

* 2015 第11回統計教育の方法論ワークショップ

大学受験生を対象とした統計教育実践例

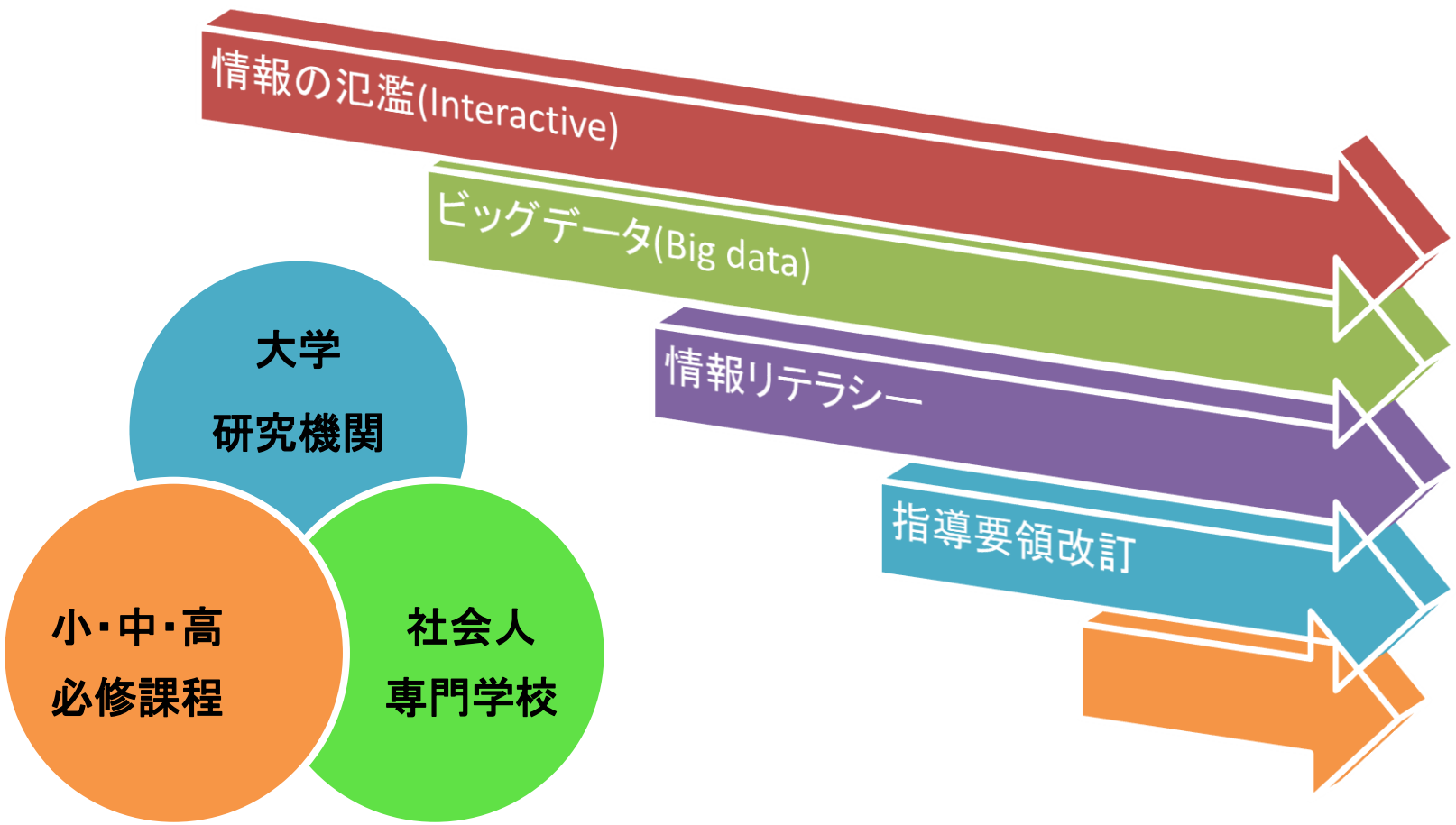
北口 正敏† 吉田 良之‡

† 大阪大学医学部 〒560-0043 大阪府豊中市待兼山町 1-17

‡ ヨシダゼミナール 〒543-0002 大阪府大阪市天王寺区上汐 4-5-21

E-mail † kitagucci@vision.hss.osaka-u.ac.jp, ‡ yoshizemil@earth.ocn.ne.jp

1,Introduction



2,題材-Material-

2.1 授業で扱った分野と背景

2.1.1 最適停止

事前情報のない複数の相手と1人ずつ順々に「お見合い」していくときに、「何人目の見合い相手までを見送り、何人目から真剣に結婚相手を探し始めるべきか」

ans.)

・最適ポリシー (stopping rule) の導出

最初の $r-1$ 人をスキップし、 r 人目が候補者 (すなわち、それまで面接した中で最もよい相手) なら婚約する。任意の r について最善のを選択する確率は次の通りである。

$$P(r) = \sum_{j=r}^n \left(\frac{1}{n}\right) \left(\frac{r-1}{j-1}\right) = \left(\frac{r-1}{n}\right) \sum_{j=r}^n \left(\frac{1}{j-1}\right)$$

where; $n \rightarrow \infty, \quad \frac{r}{n} \rightarrow x, \quad \frac{j}{n} = t, \quad \frac{1}{n} = dt$

$$P(r) = x \int_x^1 \frac{1}{t} dt = -x \log(x)$$

$P(r)$ の x についての導関数を 0 とすることで

$$P(r) = \frac{1}{e}$$

≈ 0.368

$$P(r) = x \int_x^1 \frac{1}{t} dt = -x \log(x)$$

そこで、 $P(S)$ の近似値 $P(X)$ は $P(x) = -x \log(x)$
