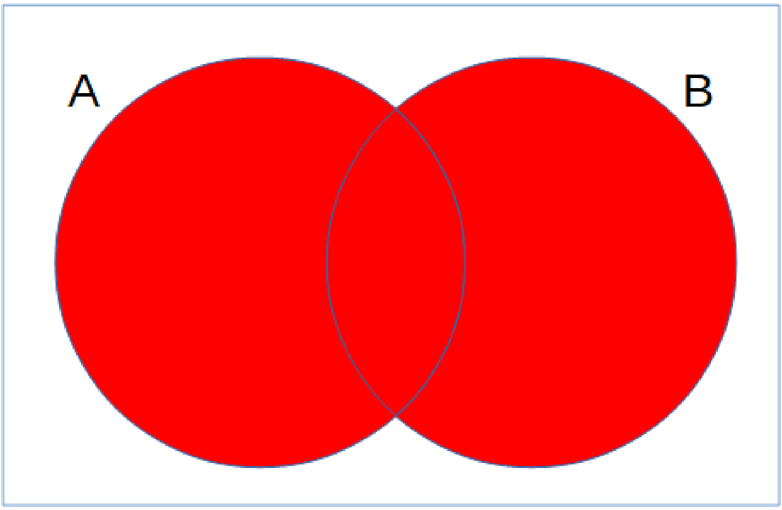
応用数学 第2章：確率・統計

はじめに

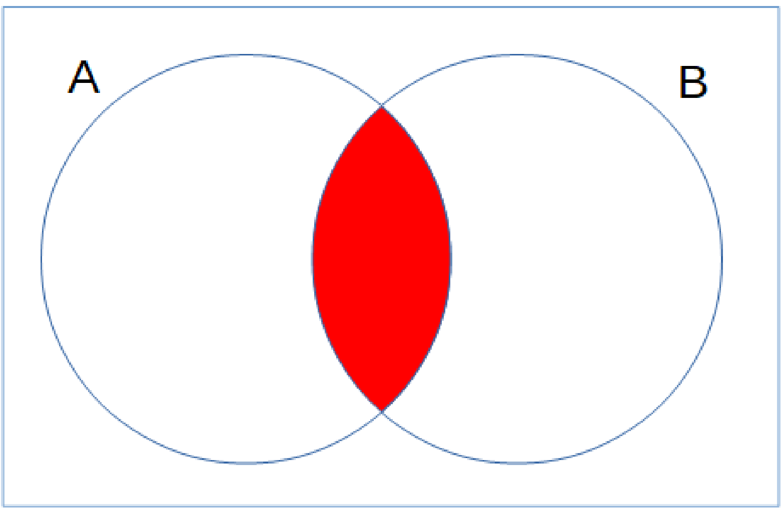
「応用数学 第2章：確率・統計」について、動画を見ながら公式を記載したり例題を解いたりしながら理解を深める。

