応用システム工学
第一回　確率統計の復習

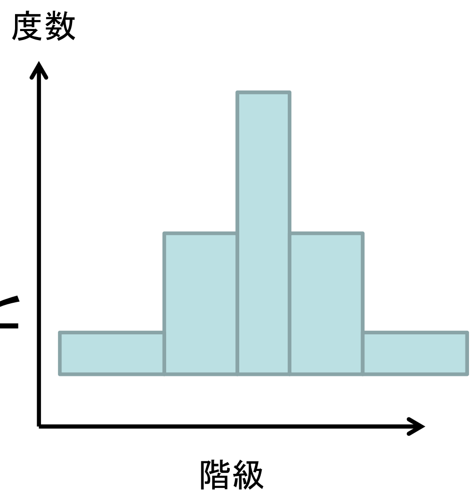
平成24年04月27日

統計

- 統計的推測　→　標本調査で(データの一部を調べて)全体を推測する
- 分布　→　標本調査したデータのバラつき
- 度数　→　データの現れる頻度を表す量
  - 相対度数　→　%で表したもの
- 度数分布　→　度数で表した分布
- 階級　→　データの値が入る区間
  - 区間の大きさを階級幅
  - 階級の上限・下限の中間値を階級値
- 代表値　→　分布を一つの数字で表す

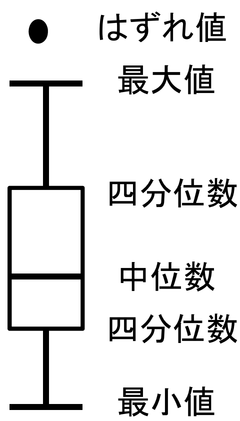
ヒストグラム

- 単なる棒グラフとは異なる
- 横軸　階級
- 縦軸　度数(相対度数)
- 階級幅を底辺・度数を面積とする長方形で，各階級の度数を表す
  - 面積で度数を表すことで，階級の取り方の自由度が高くなる



箱ひげ図(ボックスプロット・四分位図)

- ヒストグラムの発展系
- 下記の値のデータで表す
  - 最小値(小さい方から0%)
  - 第一四分位数(小さい方から25%)
  - 中位数(小さい方から50%)
    - 中央値，メディアン
  - 第三四分位数(小さい方から75%)
  - 最大値(小さい方から100%)



飛び離れた値を外れ値として，最大値・最小値から除外することもあり

分布を表す量

- 算術平均(相加平均)=$\frac{\text{データの合計}}{\text{データの総数}}$
  - $\bar{x} = \frac{x_1+x_2+\cdots+x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$
- 中央値(中位数，メジアン)=データを大きさ順に並べた時の中央の順位の値
  - データ数が偶数の場合は，中央の二つの算術平均をとる
- モード(最頻値)　→　度数が最も多い階級の階級値

分布を表す量

- 偏差　→　データ値と平均との差
  - 偏差の平均は0
- 分散　→　偏差の二乗平均
  - $\sigma^2 = V(x) = \frac{1}{n}\{(x_1-\bar{x})^2 + \cdots + (x_n-\bar{x})^2\}$

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

  - 標準偏差　→　分散の平方根 (もとの単位にもどる)

分布を表す量
モーメント

- モーメント(積率) → 分散の発展形
- 変数Xの度数分布を考える
  - 階級値xの階級の相対度数f(x)
    - 平均$\mu = E(X) = \sum_x xf(x)$
    - 分散$\sigma^2 = V(X) = E\{(x-\mu)^2\} = \sum_x (x-\mu)^2 f(x)$
    - 一般形 変数Xの関数g(x)の平均

$$E\big(g(x)\big) = \sum_x g(x)f(x)$$

  - 関数g(x)がk次 → 変数Xのk次のモーメント

分布を表す量
モーメント

- 原点の周りのk次のモーメント $\mu'_k = E\left(X^k\right)$
  - 平均μは原点の周りの1次のモーメントに相当
- 平均の周りのk次のモーメント $\mu_k = E\left((X-\mu)^k\right)$
  - 分散σ^2は平均の周りの2次のモーメントに相当
- 高次のモーメントを用いた分布の特徴量
  - 3次のモーメント 歪度 $\alpha_3 = \frac{\mu_3}{\sigma_3}$
    - ヒストグラムの正または負への偏り具合
  - 4次のモーメント 尖度 $\alpha_4 - 3 = \frac{\mu_4}{\sigma_4} - 3$ (正規分布の場合)
    - 平均値付近での分布の集中の度合い

確率について

- 確率変数
  - 値がxとなる確率pが与えられている変数X
    - Xの取り得る値は離散値・連続値どちらでも可
- 事象
  - 確率変数Xが取り得る値の部分集合
  - 事象に対する確率の表し方
    - P(X=1)　　Xが1となる確率:連続値・離散値
    - P(5≦X<10)　Xが5以上10未満となる確率:連続値
    - P(X∈A)　XがAに含まれる確率:離散値

確率分布関数

- 確率分布
  - 確率変数Xと，対応する確率P(X=x)の対応関係
- 確率分布関数
  - 確率変数Xが，x以下の値をとる確率

$$F_X(x) = P(X \leq x)$$

- 性質
  - 単調非減少
  - $F_X(-\infty) = 0$
  - $F_X(\infty) = 1$

※確率変数の種類(連続・不連続)
に関係しない

確率密度関数

- 連続な確率変数Xに対する確率密度関数:$f_X(x)$
  - $x \le X \le x + dx$ となる確率 $f_X(x)dx$
  - 確率密度関数の性質

$$F_X(x) = \int_{-\infty}^{x} f_X(x')dx'$$

$$f_X(x) = \frac{d}{dx} F_X(x)$$

$$f_X(x) \ge 0$$

$$\int_{-\infty}^{\infty} f_X(x)dx = 1$$