The Magmoid of Normalized Stochastic Kernels

(j.w.w. Elena Di Lavore and Márk Széles)

Mario Román

TallCat Seminar, 9th April 2026

PART 0: PROBABILITY IS UNINTUITIVE

> "It should be said: for someone trained in formal methods, the area of probability theory can be rather sloppy [...] this 'sloppiness' is ultimately a hindrance to further development of the field in computer science."
>
> — JACOBS, 2025.

SINGLE TEST PROBLEM. Consider an illness with 33% prevalence ($\frac{1}{3}|I\rangle + \frac{2}{3}|H\rangle$) for which we have a test with 75% sensitivity ($\frac{3}{4}|P\rangle + \frac{1}{4}|N\rangle$ when ill) and 50% specificity ($\frac{1}{2}|P\rangle + \frac{1}{2}|N\rangle$ when healthy). What is the probability of illness after a positive test?

PREVALENCE : 33%	$\frac{1}{3}\lvert I\rangle + \frac{2}{3}\lvert H\rangle$	
SENSITIVITY : 75%	$\frac{3}{4}\lvert P\rangle + \frac{1}{4}\lvert N\rangle$	when ill
SPECIFICITY : 50%	$\frac{1}{2}\lvert P\rangle + \frac{1}{2}\lvert N\rangle$	when healthy

SOLUTION. (~43% ill, ~57% healthy)

$\frac{1}{3}\lvert I\rangle + \frac{2}{3}\lvert H\rangle$	PREVALENCE
$\frac{1}{3}\lvert I\rangle\left(\frac{3}{4}\lvert P\rangle + \frac{1}{4}\lvert N\rangle\right) + \frac{2}{3}\lvert H\rangle\left(\frac{1}{2}\lvert P\rangle + \frac{1}{2}\lvert N\rangle\right)$	TESTING
$\frac{1}{4}\lvert IP\rangle + \frac{1}{12}\lvert IN\rangle + \frac{1}{3}\lvert HP\rangle + \frac{1}{3}\lvert HN\rangle$	MULTIPLY
$\frac{1}{4}\lvert IP\rangle + \frac{1}{3}\lvert HP\rangle$	OBSERVE (P)
$\frac{3}{7}\lvert I\rangle + \frac{4}{7}\lvert H\rangle$.	NORMALIZE

The phenomenon we seek to study occurs at the last two steps. We obtain something that is not a full distribution but a subdistribution; then, we multiply by a constant ($^{12}/_{7}$) to obtain again a distribution — this is normalization.

UNCLEAR TEST PROBLEM. Imagine the test result was uncertain — e.g. it was dark!

PREVALENCE : 33%	$\frac{1}{3}\|I\rangle + \frac{2}{3}\|H\rangle$	
SENSITIVITY : 75%	$\frac{3}{4}\|P\rangle + \frac{1}{4}\|N\rangle$	when Ill
SPECIFICITY : 50%	$\frac{1}{2}\|P\rangle + \frac{1}{2}\|N\rangle$	when Healthy
UNCERTAINTY : 75%	$\frac{3}{4}\|P\rangle + \frac{1}{4}\|N\rangle$	

Solution 1. (~38% ill, ~62% healthy)

~38.5%

$\frac{1}{3}\|I\rangle + \frac{2}{3}\|H\rangle$	PREVALENCE
$\frac{1}{3}\|I\rangle(\frac{3}{4}\|P\rangle + \frac{1}{4}\|N\rangle) + \frac{2}{3}\|H\rangle(\frac{1}{2}\|P\rangle + \frac{1}{2}\|N\rangle)$	CHANNEL
$\frac{1}{3}\|I\rangle(\frac{3}{4}\|P\rangle + \frac{1}{4}\|N\rangle)(\frac{3}{4}\|P\rangle + \frac{1}{4}\|N\rangle) + \frac{2}{3}\|H\rangle(\frac{1}{2}\|P\rangle + \frac{1}{2}\|N\rangle)(\frac{3}{4}\|P\rangle + \frac{1}{4}\|N\rangle)$	UNCERTAINTY
$\frac{3}{16}\|IP\rangle + \frac{1}{48}\|IN\rangle + \frac{1}{4}\|HP\rangle + \frac{1}{12}\|HN\rangle$	OBSERVE
$\frac{5}{24}\|I\rangle + \frac{1}{3}\|H\rangle$	PROJECT
$\frac{5}{13}\|I\rangle + \frac{8}{13}\|H\rangle$	NORMALIZE

