九州大学 工学部地球環境工学科
船舶海洋システム工学コース

海事統計学　第９回 （担当：木村）

標本分布２：正規分布の区間推定

授業の資料および演習・例題等の解答は

http://sysplan.nams.kyushu-u.ac.jp/gen/index.html

標本から正規分布の母数に関する情報を得る（1）

$x_1, x_2, \cdots, x_n, \cdots$ が互いに独立確率変数で、 それらが全て

期待値 μ 分散 σ^2 の**正規分布**に従うとき、

平均値 $\bar{x} = \dfrac{x_1 + x_2 + \cdots + x_n}{n}$ の分布は

正規分布 $N(\mu, \dfrac{\sigma^2}{n})$ になる

未知の期待値μの範囲を推定できる？

すなわち $z = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}}$ は標準正規分布 N(0,1) になる

正規分布表から読み取る

標準正規分布 z が -1.96 と +1.96 の間の値をとる確率は 95% つまり

$$\Pr\left\{ \qquad \right\} = 0.95$$

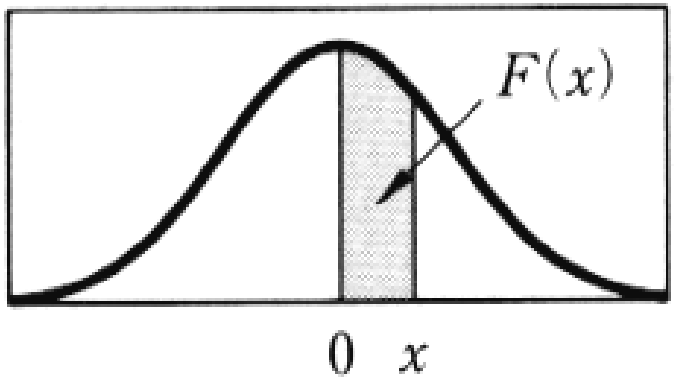
$$\Pr\left\{ \qquad \right\} = 0.95$$

分散σは既知でなければならない

期待値μがこの範囲である可能性は 95%

このままでは現実的には使えない

正規分布表



カゲの部分の確率 $F(x)$ を示す。

$$0.95\times\frac{1}{2}=0.4750$$

x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998

標本から正規分布の母数に関する情報を得る（1）

$x_1, x_2, \cdots, x_n, \cdots$ が互いに独立確率変数で、 それらが全て期待値 μ 分散 σ^2 の**正規分布**に従うとき、

平均値 $\bar{x} = \dfrac{x_1 + x_2 + \cdots + x_n}{n}$ の分布は

正規分布 $N(\mu, \dfrac{\sigma^2}{n})$ になる

未知の期待値μの範囲を推定できる？

すなわち $z = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}}$ は標準正規分布 N(0,1) になる

正規分布表から読み取る

標準正規分布 z が -1.96 と +1.96 の間の値をとる確率は 95% つまり

$$\Pr\left\{-1.96 < \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} < 1.96\right\} = 0.95$$

標準偏差σは既知でなければならない

このままでは現実的には使えない

$$\Pr\left\{\underline{\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}}\right\} = 0.95$$

期待値μがこの範囲である可能性は 95%

標本から正規分布の母数に関する情報を得る（2）

平均値 $\dfrac{x_1 + x_2 + \cdots + x_n}{n}$ の分布について標準化した分布 $z = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}}$

分散σ^{2}は未知なので n個の標本からの推定値$\hat{\sigma}^2$で代用

ただし

不偏分散

データ数 n に対して n-1 であることに注意

このとき、分布 $t = \dfrac{\bar{x} - \mu}{\hat{\sigma} / \sqrt{n}}$ は**データ数 n に依存**した関数 **正規分布にならない**

自由度 n-1 の に従う

自由度 n の t分布の確率密度関数 $$f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$

n→∞の極限では 正規分布になる

$\Gamma(x)$:ガンマ関数

標本から正規分布の母数に関する情報を得る（2）

平均値 $\dfrac{x_1+x_2+\cdots+x_n}{n}$ の分布について標準化した分布 $z=\dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}}$

分散σ²は未知なので
n個の標本からの推定値$\hat{\sigma}^2$で代用

ただし

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

不偏分散

データ数 n に対して
n-1 であることに注意

このとき、分布 $t=\dfrac{\bar{x}-\mu}{\hat{\sigma}/\sqrt{n}}$ は**データ数 n に依存**した関数
正規分布にならない

