Enrichment Designs for the Development of Personalized Medicines

Martin Posch, Alexandra Graf, Franz Koenig

Section of Medical Statistics
Center for Medical Statistics, Informatics, and Intelligent Systems
Medical University of Vienna
www.meduniwien.ac.at/medstat
martin.posch@meduniwien.ac.at

IGES 2014, Vienna

This project has received funding from the European Union's 7th Framework Programme for research, technological development and demonstration under the InSPiRe Grant Agreement no 602144.

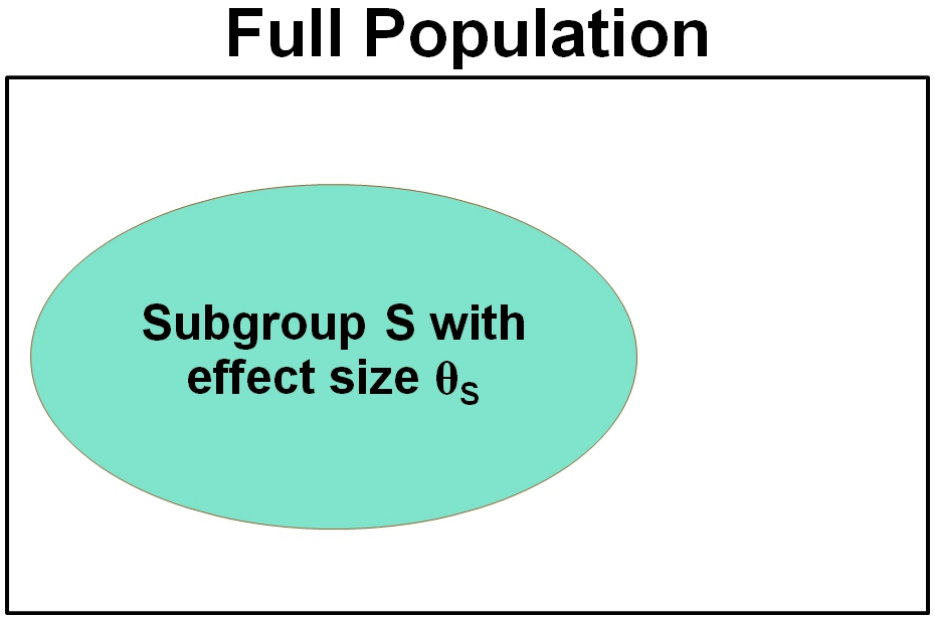
Identifying Target Populations

In recent years, clinical trials with more complex objectives, confirming treatment effects in sub-populations and/or in the overall populations, have raised more and more attention.

- The knowledge on the genetic basis of many diseases is increasing rapidly and therapies are developed that target underlying molecular mechanisms.
- Patients' responses are predicted to targeted treatments based on genetic features or other biomarkers.

Objective: Identify subgroups (based on biomarkers) where the treatment has a positive benefit risk balance.

Example



- Overall treatment effect

$$\theta_F = \lambda\theta_S + (1-\lambda)\theta_{S^c}$$

  where λ is the prevalence of subgroup S.
- Test the null hypotheses $H_F : \theta_F \leq 0$ and $H_S : \theta_S \leq 0$.
- If $\theta_{S^c} \ll \theta_S$ and the prevalence of the subgroup is small, the power to reject H_F is low.

Enrichment and Stratification Design

Enrichment Design: Randomize only patients of subgroup S (say Biomarker +). Patients of the complement S^C are excluded from the trial (Biomarker –).

Stratification Design: Include Biomarker + and Biomarker – patients. Stratify randomization by biomarker status.

- With both designs one can test H_S, i.e., for a treatment effect in the subpopulation.
- With the stratification design one can test in addition H_F.
- With equal overall sample size n, the enrichment design has a larger number of biomarker+ patients.

Testing Problem

- Parallel group comparison of the means of normal distributions.
- Total sample size n is chosen to detect an effect size Δ with a two sample one-sided z-test with $\alpha = 0.025$ and power of about 90%.

Enrichment Design:

- Test H_S with a z-test.

Stratification Design:

- Test H_S with a z-test
- Test H_F with a stratified z-test
- Correct for multiplicity with the Hochberg test.

To Enrich or not to Enrich?

- The power to reject any hypothesis depends on the effect sizes $\Theta = (\theta_S, \theta_{S^c})$
- Assume we suspect that $\theta_{S^c} \leq \theta_S$ but believe that $\theta_{S^c} > \theta_S$ is not plausible.
- Then, the enrichment design (recruiting only patients in S) always leads to the highest power:
  - If $\theta_{S^c} = \theta_S$ the enrichment design has the larger power than the stratification design using a multiple testing procedure.
  - If $\theta_{S^c} < \theta_S$ the enrichment design has larger power.

Thus, is enrichment always preferable?

To Enrich or not to Enrich?

- The power to reject any hypothesis depends on the effect sizes $\Theta = (\theta_S, \theta_{S^c})$
- Assume we suspect that $\theta_{S^c} \leq \theta_S$ but believe that $\theta_{S^c} > \theta_S$ is not plausible.
- Then, the enrichment design (recruiting only patients in S) always leads to the highest power:
  - If $\theta_{S^c} = \theta_S$ the enrichment design has the larger power than the stratification design using a multiple testing procedure.
  - If $\theta_{S^c} < \theta_S$ the enrichment design has larger power.

Thus, is enrichment always preferable?

Different Gains, different Costs...

The argument results from an oversimplification. The Power to reject any null hypotheses is not the only criteria.

- The stratification design tests the full population $H_F : \theta_F \leq 0$, demonstrating that the treatment works "on average".
- The enrichment design tests a limited null hypothesis $H_S : \theta_S \leq 0$ leading to a limited indication.
- Ethical problem if patients that may benefit are excluded.
- Enrichment maybe costly (e.g. due to longer patient recruitment, ...).

To account for these aspects we may use an approach based on utility functions.

The utility function approach

A sponsor view:

Rejection of	Gain
H_F	G_F
H_S only	G_S
none	0

$G_S \leq G_F$ (for example, $G_S = \lambda G_F$)

$$U_C(\Theta) = G_F P_\Theta(\text{reject } H_F) + G_S P_\Theta(\text{reject only} H_S)$$

A public health view:

θ_S	θ_{S^c}	Rejection of	Gain
+	+	H_F	G_F
+	+	H_S only	G_S
+	0	H_F	G_S
+	0	H_S only	G_S
0	0	H_S, H_F	0
0	0	None	0

$$U_P(\Theta) = G_F 1_{\{\theta_S, \theta_{S^c} > 0\}} P_\Theta(\text{reject } H_F) + G_S 1_{\{\theta_S, \theta_{S^c} > 0\}} P_\Theta(\text{reject only } H_S) + G_S 1_{\{\theta_S > 0, \theta_{S^c} \leq 0\}} P_\Theta(\text{reject } H_F) + G_S 1_{\{\theta_S > 0, \theta_{S^c} \leq 0\}} P_\Theta(\text{reject only } H_S)$$

The utility function approach

A sponsor view:

Rejection of	Gain
H_F	G_F
H_S only	G_S
none	0

$G_S \leq G_F$ (for example, $G_S = \lambda G_F$)

$$U_C(\Theta) = G_F P_\Theta(\text{reject } H_F) + G_S P_\Theta(\text{reject only} H_S)$$

A public health view:

θ_S	θ_{S^c}	Rejection of	Gain
+	+	H_F	G_F
+	+	H_S only	G_S
+	0	H_F	G_S
+	0	H_S only	G_S
0	0	H_S, H_F	0
0	0	None	0

$$U_P(\Theta) = G_F 1_{\{\theta_S, \theta_{S^c} > 0\}} P_\Theta(\text{reject } H_F) + G_S 1_{\{\theta_S, \theta_{S^c} > 0\}} P_\Theta(\text{reject only } H_S) +$$
$$G_S 1_{\{\theta_S > 0, \theta_{S^c} \leq 0\}} P_\Theta(\text{reject } H_F) + G_S 1_{\{\theta_S > 0, \theta_{S^c} \leq 0\}} P_\Theta(\text{reject only } H_S)$$

The utility function approach

A sponsor view:

Rejection of	Gain
H_F	G_F
H_S only	G_S
none	0

$G_S \leq G_F$ (for example, $G_S = \lambda G_F$)

$U_C(\Theta) = G_F P_\Theta(\text{reject } H_F) + G_S P_\Theta(\text{reject only} H_S)$

A public health view:

θ_S	θ_{S^c}	Rejection of	Gain
+	+	H_F	G_F
+	+	H_S only	G_S
+	0	H_F	G_S
+	0	H_S only	G_S
0	0	H_S, H_F	0
0	0	None	0

$U_P(\Theta) = G_F 1_{\{\theta_S, \theta_{S^c} > 0\}} P_\Theta(\text{reject } H_F) + G_S 1_{\{\theta_S, \theta_{S^c} > 0\}} P_\Theta(\text{reject only } H_S) +$
$G_S 1_{\{\theta_S > 0, \theta_{S^c} \leq 0\}} P_\Theta(\text{reject } H_F) + G_S 1_{\{\theta_S > 0, \theta_{S^c} \leq 0\}} P_\Theta(\text{reject only } H_S)$

Bayesian expected Utility

If $G_S = G_F = 1$ then $U_C = U_P$ is equal to the power of rejecting any hypothesis.

Consider a prior π on $\Theta = (\theta_S, \theta_{S^c})$. Then the expected utilities are

$$U_{\pi,C} = E_\pi(U_C(\Theta)), \quad U_{\pi,P} = E_\pi(U_P(\Theta))$$

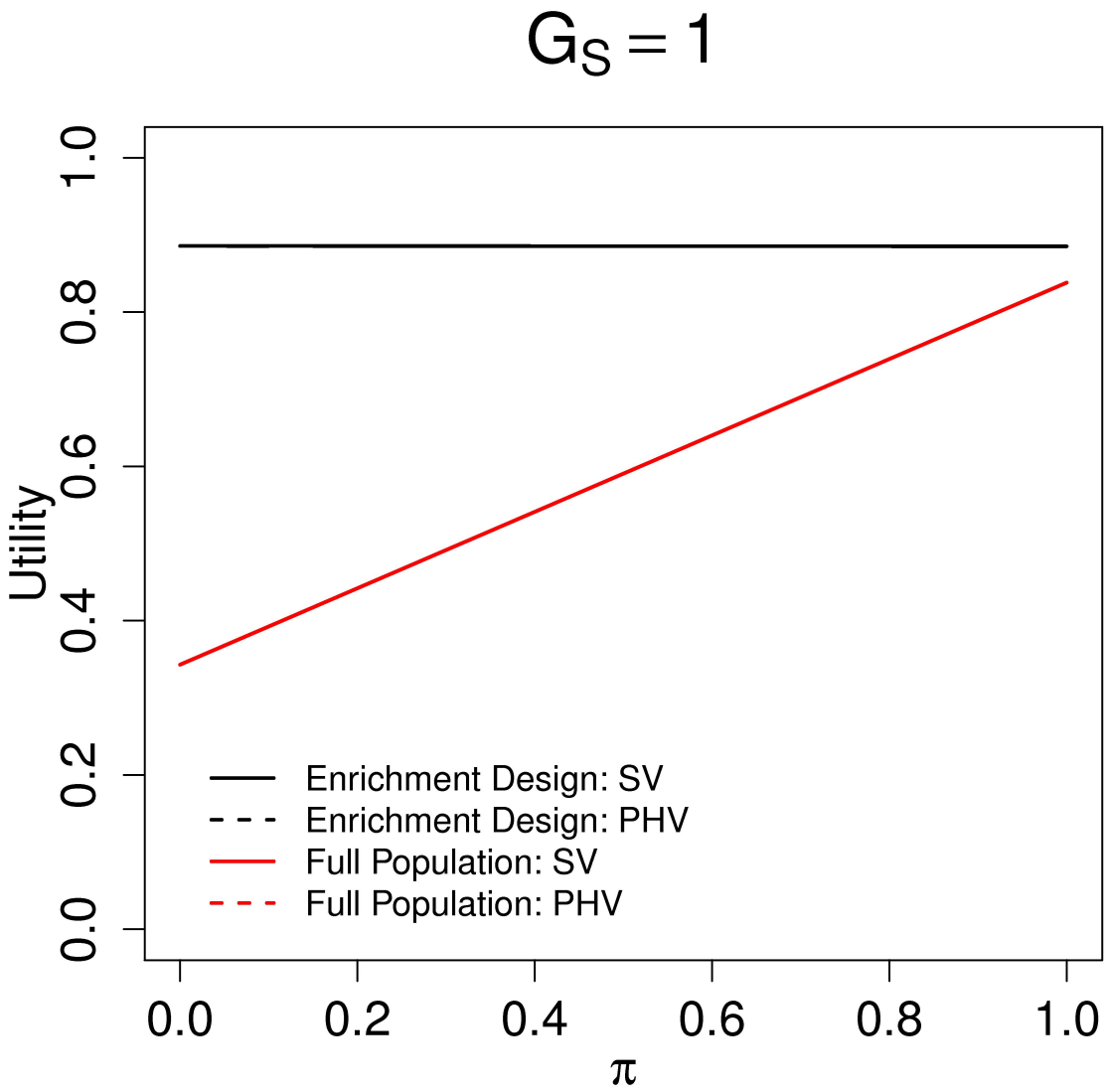
We use the simple prior

$$\begin{aligned} P\{\Theta = (\Delta, \Delta)\} &= \pi \\ P\{\Theta = (\Delta, 0)\} &= (1 - \pi) \end{aligned}$$

setting $G_F = 1$.

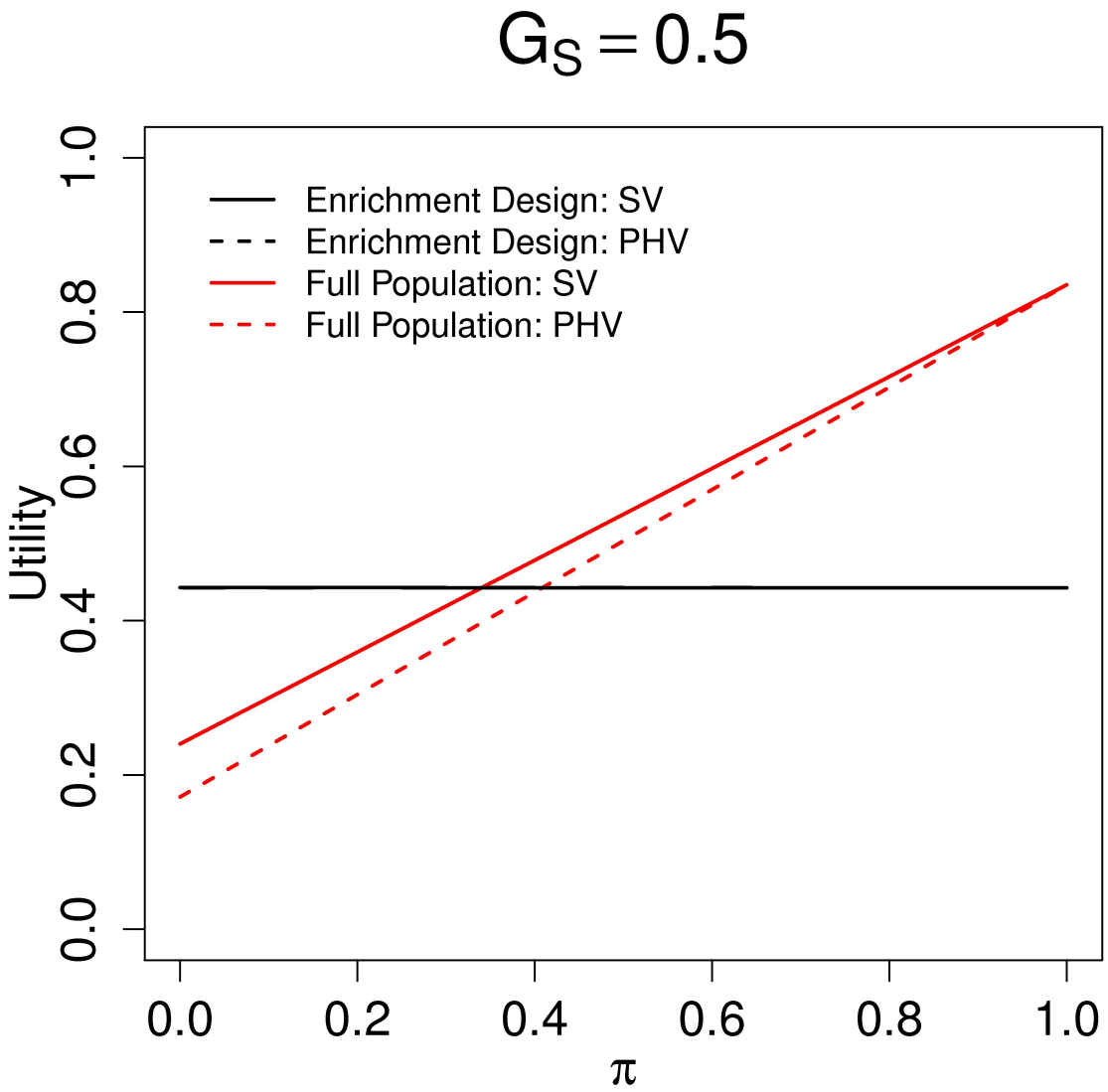
Utility for different Priors and Gains G_S