Solution 2. (~37% ill, ~63% healthy)

~37.1%

$\frac{3}{4}\|P\rangle + \frac{1}{4}\|N\rangle$	UNCERTAINTY
$(\frac{3}{4}\|P\rangle + \frac{1}{4}\|N\rangle)(\frac{1}{3}\|I\rangle + \frac{2}{3}\|H\rangle)$	PREVALENCE
$(\frac{3}{4}\|P\rangle + \frac{1}{4}\|N\rangle)(\frac{1}{3}\|I\rangle(\frac{3}{4}\|P\rangle + \frac{1}{4}\|N\rangle) + \frac{2}{3}\|H\rangle(\frac{1}{2}\|P\rangle + \frac{1}{2}\|N\rangle))$	CHANNEL
$\frac{3}{4}\|P\rangle(\frac{1}{4}\|IP\rangle + \frac{1}{3}\|HP\rangle) + \frac{1}{4}\|N\rangle(\frac{1}{12}\|IN\rangle + \frac{1}{3}\|H\rangle)))$	OBSERVE
$\frac{3}{4}\|P\rangle(\frac{3}{7}\|I\rangle + \frac{4}{7}\|H\rangle) + \frac{1}{4}\|N\rangle(\frac{1}{5}\|I\rangle + \frac{4}{5}\|H\rangle)$	NORMALIZE
$\frac{9}{28}\|PI\rangle + \frac{3}{7}\|PH\rangle + \frac{1}{20}\|NI\rangle + \frac{1}{5}\|NH\rangle$	PROJECT/MULTIPLY
$\frac{13}{35}\|I\rangle + \frac{22}{35}\|H\rangle$	PROJECT

We get two different numbers: this is terrible. (c.f. Jacobs 2018).

SINGLE TEST SEMANTICS. Let us go back to the original problem.

PREVALENCE:	33%	$\frac{1}{3}\lvert I\rangle + \frac{2}{3}\lvert H\rangle$	$\frac{1}{3}\lvert I\rangle + \frac{2}{3}\lvert H\rangle$	(P)
SENSITIVITY:	75%	$\frac{3}{4}\lvert P\rangle + \frac{1}{4}\lvert N\rangle$ when ill	$\frac{1}{4}\lvert IP\rangle + \frac{1}{12}\lvert IN\rangle + \frac{1}{3}\lvert HP\rangle + \frac{1}{3}\lvert HN\rangle$	(C)
SPECIFICITY:	50%	$\frac{1}{2}\lvert P\rangle + \frac{1}{2}\lvert N\rangle$ when healthy	$\frac{1}{5}\lvert IN\rangle + \frac{4}{5}\lvert HN\rangle$	(O)

Let us propose monadic semantics. The distribution monad (D) can compose probabilities, but cannot apply observations.

$$DX = \left\{ \sum_{i=0}^{n} \lambda_i \lvert x_i\rangle \;\middle|\; \sum_{i=0}^{n} \lambda_i = 1,\; x_i \in X,\; \lambda_i \in \mathbb{R}^+ \right\}.$$

Observations yield subdistributions: distributions adding to less then 1. The maybe monad (M) enables them.

$$DMX \cong \left\{ \sum_{i=0}^{n} \lambda_i \lvert x_i\rangle \;\middle|\; \sum_{i=0}^{n} \lambda_i \leq 1,\; x_i \in X,\; \lambda_i \in \mathbb{R}^+ \right\}.$$

Subdistributions provide a monadic semantics for probabilistic inference; normalization remains ad-hoc.

$$\begin{aligned} P &: 1 \rightarrow\!\!\!\!\!\!\!+\;\; \{I,H\} \\ C &: \{I,H\} \rightarrow\!\!\!\!\!\!\!+\;\; \{P,N\} \\ O &: \{P,N\} \rightarrow\!\!\!\!\!\!\!+\;\; 1 \end{aligned}
\qquad
\begin{aligned} P &= \tfrac{1}{3}\lvert I\rangle + \tfrac{2}{3}\lvert H\rangle \\ C(I) &= \tfrac{3}{4}\lvert P\rangle + \tfrac{1}{4}\lvert N\rangle;\; C(H) = \tfrac{1}{2}\lvert P\rangle + \tfrac{1}{2}\lvert N\rangle. \\ O(P) &= 0;\quad O(N) = 1. \end{aligned}$$

$$\begin{aligned} P &: 1 \longrightarrow DM\{I,H\} \\ C &: \{I,H\} \longrightarrow DM\{P,N\} \\ O &: \{P,N\} \longrightarrow DM1. \end{aligned}$$

