Approximate Inference via Fibrations of Statistical Games

Toby St Clere Smithe
VERSES Research
Topos Institute
q@tsmithe.net

We characterize a number of well known systems of approximate inference as *loss models*: lax sections of 2-fibrations of statistical games, constructed by attaching internally-defined loss functions to Bayesian lenses. Our examples include the relative entropy, which constitutes a *strict* section, and whose chain rule is formalized by the horizontal composition of the 2-fibration. In order to capture this compositional structure, we first introduce the notion of 'copy-composition', alongside corresponding bicategories through which the composition of copy-discard categories factorizes. These bicategories are a variant of the **Copara** construction, and so we additionally introduce coparameterized Bayesian lenses, proving that coparameterized Bayesian updates compose optically, as in the non-coparameterized case.

1 Introduction

In previous work [1], we introduced *Bayesian lenses*, observing that the Bayesian inversion of a composite stochastic channel is (almost surely) equal to the 'lens composite' of the inversions of the factors; that is, *Bayesian updates compose optically* ('BUCO') [2]. Formalizing this statement for a given category $\mathscr{C}$ all of whose morphisms ('channels') admit Bayesian inversion, we can observe that there is (almost surely) a functor $(-)^\dagger : \mathscr{C} \to \mathbf{BayesLens}(\mathscr{C})$ from $\mathscr{C}$ to the category $\mathbf{BayesLens}(\mathscr{C})$ whose morphisms $(X,A) \nrightarrow (Y,B)$ are Bayesian lenses: pairs (c,c') of a channel $X \twoheadrightarrow\!\!\!\!\!\!\bullet\; Y$ with a 'state-dependent' inverse $c' : \mathscr{C}(I,X) \to \mathscr{C}(B,A)$. Bayesian lenses constitute the morphisms of a fibration $\pi_{\mathsf{Lens}} : \mathbf{BayesLens}(\mathscr{C}) \to \mathscr{C}$, since $\mathbf{BayesLens}(\mathscr{C})$ is obtained as the Grothendieck construction of (the pointwise opposite of) an indexed category $\mathsf{Stat} : \mathscr{C}^{\,\mathrm{op}} \to \mathbf{Cat}$ of 'state-dependent channels' (recalled in Appendix A), and the functor $(-)^\dagger$ is in fact a section of π_{Lens}, taking $c : X \twoheadrightarrow\!\!\!\!\!\!\bullet\; Y$ to the lens $(c,c^\dagger) : (X,X) \nrightarrow (Y,Y)$, where $c^\dagger$ is the almost-surely unique Bayesian inversion of c (so that the projection π_{Lens} can simply forget the inversion, returning again the channel c).

The functor $(-)^\dagger$ picks out a special class of Bayesian lenses, which we may call *exact* (as they compute 'exact' inversions), but although the category $\mathbf{BayesLens}(\mathscr{C})$ has many other morphisms, the construction is not extravagant: by comparison, we can think of the non-exact lenses as representing *approximate* inference systems. This is particularly necessary in computational applications, because computing exact inversions is usually intractable, but this creates a new problem: choosing an approximation, and measuring its performance. In this paper, we formalize this process, by attaching *loss functions* to Bayesian lenses, thus creating another fibration, of *statistical games*. Sections of this latter fibration encode compositionally well-behaved systems of approximation that we call *loss models*.

A classic example of a loss model will be supplied by the relative entropy, which in some sense measures the 'divergence' between distributions: the game here is then to minimize the divergence between the approximate and exact inversions. If π and π' are two distributions on a space X, with corresponding density functions p_π and $p_{\pi'}$ (both with respect to a common measure), then their relative

Submitted to:
ACT 2023

entropy $D(\pi, \pi')$ is the real number given by $\mathbb{E}_{x\sim\pi}[\log p_\pi(x) - \log p_{\pi'}(x)]$[1]. Given a pair of channels $\alpha, \alpha' : A \rightsquigarrow B$ (again commensurately associated with densities), we can extend D to a map $D_{\alpha,\alpha'} : A \to \mathbb{R}_+$ in the natural way, writing $a \mapsto D\big(\alpha(a), \alpha'(a)\big)$. We can assign such a map $D_{\alpha,\alpha'}$ to any such parallel pair of channels, and so, following the logic of composition in $\mathscr{C}$, we might hope for the following equation to hold for all $a : A$ and composable parallel pairs $\alpha, \alpha' : A \rightsquigarrow B$ and $\beta, \beta' : B \rightsquigarrow C$,:

$$D_{\beta\bullet\alpha,\beta'\bullet\alpha'}(a) = \mathop{\mathbb{E}}_{b\sim\alpha(a)}\big[D_{\beta,\beta'}(b)\big] + D_{\alpha,\alpha'}(a)$$

The right-hand side is known as the *chain rule* for relative entropy, but, unfortunately, the equation does *not* hold in general, because the composites $\beta \bullet \alpha$ and $\beta' \bullet \alpha'$ involve an extra expectation (by the 'Chapman-Kolmogorov' rule for channel composition). However, we *can* satisfy an equation of this form by using 'copy-composition': if we write $\curlyvee_B$ to denote the canonical 'copying' comultiplication on B, and define $\beta \bullet^2 \alpha := (\mathsf{id}_B \otimes \beta) \bullet \curlyvee_B \bullet \alpha$, then $D_{\beta\bullet^2\alpha,\beta'\bullet^2\alpha'}(a)$ *does* equal the chain-rule form on the right-hand side. This result exhibits a general pattern about "copy-discard categories" [3] such as $\mathscr{C}$: composition $\bullet$ can be decomposed into first copying $\curlyvee$, and then discarding $\top$. If we don't discard, then we retain the 'intermediate' variables, and this results in a functorial assignment of relative entropies to channels.

The rest of this paper is dedicated to making use of this observation to construct loss models, including (but not restricted to) the relative entropy. The first complication that we encounter is that copy-composition is not strictly unital, because composing with an identity retains an extra variable. To deal with this, in §2, we construct a *bicategory* of copy-composite channels, extending the **Copara** construction [4, §2], and build coparameterized (copy-composite) Bayesian lenses accordingly; we also prove a corresponding BUCO result. In §3, we then construct 2-fibrations of statistical games, defining loss functions internally to any copy-discard category $\mathscr{C}$ that admits "bilinear effects". Because we are dealing with approximate systems, the 2-dimensional structure of the construction is useful: loss models are allowed to be *lax* sections. We then characterize the relative entropy, maximum likelihood estimation, the free energy, and the 'Laplacian' free energy as such loss models.

Assuming $\mathscr{C}$ is symmetric monoidal, the constructions here result in monoidal (2-)fibrations, but due to space constraints we defer the presentation of this structure (and its exemplification by the foregoing loss models) to Appendix B. For the same reason, we defer comprehensive proofs to Appendix C.

Remark 1.1. Much of this work is situated amongst monoidal fibrations of bicategories, the full theory of which is not known to the present author. Fortunately, enough structure is known for the present work to have been possible, and where things become murkier—such as in the context of monoidal indexed bicategories and their lax homomorphisms—the way largely seems clear. For this, we are grateful to Baković [5], Johnson and Yau [6], and Moeller and Vasilakopoulou [7] in particular for lighting the way; and we enthusiastically encourage the further elucidation of these structures by category theorists.

2 'Copy-composite' Bayesian lenses

2.1 Copy-composition by coparameterization

In a locally small copy-discard category $\mathscr{C}$, every object A is equipped with a canonical comonoid structure $(\curlyvee_A, \top_A)$, and so, by the comonoid laws, it is almost a triviality that the composition function

[1] For details about this 'expectation' notation $\mathbb{E}$, see 3.11.

$\bullet : \mathscr{C}(B,C) \times \mathscr{C}(A,B) \to \mathscr{C}(A,C)$ factorizes as

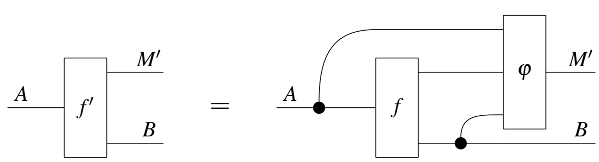
$$\mathscr{C}(B,C) \times \mathscr{C}(A,B) \xrightarrow{(\mathsf{id}_B \otimes -) \times \mathscr{C}\left(\mathsf{id}_A, \curlyvee_B\right)} \mathscr{C}(B \otimes B, B \otimes C) \times \mathscr{C}(A, B \otimes B) \cdots$$
$$\cdots \xrightarrow{\bullet} \mathscr{C}(A, B \otimes C) \xrightarrow{\mathscr{C}(\mathsf{id}_A, \mathsf{proj}_C)} \mathscr{C}(A,C)$$

where the first factor copies the B output of the first morphism and tensors the second morphism with the identity on B, the second factor composes the latter tensor with the copies, and the third discards the extra copy of B^2. This is, however, only *almost* trivial, since it witnesses the structure of 'Chapman-Kolmogorov' style composition in categories of stochastic channels such as $\mathscr{K}\ell(\mathscr{D})$, the Kleisli category of the (finitary) distributions monad $\mathscr{D} : \mathbf{Set} \to \mathbf{Set}$. There, given channels $c : A \twoheadrightarrow B$ and $d : B \twoheadrightarrow C$, the composite $d \bullet c$ is formed first by constructing the 'joint' channel $d \bullet^2 c$ defined by $(d \bullet^2 c)(b,c|a) := d(c|b)c(b|a)$, and then discarding (marginalizing over) $b : B$, giving

$$(d \bullet c)(c|a) = \sum_{b:B} (d \bullet^2 c)(b,c|a) = \sum_{b:B} d(c|b)c(b|a)\,.$$

Of course, the channel $d \bullet^2 c$ is not a morphism $A \twoheadrightarrow C$, but rather $A \twoheadrightarrow B \otimes C$; that is, it is *coparameterized* by B. Moreover, as noted above, $\bullet^2$ is not strictly unital: we need a 2-cell that discards the coparameter, and hence a bicategory, in order to recover (weak) unitality. We therefore construct a bicategory $\mathbf{Copara}_2(\mathscr{C})$ as a variant of the **Copara** construction [4, §2], in which a 1-cell $A \to B$ may be any morphism $A \twoheadrightarrow M \otimes B$ in $\mathscr{C}$, and where horizontal composition is precisely copy-composition.

Theorem 2.1. Let $(\mathscr{C}, \otimes, I)$ be a copy-discard category. Then there is a bicategory $\mathbf{Copara}_2(\mathscr{C})$ as follows. Its 0-cells are the objects of $\mathscr{C}$. A 1-cell $f : A \xrightarrow[M]{} B$ is a morphism $f : A \to M \otimes B$ in $\mathscr{C}$. A 2-cell $\varphi : f \Rightarrow f'$, with $f : A \xrightarrow[M]{} B$ and $f' : A \xrightarrow[M']{} B$, is a morphism $\varphi : A \otimes M \otimes B \to M'$ of $\mathscr{C}$, satisfying the *change of coparameter* axiom:

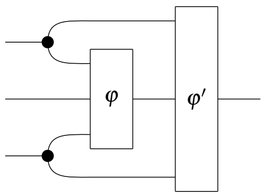


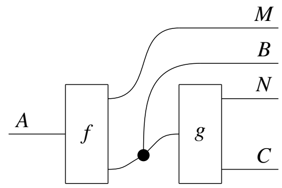
The identity 2-cell $\mathsf{id}_f : f \Rightarrow f$ on $f : A \xrightarrow[M]{} B$ is given by the projection morphism $\mathsf{proj}_M : A \otimes M \otimes B \to M$ obtained by discarding A and B, as in footnote 2. The identity 1-cell id_A on A is given by the inverse of the left unitor of the monoidal structure on $\mathscr{C}$, *i.e.* $\mathsf{id}_A := \lambda_A^{-1} : A \xrightarrow[I]{} A$, with coparameter thus given by the unit object I.

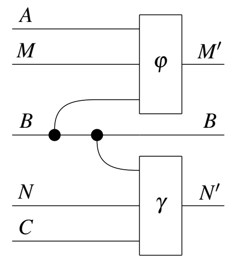
Given 2-cells $\varphi : f \Rightarrow f'$ and $\varphi' : f' \Rightarrow f''$, their vertical composite $\varphi' \odot \varphi : f \Rightarrow f''$ is given by the string diagram on the left below. Given 1-cells $f : A \xrightarrow[M]{} B$ then $g : B \xrightarrow[N]{} C$, the horizontal composite $g \circ f : A \xrightarrow[(M \otimes B) \otimes N]{} C$ is given by the middle string diagram below. Given 2-cells $\varphi : f \Rightarrow f'$ and $\gamma : g \Rightarrow g'$ between 1-cells $f, f' : A \xrightarrow[M]{} B$ and $g, g' : B \xrightarrow[N]{} C$, their horizontal composite $\gamma \circ \varphi : (g \circ f) \Rightarrow (g' \circ f')$ is

[2] We define $\mathsf{proj}_C := B \otimes C \xrightarrow{\bar{\top}_B \otimes \mathsf{id}_C} I \otimes C \xrightarrow{\lambda_C} C$, using the comonoid counit and the left unitor of $\mathscr{C}$'s monoidal structure.

defined by the string diagram on the right below.







Remark 2.2. When $\mathscr{C}$ is symmetric monoidal, $\mathbf{Copara}_2(\mathscr{C})$ inherits a monoidal structure, elaborated in Proposition B.1.

Remark 2.3. In order to capture the bidirectionality of Bayesian inversion we will need to conside left- and right-handed versions of the $\mathbf{Copara}_2$ construction. These are formally dual, and when $\mathscr{C}$ is symmetric monoidal (as in most examples) they are isomorphic. Nonetheless, it makes formalization easier if we explicitly distinguish $\mathbf{Copara}_2^l(\mathscr{C})$, in which the coparameter is placed on the left of the codomain (as above), from $\mathbf{Copara}_2^r(\mathscr{C})$, in which it is placed on the right. Aside from the swapping of this handedness, the rest of the construction is the same.

We end this section with three easy (and ambidextrous) propositions relating $\mathscr{C}$ and $\mathbf{Copara}_2(\mathscr{C})$.

Proposition 2.4. There is an identity-on-objects lax embedding $\iota : \mathscr{C} \hookrightarrow \mathbf{Copara}_2(\mathscr{C})$, mapping $f : X \to Y$ to $f : X \underset{I}{\to} Y$ (using the unitor of the monoidal structure on $\mathscr{C}$). The laxator $\iota(g) \circ \iota(f) \to \iota(g \circ f)$ discards the coparameter obtained from copy-composition.

Proposition 2.5. There is a 'discarding' functor $(-)^{\dagger} : \mathbf{Copara}_2(\mathscr{C}) \to \mathscr{C}$, which takes any coparameterized morphism and discards the coparameter.

Proposition 2.6. ι is a section of $(-)^{\dagger}$. That is, $\mathsf{id}_{\mathscr{C}} = \mathscr{C} \xhookrightarrow{\iota} \mathbf{Copara}_2(\mathscr{C}) \xrightarrow{(-)^{\dagger}} \mathscr{C}$.

2.2 Coparameterized Bayesian lenses

In order to define (bi)categories of statistical games, coherently with loss functions like the relative entropy that compose by copy-composition, we first need to define coparameterized (copy-composite) Bayesian lenses. Analogously to non-coparameterized Bayesian lenses, these will be obtained by applying a Grothendieck construction to an indexed bicategory [5, Def. 3.5] of state-dependent channels.

Definition 2.7. We define the indexed bicategory $\mathsf{Stat}_2 : \mathbf{Copara}_2^l(\mathscr{C})^{\mathrm{coop}} \to \mathbf{Bicat}$ fibrewise as follows.

(i) The 0-cells $\mathsf{Stat}_2(X)_0$ of each fibre $\mathsf{Stat}_2(X)$ are the objects $\mathscr{C}_0$ of $\mathscr{C}$.

(ii) For each pair of 0-cells A, B, the hom-category $\mathsf{Stat}_2(X)(A,B)$ is defined to be the functor category $\mathbf{Cat}\big(\mathsf{disc}\,\mathscr{C}(I,X), \mathbf{Copara}_2^r(\mathscr{C})(A,B)\big)$, where disc denotes the functor taking a set to the associated discrete category.

(iii) For each 0-cell A, the identity functor $\mathsf{id}_A : \mathbf{1} \to \mathsf{Stat}_2(X)(A,A)$ is the constant functor on the identity on A in $\mathbf{Copara}_2^r(\mathscr{C})$; *i.e.* $\mathsf{disc}\,\mathscr{C}(I,X) \xrightarrow{!} 1 \xrightarrow{\mathsf{id}_A} \mathbf{Copara}_2^r(\mathscr{C})(A,A)$.

(iv) For each triple A,B,C of 0-cells, the horizontal composition functor $\circ_{A,B,C}$ is defined by

$$\begin{aligned}
&\circ_{A,B,C} : \mathsf{Stat}_2(X)(B,C) \times \mathsf{Stat}_2(X)(A,B) \cdots \\
&\cdots \xrightarrow{=} \mathbf{Cat}\big(\mathsf{disc}\,\mathscr{C}(I,X), \mathbf{Copara}_2^r(\mathscr{C})(B,C)\big) \times \mathbf{Cat}\big(\mathsf{disc}\,\mathscr{C}(I,X), \mathbf{Copara}_2^r(\mathscr{C})(A,B)\big) \cdots \\
&\cdots \xrightarrow{\times} \mathbf{Cat}\big(\mathsf{disc}\,\mathscr{C}(I,X)^2, \mathbf{Copara}_2^r(\mathscr{C})(B,C) \times \mathbf{Copara}_2^r(\mathscr{C})(A,B)\big) \cdots \\
&\cdots \xrightarrow{\mathbf{Cat}(\curlyvee,\circ)} \mathbf{Cat}\big(\mathsf{disc}\,\mathscr{C}(I,X), \mathbf{Copara}_2^r(\mathscr{C})(A,C)\big) \cdots \\
&\cdots \xrightarrow{=} \mathsf{Stat}_2(X)(A,C)
\end{aligned}$$

where $\mathbf{Cat}(\curlyvee,\circ)$ indicates pre-composition with the universal (Cartesian) copying functor in $(\mathbf{Cat},\times,\mathbf{1})$ and post-composition with the horizontal composition functor of $\mathbf{Copara}_2^r(\mathscr{C})$.

For each pair of 0-cells X,Y in $\mathbf{Copara}^l(\mathscr{C})$, we define the reindexing pseudofunctor $\mathsf{Stat}_{2,X,Y}$: $\mathbf{Copara}^l(\mathscr{C})(X,Y)^{\mathrm{op}} \to \mathbf{Bicat}\big(\mathsf{Stat}_2(Y), \mathsf{Stat}_2(X)\big)$ as follows.

(a) For each 1-cell f in $\mathbf{Copara}^l(\mathscr{C})(X,Y)$, we obtain a pseudofunctor $\mathsf{Stat}_2(f) : \mathsf{Stat}_2(Y) \to \mathsf{Stat}_2(X)$ which acts as the identity on 0-cells.

(b) For each pair of 0-cells A,B in $\mathsf{Stat}_2(Y)$, the functor $\mathsf{Stat}_2(f)_{A,B}$ is defined as the precomposition functor $\mathbf{Cat}\big(\mathsf{disc}\,\mathscr{C}(I,f^{\dagger}), \mathbf{Copara}_2^r(\mathscr{C})(A,B)\big)$, where $(-)^{\dagger}$ is the discarding functor $\mathbf{Copara}_2^l(\mathscr{C}) \to \mathscr{C}$ of Proposition 2.5.

