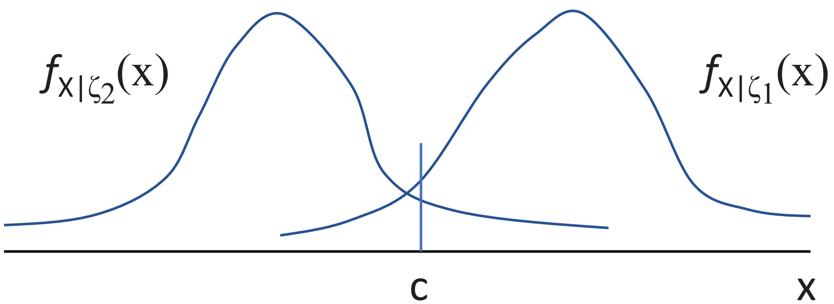
Statistics: Hypothesis Testing

- Bayesian Decision Theory
- Likelihood Ratio Test
- Neyman-Pearson Theorem
- Composite Hypotheses and GLRT
- Test for $\mu_1 \stackrel{?}{=} \mu_2$ (T-Test)
- Test for $\sigma_1^2 \stackrel{?}{=} \sigma_2^2$ (F-Test)
- Test for $\sigma^2 \stackrel{?}{=} \sigma_0^2$ (χ^2Test)
- Goodness of Fit Test (Pearson/χ^2 Test)
- Test for $P[E_1] \stackrel{?}{=} P[E_2]$ (Pearson/χ^2 Test)
- Run Test for $F_X \stackrel{?}{=} F_Y$
- Ranking Test for $F_X \stackrel{?}{=} F_Y$

Bayesian Decision Theory

- Underlying State: ζ_i with $P[\zeta = \zeta_i] = P_i$. e.g., $\zeta_1 =$ cancer, $\zeta_2 =$ benign
- Based on observation RV X, make decision of ζ denoted by $\hat{\zeta}$. e.g., $X =$ ratio of square of the boundary length of the nodule to the area of the nodule.
- Actions: a_i if $\hat{\zeta} = \zeta_i$. e.g., $a_1 =$ surgical operation, $a_2 =$ no operation
- Loss associated with a_i when the state is ζ_j: $l(a_i, \zeta_j)$. e.g., loss $= \#$ of years lost from a normal life span



$$l(a_2; \zeta_2) = 0$$
$$l(a_2; \zeta_1) = 35$$
$$l(a_1; \zeta_2) = 5$$
$$l(a_1; \zeta_1) = 5$$

The more irregular the edges of the nodule (the larger X), the more likely the nodule is a cancerous lesion.

If $X \in [c, \infty)$, $\hat{\zeta} = \zeta_1$ and take action a_1. Otherwise, a_2.

There is a value of c (to be determined) that minimizes the expected risk.

- Consider ζ_1 and ζ_2 with $X|\zeta_1 > X|\zeta_2$ most of the time.
- Decision function d: if $X \in [c, \infty)$, then $\hat{\zeta} = \zeta_1$, and a_1 is performed. Otherwise, $\hat{\zeta} = \zeta_2$, and a_2 is performed.
- Risk $R(d;\zeta)$: Conditional expectation of the loss when the underlying state is ζ and the decision rule is d.

$$R(d;\zeta_i) = l(a_1;\zeta_i)P[a_1|\zeta_i] + l(a_2;\zeta_i)P[a_2|\zeta_i],$$

$$\text{where } P[a_1|\zeta_2] = \int_c^{\infty} f_{X|\zeta_2}(x)dx \quad \text{and} \quad P[a_2|\zeta_1] = \int_{-\infty}^{c} f_{X|\zeta_1}(x)dx$$

- Expected Risk: $B(d) = R(d;\zeta_1)P[\zeta = \zeta_1] + R(d;\zeta_2)P[\zeta = \zeta_2]$
- Bayes Strategy: $B(d^*) = \min_d\{B(d)\}$, i.e., choose c to minimize $B(d)$
- Using $P[a_2|\zeta_1] = 1 - P[a_1|\zeta_1]$ and $P[a_2|\zeta_2] = 1 - P[a_1|\zeta_2]$,

$$B(d) = P_1\ l(a_2;\zeta_1) + P_2\ l(a_2;\zeta_2)$$
$$+ \int_c^{\infty} \left\{P_2 f_{X|\zeta_2}(x)[l(a_1;\zeta_2) - l(a_2;\zeta_2)] - P_1 f_{X|\zeta_1}(x)[l(a_2;\zeta_1) - l(a_1;\zeta_1)]\right\} dx$$

- Minimum of $B(d)$ is achieved when the integral yields a negative number with largest magnitude, i.e., when $c = c^*$ so that (c^*, ∞) includes all negative points and leaves out all positive points.

  For $X \in (c^*, \infty)$,

  $P_2 f_{X|\zeta_2}(x)[l(a_1;\zeta_2) - l(a_2;\zeta_2)] < P_1 f_{X|\zeta_1}(x)[l(a_2;\zeta_1) - l(a_1;\zeta_1)]$, or

$$\frac{f_{X|\zeta_1}(x)}{f_{X|\zeta_2}(x)} > \frac{[l(a_1;\zeta_2) - l(a_2;\zeta_2)]P_2}{[l(a_2;\zeta_1) - l(a_1;\zeta_1)]P_1} \triangleq k_b \quad \text{(Bayes threshold)}$$

- At $X = c^*$, $\frac{f_{X|\zeta_1}(c^*)}{f_{X|\zeta_2}(c^*)} = k_b$.

- Bayes Decision Rule: If $\frac{f_{X|\zeta_1}(x)}{f_{X|\zeta_2}(x)} > k_b$ (i.e., when $X \in (c^*, \infty)$), then $\hat{\zeta} = \zeta_1$ and take action a_1. Otherwise, $\hat{\zeta} = \zeta_2$ and take action a_2.

- For the case with n i.i.d observation RVs $\{X_1, \cdots, X_n\}$, the Bayes decision rule is

$$\text{if} \quad \frac{\prod_{i=1}^{n} f_{X_i|\zeta_1}(x_i)}{\prod_{i=1}^{n} f_{X_i|\zeta_2}(x_i)} > \frac{[l(a_1;\zeta_2) - l(a_2;\zeta_2)]P_2}{[l(a_2;\zeta_1) - l(a_1;\zeta_1)]P_1} \triangleq k_b,$$

$$\text{then} \quad \hat{\zeta} = \zeta_1, \quad \text{take action } a_1$$

$$\text{else } \hat{\zeta} = \zeta_2, \quad \text{take action } a_2.$$

- Problems with Bayes approach:

  i) *a priori* probabilities P_1 and P_2 are often unknown.

  ii) assigning a reasonable loss to an action is difficult.

Likelihood Ratio Test

- Not requiring *a priori* probabilities and loss functions,

$$\text{if} \quad \frac{\prod_{i=1}^{n} f_{X_i|\zeta_1}(x_i)}{\prod_{i=1}^{n} f_{X_i|\zeta_2}(x_i)} > k, \qquad \text{then} \quad \hat{\zeta} = \zeta_1$$

$$\text{else } \hat{\zeta} = \zeta_2.$$

- ζ_1 versus $\zeta_2 \quad \rightarrow \quad$ Hypothesis H_1 versus H_2
- The threshold k is based on some criteria, e.g.,

  P[rejecting a claim when the claim is true]$=\alpha$ or

  P[accepting the counter claim when the counterclaim is true] $= 1 - \beta$.
- Notation: $L(\zeta_1) \triangleq \prod_{i=1}^{n} f_{X_i|\zeta_1}(x_i)$, $L(\zeta_2) \triangleq \prod_{i=1}^{n} f_{X_i|\zeta_2}(x_i)$, and $\Lambda \triangleq L(\zeta_1)/L(\zeta_2)$
- Note: Every Bayes strategy leads to an LRT but not every LRT is the result of Bayes strategy

Terminologies for Hypothesis H_1 versus H_2:

- $\alpha \triangleq$ P[based on our test we decide H_2 is true $|$ H_1 is true]

  $\alpha =$ P[type I error] $=$ significance level of the test $=$ size of the test

- $\beta \triangleq$ P[based on our test we decide H_1 is true $|$ H_2 is true]

  $\beta =$ P[type II error]

  $1 - \beta =$ the power of the test

Summary of LRT for μ_1 versus μ_2

- $H_1 : \{X_i\} \sim$ i.i.d. $N(\mu_1, \sigma^2)$ versus $H_2 : \{X_i\} \sim$ i.i.d. $N(\mu_2, \sigma^2)$
- $\Lambda = \exp\{\frac{1}{2\sigma^2}\sum_{i=1}^n[(x_i-\mu_2)^2 - (x_i-\mu_1)^2]\}$
  $\ln(\Lambda) = \frac{1}{2\sigma^2}\sum_{i=1}^n(2x_i - \mu_1 - \mu_2)(\mu_1-\mu_2)$ $\ln(\Lambda) = \frac{n(\mu_1-\mu_2)}{\sigma^2}[\hat{\mu} - \frac{1}{2}(\mu_1+\mu_2)]$,
  where $\hat{\mu} \triangleq \frac{1}{n}\sum_{i=1}^n x_i$.
  LRT: Accept H_1 if $\ln(\Lambda) > \tau$, and Reject otherwise.
- Equivalently, if $\mu_1 > \mu_2$, accept H_1 if $\hat{\mu} > c$, and reject otherwise, where $c = \frac{\tau\sigma^2}{n(\mu_1-\mu_2)} + \frac{1}{2}(\mu_1+\mu_2)$.
- If $\mu_1 < \mu_2$, accept H_1 if $\hat{\mu} < c$, and reject otherwise.
- PDF of $\hat{\mu}$ given H_i is true is $f_{\hat{\mu}|H_i} = N(\mu_i, \sigma^2/n)$
- If $\mu_1 > \mu_2$, we have $\alpha = P[\hat{\mu} < c|H_1] = P[Z < \frac{c-\mu_1}{\sigma/\sqrt{n}}]$ where $Z \sim N(0,1)$, and $\frac{c-\mu_1}{\sigma/\sqrt{n}} = z_{[\alpha]}$. So, $c = \mu_1 + (\sigma/\sqrt{n})z_{[\alpha]}$.
  Power of the test $= P[\hat{\mu} < c|H_2] = P[Z < \frac{c-\mu_2}{\sigma/\sqrt{n}}]$.
- If $\mu_1 < \mu_2$, we have $\alpha = P[\hat{\mu} > c|H_1] = 1 - P[Z \leq \frac{c-\mu_1}{\sigma/\sqrt{n}}]$ and $\frac{c-\mu_1}{\sigma/\sqrt{n}} = z_{[1-\alpha]}$. So, $c = \mu_1 + (\sigma/\sqrt{n})z_{[1-\alpha]}$.
  Power of the test $= P[\hat{\mu} > c|H_2] = 1 - P[Z \leq \frac{c-\mu_2}{\sigma/\sqrt{n}}]$.

