Math 502 Statistical Inference.
Spring 2015

- **Instructor:** Xingye Qiao
- **Email:** qiao@math.binghamton.edu
- **Phone number:** (607) 777-2593
- **Office:** WH 134

- **Meeting time & location:** MWF 8:30 - 9:30 at WH 100E.
- **Office hours:** MW 3:00 - 5:00

Prerequisite

Math 501.

Learning Objectives

1. Coverage: Chapters 6 through 10 in Casella & Berger.
2. Sufficiency, completeness, likelihood, estimation, testing, decision theory, Bayesian inference, sequential procedures, multivariate distributions and inference, nonparametric inference, asymptotic theory.

Recommended Texts

The required text is **Casella & Berger** (see below). Some reference texts are listed below as well.

- **Casella, G., & Berger, R. L. (2002)**. *Statistical inference.* Australia: Thomson Learning.
- Lehmann, E. L. (1999). Elements of large-sample theory. New York: Springer.
- Lehmann, E. L., & Casella, G. (1998). *Theory of point estimation.* New York: Springer.
- Shao, J. (1999). *Mathematical statistics.* New York: Springer.
  - Shao, J. (2005). *Mathematical Statistics: Exercises and Solutions.* New York, NY: Springer.
- Hogg, R. V. and Craig, A. (1995). *Introduction to Mathematical Statistics.* Prentice Hall, Englewood Cliffs, NJ

Grading

- Homework (40%): there will be weekly homework assignments, due at the beginning of each Wednesday class.
- Midterm exams (20%+20%): there will be two midterm exams.
  - Midterm exam 1: Friday, February 27, 2015
  - Midterm exam 2: Friday, April 3, 2015
- Final exam (20%): Wednesday, May 13, 2015 from 8:00 AM to 10:00 AM.

Homework assignments

Week 1

- 01/28: Notes, Example (2), (a)-(e); Textbook, Exercises 6.1, 6.3.
- 01/30: Notes, Example (4), b and c. Use both methods and for each method try to use different representations (so that your answers are not unique). Textbook, Exercises 6.2, 6.5, 6.8, 6.9.

Due on 02/04

Week 2

- 02/04 & 06: Textbook, Exercises 6.10, 6.11, 6.13, 6.14, 6.15, 6.18, 6.19, 6.20, 6.23, 6.25.
- 02/09:
  1. Let $X_{1}\cdots X_{n}$ be i.i.d with density function defined as $f(x|\theta)=e^{-\lambda(x-\mu)},~x>\mu,~\lambda>0,~\mu\in \mathbb{R}$. Prove that $(X_{(1)},W)$ is the sufficient statistic of $\theta=(\mu,\lambda)$, where $W=\sum^{n}_{i=2}(X_{(i)}-X_{(1)})$.
  2. Textbook: Exercises 7.1, 7.2, 7.6.

Due on 02/11

Week 3

- 02/11: Textbook, Exercises 7.7, 7.8, 7.10, 7.11.
- 02/13:
  1. Notes: Show that a Bayes estimator depends on the data through a sufficient statistic.
  2. Notes: if X_i's are iid given θ, are they iid marginally? Why?
  3. Textbook, Exercises 7.14, 7.22, 7.23 (read “conjugate prior” as “prior”), 7.24, 7.25. Solutions to some of the questions may appear clearer after you read the lecture notes for Monday's class.
- 2/16: Textbook: Exercises 7.9, 7.12, 7.50.

Due on 02/18

Week 4

- 02/18: Textbook, Exercises 7.19, 7.20, 7.21.
- 02/20: Textbook, Exercises 7.37, 7.46, 7.49, 7.51, 7.52.
- 02/23: Textbook, Exercises 7.53, 7.57, 7.58.

Due on 02/25

Week 5

- 02/27: Textbook, Exercises 7.59, 7.44, 7.48, 7.60, 7.63.
- 03/02: Textbook, Exercises 7.40, 7.65, 7.66.

Due on 03/06

Week 6

- 03/06: Textbook, Exercises 8.1, 8.2, 8.3, 8.5, 8.6.
- 03/09: Textbook, Exercises 8.7, 8.8, 8.9.

Due on 03/11

Week 7

- 03/11: Textbook, Exercises 8.12, 8.13, 8.15, 8.17, 8.20.
- 03/13: Textbook, Exercises 8.19, 8.21.
- 03/16:
  - Textbook, Exercises 8.25, 8.27.
  - In the class, I showed that if we remove $k>0$ from the necessity condition of the NPL, then when $k=0$, we must have $\beta_\phi(\theta_1)=1$ for the UMP level α test ϕ. Complete my work by arguing why in this case ϕ still must satisfy equation (2) in the NPL, that is, why it must be the case that $\phi=1$ when $f(x|\theta_1)>0$ wp1. [Hint: $\phi\le 1$. This proof should not last more than 2 lines.]

