The Multivariate Normal distribution

1) We say that a random vector has the standard normal distribution in $\mathbb{R}^n$ if the coordinates are independent standard normal RVs.

   \[ \mathbf{g} = (g_1, \ldots, g_n)^T, \quad g_i \sim N(0,1), \]

   \[ \mathbb{E}\mathbf{g} = (\mathbb{E}g_1, \ldots, \mathbb{E}g_n)^T = (0,\ldots,0)^T \]

   \[ \text{cov}(\mathbf{g})_{ij} = \mathbb{E}((g_i - \mathbb{E}g_i)(g_j - \mathbb{E}g_j)) = \mathbb{E}g_i g_j = \begin{cases} \frac{\mathbb{E}g_i^2}{\mathbb{E}g_i^2} & \text{if } i=j \\ 0 & \text{if } i \neq j \end{cases} \]

   So \( \text{cov}(\mathbf{g}) = \mathbf{I}_n \), the identity matrix.

   We write \( \mathbf{g} \sim N(0,\mathbf{I}_n) \).

2) General normal random vectors.

   Say \( X \) is a normal RV. \( \mathbf{g} \sim N(\mu,\Sigma) \).

   Then \( \frac{X - \mu}{\sqrt{\Sigma}} \sim N(0,1) \).

   If we denote \( \mathbf{Z} = \frac{X - \mu}{\sqrt{\Sigma}} \), then

   \[ X = \mu + \sqrt{\Sigma} \mathbf{Z} \]

   i.e. any normal RV is a linear transformation of a standard normal.

   We can use this to generalize the notion of a standard normal RV in $\mathbb{R}^n$. 
We will say a random vector $X \in \mathbb{R}^n$
has the (multi-variate) normal distribution if
\[ Z \sim \mathcal{N}(0, I_n) \text{ and } \Sigma \text{ a symmetric matrix.} \]
\[ X = \mu + QZ. \]

**Rule:** The book is not consistent whether vectors in $\mathbb{R}^n$
might be columns $(n \times 1)$ or rows $(1 \times n)$, but you can
figure it out from the context.

**Rule:** The book’s def is a bit different, it requires
the matrix $Q$ to be invertible, but it doesn’t have
to be. What the book defines should be called
the non-degenerate normal distribution.

\[ E(X) = \mu, \quad \text{Cov}(X) = E((X-\bar{X})(X-\bar{X})^T) \]
\[ = E(QZ(QZ)^T) = QE(ZZ^T)Q^T \]
\[ = Q\Sigma Q^T = : \Sigma \]
\[ Z \sim \mathcal{N}(0, I_n) \]

So $X$ has mean $\mu$ & cov $\Sigma$. Write $X \sim \mathcal{N}(\mu, \Sigma)$.

**Density**
\[ g \sim \mathcal{N}(0, I_n). \]
Since the coordinates of $g$ are independent $\mathcal{N}(0, 1)$’s,
g has density which is given by the product of the densities of the components:
\[ f(x) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi \sigma_{x_{ij}}^2}} e^{-\frac{x_i^2}{2\sigma_{x_{ij}}^2}} = \frac{1}{(2\pi)^{n/2}} e^{-1/2\Sigma_{x_{ij}}^2}. \]
By a change of variables, one can show that if 
\[ \Sigma \text{ is invertible, then } X \text{ has density} \]
\[ f_X(x) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} e^{-(x-\mu)^T \Sigma^{-1} (x-\mu)/2}. \]

If \( \Sigma \) is not invertible, \( X \) does not have density.

**Remark:** Recall that if \( X \sim N(0, I) \), then \( M_X(t^T) = e^{t^T I} \).

Check that if \( X \sim N(\mu, \Sigma) \), then \( M_X(t^T) = e^{t^T \mu + \frac{1}{2} t^T \Sigma t} \).

Check that if \( X \sim N(\mu, \Sigma) \), then \( M_X(t^T) = E(e^{t^T X}) = e^{t^T \mu + \frac{1}{2} t^T \Sigma t} \).

Sometimes this is used as the definition of a multivariate normal. Note that
\[ M_X(t^T) = e^{t^T \mu + \frac{1}{2} t^T \Sigma t} \text{ is defined even when } \Sigma \text{ is not invertible, so } X \text{ has no density.} \]

**Remark:** If \( X \sim N(\mu, \Sigma) \) and \( t \in \mathbb{R}^n \) is arbitrary, then \( \langle X, t \rangle \) is actually normal.

In fact, any linear combination of the coordinates of a multivariate normal is normal.

In fact, the converse is also true. \( X \) is multivariate normal if and only if \( \langle X, t \rangle \) is normal \( t \in \mathbb{R}^n \).

**Remark:** This would not be true if we only restricted normals to those invertible covariance matrices \( \Sigma \).
In particular, it follows that if $X \sim N(M, \Sigma)$

then every component of $X$ is normal.

The converse is not true.

**Exercise:** Construct RVs $X, Y$ such they are both normal, but $(X, Y)$ is not a multivariate normal.

Q: What does a multidimensional normal "look like"?

Let $X \sim N(0, I_n)$.

1) Since $E(X|Z) = Cov(X|Z) = 0$, we get $X$ is isotropic.

2) 

If $n = 1$, 

If $n = 2$

What if $n$ is large?

