１年夏学期Ｆ系列「基礎統計」試験対策　（その１）
作成者 神吉【連絡先：masataka.kanki@ezweb.ne.jp】
試験予定日７／26【月】一限
電卓のみ持込可能
統計数値表は配布する。

この授業では毎回プリントを配っています。プリントがない場合は、２号館６階、605 Ｂ研究室の前でもらえるらしいですが、私もすべて持っているので、言ってくれれば渡せます。ノートのコピーがいる人も言ってください。ノートは 6/28 現在見開き 22 枚あります。これとプリントを読めばすべての問題が解けるのですが、多少量が多いので、重要な部分だけまとめて、試験対策プリントとしたいと思います。

1　範囲について

教科書は、「統計学入門」東大出版会 2800 円です。持っていない人は買う必要はないです。プリントで扱っていない難しいところは、範囲外です。さらに、プリント 3.3(Bayes の定理),7.1 も範囲外です。

プリントにミスがありました。p11 の（5）式の $P(A_0)$ は $P(A_1)$ の誤りです。また、p20 問２は、$P(X \geq x + y \mid X > x) = P(X \geq y)$ の誤りです。

2　１次元データ

$$\text{分散 } S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$\text{標準偏差 } S = \sqrt{\text{分散}}$$

は、知っているものとします。

次の変換は、標準化と呼ばれています。

$$z_i = \frac{x_i - \bar{x}}{S}$$

z_i の平均は 0、標準偏差は 1 となります。(←重要)
偏差値は、

$$50 + 10z_i$$

で定義します。
また、

$$変動係数 \frac{S}{\bar{x}}$$

が、データの散らばりを表すのに用いられることがあります。これはデータの大きさによらない指標です。(分散はデータの大きさに依存します。たとえば、おなじ $S = 2$（ｃｍ）といっても、データが「人間の身長」か「鉛筆の長さ」かでは散らばり具合はまったく違いますね。)
(ただし、$\bar{x}$ は x の平均を表します。)

3 回帰直線

2 次元データ
$(x_1, y_1), (x_2, y_2) \cdots, (x_n, y_n)$
が与えられたとき、x_i と y_i の間の直線的な関係 $y_i = a + bx_i$ をできるだけうまく構成したいとしましょう。
まず準備として、

$$共分散\ C_{xy} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

が重要です。これは、データ x, y の間の関係を表します。
$C_{xy} > 0$ のとき正の相関があり、
$C_{xy} < 0$ のとき負の相関がある。
といいます。
これを展開して、

$$C_{xy} = \frac{1}{n} \left(\sum_{i=1}^{n} x_i y_i \right) - \bar{x}\bar{y}$$

が分かります。

さらに、これをデータの大きさによらないように変換し、異なったサンプル間で比較できるようにしたのが相関係数です。

$$\text{相関係数}\ r_{xy} = \frac{C_{xy}}{S_x S_y}$$

S_x は、x の標準偏差です。y も同じです。

r_{xy} は、$-1 \leq r_{xy} \leq 1$ を満たし、絶対値が大きいほど x, y の関係は強く、$r_{xy} = \pm 1$ のときは、完全な直線関係です。

$r_{xy} = 0$ のときはまったく関係がなく、無相関であるといいます。

それでは、a, b を求めましょう。$d_k = y_i - (a + bx_i)$ として、$d_k^2 (k = 1, 2, \cdots n)$ の和を最小にするようにきめます。

つまり、

$$L = \sum_{i=1}^{n} \{y_i - (a + bx_i)\}^2$$

として、L を最小にすればいいのです。

$$L = \sum_{i=1}^{n} \{(y_i - \bar{y}) - b(x_i - \bar{x}) + (\bar{y} - a - b\bar{x})\}^2 \leftarrow \text{このようにすると計算が楽}$$

$$= \sum_{i=1}^{n} (y_i - \bar{y})^2 + b^2 \sum_{i=1}^{n} (x_i - \bar{x})^2 + n(\bar{y} - a - b\bar{x})^2 - 2b \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

$$(\leftarrow \sum_{i=1}^{n} (x_i - \bar{x}) = 0\ \text{に注意})$$

$$= nS_y^2 + b^2 nS_x^2 + n(\bar{y} - a - b\bar{x})^2 - 2nbC_{xy}$$

$$= nS_x^2 \{b - \frac{C_{xy}}{S_x^2}\}^2 + n(\bar{y} - a - b\bar{x})^2 + nS_y^2 - \frac{nC_{xy}^2}{S_x^2}$$

