第4章　統　　計

目　標　確率分布の中で最も重要な分布は，二項分布と正規分布である．ラプラスは試行の回数nが十分に大きいとき，二項分布は正規分布に近似できることを証明した．二項分布は有限確率空間における代表的な分布であるが二項係数や指数計算が多く計算がメンドウになることが多い．nが大きいときは正規分布に置きかえて計算してもよいわけだ．また，正規分布自体も理論的，実用的により大切なものであり，統計学では不可欠なものである．

母集団の中から標本を抜き出し，その標本を調べることによって母集団の特性を知ることは重要である．nが十分大きければ標本平均の分布は正規分布に近似できることも知られている．

この章では，次のテーマを扱う．

Ⅰ．二項分布・正規分布　　**Ⅱ．推定・検定**

Ⅰ．二項分布・正規分布

1°　**二項分布**　同じ条件のもとでくり返し行われるn回の独立試行において，事象Aの起こる確率をp，Aが起こる回数をXとすれば，Xは確率変数となり，その確率分布は

$$P(X=r)={}_n\mathrm{C}_r p^r(1-p)^{n-r}$$

となる．

この分布を二項分布といい，$\mathrm{B}(n,\ p)$で表す．

2°　**二項分布の平均・分散**　確率変数Xが二項分布$\mathrm{B}(n,\ p)$に従うとき，

平均：　$E(X)=np$,

分散：　$V(X)=npq\quad(q=1-p)$

標準偏差：　$\sigma=\sqrt{npq}$

である．

3°　**正規分布**　連続型確率分布Xの確率密度関数$f(x)$が

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-m)^2}{2\sigma^2}} \quad (-\infty<x<\infty)$$

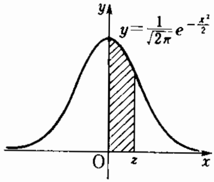
（m は任意の実数，σ は任意の正数）

である確率分布を**正規分布**（ガウス分布）といい，$\mathrm{N}(m,\ \sigma^2)$ で表す．

確率変数Xを標準化した $Z=\dfrac{X-m}{\sigma}$ の分布を**標準正規分布**といい，

$\mathrm{N}(0,\ 1)$ で表す．

（標準化については標問 **75** 研究を参照）



4° **正規分布表**　正規分布の確率計算を行うときは，右図の斜線部分の面積

$$P(0\leqq Z\leqq z)=\int_0^z f(x)\,dx$$

を計算してつくった正規分布表を使う．

5° **正規分布の平均・分散**　確率変数Xが正規分布 $\mathrm{N}(m,\ \sigma^2)$ に従うとき，

平均：　$E(X)=m$

分散：　$V(X)=\sigma^2$

6° **二項分布と正規分布の近似**　確率変数Xが二項分布 $\mathrm{B}(n,\ p)$ に従うとき，n を十分大きくとれば，

$$Z=\frac{X-np}{\sqrt{npq}} \quad (q=1-p)$$

は標準正規分布 $\mathrm{N}(0,\ 1)$ に近似的に従う（ラプラスの定理）．

標問 80.　〈二項分布 $\mathrm{B}(n,\ p)$〉

(1)　確率変数Xは二項分布に従い，$X=0$ となる確率は正である．また，$X=1$ となる確率は，$X=0$ となる確率の 6 倍であり，$X=2$ となる確率は，$X=1$ となる確率の 2 倍である．$X=2$ となる確率を求めよ．　（室蘭工大）

(2)　さいころ 1 個を n 回続けて投げるとき，その n 回のうち 4 以下の目が奇数回出る確率を $P(n)$ とする．このとき，次の各問いに答えよ．

（i）　$P(5)$ の値を分数で求めよ．

（ii）　n が限りなく大きくなるとき，$P(n)$ はどんな値に近づくか．　（宮崎大）

精講　(1)　確率変数Xが二項分布に従うとは，Xとその確率$P(X=r)$との関係が$P(X=r)={}_n\mathrm{C}_r p^r(1-p)^{n-r}$であることにほかならない．求める確率は$P(X=2)$であり，このときの未知数は$n$と$p$である．条件が2つあればこれらは決まる．

(2)　二項分布の確率${}_n\mathrm{C}_r p^r q^{n-r}$ $(q=1-p)$ は $(p+q)^n$ の展開式

$${}_n\mathrm{C}_0 q^n+{}_n\mathrm{C}_1 pq^{n-1}+\cdots\cdots+{}_n\mathrm{C}_r p^r q^{n-r}+\cdots\cdots+{}_n\mathrm{C}_n p^n$$

の項であり，二項定理と深く関連している．(☞ 標問 **22**)

解答　(1)　$X=k$ となる確率を p_k とおくと $p_k={}_n\mathrm{C}_k p^k(1-p)^{n-k}$

条件より

$$\frac{p_1}{p_0}=\frac{np(1-p)^{n-1}}{(1-p)^n}=\frac{np}{1-p}=6 \quad\cdots\cdots①$$

$$\frac{p_2}{p_1}=\frac{{}_n\mathrm{C}_2 p^2(1-p)^{n-2}}{np(1-p)^{n-1}}=\frac{(n-1)p}{2(1-p)}=2 \quad\cdots\cdots②$$

①，②より　$p=\dfrac{2}{3}$，$n=3$　　$\therefore$　$p_2={}_3\mathrm{C}_2\left(\dfrac{2}{3}\right)^2\cdot\dfrac{1}{3}=\dfrac{4}{9}$　　……(答)

(2)　(i)　1回の試行で4以下の目が出る確率は $\dfrac{2}{3}$ であるから

$$P(5)={}_5\mathrm{C}_1\left(\frac{2}{3}\right)\left(\frac{1}{3}\right)^4+{}_5\mathrm{C}_3\left(\frac{2}{3}\right)^3\left(\frac{1}{3}\right)^2+{}_5\mathrm{C}_5\left(\frac{2}{3}\right)^5=\frac{122}{243}\cdots\cdots(\text{答})$$

(ii)　$p=\dfrac{2}{3}$，$q=\dfrac{1}{3}$ とおくと

$$P(n)=\begin{cases}{}_n\mathrm{C}_1 pq^{n-1}+{}_n\mathrm{C}_3 p^3q^{n-3}+\cdots\cdots+{}_n\mathrm{C}_n p^n & (n:\text{奇数})\\ {}_n\mathrm{C}_1 pq^{n-1}+{}_n\mathrm{C}_3 p^3q^{n-3}+\cdots\cdots+{}_n\mathrm{C}_{n-1} p^{n-1}q & (n:\text{偶数})\end{cases}$$

二項定理より

$$\begin{array}{rl}(q+p)^n=&{}_n\mathrm{C}_0 q^n+{}_n\mathrm{C}_1 pq^{n-1}+{}_n\mathrm{C}_2 p^2q^{n-2}+\cdots\cdots+{}_n\mathrm{C}_n p^n\\ -)\ (q-p)^n=&{}_n\mathrm{C}_0 q^n-{}_n\mathrm{C}_1 pq^{n-1}+{}_n\mathrm{C}_2 p^2q^{n-2}+\cdots\cdots+(-1)^n{}_n\mathrm{C}_n p^n\\ \hline 1-\left(-\dfrac{1}{3}\right)^n=&2({}_n\mathrm{C}_1 pq^{n-1}+{}_n\mathrm{C}_3 p^3q^{n-3}+\cdots\cdots)\end{array}$$