Prevalence $\lambda = 0.3$



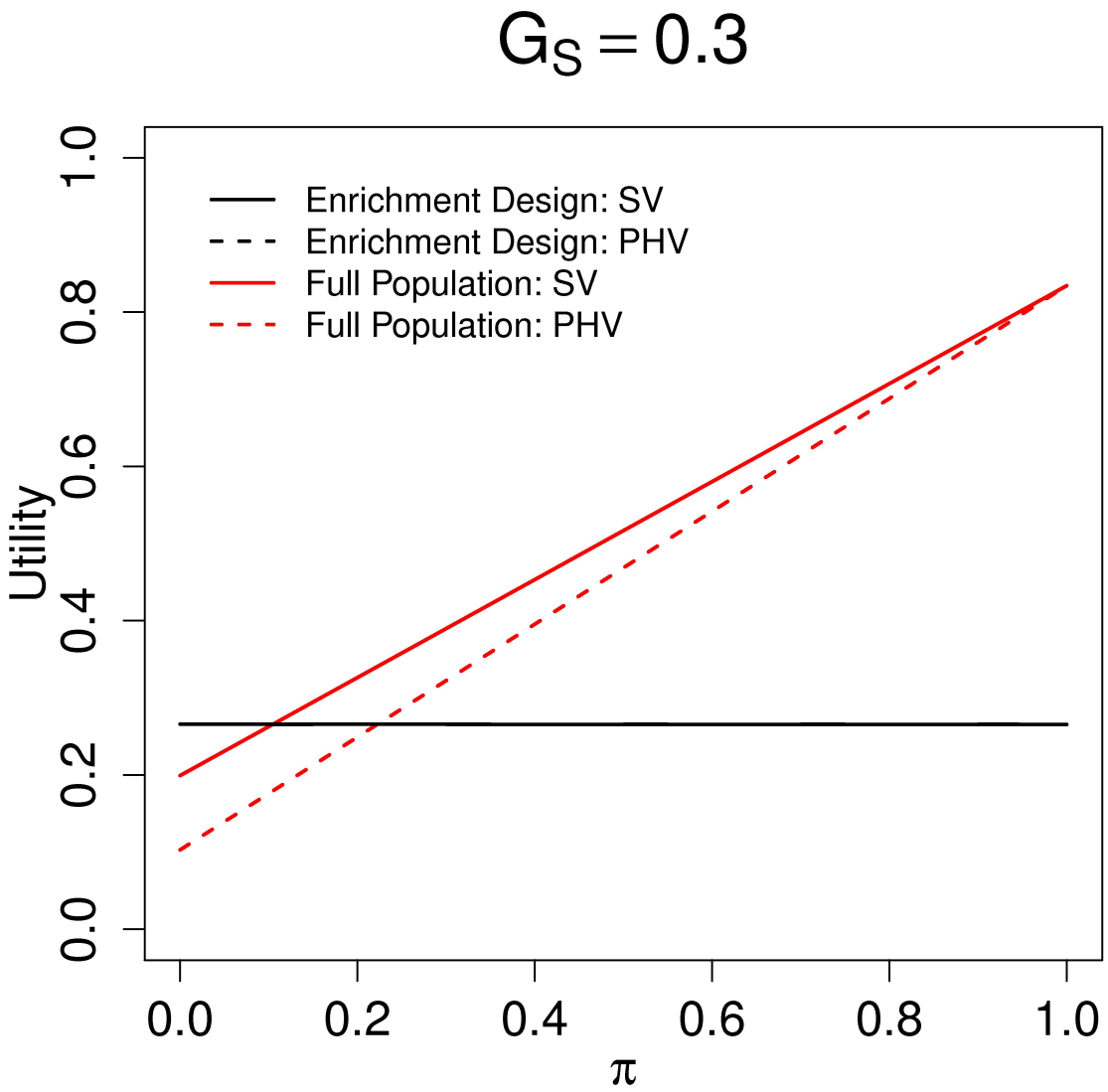
Utility for different Priors and Gains G_S

Prevalence $\lambda = 0.3$



Utility for different Priors and Gains G_S

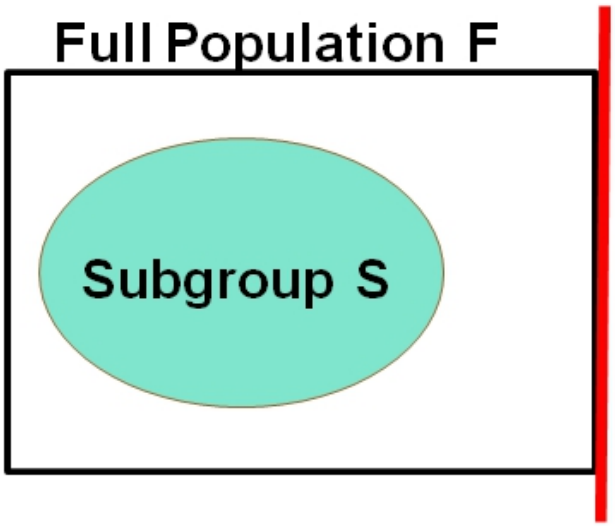
Prevalence $\lambda = 0.3$



Adaptive Trials with Population Enrichment

Fixed Overall Sample Size: n

Sample size Stage 1:
F: $n_1 = r*n$
S: $\lambda * n_1$



Decision based on interim results:

→ Continue only with S
→ Continue with F
(testing both: F and S)

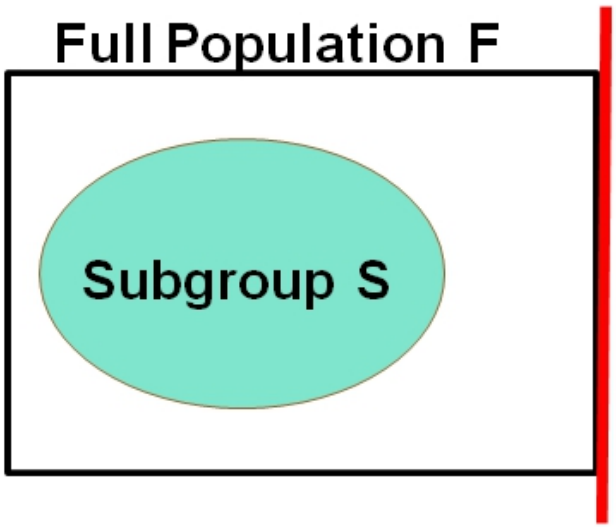
Interim Analysis
Planning of Second Stage
Subgroup Selection

- First stage data is used to choose the second stage population

Adaptive Trials with Population Enrichment

Fixed Overall Sample Size: n

Sample size Stage 1:
F: $n_1 = r*n$
S: $\lambda * n_1$



Example:

Decision based on the interim p-value p_{SC} of the Complement S^C

Introducing a „stopping-for-futility"-boundary α_C for S^C

Interim Analysis
Planning of Second Stage
Subgroup Selection

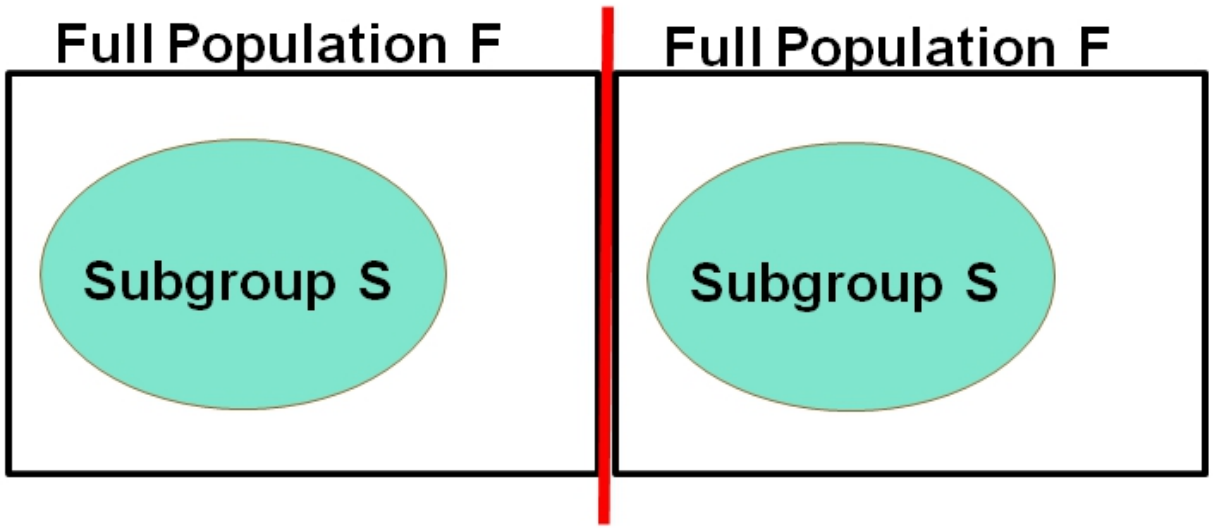
- First stage data is used to choose the second stage population

Adaptive Trials with Population Enrichment

Fixed Overall Sample Size: n

Sample size Stage 1:
F: $n_1 = r*n$
S: $\lambda * n_1$

Sample size Stage 2:
F: $n_2 = (1-r)*n$
S: $\lambda * n_2$



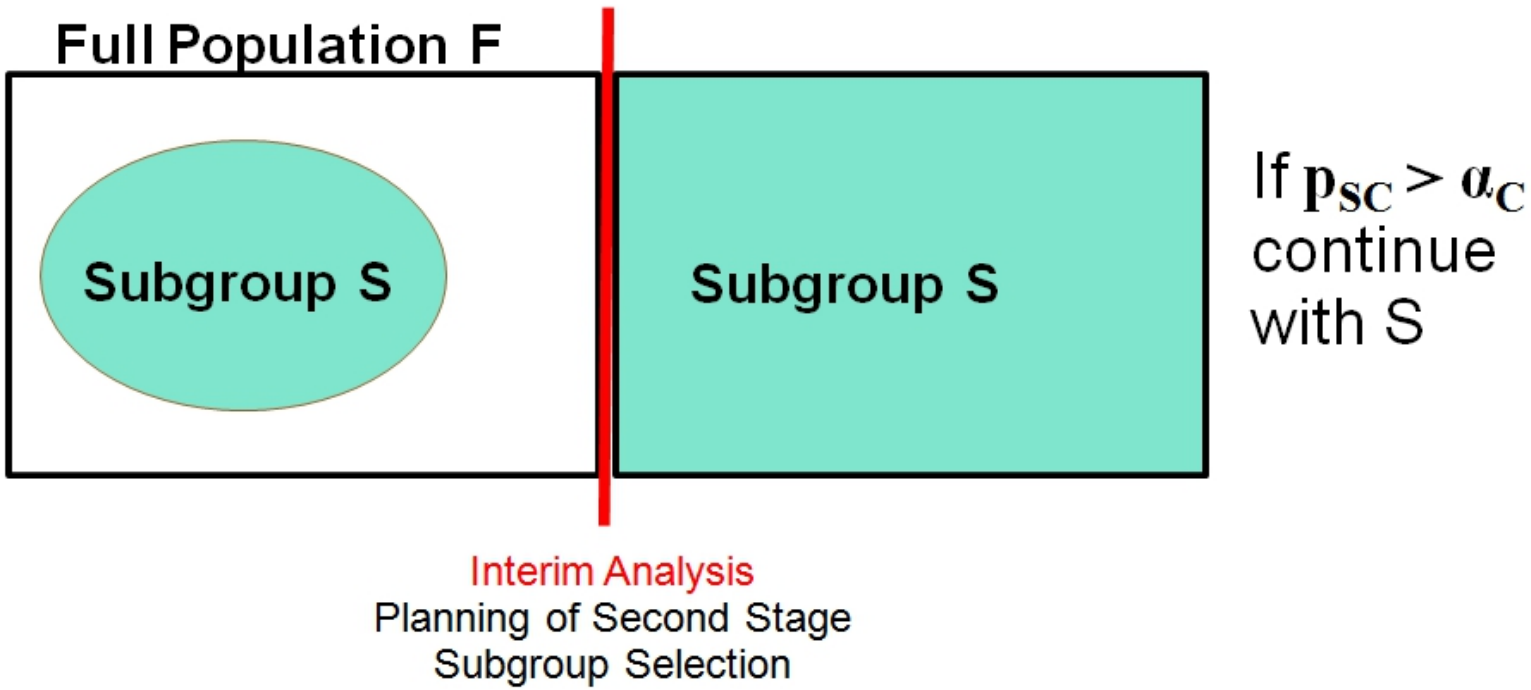
- First stage data is used to choose the second stage population
- Efficacy is demonstrated with Stage 1 and 2 data

Adaptive Trials with Population Enrichment

Fixed Overall Sample Size: n

Sample size Stage 1:
F: $n_1 = r*n$
S: $\lambda * n_1$

Sample size Stage 2:
S^C: -
S: $n_2 = (1-r)*n$



- First stage data is used to choose the second stage population
- Efficacy is demonstrated with Stage 1 and 2 data

The Adaptive Closed Test

- To control the family wise error rate apply the closure principle using adaptive combination tests at level α for

$$H_S, \quad H_F, \quad H_{FS} = H_S \cap H_F.$$

(Bauer and Kieser, 1999, Hommel, 2001)

- Reject $H_j, j \in \{S, F\}$ if H_{FS} and H_j are rejected at local level α.

Combination tests for elementary hypotheses H_S, H_F

- Compute stage wise p-values
  - First stage elementary p-values p_S, p_F
  - Second stage elementary p-values q_S, q_F (computed from second stage data only)
- Define a combination function $C(p, q)$ and critical value c such that for independent and uniformly distributed p-values

$$P(C(p, q) \leq c) = \alpha.$$

- Reject H_S if

$$C(p_S, q_S) \leq c$$

- If the trial continues in F, reject H_F if

$$C(p_F, q_F) \leq c,$$

otherwise retain H_F.

Combination tests for elementary hypotheses H_S, H_F

- Compute stage wise p-values
  - First stage elementary p-values p_S, p_F
  - Second stage elementary p-values q_S, q_F (computed from second stage data only)
- Define a combination function $C(p, q)$ and critical value c such that for independent and uniformly distributed p-values

$$P(C(p, q) \leq c) = \alpha.$$

- Reject H_S if

$$C(p_S, q_S) \leq c$$

- If the trial continues in F, reject H_F if

$$C(p_F, q_F) \leq c,$$

otherwise retain H_F.

Combination tests for elementary hypotheses H_S, H_F

- Compute stage wise p-values
  - First stage elementary p-values p_S, p_F
  - Second stage elementary p-values q_S, q_F (computed from second stage data only)
- Define a combination function $C(p, q)$ and critical value c such that for independent and uniformly distributed p-values

$$P(C(p, q) \leq c) = \alpha.$$

- Reject H_S if

$$C(p_S, q_S) \leq c$$

- If the trial continues in F, reject H_F if

$$C(p_F, q_F) \leq c,$$

otherwise retain H_F.

Combination tests for elementary hypotheses H_S, H_F

- Compute stage wise p-values
  - First stage elementary p-values p_S, p_F
  - Second stage elementary p-values q_S, q_F (computed from second stage data only)
- Define a combination function $C(p, q)$ and critical value c such that for independent and uniformly distributed p-values

$$P(C(p, q) \leq c) = \alpha.$$

- Reject H_S if

$$C(p_S, q_S) \leq c$$

- If the trial continues in F, reject H_F if

$$C(p_F, q_F) \leq c,$$

otherwise retain H_F.

Test of the intersection hypothesis $H_F \cap H_S$

- First stage p-value for $H_F \cap H_S$ with Simes test:

$$p_{FS} = \min[\max(p_F, p_S), 2\min(p_F, p_S)]$$

- Second stage p-value for $H_F \cap H_S$
  - If both populations are continued with Simes test:

$$q_{FS} = \min[\max(q_F, q_S), 2\min(q_F, q_S)]$$

  - If only H_S is selected:

$$q_{FS} = q_S$$

- Final Analysis: Reject H_{FS} if

$$C(p_{FS}, q_{FS}) \leq c$$

Test of the intersection hypothesis $H_F \cap H_S$

- First stage p-value for $H_F \cap H_S$ with Simes test:

$$p_{FS} = \min[\max(p_F, p_S), 2\min(p_F, p_S)]$$

- Second stage p-value for $H_F \cap H_S$
  - If both populations are continued with Simes test:

$$q_{FS} = \min[\max(q_F, q_S), 2\min(q_F, q_S)]$$

  - If only H_S is selected:

$$q_{FS} = q_S$$

- Final Analysis: Reject H_{FS} if

$$C(p_{FS}, q_{FS}) \leq c$$

Test of the intersection hypothesis $H_F \cap H_S$

- First stage p-value for $H_F \cap H_S$ with Simes test:

$$p_{FS} = \min[\max(p_F, p_S), 2\min(p_F, p_S)]$$

- Second stage p-value for $H_F \cap H_S$
  - If both populations are continued with Simes test:

$$q_{FS} = \min[\max(q_F, q_S), 2\min(q_F, q_S)]$$

  - If only H_S is selected:

$$q_{FS} = q_S$$

- Final Analysis: Reject H_{FS} if

$$C(p_{FS}, q_{FS}) \leq c$$

The Adaptive Closed Test

Adaptive Closed Test

Reject $H_i, i \in \{F, S\}$ if

- $C(p_{FS}, q_{FS}) \leq c$ and
- $C(p_i, q_i) \leq c$.

- The population selection rule may depend on the interim data and external data in any way.
- The selection rule needs not to be specified in detail.
- Sample sizes may be adapted based on unblinded interim data
- The familywise error rate is controlled in the strong sense.