よって、

$$b = \frac{C_{xy}}{S_x^2},\ a = \bar{y} - b\bar{x} \leftarrow \text{重要}$$

のとき、L は最小になります。よって、これが求める a, b だと分かります。

補足として、決定係数に注意してください。詳しくは述べませんが、

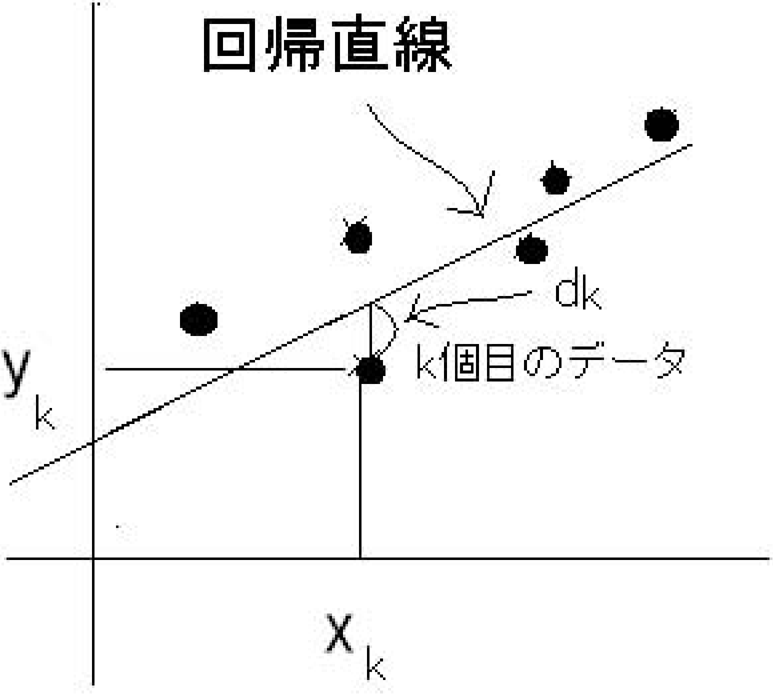
$$\text{決定係数} = r_{xy}^2$$

で定義され、直線の当てはまりのよさの指標です。直線回帰においては、相関係数よりもよく使われます。

決定係数 $= 1 \Longleftrightarrow$ データは同一直線上

が成り立ちます。■

次は、ポアソン分布、正規分布などについての予定。


回帰直線
dk
k個目のデータ
y_k
x_k

１年夏学期Ｆ系列「基礎統計」試験対策　（その２）
作成者 神吉【連絡先：masataka.kanki@ezweb.ne.jp】
試験予定日７／26 $\boxed{\text{火}}$ 一限
↑訂正 : 前のプリントでは 7/26(月）などと間抜けなことを書いていました。申し訳ありません。

電卓のみ持込可能

統計数値表は配布する。

1　過去問 2004 年の解答

気になっている人もいるだろうと思うので、まず、過去問に答えをつけます。間違っている部分を見つけたら、ご面倒でも上の連絡先までお願いします。

※グラフが別ファイルであります。参考にしてください。

問一

（１）

$X \sim N(50, 100)$ なので、母平均 $\mu = 50$, 母分散 $\sigma = \sqrt{100} = 10$ である。テキスト p280 の表を見て、

( )

$Z = \dfrac{X-\mu}{\sigma} = \dfrac{X-50}{10}$ と標準化して、$P(70 \leq X) = P(2 \leq Z) = Q(2) = 0.022750 = \mathbf{0.0228}$

( )

対称性より $P(40 \leq X \leq 60) = 2P(50 \leq X \leq 60) = 2\{0.5 - Q(1)\} = 2(0.5 - 0.15866) = 0.68268 = \mathbf{0.683}$

( )

$P(X \leq 55) = 1 - P(Z \leq 0.5) = 1 - Q(0.5) = 1 - 0.30854 = 0.69146 = \mathbf{0.691}$

（２）

16 人の体重を $X_1, X_2, \cdots, X_{16}$ として、$X_i \sim N(3.2, 0.4^2)$, $(1 \leq i \leq 16)$ だから、

$$\bar{X} = \frac{X_1 + X_2 + \cdots X_{16}}{16} \sim N\left(3.2, \frac{0.4^2}{\mathbf{16}}\right)$$

よって、$\bar{X}$ の標準偏差は、$\sqrt{0.4^2/16} = 0.1$

求める確率は、$P(\bar{X} \leq 3.35) = P\left(\dfrac{\bar{X} - 3.2}{0.1} \leq 1.5\right) = 1 - Q(1.5) = 1 - 0.066807 = \mathbf{0.933}$

（３）

$$E(X) = \sum_{k=1}^{N} k\frac{1}{N} = \mathbf{\frac{N+1}{2}}$$

$$V(X) = E(X^2) - \{E(X)\}^2 = \sum_{k=1}^{N} k^2\frac{1}{N} - \frac{1}{4}(N+1)^2 = \mathbf{\frac{N^2\text{-}1}{12}}$$

