4

Fiducial inference; a review

In this chapter, a brief review about broad lines of the historical development of the fiducial inference will be given. The concept of fiducial probability was introduced by Fisher [37] in his paper 'Inverse Probability'. The idea behind fiducial inference, is the following. Suppose that there is *no* prior information available about the true value t of the unknown parameter. Given the observation x, one wants to assign epistemic probabilities to subsets of Θ, indicating the belief that the true value of the parameter is contained in this subset. If no particular subset is of special interest, then this boils down to specifying a probability distribution $Q(x)$ on Θ, provided that these epistemic probabilities are assigned coherently to all subsets. The 'classical' method of deriving such distributional inferences $Q(x)$ is by applying Bayes's Theorem. The drawback of this method, however, is that it requires the specification of a prior distribution. Fisher regarded the specification of prior probabilities as being in conflict with the assumption that no prior information is available. His fiducial argument provides an alternative method to generate distributional inferences, which can be applied without specifying a prior. Fisher himself derived a number of fiducial inferences for various problems, without being sufficiently clear about the underlying principles; Buehler [18] wrote:

> 'Fisher never gave an acceptable general definition of fiducial probability. For the case of one observation x and one parameter θ, with cumulative sampling distribution $F_\theta(x)$ monotone decreasing in θ, Fisher defined the fiducial density for θ to be
>
> $$g_x(\theta) = -dF_\theta(x)/d\theta.\text{'}$$

(A slight notational modification has been made.) The key question is of course: what is the rationale behind this formula? Various authors have given their interpretation of what they regarded as the essence of the fiducial argument. Using these interpretations, they formalized and extended what can in essence be thought of as fiducial inference. In this chapter, these interpretations will be reviewed and presented in terms of so called structured or functional models. This approach encompasses most of the results known about fiducial inference. In Chapter 5, a different approach towards fiducial inference is adopted. This approach builds on the theory developed in Chapter 3.

4.1 Introduction

In the previous two chapters the attention was restricted to making inference about the truth or falsity of statistical hypotheses. This problem could be regarded as a problem of estimating the true value of some indicator function, i.e., $\mathbb{I}_{\Theta_{\mathrm{H}}}(t)$. The true value of $\mathbb{I}_{\Theta_{\mathrm{H}}}(t)$ can only take on two possible values, namely $\{0,1\}$. Now, recall from Section 3.1, that there is an equivalent representation of an estimator $\alpha : \mathcal{X} \mapsto [0,1]$ of $\mathbb{I}_{\Theta_{\mathrm{H}}}(t)$, i.e., the procedure $Q : \mathcal{X} \mapsto \{0,1\}^*$, where $\{0,1\}^*$ denotes the space of all probability measures on $\{0,1\}$. This procedure was defined by

$$Q(x) = (1 \Leftrightarrow \alpha(x))\delta_0 + \alpha(x)\delta_1 \in \{0,1\}^*.$$

In other words, an estimator α can also be regarded as a map Q from the outcome space to the space of probability distributions on the theoretically possible values that the unknown of interest can attain.

In this chapter the more general problem of making inference about $t \in \Theta$, or a real–valued function thereof $\psi(t) \in \Psi$, will be considered. It is assumed that the same regularity conditions as in Sections 3.1 and 3.4 hold. The idea of making inference in the form of a probability distribution on the space of theoretically possible values of the unknown of interest, can also be used in this context; take for example instead of the indicator function $\mathbb{I}_{\Theta_{\mathrm{H}}} : \Theta \mapsto \{0,1\}$ the function $\psi : \Theta \mapsto \Psi \subset \mathbb{R}$, and let $Q = Q(x)$ in this case be a probability distribution with values in Ψ. Such procedure $Q : \mathcal{X} \mapsto \Psi^*$, where Ψ^* denotes the space of all probability measures on Ψ, for making such distributional inferences can be specified in many different but equivalent ways, e.g. for each x one could prescribe how to construct the density function g_x, or alternatively, how to construct the distribution function G_x of $Q(x)$.

To start with, consider the following canonical situation, which was considered by Fisher. Suppose that one is given the outcome x of a random variable

X, with values in $\mathcal{X} \subset \mathbb{R}$, and with probability distribution function F which is known to be a member some family $\mathcal{F} = \{F_\theta : \theta \in \Theta\}$. It is assumed that this family is such that it has the following properties:

(i) $F_\theta(x)$ is a continuous function of (x, θ),
(ii) $\Theta \subset \mathbb{R}$ is convex, i.e., $\Theta = (\underline{\theta}, \bar{\theta})$, and
(iii) $F_\theta(x)$ is nondecreasing in x and nonincreasing in θ,
(iv) the densities $f_\theta(x) = F'_\theta(x)$ exist and $\mathcal{F}$ has monotone likelihood ratio.

Notice that the likelihood ratio assumption (iv) implies (iii). According to the ideas of Fisher it is not necessary to assume (iv) to conclude that the fiducial argument produces a valid distributional inference; condition (iii) suffices is this respect. However, in the next chapter it will be shown that if one wants to establish that fiducial inference is optimal in some sense, then one needs the Neyman–Pearson assumption (iv). The objective is to make a distributional inference about the true value t of the unknown parameter θ, which is implicitly defined by $F = F_t$. Fisher's fiducial argument prescribes how to construct $Q(x)$ by specifying its density g_x, i.e.,

$$g_x(\theta) = \Leftrightarrow dF_\theta(x)/d\theta.$$

Notice by looking at this prescription that the fiducial argument does not need the input of prior distributions, loss functions, etc., to derive a distributional inferences.

Taking $g_x(\theta) = \Leftrightarrow dF_\theta(x)/d\theta$, the fiducial probability of an interval (θ_1, θ_2) can be obtained by integration, and equals $F_{\theta_2}(x) \Leftrightarrow F_{\theta_1}(x)$. Notice that the fiducial density g_x does not necessarily integrate to 1, i.e., in case that $\lim_{\theta \to \underline{\theta}} F_\theta(x) \neq 1$, or $\lim_{\theta \to \bar{\theta}} F_\theta(x) \neq 0$. In these cases credibility mass $\lim_{\theta \downarrow \underline{\theta}} 1 \Leftrightarrow F_\theta(x)$ is assigned to $\{\underline{\theta}\}$, and $\lim_{\theta \uparrow \bar{\theta}} F_\theta(x)$ is assigned to $\{\bar{\theta}\}$. This provides that the distribution function G_x of the fiducial inference is given by

$$G_x(\theta) = \begin{cases} 0 & \text{if} \quad \theta < \underline{\theta}, \\ \lim_{\theta \downarrow \underline{\theta}}(1 \Leftrightarrow F_\theta(x)) & \text{if} \quad \theta = \underline{\theta}, \\ 1 \Leftrightarrow F_\theta(x) & \text{if} \quad \underline{\theta} < \theta < \bar{\theta}, \\ 1 & \text{if} \quad \theta \geq \bar{\theta}. \end{cases}$$

As it could well be possible that the fiducial argument assigns credibility mass to a singleton, the representation of the fiducial inference in terms of the distribution function G_x is preferred to the representation in terms of the density g_x. To illustrate the usage of the fiducial argument, and to get a better understanding of what it does geometrically, consider the following example.

