1 Probability basics

In statistical signal processing, manipulation of probability expressions is very important. The two tools from probability theory that we will be using frequently to manipulate the expressions are:

- Law of total probability
- Probability chain rule

1.1 Law of total probability

If $A_1, A_2, \ldots, A_n$ forms a partition of the sample space S, then the probability of an event B is given by

$$P(B) = \sum_{i=1}^{n} P(B \cap A_i)$$

Using the definition of conditional probability, we can rewrite the above expression as

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$$

Thus, the probability of the event B is the sum of the conditional probabilities $P(B|A_i)$ weighted by the probability $P(A_i)$.

1.2 Probability chain rule

For any events A and B, we have

$$P(A \cap B) = P(A)P(B|A)$$

More generally, for any events $A_1, A_2, \ldots, A_n$ we get

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = \prod_{k=1}^{n} P(A_k| \cap_{j=1}^{k-1} A_j)$$

You can use induction to show the general version of the probability chain rule.

1.3 Practice problems on probability basics

1. A computer manufacturer uses chips from three sources. Chips from sources A, B, and C are defective with probabilities 0.005, 0.001, and 0.010, respectively. If a randomly selected chip is found to be defective, find:

   (a) The probability that the manufacturer was A

   (b) The probability that the manufacturer was C

   Assume that the proportions of chips from A, B, and C are 0.5, 0.1, and 0.4, respectively.

2. A box contains three coins. One of the coins is a two-headed coin, the second is a fair coin, and the third is a biased coin with $P(head) = p$. One of the coins is picked at random and flipped:

   (a) Find the probability that the coin shows head. What is it's maximum value?

   (b) If the coin shows head, find the probability that it is the fair coin. What are it's maximum and minimum values?

3. There is a screening test for prostate cancer that looks at the level of PSA (prostate specific antigen) in the blood. There are a number of reasons besides prostate cancer that a man can have elevated PSA levels. In addition, many types of prostate cancer develop so slowly that they are never a problem. Unfortunately, there is currently no test to distinguish the different types and using the test is controversial because it's hard to quantify the accuracy rates and the harm done by false positives. For this problem, we will call a positive test a true positive if it catches a dangerous type of prostate cancer. Also, we will assume the following numbers:

   - Rate of dangerous type of prostate cancer among men over 50 = 0.0005
   - True positive rate for the test = 0.9
   - False positive rate for the test = 0.01

   Suppose you randomly select a man over 50 and perform a screening test.

   (a) What is the probability that the man has a dangerous type of the disease given that he had a positive test?

   (b) What is the probability that the man has a dangerous type of the disease given that he had a negative test?

4. Jack and Jill are involved in a duel. The rules of the duel are that they are to pick up their arrows and shoot at each other simultaneously. If one or both are hit, then the duel is over. If both shots miss, then they repeat the process. Suppose that the results of the shots are independent and that each shot of Jack will hit Jill with probability p_1 and each shot of Jill will hit Jack with probability p_2. What is:

   (a) the probability that Jack is not hit?

(b) the probability that both duelists are hit?

(c) the probability that the duel ends after the n^{th} round of shots?

5. Consider a batch of 15 plates containing 5 blue plates, 5 red plates, and 5 green plates. Jumble up the batch of plates and place all 15 of the plates around a circular table, with one plate per seat.

   (a) A plate is called "isolated" if its color does not agree with either of the nearby plates (i.e., if it has a different color than the plate to its right and a different color than the plate to its left). Let X denote the number of isolated plates. Find $E(X)$.

   (b) A plate is called "semi-happy" if its color agrees with exactly one (but not both) of the nearby plates (i.e., if its color agrees with the color of the plate on its left or on its right, but not both). Let Y denote the number of semi-happy plates. Find $E(Y)$.

   (c) A plate is called "joyous" if its color agrees with both of the nearby plates (i.e., if its color agrees with the color of the plate on its left and on its right). Let Z denote the number of joyous plates. Find $E(Z)$.

2 Linear Algebra basics

Proficiency in linear algebra is required to have a solid foundation in machine learning. The two linear algebra concepts that we will be using repeatedly in this course are:

- Decomposition of matrices
- Derivatives of vectors and matrices

In this discussion we will focus on the decomposition of matrices, namely:

- Eigendecomposition
- Singular value decomposition

2.1 Eigendecomposition

Every real symmetric matrix, $A \in \mathbf{R}^{n \times n}$, can be factored as

$$A = Q \Delta Q^T$$

where $\Delta = diag(\lambda_1, \lambda_2, \cdots, \lambda_n) \in \mathbf{R}^{n \times n}$. More specifically we have,

- The set $\{\lambda_1, \lambda_2, \cdots, \lambda_n\}$ are called the eigenvalues of A and can be found by solving the equation

$$det(A - \lambda I) = 0$$

- The columns of Q are an orthonormal set of n eigenvectors and can be found by solving the equation

$$Av_i = \lambda_i v_i$$

2.2 Singular value decomposition

Let $A \in \mathbf{R}^{m\times n}$ be a matrix with rank r, then there exists orthogonal matrices $U \in \mathbf{R}^{m\times m}$ and $V \in \mathbf{R}^{n\times n}$ such that

$$A = U\Sigma V^T$$

where $\Sigma = \begin{bmatrix} S & 0 \\ 0 & 0 \end{bmatrix}$, $S = diag(\sigma_1, \sigma_2, \cdots, \sigma_r) \in \mathbf{R}^{r\times r}$, and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$. More specifically, we have

- The set $\{\sigma_1, \sigma_2, \cdots, \sigma_r\}$ are called the non-zero singular values of A and can be calculated as

$$\sigma_i(A) = \lambda_i^{\frac{1}{2}}(A^T A) = \lambda_i^{\frac{1}{2}}(AA^T)$$

- The columns of U are called the left singular vectors of A (and are the orthonormal eigenvectors of AA^T).
- The columns of V are called the right singular vectors of A (and are the orthonormal eigenvectors of $A^T A$).

2.3 Practice problems on linear algebra basics

1. Show the following properties for matrices
   (a) If $b^T Ab > 0$ for all $b \in \mathbf{R}^n$, then all eigenvalues of A are positive.
   (b) If $A \in \mathbf{R}^{n\times n}$ is an orthogonal matrix, then all eigenvalues of A have norm 1.
   (c) If $A \in \mathbf{R}^{m\times n}$ is a matrix with rank r, then

$$\sigma_i(A) = \lambda_i^{\frac{1}{2}}(AA^T)$$

2. Determine if the following statements are true or false. If they are true, give a justification. If they are false, give a counterexample.
   (a) If $\mathbf{A}$ is an orthogonal matrix, then $\mathbf{A}^T$ is also an orthogonal matrix.
   (b) If $\mathbf{A}$ and $\mathbf{B}$ are orthogonal $n \times n$ matrices, then $\mathbf{AB}$ is orthogonal.
   (c) If $\mathbf{A}$ and $\mathbf{B}$ are orthogonal $n \times n$ matrices, then $\mathbf{A} + \mathbf{B}$ is orthogonal.
   (d) If the column vectors of $\mathbf{A}$ are orthonormal, then the row vectors of $\mathbf{A}$ must also be orthonormal.