問二

（１）

$X \sim Bi(6, 0.30)$ なので、求める確率は

$$P(X = 2) = \frac{6!}{2!4!}0.3^2 0.7^4 = 0.324135 = \mathbf{0.324}$$

（２）

これは、近似的に**ポアソン分布**と考えられる。つまり、$X \sim Po(2)$ である。（確率が小さく、人数が多いから）（$\dfrac{1000!}{3!997!}0.002^3 0.998^{997}$ などすると日が暮れる。）

$$P(X = k) = \frac{2^k e^{-2}}{k!}$$

よって、

$$P(X = 3) = \frac{2^3 e^{-2}}{3!} = \frac{4}{3e^2} = \mathbf{0.180}$$

（３）

$X \sim Ge(0.4)$ なので、（幾何分布と呼ばれる。）（ただし、知らなくても答えは出る。）

$$P(X \leq 3) = P(X = 1) + P(X = 2) + P(X = 3) = 0.4 + 0.6 \cdot 0.4 + 0.6^2 \cdot 0.4 = \mathbf{0.784}$$

問三

（１）

（　）

チェビシェフの不等式は範囲外ですが、一応解いておきます。

コイン投げは、$Bi(10000, 0.5)$ に従うので $\mu = 5000, \sigma = \sqrt{np(1-p)} = 50$。ゆえに、

$$P(4850 \leq X \leq 5150) = P(|X - 5000| \leq 3 \times 50) > 1 - \frac{1}{3^2} = 0.8888$$

よって、問いの事実は確率 **0.889** 以上で起こるとわかる。

（ ）

i 回目に投げたコインが表のとき 1, 裏のとき 0, となる確率変数 X_i を導入する。

$$X_i \sim Bi(1, 0.5) \ (i = 1, 2, \cdots, 10000)$$

よって、

$$X_1 + \cdots + X_{10000} \sim Bi(10000, 0.5)$$

中心極限定理の主張より、

$$\text{上の式} \sim N(5000, 50^2) \text{ としてよい。}$$

よって、

$$P(4850 \leq X \leq 5150) = P(-3 \leq \frac{X - 5000}{50} \leq 3) = 1 - 2Q(3) = 1 - 2 \times 0.0013499 = 0.9973$$

ゆえに、確率がおよそ **0.997** だと分かる。

問四

※こちらもグラフがあります。

（１）

t 分布を使う。A の母平均を μ_1、母分散を σ_1^2 として、

$$\frac{\bar{X} - \mu_1}{\sqrt{\frac{s_1^2}{n}}} \sim t(n-1) \ (s_1^2 \text{は標本不偏分散})$$

であり、ここでは、$n = 10$ なので、

$$Z = \frac{14.4 - \mu_1}{\sqrt{4.9/10}}$$

として、

$$Z \sim t(9)$$

p281 の表から、$t_{0.025}(9) = 2.262$ なので、

$$P(-2.262 \leq Z \leq 2.262) = 0.95$$

$$P(12.8166 \leq \mu_1 \leq 15.9834) = 0.95$$

よって、**[12.8,16.0]** が求めたい信頼区間である。

（2）

χ^2 **分布**を使う。

$$\frac{(n-1)s_1^2}{\sigma_1^2} \sim \chi^2(n-1)$$

なので、

$$Z = \frac{9 \times 4.9}{\sigma_1^2}$$

として、

$$Z \sim \chi^2(9)$$

p282,p283 の表を用いて、

$$P(2.70039 \leq Z \leq 19.0228) = 0.95$$

$$P(2.3182 \leq \sigma_1^2 \leq 16.330) = 0.95$$

ゆえに、**[2.32,16.3]** が求める信頼区間。

（3）、（4）はまだ習っていません。(7/5 現在）最後二回の授業でやるのでは。

問五

(1)

回帰直線の公式から、

$$b = \frac{12.0}{38.7} = 0.3100,\ a = 172.4 - 167.2b = 120.5$$

よって、**y=121+0.310x**

（2）

どんなことって言われても、、、。傾きが正なので、父親と子供の身長には正の相関関係がある。しかし、b がそれほど大きくないから、関係性は薄い？

（３）

決定係数＝r_{xy}^2 および、$r_{xy} = \dfrac{C_{xy}}{S_x S_y}$ より、

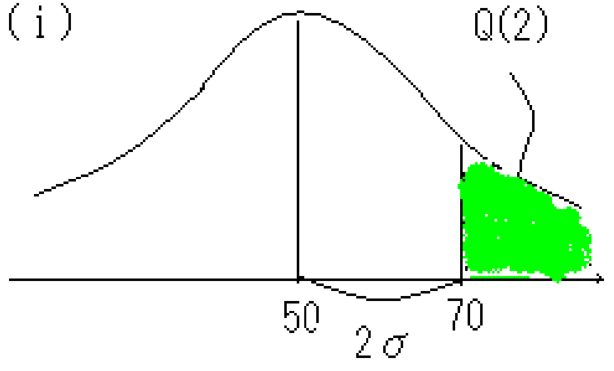
$$r_{xy}^2 = \frac{12.0^2}{38.7 \times 31.8} = \mathbf{0.117}$$

