Hypothesis Testing

- Suppose that $X = (X_1, \ldots, X_n)$ is a sample and the distribtion of X is determined by θ, where $\theta \in \Theta$. Consider the testing problem

$$H_0 : \theta \in \Theta_0 \text{ v.s. } H_1 : \theta \in \Theta_0^c. \tag{1}$$

A test for the testing problem in (1) is characterized by its rejection region (critical region). Suppose $\mathcal{C}$ is the rejection region of a test ϕ, then

$$\text{test } \phi \text{ rejects } H_0 \ \Leftrightarrow X \in \mathcal{C}.$$

- Example 1. Suppose that $(X_1, \ldots, X_n)$ is a random sample from $N(\mu, \sigma^2)$, where $\mu \in (-\infty, \infty)$ and $\sigma > 0$. Consider the testing problem

$$H_0 : \mu \leq \mu_0 \text{ v.s. } H_1 : \mu > \mu_0,$$

where μ_0 is a given constant. Suppose that $\alpha \in (0, 1)$ is a given constant and test ϕ rejects H_0 if and only if

$$\frac{\sqrt{n}(\mu(X) - \mu_0)}{S(X)} > t_{\alpha,n-1},$$

where μ and S are function on R^n defined by

$$\mu(x_1, \ldots, x_n) = \frac{\sum_{i=1}^n x_i}{n}$$

and

$$S(x_1, \ldots, x_n) = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu(x_1, \ldots, x_n))^2}{n-1}},$$

and $t_{\alpha,n-1}$ is the $(1 - \alpha)$ quantile of $t(n - 1)$: the t distribution of $(n - 1)$ degrees of freedom. That is, $t_{\alpha,n-1}$ is the constant such that

$$P(t(n - 1) > t_{\alpha,n-1}) = \alpha.$$

Write down the rejection region of ϕ.

Sol. The rejection region of ϕ is

$$\left\{ x \in R^n : \frac{\sqrt{n}(\mu(x) - \mu_0)}{S(x)} > t_{\alpha,n-1} \right\}.$$

- Power function. Suppose that ϕ is a test for the testing problem in (1) with rejection region $\mathcal{C}$. Define a function β by

$$\beta(\theta) = P_\theta(X \in \mathcal{C})$$

for $\theta \in \Theta$. Then β is called the power function of ϕ, and

$$\sup_{\theta \in \Theta_0} \beta(\theta)$$

is called the size of the test ϕ. If the size of ϕ is less than or equal to α, ϕ is called a test of level α.

- Example 2. Show that the test in Example 1 is of size α.

  Sol. Let β be the power function of the test, then

$$\begin{aligned}\beta(\mu,\sigma) &= P_{\mu,\sigma}\left(\frac{\sqrt{n}(\mu(X)-\mu_0)}{S(X)} > t_{\alpha,n-1}\right)\\ &= P_{\mu,\sigma}\left(\frac{\sqrt{n}(\mu(X)-\mu)+\sqrt{n}(\mu-\mu_0)}{S(X)} > t_{\alpha,n-1}\right).\end{aligned}$$

  Under H_0, $\mu \le \mu_0$, so

$$\begin{aligned}\beta(\mu,\sigma) &= P_{\mu,\sigma}\left(\frac{\sqrt{n}(\mu(X)-\mu)+\sqrt{n}(\mu-\mu_0)}{S(X)} > t_{\alpha,n-1}\right) \quad (2)\\ &\le P_{\mu,\sigma}\left(\frac{\sqrt{n}(\mu(X)-\mu)}{S(X)} > t_{\alpha,n-1}\right) = \alpha,\end{aligned}$$

  where the last equality holds since we have shown that

$$\frac{\sqrt{n}(\mu(X)-\mu)}{S(X)} \sim t(n-1)$$

  in Example 4 in the handout "Confidence intervals". Therefore, for all (μ,σ) such that $\mu \le \mu_0$ and $\sigma > 0$, we have

$$\beta(\mu,\sigma) \le \alpha,$$

  which implies that

$$\sup_{(\mu,\sigma):\mu\le\mu_0,\sigma>0} \beta(\mu,\sigma) \le \alpha. \quad (3)$$

  Moreover, when $\mu = \mu_0$ and $\sigma > 0$, from (2), we have

$$\beta(\mu_0,\sigma) = P_{\mu,\sigma}\left(\frac{\sqrt{n}(\mu(X)-\mu)}{S(X)} > t_{\alpha,n-1}\right) = \alpha,$$

  which, together with (3), implies that

$$\sup_{(\mu,\sigma):\mu\le\mu_0,\sigma>0} \beta(\mu,\sigma) = \alpha,$$

  so the size of the test is α.