- Example 7.2-1 A food manufacturer claims its snack bar can reduce childhood obesity. To test this claim, a group of n children take the weight-controlling snack bar, while the second group of n children do not. After a month, the average weight for the first (second) group are 98 lbs (102 lbs) with a standard deviation of 5 lbs (5 lbs). We denote hypothesis H_1 to be the weight-controlling snack bar has no effect, and alternative H_2 to be the weight-controlling snack is helpful. How to decide if H_1 is true (with $X_i, \ i = 1, \ldots, n$ being i.i.d. RVs denoting the weights of the n children)?

Solution: Assuming the weights are Normally distributed,

$f(x_i|H_1) = \frac{1}{5\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x_i-\mu_1}{5}\right)^2\right], \quad f(x_i|H_2) = \frac{1}{5\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x_i-\mu_2}{5}\right)^2\right].$

$$\Lambda = \prod_{i=1}^{n} \frac{f(x_i, H_1)}{f(x_i, H_2)} = \exp\left(\frac{1}{2}\sum_{i=1}^{n}\left[\left(\frac{x_i-\mu_2}{5}\right)^2 - \left(\frac{x_i-\mu_1}{5}\right)^2\right]\right)$$

$$= K_n \ \exp\left(\frac{4n}{25}\ \hat{\mu}_X(n)\right), \quad \text{where} \ \ K_n = \text{constant}$$

and $\hat{\mu}(n) = (1/n) \ \sum_{i=1}^{n} x_i$.

The decision function is

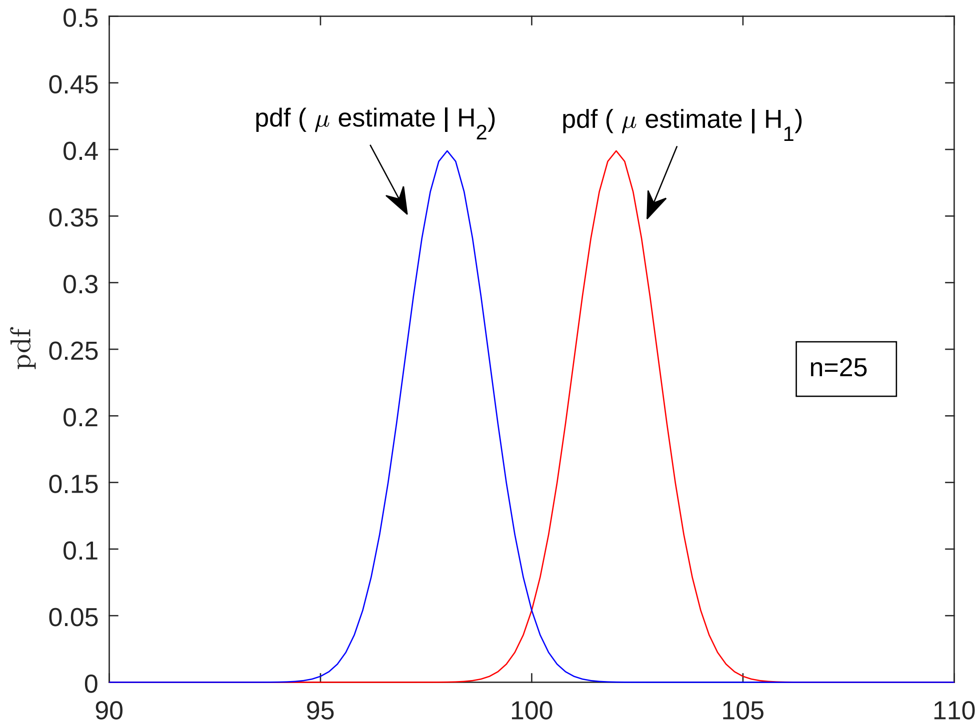
$$\text{if } K_n \exp\left(\frac{4n}{25}\,\hat{\mu}_X(n)\right) > k_n, \text{ accept } H_1. \quad \text{Otherwise, accept } H_2.$$

Simplifying the decision function by using logs, we have

$$\text{if } \hat{\mu}_X(n) > c_n, \text{ accept } H_1. \quad \text{Otherwise, accept } H_2.$$

c_n is the decision threshold to be decided.

$$f(x_i|H_1) = N(102, 25)$$
$$f(x_i|H_2) = N(98, 25)$$
$$f(\hat{\mu}|H_1) = N(102, 25/n)$$
$$f(\hat{\mu}|H_2) = N(98, 25/n)$$



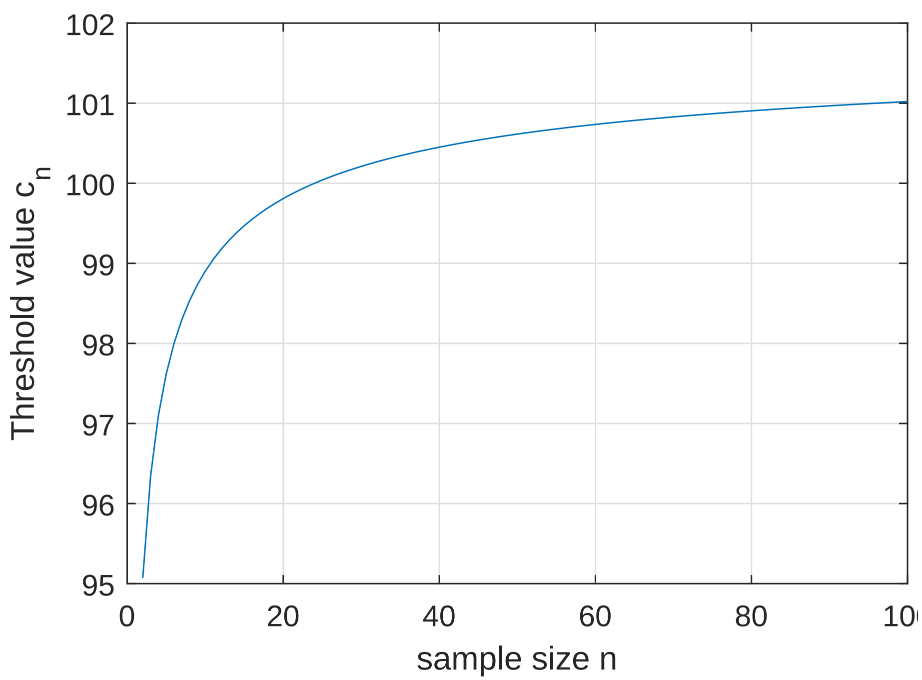
$$\text{Significance} = \alpha = P[\text{accepting } H_2|H_1] = F_{SN}\left(\frac{c_n - 102}{5/\sqrt{n}}\right) = F_{SN}(z_{[\alpha]})$$

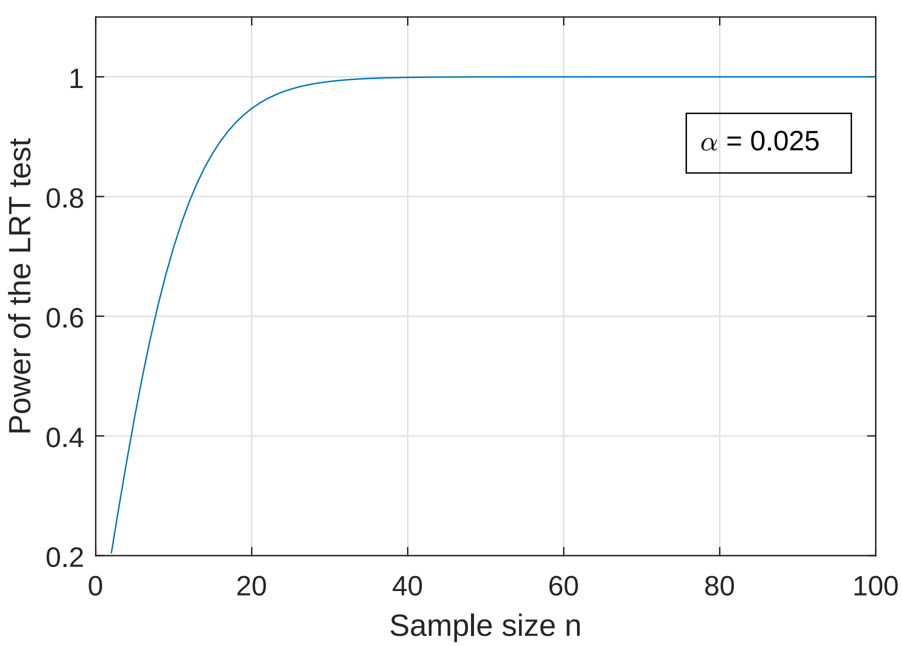
$$c_n = 102 + (5/\sqrt{n})\ z_{[\alpha]}$$

For $\alpha = 0.025$, we have $c_n = 102 - (9.8/\sqrt{n})$.

$$\text{Power of the test} = P[\text{accepting } H_2|H_2] = P[\hat{\mu} < c_n|H_2]$$

$$= F_{SN}\left(\frac{c_n - 98}{\sqrt{25/n}}\right) = F_{SN}(0.8\sqrt{n} - 1.96)$$





Neyman-Pearson Theorem (NPT)

- Theorem:

  Denote the set of points in the critical region by R_k (i.e., the region of outcomes where we reject the hypothesis H_1).

  Denote the significance of the test as α, meaning $P[\text{accept } H_2|H_1 \text{ is true}] \leq \alpha$.