集合

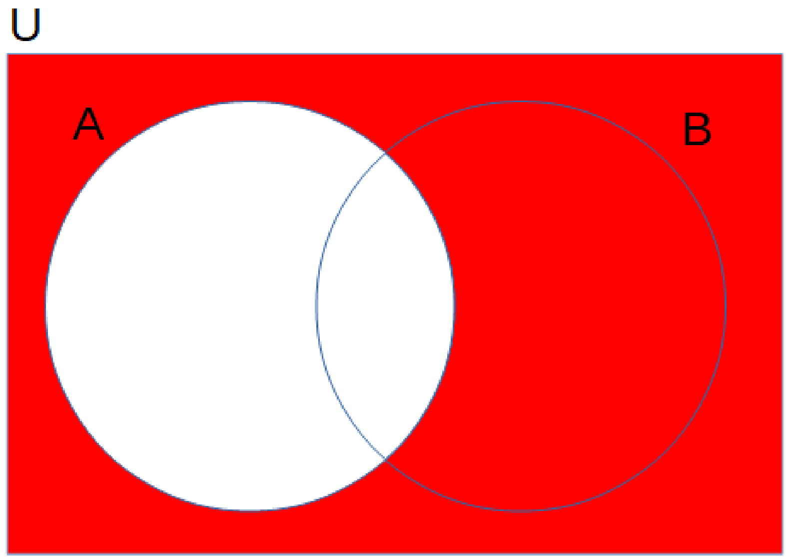
- A ∪ B （和集合）



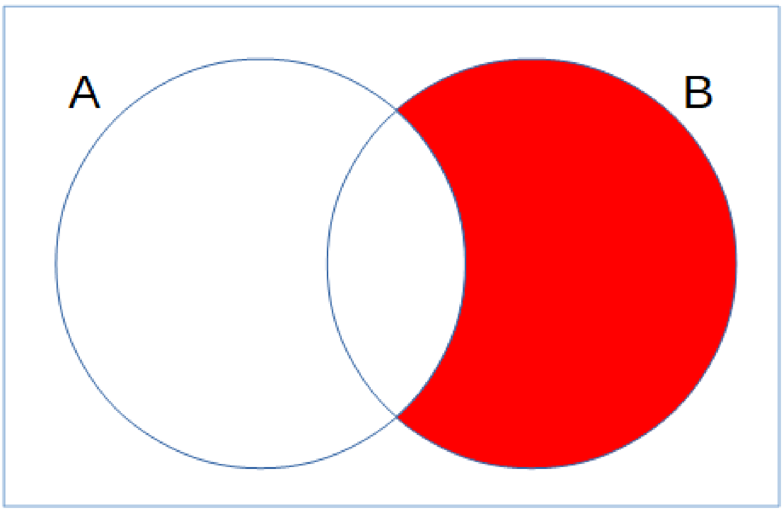
- A ∩ B（共通部分）



- U \ A = $\bar{A}$（絶対補）・・・UからAを除いた部分



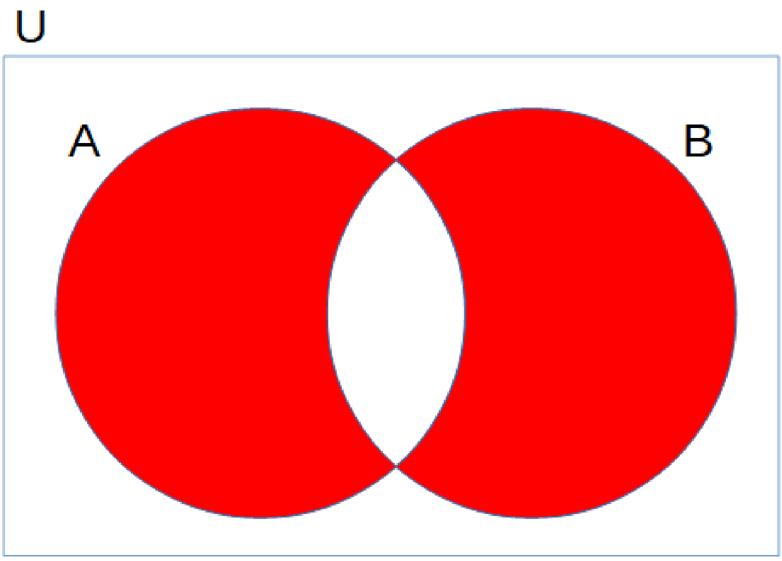
- B \ A（相対補）・・・BからAを除いた部分



問題

次の図を表現する式はどれか？

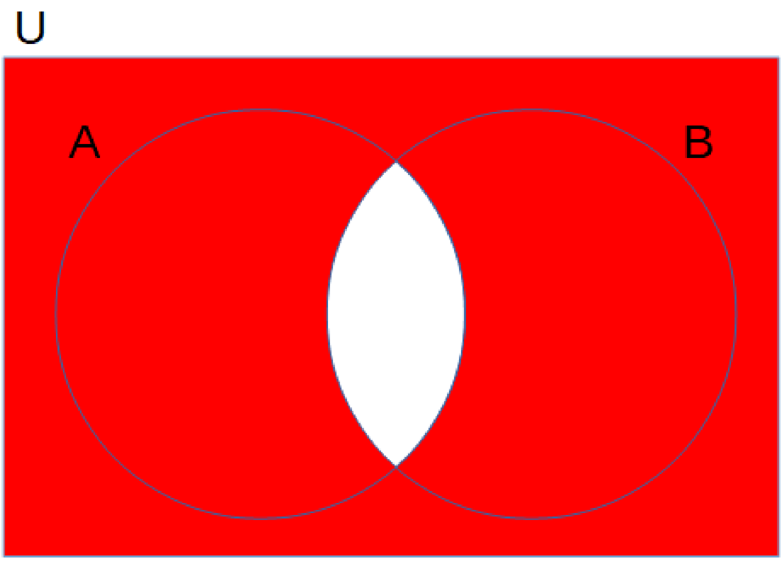
1. $\overline{A \cup B}$
2. $\overline{A \cap B}$
3. $(B \backslash A) \cap (A \backslash B)$
4. $(B \backslash A) \cup (A \backslash B)$



