Maximum Spacing Estimates Based on Different Metrics

Bo Ranneby
Department of Forest Resource Management and Geomatics
Swedish University of Agricultural Sciences
S-901 83 Umeå, Sweden

Magnus Ekström
Department of Mathematical Statistics
Umeå University
S-901 87 Umeå, Sweden

Abstract

The maximum spacing (MSP) method, introduced by Cheng and Amin (1983) and independently by Ranneby (1984), is a general method for estimating parameters in univariate continuous distributions and is known to give consistent and asymptotically efficient estimates under general conditions. This method, which is closely related to the maximum likelihood (ML) method, can be derived from an approximation based on simple spacings of the Kullback-Leibler information. In the present paper, the ideas behind the MSP method are extended and a class of estimation methods is derived from approximations of certain information measures, i.e. the $\phi$-divergences introduced by Csiszár (1963). We call these methods generalized maximum spacing (GMSP) methods, and it will be shown under general conditions that they give consistent estimates. GMSP methods have the advantage that they work also in situations where the ML method breaks down, e.g. due to an unbounded likelihood function. Other properties, such as asymptotic normality and the behaviour of the estimates when the assigned model is only approximately true, will be discussed.¹

Key words and phrases: Estimation, Spacings, Maximum spacing method, Consistency, $\phi$-divergence,

1991 AMS subject classification: 62F12, 62F10

¹Research was supported by MISTRA, the Foundation for Strategic Environmental Research.
1 Introduction

For independent, identically distributed (i.i.d.) univariate observations an estimation method, called the maximum spacing (MSP) method, is defined by Cheng and Amin (1983) and independently by Ranneby (1984). In Ranneby (1984) the method is obtained by an approximation of the Kullback-Leibler information such that each contribution is bounded from above.

The idea behind the method is as follows: Given a measure (a metric), e.g. the Kullback-Leibler information, of the distance between the distributions in our statistical model and the true underlying distribution, a good inference method ought to make this distance as small as possible. Since the true distribution is not completely known we have to use our prior knowledge and the observations to approximate the distance. Then we obtain our method for statistical inference by making this distance, in the approximation, as small as possible. Approximating different "metrics" we get different methods for statistical analysis. For example, the maximum likelihood (ML) method can be obtained from an approximation of the Kullback-Leibler information. The ML method often works properly for discrete distributions in particular, but not for instance for all mixtures of continuous distributions.

Two natural questions arise. Is it possible to obtain better methods than the ML method by different approximations of the Kullback-Leibler information? Is it possible to obtain better methods by approximations of other information-measures?

The first question led to the MSP method, and has been further elaborated by Ekström (1994, 1996) and Shao and Hahn (1996), among others. In the present paper we will deal with the second question, deriving different estimation methods from a class of information measures.

Let $F_\theta(x)$, $\theta \in \Theta$, denote a family of distribution functions, absolutely continuous with respect to the Lebesgue measure $\mu$, where $\Theta$ is a subset of a finite dimensional Euclidian space. Denote the corresponding density functions by $f_\theta(x)$. Assume that $\xi_1, \ldots, \xi_n$ is a sample of i.i.d. random variables with distribution function $F_{\theta^0}(x)$, $\theta^0 \in \Theta$, and let

$$-\infty = \xi(0) \leq \xi(1) \leq \ldots \leq \xi(n) \leq \xi(n+1) = +\infty$$

be the order statistics.

The MSP-method as introduced by Ranneby (1984) is based on the minimizing of an approximation of the Kullback-Leibler information

$$I(f_{\theta^0}, f_\theta) = \int_{-\infty}^{\infty} \log \left( \frac{f_{\theta^0}(x)}{f_\theta(x)} \right) f_{\theta^0}(x) dx.$$
The approximation used is

\[
\frac{1}{n+1} \sum_{j=1}^{n+1} \log \left( \frac{F_{\theta_0}(\xi_{(j)}) - F_{\theta_0}(\xi_{(j-1)})}{F_{\theta}(\xi_{(j)}) - F_{\theta}(\xi_{(j-1)})} \right).
\]

The minimizing of this sum with respect to \(\theta\) is equivalent to the maximizing of

\[
S_n(\theta) = \frac{1}{n+1} \sum_{j=1}^{n+1} \log \left( (n+1) \left( F_{\theta}(\xi_{(j)}) - F_{\theta}(\xi_{(j-1)}) \right) \right),
\]

which defines the MSP-estimate of \(\theta^0\), i.e. the MSP-estimate is defined as any \(\theta \in \Theta\) that maximizes \(S_n(\theta)\). If we define

\[\eta_i(n) = (n+1) \cdot \text{"the distance from } \xi_i \text{ to the nearest observation to the right of } \xi_i, \text{" (this distance is defined as } +\infty \text{ if } \xi_i = \max_{1 \leq j \leq n} \xi_j, \text{)}\]

then \(S_n(\theta)\) may be rewritten as

\[
S_n(\theta) = \frac{1}{n+1} \sum_{j=0}^{n} \log z_j(n, \theta),
\]

where \(z_0(n, \theta) = (n+1)F_{\theta}(\min_{1 \leq i \leq n} \xi_i)\) and

\[z_i(n, \theta) = (n+1) \left( F_{\theta} \left( \xi_i + \frac{\eta_i(n)}{n+1} \right) - F_{\theta}(\xi_i) \right), \quad i = 1, ..., n.\]

Let \(P_n\) denote the joint probability distribution of \((\xi_1, \eta_1(n))\). By conditioning on \(\xi_1 = s\) it is easily seen that

\[
P_n(x, y) = P_n(\xi_1 \leq x, \eta_1(n) \leq y)
\]

\[
= \begin{cases} 
\int_{-\infty}^{x} \left[ 1 - \left( 1 - \left( F_{\theta_0} \left( s + \frac{y}{n+1} \right) - F_{\theta_0}(s) \right) \right)^{n-1} \right] f_{\theta_0}(s) \, ds & \text{if } 0 < y < \infty \\
\int_{-\infty}^{x} f_{\theta_0}(s) \, ds & \text{if } y = \infty.
\end{cases}
\]

In Ekström (1994) it is shown (under the assumption that \(f_{\theta_0}(x)\) is right-continuous \(\mu\text{-a.e.}\)) that \(P_n\) converges weakly to \(P\), where

\[
P(x, y) = \int_{-\infty}^{x} \left( 1 - e^{-y f_{\theta_0}(u)} \right) f_{\theta_0}(u) \, du, \quad y > 0.
\]

The density function \(p(x, y)\) of \(P(x, y)\) is given by

\[
p(x, y) = f_{\theta_0}^2(x)e^{-y f_{\theta_0}(x)}, \quad y > 0.
\]
From this it follows that \((\xi_i, \eta_i(n) f_{\theta}(\xi_i))\), \(i = 1, \ldots, n\), converges in distribution to \((X_i, W_i)\), where \(X_i\), with distribution function \(F_{\theta}\), is independent of \(W_i\), with an exponential distribution with mean 1.

