Alain Oliviero-Durmus alain.durmus@polytechnique.edu
École Polytechnique
Stochastic calculus for machine learning, Master MVA

1 Reminders from probability

1.1 Conditional expectation

Proposition 1.1. Let $\mathcal{F} = \sigma(Y_i : i \in \mathsf{I})$ where $\{Y_i :, i \in \mathsf{I}\}$ are random variable valued in $(\mathsf{Y}, \mathcal{Y})$. Then Z a F-measurable random variable is equal to $\mathbb{E}[X|\mathcal{F}]$ if and only if for any $\mathsf{J} \subset \mathsf{I}$ finite and $\{f_j\}_{j\in\mathsf{J}}$ measurable and bounded,

$$\mathbb{E}\left[X \prod_{j\in\mathsf{J}} f_j(Y_j)\right] = \mathbb{E}\left[Z \prod_{j\in\mathsf{J}} f_j(Y_j)\right] . \tag{1}$$

1.2 Gaussian random variables

Définition 1.2. 1. For any $\mathrm{m} \in \mathbb{R}$ and $\sigma > 0$, we denote by $\mathrm{N}(\mathrm{m}, \sigma^2)$ the Gaussian distribution on $\mathbb{R}$ which has density with respect to the Lebesgue measure given by

$$x \mapsto (2\pi\sigma^2)^{-1/2} \exp(-(x-\mathrm{m})^2/(2\sigma^2)) . \tag{2}$$

We can extend this definition for $\sigma = 0$ by setting for any $\mathrm{m} \in \mathbb{R}$, $\mathrm{N}(\mathrm{m}, 0) = \updelta_{\mathrm{m}}$, where $\updelta_{\mathrm{m}}$ is the Dirac distribution at m.

2. A real-valued random variable G is said to follow a Gaussian distribution if there exists $\mathrm{m} \in \mathbb{R}$ and $\sigma \geq 0$, such that G has distribution $\mathrm{N}(\mathrm{m}, \sigma^2)$.

3. A random variable X on $\mathbb{R}^d$ is said to be a Gaussian random variable if for any $t \in \mathbb{R}^d$, the real-valued random variable $\langle t, X\rangle$ follows a Gaussian distribution.

Proposition 1.3. Let G a standard normal random variable. Then, for any $x > 0$,

$$\frac{1}{\sqrt{2\pi}}\left(\frac{1}{x} - \frac{1}{x^3}\right) \mathrm{e}^{-x^2/2} \leq \mathbb{P}(G > x) \leq \frac{1}{\sqrt{2\pi}} \frac{1}{x} \mathrm{e}^{-x^2/2} \tag{3}$$

$$\mathbb{P}(G > x) \leq \mathrm{e}^{-x^2/2} . \tag{4}$$

Théorème 1.4. Let $(G_n)_{n\in\mathbb{N}}$ be a sequence of random variables such that for any n, G_n is follows the one-dimensional Gaussian distribution $\mathrm{N}(m_n, \sigma_n^2)$.

(1) Suppose that that $(G_n)_{n\in\mathbb{N}}$ converges in distribution to G. Then G is Gaussian.

(2) In addition, if $(G_n)_{n\in\mathbb{N}}$ converges in probability to G, then it converges in L^p for any $p \geq 1$.

Proposition 1.5. The random variable X on $\mathbb{R}^d$ is Gaussian if and only if there exists $\mathrm{m} \in \mathbb{R}^d$ and a semi-definite positive matrix Σ such that for any $t \in \mathbb{R}^d$, $\mathbb{E}[\mathrm{e}^{\mathrm{i}\langle t, X\rangle}] = \exp(\mathrm{i}t\mathrm{m} - (1/2)\langle \Sigma t, t\rangle)$. In that case, m and Σ are the mean and covariance matrix of the vector X, i.e. $\mathrm{m} = \mathbb{E}[X]$ and $\Sigma = \mathbb{E}\left[XX^{\mathrm{T}}\right]$. We say then that X follows a d-dimensional Gaussian distribution with mean m and covariance matrix Σ, denoted by $\mathrm{N}(\mathrm{m}, \Sigma)$.

Proposition 1.6. Let X be d-Gaussian random variable with distribution $\mathrm{N}(\mathrm{m}, \Sigma)$. Let $m \in \mathbb{R}^n$ and $\mathrm{M} \in \mathbb{R}^{n\times d}$. Then $Y = m + \mathrm{M}X$ is a n-dimensional Gaussian random variable with distribution $\mathrm{N}(m + \mathrm{M}\mathrm{m}, \mathrm{M}\Sigma\mathrm{M}^{\mathrm{T}})$.

