Otto J.W.F. Kardaun; Willem Schaafsma

Exactly optimal q-values for Student problems

IPP 5/137
August, 2015
Abstract

Classical-statistical $p$–values are useful characteristics that express the degree of statistical significance of deviations from the null-hypothesis, small values of $p$ corresponding to significance. They are not necessarily appropriate, however, if one needs an assessment of the probability that $H_0$ is true. Neo-Bayesians require that such assessments should have the form of a posterior probability. The specification of a prior distribution being controversial, we use a different approach where the assessment is regarded as an estimate of the truth value of $H_0$. Such assessments are called $q$–values if the underlying estimator is based on quadratic loss. After the introduction, we describe a crucial lemma in Section 2. This leads to easily applicable ‘optimal’ procedures for assigning $q$–values in problems governed by non-central $t$–distributions. A comparative analysis of these procedures (also with the classical $p$–value) is made. Conclusions are formulated in Section 5. The Appendix contains some additional results of specific interest.

1 Introduction

During the process of evaluating statistical data it may happen that, after data reduction and some mathematical idealisation, one wants to discuss the question whether or not some population mean $\mu$ or –more generally– some regression coefficient $\beta$ (e.g. an element of a vector of regression coefficients arising from a multiple regression problem) is practically equal to some predetermined value. Focussing on a regression coefficient, we note that standard regression assumptions imply that the least-squares estimator $\hat{\beta}$ of $\beta$ has a normal distribution with expectation $\beta$ and variance $c\sigma^2$ where $c$ is determined by the (fixed) values of the explanatory variables, see e.g. [19], and $\sigma^2$ allows an unbiased estimator $\hat{\sigma}^2$ such that $\hat{\beta}$ and $\hat{\sigma}^2$ are independent and $\hat{\sigma}^2$ is distributed as $\sigma^2 \chi^2_\nu$ where $\nu$ denotes the number of degrees of freedom (usually $\nu = n - p$, where $n$ equals the number of observations and $p$ the number of fitted parameters). Fisher proved in this situation that $(\hat{\beta} - \beta) / \hat{\sigma} \sqrt{c}$ (with $\hat{\sigma} = \sqrt{\hat{\sigma}^2}$) has a central Student $t_\nu$-distribution [25], the density of which may be written as

$$f_\nu(x) = \frac{d_\nu}{\sqrt{2\pi}} \left( 1 + \frac{x^2}{\nu} \right)^{-\frac{1}{2}(\nu+1)}$$

where the normalisation constant $d_\nu = \frac{\Gamma\left(\frac{1}{2}(\nu+1)\right)}{\Gamma\left(\frac{1}{2}\nu\right)\sqrt{\frac{\pi}{2}\nu}}$ is an increasing function of $\nu$ which tends to 1 for $\nu \rightarrow \infty$ (while $d_1 = \sqrt{2/\pi} \simeq 0.8$), expressing that $f_\nu \rightarrow \varphi$.
(pointwise), where $\varphi$ is the density of $N(0, 1)$. (See [32] for a historical account.)

The test statistic

$$X = \frac{(\hat{\beta} - \beta_0)}{\hat{\sigma}} \sqrt{c} \tag{2}$$

is used to discuss truth or falsity of $H_0$: $\beta = \beta_0$ (with $\beta_0$ some predetermined value). Note that $X$ has a non-central $t_{\nu, \delta}$ distribution, see [10, Chap. 31], its density being denoted by $f_{\nu, \delta}$, where $\delta = (\beta - \beta_0)/\sigma \sqrt{c}$. If $H_0$: $\delta = 0$ is true, then $X$ has density $f_{\nu, 0} = f_{\nu}$.

It is good statistical practice to reject $H_0$: $\delta = 0$ at a (predetermined) nominal level of significance $\alpha$ (e.g. 5%) if and only if the outcome $x$ of $X$ satisfies $|x| \geq t_{\nu, 1/2, \alpha}$, where $t_{\nu, 1/2, \alpha}$ is such that the 'probability of an error of the first kind' is equal to $\alpha$, or, equivalently, if and only if the two-sided $p$–value (or significance probability)

$$\alpha^{(p)}(x) = P(|X_0| \geq |x|) \tag{3}$$

is less than $\alpha$. Being influenced by the decision-theoretic chapters in [6], we use $X_\theta$ to denote any random variable with density $f_{\nu, \theta}$, and $\theta$ as a general notation for the non-centrality parameter. A basic rule in statistical inference is that statistical uncertainties should be expressed when they are not ignorable. If $|x| \geq t_{\nu, 1, \alpha}$ then we should worry about the possibility that an error of the first kind has been committed, i.e. that $H_0$: $\delta = 0$ is actually true while $H_0$ is rejected. If, on the other hand, $|x| < t_{\nu, 1, \alpha}$ then we refrain from rejecting $H_0$, but have to worry about the possibility that $H_0$ is actually false. This leads to the following

**Problem.** Given the outcome $x$ of a test statistic $X$ with density $f_{\nu, \delta}$ where $\nu \in \{1, 2, \ldots, \infty\}$ is known\(^1\) and $\delta$ is unknown. Required a number $a = \alpha(x) \in [0, 1]$ specifying the ‘probability’ of $H_0$: $\delta = 0$ being true.

Fisher taught that the two-sided $p$–value $\alpha^{(p)}(x)$ is satisfactory in this respect. He noted that $\alpha^{(p)}(X_0)$ is uniformly distributed on $[0, 1]$. Note that $\alpha^{(p)}(x)$ is the smallest value of the nominal level of significance for which $H_0$ is rejected. This suggests that, for small values of $\alpha^{(p)}(x)$, a downward bias is involved. On the other hand, for $x$ close to 0, $\alpha^{(p)}(x)$ is close to 1 while such outcomes are not in conflict with values of $\theta$ that are close to zero. Hence, in such situation an upward bias is suspected to hold. In conclusion, it is questionable to use the two-sided $p$–value (beyond its proper and formal definition) as a measure of the ‘probability’ that $H_0$ is true. Several Bayesian statisticians, see e.g. [3], emphasised the idea that $p$–values should be replaced by posterior probabilities. These depend, of course, on the specification of a prior distribution, the prior probability of $H_0$ in particular.

To clarify the issue, in Ref. [21] the idea was put forward that the number $\alpha(x)$ required should be regarded as an estimate of the truth value $1_{\{0\}}(\delta)$ of the null-

\(^1\)when $\nu = \infty$ we have $f_{\infty, \delta}(t) = \varphi(t - \delta)$ where $\varphi$ is the density of $N(0, 1)$.
hypothesis and that this estimator should have optimum properties with respect to the *quadratic loss function* \( L(\theta, a) = (a - 1_{(0)})^2 \) in which case the estimate \( \alpha(x) \), as well as the underlying estimator \( \alpha \), is called a *q*-value. For further motivation see e.g. [7]. Note that classical estimation principles cannot be satisfied (e.g. the unbiasedness requirement) or provide useless results (e.g. the maximum-likelihood principle and Wald’s minimax risk principle). What we can do, however, is to study the expectation \( E\alpha(X_0) \) and the risk \( R(\theta, \alpha) = E L(\theta, \alpha(X_\theta)) \) of the estimator \( \alpha \), the values at the origin in particular, together with the integrated risk \( I(\alpha) = \int_{-\infty}^{+\infty} R(\theta, \alpha) d\theta \). Note that \( R(\theta, \alpha) \) is continuous (in 0) if and only if \( E\alpha(X_0) = \frac{1}{2} \).

The entire risk functions of the *p*-value \( \alpha^{(p)} \) for \( \nu = 1(1)6, 10, 20, 30 \) are given in the Appendix (see Fig. 9). Their continuity follows from the fact that \( \alpha^{(p)}(X_0) \) is uniformly distributed on (0,1). Some essential risk characteristics of \( \alpha^{(p)} \) are given in Table 1. We shall construct easily applicable exactly optimal *q*-values \( \alpha^{(0)}, \ldots, \alpha^{(5)} \), the first two of which having a continuous risk function, just like \( \alpha^{(p)} \). In [21], only \( \alpha^{(0)} \) was explicitly determined in case \( \nu = \infty \), whereas \( \alpha^{(1)} \) was briefly discussed. For a long time, generalisation to \( \nu < \infty \) was deemed to be too complicated. This changed when we discovered the lemma of Section 2.

