応用システム工学
第七回　回帰分析

平成24年07月20日
標本平均

確率変数の性質

- 確率変数Xの関数g(X)の期待値E[g(X)]
  - 確率密度関数:確率変数Xが値xをとる確率f(x)

$$E[g(X)] = \sum_x g(x) f(x)$$

    - 確率変数の期待値E[X]→g(x)=x
      (原点の周りの1次のモーメント(積率))
    - 確率変数の分散V[X]→g(x)=(x-μ)^2
      (期待値の周りの2次のモーメント)
  - 定数cに対する性質
    - 期待値 $E[cX] = \sum_x cxf(x) = c\sum_x xf(x) = cE[X]$
    - 分散$V[cX] = \sum_x (cx - c\mu)^2 f(x) = c^2 \sum_x (x-\mu)^2 f(x) = c^2 V[X]$

同時確率分布

- 同時確率分布
  - 確率変数Xの値がxとなり，同時に確率変数Yの値がyとなる確率分布
  - X,Yの同時確率密度関数　$f_{XY}(x,y)$
- 周辺確率分布
  - 同時確率分布から，ある確率変数の分布を抽出したもの
  - 確率変数Xの周辺確率密度関数　$f_X(x)$

$$f_X(x)=\int_{-\infty}^{\infty} f_{XY}(x,y)dy$$

複数の確率変数の性質

- 2つの確率変数の和に対する期待値

$$\begin{aligned}
E[X+Y] &= \int_x\int_y (x+y) f_{XY}(x,y)dxdy \\
&= \int_x\int_y xf_{XY}(x,y)dxdy + \int_x\int_y yf_{XY}(x,y)dxdy \\
&= \int_x x\int_y f_{XY}(x,y)dydx + \int_x y\int_y f_{XY}(x,y)dxdy \\
&= \int_x xf_X(x)dx + \int_x yf_Y(y)dy \\
&= E[X]+E[Y]
\end{aligned}$$

複数の確率変数の性質

- 確率変数X,Yが独立である場合における，
同時確率密度関数$f_{XY}(x,y)$と，
周辺確率密度関数 $f_X(x), f_Y(y)$ の関係

$$f_{XY}(x,y) = f_X(x) f_Y(y)$$

- 独立な確率変数の積の期待値

$$E[XY] = \int_x\int_y xyf_{XY}(x,y)dxdy = \int_x\int_y xf_X(x)yf_Y(y)dxdy$$

$$= \int_x xf_X(x)dx\int_y yf_Y(y)dy = E[X]E[Y]$$

  – この時共分散は

$$Cov[X,Y] = E[(X-E[X])(Y-E[Y])] = E[XY - E[X]Y - XE[Y] + E[X]E[Y]]$$

$$= E[XY] - E[X]E[Y] - E[X]E[Y] + E[X]E[Y]$$

$$= E[XY] - E[X]E[Y] = 0$$

複数の確率変数の性質

- 分散と期待値の関係

$$V[X] = E\left[(X-E[X])^2\right] = E\left[X^2 - 2XE[X] + E[X]^2\right]$$

$$= E\left[X^2\right] - 2E[X]E[X] + E[X]^2 = E\left[X^2\right] - E[X]^2$$

- 2つの確率変数の和に対する分散

$$V[X+Y] = E\left[(X+Y)^2\right] - E[X+Y]^2 = E\left[X^2 + 2XY + Y^2\right] - (E[X] + E[Y])^2$$

$$= E\left[X^2\right] + 2E[XY] + E\left[Y^2\right] - \left(E[X]^2 + 2E[X]E[Y] + E[Y]^2\right)$$

$$= \left(E\left[X^2\right] - E[X]^2\right) + \left(E\left[Y^2\right] - E[Y]^2\right) + 2(E[XY] - E[X]E[Y])$$

