中間試験のお知らせ

日時: 6 月 5 日 (金) 10:40〜
場所: いつもの講義室
範囲: 第 6 回講義まで
持ち込み禁止
遅刻・欠席はその原因を証明する公的書類を提出した場合だけ追試験または試験時間の延長を行う
試験の傾向：演習問題の類題
暗記するべき定義：累積分布関数、同時確率密度関数、同時確率質量関数、条件付き確率、期待値、分散、標準偏差、標準化変数、共分散、相関係数、統計的独立性
他に暗記するべき事項：大数の法則、チェビシェフの不等式

2009-6-1

正規分布とその性質 (教科書の6.6節)

確率密度関数が

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\{-(x-m)^2/2\sigma^2\}$$

で与えられる確率分布を正規分布と呼び $N(m, \sigma^2)$ と表記する。正規分布に従う確率変数 X の期待値は m で分散は σ^2 である。この確率密度関数のかたちは教科書 121 ページを参照。
また以下の事実が知られている

$$\begin{aligned} P(m-\sigma \le X \le m+\sigma) &\simeq 0.6827, \\ P(m-2\sigma \le X \le m+2\sigma) &\simeq 0.9545, \\ P(m-3\sigma \le X \le m+3\sigma) &\simeq 0.9973, \\ P(m-4\sigma \le X \le m+4\sigma) &\simeq 0.9999 \end{aligned}$$

$[m-3\sigma, m+3\sigma]$ を 3 シグマ範囲と呼び、確率変数が正規分布に従うなら実現値は通常この範囲に入る。また正規分布の特徴として、期待値から離れた実現値が得られる確率は、期待値と実現値の差に対して指数関数的に減少するため、めったに起こらない。

2009-6-2

この分布は極めて重要である。どんな確率変数列 X_1, X_2, ... でも X_i が同一かつ独立な確率分布に従うときには $X_1 + \cdots + X_n$ の分布の形が正規分布に似てくるという驚異的な事実が知られている。

分布の形が似てくるとは何か? (教科書の162ページの下)

確率変数列 $X_1, X_2, \ldots$ と確率変数 Y があるとする。すべての実数 x について

$$\lim_{n\to\infty} F_{X_n}(x) = F_Y(x)$$

が成立するとき、X_n が Y に法則収束すると言う。

ここで X_n は離散型確率変数で Y が連続型確率変数であっても構わない。中心極限定理ではそのようになっている。その場合、X_n には確率密度関数は存在せず、Y の確率質量関数は常に0である。従って「分布の形が似てくる」ことを定義するために、確率密度関数または確率質量関数を使用することができない。このことが、上記の定義で累積分布関数を使用することの理由である。

2009-6-3

中心極限定理 (教科書の 8.2 節)

$X_1,\ X_2,\ \ldots$ が統計的に独立な確率変数で同一の分布に従うとする。X_i の標準化変数を Z_i とする。つまり

$$Z_i = \frac{X_i - E[X_i]}{\sqrt{V[X_i]}}$$

このとき

$$A_i = \frac{Z_1 + \cdots + Z_n}{\sqrt{n}} = \frac{X_1 + \cdots + X_n - nE[X_1]}{\sqrt{V[X_1]}\sqrt{n}}$$

は平均 0 分散 1 の正規分布を持つ確率変数に法則収束する。
このことは、大雑把に言えば、独立かつ同一の分布に従う確率変数をたくさん足すと、分布の形が正規分布に近づいていくことをしめす。現実世界にこのような多数の確率変数の和として表わされる事象は多いので、中心極限定理は分布に正規分布を仮定する根拠になっている。

2009-6-4

中心極限定理の実際の例

室内で煙が立ち昇るとき、風が無いのも関わらず煙はまっすぐ昇らない。煙が斜めに昇る理由は、煙の粒子に多数の空気の分子が衝突して軌道を変えているからである。一つの空気の分子の衝突による煙の粒子の水平方向の座標の変化を確率変数 X_i で表現すると、ある程度煙が昇ったときの水平方向の座標は多数の X_i の和になるから、煙の粒子の水平座標は中心極限定理より正規分布に従うことが予想される。実際に煙の粒子の水平座標の分布を調べてみると正規分布に近い形をしている。

2009-6-5

確率変数の和の分布 (教科書150ページ)

X と Y が確率変数とする (独立性はまだ仮定しない)。$X+Y$ の確率質量関数または確率密度関数を求めることを考えたい。まず X, Y は離散型とする。
$X+Y$ が z になる場合は X が x になり Y が $z-x$ になるときである。これをすべての x について総和をとり

$$P(X+Y=z) = P_{X+Y}(z) = \sum_x P_{XY}(x, z-x)$$

となる。ここで x はすべての X の実現値を動く。
とくに X と Y が統計的に独立であるときには

$$\sum_x P_{XY}(x, z-x) = \sum_x P_X(x) P_Y(z-x)$$

である。P_X と P_Y から上記の右辺を計算する操作を畳み込みと呼ぶ。

2009-6-6

確率変数の和の分布 (連続型)

