Characterizing Density Crossing Points

By Helmut Finner, Markus Roters and Thorsten Dickhaus [1]

December, 2006

Extended Online Version of

Finner, H., Roters, M. and Dickhaus, T. (2007). Characterizing Density Crossing Points. The American Statistician 61, 28-33.

This article addresses the following question: Where and how often do distinct probability density functions cross? For example, what can be said about crossing points between normal and Student's t or chi-square densities, respectively, and what are the asymptotic crossing points if the degrees of freedom tend to infinity? Crossing points are especially of interest if, for instance, the normal is used as an approximation according to the central limit theorem because they determine the total variation distance and the areas of over- and underestimation of actual probabilities. They are strongly related to the behavior of the density ratio, a ubiquitous quantity in many fields of statistics like Neyman-Pearson or decision theory. We discuss several examples with respect to the standard normal density φ and provide elegant and appealing limiting forms for asymptotic crossing points of φ with respect to densities of standardized sums appearing in the central limit theorem. Finally, we derive and investigate the crossing points of φ with the t-density with ν degrees of freedom.

KEY WORDS: Chi-square distribution; Implicit function theorem; Gamma distribution; Local limit theorem; Monotone likelihood ratio; Variation diminishing transformation.

1. INTRODUCTION

In statistical practice, the normal distribution is often used as an approximation for other probability distributions, for example, for the t-distribution

[1] Helmut Finner is Apl. Professor (E-mail: *finner@ddz.uni-duesseldorf.de*) and Thorsten Dickhaus is Ph.D. student (E-mail: *dickhaus@ddz.uni-duesseldorf.de*), German Diabetes Center, Leibniz Center at Heinrich-Heine-University Duesseldorf, Institute of Biometrics and Epidemiology, Duesseldorf, Germany. Markus Roters is Apl. Professor and Senior Statistician at the Biometrics Department of Omnicare Clinical Research, Koeln, Germany (E-Mail: *Markus.Roters@omnicarecr.com*). The authors are grateful to two referees, an associate editor and the editor for their helpful comments and suggestions.

with large degrees of freedom or if the central limit theorem (CLT) applies. Then the question arises whether the normal approximation works well and how large the approximation error can get. An apparently simple question is for which areas the normal approximation leads to under- or overestimation of the true probabilities. Denote the probability density function (pdf) of the standard normal distribution with mean 0 and variance 1 by φ and suppose that the distribution to be approximated has a Lebesgue density g. Then the issues raised before immediately lead to the question

"Where and how often do the underlying pdf's g and φ cross?"

Hence, we ask for all solutions of the equation

$$g(z) = \varphi(z). \tag{1}$$

Any solution of (1) will be called a crossing point (CP). Alternatively, one may also ask where the likelihood ratio $g(z)/\varphi(z)$ is greater than, less than or equal to 1.

In order to understand the behavior of a family of univariate distributions with pdf's g_ϑ depending on a parameter $\vartheta \in \Theta \subseteq \mathbb{R}$, crossing points between distinct g_ϑ's are of interest, too. For example, suppose we fix some reference point $\vartheta_0 \in \Theta$. Then it is near at hand to study all solutions of the equation

$$g_{\vartheta_0}(z) = g_\vartheta(z). \tag{2}$$

Consider the following simple example with $g_{\vartheta_0} = \varphi$ and $g_\vartheta = \varphi_\sigma$ (say), where φ_σ denotes the pdf of a normal distribution with mean 0 and standard deviation σ with $0 < \sigma \neq 1$. Here, an explicit formula for the crossing points, which is analytically easy to handle, can be given, and we obtain exactly two crossing points $z_i = z_i(\sigma)$, $i = 1, 2$, given by

$$z_1 = -z_2 = \sqrt{\frac{2\sigma^2}{\sigma^2 - 1} \ln(\sigma)}.$$

We note that $z_1(\sigma)$ is strictly increasing in σ with $\lim_{\sigma \to 0+} z_1(\sigma) = 0$, $\lim_{\sigma \to \infty} z_1(\sigma) = \infty$ and $\lim_{\sigma \to 1} z_1(\sigma) = 1$. Setting $B_\sigma = \{z \in \mathbb{R} : \varphi(z) > \varphi_\sigma(z)\}$, we obtain

$$B_\sigma = \begin{cases} (-\infty, -z_1(\sigma)) \cup (z_1(\sigma), \infty), & \text{if } 0 < \sigma < 1, \\ (-z_1(\sigma), z_1(\sigma)), & \text{if } \sigma > 1. \end{cases}$$

Obviously, B_σ increases if σ moves away from 1. It should be mentioned at this point that this monotonicity behavior is a special case of the more general situation in one-parametric exponential families if the parameter is a strictly monotonic function of the natural parameter of the family. The monotonicity behavior of the CPs and the sets B_σ in σ have some intuitive interpretation, for example, in the hypotheses testing problem

$$H = \{\varphi\} \text{ versus } K = \{\varphi_\sigma\}$$

for some fixed σ with $0 < \sigma \neq 1$. A common interpretation of the likelihood ratio is that H is more likely if $\varphi(z)/\varphi_\sigma(z) > 1$ and that K is more likely if $\varphi(z)/\varphi_\sigma(z) < 1$. Therefore, crossing points of densities play an important role in hypotheses testing. According to the Neyman-Pearson lemma, the best level α test for testing H versus K is of the type

$$\phi_c(z) = \begin{cases} 1, & \text{if } \varphi(z)/\varphi_\sigma(z) \quad \leq c, \\ 0, & \text{if } \varphi(z)/\varphi_\sigma(z) \quad > c, \end{cases}$$

where c is chosen such that $\int \phi_c \varphi d\lambda = \alpha$. A value of $c > 1$ indicates that the test is not very specific. Anyhow, if we take $c = 1$ as a threshold for or against H, one would expect that more z-values lead to a decision in favor of H if σ moves away from 1. Noting that $B_\sigma = \{z : \phi_1(z) = 0\}$, this intuitive reasoning is confirmed by the monotonicity behavior of the crossing points $z_1(\sigma) = -z_2(\sigma)$. Finally, it can be shown that ϕ_1 is a Bayes test for H versus K if we specify equal prior probabilities and equal weights for the losses due to wrong decisions.

