Positive Definite and Unimodal Gauss-Hermite Expansion of Probability Density Functions

Wolfgang Freude

Summary - Zusammenfassung

The approximation of Gaussian-like probability density functions (p.d.f.) by Gauss-Hermite series of Gram-Charlier and Edgeworth type, or by a Cornish-Fisher expansion frequently violates two constraints: the curves should be positive definite, and unimodal. A new standardization allows the choice of the basic Gauss-Hermite system such that these difficulties are avoided. If the p.d.f. is given, the expansion coefficients of the approximating series may be computed by a relative least-squares error fit; usually, the two p.d.f. constraints are not violated. If the moments are known, the expansion coefficients are computed by a matrix inversion, where an iterative procedure adjusts the standardizing parameters according to the two p.d.f. constraints. The method is compared to the approximation by a Pearson function.

1. Introduction

When simulating coherent optical receivers numerically, it is often the case that in calculating the bit error probability \( BER = \int p_e(x) \, dx \) either the probability density function (p.d.f.) \( p_e(x) \) of the received signal \( x \) is known as a numerical table, or the moments of the signal may be estimated. In both cases it is useful to reconstruct an analytical formula for the p.d.f. with the constraints that it must be non-negative and with only one maximum, i.e., positive definite and unimodal.
In the past 40 years there have been numerous attempts to approximate p.d.f.'s in terms of the moments of the process. Special attention was paid to Gaussian-like p.d.f.'s, which where either represented by Gauss-Hermite series of the Gram-Charlier and Edgeworth type, or, for the Cornish-Fisher expansion, by approximating the standardized argument $z$ with normal-like p.d.f.'s, which where either represented by Hermite series of the Gram-Charlier and Edgeworth type, or, for the Cornish-Fisher expansion, by non-normally distributed argument $(Gauss-)Hermite function system was uniquely diepoint approximation [5], which in effect differs in the standardization of the random variable.

For all of these approximations the fundamental (Gauss-)Hermite function system was uniquely defined by the very moments of the process, so it happened not infrequently that the two basic constraints of a p.d.f. were violated. As no degree of freedom was left, one had to take or to leave whatever the method presented for the chosen expansion type and the given process. Especially for long-tailed p.d.f., or for the calculation of very low error probabilities, well-behaved representations of p.d.f. are requested.

The importance of the problem has been recognized long since in connection with the expansions of Gram-Charlier and Edgeworth. An early remark is found in [6]. First in [7], later on in [8], [9] moment regions were calculated (for the first four moments only) in which the constraints above were violated; also, there are notes in [10], [11].

A different approach is to approximate a p.d.f. by the family of the p.d.f.'s associated with the Pearson distributions [12, Section 6]. e.g., the beta density function. The parameters of these p.d.f. may be calculated from the moments. For a specified region of moments the functions are positive definite and unimodal. The relations are published for a set of four moments.

No further attempts to tackle the problem are known to the author. The reason for this deficiency might be that the statistics community lacked interest in esoteric Gaussian-like p.d.f. tail regions where for optical coherent receivers [13] symbol transmission error probabilities as low as $10^{-9}$ are to be calculated.

The present paper discusses a Gauss-Hermite approximation, and proposes a different standardization of the random variable with two or three process-independent parameters, which may be varied in an iterative manner to find a positive definite and unimodal representation out of the infinite manifold of density functions with a given finite number of identical moments.

For analytically or numerically given p.d.f. a relative least-squares fit (r.l.s.f.) is applied to compute the expansion coefficients of a Gauss-Hermite series, a method which allows a simple weighting of the available data, and needs not the problematic direct calculation of moments [10], [11]. If the moments of the process can be estimated, a method is proposed to reconstruct the p.d.f. of the process as a Gauss-Hermite series by a simple matrix inversion. The techniques described are demonstrated for non-Gaussian p.d.f. and compared to existing methods.