An application of the mean value theorem, for \(i = 1, 2, \ldots, n\), gives

\[
z_i(n, \theta) \approx \eta_i(n) f_{\theta}(\xi_i) = \eta_i(n) f_{\theta}(\xi_i) \frac{f_{\theta}(\xi_i)}{f_{\theta}(\xi_i)},
\]

where the right hand side above has approximately the same distribution as \(W_i f_{\theta}(\xi_i)/f_{\theta}(\xi_i)\), where \(W_i\) is independent of \(\xi_i\). Thus, it is intuitively clear that \(S_n(\theta)\) converges in probability to \(E[\log W_1] - I(f_{\theta}, f_{\theta})\), which is maximized if and only if \(f_{\theta}(x) = f_{\theta}(x)\) \(\mu\)-a.e. Thus, by heuristic arguments, the distributions that are most likely to have generated the sample should be found by the maximizing of \(S_n(\theta)\).

The Kullback-Leibler distance is a special case of Csiszar's \(\phi\)-divergence (introduced by Csiszar (1963)) with \(\phi(x) = -\log x\). In the present paper the MSP-method is extended to approximations of Csiszar's \(\phi\)-divergences, defined by

\[
\int_{-\infty}^{\infty} \phi \left( \frac{f_{\theta}(x)}{f_{\theta}(x)} \right) f_{\theta}(x) dx,
\]

where \(\phi(x)\) is a convex function. Note that the general Hellinger distance, the Kullback-Leibler information and Jeffreys divergence all are \(\phi\)-divergences or functions of a \(\phi\)-divergence. The minimizing of an approximation to (1) is equivalent to the maximizing of an approximation to

\[
\int_{-\infty}^{\infty} \Psi \left( \frac{f_{\theta}(x)}{f_{\theta}(x)} \right) f_{\theta}(x) dx,
\]

where \(\Psi(x) = -\phi(x)\).

As \(f_{\theta}(\xi_i)\) is unobservable, but \(z_i(n, \theta)\) is observable and in addition has approximately the same distribution as \(W_i f_{\theta}(\xi_i)/f_{\theta}(\xi_i)\), where \(W_i\) is independent of \(\xi_i\) and \(E[W_i] = 1\), a possible approximation of (2) would be to substitute \(f_{\theta}/f_{\theta}\) by the spacing values \(z_i(n, \theta)\) and the expected value by the mean. Thus

\[
S_{\Psi, n}(\theta) = \frac{1}{n + 1} \sum_{i=0}^{n} \Psi(z_i(n, \theta))
\]

may be used as a possible approximation of (2). But it remains to check what \(S_{\Psi, n}(\theta)\) will converge to, and if this limit is maximized when \(\theta = \theta^0\).

Again since \(z_i(n, \theta)\) has approximately the same distribution as \(W_i f_{\theta}(\xi_i)/f_{\theta}(\xi_i)\), \(W_i\) independent of \(\xi_i\), it follows heuristically that \(S_{\Psi, n}(\theta)\) will converge in probability to

\[
E \left[ \Psi \left( W_1 \frac{f_{\theta}(\xi_1)}{f_{\theta}(\xi_1)} \right) \right] = \int_0^{\infty} \left\{ \int_{-\infty}^{w} \Psi \left( \frac{w f_{\theta}(x)}{f_{\theta}(x)} \right) f_{\theta}(x) dx \right\} e^{-w} dw \\
\leq \int_0^{\infty} \Psi(w) e^{-w} dw = E[\Psi(W_1)].
\]
If $\Psi$ is strictly concave, the inequality holds if and only if $f_\theta(x) = f_{\theta_0}(x)$ $\mu$-a.e..

Thus it seems reasonable to base an estimation method on the maximization of

$$S_{\Psi,n}(\theta) = \frac{1}{n+1} \sum_{i=0}^{n} \Psi(x_i(n, \theta)).$$

The obtained estimate of $\theta_0$, i.e. any parameter value in $\Theta$ maximizing $S_{\Psi,n}(\theta)$, denoted $\hat{\theta}_{\Psi,n}$, is called a generalized maximum spacing (GMSP) estimate. If $\sup S_{\Psi,n}(\theta)$ is not attained for any $\theta \in \Theta$ we define $\hat{\theta}_{\Psi,n}$ as any point belonging to the set $\Theta$ satisfying

$$S_{\Psi,n}(\hat{\theta}_{\Psi,n}) \geq -c_n + \sup_{\theta \in \Theta} S_{\Psi,n}(\theta),$$

where $c_n > 0$ is a sequence of constants such that $c_n \to 0$ as $n \to \infty$.

2 Consistency of GMSP-estimators

Before stating the assumptions under which consistency of $\hat{\theta}_{\Psi,n}$ will be proved, some new notation will be introduced.

Define,

$$T_\Psi(\theta) = E[\Psi(W_1 f_\theta(\xi_1)/f_{\theta_0}(\xi_1))] = \int_{-\infty}^{\infty} \int_{0}^{\infty} \Psi(y f_\theta(x)) p(x, y) dy dx,$$

$$T_\Psi(M, \theta) = \int_{-\infty}^{\infty} \int_{0}^{\infty} \max \left( -M, \Psi(y f_\theta(x)) \right) p(x, y) dy dx.$$

and

$$z(n, \theta, x, y) = (n + 1) \left( F_\theta \left( x + \frac{y}{n+1} \right) - F_\theta (x) \right).$$

Consistency of $\hat{\theta}_{\Psi,n}$ will be proved under the following set of assumptions:

**Assumption A$_1$:** The true underlying density $f_{\theta_0}(\cdot)$ is right-continuous $\mu$-a.e..

**Assumption A$_2$:** The function $\Psi(t), t \in R^+$, satisfies the following conditions:

(i) $\Psi$ is strictly concave,

(ii) $\Psi(t)^+/t \to 0$, as $t \to \infty$, where $\Psi(t)^+ = \max(0, \Psi(t))$,

(iii) $\lim_{b \downarrow 1} \int_0^{\infty} \Psi(bt) e^{-t} dt = \int_0^{\infty} \Psi(t) e^{-t} dt > -\infty$. 


