Multivariate distributions

Gabriel T. Landi

References:

- Tomé, de Oliveira, chapter 1
- Ross, chapters 6 and 7

Covariance

Let X, Y be two r.v.s. We define their covariance as

$$\boxed{\mathrm{Cov}(X, Y) = \langle XY\rangle - \langle X\rangle\langle Y\rangle} \qquad (1)$$

If X and Y are independent, then $\langle XY\rangle = \langle X\rangle\langle Y\rangle$ and $\mathrm{Cov}(X,Y) = 0$. Thus, the covariance gives a measure of the degree of correlation between two r.v.s.

The covariance satisfies the following properties, which you may verify directly from (1):

(a) $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$ (2a)

(b) $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$ (2b)

(c) $\mathrm{Cov}(X, c) = 0$ when $c = \text{const.}$ (2c)

(d) $\mathrm{Cov}(cX, Y) = c\,\mathrm{Cov}(X, Y)$ (2d)

(e) $\mathrm{Cov}(X, Y+Z) = \mathrm{Cov}(X,Y) + \mathrm{Cov}(X, Z)$ (2e)

Properties (d) and (e) say that cov is bilinear: with one r.v. held constant, cov is linear in the other.

We may generalize (e) as follows

$$\begin{aligned}\mathrm{Cov}(X+Y, Z+W) &= \mathrm{Cov}(X+Y, Z) + \mathrm{Cov}(X+Y, W) \qquad (3)\\ &= \mathrm{Cov}(X, Z) + \mathrm{Cov}(Y, Z) + \mathrm{Cov}(X, W) + \mathrm{Cov}(Y, W)\end{aligned}$$

Or, more generally,

$$\boxed{\mathrm{Cov}\left(\sum_{i=1}^{m} a_i X_i \, , \; \sum_{j=1}^{m} b_i Y_i\right) = \sum_{i=1}^{m}\sum_{j=1}^{m} a_i b_i \,\mathrm{Cov}(X_i, Y_i)} \qquad (4)$$

This property may be combined with (2a) to give the following important result:

$$\begin{aligned} \mathrm{Var}(X_1 + X_2) &= \mathrm{Cov}(X_1 + X_2 \, , \, X_1 + X_2) \\ &= \mathrm{Cov}(X_1, X_1) + \mathrm{Cov}(X_1, X_2) + \mathrm{Cov}(X_2, X_1) + \mathrm{Cov}(X_2, X_2) \\ &= \mathrm{Var}(X_1) + 2\,\mathrm{cov}(X_1, X_2) + \mathrm{Var}(X_2) \end{aligned}$$

Thus

$$\boxed{\mathrm{Var}(X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + 2\,\mathrm{Cov}(X_1, X_2)} \qquad (5)$$

If X_1 and X_2 are independent, $\mathrm{cov}(X_1, X_2) = 0$ and we recover our previous result that the variance is additive. When X_1 and X_2 are not independent, the variance will no longer be additive.

Generalizing (5):

$$\boxed{\mathrm{Var}(X_1 + \ldots + X_N) = \sum_{i=1}^{N} \mathrm{Var}(X_i) + 2\sum_{i<j} \mathrm{Cov}(X_i, X_j)} \qquad (6)$$

In the last term we restrict the sum to $i<j$ and multiply by 2.

Covariance and independence

If X, Y are independent $\Rightarrow$ $\text{Cov}(X,Y) = 0$. But the converse is in general false: if $\text{Cov}(X,Y) = 0$ it does not imply that X, Y are independent.

Ex: let $Z \sim N(0,1)$ and define $X = Z$ and $Y = Z^2$. These are very dependent r.v.s: if you know one, you completely know the other. Yet

$$\text{cov}(X,Y) = \langle XY \rangle - \langle X \rangle \langle Y \rangle = \underbrace{\langle Z^3 \rangle}_{=0} - \underbrace{\langle Z \rangle}_{=0} \langle Z^2 \rangle = 0.$$

Correlation

If X, Y have units, $\text{cov}(X,Y)$ will also have units. It is convenient to introduce a dimensionless measure of correlation. A convenient definition is

$$\boxed{\text{Corr}(X,Y) = \frac{\text{Cov}(X,Y)}{SD(X)\, SD(Y)}} \tag{7}$$

where $SD(X) = \sqrt{\text{Var}(X)}$ is the standard deviation. The correlation will be a number between -1 and 1 (proof below):

$$-1 \leqslant \text{Corr}(X,Y) \leqslant 1 \tag{8}$$

If $\text{Corr} = 0$ they are uncorrelated. If $\text{corr} \geqslant 0$ they are positively correlated. In this case $X > 0$ makes it more likely that $Y > 0$. If $\text{corr} \leqslant 0$ they are anti-correlated (or negatively correlated). In this case $X > 0$ makes it likely that $Y < 0$.

