6

Some extensions and applications

At the end of the previous chapter several examples were considered where best strongly unbiased procedures could be derived by applying the fiducial argument. These fiducial procedures were compared to alternatives arising from the application of other paradigms. In this chapter some problems that are connected with Bayes’s problem will be discussed. As Bayes’s problem is discrete, no strongly unbiased procedure will exist, and hence optimality of the fiducial procedure, in the sense of uniformly minimum risk within this restricted class of procedures, cannot be obtained. Nevertheless, it will be shown that the fiducial procedure is certainly not unreasonable, and can be used as a starting point in the related problems. Besides investigating the exact (small sample) properties, also some attention will be paid to asymptotic (large sample) properties of the procedures.

6.1 Bayes’s problem

A crucial role in the history of statistical inference has been played by Bayes’s problem, posthumously published by Price in 1763 and brought to the attention of modern generations by Barnard [2]. The problem is formulated as follows.

Given the number of times in which an unknown event has happened and failed.
Required the chance that the probability of its happening in a single trial lies somewhere between any two degrees of probability that can be named.

Bayes himself invoked a physical context with a compelling solution. With respect to this context the chance \( q_{a,b} \) of \( a < p < b \) is given by

\[
q_{a,b} = \binom{n}{x} \int_a^b \theta^x (1 - \theta)^{n-x} d\theta. \tag{6.1}
\]

Bayes worried about the applicability of this result if the physical context was different from the one he specified. Fisher suggested that this was the reason for delaying publication. Writing under the pseudonym John Noon, however, Bayes remarked that the mathematician is free to specify any contextual ingredient as long as this is not in conflict with the facts. The physical context chosen by Bayes is only one of the many possibilities one could invoke. In the absence of any further information, Bayes’s problem does not have a compelling solution. The situation described by it is paradigmatic for almost every situation from statistical practice. The fact that compelling solutions do not exist, does not imply that the task of making statistical inferences should be ignored. One should realize, however, that all one can achieve is to have a discussion involving various proposals, a comparative analysis and, at the end of the day, a choice of some specific procedure. The first step towards a comparative analysis, is to rephrase Bayes’s problem into modern terms:

Given the outcome \( x \in \{0, \ldots, n\} \) of \( X \), where \( \mathcal{L}X = \text{Bin}(n,p) \).

Required a distributional inference about \( p \), i.e., an opinion in the form of a probability measure \( Q = Q(x) \) on \([0,1]\).

With respect to the original formulation, two modifications have been made: the design is specified and, more interestingly, the requirement is made that the chances \( q_{a,b} \) of \( a < p \leq b \) are probabilistically coherent; \( q_{a,b} = G(b) \Leftrightarrow G(a) \) for some probability distribution function \( G = G_x \).

The procedures When constructing such a probability distribution function \( G \), there are two natural requirements that should be taken into account. The first one is that the procedure \( Q \) should be equivariant under inversion, i.e.,

\[
G_x(\theta) + G_{n-x}(1 \Leftrightarrow \theta) = 1,
\]
for \( \theta \in (0,1) \); it should not matter whether successes are counted or failures. The second requirement is that no positive credibility should be assigned to impossible cases, i.e.,

\[
x > 0 \Rightarrow G_x(0) = 0 \quad \text{and} \quad x < n \Rightarrow G_x(1-) = 1.
\]

Keeping these two requirements in mind, various procedures will be considered.

Bayes's physical context implies that, a priori, \( p \) is the outcome of a uniformly distributed random variable \( T \). Such uniform prior probability provides the procedure

\[
Q_{\text{Bayes}}(x) = \text{Beta}(x + 1, n \Leftrightarrow x + 1),
\]

because this is the posterior distribution of \( T \) given \( x \), which, of course, corresponds to (6.1). In the Bayesian literature such uniform prior is motivated by the fact that it expresses the lack of prior information about the location of the true success probability \( p \). Notice that (6.2) appears by normalizing the likelihood function. As the uniform distribution can be questioned, a variation on this idea was given by Jeffreys [50], who suggested to use the square root of the expected Fisher information as noninformative prior. An interesting feature of applying Jeffreys's principle is that, in contrast to taking a uniform prior, it is not affected by smooth reparameterizations. This follows from the fact that, for smooth changes of parameterization \( \psi(\theta) \), the prior density satisfies

\[
n(\theta) = n(\psi(\theta))|\det(d\psi/d\theta)|.
\]

Although Jeffreys himself did not like the improper prior \( \text{Beta}(\frac{1}{2}, \frac{1}{2}) \) for Bayes's problem (Zellner, personal communication), the procedure obtained by applying Jeffreys's principle to obtain a prior and then computing the posterior probability, will be called

\[
Q_{\text{Jeffreys}}(x) = \text{Beta} \left( x + \frac{1}{2}, n \Leftrightarrow x + \frac{1}{2} \right).
\]

Notice that if a proper loss function is used to evaluate the inferences, then \( Q_{\text{Bayes}} \) and \( Q_{\text{Jeffreys}} \) are (formal) Bayes rules, and hence, provided that their Bayes risks are finite, they are admissible. Neither \( Q_{\text{Bayes}} \) nor \( Q_{\text{Jeffreys}} \) assigns positive credibility to any particular point \( \theta \). Hence, the second requirement of not assigning positive credibility to impossible cases is automatically satisfied. Another consequence of the fact that they do not assign positive credibility to any particular point \( \theta \) is that these procedures cannot be weakly unbiased. By a symmetry argument it follows that both procedures are equivariant under inversion, which qualifies them at least as reasonable candidates.