自由度 n-1 の **t分布（t-distribution）** に従う

自由度 n の
t分布の確率密度関数

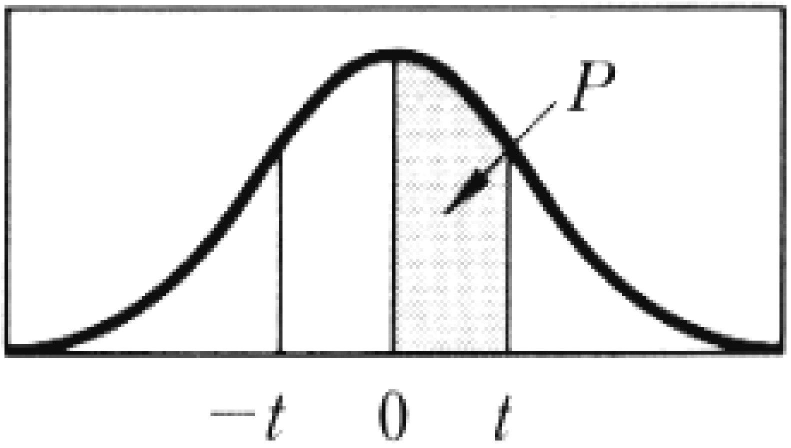
$$f(x)=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$

$\Gamma(x)$:ガンマ関数

n→∞の極限では
正規分布になる

t 分布表

自由度
（Degree of freedom）
(n-1)



カゲの部分の確率 P に対する t の値を示す。

$d.f.$ \ P	.25	.40	.45	.475	.49	.495	.4995
1	1.000	3.078	6.314	12.706	31.821	63.657	636.619
2	.816	1.886	2.920	4.303	6.965	9.925	31.598
3	.765	1.638	2.353	3.182	4.541	5.841	12.941
4	.741	1.533	2.132	2.776	3.747	4.604	8.610
5	.727	1.476	2.015	2.571	3.365	4.032	6.859
6	.718	1.440	1.943	2.447	3.143	3.707	5.959
7	.711	1.415	1.895	2.365	2.998	3.499	5.405
8	.706	1.397	1.860	2.306	2.896	3.355	5.041
9	.703	1.383	1.833	2.262	2.821	3.250	4.781
10	.700	1.372	1.812	2.228	2.764	3.169	4.587
11	.697	1.363	1.796	2.201	2.718	3.106	4.437
12	.695	1.356	1.782	2.179	2.681	3.055	4.318
13	.694	1.350	1.771	2.160	2.650	3.012	4.221
14	.692	1.345	1.761	2.145	2.624	2.977	4.140
15	.691	1.341	1.753	2.131	2.602	2.947	4.073
16	.690	1.337	1.746	2.120	2.583	2.921	4.015
17	.689	1.333	1.740	2.110	2.567	2.898	3.965
18	.688	1.330	1.734	2.101	2.552	2.878	3.922
19	.688	1.328	1.729	2.093	2.539	2.861	3.883
20	.687	1.325	1.725	2.086	2.528	2.845	3.850
21	.686	1.323	1.721	2.080	2.518	2.831	3.819
22	.686	1.321	1.717	2.074	2.508	2.819	3.792
23	.685	1.319	1.714	2.069	2.500	2.807	3.767
24	.685	1.318	1.711	2.064	2.492	2.797	3.745
25	.684	1.316	1.708	2.060	2.485	2.787	3.725
26	.684	1.315	1.706	2.056	2.479	2.779	3.707
27	.684	1.314	1.703	2.052	2.473	2.771	3.690
28	.683	1.313	1.701	2.048	2.467	2.763	3.674
29	.683	1.311	1.699	2.045	2.462	2.756	3.659
30	.683	1.310	1.697	2.042	2.457	2.750	3.646
40	.681	1.303	1.684	2.021	2.423	2.704	3.551
60	.679	1.296	1.671	2.000	2.390	2.660	3.460
120	.677	1.289	1.658	1.980	2.358	2.617	3.373
∞	.674	1.282	1.645	1.960	2.326	2.576	3.291

