第４章　確率分布と統計的な推測

第１節　確率分布

１　確率変数と確率分布

1　確率変数と確率分布

確率変数　試行の結果によってその値が定まり，各値に対応して確率が定まるような変数を確率変数という。

確率分布　確率変数 X のとりうる値とその値のとる確率との対応関係を，X の確率分布という。確率変数 X のとりうる値が x_1，x_2，……，x_n であり，それぞれの値をとる確率が p_1，p_2，……，p_n であるとき，次が成り立つ。

$p_1 \geqq 0$，$p_2 \geqq 0$，……，$p_n \geqq 0$　　　$p_1+p_2+\cdots\cdots+p_n=1$

２　確率変数と期待値と分散　　３　確率変数の和と積

1　確率変数と期待値

確率変数 X の確率分布が右の表で与えられているとき，X の期待値 $E(X)$ は

X	x_1	x_2	$\cdots$	x_n	計
P	p_1	p_2	$\cdots$	p_n	1

$$E(X)=x_1p_1+x_2p_2+\cdots\cdots+x_np_n=\sum_{k=1}^{n}x_kp_k$$

2　確率変数の分散と標準偏差

確率変数 X の確率分布が右上の表で与えられているとする。

分散　$E(X)=m$ とするとき，分散 $V(X)$ は

$$V(X)=(x_1-m)^2p_1+(x_2-m)^2p_2+\cdots\cdots+(x_n-m)^2p_n$$
$$=\sum_{k=1}^{n}(x_k-m)^2p_k=E((X-m)^2)$$
$$V(X)=E(X^2)-m^2=E(X^2)-\{E(X)\}^2$$

標準偏差　標準偏差 $\sigma(X)=\sqrt{V(X)}$

3　$aX+b$ の期待値，分散，標準偏差

X を確率変数，a，b を定数とするとき

$$E(aX+b)=aE(X)+b$$
$$V(aX+b)=a^2V(X)$$
$$\sigma(aX+b)=\sqrt{V(ax+b)}=|a|\sigma(X)$$

4　同時分布

2つの確率変数 X，Y について，$P(X=x_i,\ Y=y_j)=p_{ij}$ とするとき，$(x_i,\ y_j)$ と p_{ij} の対応を X と Y の **同時分布** という。

$P(X=x_i,\ Y=y_j)=P(X=x_i)P(Y=y_j)$

が常に成り立つとき，X，Y は互いに **独立** であるという。

$X \backslash Y$	y_1	y_2	$\cdots$	y_m	計
x_1	p_{11}	p_{12}	$\cdots$	p_{1m}	p_1
x_2	p_{21}	p_{22}	$\cdots$	p_{2m}	p_2
$\vdots$		………			$\vdots$
x_n	p_{n1}	p_{n2}		p_{nm}	p_n
計	q_1	q_2	$\cdots$	q_n	1

5　確率変数の和の期待値

2つの確率変数 X，Y について

$$E(X+Y)=E(X)+E(Y)$$
$$E(aX+bY)=aE(X)+bE(Y)\qquad(a,\ b \text{ は定数})$$

注　3つ以上の確率変数でも同様のことが成り立つ。

6　独立な２つの確率変数

2つの確率変数 X，Y が互いに独立であるとき

$$E(XY)=E(X)E(Y)$$
$$V(X+Y)=V(X)+V(Y)$$
$$V(aX+bY)=a^2V(X)+b^2V(Y)\qquad(a,\ b \text{ は定数})$$

注　3つ以上の確率変数でも同様のことが成り立つ。

7　事象の独立

2つの事象 A，B が互いに独立 $\Longleftrightarrow$ $P(A\cap B)=P(A)P(B)$

４　二項分布

1　二項分布

二項分布 $B(n,\ p)$ とは，次の式で与えられる確率分布である。

$P(X=r)={}_n\mathrm{C}_r p^r q^{n-r}$　$(r=0,\ 1,\ 2,\ \cdots\cdots,\ n)$　　ただし，$q=1-p$

2　二項分布に従う確率変数の期待値，分散，標準偏差

確率変数 X が二項分布 $B(n,\ p)$ に従うとき

期待値は　　$E(X)=np$

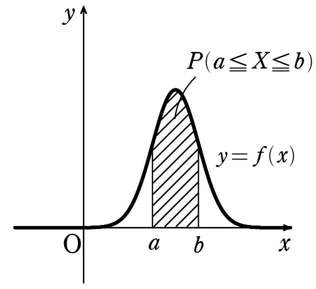
分散は　　$V(X)=npq$　　ただし，$q=1-p$

標準偏差は　$\sigma(X)=\sqrt{npq}$

５　正規分布

1　連続した値をとる確率変数

X を連続した値をとる確率変数，$f(x)$ を X の確率密度関数とすると，$f(x)$ は次のような性質をもつ。



1　常に $f(x)\geqq 0$ で

$$P(a\leqq X\leqq b)=\int_a^b f(x)\,dx$$

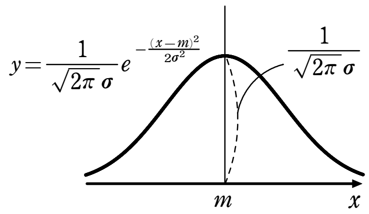
2　X のとる値の範囲が $\alpha\leqq X\leqq\beta$ のとき

$$\int_\alpha^\beta f(x)\,dx=1$$

期待値　$E(X)=m=\int_\alpha^\beta xf(x)\,dx$　　　分散　$V(X)=\int_\alpha^\beta (x-m)^2f(x)\,dx$

