

2019年度

S 1

## 総 合

2月25日(月)

情 報 学 部 (行動情報学科)

15 : 20 ~ 17 : 10

【前 期 日 程】

### 注 意 事 項

#### 試験開始前

- 1 監督者の指示があるまで、問題冊子、解答用紙、下書き用紙に手を触れてはいけません。
- 2 監督者の指示に従って、全部の解答用紙(4枚)に受験番号を記入しなさい。

#### 試験開始後

- 3 この問題冊子は、9ページあります。はじめに、問題冊子、解答用紙、下書き用紙(3枚)を確かめ、枚数の不足や、印刷の不鮮明なもの、ページの落丁・乱丁があった場合は、手をあげて監督者に申し出なさい。
- 4 解答はすべて解答用紙に記入しなさい。(下書き用紙と間違わないよう十分注意してください。下書き用紙は採点対象となりません。)
- 5 問題は、声を出して読んではいけません。
- 6 配点は、比率(%)で表示してあります。

#### 試験終了後

- 7 問題冊子と下書き用紙は、必ず持ち帰りなさい。

1 次の各問に答えなさい。(配点 25 %)

いくつかの星が書かれたカードを一度に1枚引くことのできる2種類のくじXとYがある。Aさんは、くじXとYをそれぞれ5回引いた。表1は、Aさんが引いたカードに書かれた星の数を、くじXとYについて記録したものである。

表1：くじXとYを5回引いたときに出的カードに書かれた星の数

くじX	2	2	2	2	4
くじY	1	2	3	3	3

問1. 表1から、Aさんが引いたカードに書かれた星の数の平均値、中央値、および分散を、くじXとYそれぞれに対して求めよ。

問2. 表2は、くじで引いたカードに書かれている星の数によって得られる点数を示している。くじXとYをそれぞれ5回引き表1のような結果が出た場合において、くじを1枚引いて得られる点数の平均値を、くじXとYそれぞれに対して求めよ。

表2：くじで引けるカードの星の数とそれで得られる点数

星の数	1	2	3	4
点数	5	10	50	500

問3. くじXは1回あたり100円、くじYは1回あたり40円で引くことができ、Aさんの手元に200円があるとす。ここで、Aさんは表1の結果から、くじXおよびYでそれぞれの星の数が書かれたカードの出る確率について、「くじXでは星が4つ書かれたカードは5回引いたうち1回出たのでその確率は0.2であろう」というように、その確率は表1での出現割合にしたがうものと考えた。この考えが正しいと仮定したとき、Aさんが、手元の200円を使ってくじを引いて得られる点数の合計の期待値ができるだけ高くなるようにするには、くじXとYをそれぞれ何回引けばよいかを求めよ。ただし、くじXおよびYにはカードが十分な数あり、くじを引いてもカードが出る確率は変わらないと仮定する。

問 4. くじ X には星が 2 つまたは 4 つ書かれたカードのみが入っており、星が 4 つのカードを引き当てる確率を  $p$  とする。このくじ X を 1 回引いた時の点数の期待値を  $p$  を用いて表せ。ただし、星の数と点数は表 2 のとおりとする。

問 5. B さんは問 4 の条件でくじ X を売っており、A さんにその条件を説明した後にくじを 1 回 100 円で引いてもらった。その際に、実際には  $p = 0.1$  であるところを、誤って  $p = 0.2$  であると、A さんに伝えてしまっていた。B さんは、くじ X を  $p = 0.1$  で 1 回引くときの 1 円あたりに得られる点数の期待値が、 $p = 0.2$  で 1 回くじを引いたときの値と同じになるように、A さんが払い過ぎた金額を払い戻したいと考えている。B さんは A さんに対していくら払い戻せばよいか求めよ。ただし、くじを引いたときに出るカードの星の数と得られる点数の関係は表 2 のとおりであり、払い戻し金額が 1 円を下回る部分は切り上げて 1 円を払うものとする。

2 次の各問に答えなさい。(配点 15%)

