母平均値と母比率の区間推定

樋口さぶろお

龍谷大学理工学部数理情報学科

確率統計☆演習 I L11(2015-12-11 Fri)

最終更新: Time-stamp: "2015-12-19 Sat 11:23 JST hig"

今日の目標

- 標本から母平均値を区間推定できる (母分散既知, 母分散未知)
- t 分布の確率が求められる
- 標本から母比率を区間推定できる

http://hig3.net

L10-Q1

Quiz 解答:母平均値, 母分散の点推定

1. 標本平均値は, $\frac{1}{6}(117+\cdots+112)=111\mathrm{g}$ なので, 母平均値は $111\mathrm{g}$ と推定できる.
2. 不偏標本分散は, $\frac{1}{6-1}[(117-111)^2+\cdots+(112-111)^2]=46\mathrm{g}^2$ なので, 母分散は $46\mathrm{g}^2$ と推定できる.

L10-Q2

Quiz 解答:推定

これはサイズ 10 の標本.
標本平均値は

$$\frac{1}{10}[0+0+0+0+0+0+10+10+30+100]=15(\text{円}).$$

よって, 母平均値は 15 円と推定される.

不偏標本分散は

$$\frac{1}{10-1}[(0-15)^2\times 6+(10-15)^2\times 2+(30-15)^2+(100-15)^2]=983.3(\text{円}^2)$$

. よって, 母分散は 983.3 円^2 と推定される.
母標準偏差は $\sqrt{983.3}=31.4$ 円と推定される.

L10-Q4

Quiz 解答:母比率の点推定

1. 母平均値 $p=\mathrm{E}[Y]$ を $\overline{Y}=\frac{12}{20}=\frac{3}{5}$ と推定する.
2. 母分散 $\mathrm{V}[Y]=p(1-p)$ を, $\frac{3}{5}\cdot\frac{2}{5}$ と推定する.

L10-Q5

Quiz 解答:母比率の点推定

標本比率 $\hat{p}$ は $\frac{12}{20}=\frac{3}{5}=0.6$ なので母比率 p を 0.6 と推定する.

ここまで来たよ

点推定 対 区間推定

点推定

真の母平均値はわからないが, 標本平均値を使って,

「母平均値は A 円と推定される」

それどのくらい正確なの? 実は

区間推定

「母平均値が, B 円以上 C 円以下である '確率' は $1-\alpha$」

ここで '確率' というのは不誠実.

「母平均値の信頼係数 $1-\alpha$ の信頼区間は B 円以上 C 円以下」

というのが正しい言葉遣い. 以下でその意味と B, C の求め方.

母平均値の区間推定 (母分散既知)

母平均値 μ, 母分散 σ^2 の母集団から, サイズ n の標本を何回も取り出して, 毎回, 標本平均値 $\overline{X}_{(n)}$ を計算する.

n が大きければ, $U_n = \overline{X}_{(n)}$ は $\mathrm{N}(\mu, \sigma^2/n)$ にしたがう. $Z = \frac{\overline{X}_{(n)}-\mu}{\sqrt{\sigma^2/n}}$ は $\mathrm{N}(0, 1^2)$ にしたがう.

上がいい近似であるためには, 本当は標本サイズ n は 30 以上くらい必要. また, 母集団が正規分布にしたがうなら n によらず厳密に正しい.

母平均値から遠く外れる, 確率 $\alpha = 0.05$ の例外的事象が起きない確率は

$$P\left(-1.96 < \frac{\overline{X}_{(n)} - \mu}{\sqrt{\sigma^2/n}} < +1.96\right) = 0.95.$$

$$P\left(\mu - 1.96 \times \sqrt{\sigma^2/n} < \overline{X}_{(n)} < \mu + 1.96 \times \sqrt{\sigma^2/n}\right) = 0.95.$$

μ について不等式を解くと,

$$P\left(\overline{X}_{(n)} - 1.96 \times \sqrt{\sigma^2/n} < \mu < \overline{X}_{(n)} + 1.96 \times \sqrt{\sigma^2/n}\right) = 0.95.$$

区間推定と信頼区間

母平均値 μ の 信頼係数 $1-\alpha=0.95$ の信頼区間 (95%信頼区間) は

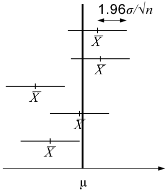
$$\overline{X}_{(n)}-1.96\times\sqrt{\sigma^2/n}<\mu<\overline{X}_{(n)}+1.96\times\sqrt{\sigma^2/n}$$

母平均値 μ の 信頼係数 $1-\alpha=0.99$ の信頼区間 (99%信頼区間) はは

$$\overline{X}_{(n)}-2.58\times\sqrt{\sigma^2/n}<\mu<\overline{X}_{(n)}+2.58\times\sqrt{\sigma^2/n}$$

と

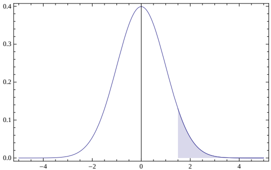
き, 信頼区間が μ を含む確率は 0.95 or 0.99.



