N-DIMENSIONAL CUMULATIVE FUNCTION, AND OTHER USEFUL FACTS ABOUT GAUSSIANS AND NORMAL DENSITIES

by Michaël Bensimhoun (reviewed by David Arnon)
Jerusalem, 06/2009

A. 1-d Gaussians and normal densities

The 1-d normal density with mean μ and standard deviation $\sigma > 0$ is given by

$$\boxed{N(x,\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}.}$$

Its variance can be shown to be σ^2.

The cumulative function of $N(x,0,1)$ cannot be computed explicitly, and is usually denoted by Φ:

$$\boxed{\Phi(t) = \int_{-\infty}^{t} N(x,0,1)\mathrm{d}x.}$$

In the general case, performing the change of variable $x' = (x-\mu)/\sigma$ in the integral $F(t) = \int_{-\infty}^{t} N(x,\mu,\sigma)\mathrm{d}x$, the cumulative function of $N(x,\mu,\sigma)$ is easily seen to be

$$\boxed{F(x) = \Phi((x-\mu)/\sigma).}$$

Notice that F is invertible, with inverse equal to

$$\boxed{F^{-1}(y) = \sigma\Phi^{-1}(y) + \mu,}$$

as can be easily checked.

The Mahalanobis distance of a point x, to the Gaussian that composes the normal density, is by definition

$$\boxed{r = \frac{|x-\mu|}{\sigma};}$$

it measures how many standard deviations is the point x far from the Gaussian.

Given $R > 0$, let us compute the probability that a point falls at a Mahalanobis distance $r \leq R$ from the Gaussian above. This probability is equal to

$$\int_{\mu-\sigma R}^{\mu+\sigma R} N(x,\mu,\sigma)\mathrm{d}x,$$

and since a Gaussian is symmetric about its mean, it is equal to

$$2\int_{\mu}^{\mu+\sigma R} N(x,\mu,\sigma)\mathrm{d}x = 2\big[F(\mu+\sigma R) - F(\mu)\big] = \boxed{2F(\mu+\sigma R) - 1.}$$

Conversely, if the probability p that a point falls within a certain Mahalanobis distance R from the Gaussian is known, one can compute R explicitly:

$$p = 2F(\mu+\sigma R) - 1 \Longrightarrow F(\mu+\sigma R) = \frac{p+1}{2} \Longrightarrow R = \Big[F^{-1}\Big(\frac{p+1}{2}\Big) - \mu\Big]/\sigma$$
$$\Longrightarrow \boxed{R = \Phi^{-1}\Big(\frac{p+1}{2}\Big).}$$

B. <u>2-D GAUSSIANS AND BIVARIATE NORMAL DENSITIES</u>

The bivariate normal density with mean $\mu = (\mu_1, \mu_2)$ and covariance matrix Σ is (with $z = (x, y)$)

$$\boxed{p(x,y) = \frac{1}{2\pi\sqrt{|\Sigma|}}\exp[-\frac{1}{2}(z-\mu)\Sigma^{-1}(z-\mu)^t],}$$

where Σ is assumed to be symmetric and positive definite. It is not difficult to show that the most general form of such two-dimensional a matrix is

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

where σ_1 and σ_2 are different from 0, and $|\rho| < 1$. Its inverse matrix Σ^{-1} can be shown to be

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

The Mahalanobis distance of a point $z = (x, y)$, to the Gaussian that composes the bivariate law above, is by definition given by r such that

$$\boxed{r^2 = (z-\mu)\Sigma^{-1}(z-\mu)^t.}$$

Explicitly,

$$r^2 = \frac{1}{1-\rho^2}\Big[\Big(\frac{x-\mu_1}{\sigma_1}\Big)^2 - 2\rho\Big(\frac{x-\mu_1}{\sigma_1}\Big)\Big(\frac{y-\mu_2}{\sigma_2}\Big) + \Big(\frac{y-\mu_2}{\sigma_2}\Big)^2\Big].$$

