9. Gaussian random sequences

We start with recalling of the definition for Gaussian random variable which is completely defined by a density $f(x)$ of its distribution function $F(x)$:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}, \tag{9.1}$$

where

$$m = \int_{\mathbb{R}} x f(x) dx \quad \text{and} \quad \sigma^2 = \int_{\mathbb{R}} (x-m)^2 f(x) dx$$

are referred as the expectation and variance. So, if ξ is Gaussian random variable with the expectation m and variance σ^2 its density function is given by (9.1). If $\sigma^2 = 0$, then, obviously, $\xi = m$ and a sense of (9.1) is lost. Parallel to the density $f(x)$ let us introduce the characteristic function of ξ: (henceforth $\imath = \sqrt{-1}$, $\lambda \in \mathbb{R}$)

$$\Phi(\lambda) := \boldsymbol{E} e^{\imath\lambda\xi} = \int_{\mathbb{R}} e^{\imath\lambda x} f(x) dx. \tag{9.2}$$

It is well known, and readily verified, that

$$\Phi(\lambda) = e^{\imath m\lambda - \lambda^2\sigma^2/2}. \tag{9.3}$$

Notice, this formula is well defined whatever σ^2 is: positive or zero. By the way, the latter comment emphasize that the characteristic function not always guarantees the existence of the distribution density function. It is well known from the Fourier analysis that under $\int_{\mathbb{R}} |\Phi(\lambda)| d\lambda < \infty$, the backward Fourier transform is well defined and so the density is well defined. If $\sigma^2 > 0$, the the above-mentioned holds true and density exists.

Naturally, the next object is Gaussian vector. Taking into the consideration the fact that characteristic function is not sensitive to "degeneration of variance", we give a definition for a Gaussian vector in term of the characteristic function. Henceforth, all vectors are column-vectors and "^{T}" is transposition symbol.

The random vector $\xi = (\xi_1, ..., \xi_n)$ with $m = \boldsymbol{E}\xi$ is the expectation vector and $\Gamma = \boldsymbol{E}(\xi - m)(\xi - m)^T$ is the covariance matrix (m_i and Γ_{ij}, $i, j = 1, \dots, m$ are entries of m and Γ respectively) is Gaussian, if for any vector λ with entries $\lambda_1, ..., \lambda_n$ the characteristic function $\Phi(\lambda) = \boldsymbol{E} \exp\big(\langle\lambda, \xi\rangle\big)$ (here $\langle a, b\rangle$ is the inner product for vectors a, b with entries a_i and b_i respectively, i.e. $\langle a, b\rangle = \sum_i a_i b_i$) is defined as follows:

$$\Phi(\lambda) = \exp\Big(\imath\langle\lambda, m\rangle - \frac{1}{2}\langle\lambda, \Gamma\lambda\rangle\Big). \tag{9.4}$$

If Γ is non-singular matrix, the distribution function of ξ possesses the density (here x is the vector with entries $x_1, \dots, x_n$)

$$f(x) = \frac{1}{(2\pi \mathsf{det})^{n/2}\Gamma^{1/2}} \exp\Big(-\frac{1}{2}\big\langle x-m, \Gamma^{-1}(x-m)\big\rangle\Big). \tag{9.5}$$

A random sequence $\xi = (\xi_1, \xi_2, \dots, \xi_n, \dots,)$ with well defined $m_i = \boldsymbol{E}\xi_i$ and $\Gamma_{ij} = \boldsymbol{E}(\xi_i - m_i)(\xi_j - m_j)$ for any i, j is called Gaussian, if any finite subvector of ξ is Gaussian.

We discuss now a few important properties of Gaussian sequences.

1. Linear transformation preserves Gaussian property.

Proof. Let ξ by a finite Gaussian vector and A is a matrix of an appropriate size entries of which are numbers. Set $\eta = A\xi$ and show that η is Gaussian vector as well.