以上

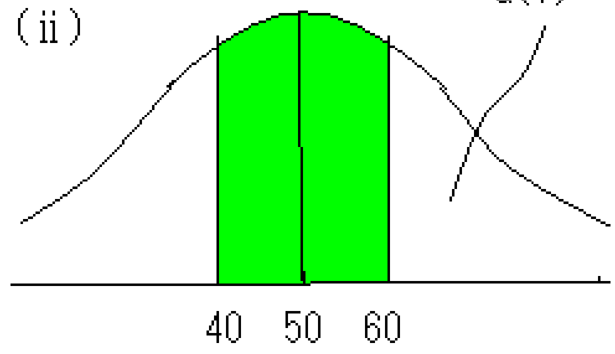
2004年度過去問の参考にしてください。

問一　1　求める部分に色を塗っています。

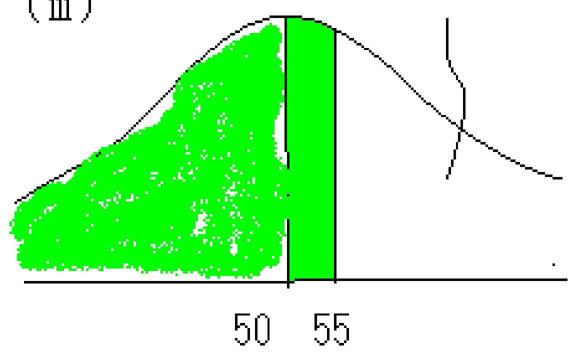
（ｉ）



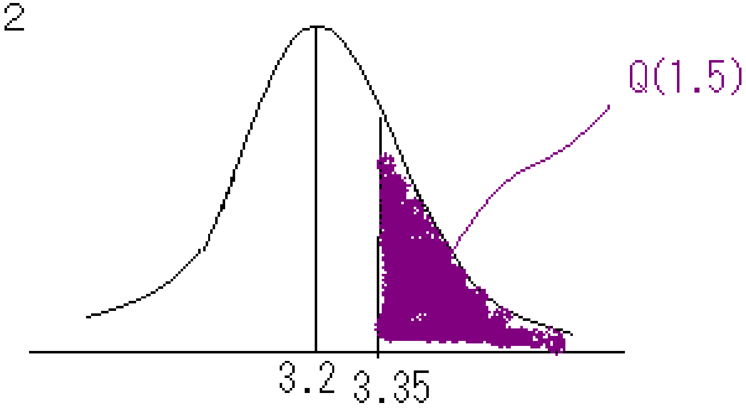
（ii）



（iii）



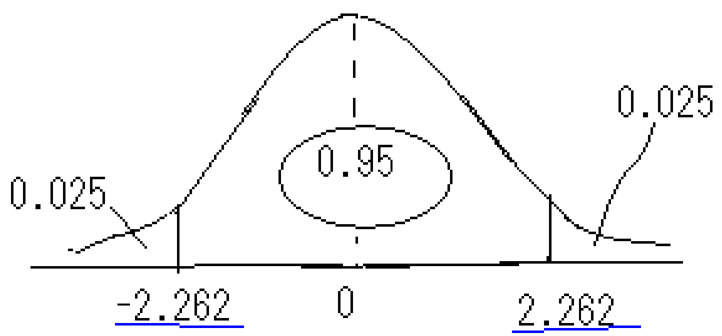
2



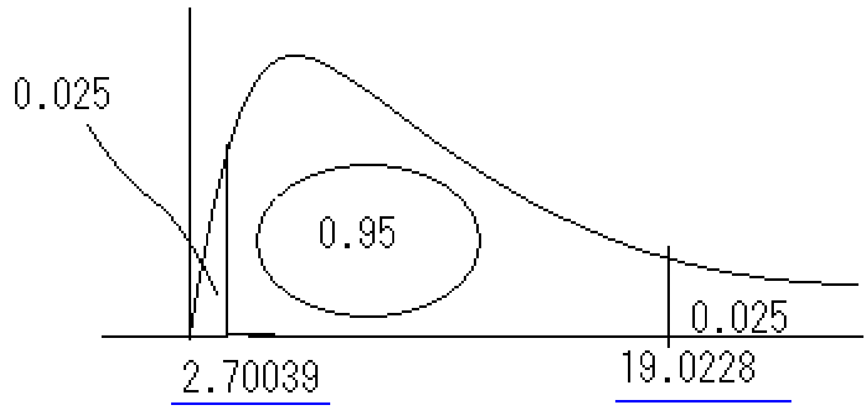
以上4つは正規分布　(左右対称)