(nの奇数・偶数により右辺の末尾が変わる．)

nが奇数，偶数いずれの場合も

$$1-\left(-\frac{1}{3}\right)^n=2P(n)\qquad\therefore\quad P(n)=\frac{1}{2}\left\{1-\left(-\frac{1}{3}\right)^n\right\}$$

よって，求める値は

$$\lim_{n\to\infty}P(n)=\frac{1}{2}\qquad\cdots\cdots(\text{答})$$

標問 81. 〈二項分布の期待値・分散〉

袋の中に3つの白球と6つの黒球が入っている．この袋から同時に5つの球を取り出したときの白球の数をXとする．また，この袋から1つ取り出してはもとにもどすことを5回くり返したときに白球が出てきた回数をYとする．

(1) 確率変数Xの確率分布，期待値，分散を求めよ．

(2) 確率変数Yの期待値，分散を求めよ．　(滋賀医大)

精講　(1)は球を1つずつ取り出すのと同じで非復元抽出，(2)は復元抽出となっている．復元抽出の試行のくり返し(独立試行)による確率変数は二項分布に従う．

解答　(1) 確率変数Xは0，1，2，3の値をとり

$$P(X=k)=\frac{{}_3C_k\cdot{}_6C_{5-k}}{{}_9C_5}$$

X	0	1	2	3
確率	$\frac{1}{21}$	$\frac{5}{14}$	$\frac{10}{21}$	$\frac{5}{42}$

であるから確率分布は右の表のようになる．　……(答)

期待値 $E(X)=1\cdot\frac{5}{14}+2\cdot\frac{10}{21}+3\cdot\frac{5}{42}=\frac{70}{42}=\frac{5}{3}$　……(答)

また，$E(X^2)=1^2\cdot\frac{5}{14}+2^2\cdot\frac{10}{21}+3^2\cdot\frac{5}{42}=\frac{140}{42}=\frac{10}{3}$ より

分散 $V(X)=E(X^2)-(E(X))^2=\frac{10}{3}-\left(\frac{5}{3}\right)^2=\frac{5}{9}$　……(答)

(2) 確率変数Yは二項分布 $\mathrm{B}\left(5,\ \frac{1}{3}\right)$ に従うから

$$E(Y)=5\cdot\frac{1}{3}=\frac{5}{3},\quad V(Y)=5\cdot\frac{1}{3}\cdot\frac{2}{3}=\frac{10}{9}$$　……(答)

研究　1° 確率変数Xが二項分布 $\mathrm{B}(n,\ p)$ に従うとき，

$$\boldsymbol{E(X)=np,\quad V(X)=npq\quad (q=1-p)}$$

であることは知っておくとよい．また，このまま使ってよい公式である．

2° $E(X+Y)=E(X)+E(Y)$ という性質から各回の平均（期待値）がいずれもpなら，$E(X)=np$ となるのは当然であるが計算で示しておこう．

$r{}_nC_r=n{}_{n-1}C_{r-1}$ であることを使うと（☞ 標問 25）

$$E(X)=\sum_{r=0}^{n} r\,{}_n\mathrm{C}_r p^r q^{n-r}=\sum_{r=1}^{n} n\,{}_{n-1}\mathrm{C}_{r-1} p^r q^{n-r}$$

$$=np\sum_{k=0}^{n-1} {}_{n-1}\mathrm{C}_k p^k q^{n-1-k} \quad (k=r-1 \text{ とおいた})$$

$$=np(p+q)^{n-1} \quad (\because \text{ 二項定理})$$

$$=np$$

3° 分散については，二項係数の次の性質を使って計算する．

$$r^2{}_n\mathrm{C}_r=\{r(r-1)+r\}{}_n\mathrm{C}_r=n(n-1)_{n-2}\mathrm{C}_{r-2}+n_{n-1}\mathrm{C}_{r-1}$$

$$V(X)=E(X^2)-(E(X))^2=\sum_{r=0}^{n} r^2{}_n\mathrm{C}_r p^r q^{n-r}-(np)^2$$

$$=\sum_{r=2}^{n} n(n-1)_{n-2}\mathrm{C}_{r-2} p^r q^{n-r}+\sum_{r=1}^{n} n_{n-1}\mathrm{C}_{r-1} p^r q^{n-r}-(np)^2$$

$$=n(n-1)p^2(p+q)^{n-2}+np(p+q)^{n-1}-(np)^2$$

$$=n(n-1)p^2+np-(np)^2$$

$$=np-np^2=npq$$

4° $(q+px)^n=\sum_{r=0}^{n} {}_n\mathrm{C}_r p^r q^{n-r} x^r$ の微分を使っても証明できる．

▶演　習◀

【60】 数列 $\{a_n\}$，$\{b_n\}$ はそれぞれ1枚の銅貨を投げて表が出たら1，裏が出たら0とすることによって得られる0，1からなる数列である．

$S_n=\sum_{i=1}^{n} a_i b_i$ の値が k $(0\leqq k\leqq n)$ となる確率を $P_n(k)$ とする．

(1) $P_2(k)$ $(k=0,\ 1,\ 2)$ を求めよ．

(2) $P_n(k)$ を求めよ．

(3) S_n の平均（期待値）E_n を求めよ．　（横浜市大）

【61】 大小2個のさいころを同時に投げる試行をTとする．1回の試行Tでさいころが2個とも偶数の目の出る事象をAとする．

(1) 試行Tをくり返すとき，n 回目に初めて事象Aが起こる確率が 0.01 以下となる最小の n を求めよ．ただし，$\log_{10}2=0.301$，$\log_{10}3=0.477$ として計算せよ．

(2) 試行Tをくり返すとき，n 回目に事象Aが起これば $X_n=1$，事象Aが起こらなければ $X_n=0$ とし，$S=X_1+X_2+X_3+X_4$ とする．このとき，S の期待値（平均値）$E(S)$ を求めよ．　（北　大）

標問 82. 〈二項分布・グラフ〉

確率変数Xが二項分布

$$P(X=k)={}_2C_k p^k(1-p)^{2-k}\quad (k=0,\ 1,\ 2)$$

に従うとする．ただし，$0\leqq p\leqq 1$ である．

(1) Xの平均値mおよび標準偏差σを求めよ．

(2) $1>m+\sqrt{6}\sigma$ となる p の範囲および，$2>m+\sqrt{6}\sigma$ となる p の範囲を求めよ．

(3) 確率 $P(X>m+\sqrt{6}\sigma)$ を p の関数と考え $f(p)$ とおく．区間 $0\leqq p\leqq 1$ における $f(p)$ の変化を調べ，グラフをかけ．

（信州大）

解答 (1) $m=2p,\ \sigma=\sqrt{2p(1-p)}$ ……(答)