【参考：ガンマ関数】

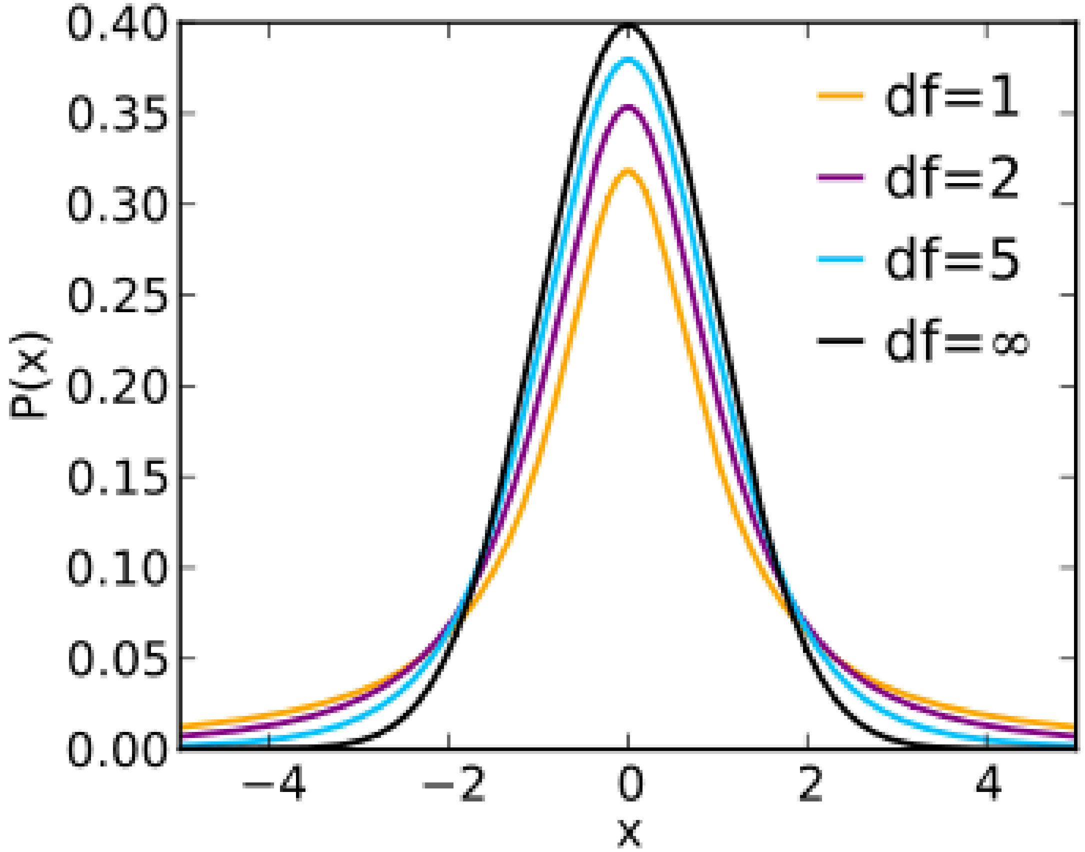
$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt$$

