Exact Evaluation of Marginal Likelihood Integrals

Shaowei Lin (Joint work with B. Sturmfels, Z. Xu)

shaowei@math.berkeley.edu

27 Aug 2008, UC Davis

Menu

Appetizer

The Occasionally Dishonest Coin-Tosser

Main Course

Marginal Likelihood Integrals

Mixtures of Independence Model

Exact Formula for the Integral

Approximations of the Integral

Dessert

Two Different Examples

Occasionally Dishonest Coin-Tosser

- *The Deal*:
  Four coin tosses. If all are equal, you lose.

Occasionally Dishonest Coin-Tosser

- *The Deal*:
  Four coin tosses. If all are equal, you lose.

- *The Scam*:
  Two coins are involved, one fair and one biased.

Occasionally Dishonest Coin-Tosser

- *The Deal*:
  Four coin tosses. If all are equal, you lose.

- *The Scam*:
  Two coins are involved, one fair and one biased.

- *The Data*:

#Heads	0	1	2	3	4
#Occurrences	51	18	73	25	75

Occasionally Dishonest Coin-Tosser

- *The Deal*:
  Four coin tosses. If all are equal, you lose.

- *The Scam*:
  Two coins are involved, one fair and one biased.

- *The Data*:

#Heads	0	1	2	3	4
#Occurrences	51	18	73	25	75

- *The Burning Question*:
  How many coins did he use?

Occasionally Dishonest Coin-Tosser

- Model *One*:

Parameters	Coin:	$0 \leq \theta_h, \theta_t \leq 1,\ \theta_h + \theta_t = 1$
Prob(i heads)	p_i =	$\binom{4}{i}\theta_h^i \theta_t^{4-i}$
Likelihood of data U	$L_U(\theta)$ =	$zp_0^{51}p_1^{18}p_2^{73}p_3^{25}p_4^{75}$ = $z4^{43}6^{73}\theta_h^{539}\theta_t^{429}$
	where z =	$242!/(51! \cdot 18! \cdot 73! \cdot 25! \cdot 75!)$

Occasionally Dishonest Coin-Tosser

- Model *One*:

  Parameters — Coin: $0 \leq \theta_h, \theta_t \leq 1,\ \theta_h + \theta_t = 1$

  Prob(i heads) — $p_i = \binom{4}{i} \theta_h^i \theta_t^{4-i}$

  Likelihood of data U — $L_U(\theta) = z p_0^{51} p_1^{18} p_2^{73} p_3^{25} p_4^{75} = z 4^{43} 6^{73} \theta_h^{539} \theta_t^{429}$

  where $z = 242!/(51! \cdot 18! \cdot 73! \cdot 25! \cdot 75!)$

- Model *Two*:

  Parameters — Coin 0: $0 \leq \theta_h, \theta_t, \leq 1,\ \ \theta_h + \theta_t = 1$

  Coin 1: $0 \leq \rho_h, \rho_t \leq 1,\ \ \rho_h + \rho_t = 1$

  Choice of coin: $0 \leq \sigma_0, \sigma_1 \leq 1,\ \ \sigma_0 + \sigma_1 = 1$

  Prob(i heads) — $p_i = \binom{4}{i} (\sigma_0 \theta_h^i \theta_t^{4-i} + \sigma_1 \rho_h^i \rho_t^{4-i})$

  Likelihood of data U — $L_U(\theta) = z p_0^{51} p_1^{18} p_2^{73} p_3^{25} p_4^{75}$

Occasionally Dishonest Coin-Tosser

- **Question:** How do we do model selection?

Occasionally Dishonest Coin-Tosser

- **Question:** How do we do model selection?
- **Method 1:** Maximum Likelihood

  Compare the maximum values of the likelihood functions.

$$\max_{\theta \in \Theta} L_U(\theta)$$

Occasionally Dishonest Coin-Tosser

- **Question:** How do we do model selection?
- **Method 1:** Maximum Likelihood

  Compare the maximum values of the likelihood functions.

$$\max_{\theta \in \Theta} L_U(\theta)$$

- **Method 2:** Marginal Likelihood

  Integrate the likelihood functions over the parameter space.

$$\int_{\Theta} L_U(\theta) d\theta$$

Marginal Likelihood Integrals

$$\int_{\Theta} L_U(\theta) p(\theta) d\theta$$

Prior Beliefs

- Probability measures $p(\theta)$ on the parameter space represent prior beliefs.
- Can be viewed as probability of model given prior beliefs about parameters.
- Maximum likelihood represents the prior belief that the parameters are optimal.
- Common priors: the uniform prior, the Dirichlet prior.

