STATS 200: Solutions to Homework 3*

1 The t_1 distribution.

(a) By definition of the t distribution, T has the same distribution as $\frac{X}{\sqrt{U}}$ where $X \sim \mathcal{N}(0,1)$, $U \sim \chi^2_1$, and X and U are independent. By definition of the χ^2 distribution, U has the same distribution as Y^2 where $Y \sim \mathcal{N}(0,1)$. Therefore T has the same distribution as $\frac{X}{\sqrt{Y^2}} = \frac{X}{|Y|}$.

Here are two different methods to show $\frac{X}{|Y|}$ has the same distribution as $\frac{X}{Y}$:

(1) Let $A = \frac{|X|}{|Y|}$, $B = \mathrm{sign}(X)$, $C = \frac{\mathrm{sign}(X)}{\mathrm{sign}(Y)}$. (Here $\mathrm{sign}(x) = 1$ if $x > 0$ and $\mathrm{sign}(x) = -1$ if $x < 0$.) Then $\frac{X}{|Y|} = AB$ and $\frac{X}{Y} = AC$. $|X|$ and $\mathrm{sign}(X)$ are independent, because $\mathrm{sign}(X) = \pm 1$ each with probability 1/2 independently of $|X|$. Similarly $|Y|$ and $\mathrm{sign}(Y)$ are independent, so $|X|$, $\mathrm{sign}(X)$, $|Y|$, and $\mathrm{sign}(Y)$ are all independent. Then A is independent of B, and A is also independent of C. But B and C have the same distribution (they are both ± 1 with probability 1/2), so therefore AB and AC have the same distribution.

(2) For any $t \in \mathbb{R}$,

$$\mathbb{P}\left[\frac{X}{|Y|} \le t\right] = \mathbb{P}[X \le t|Y|] = \mathbb{P}[X \le tY,\ Y > 0] + \mathbb{P}[X \le -tY,\ Y < 0],$$

and

$$\begin{aligned}\mathbb{P}\left[\frac{X}{Y} \le t\right] &= \mathbb{P}[X \le tY,\ Y > 0] + \mathbb{P}[X \ge tY,\ Y < 0] \\ &= \mathbb{P}[X \le tY,\ Y > 0] + \mathbb{P}[-X \le -tY,\ Y < 0].\end{aligned}$$

Since (X,Y) has the same distribution as $(-X,Y)$, the above implies

$$\mathbb{P}\left[\frac{X}{|Y|} \le t\right] = \mathbb{P}\left[\frac{X}{Y} \le t\right].$$

So the CDFs of $\frac{X}{|Y|}$ and $\frac{X}{Y}$ are the same.

*Edited from the solutions by Zhenpeng Zhou; thanks to Zhenpeng for sharing.

(b) Noting $f(x) = f(-x)$, we compute

$$\mathbb{E}[|T|] = \int_{-\infty}^{\infty} |x| f(x) \mathrm{d}x = 2\int_0^{\infty} x \frac{1}{\pi} \frac{1}{x^2+1} \mathrm{d}x$$
$$= \frac{1}{\pi} \int_0^{\infty} \frac{1}{x^2+1} \mathrm{d}(x^2+1) = \frac{1}{\pi} \ln\left(x^2+1\right)\Big|_0^{\infty} = \infty.$$

Also,

$$\mathbb{E}[T^2] = \int_{-\infty}^{\infty} x^2 f(x) \mathrm{d}x = 2\int_0^{\infty} x^2 \frac{1}{\pi} \frac{1}{x^2+1} \mathrm{d}x = \frac{2}{\pi} \int_0^{\infty} \frac{x^2}{x^2+1} \mathrm{d}x,$$

which equals ∞ since $\frac{x^2}{x^2+1} \to 1$ as $x \to \infty$.

So T does not have a well-defined (finite) mean or variance, and the LLN and CLT both do not apply. (In fact, it may be shown that $\frac{1}{n}(T_1 + \ldots + T_n)$ does not converge to 0 but rather $\frac{1}{n}(T_1 + \ldots + T_n) \sim t_1$ for any n.)

