数理解析研究所講究録
第1857巻 2013年 84-94

マルコフ決定過程における統計的手法について (Statistical Methods in Markov Decision Processes)

神奈川大学・理学部　堀口　正之 (Masayuki HORIGUCHI)
Faculty of Science, Kanagawa University

1 Introduction

本発表では, 推移確率行列が未知であるマルコフ決定過程 (Uncertain Markov Decision Processes) において, 推移確率行列の推定と行動 (action) の選択に関する考察を行う.

扱うモデルは, 有限状態空間をもち状態観測は確実にできる場合である. また, 停止問題に関しては, いつかは必ず停止する, すなわち有限期間内において停止の行動が取られるものとする. 推移確率行列の推定は, 古典的な問題として扱われる.

例えば, 状態推移法則のマルコフ性を調べる (検定する) ために, 初期分布 (p_i) と推移状態の頻度 (f_{ij}) の十分統計量を使って, 推移確率行列の χ^2 検定などを行う手法がある (Billingsley(1961)). ひとたび, マルコフ連鎖であることが過程できれば, 適切な条件下で定常分布が求まり最適化問題を容易に解くことができる.

また, 行動空間 (action space) が加わりマルコフ決定過程での同様の推移確率行列が未知の最適化問題については, 適応政策 (adaptive policy) の構成について, 最尤推定法によるもの (Kurano(1972), Mandl(1974) など) や reward-penalty type によるもの (Kurano(1987), Iki-Horiguchi-Yasuda-Kurano(2007)) などの先行研究がある. 例えば, 次のような統計量がある (Billingsley(1961)):

$S = \{1, 2, \ldots, s\}$,
$x_t = a_t$: t 期の状態,
$p_{a_1}p_{a_1a_2}\cdots p_{a_na_{n+1}}$: $n+1$ 期までの状態推移の確率
$F = (f_{ij})$: i から j への推移の回数表す $s \times s$ 行列

このとき,

$$p_{a_1}p_{a_1a_2}\cdots p_{a_na_{n+1}} = \Pi_{ij}p_{ij}^{f_{ij}}$$

と表すことができ,

$$\sum_j \frac{(f_{ij} - f_i p_{ij})^2}{f_i p_{ij}} \sim \chi^2(d_i - 1)$$

が成り立つ. ただし, d_i は $p_{ij} > 0$ を満たす状態 i の個数である.

また、マルコフ決定過程においては, 例えば, 平均期待利得 (Average reward criterion)

の場合には、Bellman 方程式:

$$g+h(i)=\max_{a\in A}\left(r(i,a)+\sum_{j=0}^{\infty}p_{ij}(a)h(j)\right),i\geqq 0$$

を満たすような optimal value g と relative value $h(i)$ が推移確率行列 $p_{ij}(a)$ から求められる. ここで, A は行動空間を表し, a は A の元で行動を表す. また, $r(i,a)$ は, 状態 i で行動 a を選択したときの 1 期間での利得の期待値を表す. マルコフ決定過程における各種の評価関数に対する Bellman 方程式については, 詳しくは [10] などを参照されたい.

推移確率行列が未知の場合には, 状態観測からの頻度分布をもとに, 推移法則や定常分布の推定と successive approximation によって評価関数値 (optimal value) の近似が行われる. また, 推移法則の逐次推定としてベイズ推定を用いることも有効であり, そのときには, 上記のような Bellman 方程式が導かれる (White(1969), Horiguchi-Piunovskiy([5],submitted)).

先行研究では, 推移確率行列に対して, 事前測度区間による区間ベイズ法 (De Robertis and Hartigan 1981[1]) の考え方で, 各成分が閉区間で表現される推移行列をもつ区間ベイズ MDPs(Interval Bayesian estimated MDPs) を構成した ([6]). さらに, 逐次抜き取り問題について考察し, リスク関数の評価の区間表現を行った ([4]). 本報告では, 論文 [4] での逐次抜き取り問題について, 損失関数を特徴づけている閾値を, 感度解析的な観点で, ベイズリスクがある一定の値以下となる場合について調べる.

