

分散の加法性を視覚的に理解する（その3）

Sampo Suzuki, CC 4.0 BY-NC-SA

2021-06-01

はじめに

分散の加法性を視覚的に理解する（その2）において、データが独立であれば分散の加法性がなりたつことがわかりました。では、同一正規分布から取り出した二つ、および、三つの値の平均値の場合はどうなるか、その2と同様の手段で確認してみます。

同一データからサンプリングした二つの値を平均した場合

最初に以下の処理を行う関数を定義します。

- データを乱数生成する¹
- 乱数生成したデータをランダムサンプリングする
- 作成したデータの統計量を求める
- 無相関検定の結果と統計量をデータフレームにまとめる

¹ 今回は `rnorm()` 関数による分散が 100 となる正規分布

```
1 f2 <- function(i = NA, n = 5000000) {  
2   # データを乱数生成する  
3   x <- rnorm(n = n, mean = 10, sd = 10)  
4   # 乱数生成したデータから二つのデータを取り出す  
5   a <- sample(x, n, replace = TRUE)  
6   b <- sample(x, n, replace = TRUE)  
7   num <- 2  
8   # 統計量を求める  
9   df <- data.frame(no = i,  
10                    var.x = var(x),  
11                    var.a = var(a), var.b = var(b),  
12                    var.ab = var((a + b) / num), var.sum = (var(a / num) + var(b / num)),  
13                    cov = cov(a / num, b / num),  
14                    cov2 = cov(a / num, b / num) * 2)  
15   # 無相関の検定結果と統計量をデータフレームにまとめる  
16   df <- cor.test(a, b) %>% broom::tidy() %>% dplyr::bind_cols(df)  
17   return(df)  
18 }
```

