Non-Linearity of Scorecard Log-Odds

Ross McDonald, Keith Smith, Matthew Sturgess, Edward Huang
Retail Decision Science, Lloyds Banking Group

Edinburgh Credit Scoring Conference
26th August 2009

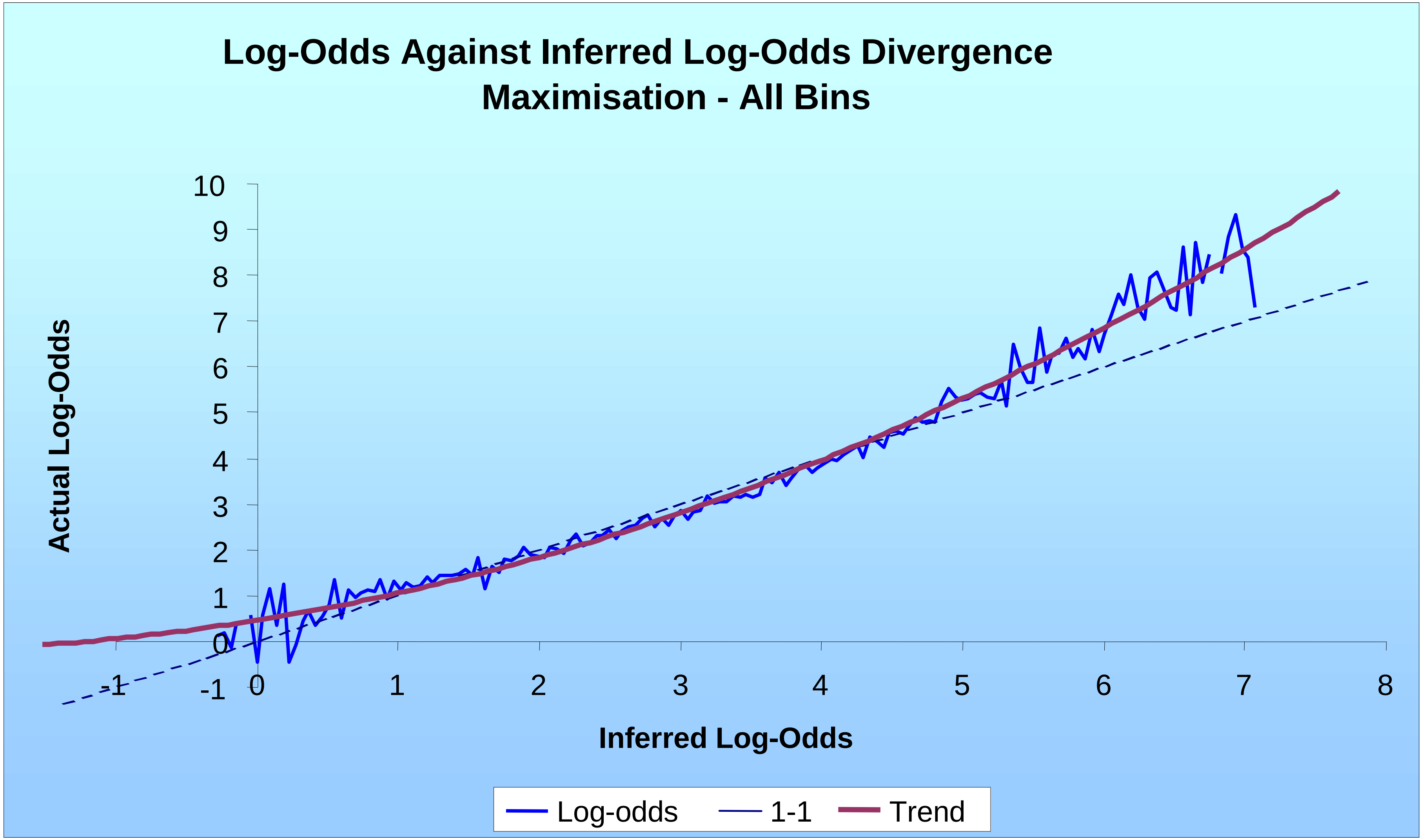
Lloyds Banking Group

- **Lloyds Banking Group was formed by the acquistion of HBOS by Lloyds TSB**
- **Three key relationship brands:**
- **The retail bank currently has**
  - **30 million customers**
  - **23 million current accounts**
  - **Number one positions in Mortgages and Savings**
- **Retail Decision Science employs ~150 specialists and modellers in Acquisition, Account Management, Collections and Recoveries, Basel and Impairment, Infrastructure, Portfolio Modelling, Fraud and Marketing.**

Outline

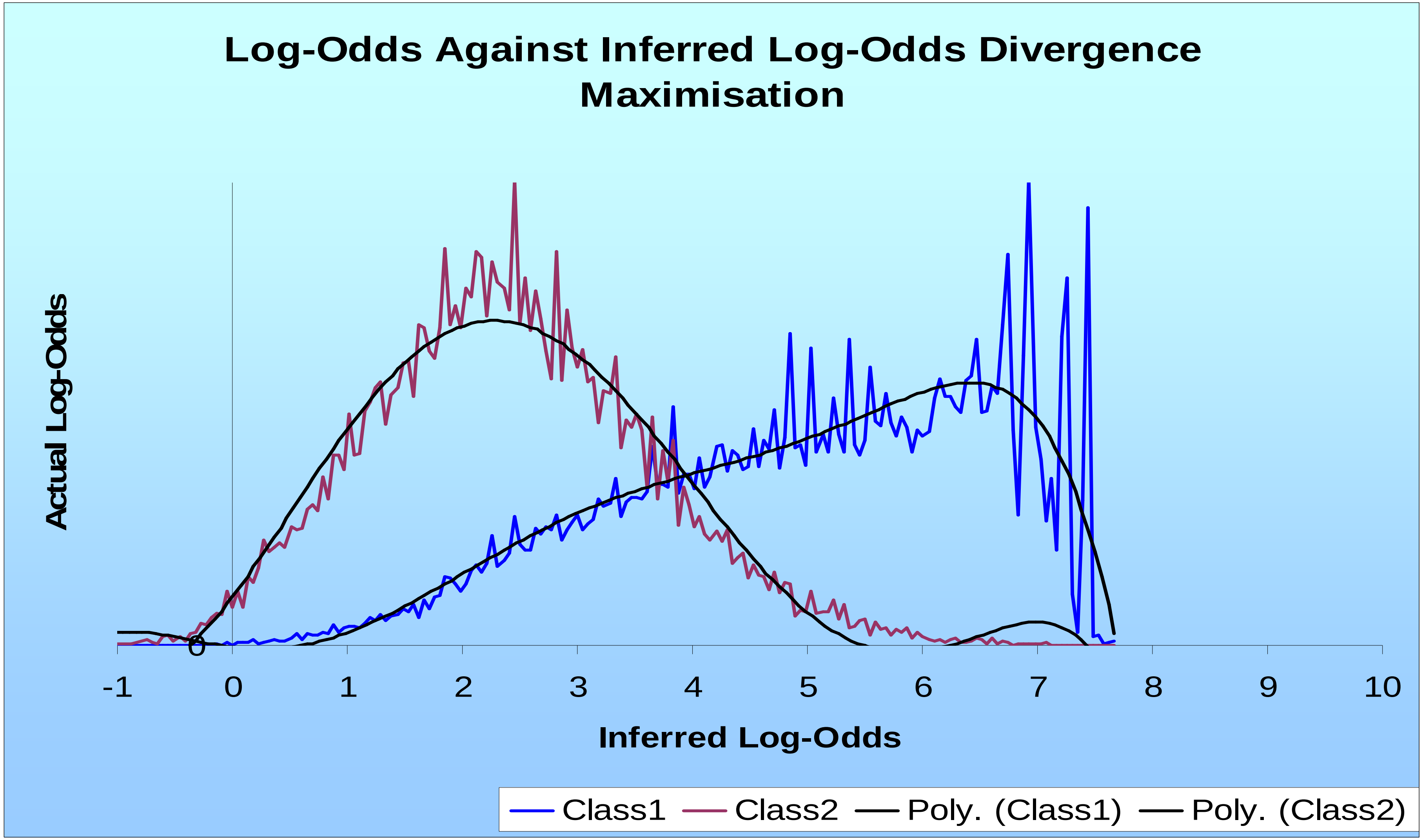
- **The Problem –** Non-linear score to log-odds relationship in scorecard models
- **Why does it matter? –** Uses of scorecard models, where do we need pr(event)?
- **Why does it happen? –** Uneven data distributions, correlation across bins
- **How can we fix it? –** Some remedies, pros and cons

The Problem



Log-Odds Against Inferred Log-Odds Divergence Maximisation - All Bins
Actual Log-Odds
10
9
8
7
6
5
4
3
2
1
0
-1
-1
0
1
2
3
4
5
6
7
8
Inferred Log-Odds
Log-odds
1-1
Trend

The Problem



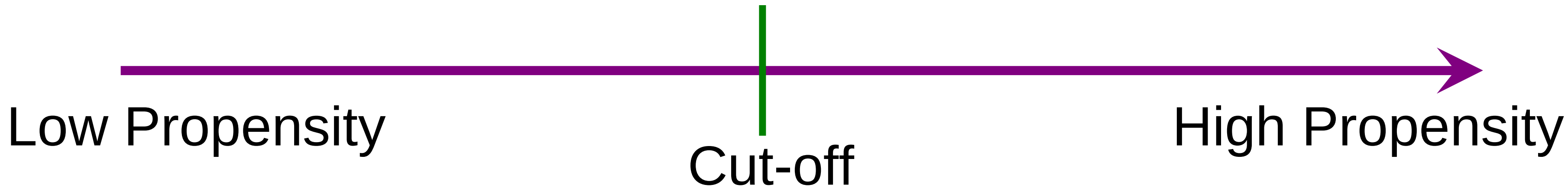
Log-Odds Against Inferred Log-Odds Divergence Maximisation
Actual Log-Odds
0
-1
0
1
2
3
4
5
6
7
8
9
10
Inferred Log-Odds
Class1
Class2
Poly. (Class1)
Poly. (Class2)

The Problem

- Scorecard models - we frequently notice a curvature in the relationship between actual (data-estimated) log-odds and the log-odds inferred from the model predictions
- Seen in many different modelling scenarios (default, marketing propensity, fraud etc.)
- The curvature is approximately quadratic
- Common to many linear modelling techniques
- Gini measures ranking performance - very good Ginis
- Deviation at the lower and upper ends of the range – low odds and high odds
- Associated with irregular score distributions

Does it Matter?

NO – If we only care about rank order of accounts, bad rates above and below the cut-off.



- The traditional ways of assessing scorecard performance
  - Gini
  - K-S statistic
  - ROC curve
  - Concordance
  - Log-odds to score plot
  - Log-Likelihood

- Only the log-odds plot reveals incorrect p(event)

Does it Matter?

- ‘Traditional’ scorecard development
- Fit model to training data
  - Validate ranking performance of model on test set
  - Select cutoff
  - Deploy scorecard – use cutoff to make automated decisions (eg. grant credit, send mail, flag as fraud), vary as required to control volumes

- As long as the model’s ranking performance is good, we do not really care how good the model’s predictions *p(event)* are

- Any monotonic transformation of the model’s predictions will preserve ranking performance, Gini

- We make use of this when we scale scorecard coefficients to a standard form (eg, 600 for equal log odds, every +10 doubles log-odds)

Does it Matter?

- **YES** In some scenarios, model accuracy p(event) does matter
- Eg:
  - Application scorecard built on historic data, predicted bad rates for applicants today used to monitor business strategy and vary cut-off
  - Basel 2 – estimated point-in-time PDs by segment based on scorecard default rate estimates
  - Capital allocation (based on profitability of a campaign / action)
  - Targeting marketing campaigns on multiple dimensions
  - Account level strategies that use estimated default rate derived from a scorecard, eg. risk-based repricing, account closure, credit line increase or decrease, collections processes
  - Where we use estimated PD as one component to calculate NPV, expected loss, expected value

Does it Matter?

