詳細解答

STEP III 2004，第 14 題

概率與統計 • 假設檢驗的錯誤

> **STEP III 2004，第 14 題**
>
> 繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：
>
> step.maths.org/questions/04-s3-q14

關鍵思路

在兩個假設之下，$\overline{X}$ 的中心同樣是 μ；兩者的分別只在於 $\overline{X}$ 有多分散。因此，否定 H_0 的證據就是 $\overline{X}$ 落在遠離 μ 的位置，不論在哪一邊，而對稱的雙側拒絕域（rejection region）正好能偵測這種情況。之後的一切，都歸結為把 $\overline{X}$ 標準化（standardise），在每個假設之下各做一次。

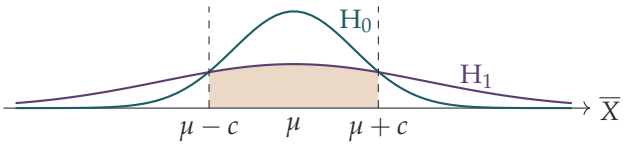
解答

為何採用雙側檢驗

由於 $X \sim N(\mu, \sigma^2)$，而且各樣本值互相獨立，樣本平均值（sample mean）滿足

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

無論在哪一個假設之下，$\overline{X}$ 都以已知的值 μ 為中心；唯一的分別是在 H_1 之下，它的標準差（standard deviation）$\sigma_1/\sqrt{n}$ 比 $\sigma_0/\sqrt{n}$ 大。$\overline{X}$ 接近 μ 的值在 H_0 之下相對較有可能出現，而遠離 μ 的值則在 H_1 之下相對較有可能出現（較闊的曲線尾部較厚；見附圖）。由於兩個分佈都關於 μ 對稱，一個很大的 $\overline{X}$ 值與一個同樣程度地小的值相比，並不是支持 H_1 的更強證據。所以，自然的檢驗方法是當 $|\overline{X} - \mu|$ 很大時拒絕 H_0，不論方向；而根據對稱性，接受區間應以 μ 為中心：當 $\mu - c < \overline{X} < \mu + c$ 時接受 H_0。在這裏採用單側檢驗（one-sided test）是錯的，因為這樣會忽略一半支持分散程度較大的證據。



陰影部分的面積（H_1 為真卻接受 H_0）是 β；兩條虛線以外的藍綠色曲線下面積是 α。

求 c

在 H_0 之下，$Z = \dfrac{\sqrt{n}\,(\overline{X} - \mu)}{\sigma_0}$ 服從標準正態分佈（standard normal distribution）。當 $|\overline{X} - \mu| \geq c$ 時，即當 $|Z| \geq c\sqrt{n}/\sigma_0$ 時，我們接受 H_1。根據標準正態分佈的對稱性，

$$\alpha = \mathrm{P}\left(|Z| \geq \frac{c\sqrt{n}}{\sigma_0}\right) = 2\left(1 - \Phi\left(\frac{c\sqrt{n}}{\sigma_0}\right)\right), \qquad \text{所以} \qquad \Phi\left(\frac{c\sqrt{n}}{\sigma_0}\right) = 1 - \tfrac{1}{2}\alpha.$$

由於 Φ 嚴格遞增，這表示 $c\sqrt{n}/\sigma_0 = z_\alpha$，因此

$$c = \frac{z_\alpha\,\sigma_0}{\sqrt{n}}.$$

β 與 n 無關

在 H_1 之下，$W = \dfrac{\sqrt{n}\,(\overline{X} - \mu)}{\sigma_1}$ 服從標準正態分佈，而

$$\begin{aligned}\beta = \mathrm{P}(|\overline{X} - \mu| < c) &= \mathrm{P}\left(|W| < \frac{c\sqrt{n}}{\sigma_1}\right) = \mathrm{P}\left(|W| < \frac{z_\alpha\sigma_0}{\sigma_1}\right) \\ &= 2\,\Phi\left(\frac{z_\alpha\sigma_0}{\sigma_1}\right) - 1,\end{aligned}$$

其中我們代入了上一部分得到的 $c\sqrt{n} = z_\alpha\sigma_0$。$c$ 之中的 $\sqrt{n}$ 與標準化時出現的 $\sqrt{n}$ 互相抵消，所以 β 只取決於 α 和 σ_0/σ_1，與 n 無關。