Recall that $X$ has density

$$f(x) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-x_i^2/2} = \frac{1}{(2\pi)^{n/2}} e^{-\|x\|^2/2}.$$ 

Notice that the density only depends on the length $\|x\|^2$ and not the direction of $x$, so the density is rotation invariant: if $U$ is an orthogonal matrix (i.e. multiplying a vector by $U$ simply rotates it), we have
\[ P(Y \in A) = P(y \in U^{-1}A) = \int_{U^{-1}A} \frac{1}{(2\pi)^{\frac{m}{2}}} e^{-\frac{1}{2} \|y\|^2} \, dy \]

Change of var., \[ y = Ux \quad \Rightarrow \quad \|y\|^2 = \|Ux\|^2 = \sum_{i} x_i^2 \quad \text{det}(U) = \frac{1}{\sqrt{\det(U)}} \]

It follows that \[ P(Y \in A) = P(G \in U^{-1}A) \quad \text{and} \quad \mathcal{U} \sim N(0, I_m) \quad \text{and} \quad U \mathcal{U} \]

So the distribution of \( Y \) is rotation invariant \( \mathcal{U} \).

What can we say about the length of \( X \)?

Last time you saw that \( \|X\|_2 \) concentrates around \( \sqrt{n} \):

\[ \|X\|_2 - \sqrt{n} \leq C \sqrt{n}, \quad \text{where} \quad K = \|X\|_2 \]

or equivalently that

\[ P \left( \|X\|_2 - \sqrt{n} > t \right) \leq 2 e^{-\frac{ct^2}{n}} \quad \text{for} \quad t > 0. \]

So with very high probability \( \|X\|_2 \) is within a constant of \( \sqrt{n} \).

Let's write \( X \) as

\[ X = \frac{X}{\|X\|_2} \cdot \frac{\|X\|_2}{\sqrt{n}}. \]

The distribution of \( X \) is rotationally invariant \( \mathcal{U} \) so is the distribution of \( \frac{X}{\|X\|_2} \), but \( \frac{X}{\|X\|_2} \) has length 1, so
\[
\text{\ exists distributed uniformly over the sphere of radius one.}
\]
On the other hand \(\forall x \in S_n\) is within a cost of \(\approx n\) of high prob., so \(\frac{\|x\|_2}{\sqrt{n}} \approx 1\), so \(x\) is roughly uniformly distributed over the sphere of radius \(1\), so \(x \approx \sqrt{n} \text{Unif}(S^{n-1})\).

\[
\text{i.e. } N(0, I_n) \approx \text{Unif}(\sqrt{n} S^{n-1}).
\]

Sub-gaussian distributions in higher dimensions

Recall that a random vector \(X \in \mathbb{R}^n\) is gaussian if \(\langle X, t \rangle\) is gaussian \(\forall t \in \mathbb{R}^n\). We can use this characterization of the gaussian distr. to define the notion of a multivariate sub-gaussian.

Definition: A random vector \(X \in \mathbb{R}^n\) is sub-gaussian if \(\forall t \in \mathbb{R}^n\), \(\langle X, t \rangle\) is sub-gaussian.

Q: Can we extend the notion of a sub-gaussian norm?

Could take \(\sup_{t \in \mathbb{R}^n} \|\langle X, t \rangle\|_{\ell_2}\); but that wouldn't be so restricted only to \(t \in S^{n-1}\).

The sub-gaussian norm of \(X\) is defined to be

\[
\|X\|_{\ell_2} := \sup_{t \in S^{n-1}} \|\langle X, t \rangle\|_{\ell_2}.
\]
Examples of sub-gaussian distributions

1) Suppose $X$ is an $n$-dimensional sub-gaussian random vector.

What can we say about its coordinates?

$\langle X, t \rangle$ is sub-gaussian $\forall t \in \mathbb{R}^n$. Taking $t = (0, -1, 0, 0)$ we see that all coordinates have to be sub-gaussian.

What if we know $X = (X_1, \ldots, X_n)$ if $X_1, \ldots, X_n$ are sub-gaussian.

Can we claim $X$ is sub-gaussian as well?

Let $t \in \mathbb{R}^n$. Need to check $\langle X, t \rangle$ is sub-g. $\langle X, t \rangle = X_1 t_1 + \cdots + X_n t_n$.

If $X_i$ is sub-g then $X_i t_i$ is also, thus $X_1 t_1 + \cdots + X_n t_n$ is also sub-gaussian. $\Rightarrow X$ is sub-g.

So $X$ is sub-gaussian if all its components are.

However, the sub-gaussian norm of $X$ might be much larger than that of its components.

That’s not the case if the components are indep.

Lemma: Let $X_1, \ldots, X_n$ be $n$ indep. mean-zero sub-gaussian RVs, then $X = (X_1, \ldots, X_n) \in \mathbb{R}^n$ is sub-gaussian $\&$

$\|X\|_{\psi_2} \leq C \max_{i \in [n]} \|X_i\|_{\psi_2}$ for some absolute constant $C$.

Proof: Let $t \in \mathbb{S}^{n-1}$.

shown previously w.r.t. sub-gaussian $\sum_{i \in [n]} (X_i t_i)^2 \leq C \sum_{i \in [n]} \|X_i\|_{\psi_2}^2$

$\leq C \max_{i \in [n]} \|X_i\|_{\psi_2} \sum_{i \in [n]} t_i^2 = C \max_{i \in [n]} \|X_i\|_{\psi_2}^2$.
2) $X \sim N(0, I_n)$. Of course, $A$ is sub-g. What is its sub-g norm? If $t \in \mathbb{S}^{n-1}$, then

$$
(X, t) = X_1 t_1 + \cdots + X_n t_n \sim N(0, t^T t) = N(0, 1)
$$

So

$$
\|X\|_2 = \|N(0, 1)\|_2 \leq C \text{ as } n \to \infty.
$$