5

Neyman–Pearson interpretation of the fiducial argument

In the previous chapter the extension of Fisher’s fiducial argument via structured models was reviewed. Using this theory, fiducial distributions could be obtained for more general problems than by using the original formulation of the fiducial argument. A fiducial or structural distribution is a probability distribution on the parameter space. In the case that $\Theta \subset \mathbb{R}$, such an epistemic probability distribution $Q$ can be regarded as a probabilistically coherent collection of degrees of belief $q_{a,b}$ for hypotheses of the form $H_{a,b}: a < t \leq b$. In this chapter, this idea will be elaborated upon to construct distributional inferences. In this respect, a particular Neyman–Pearson interpretation of the fiducial argument will be the main tool. The relation between the fiducial distribution and a family of one-sided testing problems will play a central role in this theory. The decision theoretic properties that were derived in Chapter 3 for estimators $\alpha: \mathcal{X} \mapsto [0,1]$ of $\mathbb{I}_{a}(t)$, will be extended to decision theoretic properties of procedures $Q: \mathcal{X} \mapsto \Psi$ for making distributional inferences about $\psi(t)$.  

85
5.1 Introduction

The theory of statistical inference deals with the question of how to use the available data $x$ to discuss the unknown of interest. In Chapters 2 and 3 this unknown had the form of an indicator function on the parameter space, i.e., $I_{\Theta_n}(t)$. In Chapter 4 a more general function of the unknown parameter was discussed, i.e., $\psi(t)$ not necessarily real-valued. The scope of this chapter is the same as that of Chapter 4, with the difference that the attention will be restricted to situations where $\psi(t) \in \mathbb{R}$. Some of the definitions and results for this case can be extended to a more general context. The total ordering of the real-line, and hence that of $\Psi \subset \mathbb{R}$, will be essential for our interpretation of the fiducial argument, which will be given in Section 5.2. The notation, and the regularity conditions on $P$, $\psi$, and $\gamma$, will be the same as those used in Chapter 3.

Definition 5.1 A procedure $Q$ for making distributional inferences about $\psi(t) \in \Psi \subset \mathbb{R}$, on the basis of $x \in X$, is a mapping $Q : X \mapsto \Psi^*$, where $\Psi^*$ denotes the space of all probability distributions defined on $\Psi$. A procedure $Q$ is completely determined by the corresponding family of probability distributions $\mathcal{G} = \{G_x : x \in X\}$, where

$$G_x(z) = Q(x, (\leftrightarrow, z)).$$

The notation $Q(x)$ will be used to denote the distributional inference based on $x$, and $Q(x, B)$ to denote the epistemic probability of $B \in \sigma(\Psi)$ under the probability measure $Q(x)$. Hence, $Q(\cdot, \cdot)$ is a Markov kernel (recall the notation used in Section 3.1). A procedure $Q$ has the same mathematical form as a behavioral randomized rule for estimating $\psi(t)$. Their interpretation but mainly their evaluation is, however, different. Whereas in the theory of point estimation randomized rules are a mathematical artifact, which are useful for proving theorems, but should not be used in practice because they are suboptimal if evaluated as a point estimator, procedures for making distributional inference are designed such that they express the statistical uncertainty in their dispersion. Hence, it follows that they should be evaluated differently. This topic will be touched upon in Section 5.3.

Distributional inference has some of its roots in Bayes’s famous scholium published posthumously by Price in 1763, and brought to the attention of modern generations by Barnard [2]. The essence of the approach presented in this chapter to distributional inference is given by the following line of arguments. For any pair of real-valued constants $a \leq b$, the degree of belief in the hypothesis $H_{a,b} : a < \psi(t) \leq b$ can be expressed by a number $q_{a,b} \in [0,1]$. The problem
of how to assign such a number \( q_{a,b} \) was discussed in Chapter 3. If, in addition, the requirement is made that all \( \{ q_{a,b} : a \leq b \in \mathbb{R} \} \) have to be probabilistically coherent, then they correspond to some probability measure \( Q \) on \( \Psi \), i.e., \( q_{a,b} = Q((a, b]) \). This requirement of coherence seems reasonable if one is not interested in a particular pair \((a, b)\). It is important, however, to keep in mind that the \( q_{a,b} \) are epistemic probabilities, not to be confused with frequency-theoretic probabilities. As the construction of such epistemic probabilities uses contextual ingredients (think of prior distributions, loss functions, etc.), their probabilistic coherence is neither natural nor logical; it has to be agreed upon or imposed. In the Bayesian literature, a lot of attention is paid to the requirement of probabilistic coherence. A mathematical foundation, based on axioms of rational behavior, can be found in, e.g., Von Neumann–Morgenstern [65], or Ferguson [36]. The axioms of rational behavior, however, refer to the decision making of a single person; they prescribe how a rational person should process his opinions to arrive at optimal decisions. Hence, their applicability in the context of an inter-subjective theory of statistical inference is questionable. On the other hand, allowing deviations from coherency is like opening a Pandora Box. Anyway, if no particular pair \((a, b)\) is of special interest, then the coherency of the \( q_{a,b} \) may be required, and this is done explicitly if a distributional inference about \( \psi(t) \) is required.

As it is assumed that \( \Psi \subset \mathbb{R} \), the procedure \( Q \) is completely determined by the family of corresponding distribution functions \( \mathcal{S} = \{ G_{x} : x \in \mathbb{X} \} \). Thinking in terms of such a family of distribution functions resembles Neyman’s thinking about confidence regions in terms of a family of sets \( \mathcal{A}_{\epsilon} = \{ A_{x,\epsilon} : x \in \mathbb{X} \} \), where \( A_{x,\epsilon} \subset \Psi \), and \( \epsilon \in [0,1] \). Analogous to the theory of confidence regions, the decision theoretic properties of the procedure \( Q \) will be defined via its family of distribution functions \( \mathcal{S} \). Notice that, in some sense, making distributional inference can be regarded as giving confidence statements of all level simultaneously, i.e., the equitailed \( 1 \Leftrightarrow \epsilon \) confidence region generated by \( Q \) is given by

\[
A_{x,\epsilon} = (G_{x}^{-1}(\frac{\epsilon}{2}), G_{x}^{-1}(1 \Leftrightarrow \frac{1}{2}\epsilon)) \quad \forall \ \epsilon \in [0,1].
\]

Notice, however, that these confidence regions need not necessarily have the right coverage probability.

The usual approach to the making of distributional inferences is by applying Bayes’s Theorem. The major criticism on this Bayesian approach is that it requires the specification of a prior measure \( \nu \). This criticism applies especially in the situation that is considered, namely that where practically no reliable information is available a priori for choosing \( \nu \). Modern Bayesian theory deals
with the question of how one might nevertheless choose \( \nu \) in this situation; for an extensive review on this topic see, e.g. Berger [9], or Kass–Wasserman [52]. Fisher rejected this Bayesian approach to distributional inference. His fiducial argument was in essence an attempt to replace the Bayesian paradigm by a method to generate distributional inferences without specifying a prior. It is a deplorable fact that differences of opinion about the Behrens–Fisher problem caused some alienation between Fisher and Neyman. Adherents of the Neyman–Pearson–Wald school started to believe that only the making of distributional inferences is too ambitious and that confidence intervals are appropriate because they allow a direct verification and evaluation. The theory of this chapter provides tools to repair this situation.

5.2 The Neyman–Pearson interpretation

In Chapter 4 it was shown that the extension of the fiducial argument via structured models gives an elegant paradigm for generating distributional inferences. The idea behind these structured models is that by adding some structure to the model (the structural equation) no input of additional ingredients (priors, loss functions, etc.) are needed to generate distributional inferences. Many people have criticized the fiducial, structural, or pivotal methods, because they provide solutions to problems which, in principle, do not allow a unique solution. Prior distributions, loss functions, etc., should be allowed if they are realistic. In this sense fiducial, structural, pivotal methods are somewhat dogmatic. Perhaps they correspond to optimal solutions given certain prior distributions, or loss functions. The specification of these will display something of the real nature behind the methods indicated. Therefore an alternative extension of the fiducial argument will be given, aimed at deriving the decision-theoretic properties of fiducial procedures.

Suppose that a distributional inference \( Q(x) \) has to be given about \( \psi(t) \), in the absence of any prior information about \( t \). Recall from Section 5.1 that any \( Q(x) \) can be identified with its distribution function \( G_x \). Such a distribution function \( G_x \) should, of course, be such that \( G_x(z) \) is an assessment of the ‘probability’ that the hypothesis \( H_z : \psi(t) \leq z \) is true, given that \( X = x \). In Chapter 3 it was investigated how such a ‘probability’ should be assessed in the light of the data, if \( H_z \) is tested against its logical negation \( A_z : \psi(t) > z \). Taking \( \alpha_z : \mathcal{X} \mapsto [0, 1] \) as degree of belief in \( H_z \), the essence of the fiducial argument is to take

\[
G_x(z) = \alpha_z(x).
\]
This means that the distribution function $G_x$ can hopefully be constructed by finding a suitably cohering family $\{\alpha_z : z \in \Psi\}$ of estimators of $\mathbb{E}_{H_z}(t)$, for all problems $\{(H_z, A_z) : z \in \Psi\}$. It is important to realize that first one has to verify that, for each fixed $x \in X$, $\alpha_z(x)$ is an nondecreasing right-continuous function of $z$ such that $\lim_{z \to -\infty} = 0$ and $\lim_{z \to \infty} = 1$. The identification of $G_x(z)$ with $\alpha_z(x)$ is natural, in our opinion, and has a direct connection with the way confidence intervals are usually constructed. Nevertheless, we did not find many suggestions of this kind in the literature referring to the making of distributional inferences. Fisher regarded the P-value as an intrinsic measure of evidence; he wrote that:

'It may be noticed (...) in (...) cases in which a probability distribution can be calculated by the fiducial argument, that the region containing, for example, the lowest 1% of the frequency distribution $\{x : F_\theta(x) \leq 1\%\}$, author’s interpretation) is exactly that comprising values of the parameter which would have been rejected as too low by a valid test of significance at the 1% level (i.e. $\{\theta : 1 \Leftrightarrow \alpha_\theta(x) \leq 0.01\}$, author’s interpretation). The two concepts should none the less be distinguished, for valid tests of significance at all levels may exist without the possibility of deducing by an accurate argument, a probability distribution for the unknown parameter. The direct step from the test of significance to a probability distribution cannot be sustained, and this circumstance has been responsible for some misunderstanding, and conflict of terminology.'

Notice that Fisher wrote about testing $H_\theta : t = \theta$ against $A_\theta : t < \theta$; he rejects for small values of $x$ and obtains the P-value $\alpha_\theta(x) = P(X_\theta \leq x) = F_\theta(x)$ in favor of $H_\theta$ when tested against $A_\theta$. In the canonical situation described in Section 4.1 by conditions (i)–(iv), choosing the P-value as degree of belief in $H_\theta : t \leq \theta$ provides that $G_x(\theta) = 1 \Leftrightarrow F_\theta(x)$, and hence $g_x(\theta) = \alpha_\theta(x)/d\theta$.

**Definition 5.2** Let $H_z : \psi(t) \leq z$, $A_z : \psi(t) > z$, for all $z \in \Psi$, and suppose that $\{\phi_{x, \epsilon} : \epsilon \in E \subset [0, 1]\}$ are families of nonrandomized UMPU tests satisfying the conditions of Lemma 3.7. A procedure $Q : X \mapsto \Psi^*$ will be called fiducial if

$$G_x(z) = \alpha_{sp, z}(x) \quad \forall \ z \in \Psi \quad \forall \ x \in X,$$

where $\alpha_{sp, z}(x)$ is the (symmetrized) P-value given by (3.5).

Notice that in this definition, by starting with a procedure $Q : X \mapsto \Psi^*$ it is assumed implicitly that $\alpha_z(x)$ is a probability distribution function, for every
fixed $x$. The adjective *fiducial* refers to the interpretation of the fiducial argument in the narrow sense, that is identified with the idea of equating $G_x(z)$ with a particular estimator of $I_{\Theta_R}(t)$, namely the (symmetrized) $P$-value w.r.t. a family of nonrandomized UMPU tests.

Recall the situation described at the end of Section 3.5, where it was shown in Theorem 3.5 that under certain conditions the $P$-value coincides with the posterior probability of the hypothesis. Extending this result to the context of making distributional inference about a real-valued parameter provides that sometimes the fiducial distribution coincides with a posterior distribution, i.e., under the conditions of Theorem 3.5 it follows that

$$G_{\text{fiducial},x}(z) = \alpha_{sp,z}(x) = \alpha_{\nu,z}(x) = G_{\nu,x}(z),$$

and hence that $Q_{\text{Fiducial}}$ and $Q_{\nu}$ coincide. This corresponds to the results that can be found in, e.g., Lindley [59] and Fraser [38], which indicate that in group models fiducial and posterior distributions w.r.t. right Haar measure coincide.

