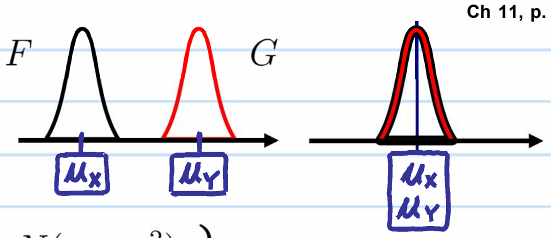
- Methods based on normality assumptions



•variable: data •parameter is fixed

- Assume that (1) F and G are normal, and (2) F and G have same variance.
- Thus, the statistical model is:

joint pdf ↕cf likelihood

$$\left.\begin{array}{lll} \text{1st sample:} & X_1,\ldots,X_n & \sim \text{ i.i.d. } N(\mu_X,\sigma^2) \\ \text{2nd sample:} & Y_1,\ldots,Y_m & \sim \text{ i.i.d. } N(\mu_Y,\sigma^2) \end{array}\right\} \Leftarrow \text{independent} \quad (*)$$

•variable: parameter •data is fixed

– This model contains three parameters: μ_X $(\in\mathbb{R})$, μ_Y $(\in\mathbb{R})$, σ^2 (>0).

– Under this model, the "difference" between F and G is simplified to be the difference between μ_X and μ_Y, i.e., $\Delta \equiv \mu_X - \mu_Y$ ($\Leftarrow$ called "effect"), and $\quad \mu_X - \mu_Y = 0 \;\Leftrightarrow\;$ no difference or no effect

10/18 a parameter estimation testing

Review 1 (estimation of the parameters in one-sample normal model)

Consider $X_1,\ldots,X_n \sim$ i.i.d. $N(\mu,\sigma^2)$, and the statistics
2 parameters cf → Data from S.r.s. (LN, CH7, p11)

Textbook Sec. 6.3

$$\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i \xrightarrow[\text{unbiased}]{e} \mu \quad \text{and} \quad s_X^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i-\overline{X})^2 \xrightarrow[\text{unbiased}]{e} \sigma^2$$

same as in survey sampling not S_X^2

- distribution (exercise)

$(x_1,\cdots,x_n)\in\mathbb{R}^n$
$(\underbrace{x_1-\overline{x}}_{u_1},\cdots,\underbrace{x_n-\overline{x}}_{u_n})\in$ an $(n-1)$-dim subspace of $\mathbb{R}^n$
$\because u_1+\cdots+u_n=0$

– $\overline{X}$ and s_X^2 are independent → joint = $\prod$ marginals

r.v. $\sigma^2 \times Z$, $Z\sim\chi^2_{n-1}$

– $\overline{X} \sim N(\mu,\sigma^2/n) \Rightarrow \sqrt{n}(\overline{X}-\mu)/\sigma \sim N(0,1)$ ← standardization

– $(n-1)s_X^2 \sim \sigma^2\chi^2_{n-1} \Rightarrow (n-1)s_X^2/\sigma^2 \sim \chi^2_{n-1}$; $n-1$: degrees of freedom

- $(T_1 = \sum_{i=1}^n X_i,\; T_2=\sum_{i=1}^n X_i^2)$ is a sufficient and complete statistic (exercise, **Hint**. 2-parameter exponential family) MATH2820, ⊠ (統計學)
- Optimality ⊠, LN, CH8, p57 ⊠, LN, CH8, p72-73

T_1/n

– $\overline{X}$ is the uniformly minimum variance unbiased estimator (UMVUE) of μ (exercise, **Hint**. Lehmann-Scheffe Thm)

– $\overline{X}$ is the maximum likelihood estimator (MLE) of μ (exercise, **Hint**.

joint pdf: $\prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$

$$\text{log-likelihood} \propto -\frac{n}{2}\log(\sigma^2) - \frac{\sum_{i=1}^n (X_i-\mu)^2}{2\sigma^2}\Bigg)$$

⊠, LN, CH8, P20-21

$\frac{1}{n-1}\left(T_2 - \frac{T_1^2}{n}\right)$

– s_X^2 is the UMVUE of σ^2 (exercise, **Hint**. Lehmann-Scheffe Thm)

– The MLE of σ^2 is $\dfrac{n-1}{n}s_X^2 = \dfrac{1}{n}\sum_{i=1}^n (X_i-\overline{X})^2$ (exercise)

Definition 1 (estimators of the parameters in the 2-sample normal model)