2 The t_n distribution for large n.

(a) We may write $U_n = \sum_{i=1}^n X_i^2$, where $X_1, \ldots, X_n \overset{IID}{\sim} \mathcal{N}(0,1)$. Then the LLN implies $\frac{1}{n}U_n \to \mathbb{E}[X_i^2] = 1$ in probability as $n \to \infty$. The function $x \mapsto 1/\sqrt{x}$ is continuous (at every $x > 0$), so by the Continuous Mapping Theorem

$$\frac{1}{\sqrt{\frac{1}{n}U_n}} \to 1$$

in probability as $n \to \infty$.

(b) We may write $T_n = \frac{1}{\sqrt{\frac{1}{n}U_n}} Z_n$ where $Z_n \sim \mathcal{N}(0,1)$, $U_n \sim \chi_n^2$, and Z_n and U_n are independent. By part (a), $\frac{1}{\sqrt{\frac{1}{n}U_n}} \to 1$ in probability. Clearly $Z_n \to \mathcal{N}(0,1)$ in distribution, since the distribution of Z_n does not change with n. Then $T_n \to \mathcal{N}(0,1)$ in distribution by Slutsky's lemma.

(c) The T statistic may be written as

$$T = \frac{\sigma}{S} \frac{\sqrt{n}\bar{X}}{\sigma}.$$

By the CLT, $\frac{\sqrt{n}\bar{X}}{\sigma} \to \mathcal{N}(0,1)$ in distribution. As $\frac{S}{\sigma} \approx 1$, the distribution of T is approximately $\mathcal{N}(0,1)$ for large n. (Formally, $\frac{S}{\sigma} \to 1$ in probability, so by Slutsky's lemma $T \to \mathcal{N}(0,1)$ in distribution.) The one-sided t-test rejects when $T > t_{n-1}(\alpha)$; by part (b), the t_{n-1} distribution is close to $\mathcal{N}(0,1)$ for large n, so $t_{n-1}(\alpha)$ is close to $z(\alpha)$. Then $\mathbb{P}[T > t_{n-1}(\alpha)] \approx \alpha$. The argument for a two-sided test is the same.

3 Comparing binomial proportions.

(a) Let $X_1, \ldots, X_n \overset{IID}{\sim} \text{Bernoulli}(p_A)$ and $Y_1, \ldots, Y_m \overset{IID}{\sim} \text{Bernoulli}(p_B)$ be indicators of whether each visitor clicked on the ad. We wish to test

$$\begin{aligned} H_0 &: p_A = p_B \\ H_1 &: p_A > p_B \end{aligned}$$

Both hypotheses are composite, as they do not specify the exact value of p_A or p_B.

(b) As $n\hat{p}_A \sim \text{Binomial}(n, p_A)$, we have $\text{Var}[n\hat{p}_A] = np_A(1-p_A)$ so $\text{Var}[\hat{p}_A] = \frac{p_A(1-p_A)}{n}$. Similarly $\text{Var}[\hat{p}_B] = \frac{p_B(1-p_B)}{m}$. Since $\hat{p}_A$ and $\hat{p}_B$ are independent,

$\text{Var}[\hat{p}_A - \hat{p}_B] = \text{Var}[\hat{p}_A] + \text{Var}[-\hat{p}_B] = \text{Var}[\hat{p}_A] + \text{Var}[\hat{p}_B] = \frac{p_A(1-p_A)}{n} + \frac{p_B(1-p_B)}{m}$.

Under H_0, $p_A = p_B = p$ for some $p \in (0,1)$, and this variance is $p(1-p)(\frac{1}{n} + \frac{1}{m})$. This is not the same for all data distributions in H_0, as it depends on p. (So we cannot perform a test of H_0 directly using the test statistic $\hat{p}_A - \hat{p}_B$.)