回答

4（$(B\backslash A)\cup(A\backslash B)$）

※2（$\overline{A\cap B}$）は以下になるので違う



確率の定義

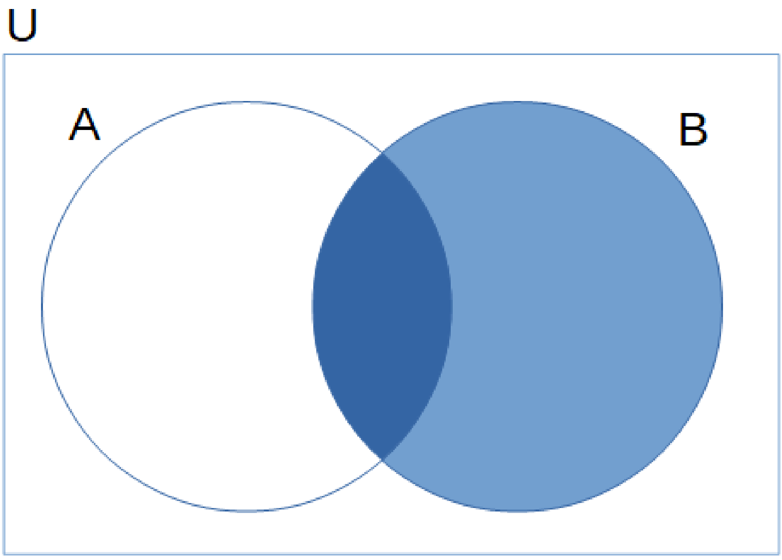
$$P(A)=\frac{n(A)}{n(B)}=\frac{\text{事象}A\text{が起こる数}}{\text{すべての事象の数}}$$

※確率は0から1の間の値をとる

- 例題1
  $P(\bar{A})$を$P(A)$を使って表せ

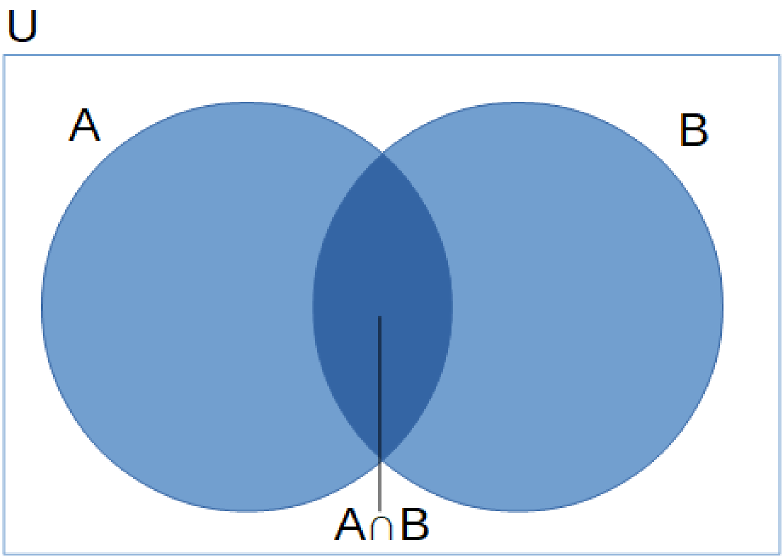
条件付き確率

- ある事象Bが与えられた下で、Aとなる確率
  例）雨が降っている条件下で交通事項に合う確率
  → 同じ「Aとなる確率」でも「全体UのうちAとなる確率」よりも確率の値は大きくなる



