Section 8

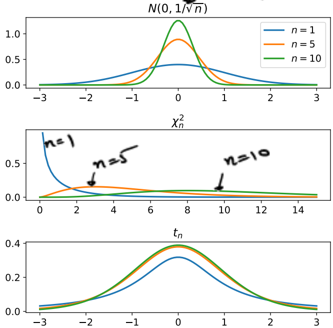
t-test, χ^2-test, two-sample z/F-test

Normal samples and statistics

$$X_1, \ldots, X_n \sim \text{iid Normal}(\mu, \sigma^2)$$

- Sample mean $\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n)$, $E[\overline{X}] = \mu$
- Sample variance $S^2 = \frac{1}{n-1}((X_1 - \overline{X})^2 + \cdots + (X_n - \overline{X})^2)$, $E[S^2] = \sigma^2$

σ $\sigma^2 = 1/n$



PDFs

$$\overline{X} \sim \text{Normal}(\mu, \sigma^2/n)$$

independent

$$\frac{(n-1)}{\sigma^2} S^2 \sim \chi^2_{n-1}$$

Chi-squared with n-1 degrees of freedom

→ only values the

$$\boxed{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1) \text{ `Z'}}$$

$$\frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

→ t distribution with n-1 degrees of freedom

difference? 'S' instead of 'σ'

t-test for mean (variance unknown)

$$X_1, \ldots, X_n \sim \text{iid Normal}(\mu, \sigma^2), \sigma^2 \text{unknown}$$

- Null $H_0 : \mu = \mu_0$, Alternative $H_A : \mu > \mu_0$
- $T = \overline{X}$, Test: Reject H_0 if $T > c$

critical value

t-test for mean (variance unknown)

$$X_1, \ldots, X_n \sim \text{iid Normal}(\mu, \sigma^2), \sigma^2\text{unknown}$$

- Null $H_0 : \mu = \mu_0$, Alternative $H_A : \mu > \mu_0$
- $T = \overline{X}$, Test: Reject H_0 if $T > c$

How to compute significance level?

- Sample variance $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$
- Given H_0, $\frac{T - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$
- For a given sampling, let $S^2 = s^2$

P(reject H0 | H0 is true)

$$\alpha = P(T > c | \mu = \mu_0) = P(t_{n-1} > \frac{c - \mu_0}{s/\sqrt{n}})$$

$$= 1 - F_{t_{n-1}}\left(\frac{c-\mu_0}{s/\sqrt{n}}\right)$$

Contrast with Z-test

σ : known

$\frac{T-\mu_0}{\sigma/\sqrt{n}} \sim \text{Normal}(0,1)$ 'Z'

$T > c \Leftrightarrow \frac{T-\mu_0}{(s/\sqrt{n})} > \frac{c-\mu_0}{s/\sqrt{n}}$

$\sim t_{n-1}$

$F_{t_{n-1}}$: CDF of t distribution with n-1 deg of freedom

χ^2-test for variance

$$X_1, \ldots, X_n \sim \text{iid Normal}(\mu, \sigma^2)$$

- Null $H_0 : \sigma = \sigma_0$, Alternative $H_A : \sigma > \sigma_0$
- $S^2 = \dfrac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$, Test: Reject H_0 if $S > c$

Test statistic

χ^2-test for variance

$$X_1, \ldots, X_n \sim \text{iid Normal}(\mu, \sigma^2)$$

- Null $H_0 : \sigma = \sigma_0$, Alternative $H_A : \sigma > \sigma_0$
- $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$, Test: Reject H_0 if $S > c$

How to compute significance level?

- Given H_0, $\dfrac{(n-1)}{\sigma_0^2}S^2 \sim \chi^2_{n-1}$

reject H_0

$$\alpha = P(S > c|H_0) = P\left(\frac{(n-1)}{\sigma_0^2}S^2 > \frac{(n-1)}{\sigma_0^2}c^2\right) = P\left(\chi^2_{n-1} > \frac{(n-1)}{\sigma_0^2}c^2\right)$$

$\sigma = \sigma_0$

$\Updownarrow$

$S > c$

$$= 1 - F_{\chi^2_{n-1}}\left(\frac{(n-1)c^2}{\sigma_0^2}\right)$$

Two samples from normal distribution

$$X_1, \ldots, X_{n_1} \sim \text{iid Normal}(\mu_1, \sigma_1^2)$$
$$Y_1, \ldots, Y_{n_2} \sim \text{iid Normal}(\mu_2, \sigma_2^2)$$

Two samples from normal distribution

$$X_1, \ldots, X_{n_1} \sim \text{iid Normal}(\mu_1, \sigma_1^2)$$
$$Y_1, \ldots, Y_{n_2} \sim \text{iid Normal}(\mu_2, \sigma_2^2)$$

- $\overline{X} \sim \text{Normal}(\mu_1, \sigma_1^2/n_1)$, $\overline{Y} \sim \text{Normal}(\mu_2, \sigma_2^2/n_2)$
- $\frac{(n_1-1)}{\sigma_1^2} S_X^2 \sim \chi^2_{n_1-1}$, $\frac{(n_2-1)}{\sigma_2^2} S_Y^2 \sim \chi^2_{n_2-1}$

Two samples from normal distribution

$$X_1, \ldots, X_{n_1} \sim \text{iid Normal}(\mu_1, \sigma_1^2)$$
$$Y_1, \ldots, Y_{n_2} \sim \text{iid Normal}(\mu_2, \sigma_2^2)$$

independent of each other.

