Cramér-Rao Lower Bound for Independent Component Analysis

Zbyněk Koldovský and Petr Tichavský

October 21, 2004

Abstract

Instantaneous linear model is well-studied problem in Blind Source Separation. Many algorithms have been developed [2] for this purpose, therefore, their mutual comparison [3] or analysis of their efficacy is practical issue. In this paper, a detailed derivation of the Rao-Cramer lower bound for Independent Component Analysis is provided, which gives the theoretical lower bound of separation performance.

1 Introduction

Independent Component Analysis (ICA) is a well-known method for Blind Source Separation. Its purpose is to recover unknown original sources from their linear mixtures without any a priori information. The linear ICA model is

$$\mathbf{X} = \mathbf{A}\mathbf{S}, \tag{1}$$

where

$$\mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1N} \\ \vdots & \ddots & \vdots \\ x_{d1} & \cdots & x_{dN} \end{pmatrix}$$

denotes a matrix of the mixed signals,

$$\mathbf{S} = \begin{pmatrix} s_{11} & \cdots & s_{1N} \\ \vdots & \ddots & \vdots \\ s_{d1} & \cdots & s_{dN} \end{pmatrix}$$

is a matrix of the original signals, and $\mathbf{A}$ is an unknown regular $d \times d$ mixing matrix. Basic assumption of ICA is the independence of the original signals. In this article, we consider a model when s_{ij} are i.i.d. random variables with probability density function $f_i(s_{ij})$.

When introducing a new algorithm it is always important to show its performance properties. This can be provided through some theoretical analysis or computer simulations. It is also usual to compare the results with another methods. There are many algorithms for ICA which has been developed in last two decades because of increasing attention to this problem in the signal processing community. In this article, we derive the Rao-Cramer lower bound for unbiased estimation of the elements of the mixing matrix $\mathbf{A}$(or the de-mixing matrix $\mathbf{W} = \mathbf{A}^{-1}$), which can give us a general performance criterion for ICA.

2 Cramér-Rao lower bound for ICA

When a vector of parameters θ is estimated from a data vector x that have probability density $f_{x|\theta}(x|\theta)$, the Cramér-Rao lower bound (CRB) is the lower bound for the variance of any unbiased estimator $\hat{\theta}$. Assume that the following Fisher information matrix and its inversion exist:

$$\mathbf{F} = \mathrm{E}_\theta \left[\frac{1}{f_{x|\theta}^2} \frac{\partial f_{x|\theta}(x|\theta)}{\partial \theta} \left(\frac{\partial f_{x|\theta}(x|\theta)}{\partial \theta} \right)^T \right] \tag{2}$$

Then, under the regularity condition [4], i.e.,

1. $\mathrm{supp} f_{x|\theta} = \{x | f_{x|\theta}(x|\theta) \neq 0\}$ is independent of θ
2. $\exists \frac{\partial f_{x|\theta}}{\partial \theta}$ for almost all x.
3. $\mathrm{E}\left[\frac{\partial \ln f_{x|\theta}}{\partial \theta}\right] = 0$

it holds

$$\mathrm{cov}\, \hat{\theta} \geq \mathrm{CRB}_\theta = \mathbf{F}^{-1}$$

If $\boldsymbol{\varphi} = \boldsymbol{\varphi}(\theta)$ is a differentiable function of θ, then the Fisher information matrix for $\boldsymbol{\varphi}$ exists as well and is equal to

$$\mathbf{F}_{\boldsymbol{\varphi}} = \mathbf{T}^{-1} \mathbf{F} \mathbf{T}^{-T}. \tag{3}$$

where $\mathbf{T}$ is the Jacobian of the mapping $\boldsymbol{\varphi}(\theta)$. If the mapping is linear, $\boldsymbol{\varphi}(\theta) = \mathbf{M}\theta$ for some regular matrix $\mathbf{M}$, then $\mathbf{T} = \mathbf{M}^T$. We shall use all these general statements in case of ICA, where $x = \mathbf{X} = \mathbf{W}^{-1}\mathbf{S}$ and $\theta = \mathrm{vec}[\mathbf{W}]$.