これを x で微分して0と置くことで、最適な $x^*=S^*/N$ を求められる。
 $P'(x) = -\log(x) - 1 = 0$ となるので、
 $S^*/N = 1/e$ e はネイピア数

2.1.2 期待値

“宝くじ” から始まり、身近ではあるけれどあまり知られていない天候デリバティブを取り上げ、通常为天候保険との比較を紹介した。また、損害保険会社などが保有する株式ヘッジの一例であるソルベンシーマージンについても紹介した。

2 天候デリバティブ

・条件設定

- 1) 持兼山スーパー...8月「猛暑がんばろうキャンペーン」
- 2) マイナス要件＝気温が 24°C を下回ると1,000千円の損害が発生！
- 3) 天候デリバティブの活用で気温リスクをヘッジしたい

問) 次の条件の天候デリバティブを金融機関にと合わせた。

・天候デリバティブの条件

→8月期間中、日次平均気温が 24° を下回った日数に対して、1日あたり1,000円の保証がなされる

1)天候保険と天候デリバティブの違いについて下表(table.2-①)に示す。 table.2-①

No	項目	天候保険	天候デリバティブ
1	事故査定・損害証明	・必要→手続きが煩雑 ・契約者に支払われる金額は査定に基づく損害額が上限 ・加入者が気象変化と収益減少の因果関係を証明し損害額を提示する	・不要→手続きが比較的簡便 ・あらかじめ約定した条件に合致すれば決済金を受け取ることができる ・気象条件は公的機関により公表される客観的な数値である
2	オーバーヘッジ	・実際に損害が発生していなければ支払われない	・実損害の有無にかかわらず契約条件をみたせば支払われる→決済金を実際の損害額を上回るオーバーヘッジが発生する可能性もある
3	契約・支払・価格	・契約者は企業および一部の団体、個人、個人事業主等である ・上記手続き理由により支払には時間を要する場合があり手続きも煩雑となる ・同一条件の場合、価格は理論上デリバティブのプレミアムより低いと考えられる	・契約は企業および一部の団体のみである ・上記理由により支払は比較的早く、手続きは簡便である ・同一条件の場合、価格は理論上デリバティブのプレミアムのほうが高い