問四

（1）t分布（左右対称）　t（9）



（2）χ２乗分布　χ^2（9）



１年夏学期Ｆ系列「基礎統計」試験対策　（その３）
作成者 神吉【連絡先：masataka.kanki@ezweb.ne.jp】

1　訂正

過去問解答で、チェビシェフの不等式は範囲外と書きましたが、確認したところ、チェビシェフの不等式はテスト範囲に入っていました。すみません。

2　確率、確率変数の補足

条件付確率は、

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

で定義されます。高校で、$P_B(A)$ と書いていたものと同じです。

3　確率変数

離散型と連続型の違いについてまとめてみました。

	離散型	連続型
確率変数	$x_1, x_2, \cdots$ のようなとびとびの値	$(-\infty, \infty)$ などの実数区間
確率の求め方	$P(X = x_k) = f(x_k)$	$P(a \leq X \leq b) = \int_a^b f(x)dx$
期待値の求め方	$\sum_k x_k f(x_k)$	$\int_{-\infty}^{\infty} xf(x)dx$
具体例	二項分布など	正規分布など

4　期待値と分散の補足

連続型でも、離散型でも次の式が成り立ちます。

$$E(aX+b)=aE(X)+b$$

$$V(X)=E(X^2)-\{E(X)\}^2$$

$$V(aX+b)=a^2V(X)$$

5　Chebyshev の不等式（チェビシェフ）

いかなる確率分布に対しても、

$$P(|X-\mu|\geq k\sigma)\leq\frac{1}{k^2}$$

が成り立つ。

どんな確率分布でも絶対に成り立つことが重要です。また、これは、

$$P(|X-\mu|\leq k\sigma)\geq 1-\frac{1}{k^2}$$

ともかけます。

6　二項分布 Bi(n,p)

X の確率分布が

$$f(x)=\frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}\ (x=0,1,\cdots,n)$$

で与えられるとき、

$$X \text{ は } Bi(n,p) \text{ に従う}$$

といい、

$$X\sim Bi(n,p)$$

と書きます。

$$E(X) = np,\ V(X) = np(1-p)$$

は重要です。

（例）表の出る確率 p のコインを n 回投げるとき n 回中表の出る回数を X として、$X \sim Bi(n,p)$

7　Poisson 分布 $Po(\lambda)$(ポアッソン)

$$f(x) = e^{-\lambda}\frac{\lambda^x}{x!}\ (x = 0, 1, 2, \cdots) \leftarrow \text{重要}$$

のとき、

$$X \sim Po(\lambda)$$

と定義します。また、

$$E(X) = \lambda,\ V(X) = \lambda$$

です。

（由来）2 項分布で、n がとても大きく、p が小さいときを考える。$np = \lambda$(一定）として、$n \to \infty (p \to 0)$ の極限を考えると、

$$\frac{n!}{x!(n-x)!}p^x(1-p)^{n-x} = \frac{n(n-1)\cdots(n-x+1)}{x!}\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{n-x} \approx \frac{\lambda^x}{x!}\left(1-\frac{\lambda}{n}\right)^n \approx \frac{\lambda^x e^{-\lambda}}{x!}$$

8　幾何分布 Ge(p)

$$f(x) = p(1-p)^{x-1}\ (x = 1, 2, \cdots)$$

のとき、

$$X \sim Ge(p)$$

と定義します。

$$E(X) = \frac{1}{p},\ V(X) = \frac{1-p}{p^2}$$

となります。

（例）さっきのコイン投げで、投げ始めてから初めて表の出る回数を X として、$X \sim Ge(p)$

9　正規分布 $N(\lambda, \sigma^2)$

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

のとき、

$$X \sim N(\mu, \sigma^2)$$

と定義します。(f(x) の式は覚えなくてよい。)

