モラルハザード

渡辺　隆裕

1. はじめに

依頼人が代理人の行動を観察できないときに，どのようにインセンティブ報酬を設計すれば代理人の**モラルハザード**（moral hazard）を防げるかという問題は，1970年代から研究され**依頼人と代理人の問題**（principle and agent problem）や**契約理論**（contract theory）と呼ばれる分野で盛んに研究されてきた．

この問題の応用例は広く，モラルハザードの語源となった保険会社と被保険者の問題，雇用主の雇用者に対する報酬，政府による公益企業や銀行の規制，株主による経営者のコントロールなどに応用されている．近年は，多くのテキストに取り上げられるようになったため，本理論に対する注目度が高くなってきている．私が教えるビジネススクールの学生は，投資家とベンチャーキャピタルの問題，成果主義報酬，保険会社と代理店との手数料契約，などに応用を試みているし，学部のゼミ生もこの理論により，アルバイトと店長の問題，フランチャイズと親会社とのロイヤリティなどの分析を試みている．

しかしながら，この理論を現実の最適報酬設計に用いて，具体的な報酬額を求めようとすると，代理人の効用関数の推定やプロジェクトの成否確率の計算の精度が悪く，困難が生じてうまくいかない．現時点においては，誰もがこの理論を応用し，現実の答を出せるような「道具」がまだ整備されていないのだ．したがって現在は，過去の事例研究や定性的な分析に留まっている感があり，現実に数値を出したいエンジニアやOR研究者には敬遠されることも多い．特に1980年代にこの問題を現実に応用しようとして手を出して失敗した研究者は，この理論の応用可能性に否定的である．

しかし最近の研究は「このような推定や計量をいかに行うか」という計量的な「道具」の開発に焦点が移ってきている感がある．多くの製品開発がそうであるように，この「基礎理論」が実際に使える「技術」として実現するのは，長い年月と多くの人の手による研究が必要であったのではないだろうか．私は間もなく，「道具」が揃い，この理論が実際に使える技術となる日が近いと考えている．

このような理由から，ここでは簡単なモラルハザードの問題をとりあげて図によって考察し，紹介したいと思う．これによって，理解を深め興味をもっていただければ幸いである．

2. 問題設定

投資家P（Principle）があるプロジェクトを立ち上げ，そのマネージャーとしてA（Agent）を起用しようとしている．ここで，プロジェクトが成功すればb_s，失敗すればb_fの利益が得られるとし，投資家はその利益からAへの報酬を支払い，その残りを自分の利益とする．プロジェクトの成否はAの努力に依存するが，努力するには，Aはより高いコストをかけなければならない．また，努力しても必ずしも成功するわけではないし，努力しなくても成功する可能性はある．Aが努力しないときのプロジェクトの成功確率をp_0でコストをc_0とし，努力するときの成功確率をp_1でコストをc_1とする．努力した方が成功確率は高い（すなわち$p_0<p_1$）とし，努力したときのコストは大きい（すなわち$c_0<c_1$）と仮定する．

Pは，プロジェクトが成功したか失敗したかは観察できるが，Aが努力したかどうかは観察できないものとする．もし成功と失敗に関わらず報酬が一定ならば，Aは努力せずに，確率p_0で運良くプロジェクトが成功する可能性を狙うだろう（モラルハザード）．そこでPは，プロジェクトが成功したときの報酬をw_s，失敗したときの報酬をw_fとしたインセンティブ報酬を用いて，Aを努力させようと試みる．

ここでPはリスク中立的であるが，Aはリスク回避的であるとする．報酬をw，費用をcとすると，

わたなべ　たかひろ
首都大学東京 大学院社会科学研究科
〒192-0397 八王子市南大沢1-1

Aの効用は$u(w)-g(c)$で表されるものとする．これはAの効用が，報酬に対する効用$u(w)$と，費用に対する負の効用関数$g(c)$に分離可能であることを表している．ここでuは単調増加で連続な凹関数であると仮定し，$u(0)=0$とする．

Aはあまりに報酬が低ければ別の投資家と契約を結ぶ．問題を簡単にするために，報酬から費用を差し引いた期待効用が非負であればPと契約を結ぶものとしよう．Pの期待利益が最大になる報酬契約(w_s, w_f)は，どのようなものだろうか．

3. ファーストベスト—Aの行動が観察できる場合

Aが努力したかどうかをPが観察できない場合は，Aのモラルハザードを防ぐために，Pはコストを支払わなければならない．この費用は，PがAの努力を観察できる「理想的な状態」との比較で語られる．この状況においてPの期待利益を最大にする報酬は**ファーストベスト**（first best）の報酬と呼ばれる．これに対し，PがAの行動を観察できない場合にPの期待利益を最大にする報酬は**セカンドベスト**（second best）と呼ばれる．そこでまずPがAの努力を観察できるようなファーストベストの場合について考察をしてみよう．