Unfortunately, (1) or (2) often cannot be solved analytically. However, a numerical solution should not be a big deal, provided that there are not too many CPs. For instance, if the family of distributions under consideration has a strict monotone likelihood ratio (see, e.g., Lehmann and Romano 2005, p. 65), that is, $g_{\vartheta_0}(z)/g_\vartheta(z)$ is strictly increasing (or strictly decreasing) in z, then there obviously exists a unique solution of (2), hence a unique CP.

There are several ways to visualize CP behavior. One may plot (i) $g_{\vartheta_1}(z)$ and $g_{\vartheta_2}(z)$ in one picture, (ii) the difference $g_{\vartheta_1}(z) - g_{\vartheta_2}(z)$, (iii) the likelihood ratio $\lambda(z) = g_{\vartheta_1}(z)/g_{\vartheta_2}(z)$, or, (iv) the log-likelihood ratio $\ln(\lambda(z))$. In this article we mainly display the likelihood ratio with one exception.

Figure 1 displays some likelihood ratios $\lambda_\nu(z) = f_\nu(z)/\varphi(z)$ as functions of z, where f_ν denotes the pdf of a t-distribution with ν degrees of freedom. Apparently, for any ν there seem to be exactly two crossing points $z_{\nu,1} = -z_{\nu,2}$ each of which seems to converge monotonically to some fixed value. In Section 3 we will proof this crossing point behavior and give an elegant limiting form for the likelihood ratio which immediately yields the

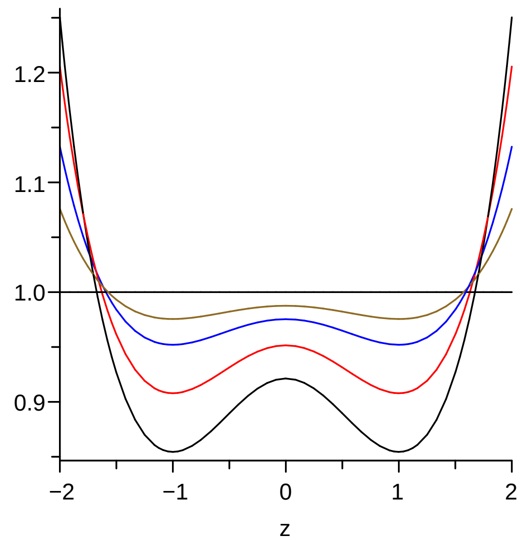

Figure 1: $\lambda_\nu(z) = f_\nu(z)/\varphi(z)$ *for* $\nu = 3, 5, 10, 20$, *and* $-2 \le z \le 2$. *The curves can be identified by noting that* $\lambda_\nu(0)$ *is increasing in* ν. *Asymptotic crossing points (ACPs) are* $\pm\sqrt{1+\sqrt{2}}$ *as pointed out in Section 3.*

asymptotic crossing points (ACPs). However, the following example shows that the behavior of CPs can be rather complicated.

Example 1. Let $X_i,\ i \in \mathbb{N}$, denote a sequence of independent identically distributed (iid) random variables with pdf

$$h_\mu(z) = p_1\varphi(z+\mu) + p_2\varphi(z-\mu) + p_3\varphi(z) \tag{3}$$

for some fixed $\mu \neq 0$ and $p_i \in [0,1]$ with $p_1 + p_2 + p_3 = 1$, hence a mixture of at most three normals with means 0 and $\pm\mu$ and variance 1. Notice that $\mathrm{E}X_i = (p_2 - p_1)\mu$ and $\mathrm{var}X_i = 1 + \mu^2(p_1 + p_2 - (p_1 - p_2)^2)$. Consider the standardized sum

$$S_n = \frac{\sum_{i=1}^n (X_i - \mathrm{E}X_1)}{\sqrt{n\mathrm{var}X_1}},\ n \in \mathbb{N}.$$

We denote the pdf of S_n by $h_n(\cdot|\mu, p_1, p_2)$. Clearly, the CLT applies for S_n and $\lim_{n\to\infty} h_n(z|\mu, p_1, p_2) = \varphi(z)$ for all $z \in \mathbb{R}$. We first consider a symmetric mixture of two normals by choosing $p_1 = p_2 = 1/2$ and $p_3 = 0$.

Figure 2 displays the likelihood ratio $h_{50}(z|10, 1/2, 1/2)/\varphi(z)$ and shows that there may be many CPs (exactly 50 in the displayed area), possibly some more in tail areas which are not displayed here. What happens if n becomes larger and/or the weights p_i change?

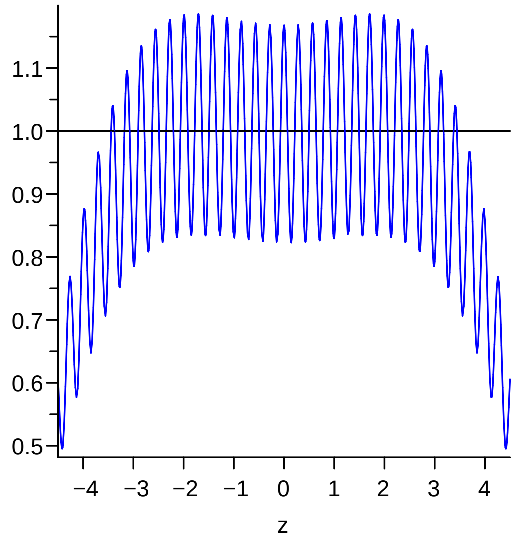

Figure 2: *Likelihood ratio behavior of* $h_{50}(z|10,1/2,1/2)/\varphi(z)$.

The picture on the left hand side of Figure 3 displays the likelihood ratios $h_n(z|10,1/2,1/2)/\varphi(z)$ for $n = 250$ and $n = 500$. Now it seems that there are only four CPs left in the displayed area. In Section 2, it will be shown that $\pm\sqrt{3 \pm \sqrt{6}}$ are the asymptotic crossing points for $n \to \infty$ (independently of $\mu \neq 0$). The picture on the right hand side of Figure 3 displays the likelihood ratios $h_n(z|10,1/6,1/6)/\varphi(z)$ for $n = 100$ and $n = 200$. Now it seems that there are six ACPs in the displayed area. In fact, the ACPs are given by $\pm\sqrt{5+2b}, \pm\sqrt{5 \pm a - b}$, where $c = \arctan(\sqrt{6}/2)/3$, $a = \sqrt{30}\sin(c)$, $b = \sqrt{10}\cos(c)$; see Section 2.