(c) For each 2-cell $\varphi : f \Rightarrow f'$ in $\mathbf{Copara}_2^l(\mathscr{C})(X,Y)$, the pseudonatural transformation $\mathsf{Stat}_2(\varphi)$: $\mathsf{Stat}_2(f') \Rightarrow \mathsf{Stat}_2(f)$ is defined on 0-cells $A : \mathsf{Stat}_2(Y)$ by the discrete natural transformation with components $\mathsf{Stat}_2(\varphi)_A := \mathsf{id}_A$, and on 1-cells $c : \mathsf{Stat}_2(Y)(A,B)$ by the substitution natural transformation with constitutent 2-cells $\mathsf{Stat}_2(\varphi)_c : \mathsf{Stat}_2(f)(c) \Rightarrow \mathsf{Stat}_2(f')(c)$ in $\mathsf{Stat}_2(X)$ which acts by replacing $\mathbf{Cat}\big(\mathsf{disc}\,\mathscr{C}(I,f^{\dagger}), \mathbf{Copara}_2^r(\mathscr{C})(A,B)\big)$ by $\mathbf{Cat}\big(\mathsf{disc}\,\mathscr{C}(I,f'^{\dagger}), \mathbf{Copara}_2^r(\mathscr{C})(A,B)\big)$; and which we might alternatively denote by $\mathbf{Cat}\big(\mathsf{disc}\,\mathscr{C}(I,\varphi^{\dagger}), \mathbf{Copara}_2^r(\mathscr{C})(A,B)\big)$.

Notation 2.8. We will write $f : A \overset{X}{\underset{M}{\rightsquigarrow}} B$ to denote a state-dependent coparameterized channel f with coparameter M and state-dependence on X.

In 1-category theory, lenses are morphisms in the fibrewise opposite of a fibration [8]. Analogously, our bicategorical Bayesian lenses are obtained as 1-cells in the bicategorical Grothendieck construction [5, §6] of (the pointwise opposite of) the indexed bicategory Stat_2.

Definition 2.9. Fix a copy-discard category $(\mathscr{C},\otimes,I)$. We define the bicategory of coparameterized Bayesian lenses in $\mathscr{C}$, denoted $\mathbf{BayesLens}_2(\mathscr{C})$ or simply $\mathbf{BayesLens}_2$, to be the bicategorical Grothendieck construction of the pointwise opposite of the corresponding indexed bicategory Stat_2, with the following data:

(i) A 0-cell in $\mathbf{BayesLens}_2$ is a pair (X,A) of an object X in $\mathbf{Copara}_2^l(\mathscr{C})$ and an object A in $\mathsf{Stat}_2(X)$; equivalently, a 0-cell in $\mathbf{BayesLens}_2$ is a pair of objects in $\mathscr{C}$.

(ii) The hom-category $\mathbf{BayesLens}_2\big((X,A),(Y,B)\big)$ is the product category $\mathbf{Copara}_2^l(\mathscr{C})(X,Y) \times \mathsf{Stat}_2(X)(B,A)$.

(iii) The identity on (X,A) is given by the pair $(\mathsf{id}_X, \mathsf{id}_A)$.

(iv) For each triple of 0-cells $(X,A),(Y,B),(Z,C)$, the horizontal composition functor is given by

$$\begin{aligned}
&\mathbf{BayesLens}_2\big((Y,B),(Z,C)\big) \times \mathbf{BayesLens}_2\big((X,A),(Y,B)\big) \\
= \;& \mathbf{Copara}_2^l(\mathscr{C})(Y,Z) \times \mathsf{Stat}_2(Y)(C,B) \times \mathbf{Copara}_2^l(\mathscr{C})(X,Y) \times \mathsf{Stat}_2(X)(B,A) \\
\xrightarrow{\sim} \;& \sum_{g:\mathbf{Copara}_2^l(\mathscr{C})(Y,Z)} \; \sum_{f:\mathbf{Copara}_2^l(\mathscr{C})(X,Y)} \mathsf{Stat}_2(Y)(C,B) \times \mathsf{Stat}_2(X)(B,A) \\
\xrightarrow{\sum_g \sum_f \mathsf{Stat}_2(f)_{C,B} \times \mathsf{id}} \;& \sum_{g:\mathbf{Copara}_2^l(\mathscr{C})(Y,Z)} \; \sum_{f:\mathbf{Copara}_2^l(\mathscr{C})(X,Y)} \mathsf{Stat}_2(X)(C,B) \times \mathsf{Stat}_2(X)(B,A) \\
\xrightarrow{\sum_{\circ_{\mathbf{Copara}_2^l(\mathscr{C})}} \circ^{\mathsf{Stat}_2(X)}} \;& \sum_{g \circ f:\mathbf{Copara}_2^l(\mathscr{C})(X,Z)} \mathsf{Stat}_2(X)(C,A) \\
\xrightarrow{\sim} \;& \mathbf{BayesLens}_2\big((X,A),(Z,C)\big)
\end{aligned}$$

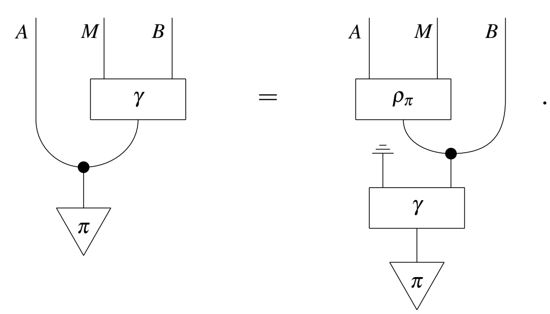
where the functor in the penultimate line amounts to the product of the horizontal composition functors on $\mathbf{Copara}_2^l(\mathscr{C})$ and $\mathsf{Stat}_2(X)$.

Remark 2.10. When $\mathscr{C}$ is symmetric monoidal, Stat_2 acquires the structure of a monoidal indexed bicategory (Definition B.2 and Theorem B.3), and hence $\mathbf{BayesLens}_2$ becomes a monoidal bicategory (Corollary B.4).

2.3 Coparameterized Bayesian updates compose optically

So that our generalized Bayesian lenses are worthy of the name, we should also confirm that Bayesian inversions compose according to the lens pattern ('optically') also in the coparameterized setting. Such confirmation is the subject of the present section, and so we first introduce a new "coparameterized Bayes' rule".

Definition 2.11. We say that a coparameterized channel $\gamma : A \mathrel{\bullet\!\!\!\rightarrow} M \otimes B$ *admits Bayesian inversion* if there exists a dually coparameterized channel $\rho_\pi : B \mathrel{\bullet\!\!\!\rightarrow} A \otimes M$ satisfying the graphical equation (with string diagrams read from bottom to top)



In this case, we say that ρ_π is the *Bayesian inversion of* γ *with respect to* π.

With this definition, we can supply the desired result that "coparameterized Bayesian updates compose optically".

Theorem 2.12. Suppose $(\gamma,\gamma^\dagger) : (A,A) \underset{M}{\twoheadrightarrow} (B,B)$ and $(\delta,\delta^\dagger) : (B,B) \underset{N}{\twoheadrightarrow} (C,C)$ are coparameterized Bayesian lenses in $\mathbf{BayesLens}_2$. Suppose also that $\pi : I \mathrel{\bullet\!\!\!\rightarrow} A$ is a state on A in the underlying category of channels $\mathscr{C}$, such that $\gamma_\pi^\dagger$ is a Bayesian inversion of γ with respect to π, and such that $\delta_{\gamma\pi}^\dagger$ is a Bayesian inversion of δ with respect to $(\gamma\pi)^{\bar{\top}}$; where the notation $(-)^{\bar{\top}}$ represents discarding coparameters. Then $\gamma_\pi^\dagger \bullet \delta_{\gamma\pi}^\dagger$ is a Bayesian inversion of $\delta \bullet \gamma$ with respect to π. (Here $\bullet$ denotes copy-composition.) Moreover, if $(\delta \bullet \gamma)_\pi^\dagger$ is any Bayesian inversion of $\delta \bullet \gamma$ with respect to π, then $\gamma_\pi^\dagger \bullet \delta_{\gamma\pi}^\dagger$ is $(\delta\gamma\pi)^{\bar{\top}}$-almost-surely equal to $(\delta \bullet \gamma)_\pi^\dagger$: that is, $(\delta \bullet \gamma)_\pi^\dagger \overset{(\delta\gamma\pi)^{\bar{\top}}}{\sim} \gamma_\pi^\dagger \bullet \delta_{\gamma\pi}^\dagger$.

In order to satisfy this coparameterized Bayes' rule, a Bayesian lens must be of 'simple' type.

Definition 2.13. We say that a coparameterized Bayesian lens (c, c') is *simple* if its domain and codomain are 'diagonal' (duplicate pairs of objects) and if the coparameter of c is equal to the coparameter of c'. In this case, we can write the type of (c, c') as $(X,X) \underset{M}{\leftrightarrow\!\!\!\rightarrow} (Y,Y)$ or simply $X \underset{M}{\leftrightarrow\!\!\!\rightarrow} Y$.

3 Statistical games for local approximate inference

3.1 Losses for lenses

Statistical games are obtained by attaching to Bayesian lenses *loss functions*, representing 'local' quantifications of the performance of approximate inference systems. Because this performance depends on the system's context (*i.e.*, the prior $\pi : I \to\!\!\!\!\bullet\, X$ and the observed data $b : B$), a loss function at its most concrete will be a function $\mathscr{C}(I,X) \times B \to \mathbb{R}_+$. To internalize this type in $\mathscr{C}$, we may recall that, when $\mathscr{C}$ is the category **sfKrn** of s-finite kernels or the Kleisli category $\mathscr{K}\ell(\mathscr{D}_{\leqslant 1})$ of the subdistribution monad, a density function $p_c : X \times Y \to [0,1]$ for a channel $c : X \to\!\!\!\!\bullet\, Y$ corresponds to an *effect* (or *costate*) $X \otimes Y \to\!\!\!\!\bullet\, I$. In this way, we can see a loss function as a kind of *state-dependent effect* $B \overset{X}{\to\!\!\!\!\bullet} I$.

Loss functions will compose by sum, and so we need to ask for the effects in $\mathscr{C}$ to form a monoid. Moreover, we need this monoid to be 'bilinear' with respect to channels, so that Stat-reindexing (*cf.* Definition A.1) preserves sums. These conditions are formalized in the following definition.

Definition 3.1. Suppose $(\mathscr{C}, \otimes, I)$ is a copy-discard category. We say that $\mathscr{C}$ *has bilinear effects* if the following conditions are satisfied:

(i) *effect monoid*: there is a natural transformation $+ : \mathscr{C}(-,I) \times \mathscr{C}(=,I) \Rightarrow \mathscr{C}(- \otimes =, I)$ making $\sum_{A:\mathscr{C}} \mathscr{C}(A,I)$ into a commutative monoid with unit $0 : I \to\!\!\!\!\bullet\, I$;

(ii) *bilinearity*: $(g + g') \bullet \curlyvee \bullet f = g \bullet f + g' \bullet f$ for all effects g, g' and morphisms f such that $(g + g') \bullet \curlyvee \bullet f$ exists.

A trivial example of a category with bilinear effects is supplied by any Cartesian category, such as **Set**. If M is any monoid in **Set**, then a less trivial example is supplied by the Kleisli category of the corresponding free module monad; bilinearity follows from the module structure. A related non-example is $\mathscr{K}\ell(\mathscr{D}_{\leqslant 1})$: the failure here is that the effects only form a *partial* monoid[3]. More generally, the category **sfKrn** of s-finite kernels [10] has bilinear effects (owing to the linearity of integration), and we will assume this as our ambient $\mathscr{C}$ for the examples below.

Given such a category $\mathscr{C}$ with bilinear effects, we can lift the natural transformation $+$, and hence the

[3] Indeed, an *effect algebra* is a kind of partial monoid [9, §2], but we do not need the extra complication here.

bilinear effect structure, to the fibres of $\mathsf{Stat}_{\mathscr{C}}$, using the universal property of the product of categories:

$$\begin{aligned}
+_X : \mathsf{Stat}(X)(-,I) \times \mathsf{Stat}(X)(=,I) &= \mathbf{Set}\big(\mathscr{C}(I,X),\mathscr{C}(-,I)\big) \times \mathbf{Set}\big(\mathscr{C}(I,X),\mathscr{C}(=,I)\big) \\
&\xrightarrow{(\cdot,\cdot)} \mathbf{Set}\big(\mathscr{C}(I,X),\mathscr{C}(-,I)\times\mathscr{C}(=,I)\big) \\
&\xrightarrow{\mathbf{Set}(\mathscr{C}(I,X),+)} \mathbf{Set}\big(\mathscr{C}(I,X),\mathscr{C}(-\otimes=,I)\big) \\
&\xrightarrow{=} \mathsf{Stat}(X)(-\otimes=,I)
\end{aligned}$$

Here, $(\cdot,\cdot)$ denotes the pairing operation obtained from the universal property. In this way, each $\mathsf{Stat}(X)$ has bilinear effects. Note that this lifting is (strictly) compatible with the reindexing of Stat, so that $+_{(-)}$ defines an indexed natural transformation. This means in particular that *reindexing distributes over sums*: given state-dependent effects $g,g' : B \overset{Y}{\to\!\!\!\bullet} I$ and a channel $c : X \to\!\!\!\bullet Y$, we have $(g +_Y g')_c = g_c +_X g'_c$. We will thus generally omit the subscript from the lifted sum operation, and just write $+$.

We are now ready to construct the bicategory of statistical games.

Definition 3.2. Suppose $(\mathscr{C},\otimes,I)$ has bilinear effects, and let $\mathbf{BayesLens}_2$ denote the corresponding bicategory of (copy-composite) Bayesian lenses. We will write $\mathbf{SGame}_{\mathscr{C}}$ to denote the following *bicategory of (copy-composite) statistical games* in $\mathscr{C}$:

- The 0-cells are the 0-cells (X,A) of $\mathbf{BayesLens}_2$;
- the 1-cells, called *statistical games*, $(X,A) \to (Y,B)$ are pairs (c,L^c) of a 1-cell $c : (X,A) \nrightarrow (Y,B)$ in $\mathbf{BayesLens}_2$ and a *loss* $L^c : B \overset{X}{\to\!\!\!\bullet} I$ in $\mathsf{Stat}(X)(B,I)$;
- given 1-cells $(c,L^c),(c',L^{c'}) : (X,A) \to (Y,B)$, the 2-cells $(c,L^c) \Rightarrow (c',L^{c'})$ are pairs (α,K^α) of a 2-cell $\alpha : c \Rightarrow c'$ in $\mathbf{BayesLens}_2$ and a loss $K^\alpha : B \overset{X}{\to\!\!\!\bullet} I$ such that $L^c = L^{c'} + K^\alpha$;
- the identity 2-cell on (c,L^c) is $(\mathsf{id}_c,0)$;
- given 2-cells $(\alpha,K^\alpha) : (c,L^c) \Rightarrow (c',L^{c'})$ and $(\alpha',K^{\alpha'}) : (c',L^{c'}) \Rightarrow (c'',L^{c''})$, their vertical composite is $(\alpha' \circ \alpha, K^{\alpha'} + K^\alpha)$, where $\circ$ here denotes vertical composition in $\mathbf{BayesLens}_2$;
- given 1-cells $(c,L^c) : (X,A) \to (Y,B)$ and $(d,L^d) : (Y,B) \to (Z,C)$, their horizontal composite is $(c \mathbin{\varphi} d, L^d_c + L^c \circ \overline{d}_c)$; and
  - given 2-cells $(\alpha,K^\alpha) : (c,L^c) \Rightarrow (c',L^{c'})$ and $(\beta,K^\beta) : (d,L^d) \Rightarrow (d',L^{d'})$, their horizontal composite is $(\beta \mathbin{\varphi} \alpha, K^\beta_c + K^\alpha \circ \overline{d}_c)$, where φ here denotes horizontal composition in $\mathbf{BayesLens}_2$.

Theorem 3.3. Definition 3.2 generates a well-defined bicategory.

The proof of this result (given in §C.3) is that $\mathbf{SGame}_{\mathscr{C}}$ is obtained via a pair of bicategorical Grothendieck constructions [5]: first to obtain Bayesian lenses; and then to attach the loss functions. The proof depends on the intermediate result that our effect monoids can be 'upgraded' to monoidal categories; we then use the delooping of this structure to associate (state-dependent) losses to (state-dependent) channels, after discarding the coparameters of the latter.

Lemma 3.4. Suppose $(\mathscr{C},\otimes,I)$ has bilinear effects. Then, for each object B, $\mathscr{C}(B,I)$ has the structure of a symmetric monoidal category. The objects of $\mathscr{C}(B,I)$ are its elements, the effects. If g,g' are two effects, then a morphism $\kappa : g \to g'$ is an effect such that $g = g' + \kappa$; the identity morphism for each effect id_g is then the constant 0 effect. Likewise, the tensor of two effects is their sum, and the corresponding unit is the constant 0. Precomposition by any morphism $c : A \to\!\!\!\bullet B$ preserves the monoidal category structure, making the presheaf $\mathscr{C}(-,I)$ into a fibrewise-monoidal indexed category $\mathscr{C}^{\mathrm{op}} \to \mathbf{MonCat}$.

As already indicated, this structure lifts to the fibres of Stat.

Corollary 3.5. For each object X in a category with bilinear effects, and for each object B, $\mathsf{Stat}(X)(B,I)$ inherits the symmetric monoidal structure of $\mathscr{C}(B,I)$; note that morphisms of state-dependent effects are likewise state-dependent, and that tensoring (summing) state-dependent effects involves copying the parameterizing state. Moreover, $\mathsf{Stat}(-)(=,I)$ is a fibrewise-monoidal indexed category $\sum_{X:\mathscr{C}^{\mathrm{op}}} \mathsf{Stat}(X)^{\mathrm{op}} \to \mathbf{MonCat}$.

3.2 Local inference models

In the context of approximate inference, one often does not have a single statistical model to evaluate, but a whole family of them. In particularly nice situations, this family is actually a subcategory $\mathscr{D}$ of $\mathscr{C}$, with the family of statistical models being all those that can be composed in $\mathscr{D}$. The problem of approximate inference can then be formalized as follows. Since both $\mathbf{BayesLens}_2$ and $\mathbf{SGame}_{\mathscr{C}}$ were obtained by bicategorical Grothendieck constructions, we have a pair of 2-fibrations $\mathbf{SGame}_{\mathscr{C}} \xrightarrow{\pi_{\mathsf{Loss}}} \mathbf{BayesLens}_2 \xrightarrow{\pi_{\mathsf{Lens}}} \mathbf{Copara}_2^l(\mathscr{C})$. Each of π_{Loss}, π_{Lens}, and the discarding functor $(-)^{\bar{\top}}$ can be restricted to the subcategory $\mathscr{D}$. The inclusion $\mathscr{D} \hookrightarrow \mathbf{Copara}_2^l(\mathscr{D})$ is a section of this restriction of $(-)^{\bar{\top}}$; the assignment of inversions to channels in $\mathscr{D}$ then corresponds to a 2-section of the 2-fibration π_{Lens} (restricted to $\mathscr{D}$); and the subsequent assignment of losses is a further 2-section of π_{Loss}. This situation is depicted in the following diagram of bicategories:

$$\begin{array}{ccc}
\mathbf{SGame}_{\mathscr{D}} & \hookrightarrow & \mathbf{SGame}_{\mathscr{C}} \\
\curvearrowright\ \big\downarrow \pi_{\mathsf{Loss}}|_{\mathscr{D}} & & \big\downarrow \pi_{\mathsf{Loss}} \\
\mathbf{BayesLens}_2|_{\mathscr{D}} & \hookrightarrow & \mathbf{BayesLens}_2 \\
\curvearrowright\ \big\downarrow \pi_{\mathsf{Lens}}|_{\mathscr{D}} & & \big\downarrow \pi_{\mathsf{Lens}} \\
\mathbf{Copara}_2^l(\mathscr{D}) & \hookrightarrow & \mathbf{Copara}_2^l(\mathscr{C}) \\
\curvearrowright\ \big\downarrow \bar{\top}|_{\mathscr{D}} & & \big\downarrow \bar{\top} \\
\mathscr{D} & \hookrightarrow & \mathscr{C}
\end{array}$$

This motivates the following definitions of *inference system* and *loss model*, although, for the sake of our examples, we will explicitly allow the loss-assignment to be lax. Before giving these new definitions, we recall the notion of *essential image* of a functor.