## 2 Integrated mean squared error

For \( \nu \in \{1, 2, \ldots, \infty\} \), the integrated risk \( I(\alpha) = \int_{-\infty}^{+\infty} R(\theta, \alpha) d\theta \) is equal to

\[
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\alpha(x))^2 f_{\nu, \theta}(x) dx d\theta = \int_{-\infty}^{\infty} (\alpha(x))^2 h_{\nu}(x) dx
\]

with

\[
h_{\nu}(x) = \int_{-\infty}^{\infty} f_{\nu, \theta}(x) d\theta
\]

Interestingly, \( h_{\nu}(x) \) does not depend on \( x \), which is derived in the following

**Lemma.** The integrated density \( h_{\nu}(x) \) satisfies

\[
h_{\nu}(x) = d_{\nu}
\]

where \( d_{\nu} \) is the constant mentioned in the introduction.

**Proof.** Let us denote by \( X_{\nu, \theta} \) a random variable with density \( f_{\nu, \theta} \). It follows from the definition of the non-central \( t \) distribution that

\[
P(X_{\nu, \theta} \leq x) = P(N(\theta, 1) \leq \frac{x}{\sqrt{d}} \sqrt{X_{\nu}^2})
\]
By writing –with some abuse of notation– Eq. (7) as a double integral of the joint probability density of $X \sim N(0, 1)$ and $Y \sim \sqrt{\chi^2_\nu}$ (see e.g. [1, 9]), and differentiating this integral with respect to $x$, one obtains

$$f_{\nu,\theta}(x) = C_{\nu}^{-1} \int_0^\infty y^\nu e^{-\frac{1}{2}y^2} e^{-\frac{1}{2} \left( \frac{x^2}{\nu} - \theta \right)} dy$$

(8)

where the normalisation constant $C_{\nu} = \Gamma\left(\frac{1}{2}\nu\right)2^{\frac{1}{2}\nu-1}\sqrt{2\pi\nu}$ is chosen such that $\int_{-\infty}^{+\infty} f_{\nu,\theta}(x)dx = 1$. The proof is completed by integrating Eq. (8) with respect to $\theta$.

### 3 Comparing the $p$-value with the $q$-value $\alpha^{(0)}$

The two-sided $p$–value (see Section 1) is such that $\alpha^{(p)}(X_0)$ is uniformly distributed on $[0, 1]$. Hence, $E\alpha^{(p)}(X_0) = \frac{1}{2}$ and $R(0, \alpha^{(p)}) = R(0^\pm, \alpha^{(p)}) = 1/3$. Some essential summary characteristics of $\alpha^{(p)}(x)$ are given in Table 1. The other risk characteristics we consider in this paper, $R(0, \alpha)$ and $\max\{R(0, \alpha), R(0^\pm, \alpha)\}$ –the latter expression being equal to $M(\alpha) = \sup_{\theta} R(\theta, \alpha)$– are determined by $\psi$ and $R(0, \alpha)$. In Fig. 9, the entire risk function (based on quadratic loss) of the two-sided $p$–value, which corresponds to the procedure $\alpha^{(p)}(x) = \min\{2F_{\nu,\theta}(x), 2(1 - 2F_{\nu,\theta}(x))\}$, with $F_{\nu,\theta}(x) = \int_{-\infty}^x f_{\nu,\theta}(u)du$, has been plotted for various degrees of freedom $\nu$ as a function of the non-centrality parameter $\theta$. For $\nu < \infty$, a numerical evaluation\(^2\) was carried out to calculate the integrated risks in Table 1 and to generate the graphs in Fig. 9.

If one requires that $\alpha$ is both symmetric (i.e., $\alpha(x) = \alpha(-x)$) and ‘strongly similar’ (i.e., $\alpha(X_0) \sim U(0, 1)$) then $\alpha^{(p)}$ is, for a large class of loss functions, optimal in the sense that $R(\theta, \alpha^{(p)}) \leq R(\theta, \alpha)$ holds for all $\theta \in \mathbb{R}$.

\(^2\) using QUADPACK routines [17] as implemented in IMSL [26]

<table>
<thead>
<tr>
<th>$\nu$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>$\infty$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E\alpha^{(p)}(X_0)$</td>
<td>.5</td>
<td>.5</td>
<td>.5</td>
<td>.5</td>
<td>.5</td>
<td>.5</td>
<td>.5</td>
<td>.5</td>
<td>.5</td>
<td>.5</td>
</tr>
<tr>
<td>$R(0, \alpha^{(p)})$</td>
<td>1/3</td>
<td>1/3</td>
<td>1/3</td>
<td>1/3</td>
<td>1/3</td>
<td>1/3</td>
<td>1/3</td>
<td>1/3</td>
<td>1/3</td>
<td>1/3</td>
</tr>
<tr>
<td>$I$</td>
<td>1.41</td>
<td>1.076</td>
<td>1.016</td>
<td>0.992</td>
<td>0.978</td>
<td>0.970</td>
<td>0.955</td>
<td>0.944</td>
<td>0.941</td>
<td>0.935</td>
</tr>
</tbody>
</table>

Table 1:
Risk characteristics of the procedure $\alpha^{(p)} \equiv \alpha^{(p)}_\nu$ for various degrees of freedom $\nu$. The quantities $E(\alpha^{(p)}(X_0)) = \psi$, $R(0, \alpha^{(p)})$ and the integrated risk $I = d_\nu||\alpha^{(p)}||^2_2$ are described in the main text. The last column corresponds to the normal limit.
Table 2:
Risk characteristics for the procedures \( \alpha^{(0)} \equiv \alpha_{\nu}^{(0)} \) that have a continuous risk function, i.e., for which \( R(0) = R(0, \alpha^{(0)}) \), and minimise integrated quadratic loss (see text). The last column gives, for \( \nu = \infty \), the rounded exact values \( R(0, \alpha^{(0)}) = \sqrt{\frac{3}{2}} \) and \( I = \sqrt{\frac{\pi}{2}} \).

<table>
<thead>
<tr>
<th>( \nu )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>( \infty )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d_\nu )</td>
<td>.798</td>
<td>.886</td>
<td>.921</td>
<td>.940</td>
<td>.952</td>
<td>.959</td>
<td>.975</td>
<td>.988</td>
<td>.992</td>
<td>1</td>
</tr>
<tr>
<td>( E\alpha^{(0)}(X_0) )</td>
<td>.500</td>
<td>.500</td>
<td>.500</td>
<td>.500</td>
<td>.500</td>
<td>.500</td>
<td>.500</td>
<td>.500</td>
<td>.500</td>
<td>.500</td>
</tr>
<tr>
<td>( R(0, \alpha^{(0)}) )</td>
<td>.375</td>
<td>.329</td>
<td>.315</td>
<td>.308</td>
<td>.304</td>
<td>.301</td>
<td>.296</td>
<td>.292</td>
<td>.291</td>
<td>.2887</td>
</tr>
<tr>
<td>( I )</td>
<td>1.253</td>
<td>1.064</td>
<td>1.003</td>
<td>.973</td>
<td>.955</td>
<td>.943</td>
<td>.920</td>
<td>.903</td>
<td>.897</td>
<td>.886</td>
</tr>
</tbody>
</table>

In our opinion, it is too dogmatic to require strong similarity. It is much more reasonable to require ‘weak similarity’ (i.e., \( E\alpha(X_0) = \frac{1}{2} \)) because this is satisfied if and only if \( R(\theta, \alpha) \) is continuous for squared error loss. The lemma of Section 2 enables us to derive the estimator \( \alpha^{(0)} \) of \( \frac{1}{2} \{ \delta \} \) which minimizes \( I(\alpha) \) under the restriction \( E\alpha(X_0) = \frac{1}{2} \) (see Theorem 1). As \( \alpha^{(0)}(x) \equiv \alpha_{\nu}^{(0)}(x) = cf_\nu(x) \), this procedure is easier to apply than \( \alpha^{(p)} \). Both \( \alpha^{(p)} \) and \( \alpha^{(0)} \) have a continuous mean squared error. It follows from Theorem 1 that \( I(\alpha^{(0)}) < I(\alpha^{(p)}) \). From Table 2 one can see that \( R(\theta, \alpha^{(0)}) < R(0, \alpha^{(0)}) = 1/3 \) holds for \( \nu = 2, \ldots, \infty \). It is noted that, with an absolute error less than about 0.5%, \( (R(0, \alpha^{(0)}))^{-1} \) can be parametrically approximated by \( 3.47 - 0.917/\nu + 0.115/\nu^2 \), something we shall use for the evaluation of procedure \( \alpha^{(5)} \) later in this paper.