DISTRIBUTIVE LAWS. A distributive law between two monads on the same category, (S, η^S, μ^S) and (T, η^T, μ^T), is a natural transformation $\psi_X : TSX \longrightarrow STX$, making the following diagrams commute.

$$\begin{tikzcd}
TTSX \arrow[r, "\mu^T S"] \arrow[d, "T\psi"'] & TSX \arrow[r, "\psi"] & STX \\
TSTX \arrow[r, "\psi T"] & STTX \arrow[ur, "S\mu^T"] &
\end{tikzcd}
\qquad
\begin{tikzcd}
 & STX \\
TSX \arrow[ur, "\psi"] & SX \arrow[l, "\eta^T S"'] \arrow[u, "S\eta^T"']
\end{tikzcd}$$

$$\begin{tikzcd}
TSSX \arrow[r, "T\mu^S"] \arrow[d, "\psi S"'] & TSX \arrow[r, "\psi"] & STX \\
STSX \arrow[r, "S\psi"] & SSTX \arrow[ur, "\mu^S T"] &
\end{tikzcd}
\qquad
\begin{tikzcd}
 & STX \\
TSX \arrow[ur, "\psi"] & TX \arrow[l, "T\eta^S"] \arrow[u, "\eta^S T"']
\end{tikzcd}$$

EXAMPLE. The distribution monad (D) and the maybe monad (M) admit a distributive law: inclusion of normalized kernels into subdistributions,

$$(-)^{\bullet}_X : MDX \longrightarrow DMX.$$

Indeed, that induces the subdistribution monad (DM).

SINGLE TEST SEMANTICS. Let us go back to the original problem.

PREVALENCE:	33%	$\frac{1}{3}\lvert I\rangle + \frac{2}{3}\lvert H\rangle$	$\frac{1}{3}\lvert I\rangle + \frac{2}{3}\lvert H\rangle$	(P)
SENSITIVITY:	75%	$\frac{3}{4}\lvert P\rangle + \frac{1}{4}\lvert N\rangle$ when ill	$\frac{1}{4}\lvert IP\rangle + \frac{1}{12}\lvert IN\rangle + \frac{1}{3}\lvert HP\rangle + \frac{1}{3}\lvert HN\rangle$	(C)
SPECIFICITY:	50%	$\frac{1}{2}\lvert P\rangle + \frac{1}{2}\lvert N\rangle$ when healthy	$\frac{1}{5}\lvert IN\rangle + \frac{4}{5}\lvert HN\rangle$	(O)

Let us propose monadic semantics. The distribution monad (D) can compose probabilities, but cannot apply observations.

$$DX = \left\{ \sum_{i=0}^{n} \lambda_i \lvert x_i\rangle \;\middle|\; \sum_{i=0}^{n} \lambda_i = 1,\ x_i \in X,\ \lambda_i \in \mathbb{R}^+ \right\}.$$

Observations yield subdistributions: distributions adding to less then 1. The maybe monad (M) enables them.

$$DMX \cong \left\{ \sum_{i=0}^{n} \lambda_i \lvert x_i\rangle \;\middle|\; \sum_{i=0}^{n} \lambda_i \leq 1,\ x_i \in X,\ \lambda_i \in \mathbb{R}^+ \right\}.$$

Subdistributions provide a monadic semantics for probabilistic inference; normalization remains ad-hoc.

$$P : 1 \nrightarrow \{I, H\} \qquad P = \tfrac{1}{3}\lvert I\rangle + \tfrac{2}{3}\lvert H\rangle$$

$$C : \{I, H\} \nrightarrow \{P, N\} \qquad C(I) = \tfrac{3}{4}\lvert P\rangle + \tfrac{1}{4}\lvert N\rangle;\ C(H) = \tfrac{1}{2}\lvert P\rangle + \tfrac{1}{2}\lvert N\rangle.$$

$$O : \{P, N\} \nrightarrow 1 \qquad O(P) = 0;\quad O(N) = 1.$$

$$P : 1 \longrightarrow DM\{I, H\}$$

$$C : \{I, H\} \longrightarrow DM\{P, N\}$$

$$O : \{P, N\} \longrightarrow DM1.$$

PART 1: NORMALIZED SEMANTICS

"Whenever there is [an observation], it is good to renormalize and resample[...] avoiding too many program executions with low [probability]."