### 5.3 Evaluating distributional inferences

In Section 3.2 loss functions were given for evaluating estimators of $I_{\Theta_R}(t)$. These loss functions will be taken as the starting point for developing loss functions for evaluating distributional inferences about the true value $\psi(t)$ of real-valued parameter functions $\psi : \Theta \mapsto \Psi \subset \mathbb{R}$. A loss function measures how close the inference is to the true value of the unknown of interest. It can be represented as a mapping $L : \Theta \times \Psi^* \mapsto [0, \infty]$. Notice that the convention that the loss cannot be negative is followed; it has the interpretation that the loss is the penalty for not assessing $\psi(t)$ correctly. Hence, the loss of a correct inference will taken to be zero, i.e.,

$$L(\theta, \delta_{\psi(\theta)}) = 0 \quad \forall \theta \in \Theta. \quad (5.1)$$

A first idea might be to construct loss functions by taking a loss function to evaluate point estimators as a starting point, and then integrate this loss w.r.t. $Q(x)$. This would correspond to evaluating the distributional inference as if it was a behavioral randomized point estimator. However, if a convex loss functions is taken, then it follows by Jensen’s that by concentrating all probability mass in a single point, i.e., the expectation of the distributional inference, the loss will be decreased. Hence, such loss functions give an incentive to suppress the statistical uncertainty. As it should be the objective is to express the statistical uncertainty as accurately as possible by a distributional inference, different
loss functions will be needed. Recall from Section 3.2 that the most important concept in this respect is properness.

**Definition 5.3** Let $T$ be a random variable with values in $\Theta$. A loss function $L : \Theta \times \Psi^* \mapsto [0, \infty]$ is said to be proper if $L_\psi(T) = P$ implies that

$$E L(T, P) \leq E L(T, Q) \quad \forall P, Q \in \Psi^*.$$  

It is said to be strictly proper if equality holds only if $P = Q$.

There are two things about this definition that need special attention. The first is that both sides of the inequality are allowed to be infinite. The second, and more important one, is that this definition of properness is different from the ordinary definition of properness. The definition given here is specially tailored for the situation that a parameter of interest, and a nuisance parameter can be distinguished. In the case that $\psi$ is the indicator function, the two definitions, of course, coincide. An important consequence of this different definition is that the properness according to this definition does not automatically imply that the marginal posterior distribution of the parameter of interest minimizes the Bayes risk.

When using a strictly proper loss function, the expected loss $E L(T, Q)$ can be decomposed into two parts. For any strictly proper loss function $L$, the corresponding entropy or information function is defined by

$$I(P) = E L(T, P),$$

where $L_\psi(T) = P$, and the corresponding dissimilarity coefficient between $P, Q \in \Psi^*$ by

$$K(P, Q) = E L(T, Q) \leftrightarrow I(P).$$

The entropy measures the intrinsic uncertainty of the problem, i.e., the uncertainty that remains even when one has complete knowledge of $L_\psi(T)$. In the present context of making inference about a deterministic constant, perfect knowledge would correspond to knowing $\delta_\psi(t)$, and consequently the entropy will always be zero. The dissimilarity coefficient measures the shortcoming of the inference $Q$ w.r.t. the ideal inference $P$, which is $\delta_\psi(t)$ in the present context. Only very few dissimilarity coefficients $K : \Psi^* \times \Psi^* \mapsto [0, \infty]$ can be used to generate a proper loss function by taking

$$L_\theta(Q) = K(\delta_\psi(\theta), Q) \quad \forall \theta \in \Theta \quad \forall Q \in \Psi^*.$$
It does not help very much if $K : \Psi^* \times \Psi^* \mapsto [0, \infty]$ is required to be a metric. For details see, e.g. Kroese [54]. To see how proper loss functions can be constructed, the history of proper loss functions will be briefly reviewed. The theory of proper loss functions was triggered by the meteorologist Brier [14], who suggested that weather predictions in the form of forecast probabilities $q_1, \ldots, q_m$ for $m$ categories $\theta_1, \ldots, \theta_m$, should be evaluated on the basis of the loss function

$$L(\theta_i, Q) = (1 \Leftrightarrow q_i)^2 + \sum_{j \neq i} q_j^2 \quad i = 1, \ldots, m,$$

where $q_j = Q(\{\theta_j\})$. Brier noted that this loss function is strictly proper. The corresponding entropy function $I(P) = 1 \Leftrightarrow \sum_i p_i^2$ is known as Gini’s concept of diversity. Notice that in the case of two categories, this loss function coincides with the quadratic loss function (3.4) given in Section 3.2. Another strictly proper loss function for this problem was introduced by Good [43]. He noted that the logarithmic loss function

$$L(\theta_i, Q) = \Leftrightarrow \log(q_i) \quad i = 1, \ldots, m,$$

corresponds to the Kullback–Leibler distance $K(P, Q) = \sum_i p_i \log(p_i/q_i)$ and Shannon’s information $I(P) = \Leftrightarrow \sum_i p_i \log(p_i)$. In the case of two categories this loss function is clearly a member of the class of proper and convex loss functions given by (3.1). The meteorologist Epstein [35] pointed out that these two scoring rules are not satisfactory if forecasts are made for $m$ ranked events. He proposed for the forecast problem, with $\theta_1 < \theta_2 < \cdots < \theta_m$, to use the scoring rule

$$L(\theta_i, Q) = \sum_j \Pi_{\{i, \ldots, m\}}(j) \Leftrightarrow G(j)^2 \quad i = 1, \ldots, m,$$

where $G(j) = \sum_{k=1}^j q_k$. Notice that this loss function is the squared Euclidean distance $\|\Pi_{\{i, \ldots, m\}} - G\|^2$ between the distribution functions of the inference and the ideal inference.

As $\Psi$ is a totally ordered set, and hence it seems natural to somehow incorporate this ordering in the loss function. This can be done by basing the loss functions on Epstein’s scoring rule. Using the notation from Section 3.4, where $(\psi, \gamma)$ was taken to be an equivalent parameterization of $\Psi$, such that $\psi$ is the parameter of interest, and $\gamma$ denotes the nuisance parameter, a direct extension of Epstein’s scoring rule that can be used in the present context is given by

$$L_\tau(\theta, Q) = f_\psi(\gamma(\theta)) \times \int_\psi \mathbb{E}_{(\psi(\theta), \gamma)}(z)(G(z))^2 + \mathbb{E}_{\gamma}(z)(1 \Leftrightarrow G(z))^2 \tau(dz), \quad (5.2)$$
where \( \tau \) is some \( \sigma \)-finite measure and \( f_w > 0 \). Notice that this loss function is proper, and even strictly proper if \( \tau \) assigns positive mass to each nondegenerate interval. The loss function with \( \tau = \lambda \) and \( f_w \equiv 1 \) is the distributional inference analogue of squared error loss, see Dehling et al. [27]. It is closely related to the quadratic loss function (3.4) for evaluating estimators of \( \mathbb{I}_{A}(t) \); it originally appeared by integrating the loss of \( Q \)-values if a family \( \{ (H_z, A_z) : z \in \Psi \} \) of one-sided problems was considered, and squared error loss was used. Similarly to Epstein’s scoring rule it explicitly uses the ordering of \( \Psi \subset \mathbb{R} \).

**Definition 5.4** Let \( \Psi \) be a totally an ordered set, and define, for each \( z \in \Psi \),

\[
\Psi_{1,z} = \{ \psi \in \Psi : \psi < z \}, \quad \text{and} \quad \Psi_{2,z} = \{ \psi \in \Psi : \psi \geq z \}.
\]

Accordingly, define, for each \( z \in \Psi \),

\[
\Psi^*_{1,z} = \{ Q \in \Psi^* : Q(\Psi_{1,z}) = 1 \}, \quad \text{and} \quad \Psi^*_{2,z} = \{ Q \in \Psi^* : Q(\Psi_{2,z}) = 1 \}.
\]

A loss function \( L : \Theta \times \Psi^* \mapsto [0, \infty] \) is said to take the ordering of \( \Psi \) into account if

\[
P \leq Q \Rightarrow L(\theta, P) \geq L(\theta, Q) \quad \forall P, Q \in \Psi^*_{1,z},
\]

\[
P \leq Q \Rightarrow L(\theta, P) \leq L(\theta, Q) \quad \forall P, Q \in \Psi^*_{2,z},
\]

for all \( \theta \in \Theta \), such that \( \psi(\theta) = z \), and all \( z \in \Psi \).

Notice that loss function (5.2) takes, indeed, the ordering of \( \Psi \) into account. Similar to the extension of the quadratic loss function (3.4) to a quadratic loss function (5.2) for evaluating distributional inferences about \( \psi(t) \) that takes the ordering of \( \Psi \) into account, it is possible to extend the entire class of loss functions, given by (3.1), i.e., by taking

\[
L_{\tau,f}(\theta, Q) = f_w(\gamma(\theta)) \times \int_{\Psi} \mathbb{I}_{(-\infty, \psi(\theta))}(z)f_0(G(z)) + \mathbb{I}_{[\psi(\theta), \infty)}(z)f_1(1 \Leftrightarrow G(z)) \tau(dz). \quad (5.3)
\]

The loss function \( L_{\tau,f} \) will be proper if and only if the pair \( (f_0, f_1) \) satisfies the condition (3.2). If this condition is fulfilled with a unique minimum and if \( \tau \) assigns positive mass to any nondegenerate interval, then \( L_{\tau,f} \) will be strictly proper. In Section 3.2 it was shown that (3.2) implies that both \( f_0 \) and \( f_1 \) are nondecreasing, which in its turn is a necessary and sufficient condition for \( L_{\tau,f} \) to take the ordering of \( \Psi \) into account. It is important to realize that is possible
that proper loss functions exist that cannot be represented as (5.3), consider for example loss functions that are based on the density \( g \) of \( Q \), instead of on the distribution function \( G \). The question of interest, however, is whether it is possible that these loss functions also take the ordering of \( \Psi \) into account.

**Conjecture 5.1** Every proper loss function \( L : \Theta \times \Psi^* \mapsto [0, \infty] \) that takes the ordering of \( \Psi \) into account, can be represented as

\[
f_w(\gamma(\theta)) \int_{\Psi} \Pi_{(-\infty, \psi(\theta))}(z) f_{0, z}(G(z)) + \Pi_{[\psi(\theta), \infty]}(z) f_{1, z}(1 \Leftrightarrow G(z)) \tau(dz), \tag{5.4}
\]

where each pair \((f_{0, z}, f_{1, z})\) satisfies the requirement (3.2).

**Idea of the proof.** Let \( \mathcal{L} \) denote the space of all loss functions \( L : \Theta \times \Psi^* \mapsto [0, \infty] \). Notice that the collection of all loss functions that take the ordering of \( \Psi \) into account, constitutes a convex cone \( \mathcal{C} \subset \mathcal{L} \). Now, define

\[
L_{0, e, z}(\theta, Q) = f_w(\gamma(\theta)) \int_{[e, 1]} (G(z)) \Pi_{(-\infty, \psi(\theta))}(z) \\
L_{1, e, z}(\theta, Q) = f_w(\gamma(\theta)) \int_{[e, 1]} (1 \Leftrightarrow G(z)) \Pi_{[\psi(\theta), \infty]}(z),
\]

for all \( e \in [0, 1] \), and all \( z \in \Psi \). It is easy to check that \( L_{i, e, z} \in \mathcal{C}, \ i = 0, 1 \). If it can be shown that these loss functions are all the extremals of \( \mathcal{C} \), and that \( \mathcal{C} \) can be suitably metrized, then every \( L \in \mathcal{C} \) can be represented as

\[
L(\theta, Q) = \int_{[0, 1] \times \Psi} L_{0, e, z}(\theta, Q) \tau(de, dz) + \int_{[0, 1] \times \Psi} L_{1, e, z}(\theta, Q) \tau(de, dz),
\]

where \( \tau \) is some \( \sigma \)-finite measure on \([0, 1] \times \Psi\). By integrating out \( e \), one obtains that \( L(\theta, Q) \) becomes (5.4), with all the \( f_{0, z} \) and \( f_{1, z} \) nondecreasing. The requirement of properness in its turn, implies that each pair \((f_{0, z}, f_{1, z})\) has to satisfy the requirement (3.2).

By taking \((f_{0, z}, f_{1, z})\), for each \( z \in \Psi \), the same pair \((f_0, f_1)\), the class (5.3) is recovered. As it is also possible to adjust \( \tau \), practically nothing can be gained by taking the \((f_{0, z}, f_{1, z})\) to be different for different values of \( z \). Hence, if Conjecture 5.1 holds, then it would be safe to say that the class (5.3) contains about all proper loss functions of interest.

### 5.4 Unbiasedness

Using the loss functions defined in the previous section, the concepts of unbiasedness and corresponding results from Section 3.3 can be extended such that
they apply to procedures for making distributional inferences about $\psi(t)$. To
start with, however, notice that, in contrast with the situation in Chapter 3, it is
possible to use the concept of mean unbiasedness, commonly used in the theory
of point estimation. It can be applied to procedures for making distributional
inference, provided that the following straightforward adjustment is made.

