確率・統計処理 & 真値推定

カルマン・フィルタ 入門

Part 5: 確率・統計

リニア・テック　別府 伸耕

linear tec : Nobuyasu Beppu

確率・統計

- イントロ
- 記述統計
- 確率の基礎
- 確率変数と確率分布
- 2次元の確率分布
- 正規分布
- 中心極限定理

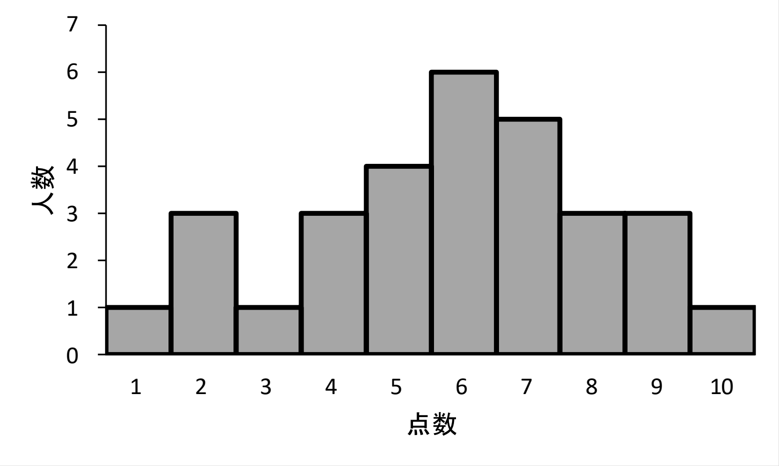
ヒストグラム：histogram

Sample

データを「階級」(calss)に分けて，
階級ごとのデータの数を表したものを「度数分布表」(frequency distribution table)という．
また，度数分布表を棒グラフで表したものを「ヒストグラム」という．

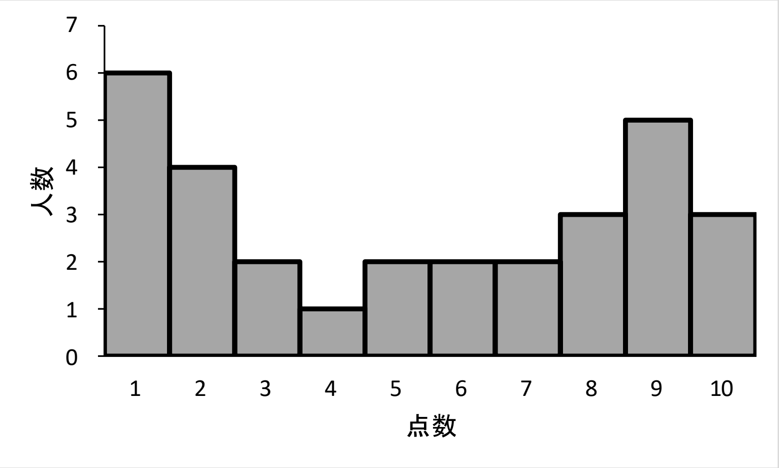
「ゲーム1」

スコア	人数
1	1
2	3
3	1
4	3
5	4
6	6
7	5
8	3
9	3
10	1
合計	30



「ゲーム2」

スコア	人数
1	6
2	4
3	2
4	1
5	2
6	2
7	2
8	3
9	5
10	3
合計	30



平均，分散，標準偏差

Sample

入手したデータの傾向を1つの値で表したものを「代表値」(representative value)という.
代表値の中でよく使われるものとして「平均」,「分散」,「標準偏差」がある.

● 平均（mean）

全データの合計値をデータの個数で割り算したもの．いわゆる「真ん中の値」.

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n}\sum_{k=1}^{n} x_k$$

● 分散（variance）

全データの「平均との差の2乗」を合計し，個数で割り算したもの．いわゆる「データのばらつき」.

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n} = \frac{1}{n}\sum_{k=1}^{n} (x_k - \bar{x})^2$$

● 標準偏差（standard deviation）

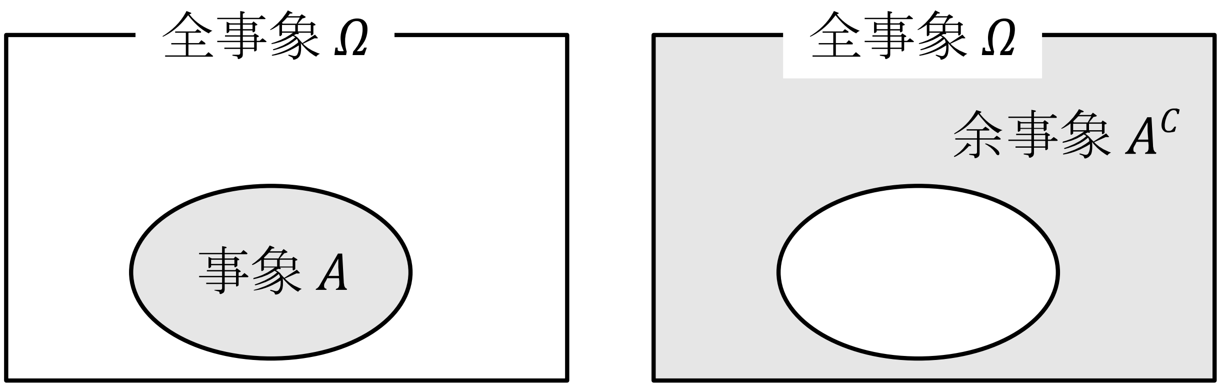
分散の平方根をとったもの.

