STATISTICS: Problem Set 1-Solutions

September 21, 2007

1. Suppose the probability of an individual being born on any particular day of the year is given by $\frac{1}{365}$.
(a)What is the probability that 2 people meeting at random have the same birthday?
(b)Suppose now that a group has three individuals. What is the probability that a least two of these individuals will share a birthday? What if the group has four individuals?
(c) How large must a group be such that the probability of finding at least two people with the same birthday is close to 50% (for this you will need to obtain an expression for the probability of at least two people sharing a birthday for an arbitrary group of size n)

(a) To compute the probability P_a that 2 individuals were born on the same day, we can first say that P_a is the sum of the probability P_i of being born the same particular day i. Indeed, the events "being born on a given day i" and "being born a given day $j \neq i$" are exclusive. Moreover, there are 365 days in a year. Finally, assuming the independence between births, we know that the probability of both being born on a particular day i is the product of the probability of each being born on that given day, $i.e$ $P_i = \frac{1}{365^2}$. Therefore, we have:

$$P_a = 365P_i = \frac{1}{365}$$

(b)

n=3

Let us define the event A as "at least 2 share the same birthday". Thus, A^c is the event "no one shares a birthday with no one else". We pass through the complementary event because it is, in my opinion, less complicated: we just have to compute the number of cases n_{A^c} in which no one share a birthday. First, there are 365 possibilities for the first to be born. For the second, it remains only 364 *different* days of birth. For the third one, 363 are left not to be born the same day as one of the two others. Thus, $n_{A^c} = 365 * 364 * 363$. Reminding that the total number of possibilities is 365^3, we have that:

$$P(A) = 1 - P(A^c) = 1 - \frac{365 * 364 * 363}{365^3} \simeq 0.0082042$$

n=4

Let denote P_b the probability that 2 people have the same birthday. By analogy with the first part of the question, we can write that:

$$P_b = 1 - \frac{365 * 364 * 363 * 362}{365^4} \simeq 0.01636$$

(c) What we are asked to do here is to find the number n of people such that the probability that at least 2 people have the same birthday be close to 0.5. Formally, we have to find n such that:

$$P = 1 - \frac{365 * 364 * 363 * 362 * ... * (365 - n + 1)}{365^n} \simeq 0.5$$

Numerically, n=23 ($P = 0.4927$).

2. A coin is to be tossed as many times as necessary to turn up one head. Thus the elements c of the sample space are $H, TH, TTH, TTTH$ and so forth. Let the probability set function P assign to these elements the respective probabilities $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}$, and so forth. Show that $P(C) = 1$. Let $C_1 = \{c : c \ is \ H, TH, TTH, TTTH, or TTTTH\}$. Compute $P(C_1)$. Next, suppose that $C_2 = \{c : c \ is \ TTTTHor \ TTTTTH\}$. Compute $P(C_2), P(C_1 \cap C_2), P(C_1 \cup C_2)$.

The experiment is to toss a coin n times. Let us first remark that tosses are pairwise independent. Moreover, we are told that the coin is tossed as many times as necessary to turn up *one* HEAD (henceforth "H"): if we define X to be the number of trials we need to turn up one H, X is distributed given a *negative* binomial distribution. Defining p as the probability of turning up H while tossing the coin *once*, we have:

$$P(X = x) = p(1-p)^{x-1}$$

In our case, assuming implicitly that H and T (i.e "tail") have the same probability of occurrence being equal to $\frac{1}{2}$, we obtain:

$$P(X = x) = \left(\frac{1}{2}\right)^x$$

First, we are asked to show that $P(C) = 1$. By definition, we have:

$$P(C) = \sum_{x=1}^{\infty} P(X = x) = \sum_{x=1}^{\infty} \left(\frac{1}{2}\right)^x$$

This a sum of terms of a geometric sequence. Moreover, $|\frac{1}{2}|$ is lower than one. Thus, we have:

$$P(C) = \left(\frac{1}{2}\right) \frac{1}{1 - \frac{1}{2}} = 1$$

We are asked to compute different probabilities, among which that of C_1. One can notice that the elements of C_1 are pairwise disjoint, the probability of the outcome being in the set C_1 is equal to the sum of the probabilities of occurrence of each element of the C_1. Thus, we have:

$$P(C_1) = \sum_{x=1}^{5} P(X = x) = \sum_{x=1}^{5} \left(\tfrac{1}{2}\right)^x = \left(\tfrac{1}{2}\right) \sum_{i=0}^{4} \left(\tfrac{1}{2}\right)^i$$
$$P(C_1) = \left(\tfrac{1}{2}\right) \frac{1-\left(\frac{1}{2}\right)^5}{1-\frac{1}{2}} = \frac{31}{32}$$

