Statistics 135

Chapter 4

The Multivariate Normal

Chris Drake

Department of Statistics

University of California, Davis

The univariate normal distribution

The normal density

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \; exp\frac{1}{2}\Big(-(\frac{x-\mu}{\sigma})^2\Big) \qquad -\infty < x < \infty$$

1 The mean is μ and the variance is σ^2.

2 Approximately 68% of values (.683) fall in the interval $\mu \pm \sigma$, approximately 95% of values (.954) fall in the interval $\mu \pm 2\sigma$.

3 The normal distribution is symmetric about the mean..

4 The third central moment, the skewness $E[(X-\mu)^3] = 0$ since the normal is symmetric and the fourth central moment, the kurtosis is $E[(X-\mu)^4] = 3$; a kurtosis greater than 3 indicates a more peaked distribution, a kurtosis less than 3 indicates a flatter distribution.

5 Extreme values more than 3 standard deviations from the mean are rare for the normal distribution.

The multivariate normal distribution

the multivariate density:

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2} \mid \mathbf{\Sigma} \mid^{1/2}} \; e^{\frac{1}{2}\left(-(\mathbf{x}-\mu)'\mathbf{\Sigma}^{-1}(\mathbf{x}-\mu)\right)}$$

1 for $x_1, .., x_p$ we have $-\infty < x < \infty$

2 μ is the mean vector and $\mathbf{\Sigma}$, positive definite, is the covariance matrix.

3 $\mathbf{\Sigma^{-1}}$ is also positive definite and has the same eigenvectors as $\mathbf{\Sigma}$; the eigenvalue associated with $\mathbf{e_i}$ is λ_i^{-1}.

4 Since $\mathbf{\Sigma}$ is positive definite, the eigenvalues are all positive.

5 $(\mathbf{x}-\mu)'\mathbf{\Sigma^{-1}}(\mathbf{x}-\mu) = c^2$ is a quadratic form, an ellipsoid centered at μ with axis $\pm c\sqrt{\lambda_i}\mathbf{e_i}$, the axis of this ellipsoid are in the direction of the eigenvectors of $\mathbf{\Sigma}$. The distribution of the quadratic form $(\mathbf{X}-\mu)'\mathbf{\Sigma^{-1}}(\mathbf{X}-\mu)$ is a χ^2_p where p is the rank of $\mathbf{\Sigma}$.

Important properties

$\mathbf{X} = (X_1, .., X_p)$ a p-variate normal vector with mean μ and covariance matrix $\mathbf{\Sigma}$ then

1 All linear combinations of $\mathbf{X}$ are normally distributed; this implies that all components X_i are normally distributed. The linear combination is $\mathbf{c}'\mathbf{X} = X_i$ with $\mathbf{c}' = (0, .., 0, 1, 0, .., 0)$ all zeros except a 1 in the i^{th} position. The proof can be done through a transformation of variables integral.

2 From (1) you can deduce that all subsets of components of $\mathbf{X}$ are normally distributed. For the subset X_i, X_j, X_k use $\mathbf{C}$ with row 1 all zeros except 1 in the i^{th} position, row 2 all zeros except in the j^{th} position and row 3 all zeros except 1 in the k^{th} position.

3 If $\mathbf{\Sigma}$ is diagonal, all components of $\mathbf{X}$ are independent. Note, for a diagonal covariance matrix the covariances and therefore, correlations, are all zero. For the multivariate normal, in this case the density factors into p components and therefore, for the multivariate normal uncorrelated implies independence.

4 For the partition $\mathbf{X} = (\mathbf{X_1}, \mathbf{X_2})$ the conditional distribution of $\mathbf{X_1} \mid \mathbf{X_2} = \mathbf{x_2}$ is normal with mean and covariance

$$\mu_{\mathbf{X_1}|\mathbf{X_2}=\mathbf{x_2}} = \mu_1 + \mathbf{\Sigma_{12}\Sigma_{22}^{-1}}(\mathbf{x_2} - \mu_\mathbf{2})$$

and

$$\Sigma_{\mathbf{X_1}|\mathbf{X_2}=\mathbf{x_2}} = \mathbf{\Sigma_{11}} - \mathbf{\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}}$$

5 if X_1 is univariate the above simplifies to

$$\mu_{X_1|\mathbf{X_2}=\mathbf{x_2}} = \mu_1 + (\sigma_{12}, .., \sigma_{1p})\mathbf{\Sigma_{22}^{-1}}(\mathbf{x_2} - \mu_\mathbf{2})$$

and

$$\sigma_{X_1|\mathbf{X_2}=\mathbf{x_2}} = \sigma_\mathbf{11} - (\sigma_\mathbf{12}, .., \sigma_\mathbf{1p})\mathbf{\Sigma_{22}^{-1}}(\sigma_\mathbf{12}, .., \sigma_\mathbf{1p})'$$

note: $(\sigma_{12}, .., \sigma_{1p})\Sigma_{22}^{-1}(\sigma_{12}, .., \sigma_{1p})' > 0$ since $\mathbf{\Sigma_{22}}$ is positive definite, so the variance of $X_1 | \mathbf{X_2} = \mathbf{x_2}$ is smaller than the variance of the marginal distribution. If $X_1, \mathbf{X_2}$ are uncorrelated then $(\sigma_{12}, .., \sigma_{1p})$ are zero.