- $\overline{X} \sim \text{Normal}(\mu_1, \sigma_1^2/n_1)$, $\overline{Y} \sim \text{Normal}(\mu_2, \sigma_2^2/n_2)$
- $\frac{(n_1-1)}{\sigma_1^2} S_X^2 \sim \chi^2_{n_1-1}$, $\frac{(n_2-1)}{\sigma_2^2} S_Y^2 \sim \chi^2_{n_2-1}$

$$\frac{\chi^2_{n_1}/n_1}{\chi^2_{n_2}/n_2} \sim F(n_1, n_2)$$
$$\Downarrow$$
$$\frac{S_X^2/\sigma_1^2}{S_Y^2/\sigma_2^2} \sim F(n_1-1, n_2-1)$$

linear combination of independent normals

$E[\overline{X}-\overline{Y}] = E[\overline{X}] - E[\overline{Y}]$

$\text{Var}(\overline{X}-\overline{Y}) = \text{Var}(\overline{X}) + (-1)^2\,\text{Var}(\overline{Y})$

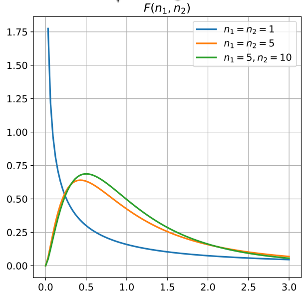
$$\overline{X} - \overline{Y} \sim \text{Normal}(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$$

F distribution

$\sigma_1 = \sigma_2$

$$\frac{S_X^2}{S_Y^2} \sim F(n_1 - 1, n_2 - 1)$$

PDF of



Two sample z-test (known variances)

$$\left.\begin{aligned} X_1, \ldots, X_{n_1} &\sim \text{iid Normal}(\mu_1, \sigma_1^2) \\ Y_1, \ldots, Y_{n_2} &\sim \text{iid Normal}(\mu_2, \sigma_2^2) \end{aligned}\right\} \text{independent}$$

- Null $H_0 : \mu_1 = \mu_2$, Alternative $H_A : \mu_1 \neq \mu_2$
- $\underset{\text{Test statistic}}{T} = \overline{Y} - \overline{X}$, Test: Reject H_0 if $|T| > \underset{\text{critical value}}{c}$

Two sample z-test (known variances)

$$X_1, \ldots, X_{n_1} \sim \text{iid Normal}(\mu_1, \sigma_1^2)$$
$$Y_1, \ldots, Y_{n_2} \sim \text{iid Normal}(\mu_2, \sigma_2^2)$$

- Null $H_0 : \mu_1 = \mu_2$, Alternative $H_A : \mu_1 \neq \mu_2$
- $T = \overline{Y} - \overline{X}$, Test: Reject H_0 if $|T| > c$

How to compute significance level?

- Given H_0, $T \sim \text{Normal}(0, \sigma_T^2)$, where $\sigma_T^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$

reject H_0

$$\alpha = P(|T| > c | H_0) = P(|\text{Normal}(0,1)| > \frac{c}{\sigma_T}) = 1 - F_Z\left(\frac{c}{\sigma_T}\right)$$

$|T| > c \Leftrightarrow \frac{|T|}{\sigma_T} > \frac{c}{\sigma_T}$

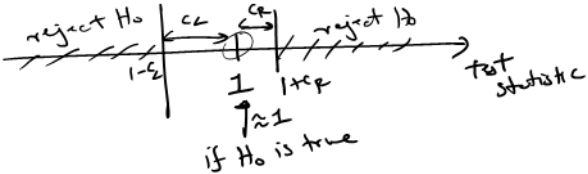
Two sample F-test

$$X_1, \ldots, X_{n_1} \sim \text{iid Normal}(\mu_1, \sigma_1^2)$$
$$Y_1, \ldots, Y_{n_2} \sim \text{iid Normal}(\mu_2, \sigma_2^2)$$

- Null $H_0 : \sigma_1 = \sigma_2$, Alternative $H_A : \sigma_1 \neq \sigma_2$
- $T = \dfrac{S_X^2}{S_Y^2}$, Test: Reject H_0 if $T > 1 + c_R$ or $T < 1 - c_L$

Test statistic

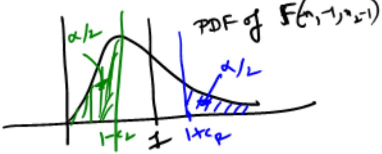
why not $S_X^2 - S_Y^2$?



Two sample F-test

$$X_1, \ldots, X_{n_1} \sim \text{iid Normal}(\mu_1, \sigma_1^2)$$
$$Y_1, \ldots, Y_{n_2} \sim \text{iid Normal}(\mu_2, \sigma_2^2)$$

- Null $H_0 : \sigma_1 = \sigma_2$, Alternative $H_A : \sigma_1 \neq \sigma_2$
- $T = \dfrac{S_X^2}{S_Y^2}$, Test: Reject H_0 if $T > 1 + c_R$ or $T < 1 - c_L$



How to compute significance level?

- Given H_0, $T \sim F(n_1 - 1, n_2 - 1)$

$$\alpha/2 = \underbrace{P(T < 1 - c_L|H_0)}_{\text{green}} = \underbrace{P(T > 1 + c_R|H_0)}_{\text{blue}}$$

Significance level = green + blue.

