STAT 346/446 Review for Final Exam

Material

- Lectures 1 - 12
  - Skip slides 9-18 of Lecture 8 and slides 21-24 of Lecture 11
- Chapters 5, 6, 7, 8, and 10
  - Chapter 5: Section 5.5 (Convergence, CLT, Slutsky, delta method...)
  - Chapter 6: Sections 6.1 and 6.2
  - Chapter 7: All, except 7.2.4 (EM algorithm)
  - Chapter 8: All, except 8.3.3 and 8.3.4 (Sizes of UIT/IUT and p-values)
  - Chapter 10: 10.1.1

Bring **2** cheat sheets and a calculator

Lecture 1 - Review

Sections 5.1-5.4

- Random samples and statistics
- The χ^2_ν, t_p, and $F_{p,q}$ distributions and their relationship to the normal distribution
- Make sure you know the facts on slides 15 - 18
- Order Statistics
  - Distribution of order statistics
  - Formula for the pdf of order statistics from a continuous distribution
  - $X_{(1)}$ and $X_{(n)}$ most used

Lecture 2 - Convergence concepts 1/2

Sections 5.5.1 - 5.5.3

- Convergence in probability and in distribution
- Understand difference
  - $X_n \xrightarrow{p} X$: $\quad \lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0$
  - $X_n \xrightarrow{D} X$: $\quad \lim_{n\to\infty} F_{X_n}(x) = F_X(x)\ \forall x$ where $F(x)$ is continuous
- Connections and results
  - $X_n \xrightarrow{p} X \quad \Rightarrow \quad X_n \xrightarrow{D} X$
  - $X_n \xrightarrow{D} a \quad \Rightarrow \quad X_n \xrightarrow{p} a$ if a is a constant
- $\xrightarrow{p}$ and continuous functions:

$$X_n \xrightarrow{p} X \quad \Rightarrow \quad h(X_n) \xrightarrow{p} h(X)$$

Lecture 2 - Convergence concepts 2/2

Sections 5.5.1 - 5.5.3

- Weak law of large numbers (WLLN):

$$\overline{X}_n \xrightarrow{p} \mu$$

- STAT 446: Convergence almost surely and SLLN:

$$\overline{X}_n \xrightarrow{\text{as}} \mu$$

  - $X_n \xrightarrow{\text{as}} X$: $\quad P(\lim_{n\to\infty} |X_n - X| < \epsilon) = 1$

Lecture 3 - Convergence theorems 1/2

Sections 5.5.3 and 5.5.4

- Slutsky's Theorem: If $X_n \xrightarrow{D} X$ and $Y_n \xrightarrow{p} a$ then
  - $X_n Y_n \xrightarrow{D} aX$
  - $X_n + Y_n \xrightarrow{D} X + a$
- Central Limit theorem
$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{D} N(0, \sigma^2) \qquad \text{as } n \to \infty$$
- Delta Method
$$\sqrt{n}(g(\overline{X}_n) - g(\mu)) \xrightarrow{D} N(0, \sigma^2 g'(\mu)^2) \qquad \text{as } n \to \infty$$
- STAT 446: Second order Delta Method
$$\sqrt{n}(g(\overline{X}_n) - g(\mu)) \xrightarrow{D} \sigma^2 \frac{g''(\mu)}{2} \chi^2_1 \qquad \text{as } n \to \infty$$

Lecture 3 - Convergence theorems 2/2

Sections 5.5.3 and 5.5.4

- Approaching asymptotic problems:
  - Remember that the CLT only applies to the sample mean $\overline{X}_n$.

$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{D} N(0, \sigma^2) \qquad \text{as } n \to \infty$$

  - If instead of $\overline{X}_n$ you have some $g(\overline{X}_n)$ think: delta method

$$\sqrt{n}(g(\overline{X}_n) - g(\mu)) \xrightarrow{D} N(0, \sigma^2 g'(\mu)^2) \qquad \text{as } n \to \infty$$

  - If you need further massaging of the left hand side think:
    - Slutsky
    - WLLN
    - $\xrightarrow{p}$ and continuous functions
- See e.g. problem 2 (c) on the midterm (346 and 446).

