Notes on the Spectral Aspects of Linear Prediction of Speech Based on the Absolute Error Minimization Criterion

Daniele Giacobello

February 10, 2008

Abstract

The standard linear prediction method exhibits spectral matching properties in the frequency domain due to Parseval's theorem [1]:

$$\sum_{n=-\infty}^{\infty} |e(n)|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |E(e^{j\omega})|^2 d\omega. \tag{1}$$

It is also interesting to note that minimizing the squared error in the time domain and in the frequency domain leads to the same set of equations, namely the Yule-Walker equations [3]. To the best of our knowledge, the only relation existing between the time and frequency domain error using the 1-norm is the trivial Hausdorff-Young inequality [2]:

$$\sum_{n=-\infty}^{\infty} |e(n)| < \frac{1}{2\pi} \int_{-\pi}^{\pi} |E(e^{j\omega})| d\omega, \tag{2}$$

which implies that time domain minimization does not corresponds to frequency domain minimization. It is therefore difficult to say if the 1-norm based approach is always advantageous compared to the 2-norm based approach for spectral modeling, since the statistical character of the frequency errors is not clear. In this notes, we provide a proof sketch for a possible spectral interpretation of the linear prediction based on the 1-norm error minimization criterion.

1 Linear Prediction of Speech

Linear prediction of speech assumes that a sample of the time serie $x(n)$, assumed to be reduntant and stationary, obtained by sampling a continuous speech signal $x(t)$ can be represented as a linear combination of the previous samples and some error signal $e(n)$ [4, 1]:

$$x(n) = \sum_{k=1}^{K} a_k x(n-k) + e(n), \tag{3}$$

In other words, we can consider the time series $x(n)$ as generated by all-pole filtering an excitation signal $e(n)$ through the filter:

$$H(z) = \frac{1}{1 - \sum_{k=1}^{K} a_k z^{-k}} = \frac{1}{A(z)}, \tag{4}$$

Given the signal $x(n)$ the problem is to determine the prediction coefficients vector $\mathbf{a} = [a_1, a_2, \ldots, a_K]$: this is usually done by minimizing the error according to some criterion. We can construct the cost function as depending from the coefficient vector:

$$e(n) = x(n) - \sum_{k=1}^{K} a_k x(n-k) \quad \text{for} \quad n = N_1, \ldots, N_2 \tag{5}$$

therefore the problem in (5) can be rewritten as a minimization problem:

$$\min_{\mathbf{a}} \|\mathbf{e}\|_p^p = \min_{\mathbf{a}} \|\mathbf{x} - \mathbf{X}\mathbf{a}\|_p^p \tag{6}$$

having:

$$\mathbf{x} = \begin{bmatrix} x(N_1) \\ \vdots \\ x(N_2) \end{bmatrix}, \mathbf{X} = \begin{bmatrix} x(N_1-1) & \cdots & x(N_1-K) \\ \vdots & & \vdots \\ x(N_2-1) & \cdots & x(N_2-K) \end{bmatrix} \tag{7}$$

and $\|\cdot\|_p$ is the p-norm defined as $\|\mathbf{x}\|_p = (\sum_{n=1}^{N} |x(n)|^p)^{\frac{1}{p}}$ for $p \geq 1$.

Even if we did not make any statistical assumption about the signal, by doing this we have actually assumed that the error vector has a generalized Gaussian distribuition [5] with variables indipendent and identically distributed:

$$p(\mathbf{e}) \propto \exp^{-(\lambda \|\mathbf{e}\|_p^p)} \tag{8}$$

We can see this clearly by approaching the linear prediction problem as a maximum-likelihood (ML) estimation of parameters (8):

$$\max_{\mathbf{a}} p(\mathbf{e}) = \max_{\mathbf{a}} p(\mathbf{x}|\mathbf{a}) = \min_{\mathbf{a}} \ln(p(\mathbf{x}|\mathbf{a})) = \min_{\mathbf{a}} \|\mathbf{x} - \mathbf{X}\mathbf{a}\|_p^p = \min_{\mathbf{a}} \|\mathbf{e}\|_p^p \quad (9)$$

same conclusion as in (6).