In the same way, we get:

$$P(C_2) = \sum_{x=5}^{6} P(X = x) = \frac{3}{2^6}$$
$$P(C_1 \cap C_2) = P(X = 5) = \left(\tfrac{1}{2}\right)^5$$
$$P(C_1 \cup C_2) = P(C_1) + P(C_2) - P(C_1 \cap C_2) = \frac{63}{64}$$

3. Examine whether functions F_1 and F_2 are valid distribution functions. Explain your answers.

$$F_1(x) = \begin{cases} 0 & \text{if } x < 0 \\ 2x^2 - \frac{2}{3}x^3 - x & \text{if } 0 \leq x \leq 2 \\ 1 & \text{if } 2 < x \end{cases}$$

$$F_2(x) = \begin{cases} 0 & \text{if } x < 0 \\ 2x^2 - \frac{2}{3}x^3 & \text{if } 0 \leq x \leq 2 \\ 1 & \text{if } 2 < x \end{cases}$$

The following theorem can be used to solve the exercise:

Theorem *Any function F with domain $\mathbb{R}$ and counterdomain [0,1] is a c.d.f. iff*

1. $lim_{x\to-\infty}F(x) = 0$, $\quad lim_{x\to+\infty}F(x) = 1$

2. $F(x)$ is non-decreasing function of x

3. $F(x)$ is right-continuous: $\forall x \; lim_{h\to 0}F(x + h) = F(x)$.

We need to check on whether $F_1(x)$, $F_2(x)$ satisfy conditions of the theorem.

Obviously, condition 1 holds for both functions, $F_1(x)$ and $F_2(x)$.

To check condition 2, differentiate both functions with respect to x for $x \in (0,2)$.

$$F_1'(x) = 4x - 2x^2 - 1$$
$$F_2'(x) = 4x - 2x^2$$

It is straightforward to check that $F_1'(x) \geq 0$ on $[1 - \frac{1}{2}\sqrt{2}, 1 + \frac{1}{2}\sqrt{2}]$ and $F_1'(x) < 0$ on $(0, 1 - \frac{1}{2}\sqrt{2}) \cup (1 + \frac{1}{2}\sqrt{2}, 2)$. Hence, F_1 is non-decreasing on $[1 - \frac{1}{2}\sqrt{2}, 1 + \frac{1}{2}\sqrt{2}]$ and is decreasing on $(0, 1 - \frac{1}{2}\sqrt{2}) \cup (1 + \frac{1}{2}\sqrt{2}, 2)$. So, condition 2 of the theorem does not hold and F_1 *is not a valid distribution function.*

On the other hand, F_2' is positive everywhere on $x \in (0,2)$, therefore, it is increasing on $(0,2)$. However, $F_2(2) = \frac{8}{3} > 1$ so that F_2 is *not* non-decreasing everywhere on its domain. Hence, condition 2 of the theorem does not hold. Therefore, F_2 *is not a valid distribution function.*

Additionally, note that neither F_1 nor F_2 satisfy the third condition of the theorem:

$$lim_{h \to 0} F_1(2+h) = 1 \neq \tfrac{2}{3} = F_1(2)$$
$$lim_{h \to 0} F_2(2+h) = 1 \neq \tfrac{8}{3} = F_2(2)$$

4. Suppose a certain accident annually kills 0.005% of the population. An insurance company provides insurance for 25,000 individuals.

(a) What is the probability that in a given year more than 5 of the insured individuals die due to the accident?

(b) Can you work out an estimate of the population mean death rate? How good an estimate is this likely to be?

(c) Can you provide a confidence interval? [Justify briefly all your modelling assumptions in answering this question.]

(a) Let X be the number of insured people who die due to the accident in a given year.

First, note that the probability of being killed by the accident is the same for the insured and uninsured individual and is equal to $p = 0.00005$. Assuming the independence between accidents, we know that the probability that k randomly chosen people die due to the accident is the product of the probabilities that each person dies due to the accident. Therefore, k randomly chosen people die with probability $(0.00005)^k$. We also know that the total number of unordered possibilities to "choose" k insured people out of 25,000 is $C^k_{25,000}$.