- Eg. we predict propensity to respond to a marketing campaign, divide the scores into segments, and target based on predicted response rates

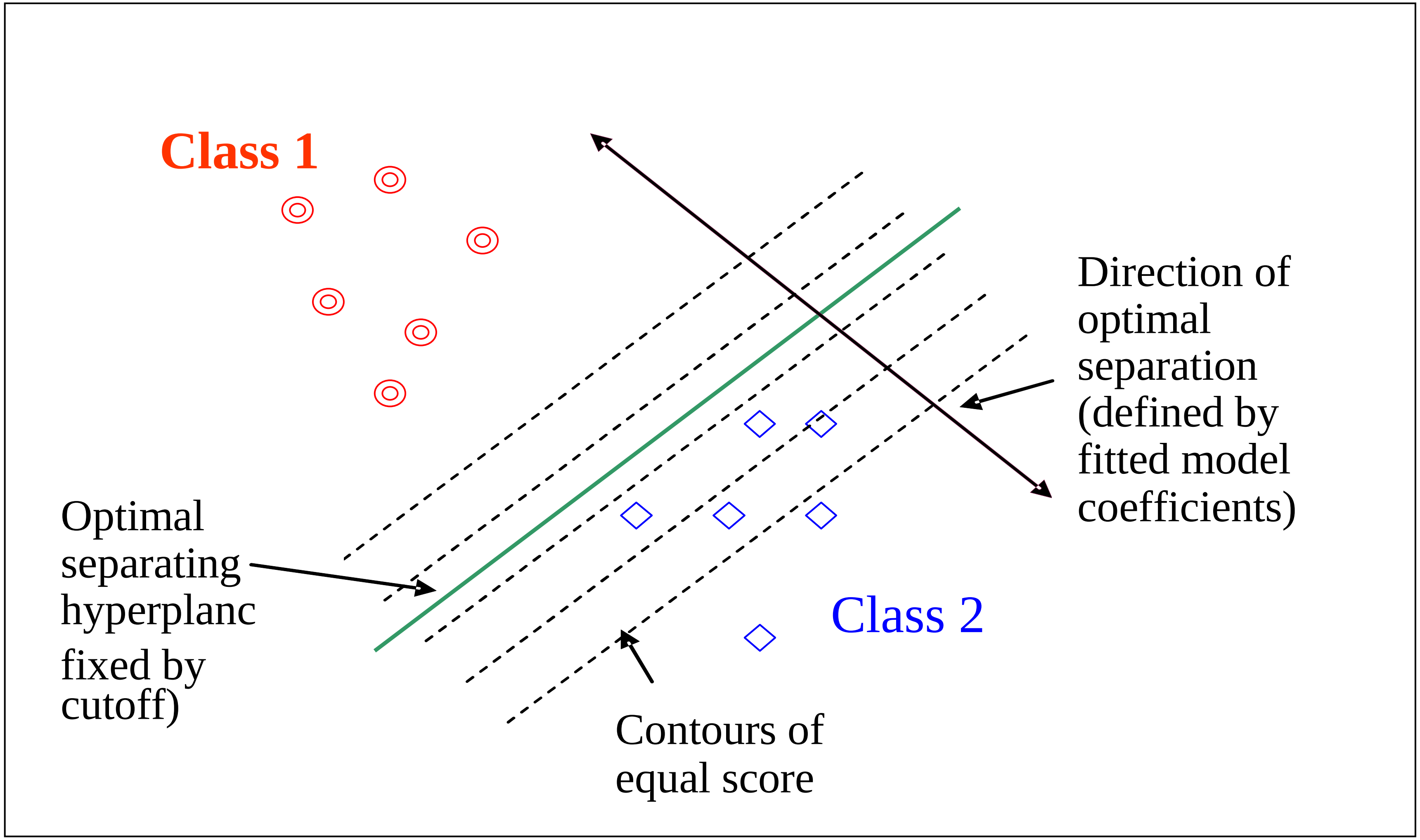
1	2	3	4	5	6	7	8	9	10
15%	12%	9%	6%	2%	1%	0.5%	0.25%	0.15%	0.05%

Expected response rate **10.5%**
acceptable

Why Does it Happen?

- Scorecards are <u>usually</u> linear or generalised linear (ie. logistic) models
- Advantages:
  - 1 coefficient per variable (or bin)
  - Interpretability – both model coefficients and output
  - Robustness
  - Easy to implement
  - Output can be linearly scaled to log-odds
  - Do not need technical understanding to use scorecards once built
- Disadvantages:
  - May not capture complex interactions easily
  - Possible instability of fitting method
  - <u>Non-linearity of log-odds</u>

Why Does it Happen?



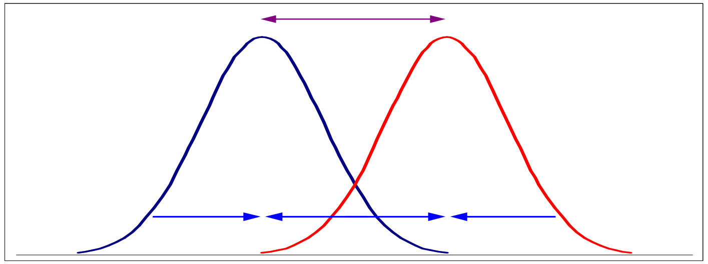
Why Does it Happen?

- Model-building methods:
  - Maximisation of Fisher Distance / Divergence

$$\frac{(\vec{\beta}\cdot(\vec{\mu}_1-\vec{\mu}_2))^2}{\vec{\beta}'(\Sigma_1+\Sigma_2)\vec{\beta}}$$

Ronald Fisher

- 'Maximise the separation of the means of score distributions while minimising the total variance'



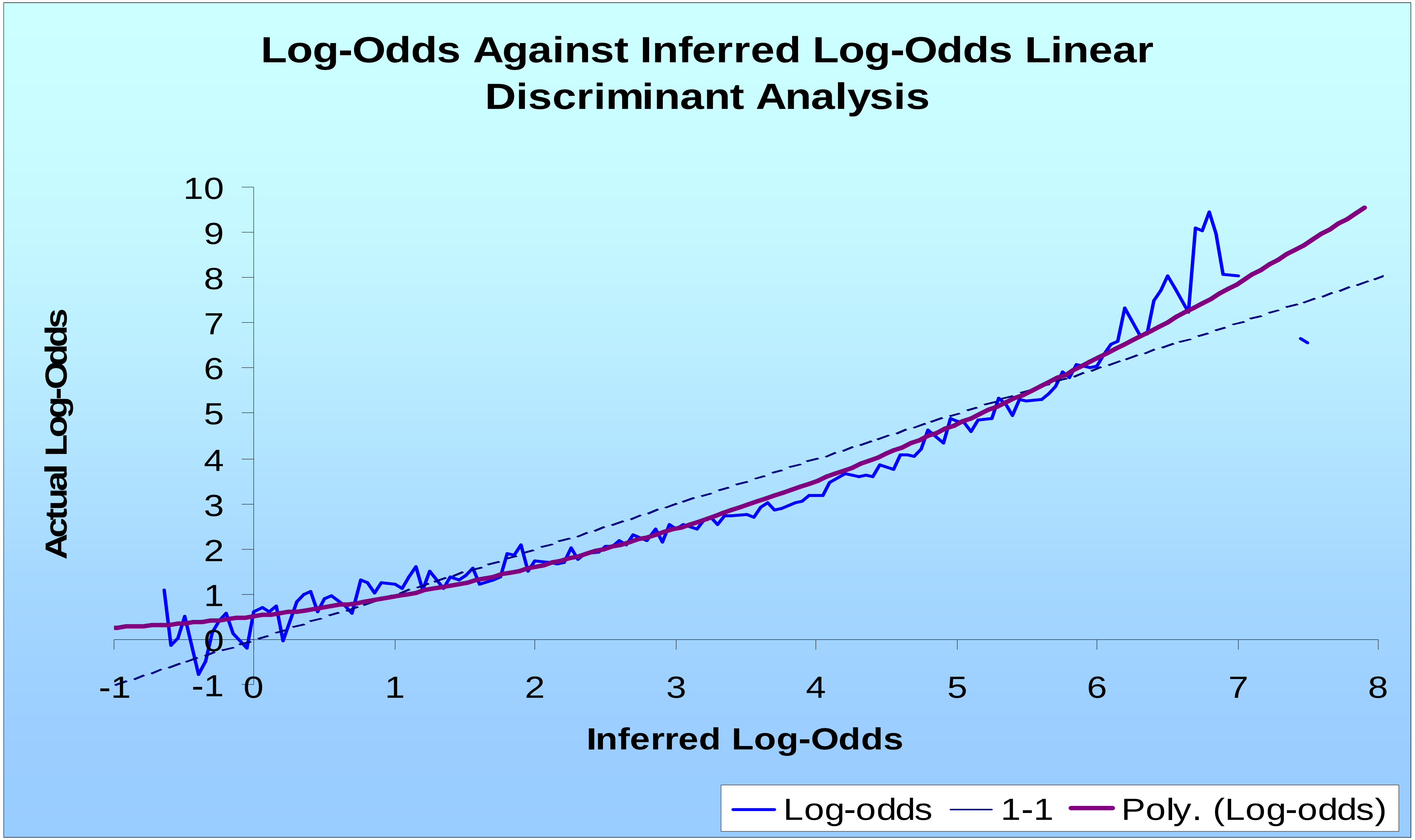
Why Does it Happen?

- Maximisation of Fisher Distance / Divergence -
- 2 options:
  - Linear Discriminant Analysis
    - Assume equal covariance matrix for both classes - then differentiate and set equal to zero.
    - Optimal solution is then given by

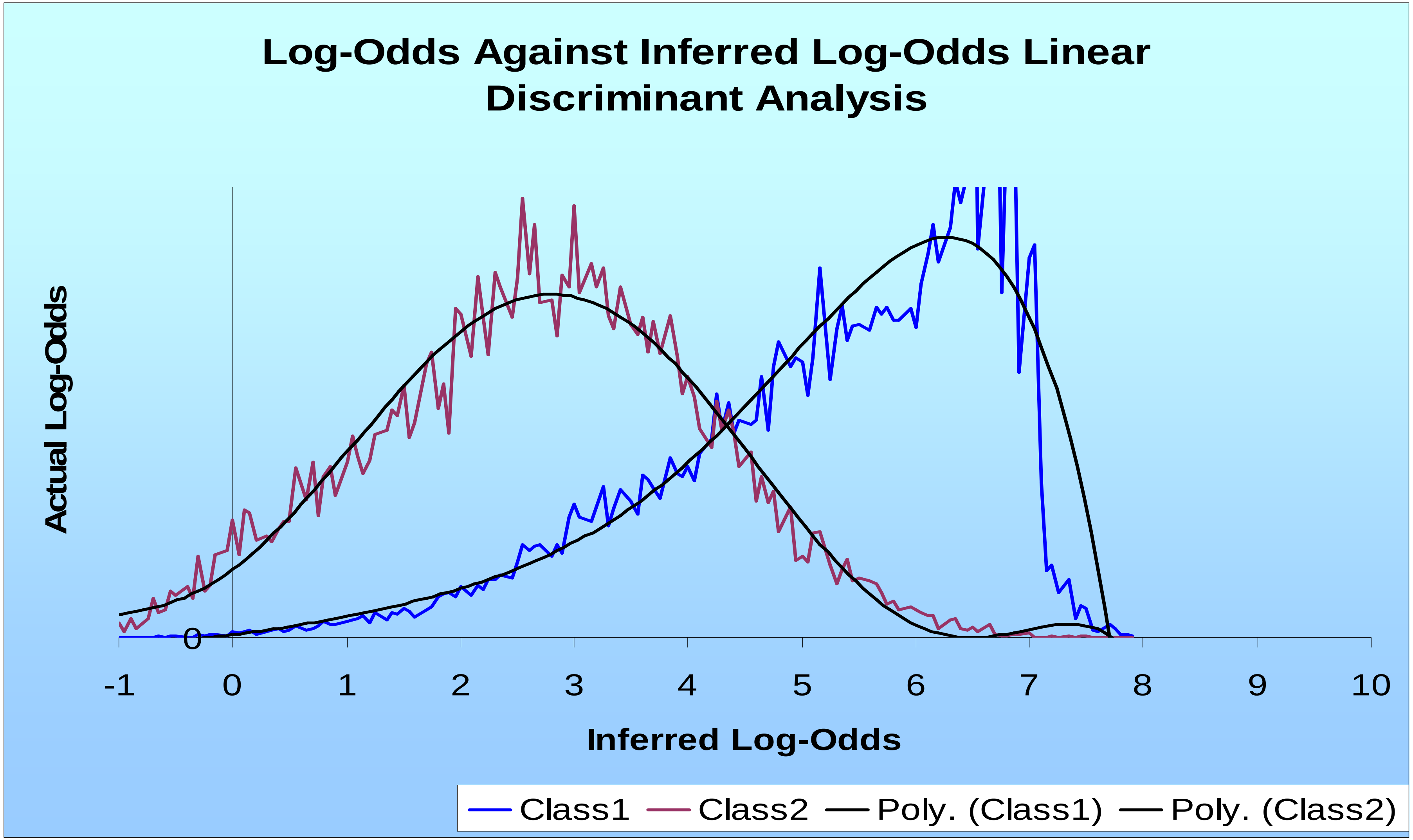
$$\vec{\beta} = \Sigma^{-1}\left(\vec{\mu}_{y=1} - \vec{\mu}_{y=2}\right)$$

  - Direct optimisation – Use an optimisation method to solve directly.