2 Notation

以下のような, 逐次抜き取り検査問題について考える ([15]). 非常に多量な同製品のまとまり (製品群) からの抜き取り問題で, 一回ごと一つの製品を抜き取り不良品 (defective item) か良品 (non-defective item) かを検査する. 一回あたりの検査費用は c とする. ここでの目的は, 不良品を含む製品群の出荷を回避し, また, 良品を含む製品群を検査によって不良品と判断することをできるだけ回避することである. そこで, 製品群の不良率 p に対して, 損失関数 $a(p), r(p)$ をそれぞれ以下のように定義する:
未知の不良率 $p\in(0,1)$ があり,

$$\begin{cases} a(p)\text{: 不良率 } p \text{ の製品群を accept する時の損失} \\ r(p)\text{: 不良率 } p \text{ の製品群を reject する時の損失} \end{cases}$$

とする ($p\cong 0$ なら $a(p)=0$, $p\cong 1$ なら $r(p)=0$ とする. 一般には下に有界であれば良い).

抜き取り検査問題としては, 母不良率 p の検定として検出力に対するサンプルサイズの決定問題も一つの手法といえるが, ここでは不良率 p は非常に小さな値も取りえると

考えて逐次抜き取りによる検査費用とリスク評価での停止問題を考える.

今, $G_0(p)$ を不良率 p の事前分布, 観測度数 N において m 個の不良品と $n=(N-m)$ 個の良品を観測しているとする. m は二項分布に従うから, このときの p の事後分布 $G_{m,n}(p)$ は次の (2.1) のようになる.

記号の説明:

c : 一回当たりの検査費用,

$G_0(p)$: 不良率 p の事前分布,

$\{x_1, x_2, \ldots, x_N\}$: N 個の観測値,

$$dG_{m,n}(p) = \frac{p^m(1-p)^n\, dG_0(p)}{\int_0^1 p^m(1-p)^n\, dG_0(p)}. \tag{2.1}$$

$v(m,n)$ は, これまでの N 回の検査で m 個の不良品と n 個の良品であったという結果のもとで p の確率分布が $G_{m,n}(p)$ であり以後は最適政策を用いて得られる期待損失を表すとする. このとき, 簡単に (m,n) が十分統計量であることがわかり, 次の関数方程式が得られる.

(2.2)

$$\begin{aligned} v(m,n) &= \min \begin{cases} \text{Stop(Accept)}: \ \int_0^1 a(p)\, dG_{m,n}(p), \\ \text{Stop(Reject)}: \ \int_0^1 r(p)\, dG_{m,n}(p), \\ \text{Continue}: \ c + \int_0^1 \left(pv(m+1,n) + (1-p)v(m,n+1)\right)\, dG_{m,n}(p) \end{cases} \\ &= \min\left\{\psi(m,n), c + b_{m,n}v(m+1,n) + (1-b_{m,n})v(m,n+1)\right\}. \end{aligned}$$

ただし, $\psi(m,n) = \min\left\{\int_0^1 a(p)\, dG_{m,n}(p), \int_0^1 r(p)\, dG_{m,n}(p)\right\}$, $b_{m,n} = \int_0^1 p\, dG_{m,n}(p)$ であるとする.

以下の定理が成り立つことが知られている ([16]).

Theorem 2.1. 次の二つの条件

(i) $\psi(m,n) \geqq 0$ *for all* $m,n = 0,1,2,\ldots,$

(ii) $\psi(m,n) \to 0$ *as* $m+n \to \infty$

が成り立つならば, $v(m,n)$ は関数方程式 *(2.2)* の唯一の解である.

ここで, 事前分布を事前測度区間 $I(L,U)$ に置き換えた場合を考える (cf. [6]). 簡単のため, $I(L,U) = [L,kL]$ であって, L は $[0,1]$ 上のルベーグ測度であるとする.

リスク関数について, 簡単に以下のようにまとめておく:
危険関数 (risk function):

$$r(\theta,\delta)=\int_X \ell(\theta,\delta(\boldsymbol{x}))f_n(\boldsymbol{x}|\theta)\,dx.$$

ただし, $\ell(\theta,\delta(\boldsymbol{x}))$ は損失関数 (loss function) を表し, θ は母数パラメータ, $\delta(\boldsymbol{x})$ は決定関数を表す.

