Statistics 582, Problem Set 9 Solutions

Wellner; 3/11/2015

1. For observations $\underline{X} = (X_1, \ldots, X_n)$, let $X_{(1)} \le \ldots \le X_{(n)}$ denote the *order statistics* of the X_i's ($X_{(i)} \equiv \mathbb{F}_n^{-1}(i/n)$, $i = 1, \ldots, n$) and let $\underline{R} = (R_1, \ldots, R_n)$ denote the *ranks* defined by $X_i = X_{(R_i)}$, $i = 1, \ldots, n$ (if $X_i = X_j$ for some $i < j$, define the ranks by $R_i < R_j$ and $X_i = X_{(R_i)}$).
   (a) Suppose that $X_1, \ldots, X_n$ are i.i.d. $F \in \mathcal{F}_{ac}$ (the absolutely continuous df's F on R) with density f. Show that the order statistics $\underline{X}_{(\cdot)} \equiv (X_{(1)}, \ldots, X_{(n)})$ are independent of the ranks $\underline{R}$ and that the order statistics have joint density $\bar{p}$ given by

   $$\bar{p}(\underline{x}_{(\cdot)}) = n! \prod_{i=1}^{n} f(x_{(i)}) \, , \qquad -\infty < x_{(1)} < \ldots < x_{(n)} < \infty$$

   while

   $$P(\underline{R} = \underline{r}) = \frac{1}{n!} \, , \qquad \underline{r} \in \Pi \equiv \{ \text{ all permutations of } \ \{1, \ldots, n\}\} \ .$$

   (b) Show that if the density f of the X_i's is log-concave, then the joint density $\bar{p}$ of the order statistics $\underline{X}_{(\cdot)}$ is log-concave; i.e. show that if $f((x+y)/2)^2 \ge f(x)f(y)$ for all $x, y \in \mathbb{R}$, then $\bar{p}((\underline{x}+\underline{y})/2)^2 \ge \bar{p}(\underline{x})\bar{p}(\underline{y})$ for all $\underline{x}, \underline{y} \in \mathcal{O}_n \equiv \{\underline{x} \in \mathbb{R}^n : \ x_1 \le x_2 \le \cdots \le x_n\}$.
   (c) Show that (a) continues to hold for any joint distribution p of the $\underline{X}$ which is symmetric with respect to permutation of its coordinates: $p(\pi\underline{x}) = p(\underline{x})$ for all $\underline{x}$ and $\pi \in \Pi$ where $\pi x \equiv (x_{\pi(1)}, \ldots, x_{\pi(n)})$.
   (d) If the joint density p of $\underline{X}$ is general (not permutation symmetric), show that the joint density $\bar{p}$ of the order statistics is given by

   $$\bar{p}(\underline{x}_{(\cdot)}) = \sum_{\pi \in \Pi} p(\pi \underline{x}_{(\cdot)}) \ ,$$

   and

   $$P(\underline{R} = \underline{r} | \underline{X}_{(\cdot)} = \underline{x}_{(\cdot)}) = \frac{p(r\underline{x}_{(\cdot)})}{\bar{p}(\underline{x}_{(\cdot)})} \ .$$

   Hint: Do (d) first by computing $P(\underline{X}_{(\cdot)} \in A)$ and $P(\underline{R} = \underline{r}, \underline{X}_{(\cdot)} \in A)$ for a Borel set $A \subset \{\underline{x} \in \mathbb{R}^n : \ x_1 < x_2 < \cdots < x_n\}$ and a fixed permutation $\underline{r}$ of $\{1, \ldots, n\}$.

Solution: I will prove (d) first; then (a) and (c) follow as corollaries:
(d) Suppose that $\underline{X}$ has joint density p. Then for any set Borel set $A \subset \{\underline{x} \in \mathbb{R}^n : \ x_1 < x_2 < \cdots < x_n\}$

$$\begin{aligned}
P(\underline{X}_{(\cdot)} \in A) &= \int_{[\underline{x}_{(\cdot)} \in A]} p(x_1, \ldots, x_n) \ dx_1 \ldots dx_n \\
&= \sum_{r \in \Pi} \int_{[R(\underline{x})=r, \ \underline{x}_{(\cdot)} \in A]} p(x_1, \ldots, x_n) \ dx_1 \ldots dx_n \\
&= \sum_{r \in \Pi} \int_A p(x_{(r_1)}, \ldots, x_{(r_n)}) \ dx_{(1)} \ldots dx_{(n)} \\
&= \int_A \overline{p}(x_{(1)}, \ldots, x_{(n)}) \ dx_{(1)} \ldots dx_{(n)}
\end{aligned}$$