Using \( \psi = E\alpha(X_0) \) as notation for the ‘size’ of \( \alpha \), we derive the following direct generalisation of the case \( \psi = \frac{1}{2} \) considered before.

**Theorem 1** For \( \psi < d_\nu d_{2\nu}/\sqrt{2} \) the integrated risk \( I(\alpha) \) assumes a minimum under the restriction \( E\alpha(X_0) = \psi \) for \( \alpha = \alpha^{(0)}_{\psi} \equiv \alpha^{(0)}_{\psi,\nu,\psi} \), where

\[
\alpha^{(0)}_{\psi}(x) = \frac{\psi}{d_\nu(d_{2\nu}/\sqrt{2})} \left( 1 + \frac{x^2}{\nu} \right)^{-\frac{1}{2}(\nu+1)} \tag{9}
\]

The risk function \( R(\theta, \alpha^{(0)}_{\psi}) = EL(\theta, \alpha^{(0)}_{\psi}(X_\theta)) \) satisfies

\[
I(\alpha^{(0)}_{\psi}) = \frac{\psi^2 \sqrt{2\pi}}{d_\nu(d_{2\nu}/\sqrt{2})} \tag{10}
\]

and

\[
R(0, \alpha^{(0)}_{\psi}) = \frac{\psi^2}{(3 + 1/\nu)d_\nu(d_{2\nu}/\sqrt{2})^2(d_{3\nu}/\sqrt{3})} \tag{11}
\]
Remarks. (a) For $\psi = \frac{1}{2}$, the risk function is continuous (as a function of $\theta$) while $R(0, \alpha_{\psi}^{(0)})$ equals the maximum risk $\sup_{\theta} R(\theta, \alpha_{\psi}^{(0)})$. Note that $R(0^{\pm}, \alpha_{\psi}^{(0)}) = R(0, \alpha_{\psi}^{(0)}) + 2\psi - 1$ and that $R(0, \alpha_{\psi}^{(0)}) = R(0^{\pm}, \alpha_{\psi}^{(0)}) = M(\alpha_{\psi}^{(0)})$. The latter feature can be rigorously derived from the fact that the risk function is the expectation value of a monotonic function of $U \sim F_{1, \delta}$, while (as a function of $u$ and $\delta$) the non-central $F$-distributions possess the monotonic likelihood ratio property, see e.g. [14]. (b) The condition of the theorem is needed to ensure $0 \leq \alpha_{\psi}^{(0)}(x) \leq 1$ ($x \in \mathbb{R}$). The upper bound for $\psi$ varies from $\frac{1}{2}$ (for $\nu = 1$) to $\frac{1}{2}\sqrt{2}$ (for $\nu = \infty$). For the situation that $\psi$ is larger than $d_{\nu}d_{2\mu}/\sqrt{2}$ a more general result has been derived by the second author (in discussion with colleagues from Groningen University) which states that, under the inequality constraint $0 \leq \alpha \leq 1$, $\alpha^{(0)}$ should be equal to $\min(1, c_{f_{\nu,0}})$ where $c$ has to be determined such that $\alpha^{(0)}$ is of size $\psi$.

Proof. The quantity $I(\alpha)$ equals $d_{\nu}||\alpha||_{2}^{2} = d_{\nu}\int_{-\infty}^{+\infty} (\alpha(x))^{2} dx$. An elementary Hilbert-space argument can be used to derive that $||\alpha||_{2}^{2}$ is minimal under the restriction $(\alpha, f_{\nu,0}) = \int_{-\infty}^{+\infty} \alpha(x)f_{\nu,0}(x) dx = \psi$ if and only if $\alpha = c_{f_{\nu,0}}$ with $c = \psi/||f_{\nu,0}||_{2}^{2}$, where $||f_{\nu,0}||_{2}^{2} = \frac{d_{\nu}^{2}}{2\nu} \int_{-\infty}^{+\infty} (1 + \frac{x^{2}}{\nu})^{-(\nu+1)} dx$. An elaboration (see, e.g., [31, 33] for basic properties of Beta functions) using

$$\int_{-\infty}^{+\infty} (1 + \frac{x^{2}}{\nu})^{-(\nu+1)} dx = \sqrt{\nu} B(\frac{1}{2}, \mu + \frac{1}{2}) = \sqrt{\frac{\nu}{\mu}} d_{2\mu} \quad (12)$$

for $\mu = \frac{1}{2}(\nu + 1), \nu + 1$ and $\frac{3}{2}(\nu + 1)$, respectively, provides the above mentioned expressions.

4 Optimal semi-Bayesian $q$-values $\alpha^{(1)}, \ldots, \alpha^{(5)}$

The estimator $\alpha^{(0)} = \alpha^{(0)}_{\frac{1}{2}}$ is less compelling than it might appear at first sight: A procedure $\alpha$ may exist such that $R(\theta, \alpha) \leq R(\theta, \alpha^{(0)})$ holds for all $\theta$, with strict inequality for some $\theta$. This issue was brought up and investigated by J. Tolboom around 1989. It is not difficult to prove that all three risk characteristics $R(0, \alpha^{(0)}), R(0^{\pm}, \alpha^{(0)})$ and $I(\alpha^{(0)})$ can be simultaneously decreased. In fact, Tolboom gave a complete ‘inadmissibility’ proof for the case $\nu = \infty$ by presenting an estimator $\alpha$ of size less than $\frac{1}{2}$ such that $R(\theta, \alpha) < R(\theta, \alpha^{(0)})$ for all $\theta$. To remove these difficulties we elaborate on the semi-Bayesian approach initiated in [21].

Theorem 2 For each $w \geq 0$, the weighted risk $wR(0, \alpha) + I(\alpha)$ is minimal as a
function of $\alpha$ if and only if $\alpha = \alpha_{\nu,w}$ where

$$\alpha_{\nu,w}(x) = \frac{wf_{\nu,0}(x)}{wf_{\nu,0}(x) + d_{\nu}}$$  \hspace{1cm} (13)

Proof. From Fubini’s theorem it follows that

$$wR(0, \alpha) + I(\alpha) = \int_{-\infty}^{\infty} \left[ w(1 - \alpha(x))^2 f_{\nu,0}(x) + d_{\nu}(\alpha(x))^2 \right] dx$$  \hspace{1cm} (14)

The proof is completed by minimising the integrand for each $x$ separately.

Remark. There are several obvious re-expressions of Eq. (13). By writing $r_{\nu,0} = d_{\nu}^{-1}f_{\nu,0}(1 + x^2/\nu)^{-\nu/2}$, one obtains

$$\alpha_{\nu,w}(x) = \frac{wr_{\nu,0}(x)}{wr_{\nu,0}(x) + 1}$$  \hspace{1cm} (15)

where $r_{\nu,0}(x)$ stands for the ‘Bayes factor’ $f_{\nu,0}(x)/h_{\nu}(x)$. Alternatively, using $z = w/\sqrt{2\pi}$ instead of $w$, Eq. (13) can be rewritten as

$$\alpha_{\nu,w}(x) = \frac{1}{1 + z^{-1}(1 + x^2/\nu)^{\frac{1}{2}(\nu + 1)}}$$  \hspace{1cm} (16)

the variable $z$ occurring also as a standard parameter of the polylogarithmic function $Li_{\frac{1}{2}}(x)$, see Appendix B.