想一想 樣本愈大，檢驗理應愈好吧？

通常確是如此，因為隨着 n 增大，兩個抽樣分佈（sampling distribution）會互相分開。但在這裏它們永遠不會分開：兩者都以 μ 為中心，而增大 n 會令兩個標準差按同一個因子 $\sqrt{n}$ 縮小。整幅圖只是按比例縮放，所以面積 α 和 β 都不變。用樣本平均值來檢驗方差（variance），是一個差劣的統計量；它把各個數值所包含的關於分散程度的資訊都丟掉了。

比值 σ_1/σ_0 的最小值

設 $r = \sigma_1/\sigma_0 > 1$。

條件 $\alpha < 0.05$。這表示 $\Phi(z_\alpha) = 1 - \frac{1}{2}\alpha > 0.975 \approx \Phi(1.960)$，所以，由於 Φ 遞增，$z_\alpha > 1.960$。

條件 $\beta < 0.05$。這表示 $2\Phi(z_\alpha/r) - 1 < 0.05$，即 $\Phi(z_\alpha/r) < 0.525 \approx \Phi(0.063)$，所以 $z_\alpha/r < 0.063$，亦即

$$r > \frac{z_\alpha}{0.063}.$$

綜合兩者，我們需要 $r > z_\alpha/0.063 > 1.960/0.063$。反過來說，若 $r > 1.960/0.063$，我們可以選取剛好大於 1.960 且滿足 $z_\alpha < 0.063r$ 的 z_α，這樣兩個條件都成立。因此這個比值必須大於

$$\frac{1.960}{0.063} \approx 31.1,$$

所以 σ_1/σ_0 的最小值約為 31。兩個標準差必須相差約三十倍，樣本平均值才能如此可靠地把它們區分開來，而且不論樣本大小如何都是這樣。

驗算

- 以 $\mu = 3$、$\sigma_0 = 1$、$\sigma_1 = 2.5$、$\alpha = 0.1$ 及 $n = 1, 4, 9, 25$ 進行模擬，在 H_0 之下的拒絕率每次都約為 0.10，而在 H_1 之下的接受率對每個 n 都約為 0.489，與 $2\Phi(1.645/2.5) - 1 \approx 0.489$ 吻合。
- 當 $r = 1.960/0.063$ 而 $\alpha = 0.05$ 恰好成立時，$\beta = 2\Phi(0.063) - 1 \approx 0.050$，一如所料正好在邊界上。若使用 Φ 的精確反函數值，會得到 $r \approx 31.3$，所以答案 31 不受題目所給表值的捨入誤差影響。
- 合理性檢查：若 $\sigma_1 = \sigma_0$，則 $\beta = 2\Phi(z_\alpha) - 1 = 1 - \alpha$，當兩個假設重合時必然如此。

常見錯誤

- 使用單側的拒絕域，例如 $\overline{X} > \mu + c$：在 H_1 之下平均值並沒有移動，所以沒有理由偏重其中一邊的尾部。
- 寫成 $\alpha = 1 - \Phi(c\sqrt{n}/\sigma_0)$，忘記了拒絕域有兩個尾部。
- 把關於 r 的不等式方向弄反：β 小需要 z_α/r 小，所以 r 要大。

STEP 之外

這道題展示了兩類錯誤之間的取捨，以及檢驗統計量（test statistic）的選擇為何重要。奈曼–皮爾遜引理（Neyman–Pearson lemma）指出，最強大的檢驗是比較兩個假設之下的似然（likelihood）；在這裏，這會引出統計量 $\sum(X_i - \mu)^2$，它服從的 χ^2_n 分佈確實會隨

着 n 增大而把兩個假設區分開來，這與 $\bar{X}$ 不同。想找一本易讀的入門書了解這些概念，可參閱 Blitzstein 與 Hwang 的 *Introduction to Probability*（概率論入門），或 Grimmett 與 Welsh 的 *Probability: An Introduction*（概率導論）。

Beyond STEP 詳細解答：https://beyondstep.keithmcnulty.org/zh-hant/。解答文字採用 CC0 1.0；題目本身版權屬 Cambridge University Press & Assessment 所有（©），本文件不予轉載。由機器從英文翻譯而成；尚未經審閱。