Hypothesis Testing

Day 11 (2/11/20)

4.5 Introduction to Hypothesis Testing

Let $X_1, X_2, \ldots, X_n$ be a random sample from a population with density $f(x;\theta)$, for θ in the parameter space Ω. Suppose we partition Ω into ω_0 and ω_1, i.e. $\omega_0 \cap \omega_1 = \phi$ and $\omega_0 \cup \omega_1 = \Omega$. Consider the two hypotheses

$$H_0 : \theta \in \omega_0 \text{ versus } H_1 : \theta \in \omega_1$$

called the *null* and *alternative* hypotheses, respectively. Let $\mathcal{D}$ be the domain of $(X_1, X_2, \ldots, X_n)$. Let C be a subset of $\mathcal{D}$. Consider the rejection rule

$$\text{Reject } H_0 \text{ if } (x_1, x_2, \ldots, x_n) \in C$$

and do not reject H_0 otherwise. Then C is called the *critical region* or *rejection region* of the test. Since a test of hypothesis is completely determined by its critical region, we will treat the two terms as synonymous.

Example: Suppose we want to test $H_0 : \mu \leq 0$ versus $H_1 : \mu > 0$. The following statements are equivalent:

- Reject H_0 if $\overline{x} > 2\frac{S}{\sqrt{n}}$
- Critical region: $\overline{x} \in \left(2\frac{S}{\sqrt{n}}, \infty\right)$

Table 1: Type I and Type II errors

	True State of Nature	
Decision	H_0 is true	H_0 is false
Reject H_0	Type I error	Correct decision
Accept H_0	Correct decision	Type II error

Comments:

- If the critical region C is the empty set ϕ, then the test never rejects and $P[\text{Type I error}] = 0$.
- If critical region C is the whole domain $\mathcal{D}$, then test always rejects and $P[\text{Type II error}] = 0$.

Definition: A critical region C is of **size** α if

$$\alpha = \max_{\theta \in \omega_0} P_\theta[(X_1, X_2, \ldots, X_n) \in C] \tag{4.5.4}$$

Definition: The **power function** of a test with critical region C is

$$\gamma_C(\theta) = P_\theta[(X_1, X_2, \ldots, X_n) \in C], \text{ for } \theta \in \omega_1 \tag{4.5.5}$$

i.e. the probability that the test correctly rejects H_0 when the true value of θ is in H_1. If two critical regions C_1 and C_2 have the same size, we say that C_1 is better than C_2 if $\gamma_1(\theta) \geq \gamma_2(\theta)$ for all $\theta \in \omega_1$.

Example 4.5.2 (Binomial proportion) Let $X_1, X_2, \ldots, X_{20}$ be a random sample of size $n = 20$ from a Bernoulli distribution with probability of success p. The total number of successes $S = \sum_{i=1}^{20} X_i$ has a binomial distribution $S \sim \text{Bin}(n = 20, p)$. Now consider testing

$$H_0 : p = .70 \text{ versus } H_1 : p < .70$$

Since fewer successes provide evidence toward H_1, it makes sense that the critical region should be of the form $S \leq k$, for k some integer less than 14. For a given k, the size of the test is

$$\alpha = P_{p=.70}\left[S \leq k\right]$$

The following table gives α for different values of k, using the R function *pbinom(k,20,p=.70)*.

Table 2: Values of α for various rejection regions

k	$P[S \leq k]$
7	0.0013
8	0.0051
9	0.0171
10	0.0480
11	0.1133
12	0.2277
13	0.3920
14	0.5836

Now consider two possible tests: $S \leq 11$ or $S \leq 12$. The following R output calculates $P(S \leq 11)$ and $P(S \leq 11)$ for different values of p.

```
> #### R code
> p<-seq(.4,.8,by=.05)
> cbind(p,pbinom(11,20,p), pbinom(12,20,p))
         p
 [1,] 0.40 0.943473633 0.97897107
 [2,] 0.45 0.869235029 0.94196590
 [3,] 0.50 0.748277664 0.86841202
 [4,] 0.55 0.585693766 0.74799414
 [5,] 0.60 0.404401275 0.58410706
 [6,] 0.65 0.237622357 0.39897340
 [7,] 0.70 0.113331463 0.22772820
 [8,] 0.75 0.040925168 0.10181186
 [9,] 0.80 0.009981786 0.03214266
```

Comments:

1. • $S \leq 11$ has size $\alpha_1 = P_{.70}\left[S \leq 11\right] = .1133$

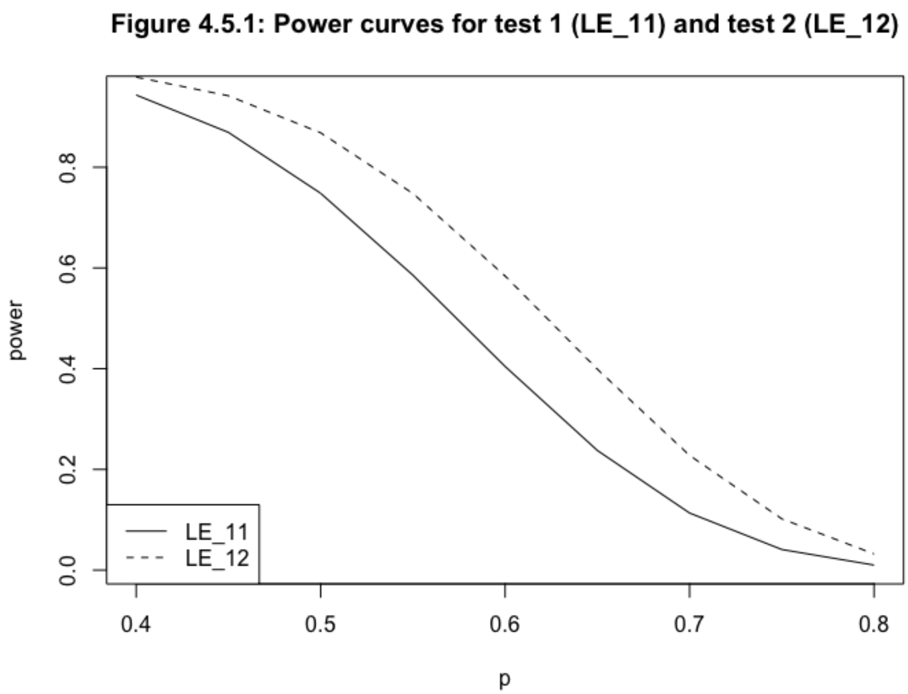
- $S \leq 12$ has size $\alpha_2 = P_{.70}\,[S \leq 12] = .2277$

2. Test 1 has smaller probability of wrongly rejecting H_0. The tradeoff is that Test 1 also has smaller probability of correctly rejecting H_0. For example

$$P_{.65}[S \leq 11] = .2376 \text{ and } P_{.65}[S \leq 12] = .3990$$

$$P_{.60}[S \leq 11] = .4044 \text{ and } P_{.60}[S \leq 12] = .5841$$

The two power functions are plotted below. The R code that generated the plot is provided.




```
> #### R code
> p<-seq(.4,.8,by=.05)
> power1<-pbinom(11,20,p)
> power2<-pbinom(12,20,p)
> plot(power1~p,type="l", lty=1,ylab="power",
  main="Figure 4.5.1: Power curves for test 1 (LE_11) and test 2 (LE_12)")
> lines(power2~p,type="l", lty=2)
> legend("bottomleft", legend = c("LE_11", "LE_12"),lty=1:2)
```

3. The null hypothesis $H_0 : p = .70$ is an example of a *simple* null hypothesis because it contains only one point. This simplifies calculation of size of the test, for example

$$\alpha = P_{p=.70}[S \leq 11] = .1133$$

Often, the null hypothesis is written as a *composite* $H_0 : p \geq .70$ versus $H_1 : p < .70$ in which case definition (4.5.4) applies

$$\alpha = \max_{p \geq .70} P_p[S \leq 11] = P_{p=.70}[S \leq 11] = .1133$$

since the power curves in Figure 4.5.1 are monotone decreasing which means the maximum over $\omega_0 = [.70, 1]$ occurs at the boundary $p = .70$.