Section 9

Problems on t-test, χ^2-test and two-sample z/F-test

Problem 1

Suppose $X \sim \text{Normal}(\mu, \sigma^2)$ with unknown σ. For $n = 16$ iid samples of X, the observed sample mean is 10.2 and the sample standard deviation is 3. What conclusion would a t-test reach if the null hypothesis assumes $\mu = 9.5$ (against an alternative hypothesis $\mu > 9.5$) at a significance level of $\alpha = 0.05$?

$X_1, \ldots, X_{16} \sim N(\mu, \sigma^2) \quad \bar{X} = 10.2, \quad S^2 = 3^2 = 9$

$H_0: \mu = 9.5, \quad H_A: \mu > 9.5$

Test: Reject H_0 if $\bar{X} > c$

$$\alpha = P(\bar{X} > c \mid H_0) = P\left(\frac{\bar{X} - 9.5}{3/\sqrt{n}} > \frac{c - 9.5}{3/\sqrt{16}}\right) \approx 1 - F_{t_{15}}\left(\frac{c - 9.5}{(3/4)}\right)$$

$$c = 9.5 + \frac{3}{4} F^{-1}_{t_{15}}(1 - \alpha) = 10.81$$

Since $\bar{X} < 10.81$, Accept H_0.

Problem 2

Suppose X is normally distributed with unknown standard deviation σ. For $n = 16$ iid samples of X, the sample standard deviation is 3.5. What conclusion would a χ^2-test reach if the null hypothesis assumes $\sigma = 3$, with an alternative hypothesis that $\sigma > 3$, and a signifiance level of $\alpha = 0.05$?

$$X_1, \ldots, X_{16} \sim N(\mu, \sigma^2), \quad S^2 = 3.5^2$$

$$H_0: \sigma^2 = 3^2, \quad H_A: \sigma^2 > 3^2 \qquad \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

Test: Reject H_0 if $S^2 > c^2$

$$\alpha = P(S^2 > c^2 \mid H_0) = P\left(\frac{15 S^2}{3^2} > \frac{15 c^2}{3^2}\right) = 1 - F_{\chi^2_{15}}\left(\frac{15 c^2}{3^2}\right)$$

$$c^2 = \frac{3^2}{15} F^{-1}_{\chi^2_{15}}(1-\alpha) = 14.997474$$

Since $3.5^2 = 12.25 < 14.99\ldots$, Accept H_0.

Problem 3

Suppose $X \sim \text{Normal}(\mu_1, 3)$ and $Y \sim \text{Normal}(\mu_2, 4)$. For $n_1 = 16$ iid samples of X and $n_2 = 8$ samples of Y, the observed sample means are 10.2 and 8.2, respectively. What conclusion would a two-sample z-test reach if the null hypothesis assumes $\mu_1 = \mu_2$ (against an alternative hypothesis $\mu_1 \neq \mu_2$) at a significance level of $\alpha = 0.05$?

$X_1, \ldots, X_{16} \sim N(\mu_1, 3)$ $\qquad$ $Y_1, \ldots, Y_8 \sim N(\mu_2, 4)$

$\bar{X} = 10.2$ $\qquad$ $\bar{Y} = 8.2$

$H_0: \mu_1 = \mu_2$, $H_A: \mu_1 \neq \mu_2$ $\qquad$ $T = \bar{X} - \bar{Y} \sim N\Big(\mu_1 - \mu_2, \underbrace{\tfrac{3}{16} + \tfrac{4}{8}}_{= 11/16}\Big)$

Test: Reject H_0 if $|T| > c$

$$\alpha = P(|T| > c \mid H_0) = P\Big(\underbrace{\frac{|T|}{\sqrt{11/16}}}_{\sim |Z|} > \frac{c}{\sqrt{11/16}}\Big) = P\Big(|Z| > \frac{c}{\sqrt{11/16}}\Big)$$

$$= 2 F_Z\Big(-c/\sqrt{11/16}\Big)$$

$$c = -\sqrt{\tfrac{11}{16}} \cdot F_Z^{-1}(\alpha/2) = 1.625$$

Since $|\bar{X} - \bar{Y}| = 10.2 - 8.2 = 2 > 1.625$, Reject H_0.

Problem 4

Suppose $X_1, X_2, \ldots, X_{30}$ is an i.i.d. sample from a distribution $X \sim \text{Normal}(\mu_1, \sigma_1^2)$ and suppose $Y_1, Y_2, \ldots, Y_{25}$ is an i.i.d. sample from a distribution $Y \sim \text{Normal}(\mu_2, \sigma_2^2)$ independent of the X_j variables. If $S_X^2 = 11.4$ and $S_Y^2 = 5.1$, what conclusion would an F-test reach for null hypothesis suggesting $\sigma_1 = \sigma_2$, an alternative hypothesis suggesting $\sigma_1 \neq \sigma_2$, and a signifiance level of $\alpha = 0.05$?