Definition 3.6 ([11])**.** Suppose $F : \mathscr{C} \to \mathscr{D}$ is an n-functor (a possibly weak homomorphism of weak n-categories). The *image* of F is the smallest sub-n-category of $\mathscr{D}$ that contains $F(\alpha)$ for all k-cells α in $\mathscr{C}$, along with any $(k+1)$-cells relating images of composites and composites of images, for all $0 \leqslant k \leqslant n$. We say that a sub-n-category $\mathscr{D}$ is *replete* if, for any k-cells α in $\mathscr{D}$ and β in $\mathscr{C}$ (with $0 \leqslant k < n$) such that $f : \alpha \Rightarrow \beta$ is a $(k+1)$-isomorphism in $\mathscr{C}$, then f is also a $(k+1)$-isomorphism in $\mathscr{D}$. The *essential image* of F, denoted $\mathsf{im}(F)$, is then the smallest replete sub-n-category of $\mathscr{D}$ containing the image of F.

Definition 3.7. Suppose $(\mathscr{C}, \otimes, I)$ is a copy-delete category. An *inference system* in $\mathscr{C}$ is a pair $(\mathscr{D}, \ell)$ of a subcategory $\mathscr{D} \hookrightarrow \mathscr{C}$ along with a section $\ell : \mathsf{im}(\iota) \to \mathbf{BayesLens}_2|_{\mathscr{D}}$ of $\pi_{\mathsf{Lens}}|_{\mathscr{D}}$, where $\mathsf{im}(\iota)$ is the essential image of the canonical lax inclusion $\iota : \mathscr{D} \hookrightarrow \mathbf{Copara}_2^l(\mathscr{D})$.

Definition 3.8. Suppose $(\mathscr{C}, \otimes, I)$ has bilinear effects and $\mathscr{B}$ is a subbicategory of $\mathbf{BayesLens}_2$. A *loss model* for $\mathscr{B}$ is a lax section L of the restriction $\pi_{\mathsf{Loss}}|_{\mathscr{B}}$ of π_{Loss} to $\mathscr{B}$. We say that L is a *strict* loss model if it is in fact a strict 2-functor, and a *strong* loss model if it is in fact a pseudofunctor.

Remark 3.9. We may often be interested in loss models for which $\mathscr{B}$ is in fact the essential image of an inference system, but we do not stipulate this requirement in the definition as it is not necessary for the following development.

Since lax functors themselves collect into categories, and using the monoidality of $+$, we have the following easy proposition that will prove useful below.

Proposition 3.10. Loss models for $\mathscr{B}$ constitute the objects of a symmetric monoidal category $\big(\mathsf{Loss}(\mathscr{B}),+,0\big)$. The morphisms of $\mathsf{Loss}(\mathscr{B})$ are icons (identity component oplax transformations [6, §4.6]) between the corresponding lax functors, and they compose accordingly. The monoidal structure is given by sums of losses.

3.3 Examples

Each of our examples involves taking expectations of log-densities, and so to make sense of them it first helps to understand what we mean by "taking expectations".

Notation 3.11 (Expectations)**.** Written as a function, a density p on X has the type $X \to \mathbb{R}_+$; written as an effect, the type is $X \twoheadrightarrow I$. Given a measure or distribution π on X (equivalently, a state $\pi : I \twoheadrightarrow X$), we can compute the expectation of p under π as the composite $p \bullet \pi$. We write the resulting quantity as $\mathbb{E}_\pi[p]$, or more explicitly as $\mathbb{E}_{x\sim\pi}\big[p(x)\big]$. We can think of this expectation as representing the 'validity' (or truth value) of the 'predicate' p given the state π [12].

3.3.1 Relative entropy and Bayesian inference

For our first example, we return to the subject with which we opened this paper: the compositional structure of the relative entropy. We begin by giving a precise definition.

Definition 3.12. Suppose α,β are both measures on X, with α absolutely continuous with respect to β. Then the *relative entropy* or *Kullback-Leibler divergence* from α to β is the quantity $D_{KL}(\alpha,\beta) := \mathbb{E}_\alpha\left[\log\frac{\alpha}{\beta}\right]$, where $\frac{\alpha}{\beta}$ is the Radon-Nikodym derivative of α with respect to β.

Remark 3.13. When α and β admit density functions p_α and p_β with respect to the same base measure $\mathrm{d}x$, then $D_{KL}(\alpha,\beta)$ can equally be computed as $\mathbb{E}_{x\sim\alpha}\big[\log p_\alpha(x) - \log p_\beta(x)\big]$. It it this form that we will adopt henceforth.

Proposition 3.14. Let $\mathscr{B}$ be a subbicategory of simple lenses in $\mathbf{BayesLens}_2$, all of whose channels admit density functions with respect to a common measure and whose forward channels admit Bayesian inversion (and whose forward and backward coparameters coincide), and with only structural 2-cells. Then the relative entropy defines a strict loss model $\mathsf{KL} : \mathscr{B} \to \mathbf{SGame}$. Given a lens $(c,c') : (X,X) \nrightarrow (Y,Y)$, KL assigns the loss function $\mathsf{KL}(c,c') : Y \xrightarrow{X} I$ defined, for $\pi : I \twoheadrightarrow X$ and $y : Y$, by the relative entropy $\mathsf{KL}(c,c')_\pi(y) := D_{KL}\big(c'_\pi(y), c^\dagger_\pi(y)\big)$, where $c^\dagger$ is the exact inversion of c.

Successfully playing a relative entropy game entails minimizing the divergence from the approximate to the exact posterior. This divergence is minimized when the two coincide, and so KL represents a form of approximate Bayesian inference.

3.3.2 Maximum likelihood estimation

A statistical system may be more interested in predicting observations than updating beliefs. This is captured by the process of *maximum likelihood estimation*.

Definition 3.15. Let $(c,c') : (X,X) \nrightarrow (Y,Y)$ be a simple lens whose forward channel c admits a density function p_c. Then its *log-likelihood* is the loss function defined by $\mathsf{MLE}(c,c')_\pi(y) := -\log p_{c^\dagger \bullet \pi}(y)$.

Proposition 3.16. Let $\mathscr{B}$ be a subbicategory of lenses in $\mathbf{BayesLens}_2$ all of which admit density functions with respect to a common measure, and with only structural 2-cells. Then the assignment $(c, c') \mapsto \mathsf{MLE}(c, c')$ defines a lax loss model $\mathsf{MLE} : \mathscr{B} \to \mathbf{SGame}$.

Successfully playing a maximum likelihood game involves maximizing the log-likelihood that the system assigns to its observations $y : Y$. This process amounts to choosing a channel c that assigns high likelihood to likely observations, and thus encodes a valid model of the data distribution.

3.3.3 Autoencoders via the free energy

Many adaptive systems neither just infer nor just predict: they do both, building a model of their observations that they also invert to update their beliefs. In machine learning, such systems are known as *autoencoders*, as they 'encode' (infer) and 'decode' (predict), 'autoassociatively' [13]. In a Bayesian context, they are known as *variational autoencoders* [14], and their loss function is the *free energy* [15].

Definition 3.17. The *free energy* loss model is the sum of the relative entropy and the likelihood loss models: $\mathsf{FE} := \mathsf{KL} + \mathsf{MLE}$. Given a simple lens $(c, c') : (X, X) \nrightarrow (Y, Y)$ admitting Bayesian inversion and with densities, FE assigns the loss function

$$\begin{aligned} \mathsf{FE}(c, c')_\pi(y) &= (\mathsf{KL} + \mathsf{MLE})(c, c')_\pi(y) \\ &= D_{KL}\big(c'_\pi(y), c^\dagger_\pi(y)\big) - \log p_{c^{\mp} \bullet \pi}(y) \end{aligned}$$

Remark 3.18. Beyond its autoencoding impetus, another important property of the free energy is its improved computational tractability compared to either the relative entropy or the likelihood loss. This property is a consequence of the following fact: although obtained as the sum of terms which both depend on an expensive marginalization[4], the free energy itself does not. This can be seen by expanding the definitions of the relative entropy and of $c^\dagger_\pi$ and rearranging terms:

$$\begin{aligned} \mathsf{FE}(c, c')_\pi(y) &= D_{KL}\big(c'_\pi(y), c^\dagger_\pi(y)\big) - \log p_{c^{\mp} \bullet \pi}(y) \\ &= \mathop{\mathbb{E}}_{(x,m) \sim c'_\pi(y)} \big[\log p_{c'_\pi}(x, m|y) - \log p_{c^\dagger_\pi}(x, m|y)\big] - \log p_{c^{\mp} \bullet \pi}(y) \\ &= \mathop{\mathbb{E}}_{(x,m) \sim c'_\pi(y)} \big[\log p_{c'_\pi}(x, m|y) - \log p_{c^\dagger_\pi}(x, m|y) - \log p_{c^{\mp} \bullet \pi}(y)\big] \\ &= \mathop{\mathbb{E}}_{(x,m) \sim c'_\pi(y)} \left[\log p_{c'_\pi}(x, m|y) - \log \frac{p_c(m, y|x) p_\pi(x)}{p_{c^{\mp} \bullet \pi}(y)} - \log p_{c^{\mp} \bullet \pi}(y)\right] \\ &= \mathop{\mathbb{E}}_{(x,m) \sim c'_\pi(y)} \big[\log p_{c'_\pi}(x, m|y) - \log p_c(m, y|x) - \log p_\pi(x)\big] \\ &= D_{KL}\big(c'_\pi(y), \pi \otimes \mathbf{1}\big) - \mathop{\mathbb{E}}_{(x,m) \sim c'_\pi(y)} \big[\log p_c(m, y|x)\big] \end{aligned}$$

Here, $\mathbf{1}$ denotes the measure with density 1 everywhere. Note that when the coparameter is trivial, $\mathsf{FE}(c, c')_\pi(y)$ reduces to

$$D_{KL}\big(c'_\pi(y), \pi\big) - \mathop{\mathbb{E}}_{x \sim c'_\pi(y)} \big[\log p_c(y|x)\big] .$$

Remark 3.19. The name *free energy* is due to an analogy with the Helmholtz free energy in thermodynamics, as we can write it as the difference between an (expected) energy and an entropy

[4] Evaluating the pushforward $c^{\mp} \bullet \pi$ involves marginalizing over the intermediate variable; and evaluating $c^\dagger_\pi(y)$ also involves evaluating $c^{\mp} \bullet \pi$.

term:

$$\begin{aligned}\mathsf{FE}(c,c')_\pi(y) &= \mathop{\mathbb{E}}_{(x,m)\sim c'_\pi(y)}\left[-\log p_c(m,y|x) - \log p_\pi(x)\right] - S_{X\otimes M}\left[c'_\pi(y)\right] \\ &= \mathop{\mathbb{E}}_{(x,m)\sim c'_\pi(y)}\left[E_{(c,\pi)}(x,m,y)\right] - S_{X\otimes M}\left[c'_\pi(y)\right] \; = U - TS\end{aligned}$$

where we call $E_{(c,\pi)} : X\otimes M\otimes Y \xrightarrow{X} I$ the *energy*, and where $S_{X\otimes M} : I \xrightarrow{X\otimes M} I$ is the Shannon entropy. The last equality makes the thermodynamic analogy: U here is the *internal energy* of the system; $T = 1$ is the *temperature*; and S is again the entropy.

3.3.4 The Laplace approximation

Although optimizing the free energy does not necessitate access to exact inversions, it is still necessary to compute an expectation under the approximate inversion, and unless one chooses wisely[5], this might also be difficult. One such wise choice established in the computational neuroscience literature is the Laplace approximation [17], in which one assumes Gaussian channels and posteriors with small variance. Under these conditions, the expectations can be approximated away.

Definition 3.20. We will say that a channel c is *Gaussian* if $c(x)$ is a Gaussian measure for every x in its domain. We will denote the mean and variance of $c(x)$ by $\mu_c(x)$ and $\Sigma_c(x)$ respectively.

Proposition 3.21 (Laplace approximation)**.** Let the ambient category of channels $\mathscr{C}$ be restricted to that generated by Gaussian channels between finite-dimensional Cartesian spaces, and let $\mathscr{B}$ denote the corresponding restriction of $\mathbf{BayesLens}_2$. Suppose $(\gamma,\rho) : (X,X) \twoheadrightarrow (Y,Y)$ is such a lens, for which, for all $y : Y$ and Gaussian priors $\pi : I \to X$, the eigenvalues of $\Sigma_{\rho_\pi}(y)$ are small. Then the free energy $\mathsf{FE}(\gamma,\rho)_\pi(y)$ can be approximated by the *Laplacian free energy*

$$\begin{aligned}\mathsf{FE}(\gamma,\rho)_\pi(y) &\approx \mathsf{LFE}(\gamma,\rho)_\pi(y) &(1)\\ &:= E_{(\gamma,\pi)}\left(\mu_{\rho_\pi}(y),y\right) - S_{X\otimes M}\left[\rho_\pi(y)\right] &(2)\\ &= -\log p_\gamma(\mu_{\rho_\pi}(y),y) - \log p_\pi(\mu_{\rho_\pi}(y)|_X) - S_{X\otimes M}\left[\rho_\pi(y)\right]\end{aligned}$$

where we have written the argument of the density p_γ in 'function' style; where $(-)_X$ denotes the projection onto X; and where $S_{X\otimes M}[\rho_\pi(y)] = \mathbb{E}_{(x,m)\sim\rho_\pi(y)}[-\log p_{\rho_\pi}(x,m|y)]$ is the Shannon entropy of $\rho_\pi(y)$. The approximation is valid when Σ_{ρ_π} satisfies

$$\Sigma_{\rho_\pi}(y) = \left(\partial^2_{(x,m)} E_{(\gamma,\pi)}\right)\left(\mu_{\rho_\pi}(y),y\right)^{-1}. \qquad (3)$$

We call $E_{(\gamma,\pi)}$ the *Laplacian energy*.

Remark 3.22. The usual form of the Laplace model in the literature omits the coparameters. It is of course easy to recover the non-coparameterized form by taking $M = 1$.

Proposition 3.23. Let $\mathscr{B}$ be a subbicategory of $\mathbf{BayesLens}_2$ of Gaussian lenses whose backward channels have small variance. Then LFE defines a lax loss model $\mathscr{B} \to \mathbf{SGame}$.

Effectively, this proposition says that, under the stated conditions, the free energy and the Laplacian free energy coincide. Consequently, successfully playing a Laplacian free energy game has the same autoencoding effect.

[5] In machine learning, optimizing variational autoencoders using stochastic gradient descent typically involves a "reparameterization trick" [16, §2.5].

4 Future work

This paper only scratches the surface of the structure of statistical games. One avenue for further investigation is the link between this structure and the similar structure of diegetic open (economic) games [18], a recent reformulation of compositional game theory [19]. Notably, the composition rule for loss functions appears closely related to the Bellman equation, suggesting that algorithms for approximate inference (such as expectation-maximization) and reinforcement learning (such as backward induction) are more than superficially similar.

Another avenue for further investigation concerns mathematical neatness. First, we seek an abstract characterization of copy-composition and $\mathbf{Copara}_2$; it has been suggested to us[6] that the computation by compilers of "static single-assignment form" [20] by compilers may have a similar structure. Second, the explicit constraint defining simple coparameterized Bayesian lenses is inelegant; we expect that using dependent optics [21, 22, 23] may help to encode this constraint in the type signature, at the cost of higher-powered mathematical machinery. Finally, we seek further examples of loss models, and more abstract (and hopefully universal) characterizations of those we already have; for example, it is known that the Shannon entropy has a topological origin [24] via a "nonlinear derivation" [25], and we expect that we can follow this connection further.

5 References

1. St Clere Smithe, T.: Cyber Kittens, or Some First Steps Towards Categorical Cybernetics. In: Proceedings 3rd Annual International Applied Category Theory Conference 2020 (ACT 2020) (2020). DOI: 10.4204/EPTCS.333.8

2. St Clere Smithe, T.: Bayesian Updates Compose Optically, (2020). arXiv: 2006.01631 [math.CT]. preprint

3. Cho, K., and Jacobs, B.: Disintegration and Bayesian Inversion via String Diagrams. Math. Struct. Comp. Sci. 29 (2019) 938-971 (2017). DOI: 10.1017/S0960129518000488. arXiv: 1709.00322v3 [cs.AI]

4. Capucci, M., Gavranović, B., Hedges, J., and Rischel, E.F.: Towards Foundations of Categorical Cybernetics, (2021). arXiv: 2105.06332 [math.CT]. preprint

5. Baković, I.: Fibrations of Bicategories. (2010). https://www2.irb.hr/korisnici/ibakovic/groth2fib.pdf

6. Johnson, N., and Yau, D.: 2-Dimensional Categories, (2020). arXiv: 2002.06055 [math]. http://arxiv.org/abs/2002.06055 (visited on 11/03/2022). preprint

7. Moeller, J., and Vasilakopoulou, C.: Monoidal Grothendieck Construction. Theory and Applications of Categories **35**(31), 1159–1207 (2020). arXiv: 1809.00727v2 [math.CT]

8. Spivak, D.I.: Generalized Lens Categories via Functors $C^{op} \to$ Cat, (2019). arXiv: 1908.02202 [math.CT]. preprint

9. Jacobs, B., and Westerbaan, B.: An Effect-Theoretic Account of Lebesgue Integration. Electronic Notes in Theoretical Computer Science **319**, 239–253 (2015). DOI: 10.1016/j.entcs.2015.12.015

10. Vákár, M., and Ong, L.: On S-Finite Measures and Kernels, (2018). arXiv: 1810.01837 [math.PR]. preprint

11. nLab authors: *Essential Image*, (2023). https://ncatlab.org/nlab/show/essential+image. 05/2023

[6]This suggestion is due to Owen Lynch.

12. Jacobs, B., and Zanasi, F.: The Logical Essentials of Bayesian Reasoning, (2018). arXiv: 1804.01193 [cs.AI]. preprint

13. Kramer, M.A.: Nonlinear Principal Component Analysis Using Autoassociative Neural Networks. AIChE J. **37**(2), 233–243 (1991). DOI: 10.1002/aic.690370209. (Visited on 05/02/2023)

14. Kingma, D.P., and Welling, M.: Auto-Encoding Variational Bayes, (2013). arXiv: 1312.6114 [stat.ML]. preprint

15. Dayan, P., Hinton, G.E., Neal, R.M., and Zemel, R.S.: The Helmholtz Machine. Neural Computation **7**(5), 889–904 (1995). DOI: 10.1162/neco.1995.7.5.889. http://dx.doi.org/10.1162/neco.1995.7.5.889

16. Kingma, D.P.: Variational Inference & Deep Learning. A New Synthesis. University of Amsterdam (2017)

17. Friston, K., Mattout, J., Trujillo-Barreto, N., Ashburner, J., and Penny, W.: Variational Free Energy and the Laplace Approximation. Neuroimage **34**(1), 220–234 (2007). DOI: 10.1016/j.neuroimage.2006.08.035. pmid: 17055746

18. Capucci, M.: Diegetic Representation of Feedback in Open Games, (2022). arXiv: 2206.12338 [cs.GT]. preprint

19. Ghani, N., Hedges, J., Winschel, V., and Zahn, P.: Compositional Game Theory. Proceedings of Logic in Computer Science (LiCS) 2018 (2016). arXiv: 1603.04641 [cs.GT]

20. Kelsey, R.A.: A Correspondence between Continuation Passing Style and Static Single Assignment Form. In: Papers from the 1995 ACM SIGPLAN Workshop on Intermediate Representations, pp. 13–22. ACM, San Francisco California USA (1995). DOI: 10.1145/202529.202532. https://dl.acm.org/doi/10.1145/202529.202532 (visited on 05/02/2023)

21. Braithwaite, D., Capucci, M., Gavranović, B., Hedges, J., and Rischel, E.F.: Fibre Optics, (2021). arXiv: 2112.11145 [math.CT]. preprint

22. Vertechi, P.: Dependent Optics, (2022). arXiv: 2204.09547 [math.CT]. preprint

23. Capucci, M.: Seeing Double through Dependent Optics, (2022). arXiv: 2204.10708 [math.CT]. preprint

24. Bradley, T.-D.: Entropy as a Topological Operad Derivation. Entropy **23**(9), 1195 (2021). DOI: 10.3390/e23091195. https://www.mdpi.com/1099-4300/23/9/1195 (visited on 05/02/2023)

25. Leinster, T.: Entropy and Diversity: The Axiomatic Approach, (2022). arXiv: 2012.02113 [cs, math, q-bio]. http://arxiv.org/abs/2012.02113 (visited on 05/02/2023). preprint

26. St Clere Smithe, T.: Mathematical Foundations for a Compositional Account of the Bayesian Brain, (2022). arXiv: 2212.12538 [cs, math, q-bio, stat]. http://arxiv.org/abs/2212.12538 (visited on 01/09/2023). preprint

A State-dependent channels

In this section, we review the indexed category $\mathsf{Stat} : \mathscr{C}^{\mathrm{op}} \to \mathbf{Cat}$ of state-dependent channels in $\mathscr{C}$, from which Bayesian lenses are obtained. We can think of Stat as a decategorified, non-coparameterized, version of Stat_2, in which the hom-sets $\mathsf{Stat}(X)(A,B)$ of each fibre are given by $\mathbf{Set}\big(\mathscr{C}(I,X), \mathscr{C}(A,B)\big)$. Reindexing is again by pre-composition, although simplified as there are now no coparameters to discard.

