Key Ideas : SOR3012

The baloons below contain many of the important ideas and theorems that are covered in this module. If you have a good understanding of what everything on this sheet means then you have a good understanding of the module content. I would recommend that you stick these sheets in the first few pages of the hardback book that you keep your notes inside and that you consult these notes regularly as you work through the module.

How do we use capital letters in statistics?

In statistics we use capital letters to denote the random outcomes from experiments performed in the future.

How do we encode the information we have about what might happen in those experiments in the future mathematically?

We encode our understanding of what might happen in those experiments in the future in a function known as the cumulative probability distribution function. The value of this function, $F_X(x)$, at small x tells you the probability that the random variable will take a value less than or equal to x.

Give three properties that all cumulative probability distribution functions must have?

The cumulative probability distribution must have the following three properties: $\lim_{x\rightarrow -\infty} P(X \le x) = 0$, $\lim_{x\rightarrow +\infty} P(X \le x) = 1$ and $\lim_{\epsilon\rightarrow 0} P(X \le (x+\epsilon)) = P(X \le x)$

What is the difference between discrete and continuous random variables?

Discrete random variables cannot take any real value on the real axis. They can only take particular (usually integer) values. Continuous random variables can take any real value in a particular range.

What does the probability mass function $f_X(x)$ measure?

The probability mass function $f_X(x)$ tells one the probability that the **discrete** random variable (capital) X will take a value of (small) x.

What does the probability density function $f_X(x)$ measure?

The probability density function $f_X(x)$ is equal to the derivative of the cumulative probability distribution function for the **continuous** random variable (capital) X evaluated at the point (small) x.

How do you calculate the expectation of a discrete random variable?

The expectation of a discrete random variable is equal to the sum over all the possible values that the random variable can take of x_i multiplied by the probabiltiy mass function $f_X(x_i)$. In other words, $\mathbb{E}[X] = \sum_{i=0}^{\infty} x_i f_X(x_i)$

How do you calculate the expectation of a continuous random variable?

The expectation of a continuous random variable is equal to the integral over all possible x values of x multiplied by the probability density function, $f_X(x)$. In other words, $\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) \mathrm{d}x$.

How do you calculate the variance of a random variable?

The variance of a random variable, X, can be calculated by taking the expectation of $[X - \mathbb{E}(X)]^2$ or by computing the expectation of the square of the random variable, $\mathbb{E}(X^2)$, and by subtracting $\mathbb{E}(X)^2$.

How is the moment generating function calculated and explain how one can calculate moments if one is given this function

The moment generating function, $M_X(t)$ for a random variable, X, is $M_X(t) = \mathbb{E}(e^{tX})$. If one evaluates the nth derivatives of this function at $t = 0$ one gets the nth moment of the distribution

Why is the expectation an important quantity?

The expectation is important because the sum of n independent and identically distributed random variables divided by n converges towards this particular value because of a result known as the law of large numbers. The law of large number states:

$$\lim_{n\to\infty} P\left(\left| \frac{S_n}{n} - \mathbb{E}(X) \right| > \epsilon \right) = 0$$

where n is the number of independent random variables with expectation $\mathbf{E}(X)$ that have been added together to give S_n and where ϵ is a small number.

What does the central limit theorem state?

The central limit theorem states that the cumulative probability distribution function for a sum of independent and identically distributed random variables of most types can be approximated using the cumulative probability distribution function of a normal distribution. More precisely it states:

$$\lim_{n\to\infty} P\left(\frac{S_n/n - \mu}{\sigma/\sqrt{n}} \le z \right) = \Phi(z)$$

where n is the number of independent random variables with expectation μ and variance σ^2 that have been added together to give S_n and where $\Phi(z)$ is the cumulative probability distribution function for the standard normal distribution with expectation 0 and variance 1.

What can we use to understand if the result from one experiment affects the outcome of a second, different experiment?

To understand if the result from one experiment, (capital) X, affects the outcome of a second, different experiment, (capital) Y we use the conditional probability. The conditional probability that $X = 3$ given $Y = 2$ is equal to the probability that $X = 3$ AND $Y = 2$ divided by the probability that $Y = 2$. In other words:

$$P(X = 3|Y = 2) = \frac{P(X = 3 \wedge Y = 2)}{P(Y = 2)}$$

What does Bayes theorem state?

Bayes theorem states that $P(X = x|Y = y)P(Y = y) = P(Y = y|X = x)P(X = x)$

If $X = 2$ whenever $Y = 4$ what can we say about the events $X = 2$ and $Y = 4$?

If $X = 2$ whenever $Y = 4$ then the events $X = 2$ and $Y = 4$ are concurrent. These two events always happen at the same time and the conditional probablity $P(X = 2|Y = 4)$ is equal to one.

If X is never equal to 2 whenever $Y = 4$ what can we say about the events $X = 2$ and $Y = 4$?

If X is never equal to 2 whenever $Y = 4$ then the events $X = 2$ and $Y = 4$ are mutually exclusive. Y's equalling 4 somehow prevents X from equalling two and the conditional probability $P(X = 2|Y = 4)$ is equal to zero.

If the value the random variable X takes has no effect on the value on the value the random variable Y takes what can we say about the random variables X and Y

If the value the random variable X takes has no effect on the value on the value the random variable Y the two random variables are said to be indepdent. For all possible values of x and y the conditional probability $P(X = x|Y = y) = P(X = x)$ and the conditional probability $P(Y = y|X = x) = P(Y = y)$.

Can two events be both independent and mutually exclusive?