Adaptation Rule: Example

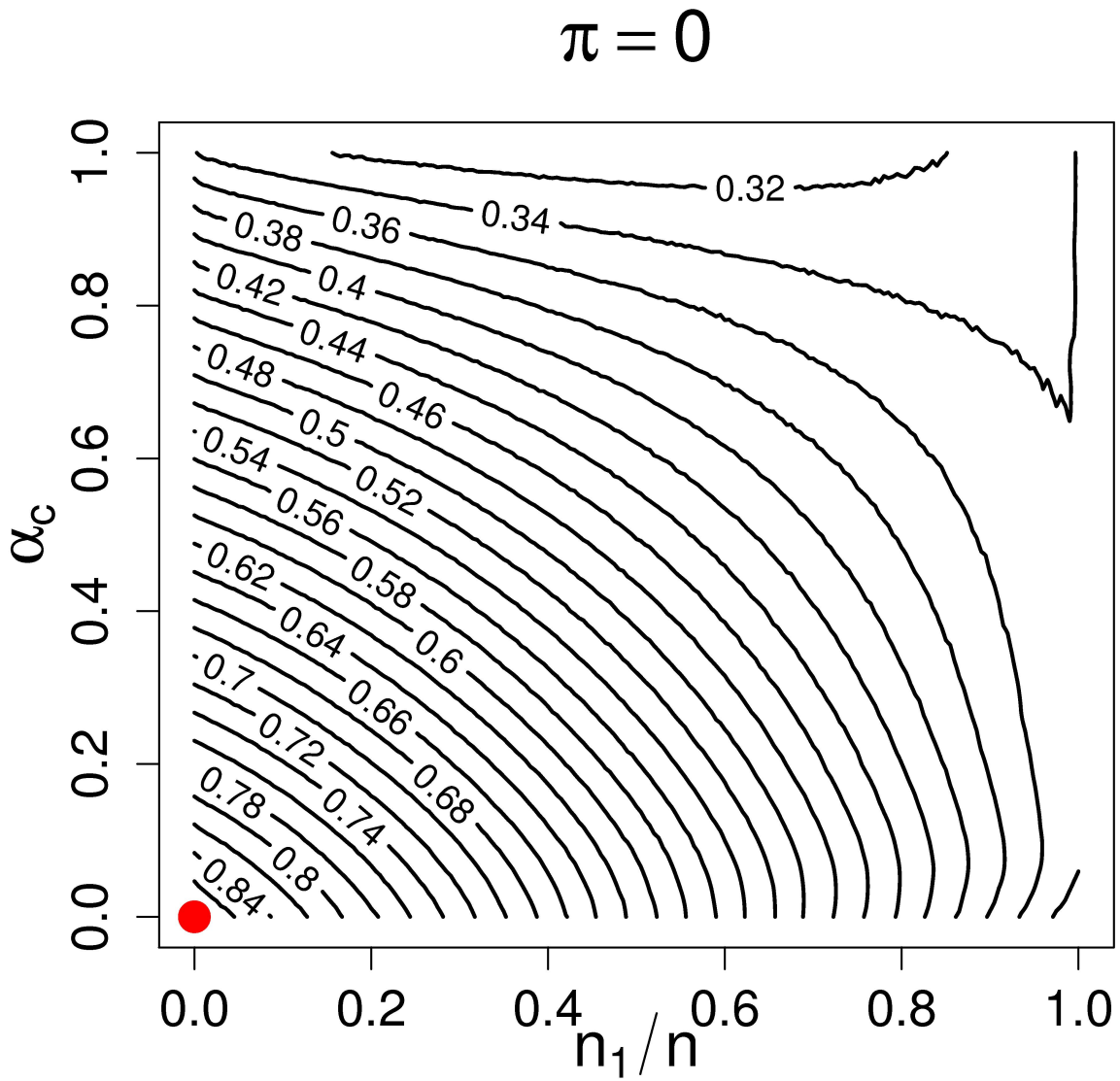
- Selection rule:
  - p_{SC} . . . the interim p-value for the z-test in the complement of the subpopulation.
  - α_C. . . selection threshold
  - Continue with F if $p_{SC} < \alpha_C$, otherwise enrich and continue with S only.
- Two types of adaptation: If the trial continues in the subpopulation only
  - Selection of hypothesis: H_F is dropped.
  - Reassessment of sample size: The sample size for H_S is increased.
- Combination Function: Inverse normal method (Lehmacher and Wassmer, 1999)

$\rightarrow$ $r = 0, \alpha_C = 0$ Fixed Sample Trial in S only

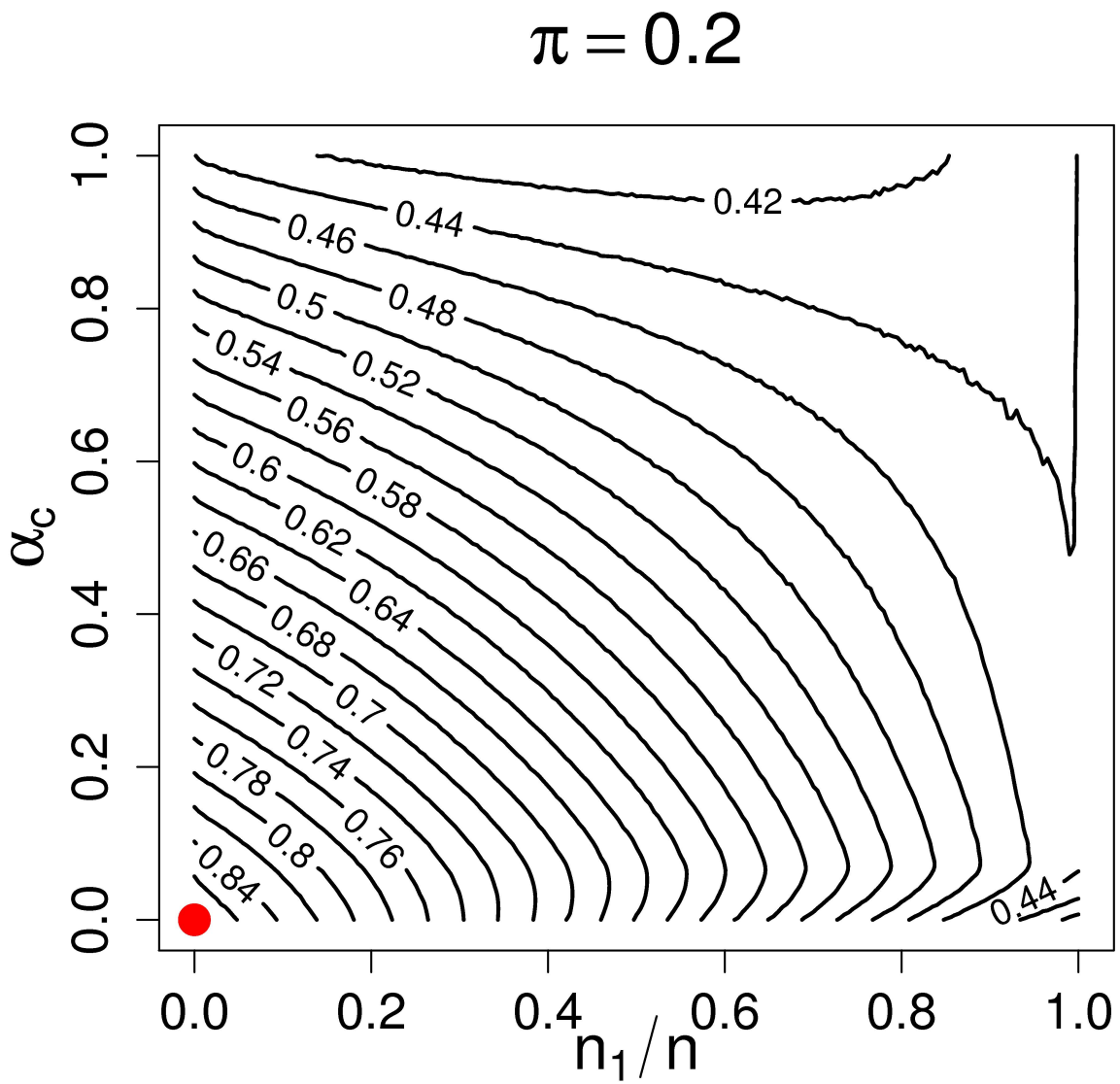
$\rightarrow$ $0 < r, \alpha_C < 1$ Adaptive Design integrating both phases

$\rightarrow$ $r = 1, \alpha_C = 1$ Fixed Sample Trial in F

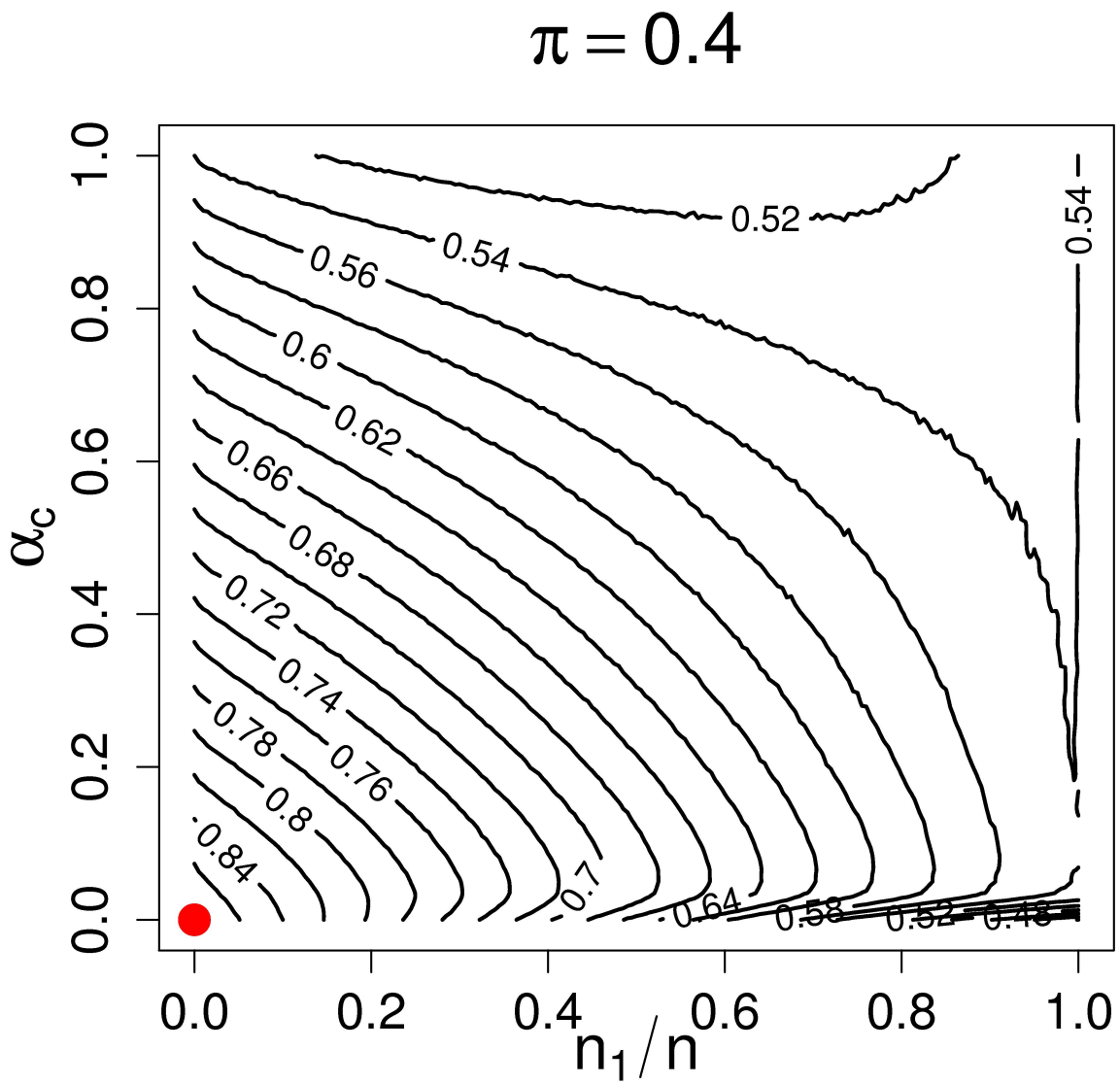
Utility $U_{\pi,C} = U_{\pi,P}$ for $\mathbf{G_S = 1}$



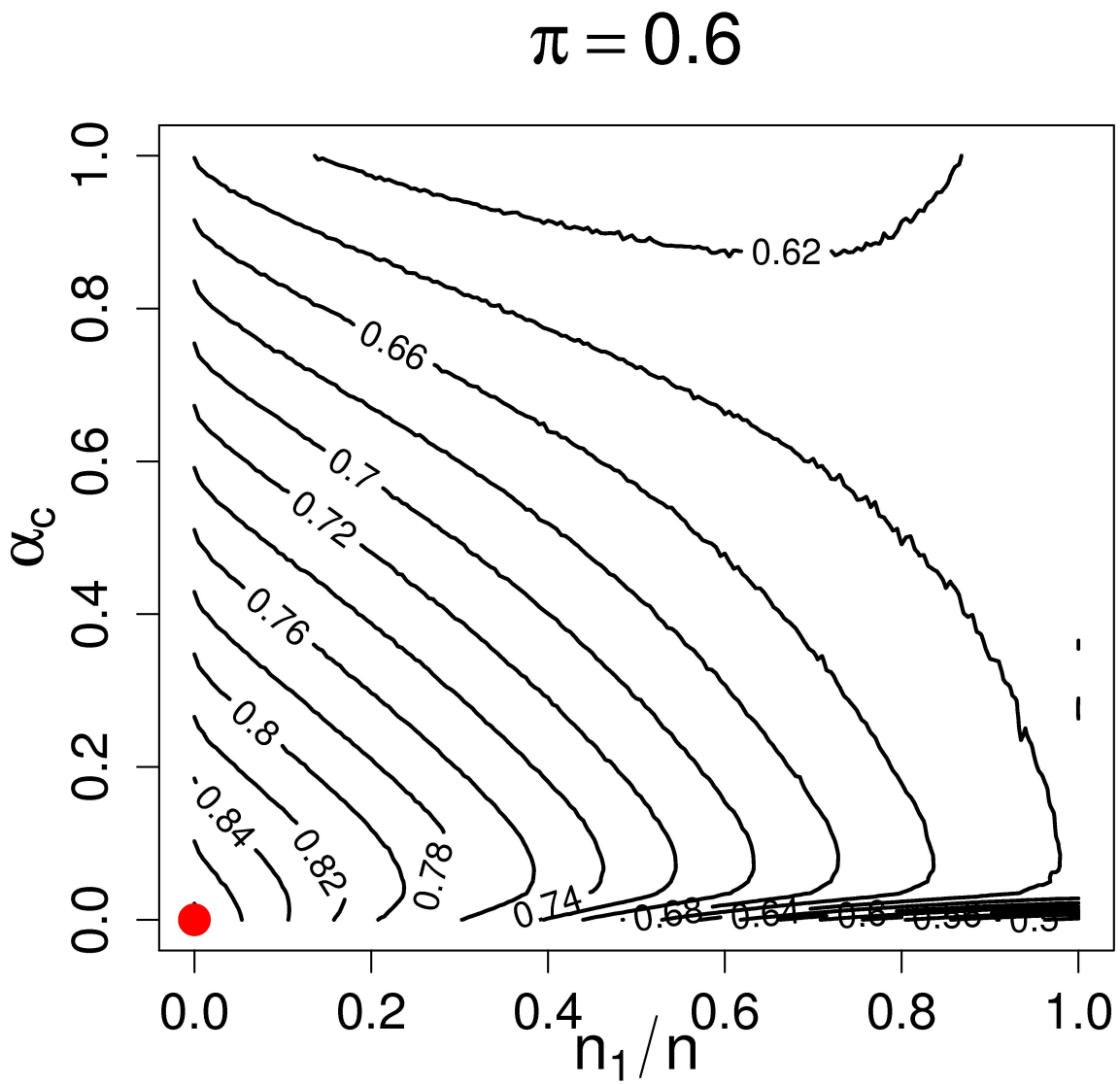
Utility $U_{\pi,C} = U_{\pi,P}$ for $\mathbf{G_S = 1}$



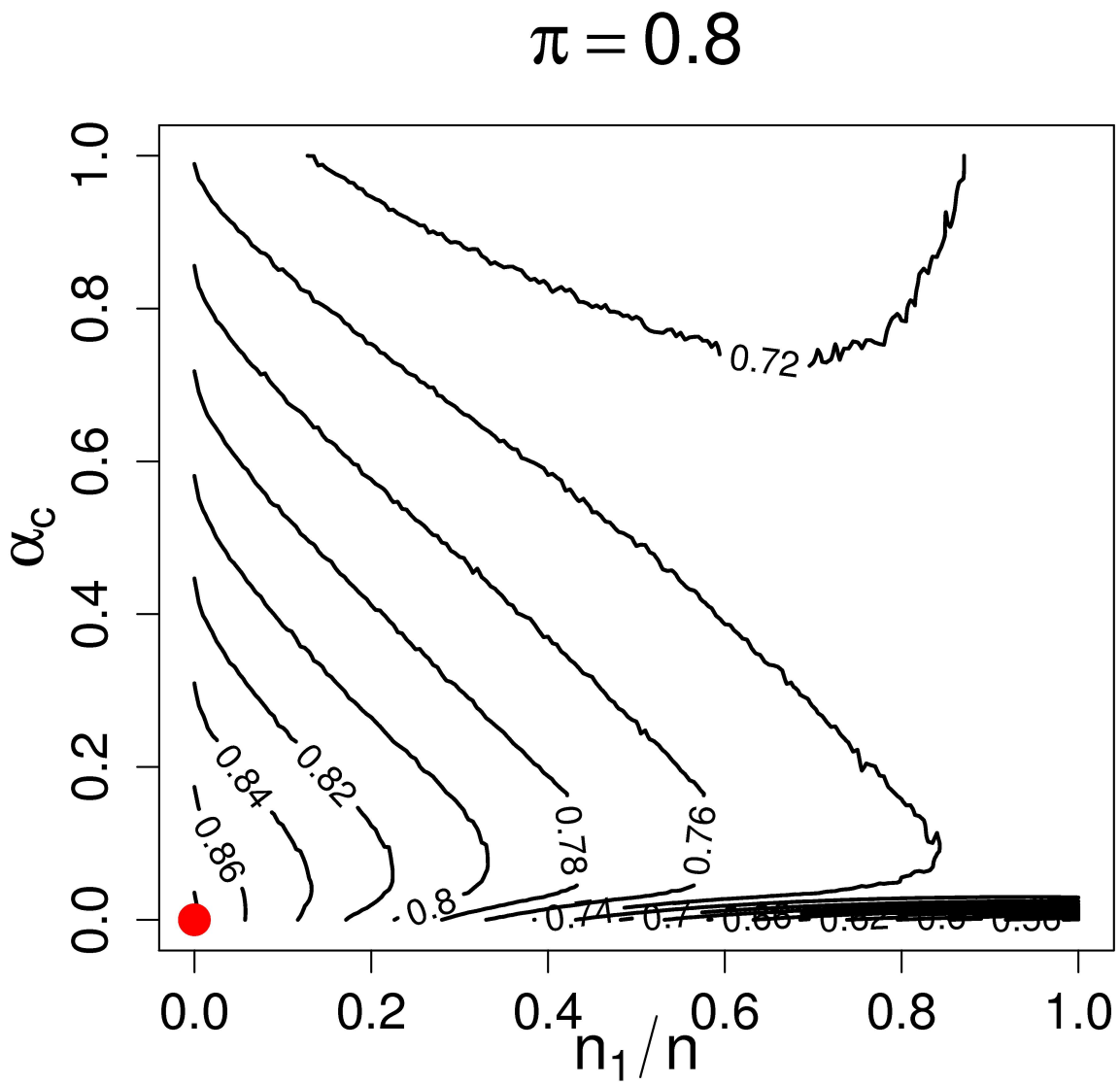
Utility $U_{\pi,C} = U_{\pi,P}$ for $\mathbf{G_S = 1}$



