i. True distance still depends on which member of $\Omega_0$

   i. Dependence is weak.

   ii. For $n$ large, $F_{1,n-1} \approx \chi^2_1$

7. Inference about two normal populations, equal variance:

   a. $X_1, \ldots, X_m \sim N(\mu, \sigma^2), Y_1, \ldots, Y_n \sim N(\nu, \tau^2)$ all ind.

   b. Log likelihood $l(\mu, \nu, \sigma, \tau)$

   \[
   = -\frac{\sum_j (X_j - \mu)^2}{2\sigma^2} - \frac{\sum_j (Y_j - \nu)^2}{2\tau^2} - m \log(\sigma) - n \log(\tau).
   \]

   c. $H_0: \mu = \nu$, $H_A: \mu \neq \nu$.

   d. Assume $\sigma = \tau$.

   e. M.L.E. for $\mu$ and $\nu$ under $H_A$: $\bar{X}$ and $\bar{Y}$.

   f. M.L.E. for $\mu$ and $\nu$ under $H_0$:

   \[
   \ddot{Z} = (\sum_j X_j + \sum_j Y_j)/(m + n).
   \]

   g. $\hat{\sigma}^2 =$ M.L.E. for $\sigma^2$ under $H_A$ solves

   \[
   0 = \frac{\sum_j (X_j - \bar{X})^2 + \sum_j (Y_j - \bar{Y})^2}{\hat{\sigma}^3} - \frac{m + n}{\hat{\sigma}}
   \]

   \[
   \hat{\sigma}^2 = \frac{\sum_j (X_j - \bar{X})^2 + \sum_j (Y_j - \bar{Y})^2}{m + n}
   \]

   h. As before M.L.E. for $\sigma^2$ under $H_0$ is

   \[
   \hat{\sigma}^2 = \frac{\sum_j (X_j - \bar{Z})^2 + \sum_j (Y_j - \bar{Z})^2}{m + n},
   \]
i. g.l.r.t. for $\mu = \nu$ vs $\mu \neq \nu$:

$$\Lambda = \exp\left(\frac{-\sum_j (X_j - \bar{Z})^2}{2\tilde{\sigma}^2} - \frac{\sum_j (Y_j - \bar{Z})^2}{2\tilde{\sigma}^2}\right)\tilde{\sigma}^{m-n} / \left[ \exp\left(\frac{\sum_j (X_j - \bar{X})^2}{2\hat{\sigma}^2} + \frac{\sum_j (Y_j - \bar{Y})^2}{2\hat{\sigma}^2}\right) \hat{\sigma}^{m-n} \right]$$

$$= \exp\left(-\frac{m+n}{2}\right)\tilde{\sigma}^{-m-n} \exp\left(\frac{m+n}{2}\right)\hat{\sigma}^{m+n}$$

$$= \left(\frac{\hat{\sigma}}{\tilde{\sigma}}\right)^{m+n}$$

i. Test statistic is ratio of powers of sum of squares of deviations from means.

ii. The only such ratios whose dist$
$ns we can handle

- are to power 1.
- with sums ind.

iii. Noting that test is equivalent to rejecting when $\hat{\sigma}^2/\tilde{\sigma}^2$ small goes half way.

iv. Noting that sum of squares about $\bar{Z}$ can be expressed as sum of squares about $\bar{X}$ or $\bar{Y}$ plus an ind. bit does rest.

j. Note

i. $\sum_j (X_j - \bar{Z})^2 = \sum_j (X_j - \bar{X} + \bar{X} - \bar{Z})^2 = \sum_j (X_j - \bar{X})^2 + m(\bar{X} - \bar{Z})^2$
ii. \( \sum_j (Y_j - \bar{Z})^2 = \sum_j (Y_j - \bar{Y})^2 + n(\bar{Y} - \bar{Z})^2. \)

iii. \( \bar{Y} - \bar{Z} = \frac{(m+n)\bar{Y} - n\bar{Y} - m\bar{X}}{m+n} = \frac{m(\bar{Y} - \bar{X})}{m+n}. \)

iv. \( \bar{X} - \bar{Z} = n(\bar{X} - \bar{Y})/(m+n). \)