The remaining part of the article is organized as follows. In Section 2 we are first concerned with ACPs between normal and the pdf of the standardized sum if the CLT applies and provide elegant and appealing limiting forms for ACPs. Then we pick up Example 1 with various choices of the weights p_i and derive the corresponding ACPs. The gamma distribution including the χ^2-distribution will be studied in more detail. Finally, we study the CP behavior between normal and t-densities in Section 3, which is not covered by the theory in Section 2. Among others, it will be shown that the ACPs are given by $\pm\sqrt{1+\sqrt{2}}$ if the degrees of freedom ν tend to infinity. The standardized t-distribution with mean zero and variance 1 leads to four ACPs given by $\pm\sqrt{3 \pm \sqrt{6}}$. In Appendix A we give a proof for the number of CPs between gamma and normal pdf's and a proof for the monotonicity of the CPs between normal and t_ν pdf's. Appendix B deals with the total variation distance and its connection to CPs.

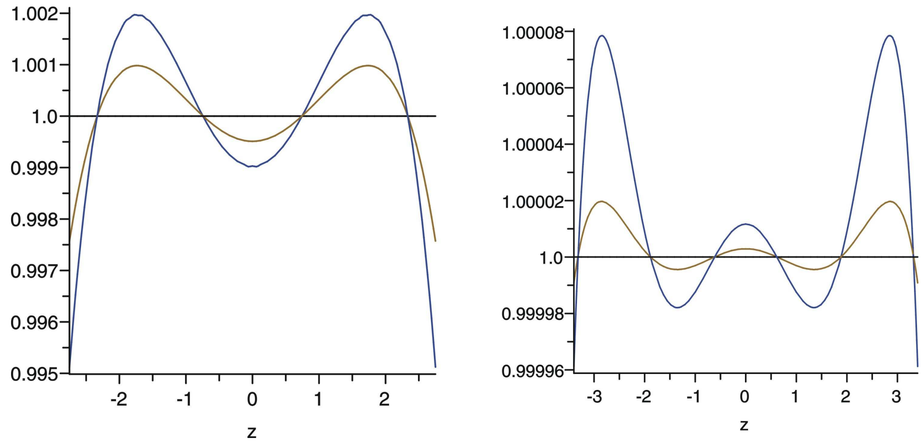

Figure 3: *Likelihood ratio behavior of (i)* $h_{250}(z|10, 1/2, 1/2)/\varphi(z)$ *and* $h_{500}(z|10, 1/2, 1/2)/\varphi(z)$ *(left picture), and, (ii)* $h_{100}(z|10, 1/6, 1/6)/\varphi(z)$ *and* $h_{200}(z|10, 1/6, 1/6)/\varphi(z)$ *(right picture). The curves can be distinguished by noting that the likelihood ratio flattens for increasing* n*. ACPs are given by* $\pm\sqrt{3 \pm \sqrt{6}}$ *(left picture), and* $\pm 0.617, \pm 1.889, \pm 3.324$ *(right picture).*

2. CENTRAL LIMIT THEOREM AND ASYMPTOTIC CROSSING POINTS

Of general interest is the crossing point behavior between the normal pdf φ and pdf's g_n, where g_n is the pdf of standardized sums

$$S_n = n^{-1/2} \sum_{i=1}^{n} (X_i - \mu)/\sigma$$

of n iid real-valued continuous random variables X_i with $\mu = \mathrm{E}X_1$ and $\sigma^2 = \mathrm{var}X_1$. In this section we derive a general solution based on a local limit theorem for g_n. To this end, we need some notation. Let

$$H_m(x) = m! \sum_{k=0}^{[m/2]} \frac{(-1)^k x^{m-2k}}{k!(m-2k)!2^k}, \ m \in \mathbb{N}_0,$$

denote the Chebyshev-Hermite polynomials as defined by Petrov (1975, p. 137), where $[a]$ denotes the largest integer $\leq a$. For example, $H_0(x) = 1$, $H_1(x) = x$, $H_2(x) = x^2 - 1$, $H_3(x) = x^3 - 3x$, $H_4(x) = x^4 - 6x^2 + 3$, $H_5(x) = x^5 - 10x^3 + 15x$ and $H_6(x) = x^6 - 15x^4 + 45x^2 - 15$. Moreover, for $r \in \mathbb{N}$, let

$$q_r(x) = (2\pi)^{-1/2} \exp(-x^2/2) \sum H_{r+2s}(x) \prod_{m=1}^{r} \frac{1}{k_m!} \left(\frac{\gamma_{m+2}}{(m+2)!\sigma^{m+2}} \right)^{k_m},$$

where the summation on the right hand side is carried out over all non-negative integer solutions $(k_1, \ldots, k_r)$ of the equality $\sum_{i=1}^{r} ik_i = r$. In this formula, $s = s(k_1, \ldots, k_r) = \sum_{i=1}^{r} k_i$ and γ_k denotes the cumulant of order k of the distribution of X_1. In terms of central moments $\mu_i = \mathrm{E}(X_1 - \mu)^i$, $i \in \mathbb{N}$, $\gamma_1, \ldots, \gamma_6$ are given by $\gamma_1 = 0$, $\gamma_2 = \mu_2$, $\gamma_3 = \mu_3$, $\gamma_4 = \mu_4 - 3\mu_2^2$, $\gamma_5 = \mu_5 - 10\mu_2\mu_3$, $\gamma_6 = \mu_6 - 15\mu_2\mu_4 + 30\mu_2^3 - 10\mu_3^2$.

The next well-known theorem provides the basis for the computation of ACPs with respect to φ and g_n.

Theorem 1. *(Gnedenko 1948; see also Petrov 1975, pp. 206-207). Let X_i, $i \in \mathbb{N}$, denote a sequence of iid real-valued random variables with $EX_1 = \mu$, $varX_1 = \sigma^2 \in (0, \infty)$, and $E|X_1 - \mu|^k < \infty$ for some integer $k \geq 3$. Let g_n denote the pdf of the standardized sum S_n and assume that g_n is bounded for some $n \in \mathbb{N}$. Then*

$$g_n(x) - \varphi(x) = \sum_{r=1}^{k-2} \frac{q_r(x)}{n^{r/2}} + o(n^{-(k-2)/2}) \quad \text{uniformly in } x \in \mathbb{R}. \tag{4}$$

With this result ACPs are handed to us on a silver platter.