- A test can be constructed using a confidence interval.

  Fact 1 *Suppose that X is a sample and $[L(X), U(X)]$ is a $(1-\alpha)$ C.I. of $g(\theta)$ based on X. Consider the testing problem*

$$H_0 : g(\theta) = \tau_0 \text{ v.s. } H_1 : g(\theta) \ne \tau_0. \quad (4)$$

  Define ϕ to be the test with rejection region

$$\{x : \tau_0 \notin [L(x), U(x)]\},$$

  then ϕ is a level α test for the testing problem in (4).

Fact 1 holds true since under $H_0 : g(\theta) = \tau_0$, the probabilty that ϕ rejects H_0 is

$$
\begin{aligned}
& P_\theta\left(\tau_0 \notin [L(x), U(x)]\right) \\
&= P_\theta\left(g(\theta) \notin [L(x), U(x)]\right) \qquad\qquad (g(\theta) = \tau_0) \\
&= 1 - P_\theta(g(\theta) \in [L(x), U(x)]) \\
&\leq 1 - (1 - \alpha) = \alpha.
\end{aligned}
$$

- Example 3. Suppose that $(X_1, \ldots, X_n)$ is a random sample from $N(\mu, 1)$ and $\alpha \in (0, 1)$. Let $z_{\alpha/2}$ be the $(1 - \alpha/2$ quantile of $N(0, 1)$. That is,

  $$P(N(0,1) > z_{\alpha/2}) = \alpha/2.$$

  Let $\bar{X}$ be the sample mean $\sum_{i=1}^n X_i/n$.

  (a) Show that $(\bar{X} - z_{\alpha/2}/\sqrt{n}, \bar{X} + z_{\alpha/2}/\sqrt{n})$ is a $(1 - \alpha)$ C.I. of μ.

  (b) Consider the testing problem

  $$H_0 : \mu = 1 \text{ v.s. } H_1 : \mu \neq 1.$$

  Construct a level α test based on the C.I. in Part (a).

  Sol.

  (a) Since $\sqrt{n}(\bar{X} - \mu) \sim N(0, 1)$, we have

  $$P\left(-z_{\alpha/2} < \sqrt{n}(\bar{X} - \mu) < z_{\alpha/2}\right) = 1 - \alpha,$$

  which implies that

  $$P(\mu \in (\bar{X} - z_{\alpha/2}/\sqrt{n}, \bar{X} + z_{\alpha/2}/\sqrt{n})) = 1 - \alpha,$$

  so $(\bar{X} - z_{\alpha/2}/\sqrt{n}, \bar{X} + z_{\alpha/2}/\sqrt{n}))$ is a $(1 - \alpha)$ C.I. of μ.

  (b) Let

  $$\mathcal{C} = \left\{(x_1, \ldots, x_n) \in R^n : 1 \notin \left(\frac{\sum_{i=1}^n x_i}{n} - \frac{z_{\alpha/2}}{\sqrt{n}}, \frac{\sum_{i=1}^n x_i}{n} + \frac{z_{\alpha/2}}{\sqrt{n}}\right)\right\},$$

  then by Fact 1, the test with rejection region $\mathcal{C}$ is a level α test.

- Most powerful test. Suppoes that ϕ is a level α test for the testing problem in (1). If for every ϕ^* that is a level α test for (1),

  $$\beta_\phi(\theta) \geq \beta_{\psi^*}(\theta) \text{ for every } \theta \in \Theta_0^c,$$

  then ϕ is called a most powerful test. Here β_ϕ and β_{ϕ^*} denote the power functions of ϕ and ϕ^* respectively.