Table 1: 二つのサンプルを平均した場合の分散

No	相関係数	p 値	母集団	標本 a	標本 b	加法 1	加法 2	差異	母集団比	cov2
1	-0.0004227	0.3446167	99.87213	99.87870	99.96647	49.94017	49.96129	-0.0211163	0.5000411	-0.0211163
2	0.0006894	0.1231613	100.07524	99.96190	100.07439	50.04355	50.00907	0.0344784	0.5000593	0.0344784
3	-0.0002985	0.5044151	100.03005	100.12848	100.12806	50.04919	50.06414	-0.0149462	0.5003415	-0.0149462
4	-0.0005679	0.2041367	100.05189	99.95022	100.02934	49.96650	49.99489	-0.0283919	0.4994058	-0.0283919
5	-0.0007649	0.0871787	100.02386	99.96435	99.96240	49.94345	49.98169	-0.0382333	0.4993154	-0.0382333
6	0.0004128	0.3560166	99.96980	100.00515	100.05158	50.03483	50.01418	0.0206444	0.5004994	0.0206444
7	-0.0003800	0.3955490	99.98493	99.96132	99.98555	49.96773	49.98672	-0.0189925	0.4997526	-0.0189925
8	-0.0004313	0.3348360	100.02496	100.04873	99.96994	49.98310	50.00467	-0.0215671	0.4997063	-0.0215671
9	0.0001845	0.6798832	99.99130	99.91004	100.03260	49.99488	49.98566	0.0092239	0.4999924	0.0092239
10	-0.0006684	0.1350209	100.09096	100.13762	100.01318	50.00425	50.03770	-0.0334453	0.4995881	-0.0334453
11	0.0005206	0.2444083	99.92114	99.93640	99.94230	49.99569	49.96968	0.0260129	0.5003515	0.0260129
12	0.0007688	0.0856032	100.03008	100.12376	99.95333	50.05773	50.01927	0.0384542	0.5004268	0.0384542
13	-0.0001434	0.7484204	100.01911	99.99394	100.09906	50.01607	50.02325	-0.0071749	0.5000652	-0.0071749
14	0.0000048	0.9914207	100.10705	100.24055	100.10889	50.08760	50.08736	0.0002409	0.5003404	0.0002409
16	0.0001215	0.7857846	100.05458	100.13252	100.17808	50.08374	50.07765	0.0060868	0.5005642	0.0060868
17	-0.0006408	0.1518697	99.85948	99.88638	99.86088	49.90481	49.93681	-0.0320014	0.4997504	-0.0320014
18	0.0001450	0.7456971	100.02241	99.93323	99.93438	49.97415	49.96690	0.0072472	0.4996295	0.0072472
19	-0.0003088	0.4899390	100.12372	100.02042	100.19305	50.03791	50.05337	-0.0154544	0.4997608	-0.0154544
20	-0.0001323	0.7673047	99.92560	99.96056	99.95999	49.97352	49.98014	-0.0066140	0.5001073	-0.0066140
21	-0.0002764	0.5365530	100.02100	100.03889	99.93493	49.97964	49.99345	-0.0138179	0.4996915	-0.0138179
22	0.0005443	0.2235309	100.07907	100.06634	100.10017	50.06887	50.04163	0.0272399	0.5002931	0.0272399
23	-0.0000553	0.9015291	99.92606	100.04581	99.91932	49.98852	49.99128	-0.0027662	0.5002551	-0.0027662
24	-0.0006679	0.1353417	100.00815	99.94482	99.99480	49.95152	49.98491	-0.0333826	0.4994745	-0.0333826
25	0.0005606	0.2100268	100.07681	99.99547	100.04771	50.03883	50.01080	0.0280350	0.5000043	0.0280350
26	0.0001202	0.7881156	100.16421	100.13795	100.28389	50.11148	50.10546	0.0060223	0.5002933	0.0060223
27	0.0002696	0.5466506	100.00794	100.10217	100.04631	50.05061	50.03712	0.0134887	0.5004664	0.0134887
28	0.0002009	0.6532036	99.84813	99.85202	99.85932	49.93787	49.92784	0.0100325	0.5001382	0.0100325
29	-0.0002756	0.5377413	99.94330	99.96889	99.89002	49.95096	49.96473	-0.0137697	0.4997930	-0.0137697
31	-0.0002466	0.5813051	100.01894	100.03143	100.01167	49.99844	50.01078	-0.0123341	0.4998897	-0.0123341
32	-0.0006443	0.1496653	99.89803	99.91691	99.90149	49.92241	49.95460	-0.0321861	0.4997337	-0.0321861
33	-0.0001251	0.7797496	100.01071	99.98714	100.11296	50.01877	50.02502	-0.0062562	0.5001341	-0.0062562
34	-0.0000266	0.9525761	99.96254	99.89658	99.96925	49.96513	49.96646	-0.0013289	0.4998385	-0.0013289
35	-0.0000921	0.8368399	99.92419	99.87316	99.96447	49.95481	49.95941	-0.0046012	0.4999271	-0.0046012
36	0.0003492	0.4349171	100.01245	100.05928	99.96914	50.02457	50.00711	0.0174618	0.5001834	0.0174618
38	-0.0006341	0.1562326	99.96110	100.07149	99.82457	49.94233	49.97401	-0.0316877	0.4996176	-0.0316877
39	-0.0008329	0.0625534	99.94755	99.99783	99.95544	49.94668	49.98832	-0.0416337	0.4997289	-0.0416337
40	0.0003846	0.3897481	100.00907	99.92099	100.02583	50.00593	49.98670	0.0192267	0.5000140	0.0192267
41	0.0002062	0.6447010	99.95999	100.00154	100.01233	50.01378	50.00347	0.0103120	0.5003380	0.0103120
42	0.0001978	0.6582189	100.08461	100.20002	100.07181	50.07786	50.06796	0.0099052	0.5003553	0.0099052
44	0.0007321	0.1016436	100.01251	99.96788	100.08360	50.04948	50.01287	0.0366124	0.5004322	0.0366124
45	0.0004272	0.3394427	99.99145	100.02156	99.93629	50.01082	49.98946	0.0213559	0.5001509	0.0213559

