§8.3　大標本による母比率の検定

母比率つまり母平均が ρ である無限2項母集団から無作為抽出された標本の大きさ n が充分大きいとします．5.4節で述べたように母分散は $\rho(1-\rho)$ です． 標本比率を表す確率変数を R とおきます． 中心極限定理より導かれた定理5.4より次のことが成り立ちます：

標本比率を表す確率変数 R は近似的に正規分布 $\mathrm{N}\left(\rho, \dfrac{\rho(1-\rho)}{n}\right)$ に従う．

よって，R を標準化した確率変数

$$Z = \frac{R-\rho}{\sqrt{\dfrac{\rho(1-\rho)}{n}}} = \sqrt{\frac{n}{\rho(1-\rho)}}\,(R-\rho)$$

は標準正規分布 $\mathrm{N}(0,1)$ に従います． この確率変数 Z の累積分布関数を $\Phi(x)$ とおきます：

$$\Phi(x) = \mathrm{P}(\,Z \le x\,) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt \ .$$

各実数 a について，

$$\mathrm{P}(\,Z \le a\,) = \Phi(a)\ ,$$
$$\mathrm{P}(\,Z \ge a\,) = 1 - \mathrm{P}(\,Z \le a\,) = 1 - \Phi(a)\ ,$$
$$\Phi(-a) = 1 - \Phi(a)\ .$$

与えられた実数 p に対して，仮説 $\rho \ne p$ 或いは仮説 $\rho > p$ 或いは仮説 $\rho < p$ を検討します． 標本比率 R を標準化した確率変数 $Z = \sqrt{\dfrac{n}{\rho(1-\rho)}}\,(R-\rho)$ は標準正規分布 $\mathrm{N}(0,1)$ に従います． しかし，母比率 ρ は未知ですから，$Z = \sqrt{\dfrac{n}{\rho(1-\rho)}}\,(R-\rho)$ は実現値を計算できないので検定統計量にできません． そこで，ρ を与えられた実数 p で置き換えた確率変数

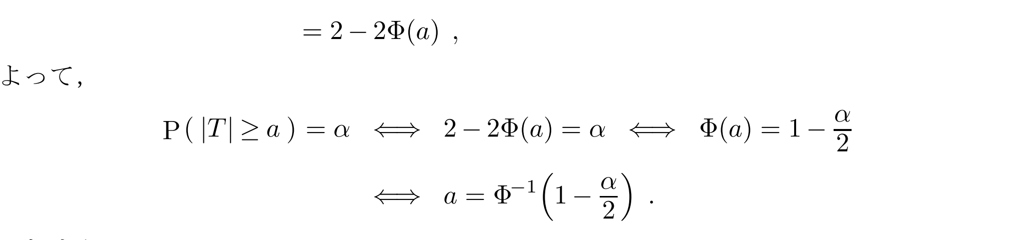
$$T = \sqrt{\frac{n}{p(1-p)}}\,(R-p)$$

を検定統計量とします． 0 に近い正の実数 α を有意水準とします．

仮説 $\rho \ne p$ を検討します． 帰無仮説 $\rho = p$ を立てます． 検定統計量 $T = \sqrt{\dfrac{n}{p(1-p)}}\,(R-p)$ について，

$$Z = \sqrt{\frac{n}{\rho(1-\rho)}}\,(R-\rho) = \sqrt{\frac{n}{p(1-p)}}\,(R-p) = T\ .$$

よって検定統計量 T は標準正規分布 $\mathrm{N}(0,1)$ に従います．



検定統計量 T の実現値が 0 から離れるほど帰無仮説 $\rho = p$ は疑わしくなります． そこで，T の実現値と 0 との間の間隔，つまり T の実現値の絶対値が，ある基準値以上のとき帰無仮説を棄却します． その基準値を考えます． 帰無仮説 $\rho = p$ より $T = Z$ なので，正の各実数 a について，