- The existence of most power test is guaranteed by for a testing problem in (5).

Theorem. *(Neyman-Pearson Lemma) Suppose that X is a sample with PDF (or PMF) f_θ, where $\theta \in \{\theta_0, \theta_1\}$. Consider the testing problem*

$$H_0 : \theta = \theta_0 \text{ v.s. } H_1 : \theta = \theta_1. \quad (5)$$

For a constant $k > 0$, let ϕ to be a test for (5) with rejection region including

$$\{x : f_{\theta_1}(x) - k f_{\theta_0}(x) > 0\}$$

but not including

$$\{x : f_{\theta_1}(x) - k f_{\theta_0}(x) < 0\}.$$

Suppose that ϕ is of size α, then ϕ is a most powerful level α test.

The proof of Neyman-Pearson Lemma is based on the following inequaliaty:

$$(I_{\mathcal{C}}(x) - I_{\mathcal{C}^*}(x))(f_{\theta_1}(x) - k f_{\theta_0}(x)) \geq 0,$$

where $\mathcal{C}$ and $\mathcal{C}^*$ are rejection region of ϕ and another level α test ϕ^* respectively.

- Example 4. Suppose that $(X_1, \ldots, X_n)$ is a random sample from $N(\mu, 1)$, where $\mu \in \{1, 1.2\}$. Consider the testing problem

$$H_0 : \mu = 1 \text{ v.s. } H_1 : \mu = 1.2. \quad (6)$$

(a) Find a level α most powerful test for the testing problem in (6). Denote the most powerful test by ϕ_1.

(b) Let ϕ_2 be the test in in Example 3. Is ϕ_2 also a level α test for the testing problem in (6)?

(c) Let β_1 and β_2 be the power functions of ϕ_1 and ϕ_2 respectively. Find $\beta_1(1.2)$ and $\beta_2(1.2)$. Write down the R commands for computing $\beta_1(1.2) - \beta_2(1.2)$ when $n = 100$ and $\alpha = 0.05$.

Ans.

(a) Let z_α be the $(1-\alpha)$ quantile of $N(0,1)$ and let

$$\mathcal{C} = \left\{(x_1, \ldots, x_n) \in R^n : \sqrt{n}\left(\frac{\sum_{i=1}^n x_i}{n} - 1\right) > z_\alpha\right\}.$$

Then the test with rejection region $\mathcal{C}$ is a level α most powerful test for the testing problem in (6).

(b) Yes, since the H_0 in this example is the same as the H_0 in Example 3.

(c) $\beta_1(1.2) = P(N(0,1) > -0.2\sqrt{n} + z_\alpha)$ and

$$\beta_2(1.2) = 1 - P(-0.2\sqrt{n} - z_{\alpha/2} < N(0,1) < -0.2\sqrt{n} + z_{\alpha/2}).$$

The R commands for computing $\beta_1(1.2) - \beta_2(1.2)$ when $n = 200$ and $\alpha = 0.05$ are given below.

```
n <- 200
alpha <- 0.05
z1 <- qnorm(1-alpha)   #qnorm is the quantile function of N(0,1)
z2 <- qnorm(1-alpha/2)
c0 <- -0.2*sqrt(n)
c1 <- c0 + z1
c2 <- c0 - z2
c3 <- c0 + z2
beta1 <- 1-pnorm(c1)   #pnorm is the CDF of N(0,1)
beta2  <- 1-(pnorm(c3)-pnorm(c2))
beta1 - beta2          #answer
```

- Approximate most powerful test. Suppose that $(X_1, \ldots, X_n)$ is a random sample from a distribution with PDF g_θ. Let $f_\theta(x_1, \ldots, x_n) = \prod_{i=1}^n g_\theta(x_i)$ for $(x_1, \ldots, x_n) \in R^n$, then f_θ is a PDF of the sample. For $k > 0$, let

$$\mathcal{C} = \{x : f_{\theta_1}(x) - k f_{\theta_0}(x) > 0\}$$

Then the test with rejection region $\mathcal{C}$ is a most powerful level α test for (5) if the size of the test is α. In the case where $f_{\theta_0}(x) > 0$ for $x \in R^n$, we can define