2. Relative Least-Squares Error Fit

A statistical process with p.d.f. $p_\text{x}(x)$ has an outcome which is described by the random variable $x$. If $\mu_1$ and $\sigma$ are mean and standard deviation of the process, the corresponding standardized random variable $X'$ is usually defined by

$$X' = \frac{x - \mu_1}{\sigma}.$$  \hspace{1cm} (1)

Here, a different standardization is chosen as was already done in characterizing single-mode optical waveguides [14], [15], where the (not yet determined) parameters $x_0$ and $\chi$ resemble $\mu_1$ and $\sigma^{-1}$, respectively,

$$X = \chi (x - x_0).$$  \hspace{1cm} (2)

Consequently, a series expansion of $p_\text{x}(x)$ in terms of the Gauss-Hermite function system $R_\chi$ with the parameters $x_0$, $x$, and the Hermite functions $H_n$ [3], may be written as

$$p_\text{x}(x) \approx \sum_{n=1}^{N} c_n R_{\chi-1}(x) \sum_{-\infty}^{\infty} \sum_{n=1}^{N} c_n R_{\chi-1}(x) \, dx,$$

$$R_{\chi}(x) = \exp \left(-\chi^2/2\right) H_n(x).$$  \hspace{1cm} (3)

Aside from the new standardization, eq. (3) may equally well represent the Gram-Charlier, Edgeworth, or saddlepoint expansion, if the coefficients $c_n$ are computed properly. It is interesting to note that the structure of the corresponding characteristic function $C_{\chi}(\xi)$ with the standardized variable $\xi = x - x_0$ is invariant with respect to a Fourier transformation,

$$C_{\chi}(\xi) = \exp(-\pi \chi^2) = \int_{-\infty}^{\infty} p_\text{x}(x) \exp(-\pi \chi^2) \, dx$$

$$= \left(\sqrt{\frac{2}{\pi}}\chi\right) \exp\left(-i \chi \sum_{n=1}^{N} (-i)^{n-1} c_n R_{\chi-1}(\Xi)\right),$$

$$\Xi = 2\pi \chi / \chi.$$  \hspace{1cm} (4)

If the p.d.f. $p_\text{x}(x)$ is given analytically or numerically, the basic parameters $x_0$, $x$, and $\chi$ may be found either by computing the first ordinary and the second central moment, or, as it is done here, by least-squares fitting a Gaussian $R_\chi(x)$ to the values $p_\text{x}(x)$ with $x = x_0$, $X = X_1$ inside a reduced range $K_1 \leq k \leq K_2$ of the main lobe,

$$\sum_{k=K_1}^{K_2} \left[ p_\text{x}(x_k) - A_0 R_\chi(X_k) \right]^2 = \min,$$

$$\ln \{ A_0 R_\chi(x - x_0) \} = A_1 + A_2 x_0 + A_3 x_0^2.$$  \hspace{1cm} (5)

With the abbreviations $\beta$, $\delta$ for column vectors with the elements $B_i$, $A_i \neq 0$, and $S$ for Hermitian matrices with the matrix elements $S_{jk}$, one has to solve the
system of equations for $A$,

$$ B = S A, \quad A = S^{-1} B, $$

$$ B_i = \sum_{k=K}^{N} x_i^{k-1} \ln p_\alpha(x_k), \quad (6) $$

$$ S_{ll} = \sum_{k=K}^{N} x_i^{4+k-2} = S_{ll}. $$

For the parameters $A_\alpha, x_0$ and $\chi$ one finds

$$ A_0 = \exp\{A_1 - A_2/(4A_3)\}, \quad x_0 = -A_2/(2A_3), \quad \chi = (-2A_3)^{1/2}. $$

The expansion coefficients for $p_\alpha(x)$ in the chosen function system are computed by minimizing the sum of the squared relative errors for all values $p_\alpha(x)$ with $x = x_k, X = X_k, 1 \leq k \leq K$,

$$ \sum_{k=K}^{N} W_k^2 [p_\alpha(x_k) - \sum_{n=1}^{N} c_n R_{n-1}(X_k)]^2 = \min, $$

$$ W_k = 1/p_\alpha(x_k). \quad (8) $$

For the coefficient vector $C$ with the elements $c_n$ one has to solve the system of equations

$$ B = SC, \quad C = S^{-1} B, $$

$$ B_i = \sum_{k=1}^{K} W_k^2 p_\alpha(x_k) R_{n-1}[x(x_k-x_0)], \quad (9) $$

$$ S_{ll} = \sum_{k=1}^{K} W_k^2 R_{n-1}[x(x_k-x_0)] R_{n-1}[x(x_k-x_0)] = S_{ll}. $$

The normalization integral in eq. (3) modifies in effect the $c_n$ computed from eq. (9) by a constant factor, which usually deviates only by small amounts from one.

