Study Guide – Math 445, Final Exam

Prof. Adam Loy

1 General Information

- The final will be on Monday 6/6 from 3 to 5:30 pm.
- This exam is closed book and closed notes, except for one standard size sheet of paper (8.5" by 11") which can have notes on both sides.
- No copying, cheating, collaborations, computers, or cell phones are allowed.
- Show your work and write complete and coherent answers. Partial credit will not be given for unsupported or incoherent answers.
- Illegible or unjustified solutions will not receive full credit.
- To study, I recommend carefully going through class notes, homework problems, and this handout actively (intermixing reading, thinking, solving problems, and asking questions). After reviewing those materials I recommend solving lots and lots practice problems.

2 Topics

Material before the midterm

- Statistical graphics: quantile-quantile plot, density curve, histogram, box plot
- Hypothesis testing: null hypothesis, alternative hypothesis, permutation tests, χ^2 test of independence, test of homogeneity, goodness-of-fit tests, p-value, assumptions for each test
- Bootstrap CIs: one-sample bootstrap, two-sample bootstrap, bootstrap standard error, the plug-in principle, bootstrap percentile CIs, bootstrap bias
- Point estimation: method of moments estimators, maximum likelihood estimators, estimate vs. estimator, bias, mean square error, consistency, asymptotically unbiased, likelihood, log-likelihood, efficiency, relative efficiently, Cramér-Rao lower bound, Fisher information, minimum variance unbiased estimator (MVUE)
- Sampling distributions: Central Limit Theorem, χ^2 distribution

Material after the midterm

- Sampling distributions: χ^2 distribution, t distribution, Central limit theorem
- Interval estimation: pivotal quantity, interpretations, confidence level, normal-based confidence intervals, one-sample t intervals, two-sample t intervals, pooled sample variance, Welch-Satterthwaite approximate degrees of freedom, assumptions for each procedure
- Hypothesis testing: null hypothesis, alternative hypothesis, test statistic, p-value, type I error, type II error, power, one-sample t-test, pooled t-test, two-sample t-test (Welch's t-test), parallel between a test and an interval, assumptions for each procedure

- Likelihood based testing: likelihood ratio tests, asymptotic likelihood ratio test, the Neyman-Pearson Lemma
- Bayesian inference: Bayes' rule, prior distribution, posterior distribution, likelihood, conjugacy, credible intervals, Monte Carlo simulation

3 Important Distributions

Name	Param.	PMF or PDF	Mean	Variance	MGF
Bernoulli	p	$P(X=1)=p$ $P(X=0)=q$	p	pq	pe^t+q
Binomial	n,p	$\binom{n}{k}p^k q^{n-k}$ $k\in\{0,1,\dots,n\}$	np	npq	$(pe^t+1-p)^n$
Geometric	p	pq^k $k\in\{0,1,2,\dots\}$	$\frac{q}{p}$	$\frac{q}{p^2}$	$\frac{p}{1-qe^t}$
Negative Binomial	r,p	$\binom{r+n-1}{r-1}p^r q^n$ $k\in\{0,1,2,\dots\}$	$\frac{rq}{p}$	$\frac{rq}{p^2}$	$\left(\frac{p}{1-qe^t}\right)^r$
Hypergeometric	w,b,n	$\frac{\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}}$ $k\in\{0,1,\dots,n\}$	$\mu=\frac{nw}{w+b}$	$\left(\frac{w+b-n}{w+b-1}\right)n\frac{\mu}{n}\left(1-\frac{\mu}{n}\right)$	
Poisson	λ	$\frac{e^{-\lambda}\lambda^k}{k!}$ $k\in\{0,1,2,\dots\}$	λ	λ	$e^{\lambda(e^t-1)}$
Uniform	$a<b$	$\frac{1}{b-a},\ x\in(a,b)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{tb}-e^{ta}}{t(b-a)}$
Normal	μ,σ^2	$\frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/(2\sigma^2)}$ $x\in\mathbb{R}$	μ	σ^2	$e^{\mu t+(\sigma^2t^2)/2}$
Exponential	λ	$\lambda e^{-\lambda x},\ x>0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{\lambda}{\lambda-t}$
Gamma	a,λ	$\frac{\lambda^a}{\Gamma(a)}x^{a-1}e^{-\lambda x}$ $x>0$	$\frac{a}{\lambda}$	$\frac{a}{\lambda^2}$	$\left(\frac{\lambda}{\lambda-t}\right)^a$
Beta	a,b	$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}$ $0<x<1$	$\mu=\frac{a}{a+b}$	$\frac{\mu(1-\mu)}{a+b+1}$	$1+\sum_{k=1}^{\infty}\left(\prod_{r=0}^{k-1}\frac{a+r}{a+b+r}\right)\frac{t^k}{k!}$
Chi-Square	n	$\frac{1}{2^{n/2}\Gamma(n/2)}x^{n/2-1}e^{-x/2}$ $x>0$	n	$2n$	$\left(\frac{1}{1-2t}\right)^{n/2}$

4 Strategic Practice Problems

The below problems are intended to give you an idea about the format of questions that I might asked, they are not intended to cover all of the material. To fully study for the exam be sure to work on these problems, review your notes, and your homework assignments.