問 1.  $n$  人が受験したある試験の得点は平均 56, 分散 25 の正規分布に従うという。A さんのこの試験の得点の偏差値は 64, B さんのこの試験の得点は 50 点であった。また, 全受験者のうち, 得点が B さんの得点以上かつ A さんの得点以下であった人数は 8041 人であった。このとき, 受験者数  $n$  を求めなさい。必要であれば, 正規分布表(次ページ参照)を用いてもよい。

問 2. 対応する 2 つの変量  $x, y$  の値の組を  $(x_1, y_1), \dots, (x_n, y_n)$  とし,  $x$  の分散を  $s_x^2$ ,  $y$  の分散を  $s_y^2$  とする。このとき,  $x + y$  と  $x - y$  の共分散を  $s_{x+y}^2$  と  $s_{x-y}^2$  を用いて表しなさい。

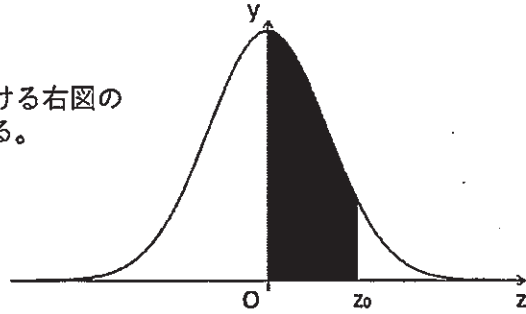
問 3. 確率変数  $X$  のとり得る値  $x$  の範囲が  $-3 \leq x \leq 3$  で, その確率密度関数  $f(x)$  が

$$f(x) = \frac{4-x}{24} \quad (-3 \leq x \leq 3)$$

で表されるとき, 平均  $E(X)$  と分散  $V(X)$  を求めよ。

# 正規分布表

次の表は、標準正規分布の分布曲線における右図の灰色部分の面積の値をまとめたものである。



0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0477	0.0517	0.0557	0.0596	0.0635	0.0675	0.0714	0.0753
0.2	0.0743	0.0782	0.0821	0.0859	0.0897	0.0935	0.0973	0.1011	0.1049	0.1087
0.3	0.1179	0.1217	0.1255	0.1292	0.1329	0.1365	0.1401	0.1437	0.1472	0.1507
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
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
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4405	0.4416	0.4427	0.4437
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4908	0.4910	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4947	0.4948	0.4949	0.4950
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998
3.5	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998

- 3 次の文章を読んで各問に答えなさい。解答の分量は任意である。分量が多ければ高得点というわけではない。問に対して、適切と思われる分量を自身で判断し、解答を解答用紙に記述しなさい。(配点 60%)

近年、Twitter、Facebook、LINE といったソーシャル・ネットワーキング・サービス(以下、SNS)の普及により、簡単かつ広範囲に情報を発信できるようになった。一方で、SNS 上での誤った情報の蔓延が社会問題となっている。

Vosoughi らの研究グループは、SNS の一種である Twitter を対象に、実際に正しいことが確認されている風説<sup>(1)</sup>(以下、“正しい”風説)と誤っていることが確認されている風説(以下、“誤った”風説)の拡散傾向の違いについて調査を行った<sup>(2)</sup>。

(1) 風説とは「世間でとりざたすること。また、そのうわさ」を意味する。

(2) Vosoughi, *et al.* (2018). The spread of true and false news online, *Science*, 359 (6380), pp.1146-1151, DOI : 10.1126/science.aap 9559.

From “The spread of true and falsenews online”. Reprinted with permission from AAAS.

Twitter とは、「ツイート」と呼ばれる短いメッセージや画像、動画、ウェブページへのリンクをオンライン上に投稿できるサービスである。Twitter を利用するユーザ(以下、Twitter ユーザ)は、タイムラインと呼ばれる画面で、自身の投稿および特定の Twitter ユーザの投稿を時系列順に閲覧することができる。図 1 はある Twitter ユーザのタイムラインの例である。Twitter ユーザは他の Twitter ユーザを「フォロー」することで、自身のタイムライン上に他ユーザのツイートを表示させることができる(図 1 の黒丸 a, b, c)。また、ユーザは他ユーザのツイートを「リツイート」することで、自身をフォローするユーザ宛に他ユーザのツイートを再配信することができる(図 1 の黒丸 b)。