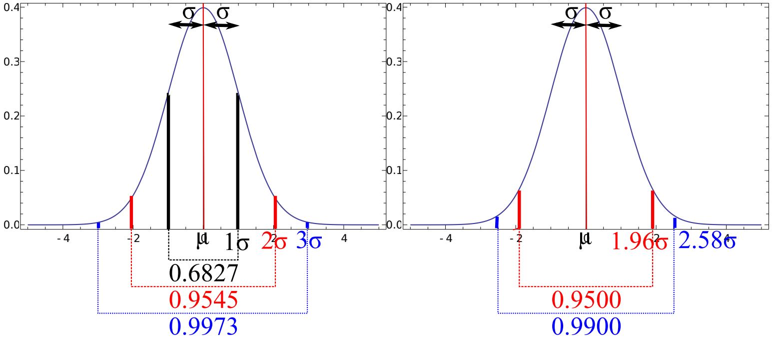
標準正規確率表 (上側確率$=Q(z) = 1 - F(z)$)

z に対する $Q(z) = P(Z > z) = 1 - F(z)$ の値の表.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010



正規分布 (ガウス分布) のグラフに関係した面積



L11-Q1

Quiz(母平均値の区間推定 (母分散既知))

あるドーナツ製造マシンが次々に製造するドーナツの重さ X_ig は, 独立同分布にしたがう確率変数である. あらかじめ行った調査により, X_i の母分散は $\sigma^2 = 9\mathrm{g}^2$ であることがわかっている.
製造された 4 個のドーナツの重さを測定したところ, 次のようだった.
$51\mathrm{g}, 52\mathrm{g}, 47\mathrm{g}, 50\mathrm{g}.$

1. 母平均値 $\mu = \mathrm{E}[X_i]$ を, 信頼係数 $1 - \alpha = 0.95$ で区間推定しよう.
2. 母平均値 $\mu = \mathrm{E}[X_i]$ を, 信頼係数 $1 - \alpha = 0.99$ で区間推定しよう.

推定が正確であるとは 信頼区間が $\boxed{}$ であること.

Quiz(区間推定の性質)

標本からの母平均値の区間推定について, 正しいのはどれ?

1. 母分散が大きいほど, 信頼区間は大きくなる
2. 標本サイズが大きいほど, 信頼区間は大きくなる
3. 母平均値が大きいほど, 信頼区間は小さくなる
4. 信頼係数が大きいほど, 信頼区間は小さくなる

ここまで来たよ

3. 母集団・標本抽出・推定
4. 母平均値と母比率の区間推定
   - 母平均値の区間推定 (母分散既知の場合:標準正規分布)
   - 母平均値の区間推定 (正規分布にしたがう母分散未知の母集団の場合:t 分布)
   - 母比率の区間推定

母分散 σ^2 なんてわかんないんですけど?

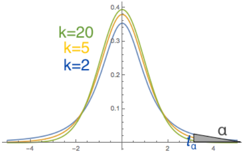
μ はわからないのに σ^2 がわかってるケースはあまりない. ふつうはどちらもわからない.
σ^2 のかわりに不偏標本分散 S^2 (それ自身確率変数) を使っちゃいたい.
母集団が正規分布のときは, 使っちゃた量 $T_{(n)} = \frac{\overline{X}_{(n)}-\mu}{\sqrt{S^2/n}}$ が, 正規分布 $\mathrm{N}(0,1^2)$ からちょっとずれた 自由度 $n-1$ の Student の t 分布にしたがうとわかる.
母集団が厳密に正規分布でなくても近似的に正しいことが多い.
t 分布は,

- 自由度 $k \to +\infty$ で $\mathrm{N}(0,1^2)$ に一致する.
- 自由度 k が小さいとき, $\mathrm{N}(0,1^2)$ より低く広い.

t 分布表

有意水準 α, 自由度 k に対して, $\alpha = P(T > t_\alpha(k))$ となる $t_\alpha(k)$ の値の表.

$k \backslash \alpha$	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.0025	0.001	0.00025
1	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	127.3	318.3	636.6
2	0.816	1.080	1.386	1.886	2.920	4.303	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	2.871	3.174	3.390
$+\infty$	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291



L11-Q2

Quiz(母平均値の区間推定 (母分散未知))

あるドーナツ製造マシンが次々に製造するドーナツの重さ Xg は, 独立同分布にしたがう確率変数である

製造された 4 個のドーナツの重さを測定したところ, 次のようだった.

$51\mathrm{g}, 52\mathrm{g}, 47\mathrm{g}, 50\mathrm{g}.$

1. 母平均値 $\mu = \mathrm{E}[X]$ を, 信頼係数 $1 - \alpha = 0.95$ で区間推定しよう.
2. 母平均値 $\mu = \mathrm{E}[X]$ を, 信頼係数 $1 - \alpha = 0.99$ で区間推定しよう.