  – 2つの確率変数が独立の場合

$$V[X+Y] = V[X] + V[Y] + 2(E[X]E[Y] - E[X]E[Y])$$

$$= V[X] + V[Y]$$

標本平均

- 標本平均の期待値

$$E\left[\overline{X}_n\right]=E\left[\frac{X_1+X_2+\cdots+X_n}{n}\right]=\frac{1}{n}\left\{E\left[X_1\right]+E\left[X_2\right]+\cdots+E\left[X_n\right]\right\}$$

$$=\frac{1}{n}\left\{\mu+\mu+\cdots+\mu\right\}=\mu$$

- 標本平均の分散

$$V\left[\overline{X}_n\right]=V\left[\frac{X_1+X_2+\cdots+X_n}{n}\right]=\frac{1}{n^2}V\left[X_1+X_2+\cdots+X_n\right]$$

$$=\frac{1}{n^2}\left\{V\left[X_1\right]+V\left[X_2\right]+\cdots+V\left[X_n\right]\right\}=\frac{1}{n^2}\left\{\sigma^2+\sigma^2+\cdots+\sigma^2\right\}$$

$$=\frac{1}{n^2}n\sigma^2=\frac{1}{n}\sigma^2$$

→標本の数nが小さいと分散が大きい
標本の数を大きくすると期待値に近づく

チェビシェフの不等式

- 集合 $I=\left\{x:\left|x-E\left[x\right]\right|\geq k\sqrt{V\left[x\right]}\right\}$ ただしk>0
- 確率変数Xがxとなる確率P(X=x)の確率頻度関数f(x) $P\left[\left|X-E\left[X\right]\right|\geq k\sqrt{V\left[X\right]}\right]=\sum_{x\in I}f(x)$
  - 確率Pで分散を表す

$$V\left[X\right]=\sum_{x}\left(x-E\left[X\right]\right)^2 f(x)\geq\sum_{x\in I}\left(x-E\left[X\right]\right)^2 f(x)$$

$$\geq\sum_{x\in I}\left(k\sqrt{V\left[X\right]}\right)^2 f(x)$$

$$=k^2V\left[X\right]\sum_{x\in I}f(x)=k^2V\left[X\right]P\left[\left|X-E\left[X\right]\right|\geq k\sqrt{V\left[X\right]}\right]$$

チェビシェフの不等式

– 確率Pで分散を表す

$$V[X] \geq k^2 V[X] P\left[\left|X - E[X]\right| \geq k\sqrt{V[X]}\right]$$

$$\frac{1}{k^2} \geq P\left[\left|X - E[X]\right| \geq k\sqrt{V[X]}\right]$$

- 確率変数と期待値の差が分散の根より大きくなる確率は小さい(1/k^2)

確率収束

- 確率変数として標本平均を考える
  – 期待値と分散 $E[\bar{X}_n] = \mu$ $V[\bar{X}_n] = \frac{1}{n}\sigma^2$
  – チェビシェフの不等式

$$\frac{1}{k^2} \geq P\left[\left|\bar{X}_n - E[\bar{X}_n]\right| \geq k\sqrt{V[\bar{X}_n]}\right] = P\left[\left|\bar{X}_n - \mu\right| \geq k\sqrt{\frac{\sigma^2}{n}}\right]$$

$\varepsilon = k\sqrt{\frac{\sigma^2}{n}}$ とすると $\frac{1}{k^2} = \frac{\sigma^2}{n\varepsilon^2} \geq P\left[\left|\bar{X}_n - \mu\right| \geq k\sqrt{\frac{\sigma^2}{n}}\right]$

$$\lim_{n\to\infty} P\left[\left|\bar{X}_n - \mu\right| \geq k\sqrt{\frac{\sigma^2}{n}}\right] = \lim_{n\to\infty} \frac{\sigma^2}{n\varepsilon^2} = 0$$

