i. True distribution still depends on which member of \( \Omega_0 \)
ii. Dependence is weak.
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 independent.
   b. Log likelihood:
      \[
      l(\mu, \nu, \sigma, \tau) = -\frac{1}{2\sigma^2} \sum_j (X_j - \mu)^2 - \frac{1}{2\tau^2} \sum_j (Y_j - \nu)^2 - m \log(\sigma) - n \log(\tau).
      \]
   c. \( H_0: \mu = \nu, \ H_A: \mu \neq \nu \).
   d. Assume \( \sigma = \tau \).
   e. MLE for \( \mu \) and \( \nu \) under \( H_A: \bar{X} \) and \( \bar{Y} \).
   f. MLE for \( \mu \) and \( \nu \) under \( H_0: \bar{Z} \) solves
      \[
      0 = \sum_j (X_j - \bar{Z})^2 + \sum_j (Y_j - \bar{Z})^2 - \frac{m+n}{\sigma^2}.
      \]
   g. \( \hat{\sigma}^2 = \text{MLE for } \sigma^2 \) under \( H_A \) solves
      \[
      \hat{\sigma}^2 = \frac{\sum_j (X_j - \bar{X})^2 + \sum_j (Y_j - \bar{Y})^2}{m+n}.
      \]
   h. As before MLE 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 \( \mu \neq \nu \):
      \[
      \Lambda = \exp \left( -\frac{\sum_{j=1}^m (X_j - \bar{Z})^2}{2\sigma_0^2} - \frac{\sum_{j=1}^n (Y_j - \bar{Z})^2}{2\tau^2} \right) \frac{\hat{\sigma}^{m-n}}{\hat{\sigma}^{m+n}}
      \]
      \[
      = \exp \left( -\frac{m+n}{2} \right) \frac{\hat{\sigma}^{m-n}}{\hat{\sigma}^{m+n}} \exp \left( \frac{m+n}{2} \right) \hat{\sigma}^{m+n}.
      \]
      i. Test statistic is ratio of powers of sum of squares of deviations from means.
   j. The only such ratios whose distributions we can handle
      • are to power 1.
      • with sums independent.
   k. Noting that test is equivalent to rejecting when \( \hat{\sigma}^2/\hat{\sigma}^2 \) small goes half way.
   l. Noting that sum of squares about \( \bar{Z} \) can be expressed as sum of squares about \( \bar{X} \) or \( \bar{Y} \) plus an independent bit does rest.
   m. Then
      \[
      \hat{\sigma}^2 - \hat{\sigma}^2 = n(\bar{Y} - \bar{Z})^2 + m(\bar{X} - \bar{Z})^2
      = (m+n)^{-2} [mn^2 + mn^2](\bar{X} - \bar{Y})^2
      = \left( \frac{1}{m} + \frac{1}{n} \right)^{-1} (1/m + 1/n)^{-1} (\bar{X} - \bar{Y})^2.
      \]
   l. Reject when large:
      \[
      \frac{m+n}{m+n-2} \frac{\hat{\sigma}^2 - \hat{\sigma}^2}{\hat{\sigma}^2} = \left( \frac{1}{m} + \frac{1}{n} \right)^{-1} (\bar{X} - \bar{Y})^2.
      \]
   m. Multiplying the test statistic by \( \frac{m+n}{m+n-2} \) to give right numerator:
      \[
      \frac{(\bar{X} - \bar{Y})^2}{(m+n-2) S_p^2}
      \]
      for \( S_p^2 = \sum_j (X_j - \bar{X})^2 + \sum_j (Y_j - \bar{Y})^2 \).
   n. Equivalently, 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. \( X_1, \ldots, X_m \sim N(\mu, \sigma^2) \)
   i. \( H_0: \sigma = \sigma_0 \) vs \( H_A: \sigma \neq \sigma_0 \)
   b. \( \hat{\mu} = \bar{X} \)
   c. \( \hat{\sigma} = \sigma_0, \hat{\sigma} = \sqrt{\sum_{j=1}^m (X_j - \bar{X})^2/m} \)
   d. Likelihood ratio statistic is
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.

Fig. 16: Example where smaller p-values do not indicate important evidence against null

Ordinate
\[ \gamma = 0.5 \quad \epsilon = 0.05 \]

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 \( \Theta \subset \Re \).
   b. Write \( g(x) = \exp(-s(x))I_A(x) \) for \( A \) free of \( \theta \).
2. Examples:
   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 exceeds this
   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.

Fig. 17: Example where smaller p-values do not indicate important evidence against null

Ordinate
\[ \gamma = 0.5 \quad \epsilon = 0.05 \]

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. $c(\mu) = \mu$, $T(x) = \sum_i x_i$, $b(\alpha) = n\mu^2/2$, $s(x) = (n/2)\log(2\pi) + \sum_i x_i^2/2$.

b. 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.

c. Choose $\eta \in c(\Theta) = \{c(\theta) | \theta \in \Theta\}$, and let $\theta = c^{-1}(\eta)$.

d. Suppose $X$ satisfies general exponential family for $\theta \in \Theta$.

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 $\eta \in H \supset c((0, \infty)) = (-\infty, 0)$.
   iii. Hence $T \sim \Gamma(\alpha, n)$.

b. $X_1, \ldots, X_n \sim \mathcal{N}(\mu, 1)$.
   i. $f_X(x; \mu) = \exp(\mu \sum_j x_j - \frac{1}{2} \sum_j x_j^2 - \frac{1}{2}n\mu^2 - \frac{n/2}{2}\log(2\pi))$ for $\mu \in \mathbb{R}$.
   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 $\eta \in H \supset \mathbb{R}$.
   iii. Hence $T \sim \mathcal{N}(n\mu, n)$.

c. $X_1, \ldots, X_n \sim \mathcal{B}(n, \pi)$ i.i.d.