$$s = \sqrt{\frac{1}{n}\sum_{k=1}^{n} (x_k - \bar{x})^2}$$

ベン図

Sample

「確率」とは，何らかの事象が全事象の中でどれだけ起こりやすいかを示す割合である．
このような議論では，「集合」の考え方が重要となる．
集合を視覚化して扱う上で便利なのが，次図に示す「ベン図」（Venn diagram）である．



通常，全事象は“Ω”（オメガ）という文字で表記される．
ベン図では，図の中の全体が全事象に対応する．
一般の事象は全事象 Ω の部分集合であり，上図のように全事象の一部を囲って表現する．

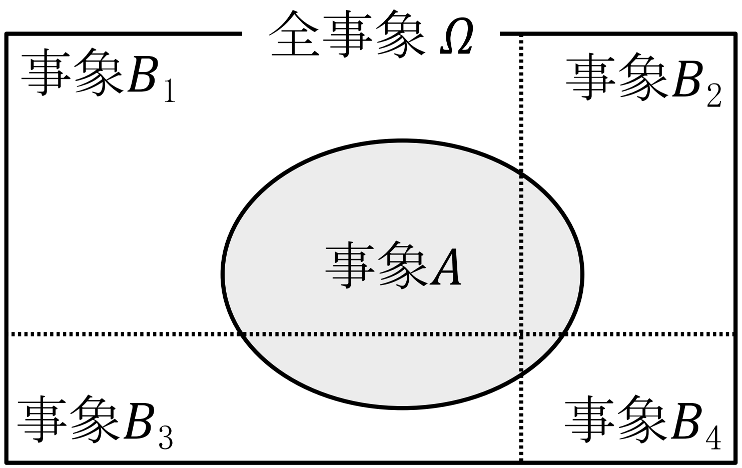
ある事象Aに対して，
A以外の事象のことを「余事象」あるいは「補事象」（complementary event）という．
補事象は“A^C”と表記される．

ベイズの定理：Bayes' theorem

Sample

次図のように全事象Ωが事象B_1, B_2, B_3, B_4に分割されており，互いに「排反」だとする.
このとき，次式が成り立つ.

$$\Omega = B_1 \cup B_2 \cup B_3 \cup B_4$$



ここで，別の「事象A」を考える.
事象Aが起こった前提で事象B_i（$1 \leq i \leq 4$）が起こる条件付き確率"$P(B_i|A)$"は次のとおり.
ただし，「乗法定理」より"$P(B_i \cap A) = P(B_i|A) \cdot P(A) = P(A|B_i) \cdot P(B_i)$"が成り立つ.

この式を「ベイズの定理」という.（実質，条件付き確率の定義にすぎない.）

$$\boxed{P(B_i|A) = \frac{P(B_i \cap A)}{P(A)} = \frac{P(A|B_i) \cdot P(B_i)}{P(A)}}$$

モーメント母関数：moment generation function

確率密度関数 “$p(x)$” に従う確率変数“X”および，一般の変数“t”を用意する．
変数tの関数“$M_X(t)$”を次式で定義し，これを「モーメント母関数」と呼ぶ．

$$M_X(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} \cdot p(x)\, dx$$

$p(x)$が正規分布のときに役立つ．

ここで，指数関数 e^{tx} の「マクローリン展開」を考える．

$$e^{tx} = 1 + \frac{1}{1!}tx + \frac{1}{2!}(tx)^2 + \frac{1}{3!}(tx)^3 + \cdots$$

上式をモーメント母関数 “$M_X(t)$” の定義式に代入する．
なおxに関する積分では，“t” は定数扱いであることに注意する．

$$M_X(t) = \int_{-\infty}^{\infty} \left\{1 + \frac{1}{1!}tx + \frac{1}{2!}(tx)^2 + \frac{1}{3!}(tx)^3 + \cdots\right\} \cdot p(x)\, dx$$

$$= \int_{-\infty}^{\infty} p(x)dx + \frac{1}{1!}t\int_{-\infty}^{\infty} xp(x)dx + \frac{1}{2!}t^2\int_{-\infty}^{\infty} x^2p(x)\, dx + \frac{1}{3!}t^3\int_{-\infty}^{\infty} x^3p(x)dx + \cdots$$

$$= 1 + \frac{1}{1!}t\, E[X] + \frac{1}{2!}t^2\, E[X^2] + \frac{1}{3!}t^3\, E[X^3] + \cdots$$