Two events $X = 3$ and $Y = 2$ cannot be independent and mutually exclusive as indepdences implies that the conditional probability $P(X = 3|Y = 2) = P(X = 3)$, while mutual exclusivity implies that the conditional probability $P(X = 3|Y = 2) = 0$. We can conclude from these two equations that $P(X = 3) = 0$ and hence that the event $X = 3$ is impossible.

What is a Bernoulli random variable?

A bernoulli random variable, X, is a discrete random variable that is used to model an experiment with two outcomes success and failure. For this variable failure is given a value of 0 and success a value of 1. The probability of success $(X = 1)$ is p. The expectation of this random variable is p and the variance is p times $(1 - p)$

What is a Binomial random variable?

A Binomial random variable is a discrete random variable that is used to model the number of successes amongst n independent Bernoulli trials. The probability mass function for this random variable is equal to $\binom{n}{p} p^x (1-p)^{n-x}$ The expectation of this random variable is np, while the variance is $np(1-p)$

What is a Geometric random variable?

A geometric random vairable is a discrete random variable that used to model the number of independent Bernoulli trials that need to be performed before you get a success. The probability mass function for this random variable is equal to $(1 - p)^{x-1} p$ The expectation of this random variable is $1/p$ while the variance is $(1 - p)/p^2$.

What is an exponential random variable?

An exponential random variable is a continuous random variable that can be used to model the process of waiting for something to happen. This random variable is unique in that it has no memory. The cumulative probability distribution function for this random variable is equal to $1 - e^{-\lambda t}$ The expectation of this random variable is $1/\lambda$ and the variance is $1/\lambda^2$.

What is a poisson random variable?

A Poisson random variable is a discrete random variable, which can be thought of as a large n limit for the binomial random variable. The probability mass function for this random variable is equal to $\frac{\lambda^x}{x!} e^{-\lambda}$. The mean and variance of this random variable are both equal to λ.

What does the joint probability mass function, $f_{XY}(x, y)$, measure?

The joint probability mass function $f_{XY}(x, y)$ measures the probability that the random variable X is equal to x and the random variable Y equals y.

How do we calculate the covariance of a pair of random variables?

The covariance of a pair of random variables X and Y, is calculated as $\mathbb{E}\{[X - \mathbb{E}(X)][Y - \mathbb{E}(Y)]\}$ or as $\mathbb{E}(XY) - \mathbb{E}(Y)\mathbb{E}(Y)$.

What is a stochastic process?

A stochastic process is a time series of random variables.

What is the simplest kind of stochastic process?

The simplest kind of stochastic process is a Markov chain. A time series of random variables is said to have to have the Markov property if the values the random variable take at future times depends only on the current value the random variable. In other words, the values the random takes in the future does not depend on the values random variables took during the past. This is a rather colloquial definition a more formal definition for a Markov chain is a time series of random variables whose probability distribution functions have the following property:

$$P(X_{t+1} = x_{t+1}|X_0 = x_0 \wedge X_1 = x_1 \wedge \cdots \wedge X_t = x_t) = P(X_{t+1} = x_{t+1}|X_t = x_t)$$

What mathematical object do we use to represent the one step transition probabilities for a Markov chain?

We use a matrix to represent the one-step transition probabilities. Element (i, j) of this matrix gives the probability that the system will transition from state i to state j in a single timestep.

How do we calculate the probability that the system will transition from state i to state j over the course of n timesteps?

Element (i, j) of the nth power of the one-step transition probability matrix is equal to the probability that the system will transition from state i to state j over the course of n timesteps. This result is known as the Chapman-Kolmogorov relation.

Draft

What is the difference between a recurrent and a transient state of a Markov chain?

Any recurrent state is guaranteed to have a finite return time. A transient state is not guaranteed to have a finite return time. More formally if a state is recurrent $\sum_{n=1}^{\infty}(\mathbf{P}^n)_{ii} = 1$ while if a state is transient $\sum_{n=1}^{\infty}(\mathbf{P}^n)_{ii} < 1$.

How do we measure the period of a state in a Markov chain?

The period of a state is equal to the greatest common divisor of the set of possible return times to that state.

When does a Markov chain have a limiting stationary distribution?

A Markov chain has a limiting stationary distribution when all the states in the chain are recurrent. A markov chain with a limiting stationary distribution is said to be ergodic. This stationary distribution can be found by finding the top left eigenvector of the transition probability matrix. Furthermore, a Markov chain which has only recurrent states satisfies the ergodic theorem which tells us 1 over the expected return time to a state is equal to the fraction of time the system stays in that state.

What does the ergodic theorem state and when does it hold?

The ergodic theorem states that

$$\lim_{n\to\infty} \frac{M_k(n)}{n} = \frac{1}{\mathbb{E}(T_k)}$$

where $M_k(n)$ is the number of visits the system makes to the kth state in a n step chain and where $\mathbb{E}(T_k)$ is the expected return time to state k. This theorem holds for Markov chains that have a finite number of recurrent states.

What is the Kolmogorov equation?

The Kolmogorov equation is a differential equation that is at the heart of the theory of Markov chains in continuous time. The Kolmogorov equation tells us that derivative of the transition probability matrix with respect to time is equal to the product of the jump rate matrix, $\mathbf{Q}$, with the transition probability matrix, $\mathbf{P}(t)$. In other words, $\frac{\mathrm{d}P(t)}{\mathrm{d}t} = \mathbf{QP}(t)$.

How is the jump rate matrix of a continuous time Markov chain defined?

The jump rate matrix, $\mathbf{Q}$, of a continuous time Markov chain is equal to $\mathbf{P}(t)$ minus the identity over t in the limit as t tends to zero. In other words, $\mathbf{Q} = \lim_{t\to 0} \frac{\mathbf{P}(t)-\mathbf{I}}{t}$.