2.1 The Fisher Information Matrix

Due to the independence, the mutual probability density function (pdf) of the original signals $\mathbf{S}$ is $f(\mathbf{S}) = \prod_{i=1}^{d} \prod_{j=1}^{N} f_i(s_{ij})$. Than the transformation theorem claims that

$$f_{\mathbf{X}}(\mathbf{X}) = |\det \mathbf{W}| f(\mathbf{W}\mathbf{X}). \tag{4}$$

Let denote the elements of matrices $\mathbf{A}$ and $\mathbf{W}$ by a_{ij} and w_{ij}, respectively. Note that m-th element of the parameter vector θ is w_{uv} for $m = u(d-1)+v$. Using the definition (2) the mn-th entry of the Fisher information matrix, where $m = u(d-1)+v$ and $n = p(d-1)+r$, is

$$\mathbf{F}_{mn} = \mathrm{E}\left[\frac{1}{f_{\mathbf{X}}^2}\frac{\partial f_{\mathbf{X}}}{\partial w_{uv}}\frac{\partial f_{\mathbf{X}}}{\partial w_{pr}}\right] = \mathrm{E}\left[\left(\frac{|\det \mathbf{W}|^{-2}}{f^2}\right)\frac{\partial f_{\mathbf{X}}}{\partial w_{uv}}\frac{\partial f_{\mathbf{X}}}{\partial w_{pr}}\right]. \tag{5}$$

To derive the latter relation (see Appendix A) we shall use following rules [5] and assumptions:

$$\frac{\partial \det \mathbf{W}}{\partial w_{ij}} = a_{ji} \det \mathbf{W} \tag{6}$$

$$\frac{\partial \mathbf{W}}{\partial w_{ij}} = \mathbf{E}_{ij} \tag{7}$$

$$\lim_{t\to\pm\infty} t f_i(t) = 0 \quad \Leftrightarrow \quad \int_R t f_i'(t)dt = -1 \tag{8}$$

$$\int_R t^2 f_i(t)dt = \sigma_i^2 < +\infty \tag{9}$$

Here $i, j = 1, \ldots, d$ and $\mathbf{E}_{ij}$ denotes a zero matrix with ij-th element equal to one. Due to the indeterminacy of the variance of the original signals [1] we can assume that $\sigma_i^2 = 1$. Also, for simplicity, zero mean value of the original signal is assumed, i.e.,

$$\int_R t f_i(t)dt = 0. \tag{10}$$

After lengthy computation (see Appendix A) we get

$$\mathbf{F}_{mn} = (N-1)^2 a_{vu}a_{rp} + N a_{vp}a_{ru} + \\ + \delta_{up} N a_{vu}a_{ru}\left(\mathrm{E}\left[s_u\psi_u(s_u)\right]^2 - 2\right) + \delta_{up} N \mathrm{E}\left[\psi_u^2(s_u)\right]\sum_{\substack{j=1\\ j\neq u}}^{d} a_{vj}a_{rj}, \tag{11}$$

where ψ_u denotes the score function of the corresponding pdf, i.e., $\psi_u(x) = -\frac{f_u'(x)}{f_u(x)}$. The assumption of existence of the score function and of finiteness of the expectation values in (11) follows:

$$\mathrm{E}\left[\psi_u^2(s_u)\right] = \int_R \psi_u^2(t) f_u(t)dt < +\infty \tag{12}$$

$$\mathrm{E}\left[s_u\psi_u(s_u)\right]^2 = \int_R t^2\psi_u^2(t) f_u(t)dt < +\infty. \tag{13}$$

2.2 Derivation of $\mathbf{F}^{-1}$

From the basic model (1) follows that the separation may consists in the estimation of the de-mixing matrix $\mathbf{W}$. Let $\hat{\mathbf{S}}$ denotes the estimated signals and $\widehat{\mathbf{W}}$ denotes the estimated de-mixing matrix. Then, $\hat{\mathbf{S}} = \widehat{\mathbf{W}}\mathbf{X} = \widehat{\mathbf{W}}\mathbf{A}\mathbf{S}$. It is known that the blind separation can be performed up to an indeterminacy of the variances, the signs, and the permutation of the original signals [1]. Because the indeterminacy of the variance of each signal is equivalent to the indeterminacy of the norm of corresponding row of the de-mixing matrix, the accuracy of the estimation of the de-mixing matrix can not be used when deriving the accuracy of the estimation of the original signals. Therefore, we will focus on the estimation of the so-called *gain* matrix $\mathbf{G} = \widehat{\mathbf{W}}\mathbf{A}$, which is just linear transformation of $\widehat{\mathbf{W}}$. Also the estimation of $\widehat{\mathbf{W}}$ can be viewed as a transformed estimation of $\widehat{\mathbf{W}}$ for a case $\mathbf{A} = \mathbf{I}$ (identity matrix), denoted by $\widehat{\mathbf{W}}_{\mathbf{I}}$. Of course, the latter transformation cancels the former, consequently, we can see that the matrix $\mathbf{G}$ equals to $\widehat{\mathbf{W}}_{\mathbf{I}}$ and is independent of the mixing matrix $\mathbf{A}$. The inversion of the Fisher information matrix can be derived just for the case $\mathbf{A} = \mathbf{I}$, thus, (11) simplifies to