$$\Lambda(X_1, \ldots, X_n) = \log \frac{f_{\theta_1}(X_1, \ldots, X_n)}{f_{\theta_0}(X_1, \ldots, X_n)},$$

then the most powerful test rejects H_0: $\theta = \theta_0$ when Λ is large. Under H_0, when n is large, it follows from CLT that

$$\frac{\sqrt{n}(\Lambda(X_1, \ldots, X_n)/n - \mu_\Lambda)}{\sigma_\Lambda} \xrightarrow{\mathcal{D}} N(0,1),$$

where

$$\mu_\Lambda = E_{\theta=\theta_0}\left(\log \frac{g_{\theta_1}(X_1)}{g_{\theta_0}(X_1)}\right)$$

and

$$\sigma_\Lambda = \sqrt{Var_{\theta=\theta_0}\left(\log \frac{g_{\theta_1}(X_1)}{g_{\theta_0}(X_1)}\right)}.$$

Let z_α be the $(1-\alpha)$ quantile of $N(0,1)$ and let

$$\mathcal{C}^* = \left\{(x_1, \ldots, x_n) \in R^n : \frac{\sqrt{n}(\Lambda(x_1, \ldots, x_n)/n - \mu_\Lambda)}{\sigma_\Lambda} > z_\alpha\right\}, \qquad (7)$$

then the test with rejection region $\mathcal{C}^*$ is an approximate most powerful level α test for (5).

- Example 5. Suppose that $(X_1, \ldots, X_n)$ is a random sample from $N(\mu, 1)$, where $\mu \in \{1, 1.2\}$. Consider the testing problem in (6).

(a) Consider the approximate most powerful level α test with rejection region $\mathcal{C}^*$ in (7). Write down R scripts for computing μ_Λ and σ_Λ in (7).

(b) Write down R scripts for estimating the size of the test in Part (a) based on 10^5 simulated samples of size n with $n = 200$ and $\alpha = 0.05$.

(c) Write down R scripts for estimating the power of the test in Part (a) based on 10^5 simulated samples of size n with $n = 200$ and $\alpha = 0.05$. Compare the estimated power with the power of ϕ_1.

Sol.

(a) Let $g_\mu(x) = e^{-(x-\mu)^2/2}/\sqrt{2\pi}$ for $x \in (-\infty, \infty)$, then g_μ is a PDF of X_1. Let

$$\mu_k = \int_{-\infty}^{\infty} g_1(x) \left(\log \frac{g_{1.2}(x)}{g_1(x)} \right)^k dx.$$

Then

$$\mu_\Lambda = \mu_1$$

and

$$\sigma_\Lambda = \sqrt{\mu_2 - \mu_1^2}$$

The R scripts for computing μ_Λ and σ_Λ are given below.

```
log.g1.fun <- function(x){ dnorm(x, mean=1, sd=1, log=TRUE) }
log.g1.2.fun <- function(x){ dnorm(x, mean=1.2, sd=1, log=TRUE) }
mu.fun <- function(k){
 f <- function(x){
   a <- log.g1.fun(x)
   return( exp(a)*(log.g1.2.fun(x) -a)^k )
  }
  return(integrate(f,-Inf, Inf)$value)
}
mu1 <- mu.fun(1)
mu2 <- mu.fun(2)
mu.Lambda <- mu1
sigma.Lambda <- sqrt(mu2-mu1^2)
```

`mu.Lambda` and `sigma.Lambda` are μ_Λ and σ_Λ respectively.

(b) Below are the R scripts for estimating the size of the test in Part (a) based on 10^5 simulated samples of size n with $n = 200$. Here it is assumed that the functions `log.g1.fun` and `log.g1.2.fun` and the variables `mu.Lambda` and `sigma.Lambda` in the solution to Part (a) have been stored in R.

```
## define a function to compute the normalized LRT test statistic
lrt_normalized.fun <- function(x){
 n <- length(x)
```

```
 Lambda <- sum(log.g1.2.fun(x) - log.g1.fun(x))
 ans <- sqrt(n) *(Lambda/n - mu.Lambda)/sigma.Lambda
 return(ans)
}

#### generate data under H0 and compute the relative frequency of rejecting H0
set.seed(1)
m <- 10^5
n <- 200
ans <- rep(0, m)
for (i in 1:m){
 x <- rnorm(n, mean=1, sd=1)
 ans[i] <- lrt_normalized.fun(x)
}
length(ans[ans>qnorm(0.95)])/m    #relative frequency of rejecting H0
```

The R output after running the above scripts is

```
0.05066
```

and this means the estimated probability of rejecting H_0 is 0.05066.