$$P(\bar{A}) = 1 - P(A)$$

- $P(A \cap B)$



$$P(A \cap B) = P(A)P(B|A) = P(B)P(A|B) = P(B \cap A)$$
※$P(B|A) \cdots$　Aという条件下でのB

事象Aと事象Bが互いに独立している場合
$$P(A \cap B) = P(A)P(B)$$

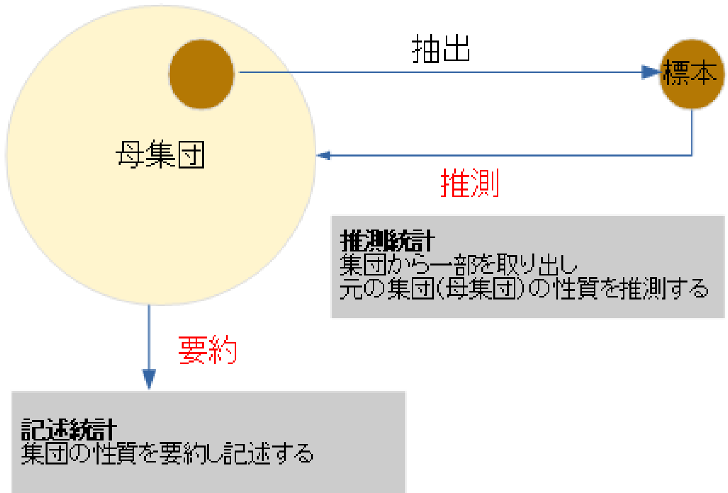
- $P(A \cup B)$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

ベイズ則

$$P(A)P(B|A) = P(B)P(A|B)$$

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
- 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
  - 事象と結びつけらた数値
  - 事象そのものと解釈する場合もある
- 確率分布（次の表）
  - 事象の発生する確率の分布
  - 離散値であれば表に示せる

事象	裏0枚、表4枚	裏1枚、表3枚	裏2枚、表2枚	裏が3枚、表1枚	裏4枚、表0枚
確率変数（裏を0, 表を1と対応させ和をとる）	4	3	2	1	0
事象が発生した回数	75	300	450	300	75
事象と対応する確率	1/16	4/16	6/16	4/16	1/16

期待値

事象X	x_1	x_2	$\cdots$	x_n
確率変数f(X)	$f(x_1)$	$f(x_2)$	$\cdots$	$f(x_n)$
確率P(X)	$P(x_1)$	$P(x_2)$	$\cdots$	$P(x_n)$

- 期待値E(f)

$$\sum_{k=1}^{n} P(X = x_k) f(X = x_k)$$

- 連続する値の期待値E(f)

$$P(X = x_k) f(X = x_k) dx$$

分散と共分散

- 分散
  - データの散らばり具合
  - 各値が期待値からどれだけずれているかを平均したもの

$$\begin{aligned}&\text{分散}Var(f)\\&= E((f_{(X=x)} - E_{(f)})^2)\\&= E(f^2_{(X=x)}) - (E_{(f)})^2\end{aligned}$$

※<u>「fの2乗の平均」-「平均の2乗」</u>

- 共分散
  - 2つのデータ系列の傾向の違い
    - 正の値ならば似た傾向
    - 負の値ならば逆の傾向
    - ゼロならば関係性に乏しい

$$\begin{aligned}&\text{共分散}Cov(f, g)\\&= E((f_{(X=x)} - E(f))(g_{(Y=y)} - E(g))\\&= E(fg) - E(f)E(g)\end{aligned}$$

