# MAT120：数量的推論　第15講ハンドアウト
仮説検定

---

前回の講義では，正規分布，標準正規分布，そして$z$-score を使ってそれらのあいだを移る方法を学びました。今回の講義では，その正規分布の道具立てと中心極限定理を用いて，データに基づいて意思決定を行います。中心となる考え方は仮説検定 であり，これは，ある観測結果があまりにも起こりにくいので，もとの仮定を疑い始めるべきかどうかを形式的に判断する方法です。ここでは，コイン投げやお菓子の取り出しのような二項実験に対する仮説検定に焦点を当てます。

# 1　導入

いかにも怪しそうな人たちの集団があなたにコインを渡して，次のようなゲームを持ちかけてきたとします。

> コインを投げ続けてください。表が出るたびに，私たちはあなたに$2 を払います。
> ですが，裏が出るたびに，あなたは私たちに$1 払わなければなりません。

最初は気前のよい話に思えたので，あなたは遊ぶことにします。ところが30 回投げたあとで，表はたった3 回しか出ず，裏が27 回出たことに気づきます。結局のところ，あなたはその怪しい集団に$21 支払うことになり，そのまま立ち去ります。

家に帰ったあと，あなたはこのありそうもない出来事が起こる確率を計算してみることにします。コイン投げは二項実験なので，次の計算を行います。

$$P(3 \text{ heads in } 30 \text{ flips}) = C_{30,3}(0.5)^3(0.5)^{27} \approx 3.78 \times 10^{-6}.$$

ここで$n = 30$ であり，コインは公平だと仮定しているので$p = 0.5$ です。これはこのちょうどその結果 の確率であって，まだ完全な仮説検定におけるP-value ではありません。この確率は非常に小さいので，コインが公平ではないと疑う非常に強い理由を与えます。そして，表が出る確率は0.5 よりかなり小さいのではないかと考えたくなります。

さて次の日，散歩に出かけたときに別の集団に出会ったとしましょう。この集団は前日の人たちほど怪しくは見えなかったので，あなたは話をしてみることにします。ところが奇妙なことに，この二つ目の集団もまったく同じゲームを持ちかけてきます。彼らはこう言います。

> コインを投げ続けてください。表が出るたびに，私たちはあなたに$2 を払います。
> ですが，裏が出るたびに，あなたは私たちに$1 払わなければなりません。

本当はやめておくべきだと思いながらも，あなたはもう一度30 回コインを投げて遊ぶことにします。この場合，驚いたことに，30 回のうち表が12 回出たので，あなたは$6 勝つことになります。

再びコインが公平だと仮定して，この結果が起こる確率を計算してみることにします。

$$P(12 \text{ heads in } 30 \text{ flips}) = C_{30,12}(0.5)^{12}(0.5)^{18} = C_{30,12}(0.5)^{30} \approx 0.081.$$

したがって，この結果そのものの確率は約8% です。実際，これはそれほど不合理な数字ではありません。というのも，30 回中の期待される表の回数である15 回でさえ，そのちょうどの確率はせいぜい14% しかないからです。この情報からは，不公平なコインで遊んでいたと考える十分な理由はありません。

### 仮説検定

今見たのは，非常に重要な統計的手続きである仮説検定の非形式的な考え方です。どちらの場合にも，私たちはあるシステム（この場合はコイン投げを含む二項システム）に直面し，小さ

な観測結果に基づいて「使っているコインは公平である」という仮定に異議を唱えました。最初の例では，自分が置かれた状況があまりにも起こりにくいことから，コインについての推論を行うことができました。二つ目の例では，もとの仮定を退けるような証拠は見つかりませんでした。実際，多くの仮説検定はこの形に従います。まず期待があり，次に観測があり，それが疑いを生み，その後に確率計算を行って，もとの仮定を退けるのがどれほど妥当かを判断するのです。この概念的な流れは次のように要約できます。

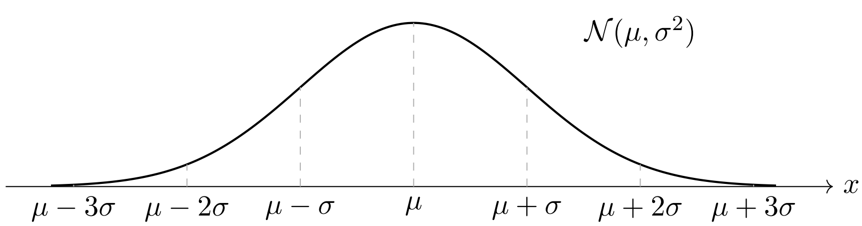

この講義では，二項実験を検定するために必要な道具を作り上げながら，この過程を詳しく見ていきます。実際には，単に二項分布だけでは少し数学が足りません。なぜなら，大きな系に対応する確率は計算が非常に難しいからです。実際，先ほどの二項係数$C_{30,12}$ には，$30!$ にまで及ぶ数の計算が含まれており，これは非常に大きいです。実用上は，この系を連続型確率変数へと持ち上げて，正規分布を自由に使えるようにしたほうが楽です。それに伴って確率計算の意味も変わります。個々の結果そのものが起こる確率を計算する代わりに，ある結果，あるいはそれよりさらに極端なもの が起こる確率を計算するのです。そこで，まずは正規分布のいくつかの特徴を復習します。

## 2　正規分布の簡単な復習

現実世界の多くのデータは，正規分布のもつ釣鐘型のパターンに従います。この分布がとくに便利なのは，たった二つの重要な情報によって決まるからです。

- **母平均** $\mu$，これは中心がどこにあるかを示す。
- **母標準偏差** $\sigma$，これはデータの広がり具合を示す。

都合のよいことに，この二つの数だけで正規分布は一意に特徴づけられます。そのため，正規分布は$\mathcal{N}(\mu, \sigma^2)$ という記法で表します。この記法は，「平均$\mu$，分散$\sigma^2$ をもつ正規分布」と読むとよいでしょう。標準偏差は平均と同じ物理単位をもつので，正規分布の$x$ 軸は，中心の$\mu$ から出発して，左右に$\sigma$ ずつ刻みながら記述することができます。



