Win Nodes for Factor-Structured Gaussian Bayes Nets

Peter Cotton

First version: September 1, 2026.
This version: November 9, 2026.

Abstract

Order-statistic queries and evidence, such as which of several correlated variables is largest or who won a contest, have long resisted exact treatment in Bayes nets. For correlated Gaussian variables the probability that one is the maximum is an orthant probability, which Bayesian optimization and reinforcement learning call intractable and approximate, and which rating systems reach only through expectation propagation. We show that when the dependence among Gaussian variables runs through a few shared factors, the probability that each variable is the largest can be computed deterministically, at fixed factor rank and quadrature resolution, at a cost linear in the number of variables, by a recent algorithm for correlated contests, and that the same computation gives the derivatives of these probabilities, the likelihood of an observed winner, and the likelihoods of a quoted price and of a published place. Under a posterior, the probabilities are the action probabilities of Thompson sampling.

1 Introduction

The probability that one of several correlated Gaussian quantities is the largest has a reputation for being intractable. Tarbouriech et al. (2023) call it the key object of Bayesian reinforcement learning and say that computing it "involves computing several complicated integrals with respect to the posterior and is intractable in most cases"; their Section 4 opens with "computing this probability is intractable in general". Hennig and Schuler (2012) need the same probability over the candidate points of Bayesian optimization. They too call it intractable, citing Plackett (1954), and approximate it by expectation propagation at a cost of $O(n^4)$, which limits them to about fifty candidates.

Using an observed ranking or winner as evidence has also relied on approximation. TrueSkill (Herbrich et al., 2007) writes the likelihood of a multi-player ranking as a product of pairwise comparisons between adjacent finishers. That factorization is exact, but the posterior it produces is not Gaussian, and TrueSkill approximates it by expectation propagation. The general-purpose alternative is simulation. GHK is the standard sampler for Gaussian orthant probabilities, and Ridgway (2016) shows that on an AR(1) covariance its relative variance grows exponentially with the dimension. In practice, bandit algorithms, ranking-and-selection procedures and rating systems approximate these probabilities from the marginal means and variances, or estimate them by counting posterior draws.

The problem is easy to state, and the difficulty is computational. For jointly Gaussian $X = (X_1, \ldots, X_n)$ with independent components, $p_i = \Pr(X_i = \max_j X_j)$ is a one-dimensional integral, the random-utility representation of Domencich and McFadden (1975). With correlated

components it becomes an orthant probability in $n-1$ dimensions, and no earlier method evaluates all n of them deterministically in time linear in n.

We show that for factor-structured Gaussian Bayes nets this can be done. We add to the Gaussian variables a *win node*, the deterministic node $W = \arg\max_i X_i$ with values in $\{1, \dots, n\}$, and ask for what any node in a graphical model must supply: its distribution given its parents, the derivative of that distribution with respect to the parents' means, and the likelihood of an observed value. When the dependence among the X_i runs through a few shared factors, the lattice algorithm of Cotton (2026) for correlated contests gives all n probabilities $\Pr(W = i)$ at once, and Section 3 derives from the same computation the Jacobian, the exact likelihood of an observed winner, and the likelihoods of a quoted price and of a published place. The cost is linear in n at fixed numerical resolution.

None of these queries is answered by the univariate marginals alone. Which alternative is best, which sensor reads the extreme, and which entrants make a shortlist all depend on the joint distribution. Section 1.1 gives a network in which every marginal is the same and the answer is far from uniform.

In applications X is usually a posterior. When $X = \theta \mid \mathcal{D}$ is the belief about the values of n alternatives after seeing data $\mathcal{D}$, p_i is the posterior probability that alternative i is the best, which is also the probability that Thompson sampling chooses i. Section 4 shows that two standard models, simulation under common random numbers and a bandit with a factor prior, keep their posteriors in the factor form of Section 3 after every update, so the win node applies to them without approximation. Section 6 uses it to stop a simulation study and to check it against Thompson sampling on a correlated bandit.

1.1 A network with identical marginals

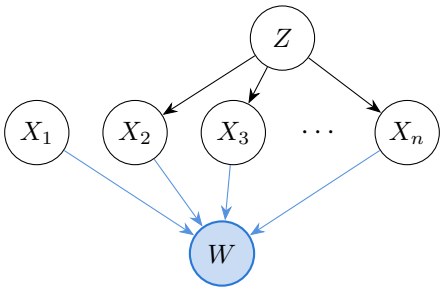


Figure 1: The Bayes net of Section 1.1. Every X_i is $N(0,1)$. A black arrow is a Gaussian conditional distribution; the blue arrows make $W = \arg\max_i X_i$ a deterministic child of every X_i. The latent Z is the parent of $X_2, \dots, X_n$ and X_1 has no parent.

Figure 1 shows a Gaussian Bayes net with a win node W added as a child of all of its variables. Nothing is observed, and the query is the distribution of W: for each i, the probability p_i that X_i is the largest. The net is built so that every X_i is $N(0,1)$. A method that sees only these marginals, for example one that treats the X_i as independent, must give every i the same probability $1/n$. The net shows how far the true distribution of W can be from that.

In Figure 1 the variables share one latent parent. $Z \sim N(0,1)$ is the parent of $X_2, \dots, X_n$, with $X_i \mid Z \sim N(\sqrt{1-\epsilon}\, Z, \epsilon)$, and $X_1 \sim N(0,1)$ has no parent. Read the X_i as a year of returns on n funds: funds $2, \dots, n$ track the same index Z with tracking variance ϵ, fund 1 follows its own strategy, and the query asks which fund returns the most.