$$\Gamma(1) = 1 \quad \Gamma(2) = 1 \quad \Gamma(3) = 2$$

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

x が正の整数のとき

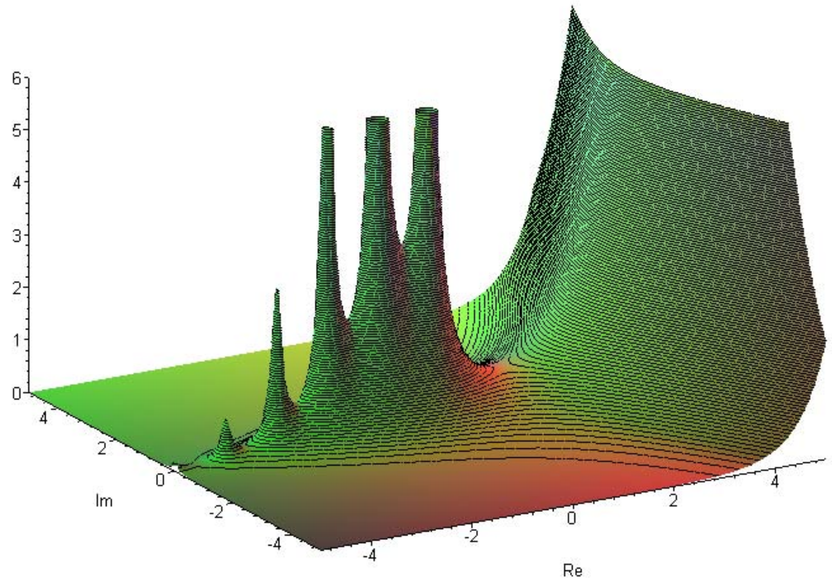
$$\Gamma(x) = (x-1)!$$



t分布のグラフ

【参考】ガンマ関数

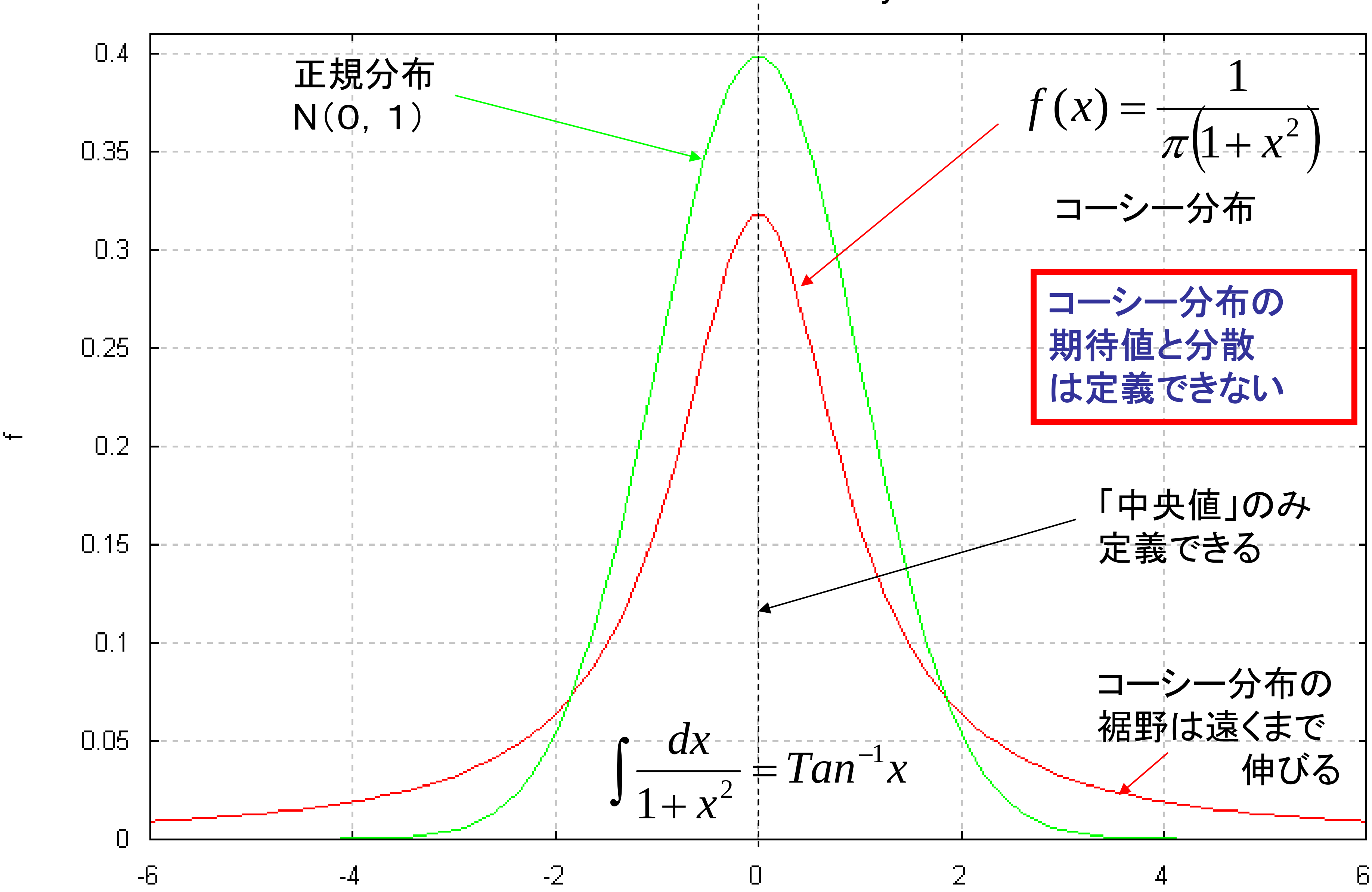
$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$$