Instead of using Bayes's Theorem to generate distributional inferences about \( p \), one could alternatively apply the fiducial argument using the symmetrized
P-value. This provides the procedure

\[ Q_{\text{Fiducial}}(x) = \frac{1}{2}\text{Beta}(x + 1, n \Leftrightarrow x) + \frac{1}{2}\text{Beta}(x, n \Leftrightarrow x + 1), \]  

which by Lemma 3.7 can be shown to be weakly unbiased. By definition, \( \text{Beta}(0, b) = \delta_0 \), for \( b > 0 \), and \( \text{Beta}(a, 0) = \delta_1 \), for \( a > 0 \). Notice that the constant weights \( \frac{1}{2} \) of the beta distributions imply that \( Q_{\text{Fiducial}} \) cannot be obtained as a posterior distribution with respect to any prior. Positive credibility is assigned to the point \( \{0\} \) if \( x = 0 \), and to the point \( \{1\} \) if \( x = n \); this is not unreasonable because if only successes or failures are observed then the possibility that one is observing a deterministic sequence cannot be excluded. Obviously, \( Q_{\text{Fiducial}} \) does not assign positive credibility to impossible cases, and it is not difficult to see that it is equivariant under inversion.

It might be an interesting idea to combine the Fisherian and Bayesian approaches. This means that instead of equating \( G_x(\theta) = \alpha_{sp, \theta}(x) \), one could also use some posterior probability as a degree of belief in the hypothesis. An example of such an approach is the following. Assign prior probability \( \frac{1}{2} \) to both \( H_0 : p \leq \theta \) and \( A_0 : p > \theta \), and spread this prior probability mass according to a uniform distribution on both \( \Theta_{H_0} = [0, \theta] \) and \( \Theta_{A_0} = (\theta, 1] \). This prior probability measure will be denoted by \( \nu_\theta \). Equating \( G_x(\theta) = \alpha_{\nu_\theta}(x) \), where \( \alpha_{\nu_\theta}(x) \) is the corresponding posterior probability of the hypothesis, provides

\[ G_{\text{semi-Bayes}, x}(\theta) = \frac{1}{1 + B_x(\theta)}, \]  

where \( B_x(\theta) = \frac{\theta}{1 - \theta} \frac{1 - G_{\text{Bayes}, x}(\theta)}{G_{\text{Bayes}, x}(\theta)} \). It, of course, remains to verify that the procedure \( Q_{\text{semi-Bayes}} \), defined by (6.5), is really a procedure for making distributional inferences, i.e., it remains to verify that \( G_{\text{semi-Bayes}, x} \) is, indeed, a distribution function, for all \( x = \{0, 1, \ldots, n\} \).

**Lemma 6.1** Let \( G \) be the distribution function of the \( \text{Beta}(a, b) \) distribution, with \( a, b \geq 1 \) integer valued. Then

\[ \tilde{G}(z) = \frac{1}{1 + B(z)}, \]  

where \( B(z) = \frac{z}{1 - z} \frac{1 - G(z)}{G(z)} \), is a probability distribution function on \([0, 1]\).

The proof of this lemma can be found in the appendix. From this lemma it follows immediately that \( Q_{\text{semi-Bayes}} \) is, indeed, a procedure for making distributional inferences. Clearly \( Q_{\text{semi-Bayes}} \) is not a Bayes procedure because the
prior depends on $\theta$. It turns out that it is not weakly unbiased. It is, however, equivariant under inversion, and does not assign positive credibility to impossible cases. Using the same idea, but instead of assigning prior probabilities $\frac{1}{2}$ to both $H_0$ and $A_0$, construct a weakly unbiased procedure by assigning appropriate prior probabilities. This can be done by employing the following strategy. Assigning prior probability $c(\theta) \in (0,1)$ to $H_0$, $1 \leftrightarrow c(\theta)$ to $A_0$, spreading this mass uniformly, and equating $G_x(\theta)$ to the corresponding posterior probability of $H_0$, provides

$$G_{\text{Unbiased},x}(\theta) = \frac{1}{1 + B_x(\theta)},$$

(6.7)

where $B_x(\theta) = \frac{1-c(\theta)}{c(\theta)} \frac{\theta}{1-\theta} \frac{1-G_{\text{Bayes},x}(\theta)}{G_{\text{Bayes},x}(\theta)}$. Now, define $c(\theta) = c_n(\theta)$ by the equation

$$\mathbb{E}G_{\text{Unbiased},x}(\theta) = \frac{1}{2} \quad \forall \theta \in (0,1).$$

Again, it is necessary to verify that $Q_{\text{Unbiased}}$ defined by (6.7) is, indeed, a procedure for making distributional inferences or, in other words, that $G_{\text{Unbiased},x}$ is a distribution function, for all $x = \{0,1,\ldots,n\}$. A slight modification of the proof of Lemma 6.1 provides that it is sufficient to show that $c(\theta)$ is nondecreasing.