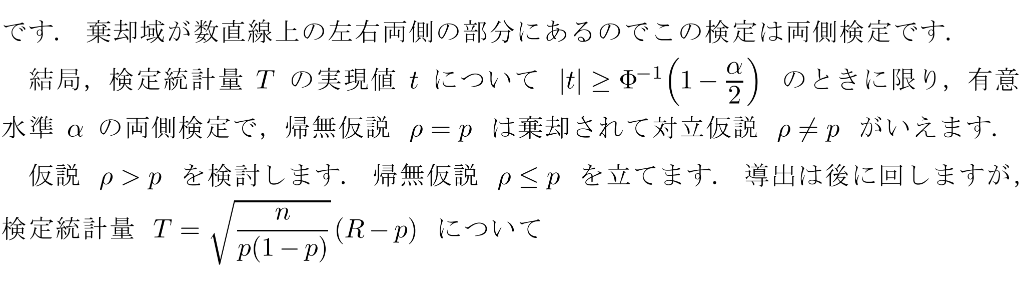
$$\begin{aligned}
\mathrm{P}(\,|T| \ge a\,) &= \mathrm{P}(\,|Z| \ge a\,) = \mathrm{P}(\,Z \le -a\,) + \mathrm{P}(\,Z \ge a\,) \\
&= \Phi(-a) + 1 - \mathrm{P}(\,Z \le a\,) = 1 - \Phi(a) + 1 - \Phi(a) \\
&= 2 - 2\Phi(a)\ ,
\end{aligned}$$

よって，

$$\begin{aligned}
\mathrm{P}(\,|T| \ge a\,) = \alpha \iff 2 - 2\Phi(a) = \alpha \iff \Phi(a) = 1 - \frac{\alpha}{2} \\
\iff a = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\ .
\end{aligned}$$

これより

$$\mathrm{P}\left(\ |T| \ge \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\ \right) = \alpha\ .$$



検定統計量 T の実現値 t について， $|t| \ge \Phi^{-1}\left(1 - \dfrac{\alpha}{2}\right)$ のときに限り，つまり $t \le -\Phi^{-1}\left(1 - \dfrac{\alpha}{2}\right)$ または $t \ge \Phi^{-1}\left(1 - \dfrac{\alpha}{2}\right)$ のときに限り，帰無仮説 $\rho = p$ は非常に疑わしいと考えてこれを棄却します． 棄却域は区間の合併

$$\left(-\infty, -\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\right] \cup \left[\Phi^{-1}\left(1-\frac{\alpha}{2}\right), \infty\right)$$

です． 棄却域が数直線上の左右両側の部分にあるのでこの検定は両側検定です．

結局，検定統計量 T の実現値 t について $|t| \ge \Phi^{-1}\left(1 - \dfrac{\alpha}{2}\right)$ のときに限り，有意水準 α の両側検定で，帰無仮説 $\rho = p$ は棄却されて対立仮説 $\rho \ne p$ がいえます．

仮説 $\rho > p$ を検討します． 帰無仮説 $\rho \le p$ を立てます． 導出は後に回しますが，検定統計量 $T = \sqrt{\dfrac{n}{p(1-p)}}\,(R-p)$ について

$$Z = \sqrt{\frac{n}{\rho(1-\rho)}}\,(R-\rho) \ge \sqrt{\frac{n}{p(1-p)}}\,(R-p) = T\ .$$

確率変数 Z は標準正規分布 $\mathrm{N}(0,1)$ に従うので， $\mathrm{E}[T] \le \mathrm{E}[Z] = 0$ ． T の実現値の分布の中心は 0 以下なので，T の実現値が 0 より大分大きいとき帰無仮説は疑わしいといえます． そこで，T の実現値がある正の基準値以上のとき帰無仮説を棄却します． その基準値を考えます． 帰無仮説 $\rho \le p$ より $Z \ge T$ なので，各実数 a について， $T \ge a$ ならば $Z \ge a$ ，よって

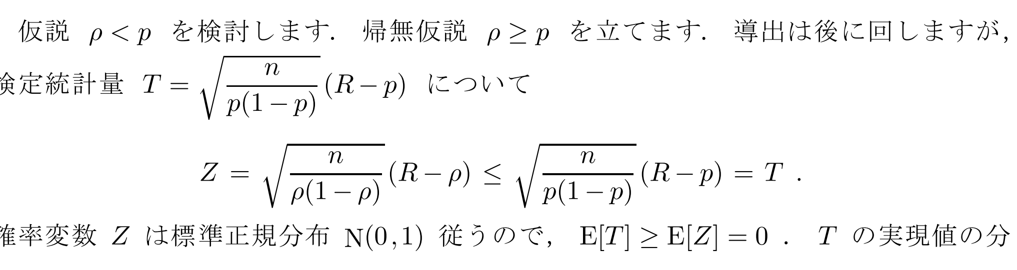
$$\mathrm{P}(\,T \ge a\,) \le \mathrm{P}(\,Z \ge a\,)\ ,$$