※損害保険とデリバティブという取引形態の違いから、会計処理・契約書類についても両者は異なる。特に、単に会計処理のみならずソルベンシーマージン等にも影響を与えることも考慮する必要がある。

[課題 3] 【気温データについて】

i)観測期間；1965年8月1日～2008年8月31日（うち8月分標本数 1364 日）

ii)基本統計量； table3-①基本統計量

1965/8/1－2008/8/31(8 月)	
平均	28.42617302
標準誤差	0.046359518
中央値（メジアン）	28.6
最頻値（モード）	29.5
標準偏差	1.712166901
分散	2.931515498
尖度	0.710445553
歪度	-0.691714464
範囲	11.7
最小	21.2
最大	32.9
合計	38773.3
標本数	1364

iii)関連数値

table3-②関連数値

総日数	①	②
	日気温 24℃未満日数	①かつ日降雨 0.5mm 以上日数
1364	22	16
発生確率%	1.61%	1.17%

- (1) 条件：8月において、日次平均気温が24℃を下回った日数に対し、1日当たり1,000千円の保証がなされる
題意に適う気温ディスカウントオプション(CTD ; Critical Temperature Day Call)について考える。
いま金融機関のコストを考えないとき、プレミアムはその期待値に等しいからこの CTD プットの取引条件は下表。

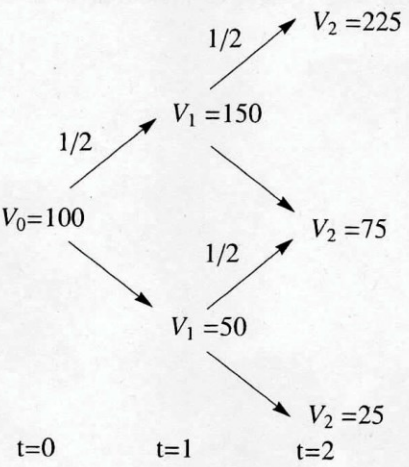
table3-③CTD プットオプションの取引条件

オプションの買い手	持兼山スーパー
観測地点	大阪
観測期間	200X/8/1～200X/8/31(31日)
インデックス(CTD)	観測期間中の日平均気温が24℃を下回った日の日数
行使値	1日
単位支払額	1,000 千円/日
最大支払額	31,000 千円(31 日分)
当初支払プレミアム	499.1 千円(+コスト・利益)

