コインなげと中心極限定理

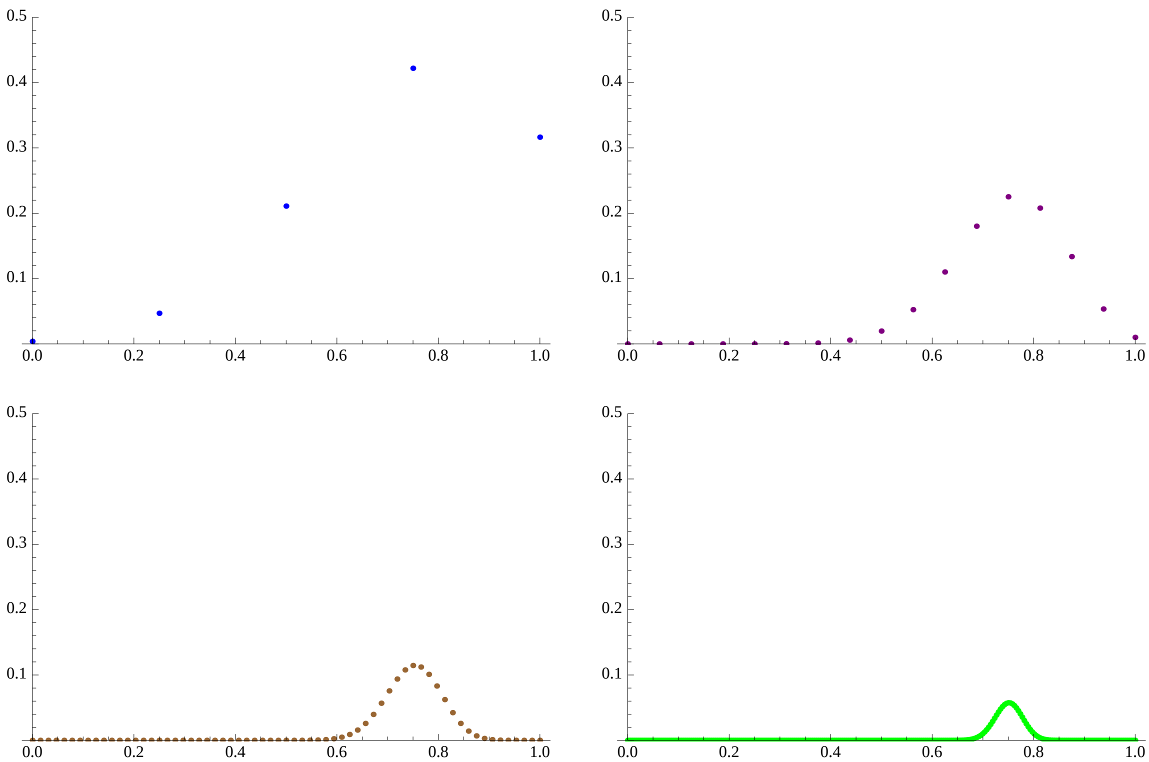
2010 年度前期「数学 I」補足資料（原　隆, 2010.06.30）

以下は，6/30 の講義で紹介した，コイン投げの結果と中心極限定理の関係を示す資料である．講義で説明したこともあって，全般的に最小限の説明しか書いていない．
考える問題は以下の通り：

- コインを N 回投げる．表の出る確率 p とする．
- N 回中，表の出た回数を n として，表の出た割合 n/N と，その「N 回中 n 回表の出る確率」$= P_{n,N}$ を考える．
- X_i を,「i 回めに表なら 1，裏なら 0」となる確率変数として定義する．これまで何度かやった通り，$n = \sum_{i=1}^{N} X_i$ である．
- また，X_i は i.i.d. で，その期待値は $\mu = p$，標準偏差は $\sigma = \sqrt{p(1-p)}$ であることに注意しておく．

1．$p = 0.75$ の場合

$p = 0.75$ の時を，横軸は n/N，縦軸は $P_{n,N}$ としてプロットした．$N = 4, 16, 64, 256$ をそれぞれ描いてある．



読み取れること：

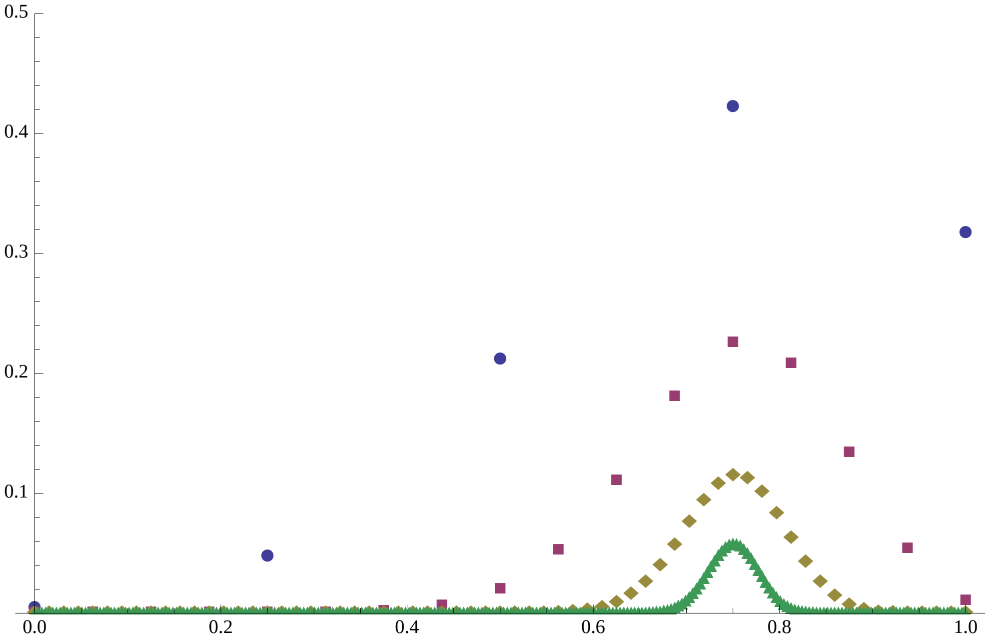
- N が増えるにつれて，点の数は多くなる（$N+1$ 通りの n があり得るから，アタリマエ）．
- 各点 n/N での確率の値 $P_{n,N}$ は小さくなる.（全確率が 1 だから，点の数が多くなれば，各点での値は小さくなるのは自然.）
- そうではあるが，N を増やして行くと，$\mu = p = 0.75$ 付近に集まっては来ているようだ.

以下では，$\mu = 0.75$ の周りにどのように集まっているのか，その規則性を見つけて，中心極限定理との関連をつけよう.