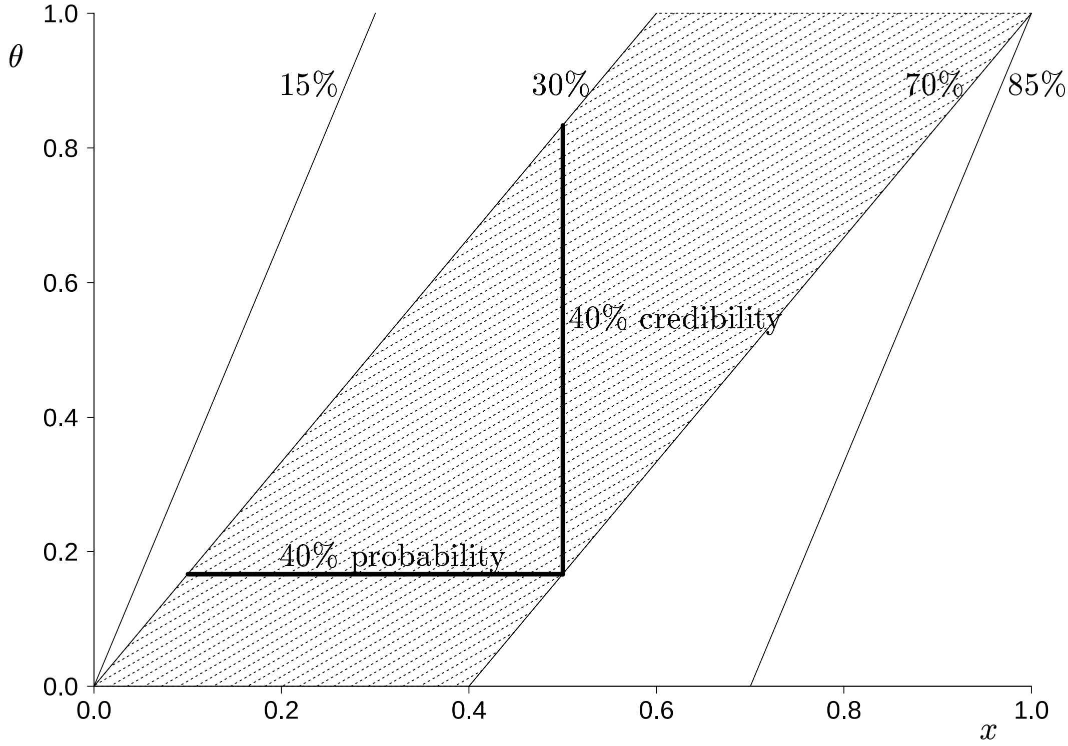

Figure 4.1: *The geometrical interpretation of the fiducial argument.*

The geometrical interpretation of the fiducial argument. Let X_θ be a random variable with density

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

The family $\mathcal{F} = \{F_\theta : \theta \in \Theta\}$ is stochastically increasing, and hence applying the fiducial argument provides a valid distributional inference. To see how the fiducial inference is constructed, break the argument into the following steps. First, draw in the (x, θ) plane the lines corresponding to fixed quantiles. In Figure 4.1 this is done for the 15%, 30%, 70% and 85% quantiles. Horizontally, one can read off, for a fixed θ, the probability of having an observation between two points. Similarly, one can read off vertically, for a fixed x, the credibility of t lying between two points. In this fashion, the fiducial distribution can be computed, and its distribution function is given by

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

Notice that credibility mass $\frac{1-x}{2}$ is assigned to $\{0\}$, and $\frac{x}{2}$ to $\{1\}$. ♣

Suppose that both $\psi : \Theta \mapsto \Psi \subset \mathbb{R}$ and $v : \mathcal{X} \mapsto \mathcal{V} \subset \mathbb{R}$ are continuous and strictly increasing functions. Then reparameterization of $\mathcal{P}$ into $\widetilde{\mathcal{P}} = \{\widetilde{\mathrm{P}}_\psi : \psi \in \Psi\}$, where $\widetilde{\mathrm{P}}_\psi = \mathcal{L}v(X_\theta)$, and a subsequently applying the fiducial argument, provides the same fiducial inference $\widetilde{Q}(v(x))$ as first applying the fiducial argument to $\mathcal{P}$, to obtain $Q(x)$, and then taking the induced measure $\widetilde{Q}(v(x)) = Q(x) \circ \psi^{-1}$ on Ψ. In other words, in the canonical situation, fiducial inferences are probabilistically coherent under monotone continuous transformations. This coherence of the inferences is not more than reasonable, because the monotone continuous transformations leave the problem virtually unchanged; one remains within the context described by Buehler. That, in contrast with Bayesian inferences, fiducial inferences are not probabilistically coherent under all transformations, will be shown in the next section.

To extend the fiducial argument beyond the limited scope of the context, as described by Buehler, one first has to recognize the idea underlying fiducial inference. In this respect, it can be helpful to look at the historical meaning of the word 'fiducial'. According to Stigler [86], the origin this word stems from surveying and astronomy:

> 'The inversion of probability statements involving the binomial distribution had proved to be a difficult step, but the same was far less true for problems of astronomical observation. (...) If e represents the error, x the observation, and θ the point observed, then $x = \theta + e$ implies equally as well that $\theta = x \Leftrightarrow e$. If e is taken as randomly and symmetrically distributed, then supposing θ fixed gives a distribution for x; and conversely, taking x as given leads to a distribution for θ. (...) R.A. Fisher was to call this a fiducial argument, borrowing a term for a fixed point from surveying and astronomy that suggested that a distance is the same regardless of which end point is fixed.'

(A slight notational modification has been made.) This means that the fiducial argument uses in essence the internal structure of the problem. That is, the fiducial argument uses explicitly the relation of how the observation x is constructed from a deterministic part θ and a random part e. An illustrative example is the following well–known problem.

The measurement problem. Suppose that the outcome $x \in \mathbb{R}^k$ of a vector a measurements X of a vector of physical constants $t \in \mathbb{R}^k$ has to be used to formulate an opinion about these constants. It is assumed that the measurement device is tested in a lab, and hence that it is known that its measurement error $e = x \Leftrightarrow t$ can be regarded as the outcome of a random variable E with a known distribution, say $N_k(0, \Sigma)$ with Σ known. Assume in addition that the distribution of E is the same, regardless of the value of t. This is the crucial assumption in this model. Rewriting the measurement equation $x = t + e$ to $t = x \Leftrightarrow e$ provides that, after observing x, any probability statement about $x \Leftrightarrow E$ implies a probability statement about t. Hence, the knowledge about t can be expressed in terms of the distributional inference

$$Q(x) = \mathcal{L}(x \Leftrightarrow E) = N_k(x, \Sigma).$$

Notice that in the case $k = 1$, the density g_x corresponding to this fiducial distribution, or structural distribution in the terminology of Fraser, satisfies, indeed,

$$g_x(\theta) = \Leftrightarrow d\Phi((x \Leftrightarrow \theta)/\sigma)/d\theta,$$

which is according to Fisher's prescription. ♣

Clearly, the idea of using the internal structure of the problem, represented by the measurement equation, to transfer the knowledge about a known error distribution and the observed outcome x to a probability distribution on Θ, is not restricted to the measurement problem. The general case will be worked out in Section 4.3. This approach to making inference is also known under the name functional approach, which refers to the specification of a functional equation, in contrast to the distributional approach, which refers to the specification of a family of distribution functions, that is commonly adopted in statistical theory. Before exploring this functional approach, first some problems concerning the noncoherence of fiducial probability will be investigated.