The moments $\mu_m$ of order $m$ about the origin are

$$ \mu_m(x_0, \chi) = \int_{-\infty}^{\infty} x^m p_\alpha(x) \, dx = \sum_{n=1}^{N} d_m(x_0, \chi) c_n, $$

$$ d_m(x_0, \chi) = (\sqrt{2}/\chi) \int_{-\infty}^{\infty} (\sqrt{2} u/\chi + x_0)^m \exp(-u^2) H_{n-1}(\sqrt{2} u) \, du. \quad (11) $$

The integration variable $u$ is chosen to facilitate the application of the Hermite integration technique, a numerically very fast formula of Gaussian type [3, Section 25, Formula 25.4.46, Table 25.10] for $u$ and $w$,

$$ \begin{aligned}
\int_{-\infty}^{\infty} \exp(-u^2) f(u) \, du = & \\
\left\{ \begin{array}{ll}
\sum_{i=1}^{l} w_i [f(u_i) + f(-u_i)] & \text{if even}, \\
\sum_{i=2}^{l} w_i [f(u_i) + f(-u_i)] & \text{if odd}.
\end{array} \right.
\end{aligned} \quad (12) $$

The result is exact, if (as it is the case here) $f(u)$ is a polynomial of maximum degree $n - 1 + m$, and if $l = [(n - 1 + m)/2] + 1$ is valid ($[x]$ stands for the integer part of $x$).

The advantages of the proposed r.l.s.f. method over the Gram-Charlier, Edgeworth and Cornish-Fisher expansions will be demonstrated for the p.d.f.

$$ p_\alpha(x) = \frac{1}{(2\pi 0.5^2)^{1/2}} \cdot \exp[-x^2/(2 \cdot 0.5^2)] \left[ 1 + \frac{2}{\pi} \arctan(5x) \right]. \quad (13) $$

Fig. 1 shows this function with dashed lines. Fig. 1a displays as a heavy line the r.l.s.f. approximation with $N = 3$ coefficients in eq. (3), i.e., for moments up to a maximum order $m = 3$. The equivalent curves for the approximations Gram-Charlier, Edgeworth, and Cornish-Fisher are taken from [16], [17] (both for Fig. 1a and Fig. 1b) and are for $m = 3$ not to be discriminated from the r.l.s.f. graph. As to be expected from the low number of expansion terms, the quality of the approximations is poor.

The various expansions differ much more, if nine-term series $N = 9$ with moments up to the maximum order $m = 9$ are evaluated, Fig. 1b. For the r.l.s.f. curve the relative root mean square (r.r.m.s.) error for the fitting range $-1 \leq x \leq 1.5$ in comparison to the ideal curve eq. (13) amounts to 20%; the deviation is hardly to be seen left of the peak. The three other expansions exhibit large errors, especially negative p.d.f. ranges.
These irregularities increase with increasing maximum order of the moments, as opposed to the r.l.s.f. technique. This is partly due to the fact that in performing the moment integral eq. (10) a sum of Gaussian-like decaying terms has to be evaluated, and that, therefore, computers with finite accuracy evaluate the tail contributions erroneously [14], but partly the cause for this behavior lies also in the fixed Hermite function system, which is defined by the moments themselves.

The r.l.s.f. method guarantees that the p.d.f. tails are as much taken into account as the central parts. No time-consuming and error-prone numerical integration is required, only the solution of a system of equations where fast and accurate standard algorithms are available. Therefore, it is unlikely (but not impossible!) that negative p.d.f. regions occur. As an example, Fig. 2 shows the ideal (eq. (13)) and the r.l.s.f. curves for \( N = 10 \) (as in Fig. 1b) and \( N = 22 \) coefficients in a logarithmic display, where \( p_s(x) \), measured in dB (decibel), is given by 10 \( \log |p_s(x)| \). The approximations are positive definite and unimodal. On the left and right hand side of the fitting range the curve is lower by about 25 dB and 21 dB, respectively, as compared to eq. (14).

