2

The truth or falsity of point null hypotheses

The theory of testing statistical hypotheses originated from discussions about truth or falsity. Astronomers and physicists computed significance probabilities to assess their models. From the side of the Bayesian school the computation of these significance probabilities was critized. It is well known that a null hypothesis may have a very small P-value, and yet be assigned a considerable posterior probability. This disagreement between objectivistic and subjectivistic degrees of belief was first studied by Jeffreys [50]. Lindley was the first to call this disagreement a paradox (see, Shafer [83]), and it is now generally known as Lindley’s Paradox, though Lindley himself refers to it as Jeffreys’s Paradox. As pointed out in Zellner [97], the paradox disappears if the Neyman–Pearson tester adjusts the level of his test in such a way that it ‘balances’ the type I and II errors. Adjusting the level of an accept–reject tests does, however, not affect the P-value, and the question arises whether or not the P-value is the most appropriate degree of belief in the hypothesis. Perhaps, some transformation $Q = \psi(P)$ should be used. To investigate the discrepancy between objectivistic and subjectivistic methods, the corresponding procedures will be regarded as estimators of the indicator of the hypothesis. They will be evaluated, in line with the theory of point estimation, in terms of both risk and bias.
2.1 Introduction

The problem to be considered can be formulated as follows: suppose that the outcome $x$ of some random variable $X$ has to be used to discuss the truth or falsity of the null hypothesis $H_0 : \mathbb{L}X = P_{\theta_0}$ if it is tested against the alternative $A : \mathbb{L}X = P_{\theta}$, where $\theta \neq \theta_0$. The notation $X_{\theta}$ will be used to denote a random variable satisfying $\mathbb{L}X_{\theta} = P_{\theta}$, and the true value $t$ of the unknown parameter $\theta$ is implicitly defined by $\mathbb{L}X = \mathbb{L}X_t$. The parameter space is denoted by $\Theta$, and is supposed to be a subset of $\mathbb{R}^k$ such that $\Theta_{H_0} \cap \Theta_A = \Theta_{H_0}$, where $\Theta_{H_0} = \{ \theta_0 \}$ and $\Theta_A = \Theta \backslash \Theta_{H_0}$. Or in words, the problems that will be considered are such that the hypothesis and the alternative have a nonempty common boundary consisting of only one point, namely $\theta_0$. The truth or falsity of $H_0$ is discussed on the basis of a number $\alpha = \alpha(x) \in [0, 1]$ indicating the degree of belief in $H_0$, in the light of the data.

Let $\mathcal{X}$ denote the outcome space, and suppose that $\mathcal{X} \subset \mathbb{R}^m$. To construct such degree of belief $\alpha(x)$, the attention will be focused on constructing an estimator $\alpha : \mathcal{X} \mapsto [0, 1]$ of the indicator of the hypothesis or, more precisely, of its true value $\mathbb{I}_{\{t_0\}}(t)$. Motivated by the work of J.O. Berger, Casella, a.o., and having its roots in the theory of Kiefer about confidence intervals, this approach was initiated by Schaalma [79], adopted by, e.g. Van der Meulen [63], and also developed by Hwang et al. [48], and Salomé et al. [74]. In this form, the problem can be regarded as a problem of point estimation. Hence, a first idea would be to apply the well-studied approach towards problems of point estimation. Notice, however, that the concept of mean-unbiasedness as used in the theory of point estimation is useless, because in the present context unbiased estimators can only exist in degenerate situations. Reduction by mean-unbiasedness is used in the theory of point estimation as a guarantee for impartiality. Consequently, some further thought has to be given to the problem of finding an appropriate alternative to the concept of mean-unbiasedness.

Fisher noted that if $P_{\theta_0} \ll \lambda$, where $\lambda$ denotes Lebesgue measure, and the test statistic is chosen appropriately, then the P-value $\alpha_\lambda$ satisfies $\mathcal{L}\alpha_\lambda(X_{\theta_0}) = U(0,1)$. Assuming that the parameterization of $\mathcal{P} = \{ P_{\theta} : \theta \in \Theta \}$ is continuous, in the sense that for $P_\theta(B)$ is continuous for $B \in \sigma(\mathcal{X})$, and taking into account the fact that the hypothesis and the alternative have $\{\theta_0\}$ as common boundary, the requirement that a procedure $\alpha$ should satisfy $\mathcal{L}\alpha(X_{\theta_0}) = U(0,1)$ can be regarded as a very strong requirement of impartiality. That is, if some procedure $\alpha$ satisfies this requirement then, in a certain sense, it neither favors the hypothesis nor the alternative. This suggests that the idea of impartiality can be made operational by using the following mathematical concept of
unbiasedness. A procedure \( \alpha : \mathcal{X} \mapsto [0, 1] \) is said to be strongly unbiased if it satisfies
\[
\mathcal{L}(\alpha(X_{\theta})) = U(0, 1) \quad \text{and} \quad \mathcal{L}(\alpha(X)) \leq U(0, 1) \quad \forall \theta \in \Theta_A.
\]
The class of all strongly unbiased procedures will be denoted by \( \mathcal{D}_s \). Notice that the requirement of strong unbiasedness is very restrictive: in many situations no strongly unbiased procedure will exist, e.g., in the case that \( P_{\theta_0} \not\ll \lambda \). Moreover, it will be shown that whenever \( \mathcal{D}_s \neq \emptyset \) it often only contains one procedure that is worthwhile considering, i.e., the \( P \)-value.

A less restrictive requirement of impartiality is that of weak unbiasedness. Instead of requiring a uniform distribution on the common boundary of the hypothesis and the alternative it is required that, on average, it points as much towards the hypothesis as to the alternative. To be precise, a procedure \( \alpha : \mathcal{X} \mapsto [0, 1] \) is said to be weakly unbiased if it satisfies
\[
\mathbb{E}(\alpha(X_{\theta_0})) = \frac{1}{2} \quad \text{and} \quad \mathbb{E}(\alpha(X_{\theta})) \leq \frac{1}{2} \quad \forall \theta \in \Theta_A.
\]
Notice that strong unbiasedness implies weak unbiasedness. The class of all weakly unbiased procedures will be denoted by \( \mathcal{D}_w \). It easily seen that \( \mathcal{D}_w \) as always nonempty, i.e., it contains \( \alpha \equiv \frac{1}{2} \). In contrast to \( \mathcal{D}_s \), the class \( \mathcal{D}_w \) will never contain a uniformly best procedure.

When evaluating the estimates of \( \mathbb{1}_{\{\theta_0\}}(t) \), generated by the procedure \( \alpha \), with squared error loss, one finds that the risk function of \( \alpha \) given by
\[
R(\theta, \alpha) = \begin{cases} 
\mathbb{E}(1 \Leftrightarrow \alpha(X_{\theta}))^2 & \text{if} \quad \theta = \theta_0, \\
\mathbb{E}(\alpha(X_{\theta}))^2 & \text{if} \quad \theta \in \Theta_A.
\end{cases}
\]
If the parameterization of \( \mathcal{P} = \{P_{\theta} : \theta \in \Theta \} \) is continuous in the sense mentioned earlier in this section, then it is straightforward to infer from (2.1) that the risk function of \( \alpha \) is continuous if and only if \( \alpha \in \mathcal{D}_w \). Continuity of the risk function on the common boundary of the hypothesis and the alternative shows that the restriction to \( \mathcal{D}_w \) or \( \mathcal{D}_s \) can, indeed, be regarded as a requirement of impartiality.

2.2 Posterior probabilities

Posterior probabilities are commonly associated with the Bayesian approach. The underlying idea of this approach is that the true value \( t \) of the unknown parameter \( \theta \) itself can be regarded as the outcome of a random variable \( T \),
with values in $\Theta$. After specifying some prior probability measure $\nu$ on $\Theta$, the posterior degree of belief in $H_0$ can be computed as the posterior probability of $\Theta_{H_0} = \{\theta_0\}$, i.e.,

$$\alpha_\nu(x) = \frac{\nu(\{\theta_0\})p_{\theta_0}(x)}{\nu(\{\theta_0\})p_{\theta_0}(x) + \int_{\Theta} p_\theta(x) \nu(d\theta)},$$

where $p_\theta$ is the density of $X_\theta$, or more precisely the Radon–Nikodym derivative of $P_\theta$ w.r.t. some measure $\mu$ dominating $P = \{P_\theta : \theta \in \Theta\}$. The lack of knowledge a priori about the truth or falsity of $H_0$ is reflected in the prior which is usually chosen such that prior probability $\frac{1}{2}$ is assigned to $\{\theta_0\}$, i.e., $\nu(\{\theta_0\}) = \frac{1}{2}$, and consequently $\frac{1}{2}$ to $\Theta_A$, i.e., $\nu(\Theta_A) = \frac{1}{2}$. Such choice of the prior measure can be regarded as the classical Bayesian one. In the situation that $p_{\theta_0}(x) > 0$, for all $x \in \mathcal{X}$, such choice of prior will provide that $\alpha_\nu \notin \mathcal{D}_w$, because the convexity of $1/(1 + x)$ on $[0, \infty)$ implies that

$$\mathbb{E}_{\alpha_\nu}(X_{\theta_0}) = \mathbb{E} \frac{1}{1 + 2 \int_{\Theta_A} p_\theta(x_{\theta_0}) \nu(d\theta)} \geq \frac{1}{1 + 2 \int_{\Theta_A} \mathbb{E} p_\theta(x_{\theta_0}) \nu(d\theta)} = \frac{1}{2},$$

where equality holds only if $\alpha_\nu \equiv \frac{1}{2}$. This shows that there is a conflict between the Bayesian concept of impartiality, i.e., $\nu(\{\theta_0\}) = \frac{1}{2}$, and the Fisherian concepts of weak and strong unbiasedness considered earlier. Notice that on average the Bayesian procedures point more towards $H_0$, i.e., they assign a larger degree of belief to $H_0$, than to the alternative.

