

Getting to Grips with Sweedler Runners for Systems

A conversational tutorial, Part XA: probabilistic worlds and quasi-Borel runners
Residual randomness, hidden Markov models, conditioning, and the road to filtering–control
duality

October 2026

Aim. Probability can enter the Sweedler programme in two different places. We may keep the ambient world deterministic and use a probabilistic *residual theory*, as in the **rand** examples of Part II. Or we may change the ambient world itself, so that state spaces, parameter spaces, function spaces and theories all live in a category designed for continuous probability. This Part develops the distinction. We first identify finite edge-emitting hidden Markov dynamics with residual Mealy runners, while keeping track of the missing initial law and of the obstruction imposed by Moore reader equations. We then investigate quasi-Borel spaces as a candidate probabilistic base. The middle of the Part develops the finite filter in two stages—linear determinisation into finite measures, followed by partial normalisation—and only then asks what disintegration must supply for non-discrete observations. This exposes an interface reversal from a stochastic server to a deterministic observer, and a forward/backward pairing $L_o \dashv L_o^*$ which gives a concrete finite test for the filtering–control duality raised in Part IX.

Contents

1	Why another change of ambient world?	2
2	Residual probability before changing base	3
2.1	A warning from Part II: Moore observation is not arbitrary noisy observation . . .	3
3	The correct elementary HMM runner	4
3.1	From transition and emission kernels to the joint runner	5
3.2	Controlled HMMs and the Mealy interface	6
4	So why not remain in Set?	6
5	Quasi-Borel spaces in one page	6
6	What should the QBS Sweedler universe be?	7
6.1	What is theorem-level and what remains a checkpoint?	8
7	The QBS probability monad as a residual theory	8
8	Hidden Markov models in QBS	9

9	Generative semantics is not yet filtering	9
10	Finite filtering by determinisation	10
11	Normalisation and the ordinary Bayes filter	10
12	Where genuine conditioning enters	11
13	Filtering reverses the interface	12
14	A second object: the belief-state HMM	12
15	Probabilistic handlers and the adjoint operators	13
16	Reachability, observability and weighted minimisation	14
17	Returning to the Kim–Mehta challenge	14
18	What QBS buys, and what it does not	15
19	The relation with Part X: smoothness and probability	15
20	A staged research programme	16
20.1	Stage 1: finite residual probability	16
20.2	Stage 2: finite determinisation and duality	16
20.3	Stage 3: the QBS ambient theorem	16
20.4	Stage 4: continuous observations	16
20.5	Stage 5: probabilistic behaviour	16
20.6	Stage 6: the Wonham bridge	17
20.7	Stage 7: continuous-time duality	17
21	What is established and what remains conjectural	17
22	Where Part XA leaves us	17
	References and orientation	18

1 Why another change of ambient world?

Part X changed the geometry of state by passing from sets to a smooth setting. That move did more than decorate ordinary state spaces with differentiability: it changed which parameter dependences and infinitesimal evolutions were legitimate. Probability invites the same question.

We already know one way to add randomness. In the original Set-based universe

$$\mathcal{D}_{\mathbf{Set}} = [\mathbf{Set}, \mathbf{Set}]_{\text{acc}},$$

we may choose a probability monad as the residual theory. A runner

$$Y : T \rightsquigarrow \text{rand}$$

then implements an abstract T -operation by a probabilistic residual computation. This was the point of the **rand** examples in Part II.

But there is a second possibility. Replace **Set** itself by a category in which continuous random variables and higher-order function spaces coexist. Then the state object Y , observation object O , parameter object A , theories acting on them, and residual probability monad can all be probabilistic objects in the same ambient semantics.

The distinction is worth keeping visible:

probability in the residual $T \rightsquigarrow \mathbf{rand}$ in a Set -based universe	versus	probability in the ambient world $T \rightsquigarrow \mathbf{P}$ in a QBS-based universe.
---	--------	--

The first is already useful and should not be treated as a deficient approximation to the second. The question is what extra mathematics the second route buys.

2 Residual probability before changing base

Let **Dist** denote the finite-distribution monad on **Set**. Concretely,

$$\mathbf{Dist}X$$

is the set of finitely supported probability distributions on X , with unit given by Dirac measures and multiplication by averaging a distribution of distributions. Its commutativity expresses the fact that independent finite probabilistic choices may be interchanged.

