18.443 Problem Set 3 Spring 2015

Statistics for Applications

Due Date: 2/27/2015

prior to 3:00pm

Derive the mean and variance for a binomial distribution.

The mean () and variance (^2) of a binomial distribution can be derived from its parameters: the number of trials (n) and the probability of success on each trial (p). The mean of a binomial distribution is calculated as the product of the number of trials and the probability of success, = np. The variance is calculated as the product of the number of trials, the probability of success, and the probability of failure (1-p), ^2 = np(1-p).

How do these values change when the distribution is skewed?

In a binomial distribution, skewness is determined by the probability of success (p). If p > 0.5, the distribution is negatively skewed, and if p < 0.5, it is positively skewed. The mean shifts towards the tail of the distribution, but the variance remains the same as it depends on both p and (1-p). However, the distribution becomes less dispersed (i.e., the variance decreases) as the distribution becomes more skewed, because the outcomes become more predictable.

Explain the basics of a fractal model.

A fractal model is a mathematical construct that is self-similar, meaning it appears the same at any scale. Fractals are complex shapes made up of an infinite number of similar, smaller copies of the whole, each of which is a reduced-scale copy of the whole. They are often used to model structures in which similar patterns recur at progressively smaller scales, and in describing partly random or chaotic phenomena such as crystal growth, fluid turbulence, and galaxy formation.

How could you apply this model to understand the structure of Chromatin?

The fractal model can be applied to understand the structure of Chromatin by considering the

self-similar nature of fractals. Chromatin, a complex of DNA and protein, has a hierarchical structure that can be considered self-similar at different scales. The basic unit of chromatin is the nucleosome, which repeats along the DNA molecule. Higher levels of organization involve the packing of these nucleosomes into more compact structures, and this packing can also be considered to repeat in a self-similar manner. Therefore, a fractal model could potentially capture this complex, multi-scale organization.

Define kurtosis in statistical terms.

Kurtosis is a statistical measure that describes the shape of a probability distribution. Specifically, it measures the 'tailedness' and sharpness of the peak of a distribution. A high kurtosis indicates a distribution with heavy tails and a sharp peak, while a low kurtosis indicates a distribution with light tails and a flat peak. The standard normal distribution has a kurtosis of 3.

How does the kurtosis of a Laplace distribution differ from that of a normal distribution?

The kurtosis of a Laplace distribution is higher than that of a normal distribution. This means that a Laplace distribution has heavier tails and a sharper peak than a normal distribution. Specifically, the kurtosis of a Laplace distribution is 6, while the kurtosis of a normal distribution is 3. This indicates that extreme values are more likely in a Laplace distribution than in a normal distribution.

State the Hardy-Weinberg principle.

The Hardy-Weinberg principle, also known as Hardy-Weinberg equilibrium, is a principle in population genetics. It states that both allele and genotype frequencies in a population remain constant from generation to generation unless specific disturbing influences are introduced. These influences can include mutation, gene flow, genetic drift, non-random mating, and selection. The principle is based on a set of assumptions about an idealized population, including infinitely large population size, random mating, no mutation, no migration, and no natural selection.

Describe a situation where the principle would not hold for the Haptoglobin gene frequencies.

The Hardy-Weinberg principle would not hold for the Haptoglobin gene frequencies if any of the assumptions of the principle are violated. For instance, if there is non-random mating in the population, the frequencies of the Haptoglobin gene may change over generations. Similarly, if there are mutations in the Haptoglobin gene, or if individuals with certain Haptoglobin gene types have a survival or reproductive advantage (natural selection), the gene frequencies may not remain constant. Migration of individuals into or out of the population can also disrupt the equilibrium.