**Conjecture 6.1** Let $c : [0,1] \mapsto [0,1]$ be implicitly defined by $\mathbb{E}G_{\text{Unbiased},x}(\theta) = \frac{1}{2}$, for all $\theta \in \Theta$. Then $c$ is continuous and monotonically increasing.

A partial proof of this lemma can be found in the appendix. Notice that by definition $Q_{\text{Unbiased}}$ is weakly unbiased, that it is equivariant under inversion, and that it does not assign positive credibility to impossible cases.

**The bias functions** So far, five different procedures have been proposed for making inference about the unknown success probability $p$. To determine which of those is most appropriate, a comparative analysis is needed. This comparative study of the procedures is started by looking at the bias functions. As both $Q_{\text{Fiducial}}$ and $Q_{\text{Unbiased}}$ are by construction weakly unbiased, it follows by definition that

$$B(\theta,Q_{\text{Fiducial}}) \equiv 0 \quad \text{and} \quad B(\theta,Q_{\text{Unbiased}}) \equiv 0.$$  

The requirement of equivariance under inversion of a procedure implies that a procedure has a bias function that is skew symmetric about $\frac{1}{2}$, i.e.,

$$B(\theta,Q) = \leftrightarrow B(1 \leftrightarrow \theta,Q) \quad \text{and} \quad B\left(\frac{1}{2},Q\right) = 0.$$
Both $Q_{\text{Bayes}}$ and $Q_{\text{Jeffreys}}$ do not assign positive credibility, for any outcome $x$. As a consequence, their bias functions have the property that the limits are $\equiv B(0, Q) = B(1, Q) = \frac{1}{2}$. The bias function of $Q_{\text{Bayes}}$ can be computed explicitly by using the fact that

$$E G_{\text{Bayes},X}(\theta) = \frac{1}{2} + \left( \theta \leftrightarrow \frac{1}{2} \right) P(X_\theta = \tilde{X}_\theta),$$

where $X_\theta$ and $\tilde{X}_\theta$ are i.i.d. $\text{Bin}(n, \theta)$. This provides that

$$B(\theta, Q_{\text{Bayes}}) = \left( \theta \leftrightarrow \frac{1}{2} \right) P(X_\theta = \tilde{X}_\theta).$$

For large $n$, this bias function may be tedious to compute, but it can be approximated very accurately by using the relation

$$P(X_\theta = \tilde{X}_\theta) = P(S_{\theta,n} = 0),$$

where $S_{\theta,n}$ is the sum of i.i.d. random variables $Z_{\theta,i}$ defined on $\{\leftrightarrow 1, 0, 1\}$. Now, for each fixed $\theta \in (0, 1)$, the Local Central Limit Theorem can be applied.

**Theorem 6.1** Let $S_{\theta,n} = \sum_{i=1}^{n} Z_{\theta,i}$, with $P(Z_{\theta,i} = \leftrightarrow 1) = P(Z_{\theta,i} = 1) = \theta(1 \leftrightarrow \theta)$ and $P(Z_{\theta,i} = 0) = \theta^2 + (1 \leftrightarrow \theta)^2$. Then

$$\sup_n \left| \sqrt{2\theta(1 \leftrightarrow \theta)n}P(S_{\theta,n} = 0) \Rightarrow \frac{1}{\sqrt{2\pi}} \right| = O\left( \frac{1}{\sqrt{n}} \right).$$

The proof can be found in, e.g. Petrov [71]. Using this LCLT, the following approximation is obtained

$$B(\theta, Q_{\text{Bayes}}) \approx \frac{\theta \leftrightarrow \frac{1}{2}}{2\sqrt{\pi \theta(1 \leftrightarrow \theta)n}}.$$ 

From this asymptotic approximation it can be seen immediately that the bias $B(\theta, Q_{\text{Bayes}})$ tends to 0 as $n$ tends to infinity, for each fixed $\theta \in (0, 1)$. From numerical calculations it follows that this asymptotic approximation is quite accurate if $\theta$ is not too close to either 0 or 1, even for small $n$, say $n \geq 3$. As $\text{Beta}(x + \frac{1}{2}, n \leftrightarrow x + \frac{1}{2})$ approximates $\frac{1}{2}\text{Beta}(x+1, n \leftrightarrow x) + \frac{1}{2}\text{Beta}(x, n \leftrightarrow x+1)$ quite accurately, provided that $x \neq 0$ or $x \neq n$, the bias function of $Q_{\text{Jeffreys}}$ is close to that of $Q_{\text{Fiducial}}$, which is identically equal to 0. The differences for $x = 0$ and $x = n$ become visible for $\theta$ close to either 0 or 1, where the bias function
6.1. Bayes's problem