A residual runner with state Y

$$Y : T \rightsquigarrow \mathbf{Dist}$$

is a natural family

$$\theta_X : TX \times Y \longrightarrow \mathbf{Dist}(X \times Y)$$

compatible with the unit and multiplication of T . Thus an abstract deterministic theory can be implemented by a stochastic lower-level mechanism.

This distinction between abstract interface and residual effect is exactly what runners were built to express. The theory T says what operation the environment is allowed to request. The residual monad says what effects the implementation may use while answering that request.

2.1 A warning from Part II: Moore observation is not arbitrary noisy observation

It is tempting to say immediately that

$$T_{\mathbf{Moore}}(1, O) \rightsquigarrow \mathbf{Dist}$$

is an arbitrary hidden Markov model. That is too quick. The observation operation in the Moore theory is a *reader*: its equations say that reading is a pure observation of the present abstract state. Those equations remain equations after we pass to a residual runner.

In the finite-distribution case one can say something stronger than “the runner cannot resample on every read.” It cannot randomise the read at all.

Reader rigidity. Let

$$Y : T_{\text{rd}}(O) \rightsquigarrow \text{Dist}$$

be a residual runner for the reader theory, with cooperation

$$\kappa : Y \longrightarrow \text{Dist}(O \times Y).$$

Then there is a unique function $f : Y \rightarrow O$ such that

$$\kappa(y) = \delta_{(f(y), y)}.$$

Indeed, the reader law

$$\text{get}(\lambda o. x) = x$$

forces the Y -marginal of $\kappa(y)$ to be δ_y . Hence

$$\kappa(y) = \rho(y) \otimes \delta_y$$

for some $\rho(y) \in \text{Dist}O$. Instantiating the diagonal reader law

$$\text{get}(\lambda o. \text{get}(\lambda o'. k(o, o')))) = \text{get}(\lambda o. k(o, o))$$

at $X = O \times O$ with $k = \eta$ gives

$$\rho(y) \otimes \rho(y) = \Delta_* \rho(y),$$

which for a finite probability distribution holds only when $\rho(y)$ is Dirac. Thus the read neither moves the state nor adds fresh randomness.

A Moore presentation is still available if one is willing to enlarge the hidden state. For example, one may store the currently sampled emission in a state $Y \times O$ and sample the next emission during the transition. The observation map is then deterministic on the enlarged state. What fails is not the interface shape but the attempt to identify an arbitrary noisy emission directly with the reader cooperation.

For general QBS objects one should phrase the corresponding claim more carefully. The same argument forces the induced law on observations to be $\{0, 1\}$ -valued; for standard Borel observation objects this is Dirac, while pathological non-standard measurable structures can admit non-Dirac 0-1 laws. The finite and standard-Borel cases used below have no such ambiguity.

This is a useful lesson for the whole programme. Choosing a residual effect changes how an operation is implemented; it does not change what equations the operation is required to satisfy.

3 The correct elementary HMM runner

For a hidden Markov step we want one round of interaction to return an observation and update a hidden state. The deterministic skeleton is therefore not the pure-reader half of a Moore machine but the autonomous Mealy operation

$$Y \longrightarrow O \times Y.$$

Equivalently, take the free theory on one O -ary operation, or the one-input instance of the Mealy theory. We write it schematically as

$$T_{\text{Mealy}}(1, O).$$

A comodel has one cooperation $Y \rightarrow O \times Y$, with no reader equations forcing the output to be repeatable without state change.

Replacing Id by Dist , a residual runner

$$Y : T_{\text{Mealy}}(1, O) \rightsquigarrow \text{Dist}$$

is determined on the generating operation by

$$\boxed{\kappa : Y \longrightarrow \text{Dist}(O \times Y).}$$

This is precisely the one-step kernel of an *edge-emitting* finite hidden Markov system: given the present hidden state, the emitted symbol and next hidden state may be correlated.

