The Multinomial and Multivariate Normal Distributions

Guy Lebanon

August 24, 2010

The two most important vector RVs are the multinomial (discrete) and the multivariate normal (continuous).

The Multinomial Distribution

Definition 1. *The vector RV* $\vec{\boldsymbol{X}} = (X_1, \ldots, X_n)$ *has a multinomial distribution with parameters* $N \in \{1, 2, \ldots\}$ *and* $\theta \in \mathbb{R}^n$ *where* $\theta_i \geq 0$ *for all* i *and* $\sum_{i=1}^n \theta_i = 1$ *if*

$$p_{\vec{\boldsymbol{X}}}(x_1, \ldots, x_n) = \begin{cases} \binom{N}{x_1, \ldots, x_n} \theta_1^{x_1} \cdots \theta_n^{x_n} & \textit{if } x_1, \ldots, x_n \textit{ are non-negative integers that sum to } N \\ 0 & \textit{otherwise} \end{cases}.$$

Here $\binom{N}{x_1, \ldots, x_n} = \frac{N!}{x_1! \cdots x_n!}$ *is the multinomial coefficient.*

The multinomial distribution applies when we have a random experiment with n possible results, each occurring with probability θ_i. The experiment is repeated N times and $X_1, \ldots, X_n$ measure the number of times the different outcomes occurred. Since there are N experiment the total number of outcomes has to be $x_1 + \cdots + x_n = N$ and since θ_i are the probability of getting outcome i in one experiment, $\sum_i \theta_i = 1$.

To see why the pmf follows from the above description consider $p_{\vec{\boldsymbol{X}}}(x_1, \ldots, x_n)$ which is the probability of getting x_1 times outcome 1, and so on until x_n times outcome n in a series of N independent experiments. $p_{\vec{\boldsymbol{X}}}(x_1, \ldots, x_n)$ is $\theta_1^{x_1} \cdots \theta_n^{x_n}$ (which is the probability of an ordered sequence of outcomes with the necessary property - x_1 times result 1 and so on) times the number of ways to obtain ordered sequences of x_1 times outcome 1 etc. That number is precisely the multinomial coefficient

$$= \binom{N}{x_1} \binom{N - x_1}{x_2} \binom{N - x_1 - x_2}{x_3} \cdots \binom{x_n}{x_n} = \frac{N!}{x_1!(N - x_1)!} \frac{(N - x_1)!}{x_2!(N - x_1 - x_2)!} \frac{(N - x_1 - x_2)!}{x_3!(N - x_1 - x_2 - x_3)!} \cdots \frac{1}{1}$$
$$= \frac{N!}{x_1! \cdots x_n!} = \binom{N}{x_1, \ldots, x_n}$$

Example: The roulette has 38 possible outcomes, 18 red, 18 black and 2 green. Thus playing the roulette is an experiment with $\theta_1 = \theta_2 = 18/38$ and $\theta_3 = 2/38$. If we play the roulette 10 times, the probability that we get 4 red outcomes, 2 black outcomes and 4 green is

$$p_{X_1, X_2, X_3}(4, 2, 4) = \frac{10!}{4!2!4!} (18/38)^4 (18/38)^2 (2/38)^4$$

The multinomial coefficient is present since there are $\frac{10!}{4!2!4!}$ ways to play 10 times and obtain 4 red 2 black and 4 green outcomes.

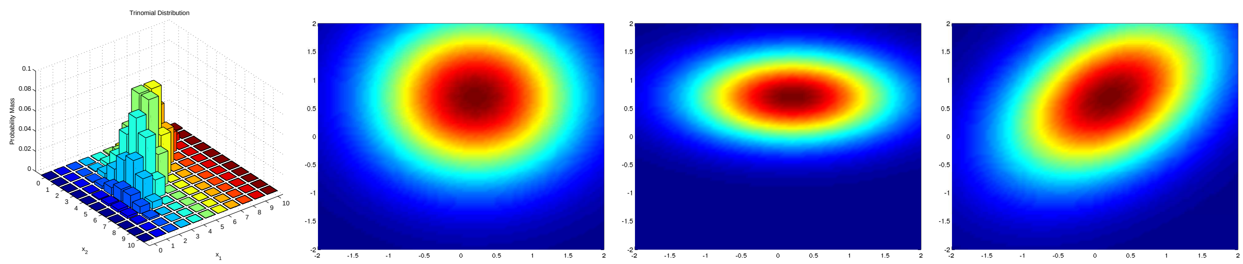

Figure 1: Probability of the multinomial ($n = 3$) distribution as a function of x_1, x_2 (left) and density of the multivariate normal for the three special cases.