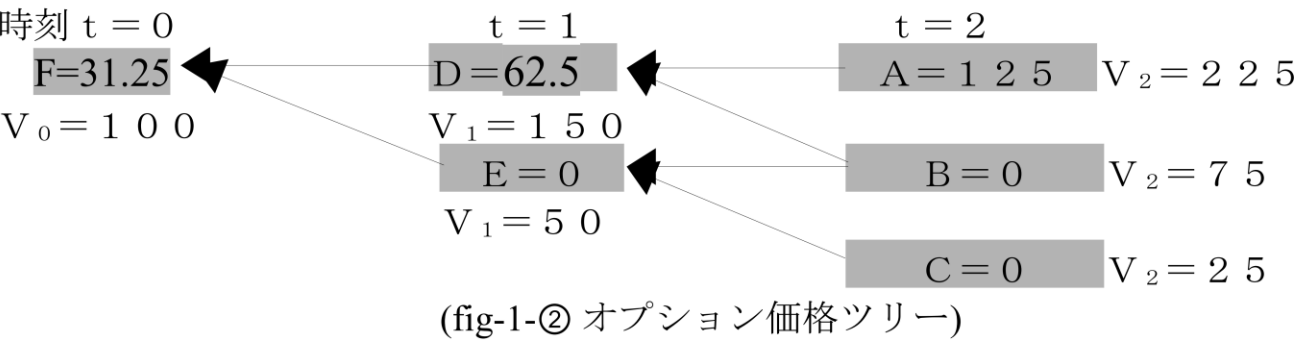
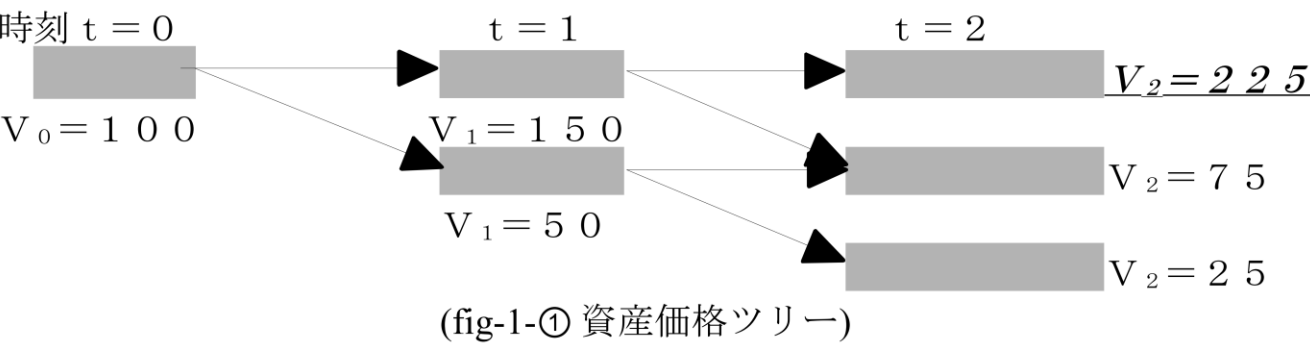
※算出根拠；日次平均気温が24℃以下になる確率は0.0161(22/1364)、観測期間が31日で、インデックスあたりの受取(支払額)が1,000千円であるからオプション価格はこれらを掛け合わせて期待値をおよそ499.1千円と算出。

1: 離散時間ダイナミックプログラミングによるプロジェクト評価

時刻 t において, 固定費用 100 を支払って投資を行うと, 利益 V_t (フローではない) が得られる投資プロジェクトを考える. 企業は, 時刻 $t = 2$ まで投資を延期することができるとし, 割引率は 0 (時間が経っても価値が減少しない) とする. V_t が, 確率 $1/2$ で上昇または下落する以下の 2 項モデルに従うとする.



課題 1 : 離散時間ダイナミックプログラミングによるプロジェクト評価
(Ans.) 題意の資産価格ツリーよりオプション価格ツリー(下図 ; fig-1-①、fig-1-②)を考える。
(A, B, C, D, E, Fは各時点におけるオプション価格)



$$(A) = \max\{0, -100+225\} = 125$$

$$(B) = \max\{0, -100+75\} = 0$$

$$(C) = \max\{0, -100+25\} = 0$$

$$(D) = \max\left\{ \frac{(1/2 \times (A) + 1/2 \times (B))}{1}, -100+150 \right\} = 62.5$$

$$(E) = \max\left\{ \frac{(1/2 \times (B) + 1/2 \times (C))}{1}, -100+50 \right\} = 0$$

$$(F) = \max\left\{ \frac{(1/2 \times (D) + 1/2 \times (E))}{1}, -100+100 \right\} = \underline{\underline{31.25}}$$

以上より、最適投資タイミングは、2回続けて価格上昇した場合（V₂）で、このとき時刻 t = 0 におけるプロジェクト価値 V_(F)=31.25 となる。 ////

2.1.3 条件付確率

有名問題である “モンティホール “は次のような内容であった。

結論：ドアを変えたら、勝率が 2 倍

モンティ・ホールのルール

「プレイヤーの前に 3 つのドアがあって、1 つのドアの後ろには景品の新車が、2 つのドアの後ろには、はずれを意味するヤギがいる。プレイヤーは新車 のドアを当てると新車がもらえる。プレイヤーが 1 つのドアを選択した後、モンティが残りのドアのうちヤギがいるドアを開けてヤギを見せる。ここでプレイヤーは、最初に選んだドアを、残っている開けられていないドアに変更してもよいと言われる。プレイヤーはドアを変更すべきだろうか？」