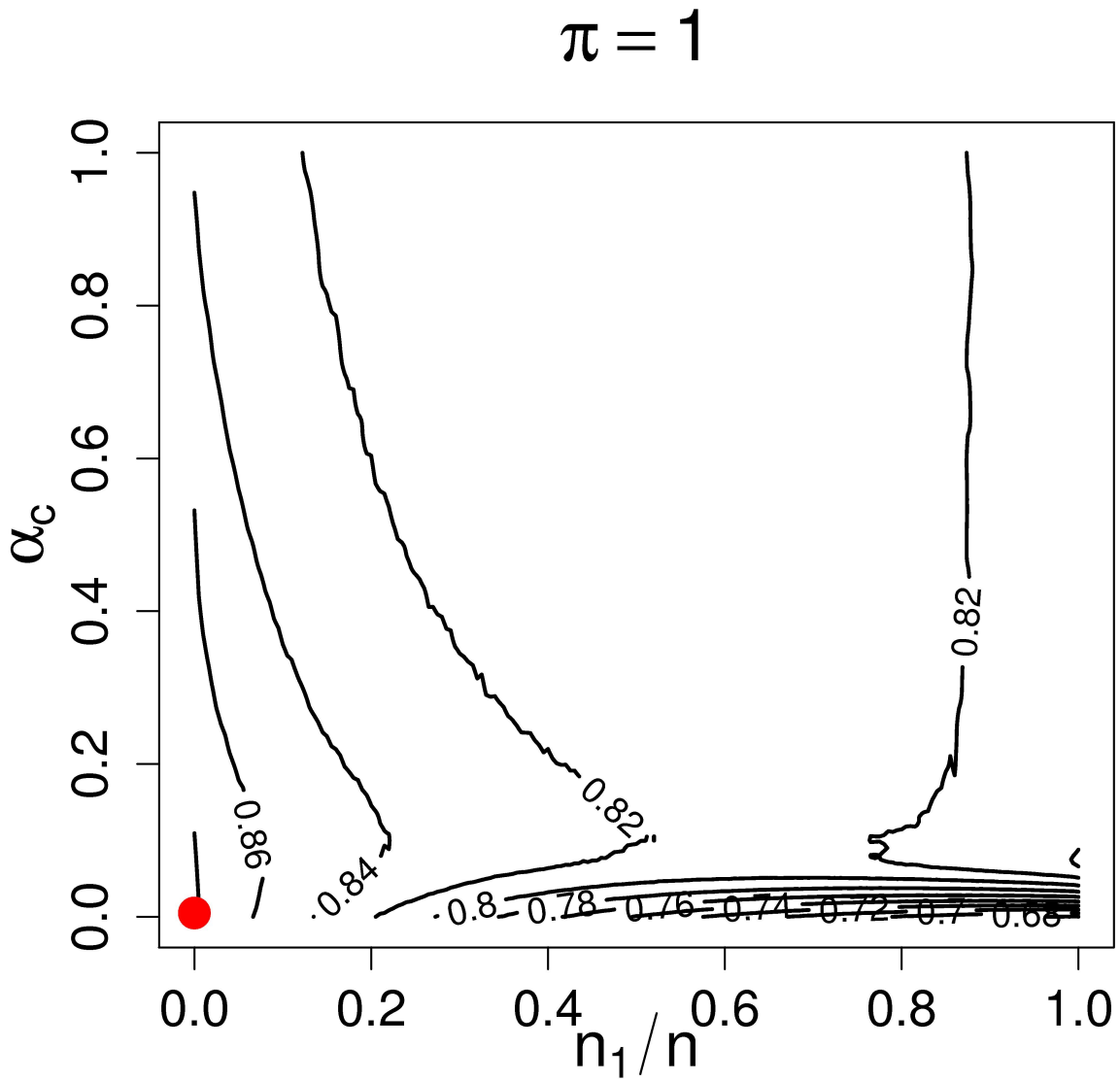
Utility $U_{\pi,C} = U_{\pi,P}$ for $\mathbf{G_S = 1}$



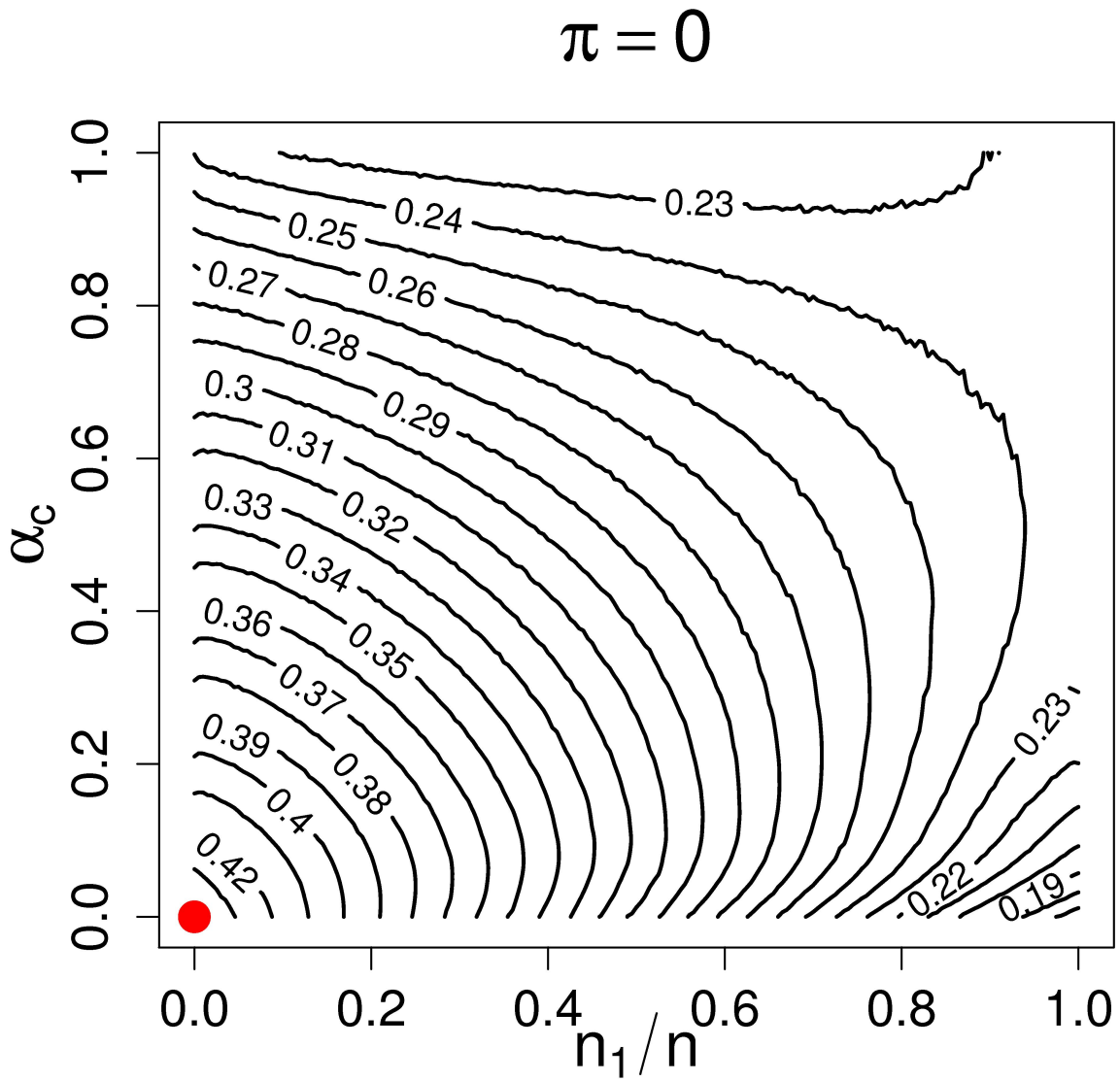
Utility $U_{\pi,C} = U_{\pi,P}$ for $\mathbf{G_S = 1}$



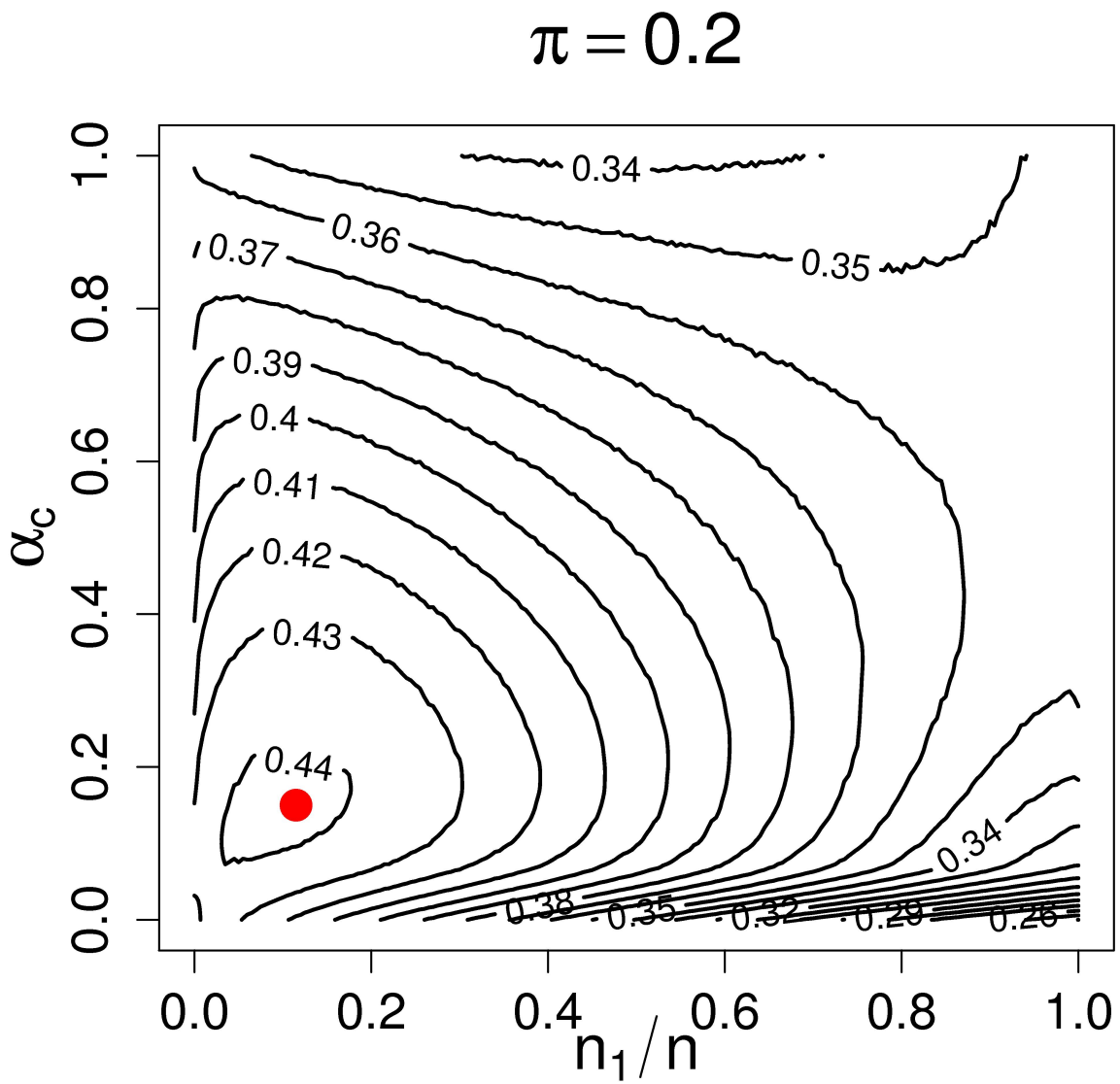
Utility $U_{\pi,C} = U_{\pi,P}$ for $\mathbf{G_S = 1}$



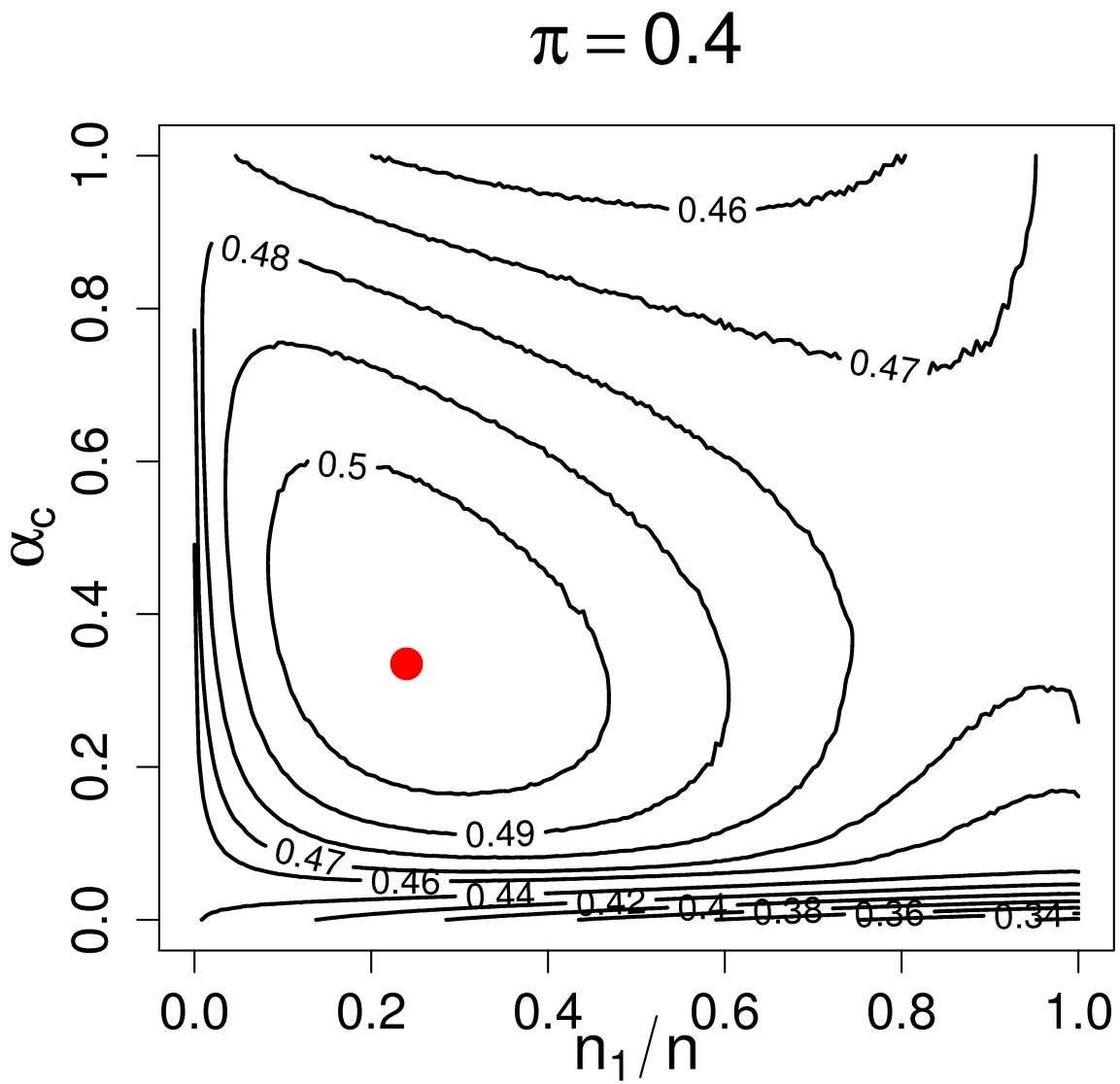
Utility $U_{\pi,C}$ for $\mathbf{G_S = .5}$