つまり、$N(\lambda, \sigma^2)$ は、平均 λ、分散 σ^2 の正規分布です。

f(x) のグラフは別ファイルを参照してください。

$X \sim N(\lambda, \sigma^2)$ のとき、

$$aX + b \sim N(a\mu + b, a^2\sigma^2)$$

が成立します。そこで、

$$Z \equiv \frac{X - \mu}{\sigma} \text{←標準化}$$

とすると、上の式で、$a = 1/\sigma,\ b = -\mu/\sigma$ としたものなので、

$$Z \sim N(0, 1)$$

を得ます。$N(0, 1)$ を標準正規分布といいます。

実際の確率計算では、標準化した後 p280 の表を見ます。

10　指数分布 $Ex(\lambda)$

$$f(x) = \lambda e^{-\lambda x}\ (x \geq 0)$$

$$= 0 \ (x < 0)$$

のとき、

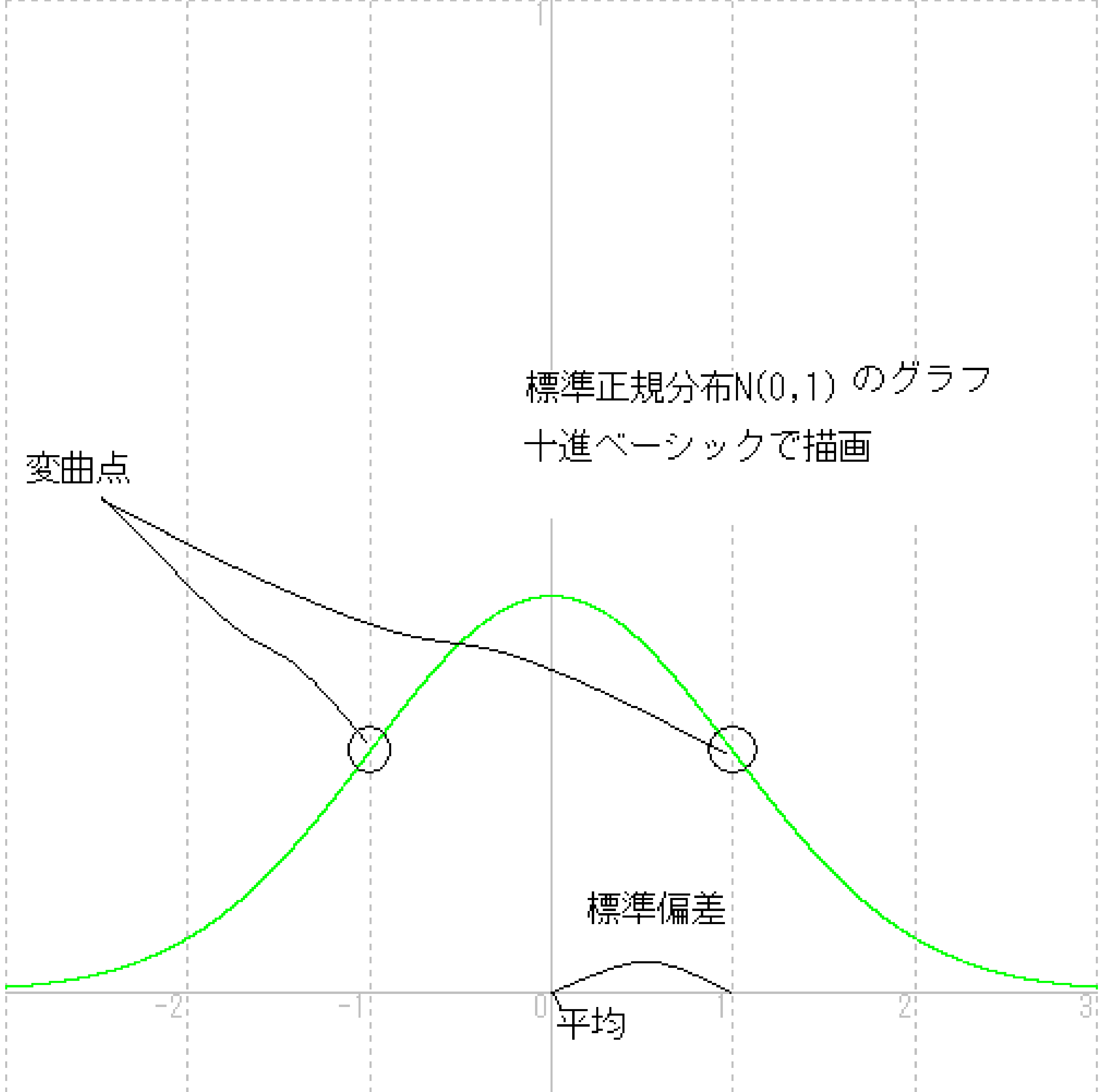
$$X \sim Ex(\lambda)$$

と定義します。

$$E(X) = \frac{1}{\lambda},\ V(X) = \frac{1}{\lambda^2}$$

となります。

つぎは、標本分布などについて扱います。



標準正規分布N(0,1) のグラフ
十進ベーシックで描画
変曲点
標準偏差
平均
1
-2
-1
0
1
2
3

１年夏学期Ｆ系列「基礎統計」試験対策　（その４）
作成者 神吉【連絡先：masataka.kanki@ezweb.ne.jp】

1　多次元の確率変数

1.1　同時分布、周辺分布

X, Y を確率変数とする。【離散型を考える。】
$X = x$ かつ $Y = y$ となる確率

$$P(X = x, Y = y) = f(x, y)$$

を二次元確率変数 (X,Y) の**同時分布**といいます。
また、$X = x$ となる確率は、

$$g(x) = P(X = x) = \sum_y f(x, y)$$

であり、これを、X の**周辺分布**といいます。
Y についても同様です。
$\phi(X, Y)$ を X と Y の函数として、$\phi(X, Y)$ の期待値は、

