『計量経済学の第一歩』

田中 隆一（著）

ウェブサポートページ補論

発行所　株式会社有斐閣
2015 年 12 月 20 日　初版第 1 刷発行

ISBN 978-4-641-15028-7

補論１　繰り返し期待値の法則

第 3 章で期待値の性質について説明しましたが，そのひとつとして**繰り返し期待値の法則**とよばれる，期待値の便利な性質を紹介しました。ここではこの繰り返し期待値の法則についてもう少し詳しく説明します。

「繰り返し期待値の法則」のうちの 1 つの性質は，「条件付き期待値の期待値は条件なし期待値になる」というものです。具体的にはXで条件付けしたUの期待値$\mathrm{E}[U|X]$の（Xに関する）期待値をとると，

$$\mathrm{E}[\mathrm{E}[U|X]] = \mathrm{E}[U]$$

となります。

たとえば，2 つの離散的な確率変数(U, X)を考えます。確率変数Xは(x_1, x_2)のいずれかの値をとり，Uは(u_1, u_2)のいずれかの値をとるとします。$X = x_1, X = x_2$それぞれで条件付けたUの期待値$E[U|X]$は，条件付き確率関数$p_U(u|x)$を使って次のように書けます。

$$\begin{aligned}
\mathrm{E}[U|X = x_1] &= u_1 \times p_U(u_1|x_1) + u_2 \times p_U(u_2|x_1) \\
\mathrm{E}[U|X = x_2] &= u_1 \times p_U(u_1|x_2) + u_2 \times p_U(u_2|x_2)
\end{aligned}$$

さらにXを確率変数としてXの確率関数$p_X(x)$を使って期待値を計算すると，次のようになります。

$$\begin{aligned}
&\mathrm{E}\big[\mathrm{E}[U|X]\big] \\
&\quad = \mathrm{E}[U|X = x_1] \times p_X(x_1) + \mathrm{E}[U|X = x_2] \times p_X(x_2) \\
&\quad = [u_1 \times p_U(u_1|x_1) + u_2 \times p_U(u_2|x_1)] \times p_X(x_1) \\
&\qquad + [u_1 \times p_U(u_1|x_2) + u_2 \times p_U(u_2|x_2)] \times p_X(x_2) \\
&\quad = \big[u_1 \times p_{U,X}(u_1, x_1) + u_2 \times p_{U,X}(u_2, x_1)\big] \\
&\qquad + \big[u_1 \times p_{U,X}(u_1, x_2) + u_2 \times p_{U,X}(u_2, x_2)\big] \\
&\quad = u_1 \times \big[p_{U,X}(u_1, x_1) + p_{U,X}(u_1, x_2)\big] + u_2 \times \big[p_{U,X}(u_2, x_1) + p_{U,X}(u_2, x_2)\big] \\
&\quad = u_1 \times p_U(u_1) + u_2 \times p_U(u_2) \\
&\quad = \mathrm{E}[U]
\end{aligned}$$

つまり，繰り返し期待値の法則が成り立っていることがわかります。

補論 2　最小 2 乗推定量の分散の推定方法

傾きパラメターβ_1の最小 2 乗推定量$\hat{\beta}_1$の分散が

$$\mathrm{V}[\hat{\beta}_1 | x_1, \dots, x_n] = \frac{s^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

となることを確かめましょう。β_1の最小 2 乗推定量（本文の (5.5) 式）に単回帰モデル

$$Y_i = \beta_0 + \beta_1 X_i + U_i$$

を代入して整理すると，

$$\hat{\beta}_1 = \beta_1 + \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})U_i}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

になります。不偏性の証明と同様に，$(x_1, \dots, x_n)$で条件付けした分散をはじめに見ると，

$$\mathrm{V}[\hat{\beta}_1 |\, x_1, \dots, x_n] = \mathrm{E}[\left(\hat{\beta}_1 - \beta_1\right)^2 |\, x_1, \dots, x_n]$$

になります（すでに$\mathrm{E}[\hat{\beta}_1] = \beta_1$の関係を使いました）。この式に$\hat{\beta}_1$を代入すると，

$$\mathrm{E}[\left(\hat{\beta}_1 - \beta_1\right)^2 |\, x_1, \dots, x_n] = \mathrm{E}\left[\left[\frac{\sum_{i=1}^{n}(X_i - \bar{X})U_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right]^2 |\, x_1, \dots, x_n\right]$$