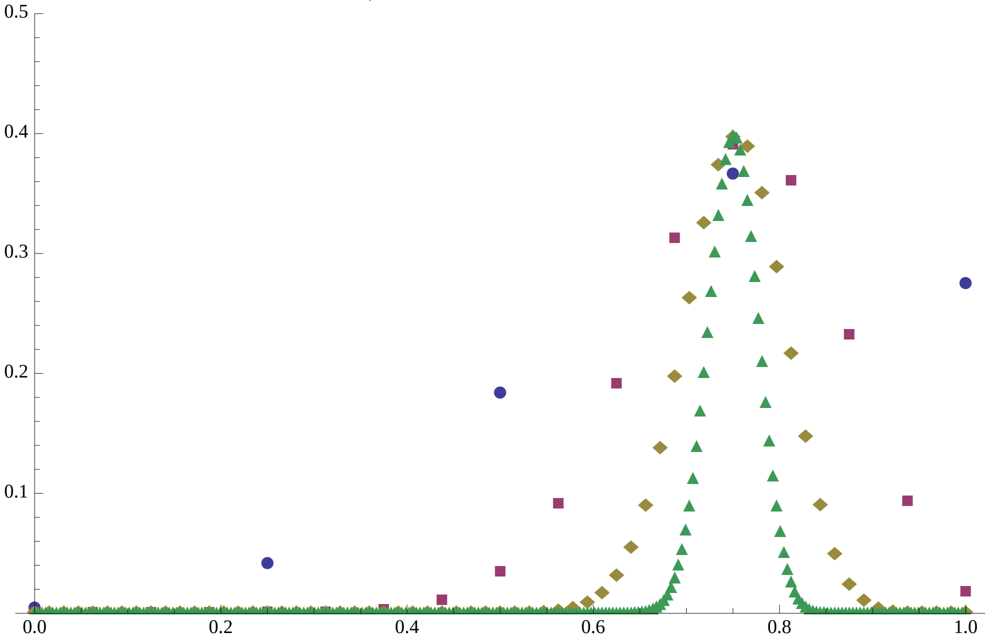
$p = 0.75$ **の場合の** $N = 4, 16, 64, 256$

眺めていても仕方ないので，$N = 4, 16, 64, 256$ のグラフを工夫しながら重ねて描いてみる．青丸が $N = 4$，紫の四角が $N = 16$，黄土色のダイヤモンド型が $N = 64$，緑の三角が $N = 256$ である．(以下，この色と記号の使い方は同じものを使い続ける.)

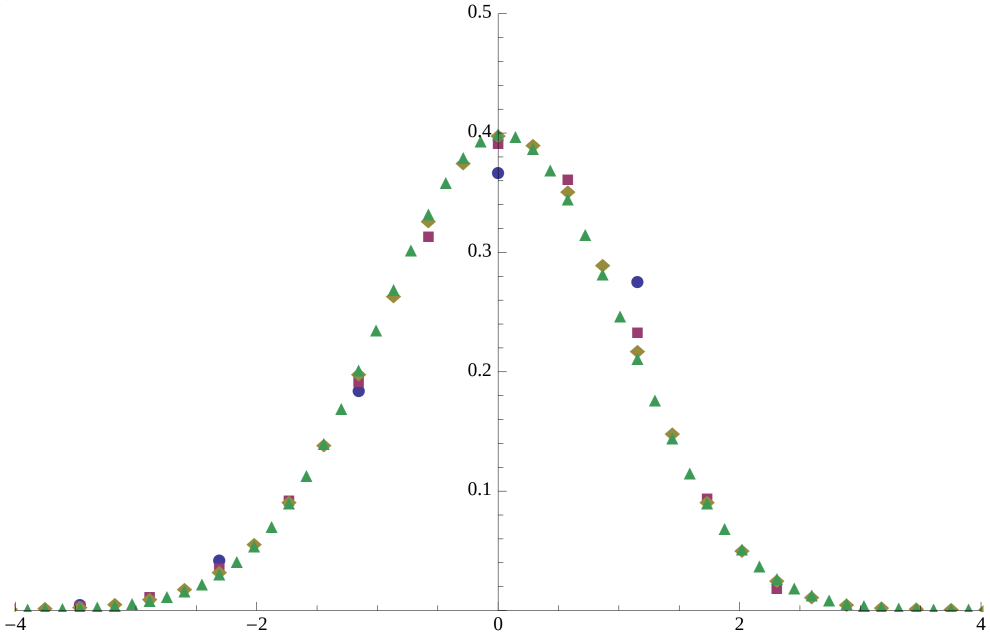
(1) まずは横軸を n/N，縦軸は $P_{n,N}$ とする —— 前ページと全く同じだが，$N = 4, 16, 64, 256$ を重ねてみた．0.75 付近に集中している様子がよりよくわかる．



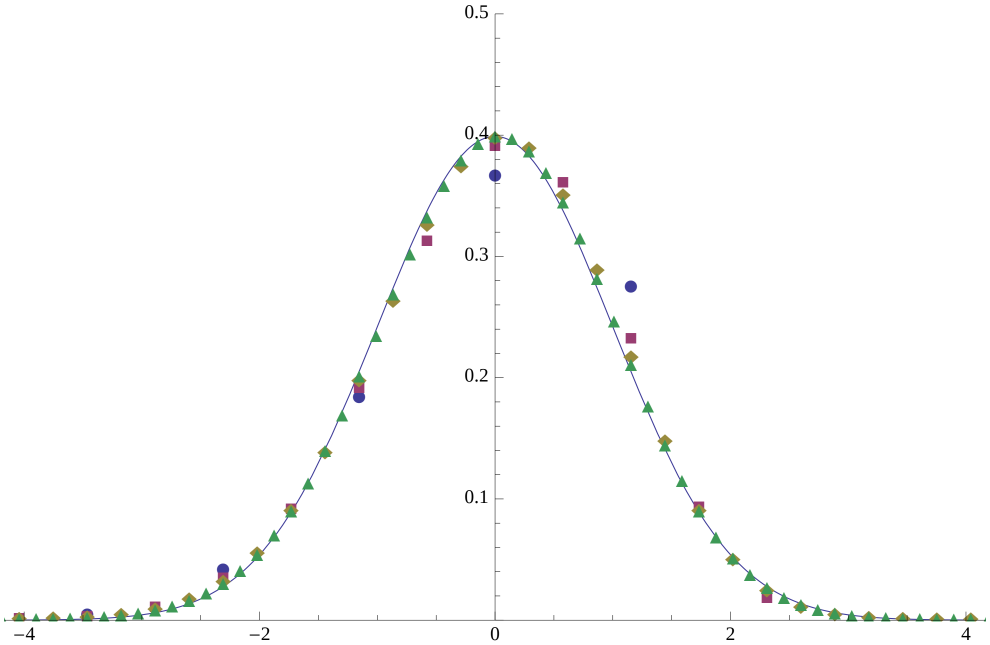
(2) でも，(1) では高さも幅も小さくなって行くので，そこを補正することを考えよう．まずは高さの補正のために，縦軸を引き延ばす —— 縦軸を $P_{n,N} \times \sigma\sqrt{N}$ とする．結果として，4 つの場合が同じような高さになった．