Although it is not completely unrelated, Lindley’s paradox does not refer to the difference in expectations, but to the difference of the procedures in the tails of the null distribution. Several lower bounds for $\alpha_\nu(x)$, corresponding to different classes of prior measures $\nu$, were derived in, e.g. Casella–Berger [20] and Berger–Sellke [11]. In many concrete examples the $P$–value is smaller than these lower bounds, for outcomes $x$ that are ‘unlikely’ under $H_0$. This illustrates the discrepancy between $P$–values and posterior probabilities based on classical Bayesian priors. A trivial lower bound for all $\alpha_\nu(x)$, corresponding to the class of all posterior probabilities based on prior probability measures satisfying $\nu(\{\theta_0\}) \geq \varepsilon > 0$, is given by

$$\alpha_\nu(x) \geq \text{LB}_\varepsilon(x) = \frac{\varepsilon p_{\theta_0}(x)}{\varepsilon p_{\theta_0}(x) + (1 \Leftrightarrow \varepsilon) p_{\theta_{\text{ml}}(x)}(x)},$$

(2.2)

where $\hat{\theta}_{\text{ml}}(x)$ is the maximum likelihood estimator of $\theta$. Notice that $\text{LB}_\varepsilon(x) = \alpha_\nu(x)$, where $\nu_\varepsilon(x) = \varepsilon \delta_{\theta_0} + (1 \Leftrightarrow \varepsilon) \delta_{\hat{\theta}_{\text{ml}}(x)}$ and $\delta$ denotes Dirac measure. Moreover, in the case that $\hat{\theta}_{\text{ml}}(x) = \theta_0$, the lower bound degenerates into $\text{LB}_\varepsilon(x) = \varepsilon$. 

The restriction to the class $D_w$ by no means excludes posterior probabilities from the discussion. If one takes the class $D_p$ of all (formal) posterior probabilities and intersects it with $D_w$ then, although this intersection will not include elements from $D_p$ with $\nu(\{\theta_0\}) = \nu(\Theta_A) = \frac{1}{2}$, it is still a very rich class. This can be seen by the following argument. With every $\sigma$–finite measure $\tilde{\nu}$ concentrated on $\Theta_A$, i.e., $\tilde{\nu}(\Theta) = \tilde{\nu}(\Theta_A)$, there can be associated a class of prior measures $\{\nu_c : c \in (0, \infty)\}$, where $\nu = c\delta_{\theta_0} + \tilde{\nu}$. The family $\{\alpha_{\nu_c} : c \in (0, \infty)\}$ of the corresponding posterior probabilities will always contain one element $\alpha_{\nu_c^*} \in D_w$, because the equation
\[
f(c) = \mathbb{E}_{\alpha_{\nu_c}(X_{\theta_0})} = \frac{1}{2},
\]
will always have a unique solution $c^*$. Notice that applying this recipe is a method for constructing weakly unbiased procedures.

Besides the connection of posterior probabilities with the Bayesian approach, there is a link between them and the Neyman–Pearson–Wald approach. This link is provided by the Complete–Class Theorem. An attractive feature of posterior probabilities is that the Bayes risk $r(\alpha) = \int_{\Theta} R(\theta, \alpha) \nu(d\theta)$, (2.3)
is minimum if $\alpha = \alpha_{\nu}$. If $r(\alpha_{\nu}) < \infty$, then this implies the admissibility of $\alpha_{\nu}$, i.e. $R(\theta, \alpha) \leq R(\theta, \alpha_{\nu})$, for all $\theta \in \Theta$, implies that $R(\theta, \alpha) = R(\theta, \alpha_{\nu})$, for all $\theta \in \Theta$. Now the Complete–Class Theorem tells that the closure of $D_p$ (a topological argument is involved) contains all admissible procedures. The following formulation of this result can be found together with its proof in Brown [16].

**Theorem 2.1** Let $X \subset \mathbb{R}^k$ and assume that the densities $p_\theta$ satisfy $p_\theta(x) > 0$ for all $x \in X$ and $\theta \in \Theta$. Let $\zeta : \Theta \mapsto \mathbb{R}$ be a measurable mapping. If $\zeta(\theta)$ is to be estimated under squared error loss, then every admissible estimator is a limit of Bayes estimators based on priors with finite support. More precisely, if $\alpha$ is admissible then there exists a sequence of prior measures $\{\nu^{(n)}\}_{n \in \mathbb{N}}$ each $\nu^{(n)}$ being supported on a finite set $\\{\theta_1^{(n)}, \ldots, \theta_{N_n}^{(n)}\}$ such that the corresponding sequence of Bayes estimators $\{\alpha_{\nu^{(n)}}, \alpha_n \}_{n \in \mathbb{N}}$ satisfies $\alpha_n \to \alpha$ a.e. ($\mu$).

A procedure is called inadmissible if it is not admissible. Notice that this use of language is somewhat misleading, because it refers to only one characteristic of a procedure. There are many examples in mathematical statistics where useful procedures are discredited as inadmissible without offering an acceptable alternative.
2.3 Optimal procedures

In the theory of point estimation, an estimator is evaluated on the basis of both its bias and its risk. Often the restriction to the class of mean-unbiased estimators is so restrictive that it contains a uniformly best element. A similar situation appears in the context of discussing the truth or falsity of hypotheses, if the family $\mathcal{P} = \{ P_\theta : \theta \in \Theta \subset \mathbb{R} \}$ has monotone likelihood ratio, and non-randomized UMP or UMPU Neyman–Pearson tests exist of all levels. In this case the restriction to $\mathcal{D}_s$ provides a uniformly best procedure, in the sense that it has uniformly minimum risk. The multidimensional case with UMPU tests will be postponed until the next chapter.

**Theorem 2.2** Suppose that $\mathcal{P} = \{ P_\theta : \theta \in \Theta \subset \mathbb{R} \}$ is dominated by Lebesgue measure, and that $\{ \phi_\epsilon : \epsilon \in [0, 1] \}$ is a family of UMP size-$\epsilon$ tests with nested critical regions for testing $H_0$ against $A$. Let $\alpha_p(x) = \inf \{ \epsilon : \phi_\epsilon(x) = 1 \}$ be the $P$-value corresponding to this family of tests. Then

$$R(\theta, \alpha_p) \leq R(\theta, \alpha) \quad \forall \theta \in \Theta,$$

for every $\alpha \in \mathcal{D}_s$.

**Proof.** The original proof was given by Schaafsma [79]. The conditions of the theorem imply that $\alpha_p \in \mathcal{D}_s$, and hence that $R(\theta_0, \alpha_p) = R(\theta_0, \alpha) = \int_{[0,1]} a^2 da = \frac{1}{2}$. For $\theta \in \Theta_A$, the fact that $\alpha_p(x) \leq \epsilon \Leftrightarrow \phi_\epsilon(x) = 1$ implies

$$P(\alpha_p(X_\theta) \leq \epsilon) = \mathbb{E}\phi_\epsilon(X_\theta) = \beta_\epsilon(\theta),$$

where $\beta_\epsilon(\theta)$ denotes the power of $\phi_\epsilon$. Similarly, $P(\alpha(X_\theta) \leq \epsilon) = \tilde{\beta}_\epsilon(\theta)$, where $\tilde{\beta}_\epsilon(\theta)$ is the power of the size-$\epsilon$ test $\tilde{\phi}_\epsilon = 1_{\{x : \alpha_\epsilon(x) \leq \epsilon\}}$. Combining the fact that $\phi_\epsilon$ is UMP size-$\epsilon$, which implies

$$P(\alpha_p(X_\theta) \leq \epsilon) \geq P(\alpha(X_\theta) \leq \epsilon) \quad \forall \epsilon \in [0, 1] \quad \forall \theta \in \Theta_A,$$

and the fact that $R(\theta, \alpha) = \mathbb{E}\alpha(X_\theta)^2$, for $\theta \in \Theta_A$, one obtains that $R(\theta, \alpha_p) \leq R(\theta, \alpha)$. 

Although from a mathematical point of view it is nice to obtain the $P$-value as best procedure within some restricted class, it does not imply that the $P$-value is automatically the most appropriate procedure; it could well be that the restriction to $\mathcal{D}_s$ excludes superior competitors.

The restriction to the class $\mathcal{D}_w$ will not be too restrictive; it contains many interesting procedures like, e.g., the $P$-value and many posterior probabilities.
The class $D_w$ will not contain a uniformly best element. Take for example the
minimax procedure $\alpha \equiv 1/2 \in D_w$ that is almost perfect in the neighborhood of
$\theta_0$ but has a relatively large risk for $\theta \in \Theta_\lambda$ far from $\theta_0$. Hence, to find good
procedures $\alpha \in D_w$, additional optimality principles are needed. In the previous
section it was mentioned that posterior probabilities minimize the Bayes risk w.r.t. some
measure $\nu$ over the class $D$ of all procedures. The following theorem specifies how to find procedures that minimize the Bayes risk w.r.t. some
measure $\nu$ over the smaller class $D_w$.