となります。期待値記号の中にある分数の分子を展開すると，

$$\begin{aligned}
&\left[\sum_{i=1}^{n}(X_i - \bar{X})U_i\right]^2 \\
&= [(X_1 - \bar{X})^2 \times U_1^2 + (X_1 - \bar{X})(X_2 - \bar{X}) \times U_1 \times U_2 + \cdots \\
&\quad +(X_1 - \bar{X})(X_n - \bar{X}) \times U_1 \times U_n] \\
&\quad +[(X_2 - \bar{X})(X_1 - \bar{X}) \times U_2 \times U_1 + (X_2 - \bar{X})^2 \times U_2^2 + \cdots \\
&\quad +(X_2 - \bar{X})(X_n - \bar{X}) \times U_1 \times U_n] + \cdots \\
&\quad +[(X_n - \bar{X})(X_1 - \bar{X}) \times U_n \times U_1 + (X_n - \bar{X})^2 \times U_n^2]
\end{aligned}$$

となります。ここで，分散均一の仮定のもとでは

$$\mathrm{E}[(X_i-\bar{X})^2\times U_i^2|x_1,\ldots,x_n]=(x_i-\bar{x})^2\mathrm{E}[U_i^2|\,x_1,\ldots,x_n]=\ (x_i-\bar{x})^2\times s^2$$

となることと，仮定 2 のもとでは$i\neq j$について

$$\begin{aligned}&\mathrm{E}\left[(X_i-\bar{X})\left(X_j-\bar{X}\right)\times U_i\times U_j\right|x_1,\ldots,x_n\\&=\ (x_i-\bar{x})\,(x_j-\bar{x})\times\mathrm{E}[U_i\times U_j|\,x_1,\ldots,x_n]\\&=\ 0\end{aligned}$$

となることに注意して整理すると，

$$\mathrm{V}[\hat{\beta}_1|x_1,\ldots,x_n]=\mathrm{E}[\left(\hat{\beta}_1-\beta_1\right)^2|\,x_1,\ldots,x_n]=\frac{s^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

となります。

次に切片パラメターβ_0の最小 2 乗推定量$\hat{\beta}_0$の分散が

$$\mathrm{V}[\hat{\beta}_0|x_1,\ldots,x_n]=s^2\left(\frac{1}{n}+\frac{\bar{x}^2}{\sum(x_i-\bar{x})^2}\right)$$

となることを確かめましょう。β_0の最小 2 乗推定量

$$\hat{\beta}_0=\bar{Y}-\hat{\beta}_1\bar{X}$$

に単回帰モデル

$$Y_i=\beta_0+\beta_1X_i+U_i$$

を代入すると，

$$\begin{aligned}\hat{\beta}_0&=\frac{1}{n}\sum_{i=1}^{n}(\beta_0+\beta_1X_i+U_i)-\hat{\beta}_1\bar{X}=\beta_0-\left(\hat{\beta}_1-\beta_1\right)\bar{X}+\frac{1}{n}\sum_{i=1}^{n}U_i\\&=\beta_0-\frac{\frac{1}{n}\sum(X_i-\bar{X})U_i}{\frac{1}{n}\sum(X_i-\bar{X})^2}\bar{X}+\frac{1}{n}\sum_{i=1}^{n}U_i\end{aligned}$$

になります。不偏性の証明と同様に，（$x_1,\ldots,x_n$）で条件付けした分散をはじめに見ると，

$$\mathrm{V}[\hat{\beta}_0|x_1,\ldots,x_n]=\mathrm{E}[\left(\hat{\beta}_0-\beta_0\right)^2|\,x_1,\ldots,x_n]$$
$$=\mathrm{E}\left[\left[-\frac{\frac{1}{n}\sum_{i=1}^n(X_i-\bar{X})U_i}{\frac{1}{n}\sum_{i=1}^n(X_i-\bar{X})^2}\bar{X}+\frac{1}{n}\sum_{i=1}^n U_i\right]^2|\,x_1,\ldots,x_n\right]$$

