Machine Intelligence 2 **Exercise Sheet 4**
SoSe 2022, Obermayer/Kashef/Strömsdörfer due: 2022-06-02 23:55

Density Transformations & random number generation

Exercise T4.1: From Pseudo-random number generators to density transformation (tutorial)

(a) How do you build a *pseudo-random number generator (PRNG)*?

(b) What is the *Inverse CDF* method?

(c) How do you transform densities while conserving probabilities?

Exercise H4.1: The Inverse CDF method (homework, 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 $\tilde{z}$ in $[0, 1]$.
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 $b = 2$.

(c) Plot a density estimate (e.g. normalized histogram) for these samples overlayed with the pdf $p_X(x)$ from above.

Exercise H4.2: Density Transformations (homework, 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 $\underline{\mathbf{g}} : \underline{\mathbf{x}} \mapsto \underline{\mathbf{g}}(\underline{\mathbf{x}})$ be a mapping with which we change the variables $\underline{\mathbf{x}}$ to a new coordinate system with coordinates $\underline{\mathbf{u}} = (g_1(\underline{\mathbf{x}}), \ldots, g_N(\underline{\mathbf{x}})) = (u_1, \ldots, u_N)^\top$, whose inverse mapping $\underline{\mathbf{g}}^{-1}(\underline{\mathbf{u}}) = \underline{\mathbf{x}}$ exists and is differentiable.
As we change the coordinate system, the integral over $f(\cdot)$ changes according to

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

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

$$\frac{\partial \underline{\mathbf{x}}}{\partial \underline{\mathbf{u}}} = \frac{\partial \underline{\mathbf{g}}^{-1}(\underline{\mathbf{u}})}{\partial \underline{\mathbf{u}}} = \begin{pmatrix} \frac{\partial x_1}{\partial u_1} & \cdots & \frac{\partial x_1}{\partial u_N} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_N}{\partial u_1} & \cdots & \frac{\partial x_N}{\partial u_N} \end{pmatrix}$$

and whose determinant $\det \frac{\partial \underline{\mathbf{x}}}{\partial \underline{\mathbf{u}}} = \left(\det \frac{\partial \underline{\mathbf{u}}}{\partial \underline{\mathbf{x}}} \right)^{-1}$ 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 inverse mapping function $\underline{\mathbf{g}}^{-1}(\underline{\mathbf{u}})$ 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 $(X_1, \ldots, X_N)^\top$ then $f(\underline{\mathbf{g}}^{-1}(\underline{\mathbf{u}})) \left| \det \frac{\partial \underline{\mathbf{x}}}{\partial \underline{\mathbf{u}}} \right|$ is the pdf of the random vector $(U_1, \ldots, U_N)^\top$.

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 $u = e^{-x}$ calculate the density $p_U(u)$ of the random variable U.

(b) (4 points) Consider two independent and uniformly in the interval $[0, 1]$ distributed random variables $(X_1, X_2)^\top$. The pdf is given by $p_{X_1,X_2}(x_1, x_2) = 1$ in $[0, 1]^2$ and zero otherwise.
Consider the variable transformation $(X_1, X_2)^\top \to (U_1, U_2)$ with
$u_1 = \sqrt{-2 \ln x_1} \cos(2\pi x_2)$ and
$u_2 = \sqrt{-2 \ln x_1} \sin(2\pi x_2)$.
Show that $(U_1, U_2)^\top$ 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 previous 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 <u>not</u> necessary).

Total 10 points.

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