Math 30440 — Probability and Statistics

Spring 2010 final exam practice problems

Instructors: David Galvin and Daniel Cibotaru

1. At the end of the day in the amusement park you want to try out both the white water rafting ride and the roller coaster. The time it takes to stand in line for and ride the white water ride is normally distributed with mean 30 minutes, variance 9. The time for the roller coaster is normally distributed with mean 50 minutes and variance 16. If the park closes in 90 minutes, what's the probability you will get to go on both rides? Assume it takes no time to go from one ride to the other.

   Solution: Total time for both rides is

   $$\text{Normal}(30, 9) + \text{Normal}(50, 16) = \text{Normal}(80, 25).$$

   So probability of getting in both rides is probability $\text{Normal}(80, 25) \leq 90$ or probability $Z < 2$.

2. During the slack time from 1pm to 5pm, customers arrive at Reckers at a constant rate of 11 per hour. A customer has just arrived. Let T be the time until the next customer arrives.

   (a) Using an exponential random variable to model T, calculate the probability that it will be more than 10 minutes until the next customer arrives.

   Solution: Model T by exponential with $\lambda = 11$ (average number of occurrences per unit time, unit being 1 hour). Then

   $$P(T > 10 \text{ minutes}) = P(T > 1/6) = \int_{1/6}^{\infty} 11e^{-11x}\, dx.$$

   Alternatively, taking units as minutes, we should take $\lambda = 11/60$ to get

   $$P(T > 10) = \int_{10}^{\infty} \frac{11}{60} e^{-\frac{11x}{60}}\, dx.$$

   (b) Suppose that 5 minutes have passed, and no customer has arrived. Calculate the probability that it will be more than a further 10 minutes until the next customer arrives.

Solution: By memorylessness, this will exactly the same as the probability calculated in part a).

3. A random variable has the following density function:

$$f(x) = \begin{cases} \frac{C}{x^4} & x \geq 1 \\ 0 & x < 1. \end{cases}$$

Find C, and compute the expectation and variance of the random variable.

Solution: To find C, solve

$$\int_1^\infty \frac{C}{x^4}\, dx = 1$$

to get $C = 3$. Then

$$E(X) = \int_1^\infty x\frac{3}{x^4}\, dx = \int_1^\infty \frac{3}{x^3}\, dx, \quad E(X^2) = \int_1^\infty \frac{3}{x^2}\, dx$$

and $\mathrm{Var}(X) = E(X^2) - E(X)^2$.

4. Your boss needs to know the average compressive force (in pounds per square inch) the type of concrete you are using for a new bridge will support before failing. Using the company laboratory you test ten samples and discover they fail at the following pressures:

3000, 3300, 3180, 3360, 3240, 3420, 3120, 3600, 3300, 3480

Assuming the pressure at which a test pieces fail is normally distributed give your boss a 95% confidence interval for the mean pressure this kind of concrete can withstand.

Solution: $n = 10$, $\bar{X} = 3300$, $S = 176.64$, $t_{09,.025} = 2.262$, so confidence interval is

$$\left(3300 - 2.262 \times \frac{176.64}{\sqrt{10}}, 3300 + 2.262 \times \frac{176.64}{\sqrt{10}}\right)$$

5. The lightbulb in the reading lamp by your bed has a lifetime that is exponentially distributed with average lifetime 8 months. Your roommate uses a bulb in his reading lamp that has lifetime that is exponentially distributed with mean 6 months.

   (a) What is the probability that you will not have to replace your lightbulb anytime in the next 12 months?

   Solution: X is lifetime of your lightbulb; X is exponential with $\lambda = 1/8$ (so its expected value is 8).

$$P(X > 12) = \int_{12}^\infty \frac{1}{8} e^{-\frac{x}{8}}\, dx.$$

(b) What is the probability that you will have to replace your lightbulb sooner that your roommate has to replace his?

Solution: Y is lifetime of roommates lightbulb; Y is exponential with $\lambda = 1/6$. Assuming independence, joint density of X and Y is

$$f(x,y) = \begin{cases} \frac{1}{48}e^{-\frac{x}{8}-\frac{y}{6}} & x, y \geq 0 \\ 0 & otherwise. \end{cases}$$

We want $P(X < Y)$, which is

$$\int_{x=0}^{\infty} \int_{y=x}^{\infty} \frac{1}{48} e^{-\frac{x}{8}-\frac{y}{6}}\ dydx.$$

6. Give short answers to each of the following questions.

(a) What is a Type I error in a hypothesis test?