4.2 Noncoherence of fiducial probability

The fiducial argument produces distributional inferences, i.e., inferences in the form of probability distributions on Ψ. As such probability distributions do not have a direct frequency interpretation, it is natural to ask the question whether or not these fiducial probabilities possess the same coherency properties as ordinary probabilities. Recall from the previous section that, if $\psi : \Theta \mapsto \Psi$ is continuous and strictly increasing, then the fiducial distribution about $\psi(t)$ can be obtained from the fiducial distribution about t, by taking the measure induced by this transformation. In this sense, fiducial probabilities behave like ordinary probabilities. This, however, is not the case in general. From a mathematical point of view this form of probabilistic coherence of the inferences is elegant, and might seem natural. Making this requirement for a theory for making distributional inferences has, however, far going consequences. It should be noted, for example, that by applying nonmonotonic transformations, stochastic ordering relations can be destroyed. In such cases the inference problems change substantially, and hence the requirement of coherence under such transformations does not make any sense. Most problems occur in higher–dimensional problems. An extreme example of how the requirement of probabilistic coherence can lead to absurd results, was given by Stein [85].

The Stein example. Suppose that inference has to be made about the mean $t \in \mathbb{R}^k$ of a k–dimensional spherical normal distribution, i.e., it is assumed that $\mathcal{L}X_\theta = N_k(\theta, \mathrm{I}_k)$. On the basis of an observation $x \in \mathbb{R}^k$ the natural distributional inference about t can be derived, and is given by $Q(x) = N_k(x, \mathrm{I}_k)$. This distributional inference corresponds both to the fiducial inference if the fiducial argument is identified with the argument used in the measurement problem, and to the posterior distribution w.r.t. a noninformative prior, i.e., Lebesgue measure on Θ. Moreover, it can be obtained as a UMR equivariant procedure if an appropriate loss function is used. All together one can conclude that this inference is perfectly valid for this situation, although there could be some discussion about using improper priors.

Now, suppose that one is not interested in t itself, but in $\psi(t)$, where $\psi : \Theta \mapsto \mathbb{R}^+$ is given by $\psi(\theta) = \theta'\theta$. Accepting $Q(x) = N_k(x, \mathrm{I}_k)$ as appropriate inference about t, and conforming oneself to the requirement of probabilistic coherence, provides that one has to take the induced measure $Q(x) \circ \psi^{-1}$ as distributional inference about $\psi(t)$, i.e.,

$$\widetilde{Q}_\psi(x) = \chi'^2_{k,x'x}.$$

All information carried in x about $\psi(t) = t't$ is contained in $v = x'x$. Applying the fiducial argument to $V_\psi = X'_\theta X_\theta$ provides $Q_\psi(x)$ defined by its distribution functions

$$G_x(\theta) = \mathrm{P}(\chi'^2_{k,\psi} > x'x).$$

Now, notice that $\widetilde{Q}_\psi(x)$ and $Q_\psi(x)$ are clearly distinct, or more specifically $\widetilde{Q}_\psi(x,(0,\psi]) < Q_\psi(x,(0,\psi])$, for every $\psi \in \mathbb{R}^+$. If k increases, then the difference between the two inferences increases. Notice that Using the theory of the next chapter it can be shown that $\widetilde{Q}_\psi$ systematically overestimates the true value $\psi(t)$. This is a consequence of the improper prior on Θ, which, after the transformation ψ, assigns on Ψ more and more prior mass towards infinity. This overshadows the evidence. ♣

The reason why probabilistic coherence of distributional inferences can lead to absurd results is as follows. In the theory of statistical inference facts (the data) have to be intermingled with fictions (priors, optimality principles, etc.) in order to arrive at some solution to the problem. The choice of the additional ingredients (the fictions) is based on the fact that they are reasonable within the given context. Transforming the problem may change the context in such a way that these additional ingredients become totally unreasonable within the new context. Other inconsistencies of fiducial inferences in higher–dimensional problems can be found in Dempster [29], or concerning the parameters of multivariate normal distribution in Geisser–Cornfield [42].

Inconsistencies, however, do not only occur in higher–dimensional problems. A second example of the fact that fiducial probability is different from ordinary probability can be given in the context of incorporating additional information by restricting the parameter space. That is, suppose that additional information becomes available in the form $t \in \widetilde{\Theta} \subset \Theta$. If $Q(x) \in \Theta^*$ is regarded as a frequency–theoretic probability distribution, then by the axioms of probability theory it follows that given this information $\widetilde{Q}(x) \in \widetilde{\Theta}^*$ should be the renormalized restriction of $Q(x)$ to $\widetilde{\Theta}$. However, applying the fiducial argument, after obtaining this information, will provide that all credibility mass assigned by $Q(x)$ to the complement of $\widetilde{\Theta}$ will be concentrated on the boundary of $\widetilde{\Theta}$. This will be illustrated by the following example.

Incorporating additional information. Let $\mathcal{L}X_\theta = N(\theta,1)$, and suppose that on the basis of x inference has to be made about the unknown mean. Applying the fiducial argument provides, of course, $Q(x) = N(x,1)$. Now, suppose that additionally it is known that the mean is nonnegative, i.e., $t \in \widetilde{\Theta} = [0,\infty)$. Renormalizing the restriction of $Q(x)$ to $[0,\infty)$ provides the distributional inference $\widetilde{Q}(x)$, defined by the distribution function

$$\widetilde{G}_x(\theta) = \begin{cases} 0 & \text{if} \quad \theta < 0, \\ \frac{\Phi(\theta - x) - \Phi(-x)}{\Phi(x)} & \text{if} \quad \theta \geq 0. \end{cases}$$

Applying the fiducial argument w.r.t. $\mathcal{P} = \{\mathrm{P}_\theta : \theta \in [0, \infty)\}$ provides the distributional inference $Q(x)$, defined by the distribution function

$$G_x(\theta) = \begin{cases} 0 & \text{if} \quad \theta < 0, \\ \Phi(\theta \Leftrightarrow x) & \text{if} \quad \theta \geq 0. \end{cases}$$

Notice that this fiducial distribution assigns credibility mass $\Phi(\Leftrightarrow x)$ to $\{0\}$. In Section 5.6 this example will be revisited. ♣

To make the discussion about coherence somewhat more precise, the following notion of inconsistency, that was introduced by Stone [89] and further developed by Heath–Sudderth–Lane [45], [46], and [57], will be employed.

