# Homework 6: 2/27 - 3/10

STA 336 Due 11:59pm Monday, March 13 Name:

**Instructions:** Write-up complete solutions to the following problems and submit answers on Gradescope. Your solutions should be neatly-written, show all work and computations, include figures or graphs where appropriate, and include some written explanation of your method or process (enough that I can understand your reasoning without having to guess or make assumptions). A rubric for homework problems appears on the final page of this assignment.

- Unless otherwise noted, problem numbers are taken from the 4th edition of DeGroot and Schervish's *Probability and Statistics.*

# Monday 2/27

## Additional Problems

AP1. Suppose $T$ has the $t$ distribution with $m$ degrees for freedom. Show that $E[T] = 0$ and $\text{Var}(T) = \frac{m}{m-2}$.

*Hint: To evaluate* $E[T^2]$*, express* $T$ *as the quotient* $Z/\sqrt{Y/m}$ *where* $Z \sim N(0,1)$ *and* $Y \sim \chi^2(m)$ *are independent. Then use LOTUS to calculate* $E[Y^{-1}]$.

AP2. The $t$ distribution with 1 degree of freedom is also known as the **Cauchy** distribution, and has pdf

$$f(x) = \frac{1}{\pi(1+x^2)} \qquad x \in \mathbb{R}$$

(a) Verify that $f$ is indeed a valid PDF by showing $\int_{-\infty}^{\infty} f(x)\,dx = 1$. *Hint:* The derivative of the arctan($x$) function is $\frac{d}{dx}\arctan(x) = \frac{1}{1+x^2}$.

(b) Show that both the median and mode of the Cauchy distribution are 0.

(c) Show that if $X$ is Cauchy distributed, then the expected value of $X$ does not exist. *Hint:* Since

$$E[X] = \int_{-\infty}^{\infty} x f(x)\,dx$$

is an improper integral with both upper and lower bounds of infinity, you will need to show that both

$$\int_0^{\infty} xf(x)dx \qquad \text{and} \qquad \int_{-\infty}^{0} xf(x)dx$$

are divergent. Use $u$-substitution to assist with the integration.

AP3. In part (c) of the previous problem, we might be tempted to say that since the Cauchy distribution is symmetric around 0, then it's mean should be 0, since the integrals over the positive and negative parts of the real line "cancel" out. But if this were true, then by the law of large numbers, the sample mean $\bar{X}$ should be close to 0, for a sample of a large number of independent variables $X_1, \dots, X_n$ from the Cauchy distribution. We'll show empirically that this is not the case:

(a) Use the `rcauchy` function in R to simulate a sample of $n = 10^4$ Cauchy variables. Compute the mean of your sample. Then repeat two more times. Were any of your means "close" to 0? (Were any within 0.5 of 0?)

(b) Repeat part (a), but this time with a sample of $n = 10^5$ and then $10^6$ Cauchy variables. This time, were any of your means "close" to 0?

(c) Of course, even if a Cauchy distribution did have a mean of 0, its possible just due to random chance that a few randomly generated sample means could be far from 0. To investigate this possibility, we need to simulate many more samples. Simulate 1000 samples, each of size $10^5$, and compute the mean of each sample.

(d) Calculate...

i. The mean of your sample means.

ii. The variance of your sample means.

iii. The proportion of your sample means that were at least a distance of 0.5 from 0.

(e) Based on the results you calculated in the previous part, does it seem plausible that the mean of a Cauchy distribution is 0? Explain.

# Wednesday 3/8

Section 8.5: 1, 4

## Additional Problems

AP4. Suppose $\mathbf{X}$ is a sample from a certain population with parameter $\theta$, and suppose $A$ and $B$ with $A < B$ are statistics based on the sample. The **coverage rate** of the random interval $(A, B)$ is the probability that the interval contains $\theta$. In order for $(A, B)$ to be $\gamma$-level confidence interval, it must be the case that the coverage rate of $(A, B)$ is greater than or equal to $\gamma$. In this problem, we'll investigate coverage rates for a variety of random intervals.