(c) Below are the R scripts for estimating the power of the test in Part (a) based on 10^5 simulated samples of size n with $n = 200$. Here it is assumed that the functions `log.g1.fun` and `log.g1.2.fun`, the variables `mu.Lambda` and `sigma.Lambda` in the solution to Part (a) have been stored in R.

```
lrt_normalized.fun <- function(x){
 n <- length(x)
 Lambda <- sum(log.g1.2.fun(x) - log.g1.fun(x))
 ans <- sqrt(n) *(Lambda/n - mu.Lambda)/sigma.Lambda
 return(ans)
}

set.seed(1)
m <- 10^5
n <- 200
ans <- rep(0, m)
for (i in 1:m){
 x <- rnorm(n, mean=1.2, sd=1)
 ans[i] <- lrt_normalized.fun(x)
}

length(ans[ans>qnorm(0.95)])/m
```

The estimated power for the approximate test is 0.88169. The power of ϕ_1 in Example 4) can be obtained by running the R scripts in the solution to Example 4 to obtain `beta1`, which is the power of ϕ_1. The power of ϕ_1 is 0.881709. The estimated power for the approximate test is close to the power of ϕ_1.

- Note. Suppose that $(X_1, \ldots, X_n)$ is a random sample from $Bin(1, p)$, where $p \in (0, 1)$. Then an approximate $(1-\alpha)$ C.I. for p is

$$\left(\bar{X} - \frac{z_{\alpha/2}\sqrt{\bar{X}(1-\bar{X})}}{\sqrt{n}}, \bar{X} + \frac{z_{\alpha/2}\sqrt{\bar{X}(1-\bar{X})}}{\sqrt{n}}\right),$$

where $\bar{X} = \sum_{i=1}^{n} X_i/n$ is the sample mean and $z_{\alpha/2}$ is the $(1-\alpha/2)$ quantile of $N(0, 1)$. In Example 5, the estimated Type I error probability is 0.05066 and the endpoints for the observed 95% approximate C.I. are

$$0.05066 \pm \frac{z_{0.025}\sqrt{0.05066(1-0.05066)}}{\sqrt{10^5}},$$

where $z_{0.025}$ can be obtained by running the R command `qnorm(0.975)`, so the observed 95% approximate C.I. is $(0.04930077, 0.05201923)$. Since 0.05 is in the observed 95% approximated C.I., we do not have strong evidence to say that the Type I error probability is not 0.05. Similary, based on the estimated power in Part (c) of Example 5, the observed 95% approximated C.I. of the power of the test contains 0.881709, so we do not have strong evidence to say that the power of the test is different from 0.881709.

- Likelihood ratio test. Suppose that X is a sample with PDF (or PMF) f_θ, where $\theta \in \Theta$. Consider the testing problem in (1). A likelihood ratio test is based on the statistic

$$\Lambda_0(X) = \log\left(\frac{f_{\hat{\theta}_0}(X)}{f_{\hat{\theta}}(X)}\right),$$

where $\hat{\theta}$ is the MLE of θ and $\hat{\theta}_0$ is the MLE of θ under $H_0 : \theta \in \Theta_0$. Suppose that $\Theta \subset R^k$ and Θ contains an open set in R^k. Suppose that

$$\Theta_0 = \{\theta = (g_1(\tau), \ldots, g_k(\tau)) : \tau \in S\}$$

where $S \subset R^{k-d}$, $d > 0$, and S contains an open set in R^{k-d}. Let J be the $k \times (k-d)$ matrix of functions such that the (i, j)-th component of J is the partial derivative of g_i with respect to its j-th component. Suppose that J is of rank $(k-d)$ on S. Under some regularity conditions, we have

$$-2\Lambda_0(X) \xrightarrow{\mathcal{D}} \chi^2(d)$$

as $n \to \infty$, so we can construct an approximate level α test with rejection region

$$\mathcal{C}_0 = \{x \in R^n : -2\Lambda_0(x) > k_{\alpha,d}\}$$

where $k_{\alpha,d}$ is the $(1-\alpha)$ quantile of $\chi^2(d)$.