### 3. Iteration Procedure to Satisfy p.d.f. Constraints

If with \( x_0, \chi \) a certain function system is chosen, the elements of matrix \( D \) are determined independently of the elements of \( b \), eq. (10), which constitute the moment vector \( \mu' \). Consequently, it makes sense to write eq. (10) in matrix form,

\[
M' = D C, \quad C = D^{-1} M',
\]

so that the elements \( c_i \) of the coefficient vector \( C \) in the p.d.f. expansion eq. (1) are to be computed by a simple matrix inversion. \( D \) must be square for that purpose, i.e., the maximum number of moments must equal the number \( N \) of coefficients.

In the conventional case of the Gram-Charlier or Edgeworth series, where \( x_0 = 0, \, \chi = 1/\sigma = (\mu'_0 - \mu'_1^2)^{1/2} \), are valid by definition. \( C \) is completely fixed by \( M' \), and if the resulting p.d.f. eq. (1) derived from, say, the first ten moments is not positive definite by chance (which is very unlikely for an infinite range of the random variable \( x \)), the expansion is physically meaningless.

With the present method, the system parameters \( x_0 \) and \( \chi \) may be adjusted starting from their initial values above, so that the p.d.f. obeys the constraints. If for a given \( M' \) this cannot be achieved, the number of relevant moments (and coefficients) may be reduced by one, and the procedure repeated.

For adjusting the parameters \( x_0 \) and \( \chi \), an objective function \( g(x_0, \chi) \) is to be minimized. Most important, the positive-definiteness of \( p_s(x) \) has to be tested. This may be done efficiently by computing the real zeros of the (for \( N \leq 10 \) explicitly programmed) polynomial

\[
S = \sum_{n=0}^{N} c_n H_{n-1}(X) \quad \text{in eq. (1).}
\]

The ranges \( X_{\text{exp}} \) for a negative p.d.f. are checked, and \( g \) is defined as the maximum of the dominant exponential for the region

---

**Fig. 2.** Non-Gaussian p.d.f. eq. (13) for \( N = 10 \) (r.m.s. error 41%, see Fig. 1b) and \( N = 22 \) (r.m.s. error 50%) coefficients. The fitting range is \(-4.6 \leq x \leq 5.62 \).

**Fig. 3.** Slowly decaying non-Gaussian p.d.f. eq. (14) for \( N = 48 \). The r.m.s. error is 31% in the fitting range \(-9.58 \leq x \leq 10.74 \).
4. Double-Gauss-Hermite Function System

So far, the basic function system has a Gaussian decay factor which is identical for both the $+X$ and the $-X$ directions in eq. (3). As a consequence, only asymptotically symmetric p.d.f. will be approximated with low errors. If this symmetry is strongly disturbed as for phase noise [19], [20, Section 3.3.4] $x=\cos z$ with a Gaussian density of $z$, the convergence properties of the iteration procedure become worse.

Therefore, $R_n(X)$ eq. (3) was generalized using different Gaussian functions for different signs of $X$, namely

$$X_1 = x_1(x-x_0), \quad X_2 = x_2(x-x_0),$$

$$R_n(X) = \begin{cases} R_n(X_1) & \text{for } x \geq x_0 \\ R_n(X_2) & \text{for } x \leq x_0 \end{cases}$$

Now, the basic function system is characterized by the three parameters $x_0$, $x_1$, and $x_2$. As before, the integrals in eq. (11) (this time of the type $\int_0^\infty u^n \exp(-u^2) \, du$) may be solved analytically, so $d_{xy}$ can be calculated efficiently by summing up precomputed factors which are weighted by powers of $x_0$, $x_1$, and $x_2$. As a consequence, the computational effort does not increase compared with the single-Gauss-Hermite expansion.

