4. Easy composite hypothesis:
   a. Composite alternative, with alternatives positioned to keep test the same:
      i. Form \( H_0 : \theta = \theta_0 \) vs. \( H_A : \theta > \theta_0 \).
      ii. Same test is most powerful for this compound alternative.
   b. Composite null, with nulls positioned to keep test the same:
      i. Form \( H_0 : \theta \leq \theta_0 \) vs. \( H_A : \theta > \theta_0 \).
      ii. Same test is most powerful for this compound null.
      iii. Test levels are generally bounded by value at the element of the null making it look most like the alternative: Worst case

5. Testing for composite hypotheses: Harder Compound Null
   a. Example: \( X_1, \ldots, X_n \sim N(\mu, \sigma^2) \)
   b. Test
      i. \( H_0 : \mu = 0 \), no restriction on \( \sigma \)
      ii. \( H_A : \mu \geq 0 \), no restriction on \( \sigma \)
   c. Complications:
      i. Test depends on value of \( \sigma \)
      ii. Worst case is unrealistically bad:
         - \( \sigma = +\infty \)
         - Never reject the null.
   d. Solution: A similar test:
      i. Definition: same probability of rejection for every value of nuisance parameter
      e. Construct via Conditioning:

WMS: 10.11

K. Testing for composite hypotheses: Formal Definition:
   1. Assume \( X \sim f(x; \theta) \) for some \( \theta \).
   2. Test \( \theta \in \Omega_0 \) vs. \( \theta \in \Omega_a \).
   a. Define \( \Lambda = (\max_{\theta \in \Omega_0} L(\theta; X))/(\max_{\theta \in \Omega} L(\theta; X)) \).
   b. Reject when small.
   c. For example, in binomial case, when testing \( H_0 : \pi = \pi_0 \) vs. \( H_A : \pi \neq \pi_0 \):
      i. Since the l.r.t.s against different alternatives take different forms, there won’t be a uniformly most powerful test.
   d. Hypotheses
      i. \( H_0 : \theta \in \Omega_0 \)
      ii. \( H_A : \theta \in \Omega_a \)
      iii. Let \( \Omega = \Omega_0 \cup \Omega_a \).
   e. Modify the likelihood ratio approach
      i. also called the likelihood ratio test
      ii. Substitute the \( \theta \) value in each hypothesis maximizing \( L(\theta) \)
         - \( \hat{\theta} = \arg\max_{\theta \in \Omega_0} L \)
         - \( \tilde{\theta} = \arg\max_{\theta \in \Omega} L \)
      iii. and then taking ratios and inverting.
      iv. reject \( H_0 \) when small.
   3. Example: \( X \sim \text{Bin}(m, \pi) \), \( H_0 : \pi = \pi_0 \), \( H_A : \pi \neq \pi_0 \).
      a. Then \( \hat{\pi} = \pi_0 \) maximizes \( L \) over \( \Omega_0 \).
      b. then \( \tilde{\pi} = X/m \) maximizes \( L \) over \( \Omega \),
      i. condition on sufficient statistic \( W \) for nuisance parameter at null.
      ii. By sufficiency data|\( W \) does not depend on nuisance parameter
      iii. Build rejection region \( R \) so that
         \[ P_0 \{ \text{data} \in R | W \} = \alpha \]
      iv. Then \( P_0 \{ \text{data} \in R \} = \mathbb{E}[P_0 \{ \text{data} \in R | W \} | W = \alpha] \)
   4. Compound Null Example:
      a. Consider observing \( X \sim \text{Bin}(m, \pi) \) and \( Y \sim \text{Bin}(n, \rho) \),
      b. and we wish to test \( H_0 : \pi = \rho \) vs. \( H_A : \pi \neq \rho \).
      c. Likelihood is
         \( L(\pi, \rho; X, Y) = \pi^X (1 - \pi)^{m - X} \rho^Y (1 - \rho)^{n - Y} \)
      d. Note that \( \hat{\pi} = X/m \) and \( \hat{\rho} = Y/n \) maximize the top
         while \( \hat{\pi} = \hat{\rho} = (X + Y)/(m + n) \) maximize the bottom.
      e. g.l.r.t. statistic is
         \( \Lambda = \frac{\pi^X (1 - \pi)^{m - X} \rho^Y (1 - \rho)^{n - Y}}{\hat{\pi}^X + \hat{\rho}^Y (1 - \hat{\pi})^{m - X} (1 - \hat{\rho})^{n - Y}} \)
      f. Present two special related problems not yet encountered:
6. Compound Null Example:

- Unlike in the simple null hypothesis case.
- In this case, the common unknown proportion is known as a nuisance parameter.
- Can’t express g.l.r.t. as a simple function of the one sufficient statistic.
- Hence distn is not easily obtained.

