

金曜 5 限基礎統計 2008 年期末試験

これは、講義のホームページにも載っていますが、一応載せておきます。

カイ二乗分布の下側確率については、次の数値を用いてよい。

自由度	下側		下側		上側		上側	
	0.5%	下側 1%	2.5%	下側 5%	上側 5%	2.5%	上側 1%	0.5%
1	0.00004	0.0003	0.001	0.003	3.84	5.02	6.63	7.88
2	0.01	0.02	0.051	0.103	5.99	7.38	9.21	10.6
3	0.072	0.115	0.216	0.352	7.81	9.35	11.34	12.84
4	0.207	0.297	0.484	0.711	9.49	11.14	13.28	14.86
5	0.412	0.554	0.831	1.145	11.07	12.83	15.09	16.75

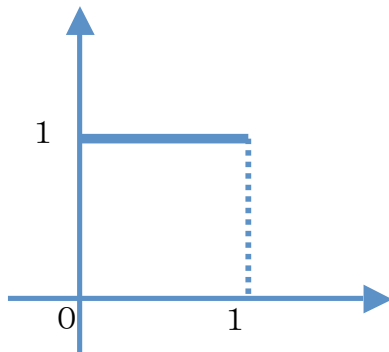
- 1, 互いに独立な確率変数  $X_1, X_2$  が、区間  $(0, 1)$  の一様分布に従うとき、 $Y = \min(X_1, X_2)$  で定義される確率変数  $Y$  の累積分布関数と密度関数を求めよ。 $\min(a, b)$  は、 $a, b$  の小さい方の値を表す。 $(a=b)$  のときは、どちらをとるとしても同じ。

【解答】  $(0, 1)$  の一様分布に従う確率変数  $X$  について、 $P(X < t) = t$  だから、

累積分布関数  $F(t)$  は、

$$\begin{aligned}
 F(t) &= P(\min(X_1, X_2) < t) \\
 &= 1 - P(\min(X_1, X_2) > t) \\
 &= 1 - P(X_1 > t, X_2 > t) && \because X_1, X_2 \text{ は互いに独立} \\
 &= 1 - P(X_1 > t)P(X_2 > t) \\
 &= 1 - (1-t)(1-t) \\
 &= \underline{\underline{1 - (1-t)^2}}
 \end{aligned}$$

密度関数は、 $f(t) = F'(t) = \underline{\underline{2(1-t)}}$



これが、 $(0, 1)$  の一様分布のグラフです。確かに面積（全確率）が1になっています。

- 2, 確率変数  $X$  が区間  $(0, 1)$  の上の一様分布に従うとき、 $Y = \frac{X}{1-X}$  で定義される確率変数  $Y$  の累積分布関数  $F(t)$  と密度関数  $f(t)$  を求めよ。

【解答】

$$\begin{aligned}
 F(t) &= P(Y < t) \\
 &= P\left(\frac{X}{1-X} < t\right) = P\left(\frac{1-X}{X} > \frac{1}{t}\right) \\
 &= P\left(\frac{1}{X} > \frac{1}{t} + 1\right) = P\left(X < \frac{t}{t+1}\right) = \frac{t}{t+1} \\
 f(t) &= F'(t) = \frac{1}{(t+1)^2}
 \end{aligned}$$

- 3,  $X$ 、 $Y$  を確率変数とする。ただし、 $V(X) > 0$ 、 $V(Y) > 0$  と仮定する。

- (a)  $V(X+cY)$  を最小化する定数  $c$  の値を、 $V(X)$ 、 $V(Y)$ 、 $Cov(X,Y)$  で表せ。

【解答】

$$\begin{aligned}
 V(X+cY) &= V(X) + V(cY) + 2Cov(X,Y) \\
 &= V(X) + c^2V(Y) + 2cCov(X,Y)
 \end{aligned}$$

これを  $c$  で微分してイコール0とおくと、

$$2cV(Y) + 2Cov(X,Y) = 0$$

$$\therefore c = -\frac{Cov(X,Y)}{V(Y)}$$

(b)上の(a)で求めた  $c$  の値を  $c_0$  とする。  $V(X+c_0Y)=V(X)$  となる条件を求めよ。

【解答】  $c_0 = -\frac{Cov(X,Y)}{V(Y)}$  とおく。

$$\begin{aligned} V(X+c_0Y) &= V(X) \\ \Leftrightarrow c_0^2V(Y) + 2c_0Cov(X,Y) &= 0 \\ \Leftrightarrow -Cov(X,Y) + 2Cov(X,Y) &= 0 \\ \Leftrightarrow \underline{\underline{Cov(X,Y) = 0}} \end{aligned}$$