where we have used the fact that the correspondence between $(x_1, ..., x_n)$ and $(x_{(1)}, \ldots, x_{(n)})$ is one-to-one and linear with Jacobian $= 1$ on each subset $[R = r]$, $r \in \Pi$. This proves that

$$\overline{p}(\underline{x}_{(\cdot)}) = \sum_{\pi \in \Pi} p(\pi \underline{x}_{(\cdot)}) \ .$$

Similarly,

$$\begin{aligned}
P(R = r, \ \underline{X}_{(\cdot)} \in A) &= \int_{[R=r, \ \underline{x}_{(\cdot)} \in A]} p(x_1, \ldots, x_n) \ dx_1 \cdots dx_n \\
&= \int_A p(x_{(r_1)}, \ldots, x_{(r_n)}) \ dx_{(1)} \ldots dx_{(n)} \\
&= \int_A \frac{p(x_{(r_1)}, \ldots, x_{(r_n)})}{\overline{p}(x_{(1)}, \ldots, x_{(n)})} \overline{p}(x_{(1)}, \ldots, x_{(n)}) \ dx_{(1)} \ldots dx_{(n)}
\end{aligned}$$

since $\overline{p}(x_{(1)}, \ldots, x_{(n)}) = 0$ implies $p(x_{r(1)}, \ldots, x_{r(n)}) = 0$ for each $r \in \Pi$. This implies that

$$P(\underline{R} = \underline{r} | \underline{X}_{(\cdot)} = \underline{x}_{(\cdot)}) = \frac{p(r\underline{x}_{(\cdot)})}{\overline{p}(\underline{x}_{(\cdot)})} \ .$$

(c) When $p(\underline{x}) = p(\pi \underline{x})$ for all $\pi \in \Pi$, then

$$\overline{p}(\underline{x}_{(\cdot)}) = n! p(\underline{x}_{(\cdot)}),$$

and

$$P(\underline{R} = \underline{r} | \underline{X}_{(\cdot)} = \underline{x}_{(\cdot)}) = \frac{p(r\underline{x}_{(\cdot)})}{\overline{p}(\underline{x}_{(\cdot)})} = \frac{p(r\underline{x}_{(\cdot)})}{n! p(\underline{x}_{(\cdot)})} = \frac{1}{n!}.$$

Hence R is independent of $\underline{X}_{(\cdot)}$, and $P(R = r) = 1/n!$ for each $r \in \Pi$.
(a) This follows easily from (c) since, in this case, for any permutation π

$$p(\pi\underline{x}) = \prod_{i=1}^{n} f(x_{\pi(i)}) = \prod_{i=1}^{n} f(x_i) = p(\underline{x}).$$

(b) If the marginal density f of the X_i's is log-concave, then

$$\begin{aligned}
\overline{p}((\underline{x}_{(\cdot)} + \underline{y}_{(\cdot)})/2)^2 &= n!^2 \prod_{i=1}^{n} f((\underline{x}_{(i)} + \underline{y}_{(i)})/2)^2 \\
&\geq n!^2 \prod_{i=1}^{n} f(\underline{x}_{(i)}) f(\underline{y}_{(i)}) \\
&= \left(n! \prod_{i=1}^{n} f(\underline{x}_{(i)}) \right) \left(n! \prod_{i=1}^{n} f(\underline{y}_{(i)}) \right) \\
&= \overline{p}(\underline{x}_{(\cdot)}) \cdot \overline{p}(\underline{y}_{(\cdot)}).
\end{aligned}$$

Thus the joint density of the order statistics is also log-concave.

2. In a comparison of the effect on growth of two diets B and C, a number of growing rats were placed on these two diets, and the following growth figures were observed after 7 weeks:

$$B : 156, 183, 120, 113, 138, 145, 142$$
$$C : 109, 107, 119, 162, 121, 123, 76, 111, 130, 115.$$

(This data is from Lehmann (1975), *Nonparametrics: Statistical Methods based on Ranks*, problem 20, page 108. Lehmann references the original article in a footnote on the same page.)
(a) Let the growth figures from the B diet be denoted by Y_j, $j = 1, \ldots, 7$, and let the growth figures from the C diet be denoted by X_i, $i = 1, \ldots, 10$. Assuming that the X_i's are i.i.d. with mean μ and that the Y_j's are i.i.d. with mean ν and that the X's and Y's have a common variance, use a two-sample t-test to test $H : \mu \geq \nu$ versus $K : \mu < \nu$. What is the p-value for your test?
(b) Now use a two-sample permutation test (or approximation thereof via sampling) to test the same hypotheses as in (a): what is the p-value (or approximate p-value) for your permutation test? Is this in (approximate) agreement with the limit theorem we proved in class?