Marginal Likelihood Integrals

$$\int_{\Theta} L_U(\theta) p(\theta) d\theta$$

Currently

- Very difficult to compute exactly.
- Tackled using MCMC, importance sampling methods.
- Approximation formulas limited to special cases.
- Accuracy of above methods and formulas questionable.

Our Goal

- Show that they can be computed *exactly* in *many* cases previously thought impractical.

Marginal Likelihood Integrals

$$\int_{\Theta} L_U(\theta) p(\theta) d\theta$$

What exactly is Exact Evaluation?!

- When $L_U(\theta)$ is a polynomial, $p(\theta) = 1$, Θ is a polytope, the integral is a *rational* number.
- Exact evaluation is computing this rational number, not its floating point approximation!
- e.g. Coin Toss Model Two

$$z' \cdot \int_{\Delta_1 \times \Delta_1 \times \Delta_1} \prod_{i=0}^{4} (\sigma_0 \theta_h^i \theta_t^{4-i} + \sigma_1 \rho_h^i \rho_t^{4-i})^{U_i} d\sigma d\theta d\rho \quad =$$

Marginal Likelihood Integrals

$$\frac{2805748035222313067135398014075361975978864622352256160544759816747367817994434767196492009426285781414295477891948457579449463459708735310230424897127628337608457740525732502310552980846527032258197855156758075892511025767529711754486138526055065915281254761412080217673204703018187910949369084430474540784253322654356704060651978380627529093477438708340212046389726976493345195544134714220439905754357896320656893049737172976960604156324007410505634773422386363996473847553080097785724548383890969259688769804869503436965543936}{3602324071338125874577562671962054628339147256791746496077298664579499436836889049486689507051463879264328153845162002285178224453663460279080758904156945946390977724512859312036096765746313969020541775346907766998180397769609299339804266010207548603870980861129358173839607260454683402083005508959248902903340347663670605747176619993139607889832999867603350320070482837740687067608852004726493742428623588390160566874549440724360484442163404900024396516685851371805424013821775746444698614706300105139962637751537933349768190601412833540994898650618 75.}{}$$

Mixtures of Independence Models

Coin Toss Example

Random Variables

$X_1, X_2, \ldots, X_4 \in \{0, 1\}$ identically distributed.

Model Parameters

$\theta_0, \theta_1, \quad \theta \in \Delta_1.$

Independence Model

$p_v = \theta^{a_v}$, where a_v are the columns of a 2×16 matrix

$$A \quad = \quad \begin{matrix} \\ \theta_0 \\ \theta_1 \end{matrix} \begin{matrix} \begin{matrix} p_{0000} & p_{0001} & p_{0010} & \cdots & p_{1101} & p_{1110} & p_{1111} \end{matrix} \\ \begin{pmatrix} 4 & 3 & 3 & \ldots & 1 & 1 & 0 \\ 0 & 1 & 1 & \ldots & 3 & 3 & 4 \end{pmatrix} \end{matrix}$$

Two-Mixture $\quad p_v = \sigma_0 \theta^{a_v} + \sigma_1 \rho^{a_v}, \quad \sigma \in \Delta_1.$

Three-Mixture $\quad p_v = \sigma_0 \theta^{a_v} + \sigma_1 \rho^{a_v} + \sigma_2 \tau^{a_v}, \quad \sigma \in \Delta_2.$

Mixtures of Independence Models

- Random Variables

$$X_1^{(1)}, X_2^{(1)}, \ldots, X_{s_1}^{(1)} \in \{0, \ldots, t_1\} \text{ identically distributed,}$$
$$\ldots$$
$$X_1^{(k)}, X_2^{(k)}, \ldots, X_{s_k}^{(k)} \in \{0, \ldots, t_k\} \text{ identically distributed.}$$

Mixtures of Independence Models

- Random Variables

$$X_1^{(1)}, X_2^{(1)}, \ldots, X_{s_1}^{(1)} \in \{0, \ldots, t_1\} \text{ identically distributed,}$$
$$\ldots$$
$$X_1^{(k)}, X_2^{(k)}, \ldots, X_{s_k}^{(k)} \in \{0, \ldots, t_k\} \text{ identically distributed.}$$