(2) $1>m+\sqrt{6}\sigma$ に(1)の $m,\ \sigma$ を代入すると

$$1-2p>\sqrt{12p(1-p)}$$

$$\therefore\ (1-2p)^2>12p(1-p) \text{ かつ } 1-2p>0$$

$$16p^2-16p+1>0 \text{ かつ } 1-2p>0$$

また，$0\leqq p\leqq 1$ であるから $0\leqq p<\dfrac{2-\sqrt{3}}{4}$ ……(答)

$2>m+\sqrt{6}\sigma$ についても $1-p>\sqrt{3p(1-p)}$ を解いて

$$0\leqq p<\frac{1}{4}$$ ……(答)

(3) $m+\sqrt{6}\sigma$ なので(2)より

$0\leqq p<\dfrac{2-\sqrt{3}}{4}$ のとき，$0\leqq m+\sqrt{6}\sigma<1$ ゆえ

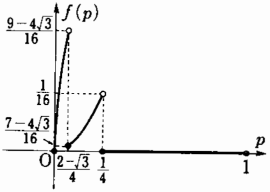
$$f(p)=P(X\geqq 1)=P(X=1)+P(X=2)$$
$$=2p(1-p)+p^2=-(p-1)^2+1$$

$\dfrac{2-\sqrt{3}}{4}\leqq p<\dfrac{1}{4}$ のとき，

$$f(p)=P(X>1)$$
$$=P(X=2)=p^2$$

$\dfrac{1}{4}\leqq p\leqq 1$ のとき，$f(p)=0$

グラフは右図のようになる．



標問 83.　〈二項分布・$E(X^2)$〉

さいころを n 回投げ，6の目が r 回出れば，ar^2+br+c（a，b，c は定数で $a>0$）の金額のお金が得られるものとする．得られる金額の平均値（期待値）m と $\dfrac{an^2}{36}+\dfrac{bn}{6}+c$ のどちらが大きいか．　（名　大）

精講　1°　6の目が出る回数をXとすると

$$P(X=r)={}_n\mathrm{C}_r\left(\frac{1}{6}\right)^r\left(\frac{5}{6}\right)^{n-r}$$

であり，Xは二項分布に従う．求める値は平均の性質より

$$E(aX^2+bX+c)=aE(X^2)+bE(X)+c$$

とほぐして計算するとよい．

2°　$E(X)$ は二項分布の公式からすぐわかるが，$E(X^2)$ はどうするか，直接計算してもできるが，分散を利用して

$$V(X)=E(X^2)-(E(X))^2 \quad \therefore\ E(X^2)=V(X)+(E(X))^2$$

とするのがはやい．こう考えていくと，

$$\frac{an^2}{36}+\frac{bn}{6}+c=a(E(X))^2+bE(X)+c$$

であるから，本問は $E(X^2)$ と $(E(X))^2$ の大小比較の問題である．

解答　6の目が出る回数をXとすると，Xは確率変数であり二項分布 $\mathrm{B}\left(n,\ \dfrac{1}{6}\right)$ に従う．

$$\therefore\ E(X)=n\cdot\frac{1}{6}=\frac{n}{6},\ V(X)=n\cdot\frac{1}{6}\cdot\frac{5}{6}=\frac{5n}{36}$$

ここで，$V(X)=E(X^2)-(E(X))^2$ より $E(X^2)=\dfrac{5n}{36}+\dfrac{n^2}{36}$

$$\begin{aligned}\therefore\ m&=E(aX^2+bX+c)\\&=aE(X^2)+bE(X)+c\\&=a\left(\frac{5n}{36}+\frac{n^2}{36}\right)+b\cdot\frac{n}{6}+c\\&>\frac{an^2}{36}+\frac{bn}{6}+c\end{aligned}$$

よって，mの方が大きい．　……(答)

標問 84.　〈正規分布曲線〉

(1) 標準正規分布の曲線 $y=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ の増減を調べ，かつ変曲点を求め，このグラフの概形をかけ．

(2) Xは，平均値がm，標準偏差がσの正規分布に従う確率変数とする．

(a) 確率 $P(\alpha\leqq X\leqq\beta)$ を，関数 $\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ の定積分の形で表せ．ただし，$\alpha<\beta$ とする．

(b) a を正の定数とするとき，次の不等式を満足する正数 ε の範囲を求めよ．

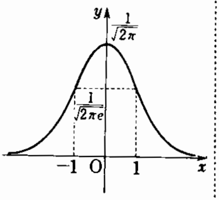
$$P(|X-m|\leqq\varepsilon)>\sqrt{\frac{2}{\pi}}\int_0^a e^{-\frac{x^2}{2}}dx$$

（岩手大）

解答 (1) $y=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$，$y'=-\frac{1}{\sqrt{2\pi}}xe^{-\frac{x^2}{2}}$，

$$y''=-\frac{1}{\sqrt{2\pi}}(1-x^2)e^{-\frac{x^2}{2}}$$

x	$-\infty$	$\cdots$	-1	$\cdots$	0	$\cdots$	1	$\cdots$	$+\infty$
y'		$+$		$+$	0	$-$		$-$	
y''		$+$	0	$-$		$-$	0	$+$	
y	0	$\nearrow$	$\frac{1}{\sqrt{2\pi e}}$	$\nearrow$	$\frac{1}{\sqrt{2\pi}}$	$\searrow$	$\frac{1}{\sqrt{2\pi e}}$	$\searrow$	0



(2) (a) 平均m，標準偏差σの正規分布曲線は

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-m)^2}{2\sigma^2}}$$

$$\therefore\quad P(\alpha\leqq X\leqq\beta)=\int_\alpha^\beta\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(X-m)^2}{2\sigma^2}}dX$$

$$=\int_{\frac{\alpha-m}{\sigma}}^{\frac{\beta-m}{\sigma}}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx \qquad \cdots\cdots(答)$$

$$\left(\because\ \frac{X-m}{\sigma}=x\ とおく\right)$$

(b)　(a)より　$P(|X-m|\leqq\varepsilon)=P(m-\varepsilon\leqq X\leqq m+\varepsilon)$

$$=2\int_0^{\frac{\varepsilon}{\sigma}}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx$$

$$>\sqrt{\frac{2}{\pi}}\int_0^{a}e^{-\frac{x^2}{2}}dx$$

$a>0$ ゆえ　$\frac{\varepsilon}{\sigma}>a$　　$\therefore$　$\varepsilon>a\sigma$　　……(答)

研究

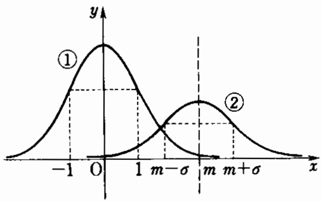
1°　(1)でえがいた標準正規分布曲線

$$y=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\quad\cdots\cdots①$$

の概形は頭に入れておくとよい．この曲線をx軸方向にσ倍，y軸方向に$\frac{1}{\sigma}$倍し，さらにx軸方向にmだけ平行移動すると，一般の正規分布曲線

$$y=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-m)^2}{2\sigma^2}}\quad\cdots\cdots②$$