$\mathrm{P}(\,Z \ge a\,) = 1 - \mathrm{P}(\,Z \le a\,) = 1 - \Phi(a)$ なので，

$$\begin{aligned}
\mathrm{P}(\,Z \ge a\,) = \alpha \iff 1 - \Phi(a) = \alpha \iff \Phi(a) = 1 - \alpha \\
\iff a = \Phi^{-1}(1-\alpha)\ .
\end{aligned}$$

これより

$$\mathrm{P}\left(\,T \ge \Phi^{-1}(1-\alpha)\,\right) \le \alpha\ .$$



検定統計量 T の実現値 t について $t \ge \Phi^{-1}(1-\alpha)$ のときに限り，帰無仮説 $\rho \le p$ は非常に疑わしいと考えてこれを棄却します． 棄却域は区間 $[\Phi^{-1}(1-\alpha), \infty)$ です． 棄却域が数直線上の右側の部分にあるのでこの検定は右側検定です．

結局，検定統計量 T の実現値 t について $t \le -\Phi^{-1}(1-\alpha)$ のときに限り，有意水準 α の左側検定で，帰無仮説 $\rho \le p$ は棄却されて対立仮説 $\rho > p$ がいえます．

仮説 $\rho < p$ を検討します． 帰無仮説 $\rho \ge p$ を立てます． 導出は後に回しますが，検定統計量 $T = \sqrt{\dfrac{n}{p(1-p)}}\,(R-p)$ について

$$Z = \sqrt{\frac{n}{\rho(1-\rho)}}\,(R-\rho) \le \sqrt{\frac{n}{p(1-p)}}\,(R-p) = T\ .$$

確率変数 Z は標準正規分布 $\mathrm{N}(0,1)$ 従うので， $\mathrm{E}[T] \ge \mathrm{E}[Z] = 0$ ． T の実現値の分布の中心は 0 以上なので，T の実現値が 0 より大分小さいとき帰無仮説は疑わしいといえます． そこで，T の実現値がある負の基準値以下のとき帰無仮説を棄却します． その基準値を考えます． 帰無仮説 $\rho \ge p$ より $Z \le T$ なので，各実数 a について， $T \le -a$ ならば $Z \le -a$ ，よって

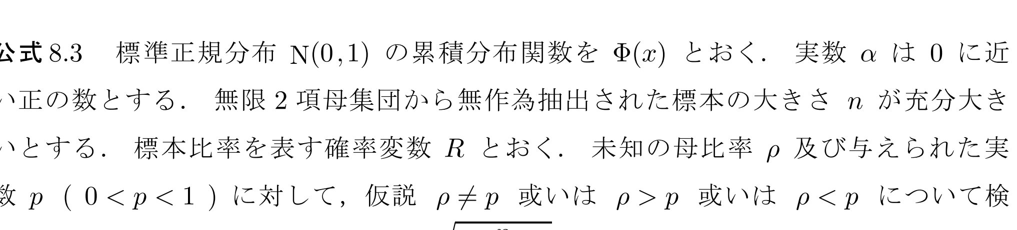
$$\mathrm{P}(\,T \le -a\,) \le \mathrm{P}(\,Z \le -a\,)\ ,$$

$\mathrm{P}(\,Z \le -a\,) = \Phi(-a) = 1 - \Phi(a)$ なので，

$$\begin{aligned}
\mathrm{P}(\,Z \le -a\,) = \alpha \iff 1 - \Phi(a) = \alpha \iff \Phi(a) = 1 - \alpha \\
\iff a = \Phi^{-1}(1-\alpha)\ .
\end{aligned}$$

これより

$$\mathrm{P}\left(\,T \le -\Phi^{-1}(1-\alpha)\,\right) \le \alpha\ .$$



検定統計量 T の実現値 t について $t \le -\Phi^{-1}(1-\alpha)$ のときに限り，帰無仮説 $\rho \ge p$ は非常に疑わしいと考えてこれを棄却します． 棄却域は区間 $(-\infty, \Phi^{-1}(1-\alpha)]$ です． 棄却域が数直線上の左側の部分にあるのでこの検定は左側検定です．

結局，検定統計量 T の実現値 t について $t \le -\Phi^{-1}(1-\alpha)$ のときに限り，有意水準 α の左側検定で，帰無仮説 $\rho \ge p$ は棄却されて対立仮説 $\rho < p$ がいえます．