$$E(\phi(X, Y)) = \sum_x \sum_y \phi(X, Y) f(X, Y)$$

となります。

1.2　和の分布

$$E(X + Y) = E(X) + E(Y)$$
$$V(X + Y) = V(X) + V(Y) + 2Cov(X, Y)$$

が成立します。

※一般には、$V(X+Y)=V(X)+V(Y)$ とは限らない。X,Y が独立ならこれも成り立つ。

ここで、

$$Cov(X,Y)=E\{(X-\mu_X)(Y-\mu_Y)\}$$

で定義され、X と Y の**共分散**といいます。

上の式を展開して、

$$Cov(X,Y)=E(XY)-E(X)E(Y)$$

が得られます。これは計算に便利です。

X と Y の**相関係数**は、

$$\rho_{XY}=\frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}}$$

で定義されます。

1.3 独立性

X,Y が独立であるとは、すべての $i,j=1,2,\cdots$ に対し、

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

が成り立つことと同値です。

X,Y が独立ならば、

$$E(XY)=E(X)E(Y)$$

$$V(X+Y)=V(X)+V(Y)\ \text{つまり、}Cov(X,Y)=0$$

が成り立ちます。

1.4 n 個への拡張

$X_1,\cdots X_n$ なる互いに独立に同一の分布に従う確率変数があり、、

$$E(X_i)=\mu,\ V(X_i)=\sigma^2\ (1\leq i\leq n)$$

を満たすとする。【この条件は後でも使うので、※とおきます。】

さらに、

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

とおく。

このとき次が成り立つ。

$$E(\bar{X}) = \mu$$

$$V(\bar{X}) = \frac{\sigma^2}{n} \text{ ←きわめて重要！}$$

1.5　再生性

要するに、

ベルヌーイ分布の和もまたベルヌーイ分布（Bi(1,p）をベルヌーイ分布という)

二項分布、ポアッソン分布、正規分布でも同様

ということ。

（ただし、確率変数は独立でないといけない。）

2　大数法則　LLN（<u>L</u>aw of <u>L</u>arge <u>N</u>umbers)

法則の内容は、

※のとき、

$\epsilon > 0$ をどのように【小さく】とっても、

$$\lim_{n\to\infty} P(|\bar{X} - \mu| \geq \epsilon) = 0$$

となる。

というものです。証明は、チェビシェフの不等式を使うが、ここでは省略します。

【例】$X_i \sim Bi(1, 0.5)$ のとき、n を大きくしていくと、$\bar{X}$ がどんどん 0.5 に近づいていく。

3　中心極限定理　CLT（$\underline{\text{C}}$entral $\underline{\text{L}}$imit $\underline{\text{T}}$heorem）

定理の内容は、
※のとき、n が大ならば、$\bar{X}$ は近似的に

$$N(\mu, \frac{\sigma^2}{n})$$

に従う。
というものです。
【例】$X_1, \cdots X_n$ が独立に $Bi(1, p)$ に従うとき、

$$\bar{X} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

4　標本分布

ある母集団があり、

$$母平均\mu$$

$$母分散\sigma^2$$

が分かっていないとする。これを推測したい。
そこで、母集団から、n 個の標本 $X_1, \cdots X_n$ を抽出し、それらが従う確率分布を求めようとする。
そのとき使う統計量は、

$$標本平均\bar{X}$$

$$標本分散\ S^2$$

$$標本不偏分散\ s^2 := \frac{1}{\mathbf{n\text{-}1}} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

である。
s^2 をつかうのは、$E(s^2) = \sigma^2$ となっているから。$E(S^2) = \frac{n-1}{n}\sigma^2$ となり、S^2 では、正しく σ^2 を推定できない。

5　正規母集団からの標本

$X_1, \cdots X_n$ が互いに独立に、$N(\mu, \sigma^2)$ に従うとき、$\bar{X}$からμ を,s^2からσ^2 を推測する。

5.1　σ^2 が既知の場合

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

より、

$$\frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \sim N(0,1)$$

表から、

$$P(-1.96 \leq \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \leq 1.96) = 0.95$$

より、

$$\left[\bar{X} - 1.96\sqrt{\frac{\sigma^2}{n}}, \bar{X} + 1.96\sqrt{\frac{\sigma^2}{n}}\right]$$