が得られる．このことから，②のグラフはσが大きければ平たくなり，σが小さければとがった形になることがわかる．



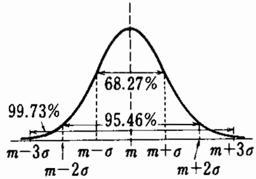
また，②の曲線は $x=m\pm\sigma$ で変曲点となることがわかる．

このσについては

$$P(a\leqq x\leqq b)=\frac{1}{\sqrt{2\pi}\sigma}\int_a^b e^{-\frac{(x-m)^2}{2\sigma^2}}dx$$

を計算すると

$$P(m-\sigma\leqq x\leqq m+\sigma)=0.6827$$
$$P(m-2\sigma\leqq x\leqq m+2\sigma)=0.9546$$
$$P(m-3\sigma\leqq x\leqq m+3\sigma)=0.9973$$



となることが知られている（この計算は高校の程度を越えるので認めてしまおう．$\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}e^{-\frac{(x-m)^2}{2\sigma^2}}dx=1$ についても，ちゃんと証明はされているので認めよう）．この値は次の節で扱う「推定・検定」で重要な意味をもつ．

2°　確率変数Xが正規分布 $\mathrm{N}(m,\ \sigma^2)$ に従うとき，

$$\text{平均}\ \boldsymbol{E(X)=m},\quad \text{分散}\ \boldsymbol{V(X)=\sigma^2}$$

であることは知っていなければいけない．これは結果を知っていて，使えればよいのである．使い方は，Xを標準化し

$$\frac{X-m}{\sigma}$$

を新しい確率変数にとれば，これは標準正規分布 N(0, 1) に従うから正規分布表より確率が計算される．本問の(2)(a)はその変形を扱ったものである．

3° 一応証明しておこう（メンドウなのでとばしてもかまわない）．

$$\text{平均}\ E(X)=\int_{-\infty}^{\infty}x\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-m)^2}{2\sigma^2}}dx$$

$$=\int_{-\infty}^{\infty}(x-m+m)\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-m)^2}{2\sigma^2}}dx$$

ここで，$\displaystyle\int_{-\infty}^{\infty}\frac{x-m}{\sqrt{2\pi}\sigma}e^{-\frac{(x-m)^2}{2\sigma^2}}dx$

$$=\frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty}ze^{-z^2}\sqrt{2}\,\sigma dz\ \left(\because\quad z=\frac{x-m}{\sqrt{2}\,\sigma}\right)$$

$$=\frac{\sqrt{2}}{\sqrt{\pi}}\sigma\left[-\frac{1}{2}e^{-z^2}\right]_{-\infty}^{\infty}=0$$

$$\therefore\quad E(X)=m\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-m)^2}{2\sigma^2}}dx$$

$$=m\quad\left(\because\quad \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-m)^2}{2\sigma^2}}dx=1\ \text{は認めている}\right)$$

$$\text{分散}\ V(X)=\int_{-\infty}^{\infty}(x-m)^2\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-m)^2}{2\sigma^2}}dx$$

$$=\int_{-\infty}^{\infty}2\sigma^2z^2\frac{1}{\sqrt{2\pi}\sigma}e^{-z^2}\sqrt{2}\,\sigma dz\quad\left(\because\quad z=\frac{x-m}{\sqrt{2}\,\sigma}\right)$$

$$=\frac{2\sigma^2}{\sqrt{\pi}}\int_{-\infty}^{\infty}z^2e^{-z^2}dz$$

$$=\frac{2\sigma^2}{\sqrt{\pi}}\left\{\left[-\frac{1}{2}ze^{-z^2}\right]_{-\infty}^{\infty}+\frac{1}{2}\int_{-\infty}^{\infty}e^{-z^2}dz\right\}\quad\text{（部分積分）}$$

$$=\frac{\sigma^2}{\sqrt{\pi}}\int_{-\infty}^{\infty}e^{-z^2}dz$$

$$=\sigma^2\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-m)^2}{2\sigma^2}}dx$$

$$=\sigma^2$$

標問 85.　〈二項分布の正規近似〉

3個のさいころを同時に投げるとき，ある1つのさいころの目が他の2個のさいころの目の和に等しい事象をE_1，3個のさいころの目の和が15以上となる事象をE_2で表すことにする．

(1)　E_1とE_2とは互いに排反な事象であるか．またE_1とE_2とは独立な事象であるか．その理由を述べて答えよ．

(2)　3個のさいころを20回投げるとき，事象E_1がちょうど5回現れる確率を小数第2位まで求めよ．ただし，事象E_1がちょうど4回現れる確率は 0.21727 である．

(3)　3個のさいころを400回投げるとき，事象E_2が少なくとも40回は現れる確率を小数第2位まで求めよ．ただし，$\int_z^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx$ の値は，$z=0.5,\ 1.0,\ 1.5,\ 2.0,\ 2.5$ について，それぞれ 0.3085，0.1587，0.0668，0.0228，0.0062 である．

（長崎大）

精講

(1)　排反と独立の区別をキチンとつけておこう．

(3)　nが十分に大きいので，“二項分布は正規分布に近似する”（これはラプラスの定理と呼ばれており証明は高校の範囲外）

解答　(1)　事象E_1における3個のさいころの目の和の最大値は12で，

$$E_1\cap E_2=\phi \qquad \therefore\ E_1 \text{ と } E_2 \text{ は排反} \qquad \cdots\cdots(\text{答})$$

また，$P(E_1)\cdot P(E_2)\neq 0$ は明らかで，$P(E_1\cap E_2)=0$ より

$$P(E_1\cap E_2)\neq P(E_1)\cdot P(E_2)$$

$\therefore$　E_1とE_2は独立な事象ではない　……(答)

(2)　1回の試行で事象E_1が起こるさいころの目の組は

$\{1,\ 1,\ 2\}$，$\{1,\ 2,\ 3\}$，$\{1,\ 3,\ 4\}$，$\{1,\ 4,\ 5\}$，$\{1,\ 5,\ 6\}$

$\{2,\ 2,\ 4\}$，$\{2,\ 3,\ 5\}$，$\{2,\ 4,\ 6\}$，$\{3,\ 3,\ 6\}$

であるから，$P(E_1)=\frac{3\cdot{}_3C_1+6\cdot 3!}{6^3}=\frac{45}{216}=\frac{5}{24}$　$(=p)$

E_1 が k 回現れる確率を $P_k(E_1)$ と表すと

$$P_5(E_1)={}_{20}C_5p^5(1-p)^{15}$$
$$=P_4(E_1)\times\frac{{}_{20}C_5}{{}_{20}C_4}\cdot p\cdot\frac{1}{1-p}=0.21727\times\frac{16}{5}\cdot\frac{5}{24}\cdot\frac{24}{19}$$
$$=0.1829\cdots\cdots$$

よって，求める値は　0.18　……(答)