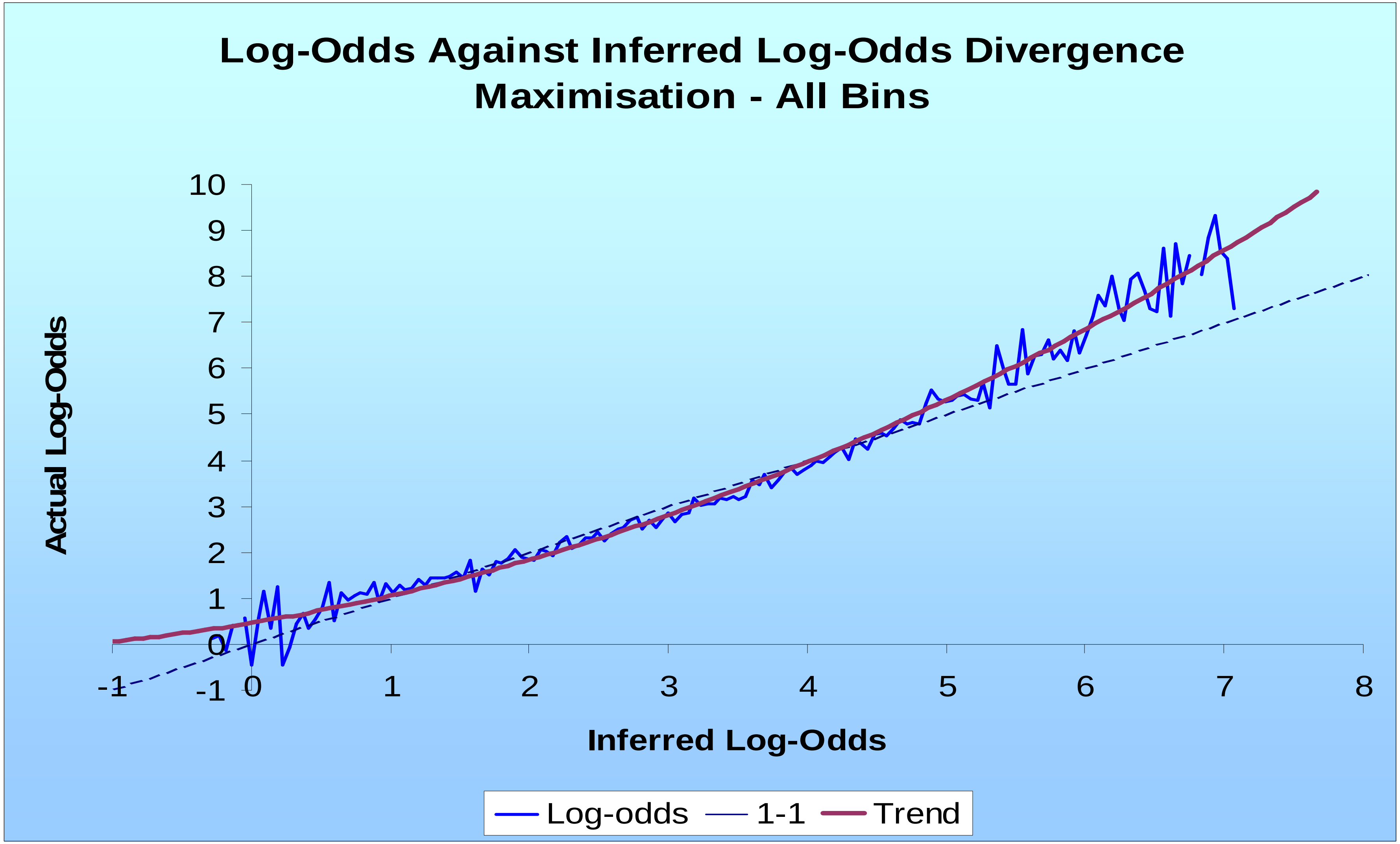
Why Does it Happen?



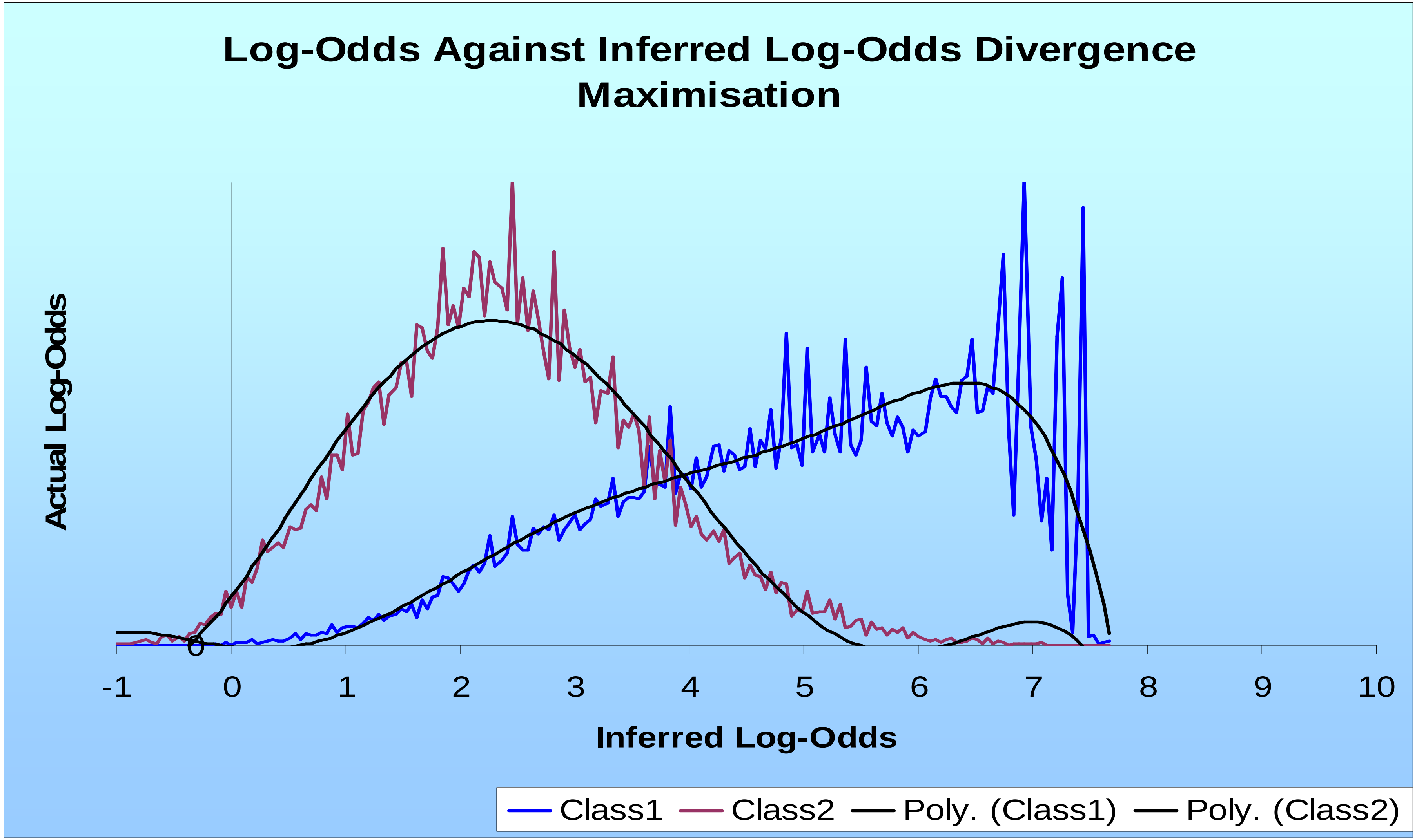
Why Does it Happen?



Why Does it Happen?



Why Does it Happen?



Why Does it Happen?

- Logistic Regression

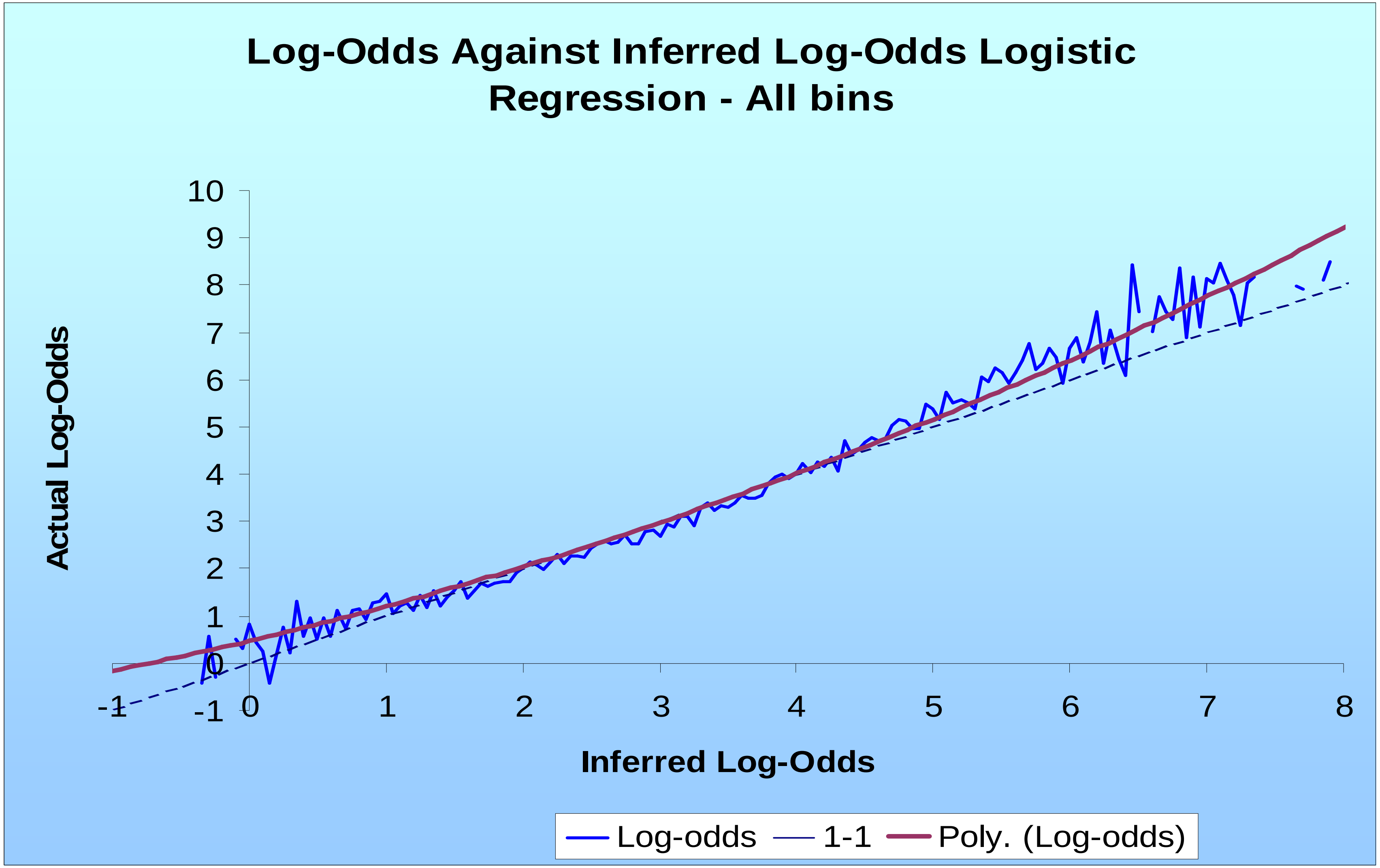
$$\log\left(\frac{p_2}{p_1}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$$