2 Spectral Matching Properites of 2-norm based linear prediction of speech

Having considered the signal $x(n)$ generated by an auto-regressive process, we can rewrite (5) in the $z-$transform domain:

$$E(z) = \left[1 - \sum_{k=1}^{K} a_k z^{-k}\right] X(z) = A(z)X(z) \quad (10)$$

Assuming $x(n)$ deterministic, we can apply the Parseval's theorem, the total error to be minimized is then given by:

$$E = \sum_{n=-\infty}^{\infty} e^2(n) = \int_{-1/2}^{1/2} |E(e^{j2\pi f})|^2 df \quad (11)$$

where $E(e^{j2\pi f})$ is obtain evaluating $E(z)$ on the unit circle $z = e^{j2\pi f}$. Denoting the power spectra of the signal as:

$$\hat{S}_{xx}(f, \mathbf{x}) = \frac{|E(e^{j2\pi f})|^2}{|A(e^{j2\pi f})|^2} \quad (12)$$

and its approximation as:

$$S_{xx}(f) = \frac{\sigma^2}{|A(e^{j2\pi f})|^2}. \quad (13)$$

We can easily see that the spectrum $|E(e^{j2\pi f})|^2$ is being modelled by a flat spectrum with magnitude σ^2, this means that the error signal obtained with 2-norm minimization is an approximation of a white noise, because of this $A(z)$ is sometimes known as "whitening filter". From (11,12,13) we obtain that the total error can be rewritten as:

$$E = \sigma^2 \int_{-1/2}^{1/2} \frac{\hat{S}_{xx}(f, \mathbf{x})}{S_{xx}(f)} df \quad (14)$$

Thus, minimizing the total error E is equivalent to the minimization of the integrated ratio of the signal spectrum $\hat{S}_{xx}(f, \mathbf{x})$ by its approximation $S_{xx}(f)$. The way the spectrum $\hat{S}_{xx}(f, \mathbf{x})$ is being approximated by $S_{xx}(f)$ is largely reflected in the relation between the corresponding autocorrelation functions. Knowing that $r(k) = \hat{r}(k)$ [1] for $k = 1, \ldots, K$ and that the autocorrelation of $x(n)$ is the fourier transform of its spectrum:

$$\hat{r}(k) = \int_{-1/2}^{1/2} \hat{S}_{xx}(f, \mathbf{x}) e^{j2\pi fk} df \tag{15}$$

and $r(k)$ is the autocorrelation of the impulse response of (4) and also the fourier transform of $S_{xx}(f)$, it follows that increasing the value of the order of the model K increases the range over $\hat{r}(k)$ and $r(k)$ are equal resulting in a better fit of $S_{xx}(f)$ to $\hat{S}_{xx}(f, \mathbf{x})$. Hence, for $K \to \infty$ the two spectra become identical:

$$S_{xx}(f) = \hat{S}_{xx}(f, \mathbf{x}) \quad \text{as} \quad K \to \infty \tag{16}$$

3 Linear Prediction Based on the Least Square Error

The most used error minimization criterion is the method of least squares ($p = 2$ in (6)), this method corresponds to the maximum likelihood approach when the error signal (or, the excitation of the filter in (4)) is considered to be a set of i.i.d. Gaussian variables:

$$\mathbf{e} \sim N(0, \mathbf{C}_e) \tag{17}$$

where $\mathbf{C}_e = \sigma^2 \mathbf{I}$ is a identity matrix multilpied by a costant that corresponds to the variance of the error. One of the reasons for the Gaussian assumption lies in the maximum entropy principle which states that for known values of the first and second moments of a random process, the specific joint probability density which has the largest entropy is the Gaussian probability density. From the definition, the log-pdf will be:

$$\ln p(\mathbf{e}) = -\frac{N}{2} \ln 2\pi - \frac{1}{2} |\mathbf{C}_e| - \frac{1}{2} \mathbf{e}^T \mathbf{C}_e^{-1} \mathbf{e} \tag{18}$$

If we solve (18) by maximazing $\ln p(\mathbf{e})$, considering that $\mathbf{e} = \mathbf{x} - \mathbf{X}\mathbf{a}$ we obtain:

$$\mathbf{a}_{ML} = \arg\min_{\mathbf{a}} \{[\,\mathbf{x} - \mathbf{X}\mathbf{a}\,]^T \mathbf{C}_e^{-1} [\,\mathbf{x} - \mathbf{X}\mathbf{a}]\} \tag{19}$$

that has a closed-form unique solution:

$$\mathbf{a}_{ML} = \left(\mathbf{X}^T \mathbf{C}_e^{-1} \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{C}_e^{-1} \mathbf{x} \tag{20}$$

