

統計学

北門 利英 (海洋生物資源学科)

Lecture 7 正規分布とその性質

Attention: 授業に関してわからないことがあれば、メールか6/24のリアルタイム接続時(13:00-14:00)に遠慮なく質問してください。

Point: 次の用語をしっかりと理解すること。

- 正規分布におけるパラメータの推定法について理解する事
- 推定の不偏性について再確認する事

T. Kitakado Statistics Lecture 07, June 24, 2020

生物資源学科の北門です。このビデオは統計学のLecture7の要約版です。今回は、正規分布におけるパラメータの推定についてお話しします。前回と今回の内容は、Lecture 8と9の基礎となりますので、必ず習得するように頑張ってください。それでは始めます。

正規分布の復習

T. Kitakado Statistics Lecture 07, June 24, 2020

まずは正規分布の復習です。

正規分布の定義

連続型分布 1 確率変数 Y の確率密度関数が

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}, \quad -\infty < y < \infty \quad \left(\begin{array}{l} -\infty < \mu < \infty \\ 0 < \sigma < \infty \end{array} \right)$$

となるとき, Y は正規分布 (normal distribution) $N(\mu, \sigma^2)$ にしたがうという。

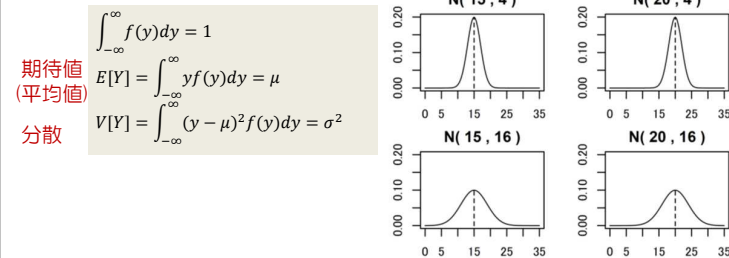


図4 正規分布の確率密度関数の概形

正規分布は連続型の確率分布で, 確率密度関数がこの式のように表される場合でした。muは期待値, sigma2は分散を意味しています。また確率密度関数の概形は平均値に対して左右対称で, 分散sigma2が大きくなると図でもわかりますように平均値の周りでの散らばりが大きくなります。

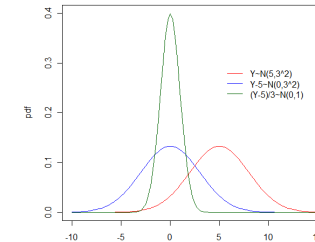
正規分布に関する性質 (1) 標準正規分布 $N(0,1)$

性質 1 確率変数 Y が $N(\mu, \sigma^2)$ にしたがうとき,

$$Z = \frac{Y - \mu}{\sigma} \quad (5)$$

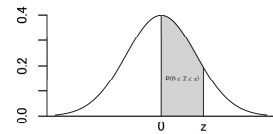
は $N(0,1)$ にしたがう。この変換を標準化 (standardization), また $N(0,1)$ を標準正規分布 (standard normal distribution) という。

$$\begin{aligned} Y &\sim N(\mu, \sigma^2) \\ Y - \mu &\sim N(0, \sigma^2) \\ Z = \frac{Y - \mu}{\sigma} &\sim N(0,1) \end{aligned}$$



正規分布に関係する大切な性質として標準化という話をしました。これは, 確率変数 Y が平均 μ , 分散 σ^2 の正規分布に従っているとき, この Y から平均値 μ を引いて, 標準偏差 σ で割った値を Z と置くと, もとものの μ と σ が何であっててもそれを使って変換すれば平均が 0, 分散が 1 の正規分布に変換できるということでした。このような変換を標準化, また平均 0, 分散 1 の正規分布を標準正規分布と呼びます。

正規分布に関する性質 (1) 標準正規分布 $N(0,1)$



$$Z \sim N(0,1)$$

$$P(0 \leq Z \leq 1.96) = 0.4750$$

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
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	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
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	0.4406	0.4418	0.4429	0.4441
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