Solution: Rejecting null when it is true

(b) What is a Type II error in a hypothesis test?

Solution: Accepting null when it is false

(c) A certain null hypothesis is accepted at 2% significance. With the same data, will it be accepted at 5% significance?

Solution: Can't say. Accepted at 2% means p-value $> .02$; accepted at 5% means p-value $> .05$. So if p-value is between .02 and .05, it will be rejected at 5%, otherwise accepted.

(d) John constructs a 90% confidence interval for a certain parameter. Mary uses the same data to construct a 95% confidence interval for the same parameter. Whose interval is shorter?

Solution: John's; Mary needs a longer interval to be more confident.

7. I want to test a hypothesis about the mean μ of a certain normal population whose variance is known to be 4. My null hypothesis is $H_0 : \mu = 16$ and my alternative is $H_1 : \mu \neq 16$. Initially I sample from the population 25 times.

(a) If I'm testing at 5% significance, what is the range of values of $\bar{X}$, my sample mean, that will lead to me accepting the null?

Solution: I will accept null as long as

$$-1.96 \leq \frac{\bar{X} - 16}{2/\sqrt{25}} \leq 1.96,$$

that is, $15.216 \leq \bar{X} \leq 16.784$.

(b) Suppose that the true mean is actually 20. What is the probability that I will incorrectly accept the null?

Solution:

$$\begin{aligned} P(\text{accept null when } \mu = 20) &= P(15.216 \leq \bar{X} \leq 16.784 \mid \bar{X} = \text{Normal}(20, 4/25)) \\ &= P\left(\frac{15.216 - 20}{2/5} \leq Z \leq \frac{16.784 - 20}{2/5}\right) \\ &= P(-11.96 \leq Z \leq -8.04) \\ &\approx P(Z \leq -8.04) \approx 0. \end{aligned}$$

(c) How large a sample would I have to take to make sure that there is only 5% probability of incorrectly accepting the null when the true mean is actually 20?

Solution: Certainly 25 is enough. Using formula on page 301,

$$n \approx \frac{4(1.96 + 1.645)^2}{(20 - 16)^2} = 3.25.$$

So even $n = 4$ would be enough.

8. I'm willing to use a certain manufacturer's theodolite (angle-measuring device) as long as I am sure that it is consistent in its readings. So I test the theodolite by measuring the same angle 20 times; the answers I get have sample variance .8 degrees. I conclude that I am 90% confident that the variance σ^2 of measurements for the theodolite is no more than c.

(a) What is a reasonable assumption to make about the distribution of measurements for this theodolite that would allow me to compute c?

Solution: Normally distributed, different readings independent of each other.

(b) What is c?

Solution: We know that $\frac{(n-1)S^2}{\sigma^2}$ is a χ-squared distribution with $n - 1$ degrees of freedom, so with probability .9

$$\chi^2_{n-1,.9} \leq \frac{(n-1)S^2}{\sigma^2}$$

or

$$\sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{n-1,.9}}.$$

In this case $n = 20$, $S^2 = .8$ and $\chi^2_{n-1,.9} = 11.65$, so we take

$$c = \frac{19 \times .8}{11.65} = 1.3.$$

9. If $X_1, \ldots, X_n$ is a sample from a distribution given by

$$f(x) = \begin{cases} \frac{4x^3}{\theta} e^{-\frac{x^4}{\theta}} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases},$$

find the maximum likelihood estimator of θ.

Solution: The joint density is

$$\frac{\prod_{i=1}^n (4X_i^3)}{\theta^n} e^{-\sum_{i=1}^n \frac{X_i^4}{\theta}}$$

as long as all $x_i > 0$, and 0 otherwise. The logarithm is

$$\ln\left(\prod_{i=1}^n (4X_i^3)\right) - n \ln\theta - \sum_{i=1}^n \frac{X_i^4}{\theta}.$$

The derivative of this is

$$-\frac{n}{\theta} + \frac{\sum_{i=1}^n X_i^4}{\theta^2}$$

which is 0 at

$$\theta = \frac{\sum_{i=1}^n X_i^4}{n}.$$

This is the MLE.

10. The following numbers are drawn independently from a uniform distribution on the interval $-\theta \le x \le \theta$:

$$-2, 3, 1, -1, -2.$$

Find the maximum likelihood estimator of θ.