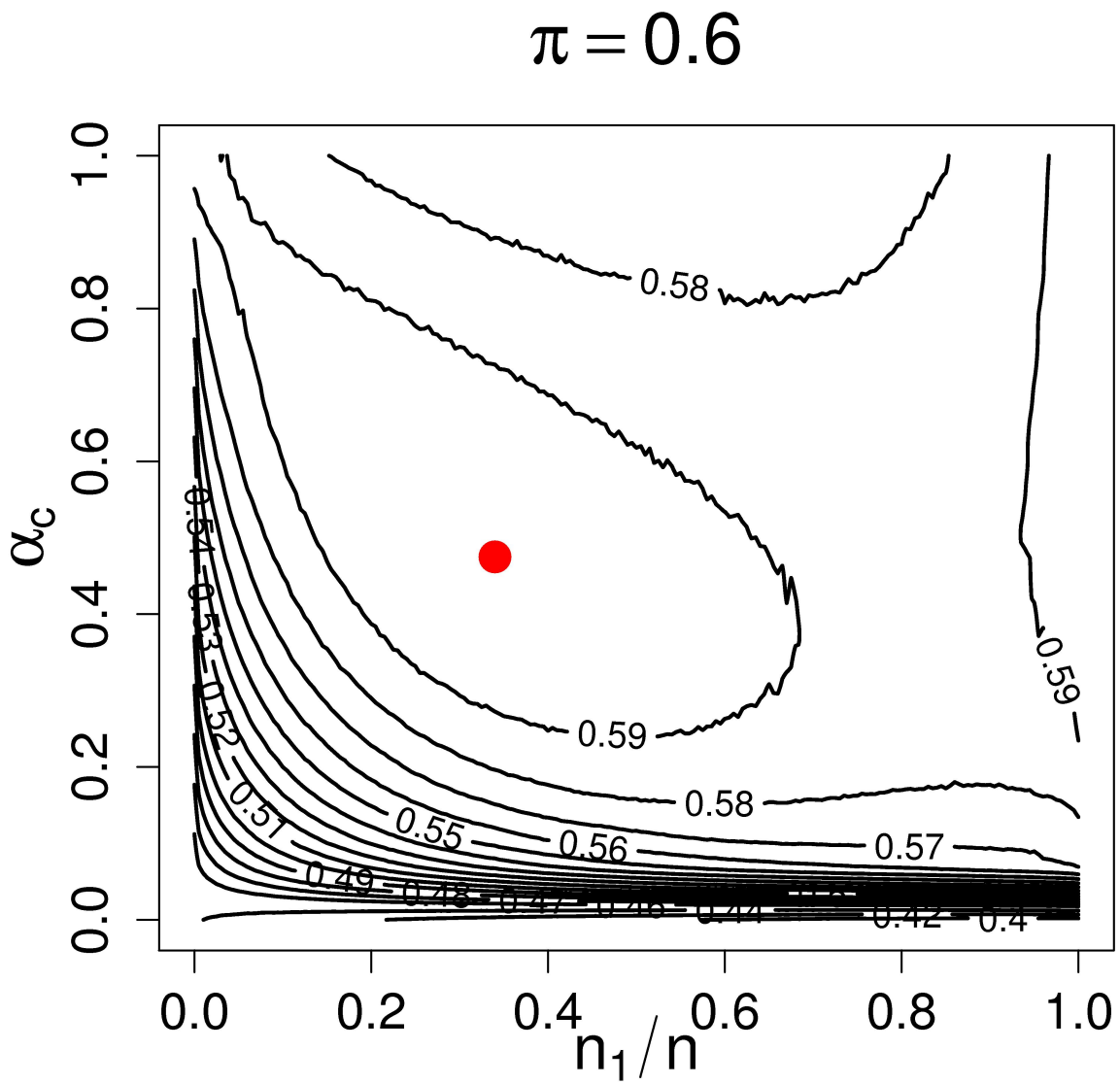
Utility $U_{\pi,C}$ for $\mathbf{G_S = .5}$



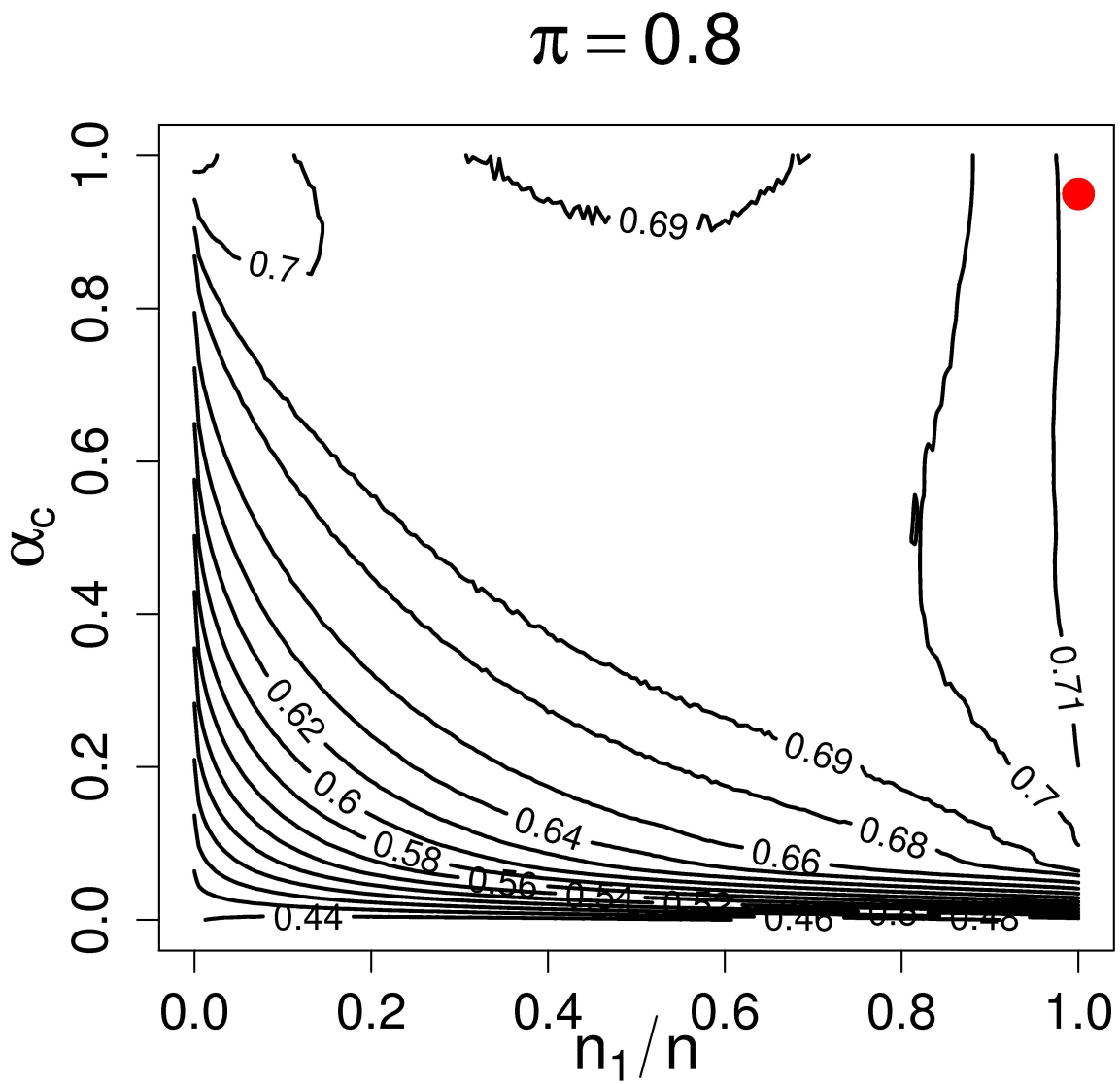
Utility $U_{\pi,C}$ for $\mathbf{G_S = .5}$



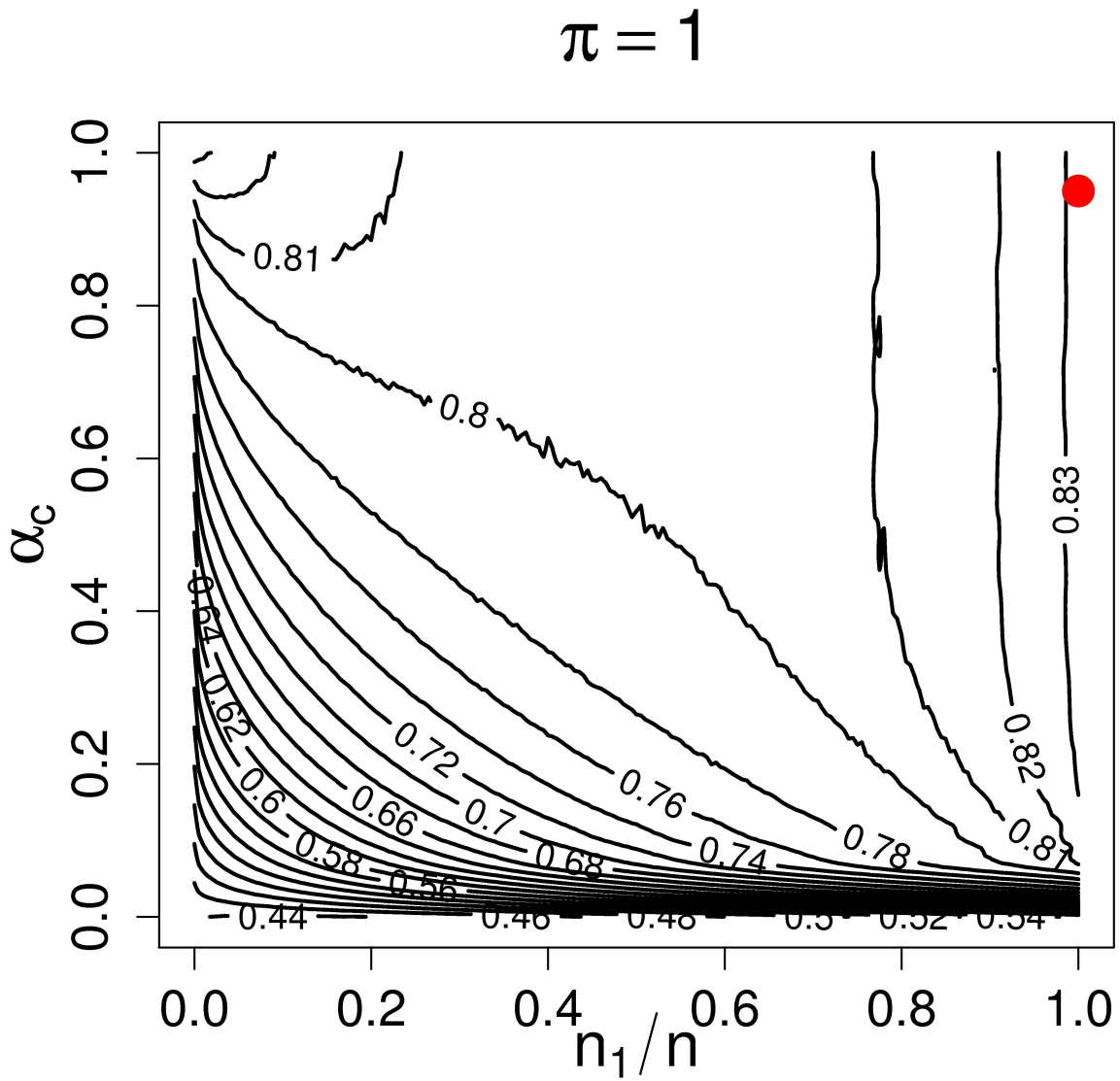
Utility $U_{\pi,C}$ for $G_S = .5$



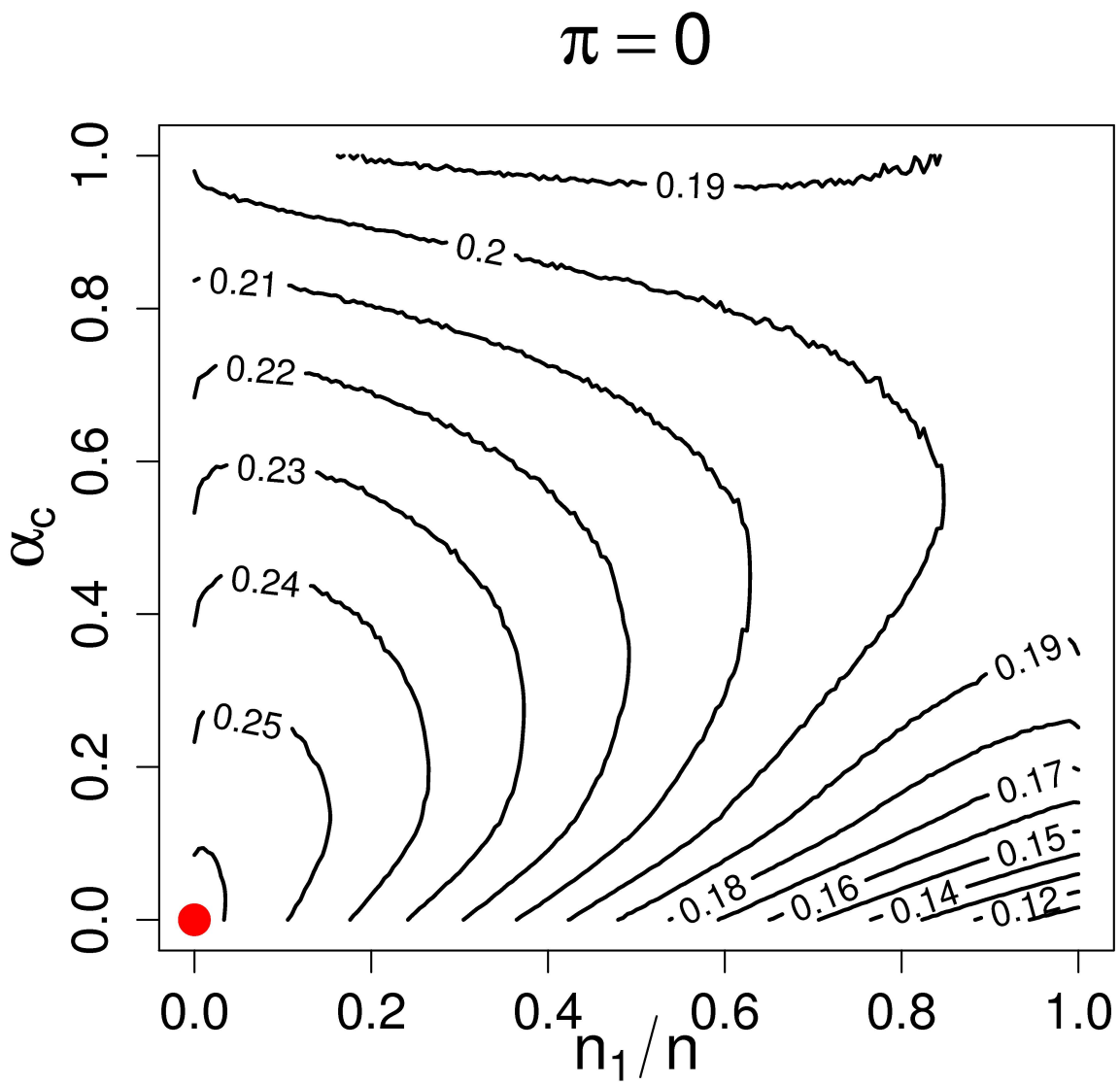
Utility $U_{\pi,C}$ for $\mathbf{G_S} = .5$