When ϵ is small the index funds rise and fall together, so the contest is close to one between two competitors, fund 1 and the index. As $\epsilon \to 0$, $p_1 \to 1/2$ and each index fund's $p_i \to 1/(2(n-1))$, so the total-variation distance between the true distribution of W and the uniform one tends to $1/2 - 1/n$. With $n = 100$ and $\epsilon = 10^{-4}$, $p_1 = 0.493$ against $1/n = 0.01$. If k independent funds take the place of fund 1, each of them is the best with probability tending to $1/(k+1)$, and the total-variation distance tends to $k/(k+1) - k/n$, which is 0.81 for $n = 100$ and $k = 9$.

2 Related methods

TrueSkill (Herbrich et al., 2007) is already a Bayes net with binary win nodes. Its evidence is a full ranking, whose likelihood factorizes exactly into greater-than factors on the performance differences of adjacent finishers; its approximation is the Gaussian moment matching of the messages. Observing only the winner is different evidence. Section 3.3 gives its likelihood, which is the multinomial probit likelihood. At $n = 2$ it is the binary probit comparison with no draw margin, whose variance is the sum of the two performance variances.

Gaussian message-passing systems and GMRF libraries compute marginals, free energies and samples for very large models, and offer no order-statistic queries. For latent Gaussian fields, Bolin and Lindgren (2015) compute excursion sets and exceedance probabilities by sequential importance sampling; those functionals are outside this paper.

GHK (Geweke, 1989; Hajivassiliou et al., 1996) is the standard method for the orthant probabilities behind p_i, and Cotton (2026) compares it in detail with the factor computation used here. GHK writes one orthant probability as a nested sequence of truncated normal probabilities and simulates it, at $O(n^2)$ per draw for a single alternative.

Within the scope of this paper, factor dependence, the win node improves on it in three respects. It gives all n probabilities in one pass, where GHK needs n runs. Its cost is linear in n. The cost of all n probabilities by GHK grows empirically as $n^{2.8}$ beyond $n = 200$, where Cotton (2026) already measures a gap of two orders of magnitude.

With fixed quadrature nodes it is deterministic: the probabilities and their Jacobian are a reproducible numerical map, with integration error but no simulation noise. Randomized nodes such as scrambled Sobol bring back an error that varies with the scramble, and the logarithm of an unbiased randomized estimate, which is what a log-likelihood needs, is biased downward. Outside this scope, for dense high-rank covariance where accuracy is paramount, Cotton (2026) recommends GHK.

For independent components, Menet et al. (2025) write p_i as the one-dimensional integral of Equation (3) without factors (their Equation 2), and their Proposition 1 evaluates it by quadrature on a shared set of integration points with a convergence guarantee. Equation (4) is that evaluator applied conditionally on the factors and averaged over them.

3 The win node on factor-structured parents

This section derives the three parts of the win node when its parents have factor structure: the forward query, which gives all n probabilities $\Pr(W = i)$ (Section 3.1); their derivative with respect to the means (Section 3.2); and the likelihood when the winner is observed (Section 3.3). All three come out of one computation. Conditional on the factors the components are independent, so each probability is a one-dimensional integral of one component's density against the product of the others' distribution functions. The algorithm of Cotton (2026) builds that product once

on a lattice at each factor node and reads every alternative's probability off it by dividing out that alternative's own term; the derivatives and the likelihood of an observed winner are further passes over the same arrays. The cost is linear in n.

3.1 The forward query

This subsection and the two that follow restate the forward query, its Jacobian and the winner likelihood of Cotton (2026) in the notation of a win node. The new material in this section is the evidence of Sections 3.4 to 3.6.

Let the parents have factor form:

$$X_i \;=\; m_i + b_i' z + \varepsilon_i, \qquad z \sim N(0, I_\rho), \quad \varepsilon_i \sim N(0, d_i), \tag{1}$$

with all the ε_i independent of each other and of z. The covariance is

$$\Sigma \;=\; BB' + \operatorname{diag}(d), \qquad B \in \mathbb{R}^{n\times\rho}, \quad \rho \ll n, \tag{2}$$

where B has rows b_i'. The variables in z are the factors and B holds the loadings. Assume every $d_i > 0$, so the maximum is almost surely unique and the p_i sum to one.

The shared factors make the components dependent. Given z, nothing is shared, and the components are independent with means $m_i + b_i' z$ and variances d_i. The probability that i is the maximum given z is then a one-dimensional integral, and p_i is its average over z. Let $F_{i|z}$ and $f_{i|z}$ be the conditional distribution and density functions of X_i given z. Component i is the maximum when every other component falls below it, so

$$p_i \;=\; \int_{\mathbb{R}^\rho} \phi_\rho(z) \int_{\mathbb{R}} f_{i|z}(x) \prod_{j\neq i} F_{j|z}(x)\, dx\, dz. \tag{3}$$

Given z, the inner integral is the probability that component i finishes at some value x with every other component below it; the outer integral averages over the factors.

The inner integrands for different i share the product $G_z(x) = \prod_j F_{j|z}(x)$. Dividing out the ith term gives every p_i from the same product:

$$p_i \;=\; \mathbb{E}_z \int f_{i|z}(x)\, \frac{G_z(x)}{F_{i|z}(x)}\, dx. \tag{4}$$

On a lattice of L points in x, with Q quadrature nodes in z, building $\log G_z$ once and subtracting $\log F_{i|z}$ for each i gives all n probabilities in $O(nLQ)$. The accuracy depends on L, on the lattice window and on Q, and the Q needed grows with the number of factors and the geometry of B. The same conditional independence gives top-k membership, with the product replaced by the distribution of the number of components above x. Top-k nodes, with their derivatives, the likelihood of an observed top-k set, are the subject of work in progress.