There are two qualifications worth recording immediately. First, the state-emitting textbook HMMs form a subclass whose joint kernel factors through separate transition and emission kernels. Up to state enlargement the two presentations generate the same kinds of output processes, but they are not the same kernels. Second, an HMM as a stochastic process also includes an initial law

$$\pi_0 \in \text{Dist}Y.$$

The runner supplies the dynamics; the basepoint π_0 supplies a particular generated process. This distinction will become important when we compare forward generation with backward observables.

3.1 From transition and emission kernels to the joint runner

A state-emitting HMM is often presented by

$$q : Y \longrightarrow \text{Dist}Y, \quad e : Y \longrightarrow \text{Dist}O.$$

If the present state emits and then transitions, strength gives maps combining the two distributions into a law on $O \times Y$, and commutativity of Dist makes the two orders of independent combination agree. Concretely,

$$Y \xrightarrow{\Delta} Y \times Y \xrightarrow{e \times q} \text{Dist}O \times \text{Dist}Y \xrightarrow{\text{dst}} \text{Dist}(O \times Y).$$

If instead the system first transitions and then emits from the new state, the joint kernel is the Kleisli composite

$$y \longmapsto q(y) \gg= (y' \mapsto e(y')) \gg= (o \mapsto \eta(o, y')).$$

Both are special cases of the arbitrary edge-emitting kernel κ .

Thus the precise elementary statement is

$$\boxed{\begin{array}{l} \text{edge-emitting finite HMM dynamics} \\ \text{without a chosen initial law} \end{array} = T_{\text{Mealy}}(1, O) \rightsquigarrow \text{Dist runner}.}$$

Choosing π_0 points the runner and determines the law of the observable process.

This is our first probabilistic appearance of the asymmetry emphasised in Part IX. The backward action of the runner on handlers is defined without a chosen state distribution. To obtain a particular forward trace law, by contrast, one must choose π_0 . We shall make this precise below for the operators L_o and L_o^* .

3.2 Controlled HMMs and the Mealy interface

Let A be a set of controls or inputs. A controlled stochastic transducer has a kernel

$$\kappa : A \times Y \longrightarrow \text{Dist}(O \times Y),$$

and is therefore the generator-level form of a residual runner

$$Y : T_{\text{Mealy}}(A, O) \rightsquigarrow \text{Dist}.$$

This is the natural setting for edge-emitting controlled HMMs and for POMDP-like generative systems. Sequencing repeated T -operations is governed by the monad-morphism laws of the runner. This should not be confused with horizontal composition of a runner $T \rightsquigarrow \text{Dist}$ with another runner out of Dist , which would require additional residual implementation data for probability itself.

4 So why not remain in Set?

For finite probability there is no reason to leave **Set** merely in order to write the HMM kernel. The residual route already does that perfectly well. The pressure to change base comes from other requirements.

Suppose that

- the hidden state is continuous;
- observations are real-valued;
- parameters themselves vary measurably;
- one wants distributions on functions or stochastic higher-order components;
- filtering manipulates spaces of probability distributions as ordinary semantic objects.

The category **Meas** of measurable spaces is the classical home for measurable maps, but it is not cartesian closed. Function spaces therefore do not interact with measurable structure in the way required by ordinary higher-order categorical semantics. This is precisely the problem that quasi-Borel spaces were designed to repair [1].

The issue matters especially for the present tutorial. Sweedler theory repeatedly asks us to form internal homs, state-indexed maps, parameterised operations, and endofunctor constructions. A probabilistic base which loses good function spaces at the outset is an awkward place in which to ask those questions.

5 Quasi-Borel spaces in one page

A quasi-Borel space X consists of an underlying set $|X|$ together with a specified collection

$$M_X \subseteq \mathbf{Set}(\mathbb{R}, |X|)$$

of admissible random elements. These contain constant maps, are closed under precomposition by measurable maps $\mathbb{R} \rightarrow \mathbb{R}$, and are closed under countable measurable case distinctions. A morphism

$$f : X \longrightarrow Y$$

is a function preserving random elements:

$$\alpha \in M_X \implies f\alpha \in M_Y.$$

The resulting category **Qbs** has exactly the properties that make it attractive here. It is cartesian closed, so function spaces are again quasi-Borel spaces; it is well-pointed; and it carries a commutative strong probability monad $\mathbf{P}$ supporting continuous probability distributions [1]. It is also a locally presentable quasitopos; one explicit result gives local $\mathfrak{c}^+$ -presentability [2]. Thus it is large enough for continuous probability while retaining substantial algebraic and categorical structure.