Since $|\rho| < 1$, the locus of the points for which r is constant are ellipses. It is not difficult to see that a parametrization of such an ellipse at level r^2 is

$$\begin{aligned} x &= r\sigma_1\cos\theta + \mu_1, \\ y &= r\sigma_2(\rho\cos\theta + \sqrt{1-\rho^2}\sin\theta) + \mu_2. \end{aligned}$$

Interestingly, if M is the Choleski decomposition of Σ (M is an upper triangular matrix such that $M^tM = \Sigma$), then this parametrization can be written

$$\boxed{(x, y) = r(\cos\theta, \sin\theta)M + (\mu_1, \mu_2);}$$

this can be shown with ease by the explicit computation of M, namely

$$M = \begin{bmatrix} \sigma_1 & \rho\sigma_2 \\ 0 & \sigma_2\sqrt{1-\rho^2} \end{bmatrix},$$

or more even simply using the fact that the parametrization above is nothing else than the locus of the points $ruM + \mu$, where $u = (x, y)$ is the set of points such that $uu^t = 1$. To see this, it suffices to put $z = ruM + \mu$ in the expression $r^2 = (z-\mu)\Sigma^{-1}(z-\mu)^t$. This Choleski form of the parametrization holds also in higher dimensions, and can be obtained efficiently in computers.

Another important fact is that the directions of the two axis of the ellipse above are given by the eigenvectors of Σ, while the extent of the great and small radius are respectively $r\sqrt{\lambda_1}$ and $r\sqrt{\lambda_2}$ (λ_1 and λ_2 being respectively the largest and smallest eigenvalues of Σ). In place of proving this directly, this can be proved using the parametrization above. To this end, one can suppose without loss of generality that $(\mu_1, \mu_2) = 0$. Denoting $u = (\cos\theta, \sin\theta)$, we have seen that the ellipse is the locus of the points $z = (x, y)$ such that

$$z = ruM, \quad \text{with} \quad uu^t = 1.$$

The great radius of the ellipse is the segment OP, where O is the center of the ellipse (that is, $(0, 0)$), and $P = (x, y)$ is one of the two points of the ellipse that are located at maximal distance from the center of the ellipse. Similarly, the small radius of the ellipse is given by one of the two points of the ellipse that are located at minimal distance from the center. In other word, we have to find z and u such that $uu^t = 1$ and

$$zz^t = x^2 + y^2 = r^2uMM^tu^t = \max, \quad \text{and similarly} \quad zz^t = r^2uMM^tu^t = \min.$$

But this is the classic Rayleigh problem, and it is well known that in the maximum case, $zz^t/r^2 = \lambda_1$ is the largest eigenvalue of MM^t, and $u^t = u_1^t$ is its corresponding normalized eigenvector. In the minimum case, $zz^t/r^2 = \lambda_2$ is the smallest eigenvalue of MM^t and $u^t = u_2^t$ is its corresponding normalized eigenvector. This gives the desired vectors $\vec{OP} = r^2u_{1,2}M$, and the extents $r\sqrt{\lambda_1}$ and $r\sqrt{\lambda_2}$ of the great and small radii of the ellipse respectively. This does not prove, yet, that $\lambda_{1,2}$ are eigenvalues of Σ, since $\Sigma = M^tM \neq MM^t$. Nevertheless, the vectors $\vec{OP}^t = r^2M^tu_{1,2}^t$ are in fact eigenvectors of Σ, and $\lambda_{1,2}$ are their corresponding eigenvalues in Σ: indeed, there holds

$$\Sigma M^tu_{1,2}^t = M^tMM^tu_{1,2}^t = M^t(\lambda_{1,2}u_{1,2}^t) = \lambda_{1,2}M^tu_{1,2}^t;$$

This shows that $\lambda_{1,2}$ are the eigenvalues of Σ, with corresponding eigenvectors proportional to the great and small radii of the ellipse, as was to be shown.