参考文献：ウィキペディア

【参考】 自由度1のt分布 ＝ 「コーシー分布」

（Cauchy distribution）



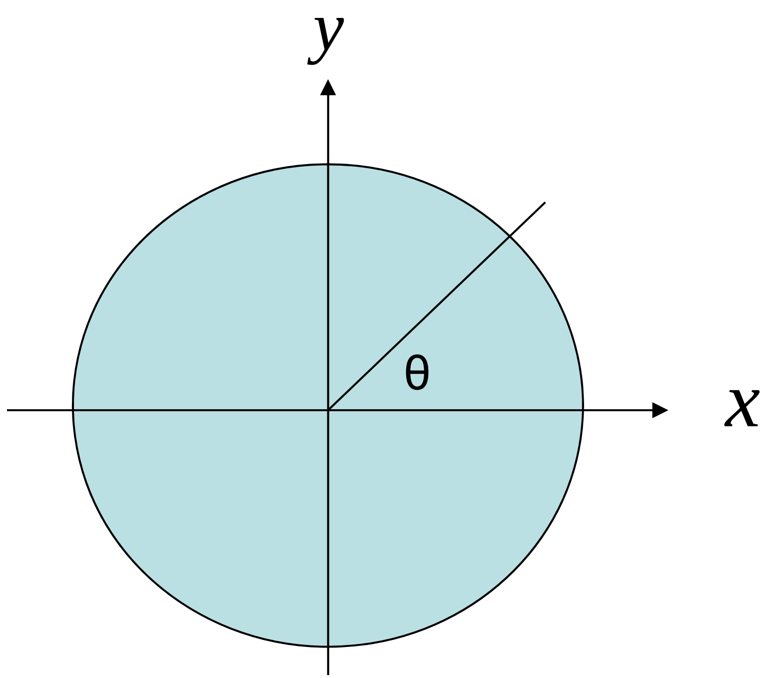
コーシー分布の期待値と分散は定義できない

⇒

中心値極限定理が成り立たない分布

【意外と身近なコーシー分布】

θ が（ - π / 2, π / 2 ）上の一様分布に従うとき, tan (θ) は **コーシー分布** になる。



【t 分布のまとめ】

標本数 n+1

自由度 n の
t分布の確率密度関数

$$f(x)=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$

ただし　$\Gamma(x)$　:ガンマ関数

n ＝1　のとき
コーシー分布

コーシー分布の
確率密度関数

$$f(x)=\frac{1}{\pi\left(1+x^2\right)}$$

コーシー分布の
期待値と分散
は定義できない

→ 中心値極限定理が
成り立たない特殊な分布

n→∞の極限では
正規分布になる

→ 標本数 n+1 が30程度までなら t分布表を利用

→ 標本数 n+1 が40以上なら正規分布表を利用

標本から正規分布母数に関する情報を得る（まとめ）

同一な**正規分布**の独立な試行で得られた n 個のデータ $x_1, x_{2,} \cdots, x_n, \cdots$ について、

平均値 $\bar{x} = \dfrac{x_1 + x_2 + \cdots + x_n}{n}$ と分散の推定値 $\boxed{\phantom{\hat{\sigma}^2}}$ を計算する

平均値 $\bar{x}$ の標準化を試みた確率変数 $t = \dfrac{\bar{x} - \mu}{\hat{\sigma}/\sqrt{n}}$ は、自由度n-1のt分布に従う

データ数 n に対して n-1 であることに注意

例） データ数 n =10 のとき、自由度9

t分布表より、自由度 9 の t が -2.262 と +2.262 の間の値をとる確率は 95% つまり

$$\Pr\left\{-2.262 < \frac{\bar{x} - \mu}{\hat{\sigma}/\sqrt{n}} < 2.262\right\} = 0.95$$

$$\Pr\left\{\bar{x} - 2.262\frac{\hat{\sigma}}{\sqrt{10}} < \mu < \bar{x} + 2.262\frac{\hat{\sigma}}{\sqrt{10}}\right\} = 0.95$$

t分布を用いれば、母集団標準偏差σが不明でも母集団平均値μについて推論できる

区間推定

$\boxed{}$ → もとの（未知な）正規分布における真の期待値μがこの範囲である可能性は95% → $\boxed{}$

標本から正規分布母数に関する情報を得る（まとめ）

同一な**正規分布**の独立な試行で得られた n 個のデータ　$x_1, x_2, \cdots, x_n, \cdots$　について、

平均値 $\overline{x} = \dfrac{x_1 + x_2 + \cdots + x_n}{n}$ と分散の推定値 $\boxed{\hat{\sigma}^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}$ を計算する