Only differences between loadings matter. Let $P = I - \mathbf{1}\mathbf{1}'/n$ and write $Bz = (PB)z + \mathbf{1}(\bar{b}'z)$, where $\bar{b}$ is the average row of B. The second term shifts every component by the same amount and changes no ranking, so $p(m, B, d) = p(m, PB, d)$. The number of factor dimensions the integral has to resolve is the rank of PB, which centering can reduce below the rank of B. The exact probabilities are also unchanged by any orthogonal rotation of the columns of B. The numerical properties of the quadrature, including its dependence on orientation, are treated in Cotton (2026).

The lattice in x and the quadrature over z have finite resolution, so the evaluation of Equation (4) has error. Section 5 reports the measured errors.

3.2 The derivative

The derivative of p_i with respect to m_j, for $j \neq i$, is minus the density of the event that i and j tie for the maximum:

$$\frac{\partial p_i}{\partial m_j} \;=\; -\,\mathbb{E}_z \int f_{i|z}(x)\, f_{j|z}(x) \prod_{k \neq i,j} F_{k|z}(x)\, dx, \qquad j \neq i, \tag{5}$$

and $\partial p_i / \partial m_i$ is the sum of the others with the sign reversed, because adding the same constant to every mean leaves p unchanged. The Jacobian is therefore a weighted graph Laplacian whose edge weights are tie densities (Cotton, 2026, §5), and it is positive semidefinite, as it must be: p is the gradient of $\mathbb{E}[\max_i X_i]$ with respect to m, and the expected maximum is convex in m. Its product with a vector costs $O(nLQ)$ on the same lattice, without forming the matrix.

Müller et al. (2019) reach the same identity for general joint laws through the surplus function. Equation (5) is the derivative of the exact probabilities. The derivative of the computed lattice probabilities also carries their normalization and, on an adaptive grid, the movement of the grid, so it need not be exactly symmetric or have zero row sums (Cotton, 2026, §§3.2, 5.2).

3.3 The likelihood of an observed winner

A node in a Bayes net must also accept evidence. The likelihood of the observation $W = k$ is the probability, under the parents, that alternative k is the largest. Write $u \sim N(0,1)$ for the winner's own standardized noise, independent of z. Conditional on z and u the winner's performance is fixed, each rival must finish below it, and the rivals are independent, so the conditional probability is a product of $n-1$ normal distribution functions. The likelihood is the average of that product over z and u:

$$\Pr(W = k \mid m) \;=\; \mathbb{E}_{z,u} \prod_{j \neq k} \Phi\left(\frac{m_k - m_j + (Bz)_k - (Bz)_j + u\sqrt{d_k}}{\sqrt{d_j}} \right). \tag{6}$$

The score of Equation (6), its gradient in m, is a weighted sum of Mills ratios over the quadrature nodes, the weight of each node being its posterior probability given the win. It is computed in one extra pass over the arrays the likelihood already builds, so the message an observed win node passes to its Gaussian parents carries exact first-order information. At $n = 2$ with no factors the likelihood is $\Phi\big((m_k - m_j)/\sqrt{d_k + d_j}\big)$, the binary probit comparison, which is TrueSkill's greater-than factor with no draw margin.

Observing the winner tells less than observing the full ranking, and Equation (6) is the likelihood of that smaller piece of evidence. The likelihood is exact; a moment approximation enters only if the user then projects the posterior onto a Gaussian family, as expectation propagation does. Equation (6) is the multinomial probit likelihood, and with its score it is the estimation core of Cotton (2026, §5).

The implementation used here agrees with finite differences and with dual-number automatic differentiation to 10^{-10}. That checks the differentiation of the implemented numerical likelihood, not its accuracy against the continuum likelihood.

3.4 A quoted price as evidence

The second kind of evidence a win node meets in practice is a price: a market's probability that each alternative wins, quoted before the outcome is known. A price is soft evidence about the parents, and the inverse of the forward query turns it into a Gaussian observation.

Model the market as pricing the same race from its own noisy view of the means, $p = \text{race}(m + \eta)$ with $\eta \sim N(0, \tau^2 I)$, where race is the forward query of Section 3.1 under the market's loadings and noise. The inversion of Cotton (2026, §5) recovers $m + \eta$ from p up to one common level, because the prices sum to one and so carry $n - 1$ numbers, not n. The observation therefore lives on the contrast space, $y = P(m + \eta)$ with $P = I - \mathbf{1}\mathbf{1}^\top/n$, and it is linear and Gaussian in m.

With a belief $N(\mu, \operatorname{diag} v)$ the posterior precision is $\operatorname{diag}(1/v) + P/\tau^2$, a rank-one correction to a diagonal, and the update is in closed form.

A price determines the differences between the entrants' means and nothing else. If every entrant is priced at one in eight, the prices are the same whether every mean is high or every mean is low, so the price cannot move the average of the means, which we call the level. When all prior variances are equal, the update leaves the average of the posterior means equal to the average of the prior means. When the prior variances differ, the average changes, and the reason is the prior, not the price.

For example, let entrant 1 have a mean known exactly and entrant 2 a mean with a wide prior. A price that makes the two equally likely to win implies that entrant 2's mean equals entrant 1's. The update moves entrant 2's mean to entrant 1's and leaves entrant 1's where it was, so the average moves.

The price model adds one parameter, the market noise τ^2. To estimate it, take a history of races with prices, compute the likelihood of all the observed prices as a function of τ^2, and choose the τ^2 that maximises it; Cotton (2026, §5) does this. A market priced under a different model from the outcome, or with a longshot bias, is inverted under its own model and then read as evidence under this one; the inversion is agnostic about where the prices came from.

3.5 Combining a price with the outcome

A win node can now take two kinds of evidence about the same race, a price before it and the outcome after it. Two questions follow. How much is each worth on its own? And does using both improve on either, as it should if the two are conditionally independent given the means? A simulation measures the posterior after each kind of evidence, and after both, against the truth, as follows.