(3)　1回の試行で事象 E_2 が起こるさいころの目の組は

{6, 6, 6}, {6, 6, 5}, {6, 6, 4}, {6, 6, 3},
{6, 5, 5}, {6, 5, 4}, {5, 5, 5}

であるから，$P(E_2)=\dfrac{2+4\cdot{}_3C_1+3!}{6^3}=\dfrac{20}{216}=\dfrac{5}{54}$

E_2 が現れる回数を X とおくと，X は二項分布 $B\left(400,\ \dfrac{5}{54}\right)$ に従う．

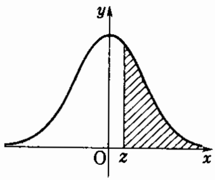
平均 $m=400\cdot\dfrac{5}{54}=\dfrac{1000}{27}$，標準偏差 $\sigma=\sqrt{400\cdot\dfrac{5}{54}\cdot\dfrac{49}{54}}=\dfrac{70\sqrt{5}}{27}$

試行回数が大きいので，

$$Z=\frac{X-m}{\sigma}=\left(X-\frac{1000}{27}\right)\times\frac{27}{70\sqrt{5}}=\frac{27X-1000}{70\sqrt{5}}$$

は標準正規分布 N(0, 1) に従うとみてよい．

$$P(X\geqq 40)=P\left(Z\geqq\frac{27\cdot 40-1000}{70\sqrt{5}}\right)$$
$$=P(Z\geqq 0.5111\cdots\cdots)$$
$$\fallingdotseq 0.3085-0.0111\times\frac{0.1587-0.3085}{1.0-0.5}$$
$$=0.3085-0.0033=0.3052$$



よって，求める値は　0.31　……(答)

▶演　習◀

【62】「次の5つの文章のうち正しいもの2つに○をつけよ.」という問題がある．いま解答者1600人が各人考えることなくでたらめに2つの文章を選んで○をつけたとする．

(1)　1600人中2つとも正しく○をつけた者が130人以上175人以下となる確率を式で表せ．

(2)　正規分布表（略）を用いて，(1)の確率を四捨五入によって小数第2位まで求めよ．(正規分布表は標問86，演習63を参照)　(広島大)

標問 86.　〈正規分布表〉

はじめの点Pの位置を数直線上の原点とする．いま次の操作に従って点Pを順次移動させる．

操作：さいころを投げて2以下の目が出ればPを負の向きに1だけ移動させる．また，3以上の目が出ればPを正の向きに1だけ移動させる．

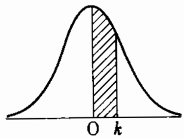
この操作を n 回行った後の点Pの座標を X_n とする．次の問いに答えよ．

(1)　確率変数 X_2 の確率分布を求めよ．

(2)　確率変数 X_n のとる値の集合は何か．

(3)　$n=450$ のとき $130 \leqq X_{450} \leqq 175$ である確率を求めよ．

(4)　$X_n \geqq \dfrac{2}{5}n$ である確率が 0.1 以下であるような n の最小値を求めよ．



（滋賀医大）

〔正規分布表の一部〕

k	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319

解答　(1)　$P(X_2=-2)=\left(\dfrac{1}{3}\right)^2=\dfrac{1}{9}$,　$P(X_2=0)={}_2\mathrm{C}_1\left(\dfrac{1}{3}\right)\left(\dfrac{2}{3}\right)=\dfrac{4}{9}$

$$P(X_2=2)=\left(\frac{2}{3}\right)^2=\frac{4}{9}$$

確率分布表にまとめると右のようになる．

X_2	-2	0	2
確率	$\dfrac{1}{9}$	$\dfrac{4}{9}$	$\dfrac{4}{9}$

(2)　n 回とも左へ移動すると $X_n=-n$ に，また右へ移動する回数が1回につき，点Pの座標は2ずつ大きくなるから，求める集合は

$$\{-n,\ -n+2,\ \cdots\cdots,\ n-2,\ n\} \qquad \cdots\cdots(\text{答})$$

(3) n 回さいころを投げて2以下の目が x 回，3以上の目が y 回出たとすると

$$\begin{cases} y-x=X_n \\ y+x=n \end{cases} \qquad \therefore\ y=\frac{X_n+n}{2}$$

$$\therefore\ P(130 \leqq X_{450} \leqq 175)=P\left(\frac{130+450}{2} \leqq y \leqq \frac{175+450}{2}\right)$$
$$=P(290 \leqq y \leqq 312)$$

y は二項分布 $\mathrm{B}\left(450,\ \frac{2}{3}\right)$ に従うから，

$$\text{平均}\ m=450 \cdot \frac{2}{3}=300,\ \text{標準偏差}\ \sigma=\sqrt{450 \cdot \frac{2}{3} \cdot \frac{1}{3}}=10$$

よって，求める確率は正規分布表より

$$P\left(-1 \leqq \frac{y-m}{\sigma} \leqq 1.2\right)=0.3413+0.3849=0.7262 \qquad \cdots\cdots(\text{答})$$

(4) $X_n=2y-n \geqq \frac{2}{5}n \qquad \therefore\ y \geqq \frac{7}{10}n$

$$\frac{y-m}{\sigma} \geqq \left(\frac{7}{10}n-\frac{2}{3}n\right) \times \frac{3}{\sqrt{2n}}=\frac{\sqrt{2n}}{20}$$

$$\therefore\ P\left(X_n \geqq \frac{2}{5}n\right)=P\left(\frac{y-m}{\sigma} \geqq \frac{\sqrt{2n}}{20}\right) \leqq 0.1$$

正規分布表より

$$\frac{\sqrt{2n}}{20} \geqq 1.28 \qquad \therefore\ n \geqq 327.68 \qquad \therefore\ n=328 \qquad \cdots\cdots(\text{答})$$

▶演　習◀

【63】 ある男子高校の生徒全体の身長は平均 165 cm，標準偏差 6 cm の正規分布に従うと仮定できるという．このとき，下の正規分布表を用いて，次の(1)，(2)に答えよ．

z	0.5	1.0	1.5	2.0	2.5	3.0
$\Phi(z)$	0.191	0.341	0.433	0.477	0.494	0.499

ただし，

$$\Phi(z)=\frac{1}{\sqrt{2\pi}}\int_0^z e^{-\frac{x^2}{2}}dx$$

である．

(1) 身長が 159 cm 以上 171 cm 以下である生徒数の全体に対する割合を求めよ．

(2) 無作為に1人の生徒を選んだとき，その生徒の身長が 180 cm 以上である確率を求めよ． （宮崎大）

II．推定・検定

1°　**標本平均**　ある母集団から抽出された n 個の変量 X_1, X_2, ……, X_n に対し

$$\overline{X}=\frac{X_1+X_2+\cdots\cdots+X_n}{n}$$

を標本平均という．母平均 m，母分散 σ^2 の母集団から大きさ n の標本を復元抽出するとき，標本平均 $\overline{X}$ の平均と分散はそれぞれ

$$E(\overline{X})=m,\quad V(\overline{X})=\frac{\sigma^2}{n}$$

である．

2°　**中心極限定理**　母平均 m，母分散 σ^2 の母集団（正規分布でなくてもよい）から抽出された大きさ n の標本平均 $\overline{X}$ は，n が大きければ

$$\text{正規分布}\quad \mathrm{N}\left(m,\ \frac{\sigma^2}{n}\right)$$

に近似できる．

3°　**推定**　母分散 σ^2 の母集団から抽出した大きさ n の標本の標本平均が $\overline{X}$ であるとき，n が十分大きければ，母平均 m の信頼区間は

$$\text{信頼度 95\% のとき}\quad \overline{X}-1.96\frac{\sigma}{\sqrt{n}}<m<\overline{X}+1.96\frac{\sigma}{\sqrt{n}},$$

$$\text{信頼度 99\% のとき}\quad \overline{X}-2.58\frac{\sigma}{\sqrt{n}}<m<\overline{X}+2.58\frac{\sigma}{\sqrt{n}}$$