**Definition 5.5** A procedure $Q : \mathcal{X} \mapsto \Psi^*$ is said to be mean unbiased if

$$\mathbb{E}_{\mathcal{X}} z Q(\theta, dz) = \psi(\theta) \quad \forall \theta \in \Theta,$$

or equivalently if its mean is a mean–unbiased point estimator of $\psi$.

This form of unbiasedness, however, depends heavily on the parameterization
of the family $\mathcal{P}$; in the case that there exists such mean–unbiased procedure for
making inference about $\psi(t)$, it is not so difficult to see that after a continuous
monotone transformation of $\Psi$ the procedure induced by this transformation
will no longer be mean unbiased, unless the transformation was affine linear.
As continuous monotone transformations do not change the ordering of $\Psi$, the
inference problem is not changed essentially if loss functions and prior distri-
butions are transformed appropriately. This means that the concept of mean
unbiasedness is not necessarily appropriate for evaluating distributional infer-
ences.

**Definition 5.6** Let $L_{\tau,f}$ be a loss function of the form (5.3). A procedure $Q : \mathcal{X} \mapsto \Psi^*$ is said to be risk or Lehmann unbiased (w.r.t. $L_{\tau,f}$) if

$$\mathbb{E}_{\mathcal{X}} (\vartheta, Q(\theta)) \geq \mathbb{E}_{\mathcal{X}} (\theta, Q(\theta)) \quad \forall \vartheta, \theta \in \Theta \text{ s.t. } \gamma(\vartheta) = \gamma(\theta).$$

Notice that this requirement of risk unbiasedness is somewhat less restrictive
than the usual requirement of risk unbiasedness where the inequality has to
hold for all $\vartheta, \theta \in \Theta$, instead of only for those that satisfy $\gamma(\vartheta) = \gamma(\theta)$. When
considering loss functions of the form (5.3), however, this definition of risk un-
biasedness is more natural. Similarly to Section 3.3 the relation between risk
unbiasedness and strong and weak unbiasedness will be investigated.

**Definition 5.7** A procedure $Q : \mathcal{X} \mapsto \Psi^*$, characterized by the family of distribution functions $\mathcal{G} = \{G_x : x \in \mathcal{X}\}$, is said to be strongly unbiased if

$$\mathcal{L} G_{X_\theta}(z) \leq U(0,1) \quad \forall \theta \in \Theta \quad \text{if } z < \psi(\theta),$$

$$\mathcal{L} G_{X_\theta}(z) \geq U(0,1) \quad \forall \theta \in \Theta \quad \text{if } z \geq \psi(\theta).$$

It is said to be strongly similar if equality holds, for all $\theta \in \Theta$ such that $\psi(\theta) = z$. 
Strong unbiasedness together with the continuity of \( G_x \), for all \( x \in \mathcal{X} \), implies strong similarity by the continuity of the parameterization of \( \mathcal{P} \). The other way around, strong similarity implies strong unbiasedness, because distribution functions are by definition nondecreasing. The term ‘similar’ stems from the expression ‘similar to the sample space’ which as a critical region in the Neyman–Pearson setting is ideal in the sense that its probability does not depend on \( \theta \). The definition of strong similarity has a motivation in the fact that a strongly similar procedure is ‘similar to the ideal procedure in the ideal situation’. The argument is as follows. Let \( T \) be a random variable with values in \( \Psi \). The ideal situation would, of course, be that where the stochastic model gives a complete specification of \( \mathcal{P} = \mathcal{L}(T, X) \), not requiring the true value of any parameter. The ideal inference corresponding to this situation would then be \( Q^*(x) = \mathcal{L}(T \mid X = x) \). If the corresponding distribution function \( G^*_x \) is continuous, for every \( x \in \mathcal{X} \), then it follows from

\[
\mathcal{L}(G^*_x(T) \mid X = x) = \mathcal{L}(G^*_X(T) \mid X = x) = U(0,1),
\]

that \( G^*_X(T) \) and \( X \) are stochastically independent, and most importantly that \( \mathcal{L}(G^*_X(T)) = U(0,1) \). A strongly similar procedure mimics this behavior. If a family \( \{\alpha_z : z \in \Psi\} \) consists of strongly unbiased estimators of \( \pi_{\theta, z}(t) \), in the sense of Definition 3.3, then the corresponding procedure \( Q \), with \( G_x(z) = \alpha_z(x) \), will be strongly unbiased in the sense of Definition 5.7. Hence, provided that the conditions of Lemma 3.4 hold, for every \( z \in \Psi \), and that \( \alpha_z(x) \) is a distribution function, for every fixed \( x \), the fiducial argument provides strongly unbiased procedures. Notice that, if \( G_x \) is continuous and strictly increasing on \( \{z : 0 < G_x(z) < 1\} \), for all \( x \in \mathcal{X} \), then it follows from the strong similarity that, for all \( \epsilon \in [0, 1] \),

\[
P(G_{X_x}^{-1}(\frac{1}{2} + \epsilon) \leq \psi(\theta) \leq G_{X_x}^{-1}(1 \Leftrightarrow \frac{1}{2} \epsilon)) = 1 \Leftrightarrow \epsilon
\]

or, in words, that the equal–tails induced confidence limits provide a confidence interval with exact coverage probability.

**Definition 5.8** A procedure \( Q : \mathcal{X} \mapsto \Psi^* \), characterized by the family of distribution functions \( \mathcal{G} = \{G_x : x \in \mathcal{X}\} \), is said to be weakly unbiased if

\[
\mathbb{E}G_{X_x}(z) \leq \frac{1}{2} \quad \forall \; \theta \in \Theta \quad \text{if} \quad z < \psi(\theta),
\]

\[
\mathbb{E}G_{X_x}(z) \geq \frac{1}{2} \quad \forall \; \theta \in \Theta \quad \text{if} \quad z \geq \psi(\theta).
\]

It is said to be weakly similar if equality holds, for all \( \theta \in \Theta \) such that \( \psi(\theta) = z \).
5.4. Unbiasedness

Weak unbiasedness together with the continuity of $G_x$, for all $x \in X$, implies weak similarity by the continuity of the parameterization of $\mathcal{P}$. The other way around, weak similarity implies weak unbiasedness, because distribution functions are by definition nondecreasing. Strong unbiasedness implies automatically weak unbiasedness. If a family $\{\alpha_z : z \in \Psi\}$ consists of weakly unbiased estimators of $\Pi_{\alpha_{H_z}}(t)$ in the sense of Definition 3.4, and if $\alpha_z(x)$ is a distribution function, for every $x \in X$, then the corresponding procedure $Q$, with $G_x(z) = \alpha_z(x)$, will be weakly unbiased in the sense of Definition 5.8. Hence, provided that the conditions of Lemma 3.7 hold, for every $z \in \Psi$, the fiducial argument provides weakly unbiased procedures.

The relation between risk unbiasedness and weak and strong unbiasedness is given by the following theorem. A version of this theorem for the case that $\Theta = \Psi$ can be found in Solomé [73].

**Theorem 5.1** Let $Q : X \mapsto \Psi^*$ be a procedure for making distributional inference about $\psi(t)$, where $\psi : \Theta \mapsto \Psi \subset \mathbb{R}$.

(i) $Q$ is strongly unbiased if and only if $Q$ is risk unbiased w.r.t. all proper loss functions $L_{\tau, f}$ of the from (5.3), with $f_0, f_1$ bounded and $\tau$ finite.

(ii) $Q$ is weakly unbiased if and only if $Q$ is risk unbiased w.r.t. all squared error loss functions $L_{\tau}$ of the from (5.2), with $\tau$ finite.

**Proof** (i) To show that risk unbiasedness, for all proper loss functions $L_{\tau, f}$ of the form (5.3), with $f_0, f_1$ bounded and $\tau$ finite, is equivalent to the requirement of strong unbiasedness, start by writing

$$\mathbb{E}L_{\tau, f}(\theta, Q(X_\theta)) = f_\psi(\gamma(\theta)) \int_\Psi \mathbb{E}f_0(G_{X_\theta}(z)) + \mathbb{E}f_1(1 \Leftrightarrow G_{X_\theta}(z)) \tau(dz).$$

Using the characterization of $f_0$ and $f_1$ from Theorem 3.1, and denoting the distribution function of $G_{X_\theta}(z)$ by $H_{\theta, z}$, provides similar to (3.7) and (3.8) that

$$\mathbb{E}f_0(G_{X_\theta}(z)) = \int_{[0,1]} (1 \Leftrightarrow H_{\theta, z}(p)) \xi(dp),$$

$$\mathbb{E}f_1(1 \Leftrightarrow G_{X_\theta}(z)) = \int_{[0,1]} H_{\theta, z}(p)(1 \Leftrightarrow p) \xi(dp),$$

and hence

$$\mathbb{E}L_{\tau, f}(\theta, Q(X_\theta)) = f_\psi(\gamma(\theta)) \int_\Psi \int_{[0,1]} \mathbb{E}(1 \Leftrightarrow H_{\theta, z}(p)) \xi(dp) + \mathbb{E}(1 \Leftrightarrow H_{\theta, z}(p))(1 \Leftrightarrow p) \xi(p)(dz).$$
The requirement \( \mathbb{E}L_{\mathcal{I}, f}(\vartheta, Q(X_0)) \Rightarrow \mathbb{E}L_{\mathcal{I}, f}(\theta, Q(X_0)) \geq 0 \), for all \( \theta, \vartheta \in \Theta \) such that \( \gamma(\theta) = \gamma(\vartheta) \), is equivalent to

\[
\begin{align*}
&f_w(\gamma(\theta)) \int_{y} \mathbb{I}_{[\psi(\theta), \psi(\vartheta)]}(z) \int_{[0,1]}(p \Leftrightarrow H_{\theta, z}(p)) \xi(dp) \tau(dz) \geq 0 \quad \text{if} \quad \psi(\vartheta) > \psi(\theta), \\
&f_w(\gamma(\theta)) \int_{y} \mathbb{I}_{[\psi(\theta), \psi(\vartheta)]}(z) \int_{[0,1]}(H_{\theta, z}(p) \Leftrightarrow p) \xi(dp) \tau(dz) \geq 0 \quad \text{if} \quad \psi(\vartheta) < \psi(\theta),
\end{align*}
\]

for arbitrary finite measures \( \tau \) and \( \xi \). These inequalities hold for all such \( \tau \) and \( \xi \) if and only if

\[
H_{\theta, z}(p) \leq p \quad \text{if} \quad z \leq \psi(\theta) \quad \text{and} \quad H_{\theta, z}(p) \geq p \quad \text{if} \quad z \geq \psi(\theta),
\]

which by definition is strong unbiasedness.

(ii) The proof is in essence the same as that of part (i). The requirement \( \mathbb{E}L_{\mathcal{I}, f}(\vartheta, Q(X_0)) \Rightarrow \mathbb{E}L_{\mathcal{I}, f}(\theta, Q(X_0)) \geq 0 \), for all \( \theta, \vartheta \in \Theta \) s.t. \( \gamma(\theta) = \gamma(\vartheta) \), is equivalent to

\[
\begin{align*}
&f_w(\gamma(\theta)) \int_{y} \mathbb{I}_{[\psi(\theta), \psi(\vartheta)]}(z)(1 \Leftrightarrow 2H_{X_0}(z)) \tau(dz) \geq 0 \quad \text{if} \quad \psi(\vartheta) > \psi(\theta), \\
&f_w(\gamma(\theta)) \int_{y} \mathbb{I}_{[\psi(\theta), \psi(\vartheta)]}(z)(2H_{X_0}(z) \Leftrightarrow 1) \tau(dz) \geq 0 \quad \text{if} \quad \psi(\vartheta) < \psi(\theta).
\end{align*}
\]

These inequalities hold if and only if \( Q \) is weakly unbiased.

Measuring the (lack of) precision by either weak or strong unbiasedness, the bias function of a procedure \( Q : X \mapsto X^* \) can be defined by

\[
B(\theta, Q) = \mathbb{E}G_{X_0}(\psi(\theta)) \Leftrightarrow \frac{1}{2}.
\]

With this definition of the bias function it is obvious that both weakly and strongly similar procedures have a bias function that is identically equal to 0. Applying the fiducial argument will typically provide weakly or even strongly similar procedures, and in these cases \( B(\theta, Q_{\text{Fiducial}}) \equiv 0 \). That applying Bayes’s Theorem will typically lead to procedures that are not weakly similar, and hence not strongly similar can be concluded from the following theorem that can be found in Kroese [54].

**Theorem 5.2** Let \( \Theta = (\underline{\theta}, \bar{\theta}) \subset \mathbb{R} \), and let \( \nu \) be a finite measure on \( \Theta \) that has no atoms. Then the posterior distribution \( Q_{\nu} \) cannot be weakly unbiased.