Write

$$\big\langle \lambda, \eta\big\rangle = \big\langle \lambda, A\xi\big\rangle = \big\langle A^T\lambda, \xi\big\rangle = \big\langle \lambda'\xi\big\rangle,$$

that is $\Phi_\eta(\lambda) = \Phi_\xi(\lambda')$ (in words: the characteristic function of η at point λ coincides with the characteristic function for ξ at the point λ'). □

2. Orthogonal Gaussian vectors are independent

Proof. Recall that any vectors ξ and η are independent, if their joint characteristic function splits into the product of the marginal characteristic functions: for any λ and μ

$$\boldsymbol{E}\exp\{\imath\langle\lambda,\xi\rangle + \langle\mu,\eta\rangle\} = \boldsymbol{E}\exp\{\imath\langle\lambda,\xi\rangle\}\boldsymbol{E}\exp\{\imath\langle\mu,\eta\rangle\}. \tag{9.6}$$

Below we verify (9.6) for Gaussian vectors. For brevity, let

$$\zeta = \begin{pmatrix} \xi \\ \eta \end{pmatrix}.$$

Then $\boldsymbol{E}\zeta = \begin{pmatrix} \boldsymbol{E}\xi \\ \boldsymbol{E}\eta \end{pmatrix}$ and $\mathsf{Cov}(\zeta,\zeta) = \begin{pmatrix} \mathsf{Cov}(\xi,\xi) & \mathsf{Cov}(\xi,\eta) \\ \mathsf{Cov}(\eta,\xi) & \mathsf{Cov}(\eta,\eta)\end{pmatrix}$. Since ξ and η are orthogonal, $\mathsf{Cov}(\xi,\eta) = 0$, so that

$$\mathsf{Cov}(\zeta,\zeta) = \begin{pmatrix} \mathsf{Cov}(\xi,\xi) & 0 \\ 0^T & \mathsf{Cov}(\eta,\eta)\end{pmatrix},$$

where 0 and 0^T are zero matrices of corresponding sizes. Hence

$$\begin{aligned}\Phi_\zeta(\lambda_1,\lambda_2) &= \exp\Big(\big\langle \imath\lambda_1, \boldsymbol{E}\xi\big\rangle - \frac{1}{2}\big\langle\lambda_1, \mathsf{Cov}(\xi,\xi)\lambda_1\big\rangle\Big) \\ &\quad\times\exp\Big(\big\langle \imath\lambda_2, \boldsymbol{E}\eta\big\rangle - \frac{1}{2}\big\langle\lambda_2, \mathsf{Cov}(\eta,\eta)\lambda_2\big\rangle\Big) \\ &= \Phi_\xi(\lambda_1)\Phi_\eta(\lambda_2),\end{aligned}$$

where $\Phi_\xi(\lambda_1)$ and $\Phi_\eta(\lambda_2)$ are the characteristic functions of ξ and η respectively.

Thus (9.6) holds true. $\square$

Remark. *The same statement holds true for infinite Gaussian sequences* $\xi = (\xi_1, \xi_2, ...)$ *and* $\eta = (\eta_1, \eta_2, ...)$ *provided that*

$$\boldsymbol{E}(\xi_i - \boldsymbol{E}\xi_i)(\eta_j - \boldsymbol{E}\eta_j) = 0, \ \forall\ i, j.$$

3. Conditional expectation coincides with conditional expectation in the wide sense.

Proof. We show that for Gaussian vector (X, Y) with subvectors X, Y and $\mathcal{M}^Y$ the linear space generated by "$1, Y$"

$$\widehat{\boldsymbol{E}}(X|\mathcal{M}^Y) = \boldsymbol{E}(X|Y) \tag{9.7}$$

holds true.

Indeed, since

$$\widehat{\boldsymbol{E}}(X|\mathcal{M}^Y) = \boldsymbol{E}X + \mathsf{Cov}(X, Y)\mathsf{Cov}^{+}(Y, Y)(Y - \boldsymbol{E}Y)$$

is the linear function of Y, then not only (X, Y) is the Gaussian vector but also $(X, Y, \widehat{\boldsymbol{E}}(X|\mathcal{M}^Y))$ is Gaussian one as well. Particularly,

$$(X - \widehat{\boldsymbol{E}}(X|\mathcal{M}^Y), Y)$$

is Gaussian vector and moreover entries of this vector are orthogonal. Consequently, they are independent. Therefore $(X - \widehat{\boldsymbol{E}}(X|\mathcal{M}^Y)$ and any random variable, being a bounded function function, say $H(Y)$ of argument Y, are independent. The latter provides

$$\boldsymbol{E}\big\{(X - \widehat{\boldsymbol{E}}(X|\mathcal{M}^Y)\big\}H(Y) = \boldsymbol{E}\big\{(X - \widehat{\boldsymbol{E}}(X|\mathcal{M}^Y)\big\}\boldsymbol{E}H(Y) = 0.$$