となります。$d_i\equiv\frac{(X_i-\overline{X})\bar{X}}{\frac{1}{n}\sum_{i=1}^n(X_i-\bar{X})^2}$とし，傾きパラメターの時と同じように展開して整理すると，

$$\left[-\frac{\frac{1}{n}\sum_{i=1}^n(X_i-\bar{X})U_i}{\frac{1}{n}\sum_{i=1}^n(X_i-\bar{X})^2}\bar{X}+\frac{1}{n}\sum_{i=1}^n U_i\right]^2=\left[\frac{1}{n}\sum_{i=1}^n U_i-\frac{1}{n}\sum_{i=1}^n d_iU_i\right]^2$$
$$=\frac{1}{n^2}\left[\sum_{i=1}^n(1-d_i)U_i\right]^2$$

となるので，$\sum_{i=1}^n d_i=0$に注意しながら展開すると，

$$\mathrm{V}[\hat{\beta}_0|x_1,\ldots,x_n]=\frac{1}{n^2}\ \mathrm{E}\left[\left[\sum_{i=1}^n(1-d_i)U_i\right]^2|\,x_1,\ldots,x_n\right]$$
$$=\frac{1}{n^2}\mathrm{E}\left[\sum_{i=1}^n(1-d_i)^2U_i^2\,|\,x_1,\ldots,x_n\right]=\frac{s^2}{n^2}\sum_{i=1}^n(1-d_i)^2$$
$$=\frac{s^2}{n^2}\sum_{i=1}^n\left(1-2d_i+d_i^2\right)=\frac{s^2}{n^2}\left(n+\frac{n\bar{x}^2}{\frac{1}{n}\sum(x_i-\bar{x})^2}\right)$$
$$=s^2\left(\frac{1}{n}+\frac{\bar{x}^2}{\sum(x_i-\bar{x})^2}\right)$$

となります。

補論 3　単回帰モデルにおける頑健な標準誤差の求め方

この補論では，単回帰モデルにおける，分散不均一性に対しても頑健な標準誤差の求め方について確認しましょう。

単回帰モデルの傾きパラメターの最小 2 乗推定量に母集団モデルを代入したものは，

$$\hat{\beta}_1 = \beta_1 + \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})U_i}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

と書くことができました（補論 2 の 1 つ目の式）。これを使って$\hat{\beta}_1$の（すべての説明変数で条件付けした）分散は，分散が均一かどうかに関係なく，

$$\begin{aligned}\mathrm{V}[\hat{\beta}_1|x_1,\dots,x_n] &= \mathrm{E}\left[\left(\hat{\beta}_1 - \beta_1\right)^2 | x_1,\dots,x_n\right] \\ &= \mathrm{E}\left[\left[\frac{\sum_{i=1}^{n}(X_i - \bar{X})U_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right]^2 \middle| x_1,\dots,x_n\right]\end{aligned}$$

と書くことができます。補論 2 で見たように，不偏性のための仮定 2 と分散均一の仮定，$\mathrm{E}[U_i^2|x_1,\dots,x_n] = s^2$のもとでは，

$$\begin{aligned}\mathrm{E}\left[\left[\frac{\sum_{i=1}^{n}(X_i - \bar{X})U_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right]^2 \middle| x_1,\dots,x_n\right] &= \frac{\mathrm{E}[\sum_{i=1}^{n}(X_i - \bar{X})^2 U_i^2 | x_1,\dots,x_n]}{[\sum_{i=1}^{n}(x_i - \bar{x})^2]^2} \\ &= \left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]\frac{\mathrm{E}[U_i^2|x_1,\dots,x_n]}{[\sum_{i=1}^{n}(x_i - \bar{x})^2]^2} \\ &= \frac{\mathrm{E}[U_i^2|x_1,\dots,x_n]}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{s^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\end{aligned}$$

と書くことができ，s^2を残差$\hat{u}_i$から推定することで分散を求めることができました。一方，分散が不均一のときは，

$$\mathrm{E}\left[\left[\frac{\sum_{i=1}^{n}(X_i - \bar{X})U_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right]^2 \middle| x_1,\dots,x_n\right] = \frac{\mathrm{E}[\sum_{i=1}^{n}(X_i - \bar{X})^2 U_i^2 | x_1,\dots,x_n]}{[\sum_{i=1}^{n}(x_i - \bar{x})^2]^2}$$