Definition A.1. Let $(\mathscr{C}, \otimes, I)$ be a monoidal category. Define the *$\mathscr{C}$-state-indexed* category $\mathsf{Stat} : \mathscr{C}^{\mathrm{op}} \to$

Cat as follows.

$$
\begin{aligned}
\mathsf{Stat} \ : \ \mathscr{C}^{\mathrm{op}} \ \rightarrow \ \mathbf{Cat}
\end{aligned}
$$

$$
X \mapsto \mathsf{Stat}(X) := \begin{pmatrix} & \mathsf{Stat}(X)_0 & := & \mathscr{C}_0 \\ & \mathsf{Stat}(X)(A,B) & := & \mathbf{Set}\big(\mathscr{C}(I,X), \mathscr{C}(A,B)\big) \\ \mathsf{id}_A : & \mathsf{Stat}(X)(A,A) & := & \begin{cases} \mathsf{id}_A : \mathscr{C}(I,X) \to \mathscr{C}(A,A) \\ \qquad \rho \quad \mapsto \quad \mathsf{id}_A \end{cases} \end{pmatrix} \tag{4}
$$

$$
f : \mathscr{C}(Y,X) \mapsto \begin{pmatrix} \mathsf{Stat}(f) : & \mathsf{Stat}(X) & \to & \mathsf{Stat}(Y) \\ & \mathsf{Stat}(X)_0 & = & \mathsf{Stat}(Y)_0 \\ & \mathbf{Set}(\mathscr{C}(I,X), \mathscr{C}(A,B)) & \to & \mathbf{Set}\big(\mathscr{C}(I,Y), \mathscr{C}(A,B)\big) \\ & \alpha & \mapsto & f^*\alpha : \big(\sigma : \mathscr{C}(I,Y)\big) \mapsto \big(\alpha(f \bullet \sigma) : \mathscr{C}(A,B)\big) \end{pmatrix}
$$

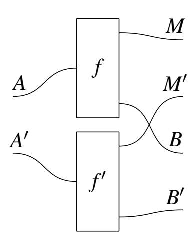
Composition in each fibre $\mathsf{Stat}(X)$ is as in $\mathscr{C}$. Explicitly, indicating morphisms $\mathscr{C}(I,X) \to \mathscr{C}(A,B)$ in $\mathsf{Stat}(X)$ by $A \overset{X}{\to\!\!\!\!\bullet} B$, and given $\alpha : A \overset{X}{\to\!\!\!\!\bullet} B$ and $\beta : B \overset{X}{\to\!\!\!\!\bullet} C$, their composite is $\beta \circ \alpha : A \overset{X}{\to\!\!\!\!\bullet} C := \rho \mapsto \beta(\rho) \bullet \alpha(\rho)$, where here we indicate composition in $\mathscr{C}$ by $\bullet$ and composition in the fibres $\mathsf{Stat}(X)$ by $\circ$. Given $f : Y \to\!\!\!\!\bullet X$ in $\mathscr{C}$, the induced functor $\mathsf{Stat}(f) : \mathsf{Stat}(X) \to \mathsf{Stat}(Y)$ acts by pre-composition.

The category of non-coparameterized Bayesian lenses is then obtained as the (1-categorical) Grothendieck construction of the pointwise opposite of Stat, following Spivak [8].

B Monoidal statistical games

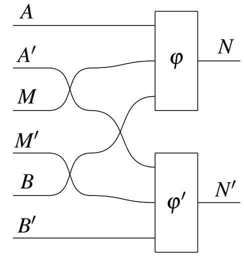
In this section, we exhibit the monoidal structures on copy-composite Bayesian lenses, statistical games, and loss models, as well as demonstrating that each of our loss models is accordingly monoidal. We begin by expressing the monoidal structure on $\mathbf{Copara}_2(\mathscr{C})$.

Proposition B.1. If the monoidal structure on $\mathscr{C}$ is symmetric, then $\mathbf{Copara}_2(\mathscr{C})$ inherits a monoidal structure $(\otimes, I)$, with the same unit object I as in $\mathscr{C}$. On 1-cells $f : A \underset{M}{\longrightarrow} B$ and $f' : A' \underset{M'}{\longrightarrow} B'$, the tensor $f \otimes f' : A \otimes A' \underset{M \otimes M'}{\longrightarrow} B \otimes B'$ is defined by



.

On 2-cells $\varphi : f \Rightarrow g$ and $\varphi' : f' \Rightarrow g'$, the tensor $\varphi \otimes \varphi' : (f \otimes f') \Rightarrow (g \otimes g')$ is given by the string diagram



.

Next, we define the notion of monoidal indexed bicategory.

Definition B.2. Suppose $(\mathscr{B}, \otimes, I)$ is a monoidal bicategory. We will say that $F : \mathscr{B}^{\mathrm{coop}} \to \mathbf{Bicat}$ is a *monoidal indexed bicategory* when it is equipped with the structure of a weak monoid object in the 3-category of indexed bicategories, indexed pseudofunctors, indexed pseudonatural transformations, and indexed modifications.

More explicitly, we will take F to be a monoidal indexed bicategory when it is equipped with

(i) an indexed pseudofunctor $\mu : F(-) \times F(=) \to F(- \otimes =)$ called the *multiplication*, *i.e.*,
   (a) a family of pseudofunctors $\mu_{X,Y} : FX \times FY \to F(X \otimes Y)$, along with
   (b) for any 1-cells $f : X \to X'$ and $g : Y \to Y'$ in $\mathscr{B}$, a pseudonatural isomorphism $\mu_{f,g} : \mu_{X',Y'} \circ (Ff \times Fg) \Rightarrow F(f \otimes g) \circ \mu_{X,Y}$;

(ii) a pseudofunctor $\eta : \mathbf{1} \to FI$ called the *unit*;

as well as three indexed pseudonatural isomorphisms — an associator, a left unitor, and a right unitor — which satisfy weak analogues of the coherence conditions for a monoidal indexed category [7, §3.2], up to invertible indexed modifications.

Using this notion, we can establish that Stat_2 is monoidal. (So as to demonstrate the structure, we do not defer the proof sketch.)

Theorem B.3. Stat_2 is a monoidal indexed bicategory.

Proof sketch. The multiplication μ is given first by the family of pseudofunctors $\mu_{X,Y} : \mathsf{Stat}_2(X) \times \mathsf{Stat}_2(Y) \to \mathsf{Stat}_2(X \otimes Y)$ which are defined on objects simply by tensor

$$\mu_{X,Y}(A,B) = A \otimes B$$

since the objects do not vary between the fibres of Stat_2, and on hom categories by the functors

$$\begin{aligned}
&\mathsf{Stat}_2(X)(A,B) \times \mathsf{Stat}_2(Y)(A',B') \\
&= \mathbf{Cat}\big(\mathsf{disc}\,\mathscr{C}(I,X), \mathbf{Copara}_2^r(\mathscr{C})(A,B)\big) \times \mathbf{Cat}\big(\mathsf{disc}\,\mathscr{C}(I,Y), \mathbf{Copara}_2^r(\mathscr{C})(A',B')\big) \\
&\cong \mathbf{Cat}\big(\mathsf{disc}\,\mathscr{C}(I,X) \times \mathsf{disc}\,\mathscr{C}(I,Y), \mathbf{Copara}_2^r(\mathscr{C})(A,B) \times \mathbf{Copara}_2^r(\mathscr{C})(A',B')\big) \\
&\xrightarrow{\mathbf{Cat}(\mathsf{disc}\,\mathscr{C}(I,\mathsf{proj}_X) \times \mathsf{disc}\,\mathscr{C}(I,\mathsf{proj}_Y), \otimes)} \mathbf{Cat}\big(\mathsf{disc}\,\mathscr{C}(I,X \otimes Y)^2, \mathbf{Copara}_2^r(\mathscr{C})(A \otimes A', B \otimes B')\big) \\
&\xrightarrow{\mathbf{Cat}(\blacktriangledown, \mathsf{id})} \mathbf{Cat}(\mathsf{disc}\,\mathscr{C}(I,X \otimes Y), \mathbf{Copara}_2^r(C)(A \otimes A', B \otimes B') \\
&= \mathsf{Stat}_2(X \otimes Y)(A \otimes A', B \otimes B') \;.
\end{aligned}$$

where $\mathbf{Cat}(\blacktriangledown, \mathsf{id})$ indicates pre-composition with the universal (Cartesian) copying functor. For all $f : X \to X'$ and $g : Y \to Y'$ in $\mathbf{Copara}_2^l(\mathscr{C})$, the pseudonatural isomorphisms

$$\mu_{f,g} : \mu_{X',Y'} \circ \big(\mathsf{Stat}_2(f) \times \mathsf{Stat}_2(g)\big) \Rightarrow \mathsf{Stat}_2(f \otimes g) \circ \mu_{X,Y}$$

are obtained from the universal property of the product $\times$ of categories. The unit $\eta : \mathbf{1} \to \mathsf{Stat}_2(I)$ is the pseudofunctor mapping the unique object of $\mathbf{1}$ to the monoidal unit I. Associativity and unitality of this monoidal structure follow from the functoriality of the construction, given the monoidal structures on $\mathscr{C}$ and $\mathbf{Cat}$. □

Just as the monoidal Grothendieck construction induces a monoidal structure on categories of lenses for monoidal pseudofunctors [7], we obtain a monoidal structure on the bicategory of copy-composite bayesian lenses.

Corollary B.4. The bicategory of copy-composite Bayesian lenses $\mathbf{BayesLens}_2$ is a monoidal bicategory. The monoidal unit is the object (I,I). The tensor $\otimes$ is given on 0-cells by $(X,A) \otimes (X',A') := (X \otimes X', A \otimes A')$, and on hom-categories by

$$\begin{aligned}
&\mathbf{BayesLens}_2\big((X,A),(Y,B)\big) \times \mathbf{BayesLens}_2\big((X,A),(Y,B)\big) \\
&= \mathbf{Copara}_2^l(\mathscr{C})(X,Y) \times \mathsf{Stat}_2(X)(B,A) \times \mathbf{Copara}_2^l(\mathscr{C})(X',Y') \times \mathsf{Stat}_2(X')(B',A') \\
&\xrightarrow{\sim} \mathbf{Copara}_2^l(\mathscr{C})(X,Y) \times \mathbf{Copara}_2^l(\mathscr{C})(X',Y') \times \mathsf{Stat}_2(X)(B,A) \times \mathsf{Stat}_2(X')(B',A') \\
&\xrightarrow{\otimes \times \mu^{\mathrm{op}}_{X,X'}} \mathbf{Copara}_2^l(\mathscr{C})(X \otimes X', Y \otimes Y') \times \mathsf{Stat}_2(X \otimes X')(B \otimes B', A \otimes A') \\
&= \mathbf{BayesLens}_2\big((X,A) \otimes (X',A'), (Y,B) \otimes (Y',B')\big) \ .
\end{aligned}$$

And similarly, we obtain a monoidal structure on statistical games.

Proposition B.5. The bicategory of copy-composite statistical games $\mathbf{SGame}$ is a monoidal bicategory. The monoidal unit is the object (I,I). The tensor $\otimes$ is given on 0-cells as for the tensor of Bayesian lenses, and on hom-categories by

$$\begin{aligned}
&\mathbf{SGame}\big((X,A),(Y,B)\big) \times \mathbf{SGame}\big((X',A'),(Y',B')\big) \\
&= \mathbf{BayesLens}_2\big((X,A),(Y,B)\big) \times \mathsf{Stat}(X)(B,I) \\
&\qquad \times \mathbf{BayesLens}_2\big((X',A'),(Y',B')\big) \times \mathsf{Stat}(X')(B',I) \\
&\xrightarrow{\sim} \mathbf{BayesLens}_2\big((X,A),(Y,B)\big) \times \mathbf{BayesLens}_2\big((X',A'),(Y',B')\big) \\
&\qquad \times \mathsf{Stat}(X)(B,I) \times \mathsf{Stat}(X')(B',I) \\
&\xrightarrow{\otimes \times \mu_{X,X'}} \mathbf{BayesLens}_2\big((X,A) \otimes (X',A'), (Y,B) \otimes (Y',B')\big) \times \mathsf{Stat}(X \otimes X')(B \otimes B', I \otimes I) \\
&\xrightarrow{\sim} \mathbf{SGame}\big((X,A) \otimes (X',A'), (Y,B) \otimes (Y',B')\big)
\end{aligned}$$

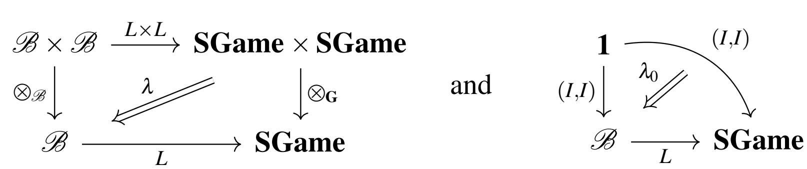
where here μ indicates the multiplication of the monoidal structure on Stat [26, Prop. 4.3.5].

We give natural definitions of monoidal inference system and monoidal loss model, which we elaborate below.

Definition B.6. A *(lax) monoidal inference system* is an inference system $(\mathscr{D}, \ell)$ for which ℓ is a lax monoidal pseudofunctor. A *(lax) monoidal loss model* is a loss model L which is a lax monoidal lax functor.

Remark B.7. We say 'lax' whenever a morphism (of any structure) only weakly preserves a monoidal operation such as composition of any order; this includes as a special case lax monoidal functors (since a monoidal category is a one-object bicategory). In this respect, we differ from [7, §2.2], who use 'weak' in the latter case; we prefer to maintain consistency. Following [6, Def. 4.2.1], we will continue to say *lax* when the witness to laxness maps composites of images to images of composites (and *oplax* when the witness maps images of composites to composites of images).

These conventions mean that a loss model $L : \mathscr{B} \to \mathbf{SGame}$ is lax monoidal when it is equipped with strong transformations



where $\otimes_{\mathscr{B}}$ and $\otimes_{\mathbf{G}}$ denote the monoidal products on $\mathscr{B} \hookrightarrow \mathbf{BayesLens}_2$ and **SGame** respectively, and when λ and λ_0 are themselves equipped with invertible modifications satisfying coherence axioms, as in Moeller and Vasilakopoulou [7, §2.2].

Note that, because L must be a (lax) section of the 2-fibration $\pi_{\mathsf{Loss}}|_{\mathscr{B}} : \mathbf{SGame}|_{\mathscr{B}} \to \mathscr{B}$, the unitor λ_0 is forced to be trivial, picking out the identity on the monoidal unit (I,I). Likewise, the laxator $\lambda : L(-)\otimes L(=) \Rightarrow L(-\otimes =)$ must have 1-cell components which are identities:

$$L(X,A)\otimes L(X',A) = (X,A)\otimes(X',A') = (X\otimes X', A\otimes A') = L\big((X,A)\otimes L(X',A)\big)$$

The interesting structure is therefore entirely in the 2-cells. We follow the convention of [6, Def. 4.2.1] that a strong transformation is a lax transformation with invertible 2-cell components. Supposing that $(c,c') : (X,A) \nrightarrow (Y,B)$ and $(d,d') : (X',A') \nrightarrow (Y',B')$ are 1-cells in $\mathscr{B}$, the corresponding 2-cell component of λ has the form $\lambda_{c,d} : L\big((c,c')\otimes(d,d')\big) \Rightarrow L(c,c')\otimes L(d,d')$, hence filling the following square in **SGame**:

$$\begin{array}{ccc} (X,A)\otimes(X',A') & \xrightarrow{L(c,c')\otimes L(d,d')} & (Y,B)\otimes(Y',B') \\ \| & \overset{\lambda_{c,d}}{\nearrow} & \| \\ (X,A)\otimes(X',A') & \xrightarrow[L((c,c')\otimes(d,d'))]{} & (Y,B)\otimes(Y',B') \end{array}$$

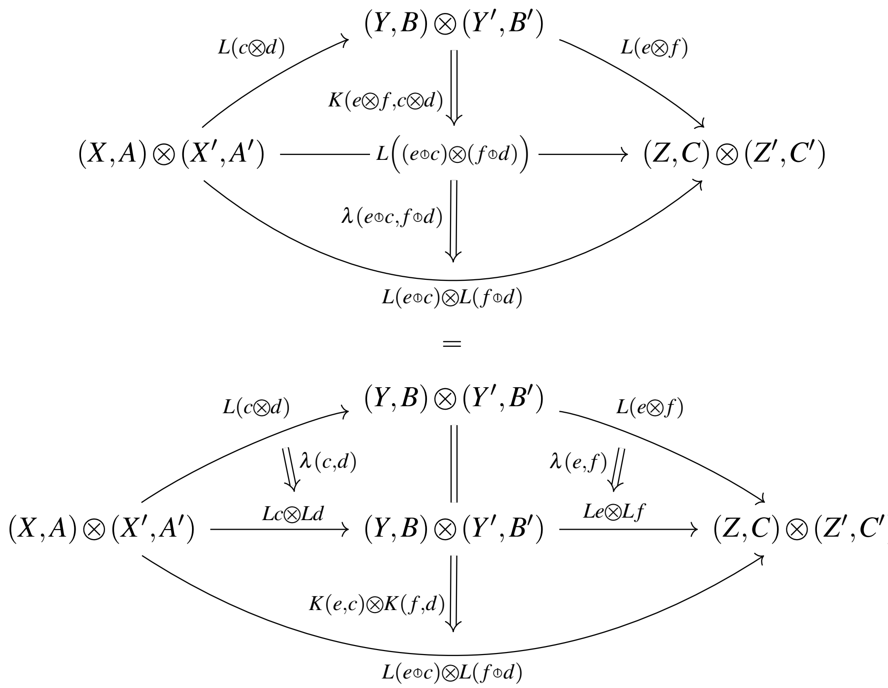
Intuitively, these 2-cells witness the failure of the tensor $L(c,c')\otimes L(d,d')$ of the parts to account for correlations that may be evident to the "whole system" $L\big((c,c')\otimes(d,d')\big)$.

Just as we have monoidal lax functors, we can have monoidal lax transformations; again, see [7, §2.2].

Proposition B.8. Monoidal loss models and monoidal icons form a subcategory $\mathsf{MonLoss}(\mathscr{B})$ of $\mathsf{Loss}(\mathscr{B})$, and the symmetric monoidal structure $(+,0)$ on the latter restricts to the former.

B.1 Examples

In this section, we present the monoidal structure on the loss models considered above. Because loss models L are (lax) sections, following Remark B.7, this monoidal structure is given in each case by a lax natural family of 2-cells $\lambda_{c,d} : L\big((c,c')\otimes(d,d')\big) \Rightarrow L(c,c')\otimes L(d,d')$, for each pair of lenses $(c,c') : (X,A) \nrightarrow (Y,B)$ and $(d,d') : (X',A') \nrightarrow (Y',B')$. Such a 2-cell $\lambda_{c,d}$ is itself given by a loss function of type $B\otimes B' \xrightarrow{X\otimes X'} I$ satisfying the equation $L\big((c,c')\otimes(d,d')\big) = L(c,c')\otimes L(d,d') + \lambda_{c,d}$. Following [6, Eq. 4.2.3], lax naturality requires that λ satisfy the following equation of 2-cells, where K denotes the laxator (with respect to horizontal composition $\diamond$) with components $K(e,c) : Le \diamond Lc \Rightarrow L(e \oplus c)$:



Since vertical composition in **SGame** is given on losses by $+$, we can write this equation as

$$\begin{aligned}&\lambda(e\oplus c, f\oplus d) + K(e\otimes f, c\otimes d)\\ &= \lambda(e,f)\diamond\lambda(c,d) + K(e,c)\otimes K(f,d)\\ &= \lambda(e,f)_{c\otimes d} + \lambda(c,d)\circ(e'\otimes f')_{c\otimes d} + K(e,c)\otimes K(f,d)\,. \end{aligned} \tag{5}$$

In each of the examples below, therefore, we establish the definition of the laxator λ and check that it satisfies equation 5.