— ☐ STATON, YANG, WOOD, HEUNEN, KAMMAR, 2016.

NORMALIZED SEMANTICS. In practice, probabilistic semantics is not concerned with validity: both $\frac{1}{3}|x\rangle + \frac{2}{3}|y\rangle$ and $\frac{1}{6}|x\rangle + \frac{1}{3}|y\rangle$ represent the same distribution. While not every distribution can be normalized, the only exception is the empty distribution. Technically, we work with subdistributions adding to exactly 1 or 0.

$$\mathsf{MD}X \cong \left\{ \sum_{i=0}^{n} \lambda_i |x_i\rangle \;\middle|\; \sum_{i=0}^{n} \lambda_i = 1 \text{ or } 0,\ x_i \in X,\ \lambda_i \in \mathbb{R}^+ \right\}.$$

PROPOSITION. Normalization defines a natural transformation, $n_X : \mathsf{DM}X \longrightarrow \mathsf{MD}X$, given by

$$n_X(0) = 0\,; \qquad n\left(\sum_{i=0}^{n>0} \lambda_i |x_i\rangle\right) = \sum_{i=0}^{n>0} \frac{\lambda_i}{\sum_i^n \lambda_i} |x_i\rangle.$$

Alternatively, calling $d(x)$ to the formal coefficient of x in the distribution d,

$$n(d)(x) = \left| \frac{d(x)}{\sum_{x'} d(x')} \right|, \text{ returning } 0 \text{ whenever } \sum_{x'} d(x') = 0.$$

PROPOSITION. Normalization defines a monoidal natural transformation.

$$n_X\left(\sum_i^n \lambda_i |x_i\rangle\right) \cdot n_Y\left(\sum_j^m \mu_j |y_j\rangle\right) = n_{X\times Y}\left(\sum_i^n \sum_j^m \lambda_i \mu_j |x_i\rangle |y_j\rangle\right).$$

NORMALIZED KERNELS. Each substochastic kernel, $f : X \to \mathsf{DM}Y$, induces $f^{\circ} : X \to \mathsf{MD}Y$, a kernel obtained by normalization. Each normalized kernel, $g : X \to \mathsf{MD}Y$, induces a substochastic kernel, $g^{\bullet} : X \to \mathsf{DM}Y$. This is a split idempotent, with $g^{\bullet\circ} = g$ and $f^{\circ\bullet\circ\bullet} = f^{\circ\bullet}$.

Normalized kernels inherit a composition, $(g_1 ⨾ g_2) = (g_1^{\bullet} ; g_2^{\bullet})^{\circ}$, but it fails to be associative. Indeed, normalization is not functorial, $(f_1^{\circ\bullet} ; f_2^{\circ\bullet}) \neq (f_1 ; f_2)^{\circ\bullet}$, but it satisfies a left absorptive property, $(f_1^{\circ\bullet} ; f_2)^{\circ\bullet} = (f_1 ; f_2)^{\circ\bullet}$.

PROPOSITION. Normalization is not a distributive law, it misses D-multiplicativity.

PROOF. Let $X = \{a, b\}$ and consider $\frac{1}{2}\big|\, \frac{1}{3}|x\rangle + \frac{2}{3}|\bot\rangle\big\rangle + \frac{1}{2}\big|\, \frac{2}{3}|y\rangle + \frac{2}{3}|\bot\rangle\big\rangle \in \mathsf{DDM}X$.

$$
\begin{array}{lll}
\frac{1}{2}\big|\frac{1}{3}|x\rangle + \frac{2}{3}|\bot\rangle\big\rangle + \frac{1}{2}\big|\frac{2}{3}|y\rangle + \frac{1}{3}|\bot\rangle\big\rangle & \in \mathsf{DDM}X & \text{NORMALIZE} \\
\frac{1}{2}\big||x\rangle\big\rangle + \frac{1}{2}\big||y\rangle\big\rangle + 0|\bot\rangle & \in \mathsf{DMD}X & \text{NORMALIZE} \\
\frac{1}{2}\big||x\rangle\big\rangle + \frac{1}{2}\big||y\rangle\big\rangle & \in \mathsf{MDD}X & \text{MULTIPLY} \\
\frac{1}{2}|x\rangle + \frac{1}{2}|y\rangle & \in \mathsf{MD}X &
\end{array}
$$