ベイズ危険 (Bayes risk): ($\pi(\theta)$:事前分布)

$$r_\beta(\delta)=\int_\Theta r(\theta,\delta)\pi(\theta)\,d\theta.$$

事後危険 (posterior risk):

$$r(\delta(\boldsymbol{x})|\boldsymbol{x})=\int_\Theta \ell(\theta,\delta(\boldsymbol{x}))\pi(\theta|\boldsymbol{x})\,d\theta.$$

損失関数を

$$\ell(d,p)=\begin{cases} a(p) & (d=d_1),\\ r(p) & (d=d_2)\end{cases}$$

とするときの, $Q\in I(L,U)$ に関するベイズリスクを $\beta(\ell,Q)$ と表して次のように定める:

$$\beta(\ell,Q)=\min_d\left\{\frac{Q(\ell(d,p))}{Q(1)}\right\}=\min\left\{\frac{Q(\ell(d_1,p))}{Q(1)},\frac{Q(\ell(d_2,p))}{Q(1)}\right\}. \tag{2.3}$$

ただし, 測度 $Q\in I(L,U)$ と可測関数 g に対して $Q(g)=\int g(p)\,dQ(p)$ と表すことにする. このとき, ベイズリスク (2.3) の区間表現は, $-\infty<\lambda<\infty$ に対して次の方程式の解 λ_1,λ_2 として得られる.

Theorem 2.2. *([1])*

(i) $\min\{\beta(\ell,\theta)|Q\in I(L,U)\}$ の下限値 λ_1 は次の方程式の唯一の解である:

$$\min_d\left\{U(\ell(d,p)-\lambda)^-+L(\ell(d,p)-\lambda)^+\right\}=0. \tag{2.4}$$

(ii) $\min\{\beta(\ell,\theta)|Q\in I(L,U)\}$ の上限値は次の方程式の唯一の解 λ_2 を超えない:

$$\min_d\left\{L(\ell(d,p)-\lambda)^-+U(\ell(d,p)-\lambda)^+\right\}=0. \tag{2.5}$$

ただし, $x^+ = \max\{0, x\}, x^- = x - x^+ = \min\{0, x\}$ とする.

ここで, 事前測度を $G_0(\cdot) \in I(L, U) = [dp, k\,dp]$ とし, そのときの事後区間測度 $G_{m,n}(\cdot) \in I(L_{m,n}, U_{m,n})$ に関して Theorem 2.2 を式 (2.2) で与えられている $\psi(m,n)$ に適用すると, $G_{m,n}(\cdot)$ に関する期待損失 (ベイズリスク) の区間表現 $[\underline{\psi}(m,n), \overline{\psi}(m,n)]$ を得る. さらに, 区間推定マルコフ決定過程 (interval estimated MDPs(Horiguchi(*preprint*))) の結果から, $b_{m,n}$ についても区間表現 $[\underline{b}_{m,n}, \overline{b}_{m,n}]$ が以下の方程式の解 $\underline{\lambda}, \overline{\lambda}$ として得られる:

$$\underline{\lambda} = \frac{Be(m+2, n+1) + (k-1)Be(m+2, n+1, \underline{\lambda})}{Be(m+1, n+1) + (k-1)Be(m+1, n+1, \underline{\lambda})}, \tag{2.6}$$

$$\overline{\lambda} = \frac{kBe(m+2, n+1) - (k-1)Be(m+2, n+1, \overline{\lambda})}{Be(m+1, n+1) - (k-1)Be(m+1, n+1, \overline{\lambda})}. \tag{2.7}$$

ただし, $Be(m+1, n+1) = \int_0^1 t^m(1-t)^n\,dt, Be(m+1, n+1, x) = \int_0^x t^m(1-t)^n\,dt$ である.

具体的な数値例は, [4] を参照のこと.

次が成り立つ.

Theorem 2.3. 区間表現 $[\underline{v}(m,n), \overline{v}(m,n)]$ の各端点について次が成り立つ:

(1) $\overline{v}(m,n) = \min\{$*Stop:* $\overline{\psi}(m,n)$, *Continue:* $\max\{c + \overline{b}_{m,n}\overline{v}(m+1,n) + (1-\overline{b}_{m,n})\overline{v}(m,n+1), c + \underline{b}_{m,n}\overline{v}(m+1,n) + (1-\underline{b}_{m,n})\overline{v}(m,n+1)\}\}$.