Prices are set before the race and the outcome is observed after it, and the two are conditionally independent given the means, so the price update is applied first and the observed-winner message of Equation (6) is applied to the updated belief.

We simulate 400 races. In each race eight alternatives have means drawn from $N(0, I)$ and share one factor; a market quotes a price for each alternative from its own noisy view of the means, with $\tau^2 = 0.25$; the race is run and the winner recorded.

Starting from the prior $N(0, I)$ we form three posteriors, one from the price, one from the winner, and one from the price followed by the winner, and compare each with the truth in two ways: the root mean square error of the posterior-mean contrasts against the true contrasts, and the log probability the posterior assigns to the winner of a second, independent running of the same race. Table 1 reports the averages over the 400 races.

Table 1 answers both questions. The price is worth far more than one observed winner, which is a single draw from the outcome, and the posterior after both is tighter and scores better than after either alone. Between messages the belief is kept in the diagonal Gaussian family: the price posterior is exact and its diagonal is carried forward, and the winner message is projected to a

evidence	contrast RMSE	posterior variance	log score
none (prior)	0.919	1.000	−2.062
winner only	0.854	0.872	−1.956
price only	0.402	0.300	−1.671
price, then winner	0.399	0.287	−1.656

Table 1: One race, two kinds of evidence, averaged over 400 seeded races with market noise $\tau^2 = 0.25$. Script: `research/win_nodes/exp1_two_evidence`.

diagonal Gaussian as in Section 3.3, so the composed row is a moment-matched posterior, not the exact one.

3.6 Places as evidence

Published results usually lie between the winner and the full order: the first k in order with the rest unplaced, the first k as a set, or the fact that a named entrant finished among the first k. Each is an event about the order of the same performances. Conditional on the shared factors each has a one-dimensional lattice form, so each passes through the same mixture over factor nodes as Equation (6).

Write f_i and Φ_i for the conditional density and distribution function of entrant i's performance, and $S_i = 1 - \Phi_i$. For the set T of the first k in unknown order, let $G = \prod_{j \notin T} \Phi_j$ be the distribution function of the best outsider and $H = \prod_{i \in T} S_i$ the probability that every member is above x. Then

$$\Pr(\text{first } k = T) \;=\; \int G'(x)\, H(x)\, dx \;=\; -\int G(x)\, H'(x)\, dx. \tag{7}$$

Differentiating either form gives the score as one lattice sum per entrant, nonnegative for a member and nonpositive for an outsider; the sums total zero because a common shift of every mean changes nothing.

For the ordered prefix $i_1, \ldots, i_k$ with everyone else behind, let $N = \{i_1, \ldots, i_k\}$ be the named entrants, let $A_k(x) = f_{i_k}(x) \prod_{l \notin N} \Phi_l(x)$ be the density that the last named entrant finishes at x with every unnamed entrant below, and recursively let $A_j(x) = f_{i_j}(x) \int_{-\infty}^{x} A_{j+1}$ be the density that entrant i_j finishes at x ahead of the rest of the prefix. The probability of the prefix is $\int A_1$, which is k cumulative sums on one lattice, for any k.

The membership of one entrant among the first k is the top-k query of Cotton (2026) and its Jacobian row is the score. At $k = 1$ all three are the observed winner, and the prefix at $k = n$ is the full order; both reductions are checked to the tolerance of the existing updates, and every message is checked against rejection-sampled posteriors of the generative model.

Published results are usually partial: a podium, a cut, a winner. The question is how much of the information in the full order each partial result keeps, and the messages above let us measure it. Table 2 does so in the simulation of Table 1, with 1,600 seeded races and one message per race. Taking the full order's reduction in contrast error as the whole, the first three in order with the rest behind keep 71 percent of it, the first three as a set 62 percent, the winner alone 34 percent, the order among three named entrants 22 percent, and the membership of one named entrant in the first three 15 percent. Every gain is at least fifteen paired standard errors. The named entrant is drawn uniformly from the first three and the three ordered entrants uniformly from the field, so the choice of who is named carries no information.

evidence	contrast RMSE	posterior variance	log score
none (prior)	0.896	1.000	−2.066
one named entrant in the first three	0.867	0.947	−2.030
order among three named	0.854	0.921	−2.015
winner only	0.830	0.872	−1.967
first three as a set	0.775	0.775	−1.918
first three in order, rest behind	0.759	0.749	−1.891
full order	0.702	0.656	−1.841

Table 2: One race, one outcome message each, ranked by information. Script: `research/win_nodes/exp2_place_evidence`.

4 Posteriors of factor form

The win node of Section 3 needs its parents in the factor form of Equation (2). In the applications we care about, the parents are a posterior over the values of n alternatives, and that posterior changes with every observation. The win node can be applied after every update without approximation only if the posterior stays in factor form. This section shows that it does in two standard settings: a simulation study run under common random numbers, and a bandit whose arms share features. In both, the posterior has no more factors than the source of the shared variation, however many alternatives there are, and the $n \times n$ posterior covariance is never formed. Section 6 uses both posteriors.

4.1 Simulation with common random numbers

The first setting is ranking and selection by simulation. The question there is which of n systems is best, which is the win node's query, and a stopping rule needs the probability that each system is best (Section 6.2). The posterior is correlated by design: common random numbers are used precisely so that the systems' outputs move together and their differences are estimated more sharply. We show that this correlated posterior has factor form.

A simulation study compares n systems, such as candidate inventory policies, by running each one many times and averaging its output. Each run is a replicate, and the unknown θ_i is system i's long-run average output. Under common random numbers every system receives the same random inputs in each replicate. A high-demand history raises every policy's profit, so the outputs move together, and the differences between systems are measured more precisely than their levels.

