1. Probability

There are several different situations to consider. They are the discrete case, the continuous case, single variables, and multi variables or the vector case. First consider some probability basics.

Let A be an event (set). Then P(A) is the probability that the event occurs. If B is another event then P(A/B) is the probability that event A occurs given that event B occurs. The probability of A and B (P(A&B)) occurring (the joint event) is $P(A \cap B)=P(A/B)P(B)=P(B/A)P(A)$ [Nadler and Smith, 1993, pp. 284; Downing and Clark, 1997, pp. 63]. If A and B are independent then $P(A \cap B) = P(A)*P(B)$.

1.1 Continuous Random Variables

Let g be a random variable (RV). That is g takes on values z a real number with a given probability. The probability density function is $p(z)$. For a continuous RV $\int p(z)dz = 1$ where the integration is over all possible values of p(z). The cumulative distribution function is $F(z)=P(g \le z)$.

The conditional probability rules apply for probability density functions. For example, $p(z_1,z_2) = p(z_1 / z_2)p(z_2) = p(z_2 / z_1)p(z_1)$ [Moon and Stirling, 2000, pp. 36, 486; Nadler and Smith, 1993, pp. 346].

$p(z_1/z_2) = p(z_1,z_2) / p(z_1)$, and $p(z_1 / z_2) = \dfrac{p(z_2/z_1)p(z_1)}{p(z_2)}$.

If c_i is a class event then $P(c_i/z) = \dfrac{p(z/c_i)P(c_i)}{p(z)}$

The expected value of g is the value we expect to get if we have many samples of g. This is written $E\{g\}=\mu$ and is called the mean value. It is given by $E\{g\} = \int_z zp(z)dz$ for the continuous case. The variance is a measure of the spread of the values of g from the mean value. This may also be called the uncertainty of g [Downing and Clark, 1997, pp. 81]. The variance is $var(g)=\sigma^2=E\{(g-\mu)^2\}$. It may be written $\sigma^2 = \int_z (z-\mu)^2 p(z)dz$ for the continuous case. The quantity σ is called the standard deviation. Another formula is $\sigma^2=E\{g^2\}-E\{g\}^2 = E\{g^2\}-\mu^2$.

Consider now the estimation of the mean and variance. If we have a finite number of samples of g then we may estimate the mean value as $\bar{E}\{g\} = \bar{\mu} = \dfrac{\sum_{i=1}^{N} z_i}{N}$. In a similar way the variance may be computed as $\bar{\sigma}^2 = \dfrac{\sum_{i=1}^{N} (z_i - \mu)^2}{N}$ [Downing and Clark, 1997, pp. 19]. An unbiased estimate of the variance is $s_2^2 = \dfrac{\sum_{i=1}^{N} (k_i - \mu)^2}{N-1}$.

These ideas may be extended to measurement vectors. Instead to RV g with values represented by z we have a vector $\mathbf{u}$=(u(1),u(2),…, u(D)) of length D. The first component of $\mathbf{u}$ is u(1). We use bold type to indicate a vector. In a similar manner the mean value is also a vector of the same length. The probability density function is $p(\mathbf{u})=(p_1, p_2, \ldots, p_D)$. Each p_i operates on its respective component of $\mathbf{u}$ namely $p_i(u(i))$. The covariance matrix is defined for $\mathbf{u}, \boldsymbol{\mu}$ both D-dimensional vectors. If $\mathbf{u}, \boldsymbol{\mu}$ are put in column matrices then $\Sigma=E\{(\mathbf{u}-\boldsymbol{\mu})(\mathbf{u}-\boldsymbol{\mu})^t\}$ is a square matrix (D by D) with elements $\sigma_{ij}=E\{(u(i)-\mu(i)(u(j)-\mu(j))^t\}$ [Nadler and Smith, 1993, pp 286]. The covariance matrix is always symmetric and positive semidefinite [Duda and Hart, 1973, pp.23].

1.2 Discrete Case

In the discrete case there are only a finite number of possible values for the RV g. For example, if g represents the number on a thrown dice then p(1)=P(g=1) and $p(3)=P(g=3)=\frac{1}{6}$. The probability density function is $p(z)=P(g=z)$ [Downing and Clark, 1997, pp. 74]. This is also called the probability mass function. For a discrete RV $\sum_{i=1}^{N} p(z_i) = 1$. The expectation is given by $E\{g\} = \sum_{i=1}^{N} z_i f(z_i)$. The variance is $\sigma^2 = \sum_{i=1}^{N} (z_i\text{-}\mu)^2 p(z_i)$ for the discrete case. For a vector valued RV the probability density function is $p(\mathbf{u})=P(g_1 = u(1),\ g_2 = u(2),\ \ldots, g_D = u(D))$.