(c) One way of estimating the variance is to take

$$\hat{V} = \frac{\hat{p}_A(1-\hat{p}_A)}{n} + \frac{\hat{p}_B(1-\hat{p}_B)}{m}.$$

Another way (since $p_A = p_B$ under H_0) is to first estimate a pooled sample proportion

$$\hat{p} = \frac{\hat{p}_A n + \hat{p}_B m}{n+m}$$

and then estimate the variance as

$$\hat{V} = \hat{p}\,(1-\hat{p}) \left(\frac{1}{n} + \frac{1}{m}\right).$$

(Both ways are reasonable under H_0.)

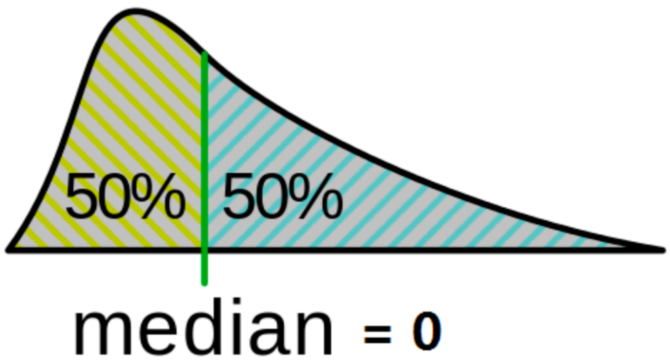
Under H_0, $p_A = p_B = p$ for some $p \in (0,1)$, so the CLT implies $\frac{\sqrt{n}(\hat{p}_A - p)}{\sqrt{p(1-p)}} \to \mathcal{N}(0,1)$ and $\frac{\sqrt{m}(\hat{p}_B - p)}{\sqrt{p(1-p)}} \to \mathcal{N}(0,1)$ in distribution as $n, m \to \infty$. So for large n and m, the distributions of $\hat{p}_A$ and $\hat{p}_B$ are approximately $\mathcal{N}(p, \frac{p(1-p)}{n})$ and $\mathcal{N}(p, \frac{p(1-p)}{m})$. $\hat{p}_A$ and $\hat{p}_B$ are independent, so their difference is distributed approximately as $\mathcal{N}(0, p(1-p)(\frac{1}{n} + \frac{1}{m}))$. We may write the test statistic T as

$$T = \frac{\sqrt{p(1-p)(\frac{1}{n} + \frac{1}{m})}}{\sqrt{\hat{V}}} \frac{\hat{p}_A - \hat{p}_B}{\sqrt{p(1-p)(\frac{1}{n} + \frac{1}{m})}}.$$

Since $\frac{\sqrt{p(1-p)(\frac{1}{n} + \frac{1}{m})}}{\sqrt{\hat{V}}} \approx 1$ with high probability for large n and m, T is approximately distributed as $\mathcal{N}(0,1)$, and an asymptotic level-α test rejects H_0 for $T > z(\alpha)$.

4 Sign test.

(a) For a density function f with median 0 but is skewed right, such as in the figure below, positive values of $X_1, \ldots, X_n$ would tend to have higher rank than negative values, so the Wilcoxon signed rank statistic would tend to take larger values under f than under any density function g that is symmetric about 0.



(b) Let $Y_i = 1$ if $X_i > 0$ and $Y_i = 0$ otherwise. Since f has median 0, $\mathbb{P}[Y_i = 1] = \mathbb{P}[X_i > 0] = 1/2$. Then

$$S = \sum_{i=1}^{n} Y_i \sim \text{Binomial}(n, \tfrac{1}{2}).$$

This distribution is the same for any PDF f with median 0. A test of H_0 versus H_1 should reject for large values of S. To achieve level-α, it should reject when $S \geq k$, where k is a value such that $\mathbb{P}[S \geq k] = \alpha$ under H_0. This is exactly the value of k given in the problem statement.