Figure 6.1: The bias functions of the ‘biased’ procedures for Bayes’s problem, with \( n = 4 \), are displayed by a solid line. Their asymptotic approximations are displayed by a dotted line.

tends to respectively \( \leftrightarrow \frac{1}{2} \) and \( \frac{1}{2} \). For large \( n \) this influence rapidly vanishes for \( \theta \) not too close to either 0 or 1. In Figure 6.1 this is clearly visible, together with the fact that the bias of \( Q_{\text{Jeffreys}} \) is considerably smaller than that of \( Q_{\text{Bayes}} \). The bias function of \( Q_{\text{semi-Bayes}} \) can be computed explicitly at the endpoints, which gives

\[
B(0, Q_{\text{semi-Bayes}}) = \leftrightarrow B(1, Q_{\text{semi-Bayes}}) = \frac{n}{2(n + 1)}.
\]

Notice that this bias is in the opposite direction of the biases of both \( Q_{\text{Bayes}} \) and \( Q_{\text{Jeffreys}} \). More interestingly, however, is the fact that in contrast to the bias functions of the Bayes procedures, the bias function of \( Q_{\text{semi-Bayes}} \) does not tend to 0 as \( n \) increases. This can be shown by using the following result.
Lemma 6.2 For every fixed \( \theta \in (0,1) \), the bias function of \( Q_{\text{semi-Bayes}} \) tends to
\[
B(\theta, Q_{\text{semi-Bayes}}) \to \frac{1}{2} \iff 4\theta \iff \frac{\theta(1 \iff \theta)}{(\theta \iff \frac{1}{4})^2} \log \left( \frac{1 \iff \theta}{\theta} \right), \tag{6.8}
\]
as \( n \) tends to infinity.

**Proof.** Fix an arbitrary \( \theta \in (0,1) \), and notice that the bias function can be written as
\[
B(\theta, Q_{\text{semi-Bayes}}) = \mathbb{E} f_{\theta}(G_{\text{Bayes}, X_s}(\theta)),
\]
where
\[
f_{\theta}(z) = \frac{(1 \iff \theta)z}{(1 \iff \theta)z + \theta(1 \iff z)} \iff \frac{1}{2}.
\]
The function \( f_{\theta} \) is bounded and continuous, and hence it suffices to show that \( G_{\text{Bayes}, X_s}(\theta) \) converges weakly to \( U(0,1) \) as \( n \to \infty \), because if \( LU = U(0,1) \) then \( \mathbb{E} f_{\theta}(U) \) is equal to the r.h.s. of (6.8). Now, notice that
\[
\mathcal{L} G_{\text{Bayes}, X_s}(\theta) = \mathcal{L} \left( \sum_{i}^{X_s+1} \theta^i (1 \iff \theta)^{n+i+1} \right).
\]
Define \( \tilde{X}_\theta \) such that \( \mathcal{L} \tilde{X}_\theta = \text{Bin}(n + 1, \theta) \), and \( I \) independent of \( \tilde{X}_\theta \) such that \( P(I = 1) = 1 \iff P(I = 0) = \theta \). Then
\[
\mathcal{L} G_{\text{Bayes}, X_s}(\theta) = \mathcal{L} \left( I \sum_{i}^{\tilde{X}_s} \theta^i (1 \iff \theta)^{n+i+1} + (1 \iff I) \sum_{i}^{\tilde{X}_s+1} \theta^i (1 \iff \theta)^{n+i+1} \right),
\]
which tends to \( IU + (1 \iff I)U \), where \( LU = U(0,1) \), as \( n \) tends to infinity. \( \blacksquare \)

In Figure 6.1 the asymptotic bias of \( Q_{\text{semi-Bayes}} \) is displayed. As the bias does not vanish for large \( n \), \( Q_{\text{semi-Bayes}} \) cannot be considered to be a very promising procedure. Provided that its risk is not superior in comparison to those of the other procedures, it can be eliminated from the list of potential candidates.

**The risk functions** To compute the risk of the procedures, the quadratic loss function \( \mathcal{L}_\tau \), given by (5.2) and with \( \tau \) Lebesgue measure on \((0,1)\), is used. It is well known that \( Q_{\text{Bayes}} \) minimizes the integrated risk if \( \nu \) is taken to be Lebesgue measure on \((0,1)\). It was established in Kroese [54] that \( Q_{\text{Fiducial}} \) minimizes this integrated risk among all weakly unbiased procedures. In Figure 6.2 the risk
functions are displayed for $n = 4$. Computations show that for other $n$ the picture stays qualitatively the same, in the sense that only the scale changes.