  Then, R_k maximizes the power of the test $\mathrm{P} \triangleq 1-\beta$ if it satisfies

$$\Lambda \triangleq \frac{\prod_{i=1}^{n} f_{X_i|\zeta_1}(x_i)}{\prod_{i=1}^{n} f_{X_i|\zeta_2}(x_i)} < k$$

  for some fixed number k, which determines R_k.

- NPT says that LRT subject to the constraint of being at significance α is the most powerful test.

  (Relationship between R_k, k, and α is not explicitly stated by NPT.)

- Example 7.2-3: It is claimed that feeding chicken with a new product 'Eggrow' will cause larger laid eggs. With ordinary feed, the average weight of the laid eggs is $\mu_2 = 60$g, with a standard deviation of 4g. 25 chickens fed on 'Eggrow' produce eggs whose average weight is $\mu_1 = 62$g, with a standard deviation of 4g. Let hypothesis be $H_1: \ \mu = \mu_1$ and the alternative be $H_2: \ \mu = \mu_2$. The significance level is 0.05. Find the decision threshold c_n and the critical region R_k.

  Solution: According to NPT, we have

$$\Lambda = \frac{\prod_{i=1}^{n}(2\pi 16)^{-1/2}\exp\left(-\frac{1}{2}\left[\frac{X_i-62}{4}\right]^2\right)}{\prod_{i=1}^{n}(2\pi 16)^{-1/2}\exp\left(-\frac{1}{2}\left[\frac{X_i-60}{4}\right]^2\right)} = \exp\left(\frac{n\hat{\mu}}{8} + \frac{n}{32}(60^2 - 62^2)\right)$$

  where $\hat{\mu} = (1/n)\ \sum_{i=1}^{n} X_i$ and $n = 25$.

  Taking logs and simplifying yields the test

  if $\hat{\mu} > c_n$, accept H_1; Otherwise, accept H_2;

  where c_n is determined by solving $\alpha = 0.05 = P[\text{accept } H_2|H_1]$.

  Since $f_{\hat{\mu}|H_1} = N(62, 16/25)$, $0.05 = P[\hat{\mu} \leq c_n|H_1] = F_{\mathrm{SN}}(\frac{c_n-62}{0.8})$.

  We find $c_n = 0.8z_{[0.05]} + 62 = 60.7$ and $R_k = (0,\ 60.7)$.

  The test is most powerful and $\mathrm{Power} = P[\hat{\mu} \leq c_n|H_2] = F_{\mathrm{SN}}(\frac{60.7-60}{0.8}) \approx 0.81$.

Composite Hypotheses

- **Example of composite hypotheses:** Suppose the parameter of interest is the mean μ, and consider $H_1 : \mu = \mu_0$ versus $H_2 : \mu \neq \mu_0$. While H_1 is a simple hypothesis, H_2 is a composite hypothesis since the likelihood function associated with H_2 is not well defined (and a search for the optimum value of μ may be needed).
- Example 7.3-1 (Testing $H_1 : \mu = \mu_1$ versus $H_2 : \mu < \mu_1$ without needing the search for optimum μ for H_2). We assume a Normal population with mean μ and variance σ^2. How to make the test with significant level $\alpha = 0.01$?

Solution: We reduce the problem of testing H_1 versus H_2 to a slightly modified problem, that is testing H_1 versus $H_2' : \mu = \mu_2 < \mu_1$. Then

$$\Lambda = \frac{\prod_{i=1}^{n}(2\pi\sigma^2)^{-1/2}\exp\left(-\frac{1}{2\sigma^2}(X_i-\mu_1)^2\right)}{\prod_{i=1}^{n}(2\pi\sigma^2)^{-1/2}\exp\left(-\frac{1}{2\sigma^2}(X_i-\mu_2)^2\right)}$$

$$= \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(X_i-\mu_1)^2 - \sum_{i=1}^{n}(X_i-\mu_2)^2\right)\right) < k$$

is the LRT for the critical region for H_1.

Next, taking logs and simplifying, we obtain the test: if $\hat{\mu} < c_n$, reject H_1.

To find c_n with significant level $\alpha = 0.01$ and the pdf $f_{\hat{\mu}|H_1}(w) = N(\mu_1; \sigma^2/n)$, we solve

$P[\hat{\mu} < c_n | H_1] = 0.01 \quad \Rightarrow \quad P[Z < \frac{c_n - \mu_1}{\sigma/\sqrt{n}}] = 0.01$

where $Z = \frac{\hat{\mu}-\mu_1}{\sigma/\sqrt{n}}$ is with pdf $N(0,1)$.

Then, we obtain $\quad \frac{c_n-\mu_1}{\sigma/\sqrt{n}} = z_{[0.01]} = -2.32 \quad$ and $\quad c_n = \mu_1 - 2.32\sigma/\sqrt{n}$.

Thus, the test is:

Reject H_1 if $\hat{\mu} < \mu_1 - 2.32\sigma/\sqrt{n}$.

(Note that we never had to specify an actual value for μ_2.)

Generalized Likelihood Ratio Test (GLRT)

- Useful for solving composite hypotheses problems
- For $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_k)$ and the global k-dimensional parameter space Θ, consider $H_1 : \boldsymbol{\theta} \in \Theta_1$ versus $H_2 : \boldsymbol{\theta} \notin \Theta_1$ where Θ_1 is a subset of Θ.
- GLR is

$$\Lambda \triangleq \frac{L_{\mathrm{LM}}(\boldsymbol{\theta}^*)}{L_{\mathrm{GM}}(\boldsymbol{\theta}^\dagger)}$$

$$\text{where} \quad L_{\mathrm{LM}}(\boldsymbol{\theta}^*) \triangleq \max_{\boldsymbol{\theta} \in \Theta_1} L(\boldsymbol{\theta}) \quad \text{and} \quad L_{\mathrm{GM}}(\boldsymbol{\theta}^\dagger) \triangleq \max_{\boldsymbol{\theta} \in \Theta} L(\boldsymbol{\theta})$$

- GLRT is

Reject H_1 if $\Lambda < c$.

- Example 7.3-2, 7.3-3
  Testing $H_1: \ \mu = \mu_1$ versus $H_2: \ \mu \neq \mu_1$ when X is Normal and σ^2 is known, with n observations. What will be the critical region ?

  Solution: The likelihood function is

  $$L(\mu) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2\right).$$

  Then $L_{LM}(\mu^*) = L(\mu_1)$.

  To get $L_{GM}(\mu^\dagger)$, solving $\frac{dL(\mu)}{d\mu} = 0$ yields $\mu^\dagger = \hat{\mu}$. Thus, $L_{GM}(\mu^\dagger) = L(\hat{\mu})$.

  Then the GLR (after simplification) is

  $$\Lambda = \frac{L(\mu_1)}{L(\hat{\mu})} = \exp\left(-\frac{n}{2\sigma^2}(\hat{\mu} - \mu_1)^2\right).$$

  The critical region is defined by $0 < \Lambda < c$.

  Taking log and simplifying gives the critical region for $\hat{\mu}$ as

  $$\hat{\mu} > \mu_1 + (2\sigma^2 \ln(1/c)/n)^{1/2}, \qquad \hat{\mu} < \mu_1 - (2\sigma^2 \ln(1/c)/n)^{1/2},$$

  where c is determined by the significance level α. (see next)

Assume that $\mu_1 = 5,\ \sigma^2 = 4,\ n = 15,\ \alpha = 0.05.$

Define $W \triangleq -2\ln\Lambda = \left(\frac{\hat{\mu}-\mu_1}{\sigma/\sqrt{n}}\right)^2.$

Then, given H_1 is true, W is χ^2 with one degree of freedom, i.e., with pdf $f_{\chi^2}(w; 1)$. The critical region $0 < \Lambda < c$ in terms of the test statistic W is

$-2\ln c < W < \infty.$

For a significance level α, we have $P[W > -2\ln c] = \alpha = 1 - F_{\chi^2}(-2\ln c; 1).$

So, from the CDF of χ^2, we obtain $-2\ln c = x_{[1-\alpha]}.$

For $\alpha = 0.05$, we have $-2\ln c = x_{[1-\alpha]} = x_{[0.95]} = 3.84.$

Therefore, the critical region for $\hat{\mu}$ is

$$(-\infty,\ \mu_1 - (2\sigma^2\ln(1/c)/n)^{1/2}) \cup (\mu_1 + (2\sigma^2\ln(1/c)/n)^{1/2},\ \infty)$$

$$= (-\infty, 3.99) \cup (6.01, \infty).$$

- Example 7.3-2, 7.3-3 (Alternative Approach)
  Testing $H_1 : \ \mu = \mu_1$ versus $H_2 : \ \mu \neq \mu_1$ when X is Normal and σ^2 is known, with n observations. What will be the critical region ?

  Solution: The GLR (after simplification) is
  $$\Lambda = \frac{L(\mu_1)}{L(\hat{\mu})} = \exp\left(-\frac{n}{2\sigma^2}(\hat{\mu} - \mu_1)^2\right).$$

  The critical region is defined by $0 < \Lambda < c$, or $(\hat{\mu} - \mu_1)^2 > \tau^2$ where $\tau > 0$.

  The critical region for the test statistic $\hat{\mu}$ is
  $$\{\hat{\mu} < \mu_1 - \tau\} \ \cup \ \{\hat{\mu} > \mu_1 + \tau\}.$$

  As the pdf of $\hat{\mu}$ given H_1 is true is $N(\mu_1, \sigma^2/n)$, for a significance level α, τ is obtained as follows:
  $$P[\hat{\mu} > \mu_1 + \tau | H_1] = 0.5\alpha$$
  $$\tau/(\sigma/\sqrt{n}) = z_{[1-0.5\alpha]} \quad \Rightarrow \quad \tau = z_{[1-0.5\alpha]}(\sigma/\sqrt{n}).$$
  Assume that $\mu_1 = 5, \ \sigma^2 = 4, \ n = 15, \ \alpha = 0.05.$
  Then, $\tau = z_{[1-0.5\alpha]}(\sigma/\sqrt{n}) = z_{[0.975]} \cdot (2/\sqrt{15}) \approx 1.01,$
  which yields $\mu_1 - \tau = 3.99$ and $\mu_1 + \tau = 6.01.$

  Thus, the critical region for the test statistic $\hat{\mu}$ is $(-\infty, 3.99) \cup (6.01, \infty)$.