- Model Parameters

$$\theta_0^{(1)}, \theta_1^{(1)}, \ldots, \theta_{t_1}^{(1)}, \quad \theta^{(1)} \in \Delta_{t_1}.$$
$$\ldots$$
$$\theta_0^{(k)}, \theta_1^{(k)}, \ldots, \theta_{t_k}^{(k)}, \quad \theta^{(k)} \in \Delta_{t_k}.$$

Mixtures of Independence Models

- Random Variables

$$X_1^{(1)}, X_2^{(1)}, \ldots, X_{s_1}^{(1)} \in \{0, \ldots, t_1\} \text{ identically distributed,}$$
$$\ldots$$
$$X_1^{(k)}, X_2^{(k)}, \ldots, X_{s_k}^{(k)} \in \{0, \ldots, t_k\} \text{ identically distributed.}$$

- Model Parameters

$$\theta_0^{(1)}, \theta_1^{(1)}, \ldots, \theta_{t_1}^{(1)}, \quad \theta^{(1)} \in \Delta_{t_1}.$$
$$\ldots$$
$$\theta_0^{(k)}, \theta_1^{(k)}, \ldots, \theta_{t_k}^{(k)}, \quad \theta^{(k)} \in \Delta_{t_k}.$$

- Independence Model

Can be represented by a $d \times n$ matrix A, where

d = #parameters = $(t_1 + 1) + (t_2 + 1) + \cdots + (t_k + 1)$,

n = #outcomes = $(t_1 + 1)^{s_1} (t_2 + 1)^{s_2} \cdots (t_k + 1)^{s_k}$.

The column a_v corresponds to the probability $p_v = \theta^{a_v}$.

Mixtures of Independence Models

- Random Variables

$$X_1^{(1)}, X_2^{(1)}, \ldots, X_{s_1}^{(1)} \in \{0, \ldots, t_1\} \text{ identically distributed,}$$
$$\ldots$$
$$X_1^{(k)}, X_2^{(k)}, \ldots, X_{s_k}^{(k)} \in \{0, \ldots, t_k\} \text{ identically distributed.}$$

- Model Parameters

$$\theta_0^{(1)}, \theta_1^{(1)}, \ldots, \theta_{t_1}^{(1)}, \quad \theta^{(1)} \in \Delta_{t_1}.$$
$$\ldots$$
$$\theta_0^{(k)}, \theta_1^{(k)}, \ldots, \theta_{t_k}^{(k)}, \quad \theta^{(k)} \in \Delta_{t_k}.$$

- Independence Model

Can be represented by a $d \times n$ matrix A, where

d = #parameters = $(t_1 + 1) + (t_2 + 1) + \cdots + (t_k + 1)$,

n = #outcomes = $(t_1 + 1)^{s_1} (t_2 + 1)^{s_2} \cdots (t_k + 1)^{s_k}$.

The column a_v corresponds to the probability $p_v = \theta^{a_v}$.

- Mixtures

$$p_v = \sigma_0 \theta^{a_v} + \ldots + \sigma_l \rho^{a_v}, \quad \sigma \in \Delta_l.$$

Mixtures of Independence Models

- Random Variables

$$X_1^{(1)}, X_2^{(1)}, \ldots, X_{s_1}^{(1)} \in \{0, \ldots, t_1\} \text{ identically distributed,}$$
$$\ldots$$
$$X_1^{(k)}, X_2^{(k)}, \ldots, X_{s_k}^{(k)} \in \{0, \ldots, t_k\} \text{ identically distributed.}$$

- Model Parameters

$$\theta_0^{(1)}, \theta_1^{(1)}, \ldots, \theta_{t_1}^{(1)}, \quad \theta^{(1)} \in \Delta_{t_1}.$$
$$\ldots$$
$$\theta_0^{(k)}, \theta_1^{(k)}, \ldots, \theta_{t_k}^{(k)}, \quad \theta^{(k)} \in \Delta_{t_k}.$$

- Independence Model

Can be represented by a $d \times n$ matrix A, where

d = #parameters = $(t_1 + 1) + (t_2 + 1) + \cdots + (t_k + 1)$,

n = #outcomes = $(t_1 + 1)^{s_1} (t_2 + 1)^{s_2} \cdots (t_k + 1)^{s_k}$.

The column a_v corresponds to the probability $p_v = \theta^{a_v}$.