Then the probability that k insured people die due to the accident is

$$P(X = k) = C^k_{25,000}(0.00005)^k(1 - 0.00005)^{25,000-k}.$$

One can easily recognize the density function of the binomial distribution. So, the number of the insured individuals dying in a given year due to the accident is distributed binomially.

Now, the probability that in a given year 5 of the insured individuals die due to the accident is

$$P(X > 5) = \sum_{k=6}^{25,000} P(X = k) = \sum_{k=6}^{25,000} C^k_{25,000}(0.00005)^k(1 - 0.00005)^{25,000-k}.$$

Note that the number 25,000 of individuals is high enough and the probability value 0.00005 is small enough to claim that the binomial distribution can be approximated by the Poisson distribution $P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$ with parameter $\lambda = 0.00005 * 25,000 = 1.25$ (see Problem 7, part (c)). Then the probability that in a given year 5 of the insured individuals die due to the accident can be computed as

$$P(X > 5) = \sum_{k=6}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} = 1 - \sum_{k=0}^{5} \frac{\lambda^k e^{-\lambda}}{k!} = 1 - e^{-\lambda}\left[1 + \lambda + \frac{\lambda^2}{2} + \frac{\lambda^3}{3!} + \frac{\lambda^4}{4!} + \frac{\lambda^5}{5!}\right] \approx 0.002.$$

(b) We need to find an estimate of the population mean death rate.

Version 1

Define the population mean death rate as $\frac{\text{number of people killed by the accident per year}}{\text{population size}}$. Then the population mean death rate is simply the probability p of being killed by the accident for the Bernoulli process Y, where Y takes on one of the two values: 1 if the person is killed or 0 if the person is not killed. As we know,

$$P(Y = y) = p^y(1 - p)^{1-y};\ y = 0, 1.$$

To find an estimate of the parameter p in the Bernoulli distribution, we use the standard maximum likelihood (ML) estimation. We draw a sample $y_1, .., y_n$ of n values of Y and compute the (multivariate) probability density function associated with this sample. We obtain the likelihood function

$$L(y_1, .., y_n|p) = P(Y_1 = y_1, .., Y_n = y_n|p) = p^{y_1}(1-p)^{1-y_1} \cdots p^{y_n}(1-p)^{1-y_n} = p^{\sum_{i=1}^n y_i}(1-p)^{n-\sum_{i=1}^n y_i},$$

where $y_i = 0$ or 1, $i = 1, .., n$.

Maximizing the likelihood function is equivalent to maximizing the log-likelihood function. The log-likelihood function in this case is

$$lnL(y_1, .., y_n|p) = ln(p)\sum_{i=1}^n y_i + ln(1 - p)(n - \sum_{i=1}^n y_i).$$

To find the value of p which maximizes the log-likelihood function, differentiate this function with respect to p and set the derivative to be equal to zero. The solution of the resulting equation is a ML estimate of p.

$$\frac{dlnL(y_1,..,y_n|p)}{dp} = \frac{\sum_{i=1}^{n} y_i}{p} - \frac{n - \sum_{i=1}^{n} y_i}{1-p} = 0$$

Rearranging gives

$$\sum_{i=1}^{n} y_i - p\sum_{i=1}^{n} y_i = np - p\sum_{i=1}^{n} y_i$$

So, ML estimate of p is $\hat{p} = \frac{\sum_{i=1}^{n} y_i}{n} = \bar{y}$.

This estimate is unbiased: since each observation has expectation p ($E(Y_i) = p$), so does the sample mean $\bar{y}$. Indeed,

$$E(\hat{p}) = E\left(\frac{\sum_{i=1}^{n} y_i}{n}\right) = \frac{\sum_{i=1}^{n} E(y_i)}{n} = \frac{np}{n} = p.$$

Version 2

Define the population mean death rate as E[number of people killed by the accident per year] (without dividing by the population size). Then the population mean death rate may be viewed as the parameter λ in the Poisson process X, where X is the number of people who die due to the accident in a given year. Recall that by the properties of the Poisson distribution, $\lambda = E(X)$ that is λ is the expected number of deaths due to the accident per year. Therefore, to estimate the population mean death rate we need to work out an estimate of the parameter λ.