Pが利益を最大にするには，Aを努力させればよいかどうかは必ずしも定かではない．もしAの努力費用がとても大きく，そのために支払わなければならない報酬が大きければ，PはAを努力させない方が良い．そこで，PはAを努力させたときとさせないときの期待利益を比較し，高い方を選ぶことになる．ここではまずAを努力させたときの最適報酬を求めてみよう．

努力を判別できる状態では，「もしAが努力しないならば高い罰金を課す」という契約をすれば，Aを必ず努力させられる．したがってAの努力を前提にプロジェクトが失敗したときの報酬w_fと，成功したときの報酬w_sがどうなるかを考えれば良い．Pの期待利益を最大にする問題は以下のように書ける．

$$\begin{aligned} max \quad & p_1(b_s-w_s)+(1-p_1)(b_f-w_f) \\ s.t. \quad & p_1(u(w_s)-g(c_1)) \\ & +(1-p_i)(u(w_f)-g(c_1))\geq 0 \end{aligned}$$

制約式は「Aは期待効用が非負でなければプロジェクトには参加しない」という条件を表すもので，**参加制約**（Participation Constraint）と呼ばれる．ここで$g(c_i)=d_i, i=0,1$という記号（費用を負の効用として置きなおす）を用いると，問題は以下のように書きなおすことができる．

$$\begin{aligned} min \quad & p_1w_s+(1-p_1)w_f \\ s.t. \quad & p_1u(w_s)+(1-p_1)u(w_f)\geq d_1 \end{aligned} \tag{1}$$

すなわち「Aの報酬から得られる期待効用が，費用の負効用より大きくなる」という制約のもとで，Aの期待報酬を最小にすれば，Pの期待利益が最大になる．

縦軸をw_s，横軸をw_fとして，この様子を図示したものが図1である．図中の曲線は$p_1u(w_s)+(1-p_1)u(w_f)=d_1$となる制約条件式を表しており，その右上の領域が，Aがプロジェクトに参加するような報酬の領域，すなわち参加制約を満たす領域である．図の平行な直線は，式(1)の目的関数の等高線，すなわちAの期待報酬$p_1w_s+(1-p_1)w_f$が無差別になる線を表していて，この平行線が左下に行くほど期待報酬は小さくなる．領域の中で，目的関数を最小にするには，目的関数の等高線が制約領域で接する点FB_1で表される報酬とすれば良い．

ここで制約条件の接線の傾き$\frac{dw_s}{dw_F}$を求めると$-\frac{1-p_1}{p_1}\frac{u'(w_s)}{u'(w_f)}$となる．目的関数の等高線の傾きは$-\frac{1-p_1}{p_1}$であるから，制約条件に目的関数が接する点を$(\hat{w}_s, \hat{w}_f)$とすると$u'(\hat{w}_s)=u'(\hat{w}_f)$とならなければならない．これより最適な報酬は，$\hat{w}_s=\hat{w}_f$として成功時も失敗時も同じ報酬を与えることである．この値をw_1^*としよう．図ではこれは点FB_1が45度線上にあることを意味している．最適な報酬は$p_1u(w_1^*)+(1$

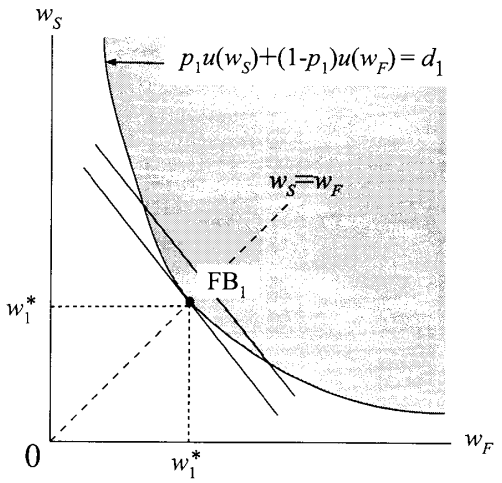

図1　ファーストベストの報酬

$-p_1)u(w_1^*)=d_1$ を満たすことより，$w_1^*=u^{-1}(d_1)$ と書くことができる．

Aを努力させないときの最適な報酬も同様に求めることができる．Aを努力させないときは，もしAが努力したら高い罰金を払うようにしておいて，努力しないもとでプロジェクトが成功したときも失敗したときも $w_0^*=u^{-1}(d_0)$ の報酬を払うようにすればよい．