This theorem corresponds to results of, e.g. Noorbaloochi-Meeden [67], or Bickel-Blackwell [12], however, the proof is based on a different argument. It can easily be shown that the theorem also holds true if \( \nu \) contains atoms in the interior of \( \Theta \). It is, however, not true that posterior distributions can never be weakly unbiased; under the conditions of Theorem 3.5, \( Q_{\text{Fiducial}} \) and \( Q_{\nu} \) coincide and are weakly unbiased or even strongly unbiased.
5.5 Optimal Procedures

Recall from Section 5.3 that if a proper loss function of the form (5.3) is used, then it does not automatically follow that the Bayes procedure minimizes the Bayes risk, provided that it is finite. However, under some extra conditions this result can still be obtained.

Lemma 5.1 Suppose that $\psi : \Theta \mapsto \Psi$ and $\gamma : \Theta \mapsto$, are logically independent, and that $\nu(d\theta) = \nu_\psi(d\psi) \times \nu_\gamma(d\gamma)$, where $\nu_\gamma(.) < \infty$. Define $\mu_B(B) = \int_{\Theta} P_\theta(B) \gamma(\theta) \nu(d\theta)$, for all $B \in \sigma(B)$, and suppose that it is \(\sigma\)-finite. Let $L_{f, \tau}$ be a proper loss function of the form (5.3). Then, provided that $r_{f, \tau, \nu}(Q, \nu) < \infty,$

$$r_{f, \tau, \nu}(Q, \nu) \leq r_{f, \tau, \nu}(Q) \quad \forall Q \in \Psi^*.$$ 

Proof. Let $Q : \mathcal{X} \mapsto \Psi^*$ be any procedure for making distributional inference about $\psi(t)$. Its Bayes risk $r_{f, \tau, \nu}(Q)$ is given by

$$\int_{\Theta} \int_{\mathcal{X}} f_w(\gamma(\theta)) \int_{\Psi} \left\{ \mathbb{P}_{[\infty, \psi(\theta)]}(z) f_0(G(z)) + \mathbb{P}_{[\psi(\theta), \infty]}(z) f_1(1 \Leftrightarrow G(z)) \right\} p_\theta(x)$$

$$\tau(dz) \mu(dx) \nu(d\theta)$$

$$= \int_{\Theta} \int_{\mathcal{X}} \int_{\Psi} f_w(\gamma) \int_{\Psi} \left\{ \mathbb{P}_{[\infty, \psi(\theta)]}(z) f_0(G(z)) + \mathbb{P}_{[\psi(\theta), \infty]}(z) f_1(1 \Leftrightarrow G(z)) \right\}$$

$$p^\nu_{U, \psi}(u) p^\nu_{V, \gamma}(v) \left[ \nu_\psi(d\psi) f_0(G(z)) + \nu_\gamma(d\gamma) \mu_U(du) \mu_V(dv) \right]$$

$$= \int_{\Psi} \int_{\mathcal{X}} \int_{\Psi} \left\{ \mathbb{P}_{[\infty, \psi(\theta)]}(z) f_0(G(z)) + \mathbb{P}_{[\psi(\theta), \infty]}(z) f_1(1 \Leftrightarrow G(z)) \right\}$$

$$\int_{\Psi} f_w(\gamma) p^\nu_{V, \gamma}(v) \left[ \nu_\psi(d\psi) f_0(G(z)) + \nu_\gamma(d\gamma) \mu_U(du) \mu_V(dv) \right] \tau(dz)$$

$$= \int_{\Psi} \int_{\mathcal{X}} \left\{ (1 \Leftrightarrow G_{\nu, \lambda}(z)) f_0(G(z)) + G_{\nu, \lambda}(z) f_1(1 \Leftrightarrow G(z)) \right\}$$

$$\int_{\Psi} p^\nu_{U, \lambda}(u) \nu_\psi(d\psi) \int_{\Psi} f_w(\gamma) p^\nu_{V, \gamma}(v) \left[ \nu_\psi(d\psi) f_0(G(z)) + \nu_\gamma(d\gamma) \mu_U(du) \mu_V(dv) \right] \tau(dz)$$

$$= \int_{\Psi} \int_{\mathcal{X}} \left\{ (1 \Leftrightarrow G_{\nu, \lambda}(z)) f_0(G(z)) + G_{\nu, \lambda}(z) f_1(1 \Leftrightarrow G(z)) \right\} \mu_B(dx) \tau(dz),$$
where

\[
G_{\nu,x}(z) = \frac{\int_{\Psi} \mathbb{I}_{(-\infty,z)}(\psi)p_{U,\psi}(u) \, \nu_{\psi}(d\psi)}{\int_{\Psi} p_{U,\psi}(u) \, \nu_{\psi}(d\psi)} = \frac{\int_{\Theta_{\Psi_{\nu}}} p_{\theta}(x) \, \nu(d\theta)}{\int_{\Theta} p_{\theta}(x) \, \nu(d\theta)}.
\]

Now, use the fact that \( f = (f_0, f_1) \) satisfy (3.2) to obtain that

\[
r_{f,\tau,\nu}(Q) = \int_{\Psi} \int_{\mathcal{X}} \left\{ (1 \Leftrightarrow G_{\nu,x}(z)) f_0(G(z)) + G_{\nu,x}(z) f_1(1 \Leftrightarrow G(z)) \right\} \mu_{\psi}(dx) \tau(dz) \leq \int_{\Psi} \int_{\mathcal{X}} \left\{ (1 \Leftrightarrow G_{\nu,x}(z)) f_0(G_{\nu,x}(z)) + G_{\nu,x}(z) f_1(1 \Leftrightarrow G_{\nu,x}(z)) \right\} \mu_{\psi}(dx) \tau(dz) = r_{f,\tau,\nu}(Q_{\nu}).
\]

As it was assumed that \( r_{f,\tau,\nu}(Q_{\nu}) < \infty \), it follows that \( Q_{\nu} \) minimizes the Bayes risk. \( \qed \)

Recall from Section 5.4 that posterior distributions \( Q_{\nu} \) will typically not be weakly, and hence not strongly unbiased. Therefore, it is interesting to see whether or not it is possible to find optimal procedures under the restriction of either weak or strong unbiasedness. The situation will be similar to that of Section 3.4, where it was shown that under the restriction of strong unbiasedness it is possible to obtain a uniformly minimum risk procedure, provided that \( \mathcal{P} \) is dominated by \( \lambda \) and has conditional monotone likelihood ratio, see e.g. Kroese et al [56], or Salomé et al [75].

To discuss optimality, consider the following line of thought. If all distributional inferences are represented by their distribution functions, then the prophet inference \( Q^* \) is represented by \( G^*(z) = \mathbb{I}_{[\psi(t),\infty)}(z) \). As it is the objective to construct \( Q(x) \) such that it is as close as possible to the prophet inference, it should be such that \( G_x(z) \) is close to 0 whenever \( z \) is smaller than
ψ(t) and, vice versa, \( G_x(z) \) should be close to 1 if \( z \) is larger than \( ψ(t) \). This is rationale behind the following ordering of the space of all procedures.

**Definition 5.9** A procedure \( Q : \mathcal{X} \mapsto ψ^* \), characterized by the family of distribution functions \( \mathcal{G} = \{ G_x : x \in \mathcal{X} \} \), is said to be better than \( \tilde{Q} : \mathcal{X} \mapsto ψ^* \), characterized by the family of distribution functions \( \mathcal{G} = \{ \tilde{G}_x : x \in \mathcal{X} \} \) (denoted by \( Q \succeq \tilde{Q} \)), if

\[
\mathcal{L}G_{X_x}(z) \leq \mathcal{L}\tilde{G}_{X_x}(z) \quad \text{if} \quad z \leq ψ(θ),
\]
\[
\mathcal{L}G_{X_x}(z) \geq \mathcal{L}\tilde{G}_{X_x}(z) \quad \text{if} \quad z \geq ψ(θ).
\]

It is said to be strictly better (denoted by \( Q \succ \tilde{Q} \)) if at least one inequality is strict.

This ordering, is a straightforward extension of the ordering that was defined in Section 3.4. It is related to the ordering defined by loss functions through the following theorem.

**Theorem 5.3** Let \( Q : \mathcal{X} \mapsto ψ^* \), characterized by the family of distribution functions \( \mathcal{G} = \{ G_x : x \in \mathcal{X} \} \), and \( \tilde{Q} : \mathcal{X} \mapsto ψ^* \), characterized by the family of distribution functions \( \mathcal{G} = \{ \tilde{G}_x : x \in \mathcal{X} \} \), be two procedures for making distributional inferences about \( ψ(t) \). Then (i) \( Q \succeq \tilde{Q} \) if and only if

\[
R_{τ, f}(θ, Q) \leq R_{τ, f}(θ, \tilde{Q}),
\]

for all proper loss functions \( L_{τ, f} \) of the form (5.3), with \( f_0, f_1 \) bounded and \( τ \) finite, and (ii) \( Q \succ \tilde{Q} \) if, in addition, strict inequality holds for the strictly proper loss functions.

**Proof.** (i) Take an arbitrary loss function \( L_{τ, f} \) of the form (5.3), with \( f_0, f_1 \) bounded and \( τ \) finite, and write (by using Fubini) the inequality \( R_{τ, f}(θ, Q) \leq R_{τ, f}(θ, \tilde{Q}) \) as follows

\[
f_w(γ(θ)) \int_Ψ \mathbb{I}_{(-∞, ψ(θ))}(z) \left( \mathbb{E}f_0(G_{X_x}(z)) \Leftrightarrow \mathbb{E}f_0(\tilde{G}_{X_x}(z)) \right) + \mathbb{I}_{[ψ(θ), ∞)}(z) \left( \mathbb{E}f_1(1 \Leftrightarrow G_{X_x}(z)) \Leftrightarrow \mathbb{E}f_1(1 \Leftrightarrow \tilde{G}_{X_x}(z)) \right) τ(dz) \leq 0.
\]

If this inequality is required to hold for all finite measures \( τ \), then this is equivalent to requiring that it must hold pointwisely, i.e.,

\[
\mathbb{E}f_0(G_{X_x}(z)) \leq \mathbb{E}f_0(\tilde{G}_{X_x}(z)) \quad \text{if} \quad z \leq ψ(θ),
\]
\[
\mathbb{E}f_1(1 \Leftrightarrow G_{X_x}(z)) \leq \mathbb{E}f_1(1 \Leftrightarrow \tilde{G}_{X_x}(z)) \quad \text{if} \quad z \geq ψ(θ).
\]
Next, use the proof of Lemma 3.8 to show that the requirement that these inequalities should hold for all \( L_{\tau,f} \) of the form (5.3), with \( f_0, f_1 \) bounded and \( \tau \) finite, is equivalent to the statement that

\[
\mathcal{L} \bar{G}_{X_\delta}(z) \leq \mathcal{L} G_{X_\delta}(z) \quad \text{if} \quad z \leq \psi(\theta),
\]

\[
\mathcal{L} G_{X_\delta}(z) \geq \mathcal{L} \bar{G}_{X_\delta}(z) \quad \text{if} \quad z \geq \psi(\theta),
\]

and all \( \theta \in \Theta \). (ii) The proof is the same as that of part (i) except that strict properness implies that both \( f_0 \) and \( f_1 \) are strictly increasing, and hence that \( \xi \) and \( \tau \) assign positive mass to any nondegenerate interval. This changes all \( \leq \) to \( < \).

The specification of a loss function does not necessarily comply with unbiasedness restrictions. One reason is that a loss function is based on information which is more or less alien to the general–purpose ideas behind the requirement of unbiasedness. As a consequence, unbiased procedures may easily be inadmissible if a specific loss function is considered. For this reason the optimality of procedures is discussed in terms of the ordering defined by Definition 5.9, and can then by Theorem 5.3 be translated to optimality properties in terms of loss functions.

**Theorem 5.4** Let \( \mathcal{P} \) be dominated by \( \mu \), and assume that \( d\mathcal{P}_\theta/d\mu(x) > 0 \), for all \( \theta \in \Theta \) and \( x \in \mathcal{X} \), and suppose that \( \{ \phi_{z,\varepsilon} : \varepsilon \in E_z \} \) is the family of all nonrandomized size-\( \varepsilon \) tests for testing \( H_z : \psi(t) \leq z \) against \( A_z : \psi(t) > z \) that are most powerful on \( \Theta_{A_z} \) and least powerful on \( \Theta_{H_z} \) among all unbiased size-\( \varepsilon \) tests, where \( E_z \subset [0,1] \) is the collection of all levels for which such nonrandomized test exist. Suppose that, for each fixed \( x \in \mathcal{X} \), the symmetrized \( P \)-value \( \alpha_{sp,z}(x) \) is a probability distribution function on \( \Psi \), and denote the fiducial procedure by \( Q_{\text{Fiducial}} \). Then

(i) there does not exist a procedure \( Q \) such that \( Q > Q_{\text{Fiducial}} \),

(ii) if, in addition, \( E_z = [0,1] \) for all \( z \in \Psi \), then \( Q_{\text{Fiducial}} \succeq Q \), for any other \( Q \in \mathcal{D}_s \).