Testing Equality of Means of Two Normal Populations (T-Test)

Problem Statement

- m i.i.d. samples $\{X_{1i}, i = 1, \cdots, m\}$ from population $P1$ with $X_1 : N(\mu_1, \sigma_1^2)$
- n i.i.d. samples $\{X_{2i}, i = 1, \cdots, n\}$ from population $P2$ with $X_2 : N(\mu_2, \sigma_2^2)$
- $\sigma_1^2 = \sigma_2^2 = \sigma^2$ and $E[(X_{1i} - \mu_1)(X_{2j} - \mu_2)] = 0$ for all i, j
- $H_1 : \mu_1 = \mu_2$ versus $H_2 : \mu_1 \neq \mu_2$, $(\sigma^2$, μ_i: unknown)

Testing Equality of Means of Two Normal Populations (T-Test)

Test Procedure

- $\hat{\mu}_1 \triangleq (1/m)\sum_{i=1}^{m} X_{1i}$ and $\hat{\mu}_2 \triangleq (1/n)\sum_{j=1}^{n} X_{2j}$
- $T^2_{m+n-2} \triangleq \frac{(m+n-2)mn}{m+n} \frac{(\hat{\mu}_1-\hat{\mu}_2)^2}{\sum_{i=1}^{m}(X_{1i}-\hat{\mu}_1)^2+\sum_{j=1}^{n}(X_{2j}-\hat{\mu}_2)^2}$
- GLR is $\Lambda = \left[1 + \frac{T^2_{m+n-2}}{m+n-2}\right]^{-(m+n)/2}$ (monotonically decreasing function of T^2_{m+n-2})
- GLRT: Reject H_1 if $T^2_{m+n-2} > t_c^2$ (i.e., $0 < \Lambda < \lambda_c$).
  Note: $\{T^2_{m+n-2} > t_c^2\} = \{T_{m+n-2} < -t_c\} \cup \{T_{m+n-2} > t_c\}$
- Under the constraint of type I error probability α, with equal probability for the two events, $P[T_{m+n-2} > t_c] = 0.5\alpha \Rightarrow F_T(t_c; m+n-2) = 1 - 0.5\alpha$. Thus, $t_c = t_{[1-0.5\alpha]}$.

Detailed Development of T-Test

- Parameter space for H_1 is $\Theta_1 = (\mu, \sigma^2)$
- Global parameter space is $\Theta = (\mu_1, \mu_2, \sigma^2)$
- Likelihood function is

$$\begin{aligned} L(\mu_1, \mu_2, \sigma) &= f_{\mathbf{x}_1 \mathbf{x}_2}(\{X_{1i}\}, \{X_{2j}\}) \\ &= (\frac{1}{2\pi\sigma^2})^{(m+n)/2} \exp\left(-\frac{\sum_{i=1}^{m}(X_{1i} - \mu_1)^2 + \sum_{j=1}^{n}(X_{2j} - \mu_2)^2}{2\sigma^2}\right) \end{aligned}$$

- To compute L_{GM}, we first find $\hat{\mu}_1^\dagger$, $\hat{\mu}_2^\dagger$ and $\hat{\sigma}^{2\dagger}$ as follows:

$$\frac{\partial L(\mu_1, \mu_2, \sigma^2)}{\partial \mu_1} = 0 \quad \Rightarrow \quad \hat{\mu}_1^\dagger = \frac{1}{m}\sum_{i=1}^{m} X_{1i} = \hat{\mu}_1$$

$$\frac{\partial L(\mu_1, \mu_2, \sigma^2)}{\partial \mu_2} = 0 \quad \Rightarrow \quad \hat{\mu}_2^\dagger = \frac{1}{n}\sum_{j=1}^{m} X_{2j} = \hat{\mu}_2$$

$$\frac{\partial L(\mu_1, \mu_2, \sigma^2)}{\partial \sigma^2} = 0 \quad \Rightarrow \quad \hat{\sigma}^{2\dagger} = \left(\sum_{i=1}^{m}(X_{1i} - \hat{\mu}_1)^2 + \sum_{j=1}^{n}(X_{2j} - \hat{\mu}_2)^2\right)/(m+n) = \hat{\sigma}^2$$

- $L_{\mathrm{GM}} = L(\hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}^2)$

$$L_{\mathrm{GM}} = \left(\frac{m+n}{2\pi(\sum_{i=1}^{m}(X_{1i} - \hat{\mu}_1)^2 + \sum_{j=1}^{n}(X_{2j} - \hat{\mu}_2)^2)} \right)^{(m+n)/2} \exp\left(-\frac{m+n}{2}\right)$$

- To compute L_{LM} with parameter space Θ_1, we first find $\hat{\mu}^*$ and $\hat{\sigma}^{2*}$ as

$$\frac{\partial L(\mu, \mu, \sigma^2)}{\partial \mu} = 0 \quad \Rightarrow \quad \hat{\mu}^* = \frac{1}{m+n}\left(\sum_{i=1}^{m} X_{1i} + \sum_{j=1}^{m} X_{2j}\right) = \frac{m}{m+n}\hat{\mu}_1 + \frac{n}{m+n}\hat{\mu}_2$$

$$\frac{\partial L(\mu, \mu, \sigma^2)}{\partial \sigma^2} = 0 \quad \Rightarrow \quad \hat{\sigma}^{2*} = \frac{\sum_{i=1}^{m}(X_{1i} - \hat{\mu}_1)^2 + \sum_{j=1}^{n}(X_{2j} - \hat{\mu}_2)^2 + \frac{mn}{m+n}(\hat{\mu}_1 - \hat{\mu}_2)^2}{m+n}$$

- $L_{\mathrm{LM}} = L(\hat{\mu}^*, \hat{\mu}^*, \hat{\sigma}^{2*})$

$$L_{\mathrm{LM}} = \left[\frac{(m+n)}{2\pi\left(\sum_{i=1}^{m}(X_{1i} - \hat{\mu}_1)^2 + \sum_{j=1}^{n}(X_{2j} - \hat{\mu}_2)^2 + \frac{mn}{m+n}(\hat{\mu}_1 - \hat{\mu}_2)^2\right)} \right]^{(m+n)/2} e^{-\frac{m+n}{2}}$$

- GLR is

$$\Lambda = \frac{L_{\mathrm{LM}}}{L_{\mathrm{GM}}} = \left[1 + \frac{\frac{mn}{m+n}(\hat{\mu}_1 - \hat{\mu}_2)^2}{\sum_{i=1}^{m}(X_{1i} - \hat{\mu}_1)^2 + \sum_{j=1}^{n}(X_{2j} - \hat{\mu}_2)^2}\right]^{-(m+n)/2}$$

- $Z \triangleq \frac{\hat{\mu}_1 - \hat{\mu}_2}{\sigma\sqrt{(m+n)/(mn)}}$ which is $N(0,1)$.
- $W_{m+n-2} \triangleq \sum_{i=1}^{m} \frac{(X_{1i}-\hat{\mu}_1)^2}{\sigma^2} + \sum_{j=1}^{n} \frac{(X_{2j}-\hat{\mu}_2)^2}{\sigma^2}$ which is Chi-square with DOF $m+n-2$.
- $T_{m+n-2} \triangleq \frac{Z\sqrt{m+n-2}}{\sqrt{W_{m+n-2}}}$ which is t-distributed with DOF $m+n-2$.
- GLR is $\Lambda = \left[1 + \frac{T_{m+n-2}^2}{m+n-2}\right]^{-(m+n)/2}$

- Example 7.3-8. 15 samples are generated from a $N(0,2)$ population (P1), $S_1 = \{2.21, 0.83, 0.393, 0.975, 0.195, 0.069, 1.91, 1.44, 3.98, 0.98,$ $2.84, 1.56, 0.4, 1.08, 0.116\}$,
  $\hat{\mu}_1' = -0.258,\ m = 15,\ \sum_{i=1}^{15} = (X_{1i}' - \hat{\mu}_1')^2 = 40.48.$
  15 samples are generated from a $N(2,2)$ population (P2),
  $S_2 = \{1.28, 0.258, 0.947, 5.85, 1.56, 1.48, 1.95, 3.22, 1.41, 1.84,$ $2.69, 3.94, 2.04, 2.08, 1.44\}$,
  $\hat{\mu}_2' = 1.801,\ n = 15,\ \sum_{i=1}^{15} = (X_{2i}' - \hat{\mu}_2')^2 = 45.66.$
  Denote $H_1 : \mu_1 = \mu_2$. With $\alpha = P(\text{reject } H_1 | H_1 \text{ true}) = 0.01$, should we accept the hypothesis H_1 ?

  Solution:
  We have
  $$T^2 \triangleq (m+n-2)\frac{mn(m+n)^{-1}(\hat{\mu}_1 - \hat{\mu}_2)^2}{\sum_{i=1}^{m}(X_{1i}' - \hat{\mu}_1')^2 + \sum_{i=1}^{n}(X_{2i}' - \hat{\mu}_2')^2} = 10.34.$$
  We find $F_T(t_{[1-0.5\alpha]}; 28) = 0.995$ (DOF $= 15 + 15 - 2 = 28$).

  From the t-distribution table, we find $t_{[1-0.5\alpha]} = 2.763$.

  Since $T^2 > t^2_{[1-0.5\alpha]}$, we reject the hypothesis.