**Theorem 2.3** Let $\nu$ be a measure on $\Theta$ and define the measure $\mu_\nu$ on $X$ by
$\mu_\nu(B) = \int_\Theta P_\theta(B) \, \nu(d\theta)$. Suppose that $\mu_\nu$ is $\sigma$–finite, that $d\mu_\nu(x)/dx > 0$ a.e. $(\mu)$, and let

$$h_\nu = \frac{P_{\theta_0}}{\int_\Theta P_\theta \, \nu(d\theta)}$$

denote the Bayes factor in favor of $H_0$. Assume that $\mathbb{E} h_\nu(X_{\theta_0}) < \infty$. Define

$$\alpha_{\nu,c} = \begin{cases} 
(\alpha_c + c h_\nu) \wedge 1 & \text{if } c \geq 0, \\
(\alpha_c + c h_\nu) \vee 0 & \text{if } c < 0.
\end{cases} \quad (2.4)$$

Then (i) there exists a $c^*$ such that $\alpha_{\nu,c^*} \in D_w$, and (ii) if $r_\nu(\alpha_{\nu,c^*}) < \infty$ then $r_\nu(\alpha_{\nu,c^*}) = \min_{\alpha \in D_w} r_\nu(\alpha)$.

The proof that will presented below, however, gives insight in the structure of the optimal procedure.

**Proof.** (i) Under the conditions of the theorem $0 < \mathbb{E} h_\nu(X_{\theta_0}) < \infty$, and hence
$f(c) = \mathbb{E} \alpha_{\nu,c}(X_{\theta_0})$ is continuous, strictly increasing, and satisfies: $\lim_{c \to -\infty} f(c) = 0$, and $\lim_{c \to \infty} f(c) = 1$. This provides that there exists a unique solution $c^*$ of the equation $f(c) = 1/2$.

(ii) To find the optimal procedure, introduce the Hilbert space $L^2(\mu_\nu)$ with inner product

$$\langle f, g \rangle_\nu = \int_X f(x) g(x) \, \mu_\nu(dx).$$

Rewriting the problem in terms of this Hilbert space provides that

$$r_\nu(\alpha) = r_\nu(\alpha_\nu) + \langle \alpha \Leftrightarrow \alpha_\nu, \alpha \Leftrightarrow \alpha_\nu \rangle_\nu \quad \text{and} \quad \alpha \in D_w \Leftrightarrow \langle h_\nu, \alpha \rangle_\nu = \frac{1}{2}.$$
Now, notice that \( r_\nu(\alpha_{\nu,c^*}) < \infty \) implies \( \alpha_{\nu,c^*} \in L^2(\mu_\nu) \). Define \( \tilde{D}_w = \{ \beta \in L^2(\mu_\nu) : \langle \beta, h_\nu \rangle_\nu = \frac{1}{2} \} \), and observe that \( D_w \subset \tilde{D}_w \). Taking the orthogonal projection
\[
\beta^\perp = \alpha_\nu + c^* h_\nu \quad \text{with} \quad c^* \text{ s.t.} \quad \mathbb{E}\beta^\perp(X_{\theta_0}) = \frac{1}{2}.
\]
of the Bayes procedure \( \alpha_\nu \) on \( \tilde{D}_w \), one minimizes \( r_\nu(\beta) \) on \( \tilde{D}_w \). Hence, \( 0 \leq \beta^\perp \leq 1 \) implies \( \beta^\perp = \alpha_{\nu,c^*} \in D_w \) which satisfies \( r_\nu(\alpha_{\nu,c^*}) = \min_{\alpha \in D_w} r_\nu(\alpha) \).

It remains to deal with the situation \( \beta^\perp \in \tilde{D}_w \setminus D_w \). In this case an argument closely related to that in the proof of the Neyman–Pearson Fundamental Lemma is needed. Notice that every \( \alpha \in D_w \) can be represented as \( \alpha = \alpha_{\nu,c^*} + \gamma \), where \( \gamma \) satisfies \( \langle \gamma, h_\nu \rangle_\nu = 0 \). Partition \( X \) into
\[
X_- = \{ x \in X : \alpha_{\nu,c^*}(x) = 0 \} \quad \text{and} \quad X_+ = \{ x \in X : \alpha_{\nu,c^*}(x) = 1 \},
\]
and observe that \( \gamma(x) \geq 0 \) if \( x \in X_- \), and \( \gamma(x) \leq 0 \) if \( x \in X_+ \). After rewriting
\[
r_\nu(\alpha) = r_\nu(\alpha_\nu) + \langle \alpha_{\nu,c^*} \Leftrightarrow \alpha_\nu, \alpha_{\nu,c^*} \Leftrightarrow \alpha_\nu \rangle_\nu + 2\langle \alpha_{\nu,c^*} \Leftrightarrow \alpha_\nu, \gamma \rangle_\nu + \langle \gamma, \gamma \rangle_\nu,
\]

it is easy to see that it suffices to show that \( \langle \alpha_{\nu,c^*} \Leftrightarrow \alpha_\nu, \gamma \rangle_\nu \geq 0 \). Now, notice that
\[
\langle \alpha_{\nu,c^*} \Leftrightarrow \alpha_\nu, \gamma \rangle_\nu = c^* \langle h_\nu, \gamma \rangle_\nu + \int_{X_-} c^* h_\nu(x) \gamma(x) \mu_\nu(dx)
+ \int_{X_+} (1 \Leftrightarrow c^* h_\nu(x)) \gamma(x) \mu_\nu(dx).
\]
The first term equals 0, and the second and third term are positive because on \( X_- \) both \( \gamma, c^* h_\nu \leq 0 \), and on \( X_+ \) both \( 1 \Leftrightarrow c^* h_\nu \), \( \gamma \leq 0 \). \( \blacksquare \)

Using this theorem, a nontrivial lower bound for the risk functions corresponding to \( \alpha \in D_w \) can be obtained. Notice that taking \( \nu = \delta_\theta \), for some fixed \( \theta \in \Theta \), and applying Theorem 2.3, provides the S(omewhere) M(imum) R(isk)(\( D_w \)) procedure, denoted by \( \alpha_0 \). This SMR\( (D_w) \) procedure satisfies the identity
\[
r_{\delta_\theta}(\alpha_\theta) = R(\theta, \alpha_\theta) = \min_{\alpha \in D_w} R(\theta, \alpha) = R_{D_w}(\theta), \tag{2.5}
\]
where \( R_{D_w}(\theta) \) denotes the envelope risk w.r.t. the class of weakly unbiased procedures. The class of all SMR\( (D_w) \) procedures will be denoted by \( C_w \subset D_w \), i.e. \( C_w = \{ \alpha_\theta : \theta \in \Theta \} \). Notice that the minimax procedure \( \alpha_{\theta_0} \equiv \frac{1}{2} \in C_w \). The regret or shortcoming function of a procedure \( \alpha \in D_w \) is defined as
\[
S(\theta, \alpha) = R(\theta, \alpha) \Leftrightarrow R_{D_w}(\theta) \geq 0. \tag{2.6}
\]
2.3. Optimal procedures

The advantage of comparing regret functions instead of risk functions is that they focus on the essential features of the procedures. Notice that \( r_\nu(\alpha_\nu) < \infty \) implies that \( \int_{\Theta} R_{D_w}(\theta) \nu(d\theta) < \infty \), and hence that the minimization of the integrated risk is equivalent to the minimization of the integrated regret.

There are two strategies for finding procedures that minimize the Bayes risk or regret over \( D_w \). The first is, of course, to take some measure \( \nu \) and to apply Theorem 2.3 to obtain \( \alpha_{\nu,c^*} \). In general, such \( \alpha_{\nu,c^*} \) will not coincide with any posterior probability, or the limit of posterior probabilities. In these cases \( \alpha_{\nu,c^*} \) will not be admissible, although there cannot exist a weakly unbiased procedure that has uniformly smaller risk. The second approach is to adjust the prior measure \( \nu_c \) such that the corresponding posterior probability \( \alpha_{\nu,c} \in D_w \). This latter method will usually result in an admissible procedure.

Instead of minimizing a Bayes risk, one could consider minimax principles. The minimax risk procedure \( \alpha \equiv \frac{1}{2} \) is not very interesting, but the \( M(\text{inimax}) R(\text{regret})(D_w) \) procedure will be, because it is nondegenerate. Like the minimax risk procedure, it may be characterized by the existence of a measure \( \nu^* \) such that \( \alpha_{\nu^*} \in D_w \) and

\[
\sup_{\theta \in \Theta} S(\theta, \alpha_{\nu^*}) = s_{\nu^*}(\alpha_{\nu^*}),
\]

where \( s_{\nu^*}(\alpha_{\nu^*}) \) denotes the Bayes regret. Such a measure \( \nu^* \) is called the least favorable prior and the corresponding \( \alpha_{\nu^*} \) is the \( MR(D_w) \) procedure. Hence, searching for the \( MR(D_w) \) procedure boils down to searching the appropriate prior \( \nu^* \). The \( MR(D_w) \) procedure can be considered as the analogue of most stringent Neyman-Pearson tests introduced by Wald during the war, see e.g. Oosterhoff-van Zwet [68] or Schaafsma [78]. In general, the problem of finding the \( MR(D_w) \) procedure is very hard. A less ambitious approach consists of finding the procedure that minimizes the maximum regret over the much smaller class \( C_w \). This procedure will be called the \( M(\text{inimax}) R(\text{regret}) S(\text{omewhere}) M(\text{inimum}) R(\text{isk})(D_w) \) procedure. It is the analogue of the most stringent somewhere most powerful Neyman-Pearson test, see e.g. Schaafsma [77]. If \( \Theta \subset \mathbb{R} \), the problem of finding the \( MRS(\text{MRSMR}) D_w \) procedure can often be solved by numerical investigation.
2.4 $\chi^2$—Problems

To illustrate the principles mentioned in this chapter, three concrete problems will be elaborated upon (recall that $t$ denotes the true value of the parameter):