These estimators $\alpha_{w}$ are admissible. They depend, however, on the weight $w$. It is interesting that (for each $\nu$) the risk characteristics $R(0, \alpha_{w})$, $I(\alpha_{w})$ and $R(0^{\pm}, \alpha_{w}) = \lim_{\theta \to 0, \theta \neq 0} R(\theta, \alpha_{w}) = R(0, \alpha_{w}) + 2E_{\alpha_{w}}(X_0) - 1$ can be expressed by the function

$$\Psi_{\nu}(w) = E_{\alpha_{\nu,w}}(X_0) = \int_{-\infty}^{+\infty} \alpha_{\nu,w}(x)f_{\nu,0}(x)dx$$

(displayed in Fig. 1) and its first derivative with respect to $w$:

$$R(0^{\pm}, \alpha_{\nu,w}) = \Psi_{\nu}(w) - w(\partial/\partial w)\Psi_{\nu}(w) = \Psi_{\nu}(w)^2 + \var{\alpha_{\nu,w}(X_0)}$$  \hspace{1cm} (17)

$$R(0, \alpha_{\nu,w}) = 1 - \Psi_{\nu}(w) - w(\partial/\partial w)\Psi_{\nu}(w) = (1 - \Psi_{\nu}(w))^2 + \var{\alpha_{\nu,w}(X_0)}$$  \hspace{1cm} (18)

and

$$I(\alpha_{\nu,w}) = d_{\nu}|\alpha_{\nu,w}|^2 = w^2(\partial/\partial w)\Psi_{\nu}(w)$$  \hspace{1cm} (19)
Table 3:
Characteristics of the weakly similar rules for \( \alpha^{(1)} = \alpha_{\nu,w}^{(1)} \) based on quadratic loss, for various degrees of freedom \( \nu \).

<table>
<thead>
<tr>
<th>( \nu )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>( \infty )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( w^{(1)} )</td>
<td>7.52</td>
<td>5.43</td>
<td>4.88</td>
<td>4.62</td>
<td>4.39</td>
<td>4.21</td>
<td>4.08</td>
<td>4.04</td>
<td>3.957</td>
<td>3.957</td>
</tr>
<tr>
<td>( \text{E} \alpha_{\nu,w}^{(1)} )</td>
<td>.500</td>
<td>.500</td>
<td>.500</td>
<td>.500</td>
<td>.500</td>
<td>.500</td>
<td>.500</td>
<td>.500</td>
<td>.500</td>
<td>.500</td>
</tr>
<tr>
<td>( R(0, \alpha_{\nu,w}^{(1)}) )</td>
<td>.313</td>
<td>.290</td>
<td>.283</td>
<td>.279</td>
<td>.277</td>
<td>.273</td>
<td>.271</td>
<td>.270</td>
<td>.2685</td>
<td>.2685</td>
</tr>
<tr>
<td>I</td>
<td>1.410</td>
<td>1.139</td>
<td>1.059</td>
<td>1.022</td>
<td>.999</td>
<td>.985</td>
<td>.957</td>
<td>.936</td>
<td>.929</td>
<td>.916</td>
</tr>
</tbody>
</table>

To facilitate their usage, the functions \( \Psi_{\nu}(w) \) have been calculated numerically by using, once more, the QUADPACK routines [17], which are available, among others, in IMSL [26] and NAG [27]. The results are shown in Fig. 3 which presents \( \Psi_{\nu}(w) \) as a function of \( w \), for \( \nu = 1(1)30 \) and \( \nu = \infty \).

While this figure illustrates the shapes of the functions \( \Psi_{\nu}(w) \), one cannot accurately read-off from it the values of \( \Psi_{\nu}^{-1}(\psi) \), even in case \( \psi \) varies over the range of primary practical interest (approximately \( 1/3 < \psi < 2/3 \)). Hence, in Fig. 2 the deviations, \( \Psi_{\nu} - \Psi_1 \), from the curve \( \Psi_1 = 1 - \frac{1}{\sqrt{1+\psi^2}} \) based on the Cauchy density \( f_1 \) are presented. We now consider the important problem how to choose \( w \). There are various possibilities.

**Choice 1.** It is in line with \( \alpha^{(p)} \) and \( \alpha^{(0)} \) to choose \( w = w^{(1)} \) such that the size \( \text{E} \alpha_w(X_0) \) of \( \alpha_w \) is equal to \( \frac{1}{2} \), which implies continuity of the risk function. Note that \( \alpha^{(1)} = \alpha_{\nu,w}^{(1)} \) minimises \( M(\alpha_w) = \max \{ R(0, \alpha_w), R(0^\pm, \alpha_w) \} \) among all extended-Bayes estimators of the form \( \alpha_w(w \geq 0) \). In this sense, it is a Bayes-Wald compromise. The results of the numerical evaluation, carried out by writing a special purpose Fortran program (‘Jonc1Student.f90’) based on QUADPACK and a one-dimensional root finder, are presented in Table 3.

**Choice 2.** We are not dogmatic about the continuous-risk requirement. If a Bayesian uses a proper prior \( \tau \) with prior probability \( \frac{1}{2} \) for \( H_0 \), then the posterior estimator \( \alpha_{\tau} \) is such that \( \text{E} \alpha_{\tau}(X_0) > \frac{1}{2} \). This motivates the idea to replace \( w^{(1)} \) by some value larger than \( w^{(1)} \) such that \( R(0, \alpha^{(1)}) \) is decreased. In Table 4, one can therefore see similar results as obtained previously, but now for \( w \) satisfying \( R(0, \alpha_{\nu,w}) = 1/4 \). This is another type of Bayes–Wald compromise, as \( 1/4 \) is the minimax risk value corresponding to the degenerate rule \( \alpha \equiv \frac{1}{2} \). It makes some sense not to reduce the risk under \( H_0 \) any further, in order to limit the ensuing increase of the risk function on the alternative.

\(^3\)The graphs were made with the special purpose package KOMPLOT [13].
Table 4: Characteristics of the rules $\alpha^{(2)} = \alpha_{\nu,w}^{(2)}$, which satisfy $R(0, \alpha_w) = 1/4$.

<table>
<thead>
<tr>
<th>$\nu$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>$\infty$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w^{(2)}$</td>
<td>11.45</td>
<td>6.77</td>
<td>5.77</td>
<td>5.34</td>
<td>5.11</td>
<td>4.96</td>
<td>4.68</td>
<td>4.49</td>
<td>4.425</td>
<td>4.31</td>
</tr>
<tr>
<td>$E\alpha_{\nu,w}^{(2)}(X_0)$</td>
<td>.576</td>
<td>.545</td>
<td>.536</td>
<td>.532</td>
<td>.529</td>
<td>.527</td>
<td>.524</td>
<td>.522</td>
<td>.521</td>
<td>.52</td>
</tr>
<tr>
<td>$R(0, \alpha_{\nu,w}^{(2)})$</td>
<td>.250</td>
<td>.250</td>
<td>.250</td>
<td>.250</td>
<td>.250</td>
<td>.250</td>
<td>.250</td>
<td>.250</td>
<td>.250</td>
<td>.25</td>
</tr>
<tr>
<td>$I$</td>
<td>1.990</td>
<td>1.379</td>
<td>1.233</td>
<td>1.166</td>
<td>1.128</td>
<td>1.103</td>
<td>1.057</td>
<td>1.024</td>
<td>1.013</td>
<td>.99</td>
</tr>
</tbody>
</table>

Table 5: Characteristics of the rules $\alpha^{(3)} = \alpha_{\nu,w}^{(3)}$, which satisfy $w R(0, \alpha_w) = I(w)$.

<table>
<thead>
<tr>
<th>$\nu$</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>$\infty$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w^{(3)}$</td>
<td>10.27</td>
<td>7.52</td>
<td>6.60</td>
<td>6.14</td>
<td>5.87</td>
<td>5.38</td>
<td>5.06</td>
<td>4.965</td>
<td>4.78</td>
</tr>
<tr>
<td>$E\alpha_{\nu,w}^{(3)}(X_0)$</td>
<td>.627</td>
<td>.592</td>
<td>.577</td>
<td>.569</td>
<td>.565</td>
<td>.555</td>
<td>.549</td>
<td>.547</td>
<td>.544</td>
</tr>
<tr>
<td>$R(0, \alpha_{\nu,w}^{(3)})$</td>
<td>.186</td>
<td>.204</td>
<td>.211</td>
<td>.215</td>
<td>.218</td>
<td>.222</td>
<td>.225</td>
<td>.226</td>
<td>.228</td>
</tr>
<tr>
<td>$I$</td>
<td>1.913</td>
<td>1.535</td>
<td>1.395</td>
<td>1.322</td>
<td>1.277</td>
<td>1.196</td>
<td>1.141</td>
<td>1.124</td>
<td>1.09</td>
</tr>
</tbody>
</table>

Choice 3. To avoid the somewhat ad-hoc character of fixing the value of $R(0, \alpha_w)$, we elaborated on the ‘intrinsic’ Bayesian idea to determine $w = w^{(3)}$ such that the contributions of $w R(0, \alpha_{\nu,w})$ and $I(w)$ to the minimum of $w R(0, \alpha_{\nu,w}) + I(w)$ are equal. In Table 5, one can see the corresponding results, while Fig. (11) illustrates how the values $w$ in Table 5 (according to choice 3) were obtained as numerical solution of the non-linear equation $w R(0, \alpha_{\nu,w}) = I(\alpha_{\nu,w})$. Unfortunately, for $\nu = 1$ no solution $w^{(3)} \in (0, \infty)$ exists. The values $w^{(3)}$ are larger than those of $w^{(2)}$ and (in contrast to the compromise at choice 2) the quantities $R(0, \alpha^{(3)})$ are smaller than $1/4$.

