Alon Rosen

Gaussian Primer

Gaussian distributions appear throughout the course: as continuous noise, as the endpoint of forward diffusion, and as the randomness from which diffusion sampling begins. We will need only a small collection of facts about them.

The main points of this chapter are:

- Linear maps and sums preserve Gaussianity.
- General Gaussians are linear images of standard Gaussians.
- Standard Gaussians are rotationally invariant, with independent orthogonal projections.

1 One-dimensional Gaussians

Write $X \sim \mathcal{N}(\mu, \sigma^2)$ for a one-dimensional Gaussian with mean μ and variance σ^2. Its density is

$$\phi_\sigma(x-\mu) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

The case $\mu = 0$ and $\sigma^2 = 1$ is the *standard Gaussian* $\mathcal{N}(0,1)$.

Fact 1. *If $X \sim \mathcal{N}(\mu, \sigma^2)$ and $Y \sim \mathcal{N}(\nu, \tau^2)$ are independent, and $a, b \in \mathbb{R}$, then*

- $aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$.
- $X + Y \sim \mathcal{N}(\mu + \nu, \sigma^2 + \tau^2)$.

2 Multi-dimensional Gaussians

A random vector $\mathbf{X} = (X_1, \ldots, X_n) \in \mathbb{R}^n$ has mean $\boldsymbol{\mu} = \mathbb{E}[\mathbf{X}]$ and covariance matrix

$$\Sigma = \mathbb{E}\big[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^\top\big].$$

The diagonal entry Σ_{ii} is the variance of X_i, while an off-diagonal entry Σ_{ij} is the covariance between X_i and X_j. The standard Gaussian in $\mathbb{R}^n$ has independent $\mathcal{N}(0,1)$ coordinates; in particular,

$$\mathbf{g} \sim \mathcal{N}(0, I_n), \qquad \mathbb{E}[\mathbf{g}] = 0, \qquad \mathbb{E}[\mathbf{g}\mathbf{g}^\top] = I_n.$$

Suppose that Σ is a covariance matrix and let A be such that $AA^\top = \Sigma$. Such a matrix A can always be found. Since Σ is a covariance matrix, it is symmetric and positive semidefinite, so it has an eigendecomposition

$$\Sigma = U \operatorname{diag}(\lambda_1, \ldots, \lambda_n) U^\top, \qquad \lambda_i \geq 0.$$

We may therefore take $A = U \operatorname{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n})$.

Fact 2. *If $\mathbf{g} \sim \mathcal{N}(0, I_n)$, then $\mathbf{X} = \boldsymbol{\mu} + A\mathbf{g}$ is a Gaussian with mean $\boldsymbol{\mu}$ and covariance Σ.*

Proof. For the mean,

$$\mathbb{E}[\mathbf{X}] = \mathbb{E}[\boldsymbol{\mu} + A\mathbf{g}] = \boldsymbol{\mu} + A\mathbb{E}[\mathbf{g}] = \boldsymbol{\mu}.$$

For the covariance,

$$\mathbb{E}[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^\top] = A\mathbb{E}[\mathbf{g}\mathbf{g}^\top]A^\top = AA^\top = \Sigma.$$

□

Thus sampling a Gaussian with a given mean and covariance reduces to sampling independent one-dimensional standard Gaussians and applying a linear map. The same calculation gives the basic rule for linear maps.

Fact 3. *If $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$, then for every matrix B and vector $\mathbf{b}$,*

$$B\mathbf{X} + \mathbf{b} \sim \mathcal{N}(B\boldsymbol{\mu} + \mathbf{b}, B\Sigma B^\top).$$

In particular, for any $\mathbf{v} \in \mathbb{R}^n$, $\mathbf{v}^\top\mathbf{X} \sim \mathcal{N}(\mathbf{v}^\top\boldsymbol{\mu}, \mathbf{v}^\top\Sigma\mathbf{v})$. Thus $\mathbf{v}^\top\Sigma\mathbf{v}$ is the variance of $\mathbf{v}^\top\mathbf{X}$.

3 The standard Gaussian

The standard Gaussian has density

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{\|\mathbf{x}\|_2^2}{2}\right).$$

This density depends only on $\|\mathbf{x}\|_2$, a property known as *rotational invariance*.

Fact 4. *If U is orthogonal and $\mathbf{g} \sim \mathcal{N}(0, I_n)$, then $U\mathbf{g} \sim \mathcal{N}(0, I_n)$.*

Fact 5. *Let $\mathbf{w}$ be any fixed unit vector, and write $\mathbf{g} = g^{\|}\mathbf{w} + \mathbf{g}_\perp$, where $g^{\|} = \mathbf{g}^\top \mathbf{w}$. Then $g^{\|} \sim \mathcal{N}(0,1)$ and $\mathbf{g}_\perp \sim \mathcal{N}(0, I_{n-1})$, independently.*

Fact 6. *For the standard Gaussian density p, we have $\log p(\mathbf{x}) = \text{constant} - \|\mathbf{x}\|_2^2/2$, and hence $\nabla \log p(\mathbf{x}) = -\mathbf{x}$. Thus the score of a standard Gaussian is simply $-\mathbf{x}$.*

4 Discrete Gaussians

A discrete Gaussian is the integer analogue of an ordinary Gaussian. For a parameter $s > 0$, an integer $k \in \mathbb{Z}$ is assigned probability proportional to

$$\exp\left(-\frac{k^2}{2s^2}\right).$$