The inequality (8) may also be written as

$$\boxed{[\text{Cov}(X,Y)]^2 \leq \text{Var}(X)\,\text{Var}(Y)} \tag{9}$$

This is a manifestation of the Cauchy - Schwarz inequality. To prove it we define a new r.v.

$$Z = X - cY$$

where c is a constant. Using (5) we then get

$$\text{Var}(Z) = \text{Var}(X) + c^2\,\text{Var}(Y) - 2c\,\text{Cov}(X,Y) \geq 0$$

since $\text{Var}(Z) \geq 0$. Now choose $c = \dfrac{\text{Cov}(X,Y)}{\text{Var}(Y)}$. Then we get

$$\text{Var}(X) + \left[\frac{\text{Cov}(X,Y)}{\text{Var}(Y)}\right]^2 \text{Var}(Y) - 2\left[\frac{\text{Cov}(X,Y)}{\text{Var}(Y)}\right]\text{Cov}(X,Y) \geq 0$$

or

$$\text{Var}(X) - \frac{[\text{Cov}(X,Y)]^2}{\text{Var}(Y)} \geq 0$$

which is precisely Eq (9). qed.

The covariance matrix

Say you have a bunch of r.v.s $X_1, \dots, X_N$, and you want to study their covariances. It is useful to group them in a matrix, with entries

$$\Theta_{ij} = \mathrm{Cov}(X_i, X_j) \tag{10}$$

This is called the <u>covariance matrix</u>. For instance, if $N=2$ we will have

$$\Theta = \begin{bmatrix} \mathrm{cov}(X_1, X_1) & \mathrm{cov}(X_1, X_2) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{cov}(X_2, X_2) \end{bmatrix} \tag{11}$$

Note that the diagonals are the variances. So if $X_1, \dots, X_N$ are independent, the covariance matrix would be diagonal. Note also that, since $\mathrm{cov}(X_2, X_1) = \mathrm{Cov}(X_1, X_2)$, Θ will be a symmetric matrix

$$\Theta^T = \Theta \tag{12}$$

It is also possible to write a neat formula for Θ using vectors. Let us group the r.v.s $X_1, \dots, X_N$ into a column vector

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_N \end{bmatrix} \qquad X^T = \begin{bmatrix} X_1 & \dots & X_N \end{bmatrix} \tag{13}$$

the <u>inner</u> (or <u>dot</u>) product may then be written as

$$\begin{aligned} X^T X &= \begin{bmatrix} X_1 & \dots & X_N \end{bmatrix} \begin{bmatrix} X_1 \\ \vdots \\ X_N \end{bmatrix} \\ &= X_1^2 + \dots + X_N^2 \end{aligned} \tag{14}$$

We may also define the <u>outer product</u>, which produces a matrix

$$x x^T = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} \begin{bmatrix} x_1 & \dots & x_N \end{bmatrix} = \begin{bmatrix} x_1^2 & x_1 x_2 & \dots & x_1 x_N \\ x_2 x_1 & x_2^2 & \dots & x_2 x_N \\ \vdots & & \ddots & \\ x_N x_1 & x_N x_2 & \dots & x_N x_N \end{bmatrix} \tag{15}$$

The outer product produces a matrix with entries

$$(x x^T)_{ij} = x_i x_j \tag{16}$$

If we now average the matrix $x x^T$, we will get

$$\langle x x^T \rangle = \begin{bmatrix} \langle x_1^2 \rangle & \langle x_1 x_2 \rangle & \dots \\ \langle x_1 x_2 \rangle & \langle x_2^2 \rangle & \dots \\ \vdots & \vdots & \end{bmatrix} \tag{17}$$

these look close to the entries of Θ in (10). What we need is to subtract $\langle x_i \rangle \langle x_j \rangle$. Thus, we finally conclude that the covariance matrix may be written as

$$\boxed{\Theta = \langle x x^T \rangle - \langle x \rangle \langle x \rangle^T} \tag{18}$$

This is a very convenient way of writing down the covariance matrix. If you ever get confused about the vector notation, than think about the problem in terms of components

$$\boxed{\Theta_{ij} = \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle} \tag{19}$$

which is just Eq (10) written down explicitly

The multivariate normal

Let $Z_1, Z_2 \sim N(0,1)$ be two iid standard normals. Their covariance matrix is

$$\Phi(Z) = \begin{bmatrix} Var(Z_1) & cov(Z_1, Z_2) \\ Cov(Z_1, Z_2) & Var(Z_2) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2 \qquad (20)$$

We say the vector $Z = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} \sim N(0, I_2)$, which means a multivariate normal with zero mean and covariance matrix I_2. It is straightforward to extend these ideas to an arbitrary number of Z_i's.