Testing Equality of Variances of Two Normal Populations (F-Test)

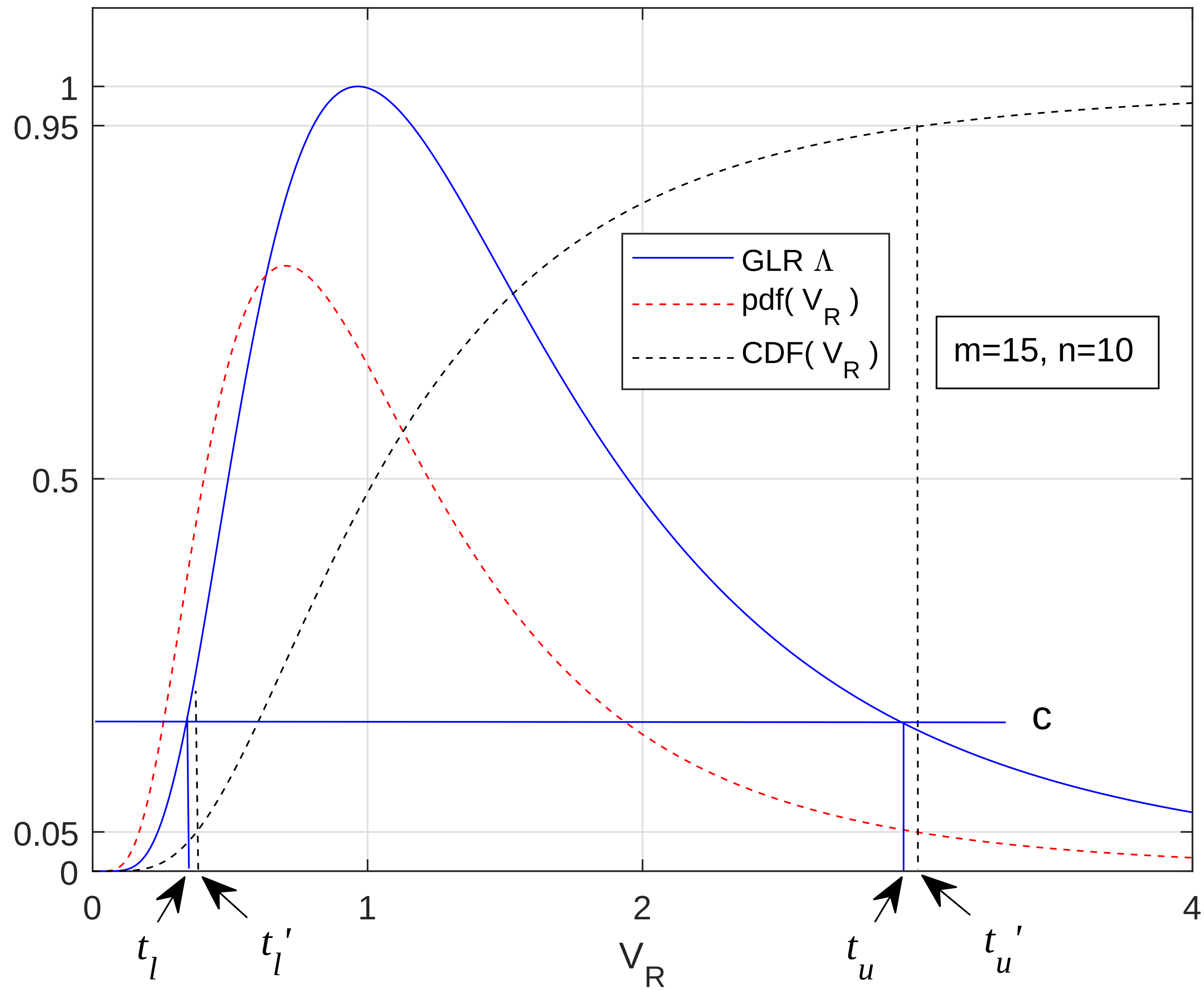
Problem Statement

- m i.i.d. samples $\{X_{1i}, i = 1, \cdots, m\}$ from population $P1$ with $X_1 : N(\mu_1, \sigma_1^2)$
- n i.i.d. samples $\{X_{2i}, i = 1, \cdots, n\}$ from population $P2$ with $X_2 : N(\mu_2, \sigma_2^2)$
- $E[(X_{1i} - \mu_1)(X_{2j} - \mu_2)] = 0$ for all i, j
- $H_1 : \sigma_1^2 = \sigma_2^2 = \sigma^2$ versus $H_2 : \sigma_1^2 \neq \sigma_2^2$, ($\mu_i, \sigma_i^2$: unknown)

Testing Equality of Variances of Two Normal Populations (F-Test)

Test Procedure

- $\hat{\mu}_1 \triangleq (1/m)\sum_{i=1}^{m} X_{1i}$ and $\hat{\mu}_2 \triangleq (1/n)\sum_{i=1}^{n} X_{2i}$
- Unbiased variance estimate ratio $V_R \triangleq \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} = \frac{(n-1)\sum_{i=1}^{m}(X_{1i}-\hat{\mu}_1)^2}{(m-1)\sum_{i=1}^{n}(X_{2i}-\hat{\mu}_2)^2}$
- GLR $\Lambda \triangleq \frac{L_{\mathrm{LM}}}{L_{\mathrm{GM}}} = \frac{(m+n)^{(m+n)/2}}{m^{m/2}n^{n/2}} \frac{[V_R(m-1)/(n-1)]^{m/2}}{[1+V_R(m-1)/(n-1)]^{(m+n)/2}}$
- When H_1 is true, $V_R = F_{m-1,n-1}$ where $F_{m-1,n-1}$ is an RV with F-distribution with $m-1$ and $n-1$ DOFs.
- Rejection/critical region of H_1: $\{0 < \Lambda < c\}$, equivalently $\{0 < V_R < t_l\} \cup \{t_u < V_R\}$. (See Fig.)
- Given a significance level α, find t_l and t_u such that $P[0 < V_R < t_l] + P[t_u < V_R] = \alpha$. Or for simplicity without much loss of accuracy, find t_l' and t_u' such that $P[0 < V_R < t_l'] = P[t_u' < V_R] = \alpha/2$ (see Fig.), i.e., $t_l' = x_{[0.5\alpha]}$ and $t_u' = x_{[1-0.5\alpha]}$ obtained from the $100a$ percentile point of $F_{m-1,n-1}$, namely $F_F(x_a; m-1; n-1) = a$.
- Reject H_1 if $\{0 < V_R < t_l\}$ or $\{t_u < V_R < \infty\}$; or if $\{0 < V_R < x_{[0.5\alpha]}\}$ or $\{x_{[1-0.5\alpha]} < V_R\}$.



1
0.95
0.5
0.05
0
0
1
2
4
V_R
GLR Λ
pdf(V_R)
CDF(V_R)
m=15, n=10
c
t_l
t_l'
t_u
t_u'

Detailed Development of F-Test

- Parameter space for H_1 is $\Theta_1 = \{\mu_1, \mu_2, \sigma^2\}$
- Parameter space for H_2 is $\Theta = \{\mu_1, \mu_2, \sigma_1^2, \sigma_2^2\}$
- Likelihood function is

$$L(\Theta) = (\frac{1}{2\pi\sigma_1^2})^{m/2} \exp\left(-\frac{\sum_{i=1}^m (X_{1i} - \mu_1)^2}{2\sigma_1^2}\right) (\frac{1}{2\pi\sigma_2^2})^{n/2} \exp\left(-\frac{\sum_{i=1}^n (X_{2i} - \mu_2)^2}{2\sigma_2^2}\right)$$

- For H_1,

$$L(\Theta_1) = (\frac{1}{2\pi\sigma^2})^{(m+n)/2} \exp\left(-\frac{\sum_{i=1}^m (X_{1i} - \mu_1)^2 + \sum_{j=1}^n (X_{2j} - \mu_2)^2}{2\sigma^2}\right)$$

and solving $\frac{\partial \ln L(\Theta_1)}{\partial \mu_1} = 0$, $\frac{\partial \ln L(\Theta_1)}{\partial \mu_2} = 0$, $\frac{\partial \ln L(\Theta_1)}{\partial \sigma^2} = 0$ yield

$$\hat{\mu}_1^* = \frac{1}{m}\sum_{i=1}^m X_{1i} = \hat{\mu}_1, \qquad \hat{\mu}_2^* = \frac{1}{n}\sum_{j=1}^m X_{2j} = \hat{\mu}_2$$

$$\hat{\sigma}^{2*} = \left(\sum_{i=1}^m (X_{1i} - \hat{\mu}_1)^2 + \sum_{j=1}^n (X_{2j} - \hat{\mu}_2)^2\right) / (m+n)$$

- Substituting μ_1^*, μ_2^*, σ^{2*} into $L(\Theta_1)$ gives

$$L_{\mathrm{LM}} = \left(\frac{m+n}{2\pi[\sum_{i=1}^m (X_{1i}-\hat{\mu}_1)^2 + \sum_{j=1}^n (X_{2j}-\hat{\mu}_2)^2]}\right)^{(m+n)/2} \exp\left(-\frac{m+n}{2}\right)$$

- To compute L_{GM}, solve $\frac{\partial \ln L(\Theta)}{\partial \mu_1} = 0$, $\frac{\partial \ln L(\Theta)}{\partial \mu_2} = 0$, $\frac{\partial \ln L(\Theta)}{\partial \sigma_1^2} = 0$, $\frac{\partial \ln L(\Theta)}{\partial \sigma_2^2} = 0$ to obtain

$$\hat{\mu}_1^\dagger = \frac{1}{m}\sum_{i=1}^m X_{1i} = \hat{\mu}_1, \quad \hat{\mu}_2^\dagger = \frac{1}{n}\sum_{j=1}^m X_{2j} = \hat{\mu}_2$$

$$\hat{\sigma}_1^{2\dagger} = \frac{1}{m}\sum_{i=1}^m (X_{1i}-\hat{\mu}_1)^2 = \hat{\sigma}_{1,\mathrm{ML}}^2, \quad \hat{\sigma}_2^{2\dagger} = \frac{1}{n}\sum_{i=1}^n (X_{2i}-\hat{\mu}_2)^2 = \hat{\sigma}_{2,\mathrm{ML}}^2$$

- Substituting $\mu_1^\dagger$, $\mu_2^\dagger$, $\sigma_1^{2\dagger}$, $\sigma_2^{2\dagger}$ into $L(\Theta)$ gives

$$L_{\mathrm{GM}} = \frac{1}{(2\pi)^{(m+n)/2}}\left[\frac{1}{m}\sum_{i=1}^m (X_{1i}-\hat{\mu}_1)^2\right]^{-m/2}\left[\frac{1}{n}\sum_{i=1}^n (X_{2i}-\hat{\mu}_2)^2\right]^{-n/2} \exp\left(-\frac{m+n}{2}\right).$$

- $\Lambda = \frac{L_{\mathrm{LM}}}{L_{\mathrm{GM}}}$

$$\Lambda = \frac{\left(\frac{m+n}{\sum_{i=1}^{m}(X_{1i}-\hat{\mu}_1)^2+\sum_{i=1}^{n}(X_{2i}-\hat{\mu}_2)^2}\right)^{(m+n)/2}}{\left(\frac{m}{\sum_{i=1}^{m}(X_{1i}-\hat{\mu}_1)^2}\right)^{m/2}\left(\frac{n}{\sum_{i=1}^{n}(X_{2i}-\hat{\mu}_2)^2}\right)^{n/2}}$$

- Using unbiased variance estimates $\hat{\sigma}_1^2 = \sum_{i=1}^{m}(X_{1i} - \hat{\mu}_1)^2/(m-1)$, $\hat{\sigma}_2^2 = \sum_{i=1}^{n}(X_{2i} - \hat{\mu}_2)^2/(n-1)$ and defining $V_R \triangleq \hat{\sigma}_1^2/\hat{\sigma}_2^2$, and after some simplification we obtain

$$\Lambda = \frac{(m+n)^{(m+n)/2}}{m^{m/2}n^{n/2}} \frac{[V_R\ (m-1)/(n-1)]^{m/2}}{[1+V_R\ (m-1)/(n-1)]^{(m+n)/2}}$$

- Note that $F_F(x_{[a]}; n_1, n_2) = a$ and $F_F(x_{[1-a]}; n_2, n_1) = 1-a$ are related by $x_{[a]} = \frac{1}{x_{[1-a]}}$.