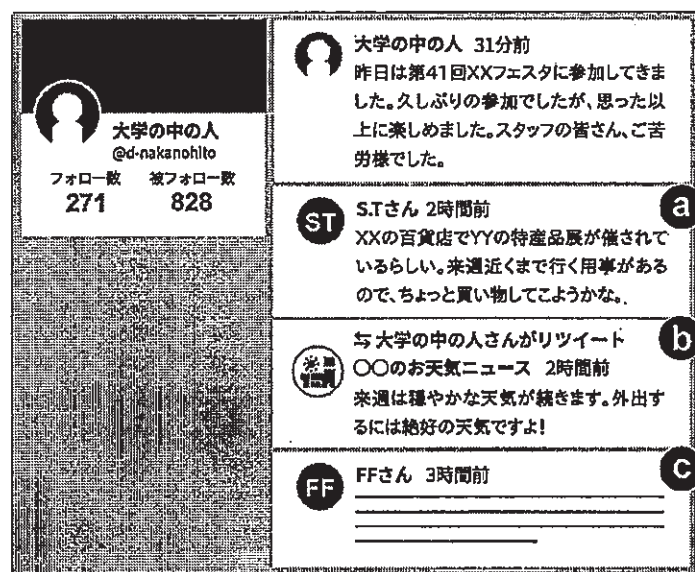


図 1. ある Twitter ユーザのタイムライン例。



図2は、Twitter ユーザ間のフォロー関係をネットワーク図として図示したものである。図中の丸印はTwitter ユーザを表す。丸印と丸印をつなぐ矢印は、あるユーザが別のユーザをフォローしていることを意味する。例えば、ユーザAはユーザGのみをフォローしていることが分かる。Twitter では、ユーザが情報を投稿すると、そのユーザをフォローするユーザに情報が拡散される。したがって、図2のようなTwitter ユーザ間のフォロー関係は「Twitter における情報の拡散経路」を表現しているとみなせる。なお、フォローは双方向性のない、投稿を受け取るだけの行為であるため、図中のユーザAとBのように、ユーザBがユーザAをフォローしている場合でも、ユーザAがユーザBをフォローしていない場合もありうる。

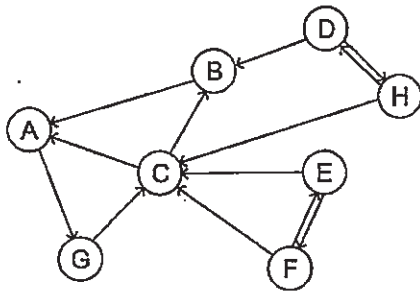


図2. Twitter ユーザのフォロー関係の図。

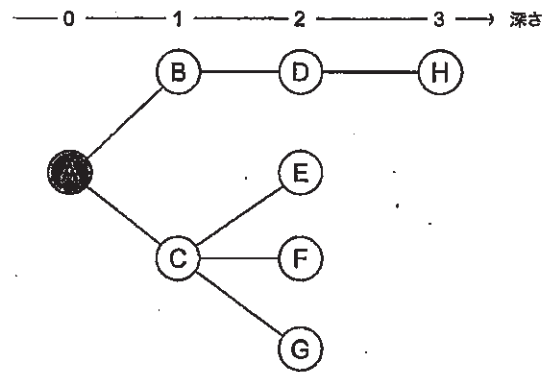


図3. 風説のカスケードを表現した図。

上記の性質をもつ Twitter において、ある風説が拡散された様子(以下、カスケード)は、図3のようなネットワーク図で表現することができる。図中の丸印は風説に関するツイートを投稿した Twitter ユーザを表す。丸印中のアルファベットの順序は風説に関するツイートの投稿順序を表している。丸印と丸印をつなぐ線は、ある Twitter ユーザが別ユーザの風説に関するツイートをリツイート(再投稿)したことを意味する。例えば図3において、ユーザAはある風説に関するツイートを最初に投稿した人物であり、ユーザBとユーザCはユーザAのツイートをリツイートしたことを示している。また、風説の拡散はユーザHの投稿を最後に止まったことを示している。

