Math 4501 - Probability and Statistics II

8.6 - Hypothesis testing: power of a statistical test

Type I error and significance level (review)

Type I error for a statistical test with critical region C:

- to reject H_0 when H_0 is true
- occurs if $(x_1, x_2, \ldots, x_n) \in C$ even if H_0 is true

The *significance level* of a statistical test is

$$\begin{aligned} \alpha &= P(\text{Type I error}) \\ &= P(\text{reject } H_0 | H_0 \text{ true}) \\ &= P((X_1, \ldots, X_n) \in C | H_0 \text{ true}) \ . \end{aligned}$$

Type II error and power of a test

Type II error for a statistical test with critical region C:

- to not reject H_0 when H_1 is true
- occurs if $(x_1, x_2, \ldots, x_n) \notin C$ even if H_1 is true

The probability of occurrence of an error of type II is denoted as β:

$$
\begin{aligned}
\beta &= P(\text{Type II error}) \\
&= P(\text{do not reject } H_0 | H_1 \text{ true}) \\
&= P((X_1, \ldots, X_n) \notin C | H_1 \text{ true}) \ .
\end{aligned}
$$

Definition

The *power* of a test is the quantity $1 - \beta$, the probability of correctly rejecting the null hypothesis H_0 when the alternative hypothesis H_1 is true.

Example

Suppose we take a random sample $X_1, X_2, \ldots, X_{20}$ from a Bernoulli distribution with unknown probability of success p.

We wish to test

$$H_0 : p = \frac{1}{2} \qquad \text{against} \qquad H_1 : p < \frac{1}{2}$$

using the critical region

$$C = \left\{ (x_1, x_2, \ldots, x_{20}) : \sum_{i=1}^{20} x_i \leq 6 \right\} .$$

Find the significance level and the power of this test.

Start by observing that since $X_1, \ldots, X_{20}$ are independent Bernoulli (p) random variables, then

$$Y = \sum_{i=1}^{20} X_i \sim bi(20, p)$$

The significance level of the test with the given critical region C is

$$\alpha = P(\text{Rej } H_0 \mid H_0 \text{ true}) = P\left((X_1, \dots, X_{20}) \in C \mid p = \tfrac{1}{2}\right)$$

$$= P\Big(\underbrace{\sum_{i=1}^{20} X_i}_{Y = \sum_{i=1}^{20} X_i \sim bi(20, \frac{1}{2})} \leq 6 \mid p = \tfrac{1}{2}\Big) = \sum_{y=0}^{6} \binom{20}{y} \underbrace{\left(\tfrac{1}{2}\right)^{y} \cdot \left(\tfrac{1}{2}\right)^{20-y}}_{\left(\frac{1}{2}\right)^{20}} = \sum_{y=0}^{6} \binom{20}{y} \left(\tfrac{1}{2}\right)^{20}$$

$$\cong 0.0577$$

The probability β of a type II error depends on the value of p because the alternative hypothesis is $H_1 : p < 1/2$:

$$\beta = P(\text{Not rej } H_0 \mid H_1 \text{ true}) = P\left((X_1, \dots, X_{20}) \notin C \mid p < \tfrac{1}{2}\right) =$$

$$= P\Big(\underbrace{\sum_{i=1}^{20} X_i > 6}_{y = \sum_{i=1}^{20} x_i \in \{7, 8, \dots, 20\}} \mid \underbrace{p < \tfrac{1}{2}}_{Y = \sum_{i=1}^{20} X_i \sim bi(20, p) \text{ with } p < \frac{1}{2}}\Big) = \sum_{y=7}^{20} \binom{20}{y} p^{y} \cdot (1-p)^{20-y} \Big\} \text{ function of the unknown parameter } p$$

The power of this test is the probability, say $K(p)$, of rejecting H_0 when H_1 is true, that is

$$K(p) = 1 - \beta = 1 - \sum_{y=7}^{20} \binom{20}{y} p^y (1-p)^{20-y} , \qquad 0 < p < \tfrac{1}{2}$$

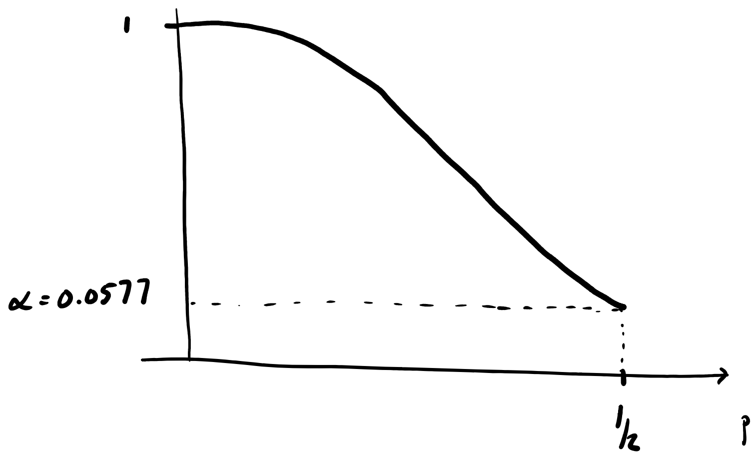
which may also be evaluated as

$$K(p) = P(\text{Rej } H_0 \mid H_1 \text{ true}) = P\left(\sum_{i=1}^{20} X_i \leq 6 \mid p < \tfrac{1}{2} \right)$$

$$= \sum_{y=0}^{6} \binom{20}{y} p^y (1-p)^{20-y} , \qquad 0 < p < \tfrac{1}{2}$$

