

第7章 (pp. 82)

正規分布とカイ二乗分布

正規分布

(1) 正規分布の確率計算

- 正規分布 （復習）
 - 平均を中心とした左右対称の連続確率分布
- 中心極限定理 （復習）
 - 標本平均は正規分布にしたがう
 - 母集団がどのような分布であっても
標本平均に関しては確率による評価ができる
 - 確率計算は簡単ではない…

$$\Pr(\alpha < X < \beta) = \int_{\alpha}^{\beta} f(x) dx = \int_{\alpha}^{\beta} \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left\{-\frac{(x - \mu_X)^2}{2\sigma_X^2}\right\} dx$$

(2) 標準正規分布

- 標準化（復習）
 - 平均値と標準偏差を統一する方法
 - 平均 = 0
 - 標準偏差 = 1
- 標準化が適用できない場合
 - データに外れ値が含まれる場合
 - 度数分布が左右対称の単峰の分布でない場合
 - 正規分布は左右対称の確率分布で外れ値が生じない形状なので標準化を適用できる！

- 正規分布の標準化
 - 標本平均 \bar{X} を標準化する
 - Z : \bar{X} を標準化した確率変数

$$Z = \frac{\bar{X} - E(\bar{X})}{\sqrt{\text{Var}(\bar{X})}} = \frac{\bar{X} - \mu_X}{\sqrt{\frac{\sigma_X^2}{n}}} = \frac{\text{標本平均} - \text{母平均}}{\sqrt{\frac{\text{母分散}}{\text{標本数}}}}$$

- 標準正規分布

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

- 確率変数 Z （標本平均 \bar{X} を標準化したもの）
- 平均 = 0
- 標準偏差 = 1（分散 = 1）
 - 確率密度関数

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{z^2}{2}\right\}$$

(3) 標準正規分布表の見方

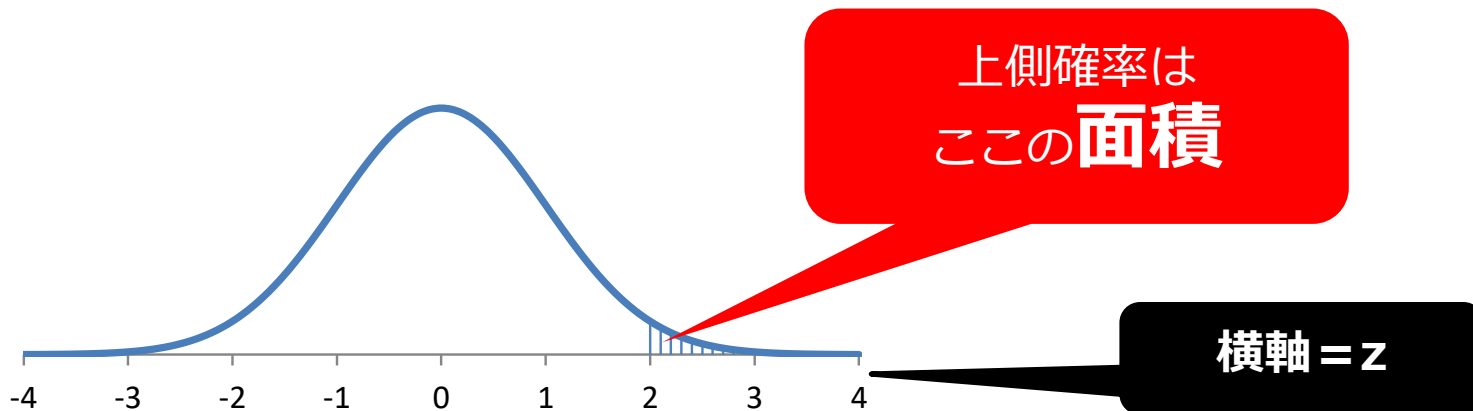
(4) 標準正規分布表に基づく確率計算

標準正規分布表の見方

！超重要！

- 標準正規分布表

- 標準正規分布にしたがう確率変数に関する確率を計算して一覧表にしたもの
- 教科書pp.162（付録2）
- Z の値 z に対応した標準正規分布 $N(0,1)$ の上側確率 $\Pr(Z > z)$ が計算されている

