Math 366: Exam 1 Review Sheet

Disclaimer: This is not meant to be an exhaustive set of examples. Study all sections we have covered in Chapters 5 and 6.

Review of Estimators

Suppose we have a set of data and we suspect that it comes from a distribution with pdf (pmf) $f(x)$ where $f(x)$ has some unknown parameters. To highlight what the parameters are we sometimes write $f(x) = f(x; \theta_1, ..\theta_n)$ to indicate that the unknown parameters are $\theta_1, .., \theta_n$. We want to estimate these parameters (for example if we suspect the distribution is binomial, the parameter is p, if its Gamma, the parameters are α and θ).

<u>Maximum Likelihood Estimate:</u> Suppose we are trying to estimate parameters $\theta_1, ..\theta_n$.

1. Form the likelihood function $L(\theta_1, .., \theta_n) = \prod_{i=1}^{n} f(x_i; \theta_1, .., \theta_n)$.
2. Simpify if possible.
3. Take $ln(L(\theta_1, .., \theta_n))$ (We will maximize this function since it is usually simpler and has the same maximum as $L(\theta_1, .., \theta_n)$).
4. If there is just one parameter θ, find the derivative of $ln(L(\theta))$ with respect to θ, set equal to zero and try to find a maximum. If there are several parameters, take the partial derivatives and try to find a maximum where the partial derivatives are simultaneously equal to zero.

<u>Method of Moments:</u> Suppose we are trying to estimate parameters $\theta_1, ..\theta_n$.

In the method of moments, we *always* use the following equations.

1. $E[X] = \frac{1}{n}\sum_{i=1}^{n} X_i$
2. $E[X^2] = \frac{1}{n}\sum_{i=1}^{n} X_i^2$
3. ...
4. $E[X^k] = \frac{1}{n}\sum_{i=1}^{n} X_i^k$

If we have only one parameter to estimate we just use the first equation, if we have 5 we use the first 5, etc. Note that the left hand side is calculated using the definition of expected value which uses $f(x)$. This is the theoretical expected value of the random variables X^i. The right hand side is the sample values. If the values of the sample are given, the right hand side can be computed to get numbers.

<u>Biased vs. Unbiased Estimators:</u> When we find estimators for our parameters, it is useful to determine if we expect the values that they give us to be close to the parameter we are trying to estimate. Suppose we find that $\hat{\theta}$ is an estimator for θ. Then $\hat{\theta}$ is *unbiased* if its expected value is actually θ. That is, $E[\hat{\theta}] = \theta$.

6.1 Concepts to know:

- Displaying data - e.g., box and whisker plots
- 5-number summary: (min, $\tilde{q}_1$, median, $\tilde{q}_3$, max)
- $\tilde{q}_i$ is the ith percentile
- Interquartile range
- Outliers, Suspected outliers

Review Problems

1. Suppose X_1,...,X_n is a random sample of size n from a distribution with pdf $f(x) = \theta x^{\theta-1}$, for $0 \leq x \leq 1$, $0 \leq \theta < \infty$.
   (a) Find the Maximum Likelihood Estimator for θ.
   (b) Find an estimator for θ using the method of moments.

2. The following is an ordered sample of 10 test scores:

$$68,\ 70,\ 72,\ 72,\ 78,\ 80,\ 81,\ 86,\ 90,\ 100$$

   (a) Find the 15th percentile:
   (b) Find the 5 number summary
   (c) Skech a box-plot - include the inner and outer fences.
   (d) Are there any outliers or suspected outliers? Explain.

3. Suppose Y is Binomial $b(n, p)$. Show that $\hat{p} = \dfrac{Y}{n}$ is an unbiased estimator for p.

4. Cola X and Cola Y are competing in a taste test. People are asked to rate the colas from 1 to 10 with 10 being the best. A survey of 50 people tasting Cola X gave $\overline{x} = 7.6$. A survey of 65 people tasting Cola Y gave $\overline{y} = 8.5$. Assume $\sigma_x^2 = 7.5$ and $\sigma_y^2 = 8.1$. The makers of Cola Y would like to make the claim that they are 95% confident that people like thier cola better. Can they make this claim based on these statistics? Use a confidence interval to justify your answer.

