11　11回目講義の補足プリント

11.1　注意

試験には電卓（平方根が計算出来れば十分。安価なものでよい。関数電卓しか持っていない者はそれを使ってもよいが、関数計算機能など通常の電卓を超える機能を用いてはならない）を持参のこと。それ以外の持ち込みは不可。必ず出題するのは　正規分布の確率計算と　正規母集団、2 項母集団に関する推測問題。数表[1]は配布する。e は 2.7 で計算する。試験時間は 90 分。

11.2　標本平均の標本分布:　電球の寿命の例

[0] **予備知識.**

話に入る前に公式の復習：$Z \sim N(0,1)$ なら、

$$P(-1.96 \leq Z \leq 1.96) = 0.95. \tag{1}$$

[1] **モデル.**

大型電球を作る工場で新型の電球が開発されたとする。新型電球は旧型に比べて寿命が長いことが期待されるとする。新型電球の寿命を調べるため、新型電球 $n = 15$ 個を取り出して、その寿命 $X_1, \cdots, X_{15}$(単位は時間) を計測する。$X_1, \cdots, X_{15}$ は互いに独立に同一の正規分布 $N(\mu, 100^2)$ に従うと考えてよいとする（$X_1, \cdots, X_{15}$ を正規母集団 $N(\mu, 100^2)$ からの無作為標本とみなす)：

$$X_1, \cdots, X_{15} \sim N(\mu, 100^2). \tag{2}$$

[2]μ **の推定.**

何度も勉強した通り、

$$E(\bar{X}) = \mu \tag{3}$$

であるから、未知の μ を $\bar{X}$ の実現値 $\bar{x}$ によって推定することは妥当であろう。例えば、$\bar{x} = 1230$ であったなら、μ は 1230(時間) と推定される。

[3] μ **に関する信頼区間.**

前回勉強した信頼区間を作ってみよう。

$$\bar{X} \sim N\left(\mu, \frac{100^2}{15}\right) \tag{4}$$

であるから、

$$\frac{\bar{X} - \mu}{\sqrt{\frac{100^2}{15}}} \sim N(0,1) \tag{5}$$

である。従って、

$$\begin{aligned} & P\left(-1.96 \leq \frac{\bar{X} - \mu}{\sqrt{\frac{100^2}{15}}} \leq 1.96\right) = 0.95 \\ \iff \quad & P\left(\bar{X} - 1.96\frac{100}{\sqrt{15}} \leq \mu \leq \bar{X} + 1.96\frac{100}{\sqrt{15}}\right) = 0.95 \end{aligned}$$

[1] http://lecture.ecc.u-tokyo.ac.jp/ chkurata/ にも置いてあります。suuhyou.pdf というファイル。

となるから。

$$\left[\bar{X} - 1.96\frac{100}{\sqrt{15}},\ \bar{X} + 1.96\frac{100}{\sqrt{15}}\right] \tag{6}$$

は確率 0.95 で未知の μ を含む。この区間の実現値は

$$\left[1230 - 1.96 \times \frac{100}{\sqrt{15}},\ 1230 + 1.96 \times \frac{100}{\sqrt{15}}\right] = [1179.4,\ 1280.6] \tag{7}$$

である。

11.3 母分散が未知の場合の標本平均の標本分布： 電球の例の続き

まず次の事実を確認してから本題に入る。$t \sim t(14)$ ならば、t 分布表より

$$P(-2.145 \leq t \leq 2.145) = 0.95 \tag{8}$$

が読み取れる (数表の記号では $t_{0.025}(14) = 2.145$)。

(7) のような区間を構成しようと考えるならば、未知の σ^2 を s^2 で置き換えた

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

を考えるのが自然だろう。ここで t は正規分布しないことに注意する。t は自由度 $n-1$ の t 分布に従う。今の場合は、$t \sim t(14)$。

従って、(8) より、

$$P\left(-2.145 \leq \frac{\bar{X} - \mu}{\sqrt{\frac{s^2}{n}}} \leq 2.145\right) = 0.95 \tag{9}$$

$$\iff \quad P\left(\bar{X} - 2.145\frac{s}{\sqrt{n}} \leq \mu \leq \bar{X} + 2.145\frac{s}{\sqrt{n}}\right) = 0.95 \tag{10}$$

即ち、区間

$$\left[\bar{X} - 2.145\frac{s}{\sqrt{n}},\ \bar{X} + 2.145\frac{s}{\sqrt{n}}\right] \tag{11}$$

は確率 0.95 で未知の μ を含む。ここで　n が大となれば区間の幅が小となることと s が小となれば区間の幅は小となることに注意しよう。区間の幅が小ということは、推定の精度が良いということである。

例えば、$s^2 = 120^2$ が得られたとすると、この区間の実現値は

$$\left[1230 - 2.145 \times \frac{120}{\sqrt{15}},\ 1230 + 2.145 \times \frac{120}{\sqrt{15}}\right] = [1163.5,\ 1296.5] \tag{12}$$