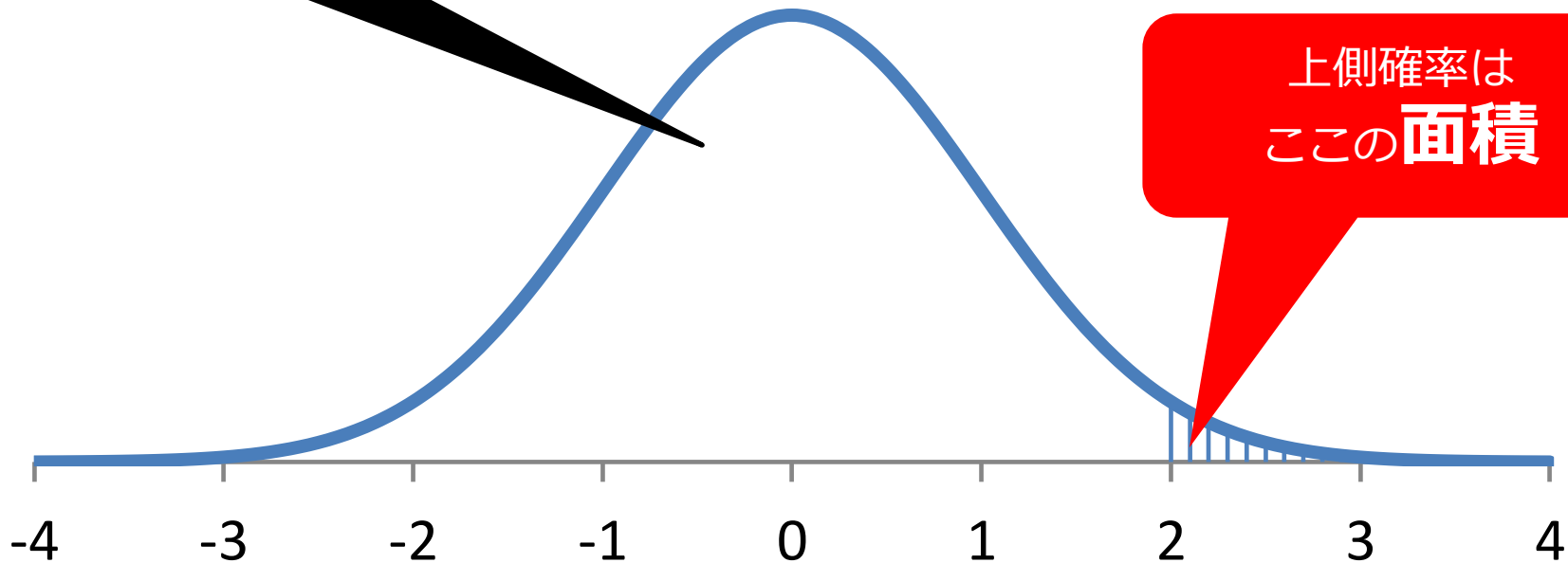


標準正規分布表の見方

！超重要！

確率なので
全体の面積は1

上側確率は
ここの面積



平均（中心）はゼロ

横軸 = z

例) $z=2.00$ のとき

！超重要！

まず縦方向から
2.0を探す

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.49601	0.49202	0.48803	0.48405	0.48006	0.47608	0.47210	0.46812	0.46414
0.1	0.46017	0.45620	0.45224	0.44828	0.44433	0.44038	0.43644	0.43251	0.42858	0.42465
0.2	0.42074	0.41683	0.41294	0.40905	0.40517	0.40129	0.39743	0.39358	0.38974	0.38590
0.3	0.38209	0.37828	0.37448	0.37070	0.36693	0.36317	0.35942	0.35569	0.35197	0.34826
.....										
1.8	0.03593	0.03515	0.03438	0.03362	0.03288	0.03216	0.03144	0.03074	0.03005	0.02937
1.9	0.02872	0.02807	0.02743	0.02680	0.02619	0.02559	0.02500	0.02442	0.02385	0.02329
2.0	0.02275	0.02222	0.02169	0.02118	0.02068	0.02018	0.01970	0.01923	0.01876	0.01830
2.1	0.01786	0.01743	0.01700	0.01659	0.01618	0.01578	0.01539	0.01500	0.01461	0.01423

例) $z=2.00$ のとき

！超重要！

次に横方向から
0.00を探す

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.49601	0.49202	0.48803	0.48405	0.48006	0.47608	0.47210	0.46812	0.46414
0.1	0.46017	0.45620	0.45224	0.44828	0.44433	0.44038	0.43644	0.43251	0.42858	0.42464
0.2	0.42074	0.41683	0.41294	0.40905	0.40517	0.40129	0.39743	0.39358	0.38974	0.38589
0.3	0.38209	0.37828	0.37448	0.37070	0.36693	0.36317	0.35942	0.35569	0.35197	0.34825
.....										
1.8	0.03593	0.03515	0.03438	0.03362	0.03288	0.03216	0.03144	0.03074	0.03005	0.02937
1.9	0.02872	0.02807	0.02743	0.02680	0.02619	0.02559	0.02500	0.02442	0.02385	0.02329
2.0	0.02275	0.02222	0.02169	0.02118	0.02068	0.02018	0.01970	0.01923	0.01876	0.01830
2.1	0.01786	0.01743	0.01700	0.01659	0.01618	0.01578	0.01539	0.01500	0.01461	0.01423