Vosoughi らの研究グループは、Twitter 上で発生したさまざまな風説のカスケードに関して、カスケードのサイズ、深さ、最大幅の観点から分析を行った。各観点の定義を以下に記す。

### カスケードのサイズ

カスケード中に含まれるユーザの数。

### カスケードの深さ

カスケード中に含まれるあるユーザXの深さを、最初にツイートを投稿したユーザAとユーザXの間に存在する線の数とする。このとき、カスケードの深さを「カスケード中に含まれるユーザの深さの最大値」と定義する(例: 図3に記した数字は、各ユーザの深さを示している。図3におけるカスケードの深さは3となる)。

### カスケードの最大幅

カスケードにおいて、深さ*i*におけるカスケード幅を「深さ*i*におけるユーザの数」と定義する。このとき、深さ*d*をもつカスケードの最大幅を「深さ0, 深さ1, ..., 深さ*d*におけるカスケード幅の中の最大値」と定義する(例: 図3におけるカスケードの最大幅は4となる)。

問 1. 図4は、Twitter上で発生した“正しい”風説に関するカスケードおよび“誤った”風説に関するカスケードの深さ(図4(a)), サイズ(図4(b)), および最大幅(図4(c))に関する図である。それぞれの図の横軸は、カスケードの深さ, サイズ, および最大幅を示している。図の縦軸は、総カスケード数に対するある深さ, サイズ, および最大幅以上のカスケードの数の割合(Complementary Cumulative Distribution Function; 以下CCDF)を示している。なお、図中の目盛は通常の日盛とは異なり、目盛の数が1, 10, 100のように等比数列となっている。例えば、図4(a)中の点線が示しているように、“正しい”風説に関するカスケードのうち深さが10以上のものは、“正しい”風説に関するカスケードの総数の0.01~0.1%であったことが分かる。図4をもとに、“正しい”風説と“誤った”風説の拡散傾向の相違点について、カスケードの深さ, サイズ, 最大幅の各々に言及しつつ分かることを述べなさい。

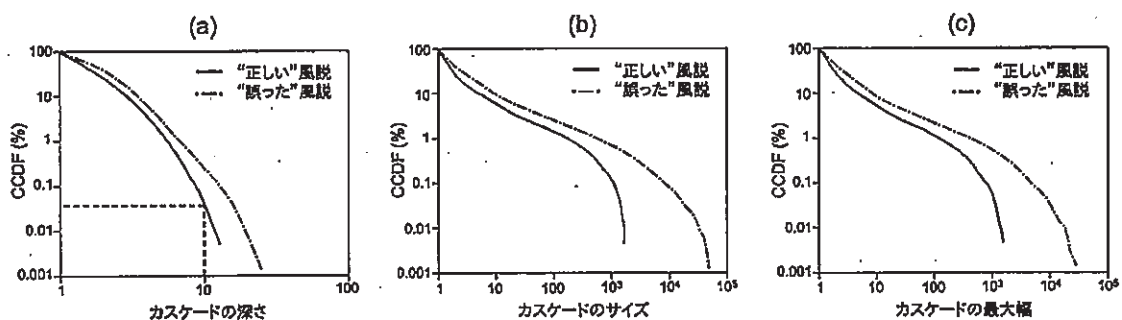


図4. “正しい”風説, “誤った”風説のカスケードの深さ, サイズ, および最大幅に関するCCDF。



問 2. 表 1 は, “正しい”風説および“誤った”風説の拡散にかかわった Twitter ユーザのフォロー数, 被フォロー数, 活動度, および Twitter 歴の中央値と平均値を示している。フォロー数は, ある Twitter ユーザがフォローしている Twitter ユーザの数である。被フォロー数とは, ある Twitter ユーザが何人の Twitter ユーザにフォローされているかを示す数である。活動度は, 投稿, リツイート, 返信といった Twitter 上での情報発信活動の積極性のある方法によって数値化したものである。Twitter 歴は, Twitter サービスの利用日数を表す。