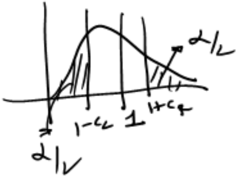
$\frac{S_X^2}{S_Y^2} \sim F(30-1, 25-1)$

$(\sigma_1 = \sigma_2) \to H_0$

Test: Reject H_0 if $\frac{S_X^2}{S_Y^2} > 1+c_R$ (or) $S_X^2/S_Y^2 < 1-c_L$

$\alpha/2 = F_{F(29,24)}(1-c_L) \Rightarrow c_L = 1 - F^{-1}_{F(29,24)}(\alpha/2) = ?$

$\alpha/2 = 1 - F_{F(29,24)}(1+c_R) \Rightarrow 1+c_R = F^{-1}_{F(29,24)}(1-\alpha/2)$



$1+c_R = 2.217$

Since $\frac{S_X^2}{S_Y^2} = \frac{11.4}{5.1} = 2.235 > 2.217$, Reject H_0.

Section 10

More problems on t-test, χ^2-test and two-sample z/F-test

Problem 1

The average annual salary of an entry-level data scientist is reported to be Rs. 8 lakhs per annum. You suspect that this seems too high, and make enquiries with 10 such persons and find that their annual salaries are

6.9, 7.2, 8.7, 7.7, 8.5, 8.0, 8.0, 7.5, 8.7, 7.4

Based on the above, what conclusion can you reach about your suspicion?

$H_0: \mu = 8$, $H_A: \mu < 8$ t-test

Test: Reject H_0 if $\bar{X} < c$ Sample std dev $= 0.631$

$$\alpha = P(\bar{X} < c \mid H_0) = P\left(\frac{\bar{X}-8}{0.631/\sqrt{10}} < \frac{c-8}{0.631/\sqrt{10}}\right) \approx F_{t_9}\left(\frac{c-8}{0.631/\sqrt{10}}\right)$$

$$c = 8 + \frac{0.631}{\sqrt{10}} \cdot F_{t_9}^{-1}(\underbrace{0.05}_{\alpha}) = 7.634$$

Since $7.86 > 7.634$, Accept H_0.

Problem 2

The weight of a cooking gas cylinder is reported to have a standard deviation of 500 g, which you suspect is too low. A sample of 10 cylinders had weights (in Kgs) of 15.1, 14.0 , 14.8, 14.8, 15.8, 14.8, 16.1, 14.3, 14.8, 14.8. Based on this data, what is your conclusion on the standard deviation?

$H_0: \sigma = 0.5$, $H_A: \sigma > 0.5$ Sample std dev $= 0.624$

$n = 10$

Test: Reject H_0 if $S^2 > c^2$.

$$\alpha = P(S^2 > c^2 | H_0) P\left(\underbrace{\frac{9S^2}{0.5^2}}_{\sim \chi^2_9} > \frac{9c^2}{0.5^2}\right) = 1 - F_{\chi^2_9}\left(\frac{9c^2}{0.5^2}\right)$$

$$\frac{9c^2}{0.5^2} = F^{-1}_{\chi^2_9}(1 - \underbrace{\alpha}_{0.05}) \Rightarrow c^2 = \frac{1}{36} F^{-1}_{\chi^2_9}(0.95) = 0.47$$

$c = 0.6855$

Since $0.624 < 0.6855$, Accept H_0.

Problem 3

Weights of a species of squirrels have a standard deviation $\sigma = 10$ grams. Suppose a sample of 30 squirrels from two different locations results in respective sample averages of 122.4 grams and 127.6 grams. Do the squirrels have the same average weight in the two locations?

Test: Reject H_0 if $|T| > c$ $\qquad T = \bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \frac{10^2}{30} + \frac{10^2}{30}\right) = N\left(\mu_1 - \mu_2, \frac{20}{3}\right)$

$$\alpha = P(|T| > c \mid H_0) = P\left(\underbrace{\frac{|T|}{\sqrt{20/3}}}_{\sim |Z|} > \frac{c}{\sqrt{20/3}}\right) = P\left(|Z| > \frac{c}{\sqrt{20/3}}\right)$$

$$\overset{0.05}{\alpha} = 2 F_Z\left(\frac{-c}{\sqrt{20/3}}\right) \Rightarrow c = -\sqrt{\frac{20}{3}}\, F_Z^{-1}(0.025) = 5.06$$

Since $|122.4 - 127.6| = 5.2 > 5.06$, Reject H_0.

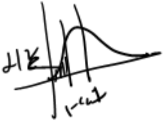
Problem 4

Two instruments for measuring resistors provide the following measurements when measuring a 1000 Ohm and a 3000 Ohm resistor, repeatedly.

$S_x^2 = 24.41$

Instrument 1	Instrument 2
1004	3005
999	2995
993	3019
1000	2993
1008	2992
1002	2991
994	3015
999	2986

$S_y^2 = 146.29$



$\alpha = 0.05$

$1 - c_L = F_{F(7,7)}(0.05/2)$

$= 0.2$

Since $\frac{24.41}{146.29} = 0.167 < 0.2$, Reject H_0.

Do the two instruments have the same variance in their measurements?

Section 11

Likelihood ratio tests

Recall: Is a coin authentic or counterfeit?

An authentic coin is known to have $P(H) = 0.5$ when tossed, while a counterfeit coin has $P(H) = 0.6$. Suppose you have a coin that could be authentic or counterfeit. You may toss the coin multiple times and observe the results. How will you test whether the coin is authentic or counterfeit?

Recall: Is a coin authentic or counterfeit?