$$
\begin{array}{lll}
\frac{1}{2}\big|\frac{1}{3}|x\rangle + \frac{2}{3}|\bot\rangle\big\rangle + \frac{1}{2}\big|\frac{2}{3}|y\rangle + \frac{1}{3}|\bot\rangle\big\rangle & \in \mathsf{DDM}X & \text{MULTIPLY} \\
\frac{1}{6}|x\rangle + \frac{1}{3}|y\rangle + \frac{1}{2}|\bot\rangle & \in \mathsf{DM}X & \text{NORMALIZE} \\
\frac{1}{3}|x\rangle + \frac{2}{3}|y\rangle & \in \mathsf{MD}X &
\end{array}
$$

SUBSTOCHASTIC KERNELS. Inclusion of normalized kernels into subdistributions, $(-)^{\bullet} \colon MDX \longrightarrow DMX$ defined by $(\emptyset)^{\bullet} = 1|\bot\rangle$ and $(\sum_{i=0}^{n>0} \lambda_i |x_i\rangle)^{\bullet} = \sum_{i=0}^{n>0} \lambda_i |x_i\rangle$, is a monoidal distributive law.
The monoidal Kleisli category it induces is the category of substochastic kernels, SubStoch.

$$f \colon X \longrightarrow DMY$$

PARTIAL STOCHASTIC KERNELS. Collapsing of non-normalized distributions, $(-)^{\bot} \colon DMX \longrightarrow MDX$ defined by $(\sum_{i=0}^{n} \lambda_i |x_i\rangle)^{\bot} = [\sum_{i=0}^{n} \lambda_i = 1] \cdot \sum_{i=0}^{n} \lambda_i |x_i\rangle$, is a monoidal distributive law.
The monoidal Kleisli category it induces is the category of partial stochastic functions, ParStoch.

$$f \colon X \longrightarrow MDY$$

NORMALIZED STOCHASTIC KERNELS. Normalization of subdistributions, $(-)^{\circ} \colon DMX \longrightarrow MDX$ defined by $(\sum_{i=0}^{n>0} \lambda_i |x_i\rangle)^{\circ} = \sum_{i=0}^{n>0} (\lambda_i / \sum_{i=0}^{n>0} \lambda_i) |x_i\rangle$ and $(\emptyset)^{\circ} = \emptyset$, is a monoidal non-$D$-multiplicative distributive law.
The monoidal non-associative Kleisli category it induces is the magmoid of normalized kernels, NORM.

$$f \colon X \longrightarrow MDY$$

☐ KOZEN, 1981. ☐ FRITZ, 2018.

PART 2 : SESQUILAWS.

"Finally, we show that there is a unique functorial extension, the so-called black-hole extension."

partial stochastic

— [] SOKOLOVA, WORACEK, 2018.

SESQUILAW. A sesquilaw (m, n, S, T) between two monads, S and T, consists of

1. a distributive law, $m_X : STX \longrightarrow TSX$, and
2. a non-T-multiplicative distributive law, $n_X : TSX \longrightarrow STX$,
3. forming a split idempotent, $\kappa_X = n_X \fatsemi m_X : TSX \longrightarrow TSX$,
4. and satisfying T-multiplicativity up to that idempotent,

$$
\begin{tikzcd}
STX \arrow[r, "m"] \arrow[dr, no head] & TSX \arrow[d, "n"] \\
& STX
\end{tikzcd}
\qquad
\begin{tikzcd}
TSTX \arrow[r, "nT"] \arrow[d, "Tm"'] & STTX \arrow[r, "S\mu"] & STX \\
TTSX \arrow[r, "\mu S"] & TSX \arrow[ur, "n"] &
\end{tikzcd}
$$

EXAMPLE. Inclusion of normalized distributions and normalization form a sesquilaw between the distribution monad (D) and the maybe monad (M). This also works for the Giry monad on standard Borel spaces.