Solution: The joint density of the readings is $1/(2\theta)^5$ as long as all five readings lie between $-\theta$ and θ, and is 0 otherwise. To maximize, we choose θ as small as possible. Subject to the condition that we must have $-\theta \le x \le \theta$ for all readings x, the smallest that we can take θ is 3.

11. Experience has shown that when I am writing reports in the morning, I tend to make on average 1 typo per page. When I write in the afternoon, I make on average a .5 typos per page (1 per two pages). On report-writing days, I write four pages in the morning and five in the afternoon. Using a Poisson distribution (or distributions) to model the number of typos I make on report-writing days, answer the following questions:

    (a) Compute the probability that I make 2 or fewer typos in the morning.

Solution: Average number of typos in a morning is 4 (1 per page, 4 pages), so we use a Poisson distribution with parameter $\lambda = 4$ to model total number.

$$P(2 \text{ of fewer typos}) = \sum_{i=0}^{2} \frac{4^i}{i!} e^{-4}.$$

(b) What is the probability that I make 2 or fewer typos during the whole day?

Solution: Average number of typos in a day is 6.5 (4 "morning errors" on average, plus 2.5 "afternoon errors" on average), so we could use a Poisson distribution with parameter $\lambda = 6.5$ to model total number.

$$P(2 \text{ of fewer typos}) = \sum_{i=0}^{2} \frac{6.5^i}{i!} e^{-6.5}.$$

12. When a stabilizer is used on poles being dropped into the ocean to act as foundations for an oil rig, the x and y coordinates of the distance from the final location of the pole to its intended location are independent and normally distributed, each with mean 0 and standard deviation σ meters. The value of σ can be set in advance, but the smaller σ is the costlier the process.

(a) If σ is set to 3, what is the probability that the pole will end up within 1 meter of its intended destination?

Solution: Let D be distance from location. We have $D^2 = x^2 + y^2$, so $D^2/\sigma^2 = Z^2 + Z^2 = \chi_2^2$.

$$P(D \leq 1) = P(D^2 \leq 1) == P(D^2/9 \leq 1/9) P(\chi_2^2 \leq 1/9) = .054.$$

(from an online χ^2 calculator).

(b) What should σ be set to so that one can be 99% certain that the pole will end up within 1 meter of its intended destination?

Solution: We want $P(D \leq 1) = .99$, that is, $P(\chi_2^2 \leq 1/\sigma^2) = .99$. From an online calculator, $P(\chi_2^2 \leq 9.21) = .99$, so we should take $1/\sigma^2 = 9.21$ and $\sigma = .329$.

13. A box of nails contains 6 produced by Sharp Co, 4 by Nail Inc and 2 by Hammermate Ltd (so 12 nails in all).

(a) You select a nail at random. What's the probability that it was produced by Hammermate Ltd?

Solution: $2/12 = .166....$

(b) You know that nails produced by Sharp Co break on first use 1% of the time, those produced by Nail Inc break 2% of the time, those produced by Hammermate Ltd break 5% of the time. What's the probability that the nail you picked breaks on first use?

Solution: By law of total probability, it's $.5 \times .01 + .333 \times .02 + .166 \times .05$.

(c) Given the information that the nail you picked does break on first use, what's the probability that is was a Hammermate Ltd nail?

Solution: By Bayes' formula, it's

$$\frac{.166 \times .05}{.5 \times .01 + .333 \times .02 + .166 \times .05}.$$

14. Suppose that 70% of the families in your (very large) city have no dogs, 22% have 1 dog and 8% have 2 dogs.

(a) Let X be the number of dogs that a randomly chosen family has. Compute $E(X)$ and $Var(X)$.

Solution: $E(X) = 0 \times .7 + 1 \times .22 + 2 \times .08 = .38$. Also, $E(X^2) = 0 \times .7 + 1 \times .22 + 4 \times .08 = .54$, so $\mathrm{Var}(X) = .54 - .38^2 = .3956$.

(b) Assuming your 200 family neighborhood constitutes a random sample and that families make their choices about dog ownership independently approximate the probability that their are more than 90 dogs in your neighborhood.