Now let $p(x, y)$ be the general bivariate normal density, with Σ as above. We hope to compute the cumulative function of this Gaussian, as a function of r. More precisely, our aim is to compute the probability that a point falls inside the ellipse given by the parametrization above, for a given Mahalanobis distance $r = R$.

To this end, let us first compute the Jacobian of the parametrization above. It is equal to

$$\begin{vmatrix} \partial x/\partial r & \partial x/\partial\theta \\ \partial y/\partial r & \partial y/\partial\theta \end{vmatrix} = \begin{vmatrix} \sigma_1 \cos\theta & -\sigma_1 r \sin\theta \\ \sigma_2(\rho\cos\theta + \sqrt{1-\rho^2}\sin\theta) & \sigma_2 r(-\rho\sin\theta + \sqrt{1-\rho^2}\cos\theta) \end{vmatrix}$$
$$= \sqrt{1-\rho^2}\sigma_1\sigma_2 r = \sqrt{|\Sigma|}\, r,$$

as can be easily checked.

To compute the integral of the Gaussian, we can suppose without loss of generality that $\mu = 0$ (since integrals are invariant under translation). The integral to compute is

$$\frac{1}{2\pi\sqrt{|\Sigma|}} \int_0^{2\pi} \mathrm{d}\theta \int_0^R \mathrm{d}r \sqrt{|\Sigma|} r e^{\frac{-r^2}{2}}.$$

It is equal to

$$\int_0^R r e^{\frac{-r^2}{2}} \mathrm{d}r = \int_0^R e^{\frac{-r^2}{2}} \mathrm{d}r^2/2 = \int_0^{R^2/2} e^{-r} \mathrm{d}r$$
$$= 1 - e^{-R^2/2}.$$

Thus, the desired cumulative function is given by

$$\boxed{F(r) = 1 - e^{-r^2/2}.}$$

It is invertible, an its inverse can be computed explicitly; in other words, if the probability p that a point falls at a certain Mahalanobis distance r from the Gaussian is known, then r can be computed explicitly: Put

$$p = 1 - e^{-r^2/2}.$$

Then

$$-r^2/2 = \ln(1-p),$$

or

$$r = \sqrt{-2\ln(1-p)}.$$

Hence,

$$\boxed{r = F^{-1}(p) = \sqrt{-2\ln(1-p)}.}$$

C. <u>n-dimensional Gaussian and multivariate normal densities</u>

Now, let us examine the n-dimensional case. The n-variate normal density with mean $\mu = (\mu_1, \mu_2, \ldots, \mu_n)$ and covariance matrix Σ is, with $z = (x_1, x_2, \ldots, x_n)$,

$$\boxed{p(z) = \frac{1}{(2\pi)^{n/2}\sqrt{|\Sigma|}} \exp\big[-\tfrac{1}{2}(z-\mu)\Sigma^{-1}(z-\mu)^t\big],}$$

where Σ is assumed symmetric and positive definite. The Mahalanobis distance of a point $z = (x_1, x_2, \ldots, x_n)$, to the Gaussian that composes the multivariate normal law above, is by definition given by r such that

$$\boxed{r^2 = (z-\mu)\Sigma^{-1}(z-\mu)^t.}$$

The arguments used above in the 2-dimensional case generalize immediately and show that:

a. The locus of the points $z = (x_1, x_2, \ldots, x_n)$ whose Mahalanobis distance is constant and equal to r is given by

$$\boxed{z = ruM + \mu,}$$

where $\Sigma = M^T M$ is the Choleski decomposition of Σ, and u is the set of points such that $uu^t = 1$ (that is, the unit sphere). This locus can be shown to be a n-dimensional ellipsoid.

b. The directions of the axis of the ellipsoid above are given by the eigenvectors of Σ, while their respective extents are given by

$$\boxed{r\sqrt{\lambda_i},}$$

$\lambda_1, \lambda_2, \ldots, \lambda_n$ being the eigenvalues of Σ.