4. セカンドベストーAの行動が観察できない場合

さて，いよいよAの行動が観察できない場合を考えてみよう．やはり最初はAを努力させる場合の報酬について考察する．Pの期待利益を最大にする報酬を求める問題は以下のように書ける．

$$\begin{aligned} min \quad & p_1 w_s+(1-p_1)w_f \\ s.t. \quad & p_1 u(w_s)+(1-p_1)u(w_f)\geq d_1 \\ & p_1 u(w_s)+(1-p_1)u(w_f)-d_1 \\ & \geq p_0 u(w_s)+(1-p_0)u(w_f)-d_0 \end{aligned}$$

これは式(1)に対して制約条件を1つ加えたものである．この2番目の条件は，Aが努力したときの期待効用が努力しなかったときの期待効用より大きくなければならないという条件で**誘因両立制約**（Incentive Compatibility Constraint）と呼ばれる．これが成り立たなければ，モラルハザードが起きる．

図2は図1に $p_0u(w_s)+(1-p_0)u(w_f)-d_0=0$ となる曲線を書き加えたものである．参加制約と誘引両立制約の両方を満たす領域である．領域の中で，目的関数を最小にするには，参加制約と誘引両立制約が交わる点 SB_1 を報酬とすれば良い．これを解くと最適な報酬 (w_s^*, w_f^*) は

$$u(w_s^*)=d_1+\frac{1-p_1}{p_1-p_0}(d_1-d_0)$$

$$u(w_f^*)=d_1-\frac{p_1}{p_1-p_0}(d_1-d_0)$$

を満たす報酬として与えられる．最適な報酬はファーストベストの場合に比べて，成功したときのAの効用を $\frac{1-p_1}{p_1-p_0}(d_1-d_0)$ だけ増加させ，失敗したときの効用を $\frac{p_0}{p_1-p_0}(d_1-d_0)$ だけ減少させるように設定し，成功することのインセンティブをより高くすれば良いことが分かる．

5. モラルハザードを防ぐためのコスト

図2において，点 FB_1，点 SB_1 を通る直線に対応する目的関数値は，それぞれファーストベストとセカンドベストにPがAに支払う報酬の期待金額を示している．この目的関数値の差は，PがAの行動を観察できない状況で，Aを努力させるために支払わなければならない追加的な費用を表していると考えることができる．この費用はどんな要因で決まるのか考えてみよう．

ここで，Aが努力しないときの費用 d_0 を Δd だけ減少させてみる（図3）．$p_0u(w_s)+(1-p_0)u(w_f)-d_0=0$ となる曲線は左下に少し移動し，最適な報酬は点 SB_1 から点 SB_2 に移動する．努力しないときの費用 d_0 が減少すれば，それだけPがAを努力させるために支払わなければならない費用は増えることが分かる．

一般的にモラルハザードを防ぐために支払わなければならない費用は，(1)Aが努力するときとしないときの費用の差 (d_1-d_0) や，(2)プロジェクトの成功確

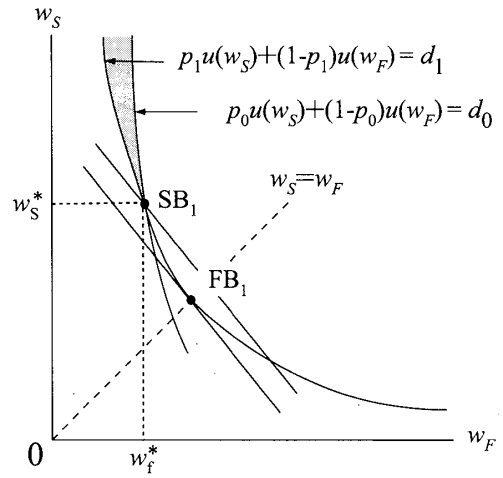

図2　セカンドベストの報酬

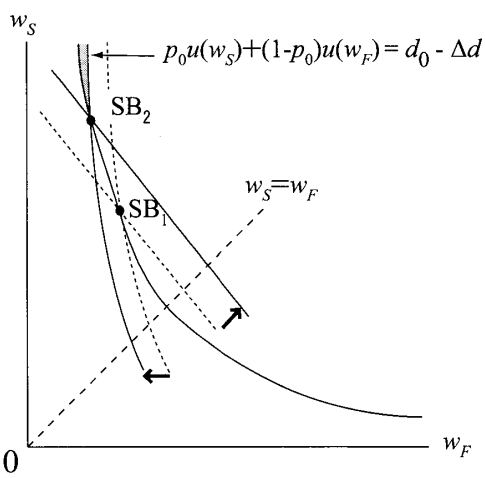

図3　努力時の費用の減少と報酬の変化

率の差 (p_1-p_0) が大きくなるほど高くなる．

6. Aを努力させない場合

さてセカンドベストの状況において，今度はAを努力させない場合を考察しよう．この場合のPの期待利益を最大にする問題は，以下のように書ける．

$$\begin{aligned} min \quad & p_0 w_s+(1-p_0)w_f \\ s.t. \quad & p_0 u(w_s)+(1-p_0)u(w_f)\geq d_0 \\ & p_0 u(w_s)+(1-p_0)u(w_f)-d_0 \\ & \geq p_1 u(w_s)+(1-p_1)u(w_f)-d_0 \end{aligned} \tag{2}$$