An authentic coin is known to have $P(H) = 0.5$ when tossed, while a counterfeit coin has $P(H) = 0.6$. Suppose you have a coin that could be authentic or counterfeit. You may toss the coin multiple times and observe the results. How will you test whether the coin is authentic or counterfeit?

Hypothesis testing

- Null H_0: $P(H) = 0.5$, Alternative H_A: $P(H) = 0.6$
- Toss the coin n times: 2^n possible outcomes
  - A: acceptance set, i.e. if outcome is in A, accept H_0; otherwise, reject H_0
- Significance level: $\alpha = P(\text{not } A | H_0)$, Power: $1 - \beta = P(\text{not } A | H_A)$

$\beta = P(A | H_A)$

Recall: Is a coin authentic or counterfeit?

An authentic coin is known to have $P(H) = 0.5$ when tossed, while a counterfeit coin has $P(H) = 0.6$. Suppose you have a coin that could be authentic or counterfeit. You may toss the coin multiple times and observe the results. How will you test whether the coin is authentic or counterfeit?

Hypothesis testing

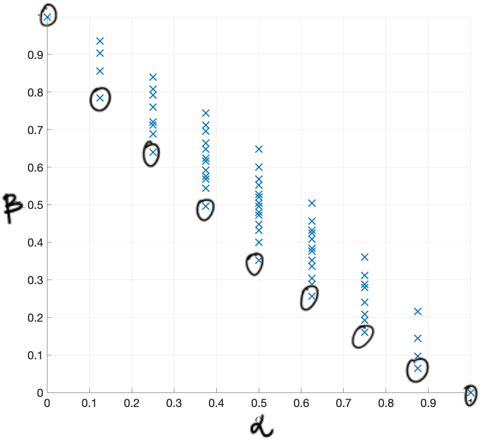
- Null H_0: $P(H) = 0.5$, Alternative H_A: $P(H) = 0.6$
- Toss the coin n times: 2^n possible outcomes
  - A: acceptance set, i.e. if outcome is in A, accept H_0; otherwise, reject H_0
- Significance level: $\alpha = P(\text{not } A | H_0)$, Power: $1 - \beta = P(\text{not } A | H_A)$

Question: How to decide acceptance subset A?

→ "good" (or) "optimal" α, β

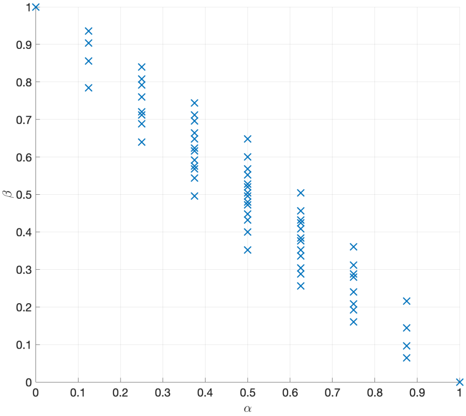
→ Works for large n

Size vs power: $n = 3$ tosses $\longrightarrow$ 8 outcomes $\longrightarrow$ 256 possible sets A



⊗: Best β at a given α.

Size vs power: $n = 3$ tosses



Goal: For a given α, find A that minimizes β or maximizes $1 - \beta$.

Is this possible for 100 tosses?

Simple hypotheses and likelihood ratio test

$$X_1, \ldots, X_n \overset{iid}{\sim} P$$

Simple null and alternative

H_0: $P = f_X$ and H_A: $P = g_X$

PMF (or) PDF

Simple hypotheses and likelihood ratio test

$$X_1, \ldots, X_n \sim P$$

Simple null and alternative

H_0: $P = f_X$ and H_A: $P = g_X$

Likelihood ratio

$$\overset{LR}{\cancel{L}}(X_1, \ldots, X_n) = \frac{\prod_{i=1}^n g_X(X_i)}{\prod_{i=1}^n f_X(X_i)}$$

Likelihood of samples given H_A is true

Likelihood of samples given H_0 is true

Simple hypotheses and likelihood ratio test

$$X_1, \ldots, X_n \sim P$$

Simple null and alternative

H_0: $P = f_X$ and H_A: $P = g_X$

Likelihood ratio

$$L(X_1, \ldots, X_n) = \frac{\prod_{i=1}^{n} g_X(X_i)}{\prod_{i=1}^{n} f_X(X_i)}$$

Likelihood ratio test

Reject H_0 if $T = L(X_1, \ldots, X_n) > c$

Recall: Is a coin authentic or counterfeit?

$$X_1, \ldots, X_n \sim \text{Bernoulli}(p) \qquad p = P(H)$$

- Null H_0: $\overset{p}{\cancel{P(H)}} = 0.5$, Alternative H_A: $\overset{p}{\cancel{P(H)}} = 0.6$

how many? = no. of T's

Likelihood ratio test Likelihood given H_A = $0.6 \times \cdots \times 0.4 \times \cdots \times 0.6 \times \cdots$

how many? = no. of H's

$$T = \frac{0.6^w \, 0.4^{n-w}}{0.5^n} > c$$

Likelihood given H_0

where w is the number of H's in the samples

Recall: Is a coin authentic or counterfeit?