The techniques described above were tested with output data $x$ from the simulation of an optical DPSK heterodyne receiver with dominant phase noise (DPSK: differential phase shift keying). For computing the BER the p.d.f. $p_n(x)$ of the time sequence $x = x_1, x_2, \ldots, x_t$ of output data in front of the decision circuit has to be reconstructed from an estimate of the associated moments. Details of the simulation and of this estimation process are of no importance for the present discussion. For a rough estimate, the p.d.f. was reconstructed directly by computing the span of the sorted sequence $x_{t-1} < x_t$ for a fixed subinterval $i + J$, i.e., $p_n[(1/2)(x_{i+J} + x_i)] \sim J/(x_{i+J} - x_i)$ (assuming that the probability of identical values $x_i$ in the subinterval $J$ is negligible). This technique [21, Section 13.3] is far superior to the usual method, where the frequency of data for a fixed interval of $x$ is counted. Next, the Pearson method for a beta-p.d.f. using four (and only four) moments was employed with $p_n(x) = (1 + x/\alpha_2)^{\beta} (1 - x/\alpha_2)^\gamma$; the parameters $\alpha_1$, $\alpha_2$, $\beta$, and $\gamma$ are uniquely determined by the moments of the process. Finally, the Gauss-Hermite and the
double-Gauss-Hermite approximations were calculated; the maximum useful number of moments to reach convergence in the sense of the minimizing criteria eq. (16) for varying $x_0$, $\chi$ and $x_1$, $x_2$ was $N = m = 5$ and $N = m = 9$. The results are shown in Fig. 5. Obviously, the double-Gauss-Hermite technique is closest (in mode and curve details) to the direct reconstruction. As is to be expected, the simple Gauss-Hermite series reacts "unwillingly" in approximating an asymptotically asymmetric p.d.f. The linear display of Fig. 5 is not very informative if low BER are to be estimated.

In a different simulation run only additive Gaussian noise was assumed to disturb the DPSK receiver. The model parameters were chosen such that a simple theory [22] yielded $BER = \int_{-\infty}^{\infty} P_e(x) \, dx \approx 10^{-5}$. The semi-logarithmic display Fig. 6 compares the p.d.f.'s from the Pearson and the double-Gauss-Hermite reconstruction; both curves reproduce per definitionem identical moments up to the order four. A strong discrepancy is to be seen. First, the limited definition range of the Pearson function with $a_1 = -0.05, a_2 = 3.1$ shows up for the lower boundary. Secondly, the Pearson function is asymptotically strongly asymmetric in contrast to the physical situation. The evaluation of the BER integral for the double-Gauss-Hermite curve yields $BER \approx 0.9 \cdot 10^{-5}$ in close agreement with the theoretical value. However, for $x \leq 0$ the order of magnitude of the Pearson function is seen to be smaller by a factor of 1000, and extended over a narrower range, so with $BER \approx 4 \cdot 10^{-11}$ the error probability is grossly underestimated.

5. Further Improvements

In reconstructing a p.d.f. from moments it was assumed that the true moments were known. However, usually the moments are only estimated from a data sample, and are therefore inaccurate. If moment error bounds can be computed, the iteration process minimizing the objective function eq. (16) can be improved by varying not only the parameters of the function system, but also the moments themselves inside their error bounds. This should add enough degrees of freedom to fulfill the p.d.f. constraints. Work is in progress to develop error bounds for various moment estimation techniques, and to add the moment variation to the iteration algorithm.

6. Conclusion

Probability density functions may be expanded in Gauss-Hermite series by a relative least-squares error fit, thereby avoiding the problems of negative regions and non-unimodality usually associated with the standard series expansions derived from moments. The moments of a statistically independent random variable may be efficiently computed, then, and transformed by linear filters in explicit form.

To reconstruct a p.d.f. from given moments the proposed iteration algorithm is well suited. Its main numerical advantages are the formulation for a fast numerical integration, the determination of the expansion coefficients by matrix inversion, and the direct search for zeros of the basic polynomials.

Acknowledgements

Thanks are due to Ir. F. Libbrecht, Laboratorium voor Electromagnetisme en Acustica, Rijksuniversiteit Gent, Belgium, for referring to the Pearson distributions. Further, the work of A. Schönhleder is acknowledged who contributed to Figs. 5 and 6.

References


