PhD Qualifying Examination–Part I

Department of Statistics
University of South Carolina
May 24, 2021 - 9:00AM–1:00PM

READ FIRST THESE INSTRUCTIONS

1. DO NOT write your name on any of your answer sheets. Instead, write your pre-assigned codename.
2. There are four (4) problems on this examination.
3. Formulas relating to distributions potentially relevant to the problems are provide in the last page.
4. You are **not allowed** to use search engines during the examination. Please adhere to the HONOR CODE in this instance. Any violation of the HONOR CODE (such as using search engines) will lead to a zero for the exam.
5. You have four hours for this examination. All four problems will be graded and are of equal weight.

The Problems

1. Let $X_1, \ldots, X_n$, where $n \geq 2$, be independent and identically distributed (i.i.d.) random variables following $\text{Exp}(\lambda)$, that is, an exponential distribution with mean equal to λ. Let $\mathbf{X} = (X_1, \ldots, X_n)$. Also, let $h(\lambda) = e^{-1/\lambda}$, which is the probability that a univariate $\text{Exp}(\lambda)$ random variable exceeds 1.

   (a) *Derive* the distribution of $W = X_2 + X_3 + \cdots + X_n$.

   (b) *Derive* the joint probability density function (pdf) of X_1 and $X_1 + W$.

   (c) *Derive* the distribution of the random variable $U = \dfrac{X_1}{\sum_{i=1}^n X_i}$.

   (d) Find the maximum likelihood estimator (MLE) of $h(\lambda)$, call it $\widehat{h}(\mathbf{X})$.

   (e) What is the limiting distribution of $\sqrt{n}\left\{\widehat{h}(\mathbf{X}) - h(\lambda)\right\}$ as $n \to \infty$?

   (f) Either *prove* that $\widehat{h}(\mathbf{X})$ is the uniform minimum variance unbiased estimator (UMVUE) of $h(\lambda)$ or *find* the UMVUE of $h(\lambda)$. If you are constructing an alternative unbiased estimator for $h(\lambda)$ besides $\widehat{h}(\mathbf{X})$, please derive a closed form expression that is as simplified as possible.

2. Two players, Alan and Bill, are playing a game consisting of a series of trials, each of which is won by either Alan or Bill. Assume that trials are independent and the probability θ (with $0 < \theta < 1$) that Alan wins a trial is constant across all trials. The winner of the overall game is the first player to win 10 trials, but he MUST win with a margin of at least two. So the game cannot be won by a score of 10-9, but possible winning scores are 11-9, 12-10, 13-11, etc.

   Now, suppose that Bill is leading the game, 8 trials to 3. At this point in time, we wish to estimate the probability p_A that Alan wins the overall game.

   (a) Calculate p_A and write it as a function of θ. Carefully show or explain your calculation.

   (b) Find the maximum likelihood estimator of p_A. You may appeal to well-established results in your answer.

   (c) There are several common methods to get a confidence interval for a binomial probability. Below is shown R output that lists 95% confidence intervals for θ in this problem, arising from five different methods. In addition, Figures 1–3 provide the five plots displaying empirical coverage probabilities and average interval widths of these five methods, based on simulations, in the case of $n = 11$ trials.

```
# Wilson score CI, with Yates continuity correction
> prop.test(x=3,n=11)$conf.int
[1] 0.07327666 0.60683390

# Wilson score CI, no continuity correction
> prop.test(x=3,n=11,correct=F)$conf.int
[1] 0.09746059 0.56564530

# Clopper-Pearson (Exact) CI
> binom.test(x=3,n=11)$conf.int
[1] 0.06021773 0.60974256

# Wald CI
> waldInterval(x=3,n=11)
[1] 0.009540121 0.535914424

# Agresti-Coull CI
> waldInterval(x=3+2,n=11+4)
[1] 0.09477411 0.57189255
```

   What are your conclusions about the quality of these five methods? Which are best? What, specifically, are the weaknesses of the other methods?

   (d) Give a 95% confidence interval for p_A. Briefly explain why your interval is a valid (at least approximately) 95% confidence interval for p_A.

