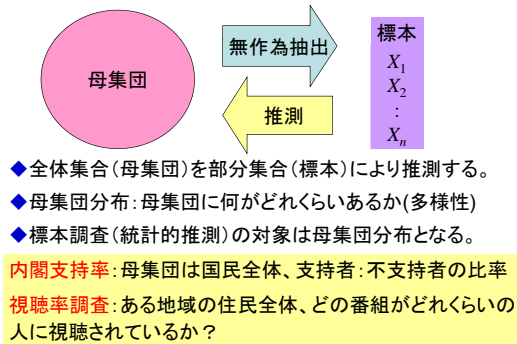


世論調査の数学 (統計的推測入門)

熊本大学工学部
数理工学科
岩佐 学

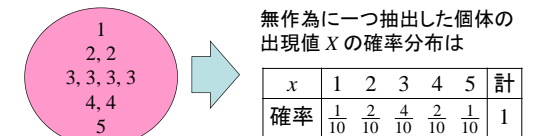
標本調査 — 母集団と標本 —



無作為抽出

- ◆ 標本抽出は無作為に行うことが必要
- ◆ 無作為とは、母集団に属する個体を等確率で選択することを意味する(特定の個体が選ばれやすくなつてはならない)

街頭インタビューは無作為か?



母集団分布(多様性) = 標本 X の確率分布(ランダム性)
無作為抽出により母集団の情報が標本に忠実にコピーされる

抽出の反復の独立性

毎回の抽出の結果に関連性がない → 独立
 $P(X=i, Y=j) = P(X=i)P(Y=j)$
 復元抽出は独立であるが、非復元抽出は独立でない。
 独立性が成り立てば、確率計算が簡単になる。

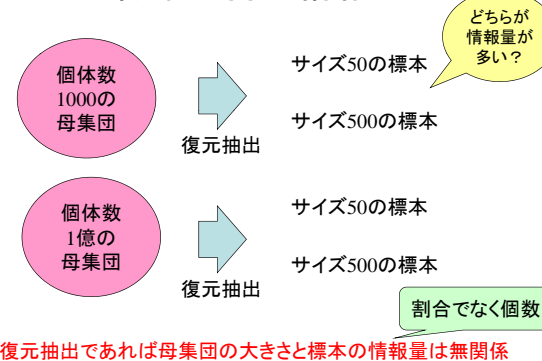
母集団: 1 2 2 2回抽出 標本: X, Y

復元抽出	$X \setminus Y$	1	2	計
	1	1/9	2/9	1/3
	2	2/9	4/9	2/3
	計	1/3	2/3	1

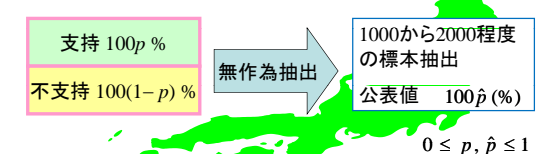
非復元抽出

$X \setminus Y$	1	2	計
1	0	2/6	1/3
2	2/6	2/6	2/3
計	1/3	2/3	1

標本の持つ情報量



世論調査(内閣支持率)



- ◆ 標本における支持率 \hat{p} を発表 (真の支持率 p との誤差を認識する必要性)
- ◆ 調査主体による支持率の差 (調査法, 抽出法の差)
- ◆ 標準誤差 $\pm 200 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ (%) 1万人の調査で±1%

視聴率調査

- ◆ 実施主体: ビデオリサーチ社 (<http://www.video.co.jp/top.htm>)
- ◆ 母集団は各地方(関東、関西、名古屋、北部九州、長崎など27地域)
- ◆ 世帯単位の標本抽出(600 or 200)
- ◆ 世帯視聴率と個人視聴率
- ◆ 機械による調査(2種類)とアンケートによる調査
- ◆ 調査世帯のローテーション(2年 or 3年)
毎週 $600/24 = 25$ 世帯ずつ調査世帯が入れ替わる

サンプリング

- ◆ 地域毎に100~600の世帯を無作為抽出(系統抽出法)
- ◆ 関東地方 15万世帯 から 600世帯 を抽出とすると
 - 15万世帯に1から150,000までの番号をつける。
 - $150,000 \div 600 = 25,000$
 - 1 から 25,000 まで整数 x を一つ無作為に選ぶ。
 - $x + 25,000n$ ($n=0, \dots, 599$) により 600 の数値を決定。

乱数は一つだけ。簡便。
 ナンバリングにより地域から万遍なく抽出可能。
 非復元抽出
 標準誤差 $\hat{p} = 0.2$ のとき $2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 2\sqrt{\frac{4}{600 \times 5^2}} = \frac{2}{25\sqrt{6}} = 0.033$

理論と経験