- Mixtures

$$p_v = \sigma_0 \theta^{a_v} + \ldots + \sigma_l \rho^{a_v}, \quad \sigma \in \Delta_l.$$

- Data

$$U = (U_v), \quad N = \textstyle\sum_v U_v.$$

Exact Formula for the Integral

Main Formula:

Integrating a monomial over a simplex

$$\int_{\Delta_m} \theta_0^{b_0} \theta_1^{b_1} \cdots \theta_m^{b_m} d\theta = \frac{m! \cdot b_0! \cdot b_1! \cdot \cdots \cdot b_m!}{(b_0 + b_1 + \cdots + b_m + m)!}$$

Sanity check: what if the monomial is 1?

Exact Formula for the Integral

Independence Model:

Let $z = N!/\prod_v U_v!$, $b = AU$, $P = \Delta_{t_1} \times \cdots \times \Delta_{t_k}$. Then,

$$\begin{aligned} L_U(\theta) &= z \cdot \theta^b \\ \int_P L_U(\theta)\, d\theta &= z \cdot \int_{\Delta_{t_1}} \theta^{b^{(1)}}\, d\theta^{(1)} \cdots \int_{\Delta_{t_k}} \theta^{b^{(k)}}\, d\theta^{(k)} \\ &= z \cdot \prod_{i=1}^{k} \frac{t_i!\, b_0^{(i)}!\, b_1^{(i)}! \cdots b_{t_i}^{(i)}!}{(s_i N + t_i)!} \end{aligned}$$

Note that the maximum and marginal likelihood of the independence model are both easy to compute.

Exact Formula for the Integral

Mixture of Independence Model:

Let $\Theta = \Delta_1 \times P \times P$. Then,

$$\begin{aligned} L_U(\sigma, \theta, \rho) &= z \cdot \prod_v (\sigma_0 \theta^{a_v} + \sigma_1 \rho^{a_v})^{U_v} \\ &= z \cdot \sum_b \phi(b) \cdot \sigma^{(b,c)/a} \cdot \theta^b \cdot \rho^c \end{aligned}$$

$$\int_\Theta L_U(\sigma, \theta, \rho)\ d\sigma d\theta d\rho = z \cdot \sum_b \phi(b) \int_{\Delta_1} \sigma^{(b,c)/a} d\sigma \int_P \theta^b d\theta \int_P \rho^c d\rho$$

where $\phi(b)$ is the coefficient of θ^b in the expansion of $\prod_v (\theta^{a_v} + 1)^{U_v}$, $c = AU - b$, and $a =$ column sum of A.

Exact Formula for the Integral

Formula:

$$\int_{\Theta} L_U(\sigma, \theta, \rho)\ d\sigma d\theta d\rho = z \cdot \sum_{b \in Z} \phi(b) \int_{\Delta_1} \sigma^{(b,c)/a} d\sigma \int_P \theta^b d\theta \int_P \rho^c d\rho$$

Computational Considerations:

- Naive estimate of number of monomials in the expansion of

$$L_U(\sigma, \theta, \rho) = z \cdot \prod_v (\sigma_0 \theta^{a_v} + \sigma_1 \rho^{a_v})^{U_v}$$

  is $\prod_v (U_v + 1)$.

- Actual number of monomials is *a lot less*.
- e.g. Coin Toss Model Two: 144,469,312 vs 48,646.
- Idea: exploit this reduction in the computation.

Exact Formula for the Integral

Formula:

$$\int_{\Theta} L_U(\sigma, \theta, \rho)\ d\sigma d\theta d\rho = z \cdot \sum_{b \in Z} \phi(b) \int_{\Delta_1} \sigma^{(b,c)/a} d\sigma \int_P \theta^b d\theta \int_P \rho^c d\rho$$

Computational Considerations:

- Monomials correspond to certain lattice points in a zonotope Z of dimension rank(A).
- In fact, these points are the image of the lattice points of the hypercuboid $\prod_v [0, U_v]$ under the linear transformation A.