Assumption $A_3$: For each $\delta > 0$ there exists a constant $M_1 = M_1(\delta)$ such that
\[
\sup_{\theta \in \Theta, \epsilon} T_{\epsilon}(M_1, \theta) < T_{\epsilon}(\theta^0),
\] where $\Theta_\delta = \Theta \setminus \{ \theta : |\theta - \theta^0| < \delta \}$.

Assumption $A_4$: Let $(X, Y)$ have the distribution function $P(x, y)$. For each $\epsilon > 0$ and $\eta > 0$ there exists a finite number $r = r(\epsilon, \eta)$ of sets $K_j = K_j(\epsilon, \eta) \subset \mathbb{R}^2$, $j = 1, 2, \ldots, r$ and a partition of $\Theta$ into disjoint sets $O_j = O_j(\epsilon, \eta)$, $j = 1, 2, \ldots, r$ such that, for each $j = 1, 2, \ldots, r$,

(i) the boundary $\partial K_j$ of the set $K_j$ has Lebesgue-measure zero,

(ii) $P((X, Y) \in K_j) > 1 - \eta$,

(iii) $\lim_{n \to \infty} \sup_{(x, y) \in K_j} \sup_{\theta, \theta' \in O_j} |z(n, \theta, x, y) - z(n, \theta', x, y)| < \epsilon$.

Theorem 1 Let $\xi_1, \ldots, \xi_n$ be a sequence of i.i.d. random variables with density function $f_{\theta_0}(x)$. Then, if assumption $A_2$ is valid, $S_{\Psi, n}(\theta^0)$ converges in probability to $T_{\Psi}(\theta^0)$ as $n \to \infty$.

Theorem 2 Let $\xi_1, \ldots, \xi_n$ be a sequence of i.i.d. random variables with distribution function $F_{\theta_0} \in \{F_\theta : \theta \in \Theta\}$, a family of probability measures dominated by $\mu$. Then, under assumptions $A_1 - A_4$, $\hat{\theta}_{\Psi, n}$ converges in probability to the parameter value $\theta^0$ as $n \to \infty$.

3 Some GMSP-estimators and information-type measures

As we have seen in the introduction, different estimators of finite dimensional parameters in univariate distributions can be derived from certain information-type measures (distance measures) of difference of probability distributions, called $\phi$-divergences and introduced by Csiszár (1963). The following measures are either $\phi$-divergences or functions of a $\phi$-divergence:

- Kullback-Leibler information: $\int f_{\theta_0}(x) \log \frac{f_{\theta_0}(x)}{f_\theta(x)} dx$,
- Jeffreys divergence: $\int \left( f_{\theta_0}(x) \log \frac{f_{\theta_0}(x)}{f_\theta(x)} + f_\theta(x) \log \frac{f_{\theta_0}(x)}{f_\theta(x)} \right) dx$,
- Rényi's divergence: $\frac{1}{\alpha - 1} \log \left( \int (f_\theta(x))^\alpha (f_{\theta_0}(x))^{1-\alpha} dx \right)$, $\alpha > 0$, $\alpha \neq 1$,
- Hellinger distance: $\left( \int \left| (f_\theta(x))^{1/p} - (f_{\theta_0}(x))^{1/p} \right|^p dx \right)^{1/p}$, $p \geq 1$,
- Vajda's measure of information: $\int f_{\theta_0}(x) \left| 1 - \frac{f_{\theta_0}(x)}{f_{\theta}(x)} \right|^\beta dx$, $\beta \geq 1$. 

5
Consequently each distance measure above can be written as a function of an integral of the form \( \int f_{\theta}(x) \Psi(f_{\theta}(x)/f_{\theta}(x))dx \), where \( \Psi \) is some concave function defined in the interval \((0, +\infty)\). The GMSP-estimators derived from these distance measures are the estimators \( \hat{\theta}_{\Psi,n} \) with

- \( \Psi(x) = \log x \) (Kullback-Leibler information),
- \( \Psi(x) = (1 - x) \log x \) (Jeffreys divergence),
- \( \Psi(x) = x^\alpha \text{sgn}(1 - \alpha), \alpha > 0, \alpha \neq 1 \) (Rényi’s divergence),
- \( \Psi(x) = -|1 - x^{1/p}|^p, p \geq 1 \) (Hellinger distance),
- \( \Psi(x) = -|1 - x|^\beta, \beta \geq 1 \) (Vajda’s measure of information).

Note that all these \( \Psi \)-functions, except \( \Psi(x) = -|1 - x| \), satisfies assumption \( A_2 \).

Note that if, for some constants \( C_1 \) and \( C_2 \), \( \Psi_2(x) = \Psi_1(x) + C_1 x + C_2 \), then \( S_{\Psi_1,n}(\theta) = S_{\Psi_2,n}(\theta) - C_1 - C_2 \). Consequently \( \Psi_1 \) and \( \Psi_2 \) will give identical GMSP-estimates. This means that the GMSP-estimates based on the Hellinger distance, with \( p = 2 \), and on the Rényi’s divergence of order \( 1/2 \) will be identical. Further, observe that concavity of \( \Psi(x) \) implies that of \( x \Psi(1/x) \). Hence the function \( \Psi^*(x) = \Psi(x) + x \Psi(1/x) \) is again a concave function and it follows that \( \int f_{\theta}(x) \Psi^*(f_{\theta}(x)/f_{\theta}(x))dx \) is symmetric in \( f_{\theta}(x) \) and \( f_{\theta}(x) \), e.g. Jeffreys divergence is a symmetrized version of the Kullback-Leibler information.

4 Discussion

In this paper the original MSP-estimator, based on an approximation of the Kullback-Leibler distance, has been extended to a whole class of estimators, by the use of distance measures other than the Kullback-Leibler information. Naturally, these estimators have different properties. So we have a problem of choosing the “best” estimation procedure. Of course, what is considered to be the best estimator depends on several criteria, such as robustness, bias, variance and mean square error, as well as on the true underlying distribution and the assigned model.