The Multivariate Normal Distribution

Definition 2. *The vector RV* $\vec{\boldsymbol{X}} = (X_1, \ldots, X_n)$ *has the multivariate normal distribution with parameters* $\vec{\boldsymbol{\mu}} \in \mathbb{R}^n$ *and* Σ *(a symmetric matrix of size* $n \times n$ *with positive eigenvalues) if*

$$f_{\vec{\boldsymbol{X}}}(x_1, \ldots, x_n) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \Sigma}} e^{-\frac{1}{2}(\vec{\boldsymbol{x}}-\vec{\boldsymbol{\mu}})^\top \Sigma^{-1}(\vec{\boldsymbol{x}}-\vec{\boldsymbol{\mu}})}.$$

Since the determinant of a matrix with all positive eigenvalues is positive - there is no problem with taking its square root. The term in the exponent may be written in scalar form as:

$$-\frac{1}{2}(\vec{\boldsymbol{x}} - \vec{\boldsymbol{\mu}})^\top \Sigma^{-1}(\vec{\boldsymbol{x}} - \vec{\boldsymbol{\mu}}) = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - \mu_i)[\Sigma^{-1}]_{ij}(x_j - \mu_j)$$

In a way similar to the one-dimensional normal RV, the vector $\boldsymbol{\mu}$ is a vector of expectations $\mathsf{E}(X_i) = \mu_i$ and the matrix Σ is the matrix of covariances and variances

$$[\Sigma]_{ij} = \begin{cases} \mathsf{Var}(X_i) & i = j \\ \mathsf{Cov}(X_i, X_j) & i \neq j \end{cases}$$

Several important special cases:

1. If Σ is the identity matrix, its determinant is 1, its inverse is the identity as well, and the exponent becomes $-\sum_{i=1}^{n}(x_i - \mu_i)^2/2$ which indicates that the pdf factors into the product of n pdf functions of normal RVs, with means μ_i and variance $\sigma_i^2 = 1$. Thus in this case, the multivariate normal vector RV is a collection of n independent RVs $X_1, \ldots, X_n$, each being normal with parameters $\mu_i, \sigma_i^2 = 1$.

2. If Σ is diagonal matrix with elements $[\Sigma]_{ij} = \sigma_i^2$, then its inverse is a diagonal matrix with elements $[\Sigma^{-1}]_{ij} = 1/\sigma_i^2$ and its determinant is the product of the diagonal elements $\prod_i \sigma_i^2$. Again, the term in the exponent of the pdf factors into a sum which indicates that the pdf factors into a product of marginal pdfs for each of the variables X_i. Thus, again we have that $X_1, \ldots, X_n$ are independent normal RV with parameters (μ_i, σ_i^2) (verify!).

3. In the general case, the shape of the pdf (its contour levels) are determined by the exponent (since the term $\frac{1}{(2\pi)^{n/2}\sqrt{\det \Sigma}}$ is constant as a function of $\vec{\boldsymbol{x}}$) which is a quadratic form $-\sum_i \sum_j (x_i - \mu_i)\Sigma_{ij}^{-1}(x_j - \mu_j)$. As a result, the contour levels of the pdf will be elliptical with a center determined by $\vec{\boldsymbol{\mu}}$ and shape determined by Σ^{-1}. If $\Sigma^{-1} = cI$ the ellipse will be spherical. If Σ^{-1} is diagonal with different elements on the diagonal we get a (potentially) non-spherical axis aligned ellipse.

As a consequence of (2) above we see that if $X_1, \ldots, X_n$ are uncorrelated multivariate normal RVs (with covariance 0) they are also independent. This is in contrast to the general case where zero covariance or correlation does not necessarily imply independence.