The Z_i are uncorrelated, which is not very interesting. We may produce correlated normal r.v.s by taking linear combinations of Z_1 and Z_2

$$\begin{aligned} X_1 &= q_{11} Z_1 + q_{12} Z_2 + \mu_1 \\ X_2 &= q_{21} Z_1 + q_{22} Z_2 + \mu_2 \end{aligned} \qquad (21)$$

where μ_i and q_{ij} are arbitrary coefficients. Now X_1 and X_2 are no longer independent

$$\begin{aligned} Cov(X_1, X_2) &= Cov(q_{11} Z_1 + q_{12} Z_2 + \mu_1, \; q_{21} Z_1 + q_{22} Z_2 + \mu_2) \\ &= q_{11} q_{21} Cov(Z_1, Z_1) + q_{12} q_{22} Cov(Z_2, Z_2) \\ &= q_{11} q_{21} + q_{12} q_{22} \end{aligned} \qquad (22)$$

By mixing the Z_i we produce correlated variables. As a limiting case we may take $X_1 = Z_1$ and $X_2 = Z_1$. In this case X_1 and X_2 are perfectly correlated.

To generalize this idea to an arbitrary number of r.v.s, it is convenient to rewrite (21) in matrix notation

$$X = Q Z + \mu \tag{23}$$

where

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \qquad Q = \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix} \tag{24}$$

The average of X is

$$\langle X \rangle = Q\langle Z \rangle + \mu = \mu \tag{25}$$

As for the covariance matrix, we start with

$$\begin{aligned} X X^T &= (QZ+\mu)(QZ+\mu)^T \\ &= (QZ+\mu)(Z^T Q^T + \mu^T) \\ &= Q Z Z^T Q^T + Q Z \mu^T + \mu Z^T Q^T + \mu \mu^T \end{aligned}$$

when we take average, the two terms in the middle will vanish since $\langle Z \rangle = 0$. We are then left with

$$\langle X X^T \rangle = Q Q^T + \mu \mu^T \tag{26}$$

This is not yet $\Theta(X)$, Eq (18). To get the actual covariance matrix we must subtract $\langle X \rangle \langle X \rangle^T = \mu \mu^T$. We then get

$$\boxed{\Theta(X) = Q Q^T} \tag{27}$$

Let us check to see if this makes sense. Using (24) we get

$$QQ^T = \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix} \begin{bmatrix} q_{11} & q_{21} \\ q_{12} & q_{22} \end{bmatrix} = \begin{bmatrix} q_{11}^2 + q_{12}^2 & q_{11} q_{21} + q_{12} q_{22} \\ q_{21} q_{11} + q_{22} q_{12} & q_{21}^2 + q_{22}^2 \end{bmatrix} \tag{28}$$

We see in entry (1,2) or (2,1) exactly the covariance $Cov(X_1, X_2)$ computed in Eq (22). You may verify that the diagonal entries are $Var(X_1)$ and $Var(X_2)$.

— // —

The general multivariate normal is characterized by a mean vector μ and a covariance matrix Θ. We write

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_N \end{bmatrix} \sim N(\mu, \Theta) \tag{29}$$

The mean vector may be arbitrary but the covariance matrix must satisfy two properties. First, it must be symmetric, $\Theta^T = \Theta$. And second, it must be positive definite; i.e., all its eigenvalues must be larger than zero. This is actually a consequence of Eq (9) [It is something which is a bit more difficult to demonstrate].

Technical comment: any positive definite matrix may be factored as in (27), which is known as the Cholesky decomposition of the matrix. A pos. def. matrix is the matrix analogue of a positive number and the factorization (27) is the matrix analogue of taking the square root.