- Example 6. Suppose that $X = (X_1, \ldots, X_n)$ is a random sample from $N(\mu, \sigma^2)$, where $\mu \in (-\infty, \infty)$ and $\sigma > 0$. Consider the testing problem

$$H_0 : \mu = 0 \text{ v.s. } H_1 : \mu \neq 0. \tag{8}$$

Find an approximate level α test based on the likelihood ratio test statistic $\Lambda_0(X)$ (assuming the regularity conditions hold).

A sketch of solution. Note that

$$\Theta = \{(\mu, \sigma) : \mu \in (-\infty, \infty), \sigma > 0\}$$

is a subset of R^2 and contains an open set in R^2.

$$\Theta_0 = \{(0, \sigma) : \sigma > 0\} = \{(g_1(\sigma), g_2(\sigma)) : \sigma \in (0, \infty)\},$$

where $g_1(\sigma) = 0$ and $g_2(\sigma) = \sigma$). $(0, \infty)$ is a subset of $R = (-\infty, \infty)$ and contains an open set in R. Let

$$J = \begin{pmatrix} g_1'(\sigma) \\ g_2'(\sigma) \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

then the matrix J is of rank 1.

Under H_0, the MLE of $(\mu, \sigma) = (0, \hat{\sigma}_0)$, where

$$\hat{\sigma}_0 = \sqrt{\frac{\sum_{i=1}^n X_i^2}{n}}.$$

Without assuming $\mu = 0$, the MLE of (μ, σ) is $(\bar{X}, \hat{\sigma})$, where $\bar{X}$ is the sample mean and

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}}.$$

The likelihood ratio test statistic

$$\Lambda_0(X) = \log \frac{\left(\frac{1}{\sqrt{2\pi\hat{\sigma}_0^2}}\right)^n e^{-\sum_{i=1}^n (X_i - 0)^2/(2\hat{\sigma}_0^2)}}{\left(\frac{1}{\sqrt{2\pi\hat{\sigma}^2}}\right)^n e^{-\sum_{i=1}^n (X_i - \bar{X})^2/(2\hat{\sigma}^2)}} = \frac{n}{2} \log\left(\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sum_{i=1}^n X_i^2}\right).$$

$$\Theta_0 = \{(g_1(\sigma), g_2(\sigma)) : \sigma > 0\}$$

An approximate level α test based on $\Lambda_0(X)$ rejects H_0 if

$$-2\Lambda_0(X) > k_{\alpha,1},$$

where $k_{\alpha,1}$ is the $(1-\alpha)$ quantile of $\chi^2(1)$.

- R scripts for an experiment of checking the distribution of $-2\Lambda_0(X)$ under $H_0 : \mu = 0$ in Example 6.

```
lrt.stat <- function(x){
 n <- length(x)
 x1 <- x - mean(x)
 lambda <- (n/2)*log( sum(x1^2)/sum(x^2) )
 ans <- -2*lambda
 return(ans)
}
density.chi <- function(x){ dchisq(x, 1) }

set.seed(1)
m <- 10^5
n <- 200
ans <- rep(0, m)
for (i in 1:m){
 x <- rnorm(n, 0, 1)
 #note: the distriubtion of lambda does not depend on sigma when mu=0
 ans[i] <- lrt.stat(x)
}
hist(ans, nclass="scott", freq=FALSE)
curve(density.chi, 0, max(ans), add=TRUE, col=2)

####Compute estimated Type I error probability when mu=0, sigma=1
length(ans[ans>qchisq(0.95, 1)])/m
```

The estimated Type I error probability is 0.05133 and the observed C.I. of the Type I error probability is $(0.0499623, 0.0526977)$.