  - Solved by maximising likelihood, usually using iterative Newton-Raphson to solve non-linear equations (or can solve directly).
  - When there is only one binned variable in the model (so no overlap of dummy variables), it can be shown that

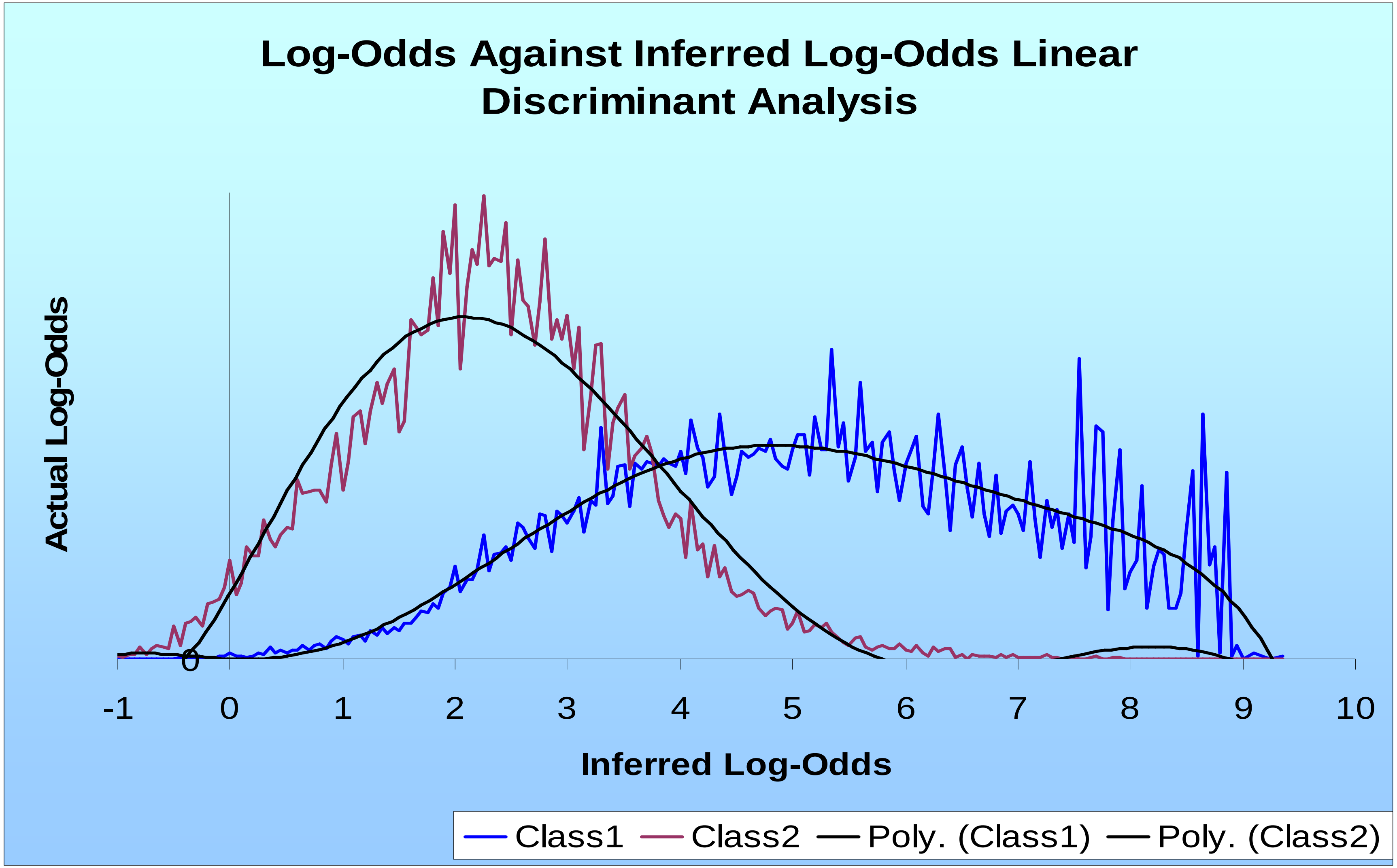
$$\beta_j = c\log\left(\frac{\sum_{x_i=1} y_i}{\sum_{x_i=1}(1-y_i)}\right)$$

← Weights of Evidence

Why Does it Happen?



Why Does it Happen?



Why Does it Happen?

- It can be shown that if the score distributions are normal with means and μ_1 and μ_2 variances σ_1^2 and σ_2^2 then the log-odds to score relationship is quadratic:

$$as^2 + bs + c$$

- Where

$$a = \frac{1}{2}\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2}\right),$$

$$b = \left(\frac{\mu_1}{\sigma_1^2} - \frac{\mu_2}{\sigma_2^2}\right),$$

$$c = \log\left(\frac{P(Class = 2)}{P(Class = 1)}\right) + \log\frac{\sigma_1}{\sigma_2} + \frac{1}{2}\left(\frac{\mu_1^2}{\sigma_1^2} - \frac{\mu_2^2}{\sigma_2^2}\right)$$

- ie. Unequal reciprocal variances in score distributions lead to curvature.
- Note that this assumes normally distributed scores.
- Why do we get unequal variances?

Why Does it Happen?

- **A simple model**
  - Let's suppose we have a linear model:

$$score = \beta_0 + \beta_{11}x_{11} + \ldots + \beta_{1m}x_{1m} + \ldots + \beta_{k1}x_{k1} + \ldots + \beta_{kn}x_{kn}$$

  - ie. we have binned all our variables 0/1 and included all bins in the model.
  - Assume each X is the outcome of a Bernouilli trial with parameter p dependent on class. ie. for class 1 we have probability p of observing a 1, (1-p) of observing a 0 etc. but p is conditional on class.
  - Across the set of bins for a single binned variable the x's will sum to exactly 1. This is a special case of the multinomial distribution with sum parameter *n* = 1.
  - Within a variable, $Var(X_i) = p_i(1-p_i)$, $Cov(X_i, X_j) = -p_i p_j$

Why Does it Happen?

LLOYDS BANKING GROUP

- A simple model, 2 variables, 2 bins:

$$score = \beta_0 + \beta_{11}x_{11} + \beta_{12}x_{12} + \beta_{21}x_{21} + \beta_{22}x_{22}$$

Within variable bins

$$Var(score \mid class) = \left(\beta_{11}{}^2 + \beta_{12}{}^2 - 2\beta_{11}\beta_{12}\right)p_{11}(1 - p_{11}) + \left(\beta_{21}{}^2 + \beta_{22}{}^2 - 2\beta_{21}\beta_{22}\right)p_{21}(1 - p_{21}) +$$

Bins across variables

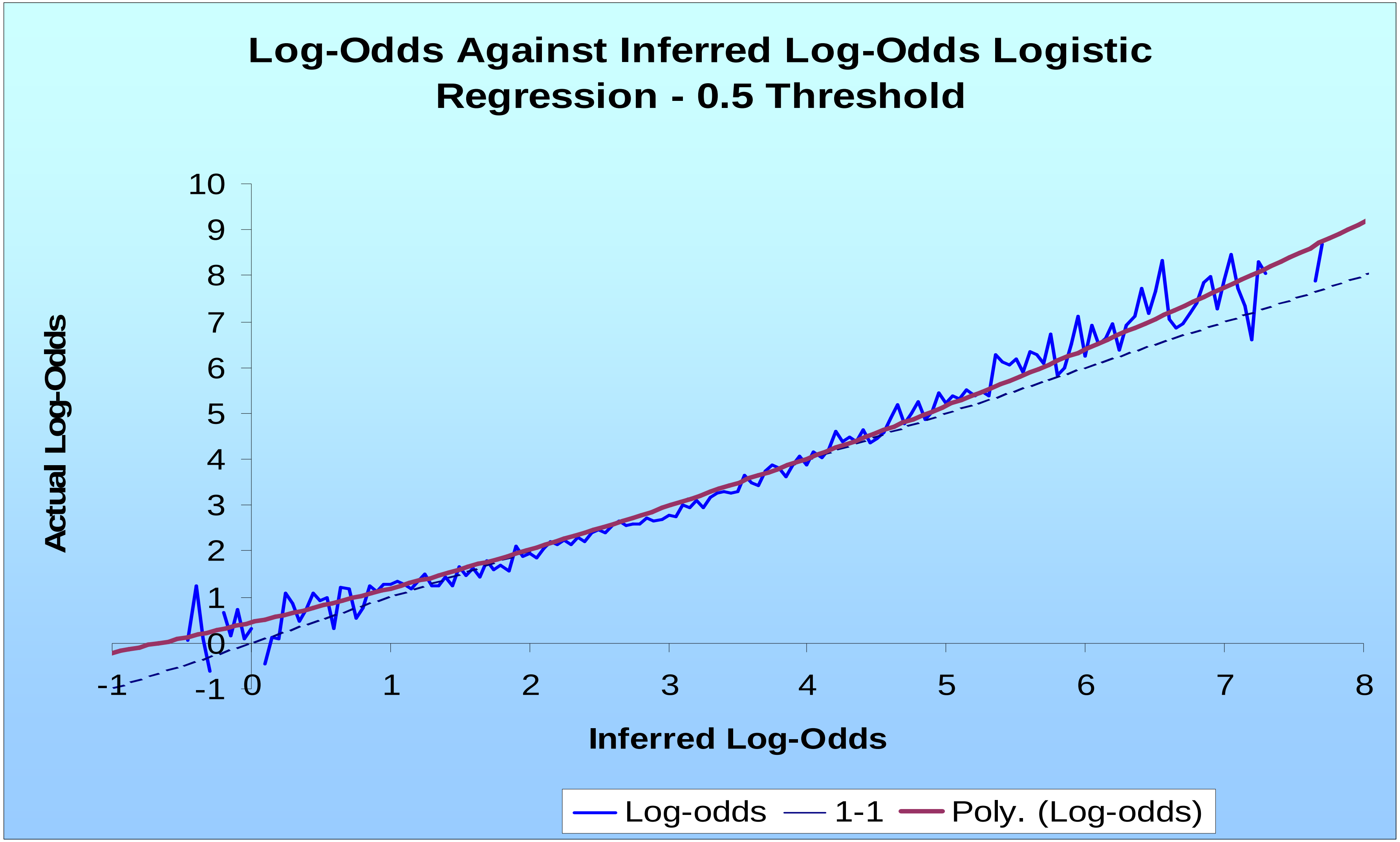
$$2\beta_{11}\beta_{21}Cov(X_{11}, X_{21}) + 2\beta_{12}\beta_{21}Cov(X_{12}, X_{21}) + 2\beta_{11}\beta_{22}Cov(X_{11}, X_{12}) + 2\beta_{11}\beta_{22}Cov(X_{12}, X_{22})$$

- For curvature, the reciprocal of these terms has to differ across classes. 2 components: covariance terms within a bin eg. $p_{11}(1\text{-}p_{11})$. These won't be equal across classes except in very rare cases (eg. perfect separation, where p_{11}=1 or 0).
- Large covariance terms for bins across different variables $Cov(X_{12},X_{21})$ etc.
- Can we reduce curvature by eliminating correlated bins?