(c) Note $\mathbb{E}[Y_i] = 1/2$, $\text{Var}[Y_i] = 1/4$, and $\frac{S}{n} = \bar{Y}$. Then by the CLT

$$\sqrt{4n}\left(\frac{S}{n} - \frac{1}{2}\right) \to \mathcal{N}(0, 1)$$

in distribution as $n \to \infty$. So for large n,

$$\begin{aligned}
\alpha &\approx \mathbb{P}\left[\sqrt{4n}\left(\frac{S}{n} - \frac{1}{2}\right) > z(\alpha)\right] \\
&= \mathbb{P}\left[\frac{S}{n} > \frac{1}{2} + \frac{1}{\sqrt{4n}} z(\alpha)\right] \\
&= \mathbb{P}\left[S > \frac{n}{2} + \sqrt{\frac{n}{4}} z(\alpha)\right],
\end{aligned}$$

and we may take as an approximate rejection threshold $\frac{n}{2} + \sqrt{\frac{n}{4}} z(\alpha)$.

5 Power comparisons.

(a) The code below runs the simulations for the null case ($\mu = 0$) as well as for $\mu = 0.1, 0.2, 0.3, 0.4$:

```
set.seed(1)
n = 100
B = 10000
for (mu in c(0,0.1,0.2,0.3,0.4)) {
  output.Z = numeric(B)
  output.T = numeric(B)
  output.W = numeric(B)
  output.S = numeric(B)
  for (i in 1:B) {
    X = rnorm(n, mean=mu, sd=1)
    if (mean(X) > 1/sqrt(n)*qnorm(0.95)) {
      output.Z[i] = 1
    } else {
      output.Z[i] = 0
    }
    T = t.test(X)$statistic
    if (T > qt(0.95,df=n-1)) {
      output.T[i] = 1
    } else {
      output.T[i] = 0
    }
    W = wilcox.test(X)$statistic
    if (W > n*(n+1)/4+sqrt(n*(n+1)*(2*n+1)/24)*qnorm(0.95)) {
      output.W[i] = 1
    } else {
      output.W[i] = 0
    }
    S = length(which(X>0))
    if (S > n/2+sqrt(n/4)*qnorm(0.95)) {
      output.S[i] = 1
    } else {
      output.S[i] = 0
    }
  }
  print(paste('mu = ', mu))
  print(paste('Z: ', mean(output.Z)))
  print(paste('T: ', mean(output.T)))
  print(paste('W: ', mean(output.W)))
  print(paste('S: ', mean(output.S)))
}
```

Under H_0 (case $\mu = 0$), we obtained the results:

Test stat	Type I Error
Likelihood ratio test	0.0507
t-test	0.0505
Wilcoxon signed rank test	0.053
Sign test	0.0441

(b) Under these alternatives, we obtained the results:

Test stat	Power over $\mathcal{N}(\mu, 1)$			
	$\mu = 0.1$	$\mu = 0.2$	$\mu = 0.3$	$\mu = 0.4$
Likelihood ratio test	0.2631	0.6356	0.913	0.9895
t-test	0.261	0.6306	0.9085	0.9885
Wilcoxon signed rank test	0.252	0.6164	0.8978	0.9847
Sign test	0.1805	0.4617	0.7521	0.9318

(c) The powers of the tests against $\mathcal{N}(\mu, 1)$ decrease as we increasingly relax the distributional assumptions (from $\mathcal{N}(0, 1)$ to $\mathcal{N}(0, \sigma^2)$ to any symmetric PDF f about 0 to any PDF f with median 0). The sign test makes the fewest distributional assumptions under H_0, but its power is substantially lower than the other three tests. Hence if we have good reason to believe that the data distribution under H_0 is symmetric (for example, if each data value is the difference of paired samples (X_i, Y_i), and (X_i, Y_i) should have the same distribution as (Y_i, X_i) under H_0), then we should at least opt for using the Wilcoxon test. The difference in powers between the Wilcoxon test, t-test, and the most-powerful likelihood ratio test is indeed very small, which supports Rice's claim (at least for the tested sample size $n = 100$).