PDF of the multivariate normal

The PDF of $Z = (Z_1, \dots, Z_N)$ is easy to find since they are independent.

$$p_{Z_1,\dots,Z_N}(z_1,\dots,z_N) = p_{Z_1}(z_1)\dots p_{Z_N}(z_N)$$

$$= \frac{1}{\sqrt{(2\pi)^N}} \exp\left\{-\frac{1}{2}(z_1^2 + \dots + z_N)^2\right\}$$

We may write this more concisely as

$$p_Z(z) = \frac{1}{\sqrt{(2\pi)^N}} \exp\left\{-\frac{1}{2} z^T z\right\}$$

This is the PDF of $N(0, I_N)$. The PDF of a general Gaussian $N(\mu, \Theta)$ is

$$\boxed{p_X(x) = \frac{1}{\sqrt{(2\pi)^N \det(\Theta)}} \exp\left\{-\frac{1}{2}(x-\mu)^T \Theta^{-1}(x-\mu)\right\}} \qquad (29')$$

I will demonstrate this formula at the end of these notes (I will not do it in class).

When $\mu = 0$ and $\Theta = I_N$ we recover the PDF of Z.

The bivariate normal

Let us practice with the 2-component normal. It's covariance matrix is, by definition

$$\Theta = \begin{bmatrix} \mathrm{cov}(X_1, X_1) & \mathrm{cov}(X_1, X_2) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{cov}(X_2, X_2) \end{bmatrix}$$

To simplify the notation let $\sigma_1^2 = \mathrm{cov}(X_1, X_1)$ and $\sigma_2^2 = \mathrm{Cov}(X_2, X_2)$. It is also convenient to define

$$\rho = \frac{\mathrm{Cov}(X_1, X_2)}{\sigma_1 \sigma_2} \in [-1, 1]$$

then we may write Θ as

$$\Theta = \begin{bmatrix} \sigma_1^2 & \sigma_1 \sigma_2 \rho \\ \sigma_1 \sigma_2 \rho & \sigma_2^2 \end{bmatrix}$$

what appears in (29') is actually Θ^{-1}. I will borrow the following result from linear algebra

det

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

We have,

$$\det \Theta = \sigma_1^2 \sigma_2^2 - \sigma_1^2 \sigma_2^2 \rho^2$$
$$= \sigma_1^2 \sigma_2^2 (1-\rho^2)$$

thus

$$\Theta^{-1} = \frac{1}{\sigma_1^2 \sigma_2^2 (1-\rho^2)} \begin{bmatrix} \sigma_2^2 & -\sigma_1 \sigma_2 \rho \\ -\sigma_1 \sigma_2 \rho & \sigma_1^2 \end{bmatrix}$$

Now we must compute $(x-\mu)^T \Theta^{-1}(x-\mu)$. For simplicity let $y = x - \mu = (x_1, x_2)$. Then

$$y^T \Theta^{-1} y = \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)} \begin{bmatrix} y_1 & y_2 \end{bmatrix} \begin{bmatrix} \sigma_2^2 & -\sigma_1\sigma_2\rho \\ -\sigma_1\sigma_2\rho & \sigma_1^2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

$$= \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)} \begin{bmatrix} y_1 & y_2 \end{bmatrix} \begin{bmatrix} \sigma_2^2 y_1 - \sigma_1\sigma_2\rho y_2 \\ -\sigma_1\sigma_2\rho y_1 + \sigma_1^2 y_2 \end{bmatrix}$$

$$= \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)} \left\{ \sigma_2^2 y_1^2 - 2\sigma_1\sigma_2\rho y_1 y_2 + \sigma_1^2 y_2 \right\}$$

$$= \frac{1}{1-\rho^2} \left\{ \frac{y_1^2}{\sigma_1^2} - \frac{2\rho y_1 y_2}{\sigma_1\sigma_2} + \frac{y_2^2}{\sigma_2^2} \right\}$$

Thus we conclude that the PDF of the bivariate normal will be

$$\boxed{P_X(x) = \frac{1}{\sqrt{(2\pi)^2\sigma_1^2\sigma_2^2(1-\rho^2)}} \exp\left\{ -\frac{1}{2(1-\rho^2)} \left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - \frac{2\rho}{\sigma_1\sigma_2}(x_1-\mu_1)(x_2-\mu_2) + \frac{(x_2-\mu_2)^2}{\sigma_2^2} \right\} \right.}$$

Transformation of variables

In a previous set of notes I showed how to find the PDF of $Y = f(X)$ given $P_X(x)$. The result was:

$$P_Y(y) = \int_{-\infty}^{\infty} dx \; \delta(y - f(x)) \; P_X(x) \tag{30}$$

This formula may be extended to more complicated transformations. It always has the same structure.