母平均値の区間推定 (母分散未知, 大標本) I

自由度 $n-1$ が大きいとき, t 分布のかわりに $\mathrm{N}(0,1^2)$ を使っても大した誤差じゃない. もともと有限の n で中心極限定理を使ってる. 物理実験 L11-Q3

Quiz(母平均値の区間推定 (母分散未知, 大標本))

あるドーナツ製造マシンが次々に製造するドーナツの重さ X_ig は, 独立同分布にしたがう確率変数である.
製造された 400 個のドーナツの重さを測定したところ, 次のようだった.
$51\mathrm{g}, 52\mathrm{g}, 47\mathrm{g}, \ldots, 50\mathrm{g}$.
ここから標本平均値, 不偏標本分散を計算したところ, $m = 51\mathrm{g}, s^2 = 4\mathrm{g}^2$ だった.

1. 母平均値 $\mu = \mathrm{E}[X_i]$ を, 信頼係数 $1-\alpha = 0.95$ で区間推定しよう.
2. 母平均値 $\mu = \mathrm{E}[X_i]$ を, 信頼係数 $1-\alpha = 0.99$ で区間推定しよう.

ここまで来たよ

3. 母集団・標本抽出・推定

4. 母平均値と母比率の区間推定
   - 母平均値の区間推定 (母分散既知の場合:標準正規分布)
   - 母平均値の区間推定 (正規分布にしたがう母分散未知の母集団の場合:t 分布)
   - 母比率の区間推定

母比率の区間推定

「…である」のダミー変数 Y

$$f_y = \begin{cases} p & (y = 1, \text{…である}) \\ 1 - p & (y = 0, \text{…である以外}) \end{cases}$$

$\mathrm{E}[Y] = p =$ 母比率, $\mathrm{V}[Y] = p(1-p)$.
標本比率 $\hat{p}$ から, $\hat{p}(1-\hat{p})$ が母分散であるかのようにして, 標準正規分布の場合の式を使う.

母比率の区間推定

母比率の信頼係数 $1 - \alpha = 0.95$ の信頼区間は

$$\hat{p} - 1.96 \times \sqrt{\frac{1}{n}\hat{p}(1-\hat{p})} < p < \hat{p} + 1.96 \times \sqrt{\frac{1}{n}\hat{p}(1-\hat{p})}$$

母比率の信頼係数 $1 - \alpha = 0.99$ の信頼区間は

$$\hat{p} - 2.58 \times \sqrt{\frac{1}{n}\hat{p}(1-\hat{p})} < p < \hat{p} + 2.58 \times \sqrt{\frac{1}{n}\hat{p}(1-\hat{p})}$$

導出

標本平均値 $\frac{K}{n}$ は, n が大きいとき, 中心極限定理から, 母平均値 p, 母分散 $\frac{1}{n}p(1-p)$ の正規分布にしたがう. よって,

$$p - 1.96 \times \sqrt{\frac{1}{n}p(1-p)} < \hat{p} < p + 1.96 \times \sqrt{\frac{1}{n}p(1-p)}$$

これを, p に関する条件に書き替えたい

とても小さい標本 中心極限定理はやめて, 2 項分布と思って扱え

小標本 p についての 2 次方程式を解け

大標本 p と $\hat{p}$ はかなり近いはず. 最左右辺の $p(1-p)$ は $\hat{p}(1-\hat{p})$ に置きかえちゃっていい.

$$\hat{p} - 1.96 \times \sqrt{\frac{1}{n}\hat{p}(1-\hat{p})} < p < \hat{p} + 1.96 \times \sqrt{\frac{1}{n}\hat{p}(1-\hat{p})}$$

L11-Q4

Quiz(母比率の区間推定)

選挙で出口調査をしたところ, 50 人中 35 人が A 候補に投票したと答えた. 母比率, すなわち有権者全体での A 候補の得票率を考える.

1. A 候補の得票率を, (点) 推定しよう
2. A 候補の得票率を, 信頼係数 $1-\alpha=0.95$ で区間推定しよう.
3. A 候補の得票率を, 信頼係数 $1-\alpha=0.99$ で区間推定しよう.

注: 下限, 上限が 0,1 を越えるときは, 0,1 に直してしまっていい.

連絡

- 配布資料は 1-503 向かいの引出, `http://hig3.net` で再配布.
- 週のタイムラインで見たように, 非参照 Quiz 予習問題を RaMMoodle に金 17:00 ごろまでに公開. これで来週の Quiz に備えてね.
- **2015-12-16 水 4 学外実習 (インターンシップ) 履修説明会**
- Math ラウンジの配布資料訂正. 正しくは, 2015-12-23 水:授業なし, 2015-12-24 木:土曜授業.
- オフィスアワー月 4 木 6(1-502)

manaba / 出席カード
`https://manaba.ryukoku.ac.jp`