分散と標準偏差

- 分散は2乗してしまうので元のデータと単位が違ってしまう
  （例えば身長をcmで計っていたとすると、cmの2乗は平方センチメートル）
  → 2乗の平方根すれば、元の単位に戻る

$$
\begin{aligned}
\sigma &= \sqrt{Var(f)} \\
&= \sqrt{E((f_{(X=x)} - E_{(f)})^2)}
\end{aligned}
$$

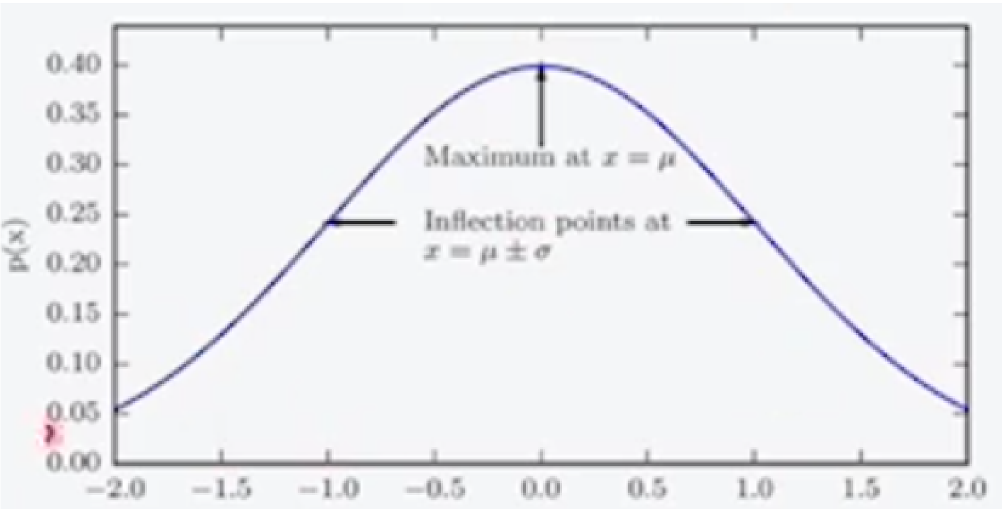
様々な確率分布

- ベルヌーイ分布
  - コイントスのようなイメージ
  - 裏と表で出る割合が等しくなくとも使える
  - $P(x|\mu) = \mu^x (1-\mu)^{1-x}$

- マルチヌーイ（カテゴリカル）分布
  - サイコロを転がすイメージ
  - 各面の出る割合が等しくなくとも使える

- 二項分布
  - ベルヌーイ分布の多試行版
  - $P(x|\lambda, n) = \frac{n!}{x!(n-x)!} \lambda^x (1-\lambda)^{n-x}$
  - 釣鐘型の分布

- ガウス分布
  - 釣鐘型の連続分布
  - $N(x; \mu, \sigma^2) = \sqrt{\frac{1}{2\pi\sigma^2}} exp(-\frac{1}{2\sigma^2}(x-\mu)^2)$
  - 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
  - それぞれの積分、面積を足し合わせると合計1になる



推定

- 母集団を特徴づける母数（パラメータ：平均など）を統計学的に推測すること
- 点推定
  - 平均値などを1つの値に推定すること
- 区間推定
  - 平均値などが存在する範囲（区間）を推定すること

推定量と推定値

- 推定量（estimator）
  - パラメータ（平均など）を推定するために利用する数値の計算方法や計算式のこと。推定関数とも呼ぶ

- 推定値（estimate）
  - 実際に施行を行った結果から計算した値
  - 真の値をθとすると、$\hat{\theta}$は推定値を表す
    - ハット（^） がついていたら、推定しているのだなと考える

標本平均

- 母集団から取り出した標本の平均値
  - 一致性
    - サンプル数が多くなれば、母集団の値に近づく
  - 不偏性
    - サンプル数がいくらであっても、その期待値は母集団の値と同様
  - $E(\hat{\theta}) = \theta$