Solution: By the central limit theorem the number of dogs in a random sample of 200 is roughly normal with mean $200 \times .38 = 76$ and variance $200 \times .3956 = 79.12$. So

$$P(\text{more than 90 dogs}) = P(\text{Normal}(76, 79.12) > 90) = P(Z > 1.57).$$

15. I have a coin that comes up heads with some (unknown) probability q, and I want to estimate q. I do the following experiment: I toss the coin repeatedly until I first see a head, and let T be the number of tosses it takes until this happens. I repeat this experiment n times, and get the n readings $t_1, t_2, \ldots, t_n$ for T. Use this data to give a maximum likelihood estimator for q. (Note that T has the following mass function: $P(T = k) = (1-q)^{k-1}q$, $k = 1, 2, 3, \ldots$.)

Solution: The joint mass function of the n readings is

$$\prod_{i=1}^{n} \left((1-q)^{t_i - 1} q\right).$$

We want to choose q to maximize this. The logarithm of the function is

$$\ln(1-q) \sum_{i=1}^{n} (t_i - 1) + n \ln q.$$

The derivative of this with respect to q is

$$\frac{-\sum_{i=1}^{n}(t_i-1)}{1-q}+\frac{n}{q}.$$

This is 0 when

$$q=\frac{n}{\sum_{i=1}^{n} t_i}.$$

This is the MLE.

16. Professor G. wants to estimate the proportion of students who prefer multiple-choice exams to partial credit exams. Among 100 students currently in his class, 56 prefer multiple-choice exams.

    (a) Find a 90% confidence interval for the proportion of students who prefer multiple-choice exams.
    Solution: $\hat{p} = .56$. Our 90% confidence interval is either

    $$.56 \pm 1.645\sqrt{\frac{.56\times .44}{100}} \quad \text{or} \quad .56 \pm 1.645\sqrt{\frac{.5\times .5}{100}}$$

    depending on whether we choose to take .56 as an estimate for p or use the worst-case estimate of .5.

    (b) If the estimate above is to be within 0.05 of the true proportion favoring multiple-choice exams (with a 90% confidence), how many students should be sampled?
    Solution: We choose n large enough to make

    $$1.645\sqrt{\frac{.56\times .44}{n}} \le .05 \quad \text{or} \quad 1.645\sqrt{\frac{.5\times .5}{n}} \le .05$$

    (again depending on whether we choose to take .56 as an estimate for p or use the worst-case estimate of .5).

17. A person's response time to a stimulus is known to be normally distributed. I test someone's response time to the same stimulus six times, and get the following six time measurements (measured in tenths of a second):

    $$8 \quad 9.5 \quad 7.8 \quad 10 \quad 9.2 \quad 9.5$$

    (a) Compute the sample mean $\bar{X}$ and sample variance S^2 of this data.
    Solution: From excel, $\bar{X} = 9$ and $S^2 = .796$.

    (b) Construct a 95% confidence interval for the person's mean response time to the stimulus.
    Solution: $9 \pm t_{5,.025}\frac{\sqrt{.796}}{\sqrt{6}}$.

(c) Suppose I wanted a 95% confidence interval estimate for the mean that had width no more than .02 (in other words, an estimate of the form $(x - .01, x + .01)$). How many samples should I take? (Assume that the value for S^2 that you calculated in part a) is in fact the variance σ^2 of the population.)

Solution: Choose n large enough that

$$1.96\frac{\sqrt{.796}}{\sqrt{n}} \leq .01$$

(since the sample will need to be large enough that we will be reading t-values off the ∞ row).

18. Is there a difference between reading or listening when it comes to retaining information? 20 randomly selected people are broken into two groups. The first group (group A, consisting of 8 people) spend a week reading factsheets about US history; the second group (group B, 12 people) listen to the same factsheets being read on an audio book. At the end of a week all 20 people take a standardized test on which the results are assumed to be normally distributed. The mean for group A was 75 with sample variance 16, and for group B was 80 with sample variance 25.

(a) Assuming that the actual variances for the two types of groups are equal, use the data to compute S_p^2, the pooled estimator for the common variance.

Solution:

$$S_p^2 = \frac{7 \times 16 + 11 \times 25}{18} = 21.5.$$

(b) Test at 5% significance the null hypothesis that the there is no difference in the mean score between readers and listeners, against the alternative that there is a difference.

Solution: The test statistic we compute

$$\frac{75 - 80}{\sqrt{\frac{21.5}{8} + \frac{21.5}{12}}} = -2.3625.$$

Since $t_{18,.025} = 2.101$, we have seen an unusual extreme reading, and reject the null.