This becomes, considerng $\mathbf{C}_e = \sigma^2 \mathbf{I}$:

$$\mathbf{a}_{ML} = \left(\mathbf{X}^T \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{x} \tag{21}$$

We would like to calculate the probability density function (pdf) as a function of the power spectral density (PSD). Knowing that filtering linearly a white Gaussian process outputs a signal that is still Gaussian process but not (or not necessarly) white, we can model the signal pdf as:

$$\mathbf{x} \sim N(0, \mathbf{C}_{xx}) \tag{22}$$

with $\mathbf{C}_{xx}$ that is no more a diagonal matrix (variables not uncorrelated and not indipendent). The log-pdf would be:

$$\ln p(\mathbf{x}) = -\frac{N}{2} \ln 2\pi - \frac{1}{2} |\mathbf{C}_{xx}| - \frac{1}{2} \mathbf{x}^T \mathbf{C}_{xx}^{-1} \mathbf{x} \tag{23}$$

and each term can be made dependent from the PSD thanks to the asymptotic relations (for $N \to \infty$) [6]:

$$|\mathbf{C}_{xx}| = \prod_{k=1}^{N} \lambda_k(\mathbf{C}_{xx}) \simeq \prod_{k=0}^{N-1} S_{xx}\left(\frac{2\pi}{N} k\right) \tag{24}$$

and:

$$\mathbf{C}_{xx}^{-1} = \sum_{k=1}^{N} \frac{1}{\lambda_k(\mathbf{C}_{xx})} \mathbf{q}_k \mathbf{q}_k^H \simeq \sum_{k=0}^{N-1} \frac{1}{S_{xx}(\frac{2\pi}{N} k)} \mathbf{v}_k \mathbf{v}_k^H \tag{25}$$

with $\mathbf{v}_k$ being a sinusoid that makes k cycles in N samples:

$$\mathbf{v}_k = \frac{1}{\sqrt{N}} \left[1, \exp\left(j \frac{2\pi k}{N}\right), \ldots, \exp\left(j \frac{2\pi (N-1)}{N}\right)\right]^T \tag{26}$$

Substituting the relations into (23) we obtain:

$$\ln p(\mathbf{x}) = -\frac{N}{2}\ln 2\pi - \frac{1}{2}\sum_{k=0}^{N-1}\left(\ln S_{xx}\left(\frac{2\pi}{N}k\right) + \frac{\left|\mathbf{v}_k^H\mathbf{x}\right|^2}{S_{xx}(\frac{2\pi}{N}k)}\right) \tag{27}$$

Noting that:

$$\mathbf{v}_k^H\mathbf{x} = DFT_N(\mathbf{x})|_{\omega_k} = \frac{1}{\sqrt{N}}X(\omega_k) \tag{28}$$

and

$$\left|\mathbf{v}_k^H\mathbf{x}\right|^2 = \frac{1}{N}\left|X(\omega_k)\right|^2 = \hat{S}_x(\omega_k, \mathbf{x}) \tag{29}$$

represent a trasformation of the observations (the DFT) that corresponds with the periodogram, we can rewrite (27) as:

$$\ln p(\mathbf{x}) = -\frac{N}{2}\ln 2\pi - \frac{1}{2}\sum_{k=0}^{N-1}\left(\ln S_{xx}\left(\frac{2\pi}{N}k\right) + \frac{\hat{S}_{xx}(\omega_k, \mathbf{x})}{S_{xx}(\frac{2\pi}{N}k)}\right) \tag{30}$$

In this form, it can result hard to understand, so multiplicating and dividing the second term for the band unit $1/N$ we have:

$$\ln p(\mathbf{x}) = -\frac{N}{2}\ln 2\pi - \frac{N}{2}\sum_{k=0}^{N-1}\frac{1}{N}\left(\ln S_{xx}\left(\frac{2\pi}{N}k\right) + \frac{\hat{S}_{xx}(\omega_k, \mathbf{x})}{S_{xx}(\frac{2\pi}{N}k)}\right) \tag{31}$$

that for $N \to \infty$ becomes:

$$\ln p(\mathbf{x}) \simeq -\frac{N}{2}\ln 2\pi - \frac{N}{2}\int_{-1/2}^{1/2}\ln S_{xx}\left(f\right) + \frac{\hat{S}_{xx}(f, \mathbf{x})}{S_{xx}(f)}df \tag{32}$$

This is the asymptotic relation that holds until N is sufficiently large (ideally $N \to \infty$).