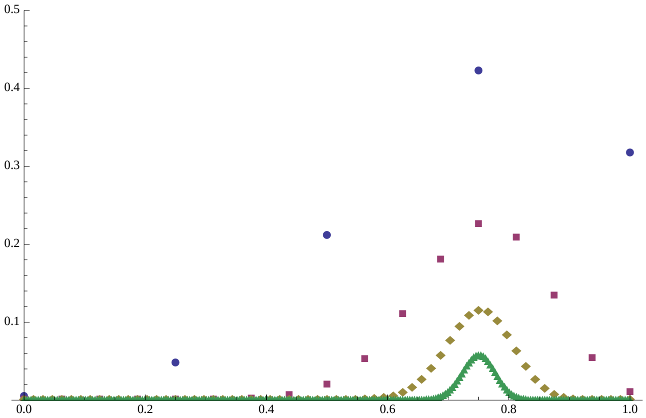
(3) でも，(2) では幅が小さくなり過ぎなので，今度は横軸も引き延ばすことを考える．ただし，単純に引き延ばすと，分布の中心 0.75 が右の方にとんで行ってしまって困る．そこで，分布の中止を原点に移動し，その上で横軸を引き延ばそう．具体的には，縦軸は上のママ $P_{n,N} \times \sigma\sqrt{N}$ だが，横軸を $\dfrac{\sqrt{N}}{\sigma} \times \left(\dfrac{n}{N} - \mu\right)$ としてみた．結果として，4 つの場合がほとんど重なっている．



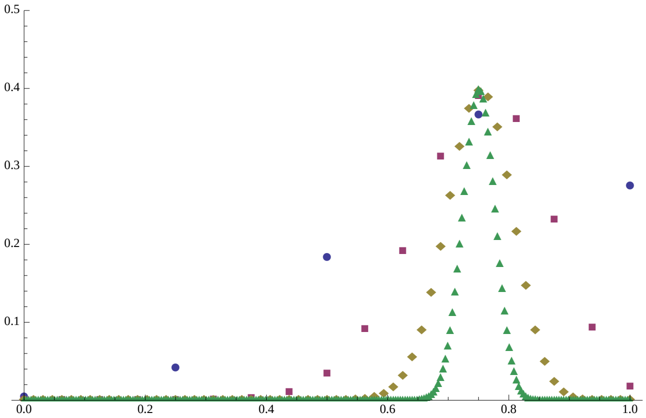
(4) 最後に，上のプロットと正規分布のグラフを重ねてみた．非常に良く重なっている．特に $N = 64, 256$ では正規分布のグラフからのずれがほとんどわからない．



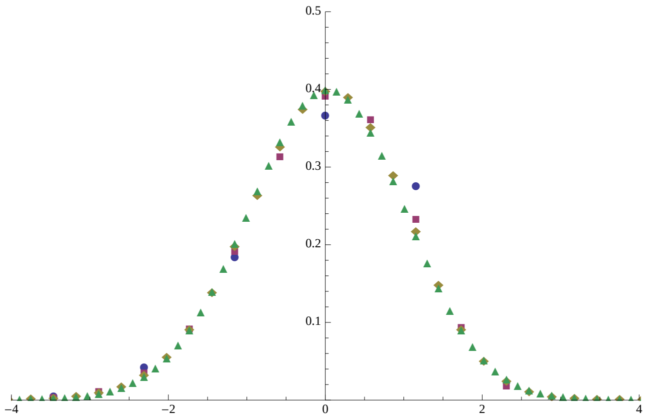
$p = 0.75$ の場合のまとめ．ちゃんとスケールすると，正規分布のグラフに重なって行くことがわかる．



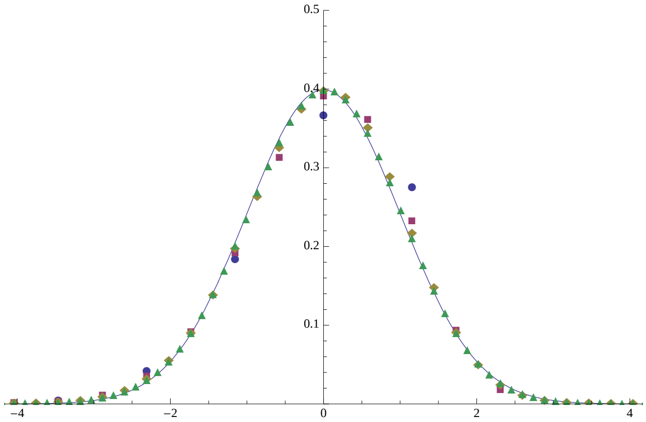
4 つを単純に重ねたもの：



縦軸を引き延ばしたもの:



横軸も引き延ばしたもの：

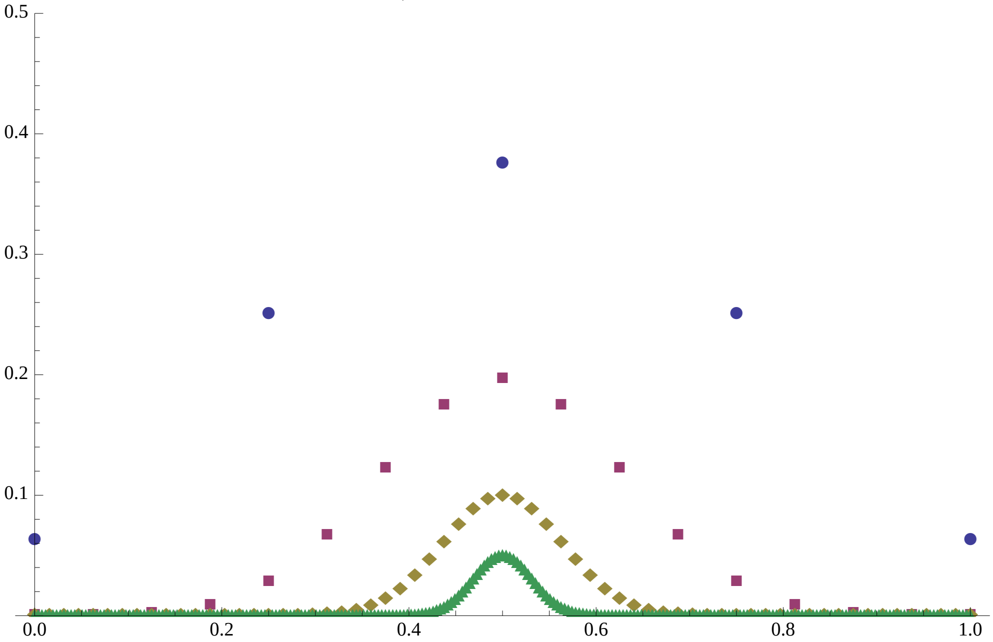


正規分布のグラフを重ねたもの：

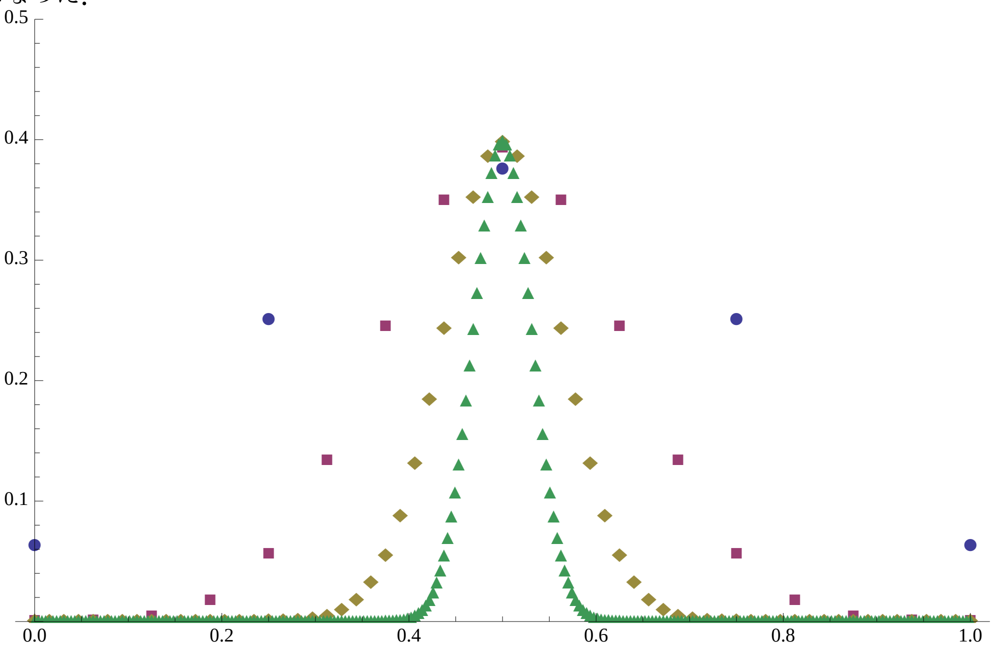
$p = 0.5$ **の場合の** $N = 4, 16, 64, 256$

$p = 0.75$ だけでは説得力がないので，$p = 0.5$ の場合の n/N vs. $P_{n,N}$ のプロットも同様にやってみた．

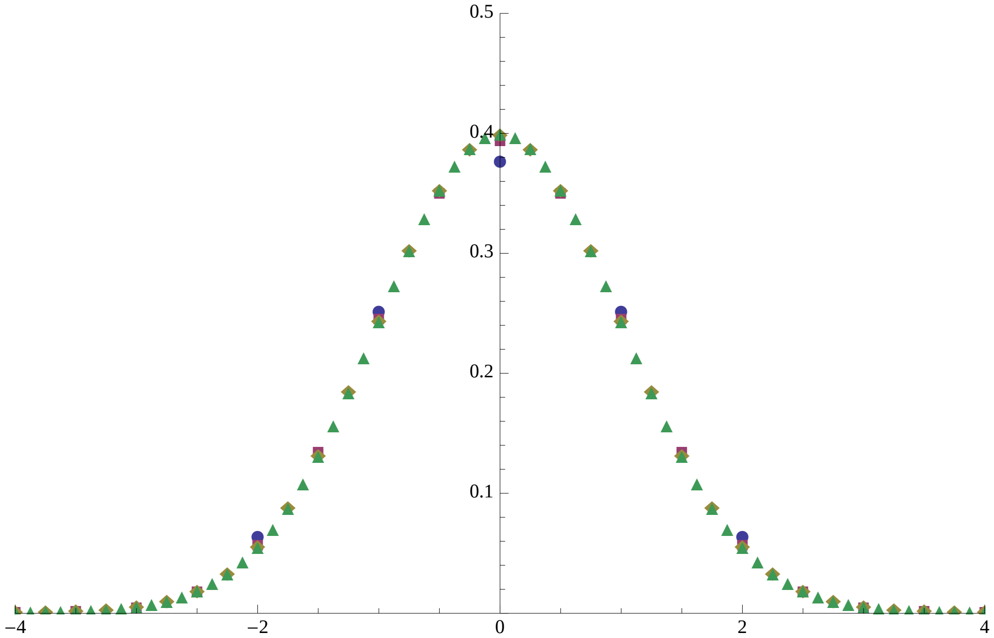
(1) まずは横軸は n/N，縦軸は $P_{n,N}$ —— 何の手も加えていない．



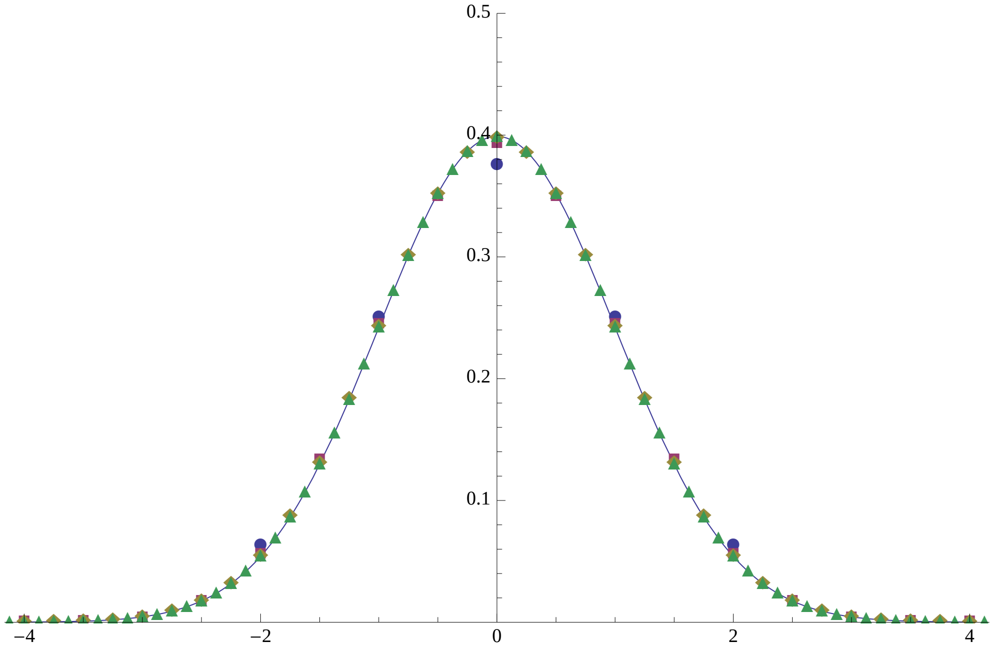
(2) 次に，縦軸を $P_{n,N} \times \sigma\sqrt{N}$ としてみた —— 縦軸を引き延ばした．結果として，4 つの場合が同じような高さになった．