Finally, note that as $p \to \frac{1}{2}^-$, we have $K(p) \longrightarrow \alpha$

Graph of $K(p)$



good features for a power function:

1. $K(p)$ small when H_0 true
2. $K(p)$ large when p is far away from values where H_0 is true

Notes:

- The power of the test is a function of the parameter we are testing, ranging over all possible values in the corresponding parameter space.
- The sample size can be selected to create a test with some desired significance and power (at a given parameter value).

Example

To test

$$H_0 : p = \frac{1}{2} \qquad \text{against} \qquad H_1 : p < \frac{1}{2}$$

we take a random sample of Bernoulli trials, $X_1, X_2, \ldots, X_n$, and use the test statistic $Y = \sum_{i=1}^{n} X_i$ to construct a critical region of the form

$$C = \{y : y \leq c\} \ .$$

Determine the value of c and the sample size n so that the test significance level is 0.05 and its power when $p = 1/4$ is 0.90.

Let $K(p)$ be the power of this test. We want to determine c and m so that

$$\alpha = K\left(\tfrac{1}{2}\right) = 0.05 \quad \text{and} \quad K\left(\tfrac{1}{4}\right) = 0.9$$

We may proceed as follows:

$$\underbrace{0.05}_{\alpha} = P(\text{Rej } H_0 \mid H_0 \text{ true}) = P\left(\sum_{i=1}^{m} X_i \leq c \;\middle|\; p = \tfrac{1}{2}\right)$$

We now recall that, for large enough m, we can approximate $Y = \sum_{i=1}^{m} X_i$ by a normal distribution. Indeed, we have that

$$\frac{Y - \overbrace{m/2}^{E[Y] \text{ under } H_0}}{\sqrt{\underbrace{m \cdot \tfrac{1}{2} \cdot \tfrac{1}{2}}_{Var(Y) \text{ under } H_0}}} \text{ is approximately } N(0,1)$$

Hence, we find that

$$0.05 = P\Big(\underbrace{\sum_{i=1}^{m} X_i}_{Y \sim bi(m, \frac{1}{2})} \leq c \;\Big|\; p = \tfrac{1}{2} \Big) = P\left(\underbrace{\frac{Y - m/2}{\sqrt{m \cdot \frac{1}{2} \cdot \frac{1}{2}}}}_{\text{approx. } N(0,1)} \leq \frac{c - m/2}{\sqrt{m \cdot \frac{1}{2} \cdot \frac{1}{2}}} \right)$$

half-unit continuity correction

$$\approx \Phi\left(\frac{c + 1/2 - m/2}{\sqrt{m \cdot \frac{1}{2} \cdot \frac{1}{2}}} \right)$$

and so

$$\frac{c + \frac{1}{2} - \frac{m}{2}}{\sqrt{\frac{m}{4}}} \approx -1.645$$

from table

Similarly, we note that

$$0.9 = K\left(\tfrac{1}{4}\right) = P\Big(\underbrace{\sum_{i=1}^{m} X_i}_{Y} \le c \,\Big|\, p = \tfrac{1}{4}\Big) = P\left(\frac{Y - m/4}{\sqrt{m\cdot\frac{1}{4}\cdot\frac{3}{4}}} \le \frac{c - m/4}{\sqrt{m\cdot\frac{1}{4}\cdot\frac{3}{4}}}\right)$$

and we that

$$\frac{Y - \overbrace{m/4}^{E[Y] \text{ at } p = 1/4}}{\sqrt{\underbrace{m\cdot\frac{1}{4}\cdot\frac{3}{4}}_{\mathrm{Var}(Y) \text{ at } p = 1/4}}}$$

is approx. $N(0,1)$ to conclude that

continuity correction

$$0.9 \approx \Phi\left(\frac{c + 1/2 - m/4}{\sqrt{m\cdot\frac{1}{4}\cdot\frac{3}{4}}}\right), \quad \text{that is,} \quad \frac{\overbrace{c + 1/2}^{\text{continuity correction}} - m/4}{\sqrt{m\cdot\frac{3}{16}}} \approx \overbrace{1.282}^{\text{table}}$$

Finally, we solve

$$\begin{cases} \dfrac{c + 1/2 - m/2}{\sqrt{\frac{m}{4}}} \approx -1.645 \\[2ex] \dfrac{c + 1/2 - m/4}{\sqrt{\frac{m \cdot 3}{16}}} \approx 1.282 \end{cases}$$

for c and m to get the desired (approximated) values for m and c:

Subtracting the 1st equation from the 2nd results in

$$\frac{m}{4} \approx 1.645 \sqrt{\frac{m}{4}} + 1.282 \sqrt{\frac{m \cdot 3}{16}} .$$

from where we get that $\sqrt{m} \approx 5.512$ or $m \approx 30.4$

We take $m = 31$ (the least integer greater than 30.4) and substitute into any of the two equations to get

$$c \approx \frac{\overset{31}{m}}{4} - \frac{1}{2} + 1.282\sqrt{\frac{\overset{31}{m}\cdot 3}{16}} \approx 10.9$$

Since $Y = \sum_{i=1}^{m} X_i$ is an integer, we could take either $c = 10$ or $c = 11$.

