Chapter 6

Gaussian Random Variables

6.1 Random Variables in $\mathbb{R}^d$

For $x \in \mathbb{R}^d$, we choose to write x as a $d \times 1$ column vector. Let $\|\cdot\|_2$ be the Euclidean norm on $\mathbb{R}^d$ given by

$$\|x\|_2 = \sqrt{\sum_{i=1}^{d} x_i^2}$$

Suppose Z is an $\mathbb{R}^d$-valued random vector satisfying $\mathrm{E}\left[\|\right] Z\|_2^2 < \infty$.

Definition 6.1: The *mean* of Z is the $d \times 1$ column vector defined by $\mu = \mathrm{E}[Z]$. The *covariance matrix* C of Z is the $d \times d$ matrix given by

$$C = \mathrm{E}\left[(Z-\mu)(Z-\mu)^T\right]$$

Note that if $W = BZ$, where B is a (deterministic) $n \times d$ matrix, then the mean of W is $B\mu$ and its covariance matrix is BCB^T.
Any covariance matrix is automatically symmetric and non-negative definite. When C is positive definite, C can be expressed as

$$C = LL^T \tag{6.1}$$

where L is lower triangular and non-singular. (This is the so-called *Cholesky factorization* of C). If $W = L^{-1}Z$, W has the identity as its covariance matrix, so that the components of $W - \mathrm{E}[W]$ are orthogonal r.v's.

6.2 Gaussian Random Variable in $\mathbb{R}^d$

Let C be any symmetric non-negative definite $d \times d$ matrix, and let μ be a $d \times 1$ vector. The r.v Z is said to be a Gaussian r.v with covariance C and mean μ (written $Z \stackrel{\mathcal{D}}{=} N(\mu, C)$) if its characteristic function (essentially, the Fourier transform of its density) is given by

$$\mathrm{E}\left[\exp(i\theta^T Z)\right] = \exp\left(i\theta^T\mu - \frac{\theta^T C\theta}{2}\right) \tag{6.2}$$

If C is positive definite, Z has a density on $\mathbb{R}^d$ given by

$$f(z) = (2\pi)^{-d/2}|\det C|^{-1/2}\exp\left(-(z-\mu)^T\frac{C^{-1}}{2}(z-\mu)\right)$$

If C is singular, Z has a density that is concentrated on a lower dimensional affine subspace of $\mathbb{R}^d$. In particular, C can then be expressed as

$$C = B\tilde{C}B^T \tag{6.3}$$

where $\tilde{C}$ is a $k \times k$ symmetric positive definite matrix with $k < d$ and B is $d \times k$. If $X \stackrel{\mathcal{D}}{=} N(0, \tilde{C})$, then $Z = BX + \mu$ has the required characteristic function corresponding to (6.2).
The key properties of Gaussian random variables are as follows:

1. The family of Gaussian r.v's is closed under affine operations.

   If $Z \stackrel{\mathcal{D}}{=} N(\mu, C)$, then $W = BZ + \nu \stackrel{\mathcal{D}}{=} N(B\mu + \nu,\ BCB^T)$.

2. If C is diagonal, then the components of $Z \stackrel{\mathcal{D}}{=} N(\mu, C)$ are independent r.v's (so their uncorrelatedness implies independence when Z is Gaussian).

3. Any $\mathbb{R}^d$-valued Gaussian r.v can be written as a linear combination of independent scalar Gaussian r.v's.

   In particular, if $Z \stackrel{\mathcal{D}}{=} N(\mu, C)$ then we first write $C = B\tilde{C}B^T$, where $\tilde{C}$ is a $k \times k$ positive definite covariance matrix with $k \leq d$ (see (6.3)). We next express $\tilde{C}$ as $\tilde{C} = LL^T$ (see (6.1)). If $X \stackrel{\mathcal{D}}{=} N(0, I)$ is a $\mathbb{R}^k$-valued r.v, then $Z \stackrel{\mathcal{D}}{=} BLX + \mu$.