Solution: (1) I compute, with $m = 10$, $n = 7$,

$$\tau_{m,n} = \frac{\sqrt{\frac{mn}{N}}(\overline{Y}_n - \overline{X}_m)}{\sqrt{\frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}}} = 2.302,$$

and the $p-$value is $P(t_{10,7} \geq 2.302) = .018$.
(b) Here the number of elements in the permutation distribution of $\tau_{m,n} = \tau_{10,7}$ is $\binom{17}{7} = 19448$. By sampling 3000 permutations (or samples without replacement from the pooled data), I find an estimated $p-$value of 0.0183, which is in close agreement with the $p-$value of 0.018 for the two-sample (normal theory) $t-$test we obtain in (a). Figure 1 gives the histogram of values of $\overline{Y}$ for the 3000 sampled labelling of the pooled data as X's and Y's. By enumerating the entire list of all 19448 values of $\tau_{m,n}$ obtained from all possible relabelling of the data as X's and Y's I get a $p-$value of 0.0186.

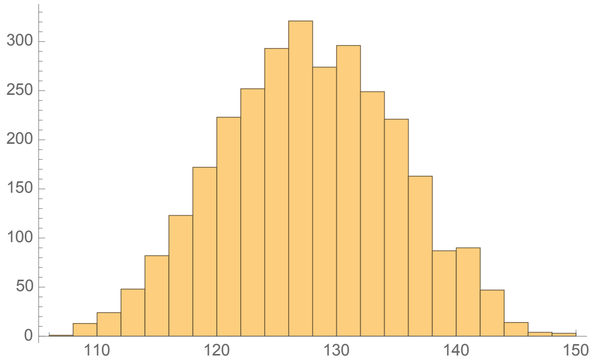

Figure 1: Sampled permutation distribution, two-sample $t-$statistic $\tau_{m,n}$, 3000 "chooses" or "re-labellings"

3. Read TPE (Lehmann and Casella) pages 160 - 162 concerning the notion of *equivariance* of an estimator $\delta = \delta(X)$ under a group of transformations G. Relate this to *invariance* of a (test) function ϕ under a group of transformations G. Illustrate equivariance with two examples.

   Solution: Let $X \sim P_\theta$ for $\theta \in \Theta$, and suppose that we want to estimate some function $h(\theta) \in \mathcal{H}$. For a group of transformations G on the sample space $\mathcal{X}$, we typically induce a group $\overline{G}$ on the parameter space Θ via the correspondence $gX \sim P_{\overline{g}\theta}$. Suppose, moreover, that $\overline{G}$ induces a group G^* on $\mathcal{H}$ via $h(\overline{g}\theta) = g^* h(\theta)$. If $\delta : \mathcal{X} \to \mathcal{H}$ yields an estimator $\delta(X)$ of $h(\theta)$, then we expect to use $g^* \circ \delta(X)$ or $\delta(gX)$ to estimate $h(\overline{g}\theta)$. Thus equivariance is just the requirement that $g^* \circ \delta(X) = \delta(gX)$.

   It is fairly straightforward to relate this to the testing situation, in which case $h(\theta) = 1_{\Theta_K}(\theta)$ and the induced group G^* reduces to the trivial group $G = \{e\}$.

   Here are two examples of equivariance in estimation:

   Example 1. **Location** Suppose that $X = (X_1, \ldots, X_n)$ where the X_i's are i.i.d. $N(\theta, \sigma^2)$ where $\sigma^2 > 0$ is known. We want to estimate $h(\theta) = \theta$. If

$G = \{g_c : g_c(x) = x + c\underline{1},\ c \in R\}$, then the induced group on the parameter space is $\overline{G} = \{\overline{g}_c : g_c(\theta) = \theta + c,\ c \in R\}$, and this is also the group G^* in the discussion above. Note that for the usual estimator $\delta(X) = \overline{X} = n^{-1}\sum_1^n X_i$ we have

$$\delta(g_c X) = \overline{X} + c = g_c^*(\overline{X}),$$

i.e. $\delta = \overline{X}$ is (location) equivariant.