5. A certain medication is supposed to contain 20% of the active ingredient. As a quality control measure, a sample of 12 was taken and the amount of active ingredient measured. The sample mean was $\overline{x} = 20$ and the sample variance was $s^2 = .1$. The company claims the true mean lies in the interval $(19.836, 20.164)$. What level of confidence is there in this interval?

6. In 2000, a poll showed that 79% of children entering Kindergarten had up-to-date immunizations. To update the estimate, how large a sample size is required to be 98% confident that the new estimate has an error no more that $\pm .05$?

7. (a) A 2005 survey showed that 37% of the American viewing audience watched NBC nightly news. If 2000 people were surveyed, find a 90% confidence interval for the proportion of the American viewing audience watching NBC nightly news.

   (b) If NBC wants to repeat the survey to confirm that the percentage still holds, what sample size should they use to maintain the 90% confidence level and have the error be no bigger that $\epsilon = .02$?

8. Suppose Y has the binomial distribution $b(100, .7)$. Use the normal distribution to estimate $P(80 \leq Y \leq 90)$.

9. (§5.4#15) The tensile strength X of paper has $\mu = 30$ and $\sigma = 3$ (pounds per square inch). A random sample of size $n = 100$ is taken from the distribution of tensile strengths. Compute the probability that the sample mean $\overline{X}$ is greater than 29.5 pounds per square inch.

10. (§6.4#7) Thirteen tons of cheese is stored in some old gypsum mines, including "22-pound" wheels (label weight). A random sample of $n = 9$ of these wheels yielded the following weights in pounds:

21.50	18.95	18.55	19.40	19.15
22.35	22.90	22.20	23.10	

Assuming that the distribution of the weights of the wheels of cheese is $N(\mu, \sigma^2)$, find a 96% confidence interval for μ.

11. (§6.7#13) In developing countries in Africa and the Americas, let p_1 and p_2 be the respective proportions of women with nutritional anemia. Find an approximate 90% confidence interval for $p_1 - p_2$, given that a random sample of $n_1 = 2100$ African women yielded $y_1 = 840$ with nutritional anemia and a random sample of $n_2 = 1900$ women from the Americas yielded $y_2 = 323$ women with nutritional anemia.

12. (§5.5#13) About 60% of all Americans have a sedentary life style. Select $n = 96$ Americans at random (assume independence). What is the probability that between 50 and 60 inclusive, do not exercise regularly?

13. (§6.8#9) A die has been loaded to change the probability of rolling a 6. In order to estimate p, the new probability of rolling a 6, how many times must the die be rolled so that we are 99% confident that the maximum error of the estimate of p is $\epsilon = 0.02$?

14. (§6.5#3) Independent random samples of the heights of adult males living in two countries yielded the following results: $n = 12$, $\overline{x} = 65.7$ inches, $s_x = 4$ inches and $m = 15$, $\overline{y} = 68.2$ inches, $s_y = 3$ inches. Find an approximate 98% confidence interval for the difference $\mu_X - \mu_Y$ of the means of the populations of heights. Assume that $\sigma_X^2 = \sigma_Y^2$.

15. (§6.2#7*b*) Let $f(x; \theta) = \theta x^{\theta - 1}, 0 < x < 1, \theta \in \Omega = \{\theta : 0 < \theta < \infty\}$. Let $X_1, X_2, ..., X_n$ denote a random sample of size n from this distribution.

    b) Show that $\hat{\theta} = -n/ln(\prod_{i=1}^{n} X_i)$ is the maximum likelihood estimator of θ.

16. THIS IS LACKING IN SECTION 5.4 AND 5.5 PROBLEMS