である．（n が十分大きいとき，σ を標本の標準偏差で代用してもよい．）

4°　**検定**　母平均が m であるという仮説をたてたとき，母集団から任意抽出した大きさ n の標本平均が $\overline{X}$ であったとする．

$$Z=\frac{\overline{X}-m}{\frac{\sigma}{\sqrt{n}}}\quad \text{は}\ \mathrm{N}(0,\ 1)\ \text{に従うので}$$

有意水準5％のとき　$|Z|\geqq 1.96$ なら仮説を捨てる．

有意水準1％のとき　$|Z|\geqq 2.58$ なら仮説を捨てる．

（注）　仮説を捨てるとき，検定の結果は有意であるという．有意水準5％というのは，仮説を捨てたときの危険率（仮説が正しいにもかかわらずこれを捨てる）が5％より小さいということである．正しい仮説を捨ててしまう誤りを第1種の過誤という．

標問 87.　〈標本平均〉

ある県の20歳の男子を母集団とするとき，その身長は平均165cm，標準偏差 4cm の正規分布をするという．いま，

$F(x)=\int_0^x \frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt$ とするとき，x と $F(x)$ の対応する値が右のように与えられているものとする．このとき，次の各問いに答えよ．

x	0.126	1.000	1.645	2.000
$F(x)$	0.050	0.341	0.450	0.477

(1)　この母集団で身長の上位5％の者は何cm以上と考えられるか．

(2)　この母集団から無作為に64人の標本を抽出したとき，その標本平均が 164cm 以上でかつ 166cm 以下である確率を求めよ．

（宮崎大）

解答　(1)　求める身長をxとすると

$$P(X\geqq x)=P\left(\frac{X-165}{4}\geqq\frac{x-165}{4}\right)\leqq 0.05$$

$\frac{X-165}{4}$ は N(0, 1) に従うので，表より $\frac{x-165}{4}\geqq 1.645$

$\therefore$　$x=165+6.580=171.58$cm 以上　……(答)

(2)　標本平均$\overline{X}$の標準偏差は　$\frac{4}{\sqrt{64}}=0.5$

$$\therefore\quad P(164\leqq\overline{X}\leqq 166)=P\left(-2\leqq\frac{\overline{X}-165}{0.5}\leqq 2\right)$$

$$=2P\left(0\leqq\frac{\overline{X}-165}{0.5}\leqq 2\right)=2\times 0.477=0.954$$　……(答)

研究　1°　用語が正確に理解されているか，正規分布表が使えるかをためす問題である．

2°　標本平均$\overline{X}$の平均は $E(\overline{X})=m$（母平均），分散は $V(\overline{X})=\frac{\sigma^2}{n}$（$\sigma^2$ は母分散，n は標本の大きさ）である．分散について示すと

$$V(\overline{X})=V\left(\frac{X_1+X_2+\cdots+X_n}{n}\right)=\frac{1}{n^2}(V(X_1)+V(X_2)+\cdots+V(X_n))=\frac{\sigma^2}{n}$$

標問 88. 〈度数分布表・母平均の推定〉

高校3年生の野球選手から任意に100人抽出して，その身長の測定値（単位は cm）を8つの階級に整理したものが下の表である．

階級番号 i	身長の階級	階級値 x_i	人数 f_i	$u_i=(x_i-170)/3$	u_if_i	$u_i^2f_i$
1	160～162	161	1	-3	-3	9
2	163～165	164	13	-2	-26	52
3	166～168	167	17	-1	-17	17
4	169～171	170	19	0	0	0
5	172～174	173	23	1	23	23
6	175～177	176	12	2	24	48
7	178～180	179	11	3	33	99
8	181～183	182	4	4	16	64
計			100		50	312

(1)　u_i の平均 $\overline{u}$ を用いて標本平均 $\overline{x}$ を表し，$\overline{x}$ を計算せよ．

(2)　$\overline{u}$，$\displaystyle\sum_{i=1}^{8} u_i{}^2 f_i$ を用いて標本分散 s^2 を表し，s^2 を計算せよ．

(3)　高校3年生の野球選手の身長は平均m，分散 26.01 の正規分布に従っていると仮定できるとき，平均mの信頼区間を信頼度95％で（小数第1位まで）求めよ．ただし，

$\displaystyle\int_0^{1.96}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=0.475$ である．　（長崎大）

解答　(1)　仮平均が 170cm，階級幅が 3cm であるから

$$\overline{x}=170+3\overline{u}=170+3\times\frac{50}{100}=171.5\,\text{cm}$$ ……(答)

(2)　$$s^2=E(\overline{x}^2)-(E(\overline{x}))^2=3^2\left(\frac{1}{100}\sum_{i=1}^{8}u_i{}^2f_i-\overline{u}^2\right)=25.83$$ ……(答)

(3)　$2\displaystyle\int_0^{1.96}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=0.95$ より 95％ の信頼度でのmの信頼区間は

$$\overline{x}-1.96\sqrt{\frac{26.01}{100}}<m<\overline{x}+1.96\sqrt{\frac{26.01}{100}}$$

$$\therefore\quad 170.5<m<172.5$$ ……(答)

研究

1° 度数分布表から平均 $\bar{x}$, 分散 s^2 を計算するには, 階級値, 度数をそれぞれ x_i, f_i $(1 \leqq i \leqq k)$ とおけば

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{k} x_i f_i,\quad s^2=\frac{1}{n}\sum_{i=1}^{k} x_i^2 f_i-\bar{x}^2 \quad \left(n=\sum_{i=1}^{k} f_i\right)$$

として求められるが, $x_i f_i$, $x_i^2 f_i$ の値が大きくなり計算がメンドウになる. 簡便計算法として, 本問のように $\bar{x}$ の近似値 x_0 を適当に選んで

$$u_i=\frac{x_i-x_0}{c} \quad (c\text{ は階級幅})$$

として変数をとりかえると, u_i の平均を $\bar{u}$ として

$$\bar{x}=x_0+c\bar{u}=x_0+\frac{c}{n}\sum_{i=1}^{k} u_i f_i$$

$$s^2=c^2\left(\frac{1}{n}\sum_{i=1}^{k} u_i^2 f_i-\left(\frac{1}{n}\sum_{i=1}^{k} u_i f_i\right)^2\right)$$