As before, we use the maximum likelihood (ML) estimation. We draw a sample $k_1,..,k_T$ of T values of X (k_t is the number of people killed by the accident in the year t) and compute the (multivariate) probability density function associated with this sample. We obtain the likelihood function

$$L(k_1,..,k_T|\lambda) = \prod_{i=1}^{T} \frac{\lambda^{k_i} e^{-\lambda}}{k_i!} = \frac{\lambda^{\sum_{i=1}^{T} k_i} e^{-T\lambda}}{\prod_{i=1}^{T} k_i!}.$$

The corresponding log-likelihood function is

$$lnL(k_1,..,k_T|\lambda) = -T\lambda + \sum_{i=1}^{T} k_i ln\lambda - \sum_{i=1}^{T} ln(k_i!).$$

Taking the derivative of $lnL(k_1,..,k_T|\lambda)$ with respect to λ we obtain

$$\frac{dlnL(k_1,..,k_T|\lambda)}{d\lambda} = -T + \frac{\sum_{i=1}^{T} k_i}{\lambda}$$

So, the ML estimate of λ is $\hat{\lambda} = \frac{\sum_{i=1}^{T} k_i}{T} = \bar{k}$.

This estimate is unbiased:

$$E(\hat{\lambda}) = E\left(\frac{\sum_{i=1}^{T} k_i}{T}\right) = \frac{\sum_{i=1}^{T} E(k_i)}{T} = \frac{T\lambda}{T} = \lambda.$$

(c) We need to find a confidence interval for the mean death rate. Recall that the confidence interval for a parameter is the interval around the sample estimate of this parameter within which, if the parameter actually takes a value in this range, the observed data would not be considered particularly unusual. Alternatively, one may say that a confidence interval gives an estimated range of values which is likely to include an *unknown population parameter*, the estimated range being calculated from a given set of sample data.

The confidence interval for a parameter is calculated for a certain confidence level. The chosen level depends on how precise we want to be. We interpret an interval calculated at a 95% level as, *we are 95% confident that the interval contains the true parameter value.* We could also say that 95% of all confidence intervals formed in this manner (from different samples of the population) will include the true parameter.

Version 1

Consider the first definition of the mean death rate, where it is represented by the parameter p in the Bernoulli distribution. How do we find a confidence interval for p?

First of all, the sample estimate of the parameter p in the Bernoulli distribution is $\hat{p} = \frac{\sum_{i=1}^{n} y_i}{n} = \bar{y}$. Secondly, the endpoints of the interval have to be calculated from the sample, so they are statistics, functions of the sample $Y_1, .., Y_n$ and hence random variables themselves. We determine the endpoints as follows.

Since the size n of the sample considered in this problem is large, we can use the Central limit theorem (CLT) which says that the sample mean $\bar{y}$ is distributed normally. By standardizing we get a random variable

$$Z = \frac{\bar{y} - E(\bar{y})}{\sqrt{var(\bar{y})}} \sim N(0, 1)$$

$\bar{y} = \hat{p}$, therefore, we obtain

$$Z = \frac{\hat{p} - E(\hat{p})}{\sqrt{var(\hat{p})}} \sim N(0, 1)$$

As calculated in part (b), $E(\hat{p}) = p$. The variance of $\hat{p}$ is calculated as follows:

$$var(\hat{p}) = var\left(\frac{\sum_{i=1}^{n} y_i}{n}\right) = \frac{\sum_{i=1}^{n} var(y_i)}{n^2} = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}.$$

Substituting these values for $E(\hat{p})$ and $var(\hat{p})$ in the expression for Z, we receive

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}})} \sim N(0, 1)$$

It is now possible to find numbers $-z$ and z, such that Z lies in between with probability $1 - \alpha$, the confidence level. Taking $1 - \alpha = 0.95$, we obtain

$$P(-z \leq Z \leq z) = 1 - \alpha = 0.95$$

The number z is a $100*(\alpha/2)\% = 2.5\%$ critical value for the standard normal distribution, and it is equal to 1.96. Then we get:

$$0.95 = 1 - \alpha = P(-z \leq Z \leq z) = P\left(-1.96 \leq \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \leq 1.96\right) =$$
$$= P\left(\hat{p} - 1.96\sqrt{\frac{p(1-p)}{n}}) \leq p \leq \hat{p} + 1.96\sqrt{\frac{p(1-p)}{n}})\right)$$

In the expression above $\sqrt{p(1-p)}$ is a *population* standard deviation. Since we don't know the true (population) parameter p, we can substitute it for the sample parameter $\hat{p}$, and thus obtain the *sample* standard deviation $\sqrt{\hat{p}(1-\hat{p})}$. Then we will get the following (good) approximation of the 95% confidence interval for p:

$$\left[\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}), \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}})\right],$$