It is more difficult (but not impossible) to generalize the 2-d argument above in order find the cumulative function of the multivariate normal law. Instead, we shall find another way. Here is a well known fact about normal laws: Let $X = (X_1, X_2, \ldots, X_n)$ be a n-dimensional random normal variable, with covariance matrix Σ. Assume furthermore that X is centered (that is, its mean is 0). Then

$$\boxed{\Sigma = E(X^T X).}$$

Indeed, if we denote $Z = (x_1, x_2, \ldots, x_n)$ and $C = \frac{1}{(2\pi)^{n/2}\sqrt{|\Sigma|}}$, we have to compute the integral

$$E(X^T X) = C \int\int \cdots \int Z^T Z \exp(-\frac{1}{2} Z\Sigma^{-1} Z^T) dx_1 dx_2 \cdots dx_n.$$

Let $\Sigma = M^T M$ be the Choleski decomposition of Σ. Notice that M is invertible since

$$\det(\Sigma) = \det(M^T)\det(M) = \det(M)^2,$$

and since Σ is positive by hypothesis (i.e. $\det(\Sigma) > 0$). Furthermore, $\det(M) = \sqrt{\det(\Sigma)}$. Let us put $Z = YM$, or what is the same, $Y = ZM^{-1}$ (with $Y = (y_1, \ldots, y_n)$). Thus,

$$Z\Sigma^{-1}Z^T = YMM^{-1}(M^T)^{-1}M^TY^T = YY^T.$$

The Jacobian of the transformation $Z = YM$ is constant and equal to $|M|$, hence the integral above becomes

$$C\int\int\cdots\int M^TY^TYM\exp(-\frac{1}{2}YY^T)|M|dy_1dy_2\cdots dy_n.$$

Extracting the matrices M^T and M from the integral, and taking into account that $|M| = \sqrt{\Sigma}$; the expression above becomes

$$M^T\left(\frac{1}{(2\pi)^{n/2}}\int\int\cdots\int Y^TY\exp(-\frac{1}{2}YY^T)dy_1dy_2\cdots dy_n\right)M.$$

The expression inside the parentheses can be shown to be the identity matrix (sketch of the proof: non-diagonal components cancel by symmetry, and diagonal elements can be computed using the volume of the k-dimensional sphere). Hence the whole expression is equal to $M^TIM = M^TM = \Sigma$, as was claimed.

Notice that the argument above shows that if a n-dimensional centered stochastic variable X is normally distributed, with covariance matrix Σ, and if $\Sigma = M^TM$ is the Choleski decomposition of M, then $Y = XM^{-1}$ is normally distributed with covariance matrix equal to I (that is, its gaussian is the product of n independent Gaussians with standard deviation $\sigma = 1$). By the way, this gives a good mean to implement stochastic generators of the variable X: it suffices to generate n samples $y_1, y_2, \ldots, y_n$ following separately the 1-d normal law with standard deviation $\sigma = 1$, and to multiply $Y = (y_1, y_2, \ldots, y_n)$ by M from the right; this gives a n-dimensional stochastic sample of the n-variate normal law with covariance matrix Σ.

Now, our aim is to find the cumulative function of the Gaussian which composes X as a function of the Mahalanobis distance r, that is, to find for each r, the proportion of instances of X that falls statistically at a Mahalanobis distance $r' \leq r$ from the Gaussian composing X. Without loss of generality, the variable X can be assumed to be centered. Recall that the Mahalanobis distance of a point $p = (p_1, \ldots, p_n)$ to this Gaussian is given by $r^2 = p\Sigma^{-1}p^T$. Letting p be an instance of X, we can see r^2 itself as as random variable, that is, $r^2 = X\Sigma^{-1}X^T$, giving rise the variable