Choice 4. In Table 6, one can see the results for rules $\alpha_w^{(4)}$ that satisfy $R(0, \alpha_{\nu,w}) = 1/3$, the rationale for this procedure being that 1/3 is the maximum risk $R(0, \alpha^{(p)})$ of the $p$-value (independent of $\nu$). The entire risk functions $R(\theta, \alpha_{\nu,w}^{(4)})$ are displayed in Fig. 5. As $R(0, \alpha^{(1)}) < 1/3$, we have $w^{(4)} < w^{(1)}$.

Choice 5. It is recalled that, by minimising $I$, the procedure $\alpha^{(0)}$ (with $\psi = \frac{1}{2}$) constitutes a first improvement over the two-sided $p$-value. Therefore, in Table 7 we display the results for $\alpha_{\nu,w}^{(5)}$ that satisfy $R(0, \alpha_{\nu,w}^{(5)}) = R(0, \alpha^{(0)})$. The risk characteristics in this table are, for $\nu > 2$, intermediate between those of the procedures $\alpha^{(3)}$ and $\alpha^{(4)}$. 

9
Table 6:
Characteristics of the rules $\alpha_{\nu,w(4)}$, which satisfy $R(0, \alpha_{\nu,w(4)}) = 1/3$.

<table>
<thead>
<tr>
<th>$\nu$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>$\infty$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w^{(4)}$</td>
<td>6.65</td>
<td>4.385</td>
<td>3.85</td>
<td>3.61</td>
<td>3.48</td>
<td>3.39</td>
<td>3.23</td>
<td>3.12</td>
<td>3.08</td>
<td>3.01</td>
</tr>
<tr>
<td>$E\alpha_{\nu,w(4)}(X_0)$</td>
<td>.476</td>
<td>.458</td>
<td>.448</td>
<td>.445</td>
<td>.4434</td>
<td>.4422</td>
<td>.4399</td>
<td>.4383</td>
<td>.4378</td>
<td>.4367</td>
</tr>
<tr>
<td>$R(0, \alpha_{\nu,w(4)})$</td>
<td>.333</td>
<td>.333</td>
<td>.333</td>
<td>.333</td>
<td>.333</td>
<td>.333</td>
<td>.333</td>
<td>.333</td>
<td>.333</td>
<td>.333</td>
</tr>
<tr>
<td>$I$</td>
<td>1.263</td>
<td>.928</td>
<td>.840</td>
<td>.7995</td>
<td>.7663</td>
<td>.7125</td>
<td>.7118</td>
<td>.7051</td>
<td>.692</td>
<td></td>
</tr>
</tbody>
</table>

Table 7:
Characteristics of the rules $\alpha_{\nu,w(5)}$ satisfying $R(0, \alpha_{\nu,w}) = R(0, \alpha^{(0)})$, where $\alpha^{(0)}$ minimises the integrated risk $I$, are shown for various degrees of freedom $\nu$.

<table>
<thead>
<tr>
<th>$\nu$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>$\infty$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E\alpha_{\nu,w(5)}(X_0)$</td>
<td>.432</td>
<td>.459</td>
<td>.467</td>
<td>.470</td>
<td>.4722</td>
<td>.4735</td>
<td>.4762</td>
<td>.4781</td>
<td>.4788</td>
<td>.4776</td>
</tr>
<tr>
<td>$R(0, \alpha_{\nu,w(5)})$</td>
<td>.375</td>
<td>.329</td>
<td>.315</td>
<td>.308</td>
<td>.304</td>
<td>.301</td>
<td>.296</td>
<td>.292</td>
<td>.2907</td>
<td>.2906</td>
</tr>
<tr>
<td>$I$</td>
<td>1.016</td>
<td>.948</td>
<td>.915</td>
<td>.8967</td>
<td>.8857</td>
<td>.8783</td>
<td>.8634</td>
<td>.8523</td>
<td>.8486</td>
<td>.8324</td>
</tr>
</tbody>
</table>

Analytic expressions for the median losses

Since the calculation of $E\alpha_w(X)$, $R(0^\pm) = E(\alpha_w(X))^2$ and $R(0) = E(1 - \alpha_w(X))^2$ requires a series of numerical integrations, it is of some interest to develop explicit formulae for the quantiles of the distributions of $\alpha_w(X)$, $\alpha_w^2(X)$ and $(1 - \alpha_w(X))^2$. The median values, even though they do not give the exact results described earlier in this paper, can be used as zeroth-order approximation to the expectation values.

**Theorem 3** The quantile probabilities $P\{\alpha_w(X)^2 \leq \beta_0^2(w)\} = P\{0 \leq \alpha_w(X) \leq \beta_0(w)\} = b_0$ with $X \sim t_{\nu,\delta}$ can be inverted into

\[
\frac{1 - \beta_0(w)}{\beta_0(w)} = \frac{1}{w/\sqrt{2\pi}} \left(1 + \frac{1}{\nu} F_{1,\nu,\delta;1-b_0} \right)^{\frac{1}{2}(\nu+1)}
\]

and, likewise, $P\{(1 - \alpha_w(X))^2 \leq \gamma_0^2(w)\} = c_0$ into

\[
\frac{\gamma_0(w)}{1 - \gamma_0(w)} = \frac{1}{w/\sqrt{2\pi}} \left(1 + \frac{1}{\nu} F_{1,\nu,\delta;1-c_0} \right)^{\frac{1}{2}(\nu+1)}
\]

Here, $F_{1,\nu,\delta;1-c}$ stands for the critical value corresponding to the right-tail probability $c$ of random variable $X^2 \sim F_{1,\nu,\delta}$. 

---

10
Proof. By direct calculation, (Hint: Since $\alpha_w(X) \sim (1 + z^{-1}(1 + \frac{X^2}{\nu})^{\frac{\nu+1}{2}})^{-1}$, with $X \sim t_{\nu, \delta}$ and $z = w/\sqrt{2\pi}$, it is useful to consider the monotonic transformation to the ‘intermediate’ quantity $Y_w(X) = \left(1 - \frac{1 - \alpha_w(X)}{\alpha_w(X)}\right)^2 = \frac{1}{z^2}(1 + \frac{X^2}{\nu})^{\nu+1}$, where $X^2 \sim F_{1, \nu, \delta}$. Since the quantile probabilities $P\{0 \leq \alpha_w(X) \leq \beta_0\} = b_0$ can be written as $P\{z^{-2}(1 + \frac{X^2}{\nu})^{\nu+1} \leq \left(\frac{1 - \beta_0}{\beta_0}\right)^2\} = P\{X^2 \leq \nu\left((\frac{z(1 - \beta_0)}{\beta_0})^{\frac{2}{\nu}} - 1\right)\} = b_0$, solving for $\beta_0(w)$ leads to the desired formula, and similarly for $\gamma_0(w)$. (Note that the right-hand sides of Eqs. (20) and (21) do have the same form.)

Given the extensive numerical availability of the critical values of the $F_{1, \nu, \delta}$ distributions in statistical software [5, 15, 26], the above expressions conveniently provide (for $\delta = 0$) the median losses $\bar{R}(0) = \text{median}(1 - \alpha_w(X))^2$, $\bar{R}(0^\pm) = \text{median}(\alpha_w(X))^2$, which can be used as rough approximations to the expected losses that have the theoretical feature of being proper, but require numerical integration. It is noted that the median curves provide exact straight lines in Figs. (7) and (8), with slopes that depend on $\nu$ according to Eqs. (20) and (21).