So, with probability 0.95 the true parameter p lies within this confidence interval.
Version 2
The similar arguments may be employed to construct a confidence interval for the mean death rate represented by the parameter λ from the Poisson distribution. Particularly, as soon as the size T of the sample used to calculate $\hat{\lambda}$ is sufficiently large(the rule of thumb for practical purposes is that $T > 30$), we can use the Central limit theorem. According to this theorem,

$$\frac{\hat{\lambda} - E(\hat{\lambda})}{\sqrt{var(\hat{\lambda})}} \sim N(0, 1)$$

and since $E(\hat{\lambda}) = \lambda$ and $var(\hat{\lambda}) = \frac{\lambda}{n}$, the 95% confidence interval for λ is

$$\left[\hat{\lambda} - 1.96\sqrt{\frac{\hat{\lambda}}{n}}, \hat{\lambda} + 1.96\sqrt{\frac{\hat{\lambda}}{n}}\right],$$

5. Suppose X is a Poisson random variable with an (average) arrival rate of λ counts (occurrences of a particular event, such as the arrival of a bus) per unit of time. Thus, $E(X) = \lambda$. The probability of seeing a particular number k of counts in *one* unit of time is $P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$, $k = 0, 1, 2, 3, ...$

For the same random variable, if we want to answer a question that involves an interval of $t > 0$ units of time rather than 1 unit of time, then we have a new random variable X_t that is Poisson distributed with mean λt. Now the probability of seeing exactly k counts in t units of time is $P(X_t = k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$.

(a) Now consider a Poisson process with rate λ per unit time and the random variable W, which is the time one must wait to see the first count.

Explain briefly why the following two events are equivalent: $\{W > t\} = \{X_t = 0\}$.

(b) Show that the cumulative distribution of W is $P(W < t) = 1 - e^{-\lambda t}$ and also work out the density function of W. This is the so-called exponential density with rate λ. (Use the complement rule.)

(c) Can you generalize the argument to work out the density function of the random variable Z for the time one must wait to see the n^{th} count?

(a) By definition, events A and B in a random experiment are said to be equivalent if the probability of the symmetric difference is 0:

$$P(\{A \setminus B\} \cup \{B \setminus A\}) = P(\{A \setminus B\}) = P(\{B \setminus A\}) = 0$$

Intuitively, equivalent events are indistinguishable from a probability point of view. As soon as one of the events occurs, the other occurs as well.

In our problem events $\{W > t\}$, $\{X_t = 0\}$ are equivalent because both conditions, $W > t$ and $X_t = 0$, say that no counts arrive in the first t units of time. As soon as the time of waiting for the first count is longer than t units of time, no counts occurs per time period t. And the other way round, as soon as no counts arrive during the time interval t, the time one must wait to see the first count is longer than t units.

(b) Using the result of part (a), the cumulative distribution function of W can be derived as follows:

$$F(t) = P(W < t) = 1 - P(W > t) = 1 - P(X_t = 0) = 1 - (\lambda t)^0 \frac{e^{-\lambda t}}{0!} = 1 - e^{-\lambda t}.$$

The density function f of W can be found by differentiating the cumulative distribution function F with respect to t. We obtain

$$f(t) = \frac{dF(t)}{dt} = \lambda e^{-\lambda t} \text{for } t \geq 0.$$

Function

$$f(t) = \begin{cases} \lambda e^{-\lambda t} & \text{if } t \geq 0 \\ 0 & \text{if } t < 0 \end{cases}$$

is called the exponential density function with the rate λ.

(c) Let Z be the time one must wait to see the n^{th} count. Generalizing the argument used above, events $\{Z > t\}$, $\{X_t \leq n-1\}$ are equivalent. Therefore, the cumulative distribution function of Z is

$$G(t) = P(Z < t) = 1 - P(Z > t) = 1 - P(X_t \leq n-1) = 1 - \sum_{k=0}^{n-1} (\lambda t)^k \frac{e^{-\lambda t}}{k!}.$$

Differentiating G with respect to t we obtain the density function g of Z:

$$g(t) = \frac{dG(t)}{dt} = -\sum_{k=0}^{n-1} \left[k\lambda^k t^{k-1} \frac{e^{-\lambda t}}{k!} - \lambda(\lambda t)^k \frac{e^{-\lambda t}}{k!} \right] = -\lambda \left[\sum_{k=1}^{n-1} (\lambda t)^{k-1} \frac{e^{-\lambda t}}{(k-1)!} - \sum_{k=0}^{n-1} (\lambda t)^k \frac{e^{-\lambda t}}{k!} \right] =$$

$$= -\lambda \left[\sum_{k=0}^{n-2} (\lambda t)^k \frac{e^{-\lambda t}}{k!} - \sum_{k=0}^{n-1} (\lambda t)^k \frac{e^{-\lambda t}}{k!} \right] = -\lambda \left[-(\lambda t)^{n-1} \frac{e^{-\lambda t}}{(n-1)!} \right] = \lambda^n t^{n-1} \frac{e^{-\lambda t}}{(n-1)!}$$