$$(\mathbf{F_I})_{mn} = (N-1)^2\delta_{vu}\delta_{rp} + N\delta_{vp}\delta_{ru} + N\Big(\delta_{up}\delta_{vu}(\delta_{ru}\eta_u - \delta_{rv}\kappa_u) + \delta_{up}\kappa_u\delta_{rv}\Big), \quad (14)$$

where $\eta_u = \mathrm{E}\left[s_u\psi_u(s_u)\right]^2 - 2$ and $\kappa_u = \mathrm{E}\left[\psi_u^2(s_u)\right]$. Note that we are interested only in the CRB of the variance of the non-diagonal elements of $\mathbf{G}$, which cause the inter-signal interference and correspond to the diagonal elements of $(\mathcal{F}^{-1})_{mm}$, where $m = (i-1)d + j$, $i \neq j$. Using this, in Appendix B we derive that

$$(\mathbf{F_I}^{-1})_{mm} = \frac{1}{N}\frac{\kappa_j}{\kappa_i\kappa_j - 1} = \frac{1}{N}\frac{\mathrm{E}\left[\psi_j^2(s_j)\right]}{\mathrm{E}\left[\psi_i^2(s_i)\right]\mathrm{E}\left[\psi_j^2(s_j)\right] - 1}, \quad (15)$$

which give us the desired lower bound

$$\mathrm{var}(\widehat{\mathbf{G}}_{ij}) \geq \mathrm{CRB}(\widehat{\mathbf{G}}_{ij}) = (\mathbf{F_I}^{-1})_{mm}. \quad (16)$$

3 Conclusion

In this paper, we derive the CRB for ICA under general assumptions. It is shown that the latter depends only on the probability density function of the original signals. Since $\mathrm{E}[\psi^2(s)] \geq 1$ (see Appendix C), where the equality holds if and only if s has standard Gaussian distribution function, the separation can be performed under assumption that at most one of the original signals is Gaussian. This is, indeed, well-known feature [2] in ICA.

References

[1] P. Tichavský and Z. Koldovský, "Optimal Pairing of Signal Components Separated by Blind Techniques", *IEEE Signal Processing Letters*, vol. 11, no. 2, pp. 119–122, 2004.

[2] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, Wiley-Interscience, New York, 2001.

[3] X. Giannakopoulos, J. Karhunen, and E. Oja, "Experimental Comparison of Neural Algorithms for Independent Component Analysis and Blind Separation," *Int. J. of Neural Systems*, vol. 9, pp. 651-656, 1999.

[4] R. C. Rao, *Linear Statistical Inference and Its Applications*, 2nd ed. Wiley, New York, 1973.

[5] T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*, Prentice Hall, New Jersey, 2000.

Appendix A

First, we derive $\frac{\partial f_{\mathbf{X}}}{\partial w_{uv}}$ in (5) using (4) and (6)-(10)

$$\frac{\partial f_x}{\partial w_{uv}} = \frac{\partial f(\mathbf{WX})|\det \mathbf{W}|}{\partial w_{uv}} = \frac{|\det \mathbf{W}|}{\partial w_{uv}} f(\mathbf{WX}) + |\det \mathbf{W}| \frac{\partial f(\mathbf{WX})}{\partial w_{uv}} =$$

$$= |\det \mathbf{W}| a_{vu} f(\mathbf{WX}) + |\det \mathbf{W}| \sum_{k=1}^{d} \sum_{l=1}^{N} \Big(\prod_{\substack{i=1,\dots,d \\ j=1,\dots,N \\ \neg(i=k \wedge j=l)}} f_i((\mathbf{WX})_{ij}) \Big) \frac{\partial f_k((\mathbf{WX})_{kl})}{\partial w_{uv}} =$$

$$= |\det \mathbf{W}| a_{vu} f(\mathbf{WX}) + |\det \mathbf{W}| \sum_{k=1}^{d} \sum_{l=1}^{N} f(\mathbf{WX}) \frac{\frac{\partial}{\partial w_{uv}} f_k((\mathbf{WX})_{kl})}{f_k((\mathbf{WX})_{kl})}.$$