Thus, (9.7) holds and particularly

$$\boldsymbol{E}(X|Y) = \boldsymbol{E}X + \mathsf{Cov}(X, Y)\mathsf{Cov}^{+}(Y, Y)(Y - \boldsymbol{E}Y). \tag{9.8}$$

In other words, the optimal in the mean square sense estimate of X given Y is linear function in Y. $\square$

4. The conditional distribution is Gaussian with probability one.

Proof. We show that for (X, Y) the Gaussian vector with subvectors with X, Y, the conditional distribution

$$F_{\xi|\eta}(x) = \boldsymbol{P}(X \leq x|Y)$$

is Gaussian (P-a.s.)

Denote $\xi = X - \boldsymbol{E}(X|Y)$. A was mentioned above ξ and and $\boldsymbol{E}(X|Y)$ are independent. Hence, since $X = \boldsymbol{E}(X|Y) + \xi$, the conditional characteristic function $\Phi_{X|Y}(\lambda) = \boldsymbol{E}\big(e^{\imath\lambda X}|Y\big)$ is defined as follows (here $\widehat{X} = \boldsymbol{E}(X|Y)$)

$$\begin{aligned}\Phi_{X|Y}(\lambda) &= \boldsymbol{E}\Big(e^{\imath\lambda(\widehat{X}+\xi)}\big|Y\Big) \\ &= e^{\imath\lambda\widehat{X}}\boldsymbol{E}\big(e^{\imath\lambda\xi}|X\big) \\ &= e^{\imath\lambda\widehat{X}}\boldsymbol{E}e^{\imath\lambda\xi}\end{aligned}$$

and it remains to recall that ξ is Gaussian vector. □

5. Structure of Gaussian Markov sequence. Let $(X_n)_{n=0,1,\dots,}$ be Gaussian Markov sequence with $\mathsf{Var}(X_n) > 0$, $n \ge 0$. Set $\varepsilon_n = X_n - \boldsymbol{E}(X_n|X_{n-1})$. Then

$$X_n = \boldsymbol{E}(X_n|X_{n-1}) + \varepsilon_n.$$

By (9.8), we have $\boldsymbol{E}(X_n|X_{n-1}) = a_0(n) + a_1(n)X_{n-1}$, where

$$a_1(n) = \frac{\mathsf{Cov}(X_n, X_{n-1})}{\mathsf{Var}(X_{n-1})}, \quad a_0(n) = \boldsymbol{E}X_n - a_1(n)\boldsymbol{E}X_{n-1}.$$

Further, by Markov property, we have

$$\boldsymbol{E}(X_n|X_{n-1}, \dots, X_1, X_0) = \boldsymbol{E}(X_n|X_{n-1}),\ n \ge 1,$$

that is $\varepsilon_n = X_n - \boldsymbol{E}(X_n|X_{n-1}, \dots, X_1, X_0)$ and therefore is orthogonal to any linear combination of $X_{n-1}, \dots, X_1, X_0$. Particularly ε_n is then orthogonal to any ε_m, $m < n$. So by Gaussian property, when the orthogonality is equivalent to the independence, $(\varepsilon_n)_{n\ge 1}$ forms Gaussian sequence of zero mean independent random variables with

$$\boldsymbol{E}\varepsilon_n^2 = \mathsf{Var}(X_n) - \frac{\mathsf{Cov}^2(X_n, X_{n-1})}{\mathsf{Var}(X_{n-1}}. \tag{9.9}$$

Thus, any Gaussian Markov process sequence (X_n) is defined by the recursion

$$X_n = a_0(n) + a_1(n)X_{n-1} + \varepsilon \tag{9.10}$$

subject to the initial condition X_0.