(c)定数  $m$  に対して次の式を示せ。  $E((X-m)^2) = V(X) + (m - E(X))^2$

【解答】

4, 連続型確率変数  $X$ 、 $Y$  が同時確率分布

$$f(x,y) = \frac{4}{5}(xy + x + y) \quad (0 < x < 1, 0 < y < 1)$$

をもつとき、周辺密度関数  $f(x)$ 、 $f(y)$  と、条件付き密度  $f(y | x)$  を求めよ。

【解答】

$$\begin{aligned} f(x) &= \int_0^1 f(x,y) dy \\ &= \int_0^1 \frac{4}{5}(xy + x + y) dy = \frac{4}{5} \left[ \frac{1}{2}x^2 + xy + \frac{1}{2}y^2 \right]_0^1 \\ &= \underline{\underline{\frac{2}{5}(3x+1)}} \quad (0 < x < 1) \end{aligned}$$

$$\text{同様にして、} \quad f(y) = \underline{\underline{\frac{2}{5}(3y+1)}} \quad (0 < y < 1)$$

条件付き密度関数は、

$$f(y|x) = \frac{f(x,y)}{f(x)} = \frac{2(xy + x + y)}{\underline{\underline{3x+1}}}$$

5, 確率変数  $X$  が次の密度関数をもつとき, 期待値と分散を求めよ。

$$f(x) = \begin{cases} x & (0 < x < 1) \\ 2-x & (1 \leq x < 2) \\ 0 & (x \leq 0, 2 \leq x) \end{cases}$$

【解答】

$$\begin{aligned} E(x) &= \int_{-\infty}^{\infty} xf(x)dx \\ &= \int_0^1 x \cdot x dx + \int_1^2 x(2-x) dx \\ &= \underline{\underline{1}} \end{aligned}$$

よく考えると, 密度関数が  $x=1$  について対称なので自明。

分散を定義どおり計算すると,

$$\begin{aligned} V(X) &= E(x - E(X))^2 \\ &= E((x-1)^2) \\ &= \int_{-\infty}^{\infty} (x-1)^2 f(x) dx \\ &= \int_0^1 (x-1)^2 \cdot x dx + \int_1^2 (x-1)^2 (2-x) dx = \underline{\underline{\frac{1}{6}}} \end{aligned}$$

となるが、計算が複雑。公式  $V(X) = E(X^2) - (E(X))^2$  を使うと、

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{\infty} x^2 f(x) dx \\ &= \int_0^1 x^3 dx + \int_1^2 x^2 (2-x) dx = \frac{7}{6} \end{aligned}$$

より、 $V(X) = \frac{7}{6} - 1^2 = \frac{1}{6}$  と求まる。

6, 離散型の確率変数  $X, Y$  はいずれも  $-1, 0, 1$  という値のみをとる。確率変数  $X, Y$  は独立でないが無相関となるように  $\Pr(X=x, Y=y)$  ( $x, y = -1, 0, 1$ ) を定めて, 表に表せ。

【解答】 他にもいろいろできるはず。

	Y=1	Y=0	Y=-1	
X=1	0	1/4	0	1/4
X=0	1/4	0	1/4	1/2
X=-1	0	1/4	0	1/4
	1/4	1/2	1/4	1

7, 未知の期待値  $\mu$ 、分散  $\sigma^2$  の正規分布から、次のような標本の大きさ  $N=5$  のデータを得た。

$$X_1 = 1, X_2 = 1, X_3 = 3, X_4 = 3, X_5 = 2$$

(1) 期待値  $\mu$  の 99% 信頼区間を求めよ。

【解答】 標本平均  $N=2$ 、

$$\text{不偏分散 } s^2 = \frac{1}{5-1} \left( (1-2)^2 + (1-2)^2 + (3-2)^2 + (3-2)^2 + (2-2)^2 \right) = 1$$

$\frac{m-\mu}{\sqrt{s^2/N}}$  は、自由度  $N-1=4$  の  $t$  分布にしたがう。この分布の両側 1%

点は 4.604 であるので、 $\sqrt{s^2/N} = \frac{1}{2.236}$  より、

$$P\left(-4.604 < \frac{\mu - m}{\sqrt{s^2/N}} < 4.604\right) = P\left(2 - \frac{4.604}{2.236} < \mu < 2 + \frac{4.604}{2.236}\right) = 0.99$$

したがって、99% 信頼区間は  $[-0.059, 4.059]$

(2) 仮説  $\mu=0$  を有意水準 5% で棄却せよ。

【解答】  $\frac{m-\mu}{\sqrt{s^2/N}}$  は、自由度  $N-1=4$  の  $t$  分布にしたがう。 $\mu=0$  のとき、この

値は、4.472 となり、自由度 4 の  $t$  分布の両側 5% 点 2.276 を越えている。したがって、 $\mu=0$  である限り、このような値が出ることはまれであり、仮説  $\mu=0$  は棄却される。

(3)分散 $\sigma^2$ の95%信頼区間を求めよ。

【解答】 $\frac{(N-1)s^2}{\sigma^2}$ が自由度 $N-1=4$ のカイ二乗分布にしたがうことより、

与えられた表の上側下側それぞれ2.5%の値を使って、

$$\begin{aligned} P\left(0.484 < \frac{4s^2}{\sigma^2} < 11.14\right) &= P\left(\frac{1}{0.484} > \frac{\sigma^2}{4} > \frac{1}{11.14}\right) \\ &= P\left(\frac{4}{0.484} > \sigma^2 > \frac{4}{11.14}\right) = 0.95 \end{aligned}$$