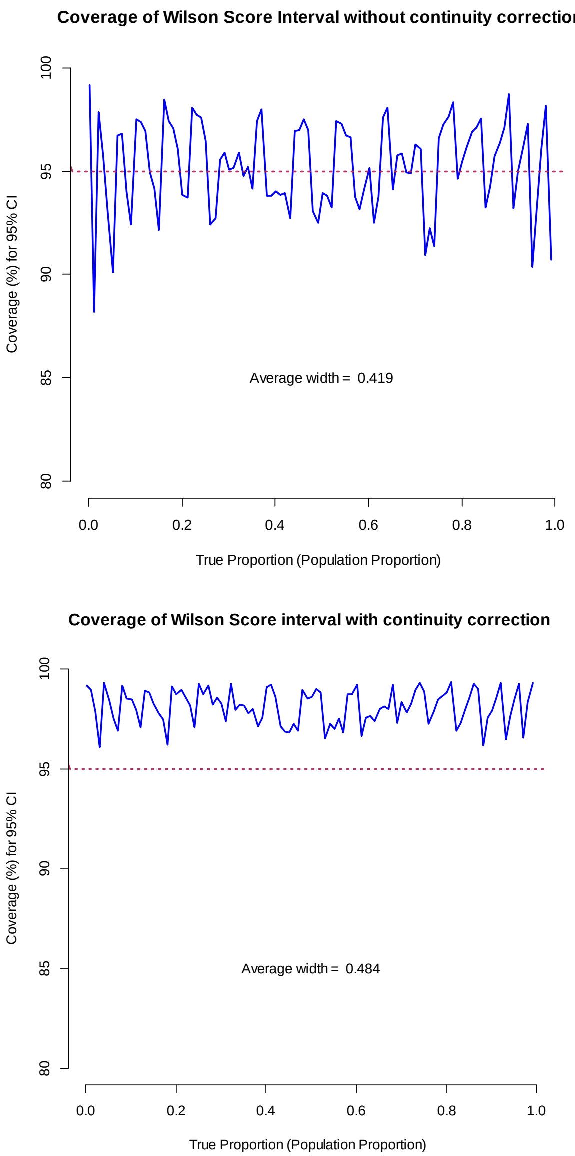

Figure 1: Empirical coverage probabilities and average interval width of Wilson score interval estimates, without continuity correction (in the top panel) and with continuity correction (in the bottom panel).

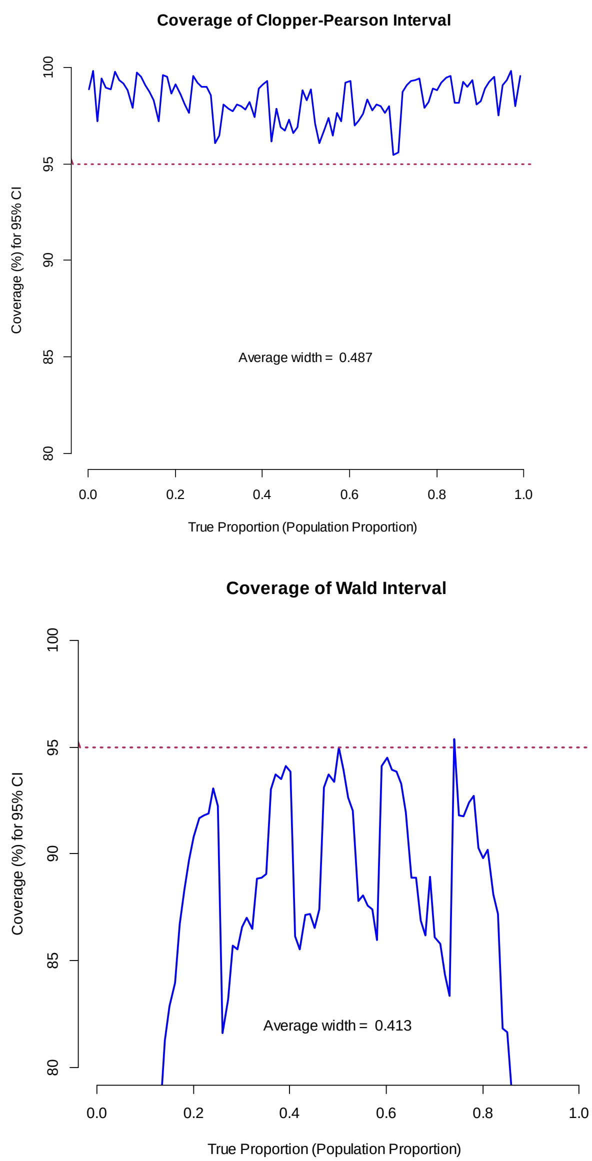

Figure 2: Empirical coverage probabilities and average interval width of Clopper-Pearson interval estimates (in the top panel) and those of Wald interval estimates (in the bottom panel).