4. Suppose $Z = (Z_1, Z_2)^T$ is Gaussian with mean μ and covariance C. Then, the conditional distribution of Z_1 given Z_2 is Gaussian with a mean that is affine in Z_2.

   Specifically, put $\tilde{Z}_i = Z_i - \mathrm{E}[Z_i]$ and write C in the partitioned form

   $$C = \begin{pmatrix} \mathrm{E}\left[\tilde{Z}_1\tilde{Z}_1^T\right] & \mathrm{E}\left[\tilde{Z}_1\tilde{Z}_2^T\right] \\ \mathrm{E}\left[\tilde{Z}_2\tilde{Z}_1^T\right] & \mathrm{E}\left[\tilde{Z}_2\tilde{Z}_2^T\right] \end{pmatrix}$$

   If C is positive definite, then the conditional distribution of Z_1 given Z_2 is Gaussian with mean

   $$\mathrm{E}[Z_1] + \mathrm{E}\left[\tilde{Z}_1\tilde{Z}_2^T\right](\mathrm{E}\left[\tilde{Z}_2\tilde{Z}_2^T\right])^{-1}(Z_2 - \mathrm{E}[Z_2])$$

   and covariance matrix

   $$\mathrm{E}\left[\tilde{Z}_1\tilde{Z}_1^T\right] - \mathrm{E}\left[\tilde{Z}_1\tilde{Z}_2^T\right](\mathrm{E}\left[\tilde{Z}_2\tilde{Z}_2^T\right])^{-1}\mathrm{E}\left[\tilde{Z}_2\tilde{Z}_1^T\right]$$

   In particular, the *best affine predictor coincides with the best non-linear predictor* (i.e., the conditional expectation).

5. Sums of independent Gaussian r.v's are Gaussian.

   In particular, suppose that $Z_1, \ldots, Z_n$ are n independent $\mathbb{R}^d$-valued r.v's with means $\mu_1, \ldots, \mu_n$ and covariance matrices $C_1, \ldots, C_n$. Then

   $$Z_1 + \ldots + Z_n \stackrel{\mathcal{D}}{=} N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n C_i\right)$$

6.3 Gaussian Processes

Gaussian processes are widely used in many applied settings. A stochastic process $X = (X(t) : t \in T)$ is said to be *Gaussian* if $(X(t_1), \ldots, X(t_n))$ is Gaussian for any finite collection $t_1, \ldots, t_n$ of time points chosen from T.
A key advantage of the Gaussian assumption is the ease with which it can be characterized. In particular, if X is real-valued, one need only specify the mean function $m(t)(= \mathrm{E}[X(t)])$ for $t \in T$ and the covariance function $c(s,t)(= \mathrm{cov}(X(s), X(t),))$ for $s, t \in T$. Of course, in order that $c : T \times T \to I\!R$ must be a legitimate covariance function, it must that for each finite subset $\{t_i : 1 \leq i \leq n\} \subseteq T$, the $n \times n$ matrix

$$c(t_i, t_j) : 1 \leq i, j \leq n$$

is symmetric and non-negative definite.

Proposition 6.1: If X is stationary, then $m(t) = m$ for $t \in T$ and $c(s,t) = c(|t-s|)$ for $s, t \in T$.

Here are some examples of Gaussian processes.

Example 6.1: Suppose that X is stationary with $c(t) = \sigma^2 \exp(-r|t|)$ for $r > 0$. Then X is called the (stationary) Ornstein-Uhlenbeck process. In this case, the autocorrelation function of X (and its "memory") decays exponentially in time.

Exercise 6.1: Suppose that X is as in Example Example 6.1. Compute the conditional distribution of $X(t)$ for $0 \leq t \leq 1$, conditional on $X(0)$ and $X(1)$.