Next, let derive $\frac{\partial f_k((\mathbf{WX})_{kl})}{\partial w_{uv}}$

$$\frac{\partial f_k((\mathbf{WX})_{kl})}{\partial w_{uv}} = f_k'((\mathbf{WX})_{kl}) \frac{\partial (\mathbf{WX})_{kl}}{\partial w_{uv}} = f_k'((\mathbf{WX})_{kl}) \sum_{\kappa=1}^{d} \frac{\partial w_{k\kappa}}{\partial w_{uv}} x_{\kappa l} =$$

$$= f_k'((\mathbf{WX})_{kl}) \delta_{ku} x_{vl} = f_k'((\mathbf{WX})_{kl}) \delta_{ku} \sum_{k=1}^{d} a_{vk} s_{kl}$$

Returning to the above formula we get

$$\frac{\partial f_{\mathbf{X}}}{\partial w_{uv}} = |\det \mathbf{W}| f(\mathbf{WX}) \left[a_{vu} + \sum_{l=1}^{N} \sum_{k=1}^{d} \frac{f_u'((\mathbf{WX})_{ul})}{f_u((\mathbf{WX})_{ul})} a_{vk} s_{kl} \right].$$

From (1) follows that $\mathbf{S} = \mathbf{A}^{-1}\mathbf{X} = \mathbf{W}\mathbf{X}$, consequently,

$$\frac{\partial f_{\mathbf{X}}}{\partial w_{uv}} = |\det \mathbf{W}| f(\mathbf{S}) \left[a_{vu} + \sum_{l=1}^{N} \sum_{k=1}^{d} \frac{f'_u(s_{ul})}{f_u(s_{ul})} a_{vk} s_{kl} \right]$$

Using this we can directly compute the mn-th entry of the Fisher information matrix.

$$\mathbf{F}_{mn} = \mathrm{E}\left[\left(\frac{|\det \mathbf{W}|^{-2}}{f^2} \right) \frac{\partial f_{\mathbf{X}}}{\partial w_{uv}} \frac{\partial f_{\mathbf{X}}}{\partial w_{pr}} \right] = a_{vu} a_{rp} + a_{rp} \mathrm{E}\left[\sum_{l=1}^{N} \sum_{k=1}^{d} \frac{f'_u(s_{ul})}{f_u(s_{ul})} a_{vk} s_{kl} \right] +$$

$$+ a_{vu} \mathrm{E}\left[\sum_{i=1}^{N} \sum_{j=1}^{d} \frac{f'_p(s_{pi})}{f_p(s_{pi})} a_{rj} s_{ji} \right] + \mathrm{E}\left[\sum_{l=1}^{N} \sum_{i=1}^{N} \sum_{k=1}^{d} \sum_{j=1}^{d} \frac{f'_u(s_{ul})}{f_u(s_{ul})} \frac{f'_p(s_{pi})}{f_p(s_{pi})} s_{kl} s_{ji} a_{vk} a_{rj} \right]$$

The second and the third term are equal to $-N a_{vu} a_{rp}$, because $\mathrm{E}\left[\frac{f'_u(s_{ul})}{f_u(s_{ul})} s_{kl} \right] = -\delta_{ku}$. To simplify the last term, we shall consider two cases:

1. $u \neq p$, then

$$\sum_{l=1}^{N} \sum_{i=1}^{N} \sum_{k=1}^{d} \sum_{j=1}^{d} \underbrace{\mathrm{E}\left[\frac{f'_u(s_{ul})}{f_u(s_{ul})} \frac{f'_p(s_{pi})}{f_p(s_{pi})} s_{kl} s_{ji} \right]}_{\delta_{ku}\delta_{jp} + \delta_{kp}\delta_{ju}\delta_{il}} a_{vk} a_{rj} = N^2 a_{vu} a_{rp} + N a_{vp} a_{ru}$$