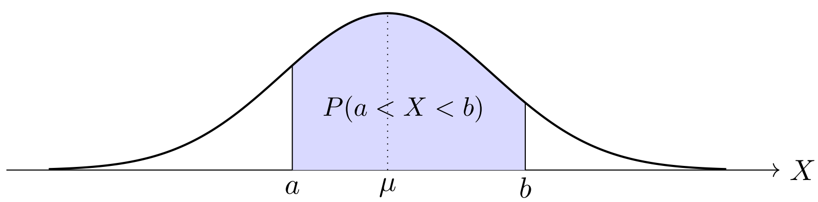
### 2.1　確率は曲線の下の面積として表される

確率分布の$x$ 軸は，確率変数が取りうる値を表します。二項分布の場合，この確率変数は離散的な個数ですが，正規分布の場合，この確率変数は連続的な尺度です。実際，連続型確率変数から作られるどんな確率分布においても，確率はある一点での曲線の高さではなく，曲線の下の面積 によって表されます。

連続型確率変数$X$ に対して，

$$P(a < X < b)$$

とは，$X$ がランダムに$a$ と$b$ のあいだのどこかの値を取る確率を意味します。この確率は，点$a$ と$b$ のあいだでの曲線の下の面積によって与えられます。正規分布では，これは次のように表せます。




### 2.2　標準正規分布

数学的には，こうした「曲線の下の面積」を正確に計算するのは難しいことがよくあります。そのため，$x$ 軸のある領域に対応する確率を厳密に記述するのは厄介です。実際には，私たちはもっとも単純な場合に注目することが多いです。これがいわゆる標準正規分布 です。この分布に対応する確率変数には特別に$Z$ という文字を使い，標準正規分布を次のように書きます。

$$Z \sim \mathcal{N}(0, 1).$$

標準正規分布は，平均が0 で，標準偏差が1 であることに注意してください。

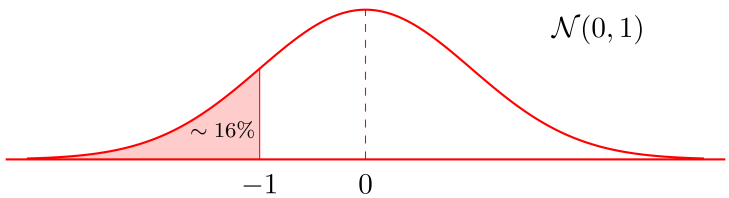
関係する確率を手で計算するのは難しいので，私たちはしばしば

$$P(Z < z),$$

という形のよく使われる値を表にまとめたものを使います。これは，確率変数がある固定された入力$z$ の左側 の値を取る確率を意味します。これはそのまま，$z$ の左側 にある面積にも対応します。この講義では，次の表を用います。

| $z$ | $-2.0$ | $-1.5$ | $-1.0$ | $-0.5$ | $0.0$ | $0.5$ | $1.0$ | $1.5$ | $2.0$ |
|---|---|---|---|---|---|---|---|---|---|
| $P(Z < z)$ | 0.02 | 0.07 | 0.16 | 0.31 | 0.50 | 0.69 | 0.84 | 0.93 | 0.98 |

$z$ の値を一つ選ぶと，その値の左側にある対応する面積が得られます。たとえば$z = -1$ を選ぶと，標準正規分布の全体の面積のうち約16% が$-1$ の左側にあることが分かります。

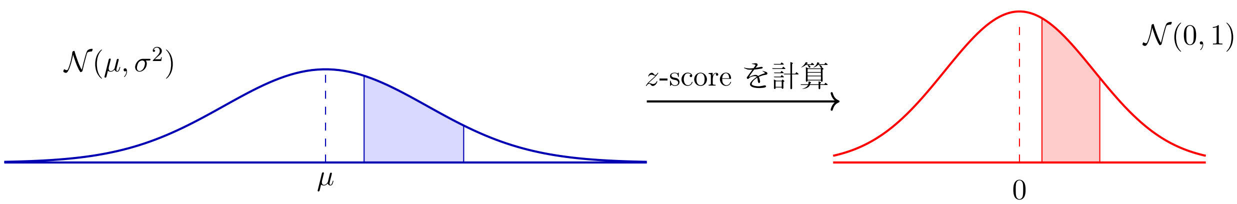



### 2.3 他の正規分布を標準正規分布へ変換する

任意の正規分布$X \sim \mathcal{N}(\mu, \sigma^2)$ に対して，その分布の確率は，まずそれを標準正規分布へ変換することで計算できます。ある値$x$ を標準正規の世界へ移すには，*z-score* を計算します。

$$z = \frac{x - \mu}{\sigma}.$$

この公式は単に，値$x$ が$\mu$ からどれだけ離れているかを，$\sigma$ を単位として測っていることを表しています。この$z$-score の公式には便利な性質があり，面積を保つ のです。つまり，$\mathcal{N}(\mu, \sigma^2)$ における確率計算を，$\mathcal{N}(0,1)$ における確率計算へ変換できるのです。