いま、 $P(B)$ = 事象 B が発生する確率（事前確率, prior probability） $P(B | A)$ = 事象 A が起きた 後 での、事象 B の 確率（事後確率 , posterior probability）とすると、ベイズの定理によれば、 $P(A) > 0$ の条件の下次式が成立する。

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

2.1.4 統計学習を支える数学的基盤

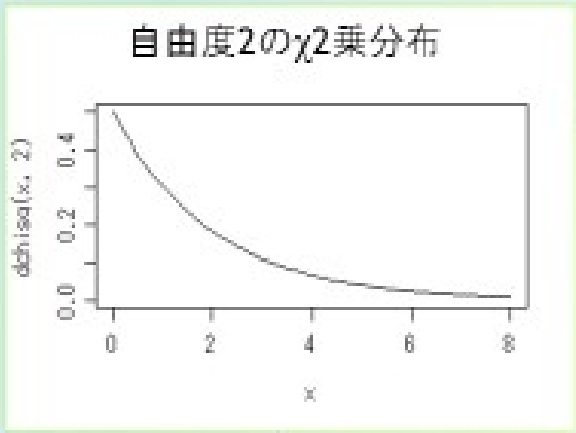
定理 (大数の (強) 法則) $X_1, X_2, \dots$ を互いに独立で同一の分布にしたがう確率変数とする, $E[X_1] = \mu(< \infty)$, $V[X_1] < \infty$ ならば

$$\frac{1}{n} \sum_{i=1}^n X_i \rightarrow \mu \text{ a.s. as } n \rightarrow \infty. \quad (\text{概収束})$$

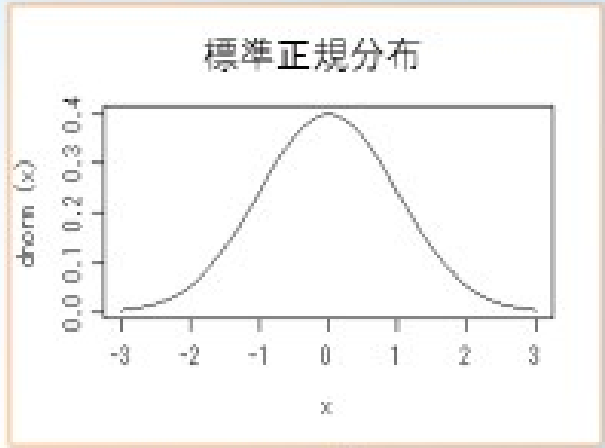
定理 (中心極限定理) $X_1, X_2, \dots$ を互いに独立で同一の分布に従う確率変数とする, $E[X_1] = \mu(< \infty)$, $V[X_1] = \sigma^2(< \infty)$ ならば

$$P \left[\frac{1}{\sigma \sqrt{n}} \left(\sum_{i=1}^n X_i - n\mu \right) \leq x \right] \rightarrow N(x) \quad \text{as } n \rightarrow \infty.$$

データ(サンプル)数と分布の推移 (中心極限定理)



