Math 10B with Professor Stankova
Worksheet, Midterm #2; Wednesday, 3/21/2018
GSI name: Roy Zhao

1 Concepts

We use **Bayes theorem** when we want to find the probability of A given B but we are told the opposite probability, the probability of B given A. There are several forms of Bayes Theorem as follows:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\bar{A})P(\bar{A})} = \frac{1}{1 + \frac{P(B|\bar{A})P(\bar{A})}{P(B|A)P(A)}}.$$

In order to discern which form to use, look at the information you are given. If you are told $P(B|A)$ as well as $P(B|\bar{A})$, use the latter two methods but if you are only told $P(B)$, then use the first form.

We say that two events A, B are **independent** if $P(A \cap B) = P(A)P(B)$.

A **random variable** is any function $X : \Omega \to \mathbb{R}$. It isolates some concept that we care about. For example, when we flip a coin 20 times, then we can define a random variable which is the number of heads that we flip.

A **probability mass function (PMF)** is a function from $\mathbb{R}$ to $[0, 1]$ that is associated to a random variable X. We define $f(x) = P(X = x) = P(X^{-1}(\{x\}))$.

Two random variables X, Y are called **independent** if for any subsets $E, F \subset \mathbb{R}$, the subsets $X^{-1}(E), Y^{-1}(F) \subset \Omega$ are independent. To prove that two random variables are independent, we need to show that those two sets are independent for any two choices of E, F (actually, it suffices to only consider E, F as one point sets or that $P(X = x, Y = y) = P(X = x)P(Y = y)$ for any $x, y \in \mathbb{R}$). To prove that they are not independent, we only need to find one counterexample pair E, F.

Distribution	PMF	$E(X)$	Variance
Uniform	If $\#R(X) = n$, then $f(x) = \frac{1}{n}$ for all $x \in R(X)$.	$\sum_{i=1}^{n} \frac{x_i}{n}$	$\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{n}$
Bernoulli Trial	$f(0) = 1 - p$, $f(1) = p$	p	$Var(X) = p(1-p)$
Binomial	$f(k) = \binom{n}{k} p^k (1-p)^{n-k}$	np	$np(1-p)$
Geometric	$f(k) = (1-p)^k p$	$\frac{1-p}{p}$	$Var(X) = \frac{1-p}{p^2}$
Hyper-Geometric	$f(k) = \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}$	$\frac{nm}{N}$	$\frac{nm(N-m)(N-n)}{N^2(N-1)}$
Poisson	$f(k) = \frac{\lambda^k e^{-\lambda}}{k!}$	λ	λ

The **Expected Value** is the weighted average of all the values the random variables can take on. By definition, it satisfies some properties:

- $E[c] = c$
- $E[cX] = cE[X]$

- $E[X+Y] = E[X] + E[Y]$ for **all** random variables
- $E[XY] = E[X]E[Y]$ for **independent** random variables.

The **Covariance** is defined as $Cov(X,Y) = E[XY] - E[X]E[Y]$. It measures how "independent" two random variables are. For **independent** random variables, we have $Cov(X,Y) = 0$. Note that we can recover the definition of regular variance because the covariance of a random variable with itself is $Cov(X,X) = E[X^2] - E[X]^2 = Var(X)$. We can update the formula for the variance of the sum of two random variables as $Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$ which holds for **all** random variables. Properties that hold for the random variable are:

- $Cov(X,Y) = Cov(Y,X)$
- $Cov(X,Y+Z) = Cov(X,Y) + Cov(X,Z)$
- $Cov(X,cY) = cCov(X,Y)$ for any constant c
- $Cov(X,c) = 0$ for any constant c

The **Variance** is defined as $Var(X) = E((X-\mu)^2)$. An easier form is $E(X^2) - E(X)^2$. It satisfies some properties:

- $Var(c) = 0$
- $Var(cX) = c^2 Var(X)$
- $Var(X+Y) = Var(X) + V(Y)$ for **independent** random variables.

In order to compute the probability $P(a \leq X \leq b)$ for a normal distribution, we need to take an integral $\int_a^b \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/\sigma^2}$ and this integral is almost impossible to do without a calculator. So, what we do is have a table of values for this integral and look up the value that we need. Given a z score such as 1.5, when we look it up in the table, $z(1.5) = P(0 \leq Z \leq 1.5)$, where Z is the standard normal distribution; the bell curve with mean $\mu = 0$ and standard deviation $\sigma = 1$.