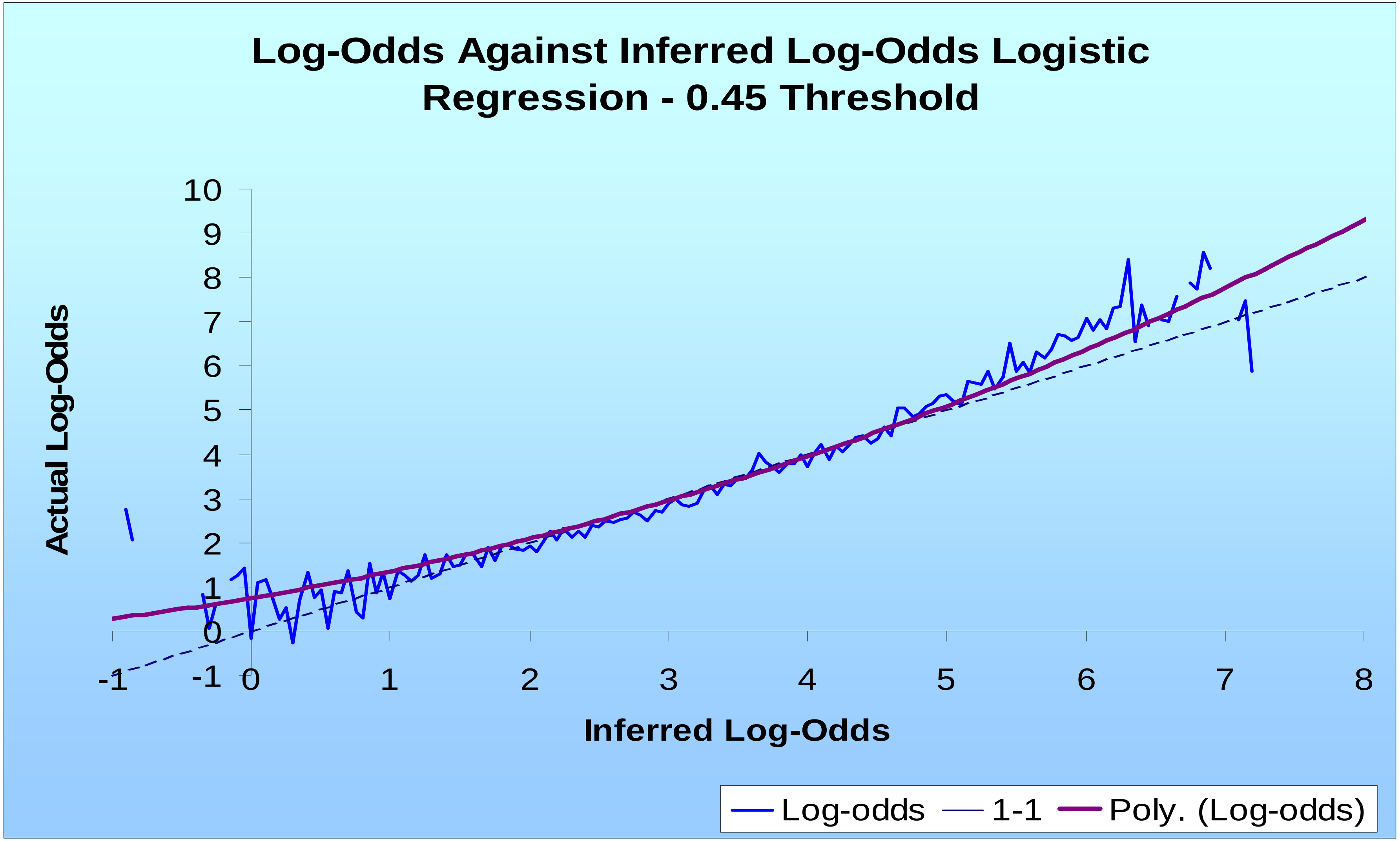
How Can We Fix It?

- Can we improve the curvature by eliminating correlated bins?
- High bin correlation occurs for several reasons:
  - High correlation of ‘missing’ category. Eg. Scorecard models often incorporate bureau data. If one field is missing, others are often missing for the same reason (no relevant data, failure to match address, etc.)
  - Correlation of values that have an underlying connection – eg. high credit balance usually means high interest charges.
  - Correlation of values where there is no qualifying data – eg. variables summarising arrears behaviour for someone who has never been in arrears.
- It is common practice when building a scorecard to monitor variable correlation.
- Correlation can lead to instability, since we may end up with multiple optimal solutions for coefficients. Reduces interpretability – coefficients may change sign
- **Risk:** even highly correlated variables can add useful information.

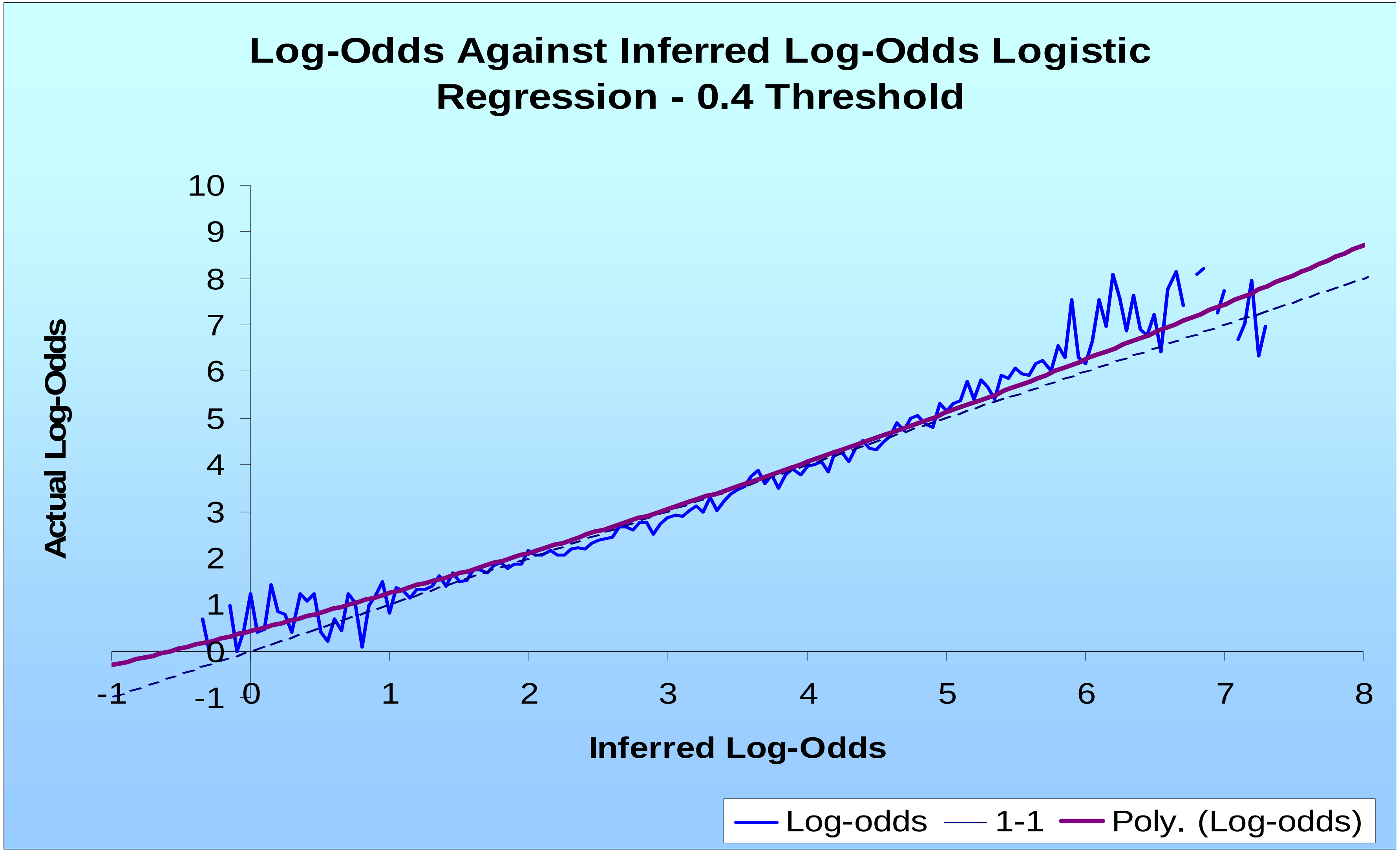
How Can We Fix It?



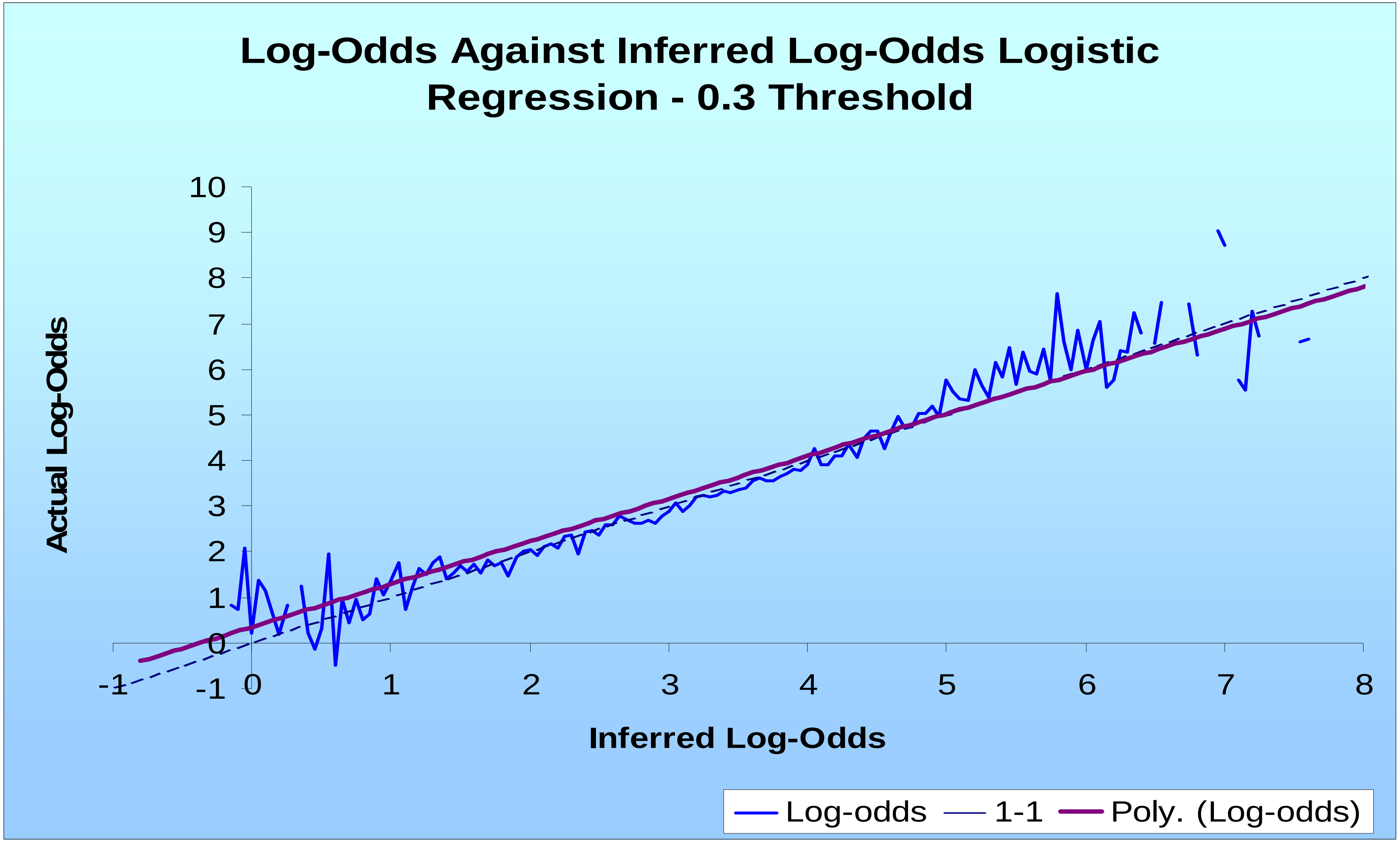
How Can We Fix It?