2　正規分布 $N(m,\ \sigma^2)$

確率変数 X が正規分布 $N(m,\ \sigma^2)$ に従うとき



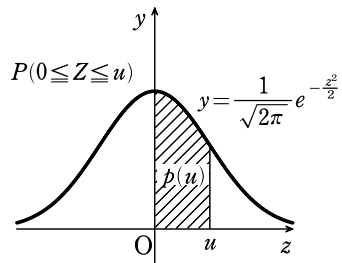
1　期待値は　　$E(X)=m$
　標準偏差は　$\sigma(X)=\sigma$

2　分布曲線 $y=f(x)$ は直線 $x=m$ に関して対称であり，y は $x=m$ で最大値をとる。
　また，x 軸を漸近線とし，標準偏差 σ が大きくなると曲線の山は低くなって横に広がる。

3　標準正規分布 $N(0,\ 1)$

1　確率変数 X が正規分布 $N(m,\ \sigma^2)$ に従うとき，$Z=\dfrac{X-m}{\sigma}$ とおくと，確率変数 Z は標準正規分布 $N(0,\ 1)$ に従う。

2　確率変数 Z が標準正規分布 $N(0,\ 1)$ に従うとする。
　$u\geqq 0$，$v\geqq 0$ とし，$P(0\leqq Z\leqq u)=p(u)$ とすると，たとえば，次のことが成り立つ。



$P(u\leqq Z\leqq v)=p(v)-p(u)$
$P(Z\geqq 0)=0.5$，$P(Z\leqq 0)=0.5$
$P(Z\geqq u)=0.5-p(u)$
$P(-v\leqq Z\leqq -u)=p(v)-p(u)$
$P(-u\leqq Z\leqq v)=p(u)+p(v)$

4　二項分布の正規分布による近似

1　二項分布 $B(n,\ p)$ に従う確率変数 X は，n が大きいとき，近似的に正規分布 $N(np,\ npq)$ に従う。ただし，$q=1-p$ である。

2　二項分布 $B(n,\ p)$ に従う確率変数 X に対し，$Z=\dfrac{X-np}{\sqrt{npq}}$ は，n が大きいとき，近似的に標準正規分布 $N(0,\ 1)$ に従う。ただし，$q=1-p$ である。

第２節　統計的な推測

６　母集団と標本　　７　標本平均の分布

1　全数調査と標本調査

全数調査　調査の対象全体からデータを集めて調べる方法

標本調査　調査の対象全体からその一部を抜き出して調べる方法

2　母集団分布

母集団における変量 x の分布を母集団分布，その平均値を母平均，標準偏差を母標準偏差という。大きさ1の無作為標本における変量 x の値を X とすると，X は確率変数であり，この X の確率分布，期待値，標準偏差は，それぞれ，母集団分布，母平均，母標準偏差と一致する。

3　標本平均の期待値と標準偏差

母平均 m，母標準偏差 σ の母集団から大きさ n の無作為標本を抽出するとき

標本平均 $\overline{X}=\dfrac{X_1+X_2+\cdots\cdots+X_n}{n}$，　**期待値** $E(\overline{X})=m$，　**標準偏差** $\sigma(\overline{X})=\dfrac{\sigma}{\sqrt{n}}$

4　標本平均の分布と正規分布

母平均 m，母標準偏差 σ の母集団から抽出された大きさ n の無作為標本について，

1　標本平均 $\overline{X}$ は，n が大きいとき，近似的に正規分布 $N\left(m,\ \dfrac{\sigma^2}{n}\right)$ に従うとみなすことができる。

2　標本平均 $\overline{X}$ に対して，$Z=\dfrac{\overline{X}-m}{\dfrac{\sigma}{\sqrt{n}}}$ は n が大きいとき，近似的に標準正規分布 $N(0,\ 1)$ に従う。

5　大数の法則

母平均 m の母集団から大きさ n の無作為標本を抽出するとき，n が大きくなるに従って，その標本平均 $\overline{X}$ は母平均 m に近づく。

８　推定

1　母平均の推定

母標準偏差を σ とする。標本の大きさ n が大きいとき，母平均 m に対する信頼度 95 % の信頼区間は　$\left[\overline{X}-1.96\cdot\dfrac{\sigma}{\sqrt{n}},\ \overline{X}+1.96\cdot\dfrac{\sigma}{\sqrt{n}}\right]$

注　上の σ は母標準偏差であるが，これが不明の場合は，代わりに標本の標準偏差 S を用いてよい。

2　母比率の推定

標本の大きさ n が大きいとき，標本比率を R とすると，母比率 p に対する信頼度 95 % の信頼区間は　$\left[R-1.96\sqrt{\dfrac{R(1-R)}{n}},\ R+1.96\sqrt{\dfrac{R(1-R)}{n}}\right]$