(2) $\underline{v}(m,n) = \min\{$*Stop:* $\underline{\psi}(m,n)$, *Continue:* $\min\{c + \underline{b}_{m,n}\underline{v}(m+1,n) + (1-\underline{b}_{m,n})\underline{v}(m,n+1), c + \overline{b}_{m,n}\underline{v}(m+1,n) + (1-\overline{b}_{m,n})\underline{v}(m,n+1)\}\}$.

(3) $\underline{v}(m,n), \overline{v}(m,n)$ のそれぞれに対する最適政策は, それぞれの関数方程式を満たす決定政策である.

3 Examples

[4] では, [11] の例題に沿って次のように条件設定をして, 区間ベイズ推定を適用した場合の数値例を考えた. 損失関数 $a(p), b(p)$ を

$$a(p) = \begin{cases} 0, & 0 \leqq p \leqq \frac{3}{4}, \\ 1, & \frac{3}{4} < p \leqq 1, \end{cases} \qquad r(p) = \begin{cases} 1, & 0 \leqq p \leqq \frac{1}{4}, \\ 0, & \frac{1}{4} < p \leqq 1, \end{cases}$$

とし, $G_0(p) = p$(一様), $I(L, U) = [L, kL] = [dp, k\,dp]$, $c = 0.04$ とする.

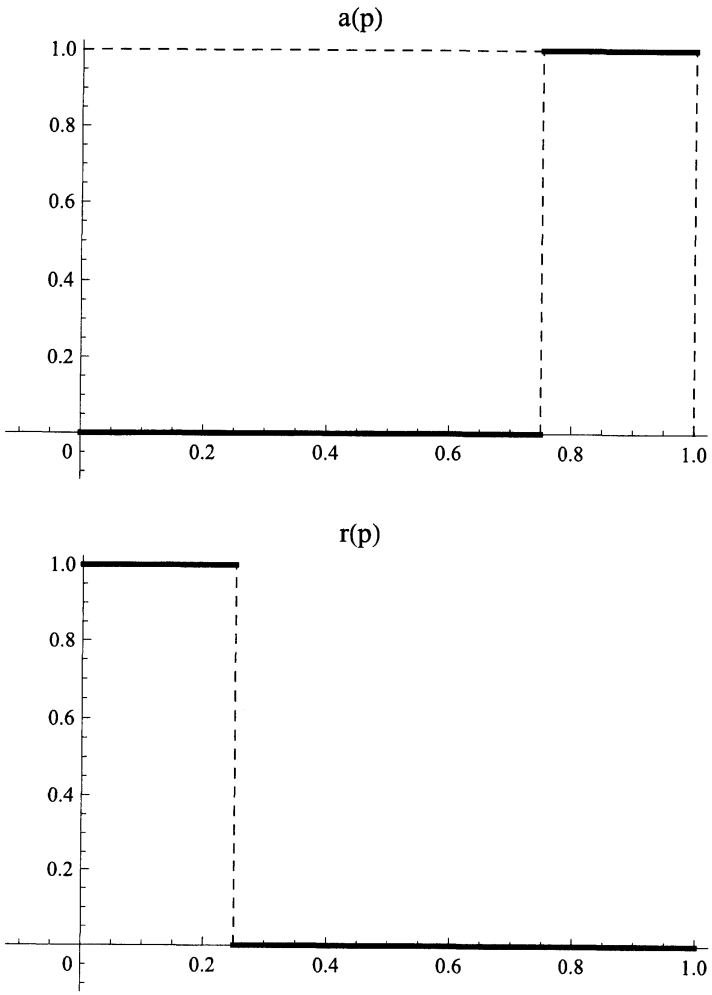

Figure 3.1: loss functions $a(p), r(p)$

ベイズリスクの区間表現$[\lambda_1, \lambda_2]$に関して, その下限値と上限値を求めるとLower Bayes risk (λ_1)に関しては, まずd_1とd_2のそれぞれについて期待損失に関する解として,

$$\lambda_1^{d_1} = \frac{1 - Be(m+1, n+1, \frac{3}{4})}{1 + (k-1)Be(m+1, n+1, \frac{3}{4})}, \tag{3.1}$$

$$\lambda_1^{d_2} = \frac{Be(m+1, n+1, \frac{1}{4})}{k + (1-k)Be(m+1, n+1, \frac{1}{4})}, \tag{3.2}$$