k. Then
\[
\tilde{\sigma}^2 - \hat{\sigma}^2 = n(\bar{Y} - \bar{Z})^2 + m(\bar{X} - \bar{Z})^2 = (m+n)^{-2}[nm^2 + mn^2](\bar{X} - \bar{Y})^2 = \left(\frac{1}{m} + \frac{1}{n}\right)^{-1} (1/m + 1/n)^{-1}(\bar{X} - \bar{Y})^2.
\]

l. Reject when large:
\[
\frac{m+n}{m+n-2} \frac{\tilde{\sigma}^2 - \hat{\sigma}^2}{\hat{\sigma}^2} = \left(\frac{1}{m} + \frac{1}{n}\right)^{-1} \frac{(\bar{X} - \bar{Y})^2}{\hat{\sigma}^2}.
\]

m. Multiplying the test statistic by \( \frac{m+n}{m+n-2} \) to give right numerator:
\[
\frac{(\bar{X} - \bar{Y})^2}{\left(\frac{1}{m} + \frac{1}{n}\right) S_p^2} \text{large, for } S_p^2 = \frac{\sum_j (X_j - \bar{X})^2 + \sum_j (Y_j - \bar{Y})^2}{m+n-2}.
\]

n. Equiv., reject when \( |T| \) large, for \( T = \frac{(\bar{X} - \bar{Y})}{S_p \sqrt{1/m + 1/n}}. \)

i. \( T \sim t(m+n-2). \)

ii. Reject when \( |T| \geq t_{1-\alpha/2}(m+n-2). \)

WMS: 10.9

8. Inference about normal variance:

a. \( X_1, \ldots, X_m \sim \mathcal{N}(\mu, \sigma^2) \)

i. \( H_0 : \sigma = \sigma_0 \) vs. \( H_A : \sigma \neq \sigma_0 \)
b. \( \hat{\mu} = \tilde{\mu} = \bar{X} \)

c. \( \hat{\sigma} = \sigma_0, \tilde{\sigma} = \sqrt{\frac{\sum_{j=1}^{m} (X_j - \bar{X})^2}{m}} \)

d. Likelihood ratio statistic is
\[
\Lambda = \frac{\exp \left( -\frac{\sum_{j=1}^{m} (X_j - \bar{X})^2}{2\sigma_0^2} \right) \sigma_0^{-m}}{\exp \left( -\frac{\sum_{j=1}^{m} (X_j - X)^2}{2\sum_{j=1}^{m} (X_j - X)^2/m} \right) \left( \sum_{j=1}^{m} (X_j - \bar{X})^2/m \right)^{-m/2}}
\]
\[
= \exp \left( -\frac{\sum_{j=1}^{m} (X_j - \bar{X})^2}{2\sigma_0^2} \right) \left( \frac{\sum_{j=1}^{m} (X_j - \bar{X})^2}{\sigma_0^2} \right)^{m/2}
\]
\[
\times \exp \left( \frac{m}{2} \right) m^{m/2}
\]
i. Fig. 15 compares the likelihood ratio and equal tailed tests.

CH: 9.3 (ii)

L. Summary: Three tests \( H_0 : \theta = \theta_0 \) vs. \( H_0 : \theta \neq \theta_0 \), all approximately \( \chi^2 \)

1. Wald test: \( (\hat{\theta} - \theta_0)^\top I(\theta_0)(\hat{\theta} - \theta_0) \),
   a. for \( I(\theta) = -\ell''(\theta) \), or \( \mathbb{E}[-\ell''(\theta)] \)
   b. For iid observations, \( \mathbb{E}[-\ell''(\theta)] = n\hat{i}(\theta) \).

2. Score test \( \ell'(\theta)^\top I(\theta_0)^{-1} \ell'(\theta) \)

3. Likelihood ratio test \( 2(\ell(\hat{\theta}) - \ell(\theta_0)) \).

WMS: 10.6

M. Rather than report just \( H_0 \) or \( H_a \), report significance level at
Fig. 15: Likelihood Ratio Statistic for Testing a Variance

\[ \sqrt{\sum_{j=1}^{m}(X_j - \bar{X})^2 / \sigma_0^2} \]

Straight lines connect critical values for equal tailed test. Slope of line indicates departure of standard equal-tailed test from generalized likelihood ratio test.
Lecture 9

which test would just reject $H_0$: Report $p$-value or attained significance level.

