

# Math 106 – Notes

## Week 3: January 22, 2026

Ethan Levien

### Gaussian process (5.4)

#### 0.1 Review of multivariate Gaussians

Throughout this section we assume all random variables and stochastic processes have mean zero. This assumption simplifies notation and calculations and does not affect any of the conceptual conclusions.

**Maximum entropy principle (informal statement).** Among all probability distributions on  $\mathbb{R}^n$  with a prescribed covariance matrix  $\Sigma$ , the mean-zero multivariate Gaussian distribution with covariance  $\Sigma$  is the unique distribution of maximal entropy (see Appendix A for details). In this sense it is the “least informative” or “most random” choice compatible with the covariance  $\Sigma$ .

This maximum-entropy characterization provides a natural motivation for Gaussian processes. Suppose we wish to model a random function  $X = \{X_t\}_{t \in \mathbb{R}}$  while specifying only its second-order statistics, encoded by a covariance function

$$K(s, t) = \text{cov}(X_s, X_t), \quad s, t \in \mathbb{R}. \quad (1)$$

The guiding principle is to choose a stochastic process that is otherwise as unstructured as possible, subject only to this covariance constraint.

**Definition 1** (Gaussian process). *A stochastic process  $\{X_t\}_{t \in \mathbb{R}}$  is called a Gaussian process with covariance function  $K$  if for every finite collection of indices  $t_1, \dots, t_n \in \mathbb{R}$ , the random vector*

$$(X_{t_1}, \dots, X_{t_n}) \quad (2)$$

*is multivariate normal with mean zero and covariance matrix*

$$\Sigma_{i,j} = K(t_i, t_j). \quad (3)$$

From the maximum-entropy perspective, a Gaussian process is the unique stochastic process whose finite-dimensional distributions maximize entropy subject to the constraint that their covariances are prescribed by  $K$ . The existence and consistency of such a process follow from the Kolmogorov extension theorem.

## A Properties of Gaussian

Recall some basic properties of (mean-zero) multivariate Gaussian vectors. If  $\mathbf{X} \sim \text{Normal}(0, \Sigma)$ , then its moment-generating function (or equivalently its characteristic function) has the closed form

$$M_{\mathbf{X}}(t) = \mathbb{E}\left[e^{t^\top \mathbf{X}}\right] = \exp\left(\frac{1}{2}t^\top \Sigma t\right), \quad t \in \mathbb{R}^n. \quad (4)$$

Equivalently, the characteristic function is

$$\varphi_{\mathbf{X}}(t) = \mathbb{E}\left[e^{it^\top \mathbf{X}}\right] = \exp\left(-\frac{1}{2}t^\top \Sigma t\right). \quad (5)$$

Several standard consequences follow immediately:

- All mixed moments of  $\mathbf{X}$  can be computed by differentiating  $M_{\mathbf{X}}(t)$  at  $t = 0$ .
- The covariance matrix satisfies

$$\text{cov}(X_i, X_j) = \frac{\partial^2}{\partial t_i \partial t_j} \log M_{\mathbf{X}}(t) \Big|_{t=0} = \Sigma_{i,j}. \quad (6)$$

- Any linear transformation of a Gaussian vector is again Gaussian: if  $\mathbf{Y} = A\mathbf{X}$ , then

$$\mathbf{Y} \sim \text{Normal}\left(0, A\Sigma A^\top\right). \quad (7)$$

These facts will be used repeatedly when we construct Gaussian processes and analyze their finite-dimensional distributions.

## B Maximum Entropy Characterization of the Gaussian (optional)

A fundamental characterization of the Gaussian is the following:

**Theorem 1** (Maximum Entropy Characterization). *Let  $\mathcal{P}$  be the class of all absolutely continuous probability distributions  $\mu$  on  $\mathbb{R}^n$  with mean zero and covariance*

$$\int_{\mathbb{R}^n} xx^\top d\mu(x) = \Sigma, \quad (8)$$

*where  $\Sigma$  is a fixed positive-definite matrix. Among all  $\mu \in \mathcal{P}$ , the differential entropy*

$$h(\mu) = - \int_{\mathbb{R}^n} f(x) \log f(x) dx, \quad (9)$$

*where  $f$  is the density of  $\mu$ , is uniquely maximized by the Gaussian density*

$$\phi_\Sigma(x) = \frac{1}{(2\pi)^{n/2}(\det \Sigma)^{1/2}} \exp\left(-\frac{1}{2}x^\top \Sigma^{-1}x\right), \quad (10)$$

*and the maximal entropy equals*

$$h(\phi_\Sigma) = \frac{1}{2} \log((2\pi e)^n \det \Sigma). \quad (11)$$

*Proof.* Let  $\mu$  be any distribution in  $\mathcal{P}$  with density  $f$ , and let  $\phi_\Sigma$  be the mean-zero Gaussian density with covariance  $\Sigma$ . Consider the Kullback–Leibler divergence

$$D(f\|\phi_\Sigma) = \int_{\mathbb{R}^n} f(x) \log \frac{f(x)}{\phi_\Sigma(x)} dx. \quad (12)$$

Since  $D(f\|\phi_\Sigma) \geq 0$  for all  $f$ , and equality holds if and only if  $f = \phi_\Sigma$  almost everywhere, we expand the divergence:

$$D(f\|\phi_\Sigma) = \int f(x) \log f(x) dx - \int f(x) \log \phi_\Sigma(x) dx. \quad (13)$$

The first term equals  $-h(f)$ . To compute the second term, write

$$\log \phi_\Sigma(x) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log(\det \Sigma) - \frac{1}{2} x^\top \Sigma^{-1} x. \quad (14)$$

Taking expectations under  $f$  and using that  $\mu$  has covariance  $\Sigma$ ,

$$\int f(x) x^\top \Sigma^{-1} x dx = \text{tr}(\Sigma^{-1} \Sigma) = n. \quad (15)$$

Hence

$$\int f(x) \log \phi_\Sigma(x) dx = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log(\det \Sigma) - \frac{1}{2} n. \quad (16)$$

Substitute this into the expression for  $D(f\|\phi_\Sigma)$ :

$$D(f\|\phi_\Sigma) = -h(f) + \frac{n}{2} \log(2\pi) + \frac{1}{2} \log(\det \Sigma) + \frac{1}{2} n. \quad (17)$$

Rearranging yields

$$h(f) = \frac{n}{2} \log(2\pi e) + \frac{1}{2} \log(\det \Sigma) - D(f\|\phi_\Sigma). \quad (18)$$

Since  $D(f\|\phi_\Sigma) \geq 0$ , we obtain

$$h(f) \leq \frac{1}{2} \log((2\pi e)^n \det \Sigma), \quad (19)$$

with equality if and only if  $f = \phi_\Sigma$  almost everywhere. Thus  $\phi_\Sigma$  is the unique maximizer of entropy subject to the covariance constraint.  $\square$