Assume that the postulated model is true. By Theorem 2 we know that \( \hat{\theta}_{\Psi,n} \) tends to \( \theta^0 \) in probability as \( n \to \infty \). In Nordahl (1994, unpublished) it turns out under general conditions (in the single parameter case) that

\[
\sqrt{n} (\hat{\theta}_{\Psi,n} - \theta^0) \overset{D}{\to} N \left(0, \sigma_{\Psi}^2 \right)
\]

where \( \sigma_{\Psi}^2 \) equals

\[
\frac{\int_0^\infty (\Psi'(w)w)^2 e^{-w}dw + 2(\int_0^\infty \Psi'(w)we^{-w}dw)^2 - 2\int_0^\infty \Psi'(w)we^{-w}dw \int_0^\infty \Psi'(w)w^2e^{-w}dw}{I(\theta^0) \left( \int_0^\infty \Psi''(w)w^2e^{-w}dw \right)^2}
\]
and where the quantity $I(\theta)$ is the Fisher information.

Suppose that $W$ is an exponential random variable with expectation $1$ and that $\lim_{w \to 0} \Psi'(w)w^2e^{-w} = \lim_{w \to \infty} \Psi'(w)w^2e^{-w} = 0$. Then by an integration by parts,

$$\int_0^\infty \Psi''(w)w^2e^{-w}dw = E[\Psi''(W)W^2] = E[\Psi'(W)W^2] - 2E[\Psi'(W)W].$$

Hence, $\sigma_\Psi^2$ is equal to


By the Cauchy-Schwarz inequality,

$$V[\Psi'(W)W] \geq (Cov[\Psi'(W)W, W])^2,$$

with equality if and only if, for some constants $C_1$ and $C_2$, $\Psi'(w) = C_1 + C_2/w$. So here, an asymptotically efficient estimation procedure is obtained if and only if $\Psi(w) = C_1w + C_2\log w + C_3$, for some constants $C_1$, $C_2 > 0$ and $C_3$. It should be observed that this estimator $\hat{\theta}_{\Psi,n}$ does not depend on the chosen values of $C_1$, $C_2 > 0$ and $C_3$. Thus, we may choose $C_2 = 1$ and $C_1 = C_3 = 0$, which gives us the original MSP-estimator proposed by Cheng and Amin (1983) and independently by Ranneby (1984).

However the original MSP-estimator has the disadvantage that, just as for the ML-method, it does not generally possess the property of stability under small perturbations of the underlying model. That is, the distribution of the estimator can be greatly perturbed if the assumed model is only approximately true. A simulation study by Nordahl (1992), where the ML-method is compared with different GMSP-estimators, illustrates this phenomenon. The model there is a normal distribution with an unknown location parameter, $F_\theta(x) = \Phi(x - \theta)$, but the samples are generated from an $\varepsilon$-contaminated normal distribution $G(x - \theta^0)$, where $G(x) = (1 - \varepsilon)\Phi(x) + \varepsilon H(x)$. (Suggested interpretation: most observations are distributed according to $\Phi(x - \theta^0)$, but a small fraction $\varepsilon$ of them are distributed according to $H(x - \theta^0)$. This might happen if the observations are made with an instrument that jams with probability $\varepsilon$.) In the simulations, $H(x) = \Phi((x - \rho)/\tau)$, $\tau > 0$. It is seen that the original MSP-method and the ML-method give nearly identical results, as is to be expected, and that the contaminating distribution highly influences these two estimators. For instance, if $\rho \neq 0$, then these two estimators will have the asymptotic bias $\varepsilon \rho$.

An interesting alternative is the estimator based on the Hellinger distance with $p = 2$, i.e. the estimator $\hat{\theta}_{\Psi,n}$ with $\Psi(y) = -(1 - \sqrt{x})^2$. In this case,

$$S_{\Psi,n}(\theta) \overset{P}{\to} -\frac{1}{2\sqrt{\pi}} \int \left(\sqrt{g(x)} - \sqrt{f_\theta(x)}\right)^2 dx - \frac{1}{\sqrt{\pi}},$$
where \( g \) is the density function corresponding to the underlying distribution \( G \) that has generated the sample. It follows that \( \hat{\theta}_{\psi,n} \) tends in probability to the parameter value \( \theta^* \in \Theta \) that minimizes the Hellinger distance

\[
\left( \int (\sqrt{g(x)} - \sqrt{f_{\theta}(x)})^2 \, dx \right)^{1/2},
\]

which gives \( \theta^* = \theta^0 \) if \( g(x) = f_{\theta^0}(x) \) a.e. and \( \theta^0 \in \Theta \). For this estimator the contaminating distribution has a comparatively low influence of the mean square error of \( \hat{\theta}_{\psi,n} \).

In Figure 1 below the asymptotic bias of \( \hat{\theta}_{\psi,n} \) is compared with the bias of the ML-estimator.

\[\text{(a): } \varepsilon = 0.1, \tau = 1 \]
\[\text{(b): } \varepsilon = 0.1, \rho = 2 \]

Figure 1. Asymptotic bias of the ML-estimator (dotted lines) and of the GMSP-estimator based on the Hellinger distance with \( p=2 \) (solid lines) when the data are generated from \( G(x-\theta^0) \).

Unfortunately this procedure entails a loss of asymptotic efficiency under the true model. It is not even the estimator with the lowest asymptotic variance among estimators based on the family of Hellinger distances. However, its asymptotic variance is very close to the lowest among the Hellinger based estimators. (The lowest asymptotic variance is approximately \( 1.09269/I(\theta^0) \) and is obtained when \( p \approx 2.04726 \). The asymptotic variance when \( p = 2 \) equals \( (16 - 4\pi)/(I(\theta^0)\pi) \approx 1.09296/I(\theta^0) \). See Figure 2(c).) Moreover the estimator with \( p = 2 \) is certainly the easiest to work with in the class of Hellinger based estimators.
The properties of the Hellinger distance, with $p = 1$ and $p = 2$, have been used before in problems in the theory of statistical estimation, and in the construction of consistent estimators, see e.g. Ibragimov and Has'minskii (1981). Of particular interest is the parametric estimation procedure constructed by Beran (1977), the minimum Hellinger distance estimator. The estimator is defined as the value $\hat{\theta}_n$ in the parameter space which minimizes the Hellinger distance ($p = 2$) between $f_{\theta_n}$ and $\hat{g}_n$, where $\hat{g}_n$ is a suitable nonparametric density estimator, e.g. a kernel density estimator. This estimator is asymptotically efficient and minimax robust in Hellinger metric neighbourhoods of the given model. It should be noted that the ideas behind Beran's method are related to those in this paper. The difference is that here we do not need to perform a density estimation before the Hellinger distance is calculated, since we use an approximation of this distance.

