單元 2: 與常態分布相關的抽樣分布
(課本 §7.2)

定理 7.1. 若隨機樣本

$$Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$$

則統計量 (樣本期望值)

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \sim N(\mu, \sigma^2/n)$$

註. 標準化 (常態化, normalizing), 得

$$\sqrt{n}\left(\frac{\overline{Y}-\mu}{\sigma}\right) = \frac{\overline{Y}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

例 1. 設一裝瓶機被設定成每瓶的平均注入量爲 μ 盎司且標準差 σ 爲 1 盎司的常態分布. 今隨機選取裝好的 9 瓶, 其容量分別爲 $y_1, \ldots, y_9$, 並令樣本期望值

$$\overline{y} = \frac{1}{9}(y_1 + \cdots + y_9)$$

(a) 試求樣本期望值與眞實期望值 μ 的差在 0.3 盎司內的機率.

(b) 試問需取多大的樣本以致於誤差在 0.3 盎司內的機率爲 0.95?

定理 7.2. 若隨機樣本

$$Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$$

則

$$\sum_{i=1}^{n} \left(\frac{Y_i - \mu}{\sigma} \right)^2 \sim \chi^2(n)$$

稱作自由度爲 n 的卡方分布.

註. 由表 6, 得

$$P(\chi^2(n) \geq \chi^2_\alpha) = \alpha$$

如圖示. 如, $n = 10$,

$$P(\chi^2(10) > 4.86518) = 0.90$$

或由

$$\text{R}: \texttt{pchisq}(\texttt{y}_0, \texttt{n}) = P(\chi^2(n) \leq y_0) \overset{\text{記作}}{=} p$$

計算 $\chi^2(n)$ 的 cdf, 即給定 y_0, 求 p.

另 p-th quantile

$$\phi_p = \mathtt{qchisq(p, n)}$$

如圖示, 即

$$P(\chi^2(n) \leq \phi_p) = p$$

計算 inverse cdf 的值, 即給定 p, 求 ϕ_p.

例 2. 令

$$Z_1, \ldots, Z_6 \overset{\text{iid}}{\sim} N(0, 1)$$

試求 b 使得

$$P\left(\sum_{i=1}^{6} Z_i^2 \leq b\right) = 0.95$$

定義. 樣本變異數

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \overline{Y})^2$$

可作爲母體變異數 σ^2 的一個好估計 (詳述於第 8 章).

定理 7.3 若隨機樣本

$$Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$$

則

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i-\overline{Y})^2=\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$$

且 $\overline{Y}$ 與 S^2 相互獨立.

註. 由定理 7.2,

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i-\mu)^2\sim\chi^2(n)$$

但

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i-\overline{Y})^2\sim\chi^2(n-1)$$

自由度少 1, 因爲 μ 被已知的 $\overline{Y}$ 取代.

例 3. 續例 1 的裝瓶機問題. 今隨機選取 10 瓶, 並計算 S^2. 試求 b_1 與 b_2 使得

$$P(b_1\le S^2\le b_2)=0.90$$

定義. 令 $Z\sim N(0,1)$ 且考慮自由度爲 ν 的卡方分布隨機變數 $\chi^2(\nu)$. 若 Z 與 $\chi^2(\nu)$ 相互獨立, 則定義

$$T(\nu)=\frac{Z}{\sqrt{\chi^2(\nu)/\nu}}$$

並稱其分布爲自由度爲 ν 的 t 分布.

註 1. 若

$$Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$$

則

$$\sqrt{n}\left(\frac{\overline{Y} - \mu}{S}\right) \sim T(n-1)$$

註 2. $N(0,1)$ 與 $T(\nu)$ 的比較:

(1) $T(\nu)$ 的 pdf (參看練習題 7.98).

(2) 當 $\nu > 1$ 時,

$$E(T) = 0$$

且當 $\nu > 2$ 時,

$$\text{Var}(T) = \frac{\nu}{\nu - 2} = 1 + \frac{2}{\nu - 2}$$

恆大於 1, 且爲 ν 的遞減函數 (參看練習題 7.30).

(3) 由圖形, $N(0,1)$ 較集中, $T(\nu)$ 有較多的機率質量在尾部 (或稱有肥大的尾部, fat tails).

註 3. 由表 5,

$$P(T(\nu) > t_\alpha) = \alpha$$

如圖示.