となる。

12　信頼区間

12.1　正規母集団の場合

$X_1, \cdots, X_n$ は互いに独立に同一の正規分布 $N(\mu, \sigma^2)$ に従っているとする。

(1) 正規母集団 $N(\mu, \sigma^2)$ の母平均 μ の信頼区間 (母分散 σ^2 既知の場合):

$Z \sim N(0,1)$ のとき、Z の上側 $100\alpha\%$点を Z_α と書く。即ち、

$$P(Z > Z_\alpha) = \alpha. \tag{13}$$

既に学んだ通り、

$$\frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \sim N(0,1) \tag{14}$$

であるから、

$$P\left(-Z_{\alpha/2} \leq \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \leq Z_{\alpha/2}\right) = 1 - \alpha \tag{15}$$

$$\iff \quad P\left(\bar{X} - Z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}} \leq \mu \leq \bar{X} + Z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}}\right) = 1 - \alpha \tag{16}$$

が成立する。従って、

$$\left[\bar{X} - Z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}}, \quad \bar{X} + Z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}}\right] \tag{17}$$

は μ に関する信頼係数 $1-\alpha$ の信頼区間である。

(2) 正規母集団 $N(\mu, \sigma^2)$ の母平均 μ の信頼区間 (母分散 σ^2 未知の場合):

$t \sim t(n-1)$ のとき、t の上側 $100\alpha\%$点を $t_\alpha(n-1)$ と書く。即ち、

$$P(t > t_\alpha(n-1)) = \alpha. \tag{18}$$

既に学んだ通り、$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ とおけば

$$\frac{\bar{X} - \mu}{\sqrt{s^2/n}} \sim t(n-1) \tag{19}$$

であるから、

$$P\left(-t_{\alpha/2}(n-1) \leq \frac{\bar{X} - \mu}{\sqrt{s^2/n}} \leq t_{\alpha/2}\right) = 1 - \alpha \tag{20}$$

$$\iff \quad P\left(\bar{X} - t_{\alpha/2}(n-1)\sqrt{\frac{s^2}{n}} \leq \mu \leq \bar{X} + t_{\alpha/2}(n-1)\sqrt{\frac{s^2}{n}}\right) = 1 - \alpha \tag{21}$$

が成立する。従って、

$$\left[\bar{X} - t_{\alpha/2}(n-1)\sqrt{\frac{s^2}{n}}, \quad \bar{X} + t_{\alpha/2}(n-1)\sqrt{\frac{s^2}{n}}\right] \tag{22}$$

は μ に関する信頼係数 $1-\alpha$ の信頼区間である。

(3) 正規母集団 $N(\mu, \sigma^2)$ の母分散 σ^2 の信頼区間:

$Y \sim \chi^2(n-1)$ のとき、Y の上側 100α%点を $\chi^2_\alpha(n-1)$ と書く。即ち、

$$P(Y > \chi^2_\alpha(n-1)) = \alpha. \tag{23}$$

既に学んだ通り、$s^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$ とおけば

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1) \tag{24}$$

であるから、

$$P\left(\chi^2_{1-\alpha/2}(n-1) \le \frac{(n-1)s^2}{\sigma^2} \le \chi^2_{\alpha/2}(n-1)\right) = 1-\alpha \tag{25}$$

$$\iff \quad P\left(\frac{(n-1)s^2}{\chi^2_{\alpha/2}(n-1)} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}(n-1)}\right) = 1-\alpha \tag{26}$$

が成立する。従って、

$$\left[\frac{(n-1)s^2}{\chi^2_{\alpha/2}(n-1)},\ \ \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}(n-1)}\right] \tag{27}$$

は σ^2 に関する信頼係数 $1-\alpha$ の信頼区間である。

12.2 中心極限定理に基づく信頼区間

例 12.1. **(支持率の推定)** 国民の何パーセントが国債の新規発行を支持しているかを調べるため、400 人を無作為に選び、支持不支持の如何について尋ねたところ、$\bar{X} = 0.6$ であった。国民全体における支持率 p の 99%信頼区間は

$$\left[0.6 \pm 2.58 \times \sqrt{\frac{0.6(1-0.6)}{400}}\right] = [0.537, 0.663] = [53.7, 66.3](\%)$$

である。□

例 12.2. **(感染率の推定)** ある地域でのある病原虫の感染率を調べるため、その地域から 120 人を無作為に抽出して感染の有無を調べたところ 11 人の感染者がいた。感染率は幾らと推定されるか。

例 12.3. **(自動販売機)** ある自動販売機で売られているジュースの 1 日当たりの販売数を冬季 45 日間調べたところ、平均 $\bar{X} = 11$(本/日)であった。45 日間の販売数はポアソン母集団 $Po(\lambda)$ からの無作為標本と見なせるとする。母平均 λ の 95%信頼区間を求めると、

$$\left[11 \pm 1.96 \times \sqrt{11/45}\right] = [10.0,\ 12.0]$$

となる。□