10 Multivariate Gaussian

Generalizes a univariate Gaussian.

Consider a vector random variable $X = [X_1, X_2, \cdots, X_D]^\top$. Nothing but a joint RV with d RVs. Represent as a $d \times 1$ vector.

Definition: The RV X has a multivariate (jointly) Gaussian PDF if $\exists$ a finite set of i.i.d. univariate standard-normal RVs $W_1, \cdots, W_N$ (with $D \le N$) such that each X_d can be expressed as $X_d = \mu_d + \sum_n A_{dn} W_n$ (i.e., $X = AW + \mu$).

Example 1 (Zero Mean + Isotropic): The case of independent standard-normal RVs $W_1, \cdots, W_D$ with $A = I_{D \times D}$ and $\mu = 0$, i.e. $X = W$
Then, the Gaussian PDF is $p(w) = \prod_d p(w_d) = \frac{1}{(2\pi)^{d/2}} \exp(-0.5 w^\top w)$

Formula for the PDF using Transformation of RVs

Example 2 (Zero Mean + Anisotropic): What is the PDF $q(X)$ for arbitrary non-singular SQUARE A and $\mu = 0$?

– Recall: Given PDF $p(w)$ and the transformation $X = g(W)$, the PDF $q(x) = p(g^{-1}(x)) \left| \frac{d}{dx} g^{-1}(x) \right|$
– In our case, $X = g(W) = AW$
– The inverse transformation $g^{-1}(X) = W = A^{-1}X$
– In the univariate case, we wanted the *magnitude* of the *derivative* of the inverse transformation: $\frac{\partial}{\partial y} g^{-1}(y)$
– In the multivariate case, we want the *volume* captured by the columns of the *Jacobian* of the inverse transformation: vol$(\frac{d}{dX} A^{-1}X) =$ vol$(A^{-1}) = \det(A^{-1}) = 1/\det(A)$

** Geometric intuition for vol$(A^{-1}) = 1/\det(A)$ (Note: determinant is defined only for a square matrix)
** Observe that the linear transformation A maps an infinitesimal hyper-cube $\delta \times \cdots \times \delta$ to an infinitesimal hyper-parallelepiped. If the axes of the hyper-cube were the cardinal axes, then the axes of the hyper-parallelepiped are the columns of A !!
** The volume of the hyper-parallelepiped is $\delta^d \det(A)$. In 3D, the volume can also be written as the scalar triple product $a_1 \cdot (a_2 \times a_3)$ where a_i is the i-th column of A

** Why is the volume equal to the determinant ?
The following is some intuition (not a proof; a separate inductive proof exists):
Adding multiples of one column to another:
1) keeps the determinant unchanged because the determinant function is multi-linear.
2) corresponds to a skew translation of the parallelepiped, which does not affect its volume.
Using Gram-Schmidt orthogonalization, we can transform matrix A to an orthogonal matrix A_{ortho} (NOT orthonormal; that would have determinant 1). This doesn't change the determinant or the volume.
We can rotate A_{ortho} to make it to diagonal form. Rotation doesn't change the determinant or the volume.
For this diagonal matrix, the determinant (= product of diagonal entries) equals the volume of a "rectangle" (= product of side lengths).

** Thus, $|dw| = \delta^d \implies |dx| = \delta^d \det(A)$
** Thus, $\frac{|dw|}{|dx|} = 1/\det(A)$

– Finally, the transformation of variables gives :
$q(X) = p(A^{-1}X) \frac{1}{\det(A)} = \frac{1}{(2\pi)^{d/2} \det(A)} \exp(-0.5 X^\top (A^{-1})^\top A^{-1} X)$
– Simplify: Let $C := AA^\top$. Then, $C^{-1} = A^{-T} A^{-1}$ and $\det(C) = \det(A)\det(A^\top) = (\det(A))^2$
– So, the multivariate-Gaussian PDF $q(X) = \frac{1}{(2\pi)^{d/2} |C|^{0.5}} \exp(-0.5 X^\top C^{-1} X)$, where C has a special name.

Property: The mean of $X = AW$ is zero

Proof: $E[AW] = AE[W] = A \cdot 0 = 0$
Note: $E[X] = [E[X_1], E[X_2], ..., E[X_d]]^\top$ (recall: all X_i share the same probability space).

Example 3 (Nonzero mean + Anisotropic): If X is multivariate Gaussian with zero mean, then $Y = X + \mu$ is multivariate Gaussian with PDF $p(y) = \frac{1}{(2\pi)^{d/2}|C|^{0.5}} \exp(-0.5(y-\mu)^\top C^{-1}(y-\mu))$
Proof:
Y is multivariate Gaussian because Y can be expressed as $AW + \mu$, where W_n is i.i.d. standard normal.
PDF $p(y) = \frac{1}{(2\pi)^{d/2}|C|^{0.5}} \exp(-0.5(y-\mu)^\top C^{-1}(y-\mu))$ because of the transformation of the variables $Y = X + \mu$

Property: The *mean vector* of $X = AW + \mu$ is μ.
Proof: $E[AW + \mu] = AE[W] + \mu = \mu$

Property: If Y is multivariate Gaussian, then $Z = BY + c$ is multivariate Gaussian, where A, BB are square invertible.
Proof: Because Y is multivariate Gaussian, $Y = AW + \mu$. Thus, $Z = B(AW + \mu) + c = (BA)W + (B\mu + c)$

Covariance Matrix

For any multivariate RV X, the definition of covariance is $C := E[(X - E[X])(X - E[X])^\top]$. This leads to a matrix C, where the outer-product structure implies that $C_{ij} = E[(X_i - E[X_i])(X_j - E[X_j])]$ which equals Cov(X_i, X_j).

Cov$(W) = E[WW^\top] = I$ because:
(i) Cov$(W_i, W_i) = 1$ and
(ii) Cov$(W_i, W_{j \neq i}) = 0$ because of independence of W_i and W_j

Cov$(X) = E[(X - E[X])(X - E[X])^\top] = E[(AW)(AW)^\top] = E[AWW^\top A^\top] = AE[WW^\top]A^\top = AA^\top$
Thus, the RV $X = AW + \mu$ has covariance $C = AA^\top$, where $C_{ij} =$ Cov(X_i, X_j).

More properties of C:

1) $C = E[XX^\top] - E[X](E[X])^\top$
Proof: Expand the terms in the definition.

2) C is symmetric
Proof: $C_{ij} =$ Cov$(X_i, X_j) =$ Cov$(X_j, X_i) = C_{ji}$

3) C is positive semi-definite (PSD)
Proof: For any $d \times 1$ non-zero vector a, $a^\top Ca = E[a^\top(X - E[X])(X - E[X])^\top a] = E[(f(X))^\top f(X)] \geq 0$ that is the variance of a scalar RV $f(X) = (X - E[X])^\top a$

Marginal PDFs

Property: The 1D marginal PDF of the multivariate Gaussian Z, for any single variable, is (univariate) Gaussian.

Proof: From the definition, we know that:
(1) $X_d = \mu_d + \sum_n A_{dn} W_n$, where W_n are i.i.d. standard Normal,
(2) the transformations of scaling and translation on a univariate Gaussian RV leads to another univariate Gaussian RV,