Another interesting family of GMSP-estimators are those based on Rényi's divergence of order $\alpha$. For $\alpha = 1/2$ the estimator is identical to the Hellinger based estimator with $p = 2$. At a first sight it is appealing to choose $\alpha$ as small as possible (see figure 2(b)). But this gives an estimator which is extremely unstable,
even for very small perturbations in the underlying model. This is not surprising since $\Psi(x) = x^\alpha \text{sgn}(1 - \alpha)$ is a very flat function when the constant $\alpha$ is close to zero. Further simulations of the $\varepsilon$-contaminated normal distribution $G(x - \theta^0)$, under the normal distribution model $F_\theta(x) = \Phi(x - \theta)$, has shown that $\alpha = 0.1$ seems to be a good compromise between the desired properties of efficiency and stability. This estimator has a low asymptotic variance $(1.00552/I(\theta^0))$ under the true model and behaves similarly to the Hellinger based estimator, with $p = 2$, when the underlying distribution is disturbed.

The estimator based on Jeffreys divergence has the asymptotic variance $\approx 1.07247/I(\theta^0)$ under the true model, but is not as stable as the “best” GMSP-estimators based on the Hellinger distance or the Rényi divergence.

### 5 Proofs

**Lemma 1** Let $W_1, \ldots, W_{n+1}$ be an i.i.d. sample of exponentially distributed random variables with mean 1. Set $Z_j = W_j / (W_1 + \ldots + W_{n+1})$, $j = 1, \ldots, n + 1$. Then $(Z_1, \ldots, Z_{n+1})$ is distributed as the set of $n+1$ simple spacings determined by $n$ independent uniform random variables $U_1, \ldots, U_n$, on $[0,1]$, i.e.

$$\{U(i) - U(i_{i-1}) : 1 \leq i \leq n + 1\} \overset{\text{D}}{=} \left\{ W_i / \sum_{j=1}^{n+1} W_j : 1 \leq i \leq n + 1 \right\}.$$

**Proof.** See Pyke (1965).

**Proof of Theorem 1.** For a positive number $\delta$, let

$$\overline{T}_\delta = \int_0^\infty \sup_{-\delta \leq d \leq \delta} \Psi \left( \frac{w}{1+d} \right) e^{-w} dw \quad \text{and} \quad \underline{T}_\delta = \int_0^\infty \inf_{-\delta \leq d \leq \delta} \Psi \left( \frac{w}{1+d} \right) e^{-w} dw.$$

By assumption $A_2$,

$$\lim_{\delta \downarrow 0} \overline{T}_\delta = \lim_{\delta \downarrow 0} \underline{T}_\delta = \int_0^\infty \Psi(w) e^{-w} dw = T_\Psi(\theta^0).$$

Hence for each $\varepsilon > 0$, there exists $\delta_0 > 0$ such that

$$|T_\Psi(\theta^0) - \overline{T}_{\delta_0}| < \varepsilon / 2 \quad \text{and} \quad |T_\Psi(\theta^0) - \underline{T}_{\delta_0}| < \varepsilon / 2. \quad (3)$$

Let $W_1, \ldots, W_{n+1}$ be independent standard exponential random variables. Then, using the representation of uniform spacings in Lemma 1 (note that $F_\theta(\xi_i)$ is
uniformly distributed), and Bonferroni’s inequality,

\[ P \left( S_{\Psi,n}(\theta^0) > T_\Psi(\theta^0) - \varepsilon \right) \]

\[ = P \left( \frac{1}{n+1} \sum_{j=1}^{n+1} \Psi \left( \frac{W_j}{n+1} \right) > T_\Psi(\theta^0) - \varepsilon \right) \]

\[ \geq P \left( \frac{1}{n+1} \sum_{j=1}^{n+1} \inf_{-\delta_0 \leq d \leq \delta_0} \Psi \left( \frac{W_j}{1+d} \right) > T_\Psi(\theta^0) - \varepsilon, \left| \frac{1}{n+1} \sum_{i=1}^{n+1} W_i - 1 \right| < \delta_0 \right) \]

\[ \geq 1 - P \left( \frac{1}{n+1} \sum_{j=1}^{n+1} \inf_{-\delta_0 \leq d \leq \delta_0} \Psi \left( \frac{W_j}{1+d} \right) \leq T_\Psi(\theta^0) - \varepsilon \right) - P \left( \left| \frac{1}{n+1} \sum_{i=1}^{n+1} W_i - 1 \right| \geq \delta_0 \right) \]

\[ \rightarrow 1 \text{ as } n \to \infty, \quad (4) \]

where the convergence follows by the law of large numbers. Likewise,

\[ P \left( S_{\Psi,n}(\theta^0) < T_\Psi(\theta^0) + \varepsilon \right) \]

\[ \geq 1 - P \left( \frac{1}{n+1} \sum_{j=1}^{n+1} \sup_{-\delta_0 \leq d \leq \delta_0} \Psi \left( \frac{W_j}{1+d} \right) \geq T_\Psi(\theta^0) + \frac{\varepsilon}{2} \right) - P \left( \left| \frac{1}{n+1} \sum_{i=1}^{n+1} W_i - 1 \right| \geq \delta_0 \right) \]

\[ \rightarrow 1 \text{ as } n \to \infty. \quad (5) \]

Finally (4) and (5) give

\[ P \left( \left| S_{\Psi,n}(\theta^0) - T_\Psi(\theta^0) \right| > \varepsilon \right) \to 0, \quad n \to \infty, \]

which completes the proof of the theorem. \( \square \)

**Lemma 2** Under assumptions A2(i) and A2(ii), the random function

\[ V_{\Psi,n}(N, \theta) = \frac{1}{n+1} \sum_{i=1}^{n} \max(0, \Psi(z_i(n, \theta)) - N) \]

converges to zero for all elementary events, uniformly in \( n \) and \( \theta \), as \( N \to \infty \) (i.e. \( \sup_{n \geq 1, \theta \in \Theta} V_{\Psi,n}(N, \theta) \to 0 \)).

**Proof.** Assume that \( \Psi \) is strictly increasing (if not, \( \Psi \) is bounded from above and the lemma follows trivially).