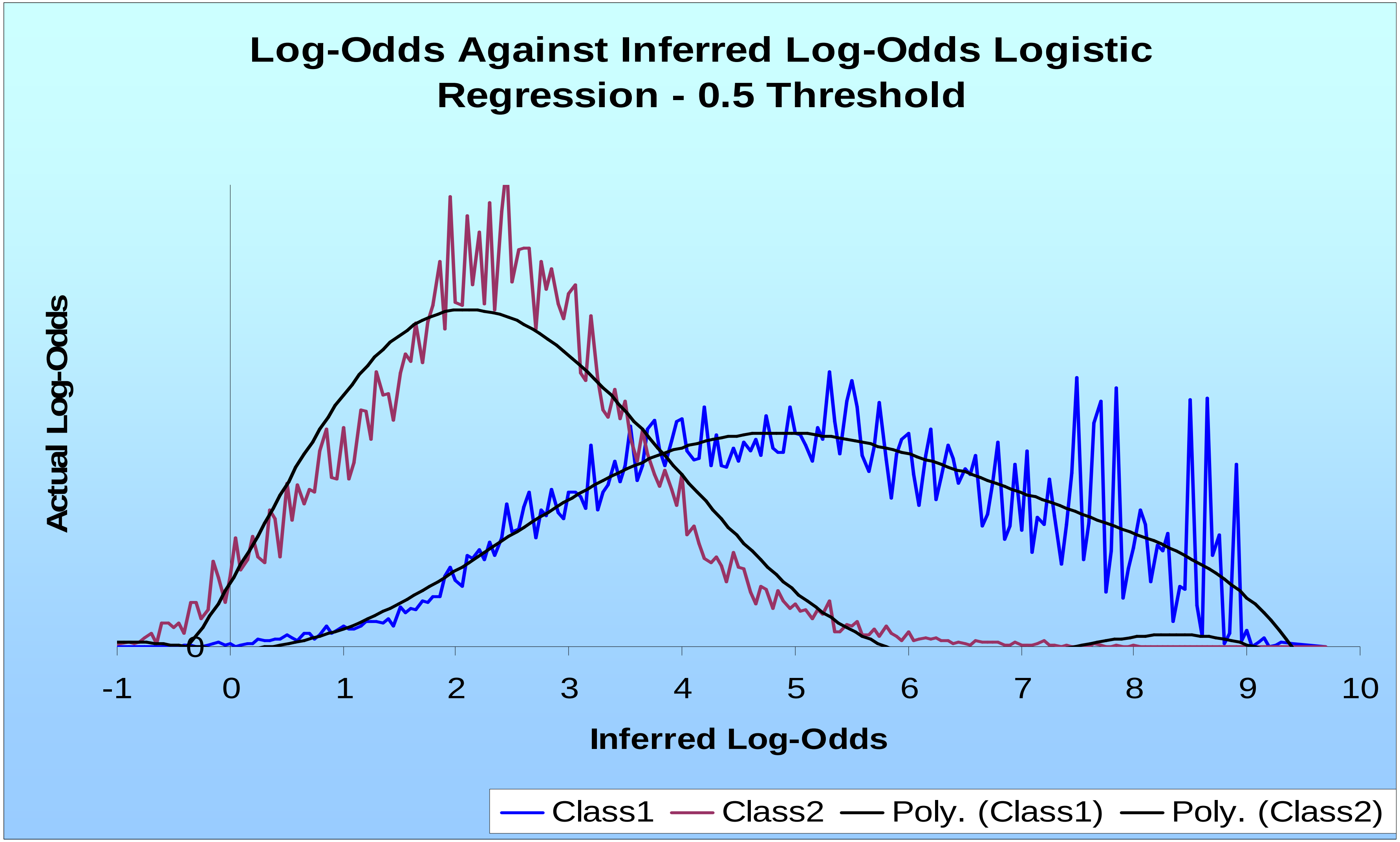
How Can We Fix It?



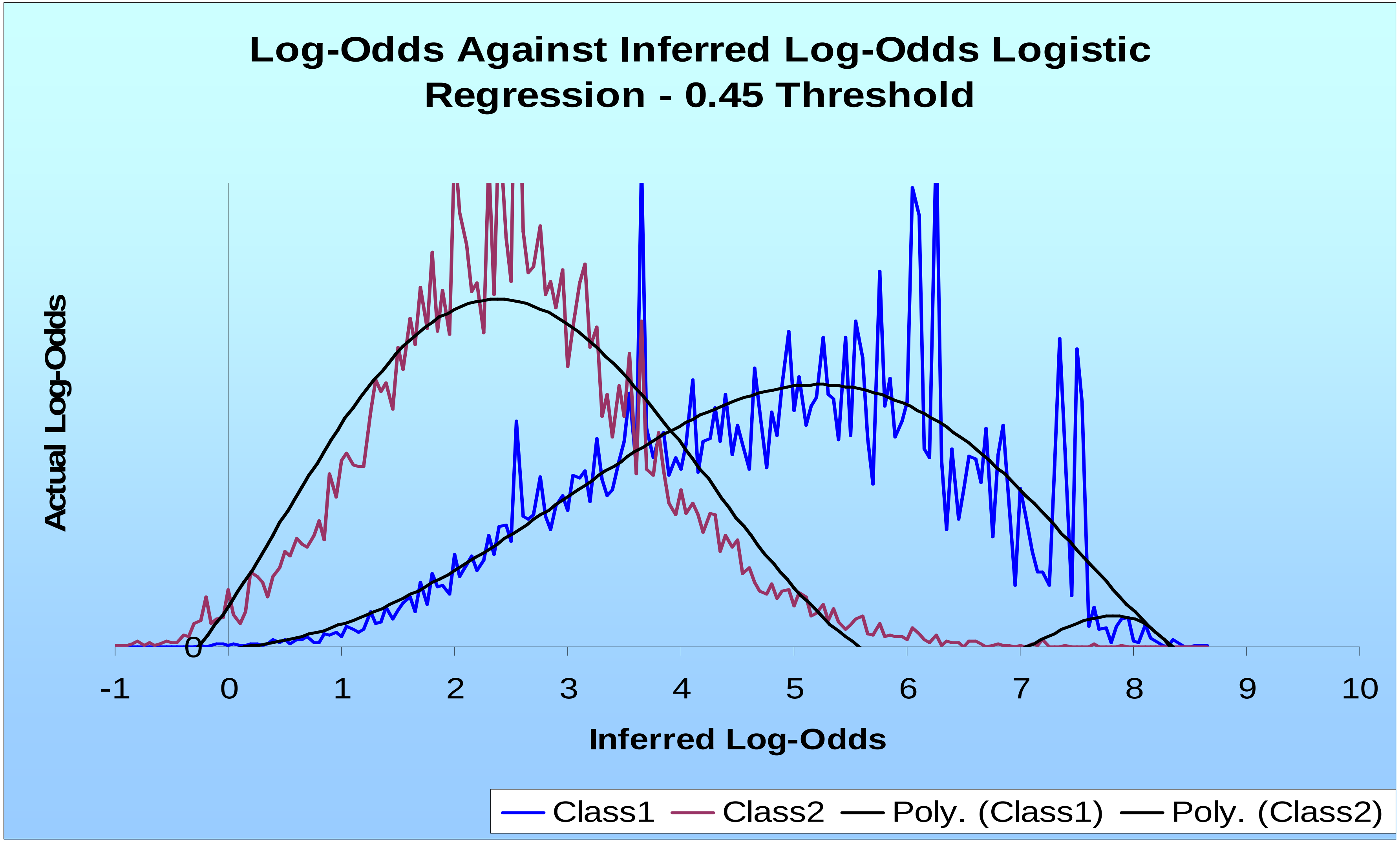
How Can We Fix It?



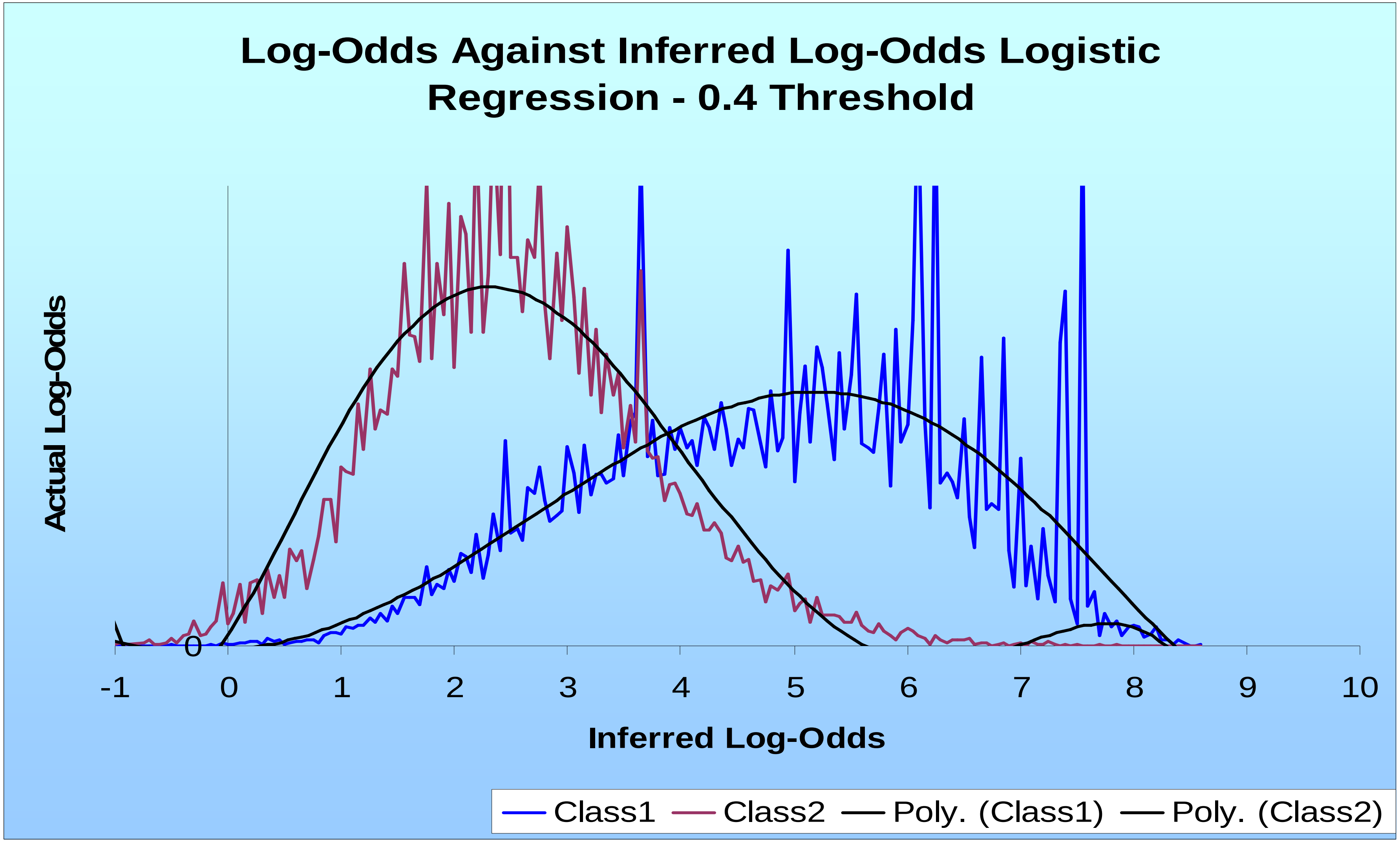
How Can We Fix It?