(3) 縦軸は上のママ $P_{n,N} \times \sigma\sqrt{N}$，更に横軸を $\dfrac{\sqrt{N}}{\sigma} \times \left(\dfrac{n}{N} - \mu\right)$ としてみた — 横軸の中心を $p = 0.5$ に持って来た上で，更に引き延ばした．結果として，4 つの場合がほとんど重なっている．

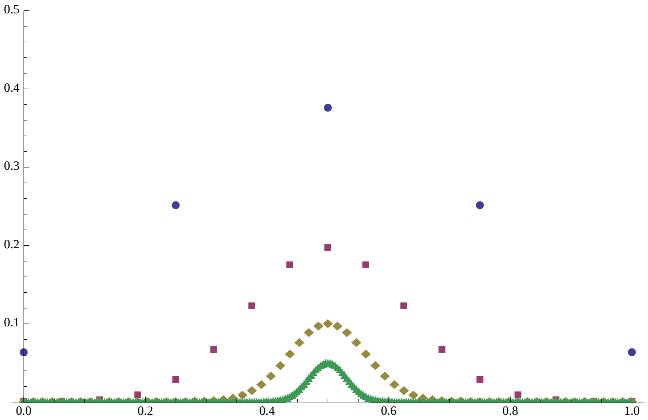


(4) 最後に，上のプロットとガウス分布のグラフを重ねてみた．非常に良く重なっている（$n = 4$ でも！）

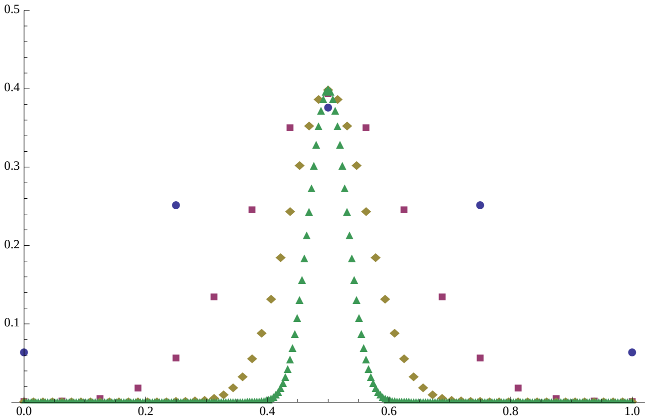


$p = 0.5$ の方が，$p = 0.75$ よりも収束が良さそうだ．

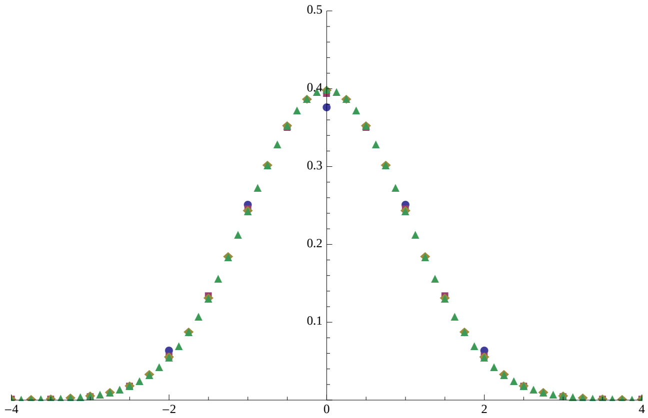
$p = 0.5$ の場合のまとめ



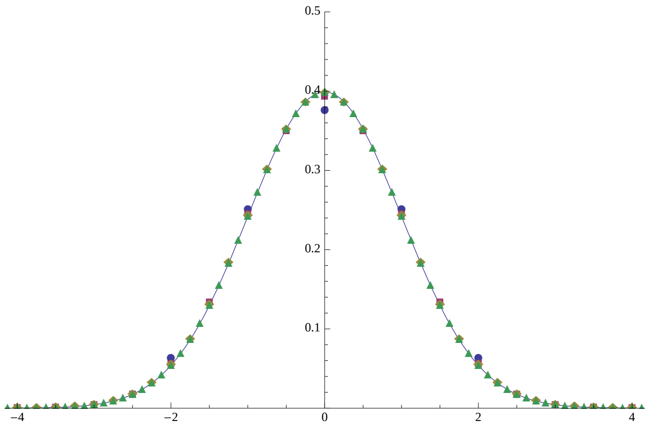
0.5
0.4
0.3
0.2
0.1
0.0
0.2
0.4
0.6
0.8
1.0