6. The table shows the number of newborn boys and girls in the UK in 2003 and 2004.

(a) Set up a Bernoulli model for the 2003 data and estimate (by maximum likelihood) the parameter p giving the probability of a male birth.

(b) Consider a joint model for the data for 2003 and 2004, where p can vary across the two years (denote this by p_1 and p_2 respectively), and where all the observations are independent. Assuming that the joint likelihood is found by multiplying the two marginal likelihoods for 2003 and 2004, estimate p_1 and p_2.

(c) Finally, formulate (within the likelihood framework) a test of the null hypothesis $p_1 = p_2$ and test the hypothesis by computing a likelihood ratio statistic and comparing with a $\chi^2(1)$ density function.

(a)Let X be a random variable such that $X = 1$ indicates a male birth and $X = 0$ indicates a female birth. A Bernoulli density function is

$$P(X = x) = p^x (1-p)^{1-x}, x = 0 \text{ or } 1.$$

The method of maximum likelihood estimates p by finding the value of p that maximizes the likelihood function $L(x_1, .., x_n | p)$. To construct the likelihood function associated with a *known* probability density function we draw a sample $x_1, .., x_n$ (the data points) of n values from this distribution and compute the (multivariate) probability density function using our observed data.

As we already saw in Problem 4, likelihood function associated with the Bernoulli distribution is

$$L(x_1,..,x_n|p) = P(X_1 = x_1,..,X_n = x_n|p) = p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i},$$

where $x_i = 0$ or 1, $i = 1,..,n$, and $n = \#girls_{2003} + \#boys_{2003} = 695549$.

The corresponding log-likelihood function is

$$lnL(x_1,..,x_n|p) = ln(p)\sum_{i=1}^{n} x_i + ln(1-p)(n - \sum_{i=1}^{n} x_i).$$

Taking the derivative of this function with respect to p and equating it with zero, we obtain the ML estimate of p:

$$\hat{p} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum_{i=1}^{\#boys_{2003}} 1}{n} = \frac{356578}{695549} \approx 0.512657$$

(b) Since all observations in the data for 2003 and 2004 are independent, the joint density function is a product of the two marginal density functions:

$$P(X_1 = x, X_2 = y) = p_1^x(1-p_1)^{1-x}p_2^y(1-p_2)^{1-y},$$

where $x = 0$ or 1 and $y = 0$ or 1.

Similarly, the joint likelihood function is a product of the two marginal likelihoods.

$$L(x_1,..,x_n,y_1,..,y_m|p_1,p_2) = P((X_1)_1 = x_1,..,(X_1)_n = x_n,(X_2)_1 = y_1,..,(X_2)_m = y_m|p_1,p_2) =$$

$$= p_1^{\sum_{i=1}^{n} x_i}(1-p_1)^{n-\sum_{i=1}^{n} x_i}p_2^{\sum_{j=1}^{m} y_j}(1-p_2)^{m-\sum_{j=1}^{m} y_j},$$

where $x_i = 0$ or 1, $i = 1,..,n$, $\quad y_j = 0$ or 1, $j = 1,..,m$,
$n = \#girls_{2003} + \#boys_{2003} = 695549$ (number of 2003 observations), and
$m = \#girls_{2004} + \#boys_{2004} = 715996$ (number of 2004 observations).

The corresponding joint log-likelihood function is

$$lnL(x_1,..,x_n,y_1,..,y_m|p_1,p_2) = ln(p_1)\sum_{i=1}^{n} x_i + ln(1-p_1)(n-\sum_{i=1}^{n} x_i) + ln(p_2)\sum_{j=1}^{m} y_j + ln(1-p_2)(m-\sum_{j=1}^{m} y_j).$$

Differentiate lnL with respect to p_1 and with respect to p_2 and set both derivatives equal to zero. We obtain

$$\frac{dlnL}{dp_1} = \frac{\sum_{i=1}^{n} x_i}{p_1} - \frac{n - \sum_{i=1}^{n} x_i}{1-p_1} = 0$$
$$\frac{dlnL}{dp_2} = \frac{\sum_{j=1}^{m} y_j}{p_2} - \frac{m - \sum_{j=1}^{m} y_j}{1-p_2} = 0$$

So, the ML estimate of p_1 is $\hat{p_1} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum_{i=1}^{\#boys_{2003}} 1}{n} = \frac{356578}{695549} \approx 0.512657$, and the ML estimate of p_2 is $\hat{p_2} = \frac{\sum_{j=1}^{m} y_j}{m} = \frac{\sum_{j=1}^{\#boys_{2004}} 1}{m} = \frac{367586}{715996} \approx 0.513391$.