例) $z=2.00$ のとき

！超重要！

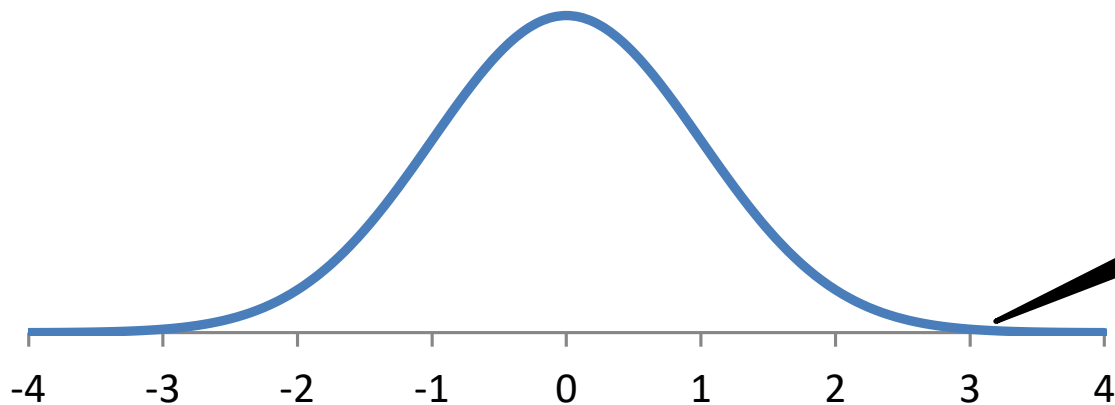
この値が $z=2.00$ のときの
上側確率

$$\Pr(Z > 2.00) = 0.02275$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.49601	0.49202	0.48803	0.48404	0.48005	0.47606	0.47207	0.46808	0.46409
0.1	0.46017	0.45620	0.45224	0.44828	0.44431	0.44035	0.43639	0.43243	0.42847	0.42451
0.2	0.42074	0.41683	0.41291	0.40899	0.40507	0.40115	0.39723	0.39331	0.38939	0.38547
0.3	0.38209	0.37828	0.37446	0.37065	0.36683	0.36301	0.35919	0.35538	0.35156	0.34774
0.4	0.34393	0.34013	0.33632	0.33251	0.32870	0.32489	0.32108	0.31727	0.31346	0.30965
0.5	0.30583	0.30203	0.29821	0.29439	0.29058	0.28676	0.28295	0.27913	0.27531	0.27149
0.6	0.26768	0.26387	0.26005	0.25623	0.25241	0.24859	0.24477	0.24095	0.23713	0.23331
0.7	0.22949	0.22567	0.22185	0.21803	0.21421	0.21039	0.20657	0.20275	0.19893	0.19511
0.8	0.19129	0.18747	0.18365	0.17983	0.17601	0.17219	0.16837	0.16455	0.16073	0.15691
0.9	0.15309	0.14927	0.14545	0.14163	0.13781	0.13399	0.13017	0.12635	0.12253	0.11871
1.0	0.11489	0.11107	0.10725	0.10343	0.09961	0.09579	0.09197	0.08815	0.08433	0.08051
1.1	0.07669	0.07287	0.06905	0.06523	0.06141	0.05759	0.05377	0.04995	0.04613	0.04231
1.2	0.03849	0.03467	0.03085	0.02703	0.02321	0.01939	0.01557	0.01175	0.00793	0.00411
1.3	0.00227	0.00145	0.00063	0.00021	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1.4	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1.5	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1.6	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1.7	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1.8	0.03593	0.03515	0.03437	0.03362	0.03288	0.03216	0.03144	0.03074	0.03005	0.02936
1.9	0.02872	0.02794	0.02717	0.02643	0.02568	0.02494	0.02421	0.02348	0.02275	0.02202
2.0	0.02275	0.02222	0.02169	0.02118	0.02068	0.02018	0.01970	0.01923	0.01876	0.01830
2.1	0.01786	0.01743	0.01700	0.01659	0.01618	0.01578	0.01539	0.01500	0.01461	0.01423