1. Consider the construction of a confidence interval for a population mean based on a sample of size 6 (you may assume that the data comes from a population that is roughly Normal). The "typical" small sample 95% confidence interval has the form $\left(\overline{x} - 2.571\frac{s}{\sqrt{n}},\ \overline{x} + 2.571\frac{s}{\sqrt{n}}\right)$; the ± 2.571 are chosen because 95% of the t_5 distribution is between those values. Another valid 95% confidence interval would be $\left(\overline{x} - 2.191\frac{s}{\sqrt{n}},\ \overline{x} + 3.365\frac{s}{\sqrt{n}}\right)$ because 95% of the t_5 distribution is between -2.191 and 3.365. Is one confidence interval better than the other? If so, which one? Justify your answer.

2. Let $X_1, \ldots, X_n$ be independent random variables with $X_i \sim \text{Poisson}(i\theta)$, $\theta > 0$, for $i = 1, \ldots n$.

   (a) Write down the joint probability mass function of $X_1, \ldots, X_n$.

   (b) Derive the likelihood ratio test statistic.

   (c) Suppose that $n = 5$. Specify the rejection region used to test $H_0 : \theta \geq 1/5$ vs. $H_a : \theta < 1/5$ at the $\alpha = e^{-3}$ significance level. To receive full credit, you must determine the numerical values of any constants in the test.

3. Let $X_1, X_2, \ldots, X_n$ be a random sample from the Poisson distribution with PDF

$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!},\ x = 0, 1, 2, \ldots$$

   (a) Write down the likelihood function of λ.

   (b) Suppose that you decide to use a Gamma(a, b) prior distribution for λ with PDF

$$p(\theta) = \frac{b^a}{\Gamma(a)}\lambda^{a-1}e^{-b\lambda},\ \lambda > 0.$$

   Find the posterior density of λ.

   (c) Is the gamma prior a conjugate family to the Poisson likelihood?

   (d) Describe how you would create a 95% credible interval for λ.

4. Let $X_1, \ldots, X_n$ be i.i.d. Exponential random variables with parameter λ, and suppose we would like to test $H_0 : \lambda = \lambda_0$ vs. $H_a : \lambda > \lambda_0$.

   (a) Find the likelihood ratio statistic.

   (b) Clearly define the rejection region for this test.

5. Below is the output from R's optim command based on two models:

   Model A: $Y_i \sim \text{LogNormal}(\mu_i = \beta_0,\ \sigma^2)$
   Model B: $Y_i \sim \text{LogNormal}(\mu_i = \beta_0 + \beta_1 X_i,\ \sigma^2)$

   for some known X_i for $n = 81$ observations. Here β_0, β_1, and σ^2 are all unknown parameters. The variables are:

   $Y =$ amount spent on last haircut plus one dollar (so that there are no zeroes in the data set)
   $X =$ a binary variable: 1 for females, 0 for males

Below is the relevant R output:

```
> modela <- optim(par = c(mean(log(y)), sd(log(y))), fn = lognorm.loglik.null,
                  x = x, y = y, control = list(fnscale=-1))
> modelb <- optim(par = c(modela$par[1], 0, modela$par[2]), fn = lognorm.loglik,
                  x = x, y = y,control = list(fnscale=-1))
```

```
> modela
$par
[1] 3.1521691 0.9905972

$value
[1] -374.4318

$counts
function gradient
      43       NA

$convergence
[1] 0

$message
NULL
```

```
> modelb
$par
[1] 2.8937810 0.7059336 0.8744568

$value
[1] -369.3333

$counts
function gradient
      74       NA

$convergence
[1] 0

$message
NULL
```

(a) Why is it important that there are no zeroes in the data set?

(b) We'd like to test $H_0 : \beta_1 = 0$ vs. $H_a : \beta_1 \neq 0$. Perform a likelihood ratio test to test the above hypotheses.

6. A study evaluated an experimental genetic treatment thought to increase litter size (the number of offspring) of sows (female pigs). Thirty sows were chosen from a pig research facility, and 10 were randomly assigned to receive the genetic treatment; the remaining 20 served as controls. The size of each sow's first litter was recorded. The data are summarized below.

	Control	Experiment
average	10.3	10.7
s.d.	2.4	1.5
n	20	10

(a) Which average litter size (control or experimental group) is more precisely known? Make as few assumptions as possible for this part. Explain.

(b) Use a two-tailed t-test to test the null hypothesis of no effect of the experimental genetic treatment on the mean litter size. Please report your test statistic, an approximate p-value, and a one-sentence conclusion.

(c) List the assumptions underlying the t-test used in the previous question. For each assumption, list one reasonable method of assessing the validity of that assumption.

7. Suppose that $X_1, X_2, \ldots, X_n$ form a random sample from an exponential distribution with mean $1/\lambda$. Last term we learned that the sum of i.i.d. exponential(λ) random variables follows a $\mathrm{Gamma}(n, \lambda)$ distribution. Based on this fact, you may assume that

$$\frac{\lambda}{2}\sum_{i=1}^{n} X_i \sim \mathrm{Gamma}\left(n, \frac{1}{2}\right).$$

Explain why the above quantity is a pivotal quantity, and derive a formula for a $(1-\alpha)100\%$ confidence interval for λ.

8. Chapter 8, exercise 23