Theorem 2. *(ACP Theorem for CLT pdf's.) Let the conditions of Theorem 1 be satisfied and let k denote the smallest integer such that $E|X_1|^k$ is finite and $\gamma_k \neq 0$. Then the roots of the Chebyshev-Hermite polynomial H_k are ACPs and the sign of $\gamma_k H_k(x)$ determines where $g_n(x)/\varphi(x)$ is < 1 (> 1) on suitably chosen intervals between the ACPs.*

Proof: The assertion follows immediately by verifying that, under the assumptions of the theorem, equation (4) is equivalent to

$$n^{(k-2)/2}(g_n(x) - \varphi(x)) = \varphi(x) H_k(x) \frac{\gamma_k}{k!\sigma^k} + o(1) \quad \text{uniformly in } x \in \mathbb{R}.$$

Notice that the equation $H_k(x) = 0$ has exactly k distinct solutions in $\mathbb{R}$, hence the ACP theorem yields a set C_k (say) of k distinct ACPs. We get

$$\begin{aligned} C_3 &= \{0, \pm\sqrt{3}\} \cong \{0, \pm 1.732\}, \\ C_4 &= \{\pm\sqrt{3 \pm \sqrt{6}}\} \cong \{\pm 0.742, \pm 2.334\}, \\ C_5 &= \{0, \pm\sqrt{5 \pm \sqrt{10}}\} \cong \{0, \pm 1.356, \pm 2.857\}, \end{aligned}$$

and, with $c = \arctan(\sqrt{6}/2)/3$, $a = \sqrt{30}\sin(c)$, $b = \sqrt{10}\cos(c)$,

$$C_6 = \{\pm\sqrt{5 + 2b}, \pm\sqrt{5 \pm a - b}\} \cong \{\pm 0.617, \pm 1.889, \pm 3.324\}.$$

Moreover, on every finite but fixed interval each CP between g_n and φ is close to an ACP if n is sufficiently large. However, under the assumptions of Theorem 2 there may be more crossing points x_n depending on n with $x_n \to \pm\infty$. These CPs cannot be determined with the method described before.

We now apply Theorem 2 to three distributional settings. The main task consists in determining the cumulants of the distributions under consideration.

Example 2. (Example 1 continued). (a) For $p_1 = p_2 = 1/2$, $p_3 = 0$ we obtain $\gamma_3 = 0$ and $\gamma_4 = -2\mu^4$, hence, four ACPs are given by $C_4 = \{\pm\sqrt{3 \pm \sqrt{6}}\}$. (b) For $p_1 = p_2 = 1/6$, $p_3 = 2/3$ we obtain $\gamma_3 = \gamma_4 = \gamma_5 = 0$ and $\gamma_6 = -2\mu^6/9$, hence, six ACPs are given by $C_6 \cong \{\pm 0.617, \pm 1.889, \pm 3.324\}$. (c) For $p_1 = 1/6$, $p_2 = 1/3$, $p_3 = 1/2$ we obtain $\gamma_3 = -2\mu^3/27 \neq 0$, hence, three ACPs are given by $C_3 = \{0, \pm\sqrt{3}\}$.

Example 3. Suppose the X_i's in Theorem 2 follow a uniform distribution on $[-a, a]$, $a > 0$ fixed. Then $\gamma_3 = 0$ and $\gamma_4 = a^4/5 \neq 0$, hence C_4 applies for ACPs.

Example 4. Suppose the X_i's in Theorem 2 follow a gamma distribution with pdf $f_{\tau,\theta}$ (with $\tau, \theta > 0$) given by

$$f_{\tau,\theta}(x) = \frac{x^{\tau-1}\exp(-x/\theta)}{\Gamma(\tau)\theta^{\tau}} I_{[0,\infty)}(x). \tag{5}$$

Since the gamma distribution has mean $\mu \equiv \mu(\tau, \theta) = \tau\theta$ and standard deviation $\sigma \equiv \sigma(\tau, \theta) = \sqrt{\tau}\theta$, the modified density $\tilde{f}_\tau(x) = \sigma f_{\tau,\theta}(\sigma x + \mu)$ given by

$$\tilde{f}_\tau(x) \quad = \quad \frac{\sqrt{\tau}(\tau + \sqrt{\tau}x)^{\tau-1}\exp(-(\tau + \sqrt{\tau}x))}{\Gamma(\tau)} \, I_{[-\sqrt{\tau},\infty)}(x)$$

corresponds to a standardized gamma distribution with mean 0 and standard deviation 1 and is independent of θ. For $\tau = \nu/2$ we obtain the standardized χ^2-distribution with ν degrees of freedom. Moreover, S_n has pdf $\tilde{f}_{n\tau}$. Figure 4 shows the shape of $\tilde{f}_{2.5}(x)$ and $\tilde{f}_{20}(x)$ together with $\varphi(x)$. Noting that $\gamma_3 = 2\tau\theta^3 \neq 0$ for the gamma distribution with pdf $f_{\tau,\theta}$, three ACPs are given by $C_3 = \{0, \pm\sqrt{3}\}$.

For $\tau > 1$, the equation $\tilde{f}_\tau(x) = \varphi(x)$ has exactly three distinct solutions $x_i(\tau) \in \mathbb{R}$, $i = 1, 2, 3$, with $-\sqrt{\tau} < x_1(\tau) < -1 < x_2(\tau) < 0$ and $1 < x_3(\tau)$. The proof for this assertion is given in Appendix A. We conjecture that the crossing points $x_1(\tau)$ and $x_3(\tau)$ are strictly decreasing in $\tau > 1$ and

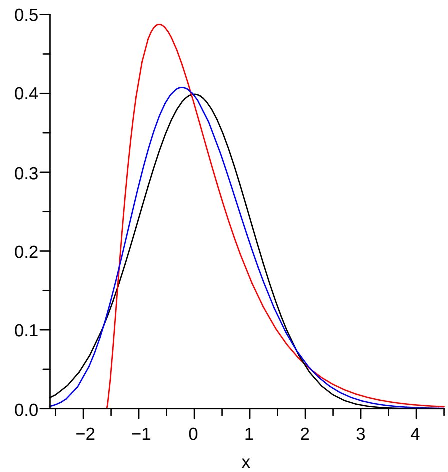

Figure 4: $\tilde{f}_{2.5}(x)$, $\tilde{f}_{20}(x)$ *and* $\varphi(x)$. *The curves can be identified by noting that the mode of the standardized gamma distribution is decreasing in* τ. *The ACPs are given by* 0 *and* $\pm\sqrt{3}$.

the crossing points $x_2(\tau)$ are strictly increasing in $\tau > 1$. If this conjecture holds true we get that $x_1(\tau) \in (-\sqrt{\min\{3,\tau\}}, -1)$, $x_2(\tau) \in (-\kappa, 0)$, $x_3(\tau) \in (\sqrt{3}, 2+\kappa)$ for $\tau > 1$, where $\kappa = (3 - \ln(2) - \ln(\pi))^{1/2} - 1 \approx 0.078$.

