Math 10B with Professor Stankova
Worksheet, Discussion #22; Tuesday, 4/16/2019
GSI name: Roy Zhao

Maximum Likelihood Estimation (MLE)

Concepts

1. We often are not given the actual values for the parameters of different discrete random variables and instead have to estimate what they are. To do this, we calculate the **likelihood function** $L(\theta)$ which is the probability that we see the data we see if we set the parameter equal to θ. Namely, $L(\theta|x_1, \ldots, x_n) = P(x_1, \ldots, x_n|\theta)$, the probability we see $x_1, \ldots, x_n$ if our parameter is equal to θ. Then we choose the value of θ that maximizes this function by taking the derivative and setting it equal to 0.

Examples

2. Let $f(x) = \frac{1}{2\sqrt{2\pi}} e^{-(x-5)^2/8}$ be a PDF. Calculate the probability $P(3 \leq X \leq 7)$.

Solution: The formula for a normal distribution is $\frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$, and so we see that $\mu = 5$ and $\sigma = 2$. We split up the probability $P(3 \leq X \leq 7)$ at the mean to get $P(3 \leq X \leq 7) = P(3 \leq X \leq 5) + P(5 \leq X \leq 7)$. We use z scores. In order to calculate z scores, we take $\frac{|a-\mu|}{\sigma}$ where a is the value you want to take the z score of. So, the z score of 3 is $\frac{|3-5|}{2} = 1$ and the z score of 7 is $\frac{|7-5|}{2} = 1$. So the probability is

$$z(1) + z(1) = 0.3413 + 0.3413 = 0.6826.$$

3. A Pareto distribution is given by the PDF $f(x) = \frac{p}{x^{p+1}}$ for $x \geq 1$ and 0 for $x < 1$ for some parameter p. Suppose I draw from this distribution 4 times and get the values $1, 2, 1, 1, 1$. What is the 95% confidence interval for μ?

Solution: The mean of the Pareto distribution is $\frac{p}{p-1}$. The standard deviation is $\sqrt{\frac{p}{(p-1)^2(p-2)}}$. So again we calculate $\hat{\mu} = \frac{\hat{p}}{\hat{p}-1} = \frac{1+2+1+1+1}{5} = \frac{6}{5}$ so we see $\hat{p} = 6$. Then $\hat{\sigma} = \sqrt{\frac{6}{5^2 \cdot 4}} = \frac{\sqrt{6}}{10}$. So the 95% confidence interval for μ is

$$(\hat{\mu} - 2\frac{\hat{\sigma}}{\sqrt{n}}, \hat{\mu} + 2\frac{\hat{\sigma}}{\sqrt{n}}) = (\frac{6}{5} - \frac{2 \cdot \sqrt{6}/10}{\sqrt{5}}, \frac{6}{5} + \frac{2 \cdot \sqrt{6}/10}{\sqrt{5}}).$$

4. I have a bag with 5 red and blue balls. I pull out a ball and it is red. I put it back and I add 3 blue balls and pull out another ball, which is blue. What is the maximum likelihood for the original number of blue balls.

Solution: Let n be the given number of blue balls. Then the probability for picking a red is $1-n/5$. After, the probability of getting a blue ball after adding 3 blue balls is $(n+3)/8$. So, we want to maximize

$$\frac{5-n}{5}\cdot\frac{n+3}{8}=\frac{-n^2+2n+15}{40}.$$

Taking the derivative and setting it equal to 0, we get that $2n=2$ or $n=1$ is the maximum likelihood.

5. You assume that the lifespan of lightbulbs are exponentially distributed (PDF is $\lambda e^{-\lambda t}$ for $t\geq 0$) and notice that your three light bulbs go out in 1, 2, and 3 years. What is the maximum likelihood estimator for λ?

Solution: We want to find the maximum likely λ given our sample of $1,2,3$. So, we want to maximize $L(\lambda|x_1,x_2,x_3)$ where $x_i=i$. By definition, we have that $L(\lambda|x_1,x_2,x_3)=P(x_1,x_2,x_3|\lambda)=P(x_1|\lambda)P(x_2|\lambda)P(x_3|\lambda)$ by independence. We calculate that as $\lambda^3e^{-6\lambda}$. In order to find the maximum, we take the derivative and set it equal to 0 to get

$$3\lambda^2e^{-6\lambda}-6\lambda^3e^{-6\lambda}=0\implies\lambda=0,\frac{1}{2}.$$

The solution $\lambda=0$ doesn't make sense and hence $\lambda=1/2$.

Problems

6. True **FALSE** In MLE, you calculate the probability that your parameter θ is a given value.

Solution: You calculate the likelihood that the parameter is a given value by calculating the probability of the outcomes given the outcome.

