* Mathematical Statistics:-

- Application of mathematics to statistics for data analysis and interpretation.

- We would be dealing mostly with continuous RV.

* Transformation of Random Variables -

Let X be an RV with pdf $p(x)$. Let $g(x)$ be a strictly ↑ function

Now define $Y = g(X)$. we wish to find this pdf.

- Principle of Probability mass conservation

In simple terms; $\boxed{P(a \le X \le b) = P(g(a) \le Y \le g(b))}$

$$\Rightarrow \int_a^b p(x)dx = \int_{g(a)}^{g(b)} q(y)dy$$

$$\Rightarrow \int_{g(a)}^{g(b)} p(g^{-1}(y))\left[\frac{d}{dy} g^{-1}(y)\right]dy = \int_{g(a)}^{g(b)} q(y)dy \qquad (\text{put } x = g^{-1}(y))$$

This holds for all intervals $\Rightarrow$ $\boxed{q(y) = p(g^{-1}(y)) \cdot \left|\frac{d}{dy} g^{-1}(y)\right|}$

→ modulus present to take care of ↓ g

- If g wasn't strictly monotonic;- split it into piecewise monotonic and apply conservation of probability mass.

- If $g(x)$ is of the form $ax+b$ ⇒ there is just Linear Scaling happening.

Multivariate Gaussian :-

Consider a vector RV $X = [x_1, .., x_D]$ of length 'D'.

Definition X has a multivariate joint gaussian pdf if $\exists$ finite set of i.i.d univariate standard normal RVs $W_1, ..., W_n$ $(n \geq D)$ such that each x_d can be represented as

$$x_d = \mu_d + \sum_n A_{nd} W_n \; .$$

Example :- Zero mean + Isotropic/Spherical gaussian

Defined as $\mu = 0$, $A = I_{D \times D}$ $\Rightarrow X = W$ (all are independent!)
$(D = N)$

$$\Rightarrow p(X) = p(W) = \frac{1}{(2\pi)^{\frac{N}{2}}} e^{-\frac{1}{2}(\sum \omega_i^2)} = \frac{1}{(2\pi)^{N/2}} e^{-\frac{1}{2}(\omega \times \omega^T)}$$

Definition * Level set — Essentially a contour; $L_c(f) = \{(x_1, .., x_n) \mid f(x_1, .., x_n) = c\}$

- Therefore, the level set of above gaussian is when $\sum \omega_i^2 = K \Rightarrow$ in 3D we get a sphere — hence the name!

- We shall get to the most general case by making our analysis more "wide".

Generalization 1 - 'A' is Singular and Diagonal

- A is singular $\Rightarrow$ no diagonal element is zero.

$X = \mu + AW \Rightarrow x_i = \mu_i + A_{ii} \omega_i \Rightarrow$ As all ω_i are independent with $\mu = 0, \sigma = 1$

x_i are independant with $\mu = \mu_i, \sigma^2 = A_{ii}^2$

$$\Rightarrow P(x) = \frac{1}{(2\pi)^{N/2}} \cdot \frac{1}{(\Pi A_{ii})} \cdot \exp\left[-0.5 \sum \left(\frac{x_i - \mu_i}{A_{ii}}\right)^2\right]$$

This $P(x)$ is a Hyper-Ellipsoid with mean at μ; Axes aligned with cardinal axes.

- In two dimensions, we get an ellipse centered at (μ_1, μ_2); with major/minor axes' lengths being $2A_{11}/2A_{22}$

Generalization 2 - 'A' is non-Singular, $\mu=0$

$X = AW \Rightarrow X = g(W) \Rightarrow$ transformation of variables!

$g^{-1}(W) = A^{-1}W \Rightarrow$ in 2D, this depends on the magnitude of derivative of g^{-1}.

We measured how g^{-1} scaled the values.

In general, this would depend on the determinant of the Jacobian of g^{-1}.

and in 3D, g^{-1} refers to how volumes are scaled between the "axes",

<u>Reminder</u> Jacobian $\Rightarrow \nabla f = \dfrac{\partial(f_1, \dots, f_m)}{\partial(x_1, \dots, x_n)} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$

In 2D :- $dx \rightarrow dy \Rightarrow$ Linear to Linear

3D :- $dx \cdot dy \rightarrow dx' \cdot dy' \Rightarrow$ Cube to parallelopiped

nD :- $dx_{\dots} \rightarrow dx'_{\dots} \Rightarrow$ Hyper Cube to hyper parallelopiped.

- To calculate the value of $p(X)$, we would need to find $|\nabla A^{-1}|$. However, understand that this is just a scaling factor, between an infinite-simal Hyper-Parallelopiped and an infinitesimal Hypercube.

- Without proof, we state that the volume of a Hyper-parallelopiped is determinant of of the sides of the hyper-parallelopiped.