T. Kitakado

Statistics

Lecture 07, June 24, 2020

この標準正規分布には数表が与えられていて、これを用いた計算の練習問題に前回取り組まれたかと思います。今回の提出課題でも利用します。

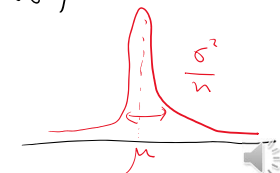
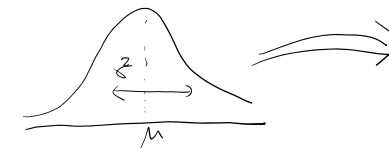
正規分布に関する性質 (2) 再生性

正規分布にしたがう独立な確率変数の和もまた正規分布にしたがう。独立同一な n 個の正規分布にしたがう確率変数の和もまた正規分布にしたがう。これらの性質は区間推定と仮説検定のところで再度説明するので、今回は定理の存在だけ覚えておいてください。

定理 2 確率変数 Y_1, Y_2 が独立でそれぞれ $N(\mu_1, \sigma_1^2), N(\mu_2, \sigma_2^2)$ にしたがうとき、 $aY_1 + bY_2$ は $N(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$ にしたがう。

定理 3 確率変数 $Y_1, Y_2, \dots, Y_n$ が独立同一に $N(\mu, \sigma^2)$ にしたがうとき、 $\bar{Y} = \sum_{i=1}^n Y_i/n$ は $N(\mu, \sigma^2/n)$ にしたがう。

$$Y_1, Y_2, \dots, Y_n \sim \text{i.i.d. } N(\mu, \sigma^2)$$



T. Kitakado

Statistics

Lecture 07, June 24, 2020

もう一つの大変な性質として、正規分布の再生性があげられます。この定理3を覚えてください。これは n 個の独立同一な正規分布に従う確率変数を考えるとき、それを1次元にまとめた標本平均もまた正規分布に従い、その期待値が μ 、分散は元の分散の n 分の1になります。すなわち、標本平均にまとめるとちがりが小さくなることを意味します。

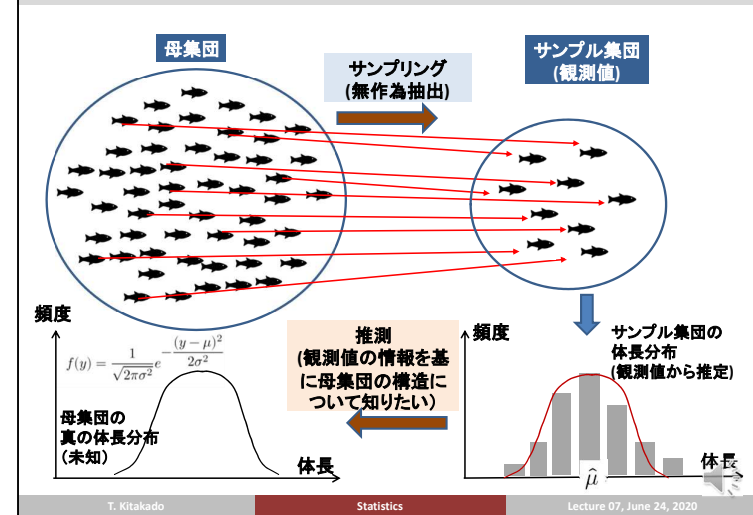
パラメータの推定

T. Kitakado

Statistics

Lecture 07, June 24, 2020

母集団とサンプル集団（観測値）



T. Kitakado

Statistics

Lecture 07, June 24, 2020

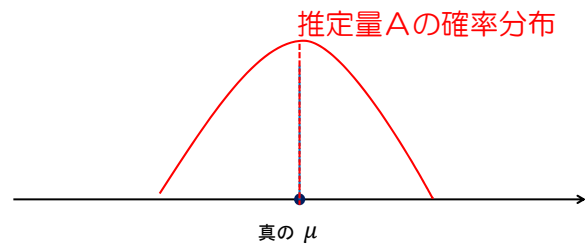
これは前回用いた模式図ですが、いま何か知りたい対象となる母集団があり、それが正規分布に分布に従っているとします。そこから無作為にサンプリングを行って観測値を得て、その観測値を基に母集団のパラメータを推定します。