Definition 4.1 *A procedure $Q : \mathcal{X} \mapsto \Theta^*$ for making distributional inferences is said to be strongly inconsistent if*

$$\inf_{x \in \mathcal{X}} \int_\Theta \phi(x, \theta) \; Q(x, d\theta) > \sup_{\theta \in \Theta} \int_\mathcal{X} \phi(x, \theta) \; \mathrm{P}_\theta(dx),$$

for some bounded and measurable function $\phi : \mathcal{X} \times \Theta \mapsto \mathbb{R}$.

It can be shown that $Q : \mathcal{X} \mapsto \Theta^*$ cannot be strongly inconsistent if it can be approximated by posterior distributions w.r.t. proper priors in the following way. Let ν be a probability measure on Θ, and define $\mu_\nu(B) = \int_\Theta \mathrm{P}_\theta(B) \; \nu(d\theta)$, for all $B \in \sigma(\mathcal{X})$, then Q can be approximated by posterior distributions w.r.t. proper priors if

$$\inf_{\nu \in \Theta^*} \int_\mathcal{X} \|Q(x) \Leftrightarrow Q_\nu(x)\|_{\mathrm{TV}} \, \mu_\nu(dx) = 0,$$

where Q_ν is the posterior w.r.t. to the proper prior measure ν. Hence, fiducial inference will be coherent in this sense, if it coincides with a posterior distribution Q_ν, where in this case ν cannot be proper (see Section 4.4), and Q_ν is in its turn approximable by posterior distributions w.r.t. proper priors. Techniques for showing whether or not a posterior distribution w.r.t. an improper prior is approximable by posterior distributions w.r.t. proper priors, are usually based on truncation of improper priors, see e.g. Stone [87]. This extends results that Jeffreys [51] obtained for Student's problem. Another possibility is to use the invariance of the statistical model, in the case that this possible. In Section 5.6 this topic will be revisited.

4.3 Structured models

In statistical inference it is usual to start with the assumption that $\mathcal{L}X = \mathrm{P}$ is a member of some known family of probability distributions $\mathcal{P} = \{\mathrm{P}_\theta : \theta \in \Theta\}$. To obtain from this starting point a unique method of inference one has to adopt certain principles and add several additional ingredients. In the measurement problem it looked as if no such inputs were needed to obtain a unique method of inference. This was due to the extra information that was given a priori, namely the functional or structural equation that specified how the observation was constructed of an unknown deterministic component (the parameter) and a random component (the error) from a known distribution irrespective of the true value t of θ. Such a statistical model will be called a structured model. Notice that given such a functional equation and the error distribution one could uniquely determine $\mathcal{P}$, whereas given some $\mathcal{P}$ it is impossible to extract the functional equation. The term functional model was introduced by Bunke [19], but the underlying ideas were already implicitly present in the work of Fraser [40] on what he called structural inference. An extensive review of the functional approach to fiducial inference can be found in Dawid–Stone [25], and Dawid–Wang [26]. The ideas and terminology used in this section, are based on these two articles.

The ideas underlying the measurement problem will now be used to formulate general structured models. Keeping the measurement problem in mind, one can consider a model that states that the outcome x of the random variable X_θ (the measurement) is uniquely determined as a function of $\theta \in \Theta$ (the parameter) and the outcome e of the random variable E (the measurement error). Notice that the crucial underlying assumption of such a structured model is that the distribution of the measurement error does not depend on the value of the unknown parameter that is to be measured. To make this more precise, E is some random variable that assumes values in $\mathcal{E}$, and its distribution $\mathrm{P} = \mathcal{L}E$ is assumed to be known. For each pair of given $\theta \in \Theta$ and $e \in \mathcal{E}$, the outcome $x \in \mathcal{X}$ will be uniquely determined by some known function that maps $\Theta \times \mathcal{E}$ into $\mathcal{X}$. The notation $x = \theta e$ is used to denote the functional equation and reflects the algebraic structure of the problem, i.e., e is the function $e : \Theta \mapsto \mathcal{X}$ that acts on the right of θ and maps the parameter space into the outcome space, and similarly θ is the function $\theta : \mathcal{E} \mapsto \mathcal{X}$ that acts on the left of e and maps the measurement error space into the outcome space. Notice that in the case that $\mathcal{E} = \mathcal{X} = \Theta$ both e and θ can be regarded as transformations.

A structured model is said to be simple if for every pair x and e there exists a θ such that $x = \theta e$. It is said to be invertible if the functional equation is invertible, i.e., $x = \theta e$ has, for all pairs e and x, a unique solution θ. This solution θ will then be denoted by $\theta = xe^{-1}$. If a simple structured model is invertible, then given the outcome x the probability distribution P of E can be transfered to a fiducial distribution $Q_{\text{Fiducial}}(x)$ for t by defining

$$Q_{\text{Fiducial}}(x, B) = \mathrm{P}(\{e : xe^{-1} \in B\}) \quad \forall\, B \in \sigma(\mathcal{E}). \tag{4.1}$$

This will be abbreviated by using the notation $Q_{\text{Fiducial}}(x) = \mathcal{L}(xE^{-1})$. Notice that the measurement problem is an example of such a simple invertible structured model, and that the fiducial distribution that was, indeed, obtained by the same line of reasoning.

Now, consider the case that ψ is not the identity but any bijection $\psi : \Theta \mapsto \Psi$. Then the fiducial distribution is given by $Q_{\text{Fiducial}}(x) = \mathcal{L}\psi(xE^{-1})$. That the fiducial distribution cannot be defined in this fashion, if ψ is not a bijection, is illustrated by the Stein example. This example indicates that it is dangerous to treat fiducial probabilities carelessly as if they were frequency–theoretic probabilities. In Dawid–Stone [25], it is shown how the functional structure can be used to avoid these incoherences. For simple functional models, marginalization does not lead to inconsistencies provided that the following condition holds. Suppose that inference has to be made about $\psi(t)$, where $\psi : \Theta \mapsto \Psi$, and that there exists a function $v : \mathcal{X} \mapsto \mathcal{V}$ such that $v = \psi e$ is again a simple functional model. In this case it follows that $\psi(xE^{-1}) = v(x)E^{-1}$.

The question whether fiducial inference via the functional approach is, indeed, an extension of the fiducial argument, as given by Fisher, can be answered by checking that the fiducial distributions obtained by the functional approach, indeed, coincide with those derived by Fisher, in the case that both methods can be applied. The following simple structured models will reproduce the Fisher's original setting.