9.1. **Gaussian analog of non-Gaussian random sequence.** Let $(X_n)_{n=0,1,\dots,}$ be a sequence of random variables with well defined $\boldsymbol{E}X_n$ and $\mathsf{Cov}(X_n, X_m)$. Since any Gaussian sequence is completely defined by its expectation and correlation functions, the Gaussian sequence $(X'_n)_{n=0,1,\dots,}$ with $\boldsymbol{E}X'_n \equiv \boldsymbol{E}X_n$ and $\mathsf{Cov}(X'_n, X'_m) \equiv \mathsf{Cov}(X_n, X_m)$ is said to be a Gaussian analog of $(X_n)_{n=0,1,\dots,}$.

If the Gaussian analog X'_n is Markov sequence, i.e. it is generated by the recursion given in (9.10), then, obviously, X_n is generated by the same recursion in which $(\varepsilon_n)_{n\ge 1}$ is a sequence of zero mean orthogonal random variables satisfying (9.9). An inverse statement is also valid.

Namely, If non-Gaussian random sequence is defined by recursion given in (9.10) subject to X_0 orthogonal to $(\varepsilon_n)_{n\ge 1}$, then (X_n') is defined by the same recursion in which $(\varepsilon_n)_{n\ge 0}$ is zero mean sequence of independent Gaussian random variables satisfying (9.9). In additional, X_0' is independent of $(\varepsilon_n)_{n\ge 0}$ and $\boldsymbol{E}X_0' = \boldsymbol{E}X_0$, $\mathsf{Var}(X_0') = \mathsf{Var}X_0$.

EXAMPLE. Let $(X_n)_{n=0,1\dots,}$ be Markov process valued in a finite alphabet $\{x_1,\dots,x_m\}$ and a transition probability matrix Λ with entries λ_{ij} and a distribution $p_0(1),\dots,p_0(m)$ for X_0 (see, Lect. 8).

We give now the Gaussian analog of that process. More exactly, we find the Gaussian analog for a sequence $I_n = \begin{pmatrix} I(X_n = x_1) \\ \vdots \\ I(X_n = x_m) \end{pmatrix}$, $n = 0, 1, \dots,$. Set

$$\varepsilon_n = I_n - \boldsymbol{E}(I_n|I_{n-1}) \tag{9.11}$$

and notice that $\boldsymbol{E}(I_n|I_{n-1}) = \begin{pmatrix} P(X_n = x_1|I_{n-1}) \\ \vdots \\ P(X_n = x_m|I_{n-1}) \end{pmatrix}$. Hence

$$\begin{aligned}
\boldsymbol{E}(I_n|I_{n-1}) = \begin{pmatrix} P(X_n = x_1|I_{n-1}) \\ \vdots \\ P(X_n = x_m|I_{n-1}) \end{pmatrix} \\
= \begin{pmatrix} \sum_{j=1}^m \lambda_{j1} I(X_{n-1} = x_j) \\ \vdots \\ \sum_{j=1}^m \lambda_{jm} I(X_{n-1} = x_j) \end{pmatrix} = \Lambda^T I_{n-1}.
\end{aligned}$$

The Markov property $\boldsymbol{E}(I_n|I_{n-1}) = \boldsymbol{E}(I_n|I_{n-1},\dots,I_1,I_0)$ provides that $(\varepsilon_n)_{n\ge 1}$ is a random sequence of zero mean orthogonal random vectors ($\boldsymbol{E}\varepsilon_n\varepsilon_m = 0$ for $n \ne m$) with the matrix $D_n = \mathsf{Cov}(\varepsilon_n, \varepsilon_n)$ defined as:

$$\begin{aligned}
D_n &= \boldsymbol{E}(I_n - \Lambda I_{n-1})(I_n - \Lambda I_{n-1})^T \\
&= \boldsymbol{E}(I_n I_n^T - I_n I_{n-1}^T \Lambda^T - \Lambda I_{n-1} I_n^T + \Lambda I_{n-1} I_{n-1}^T \Lambda^T).
\end{aligned}$$