が得られ, Upper Bayes risk (λ_2) に関しては,

$$\lambda_2^{d_1} = \frac{k(1 - Be(m+1, n+1, \frac{3}{4}))}{k + (1-k)Be(m+1, n+1, \frac{3}{4})}, \tag{3.3}$$

$$\lambda_2^{d_2} = \frac{kBe(m+1, n+1, \frac{1}{4})}{1 + (k-1)Be(m+1, n+1, \frac{1}{4})} \tag{3.4}$$

を得る. 従って, $\lambda_1 = \min\{\lambda_1^{d_1}, \lambda_1^{d_2}\}$ と $\lambda_2 = \min\{\lambda_2^{d_1}, \lambda_2^{d_2}\}$ とからベイズリスクの下限値と上限値をそれぞれ具体的に求めることができる. また, 最適政策の期待損失に関する区間表現 $[\underline{v}(m,n), \overline{v}(m,n)]$ も Theorem 2.3 によって求めることができる.

ここで, $a(p), r(p)$ を次のように変えて区間表現を調べてみる.

$$a(p) = \begin{cases} 0, & 0 \leqq p \leqq a_1 \\ 1, & a_1 < p \leqq 1 \end{cases}, \quad r(p) = \begin{cases} 1, & 0 \leqq p \leqq r_1 \\ 0, & r_1 < p \leqq 1 \end{cases}. \tag{3.5}$$

すなわち, $a(p), r(p)$ のグラフにおけるジャンプの点 (不連続点) を a_1, r_1 とする. 上の式 (3.1), (3.2), (3.3), (3.4) と同様に, ベイズリスクの下限値と上限値が以下のように得られる.

Lower Bayes risk (λ_1):

$$\lambda_1^{d_1} = \frac{1 - Be(m+1, n+1, a_1)}{1 + (k-1)Be(m+1, n+1, a_1)}, \tag{3.6}$$

$$\lambda_1^{d_2} = \frac{Be(m+1, n+1, r_1)}{k + (1-k)Be(m+1, n+1, r_1)}. \tag{3.7}$$

Upper Bayes risk (λ_2):

$$\lambda_2^{d_1} = \frac{k(1 - Be(m+1, n+1, a_1))}{k + (1-k)Be(m+1, n+1, a_1)}, \tag{3.8}$$

$$\lambda_2^{d_2} = \frac{kBe(m+1, n+1, r_1)}{1 + (k-1)Be(m+1, n+1, r_1)}. \tag{3.9}$$

$\lambda_1^{d_1}, \lambda_2^{d_1}$ を a_1 についての関数, $\lambda_1^{d_2}, \lambda_2^{d_2}$ を r_1 についての関数とするとき, 次の補題のもとで, それぞれの関数の単調性が示される.

Lemma 3.1. $0 < x < x'$ となる実数 x, x' と実数 $k > 0$ に対して次が成り立つ.

$$\frac{x}{k+(1-k)x} < \frac{x'}{k+(1-k)x'}$$

不完全ベータ関数 $Be(m, n, x)$ は x に関して単調増加であるので, 上の Lemma 3.1 から次を得る.

Theorem 3.1. $\lambda_1^{d_1}, \lambda_2^{d_1}$ は a_1 に関して単調減少関数, $\lambda_1^{d_2}, \lambda_2^{d_2}$ は r_1 に関して単調増加関数である.

Remark: a_1, r_1 の値を行動空間の値と考えて, Bayes risk $\lambda_i^{d_j} \leqq c$ (ただし c は一回当たりの抜き取り費用を表す) となるような最小の a_1 の値と最大の r_1 の値を調べる (Table 3.1, Table 3.2).