Sesquilaws induce actions. Any sesquilaw, (m, n, S, T), induces a right action of the monad TS into the functor ST: a natural transformation $u_X : STTSX \longrightarrow STX$ defined by either side of this commutative diagram,

$$\begin{tikzcd}
STTSX \arrow[r, "S\mu^T S"] \arrow[d, "mTS"'] \arrow[rrrd, bend right=10, "u_X"] & STSX \arrow[r, "mS"] & TSSX \arrow[d] & \\
TSTSX \arrow[r, "TmS"] & TTSSX \arrow[r, "\mu^T\mu^S"] & TSX \arrow[r, "n"] & STX
\end{tikzcd}$$

making the following two diagrams commute

$$\begin{tikzcd}
STX \arrow[r, "ST\eta^{TS}_X"] \arrow[dr, no head] & STTSX \arrow[d, "u_X"] \\
 & STX
\end{tikzcd}
\qquad
\begin{tikzcd}
STTSTSX \arrow[r, "ST\mu^{TS}"] \arrow[d, "uTS"'] & STTSX \arrow[d, "u_X"] \\
STTSX \arrow[r, "u_X"] & STX
\end{tikzcd}$$

Corollary. Normalized stochastic kernels admit an action from substochastic kernels

$$(\triangleleft) : \textsc{Norm}(X;Y) \times \textsc{SubStoch}(Y;Z) \longrightarrow \textsc{Norm}(X;Z),$$

defined by $(p \triangleleft f) = (p^\bullet ; f)^\circ$ and satisfying $p \triangleleft id = p$ and $p \triangleleft (f ; g) = p \triangleleft f \triangleleft g$.

MONOIDAL MAGMOIDS I: ASSOCIATIVE CENTER. Monoidal coherence needs unitors, associators, and their tensors, and units to satisfy some instances of associativity.

DEFINITION. In a non-associative category, a morphism $f : X \to Y$ is <u>associative</u> whenever, for each $g_1 : X_1 \to X$, each $g_2 : X_2 \longrightarrow X_1$, each $h_1 : Y \longrightarrow Y_1$ and each $h_2 : Y_1 \longrightarrow Y_2$, we have that

$$(g_2 ⨾ g_1) ⨾ f = g_2 ⨾ (g_1 ⨾ f); \quad (g_1 ⨾ f) ⨾ h_1 = g_1 ⨾ (f ⨾ h_1); \quad (f ⨾ h_1) ⨾ h_2 = f ⨾ (h_1 ⨾ h_2).$$

REMARK. Associative morphisms are closed under composition but not necessarily closed under tensoring; we must explicitly pick a closed subclass of associative morphisms.

DEFINITION (Monoidal magmoid). A monoidal magmoid consists of a monoidal category $(\mathbb{A}, \otimes, I)$ and a non-associative category with tensor and unit $(\mathbb{M}, \otimes, I)$ linked by an identity-on-objects functor, $(-)_\uparrow : \mathbb{A} \longrightarrow \mathbb{M}$, whose image is associative, that preserves the tensor structure.

Equivalently, a monoidal magmoid is a <u>non-associative monoidal promonad</u>.

MONOIDAL MAGMOIDS II: COMMUTATIVITY. Commutativity means monoidality in the associative case, but it splits into three notions in the non-associative setting.

1. Monoidality, $(f \otimes id) ⨾ (id \otimes g) = f \otimes g = (id \otimes g) ⨾ (f \otimes id)$.
2. Left commutativity, $(f ⨾ (g \otimes id)) ⨾ (id \otimes h) = f ⨾ (g \otimes h) = (f ⨾ (id \otimes h)) ⨾ (g \otimes id)$.
3. Right commutativity, $(f \otimes id) ⨾ ((id \otimes g) ⨾ h) = (f \otimes g) ⨾ h = (id \otimes g) ⨾ ((f \otimes id) ⨾ h)$.

$$
\begin{tikzcd}
 & T(TA \otimes B) \arrow[dr, "R"] & & \\
TA \otimes TB \arrow[ur, "L"] \arrow[dr, "R"'] & & TT(A \otimes B) \arrow[r, "\mu"] & T(A \otimes B) \\
 & T(A \otimes TB) \arrow[ur, "L"'] & &
\end{tikzcd}
$$

MONOIDALITY

$$
\begin{tikzcd}
 & TT(TX \otimes Y) \arrow[r, "\mu"] & T(TX \otimes Y) \arrow[dr, "R"] & & \\
T(TX \otimes TY) \arrow[ur, "L"] \arrow[dr, "R"'] & & & TT(X \otimes Y) \arrow[r, "\mu"] & T(X \otimes Y) \\
 & TT(X \otimes TY) \arrow[r, "\mu"'] & T(X \otimes TY) \arrow[ur, "L"'] & &
\end{tikzcd}
$$

(LEFT) COMMUTATIVITY

In the associative case, MONOIDALITY implies both forms of COMMUTATIVITY, which coincide.