Consider the following example.

Example

The following are samples of two measurements taken from class A and class B

Class A

0	5
1	4
2	7
1	4
4	3
2	4
1	2
3	2

Class B

8	3
7	2
6	2
5	1
8	2
7	0
4	3
8	1

The mean vectors of classes A and B are given below.

$\boldsymbol{\mu}_A = (1.7500,\ 3.8750)$

$\boldsymbol{\mu}_B = (6.6250,\ 1.7500)$

The covariance matrix is $\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_1\,\sigma_2 \\ \sigma_1\,\sigma_2 & \sigma_2^2 \end{bmatrix}$. The covariance matrices of classes A and B are given below.

$\Sigma_A = \begin{bmatrix} 1.4375 & -.53125 \\ -.53125 & 2.3594 \end{bmatrix}$. The inverse is $\Sigma_A^{-1} = \begin{bmatrix} .75879 & .17085 \\ .17085 & 4.6231 \end{bmatrix}$.

The determinate of the cov matrix is $|\Sigma_A| = 3.1094$.

$\Sigma_B = \begin{bmatrix} 1.9844 & -.21875 \\ -.21875 & .93750 \end{bmatrix}$. The inverse is $\Sigma_B^{-1} = \begin{bmatrix} .51724 & .12069 \\ .12069 & 1.0948 \end{bmatrix}$

1.3 Means and Variance Calculation

Consider the calculation of the mean and covariance matrices when we have a finite number of samples. For the case of the single variable we

Consider now the estimation of the mean and variance. We may estimate the mean value as $\bar{E}\{g\} = \bar{\mu} = \frac{\sum_{i=1}^{N} z_i}{N}$. In a similar way the variance may be computed as $\bar{\sigma}^2 = \frac{\sum_{i=1}^{N}(z_i - \bar{\mu})^2}{N}$ [Downing and Clark, 1997, pp. 19]. An unbiased estimate of the variance is $s_2^2 = \frac{\sum_{i=1}^{N}(z_i - \bar{\mu})^2}{N\text{-}1}$. Note that $s_2^2 = \frac{N}{N\text{-}1}\sigma^2$. As N increases the two values become close.

Consider a vector measurement vector. The maximum likelihood estimation of the covariance matrix is $\bar{\Sigma} = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{u}_i - \bar{\boldsymbol{\mu}})(\mathbf{u}_i - \bar{\boldsymbol{\mu}})^t$ where $\bar{\boldsymbol{\mu}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{u}_i$ and N is the number of samples [Duda and Hart, 1973, pp. 49]. The unbiased estimate of the covariance matrix is $S = \frac{1}{N\text{-}1}\sum_{i=1}^{N}(\mathbf{u}_i - \hat{\boldsymbol{\mu}})(\mathbf{u}_i - \hat{\boldsymbol{\mu}})^t$ [Nadler and Smith, 1993, pp. 287].

1.4 Normal Density Function

The normal density function with one variable is $p(z)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{(z\text{-}\mu)^2}{2\sigma^2}\right]$. The notation $N(z,\mu,\sigma)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{(z\text{-}\mu)^2}{2\sigma^2}\right]$ maybe used to denote the normal density function where the parameters of the function are μ,σ. The multivariate normal density function is $p(\mathbf{u})=\frac{1}{(2\pi)^{\frac{D}{2}}|\Sigma|^{\frac{1}{2}}}\exp\left[-\frac{1}{2}(\mathbf{u}\text{-}\boldsymbol{\mu})^t\Sigma^{-1}(\mathbf{u}\text{-}\boldsymbol{\mu})\right]$ where D is the length of the **u** vector. Here the quantities are in column matrices representing vectors, $|\Sigma|$ is the determinate of the covariance matrix and Σ^{-1} is its inverse [Nadler and Smith, 1993, pp. 335]. The quantity $(\mathbf{u}\text{-}\boldsymbol{\mu})^t\Sigma^{-1}(\mathbf{u}\text{-}\boldsymbol{\mu})$ from the equation $p(\mathbf{u})=\frac{1}{(2\pi)^{\frac{D}{2}}|\Sigma|^{\frac{1}{2}}}\exp\left[-\frac{1}{2}(\mathbf{u}\text{-}\boldsymbol{\mu})^t\Sigma^{-1}(\mathbf{u}\text{-}\boldsymbol{\mu})\right]$ is called the squared Mahalanobis distance from $\mathbf{u}$ to $\boldsymbol{\mu}$ [Duda and Hart, 1973, pp. 24; Nadler and Smith, 1993, pp. 351]