#### 2.3.1 例：身長

男性の身長が平均171 cm，標準偏差5.6 cm の正規分布にほぼ従うとしましょう。身長が180 cm より高い人に出会う確率はどれくらいでしょうか。

まず$z$-score を計算します。

$$z = \frac{180 - 171}{5.6} \approx 1.61.$$

したがって，

$$P(X > 180) = P(Z > 1.61).$$

あとは標準正規表や計算機を使って右側の裾の面積を見積もることができます。

ここで重要なのは厳密な数値そのものではありません。重要なのは，どんな 正規分布に関する問題でも，単一の普遍的な分布$\mathcal{N}(0,1)$ に関する問題へ翻訳できるという点です。

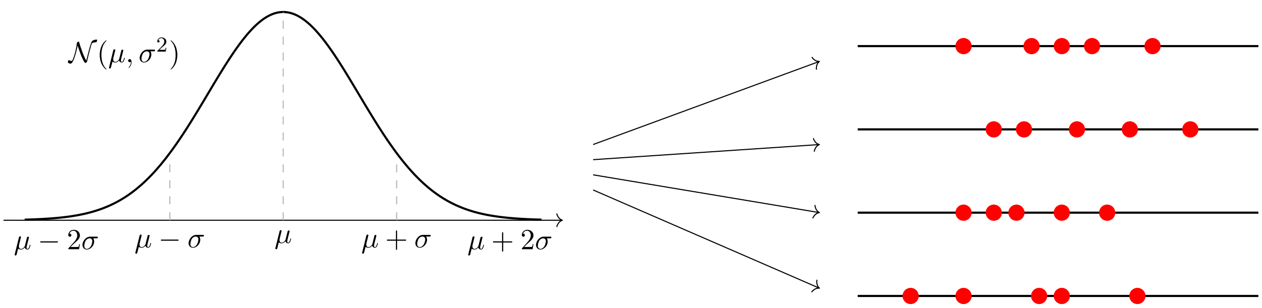
## 3 中心極限定理

### 3.1 標本平均の標本分布

ある正規分布，すなわち$\mathcal{N}(\mu, \sigma^2)$ があるとしましょう。これまでの議論によれば，その変数の値をランダムに一つ選ぶことは，$x$ 軸上のどこかの点をランダムに引くことに対応します。今見たように，その点がある領域の中に入る確率は，その領域の面積 によって与えられます。した

がって，面積の大きい領域ほど，ランダムに選ばれた点を含む確率が高いということになります。

では今度は，$x$ 軸上からランダムに5 個の点を選ぶとしましょう。第13講の用語では，これはサイズ5 の単純無作為標本 と呼ばれます。下の図は，サイズ5 の標本をいくつかランダムに選んだ様子を表しています。




これらの各標本について，標本平均$\bar{x}$ を計算することができます。とくに，サイズ5 のすべての単純無作為標本に対応するすべての 標本平均の集まりを考えることもできます。標本はランダムに取られているので，$\bar{x}$ そのものも一つの確率変数とみなすことができます。こうして得られるすべての$\bar{x}$ の集まりは，それ自体で一つの確率分布を形成します。このとき$x$ 軸は$\bar{x}$ で書かれます。この新しい確率分布は，$n = 5$ に対する標本平均の標本分布 と呼ばれます。

### 3.2 定理そのもの

任意の連続型確率分布に対して$\bar{x}$ の標本分布を作れるので，次のような自然な疑問が生じます。

> 一般に，こうした標本分布はいったいどのような形をしているのか。

この問いへの答えは，初等統計学の中でもっとも重要な事実の一つです。それが中心極限定理です。

**中心極限定理**

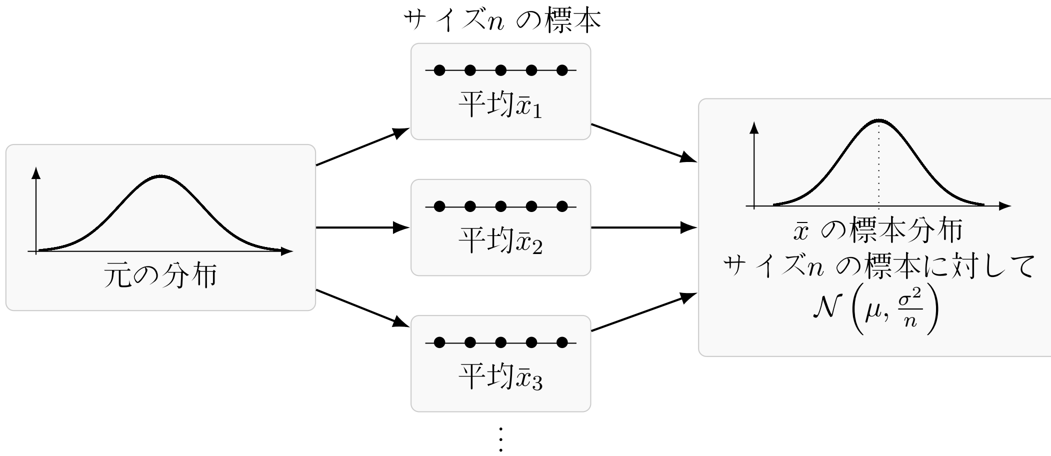
母集団が平均$\mu$，分散$\sigma^2$ をもつとき，標本サイズ$n$ が十分大きければ，標本平均$\bar{x}$ の標本分布は近似的に正規分布になり，

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

したがって，

- 標本分布の中心は依然として母平均$\mu$ であり，
- 広がりは小さく なる。なぜなら標準偏差は$\sigma$ ではなく$\sigma/\sqrt{n}$ になるからである。

中心極限定理を視覚的に表す一つの方法は次の通りです。




### 3.3 なぜこれが重要なのか

中心極限定理の主張では，もとの確率分布が正規分布であるとは仮定していないことに注意してください。実際，これこそが中心極限定理の偉大な力です。非常に広い範囲の分布に対して成り立つのです。言い換えると，中心極限定理は，元の母集団が正規分布でなくても，標本サイズが十分大きければ標本平均の標本分布はしばしば近似的に正規分布になると言っています。