よって、 $[0.359, 8.264]$

(4)仮説 $\sigma^2=2$ を有意水準1%で両側検定せよ。

【解答】仮説のもと、 $\frac{(N-1)s^2}{\sigma^2} = \frac{4}{2} = 2$ は自由度4のカイ二乗分布の実現値だが、

この両側1%点は0.207と14.86であり、仮説は棄却できない。

8, コインを25枚投げたところ、表が10枚出た。表の割合が正規分布をすると仮定する。

(1) 仮説「表の出る確率が $1/2$ であり、正しく鑄造されている」を有意水準5%で検定を行え。両側検定か片側検定どちらを選んだかを述べ、その理由も記せ。

【解答】コインの表裏の出方は二項分布にしたがうので、コイン1枚について、分散 $V(X) = Np(1-p)$ である(教科書 p112 参照)。

推定量は $10 \div 25 = 0.4$ であるから、仮説 $p=0.5$ のもと、

$$\frac{0.4 - 0.5}{\sqrt{p(1-p)/N}} = \frac{-0.1}{\sqrt{0.25/25}} = \frac{-0.1}{0.1} = -1$$

仮説のもと統計量は標準正規分布にしたがうはずなので、この側5±1.96%を超えないので、仮説は棄却されない。正しく鑄造されないこ

とは、表が出やすい場合と裏が出やすい場合が考えられるので、両側検定を用いる。

(2) 信頼係数 95% の区間推定を  $p$  について求めよ。標準誤差は  $p$  の推定値を使って計算すること。

【解答】

$p$  の推定量を  $m (=0.4)$  とすると、

$$\begin{aligned} P\left(-1.96 < \frac{p-m}{\sqrt{m(1-m)/N}} < 1.96\right) &= P\left(-1.96 < \frac{p-0.4}{0.098} < 1.96\right) \\ &= P(0.4 - 1.96 \times 0.098 < p < 0.4 + 1.96 \times 0.098) \end{aligned}$$

よって信頼区間は、 $[0.208, 0.592]$

9, アカサギ (架空の鳥) の 25% はある遺伝子  $X$  を 2 つもち、50% が遺伝子  $X$  を 1 つもち、25% が遺伝子  $X$  をもたない。遺伝子  $X$  を 2 つもつアカサギの 80% は目が黒く、残り 20% は白い。遺伝子  $X$  を 1 つだけもつアカサギの 50% は目が黒く、50% は目が白い。遺伝子  $X$  をもたないアカサギの 20% は目が黒く、80% は目が白い。目の白いアカサギを捉えたとき、この個体が遺伝子  $X$  を 2 つもつ条件付き確率、 $X$  を 1 つもつ条件付き確率、 $X$  を 1 つもたない条件付き確率をそれぞれ求めよ。

【解答】 アカサギの目が白い確率は、

$$0.25 \times 0.2 + 0.5 \times 0.5 + 0.25 \times 0.8 = 0.5$$

したがって、目の白いアカサギが遺伝子  $X$  を 2 つもつ、1 つだけもつ、もたない確率はそれぞれ、

$$\frac{0.05}{0.5} = 0.1, \quad \frac{0.25}{0.5} = 0.5, \quad \frac{0.2}{0.5} = 0.4 \quad \text{となる。}$$

第 1 問にでてほしいね。

10, 100 人について、喫煙と呼吸器疾患の有無の調査を行い、次のようなデータを得た。喫煙と呼吸器疾患の間に関連は認められるか。有意水準は自分で設定せよ。

	喫煙する	喫煙せず
呼吸器疾患あり	20	30
呼吸器疾患なし	30	20

**【解答】**

上の表から喫煙と呼吸器疾患が無関係であるときの分布との差を示すカイ二乗統計量をつくると、

$$\frac{5^2}{25} + \frac{(-5)^2}{25} + \frac{5^2}{25} + \frac{(-5)^2}{25} = 4$$

となる。これは自由度1のカイ二乗分布の上側5%点3.84を超えるので、仮説は有意水準5%で棄却され、関係は認められる。

なぜ喫煙しない人の方が、呼吸器疾患にかかっている人が多いのか？  
問題の作成ミスか、意図的にやったのかどっちだろう・・・。

11, 互いに独立な確率変数 X、Y が区間 (0, 1) の一様分布にしたがうとき、 $Z=X+2Y$  の累積分布関数と密度関数を求めよ。

**【解答】**

累積分布関数

$$F(t) = P(X + 2Y < t) = \begin{cases} \frac{1}{4}t^2 & (t < 1) \\ \frac{1}{4} + \frac{1}{2}(t-1)^2 & (1 < t < 2) \\ 1 - \left(\frac{3}{2} - \frac{t}{2}\right)^2 & (2 < t < 3) \end{cases}$$



密度関数は、累積分布関数を微分して、

$$f(t) = \begin{cases} \frac{1}{2}t & (t < 1) \\ \frac{1}{2} & (1 < t < 2) \\ \frac{3}{2} - \frac{t}{2} & (2 < t < 3) \end{cases}$$