Lemma 4.1 *Let $\mathcal{X} \subset \mathbb{R}$, $\Theta \subset \mathbb{R}$, and $\mathcal{L}E = U(0,1)$. Assume that the functional relation $x = \theta e$ is such that θe is continuous, and strictly increasing in e, for all $\theta \in \Theta$, and continuous and strictly increasing in θ, for all $e \in \mathcal{E}$. Define $F_\theta(x) = \theta^{-1}x$. Then $Q_{\text{Fiducial}}(x)$ defined by (4.1) has a density, given by $g_x(\theta) = \Leftrightarrow dF_\theta(x)/d\theta$.*

Proof. Under the conditions of the lemma, $F_\theta(x)$ is continuous and strictly increasing in x, and continuous and strictly decreasing in θ. The distribution function corresponding to $Q_{\text{Fiducial}}(x)$ is given by

$$G_x(\theta) = \mathrm{P}(\{e : xe^{-1} \leq \theta\}) = \mathrm{P}(\{e : e \geq F_\theta(x)\}) = 1 \Leftrightarrow F_\theta(x).$$

Hence, $G_x(\theta)$ is continuous and strictly increasing in θ. Notice that the fact that for every pair x and e there exists a solution to the functional equation implies that $G_x(\theta) \to 0$ as $\theta \to \Leftrightarrow\infty$ and $G_x(\theta) \to 1$ as $\theta \to \infty$. Hence, $G_x(\theta)$ is a continuous distribution function, and its density w.r.t. Lebesgue measure $g_x(\theta)$ is given by $g_x(\theta) = dG_x(\theta)/d\theta = \Leftrightarrow dF_\theta(x)/d\theta$, which coincides with the fiducial density provided by Fisher. ■

To go beyond the context of simple structured models without making any essential adaptations, the following definitions are introduced. If there exists a $\theta \in \Theta$ such that the functional equation $x = \theta e$ has a solution for given $x \in \mathcal{X}$ and $e \in \mathcal{E}$, then x and e are called compatible. The set of all e that are compatible with a fixed x will be denoted by $\mathcal{E}_x = \{e \in \mathcal{E} : \exists\, \theta \in \Theta \text{ s.t. } x = \theta e\}$. Similarly, one can define $\mathcal{X}_\theta$ as the set of all x compatible with θ, and Θ_x as the set of all θ compatible with x. To be able to derive the fiducial distribution, it is essential that the functional equation is invertible, i.e., for all $x \in \mathcal{X}$ and $e \in \mathcal{E}_x$ the solution $\theta = xe^{-1}$ of $x = \theta e$ has to be unique. Given some invertible structured model, the fiducial 'inversion' can be made as follows. By observing x, one can logically conclude that $e \in \mathcal{E}_x$, and hence one has to condition on this event. Notice that this is also the only information about the outcome of E that can be obtained from observing x. Hence, define, for every $x \in \mathcal{X}$, a random variable E^x on $\mathcal{E}_x$, such that $\mathcal{L}E^x = \mathrm{P}^x$, where P^x denotes the restriction of P to $\mathcal{E}_x$. The fiducial distribution generated by the functional model is then defined on Θ_x and given by $Q_{\text{Fiducial}}(x) = \mathcal{L}(x[E^x]^{-1})$, i.e.,

$$Q_{\text{Fiducial}}(x, B) = \mathrm{P}^x(\{e : xe^{-1} \in B\}) \quad \forall\, B \in \sigma(\mathcal{E}_x). \tag{4.2}$$

A natural generalization of the simple structured models are the so called partitionable structured models. Not much is changed if the assumption that $\mathcal{E}_x = \mathcal{E}$, for all $x \in \mathcal{X}$, is replaced by the assumption that for $x_1, x_2 \in \mathcal{X}$ either $\mathcal{E}_{x_1} = \mathcal{E}_{x_2}$, or $\mathcal{E}_{x_1} \cap \mathcal{E}_{x_2} = \emptyset$. A structured model satisfying this condition is said to be partitionable. In the case that a functional model is partitionable, there exist functions a defined on $\mathcal{X}$ and b defined on $\mathcal{E}$ such that $a(x) = b(e)$ if and only if $e \in \mathcal{E}_x$. Notice that, for such a function a, $\mathcal{L}a(X_\theta)$ does not depend on θ, and can be called a functional ancillary. The idea underlying the analysis of such partitionable functional models is that, conditionally on $a(X_\theta) = a$, the situation is the same as in the simple functional model. Denote $\mathrm{P}^a = \mathcal{L}(E|b(E) = a)$, and define a random variable E^a on Θ_x such that $\mathcal{L}E^a = \mathrm{P}^a$, then the fiducial distribution for the partitionable functional model is given by $Q_{\text{Fiducial}}(x) = \mathcal{L}(x[E^a]^{-1})$.

There are two important special cases of structured models, each highlighting different aspects of structured models. The first class consists of the so called pivotal models, see e.g. Barnard [3], the second class of the so called structural models, see e.g. Fraser [40]. These models will be treated in more detail in the next section.

4.4 Pivotal and structural models

Pivotal methods are commonly used in statistics. A statistical methodology based on pivotal methods was introduced by Barnard [3], [4]. A function u of both the parameter and the observation is said to be a pivot or pivotal function if its distribution does not depend on the unknown parameter θ, i.e., $\mathcal{L}u(X_\theta, \theta) = \mathrm{P}$. The idea behind pivotal inference is to introduce a random variable U, such that $\mathcal{L}U = \mathrm{P}$, then equate $u(X_\theta, \theta) = U$, take l.h.s. of the equation with x fixed, i.e., $u(x, \theta) = U$ for the observed x, and finally invert the function u in its second argument, i.e., $\theta = u^{-1}(x, U)$. The distributional inference arising from such a pivotal argument is then, of course,

$$Q(x) = \mathcal{L}u^{-1}(x, U).$$

To relate this pivotal argument to structured models, consider the following line of thought. Consider the set of all pairs (x, θ) that are compatible, i.e., $\cup_{x \in \mathcal{X}}(x, \Theta_x) = \cup_{\theta \in \Theta}(\mathcal{X}_\theta, \theta) \subset \mathcal{X} \times \Theta$. Assume that, for all these pairs (x, θ), the functional equation $x = \theta e$ can be solved explicitly for e, and denote this solution by $e = u(x, \theta)$. Now, notice that this function $u : \cup_{x \in \mathcal{X}}(x, \Theta_x) \mapsto \mathcal{E}$ is a pivotal function, because $\mathcal{L}u(X_\theta, \theta) = \mathcal{L}E = \mathrm{P}$, and hence does not depend on the unknown parameter θ. In case that u can be inverted explicitly in its second argument, the fiducial distribution is given by

$$Q_{\text{Fiducial}}(x) = \mathcal{L}(u^{-1}(x, E)). \tag{4.3}$$

An extension of the use of pivotal methods is given in Weerahandi [93], who considers so called generalized pivotal functions which are functions of the unknown parameter θ, the random variable X_θ, and the observed value of the random variable x. The problem with applying pivotal methods, if no structural equation is given, is that in higher–dimensional problems, the choice of pivots is often nonunique. In these cases, different pivotal functions will often lead to different inferences. Fraser [39], [38] investigated under which conditions the choice of pivot is essentially unique. This resulted in the theory of the so called structural models. These structural models are the basis for the idea behind functional models. Structural models were introduced by Fraser [39], and for an extensive treatment of these structural models one can consult Fraser [40].

