Stochastic Processes:
A review of probability theory
CIS002-2 Computational Alegrba and Number Theory

David Goodwin

david.goodwin@perisic.com

University of Bedfordshire

10:00, Friday 09th March 2012

Outline

Outline

Random variables and mutually exclusive events

- Probability theory is used to describe a situation in which we do not know the precise value of a variable, but may have an idea of the likelihood that it wil have one of a number of possible values.
- let us call the unknown quantity X, referred to as a random variable.
- We describe the likelihood X will have one of all the possible values as the probability, $0 < X < 1$.
- The various values of X, and of any random variable, are an example of mutually exclusive events.
- The total probability that one of two or more mutually exclusive events occurs is the sum of the probabilities for each event

Random variables: roll of dice.

- The sum of the probabilities for all the mutually exclusive possible values must always be unity.
- If a die is fair, then all the possible values are equally likely, therefore the probability for each event is 1/6.
- in this example, x is a discrete random variable.

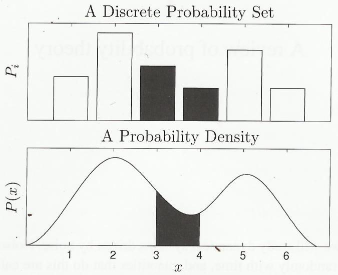

Figure 1.1. An illustation of summing the probabilities of mutually exclusive events, both for discrete and continuous random variables.

If we want to know the probability for X, being the roll of a die, being in the range from 4 to 6, we sum all the probabilities for the values from 4 to 6, illustrated in the figure 1.1 above.

University of Bedfordshire

Continuous random variables

- If X could take the value of any real number, then we say X is a continuous random variable.
- If X is a continuous random variable, the probability is now a function of x, where x ranges over the values of X.
- This type of probability is called a probability density, denoted $P(x)$.
- The probability for X to be in the range $x = a$ to $x = b$ is now thw area under $P(x)$ from $x = a$ to $x = b$

$$Prob(a < X < b) = \int_a^b P(x)dx$$

- Thus, the integration (area under the curve) of $P(x)$ over the whole real number line (from $-\infty$ to ∞) must be untiy, since X must take on one of these values.

$$\int_{-\infty}^{\infty} P(x)dx = 1$$

Statistical definitions

- The average of X, also known as the mean, or expectation value, of X is defined by

$$\langle X \rangle \equiv \int_{-\infty}^{\infty} P(x) x dx$$

- If $P(x)$ is symmetric about $x = 0$, then it is not difficult to see that the mean of X is zero.
- If the density is symmetric about any other point then the mean is the value at this point.
- The varience of X is defined as

$$V_X \equiv \int_{-\infty}^{\infty} P(x)(x - \langle X \rangle)^2 dx = \langle X^2 \rangle - \langle X \rangle^2$$

- The standard deviation of X, denoted by σ_X and defined as the square root of the varience, is a measure of how broad the probability density for X is - that is, how much we expect X to deviate from the mean value.

The Gaussian

- An important example of a probaability density is the Gaussian, given by

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- The mean of this Gaussian is $\langle X \rangle = \mu$ and the variance is $V(x) = \sigma^2$.
- A plot of this probability density is shown in the figure 1.2 below.

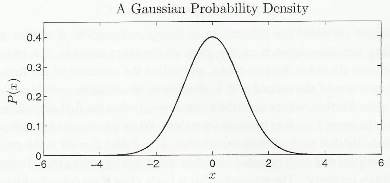

Figure 1.2. A Gaussian probability density with variance $V = 1$, and mean $\langle X \rangle = 0$.