Comparing this result with the result in part (a), one can see that the estimates of p_1 and p_2 maximizing the joint likelihood are identical to the corresponding estimates maximizing the marginal likelihoods for 2003 and 2004.

(c) The hypotheses we need to test are $H_0 : p_1 = p_2$ versus $H_1 : p_1 \neq p_2$. We use a likelihood-ratio test. First, compute a likelihood-ratio statistic Λ associated with the pooled data on the child's sex for the years 2003 and 2004:

$$0 \leq \Lambda = \frac{\max L(x_1, .., x_n, y_1, .., y_m | p_1 = p_2 = p)}{\max L(x_1, .., x_n, y_1, .., y_m | p_1 \neq p_2)} \leq 1$$

The numerator is the maximized value of the joint likelihood function *under the null hypothesis*. Being maximum likelihood, it provides the most favorable evidence for H_0. We contrast that with the best evidence over all possible values of p_1, p_2 by comparing the relative likelihood of $p_1 = p_2$ with all other alternatives, and do not reject H_0 when Λ is close to unity. Conversely, when the numerator (restricting p_1 to be equal p_2) is a small percentage of the denominator, the evidence seems strongly against the null being true.

Now, to test the hypotheses formally, first compute the value of the likelihood ratio.

The numerator Consider the joint likelihood function under H_0.

$$L(x_1, .., x_n, y_1, .., y_m | p_1 = p_2 = p) = p^{\sum_{i=1}^{n} x_i + \sum_{j=1}^{m} y_j} (1-p)^{n+m-\left(\sum_{i=1}^{n} x_i + \sum_{j=1}^{m} y_j\right)},$$

where $n = 695549$ (number of 2003 observations), and $m = 715996$ (number of 2004 observations).

$maxL(x_1, .., x_n, y_1, .., y_m | p_1 = p_2 = p)$ is then provided by the ML estimate of p. From parts (a), (b) of the solution above we know that ML estimate $\hat{p}$ is computed as $\hat{p} = \frac{\sum_{i=1}^{n} x_i + \sum_{j=1}^{m} y_j}{n+m} = \frac{\sum_{i=1}^{\#boys_{2003}+\#boys_{2004}} 1}{n+m} = \frac{724164}{1411545} \approx 0.513029$. So,

$$\max L(x_1, .., x_n, y_1, .., y_m | p_1 = p_2 = p) \approx 0.513029^{724164} (0.486971)^{687381}$$

The denominator The unrestricted joint likelihood function is

$$L(x_1, .., x_n, y_1, .., y_m | p_1, p_2) \approx p_1^{\sum_{i=1}^{n} x_i} (1-p_1)^{n-\sum_{i=1}^{n} x_i} p_2^{\sum_{j=1}^{m} y_j} (1-p_2)^{m-\sum_{j=1}^{m} y_j}$$

(see part (b)).

The maximum of this function is achieved at $\hat{p_1} = \frac{\sum_{i=1}^{n} x_i}{n} =\approx 0.512657$, and $\hat{p_2} = \frac{\sum_{j=1}^{m} y_j}{m} =\approx 0.513391$. So,

$$\max L(x_1, .., x_n, y_1, .., y_m | p_1, p_2) \approx$$
$$\approx 0.512657^{356578} (0.487343)^{338971} 0.513391^{367586} (0.486609)^{348410}$$

Hence, the likelihood ratio $\Lambda = \frac{0.513029^{724164}(0.486971)^{687381}}{0.512657^{356578}(0.487343)^{338971}0.513391^{367586}(0.486609)^{348410}}$.