つぎに X, Y が連続型のとき $Z = X + Y$ の確率密度関数を求めたい。
Z が z 以下の値をとることと $X + Y \leq z$ は同値条件である。従って

$$\begin{aligned} & P(X + Y \leq z) \\ = & \int_{x+y \leq z} f_{XY}(x, y) dx dy (\text{面積分}) \\ = & \int_{-\infty}^{z} \int_{-\infty}^{\infty} f_{XY}(x, z' - x) dx dz' \end{aligned}$$

上の式を図を書いて説明する。
上の式を z で微分すると $X + Y$ の確率密度関数が得られるので、$Z = X + Y$ の確率密度関数は

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_{XY}(x, z - x) dx$$

になる。
とくに X と Y が統計的に独立であるときに上の式は

$$\int_{-\infty}^{\infty} f_X(x) f_Y(z - x) dx$$

確率変数の和の分布 (連続型) の例

U と V を $[0,1]$ に値をとる一様連続確率変数とし、U と V は統計的に独立であるとする。U の確率密度関数は

$$f_U(u) = \begin{cases} 1 & \text{if } 0 \leq u \leq 1, \\ 0 & \text{その他} \end{cases}$$

となり V の確率密度関数も同じである。このとき $U+V$ の確率密度関数 $f_{U+V}(z)$ は

$$\int_{-\infty}^{\infty} f_U(u) f_V(z-u) du \tag{1}$$

である。このとき被積分関数が非0になる条件は「$0 \leq u \leq 1$ かつ $0 \leq z-u \leq 1$」である。これは「$0 \leq u \leq 1$ かつ $z-1 \leq u \leq z$」と同値である。従って被積分関数が非0になる u の範囲の長さは

$$\begin{array}{ll} z & \text{if } 0 \leq z \leq 1, \\ (2-z) & \text{if } 1 \leq z \leq 2, \\ 0 & \text{その他} \end{array}$$

である。被積分関数は非0のときつねに1なので上記の関数は式 (1) の積分値に等しい。つまり $U+V$ の確率密度関数である。

2009-6-8

正規分布の和の確率密度関数 (151ページ)

確率変数 X が $N(0,1)$, Y が $N(0,1)$ に従い統計的に独立であるとする。このとき $Z = X + Y$ の確率密度関数は

$$
\begin{aligned}
& \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx \\
= & \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2}\right) \exp\left(-\frac{(z-x)^2}{2}\right) dx \\
= & \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\left(-\frac{x^2+(z-x)^2}{2}\right) dx \\
= & \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\left(-\frac{2x^2-2zx+z^2}{2}\right) dx \\
= & \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\left(-\frac{2(x-z/2)^2+z^2/2}{2}\right) dx \\
= & \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\left(-\frac{2(x-z/2)^2}{2}\right) dx \exp\left(-\frac{z^2}{4}\right) \\
= & \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\left(-x^2\right) dx \exp\left(-\frac{z^2}{4}\right) \\
= & \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \exp\left(-x^2\right) dx \frac{1}{2\sqrt{\pi}} \exp\left(-\frac{z^2}{4}\right)
\end{aligned}
$$

2009-6-9

$$= \frac{1}{2\sqrt{\pi}} \exp\left(-\frac{z^2}{4}\right)$$

最後の式は、分散 2 平均 0 の正規分布の確率密度関数である。正規分布に従う独立確率変数の和は正規分布に従うが、これを正規分布の再生性と呼ぶ。

演習問題

51. U を $[-2,2]$ に値をとる一様連続確率変数, V を $[-3,3]$ に値をとる一様連続確率変数とし、U と V は統計的に独立とする。$U+V$ の確率密度関数を求め図示せよ。 値がマイナスになる関数を答えないこと。

52. $U+V$ の累積分布関数を求め図示せよ。引数の値が大きいときに関数の値が 1 になる関数を答えること。

53. X を期待値 m_X 分散 σ_X^2, Y を期待値 m_Y, 分散 σ_Y^2 の正規確率変数とし、X と Y は統計的に独立と仮定する。$X+Y$ の確率密度関数を求め $X+Y$ が期待値 m_X+m_Y, 分散 $\sigma_X^2+\sigma_Y^2$ の正規分布に従っていることを確認せよ。

54. 授業でわかりにくかった所や要望を書いて下さい。また前回の答案で×になったところで、理由がわからないところは申し出て下さい。