清华大学电子工程系

概率论与数理统计

2020 年春季学期

作业 13

王哲凡 2019011200 2020 年 5 月 21 日

13.1. *Solution.* 设 $H_0 : X \sim P(\lambda), H_1 :$ 总体不服从 Poisson 分布.

假设 H_0 成立，对于参数 λ，其似然函数为：

$$L(\lambda) = \frac{\lambda^{\sum\limits_{i=1}^{n} X_i}}{X_1! \cdots X_n!} e^{-n\lambda}$$

故：

$$\frac{\partial \ln L}{\partial \lambda} = -n + \frac{1}{\lambda} \sum_{i=1}^{n} X_i = 0 \Rightarrow \lambda^* = \overline{X}$$

因此 $\lambda^* = \overline{x} = 0.805$.

分为 $0, 1, 2$ 和 3 以及上四个单元，E_i 分别为：

$$E_1 \approx 89.4176, E_2 \approx 71.9812, E_3 \approx 28.9724, E_4 \approx 9.6288$$

可得：

$$\chi_0^2 \approx 0.9113 < \chi_{0.05}^2(4 - 1 - 1) \approx 5.99$$

也就是不拒绝 H_0，则可以认为其服从泊松分布.

13.2. *Solution.* 将数据等分成五个单元，E_i 均为 5，计算：

$$\chi_0^2 = \sum_{i=1}^{5} \frac{(O_i - 5)^2}{5} = 1.2$$

其分布应该趋向于 $\chi^2(1)$，而 $\chi_{0.25}^2(1) \approx 1.32, \chi_{0.5}^2(1) \approx 0.45$，即 P 值在 $0.5 \sim 0.75$ 之间，因此不拒绝原假设.

即数据不是伪造的.

13.3. *Solution.* 计算其 χ_0^2：

$$\chi_0^2 \approx 0.5103$$

其分布应该趋向于 $\chi^2(3)$，故其 P 值约为 $0.0834 > \alpha$.

因此证据指出其不是伪造的，不够确凿.

13.4. *Solution.* 设 $H_0 :$ 独立，$H_1 :$ 不独立.

计算 P_{ij}^*，并由此得到 χ_0^2：

$$\chi_0^2 \approx 3.1892$$

自由度为 $(a - 1)(b - 1) = 2$.

故 P 值 $= P(\chi^2 \geq \chi_0^2) \approx 0.2030 > \alpha$.

因此不拒绝原假设，即慢性支气管炎与吸烟量无关.

13.5. *Solution.* 设 H_0 : 想法相同，H_1 : 想法不同.

计算 P_i^* 和 E_{ij}，并由此得到 χ_0^2：

$$\chi_0^2 \approx 91.7155$$

自由度为 $(a-1)(b-1) = 2$.

故 P 值 $= P(\chi^2 \geq \chi_0^2) \approx 0 < \alpha$，因此拒绝 H_0.

即不同组别的居民对于政府议题有不同想法.

13.6. *Solution.* Fisher 精确检验用于分析列联表统计显著性的检验方法，也就是用于检验两个分类的关联.

实验中常用于小数据情况，但大样本也适用.

对于一个 2×2 的列联表：

	A	**B**	**Row Total**
C	a	b	$a+b$
D	c	d	$c+d$
Column Total	$a+c$	$b+d$	$n = a+b+c+d$

根据超几何分布可得这种情况出现的概率：

$$P_1 = \frac{\binom{a+b}{a}\binom{c+d}{c}}{\binom{n}{a+c}}$$

比如具体的：

	A	**B**	**Row Total**
C	1	9	10
D	11	3	14
Column Total	12	12	24

其概率为：

$$P_1 \approx 0.001346$$

在行列和不变的前提下，出现比其更极端的可能情况为：

	A	**B**	**Row Total**
C	0	10	10
D	12	2	14
Column Total	12	12	24

概率为：

$$P_2 \approx 0.000034$$

因此 P 值 ≈ 0.001380.

P 值越小，越倾向于拒绝原假设，比如此处就可以拒绝两者无关.

13.7. *Solution.* 似然比为：

$$\Lambda = \frac{\sup\limits_{\mu \in M_0} L(\mu)}{\sup\limits_{\mu \in M_0 \cup M_1} L(\mu)} = \frac{\prod\limits_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(X_i-\mu_0)^2}{2\sigma^2}} \right)}{\prod\limits_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(X_i-\overline{X})^2}{2\sigma^2}} \right)} = \exp\left\{ -\frac{1}{2} \left[\frac{\sqrt{n}(\overline{X}-\mu_0)}{\sigma} \right]^2 \right\}$$

其中 H_0 成立时，$Z = \dfrac{\sqrt{n}(\overline{X}-\mu_0)}{\sigma} \sim N(0,1)$，$\Lambda = \exp\left(-\frac{1}{2}Z^2\right)$.

因此 $P(\Lambda \le \lambda_0 | H_0) \le \alpha \Leftrightarrow Z > z_{\frac{\alpha}{2}} \vee Z < -z_{\frac{\alpha}{2}}$ 时拒绝 H_0.

故与 z 检验等价.

13.8. *Solution.* 计算卡方统计值得：

$$\chi_0^2 = \sum_{i=1}^{4} \frac{(O_i - E_i)^2}{E_i} = 0.47$$

其自由度为 $4-1=3$，可知其 P 值 ≈ 0.9254，故不拒绝原假设，即 Mendel 理论正确.

其似然比为：

$$\Lambda = \frac{\sup\limits_{p \in P_0} L(p)}{\sup\limits_{p \in P_0 \cup P_1} L(p)} = \frac{p_1^{O_1} p_2^{O_2} p_3^{O_3} p_4^{O_4}}{\left(\frac{O_1}{n}\right)^{O_1} \left(\frac{O_2}{n}\right)^{O_2} \left(\frac{O_3}{n}\right)^{O_3} \left(\frac{O_4}{n}\right)^{O_4}}$$

则：

$$-2\ln\Lambda = 2\sum_{i=1}^{4} O_i(\ln O_i - \ln p_i) - 2n\ln n$$

分布趋向于 $\chi^2(3)$. 计算可得：

$$-2\ln\Lambda \approx 0.4754$$

可知其 P 值 ≈ 0.9243，同样不拒绝原假设，即 Mendel 理论正确.