Transformation of R.V. and Multivariate Gaussian

Your Name(?)

Contents

1 Introduction

In this article, we will study about the following topics of statistics:

- Transformation of random variables
- Multivariate Gaussian random variable

2 Transformation of Random Variable

Given any continuous r.v. X with PDF $P_X(x)$ and given any function $g(X)$(defined on range of X) we intend to find PDF associated with the r.v. $Y = g(X)$.
For simplicity, let's assume $g(.)$ is monotonic increasing.
Then by probability mass conservation,

$$P(a < X < b) = P(g(a) < Y < g(b)) = \int_{g(a)}^{g(b)} Q(y)dy$$

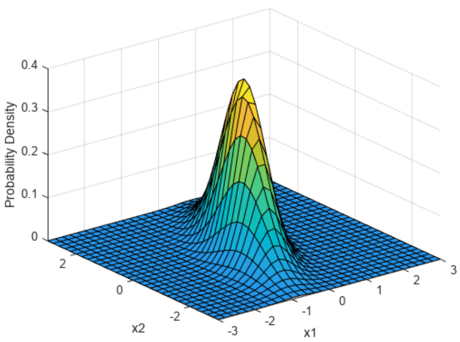
Using $y = g(x)$, we get the below relation upon simplification

$$Q(y) = P(g^{-1}(y))\frac{d(g^{-1}(y))}{dy}$$

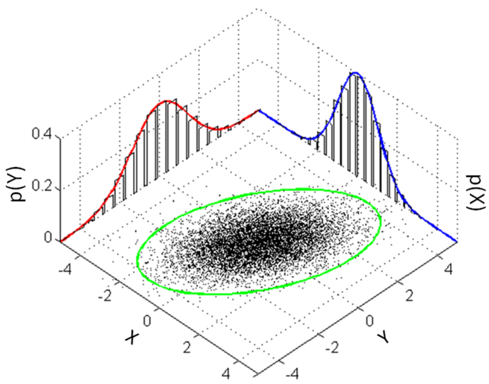
To handle monotonically decreasing $g(.)$ as well[1],

$$Q(y) = \begin{cases} +P(g^{-1}(y))\frac{d(g^{-1}(y))}{dy} & \text{for } g(.) \text{ monotonically increasing} \\ -P(g^{-1}(y))\frac{d(g^{-1}(y))}{dy} & \text{for } g(.) \text{ monotonically decreasing} \end{cases} \tag{1}$$

For more information, refer [1]



(a) Example 1



(b) Example 2

3 Multi-variate Gaussian Disribution

3.1 Definition

Let X be a vector of random variables of dimension D.
A r.v. X has a joint PDF as multi-variate Gaussian distribution $\exists$ finite i.i.d. standard Gaussian

[1]we could have used modulus operator but I wanted things to look more complicated

r.v. $W_1, W_2, \ldots W_N$ with $N > D$ such that

$$X = AW + \mu$$

Refer fig[1a] and fig[1b] for visual examples. This has many applications in machine learning, refer [3] and [2].

3.2 A is diagonal

In this case, the X_i are independent. The standard deviation of distribution of X_i is A_{ii}.

3.3 A is non-singular square matrix

Let's take $\mu = 0$ for simplicity.
Similar to univariate case, where scaling was determined by $\left|\frac{d(g^{-1}(y))}{dy}\right|$, the scaling for multi-variate case is determined by determinant of matrix of derivatives, Jacobian matrix.
Also, $W = A^{-1}X$, which is a linear transformation of vector X. A^{-1} maps a hypercube to parallelepiped. If the vectors describing the hypercube are along cardinal axis, then the parallelepiped is described by vectors which are columns of A^{-1}.
We intend to find the volume of the parallelepiped formed due to this transformation.
Claim: The volume of parallelepiped described by column vectors of matrix A^{-1} is given by $det(A^{-1})$
Proof: Addition of any scaled column of a matrix M to another column does not change the determinant.
Therefore by Gram-Schmidt orthogonalization process the columns of A^{-1} can be constructed to be orthogonal to each other, without changing the determinant. Then multiplying by an orthogonal matrix would rotate the orthogonal vectors(to align them with cardinal axis), and this operation would not change the determinant as well. Now the result matrix is diagonal square matrix and the volume of the parallelepiped described by the column vectors is given by product of diagonal elements.

From the above result, an infinitesimal volume δ^D after transformation becomes $\delta^D \cdot det(A^{-1})$.

Let $C = A \cdot A^T$. Then $det(A) = \sqrt{det(C)}$. The above expression can we rewritten as

$$P(X) = \frac{1}{(2\pi)^{D/2}} \cdot \frac{1}{\sqrt{det(C)}} \cdot exp(0.5 \cdot X^T \cdot C^{-1} \cdot X) \tag{2}$$

Sample Values of bivariate normal distribution		
x	y	f(x,y)
0	0	1.6
0	1	0.096
$\sqrt{2}$	$\sqrt{2}$	0.02

References

[1] URL: https://bookdown.org/pkaldunn/DistTheory/Transformations.html.

[2] URL: https://distill.pub/2019/visual-exploration-gaussian-processes/.

[3] Carl Edward Rasmussen. *Gaussian Processes for Machine Learning*. The MIT Press.