Due on 03/18

Week 8

- 03/18: Textbook, Exercises 8.28, 8.29, 8.30, 8.33.
- 03/20: Textbook, Exercises 8.37, 8.38, 8.47.
- 03/23: Textbook, Exercises 9.1, 9.2, 9.3.
- Additional: Carry out the following simulation project. Submit the R code and report the result properly.
  1. Use R to generate 10 observations from $N(1,4)$.
  2. Now pretend that you only known that the data were from $N(\mu,4)$ without knowing μ and construct a 80% confidence interval for μ.
  3. Repeat Steps 1 and 2 100 times.
  - Count the proportion among the 100 trials where the C.I. contains the true mean?
  - What is the relation between the proportion and the confidence coefficient?
  - Repeat Steps 1, 2 and 3, but pretend that you know neither the mean μ nor the variance σ^2. Then compare the lengths of the confidence intervals between the current and the previous settings. Make

comments on the lengths and discuss why there is a difference.

Due on 03/25

Week 9

- 03/25: Submit all your code and output, preferably using LaTex. In a numerical problem, unless stated otherwise, $1-\alpha=0.95$. Textbook, Exercises
  1. 9.4; In addition, assume that $n=10$ and $m=15$ and that $\sigma_X^2=1$ and $\sigma_Y^2=3$, generate some X_i's and Y_j's. Then use a numerical method to provide a CI based on the generated (observed) data. Then repeat the whole process for 1000 times. Report the number of time that the true $\lambda=3$ is covered by the CIs.
  2. 9.6; here assume that $X\sim bin(n,p)$ is observed and n is known. Next, let $n=50$ and generate X with $p=0.3$. Numerically provide the CI for the observed X. Repeat for 1000 times and report the number of times that the true $p=0.3$ is covered by the CI.
  3. 9.12.
  4. 9.13(b).
- 03/27: Textbook, Exercises: 9.16, 9.17 and 9.23. In 9.17, you need to find the shortest confidence interval using the pivotal method (and prove it using a result in class). Moreover, find in addition a second CI using pivotal method with equal left and right probabilities. Assume that $\alpha=0.05$ and verify that your shortest confidence interval is indeed shorter than the second one.
- 03/30:
  1. Textbook, Exercise: 9.37
  2. Assume that $X_1,\dots,X_n$ are iid from Cauchy, where $f(x)=[\pi (1+x^2)]^{-1}$.
     I. Calculate $\int_{-\infty}^\infty |x|f(x)dx$.
     II. What is the mean of X_1?
     III. Can we apply the SLLN to prove that $\overline X_n\rightarrow \mu_X$ a.s.?
     IV. Let $n=100$, simulate the sample and calculate $\overline X_n$. Then repeat this for 500 times. Collect all the $\overline X_n$'s and sort them (from the smallest to the greatest) and plot the sorted $\overline X_n$ values.

Due on 04/1

Week 10/11/12

- 03/27: Textbook, Exercises: 10.1, 10.2
- Spring break
- 04/17:
  1. Let W_n be a random variable with mean μ and variance C/n^ν with $\nu>0$. Prove that W_n is consistent with μ.
  2. Let Y_n be the nth order statistic of a random sample of size n from uniform$(0,\theta)$. Prove that $\sqrt{Y_n}$ is consistent with $\sqrt{\theta}$. Can you use Theorem 1 on page 51 of the lecture notes?
  3. Let Y_n be the nth order statistic of a random sample of size n with continuous CDF $F(\cdot)$. Define $Z_n=n[1-F(Y_n)]$. Find the limiting distribution of Z_n. That is, is Z_n convergent to some random

variable, in what mode?
4. In the question above, let F be the CDF for standard normal. Let n be a large number. Then numerically verify you claim of the limiting distribution above by comparing $P(Z_n\le t)$ with $P(Z\le t)$ for arbitrary t where Z is the limiting random variable of Z_n.
5. In general, $X_n\Rightarrow X$ and $Y_n\Rightarrow Y$ cannot imply $X_n+Y_n\Rightarrow X+Y$. Please give a counterexample to illustrate this. The symbol $\Rightarrow$ means convergence in distribution.