5 Conclusions

The problem of estimating $I_{\{0\}}(\delta)$ with respect to squared-error loss allows many different approaches. The estimators $\alpha^{(p)}$, $\alpha^{(0)}$, $\ldots$, $\alpha^{(5)}$ are all optimal in a certain sense. They provide compromises between $\alpha \equiv 0$ and $\alpha \equiv 1$. The minimax estimator $\alpha \equiv \frac{1}{2}$ is also a compromise, but not a reasonable one because $I(\alpha) = \infty$ and $R(0, \alpha) = R(0^\pm, \alpha) = \frac{1}{2}$.

From the analysis in this paper, it appears that $\alpha^{(1)}(w) = \alpha_{\nu, w(\nu)}$ constitutes a definite improvement over using the $p$-value. This procedure has a continuous risk function, and is, in contrast to $\alpha^{(0)}$, admissible in the sense of Wald. Furthermore, on the interval $\delta \in (-1.2, 1.2)$, it exhibits a smaller risk function than the $p$-value.

The other three rules that have been considered go beyond the class of procedures with a continuous risk function. Their preferential usage may depend on the concrete situation. When it is considered important to minimise the risk under the point null-hypothesis $H_0$, the rule $\alpha^{(2)}$ with $R(0) = 1/4$ is useful, while in other situations (e.g. when the equivalence of two drugs is being investigated, exact equivalence being practically unattainable), usage of $\alpha^{(4)}$ with $R(0) = 1/3$ would be more indicated. The latter rule has the same risk as the $p$-value for $\delta = 0$, but has a smaller quadratic loss over the range $\delta \in (-1.65, 1.65)$. (Moreover, in that
situation, modelling can sensibly be based on probability assignments to three hypotheses: A\(_{-1}\): \(\theta < \delta_0\), H: \(\theta \in [-\delta_0, +\delta_0]\), and A\(_{+1}\): \(\theta > \delta_0\). This approach, an extension of the present work, is considered elsewhere.) For \(\nu > 2\), rule \(\alpha^{(5)}\) is a practical compromise between \(\alpha^{(3)}\) and \(\alpha^{(4)}\).

In order to apply a further principle to order the estimators \(\alpha^{(p)}, \alpha^{(0)}, \ldots, \alpha^{(5)}\) we regard \(\alpha\) as better than \(\alpha'\) if \(I(\alpha)(M(\alpha) - \frac{1}{4}) < I(\alpha')(M(\alpha') - \frac{1}{4})\) where \(M(\alpha) = \max(R(0, \alpha), R(0 \pm, \alpha))\). Note that this criterion is independent on whether \(I\) is replaced by \(cI\), \(c > 0\). (It makes little sense to apply this ‘secondary’ criterion to all estimators \(\alpha_w\) since that would lead to degenerate solutions.) For \(\nu = 1(1)5\) and \(\nu = 10\), the relevant computations are presented in Table 8. We conclude from this table that \(\alpha^{(1)}\) is recommendable also from the viewpoint that its value \(C(\alpha)\) is lower than that of the other procedures.

6 Appendix

A. Properties of and approximations to \(d_\nu\). The quantity \(d_\nu = \sqrt{\frac{2\nu}{\pi}} B(\frac{1}{2}, \frac{1}{2}\nu) = \frac{\Gamma(\frac{1}{2}\nu + \frac{1}{2})}{\Gamma(\frac{1}{2}\nu)}\sqrt{\frac{1}{2}\nu}\) and also its inverse –for which, from the duplication formula (see [2, 31]), it can be derived that \((\frac{d_\nu}{\sqrt{2\pi}})^{-1} = 2^{\frac{\nu}{2} - 1}) \sqrt{\frac{1}{2}\nu B(\frac{1}{4}\nu, \frac{1}{4}\nu + \frac{1}{2})}\) – occurs in several situations. While simplifying expressions that contain the just mentioned Gamma function ratio, it has been extensively utilised in [10] and also in [29] (where it is called \(\beta_f\)). The first moment of \(X \sim t_{\nu, \delta}\) equals \(E X = \frac{\delta d_\nu}{1 - 1/\nu},\)

<table>
<thead>
<tr>
<th>(\nu)</th>
<th>(\alpha^{(4)})</th>
<th>(\alpha^{(5)})</th>
<th>(\alpha^{(0)})</th>
<th>(\alpha^{(p)})</th>
<th>(\alpha^{(1)})</th>
<th>(\alpha^{(2)})</th>
<th>(\alpha^{(3)})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.109</td>
<td>.128</td>
<td>.157</td>
<td>.117</td>
<td>.101</td>
<td>.262</td>
<td>.172</td>
</tr>
<tr>
<td>2</td>
<td>.077</td>
<td>.075</td>
<td>.084</td>
<td>.089</td>
<td>.046</td>
<td>.262</td>
<td>.211</td>
</tr>
<tr>
<td>3</td>
<td>.070</td>
<td>.059</td>
<td>.056</td>
<td>.084</td>
<td>.035</td>
<td>.089</td>
<td>.227</td>
</tr>
<tr>
<td>4</td>
<td>.066</td>
<td>.052</td>
<td>.065</td>
<td>.082</td>
<td>.030</td>
<td>.075</td>
<td>.160</td>
</tr>
<tr>
<td>5</td>
<td>.064</td>
<td>.048</td>
<td>.052</td>
<td>.081</td>
<td>.028</td>
<td>.064</td>
<td>.136</td>
</tr>
<tr>
<td>10</td>
<td>.059</td>
<td>.040</td>
<td>.042</td>
<td>.079</td>
<td>.022</td>
<td>.051</td>
<td>.08</td>
</tr>
</tbody>
</table>

Table 8: Entries of \(C(\alpha) = I(\alpha)(M(\alpha) - \frac{1}{4})\) according to the rules \(\alpha^{(p)}, \alpha^{(0)}, \ldots, \alpha^{(5)}\) for \(\nu = 1, \ldots, 5\) and \(\nu = 10\) degrees of freedom.
follows (see, e.g., [30]) that 

\[ d \sim \frac{1}{\sqrt{1 - d^2}} (\frac{1}{\nu(1 - d^2)} - 2), \]

respectively.

Geometrically, 

\[ \frac{d_{\nu+1}}{d_{\nu}} = \frac{V_{\nu+1}(r)}{V_{\nu}(r)} = \frac{1}{\nu + 1} V_{\nu+1}(r), \]

where \( V_{\nu}(r) \) is the volume and \( S_{\nu}(r) \) the surface area of the \( \nu \)-dimensional sphere. As is well known (see e.g. [8, 16]), the circumference, \( C \), of an ellipse with elongation \( \kappa = b/a > 1 \) and squared eccentricity \( e^2(\kappa) = 1 - \kappa^{-2} \) is expressed by a complete elliptic integral \( E(\kappa) \), the usual series expansion of which reads

\[ C = 4aE(\kappa) = 2\pi a \kappa \left( 1 - \sum_{m=1}^{\infty} \frac{(d_{2m} \sqrt{m})^2}{2m - 1} [e^2(\kappa)]^m \right), \tag{23} \]

Moreover, since \( \sqrt{\nu} d_{\nu} \) is equal to the first moment of the \( \chi_{\nu} \) distribution [9], a statistical interpretation of \( d_{\nu} \) is that, for \( X_1, X_2, \ldots, X_n \), i.i.d. \( N(\mu, \sigma^2) \), among all estimators of the type \( c \sqrt{(n - 1)^{-1} \sum_{i=1}^{n} (X_i - X)^2} \), \( c = \frac{1}{d_{n-1}} \) gives the uniformly minimum variance unbiased (UMVU) estimator for \( \sigma \), while \( c = d_{n-1} \) yields the estimator that minimises the mean squared error. In a similar vein,

\[ \frac{cX}{\sqrt{(n - 1)^{-1} \sum_{i=1}^{n} (X_i - X)^2}} \]

is an UMVU estimator for \( \frac{\mu}{\sigma} \) if \( c = d_{n-2} \sqrt{\frac{n-2}{n-1}} = \frac{1}{d_{n-1}} \frac{n-2}{n-1} \).