標本分散

サンプルサイズをnとすると

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

- 一致性は満たすが、不偏性は満たさない！

不偏分散

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

- 標本分散に$\frac{n}{n-1}$をかけて修正している
- サンプルが十分に取れない場合は、このようなやり方もある

演習問題

問 3.1

「異なる絵柄の描かれた 5 枚のカードから 1 枚を引く」というような，同じ条件下で繰り返し行うことのできる実験や観測などのことを「試行」といい，それによって起こる「★印の描かれたカードを引いた」というような，結果を「事象」という。
試行の結果として起こる事象に整数や実数の数値が結びつけられているときに，その数値を「確率変数」という。
次の選択肢のうち，確率変数として適当なものはどれか。すべて選べ。

a.　さいころを振ったときに出た目の数。
b.　3 枚のコインを同時に投げて裏表どちらが出るか試行したときのコインの枚数。
c.　赤，青，黄の 3 色のボールが壷の中に入っている。ここから 1 個を取り出したときの色。
d.　8 本のうち 1 本が当たりであるクジを，当たりが出るまで抽選し続けたときの回数。

問3.1

- 回答（★ 誤り ★）
  - a
- 考察（★ 誤り ★）
  - bは、常に3枚で試行とは関係なし。cは事象、dは事象が発生した回数
- 正答
  - a, d
  - 解説には「試行の結果として起こる事象に数値が結びつけられているときに、その数値を「確率変数」という」とある。
    dはまさに回数なので確率変数となる。

問 3.2

試行の結果に生じる様々な事象は，すべてが同じ割合で生じるとは限らない。たとえば，「4 枚のコインを同時に投げて裏表どちらが出るか」試行した場合「すべて表であったり，すべて裏である場合は少なく，2 枚は裏で 2 枚が表といった裏表が混合している場合の方がより多く発生するだろう」ということは計算しなくとも直観的に理解できるかと思う。事象は様々な割合で（すなわち様々な確率で）発生するのだから，事象と結びつけられた数値（たとえば表という事象に 1 という数値を，裏という事象に 0 という数値を対応させる etc）である確率変数もまた様々な確率で発生することになる。この確率変数と対応する確率との関係を表した関数を「確率分布」という。確率変数が離散値（とびとびの値）ならば対応

する確率もまた離散値となる。つまり確率変数と対応する確率との関係（すなわち確率分布）を表にまとめることができるということである。
上述のコインを同時に投げる試行を 1200 回行ったとして，下の表を作った。空欄を埋めよ。

事象	裏が 0 枚，表が 4 枚	裏が 1 枚，表が 3 枚	裏が 2 枚，表が 2 枚	裏が 3 枚，表が 1 枚	裏が 4 枚，表が 0 枚
確率変数（裏を 0，表を 1 と対応させ和をとった）	4	3	2	1	0
事象が発生した回数	75	300	450		75
事象と対応する確率	1/16				

問3.2

- 回答

事象	裏0枚、表4枚	裏1枚、表3枚	裏2枚、表2枚	裏が3枚、表1枚	裏4枚、表0枚
確率変数（裏を0, 表を1と対応させ和をとった）	4	3	2	1	0
事象が発生した回数	75	300	450	**300**	75
事象と対応する確率	1/16	**1/4(4/16)**	**3/8(6/16)**	**1/4(4/16)**	**1/16**

- 考察
  - 「裏が3枚、表1枚」の回数 = 1200 - (75 + 300 + 450 + 75) = 300
  - 「裏1枚、表3枚」の確率 = 300 / 1200 = 1/4(4/16)、他同様に計算