これは，正規分布が現れるのは単に世界がいつも正規的だからではなく，ランダムに選ばれた平均値 が自然に近似的な正規分布の中へ並んでいくからだ，ということを意味します。

**推測統計**

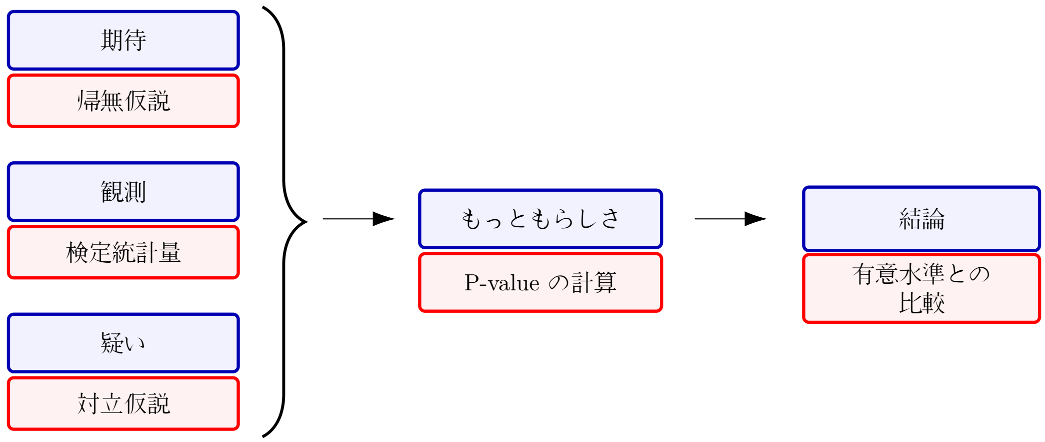
いくつかの基本的な仮定のもとでは，母平均$\mu$ が実際には何であるか分からなくても，中心極限定理は成り立ちます。この定理が適用できるとき，$\bar{x}$ の標本分布は真の 母平均$\mu$ を中心にもちます。したがって，標本データと母集団データとのあいだに橋をかけ，完全には分からないより大きな母集団について何かを言うことができるのです。

現実には，真の母平均$\mu$ や母標準偏差$\sigma$ を知っていることはほとんどありません。さらに，資源の制約があるため，すべての情報を記録したとしても，これらの母数を知ることができないことが多いです。その代わりに，私たちが持っているのは標本データと，そこから得られる標本統計量だけです。中心極限定理は，標本統計量が予測可能な正規的構造をもつことを教えてくれます。そして，これが標本データと母集団母数とのあいだに橋をかけるのです。ここから先は，この構造を使って母平均$\mu$ に関する主張を検定する方法を学びます。

## 4 仮説検定

### 4.1 統計的検定の一般的な構造

導入のところで，仮説検定を行うための一般的な流れに触れました。以下に，同じ流れを正しい専門用語を付けて書き直します。




これらを順に説明していきます。

- どんな仮説検定でも出発点になるのは，私たちが検定したい主張です。これを帰無仮説 と呼び，数式では$H_0$ と書きます。帰無仮説の形は，扱うデータによってさまざまです。この講義では入門レベルにとどめ，帰無仮説が母平均$\mu$ の値について主張している場合だけを考えます。
- 次に対立仮説 があります。これは，帰無仮説に対して検定したい主張です。実際には，これはしばしば帰無仮説の主張とは異なる，何らかの疑われる値になります。帰無仮説が「平均$\mu$ はある特定の値$c$ に等しい」と主張しているなら，対立仮説は「$\mu$ は何らかの意味で$c$ に等しくない」という主張になります。これは三通りあり，$\mu < c$，$\mu > c$，$\mu \neq c$ です。これら三種類の対立仮説によって，検定はそれぞれ左片側検定，右片側検定，両側検定 になります。
- どんな仮説検定にも，決定的に重要な標本データがあります。これにより標本平均$\bar{x}$ を計算し，その平均を標本分布$\mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$ の上に配置することができます。ここで$n$ は標本サイズです。この標本平均$\bar{x}$ は標本分布のどこかに存在するので，それに対応する$z$-score をもちます。この$z$-score が，確率を計算するために使う検定統計量 になることが多いです。
- 以上の三点が，仮説検定を行うための準備です。次に，私たちは標本データの*P-value* を求めなければなりません。これは，「帰無仮説が正しいと仮定したときに」，この標本平均，あるいはそれよりも極端な値が得られる確率です。数学的には，左片側検定であれば，$z$-score を使って確率$P(Z < z)$ を求めることに対応します。
- P-value は，「帰無仮説が正しいと仮定したときに」，標本データ，あるいはそれよりさらに極端なもの が得られる近似確率を与えます。この確率が非常に小さい なら，その観測はモデルのもとでは説明しにくいことになります。これを明確にするために，「許容できる確率」の切断点を設けます。これを有意水準 と呼びます。これを文字$\alpha$ で表し，典型的には$\alpha$ は5% か2% です。ただし，その正確な値は最終的には検定の入力として与えるものです。

仮説検定の最終判断は次のように行われます。

**判定規則**

- 計算された**P-value** $< \alpha$ なら，その観測は$H_0$ のもとではあまりにも起こりにくいとみなされるので，**$H_0$ を棄却する**。
- **P-value** $\geq \alpha$ なら，その観測はまだ十分には起こりにくくないので，**$H_0$ を棄却しない**。