Using the same kind of arguments as above (ie, normal approximation and continuity correction), we would get:

① if $m = 31$ and $c = 10$

$$\left.\begin{aligned} \alpha = K\left(\tfrac{1}{2}\right) &= P\left(Y \le 10 \mid p = \tfrac{1}{2}\right) \approx 0.0362 \\ K\left(\tfrac{1}{4}\right) &= P\left(Y \le 10 \mid p = \tfrac{1}{4}\right) \approx 0.8730 \end{aligned}\right\}$$

the exact values would be 0.0354 and 0.8716

(2) if $n = 31$ and $c = 11$

$$\alpha = K\left(\tfrac{1}{2}\right) = P\left(Y \leq 11 \mid p = \tfrac{1}{2}\right) \approx 0.0754$$

$$K\left(\tfrac{1}{4}\right) = P\left(Y \leq 11 \mid p = \tfrac{1}{4}\right) \approx 0.9401$$

the exact values would be

0.0748

and

0.9356

Example

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from the $N(\mu, 100)$ distribution.

We wish to test

$$H_0 : \mu = 60 \qquad \text{against} \qquad H_1 : \mu > 60$$

using a test of the form

$$\textit{Reject } H_0 \textit{ if and only if } \bar{X} > c$$

Determine the value of c and the sample size n so that $\alpha = 0.025$ and, when $\mu = 65$, $\beta = 0.05$.

Let $K(\mu)$ be the power function for this test. We want to find $c \in \mathbb{R}$ and $n \in \mathbb{N}$ such that

$$\alpha = K(60) = 0.025 \quad \text{and} \quad K(65) = 1 - \beta = 0.95$$

We observe that

$$0.025 = \alpha = P(\text{Rej } H_0 \mid H_0 \text{ true}) = P(\bar{X} > c \mid \mu = 60) = P\Bigg(\overbrace{\frac{\bar{X} - 60}{10/\sqrt{n}}}^{Z} > \frac{c-60}{10/\sqrt{n}}\Bigg)$$

$$= 1 - P\left(Z \leq \frac{c-60}{10/\sqrt{n}}\right)$$

since, under H_0, we have $\mu = 60$, then $X_1, \ldots, X_n \sim N(60, 100)$ and $\bar{X} \sim N(60, \frac{100}{n})$ so that $Z = \dfrac{\bar{X} - 60}{10/\sqrt{n}} \sim N(0,1)$

that is

$$1 - \Phi\left(\frac{c-60}{10/\sqrt{n}}\right) = 0.025 \quad \text{or}$$

$$\Phi\left(\frac{c-60}{10/\sqrt{n}}\right) = 0.975 \quad \text{and so} \quad \frac{c-60}{10/\sqrt{n}} = 1.96$$

from table

Similarly, from $K(65) = 0.95$, we find

$$0.95 = P(\text{Rej } H_0 \mid \mu = 65) = P(\bar{X} > c \mid \mu = 65) = P\left(\frac{\bar{X} - 65}{10/\sqrt{n}} > \frac{c - 65}{10/\sqrt{n}} \right)$$

But for $\mu = 65$, we have $\bar{X} \sim N\left(65, \frac{100}{n}\right)$ and $Z = \frac{\bar{X} - 65}{10/\sqrt{n}} \sim N(0,1)$

Thus, we have

$$0.95 = P\left(Z > \frac{c - 65}{10/\sqrt{n}} \right) \iff 1 - \Phi\left(\frac{c - 65}{10/\sqrt{n}} \right) = 0.95$$

$$\iff \Phi\left(\frac{c - 65}{10/\sqrt{n}} \right) = 0.05 \iff \frac{c - 65}{10/\sqrt{n}} = -1.645$$

from table

We now solve

$$\begin{cases} \dfrac{c-60}{10/\sqrt{n}} = 1.96 \\[2ex] \dfrac{c-65}{10/\sqrt{n}} = -1.645 \end{cases}$$

for c and n. Subtracting the 2^{nd} equation from the 1^{st}, yields:

$$\frac{5}{10/\sqrt{n}} = 1.96 + 1.645$$

from which we get $\dfrac{10}{\sqrt{n}} = \dfrac{5}{3.605}$ and $c = 60 + 1.96 \cdot \left(\dfrac{10}{\sqrt{n}}\right) = 62.718$

Finally, we also get that $\sqrt{n} = 7.21$ and so $n = 51.98$

Since n must be an integer, we set $n = 52$ to obtain a test with $\alpha \approx 0.025$ and $\beta \approx 0.05$.