The Multivariate Normal distribution

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

$$g = (g_1, \dots, g_n)^{tr}, \qquad g_1, \dots, g_n \text{ - indep}$$
$$g_i \sim N(0,1).$$

$$Eg = (Eg_1, \dots, Eg_n)^{tr} = (0, \dots, 0)^{tr}$$

$$(\mathrm{Cov}(g))_{ij} = E((g_i - Eg_i)(g_j - Eg_j)) = Eg_i g_j = \begin{cases} Eg_i^2 & i=j \\ Eg_i Eg_j & i \neq j \end{cases}$$
$$= \begin{cases} 1 & i=j \\ 0 & i \neq j \end{cases}$$

So $\mathrm{cov}(g) = I_n$ - the identity matrix.

We write $g \sim N(0, I_n)$.

2) General normal random vectors.

Say X is a normal RV. Eg $X \sim N(\mu, \sigma^2)$.

Then $\dfrac{X-\mu}{\sigma} \sim N(0,1)$.

If we denote $Z = \dfrac{X-\mu}{\sigma}$, then

$$X = \mu + \sigma Z$$

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 $\exists\, \mu \in \mathbb{R}^n$, $Z \sim N(0, I_n)$ & an $n\times n$ matrix Q s.t.

$$X = \mu + QZ.$$

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

Rmk: 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.

$$E(X) = \mu, \quad \mathrm{Cov}(X) = E\left((X - EX)(X-EX)^T\right)$$

$$= \underbrace{E\left(QZ(QZ)^T\right) = QE(ZZ^T)Q^T}_{\text{(multi) linearity of expectation}}$$

$$\underbrace{=}_{Z\sim N(0,I_n)} QIQ^T = QQ^T =: \Sigma$$

So X has mean μ & Cov. Σ. Write $X \sim N(\mu, \Sigma)$.

Density

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

Since the coordinates of g are indep $N(0,1)$'s, g has density, which is given by the product of the densities of the components:

$$f(x) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-x_i^2/2} = \frac{1}{(2\pi)^{n/2}} e^{-\|x\|_2^2/2}.$$

By a change of variables, can show that if Σ is invertible, then X has density

$$f_X(x) = \frac{1}{(2\pi)^{n/2} (\det \Sigma)^{1/2}} e^{-(x-\mu)^T \Sigma^{-1} (x-\mu)/2}.$$

If Σ is not invertible, X does not have a density.

Rmk: Recall that if $X \sim N(0,1)$, then $M_X(t) = e^{t^2/2}$

Check that if $X \sim N(\mu, \sigma^2)$, then $M_X(t) = e^{\mu t + \sigma^2 t^2/2}$.

Check that if $X \sim N(\mu, \Sigma)$, then $M_X(t) := E(e^{\langle X, t\rangle}) = e^{\langle \mu, t\rangle + \frac{1}{2} t^{tr} \Sigma t}$

Sometimes this is used as the defn of a multivariate normal. Note that

$$M_X(t) = e^{\langle \mu, t\rangle + \frac{1}{2} t^{tr} \Sigma t}$$ & is defined even when Σ is not invertible, so X has no density.

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

I.e. any linear combination of the coordinates of a multivariate normal is normal

In fact the converse is also true.

X is multivariate normal iff $\langle X, t\rangle$ is normal $\forall\, t \in \mathbb{R}^n$.

Rmk: This would not be true if we only restricted normals to those invertible covariance matrices Σ

Rmk: In particular, it follows that if $X \sim N(\mu, \Sigma)$ then every component of X is normal.
The converse is not true.

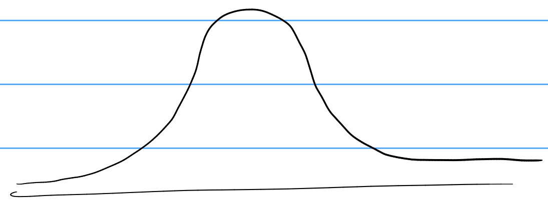
Exercise: Construct RVs X, Y s.t. they are both normal, but (X, Y) is not a multivariate normal.

Q: What does a multidimensional normal "look like"?
Let $X \sim N(0, I_n)$.

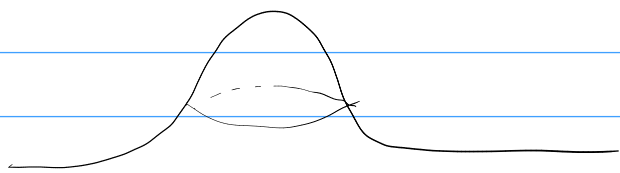
1) Since $\mathbb{E}(X) = 0$, $Cov(X) = I_n$, we get X is isotropic.