For the likelihood-ratio test we now need to compute $-2log\Lambda$ and compare it to a $\chi^2(1)$ critical value. Since $log\Lambda$ is a negative number close to zero when $\Lambda \approx 1$, we do not reject H_0 if $-2log\Lambda$ is small enough. In our case $-2log\Lambda \approx 0.76$ and the 5% critical value from the table for the χ^2-statistic with 1 degree of freedom is 3.84. $0.76 < 3.84$, therefore, based on the likelihood-ratio test we *cannot reject* the null hypothesis at 5% significance level.

7.

(a) Defining $B(n,k,p) = \frac{n!}{(n-k)!k!}p^k(1-p)^{n-k}$; $k = 0,1,2,...n$, show that:
$lim_{n\to\infty;np=\lambda} B(n,k,p) = \frac{e^{-\lambda}\lambda^k}{k!}$

(b) Derive the m.g.f. for a random variable with probability distribution function $B(n,k,p)$ and show it is equal to $M_{B(n,k,p)}(t) = ((1-p)+pe^t)^n$

(c) Work out the m.g.f. of a Poisson distribution with parameter λ, and show that under the same limiting process as in (a) above, the m.g.f. derived in (b) coincides with the m.g.f. for the Poisson.

(a) In the definition equation of $B(n,k,p)$, the first thing to do is to replace p by $\frac{\lambda}{n}$. We obtain:

$$B(n,k,p) = \frac{n!}{(n-k)!k!}(\frac{\lambda}{n})^k(1-(\frac{\lambda}{n}))^{n-k} = \frac{\lambda^k}{k!}\frac{n(n-1)\cdots(n-k+1)}{n^k}(1-\frac{\lambda}{n})^n(1-\frac{\lambda}{n})^{-k}$$

Let us compute the limit of each of the terms the global expression is made up of.

$lim_{n\to\infty}\frac{n(n-1)...(n-k+1)}{n^k}$ To determine such a limit, one can use the following result:$lim_{n\to\infty}\frac{n-k}{n} = 1$, for any $k < \infty$. We can write:

$$lim_{n\to\infty}\frac{n(n-1)...(n-k+1)}{n^k} = lim_{n\to\infty}\frac{n}{n}lim_{n\to\infty}\frac{n-1}{n}...lim_{n\to\infty}\frac{n-k+1}{n} = 1*1*\cdots*1 = 1$$

$lim_{n\to\infty}(1-\frac{\lambda}{n})^k$

$$lim_{n\to\infty}(1-\frac{\lambda}{n})^k = 1^k = 1$$

$lim_{n\to\infty}(1-\frac{\lambda}{n})^n$ We can use the following result:

$$lim_{n\to\infty}(1-\frac{x}{n})^n = e^{-x}$$

Applied to our case, we have:

$$lim_{n\to\infty}(1-\frac{\lambda}{n})^n = e^{-\lambda}$$

In the end, we obtain:

$$lim_{n\to\infty} B(n,k,p) = \frac{\lambda^k}{k!} * 1 * 1 * e^{-\lambda} = \frac{\lambda^k e^{-\lambda}}{k!} = P(X = k),$$

where X is distributed given a Poisson distribution.
In practice, when $p < .1$ and $n > 50$, the Poisson is a good approximation of the binomial distribution.

(b)We have:

$$M_B(t) := E(e^{tx}) = \sum_{x=0}^{\infty} e^{tx} P(X = x) = \sum_{x=0}^{\infty} \frac{n!}{x!(n-x)!} e^{tx} p^x (1-p)^{n-x} =$$
$$= \sum_{x=0}^{\infty} \frac{n!}{x!(n-x)!} (e^t p)^x (1-p)^{n-x} = (e^t p + (1-p))^n$$

(c) First, derive the m.g.f. of a Poisson distribution with parameter λ.

$$M_{poi}(t) := E(e^{tX}) := \sum_{x=0}^{\infty} \frac{e^{tx} \lambda^x e^{-\lambda}}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(e^t \lambda)^x}{x!}$$

One can easily notice that the sum is the series development of $e^{e^t \lambda}$. Therefore, we obtain:

$$M_{poi}(t) = e^{-\lambda} e^{e^t \lambda} = e^{\lambda(e^t - 1)}$$

Now, show that the limit of the m.g.f. derived in (b) coincides with the m.g.f. for the Poisson.

$$lim_{n\to\infty \ np=\lambda} M_B(t) := lim_{n\to\infty \ np=\lambda} (e^t \frac{\lambda}{n} + (1 - \frac{\lambda}{n}))^n = lim_{n\to\infty \ np=\lambda} (1 + (e^t - 1)\frac{\lambda}{n})^n = e^{\lambda(e^t - 1)}$$

Both MGF coincide.