確率・統計・モデリング問題集まとめ・補足

6.1 確率の基礎

測度論の辺りの話は実際には試験に必要なし．興味のある人は [1] 伊藤清三，ルベーグ積分入門 [2] 猪狩惺，実解析入門 等を参照すると良いと思います．また可算や非加算の話は集合・位相の本を開けば載っているので各人で参照してください．

• **テイル確率** (tail probability)　(テキスト P59)

テイル確率 (tail probability) とは $P(X \geq k)$ となる確率．しっぽ確率ともよばれる．以下に性質を．

$$(i) P(X \geq 0) = 1 \quad (\text{ただし，確率変数 } X \text{ は非負の値をとるものとする})$$

$$(ii) E(X) = \begin{cases} \sum_{k=1}^{\infty} P(X \geq k) = \sum_{k=0}^{\infty} P(X > k) & (\text{離散の場合}) \\ \\ \int_0^{\infty} P(X \geq t)dt & (\text{連続の場合}) \end{cases}$$

$$(iii)(\text{離散の場合}) \quad \sum_{k=1}^{\infty} kP(X \geq k) = E(\frac{X(X+1)}{2})$$

$$(\text{連続の場合}) \quad \int_0^{\infty} tP(X \geq t)dt = E(\frac{X^2}{2})$$

大事なのは (ii) と (iii) の性質で，これらの性質から期待値や分散が計算できます．和や積分の範囲に注意が必要です．(ii) のテキストの証明を補足以下にしておきます．(ゼミでやりました.) まず離散の場合の性質 (ii) から．一段目の証明はわかると思います．二段目の指示関数を用いた証明の復習.

指示関数の定義.

$$1_{X \geq k} = \begin{cases} 1 & if\ X \geq k \\ 0 & otherwise \end{cases}$$

つぎに

$$\sum_{k=1}^{\infty} P(x \geq k) = E(\sum_{k=1}^{\infty} 1_{X \geq k})$$

これは,

$$\begin{aligned} E(\sum_{k=1}^{\infty} 1_{X \geq k}) &= \sum_{l} (\sum_{k=1}^{\infty} 1_{X \geq k}) P(X = l) \quad (P59 \text{ 期待値の定義式で } h(k) = \sum_{k=1}^{\infty} 1_{X \geq k} \text{としたもの}) \\ &= \sum_{k=1}^{\infty} (\sum_{l} 1_{X \geq k} P(X = l)) \quad (\textstyle\sum \text{ について交換}) \\ &= \sum_{k=1}^{\infty} \sum_{l=k}^{\infty} P(X = l) \quad (\text{指示関数の意味は } X \geq k \text{ なら } 1，\text{それ以外は } 0) \\ &= \sum_{k=1}^{\infty} P(X \geq k) \end{aligned}$$

その次.

$$E(\sum_{k=1}^{\infty} 1_{X \geq k}) = E(\sum_{k=1}^{X} 1)$$

これは指示関数の範囲の意味を"逆に"考え，k が X 以下なら 1，それ以外では 0 (cf 今までは X が k 以上なら 1，それ以外では 0 と考えていた) と考えれば理解できます.

今度は連続の場合の性質 (ii) の補足.

$$\begin{aligned}
\int_0^\infty P(X \ge t)dt &= \int_{t=0}^{t=\infty} \left\{ \int_{x=t}^{x=\infty} f(x)dx \right\} dt \\
&= \int_{t=0}^{t=\infty} \left\{ \int_{x=0}^{x=\infty} 1_{X \ge t} f(x)dx \right\} dt \quad \text{(指示関数を用いて積分範囲を変更)} \\
&= \int_{x=0}^{x=\infty} \left\{ \int_{t=0}^{t=\infty} 1_{X \ge t} dt \right\} f(x)dx \quad \text{(積分順序の変更．先に } t \text{ で積分する.)} \\
&= E\left(\int_{t=0}^{t=\infty} 1_{X \ge t} dt \right) \text{(}P59\text{ 期待値の定義式で } h(x) = \int_{t=0}^{t=\infty} 1_{X \ge t} dt \text{ としたもの)} \\
&= E\left(\int_0^X 1dt \right) \quad \text{(また指示関数を”逆”の見方でみる)} \\
&= E(X)
\end{aligned}$$

性質 (iii) はどちらの場合も指示関数の見方を変更するという考え方を用いれば理解できます.

• **母関数** (generating function)

定義や性質は教科書にある通りなので割愛．母関数を用いればXの n 次のモーメントを計算できます．アクチュアリーの試験では母関数を計算させる問題も出題されているので要練習です.

• **確率変数の変換**

ここは非常に重要でアクチュアリーの試験でも良く出題されるので，練習を積みましょう.

6.2 確率分布を表す母数 (パラメーター)

歪度，尖度の定義・性質等を確認しておいてください.