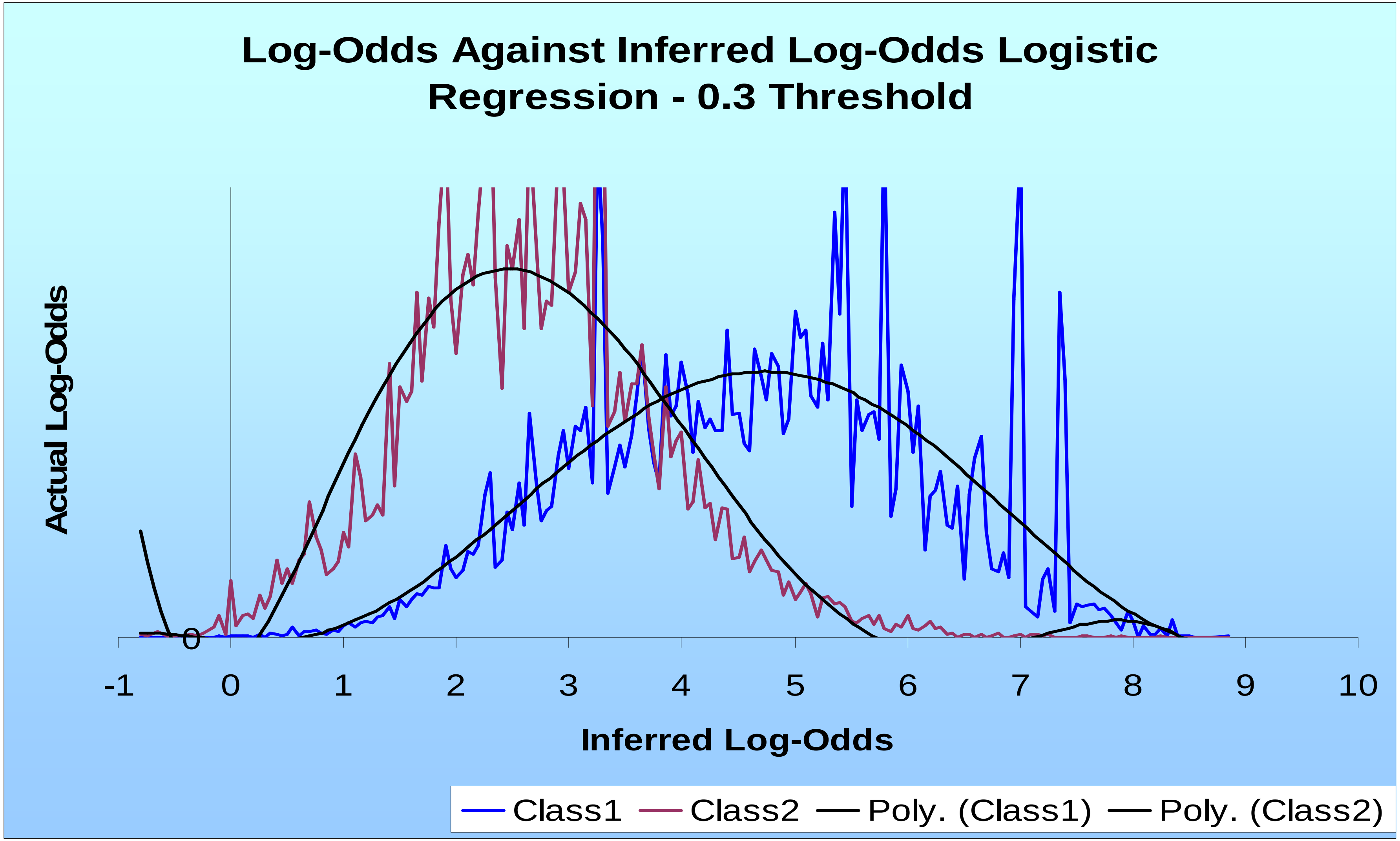
How Can We Fix It?



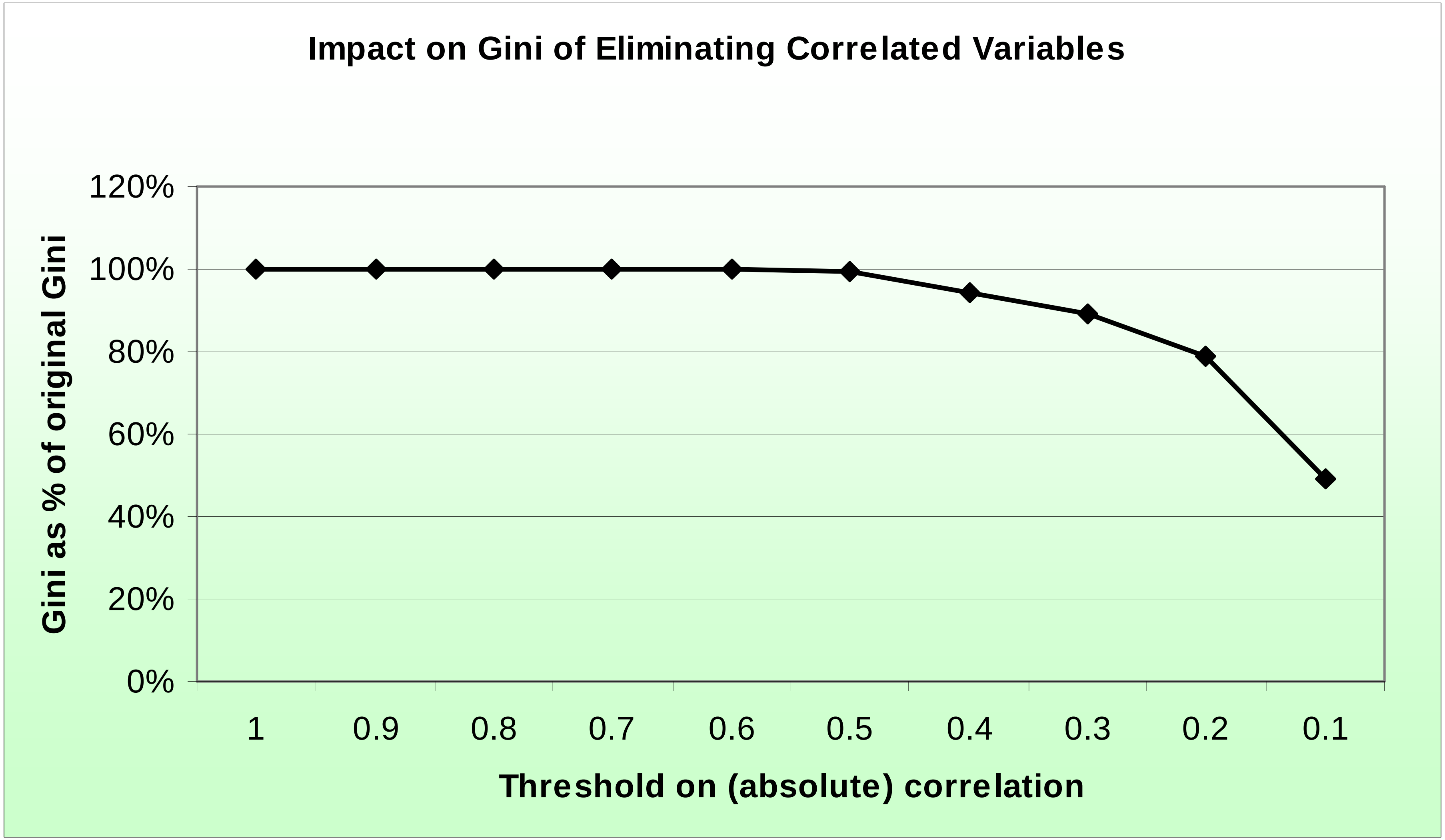
How Can We Fix It?



How Can We Fix It?



How Can We Fix It?

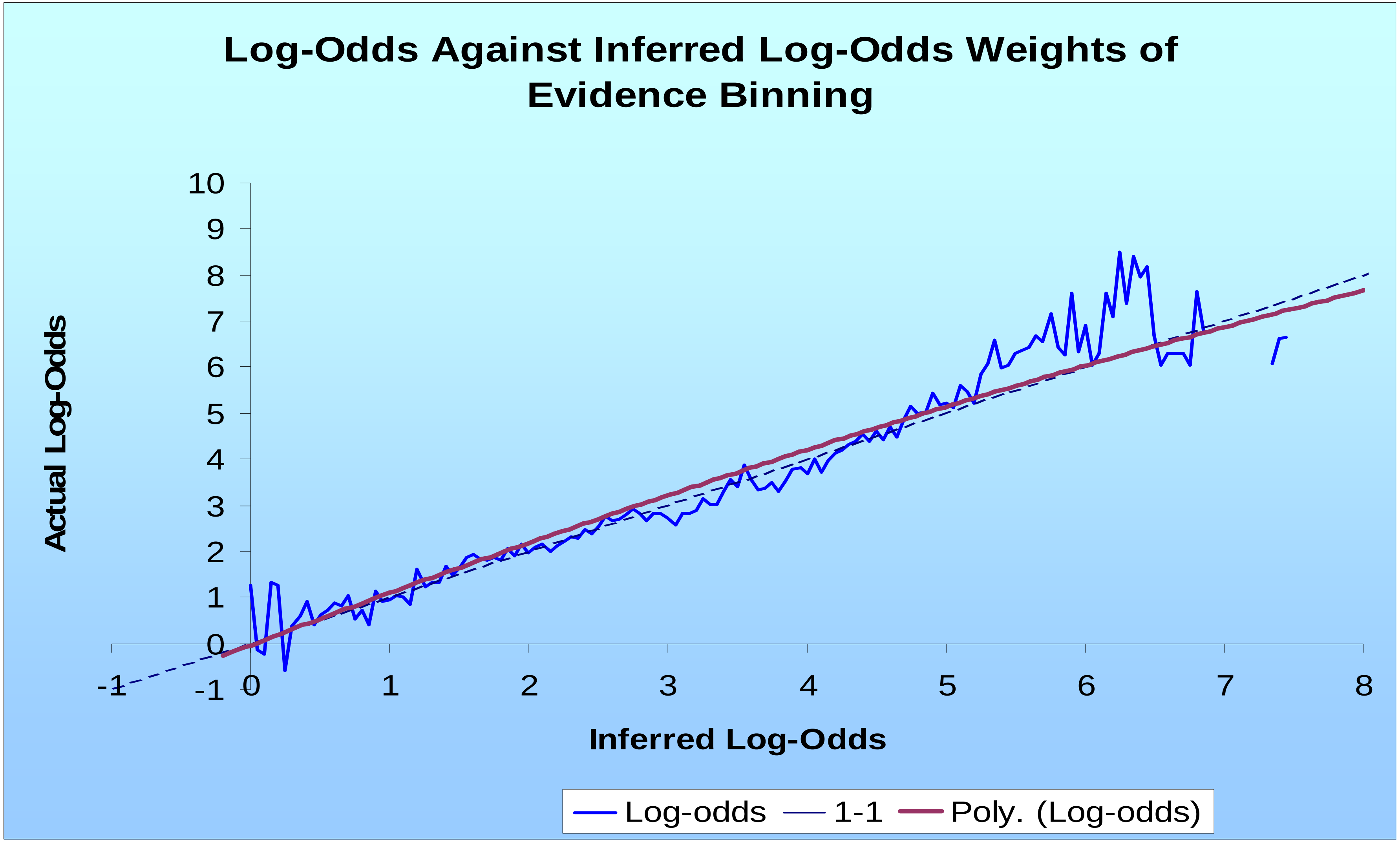


How Can We Fix It?