- Example 7.3-9

  Test the hypothesis that the variances of two populations are the same at the significance level of $\alpha = 0.05$. The two sets of samples are:
  N(0,1) : { 0.436, -1 .06, -1 .11, 0.46, 0.491, -1 .05, 0.502, 0.598, 1.61, -0 .981, -0 .021, 0.253, -1 .24, 0.059, 2.12 }, with $\hat{\mu}_1 = 0.074$, $\hat{\sigma}_1^2 = 1.02$.
  $N(0,4)$: $\{0.634, 0.0818, -1.32, 2.96, 3.11, 3.13, 2.62, -1.96, 0.85, -6.51, -3.39, 4.25, -1.08, 3.42, 2.72\}$, with $\hat{\mu}_2 = 0.54, \hat{\sigma}_2^2 = 9.25$.

  Solution:

  We compute

  $$V_R = \frac{(15-1)\sum_{i=1}^{15}(X_{1i}-0.54)^2}{(15-1)\sum_{i=1}^{15}(X_{2i}-0.074)^2} = \frac{9.25}{1.02} = 9.06.$$

  For $\alpha = 0.05$, using 'equal area' approach, we seek $x_{[0.025]}$ and $x_{[0.975]}$ such that $F_F(x_{[0.025]}; 14; 14) = 0.025$ and $F_F(x_{[0.975]}; 14; 14) = 0.975$.

  We obtain $x_{[0.025]} = 0.34$ and $x_{[0.975]} = 2.98$, which give the acceptance region $(0.34, 2.98)$.

  Since $V_R \notin (0.34, 2.98)$, the hypothesis is rejected.

F-Test for Testing If Multiple Groups Are Statistically Alike

- k groups; group i has n_i i.i.d. samples with variance σ_i^2; $\sum_{i=1}^{k} n_i = n$
- H_1: Statistics (e.g., mean and variance) of k groups are the same
- $Y_{ij} = j$th sample of group i
- $Z_i \triangleq \sum_{j=1}^{n_i} Y_{ij}/n_i$ and $\hat{\mu}_Z \triangleq \sum_{i=1}^{k} Z_i/k$. Then $\sigma_{Z_i}^2 = \sigma_{Y_i}^2/n_i$.
- If $n_i \gg 1$ and $\{Z_i\}$ are i.i.d. (i.e., under H_1), then $V \triangleq \sum_{i=1}^{k} \left(\frac{Z_i - \hat{\mu}_Z}{\sigma_{Z_i}}\right)^2$ is χ^2_{k-1}.
  Note $\sum_{i=1}^{k}(Z_i - \hat{\mu}_Z)^2$ is sometimes called inter-group variability.
- If $n_i \gg 1$ and $\{Z_i\}$ are independent, $W \triangleq \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(\frac{Y_{ij} - Z_i}{\sigma_{Y_i}}\right)^2$ is χ^2_{n-k}.
- Under H_1, $\frac{V/(k-1)}{W/(n-k)}$ is distributed as $F_{k-1,n-k}$:
$$\frac{V/(k-1)}{W/(n-k)} = \frac{(n-k)\sum_{i=1}^{k} n_i(Z_i - \hat{\mu}_Z)^2}{(k-1)\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - Z_i)^2}$$
- The F test: Accept H_1 if $\frac{V/(k-1)}{W/(n-k)} < c$. Reject H_1 otherwise.
- For a significance level α, $c = f_{[1-\alpha]}$ such that $F_F(f_{[1-\alpha]}; k-1, n-k) = 1 - \alpha$.

Testing Whether $\sigma^2 = \sigma_0^2$ for a Normal Population

- m i.i.d. samples $\{X_i,\ i =, \cdots, m\}$ from $N(\mu, \sigma^2)$
- $H_1 : \sigma^2 = \sigma_0^2$ versus $H_2 : \sigma^2 \neq \sigma_0^2$
- $\Theta_1 = \{\mu, \sigma_0^2\}$ and $\Theta = \{\mu, \sigma^2\}$
- $L(\Theta_1) = (2\pi\sigma_0^2)^{-m/2} \exp\left(-\frac{1}{2}\sum_{i=1}^m (\frac{X_i-\mu}{\sigma_0})^2\right)$ which is maximized when $\hat{\mu}^* = \hat{\mu} \triangleq \frac{1}{m}\sum_{i=1}^m X_i$. Thus,

$$L_{\rm LM} = \frac{1}{(2\pi\sigma_0^2)^{m/2}} \exp\left(-\frac{1}{2}\sum_{i=1}^m \left(\frac{X_i - \hat{\mu}}{\sigma_0}\right)^2\right)$$

- $L(\Theta) = (2\pi\sigma^2)^{-m/2} \exp\left(-\frac{1}{2}\sum_{i=1}^m (\frac{X_i-\mu}{\sigma})^2\right)$ which is maximized when $\hat{\mu}^\dagger = \hat{\mu}$ and $\hat{\sigma}^{2\dagger} = \hat{\sigma}_{\rm ML}^2 \triangleq \frac{1}{m}\sum_{i=1}^m (X_i - \hat{\mu})^2$. Thus,

$$L_{\rm GM} = \frac{1}{(2\pi\hat{\sigma}_{\rm ML}^2)^{m/2}} \exp\left(-\frac{1}{2}\sum_{i=1}^m \left(\frac{X_i - \hat{\mu}}{\hat{\sigma}_{\rm ML}}\right)^2\right) = \frac{\exp(-m/2)}{(2\pi\hat{\sigma}_{\rm ML}^2)^{m/2}}$$

- $\Lambda = L_{\rm LM}/L_{\rm GM}$ is

$$\Lambda = \left(\frac{1}{m}\sum_{i=1}^{m}\left(\frac{X_i-\hat{\mu}}{\sigma_0}\right)^2\right)^{m/2} \exp\left(\frac{1}{2}\left[m-\sum_{i=1}^{m}\left(\frac{X_i-\hat{\mu}}{\sigma_0}\right)^2\right]\right)$$

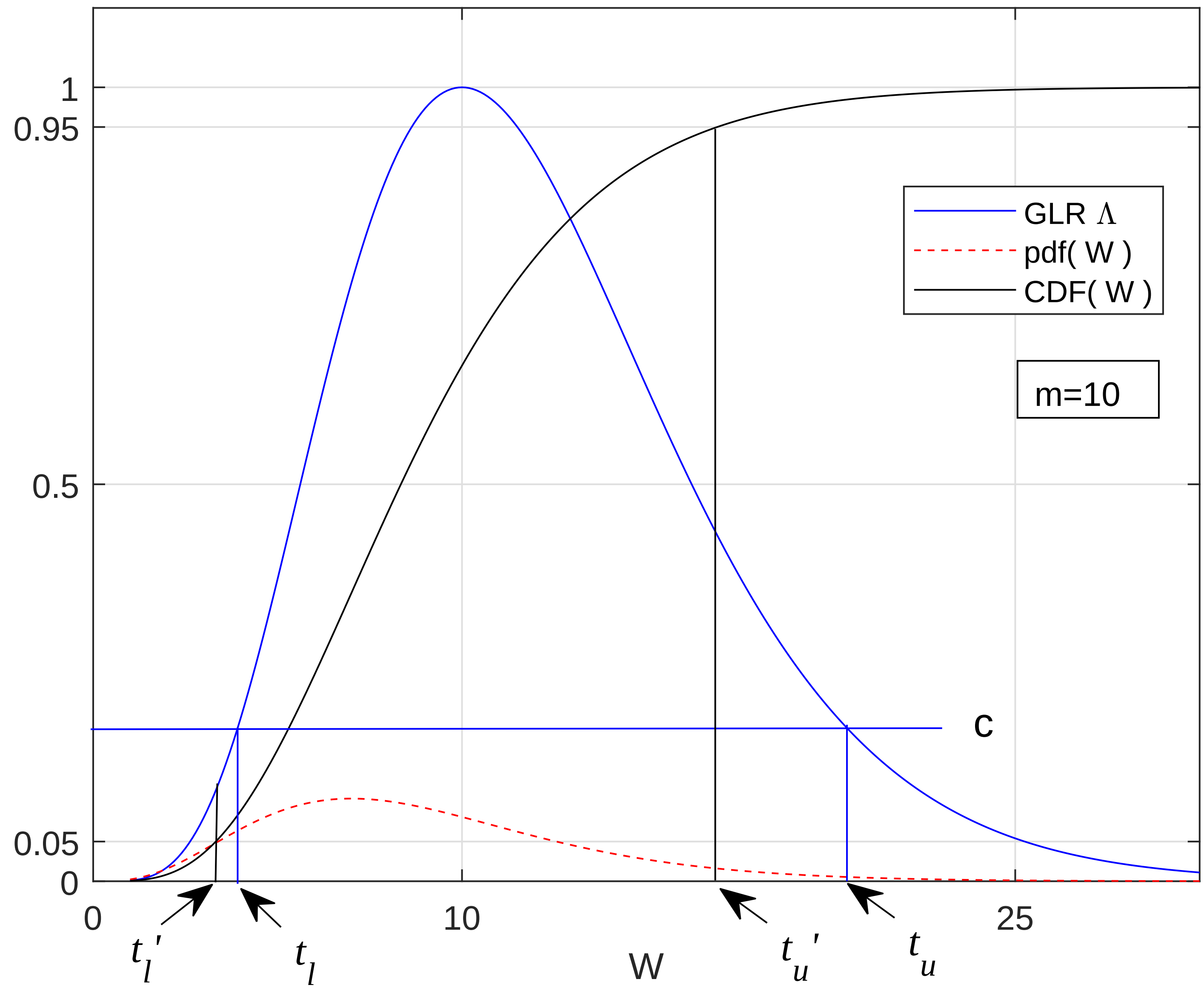
- Note that $W \triangleq \sum_{i=1}^{m}\left(\frac{X_i-\hat{\mu}}{\sigma_0}\right)^2$ is χ^2_{m-1}. Then

$$\Lambda = \left(\frac{W}{m}\right)^{m/2} \exp\left(-0.5(W-m)\right)$$

- The critical event $\{0 < \Lambda < c\}$ is equivalent to $\{0 < W < t_l\} \cup \{t_u < W\}$ where $\Lambda(t_l) = \Lambda(t_u)$ and $t_l < t_u$. (See Fig.)