Under the two-sample normal model $(*)$ in LNp.6, → 3 parameters: μ_X, μ_Y, σ^2

- an intuitive estimator of μ_X is $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$,
- an intuitive estimator of μ_Y is $\overline{Y} = \frac{1}{m}\sum_{j=1}^m Y_j$,

treated as one-sample $\because \overline{X} \xrightarrow{e} \mu_X$, $\overline{Y} \xrightarrow{e} \mu_Y$

- since $s_X^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i-\overline{X})^2$ and $s_Y^2 = \frac{1}{m-1}\sum_{j=1}^m (Y_j-\overline{Y})^2$ estimate the same parameter σ^2, we can pool them to get a better estimator:

$n-1$, $m-1$: d.f.s of S_X^2, S_Y^2

$$s_p^2 = \underbrace{\frac{(n-1)}{(n-1)+(m-1)}s_X^2 + \frac{(m-1)}{(n-1)+(m-1)}s_Y^2}_{=1} = \frac{(n-1)s_X^2+(m-1)s_Y^2}{m+n-2}.$$

Note 1 (Some notes about the estimator of σ^2)

- s_p^2 is called the **pooled sample variance**
- s_p^2 is a weighted average of the sample variances of the X_i's and Y_j's, where
  - the weights are proportional to the degrees of freedom, it is appropriate since if one sample is of much larger size than the other, the estimate of σ^2 from that sample is more reliable $\Rightarrow$ it receives greater weight
  - since $E(s_X^2) = \sigma^2$ and $E(s_Y^2) = \sigma^2 \Rightarrow s_p^2$: an unbiased estimator of σ^2

from one-sample property in Review1 (LNp.6)

Theorem 1 (distributions of the parameter estimators, 2-sample normal model)

- Since $(X_1, \ldots, X_n)$, $(Y_1, \ldots, Y_m)$ are independent random variables

$\Rightarrow (\overline{X},\ s_X^2,\ \overline{Y},\ s_Y^2)$ are independent random variables

$\Rightarrow (\overline{X},\ \overline{Y},\ s_p^2)$ are independent random variables

$\because \overline{X}, \overline{Y}$ independent. $Var(\overline{X}-\overline{Y}) = Var(\overline{X}) + Var(\overline{Y}) = \sigma^2/n + \sigma^2/m$

- $\overline{X} \sim N(\mu_X, \sigma^2/n) \Rightarrow \sqrt{n}(\overline{X}-\mu_X)/\sigma \sim N(0,1)$
- $\overline{Y} \sim N(\mu_Y, \sigma^2/m) \Rightarrow \sqrt{m}(\overline{Y}-\mu_Y)/\sigma \sim N(0,1)$

standardization

- $\overline{X}-\overline{Y} \sim N\left(\mu_X - \mu_Y, \frac{\sigma^2}{n} + \frac{\sigma^2}{m}\right) \Rightarrow \frac{(\overline{X}-\overline{Y})-(\mu_X-\mu_Y)}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}} \sim N(0,1)$

$\Delta = \mu_X - \mu_Y \leftarrow e$

- Since **(i)** $(n-1)s_X^2/\sigma^2 \sim \chi^2_{n-1}$, **(ii)** $(m-1)s_Y^2/\sigma^2 \sim \chi^2_{m-1}$, and **(iii)** s_X^2 and s_Y^2 are independent,

$$\frac{(n-1)s_X^2 + (m-1)s_Y^2}{\sigma^2} = \frac{(m+n-2)s_p^2}{\sigma^2} \sim \chi^2_{m+n-2}$$

$S_p^2 \sim \frac{\sigma^2}{m+n-2}\chi^2_{m+n-2} \Leftarrow$

Theorem 2 (log-likelihood, 2-sample normal model)

testing $\rightarrow$ likelihood ratio
estimation $\rightarrow$ maximum likelihood

Under the two-sample normal model ($*$) in LNp.6, the log-likelihood is proportional to (exercise)

$$l(\mu_X, \mu_Y, \sigma^2) \propto -\frac{m+n}{2}\log(\sigma^2) - \frac{\sum_{i=1}^n (X_i-\mu_X)^2}{2\sigma^2} - \frac{\sum_{j=1}^m (Y_j-\mu_Y)^2}{2\sigma^2}$$

$$= -\frac{1}{2\sigma^2}\left(\sum_{i=1}^n X_i^2 + \sum_{j=1}^m Y_j^2\right) + \frac{\mu_X}{\sigma^2}\left(\sum_{i=1}^n X_i\right) + \frac{\mu_Y}{\sigma^2}\left(\sum_{j=1}^m Y_j\right)$$

$$-[(m+n)/2]\log(\sigma^2) - (n\,\mu_X^2)/(2\sigma^2) - (m\,\mu_Y^2)/(2\sigma^2)$$