重要なのは，「$H_0$ を棄却しない」というのは，$H_0$ が真であることを証明した という意味では

ないことです。それは単に，標本がそれに反する十分に強い証拠を与えていない，という意味にすぎません。

### 4.2 用語のまとめ

この節で出てきた重要な用語は次のように要約できます。

**重要な用語**

- **帰無仮説($H_0$)：** 最初に検定する主張。
- **対立仮説($H_1$)：** 私たちの疑いを表す競合する主張。
- **検定統計量：** 標本データから得られる数値情報。
- **P-value：** $H_0$ が真であると仮定したとき，観測と同じくらい，あるいはそれ以上に極端な結果が出る確率。
- **有意水準($\alpha$)：** 何を「起こりにくすぎる」とみなすかを決める切断確率。

このレベルでは，直感的な言い換えも役に立ちます。

- $H_0$：私たちが検定しているもの
- $H_1$：私たちが抱いている疑い
- 検定統計量：実際に観測したもの
- P-value：もし$H_0$ が真なら，その観測を見るためにどれだけ運が必要か
- 有意水準：私たちが許容する運の大きさ

### 4.3 例

導入では，非形式的な原型となる例を二つ見ました。30 回コインを投げて，その確率を計算したのでした。

**例1：** 怪しいコイン。この場合，私たちの統計的検定は次のようになります。

- **帰無仮説：** 私たちは使っているコインが公平だと期待していました。つまり，表が出る確率は裏が出る確率と同じだと期待していました。したがって，帰無仮説は$H_0 : p = 0.5$ となります。
- **検定統計量：** 私たちは30 回の投げで表を3 回観測しました。当時は明示しませんでしたが，これが検定統計量でした。標本データは成功確率が10% であることを示唆しています。
- **対立仮説：** 30 回中27 回が裏だったので，表が出る確率は実際には0.5 より小さいのではないかと疑いました。したがって，対立仮説は$H_1 : p < 0.5$ となります。
- **確率計算：** この場合，私たちはこのちょうどの結果の確率を直接計算しました。式$C_{30,3}(0.5)^{30} \approx 0$ を用いました。これはまだ正式な検定の完全なP-value ではありませんが，それでも，もしコインが公平ならこの観測が異常なほど起こりにくいことをすでに示しています。
- **結論：** 私たちは，この特定の結果は異常なほど起こりにくいと結論し，したがって公平なコインという仮定を棄却する非常に強い理由があると判断しました。結論として，表が出る確率がより小さいような不公平なコインを扱っている可能性を，データは強く示していると言えます。

**例2：** もっと普通のコイン

- **帰無仮説：** 再びコインが公平であるという主張を検定したので，帰無仮説は$H_0 : p = 0.5$ でした。
- **検定統計量：** 30 回の投げで表を12 回観測しました。これも当時は明示しませんでしたが，標本データは確率が$\frac{12}{30} = 0.4$ であることを示唆していました。
- **対立仮説：** それでも，コインが公平ではないかもしれないことを検定するために，表が出る確率は0.5 より小さいという対立仮説を採用しました。記号では$H_1 : p < 0.5$ です。
- **確率計算：** 再び，コインが公平であるという仮定のもとで，この特定の結果の確率を考えました。その計算は$C_{30,12}(0.5)^30 \approx 8\%$ でした。この観点から見ると，その結果は特に怪しいようには見えません。
- **結論：** 私たちは，このちょうどの結果の確率が8% であることは十分にありうると結論し，したがってコインが不公平だと考える強い理由はないと判断しました。

## 4.4 母平均を検定するという考え方

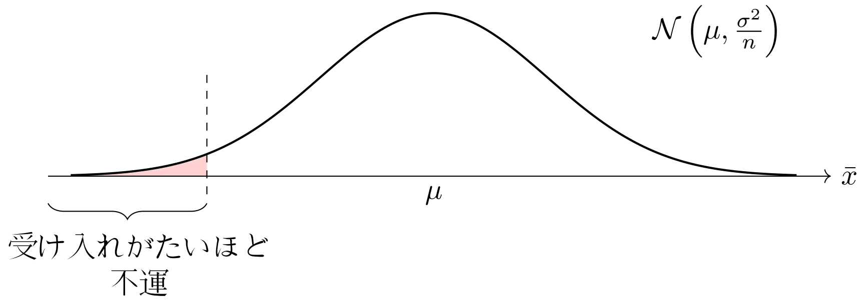
仮説検定の考え方は，帰無仮説が正しいと仮定して，その仮定の論理的帰結を探ることです。ある母集団の平均について明確な主張をする仮説を扱うとき，中心極限定理を用いて，ある標本がランダムに現れる確率を計算することができます。通常，どんな単純無作為標本でも，その標本平均は正規分布$\mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$ の中心の比較的近くにあるだろうと期待されます。ここで$\mu$ は，帰無仮説が主張している母平均の値です。実際，第14講での議論によれば，標本平均の$z$-score が$-1$ から$+1$ のあいだに入る確率は$\sim 68\%$ 程度と期待されます。すなわち，標本平均は分布$\mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$ において

$$\left[\mu - \frac{\sigma}{\sqrt{n}},\ \mu + \frac{\sigma}{\sqrt{n}}\right]$$

の区間に入ると期待されるのです。

たとえば，もし単純無作為標本を取って，得られた標本平均が主張された母平均$\mu$ から非常に遠い値だったとすると，それは現実には非常に起こりにくいことに対応します。もちろん，ランダム性の本質上，とくに運の悪い標本を引いて，母平均$\mu$ からかなり離れた標本平均が得られることは常にありえます。しかし，どれほど運が悪ければ，モデルに何か問題があると考え始めるべきなのでしょうか。たとえば，正規分布の中心から5 標準偏差も離れた標本を得ることは統計的には可能です。しかし，それが最初の標本 で起こったなら，それは単にモデルが間違っていると考えるほうがずっと自然です。

