目　次

内容

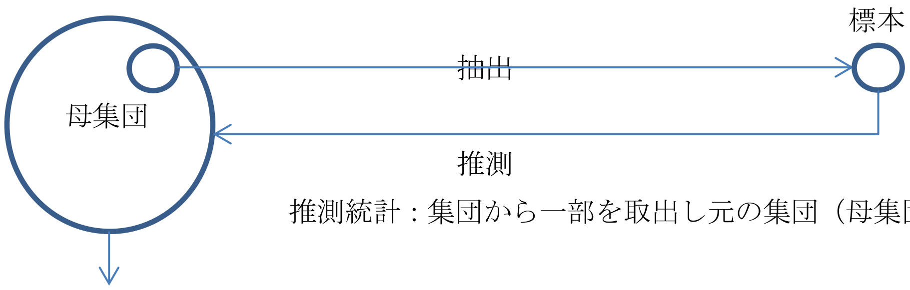
１．統計学２

１．１　統計学 2-1（プロローグ）

1）記述統計

・木だけを見づに森を見る。

2）推測統計



推測統計：集団から一部を取出し元の集団（母集団）の性質を推測する

推測統計：集団の性質を要約し記述する

１．２　統計学 2-2（確率変数と確率分布）

１）確率変数

・確率的な変数

・事象と結び付けられた数値

・当たりくじの金額

・事象そのものを指すと解釈する場合も多い

２）確率分布

・事象の発生する確率分布

・離散値であれば表に表せる。

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と対応させ和を取った）	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	４／１６	６／１６	４／１６	１／１６

１．３　統計学 2-3（期待値）

期待値はおおむね平均値

・平均は足し合わせて個数で割る

・期待値はその分布における、確率変数の平均の値又は「ありそう」な値

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

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

連続する値ならば

期待値$$E(f)=\int P(X=x_k)\, f(X=x_k)\, dx$$

１．４　統計学 2-4（分散と共分散）

１）分散＝平均との差の二乗

・データの散らばり具合

・データの各々の値が、期待値からどれだけズレているのか平均したもの

分散Ｖａｒ（ｆ）＝Ｅ（(ｆ（Ｘ＝ｘ）－Ｅ（ｆ))２乗)

＝Ｅ（ｆ（Ｘ－ｘ）２乗)－（Ｅ（ｆ))２乗

もともとは絶対値であったが、計算が面倒なので２乗した。

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

２）共分散

・２つのデータ系列の傾向の違い

・正の値を取れば似た傾向

・負の値を取れば逆の傾向

・ゼロを取れば関係性に乏しい

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

１．５　統計学 2-5（分散と標準偏差）

分散は２乗してしまっているので元のデータと単位が違う
２乗することの逆演算（つまり平行根を求める）をすれば元の単に戻る
標準偏差

$$\sigma = \sqrt{\mathrm{Var}(f)}$$
$$= \sqrt{\mathrm{E}\Big(\big(f_{(X=x)} - \mathrm{E}_{(f)}\big)^2\Big)}$$

１．６　統計学 2-6（様々な確率分布Ｉ・ベルヌーイ分布）

確率分布
１）ベルヌーイ分布
・コイントスのイメージ
・裏と表で出る割合が等しくなくとも扱える
μ：平均の値－＞x が１の時（表の確率）

$$P(x|\mu) = \mu^x(1-\mu)^{1-x}$$

x＝０：裏の確率

２）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値くても扱える

１．７　統計学 2-7（様々な確率分布Ⅱ）

１）二項分布
・ベルヌーイ分布の多試行版

$$P(x|\lambda, n)$$
$$= \frac{n!}{x!\,(n-x)!}\lambda^x(1-\lambda)^{n-x}$$

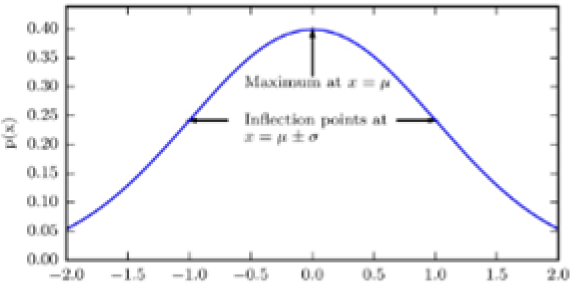
２項係数　　　　　　表　　　　裏
全てが表、裏の場合は係数はいらない。

パターン分係数をかける。
ｘとｎ－ｘが同じぐらいになった時一番大きい

２）ガウス分布
・釣鐘型の連続分布
・人工的に２項分布を作ったもの

$$\mathcal{N}(x;\mu,\sigma^2)=\sqrt{\frac{1}{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$$

・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
・全て足し合わせると１になるように係数を付けている



１．８　統計学 2-8（推定）

推定：母集団を特徴づける母数（パラメーター：平均など）を統計的に推測すること。
　母数：母集団が持っている数

１）点推定
　平均値などを１つの値に推定すること。

２）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

１．９　統計学 2-9（推定量と推定値）

１）推定量（ｅｓｔｉｍａｔｏｒ）
　パラメータ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（関数)。
　推定関数ともいう。