Now let \( t = (t_1, t_2, \ldots, t_n) \) and

\[ A_n = \left\{ t : \sum_{i=1}^{n} t_i = n + 1, \ t_i > 0, \ i = 1, \ldots, n \right\}. \]
Then by the definition of the $z_i(n, \theta)$'s,

$$
\sup_{n \geq 1, \delta \in \Theta} V_{\Psi, n}(N, \theta) \leq \sup_{n \geq 1, t \in \mathcal{A}_n} \frac{1}{n+1} \sum_{i=1}^{n} \max(0, \Psi(t_i) - N).
$$

Let $L_n = \{i \in \{1, 2, \ldots, n\} : \Psi(t_i) > N\}$ and let $l_n$ denote the number of elements in the set $L_n$. Using the fact that $\Psi$ is a concave function, we get

$$
\sup_{n \geq 1, t \in \mathcal{A}_n} \frac{1}{n+1} \sum_{i=1}^{n} \max(0, \Psi(t_i) - N) \leq \sup_{n \geq 1, t \in \mathcal{A}_n} \frac{l_n}{n+1} \sum_{i \in L_n} \frac{\Psi(t_i) - N}{l_n}
$$

$$
\leq \sup_{n+1 \geq \psi^{-1}(N)} \frac{l_n}{n+1} \left( \frac{n+1}{l_n} - N \right) = \sup_{s \geq \psi^{-1}(N)} \frac{1}{s} (\Psi(s) - N)
$$

where the right hand side tends to zero as $N \to \infty$ since $\psi^{-1}(N) \to \infty$ as $N \to \infty$ and since $\Psi(t)/t \to 0$ as $t \to \infty$. This establishes the lemma. \(\square\)

**Lemma 3** Under assumptions $A_2(i)$ and $A_2(ii)$ the function

$$
V_{\Psi}(N, \theta) = \int_{-\infty}^{\infty} \int_{0}^{\infty} \max(0, \Psi(yf_\theta(x)) - N) p(x, y) dy dx
$$

converges to zero, uniformly in $\theta$, as $N \to \infty$ (i.e. $\sup_{\theta \in \Theta} V_{\Psi}(N, \theta) \to 0$).

**Proof.** Assume that $\Psi$ is strictly increasing (if not, the lemma follows immediately). A change of variables establishes that

$$
V_{\Psi}(N, \theta) = \int_{-\infty}^{\infty} \int_{0}^{\infty} \max(0, \Psi(yf_\theta(x)) - N) f_{\phi_\theta}(x) e^{-yf_\theta(x)} dy dx
$$

$$
= \int_{-\infty}^{\infty} \int_{0}^{\infty} \max(0, \Psi(z) - N) \frac{f_{\phi_\theta}(x)}{f_\theta(x)} \exp\left(\frac{-zf_{\phi_\theta}(x)}{f_\theta(x)}\right) dz dx
$$

$$
= \int_{-\infty}^{\infty} f_{\phi_\theta}(x) \exp\left(-a_{\theta, \phi_\theta}(N, x)\right) \left(\int_{\psi^{-1}(N)}^{\infty} (\Psi(z) - N) \frac{f_{\phi_\theta}(x)}{f_\theta(x)} dz\right) dx
$$

$$
\cdot \exp\left(\frac{-zf_{\phi_\theta}(x)}{f_\theta(x)} + a_{\theta, \phi_\theta}(N, x)\right) dx
$$

where $a_{\theta, \phi_\theta}(N, x) = \psi^{-1}(N)f_{\phi_\theta}(x)/f_\theta(x)$. Further, let $b_{\theta, \phi_\theta}(N, x) = \psi^{-1}(N) + f_\theta(x)/f_{\phi_\theta}(x)$. Now,

$$
\int_{\psi^{-1}(N)}^{\infty} \frac{f_{\phi_\theta}(x)}{f_\theta(x)} \exp\left(\frac{-zf_{\phi_\theta}(x)}{f_\theta(x)} + a_{\theta, \phi_\theta}(N, x)\right) dz = 1
$$

and

$$
\int_{\psi^{-1}(N)}^{\infty} \frac{zf_{\phi_\theta}(x)}{f_\theta(x)} \exp\left(\frac{-zf_{\phi_\theta}(x)}{f_\theta(x)} + a_{\theta, \phi_\theta}(N, x)\right) dz = b_{\theta, \phi_\theta}(N, x).
$$
Since \(- (\Psi(z) - N)\) is a convex function, an application of Jensen’s inequality gives
\[
\int_{\Psi^{-1}(N)}^{\infty} (\Psi(z) - N) \frac{f_{\theta}(x)}{f_{\theta}(x)} \exp \left( -z \frac{f_{\theta}(x)}{f_{\theta}(x)} + a_{\theta,\phi}(N, x) \right) \, dz \leq \Psi(b_{\theta,\phi}(N, x)) - N.
\]

Therefore, for \(K\) chosen such that \(\sup_{x \in \mathbb{R}^+} e^{-x}(1 + x) \leq K\),
\[
\sup_{\theta \in \Theta} V_\Psi(N, \theta) \leq \sup_{\theta \in \Theta} \int_{-\infty}^{\infty} f_{\theta}(x) \exp \left( - a_{\theta,\phi}(N, x) \right) \left( \Psi(b_{\theta,\phi}(N, x)) - N \right) \, dx
\]
\[
= \sup_{\theta \in \Theta} \int_{-\infty}^{\infty} f_{\theta}(x) \exp \left( - a_{\theta,\phi}(N, x) \right) \left( 1 + a_{\theta,\phi}(N, x) \right) \frac{\Psi(b_{\theta,\phi}(N, x)) - N}{b_{\theta,\phi}(N, x)} \, dx
\]
\[
\leq K \cdot \sup_{s \geq \Psi^{-1}(N)} \frac{1}{s} (\Psi(s) - N) \cdot \sup_{\theta \in \Theta} \int_{-\infty}^{\infty} f_{\theta}(x) \, dx \to 0 \text{ as } N \to \infty,
\]
where the convergence follows since \(\Psi^{-1}(N) \to \infty\) as \(N \to \infty\), and since \(\Psi(t)/t \to 0, t \to \infty\). This establishes the lemma. \(\square\)

**Lemma 4** Suppose that assumptions \(A_1, A_2(i)\) and \(A_2(ii)\) are satisfied. Then for every fixed \(\theta \in \Theta\) and every fixed \(M \in \mathbb{R}^+\), the random function
\[
T_{\Psi, n}(M, \theta) = \frac{1}{n + 1} \sum_{i=1}^{n} \max(-M, \Psi(z_i(n, \theta)))
\]
converges in probability to \(T_{\Psi}(M, \theta)\), as \(n \to \infty\). If in addition assumption \(A_4\) is satisfied, then the convergence is uniform in \(\Theta\), that is \(\sup_{\theta \in \Theta} |T_{\Psi, n}(M, \theta) - T_{\Psi}(M, \theta)| \xrightarrow{P} 0, n \to \infty\).