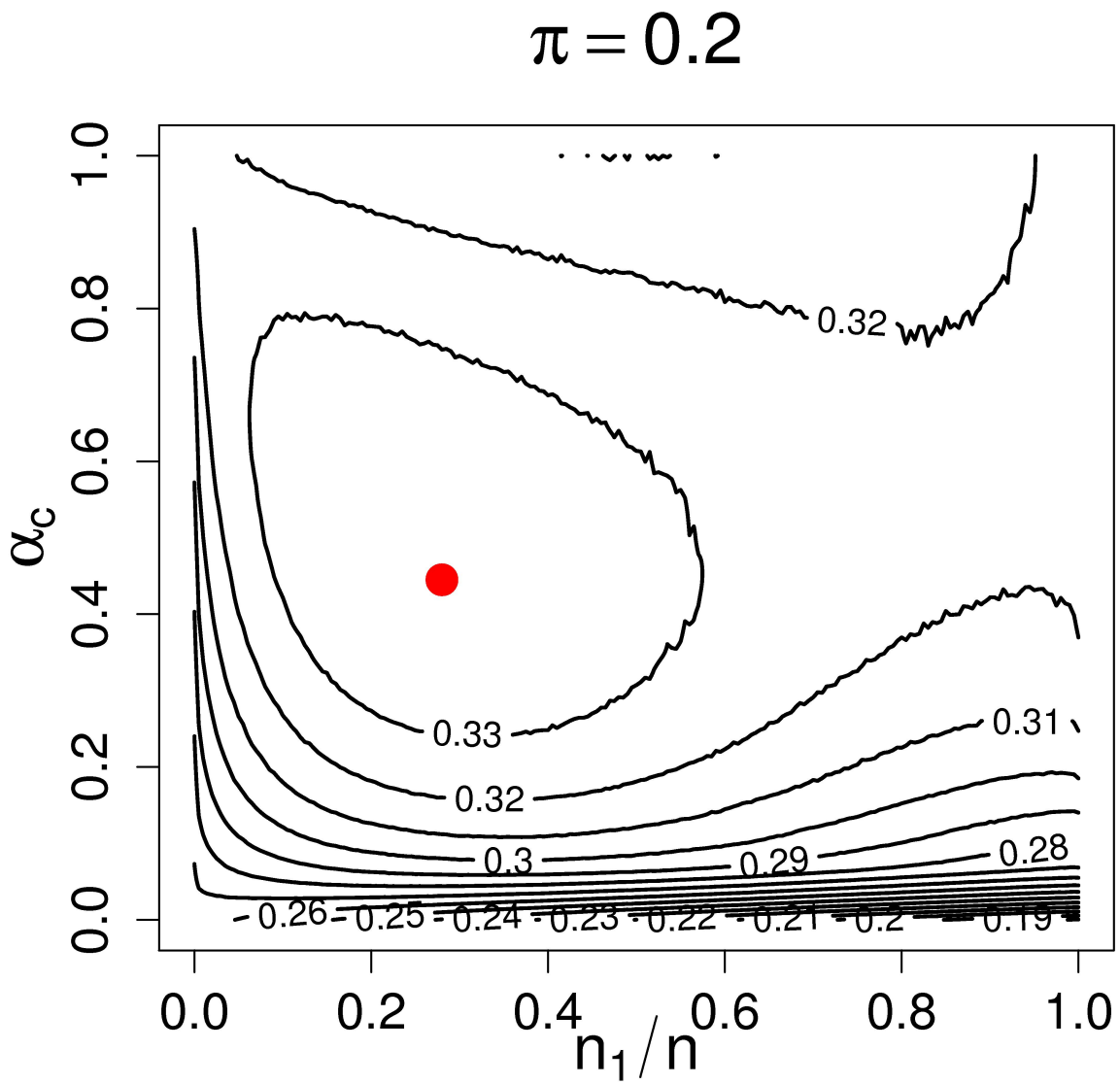
Utility $U_{\pi,C}$ for $G_S = .5$



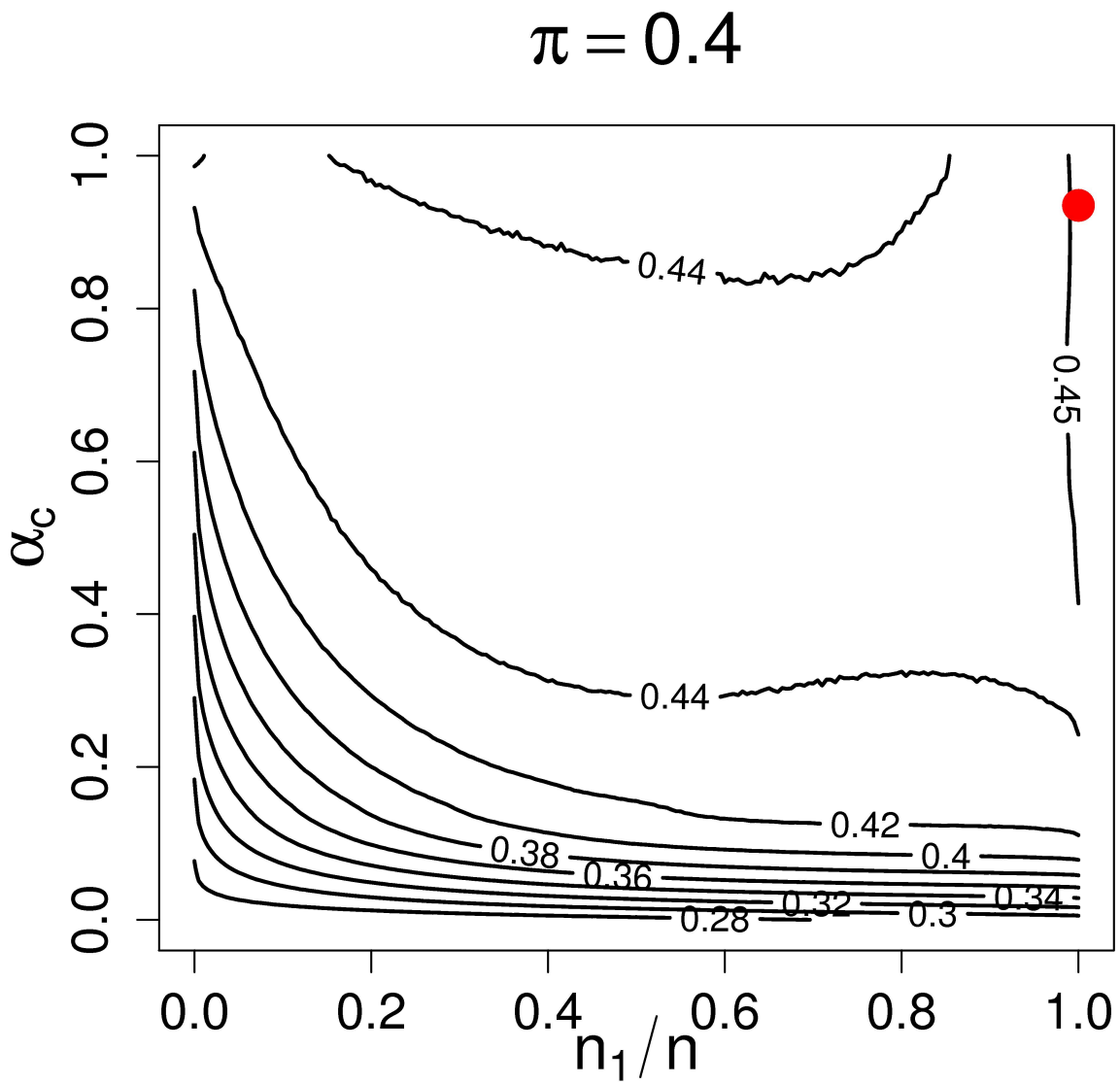
Utility $U_{\pi,C}$ for $\mathbf{G_S = .3}$



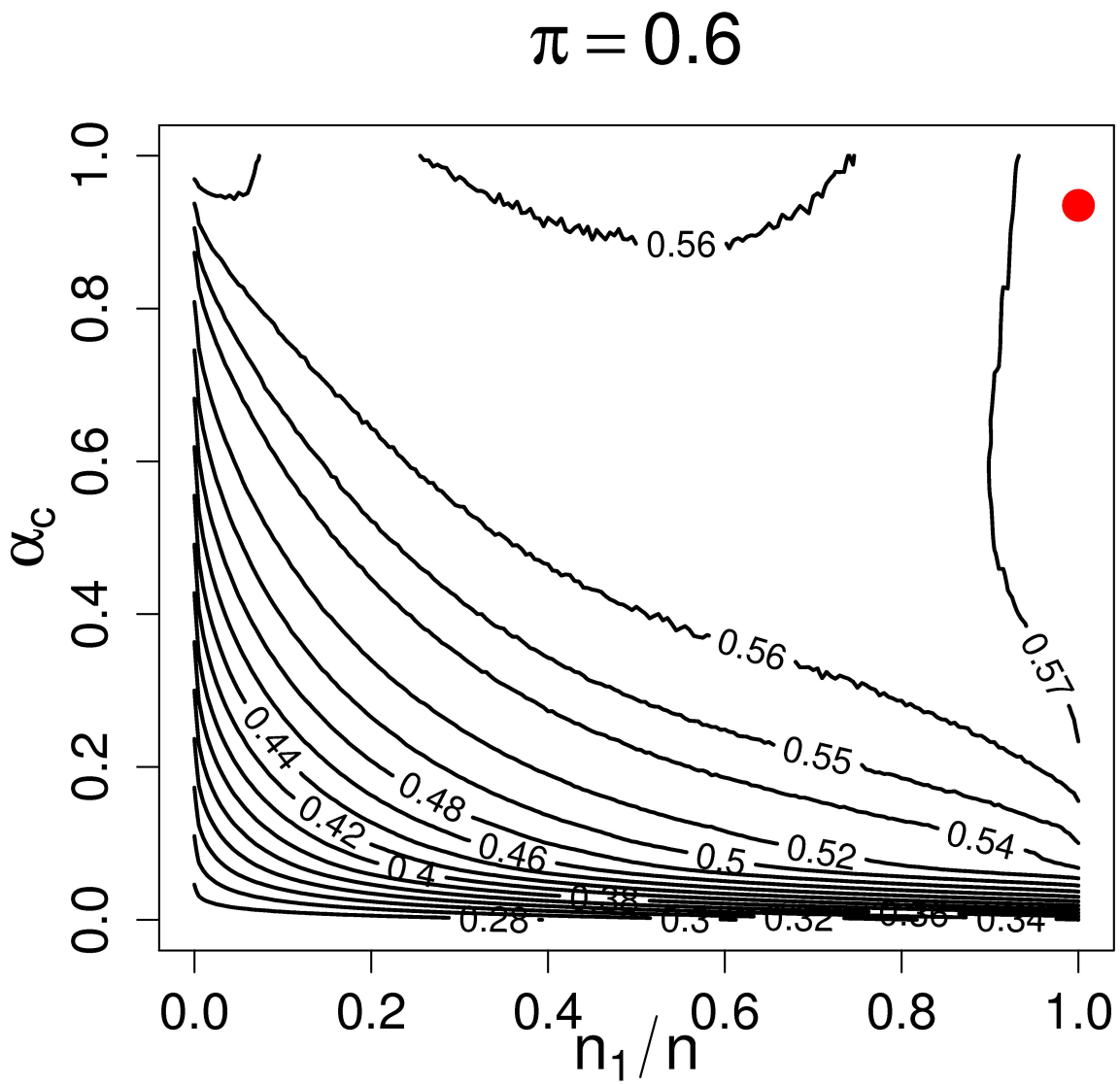
Utility $U_{\pi,C}$ for $\mathbf{G_S = .3}$



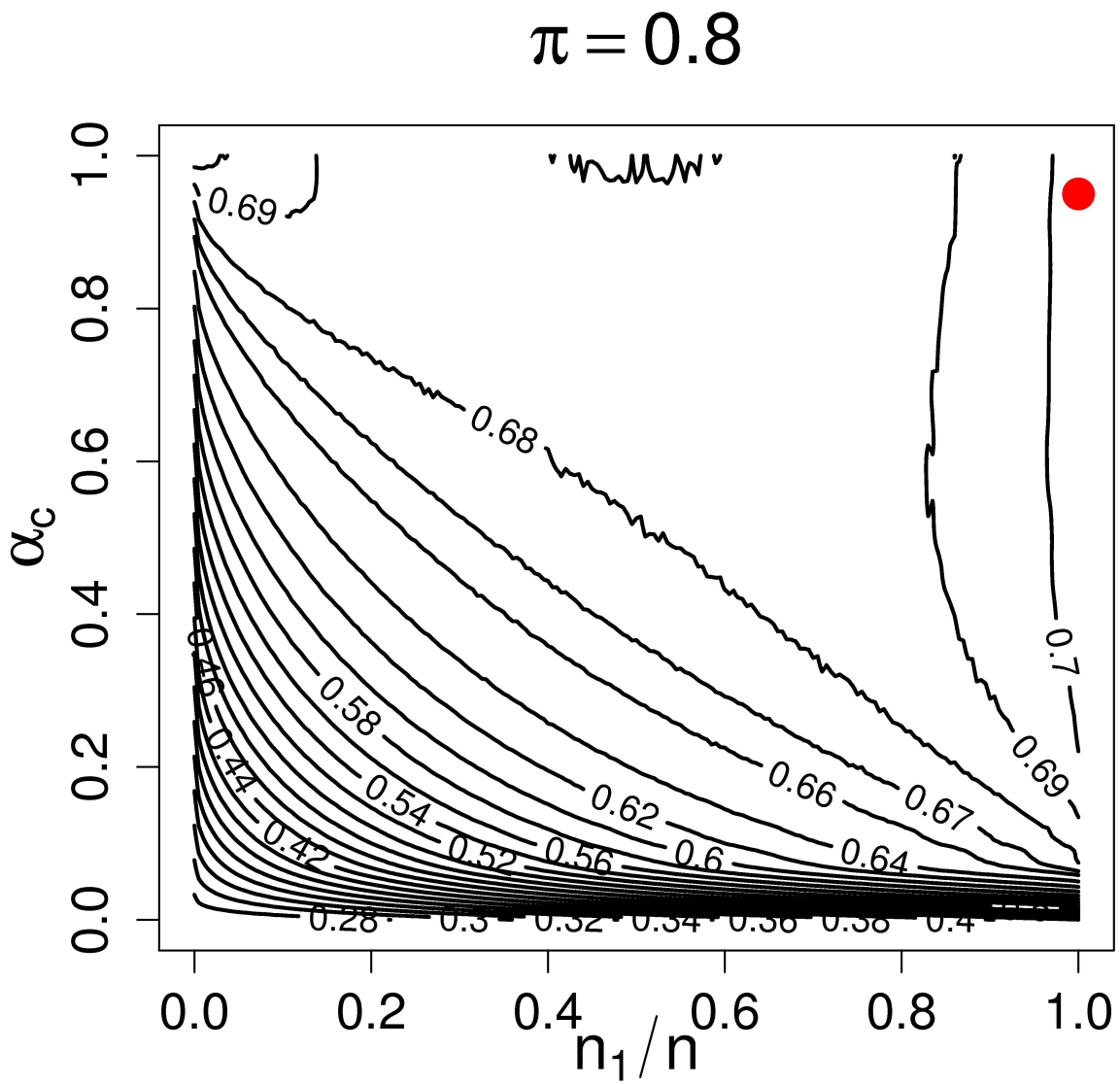
Utility $U_{\pi,C}$ for $\mathbf{G_S} = .3$



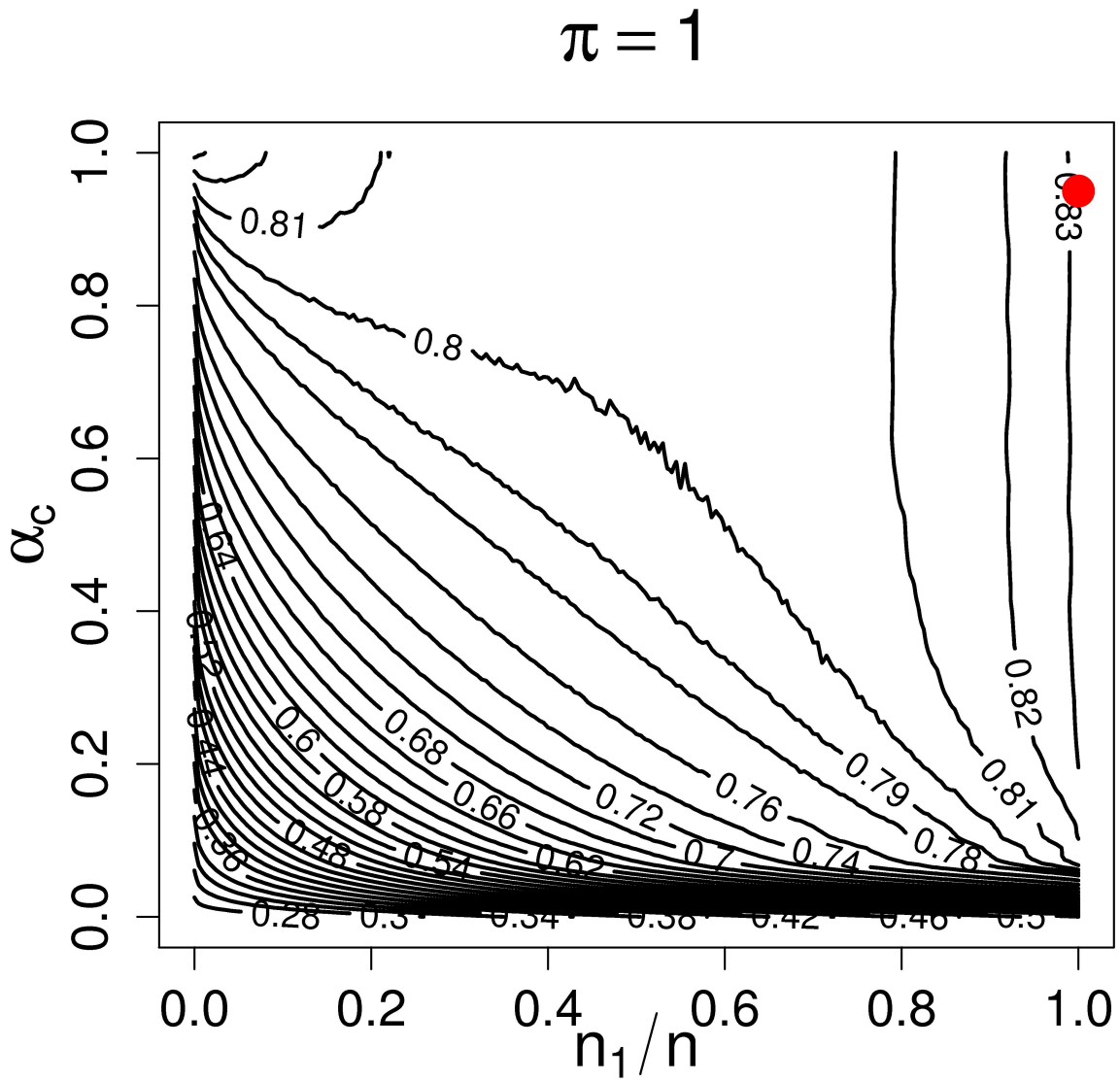
Utility $U_{\pi,C}$ for $\mathbf{G_S = .3}$



