

## 2章: 確率・統計

### 集合

$S=\{a, b, c, d, e, f, g\}$  ※ $a\sim g$ は各要素、 $S$ が集合を表す。この時、要素 $a$ が集合 $S$ に属することを  $a \in S$ と表す。  
 $S$ の部分集合 $M=\{c, d, g\}$ がある場合、  $M \subset S$ と表す。

- 集合Aと集合BのOR:  $A \cup B$
- 集合Aと集合BのAND:  $A \cap B$

### 確率

- 頻度確率（客観確率）
  - 発生する頻度
  - 例：10本のうち、1本が当たりのくじは当選確率10%
- ベイズ確率（主観確率）
  - 信念の度合い
  - 例：「あなたは40%の確率でインフルエンザ」という診断
  - 全数調査などを行わずに、色々な条件を加味して確率を出す。

### 条件付き確率

ある事象Bが与えられた下でAとなる確率  $P(A|B) = \frac{P(A \cap B)}{P(B)}$  これを変形すると  $P(A \cap B) = P(B) * P(A|B)$  となる。

### ベイズ則

ある事象A,Bがある時、以下が成り立つ。  $P(A)P(B|A) = P(B)P(A|B)$

### 記述統計と推測統計

- 記述統計： 集団の性質を要約し、記述するもの。AIの扱うデータはこちらがメイン
- 推測統計： 集団から一部を取り出し、元の集団（母集団）の性質を推測する

### 確率変数と確率分布

- 確率変数
  - 事象と結びつけられた数値。賞金みたいなもの。
  - 事象そのものを指すと解釈する場合も多い
- 確率分布
  - 事象の発生する確率の分布。
  - 離散値であれば表に示せる

項目	説明	式
期待値	その分布における確率変数の平均の値or「ありえそう」な値	$E(f) = \sum_{k=1}^n P(X = x_k)f(X = x_k)$

項目	説明	式
分散	データの各々の値が期待値から どれだけズレているかを平均したもの	$Var(f) = E(f^2(X = x)) - (E(f))^2$
共分散	2つのデータの傾向の違い。 正の値は似た傾向、負の値は逆の傾向を示す	$Cov(f, g) = E(fg) - E(f)E(g)$
標準偏差	分散の平方根を取って、元の単位に戻したもの	$\sigma = \sqrt{Var(f)} = \sqrt{E((f(x) - E(f))^2)}$

## 様々な分布

分布	説明	式
ベルヌーイ分布	コイントスなど0, 1が決まるもの。 裏と表で出る割合が等しくなくとも扱える	$P(x u) = u^x(1 - u)^{1-x}$
マルチヌーイ分布（カテゴリカル分布）	さいころを転がすイメージ。 各面の出る割合が等しくなくとも扱える	
二項分布	ベルヌーイ分布の多試行版	$P(x \lambda, n) = \frac{n!}{x!(n-x)!} \lambda^x (1-\lambda)^{n-x}$
ガウス分布	釣り鐘型の連続分布。 真の分布がわからなくてもサンプルが多ければ正規分布に近づく	$N(x; \mu, \sigma^2) = \sqrt{\frac{1}{2\pi\sigma^2}} \exp(-\frac{1}{2\sigma^2}(x - \mu)^2)$