0.5
0.4
0.3
0.2
0.1
0.0
0.2
0.4
0.6
0.8
1.0



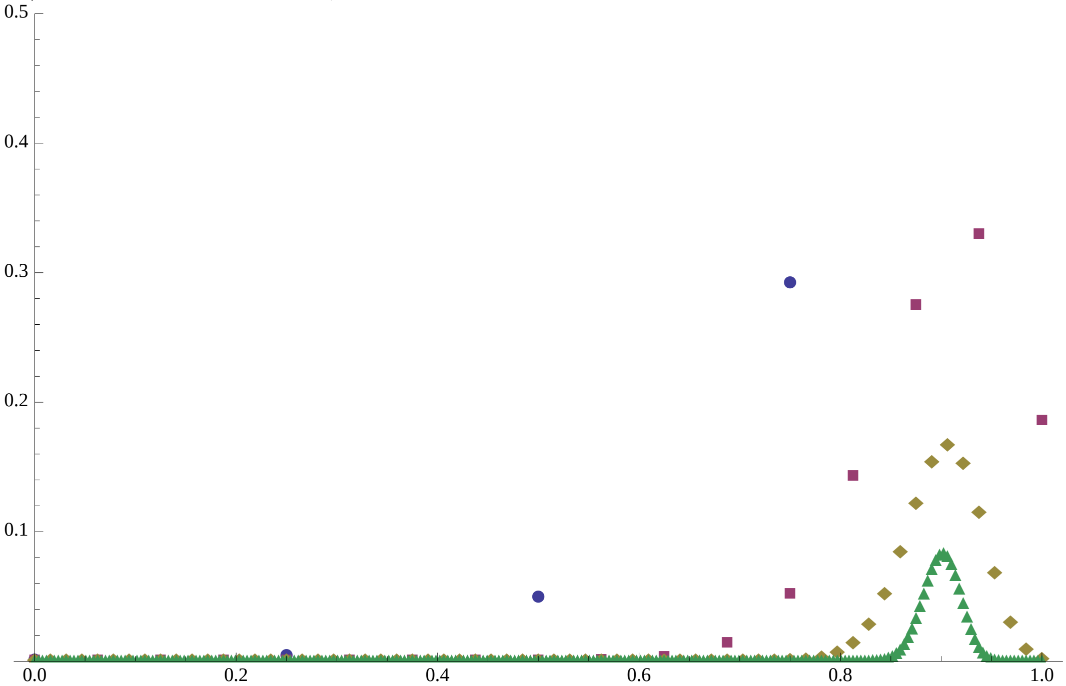
0.5
0.4
0.3
0.2
0.1
−4
−2
0
2
4

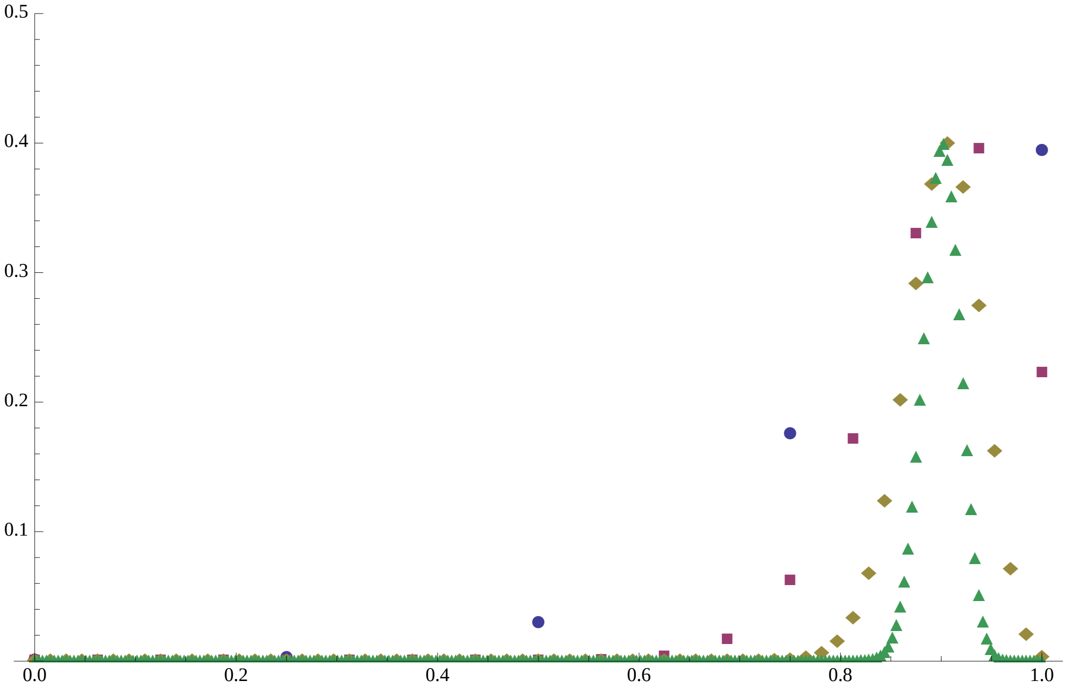


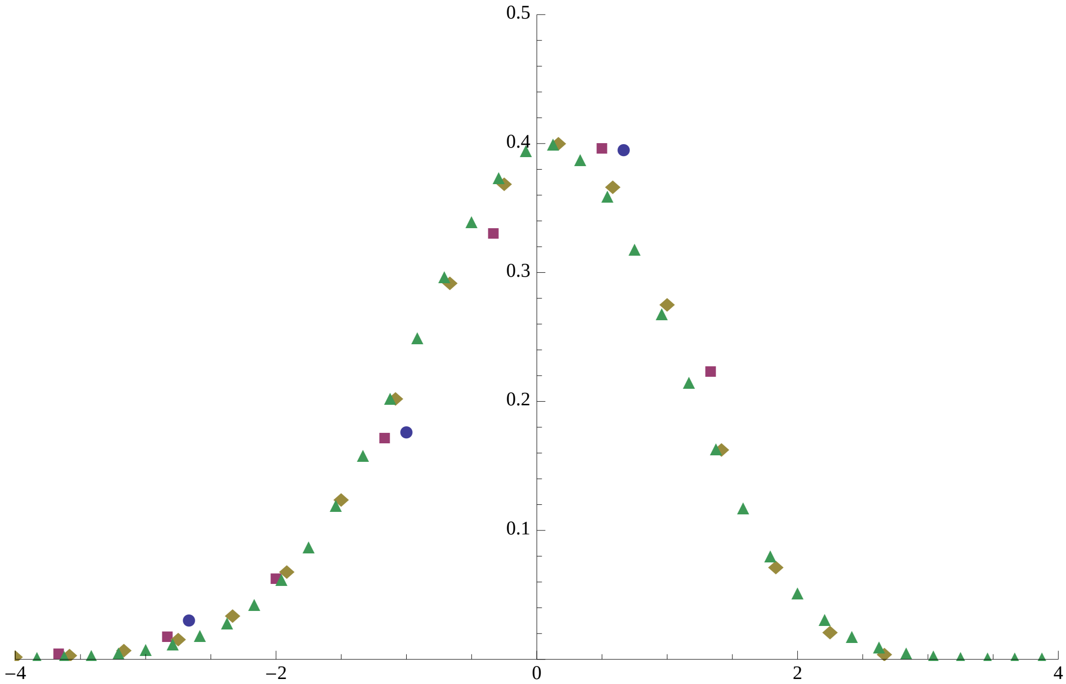
0.5
0.4
0.3
0.2
0.1
−4
−2
0
2
4

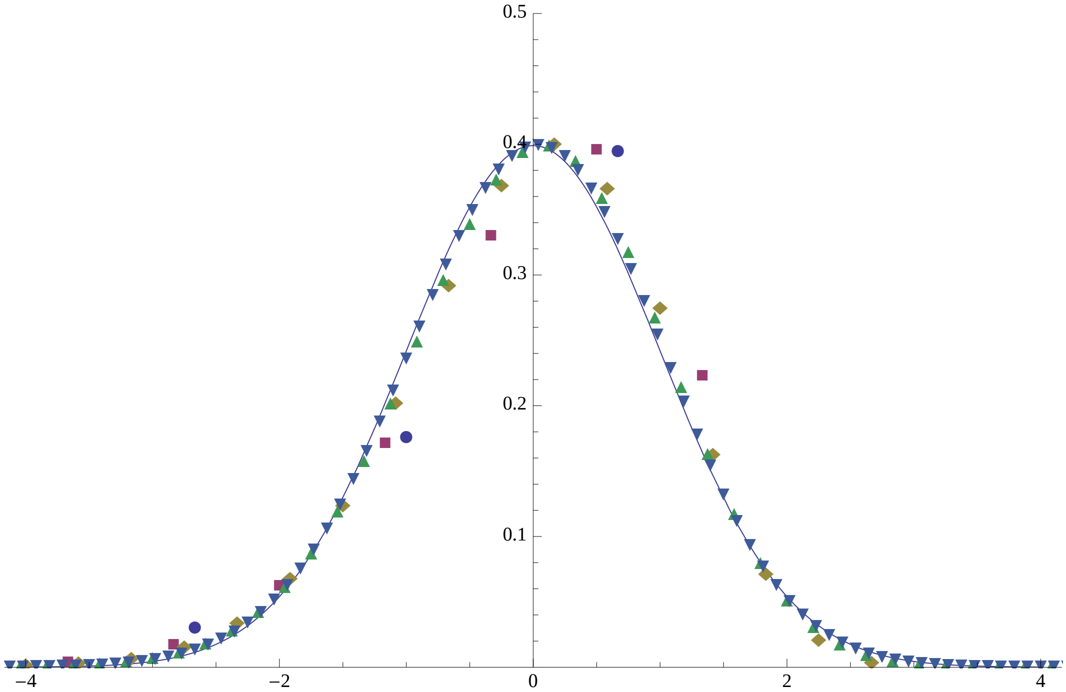
$p = 0.9$ **の場合の** $N = 4, 16, 64, 256$

今度は流石につらそうであるが，でも，N が大きくなると正規分布に重なって行くことが見て取れる．最後の図のみ，$N = 1024$ も描いてある．