- 04/20: Textbook,
  1. Exercises: 10.4, 10.5, 10.6.
  2. For $X\sim bin(n,p)$, let $\tau(p)=1/(1-p)$. What can we say about $\hat{\tau}$ for $p\ne 1$?
  3. For $X_1,\dots,X_n\sim Unif(0,\theta)$, find the MLE of θ. Find an unbiased estimator based which is a function of the MLE. Calculate the variance of this unbiased estimator. Calculate the theoretical optimal variance due to the CRLB. Compare them.

Due on 04/22

Week 13

- 4/22: Textbook. Exercises: 10.8, 10.19(a), 10.35.
- 4/24: Textbook. Exercises: 10.31, 10.32, 10.33, 10.34, 10.36, 10.37
- 4/27: In exercise 10.36, you were asked to derive two Wald statistics to run approximate large sample test. Now let $n=25$, $\alpha=1$, $H_0:\beta=\beta_0=2$. Please numerically compare the power of these two test when the true value of β is 3, by running the test on the data for 10,000 times, and see which one rejects the null hypothesis more often. Try to interpret the result.

I am not satisfied with some of your answers to 9.23 in the homework returned today. I am giving a second chance for those who lost points for 9.23. You may submit your new answers (especially the numerical answers) along with this homework. I will consider adding back some points to that homework assignment. Please indicate that how many points you lost for 9.23. For the numerical answer, I have provided a Monte Carlo method to calculate the p value in the solution. You should use some other approach. For example, you can calculate the p value by taking the sum of the probabilities of x which satisfies $LR(x)<LR(x_0)$ for $x=0,1,2,\dots,10000$ (instead of ∞) to approximate the p value, where x_0 is the observed data. This is just one suggestion and there are other approaches.

Due on 05/01

Week 14

- 4/29:
  - Textbook. Exercises: 10.38.
  - Suppose that a random variable X has a Poisson distribution for which the mean θ is unknown. Find the Fisher information $I(\theta)$ in X.
  - Suppose $X_1,\dots,X_n\sim Pois(\theta)$. Find the large sample Z test, score test and LRT for testing $H_0:\theta=2$ vs $H_a:\theta\neq 2$.
  - Simulate the distribution of $-2\log(\lambda_n)$ using the empirical distribution function (EDF) and compare it with the CDF of $\chi^2(1)$ distribution. You may revise the following code shown in the class to draw the EDF

and CDF. Simulate a large number of data samples (say 5000), where each sample has size n. Make the case for $n=5$ and $n=100$.

- 5/1:
  - Read Example 10.4.5 and finish exercise 10.40; finish exercise 10.41, 10.47 and 10.48.
  - As in Example 10.3.4, with $\mathbf{X}\sim \textrm{Multinomial}(n,p_1,\ldots,p_5)$. Compare $H_0: p_1=p_2=p_5=0.01, p_3=0.5$ v.s. H_1: H_0 is not true.
    1. Derive the likelihood ratio test for $n=1$ and $n=100$ with level $\alpha=0.05$.
    2. Give an estimate of $P(H_o|H_1)$ when $p_1=p_2=p_5$, $p_3=0.3$, $n=100$, using simulation. Note that this is the probability of making type II error. Present the program.
    3. Compute $P(H_o|H_1)$ when $p_1=p_2=p_5$, $p_3=0.3$, $n=1$.
    4. Remark 1: in computing, sometimes it is better to use $\log(0^0)$ instead of $0*\log(0)$ as the latter can cause numerical trouble.
    5. Remark 2: What is the difference in degrees of freedom? Think how many additional constraints are imposed.
    6. Remark 3: You can try several combinations of p_k's that satisfy $p_1=p_2=p_5$, $p_3=0.3$.

R code notes pp. 60, *fig10.r*

```
myfun=function(n){
    m=1000
    x=rgamma(m,n,1)/n # m X's
    y=-2*(n*log(x)+n*(1-x)) # m λ's
    u=rchisq(m,1)

    qqplot(y,u,main=paste("QQ plot, n=",n))
    lines(y,y)

    sy=sort(y)
    plot(sy,ppoints(sy), xlim=c(0.5,2), ylim=c(0.4,0.9), type="l", lty=1,
main=paste("CDF, n=",n))
    lines(sy,pchisq(sy,1), xlim=c(0.5,2), ylim=c(0.4,0.9), type="l", lty=2)
}

pdf("fig10.pdf",height=9.0, width=6.5)
par(mfrow=c(2,2))
n=1
myfun(n)
n=100
myfun(n)
dev.off()
```

Due on 05/06