- **Problem 1**: $\mathcal{L}X_\theta = \theta \chi^2_{V,\theta}$, $H_0 : t = 1$, $A_1 : t > 1$,
- **Problem 2**: $\mathcal{L}X_\theta = \chi^2_{V,\theta}$, $H_0 : t = 0$, $A_2 : t > 0$,
- **Problem 3**: $\mathcal{L}X_\theta = \theta + \chi^2_{V}$, $H_0 : t = 0$, $A_3 : t > 0$,

where $\chi^2_{V,\theta}$ denotes a noncentral $\chi^2$ distribution with $v$ degrees of freedom and noncentrality parameter $\theta$. There are several interesting features of these problems. First notice that the distribution of $X_\theta$ is a $\chi^2_v$ distribution under $H_0$, for all three problems. The alternatives $A_i$, $i = 1, 2, 3$ all specify stochastically larger distributions, even with monotone likelihood ratio. Consequently, the Neyman–Pearson size-$\epsilon$ tests that reject for large values of $x$ coincide and are UMP. The $P$-value corresponding to these tests is given by

$$
\alpha_p(x) = P(\chi^2_v \geq x) = 1 - \frac{\left(\frac{1}{v+1}x\right)^v \exp\left\{-\frac{1}{v+1}x\right\} \Gamma\left(1; \frac{1}{2}v + 1; \frac{1}{2}x\right)}{\left(\frac{v}{v+1}\right)^v},
$$

(2.7)

which specializes to $\exp\left\{-\frac{1}{v+1}x\right\} \sum_{n=0}^{v-1} \frac{\left(\frac{1}{v+1}x\right)^n}{n!}$ for $v$ even. Notice that

$$
i\Gamma_{a_1, \ldots, a_i; b_1, \ldots, b_j; x}^{k} = \sum_{k=0}^{\infty} \frac{(a_1)_k \cdots (a_i)_k x^k}{(b_1)_k \cdots (b_j)_k} \frac{1}{k!}.
$$

The conditions of Theorem 2.2 are satisfied, and hence the class $D_s$ contains a uniformly best element: the $P$-value. In contrast to the $P$-value, the procedures arising from other optimality principles will depend on the explicit formulation of the alternative.

A second feature of the three problems is their ordering. Starting with $(H_0, A_3)$ and adding the assumption that the unknown true value $t$ of the shift $\theta$ is itself the outcome of a random variable $T$, such that, given the outcome $k$ of some other random variable $K$, with $\mathcal{L}K = \text{Poisson}(\frac{1}{2}t)$, the distribution of $T$ is $\chi^2_{V+2K}$, the marginal distribution becomes $\mathcal{L}X_\theta = \chi^2_{V,\theta}$, and $(H_0, A_2)$ is obtained. Similarly, starting with $(H_0, A_2)$ and adding the assumption that the unknown true value $t$ of the noncentrality parameter $\theta$ is the outcome of a random variable $T$, such that $\mathcal{L}T = (\theta \leftrightarrow 1) \chi^2_{V}$, the marginal distribution becomes $\mathcal{L}X_\theta = \theta \chi^2_{V}$, and $(H_0, A_1)$ is obtained. It is to be expected that by
adding extra assumptions, the uncertainty will be decreased, and hence that the procedures will have smaller risk at \( \Theta_H \). Next Problems 1, 2 and 3 will be dealt with. The feature just mentioned appears in the scale of the regret functions in Figures 2.1, 2.2 and 2.3.

**Remark 2.1** At this point the author would like to thank R. de Bruin, who wrote the computer programmes that were needed for the computation of all regret functions. These regret functions are displayed in Figures 2.1, 2.2 and 2.3.

**Problem 1** The testing problem \((H_0, A_1)\) arises, for example, if one wants to test the null hypothesis that the variance of i.i.d. normal variates is 1, against the alternative that it is larger than 1. The lower bound (2.2) for all posterior probabilities based on prior probabilities that satisfy \( \nu(\{1\}) = \nu((1, \infty)) = \frac{1}{2} \), is given by

\[
LB(x) = \begin{cases} 
\frac{1}{2} & \text{if } x \leq v, \\
\frac{1}{1+(\frac{x}{v})\frac{1}{2v} \exp\left(\frac{1}{2}(x-v)\right)} & \text{if } x > v.
\end{cases}
\] (2.8)

The discrepancy between the P-value and any Bayesian procedure is illustrated by the fact that the P-value is smaller than this lower bound for large values of \( x \). To state it even stronger: for large values of \( x \), the P-value will be smaller than any lower bound for posterior probabilities based on finite priors, as will be shown in the following lemma.

**Lemma 2.1** Let \( \nu(\Theta) = 1 \) be such that \( \alpha_\nu \neq 0 \). Then there exists a \( x_0 \in (0, \infty) \) such that \( x > x_0 \) implies that \( \alpha_p(x) < \alpha_\nu(x) \), while \( \alpha_p(x) \) decreases faster than any posterior probability as \( x \) increases.

**Proof.** The requirement that \( \alpha_\nu \neq 0 \), implies that there exists an \( \epsilon > 0 \) such that \( \nu(\{\theta_0\}) \geq \epsilon \). Hence, \( \alpha_\nu \geq LB_\epsilon \), where \( LB_\epsilon \) is given by (2.2). If \( x \to \infty \), then both \( \alpha_p(x) \) and \( LB_\epsilon(x) \) converge monotonically to 0. As both \( \alpha_p \) and \( LB_\epsilon \) are finite, it suffices to show that the rate of convergence of \( \alpha_p \) is larger than that of \( LB_\epsilon \). Using the fact that

\[
iF_1 \left( 1; \frac{1}{2}v + 1; \frac{1}{2}x \right) = \frac{(\frac{1}{2}v + 1)}{(\frac{1}{2}x)^{\frac{1}{2}v} \exp\left(\frac{1}{2}(x-v)\right)} \left(1 + O\left(|x|^{-1}\right)\right),
\]

one obtains that \( \alpha_p(x) \sim \alpha_\nu \frac{1}{2v-1} \exp\left\{\frac{1}{2}x\right\} \to 0 \). Adjusting (2.8) slightly to incorporate that \( \nu(\{1\}) = \epsilon \) instead of \( \frac{1}{2} \), one can see that \( LB_\epsilon(x) \sim \epsilon \frac{1}{2v} \exp\left\{\frac{1}{2}x\right\} \to 0 \).
Consequently, the P-value cannot correspond to any posterior probability w.r.t. a finite measure \( \nu \). This, however, does not exclude the possibility that it can be obtained as the limit of a sequence of Bayes procedures. To show that this possibility does not exist, one can make use of another property of the P-value. Using Theorem 2.1, one can show that any procedure that attains 1 in some point of continuity is inadmissible.

**Lemma 2.2** Let \( X = [x_0, \infty) \), \( \Theta = [\theta_0, \infty) \), and \( \mathcal{P} \) be a family of probability distributions with monotone likelihood ratio, and probability densities \( p_\theta(x) > 0 \) a.e. \((\mu)\), for all \( \theta \in \Theta \). Let \( \alpha : X \mapsto [0, 1] \) be non-constant, with \( \alpha(x_0) = 1 \) and right continuous at \( x_0 \). If there exists a function \( b \) such that

\[
\lambda_\theta(x) = \frac{p_{\theta_0}(x)}{p_{\theta_0}(x) + \sum_{i=1}^{N_n} p_{\theta_i}(x)} \leq b(x) < \infty \quad \text{a.e.}(\mu),
\]

then \( \alpha \) is inadmissible for estimating \( \mathbb{I}_{\{\theta_0\}}(t) \) under squared error loss.

**Proof.** The conditions of Theorem 2.1 are satisfied, and hence it has to be shown that there cannot exist a sequence \( \{\alpha_n\}_{n \in \mathbb{N}} \) such that \( \alpha_n \to \alpha \) a.e. \((\mu)\). The notations used in the proof of this theorem correspond to those used in Theorem 2.1. If the support of the prior does not contain \( \{\theta_0\} \), then \( \alpha_n \equiv 0 \). Hence it can be assumed w.l.o.g. that \( \nu^{(n)}_0 > 0 \). Write

\[
\alpha_n(x) = \frac{\nu^{(n)}_0 p_{\theta_0}(x)}{\nu^{(n)}_0 p_{\theta_0}(x) + \sum_{i=1}^{N_n} \nu^{(n)}_i p_{\theta_i}(x)} \leq \frac{\varphi^{(n)}_0}{\varphi^{(n)}_0 + \sum_{i=1}^{N_n} \varphi^{(n)}_i \lambda_{\theta_i}(x)},
\]

where \( \varphi^{(n)}_i \propto \nu^{(n)}_i p_{\theta^{(n)}_i}(x_0)/p_{\theta_0}(x_0) \). It follows from the monotone likelihood ratio that \( \lambda_{\theta_i}(x) \) is an increasing function, with \( \lambda_{\theta_i}(x_0) = 1 \). Now, define a random variable \( \Delta_n \) such that \( P(\Delta_n = \theta^{(n)}_i) \propto \varphi^{(n)}_i \), and rewrite

\[
\alpha_n(x) = \frac{P(\Delta_n = \theta_0)}{\mathbb{E}\lambda_{\Delta_n}(x)}. \quad (2.9)
\]

The fact that \( \lambda_{\theta_i}(x) \geq 1 \), provides \( \alpha_n(x) \leq P(\Delta_n = \theta_0) \leq 1 \) a.e. As \( \alpha(x_0) = 1 \) and \( \alpha \) is right continuous at \( x_0 \), it follows that \( P(\Delta_n = \theta_0) \to 1 \), which means that \( \Delta_n \to \theta_0 \) in probability. Consequently, \( \lambda_{\Delta_n}(x) \to 1 \) in probability. As \( \lambda_{\theta}(x) \leq b(x) < \infty \) a.e. \((\mu)\), the sequence \( \{\lambda_{\Delta_n}(x)\}_{n \in \mathbb{N}} \) is uniformly integrable a.e. \((\mu)\) and, hence \( \mathbb{E}\lambda_{\Delta_n}(x) \to 1 \) a.e. \((\mu)\). This in its turn implies \( \alpha_n \to 1 \) a.e. \((\mu)\), which leads to a contradiction. \( \blacksquare \)
The family of $\chi^2_\nu$ distributions and the P-value $\alpha_\theta$ satisfy the conditions of Lemma 2.2, with
\[
\lambda_\theta(x) = \exp \left\{ \frac{\theta}{2\theta} x \right\} \leq \exp \left\{ \frac{x}{2} \right\} = b(x),
\]
and hence the P-value is inadmissible. Hence, it can be concluded that the reduction by strong unbiasedness is so restrictive that it does not allow any admissible procedures. To see whether this implies that the P-value is an inappropriate degree of belief for this problem, a comparative analysis is made.

