

11 母比率の区間推定

11.1 2項分布

$P[X = 1] = p$, $P[X = 0] = q = 1 - p$ である様な確率変数 X :

$$E[X] = p, \quad \text{Var}[X] = E[X^2] - p^2 = p(1 - p) = pq$$

母集団 X からとった大きさ n の標本の和 $S = X_1 + \dots + X_n$ に対して

$$P[S = k] = \binom{n}{k} p^k q^{n-k}, \quad k = 0, 1, 2, \dots, n$$

k	0	1	2	...	k	...	n
$P[S = k]$	q^n	npq^{n-1}	$\binom{n}{2}p^2q^{n-2}$	...	$\binom{n}{k}p^kq^{n-k}$	...	p^n

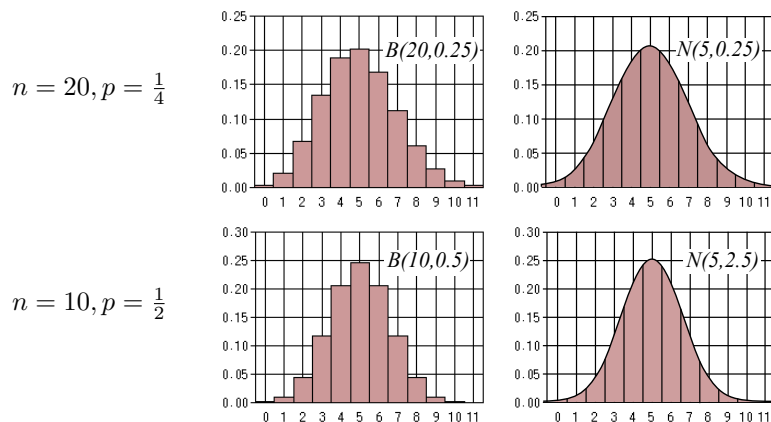
このような分布を 2 項分布 $B(n, p)$ と言います。平均値は np 、分散は npq です。

11.2 2項分布の正規分布近似

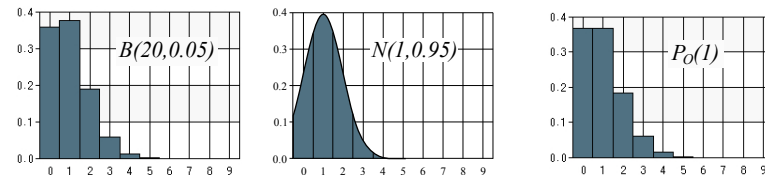
事実 11.2.1 [2 項分布の正規分布による近似 (半整数補正あり)] 2 項分布 $B(n, p)$ は、 $np \geq 5, nq \geq 5$ を満たしていれば正規分布 $N(np, npq)$ で次のように近似されます ($q = 1 - p$) :

$$P[v \leq B(n, p) \leq w] \approx P\left[v - \frac{1}{2} \leq N(np, npq) \leq w + \frac{1}{2}\right].$$

$np \geq 10$ ならかなり良く近似されます。特に p が $\frac{1}{2}$ に近ければ精度が良くなります。



p がとても小さい場合には正規分布近似は粗すぎるので Poisson 分布で近似します。



問題 11.2.2 [教科書 例題 15.4] サイコロを n 回繰り返し投げた時に 6 の目が出る回数を X_n とします。600 回繰り返し投げたとき、 $P[90 \leq X_{600} \leq 110]$ を求めて下さい。

11.3 母比率の区間推定

問題 11.3.1 ある番組の視聴率を調べるために、500 世帯を無作為抽出して調査したところ 45 世帯がその番組を見ていたことが分かりました。この番組の視聴率 (全世帯中の視聴世帯の割合) はどれくらいであると考えられるでしょうか。信頼度 95 % で区間推定してみましょう。

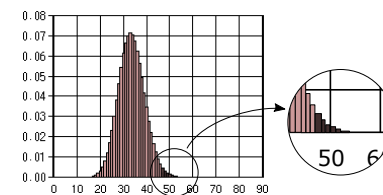
視聴率を p と置くと 500 世帯抽出した時の視聴数 X は 2 項分布 $B(500, p)$ に従います。

信頼度 95 % で考える場合は、大きい方と小さい方それぞれに 2.5 % の区域が設定され、45 はそこに入らないと考えられます。

[p が小さいとき]

$$P[45 \leq B(500, p)] = 0.025$$

となる p の値を求めます。



$$0.25 = P[45 \leq B(500, p)] \approx P[44.5 \leq N(500p, 500p(1-p))] = P\left[\frac{44.5 - 500p}{\sqrt{500p(1-p)}} \leq N(0, 1)\right]$$

$$0.475 \approx P\left[0 \leq N(0, 1) \leq \frac{44.5 - 500p}{\sqrt{500p(1-p)}}\right] \Rightarrow \frac{44.5 - 500p}{\sqrt{500p(1-p)}} \approx 1.96$$

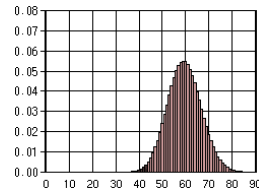
$$251920.8p^2 - 46420.8p + 1980.25 = 0$$