- For $P[\text{reject } H_1 | H_1 \text{ true}] = \alpha$, using equal area rule for simplicity, we can use
$t_l' = x_{[0.5\alpha]}$ and $t_u' = x_{[1-0.5\alpha]}$
where $F_{\chi^2}(x_{[0.5\alpha]}; m-1) = 0.5\alpha$ and $F_{\chi^2}(x_{[1-0.5\alpha]}; m-1) = 1-0.5\alpha$.

- Reject H_1 if $0 < W < t_l'$ or $W > t_u'$.



1
0.95
0.5
0.05
0
0
10
25
W
GLR Λ
pdf(W)
CDF(W)
m=10
c
t_l'
t_l
t_u'
t_u

- Example 7.3-10 Two sets of samples, P1 and P2 are drawn from $N(1,1)$ and $N(1,4)$, respectively. Test the hypothesis $H_1 : \sigma^2 = \sigma_0^2 = 1$ versus $H_2 : \sigma^2 \neq \sigma_0^2$ for each set using $\alpha = 0.05$ and $\alpha = 0.2$.
  N(1,1) [P1] -0.0644 2.91 -0.323 1.21 2.66 0.45 1.26 0.923 1.96 1.62
  N(1,4) [P2] 0.705 0.685 0.718 1.03 2.52 1.96 0.417 2.69 -1.52 2.98

  Solution: From P1 and P2, we compute $W_1 = 10.3$ and $W_2 = 16.5$, according to $W \triangleq \sum_{i=1}^{10} \left(\frac{X_i - \hat{\mu}}{\sigma_0}\right)^2$. Note $W \sim \chi_9^2$.

  For $\alpha = 0.05$, from χ^2 CDF table,
  we have $t_l' = x_{[0.5\alpha]} = 2.7$ and $t_u' = x_{[1-0.5\alpha]} = 19$.

  Thus, the critical region is $(0, 2.7) \cup (19, \infty)$.

  Since $W_1, W_2 \notin (0, 2.7) \cup (19, \infty)$, we accept H_1 for both P1 and P2.

  For $\alpha = 0.2$, the critical region is $(0, 4.17) \cup (14.7, \infty)$.

  Since $W_1 \notin (0, 4.17) \cup (14.7, \infty)$ but $W_2 \in (0, 4.17) \cup (14.7, \infty)$, H_1 is rejected for P2 but accepted for P1.

  Note: (1) Small sample sizes lead to errors; (2) Small α leads to small critical region.

Goodness of Fit (Pearson's Test or Chi-Square Test)

- H_1 : a set of probabilities $\{p_i, i = 1, \cdots, L\}$ satisfy $\{p_i = p_{0i}, i = 1, \cdots, L\}$ for the predetermined values $\{p_{0i}\}$
- We can test whether data $\{X_i\}$ come from an assumed distribution (CDF), pdf, or PMF by defining $p_i = P[x_i < X \leq x_{i+1}]$ where the range of X is divided into L bins defined by $(x_i, x_{i+1}]$, $i = 1, \cdots, L$.
- Define an RV X_{ij} as $X_{ij} \triangleq 1$ if jth observation of X is in bin i and $X_{ij} \triangleq 0$ otherwise. Define $P[X_{ij} = 1] \triangleq p_i$. Note $\sum_{i=1}^{L} p_i = 1$.
- Number of outcomes in bin i from n trials is $Y_i = \sum_{j=1}^{n} X_{ij}$, $i = 1, \cdots, L$. Note $\sum_{i=1}^{L} Y_i = n$.
- $\{Y_i\}$ have multinomial PMF, but for $n \gg 1$, Y_i is (approx) $N(np_i, np_i)$.
- Pearson's test statistics is defined as $V \triangleq \sum_{i=1}^{L} \left(\frac{Y_i - np_{0i}}{\sqrt{np_{0i}}}\right)^2$. Under H_1, V is (approx) χ^2 with $L - 1$ DOF.
- Pearson's test or Chi-Square test: Reject H_1 if $V > c$.
- For significance α, $c = x_{[1-\alpha]}$ where $F_{\chi^2}(x_{[1-\alpha]}, L - 1) = 1 - \alpha$.

- Example 7.4-1

  We flip a coin 100 times and observe 61 heads and 39 tails. Test the hypothesis H_1 : the coin is fair, i.e., p_{01}=P[head]=0.5=p_{02}=P[tail]. (Set significant level $\alpha = 0.05$)

  Solution:

  We have

  $$V = \sum_{i=1}^{2} \left(\frac{Y_i - np_{0i}}{\sqrt{np_{0i}}} \right)^2 = \frac{(61-50)^2}{100\times 0.5} + \frac{(39-50)^2}{100\times 0.5} = 4.84.$$

  Computing the critical value from $0.95 = F_{\chi^2}(x_{[0.95]}; 1)$ yields $x_{[0.95]} = 3.84$.

  Since $V = 4.84 > 3.84$, we reject the hypothesis that the coin is fair.

- Example 7.4-2

  Test the hypothesis that a six-faced die is fair at $\alpha = 0.05$. Let $Y_i, i = 1, \ldots, 6$ denote the number of times face i shows up. Casting the die 1000 times, we observe
  $Y_1 = 152$, $Y_2 = 175$, $Y_3 = 165$, $Y_4 = 180$, $Y_5 = 159$, $Y_6 = 171$.

  Solution:

  We have
  $V = \sum_{i=1}^{6} \left(\frac{Y_i - np_{0i}}{\sqrt{np_{0i}}} \right)^2 = \frac{6}{1000} \sum_{i=1}^{6} (Y_i - 1000/6)^2 = 3.25.$

  Computing the critical value from $0.95 = F_{\chi^2}(x_{[0.95]}; 5)$ yields $x_{[0.95]} = 11.1$.

  Since $V = 3.25 < 11.1$, we accept the hypothesis.

- Example 7.4-3 (Test of Normality)
  Let H_1 be the hypothesis that X is distributed as $N(0,1)$ and H_2 be the alternative. In 1000 observations, for the ten intervals defined as $R_1 \triangleq (-\infty,-2]$, $R_2 \triangleq (-2,-1.5]$, $R_3 \triangleq (-1.5,-1]$, $R_4 \triangleq (-1,-0.5]$, $R_5 \triangleq (-0.5,0]$, $R_6 \triangleq (0,0.5]$, $R_7 \triangleq (0.5,1]$, $R_8 \triangleq (1,1.5]$, $R_9 \triangleq (1.5,2]$, and $R_{10} \triangleq (2,\infty)$, we observe that the number of outcomes in R_i denoted by Y_i for $i = 1, \cdots, 10$, are $Y_1 = 19$, $Y_2 = 42$, $Y_3 = 96$, $Y_4 = 135$, $Y_5 = 202$, $Y_6 = 193$, $Y_7 = 155$, $Y_8 = 72$, $Y_9 = 53$, and $Y_{10} = 33$.
  Test H_1 at $\alpha = 0.05$.

  Solution: Define $p_{0i} = P[X_{SN} \in R_i]$. Then, we have $p_{01} = 0.023$, $p_{02} = 0.044$, $p_{03} = 0.092$, $p_{04} = 0.145$, $p_{05} = 0.1915$, $p_{06} = 0.1915$, $p_{07} = 0.15$, $p_{08} = 0.092$, $p_{09} = 0.044$, $p_{010} = 0.023$.

  We have $V = \sum_{i=1}^{10} \left(\frac{Y_i - np_{0i}}{\sqrt{np_{0i}}} \right)^2 = 12.9$.

  Computing the critical value from $1 - \alpha = 0.95 = F_{\chi^2}(x_{[0.95]}; 9)$ yields $x_{[0.95]} = 16.92$.

  Since $V < x_{[0.95]}$, we accept the hypothesis H_1.

Testing If $P[E_1] = P[E_2]$ Using Pearson's Test

- Z_1 = number of times E_1 occurs in m i.i.d. trials
- Z_2 = number of times E_2 occurs in n i.i.d. trials
- Define $P[E_i] \triangleq p_i$, $q_i = 1 - p_i$, $i = 1, 2$.
- Define $Y = (Z_1/m) - (Z_2/n)$. Note for $m \gg 1, n \gg 1$, by CLT, we have $Z_1/m : N(p_1, p_1 q_1/m)$ and $Z_2/n : N(p_2, p_2 q_2/n)$.
- Under $H_1 \triangleq \{p_1 = p_2\}$, $Y : N(0, p_1 q_1 (m+n)/(mn))$, and the Pearson test statistics is $V = \left(\frac{Y}{\sigma_Y}\right)^2$ which is χ_1^2 (with 1 DoF).
- At significance α, find $x_{[1-\alpha]}$ which satisfies $F_{\chi^2}(x_{[1-\alpha]}; 1) = 1 - \alpha$. Then, the test is

  Accept H_1 if $V < x_{[1-\alpha]}$. Reject H_1 otherwise.
- If σ_Y is unknown, replace it with its estimate $\hat{\sigma}_Y = \sqrt{\hat{p}\hat{q}(m+n)/(mn)}$ where $\hat{p} = (Z_1 + Z_2)/(m+n)$ and $\hat{q} = 1 - \hat{p}$. Recall that under H_1, $p_1 = p_2 = p$ and $\sigma_Y^2 = pq(m+n)/(mn)$.