**Proof.** Define
\[
H_{\Psi, n}(M, N, \theta) = \frac{1}{n + 1} \sum_{i=1}^{n} \Psi_{M,N}(z_i(n, \theta))
\]
and
\[
H_{\Psi}(M, N, \theta) = \int_{-\infty}^{\infty} \int_{0}^{\infty} \Psi_{M,N}(yf_{\theta}(x)) p(x, y) \, dy \, dx
\]
where
\[
\Psi_{M,N}(x) = \begin{cases} 
-M & \text{if } \Psi(x) \leq -M \\
\Psi(x) & \text{if } -M < \Psi(x) < N \\
N & \text{if } N \leq \Psi(x)
\end{cases}
\]
Recall that

\[ P_n(x, y) = P_n(\xi_1 \leq x, \eta_1(n) \leq y) = \begin{cases} \int_{-\infty}^x \left[ 1 - \left( 1 - \left( F_{\varphi}(s + \frac{y}{n+1}) - F_{\varphi}(s) \right) \right)^n \right] f_{\varphi}(s) \, ds & \text{if } 0 < y < \infty, \\ \int_{-\infty}^y f_{\varphi}(s) \, ds & \text{if } y = \infty. \end{cases} \]

The density function of \( P_n(x, y) \) is given by

\[ p_n(x, y) = \begin{cases} \frac{n-1}{n+1} f_{\varphi}(x) f_{\varphi}(x + \frac{y}{n+1}) \left( 1 - \left( F_{\varphi}(x + \frac{y}{n+1}) - F_{\varphi}(x) \right) \right)^{-n} & \text{if } 0 < y < \infty, \\ f_{\varphi}(x) (F_{\varphi}(x))^{n-1} & \text{if } y = \infty \end{cases} \]

with respect to the measure \( \lambda \) which is the two-dimensional Lebesgue measure \( \mu \times \nu \) on \( \mathbb{R} \times \mathbb{R}^+ \), and the Lebesgue measure \( \mu \) on \( \mathbb{R} \times \{+\infty\} \). The random variables \( z_i(n, \theta), i = 1, 2, \ldots, n \), are exchangeable. Hence

\[ E[H_{\psi,n}(M, N, \theta)] = \frac{n}{n+1} E[H_{\psi,M,N}(z_1(n, \theta))], \]
\[ Var[H_{\psi,n}(M, N, \theta)] = \frac{n}{(n+1)^2} Var[H_{\psi,M,N}(z_1(n, \theta))] + \frac{n^2-n}{(n+1)^2} Cov[H_{\psi,M,N}(z_1(n, \theta)), H_{\psi,M,N}(z_2(n, \theta))]. \]

As \( \Psi_{M,N}(x), x > 0 \), is a bounded continuous function, and since by assumption \( A_1, z(n, \theta, x, y) \rightarrow y f_{\theta}(x) \) and \( p_n(x, y) \rightarrow p(x, y) \) as \( n \rightarrow \infty \) for almost all \((x, y) \in \mathbb{R} \times \mathbb{R}^+\), it follows from Lebesgue’s dominated convergence theorem that

\[ E[\Psi_{M,N}(z_1(n, \theta))] = \int_{-\infty}^\infty \int_0^\infty \Psi_{M,N}(z(n, \theta, x, y)) p_n(x, y) \, dy \, dx \\
+ \int_{-\infty}^\infty \lim_{y \to \infty} \Psi_{M,N}(z(n, \theta, x, y)) f_{\theta}(x) (F_{\varphi}(x))^{n-1} \, dx \\
= H_\psi(M, N, \theta). \]

Similarly, it can be shown that

\[ E[\Psi_{M,N}(z_1(n, \theta)) \cdot \Psi_{M,N}(z_2(n, \theta))] \rightarrow H_\psi^2(M, N, \theta), \quad n \rightarrow \infty, \]

so that

\[ Cov[\Psi_{M,N}(z_1(n, \theta)), \Psi_{M,N}(z_2(n, \theta))] \rightarrow 0, \quad n \rightarrow \infty. \]
Thus $E [H_{\Psi,n}(M,N,\theta)] \to H_{\Psi}(M,N,\theta)$ and $\text{Var} [H_{\Psi,n}(M,N,\theta)] \to 0$ and hence

$$H_{\Psi,n}(M,N,\theta) \overset{P}{\to} H_{\Psi}(M,N,\theta) \quad \text{as} \quad n \to \infty.$$  \hfill (6)

The first part of the lemma, the pointwise convergence of $T_{\Psi,n}(M,\theta)$, now follows as a consequence of the inequalities

$$H_{\Psi,n}(M,N,\theta) - H_{\Psi}(M,N,\theta) - V_{\Psi}(N,\theta)$$

$$\leq T_{\Psi,n}(M,\theta) - T_{\Psi}(M,\theta)$$

$$\leq H_{\Psi,n}(M,N,\theta) + V_{\Psi,n}(N,\theta) - H_{\Psi}(M,N,\theta),$$

(7)

together with Lemma 2 and 3.