例) $z=5.00$ のとき

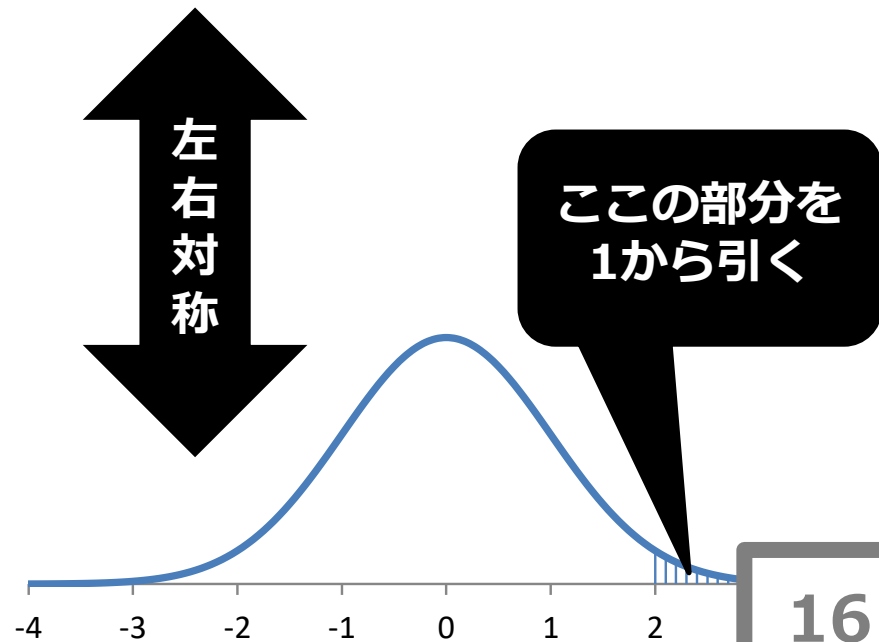
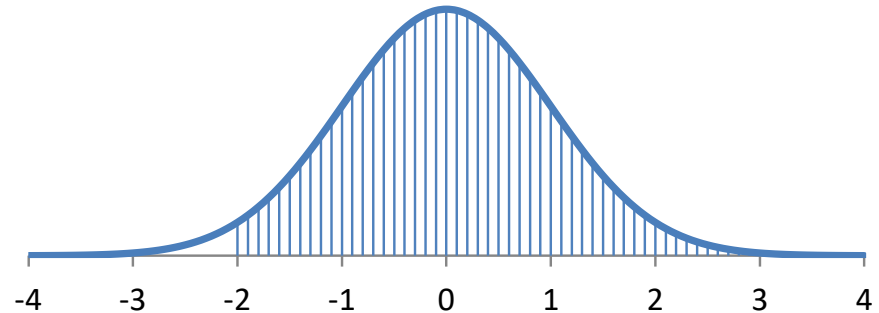
- 標準正規分布表には $z=3.09$ までしかない…
 - もっとも近い値で計算する
 - $z=3.09$ の確率をみると
0.00100とものすごく小さい値
 - $z>3.09$ のときはどんどん0に近似するので
ほぼ0とみなすことが多い
 - 教科書によってはさらに細かい表まであることもある



$z>3.09$ は
ほぼゼロ

例) $z = -2.00$ のとき

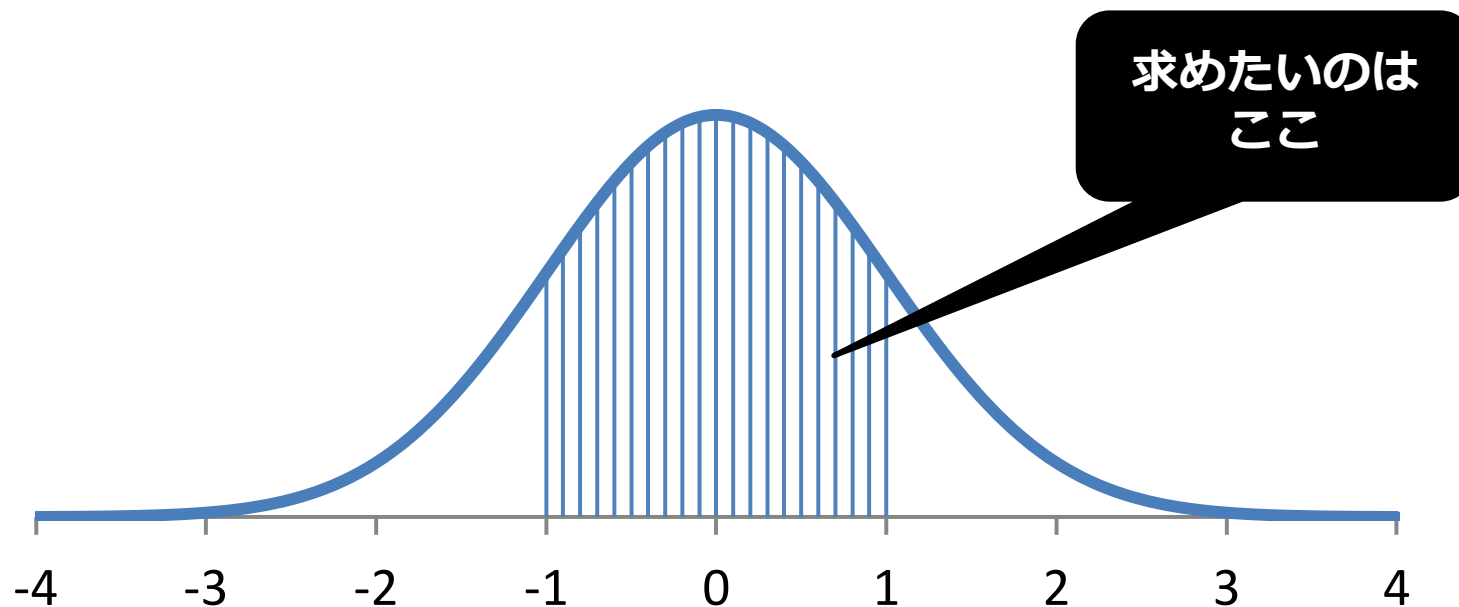
- もし z がマイナスの値だったら…？
 - 左右対称
 - 面積全体は1
(確率の合計は1)
 - つまり、
絶対値の上側確率を
計算して
1から引けば
求めることができる