Example 2. Scale Suppose that $X = (X_1, \ldots, X_n)$ where the X_i's are i.i.d. $N(0, \theta^2)$ where $\theta > 0$ is unknown. We want to estimate $h(\theta) = \theta^2$. If $G = \{g_c :\ g_c(x) = cx,\ x \in R^n,\ c > 0\}$, then the induced group on the parameter space $\Theta = \{\theta :\ \theta > 0\}$ is $\overline{G} = \{\overline{g}_c :\ \overline{g}_c(\theta) = c\theta\}$, and in this case the group $G^* = \{g_c^* :\ g_c^*(h) = c^2 h :\ c > 0\}$ since $h(\overline{g}\theta) = (c\theta)^2 = c^2 h(\theta)$. For the natural (consistent) estimator $\delta(X) = S_X^2 = n^{-1}\sum_1^n (X_i - \overline{X})^2$ of $h(\theta) = \theta^2$, we have

$$\delta(g_c(X)) = c^2 S_X^2 = g_c^*(\delta(X)),$$

i.e. $\delta = S_X^2$ is (scale)-equivariant.

4. Let X and Y be independent exponential random variables with parameters λ and μ respectively: thus $P(X > x) = \exp(-\lambda x)$ and $P(Y > y) = \exp(-\mu y)$ for $x, y \geq 0$. Let $\theta \equiv \lambda/\mu$.
(a) Show that the problem of testing $H_0 : \theta \leq 1$ versus $H_1 : \theta > 1$ is invariant under the group G of transformations $g_c(x, y) = (cx, cy)$, $c > 0$.
(b) Find a maximal invariant.
(c) Find a UMP invariant test of size α.
(d) Show that the problem of testing $H_0' : \theta = 1$ versus $H_1' : \theta \neq 1$ is invariant *in addition* under the transformation $g(x, y) = (y, x)$.
(e) Find a UMP invariant test of size α of H_0' versus H_1'.
(f) Find UMP invariant tests of the hypotheses in (a) and (d) when $X_1, \ldots, X_m$ are i.i.d. Exponential(λ) and $Y_1, \ldots, Y_n$ are i.i.d. Exponential(μ).

Solution: (a) Now X, Y have joint density

$$p_{\lambda,\mu}(x, y) = \lambda e^{-\lambda x} \mu e^{-\mu y} 1_{[0,\infty)}(x) 1_{[0,\infty)}(y)$$

so that if $c > 0$, $g_c(X, Y) = (cX, cY) \sim p_{\lambda/c, \mu/c}(x, y)$ and hence $\overline{g}(\lambda, \mu) = (\lambda/c, \mu/c)$. Note that $\delta(\lambda, \mu) = \lambda/\mu = (\lambda/c)/(\mu/c) = \delta(\overline{g}(\lambda, \mu))$, so that the hypotheses $H : \delta \leq 1$ and $K :\ \delta > 1$ are invariant. $T(X, Y) = X/Y$ and $\delta(\lambda, \mu) = \lambda/\mu$ are maximmal invariants on the sample space and parameter space respectively, and since $2\mu Y \sim \chi_2^2$ and $2\lambda X \sim \chi_2^2$ are independent

$$T(X, Y) = \frac{2\lambda X}{2\mu Y}\frac{\mu}{\lambda} \sim \delta^{-1} F_{2,2}.$$

Since this family of distributions has monotone (decreasing) likelihood ratio, it follows that the UMP $G-$invariant test of H versus K rejects H when $T < F_{2,2,\alpha}$ (where $P(F_{2,2} \leq F_{2,2,\alpha}) = \alpha$. or when $T^{-1} = Y/X > F_{2,2,1-\alpha} = (1-\alpha)/\alpha$.
(b) If $g(x,y) = (y,x)$ is also considered, then the transformation induced on the maximal invariant of A is given by

$$g(T) = T(g(X,Y)) = T(Y,X) = \frac{Y}{X} = \frac{1}{T(X,Y)}.$$

Claim: $T \vee T^{-1}$ is the maximal invariant with respect to this new group.
Proof. $T \vee T^{-1}$ is invariant since $T^{-1} \vee T = T \vee T^{-1}$; and $T \vee T^{-1}$ is maximal since $T \vee T^{-1} = T^* \vee T^{*-1}$ implies that either $T = T^*$ or $T = T^{*-1}$.

