Recall: hypothesis testing:

\[ H_0 : \theta = \theta_0 \]
\[ H_1 : \theta \neq \theta_0 \]

\[ \theta \in \Theta, \theta_0 \in \Theta, \theta_0 = (\theta_{01}, \ldots, \theta_{0k}) \]

Procedure:

1. Collect data \( X_1, \ldots, X_n = X \)
2. Compute an estimate \( \hat{\theta} \) (or some other statistic \( T \)).
3. Determine:
   \[ R_0 = \text{rejection region} \]
   \[ \beta(\theta) = \text{power} \]
   \[ = \Pr_{\theta} (X \in R) \]
   \[ (R \subset \mathbb{R}^k) \]
   \[ \text{in general} \]

\[ \text{size } \alpha = \sup_{\theta \in \Theta_0} \beta(\theta) \]

**Wald Test**

\[ H_0 : \theta = \theta_0 \]
\[ H_1 : \theta \neq \theta_0 \]

Assume that \( \hat{\theta} \sim N(\theta_0, \text{var} \hat{\theta}) \)

Collect data:

\[ X_1, \ldots, X_n \text{ such that } \mathbb{E}(X_j) = \theta_0 \]
under \( H_0 \).
Size of Wald test:

Reject $H_0$ when

$|W| = \left| \frac{\hat{\theta} - \theta_0}{\hat{s}_e} \right| > z_{\alpha/2}$

Recall:

$\mathbb{P}(z > z_{\alpha/2}) = \frac{\alpha}{2}$

Theorem: Asymptotically, the Wald test has size $\alpha$:

$\mathbb{P}_{\theta_0}(|W| > z_{\alpha/2}) \rightarrow \alpha$ as $n \rightarrow \infty$, (as our sample size increases.)

Proof: Under $H_0$, $\hat{\theta} - \theta_0 \sim N(0,1)$, then

$\mathbb{P}_{\theta_0}(|W| > z_{\alpha/2}) = \mathbb{P}_{\theta_0}\left( \left| \frac{\hat{\theta} - \theta_0}{\hat{s}_e} \right| > z_{\alpha/2} \right)$

$\rightarrow \mathbb{P}(|Z| > z_{\alpha/2}) = \alpha$.

Note: For practical considerations, $\hat{s}_e$ is used. But $s_{\theta_0}$ is also ok, if you're able to compute it ($s_{\theta_0}^2 = \text{Var}_{\theta_0}(\hat{\theta})$).

What is the power of the Wald test?

$\beta(\theta) = \mathbb{P}_\theta(X \in \mathcal{R})$

$\quad = \mathbb{P}_\theta\left( \left| \frac{\hat{\theta} - \theta_0}{\hat{s}_e} \right| > z_{\alpha/2} \right)$

$\quad = 1 - \mathbb{P}_\theta\left( -z_{\alpha/2} \leq \frac{\hat{\theta} - \theta_0}{\hat{s}_e} \leq z_{\alpha/2} \right)$

$\quad = 1 - \mathbb{P}_\theta\left( -z_{\alpha/2} \leq \frac{\hat{\theta} - \theta + \theta - \theta_0}{\hat{s}_e} \leq z_{\alpha/2} \right)$

$\quad = 1 - \mathbb{P}_\theta\left( -z_{\alpha/2} \leq \frac{\theta - \theta_0 - \theta_0 + \theta}{\hat{s}_e} \leq z_{\alpha/2} \right)$

$\quad = 1 - \Phi\left( z_{\alpha/2} + \frac{\theta - \theta_0}{\hat{s}_e} \right) + \Phi\left( -z_{\alpha/2} + \frac{\theta - \theta_0}{\hat{s}_e} \right)$

$\quad = 1 - \Phi\left( z_{\alpha/2} + \frac{\theta - \theta_0}{\hat{s}_e} \right) + \Phi\left( -z_{\alpha/2} + \frac{\theta - \theta_0}{\hat{s}_e} \right)$
Example Comparing two means

Data: \( X_1, \ldots, X_m \sim N \) \( \text{Samples from two } \)
\( Y_1, \ldots, Y_n \sim N \) different populations.

\[ \mathbb{E}(X_j) = \mu_1 \]
\[ \mathbb{E}(Y_j) = \mu_2. \]

Want to test
\[ H_0: \mu_1 = \mu_2 \]
\[ \text{vs. } H_1: \mu_1 \neq \mu_2 \]

Let \( s = \mu_1 - \mu_2 \)
\[ H_0: s = 0 \]
\[ H_1: s \neq 0. \]

Use the Wald test:

Let \( \hat{s} = \bar{X} - \bar{Y} \) Under suitable conditions
on \( \text{Var}(X_i), \text{Var}(Y_i) \).