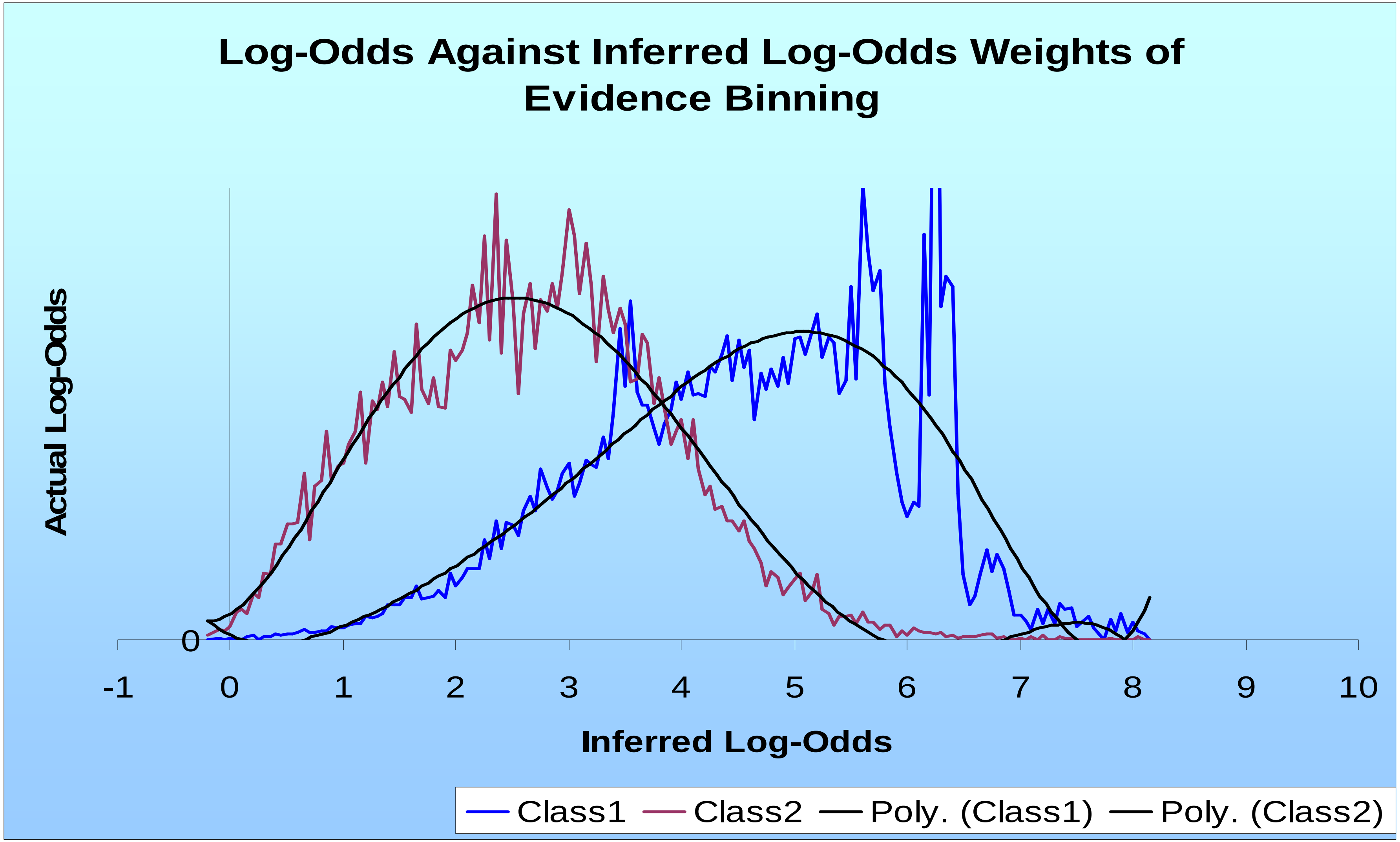
- **Other ideas for fixing non-linearity:**
  - Adopt a different binning method. Rather than using dummy variables, each variable takes a value equal to the weights of evidence value for the current bin. This is equivalent to using dummy variables and enforcing the condition that all model coefficients β must be proportional to the weights of evidence. In this case leads to a reduction of 10% in Gini.

  - Apply a retrospective, non-linear scaling to the model scores to align to log-odds (eg. apply a quadratic transformation). No impact on Gini. Disadvantage in that an extra transformation must be applied – no longer derive score by summing values

How Can We Fix It?

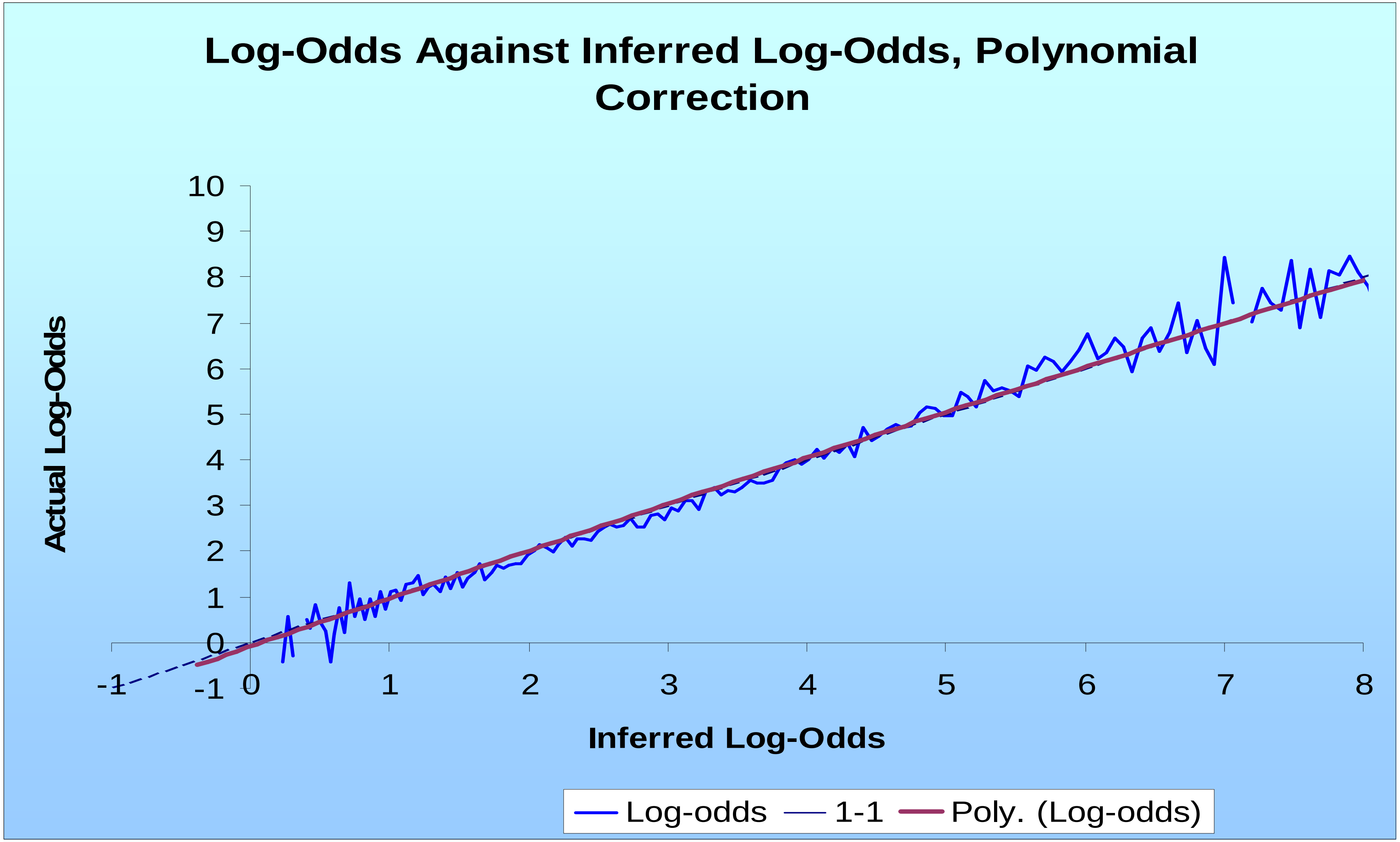


How Can We Fix It?

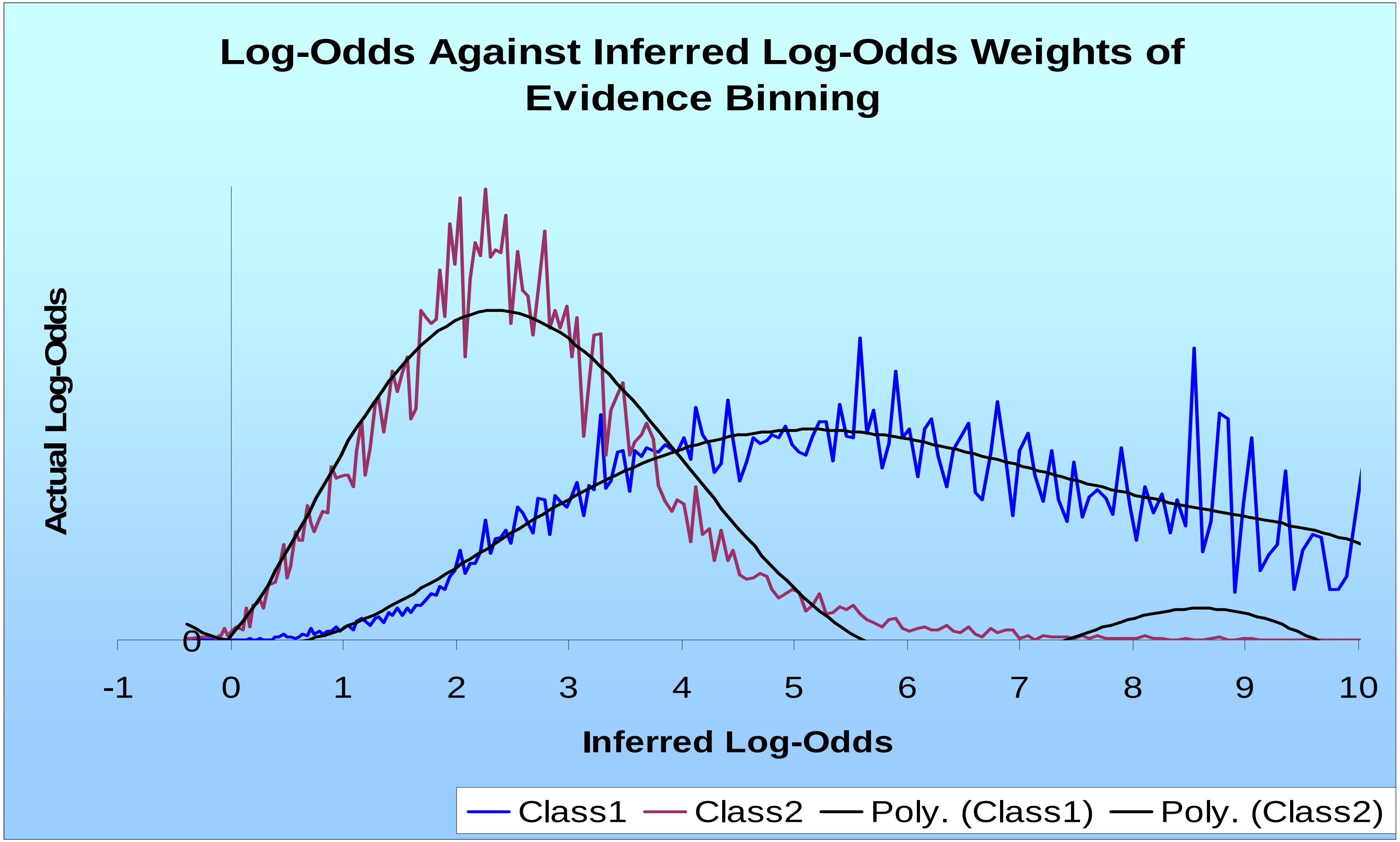


How Can We Fix It?

LLOYDS BANKING GROUP



How Can We Fix It?



Conclusion

- **Causes:**
  - Non-linear (quadratic) curvature caused by uneven score distributions.
  - This caused by underlying (non-normal) qualities of the data between classes, exacerbated by collinearity across variable bins
  - It has no adverse effect on ranking performance or Gini, but is a problem where we rely on accurate estimation of p(event)
- **Solutions:**
  - Eliminate collinear bins when building the model
  - Enforce weights of evidence proportionality in binning
  - Try different modelling techniques
  - Non-linear rescaling of output model scores

LLOYDS BANKING GROUP

Questions?