Hand out

due	15 mai 2014
cur	22 mai 2014
ver	1
rev	0

確率論 (*Probability*) 第 6 週

w.yamamoto

1　期待値

確率分布 F に従う確率変数 X が、確率分布の標本空間 Ω のどの点を中心に出現するかを表す量を「確率変数 X の期待値」という。期待される X の値、という意味の英語 “expected value” の訳である。この X の期待値を得る操作 (計算) を $E[\cdot]$ と書く。期待値を平均と言うこともあるが、「何かの期待値」もしくは「何かの平均」と「確率分布の期待値」もしくは「確率分布の平均」とは区別しておく。

F が離散分布および Ω が離散集合の場合に、確率関数 $p(x) = Pr[X = x]$ を用いて

$$E[X] = \sum_{x=-\infty}^{\infty} xp(x) \tag{1}$$

と定義される。F が連続分布および Ω が連続集合の場合には、累積分布関数 $F(x) = Pr[X \le x]$ の導関数である確率密度関数 $f(x) = dF(x)/dx$ を用いて

$$E[X] = \int_{-\infty}^{\infty} xf(x)\,dx \tag{2}$$

と定義される。離散集合の場合には確率 $p(x)$ や重み付きの値 $xp(x)$ を総和 Σ で集めて平均を計算し、連続集合の場合には密度 $f(x)$ や重み付きの値 $xf(x)$ を積分 $\int$ で集めて平均を計算する。

いずれの場合にも「$X = x$ の確率 (あるいは密度) を重みにした Ω の値の平均」という意味の量となるのは、重み付き平均を

$$\frac{\displaystyle\sum_{x=-\infty}^{\infty} xp(x)}{\displaystyle\sum_{x=-\infty}^{\infty} p(x)} = \sum_{x=-\infty}^{\infty} xp(x)$$

$$\frac{\displaystyle\int_{-\infty}^{\infty} xf(x)\,dx}{\displaystyle\int_{-\infty}^{\infty} xf(x)\,dx} = \int_{-\infty}^{\infty} xf(x)\,dx$$

と計算してみれば明らか。

ところで $E[\cdot]$ の中に入るのは、確率変数そのものだけでなく、確率変数の関数も入ってよい。ある関数 $g(x)$ に確率変数 X を代入して得た $g(X)$ も、確率変数であり、期待値を次のように計算できる。

$$E[g(X)] = \begin{cases} \displaystyle\sum_{x=-\infty}^{\infty} g(x)\,p(x) \\ \displaystyle\int_{-\infty}^{\infty} g(x)\,f(x)\,dx \end{cases}$$

これを $g(X)$ の期待値と呼ぶ。

また期待値計算には任意の定数 a と b による $aX + b$ の変換に関して

$$E[aX + b] = aE[X] + b \tag{3}$$

という性質もある。これは確率分布が連続の場合にも、離散の場合にも簡単に証明できる。

2　確率分布の特徴量

確率分布には幾つかの特徴量がある。

h location-shift の図を入れる。(対称な分布と非負の分布両方)

h scale-change の図を入れる。(対称な分布と非負の分布両方)

2.1　確率分布の平均

$g(X) = X$ としたときの $E[g(X)]$ を確率分布 F の平均という。これは確率あるいは密度を重みとしたときの確率分布の重心であり、μ、μ_1 あるいは m_1 と書く。

$$\mu = \begin{cases} \displaystyle\sum_{x=-\infty}^{\infty} xp(x) \\ \displaystyle\int_{-\infty}^{\infty} xf(x)\,dx \end{cases}$$

期待値は確率分布の位置を表すひとつの特徴量である。期待値が大きい確率分布は、期待値が小さい確率分布と比較して、確率関数や確率密度が数直線上の右側に寄る。

2.2　確率分布の分散と標準偏差

2.2.1　分散

$g(X) = (X - \mu)$ としたときの $E[g(X)]$ を確率分布 F の分散という。これは X の μ からの 2 乗距離の平均であり、σ^2、m_2 などと書く。

$$\sigma^2 = \begin{cases} \displaystyle\sum_{x=-\infty}^{\infty} (x-\mu)^2\, p(x) \\ \displaystyle\int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx \end{cases}$$

分散は確率分布の広がり具合 (ばらつきの大きさ) を表すひとつの特徴量である。分散が大きい確率分布は、分散が小さい確率分布と比較して、確率関数や確率密度の背が低く、また一定以上の確率や密度の範囲が標本空間の広範囲に渡る。