Suppose that G is a unitary transformation group working on the left of $\mathcal{X}$, and that $\mathcal{P} = \{\mathrm{P}_\theta : \theta \in \Theta\}$ is generated by this group, i.e., it is possible to identify $\mathcal{X}$ with G, and Θ with $\overline{\mathrm{G}}$. To see that such an invariant model is a functional model, notice that $\mathcal{X} = \mathcal{E}$, and that the functional equation $x = \theta e$, which is called the structural equation in this case, corresponds to a transformation of $\mathcal{X}$ by one of the elements of the group. Hence, structural models are in some sense also pivotal models; the pivot is given by $u(x, \theta) = \mathrm{g}_{\theta^{-1}} x$. The fiducial or in this case structural distribution is given by

$$Q_{\text{Structural}}(x) = \mathcal{L}(\mathrm{g}_x \mathrm{g}_{E^{-1}}(x_{\mathrm{e}})),$$

which is clearly equivariant under the working of G. In general, the analysis of structural models proceeds as follows. First, notice that structural models are partitionable functional models; the orbits of G form a disjunct partition of $\mathcal{X}$. By definition x and e are on the same orbit of G. Hence, after observing x one has to condition on the orbit, because this is the only information that is obtained about the outcome of E. A maximal invariant statistic $v : \mathcal{X} \mapsto \mathcal{V}$ is a functional ancillary in this context, and the analysis proceeds as prescribed in the previous section for partitionable functional models.

Now, suppose that the structural model satisfies all regularity conditions from Section 3.5 (the case that $\overline{\mathrm{G}}$ works transitively on Θ). In Fraser [38] it is shown that

$$Q_{\text{Fiducial}} = Q_{\text{Structural}} = Q_{\nu^r},$$

where ν^r denotes the induced right Haar measure. Or in other words, for an invariant problem where $\overline{\mathrm{G}}$ works transitively on Θ, the fiducial inference coincides with a posterior distribution w.r.t. the right Haar measure. An interesting question in this respect is whether or not there can be found other situations where the fiducial distribution coincides with a (formal) posterior distribution. For the one–dimensional problem the answer to this question was given by Lindley [59].

Theorem 4.1 *Suppose that $\mathcal{X} \subset \mathbb{R}$ and $\Theta = (\underline{\theta}, \bar{\theta}) \subset \mathbb{R}$. Let the distribution functions $F_\theta(x)$ be such that $dF_\theta(x)/d\theta$ exists and $\lim_{\theta \to \bar{\theta}} F_\theta(x) = 0$ and $\lim_{\theta \to \underline{\theta}} F_\theta(x) = 1$. Then there exists a measure ν such that $Q_\nu(x) = Q_{\text{Fiducial}}(x)$ if and only if there exist monotone transformations $u : \mathcal{X} \mapsto \mathbb{R}$ and $\eta : \Theta \mapsto \mathbb{R}$ such that $\eta(\theta)$ is a location parameter for $u(X_\theta)$, i.e., $\mathcal{L}u(X_\theta) = \mathcal{L}(\eta(\theta) + U)$ where U is a random variable with a fixed distribution.*

That this requirement is rather restrictive can be illustrated by considering the case of 1–parameter exponential families. It can be shown that the only two exponential families that satisfy the conditions of the theorem are the normal distributions with known variance (location family), and the gamma distributions (scale family). An immediate consequence of the fact that often fiducial inference does not coincide with Bayesian inference is that, in general, it should not be expected that fiducial inference produces admissible procedures. Lindley's result does not extend to higher dimensions: in Brillinger [15] an example is given, where there exits a measure ν such that Q_ν and Q_{Fiducial} coincide, but where it can be shown that there does not exist a group G which leaves the model $\mathcal{P}$ invariant.

If the group $\overline{\mathrm{G}}$ is noncompact, then $\nu^r(\Theta)$ is infinite. Hence, in the case that the model is invariant under a noncompact transitive group $\overline{\mathrm{G}}$, the question whether or not $Q_{\text{Fiducial}} = Q_{\nu^r}$ is approximable by posteriors w.r.t. proper priors, and hence whether it is consistent in the sense of Heath–Sudderth–Lane, is of interest. Provided that the sampling densities $\mathrm{p}_\theta(x) > 0$, for all $x \in \mathcal{X}$, and $\theta \in \Theta$, it can be shown that Q_{ν^r} is consistent if and only if G is amenable, see Stone [88]. Moreover, it is shown in this paper that Q_{ν^r} is the only equivariant procedure that is consistent. To conclude with, all approaches will be illustrated by the following well–known problem.

Student's problem The following problem initiated the modern theory of exact small–sample inference. Suppose that the outcome $x = (x_1, \ldots, x_n)$ of an independent random sample $X = (X_1, \ldots, X_n)$ from $N(\mu, \sigma^2)$ is observed, where it is assumed that a priori nothing is known about the true value (μ, σ^2) of the unknown parameter $(\theta_1, \theta_2) \in \mathbb{R} \times \mathbb{R}^+$. And suppose that distributional inference has to be made about the unknown mean μ. Use the following standard notation:

$$\bar{X} = n^{-1} \textstyle\sum_{i=1}^n X_i, \quad S^2 = (n \Leftrightarrow 1)^{-1} \sum_{i=1}^n (X_i \Leftrightarrow \bar{X})^2 \ ,$$

and small letters to denote the corresponding observed outcomes of these statistics. Student stipulated and Fisher proved that the random variable $T = \sqrt{n}(\bar{X} \Leftrightarrow \mu)/S$ has a distribution which is know as Student's t_{n-1}, at least if $n \geq 2$, and hence is a pivotal quantity. Student used this result to think of μ in terms of the distributional inference

$$Q_{\text{Student}}(x) = \mathcal{L}(\bar{x} + n^{-\frac{1}{2}} s T_{n-1}),$$

where T_{n-1} is a random variable such that $\mathcal{L}T_{n-1} = t_{n-1}$, see Student [91]. This, of course, coincides with the fiducial distribution obtained by a pivotal argument (4.3). Jeffreys [51] obtained the distributional inference $Q_{\text{Student}}(x)$ as the marginal posterior distribution of μ by taking the limit of posterior distributions, w.r.t. uniform priors on growing rectangles of μ and $\log(\sigma)$, for a fixed x. Stone [88] showed that this approximation also holds in terms of the definition given in Section 4.2. This is, of course, not surprising because Q_{Student} is the marginal posterior distribution of μ w.r.t. the right Haar measure σ^{-1} on the location scale group works transitively on the parameter space. Moreover, the location scale group is amenable, see e.g. Bondar–Milnes [13]. So it follows that Student's inference is consistent in the sense of Heath–Sudderth–Lane.