2° 母集団の母平均, 母分散をそれぞれ m, σ^2 とするとき, 大きさ n の標本平均 $\overline{X}$ は

$m-\sigma<\overline{X}<m+\sigma$ のとき 68.27%

$m-2\sigma<\overline{X}<m+2\sigma$ のとき 95.46%

$m-3\sigma<\overline{X}<m+3\sigma$ のとき 99.73%

の区間に存在することは標問 **84** の研究でみた. 99.73%ならほぼ確実にこの区間にあるといえるが, 区間の幅が広すぎる. 幅をせまくしようとすると確率がさがる. このかねあいを**信頼度**ということばで表す. 信頼度は95%と99%が使われることが多い. 信頼度95%の区間(信頼区間)は, $\overline{X}$ を標準化して $Z=\dfrac{\overline{X}-m}{\frac{\sigma}{\sqrt{n}}}$ とおくと $P(|Z|<\alpha)=0.95$ であり, 正規分布表から $\alpha=1.96$ であるから

$$\therefore\quad -1.96<Z<1.96$$

$$\therefore\quad m-1.96\frac{\sigma}{\sqrt{n}}<\overline{X}<m+1.96\frac{\sigma}{\sqrt{n}}$$

m について整理し直して

$$\overline{X}-1.96\frac{\sigma}{\sqrt{n}}<m<\overline{X}+1.96\frac{\sigma}{\sqrt{n}}$$

これは母平均 m が確率95%でこの区間に含まれていることを意味する.

信頼度を99%にとると $\overline{X}-2.58\dfrac{\sigma}{\sqrt{n}}<m<\overline{X}+2.58\dfrac{\sigma}{\sqrt{n}}$ となる.

標問 89. 〈標本個数〉

弱い酸による布地の損傷を実験するのに，その酸につけた布地が使用に耐えなくなるまでの時間を測ることにした．このようにして与えられる実験データの平均が，真の値と，母集団標準偏差の10%以上違わないことが，95%以上の確率で正しいといえるためには，何個のデータをとればよいか．ただし，時間は正規分布に従うものとする．また，$\int_0^{1.96}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=0.47500$ である．

（宮崎医大）

解答　母集団の平均時間を m，標準偏差を σ とする．大きさ n の標本の平均時間を $\overline{X}$ とすると

$$P(|\overline{X}-m|<0.1\sigma)\geqq 0.95$$

$$\therefore\quad P\left(\frac{|\overline{X}-m|}{\frac{\sigma}{\sqrt{n}}}<0.1\sqrt{n}\right)\geqq 0.95$$

$2\int_0^{1.96}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=0.95$ であるから

$$\frac{|\overline{X}-m|}{\frac{\sigma}{\sqrt{n}}}\leqq 1.96<0.1\sqrt{n}$$

$$\therefore\quad n>(19.6)^2=384.16$$

よって，385個以上のデータをとればよい．　……(答)

▶演　習◀

【64】 ある年の20歳の男子学生（母集団）の体重 X kg は，平均値 59.8，標準偏差 6.9 の正規分布に従うという．この母集団からランダムに25人からなる標本を何度も抽出するとき，標本平均 $\overline{X}$ は平均値$\square$，標準偏差$\square$の$\square$分布に従う．また n 人からなる標本をランダムに抽出するとき，その体重の平均値が95%の信頼度で，幅 1kg 以内（誤差 0.5kg 以内）の信頼区間に入るようにするには，n を$\square$とすればよい．

（旭川医大）

標問 90. 〈母比率の推定〉

ある原野には，A，B 2種の野ねずみが生息しているという．任意に300匹の野ねずみを捕えたところ，A種が90匹いた．A種の野ねずみは，この原野全体で何%生息していると考えられるか．信頼度95%で推定せよ． (旭川医大)

精講 **比率の推定** n 個の標本の中に，性質Aをもつものが r 個あれば，$\dfrac{r}{n}$ を**標本比率**といい $\overline{p}$ で表す．母集団における性質Aの比率 p を**母比率**という．信頼度95%で p の信頼区間を推定してみよう．

n 個の標本のうち性質Aをもつものの個数を X とすれば，X は二項分布 $B(n,\ p)$ をなす．n が十分大きいときには正規分布 $N(np,\ np(1-p))$ で近似できるから

$$P\left(\frac{|X-np|}{\sqrt{np(1-p)}}<1.96\right)=0.95$$

$$\therefore\quad X-1.96\sqrt{np(1-p)}<np<X+1.96\sqrt{np(1-p)}$$

$$\therefore\quad \overline{p}-1.96\sqrt{\frac{p(1-p)}{n}}<p<\overline{p}+1.96\sqrt{\frac{p(1-p)}{n}}$$

n が十分に大きいときは根号内の p を $\overline{p}$ にかえてもよく

$$\overline{p}-1.96\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}<p<\overline{p}+1.96\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}$$

信頼度99%では $\overline{p}-2.58\sqrt{\dfrac{\overline{p}(1-\overline{p})}{n}}<p<\overline{p}+2.58\sqrt{\dfrac{\overline{p}(1-\overline{p})}{n}}$ となる．

解答 $\overline{p}=\dfrac{90}{300}=\dfrac{3}{10}$ より $1.96\sqrt{\dfrac{\overline{p}(1-\overline{p})}{300}}=\dfrac{1.96\sqrt{7}}{100}=0.0518\cdots\cdots$

よって，求める区間は 24.82% から 35.18% ……(答)

▶演 習◀

【65】 新しい薬を作っているある工場で，大量の製品全体の中から任意に1000個を抽出して検査を行ったところ，20個の不良品があった．この製品全体について不良率を，二項分布の計算には正規分布を用い，95%の信頼度で推定せよ． (山梨医大)

標問 91.　〈検定〉

(1) あるところにきわめて多くの白球と黒球がある．400個の球を無作為に取り出したとき，白球が222個，黒球が178個あった．白球と黒球との割合は同じであるという仮説を有意水準5％で検定せよ．ただし，Zが標準正規分布に従うとき$P(|Z|\geqq 1.96)=0.05$ とする．　（中央大）

(2) ある種のメダカの黒色個体と白色個体とを交配させたところ，黒色個体ばかりを得た．この第2代の黒色個体どうしを交配させた結果，黒色個体162尾，白色個体63尾が生じた．このメダカの体色の遺伝が，メンデルの法則に従うとすれば，第3代の体色の分離比は 3：1 となるはずである．この実験結果がメンデルの法則に矛盾するか，しないかを危険率5％で検定せよ．　（旭川医大）

解答 (1)「白球と黒球の割合は同じである」と仮定する．400 個の球を取り出したときの白球の個数をXとすると，Xは

$$\text{正規分布}\ \mathrm{N}\left(400\times\frac{1}{2},\ 400\times\frac{1}{2}\times\frac{1}{2}\right)=\mathrm{N}(200,\ 100)$$

に従う．$Z=\dfrac{X-200}{10}$ は N(1, 0) に従うので，有意水準5％では

$Z=\dfrac{222-200}{10}=2.2>1.96$　∴ 仮説は棄却される．　……(答)

(2)「第3代の体色の分離比は 3：1 である」と仮定する．
225尾中の黒色個体の個数をXとすると，Xは

$$\text{正規分布}\ \mathrm{N}\left(225\times\frac{3}{4},\ 225\times\frac{3}{4}\times\frac{1}{4}\right)=\mathrm{N}\left(\frac{675}{4},\ \left(\frac{15\sqrt{3}}{4}\right)^2\right)$$

に従う．$Z=\dfrac{X-\dfrac{675}{4}}{\dfrac{15\sqrt{3}}{4}}$ は N(1, 0) に従う．$X=162$ では

$$|Z|=\left|162-\frac{675}{4}\right|\times\frac{4}{15\sqrt{3}}=\frac{3\sqrt{3}}{5}\fallingdotseq 1.04<1.96\quad(\text{危険率}5\%)$$

仮説は捨てられないから，矛盾しているとはいえない．　……(答)

研究

1° 検定の手順をまとめておこう.