Exact Formula for the Integral

Formula:

$$\int_{\Theta} L_U(\sigma, \theta, \rho)\ d\sigma d\theta d\rho = z \cdot \sum_{b \in Z} \phi(b) \int_{\Delta_1} \sigma^{(b,c)/a} d\sigma \int_P \theta^b d\theta \int_P \rho^c d\rho$$

Computational Considerations:

- Bottleneck is in computing $\phi(\cdot)$

  Naive method:

$$\phi_A(b, U) \ = \ \sum_{Ax=b} \prod_{v=1}^{n} \binom{U_v}{x_v}$$

  Instead, use recurrence formula:

$$\phi_A(b, U) \ = \ \sum_{x_n=0}^{U_n} \binom{U_n}{x_n} \phi_{A \setminus a_n}(b - x_n a_n, U \setminus U_n)$$

  Exploit low rank of A to store $\phi(\cdot)$ efficiently.

Exact Formula for the Integral

Formula:

$$\int_{\Theta} L_U(\sigma, \theta, \rho)\ d\sigma d\theta d\rho = z \cdot \sum_{b \in Z} \phi(b) \int_{\Delta_1} \sigma^{(b,c)/a} d\sigma \int_P \theta^b d\theta \int_P \rho^c d\rho$$

Computational Considerations:

- Only need to sum half the terms because of symmetry.
- Precompute and look-up values of factorials.
- Computation is highly parallelizable.
- Maple library:

  `http://math.berkeley.edu/~shaowei/integrals.html`

A Maple Demo

A Maple Demo

	Time(seconds)	Memory(bytes)
Ignorant Integration	16.331	155,947,120
Naive Expansion	0.007	458,668

	Time(minutes)	Memory(bytes)
Naive Expansion	43.67	9,173,360
Fast Integral (m=1)	1.76	13,497,944

Approximations of the Integral

Question:

Suppose $U = NY$ where Y is a fixed vector with $\sum_v Y_v = 1$.

As $N \to \infty$, how does the log marginal likelihood behave?

$$\log \int_\Theta L_U(\theta) d\theta$$

Approximations of the Integral

Question:

Suppose $U = NY$ where Y is a fixed vector with $\sum_v Y_v = 1$.

As $N \to \infty$, how does the log marginal likelihood behave?

Answer 1:

$$\log \int_\Theta L_U(\theta) d\theta \to -\infty$$

.

Approximations of the Integral

Question:

Suppose $U = NY$ where Y is a fixed vector with $\sum_v Y_v = 1$.

As $N \to \infty$, how does the log marginal likelihood behave?

Answer 2: BIC Score

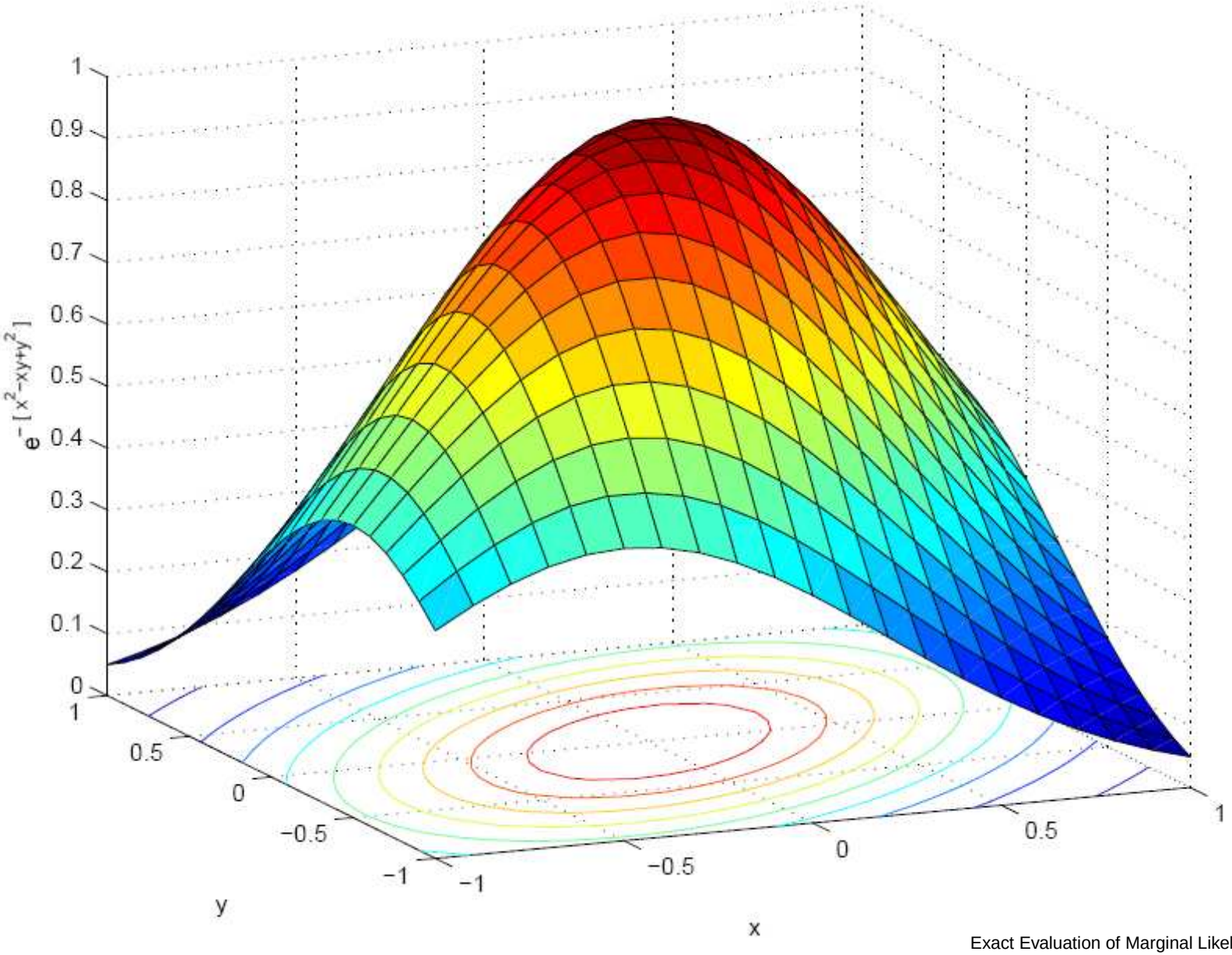
$$\log \int_\Theta L_U(\theta) d\theta \quad \approx \quad \log L(\hat{\theta}) - \frac{d}{2} \log N$$