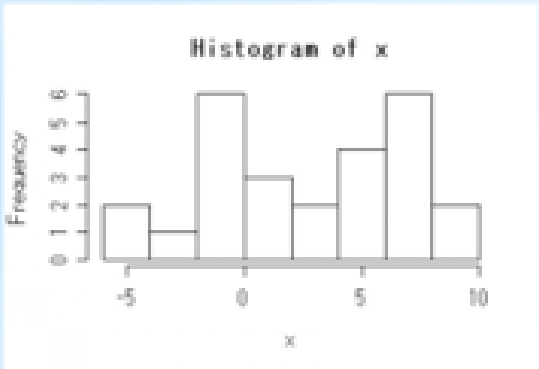
本当？



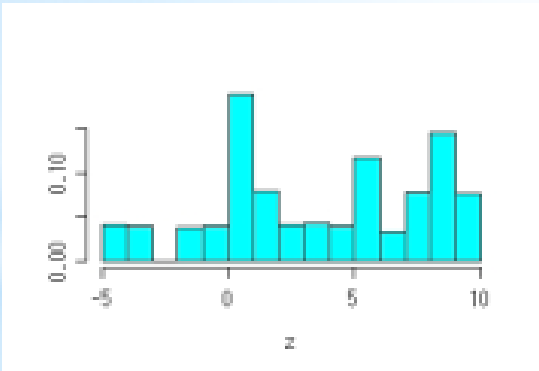
適当なデータからサンプルを繰り返し抽出

```
X <- C(1,1,2,3,4,5,5,5,6,7,7,8,8,8,8,9,9,0,0,0,0,0,-1,-2,-4,-5)
hist(x)
```

抽出したサンプルをヒストグラム化して比較

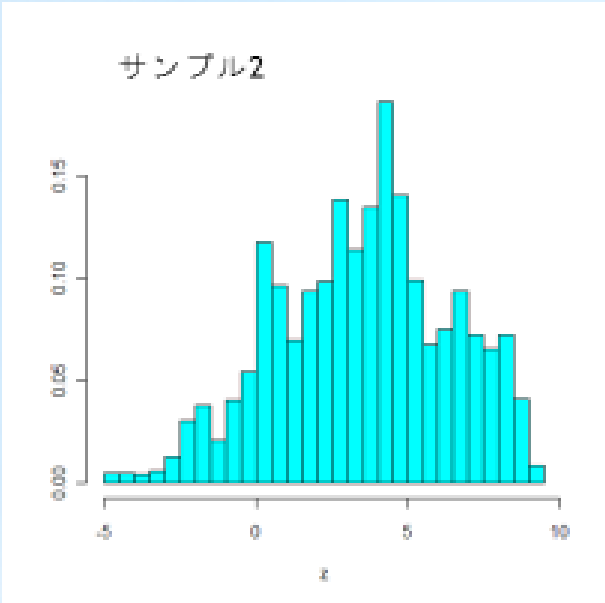


```
> mean(x)
[1] 3.230769
> var(x)
[1] 17.30462
```

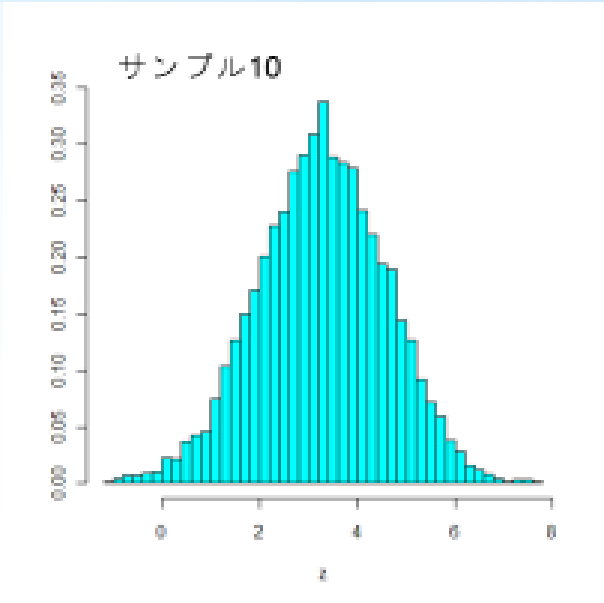


```
z <- numeric(0)
fcl1 <- function(s) {
  for(i in 1:5000){
    y <- sample(x, 「サンプリング
    数s」,replace=T)
    z[i] <- mean(y)
  }
  truehist(z)
}
```