Summarise the shared random inputs of replicate r by ρ variables F_r, such as the demand level and its volatility. The analyst does not record F_r. System i responds to it with known loadings v_i and has its own noise ε_{ir}:

$$Y_{ir} \;=\; \theta_i + v_i' F_r + \varepsilon_{ir}, \qquad F_r \sim N(0, I_\rho), \quad \varepsilon_{ir} \sim N(0, \sigma_i^2). \tag{8}$$

The F_r and ε_{ir} are independent of each other and across replicates, and the noise scales $\sigma_i > 0$ are known. If F_r were recorded, subtracting $v_i' F_r$ from each output would remove the correlation and leave a diagonal posterior. The prior is $\theta \sim N(0, s_0^2 I)$ with $s_0 > 0$.

The covariance of one replicate across systems is $\Lambda = VV' + D$ with $D = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_n^2)$.

After $n_r \geq 1$ replicates with mean vector $\bar{Y}$, the posterior of θ is Gaussian with precision

$$A \;=\; s_0^{-2} I + n_r \Lambda^{-1}. \tag{9}$$

The Woodbury identity gives

$$\Lambda^{-1} = D^{-1} - D^{-1} V M^{-1} V' D^{-1}, \qquad M = I_\rho + V' D^{-1} V. \tag{10}$$

With $U = D^{-1}V$ and $S = n_r M^{-1}$, the precision is

$$A = \mathrm{diag}(a) - USU', \qquad a_i = s_0^{-2} + n_r \sigma_i^{-2}, \tag{11}$$

a positive diagonal minus a positive semidefinite term of rank at most ρ. The Woodbury identity again gives

$$A^{-1} = \mathrm{diag}(a)^{-1} + \mathrm{diag}(a)^{-1} U\, C\, U' \mathrm{diag}(a)^{-1}, \qquad C = \left(S^{-1} - U' \mathrm{diag}(a)^{-1} U\right)^{-1}. \tag{12}$$

C is positive definite. Let $K = \left(\begin{smallmatrix} \mathrm{diag}(a) & U \\ U' & S^{-1} \end{smallmatrix}\right)$. The Schur complement of S^{-1} in K is A, which is positive definite, and S^{-1} is positive definite, so K is positive definite. The Schur complement of $\mathrm{diag}(a)$ in K is C^{-1}, so C^{-1} is positive definite too. Hence

$$\Sigma_{\text{post}} \;=\; \mathrm{diag}(d) + BB', \qquad d_i = a_i^{-1}, \quad B = \mathrm{diag}(a)^{-1} U\, C^{1/2}, \tag{13}$$

exactly, for every $n_r \geq 1$.

The posterior mean is $m = A^{-1}h$ with $h = n_r \Lambda^{-1} \bar{Y}$, and both factors are available from Equations (10) and (12). No step is an approximation. The posterior has factor form because the shared random inputs have factor form.

4.2 A bandit whose arms share features

The second setting is a bandit whose arms are correlated through shared features, the model behind linear Thompson sampling. The probability that each arm is best is the probability that Thompson sampling pulls it (Section 6.1), and we show that the posterior stays in factor form after every pull.

In a bandit, a decision maker repeatedly picks one of n arms and observes a noisy reward. Here θ_i is arm i's mean reward, and pulling arm a returns $\theta_a + N(0, \sigma_n^2)$ with $\sigma_n > 0$. Arm i has known features v_i, and $\theta_i = v_i'\beta + e_i$ with a shared $\beta \sim N(0, I_\rho)$ and independent $e_i \sim N(0, d_{0i})$, $d_{0i} > 0$. The prior is $\theta \sim N(0, \Sigma_0)$ with $\Sigma_0 = VV' + \mathrm{diag}(d_0)$, which has factor form.

Linear Thompson sampling (Agrawal and Goyal, 2013), the version Chapelle and Li (2011) evaluated on display advertising and news recommendation, is this model without the residual, with factor rank equal to the number of features. A Gaussian-process prior with a kernel of finitely many features and an independent residual has this form exactly. A dense kernel does not, and an inducing-point approximation to one has a dense residual (Titsias, 2009), so for those a diagonal residual is an added approximation.

With per-arm pull counts c, the posterior precision is $A = \Sigma_0^{-1} + \sigma_n^{-2}\,\mathrm{diag}(c)$. Equation (10) applied to Σ_0 puts A in the form of Equation (11), with $a_i = d_{0i}^{-1} + c_i \sigma_n^{-2}$, $S = M_0^{-1}$ and $M_0 = I_\rho + V' \mathrm{diag}(d_0)^{-1} V$, and the rest of Section 4.1 applies unchanged. The posterior is exactly factor plus diagonal after every pull, with no more factors than the prior.

The same posterior has a form with no subtraction inside an inverse. Let $\lambda_i = c_i/\sigma_n^2$ be the observation precision accumulated on arm i, $r_i = 1/(1 + d_{0i}\lambda_i)$, $R = \mathrm{diag}(r)$ and $H = I_\rho + V'\mathrm{diag}(\lambda_i r_i)V$. Then

$$\Sigma_{\text{post}} \;=\; \mathrm{diag}(d_{0i}r_i) + RVH^{-1}V'R, \qquad B_{\text{post}} = RVH^{-1/2}. \tag{14}$$

H is the identity plus a positive semidefinite matrix, so it is positive definite. An arm never pulled has $\lambda_i = 0$, and arms observed with different noise levels enter through their own λ_i. Equation (14) agrees with direct inversion of the dense precision to machine precision, and is the form to implement.

5 Accuracy

All numbers in this section come from committed, seeded scripts in the public repository. Monte Carlo is used only as a check.

