**Cousera Data Science Course 6 Week 1: Statistical Inference**

**1. Probability**

Given a random experiment, a probability measure a population quantity that summarizes the randomness.   
Specifically, probability takes a possible outcome from the experiment and:  
so that the probability of the union of any two sets of outcomes that have nothing in common must be the sum of their probabilities

**Rules**Probability that nothing occurs is 0  
probability that something occurs is 1  
Probability of something is 1 minus the probability that the opposite occur  
Probability of at least one or two or more things that cannot simultaneously occur is the sum of their respective probabilities  
If an event A implies event B, then the probability of A occurring is less than the probability that B occurs  
For any two events the probability of at least one occurs is the sum of their probabilities minus their intersection

Ex: 3% sleep problem A, 10% sleep problem B, can’t just add them up

**Probability mass functions  
Random Variables**- numerical outcome of an experiment  
-can be discrete or continuous (values vs ranges)  
Ex of discrete random variables: flip of a code, outcome from the roll of a die, the website traffic on a given day  
Ex of continuous random variable: BMI of a subject after a time, intelligence quotients of a sample of children

**PMF**Evaluated at a value corresponds to the probability that a random variable takes that value. To be a valid pmf a function, p, must satisfy  
1. It must always be larger than or equal to 0  
2. The sum of the possible values that the random variable can take has to add up to one

Ex: Bernoulli distribution  
X=0 -> tails, X = 1 -> heads  
p(x) = (1/2)x(1/2)1-x for x = 0, 1

**Probability Density Function**-associated with continuous random variable  
1. it must be larger than or equal to zero everywhere  
2. Total area must be equal to 1

F(x) = 2x for 0< x<1  
 = 0 else

What’s the probability that 75% of the calls get addressed?  
-> it’s just another right triangle

**CDF and survival function**The **Cumulative distribution function (CDF)** of a random variable, X, returns the probability that the random variable is less than or equal to the value x  
F(x) = P(X <= x)  
Survival function is P(X >= x)

**Quantile**If you were the 95th percentile on an exam, you know that 95% of people scored worse than you and 5% scored better.   
The alpha-th quantile of a distribution with distribution function F is the point xalpha so that F(xalpha) = alpha -> move x around until F(x) = alpha  
-a percentile

On 50% of the days: 0.5 = x^2 , x = 0.7071

Qbeta(0.5, 2, 1) gives us relevant quantiles

**Conditional Probability**Let B be an event so that P(B) > 0  
Statistical “Independence”

Bayes’ Rule  
-allows us to reverse the conditioning set provided that we know some marginal probabilities

Diagnostic Tests  
Let + and – be the events that the result of a diagnostic test is positive or negativey respectively  
Let D and DC be the event that the subject of the test has or does not have the disease respectively  
The sensitivity is the probability that the test is positive given that the subject actually has the disease  
The specificity is the probability that the test is negative given that the subject does NOT have the disease

The positive predictive value is the probability that the usbjecct has the disease given that the test is positive  
The negative predictive value is the probability that the subject does not have the disease given that the test is negative  
The prevalence of the disease is the marginal probability of disease

The diagnostic likelihood ratio of a positive test is sensitivity / (1- specificity)  
The diagnostic likelihood ratio of a negative test, is (1- sensitivity) / specificity

**IID random variables**-random variables are said to be iid if they are independent and identically distributed  
independent – statistically unrelated from one and another  
identically distributed – all having been drawn from the same population distribution

**Expected values**  
-useful for characterizing a distribution  
-mean is characterization of its center  
-the variance and standard deviation are characterizations of how spread out it is  
-our sample expected values will estimate the versions

**The Population Mean**-The expected value or mean of a random variable is the center of its distribution  
-for discrete random variable X with PMF p(x) it is defined as E[X] = sigmaxp(x)  
E[X] is the center of mass of a collection of locations and weights

**The Sample Mean**  
-the sample mean estimates this population mean  
-the center of mass of the data is the empirical mean  
- X = sigma xi p(xi) where p(xi) = 1/n

**The Variance**  
-The variance of a random variable is measure of spread  
-If X is a random variable with mean u, the variance of X is defined as:  
Var(x) = E[(x-u)^2] = E[X^2] – E[X]^2  
-The expected distance from the mean  
-Densities with a higher variance are more spread out than densities with a lower variance  
-The square root of the variance is called the standard deviation  
-The standard deviation has the same units as X

**The Sample Variance**-The sample variance is   
S^2 = sigma(Xi – X)^2 / n-1 (almost, but not quite, the average squared deviation from the sample mean)  
-It is also a random variable  
 -It has an associate population distribution  
 -Its expected value is the population variance  
 -Its distribution gets more concentrated around the population variance with more data  
-Its square root is the sample standard deviation

**Distributions**

**Bernoulli distribution**  
-result of a binary outcome  
-Bernoulli random variables only take the values 1 and 0 with probabilities of p and 1-p  
-The PMF of a Bernoulli random variable X is   
P(X = x) = p^x (1-p) ^ (1-x)  
-The mean of a Bernoulli random variable is p and the variance is p(1-p)  
-If we let X be a Bernoulli random variable, it is typical to call X = 1 as a “success” and X = 0 as a “failure”

**Binomial trials**-The binomial random variables are obtained as the sum of iid Bernoulli trials  
-in specific, let X1 …. Xn be iid Bernoulli(p); then X = sigma(Xi) is a binomial random variable  
-The binomial mass function is P(X = x) = (n x) p^x (1-p) ^ (n-x)  
for x = 0…. N

**Choose**: Recall (“n choose x”) - > n choose 0 = n choose n = 1

**The Normal Distribution**A random variable is said to follow a normal or Gaussian distribution with mean u and variance alpha^2 if the associated density is (check slides)  
Note that…  
1. Approximately 68%, 95% and 99% of the normal density lies within 1, 2, and 3 standard deviation from the mean, respectively  
2. -1.28, -1.645, -1.96, -2.33 are the 10th, 5th, 2.5th, and 1st percentiles of the standard normal distribution respectively  
3. By symmetry, the other side holds true as well

**The Poisson Distribution**-modeling count data  
-modeling event-time or survival data  
-modeling contingency tables  
-approximating binomials when n is large and p is small

**Asymptotics**  
-The term for the behavior of statistics as the sample size limits to infinity  
-Asymptopia is my name for the land of asymptotics, where everything works out well and there are no messes. The land of infinite data is nice that way.  
-Asymptotics are incredibly useful for simple statistical inference and approximations  
-Asymptotics generally give no assurances about finite sample performance  
-Asymptotics form the basis of frequency interpretation of probabilities

**Limits of random variables**-fortunately for the sample mean there’s a set of powerful results  
-these results allow us to talk about the large sample distribution of sample means of a collection of iid observations  
-the first of these results we intuitively know  
 -it says that the average limits to what its estimating the population mean  
 -its called the law of large numbers

**The Central Limit Theorem**-CLT states that the distribution of averages of iid variables (properly normalized) becomes that of a standard normal as the sample size increases  
-CLT applies in aan endless variety of settings  
- (Estimate – Mean of estimate) / Std. Err of estimate  
-The useful way to think about the CLT is that Xn is approximately N(u, alpha^2/n)