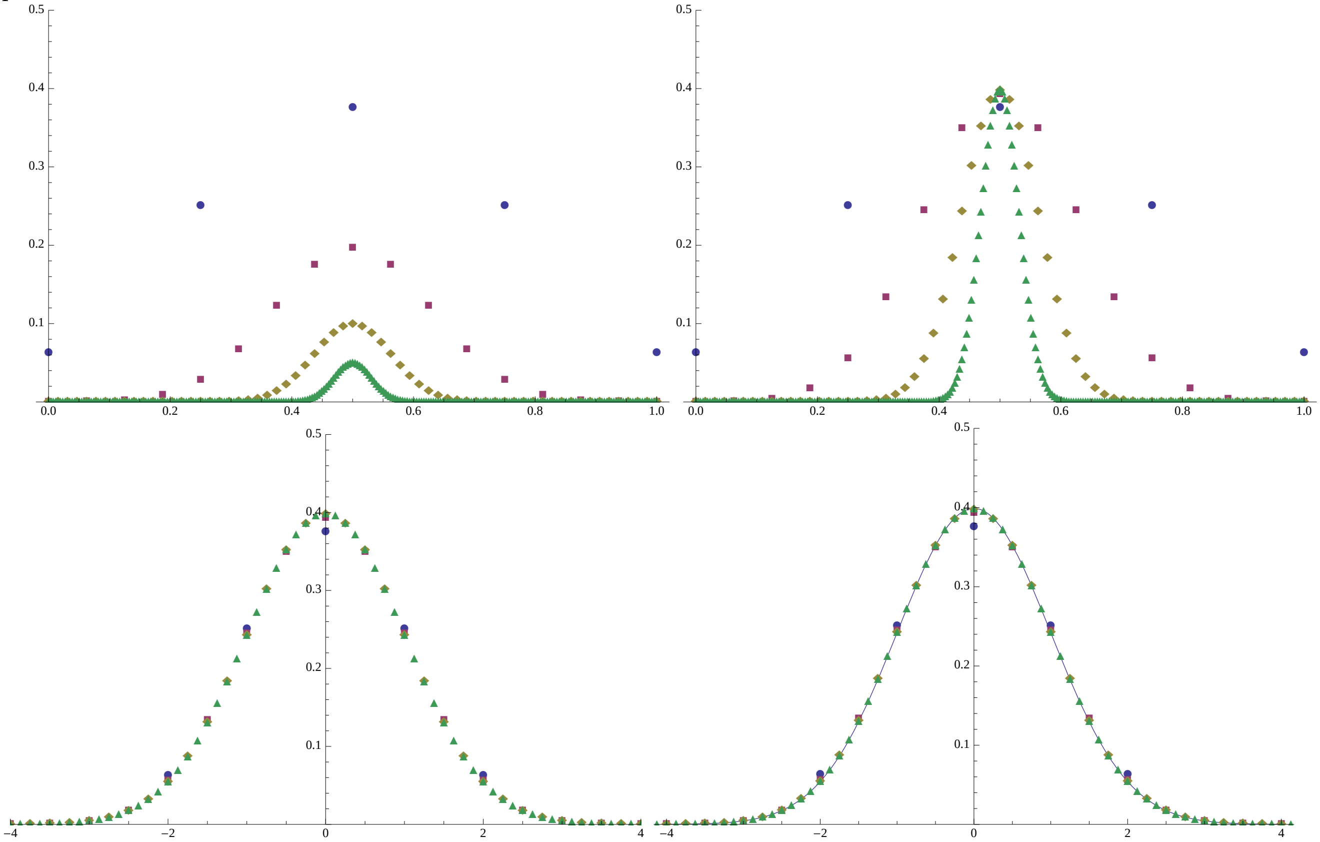



0.5
0.4
0.3
0.2
0.1
−4
−2
0
2
4

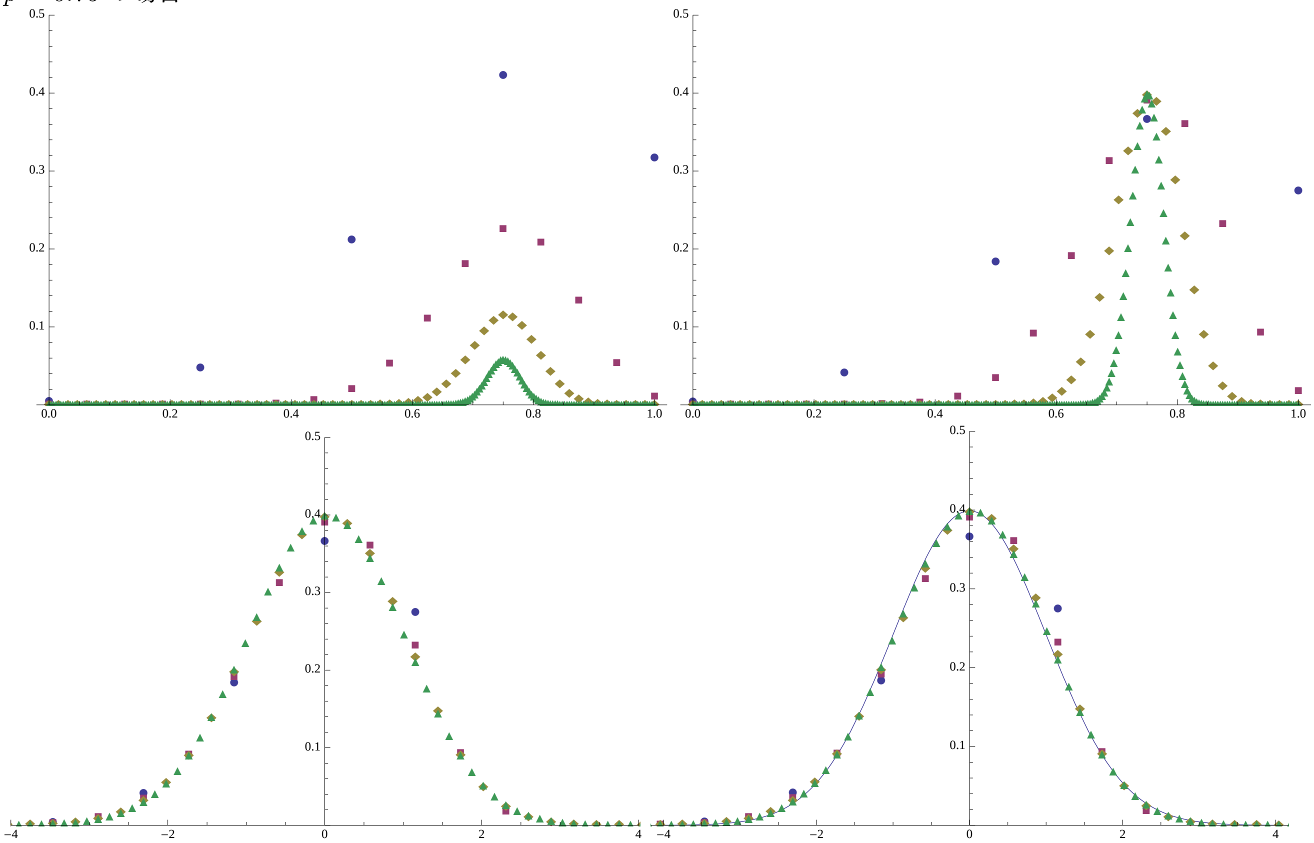

0.5
0.4
0.3
0.2
0.1
−4
−2
0
2
4

もう一度，$p-0.5, 0.75, 0.9$ をまとめておく．

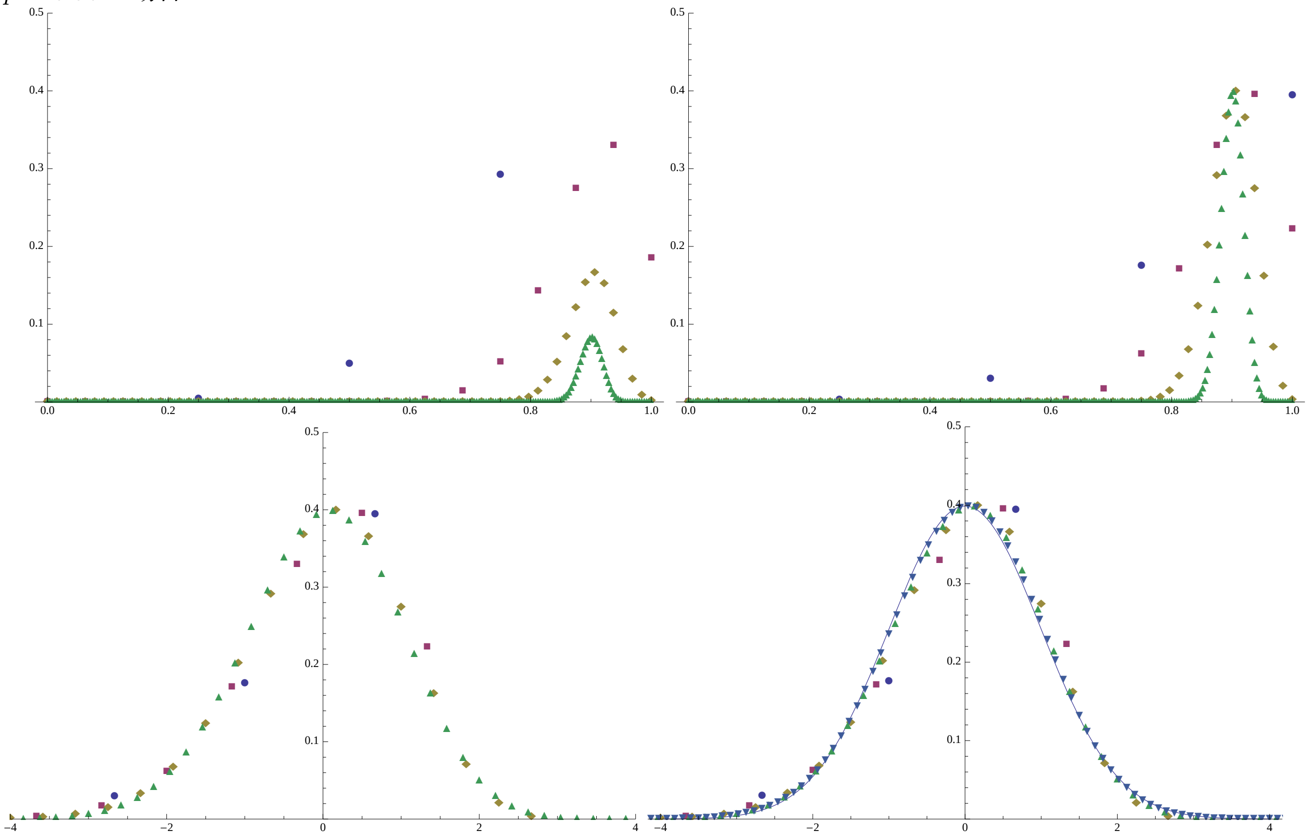
$p=0.5$ の場合



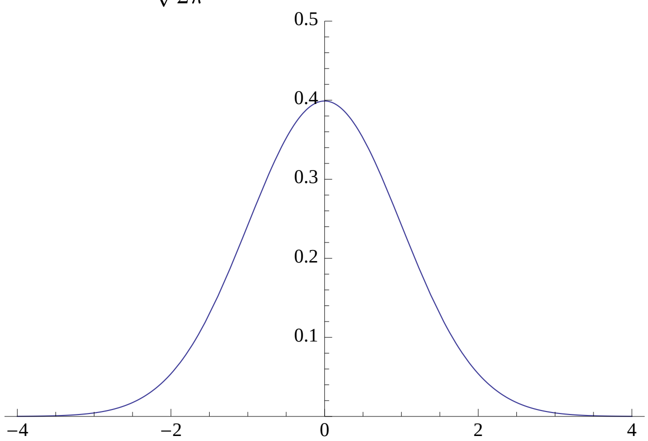
$p=0.75$ の場合



$p = 0.90$ の場合



正規分布 $y = \dfrac{e^{-x^2/2}}{\sqrt{2\pi}}$ のグラフは以下のようである．



通常

$$\Phi(a) := \int_{-\infty}^{a} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx$$

と書く．以下に $1 - \Phi(a) = \displaystyle\int_{a}^{\infty} \frac{e^{x^2/2}}{\sqrt{2\pi}} dx$ のいくつかの値を載せておく：

a	0	1	1.645	1.960	2	2.326	2.576	3	4
$1-\Phi(a)$	$\frac{1}{2}$	0.1587	$\frac{1}{20}$	$\frac{1}{40}$	0.02275	$\frac{1}{100}$	$\frac{1}{200}$	1.350×10^{-3}	3.167×10^{-5}