PROPOSITION. Monoidal sesquilaws induce monoidal left commutative non-associative monads.

PART 3: DO-NOTATION

```
(leftDo NORM
   patient ← prevalence
   test ← (channel patient)
   result ← uncertainty
   () ← (observe test result)
   return patient)
```

```
(leftDo NORM
   result ← uncertainty
   patient ← (leftDo NORM
                 patient ← prevalence
                 test ← (channel patient)
                 () ← (observe test result)
                 return patient)
   return patient)
```

DO-NOTATION. Moggi's system PL is a metalanguage for monadic computation popularized as DO-notation; it takes semantics over cartesian closed categories with a monad. It consists of two syntax rules and leaves variable management to the metalevel.

$$\left[\!\left[\begin{array}{l}\mathrm{do}\ T \\ \quad \mathrm{return}\ x\end{array}\right]\!\right] = \eta_T(x)$$

$$\left[\!\left[\begin{array}{l}\mathrm{do}\ T \\ \quad x \leftarrow e \\ \quad \mathrm{rest}\ldots\end{array}\right]\!\right] = \beta_T(e)\left(\lambda x. \left[\!\left[\begin{array}{l}\mathrm{do}\ T \\ \quad \mathrm{rest}\ldots\end{array}\right]\!\right]\right)$$

The monad T is presented in relative form: $\eta_T : X \to TX$ and $\beta_T : TX \times (X \to TY) \to TY$.

As an example, consider the single test problem over the subdistribution monad.

```
(do SUBDIST
    patient ⟵ prior
    result  ⟵ channel(patient)
    ()      ⟵ observe(Negative, result)
    RETURN(patient))
```

```
β(prior)(λ patient.
   β(channel(patient))(λ result.
      β(observe(Negative, result))(λ.
          patient))).
```

LEFT DO. The monadic metalanguage is naturally right associative: variables bind to the right. For non-associative monads, this becomes relevant: we may associate to the left.

$$\left[\begin{array}{l} \mathrm{leftDo}\ \mathbb{T} \\ \quad x_1 \leftarrow e_1 \\ \quad x_2 \leftarrow e_2 \\ \quad \mathrm{rest}\ldots \end{array}\right] = \left[\begin{array}{l} \mathrm{leftDo}\ \mathbb{T} \\ \quad (x_1, x_2) \leftarrow \left[\begin{array}{l} \mathrm{do}\ \mathbb{T} \\ \quad x_1 \leftarrow e_1 \\ \quad x_2 \leftarrow e_2 \\ \quad \mathrm{return}\ (x_1, x_2) \end{array}\right] \\ \quad \mathrm{rest}\ldots \end{array}\right];$$

$$\left[\begin{array}{l} \mathrm{leftDo}\ \mathbb{T} \\ \quad x \leftarrow e \\ \quad \mathrm{return}\ v \end{array}\right] = \left[\begin{array}{l} \mathrm{do}\ \mathbb{T} \\ \quad x \leftarrow e \\ \quad \mathrm{return}\ v \end{array}\right]; \qquad \left[\begin{array}{l} \mathrm{leftDo}\ \mathbb{T} \\ \quad \mathrm{return}\ v \end{array}\right] = \left[\begin{array}{l} \mathrm{do}\ \mathbb{T} \\ \quad \mathrm{return}\ v \end{array}\right].$$

Again, assuming a non-associative monad presented in relative form.

$$\begin{array}{l} (\mathrm{leftDo}\ \mathrm{SUBDIST/NORM} \\ \quad \mathrm{patient} \longleftarrow \mathrm{prevalence} \\ \quad \mathrm{result} \longleftarrow \mathrm{channel}\,(\mathrm{patient}) \\ \quad () \longleftarrow \mathrm{observe}\,(\mathrm{Negative}, \mathrm{result}) \\ \quad \mathrm{RETURN}\,(\mathrm{patient})) \end{array}$$

$$\begin{array}{l} \beta(\beta(\beta(\mathrm{prevalence})(\lambda\,\mathrm{patient}. \\ \qquad \beta(\mathrm{channel}(\mathrm{patient}))(\lambda\,\mathrm{result}. \\ \qquad\quad (\mathrm{patient}, \mathrm{result}))) \\ \quad (\lambda(\mathrm{patient}, \mathrm{result}).\ \beta(\mathrm{observe}(\mathrm{Negative})(\mathrm{result}) \\ \qquad\qquad (\lambda.\ (\mathrm{patient}, \mathrm{result}))) \\ \quad (\lambda(\mathrm{patient}, \mathrm{result}).\ \mathrm{patient}) \end{array}$$

UNCLEAR TEST PROBLEM. Imagine the test result was uncertain — e.g. it was dark!