University of Bedfordshire

Outline

Independence

- Two random variables are referred to as independent if neither of their probability densities depends on the value of the other variable.
- The probability that two independent random events occur is the product of their probabilities.
- This is true for discrete and continuous independent random variables.
- In the case of continuous independent random variables we speak of the joint probability density.

$$P(x, y) = P_X(x)P_Y(y)$$

- We can take this further and ask what the probability that X falls within the interval $[a, b]$ and Y falls in the interval $[c, d]$. This is

$$\int_a^b \int_c^d P(x, y)dydx = \int_a^b P_X(x)dx \int_c^d P_Y(y)dy$$

- It is also worth noting that when two variables are independent, then the expectation value of their product is simply the product of their expectation values

$$\langle XY \rangle = \langle X \rangle \langle Y \rangle$$

Outline

Dependent random variables

- Two random variables are referred to as dependent if their joint probability density, $P(x, y)$, does not factor into the product of their respective probability densities.
- To obtain the probability density for one variable alone (say X), we integrate the joint probability density over all values of the other variable (in this case Y).
- For each value of X, we want to know the total probability summed over all the mutually exclusive values that Y can take.
- In this context, the probability densities for a single variable are referred to as the marginals of the joint density.
- If we know nothing about Y, then our probability density for X is just the marginal

$$P_X(x) = \int_{-\infty}^{\infty} P(x, y) dy$$

- If X and Y are dependent, adn we learn the value of Y, then in general this will change our probability density for X (and vice versa). The probability density for X given that we know that $Y = y$, is written $P(x \mid y)$ and is referred to as the conditional probability density for X given Y.

Dependent random variables

- To see how to calculate this conditional probability, we note first that $P(x, y)$ with $y = a$ gives a relative probability for different values of x gievn that $Y = a$.
- To obtain the conditional proability density for X given that $Y = a$, all we have to do is divide $P(x, a)$ by its integral over all values of x. This ensures that the integral of the conditional probability is unity

$$P(x \mid y) = \frac{P(x, y)}{\int_{-\infty}^{\infty} P(x, y) dx}$$

- If we substitute

$$P_Y(y) = \int_{-\infty}^{\infty} P(x, y) dx$$

into this equation for the conditional probability we have

$$P(x \mid y) = \frac{P(x, y)}{P_Y(y)}$$

- Further than this, we also see

$$P(x, y) = P(x \mid y) P_Y(y)$$

- Generally when two random variables are dependent $\langle XY \rangle \neq \langle X \rangle \langle Y \rangle$

Outline

Correlations and correlation coefficients

- The expectation value of the product of two random variables is called the correlation of the two variables.
- Item the correlation is a measure of how correlated two variables are.
- For a measure of how mutually dependent two variables are we divide the correlation by the square root of the product of the variances

$$C_{XY} \equiv \frac{\langle XY \rangle}{\sqrt{V(X)V(Y)}}$$

where C_{XY} is called the correlation coefficient of X and Y.
- If the means of X and Y are not zero, we can remove these when calculating the correlation coefficient and preserve its properties, we can find in general the correlation coefficient as

$$C_{XY} \equiv \frac{\langle (X - \langle X \rangle)(Y - \langle Y \rangle) \rangle}{\sqrt{V(X)V(Y)}} = \frac{\langle XY \rangle - \langle X \rangle \langle Y \rangle}{\sqrt{V(X)V(Y)}}$$

Correlations and correlation coefficients

- The quantity $\langle XY \rangle - \langle X \rangle \langle Y \rangle$ is called the covariance of X and Y and is zero if X and Y are independent.
- The correlation coefficient is zero if X and Y are independent.
- The correlation coefficient is unity if $X = cY$ (c being some positive constant).
- If $X = -cY$, then the correlation coefficient is -1, and we say that the two variables are perfectly anti-correlated.
- The correlation coefficient provides a rough measure of the mutual dependence of two random variables, and is one that is relatively easy to calculate.

Outline

University of Bedfordshire

Adding random variables together

- The proability density for $Z = X + Y$ is given by
$$P_Z(z) = \int_{-\infty}^{\infty} P_X(s-z)P_Y(s)ds \equiv P_X * P_Y$$
which is called the convolution of P_X and P_Y, and is denoted by another function "$*$".
- The mean and the variance are defined as follows, for $X = X_1 + X_2$
$$\langle X \rangle = \langle X_1 \rangle + \langle X_2 \rangle$$
$$V_X = V_1 + V_2$$
where the two events are independent.
- The notion that averaging the results of a number of independent measurements producing a more accurate results is an important one here. If we sum the avereges of a number of experiments, N, the mean will not change, however, because we are dividing each of the variable by N, the variance goes down by $1/N^2$.
- Because it is the variances that add together, the variance of the sum is V/N. Thus the variance gets smaller as we add more results together.
- The uncertainty of the results is the standard deviation, and the standard deviation of the average is $\sigma/\sqrt{N}$