平均値 $\overline{x}$ の標準化を試みた確率変数 $t = \dfrac{\overline{x} - \mu}{\hat{\sigma}/\sqrt{n}}$ は、自由度n-1のt分布に従う

データ数 n に対して
n-1 であることに注意

例）　データ数 n =10 のとき、自由度9

t分布表より、自由度 9 の t が -2.262　と +2.262
の間の値をとる確率は 95%　つまり

$$\Pr\left\{-2.262 < \frac{\overline{x} - \mu}{\hat{\sigma}/\sqrt{n}} < 2.262\right\} = 0.95$$

$$\Pr\left\{\overline{x} - 2.262\frac{\hat{\sigma}}{\sqrt{10}} < \mu < \overline{x} + 2.262\frac{\hat{\sigma}}{\sqrt{10}}\right\} = 0.95$$

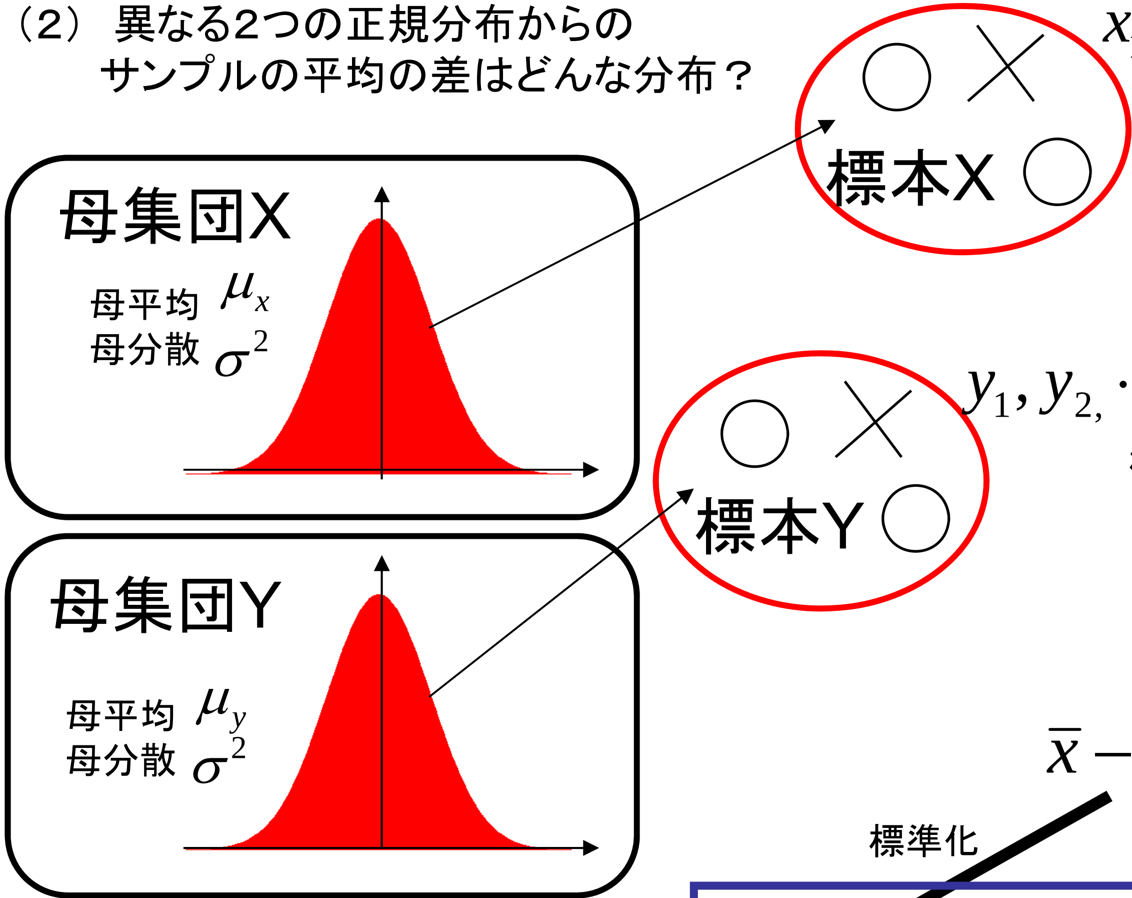
t分布を用いれば、母集団標準偏差σが不明でも母集団平均値μについて推論できる

区間推定

もとの（未知な）正規分布における真の期待値μが
信頼区間 → この範囲である可能性は 95%　→ **信頼係数**

正規母集団の母平均の差の分布

（2） 異なる2つの正規分布からの
サンプルの平均の差はどんな分布？



$x_1, x_2, \cdots, x_m$

標本平均 $\bar{x} = \dfrac{x_1 + x_2 + \cdots, + x_m}{m}$

標本不偏分散 $U_x^2 = \dfrac{1}{m-1}\sum_{i=1}^{m}(x_i - \bar{x})^2$

$y_1, y_2, \cdots, y_n$

標本平均 $\bar{y} = \dfrac{y_1 + y_2 + \cdots, + y_n}{n}$

標本不偏分散 $U_y^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$

$\bar{x} - \bar{y}$ の分布は、平均値 $\mu_x - \mu_y$
分散 $\sigma^2\left(\dfrac{1}{m} + \dfrac{1}{n}\right)$