右側検定と左側検定とを併せて片側検定といいます．

公式 8.3　標準正規分布 $\mathrm{N}(0,1)$ の累積分布関数を $\Phi(x)$ とおく． 実数 α は 0 に近い正の数とする． 無限2項母集団から無作為抽出された標本の大きさ n が充分大きいとする． 標本比率を表す確率変数 R とおく． 未知の母比率 ρ 及び与えられた実数 p ($0 < p < 1$) に対して，仮説 $\rho \ne p$ 或いは $\rho > p$ 或いは $\rho < p$ について検討するとき，検定統計量 T を $T = \sqrt{\dfrac{n}{p(1-p)}}\,(R-p)$ と定めると次のようになる．

(1) 検定統計量 T の実現値 t について $|t| \ge \Phi^{-1}\left(1 - \dfrac{\alpha}{2}\right)$ のとき，有意水準 α の両側検定で，帰無仮説 $\rho = p$ は棄却されて対立仮説 $\rho \ne p$ がいえる．

(2) 検定統計量 T の実現値 t について $t \ge \Phi^{-1}(1-\alpha)$ のとき，有意水準 α の右側検定で，帰無仮説 $\rho \le p$ は棄却されて対立仮説 $\rho > p$ がいえる．

(3) 検定統計量 T の実現値 t について $t \le -\Phi^{-1}(1-\alpha)$ のとき，有意水準 α の左側検定で，帰無仮説 $\rho \ge p$ は棄却されて対立仮説 $\rho < p$ がいえる．

$\boxed{\text{例題 8.3}}$　大きさの約 53 万の 2 項母集団から無作為抽出された大きさ 200 の標本の標本比率が 0.25 であったとする． 母比率は 0.3 より小さいといえるかどうか，有意水準 5% で検定する．

大きさ約 53 万の 2 項母集団から大きさ 200 の標本を抽出するので，無限母集団とみなす． 母比率 ρ の無限2項母集団から無作為抽出した大きさ 200 の標本の標本比率を表す統計量 R は近似的に正規分布 $\mathrm{N}\left(\rho, \dfrac{\rho(1-\rho)}{200}\right)$ に従うので，R を標準化した確率変数

$$Z = \frac{R-\rho}{\sqrt{\dfrac{\rho(1-\rho)}{200}}} = \sqrt{\frac{200}{\rho(1-\rho)}}\,(R-\rho)$$

は近似的に標準正規分布 $\mathrm{N}(0,1)$ に従う．

母比率 ρ について $\rho < 0.3$ ではないかと考えているので，帰無仮説 $\rho \ge 0.3$ を立てる． 検定統計量 T を次のように定める：

$$T = \sqrt{\frac{200}{0.3(1-0.3)}}\,(R-0.3) = \sqrt{\frac{20000}{21}}\,(R-0.3)\ .$$

帰無仮説 $\rho \ge 0.3$ より

$$Z = \sqrt{\frac{200}{\rho(1-\rho)}}\,(R-\rho) \le \sqrt{\frac{200}{0.3(1-0.3)}}\,(R-0.3) = T\ .$$

確率変数 Z は標準正規分布 $\mathrm{N}(0,1)$ に従うので $\mathrm{E}[T] \ge \mathrm{E}[Z] = 0$ ． 検定統計量 T の実現値が小さいほど帰無仮説は疑わしいので左側検定を行う． 確率変数 Z は近似的に標準正規分布 $\mathrm{N}(0,1)$ に従うので

$$\mathrm{P}(\,Z \le -1.645\,) = 0.05\ .$$

帰無仮説より $Z \le T$ なので， $T \le -1.645$ ならば $Z \le -1.645$ ，よって，

$$\mathrm{P}(\,T \le -1.645\,) \le \mathrm{P}(\,Z \le -1.645\,) = 0.05\ .$$

検定統計量 T の実現値 t について $t \le -1.645$ のときに限り，有意水準 5% の左側検定で，帰無仮説 $\rho \ge 0.3$ は棄却されて対立仮説 $\rho < 0.3$ がいえる．

いま，確率変数 $\overline{R}$ の実現値が 0.25 なので，検定統計量 $T = \sqrt{\dfrac{20000}{21}}\,(R-0.3)$ の実現値 t は

$$t = \sqrt{\frac{20000}{21}}\,(0.25-0.3) \fallingdotseq -1.543\ .$$

$t \not\le -1.645$ なので，帰無仮説 $\rho \ge 0.3$ は棄却できない． つまり，有意水準 5% の左側検定では，母比率は 0.3 より小さいといえない．　$\boxed{\text{終}}$

$\boxed{\text{問題 8.3.1}}$　無限2項母集団から無作為抽出された大きさが 500 である標本の比率が 0.83 であるとします． 母比率は 0.8 より大きいといえるかどうか，有意水準 5% で仮説検定しなさい．