2)

If $n=1$,



If $n=2$



What if n is large?

Recall that X has density

$$f(x) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-x_i^2/2} = \frac{1}{(2\pi)^{n/2}} e^{-\|x\|_2^2/2}.$$

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

$$P(\underset{\subseteq \mathbb{R}^n}{Ug \in A}) = P(g \in U^{-1}A) = \int_{U^{-1}A} \frac{1}{(2\pi)^{n/2}} e^{-\|x\|_2^2/2} dx$$

ch. of. var. $y = Ux$

$$= \int_A \frac{1}{(2\pi)^{n/2}} e^{-\overbrace{\|U^{-1}y\|_2^2}^{\|y\|_2^2}/2} \underbrace{|\det U^{-1}|}_{1} dy = P(g \in A).$$

It follows that $P(g \in A) = P(g \in U^{-1}A)$ $\forall$ ognl U,

& $Ug \sim N(0, I_n)$ $\forall$ ognl U.

So the distr of X is rotation invariant $\forall n$.

What can we say about the length of X?

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

$$\big\| \|X\|_2 - \sqrt{n} \big\|_{\psi_2} \leq CK^2 \quad \text{, where } K = \|X_i\|_{\psi_2}$$

or, equivalently that

$$P(|\|X\|_2 - \sqrt{n}| \geq t) \leq 2e^{-\frac{ct^2}{K^4}} \quad \forall t \geq 0.$$

so with very high probability $\|X\|_2$ is within a constant of $\sqrt{n}$.

Let's write X as

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

The distribution of X is rotationally invariant $\Rightarrow$ so is the distr of $\frac{X}{\|X\|_2}$, but $\frac{X}{\|X\|_2}$ has length 1, so

$\frac{X}{\|X\|_2}$ is distributed uniformly over the sphere of radius one.

On the other hand $\|X\|_2$ is within a const of $\sqrt{n}$ w/ high prob., so $\frac{\|X\|_2}{\sqrt{n}} \approx 1$, so $\frac{X}{\sqrt{n}}$ is roughly uniformly distributed over the sphere of radius 1, so $X \approx \sqrt{n}\, Unif(S^{n-1})$.

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

Sub-gaussian distributions in higher dimensions

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

Def: 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 $\sup \|\langle X, t\rangle\|_{\psi_2}$, but that would be ∞.

Restrict only to $t \in S^{n-1}$:

The sub-gaussian norm of X is defined to be

$$\|X\|_{\psi_2} := \sup_{t \in S^{n-1}} \|\langle X, t\rangle\|_{\psi_2}.$$

Examples of sub-gaussian distributions

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

Q: what can we say about its coordinates?

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

What if we know $X = (X_1, \ldots, X_n)$ w/ $X_1, \ldots, X_n$ - subgaussian. 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.

$\langle X, t\rangle = X_1 t_1 + \ldots + X_n t_n$.

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

So X is subgaussian iff 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: If $X_1, X_2, \ldots, X_n$ are indep. mean-zero sub-gaussian RVs, then

$X = (X_1, \ldots, X_n) \in \mathbb{R}^n$ is sub-gaussian &

$$\|X\|_{\psi_2} \le C \max_{i \le n} \|X_i\|_{\psi_2} \quad \text{for some}$$

absolute constant C.

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

$$\|\langle X, t\rangle\|_{\psi_2}^2 = \|\sum t_i X_i\|_{\psi_2}^2 \overset{\text{shown previously; uses indep'ce}}{\le} C \sum_{i=1}^{n} \|t_i X_i\|_{\psi_2}^2 = C \sum_{i=1}^{n} t_i^2 \|X_i\|_{\psi_2}^2$$

$$\le C \max_{i=1}^{n} \|X_i\|_{\psi_2}^2 \sum_{i=1}^{n} t_i^2 = C \max_{i=1}^{n} \|X_i\|_{\psi_2}^2 \quad \triangleright$$

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

Of course it is sub-g. What is its sub-g norm?

If $t \in S^{n-1}$, then

$\langle X, t \rangle = X_1 t_1 + \dots + X_n t_n \sim N(0, t_1^2 + \dots + t_n^2) = N(0,1)$

so $\|X\|_{\psi_2} = \|N(0,1)\|_{\psi_2} < C$ indep of n.