A corresponding maximal invariant on the parameter space is $\delta \vee \delta^{-1} = \frac{\lambda}{\mu} \vee \frac{\mu}{\lambda} \equiv \nu$, and the hypotheses $H : \delta = 1$ and $K : \delta \neq 1$ are clearly invariant. When expressed in terms of ν the hypotheses become $H : \nu = 1$ versus $K : \nu > 1$. It remains to show that the maximal invariant has monotone likelihood ratio (MLR).

By direct calculation, for $t > 1$,

$$1 - F_\nu(t) \quad = \quad P_{\lambda,\mu}(T \vee T^{-1} > t) = \frac{2 + \eta t}{1 + \eta t + t^2} \tag{1}$$

where $\eta \equiv \eta(\nu) = \delta + 1/\delta = \nu + 1/\nu$ is a monotone function of $\nu \geq 1$. Hence $T \vee T^{-1}$ has density

$$f_\nu(t) = \frac{\eta(t^2+1) + 4t}{(t^2 + \eta(\nu)t + 1)^2} 1_{[1,\infty)}(t).$$

so that for $\eta_1 < \eta_2$

$$\frac{f_{\eta_2}}{f_{\eta_1}}(t) = \left(\frac{t^2 + \eta_1 t + 1}{t^2 + \eta_2 t + 1}\right)^2 \frac{\eta_2(t^2+1) + 4t}{\eta_1(t^2+1) + 4t} \equiv g(t)^2 h(t)$$

where

$$g'(t) = \frac{(\eta_2 - \eta_1)(t^2 - 1)}{(t^2 + \eta_2 t + 1)^2} \geq 0 \qquad \text{for } \; t \geq 1$$

and

$$h'(t) = \frac{(\eta_2 - \eta_1)(4t^2 - 1)}{(\eta_1(t^2+1) + 4t)^2} \geq 0 \qquad \text{for } \; t \geq 1.$$

Hence the distribution of $T \vee T^{-1}$ has MLR and the UMP $G-$invariant test rejects H when $M \equiv T \vee T^{-1} > \frac{2-\alpha}{\alpha}$. Then, when $\delta = 1$, $\eta = 2$, and

$$P_{\nu=1}(M > (2-\alpha)/\alpha)) = \alpha.$$

[Proof of (1): Since $T \sim \delta^{-1} F_{2,2}$ where $P(F_{2,2} \le x) = x/(1+x)$,

$$\begin{aligned} P(T \vee T^{-1} > t) &= P(T > t) + P(T < t^{-1}) = P(F_{2,2} > \delta t) + P(F_{2,2} < \delta/t) \\ &= \frac{1}{1+\delta t} + \frac{\delta/t}{1+\delta/t} = \frac{1/\delta}{1/\delta + t} + \frac{\delta}{\delta + t} \\ &= \frac{2 + (\delta + 1/\delta)t}{1 + (\delta + 1/\delta)t + t^2} = \frac{2 + (\nu + 1/\nu)t}{1 + (\nu + 1/\nu)t + t^2}.] \end{aligned}$$

(c) When $X_1, \ldots, X_m$ are i.i.d. exponential(λ) and $Y_1, \ldots, Y_n$ are i.i.d. exponential(μ) respectively, and $G = \{g : g(\underline{x}, \underline{y}) = (c\underline{x}, c\underline{y}), c > 0\}$, then we reduce by sufficiency first: $(\sum_1^m X_i, \sum_1^n Y_j)$ is sufficient for (λ, μ). Then $T = \sum_1^m X_i / \sum_1^n Y_j$ is a maximal invariant with respect to $G^* = \{g^* : g^*(x, y) = (cx, cy),\ c > 0\}$ acting on the space of the sufficient statistic. Moreover,

$$T = \frac{\mu}{\lambda} \frac{m}{n} \frac{2\lambda \sum_1^m X_i/(2m)}{2\mu \sum_1^n Y_j/(2n)} \sim \frac{1}{\delta n/m} F_{2m,2n}$$

where $\delta \equiv \delta(\lambda, \mu) = \lambda/\mu$ is a maximal invariant with respect to $\overline{G}$ on the parameter space, and the density of $F_{2m,2n}$ is given by

$$f_{F_{2m,2n}}(x) = c_{m,n} \frac{x^{m-1}}{(1 + (m/n)x)^{m+n}} 1_{(0,\infty)}(x).$$