（i） 仮説をたてる.

（ii） 有意水準（危険率）から, 棄却域を求める.

（iii） 実現値が棄却域の中にあるかどうかを調べる.

2° まず, 仮説のたて方であるが, 検定とはそもそも確率的背理法を基礎に議論されるものであって“仮説は捨てられる”方向にたてるものである. 本問(2)のように仮説が捨てられない場合もあるが, ここで出す結論は積極的に仮説を採用するというのではなく, 捨てるだけの理由がないというだけである. この意味で仮説を**帰無仮説**ともいう.

3° 次に, 仮説を正しいとして, 信頼度95%で標本平均$\overline{X}$のとり得る範囲を求める. すなわち

$$-1.96 < Z = \frac{\overline{X}-m}{\frac{\sigma}{\sqrt{n}}} < 1.96 \qquad \therefore \quad |Z| < 1.96$$

である. Zがこの範囲に入っているなら仮説と標本結果は矛盾しないと判断し仮定は捨てられない. この範囲を越えたときに仮定は捨てられるが, そこには正しい仮説であるにもかかわらず捨てる可能性が5%残っている. この意味から“危険率5%”といういい方をする. また, 仮説を捨てると結論したときは, この検定に意味があったわけで“有意水準5%”ともいう.

$$\textbf{棄却域}\begin{cases} |Z| \geqq 1.96 \ （\text{有意水準}\ 5\%） \\ |Z| \geqq 2.58 \ （\text{有意水準}\ 1\%） \end{cases}$$

4° 実現値が棄却域にあるとき仮説は捨てられるが, 3°で述べたように“正しい仮説を捨てる危険”がいつもあり, この誤りを**第一種の過誤**という. これに対し“正しくない仮説を採択する危険”もあるわけで, この誤りを**第二種の過誤**という. この誤りは危険率(有意水準)の中には含まれない. 危険率というのは第一種の過誤に関するものだけである. 第二種の過誤は棄却域のとり方に依存する.

▶演　習◀

【66】 B型の薬の有効率（服用して効き目のある確率）は0.6であるといわれている. A型の薬を200人の患者に与えたところ, 134人の患者に効き目があったという. A型の薬はB型の薬より, 優れているといえるか. 有意水準5%で検定せよ.　（旭川医大）

標問 92.　〈検定〉

(1)　確率変数Xのとり得る値が 0, 1, 2, ……, n であり，その確率分布が二項分布

$$P(X=k)={}_nC_k p^k(1-p)^{n-k} \quad (0\leqq p\leqq 1)$$

に従うとき，Xの平均値mと標準偏差σとを求めよ．

(2)　n が大きいとき(1)の確率変数Xは，ほぼ正規分布 $N(m,\ \sigma^2)$ に従うとみて，右の表を用いて次の検定をせよ．

1枚の硬貨を800回投げたら表が430回出た．危険率（有意水準）5％で考えるとき，この硬貨は正しく作られていないといえるか．また，危険率1％ではどうか．ただし，正しい硬貨を投げたとき表の出る確率は $\dfrac{1}{2}$ とする．

（山梨医大）

正規分布 $N(0,\ 1)$ の表の一部

k	$P(0\leqq X\leqq k)$
1.6	0.445
1.7	0.455
1.8	0.464
1.9	0.471
2.0	0.477
2.1	0.482
2.2	0.486
2.3	0.489
2.4	0.492
2.5	0.494

解答　(1)　$m=np,\ \sigma=\sqrt{np(1-p)}$　……(答)

(2)　「硬貨は正しく作られている」と仮定する．

表が出る回数をXとするとXは二項分布 $B\left(800,\ \dfrac{1}{2}\right)$ に従うから

$$m=800\times\frac{1}{2}=400,\quad \sigma=\sqrt{800\times\frac{1}{2}\times\frac{1}{2}}=10\sqrt{2}$$

Xは正規分布 $N(400,\ 200)$ に従い，$Z=\dfrac{X-400}{10\sqrt{2}}$ は $N(0,\ 1)$ に従う．$X=430$ のとき　$|Z|=\dfrac{430-400}{10\sqrt{2}}\fallingdotseq 2.12$

表より，$P(|Z|\geqq 2.0)=0.046<0.05$，$P(|Z|\geqq 2.5)=0.012>0.01$

仮説は危険率5％では棄却され，危険率1％では採択される．

……(答)

標問 93. 〈片側検定〉

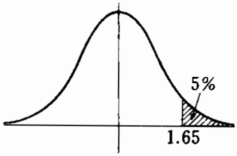
A市内の高校3年男子の，昨年度の平均身長は 168.2cm であり，これは一昨年度の平均身長よりも高かった．今年度もこの傾向があるか否かを調べるために100名を任意抽出して平均身長を求めたところ 169.3cm であり，また，その標準偏差は 5.8cm であった．今年度も平均身長増加の傾向が続いていると判定されるか．有意水準5％で（片側）検定せよ．

精講

1° 帰無仮説が否定されたとき採用される仮説を対立仮説というが，対立仮説には2種類ある．帰無仮説 $m=m_1$ に対して対立仮説は （i） $m \neq m_1$，（ii） $m > m_1$（または $m < m_1$）が考えられる．（i）は両側検定，（ii）は片側検定を行うが，どちらを使うかは問題によって判断する．

2° 片側検定では次のように棄却域が変わる．

計算方法は両側検定の場合と同じである．



図のような**右片側検定では**

有意水準5％のとき $Z \geqq 1.65$ なら仮説を捨てる．

有意水準1％のとき $Z \geqq 2.33$ なら仮説を捨てる．

左片側検定では

有意水準5％のとき $Z \leqq -1.65$ なら仮説を捨てる．

有意水準1％のとき $Z \leqq -2.33$ なら仮説を捨てる．

解答 「今年度の平均身長も昨年度と同じ 168.2cm である」と仮定．

n が十分大きいので母標準偏差を標本標準偏差 5.8 で代用し，標本平均を $\overline{X}$ で表すと

$$Z=\frac{\overline{X}-168.2}{\frac{5.8}{\sqrt{100}}}$$ は正規分布 N(0, 1) に従う．

$\overline{X}=169.3$ のとき

$$Z=\frac{10(169.3-168.2)}{5.8}=1.90>1.65$$

よって，仮説は棄却される．すなわち，今年度も平均身長増加の傾向は続いていると判定される． ……(答)