

分散の加法性を視覚的に理解する（その2）

Sampo Suzuki, CC 4.0 BY-NC-SA

2021-06-02

はじめに（小室先生のアドバイスから）

分散の加法性が成り立つには「データが独立」であるという前提条件があります。乱数生成した二つのデータ¹が本当に独立なのかを確認すると共に分散の加法性も確認してみます。

¹ 各々 5×10^6 個のデータ

関数の定義

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³
- 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

³ `sampling = TRUE` の場合のみ

```
1 f <- function(i = NA, sampling = FALSE, n = 5000000) {
2   # データを乱数生成する
3   x <- rnorm(n = n, mean = 10, sd = 10)
4   y <- rnorm(n = n, mean = 30, sd = 10)
5   # 乱数生成したデータからサンプリングする場合
6   if (sampling == TRUE) {
7     x <- sample(x, n, replace = TRUE)
8     y <- sample(y, n, replace = TRUE)
9   }
10  # 統計量を求める
11  df <- data.frame(no = i, var.x = var(x), var.y = var(y),
12                  var.xy = var(x + y), var.sum = var(x) + var(y),
13                  cov = cov(x, y), cov2 = cov(x, y) * 2)
14  # 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
15  df <- cor.test(x, y) %>% broom::tidy() %>% dplyr::bind_cols(df)
16  return(df)
17 }
```

この関数を for ループで 45 回繰り返し、その結果をデータフレームにまとめ、分散がどのようなになっているかを比較します。

Table 1: 変数の意味

変数名	その意味	備考
var.x	データ x の分散	
var.y	データ y の分散	
var.xy	データ x と y を加算したものの分散 ($\text{var}(x + y)$)	加法 1
var.sum	データ x, y の分散を加算したもの ($\text{var}(x) + \text{var}(y)$)	加法 2
var.diff	var.xy から var.sum を減算したもの	加法 1 と加法 2 の差異
cov2	データ x, y の共分散の 2 倍数	
cov	データ x, y の共分散	計算のみで未出力

乱数生成したデータの場合

No	相関係数	p 値	標本 x	標本 y	加法 1	加法 2	差異	cov2
1	0.0002405	0.5906758	100.03355	99.95040	200.0321	199.9840	0.0481035	0.0481035
2	0.0007900	0.0773232	100.04689	100.11172	200.3167	200.1586	0.1581199	0.1581199
3	0.0001524	0.7333445	100.00385	99.94863	199.9829	199.9525	0.0304641	0.0304641
4	0.0003742	0.4027862	100.06641	99.97769	200.1189	200.0441	0.0748493	0.0748493
5	-0.0000289	0.9484144	99.92723	99.99781	199.9193	199.9250	-0.0057846	-0.0057846
6	-0.0008565	0.0554575	99.96148	99.96261	199.7528	199.9241	-0.1712423	-0.1712423
7	0.0001456	0.7447831	100.15102	100.03861	200.2188	200.1896	0.0291435	0.0291435
8	-0.0002626	0.5570963	99.93554	99.91521	199.7983	199.8508	-0.0524780	-0.0524780
9	-0.0002504	0.5754730	99.92129	100.03234	199.9036	199.9536	-0.0500772	-0.0500772
11	-0.0004257	0.3412065	99.97925	100.03056	199.9247	200.0098	-0.0851343	-0.0851343
12	-0.0004302	0.3360899	100.09555	100.03086	200.0403	200.1264	-0.0860911	-0.0860911
13	-0.0001740	0.6972663	100.13433	99.94565	200.0452	200.0800	-0.0348083	-0.0348083
15	-0.0002304	0.6064396	100.03044	99.97619	199.9606	200.0066	-0.0460791	-0.0460791
16	-0.0004032	0.3672358	99.96101	100.04571	199.9261	200.0067	-0.0806499	-0.0806499
17	-0.0002355	0.5985000	99.86689	99.91427	199.7341	199.7812	-0.0470453	-0.0470453
18	-0.0001552	0.7285535	100.04952	99.95039	199.9689	199.9999	-0.0310411	-0.0310411
19	-0.0003013	0.5004650	99.93607	99.93441	199.8103	199.8705	-0.0602238	-0.0602238
20	0.0003161	0.4797113	99.95667	100.08492	200.1048	200.0416	0.0632283	0.0632283
21	-0.0008416	0.0598590	100.04632	100.12057	199.9984	200.1669	-0.1684565	-0.1684565
22	-0.0000757	0.8656125	100.06773	100.07331	200.1259	200.1410	-0.0151474	-0.0151474
24	0.0002168	0.6278622	99.99654	100.03160	200.0715	200.0281	0.0433624	0.0433624
25	0.0003096	0.4887490	99.98423	99.94195	199.9881	199.9262	0.0618983	0.0618983
26	-0.0003628	0.4171664	99.91858	99.99546	199.8415	199.9140	-0.0725379	-0.0725379
27	0.0004670	0.2964230	99.84074	99.95558	199.8896	199.7963	0.0932952	0.0932952
28	0.0001997	0.6551878	100.06020	99.96489	200.0650	200.0251	0.0399472	0.0399472
29	-0.0004876	0.2756164	100.07406	99.98583	199.9623	200.0599	-0.0975414	-0.0975414
30	0.0001878	0.6745860	100.10249	99.97609	200.1162	200.0786	0.0375684	0.0375684
31	-0.0005518	0.2127930	100.09740	100.05732	200.0443	200.1547	-0.1104361	-0.1104361
32	0.0001833	0.6818780	100.14285	100.04816	200.2277	200.1910	0.0366978	0.0366978
33	-0.0004313	0.3348183	99.95623	99.95133	199.8213	199.9076	-0.0862235	-0.0862235
34	-0.0008274	0.0642915	99.97238	100.05238	199.8593	200.0248	-0.1655032	-0.1655032
35	0.0002048	0.6469738	100.02357	99.93272	199.9972	199.9563	0.0409532	0.0409532
36	0.0004283	0.3382556	100.08332	100.04006	200.2091	200.1234	0.0857048	0.0857048
37	-0.0003048	0.4954516	99.98206	99.92530	199.8464	199.9074	-0.0609416	-0.0609416
38	0.0001726	0.6995903	100.01911	99.99586	200.0495	200.0150	0.0345162	0.0345162
39	0.0007299	0.1026352	99.95899	99.95363	200.0585	199.9126	0.1459256	0.1459256
40	-0.0002076	0.6425365	100.01073	99.93473	199.9040	199.9455	-0.0415040	-0.0415040
41	-0.0001471	0.7422250	100.05012	99.97362	199.9943	200.0237	-0.0294220	-0.0294220
42	0.0001196	0.7892026	99.91426	99.99920	199.9374	199.9135	0.0239018	0.0239018
43	-0.0006294	0.1592944	99.98230	100.04003	199.8964	200.0223	-0.1259001	-0.1259001
44	-0.0005545	0.2150104	99.96039	100.02832	199.8778	199.9887	-0.1108943	-0.1108943
45	0.0002047	0.6471306	100.10302	100.06351	200.2075	200.1665	0.0409767	0.0409767