1. Answers question: How often would we see data this unfavorable or more unfavorable to the $H_0$ if $H_0$ true?

2. Keep $H_0$ unless $p$-value $\leq .05$ gives significance level of .05

3. Why?
   a. Allows others to use other significance levels.
   b. Does not relieve you of the necessity of choosing significance level in advance.

4. The $p$-value is sometimes described as measuring the evidence against the null hypothesis.
   a. Consider a null distribution that is uniform on $[0, 1]$,
   b. alternative distribution, according to the density
      \[
      f_A(x) = \begin{cases} 
      \gamma + \epsilon(x - 1/4) & \text{if } x \in [0, 1/2] \\
      2 - \gamma + \epsilon(x - 3/4) & \text{if } x \in (1/2, 1] 
      \end{cases},
      \]
      for some $\gamma \in (0, 1)$ and some $\epsilon > 0$ (Fig. 16)
   c. Consider one observation $X$ drawn from one of these two distributions.
   d. In this case, the likelihood ratio $f_0(x)/f_A(x)$ is monotone decreasing in $x$, 
e. critical region for the most powerful test is \( \{ X \geq x \} \).

f. Hence the \( p \)-value is \( 1 - x \) (Fig. 17).

g. Among those \( p \)-values less than \( 1/2 \), variation in these values has minimal value in distinguishing between the null and
alternative hypotheses.

5. How?

a. One-sided tests: If alternative is that $\mu > \mu_0$:
   i. Calculate $z$
   ii. Find probability that a standard normal random variable
b. Two-sided tests:
   i. Calculate $z$, and take abs value
   ii. Find probability that a standard normal random variable
        exceeds this
   iii. double it.

CH: 2.2 (vi)

N. Exponential family model:

1. Definition: $f_X(x; \theta) = \exp(t(x)c(\theta) - b(\theta))g(x)$
   a. $g$ and $b$ known, and $\Theta \subset \mathcal{R}$.
   b. Write $g(x) = \exp(-s(x))I_A(x)$ for $A$ free of $\theta$.

2. Examples:
   a. $X_i \sim \mathcal{E}(\alpha)$ i.i.d.
      i. $f_X(x; \alpha) = \alpha^n \exp(-\alpha \sum_i x_i)$.
      ii. $c(\alpha) = -\alpha$, $T(x) = \sum_i x_i$, $b(\alpha) = -n \log(\alpha)$, $s(x) = 0$.
   b. $X_i \sim \mathcal{N}(\mu, 1)$ i.i.d.
      i. $f_X(x; \mu) = (2\pi)^{-n/2} \exp(-n\mu^2/2 + \mu \sum_i x_i - \sum_i x_i^2/2)$.
      ii. $c(\mu) = \mu$, $T(x) = \sum_i x_i$, $b(\alpha) = n\mu^2/2$,
          $s(x) = (n/2) \log(2\pi) + \sum_i x_i^2/2$. 
Lecture 9

\[ \text{c. } X_i \sim \binom{1}{\pi} \text{ i.i.d.} \]

\[ \begin{align*}
&\text{i. } f_X(\mathbf{x}; \mu) = \exp(\sum x_i \log(\pi) + (n - \sum x_i) \log(1 - \pi)) = \\
&\quad \exp(\sum x_i [\log(\pi) - \log(1 - \pi)] + n \log(1 - \pi)) \\
&\text{ii. } c(\pi) = \log(\pi) - \log(1 - \pi), T(\mathbf{x}) = \sum x_i, \\
&\quad b(\pi) = -n \log(1 - \pi), s(\mathbf{x}) = 0. \\
\end{align*} \]

d. Gamma, P, NB in, ... work

e. \[ \text{X}_i \sim \mathcal{U}(0, \theta) \text{ i.i.d.} \]

\[ \begin{align*}
&\text{i. As before } f_X(\mathbf{x}) = \theta^{-n} I_{x(n) < \theta} \\
&\text{ii. Since } \{x(n) < \theta\} \text{ depends on } \theta, \text{ this isn’t an exponential family.} \\
\end{align*} \]

f. \[ \text{X}_i \sim \mathcal{T}(1, \theta) \text{ i.i.d.} \]

\[ \begin{align*}
&\text{i. } f_X(\mathbf{x}; \theta) = \pi^{-n} \prod_{i=1}^{n} (1 + (x_i - \theta)^2)^{-1} = \\
&\quad \exp(-n \log(\pi) - \sum_{i=1}^{n} \log(1 + (x_i - \theta)^2)) \\
&\text{ii. Not an exponential family: No way to disentagle } \mathbf{x} \text{ and } \theta. \\
\end{align*} \]