The attention will now be concentrated on finding interesting competitors for the P-value within the class $\mathcal{D}_w$. To start with, however, the envelope risk for the class $\mathcal{D}_w$ will be computed. Using Theorem 2.3 with $\nu = \delta_\theta$, the SMR($\mathcal{D}_w$) procedures can be obtained. If $\nu \leq 2$ or $\theta < 1/(2 \Leftrightarrow 4^{\frac{1}{2}})$, then no truncation is necessary and
\[
\alpha_\theta(x) = \frac{1}{2} \left( \frac{2\theta \Leftrightarrow 1}{\theta} \right)^{\nu} \exp \left\{ \frac{\theta \Leftrightarrow 1}{2\theta} x \right\}. \tag{2.10}
\]
For these $\theta$, the envelope risk is given by
\[
R_{\mathcal{D}_w}(\theta) = \frac{1}{4} \left( \frac{2\theta \Leftrightarrow 1}{\theta^2} \right)^{\nu}. \tag{2.11}
\]
If $\theta$ is larger then truncation is needed to compute $\alpha_\theta$, and one has to rely on numerical methods. Nevertheless, (2.11) still provides a lower bound for the envelope risk as can be seen from the following lemma.

**Lemma 2.3** The envelope risk satisfies $R_{\mathcal{D}_w}(\theta) \geq \frac{1}{4} \left( \frac{2\theta \Leftrightarrow 1}{\theta^2} \right)^{\nu}$, for all $\theta \in [1, \infty)$.

**Proof.** Recalling the proof of Lemma 2.3, one can see that the expression on the r.h.s. of (2.10) is the projection $\beta_\theta^\perp$ of $\alpha_\nu \equiv 0$ on $\hat{\mathcal{D}}_w$, for all $\theta \in [1, \infty)$. In the case that $\nu \leq 2$ or $\theta \leq 1/(2 \Leftrightarrow 4^{\frac{1}{2}})$, it follows that $0 \leq \beta_\theta^\perp \leq 1$, and hence $\beta_\theta^\perp$ coincides with $\alpha_\theta$ which provides that
\[
\frac{1}{4} \left( \frac{2\theta \Leftrightarrow 1}{\theta^2} \right)^{\nu} = r_{\delta_\theta}(\beta_\theta^\perp) = r_{\delta_\theta}(\alpha_\theta) = R(\theta, \alpha_\theta) = R_{\mathcal{D}_w}(\theta).
\]
On the other hand, if $\nu > 2$ and $\theta > 1/(2 \Leftrightarrow 4^{\frac{1}{2}})$, then
\[
\frac{1}{4} \left( \frac{2\theta \Leftrightarrow 1}{\theta^2} \right)^{\nu} = r_{\delta_\theta}(\beta_\theta^\perp) < r_{\delta_\theta}(\alpha_\theta) = R(\theta, \alpha_\theta) = R_{\mathcal{D}_w}(\theta),
\]
Figure 2.1: Both figures correspond to Problem 1, with \( v = 2 \). In the figure on the l.h.s. all procedures are plotted together with the lower bound on the Bayes procedures. In the figure on the r.h.s. the corresponding regret functions are displayed. Notice that the axis of the parameter has a logarithmic scale.

which proves the lemma.

To find procedures in \( D_w \) that minimize the Bayes risk, a measure \( \nu \) has to be specified. In general, there will be no particular \( \theta \) in the alternative that is of special interest, hence it seems reasonable to take a measure \( \nu \ll \lambda \). Taking \( \nu \) such that \( \nu(d\theta) = \theta^r d\theta \), where \( r \leq \frac{1}{2} v \Leftrightarrow 1 \), provides that the Bayes factor in favor of \( H_0 \) is given by

\[
h_\nu(x) = \frac{\frac{1}{2}v \Leftrightarrow r \Leftrightarrow 1}{\text{B}(1; \frac{1}{2}v \leftrightarrow r; \frac{1}{2}x)}.
\]

The choice \( r = \Leftrightarrow 1 \) is of particular interest, because it is the natural weight function for scale alternatives. As mentioned earlier, two strategies can be followed. As it is required that \( \alpha \in D_w \), no point mass needs to be assigned to \( \{\theta_0\} \) and Theorem 2.3 can be used to obtain \( \alpha_{\nu,c^*} \in D_w \). The other possibility is to change the prior \( \nu \) by adjusting the point mass in \( \{\theta_0\} \) to \( \nu_{<*} \), in such a way that the corresponding posterior probability \( \alpha_{\nu,c^*} \in D_w \).

The case \( v = 2 \) will be worked out explicitly. Notice that (2.7) simplifies to \( \alpha_p(x) = \exp\{\Leftrightarrow \frac{1}{2} x\} \). Hence, (2.10) provides that \( \lim_{\theta \to \infty} \alpha_\theta(x) = \alpha_p(x) \). In other words, the P-value corresponds to the limit of elements in \( C_w \). The risk function, and hence the regret function, of \( \alpha_p \) can be computed explicitly. The
latter is given by
\[ S(\theta, \alpha_p) = \frac{1}{8\theta^3 + 4\theta^2} . \]

The regret function of any SMR(\(\mathcal{D}_w\)) procedure \(\alpha_{\vartheta}\) is given by
\[ S(\theta, \alpha_{\vartheta}) = \frac{\frac{1}{\vartheta^2} \theta^2 \Leftrightarrow \frac{1}{\vartheta} \theta + 1}{8(\frac{\vartheta - 1}{\vartheta})\theta^3 + 4\theta^2} . \] (2.12)

It has a local maximum at \(\theta = 1\), hits 0 in the local minimum at \(\theta = \vartheta\), has a second local maximum at some \(\theta > \vartheta\), and after this second local maximum it decreases to 0. To find the MRSMR(\(\mathcal{D}_w\)) procedure, the two local maxima have to be balanced by choosing \(\vartheta\) appropriately; by increasing \(\vartheta\) the local maximum at 1 will increase and the other local maximum will decrease, and vice versa by decreasing \(\vartheta\) the local maximum at 1 will decrease and the other local maximum will increase. Numerical minimization of the maximum regret provides that the MRSMR(\(\mathcal{D}_w\)) procedure is \(\alpha_{\theta}\), with \(\theta = 1.44\), i.e.
\[ \alpha_{\text{MRSMR}}(x) = 0.65 \exp\{1.56x\} . \]

Specializing (2.12) to \(\vartheta = 1.44\) provides
\[ S(\theta, \alpha_{\text{MRSMR}}) = \frac{0.48\theta^2 \Leftrightarrow 0.69\theta + 1}{2.44\theta^3 + 4\theta^2} . \]

Minimizing the integrated risk w.r.t. the weight function \(\nu(d\theta) = \theta^{-1}\), the two procedures, corresponding to the strategies that have been proposed, can be obtained. The first approach is to apply Theorem 2.3, which provides
\[ \alpha_{\nu, c^*}(x) = \frac{c^* x}{\exp\{1.5x\} \Leftrightarrow 1} , \]
with \(c^* = 0.39\). The second approach is to appropriately choose the point mass assigned to \(\{1\}\), and compute the posterior probability. This provides
\[ \alpha_{\nu, c^*}(x) = \frac{x}{x + c^*(\exp\{1.5x\} \Leftrightarrow 1}) , \]
with \(c^* = 1.16\). In Figure 2.1 all procedures and corresponding regret functions are displayed. Notice that the regret function of \(\alpha_{\text{MRSMR}}\) displays, indeed, the qualitative behavior of a SMR(\(\mathcal{D}_w\)) procedure, and that the two local maxima
of this regret function are of the same size. In the case that \( v = 2 \), the P-value resembles a SMR(\( D_w \)) procedure in the sense that \( \alpha_p = \alpha_{\infty} \). As it minimizes the risk at infinity, the local maximum at 1 is large. Comparing the regret functions of \( \alpha_{\nu,c} \) and \( \alpha_{\nu,*} \), one sees that by assigning point mass to \{1\} the regret will be smaller for \( \theta \) close to 1 at the cost of a larger regret for \( \theta \) far from 1. Although, the choice between the different procedures will always be a matter of taste, one might conclude from looking at Figure 2.1 that \( \alpha_{\nu,*} \) should be preferred to \( \alpha_{\text{MRSRM}} \) and that \( \alpha_p \) focuses too much on large values of \( \theta \). In our opinion \( \alpha_{\nu,*} \) compares most favorably to the procedures proposed, and we suggest to assign degrees of belief to \( H_0 \) according to this procedure.