註 4. 若 σ^2 已知, 則採用

$$\sqrt{n}\left(\frac{\overline{Y} - \mu}{\sigma}\right) \sim N(0,1)$$

對 μ 推論.

若 σ^2 未知, 則採用

$$\sqrt{n}\left(\frac{\overline{Y} - \mu}{S}\right) \sim T(n-1)$$

對 μ 推論.

例 4. 若某金屬線的

$$\text{拉張強度 (tensile strength)} \sim N(\mu, \sigma^2)$$

其中 μ 與 σ^2 均未知. 隨機選取 6 段金屬線並令 Y_i 爲第 i 段的拉張強度, $i = 1, \ldots, 6$. 則

$$\mu \approx \overline{Y}, \quad \sigma^2 \approx S^2$$

且

$$\sigma^2_{\overline{Y}} = \frac{\sigma^2}{n} \approx \frac{S^2}{n}$$

試求 $\overline{Y}$ 與眞正母體期望値 μ 的差在 $2\frac{S}{\sqrt{n}}$ 個單位內的機率.

令

$$X_1, \ldots, X_{n_1} \overset{\text{iid}}{\sim} N(\mu_1, \sigma_1^2)$$

即由母體 1 所得的隨機樣本, 且

$$Y_1, \ldots, Y_{n_2} \overset{\text{iid}}{\sim} N(\mu_2, \sigma_2^2)$$

即由母體 2 所得的隨機樣本.

問. 如何比較 σ_1^2 與 σ_2^2?

答. 一個直觀的作法爲, 由 $X_1, \ldots, X_{n_1}$ 得 S_1^2 $(\approx \sigma_1^2)$ 且由 $Y_1, \ldots, Y_{n_2}$ 得 S_2^2 $(\approx \sigma_2^2)$ 並考慮 S_1^2/S_2^2. 但更好的作法爲考慮

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \left(\frac{\sigma_2^2}{\sigma_1^2}\right)\left(\frac{S_1^2}{S_2^2}\right)$$

即與各自變異數相比後的比值, 且其分布爲自由度爲 (n_1-1) 與 (n_2-1) 的 F-分布, 並記作

$$\left(\frac{\sigma_2^2}{\sigma_1^2}\right)\left(\frac{S_1^2}{S_2^2}\right) \sim F(n_1-1, n_2-1)$$

定義. 考慮隨機變數 $\chi_1^2(\nu_1)$ 與 $\chi_2^2(\nu_2)$. 若 $\chi_1^2(\nu_1)$ 與 $\chi_2^2(\nu_2)$ 相互獨立, 則 定義

$$F(\nu_1, \nu_2) = \frac{\chi_1^2(\nu_1)/\nu_1}{\chi_2^2(\nu_2)/\nu_2}$$

並稱其分布爲 (分子) 自由度爲 ν_1 且 (分母) 自由度爲 ν_2 的 F-分布.

註 1. 若

$$X_1, \ldots, X_{n_1} \overset{\text{iid}}{\sim} N(\mu_1, \sigma_1^2)$$

以及

$$Y_1, \ldots, Y_{n_2} \overset{\text{iid}}{\sim} N(\mu_2, \sigma_2^2)$$

且它們相互獨立, 則

$$(n_1-1)S_1^2/\sigma_1^2 \sim \chi_1^2(n_1-1)$$

且

$$(n_2-1)S_2^2/\sigma_2^2 \sim \chi_2^2(n_2-1)$$

以及 S_1^2 與 S_2^2 相互獨立. 故由 F 分布的定義,

$$\begin{aligned}&\frac{(n_1-1)S_1^2/[\sigma_1^2(n_1-1)]}{(n_2-1)S_2^2/[\sigma_2^2(n_2-1)]}\\ &\qquad\qquad = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1-1, n_2-1)\end{aligned}$$

其中 n_1-1 稱作分子自由度, n_2-1 稱作分母自由度.

註 2. 由表 7,

$$P(F(\nu_1,\nu_2) > F_\alpha) = \alpha$$

如圖示.

例 5. 設由兩個有相同母體變異數的常態母體中, 分別取出大小爲 $n_1 = 6$ 與 $n_2 = 10$ 的隨機樣本. 試求 b 使得

$$P(S_1^2/S_2^2 \le b) = 0.95$$