6 For a multivariate normal $\mathbf{X_1}, \mathbf{X_2}$ are independent if and only if they are uncorrelated.

Maximum likelihood Estimation

$\mathbf{X_1}, \mathbf{X_2}, .., \mathbf{X_n}$ a sample from a p-variate normal distribution. The density is $f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}|\mathbf{\Sigma}|^{1/2}} \; e^{\frac{1}{2}\left(-(\mathbf{x}-\mu)'\mathbf{\Sigma}^{-1}(\mathbf{x}-\mu)\right)}$. The likelihood is given by

$$\mathcal{L}(\mu, \mathbf{\Sigma}; \mathbf{x}) = (\mathbf{x}) = \prod_{j=1}^{n} \left\{ \frac{1}{(2\pi)^{p/2} \mid \mathbf{\Sigma} \mid^{1/2}} \; e^{\frac{1}{2}\left(-(\mathbf{x_j}-\mu)'\mathbf{\Sigma}^{-1}(\mathbf{x_j}-\mu)\right)} \right\}$$

where $\mathcal{L}$ is a function of μ and $\mathbf{\Sigma}$ and the $\mathbf{x_j}$ are the sample values (the data). Once we have obtained a sample, $\mathbf{X_j} = \mathbf{x_j}$ is a known vector of numbers, not a random quantity; the parameters are still unknown and the likelihood is the density of the sample when it is a function of the unknown parameters.

If the sample consists of independently selected observations, then the probability of the sample is just the product of the densities. The likelihood is used to obtain estimates of the unknown parameters. We choose those values as estimates that maximize the likelihood.

The estimates are called maximum likelihood estimators are.

For the multivariate normal

$$\hat{\mu}_{MLE} = \bar{\mathbf{X}} \qquad \textit{and} \qquad \hat{\mathbf{\Sigma}}_{MLE} = \mathbf{S_n}$$

and their observed values are

$$\hat{\mu}_{MLE} = \bar{\mathbf{x}} \qquad \textit{and} \qquad \hat{\mathbf{\Sigma}}_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}-\bar{\mathbf{x}})(\mathbf{x}-\bar{\mathbf{x}})'$$

The maximum likelihood estimates are also the sufficient statistics.

This can be determined by inspection of the likelihood. It depends on the data only through $\bar{\mathbf{X}}$ and $\mathbf{S} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}-\bar{\mathbf{x}})(\mathbf{x}-\bar{\mathbf{x}})'$. For sufficiency we require that the likelihood factors into two parts, one part that contains the data only through the sufficient statistics and the parameters and the other part may contain the data in other forms that does not depend on the parameters.

Sampling distribution of $\bar{\mathbf{X}}$ and $\mathbf{S}$

Recall: in univariate case

$$\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n}) \qquad and \qquad (n-1)S^2 = n\hat{\sigma}^2 \sim \sigma^2\chi^2_{n-1}$$

In the multivariate case we have

$$\bar{\mathbf{X}} \sim \mathcal{N}_p(\mu, \frac{1}{n}\mathbf{\Sigma}) \qquad and \qquad (n-1)\mathbf{S} \sim Wishart_{n-1}$$

where $(n-1)$ are the degrees of freedom; $\bar{\mathbf{X}}$ *and* $\mathbf{S}$ *are statistically independent*

Properties of the Wishart

1 $\mathbf{A_1} \sim W_{m_1}(\mathbf{A_1} \mid \mathbf{\Sigma})$ and $\mathbf{A_2} \sim W_{m_2}(\mathbf{A_2} \mid \mathbf{\Sigma})$ then $\mathbf{A_1} + \mathbf{A_2} \sim W_{m_1+m_2}(\mathbf{A_1} + \mathbf{A_2} \mid \mathbf{\Sigma})$

2 $\mathbf{A} \sim W_m(\mathbf{A} \mid \mathbf{\Sigma})$ then $\mathbf{CAC'} \sim W_m(\mathbf{CAC'} \mid \mathbf{C\Sigma C'})$

Note: the Wishart behaves similarly as the χ^2-distribution.