$$X_1, \ldots, X_n \sim \text{Bernoulli}(p)$$

- Null H_0: $P(H) = 0.5$, Alternative H_A: $P(H) = 0.6$

Likelihood ratio test

$$T = \frac{0.6^w \, 0.4^{n-w}}{0.5^n} > c \Leftrightarrow \frac{0.6^w}{0.4^w} > c \cdot \frac{0.5^n}{0.4^n}$$

$$\Updownarrow$$

$$w \log 3/2 > \log\left(\frac{5^n c}{4^n}\right)$$

$$w > \frac{\log(5^n c/4^n)}{\log(3/2)} \triangleq w_c$$

where w is the number of H's in the samples

- Likelihood ratio test is equivalent to $w > w_c$
  - Eg, $n = 100$: Reject null if number of $H > 55$

Optimality of likelihood ratio test

Theorem

Suppose both null and alternative hypotheses are simple, and there is some test that achieves a power of $1-\beta$ *at a significance level* α. *Then, there is a likelihood ratio test at significance level* α *achieving power at least as high as* $1-\beta$.

- *Good news*: For simple null and alternative, likelihood ratio tests are enough.

Optimality of likelihood ratio test

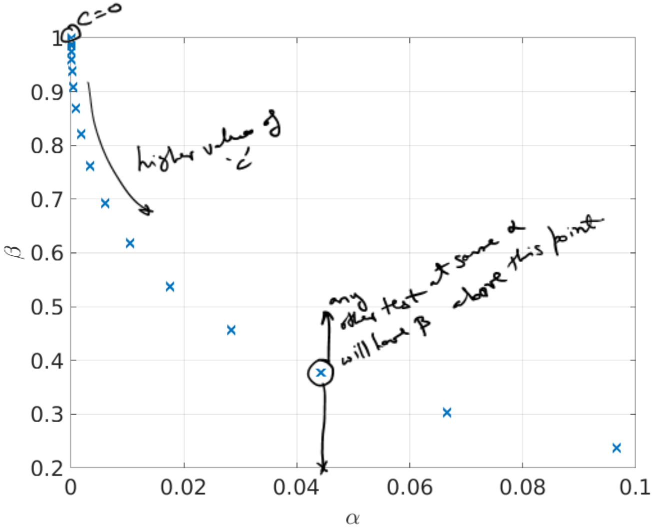
Theorem

Suppose both null and alternative hypotheses are simple, and there is some test that achieves a power of $1 - \beta$ at a significance level α. Then, there is a likelihood ratio test at significance level α achieving power at least as high as $1 - \beta$.

- *Good news*: For simple null and alternative, likelihood ratio tests are enough.
- *Bad news*
  - If any of the hypotheses are composite, then the optimality result does not hold.
  - In most situations, we will not have simple null and alternative!

Fake coin: Size vs power for $n = 100$ tosses

Optimal test: Reject H_0 if number of heads $> c$, $c = 0, 1, \ldots, 100$



Section 12

Goodness of fit tests

Example: Assessing goodness of fit

The expected distribution of grades of students in a class ($0 < p < 1$) and the actual frequencies of grades are shown below:

Grade	S	A	B	C	D	E	U
Fit	$p/32$	$p/4$	$p/2$	$1-p$	$p/8$	$p/16$	$p/32$
Observed	15	97	203	397	55	33	10

Total: 810

ML estimate: $L \propto (1-p)^{397} p^{413}$, $\hat{p}_{ML} = 413/(413+297) = 0.51$

Example: Assessing goodness of fit

The expected distribution of grades of students in a class ($0 < p < 1$) and the actual frequencies of grades are shown below:

Grade	S	A	B	C	D	E	U
Fit	$p/32$	$p/4$	$p/2$	$1-p$	$p/8$	$p/16$	$p/32$
Observed	15	97	203	397	55	33	10

ML estimate: $L \propto (1-p)^{397} p^{413}$, $\hat{p}_{ML} = 413/(413+297) = 0.51$

Expected counts of grades:

Grade	S	A	B	C	D	E	U
Expected	12.9	103.2	206.6	396.9	51.6	25.8	12.9

$\frac{0.51}{32} \times 810$ $\quad$ $\frac{0.51}{4} \times 810$ $\quad$. $\quad$ - $\quad$. $\quad$ $\frac{0.51}{32} \times 810$

Example: Assessing goodness of fit

The expected distribution of grades of students in a class ($0 < p < 1$) and the actual frequencies of grades are shown below:

Grade	S	A	B	C	D	E	U
Fit	$p/32$	$p/4$	$p/2$	$1-p$	$p/8$	$p/16$	$p/32$
Observed	15	97	203	397	55	33	10

ML estimate: $L \propto (1-p)^{397} p^{413}$, $\hat{p}_{ML} = 413/(413+297) = 0.51$

Expected counts of grades:

Grade	S	A	B	C	D	E	U
Expected	12.9	103.2	206.6	396.9	51.6	25.8	12.9

Question: Is the above a good-enough fit?