「ほどほどに不運」と「受け入れがたいほど不運」を区別するために，有意水準$\alpha$ を導入します。この値は切断点として働き，この点より先に落ちる観測値は，もはや合理的に受け入れるには不運すぎると言えるようになります。元のモデルが証明された形で間違っていると言う代わりに，私たちはそうした観測を元のモデルに対する強い証拠として扱います。だからこそ，標本統計量のP-value が有意水準より小さいときに，帰無仮説を棄却するのです。




これが，統計的検定に切断点が組み込まれている理由です。どの程度の「運」までなら，もとのモデルをまだ信頼するかを，私たちはあらかじめ決めておくのです。

### 4.5　第1種の誤りと第2種の誤り

結局のところ，仮説検定には常に不確かさが伴います。たとえば，仮説検定を行ってP-value が有意水準より小さいことが分かったとしても，それは帰無仮説が偽であることを証明した ことにはなりません。単に，標本が帰無仮説に反する証拠を与えたというだけです。さらに，そのような状況においても，対立仮説が証明された わけではありません。単に，データが帰無仮説のもとでは説明しにくいと分かっただけなのです。

仮説検定を行うときには常に不確実性があるので，誤って間違った結論を引いてしまう可能性は常にあります。主な誤りには二種類あり，第*1*種の誤り と第*2*種の誤り と呼ばれます。定義は次の通りです。

**第1種の誤りと第2種の誤り**

- **第1種の誤り** とは，実際には$H_0$ が真であるにもかかわらず，$H_0$ を棄却してしまうこと。
- **第2種の誤り** とは，実際には$H_0$ が偽であるにもかかわらず，$H_0$ を棄却しないこと。

より平易に言えば，

- **第1種の誤り** とは，帰無仮説に疑いをもち，それを棄却したが，実際には帰無仮説が真だったということ。
- **第2種の誤り** とは，標本が帰無仮説の問題を見抜けず，そのため実際には偽であるにもかかわらず棄却しなかったということ。

## 5　二項実験の検定

これまでのコイン投げの例では，確率は$C_{n,r}p^rq^{n-r}$ という二項確率公式を素朴に使って計算していました。しかし，$n = 30$ 回のコイン投げのような大きな系では，二項係数は

$$C_{30,r} = \frac{30!}{r!(30-r)!}.$$

となります。小さな$n$ に対しては，これまでパスカルの三角形を用いて二項係数を計算しており，それはここまでは非常に有用でした。しかし，$n$ が大きく なると，三角形を使うのは現実

的ではなくなります。実際，$n!$ は非常に速く大きくなるので，他の計算法も非現実的になっていきます。実用上は，この確率計算を，二項分布にだいたい合う正規分布で近似 するほうがずっと簡単です。そうすれば，$z$-score を使ってP-value を計算し，中心極限定理を使って判断を下すことができます。この近似によって，得られる確率の意味も変わります。今度は連続型確率変数を扱っているので，確率計算は二項分布の一本の棒の高さではなく，ある領域 の面積を与えます。実際には，この結果，あるいはそれよりさらに極端なもの が得られる尤もらしさを確率として考えることになります。これが，今後仮説検定を進める方法です。

### 5.1 二項実験から中心極限定理へ

コインを繰り返し投げるような二項実験を考えましょう。各試行には成功か失敗の二つの結果があり，$p$ は成功確率，$q = 1 - p$ は失敗確率です。統計学の観点から二項実験を見ると，次のように解釈できます。

- 各試行は一つ一つのデータ点であり，
- $n$ 回試行することはサイズ$n$ の標本を取ることに対応する。

実験のどんな固定された結果も，サイズ$n$ の標本とみなせます。

二項実験の結果をデータとして解釈することにすれば，その上で統計を行うことができます。そのための一つの方法は，データを数値化することです。すなわち，成功を1，失敗を0 と対応させます。すると標本平均はちょうど標本比率に一致します。なぜなら，

$$\bar{x} = \frac{\text{成功の回数}}{n} = \hat{p}.$$

だからです。したがって，二項実験では，標本平均について話すことと標本比率について話すことは同じことになります。

この0-1 化された二項実験では，平均は$p$ に等しく，分散は$pq$ に等しくなります。

### 5.2 標本比率の標本分布

二項実験では，

$$\text{母平均} = p, \qquad \text{母分散} = pq.$$

であることが分かります。こうした試行を$n$ 回平均すると，標本比率$\hat{p}$ が得られます。したがって，これは先ほどの標本平均の話を，二項実験の言葉で書き直したものにすぎません。

中心極限定理によれば，$n$ が十分大きければ，すべての標本比率$\hat{p}$ の集まりは近似的に正規分布になります。具体的には，次の公式が成り立ちます。

**二項仮説検定のための重要公式**

真の成功確率が$p$ であるならば，標本比率$\hat{p}$ の標本分布は近似的に

$$\mathcal{N}\left(p, \frac{p(1-p)}{n}\right)$$

である。

これは非常に重要な事実です。十分に大きな標本では，二項的な「運」の構造は近似的に正規的になる，ということを教えてくれます。

二項の設定では，未知の母数は真の成功確率$p$ です。

たとえば，コイン投げ実験では，

- **母集団母数** は，表が出る真の確率であり，
- **標本統計量** は，観測された表の比率です。

したがって，比較するのは

真の確率$p$　と　標本比率$\hat{p}$

です。

### 5.3　完全な検定：怪しいコイン

先ほどのいかさま話に戻りましょう。今度は，最初のコインに対して仮説検定を概略的に行います。

**Step 1:** 仮説を述べる。
もしこのコインが表に不利な 偏りをもっていると疑うなら，自然な仮説は

$$H_0 : p = 0.5, \qquad H_1 : p < 0.5.$$

です。

**Step 2:** 観測された標本比率を計算する。
私たちは30 回の投げで表を3 回観測したので，

$$\hat{p} = \frac{3}{30} = 0.1.$$

**Step 3:** $z$-score を計算する。
ここでは二項計算を使う代わりに，正規曲線で近似し，対応する$z$-score を検定統計量として使います。この場合，

$$p_0 = 0.5, \qquad q_0 = 0.5, \qquad n = 30.$$

なので，対応する$z$-score は

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n}} = \frac{0.1 - 0.5}{\sqrt{(0.5)(0.5)/30}} \approx -4.38.$$