where d is the dimension of the model and $L(\hat{\theta})$ is the *maximum* likelihood.

BIC stands for Bayesian Information Criterion.

Assumes that the model is in the exponential family. In particular, the model has one local maxima. As $N \to \infty$, the "main bulk" of the integral accumulates near the maximum likelihood.

Approximations of Integral



Approximations of the Integral

Question:

Suppose $U = NY$ where Y is a fixed vector with $\sum_v Y_v = 1$.

As $N \to \infty$, how does the log marginal likelihood behave?

Answer 3: Laplace Approximation

$$\log \int_\Theta L_U(\theta) d\theta \quad \approx \quad \log L(\hat{\theta}) - \frac{1}{2} \log |\det H(\hat{\theta})| + \frac{d}{2} \log 2\pi$$

where H is the Hessian of the log-likelihood function $\log L$.

Only assumes that L is twice differentiable, convex and achieves maximum on internal point.

Back to the Coin Tosser

- **Maximum Likelihood**

  Model One: $0.1443566234 \times 10^{-54}$

  Model Two: $0.1395471101 \times 10^{-18}$

Back to the Coin Tosser

- **Maximum Likelihood**

  Model One: $0.1443566234 \times 10^{-54}$

  Model Two: $0.1395471101 \times 10^{-18}$

- **Marginal Likelihood**

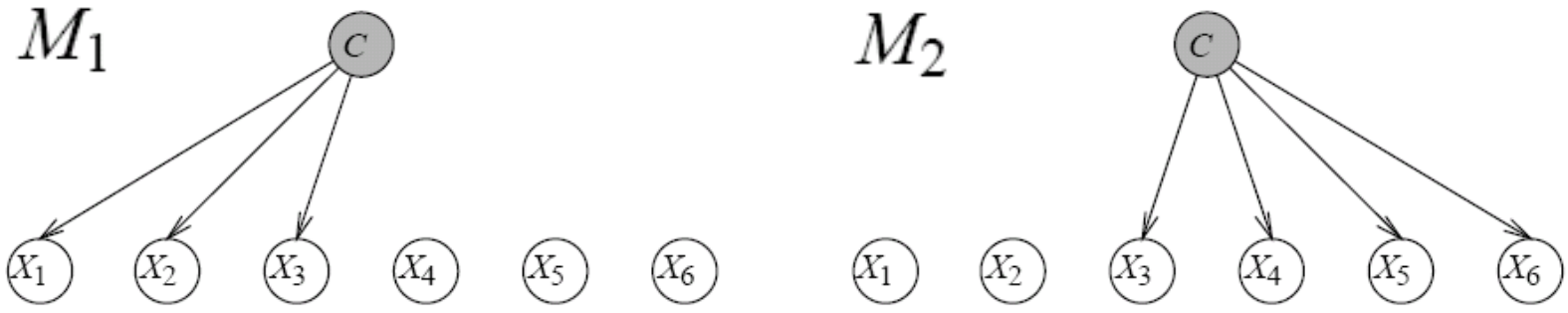
  Model One: $0.5773010423 \times 10^{-56}$