$\sqcap$: data
$\sqcup$: parameter

$\left(-\frac{1}{2\sigma^2}, \frac{\mu_X}{\sigma^2}, \frac{\mu_Y}{\sigma^2}\right) \in (-\infty, 0)\times\mathbb{R}\times\mathbb{R}$

$\in$ 3-parameter exponential family $\sum_{i=1}^n x_i^2 - 2\mu_X\left(\sum_{i=1}^n x_i\right) + n\mu_X^2$

joint pdf: $\left[\prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu_X)^2}{2\sigma^2}}\right] \times \left[\prod_{j=1}^m \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y_j-\mu_Y)^2}{2\sigma^2}}\right]$

From the log-likelihood, we have

- $\frac{\partial l}{\partial \mu_X} = \frac{1}{\sigma^2}\left[\left(\sum_{i=1}^n X_i\right) - n \times \mu_X\right] = 0 \Rightarrow \hat{\mu}_{X,MLE} = \overline{X}$
- $\frac{\partial l}{\partial \mu_Y} = \frac{1}{\sigma^2}\left[\left(\sum_{j=1}^m Y_j\right) - m \times \mu_Y\right] = 0 \Rightarrow \hat{\mu}_{Y,MLE} = \overline{Y}$
- $\frac{\partial l}{\partial \sigma^2} = -\frac{m+n}{2\sigma^2} + \frac{\sum_{i=1}^n (X_i-\mu_X)^2}{\cancel{2}\sigma^{4}} + \frac{\sum_{j=1}^m (Y_j-\mu_Y)^2}{\cancel{2}\sigma^{4}} = 0 \Rightarrow \hat{\sigma}^2_{MLE} = \frac{n+m-2}{n+m} S_p^2$

Theorem 3 (UMVUE and MLE of the parameters in the 2-sample normal model)

- $\left(R_1 = \sum_{i=1}^n X_i, R_2 = \sum_{j=1}^m Y_j, R_3 = \sum_{i=1}^n X_i^2 + \sum_{j=1}^m Y_j^2\right)$ is a sufficient and complete statistic (**Hint**. 3-parameter exponential family)
- $\overline{X}$ $(= R_1/n)$ is the UMVUE (by Lehmann-Scheffe Thm) and MLE of μ_X
- $\overline{Y}$ $(= R_2/m)$ is the UMVUE (by Lehmann-Scheffe Thm) and MLE of μ_Y

- (by Lehmann-Scheffe Thm) The pooled sample variance s_p^2 is the UMVUE of σ^2, since **(i)** s_p^2 is unbiased, and **(ii)** Check Note 1 (LNp.8)

$$(m+n-2)s_p^2 = (n-1)s_X^2 + (m-1)s_Y^2$$
$$= \sum_{i=1}^n (X_i - \overline{X})^2 + \sum_{j=1}^m (Y_j - \overline{Y})^2$$
$$= \left(\sum_{i=1}^n X_i^2\right) - n\overline{X}^2 + \left(\sum_{j=1}^m Y_j^2\right) - m\overline{Y}^2 = R_3 - (R_1^2/n) - (R_2^2/m)$$

$n\overline{X}^2 = n\left(\frac{\sum_{i=1}^n X_i}{n}\right)^2 = n\left(\frac{R_1}{n}\right)^2$

not unbiased

- The MLE of σ^2 is $\dfrac{m+n-2}{m+n}\, s_p^2 = \dfrac{(n-1)s_X^2 + (m-1)s_Y^2}{m+n}$ $\xleftarrow{\text{cf.}}$ $m+n-2$

Question 2 (how to claim Δ=0 or Δ≠0?)

Under the two-sample normal model (∗) in LNp.6, consider the parameter

$$\Delta = \mu_X - \mu_Y.$$

invariance property of MLE: ☒. LN, CH8, p.19 $\hat{\theta}_{MLE} \xrightarrow{e} \theta \to g(\theta)$ ⇒ MLE of $g(\theta)$ is $g(\hat{\theta}_{MLE})$

Notice that

$$\Delta = 0 \quad \Leftrightarrow \quad \text{no difference in the two samples}$$

estimation of Δ

- The UMVUE (by Lehmann-Scheffe Thm and $\hat{\Delta} = R_1/n - R_2/m$) and MLE of Δ is $\hat{\Delta} = \overline{X} - \overline{Y}$. ← point estimator