最初の制約式は，Aが努力しないときに契約に参加する参加制約，2つ目の制約式は，Aが努力しないときの期待効用が努力したときより大きくなければならないという誘因両立制約である．図2にこれらのすべての制約は書き表されているのだが，再度，Aが努力しなかったときに焦点を当て図を書き直してみよう（図4）．

図4において $p_0u(w_s)+(1-p_0)u(w_f)-d_0=0$ と45度線が交わる点 FB_0 は，Aを努力させないときのファーストベストにおける最適報酬 (w_0^*, w_0^*) を示している．$u^{-1}(d_1)\geq u^{-1}(d_0)$ から $w_1^*\geq w_0^*$ であり，点 FB_0 は点 FB_1 よりも左下に位置することが分かる．図に示された領域は，参加制約と誘引領域制約の両方を満たす領域であるが，図から分かるようにファーストベストにおける最適報酬 (w_0^*, w_0^*) は誘引領域制約も満たしており，セカンドベストにおいても依然として最適報酬である．すなわち，Aを努力させないときの報酬は，Aの行動が観察できても観察できなくても同じ (w_0^*, w_0^*) になる．

最後に，数値例を示す．ここで $u(w)=\sqrt{w}$，$d_0=20$，$d_1=40$，$p_0=0.25$，$p_1=0.75$，$b_f=400$ とする．このときプロジェクトが成功したときの利益である b_S を変化させたときのPの期待利益の変化が図5に示されている．ここで直線 FB_1，SB_1，FB_0 はそれぞれ，ファーストベストでAを努力させたとき，セカンドベストでAを努力させたとき，Aを努力させないとき（ファーストベストもセカンドベストも同じ）を表している．FB_1 は SB_1 を300だけ上方に移動させた直線となる．これはAの努力が観察できないときにAを努力させるためには，モラルハザードを防ぐために余計に300のコストがかかるということを示している．

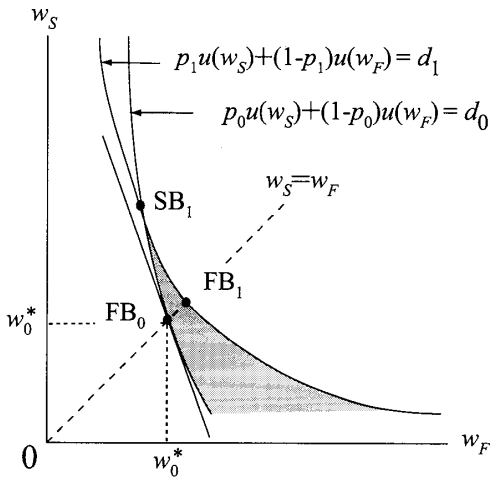

図4　Aが努力しないときの最適報酬

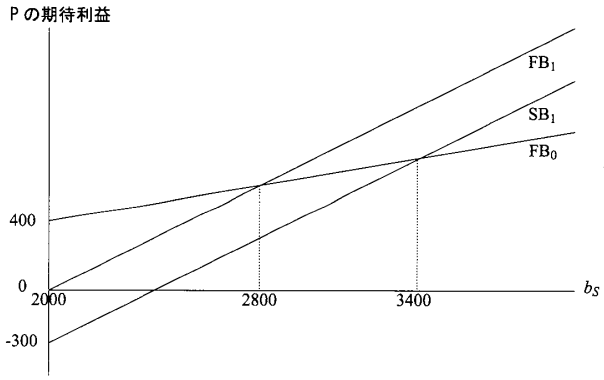

図5　b_S の変化によるPの期待利益

図より $b_S\leq 2800$ では，PはAを努力させない方が良い．$2800<b_S\leq 3400$ では，ファーストベストではAを努力させた方が良いが，セカンドベストにおいては，モラルハザードを防ぐための費用がかかるために努力させない方が良い領域である．この領域は，情報の非対称性が存在するために社会的な厚生が損なわれる領域でもある．$b_S>3400$ では，ファーストベストでもセカンドベストでもAを努力させた方が良い領域である．

以上，モラルハザードの理論について図によって解説をしてきた．さらに詳しく知りたい方は文献[1]などを参考にすると良い．

参考文献

[1] 伊藤秀史：『契約の経済理論』，有斐閣，2003.