第 5 章　条件付き確率

問 5.1

条件付き確率とは，事象 A が起こった条件の下で事象 B が起こる確率のことであり，これを$\mathrm{p}(B|A)$と表す（B が先に書かれるのは，現在の数学の記述法の発祥の地であるヨーロッパで使われている言語における語順が反映しているため。例：the probability of B under the condition A）。たとえば，ある年における，洗濯物を干していたときに，雨が降ってきた確率（洗濯物を干していた日という条件下で，雨が降ってきた日の発生する確率）を考える。前述の記述法に従うならば，

$$\mathrm{p}\left(雨が降ってきた日\middle|洗濯物を干していた日\right)$$

のように記述できる。これは洗濯物を干していてかつ雨が降ってきた確率（洗濯物を干していてかつ雨が降ってきた日の発生する確率）と同様ではない。事象 A と事象 B が同時に起こる確率を同時確率といい，$\mathrm{p}(A,B)$あるいは$\mathrm{p}(B,A)$で表され（どちらも同じ値となる），条件付き確率とは区別されなければならない。なぜなら，条件付き確率の場合

$$\mathrm{p}\left(雨が降ってきた日\middle|洗濯物を干していた日\right)$$
$$=\frac{洗濯物を干していてかつ雨が降ってきた日数}{すべての洗濯物を干していた日数}$$

を表していることになるが，同時確率の場合

$$\mathrm{p}\left(雨が降ってきた日,洗濯物を干していた日\right)$$
$$=\frac{洗濯物を干していてかつ雨が降ってきた日数}{すべての日数（つまり365日）}$$

を表していることになるからである。

1 年のうち洗濯物を干していた日数を 60 日，洗濯物を干していてかつ雨が降ってきた日数

を 12 日として，ある年における，洗濯物を干していた日という条件下で，雨が降ってきた日の発生する確率と，洗濯物を干していてかつ雨が降ってきた日の発生する確率を求めよ。

5.2
袋の中に赤い玉 3 個と白い玉 2 個が入っている。赤い玉は A，B，C の文字が，白い玉には A，B の文字が，それぞれ 1 個に対して 1 文字ずつ記されている。以下の問いに答えよ。

問 5.2.1　出てきた玉が赤色であったとき，それに記されている文字が B である確率。

問 5.2.2　出てきた玉に記されている文字が A であったとき，その玉の色が白色である確率。

問5.1

- 回答1（洗濯ものを干していた日という条件下で、雨が降ってきた日の発生する確率）

$$
\begin{aligned}
&P(\text{雨が降ってきた日}|\text{洗濯物を干していた日}) \\
&= \frac{\text{洗濯物を干していてかつ雨が降ってきた日数}}{\text{洗濯物を干していたすべての日数}} \\
&= \frac{12}{60} \\
&= \frac{1}{5}
\end{aligned}
$$

- 回答2（洗濯ものを干していてかつ雨が降ってきた日の発生する確率）

$$
\begin{aligned}
&P(\text{雨が降ってきた日}, \text{洗濯物を干していた日}) \\
&= \frac{\text{洗濯物を干していてかつ雨が降ってきた日数}}{\text{すべての日数}365\text{日}} \\
&= \frac{12}{365}
\end{aligned}
$$

- 考察
  - 問題文の中で条件付確率を丁寧に説明してくれているのが嬉しい
  - 問題を解く中で、条件付確率の意味が理解できる
  - $P(A \cap B) = P(B)P(A|B)$
    ベン図において、Aを雨が降ってきた日数、Bを洗濯物を干していた日数とすると
    $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$

問5.2.1

- 回答

$$P(\text{文字が}B|\text{赤玉が出る}) \\ = \frac{1}{3}$$

- 考察
  - 赤玉である条件下で文字がBである確率なので、赤玉3つからBが出る確率を求めることに等しい

問5.2.2

- 回答

$$P(\text{白玉が出る}|\text{文字が}A) \\ = \frac{1}{2}$$

- 考察
  - 文字がAである条件下で白玉が出る確率なので、Aの赤玉・白玉の計2つから白玉が出る確率を求めることに等しい

おわりに

本レポートでは、動画の要点を記載しながら最後に付属の演習問題を回答した。
動画では例題が少なかったが、演習問題のおかげで条件付確率の考え方がとても理解できた。
一方で、ベルヌーイ分布や二項分布などの各種確率分布や標本平均をどういう時にどのように利用するのかは
まだイメージできていない。
インターネットの記事を参照したり、確認テストを通してさらに理解を深めていく予定である。