Fisher was fascinated by the exactness of this result. He noted that the fiducial limits $\bar{x} \pm n^{-\frac{1}{2}} s t_{n-1}(\frac{1}{2}\epsilon)$ are correct in the sense that the corresponding stochastic intervals have the right coverage probability ϵ, and he regarded this as an example of the logical validity of an inductive method. Wallace wrote the following about Student's inference:

> 'Student's work is accepted and recognized as basic by all. Setting aside doubts on the Gaussian and independence assumptions, we have wide acceptance for inferential statements in the form of 95% limits (and of other levels) on the unknown mean μ even if we cannot agree on the adjective ('fiducial' or 'confidence', author's explanation) modifying 'limits'. The uncertainty about μ is conveniently and at least schematically represented by a t–distribution centered at the observed mean and scaled by the estimated standard error of the mean, with limits by the appropriate fractiles.'

Fisher knew that the distribution functions $G_{\text{Student},x}$ are such that

$$\mathcal{L} G_{\text{Student},X_\theta}(\theta_1) = U(0,1) \quad \forall\, \theta \in \Theta,$$

see Fisher [37], which is equivalent to the exactness on the confidence intervals. This pivotal relation will play an important role in the theory of fiducial inference that will be presented in the next chapter; in Section 5.7 this problem will be revisited. ♣

For Student's problem it turned out that there are different ways to obtain Student's inference as a distributional inference about μ. In other words, all different methods lead to the same result. In the next example it is shown that in higher dimensions the different approaches can, and will, all lead to different results.

Multivariate Student's problem Consider Student's problem again, with the difference that X_i is now a random variable with a $N_k(\mu, \Sigma)$ distribution, where $\mu \in \mathbb{R}^k$ is the vector of means, and Σ is the $k \times k$ variance–covariance matrix. Assume that the number of observations is lager than the dimension of the problem, i.e., $n > k$. Let $\bar{X} = (\bar{X}_1, \ldots, \bar{X}_k)'$ denote the vector of sample means, and $S = \{s_{i,j}\}_{i,j}$ the sample variance–covariance matrix. For deriving the joint fiducial distribution of the means, Fisher proposed to use in the bivariate case the following approach: (i) use the sample correlation coefficient to compute the fiducial density for the correlation coefficient, next (ii) invert the conditional distribution of S_1 and S_2 given r and ρ, then (iii) invert the distribution of $\bar{X}_1$ and $\bar{X}_2$, given Σ, and finally (iv) compute the marginal distributions of the means. Applying this scheme for the k–variate problem, Bennet–Cornish [7] obtained the following fiducial density for μ

$$\begin{aligned} q_{\mathrm{Fiducial},x}(\mu) &= \left(\frac{n}{(n \Leftrightarrow 1)\pi}\right)^{\frac{1}{2}k} \frac{,\left(\frac{1}{2}(n+k \Leftrightarrow 1)\right)}{,\left(\frac{1}{2}(n \Leftrightarrow 1)\right)|S|^{\frac{1}{2}}} \\ &\quad \times \left\{1 + \frac{n(\bar{x} \Leftrightarrow \mu)' S^{-1} (\bar{x} \Leftrightarrow \mu)}{n \Leftrightarrow 1}\right\}^{-\frac{1}{2}(n+p-1)}. \end{aligned} \tag{4.4}$$

Cornish showed that the quantity $T^2 = n(\bar{x} \Leftrightarrow \mu)' S^{-1}(\bar{x} \Leftrightarrow \mu)$, where μ is the variable and $\bar{x}$ and S are fixed, is distributed like $kF_{k,n-1}$. This differs from the density T^2 as given by Hotelling, where $\bar{X}$ and S are variable and μ is fixed, which is distributed like $k(n \Leftrightarrow 1)/(n \Leftrightarrow k)F_{k,n-k}$. Following a Bayesian approach, Geisser–Cornfield [42] obtained that, using

$$\nu_v(d\mu, d\Sigma) = |\Sigma|^{-\frac{1}{2}v} d\mu d\Sigma$$

as prior, the marginal posterior of μ is given by

$$\begin{aligned} q_{\nu_v,x}(\mu) &= \left(\frac{n}{(n \Leftrightarrow 1)\pi}\right)^{\frac{1}{2}k} \\ \times \quad & \frac{,\left(\frac{1}{2}(n+k \Leftrightarrow v+1)\right)}{,\left(\frac{1}{2}(n \Leftrightarrow v+1)\right)|S|^{\frac{1}{2}}} \left\{1 + \frac{n(\bar{x} \Leftrightarrow \mu)' S^{-1}(\bar{x} \Leftrightarrow \mu)}{n \Leftrightarrow 1}\right\}^{-\frac{1}{2}(n+p-v+1)}, \end{aligned} \tag{4.5}$$

which reduces for $v = 2$ to (4.4), and for $v = k+1$ will have the usual Hotelling density of T^2. The choice $v = k+1$ is appealing from the point of view of invariance. Let G be the group with elements (A, b), where A is a $k \times k$ nonsingular matrix, and b a $k \times 1$ vector. The group operation is

$$(A_1, b_1)(A_2, b_2) = (A_1 A_2, b_1 + b_2 A_2'),$$

where the prime denotes transpose. Then G works on the sample space via

$$(S, \bar{x}) \mapsto (ASA', \bar{x}A' + b),$$

and it is easy to check that this is, indeed, a left action. The induced left action of G on the parameter space is $\overline{\mathrm{G}}$ given by

$$(\Sigma, \mu) \mapsto (A\Sigma A', \mu A' + b).$$

With these two group actions it is clear that the model is invariant. Moreover, the group acts transitively over the parameter space. The induced invariant Haar measure on the parameter space is

$$\nu(d\mu, d\Sigma) = |\Sigma|^{-\frac{1}{2}(k+1)} d\mu d\Sigma.$$

The Bayes procedure w.r.t. to this invariant prior is given by (4.5), with $v = k + 1$. Notice that it has the usual Hotelling density of T^2.

In a series of papers by Eaton–Sudderth [32], [33], and [34] the use of this group G is criticized, because it is not amenable. Consider the subgroup H of G, with elements (T, b), where T is a lower triangular matrix whose diagonal elements are positive. It can be shown that $\overline{\mathrm{H}}$ also acts transitively on the parameter space. Notice that, for each positive definite Σ, there exists a unique $, \ \in \overline{\mathrm{H}}$ such that

$$\Sigma = \, , \, , '.$$

Hence, it is possible to reparameterize the multivariate normal distribution in terms of $(, \, , \mu)$. For this parameterization, $\overline{\mathrm{H}}$ acts on the parameter space by

$$(, \, , \mu) \mapsto (T, \, , \mu T' + b).$$

The subgroup H of G is amenable, see e.g. Bondar–Milnes [13]. The left Haar measure induced by H on the parameter space coincides with the standard Jeffreys prior, and provides (4.5) with $v = k+1$ as posterior distribution. Recall from Section 4.2, that an equivariant posterior can only be consistent in the sense of Heath–Sudderth–Lane if the right Haar measure of an amenable group is used as prior. Using the $(, \, , \mu)$ parameterization of the normal distributions, the right Haar measure of H on the parameter space is given by

$$\nu^r(d\mu, d, \,) = \frac{d\mu d,}{\prod_{i=1}^{k} \gamma_{i,i}^{k-i+1}},$$

where $\gamma_{i,i}$ denotes the ith diagonal element of , . Let G_T^+ denote the group of all lower triangular matrices with positive diagonal elements. Then the marginal posterior distribution of μ, in the case that ν^r is used as prior, is proportional to the integral

$$\int_{G_T^+} |,|^{-n} \exp\left\{\Leftrightarrow\frac{1}{2}\mathrm{tr}((,\,,')^{-1}((n\Leftrightarrow 1)S + n(\bar{x}\Leftrightarrow\mu)(\bar{x}\Leftrightarrow\mu)'))\right\}\nu^r(\mu, d,\,).$$