モーメント母関数 $M_X(t)$ には「すべての次数のモーメント」が含まれている．

アフィン変換と確率密度関数（1）

Sample

2×2行列“A”および2次元ベクトル“$\boldsymbol{b}$”を用いて,
2次元の確率変数ベクトル “$\boldsymbol{X}$” から新しい2次元確率変数ベクトル “$\boldsymbol{Y}$” を作る.

$$\boldsymbol{Y} = A\boldsymbol{X} + \boldsymbol{b}$$

ベクトル$\boldsymbol{X}$に対する上式のような変形操作を,「アフィン変換」(affine transform)という.
アフィン変換は,「線形変換＋オフセット付加」の操作と見なせる.

これから, 確率変数ベクトル“$\boldsymbol{Y}$”が従う同時確率密度関数 “$p_{\boldsymbol{Y}}(y_1, y_2)$” を求める.

ベクトル$\boldsymbol{X}$のことをベクトル$\boldsymbol{Y}$で表すと次のようになる.

$$\boldsymbol{X} = A^{-1}(\boldsymbol{Y} - \boldsymbol{b}) = A^{-1}\begin{pmatrix} Y_1 - b_1 \\ Y_2 - b_2 \end{pmatrix}$$

ここで, 逆行列 A^{-1} の要素を次のように表記する.

$$A^{-1} = \begin{pmatrix} a'_{11} & a'_{12} \\ a'_{21} & a'_{22} \end{pmatrix}$$

すると, ベクトル$\boldsymbol{X}$の成分X_1およびX_2は次のように書ける.

$$X_1 = a'_{11}(Y_1 - b_1) + a'_{12}(Y_2 - b_2)$$
$$X_2 = a'_{21}(Y_1 - b_1) + a'_{22}(Y_2 - b_2)$$

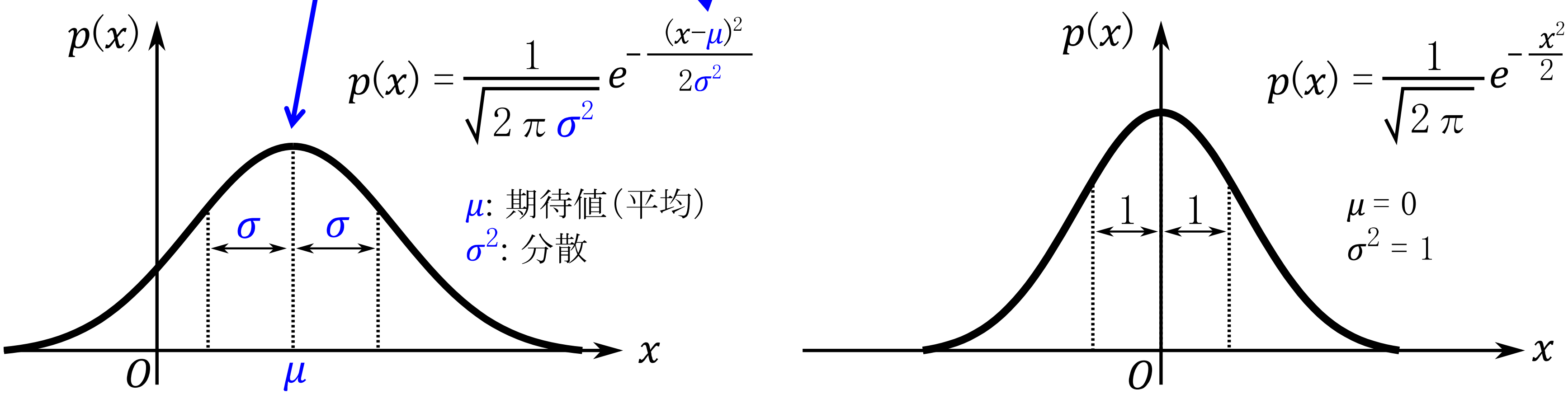
標準正規分布：standard normal distribution

期待値が“$\mu = 0$”，分散が“$\sigma^2 = 1$”である正規分布を「標準正規分布」という.

$$\boxed{p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}}$$

通常の「正規分布」と，上式の「標準正規分布」のグラフを次図に示す.

正規分布のピーク“μ”は，
「指数」の部分を見ればすぐにわかる.

