2000) A comparison on power of tests for normality among chi-square, Shapiro-Wilk W statistic and Shapiro-Francia W' statistic. M.Sc. Thesis in Biostatistics. Bangkok: Faculty of Public Health, Mahidol University