In statistical practice, the null hypothesis is rejected, or in other words an effect is considered to be significant, if the P-value is smaller than 0.05 or 0.01. In industrial statistics this level is often taken to be even smaller, e.g. 0.001. To make the link between the results obtained for Problem 1 and the Neyman–Pearson theory, one can compare the degrees of belief assigned by the various procedures for values of \( x \) such that the P-value achieves one of these levels.

<table>
<thead>
<tr>
<th>( x )</th>
<th>5.991</th>
<th>9.210</th>
<th>13.81</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha_p(x) )</td>
<td>0.050</td>
<td>0.010</td>
<td>0.001</td>
</tr>
<tr>
<td>( \alpha_{\text{MRSRM}}(x) )</td>
<td>0.261</td>
<td>0.160</td>
<td>0.079</td>
</tr>
<tr>
<td>( \alpha_{\nu,c}(x) )</td>
<td>0.123</td>
<td>0.036</td>
<td>0.005</td>
</tr>
<tr>
<td>( \alpha_{\nu,*}(x) )</td>
<td>0.214</td>
<td>0.074</td>
<td>0.012</td>
</tr>
</tbody>
</table>

Table 2.1

Table 2.1 shows that, indeed, the P-value assigns smaller degree of belief to the hypothesis than the other procedures. The difference to \( \alpha_{\text{MRSRM}}(x) \) and \( \alpha_{\nu,c}(x) \) is considerable, whereas the difference to \( \alpha_{\nu,*}(x) \) is somewhat less dramatic. If one considers \( \alpha_{\nu,*} \) to be the most appropriate degree of belief for this problem, then Table 2.1 gives an impression of how much the degrees of belief assigned by P-value have to be adjusted.

**Problem 2** The testing problem \((H_0, A_2)\) arises, for example, in problems of testing goodness-of-fit, or in the problem of testing whether or not the mean of normally distributed random variables is equal to 0. Again the investigation is started by showing the discrepancy between the P-value and Bayesian procedures. To this end the P-value is compared with the lower bound (2.2) on posterior probabilities. This lower bound for posterior probabilities based on priors that satisfy \( \nu(\{\theta_0\}) = \nu(\Theta_\lambda) = \frac{1}{2} \) can be expressed in terms of \( \theta_{\text{ml}}(x) \),
2.4. $\chi^2$-Problems

i.e.

$$
LB(x) = \frac{\exp\{\frac{1}{2}\theta_{ml}(x)\}}{\exp\{\frac{1}{2}\theta_{ml}(x)\} + \mathcal{F}_1\left(\frac{1}{2}; \frac{1}{2} \theta_{ml}(x)x\right)}.
$$

(2.13)

Similarly to the previous problem, the rate of convergence of $\alpha_p$ to 0 as $x$ increases will be compared to that of the lower bound $LB_\varepsilon$.

**Lemma 2.4** The lower bound $LB_\varepsilon(x) \sim cx^{\frac{1}{2}(v-1)}\exp\{\frac{1}{2}x\}$, as $x \to \infty$.

**Proof.** To study the asymptotic behavior of $LB_\varepsilon$, one first has to look at the asymptotic behavior of $\theta_{ml}$. Notice that

$$
\hat{\theta}_{ml}(x) = \begin{cases} 
0 & \text{solution of } \mathcal{F}_0\mathcal{F}_1(\frac{1}{2}+1; \frac{1}{2}\theta(x) \leftrightarrow \mathcal{F}_1(\frac{1}{2}; \frac{1}{2} \theta(x)) = 0 & \text{if } x \leq v \\
\text{if } x > v
\end{cases}
$$

For $x > v$ the maximum likelihood estimator behaves like $\hat{\theta}_{ml}(x) \approx (x \leftrightarrow v+1) \Rightarrow (x \leftrightarrow v+1)^{-1}$. Hence, $\hat{\theta}_{ml}(x) = x(1 + O(|x|^{-1}))$, as $x \to \infty$. Using the facts

$$
\mathcal{F}_1\left(\frac{1}{2}v; \frac{1}{2}x^2(1 + O(|x|^{-1}))\right) \to 1 \quad \text{as} \quad x \to \infty,
$$

and

$$
\mathcal{F}_1\left(\frac{1}{2}v; \frac{1}{4}x^2\right) = \exp\{x\}(2x)^{-\frac{1}{2}(v-1)}\left(\frac{v \leftrightarrow 1}{(\frac{1}{2}v)}(1 + O(|2x|^{-1}))\right),
$$

one can complete the proof. \qed

The fact that $LB_\varepsilon$ has the same rate of convergence as $\alpha_p$, does not imply that $\alpha_p$ can be represented as a posterior probability or as the limit of a sequence of posterior probabilities. On the contrary, a slight modification of Lemma 2.2 can be used to show that the converse is true.

**Lemma 2.5** Let $\mathcal{X} = [0, \infty)$, $\Theta = [0, \infty)$, and $\mathcal{P}$ be the family of non-central $\chi^2$ distributions. Let $\alpha : \mathcal{X} \mapsto [0,1]$ be strictly positive, non-constant, continuous from the right at 0, and with $\alpha(0) = 1$. Then $\alpha$ is inadmissible for estimating $\mathbb{P}_{\{\theta_0\}}(t)$ under squared error loss.

**Proof.** The idea of the proof is the same as that of Lemma 2.2. The difference, however, is that there does not exist a function $b(x)$ such that $\lambda_\theta(x) \leq b(x)$, and hence it remains to show that the sequence $\{\lambda_{\Delta_n}(x)\}_{n \in \mathbb{N}}$ is uniformly integrable a.e. ($\mu$). To do so, one can use the following inequality. For $a \geq \frac{1}{2}$,

$$
(\mathcal{F}_1(a,x))^2 = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \frac{(a)^2}{(a+i)(a+j)(i+j)!} x^{i+j}.
$$
\[
\begin{align*}
&= \sum_{k=0}^{\infty} \sum_{i=0}^{k} \binom{k}{i} \frac{(a)^2 x^k}{(a+i)(a+k)k!} \\
&= \sum_{k=0}^{\infty} \frac{(a)^2 x^k}{(a+k)k!} \sum_{i=0}^{k} \binom{k}{i} \frac{(a+k)}{(a+i)(a+k)k!} \\
&= \frac{2^{2a-2} (a)^2}{\sqrt{\pi}} \sum_{k=0}^{\infty} \frac{(a)^2 (4x)^k}{(a+k)k!} \\
&\leq \frac{2^{2a-2} (a)^2}{\sqrt{\pi}} \sum_{k=0}^{\infty} \frac{(a)^2 (4x)^k}{(a+k)k!} \\
&= \frac{2^{2a-2} (a)}{\sqrt{\pi}} F_1(a, 4x).
\end{align*}
\]

Hence, for \( v \geq 1 \), the uniform integrability follows from

\[
E|\lambda_n(x)|^2 = E\left|_0 F_1 \left( \frac{1}{2}v; \frac{1}{4}\Delta_n x \right) \right|^2 \\
= E\left( F_1 \left( \frac{1}{2}v; \frac{1}{4}\Delta_n x \right) \right)^2 \\
\leq \frac{2^{2a-2} (\frac{1}{2})v}{\sqrt{\pi}} F_1 \left( \frac{1}{2}v; \frac{1}{4}\Delta_n 4x \right) \rightarrow \frac{2^{2a-2} (\frac{1}{2})v}{\sqrt{\pi}} \frac{1}{\alpha(4x)} < \infty
\]
a.e. (\( \mu \)), because \( \alpha > 0 \) a.e. (\( \mu \)). This completes the proof. \( \blacksquare \)

It follows immediately that the P-value is inadmissible. Again, it can be concluded that the reduction by strong unbiasedness is too restrictive that it does not allow any admissible procedures. To see whether this implies that the P-value is an inappropriate degree of belief for this problem, a comparative analysis is made.

The attention will now be concentrated on finding interesting competitors for the P-value within the class \( D_w \). To start with, however, the envelope risk for the class \( D_w \) will be computed. To this end, the same strategy is adopted as in the previous problem to obtain the class \( C_w = \{ \alpha_\theta : \theta \in [0, \infty) \} \) and the envelope risk \( R_{D_w}(\theta) \). Applying Theorem 2.3 with \( \nu = \delta_\theta \) provides

\[
\alpha_\theta = \left( c(\theta) \frac{\exp\{\frac{1}{2}\theta\}}{F_1(\frac{1}{2}v; \frac{1}{4}\theta x)} \right) \wedge 1, \quad (2.14)
\]
and for \( \theta \) sufficiently small no truncation will be necessary. To evaluate the envelope risk, numerical methods had to be used because no explicit formula could be given.

To obtain procedures that minimize the Bayes regret, weight functions of the form \( \nu(d\theta) = \theta^r d\theta \) are considered. These provide that the Bayes factor in favor of \( H_0 \) is given by

\[
h_\nu(x) = \frac{1}{(r + 1)2^{r+1}I_1(r + 1; \frac{1}{2}v; \frac{1}{2}x)}.
\]

Applying Theorem 2.3 one can obtain the procedure that minimizes the Bayes risk over the class \( \mathcal{D}_w \), this procedure will be denoted by \( \alpha_{\nu_{c^*}} \). Alternatively, one can adjust the measure \( \nu \) to \( \nu_{c^*} \) by assigning point mass to \( \{0\} \), in such a way that the corresponding posterior probability \( \alpha_{\nu_{c^*}} \in \mathcal{D}_w \). Interesting choices of \( r \), can be based on the principle of insufficient reason; \( r = 0 \) corresponds to a uniform prior for the non-centrality parameter \( \theta \), \( r = \frac{1}{2} \) to a uniform prior for its square root, and \( r = \frac{1}{2}v \) to a uniform prior for the vector of means in a context with a \( v \)-dimensional vector of deviates.