Proposition 1.7. Let X be d-dimensional Gaussian random variable with distribution $\mathrm{N}(\mathrm{m}, \Sigma)$. Then if for any $i \in \{1, \ldots, d\}$, X_i is the i-th component of X, the family of one dimensional random variables $(X_i)_{i\in\{1,\ldots,d\}}$ is independent if and only if $\Sigma_{i,j} = 0$ for $i, j \in \{1, \ldots, d\}$, $i \neq j$.

Proposition 1.8. Let X be a $\sum_{i=1}^d n_i$-dimensional Gaussian random variable with distribution $\mathrm{N}(\mathrm{m}, \Sigma)$. Then for any $i \in \{1, \dots, d\}$, define $Z_i = (X_{n_{i-1}+1}, \dots, X_{n_i})$, where $n_0 = 0$ and X_i is the i-th component of X for $j \in \{1, \dots, \sum_{i=1}^d n_i$. The family of r.v.s $(Z_i)_{i\in\{1,\dots,d\}}$ are independent if and only if $\Sigma_{i,j} = 0$ for $i, j \notin \bigcup_{i=1}^d \{n_{i-1} + 1, \dots n_i\}$.

Proposition 1.9. Let

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{pmatrix} \right) ,$$

where $X \in \mathbb{R}^d$, $Y \in \mathbb{R}^p$, and Σ_{YY} is invertible. Then, the conditional expectation of X given Y is

$$\mathbb{E}[X \mid Y] = \mu_X + \Sigma_{XY}\Sigma_{YY}^{-1}(Y - \mu_Y) .$$

1.3 Exercises

Exercise 1.1. Show Proposition 1.1.

Exercise 1.2. Let $(G_i)_{i\in\{1,\dots,n\}}$ be i.i.d. one-dimensional Gaussian random variables with distribution $\mathrm{N}(0, 1)$.

1. Show that the n-dimensional vector $X = (G_1, \dots, G_n)$ is Gaussian and specify its mean and covariance matrix.
2. Deduce how to get a random variable with distribution $\mathrm{N}(\mathrm{m}, \Sigma)$ from $(G_i)_{i\in\{1,\dots,n\}}$.

Exercise 1.3. Let G be a random variable following the standard Gaussian distribution $\mathcal{N}(0, 1)$.

(1) Compute $\mathbb{E}[G^4]$ and $\mathbb{E}[|G|]$.
(2) Compute $\mathbb{E}[\mathrm{e}^{aG}]$, $\mathbb{E}[G\mathrm{e}^{aG}]$, and $\mathbb{E}[\mathrm{e}^{aG^2}]$, where $a \in \mathbb{R}$ is a real number.

Exercise 1.4. The goal of this exercise is to show Théorème 1.4-(1). As a first step, we show

Proposition 1.10. Let $(x_n)_{n\in\mathbb{N}}$ be a real sequence such that for any $t \in \mathbb{R}$, $(\mathrm{e}^{\mathrm{i}tx_n})_{n\in\mathbb{N}}$ converges, then $(x_n)_{n\in\mathbb{N}}$ converges.

(1) Suppose that $(x_n)_{n\in\mathbb{N}}$ be a real sequence such that for any $t \in \mathbb{R}$, $(\mathrm{e}^{\mathrm{i}tx_n})_{n\in\mathbb{N}}$ converges but $(x_n)_{n\in\mathbb{N}}$ does not converge.

(i) Show that $(x_n)_{n\in\mathbb{N}}$ cannot have two limit points and therefore it necessarily diverges.
(ii) Denoting for any $t \in \mathbb{R}$, $f(t) = \lim_{n\to+\infty} \mathrm{e}^{\mathrm{i}tx_n}$, show that

$$1 = \int_0^1 |f(t)|^2 \,\mathrm{d}t = \lim_{n\to\infty} \int_0^1 f(t)\mathrm{e}^{\mathrm{i}tx_n} \mathrm{d}t . \tag{5}$$

(iii) Recognizing a Fourier transform, show that $\lim_{n\to\infty} \int_0^1 f(t)\mathrm{e}^{\mathrm{i}tx_n}\mathrm{d}t = 0$.
(iv) Deduce Proposition 1.10.

(2) Using Proposition 1.10, show Théorème 1.4-(1).

Exercise 1.5. Show Proposition 1.9.