**Proof.** The proof trivially follows from results that can be found in Chapter 3. Let \( Q \) be an arbitrary procedure characterized by \( \mathcal{G} = \{ G_x : x \in \mathcal{X} \} \). Fix some \( z \in \Psi \), and define the estimator \( \alpha_z(x) : \mathcal{X} \mapsto [0,1] \) of \( \mathbb{I}_{O_{H_z}}(t) \), by equating \( \alpha_z(x) = G_x(z) \). Notice that by definition \( G_{\text{Fiducial},z}(z) = \alpha_{sp,z}(x) \). Now, apply Theorem 3.3 to obtain the desired result.

Notice that the fact that there does not exist a procedure that is better than the fiducial procedure does not imply that it is admissible w.r.t. a particular loss
function. It only tells that there cannot exist a procedure that uniformly improves the risk of the fiducial procedure w.r.t. all loss functions; it can happen that for different proper loss functions there exist different procedures that uniformly improve the risk of the fiducial procedure. In the case that $Q_{\text{Fiducial}}$ has uniformly minimum risk within $D_s$, two questions arise. The first is whether the best strongly similar procedure is unique, the second whether monotone likelihood ratio is really needed to establish the optimality. These questions are successively dealt with in the following examples.

Example 1 To establish that best strongly similar procedures need not be unique, consider the following situation. Let $X_\theta$ be a random variable with density

$$f_\theta(x) = \begin{cases} 
2(1 \Leftrightarrow \theta) & \text{if } 0 \leq x \leq \frac{1}{2}, \\
2\theta & \text{if } \frac{1}{2} < x \leq 1.
\end{cases}$$

A best strongly similar procedure exists and is given by

$$G_{\text{Fiducial},x}(\theta) = \begin{cases} 
1 \Leftrightarrow 2x(1 \Leftrightarrow \theta) & \text{if } 0 \leq x \leq \frac{1}{2}, \\
2(1 \Leftrightarrow x)\theta & \text{if } \frac{1}{2} < x \leq 1.
\end{cases}$$

Define $\pi_1 : [0, \frac{1}{2}] \mapsto [0, \frac{1}{2}]$ and $\pi_2 : (\frac{1}{2}, 1] \mapsto (\frac{1}{2}, 1]$ such that they are both 1-1 and (uniform) measure preserving, and take $\pi = (\pi_1, \pi_2)$, then the procedure defined by

$$G_x(\theta) = G_{\text{Fiducial}, \pi(x)}(\theta)$$

will also be strongly similar, and moreover $\mathcal{L}G_{\text{Fiducial}, X_\theta}(z) = \mathcal{L}G_{X_\theta}(z)$ for all $\theta$ and $z$.

Example 2 To establish that stochastic ordering of $\mathcal{F} = \{F_\theta : \theta \in \Theta\}$ is not sufficient for optimality, the previous example is modified in two different but related ways. Let $X_\theta$ have a density

$$f_\theta(x) = \begin{cases} 
2(1 \Leftrightarrow \theta) & \text{if } 0 \leq x \leq \frac{1}{3}, \\
1 & \text{if } \frac{1}{3} < x \leq \frac{2}{3}, \\
2\theta & \text{if } \frac{2}{3} < x \leq 1.
\end{cases}$$

The procedure defined by $G_{\text{Fiducial}, X}(\theta) = 1 \Leftrightarrow F_\theta(x)$ is best strongly similar. Its non-uniqueness is not of interest at the moment. Next, consider the following
modification of the problem. Let \( \tilde{X}_\theta \) have a density

\[
\tilde{f}_\theta(x) = \begin{cases} 
2(1 \Leftrightarrow \theta) & \text{if } 0 \leq x \leq \frac{1}{3}, \\
2\theta & \text{if } \frac{1}{3} < x \leq \frac{2}{3}, \\
1 & \text{if } \frac{2}{3} < x \leq 1.
\end{cases}
\]

Define \( \tilde{G}_{\text{Fiducial},x}(\theta) = 1 \Leftrightarrow \tilde{F}_\theta(x) \), and the transformation

\[
\pi(x) = \begin{cases} 
x & \text{if } 0 \leq x \leq \frac{1}{3}, \\
x + \frac{1}{3} & \text{if } \frac{1}{3} < x \leq \frac{2}{3}, \\
x \Leftrightarrow \frac{1}{3} & \text{if } \frac{2}{3} < x \leq 1.
\end{cases}
\]

Notice that \( \mathcal{L}X_\theta = F_\theta \Rightarrow \mathcal{L}\pi(X_\theta) = \tilde{F}_\theta \) and \( \mathcal{L}\tilde{X}_\theta = \tilde{F}_\theta \Rightarrow \mathcal{L}\pi(\tilde{X}_\theta) = F_\theta \). This leads to the idea of using the procedure defined by

\[
\tilde{G}_x(\theta) = G_{\text{Fiducial},\pi(x)}(\theta),
\]

which is strongly similar. Next, it has to be shown that the procedure defined by \( \tilde{G}_x \) is better than that defined by \( \tilde{G}_{\text{Fiducial},x} \). For that purpose, introduce the procedure defined by

\[
G_x(\theta) = 1 \Leftrightarrow \tilde{F}_\theta(\pi(x)),
\]

so that \( \mathcal{L}\tilde{G}_{\text{Fiducial},x}(\theta)(z) = \mathcal{L}G_{x}(\theta)(z) \) and \( \mathcal{L}\tilde{G}_{\text{Fiducial},x}(\theta)(z) = \mathcal{L}G_{\text{Fiducial},x}(\theta)(z) \), for all \( \theta \) and \( z \). It is sufficient to show that the procedure defined by \( G_x \) is not in the equivalence class of best strongly similar procedures, in the first formulation of the problem. That this is the case follows from the fact that \( G_x \) is constant on \( (0, 1) \) if \( \frac{1}{3} < x \leq \frac{2}{3} \).

\[\nabla\]

### 5.6 Invariance

Recall the assumptions and the notations of Section 3.5. In this section, two different types problems were considered, namely invariant and equivariant problems. The ideas and results developed of Section 3.5 can easily be carried over to the context of making distributional inference about \( \psi(t) \).

**Invariant problems** Analogously to section Section 3.5, invariant problems will be considered first.
A problem of making distributional inference about $\psi(t)$, where $\psi : \Theta \mapsto \Psi$ is said to be invariant under $G$ if:

(i) $\mathcal{P}$ is invariant under $G$,
(ii) $\psi(\bar{g}\theta) = \psi(\theta)$, for all $\bar{g} \in \overline{G}$ and $\theta \in \Theta$.

Now, consider an invariant problem. A loss function $L_{\tau,f}$ for such problem is said to be invariant if

$$L_{\tau,f}(\bar{g}\theta, Q) = L_{\tau,f}(\theta, Q) \quad \forall \bar{g} \in \overline{G} \quad \forall \theta \in \Theta \quad \forall Q \in \Psi^*.$$

Notice that if a problem is invariant, then this implies that the loss function $L_{\tau,f}$ is such that $f_w$ is constant. Within the context of an invariant problem, a procedure $Q : \mathcal{X} \mapsto \Psi^*$ is said to be invariant if $Q \circ g = Q$, for every $g \in G$, which automatically implies that $Q$ depends on $x$ only through a maximal invariant statistic $v : \mathcal{X} \mapsto \mathcal{V}$. The risk function of any invariant procedure is constant on the orbits of $\overline{G}$, because if $\vartheta = \bar{g}\theta$ then

$$R_{\tau,f}(\vartheta, Q) = \mathbb{E}L_{\tau,f}(\vartheta, Q(X_\vartheta)) = \mathbb{E}L_{\tau,f}(\theta, Q(g(X_\theta))) = \mathbb{E}L_{\tau,f}(\theta, Q(X_\theta))$$

$$= R_{\tau,f}(\theta, Q).$$

If $G$ is amenable, then by applying the symmetrization used in Theorem 3.4, it can be shown that, for every procedure $Q : \mathcal{X} \mapsto \Psi^*$, there always exists an (almost) invariant procedure $\tilde{Q} : \mathcal{X} \mapsto \Psi^*$ such that

$$R_{\tau,f}(\theta, \tilde{Q}) \leq \sup_{\bar{g} \in \overline{G}} R_{\tau,f}(\bar{g}\theta, Q).$$

Or in other words, in a minimax sense the invariant procedures form an essentially complete class. An example of an invariant problem was given in Section 4.2, where Stein’s problem was considered. For that problem, the family $\mathcal{P}$ is invariant under rotations, and a maximal invariant statistic is given by $v(x) = x'x$. Although it has to be admitted that there is no decision-theoretic argument forcing reduction by invariance, it is certainly elegant and helps in reducing the problem to a more manageable form.

**Equivariant problems** Now, the attention will be turned to the so called equivariant problems. These are in the context of making distributional inferences about $\psi(t)$ often more interesting. Suppose that $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ is invariant under $G$, and that $\Psi$ can be identified with $\overline{G}$. Define the group $\overline{G}$, homeomorphic to $G$ and working on $\Psi$, by $\psi(\bar{g}\theta) = \bar{g}\psi(\theta)$. 

**Definition 5.10**

A problem of making distributional inference about $\psi(t)$, where $\psi : \Theta \mapsto \Psi$ is said to be invariant under $G$ if:

(i) $\mathcal{P}$ is invariant under $G$,
(ii) $\psi(\bar{g}\theta) = \psi(\theta)$, for all $\bar{g} \in \overline{G}$ and $\theta \in \Theta$. 

Now, consider an invariant problem. A loss function $L_{\tau,f}$ for such problem is said to be invariant if

$$L_{\tau,f}(\bar{g}\theta, Q) = L_{\tau,f}(\theta, Q) \quad \forall \bar{g} \in \overline{G} \quad \forall \theta \in \Theta \quad \forall Q \in \Psi^*.$$ 

Notice that if a problem is invariant, then this implies that the loss function $L_{\tau,f}$ is such that $f_w$ is constant. Within the context of an invariant problem, a procedure $Q : \mathcal{X} \mapsto \Psi^*$ is said to be invariant if $Q \circ g = Q$, for every $g \in G$, which automatically implies that $Q$ depends on $x$ only through a maximal invariant statistic $v : \mathcal{X} \mapsto \mathcal{V}$. The risk function of any invariant procedure is constant on the orbits of $\overline{G}$, because if $\vartheta = \bar{g}\theta$ then

$$R_{\tau,f}(\vartheta, Q) = \mathbb{E}L_{\tau,f}(\vartheta, Q(X_\vartheta)) = \mathbb{E}L_{\tau,f}(\theta, Q(g(X_\theta))) = \mathbb{E}L_{\tau,f}(\theta, Q(X_\theta))$$

$$= R_{\tau,f}(\theta, Q).$$

If $G$ is amenable, then by applying the symmetrization used in Theorem 3.4, it can be shown that, for every procedure $Q : \mathcal{X} \mapsto \Psi^*$, there always exists an (almost) invariant procedure $\tilde{Q} : \mathcal{X} \mapsto \Psi^*$ such that

$$R_{\tau,f}(\theta, \tilde{Q}) \leq \sup_{\bar{g} \in \overline{G}} R_{\tau,f}(\bar{g}\theta, Q).$$

Or in other words, in a minimax sense the invariant procedures form an essentially complete class. An example of an invariant problem was given in Section 4.2, where Stein’s problem was considered. For that problem, the family $\mathcal{P}$ is invariant under rotations, and a maximal invariant statistic is given by $v(x) = x'x$. Although it has to be admitted that there is no decision-theoretic argument forcing reduction by invariance, it is certainly elegant and helps in reducing the problem to a more manageable form.

**Equivariant problems** Now, the attention will be turned to the so called equivariant problems. These are in the context of making distributional inferences about $\psi(t)$ often more interesting. Suppose that $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ is invariant under $G$, and that $\Psi$ can be identified with $\overline{G}$. Define the group $\overline{G}$, homeomorphic to $G$ and working on $\Psi$, by $\psi(\bar{g}\theta) = \bar{g}\psi(\theta)$.
Definition 5.11 A problem of making distributional inference about $\psi(t)$, where $\psi: \Theta \mapsto \Psi$ is said to be equivariant under $G$ if:

(i) $\mathcal{P}$ is invariant under $G$,
(ii) $\exists \bar{G}$ homeomorphic to $G$, such that $\psi(\bar{g}\theta) = \bar{g}\psi(\theta)$, for all $\bar{g} \in \bar{G}$, and $\theta \in \Theta$.