Lecture 4: Point estimators

Sections 7.1 and 7.2.1-7.2.3

- Method of Moments (MOM), Maximum Likelihood Estimators (MLE)
- Invariance of the MLE
  - If $\hat{\theta}$ is the MLE of θ then $\tau(\hat{\theta})$ is the MLE of $\tau(\theta)$
- Bayesian estimation
  - Bayes Estimator = Posterior mean
  - How to find the posterior distribution (proportionality argument):

$$\pi(\theta \mid \mathbf{x}) = \frac{f(\mathbf{x} \mid \theta)\pi(\theta)}{m(\mathbf{x})} \quad \propto f(\mathbf{x} \mid \theta)\pi(\theta)$$

If $f(\mathbf{x} \mid \theta)\pi(\theta)$ is recognizable as the kernel of a known distribution (remember that now θ is the random variable, not $\mathbf{X}$) then that is the posterior distribution

Lecture 5: Evaluation of point estimators 1/2

Sections 7.3.1-7.3.2

- Unbiased: $E(W_n) = \theta$
  - Bias: $\mathrm{Bias}(W_n) = E(W_n) - \theta$
- Mean Square Error (MSE):

$$\mathrm{MSE}(W_n) = E((W_n - \theta)^2) = \mathrm{Var}(W_n) + \mathrm{Bias}(W_n)^2$$

- Best Unbiased estimator
  - = Uniformly minimum variance unbiased estimator (UMVUE)
- Cramer-Rao lower bound (Fisher Information)
  - How to calculate (shortcuts for iid and for exponential families)
  - How to use to show that an estimator is UMVUE

- Cramer-Rao lower bound for $\mathrm{Var}(W(\mathbf{X}))$:

$$\text{CRLB:} \quad \frac{\left(\frac{d}{d\theta}E(W(\mathbf{X}))\right)^2}{E\left(\left(\frac{d}{d\theta}\log(f(\mathbf{X}\mid\theta))\right)^2\right)}$$

- Fisher Information - shortcuts:

$$\begin{aligned} I(\theta) &= E\left(\left(\frac{d}{d\theta}\log(f(\mathbf{X}\mid\theta))\right)^2\right) \\ &= nE\left(\left(\frac{d}{d\theta}\log(f(X\mid\theta))\right)^2\right) \quad \text{if iid} \\ &= -nE\left(\frac{d^2}{d\theta^2}\log(f(X\mid\theta))\right) \text{ if exponential family} \end{aligned}$$

- Sufficient Statistics - how to identify
  - Factorization Theorem
  - Exponential Family
- Complete Statistic - how to identify
  - Exponential Family
- For exponential families:

$$f(x \mid \boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta}) \exp\left(\sum_{j=1}^{k} w_j(\boldsymbol{\theta}) t_j(x)\right)$$

The statistic $$T(\mathbf{X}) = \left(\sum_{i=1}^{n} t_1(X_i), \sum_{i=1}^{n} t_2(X_i), \ldots, \sum_{i=1}^{n} t_k(X_i)\right)$$

is sufficient and complete (for full exponential families)

Lecture 7 - Sufficiency and Unbiasedness 1/2

Section 7.3.3

- Lehmann - Scheffé
  - A complete sufficient statistic is the UMVUE of its expected value
  - Another way to prove that we have a UMVUE (i.e. other than Cramer-Rao)
- Rao-Blackwell
  - Conditioning on a sufficient statistic always gives a better estimator (lower variance) with the same expected value.

Lecture 7 - Sufficiency and Unbiasedness 2/2

- A way to construct a UMVUE of $\tau(\theta)$
- Rao-Blackwell + Lehmann-Scheffé:
  - Find a complete statistic T. If unbiased of $\tau(\theta)$, then we are done
  - If not unbiased, find a simple unbiased estimator W and set

    $$\phi(T) = E(W \mid T)$$

    then $\phi(T)$ is unbiased and based only on a complete sufficient statistic. Therefore $\phi(T)$ is the best unbiased estimator of $\tau(\theta)$
    - Often the simple unbiased estimator W is based on only one observation, e.g. X_1
- See for example Exercise 7.57 (HW 4)

Lecture 8 - Asymptotic results for point estimation

Section 10.1.1

- Consistent estimator if $W_n \xrightarrow{P} \theta$
  - $Var(W_n) \to 0$ and $\mathrm{bias}(W_n) \to 0$
  - Most MLEs

Lecture 9 - Hypothesis Testing methods 1/2

Sections 8.1 and 8.2

- Likelihood ratio tests (LRT): Reject if

$$\lambda(\mathbf{x}) = \frac{\sup_{\theta\in\Theta_0} L(\theta \mid \mathbf{x})}{\sup_{\theta\in\Theta} L(\theta \mid \mathbf{x})} = \frac{\sup_{\theta\in\Theta_0} L(\theta \mid \mathbf{x})}{L(\hat{\theta} \mid \mathbf{x})} \leq c \qquad 0 \leq c \leq 1$$