PREVALENCE : 33%	$\frac{1}{3}\|I\rangle + \frac{2}{3}\|H\rangle$	
SENSITIVITY : 75%	$\frac{3}{4}\|P\rangle + \frac{1}{4}\|N\rangle$	when Ill
SPECIFICITY : 50%	$\frac{1}{2}\|P\rangle + \frac{1}{2}\|N\rangle$	when Healthy
UNCERTAINTY : 75%	$\frac{3}{4}\|P\rangle + \frac{1}{4}\|N\rangle$	

SOLUTION 1. (~38% ill, ~ 62% healthy)

~38.5%

$\frac{1}{3}\|I\rangle + \frac{2}{3}\|H\rangle$	PREVALENCE
$\frac{1}{3}\|I\rangle(\frac{3}{4}\|P\rangle + \frac{1}{4}\|N\rangle) + \frac{2}{3}\|H\rangle(\frac{1}{2}\|P\rangle + \frac{1}{2}\|N\rangle)$	CHANNEL
$\frac{1}{3}\|I\rangle(\frac{3}{4}\|P\rangle + \frac{1}{4}\|N\rangle)(\frac{3}{4}\|P\rangle + \frac{1}{4}\|N\rangle) + \frac{2}{3}\|H\rangle(\frac{1}{2}\|P\rangle + \frac{1}{2}\|N\rangle)(\frac{3}{4}\|P\rangle + \frac{1}{4}\|N\rangle)$	UNCERTAINTY
$\frac{3}{16}\|IP\rangle + \frac{1}{48}\|IN\rangle + \frac{1}{4}\|HP\rangle + \frac{1}{12}\|HN\rangle$	OBSERVE
$\frac{5}{24}\|I\rangle + \frac{1}{3}\|H\rangle$	PROJECT
$\frac{5}{13}\|I\rangle + \frac{8}{13}\|H\rangle$	NORMALIZE

SOLUTION 2. (~37% ill, ~ 63% healthy)

~37.1%

$\frac{3}{4}\|P\rangle + \frac{1}{4}\|N\rangle$	UNCERTAINTY
$(\frac{3}{4}\|P\rangle + \frac{1}{4}\|N\rangle)(\frac{1}{3}\|I\rangle + \frac{2}{3}\|H\rangle)$	PREVALENCE
$(\frac{3}{4}\|P\rangle + \frac{1}{4}\|N\rangle)(\frac{1}{3}\|I\rangle(\frac{3}{4}\|P\rangle + \frac{1}{4}\|N\rangle) + \frac{2}{3}\|H\rangle(\frac{1}{2}\|P\rangle + \frac{1}{2}\|N\rangle))$	CHANNEL
$\frac{3}{4}\|P\rangle(\frac{1}{4}\|IP\rangle + \frac{1}{3}\|HP\rangle) + \frac{1}{4}\|N\rangle(\frac{1}{12}\|IN\rangle + \frac{1}{3}\|H\rangle))$	OBSERVE
$\frac{3}{4}\|P\rangle(\frac{3}{7}\|I\rangle + \frac{4}{7}\|H\rangle) + \frac{1}{4}\|N\rangle(\frac{1}{5}\|I\rangle + \frac{4}{5}\|H\rangle)$	NORMALIZE
$\frac{9}{28}\|PI\rangle + \frac{3}{7}\|PH\rangle + \frac{1}{20}\|NI\rangle + \frac{1}{5}\|NH\rangle$	PROJECT/MULTIPLY
$\frac{13}{35}\|I\rangle + \frac{22}{35}\|H\rangle$	PROJECT

We get two different numbers: this is terrible. (c.f. Jacobs 2018).

PART 3 : DO-NOTATION

```
(leftDo NORM
   patient ← prevalence
   test ← (channel patient)
   result ← uncertainty
   () ← (observe test result)
   return patient)
```

```
(leftDo NORM
   result ← uncertainty
   patient ← (leftDo NORM
                patient ← prevalence
                test ← (channel patient)
                () ← (observe test result)
                return patient)
   return patient)
```