２）推定値（ｅｓｔｉｍａｔｅ）
　実際に試行を行った結果から計算した値。

真の値をθとすると、推定値はθﾊｯﾄ（$\hat{\theta}$）のように表す

１．１０　統計学 2-10（標本平均）

標本平均

　母集団から取り出した標本の平均値

１）サンプル数が大きくなれば、母集団に近づく。―――＞一致性

２）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。―――＞不遍性

$$E\left(\hat{\theta}\right) = \theta$$

１．１１　統計学 2-11（標本分散）

１）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。

たくさんのデータのバラつき具合は、少数のデータのバラつきより大きくなる。

２）不偏分散

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

　　　　　　値を修正している

　母集団の分散に近づける

１．１２　統計学 2-12（増えた量は同じなのに、何が違う）

情報に関する考え方・情報科学

１）個数の違い

　１０，１１（１個変化）　Δｗ＝１　　１／１０

　１，２（１個変化）　　　Δｗ＝１　　１／２０

　基の量に対しての増えた量の比率

２）情報の感じ方

３）１／ｗの積分ｌｏｇ（ｗ）

１．１３　統計学 2-13（自己情報量）

自己情報量

　Ｉ（ｘ）＝－ｌｏｇ（Ｐ（ｘ））＝ｌｏｇ（Ｗ（ｘ））

　　Ｗ()＝１／Ｐ()

　低により単位を変える

１）対数の底が２の時、単位はビット（bit）

２）対数の底がネイピアｅの時、単位はナット（nat）

１．１４　統計学 2-14（シャノンエントロピー）

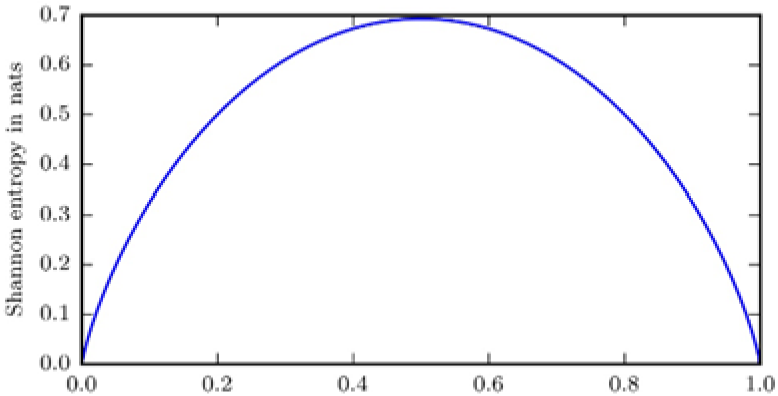
シャノンエントロピー

・微分エントロピーともいうが、部分しているわけでもない

・自己情報量の期待値

$$
\begin{aligned}
H(x) &= E\big(I(x)\big) \\
&= -E\Big(\log\big(P(x)\big)\Big) \\
&= -\sum\Big(P(x)\log\big(P(x)\big)\Big)
\end{aligned}
$$

コイン投げの一例



必ず裏　　　　　　　　　　　　　　　　　　　　　　　　必ず表

・情報量が最大になるようなところ

１．１５　統計学 2-15（カルバック・ライブラー）

カルバック・ライブラー　ダイバージェンス

・同じ事象・確率変数における異なる確率分布 P,Q の違いを表す。

$$D_{\mathrm{KL}}(P\|Q) = \mathbb{E}_{\mathrm{x}\sim P}\left[\log \frac{P(x)}{Q(x)}\right] = \mathbb{E}_{\mathrm{x}\sim P}\left[\log P(x) - \log Q(x)\right]$$

平均

コインを投げる、裏と表は同じ確率であるが、実際確率が違う場合に使用する。

ダイバージェンス：距離に近い概念

$$I(Q(x)) - I(P(x))$$
$$= (-\log(Q(x))) - (-\log(P(x))) = \log \frac{P(x)}{Q(x)}$$

I(Q)　当選最初の情報

I(P)　後になってわかった情報

期待値

$$E(f(x)) = \sum_x P(x)f(x)$$

情報利得

$$D_{\mathrm{KL}}(P\|Q) = \sum_x P(x)\,(-\log(Q(x))) - (-\log(P(x))) = \sum_x P(x)\log \frac{P(x)}{Q(x)}$$

自己情報量の期待値に似ている。

１．１６　統計学 2-16（交差エントロピー）

交差エントロピー

・KL ダイバージェンスの一部分を取り出したもの

・Q についての自己情報量を P の分布で平均している

$$D_{\mathrm{KL}}(P\|Q) = \sum_{x} P(x)\left(-\log(Q(x))\right) - \left(-\log(P(x))\right)$$

$$\boxed{H(P,Q)} = \boxed{H(P)} + D_{\mathrm{KL}}(P\|Q)$$

$$H(P,Q) = -\mathbb{E}_{\mathrm{x}\sim P}\log Q(x) = -\sum_{x} P(x)\log Q(x)$$