4. Other names for the *size* of a test

   (a) level of significance

   (b) maximum power over the null region

   (c) maximum P[Type I error]

   (d) Type I error rate

   (e) false rejection rate

Example 4.5.3 (Large sample test for the mean) Suppose $X_1, X_2, \ldots, X_{25}$ is a random sample from a population with mean μ and variance $\sigma^2 = 100$. To test $H_0 : \mu = 40.0$ versus $H_1 : \mu > 40.0$ consider the rejection region

$$\frac{\overline{X} - 40.0}{S/\sqrt{25}} \geq 1.645 \qquad (*)$$

Since the LHS is approximately $N(0,1)$ under the null, the test (*) has size

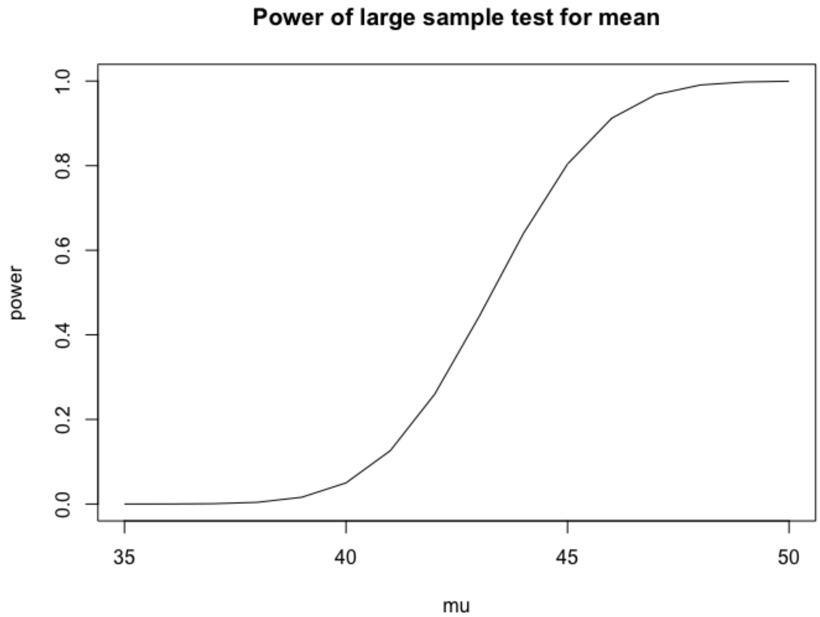
$$P_{\mu=40}\left(\frac{\overline{X} - 40.0}{S/\sqrt{25}} \geq 1.645\right) = 1 - \Phi(1.645) = .05$$

Under alternative, the LHS of (*) is not $N(0,1)$ because the mean is not 0. The power function is

$$\begin{aligned}
\gamma(\mu) = P_\mu\left(\frac{\overline{X} - 40.0}{S/\sqrt{25}} \geq 1.645\right) &= P_\mu\left(\overline{X} \geq 40.0 + 1.645\ S/\sqrt{25}\right) \\
&= P_\mu\left(\frac{\overline{X} - \mu}{\sigma/\sqrt{25}} \geq \frac{40.0 - \mu + 1.645\ S/\sqrt{25}}{10/\sqrt{25}}\right) \\
&\doteq 1 - \Phi\left(\frac{40.0 - \mu}{10/\sqrt{25}} + 1.645\right), \text{ by the CLT and since } S/\sigma \doteq 1
\end{aligned}$$

The following table and plot gives the power function for several values of μ.

μ	$\gamma(\mu)$
37	0.0008308
38	0.0040863
39	0.0159823
40	0.0500000
41	0.1261349
42	0.2595110
43	0.4424132
44	0.6387600
45	0.8037649
46	0.9123145
47	0.9682123




```
> #### R code
> mu<-seq(35,50)
> power<-1-pnorm((40-mu)/(10/sqrt(25)) + 1.645)
> cbind(mu, power)
> plot(power~mu, main="Power of large sample test for mean", type="l")
```

The test has size $\alpha = .05$, and the power increases as μ gets deeper into the alternative region. In general, let $X_1, \ldots, X_n$ be a random sample from a population density f with mean μ and variance σ^2. For testing

$$H_0 : \mu = \mu_0 \text{ versus } H_1 : \mu > \mu_0$$

the size α test

$$\frac{\overline{X} - \mu_0}{S/\sqrt{n}} \geq z_\alpha \tag{4.5.11}$$

has power function

$$\begin{aligned}
\gamma(\mu) &= P_\mu\left(\frac{\overline{X} - \mu_0}{S/\sqrt{n}} \geq z_\alpha\right) = P_\mu\left(\overline{X} \geq \mu_0 + z_\alpha S/\sqrt{n}\right) \\
&= P_\mu\left(\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \geq \frac{\mu_0 - \mu}{\sigma/\sqrt{n}} + z_\alpha \frac{S/\sqrt{n}}{\sigma/\sqrt{n}}\right) \\
&\doteq 1 - \Phi\left(\frac{\mu_0 - \mu}{\sigma/\sqrt{n}} + z_\alpha\right)
\end{aligned}$$