The case \( v = 2 \) will be worked out explicitly. The envelope risk \( R_{\mathcal{D}_w}(\theta) \) and the SMR(\( \mathcal{D}_w \)) procedures were obtained by numerical methods. The regret functions will again have the same qualitative behavior as in the previous problem. Hence, the MRSMR(\( \mathcal{D}_w \)) procedure was again obtained by balancing the two local maxima of the regret functions of the SMR(\( \mathcal{D}_w \)) procedures.

**Figure 2.2:** Both figures correspond to Problem 2, with \( v = 2 \). In the figure on the l.h.s. all procedures are plotted together with the lower bound on the Bayes procedures. In the figure on the r.h.s. the corresponding regret functions are displayed.
This resulted in the MRSMR($D_w$) procedure $\alpha_\theta$, with $\theta = 1.17$. Although no explicit formula for the envelope risk can be obtained, some results can still be computed analytically, e.g. the risk of the P-value which equals

$$R(\theta, \alpha_p) = \frac{1}{3} \exp \left\{ -\frac{1}{3} \theta \right\}.$$ 

Now, notice that in case $\nu_2(d\theta) \equiv 1$, i.e., $r = 0$, the procedure $\alpha_{\nu_2,c^*}$, obtained by applying Theorem 2.3, coincides with $\alpha_p$. Applying Theorem 2.3 with $\nu_1$ corresponding to $r = \frac{1}{2}$ provides

$$\alpha_{\nu_1,c^*}(x) = \frac{c^*}{1 F_1\left(\frac{1}{2}; 1; \frac{1}{2}x\right)},$$

with $c^* = 0.79$. Computing the posterior probability w.r.t. $\nu^1$ after assigning point mass to $\{0\}$, one gets

$$\alpha_{\nu^*,1}(x) = \frac{1}{1 + c^* 1 F_1\left(\frac{1}{2}; 1; \frac{1}{2}x\right)}$$

with $c^* = 0.58$. The posterior probability w.r.t. $\nu^2$ is given by

$$\alpha_{\nu^*,2}(x) = \frac{1}{1 + c^* \exp\left\{ \frac{1}{2}x\right\}},$$

with $c^* = 0.40$. In Figure 2.2 all proposed procedures are displayed together with the corresponding regret functions. Looking at Figure 2.2 one can see that both the pairs $\alpha_{\nu^*,1}$, $\alpha_{\text{MRSMR}}$ and $\alpha_{\nu^*,2}$, $\alpha_{\nu_1,c^*}$ are quite alike. The first pair having smallest regret near $\theta = 0$, and the second pair for large $\theta$. Similar to the situation in Problem 1, $\alpha_p$ has a relatively large regret for small $\theta$ and relatively small regret for large $\theta$. This, however, is less extreme than it was in Problem 1, and $\alpha_p$ does certainly not look unreasonable in the present situation. This is a fortunate circumstance, because many applications of $\chi^2$-tests are in this context and rely on P-values. To our taste, however, $\alpha_{\nu^*,2}$ should be preferred.

Again, it is interesting to compare the degrees of belief of the various procedures for the critical values of $x$.

\begin{center}
\begin{tabular}{l|ccc}
$x$       & 5.991 & 9.210 & 13.81 \\
\hline
$\alpha_p(x)$ & 0.050 & 0.010 & 0.001 \\
$\alpha_{\text{MRSMR}}(x)$ & 0.197 & 0.118 & 0.063 \\
$\alpha_{\nu_1,c^*}(x)$ & 0.107 & 0.028 & 0.004 \\
$\alpha_{\nu^*,1}(x)$ & 0.190 & 0.057 & 0.007 \\
$\alpha_{\nu^*,2}(x)$ & 0.111 & 0.024 & 0.002 \\
\end{tabular}
\end{center}

Table 2.2
Table 2.2 shows that the P-value, indeed, assigns smaller degree of belief to the hypothesis than the other procedures. The difference between the P-value and \( \alpha_{MRSMR}(x) \) or \( \alpha_{v_{*,1}}(x) \) is alarming. However, the difference between the P-value and \( \alpha_{v_{1,c}}(x) \) or \( \alpha_{v_{*,2}}(x) \) though considerable is not so large that we feel forced to object to the use of \( \alpha_p \).

**Problem 3** The testing problem \((H_0, A_3)\) does not directly arise from practical applications. Its relation with the previous testing problems, and its degenerate character make it interesting to study. The fact that the condition \( p_\theta(x) > 0 \), for all \( x \) and \( \theta \), is not satisfied, causes some trouble. Similarly to the previous problems the investigation is started by comparing the P-value with Bayesian procedures. This is done by comparing the P-value with the lower bound (2.2). Notice that \( p_{\theta_{ml}}(x) \) is constant, and equal to the mode size. This provides, for \( v \geq 2 \),

\[
LB(x) = \frac{1}{1 + \left(\frac{\nu-2}{x}\right)^{\frac{1}{\nu-1}} \exp\left\{\frac{1}{\nu}(x \Leftrightarrow v + 2)\right\}},
\]

with the convention for \( v = 2 \) that \( 0^0 = 1 \). Hence, \( LB(x) \sim c x^{\frac{1}{\nu-1}} \exp\{\frac{1}{\nu}x\} \rightarrow 0 \) as \( x \rightarrow \infty \), which is the same as for \( \alpha_p \). It turns out that, at least for \( v = 2 \), the P-value can be obtained as a posterior probability. To judge what value this property has, a comparative analysis is made.

The attention will now be concentrated on finding interesting competitors for the P-value within the class \( D_w \). To start with, however, the envelope risk for the class \( D_w \) will be computed. Analogously to the previous problems, first the class \( C_w = \{\alpha_\theta : \theta \in [0, \infty]\} \) is determined. Notice that Theorem 2.3 cannot be applied because it requires \( p_\theta(x) > 0 \) a.e. \((\mu)\). The adjustments that have to be made are, however, minor. If one is interested in procedures that have optimal risk at \( \theta \), then \( \alpha_\theta \) should obviously be such that \( \alpha_\theta(x) \) is as large as possible for \( x \in [0, \theta] \) and as small as possible for \( [\theta, \infty) \) with the restriction that \( \alpha_\theta \in D_w \). In the case that \( \theta \) is smaller than the median of the \( \chi^2_v \) distribution, \( \alpha_\theta \) can be obtained by formally applying Theorem 2.3. As

\[
h_{\delta_0}(x) = \begin{cases} \infty & \text{if } x \in (0, \theta), \\ \left(\frac{x}{x-\theta}\right)^{\frac{1}{\nu-1}} \exp\{\frac{1}{\nu}x\} & \text{if } x > \theta, \end{cases}
\]

one can see that truncation will always be necessary to derive \( \alpha_\theta = (ch_{\delta_0}) \wedge 1 \). In the case that \( \theta \) is larger than the median of the \( \chi^2_v \) distribution, \( \alpha_\theta \) is not uniquely determined. However, any SMR(\(D_w\)) procedure \( \alpha_\theta \), e.g. \( \alpha_\theta = \mathbb{I}_{[0, \text{median}(\chi^2_v)]}, \) will provide that \( R(\theta, \alpha_\theta) = 0 \), and hence \( R_{D_w}(\theta) = 0 \), for \( \theta \in [\text{median}(\chi^2_v), \infty) \).
Figure 2.3: Both figures correspond to Problem 3, with \( v = 2 \). In the figure on the l.h.s. all procedures are plotted together with the lower bound on the Bayes procedures. In the figure on the r.h.s. the corresponding regret functions are displayed.

For minimizing the Bayes regret, the weight function \( dv/d\theta \equiv 1 \) seems to be most convenient. Moreover, it is the natural choice for location problems. This weight function leads to the Bayes factor in favor of \( H_0 \) given by

\[
h_v(x) = \frac{v}{2x_1F_1(1; \frac{1}{v} + 1; \frac{1}{2}x)}.
\]

By applying Theorem 2.3, one obtains \( \alpha_{v,c^*} \). By computing the posterior probability and adjusting the point mass in \( \{0\} \), \( \alpha_{v,c^*} \) can be obtained.