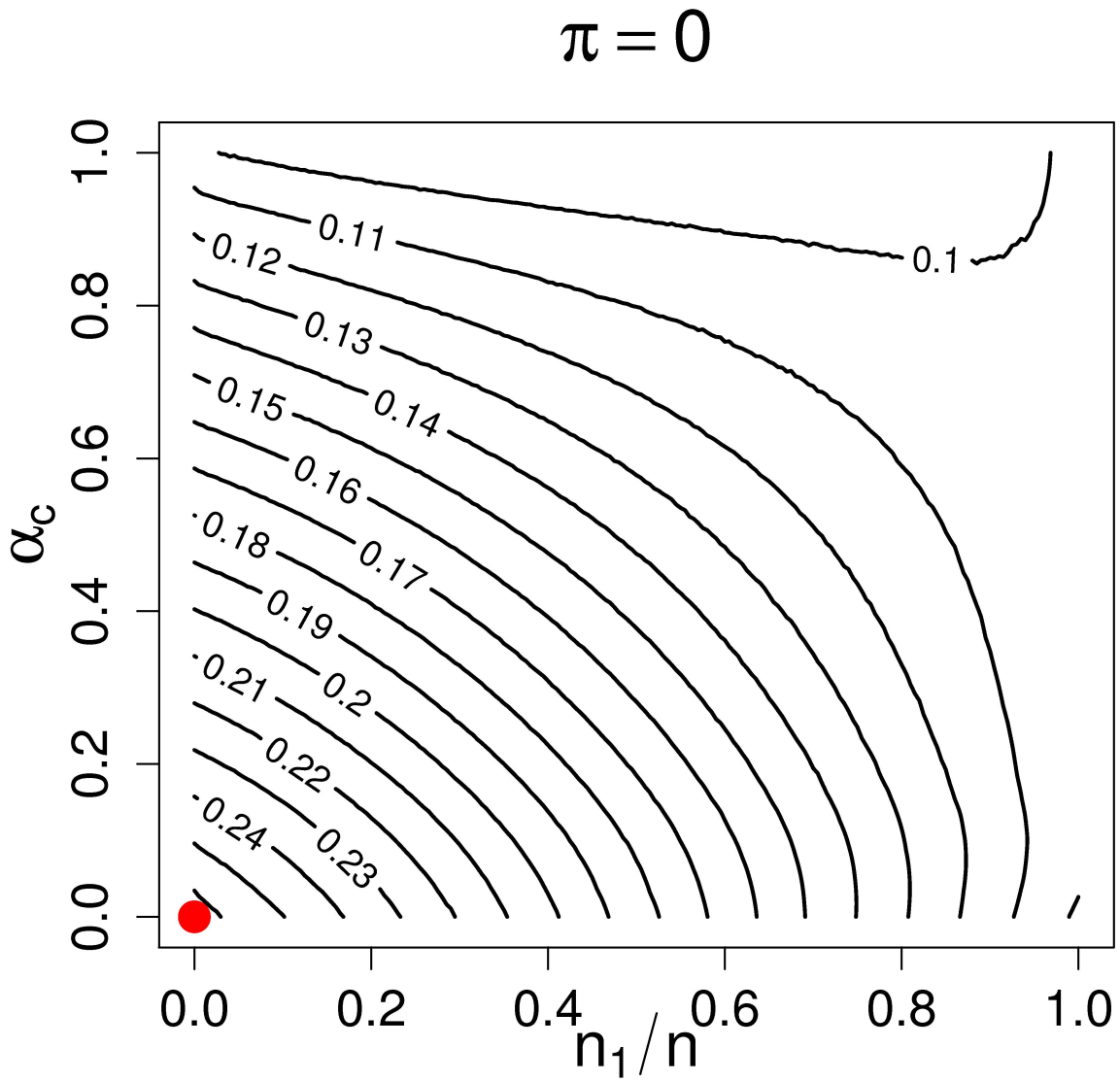
Utility $U_{\pi,C}$ for $\mathbf{G_S = .3}$



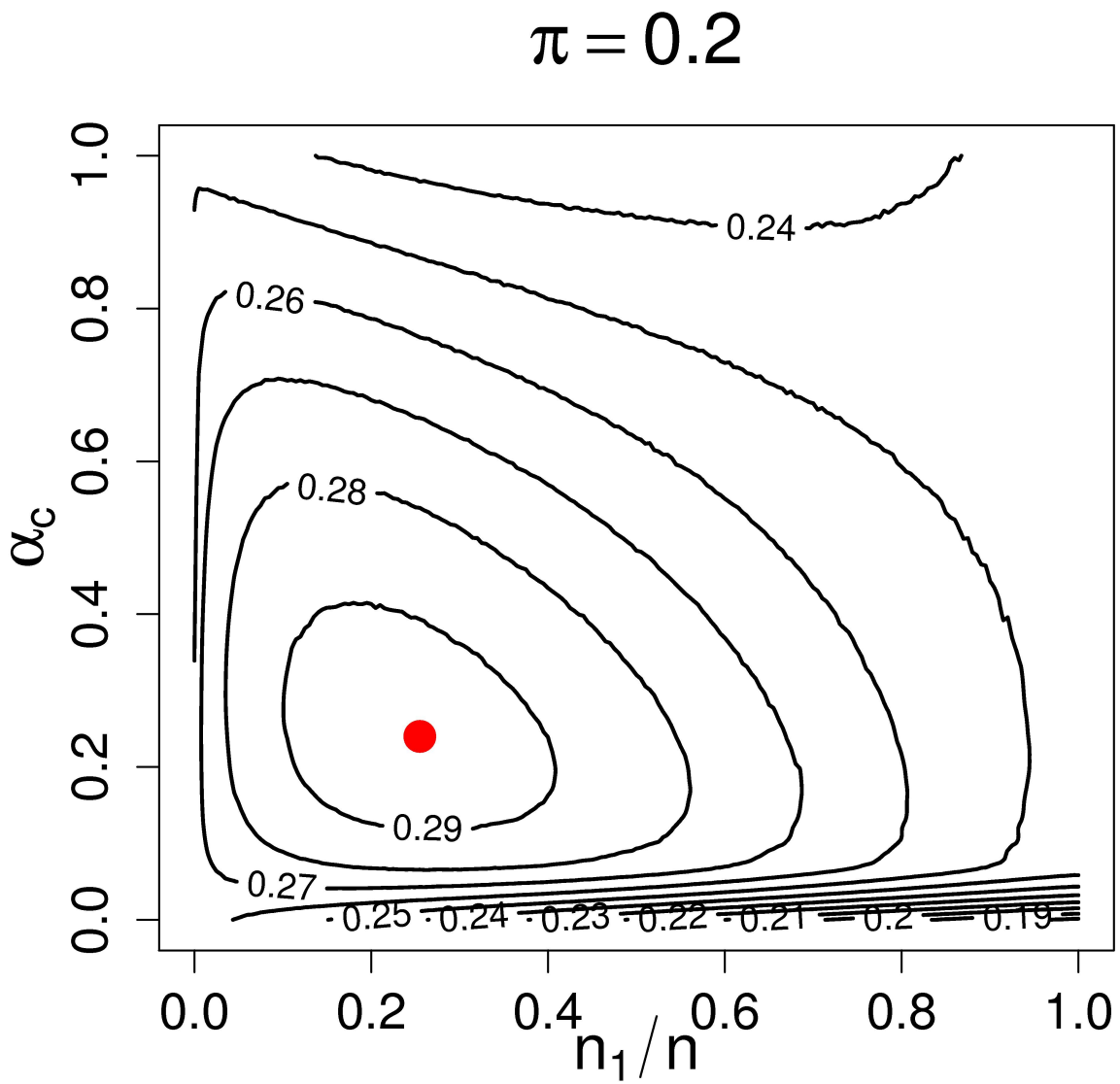
Utility $U_{\pi,C}$ for $\mathbf{G_S} = .3$



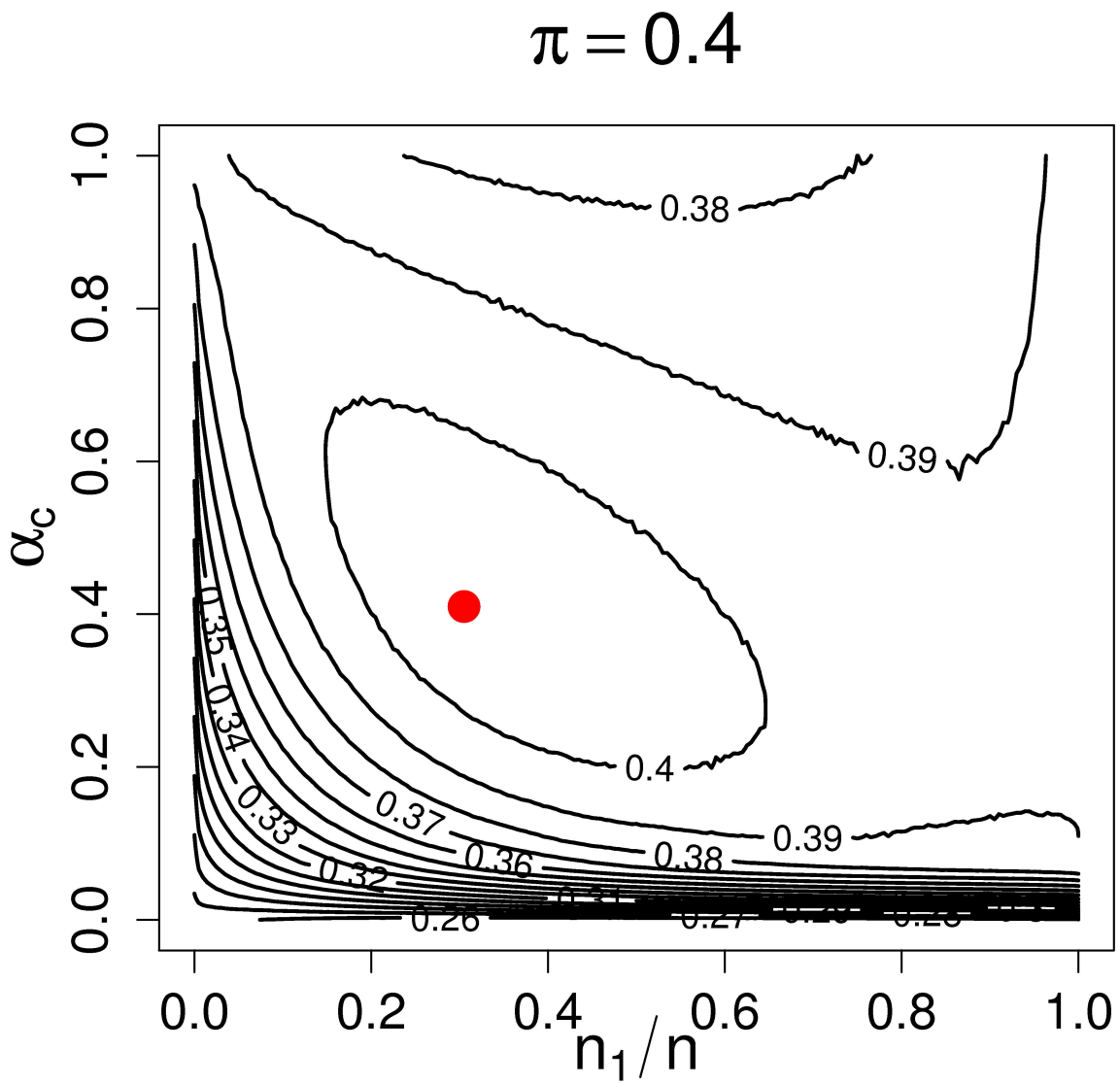
Utility $U_{\pi,P}$ for $\mathbf{G_S = .3}$



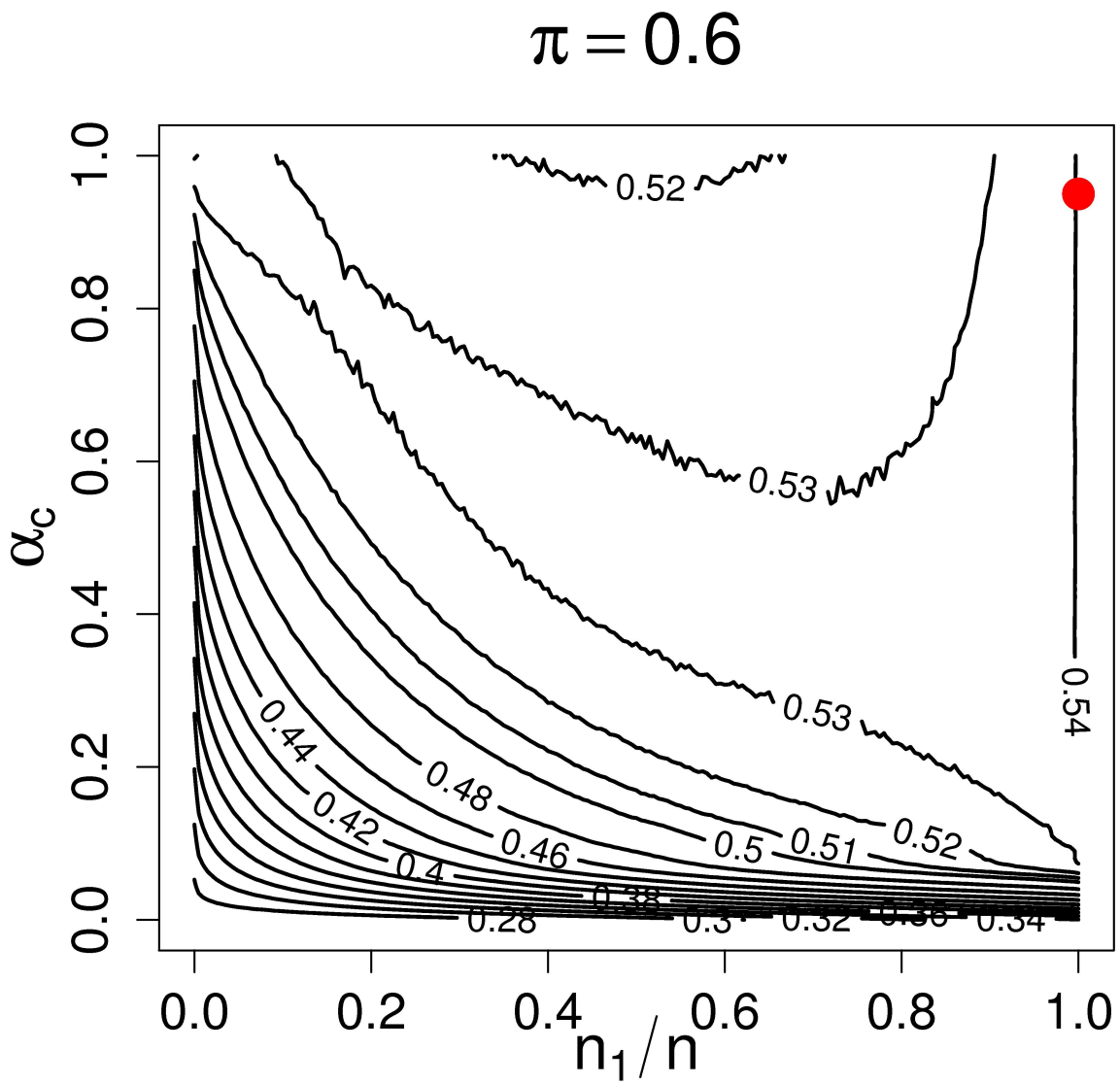
Utility $U_{\pi,P}$ for $\mathbf{G_S = .3}$



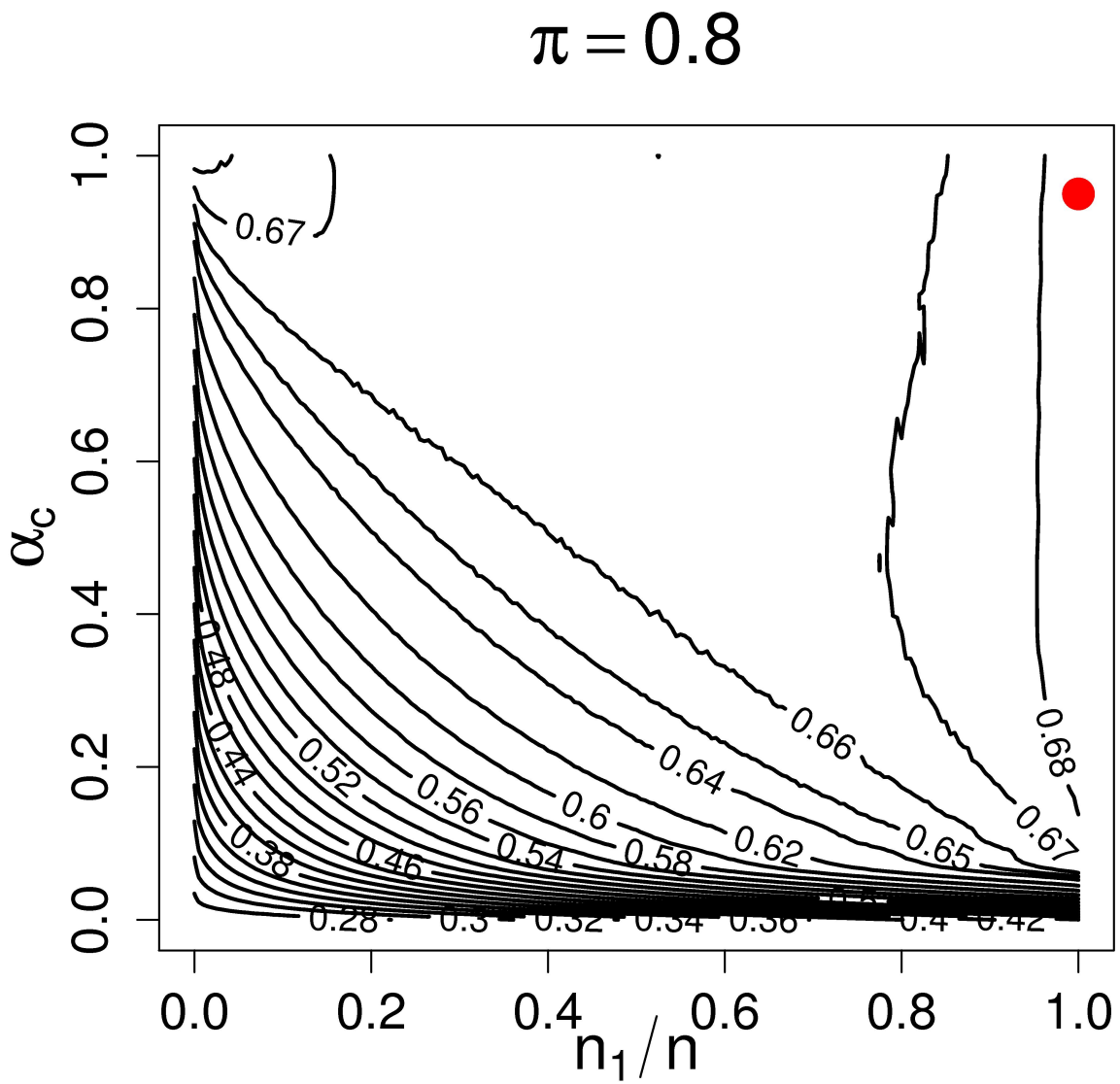
Utility $U_{\pi,P}$ for $\mathbf{G_S = .3}$