- Example 7.4-4 In a Governor's race, exit polls showed in a upstate county 167 out of 211 voters voted for the Republican, while in a downstate county 216 out of 499 voters voted Republican. Can we assume the probability, p_1, that an upstate voter will vote for Republican is the same as, p_2, that of a downstate voter ? ($\alpha = 0.05$)

  Solution: Denote $H_1 : p_1 = p_2 = p$ and $H_2 : p_1 \neq p_2$.

  Under H_1, we compute

  $\hat{p} = \frac{Z_1+Z_2}{m+n} = \frac{167+216}{211+499} = 0.54, \;\; \hat{q} = 0.46,$

  $\hat{\sigma}_Y = \sqrt{\hat{p}\hat{q}(m+n)/(mn)} = 0.041,$

  $Y = (Z_1/m) - (Z_2/n) = 0.36,$

  $V = Y^2/\sigma_y^2 = (0.36/0.041)^2 \approx 77.$

  At $\alpha = 0.05$, we find that $x_{[0.95]} = 3.84$ from $F_{\chi^2}(x_{[0.95]}; 1) = 0.95$.

  Since $x_{[0.95]} = 3.84 < V$, the hypothesis is rejected.

Run Test for Equality of Two Populations

(Distribution-Free Hypothesis Testing)

- n_1 i.i.d. observations $\{X_i^{(1)}, i = 1, \cdots, n_1\}$ from population P1
- n_2 i.i.d. observations $\{X_i^{(2)}, i = 1, \cdots, n_2\}$ from population P2
- H_1: P1 and P2 have the same distribution (or the two sets of samples come from the same population)
- H_2: the two sets of samples come from different populations (or no enough evidence for H_1)
- $\{Y_i, i = 1, \cdots, n_1\} \triangleq$ ordered samples of $\{X_i^{(1)}\}$ s.t. $Y_i < Y_{i+1}$.
- $\{Z_i, i = 1, \cdots, n_2\} \triangleq$ ordered samples of $\{X_i^{(2)}\}$ s.t. $Z_i < Z_{i+1}$.
- Mix $\{Y_i\}$ and $\{Z_i\}$ and order them similarly. Then count the total number of runs of the ordered sequence, and denote it by D.
- For example, the following ordered sequence $y_1\ y_2\ z_1\ z_2\ y_3\ y_4\ y_5\ z_3\ z_4\ z_5\ y_6\ y_7\ y_8\ z_6\ z_7\ z_8\ z_9\ y_9\ y_{10}\ z_{10}$ has $D = 8$ runs. The first run is $y_1 y_2$, the second run is $z_1 z_2$, the third is $y_3 y_4 y_5$, etc.

- For $n_1 \geq 10$ and $n_2 \geq 10$, under H_1, D can be approximated as Normal with mean and variance given by

$$\mu_D \approx \frac{2n_1 n_2}{n_1 + n_2} \quad \text{and} \quad \sigma_D^2 \approx 4(n_1 + n_2)\left(\frac{n_1}{n_1 + n_2}\right)^2 \left(\frac{n_2}{n_1 + n_2}\right)^2.$$

- For a significance level α, the test threshold d_0 can be computed from SN CDF as

$$\alpha = \sum_{\text{all } d \leq d_0} P_D(d) \approx \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_{[\alpha]}} \exp(-0.5x^2)dx = \Phi(z_{[\alpha]}) = F_{SN}(z_{[\alpha]})$$

where $\quad z_{[\alpha]} \triangleq \frac{d_0 - \mu_D}{\sigma_D} \quad$ or $\quad d_0 = \mu_D + \sigma_D z_{[\alpha]}$.

- If $D > d_0$, accept H_1. Otherwise, reject H_1.

- Example 7.5-8 (modified)
  Consider the following three sets of 10 ordered samples:
  N(1,1) $\rightarrow \{y^{(1)}: -1.4, -0.33, 0.4, 0.44, 0.7, 0.74, 1.3, 1.3, 1.7, 2.4\}$
  N(1,1) $\rightarrow \{y^{(2)}: -0.67, -0.21, 0.38, 0.38, 0.51, 0.71, 1.4, 1.5, 2, 2.9\}$
  N(1,3) $\rightarrow \{y^{(3)}: -3.8, -2.5, -0.13, 2.2, 2.8, 3, 3.8, 4.6, 5.5, 5.8\}$
  where we know $\{y^{(1)}\}$ comes from $N(1,1)$ but don't know about $\{y^{(2)}\}$ ad $\{y^{(3)}\}$.
  At a significant level $\alpha = 0.05$ and $\alpha = 0.2$, (a) test the hypothesis H_a that $\{y^{(1)}\}$ and $\{y^{(2)}\}$ come from the same distribution of $N(1,1)$, (b) test the hypothesis H_b that $\{y^{(1)}\}$ and $\{y^{(3)}\}$ come from the same distribution of $N(1,1)$,
  Solution:
  $\mu_D \approx \frac{2n_1n_2}{n_1+n_2} = 10, \quad \sigma_D^2 \approx 4(n_1+n_2)\left(\frac{n_1}{n_1+n_2}\right)^2\left(\frac{n_2}{n_1+n_2}\right)^2 = 5.$
  For $\alpha = 0.05$, $F_{\mathrm{SN}}(z_{[0.05]}) = 0.05$ yields $z_{[0.05]} = -1.65$ and
  $d_0 = \sigma_D z_{[0.05]} + \mu_D = 6.3.$
  For $\alpha = 0.2$, $F_{\mathrm{SN}}(z_{[0.2]}) = 0.2$ yields $z_{[0.2]} = -0.85$ (approx) and
  $d_0 = \sigma_D z_{[0.2]} + \mu_D = 8.1.$
  (a) After interleaving $\{y^{(1)}\}$ and $\{y^{(2)}\}$ and counting the runs, we get $D = 14$.
  Since $D > d_0$, we accept the hypothesis H_a for both cases of α.
  (b) After interleaving $\{y^{(1)}\}$ and $\{y^{(3)}\}$ and counting the runs, we get $D = 7$.
  For $\alpha = 0.05$, since $D > d_0$, we accept the hypothesis H_b.
  For $\alpha = 0.2$, $D < d_0$, we reject the hypothesis H_b.

Ranking (Rank-Sum) Test for Equality of Two Populations

- $H_1 : F_X = F_Y$, $H_2 : F_X \neq F_Y$ for populations X and Y with CDF F_X and F_Y
- Obtain n_1 i.i.d. samples from population X and order them. Denote the ordered sequence by $\{X_i, i = 1, \cdots, n_1\}$ where $X_i < X_{i+1}$.
- Obtain n_2 i.i.d. samples from population Y and order them. Denote the ordered sequence by $\{Y_i, i = 1, \cdots, n_2\}$ where $Y_i < Y_{i+1}$.
- Mix $\{X_i\}$ and $\{Y_i\}$ and order them similarly. Then, assign the rank of the ith element of the ordered sequence to be i.
- For example, if $n_1 = 3$, $n_2 = 4$ and the ordered sequence is $X_1\ Y_1\ X_2\ X_3\ Y_2\ Y_3\ Y_4$, then the Y sequence has ranks 2, 5, 6, 7.
- A suitable test statistics is $T \triangleq \sum_{Y \text{ sequence}} \text{ranks}$.
- For $n_1 > 7$, $n_2 > 7$, under H_1, T is approximately $N(\mu_T, \sigma_T^2)$ with $\mu_T = n_2(n_1 + n_2 + 1)/2$ and $\sigma_T^2 = n_1 n_2(n_1 + n_2 + 1)/12$.
- For a significance level α, the thresholds t_l and t_u can be found from CDF of Gaussian RV T as $F_T(t_l) = 0.5\alpha$ and $F_T(t_u) = 1 - 0.5\alpha$, i.e., $t_l = \mu_T + \sigma_T z_{[0.5\alpha]}$, $t_u = \mu_T + \sigma_T z_{[1-0.5\alpha]}$ where $F_{SN}(z_{[x]}) = x$.
- Accept H_1 if $t_l \leq T \leq t_u$. Reject H_1 otherwise.

- Example 7.5-9 (modified)
  Testing population sameness for $\{y^{(1)}\}$ and $\{y^{(3)}\}$ in Example 7.5-8 with ranking test for $\alpha = 0.05$ and 0.2.

  Solution:
  Co-join $\{y^{(1)}\}$ and $\{y^{(3)}\}$ and assign ranks to the elements. Then, we have $y_1^{(3)} < y_2^{(3)} < y_1^{(1)} < y_2^{(1)} < y_3^{(3)} < y_3^{(1)} < y_4^{(1)} < y_5^{(1)} < y_6^{(1)} < y_7^{(1)} < y_8^{(1)} < y_9^{(1)} < y_4^{(3)} < y_{10}^{(1)} < y_5^{(3)} < y_6^{(3)} < y_7^{(3)} < y_8^{(3)} < y_9^{(3)} < y_{10}^{(3)}$,
  where the ranks for the elements in $y^{(3)}$ are 1, 2, 5, 13, 15, 16, 17, 18, 19 and 20. We have $T = \sum \mathrm{Ranks} = 126$.

  Next, $\mu_T = n_2(n_1 + n_2 + 1)/2 = 105$ and $\sigma_T^2 = n_1 n_2(n_1 + n_2 + 1)/12 = 175$.

  From SN CDF, we obtain $t_l = \mu_T + \sigma_T z_{[0.5\alpha]}$ and $t_u = \mu_T + \sigma_T z_{[1-0.5\alpha]}$ as follows:

  At $\alpha = 0.05$, we have $z_{[0.025]} = -z_{[0.975]} = -1.96$ and $t_l = 79, t_u = 131$. Since rank-sum $T \in [t_l, t_u]$, we accept the hypothesis.

  At $\alpha = 0.2$, we have $z_{[0.1]} = -z_{[0.9]} = -1.28$ and $t_l = 88, t_u = 122$. Since rank-sum $T \notin [t_l, t_u]$, we reject the hypothesis.