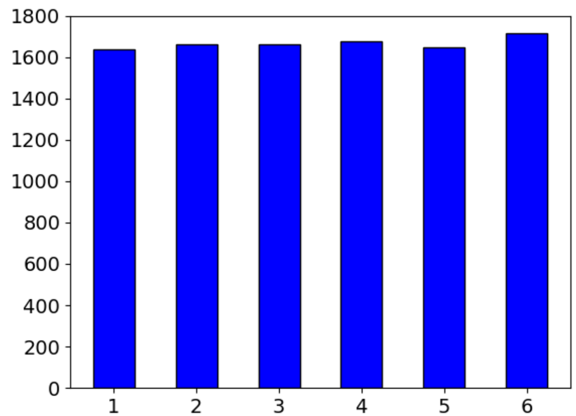


(a) 正規分布　　(b) 標準正規分布

中心極限定理の例 （すべて10,000回の試行）

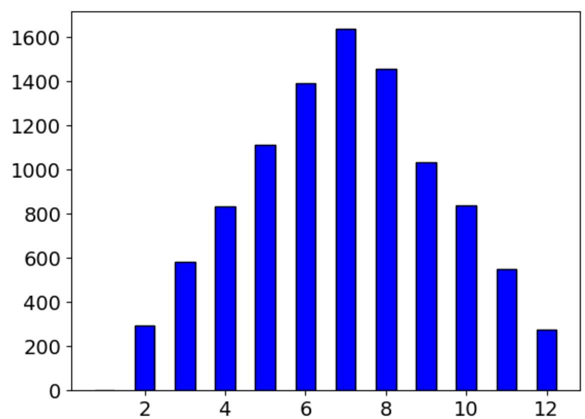
1つのサイコロの和の確率分布

X_1



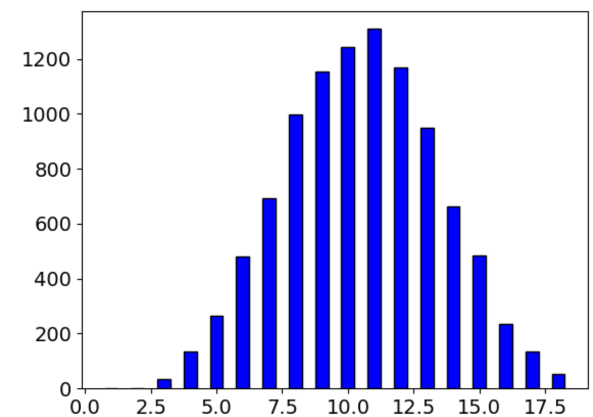
2つのサイコロの和の確率分布

$X_1 + X_2$



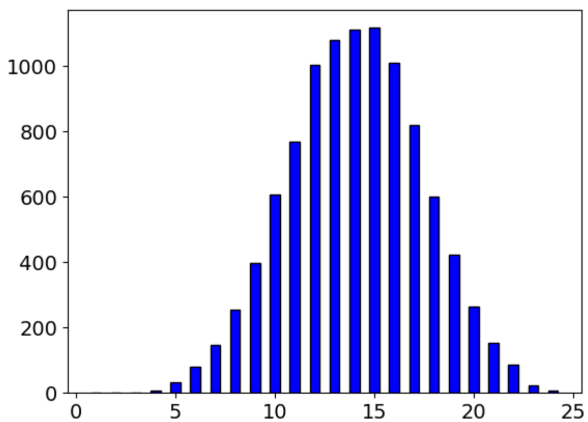
3つのサイコロの和の確率分布

$X_1 + X_2 + X_3$



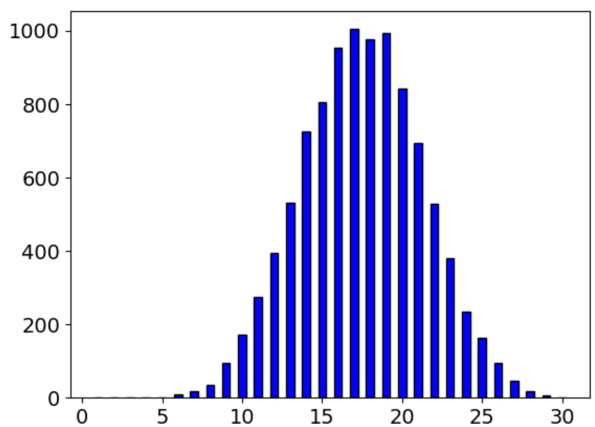
4つのサイコロの和の確率分布

$X_1 + X_2 + X_3 + X_4$



5つのサイコロの和の確率分布

$X_1 + X_2 + X_3 + X_4 + X_5$



6つのサイコロの和の確率分布

$X_1 + X_2 + X_3 + X_4 + X_5 + X_6$