

ゼミ 11月19日分

平山 奈々海

November 19, 2024

教科書の説明：黒色

補足：青色

わからなかったところ、気になったところ：赤色

1 第9章 仮説検定と信頼区間

1.1 9.1 仮説検定とは

仮説検定は2つの仮説を立ててどちらが正しいかをデータから判断する統計手法である。例えば、コインが歪みのないものか、多少の歪みがあるのかを検証したいとする。コインの表の出る確率を θ とすると、歪みがないと言う仮説は $\theta = 0.5$ と表すことができる。一方歪みがあると言う仮説は $\theta \neq 0.5$ と書くことができる。いま単純な場合として、 $\theta = 0.5$ か $\theta = 0.7$ かをデータから判断することを考えてみよう。歪みがないと言う仮説を $H_0 : \theta = 0.5$ と書いて**帰無仮説**と呼ぶ。表の出る確率が0.7の歪みを持つという仮説を $H_1 : \theta = 0.7$, もしくは $H_A : \theta = 0.7$ と書いて**対立仮説**と呼ぶ。どちらを帰無仮説に取るかについては後で説明するが、常識的な仮説や否定したい仮説を帰無仮説に取る。**コインに歪みがない($\theta = 0.5$)というのは常識的な仮説なので帰無仮説においている。**

10回コインを投げる実験を行ったところ6回表が出たとしよう。この実験結果からどちらの仮説が正しいと判断したらよいだろうか。表の出る回数を X で表すと、 X は2項分布 $Bin(10, \theta)$ に従う。 $\theta = 0.5$ と $\theta = 0.7$ のときの $X = k$ となる確率 $p(k|\theta)$ を表示すると下の表のようになる。2項分布は、 $\theta = 0.5$ のときは $k = 5$ を中心に分布し、 $\theta = 0.7$ のときには $k = 7$ を中心に分布している。**2項分布の期待値は np なので、 $Bin(10, 0.5)$ に従う時は $10 \times 0.5 = 5$ が期待値である。同様に $Bin(10, 0.7)$ に従う時は $10 \times 0.7 = 7$ が期待値である。**確率 $p(k|0.5)$ は $\theta = 0.5$ のときには k という値が現れる尤もらしさを表しているので、 $k = 6$ という値が出る尤もらしさは0.205と言う意味になる。 $\theta = 0.7$ のときに $k = 6$ という値が出る尤もらしさは0.200であるから、 $p(6|0.5) > p(6|0.7)$ となり、 $k = 6$ と言う値は $\theta = 0.7$ より $\theta = 0.5$ の時の方が起こり易いことがわかる。**よって、10回コインを投げて6回表が出るという結果が出た時にはコインに歪みがないという帰無仮説の方が正しそうである。**

| k | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|-----------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| $p(k 0.5)$ | 0.001 | 0.010 | 0.044 | 0.117 | 0.205 | 0.246 | 0.205 | 0.117 | 0.044 | 0.010 | 0.001 |
| $p(k 0.7)$ | 0.000 | 0.000 | 0.001 | 0.009 | 0.037 | 0.103 | 0.200 | 0.267 | 0.233 | 0.121 | 0.028 |
| $\frac{p(k 0.5)}{p(k 0.7)}$ | 165 | 70.8 | 30.4 | 13.0 | 5.58 | 2.39 | 1.02 | 0.439 | 0.188 | 0.081 | 0.035 |

Table 1: 条件付き確率の比較表

計算の補足

$$p(6|0.5) = (106)(0.5)^6 0.5^{10-6}$$

以上の内容から、どちらの仮説が尤もらしいかを調べるには確率関数の比 $\frac{p(k|0.5)}{p(k|0.7)}$ を用いるのがよいことがわかる。この比を**尤度比**と呼ぶ。 $k \leq 6$ のときには $\frac{p(k|0.5)}{p(k|0.7)} > 1$ となり $\theta = 0.5$ の仮説が受け入れられ、 $k \geq 7$ のときには $\frac{p(k|0.5)}{p(k|0.7)} < 1$ となり $\theta = 0.7$ の仮説が受け入れられると言うルールを考えるのが自然なように思える。2つの仮説を対称に考えるのであれば、このルールが妥当である。しかし、統計的仮説検定においては帰無仮説と対立仮説を非対称に扱う。その理由は、誤った仮説を選択してしまうリスクが帰無仮説と対立仮説では非対称に扱われる点にある。**(リスクが非対称というのは、確率が異なるという意味ですか?)** 上述の例で、 $X \leq 6$ なら仮説 $\theta = 0.5$ を選択し、 $X \geq 7$ なら仮説 $\theta = 0.7$ を選択するルールをとる場合、 $\theta = 0.5$ が正しいのに誤って $\theta = 0.7$ を選択してしまう確率は