Recall. duality between C.I. and testing

$\hat{\Delta}$ ~ normal in Thm 1 (LNp.8)

- But, $\hat{\Delta} \neq 0$ is not a strong enough evidence to reject $\Delta = 0$ (**Note.** $P(\hat{\Delta} \neq 0) = 1$). A better way is to examine if a C.I. of Δ contains 0.
- Q: how to construct an interval estimator for Δ? ← interval estimator

Review 2 (pivotal quantity of θ)

Recall, LN, CH7, p.33

A **pivotal quantity** for θ is a function of data $X_1, \ldots, X_n$ and the parameter θ, denoted by

$$Q(\mathbf{X}, \theta) = Q(X_1, \ldots, X_n, \theta),$$ ← a r.v., but not a statistic

if the distribution of $Q(\mathbf{X}, \theta)$ is irrelevant to *all* parameters.

Theorem 4 (confidence interval of Δ, 2-sample normal model)

Under the two-sample normal model (∗) in LNp.6,

- σ^2 known (σ^2 is not a parameter) — parameters μ_X, μ_Y

Recall. distribution of $\overline{X}-\overline{Y}$ in Thm 1 (LNp.8)

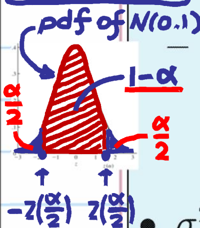
  – a pivotal quantity of Δ is

function of data & Δ only

$$Q_{Z,\Delta} = \frac{(\overline{X}-\overline{Y}) - (\mu_X - \mu_Y)}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}} = \frac{(\overline{X}-\overline{Y}) - \Delta}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}} \sim N(0,1)$$

irrelevant to μ_X, μ_Y

a known constant; std. error of $\overline{X}-\overline{Y}$



  – a $100(1-\alpha)\%$ C.I. for Δ is $(\overline{X}-\overline{Y}) \pm z(\alpha/2) \times \left(\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}\right)$ since

$$1-\alpha = P\left(\,|Q_{Z,\Delta}| < z(\alpha/2)\right)$$ ← data: fixed, Δ: changed

known

$$= P\left((\overline{X}-\overline{Y}) - z(\alpha/2)\sigma\sqrt{\tfrac{1}{n}+\tfrac{1}{m}} < \Delta < (\overline{X}-\overline{Y}) + z(\alpha/2)\sigma\sqrt{\tfrac{1}{n}+\tfrac{1}{m}}\right)$$

parameter — function of data only

- σ^2 unknown (σ^2 is a parameter)

irrelevant to μ_X, μ_Y, σ^2

  – a pivotal quantity of Δ is

function of data & Δ only

$$Q_{T,\Delta} = \frac{(\overline{X}-\overline{Y}) - (\mu_X - \mu_Y)}{s_p\sqrt{\frac{1}{n}+\frac{1}{m}}} = \frac{\left[(\overline{X}-\overline{Y}) - \Delta\right] / \left(\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}\right)}{\sqrt{\left[\frac{(m+n-2)s_p^2}{\sigma^2}\right]\frac{1}{m+n-2}}} \sim t_{m+n-2}$$

~N(0,1); a parameter; estimated std. error of $\overline{X}-\overline{Y}$; $\sim\chi^2_{m+n-2}$ (Thm 1, LNp.8); TBp.193; independent (Thm 1, LNp.8)

C.I. when σ^2 known c.f.

• a $100(1-\alpha)\%$ C.I. for Δ is $(\overline{X}-\overline{Y}) \pm t_{m+n-2}(\alpha/2) \times \left(s_P\sqrt{\frac{1}{n}+\frac{1}{m}}\right)$

$\because t$ is a distribution symmetric about 0

estimated std. error of $\overline{X}-\overline{Y}$

Note 2 (A note about the confidence intervals of Δ)

These confidence intervals are of the form

(estimate) $\pm$ (critical value) $\times$ [(estimted) standard error],

evaluates the accuracy of estimator

where the (estimated) standard error of $\overline{X}-\overline{Y}$ is $\sigma_{\overline{X}-\overline{Y}} = \sigma\sqrt{\frac{1}{n}+\frac{1}{m}}$ when σ^2 is known, and is $s_{\overline{X}-\overline{Y}} = s_P\sqrt{\frac{1}{n}+\frac{1}{m}}$ when σ^2 is unknown.