Consider an equivariant problem. A loss function $L_{\tau,f}$ is said to be equivariant if

$$L_{\tau,f}(\bar{g}\theta, Q \circ \bar{g}^{-1}) = L_{\tau,f}(\theta, Q) \quad \forall \bar{g} \in \bar{G} \quad \forall \theta \in \Theta \quad \forall Q \in \Psi^*.$$

Within the context of an equivariant problem, a procedure $Q : X \mapsto \Psi^*$ is said to be equivariant if $Q(g \cdot) = Q \circ \bar{g}^{-1}(\cdot)$, for every $g \in G$. The risk function of an equivariant procedure is constant on the orbits of $\bar{G}$, because if $\vartheta = \bar{g}\theta$ then

$$R_{\tau,f}(\vartheta, Q) = E_{\tau,f}(\theta, Q(X_{\theta})) = E_{\tau,f}(\bar{g}\theta, Q(gX_{\theta})) = E_{\tau,f}(\theta, Q \circ \bar{g}^{-1}(X_{\theta})) = R_{\tau,f}(\theta, Q).$$

Using this result, it can be shown that in the case that $\bar{G}$ works transitively on $\Theta$, then a best equivariant procedure can be obtained. Notice that in this case $\Psi = \Theta$, and $G = \bar{G}$.

Lemma 5.2 Suppose that $\mathcal{P} = \{P_{\theta} : \theta \in \Theta\}$ is invariant under $G$, and that $\bar{G}$ works transitively on $\Theta$. Write $x = (u, v)$, where $u : X \mapsto U$ can be identified with $G$, and $v : X \mapsto V$ is ancillary. Let $L_{\tau,f}$ be a proper and equivariant loss function, and define

$$Q_{\text{MRE}}(x) = L(\bar{g}_u(x), \bar{g}_u^{-1}(\theta_u), v(X_{\theta_u}) = v). \quad (5.5)$$

Then $Q_{\text{MRE}}$ has uniformly minimum risk among all equivariant procedures, provided that the risk exists, and is well-defined.

**Proof.** Let $Q$ be any equivariant procedure for which the risk exists and well-defined. As the risk of $Q$ is constant on orbits of $\bar{G}$, which works transitively on $\Theta$, it is equal to

$$R_{\tau,f}(\theta, Q(X_{\theta})) = E_{\tau,f}(\theta, Q(X_{\theta})) = E_{\tau,f}(\theta, Q(X_{\theta})) = E[ E_{\tau,f}(\theta, Q(u(X_{\theta_u}) \circ \bar{g}_u(X_{\theta_u})(v))) \mid v(X_{\theta_u}) = v ] = E[ E_{\tau,f}(\theta, Q(u(X_{\theta_u}) \circ \bar{g}_u(X_{\theta_u})(v))) \mid v(X_{\theta_u}) = v ] = E[ E_{\tau,f}(\bar{g}_u^{-1}(X_{\theta_u} \theta_u), Q(u(X_{\theta_u}) \circ \bar{g}_u(X_{\theta_u})(v))) \mid v(X_{\theta_u}) = v ].$$
The properness of $L_{r,f}$ now implies that this risk can be minimized by taking $Q((u_e, v))$ to be

$$Q_{\text{MRE}}((u_e, v)) = \mathcal{L}(\tilde{g}_n(X_{n\theta}, \theta_1| v(X_{\theta_2}) = v)).$$

The equivariance, then provides that $Q_{\text{MRE}}$ is defined by (5.5).

A general method of Stein can be used to find the MRE procedures in equivariant problems, see e.g. Zidek [98] or Eaton [31]. This method is based on the fact that, under some conditions, it can be shown that the MRE procedure coincides with the (formal) Bayes procedure w.r.t. the induced right Haar measure on $\Theta$, i.e., $Q_{\text{MRE}} = Q_{\nu^*}$. This method for finding the MRE procedure will be applied in the next section.

It is important to realize that it can be possible that the same family of probability measures $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ is generated by different groups, all working transitively on $\Theta$. Applying the theory of this section to find MRE procedures will then provide different answers, if the different groups are used. A nice example is given by McCullagh [61], [62], who considers the family location-scale family of Cauchy distributions. The Cauchy distributions are invariant under the Möbius group, but can be generated by both the subgroups translation-scale on $x$, and translation-scale on $1/x$. From a mathematical point of view neither of the two parameterizations has to be preferred over the other. Hence, the choice of parameterization has to be made on the basis of contextual ingredients.

## 5.7 Some examples

To illustrate the results of this chapter, some problems will be worked out explicitly. First, two problems from the previous chapter will be reinvestigated: making inference about the mean of a normal random variable if the mean is known to be nonnegative, and a few more words are spent on Student's problem. Then, going beyond Student's problem, the problem of making inference about the mean of a lognormal distribution is investigated. In this case a best strongly similar procedure can be obtained by applying the fiducial argument in the form that was presented in Section 5.2. Some competing procedures will be derived by applying different paradigms, and a comparative analysis on the basis of both bias and risk will be given. To conclude with the problem of making inference about the correlation coefficient of a bivariate normal distribution will be discussed.
Incorporating additional information. If a distributional inference has to be made about the unknown mean \( t \) of a normal distribution with variance 1 in the absence of any further prior information, then practically every approach will provide the MRE procedure \( Q(x) = N(x, 1) \). Using the equivariant quadratic loss function \( L_r \) given by (5.2), with \( \tau \) Lebesgue measure, provides that the risk of this procedure is constant. It is displayed by the dotted line on the l.h.s. of Figure 5.1. Now, suppose that the additional information \( t \geq 0 \) is given. Applying the fiducial argument in this new situation provides

\[
G_{\text{Fiducial}, x}(\theta) = \begin{cases} 
0 & \text{if } \theta < 0, \\
\Phi(\theta - x) & \text{if } \theta \geq 0,
\end{cases}
\]

whereas a Bayesian approach with uniform prior leads to

\[
G_{\text{Bayes}, x}(\theta) = \begin{cases} 
0 & \text{if } \theta < 0, \\
\frac{\Phi(\theta-x) - \Phi(-x)}{\Phi(x)} & \text{if } \theta \geq 0.
\end{cases}
\]

Notice that applying the fiducial argument provides a strongly similar procedure, and hence its bias function is by definition equal to 0. Theorem 5.4 provides, in addition, that it is also the best strongly unbiased procedure. The Bayesian procedure cannot be weakly unbiased because it does not assign any positive credibility to \( \{0\} \), and hence \( B(0, Q_{\text{Bayes}}) = \leq \frac{1}{2} \). This has the effect that, for small \( t \), \( Q_{\text{Bayes}} \) tends to overestimate \( t \), see Figure 5.1. For \( t \) large, the effect of conditioning to the positive part of the real-line will become negligible, which causes the bias to tend to 0. Similarly, for large \( t \), the risk of both
The risk and bias functions of both Fiducial and Bayesian are displayed in Figure 5.1. On the basis of these results, one could conclude that abandoning probabilistic coherence, and assigning credibility mass to the boundary of the parameter space is not unreasonable. Moreover, it has the advantage that it preserves the strong unbiasedness.

Student's problem. Using the notation of this chapter, Student's problem can be reformulated as follows. Suppose $x = (x_1, \ldots, x_n)$ is the outcome of an independent random sample $X = (X_1, \ldots, X_n)$ from $N(\mu, \sigma^2)$, where $\mu$ was known, and in terms of the parameter $(\psi(\theta), \gamma(\theta) \in \mathbb{R}^2$. In terms of this notation, $\Theta = \mathbb{R}^2 \times \mathbb{R}$. The posterior distribution w.r.t. the right Haar measure $d\mu/\sigma^2$ on the location/scale group. Some further developments concerning Student's problem can be found in the work of Pedersen and Buehler, reported in Seidenfeld [51], and Zabell [52], indicating that some improvement is possible by modifying the conditioning argument. As the condition of logical independence does not necessarily minimize the Bayes risk if a loss function of the form (5.3) is used. Hence, taking an equivariant procedure function of the form (5.3), it might be possible to find an equivariant procedure. Some examples procedures will tend to the constant risk of $Q(\sigma)$ in the problem without the information that $\sigma > 0$. The risk and bias functions of both $Q_{\text{Fiducial}}$ and $Q_{\text{Bayes}}$ are displayed in Figure 5.1. On the basis of these results, one could conclude that abandoning probabilistic coherence, and assigning credibility mass to the boundary of the parameter space is not unreasonable. Moreover, it has the advantage that it preserves the strong unbiasedness.
that improves the risk of \(Q_{\text{Student}}\). Taking
\[
L_{\lambda,f}(\theta,Q) = (\gamma(\theta))^{-\frac{1}{2}}
\]
\[
\int_{\psi} \Pi(-\infty,\psi(\theta))(z) f_0(G(z)) + \Pi[\psi(\theta),\infty](z) f_1(1 \Leftrightarrow G(z)) dz.
\]
as equivariant loss function, it was established in Kroese et al. [55], that the MRE procedure for this location–scale problem is given by
\[
Q_{\text{MRE}}(x) = L(\bar{x} + n^{-\frac{1}{2}} s T_n),
\]
where \(s^2 = (n \Leftrightarrow 1)/ns^2\). This inference can also be obtained by taking the posterior distribution w.r.t. Jeffreys’s prior \(\sigma^{-2}d\mu d\sigma\), which is the left Haar measure on the location–scale group. Notice, however, that although this prior is called Jeffreys’s prior, Jeffreys [51] himself writes that he prefers the prior \(\sigma^{-1}d\mu d\sigma\). As both Gauss’s and Student’s inferences are equivariant, \(Q_{\text{MRE}}\) has uniformly smaller risk. In other words, Student’s problem is an example where the best strongly unbiased procedure is inadmissible. Notice that any other \(f_w\) will preserve the ordering of the risk functions, however, the risk of equivariant procedures will not be constant anymore. The inadmissibility of Student’s inference should, however, not be taken too seriously. Of course, a price has to be paid if one requires strong unbiasedness. Notice that by symmetry arguments it follows that \(Q_{\text{Gauss}}\) and \(Q_{\text{MRE}}\) are weakly unbiased.

The MRE procedure \(Q_{\text{MRE}}\) is more concentrated than \(Q_{\text{Student}}\), and for \(n = 1\), Student’s inference is not defined whereas \(Q_{\text{MRE}}(x) = \delta_x\) prescribes that, in the absence of information about \(\sigma^2\), one should use the distributional inference where all credibility mass is concentrated in the point estimate: if no information is available about the statistical uncertainty then it will be suppressed. Barnard (personal communication), however, suggested that, in the case \(n = 1\), instead of reporting \(\delta_x\), the likelihood function should be reported, or perhaps better the likelihood function as a function of \(\mu\) and \(\log(\sigma)\).

\[\clubsuit\]

**Inference about the mean of a lognormal distribution.** Neyman and E.S. Pearson constructed critical regions similar to the sample space by using conditioning arguments of the type used by Fisher to derive his exact test for the \(2 \times 2\) table. Using these conditioning arguments it is possible to tackle the following problem of which the solution does neither belong to the ‘modest handful of fiducial distributions explicitly adduced by Fisher’, which one can find in Savage [76], nor to the approximations discussed in, e.g. Aitchinson–Brown [1],
Weerahandi [93] or DiCicco–Efron [30]. The formulation of the problem is as follows. Suppose that inference has to be made about the mean \( \exp \{ \mu + \frac{1}{2} \sigma^2 \} \) of the lognormal distribution \( LN(\mu, \sigma^2) \). After natural transformations (recall that unbiasedness, optimality, etc., are preserved under monotone transformations), the problem becomes: given the outcome \( x = (x_1, \ldots, x_n) \) of an independent random sample \( X = (X_1, \ldots, X_n) \) from \( N(\mu, \sigma^2) \), where nothing is known a priori about the true value

\[
(\psi(t), \gamma(t)) = \left( \mu + \frac{1}{2} \sigma^2, \sigma^2 \right)
\]

of the unknown parameter, a distributional inference about \( \mu + \frac{1}{2} \sigma^2 \) is required. Let \( \Theta \) denote the canonical parameter space of the exponential family. Then \( \psi : \Theta \mapsto \mathbb{R} \) is defined by \( \psi(\theta) = \L (\theta_1 + \frac{1}{2})/(2\theta_2) \) and \( \gamma \mapsto \mathbb{R}^+ \) by \( \gamma(\theta) = \L 1/(2\theta_2) \). The following standard notation will be used

\[
\bar{X} = n^{-1} \sum_i X_i, \quad \bar{S}^2 = n^{-1} \sum_i (X_i - \bar{X})^2, \quad S^2 = (n \L 1)^{-1} \sum_i (X_i - \bar{X})^2,
\]

and corresponding small letters to denote the observed outcomes.

