Machine Intelligence 2 | **Exercise Sheet 4**
SoSe 2020, Obermayer/Kashef | due: 2020-05-21 23:55

Density Transformations & random number generation

Exercise 4.1: Inverse CDF and Random Number Generation (4 points)

Background: If $F_X(x)$ is the cumulative distribution function (cdf) of a random variable X, then the random variable $Z = F_X(X)$ is uniformly distributed on the interval $[0, 1]$. This result provides a general recipe to generate samples $\tilde{x}$ of a random variable X with a desired probability density function (pdf) $p_X(x)$ from uniformly distributed random numbers $\tilde{z} \in [0, 1]$:

1. Compute the cdf $F_X(x)$ of the desired pdf $p_X(x)$
2. Determine the inverse transformation F^{-1}.
3. Sample uniformly distributed numbers (in $[0, 1]$), $\tilde{z}$.
4. Get the samples $\tilde{x} = F^{-1}(\tilde{z})$ from X.

The pdf of a Laplace distribution with location parameter μ (= mean), and scale parameter $b > 0$ (variance = $2b^2$) is given by

$$p_X(x) = \frac{1}{2b} \exp\left(-\frac{|x - \mu|}{b}\right).$$

Task:

(a) Following the procedure above, derive a formula to generate samples of a scalar random variable with a Laplacian distribution from uniformly distributed random numbers.

(b) Implement your procedure for verification and generate 500 samples for a Laplacian random variable X with a specific mean $\mu = 1$ and scale parameter $b = 2$. Plot a density estimate (e.g. normalized histogram) for these samples overlayed with the pdf $p_X(x)$ from above.

Exercise 4.2: Density Transformations (6 points)

Background: Let $f(\underline{\mathbf{x}}) = f(x_1, \ldots, x_N)$ be a function of $\underline{\mathbf{x}} \in \Omega \subset \mathbb{R}^N$ and assume we make a change of variables to a new coordinate system by a mapping $\underline{\mathbf{u}} = \underline{\mathbf{u}}(\underline{\mathbf{x}}) = (u_1(\underline{\mathbf{x}}), \ldots, u_N(\underline{\mathbf{x}}))$, whose inverse mapping $\underline{\mathbf{x}} = \underline{\mathbf{x}}(\underline{\mathbf{u}}) = (x_1(\underline{\mathbf{u}}), \ldots, x_N(\underline{\mathbf{u}}))$ exists and is differentiable. As we change the coordinate system, the integral over f changes according to

$$\int_\Omega f(\underline{\mathbf{x}})\mathbf{d}\underline{\mathbf{x}} = \int_{u(\Omega)} f(\underline{\mathbf{x}}(\underline{\mathbf{u}})) \left|\det \frac{\partial \underline{\mathbf{x}}(\underline{\mathbf{u}})}{\partial \underline{\mathbf{u}}}\right| \mathbf{d}\underline{\mathbf{u}} = \int_{u(\Omega)} f(\underline{\mathbf{x}}(\underline{\mathbf{u}})) \frac{1}{\left|\det \frac{\partial \underline{\mathbf{u}}(\underline{\mathbf{x}})}{\partial \underline{\mathbf{x}}}\right|} \mathbf{d}\underline{\mathbf{u}},$$

where $\frac{\partial \underline{\mathbf{x}}(\underline{\mathbf{u}})}{\partial \underline{\mathbf{u}}}$ is the *Jacobi* matrix, which is the matrix of the partial derivatives

$$\frac{\partial \underline{\mathbf{x}}(\underline{\mathbf{u}})}{\partial \underline{\mathbf{u}}} = \begin{pmatrix} \frac{\partial x_1(\underline{\mathbf{u}})}{\partial u_1} & \cdots & \frac{\partial x_1(\underline{\mathbf{u}})}{\partial u_N} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_N(\underline{\mathbf{u}})}{\partial u_1} & \cdots & \frac{\partial x_N(\underline{\mathbf{u}})}{\partial u_N} \end{pmatrix}$$

and whose determinant $\det \frac{\partial \underline{\mathbf{x}}(\underline{\mathbf{u}})}{\partial \underline{\mathbf{u}}} = \left(\det \frac{\partial \underline{\mathbf{u}}(\underline{\mathbf{x}})}{\partial \underline{\mathbf{x}}}\right)^{-1}$ is called the *Jacobi determinant* (also *functional determinant*).