The sample variances are given by:
\[ s_1^2 = \frac{1}{m-1} \sum_{i=1}^{m} (X_i - \bar{X})^2 \]
\[ s_2^2 = \frac{1}{n-1} \sum_{j=1}^{n} (Y_j - \bar{Y})^2 \]

\[ \Rightarrow \text{se}(\hat{s}) = \sqrt{\text{Var}(\bar{X} - \bar{Y})} \]
\[ \approx \sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}} = \hat{s}_{\text{se}} \]

\[ \Rightarrow \frac{\hat{s} - (\mu_1 - \mu_2)}{\hat{s}_{\text{se}}} \sim N(0,1) \]
Therefore the Wald statistic is given by
\[
W = \frac{\hat{\delta} - 0}{\text{se}} = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_1^2 + s_2^2}{n}}}
\]

Pick size $\alpha$, reject $H_0$ if $\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_1^2 + s_2^2}{n}}} > z_{\alpha/2}$.

For small sample sizes of normal variants:

Recall the Wald assumption:
\[
\frac{\hat{\theta} - \theta}{\text{se}} \sim N(0,1)
\]

What if the sample size is small, but normally distributed? ($X_i \sim N(\mu, \sigma^2)$, $\mu, \sigma^2$ are unknown).

Then let $\hat{\mu} = \bar{X}$ and let
\[
s^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2,
\text{se}(\hat{\mu}) = \frac{1}{\sqrt{n}} \sqrt{s^2}
\]

Consider the quantity:
\[
\frac{\bar{X} - \mu}{\frac{1}{\sqrt{n}} \sqrt{\frac{1}{n-1} \sum (X_i - \bar{X})^2}}
\]
has a "t-distribution" with $n-1$ degrees of freedom,

$\sim t_{n-1}$

\[
f(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi} \Gamma\left(\frac{n}{2}\right)\left(1 + \frac{t^2}{n}\right)^{(n+1)/2}} \approx \text{approximately normal}
\]
for large $n$. 4
**p-value**

Instead of reporting "reject $H_0"$ or 'retain $H_0", you could instead ask if $H_0$ is rejected for a particular size $\alpha$.

$$p\text{-value} = \text{smallest } \alpha \text{ at which the test reject } H_0.$$  
$$= \inf \left\{ \alpha : T(X_1, \ldots, X_n) \in R_{\alpha} \right\}$$

"the $p$-value is a measure of the evidence against $H_0"$.

- Large $p$-value: Not evidence in favor of $H_0$.
  1. $H_0$ could be true
  2. $H_0$ is false, but the test has low power.

Then If we have a test of the form:

Reject $H_0$ if and only if $T(X_1, \ldots, X_n) > C_\alpha$,

then $p\text{-value} = \sup_{\theta \in \Theta_0} \mathbb{P}(T(X_1, \ldots, X_n) \geq T(X_1, \ldots, X_n))$

= probability of observing a more extreme event purely due to random fluctuations when $H_0$ is true.
For the Wald test: Let \( w \) be observed Wald statistic
\[
\hat{\theta} - \theta_0 \approx \frac{1}{\hat{\sigma}}
\]
\[
p\text{-value} = \Phi_{\theta_0}(1/w) = \Phi(1/|w|) = 2\Phi(-1/|w|),
\]
The p-value is a random variable.

If Ho is true, then we may observe "reasonable" test statistics or some extreme ones:
\[
\begin{align*}
\mathbb{P}(p\text{-value} < p) &= \mathbb{P}_T \left( \mathbb{P}(T > T(x_nX_n)) < p \right) \\
&= \mathbb{P}_T \left( 1 - F(T(x_nX_n)) < p \right) \quad \text{Let } T \text{ have CDF } F \\
&= \mathbb{P}(F(T(x_nX_n)) > 1-p) \\
&= \mathbb{P}(T > F^{-1}(1-p)) \\
&= 1 - \mathbb{P}(T < F^{-1}(1-p)) \\
&= 1 - F(F^{-1}(1-p)) = 1 - (1-p) \\
&= p
\end{align*}
\]
\[
\Rightarrow p\text{-value} \sim \text{Uniform}(0,1) \text{ random variable.}
\]
\[ \chi^2 \text{ distribution} \]

If \( Z_1, \ldots, Z_k \sim N(0,1) \) and independent, then
\[ V = \sum_{i=1}^{k} Z_i^2 \sim \chi^2_k \]

\[ E(V) = k \]
\[ \text{Var}(V) = 2k \]

\[ f(v) = \frac{v^{k/2-1} e^{-v/2}}{2^{k/2} \Gamma(k/2)}, \quad v > 0. \]

**Pearson \( \chi^2 \) test**

**Setup:** Multinomial model/data

Multinomial random variable \( X = (X_1, \ldots, X_k) \), \( \sum_{i=1}^{k} X_i = n \)