We will often use the notation $(-)_X$ to denote projection onto a factor X of a monoidal product.

B.1.1 Relative entropy

Proposition B.9. The loss model KL of Proposition 3.14 is lax monoidal. Supposing that $(c,c') : (X,X) \twoheadrightarrow (Y,Y)$ and $(d,d') : (X',X') \twoheadrightarrow (Y',Y')$ are lenses in $\mathscr{B}$, the corresponding component $\lambda^{\mathsf{KL}}(c,d)$ of the laxator is given, for $\omega : I \to\!\!\!\bullet\, X\otimes X'$ and $(y,y') : Y\otimes Y'$, by

$$\lambda^{\mathsf{KL}}(c,d)_\omega(y,y') := \mathop{\mathbb{E}}_{\substack{(x,x',m,m')\sim\\ (c'_{\omega_X}\otimes d'_{\omega_{X'}})(y,y')}} \left[\log\frac{p_{\omega_X\otimes\omega_{X'}}(x,x')}{p_\omega(x,x')}\right] + \log\frac{p_{(c\otimes d)^\dagger\bullet\omega}(y,y')}{p_{(c\otimes d)^\dagger\bullet(\omega_X\otimes\omega_{X'})}(y,y')}\,.$$

B.1.2 Maximum likelihood estimation

Proposition B.10. The loss model MLE of Proposition 3.16 is lax monoidal. Supposing that $(c,c') : (X,X) \twoheadrightarrow (Y,Y)$ and $(d,d') : (X',X') \twoheadrightarrow (Y',Y')$ are lenses in $\mathscr{B}$, the corresponding component $\lambda^{\mathsf{MLE}}(c,d)$ of the laxator is given, for $\omega : I \to\!\!\!\bullet\, X\otimes X'$ and $(y,y') : Y\otimes Y'$, by

$$\lambda^{\mathsf{MLE}}(c,d)_\omega(y,y') := \log\frac{p_{(c\otimes d)^\dagger\bullet(\omega_X\otimes\omega_{X'})}(y,y')}{p_{(c\otimes d)^\dagger\bullet\omega}(y,y')}\,.$$

B.1.3 Free energy

Corollary B.11. The loss model FE of Definition 3.17 is lax monoidal. Supposing that $(c,c') : (X,X) \nrightarrow (Y,Y)$ and $(d,d') : (X',X') \nrightarrow (Y',Y')$ are lenses in $\mathscr{B}$, the corresponding component $\lambda^{\mathsf{FE}}(c,d)$ of the laxator is given, for $\omega : I \bullet\!\!\to X \otimes X'$ and $(y,y') : Y \otimes Y'$, by

$$\lambda^{\mathsf{FE}}(c,d)_\omega(y,y') := \mathop{\mathbb{E}}_{(x,x')\sim(c'_{\omega_X}\otimes d'_{\omega_{X'}})(y,y')}\left[\log\frac{p_{\omega_X\otimes\omega_{X'}}(x,x')}{p_\omega(x,x')}\right].$$

B.1.4 Laplacian free energy

Proposition B.12. The loss model LFE of Propositions 3.21 and 3.23 is lax monoidal. Supposing that $(c,c') : (X,X) \nrightarrow (Y,Y)$ and $(d,d') : (X',X') \nrightarrow (Y',Y')$ are lenses in $\mathscr{B}$, the corresponding component $\lambda^{\mathsf{LFE}}(c,d)$ of the laxator is given, for $\omega : I \bullet\!\!\to X \otimes X'$ and $(y,y') : Y \otimes Y'$, by

$$\lambda^{\mathsf{LFE}}(c,d)_\omega(y,y') := \log\frac{p_{\omega_X\otimes\omega_{X'}}(\mu_{(c\otimes d)'_\omega}(y,y')_{XX'})}{p_\omega(\mu_{(c\otimes d)'_\omega}(y,y')_{XX'})}$$

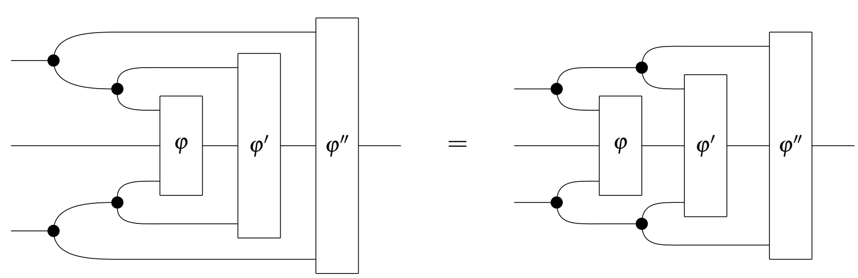
where $\mu_{(c\otimes d)'_\omega}(y,y')_{XX'}$ is the $(X\otimes X')$-mean of the Gaussian distribution $(c'_{\omega_X}\otimes d'_{\omega_{X'}})(y,y')$.

C Proofs

C.1 Proof of Theorem 2.1

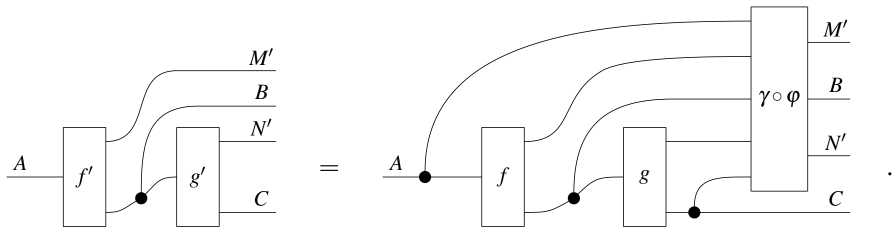
Proof. To show that $\mathbf{Copara}_2(\mathscr{C})$ is a bicategory, we need to establish the unitality and associativity of vertical composition; verify that horizontal composition is well-defined and functorial; establish the weak unitality and associativity of horizontal composition; and confirm that the corresponding unitors and associator satisfy the bicategorical coherence laws. Then, to prove that $\mathbf{Copara}_2(\mathscr{C})$ is moreover monoidal, we need to demonstrate that the tensor as proposed satisfies the data of a monoidal bicategory. However, since the monoidal structure is inherited from that of $\mathscr{C}$, we will elide much of this latter proof, and demonstrate only that the tensor is functorial; the rest follows straightforwardly but tediously.

We begin by confirming that vertical composition $\odot$ is unital and associative. To see that $\odot$ is unital, simply substitute the identity 2-cell (given by projection onto the coparameter) into the string diagram defining $\odot$ and then apply the comonoid counitality law twice (once on the left, once on the right). The associativity of $\odot$ requires that $\varphi''\odot(\varphi'\odot\varphi) = (\varphi''\odot\varphi')\odot\varphi$, which corresponds to the following graphical equation:

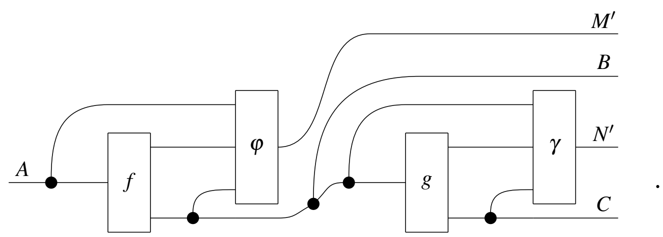


To see that this equation is satisfied, simply apply the comonoid coassociativity law twice (once left, once right).

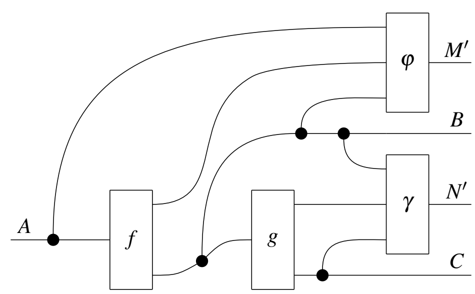
Next, we check that horizontal composition $\circ$ is well-defined, which amounts to checking whether the horizontal composite of 2-cells satisfies the change of coparameter axiom. Again, we reason graphically. Given 2-cells φ and γ between composable pairs of 1-cells f, f' and g, g', our task is to verify that



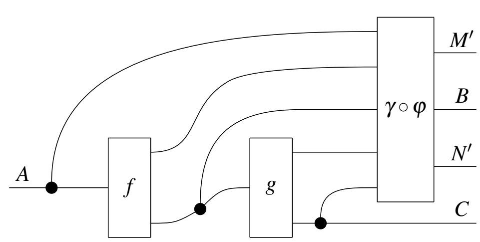
Since φ and γ satisfy change of coparameter *ex hypothesi*, the left hand side is equal to the morphism



By comonoid coassociativity, this is in turn equal to

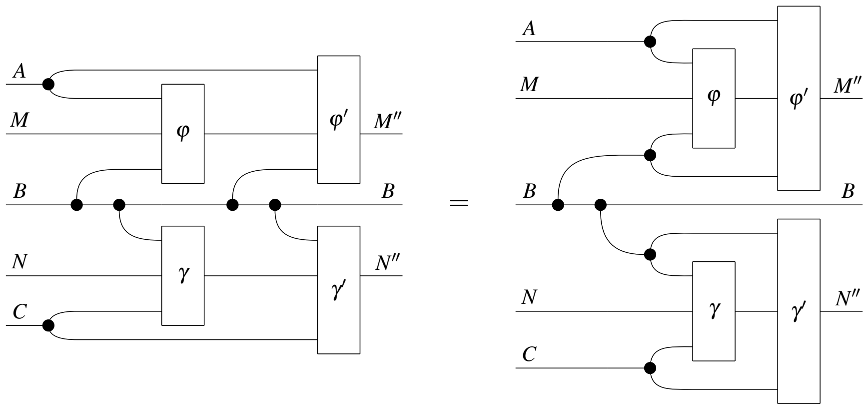


which, by the definition of $\circ$, is precisely equal to



and so this establishes the result.

We now verify that $\circ$ so defined is functorial on 2-cells, beginning with the preservation of composition. We need to validate the equation $(\gamma' \circ \varphi') \odot (\gamma \circ \varphi) = (\gamma' \odot \gamma) \circ (\varphi' \odot \varphi)$ (for appropriately composable 2-cells). This amounts to checking the following equation, which can be seen to hold by two applications of comonoid coassociativity:



It is easy to verify that $\circ$ preserves identities, *i.e.* that $\mathsf{id}_g \circ \mathsf{id}_f = \mathsf{id}_{g\circ f}$; just substitute the identity 2-cells into the definition of $\circ$ on 2-cells, and apply comonoid counitality four times.

Next, we establish that horizontal composition is weakly associative, which requires us to supply isomorphisms $\alpha_{f,g,h} : (h \circ g) \circ f \Rightarrow h \circ (g \circ f)$ natural in composable triples of 1-cells h, g, f. Supposing the three morphisms have the types $f : A \underset{M}{\longrightarrow} B$, $g : B \underset{N}{\longrightarrow} C$, and $h : C \underset{O}{\longrightarrow} D$, we can choose $a_{f,g,h}$ to be the 2-cell represented by the morphism

$$A \otimes \big((M \otimes B) \otimes ((N \otimes C) \otimes O)\big) \otimes D \xrightarrow{\mathsf{proj}} (M \otimes B) \otimes ((N \otimes C) \otimes O) \cdots$$
$$\cdots \xrightarrow{\alpha^{\mathscr{C}}_{(M\otimes B),(N\otimes C),O}} ((M \otimes B) \otimes (N \otimes C)) \otimes O \cdots$$
$$\cdots \xrightarrow{\alpha^{\mathscr{C}}_{(M\otimes B),N,C} \otimes \mathsf{id}_O} (((M \otimes B) \otimes N) \otimes C) \otimes O$$

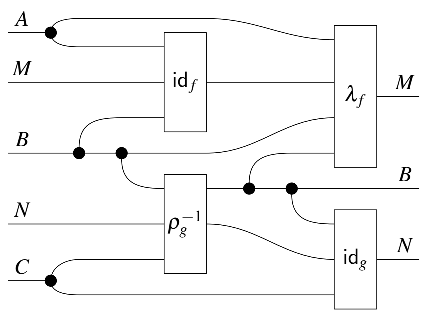
where the first factor is the projection onto the coparameter and $\alpha^{\mathscr{C}}$ denotes the associator of the monoidal structure $(\otimes, I)$ on $\mathscr{C}$. In the inverse direction, we can choose the component $\alpha^{-1}_{f,g,h} : h \circ (g \circ f) \Rightarrow (h \circ g) \circ f$ to be the 2-cell represented by the morphism

$$A \otimes \big((((M \otimes B) \otimes N) \otimes C) \otimes O\big) \otimes D \xrightarrow{\mathsf{proj}} (((M \otimes B) \otimes N) \otimes C) \otimes O \cdots$$
$$\cdots \xrightarrow{\alpha^{\mathscr{C},-1}_{(M\otimes B),N,C} \otimes \mathsf{id}_O} ((M \otimes B) \otimes (N \otimes C)) \otimes O \cdots$$
$$\cdots \xrightarrow{\alpha^{\mathscr{C},-1}_{(M\otimes B),(N\otimes C),O}} (M \otimes B) \otimes ((N \otimes C) \otimes O)$$

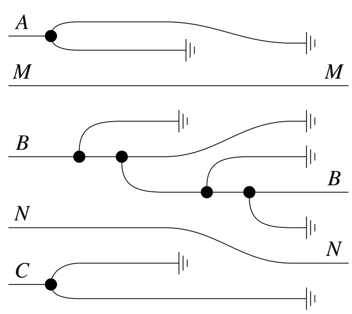
where $\alpha^{\mathscr{C},-1}$ denotes the inverse of the associator on $(\mathscr{C}, \otimes, I)$. That the pair of $\alpha_{f,g,h}$ and $\alpha^{-1}_{f,g,h}$ constitutes an isomorphism in the hom category follows from the counitality of the comonoid structures. That this family of isomorphisms is moreover natural follows from the naturality of the associator on $(\mathscr{C}, \otimes, I)$.

We come to the matter that motivated the construction of $\mathbf{Copara}_2(\mathscr{C})$: the weak unitality of copy-composition, witnessed here by the weak unitality of horizontal composition. We need to exhibit two families of natural isomorphisms: the left unitors with components $\lambda_f : \mathsf{id}_B \circ f \Rightarrow f$, and the right unitors with components $\rho_f : f \circ \mathsf{id}_A \Rightarrow f$, for each morphism $f : A \underset{M}{\longrightarrow} B$. Each such component will be defined by a projection morphism, and weak unitality will then follow from the counitality of the comonoid structures. More explicitly, λ_f is witnessed by $\mathsf{proj}_M : A \otimes M \otimes B \otimes B \to M$; its inverse λ_f^{-1} is witnessed by $\mathsf{proj}_{M\otimes B} : A \otimes M \otimes B \to M \otimes B$; ρ_f is witnessed by $\mathsf{proj}_M : A \otimes A \otimes M \otimes B \to M$; and its inverse ρ_f^{-1} is witnessed by $\mathsf{proj}_{A\otimes M} : A\otimes M\otimes B \to A\otimes M$. Checking that these definitions give natural isomorphisms is then an exercise in counitality that we leave to the reader.

All that remains of the proof that $\mathbf{Copara}_2(\mathscr{C})$ is indeed a bicategory is to check that the unitors are compatible with the associator (*i.e.*, $(\mathsf{id}_g \circ \lambda_f) \odot \alpha_{g,\mathsf{id}_B,f} = \rho_g \circ \mathsf{id}_f$) and that associativity is order-dependent (*i.e.*, the associator α satisfies the pentagon diagram). The latter follows immediately from the corresponding fact about the associator $\alpha^{\mathscr{C}}$ on $(\mathscr{C},\otimes,I)$. To demonstrate the former, it is easier to verify that $(\mathsf{id}_g \circ \lambda_f) \odot \alpha_{g,\mathsf{id}_B,f} \odot (\rho_g^{-1} \circ \mathsf{id}_f) = \mathsf{id}_{g\circ f}$. This amounts to checking that the following string diagram is equally a depiction of the morphism underlying $\mathsf{id}_{g\circ f}$:



(Note that here we have elided the associator from the depiction. This is allowed by comonoid counitality, and because string diagrams are blind to bracketing.) Substituting the relevant morphisms into the boxes, we see that this diagram is equal to

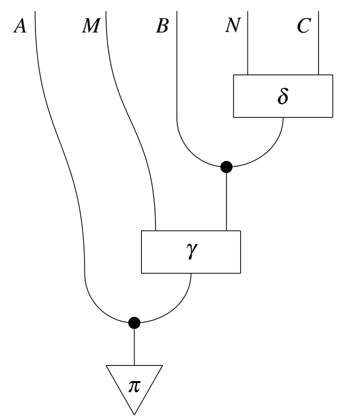


and six applications of counitality give us $\mathsf{id}_{g\circ f}$. This establishes that $\mathbf{Copara}_2(\mathscr{C})$ is a bicategory. □

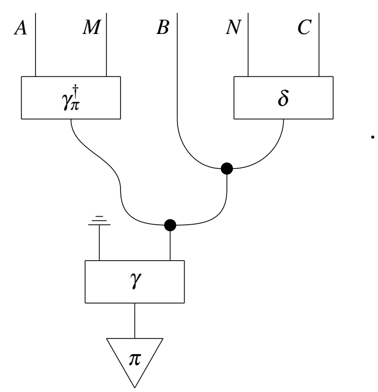
C.2 Proof of Theorem 2.12

Proof. We only need to show that $\gamma_\pi^\dagger \bullet \delta_{\gamma\pi}^\dagger$ is a Bayesian inversion of $\delta \bullet \gamma$ with respect to π; the 'moreover' claim follows immediately because Bayesian inversions are almost surely unique [26, Prop. 4.1.28]. Thus,

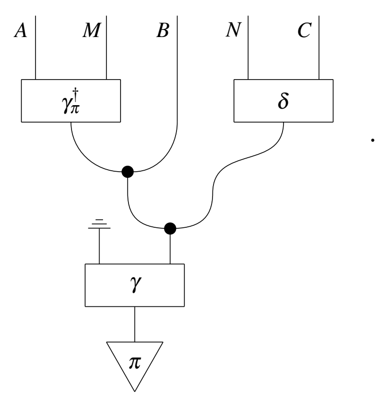
$\delta \bullet \gamma \bullet \pi$ has the following depiction;



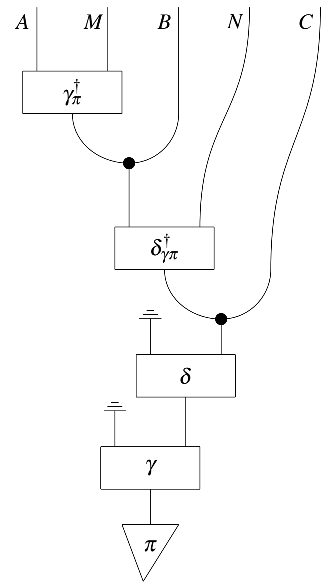
Since $\gamma^\dagger_\pi$ is a Bayesian inversion of γ with respect to π, this is equal to



By the coassociativity of copying, this in turn is equal to



And since $\delta^\dagger_{\gamma\pi}$ is a Bayesian inversion of δ with respect to $(\gamma\pi)^{\bar{\top}}$, this is equal to

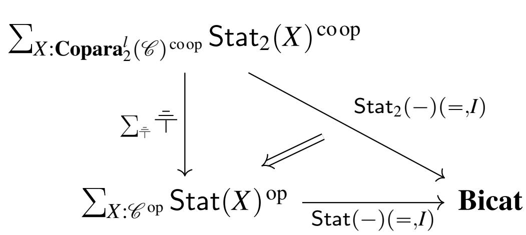


which establishes the result. □

C.3 Proofs about statistical games (§3)

Proof of Theorem 3.3. Recall that every monoidal category $\mathscr{M}$ can be transformed into a one-object bicategory, its *delooping* $\mathbf{B}\mathscr{M}$, with the 1-cells and 2-cells being the objects and morphisms of $\mathscr{M}$, vertical composition being composition in $\mathscr{M}$, and horizontal composition being the tensor. This 'delooping' is functorial, giving a 2-functor $\mathbf{B} : \mathbf{MonCat} \to \mathbf{Bicat}$ which, following Corollary 3.5, we can compose after $\mathsf{Stat}(-)(=,I)$ (taking its domain as a locally discrete 2-category) to obtain indexed bicategories; we will assume this transformation henceforth.