For instance, let X, Y be two r.v.'s with joint PDF $P_{X,Y}(x,y)$ and let $Z = f(X,Y)$. Then

$$\boxed{P_Z(z) = \int dx\,dy \; \delta(z - f(x,y)) \; P_{X,Y}(x,y)} \tag{31}$$

You may demonstrate this formula using the same trick that we used to demonstrate (39) (via the characteristic function).

Or suppose you want to change variables, from X, Y to $Z = f(X,Y)$ and $W = g(X,Y)$. In this case

$$\boxed{P_{Z,W}(z,w) = \int dx\,dy \; \delta(z - f(x,y)) \; \delta(w - g(x,y)) \; P_{X,Y}(x,y)} \tag{32}$$

If we integrate (32) over w we get (30).

Example: the Cauchy distribution

Let $X, Y \sim N(0,1)$ iid, and define

$$T = \frac{X}{Y} \tag{33}$$

We say $T \sim$ Cauchy $(0,1)$. To find the PDF we use Eq (31):

$$P_T(t) = \int dx\,dy\; \delta\left(t - \frac{x}{y}\right) P_X(x)\, P_Y(y) \tag{34}$$

We compute the integral over x first. Recall that

$$\delta(ax) = \frac{\delta(x)}{|a|} \tag{35}$$

then

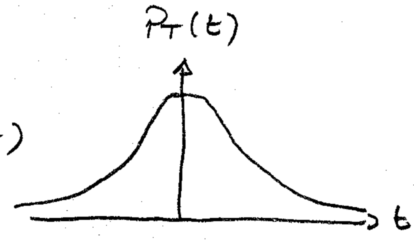
$$\delta\left(t - \frac{x}{y}\right) = \delta\left(\frac{ty - x}{y}\right) = |y|\, \delta(ty - x) \tag{36}$$

Eq (34) then becomes

$$\begin{aligned} P_T(t) &= \int dx\,dy\; |y|\, \delta(ty - x)\, P_X(x) \cdot P_Y(y) \\ &= \int dy\; |y|\, P_X(ty)\, P_Y(y) \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} dy\; |y|\, e^{-y^2(1+t^2)/2} \\ &= \frac{1}{\pi} \int_0^{\infty} dy\; y\, e^{-y^2(1+t^2)/2} \qquad u = y^2/2 \quad du = y\,dy \\ &= \frac{1}{\pi} \int_0^{\infty} e^{-u(1+t^2)}\, du \\ &= \frac{1}{\pi} \frac{1}{1+t^2} \end{aligned}$$

Thus

$$P_T(t) = \frac{1}{\pi} \frac{1}{1+t^2} \qquad (37)$$



This is the PDF of Cauchy (0,1). The more general Cauchy (a, b) is defined as

$$T \sim \text{Cauchy}(a,b): \qquad P_T(t) = \frac{1}{\pi} \frac{b}{b^2 + (t-a)^2} \qquad (38)$$

The general Cauchy is defined as $b\frac{X}{Y} + a$.

The Cauchy distribution was sent from Hell by Satan himself. At first it seems innocent; it even looks a little bit like the Gaussian. But it is fundamentally different: its tails fall off too slowly. As a consequense, all its moments are either infinite or undefined. And it does not satisfy the central limit theorem or the law of large numbers. The latter means that

$$\lim_{N \to \infty} \frac{T_1 + \cdots + T_N}{N} = \text{undefined} \qquad (39)$$

This limit does not converge to anything. It just keeps fluctuating like crazy. The first person to notice this was Poisson.

Why the mean of the Cauchy is undefined

The mean of Cauchy (0,1) is

$$\langle T \rangle = \int_{-\infty}^{\infty} t \, P_T(t) \, dt = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{t}{1+t^2} \, dt \qquad (40)$$

At first you may argue that this is zero since $\frac{t}{1+t^2}$ is an odd function of t. But to see where the trouble is, notice that we could take (40) to mean

$$\langle T \rangle = \lim_{\sigma \to \infty} \frac{1}{\pi} \int_{-\sigma}^{\sigma} \frac{t}{1+t^2} \, dt = 0 \qquad (41)$$

This is called the Cauchy principal value and it is zero. But we could also have taken

$$\langle T \rangle = \lim_{\sigma \to \infty} \frac{1}{\pi} \int_{-2\sigma}^{\sigma} \frac{t}{1+t^2} \, dt \qquad (42)$$

and this is not zero. Thus, depending on how we define the limit, we get different results. That is why we say the mean is undefined.