\( \text{# samples in bin } i \) \( \sum_{i=1}^{k} p_i = 1 \)

\[ \hat{p} = \left( \hat{p}_1, \ldots, \hat{p}_k \right) \]

(Extension of binomial random variable)

The MLEs for the \( p_j \) are:

\[ \hat{p}_j = \frac{X_j}{n} \]

**Test:** \( H_0: \hat{p} = p_0 = \left( p_{01}, \ldots, p_{0k} \right) \)

Pearson \( \chi^2 \) statistic:

\[ \chi^2 = \sum_{j=1}^{k} \frac{(X_j - E(X_j))^2}{E(X_j)} = \sum_{j=1}^{k} \left( \frac{X_j - np_{0j}}{np_{0j}} \right)^2 \]
Theorem: Asymptotically, under $H_0$, $T \sim \chi^2_{k-1}$

$k-1$ degrees of freedom

Because $p_1 + p_2 + \ldots + p_n = 1$

$\Rightarrow$ $k-1$ df's determine the $\text{ch}^2$ value.

2. The test:

Reject $H_0$ if $T > \chi^2_{k-1, \alpha}$

$P(\chi^2_{k-1} > \chi^2_{k-1, \alpha}) = \alpha$

$\Rightarrow$ Asymptotically, this test has level $\alpha$ and

$p$-value $= P(\chi^2_{k-1} > t)$, when $t = \text{observed statistic}$

This is a one-sided test:

\[
\begin{array}{c}
\chi^2_{k-1} \text{ random variable.} \\
\text{area} = p\text{-value.}
\end{array}
\]

This test can also be used for a continuous distribution

(i.e., Goodness of fit, §10.8)

Collect data $X_1, \ldots, X_n$ i.i.d

$H_0 : X_j \sim F$

$H_1 : X_j \not\sim F$.

Let $N_j = \# \text{ of } X_j$'s in bin $I_j$.

Let $N_j = \# \text{ of } X_j$'s in bin $I_j$.
\[ N = (N_1, \ldots, N_k) \sim \text{Multinomial} \left( n, p_1, p_2, \ldots, p_k \right) \]

Define test statistic similarly:

\[ Q = \sum_{j=1}^{k} \frac{(N_j - np_j)^2}{np_j} \]

Under \( H_0 \), \( Q \sim \chi^2_{k-1} \), apply Pearson test.

**The Likelihood Ratio Test**

(think how this compares with the Wald test)

Applicable to vector valued parameters:

Consider testing:

\[ \theta : (\theta_1, \theta_2, \ldots, \theta_k) \in \mathbb{R}^k \]

\[ H_0 : \theta \in \Theta_0 \quad \Theta_0 \subset \mathbb{R}^k \]

\[ H_1 : \theta \notin \Theta_0 \]

The Likelihood Ratio Statistic (LRS):

\[ \lambda = 2 \cdot \log \left( \frac{\sup_{\Theta} L(\theta)}{\sup_{\Theta_0} L(\theta)} \right) = 2 \cdot \log \frac{L(\hat{\theta})}{L(\hat{\theta}_0)} \]

where \( L(\theta) = f(x_1, \ldots, x_n; \theta) \)

= likelihood of the data with

joint pdf \( f \).
Theorem
Let \( \theta = (\theta_1, \ldots, \theta_q, \theta_{q+1}, \ldots, \theta_r) \in \mathbb{R}^r \)
\[ \mathcal{P}_0 = \{ \theta: \theta_{q+1}, \ldots, \theta_r = \theta_{q+1}, \ldots, \theta_r \} \]
Under \( H_0: \theta \in \mathcal{P}_0 \), \( \lambda \sim \chi^2_{r-q} \)
\[ p\text{-value} = \mathbb{P}(\chi^2_{r-q} > \lambda) \]
observed value.

Note: the LRT is applicable to any distribution, relies on the asymptotic normality of the MLE.

Neyman - Pearson Lemma

Consider the special case:

\[ H_0: \theta = \theta_0 \]
\[ H_1: \theta = \theta_1 \]
What is the most powerful test?

Then
\[ T = \frac{L(\theta_1)}{L(\theta_0)} = \frac{\prod_i^n f(x_i; \theta_1)}{\prod_i^n f(x_i; \theta_0)} \]

Choose \( k \) such that \( \mathbb{P}_{\theta_0}(T > k) = \alpha \).

Then, the test which rejects \( H_0 \) if \( T > k \) is the most powerful size \( \alpha \) test. (i.e., for a fixed \( \alpha \), this test maximizes \( \beta(\theta_1) \).)
Rephrase: Among all tests with the same probability of a Type I error, the Likelihood Ratio Test minimizes the probability of a Type II error.