Table 3.1: Lower (a_1, r_1)

m, n	0	1	2	3	4
0	0.9231, 0.0769	0.7227, 0.0392	0.5747, 0.0263	0.4737, 0.0198	0.4013, 0.0159
1	0.9608, 0.2774	0.8299, 0.1701	0.7092, 0.1233	0.6137, 0.0969	0.5390, 0.0798
2	0.9737, 0.4253	0.8767, 0.2908	0.7772, 0.2228	0.6920, 0.1810	0.6215, 0.1523
3	0.9802, 0.5266	0.9031, 0.3863	0.8190, 0.3079	0.7433, 0.2567	0.6779, 0.2204
4	0.9842, 0.5987	0.9202, 0.4610	0.8475, 0.3785	0.7796, 0.3221	0.7193, 0.2807

Table 3.2: Upper (a_1, r_1)

m, n	0	1	2	3	4
0	0.9796, 0.0204	0.8571, 0.0103	0.7267, 0.0068	0.6220, 0.0051	0.5408, 0.0041
1	0.9898, 0.1429	0.9151, 0.0849	0.8193, 0.0608	0.7315, 0.0474	0.6568, 0.0389
2	0.9932, 0.2733	0.9392, 0.1807	0.8637, 0.1363	0.7895, 0.1096	0.7227, 0.0918
3	0.9949, 0.3780	0.9526, 0.2685	0.8904, 0.2105	0.8263, 0.1737	0.7665, 0.1481
4	0.9959, 0.4592	0.9611, 0.3432	0.9082, 0.2773	0.8519, 0.2335	0.7980, 0.2020

損失関数の a_1, r_1 を変化させたときのベイズリスクの下限値, 上限値のグラフは以下のようになる. $m = 1, n = 2$ のときの図内 (Figure 3.2) のそれぞれの単調関数は, 単調減少関数はそれぞれ $\lambda_1^{d_1}, \lambda_2^{d_1}$ を表し, 単調増加関数はそれぞれ $\lambda_1^{d_2}, \lambda_2^{d_2}$ を表す. 例えば, 縦軸の値に一回当たりの抜き取り検査費用 c の値を取ったときの横軸に相当する値が Table 3.1, 3.2 の表内に示されている値である.