More about the transformation formula

Let us now turn to Eq (32):

$$P_{z,w}(z,w) = \int dx\,dy\; \delta(z - f(x,y))\,\delta(w - g(x,y))\, P_{x,y}(x,y)$$

Suppose that the transformation $z = f(x,y)$ and $w = g(x,y)$ is one-to-one. Then we may change variables in the integral, from (x,y) to (z,w). Recall from calculus that

$$dx\,dy = dz\,dw \left| \frac{\partial(x,y)}{\partial(z,w)} \right| \tag{43}$$

where $\left|\frac{\partial(\cdots)}{\partial(\cdots)}\right|$ is the Jacobian determinant of the transformation

$$\left| \frac{\partial(x,y)}{\partial(z,w)} \right| = \begin{vmatrix} \frac{\partial x}{\partial z} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial z} & \frac{\partial y}{\partial w} \end{vmatrix} \tag{44}$$

In this case the transformation formula simplifies to

$$\boxed{P_{z,w}(z,w) = P_{x,y}(x,y) \left| \frac{\partial(x,y)}{\partial(z,w)} \right|} \tag{45}$$

Example: the Box-Muller algorithm

The Box-Muller algorithm is a method to generate $N(0,1)$ r.v.s from $\mathrm{Unif}(0,1)$. Let $U, V \sim \mathrm{Unif}(0,1)$, iid. Define

$$X = \sqrt{-2\ln U}\,\cos(2\pi V)$$
$$Y = \sqrt{-2\ln U}\,\sin(2\pi V) \qquad (46)$$

We will show that $X, Y \sim N(0,1)$ iid. To do that we apply Eq. (45). The Jacobian of the transformation is:

$$\begin{aligned} x &= \sqrt{-2\ln u}\,\cos(2\pi v) \\ y &= \sqrt{-2\ln u}\,\sin(2\pi v) \end{aligned} \quad \longrightarrow \quad \begin{aligned} u &= e^{-(x^2+y^2)/2} \\ v &= \frac{1}{2\pi}\arctan(y/x) \end{aligned}$$

so

$$\frac{\partial u}{\partial x} = -\,x\,e^{-(x^2+y^2)/2} \qquad \frac{\partial u}{\partial y} = -\,y\,e^{-(x^2+y^2)/2} \qquad (47)$$

We also use the result

$$\frac{d}{dx}\arctan(x) = \frac{1}{1+x^2} \qquad (48)$$

then

$$\frac{\partial v}{\partial x} = \frac{1}{2\pi}\frac{\partial}{\partial x}\arctan(y/x) = \frac{1}{2\pi}\frac{-y}{x^2}\,\frac{1}{1+(y/x)^2} = \frac{1}{2\pi}\frac{-y}{x^2+y^2}$$
$$\frac{\partial v}{\partial y} = \frac{1}{2\pi}\frac{\partial}{\partial y}\arctan(y/x) = \frac{1}{2\pi}\frac{1}{x}\,\frac{1}{1+(y/x)^2} = \frac{1}{2\pi}\frac{x}{x^2+y^2} \qquad (49)$$

thus

$$\left|\frac{\partial(u,v)}{\partial(x,y)}\right| = \begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{vmatrix} = \frac{\partial u}{\partial x}\frac{\partial v}{\partial y} - \frac{\partial u}{\partial y}\frac{\partial v}{\partial x}$$
$$= \frac{1}{2\pi}\frac{e^{-(x^2+y^2)/2}}{x^2+y^2}\left\{ \underbrace{(-x)(x) - (-y)(-y)}_{-(x^2+y^2)} \right\}$$

Taking the absolute value, we get

$$\left| \frac{\partial(u,v)}{\partial(x,y)} \right| = \frac{1}{2\pi} e^{-(x^2+y^2)/2} \tag{50}$$

Thus, since the PDF of Unif (0,1) is 1, we get

$$p_{X,Y}(x,y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2} \tag{51}$$

This PDF factors into a product

$$p_{X,Y}(x,y) = \frac{e^{-x^2/2}}{\sqrt{2\pi}} \frac{e^{-y^2/2}}{\sqrt{2\pi}} = p_X(x)\, p_Y(y) \tag{52}$$

Thus, we conclude that X and Y are statistically independent $N(0,1)$.

<u>Conclusion</u>: generate two Unif (0,1) r.v.s, U and V. Then apply Eq (46) and you get two indep. $N(0,1)$.

If you want $N(\mu,\sigma^2)$ then you just need to recall that $\sigma X+\mu \sim N(\mu,\sigma^2)$ if $X \sim N(0,1)$.