2つの評価側面

(1) 不偏性

真のパラメータの値がなんであろうとも、
推定量の期待値が真値と一致する

$$E[\hat{\mu}(Y)] = \mu \text{ for all } \mu$$



T. Kitakado

Statistics

Lecture 07, June 24, 2020

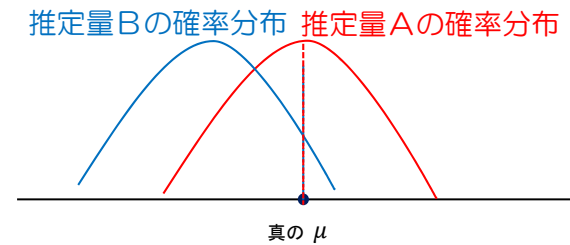
さて、その推定をする際に利用する計算方法、すなわち推定量の性質を評価する方法ですが、ここでは2点あげたいと思います。一つは不偏性と呼ばれるもので、これまでの授業でも触れてきましたが改めてまとめてみます。我々は一部を調べて全体を把握しようとするわけですが、一部しか見ていないので必ず誤差が生じます。例えば正規分布の平均値 μ を推定したい時に、その値を間違えることなく推定することはできず、ばらつきが生じます。それも確率分布として表せるわけですが、ここでは推定量Aを用いると、真の μ に対してこのように偏りなく散らばるとします。そして μ -hatの期待値をとると、ちゃんと μ になっているとします。これはよさそうな性質ですね。

2つの評価側面

(1) 不偏性

真のパラメータの値がなんであろうとも、
推定量の期待値が真値と一致する

$$E[\hat{\mu}(Y)] = \mu \text{ for all } \mu$$



T. Kitakado

Statistics

Lecture 07, June 24, 2020

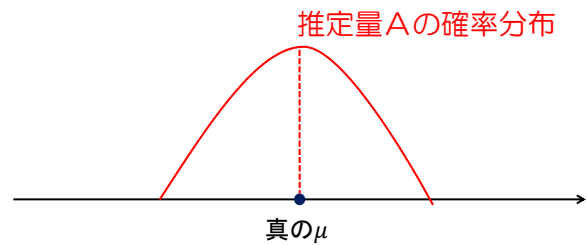
いまこれとは別に推定量Bというのがあったとしましょう。これもばらつきわけですが、青の線のようなばらつきだったとしましょう。これはなんとなくよくなさそうですね。ばらつきの程度は同じでも、過小評価する傾向があります。期待値をとると μ にはなりませんね。偏ってしまっているのです。

2つの評価側面

(1) 不偏性

真のパラメータの値がなんであろうとも、
推定量の期待値が真値と一致する

$$E[\hat{\mu}(Y)] = \mu \text{ for all } \mu$$



T. Kitakado

Statistics

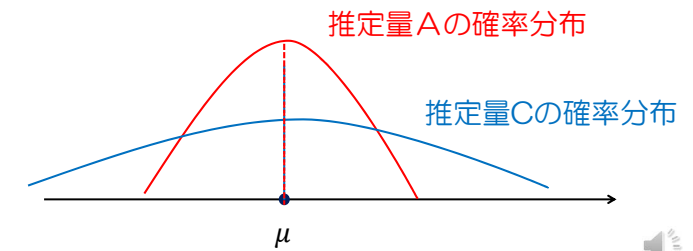
Lecture 07, June 24, 2020

ここで大事なのは、本当の値を100%的中させることはできないのですが、偏りなく推定できるという性質が大事で、これは真の μ の値が何であっても、満たされることが大切です。すなわち μ が変わればちゃんと推定量もそれに応じて変わる、という性質です。このような推定量の性質のことを不偏性と言います。

2つの評価側面

(2) 分散をなるべく小さくする

$$V[\hat{\mu}(Y)] \rightarrow \min$$



T. Kitakado

Statistics

Lecture 07, June 24, 2020

次に不偏性を満たしている2つの推定量AとCがあったとして、これだとどちらがよいでしょうか。一目瞭然、AもCも μ を中心にばらつくわけですが、ばらつきが小さいほうがよい、すなわち推定量Aの方がよいことになります。

定式化 (μ の推定)

$$Y_1, Y_2, \dots, Y_n \sim (iid) N(\mu, \sigma^2)$$

$$\hat{\mu}(Y) = \bar{Y} = \frac{1}{n}(Y_1 + \dots + Y_n)$$

推定量の不偏性

$$E[\hat{\mu}(Y)] = E\left[\frac{1}{n}(Y_1 + \dots + Y_n)\right] = \frac{1}{n}(E[Y_1] + \dots + E[Y_n]) = \mu$$