Thus the density of T is given by

$$f_T(x, \delta) = \delta \tilde{c}_{m,n} \frac{(\delta x)^{m-1}}{(1 + \delta x)^{m+n}} 1_{(0,\infty)}(x)$$

which has monotone (decreasing) likelihood ratio in $M(x) = x$. Thus the UMP-G^* invariant test rejects H_0 when $((n/m)T)^{-1} = (\sum_1^n Y_j/2n)/(\sum_1^m X_i/2m) > F_{2n,2m,\alpha}$ where $P(F_{2n,2m} > F_{2n,2m,\alpha}) = \alpha$.

For testing $H_0' : \theta = 1$ versus $H_1' : \theta \ne 1$, we find, much as in (b), that $\tilde{T} \equiv T \vee T^{-1}$ is a maximal invariant with respect to the induced group $G^* = G_2^* \oplus G_1^*$ on the space of the sufficient statistic. By a calculation similar to that of part (b),

$$\begin{aligned} 1 - F(t) &= P(T \vee T^{-1} > t) = P(T > t) + P(T < 1/t) \\ &= P(\delta^{-1}(m/n) F_{2m,2n} > t) + P(\delta^{-1}(m/n) F_{2m,2n} < 1/t) \\ &= P((m/n) F_{2m,2n} > \delta t) + P((m/n) F_{2m,2n} < \delta/t) \end{aligned}$$

and hence

$$\begin{aligned} f_{T \vee T^{-1}}(t) &= \delta \tilde{c}_{m,n} \frac{(\delta t)^{m-1}}{(1 + \delta t)^{m+n}} + \frac{\delta}{t^2} \tilde{c}_{m,n} \frac{(\delta/t)^{m-1}}{(1 + \delta/t)^{m+n}} \\ &= \tilde{c}_{m,n} \frac{t^{m-1}(\delta + t)^m (1 + t/\delta)^n + t^{n-1}(1 + \delta t)^m (1/\delta + t)^n}{(t^2 + (\delta + 1/\delta)t + 1)^{m+n}}. \end{aligned}$$

It is not yet clear to me that this density depends only on $\nu = \delta \vee 1/\delta$, so I may be making a mistake somewhere.

5. **Optional bonus problem 1:** (From Wasserman, *All of Statistics*, page 171.) In 1961, 10 essays appeared in the *New Orleans Daily Crescent*. They were signed "Quintus Curtius Snodgrass" and some people suspected they were actually written by Mark Twain. To investigate this, consider the proportion of three letter words founds in an author's work. From eight Twain essays we have

   .225, .262, .217, .240, .230, .229, .235, .217

   From 10 Snodgrass essays we have:

   .209, .205, .196, .210, .202, .207, .224, .223, .220, .201

   (a) Perform a Wald test for equality of the means. Give a $p-$value and a 95% confidence interval for the difference of means. What conclusion do you reach?
   (b) Now use a permutation test to test the equality of means. What is your conclusion?

   Solution: (a) Labelling the Twain proportions as X's and the Snodgrass proportions as Y's, we find that $\overline{X}_m = .231875$, $\overline{Y}_n = .2097$, $S_X = .01456$, and $S_Y = .00966$. Assuming that $X_i \sim N(\mu, \sigma^2)$ and $Y_j \sim N(\nu, \tau^2)$ with $\sigma \neq \tau$, the Wald statistic becomes

$$
\begin{aligned}
W_{m,n} &= \left\{ \frac{\sqrt{\frac{mn}{N}}(\overline{X}_m - \overline{Y}_n)}{\sqrt{(n/N)S_X^2 + (m/N)S_Y^2}} \right\}^2 \\
&= \left\{ \frac{\sqrt{\frac{8 \cdot 10}{18}}(.231875 - .2097)}{\sqrt{(10/18)(.000212125) + (8/18)(.0000933444)}} \right\}^2 \\
&= 3.70355^2 = 13.7163
\end{aligned}
$$

   and the (approximate) p-value is $P(\chi_1^2 > 13.7163) = .000213$. If we use Welch's approximate $t-$test (see e.g. Lehmann and Casella, TSH, page 447), then the degrees of freedom f becomes, with $R \equiv mS_X^2/(nS_Y^2)$

$$
\frac{1}{f} = \left(\frac{R}{1+R}\right)^2 \frac{1}{m-1} + \frac{1}{(1+R)^2} \frac{1}{n-1} = 1/13.6148.
$$