の式の誤差項に残差 $\hat{u}_i$を代入したもの，

$$\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2\,\hat{u}_i^2}{\left[\sum_{i=1}^{n}(x_i-\bar{x})^2\right]^2}$$

になります。

補論 4　回帰解剖

重回帰モデル

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + U$$

の$k+1$個の回帰パラメターを最小 2 乗法で推定する際には，残差 2 乗和の最小化でもモーメント法でも$k+1$本の式からなる連立方程式を解くことによって$k+1$個のパラメターの推定値を求めることができました。この補論では，別の解法として，「回帰解剖」とよばれる方法による回帰パラメターの推定値の求め方を紹介します。

ここでは説明変数が政策変数Xと共変量Cの 2 つの重回帰モデル

$$Y = \beta_0 + \beta_1 X + \beta_2 C + U$$

を考えて，政策変数Xの傾きパラメターβ_1を求めてみましょう（ほかの係数もまったく同じ方法で推定できますし，ここでの考え方は説明変数が 3 つ以上の場合にも使えます)。

まず，この説明変数Xをその他の説明変数であるCに回帰する単回帰モデル

$$X_i = \delta_0 + \delta_1 C_i + V_{Xi}$$

（V_{Xi}はこの回帰式の誤差項です）を最小 2 乗法で推定します。こうして得られた回帰パラメターの推定値を使って，残差

$$\hat{V}_{Xi} = X_i - \hat{\delta}_0 - \hat{\delta}_1 C_i$$

を求めます。最後に被説明変数 Y をこの残差に回帰する単回帰モデル

$$Y_i = b_0 + b_1 \hat{V}_{Xi} + V_i$$

の傾きパラメターb_1を最小 2 乗法で推定すると，この推定量はもともとの重回帰モデルの説明変数 X の傾きパラメターβ_1の最小 2 乗推定量と同じになる，つまり，

$$\hat{b}_1 = \hat{\beta}_1$$

となります。

この回帰解剖による傾きパラメターの求め方は，重回帰分析が「他の説明変数を一定（セ

テリス・パリブス)」とする分析であることをとてもよく表しています。重回帰分析において，説明変数Xの傾きパラメターβ_1は「共変量Cを一定」としたときに，政策変数Xが成果変数Yに与える影響を表しています。この回帰解剖における残差$\hat{V}_{Xi}$は，「共変量Cの影響を取り除いたうえでも残っている，政策変数Xの“純粋な”変化」と見ることができます。この残差は，共変量の影響を取り除いた後のものですので，この残差と成果変数との関係は，「共変量の影響を制御した」うえでの関係，つまり，共変量を一定としたときの政策変数Xと成果変数Yの関係と言えます。結果として，最小 2 乗法も回帰解剖も同じ「セテリス・パリブス」分析を行っていると言えるのです。

補論 5　最小 2 乗推定量の不偏性の証明

不偏性のための 4 つの仮定が成り立っているときには，最小 2 乗法は不偏性を持つ，つまり平均的に正しいパラメターの値を知ることのできる方法になっています。この補論では，単回帰分析においてこの 4 つの仮定が満たされていると最小 2 乗推定量が不偏性を持つことを確認しておきましょう。

まず，傾きパラメターの最小 2 乗推定量からチェックします。傾きパラメターの最小 2 乗推定量

$$\hat{\beta}_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

には被説明変数Y_iが含まれています。ここに単回帰モデル$Y_i = \beta_0 + \beta_1 X_i + U_i$を代入して説明変数と誤差項を使った形に書き直すと，

$$\hat{\beta}_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(\beta_0 + \beta_1 X_i + U_i - \bar{Y})}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

となります。$\bar{Y} = \beta_0 + \beta_1 \bar{X}$に注意しながら整理すると，

$$\hat{\beta}_1 = \beta_1 + \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})U_i}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