2. $u = p$, then

$$\mathrm{E}\left[\sum_{l=1}^{N} \sum_{i=1}^{N} \sum_{k=1}^{d} \sum_{j=1}^{d} \frac{f'_u(s_{ul})}{f_u(s_{ul})} \frac{f'_u(s_{ui})}{f_u(s_{ui})} s_{kl} s_{ji} a_{vk} a_{rj} \right] =$$

$$= \sum_{i=1}^{N} \sum_{k=1}^{d} \sum_{j=1}^{d} \underbrace{\mathrm{E}\left[\left[-\frac{f'_u(s_{ui})}{f_u(s_{ui})} \right]^2 s_{ki} s_{ji} \right]}_{\delta_{kj} \times \ldots} a_{vk} a_{rj} +$$

$$+ \sum_{\substack{i,l=1 \\ i \neq l}}^{N} \sum_{k=1}^{d} \sum_{j=1}^{d} \underbrace{\mathrm{E}\left[\frac{f'_u(s_{ul})}{f_u(s_{ul})} \frac{f'_u(s_{ui})}{f_u(s_{ui})} s_{kl} s_{ji} \right]}_{\delta_{ku}\delta_{ju} \times \ldots} a_{vk} a_{rj} =$$

$$= \sum_{i=1}^{N} \underbrace{\mathrm{E}\left[-\frac{f'_u(s_{ui})}{f_u(s_{ui})} \right]^2}_{\mathrm{E}[\psi_u^2(\xi_u)]} \sum_{\substack{j=1 \\ j \neq u}}^{d} \underbrace{\mathrm{E}[s_{ji}^2]}_{1} a_{vj} a_{rj} + a_{vu} a_{ru} \sum_{i=1}^{N} \underbrace{\mathrm{E}\left[-\frac{f'_u(s_{ui})}{f_u(s_{ui})} s_{ui} \right]^2}_{\mathrm{E}[\psi_u^2(\xi_u)\xi_u^2]} +$$

$$+ \sum_{\substack{i,l=1 \\ i \neq l}}^{N} a_{vu} a_{ru} = N \Big(\mathrm{E}[\psi_u^2(\xi_u)] \sum_{\substack{j=1 \\ j \neq u}}^{d} a_{vj} a_{rj} + \big(\mathrm{E}[\psi_u^2(\xi_u)\xi_u^2] + (N-1) \big) a_{vu} a_{ru} \Big)$$

Here ξ_u denotes a random variable with pdf f_u, and ψ_u denotes its score function, i.e., $\psi_u(x) = -\frac{f'_u(x)}{f_u(x)}$. After few simplifications (11) follows. ■

Appendix B

Definition (14) can be rewritten as $\mathbf{F_I} = (N-1)^2\mathbf{F}_1 + N(\mathbf{P}+\boldsymbol{\Sigma})$, where mn-th element of $\mathbf{F}_1$, $\mathbf{P}$ and $\boldsymbol{\Sigma}$ are $\delta_{ji}\delta_{vu}$, $\delta_{ju}\delta_{vi}$, and $\delta_{ji}\delta_{vu}\delta_{vi}(\eta_i - \kappa_i) + \delta_{iu}\delta_{vj}\kappa_i$, respectively, for $m = (i-1)d+j$ and $n = (u-1)d+v$. Note that $\mathbf{F}_1$ is a rank-one matrix, $\mathbf{F}_1 = \mathbf{e}\mathbf{e}^T$, where $\mathbf{e} = \mathrm{vec}(\mathbf{I})$. Next,

$$\mathbf{P} = \begin{pmatrix} \mathbf{e}_1\mathbf{e}_1^T & \mathbf{e}_2\mathbf{e}_1^T & \cdots & \mathbf{e}_d\mathbf{e}_1^T \\ \mathbf{e}_1\mathbf{e}_2^T & \ddots & & \vdots \\ \vdots & & & \vdots \\ \mathbf{e}_1\mathbf{e}_d^T & \cdots & & \mathbf{e}_d\mathbf{e}_d^T \end{pmatrix} \quad \text{and}$$

$$\boldsymbol{\Sigma} = \mathrm{diag}(\underbrace{\eta_1, \kappa_1, \ldots, \kappa_1}_{d}, \underbrace{\kappa_2, \eta_2, \kappa_2, \ldots, \kappa_2}_{d}, \ldots),$$

where $\mathbf{e}_i$ denotes the i-th column of $\mathbf{I}$.