Table 2: 乱数生成した二つのデータの分散

No	相関係数	p 値	標本 x	標本 y	加法 1	加法 2	差異	cov2
10	-0.0011382	0.0109276	100.02103	100.02146	199.8148	200.0425	-0.2276805	-0.2276805
14	-0.0008911	0.0463006	99.97067	99.96079	199.7533	199.9315	-0.1781663	-0.1781663
23	-0.0009108	0.0416865	100.03319	99.89816	199.7493	199.9314	-0.1820997	-0.1820997

Table 3: 乱数生成した二つのデータが独立でない場合

$$\text{加法 1} = \text{var}(x + y), \text{加法 2} = \text{var}(x) + \text{var}(y)$$

乱数生成したデータをランダムサンプリングした場合

Table 4: ランダムサンプリングしたデータの分散

No	相関係数	p 値	標本 x	標本 y	加法 1	加法 2	差異	cov2
1	0.0004992	0.2643631	100.00419	100.05990	200.1639	200.0641	0.0998624	0.0998624
2	-0.0003290	0.4618939	99.96097	99.92836	199.8236	199.8893	-0.0657695	-0.0657695
3	0.0007499	0.0935831	99.90959	99.97202	200.0315	199.8816	0.1498881	0.1498881
4	-0.0004114	0.3576327	100.05543	100.04644	200.0195	200.1019	-0.0823190	-0.0823190
5	-0.0001846	0.6796877	100.06244	99.94424	199.9697	200.0067	-0.0369312	-0.0369312
6	0.0004582	0.3055784	100.13468	99.97410	200.2005	200.1088	0.0916878	0.0916878
7	-0.0000583	0.8962583	99.93498	100.02294	199.9463	199.9579	-0.0116599	-0.0116599
8	0.0006919	0.1218079	100.01941	99.96146	200.1192	199.9809	0.1383752	0.1383752
9	0.0006845	0.1258807	99.89239	99.94080	199.9700	199.8332	0.1367824	0.1367824
10	0.0005111	0.2530770	100.13476	99.98200	200.2191	200.1168	0.1022844	0.1022844
11	0.0007520	0.0926401	99.92396	99.98816	200.0625	199.9121	0.1503437	0.1503437
12	0.0001118	0.8025082	100.05911	99.98613	200.0676	200.0452	0.0223749	0.0223749
13	-0.0002818	0.5286018	100.01455	99.94620	199.9044	199.9608	-0.0563506	-0.0563506
14	0.0002972	0.5062940	100.03436	100.06820	200.1620	200.1026	0.0594759	0.0594759
15	-0.0003046	0.4958450	99.96929	100.08177	199.9901	200.0511	-0.0609298	-0.0609298
16	-0.0005595	0.2109395	99.95303	99.85495	199.6962	199.8080	-0.1117842	-0.1117842
17	0.0000517	0.9080509	99.94353	99.91597	199.8698	199.8595	0.0103231	0.0103231
18	-0.0005386	0.2284480	100.10207	99.92016	199.9145	200.0222	-0.1077337	-0.1077337
19	0.0003708	0.4070281	99.99910	99.94069	200.0139	199.9398	0.0741378	0.0741378
20	0.0002660	0.5519554	99.97989	100.20805	200.2412	200.1879	0.0532534	0.0532534
21	0.0001940	0.6644167	99.95954	100.07311	200.0715	200.0326	0.0388087	0.0388087
22	0.0002388	0.5932892	99.84908	100.09592	199.9928	199.9450	0.0477560	0.0477560
23	-0.0000588	0.8953510	100.04059	99.88865	199.9175	199.9292	-0.0117608	-0.0117608
24	0.0000362	0.9355171	100.00611	100.06000	200.0733	200.0661	0.0072388	0.0072388
25	-0.0001666	0.7094174	99.86403	99.87116	199.7019	199.7352	-0.0332857	-0.0332857
26	-0.0003364	0.4519261	100.13461	100.07880	200.1461	200.2134	-0.0673513	-0.0673513