**Step 4:** P-value を計算する。
$z$-score が$-4.38$ であるということは，観測された結果が標本分布の中心から4 標準偏差以上も離れているということです。したがって，この領域に対応する確率は非常に小さいと期待されます。実際，この面積は中心からあまりにも左に離れているので，標準的な$z$ 表には載っていません。$z$-score 計算機を使うと，P-value は

$$P(Z < -4.38) \approx 0.00001.$$

となります。

**Step 5:** 有意水準と比較する。
もし有意水準として$\alpha = 0.05$ を使うなら，

$$0.00001 < 0.05.$$

は明らかです。したがって，$H_0$ を棄却します。

**結論**

コインが公平であるという仮説に対して，きわめて強い反証がある。

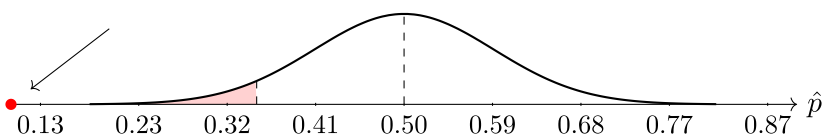
この検定では，帰無仮説は$\hat{p}$ の標本分布が0.5 を中心にもつと予測します。

$$\hat{p} \approx N\left(0.5, \frac{(0.5)^2}{30}\right).$$

したがって，この標本分布の標準偏差は

$$\sqrt{\frac{(0.5)^2}{30}} \approx 0.09$$

です。下のグラフがこの分布です。観測値$\hat{p} = 0.1$ は，帰無仮説が主張している中心から非常に遠い位置にあります。




公平なコインモデルでは，観測される比率のほとんどは0.5 の近くにあるはずです。$\hat{p} = 0.1$ のような観測値は，ありそうもない左裾のずっと先にあります。したがって，帰無仮説を棄却します。

## 5.4 正規近似を使う前の重要条件

ただし，正規近似はあらゆる 二項検定に自動的に適用できるわけではありません。標本が，帰無仮説のもとで十分大きくなければならないのです。

**正規近似の条件**

比率に対する$z$ 検定を使う前に，次を確認しなければならない。

$$np_0 > 5 \qquad \text{and} \qquad nq_0 > 5.$$

この両方が成り立てば，このレベルでは正規近似はたいてい妥当である。

怪しいコインの例では，

$$np_0 = 30(0.5) = 15 > 5, \qquad nq_0 = 30(0.5) = 15 > 5,$$

なので，近似は許されます。

# 6 例と演習

## 6.1 Haribo データ

授業では，19 袋のHaribo のお菓子について，各袋に含まれているお菓子の種類ごとの個数を数えました。

| 標本 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Rings | 3 | 7 | 4 | 2 | 6 | 6 | 7 | 3 | 7 | 4 | 2 | 6 | 6 | 4 | 7 | 5 | 5 | 2 | 6 |
| Hearts | 5 | 4 | 6 | 4 | 5 | 2 | 5 | 7 | 1 | 3 | 6 | 3 | 4 | 6 | 4 | 3 | 2 | 4 | 5 |
| Bottles | 4 | 1 | 4 | 2 | 4 | 2 | 3 | 3 | 4 | 6 | 4 | 0 | 2 | 4 | 4 | 4 | 8 | 1 | 5 |
| Eggs | 4 | 5 | 3 | 6 | 0 | 6 | 1 | 2 | 6 | 2 | 4 | 4 | 1 | 2 | 1 | 4 | 3 | 7 | 1 |
| Gummy bears | 7 | 8 | 6 | 12 | 11 | 10 | 10 | 9 | 7 | 11 | 6 | 14 | 15 | 8 | 10 | 9 | 6 | 12 | 7 |
| Total per bag | 23 | 25 | 23 | 26 | 26 | 26 | 26 | 24 | 25 | 26 | 22 | 27 | 28 | 24 | 26 | 25 | 24 | 26 | 24 |

数えたお菓子の総数は476 であり，その比率は次の通りです。

| 種類 | 総数 | 比率 |
|---|---|---|
| Rings | 92 | $92/476 \approx 19.3\%$ |
| Hearts | 79 | $79/476 \approx 16.6\%$ |
| Bottles | 65 | $65/476 \approx 13.7\%$ |
| Eggs | 62 | $62/476 \approx 13.0\%$ |
| Gummy bears | 178 | $178/476 \approx 37.4\%$ |

したがって，Gummy bears は他の種類よりもかなり頻繁に現れています。

### 6.1.1 Haribo の問いを仮説検定に変える

Gummy bears の比率が20% より大きいという仮説を検定するために，この系を二項実験として解釈します。そのためにまず次のように定義します。

- **成功** = 「gummy bear を引くこと」
- **失敗** = 「gummy bear ではないものを引くこと」

この入門レベルでは，476 個のお菓子を476 個の独立した観測として扱うので，$n = 476$ とします。また，袋ごとの差も無視します。

これから，「ランダムに選ばれたお菓子がgummy bear である真の確率は20% である」という主張を検定します。

**Step 1:** 仮説。

$p$ を，ランダムに選ばれたお菓子がgummy bear である真の確率とします。標本データに基づけば，gummy bear の真の比率は20% より大きいのではないかと疑われます。したがって，帰無仮