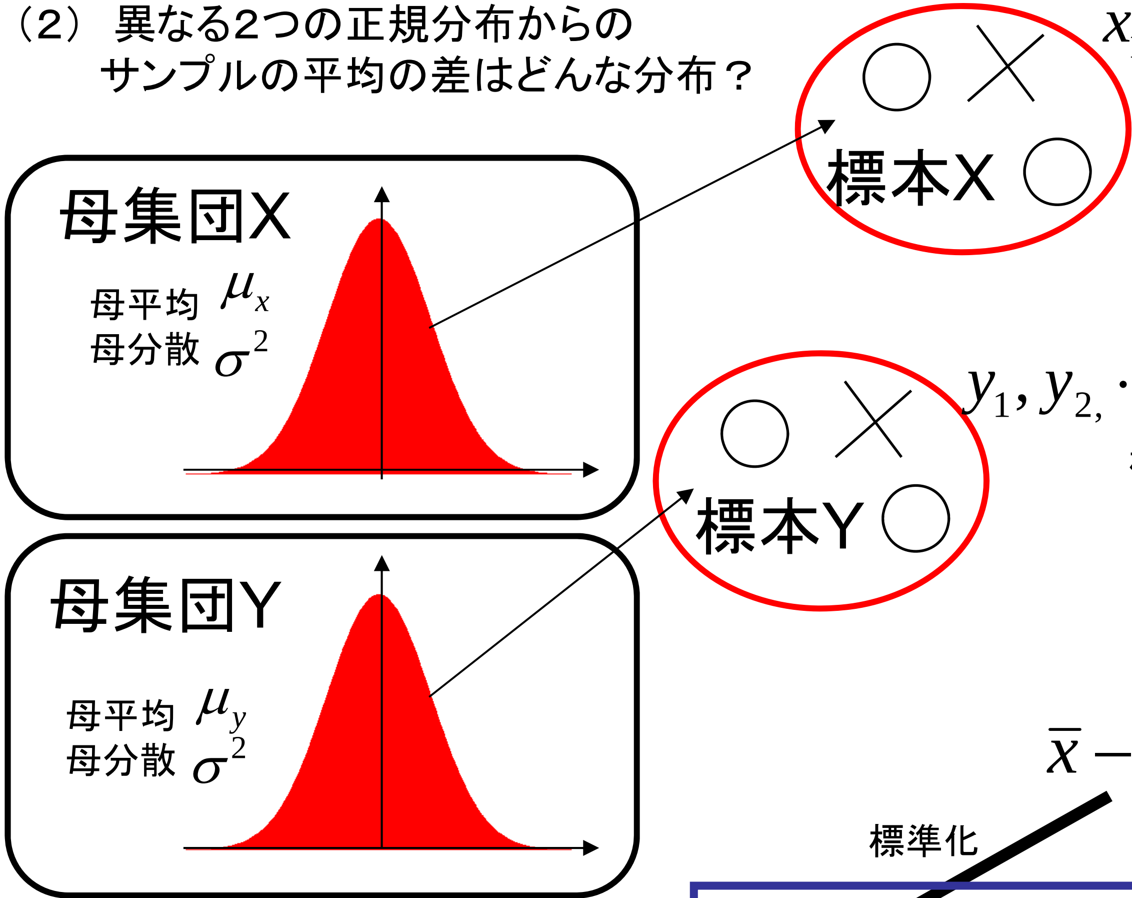
標準化

これを利用することにより、

このとき t は自由度 の t分布

正規母集団の母平均の差の分布

（2） 異なる2つの正規分布からの
サンプルの平均の差はどんな分布？



$x_1, x_2, \cdots, x_m$

標本平均 $\bar{x} = \dfrac{x_1 + x_2 + \cdots, + x_m}{m}$

標本不偏分散 $U_x^2 = \dfrac{1}{m-1}\sum_{i=1}^{m}(x_i - \bar{x})^2$

$y_1, y_2, \cdots, y_n$

標本平均 $\bar{y} = \dfrac{y_1 + y_2 + \cdots, + y_n}{n}$

標本不偏分散 $U_y^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$

$\bar{x} - \bar{y}$ の分布は、平均値 $\mu_x - \mu_y$
分散 $\sigma^2\left(\dfrac{1}{m} + \dfrac{1}{n}\right)$

標準化

$$t = (\bar{x} - \bar{y})\sqrt{\frac{(m+n-2)\,mn}{\left((m-1)U_x^2 + (n-1)U_y^2\right)(m+n)}}$$

これを利用することにより、未知数の
集団Xの母平均と集団Yの母平均に
有意な差があるかどうかを判定できる

このとき t は自由度 $\boxed{m+n-2}$ の t分布

【演習問題】　2018.05.08　学籍番号　　　　　　氏名

同一な正規分布の独立な試行で得られたｎ個のデータ　$x_1, x_{2,} \cdots, x_n$　について、

データ数ｎ＝９　のとき、もとの正規分布の期待値を信頼係数90％で区間推定せよ。

ただし　データの平均を $\overline{X}$　分散の推定値を $\hat{\sigma}^2$　とする。

【演習問題】　　　　　　　　　　　学籍番号　　　　　　　氏名