The fiducial argument prescribes that \( G_{\text{Fiducial}, x} \) should be constructed by equating \( G_{\text{Fiducial}, x}(z) = \alpha_{p,z}(x) \) where, \( \alpha_{p,z} \) is the P-value if \( H_z : \mu + \frac{1}{2} \sigma^2 \leq z \) is tested against \( A_z : \mu + \frac{1}{2} \sigma^2 \leq z \). Reformulating this in terms of the canonical parameters of the exponential family, these testing problems become

\[
H_z : t_1 + 2z t_2 \leq \L \frac{1}{2} \quad \text{versus} \quad A_z : t_1 + 2z t_2 > \L \frac{1}{2}.
\]

Notice that these testing problems are, for each fixed \( z \), linear in the canonical parameters, and hence UMP unbiased size-\( \epsilon \) tests can be obtained by applying standard Neyman–Pearson theory. The distribution functions of \( Q_{\text{Fiducial}} \) can be rewritten as

\[
G_{\text{Fiducial}, x}(z) = \frac{\int_0^1 \exp \{ \L b(x, z) r \} \L (r \L 1)^{\frac{1}{2}(n-3)} \, dr}{\int_0^1 \exp \{ \L b(x, z) r \} \L (r \L 1)^{\frac{1}{2}(n-3)} \, dr}, \quad (5.6)
\]

where

\[
a(x, z) = \text{sign}(\bar{x} \L z) \sqrt{\frac{(x - z)^2}{(x + (\bar{x} - z)^2)}}, \quad b(x, z) = \frac{1}{2} n \sqrt{\bar{s}^2 + (\bar{x} \L z)^2}.
\]

Notice that \( n = 1 \) the procedure degenerates to \( Q_{\text{Fiducial}}(x) = \delta_x \). This seems reasonable; if no information about the uncertainty is available then it is suppressed. If, for \( n \geq 2 \), it can be shown that the \( G_{\text{Fiducial}, x} \) are, indeed, distribution functions, then Lemma 3.4 provides that it is strongly unbiased, and
moreover Theorem 5.4 provides that it has uniformly minimum risk among all strongly unbiased procedures. Hence, it remains to be shown that $G_{\text{Fiducial},x}$ is, indeed, a probability distribution function.

**Lemma 5.3** If $n \geq 3$, then, for all $x \in \mathcal{X}$, $G_{\text{Fiducial},x}(z)$ as defined in (5.6) is non-decreasing in $z$, $\lim_{z \to -\infty} G_{\text{Fiducial},x}(z) = 0$ and $\lim_{z \to \infty} G_{\text{Fiducial},x}(z) = 1$.

The proof can be found in the appendix. As $Q_{\text{Fiducial}}$ is degenerate in the case $n = 1$, it cannot be strongly similar. For $n = 2$, the computations made, suggest that the $G_{\text{Fiducial},x}$ are also distribution functions.

Although $Q_{\text{Fiducial}}$ has minimum risk among all strongly similar procedures, it may well be possible to improve its risk uniformly by considering a procedure that is not strongly similar, i.e., by considering weakly unbiased procedures. But how serious should the loss in risk performance due to requiring strong similarity be taken? In the context of point estimation it is well known that a smaller MSE may be obtained by allowing some bias. However, in practice one usually keeps using the unbiased estimator. In the comparison of different procedures this bias could play a similar role as the risk. Notice that the problem is equivariant under changes of location if an equivariant loss function is chosen. Hence, for this example the quadratic loss function (5.2), with $\tau$ Lebesgue measure, is used to evaluate the inferences. As no natural weighting function $f_w$ exists, it is chosen to be constant, and equal to 1. Notice that $Q_{\text{Fiducial}}$ is equivariant, and hence has constant risk for $\theta$ such that $\gamma(\theta)$ is fixed. In general, procedures that are equivariant under changes of location, have risk and bias functions that are constant on sets of $\theta$ such that $\gamma(\theta)$ is fixed.

The question of interest is, of course, how can one find interesting competitors to $Q_{\text{Fiducial}}$? Extending the classical statistical approach based on the CLT, where a point estimator is given with its standard deviation, one can introduce $Q_{\text{Gauss}}$ which is obtained as follows. The point estimator $\bar{x} + \frac{1}{2}S^2$ is an unbiased estimator of $\mu + \frac{1}{2}\sigma^2$ and has variance $n^{-1}\sigma^2 + \frac{1}{2}(n \leftrightarrow 1)^{-1}\sigma^4$. As the value of $\sigma^2$ is unknown, one can replace it by its unbiased estimator $S^2$, and obtain

$$Q_{\text{Gauss}}(x) = N \left( \bar{x} + \frac{1}{2}s^2, n^{-1}s^2 + \frac{1}{2}(n \leftrightarrow 1)^{-1}s^4 \right).$$

Using this method of inference, the uncertainty of not knowing $\sigma^2$ is suppressed which causes the inferences to be less dispersed. Notice that $Q_{\text{Gauss}}$ is also equivariant under changes of location. Fisher used pivotal quantities to extend the fiducial argument to higher dimensions. In Weerahandi [93] this approach,
based on pivotal functions, is worked out explicitly, and he proposes the following method. Define $U = \sqrt{n}(\bar{X} \Leftarrow \mu)/\sigma$, and $V = (n \Leftarrow 1)S^2/\sigma^2$, and notice that $U$ and $V$ are independent and, moreover, that $\mathcal{L}U = N(0,1)$ and $\mathcal{L}V = \chi^2_{n-1}$. Algebraic manipulation of these two equalities provides
\[
\mu = \bar{X} \Leftarrow S\sqrt{\frac{n-1}{n}} \frac{U}{\sqrt{V}} \quad \text{and} \quad \sigma^2 = (n \Leftarrow 1)\frac{S^2}{V},
\]
and hence that $\mu + \frac{1}{2}\sigma^2 = \bar{X} \Leftarrow S\sqrt{\frac{n-1}{n}} \frac{U}{\sqrt{V}} + \frac{1}{2}(n \Leftarrow 1)\frac{S^2}{V}$. Given the observed $x$ this provides the distributional inference
\[
Q_{\text{Weerahandi}}(x) = \mathcal{L} \left( \bar{x} \Leftarrow s \sqrt{\frac{n-1}{n}} \frac{U}{\sqrt{V}} + \frac{1}{2}(n \Leftarrow 1)\frac{S^2}{V} \right),
\]
where $U$ and $V$ are independent, $\mathcal{L}U = N(0,1)$, and $\mathcal{L}V = \chi^2_{n-1}$. Notice that $Q_{\text{Weerahandi}}$ is also equivariant under changes of location. A neo-Bayesian approach to the problem would be as follows. Start by making distributional inference about the parameters of the normal distribution $(\mu, \sigma^2)$, which are regarded as the outcome of a random variable $(T_1, T_2)$. As the normal distributions constitute a location–scale family, the invariant measure $\nu(d\theta) = 1/\sqrt{\theta_2}$ (right Haar measure) is taken as improper prior. The posterior distribution $Q_{\text{MRE}}(x)$ would then be the best equivariant procedure if an invariant proper loss function is used to evaluate the inferences about $(\mu, \sigma^2)$. Notice that $Q_{\text{MRE}}(x)$ is determined by the conditional posterior distribution of $T_1$ given $T_2 = \theta_2$, which is $N(\bar{x}, n^{-1}\theta_2)$, and the marginal posterior distribution of $T_2$, which is the Inverse Gamma distribution $IG(\frac{1}{2}(n \Leftarrow 1), \frac{1}{2}(n \Leftarrow 1)s^2)$. The neo-Bayesian solution for making inference about $\mu + \frac{1}{2}\sigma^2$ would then be to take the induced measure of the mapping $\eta$, given by $\eta(\theta) = \theta_1 + \frac{1}{2}\theta_2$, i.e.,
\[
Q_{\text{Bayes}}(x) = Q_{\text{MRE}}(x) \circ \eta^{-1}.
\]
An explicit expression for the density $g_{\text{Bayes}, x} = G'_{\text{Bayes}, x}$ of $Q_{\text{Bayes}}$ is given by
\[
g_{\text{Bayes}, x}(z) = \frac{\frac{2}{\pi}(\frac{1}{2}n)^n(\frac{n-1}{2})^{\frac{1}{2}(n-1)}}{(\frac{1}{2}(n \Leftarrow 1))(\frac{n}{2}(\frac{1}{2}n \sqrt{\bar{s}^2 + (\bar{x} \Leftarrow z)^2})^{\frac{1}{2}n})\exp\left\{n \frac{1}{2}(z \Leftarrow \bar{x})\right\}} \text{BesselK} \left(\frac{1}{2}n, \frac{1}{2}n \sqrt{\bar{s}^2 + (\bar{x} \Leftarrow z)^2}\right),
\]
where BesselK($\cdot$, $\cdot$) is a modified Bessel function of the second kind. Notice that this is the density function of the generalized hyperbolic distribution $H(\frac{1}{2}(n \Leftarrow 1), \frac{n}{2}, \frac{1}{2}n, \bar{x}, \bar{s})$, see e.g. Barndorff-Nielsen [5].
Figure 5.2: Both figures correspond to the problem of making distributional inference about $\exp\{\mu = \frac{1}{2}\sigma^2\}$, with $n = 3$. In the figure on the l.h.s. the bias functions are plotted, and in the figure on the r.h.s. the corresponding risk functions are displayed.

Lemma 5.4 $Q_{\text{Weerahandi}} = Q_{\text{Bayes}}$.

Proof. It has to be show that $Q_{\text{Weerahandi}}(x) = Q_{\text{Bayes}}(x)$, for each $x \in \mathcal{X}$. So, take an arbitrary $x \in \mathcal{X}$, and define $U$ and $V$ independent such that $\mathcal{L}U = N(0,1)$ and $\mathcal{L}V = \chi^2_{n-1}$. Now use the fact that

$$\mathcal{L}(T_1 | T_2) = \mathcal{L}\left(\bar{x} + \sqrt{\frac{T_2}{n}}U\right) \quad \text{and} \quad \mathcal{L}T_2 = \mathcal{L}\left(\frac{n \leftarrow 1}{s^2}\right),$$

and hence that

$$\mathcal{L}(T_1, T_2) = \mathcal{L}\left(\bar{x} + s \sqrt{\frac{n \leftarrow 1}{U}} \frac{U}{\sqrt{V}}, \frac{n \leftarrow 1}{s^2}\right).$$

Hence, $Q_{\text{Bayes}}(x) = \mathcal{L}\left(\frac{n \leftarrow 1}{s^2}\right)$ because $\mathcal{L}U = \mathcal{L}(\leftarrow U)$, which completes the proof. \[\square\]

For $n = 3$, the bias and risk functions of $Q_{\text{Fiducial}}, Q_{\text{Gauss}}$ and $Q_{\text{Weerahandi}} = Q_{\text{Bayes}}$ are computed numerically, and displayed in Figure 5.2. As all procedures are equivariant under changes of location, and hence it suffices to study a cross section of the bias and the risk function where they are plotted for different values of $\gamma$. It can be seen that $Q_{\text{Fiducial}}$ has smaller risk than $Q_{\text{Weerahandi}}$ and
that $Q_{\text{Gauss}}$ has in its turn smaller risk than $Q_{\text{Fiducial}}$. Notice that $Q_{\text{Gauss}}$ has a positive bias which means that on average it underestimates the true value of the parameter. Conversely $Q_{\text{Weerahandi}} = Q_{\text{Bayes}}$ has a negative bias which means that on average, it overestimates the true value of the parameter.

**Inference about the correlation coefficient of a bivariate normal distribution** When Fisher [37] introduced the fiducial argument, he used the problem of making inference about the correlation coefficient $\rho$ of a bivariate normal distribution as an illustration. Nevertheless, he did not consider it necessary to write down an explicit expression for the fiducial density. One way to obtain such an expression is to start from the marginal sampling density of the correlation coefficient. Let $r$ denote the sample correlation coefficient that can be computed from an i.i.d. sample of size $n \geq 3$ from a bivariate normal distribution. The sampling density of $r$, given $\rho$, can be represented as

$$f_{\rho}(r) = \frac{n-2}{\pi} (1 \leftrightarrow \rho^2) \frac{1}{2(n-1)} (1 \leftrightarrow r^2) \frac{1}{2(n-4)} \int_0^\infty \frac{dz}{(\cosh(z) \leftrightarrow \rho r)^{n-1}}. \tag{5.9}$$

Fisher prescribed that the fiducial density of $\rho$, should be computed by inverting the sampling distribution of $r$, according to the formula

$$g_r(\rho) = \frac{dF_{\rho}(r)}{d\rho},$$

where $F_{\rho}(r)$ is the distribution function of $r$, and $g_r(\rho)$ denotes the fiducial density. By integrating (5.9) over $r$ to obtain $F_{\rho}(r)$, and subsequently differentiating w.r.t. $\rho$, Cornish–Bennet [7] established that an explicit expression for this fiducial density is given by

$$g_r(\rho) = \frac{n}{\pi} (1 \leftrightarrow r^2) \frac{1}{2(n-2)} (1 \leftrightarrow \rho^2) \frac{1}{2(n-3)} \int_0^\infty \frac{\cosh(z)dz}{(\cosh(z) \leftrightarrow \rho r)^{n-1}}. \tag{5.10}$$

This method, of course, coincides with the approach that was proposed in this chapter. Lemma 3.4 provides that this inference is strongly unbiased. As the family $\{F_{\rho} : \rho \in (\pm 1, 1)\}$ has monotone likelihood ratio, it follows from Theorem 5.4 that it has uniformly minimum risk among all strongly unbiased procedures. To obtain this result, a reduction by invariance under coordinate wise location-scale transformations was performed implicitly.