Chi-square goodness of fit test

$X \in \{a_1, \ldots, a_k\}$ with $P(X = a_i) = p_i$

discrete

Chi-square goodness of fit test

$X \in \{a_1, \ldots, a_k\}$ with $P(X = a_i) = p_i$

Observed vs expected counts: n samples

no. of a_1's in n samples

	a_1	a_2	$\ldots$	a_k
Observed	y_1	y_2	$\ldots$	y_k
Expected	np_1	np_2	$\ldots$	np_k

no. of a_k in n samples

H_0: Samples are iid X, H_A: Samples are not iid X

Chi-square goodness of fit test

$X \in \{a_1, \ldots, a_k\}$ with $P(X = a_i) = p_i$

Observed vs expected counts: n samples

	a_1	a_2	$\ldots$	a_k
Observed	y_1	y_2	$\ldots$	y_k
Expected	np_1	np_2	$\ldots$	np_k

$\frac{(y_1 - np_1)^2}{np_1} + \frac{(y_2 - np_2)^2}{np_2} + \cdots$

H_0: Samples are iid X, H_A: Samples are not iid X

Test Statistic: $T = \sum_{i=1}^{k} \frac{(y_i - np_i)^2}{np_i}$ is $\underline{\text{approx } \chi^2_{k-1}}$

Chi-squared with $k-1$ degrees of freedom (dof)

Test: Reject H_0 if $T > c$

Chi-square goodness of fit test

$X \in \{a_1, \ldots, a_k\}$ with $P(X = a_i) = p_i$

Observed vs expected counts: n samples

	a_1	a_2	$\ldots$	a_k
Observed	y_1	y_2	$\ldots$	y_k
Expected	np_1	np_2	$\ldots$	np_k

H_0: Samples are iid X, H_A: Samples are not iid X

Test Statistic: $T = \sum_{i=1}^{k} \frac{(y_i - np_i)^2}{np_i}$ is approx χ^2_{k-1}

Test: Reject H_0 if $T > c$

CDF of χ^2_{k-1}

Significance level: $\alpha = P(T > c | H_0) \approx 1 - F_{\chi^2_{k-1}}(c)$

Problem: Grades data

$n = 810$ samples, $k = 7$

Grade	S	A	B	C	D	E	U
Observed	15	97	203	397	55	33	10
Expected	12.9	103.2	206.6	396.9	51.6	25.8	12.9

Problem: Grades data

$n = 810$ samples, $k = 7$

Grade	S	A	B	C	D	E	U
Observed	15	97	203	397	55	33	10
Expected	12.9	103.2	206.6	396.9	51.6	25.8	12.9

χ^2 **test**: deg of freedom $k - 1 = 6$, $\alpha = 0.05$, $c = F^{-1}_{\chi^2_6}(1 - 0.05) = 12.59$

$$T = \frac{(15-12.9)^2}{12.9} + \frac{(97-103.2)^2}{103.2} + \frac{(203-206.6)^2}{206.6} + \frac{(397-396.9)^2}{396.9} + \frac{(55-51.6)^2}{51.6} + \frac{(33-25.8)^2}{25.8} + \frac{(10-12.9)^2}{12.9} = 3.66$$

p-value: $1 - F_{\chi^2_6}(3.66) > 0.05$

Conclusion: Since $T < c$, accept the fit.

Goodness of fit for continuous distributions

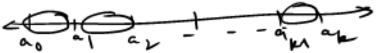
Basic idea: convert continuous to discrete by binning

Goodness of fit for continuous distributions

Basic idea: convert continuous to discrete by binning

$X \sim f_X(x)$, continuous with PDF $f_X(x)$, n samples

Bins: $[a_0, a_1]$, $[a_1, a_2]$, ..., $[a_{k-1}, a_k]$



Bin probabilities: Let $p_i = P(a_{i-1} < X < a_i) = \int_{a_{i-1}}^{a_i} f_X(x)dx$ is known

Observed vs Expected counts:

	$[a_0, a_1]$	$[a_1, a_2]$	...	$[a_{k-1}, a_k]$
Observed	y_1	y_2	...	y_k
Expected	np_1	np_2	...	np_k

no. of samples in $[a_{k-1}, a_k]$

- Pick bins such that each $y_i \geq 5$ and $\sum_i p_i = 1$

Goodness of fit for continuous distributions

Basic idea: convert continuous to discrete by binning

$X \sim f_X(x)$, continuous with PDF $f_X(x)$, n samples

Bins: $[a_0, a_1]$, $[a_1, a_2]$, $\ldots$, $[a_{k-1}, a_k]$

Bin probabilities: Let $p_i = P(a_{i-1} < X < a_i) = \int_{a_{i-1}}^{a_i} f_X(x)dx$

Observed vs Expected counts:

	$[a_0, a_1]$	$[a_1, a_2]$	$\ldots$	$[a_{k-1}, a_k]$
Observed	y_1	y_2	$\ldots$	y_k
Expected	np_1	np_2	$\ldots$	np_k

- Pick bins such that each $y_i \geq 5$ and $\sum_i p_i = 1$

Test: same as before $T = \sum_{i=1}^{k} \frac{(y_i - np_i)^2}{np_i}$ dof $= k-1$

Example: Beta(3,3) goodness of fit

Bin counts of 100 samples from Beta(3,3) distribution are given below.

	$[0.0, 0.2]$	$[0.2, 0.4]$	$[0.4, 0.6]$	$[0.6, 0.8]$	$[0.8, 1.0]$
Observed	7	23	40	28	6
Expected	5.8	25.9	36.5	25.9	5.8

CDF of Beta(3,3)

Eg: Expected count for $[0.2, 0.4] = 100(\underbrace{F_{\text{B}(3,3)}(0.4) - F_{\text{B}(3,3)}(0.2)}_{P(\text{Beta}(3,3) \in [0.2, 0.4])}) = 25.9$

Example: Beta(3,3) goodness of fit

Bin counts of 100 samples from Beta(3,3) distribution are given below.