6.3 標本統計量

• **最尤推定量** (maximum likelihood estimator)

教科書の説明では抽象的なので具体例を2つほど．1つ目の例として (竹村彰通，現代数理統計学 P138 参照) 二項分布の成功確率 p の最尤推定量をもとめる．(経験的には p=(成功回数)/(試行回数) だと思われるが，これを解析的に求める．)

今考えているのは,「ある試行の成功回数が二項分布に従うとして，その試行を行ったら n 回中 x 回成功した」という状況．まず尤度関数 (likelihood function)$L(p)$(教科書で言うパラメータ θはこの例では成功確率 p に相当する) は

$$L(p) = {}_nC_x\ p^x(1-p)^{n-x}$$

なので対数尤度関数 (log likelihood function)$l(p)(= \log L(p))$ は

$$\begin{aligned}
l(p) &= x \log p + (n-x)\log(1-p) + \log {}_nC_x \\
\frac{\partial l(p)}{\partial p} &= \frac{x}{p} - \frac{n-x}{1-p} = \frac{x-np}{p(1-p)}
\end{aligned}$$

よって

$$\frac{\partial l(p)}{\partial p} = 0 \Longleftrightarrow p = \frac{x}{n}$$

従って，p の最尤推定量 $\hat{p} = x/n$ と分かる．

次に指数分布のパラメータ λ の最尤推定量を解析的に求めることを考える．パラメータ λ の同一な指数分布に従う確率変数 $X_k(k = 1, 2, \cdots, n)$ が x_k という値をとったとする．このとき，尤度関数 $L(\lambda)$ は

$$L(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i}$$

従って対数尤度関数 $l(\lambda)$ は

$$l(\lambda) = \sum_{i=1}^{n} \log \lambda e^{-\lambda x_i} = n \log \lambda - \lambda \sum_{i=1}^{n} x_i$$

よって，

$$\frac{\partial l(\lambda)}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i$$

従って，λの最尤推定量$\hat{\lambda}$ は

$$\begin{aligned} \frac{\partial l(\lambda)}{\partial \lambda} &= 0 \\ \Longleftrightarrow \frac{1}{\lambda} &= \frac{1}{n} \sum_{i=1}^{n} x_i \\ \therefore \quad \frac{1}{\hat{\lambda}} &= \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x} \end{aligned}$$

これは，パラメータ λ の指数分布に従う確率変数の期待値 (平均) が $\dfrac{1}{\lambda}$ に従うことを考えれば妥当な結果です．

• **クラメル・ラオの不等式** (Cramer-Rao inequality)

この不等式は情報量不等式 (information inequality) とも呼ばれます．クラメル・ラオの不等式を考えるにあたって，まずフィッシャーの情報量 (Fisher information)$I_n(\theta)$ を定義します．(ただし θ は密度関数のパラメータ．例えば指数分布であれば λ に相当．)

$$\begin{aligned} I_n(\theta) &= E\left(\left\{ \frac{\partial \log f(X, \theta)}{\partial \theta} \right\}^2 \right) \\ &= \int_{\Omega} \left(\frac{\frac{\partial}{\partial \theta} f(x, \theta)}{f(x, \theta)} \right)^2 f(x, \theta) dx \\ &= \int_{\Omega} \frac{\left(\frac{\partial}{\partial \theta} f(x, \theta) \right)^2}{f(x, \theta)} dx \end{aligned}$$

上の定義は連続の場合ですが，離散の場合は $\int$ を $\sum$ に置き換えてください．また $I_n(\theta)$ の添え字 n は密度関数の標本サイズを表わします．(上の最尤推定量の例で言うと，1 つ目の例では 1，二つ目の例では n です．つまり，確率変数の数です．)

この $I_n(\theta)$ を用いてクラメル・ラオの不等式は

$$Var(\hat{\theta}) \geq 1/I_n(\theta)$$

と表わせる．詳しい証明は (竹村彰通，現代数理統計学第 7 章) を参照してください．ここで具体例を 1 つ．正規分布の平均 μ の推定において $\frac{1}{n}\sum_{i=1}^{n} x_i$ が不偏推定量であることを示します．まず対数尤度関数 $l(\mu, x)$ を求める．

$$\begin{aligned} l(\mu, x) &= -\frac{(x-\mu)^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2) \\ \frac{\partial l(\mu, x)}{\partial \mu} &= \frac{x - \mu}{\sigma^2} \end{aligned}$$

よって，フィッシャーの情報量は

$$
\begin{aligned}
I(\mu) &= E\left(\left\{\frac{\partial l(\mu,x)}{\partial \mu}\right\}^2\right) \\
&= \frac{E(X-\mu)^2}{\sigma^4} \\
&= \frac{1}{\sigma^2}
\end{aligned}
$$

従って

$$\frac{1}{I_n(\mu)} = \frac{1}{nI(\mu)} = \frac{\sigma^2}{n} = Var(\bar{X})$$

よって $\bar{X}$ はμの推定量の中で分散を最小にすることが，クラメル・ラオの不等式からわかる.