2.4.1 SAMPLING A MULTIVARIATE GAUSSIAN

Define $x \in \mathbb{R}^d$, a multivariate Gaussian:

$$p(x) = (2\pi)^{-\frac{d}{2}} |\det \Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right),$$

where $\mu \in \mathbb{R}^d$ is the mean vector and $\Sigma \in \mathbb{R}^{d\times d}$ is a $d \times d$ symmetric positive definite matrix. Recall that, in the univariate case, $Y = \mu + \sigma X$ (where μ, σ are scalars) gave us a sample from $\mathcal{N}(\mu, \sigma^2)$. The same idea works here, however, since now we have the covariance instead of variance, we need to find a notion of a "square-root" of the covariance matrix Σ. This is done using a Cholesky decomposition[5]. The algorithm is provided below.

Algorithm 6 Sampling Multivarite Gaussian

1: Input: The number of samples n.
2: **for** $i = 1, \dots, n$ **do**
3: Compute L such that $\Sigma = LL^\top$. (Cholesky decomposition)
4: Draw d univariate independent normals $v_k \sim \mathcal{N}(0,1)$ to form the vector $v = [v_1, \dots, v_d]^\top$
5: Generate $x_i = \mu + Lv$.
6: **end for**

2.5 SOLVED EXAMPLES

Example 2.15 (Rejection sampling)**.** Let us go back to $\text{Beta}(2,2)$ example we used to demonstrate the fundamental theorem of simulation. We can now formalise it. Let

$$p(x) = \text{Beta}(\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)+\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}.$$

Ignoring the normalising constant in front, we can choose

$$\bar{p}(x) = x^{\alpha-1}(1-x)^{\beta-1},$$

and given that we used uniform "box" before, we choose:

$$q(x) = \text{Unif}(0,1)$$

We would like to compute

$$M = \sup_x \frac{\bar{p}(x)}{q(x)},$$

as in our demonstration we have computed this quantity visually. For this, we compute

$$\log \bar{p}(x)/q(x) = (\alpha-1)\log x + (\beta-1)\log(1-x)$$

[5]You do not need to know how to implement or compute this, it is perfectly fine to use `numpy.linalg.cholesky`.

The derivative

$$\frac{\mathrm{d}\log \bar{p}(x)/q(x)}{\mathrm{d}x} = \frac{\alpha - 1}{x} + \frac{1-\beta}{1-x}$$

The maximum is

$$x^\star = \frac{\alpha - 1}{\alpha + \beta - 2}.$$

Finding $x^\star$, we compute the supremum by plugging $x^\star$ into the ratio $\bar{p}/q$ which is given as

$$M = \frac{\bar{p}(x^\star)}{q(x^\star)}.$$

This leads to

$$M = \frac{(\alpha - 1)^{\alpha - 1}(\beta - 1)^{\beta - 1}}{(\alpha + \beta - 2)^{\alpha + \beta - 2}}.$$

We can find our optimal M by plugging $\alpha = 2$ and $\beta = 2$. The procedure is then given by

- Sample $X' \sim q(x) = \text{Unif}(0, 1)$
- Sample $U \sim \text{Unif}(0, 1)$
- If $U \leq \bar{p}(X')/Mq(X')$,
  - Accept X'

Example 2.16 (Rejection sampling)**.** Let us prove now the fact the average acceptance probability (acceptance rate) is given as

$$\hat{a} = \mathbb{E}[a(X')] = \frac{1}{M} \tag{2.9}$$

in the normalised case. Similarly, we will also prove

$$\hat{a} = \mathbb{E}[a(X')] = \frac{Z}{M} \tag{2.10}$$

for the unnormalised case where we use $\bar{p}(x)$ instead of $p(x)$. For the first fact, we can prove (2.9) by noting

$$\begin{aligned}
\hat{a} = \mathbb{E}[a(X')] &= \int a(x')q(x')\mathrm{d}x' \\
&= \int \frac{p(x')}{Mq(x')} q(x')\mathrm{d}x' \\
&= \frac{1}{M}\int p(x')\mathrm{d}x' \\
&= \frac{1}{M}.
\end{aligned}$$