抽出したサンプル数を増やしてヒストグラム化

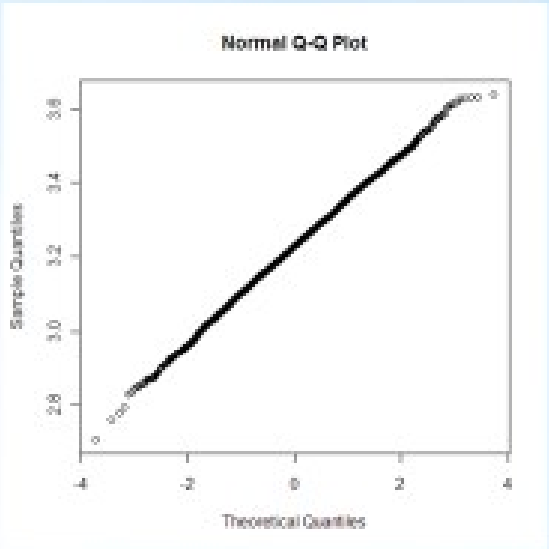
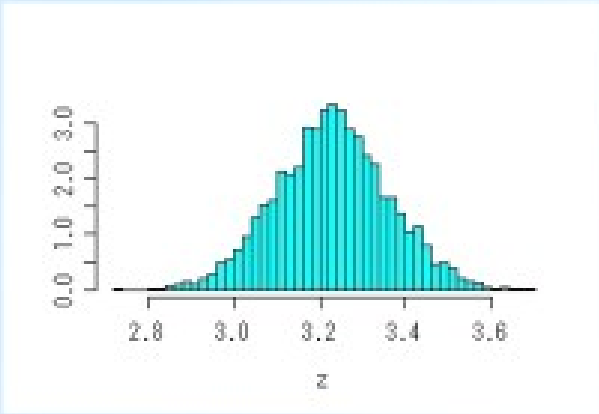


```
> for(i in 1:5000){  
+   y <- sample(x,2,replace=T)  
+   z[i] <- mean(y)  
+ }  
> truehist(z)
```



```
> for(i in 1:5000){  
+   y <- sample(x,10,replace=T)  
+   z[i] <- mean(y)  
+ }  
> truehist(z)
```

サンプル数1000の分布を可視化する



数字で確認する

サンプル1

```
> mean(x)
[1] 3.230769
> var(x)
[1] 17.30462
```



サンプル1000

```
> mean(z)
[1] 3.228538
> var(z)
[1] 0.01700635
```



分散比

```
> var(x)/var(z)
[1] 1017.538
```

受験問題と最適停止の例

- 1) サイコロをふったときの目の出方を得点とする
→1回目を振った後2回目を振るかどうかの判断は？
→さらに2回目の後3回目を振るかどうかの判断は？

- 2) 線形計画の問題

ans.)

(1) サイコロを1回ふったときの得点の期待値は $7/2 = 3.5$ 。(1回目でも2回目でも3回目でも同じ)

よって、1回目にふって出た目の数が1, 2, 3のいずれかのときは2回目をふり、
1回目にふって出た目の数が4, 5, 6のいずれかのときは2回目をふらない…

(A)

これが最も有利な戦略である。

この戦略に出た場合の得点の期待値は $1/2 \times 7/2 + 1/6 \times 4 + 1/6 \times 5 + 1/6 \times 6 = 17/4 = 4.25 \dots$ (B)

何も考えずに2回目をふる場合もしくは無条件に2回目をふらない場合の得点の期待値は $7/2 = 3.5 \dots$ (C)

(2) サイコロをふるチャンスが残り1回となった時点では上記 (A) と同じ。

さて、1回目をふった後に2回目をふるか否かの判断をどうすればいいか？

2回目をふった時点で戦法 (A) を採れば期待値は (B) になるから、1回目に出た目と (B) をくらべればいい。

よって、1回目にふって出た目の数が1, 2, 3, 4なら2回目をふり、5, 6なら2回目をふらない

2回目にふって出た目の数が1, 2, 3なら3回目をふり、4, 5, 6なら2回目をふらない… (D)

これが最も有利な戦略である。

この戦略に出た場合の得点の期待値は $4/6 \times 17/4 + 1/6 \times 5 + 1/6 \times 6 = 14/3 \approx 4.67 \dots$ (E)

実数 x, y が, $3x + y \geq 6, 2x - y \leq 4, x + 2y \leq 7$ を同時にみたすとき, 次の問いに答えよ.