5. Way around these problems: Wilks’ Theorem

a. Theorem: Under certain regularity conditions
   \[-2 \log(\Lambda) \text{ approx. } \sim \chi^2_d \text{ distn.} \]
   \[X_i \sim f(x; \theta), \text{i.i.d., } \theta \in \Omega.\]
   ii. Regularity conditions include \(\Omega_0 \subset \Omega^\circ\)
   iii. \(d\) is difference in dimension between \(\Omega\) and \(\Omega_0\).
   iv. \(\chi^2_d\) same as \(\Gamma(d/2, 1/2)\) .

b. Simplification:
   \[\Omega = \Omega_0 \times \Omega^k\]
   \[\theta = (\mu, \omega) \text{ for } \mu \in \mathbb{R}^d \text{ interest parameter, } \omega \in \Omega^k \text{ nuisance parameter.}\]

c. Proof in the easy case, \(\omega\) known.
   \[-2 \log(\Lambda) \text{ can be expressed as} \]
   \[= -2(\ell(\mu, \omega) - \ell(\mu, \omega) + (-\mu)^T \ell'(\mu, \omega)(-\mu) - \ell(\mu, \omega))\]
   \[= -\mu^T \ell'(\mu, \omega) \mu\]
   ii. \(\mu\) a column vector with \(d\) components, \(\ell'(\mu, \omega)\) a row vector with \(d\) components.
   iii. \(o = \ell'(\mu, \omega) \approx \ell'(o, \omega) + \mu^T \ell''(o, \omega)\)

6. Compound Null Example:

- \(X_1, \ldots, X_n \sim N(\mu, \sigma^2)\)

b. Test
   i. \(H_0 : \mu = 0, \text{ no restriction on } \sigma \Rightarrow \Omega_0 = \{0\} \times (0, \infty)\).
   ii. \(H_A : \mu \neq 0, \text{ no restriction on } \sigma \Rightarrow \Omega_0 = (\mathbb{R} - \{0\}) \times (0, \infty)\).

c. Likelihood \(L(\mu, \sigma) = \exp \left( -\sum_{j=1}^n (X_j - \mu)^2 / (2\sigma^2) \right) (2\pi)^{n/2} \sigma^{-n}\)
   i. \(\hat{\mu} = 0, \hat{\sigma} = \sqrt{\sum_{j=1}^n X_j^2 / n}\)
   ii. \(\hat{\mu} = \bar{X}, \hat{\sigma} = \sqrt{\sum_{j=1}^n (X_j - \bar{X})^2 / n}\)

d. Likelihood ratio statistic:

\[\Lambda = \frac{\exp(-\sum_{j=1}^n (X_j - \hat{\mu})^2 / (2\hat{\sigma}^2)) (2\pi)^{-n/2} \hat{\sigma}^{-n}}{\exp(-\sum_{j=1}^n (X_j - \mu)^2 / (2\sigma^2)) (2\pi)^{-n/2} \sigma^{-n}}\]
\[= \frac{\exp \left(-n \sum_{j=1}^n \frac{(X_j - 0)^2}{2} \right) \left(\sum_{j=1}^n X_j^2 / n\right)^{-n/2}}{\exp \left(-n \sum_{j=1}^n \frac{(X_j - \mu)^2}{2} \right) \left(\sum_{j=1}^n (X_j - \bar{X})^2 / n\right)^{-n/2}}\]
\[= \left[\sum_{j=1}^n (X_j - \bar{X})^2 / \sum_{j=1}^n X_j^2 \right]^{-n/2}\]
\[\Lambda = \left[\frac{\sum_{j=1}^n (X_j - \bar{X})^2}{n \bar{X}^2 + \sum_{j=1}^n (X_j - \bar{X})^2} \right]^{-n/2}\]
\[= \left[1 + n \bar{X}^2 / \sum_{j=1}^n (X_j - \bar{X})^2 \right]^{-n/2}\]
\[= (1 + T^2 / (n - 1))^{-n/2}\]

for \(T = \sqrt{n} \bar{X} / \sqrt{\sum_{j=1}^n (X_j - \bar{X})^2 / (n - 1)}\)

e. So \(S = -2 \log(\Lambda) = -n \log(1 + T^2 / (n - 1)) \approx T^2\) for \(T^2 \sim \chi^2_{1,n-1}\).

f. For large \(n\),
   i. \(\chi^2_{1,n-1} \approx \chi^2_1\)
ii. $T_{1,n-1} \approx N(0,1)$
g. Hence reject $H_0$ when $|T| > c$ large
h. $T \sim T_{n-1}$
i. $c = T_{n-1,\alpha/2}$. 