One key area these pop up in is when taking the average of a bunch of trials. The **Central Limit Theorem (CLT)** tells us that for X_i independent and identically distributed (i.i.d.) (e.g. rolling a die multiple times) with $E[X_i] = \mu$ and $Var(X_i) = \sigma^2$, then the average that we get (e.g. the average number that we roll) is **approximately** normal distributed with mean μ and standard deviation $\sigma/\sqrt{n}$. So

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

is approximately normally distributed with $E[\bar{X}] = \mu$ and $Var(\bar{X}) = \sigma^2/n$.

In order to compute probabilities, we compute the z score. Given a normal distribution with mean μ and standard deviation σ, the z score of a value a is $\frac{|a-\mu|}{\sigma}$. Then we look up this value in a table.

The **Law of Large Numbers** is a weaker statement that just says that as we take averages and let $n \to \infty$, then the sample mean becomes closer and closer to the actual mean μ. Namely, $E[\bar{X}] \to \mu$ and the probability that we are far away from the mean goes to 0.

Often times, we are not given the distribution or parameters of the distribution (but we know what kind of distribution it is), and we want to figure out what the parameters are. One example is if you are given a biased coin and you want to figure out how biased it is (how likely flipping heads/tails is).

The **estimator for the mean** is the **sample mean** which is given as

$$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{k=1}^{n} x_k.$$

The **biased standard deviation estimator** is given by

$$x_* = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(x_k - \bar{x})^2}.$$

The **unbiased standard deviation** or **sample standard deviation** is given by

$$s = \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}(x_k - \bar{x})^2}.$$

Given estimators for the mean and standard deviation (or the sample mean and sample standard deviation) $\hat{\mu}, \hat{\sigma}$ respectively, the 95% **confidence interval** for the expected value μ is

$$(\hat{\mu} - 2\hat{\sigma}/\sqrt{n}, \hat{\mu} + 2\hat{\sigma}/\sqrt{n}).$$

You say that you are 95% confident that μ is in that interval.

In general, statistics does not allow you to prove anything is true, but instead allows you to show that things are probably false. So when we do hypothesis testing, the **null hypothesis** H_0 is something that we want to show is false and the **alternative hypothesis** H_1 is something that you want to show is true. For example, to show that a drug cures cancer, the null hypothesis would be that the drug does nothing and the alternative hypothesis would be that the drug does help cure cancer.

A **type 1 error** is rejecting a true null which means that in our example, saying a drug cures cancer when it doesn't. A **type 2 error** is failing to reject a false null which means in our case as saying that the drug doesn't do anything when it does. The **significance level** is the probability of making a type 1 error. The **power** is 1 minus the probability of making a type 2 error.

You use a χ^2 test to determine if a distribution is how you expect it to be. Suppose that you expect it to be distributed with a different values and for each of these values, you expect to get outcome k m_k times but actually get it n_k times. Then you compare the statistic

$$r = \sum_{k=1}^{a} \frac{(n_k - m_k)^2}{m_k}$$

with the $\chi^2(a-1)$ distribution.

To test for independence, it is just a modified version of the χ^2 test. You sum up the rows to get N_i and the columns to get M_j. Let the total sum of all the elements be S. Then, your expected distribution at square ij is $\frac{N_i M_j}{S}$, and then you perform the χ^2 test. If you have r rows and c columns, then the number of degrees of freedom is $(r-1)(c-1)$.

A **homogeneous** recursion does not include any extra constants (e.g. $a_n = a_{n-1} + a_{n-2}$) and a **nonhomogeneous** recursion contains one (e.g. $a_n = a_{n-1} + 4$). The **order** of a recursion equation is the "farthest" back the relation goes. For instance, the order of $a_n = a_{n-1} + a_{n-3}$ is 3 because we need the term 3 terms back (a_{n-3}).

The general solution of a first order equation $a_n = a_{n-1} + d$ is $a_n = a_0 + nd$.

In order to solve a linear homogeneous we can replace the equation with its characteristic polynomial. For instance, the characteristic polynomial of $a_n = 2a_{n-1} + a_{n-2}$ is $\lambda^2 = 2\lambda + 1$. Then if $\lambda_1, \dots, \lambda_k$ are roots of this polynomial, then the general form of the solution is $a_n = C_1\lambda_1^n + \cdots + C_k\lambda_k^n$.

The Δ operator takes in a series and spits out a new one. By definition, we have that $\Delta a_n = a_{n+1} - a_n$. This is done to change linear non-homogeneous equations into homogeneous ones.