Another possibility to obtain the fiducial density for $\rho$ is by taking the bivariate normal distribution as a starting point. Suppose that we have an i.i.d.
sample $X_1, \ldots, X_n$ from a bivariate normal distribution, and assume that $n \geq 3$. Let
\[
\bar{X} = \left( \begin{array}{c} \bar{X}_1 \\ \bar{X}_2 \end{array} \right) \quad \text{and} \quad S = \left( \begin{array}{cc} S_1^2 & RS_1S_2 \\ RS_1S_2 & S_2^2 \end{array} \right),
\]
denote the standard sufficient statistics for the bivariate normal distribution. Similarly let
\[
\mu = \left( \begin{array}{c} \mu_1 \\ \mu_2 \end{array} \right) \quad \text{and} \quad \Sigma = \left( \begin{array}{cc} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{array} \right),
\]
denote the standard parameters of the bivariate normal distribution. Using the fact that $\Sigma^{-1} S$ is a pivotal quantity, that
\[
S = TT' = \left( \begin{array}{cc} S_1 & 0 \\ RS_2 & S_2 \sqrt{1 - \rho^2} \end{array} \right) \left( \begin{array}{cc} S_1 & 0 \\ RS_2 & S_2 \sqrt{1 - \rho^2} \end{array} \right)'
\]
where prime denotes transpose, and that
\[
\Sigma = \Sigma_1 \Sigma_2 \Sigma_3
\]
\[
\Sigma_1 = \left( \begin{array}{cc} \sigma_1^2 & 0 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{array} \right) \left( \begin{array}{cc} \sigma_1^2 & 0 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{array} \right)'
\]
it follows that
\[
\Sigma_1^{-1} T = \left( \begin{array}{cc} S_1 & 0 \\ RS_2 & S_2 \sqrt{1 - \rho^2} \end{array} \right) \left( \begin{array}{cc} \frac{\sigma_1}{\sigma_2} \sqrt{1 - \rho^2} & 0 \\ \frac{\rho \sigma_1}{\sigma_2} \sqrt{1 - \rho^2} & S_2 \sqrt{1 - \rho^2} \end{array} \right)
\]
is a pivotal quantity. Taking this as a starting point, Fraser obtained the same expression as (5.10) for the fiducial density by using the following argument. The Wishart distribution may be factorized into the product of three independent distributions; that of
\[
Y_1 = \sqrt{\frac{n}{\sigma_1}} \left[ \frac{1}{S_1} \right] \sim \chi^2_{n-1}
\]
\[
Y_2 = \sqrt{\frac{n}{\sigma_2}} \left[ \frac{1}{S_2 \sqrt{1 - \rho^2}} \right] \sim \chi^2_{n-2}
\]
\[
Z = \sqrt{\frac{n}{1 - \rho^2}} \left( \frac{RS_2}{\sigma_2} \right) \sim N(0,1).
\]
Moreover, it can be shown that these quantities satisfy the relation
\[
Z = \sqrt{\frac{r}{1 - \rho^2}} Y_2 \sim \frac{\rho}{\sqrt{1 - \rho^2}} Y_1.
\] (5.11)
Starting with the joint distribution of $Y_1, Y_2,$ and $Z$, keeping $r$ fixed, the fiducial distribution of $\rho$ can be obtained by making the observation that rewriting (5.11) provides

\[
L \left( \frac{\rho}{\sqrt{1 - \rho^2}} \right) = L \left( \frac{r \sqrt{1 - r^2} Y_2 \Leftrightarrow Z}{Y_1} \right).
\]

As $x/\sqrt{1 - x^2}$ is a monotone transformation on $(\Leftrightarrow 1, 1)$, and fiducial distributions are probabilistically coherent under monotone transformations, the fiducial distribution of $\rho$ is obtained by taking the distribution induced by the inverse of this transformation.

It can also be shown that the fiducial distribution for $\rho$ can be obtained by marginalizing an invariant posterior distribution for all parameters of the bivariate normal distribution, provided that the appropriate group is chosen to describe the symmetry of the model. To do so, start by noting that the joint sampling distribution of the five statistics of the bivariate normal distribution has a density given by

\[
f(\bar{x}, S|\mu, \Sigma) = \frac{n}{2\pi \sigma_1 \sigma_2 \sqrt{1 - \rho^2}} \exp \left\{ \frac{-n}{2(1 - \rho^2)} \left( \frac{(\bar{x}_1 \Leftrightarrow \mu_1)^2}{\sigma_1^2} \Leftrightarrow \frac{2\rho(\bar{x}_1 \Leftrightarrow \mu_1)(\bar{x}_2 \Leftrightarrow \mu_2)}{\sigma_1 \sigma_2} + \frac{(\bar{x}_2 \Leftrightarrow \mu_2)^2}{\sigma_2^2} \right) \right\}
\times \frac{(n \Leftrightarrow 1)^{n-1} \frac{n}{\pi} \frac{s_1^{-n-2} s_2^{-n-2}(1 \Leftrightarrow \rho^2)^{\frac{1}{2}(n-4)}}{s_1^{-n-1} s_2^{-n-1}(1 \Leftrightarrow \rho^2)^{\frac{1}{2}(n-1)}}}{(n \Leftrightarrow 2)\sigma_1^{-n-1} \sigma_2^{-n-1} \frac{2\rho s_1 s_2}{\sigma_1 \sigma_2} + \frac{s_2^2}{\sigma_2^2}} \right\}.
\]

Recall from Section 4.4 that the group $G$ is the semi-direct product of translations and multiplication by nonsingular matrices, and that $H$ is the subgroup of translations and multiplication by lower triangular matrices with positive diagonal elements. The induced invariant measures for the groups $G$ and $H$ can be expressed in terms of the standard parameterization of the bivariate normal distribution. This provides that the standard Jeffreys prior on $G$, which equals the left Haar measure on $H$, is given by

\[
\nu'(d\mu, d\Sigma) = \frac{d\mu d\sigma_1 d\sigma_2 d\rho}{\sigma_1 \sigma_2 (1 \Leftrightarrow \rho^2)^{\frac{1}{2}}}.
\]
and that the right Haar measure on $H$, is given by

$$
\nu^r(d\mu, d\Sigma) = \frac{d\mu d\sigma_1 d\sigma_2 d\rho}{\sigma^2(1 \leftrightarrow \rho^2)}
$$

(5.14)

where $d\mu = d\mu_1 d\mu_2$. In Geisser–Cornfield [42] it was shown that computing the posterior distribution of $\mu$ and $\Sigma$ w.r.t. the prior $\nu^l$ given by (5.13), and subsequently computing the marginal posterior of $\rho$, provides the same posterior distribution for $\rho$ as that obtained by taking the prior $\nu(d\rho) = (1 \leftrightarrow \rho^2)^{-1/2}$ and computing the posterior distribution using the marginal sampling distribution of $r$ given by (5.9). A nice feature of this posterior distribution for $\rho$ is that it has the same density as the sampling distribution of $r$, but with $\rho$ and $r$ interchanged, i.e.,

$$
g_{\nu^l, r}(\rho) = \frac{n}{\pi} (1 \leftrightarrow \rho^2)^{\frac{1}{2}(n-1)} (1 \leftrightarrow \rho^2)^{\frac{1}{2}(n-4)} \int_0^\infty \frac{dz}{(\cosh(z) \leftrightarrow \rho r)^{n-1}} d\rho.
$$

(5.15)

In Section 4.2 it is argued that $\nu^l$ might not be the most appropriate prior because it leads to inconsistent inferences. Taking $\nu^r$ as a prior this problem will be avoided.

**Lemma 5.5** Suppose that $n \geq 3$. The marginal density of $\rho$ of the posterior distribution of $\mu$ and $\Sigma$ w.r.t. the prior $\nu^r$ given by (5.14) equals the fiducial density of $\rho$ given by (5.10).

**Proof.** Take the posterior density of $\mu$ and $\Sigma$ w.r.t. the prior $\nu^r$, and integrate out $\mu$. Notice that the marginal posterior density of $\sigma_1$, $\sigma_2$, and $\rho$, is then proportional to

$$
\exp \left\{ \frac{-\frac{N}{2} \left( \frac{\sigma^2}{\sigma_1^2} \right) - \frac{2\rho r \sigma_1 s_3}{\sigma_1 \sigma_2} + \frac{s_2^2}{\sigma_2^2} }{\sigma_1^{n+1} \sigma_2^{n-1} (1 \leftrightarrow \rho^2)^{\frac{1}{2}(n+1)}} \right\}.
$$

(5.16)

Next, to integrate out $\sigma_1$ and $\sigma_2$, make the transformation of variables

$$
x = \frac{s_1 s_3}{\sigma_1 \sigma_2} \quad \text{and} \quad \exp\{y\} = \frac{s_1 \sigma_4}{\sigma_1 s_3},
$$

so that (5.16) is proportional to

$$
\frac{x^{n-2} \exp\{y\}}{(1 \leftrightarrow \rho^2)^{\frac{1}{2}(n+1)}} \exp \left\{ \frac{nx}{2(1 \leftrightarrow \rho^2)} (2 \cosh(y) \leftrightarrow 2\rho r) \right\}.
$$

Starting by integrating out $x$, provides that up to a constant

$$
(1 \leftrightarrow \rho^2)^{\frac{1}{2}(n-3)} \frac{\exp\{y\} dy d\rho}{(\cosh(y) \leftrightarrow \rho r)^{n-1}},
$$
so it remains to evaluate
\[
\int_{-\infty}^{\infty} \frac{\exp\{y\} dy}{(\cosh(y) \leftrightarrow \rho r)^{n-1}} = 2 \int_{-\infty}^{\infty} \frac{\cosh\{y\} dy}{(\cosh(y) \leftrightarrow \rho r)^{n-1}} = 2 \int_{-\infty}^{\infty} \frac{\exp\{y\} dy}{(\cosh(y) \leftrightarrow \rho r)^{n-1}},
\]

which is finite in case \( n \geq 3 \). Hence,
\[
\int_{-\infty}^{\infty} \frac{\exp\{y\} dy}{(\cosh(y) \leftrightarrow \rho r)^{n-1}} = \int_{-\infty}^{\infty} \frac{\cosh\{y\} dy}{(\cosh(y) \leftrightarrow \rho r)^{n-1}} = 2 \int_{0}^{\infty} \frac{\cosh\{y\} dy}{(\cosh(y) \leftrightarrow \rho r)^{n-1}},
\]

which completes the proof, as it shows that the marginal density of \( \rho \) is proportional to the fiducial density (5.10).

This result provides that the fiducial distribution of the correlation coefficient is consistent in the sense of Heath–Sudderth–Lane.

Using the theory of exponential families it can be shown that the fiducial distribution about the correlation coefficient can be obtained by equating \( G_{\text{Fiducial}, r}(z) \) the \( P \)-value corresponding to a family of UMPU tests for testing \( H_z : \rho \leq z \) against \( A_z : \rho > z \), see e.g. Lehmann [58]. Hence by Theorem 5.4, it follows that the fiducial procedure does not only have uniformly minimum risk among all strongly unbiased procedures based on \( r \), but it has this property among all strongly unbiased procedures.

To illustrate the trade–off between risk and bias, consider the following example. Take \( n = 11 \), let \( \nu_1 \) be Lebesgue measure on \( \mathbb{R} \), and let \( Q_{\nu_1} \) be the corresponding posterior distribution. Similarly take \( \nu_2(d\theta) = (1 \leftrightarrow \theta^2)^{-1} \), and let \( Q_{\nu_2} \) be the corresponding posterior distribution. In Figure 5.3 the risk and bias functions of the Bayes procedures \( Q_{\nu_1} \) and \( Q_{\nu_2} \) are plotted together with those of the fiducial procedure \( Q_{\text{Fiducial}} \). It can be seen in Figure 5.3 that both in terms of bias and risk, \( Q_{\text{Fiducial}} \) takes an intermediate position between \( Q_{\nu_1} \) and \( Q_{\nu_2} \). Looking at the bias, it can be seen that \( Q_{\nu_1} \) displays a central tendency, whereas \( Q_{\nu_2} \) has a tendency to assign its mass somewhat more towards the extreme points. This has its reflection in the risk functions. Notice that \( Q_{\nu_2} \) has smaller risk around \( \approx 1 \) and 1, whereas \( Q_{\nu_1} \) has better risk around
Figure 5.3: On the l.h.s. the risk functions of $Q_F$, $Q_{v_1}$, and $Q_{v_2}$, w.r.t. the quadratic loss function (3.4) with $\tau$ Lebesgue measure on $[0,1]$, are displayed. On the r.h.s. the bias functions are displayed. Notice that the bias function of $Q_F$ is identically equal to 0. The computations for both figures are based on the sample size $n = 11$.

As $Q_{\text{Fiducial}}$ compromises both in terms of risk and bias between these two procedures, it can be concluded that it is not at all a bad choice.