	$[0.0, 0.2]$	$[0.2, 0.4]$	$[0.4, 0.6]$	$[0.6, 0.8]$	$[0.8, 1.0]$
Observed	7	23	40	28	6
Expected	5.8	25.9	36.5	25.9	5.8

Eg: Expected count for $[0.2, 0.4] = 100(F_{\mathrm{B}(3,3)}(0.4) - F_{\mathrm{B}(3,3)}(0.2)) = 25.9$

$$\begin{aligned} T = &\frac{(7-5.8)^2}{5.8} + \frac{(23-25.9)^2}{25.9} + \frac{(40-36.5)^2}{36.5} \\ &+ \frac{(28-25.9)^2}{25.9} + \frac{(6-5.8)^2}{5.8} = 1.09 \end{aligned}$$

Example: Beta(3,3) goodness of fit

Bin counts of 100 samples from Beta(3,3) distribution are given below.

	$[0.0, 0.2]$	$[0.2, 0.4]$	$[0.4, 0.6]$	$[0.6, 0.8]$	$[0.8, 1.0]$
Observed	7	23	40	28	6
Expected	5.8	25.9	36.5	25.9	5.8

Eg: Expected count for $[0.2, 0.4] = 100(F_{\mathrm{B}(3,3)}(0.4) - F_{\mathrm{B}(3,3)}(0.2)) = 25.9$

$$T = \frac{(7-5.8)^2}{5.8} + \frac{(23-25.9)^2}{25.9} + \frac{(40-36.5)^2}{36.5} + \frac{(28-25.9)^2}{25.9} + \frac{(6-5.8)^2}{5.8} = 1.09$$

P-value $= 1 - F_{\chi^2_4}(1.09) = 0.895$ is quite high. So, accept fit to Beta(3,3).

Example: Test for independence

Consider the following cross-tabulation of grades across 3 different courses.

	S	A	B	C	D	E	U	Total
Math I	15	97	203	387	55	33	10	800
Stats I	28	182	381	726	103	62	19	1500
CT	47	303	634	1209	172	103	31	2500
Total	90	582	1218	2321	331	198	60	4800

#students in Math I

#S grades awarded

Are the grades independent of subjects?

Example: Test for independence

Consider the following cross-tabulation of grades across 3 different courses.

	S	A	B	C	D	E	U	Total
Math I	15	97	203	387	55	33	10	800
Stats I	28	182	381	726	103	62	19	1500
CT	47	303	634	1209	172	103	31	2500
Total	90	582	1218	2321	331	198	60	4800

Are the grades independent of subjects?

- Marginal PMF of grades: $P(\mathrm{S}) = 90/4800$, $P(\mathrm{A}) = 582/4800$ etc.
- Marginal PMF of subjects: $P(\text{Math I}) = 800/4800$, $P(\text{Stats I}) = 1500/4800$ etc.
- If independent, count of (Math I, S) $= \overbrace{\frac{800}{4800} \cdot \frac{90}{4800}}^{p_{ij}} \cdot \overbrace{4800}^{n} = 15$
- If independent, count of (Stats I, A) $= \frac{1500}{4800} \cdot \frac{582}{4800} \cdot 4800 = 181.9$ etc.

Example: Observed vs Expected

Observed

	S	A	B	C	D	E	U	Total
Math I	15	97	203	387	55	33	10	800
Stats I	28	182	381	726	103	62	19	1500
CT	47	303	634	1209	172	103	31	2500
Total	90	582	1218	2321	331	198	60	4800

$\frac{\text{row total} \times \text{col total}}{\text{grand total}} = \frac{2321 \times 1500}{4800}$

Expected, if independent

	S	A	B	C	D	E	U
Math I	15	97	203	386.8	55.2	33	10
Stats I	28.1	181.9	380.6	725.3	103.4	61.9	18.8
CT	46.9	303.1	634.4	1208.9	172.4	103.1	31.2

Example: Chi-squared test for independence

Null H_0: Joint PMF is product of marginals, H_A: It is not

- Test statistic: $T = \sum_{i,j} \frac{(y_{ij} - np_{ij})^2}{np_{ij}}$ $\sim \chi^2_{dof}$

  $dof \downarrow$ (no. of rows -1)(no of cols -1)
  - p_{ij}: product of marginals for (i, j)
  - np_{ij}: expected, if independent
  - We get $T = 0.012$
- Approximate distribution of T: Chi-squared with $(3 - 1)(7 - 1) = 12$ degrees of freedom
  - dof $=$ (no of rows -1)(no of cols -1)

Example: Chi-squared test for independence

Null H_0: Joint PMF is product of marginals, H_A: It is not

- Test statistic: $T = \sum_{i,j} \dfrac{(y_{ij} - np_{ij})^2}{np_{ij}}$
  - p_{ij}: product of marginals for (i, j)
  - np_{ij}: expected, if independent
  - We get $T = 0.012$
- Approximate distribution of T: Chi-squared with $(3-1)(7-1) = 12$ degrees of freedom
  - dof = (no of rows -1)(no of cols -1)

Test: Reject H_0 if $T > c$

CDF of χ^2_{12}

Significance level: $\alpha = 1 - F_{\chi^2_{12}}(c)$

P-**value**: $1 - F_{\chi^2_{12}}(0.012) = 0.999$, very good fit