For the unnormalised case, we can prove (2.10) as For the unnormalised case:

$$
\begin{aligned}
\hat{a} = \mathbb{E}[a(X')] &= \int a(x')q(x')\mathrm{d}x' \\
&= \int \frac{\bar{p}(x')}{Mq(x')} q(x')\mathrm{d}x' \\
&= \int Z\frac{p(x')}{Mq(x')} q(x')\mathrm{d}x' \\
&= \frac{Z}{M}\int p(x')\mathrm{d}x' \\
&= \frac{Z}{M}.
\end{aligned}
$$

Example 2.17 (Rejection sampling)**.** Consider the following example where we describe a sampling method for Gaussian using a Cauchy distribution. Let

$$
\begin{aligned}
\bar{p}(x) &= e^{-x^2/2} \\
q(x) &= \frac{1}{\pi}\frac{1}{1+x^2}.
\end{aligned}
$$

We need to compute

$$M = \sup_x \frac{\bar{p}(x)}{q(x)},$$

as usual. For this we compute

$$\log \bar{p}(x)/q(x) = -\frac{x^2}{2} + \log(1+x^2) + \log(1/\pi)$$

and find the roots Taking the derivative

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x}\log \bar{p}(x)/q(x) = -x + \frac{2x}{1+x^2} = 0 \\
x = 0, \pm 1.
\end{aligned}
$$

We have three roots to decide. Which one is the maximum? To look at the answer, we need to check second derivatives. We compute the second derivative

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}\log \bar{p}(x)/q(x) = -1 + \frac{2(1-x^2)}{(1+x^2)^2} = 0$$

- When $x = 0$, the second derivative is positive - which means $x = 0$ is a minimum.
- When $x = \pm 1$, the second derivative is negative - which means $x = \pm 1$ is a maximum.
- $x^\star = \pm 1$.

So we have

$$M = \frac{\bar{p}(1)}{q(1)} = 2\pi e^{-1/2}.$$

Example 2.18 (Marginalisation)**.** Consider

$$p(x) = \mathcal{N}(x; \mu, \sigma_0^2)$$
$$p(y|x) = \mathcal{N}(y; x, \sigma^2).$$

We aim at computing $p(y)$. The direct computation of the integral

$$p(y) = \int p(y|x)p(x)\mathrm{d}x = \int \mathcal{N}(y; x, \sigma^2)\mathcal{N}(x; \mu, \sigma_0^2)\mathrm{d}x.$$

could be tedious. Note that

$$y = (y - x) + x$$
$$y - x \sim \mathcal{N}(y - x; 0, \sigma^2)$$
$$x \sim \mathcal{N}(x; \mu, \sigma_0^2).$$

This is a sum of Gaussians. Therefore, $p(y)$ is also a Gaussian with means and variances summed:

$$p(y) = \mathcal{N}(y; \mu, \sigma_0^2 + \sigma^2).$$

Example 2.19 (Proof of Fundamental Theorem of Simulation)**.** This proof required the knowledge of marginalisation – we can now attempt at proving this theorem. For completeness, we state the theorem below.

Theorem. *Drawing samples from one dimensional random variable X with a density $p(x) \propto \bar{p}(x)$ is equivalent to sampling uniformly on the two dimensional region defined by*

$$\mathsf{A} = \{(x, y) \in \mathbb{R}^2 : 0 \leq y \leq \bar{p}(x)\}. \tag{2.11}$$

In other words, if (x', y') is uniformly distributed on A, then x' is a sample from $p(x)$.

The proof idea: Start from a uniform distribution $q(x, y)$ on A and show that the marginal in x is $p(x)$.