例) $-1.00 < z < 1.00$ のとき

$$\Pr(-1.00 < Z < 1.00)$$

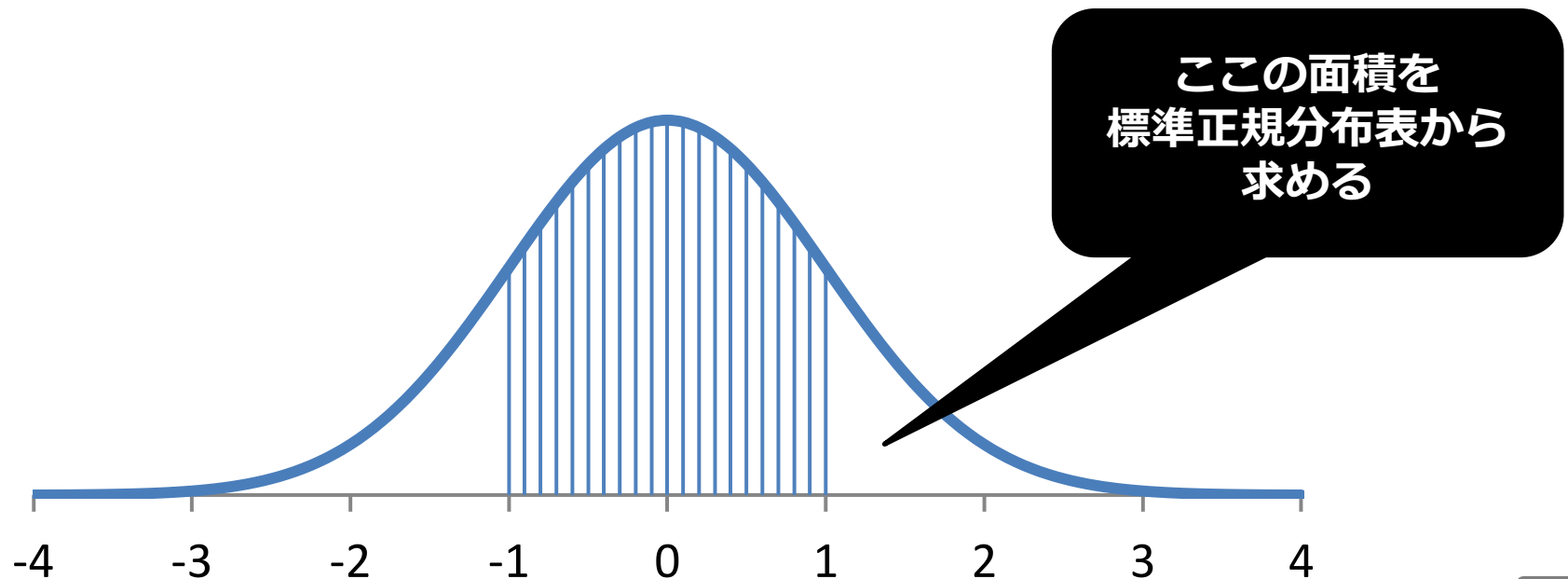
- 正規分布は左右対称



例) $-1.00 < z < 1.00$ のとき

$$\Pr(-1.00 < Z < 1.00)$$

- 正規分布は左右対称
 - 両端の白い部分の面積は同じ



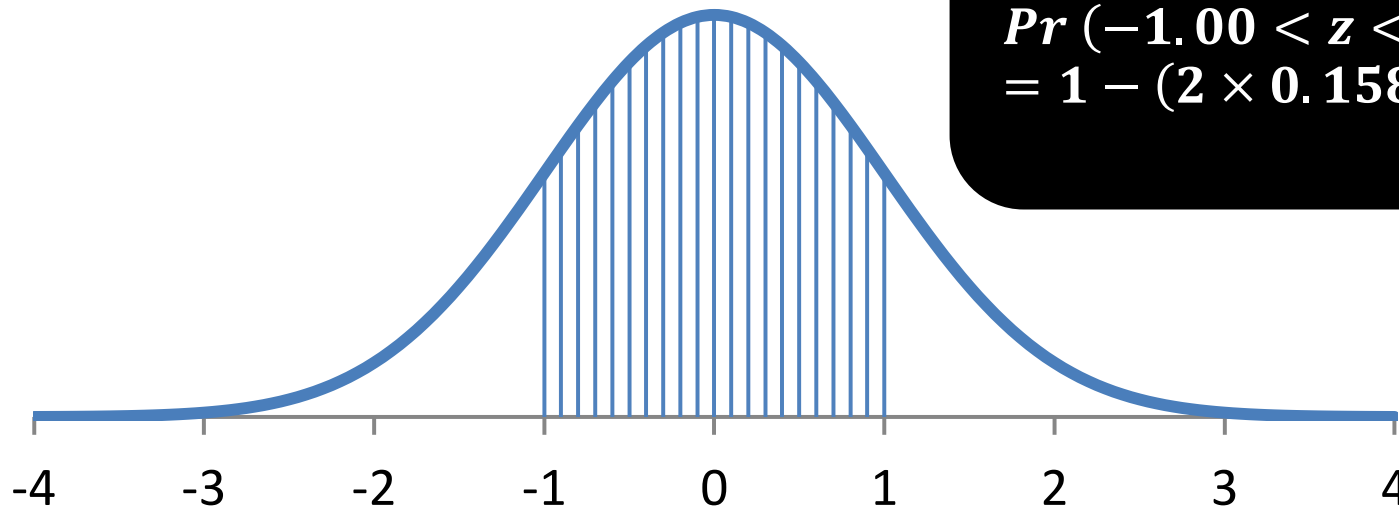
例) $-1.00 < z < 1.00$ のとき

$$\Pr(-1.00 < Z < 1.00)$$

- 正規分布は左右対称
 - 両端の白い部分の面積は同じ
- 確率の合計は1

求めた面積を2倍にして
1から引く

$$\begin{aligned} \Pr(-1.00 < z < 1.00) \\ = 1 - (2 \times 0.15866) = 0.68 \end{aligned}$$



- 母集団2万人の来店頻度
 - 平均 2.00
 - 標準偏差 1.41
 - 分散 $1.41^2 = 1.976$
- 標本平均 \bar{X} の分布⇒正規分布
 - 平均 2.00
 - 分散 $\frac{1.976}{n}$
- 標本の大きさが200のときに
母平均 ± 0.1 の範囲内に
標本平均が含まれる確率を求める

(pp. 86) 例題7-1

- 母平均 2.00
- 母分散 1.976
- 標本数 200

- 標本平均が母平均より0.1大きい場合のZ

$$Z = \frac{\bar{X} - E(\bar{X})}{\sqrt{\text{Var}(\bar{X})}} = \frac{\bar{X} - \mu_X}{\sqrt{\frac{\sigma_X^2}{n}}} = \frac{\text{標本平均} - \text{母平均}}{\sqrt{\frac{\text{母分散}}{\text{標本数}}}}$$

(pp. 86) 例題7-1

- 母平均 2.00
- 母分散 1.976