Example 6.2: Suppose that X is a stationary Gaussian process with $T = I\!R$. Then, Z is a stationary Gaussian process, provided that

$$Z(t) = \int_{-\infty}^{\infty} a(u) X(t-u)\, du$$

with $a(\cdot)$ deterministic and

$$\int_{-\infty}^{\infty} |a(u)|\, du < \infty$$

Example 6.3: Suppose that X has $m(t) = 0$ for $t \geq 0$, with $c(s,t) = \min(s,t)$ for $s, t \geq 0$. This Gaussian process is known as *standard Brownian motion*.

Exercise 6.2: Prove that if X is a standard Brownian motion, then:

1. X has independent increments (i.e., for $t_1 < t_2 < \ldots < t_n$, $X(t_1) - X(0), X(t_2) - X(t_1), \ldots, X(t_n) - X(t_{n-1})$ are independent r.v's).
2. X has stationary increments (i.e., for $0 \leq s \leq t, X(t) - X(s) \stackrel{\mathcal{D}}{=} X(t-s) - X(0)$).
3. For $\sigma^2 > 0$, $X(\sigma^2 \cdot) \stackrel{\mathcal{D}}{=} \sigma X(\cdot)$ (i.e., X is a "self-similar" process; this is related to the fractal behavior of X).
4. Put $Y(t) = tX(1/t)$ for $t \geq 0$ (Define $Y(0)$ via right continuity). Then, $Y(\cdot) \stackrel{\mathcal{D}}{=} X(\cdot)$.

Example 6.4: Suppose that X has $m(t) = 0$ for $t \geq 0$ and

$$c(s,t) = \frac{1}{2}\{|s|^{2H} + |t|^{2H} - |t-s|^{2H}\}$$

for $s, t \geq 0$ (with $0 \leq H \leq 1$). This Gaussian process is known as *standard fractional Brownian motion* with Hurst parameter H.

Exercise 6.3: Prove that:

1. If X is a standard fractional Brownian motion with $H = 1/2$, then X is standard Brownian motion.
2. If X is a standard fractional Brownian motion, then X has stationary increments.
3. If X is a standard fractional Brownian motion with Hurst parameter H, then $X(\sigma^2 \cdot) \overset{\mathcal{D}}{=} \sigma^{2H} X(\cdot)$ for $\sigma^2 > 0$. (i.e., fractional Brownian motion is self-similar, and is a fractal process).
4. Prove that $\mathrm{corr}\left(()\, X(1) - X(0), X(t+1) - X(t)\right) \sim \gamma t^{-\beta}$ as $t \to \infty$ and compute γ and β. (This calculation establishes that fractional Brownian motion has "long-range dependent" increments when $H \neq 1/2$ in the sense that the autocorrelation function is not integrable).
5. Compute the conditional distribution of $X(t)$, conditional on $X(0)$ and $X(1)$, for $0 \leq t \leq 1$. (Note the behaviour at $H = 1/2$).

Gaussian processes can have "realizations" or "sample paths" that vary from being quite irregular to rather smooth. To be specific, we are referring here to the smoothness properties of the function $X(\cdot, \omega)$ (i.e. the behavior of X as a function of time).
The first question we can ask is: "when does a Gaussian process have continuous realizations (or, equivalently, sample paths)?" The next problem addresses this issue.

Exercise 6.4: Suppose that $X = (X(t) : -\infty < t < \infty)$ is a stationary Gaussian process. Prove that $X(t+h) \to X(t)$ in L^2 as $h \to 0$ if and only if $\rho(h) \triangleq \mathrm{corr}\left(()\, X(0), X(h)\right) \to \rho(0)$ as $h \to 0$.

The next problem addresses the question of when a stationary Gaussian process is differentiable in mean-square.

Exercise 6.5: Let $X = (X(t) : -\infty < t < \infty)$ be a stationary Gaussian process. For $h \neq 0$ put $D(h) = (X(t+h) - X(t))/h$.

a.) Prove that $D(h)$ is Gaussian for each $h \neq 0$ and compute the mean and variance of $D(h)$.

b.) Prove that if $D(h) \to W$ in L^2 as $h \to 0$, then $\rho(\cdot)$ must be twice differentiable at $h = 0$ with $\rho'(0) = 0$.