University of Bedfordshire

Outline

University of Bedfordshire

Transformation of a random variable

- If we know the probability density for a random variable X, then it can be useful to know how to calculate the probability density for some random variable Y, that is a function of X. This is referred to as a transformation of a random variable.
- Consider the case where $Y = aX + b$ for constants a and b.
  1. The probability density will be stretched by a factor a.
  2. The probability density will be shifted a distance of b.

University of Bedfordshire

Transformation of a random variable

- More generally, if $Y = g(X)$, then we determine the probability density for Y by changing the variables as shown below.
- We begin by writing the expectation value of a function Y, $f(Y)$, in terms of $P(x)$.

$$\langle f(Y) \rangle = \int_{x=a}^{x=b} P(x) f(g(x)) dx$$

where a and b are the upper and lower limits on the values X can take.

- Now we transform this into an integral over the values of Y

$$\langle f(Y) \rangle = \int_{y=g(a)}^{y=g(b)} P(g^{-1}(y)) \left(\frac{dx}{dy} \right) f(y) dy$$
$$= \int_{y=g(a)}^{y=g(b)} \frac{P(g^{-1}(y))}{g'(g^{-1}(y))} f(y) dy$$

Transformation of a random variable

- We now identify the function that multiplies $f(y)$ inside the integral over y as the probability density.
- The probability density for y is therefore

$$Q(y) = \frac{P(g^{-1}(y))}{|g'(g^{-1}(y))|}$$

- One must realise that this expression for $Q(y)$ only works for functions that map a single value of x to a single value of y (invertable functions), because in the change of variables we assumed that g was invertable.

Outline

The distribution function

- The probability distribution function, which we call $D(x)$, of a random variable X is defined as the probability that X is less than or equal to x

$$D(x) = Prob(X \leq x) = \int_{-\infty}^{x} P(z)dz$$

- In addition, the fundamental theorem of calculus tells us that

$$P(x) = \frac{dD(x)}{dx}$$

Outline

The characteristic function

- Another useful definition is that of the characteristic function, $\chi(s)$.
- The function is defined as the fourier transform of the probability density.
- The Fourier transform of a function $P(x)$ is another function given by

$$\chi(s) = \int_{-\infty}^{\infty} P(x) e^{isx} dx$$

- One use of the Fourier tansform is that it has a simple inverse, allowing one to perform a transformation on $\chi(s)$ to get back $P(x)$. This inverse transform is

$$P(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \chi(s) e^{-isx} ds$$

- If we have two functions $F(x)$ and $G(x)$, then the fourier tansform of their convolution is simple the product of their respective fourier transforms.
- We now have an alternative way to find the probability density of the sum of two random variables:
  1. Convolve their two densities.
  2. Calculate the characteristic functions for each, multiply these together, and then take the inverse Fourier transform.

Outline

THE MULTIVARIATE GAUSSIAN

- It is possible to have a probability density for N variables, in which the marginal densities for each of the variables are all Gaussian, and where all the variables may be correlated.
- Defining a column vector of N random variables, $\mathbf{x} = (x_1, x_2, \ldots, x_N)^T)$, the general form of the multivariate Gaussian is

$$P(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^N det[\Gamma]}} exp[-\frac{1}{2}(\mathbf{x}-\mu)^T \Gamma^{-1}(\mathbf{x}-\mu)]$$

  where **mu** is the vector of the means of the random variables, and Γ is the matrix of covariances of the variables,

$$\Gamma = \langle \mathbf{X}\mathbf{X}^T \rangle - \langle \mathbf{X} \rangle \langle \mathbf{X} \rangle^T = \langle \mathbf{X}\mathbf{X}^T \rangle - \mu\mu^T$$

- Note that the diagonal elements of Γ are the variances of the individual variables.