Looking at Figure 6.2, the first thing that catches the eye is that the procedures $Q_{\text{Unbiased}}$ and $Q_{\text{semi-Bayes}}$ have considerably larger risk than the other procedures. Also taking into account the bias function of $Q_{\text{semi-Bayes}}$, it is safe to say that both procedures can be excluded from the list of reasonable procedures. The risk of $Q_{\text{Jeffreys}}$ and $Q_{\text{Fiducial}}$ are almost identical; that of $Q_{\text{Fiducial}}$ is slightly larger, which might be regarded as the price that has to be paid for requiring weak unbiasedness. The three candidates that are left, all have their own merits. A practical advantage of both $Q_{\text{Fiducial}}$ and $Q_{\text{Bayes}}$ is that they are based on Beta-distributions with integer parameters, which makes them easier to analyze. As $Q_{\text{Fiducial}}$ and $Q_{\text{Jeffreys}}$ are almost identical, not much is lost by excluding $Q_{\text{Jeffreys}}$ from further consideration. Hence, $Q_{\text{Fiducial}}$ and $Q_{\text{Bayes}}$ remain; the first having better bias properties, the latter having better risk properties.
6.2 Quantiles and quantile functions

The previous section was concluded by observing that both \( Q_{\text{Bayes}} \) and \( Q_{\text{Fiducial}} \) are reasonable procedures for Bayes’s problem. In this section these procedures will be used to make distributional about the \( p \)-th quantile \( \xi_p = F^{-1}(p) \), where \( p \in [0,1] \), and \( F \) denotes the unknown true distribution function of \( X \). Denote by \( x_{[1]}, \ldots, x_{[n]} \) the ordered outcomes of an independent random sample from a distribution \( F \) which has the interval \([x_{[0]}, x_{[n+1]}] = [a, b]\) as its support, where for the moment \( a \) and \( b \) are assumed to be known. The logical equivalence of the statements

\[
\xi_p \geq z \Leftrightarrow F(z) \leq p
\]

provides that, instead of considering the hypotheses of the form \( H_z : \xi_p \geq z \), one can concentrate on hypotheses of the form \( H_z : F(z) \leq p \). Notice that the true value of \( F(z) \) is unknown, because \( F \) is unknown. All information concerning \( F(z) \) seems to be contained in the random variable

\[
S_z = \#\{1 \leq i \leq n : X_{[i]} \leq z\},
\]

where \( \mathcal{L}S_z = \text{Bin}(n, F(z)) \). Using this perspective, the problem can be reformulated as follows. Given the outcome \( s_z \) of \( S_z \), a distributional inference will be made about the unknown success probability \( F(z) \). In this formulation, the problem of making inference about a quantile \( \xi_p \) is equivalent to Bayes’s problem.

For Bayes’s problem, both \( Q_{\text{Bayes}} \) and \( Q_{\text{Fiducial}} \) were considered to be reasonable procedures. Applying \( Q_{\text{Bayes}} \) provides that credibility

\[
\int_{[0,p]} \frac{u^{s_z-1}(1 \leftrightarrow u)^{n-s_z}}{\beta(s_z+1, n \leftrightarrow s_z+1)} \, du = P(B_{n+1, p} \geq s_z + 1),
\]

where \( \mathcal{L}B_{n+1, p} = \text{Bin}(n+1, p) \), is assigned to \( H_z : F(z) \leq p \) or, equivalently, to \( H_z : \xi_p \geq z \). Regarded as a function of \( z \), the credibility thus assigned to \( \xi_p \geq z \) jumps (downward) if \( z = x_{[i]} \); for \( z = x_{[i]} \) the outcome \( s_z \) of \( S_z \) is equal to \( i \), whereas for \( z = x_{[i]} \) the outcome \( s_z \) of \( S_z \) is equal to \( i \). Hence, all credibility mass of the distributional inference about \( \xi_p \) is concentrated in the order statistic \( x_{[0]}, \ldots, x_{[n+1]} \), i.e.,

\[
Q_{\text{Bayes}}(x, x_{[i]}) = P(B_{n+1, p} \geq i) \Leftrightarrow P(B_{n+1, p} \geq i + 1) = P(B_{n+1, p} = i), \quad (6.9)
\]

for \( i = 0, \ldots, n + 1 \). Similarly, applying \( Q_{\text{Fiducial}} \) provides that credibility

\[
\frac{1}{2} \int_{[0,p]} \frac{u^{s_z-1}(1 \leftrightarrow u)^{n-s_z}}{\beta(s_z, n \leftrightarrow s_z + 1)} \, du + \frac{1}{2} \int_{[0,p]} \frac{u^{s_z-1}(1 \leftrightarrow u)^{n-s_z-1}}{\beta(s_z+1, n \leftrightarrow s_z)} \, du =
\]
6.2. Quantiles and quantile functions

$$\frac{1}{2} P(B_{n,p} \geq s_z) + \frac{1}{2} P(B_{n,p} \geq s_z + 1),$$

where $L B_{n,p} = \text{Bin}(n, p)$, is assigned to $H_z : \xi_p \geq z$. Again, all credibility mass is concentrated in the order statistic $x_{[0]}, \ldots, x_{[n+1]}$, i.e.,

$$Q_{\text{Fiducial}}(x, x_{[i]}) = \frac{1}{2} P(B_{n,p} = i) + \frac{1}{2} P(B_{n,p} = i + 1),$$