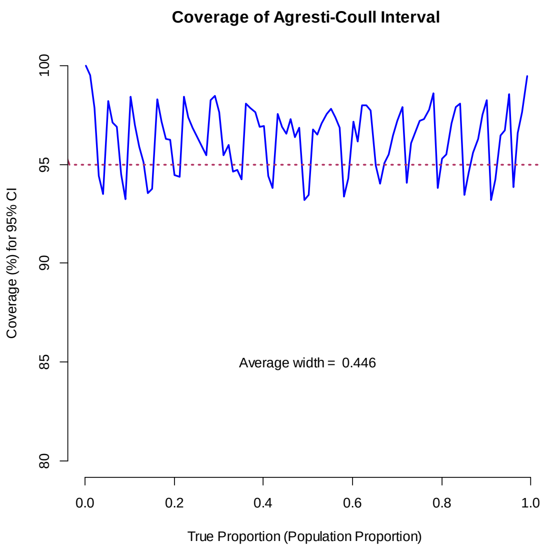

Figure 3: Empirical coverage probabilities and average interval width of Agresti-Coull interval estimates.

3. Suppose that $X_1, \ldots, X_n$ are independent and identically distributed (i.i.d.) random variables following Beta$(\gamma, 1)$ and that $Y_1, \ldots, Y_m$ are i.i.d. Beta$(\theta, 1)$. Let $\mathbf{X} = (X_1, \ldots, X_n)$ and $\mathbf{Y} = (Y_1, \ldots, Y_m)$. Assume further that $\mathbf{X}$ and $\mathbf{Y}$ are independent. We will consider testing $H_0 : \gamma = \theta$ versus $H_1 : \gamma \neq \theta$ using the statistic,

$$T = \frac{\sum_{i=1}^{n} \log X_i}{\sum_{i=1}^{n} \log X_i + \sum_{i=1}^{m} \log Y_i}.$$

(a) *Derive* the distribution of $Z = -\gamma \log X_1$.

(b) *Derive* the distribution of T under H_0. You may use properties about sums of certain i.i.d. random variables without proving these properties.

(c) Show that the maximum likelihood estimator (MLE) of γ is $\widehat{\gamma}(\mathbf{X})$, where

$$\widehat{\gamma}(\mathbf{X}) = \frac{-n}{\sum_{i=1}^{n} \log X_i}.$$

Obviously, the MLE of θ is $\widehat{\theta}(\mathbf{Y}) = -m / \sum_{i=1}^{m} \log Y_i$.

(d) We define $\widehat{\mu}_0(\mathbf{X}, \mathbf{Y})$ as follows:

$$\widehat{\mu}_0(\mathbf{X}, \mathbf{Y}) = \frac{-(n+m)}{\sum_{i=1}^{n} \log X_i + \sum_{i=1}^{m} \log Y_i}.$$

Use the fact that $\prod_{i=1}^{n} X_i = \exp(\sum_{i=1}^{n} \log X_i)$ to show that

$$\frac{\{(\prod_{i=1}^{n} X_i)(\prod_{i=1}^{m} Y_i)\}^{\widehat{\mu}_0(\mathbf{X},\mathbf{Y})-1}}{(\prod_{i=1}^{n} X_i)^{\widehat{\gamma}(\mathbf{X})-1} (\prod_{i=1}^{m} Y_i)^{\widehat{\theta}(\mathbf{Y})-1}} = 1.$$

(e) Construct the likelihood ratio test (LRT) statistic for testing $H_0 : \gamma = \theta$ against $H_1 : \gamma \neq \theta$ in terms of $\widehat{\mu}_0(\mathbf{X}, \mathbf{Y})$, $\widehat{\gamma}(\mathbf{X})$, and $\widehat{\theta}(\mathbf{Y})$.

(f) Using your answers in parts (c)-(e), show that the LRT statistic can be written in such a way that it involves the data only through the statistic T.