$\boxed{\text{問題 8.3.2}}$　無限2項母集団から無作為抽出された大きさが 3000 である標本の比率が 0.598 であるとします． 母比率は 0.6 より小さいといえるかどうか，有意水準 1% で仮説検定しなさい．

——— 標本比率に関する不等式の証明

定理　実数 ρ と p とについて $0 < \rho < 1$ かつ $0 < p < 1$ とする．

$\rho \le p$ ならば，区間 $[0,1]$ の各実数 R について $\dfrac{R-\rho}{\sqrt{\rho(1-\rho)}} \ge \dfrac{R-p}{\sqrt{p(1-p)}}$ ．

$\rho \ge p$ ならば，区間 $[0,1]$ の各実数 R について $\dfrac{R-\rho}{\sqrt{\rho(1-\rho)}} \le \dfrac{R-p}{\sqrt{p(1-p)}}$ ．

証明　変数 R の関数 $f(R)$ を次のように定める：

$$f(R) = \frac{R-\rho}{\sqrt{\rho(1-\rho)}} - \frac{R-p}{\sqrt{p(1-p)}}\ .$$

このとき，

$$\begin{aligned}
f(0)\,f(1) &= \left(\frac{-\rho}{\sqrt{\rho(1-\rho)}} - \frac{-p}{\sqrt{p(1-p)}}\right)\left(\frac{1-\rho}{\sqrt{\rho(1-\rho)}} - \frac{1-p}{\sqrt{p(1-p)}}\right) \\
&= \frac{\rho + p - 2\rho p - 2\sqrt{\rho p(1-\rho)(1-p)}}{\sqrt{\rho p(1-\rho)(1-p)}}\ .
\end{aligned}$$

ここで

$$\begin{aligned}
&(\rho + p - 2p\rho)^2 - \left\{2\sqrt{\rho p(1-\rho)(1-p)}\right\}^2 \\
&= \rho^2 + p^2 + 4\rho^2 p^2 + 2\rho p - 4\rho^2 p - 4\rho p^2 - 4\rho p + 4\rho^2 p + 4\rho p^2 - 4\rho^2 p^2 = \rho^2 - 2\rho p + p^2 \\
&= (\rho - p)^2 \ge 0\ ,
\end{aligned}$$

よって

$$(\rho + p - 2\rho p)^2 \ge \left\{2\sqrt{\rho p(1-\rho)(1-p)}\right\}^2\ ,$$

$0 < \rho < 1$ かつ $0 < p < 1$ なので， $-1 < 2\rho - 1 < 1$ かつ $-1 < 2p - 1 < 1$ ， $(2\rho-1)(2p-1) < 1$ ，よって $\rho + p - 2\rho p = \dfrac{1 - (2\rho-1)(2p-1)}{2} > 0$ ． これより，

$$\rho + p - 2\rho p \ge 2\sqrt{\rho p(1-\rho)(1-p)}\ ,$$
$$\rho + p - 2\rho p - 2\sqrt{\rho p(1-\rho)(1-p)} \ge 0\ ,$$

よって $f(0)\,f(1) \ge 0$ ． $f(R)$ は R の1次関数なので，区間 $[0,1]$ の各実数 R について $f(R) \ge 0$ か，または，区間 $[0,1]$ の各実数 R について $f(R) \le 0$ ． $\rho \le p$ ならば， $f(p) = \dfrac{p-\rho}{\sqrt{\rho(1-\rho)}} \ge 0$ なので，区間 $[0,1]$ の各実数 R について， $f(R) \ge 0$ ， $\dfrac{R-\rho}{\sqrt{\rho(1-\rho)}} - \dfrac{R-p}{\sqrt{p(1-p)}} \ge 0$ ， $\dfrac{R-\rho}{\sqrt{\rho(1-\rho)}} \ge \dfrac{R-p}{\sqrt{p(1-p)}}$ ． $\rho \ge p$ ならば， $f(p) = \dfrac{p-\rho}{\sqrt{\rho(1-\rho)}} \le 0$ なので，区間 $[0,1]$ の各実数 R について， $f(R) \le 0$ ， $\dfrac{R-\rho}{\sqrt{\rho(1-\rho)}} - \dfrac{R-p}{\sqrt{p(1-p)}} \le 0$ ， $\dfrac{R-\rho}{\sqrt{\rho(1-\rho)}} \le \dfrac{R-p}{\sqrt{p(1-p)}}$ ．　（証明終り）