同一な正規分布の独立な試行で得られたｎ個のデータ　$x_1, x_{2,} \cdots, x_n$　について、

データ数ｎ＝９　のとき、もとの正規分布の期待値を信頼係数９０％で区間推定せよ。

ただし　データの平均を $\overline{x}$　分散の推定値を $\hat{\sigma}^2$　とする。

データ数ｎ＝９のとき、自由度　？

t分布表より、自由度８のｔが $\boxed{}$ と $\boxed{}$

の間の値をとる確率は90%　つまり

$$\Pr\left\{ \boxed{} < \frac{\overline{x} - \mu}{\hat{\sigma} / \sqrt{9}} < \boxed{} \right\} = 0.9$$

$$\Pr\left\{ \underbrace{\overline{x} - \boxed{} < \mu < \overline{x} + \boxed{}}_{\text{期待値の信頼区間}} \right\} = 0.9$$

【演習問題】 学籍番号 氏名

同一な正規分布の独立な試行で得られたｎ個のデータ $x_1, x_{2,} \cdots, x_n, \cdots$ について、

データ数ｎ＝９ のとき、もとの正規分布の期待値を信頼係数９０％で区間推定せよ。

ただし データの平均を $\overline{X}$ 分散の推定値を $\hat{\sigma}^2$ とする。

データ数 n = 9 のとき、自由度 8

t分布表より、自由度 8 の t が -1.860 と +1.860

の間の値をとる確率は 90% つまり

$$\Pr\left\{-1.860 < \frac{\bar{x}-\mu}{\hat{\sigma}/\sqrt{9}} < 1.860\right\} = 0.9$$

$$\Pr\left\{\underbrace{\bar{x}-1.860\frac{\hat{\sigma}}{\sqrt{9}} < \mu < \bar{x}+1.860\frac{\hat{\sigma}}{\sqrt{9}}}_{\text{期待値の信頼区間}}\right\} = 0.9$$