以下の(ア)~(ウ)について, 表 1 に記された中央値と平均値から予想されるヒストグラムの形状として最も近いと思われるものを, 図 5 の(a)~(c)からそれぞれ選びなさい。

- (ア) “正しい”風説の拡散に関わった Twitter ユーザのフォロー数
- (イ) “誤った”風説の拡散に関わった Twitter ユーザの活動度
- (ウ) “正しい”風説の拡散に関わった Twitter ユーザの Twitter 歴

表 1. “正しい”風説および“誤った”風説を拡散した Twitter ユーザの中央値と平均値。  
表中の「正」「誤」はそれぞれ“正しい”風説, “誤った”風説を拡散したユーザ群を表す。

指 標	中央値		平均値	
	正	誤	正	誤
フォロー数	509	383	1707	1002
被フォロー数	466	410	5240	2234
活動度	9.54	9.52	24.65	19.70
Twitter 歴	1214	982	1269	1072

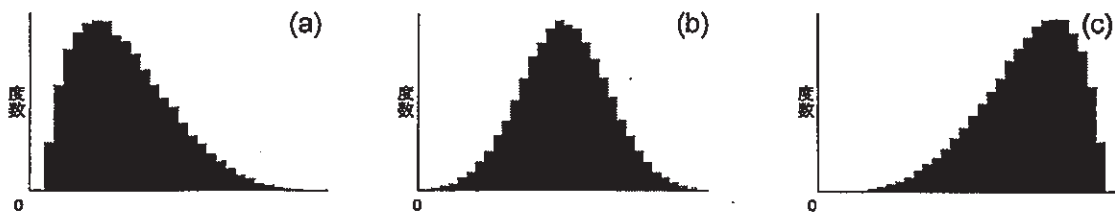


図 5. ヒストグラムの形状の候補。

問 3. Twitterにおいて、図4で観察されたような“正しい”風説と“誤った”風説の拡散傾向の違いが生じる要因として、少なくとも以下の3つが候補として考えられる：

想定要因1 Twitter ユーザの活動度

想定要因2 ツイートの拡散経路

想定要因3 風説に関するツイートから受ける印象

以下は想定要因1に関する考察例である。

想定要因1に関する考察例(考察の一部は\*\*\*に置き換えられている)

図4で観察されたような“正しい”風説と“誤った”風説の拡散傾向の違いに影響を与える要因として、Twitter ユーザの活動度は妥当性が\*\*\*と考える。

Twitter ユーザの活動度が高いということは、活動度の定義よりTwitter上での情報発信頻度が高いことを意味する。情報発信頻度が高まるほど、発信された情報が他者の目に触れる可能性は高くなるため、風説の拡散規模(カスケードの深さ、サイズ、最大幅)は発信頻度に比例して大きくなることが予想される。表1によると、“正しい”風説の拡散に関わるユーザの活動度は“誤った”風説の拡散に関わるユーザのそれよりも相対的に大きい。一方、図4によると、“正しい”風説と比べて“誤った”風説の方が\*\*\*。このことは先の予想と\*\*\*。

以上より、Twitter ユーザの活動度は、風説の拡散傾向に影響を与える要因として妥当性が\*\*\*と考える。

考察例を参考に、図4および表1を踏まえて、想定要因2の妥当性について考察しなさい。

問 4. Twitter ユーザの中には、自動的に情報を発信、共有する「ボット」と呼ばれるソフトウェアが含まれていることが分かっている。Vosoughiらは、調査データに含まれるTwitter ユーザのうち10%程度がボットであったこと、ボットは“正しい”風説に関するツイートと“誤った”風説に関するツイートをほぼ1:1の割合で行っていることを明らかにしている。また、“正しい”風説に関するツイートに比べ、“誤った”風説に関するツイートは「驚き」や「反感」を感じさせるような内容のものが多かったことも明らかにしている。このことから、Vosoughiらは、図4のように“正しい”風説と“誤った”風説の拡散傾向の違いが生じる要因の1つとして「風説に関するツイートから受ける印象」を挙げている。

これらの分析および図4、表1のデータを踏まえて、Twitterにおける“誤った”風説の拡散を防止する方法について、あなたの考えを論理的に述べなさい。