7. True **FALSE** In the ball example from above, the solutions is always found by setting the derivative to 0.

Solution: If I solve and get a noninteger number, like 2.2, then I have to plug in $n = 2$ and $n = 3$ and take the higher one.

8. **TRUE** False The likelihood function to determine the probability of flipping heads with a biased coin if we see 3 heads out of 5 flips is a polynomial of degree 5.

Solution: If p is the probability of flipping heads, then the likelihood function is $p^3(1-p)^2$.

9. The number of students in a class that can Dougie is normally distributed mean 15 and standard deviation 4. Calculate the probability that in this class between 16 and 20 students can Dougie.

Solution: We have $P(16 \le X \le 20) = P(15 \le X \le 20) - P(15 \le X \le 16)$ and we calculate the z scores. The first is $\frac{|20-15|}{4} = \frac{5}{4}$ and the second is $\frac{|16-15|}{4} = \frac{1}{4}$. Thus, the probability is $z(1.25) - z(0.25)$.

10. The number of students in a class that knows how to do “The Woah” is normally distributed with mean 5 and standard deviation 2. Calculate the probability that between 4 and 8 students know how to do it.

Solution: We have $P(4 \le X \le 8) = P(4 \le X \le 5) + P(5 \le X \le 8)$ and we calculate the z scores. The first is $\frac{|4-(5)|}{2} = \frac{1}{2}$ and the second is $\frac{|8-(5)|}{2} = \frac{3}{2}$. Thus, the probability is $z(0.5) + z(1.5)$.

11. There is a bag with 12 balls colored red and blue. You pull out three balls (with replacement) and get BBR. What is the maximum likelihood for the number of blue balls in bag?

Solution: Let n be the number of blue balls. Then the likelihood is given by $\frac{n}{12} \cdot \frac{n}{12} \cdot \frac{12-n}{12}$. Taking the derivative and setting equal to 0 gives $n = 8$.

12. I go to Kip’s and want to figure out the total number of students n there. By looking, I see that I’ve taught 2 of the students there. I pick a student at random and it turns out to be one of the students I’ve taught. Then 8 more students come in and of those I have taught 7 of them. Now I again pick a random student and I haven’t taught this second student picked. What is the most likely number n of total students at Kip’s originally?

Solution: We want to find $L(n|pick)$ the likelihood of n given what I picked. This is just the probability. So originally, there are 2 students I've taught and $n-2$ that I haven't. The probability of picking a student I've taught is $\frac{2}{n}$. Then 8 more come in and I've taught a total of $2+7=9$ of them and not taught $n-2+1=n-1$ of them out of a total of $n+8$ students. So the probability of picking someone I didn't teach is $\frac{n-1}{n+8}$. So the likelihood function is

$$L(n)=\frac{2}{n}\cdot\frac{n-1}{n+8}=\frac{2n-2}{n^2+8n}.$$

We take the derivative and set it equal to 0. This gives

$$\frac{(n^2+8n)\cdot 2-(2n-2)(2n+8)}{(n^2+8n)^2}=\frac{-2(n-4)(n+2)}{(n^2+8n)^2}=0.$$

So we get $n=4$ or $n=-2$. But $n=-2$ which doesn't make sense. Then note that f' is positive before and negative after so this is a local and global max. So the maximum likelihood estimate is $\hat{n}=4$.

13. You have a coin that you think is biased. you flip it 4 times and get the sequence $HHHT$. What is the maximum likelihood estimate for the probability of getting heads?

Solution: Let p be the probability of getting heads. Then the probability of getting $HHHT$ is $p^3(1-p)$. Taking the derivative and setting equal to zero gives $3p^2-4p^3 \implies p=\frac{3}{4}$.

14. During Cal Day, my two friends and I asked prospective students where they were from until we found someone who wasn't from California. We had to ask $23, 18, 46$ people respectively before finding someone not from California. What is the maximum likelihood estimate for the percentage of students from California?

Solution: This is a geometric distribution. Let p be the probability of success which in this case is finding someone not from California. Then, the likelihood function is

$$L(p|23,18,46)=P(23,18,46|p)=(1-p)^{23}p\cdot(1-p)^{18}p\cdot(1-p)^{46}p=(1-p)^{87}p^3.$$

We want to maximize this and hence take its derivative and set it equal to 0. Since this is a product, we take its log derivative. So we first take the log which gives $87\ln(1-p)+3\ln p$. Then we take the derivative and set it equal to 0 to get

$$\frac{-87}{1-p}+\frac{3}{p}=0.$$

So $\frac{3}{p} = \frac{87}{1-p}$ and $3 - 3p = 87p$ so $p = \frac{1}{30}$ is the maximum. Thus, the maximum likelihood for the percentage of students from California is 1 minus this or $\frac{29}{30}$.