For $\tau \in (0, 1]$ the situation is somewhat different and will not be considered here. Notice that $f_{\tau,\theta}(0) = 0$ for $\tau > 1$, but $f_{1,\theta}(0) = 1/\theta$ and $\lim_{x\to 0} f_{\tau,\theta}(x) = \infty$ for $\tau \in (0, 1)$. It even holds that $\lim_{\tau\to\infty} \tilde{f}_\tau(x) = \varphi(x)$ for all $x \in \mathbb{R}$.

3. CROSSING POINTS OF NORMAL WITH T

In this section we study the crossing points between t-densities f_ν and the standard normal density φ. The density of the t_ν-distribution is given by

$$f_\nu(z) \quad = \quad \frac{\Gamma\left(\nu/2 + 1/2\right)}{\Gamma\left(\nu/2\right)} \frac{1}{\sqrt{\nu\,\pi}} \left(1 + \frac{z^2}{\nu}\right)^{-\nu/2 - 1/2}, \; z \in \mathbb{R}.$$

It is well known that t-densities approach the standard normal density φ if the degrees of freedom tend to infinity. Although it seems common knowledge that there are exactly two crossing points between the t- and normal densities (cf. the discussion around Figure 1), some effort seems necessary to prove this fact. As main results it will be shown that the ACPs for $\nu \to \infty$ are given by $\pm\sqrt{1+\sqrt{2}}$ and that the CPs are monotone with respect to the degrees of freedom. Ingredients for proving these results are series expansions, the implicit function theorem and the theory of variation diminishing transformations. Especially the monotonicity of the CPs seems to be a hard nut to crack and is deferred to Appendix A. Similarly as in Section 1, some intuitive reasoning for the monotonicity can be given in terms of testing hypotheses by considering the sets $\{z : \varphi(z)/f_\nu(z) > 1\}$ which turn out to be decreasing in ν.

One might expect similar results for the standardized t_ν-distribution with variance 1, which differs only slightly from the original t_ν-distribution with variance $\nu/(\nu-2)$ if $\nu > 2$. Perhaps surprisingly, in the case of the standardized t_ν-distribution four ACPs occur and are given by $C_4 = \{\pm\sqrt{3 \pm \sqrt{6}}\}$.

First, we summarize some useful facts. The family of t-densities $\{f_\nu : \nu \in \Theta\}$, $\Theta = (0, \infty]$, (including the limit $\varphi = f_\infty$) does not have a monotone likelihood ratio in (z, ν) but has a strict monotone likelihood ratio on certain distinct intervals, i.e., for $0 < \nu < \mu \leq \infty$, $f_\nu(z)/f_\mu(z)$ is strictly increasing on $[-1, 0]$ and $[1, \infty)$, and strictly decreasing on $(-\infty, -1]$ and $[0, 1]$.

The following lemma gives the number of crossing points between t-densities with different degrees of freedom and between t-densities and φ.

Lemma 1. *For each pair $\nu, \mu \in \Theta$, $\nu \neq \mu$, the equation $f_\nu(z) = f_\mu(z)$ has a unique solution $z_{\nu,\mu}$ on $(0, \infty)$, and, $z_{\nu,\mu} > 1$.*

Proof: We first show that the function $g : (0, \infty) \times (0, \infty) \to (0, \infty)$ defined by

$$g(z, a) = \ln\left(\frac{\Gamma(z + a)}{\Gamma(z) z^a}\right) \tag{6}$$

is strictly increasing in $z > 0$ for $a \in (0, 1)$ and strictly decreasing in $z > 0$ for $a \in (1, \infty)$. To this end, let $\Psi(x) = (d/dx)\ln(\Gamma(x))$ for $x > 0$ denote the digamma function. With the representation

$$\Psi(x) = \int_0^\infty [\exp(-t)/t - \exp(-xt)/(1 - \exp(-t))]dt$$

we easily obtain

$$\begin{aligned} \frac{\partial}{\partial z} g(z, a) &= \Psi(z + a) - \Psi(z) - \frac{a}{z} \\ &= \int_0^\infty \exp(-zt) \left(\frac{1 - \exp(-at)}{1 - \exp(-t)} - a\right) dt. \end{aligned}$$

Obviously, the integrand of the last expression is strictly positive for $a \in (0, 1)$ and strictly negative for $a > 1$, hence, the assertion follows. Now, by choosing $a = 1/2$ and $z = \nu/2$ in (6) we immediately obtain that $f_\nu(0)$ is strictly increasing in $\nu \in \Theta$. Moreover,

$$\lim_{z\to\infty} \frac{f_\nu(z)}{f_\mu(z)} = \infty \text{ for all } 0 < \nu < \mu \leq \infty.$$

Now the assertion follows by combining these facts with the likelihood ratio properties mentioned at the beginning of this section.

In the following theorem, the ACPs between φ and f_ν are given for ν tending to infinity.

Theorem 3. *Let $z_\nu > 0$ be the positive solution of $f_\nu(z) = \varphi(z)$, $\nu > 0$. Then*

$$\lim_{\nu\to\infty} z_\nu = \sqrt{1+\sqrt{2}} = 1.553773974\ldots$$

Proof: From Fisher (1925), who gave an expansion for the logarithm of the ratio of t- and normal density, we have

$$\ln(f_\nu(z)) - \ln(\varphi(z)) = \frac{z^4 - 2z^2 - 1}{4\nu} + O\left(\nu^{-2}\right).$$

Therefore, the equation $f_\nu(z) = \varphi(z)$ is equivalent to

$$\nu(z^4 - 2z^2 - 1) = O(1).$$

From Lemma 1 we know that the latter equation has a unique solution $z_\nu \in (1, \infty)$. Noting that $O(1)$ is uniformly bounded in z on compact intervals, the assertion follows.

