

Section 4.5

Day 9 (2/12)

1 Sec. 4.5 Intro to Hypothesis Testing

Let X_1, X_2, \dots, 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 such that

- $\omega_0 \cap \omega_1 = \phi$
- $\omega_0 \cup \omega_1 = \Omega$

Consider the two hypotheses $H_0 : \theta \in \omega_0$ 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, \dots, X_n) . Let C be a subset of \mathcal{D} . Consider the rejection rule

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

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 test $H_0 : \mu = 0$ versus $H_1 : \mu > 0$. We may describe the test in two equivalent ways:

- Reject H_0 if $\bar{x} > 2(s/\sqrt{n})$
- Critical region: $[2(s/\sqrt{n}), \infty]$

Table 1: Type I and Type II errors

Decision	True State of Nature	
	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 the critical region C is the whole domain \mathcal{D} , then the test always rejects and $P[\text{Type II error}] = 0$.

Definition: A critical region C is of size α if

$$\alpha = \max_{\theta \in \omega_0} P_{\theta}[(X_1, X_2, \dots, X_n) \in C] \quad (1)$$

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

$$\gamma_C(\theta) = P_{\theta}[(X_1, X_2, \dots, X_n) \in C], \text{ for } \theta \in \omega_1 \quad (2)$$

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 $\gamma_{C_1}(\theta) \geq \gamma_{C_2}(\theta)$ for all $\theta \in \omega_1$.

1.1 Example: Binomial test

Let X_1, X_2, \dots, X_{20} be a random sample of size $n = 20$ from a Bernoulli distribution with probability of success p , i.e.

$$\begin{array}{c|cc} x & 1 & 0 \\ \hline p(x) & p & 1-p \end{array}$$

Then the total number of successes is $S = \sum_{i=1}^{20} X_i$ 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 between 0 and 20. For a given k , the size of the test is

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

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	$P[S \leq k]$
6	0.0003
7	0.0013
8	0.0051
9	0.0171
10	0.0480
11	0.1133
12	0.2277
13	0.3920
14	0.5836
15	0.7625

Now consider two possible tests

- Test 1: Reject H_0 if $S \leq 11$. This test has size $\alpha_1 = P_{.70} [S \leq 11] = .1133$

- Test 2: Reject H_0 if $S \leq 12$. This test has size $\alpha_2 = P_{.70}[S \leq 12] = .2277$

Note that Test 1 has smaller probability of wrongly rejecting H_0 (of course, since $C_1 \subset C_2$). The tradeoff is that Test 1 also has smaller probability of correctly rejecting H_0 . For example when the true value is $p = .65$

$$P_{.65}[S \leq 11] = .2376 \text{ and } P_{.60}[S \leq 12] = .3990$$

or when the true value is $p = .60$

$$P_{.60}[S \leq 11] = .4044 \text{ and } P_{.50}[S \leq 12] = .5841$$

The following table gives the power of Test 1 and Test 2 for different values of p . The two

Table 3: Power of tests

p	$P[S \leq 11]$	$P[S \leq 12]$
0.30	0.9949	0.9987
0.35	0.9804	0.9940
0.40	0.9435	0.9790
0.45	0.8692	0.9420
0.50	0.7483	0.8684
0.55	0.5857	0.7480
0.60	0.4044	0.5841
0.65	0.2376	0.3990
0.70	0.1133	0.2277

power functions are plotted in Figure 4.5.1.

Comments:

1. 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 (1) 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 $\omega_0 = [.70, 1.0]$ occurs at the boundary $p = .70$.

2. 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 discovery rate

1.2 Large sample test for the mean

Example: Suppose X_1, X_2, \dots, X_{25} is a random sample from a population with mean μ and variance $\sigma^2 = 100$. To test for

$$H_0 : \mu = 40.0 \text{ versus } H_1 : \mu > 40.0$$

consider the rejection region

$$\frac{\bar{X} - 40.0}{S/\sqrt{25}} \geq 1.645 \quad (3)$$

Since the LHS is approximately $N(0, 1)$ under the null where $\mu = 40.0$, the test (3) has size

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

If the alternative hypothesis is true, then the LHS of (3) is not $N(0, 1)$ because the mean is not 0. The power function is

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

The following table gives $\gamma(\mu)$ for several values of μ . Note that the test has size $\alpha = .05$, and increases as μ gets deeper into the alternative region. In general, let X_1, X_2, \dots, X_n be

Table 4: Power of tests

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

a random sample from a population density f with mean μ and variance σ^2 . For testing

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

consider the rejection region

$$\frac{\bar{X} - \mu_0}{S/\sqrt{n}} \geq z_\alpha$$

The power function is

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

2 Appendix

R code for Table 2

```
> k<-seq(0,20)
> options(scipen=999)
> cbind(k,pbinom(k,20,p=.70))
      k
[1,] 0 0.0000000000348678
[2,] 1 0.0000000016620339
[3,] 2 0.0000000377308814
[4,] 3 0.0000005426947468
[5,] 4 0.0000055502530783
[6,] 5 0.0000429400219536
[7,] 6 0.0002610470070595
[8,] 7 0.0012788796042202
[9,] 8 0.0051381615351214
[10,] 9 0.0171448164312585
[11,] 10 0.0479618973313436
[12,] 11 0.1133314628769786
[13,] 12 0.2277282025818398
[14,] 13 0.3919901877990766
[15,] 14 0.5836291705525188
[16,] 15 0.7624922211223985
[17,] 16 0.8929131954962690
[18,] 17 0.9645168677015313
[19,] 18 0.9923627402258000
[20,] 19 0.9992020773370238
[21,] 20 1.0000000000000000
```

R code for Table 3

```
> pvec<-seq(.10,.70,by=.05)
> options(digits=4)
> cbind(pvec,pbinom(11,20,pvec), pbinom(12,20,pvec))
      pvec
[1,] 0.10 1.0000 1.0000
[2,] 0.15 1.0000 1.0000
[3,] 0.20 0.9999 1.0000
[4,] 0.25 0.9991 0.9998
[5,] 0.30 0.9949 0.9987
[6,] 0.35 0.9804 0.9940
[7,] 0.40 0.9435 0.9790
[8,] 0.45 0.8692 0.9420
[9,] 0.50 0.7483 0.8684
[10,] 0.55 0.5857 0.7480
[11,] 0.60 0.4044 0.5841
[12,] 0.65 0.2376 0.3990
[13,] 0.70 0.1133 0.2277
```

R code for Table 4

```
> mu<-seq(35,50)
> cbind(mu,pnorm(qnorm(.05)-(40-mu)/2))
      mu
[1,] 35 0.0000170
[2,] 36 0.0001338
[3,] 37 0.0008308
[4,] 38 0.0040863
[5,] 39 0.0159823
[6,] 40 0.0500000
[7,] 41 0.1261349
[8,] 42 0.2595110
[9,] 43 0.4424132
[10,] 44 0.6387600
[11,] 45 0.8037649
[12,] 46 0.9123145
[13,] 47 0.9682123
[14,] 48 0.9907423
[15,] 49 0.9978492
```