Next, let $\rho > 0$ be arbitrary. By Lemmas 2 and 3, there exists an integer $N = N(\rho)$ such that

$$\sup_{n \geq 1, \theta \in \Theta} V_{\Psi,n}(N,\theta) \leq \frac{\rho}{4} \quad \text{and} \quad \sup_{\theta \in \Theta} V_{\Psi}(N,\theta) \leq \frac{\rho}{4}.$$  \hfill (8)

Since $\Psi$ is a concave function, it satisfies a Lipschitz condition on the set $A_M = \{ t \in R^+ : \Psi(t) \geq -M \}$, i.e. there exists a constant $C_M$ such that

$$|\Psi(x) - \Psi(y)| < C_M|x - y| \quad \text{for all} \quad x, y \in A_M.$$  

Choose the sets $K_j$ and $O_j$, $j = 1, \ldots, r$, in assumption $A_4$ such that

$$P((X,Y) \in K_j) > 1 - \frac{\rho}{16 \max(M,N)}$$

and

$$\sup_{(x,y) \in K_j, \theta, \theta' \in O_j} \sup_{\theta, \theta' \in O_j} |z(n,\theta,x,y) - z(n,\theta',x,y)| < \frac{\rho}{8C_M}$$

(9)

for all $n$ large enough. Clearly for $\theta, \theta' \in O_j$ and $n$ sufficiently large

$$|\Psi_{M,N}(z(n,\theta,x,y)) - \Psi_{M,N}(z(n,\theta',x,y))| < \begin{cases} \rho/8 & \text{on } K_j \\ \max(M,N) & \text{on } K_j^c. \end{cases}$$

Notice that

$$H_{\Psi}(M,N,\theta) = \lim_{n \to \infty} \int_{K_j} \Psi_{M,N}(z(n,\theta,x,y))p(x,y)dydx + \lim_{n \to \infty} \int_{K_j^c} \Psi_{M,N}(z(n,\theta,x,y))p(x,y)dydx.$$
Thus
\[ \max_{1 \leq j \leq r} \sup_{\theta, \theta' \in O_j} |H_{\Psi}(M, N, \theta) - H_{\Psi}(M, N, \theta')| < \frac{\rho}{8} + \frac{\rho}{8} = \frac{\rho}{4}. \] (10)

Furthermore
\[
|H_{\Psi, n}(M, N, \theta) - H_{\Psi, n}(M, N, \theta')| \\
< \frac{1}{n} \sum_{i=1}^{n} |\Psi_{M,N}(z_i(n, \theta)) - \Psi_{M,N}(z_i(n, \theta'))| I((\xi, n_i(n)) \in K_j) \\
+ \frac{1}{n} \sum_{i=1}^{n} |\Psi_{M,N}(z_i(n, \theta)) - \Psi_{M,N}(z_i(n, \theta'))| I((\xi, n_i(n)) \in K_j). 
\]

The boundary $\delta K_j$ has $P$-measure zero and thus
\[
\frac{1}{n} \sum_{i=1}^{n} I((\xi, n_i(n)) \in K_j) \quad \xrightarrow{P} \quad P \left( (X, Y) \in K_j^c \right) < \frac{\rho}{16 \max(M, N)}. 
\]

Therefore it follows from inequality (9) that
\[
P \left( \max_{1 \leq j \leq r} \sup_{\theta, \theta' \in O_j} |H_{\Psi, n}(M, N, \theta) - H_{\Psi, n}(M, N, \theta')| < \frac{\rho}{4} \right) \to 1 \quad \text{as} \quad n \to \infty. \quad (11)
\]

Choose some fixed parameter values $\theta_j \in O_j$, $j = 1, \ldots, r$. Then by combining the results (6)-(8), (10) and (11), the following relation holds with a probability tending to 1, as $n$ tends to infinity,
\[
\sup_{\theta \in \Theta} |T_{\Psi, n}(M, \theta) - T_{\Psi}(M, \theta)| \leq \sup_{\theta \in \Theta} |H_{\Psi, n}(M, N, \theta) - H_{\Psi}(M, N, \theta)| + \frac{\rho}{4} \\
\leq \max_{1 \leq j \leq r} \sup_{\theta \in \Theta_j} |H_{\Psi, n}(M, N, \theta) - H_{\Psi, n}(M, N, \theta_j)| \\
+ \max_{1 \leq j \leq r} |H_{\Psi, n}(M, N, \theta_j) - H_{\Psi}(M, N, \theta_j)| \\
+ \max_{1 \leq j \leq r} |H_{\Psi}(M, N, \theta_j) - H_{\Psi}(M, N, \theta)| + \frac{\rho}{4} \\
< \rho.
\]

This completes the proof of Lemma 4. \Box

By combining the results from Lemma 4 and Theorem 2 we are now able to prove the main theorem.

\textbf{Proof of Theorem 2.} Let $\delta > 0$ be arbitrary and denote
\[
T_{\Psi}(\theta^0) - \sup_{\theta \in \Theta_\delta} T_{\Psi}(M, \theta) = a(M, \delta).
\]
By assumption $A_3$, there exists an integer $M_1$ such that $a(M, \delta) > 0$ for all $M \geq M_1$. Put $\varepsilon = a(M_1, \delta)$. Moreover,

$$S_{\Psi,n}(\theta) = \frac{1}{n+1} \sum_{i=1}^{n} \Psi(z_i(n, \theta)) + \frac{1}{n+1} \Psi((n+1)F_{\theta}\left(\min_{1 \leq j \leq n} \xi_j\right))$$

$$\leq T_{\Psi,n}(M, \theta) + \frac{\Psi(n+1)^+}{n+1},$$

so if $n$ is chosen such that $c_n < \varepsilon/4$ and $\Psi(n+1)^+/(n+1) < \varepsilon/4$, then

$$T_{\Psi,n}(M, \hat{\theta}_{\Psi,n}) \geq S_{\Psi,n}(\hat{\theta}_{\Psi,n}) - \frac{\varepsilon}{4} \geq S_{\Psi,n}(\theta^0) - \frac{\varepsilon}{2}.$$ 

It follows from Lemma 4 and Theorem 2 that

$$T_{\Psi}(M, \hat{\theta}_{\Psi,n}) > T_{\Psi,n}(M, \hat{\theta}_{\Psi,n}) - \varepsilon/4$$

and

$$S_{\Psi,n}(\theta^0) > T_{\Psi}(\theta^0) - \varepsilon/4,$$

both hold with a probability tending to 1 as $n$ tends to infinity. Hence, for all $M \geq M_1$,

$$T_{\Psi}(M, \hat{\theta}_{\Psi,n}) > T_{\Psi}(\theta^0) - \varepsilon \geq T_{\Psi}(\theta^0) - a(M, \delta) = \sup_{\theta \in \Theta} T_{\Psi}(M, \theta)$$

holds with a probability tending to 1 as $n \to \infty$. This implies that

$$P\left(\hat{\theta}_{\Psi,n} \in \{\theta : |\theta - \theta^0| < \delta\}\right) \to 1 \quad \text{as} \quad n \to \infty,$$

and since $\delta$ is arbitrary the theorem follows. \(\square\)

**References**