(g) i. The general LRT theory tells us to reject H_0 when the LRT statistic is small. *Describe* an equivalent rejection rule in terms of T. You do not need to derive any unknown constants in explicit form.

ii. Suppose that $n = 23$ and $m = 12$. Explain how you would find the rejection region of a size-0.10 test in terms of T. Note: Again, you do not need to to derive explicit cutoffs for the rejection region, but *explain clearly* how you could determine these cutoffs for your rejection region using the given sample sizes, this significance level, and numerical software.

4. Consider a random sample consisting of n bivariate observations, $(X_1, Y_1), \ldots, (X_n, Y_n)$. The random sample is from a population that follows the joint distribution of (X, Y), where the marginal distribution of X is Poisson(λ), and conditioning on $X = x$, Y follows binomial($x + 1$, p).

   (a) Provide the maximum likelihood estimator (MLE) of p, denoted by $\hat{p}_1$. You do not need to justify that the MLE you provide here does maximize the likelihood function by checking the second-order derivatives.

   (b) Does $\hat{p}_1$ converge almost surely to p as $n \to \infty$? Justify your answer.

   (c) Suppose that the prior distribution of p is Beta(α, β), where α and β are two pre-specified constants. Provide a Bayes estimator of p, denoted by $\hat{p}_2$. Does $\hat{p}_2$ converge in probability to p as $n \to \infty$? Justify your answer.

   (d) Find the covariance between X and Y. Provide the MLE of this covariance.

   (e) Derive the marginal distribution of Y. (Hint: $e^x = \sum_{k=0}^{\infty} x^k/k!$)

Formulas relating to some distributions

- Binomial(n, p)

pmf: $f(x|n,p) = \binom{n}{x} p^x (1-p)^{n-x}$, $x = 0, 1, 2, \ldots, n$; $0 \le p \le 1$

mgf: $M(t) = (pe^t + 1 - p)^n$

moments: $E(X) = np$, $\mathrm{Var}(X) = np(1-p)$

- Poisson(λ)

pmf: $f(x|\lambda) = \dfrac{e^{-\lambda}\lambda^x}{x!}$, $x = 0, 1, \ldots$; $\lambda \ge 0$

mgf: $M(t) = \exp\{\lambda(e^t - 1)\}$

moments: $E(X) = \lambda$, $\mathrm{Var}(X) = \lambda$

- Beta(α, β)

pdf: $f(x|\alpha,\beta) = \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$, $0 \le x \le 1$, $\alpha > 0$, $\beta > 0$

moments: $E(X) = \dfrac{\alpha}{\alpha+\beta}$, $\mathrm{Var}(X) = \dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

- Exponential(β)

pdf: $f(x|\beta) = \beta^{-1}e^{-x/\beta}$, $0 \le x < \infty$, $\beta > 0$

mgf: $M(t) = (1-\beta t)^{-1}$, for $t < 1/\beta$

moments: $E(X) = \beta$, $\mathrm{Var}(X) = \beta^2$

- Gamma(α, β)

pdf: $f(x|\alpha,\beta) = \dfrac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1}e^{-x/\beta}$, $0 \le x < \infty$, $\alpha, \beta > 0$

mgf: $M(t) = (1-\beta t)^{-\alpha}$, for $t < 1/\beta$

moments: $E(X) = \alpha\beta$, $\mathrm{Var}(X) = \alpha\beta^2$

notes: Gamma(1, β) is exponential(β). Gamma($p/2$, 2) is χ^2_p

- Normal(μ, σ^2)

pdf: $f(x|\mu,\sigma^2) = \dfrac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}$, $-\infty < x < \infty$, $-\infty < \mu < \infty$, $\sigma > 0$

mgf: $M(t) = e^{\mu t + \sigma^2 t^2/2}$

- t distribution with ν degrees of freedom.

pdf: $f(x|\nu) = \dfrac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \dfrac{1}{\sqrt{\nu\pi}} \left(1 + x^2/\nu\right)^{-(\nu+1)/2}$, $-\infty < x < \infty$, $\nu = 1, 2, \ldots$

moments: $E(X) = 0$, for $\nu > 1$; $\mathrm{Var}(X) = \nu/(\nu-2)$, for $\nu > 2$