The case that \( v = 2 \) will be worked out explicitly. Notice that the median of the \( \chi^2_2 \) distribution is \( 2 \log 2 \). If \( \theta \leq 2 \log 2 \), then the SMR(\( D_w \)) procedure \( \alpha_\theta \) is given by

\[
\alpha_\theta(x) = \begin{cases} 1 & \text{if } x \in (0, \theta) , \\ 1 + \frac{1}{2} \exp\{ \frac{1}{2} \theta \} & \text{if } x \in [\theta, \infty), \\ \\
\end{cases}
\]

and if \( \theta > 2 \log 2 \), then \( \alpha_\theta = \mathbb{1}_{[0, 2 \log 2]} \). This provides that the envelope risk is equal to

\[
R_{D_w}(\theta) = \begin{cases} 1 + \exp\{ \frac{1}{2} \theta \} + \frac{1}{2} \exp\{ \theta \} & \text{if } \theta \leq 2 \log 2 , \\ 0 & \text{if } \theta > 2 \log 2 . \\ \\
\end{cases}
\]

The MRSMR(\( D_w \)) procedure is \( \alpha_\theta \), with \( \theta = 0.65 \). Notice that \( \alpha_{\text{MRSMR}}(x) \not\to 0 \) as \( x \to \infty \). This degenerate behavior of the \( \alpha_{\text{MRSMR}} \) procedure will, however,
only occur for \( v = 2 \). The regret function of \( \alpha_{\text{MRSMR}} \) is given by

\[
S(\theta, \alpha_{\text{MRSMR}}) = \begin{cases} 
0.35 \exp\{\frac{1}{2} \theta\} \Leftrightarrow 0.25 \exp\{\theta\} & \text{if } \theta \leq 0.65, \\
\exp\{\frac{1}{2} \theta\} \Leftrightarrow 0.25 \exp\{\theta\} \Leftrightarrow 0.91 & \text{if } 0.65 < \theta \leq 1.39, \\
0.09 & \text{if } \theta > 1.39.
\end{cases}
\]

By taking \( \nu(d\theta) \equiv 1 \) and assigning point mass 1 to \( \{0\} \), one obtains that the posterior probability \( \alpha_{\nu,c} \) coincides with \( \alpha_p \). As its risk \( R(\theta, \alpha_p) = \frac{1}{\theta} \exp\{\theta\} \) is integrable w.r.t. this prior measure, \( \alpha_p \) will be admissible, in contrast with the previous problems. Its regret function is given by

\[
S(\theta, \alpha_p) = \begin{cases} 
\frac{1}{3} \exp\{\theta\} \Leftrightarrow \frac{1}{3} \exp\{\theta\} + \exp\{\frac{1}{2} \theta\} \Leftrightarrow 1 & \text{if } \theta \leq 2 \log 2, \\
\frac{1}{3} \exp\{\theta\} & \text{if } \theta > 2 \log 2.
\end{cases}
\]

Minimization of the integrated risk by applying Theorem 2.3 provides,

\[
\alpha_{\nu,c^*}(x) = \left(\frac{c^*}{\exp\{\frac{1}{2} x\} \Leftrightarrow 1}\right) \wedge 1,
\]

with \( c^* = 0.40 \). In Figure 2.3 all procedures and their corresponding regret functions are displayed. In Figure 2.3 it can be seen that the degenerate character of \( \alpha_{\text{MRSMR}} \) causes that its regret will not tend to 0 as \( \theta \to \infty \). A more interesting observation is that, in contrast to the previous problems, \( \alpha_p \) has a relatively small regret for small values of \( \theta \) and a relatively large regret for large values of \( \theta \). In this example, \( \alpha_p \) can certainly be considered as a good choice.

Again, it is interesting to compare the degrees of belief of the various procedures for the critical values of \( x \).

<table>
<thead>
<tr>
<th>( x )</th>
<th>5.991</th>
<th>9.210</th>
<th>13.81</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha_p(x) )</td>
<td>0.050</td>
<td>0.010</td>
<td>0.001</td>
</tr>
<tr>
<td>( \alpha_{\text{MRSMR}}(x) )</td>
<td>0.308</td>
<td>0.308</td>
<td>0.308</td>
</tr>
<tr>
<td>( \alpha_{\nu,c^*}(x) )</td>
<td>0.022</td>
<td>0.004</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Table 2.3

It makes no sense to compare the P-value with \( \alpha_{\text{MRSMR}}(x) \), because this procedure is degenerate. A surprising feature of Table 2.3, however, is that the P-value assigns larger degree of beliefs to the hypothesis than the other procedure do.

Suggestions for \( v \neq 2 \) For all three problems, competitors for the P-value were proposed. The case \( v = 2 \), was worked out explicitly and a comparative
analysis on the basis of the regret functions was given. In the first problem
the P-value seemed unreasonable whereas it was reasonable in Problem 2, and
good in Problem 3. Although all procedures and corresponding regret functions
will change if $v$ changes, it is to be expected that the qualitative features of the
problems will remain the same; the case $v = 2$ is representative for general $v$.
If $v$ is large, then under proper rescaling of the observation and the parameter,
\[ \begin{align*}
\text{Problem 1: } & \quad \tilde{\theta} = \sqrt{\frac{1}{2} v} \log(\theta) \Leftrightarrow \log(v), \\
\text{Problem 2: } & \quad \tilde{\theta} = \sqrt{\frac{1}{2} v} \theta \Leftrightarrow 1, \\
\text{Problem 3: } & \quad \tilde{\theta} = \sqrt{\frac{1}{2} v} \theta \Leftrightarrow 1,
\end{align*} \]
the problems will be close to the problem: $\mathcal{L} = \mathcal{N}(\tilde{\theta}, 1)$, $H_0 : \tilde{t} = 0$ and $A : \tilde{t} > 0$. Tools to show whether, indeed, the procedures and corresponding
regret functions converge to those of the limiting problem, as $v \to \infty$, can be
found in Snijders [84].

The same program of comparative analyses will be carried out for this lim-
iting problem. The lower bound on posterior probabilities based on priors sat-
sifying $\nu(\{\theta_0\}) = \nu(\Theta_A) = \frac{1}{2}$, is given by
\[ LB(x) = \begin{cases} 
\frac{1}{2} & \text{if } x \leq 0, \\
\frac{1}{1+\exp\left\{ \frac{1}{2} x \right\}} & \text{if } x > 0.
\end{cases} \]
Applying Theorem 2.3, with $\nu = \delta_\theta$, provides the SMR($D_w$) procedures
\[ \alpha_\theta(x) = \left( c(\theta) \exp\left\{ \frac{1}{2} \theta^2 \right\} \exp\{ \Leftrightarrow x \} \right) \wedge 1. \]
The envelope risk $R_{D_w}$ has to be be computed numerically. The MRSMR($D_w$)
procedure turns out to be $\alpha_\theta$, with $\theta = 0.42$, i.e.
\[ \alpha_{\text{MRSMR}}(x) = (0.464 \exp\{ \Leftrightarrow 0.42 x \}) \wedge 1. \]
Taking $dv/d\theta \equiv 1$, which is the natural weight function for shift alternatives,
provides that the Bayes factor in favor of $H_0$ is given by
\[ h_\nu(x) = \frac{\phi(x)}{\Phi(x)}, \]
where $\phi$ and $\Phi$ denote respectively the density and the distribution function
of the standard normal distribution. The P-value is in this case $\alpha_p(x) = 1 \Leftrightarrow \Phi(x)$. 

Figure 2.4: Both figures correspond to the limiting testing problem. In the figure on the l.h.s. all procedures are plotted together with the lower bound on the Bayes procedures. In the figure on the r.h.s. the corresponding regret functions are displayed.

Theorem 2.3 provides

\[ \alpha_{\nu, \nu}(x) = \left( c^* \frac{\phi(x)}{\Phi(x)} \right) \wedge 1, \]

with \( c^* = 0.5825 \). Computing a weakly unbiased posterior probability gives

\[ \alpha_{\nu, \nu}^*(x) = \frac{1}{1 + c^* \frac{\phi(x)}{\Phi(x)}}, \]

with \( c^* = 0.6943 \).

Looking at Figure 2.2, one could say on the basis of the regret functions, that the situation in the limiting problem resembles the situation of Problem 2. Hence, the same conclusions can be drawn w.r.t. the various principles that lead to the procedures. It can be noticed that \( \alpha_{\nu, \nu}^* \) and \( \alpha_p \) are quite similar. Though the P-value is certainly not unreasonable for this problem, our preference goes to \( \alpha_{\nu, \nu}^*(x) \). For further discussion on this problem, see Schaafsma et al. [80].

<table>
<thead>
<tr>
<th>( x )</th>
<th>( \alpha_p(x) )</th>
<th>( \alpha_{\text{MRSRM}}(x) )</th>
<th>( \alpha_{\nu, \nu}(x) )</th>
<th>( \alpha_{\nu, \nu}^*(x) )</th>
<th>( \alpha_{\nu, \nu}^*(x) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.645</td>
<td>0.050</td>
<td>0.233</td>
<td>0.063</td>
<td>0.135</td>
<td></td>
</tr>
<tr>
<td>2.326</td>
<td>0.010</td>
<td>0.175</td>
<td>0.016</td>
<td>0.037</td>
<td></td>
</tr>
<tr>
<td>3.090</td>
<td>0.001</td>
<td>0.127</td>
<td>0.002</td>
<td>0.005</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.4
In Table 2.4 the degrees of belief assigned to the hypothesis by the various procedures are displayed for the critical values of \( x \).

**Recapitulation**  In Section 2.3 it was shown that \( P \)-values appear, if one accepts the requirement of strong unbiasedness. In the Neyman–Pearson theory they appear as the smallest significance level at which the hypothesis is still rejected. Both ways of supporting the \( P \)-value are questionable. Pearson regarded \( P \) as a fairly reasonable criterion for the probability that \( H_0 \) is true. In this section alternative criteria \( Q = \psi(P) \) have been studied, that suggest, that for small \( P \), considerable modification of \( P \) may be necessary.

The Problems 1, 2 and 3, for fixed \( v \), are displaying an increase of complexity (see the beginning of this section). This is expressed in an increase of the maximum shortcoming of the MRSMR(\( D_w \)) procedure which increases from 0.014 (Figure 2.1) via 0.026 (Figure 2.2) to 0.08 (Figure 2.3). The maximum shortcoming of the \( P \)-value is achieved at \( \theta_0 \) and is the same in all three problems (\( \frac{1}{17} \)). In our opinion \( \alpha_{\nu,*} \) is the most appealing procedure (see the of conclusion Problem 1), and we suggest to use \( Q = \alpha_{\nu,*} (x) \) as a criterion for the probability that \( H_0 \) is true.