For notational convenience, introduce $p_n = \boldsymbol{E}I_n$ the vector entries of which are $p(X_n = x_1),\dots,p(X_n = x_m)$. Noticing now that $I_n I_n^T = \mathsf{diag}(I_n)$ (here $\mathsf{diag}(\cdot)$ is diagonal matrix with the diagonal "$\cdot$"), we find $\boldsymbol{E}I_n I_n^T = \mathsf{diag}(p_n)$. Analogously, whereas $\boldsymbol{E}(I_n|I_{n-1}) = \Lambda^T I_{n-1}$, we have $\boldsymbol{E}I_n I_{n-1}^T = \mathsf{diag}(p_{n-1})\Lambda$. Hence

$$D_n = \mathsf{diag}(p_n) - \Lambda^T \mathsf{diag}(p_{n-1})\Lambda. \tag{9.12}$$

Further, taking the expectation from both sides of (9.11), we find

$$p_n = \Lambda p_{n-1}. \tag{9.13}$$

Finally, $\boldsymbol{E}I_0 = P_0$ and $\mathsf{Cov}(I_0, I_0) = \mathsf{diag}(p_0) - p_0 p_0^T$.

Thus, I'_n the Gaussian analog of I_n is defined by a linear recurrent equation

$$I'_n = \Lambda I'_{n-1} + \varepsilon'_n,$$

where I'_0 is Gaussian vector with the expectation and covariance p_0 and $\mathsf{diag}(p_0) = p_0 P_0^T$ respectively independent of $(\varepsilon'_n)_{n\ge 1}$ the sequence of zero mean independent Gaussian vectors with $\mathsf{Cov}(\varepsilon'_n \varepsilon'_n) = D_n$ (see, (9.12)).

9.2. **Continuation of Remark 8.2 from Lect. 8.** Let I_n is referred as a signal filtered from observation

$$Y_n = X_n + \eta_n$$

where $(\eta_n)_{n\ge 1}$ is zero orthogonal random sequence with $\mathsf{Var}(\eta_n) = \sigma^2$ independent of (X_n). Such a model is completely adapted to the Kalman filtering setting. In fact, for a vector $\mathfrak{x}$ with entries $x_1, \dots, x_m$ we have $X_n = \mathfrak{x}^T I_n$ and by (9.11)

$$\begin{aligned} I_n &= \Lambda^T I_{n-1} + \varepsilon_n \\ Y_n &= \mathfrak{x}^T I_n + \eta_n := \mathfrak{x}^T \Lambda^T I_{n-1} + \mathfrak{x}^T \varepsilon_n + \eta_n \end{aligned}$$

Hence, we obtain the following Kalman filter

$$\begin{aligned} \widehat{I}_n &= \Lambda \widehat{I}_{n-1} + \frac{\mathsf{Cov}(\varepsilon_n, \varepsilon_n) + \Lambda^T P_{n-1} \Lambda \mathfrak{x}}{\mathfrak{x}^T \mathsf{Cov}(\varepsilon_n, \varepsilon_n)\mathfrak{x} + \sigma^2 + \mathfrak{x}^T \Lambda^T P_{n-1} \Lambda \mathfrak{x}} (Y_n - \mathfrak{x}^T \Lambda^T \widehat{I}_{n-1}) \\ P_n &= \Lambda^T P_{n-1} \Lambda + \mathsf{Cov}(\varepsilon_n, \varepsilon_n) - \frac{\Big(\mathsf{Cov}(\varepsilon_n, \varepsilon_n) + \Lambda^T P_{n-1} \Lambda \mathfrak{x}\Big)^2}{\mathfrak{x}^T \mathsf{Cov}(\varepsilon_n, \varepsilon_n)\mathfrak{x} + \sigma^2 + \mathfrak{x}^T \Lambda^T P_{n-1} \Lambda \mathfrak{x}} \end{aligned}$$

subject to the initial conditions

$$\begin{aligned} \widehat{I}_0 &= \boldsymbol{E} I_0 + \frac{\mathsf{Cov}(I_0, \mathfrak{x}^T I_0)}{\sigma^2 + \mathsf{Cov}(\mathfrak{x}^T I_0, \mathfrak{x}^T I_0)} (Y_0 - \mathfrak{x}^T \boldsymbol{E} I_0) \\ P_0 &= \mathsf{Cov}(I_0, I_0) - \frac{\Big(\mathsf{Cov}(I_0, \mathfrak{x}^T I_0)\Big)^2}{\sigma^2 + \mathsf{Cov}(\mathfrak{x}^T I_0, \mathfrak{x}^T I_0)}. \end{aligned}$$