Applying the matrix inversion lemma

$$(\mathbf{M}+\mathbf{B}\mathbf{C})^{-1} = \mathbf{M}^{-1} - \mathbf{M}^{-1}\mathbf{B}(\mathbf{I}+\mathbf{C}\mathbf{M}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{M}^{-1},$$

for $\mathbf{M} = \mathbf{P}+\boldsymbol{\Sigma}$, $\mathbf{B} = \mathbf{e}$, and $\mathbf{C} = \mathbf{e}^T$ we get

$$\mathbf{F_I}^{-1} = \frac{1}{N}\left[(\mathbf{P}+\boldsymbol{\Sigma})^{-1} - \frac{(\mathbf{P}+\boldsymbol{\Sigma})^{-1}\mathbf{e}\mathbf{e}^T(\mathbf{P}+\boldsymbol{\Sigma})^{-1}}{\frac{N}{(N-1)^2} + \mathbf{e}^T(\mathbf{P}+\boldsymbol{\Sigma})^{-1}\mathbf{e}}\right] \tag{17}$$

To compute the inversion $(\mathbf{P}+\boldsymbol{\Sigma})^{-1}$ note that $\boldsymbol{\Sigma}$ is diagonal and $\mathbf{P}$ is a special permutation matrix such that $\mathbf{P}\mathrm{vec}(\mathbf{M}) = \mathrm{vec}(\mathbf{M}^T)$ for any $d \times d$ matrix $\mathbf{M}$. Moreover, $\mathbf{P}$ obeys $\mathbf{P}\mathbf{P} = \mathbf{I}$, and for any diagonal matrix $\mathbf{D} = \mathrm{diag}(\mathbf{d})$ it holds that

$$\mathbf{P}\mathbf{D} = \mathbf{D}'\mathbf{P},$$

where $\mathbf{D}' = \mathrm{diag}(\mathbf{P}\mathbf{d}) = \mathbf{P}\mathbf{D}\mathbf{P}$. These facts can be used to show that the inversion of $\mathbf{P}+\boldsymbol{\Sigma}$ can be written in the form $\mathbf{D}_1 + \mathbf{D}_2\mathbf{P}$ for suitable diagonal matrices $\mathbf{D}_1$ and $\mathbf{D}_2$. The equality

$$(\mathbf{P}+\boldsymbol{\Sigma})(\mathbf{D}_1+\mathbf{D}_2\mathbf{P}) = \mathbf{I}$$

is fulfilled for $\boldsymbol{\Sigma}\mathbf{D}_1 + \mathbf{D}_2' = \mathbf{I}$ and $\mathbf{D}_1' + \boldsymbol{\Sigma}\mathbf{D}_2 = \mathbf{0}$. Hence

$$\mathbf{D}_1 = (\boldsymbol{\Sigma}'\boldsymbol{\Sigma} - \mathbf{I})^{-1}\boldsymbol{\Sigma}' \quad \text{and} \quad \mathbf{D}_2 = -\boldsymbol{\Sigma}^{-1}\mathbf{D}_1'$$

where $\boldsymbol{\Sigma}' = \mathbf{P}\boldsymbol{\Sigma}\mathbf{P}$ and $\mathbf{D}_1' = \mathbf{P}\mathbf{D}_1\mathbf{P}$. Finally, to see that $(\mathbf{F_I}^{-1})_{mm} = N^{-1}(\mathbf{D}_1)_{mm}$ for $m = (i-1)d+j$, $i \neq j$, we must prove that $(\mathbf{D}_2)_{mm} = 0$ and so does the second term in (17). This, indeed, easily follows using the special properties of the matrices $\mathbf{F}_1$ and $\mathbf{P}$, therefore, the assertion (15) follows from the definition of $\mathbf{D}_1$, which concludes the appendix. ■

Appendix C

Let $f(x)$ is a differentiable probability density function for almost all x and for which the first and the second moment exists. Let the score function fulfils (12). Then we can start with the equality

$$\int_{\mathbb{R}} f(x)dx = 1.$$

Using the integration per partes and due to the existence of the first moment $\mu = \int_{\mathbb{R}} xf(x)dx$ we can write

$$-\int_{\mathbb{R}} (x-\mu)f'(x)dx = 1.$$

Next,

$$-\int_{\mathbb{R}} (x-\mu)\sqrt{f(x)}\frac{f'(x)}{\sqrt{f(x)}}dx = 1.$$

From the Cauchy-Schwartz inequality follows that

$$1 \leq \left[\int_{\mathbb{R}} (x-\mu)^2 f(x)dx\right]\left[\int_{\mathbb{R}} \frac{f'^2(x)}{f(x)}dx\right] = \mathrm{Var}(x)\mathrm{E}[\psi^2(x)].$$

Consequently,

$$\mathrm{E}[\psi^2(x)] \geq \frac{1}{\mathrm{Var}(x)}, \tag{18}$$

where the equality hold if and only if

$$\psi(x) \propto (x-\mu),$$

i.e., the density must be Gaussian. ■