  - Only depend on sufficient statistics: Reject if

$$\lambda(t) = \frac{\sup_{\theta\in\Theta_0} L(\theta \mid t)}{\sup_{\theta\in\Theta} L(\theta \mid t)} \leq c \qquad 0 \leq c \leq 1$$

  - Constrained optimization is easier if we can show that the likelihood function is a monotone or unimodal function of θ
  - Simplifying the decision rule is often easier if we can show that $\lambda(t)$ is a monotone or unimodal function of t

Lecture 9 - Hypothesis Testing methods 2/2

- Union-Intersection Test (UIT) and Intersection-Union Test (IUT)
  - $H_0 : \bigcap_k \Theta_k \quad \Rightarrow \quad$ rejection region: $\bigcup_k R_k$
  - $H_0 : \bigcup_k \Theta_k \quad \Rightarrow \quad$ rejection region: $\bigcap_k R_k$
- Bayesian Test: Reject if

$$P(\theta \in \Theta_0^c \mid \mathbf{X}) > \frac{1}{2}$$

Have to find the posterior distribution of θ

Lecture 10 - Evaluation of Hypothesis Tests 1/2

Section 8.3.1

- Power function

$$\beta(\theta) = P_\theta(\text{ reject } H_0) = P_\theta(\mathbf{X} \in R)$$

- Type I and Type II errors:
  - Type I: Reject H_0 when it is true
  - Type II: Don't reject H_0 when it is false
- Power function and types of error:
  - $\beta(\theta) =$ probability of type I error for $\theta \in \Theta_0$
  - $1 - \beta(\theta) =$ probability of type II error for $\theta \in \Theta_0^c$

Lecture 10 - Evaluation of Hypothesis Tests 2/2

- Size of a test:

$$\alpha = \sup_{\theta \in \Theta_0} \beta(\theta)$$

  Largest probability of type I error
    - The constrained maximization is easier if we can show that $\beta(\theta)$ is a monotone or unimodal function of θ

- A test is a level α test if

$$\sup_{\theta \in \Theta_0} \beta(\theta) \leq \alpha$$

- A size α test is also a level α test
    - A size 0.01 test is also a level 0.01 test
    - A size 0.01 test is also a level 0.05 test

- LRT: the constant *c* determines the size of the test

Lecture 11 - Most powerful tests 1/2

Section 8.3.2

- Uniformly Most Powerful (UMP) tests

$$\beta_\delta(\theta) \geq \beta_{\delta^*}(\theta) \qquad \forall \theta \in \Theta_0^c$$

- Neyman-Pearson Lemma for simple hypotheses
  $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1$
  UMP level α test rejects if

$$\frac{g(t \mid \theta_1)}{g(t \mid \theta_0)} > k \quad k \geq 0$$

  where $\alpha = \beta(\theta_0) = P_{\theta_0}(\text{ reject } H_0)$

  - Only depends on sufficient statistics
  - The k determines α (or vice versa)

Lecture 11 - Most powerful tests 2/2

- Monotone likelihood ratio (MLR) families: For every $\theta_2 > \theta_1$

  $$\mathrm{LR}(t) = \frac{g(t \mid \theta_2)}{g(t \mid \theta_1)}$$

  is a monotone function of t

- MLR + sufficient statistic T: Karlin-Rubin Theorem
  - $H_0 : \theta \leq \theta_0$ vs. $H_1 : \theta > \theta_0$
    - UMP level α test rejects if $T > t_0$
    - where $\alpha = \beta(\theta_0) = P_{\theta_0}(\text{ reject } H_0) = P_{\theta_0}(T > t_0)$
  - $H_0 : \theta \geq \theta_0$ vs. $H_1 : \theta < \theta_0$
    - UMP level α test rejects if $T < t_0$
    - where $\alpha = \beta(\theta_0) = P_{\theta_0}(\text{ reject } H_0) = P_{\theta_0}(T < t_0)$
  - The t_0 determines α (or vice versa)

Lecture 12 - Decision Theory

Sections 7.3.4 and 8.3.5

- Action space $\mathcal{A}$ and action rule $a = \delta(\mathbf{x}) : \mathcal{X} \to \mathcal{A}$
- Loss function: $L(\theta, a) : \Theta \times \mathcal{A} \to \mathbb{R}^+$
- Risk function: $R(\theta, \delta) = E_\theta(L(\theta, \delta(\mathbf{X})))$
- Point estimation as a decision problem
  - Minimizing risk (over all θ) under a squared error (or absolute error) loss
  - Bayes risk: $\int_\Theta R(\theta, \delta)\pi(\theta) d\theta$ and Bayes rule
- Hypothesis testing as a decision problem
  - (Generalized) 0-1 loss
  - Risk functions and relationship to power function