Remark 1. By using Fisher's (1925) higher order expansion

$$\ln(f_\nu(z)) - \ln(\varphi(z)) = \frac{z^4 - 2z^2 - 1}{4\nu} - \frac{2z^6 - 3z^4}{12\nu^2} + O\left(\nu^{-3}\right)$$

and setting $d(z, \nu) = 4\nu\left[\ln(f_\nu(z) - \ln(\varphi(z))\right]$, we get the representation

$$d(z, \nu) = (z^4 - 2z^2 - 1) + \frac{1}{\nu}(z^4 - \frac{2}{3}z^6) + O(\nu^{-2}).$$

With $u = z^2$ we obtain a cubic equation. Its unique solution on $(1, \infty)$ is asymptotically given by

$$z_\nu = \sqrt{1+\sqrt{2}} + \frac{C}{\nu} + O(\nu^{-2}) \quad \text{with} \quad C = \frac{1}{24}\frac{8+5\sqrt{2}}{\sqrt{1+\sqrt{2}}}.$$

This representation can be used as an approximation of the exact crossing point, even in the finite case.

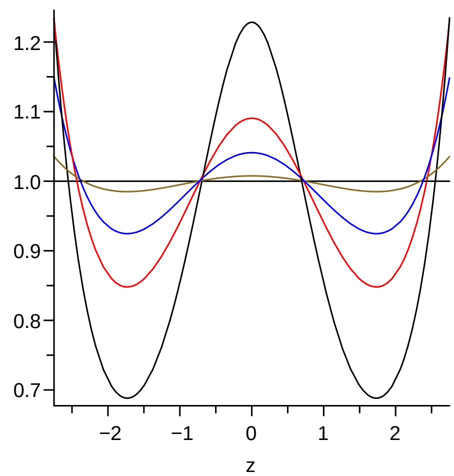

Figure 5: *Likelihood ratio* $\tilde{\lambda}_\nu(z) = \tilde{f}_\nu(z)/\varphi(z)$ *for* $\nu = 5, 10, 20, 100$. *The curves can be identified by noting that* $\tilde{\lambda}_\nu(0)$ *is decreasing in* ν.

Finally, it can be shown (see Appendix A) that the z_ν's are decreasing in ν. Setting $B_{\nu,\mu} = \{z \in \mathbb{R} : f_\nu(z)/f_\mu(z) > 1\}$, this is equivalent to

$$B_{\nu,\mu} \subsetneq B_{\nu,\eta} \text{ for } 0 < \nu < \mu < \eta \leq \infty,$$

in accordance with the hypotheses testing interpretation outlined in Section 1.

Remark 2. It is remarkable that standardization of the t_ν-distribution leads to four crossing points. Suppose the random variable X has a t_ν-distribution. Then, for $\nu > 2$, the standardized variable $Y = X/\sqrt{\nu/(\nu-2)}$ has mean 0 and variance 1 with pdf $\tilde{f}_\nu(z) = \sqrt{\nu/(\nu-2)} f_\nu(z\sqrt{\nu/(\nu-2)})$. With the method illustrated before, it can now be shown that this transformation results in exactly four CPs between φ and $\tilde{f}_\nu$ (or $\tilde{f}_\nu$ and $\tilde{f}_\mu$, $\nu \neq \mu$).

For $\nu \to \infty$, the ACPs between φ and $\tilde{f}_\nu$ are now given by $\pm\sqrt{3 \pm \sqrt{6}}$ which already appeared in connection with Theorem 2. We give a brief outline for a proof. First, for $2 < \nu < \mu \leq \infty$, the likelihood ratio $\tilde{f}_\nu(z)/\tilde{f}_\mu(z)$ is strictly decreasing on $(-\infty, -\sqrt{3}]$ and $[0, \sqrt{3}]$ and strictly increasing on $[-\sqrt{3}, 0]$ and $[\sqrt{3}, \infty)$, see Figure 5 for an illustration. Moreover, in contrast to $f_\nu(0)$, $\tilde{f}_\nu(0)$ is strictly decreasing in $\nu \in (2, \infty)$. To see this, write

$$\tilde{f}_\nu(0) = \sqrt{\frac{\nu}{\nu-2}} \frac{\Gamma\left(\nu/2 + 1/2\right)}{\Gamma\left(\nu/2\right)} \frac{1}{\sqrt{\nu\pi}} = \frac{\Gamma(z+a)}{\Gamma(z)z^a} \frac{1}{\sqrt{2\pi}}$$

with $z = \nu/2 - 1$, $a = 3/2$, and the assertion follows by applying the properties of the function g defined in (6). The asymptotic crossing points can be obtained as in the proof of Theorem 3. We conjecture that the smaller (larger) positive crossing point between φ and $\tilde{f}_\nu$ is strictly increasing (decreasing) in $\nu > 2$.

APPENDIX A

Lemma 2. *Let $\tau > 1$ in Example 4. Then the equation $\tilde{f}_\tau(x) = \varphi(x)$ has exactly three distinct solutions $x_i(\tau) \in \mathbb{R}$, $i = 1, 2, 3$, with $-\sqrt{\tau} < x_1(\tau) < -1 < x_2(\tau) < 0$ and $1 < x_3(\tau)$.*

Proof: Consider the likelihood ratio defined by $\lambda_\tau(x) = \tilde{f}_\tau(x)/\varphi(x)$ on $\mathbb{R}$. It is easily seen that $\lambda_\tau(x) = 0$ for $x \le -\sqrt{\tau}$, $\lambda_\tau(x)$ is strictly increasing on $[-\sqrt{\tau}, -1]$, strictly decreasing on $[-1, 1]$ and strictly increasing on $[1, \infty)$ with $\lim_{x\to\infty} \lambda_\tau(x) = \infty$. Consequently, it remains to show

$$\lambda_\tau(-1) > 1, \tag{7}$$

$$\lambda_\tau(0) < 1, \tag{8}$$

$$\lambda_\tau(1) < 1. \tag{9}$$

Therefore, we study the behavior of the log-likelihood ratio $g_\tau(x) = \ln(\lambda_\tau(x))$ for $x > -\sqrt{\tau}$ and note that

$$g_\tau(x) = \ln(\tau)/2 + (\tau-1)\ln(\tau+\sqrt{\tau}\,x) - (\tau+\sqrt{\tau}\,x) - \ln(\Gamma(\tau)) + \ln(2\pi)/2 + x^2/2.$$

A straightforward calculation yields

$$\frac{d^2}{d\tau^2} g_\tau(0) = \frac{1}{\tau} + \frac{1}{2\tau^2} - \Psi'(\tau). \tag{10}$$