2.2.2　２乗距離の期待値を最小にする点としての平均

確率分布の平均を中心と呼ぶのは、確率変数 X とある点 a の間の 2 乗距離の期待値 $E\left[(X-a)^2\right]$ を、$a-\mu$ が最小にするからである。実際

$$\frac{\partial}{\partial a} E\left[(X-a)^2\right] = E\left[\frac{\partial}{\partial a}(X-a)^2\right] = E[-2(X-a)] = -2E[X] + 2a \tag{4}$$

また

$$\frac{\partial^2}{\partial a^2} E\left[(X-a)^2\right] = E\left[\frac{\partial^2}{\partial a^2}(X-a)^2\right] = E[2] = 2 \tag{5}$$

より、$E\left[(X-a)^2\right]$ は $a = E[X]$ で最小値 $m_2 = \sigma^2$ をとる。ここの箇所の計算では、

$$\frac{\partial}{\partial a}\int_{-\infty}^{\infty}(x-a)^2 f(x)\,dx = \int_{-\infty}^{\infty}\frac{\partial}{\partial a}(x-a)^2 f(x)\,dx$$

などの微分と積分の交換可能性、および

$$\frac{\partial}{\partial a}\sum_{x=-\infty}^{\infty}(x-a)^2\, p(x) = \sum_{x=-\infty}^{\infty}\frac{\partial}{\partial a}(x-a)^2\, p(x)$$

などの微分と総和の交換可能性は既知とした。

2.2.3 標準偏差

さて、確率分布の分散の単位は X の単位の二乗になってしまう。そのため散らばりの大きさの指標としては、ちょっと大きな散らばりでももの凄く大きくなってしまうことがある。そのため分散よりは、ばらつきの大きさの表現として実感が沸き易い、分散の平方根を用いることがある。これは地図上の 2 点間を、実際の道路の弧や高低を辿って移動する際の総移動距離を求める際に、三平方の定理 (ピタゴラスの定理) を何度も組み合わせて計算したからといって距離の二乗を得て満足することはなく、実際に移動する距離を解としなければ、間隔の感触が掴めないのと同様である。

例えば「世帯総所得学の平均は 546.9 万円、分散は 153252 万円^2」と報告するより、「世帯総所得学の平均は 546.9 万円、標準偏差は 391 万円」と報告する方が、単位を受け入れ易い。(いずれも厚生労働省平成 22 年国民生活基礎調査より推計)

確率分布の分散の平方根のことを標準偏差という。

$$\sigma^2 = \begin{cases} \sqrt{\displaystyle\sum_{x=-\infty}^{\infty} (x-\mu)^2\, p(x)} \\ \sqrt{\displaystyle\int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx} \end{cases}$$

2.2.4 確率変数の標準化

確率変数 X に

$$Z = \frac{X-\mu}{\sigma} \tag{6}$$

という変換を施す。この変換は

$$E[Z] = E\left[\frac{X-\mu}{\sigma}\right] = \frac{E[X-\mu]}{\sigma} = \frac{E[X]-\mu}{\sigma} = 0 \tag{7}$$

および

$$E\left[Z^2\right] = E\left[\left(\frac{X-\mu}{\sigma}\right)^2\right] = \frac{E\left[(X-\mu)^2\right]}{\sigma^2} = \frac{\sigma^2}{\sigma^2} = 1 \tag{8}$$

より、平均が 0、分散が 1 となる。この変換を、確率変数の標準化という。

2.2.5 変動係数

標準偏差を平均で割った値

$$\frac{\sigma}{\mu} = \frac{\sqrt{\mu_2}}{m_1} \tag{9}$$

を変動係数と言う。

2.3 確率分布の歪度

3 次の中心モーメント μ_3 を標準偏差 $\sqrt{\mu_2}$ の 3 乗で標準化した量

$$\frac{\mu_3}{{\mu_2}^{3/2}} = \frac{\mu_3}{\sigma^3} \tag{10}$$

を確率分布の歪度と呼ぶ。確率分布の平均 μ の周りに対称な確率関数もしくは確率密度関数を持つ確率分布であれば、歪度は必ず 0 となる。

2.4 確率分布の尖度

4 次の中心モーメント μ_4 を標準偏差 $\sqrt{\mu_2}$ の 4 乗で標準化した量

$$\frac{\mu_4}{{\mu_2}^2} = \frac{\mu_4}{\sigma^4} \tag{11}$$

を確率分布の尖度と呼ぶ。この式を正規分布について計算すると、平均や分散によらず 3 となる。そこで、上の式から 3 を減じた式

$$\frac{\mu_4}{{\mu_2}^2} - 3 \tag{12}$$

を尖度と定義することもしばしばである。ソフトウェアが尖度を算出していたときには、この 2 つの定義のどちらを用いているか、確認する必要がある。