(6.10)

for $i = 0, \ldots, n$. Some adaptations are needed in case that one or both $a$ and $b$ are unknown. A number of different cases can be distinguished. If both $a$ and $b$ are finite but unknown, then they can be estimated by

$$a = x_{[0]} = x_{[1]} \leftrightarrow (x_{[2]} \leftrightarrow x_{[1]}) \quad \text{and} \quad b = x_{[n+1]} = x_{[n]} + (x_{[n]} \leftrightarrow x_{[n-1]}).$$

If the support is not finite, then, for each of the possibilities $(\sim \infty, b], [a, \infty)$, or $(\sim \infty, \infty)$, a continuous distribution function $\Psi : [a, b] \mapsto [0, 1]$ can be used to transform the problem to $[0, 1]$, i.e., take $0 = \Psi(x_{[0]}), \Psi(x_{[1]}), \ldots, \Psi(x_{[n+1]}) = 1$. For specific recommendations for $\Psi$, see De Bruin et al. [17].

The distributional inferences (6.9) and (6.10) about $\xi_p$, can be used to obtain nonparametric estimates of the quantile function $F^{-1}$. Taking the expectations of either $Q_{\text{Bayes}}$ or $Q_{\text{Fiducial}}$, and regarding this expectation as a function of $p$, provides the Bernstein polynomial

$$H_n(p) = \sum_{i=0}^{n+1} x_{[i]} \binom{n+1}{i} p^i (1 \leftrightarrow p)^{n+1-i},$$

(6.11)

in case that $Q_{\text{Bayes}}$ is used, or the Kantorovitz polynomial

$$K_n(p) = \sum_{i=0}^{n} \frac{1}{2} (x_{[i]} + x_{[i+1]}) \binom{n}{i} p^i (1 \leftrightarrow p)^{n-i},$$

(6.12)

in case that $Q_{\text{Fiducial}}$ is used. It is not difficult to see that both $H_n$ and $K_n$ are strictly increasing functions, and that $H_n(0) = K_n(0) = x_{[0]} = a$ and $H_n(1) = K_n(1) = x_{[n+1]} = b$. Notice that if $p = \frac{1}{2}$, then $H_n(\frac{1}{2})$ is a modification of the Islamic-Mean estimator of the median, see Dehling et al. [28]. For further discussion about $H_n$ or $K_n$, see e.g. Munoz et al. [64] or De Bruin et al. [17].

The estimates $H_n(p)$ and $K_n(p)$ of $\xi_p$, for a fixed value of $p \in [0, 1]$, are L-statistics. They differ slightly from the $p$th sample quantile $F_{n-1}^p(p)$, where $F_n$ denotes the empirical distribution function, that is commonly used to estimate
the population quantile. The asymptotic distribution of $F_{n}^{-1}(p)$ is well known to be normal, with the true value $\xi_{p}$ as expectation, and $n^{-1}p(1 \Leftrightarrow p)(f(\xi_{p}))^{-2}$ as variance. Using methods of Dehling et al. [28] it can be shown that $H_{n}(p)$ and $K_{n}(p)$ have the same asymptotic distribution. To analyze the asymptotic properties of $H_{n}$ and $K_{n}$, for all $p$ simultaneously, introduce the following notation. Suppose that $U_{1}, \ldots, U_{n}$ is an independent random sample from $U(0, 1)$, and let the corresponding order statistic be denoted by $U_{[0]}, \ldots, U_{[n+1]}$, where by definition $U_{[0]} = 0$ and $U_{[n+1]} = 1$. Define the interpolated uniform empirical quantile function $V_{n}(t)$, with $t \in (0, 1)$, by

$$
V_{n}(t) = \begin{cases} 
U_{[i]} & \text{if } t = \frac{i}{n+1}, \\
\alpha U_{[i-1]} + (1 \Leftrightarrow \alpha)U_{[i]} & \text{if } t = \alpha \frac{i-1}{n+1} + (1 \Leftrightarrow \alpha) \frac{i}{n+1},
\end{cases}
$$

for $i = 0, 1, \ldots, n + 1$ and $\alpha \in (0, 1)$. Using this interpolated uniform empirical quantile function, the uniform empirical quantile process can be defined by