推定量の分散

$$V[\hat{\mu}(Y)] = V\left[\frac{1}{n}(Y_1 + \dots + Y_n)\right] = \frac{1}{n^2}(V[Y_1] + \dots + V[Y_n]) = \frac{\sigma^2}{n}$$

T. Kitakado

Statistics

Lecture 07, June 24, 2020



さて正規分布の場合の平均値 μ の推定ですが、標本平均を用います。この推定量は、ここに示していますように不偏で、また証明は省略しますが分散を σ^2/n よりも小さくする不偏な推定量は存在しないことが示せます。

定式化(σ^2 の推定)

$$Y_1, Y_2, \dots, Y_n \sim (iid) N(\mu, \sigma^2)$$

$$\hat{\sigma}^2(Y) = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2$$

推定量の不偏性(不偏であることだけを覚えてください)

$$E[\hat{\sigma}^2(Y)] = E\left[\frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2\right] = \frac{1}{n-1}(n-1)\sigma^2 = \sigma^2$$

推定量の分散(これは覚えなくてよいです)

$$V[\hat{\sigma}^2(Y)] = \frac{2\sigma^4}{n-1}$$

T. Kitakado

Statistics

Lecture 07, June 24, 2020



また分散ですが、各観測値と標本平均の差の2乗を平均するわけですが、ここで n ではなく $(n-1)$ で割っているのは、その下の式に示したように、そうすることで不偏性を満たすことができるからです。

例題

ある海域に生息する××マグロの1歳魚を9個体ランダムにサンプリングし、その体長を測定した。これらが以下のように独立同一に正規分布 $N(\mu, \sigma^2)$ に従うとすると、 μ と σ^2 の推定値を求めよ。また、 μ の推定値の標準誤差を求めよ。

$Y=(42, 40, 50, 39, 40, 43, 42, 48, 43)$



例題

$Y_1, Y_2, \dots, Y_n \sim (iid)N(\mu, \sigma^2)$

$$\hat{\mu} = \bar{y} = \frac{1}{9}(42 + 40 + 50 + 39 + 40 + 43 + 42 + 48 + 43) = 43$$

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{n-1} \sum_{i=1}^n (y_i - \hat{\mu})^2 \\ &= \frac{1}{8} [(42 - 43)^2 + (40 - 43)^2 + \dots + (43 - 43)^2] = 13.75\end{aligned}$$

$\hat{\mu}$ の標準誤差 (推定量の標準偏差に値を代入したもの)

$$SE[\hat{\mu}] = \frac{\hat{\sigma}}{\sqrt{n}} = \frac{\sqrt{13.75}}{\sqrt{9}} = 1.236$$



提出課題

前回の Lecture 6 と今回の Lecture 7 の課題と併せて一つの課題とします。計算式と答えを書いて、必ずまとめて 1 つのファイルで提出してください。ワードで数式の書けない人は、手書きの計算結果を写真に撮ってワードに張り付けて下さっても構いません。答だけの回答は大きく減点します。提出期限は 7 月 3 日 23:59 とします。

Lecture06-HW1 (再掲) あるチョコレート工場で 100g の板チョコを生産しているが、製品によってばらつきが生じ、正規分布 $N(102, 2^2)$ にしたがうとされている。100g 未満のチョコレートは出荷できないとき、生産したチョコレートの何%が不良品となるか？また、不良品率を 1 パーセント以下にしたいとき、板チョコの重さの平均値をいくらにするように生産工程を変えればよいか？(平均値を変えてもばらつき、すなわち分散は変わらないとする)

Lecture07-HW1 ある航空会社の XX 路線の予定飛行時間は 100 分とされているが、過去 10 日間の記録をみると 105, 98, 110, 99, 95, 112, 120, 93, 102, 106 (分) であった。これらが独立同一に正規分布 $N(\mu, \sigma^2)$ にしたがうとすると、期待値 μ と分散 σ^2 の推定値を求めよ。また、 μ の推定値の推定誤差を求めよ。さらに、推定した μ, σ^2 の値を基にして、この路線の飛行機が予定飛行時間よりも遅れる確率も求めよ (前回の標準正規分布表を用いてください)。



T. Kitakado

Statistics

Lecture 07, June 24, 2020

最後に提出課題です。7月3日までに学務システムを通して提出してください。それでは Lecture7のダイジェストビデオを終わります。