Hence, the mean-square continuity and differentiability of a Gaussian process is intimately related to the behavior of its covariance function at $h = 0$. These results generalize, in a suitable way, to non-Gaussian processes.
Remark. The above problems discuss the issue of mean-square continuity and differentiability. The more subtle issue of whether the sample paths of a Gaussian process are (with probability one) continuous or differentiable requires more mathematical machinery to answer. (Of course, the required conditions on the covariance/correlation function will need to be at least as strong as those needed to obtain mean-square continuity and differentiability.)

6.4 Gaussian Random Fields

Gaussian random fields are frequently used in describing (random) surfaces in $\mathbb{R}^d$. In particular, a Gaussian random field in $\mathbb{R}^d$ is a real-valued collection of Gaussian r.v's indexed by points in $\mathbb{R}^d$ i.e., $X = (X(x) : x \in \mathbb{R}^d)$. As with Gaussian processes, such random fields are characterized by their means (i.e., $m(x) = \mathrm{E}\left[X(x)\right]$ for $x \in \mathbb{R}^d$) and their covariances (i.e., $c(x, y) = \mathrm{cov}\left(X(x), Y(y)\right)$ for $x, y \in \mathbb{R}^d$.
A Gaussian random field model is frequently used in the physical sciences to describe log-concentrations as a function of space (and sometimes, space / time). A common simplifying assumption is to impose the requirement that the random field is "stationary and isotropic" so that $m(x) = m$ for $x \in \mathbb{R}^d$ and

$c(x,y) = c(\|x-y\|_2)$ for $x, y \in \mathbb{R}^d$ (so that the covariance function depends only on the distance between points). A frequently used covariance specification is to require that

$$c(x,y) = \sigma^2 \exp(-r\|x-y\|_2)$$

for $x, y \in \mathbb{R}^d$ (with $r > 0$). Such random fields are used when making spatial interpolations based on experimental data gathered at various spatial reference points. Note that the use of such stochastic models provides a natural means of computing such spatial interpolations (namely, by computing conditional distributions given the observed values of the random field at the spatial reference points). This forms the basis for the widely applied method of "kriging" that originated in the geosciences.
As for Gaussian random processes, the smoothness of the Gaussian random field is determined by the smoothness of the correlation function.

6.5 Parameter Estimation for Gaussian Models

In most experimental settings, the mean and covariance structure must be estimated from observed data. There are (at least) two different contexts that arise in practice:

1. Repeated observations from a single Gaussian population (e.g. weight / height combinations of randomly selected individuals).

   Here, we model the observed data as iid samples $Z_1, Z_2, \ldots, Z_n$ from a Gaussian population with (unknown) mean μ and (unknown) covariance matrix C.

2. Observations collected from a single realization of a Gaussian process or random field (e.g. spatial measurements of log-concentrations of a groundwater contaminant).

   Here, we model the data collected as the values of the random fields $Z(x_1), \ldots, Z(x_n)$ associated with n collection points $x_1, \ldots, x_n$. The mean function $m(\cdot)$ and covariance function $c(\cdot)$ are assumed to depend on a small number of (unknown) parameters.

In either of the two above settings, maximum likelihood is typically used to develop estimators for the unknown parameters. For example, in context i) above, the maximum likelihood estimators for μ and C are, respectively,

$$\begin{aligned} \hat{\mu}_n &= \frac{1}{n}\sum_{i=1}^{n} Z_i \\ \hat{C}_n &= \frac{1}{n}\sum_{i=1}^{n} Z_i Z_i^T - \hat{\mu}_n \mu_n^T \end{aligned} \tag{6.4}$$

In context ii), the maximization of the likelihood may need to be done numerically. It should be noted that the bootstrap offers one means of computing confidence regions for the unknown parameters in such problems.