(a) Suppose $X_1, \ldots, X_n$ are a sample from $N(\mu, \sigma)$. Use R to simulate 10000 samples, each of size 5, assuming that $\mu = 0$ and $\sigma = 1$.

For each sample, compute the sample mean and sample variance, and then use these values to construct the upper and lower bounds of the 95% confidence interval for $\mu$, using Theorem 8.5.1. *Note: While you as the simulator know the values of $\mu$ and $\sigma^2$, you should not use these values to calculate the sample mean, sample variance, or bounds of the confidence interval.*

(b) Estimate the coverage rate of this confidence interval procedure by computing the proportion of samples which produced a 95% confidence interval that contained the true population mean $\mu$.

(c) Suppose that instead of using the $t(4)$ distribution to compute quantiles for our confidence interval of $\mu$, we instead used the Normal distribution. This would correspond to incorrectly treating the observed sample standard deviation $s$ as the true value of the population standard deviation $\sigma$. As shown in Exercise 8.5.1, the confidence interval takes the form

$$\left(\bar{X} - \Phi^{-1}\left(\frac{1+\gamma}{2}\right)\frac{s}{\sqrt{n}}, \bar{X} + \Phi^{-1}\left(\frac{1+\gamma}{2}\right)\frac{s}{\sqrt{n}}\right)$$

where $\Phi^{-1}$ is the quantile function for the standard Normal distribution.

Repeat part (a), but this time, using the above confidence interval formula, rather than the formula from Theorem 8.5.1. In particular, use the sample standard deviation $s$ (rather than the true population standard deviation $\sigma$) to calculate your confidence intervals.

(d) Estimate the coverage rate of this confidence interval procedure by computing the proportion of samples which produced a 95% confidence interval that contained the true population mean $\mu$. How does the coverage rate compare to the coverage rate of the interval from parts (a) and (b)?

# Friday 3/10

Section 8.5: 6

## Additional Problems

AP5. Suppose $X_1, X_2, \ldots, X_n$ are an iid sample from $\text{Unif}(0, \theta)$. Let $Y_n = \max\{X_1, \ldots, X_n\}$ be the MLE of $\theta$.

(a) Find the PDF and quantile function for $Y_n/\theta$.

(b) Show that $Y_n/\theta$ is a pivotal quantity.

(c) Use $Y_n/\theta$ to construct an exact $\gamma$-coefficient confidence interval for $\theta$.

# Homework 6: 2/27 - 3/10

STA 336 Due 11:59pm Monday, March 13 Name:

## General Rubric

| Points | Criteria |
|---|---|
| 5 | The solution is correct **and** well-written. The author leaves no doubt as to why the solution is valid. |
| 4.5 | The solution is well-written, and is correct except for some minor arithmetic or calculation mistake. |
| 4 | The solution is technically correct, but author has omitted some key justification for why the solution is valid. Alternatively, the solution is well-written, but is missing a small, but essential component. |
| 3 | The solution is well-written, but either overlooks a significant component of the problem or makes a significant mistake. Alternatively, in a multi-part problem, a majority of the solutions are correct and well-written, but one part is missing or is significantly incorrect |
| 2 | The solution is either correct but not adequately written, or it is adequately written but overlooks a significant component of the problem or makes a significant mistake. |
| 1 | The solution is rudimentary, but contains some relevant ideas. Alternatively, the solution briefly indicates the correct answer, but provides no further justification |
| 0 | Either the solution is missing entirely, or the author makes no non-trivial progress toward a solution (i.e. just writes the statement of the problem and/or restates given information) |
| **Notes:** | For problems with multiple parts, the score represents a holistic review of the entire problem.<br><br>Additionally, half-points may be used if the solution falls between two point values above. |