Note that $\Psi'(\tau) = \sum_{j=0}^{\infty} (\tau + j)^{-2}$. In the following we make use of the inequalities

$$\frac{1}{\tau} + \frac{1}{2\tau^2} < \sum_{j=0}^{\infty} \frac{1}{(\tau+j)^2} < \frac{1}{\tau - 1/2} \tag{11}$$

which apply for all $\tau > 1$. These bounds can be derived by applying the integral criterion for convergent series. From the left hand side inequality in (11) we get

$$\frac{d^2}{d\tau^2} g_\tau(0) < 0,$$

i.e., $g_\tau(0)$ is strictly concave in $\tau > 1$. Together with $g_1(0) = \ln(2\pi)/2 - 1 < 0$, $\lim_{\tau\to\infty} g_\tau(0) = 0$ and the monotonicity behavior of λ_τ stated before, we have proven (8) and (9). In order to show (7), observe that

$$\frac{d^2}{d\tau^2}\left(g_\tau(-1) - g_\tau(0)\right) = \frac{1}{2\tau^2(\sqrt{\tau}-1)}.$$

Plugging in (10) and the right hand side of (11), we obtain for $\tau > 1$ that

$$\frac{d^2}{d\tau^2} g_\tau(-1) = \frac{1}{\tau} + \frac{1}{2\tau^2} - \Psi'(\tau) + \frac{1}{2\tau^2(\sqrt{\tau}-1)} > \frac{\sqrt{\tau}(2\tau-1)+\tau}{2\tau^2(\tau-1)(2\tau-1)} > 0.$$

Hence, (7) follows.

Theorem 4. *The z_ν's defined in Theorem 3 are strictly decreasing in $\nu \in (0,\infty)$ and, for every fixed $\mu > 0$, the $z_{\nu,\mu}$'s defined in Lemma 1 are strictly decreasing in $\nu \in (0,\mu)$.*

Proof: The proof given here relies on studying the analytic properties of the log-likelihood ratio $\tilde{\lambda}_\nu(z) = \ln(f_\nu(z)/\varphi(z))$. Obviously, the monotonicity of the $z_{\nu,\mu}$'s can equivalently be expressed by the monotonicity of the roots of $\tilde{\lambda}_\nu$. For a fixed $\mu > 0$ we therefore define $F_\mu(\nu, z) = \tilde{\lambda}_\nu(z) - \tilde{\lambda}_\mu(z)$ and $z_\mu(\nu) = z_{\nu,\mu}$. Note that the function z_μ is implicitly defined by $F_\mu(\nu, z_\mu(\nu)) = 0$. In order to show the monotonicity of the function $z_\mu(\nu)$ we apply the Implicit Function Theorem, which yields

$$\frac{dz_\mu(\nu)}{d\nu} = -\frac{\partial F_\mu}{\partial z}(\nu, z_\mu(\nu))^{-1}\frac{\partial F_\mu}{\partial \nu}(\nu, z_\mu(\nu)).$$

Since the function $\partial F_\mu/\partial z$ is strictly positive for all $z > 1$ and $\nu \in (0,\mu)$, the main task of our proof consists of studying the function $\partial F_\mu/\partial\nu$ with respect to ν. For $x > 1$ we define $g(\nu, x) = 2F_\mu(\nu, \sqrt{x})$ and show (see below)

$$\nu(\nu+x)\frac{\partial}{\partial\nu}g(\nu,x) \text{ is strictly decreasing in } \nu \in (0,\infty), \tag{12}$$

$$\lim_{\nu\to 0}\left(\nu(\nu+x)\frac{\partial}{\partial\nu}g(\nu,x)\right) = 2x, \tag{13}$$

$$\lim_{\nu\to\infty}\left(\nu(\nu+x)\frac{\partial}{\partial\nu}g(\nu,x)\right) = -\frac{1}{2}(x^2-2x-1). \tag{14}$$

From the definition of F_μ we get $\lim_{\nu\to 0} F_\mu(\nu,z) = -\infty$ and $\lim_{\nu\to\mu} F_\mu(\nu,z) = 0$ for each $z > 1$. From properties (12)-(14) we see that the equation

$F_\mu(\nu, z) = 0$ has no solution if $\nu \in (0, \mu)$ and $1 < z \leq \sqrt{1+\sqrt{2}}$, thus $z_\mu(\nu) > \sqrt{1+\sqrt{2}}$. Since for fixed $0 < \nu < \mu$ we have $F_\mu(\nu, z_{\nu,\mu}) = 0$ and since there exists a unique $\nu_0 \in (0, \mu)$ with $(\partial F_\mu/\partial\nu)(\nu_0, z_\mu(\nu)) = 0$, we conclude that $\nu_0 \in (\nu, \mu)$, hence $(\partial F_\mu/\partial\nu)(\nu, z_\mu(\nu)) > 0$.

It remains to show the properties (12), (13) and (14). Let

$$\nu(\nu+x)\frac{\partial}{\partial\nu}g(\nu,x) \quad = \quad a_x(\nu) + b_x(\nu) - x(x-1),$$

where $a_x(\nu) = \nu(\nu+x)C(\nu)$ with $C(\nu) = \Psi(\nu/2+1/2) - \Psi(\nu/2) - 1/\nu$ and $b_x(\nu) = (\nu+x)(x - \nu\ln(1+x/\nu))$. First note that

$$\begin{aligned}
\frac{d}{d\nu}b_x(\nu) &= 2x - (2\nu+x)\ln\left(1+\frac{x}{\nu}\right), \\
\frac{d^2}{d\nu^2}b_x(\nu) &= -2\ln\left(1+\frac{x}{\nu}\right) + \frac{(2\nu+x)x}{\nu(\nu+x)}, \\
\frac{d^3}{d\nu^3}b_x(\nu) &= -\frac{x^3}{\nu^2\,(\nu+x)^2}.
\end{aligned}$$

Obviously, $(d^3/d\nu^3)b_x(\nu) < 0$ for all $\nu \in (0,\infty)$, hence $(d^2/d\nu^2)b_x$ is strictly decreasing. Since $\lim_{\nu\to\infty}(d^2/d\nu^2)b_x(\nu) = 0$ it follows that $(d^2/d\nu^2)b_x$ is strictly positive on $(0,\infty)$, which entails that b_x is strictly convex on $(0,\infty)$ for any fixed $x > 1$. Together with

$$\lim_{\nu\to 0} b_x(\nu) = x^2 \quad\text{and}\quad \lim_{\nu\to\infty} b_x(\nu) = x^2/2 \tag{15}$$

we furthermore get that $b_x(\nu)$ is strictly decreasing in $\nu \in (0,\infty)$ for any fixed $x > 1$. The first part of (15) is immediately proved by noting that L'Hospital's rule yields $\lim_{\nu\to 0}(\nu\ln(1+x/\nu)) = 0$ while the second part follows by inserting $\ln(1+x/\nu) = -x/\nu + x^2/(2\nu^2) + O(\nu^{-3})$ in $b_x(\nu)$. Moreover, for $a_x(\nu)$ we get

$$\lim_{\nu\to 0} a_x(\nu) = x \quad\text{and}\quad \lim_{\nu\to\infty} a_x(\nu) = 1/2. \tag{16}$$