- 標本数 200

- 標本平均が母平均より0.1大きい場合のZ

$$Z = \frac{\bar{X} - E(\bar{X})}{\sqrt{\text{Var}(\bar{X})}} = \frac{\bar{X} - \mu_X}{\sqrt{\frac{\sigma_X^2}{n}}} = \frac{2.10 - 2.00}{\sqrt{\frac{1.976}{200}}} = 1.01$$

(pp. 86) 例題7-1

- 母平均 2.00
- 母分散 1.976
- 標本数 200
- $Z=1.01$ のときの上側確率を
標準正規分布表から求める
- $\Pr(-1.01 < Z < 1.01)$ を求める

- 標本数が $n=1000$ のとき
母平均 ± 0.1 の範囲内に
標本平均が含まれる確率を求めよ

- 標本数が $n=1000$ のとき
母平均 ± 0.1 の範囲内に
標本平均が含まれる確率を求めよ

$$Z = \frac{\bar{X} - E(\bar{X})}{\sqrt{\text{Var}(\bar{X})}} = \frac{\bar{X} - \mu_X}{\sqrt{\frac{\sigma_X^2}{n}}} = \frac{2.10 - 2.00}{\sqrt{\frac{1.976}{1000}}} \approx 2.25$$

$$\Pr(Z > 2.25) = 0.01222$$

$$\Pr(-2.25 < Z < 2.25) = 1 - (2 \times 0.01222) = 0.98$$

(6) 確率分布の再生性

確率分布の再生性

- 同じ確率分布にしたがう 2 つの独立な確率変数
 - X_1 と X_2
- 2 つの確率変数の和 $X_1 + X_2$
- 同じ確率分布にしたがうならばその確率分布は再生性をもつ
 - 再生性をもつ確率分布
 - 二項分布
 - ポアソン分布
 - 正規分布
 - カイ二乗分布
 - 中心極限定理が適用される
 - 標本の大きさを大きくしたとき標準正規分布に近似