The accuracy of the factor computation is studied in Cotton (2026). On ten posteriors from the experiments of Section 6, raising the lattice from 257 to 2,049 points changes no entry by more than 4×10^{-16}, and replacing the default factor rule by 2^{15} scrambled-Sobol nodes changes no entry by more than 1.2×10^{-6}. Four million Monte Carlo draws differ from the computed probabilities by 1 to 4×10^{-4} per entry, which is their own standard error. No error bound is proved. The script is `research/rs_crn/exp3_quadrature`.

6 Applications

The three applications use the factor-form posteriors of Section 4 and concern the probability that each of several alternatives is the best: Section 6.1 relates it to Thompson sampling and to two approximations in current use, Section 6.2 uses it to stop a simulation study, and Section 6.3 checks it against Thompson sampling on a correlated bandit.

6.1 Thompson sampling probabilities and their approximations

With what probability will Thompson sampling show each of fifty advertisements next? Let $X = \theta \mid \mathcal{D}$ be a posterior over the values of n alternatives. Thompson sampling draws one sample $\tilde{\theta}$ and chooses the alternative with the largest sampled value, so it chooses i with probability $p_i = \Pr(W = i)$.

Lemma 8 of Tarbouriech et al. (2023) states the general version: the expected fraction of pulls each arm receives under Thompson sampling equals its posterior probability of being best. Choosing i with the computed probability $\hat{p}_i$ is therefore Thompson sampling up to the evaluation error of Section 5. The policy is still randomized. Only the evaluation of its probabilities is deterministic.

Thompson sampling itself never needs the numbers p_i. A stopping rule needs $\max_i p_i$. A shortlist needs the k entries most likely to be best, an allocator needs to know which p_i are close, a constraint on how often any arm may be shown needs the probabilities it constrains, and off-policy evaluation needs them as propensities. In each of these uses they are currently approximated, either by counting posterior draws or from the marginal means m_i and standard deviations $s_i = \sqrt{d_i + \|b_i\|^2}$. Two marginal approximations are compared below.

F-LITE (Menet et al., 2025) treats the alternatives as independent, $\prod_i N(m_i, s_i^2)$, and approximates the probability of being best by $\Phi((m_i - \kappa^*)/s_i)$, with a threshold κ^* chosen so that these sum to one. The threshold step is a second approximation. With two independent arms, $N(1,1)$ and $N(0,1)$, it gives 0.691 for the first arm, against the exact 0.760.

The *VBOS policy* is the variational Bayesian optimistic sampling rule of O'Donoghue and Lattimore (2021), extended by Tarbouriech et al. (2023). In the near-closed form of Menet et al. (2026) it is

$$\tilde{\pi}_i \;=\; v\!\left(\frac{m_i - \kappa^*}{s_i}\right), \qquad v(c) = \exp\!\Big(-\tfrac{1}{8}\big(\sqrt{c^2+4} - c\big)^2\Big), \tag{15}$$

with κ^* again chosen so the entries sum to one. The VBOS policy departs from p on purpose, and its regret analysis depends on that departure. Its distance from p matters because Menet et al. (2026) use it as the trainable substitute for p.

Both approximations ignore correlation, and the second example of Section 1.1 shows how large the resulting error can be. A simpler example fixes the mean gap and varies the size of a shared shock. Let $\theta_1 = 3 + Z + \varepsilon_1$ and $\theta_2 = Z + \varepsilon_2$, with $Z \sim N(0, \tau^2)$ and independent $\varepsilon_i \sim N(0,1)$. The shared Z cancels from the difference, so $p_1 = \Phi(3/\sqrt{2}) \approx 0.983$ for every τ. F-LITE sees only the marginal variances $1 + \tau^2$ and gives $\Phi\big(3/(2\sqrt{1+\tau^2})\big)$, which is 0.933 at $\tau = 0$, 0.682 at $\tau = 3$, and tends to one half as τ grows.

6.2 Stopping a simulation study

Has a simulation study run long enough to say which of a hundred inventory policies is best? The first experiment runs the simulation study of Section 4.1 and asks when it can stop. A study stops when it is confident enough that it has found the best system. The question is whether a stopping rule that uses the computed probabilities stops earlier than rules that use the marginal approximations, and whether it still selects the best system as often as its confidence level promises.

The setting is the posterior of Section 4.1 with $n = 100$ systems, $\rho = 2$ shared input streams and prior scale $s_0 = 1$. True means are drawn from the prior. In every regime each system's replicate output has variance one, and the *factor share* is the fraction of that variance that comes from the shared inputs. The posterior uses the outputs of all n systems jointly, so its marginal variances still differ between regimes: after 20 replicates each θ_i has posterior variance 0.048 in the independent regime and 0.027 in the aligned one. The comparison is between rules within a regime.

Correlation can mislead a marginal-only rule in two directions. If the leading systems move together, their difference is less uncertain than their marginals suggest, and the rule is too cautious. If they move oppositely, their difference is more uncertain, and the rule is too confident. The regimes produce each case, with an uncorrelated control.

- *Independent*: factor share zero. This is the control.
- *Aligned*: every system loads on the same stream, with factor share 0.6. Differences between systems then vary less than their marginal variances suggest. Every row of B is the same, so $PB = 0$ and the computed probabilities are those of an independent contest, which the marginal rules do not know. This is the shared shock example of Section 6.1 with a hundred systems.

- *Opposed*: half the systems load positively and half negatively on one stream, with factor share 0.6. Differences between systems in opposite halves then vary more than their marginal variances suggest.

Three stopping rules differ only in how they estimate p: the win node of Section 3, F-LITE and the VBOS policy. After every replicate a rule recomputes its estimate, and it stops the first time the largest entry reaches $1-\delta$, with $\delta = 0.05$, selecting that entry. A fourth rule, drawing M samples from the correlated posterior and counting argmax frequencies, uses the correlation and costs $O(Mn\rho)$. At $p = 0.95$, $M = 500$ draws give a standard error of 0.0097, and a 95% margin of 0.01 needs about 1,800. It was not run.