これを解くのは容易ではありません (計算機なら問題ない。解は 0.0671)。

【 p が大きいとき】同様に、

$$P[B(500, p) \leq 45] = 0.025$$

となる値 p を求めると 0.119425 が得られます。



以上から、 $0.0671 \leq p \leq 0.119$ を『母比率の信頼度 95 %の信頼区間』と言います。

手計算なら、 X の分散をサンプル値 $p = \frac{45}{500}$ を使った

$$500 \cdot \frac{45}{500} \left(1 - \frac{45}{500}\right) = 9 \cdot 4.55$$

で代用し X を正規分布 $N(500p, 9 \cdot 4.55)$ で近似します：

$$0.95 = P[|X - 500p| \leq d] \approx P\left[|N(500p, 9 \cdot 4.55) - 500p| \leq d + \frac{1}{2}\right]$$

あとは正規母平均の場合と同じように信頼区間を求めます ($0.066 \leq p \leq 0.114$)。

11.4 出口調査

問題 11.4.1 ある選挙において出口調査を実施したところ、100 人中に A 候補に投票したと言う人は 40 人、B 候補に投票したと言う人は 30 人いました。A 候補の得票率 p_A と B 候補の得票率 p_B それぞれの信頼度 99 %の信頼区間を求めてください。そしてその結果 A 候補は当選確実であると言えるでしょうか。半整数補正はしなくて結構です。

11.5 問題演習

●正規分布による近似

基本演習 11.1 2 項分布 $B(3, p)$ に従う確率変数 X の平均値 (期待値)・分散を分布表から自分の手で計算して下さい。

基本演習 11.2 [教科書 例題 15.6] 正常な硬貨を 400 回投げて表の出る回数を確率変数 X とするとき、 $P[190 \leq X \leq 210]$ を求めて下さい。

基本演習 11.3 [教科書 練習問題 16-7] 正常なサイコロをくりかえし投げて実際に 1 の目が出る割合とその数学的確率 $\frac{1}{6}$ との差が 0.1 以下になる確率を考えます。その確率が 95 % 以上であるようにするためには少なくとも何回投げれば良いでしょうか。ただし、正規分布で近似するときに $\frac{1}{2}$ の補正はしなくて結構です。

基本演習 11.4 [教科書 練習問題 15-5] 袋の中に赤玉 3 個と白玉 7 個が入っています。その中から復元抽出法で 1 個ずつ玉を繰り返し取り出すとき赤玉の出た回数を X で表すことにします。

- (1) 100 回取り出すとき $P[X < 27 \text{ 又は } 33 < X]$ を求めて下さい。
- (2) 1000 回取り出すとき $P[X < 270 \text{ 又は } 330 < X]$ を求めて下さい。

●母比率の区間推定

以下の問題においては、特記がなければ半整数補正は省略してください。

基本演習 11.5 ある選挙において出口調査を実施したところ、1000 人中に A 候補に投票したと言う人は 400 人、B 候補に投票したと言う人は 300 人いました。A 候補の得票率 p_A と B 候補の得票率 p_B それぞれの信頼度 99 %の信頼区間を求めてください。そしてその結果 A 候補は当選確実であると言えるでしょうか。

基本演習 11.6 あるお笑い芸人の認知度を調べるため街頭アンケートを行ったところ、通行人 100 人のうち 30 人がこの芸能人のことを知っていました。信頼度 95 %でこの芸人の認知度の信頼区間を求めて下さい。

基本演習 11.7 サイコロを 400 回投げたところ、6 の目が 75 回出ました。このサイコロで 6 の目が出る確率の 95 %信頼区間を求めて下さい。

基本演習 11.8 T 社は新車のデザインについて A~C の 3 つの案を持っています。社内の意見では A 案が有力ですが、メインターゲット層の反応を調査してみることにになり、ターゲットとなる年代の男性 50 人を無作為に抽出し、好みのデザインを 1 つだけ選んでもらったところ、次の結果を得ました：

A	B	C
32 人	11 人	7 人

このデータから、メインターゲット層のデザイン A への支持率 p を信頼度 90 %で区間推定して下さい。信頼区間の左端が 0.5 を超えるようであれば、T 社はデザイン A を採用する方針です。

基本演習 11.9 2020 年 4 月 21 日から 28 日にかけて、ある医療機関が都内の成人 147 人に新型コロナウイルス抗体検査を実施したところ、7 名が抗体ありとの結果でした。都内の成人全体での抗体保有率 p の 95 %信頼区間を求めてください。母集団の分散をサンプル値を使って既知としたうえで正規分布によって近似し、半整数補正も行なって下さい。

基本演習 11.10 厚生労働省は、2020 年 6 月 1 日~6 月 7 日にかけて東京都・大阪府・宮城県において無作為抽出された一般住民を対象に、新型コロナウイルス抗体検査を実施しました。東京都では観測数 1971 名に対して抗体保有者が 2 名となりました。東京都における抗体保有率の 95 %信頼区間を求めて下さい。ただし母集団の分散をサンプル値を使って既知としたうえで正規分布によって近似し、半整数補正も行なって下さい。

基本演習 11.11 広島県・広島大学は 2021 年 12 月から 2022 年 1 月にかけて無作為抽出された一般市民に対して新型コロナウイルス抗体検査を実施し、1942 人の参加者中、抗体保有者は 1789 人でした。広島県における抗体保有率の 95 %信頼区間を求めて下さい。ただし母集団の分散をサンプル値を使って既知としたうえで正規分布によって近似し、半整数補正も行なって下さい。