となります。この式の両辺をn人分全員の説明変数（$X_1 = x_1, \dots, X_n = x_n$）で条件付けし，傾きパラメターの最小 2 乗推定量の条件付き期待値$\mathrm{E}[\hat{\beta}_1 | x_1, \dots, x_n]$を計算すると，不偏性のための仮定 2（無作為抽出）と 3（平均独立$\mathrm{E}[U_i | x_i] = 0$）のもとでは$\mathrm{E}[U_i | x_1, \dots, x_n] = 0$となるので，

$$\begin{aligned}
&\mathrm{E}[\hat{\beta}_1 | x_1, \dots, x_n] \\
&= \beta_1 + \mathrm{E}\left[\frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})U_i}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2} | x_1, \dots, x_n\right]
\end{aligned}$$

$$= \beta_1 + \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})\,\mathrm{E}[U_i|x_1, \dots, x_n\,]}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$= \beta_1$$

となります。つまり，母集団分布における傾きパラメターの値になります。

最後にこの条件付き期待値の説明変数Xに関する期待値をとります。いままで見てきた条件付き期待値は, 主に条件となる説明変数Xがそれぞれの値$(X_1 = x_1, \dots, X_n = x_n)$になるという条件のもとで計算した期待値でした。この条件となっている説明変数Xをそれが実現する前の確率変数と考えると，条件付き期待値の期待値を計算することができます。具体的には，$\mathrm{E}[\hat{\beta}_1|x_1, \dots, x_n]$の条件付け変数をそれらが実現する前の確率変数に戻したもの$\mathrm{E}[\hat{\beta}_1|X_1, \dots, X_n]$にして，この条件付き期待値の$(X_1, \dots, X_n)$についての期待値$\mathrm{E}[\mathrm{E}[\hat{\beta}_1|X_1, \dots, X_n]]$を計算できます（外側の期待値記号は$(X_1, \dots, X_n)$についての期待値を計算しています)。このように繰り返し期待値を計算する際には，次の関係が成り立っていることが知られています。

$$\mathrm{E}\left[\mathrm{E}\left[\hat{\beta}_1\middle|X_1, \dots, X_n\right]\right] = \mathrm{E}\left[\hat{\beta}_1\right]$$

この関係は補論 1 で見た**繰り返し期待値の法則**ですが，この関係を使うと，

$$\mathrm{E}[\hat{\beta}_1] = \mathrm{E}\Big[\mathrm{E}[\hat{\beta}_1|X_1, \dots, X_n]\Big] = \mathrm{E}[\beta_1] = \beta_1$$

となり，傾きパラメターの最小 2 乗推定量の期待値が母集団分布における「真の」傾きパラメターと同じになる，つまり不偏性を持つことがわかります (繰り返し期待値の法則は, 補論 1 を参照してください)。

切片パラメターの最小 2 乗推定量についても，同じ方法で不偏性を持つことが確認できます。ここでは，説明変数で条件付けした切片パラメターの最小 2 乗推定量$\hat{\beta}_0$の条件付き期待値が真のパラメターβ_0と同じになることを確認しておきます。

$$\mathrm{E}\left[\hat{\beta}_0\middle|x_1, \dots, x_n\right] = \mathrm{E}\left[\bar{Y} - \hat{\beta}_1\bar{X}\middle|x_1, \dots, x_n\right]$$

$$= \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n} Y_i - \hat{\beta}_1\bar{X}\middle|x_1, \dots, x_n\right]$$

$$= \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}(\beta_0 + \beta_1 X_i + U_i) - \hat{\beta}_1\bar{X}\middle|x_1, \dots, x_n\right]$$

$$= \beta_0 + \beta_1 \bar{x} + \mathrm{E}\left[\frac{1}{n}\textstyle\sum_{i=1}^{n} U_i \,\middle|\, x_1, \dots, x_n\right] - \mathrm{E}\left[\hat{\beta}_1 \middle| x_1, \dots, x_n\right]\bar{x}$$

$$= \beta_0 + \beta_1 \bar{x} + \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}[U_i | x_1, \dots, x_n] - \beta_1 \bar{x} = \beta_0$$

以上で，単回帰分析において不偏性のための 4 つの仮定が満たされているのであれば，最小 2 乗推定量は不偏性を持つことが確認できました。重回帰分析においても同様の方法によって不偏性を確認することができます。