カイ二乗分布

(1) カイ二乗分布の定義

カイ二乗分布の定義

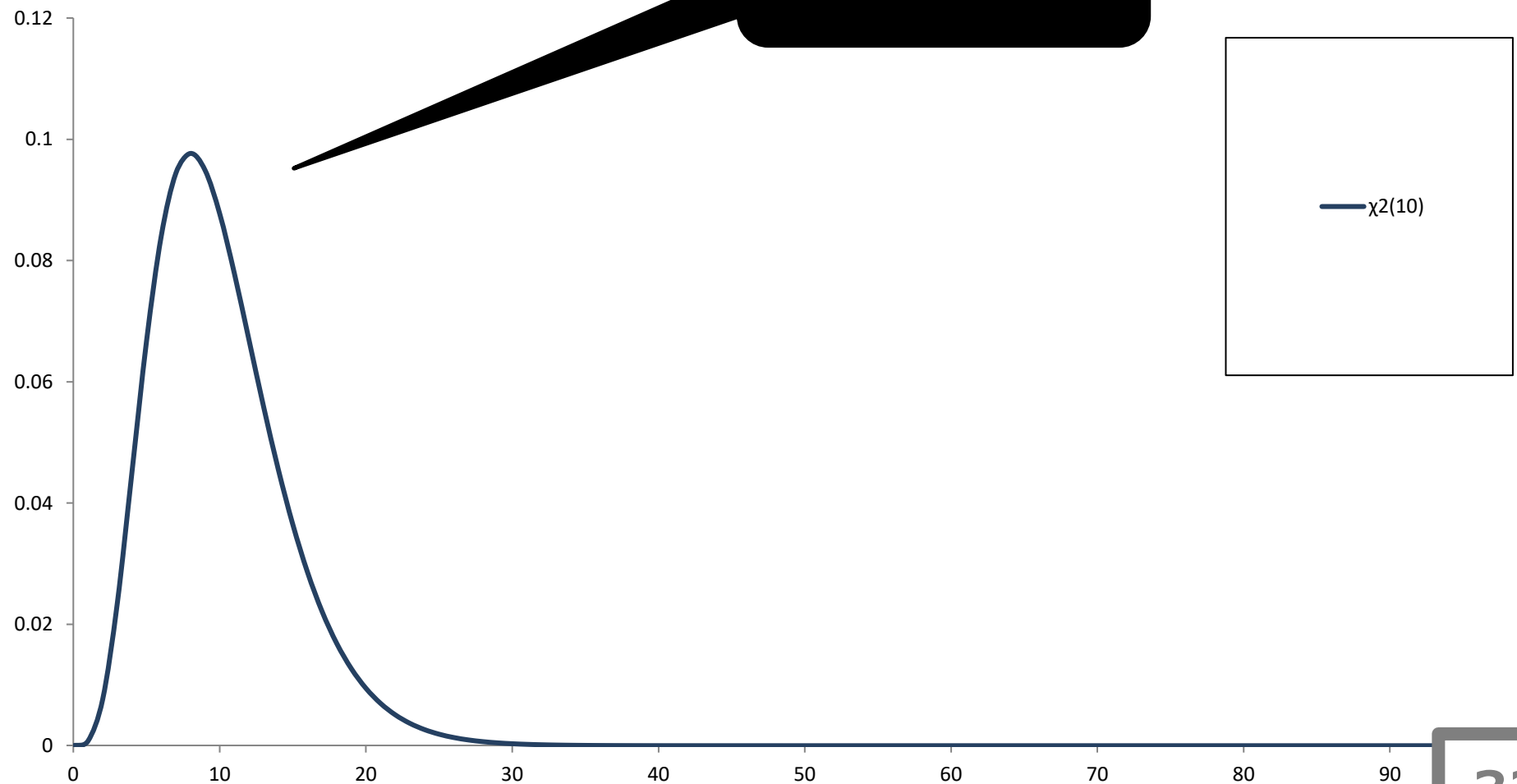
- χ^2 分布 (chi-squared distribution)
 - 自由度 ν のカイ二乗分布
 - 標準正規分布 $N(0,1)$ にしたがう独立な ν 個の確率変数 X_i の平方和に関する確率分布
$$\chi^2(\nu) \sim X_1^2 + X_2^2 + X_3^2 + \cdots + X_\nu^2$$
 - 平均値 ν 自由度と等しい
 - 分散 2ν 自由度の2倍
 - 自由度が小さいとき左右非対称
 - 自由度が大きくなるにつれて左右対称に変化
 - 自由度の値が大きいために正規分布に近似

自由度

- 自由度 (degree of freedom)
 - 自由に動くことのできる変数の数
 - 与えられた等式 (条件) の分だけ減る
- 例 データ (4, x, 1, 1) xは未知
条件 平均値は2
 $\Rightarrow 4 + x + 1 + 1 = 2 \times 4$
- データの個数n個に
平均値を与える (条件1つ) と
データ **n-1** 個がわかれば、すべてのデータがわかる
(残りの1つは必然的に決まる)