(prove by gram-Schmidt and rotate to form Hyper-Cube)

- Therefore; $p(X) = p(A^{-1}W) \cdot \text{Scaling} = p(A^{-1}W) \cdot \dfrac{1}{\det A} = \dfrac{1}{(2\pi)^{N/2}\,|\det A|} \cdot \exp\left(-0.5\, X^T A^{-1^T} A^{-1} X\right)$

- For simplicity, take $C = AA^T \rightarrow p(x) = \dfrac{1}{(2\pi)^{\frac{N}{2}} |C|^{1/2}} \exp\left[-0.5\, x^T C^{-1} x\right]$

$|C| = |A|^2$

* We can easily extend this to Singular A, non-zero μ

Let $Y = X + \mu \Rightarrow$ $\boxed{p(Y) = \dfrac{1}{(2\pi)^{D/2} |\det C^{1/2}|} \exp\left[-0.5\,(Y-\mu)^T C^{-1} (Y-\mu)\right]}$

Lemma If Y is a multi-variate gaussian; $Z = AY + c$ is also a multi-variate gaussian.

Definition :- Mean of a multi-variate gaussian $X = AW + \mu \Rightarrow \mu$,

Covariance Matrix of $X = AW + \mu$ is given by $C = AA^T$

Properties :- $C = E[xx^T] - E[x]\,E[x]^T$

C is symmetric (obviously!)

C is positive-Semi Definite matrix.

- C is said to be positive Semidefinite iff $\forall$ column vectors V,

  $V^T C V \geq 0$.

- If $V^T C V > 0 \rightarrow C$ is positive definite.

- Now that we know the joint pdf of X; we are interested in its Level Sets.

We first define a few terms.

Definition Orthogonal Matrix

- 'A' when $AA^T = A^TA$ = Identity matrix
- If $|A| = +1$; it is called as a Rotation matrix, models rotation

  $|A| = -1$; called as Reflection matrix, models reflection + rotation. They are also Symmetric.

- Lets find the Level sets for the multivariate gaussian. We start from special cases and build upto general cases.

Case 1 – $\mu = 0$; A is orthogonal

$$\Rightarrow X = AW \Rightarrow p(x) = \frac{1}{(2\pi)^{D/2}} \exp(-0.5\, x^T x) \Rightarrow \text{same as } W!$$

- This is also a "zero mean isotropic multivariate gaussian". The pdf is unchanged, because A can either rotate/reflect $p(W)$. But because $p(W)$ is spherical, it remains unchanged.

Case 2 – $\mu = 0$, A is Square diagonal with +ve entries

$$X = AW \Rightarrow p(x) = \frac{1}{(2\pi)^{D/2}} \cdot \frac{1}{|\det A|} \exp(-0.5\, x^T A^{-2} x) \quad \text{from the formula}$$

- Graphically, the value of $X_i = A_{ii} W_i$; meaning each dimension is amplified by a factor of A_{ii}. Therefore, the pdf is Zero mean anisotropic in nature.

- We can extend this case further. Suppose $A = RS$ where S is diagonal and R is orthogonal in nature.

- Without solving, we can see that $p(X)$ is just rotating the $p(X')$ where $X' = SW$ by R! Hence, it is still Zero mean Anisotropic in nature.

- However, if $S - kI \rightarrow p(x')$ is circular $\Rightarrow p(x) = p(x')$!

Case 3 :- General case

- We've already stated that $C = AA^T$ is symmetric, and positive semi-definite in nature. However, we shall look at the cases where C is positive definite.

(when C is semi-def; C is not invertible, causing problems)

Recall :- $Av = \lambda v \Rightarrow$ for a column vector v; λ is called the corresponding eigen value.

- This is possible iff A is diagonizable $\Rightarrow$ it is similar to a diagonal matrix

$\Rightarrow \exists P$ which is invertible and diagonal D

s.t $P^{-1}AP = D$

- If A is diagonizable, but is invertible, it is then called as a "Defective Matrix".

Theorem :- Every real symmetric matrix is diagonizable by an orthogonal matrix. It has N real Eigen values with N-linearly independant Eigen vectors — Spectral Theorem

- Applies for C; as it is real & symmetric.

- In mathematical terms;-

If C is real Symmetric $\rightarrow \exists V$, $V^TV = VV^T = I$ and V^TCV = Diagonal matrix

also; N-Eigen values, N-LI-Eigen Vectors.

Extending Spectral $\Rightarrow$ If C is a positive definite matrix, all the Eigen values are positive.

- Returning to the original question of finding Level sets:-

$$p(x) = \frac{1}{(2\pi)^{D/2}\ |C|^{1/2}} \exp\left(-0.5\,(x-\mu)^T C^{-1}(x-\mu)\right)$$

- From the spectral theorem, $C = V^T D V \Rightarrow C^{-1} = V^T D^{-1} V \rightarrow C^{-1}$ is SPD

- Every Level set has $(x-\mu)^T C^{-1}(x-\mu) = \text{Constant} \geq 0$ as C^{-1} is also SPD

$$\Rightarrow (x-\mu)^T\ V^T\ D^{-1} V(x-\mu) \Rightarrow \left[V(x-\mu)\right]^T D^{-1}\left[V(x-\mu)\right] = a \geq 0$$

V is orthogonal $\Rightarrow V(x-\mu) = X' - \mu'$ by changing the axes

$$\Rightarrow (x'-\mu')^T D^{-1}(x'-\mu') = a$$

- The center is at μ' in the new rotated system.

In the new system, the half-lengths are root of diagonal Elements of D^{-1}

Define A as diagonal square with $A_{ii} = (D_{ii})^{1/2}$ and write pdf to get this pdf again!