Remark: The absolute value of the Jacobi determinant at a point $\underline{\mathbf{u}}_0$ corresponds to the factor by which the function $\underline{\mathbf{x}}(\underline{\mathbf{u}})$ expands or shrinks volumes near $\underline{\mathbf{u}}_0$.

Implication: If $f(\underline{\mathbf{x}})$ is the probability density function (pdf) of the N-dimensional random vector $\underline{\mathbf{X}}$ then $f(\underline{\mathbf{x}}(\underline{\mathbf{u}})) \left|\det \frac{\partial \underline{\mathbf{x}}(\underline{\mathbf{u}})}{\partial \underline{\mathbf{u}}}\right|$ is the pdf of the random vector $\underline{\mathbf{u}}(\underline{\mathbf{X}})$.

Task:

(a) (1 point) Consider the density of a random variable X to be $p_X(x) = e^{-x},\ x \geq 0$. For the change of variables $u = u(x) = e^{-x}$ calculate the density $p_{u(X)}(u)$ of the random variable $u(X)$.

(b) (4 points) Consider two independent and uniformly in the interval $[0, 1]$ distributed random variables $(X_1, X_2)^\top =: \underline{\mathbf{X}}$. The pdf is given by $p_{\underline{\mathbf{X}}}(x_1, x_2) = 1$ in $[0, 1]^2$ and zero otherwise.
Consider the variable transformation $\underline{\mathbf{u}} = \underline{\mathbf{u}}(\underline{\mathbf{x}})$ with
$u_1(\underline{\mathbf{x}}) = \sqrt{-2 \ln x_1}\, \cos(2\pi x_2)$ and
$u_2(\underline{\mathbf{x}}) = \sqrt{-2 \ln x_1}\, \sin(2\pi x_2)$.
Show that $\underline{\mathbf{u}}(\underline{\mathbf{X}})$ corresponds to two independent unit-variance zero-mean normally distributed random variables.
Remark:
This procedure to produce Gaussian samples from uniform random numbers is called the *Box-Muller method*.

(c) (1 point) **Outline** how to generalize the last result to N dimensions[1], i.e., how to generate samples from a multidimensional Gaussian distribution with mean vector $\underline{\mu}$ and covariance matrix $\underline{\mathbf{\Sigma}}$ just from uniformly distributed random numbers in $[0, 1]^N$. Use the following:

- Any symmetric positive semidefinite matrix (such as the covariance matrix $\underline{\mathbf{\Sigma}}$) has a Cholesky decomposition $\underline{\mathbf{\Sigma}} = \underline{\mathbf{L}}\,\underline{\mathbf{L}}^\top$ (and that can be easily computed numerically).
- If $\underline{\mathbf{L}}$ is a constant matrix and $\underline{\mathbf{X}}$ a random vector then $\mathrm{Cov}(\underline{\mathbf{L}}\,\underline{\mathbf{X}}) = \underline{\mathbf{L}}\,\mathrm{Cov}(\underline{\mathbf{X}})\,\underline{\mathbf{L}}^\top$.
- The covariance matrix of independent unit-variance Gaussian variables is identity, i.e., $\mathrm{Cov}(\underline{\mathbf{X}}) = \underline{\mathbf{I}}$.

Confirm that the above properties hold for your solution (a detailed proof is not necessary).

Total 10 points.

[1] It might help to think of N as even.