In the case of auto-regressive (AR) parametric spectral estimation, the PSD depends on a set of deterministic parameters $\boldsymbol{\theta}$ that are the recursive component of the filter $\mathbf{a} = [a(1), \ldots, a(K)]^T$ and the scaling factor σ^2:

$$S_{xx}(f|\boldsymbol{\theta}) = \frac{\sigma^2}{|A(f|\boldsymbol{a})|^2} \qquad \boldsymbol{\theta} = [\mathbf{a}, \sigma^2]^T \in R^{K+1} \tag{33}$$

the log-likelihood for the ML estimation becomes substituting (33) in (32):

$$\begin{aligned}\ln p(\mathbf{x}|\boldsymbol{\theta}) \simeq & -\frac{N}{2}\ln 2\pi - \frac{N}{2}\ln(\sigma^2) + \frac{N}{2}\int_{-1/2}^{1/2} \ln |A(f|\mathbf{a})|^2 df - \\ & \frac{N}{2\sigma^2}\int_{-1/2}^{1/2} |A(f|\mathbf{a})|^2 \hat{S}_{xx}(f,\mathbf{x}) df\end{aligned} \tag{34}$$

For monic polynomials (with $a(0) = 1$) we have $\int_{-1/2}^{1/2} \ln |A(f|\mathbf{a})|^2 df = 0$, (34) therefore becomes:

$$\ln p(\mathbf{x}|\boldsymbol{\theta}) \simeq -\frac{N}{2}\ln 2\pi - \frac{N}{2}\ln(\sigma^2) - \frac{N}{2\sigma^2}\int_{-1/2}^{1/2} |A(f|\mathbf{a})|^2 \hat{S}_{xx}(f,\mathbf{x}) df \tag{35}$$

Putting the first gradient to zero in respect to σ^2:

$$\frac{\delta \ln p(\mathbf{x}|\boldsymbol{\theta})}{\delta \sigma^2} = 0 \rightarrow -\frac{N}{2\sigma^2} + \frac{N}{2\sigma^4}\int_{-1/2}^{1/2} |A(f|\mathbf{a})|^2 \hat{S}_{xx}(f,\mathbf{x}) df \tag{36}$$

and therefore:

$$\hat{\sigma}^2 = \hat{\sigma}^2(\mathbf{a}) = \int_{-1/2}^{1/2} |A(f|\mathbf{a})|^2 \hat{S}_{xx}(f,\mathbf{x}) df \tag{37}$$

we have the the power depends on the recursive part of the filter $\mathbf{a}$. Sobstituting into the log-likelihood function (35):

$$\ln p(\mathbf{x}|\mathbf{a}, \hat{\sigma}^2(\mathbf{a})) \simeq -\frac{N}{2}(1 + \ln 2\pi) - \frac{N}{2}\ln \sigma^2(\mathbf{a}) \tag{38}$$

this means that maximizing $\ln p(\mathbf{x}|\mathbf{a}, \hat{\sigma}^2(\mathbf{a}))$ corresponds to minimizing $\sigma^2(\mathbf{a})$. It is now clear that the Gaussian maximum-likelihood estimation of the parameters that generated the signal $x(n)$ corresponds to minimizing the integrated ratio of the signal spectrum $\hat{S}_{xx}(f,\mathbf{x})$ to its approximation $S_{xx}(f|\boldsymbol{\theta})$ (33). Proceeding with the calculations of the gradients (always assuming $\mathbf{a} \in R^K$):

$$\frac{\delta \sigma^2(\mathbf{a})}{\delta a(k)} = \frac{\delta}{\delta a(k)}\int_{-1/2}^{1/2} A(f|\mathbf{a}) A^*(-f|\mathbf{a}) \hat{S}_{xx}(f,\mathbf{x}) df \tag{39}$$

applying the properties of the derivative in the product of functions and developing the calculations, knowing that $\hat{S}_{xx}$ is real, we obtain that solving (39) is equivalent to solve:

$$\int_{-1/2}^{1/2} A(f|\mathbf{a})\hat{S}_{xx}(f,\mathbf{x})e^{j2\pi fn}df = 0 \tag{40}$$

developing the calculations:

$$\int_{-1/2}^{1/2} \hat{S}_{xx}(f,\mathbf{x})e^{j2\pi fn}df + \sum_{k=1}^{K} a(k)\int_{-1/2}^{1/2} \hat{S}_{xx}(f,\mathbf{x})e^{j2\pi f(n-k)}df = 0 \tag{41}$$

the periodogram $\hat{S}_{xx}(f,\mathbf{x})$ is the Fourier transform of the sampled autocorrelation function (biased), therefore through (41) we will obtain the Yule-Walker equations written now respect to the autocorrelation function [7]:

$$\hat{r}(k) + \sum_{k=1}^{K} a(k)\hat{r}(n-k) = 0 \quad \text{for} \quad k = 1,\ldots,K \tag{42}$$

We can also see that:

$$\hat{\sigma}^2(\hat{\mathbf{a}}) = \hat{r}(0) + \sum_{k=1}^{K} a(k)\hat{r}(k) \tag{43}$$

4 Linear Prediction Based on the Least Absolute Error

Assuming that the process is Gaussian is based upon the fact that the Gaussian assumption is often sufficient for tractable mathematics, but also is based upon a very liberal view of the central limit theorem, which may be loosely stated: "almost any random process put into almost any linear system will come out almost Gaussian."

The linear prediction method based on the least absolute error has only recently started to be used as it does not have a closed form solution as the least square method (20) that can be solved easily. Neverthless, it seems to be really interesting when dealing with the representation of voiced speech were the excitation can be better represented by a sparse impulsive signal.

The method introduced in this section corresponds to the assumption that the error signal has a Laplacian probability density function, then the speech signal analyzed will still have a Laplacian distribuition [8]:

$$\mathbf{x} \sim L(0, \mathbf{C}_{xx}) \tag{44}$$

The Laplacian pdf, differently from the Gaussian, does not have a simple closed form that includes the covariance matrix of the analyzed signal. Althought many studies have been made in order to fill this gap. According to [9]:

$$p(\mathbf{x}) = \frac{2}{(2\pi)^{N/2}|\mathbf{C}_{xx}|^{1/2}} \left(\frac{\mathbf{x}^T \mathbf{C}_{xx}^{-1} \mathbf{x}}{2} \right)^{-N/4+1/2} K_{N/2-1}\left(\sqrt{2\mathbf{x}^T \mathbf{C}_{xx}^{-1} \mathbf{x}} \right) \tag{45}$$

where $K_{N/2-1}\left(\sqrt{2\mathbf{x}^T \mathbf{C}_{xx}^{-1} \mathbf{x}} \right)$ denotes the modified bessel function of the second kind and order $N/2-1$ evaluated at $\sqrt{2\mathbf{x}^T \mathbf{C}_{xx}^{-1} \mathbf{x}}$. Noting that the Bessel function, for $\sqrt{2\mathbf{x}^T \mathbf{C}_{xx}^{-1} \mathbf{x}}$ sufficiently large, behaves like:

$$K_{N/2-1}\left(\sqrt{2\mathbf{x}^T \mathbf{C}_{xx}^{-1} \mathbf{x}} \right) \simeq \sqrt{\frac{\pi}{2\sqrt{2\mathbf{x}^T \mathbf{C}_{xx}^{-1} \mathbf{x}}}} \exp\left(-\sqrt{2\mathbf{x}^T \mathbf{C}_{xx}^{-1} \mathbf{x}} \right) \tag{46}$$

we can rewrite the pdf as:

$$p(\mathbf{x}) \simeq \frac{2}{(2\pi)^{N/2}|\mathbf{C}_{xx}|^{1/2}} \left(\frac{\mathbf{x}^T \mathbf{C}_{xx}^{-1} \mathbf{x}}{2} \right)^{-N/4+1/2} \sqrt{\frac{\pi}{2\sqrt{2\mathbf{x}^T \mathbf{C}_{xx}^{-1} \mathbf{x}}}} \exp\left(-\sqrt{2\mathbf{x}^T \mathbf{C}_{xx}^{-1} \mathbf{x}} \right) \tag{47}$$

to make it more clear we can rewrite ans set $G = 1/\sqrt{2\pi^{N-1}}$:

$$p(\mathbf{x}) \simeq G \frac{\left(2\mathbf{x}^T \mathbf{C}_{xx}^{-1} \mathbf{x} \right)^{-N/4+1/4} \exp\left(-\sqrt{2\mathbf{x}^T \mathbf{C}_{xx}^{-1} \mathbf{x}} \right)}{|\mathbf{C}_{xx}|^{1/2}} \tag{48}$$