   Thus the approximate p-value using Welch's approximation is $P(|t_{13.61}| \geq 13.7163) =$

.00265. A 95% confidence interval for $\mu - \nu$ based on normal theory is given by

$$\begin{aligned}
&\overline{X}_m - \overline{Y}_n \pm z_{.025}\sqrt{S_X^2/m + S_Y^2/n} \\
&= .022175 \pm 1.95996\sqrt{.000212125/8 + .0000933444/10} \\
&= (0.0104397, 0.0339103)
\end{aligned}$$

The conclusion based on either of these tests is to reject the null hypothesis: from this evidence we would conclude that the Snodgrass and Twain essays were written by different authors.

(b) If we do an exact permutation t-test using the statistic introduced in class (involving the assumption of equal variances in the alternative), there are $\binom{18}{8} = 43758$ combinations to consider, and the observed value of the statistic (in the form $(\overline{X}_m - \overline{z})/\sigma_N$) is 2.78917. By my calculations the exact one-sided p-value is 0.000525618, and the exact two-sided p-value is 0.000777. In contrast, by drawing $10^5 = 100,000$ random permutations, the estimated p-values were 0.00058 and 0.00088 respectively.

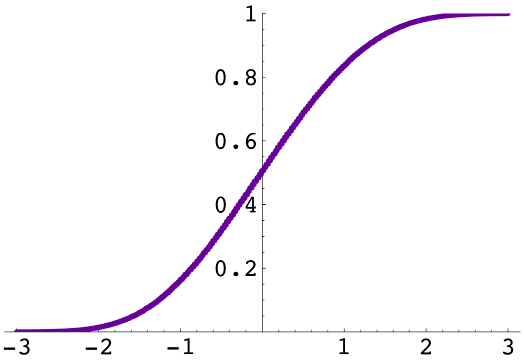

Figure 2: Exact permutation distribution, two-sample $t-$statistic $(\overline{X} - \overline{z})/\sigma_N$

I have not yet programmed the exact permutation test based on the Wald - type statistic used in part (a), but without squaring, which allows for the possibility of different variances. There are still $\binom{18}{8} = 43758$ combinations to consider, and the observed value of the test statistic is 3.70355. I have however, programmed the approximate permutation test based on sampling from the permutation distribution. By drawing $10^5 = 100,000$ random permutations, the estimated p-values I calculated are 0.00461 and 0.01064 respectively. It seems that the permutation distribution of the unequal variances version of the unsquared form of the Wald statistic is more nearly normal than that of the classical $t-$statistic. Note that while the two-sided permutation test still rejects at level $\alpha = .05$, this two-sided p-valued (0.01064)is not nearly as small as the estimated p-value of the

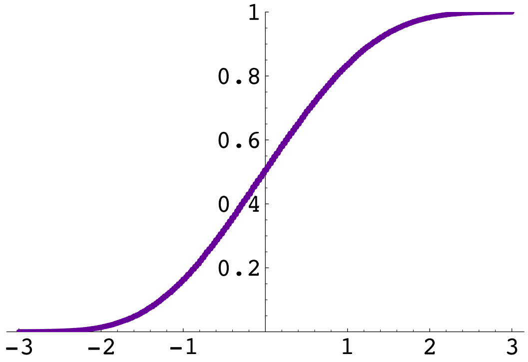

Figure 3: Approximate permutation distribution, two-sample $t-$statistic $(\overline{X} - \overline{z})/\sigma_N$, based on 10^5 random permutations

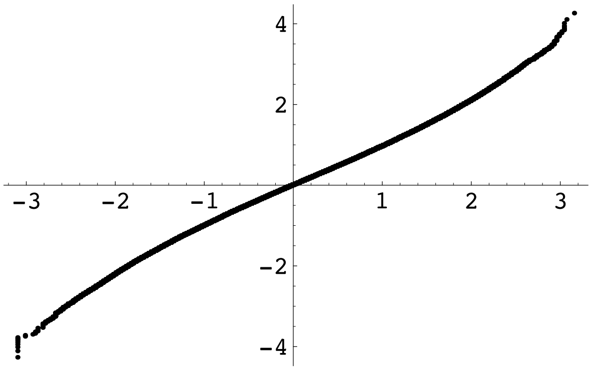

Figure 4: QQ-plot, Approximate permutation distribution, two-sample $t-$statistic $(\overline{X} - \overline{z})/\sigma_N$, based on 10^5 random permutations

permutation $t-$ test noted above (0.00088). We would continue to reject the null hypothesis at the level 0.05, but not at 0.01.