  Model Two: $0.7788716339 \times 10^{-22}$ (Actual)

  $0.3706788423 \times 10^{-22}$ (BIC)

  $0.4011780794 \times 10^{-22}$ (Laplace)

BIC can be wrong!

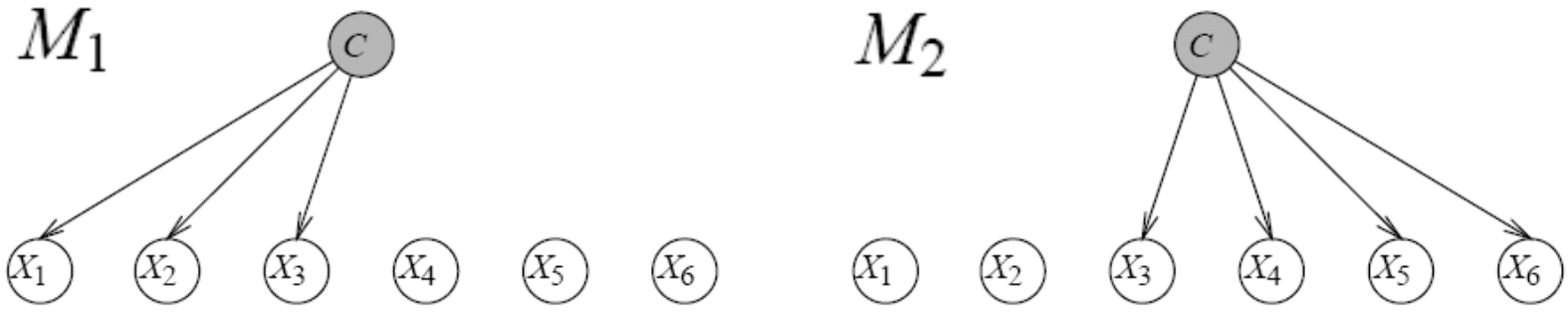


Consider the two hidden binary tree models above.

M1: $$p_v = (\sigma_0 \theta_{v_1}^{(1)} \theta_{v_2}^{(2)} \theta_{v_3}^{(3)} + \sigma_1 \rho_{v_1}^{(1)} \rho_{v_2}^{(2)} \rho_{v_3}^{(3)}) \theta_{v_4}^{(4)} \theta_{v_5}^{(5)} \theta_{v_6}^{(6)}$$

M2: $$p_v = \theta_{v_1}^{(1)} \theta_{v_2}^{(2)} (\sigma_0 \theta_{v_3}^{(3)} \theta_{v_4}^{(4)} \theta_{v_5}^{(5)} \theta_{v_6}^{(6)} + \sigma_1 \rho_{v_3}^{(3)} \rho_{v_4}^{(4)} \rho_{v_5}^{(5)} \rho_{v_6}^{(6)})$$

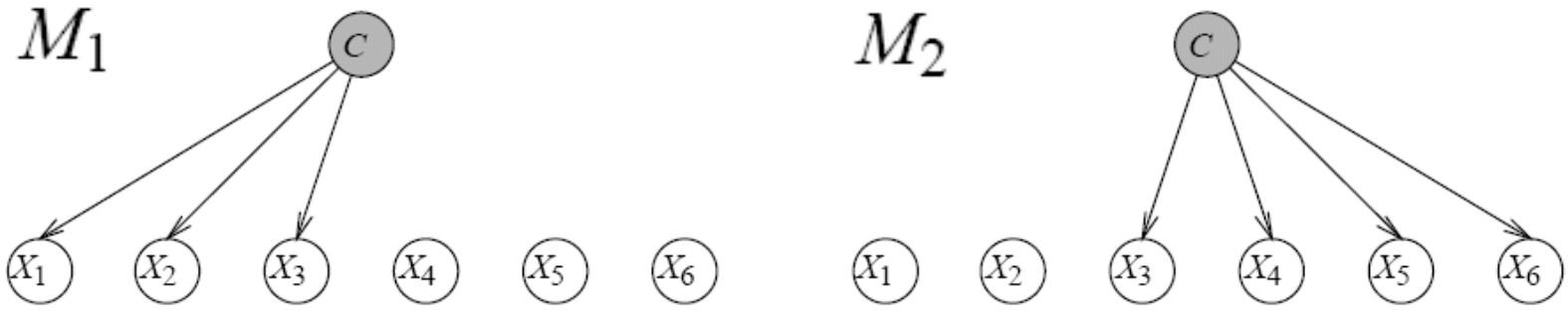
BIC can be wrong!



