详细解答

STEP III 2004，第 14 题

概率与统计 • 假设检验的错误

STEP III 2004，第 14 题

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/04-s3-q14

关键思路

在两个假设之下，$\overline{X}$ 的中心同样是 μ；两者的分别只在于 $\overline{X}$ 有多分散。因此，否定 H_0 的证据就是 $\overline{X}$ 落在远离 μ 的位置，不论在哪一边，而对称的双侧拒绝域（rejection region）正好能侦测这种情况。之后的一切，都归结为把 $\overline{X}$ 标准化（standardise），在每个假设之下各做一次。

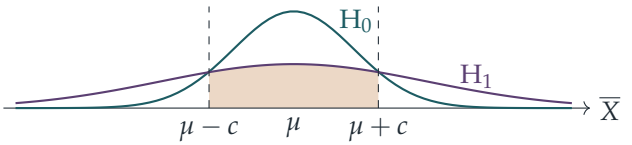
解答

为何采用双侧检验

由于 $X \sim \mathrm{N}(\mu, \sigma^2)$，而且各样本值互相独立，样本平均值（sample mean）满足

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

无论在哪一个假设之下，$\overline{X}$ 都以已知的值 μ 为中心；唯一的分别是在 H_1 之下，它的标准差（standard deviation）$\sigma_1/\sqrt{n}$ 比 $\sigma_0/\sqrt{n}$ 大。$\overline{X}$ 接近 μ 的值在 H_0 之下相对较有可能出现，而远离 μ 的值则在 H_1 之下相对较有可能出现（较阔的曲线尾部较厚；见附图）。由于两个分布都关于 μ 对称，一个很大的 $\overline{X}$ 值与一个同样程度地小的值相比，并不是支持 H_1 的更强证据。所以，自然的检验方法是当 $|\overline{X} - \mu|$ 很大时拒绝 H_0，不论方向；而根据对称性，接受区间应以 μ 为中心：当 $\mu - c < \overline{X} < \mu + c$ 时接受 H_0。在这里采用单侧检验（one-sided test）是错的，因为这样会忽略一半支持分散程度较大的证据。



阴影部分的面积（H_1 为真却接受 H_0）是 β；两条虚线以外的蓝绿色曲线下面积是 α。

求 c

在 H_0 之下，$Z = \dfrac{\sqrt{n}\,(\overline{X} - \mu)}{\sigma_0}$ 服从标准正态分布（standard normal distribution）。当 $|\overline{X} - \mu| \geq c$ 时，即当 $|Z| \geq c\sqrt{n}/\sigma_0$ 时，我们接受 H_1。根据标准正态分布的对称性，

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

由于 Φ 严格递增，这表示 $c\sqrt{n}/\sigma_0 = z_\alpha$，因此

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

β 与 n 无关

在 H_1 之下，$W = \dfrac{\sqrt{n}\,(\overline{X} - \mu)}{\sigma_1}$ 服从标准正态分布，而

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

其中我们代入了上一部分得到的 $c\sqrt{n} = z_\alpha\sigma_0$。$c$ 之中的 $\sqrt{n}$ 与标准化时出现的 $\sqrt{n}$ 互相抵消，所以 β 只取决于 α 和 σ_0/σ_1，与 n 无关。

想一想 样本愈大，检验理应愈好吧?

通常确是如此，因为随着 n 增大，两个抽样分布（sampling distribution）会互相分开。但在这里它们永远不会分开：两者都以 μ 为中心，而增大 n 会令两个标准差按同一个因子 $\sqrt{n}$ 缩小。整幅图只是按比例缩放，所以面积 α 和 β 都不变。用样本平均值来检验方差（variance），是一个差劣的统计量；它把各个数值所包含的关于分散程度的信息都丢掉了。

比值 σ_1/σ_0 的最小值

设 $r = \sigma_1/\sigma_0 > 1$。

条件 $\alpha < 0.05$。这表示 $\Phi(z_\alpha) = 1 - \frac{1}{2}\alpha > 0.975 \approx \Phi(1.960)$，所以，由于 Φ 递增，$z_\alpha > 1.960$。

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

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

综合两者，我们需要 $r > z_\alpha/0.063 > 1.960/0.063$。反过来说，若 $r > 1.960/0.063$，我们可以选取刚好大于 1.960 且满足 $z_\alpha < 0.063\,r$ 的 z_α，这样两个条件都成立。因此这个比值必须大于

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

所以 σ_1/σ_0 的最小值约为 31。两个标准差必须相差约三十倍，样本平均值才能如此可靠地把它们区分开来，而且不论样本大小如何都是这样。

验算

- 以 $\mu = 3$、$\sigma_0 = 1$、$\sigma_1 = 2.5$、$\alpha = 0.1$ 及 $n = 1, 4, 9, 25$ 进行模拟，在 H_0 之下的拒绝率每次都约为 0.10，而在 H_1 之下的接受率对每个 n 都约为 0.489，与 $2\Phi(1.645/2.5) - 1 \approx 0.489$ 吻合。
- 当 $r = 1.960/0.063$ 而 $\alpha = 0.05$ 恰好成立时，$\beta = 2\Phi(0.063) - 1 \approx 0.050$，一如所料正好在边界上。若使用 Φ 的精确反函数值，会得到 $r \approx 31.3$，所以答案 31 不受题目所给表值的舍入误差影响。
- 合理性检查：若 $\sigma_1 = \sigma_0$，则 $\beta = 2\Phi(z_\alpha) - 1 = 1 - \alpha$，当两个假设重合时必然如此。

常见错误

- 使用单侧的拒绝域，例如 $\overline{X} > \mu + c$：在 H_1 之下平均值并没有移动，所以没有理由偏重其中一边的尾部。
- 写成 $\alpha = 1 - \Phi(c\sqrt{n}/\sigma_0)$，忘记了拒绝域有两个尾部。
- 把关于 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

这道题展示了两类错误之间的取舍，以及检验统计量（test statistic）的选择为何重要。奈曼–皮尔逊引理（Neyman–Pearson lemma）指出，最强大的检验是比较两个假设之下的似然（likelihood）；在这里，这会引出统计量 $\sum(X_i - \mu)^2$，它服从的 χ^2_n 分布确实会随着

n 增大而把两个假设区分开来，这与 $\overline{X}$ 不同。想找一本易读的入门书了解这些概念，可参阅 Blitzstein 与 Hwang 的 *Introduction to Probability*（概率论入门），或 Grimmett 与 Welsh 的 *Probability: An Introduction*（概率导论）。

Beyond STEP 详细解答：https://beyondstep.keithmcnulty.org/zh-hans/。解答文字采用 CC0 1.0；题目本身版权属 Cambridge University Press & Assessment 所有（©），本文件不予转载。由机器从英文翻译而成；尚未经审阅。