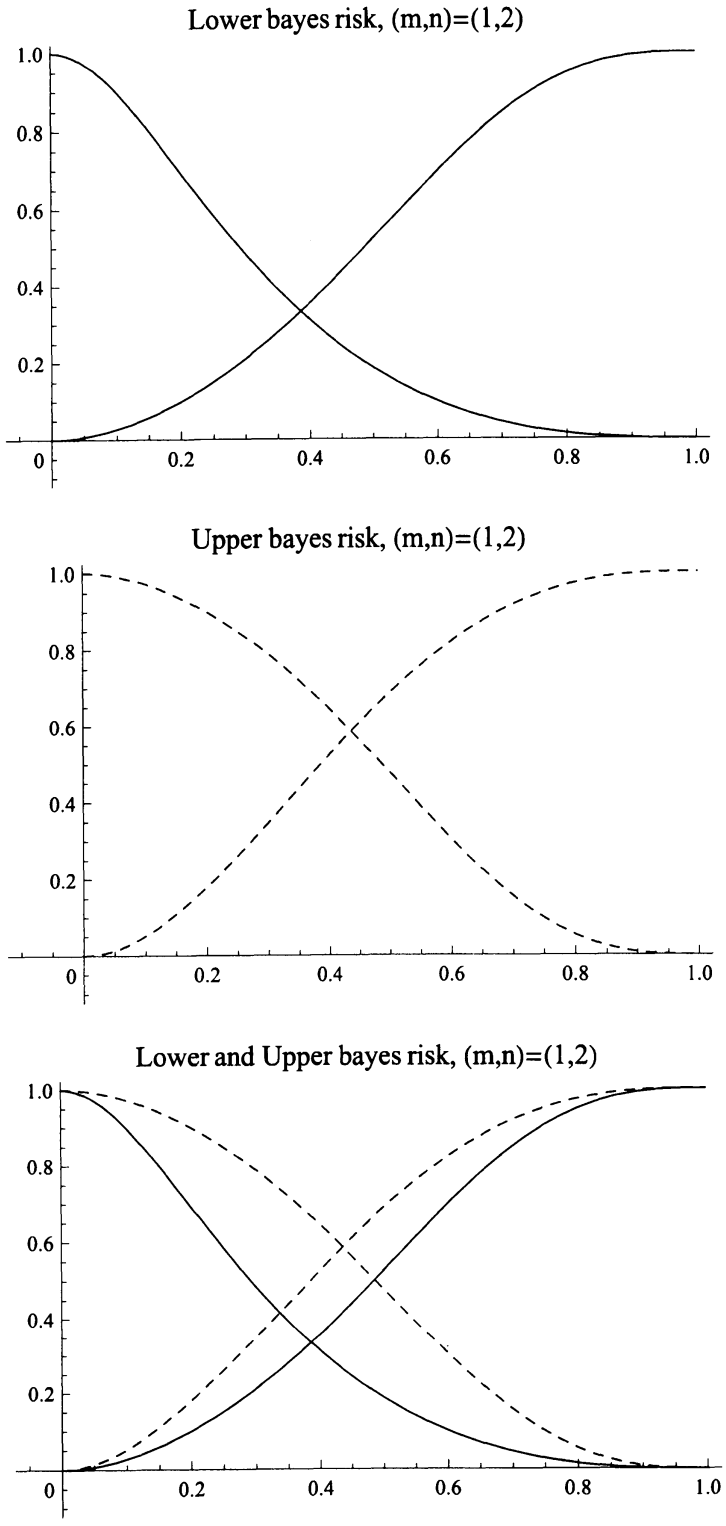

Figure 3.2: Lower and Upper bounds for $\psi(m,n) \leqq c = 0.04$

References

[1] L. De Robertis and J. A. Hartigan. Bayesian inference using intervals of measures. *Ann. Statist.*, 9(2):235–244, 1981.

[2] M. H. DeGroot. *Optimal Statistical Decisions.* Wiley Classics Library, reprint of the 1970 original, John Wiley & Sons, 2004.

[3] D. J. Hartfiel. *Markov set-chains*, volume 1695 of *Lecture Notes in Mathematics.* Springer-Verlag, Berlin, 1998.

[4] 堀口正之. 区間ベイズ手法と逐次抜き取り問題について. *RIMS* **講究録** *1802* **「不確実・不確定環境下における数理的意思決定とその周辺」**, pages 85–91, 2012.

[5] M. Horiguchi and A.B. Piunovskiy. Optimal stopping model with unknown transition probabilities. submitted, 2013.

[6] 伊喜哲一郎, 堀口正之, 安田正實, 蔵野正美. 不確実性の下でのマルコフ決定過程に対する区間ベイズ手法 (An interval bayesian method for uncertain MDPs), *RIMS* **講究録** *1636* **「不確実性と意思決定の数理」**, p.1-p.8, 2009/04.

[7] M. Kurano, J. Song, M. Hosaka, and Y. Huang. Controlled Markov set-chains with discounting. *J. Appl. Probab.*, 35(2):293–302, 1998.

[8] M. Kurano, M. Yasuda, and J. Nakagami. Interval methods for uncertain Markov decision processes. In *Markov processes and controlled Markov chains (Changsha, 1999)*, pages 223–232. Kluwer, 2002.

[9] J. J. Martin. *Bayesian decision problems and Markov chains.* Publications in Operations Research, No. 13. John Wiley & Sons Inc., 1967.

[10] M. L. Puterman. *Markov decision processes: discrete stochastic dynamic programming.* John Wiley & Sons Inc., 1994.

[11] 坂口実. **経済分析と動的計画**. 東洋経済新報社, 1970.

[12] K. M. van Hee. *Bayesian control of Markov chains*, volume 95 of *Mathematical Centre Tracts.* Mathematisch Centrum, 1978.

[13] G. B. Wetherill. Bayesian sequential analysis. *Biometrika*, 48(3):281–292, 1961.

[14] S. S. Wilks. *Mathematical statistics.* A Wiley Publication in Mathematical Statistics. John Wiley & Sons Inc., 1962. 田中英之, 岩本誠一 (訳),「数理統計学・増訂新版 1,2」, 1971,1972年, 東京図書.

[15] A. Wald. *Statistical Decision Functions.* John Wiley & Sons Inc., New York, N. Y., 1950.

[16] D. J. White. *Dynamic programming.* Oliver & Boyd, Edinburgh-London, 1969. Mathematical Economic Texts, 1.