Next, observe that we can extend the domain of $\mathsf{Stat}(-)(=,I)$ to $\sum_{X:\mathbf{Copara}^l_2(\mathscr{C})^{\mathrm{coop}}} \mathsf{Stat}_2(X)^{\mathrm{coop}}$ by discarding the coparameters of the (coparameterized) state-dependent channels as well as the coparameter on any reindexing, as in the following diagram of indexed bicategories:



Here, the 2-cell indicates also discarding the coparameters of the 'effects' in $\mathsf{Stat}_2(-)(=,I)$.

If we let L denote the composite functor in the diagram above, we can reason as follows:

$$\dfrac{\dfrac{\dfrac{\dfrac{L : \sum_{X:\mathbf{Copara}_2^l(\mathscr{C})^{\mathrm{co\,op}}} \mathsf{Stat}_2(X)^{\mathrm{co\,op}} \to \mathbf{Bicat}}{\prod_{X:\mathbf{Copara}_2^l(\mathscr{C})^{\mathrm{co\,op}}} \mathbf{Bicat}^{\mathsf{Stat}_2(X)^{\mathrm{co\,op}}}}\ \text{sum/product}}{\prod_{X:\mathbf{Copara}_2^l(\mathscr{C})^{\mathrm{co\,op}}} \mathbf{2Fib}\left(\mathsf{Stat}_2(X)\right)}\ \prod \textstyle\int}{\mathbf{Copara}_2^l(\mathscr{C})^{\mathrm{co\,op}} \to \mathbf{Bicat}}\ \text{forget}}{G : \mathbf{Copara}_2^l(\mathscr{C})^{\mathrm{co\,op}} \to \mathbf{Bicat}}\ \text{op}$$

where the first step uses the adjointness of (dependent) sums and products; the second applies the bicategorical Grothendieck construction in the codomain; the third forgets the 2-fibrations, to leave only the total bicategory; and the fourth step takes the pointwise opposite. We can thus write the action of G as $G(X) = \left(\int L(X,-)\right)^{\mathrm{op}}$.

Since each bicategory $L(X,B)$ has only a single 0-cell, the 0-cells of each $G(X)$ are equivalently just the objects of $\mathscr{C}$, and the hom-categories $G(X)(A,B)$ are equivalent to the product categories $\mathsf{Stat}_2(X)(B,A) \times \mathsf{Stat}(X)(B,I)$. That is to say, a 1-cell $A \to B$ in $G(X)$ is a pair of a state-dependent channel $B \overset{X}{\nrightarrow} A$ along with a correspondingly state-dependent effect on its domain B. We therefore seem to approach the notion of statistical game, but in fact we are already there: $\mathbf{SGame}_{\mathscr{C}}$ is simply $\int G$, by the bicategorical Grothendieck construction. To see this is only a matter of further unfolding the definition. □

Proof sketch for Proposition 3.10. From [6, Theorem 4.6.13], we have that icons compose, giving a category. Then note that, for any two loss models F and G and any n-cell α, $F(\alpha)$ and $G(\alpha)$ must only differ on the loss component, and so we can sum the losses; this gives the monoidal product. The monoidal unit is necessarily the constant 0 loss. Finally, observe that the structure is symmetric becauase effect monoids are by Definition 3.1 commutative. □

Proof of Proposition 3.14. Being a section of $\pi_{\mathsf{Loss}}|_{\mathscr{B}}$, KL leaves lenses unchanged, only acting to attach loss functions. It therefore suffices to check that this assignment of losses is strictly functorial. Writing $\bullet$ for composition in $\mathscr{C}$, $\circ$ for horizontal composition in Stat_2, φ in $\mathbf{BayesLens}_2$, and $\diamond$ for horizontal composition of losses in $\mathbf{SGame}$, we have the following chain of equalities:

$$\begin{aligned}
\mathsf{KL}\big((d,d')\varphi(c,c')\big)_\pi(z) &= \mathop{\mathbb{E}}_{(x,m,y,n)\sim(c'\circ d'_c)_\pi(z)}\Big[\log p_{(c'\circ d'_c)_\pi}(x,m,y,n|z) \\
&\qquad\qquad - \log p_{(c^\dagger\circ d_c^\dagger)_\pi}(x,m,y,n|z)\Big] \\
&= \mathop{\mathbb{E}}_{(y,n)\sim d'_{c\bullet\pi}(z)}\ \mathop{\mathbb{E}}_{(x,m)\sim c'_\pi(y)}\Big[\log p_{c'_\pi}(x,m|y)\, p_{d'_{c\bullet\pi}}(y,n|z) \\
&\qquad\qquad - \log p_{c^\dagger_\pi}(x,m|y)\, p_{d^\dagger_{c\bullet\pi}}(y,n|z)\Big] \\
&= \mathop{\mathbb{E}}_{(y,n)\sim d'_{c\bullet\pi}(z)}\Big[\log p_{d'_{c\bullet\pi}}(y,n|z) - \log p_{d^\dagger_{c\bullet\pi}}(y,n|z) \\
&\qquad\qquad + \mathop{\mathbb{E}}_{(x,m)\sim c'_\pi(y)}\Big[\log p_{c'_\pi}(x,m|y) - \log p_{c^\dagger_\pi}(x,m|y)\Big]\Big] \\
&= D_{KL}\big(d'_{c\bullet\pi}(z), d^\dagger_{\bullet\pi}(z)\big) + \mathop{\mathbb{E}}_{(y,n)\sim d'_{c\bullet\pi}(z)}\Big[D_{KL}\big(c'_\pi(y), c^\dagger_\pi(y)\big)\Big] \\
&= \mathsf{KL}(d,d')_{c\bullet\pi}(z) + \big(\mathsf{KL}(c,c')\circ d'_c\big)_\pi(z) \\
&= \big(\mathsf{KL}(d,d')\diamond\mathsf{KL}(c,c')\big)_\pi(z)
\end{aligned}$$

The first line obtains by definition of KL and $\oplus$; the second by definition of $\circ$; the third by the log adjunction ($\log ab = \log a + \log b$) and by linearity of $\mathbb{E}$; the fourth by definition of D_{KL}; the fifth by definition of KL and of $\circ$; and the sixth by definition of $\diamond$.

This establishes that $\mathsf{KL}\big((d,d')\oplus(c,c')\big) = \mathsf{KL}(d,d')\diamond\mathsf{KL}(c,c')$ and hence that KL is strictly functorial on 1-cells. Since we have assumed that the only 2-cells are the structural 2-cells (*e.g.*, the horizontal unitors), which do not result in any difference between the losses assigned to the corresponding 1-cells, the only loss 2-cell available to be assigned is the 0 loss; which assignment is easily seen to be vertically functorial. Hence KL is a strict 2-functor, and moreover a section of $\pi_{\mathsf{Loss}}|_{\mathscr{B}}$ as required. □

Proof of Proposition 3.16. We adopt the notational conventions of the proof of Proposition 3.14. Observe that

$$\mathsf{MLE}\big((d,d')\oplus(c,c')\big)_\pi(z) = -\log p_{d^\dagger \bullet c^\dagger \bullet \pi}(z) = \mathsf{MLE}(d,d')_{c\bullet\pi}(z)\ .$$

By definition, we have

$$\big(\mathsf{MLE}(d,d')\diamond\mathsf{MLE}(c,c')\big)_\pi(z) = \mathsf{MLE}(d,d')_{c\bullet\pi}(z) + \big(\mathsf{MLE}(c,c')\circ d'_c\big)_\pi(z)$$

and hence by substitution

$$\big(\mathsf{MLE}(d,d')\diamond\mathsf{MLE}(c,c')\big)_\pi(z) = \mathsf{MLE}\big((d,d')\oplus(c,c')\big)_\pi(z) + \big(\mathsf{MLE}(c,c')\circ d'_c\big)_\pi(z)\ .$$

Therefore, $\mathsf{MLE}(c,c')\circ d'_c$ constitutes a 2-cell from $\mathsf{MLE}(d,d')\diamond\mathsf{MLE}(c,c')$ to $\mathsf{MLE}\big((d,d')\oplus(c,c')\big)$, and hence MLE is a lax functor. It is evidently moreover a section of $\pi_{\mathsf{Loss}}|_{\mathscr{B}}$, and, like KL, acts trivially on the (purely structural) 2-cells. □

Proof of Lemma 3.21. We can write the density functions of Gaussian channels as:

$$\begin{aligned}
\log p_\gamma(m,y|x) &= \frac{1}{2}\big\langle \varepsilon_\gamma, {\Sigma_\gamma}^{-1}\varepsilon_\gamma\big\rangle - \log\sqrt{(2\pi)^{|Y|}\det\Sigma_\gamma} \\
\log p_{\rho_\pi}(x,m|y) &= \frac{1}{2}\big\langle \varepsilon_{\rho_\pi}, {\Sigma_{\rho_\pi}}^{-1}\varepsilon_{\rho_\pi}\big\rangle - \log\sqrt{(2\pi)^{|X|}\det\Sigma_{\rho_\pi}} \\
\log p_\pi(x) &= \frac{1}{2}\big\langle \varepsilon_\pi, {\Sigma_\pi}^{-1}\varepsilon_\pi\big\rangle - \log\sqrt{(2\pi)^{|X|}\det\Sigma_\pi}
\end{aligned} \tag{6}$$

where for clarity we have omitted the dependence of Σ_γ on x and Σ_{ρ_π} on y, and where

$$\begin{aligned}
\varepsilon_\gamma &:= (m,y) - \mu_\gamma(x)\,, \\
\varepsilon_{\rho_\pi} &:= (x,m) - \mu_{\rho_\pi}(y)\,, \\
\varepsilon_\pi &:= x - \mu_\pi\,.
\end{aligned} \tag{7}$$

Then, recall that we can write the free energy $\mathsf{FE}(\gamma,\rho)_\pi(y)$ as the difference between expected energy and entropy:

$$\begin{aligned}
\mathsf{FE}(\gamma,\rho)_\pi(y) &= \underset{(x,m)\sim\rho_\pi(y)}{\mathbb{E}}\big[-\log p_\gamma(m,y|x) - \log p_\pi(x)\big] - S_{X\otimes M}\big[\rho_\pi(y)\big] \\
&= \underset{(x,m)\sim\rho_\pi(y)}{\mathbb{E}}\big[E_{(\gamma,\pi)}(x,m,y)\big] - S_X\big[\rho_\pi(y)\big]
\end{aligned}$$

Next, since the eigenvalues of $\Sigma_{\rho_\pi}(y)$ are small for all $y : Y$, we can approximate the expected energy by its second-order Taylor expansion around the mean $\mu_{\rho_\pi}(y)$:

$$\begin{aligned}
\mathsf{FE}(\gamma,\rho)_\pi(y) \approx \mathop{\mathbb{E}}_{(x,m)\sim\rho_\pi(y)} \Bigg[& E_{(\gamma,\pi)}(\mu_{\rho_\pi}(y),y) + \left\langle \varepsilon_{\rho_\pi}(x,m,y),\, \left(\partial_{(x,m)} E_{(\gamma,\pi)}\right)\left(\mu_{\rho_\pi}(y),y\right) \right\rangle \\
& + \frac{1}{2} \left\langle \varepsilon_{\rho_\pi}(x,m,y),\, \left(\partial^2_{(x,m)} E_{(\gamma,\pi)}\right)\left(\mu_{\rho_\pi}(y),y\right) \cdot \varepsilon_{\rho_\pi}(x,m,y) \right\rangle \Bigg] \\
& - S_{X\otimes M}\left[\rho_\pi(y)\right] \\
\overset{(a)}{=} & E_{(\gamma,\pi)}(\mu_{\rho_\pi}(y),y) + \left\langle \mathop{\mathbb{E}}_{(x,m)\sim\rho_\pi(y)} \left[\varepsilon_{\rho_\pi}(x,m,y)\right],\, \left(\partial_{(x,m)} E_{(\gamma,\pi)}\right)\left(\mu_{\rho_\pi}(y),y\right) \right\rangle \\
& + \frac{1}{2} \operatorname{tr}\left[\left(\partial^2_{(x,m)} E_{(\gamma,\pi)}\right)\left(\mu_{\rho_\pi}(y),y\right) \Sigma_{\rho_\pi}(y) \right] - S_{X\otimes M}\left[\rho_\pi(y)\right] \\
\overset{(b)}{=} & E_{(\gamma,\pi)}(\mu_{\rho_\pi}(y),y) + \frac{1}{2} \operatorname{tr}\left[\left(\partial^2_{(x,m)} E_{(\gamma,\pi)}\right)\left(\mu_{\rho_\pi}(y),y\right) \Sigma_{\rho_\pi}(y) \right] - S_{X\otimes M}\left[\rho_\pi(y)\right]
\end{aligned}$$

where $\left(\partial^2_{(x,m)} E_{(\gamma,\pi)}\right)\left(\mu_{\rho_\pi}(y),y\right)$ is the Hessian of $E_{(\gamma,\pi)}$ with respect to (x,m) evaluated at $(\mu_{\rho_\pi}(y),y)$. The equality marked (a) holds first by the linearity of expectations and second because

$$\begin{aligned}
& \mathop{\mathbb{E}}_{(x,m)\sim\rho_\pi(y)} \left[\left\langle \varepsilon_{\rho_\pi}(x,m,y),\, \left(\partial^2_{(x,m)} E_{(\gamma,\pi)}\right)\left(\mu_{\rho_\pi}(y),y\right) \cdot \varepsilon_{\rho_\pi}(x,m,y) \right\rangle \right] \\
&= \mathop{\mathbb{E}}_{(x,m)\sim\rho_\pi(y)} \left[\operatorname{tr}\left[\left(\partial^2_{(x,m)} E_{(\gamma,\pi)}\right)\left(\mu_{\rho_\pi}(y),y\right) \varepsilon_{\rho_\pi}(x,m,y)\,\varepsilon_{\rho_\pi}(x,m,y)^T \right] \right] \\
&= \operatorname{tr}\left[\left(\partial^2_{(x,m)} E_{(\gamma,\pi)}\right)\left(\mu_{\rho_\pi}(y),y\right) \mathop{\mathbb{E}}_{(x,m)\sim\rho_\pi(y)} \left[\varepsilon_{\rho_\pi}(x,m,y)\,\varepsilon_{\rho_\pi}(x,m,y)^T \right] \right] \\
&= \operatorname{tr}\left[\left(\partial^2_{(x,m)} E_{(\gamma,\pi)}\right)\left(\mu_{\rho_\pi}(y),y\right) \Sigma_{\rho_\pi}(y) \right] \qquad (8)
\end{aligned}$$

where the first equality obtains because the trace of an outer product equals an inner product; the second by linearity of the trace; and the third by the definition of the covariance $\Sigma_{\rho_\pi}(y)$. The equality marked (b) above then holds because $\mathbb{E}_{(x,m)\sim\rho_\pi(y)}\left[\varepsilon_{\rho_\pi}(x,m,y)\right] = 0$.

Next, note that the entropy of a Gaussian measure depends only on its covariance,

$$S_{X\otimes M}\left[\rho_\pi(y)\right] = \frac{1}{2} \log\det\left(2\pi e\, \Sigma_{\rho_\pi}(y)\right) ,$$

and that the energy $E_{(\gamma,\pi)}(\mu_{\rho_\pi}(y),y)$ does not depend on $\Sigma_{\rho_\pi}(y)$. We can therefore write down directly the covariance $\Sigma^*_{\rho_\pi}(y)$ minimizing $\mathsf{FE}(\gamma,\rho)_\pi(y)$ as a function of y. We have

$$\partial_{\Sigma_{\rho_\pi}} \mathsf{FE}(\gamma,\rho)_\pi(y) \overset{(b)}{\approx} \frac{1}{2}\left(\partial^2_{(x,m)} E_{(\gamma,\pi)}\right)\left(\mu_{\rho_\pi}(y),y\right) + \frac{1}{2} {\Sigma_{\rho_\pi}}^{-1}$$

by equation (b) above. Setting $\partial_{\Sigma_{\rho_\pi}} \mathsf{FE}(\gamma,\rho)_\pi(y) = 0$, we find the optimum as expressed by equation (3):

$$\Sigma^*_{\rho_\pi}(y) = \left(\partial^2_x E_{(\gamma,\pi)}\right)\left(\mu_{\rho_\pi}(y),y\right)^{-1} .$$

Finally, by substituting $\Sigma^*_{\rho_\pi}(y)$ in equation (8), we obtain the desired expression, equation (1):

$$\mathsf{FE}(\gamma,\rho)_\pi(y) \approx E_{(\gamma,\pi)}\left(\mu_{\rho_\pi}(y),y\right) - S_{X\otimes M}\left[\rho_\pi(y)\right] =: \mathsf{LFE}(\gamma,\rho)_\pi(y) .$$

□

Proof of Proposition 3.23. Again we follow the notational conventions of the proof of Proposition 3.14. Additionally, if ω is a state on a tensor product such as $X \otimes Y$, we will write ω_X and ω_Y to denote its X and Y marginals. We will continue to write $c^{\dagger}$ to denote the result of discarding the coparameters of a coparameterized channel c.