Of course, \( d_{\nu} \) can be calculated directly in computer programs when gamma functions are conveniently available. We reflect a little when this is not precisely the situation. For small integer \( \nu (\nu = 2m \text{ or } \nu = 2m+1) \) the standard expressions

\[ d_{2m} = \frac{\Gamma(m + \frac{1}{2})}{\sqrt{m} \sqrt{\Gamma(m + 1)}} \]

and \( d_{2m+1} = \frac{1}{\sqrt{m + 1/2} \sqrt{\Gamma(m + 1/2)}} \), with \( \Gamma(m + 1/2) = \frac{1}{2} \cdot \frac{3}{2} \cdot \frac{5}{2} \cdots \cdot \frac{2m-1}{2} \), allow for direct calculation by elementary arithmetic using a few terms only. From Binet’s approximation [31] to \( \ln \Gamma(\nu) \), one gets the simple asymptotic approximation \( d_{\nu} \approx d_{\nu,b} = e^{-1/2} (1 + 1/\nu)^{\nu/2} \), which can be improved by expressions in powers of \( \nu^{-1} \). For instance\(^5 \), \( 1 - d_{\nu}/d_{\nu,b} \approx 0.166644 e^{-1/2} (1 + 0.01208 + 0.2373 - 0.1266 + r(\nu)) \)

with \( r(\nu) = 0.01428 \times 10^{-3} (\sin(\frac{5.2\pi}{\nu}) + \cos(\frac{5.2\pi}{\nu})) \) is an estimate accurate to about 0.0001\% for \( 2 \leq \nu \leq 90 \).

Alternatively, one can bracket \( d_{\nu} \) by inequalities. From Hölder’s inequality it follows (see, e.g., [30]) that \( \frac{1}{\sqrt{1 + 1/\nu}} < d_{\nu} < 1 \). Under the restriction \( \nu \geq 1.5 \), we derived the somewhat narrower inequality \( e^{-1/2} (1 + \frac{1}{\nu + 1/4})^{\nu/2} < d_{\nu} < e^{-1/2} (1 + \frac{1}{\nu})^{\nu/2} \).

\(^4\)It is recalled that \( \chi_{\nu} \) equals the normalised Maxwell distribution for the speed of ideal gas particles in \( \nu \) dimensions.

\(^5\)as derived by regression analysis using SAS [20]
Note, from the definition of $d_\nu$, that $d_{2\nu}d_{2\nu+1} = \sqrt{2\nu + 1}$. Therefore, the multiplication factor in Eq. (9), $\sqrt{2\nu + 1}d_{2\nu+1}$, equals $\frac{\sqrt{2\nu + 1}d_{2\nu+1}}{\sqrt{d_c}}$, which (by using elementary properties) can be re-expressed as

$$B\left(\frac{1}{2}, \frac{1}{2}\nu\right)/B\left(\frac{1}{2}, \nu + \frac{1}{2}\right) = \frac{\int_0^{\pi/2} \cos^{\nu-1}(\phi) d\phi}{\int_0^{\pi/2} \cos^{2\nu}(\phi) d\phi} = \frac{\Gamma\left(\frac{1}{2}\nu\right)}{\Gamma\left(\frac{1}{2}\nu + \frac{1}{2}\right)} \frac{\Gamma(\nu + 1)}{\Gamma(\nu + \frac{1}{2})}. \tag{24}$$

B. Computation of $\Psi_\nu$ for $\nu = 1, 3, \infty$. Except for some special situations, with a very few degrees of freedom and for $\nu = \infty$, a numeric evaluation is needed to calculate $\Psi_\nu(w) = \int_{-\infty}^{+\infty} \frac{w f(x)}{w f(x) + d_{\nu}} f_{\nu}(x) dx$. Using the abbreviation $z(w) = \frac{w}{\sqrt{2\pi}}$, one can directly derive the analytic expression

$$\Psi_1(w) = 1 - \frac{1}{\sqrt{1 + z(w)}}, \tag{25}$$

and, somewhat more elaborated,

$$\Psi_3(w) = 1 - \frac{\sqrt{2}}{z(w)} \sqrt{1 + \sqrt{1 + z(w)}} \left(1 - \frac{1}{\sqrt{1 + z(w)}}\right) \tag{26}$$

such that also $R(0)$, $R(0^\pm)$ and $I$ can be directly calculated analytically for these two special cases ($\nu = 1$ and $\nu = 3$).

For $\nu \to \infty$, the function $\Psi_\infty(w) = E_{\nu, \infty}(X_0)$ coincides with $1 - E(w \varphi(x) + 1)^{-1}$ (where $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$), and can be expressed by the polylogarithmic function $\text{Li}_{\frac{1}{2}}(z)$, which is seen as follows. By substituting $\frac{1}{2}x^2 = t$, one can rewrite $w \left(1 - \Psi_\infty(w)\right)$ as

$$\sqrt{2} \int_0^\infty \frac{t^{-\frac{1}{2}}}{1 + (z(w))^{-1}} e^t dt = -\sqrt{2\pi} \text{Li}_{\frac{1}{2}}(-z(w)) \tag{27}$$

where

$$\text{Li}_{\frac{1}{2}}(z) = \frac{1}{\Gamma\left(\frac{1}{2}\right)} \int_0^\infty \frac{t^{-\frac{1}{2}}}{z^{-1} e^t - 1} dt \tag{28}$$

is the analytic continuation of $\sum_{n=1}^\infty \frac{z^n}{\sqrt{n}}$ outside $|z| < 1$.

Having been derived and utilized in statistical mechanics to describe identical particles with integer and with half integer spin (see e.g. [12, 18, 22, 23]), the functions $B\text{E}_{\frac{1}{2}}(\alpha) = \text{Li}_{\frac{1}{2}}(e^\alpha)$ and $F\text{D}_{\frac{1}{2}}(\alpha) = -\text{Li}_{\frac{1}{2}}(-e^\alpha)$ are generally known as Bose–Einstein and Fermi–Dirac integrals, respectively. A specific physical
application is the magnetic susceptibility of a free electron gas, see [24]. The origin of these special functions can be traced back to the early XVIII century and even to Bernoulli, see, e.g., [11, 16, 28].

**C. Risk functions for \( \alpha^{(p)} \), \( \alpha^{(0)} \), \( \alpha^{(3)} \) and \( \alpha^{(5)} \).** Figs. 9 and 10 show the risk functions of the procedures \( \alpha^{(p)} \) and \( \alpha^{(0)} \) corresponding to the \( p \)-value and to the procedure that minimises the integrated risk (for quadratic loss), respectively. One can see that, for \( \nu \geq 2 \) and \( \theta \in (-1.2, 1.2) \), the risk function for \( \alpha^{(0)} \) is somewhat smaller than that for \( \alpha^{(p)} \). Likewise, the risk functions of \( \alpha^{(3)} \) and \( \alpha^{(5)} \) are plotted in Figs. 12 and 13.

**D. Connection with enzyme kinetics.** As stated above, according to the authors’ knowledge, analytic expressions for \( \Psi_{\nu}(w) \) are currently not available (when \( \nu \) is not equal to 1 or 3). It is of some interest to note that \( 1/\Psi_{\nu}(w) \) is reasonably approximated by \( a(\nu) + b(\nu)/w \). This can be seen from Fig. 7. A resemblance exists with steady state solutions of particular first order differential equations. Let us look at the following example from enzymology (which has been originally considered by Michaelis and Menten around 1913, see also [4]): The reaction \( E + S \rightarrow ES \rightarrow E + S' \) with \( E \) an enzyme and \( S \) a reactant (‘substrate’), where the reaction rates depend linearly on the concentrations \([S], [E] \) and \([SE] \). The normalised reaction speed is described by \( \frac{v}{v_{\text{max}}} = \frac{[S]}{[S] + K_M} \), where \( K_M \) is the ratio between the reaction rates of dissociation and formation of the compound \( ES \). This expression is precisely converted into \( \frac{w_{\text{max}}}{w} = 1 + \frac{K_M}{[S]} \), which corresponds to our situation when \( \Psi_{\nu} \) is identified with \( v/v_{\text{max}} \) and \( w \) with \([S] \), the difference being that we take the expectation value of \( \alpha_{\nu}(X) = \frac{w}{w+d_{\nu}/p(X)} \), which leads to deviations from straight lines in Fig. 7.