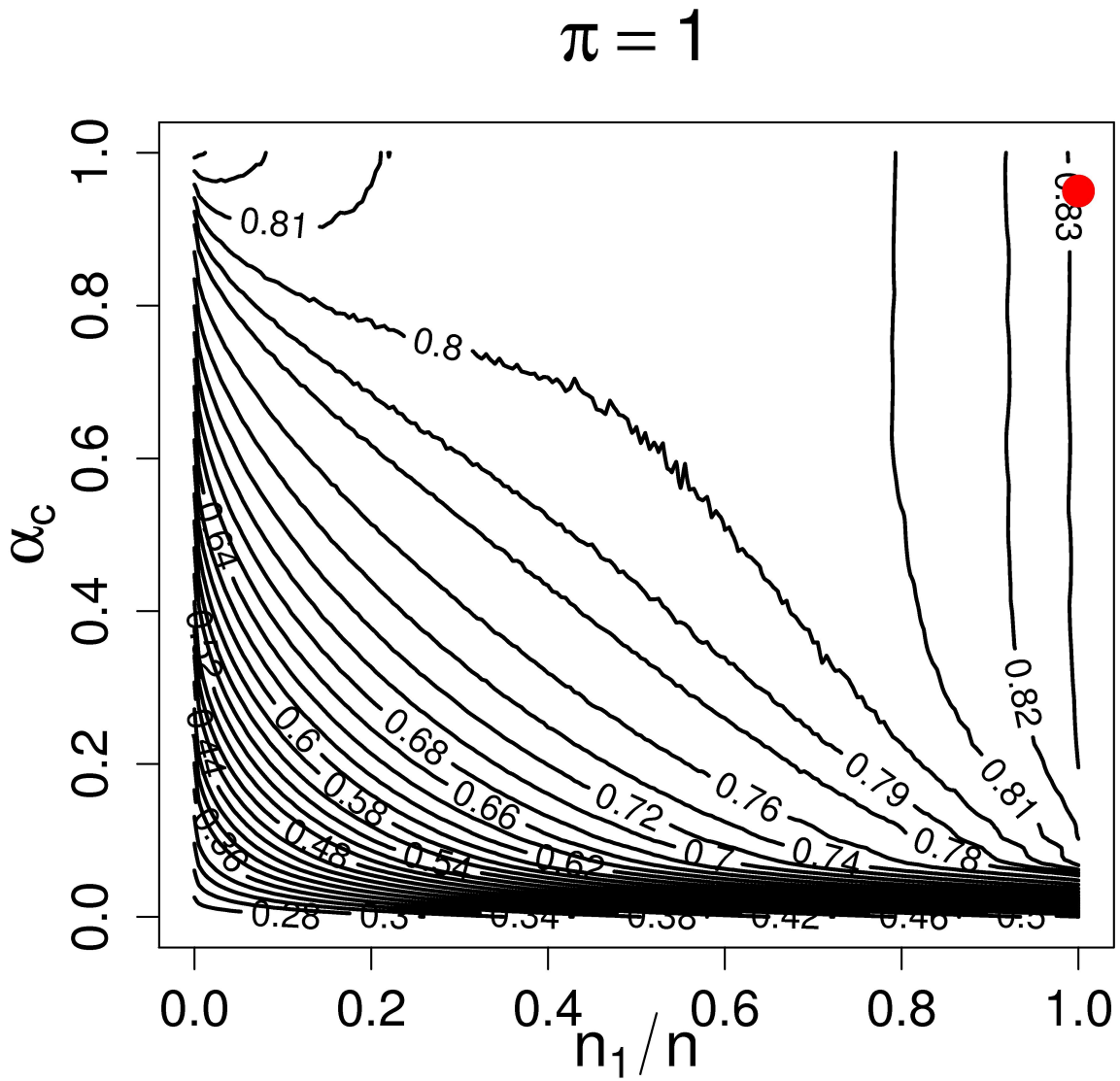
Utility $U_{\pi,P}$ for $\mathbf{G_S = .3}$



Utility $U_{\pi,P}$ for $\mathbf{G_S = .3}$

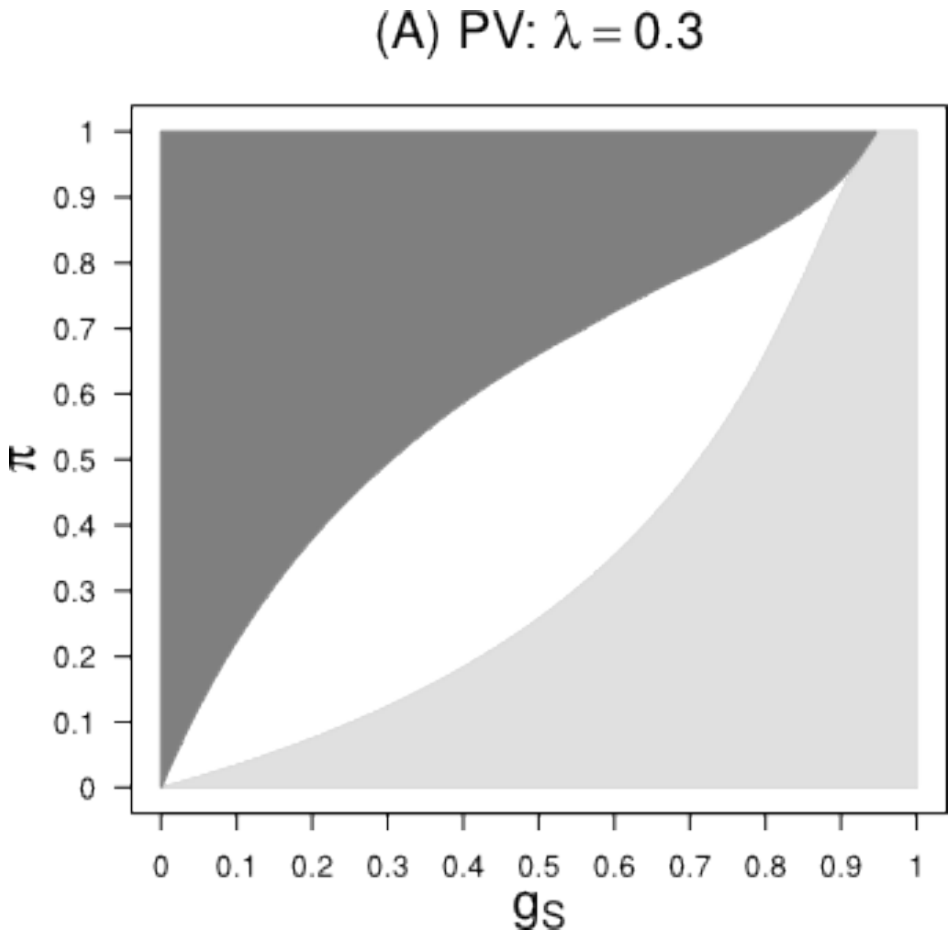


Utility $U_{\pi,P}$ for $\mathbf{G_S = .3}$



Which Design is best?

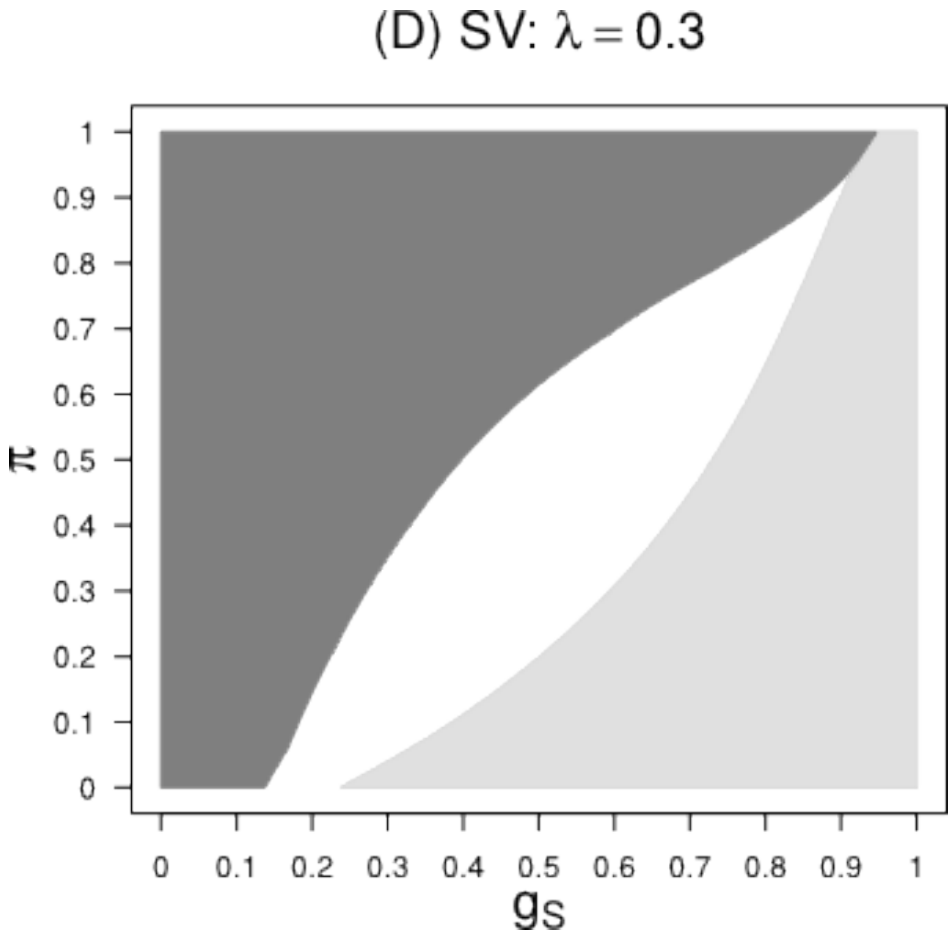
Public View Utility Function



■ Stratification □ Adaptive ■ Enrichment

Which Design is best?

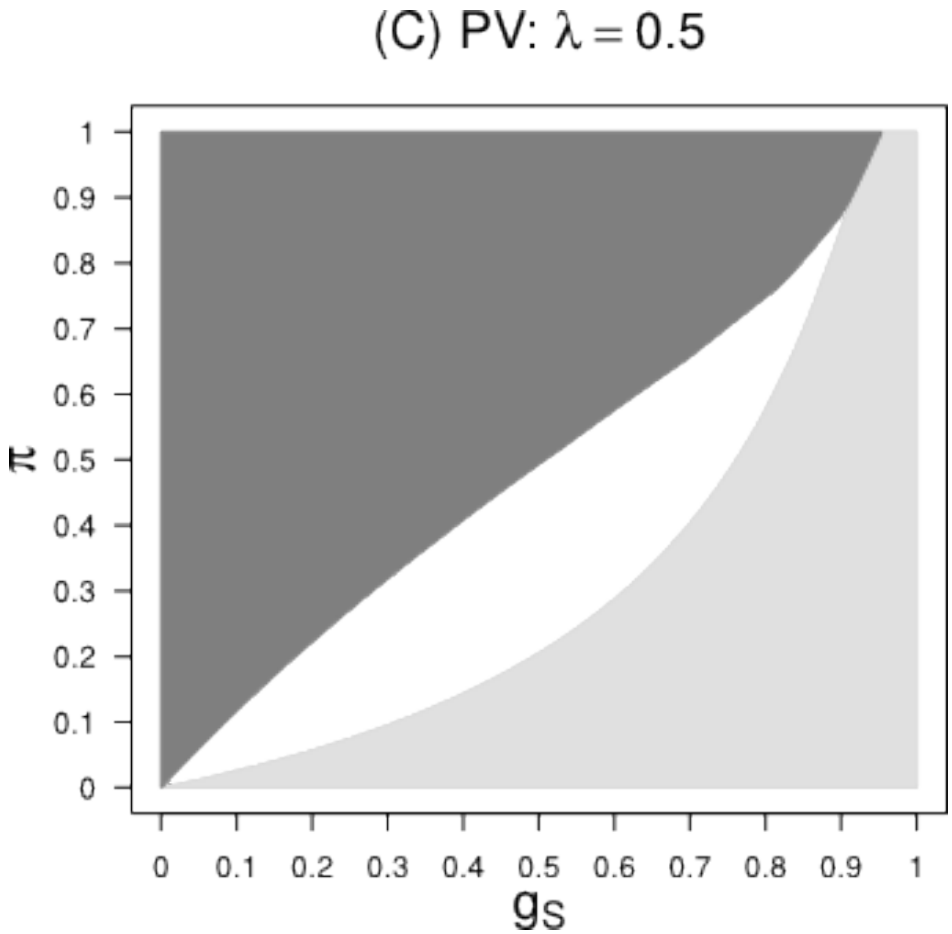
Sponsor View Utility Function



■ Stratification □ Adaptive ■ Enrichment

Which Design is best?

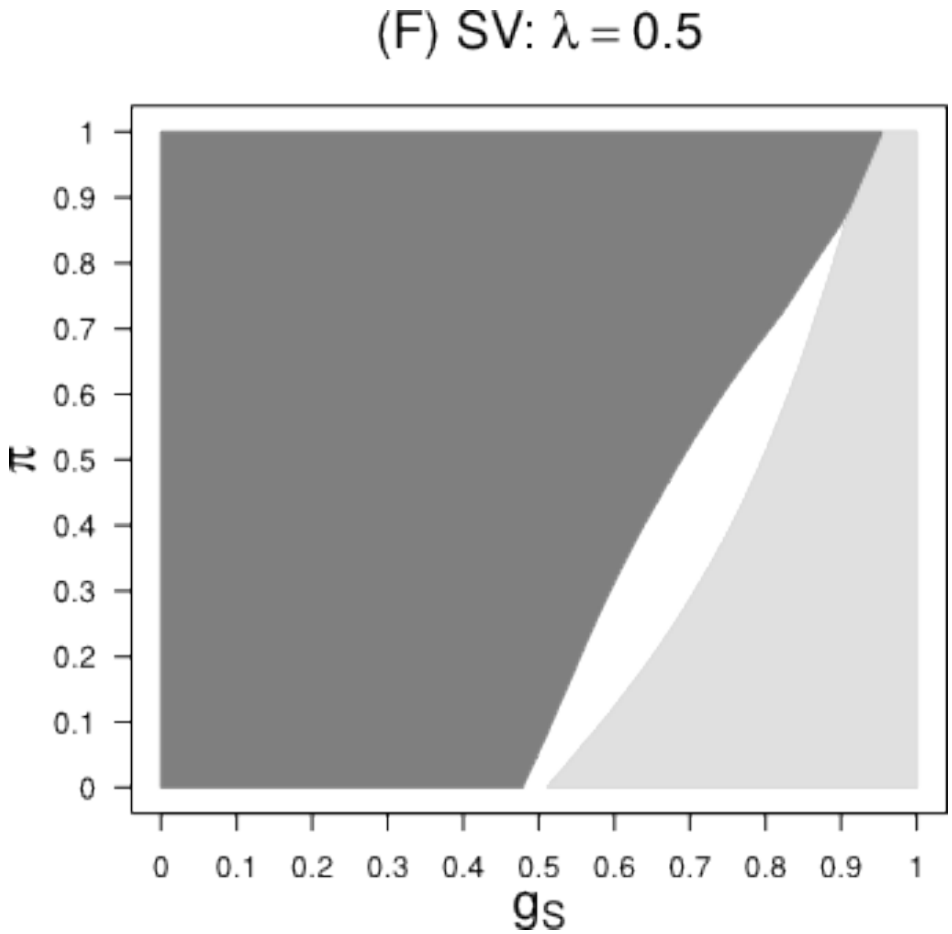
Public View Utility Function



■ Stratification □ Adaptive ■ Enrichment

Which Design is best?

Sponsor View Utility Function



■ Stratification □ Adaptive ■ Enrichment

Summary and Limitations

- For enrichment designs investigating the power to reject any null hypothesis may not be sufficient.
- The optimized design depends critically on the prior.
- We investigated a very simple adaptation rule, that depended on the effect size of the complement of S only.
- Designs can be extended to optimize the conditional expected utility, taking into account also the effect size S.
- The loss resulting from false positive rejections of H_S and H_F is accounted for only through the multiple testing procedure but not included in the utility function.

References

- Bauer P, Kieser M. Combining different phases in the development of medical treatments within a single trial. *Statistics in Medicine* 1999; **18**:1833–1848.
- Brannath, Zuber, Branson, Bretz, Gallo, Posch, Racine-Poon. Confirmatory adaptive designs with Bayesian decision tools for a targeted therapy in oncology *Statistics in Medicine* 2009; **28**:1445-63.
- Bretz F, König F, Brannath W, Glimm E, Posch M. Adaptive Designs for Confirmatory Clinical Trials. *Statistics in Medicine* 2010; **28**:1181-217.
- Hommel G. Adaptive modifications of hypotheses after an interim analysis. *Biometrical Journal* 2001; **43**:581–589.
- Jenkins M, Stone A, Jennison C. An adaptive seamless phase II/III design for oncology trials with subpopulation selection using correlated survival endpoints. *Pharmaceutical Statistics* 2010; Online preview.
- Lehmacher W, Wassmer G. Adaptive sample size calcualtions in group sequential trials. *Biometrics* 1999; **55**: 1286–1290.
- Posch M, König F, Branson M, Brannath W, Dunger-Baldauf C, Bauer P. Testing and estimation in flexible group sequential designs with adaptive treatment selection. *Statistics in Medicine* 2005; **24**:3697-3714.
- Wang SJ, Hung J., ONeil R. Adaptive patient enrichment designs in therapeutic trials. *Biometrical Journal* 2009; **51**:358–374.