(1) $3x - y$ のとりうる値の最大値, 最小値を求めよ.

(2) $x^2 + y^2$ のとりうる値の最大値, 最小値を求めよ.

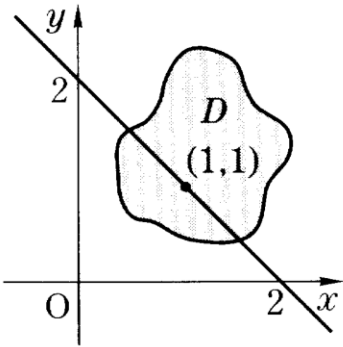
精講

領域 D 内を点 (x, y) が動くとき、 $x+y$ のとりうる値 (いいかえれば最大値, 最小値) はどのように考えればよいのでしょうか. たとえば, $(x, y)=(1, 1)$ としたときの $x+y$ は2ですが, この「2」はどこに表れているかというと, $x+y=2$ だから, **直線の y 切片** になっているのです.

(右図参照)

だから, $x+y=k$ とおいて, この直線が D と共有点をもちながら動くときの y 切片 k のとりうる値の範囲を考えればよいのです.

(右図で, $x+y=k$ は D と共有点をもっています)



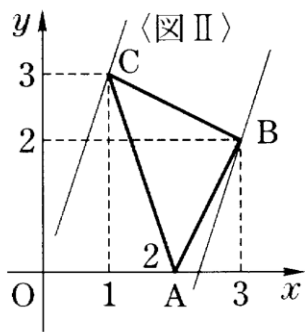
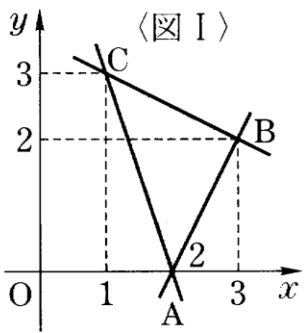
解 答

連立不等式 $\begin{cases} 3x + y \geq 6 \\ 2x - y \leq 4 \\ x + 2y \leq 7 \end{cases}$ の表す領域は

〈図 I〉の色の部分(境界も含む).

注 境界になる3つの直線の交点を先に求めておくと、領域がかきやすくなります.

- (1) $3x - y = k$ とおくと, ◀ ポイント
 $y = 3x - k$ となり, これは, 傾き 3, y 切片 $-k$ の直線を表す.
よって, この直線が, 〈図 I〉の色部分と共有点をもつように動くときの, y 切片のとりうる値の範囲を考えればよい.



4.1 学習内容と範囲について

学習要領(教科書)に記述のある数式等を用いるといった制約のもと、極力身近な題材を使うことで興味を引き、意欲を掻き立てる効果を狙った。

4.2 理解度と取組について

継続期間的な側面と、人数的な面から定量的な評価は行わなかったが、身近な数学の例が将来の学習意欲や、進路に対して前向きな影響を及ぼすとなれば本論の目的は半ば達成されたものと考えている。

4.3 問題点と今後の展望について

たとえば **BS** モデルに代表されるような数学を身に付けることは簡単ではないかもしれないが、学習を継続する切欠になったり、これまでの復習をこなして再挑戦したりすることは遅すぎるということはないと考えている。そして、学生のみならず専門学校や職業訓練などのカリキュラムの一環として組み込まれ、広く学ぶ人が増えるような活動を目指したい。

また、今後は大学生・大学院生を対象とした勉強会などの機会があれば積極的に事例を蓄積したいと考えている。

＊ 2015 第11回統計教育の方法論ワークショップ

大学受験生を対象とした統計教育実践例

北口 正敏† 吉田 良之‡

† 大阪大学医学部 〒560-0043 大阪府豊中市待兼山町 1-17

‡ ヨシダゼミナール 〒543-0002 大阪府大阪市天王寺区上汐 4-5-21

E-mail † kitagucci@vision.hss.osaka-u.ac.jp, ‡ yoshizemil@earth.ocn.ne.jp