$$
V_{n}(t) = \sqrt{n}(V_{n}(t) \Leftrightarrow t) \quad t \in (0, 1).
$$

Notice that $V_{n}(t) = \Leftrightarrow U_{n}(t)$, where $U_{n}(t)$ is the interpolated uniform empirical process. Hence, by the Donsker-Prokhorov invariance principle and Donsker's Theorem, it follows that $V_{n}(t)$ converges weakly to a standard Brownian Bridge. Next, introduce the empirical quantile function

$$
Q_{n}(t) = F^{-1}(V_{n}(t)),
$$

where $V_{n}(t)$ is the uniform empirical quantile function of the transformed sample $F(X_{1}), \ldots, F(X_{n})$. The corresponding standardized empirical quantile process $Q_{n}(t)$, with $t \in (0, 1)$, is defined as

$$
Q_{n}(t) = f(F^{-1}(t))\sqrt{n}(Q_{n}(t) \Leftrightarrow F^{-1}(t)).
$$

Using standard techniques from empirical process theory, it can be shown that $Q_{n}$ converges weakly to a standard Brownian Bridge. Using these notations

$$
H_{n}(p) = \sqrt{n}(H_{n}(p) \Leftrightarrow F^{-1}(p)) \quad \text{and} \quad K_{n}(p) = \sqrt{n}(K_{n}(p) \Leftrightarrow F^{-1}(p))
$$

can be analyzed by making the following decomposition

$$
H_{n}(p) = \int_{[0, 1]} \frac{1}{f(F^{-1}(p))} Q_{n}(t) k_{1,n}(p, dt) + \sqrt{n} \, R_{1,n}(p),
$$

$$
K_{n}(p) = \int_{[0, 1]} \frac{1}{f(F^{-1}(p))} Q_{n}(t) k_{2,n}(p, dt) + \sqrt{n} \, R_{2,n}(p),
$$

where $k_{1,n}(p, dt)$ and $k_{2,n}(p, dt)$ are kernels.
where the kernels are given by respectively

\[
\begin{align*}
k_{1,n}(p,t) &= \frac{1}{2} \left( (n+1)^t (1 \Leftrightarrow p)^{(n+1)(1-t)} \right) \\
&\quad \cdot \left( \frac{(n+1)(1 \Leftrightarrow t) + 1}{n+1 \cdots \frac{1}{n+1}} \right) (t),
\end{align*}
\]

\[
\begin{align*}
k_{2,n}(p,t) &= \frac{1}{2} \left( (n+1)^{t-1} (1 \Leftrightarrow p)^{(n+1)(1-t)-1} \right) \\
&\quad \cdot \left( \frac{(n+1)(1 \Leftrightarrow t) + 1}{n+1 \cdots \frac{1}{n+1}} \right) (t),
\end{align*}
\]

and the deterministic remainder terms by respectively

\[
\begin{align*}
R_{1,n}(p) &= \int_{[0,1]} F^{-1}(t) k_{1,n}(p, dt) \Leftrightarrow F^{-1}(p), \\
R_{2,n}(p) &= \int_{[0,1]} F^{-1}(t) k_{2,n}(p, dt) \Leftrightarrow F^{-1}(p).
\end{align*}
\]

Now, notice that both the kernels \(k_{1,n}\) and \(k_{2,n}\) satisfy the following regularity conditions: (i) for each \(p \in (0,1)\), \(k_{i,n}\) is supported on a subset of \([0,1]\), (ii) there exists a sequence \(\delta_n \downarrow 0\) such that

\[
\sup_{p \in (0,1)} \left| \int_{(p-\delta_n, p+\delta_n)} k_{i,n}(p, dt) \Leftrightarrow 1 \right| \to 0,
\]

and (iii) for an arbitrary interval \([a,b] \subset (0,1)\) and any function \(g \in C^m[a,b]\), with \(m \geq 3\), and well defined and differentiable on \((0,1)\)

\[
\sup_{p \in [a,b]} \left| g(p) \Leftrightarrow \int_{(0,1)} g(t) k_{i,n}(p, dt) \right| = O(n^{-1}),
\]

for \(i = 1, 2\). Hence, the following theorem due to Cheng [21], which is an extension of results of Csörgő et al. [24], can be applied.

**Theorem 6.2** Suppose that \(F\) satisfies the following conditions: (i) \(F^{-1}\) is differentiable on \((0,1)\), and \((F^{-1}(p))' = 1/f(F^{-1}(p)) < \infty\), (ii) there is a \(\gamma > 0\) such that

\[
\sup_{p \in (0,1)} |p(1 \Leftrightarrow p)(\log(1/f(F^{-1}(p))))|' \leq \gamma,
\]