Table 2: 二つのサンプルが独立でない場合

No	相関係数	p 値	母集団	標本 a	標本 b	加法 1	加法 2	差異	母集団比	cov2
15	0.0010751	0.0162212	100.08494	100.1292	100.13671	50.12029	50.06647	0.0538243	0.5007776	0.0538243
30	0.0009400	0.0355671	100.06054	100.1077	100.10518	50.10027	50.05322	0.0470486	0.5006996	0.0470486
37	0.0009437	0.0348487	99.96085	100.0225	99.95927	50.04263	49.99545	0.0471792	0.5006223	0.0471792
43	0.0011961	0.0074831	100.01955	100.0622	99.96043	50.06548	50.00566	0.0598113	0.5005569	0.0598113

$$\text{加法 1} = \text{var}\left(\frac{a+b}{2}\right), \text{加法 2} = \text{var}\left(\frac{a}{2}\right) + \text{var}\left(\frac{b}{2}\right)$$

同一データからサンプリングした三つの値を平均した場合

最初に以下の処理を行う関数を定義します。

- データを乱数生成する²
- 乱数生成したデータをランダムサンプリングする
- 作成したデータの統計量を求める
- 無相関検定の結果と統計量をデータフレームにまとめる

² 今回は `rnorm()` 関数による分散が 100 となる正規分布

```

1 f3 <- function(i = NA, n = 5000000) {
2   # データを乱数生成する
3   x <- rnorm(n = n, mean = 10, sd = 10)
4   # 乱数生成したデータから三つのデータを取り出す
5   a <- sample(x, n, replace = TRUE)
6   b <- sample(x, n, replace = TRUE)
7   c <- sample(x, n, replace = TRUE)
8   num <- 3
9   # 統計量を求める
10  df <- data.frame(no = i,
11                  var.x = var(x),
12                  var.a = var(a), var.b = var(b), var.c = var(c),
13                  var.abc = var((a + b + c) / num),
14                  var.sum = (var(a / num) + var(b / num) + var(c / num)),
15                  cov.ab = cov(a, b), cov.ac = cov(a, c), cov.bc = cov(b, c),
16                  cov2.ab = cov(a, b) * 2, cov2.ac = cov(a, c) * 2, cov2.bc = cov(b, c) * 2)
17  # 無相関の検定結果と統計量をデータフレームにまとめる
18  df <- cor.test(a, b) %>% broom::tidy() %>% dplyr::bind_cols(df)
19  df <- cor.test(a, c) %>% broom::tidy() %>% dplyr::bind_cols(df)
20  df <- cor.test(b, c) %>% broom::tidy() %>% dplyr::bind_cols(df)
21  return(df)
22 }
```