Proof. Consider the pair (X, Y) uniformly distributed on the region A. We denote their joint density as $q(x, y)$ as

$$q(x, y) = \frac{1}{|\mathsf{A}|}, \qquad \text{for } (x, y) \in \mathsf{A}. \tag{2.12}$$

where $|\mathsf{A}|$ is the area of the set A. We note that

$$p(x) = \frac{\bar{p}(x)}{|\mathsf{A}|}.$$

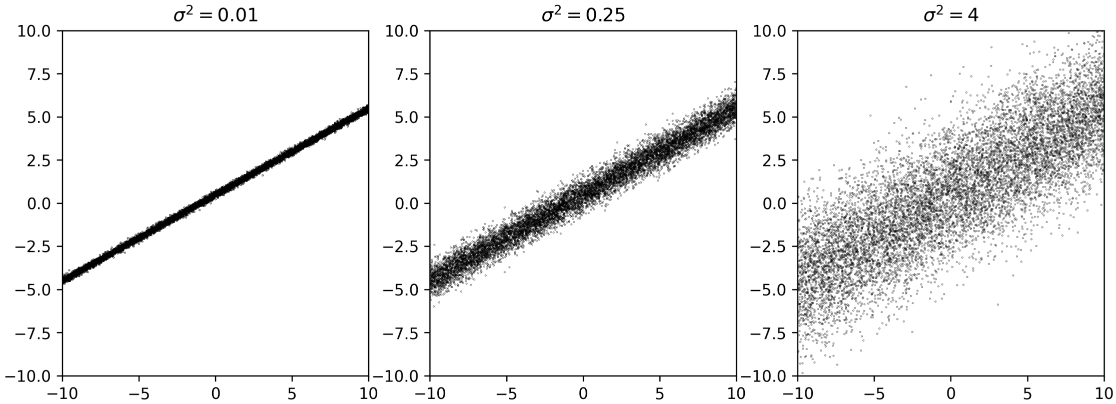

Figure 2.8: The data simulated from (2.15)–(2.16) using $a = 0.5$ and $b = 0.5$ with three different values for σ^2. As can be seen from the figures, the generated (x, y) pairs exhibit a clear linear relationship (as intended) with variance changing depending on our modelling choice.

We use the standard formula for the joint density $q(x, y) = q(y|x)q(x)$. Note that, since (X, Y) is uniform in A, for fixed x, we have

$$q(y|x) = \frac{1}{\bar{p}(x)} \quad \text{for } (x, y) \in \mathsf{A}.$$

We therefore write

$$q(x, y) = q(y|x)q(x) = \frac{q(x)}{\bar{p}(x)} \qquad \text{for } (x, y) \in \mathsf{A}. \tag{2.13}$$

We consider now (2.12) and (2.13) which are both valid on $(x, y) \in \mathsf{A}$. Combining them gives

$$q(x) = \frac{\bar{p}(x)}{|\mathsf{A}|},$$

which means $q(x) = p(x)$. □

Example 2.20 (Linear Model)**.** Linear models are of utmost importance in many fields of science. Assume that we would like to simulate (x, y) pairs that have a linear relationship. We know that we can sample $x, y \sim p(x, y)$ by sampling $x \sim p(x)$ and $y|x \sim p(y|x)$ from the last chapter. We will now use this for a linear example.

To start intuitively, a typical linear relationship is described as

$$y = ax + b, \tag{2.14}$$

which describes a line where a is the slope and b is the intercept. In order to obtain a probabilistic model and generate data, we have to simulate both x and y variables. Since, from the equation, it is clear that y is generated *given* x, we should start from defining x.

Now this depends on the application. For example, x can be a variable that may be uniform or a Gaussian. We denote its density as $p(x)$. The typical task is also to formulate $p(y|x)$. The linear equation suggests a deterministic relationship, however, real data often contains *noise*. To generate realistic data, we will instead assume

$$y = ax + b + n$$

where $n \sim \mathcal{N}(0, \sigma^2)$ is *noise* (often with small σ^2. Note that, given noise is zero mean and $ax + b$ is a deterministic number (given x), we can then write our full model

$$p(x) = \text{Unif}(x; -10, 10) \tag{2.15}$$
$$p(y|x) = \mathcal{N}(y; ax + b, \sigma^2). \tag{2.16}$$

where we chose our $p(x)$ distribution to be uniform on $[-10, 10]$. As a result, we have a full model to simulate variables with a linear relationship

$$X_i \sim p(x),$$
$$Y_i | X_i = x_i \sim p(y|x_i),$$

where $p(x)$ could be a uniform, Gaussian, truncated Gaussian etc. depending on the nature of the modelled variable. The results of this generation can be seen in the scatter plot in Fig. 2.8.