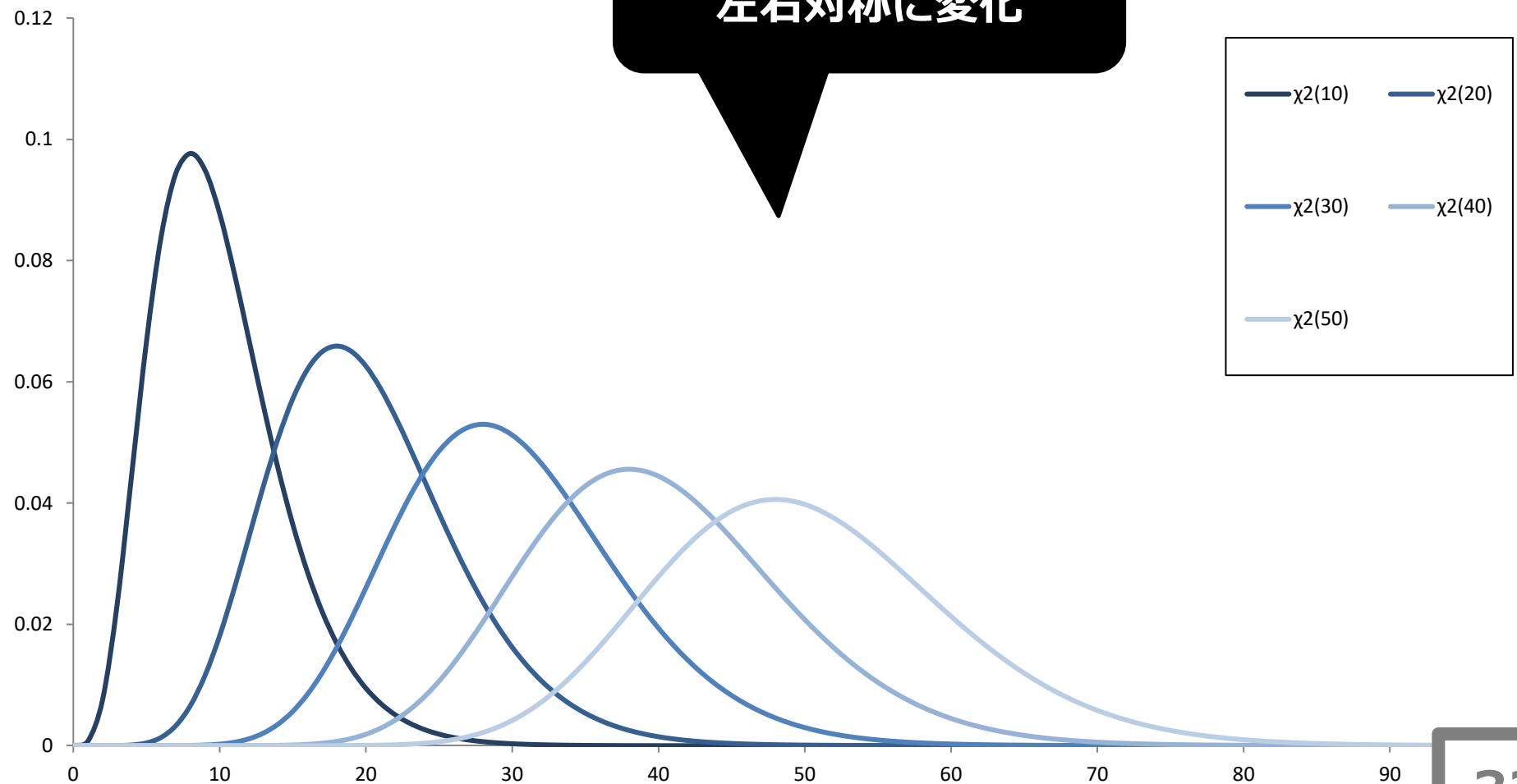
自由度10のカイ二乗分布

左右非対称



自由度10～50のカイ二乗分布

自由度が大きくなると
左右対称に変化



平方和の分解

Z^2 (標本平均 \bar{X} を標準化した Z の平方)
– 自由度1のカイ二乗分布にしたがう

$$\frac{\sum (X_i - \mu_X)^2}{\sigma_X^2} = \frac{\sum (X_i - \bar{X})^2}{\sigma_X^2} + \frac{n(\bar{X} - \mu_X)^2}{\sigma_X^2}$$

自由度n
のカイ二乗分布

自由度n-1
のカイ二乗分布

自由度1
のカイ二乗分布

第7章のまとめ

- 標準正規分布
 - 平均 0
 - 分散 1
- 標準正規分布表
 - 上側確率 $\Pr(Z > z)$ を求めることができる
 - ある区間内に含まれる確率は標準化により求める
- 再生性
 - 二項分布・ポアソン分布
 - 正規分布
 - カイ二乗分布
- カイ二乗分布
 - 標準正規分布の二乗和に関する確率分布
 - 自由度が大きくなると正規分布に近似
- 自由度
 - 自由に動くことのできる変数の数
 - 与えられた等式（条件）の分だけ減る