In order to prove (16), we use the series representation $\Psi(z+1) = -\gamma + \sum_{n=1}^{\infty} z/(n^2+nz)$. Some algebraic manipulations result in

$$a_x(\nu) \quad = \quad (x+\nu)\left(\frac{1-\nu}{\nu+1} + \nu\sum_{n=2}^{\infty}\frac{2}{(2n+\nu-1)(2n+\nu-2)}\right)$$

which immediately yields the first part of (16). For the second part we furthermore utilize the representation $\nu^{-1} = \sum_{k=0}^{\infty}(\nu+k)^{-1}(\nu+k+1)^{-1}$

and obtain with $z = \nu/2$ in a few steps

$$\begin{aligned}\lim_{\nu\to\infty} a_x(\nu) &= \lim_{z\to\infty} \sum_{k=0}^{\infty} \frac{z(z+x/2)}{(z+k)(z+k+1/2)(z+k+1)} \\ &= 1/2.\end{aligned}$$

Now, combining (15) and (16) yields (13) and (14).

The remaining property (12) will be deduced by showing that a_x is strictly decreasing in $\nu \in (0,\infty)$ for any fixed $x > 1$. For a fixed constant $K \in \mathbb{R}$, consider the function $a_x(\nu) - K$. Via Laplace transformation we obtain

$$\begin{aligned}a_x(\nu) - K &= \int_0^{\infty} \exp(-\nu t)\nu(\nu+x)\left[\frac{1-\exp(-t)}{1+\exp(-t)} - (1-\exp(-xt))\frac{K}{x}\right] dt \\ &= \int_{-\infty}^{0} \exp(\nu t)\nu(\nu+x) f(t|K,x) dt.\end{aligned}$$

with $f(t|K,x) = (1-\exp(t))/(1+\exp(t)) - (1-\exp(xt))K/x$. Note that the function $\exp(\nu t)$ is strictly totally positive of order ∞ ($\mathrm{STP}_\infty(\nu,t)$) and that $\nu(\nu+x) > 0$ for all $\nu > 0$, $x > 1$. Expressing the function f as

$$f(t|K,x) = \frac{1-\exp(xt)}{1+\exp(t)}\left[\frac{1-\exp(t)}{1-\exp(xt)} - (1+\exp(t))\frac{K}{x}\right],$$

it is easily seen that $f(\cdot|K,x)$ has at most one sign change on $(-\infty,0)$ for each $K \in \mathbb{R}$ and $x > 1$. As a consequence, the theory of *Variation Diminishing Transformations* (see Brown et al. (1981)) yields that $a_x(\nu) - K$ has at most one sign change (at an isolated zero) for each $K \in \mathbb{R}$. Together with the limiting behavior of a_x for $\nu = 0, \infty$ it follows that a_x is strictly decreasing in $\nu \in (0,\infty)$ for any fixed $x > 1$. These considerations complete the proof.

APPENDIX B

A more formal illustration of the importance of CPs between densities appears in terms of the maximum difference between two probability measures P_1 and P_2, known as *total variation distance*. This distance measure can be defined by

$$d_{\mathrm{TV}}(P_1,P_2) = \sup_A |P_1(A) - P_2(A)|,$$

where the supremum is taken over all measurable sets A. Now suppose that P_i has a pdf with respect to the Lebesgue measure λ on $\mathbb{R}$, $i = 1,2$. Then the total variation distance between P_1 and P_2 can be calculated by

$$\begin{aligned} d_{\mathrm{TV}}(P_1, P_2) &= \frac{1}{2}\int |g_1(z) - g_2(z)| d\lambda(z) \\ &= \int_B (g_1(z) - g_2(z)) d\lambda(z), \end{aligned}$$

where $B = B(g_1, g_2) = \{z \in \mathbb{R} : g_1(z) > g_2(z)\}$. Typically, the set of crossing points between g_1 and g_2, that is, $\{z \in \mathbb{R} : g_1(z) = g_2(z)\}$, can be used to describe the set B.

For example, in order to compute the total variation distance (TVD) between normal and t_ν-distribution, let Φ denote the cumulative distribution function (cdf) of the standard normal distribution and let F_ν denote the cdf of the t_ν-distribution. Denote the TVD between normal and t_ν by $d_{\mathrm{TV}}(\Phi, F_\nu)$, Then we obtain from Theorem 4 that

$$\begin{aligned} d_{\mathrm{TV}}(\Phi, f_\nu) &= 2[F_\nu(-z_\nu) - \Phi(-z_\nu)] \\ &> 2[F_\nu(-(1 + 2^{1/2})^{1/2}) - \Phi(-(1 + 2^{1/2})^{1/2})]. \end{aligned}$$

For large values of ν, the bound on the right hand side yields a good approximation for $d_{\mathrm{TV}}(\Phi, F_\nu)$. The maximum difference in one tail is $d_{\mathrm{TV}}(\Phi, F_\nu)/2$.

REFERENCES

Brown, L. D., Johnstone, I. M. and MacGibbon, K. B. (1981), "Variation diminishing transformations: A direct approach to total positivity and its statistical applications", *J. Amer. Statist. Assoc.* 76, 824-832.

Finner, H., Roters, M., Dickhaus, T. (2007), "Characterizing Density Crossing Points", *Am. Stat.* 61, 28-33.

Fisher, R. A. (1925), "Expansion of "Student's" integral in powers of n^{-1}", *Metron*, 5, 109-120.

Gnedenko, B. V. (1948), "O lokal'noĭ predel'noĭ teoreme teorii veroyatnosteĭ", *Uspehi matem*, 3, No. 3, 187-194.

Lehmann, E. L., Romano, J. P. (2005), *Testing statistical hypotheses. 3rd ed.*, Springer, New York.

Petrov, V. V. (1975), *Sums of independent random variables*, Translated from the Russian by A. A. Brown. Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 82, Springer, New York.