$$P_{\theta=0.5}(X \geq 7) = \sum_{k=7}^{10} p(k|0.5) = 0.172$$

となる。これを**第1種の誤り**の確率と呼ぶ。一方、 $\theta = 0.7$ が正しいのに誤って $\theta = 0.5$ を選択してしまう確率は

$$P_{\theta=0.7}(X \leq 6) = \sum_{k=0}^6 p(k|0.7) = 0.350$$

となる。これを**第2種の誤り**の確率と呼ぶ。

統計的仮説検定では第1種の誤りの確率を小さく抑える検定方式が望まれる。これは、帰無仮説を否定することは「重大な意味を持つ」と考え、誤って帰無仮説を否定してしまうリスクを小さくしたいことを意味する。上述の例では、コインに歪みがないと言う仮説は当然成り立つと思われる常識的な仮説であり、この仮説が否定されることは少なからず驚きを与えることになる。「このコインには歪みがあります」と言う衝撃的な主張に対しては、その主張が誤る確率を0.01, 0.05 など小さい値に抑え、高い信頼性を持ってその主張が正しいことを保証してあげることが統計的仮説検定の考え方である。例えば、第1種の誤りの確率を0.055以下に抑えるように検定のルールを作ろうとすると、上の表から $X \geq 8$ のとき帰無仮説 $H_0 : \theta = 0.5$ を否定するように取れば良い。実際、 $P_{\theta=0.5}(X \geq 8) = 0.055$ を満たしている。

1.2 統計的仮説検定の考え方

統計的仮説検定のポイントと用語をまとめておこう。わかりやすく説明するために、次のような正規母集団のモデルを想定してみる。ある地域の小学校1年生男子児童の身長が平均 μ , 分散 σ_0^2 の正規分布に従っているとし、その母集団からサイズ n のランダム標本 X_1, \dots, X_n がとられるとする。

$$X_1, \dots, X_n, i.i.d. \sim N(\mu, \sigma_0^2)$$

その実現値を x_1, \dots, x_n とし、 $\mathbf{X} = (X_1, \dots, X_n), \mathbf{x} = (x_1, \dots, x_n)$ とおく。 μ は未知のパラメーター、 σ_0^2 の値はわかっているものとする。小学校1年生男子児童の身長の全国平均の値は μ_0 として知られている時、この地域の男子児童の身長の平均 μ が全国平均 μ_0 に等しいか否かを判断したい。

1.2.1 帰無仮説と対立仮説

一般に、仮説検定では**帰無仮説**と**対立仮説**という2つの排反な仮説を設け、それぞれ H_0, H_1 という記号で表す。排反とは、2つの仮説の両方に含まれるパラメーターは存在しないという意味である。通常は否定したい仮説や自然で常識的な仮説、経験上もしくは理論上成り立つ仮説などを帰無仮説に取る。上の例ではこの地域の児童の身長の平均 μ が全国平均 μ_0 に等しいのが自然であるので、次のような仮説を帰無仮説に取る。

$$H_0 : \mu = \mu_0$$

対立仮説については、状況に応じて幾つかの設定の仕方がある。例えば、特定のわかっている値 $\mu_1 (\mu_1 \neq \mu_0)$ と比較したい時には

$$H_1 : \mu = \mu_1$$

となる。仮説は μ_1 という一点からなるので**単純仮説**と呼ばれる。仮説が複数の点から構成されている場合は**複合仮説**と呼ぶ。例えば次のような仮説の取り方が考えられる。

$$H_1 : \mu \neq \mu_0, \quad H_1 : \mu < \mu_0, \quad H_1 : \mu > \mu_0$$

最初の対立仮説は全国平均と同等でないこと、2番目は全国平均より劣っていること、3番目は優れていることを意味する。最初の検定を**両側検定**、2番目、3番目の検定を**片側検定**と呼ぶ。

帰無仮説 H_0 が否定される時、 **H_0 を棄却する**もしくは**有意である**という、帰無仮説 H_0 が棄却できない時には **H_0 を受容する**。

1.2.2 検定統計量と棄却域

仮説検定は、標本に基づいて H_0 を棄却するか否かを判断する。そこで適当な統計量 $W = W(\mathbf{X})$ と定数 C を用いて

「 $W(\mathbf{X}) > C$ のとき H_0 を棄却する」「 $W(\mathbf{X}) \leq C$ のとき H_0 を受容する」

のような形で検定を行うことができる時、 W を**検定統計量**と呼ぶ。この場合、 C の値は H_0 の棄却と受容の境界を与えており、**棄却限界**と呼ばれる。検定統計量として H_0 と H_1 を最も良く識別できるような関数 $W(\mathbf{X})$ を見つけることが大事である。

検定統計量を用いると、標本空間は、仮説 H_0 を棄却する領域と受容する領域に分割される。 $R = \{x | W(x) > C\}$ を H_0 の**棄却域**、 $A = \{x | W(x) \leq C\}$ を H_0 の**受容域**と呼ぶ。

(9.1) で与えられる正規母集団のモデルにおいて

$$H_0 : \mu = \mu_0 \quad vs. \quad H_1 : \mu = \mu_1, \quad (\mu_1 > \mu_0)$$

なる検定問題を考えてみる。標本平均 \bar{X} を用いて $W = \sqrt{n}(\bar{X} - \mu_0)/\sigma_0$ なる形の検定統計量を考えると、定数 C を適当にとつて、 $W = \sqrt{n}(\bar{X} - \mu_0)/\sigma_0 > C$ ならば H_0 を棄却し $W = \sqrt{n}(\bar{X} - \mu_0)/\sigma_0 \leq C$ ならば H_0 を受容するという方法が自然である。この検定統計量による H_0 の棄却域と受容域は $R = \{x | \sqrt{n}(\bar{X} - \mu_0)/\sigma_0 > C\}$, $A = \{x | \sqrt{n}(\bar{X} - \mu_0)/\sigma_0 \leq C\}$ となる。