To compute this integral, introduce the following notation. Let τ be the $k \times k$ lower triangular matrix, such that

$$\tau\tau' = (n\Leftrightarrow 1)S + n(\bar{x}\Leftrightarrow\mu)(\bar{x}\Leftrightarrow\mu)'.$$

Making a transformation of variables, and using a result that can be found in Eaton–Sudderth [33], it turns out that the integral is equal to

$$|(n\Leftrightarrow 1)S + n(\bar{x}\Leftrightarrow\mu)(\bar{x}\Leftrightarrow\mu)'|^{-\frac{1}{2}n}\frac{c(n)}{\Delta(\tau)},$$

where $c(n)$ is a constant that only depends on n, and Δ is the modular function of G_T^+, i.e.,

$$\Delta(\tau) = \prod_{i=1}^{k} \tau_{i,i}^{k-2i+1}.$$

The following step is, of course, to express the diagonal elements of τ in terms of x and μ. To do so, observe that $(n\Leftrightarrow 1)S + n(\bar{x}\Leftrightarrow\mu)(\bar{x}\Leftrightarrow\mu)' = Y'Y$, where

$$Y = \begin{pmatrix} x_{1,1}\Leftrightarrow\mu_1 & \cdots & x_{1,k}\Leftrightarrow\mu_k \\ \vdots & & \vdots \\ x_{n,1}\Leftrightarrow\mu_1 & \cdots & x_{n,k}\Leftrightarrow\mu_k \end{pmatrix} = (y_1, \ldots, y_k),$$

and $y_i = (x_{1,i}\Leftrightarrow\mu_i, \ldots, x_{n,i}\Leftrightarrow\mu_i)'$. Now, notice that $Y = Q\tau'$, where the columns of Q are the orthonormal vectors

$$\begin{aligned} q_1 &= \frac{y_1}{\sqrt{\langle y_1, y_1\rangle}}, \\ q_2 &= \frac{y_2 - \langle y_2, q_1\rangle q_1}{\sqrt{\langle y_2 - \langle y_2, q_1\rangle q_1, y_2 - \langle y_2, q_1\rangle q_1\rangle}}, \\ &\vdots \\ q_k &= \frac{y_k - \langle y_k, q_1\rangle q_1 - \cdots - \langle y_k, q_{k-1}\rangle q_{k-1}}{\sqrt{\langle y_k - \langle y_k, q_1\rangle q_1 - \cdots - \langle y_k, q_{k-1}\rangle q_{k-1}, y_k - \langle y_k, q_1\rangle q_1 - \cdots - \langle y_k, q_{k-1}\rangle q_{k-1}\rangle}}, \end{aligned}$$

that can be obtained by applying a Gramm–Schmidt process to the columns of Y. Now, write $Q'Y = \tau'$, to obtain that the diagonal elements of τ can be expressed as

$$\begin{aligned}
\tau_{1,1} &= \sqrt{\langle y_1, y_1\rangle}, \\
\tau_{2,2} &= \sqrt{\tfrac{\langle y_2,y_2\rangle\langle y_1,y_1\rangle - \langle y_1,y_2\rangle^2}{\langle y_1,y_1\rangle}}, \\
&\vdots \\
\tau_{i,i} &= \sqrt{\tfrac{|(y_1,\ldots,y_i)'(y_1,\ldots,y_i)|}{|(y_1,\ldots,y_{i-1})'(y_1,\ldots,y_{i-1})|}}, \\
&\vdots \\
\tau_{k,k} &= \sqrt{\tfrac{|(y_1,\ldots,y_k)'(y_1,\ldots,y_k)|}{|(y_1,\ldots,y_{k-1})'(y_1,\ldots,y_{k-1})|}}.
\end{aligned}$$

This yields that the modular function can be expressed in terms of

$$\Delta(\tau) = \frac{|(y_1)'(y_1)|\cdots|(y_1,\ldots,y_{k-1})'(y_1,\ldots,y_{k-1})|}{|(n \Leftrightarrow 1)S + n(\bar{x} \Leftrightarrow \mu)(\bar{x} \Leftrightarrow \mu)'|^{-\frac{1}{2}(k-1)}}.$$

Now use the equality

$$\frac{|(n \Leftrightarrow 1)S|}{|(n \Leftrightarrow 1)S + n(\bar{x} \Leftrightarrow \mu)(\bar{x} \Leftrightarrow \mu)'|} = \left\{1 + \frac{n(\bar{x} \Leftrightarrow \mu)'S^{-1}(\bar{x} \Leftrightarrow \mu)}{n \Leftrightarrow 1}\right\}^{-1},$$

to show that the marginal posterior distribution of μ, in the case that ν^r is used as prior distribution, is given by

$$\begin{aligned}
&q_{\nu^r,x}(\mu) = \\
&\quad \left(\frac{n}{(n \Leftrightarrow 1)\pi}\right)^{\frac{1}{2}k} \frac{,\left(\frac{1}{2}(n + k \Leftrightarrow 1)\right)}{,\left(\frac{1}{2}(n \Leftrightarrow 1)\right)|S|^{\frac{1}{2}}} \left\{1 + \frac{n(\bar{x} \Leftrightarrow \mu)'S^{-1}(\bar{x} \Leftrightarrow \mu)}{n \Leftrightarrow 1}\right\}^{-\frac{1}{2}(n+p-1)} \\
&\quad \times c(n,k) \prod_{i=1}^{k} \frac{|(n \Leftrightarrow 1)S + n(\bar{x} \Leftrightarrow \mu)(\bar{x} \Leftrightarrow \mu)'|}{(|I^{(i)})'\{(n \Leftrightarrow 1)S + n(\bar{x} \Leftrightarrow \mu)(\bar{x} \Leftrightarrow \mu)'\}I^{(i)}|}, \qquad (4.6)
\end{aligned}$$

where $I^{(i)} = (\mathrm{I}_i, 0)'$, with I_i the $i \times i$ identity, and 0 a $(k \Leftrightarrow i) \times i$ zero matrix, and $c(n,k)$ a constant that only depends on k and n. Notice that the first term of this density is equal to Fisher's fiducial density, and that the second term introduces some kind of asymmetry in the means. It can be concluded that the three different approaches (4.4), (4.5), and (4.6) only coincide in the univariate case, i.e. $k = 1$. ♣