$$r = \sqrt{X\Sigma^{-1}X^T}.$$

Thus, the problem amounts to find the cumulative function of r. It can be obtained from the cumulative function of r^2: indeed, if Φ_{r^2} is the cumulative function of r^2, that is, $\Phi_{r^2}(c) = P(r^2 \leq c)$, then the cumulative function Φ_r of r is

$$\Phi_r(c) = P(r^2 \leq c^2) = \Phi_{r^2}(c^2).$$

We need a last simplification stage: As above, put $X = YM$, where $\Sigma = M^T M$ is the Choleski decomposition of Σ. We have seen that Y is normal, with covariance matrix equal to I. Furthermore,

$$r^2 = X\Sigma^{-1}X^T = YM\Sigma^{-1}M^TY^T = YMM^{-1}(M^T)^{-1}M^TY^T = YY^T.$$

Henceforth, we have only to find the cumulative function of YY^T, or in other words, the cumulative function of the sum of the squares of n independently normally distributed Gaussian variables, each with mean 0 and standard deviation $\sigma = 1$. Such a distribution is known in the literature as the Chi-square distribution with n degrees of freedom, and its density has been obtained in an analytic form: If $N_1, \ldots, N_n$ are n independently distributed normal random variables, with mean 0 and standard deviation $\sigma = 1$, then the density distribution of $N_1^2 + \cdots + N_n^2$ is

$$f(x) = \frac{x^{(n-2)/2}e^{-x/2}}{2^{n/2}\Gamma(n/2)},$$

where Γ is the well known Euler Gamma function.

We can now conclude: The cumulative function of the n-variate normal law is

$$F(r) = \int_0^{r^2} \frac{x^{(n-2)/2}e^{-x/2}}{2^{n/2}\Gamma(n/2)}dx.$$

The change of variable $x = y^2$, $dx = 2y\,dy$, transforms this last expression into

$$F(r) = \int_0^{r} \frac{y^{n-1}e^{-y^2/2}}{2^{(n-2)/2}\Gamma(n/2)}dy.$$

If n is even, with $n = 2m$, then since $\Gamma(1) = 1$ and $\Gamma(x) = (x-1)\Gamma(x-1)$,

$$\boxed{F(r) = \frac{1}{2^{m-1}(m-1)!}\int_0^r y^{n-1}e^{-y^2/2}dy,}$$

and such an integral can be computed explicitly using integration by parts. If n is odd, with $n = 2m+1$, then since $\Gamma(1/2) = \sqrt{\pi}$ and $\Gamma(x) = (x-1)\Gamma(x-1)$,

$$F(r) = \frac{1}{2^{m-1}(m-1/2)(m-3/2)\cdots 3/2\sqrt{2\pi}}\int_0^r y^{n-1}e^{-y^2/2}dy.$$

Simplifying by the factors 2 leads to

$$\boxed{F(r) = \frac{1}{3\cdot 5\cdot 7\cdots(n-2)\sqrt{2\pi}}\int_0^r y^{n-1}e^{-y^2/2}dy.}$$

By elementary computations, this can be expressed explicitly by mean of the one-dimensional normal cumulative function $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-x^2/2} dx$.

<u>Remark:</u> In Matlab, the chi-square density with n-degrees of freedom is given by the function

$$\boxed{\text{chi2pdf(X,n),}}$$

while its cumulative distribution function is given by

$$\boxed{\text{chi2cdf(s,n).}}$$

and its inverse cumulative distribution by

$$\boxed{\text{chi2inv(t,n).}}$$

Hence, by what has been seen, the cumulative function of any n-variate normal centered law is

$$\boxed{\mathrm{F}(r) = \text{chi2cdf}(r^2,n)}$$

(r the Mahalanobis distance to the Gaussian). Conversely, the confidence level of such a law for a given confidence c (with $0 \leq c \leq 1$) is given by

$$\boxed{\sqrt{\text{chi2inv}(c,n)}.}$$

This is the Mahalanobis distance threshold under which a proportion c of points emitted by any n-variate normal law fall statistically below this threshold.