は、確率 0.95 で未知の μ を含む。これを、μ に関する信頼係数 0.95 の**信頼区間**という。

5.2　σ^2 が未知のとき

5.2.1　χ^2 分布の 定義$_{\text{ていぎ}}$

$Z_1, \cdots Z_k$ が独立で、$N(0,1)$ に従うとき、

$$Y := \sum_{i=1}^{k} Z_i^2 \sim \chi^2(k)$$

$\chi^2(k)$ を自由度 k の χ^2 分布という。

5.2.2　σ^2 の推定

このとき使われるのが次の定理である。

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$$

これは重要である。

【例】n=40 のとき、

$$\frac{39s^2}{\sigma^2} \sim \chi^2(39)$$

より、

$$P(23.6543 \leq \frac{39s^2}{\sigma^2} \leq 58.1201) = 0.95$$

p282 の表を参照。縦軸が自由度 39、横軸は、0.975 と 0.025 のところを読む。

ゆえに、

$$\left[\frac{39}{58.1201}s^2, \frac{39}{23.6543}s^2\right]$$

は、確率 0.95 で未知の σ^2 を含む信頼区間である。

5.2.3　t 分布の定義

$$X \sim N(0,1)$$

$$Y \sim \chi^2(k)$$

$$X \text{ と } Y \text{ は独立}$$

のとき、

$$\frac{X}{\sqrt{\frac{Y}{k}}} \sim t(k)$$

と定義する。

$t(k)$ を自由度 k の t 分布と呼ぶ。

5.2.4　μ の推定

ここで用いるのは、

$$\frac{\bar{X}-\mu}{\sqrt{\frac{\mathbf{s}^2}{n}}} \sim t(n-1)$$

である。これも重要。

例として、プリント 9.3 の電球の問題がわかりやすいでしょう。

6　２標本問題

２つの標本

$$X_1, \cdots X_m \sim N(\mu_1, \sigma_1^2)$$

$$Y_1, \cdots Y_n \sim N(\mu_2, \sigma_2^2)$$

のとき、$\mu_1 - \mu_2$ を推定する。

$$\bar{X} \sim N(\mu_1, \frac{\sigma_1^2}{m})$$

$$\bar{Y} \sim N(\mu_2, \frac{\sigma_2^2}{n})$$

なので、

$$\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{m} + \frac{\sigma_1^2}{n}\right)$$

となる。

σ_1^2, σ_2^2 が既知なら信頼区間は、1 標本のときと同じように求められる。

σ_1^2, σ_2^2 が未知のときは、複雑なので、授業では $\sigma_1 = \sigma_2 := \sigma$ のときのみ取り扱った。

$$s_1^2 = \frac{1}{m-1}\sum_{i=1}^{m}(X_i - \bar{X})^2$$

$$s_2^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

の両方から、σ^2 を推定するため、

合併した分散

$$s^2 = \frac{1}{(m-1)+(n-1)}\{(m-1)s_1^2 + (n-1)s_2^2\}$$

を用いる。

このとき、$E(s^2) = \sigma^2$ であり、

$$\frac{(m+n-2)s^2}{\sigma^2} \sim \chi^2(m+n-2)$$

$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{s^2\left(\frac{1}{m}+\frac{1}{n}\right)}} \sim t(m+n-2)$$

が成り立つ。

7 推定

7.1 点推定

$\hat{\theta}$ が θ の不偏推定量であるとは、

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

が成り立つことで定義される。

【例】

$\bar{X}$ は μ の不偏推定量。

s^2 は σ^2 の不偏推定量。

7.2 モーメント法

モーメントを標本モーメントで推定する方法のこと。

定義

$E(X_i^r) := \mu_r$ を r 次モーメント

$\displaystyle\frac{1}{n}\sum_{i=1}^{n} X_i^r := \hat{\mu_r}$ を r 次標本モーメント
という。
また、
$E((X_i-\mu)^r) := \mu_r'$ を平均まわりの r 次母モーメント
$\displaystyle\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^r := \hat{\mu_r}'$
を平均まわりの r 次標本モーメント
という。

7.3 区間推定

今までやってきたように、信頼区間を作る推定法のことです。一般論よりも、具体的な問題にあたってみるのがよいでしょう。

7.4 中心極限定理を用いた区間推定

これは特に新しい知識は要りません。
注意は、たとえば、$Bi(1,p)$ で、p が分からないとき、$p(1-p)$ を $\bar{X}(1-\bar{X})$ と近似してよいという点です。そうしないと信頼区間が求まらないことがあります。

8 最後に

これで、7/5 までの授業の範囲が終わりました。残り 2 回の授業は各自でお願いします。
$\bowtie$