説と対立仮説は

$$H_0 : p = 0.20, \qquad \text{and} \quad H_1 : p > 0.20.$$

です。

**Step 2:** 観測された標本比率。

標本では，合計$n = 476$ 個のお菓子のうち178 個がgummy bear でした。したがって，gummy bear の標本比率は

$$\hat{p} = \frac{178}{476} \approx 0.374,$$

であり，上の表と一致します。

**Step 3:** 近似条件の確認。
この二項実験を正規分布で近似するためには，まず第5節で述べた二つの条件を確認しなければなりません。$H_0$ のもとでは，

$$p_0 = 0.20, \qquad \text{and} \quad q_0 = 0.80.$$

二つの条件を計算すると，

$$np_0 = 476(0.20) = 95.2 > 5, \qquad nq_0 = 476(0.80) = 380.8 > 5.$$

となります。両方とも5 より大きいので，仮説検定を進めてよいことが分かります。

**Step 4:** 検定統計量を計算する。

$p_0 = 0.2$，$q_0 = 0.8$，$\hat{p} = 0.374$，$n = 476$ を用いると，

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n}} = \frac{0.374 - 0.20}{\sqrt{(0.20)(0.80)/476}} \approx 9.49.$$

となります。したがって，帰無仮説によれば，標本で観測された$\hat{p}$ は想定される標本分布の中心から9 標準偏差以上も離れています。

**Step 5:** P-value を解釈する。

対立仮説はgummy bears の比率が0.2 より大きい と主張しています。したがって，私たちは標本分布の右裾を見ます。対応するP-value は

$$P(Z > 9.49),$$

であり，これは実際上は0 とみなせるほど小さいです。

**結論**

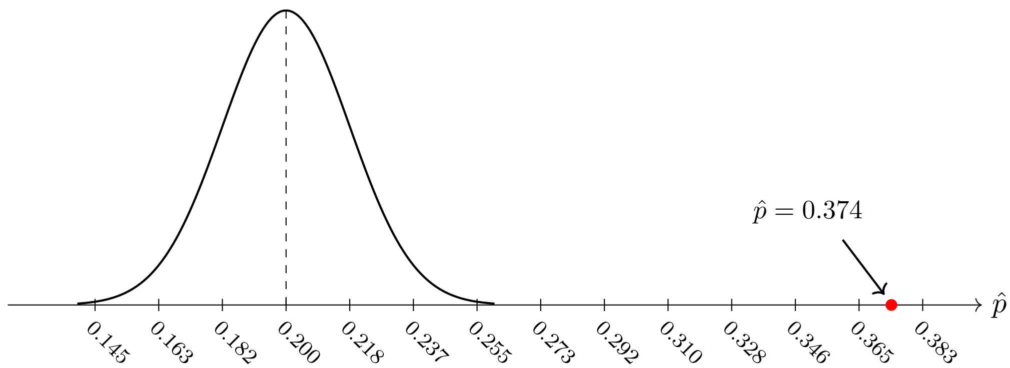
この標本は，gummy bears がたった20% しか現れないという主張に対して圧倒的な反証を与えている。少なくともこのデータセットでは，gummy bears は著しく過剰に含まれているように見える。

## 6.2 Haribo 検定の図

帰無仮説のもとでは，標本分布は近似的に

$$\hat{p} \approx N\left(0.20, \frac{(0.20)(0.80)}{476}\right).$$

です。その標準偏差はおよそ0.018 にすぎません。つまり，0.374 のような観測値は右裾のずっと先にあります。




## 6.3 例題

**Exercise**

あるニュースサイトは，訪問者の60% が少なくとも一本の記事をクリックすると主張している。サイズ$n = 96$ の単純無作為標本では，$x = 48$ 人しか記事をクリックしなかった。
$p$ を真のクリック率とする。

**(a)** 真のクリック率が60% より低いかどうかを検定するための仮説を述べよ。

**(b)** $H_0$ のもとで正規近似の条件を確認せよ。

**(c)** $\hat{p}$ と$z$-score を計算せよ。

**(d)** 下の標準正規表を用いてP-value を求めよ。

**(e)** 有意水準$\alpha = 0.05$ で判定せよ。

**(f)** 文脈に即して結論を解釈せよ。

| $z$ | $-2.0$ | $-1.5$ | $-1.0$ | $-0.5$ | 0.0 | 0.5 | 1.0 | 1.5 | 2.0 |
|---|---|---|---|---|---|---|---|---|---|
| $P(Z < z)$ | 0.02 | 0.07 | 0.16 | 0.31 | 0.50 | 0.69 | 0.84 | 0.93 | 0.98 |

**解答**

**(a)** 真のクリック率が60% より低いかどうかを検定したいので，

$$H_0 : p = 0.60, \qquad H_1 : p < 0.60.$$

**(b)** $H_0$ のもとでは，

$$p_0 = 0.60, \qquad q_0 = 0.40.$$

したがって，

$$np_0 = 96(0.60) = 57.6 > 5, \qquad nq_0 = 96(0.40) = 38.4 > 5.$$

よって正規近似は妥当である。

**(c)** 標本比率は

$$\hat{p} = \frac{48}{96} = 0.50.$$

したがって，

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n}} = \frac{0.50 - 0.60}{\sqrt{(0.60)(0.40)/96}} = -2.0.$$

**(d)** これは左片側検定なので，

$$\text{P-value} = P(Z < -2.0).$$

表より，

$$P(Z < -2.0) \approx 0.02.$$

**(e)**

$$0.02 < 0.05,$$

なので，$H_0$ を棄却する。

**(f)** 真のクリック率は60% より低いことを示す証拠がある。