Observe that, by repeated application of the linearity of $\mathbb{E}$, the log adjunction, and the definitions of $\bullet$ and $\circ$,

$$\begin{aligned}
&\big(\mathsf{LFE}(d,d') \diamond \mathsf{LFE}(c,c')\big)_\pi(z) \\
&= \mathsf{LFE}(d,d')_{c\bullet\pi}(z) + \big(\mathsf{LFE}(c,c') \circ d'_{c\bullet\pi}\big)_\pi(y) \\
&= \mathsf{LFE}(d,d')_{c\bullet\pi}(z) + \mathop{\mathbb{E}}_{(y,n)\sim d'_{c\bullet\pi}(z)}\big[\mathsf{LFE}(c,c')_\pi(y)\big] \\
&= -\log p_d\big(\mu_{d'_{c\bullet\pi}}(z), z\big) - \log p_{c^{\dagger}\bullet\pi}\big(\mu_{d'_{c\bullet\pi}}(z)_Y\big) \\
&\qquad + \mathop{\mathbb{E}}_{(y,n)\sim d'_{c\bullet\pi}(z)}\Big[\log p_{d'_{c\bullet\pi}}(y,n|z) - \log p_c\big(\mu_{c'_\pi}(y), y\big) - \log p_\pi\big(\mu_{c'_\pi}(y)_X\big) \\
&\qquad\qquad + \mathop{\mathbb{E}}_{(x,m)\sim c'_\pi(y)}\big[\log p_{c'_\pi}(x,m|y)\big]\Big] \\
&= -\log p_d\big(\mu_{d'_{c\bullet\pi}}(z), z\big) - \log p_{c^{\dagger}\bullet\pi}\big(\mu_{d'_{c\bullet\pi}}(z)_Y\big) \\
&\qquad + \mathop{\mathbb{E}}_{(y,n)\sim d'_{c\bullet\pi}(z)}\big[-\log p_c\big(\mu_{c'_\pi}(y), y\big) - \log p_\pi\big(\mu_{c'_\pi}(y)_X\big)\big] \\
&\qquad + \mathop{\mathbb{E}}_{(y,n)\sim d'_{c\bullet\pi}(z)}\Big[\log p_{d'_{c\bullet\pi}}(y,n|z) + \mathop{\mathbb{E}}_{(x,m)\sim c'_\pi(y)}\big[\log p_{c'_\pi}(x,m|y)\big]\Big] \\
&= -\log p_d\big(\mu_{d'_{c\bullet\pi}}(z), z\big) - \log p_{c^{\dagger}\bullet\pi}\big(\mu_{d'_{c\bullet\pi}}(z)_Y\big) \\
&\qquad + \mathop{\mathbb{E}}_{(y,n)\sim d'_{c\bullet\pi}(z)}\big[-\log p_c\big(\mu_{c'_\pi}(y), y\big) - \log p_\pi\big(\mu_{c'_\pi}(y)_X\big)\big] \\
&\qquad + \mathop{\mathbb{E}}_{(y,n)\sim d'_{c\bullet\pi}(z)}\mathop{\mathbb{E}}_{(x,m)\sim c'_\pi(y)}\big[\log p_{d'_{c\bullet\pi}}(y,n|z) + \log p_{c'_\pi}(x,m|y)\big] \\
&= -\log p_d\big(\mu_{d'_{c\bullet\pi}}(z), z\big) - \log p_{c^{\dagger}\bullet\pi}\big(\mu_{d'_{c\bullet\pi}}(z)_Y\big) \\
&\qquad + \mathop{\mathbb{E}}_{(y,n)\sim d'_{c\bullet\pi}(z)}\big[-\log p_c\big(\mu_{c'_\pi}(y), y\big) - \log p_\pi\big(\mu_{c'_\pi}(y)_X\big)\big] \\
&\qquad + \mathop{\mathbb{E}}_{(x,m,y,n)\sim (c'\circ d'_c)_\pi(z)}\big[-\log p_{(c'\circ d'_c)_\pi}(x,m,y,n|z)\big] \\
&= -\log p_d\big(\mu_{d'_{c\bullet\pi}}(z), z\big) - \log p_{c^{\dagger}\bullet\pi}\big(\mu_{d'_{c\bullet\pi}}(z)_Y\big) \\
&\qquad + \mathop{\mathbb{E}}_{(y,n)\sim d'_{c\bullet\pi}(z)}\big[-\log p_c\big(\mu_{c'_\pi}(y), y\big) - \log p_\pi\big(\mu_{c'_\pi}(y)_X\big)\big] - S_{XMYN}\big[(c'\circ d'_c)_\pi(z)\big] \\
&= -\log p_d\big(\mu_{d'_{c\bullet\pi}}(z), z\big) - \log p_{c^{\dagger}\bullet\pi}\big(\mu_{d'_{c\bullet\pi}}(z)_Y\big) \\
&\qquad + \mathop{\mathbb{E}}_{(y,n)\sim d'_{c\bullet\pi}(z)}\big[E_{(c,\pi)}\big(\mu_{c'_\pi}(y), y\big)\big] - S_{XMYN}\big[(c'\circ d'_c)_\pi(z)\big]
\end{aligned}$$

where $XMYN$ is shorthand for $X \otimes M \otimes Y \otimes N$.

Now, writing $E^{\mu}_{(c,\pi)}(y) := E_{(c,\pi)}\big(\mu_{c'_\pi}(y), y\big)$, by the Laplace assumption, we have

$$\mathop{\mathbb{E}}_{(y,n)\sim d'_{c\bullet\pi}(z)}\big[E^{\mu}_{(c,\pi)}(y)\big] \approx E^{\mu}_{(c,\pi)}(\mu_{d'_{c\bullet\pi}}(z)_Y) + \frac{1}{2}\operatorname{tr}\Big[\Big(\partial_y^2 E^{\mu}_{(c,\pi)}\Big)\big(\mu_{d'_{c\bullet\pi}}(z)_Y\big)\,\Sigma_{d'_{c\bullet\pi}}(z)_{YY}\Big]$$

and so we can write

$$
\begin{aligned}
&\big(\mathsf{LFE}(d,d') \diamond \mathsf{LFE}(c,c')\big)_{\pi}(z) \\
&\approx -\log p_d\big(\mu_{d'_{c\bullet\pi}}(z),z\big) - \log p_{c^{\dagger}\bullet\pi}\big(\mu_{d'_{c\bullet\pi}}(z)_Y\big) \\
&\qquad + E^{\mu}_{(c,\pi)}(\mu_{d'_{c\bullet\pi}}(z)_Y) + \frac{1}{2}\operatorname{tr}\Big[\Big(\partial^2_y E^{\mu}_{(c,\pi)}\Big)\,(\mu_{d'_{c\bullet\pi}}(z)_Y)\,\Sigma_{d'_{c\bullet\pi}}(z)_{YY}\Big] \\
&\qquad - S_{XMYN}\big[(c' \circ d'_c)_\pi(z)\big] \\
&= -\log p_d\big(\mu_{d'_{c\bullet\pi}}(z),z\big) - \log p_c\big(\mu_{c'_\pi}(\mu_{d'_{c\bullet\pi}}(z)_Y),\mu_{d'_{c\bullet\pi}}(z)_Y\big) - \log p_\pi\big(\mu_{c'_\pi}(\mu_{d'_{c\bullet\pi}}(z)_Y)_X\big) \\
&\qquad - S_{XMYN}\big[(c' \circ d'_c)_\pi(z)\big] - \log p_{c^{\dagger}\bullet\pi}\big(\mu_{d'_{c\bullet\pi}}(z)_Y\big) \\
&\qquad + \frac{1}{2}\operatorname{tr}\Big[\Big(\partial^2_y E^{\mu}_{(c,\pi)}\Big)\,(\mu_{d'_{c\bullet\pi}}(z)_Y)\,\Sigma_{d'_{c\bullet\pi}}(z)_{YY}\Big] \\
&= E_{(d\bullet c,\pi)}\big(\mu_{(c'\circ d'_c)_\pi}(z),z\big) - S_{XMYN}\big[(c' \circ d'_c)_\pi(z)\big] \\
&\qquad - \log p_{c^{\dagger}\bullet\pi}\big(\mu_{d'_{c\bullet\pi}}(z)_Y\big) + \frac{1}{2}\operatorname{tr}\Big[\Big(\partial^2_y E^{\mu}_{(c,\pi)}\Big)\,(\mu_{d'_{c\bullet\pi}}(z)_Y)\,\Sigma_{d'_{c\bullet\pi}}(z)_{YY}\Big] \\
&= \mathsf{LFE}\big((d,d') \mathbin{\Phi} (c,c')\big)_\pi(z) - \log p_{c^{\dagger}\bullet\pi}\big(\mu_{d'_{c\bullet\pi}}(z)_Y\big) \\
&\qquad + \frac{1}{2}\operatorname{tr}\Big[\Big(\partial^2_y E^{\mu}_{(c,\pi)}\Big)\,(\mu_{d'_{c\bullet\pi}}(z)_Y)\,\Sigma_{d'_{c\bullet\pi}}(z)_{YY}\Big]\ .
\end{aligned}
$$

Therefore, if we define a loss function κ by

$$\kappa_\pi(z) := \frac{1}{2}\operatorname{tr}\Big[\Big(\partial^2_y E^{\mu}_{(c,\pi)}\Big)\,(\mu_{d'_{c\bullet\pi}}(z)_Y)\,\Sigma_{d'_{c\bullet\pi}}(z)_{YY}\Big] - \log p_{c^{\dagger}\bullet\pi}\big(\mu_{d'_{c\bullet\pi}}(z)_Y\big)$$

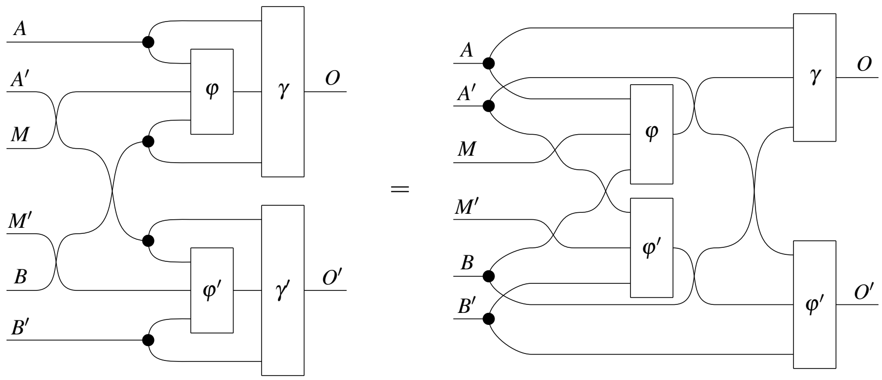
then κ constitutes a 2-cell $\mathsf{LFE}(d,d') \diamond \mathsf{LFE}(c,c') \Rightarrow \mathsf{LFE}\big((d,d') \mathbin{\Phi} (c,c')\big)$, as required. □

C.4 Proofs about monoidal statistical games (§B)

Proof of Proposition B.1. To establish that $(\mathbf{Copara}_2(\mathscr{C}),\otimes,I)$ is a monoidal bicategory, we need to show that $\otimes$ is a pseudofunctor $\mathbf{Copara}_2(\mathscr{C}) \times \mathbf{Copara}_2(\mathscr{C}) \to \mathbf{Copara}_2(\mathscr{C})$ and that I induces a pseudofunctor $\mathbf{1} \to \mathbf{Copara}_2(\mathscr{C})$, such that the pair of pseudofunctors satisfies the relevant coherence data. We will omit the coherence data, and only sketch that the pseudofunctor $\otimes$ is well defined, leaving a full proof for later work. (In the sequel here, we will not make very much use of this tensor.)

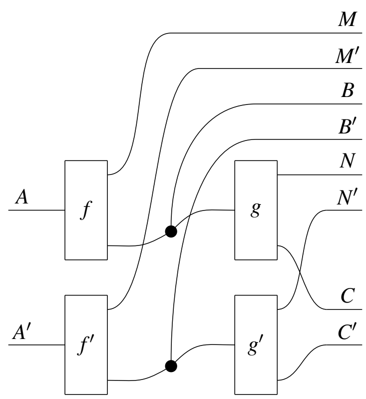
First, we confirm that $\otimes$ is locally functorial, meaning that our definition gives a functor on each pair of hom categories. We begin by noting that $\otimes$ is well-defined on 2-cells, that $\varphi \otimes \varphi'$ satisfies that change of coparameter axiom for $f \otimes f'$; this is immediate from instantiating the axiom's string diagram. Next, we note that $\otimes$ preserves identity 2-cells; again, this is immediate upon substituting identities into the defining diagram. We therefore turn to the preservation of composites, which requires that

$(\gamma \odot \varphi) \otimes (\gamma' \odot \varphi') = (\gamma \otimes \gamma') \odot (\varphi \otimes \varphi')$, and which translates to the following graphical equation:

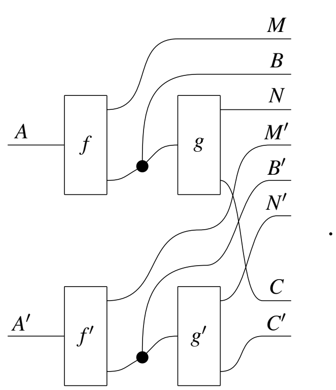


It is easy to see that this equation is satisfied: use the naturality of the symmetry of $(\mathscr{C}, \otimes, I)$. This establishes that $\otimes$ is locally functorial.

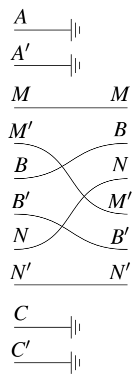
Next, we confirm that $\otimes$ is horizontally (pseudo) functorial. First, we note that $\mathsf{id}_f \otimes \mathsf{id}_{f'} = \mathsf{id}_{f \otimes f'}$ by the naturality of the symmetry of $(\mathscr{C}, \otimes, I)$. Second, we exhibit a multiplication natural isomorphism, witnessing pseudofunctoriality, with components $\mu_{g,g',f,f'} : (g \otimes g') \circ (f \otimes f') \Rightarrow (g \circ f) \otimes (g' \circ f')$ for all composable pairs of 1-cells g, f and g', f'. Let these 1-cells be such that $(g \otimes g') \circ (f \otimes f')$ has the underlying depiction



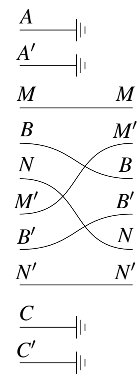
and so $(g \circ f) \otimes (g' \circ f')$ has the depiction



It is then easy to see that defining $\mu_{g,g',f,f'}$ and its inverse $\mu^{-1}_{g,g',f,f'}$ as the 2-cells with the following respective underlying depictions gives us the desired isomorphism:



and



.

The naturality of this definition is a consequence of the naturality of the symmetry of $(\mathscr{C}, \otimes, I)$.

That this tensor satisfies the monoidal bicategory axioms — of associativity, unitality, and coherence — follows from the fact that the monoidal structure $(\otimes, I)$ satisfies correspondingly decategorified versions of these axioms; we leave the details to subsequent exposition. □

Proof of Proposition B.9. We have

$$
\begin{aligned}
&\big(\mathsf{KL}(c) \otimes \mathsf{KL}(d)\big)_\omega(y, y') \\
&= \mathop{\mathbb{E}}_{(x,m) \sim c'_{\omega_X}(y)} \left[\log p_{c'_{\omega_X}}(x, m|y) - \log p_{c^\dagger_{\omega_X}}(x, m|y) \right] \\
&\qquad + \mathop{\mathbb{E}}_{(x',m') \sim d'_{\omega_{X'}}(y')} \left[\log p_{d'_{\omega_{X'}}}(x', m'|y') - \log p_{d^\dagger_{\omega_{X'}}}(x', m'|y') \right] \\
&= \mathop{\mathbb{E}}_{\substack{(x,x',m,m') \sim \\ (c'_{\omega_X} \otimes d'_{\omega_{X'}})(y,y')}} \left[\log p_{c'_{\omega_X} \otimes d'_{\omega_{X'}}}(x, x', m, m'|y, y') - \log p_{c^\dagger_{\omega_X} \otimes d^\dagger_{\omega_{X'}}}(x, x', m, m'|y, y') \right]
\end{aligned}
$$

and

$$
\begin{aligned}
&\big(\mathsf{KL}(c \otimes d)_\omega(y, y') \\
&= \mathop{\mathbb{E}}_{\substack{(x,x',m,m') \sim \\ (c'_{\omega_X} \otimes d'_{\omega_{X'}})(y,y')}} \left[\log p_{c'_{\omega_X} \otimes d'_{\omega_{X'}}}(x, x', m, m'|y, y') - \log p_{(c \otimes d)^\dagger_\omega}(x, x', m, m'|y, y') \right] .
\end{aligned}
$$

Using Bayes' rule, we can rewrite the exact inversions in these expressions, obtaining

$$
\begin{aligned}
&\big(\mathsf{KL}(c) \otimes \mathsf{KL}(d)\big)_\omega(y, y') \\
&= \mathop{\mathbb{E}}_{\substack{(x,x',m,m') \sim \\ (c'_{\omega_X} \otimes d'_{\omega_{X'}})(y,y')}} \Big[\log p_{c'_{\omega_X} \otimes d'_{\omega_{X'}}}(x, x', m, m'|y, y') - \log p_c(y, m|x) - \log p_d(y', m'|x') \\
&\qquad - \log p_{\omega_X}(x) - \log p_{\omega_{X'}}(x') + \log p_{c^\dagger \bullet \omega_X}(y) + \log p_{d^\dagger \bullet \omega_{X'}}(y') \Big]
\end{aligned}
$$

and

$$\begin{aligned}
&\big(\mathsf{KL}(c \otimes d)_\omega(y,y') \\
&= \mathop{\mathbb{E}}_{\substack{(x,x',m,m')\sim \\ (c'_{\omega_X} \otimes d'_{\omega_{X'}})(y,y')}} \Big[\log p_{c'_{\omega_X} \otimes d'_{\omega_{X'}}}(x,x',m,m'|y,y') - \log p_c(y,m|x) - \log p_d(y',m'|x') \\
&\qquad - \log p_\omega(x,x') + \log p_{(c\otimes d)^\dagger \bullet \omega}(y,y') \Big] .
\end{aligned}$$

We define $\lambda^{\mathsf{KL}}(c,d)_\omega(y,y')$ as the difference from $\big(\mathsf{KL}(c \otimes d)_\omega(y,y')$ to $\big(\mathsf{KL}(c) \otimes \mathsf{KL}(d)\big)_\omega(y,y')$, and so, with a little rearranging, we obtain the expression above:

$$\begin{aligned}
\lambda^{\mathsf{KL}}(c,d)_\omega(y,y') &:= \big(\mathsf{KL}(c \otimes d)_\omega(y,y') - \big(\mathsf{KL}(c) \otimes \mathsf{KL}(d)\big)_\omega(y,y') \\
&= \mathop{\mathbb{E}}_{\substack{(x,x',m,m')\sim \\ (c'_{\omega_X} \otimes d'_{\omega_{X'}})(y,y')}} \left[\log \frac{p_{\omega_X \otimes \omega_{X'}}(x,x')}{p_\omega(x,x')} \right] + \log \frac{p_{(c\otimes d)^\dagger \bullet \omega}(y,y')}{p_{(c\otimes d)^\dagger \bullet (\omega_X \otimes \omega_{X'})}(y,y')} .
\end{aligned}$$

Next, we need to validate lax naturality. Since KL is strict on losses, we need only check that

$$\lambda^{\mathsf{KL}}(e \,\Phi\, c, f \,\Phi\, d) = \lambda^{\mathsf{KL}}(e,f)_{c\otimes d} + \lambda^{\mathsf{KL}}(c,d) \circ (e' \otimes f')_{c\otimes d} .$$

By definition, we have

$$\begin{aligned}
&\big(\lambda^{\mathsf{KL}}(e,f)_{c\otimes d}\big)_\omega(z,z') \\
&= \mathop{\mathbb{E}}_{\substack{(y,y',n,n')\sim \\ (e'_c \otimes f'_d)_\omega(z,z')}} \left[\log \frac{p_{(c\otimes d)^\dagger \bullet (\omega_X \otimes \omega_{X'})}(y,y')}{p_{(c\otimes d)^\dagger \bullet \omega}(y,y')} \right] + \log \frac{p_{(e\otimes f)^\dagger \bullet (c\otimes d)^\dagger \bullet \omega}(z,z')}{p_{(e\otimes f)^\dagger \bullet (c\otimes d)^\dagger \bullet (\omega_X \otimes \omega_{X'})}(z,z')}
\end{aligned}$$

and

$$\begin{aligned}
&\big(\lambda^{\mathsf{KL}}(c,d) \circ (e' \otimes f')_{c\otimes d}\big)_\omega(z,z') \\
&= \mathop{\mathbb{E}}_{\substack{(y,y',n,n')\sim \\ (e'_c \otimes f'_d)_\omega(z,z')}} \left[\mathop{\mathbb{E}}_{\substack{(x,x',m,m')\sim \\ (c'_{\omega_X} \otimes d'_{\omega_{X'}})(y,y')}} \left[\log \frac{p_{\omega_X \otimes \omega_{X'}}(x,x')}{p_\omega(x,x')} \right] + \log \frac{p_{(c\otimes d)^\dagger \bullet \omega}(y,y')}{p_{(c\otimes d)^\dagger \bullet (\omega_X \otimes \omega_{X'})}(y,y')} \right] .
\end{aligned}$$

And so we also have

$$\begin{aligned}
&\lambda^{\mathsf{KL}}(e \,\Phi\, c, f \,\Phi\, d)_\omega(z,z') \\
&= \mathop{\mathbb{E}}_{\substack{(x,x',m,m')\sim \\ \big((c' \circ e'_c) \otimes (d' \circ f'_d)\big)_\omega(z,z')}} \left[\log \frac{p_{\omega_X \otimes \omega_{X'}}(x,x')}{p_\omega(x,x')} \right] + \log \frac{p_{(e\otimes f)^\dagger \bullet (c\otimes d)^\dagger \bullet \omega}(z,z')}{p_{(e\otimes f)^\dagger \bullet (c\otimes d)^\dagger \bullet (\omega_X \otimes \omega_{X'})}(z,z')} \\
&= \big(\lambda^{\mathsf{KL}}(c,d) \circ (e' \otimes f')_{c\otimes d}\big)_\omega(z,z') + \big(\lambda^{\mathsf{KL}}(e,f)_{c\otimes d}\big)_\omega(z,z')
\end{aligned}$$

thereby establishing the lax naturality of λ^{KL}, by the commutativity of $+$. □