標本サイズnを大きくすると，標本平均と母平均に差がある確率が0となる

不変分散

- 一致推定量→標本サイズが十分大きいとき，標本平均が母平均に確率収束する性質をもつ推定量→標本平均は母平均に確率収束
- 不変推定量→期待値が母数に等しい推定量
  - 標本平均 $\overline{X}$ は不偏推定量
    - 母平均 μ に対して $E\left[\overline{X}\right]=\mu$

標本平均は
母平均と等しくなる

  - 標本の分散は不偏推定量となるか？
    - 標本の分散

$$S^2=\frac{\left(X_1-\overline{X}\right)^2+\left(X_2-\overline{X}\right)^2+\cdots+\left(X_n-\overline{X}\right)^2}{n}$$

不変分散

- 標本の分散

$$S^2=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2$$
$$=\frac{1}{n}\sum_{i=1}^{n}\left[\left(X_i-\mu\right)-\left(\overline{X}-\mu\right)\right]^2$$
$$=\frac{1}{n}\sum_{i=1}^{n}\left[Y_i-\overline{Y}\right]^2$$
$$=\frac{1}{n}\sum_{i=1}^{n}Y_i^2-2\frac{\overline{Y}}{n}\sum_{i=1}^{n}Y_i+\frac{\overline{Y}^2}{n}\sum_{i=1}^{n}1$$
$$=\frac{1}{n}\sum_{i=1}^{n}Y_i^2-2\overline{Y}\overline{Y}+\frac{\overline{Y}^2n}{n}=\frac{1}{n}\sum_{i=1}^{n}Y_i^2-\overline{Y}^2$$

変数変換

$$Y_i=X_i-\mu$$
$$\overline{Y}=\overline{X}-\mu$$

不変分散

- 標本の分散の期待値

$$E\left[S^2\right]=E\left[\frac{1}{n}\sum_{i=1}^{n}Y_i^2-\overline{Y}^2\right]=E\left[\frac{1}{n}\sum_{i=1}^{n}Y_i^2\right]-E\left[\overline{Y}^2\right]$$

$$=\frac{1}{n}\sum_{i=1}^{n}E\left[Y_i^2\right]-E\left[\overline{Y}^2\right]=\frac{1}{n}\sum_{i=1}^{n}E\left[\left(X_i-\mu\right)^2\right]-E\left[\left(\overline{X}-\mu\right)^2\right]$$

  – 標本サイズが大きいときの標本と母平均の差の二乗平均

$$E\left[\left(X_i-\mu\right)^2\right]\rightarrow\sigma^2$$

  – 標本平均の分散

$$E\left[\left(\overline{X}-\mu\right)^2\right]=\frac{\sigma^2}{n}$$

不変分散

- 標本分散の期待値

$$E\left[S^2\right]=\frac{1}{n}\sum_{i=1}^{n}\sigma^2-\frac{\sigma^2}{n}$$

$$=\frac{1}{n}n\sigma^2-\frac{\sigma^2}{n}=\frac{n-1}{n}\sigma^2$$

  – 標本分散 S^2 の期待値 $E\left[S^2\right]$ は母分散 σ^2 とは等しくない
  →標本分散 S^2 は母分散の不偏推定量ではない
  $\frac{n}{n-1}S^2$ が母分散の不偏推定量に相当

- 不偏分散（自由度n-1）

$$s^2=\frac{\left(X_1-\overline{X}\right)^2+\left(X_2-\overline{X}\right)^2+\cdots+\left(X_n-\overline{X}\right)^2}{n-1}$$

レポート

- 課題内容
  - 電気システム･電子デバイス等の信頼性は，故障率曲線（バスタブカーブ）で表現される。信頼性解析に用いる確率分布である，指数分布，正規分布，対数正規分布，ワイブル分布および，バスタブカーブについて調べ報告せよ。
- 制約条件
  - 枚数　15枚　→　満点で15点
  - レポート締切　7/31
  - 提出先　E2-111