Table 3: 三つのサンプルを平均した場合の分散

No	母集団	標本 a	標本 b	標本 c	加法 1	加法 2	差異	母集団比
1	100.04444	99.99364	99.95237	99.97451	33.30230	33.32450	-0.0221988	0.3328751
3	99.97346	99.93094	99.92829	99.87890	33.30009	33.30424	-0.0041419	0.3330893
4	100.04684	99.99903	100.01119	100.09931	33.32948	33.34550	-0.0160241	0.3331388
5	99.98713	100.06519	99.94869	99.98973	33.30402	33.33373	-0.0297117	0.3330831
6	100.11323	100.05457	100.20764	100.13805	33.38514	33.37781	0.0073303	0.3334738
7	99.97585	100.06765	99.98634	99.93075	33.29689	33.33164	-0.0347474	0.3330493
8	100.03576	100.04278	100.07724	100.04879	33.36852	33.35209	0.0164251	0.3335659
10	100.02147	100.00110	100.02245	99.87074	33.36918	33.32159	0.0475916	0.3336202
12	100.05354	99.96653	100.00867	100.04385	33.29563	33.33545	-0.0398189	0.3327781
13	100.00835	99.95240	100.05820	100.07131	33.34059	33.34243	-0.0018450	0.3333781
14	100.01183	100.01245	100.00605	99.98231	33.31892	33.33342	-0.0145007	0.3331498
15	100.09281	100.10893	100.09232	99.98592	33.36939	33.35413	0.0152559	0.3333844
16	99.98783	99.87088	99.97196	99.98309	33.30840	33.31399	-0.0055924	0.3331245
17	99.95895	100.02779	100.00703	99.96990	33.33737	33.33386	0.0035097	0.3335106
18	99.97938	99.92114	99.96560	99.97708	33.30683	33.31820	-0.0113705	0.3331370
19	100.06790	100.00039	99.97305	100.11321	33.35061	33.34296	0.0076534	0.3332799
20	100.10186	100.05983	100.10586	100.10155	33.37596	33.36303	0.0129313	0.3334200
21	99.97606	99.92166	99.93785	99.97295	33.30252	33.31472	-0.0122004	0.3331049
22	100.06831	100.04468	100.06287	99.99709	33.34573	33.34496	0.0007739	0.3332297
23	99.84294	99.90149	99.71609	100.01128	33.28312	33.29210	-0.0089743	0.3333548
24	100.06130	100.12195	100.07985	100.11951	33.33344	33.36903	-0.0355897	0.3331302
25	100.09145	100.01542	100.12195	100.19300	33.37587	33.37004	0.0058327	0.3334538
26	99.94882	99.96522	99.98467	99.90111	33.31360	33.31678	-0.0031763	0.3333066
27	99.93525	99.99087	99.90116	99.96036	33.30939	33.31693	-0.0075442	0.3333097
28	100.03707	100.06852	99.91367	99.96827	33.32400	33.32783	-0.0038278	0.3331165
29	100.03579	99.96155	99.96476	100.06081	33.34713	33.33190	0.0152284	0.3333520
30	100.10607	100.21994	100.18056	99.94590	33.37703	33.37182	0.0052115	0.3334167
31	100.06611	100.07600	100.02569	100.00403	33.34717	33.34508	0.0020945	0.3332514
32	100.02972	100.02388	100.01994	99.93302	33.31065	33.33076	-0.0201115	0.3330075
33	99.93844	100.04248	99.93554	100.02755	33.31410	33.33395	-0.0198471	0.3333463
34	100.12003	99.99567	100.06442	100.11670	33.33656	33.35298	-0.0164191	0.3329659
35	99.97455	99.97492	100.06277	99.98965	33.30549	33.33637	-0.0308814	0.3331397
36	100.08121	100.14637	99.95326	100.18183	33.37774	33.36461	0.0131284	0.3335065
37	99.94658	99.81552	100.10360	100.01504	33.33012	33.32602	0.0041007	0.3334793
38	100.04696	100.11784	100.02709	100.12850	33.36266	33.36372	-0.0010547	0.3334700
39	100.08929	100.13179	100.18596	100.03494	33.38592	33.37252	0.0133982	0.3335614
40	100.04294	99.96954	99.94056	100.08237	33.32970	33.33250	-0.0027918	0.3331540
41	99.97840	100.00749	100.02821	100.00197	33.34887	33.33752	0.0113495	0.3335607
42	99.96526	99.92862	99.92791	99.98663	33.31896	33.31591	0.0030531	0.3333054
43	99.94242	99.94334	99.92731	99.84776	33.30515	33.30205	0.0031072	0.3332434
44	99.96083	99.92636	100.01802	100.00641	33.31566	33.32787	-0.0122018	0.3332872

Table 4: 三つのサンプルのどれかが独立でない場合

No	母集団	標本 a	標本 b	標本 c	加法 1	加法 2	差異	母集団比
2	100.01114	100.10941	100.10965	99.98592	33.35917	33.35611	0.0030629	0.3335546
9	99.88774	99.86397	99.84461	99.92135	33.28043	33.29221	-0.0117852	0.3331783
11	100.10944	100.05398	100.12787	100.11281	33.37623	33.36607	0.0101604	0.3333975
45	99.89511	99.76852	99.83532	99.81607	33.28638	33.26888	0.0175006	0.3332133

$$\text{加法 1} = \text{var}\left(\frac{a+b+c}{3}\right), \text{ 加法 2} = \text{var}\left(\frac{a}{3}\right) + \text{var}\left(\frac{b}{3}\right) + \text{var}\left(\frac{c}{3}\right)$$

まとめ

データが独立であれば分散の加法性が成り立っており、 n 個の平均をとった場合、分散が $\frac{1}{n}$ になることが予想できます。

About handout style

The Tufte handout style is a style that Edward Tufte uses in his books and handouts. Tufte's style is known for its extensive use of sidenotes, tight integration of graphics with text, and well-set typography. This style has been implemented in LaTeX and HTML/CSS³, respectively.

³ See Github repositories `tufte-latex` and `tufte-css`