Another example is a second order, irreversible reaction process with equal concentration (denoted by \( c \)) of both reactants. In this case, \( c(t) \) satisfies the differential equation \(-\frac{d}{dt}c(t) = kc^2(t) \) which, with the initial condition \( c(0) = 1 \), is explicitly solved by \( 1/c(t) = 1 + kt \). Identifying \( t \) with \( w \) and \( c(t) \) with \( 1 - \Psi_{\nu}(w) \), we have an analogous situation, since to a first order approximation \( \frac{1}{1-\Psi_{\nu}(w)} = \frac{1}{1-E\alpha_{\nu,\nu}(X)} \simeq E\frac{1}{1-\alpha_{\nu,\nu}(X)} = 1 + wE\frac{f_{\nu,\nu}(X)}{d_{\nu}} \). From Fig. 8, where \( \frac{1}{1-\Psi_{\nu}(w)} \) is plotted against \( w \), one can see that the deviations from straight lines are indeed rather weak, albeit not entirely negligible. For practical use, values \( w = \Psi_{\nu}^{-1}(\psi) \) (with, approximately, \( 0.3 < \psi < 0.7 \)) can be read off more accurately from this kind of plot than from Fig. 1. Obviously, the graphs in this plot are the same (except for a simple shift of the origin) as those of the ‘odds’ \( \frac{\Psi_{\nu}(w)}{1-\Psi_{\nu}(w)} \), and, as stated above, since they represent special functions (being polylogarithms for \( \nu \rightarrow \infty \)), elementary transformations to convert them exactly into straight lines are not generally available.
D. Ratio between $\alpha^{(1)}$ and $\alpha^{(p)}$. A direct comparison of $\alpha^{(p)}$ and $\alpha^{(1)}$ is displayed in Fig. 14, where $\frac{\alpha^{(1)}(x)}{\alpha^{(p)}(x)} - 1$ is plotted against $x$ for $\nu = 1$ (lower curve), 2, 3, 4, 5, 6, and 10 (upper curve). From this figure one can see that $\alpha^{(1)}(x) < \alpha^{(p)}(x)$ for $x \lesssim \frac{1}{2}$, while $\alpha^{(1)}(x) > \alpha^{(p)}(x)$ for $x \in (\frac{1}{2}, x_{\text{max},\nu})$, where $x_{\text{max},\nu}$ (with $x_{\text{max},1} \simeq 4$ and $x_{\text{max},2} \simeq 20$) is an increasing function of $\nu$. Moreover, one can see from the figure –for instance– that (at the conventional level of 5%) for $x \simeq 4.30$ and $\nu = 2$, the ratio $\alpha^{(1)}/\alpha^{(p)} \simeq 1.2$, while for $x \simeq 2.23$ and $\nu = 10$, the ratio $\alpha^{(1)}/\alpha^{(p)} \simeq 3.1$. 
Figure 1: Functions describing $\Psi_\nu(w) = E_{\alpha_{\nu,w}}(X)$ for $\nu = 1, \ldots, 30$ and $\nu = \infty$ degrees of freedom; the function $\alpha_{\nu,w}$ assigns the ‘probability’ $\alpha_{\nu,w}(x)$ to $H_0$ (‘central Student-t’), and $1 - \alpha_{\nu,w}(x)$ to $A$ (‘non-central Student-t’).
Figure 2: Functions describing $\Psi_\nu(w) - \Psi_1(w)$ against $\Psi_1(w) = E\alpha_{1,w}(X) = 1 - \frac{1}{\sqrt{1+w/\sqrt{2\pi}}}$ for $\nu = 1, \ldots, 30$ degrees of freedom, as well as for the normal limit ($\nu \to \infty$). The intersections of the straight lines with the curved ones indicate on the vertical axis the increment of $\Psi_\nu(w)$ with respect to $\Psi_1(w) = 1/3, 0.4, 0.5, 0.6$ and $2/3$, respectively; the corresponding values of $w$, for which $\Psi_\nu(w) = 1/3, 0.4, 0.5, 0.6$ and $2/3$, can be read from the upper horizontal axis.
Figure 3: A plot of the risk function: \( E(\alpha_{\nu,w}(X; \theta))^2 \) (\( \theta \neq 0 \)) and \( E(1 - \alpha_{\nu,w}(X; \theta))^2 \) (\( \theta = 0 \)), as a function of \( \theta \) for \( \nu = 1(1)6, 10, 20, 30 \) and \( w = w_{\nu}^{(1)} \).
Figure 4: A plot of the risk function: $E(\alpha_{\nu,w}^{(2)}(X;\theta))^2 (\theta \neq 0)$ and $E(1 - \alpha_{\nu,w}^{(2)}(X;\theta))^2 (\theta = 0)$, as a function of $\theta$ for $\nu = 1(1)6, 10, 20, 30$ and $w = w_{\nu}^{(2)}$. 

$R(0) = 1/4$ weights
Figure 5: A plot of the risk function: $E\left(\alpha_{\nu,w}(X;\theta)\right)^2 (\theta \neq 0)$ and $E\left(1 - \alpha_{\nu,w}(X;\theta)\right)^2 (\theta = 0)$, as a function of $\theta$ for $\nu = 1(1)6, 10, 20, 30$ and $w = w_{\nu}^{(4)}$. 

R(0) = 1/3 weights
Figure 6: A plot of $\max\{R(0), R(0^+)\}$ against I and the position of various rules $\alpha^{(j)}$ ($j = p, 0, 1, \ldots, 5$) for $\nu = 2$ (○) and for $\nu = \infty$ (♦). The two dots in the troughs correspond to the procedures $\alpha_{p,w(1)}$. 
Figure 7: Functions describing $1/\Psi_\nu(w)$ as a function of $1/w$ for $\nu = 1, \ldots, 30$ and $\nu = \infty$ degrees of freedom, where $\Psi_\nu(w) = E_{\alpha_\nu,w}(X)$ is plotted in Fig. (1).
Figure 8: A plot of $1/(1 - \Psi_\nu(w))$ as a function of $w$, for $\nu = 1, \ldots, 30$ and $\nu = \infty$ degrees of freedom, where $\Psi_\nu(w) = E_{\alpha_{\nu,w}}(X)$ is plotted in Fig. (1).
Figure 9: A plot of the risk function: $E(\alpha^{(p)}_{\nu,w}(X; \theta))^2$ ($\theta \neq 0$) and $E(1 - \alpha^{(p)}_{\nu,w}(X; \theta))^2$ ($\theta = 0$), corresponding to the two-sided $p$-value, as a function of $\theta$ for $\nu = 1(1)6, 10, 20, 30$ degrees of freedom.
Figure 10: A plot of the risk function: $E(\alpha^{(0)}_{\nu,w}(X; \theta))^2 (\theta \neq 0)$ and $E(1 - \alpha^{(0)}_{\nu,w}(X; \theta))^2 (\theta = 0)$, corresponding to the procedure that minimises the integrated risk, as a function of $\theta$ for $\nu = 1(1)6, 10, 20, 30$ degrees of freedom.
Figure 11: Functions describing the integrated Risk $I(w)$ (dashed lines), $wR(0, w)$ (solid lines) and their difference (solid lines) for $\nu = 1, 2, 3, \ldots$. The zero crossings of the latter solid lines yield the values $w$ displayed in Table 5, for which $wR(0, w) = I(w)$. Note that for $\nu = 1$ (Cauchy/Lorentz distribution), $I(w) > wR(0, w)$ for all $w \in [0, \infty)$. 
Figure 12: A plot of the risk function: $E(\alpha^{(3)}_{\nu,w}(X; \theta))^2 (\theta \neq 0)$ and $E(1 - \alpha^{(3)}_{\nu,w}(X; \theta))^2 (\theta = 0)$, as a function of $\theta$ for $\nu = 1(1)6, 10, 20, 30$. 
Figure 13: A plot of the risk function: $E(\alpha_{\nu,w}^{(5)}(X;\theta))^2$ ($\theta \neq 0$) and $E(1 - \alpha_{\nu,w}^{(5)}(X;\theta))^2$ ($\theta = 0$), as a function of $\theta$ for $\nu = 1(1)6, 10, 20, 30$. 
Figure 14: A plot of the ratio \( \frac{\alpha^{(3)}(x)}{\alpha^{(p)}(x)} - 1 \) as a function of \( x \) for \( \nu = 1(1)6, 10. \)
References