Example: the Maxwell - Boltzmann distribution

We will learn later in the course that the prob. of finding a system, which is in thermal equilibrium, at a given state is

$$p = \frac{e^{-E/k_BT}}{z} \qquad z = \text{norm. const.} \tag{53}$$

where E is the energy of that state. Now consider just a single particle in a gas and assume it has only kinetic energy, so that

$$E = \frac{1}{2} m v^2 = \frac{1}{2} m (v_x^2 + v_y^2 + v_z^2) \tag{54}$$

Then the probability of finding the particle with some velocity vector $\vec{v}$ will be

$$\boxed{p(v_x, v_y, v_z) = \frac{1}{z} \exp\left\{ -\frac{m}{2k_BT} (v_x^2 + v_y^2 + v_z^2) \right\}} \tag{55}$$

Here the 3 components of the velocity are the random variables and this is their joint distribution

It is quite clear that the 3 cartesian components are statistically independent since

$$p(v_x, v_y, v_z) = p(v_x)\, p(v_y)\, p(v_z) \tag{56}$$

This reflects the isotropy of space

We also see that each probability is a Gaussian with mean 0 and variance k_BT/m. That is, $v_i \sim N(0, k_BT/m)$. This allows us to determine the normalization constant Z. I always remember that

$$X \sim N(\mu, \sigma^2): \quad P_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$$

Thus, in our case, we will have

$$\boxed{p(v_x, v_y, v_z) = \frac{1}{(2\pi)^{3/2}} \left(\frac{m}{k_BT}\right)^{3/2} e^{-\frac{m}{2k_BT}(v_x^2 + v_y^2 + v_z^2)}} \tag{57}$$

We also know that if $X \sim N(0, \sigma^2)$, $\langle X^2 \rangle = \mathrm{Var}(X) = \sigma^2$. Thus we conclude that

$$\langle v_x^2 \rangle = \frac{k_B T}{m} \tag{58}$$

We also customarily write this as

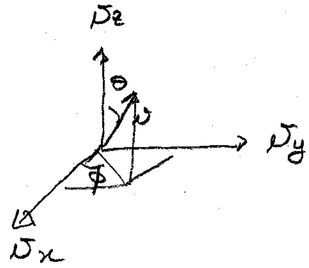
$$\frac{1}{2} m \langle v_x^2 \rangle = \frac{1}{2} k_B T \tag{59}$$

The average kinetic energy is then

$$\frac{1}{2} m \langle v^2 \rangle = \frac{1}{2} m \left[\langle v_x^2 \rangle + \langle v_y^2 \rangle + \langle v_z^2 \rangle\right] = \frac{3}{2} k_B T \tag{60}$$

This reflects an old idea of <u>equipartition</u>: attribute $1/2\, k_BT$ to each degree of freedom of your system. This is a limited result. It holds only in a few simple situations.

Now let us compute the probability distribution of the magnitude of the velocity $v = \sqrt{v_x^2 + v_y^2 + v_z^2}$. We can do that by moving to spherical coordinates



$$v_x = v \sin\theta \cos\phi$$
$$v_y = v \sin\theta \sin\phi \qquad (61)$$
$$v_z = v \cos\theta$$

The Jacobian of the transformation is $v^2 \sin\theta$, as you probably remember from the formula

$$dv_x\, dv_y\, dv_z = v^2 \sin\theta \, dv\, d\theta\, d\phi \qquad (62)$$

Thus, the distribution of v, θ, ϕ will be, according to (45) and (57):

$$\boxed{p(v,\theta,\phi) = \left(\frac{m}{2\pi k_B T}\right)^{3/2} v^2 \sin\theta \; e^{-mv^2/2k_BT}} \qquad (63)$$

This is the joint distribution to observe the velocity with magnitude v and at a certain direction θ, ϕ.

Notice that it is uniform in ϕ (it does not depend on ϕ). It is not uniform in θ since it depends on $\sin\theta$. This makes it more likely to observe values of θ near the equatorial line. The r.v.s (v, θ, ϕ) are statistically independent since (63) factors into a product

$$\boxed{p(v,\theta,\phi) = \underbrace{\frac{1}{2\pi}}_{\rho(\phi)} \; \underbrace{\frac{\sin\theta}{2}}_{\rho(\theta)} \; \underbrace{4\pi \left(\frac{m}{2\pi k_B T}\right) v\, e^{-mv^2/2k_BT}}_{\rho(v)}} \qquad (64)$$

I was careful to sepparate the numerical factors in order to ensure that each probability is correctly normalized:

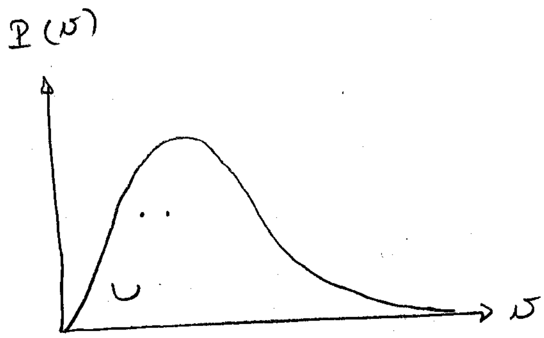
$$\int_0^{2\pi} P(\phi)\,d\phi = \frac{1}{2\pi}\, 2\pi = 1$$

$$\int_0^{\pi} P(\theta)\,d\theta = \frac{1}{2}\int_0^{\pi} \sin\theta\, d\theta = \frac{1}{2}\,(-\cos\theta)\Big|_0^{\pi} = 1 \qquad (65)$$

$$\int_0^{\infty} P(v)\,dv = 1$$ ← This is not obvious. You'd have to check.

The distribution $P(v)$ is the most important part: it is called the Maxwell-Boltzmann distribution of velocities

$$\boxed{P(v) = 4\pi \left(\frac{m}{2\pi k_B T}\right)^{3/2} v\, e^{-mv^2/2k_BT}} \qquad (66)$$



If you measure the speed of a molecule in a gas, this is the distribution (you can do this using time-of-flight spectrometers).

Example: PDF of the multivariate normal

Let $Z_1, \dots, Z_N \sim N(0,1)$ iid. We saw that a general multivariate Gaussian may be defined as

$$X = QZ + \mu \tag{67}$$

Now let us find the PDF of X, starting from the PDF of Z. Since the Z_i are independent

$$P_Z(z) = P_{Z_1,\dots,Z_N}(z_1,\dots,z_N) = P_{Z_1}(z_1) \cdots P_{Z_N}(z_N)$$

$$= \frac{1}{\sqrt{(2\pi)^N}} \exp\left\{-\frac{1}{2}(z_1^2 + \dots + z_N^2)\right\}. \tag{68}$$

We may write this more compactly as

$$P_Z(z) = \frac{1}{\sqrt{(2\pi)^N}} \exp\left\{-\frac{1}{2} z^T z\right\} \tag{69}$$

where

$$z^T z = z \cdot z = \begin{bmatrix} z_1 & \dots & z_N \end{bmatrix} \begin{bmatrix} z_1 \\ \vdots \\ z_N \end{bmatrix} = z_1^2 + \dots + z_N^2 \tag{70}$$

To find the distribution of X we use Eq (45). Notice that

$$\frac{\partial x_i}{\partial z_j} = Q_{ij} \tag{71}$$

since x and z are linearly related by (67). Consequently

$$\left| \frac{\partial(x_1,\dots,x_N)}{\partial(z_1,\dots,z_N)} \right| = \det(Q) \tag{72}$$

is simply the determinant of the matrix Q.

We saw in (27) that the covariance matrix is $\Theta = QQ^T$.

From linear algebra we have the following results

$$\det(AB) = \det(A)\det(B) \qquad (73)$$
$$\det(A^T) = \det(A)$$

consequently

$$\det(\Theta) = \det(Q)^2 \Rightarrow \det(Q) = \sqrt{\det(\Theta)} \qquad (74)$$

We may then write the PDF of $X = (X_1, \dots, X_N)$ as

$$P_X(x) = P_Z(z)\left|\frac{\partial(z)}{\partial(x)}\right| = \frac{1}{\sqrt{(2\pi)^N \det(\Theta)}} e^{-z^Tz/2} \qquad (75)$$

To finish we must express z^Tz as a function of x. We have $x = Qz + \mu$ so

$$z = Q^{-1}(x-\mu) \qquad (76)$$

consequently

$$\begin{aligned} z^Tz &= [Q^{-1}(x-\mu)]^T[Q^{-1}(x-\mu)] \\ &= (x-\mu)^T(Q^{-1})^T Q^{-1}(x-\mu) \\ &= (x-\mu)^T[QQ^T]^{-1}(x-\mu) \\ &= (x-\mu)^T\Theta^{-1}(x-\mu) \end{aligned}$$

thus, the PDF of $N(\mu, \Theta)$ is

$$\boxed{P_X(x) = \frac{1}{\sqrt{(2\pi)^N\det(\Theta)}}\exp\left\{-\frac{1}{2}(x-\mu)^T\Theta^{-1}(x-\mu)\right\}} \qquad (77)$$