6. **Optional bonus problem 2:** (From the material on consistency of Neyman Pearson tests, section 6.1; and Donoho and Jin, *Ann. Statist.* **32** (2004), 962-994)
Let $p_\mu(x) \equiv \phi(x - \mu)$ denote the density of $X \sim N(\mu, 1)$, and let P_μ denote the corresponding measure on $\mathbb{R}$.
(a) Consider testing $H: \ P = P_0$ versus $K: \ Q = P_\mu$ with $\mu > 0$ fixed. Compute $\rho(P, Q) = \rho(P_0, P_\mu) = \int \sqrt{p_0(x) p_\mu(x)} dx$ explicitly as a function of μ.
(b) Compare the power functions of the following two tests of H versus K when $\mu = 1$:
(i) The Neyman - Pearson test with $\alpha = .05$; (ii) The Neyman - Pearson type test with $k = k_n = 1$ (in the notation of Theorem 6.1.4, page 8, Chapter 6).

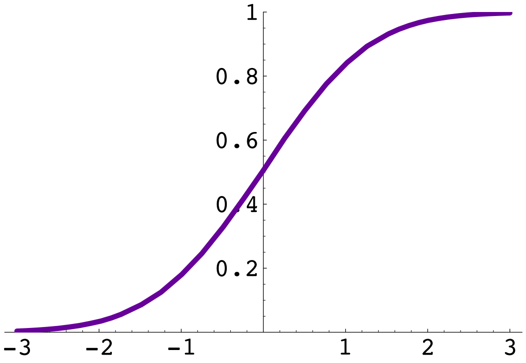

Figure 5: Approximate permutation distribution, one-sided Wald statistic $\overline{X} - \overline{Y})/\sqrt{S_X^2/m + S_Y^2/n}$ based on 10^5 random permutations

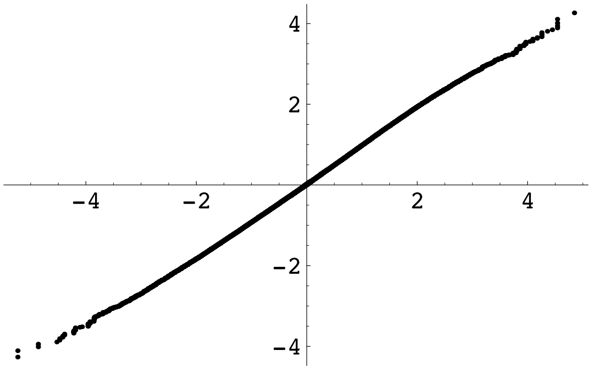

Figure 6: QQ-plot, Approximate permutation distribution, one-sided Wald statistic $\overline{X} - \overline{Y})/\sqrt{S_X^2/m + S_Y^2/n}$ based on 10^5 random permutations

(c) Now suppose $\mu = \mu_n \equiv t/\sqrt{n}$ with $t > 0$ and consider testing $H : P = P_0$ versus $K_n : \ P = P_{\mu_n}$.
(i) Find the limit of $\rho(P_0, P_{\mu_n})^n$ as a function of t.
(ii) Compare the limiting power function of the two tests in (b).
(d) Now suppose that $Q = Q_n$ is given by the mixture

$$q(x) = q(x;\mu,\epsilon) = (1-\epsilon)p_0(x) + \epsilon p_\mu(x)$$

where $\epsilon = \epsilon_n = n^{-\beta}$ with $1/2 < \beta < 1$ and $\mu = \mu_n = \sqrt{2r\log n}$ with $r > \rho^*(\beta)$ and where the function

$$\rho^*(\beta) = \begin{cases} \beta - 1/2, & 1/2 \le \beta < 3/4, \\ (1-\sqrt{1-\beta})^2, & 3/4 \le \beta < 1. \end{cases}$$

This is a model for "sparse normal means". Show that k_n can be chosen so that $E_{P_0^n}\phi_n(\underline{X}) \to 0$ and $E_{Q_n}^n(1-\phi_n(\underline{X}_n)) \to 0$.