Example 2 (confidence interval of Δ, heat of fusion of ice, cont. Ex.1 in LNp.3)

Statistical modeling: assume 2-sample normal model (*) in LNp.6

- $n=13$, $\overline{X}_A = 80.02$, $s_A = 0.024$; $m=8$, $\overline{X}_B = 79.98$, $s_B = 0.031$
  ($\overline{X}_A \xrightarrow{e} \mu_X$, $\overline{X}_B \xrightarrow{e} \mu_Y$)
- $s_P = \sqrt{\frac{12}{19}s_A^2 + \frac{7}{19}s_B^2} = 0.027 \xrightarrow{e} \sigma$, $s_{\overline{X}_A-\overline{X}_B} = s_P\sqrt{\frac{1}{13}+\frac{1}{8}} = 0.012 \xrightarrow{e} \sigma_{\overline{X}-\overline{Y}}$
- A 95% confidence interval for $\Delta = \mu_A - \mu_B$ is side-by-side box-plots in LNp.4 c.f.
  $(\overline{X}_A - \overline{X}_B) \pm t_{19}(0.025) \times s_{\overline{X}_A-\overline{X}_B} = (0.04) \pm (2.093) \times (0.012) = (0.015, 0.065)$.

Question 3 (how to perform testing of Δ=0?)

Note: $0 \notin (0.015, 0.065) \Rightarrow$ reject $\Delta = 0$

- **Recall.** duality between confidence interval and hypothesis testing
- Q: What are the hypothesis testings corresponding to these confidence intervals of Δ?

Theorem 5 (z-test and t-test for Δ=Δ_0, 2-sample normal model)

Under the two-sample normal model (*) in LNp.6, consider the null and alternative hypotheses:

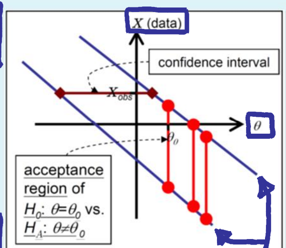
$$H_0: \mu_X - \mu_Y = \Delta = \Delta_0 \quad \text{vs.} \quad H_A: \mu_X - \mu_Y = \Delta \neq \Delta_0$$

where Δ_0 is a known constant (**Note.** if $\Delta_0 = 0$, $H_0: \mu_X = \mu_Y$ vs. $H_A: \mu_X \neq \mu_Y$), and H_A is a **two-sided alternative**. From the duality between C.I. and testing,

C.I. fixed changed | testing changed fixed

$|Q_{Z_0,\Delta}| < z(\alpha/2)$ $\xleftrightarrow{\text{cf.}}$ $|Q_{Z,\Delta_0}| < z(\alpha/2)$

$|Q_{T_0,\Delta}| < t_{m+n-2}(\alpha/2)$ $\xleftrightarrow{\text{cf.}}$ $|Q_{T,\Delta_0}| < t_{m+n-2}(\alpha/2)$



the corresponding test of these confidence intervals are:

- test statistic
  – σ^2 known: $Z = \dfrac{(\overline{X}-\overline{Y})-\Delta_0}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}}$ (cf. $Q_{Z,\Delta}$ in LNp.11) — pivotal quantity; std. error of $\overline{X}-\overline{Y}$
  – σ^2 unknown: $T = \dfrac{(\overline{X}-\overline{Y})-\Delta_0}{s_P\sqrt{\frac{1}{n}+\frac{1}{m}}}$ (cf. $Q_{T,\Delta}$ in LNp.11) — estimated std. error of $\overline{X}-\overline{Y}$

Q: Are they statistics?

$|Q_{Z,\Delta}| < z(\alpha/2)$
$|Q_{T,\Delta}| < t_{n+m-2}(\alpha/2)$

- null distribution (check Thm4 (LNp.11))
  $\Delta = \Delta_0$, $\overline{X}-\overline{Y} \sim N(\Delta_0, \sigma^2(1/n+1/m))$
  – σ^2 known: under H_0, $Z \sim N(0,1)$
  – σ^2 unknown: under H_0, $T \sim t_{m+n-2}$ $\leftarrow \dfrac{[(\overline{X}-\overline{Y})-\Delta_0]/\sigma\sqrt{1/n+1/m}}{\sqrt{(n+m-2)S_P^2/\sigma^2(n+m-2)}}$
- level-α rejection region — $\Delta = \Delta_0$, Thm4 (LNp.11)
  – σ^2 known: $|Z| > z(\alpha/2)$, called z-test (reasonable?)
  – σ^2 unknown: $|T| > t_{m+n-2}(\alpha/2)$, called t-test (reasonable?)

Why $\alpha/2$? $\because$ 2-sided alternative

Note. The t-test (or z-test) rejects H_0 if and only if its corresponding C.I. does not include Δ_0.