27	-0.0002519	0.5732147	99.97644	99.88570	199.8118	199.8621	-0.0503505	-0.0503505
28	-0.0007438	0.0962771	99.89452	99.91890	199.6648	199.8134	-0.1486203	-0.1486203
29	-0.0001734	0.6982670	99.88154	100.02434	199.8712	199.9059	-0.0346571	-0.0346571
30	-0.0007853	0.0790772	100.11982	100.05634	200.0190	200.1762	-0.1572057	-0.1572057
31	-0.0000707	0.8743035	100.13424	99.86444	199.9845	199.9987	-0.0141492	-0.0141492
32	-0.0001357	0.7614988	99.77530	100.05693	199.8051	199.8322	-0.0271243	-0.0271243
33	-0.0001069	0.8111459	99.95227	99.84470	199.7756	199.7970	-0.0213505	-0.0213505
34	-0.0000265	0.9528319	99.96384	99.88967	199.8482	199.8535	-0.0052867	-0.0052867
35	-0.0005338	0.2326680	99.99355	100.02165	199.9084	200.0152	-0.1067593	-0.1067593
36	0.0003994	0.3717949	100.07313	99.97915	200.1322	200.0523	0.0799036	0.0799036
37	0.0000662	0.8822869	100.09695	100.02205	200.1323	200.1190	0.0132517	0.0132517
39	-0.0000754	0.8661529	99.92744	100.01775	199.9301	199.9452	-0.0150712	-0.0150712
40	0.0000553	0.9015342	100.00996	100.08630	200.1073	200.0963	0.0110715	0.0110715
41	0.0007117	0.1114930	99.93513	99.98389	200.0613	199.9190	0.1422921	0.1422921
42	0.0007097	0.1125051	100.03694	100.05762	200.2366	200.0946	0.1420157	0.1420157
43	-0.0003738	0.4031844	100.10921	100.11551	200.1499	200.2247	-0.0748535	-0.0748535
44	0.0001112	0.8035807	99.91321	99.97610	199.9115	199.8893	0.0222335	0.0222335
45	0.0003250	0.4673757	100.07371	99.98714	200.1259	200.0609	0.0650228	0.0650228

Table 5: ランダムサンプリングしたデータが独立でない場合

No	相関係数	p 値	標本 x	標本 y	加法 1	加法 2	差異	cov2
38	0.0009822	0.0280692	100.1892	100.1929	200.5789	200.3821	0.19682	0.19682

$$\text{加法 1} = \text{var}(x + y), \text{ 加法 2} = \text{var}(x) + \text{var}(y)$$

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いることがわかります。データが独立とは言い難い無相関の検定が成功するケース (95% 信頼区間に 0 が入らない) では、分散の差 (共分散の 2 倍数) が一桁大きいので加法性が成り立っているとは言い難いように思えますがこのケースでは数値だけを見ている限り差はよくわかりません。

`cor.test()` 関数について

`cor.test()` 関数は無相関の検定を行う関数です。対立仮説 (H_1) は下記の出力の通り「true correlation is **not** equal to 0 (相関係数はゼロではない)」ですので、帰無仮説 (H_0) は「相関係数はゼロである (相関はない)」となります。有意水準 α で検定が失敗すれば (帰無仮説が棄却されない、 $p \geq \alpha$ である) 帰無仮説が採択されますので相関係数はゼロ (データ間には相関がない) と考えられます。

```
##
## Pearson's product-moment correlation
##
## data:  rnorm(n) and rnorm(n)
## t = -0.56198, df = 4999998, p-value = 0.5741
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.0011278474  0.0006251977
## sample estimates:
##           cor
## -0.000251325
```

Appendix

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/CSS⁴, respectively.

⁴ See Github repositories `tufte-latex` and `tufte-css`