Suppose the data for sample size $N = 36$ is

		$X_4X_5X_6$							
		000	001	010	011	100	101	110	111
$X_1X_2X_3$	000	2	3	0	1	3	5	1	1
	001	0	0	0	0	0	0	0	0
	010	0	0	0	0	0	1	0	0
	011	0	0	0	0	0	0	0	0
	100	3	4	1	1	2	3	1	1
	101	0	1	0	0	0	0	0	0
	110	1	1	0	0	0	0	0	0
	111	0	0	0	0	0	0	0	0

BIC can be wrong!



Model Selection:

- BIC Score: M1's score is better than M2's.
- Actual Marginal Likelihood:

M1

$$\frac{26736202573582791008019248300635714612982861895953897913266720923361652444310905663581365769429178055600000000}{} \approx 0.449 \times 10^{-17}$$

M2

$$\frac{48293401975547884279365197096430603703508201757248809211637315169}{873248402971499818328286563178459524881596589864311287443444152295294483200000000} \approx 0.553 \times 10^{-17}$$

Thus, the BIC score will lead a Bayesian to choose the wrong model!

Summary

The Occasionally Dishonest Coin-Tosser

Marginal Likelihood Integrals
Mixtures of Independence Model
Exact Formula for the Integral
Approximations of the Integral

Comparing approximations for coin toss example.
Some approximations can lead to wrong model selection!

Future work

- Develop faster algorithms for exact evaluation of integral.
- Develop more accurate approximations using algebraic geometric tools.

Schizophrenic Patients

Evans, Gilula and Guttman:
studied association between length of hospital stay (Y) and frequency of visits by relatives.

	$2 \leq Y < 10$	$10 \leq Y < 20$	$20 \leq Y$	$Totals$
Visited regularly	43	16	3	62
Visited rarely	6	11	10	27
Visited never	9	18	16	43
$Totals$	58	45	29	$\mathbf{132}$

Equivalent to our models, for $k = 2, s_1 = s_2 = 1, t_1 = t_2 = 2$ and $N = 132$.

Schizophrenic Patients

- “each estimate requires a 9-dimensional integration”
- “the dimensionality of the integral does present a problem”
- “all posterior moments can be calculated in closed form .... however, even for modest N these expressions are far to complicated to be useful”

Schizophrenic Patients

The integral evaluates to

$$\frac{27801948853106338912064360032498932910387614080 5}{28524283958209256935726588667532284587409752803 3}$$

$$\frac{99493069713103633199906939405711180837568853737}{12288402873591935400678094796599848745442833177572204}$$

$$\frac{50448819979286456995185542195946815073112429169997801}{33503900169921912167352239204153786645029153951176422}$$

$$\frac{43298328046163472261962028461650432024356339706541132}{34375318471880274818667657423749120000000000000000.}$$

Time taken: 13 days on a modest laptop.

References

1. D.M. Chickering and D. Heckerman: Efficient approximations for the marginal likelihood of Bayesian networks with hidden variables, *Machine Learning* **29** (1997) 181-212; Microsoft Research Report, MSR-TR-96-08.
2. D. Geiger and D. Rusakov: Asymptotic model selection for naive Bayesian networks, *Journal of Machine Learning Reseach* **6** (2005) 1–35.
3. S. Hoşten, A. Khetan and B. Sturmfels: Solving the likelihood equations, *Foundations of Computational Mathematics* **5** (2005) 389–407.
4. L. Pachter and B. Sturmfels: *Algebraic Statistics for Computational Biology*, Cambridge University Press, 2005.
5. M. Evans, Z. Gilula and I. Guttman: Latent class analysis of two-way contingency tables by Bayesian methods, *Biometrika* **76** (1989) 557–563.