3. Natural sufficient statistics

a. Theorem: If \( f_X(\mathbf{x}; \theta), \theta \in \Theta \) is an exponential family model, with \( T = t(\mathbf{X}) \) as in the definition, then \( T \) is sufficient.

b. Proof: Let \( g(t; \theta) = \exp(c(\theta)t - b(\theta)) \) and

\[ h(\mathbf{x}) = \exp(-s(\mathbf{x})) I_A(\mathbf{x}). \]
c. $T$ is called the natural sufficient statistic
d. This includes examples above: exponential, normal, and Bernoulli. In each case $t(x) = \sum_i x_i$.

4. Natural parameterization: Transform to case with $c$ identity.
   a. Restrict attention to cases with original $c$ 1-1.
   b. Suppose $X$ satisfies general exponential family for $\theta \in \Theta$.
   c. Choose $\eta \in c(\Theta) = \{c(\theta) | \theta \in \Theta\}$, and let $\theta = c^{-1}(\eta)$.
   d. Try to write $f_X(x)$ as $\exp(\eta T(x) - b^*(\eta) - s(x))I_A(x)$ for some $b^*$
      i. $b^*(\eta) = \log(\int \exp(\eta T(x) - s(x)) \, dx) = b(\theta) \log(\int \exp(\eta T(x) - b(\theta) - s(x)) \, dx) = b(\theta)$.
      ii. Hence $b^*(\eta) = b(c^{-1}(\eta))$.
      iii. Also $\exp(\eta T(x) - b^*(\eta) - s(x))I_A(x)$ for $\eta \in c(\Theta)$ gives all of the distributions that the original family did.

5. Natural Exponential Family Examples:
   a. $X_1, \ldots, X_n \sim \mathcal{E}(\alpha)$ i.i.d.
      i. $f_X(x; \alpha) = \exp(-\alpha \sum_j x_j + n \log(\alpha))$ for $\alpha \in (0, \infty)$.
      ii. $T = \sum_i X_i$, $c(\alpha) = -\alpha = \eta$, $b(\alpha) = -n \log(\alpha)$,
          $b^*(\eta) = b(c^{-1}(\eta)) = -n \log(-\eta)$ for $\eta \in H \supset c((0, \infty)) =$
Lecture 10

\((-\infty, 0)\).

iii. Hence \(T \sim \Gamma(\alpha, n)\).

b. \(X_1, \ldots, X_n \sim \mathcal{N}(\mu, 1)\).

i. \(f_X(x; \mu) = \exp(\mu \sum_j x_j - \frac{1}{2} \sum_j x_j^2 - \frac{1}{2} n \mu^2 - (n/2) \log(2\pi))\) for \(\mu \in \mathcal{R}\).

ii. \(T = \sum_i X_i, c(\mu) = \mu = \eta, b(\alpha) = \frac{1}{2} n \mu^2, b^*(\eta) = \frac{1}{2} n \eta^2\) for \(\eta \in H \supset \mathcal{R}\).

iii. Hence \(T \sim \mathcal{N}(n\mu, n)\).

c. \(X_1, \ldots, X_n \sim \text{Bin}(n, \pi)\) i.i.d.

i. \(f_X(x; \pi) = \exp(\sum x_j \log(\pi) + (n - \sum_j x_j) \log(1 - \pi)) = \exp(\sum x_j \log(\pi/(1 - \pi)) + n \log(1 - \pi))\) for \(\pi \in (0, 1)\).

ii. \(T = \sum_i X_i, c(\pi) = \log(\pi/(1 - \pi)) = \eta,\)

\(c^{-1}(\eta) = 1 - 1/(1 + \exp(\eta)), b(\pi) = -n \log(1 - \pi),\)

\(b^*(\eta) = -n \log(1 - 1 + 1/(1 + \exp(\eta))) = n \log(1 + \exp(\eta))\)

for \(\eta \in \mathcal{R}\).

iii. Hence \(T \sim \text{Bin}(n, \pi)\).