Proof of Proposition B.10. To obtain the definition of $\lambda^{\mathsf{MLE}}(c,d)$, we consider the difference from $\mathsf{MLE}(c\otimes d)$ to $\mathsf{MLE}(c)\otimes\mathsf{MLE}(d)$:

$$\begin{aligned}\lambda^{\mathsf{MLE}}(c,d)_\omega(y,y') &:= \mathsf{MLE}(c\otimes d)_\omega(y,y') - \big(\mathsf{MLE}(c)\otimes\mathsf{MLE}(d)\big)_\omega(y,y')\\ &= -\log p_{(c\otimes d)^\dagger\bullet\omega}(y,y') + \log p_{c^\dagger\bullet\omega_X}(y) - \log p_{d^\dagger\bullet\omega_{X'}}(y')\\ &= \log\frac{p_{(c\otimes d)^\dagger\bullet(\omega_X\otimes\omega_{X'})}(y,y')}{p_{(c\otimes d)^\dagger\bullet\omega}(y,y')}\,.\end{aligned}$$

To demonstrate lax naturality, recall that MLE is a lax section, so we need to consider the corresponding $\diamond$-laxator. From Proposition 3.16, the laxator $K^{\mathsf{MLE}}(e,c):\mathsf{MLE}(e)\diamond\mathsf{MLE}(c)\Rightarrow\mathsf{MLE}(e\mathbin{\varobar} c)$ is given by $K^{\mathsf{MLE}}(e,c):=\mathsf{MLE}(c)\circ e'_c$. Next, observe that

$$\begin{aligned}\lambda^{\mathsf{MLE}}(e\mathbin{\varobar} c, f\mathbin{\varobar} d)_\omega(z,z') &= \log\frac{p_{\big((e\bullet c)^\dagger\otimes(f\bullet d)^\dagger\big)\bullet(\omega_X\otimes\omega_{X'})}(z,z')}{p_{\big((e\bullet c)^\dagger\otimes(f\bullet d)^\dagger\big)\bullet\omega}(z,z')}\\ &= \log\frac{p_{(e\otimes f)^\dagger\bullet(c\otimes d)^\dagger\bullet(\omega_X\otimes\omega_{X'})}(z,z')}{p_{(e\otimes f)^\dagger\bullet(c\otimes d)^\dagger\bullet\omega}(z,z')}\\ &= \lambda^{\mathsf{MLE}}(e,f)_{(c\otimes d)\bullet\omega}(z,z')\,.\end{aligned}$$

Consequently, we need to verify the equation

$$\mathsf{MLE}(c\otimes d)\circ(e\otimes f')_{c\otimes d} = \lambda^{\mathsf{MLE}}(c,d)\circ(e'\otimes f')_{c\otimes d} + \big(\mathsf{MLE}(c)\otimes\mathsf{MLE}(d)\big)\circ(e'\otimes f')_{c\otimes d}$$

which, by bilinearity of effects, is equivalent to verifying

$$\mathsf{MLE}(c\otimes d) = \lambda^{\mathsf{MLE}}(c,d) + \mathsf{MLE}(c)\otimes\mathsf{MLE}(d)\,.$$

But, since $+$ is commutative, this is satisfied by the definition of $\lambda^{\mathsf{MLE}}(c,d)$ as a 2-cell of type $\mathsf{MLE}(c\otimes d)\Rightarrow\mathsf{MLE}(c)\otimes\mathsf{MLE}(d)$. □

Proof sketch for Corollary B.11. FE is defined as $\mathsf{KL}+\mathsf{MLE}$, and hence λ^{FE} is obtained as $\lambda^{\mathsf{KL}}+\lambda^{\mathsf{MLE}}$. Since $+$ is functorial, it preserves lax naturality, and so λ^{FE} is also lax natural. λ^{FE} is thus a strong transformation $\mathsf{FE}(-)\otimes\mathsf{FE}(=)\Rightarrow\mathsf{FE}(-\otimes=)$, and hence FE is lax monoidal by Remark B.7. □

Proof of Proposition B.12. We have

$$\begin{aligned}&\mathsf{LFE}(c\otimes d)_\omega(y,y')\\ &= -\log p_{c\otimes d}(\mu_{(c'\otimes d')_\omega}(y,y'),y,y') - \log p_\omega(\mu_{(c'\otimes d')_\omega}(y,y')_{XX'})\\ &\qquad - S_{XX'MM'}\left[(c'\otimes d')_\omega(y,y')\right]\\ &= -\log p_c(\mu_{c'_{\omega_X}}(y),y) - \log p_d(\mu_{d'_{\omega_{X'}}}(y'),y') - p_\omega(\mu_{(c'\otimes d')_\omega}(y,y')_{XX'})\\ &\qquad - S_{XM}\left[c'_{\omega_X}(y)\right] - S_{X'M'}\left[d'_{\omega_{X'}}(y')\right]\end{aligned}$$

and

$$\begin{aligned}&\big(\mathsf{LFE}(c)\otimes\mathsf{LFE}(d)\big)_\omega(y,y')\\ &= \mathsf{LFE}(c)_{\omega_X}(y) + \mathsf{LFE}(d)_{\omega_{X'}}(y')\\ &= -\log p_c(\mu_{c'_{\omega_X}}(y),y) - p_{\omega_X}(\mu_{c_{\omega_X}}(y)_X) - S_{XM}\left[c'_{\omega_X}(y)\right]\\ &\qquad -\log p_d(\mu_{d'_{\omega_{X'}}}(y'),y') - p_{\omega_{X'}}(\mu_{d'_{\omega_{X'}}}(y')_{X'}) - S_{X'M'}\left[d'_{\omega_{X'}}(y')\right]\end{aligned}$$

so that

$$\begin{aligned}\lambda^{\mathsf{LFE}}(c,d)_\omega(y,y') &= \mathsf{LFE}(c\otimes d)_\omega(y,y') - \big(\mathsf{LFE}(c)\otimes\mathsf{LFE}(d)\big)_\omega(y,y') \\ &= \log\frac{p_{\omega_X\otimes\omega_{X'}}(\mu_{(c\otimes d)'_\omega}(y,y')_{XX'})}{p_\omega(\mu_{(c\otimes d)'_\omega}(y,y')_{XX'})}\end{aligned}$$

as given above.

We need to verify lax naturality, which means checking the equation

$$\lambda^{\mathsf{LFE}}(e\diamond c, f\diamond d) + \kappa(e\otimes f, c\otimes d) = \lambda^{\mathsf{LFE}}(e,f)_{c\otimes d} + \lambda^{\mathsf{LFE}}(c,d)\circ(e'\otimes f')_{c\otimes d} + \kappa(e,c)\otimes\kappa(f,d)$$

where κ is the $\diamond$-laxator with components $\kappa(e,c) : \mathsf{LFE}(e)\diamond\mathsf{LFE}(c) \Rightarrow \mathsf{LFE}(e\diamond c)$ given by

$$\kappa(e,c)_\pi(z) = \frac{1}{2}\operatorname{tr}\Big[\Big(\partial_y^2 E^\mu_{(c,\pi)}\Big)\big(\mu_{e'_{c\bullet\pi}}(z)_Y\big)\,\Sigma_{e'_{c\bullet\pi}}(z)_{YY}\Big] - \log p_{c^\dagger\bullet\pi}\big(\mu_{e'_{c\bullet\pi}}(z)_Y\big)\ .$$

(see Proposition 3.23). We have

$$\begin{aligned}\lambda^{\mathsf{LFE}}(e\diamond c, f\diamond d) &= \log\frac{p_{\omega_X\otimes\omega_{X'}}(\mu_{(c\otimes d)'_\omega}(\mu_{(e\otimes f)'_{(c\otimes d)\bullet\omega}}(z,z')_{YY'})_{XX'})}{p_\omega(\mu_{(c\otimes d)'_\omega}(\mu_{(e\otimes f)'_{(c\otimes d)\bullet\omega}}(z,z')_{YY'})_{XX'})} \\ &= \lambda^{\mathsf{LFE}}(c,d)_\omega(\mu_{(e\otimes f)'_{(c\otimes d)\bullet\omega}}(z,z')_{YY'})_{XX'})\end{aligned}$$

and, by the Laplace approximation,

$$\begin{aligned}&\big(\lambda^{\mathsf{LFE}}(c,d)\circ(e'\otimes f')_{c\otimes d})_\omega(z,z') \\ &= \mathop{\mathbb{E}}_{\substack{(y,y',n,n')\sim\\ (e'_c\otimes f'_d)_\omega(z,z')}}\Big[\lambda^{\mathsf{LFE}}(c,d)_\omega(y,y')\Big] \\ &\approx \lambda^{\mathsf{LFE}}(c,d)_\omega(\mu_{(e\otimes f)'_{(c\otimes d)\bullet\omega}}(z,z')_{YY'}) \\ &\qquad + \frac{1}{2}\operatorname{tr}\Big[\Big(\partial^2_{(y,y')}\lambda^{\mathsf{LFE}}(c,d)_\omega\Big)\Big(\mu_{(e\otimes f)'_{(c\otimes d)\bullet\omega}}(z,z')_{YY'}\Big)\,\Sigma_{(e\otimes f)'_{(c\otimes d)\bullet\omega}}(z,z')_{(YY')(YY')}\Big]\ .\end{aligned}$$

We also have

$$\begin{aligned}&\big(\kappa(e,c)\otimes\kappa(f,d)\big)_\omega(z,z') \\ &= \kappa(e,c)_{\omega_X}(z) + \kappa(f,d)_{\omega_{X'}}(z') \\ &= \frac{1}{2}\operatorname{tr}\Big[\Big(\partial_y^2 E^\mu_{(c,\omega_X)}\Big)\big(\mu_{e'_{c\bullet\omega_X}}(z)_Y\big)\,\Sigma_{e'_{c\bullet\omega_X}}(z)_{YY}\Big] - \log p_{c^\dagger\bullet\omega_X}\big(\mu_{e'_c\bullet\omega_X}(z)_Y\big) \\ &\qquad + \frac{1}{2}\operatorname{tr}\Big[\Big(\partial_{y'}^2 E^\mu_{(d,\omega_{X'})}\Big)\big(\mu_{f'_{d\bullet\omega_{X'}}}(z')_{Y'}\big)\,\Sigma_{f'_{d\bullet\omega_{X'}}}(z')_{Y'Y'}\Big] - \log p_{d^\dagger\bullet\omega_{X'}}\big(\mu_{f'_{d\bullet\omega_{X'}}}(z')_{Y'}\big) \\ &= \frac{1}{2}\operatorname{tr}\Big[\Big(\partial^2_{(y,y')}E^\mu_{(c\otimes d,\omega_X\otimes\omega_{X'})}\Big)\big(\mu_{(e\otimes f)'_{(c\otimes d)\bullet(\omega_X\otimes\omega_{X'})}}(z,z')_{YY'}\big) \\ &\qquad\qquad \Sigma_{(e\otimes f)'_{(c\otimes d)\bullet(\omega_X\otimes\omega_{X'})}}(z,z')_{(YY')(YY')}\Big] \\ &\qquad - \log p_{(c\otimes d)^\dagger\bullet(\omega_X\otimes\omega_{X'})}\big(\mu_{(e\otimes f)'_{(c\otimes d)\bullet(\omega_X\otimes\omega_{X'})}}(z,z')_{YY'}\big)\ .\end{aligned}$$

The left-hand side of the lax naturality equation is therefore given by

$$\begin{aligned}
&\left(\lambda^{\mathsf{LFE}}(e \circ c, f \circ d) + \kappa(e \otimes f, c \otimes d)\right)_{\omega}(z,z') \\
&= \lambda^{\mathsf{LFE}}(c,d)_{\omega}(\mu_{(e\otimes f)'_{(c\otimes d)\bullet\omega}}(z,z')_{YY'}) \\
&\quad + \frac{1}{2}\operatorname{tr}\left[\left(\partial^2_{(y,y')} E^{\mu}_{(c\otimes d,\omega)}\right)\left(\mu_{(e\otimes f)'_{(c\otimes d)\bullet\omega}}(z,z')_{YY'}\right)\Sigma_{(e\otimes f)'_{(c\otimes d)\bullet\omega}}(z,z')_{(YY')(YY')}\right] \\
&\quad - \log p_{(c\otimes d)\bullet\omega}\left(\mu_{(e\otimes f)'_{(c\otimes d)\bullet\omega}}(z,z')_{YY'}\right)
\end{aligned}$$

while the right-hand side is given by

$$\begin{aligned}
&\left(\lambda^{\mathsf{LFE}}(e,f)_{c\otimes d} + \lambda^{\mathsf{LFE}}(c,d)\circ(e'\otimes f')_{c\otimes d} + \kappa(e,c)\otimes\kappa(f,d)\right)_{\omega}(z,z') \\
&= \log\frac{p_{(c\otimes d)^{\dagger}\bullet(\omega_X\otimes\omega_{X'})}(\mu_{(e\otimes f)'_{(c\otimes d)\bullet\omega}}(z,z')_{YY'})}{p_{(c\otimes d)^{\dagger}\bullet\omega}(\mu_{(e\otimes f)'_{(c\otimes d)\bullet\omega}}(z,z')_{YY'})} \\
&\quad + \lambda^{\mathsf{LFE}}(c,d)_{\omega}(\mu_{(e\otimes f)'_{(c\otimes d)\bullet\omega}}(z,z')_{YY'}) \\
&\quad + \frac{1}{2}\operatorname{tr}\left[\left(\partial^2_{(y,y')}\lambda^{\mathsf{LFE}}(c,d)_{\omega}\right)\left(\mu_{(e\otimes f)'_{(c\otimes d)\bullet\omega}}(z,z')_{YY'}\right)\Sigma_{(e\otimes f)'_{(c\otimes d)\bullet\omega}}(z,z')_{(YY')(YY')}\right] \\
&\quad + \frac{1}{2}\operatorname{tr}\Big[\left(\partial^2_{(y,y')} E^{\mu}_{(c\otimes d,\omega_X\otimes\omega_{X'})}\right)\left(\mu_{(e\otimes f)'_{(c\otimes d)\bullet(\omega_X\otimes\omega_{X'})}}(z,z')_{YY'}\right) \\
&\qquad\qquad \Sigma_{(e\otimes f)'_{(c\otimes d)\bullet(\omega_X\otimes\omega_{X'})}}(z,z')_{(YY')(YY')}\Big] \\
&\quad - \log p_{(c\otimes d)^{\dagger}\bullet(\omega_X\otimes\omega_{X'})}\left(\mu_{(e\otimes f)'_{(c\otimes d)\bullet(\omega_X\otimes\omega_{X'})}}(z,z')_{YY'}\right) \\
&= -\log p_{(c\otimes d)^{\dagger}\bullet\omega}(\mu_{(e\otimes f)'_{(c\otimes d)\bullet\omega}}(z,z')_{YY'}) + \lambda^{\mathsf{LFE}}(c,d)_{\omega}(\mu_{(e\otimes f)'_{(c\otimes d)\bullet\omega}}(z,z')_{YY'}) \\
&\quad + \frac{1}{2}\operatorname{tr}\left[\left(\partial^2_{(y,y')}\lambda^{\mathsf{LFE}}(c,d)_{\omega}\right)\left(\mu_{(e\otimes f)'_{(c\otimes d)\bullet\omega}}(z,z')_{YY'}\right)\Sigma_{(e\otimes f)'_{(c\otimes d)\bullet\omega}}(z,z')_{(YY')(YY')}\right] \\
&\quad + \frac{1}{2}\operatorname{tr}\Big[\left(\partial^2_{(y,y')} E^{\mu}_{(c\otimes d,\omega_X\otimes\omega_{X'})}\right)\left(\mu_{(e\otimes f)'_{(c\otimes d)\bullet(\omega_X\otimes\omega_{X'})}}(z,z')_{YY'}\right) \\
&\qquad\qquad \Sigma_{(e\otimes f)'_{(c\otimes d)\bullet(\omega_X\otimes\omega_{X'})}}(z,z')_{(YY')(YY')}\Big].
\end{aligned}$$

The difference from the left- to the right-hand side is thus

$$\begin{aligned}
&\frac{1}{2}\operatorname{tr}\left[\left(\partial^2_{(y,y')} E^{\mu}_{(c\otimes d,\omega)}\right)\left(\mu_{(e\otimes f)'_{(c\otimes d)\bullet\omega}}(z,z')_{YY'}\right)\Sigma_{(e\otimes f)'_{(c\otimes d)\bullet\omega}}(z,z')_{(YY')(YY')}\right] \\
&\quad - \frac{1}{2}\operatorname{tr}\Big[\left(\partial^2_{(y,y')} E^{\mu}_{(c\otimes d,\omega_X\otimes\omega_{X'})}\right)\left(\mu_{(e\otimes f)'_{(c\otimes d)\bullet(\omega_X\otimes\omega_{X'})}}(z,z')_{YY'}\right) \\
&\qquad\qquad \Sigma_{(e\otimes f)'_{(c\otimes d)\bullet(\omega_X\otimes\omega_{X'})}}(z,z')_{(YY')(YY')}\Big] \\
&\quad - \frac{1}{2}\operatorname{tr}\left[\left(\partial^2_{(y,y')}\lambda^{\mathsf{LFE}}(c,d)_{\omega}\right)\left(\mu_{(e\otimes f)'_{(c\otimes d)\bullet\omega}}(z,z')_{YY'}\right)\Sigma_{(e\otimes f)'_{(c\otimes d)\bullet\omega}}(z,z')_{(YY')(YY')}\right].
\end{aligned}$$

Now, by definition $\Sigma_{(e\otimes f)'_{(c\otimes d)\bullet\omega}} = \Sigma_{(e\otimes f)'_{(c\otimes d)\bullet(\omega_X\otimes\omega_{X'})}}$, and so by the linearity of the trace and of derivation, this difference simplifies to

$$\begin{aligned}
&\frac{1}{2}\operatorname{tr}\Big[\left(\partial^2_{(y,y')}\left(E^{\mu}_{(c\otimes d,\omega)} - E^{\mu}_{(c\otimes d,\omega_X\otimes\omega_{X'})} - \lambda^{\mathsf{LFE}}(c,d)_{\omega}\right)\right) \\
&\qquad\qquad \left(\mu_{(e\otimes f)'_{(c\otimes d)\bullet\omega}}(z,z')_{YY'}\right)\Sigma_{(e\otimes f)'_{(c\otimes d)\bullet\omega}}(z,z')_{(YY')(YY')}\Big].
\end{aligned}$$

Recall from the proof of Proposition 3.23 that $E^{\mu}_{(c,\pi)}(y) := E_{(c,\pi)}\big(\mu_{c'_\pi}(y), y\big)$, and hence

$$\begin{aligned}
&\big(E^{\mu}_{(c\otimes d,\omega)} - E^{\mu}_{(c\otimes d,\omega_X\otimes\omega_{X'})}\big)(y,y') \\
&= \big(E_{(c\otimes d,\omega)} - E_{(c\otimes d,\omega_X\otimes\omega_{X'})}\big)\big(\mu_{(c\otimes d)'_\omega}(y,y'), y, y'\big) \\
&= -\log p_\omega(\mu_{(c\otimes d)'_\omega}(y,y')_{XX'}) + \log p_{\omega_X\otimes\omega_{X'}}(\mu_{(c\otimes d)'_\omega}(y,y')_{XX'}) \\
&= \log \frac{p_{\omega_X\otimes\omega_{X'}}(\mu_{(c\otimes d)'_\omega}(y,y')_{XX'})}{p_\omega(\mu_{(c\otimes d)'_\omega}(y,y')_{XX'})} \\
&= \lambda^{\mathsf{LFE}}(c,d)_\omega(y,y')
\end{aligned}$$

so that $E^{\mu}_{(c\otimes d,\omega)} - E^{\mu}_{(c\otimes d,\omega_X\otimes\omega_{X'})} - \lambda^{\mathsf{LFE}}(c,d)_\omega = 0$. This establishes that λ^{LFE} is lax natural. □