With the true probabilities the rule has a Bayesian guarantee. Let τ be the first time $\max_i p_i \geq 1-\delta$, let B_r be the budget of replicates, and on $\{\tau \leq B_r\}$ let I be the selected alternative and I^* the best. Then

$$\Pr(I \neq I^* \mid \tau \leq B_r) \;=\; \mathbb{E}\big[1 - p_I(\mathcal{D}_\tau) \mid \tau \leq B_r\big] \;\leq\; \delta. \tag{16}$$

Optional stopping does not affect this statement, which averages over the prior. It is not a frequentist guarantee for a fixed θ.

For computed probabilities with $\|\hat{p} - p\|_\infty \leq \eta$ at stopping, the error is at most $\delta + \eta$, and a threshold of $1 - \delta + \eta$ restores δ. At the ten posteriors measured in Section 5, $\|\hat{p} - p\|_\infty$ was about 10^{-6}. That is a measurement at those posteriors and not a bound at every stopping time.

The threshold is nominal, and its true confidence level differs by rule. With two independent arms of equal posterior standard deviation s, an entry of 0.95 requires a mean gap of $2.33\,s$ for the exact probabilities, $3.29\,s$ for F-LITE and $4.84\,s$ for the VBOS policy. The true posterior probability of a correct selection at those gaps is 0.950, 0.990 and 0.9997. Part of any extra sampling by the marginal rules is this higher effective target.

regime	rule	mean replicates	realized PCS	censored of 200
independent	win node	126.5	0.971	29
	F-LITE	188.7	1.000	57
	VBOS	252.9	1.000	88
aligned	win node	81.2	0.973	17
	F-LITE	175.2	1.000	53
	VBOS	236.7	1.000	90
opposed	win node	122.7	0.970	33
	F-LITE	188.6	0.993	64
	VBOS	244.2	1.000	92

Table 3: Stopping at nominal 95% confidence. Each regime has 200 replications and a budget of 400 replicates. Mean replicates is the mean of the stopping time capped at 400, so a censored run counts 400. Realized PCS is the fraction of stopped runs that selected the true best system. The stopped runs differ by rule.

In Table 3, every rule's realized PCS among its stopped runs is above nominal. The lowest, the win node rule's 0.970 in the opposed regime, has 167 stopped runs and a standard error of 0.013.

At this common nominal threshold and budget, the two marginal rules used 1.5 to 2.9 times the capped mean replicates of the win node rule, and hit the budget far more often. The gap is largest in the aligned regime, where the win node rule stops after 81.2 replicates on average, against 175.2 and 236.7, and stops in 183 of 200 runs, against 147 and 110. The thresholds are nominal and not calibrated to a common PCS, so these ratios compare spend at a common threshold. They do not compare spend at a matched true error.

The marginal estimates are far from the computed probabilities along the way. Before stopping in the aligned regime, F-LITE is at median total-variation distance 0.116 from them, with ninetieth percentile 0.174, and the VBOS policy is at 0.133, with ninetieth percentile 0.207. In the independent regime the distances are 0.063 and 0.095, because F-LITE's threshold step is biased even for an independent posterior and the VBOS policy departs from p by design. At the experiment's Monte Carlo checks, using 2×10^5 draws at twelve posteriors, the largest total-variation distance of the computed probabilities was 0.0026, which is the noise of the check.

6.3 A bandit with correlated arms

Does choosing arms from the computed probabilities reproduce Thompson sampling when the arms are correlated? On the bandit of Section 4.2, with $n = 50$ arms, horizon 200 and 100 replications, choosing arms with the computed probabilities gives cumulative Bayes regret within two percent of Thompson sampling's, in an independent regime and in one where each half of the arms shares a factor carrying 0.75 of the prior variance. The paired differences, -3.0 ± 12.0 and $+1.9 \pm 8.0$ at 95%, are consistent with zero, which is what Lemma 8 requires. Used as the policy, the VBOS approximation has ten to twelve percent more regret than Thompson sampling, a paired difference of 16.1 ± 10.6 and 14.7 ± 7.4, with its action distribution at total-variation distance 0.13 to 0.16 from the computed probabilities. F-LITE has three to six percent more, 5.7 ± 11.3 and 6.8 ± 7.2, at total-variation distance about 0.06. The script is `research/rs_crn/exp2_bandit`.

7 Software

The factor win node, its top-k relatives, both Jacobian blocks, the calibration inverses, and the evidence messages of Sections 3.3–3.6, all through one function (`update_race`: prices, winner, order, prefix, top set, membership), are in the `winning` package of Cotton (2026): a Python reference with R, JavaScript, Julia and Rust ports, each checked against the same embedded reference vectors, most to 10^{-15}. The observed win node, Equation (6) with its analytic score, is the estimation core of `MultinomialProbit.jl`, whose score is also checked by forward-mode automatic differentiation through the same code. Chain parents, which this paper does not treat, are handled by the classical absorbing-barrier recursion in `GMRFExtremes.jl`.

8 Scope

Structure. The reduction requires factor structure. Factor dependence of rank ρ costs a Q-node quadrature and is exact to quadrature accuracy. Chain and band parents, in which each variable is correlated with its neighbours rather than through shared factors, are a different structure: there the law of the maximum, the first-passage law and the argmax law are computed by the classical forward recursion for a Markov chain with an absorbing barrier on a lattice, implemented

in `GMRFExtremes.jl`, and this paper does not treat them. General sparse precision gets neither treatment, and the excursion functionals then belong to the sampling methods of Bolin and Lindgren (2015).