The log-likelihood function becomes:

$$\ln p(\mathbf{x}) = \ln G - \frac{N-1}{4} \ln\left(2\mathbf{x}^T \mathbf{C}_{xx}^{-1} \mathbf{x} \right) - \sqrt{2\mathbf{x}^T \mathbf{C}_{xx}^{-1} \mathbf{x}} - \frac{1}{2} \ln |\mathbf{C}_{xx}| \tag{49}$$

using the asymptotic relations in (24) and (25) and multiplying and dividing by the band unit$1/N$ we can rewrite as:

$$\ln p(\mathbf{x}) \simeq \ln G - \frac{N-1}{4} \ln \left(2N \int_{-1/2}^{1/2} \frac{\hat{S}_{xx}(f,\mathbf{x})}{S_{xx}(f)} df \right) - \sqrt{2N \int_{-1/2}^{1/2} \frac{\hat{S}_{xx}(f,\mathbf{x})}{S_{xx}(f)} df} - \frac{N}{2} \int_{-1/2}^{1/2} \ln |S_{xx}(f)| df \quad (50)$$

Substituting the relations of (33) and remembering that for monic polynomials $\int_{-1/2}^{1/2} \ln |A(f|\mathbf{a})|^2 df = 0$, we obtain:

$$\ln p(\mathbf{x}|\boldsymbol{\theta}) \simeq \ln G - \frac{N-1}{4} \ln \left(\frac{2N}{\sigma^2} \int_{-1/2}^{1/2} |A(f|\mathbf{a})|^2 \hat{S}_{xx}(f,\mathbf{x}) df \right) - \sqrt{\frac{2N}{\sigma^2} \int_{-1/2}^{1/2} |A(f|\mathbf{a})|^2 \hat{S}_{xx}(f,\mathbf{x}) df} - \frac{N}{2} \ln \left(\sigma^2\right) \quad (51)$$

evaluating the first derivative of (51) with respect to σ^2 brings us to the following result:

$$\sigma^2 = \frac{2N}{N-1} \left(2N \int_{-1/2}^{1/2} |A(f|\mathbf{a})|^2 \hat{S}_{xx}(f,\mathbf{x}) df \right) \quad (52)$$

which means that spectral flatness measure for $K \to \infty$ is identical for both the 2-norm and 1-norm error minimization criterion.

References

[1] J. Makhoul, "Linear Prediction: A Tutorial Review", *Proc. IEEE*, vol. 63(4), pp. 561–580, Apr. 1975.

[2] M. Reed and B. Simon, *Methods of Modern Mathematical Physics II: Fourier Analysis, Self-adjointness*, Academic Press, 1975.

[3] P. Stoica and R. Moses, *Spectral Analysis of Signals*, Pearson Prentice Hall, 2005.

[4] J. H. L. Hansen, J. G. Proakis, and J. R. Deller, Jr., *Discrete-Time Processing of Speech Signals*, Prentice-Hall, 1987.

[5] J.-R. Ohm, *Multimedia Communication Technology: Representation, Transmission, and Identification of Multimedia Signals*, Springer-Verlag, 2004.

[6] U. Spagnolini, *Statistical Signal Processing for Telecommunications*, Politecnico Di Milano Press, 2004. (In Italian)

[7] J.D. Markel and A.H. Gray, *Linear Prediction of Speech*, Springer-Verlag, 1976.

[8] T. Eltoft, T. Kim, T. Lee, " On the Multivariate Laplace Distribution", *IEEE Signal Processing Letters*, Vol. 13, no. 5, pp. 300–303, 2006.

[9] S. Kotz, N. Balakrishnan, N. L. Johnson, *Continuous Multivariate Distributions, Volume 1, Models and Applications, 2nd edition*, Wiley, 2000