(iii) either \(1/f(F^{-1}(p)) < \infty\) or \(1/f(F^{-1}(p))\) is nonincreasing on some interval \((0,p^*)\), and either \(1/f(F^{-1}(1)) < \infty\) or \(1/f(F^{-1}(p))\) is nondecreasing on some
interval \((p^*, 1)\), (iv) for an arbitrary fixed interval \([a, b] \subset (0, 1)\), \(F^{-1} \in C^m[a, b]\), with \(m \geq 3\). Let \(I\) be an arbitrary fixed compact subset of \((0, 1)\). Then there exists a sequence of Brownian Bridges \(\{B_n(p) : 0 \leq p \leq 1\}_{n=1}^{\infty}\) such that as \(n \to \infty\),

\[
\sup_{p \in I} \left| H_n(p) - \frac{1}{f(F^{-1}(p))} B_n(p) \right| = o_p(1),
\]

\[
\sup_{p \in I} \left| K_n(p) - \frac{1}{f(F^{-1}(p))} B_n(p) \right| = o_p(1).
\]

This theorem provides that, under some smoothness conditions concerning the unknown distribution function \(F\), both \(H_n\) and \(K_n\) have the same asymptotic behavior as the ordinary quantile estimator. Notice that the advantage of using either \(H_n\) or \(K_n\) over ordinary kernel estimators of the quantile distribution, is that they have an intrinsic choice for the smoothing parameter.

6.3 Density estimation

In the previous section, two estimators \(H_n\) and \(K_n\) were obtained for estimating the quantile function \(F^{-1}\). The attractive feature of both estimators is that they are smoothed versions of the ordinary quantile function \(F_n^{-1}\) having the same asymptotic behavior, with the choice of the smoothing parameter given by an intrinsic argument. Taking into account the smoothness of \(H_n\), and the fact that it is strictly increasing, makes it possible to obtain the nonparametric density estimator

\[ g_n = (H_n^{-1})', \]

for the unknown density \(f\), with support \([a, b]\). Numerical and theoretical considerations, given in De Bruin et al. [17], and Cheng [21], provide that one should not worry too much about the potential wiggling character of the density estimators.

The price that has to be paid for having an intrinsic smoothing argument is that the method requires that \(X_1, \ldots, X_n\) is an independent random sample from a univariate distribution with a positive density on the given support \([a, b]\). The competing methods, e.g. parametric methods, kernel methods, spline approximations, nearest neighbor methods, or orthogonal expansion or wavelet methods, are more flexible and open to generalization. A numerical comparative analysis was carried out to evaluate the performance of \(g_n\) in comparison to that of competing methods. An expert in the area suggested to use a kernel estimate and to adapt the density obtained to the support \([0, 1]\) given (the comparison will be made for mixtures of Beta distributions) by reflecting the tails about 0.
Figure 6.3: On the l.h.s 100 comparisons (sample size \( n = 100 \)) of \( g_n \) with a kernel density estimator with optimal bandwidth are displayed. On the r.h.s. the same comparison is made for a kernel estimator with cross-validated bandwidth.

and about 1 and superposing. The competing kernel density estimator used the biweight \( 15(I \leq u^2)^2/16, \leq l < u < 1 \), as kernel and the bandwidth \( h \) (or rather \( 2h \)) which was taken to be optimal for the sample size \( n = 100 \) and the true distribution \( \frac{1}{2}\text{Beta}(1, 5) + \frac{1}{2}\text{Beta}(7, 2) \) to start with. Measuring the difference between the estimator of the density \( g \) and the true density \( f \) by

\[
\int_{a,b} |g(z) - f(z)|dz,
\]

and always using 100 different simulations, it was obtained that the minimum ‘average distance’ occurred at about the value \( h = 0.18 \), being about 0.18 (see also the l.h.s. of Figure 6.3 and note that the average y-coordinate is 0.18). Now the interpretation of Figure 6.3 is obvious. If the bandwidth \( h = 0.18 \) is chosen such that it is optimal for the true distribution actually approximated, then the kernel method has a better performance.

Remark 6.1 The computations for the picture on the l.h.s. in Figure 6.3 were performed by R. de Bruin.

If a different bandwidth is chosen (in practice one does not know which distribution is estimated), then the advantage of the kernel method may get lost. If,
e.g., $h = 0.06$ is chosen then the $x$-coordinates (our method) on the l.h.s. of Figure 6.3 are preserved but the $y$-coordinates (‘their’ method) become larger such that on the l.h.s. of Figure 6.3 the average $y$-coordinate increases from 0.18 to 0.26. The $x$-coordinate being about 0.21, it can be inferred that ‘our’ method is in the lead if the bandwidth $h = 0.06$ is chosen for the kernel estimate. Making a number of similar comparisons, the global conclusion was that the kernel method with optimal bandwidth is in the lead, but that ‘our’ method takes over if the bandwidth is something like a factor 2 different from the optimal one. A more reasonable comparison is obtained if the bandwidth of the kernel density estimator is selected on the basis of the data. Using likelihood cross-validation and carrying out the same comparison provides the results displayed on the r.h.s. of Figure 6.3. The average bandwidth that was selected was 0.19 with standard deviation 0.06. The average distance of the kernel density estimator was 0.19. It can be concluded that the kernel density estimator with cross-validated bandwidth has a slightly better performance, at the cost of a much larger computational effort.