Rank. The error of the marginal approximations can approach one half with every marginal held fixed, whatever the factor rank. The comparison with counted posterior draws depends on rank. On the posteriors of the two experiments, where the centered loadings have rank at most two, one evaluation for a hundred alternatives takes about ten milliseconds and is accurate to 10^{-6} per entry.

Counting draws for ten milliseconds gives about twenty thousand draws, which estimate an entry near 0.01 to seven percent of itself. A standard error of 10^{-6} on an entry near 0.01 would take about 10^{10} draws.

At factor ranks three to five, on single hundred-arm posteriors of a linear sampler, one evaluation takes 0.3 to 1 second and 200,000 draws take 0.3 second, with errors of the same order. The win node is faster than counting at low rank, and the rank that matters is that of the centered loadings.

Exactness of the posteriors. The derivations of Section 4 are exact because the observations are generated with factor structure. When shared inputs enter a simulation nonlinearly, a low-rank model of Λ is an approximation, and the posterior inherits its error. A dense covariance can be fitted to factors, blocks, nested groups or a tree on intake (Cotton, 2026). The fit introduces model error, and its effect on the probabilities depends on the covariance of the contrasts and its conditioning: a small covariance residual can still move a nearly deterministic comparison a long way.

Allocation. The probabilities say how uncertain the choice of best is. They do not say where the next measurement should go. The knowledge gradient (Frazier et al., 2008) answers that question for one objective, the expected improvement in the value of the final choice, usually the largest posterior mean, and Tuisov (2026) treats it as the myopic value of computation. Reducing the entropy of p, which is entropy search, and raising $\max_i p_i$, which aims at correct identification under zero-one loss, are different objectives. Hennig and Schuler (2012) note that both are objectives of an optimizer. A bandit is scored on the values it collects.

Entropy search. On a finite set of representer points, the probability that the minimum of a Gaussian process lies at each point is p in the minimum convention. Hennig and Schuler (2012) approximate it by expectation propagation at cost $O(n^4)$, and their implementation uses fifty representer points. For a belief of factor form the win node gives the same probabilities in time linear in n, and the closure of Section 4 is the case in which the hypothetical posterior after a measurement stays in factor form. Entropy search also uses derivatives with respect to the covariance, which the win node does not supply, and this paper does not run it.

Top-k nodes. The win node is the $k = 1$ case of a top-k node. For factor parents the forward query and its derivatives are implemented in the `winning` package; the full treatment, including observed top-k sets, is work in progress.

Off-policy evaluation. The replay method of Li et al. (2011) needs a uniformly random log, because the sampler's own propensities were not available, and discards every logged event whose random arm differs from the policy's choice. With the propensities computed, the log of a correlated Thompson sampler can be used directly. This paper does not test that.

References

S. Agrawal and N. Goyal. Thompson sampling for contextual bandits with linear payoffs. In *ICML*, 2013.

D. Bolin and F. Lindgren. Excursion and contour uncertainty regions for latent Gaussian models. *J. R. Statist. Soc. B* 77(1):85–106, 2015.

O. Chapelle and L. Li. An empirical evaluation of Thompson sampling. In *Advances in Neural Information Processing Systems 24*, 2011.

P. Cotton. Scalable inversion of contests with correlated performances, including softmax and multinomial probit. arXiv:2609.01133, 2026. Reference implementation and parity-locked ports: `github.com/microprediction/winning`.

T. Domencich and D. McFadden. *Urban Travel Demand: A Behavioral Analysis*. North-Holland, 1975.

P. I. Frazier, W. B. Powell, and S. Dayanik. A knowledge-gradient policy for sequential information collection. *SIAM Journal on Control and Optimization*, 47(5):2410–2439, 2008.

J. Geweke. Bayesian inference in econometric models using Monte Carlo integration. *Econometrica*, 57(6):1317–1339, 1989.

V. Hajivassiliou, D. McFadden, and P. Ruud. Simulation of multivariate normal rectangle probabilities and their derivatives: theoretical and computational results. *Journal of Econometrics*, 72(1–2):85–134, 1996.

P. Hennig and C. J. Schuler. Entropy search for information-efficient global optimization. *Journal of Machine Learning Research*, 13:1809–1837, 2012.

R. Herbrich, T. Minka, and T. Graepel. TrueSkill: a Bayesian skill rating system. In *Advances in Neural Information Processing Systems 19*, 2007.

L. Li, W. Chu, J. Langford, and X. Wang. Unbiased offline evaluation of contextual-bandit-based news article recommendation algorithms. In *WSDM*, 2011.

N. Menet, J. Huebotter, P. Kassraie, and A. Krause. Efficiently estimating Gaussian probability of maximality. In *AISTATS*, 2025. arXiv:2501.13535.

N. Menet, A. Terzić, M. Hersche, A. Krause, and A. Rahimi. Thompson sampling via fine-tuning of LLMs. In *ICLR*, 2026. arXiv:2510.13328.

D. Müller, Y. Nesterov, and V. Shikhman. Discrete choice prox-functions on the simplex. arXiv:1909.05591, 2019.

B. O'Donoghue and T. Lattimore. Variational Bayesian optimistic sampling. In *Advances in Neural Information Processing Systems 34*, 2021. arXiv:2110.15688.

R. L. Plackett. A reduction formula for normal multivariate integrals. *Biometrika*, 41(3–4):351–360, 1954.

J. Ridgway. Computation of Gaussian orthant probabilities in high dimension. *Statistics and Computing* 26:899–916, 2016.

J. Tarbouriech, T. Lattimore, and B. O'Donoghue. Probabilistic inference in reinforcement learning done right. In *Advances in Neural Information Processing Systems 36*, 2023. arXiv:2311.13294.

M. Titsias. Variational learning of inducing variables in sparse Gaussian processes. In *AISTATS*, 2009.

A. Tuisov. Search as computation allocation. arXiv:2607.27871, 2026.