For standard Borel spaces, the QBS probability monad agrees with the familiar measure-theoretic probability monad in the expected sense. But QBS also has function spaces such as

$$Y^X$$

and can therefore discuss random functions and probability measures on higher-order objects directly.

The slogan is not that QBS replaces all of measure theory. It is that QBS is a convenient semantic category in which continuous probability and higher-order function spaces coexist. Some measure-theoretic constructions require additional hypotheses or additional structure, and conditioning will be our main example.

6 What should the QBS Sweedler universe be?

The Set-based notation

$$[\mathbf{Set}, \mathbf{Set}]_{\text{acc}}$$

is slightly deceptive when transported to a non-Set base. Over **Set**, ordinary and Set-enriched functors coincide. Over **Qbs**, the enrichment is load-bearing.

Because **Qbs** is cartesian closed, a tensorial strength on an endofunctor is equivalent to enrichment over **Qbs** in the standard Kock sense. The strong QBS probability monad therefore belongs naturally to the self-enriched endofunctor world. Moreover, the elementary reduction of a natural family

$$X^O \times Y \longrightarrow \mathbf{P}(X \times Y)$$

to a single cooperation

$$Y \longrightarrow \mathbf{P}(O \times Y)$$

is an enriched Yoneda calculation. If O is a genuinely non-discrete QBS object, the theory itself must consequently be understood as a QBS-enriched algebraic theory with arity object O , not as the Set-theory with an underlying set of operations pasted in afterwards.

Choose a regular cardinal κ large enough for the accessibility calculations under consideration and write

$$\mathcal{D}_{\mathbf{Qbs}, \kappa} = \mathbf{Qbs}\text{-}[\mathbf{Qbs}, \mathbf{Qbs}]_{\kappa}.$$

The intended objects are κ -accessible QBS-enriched endofunctors. Composition gives the first monoidal product. Because **Qbs** is cartesian closed, the second product should be the enriched Day convolution determined by the cartesian tensor:

$$(F \star G)(X) = \int^{A, B \in \mathbf{Qbs}_{\kappa}} \mathbf{Qbs}(A \times B, X) \times F(A) \times G(B),$$

with homs, products and coends interpreted in the QBS enrichment. The interchange transformation is inherited from the cartesian monoidal structure exactly as in the enriched resource calculation of Part VI.

6.1 What is theorem-level and what remains a checkpoint?

Part IV taught us not to write $\mathbf{Run}(\mathcal{D})$ until the ambient hypotheses have been checked. QBS is unusually promising: it is cartesian closed, and the PPS analysis proves that it is locally $\mathfrak{c}^+$ -presentable, with an explicit description of presentable objects [2]. Thus fixed-rank enriched universal algebra is available.

The remaining work is more specific. The Sweedler theorem used throughout the tutorial requires a duoidal category of accessible enriched endofunctors with the relevant colimits and convolution closures, and then existence of the measuring objects/classifying comonads. We therefore still have to check:

1. closure of the chosen fixed- κ enriched endofunctors under composition and enriched Day convolution;
2. the precise interchange and cocontinuity hypotheses required by the enriched Sweedler construction;
3. accessibility of the particular probability endofunctor $\mathbf{P}$ at a workable rank;
4. the corresponding existence of the measuring objects used to classify residual runners.

Thus

$$\mathbf{Run}(\mathcal{D}_{\mathbf{Qbs}, \kappa})$$

remains a strongly motivated working universe rather than a theorem silently imported from the Set case. The point is nevertheless sharper than before: enrichment is not optional decoration, since both the probability monad and the arity calculation for continuous observations live there.

7 The QBS probability monad as a residual theory

The probability construction of Heunen–Kammar–Staton–Yang gives a strong affine commutative monad

$$\mathbf{P} : \mathbf{Qbs} \longrightarrow \mathbf{Qbs}$$

[1]. Following their definition, an element of $\mathbf{P}X$ is represented by a random element together with a probability measure on the source, modulo equality of the induced pushforward probability measure on the σ -algebra generated by the QBS structure. Equality of integration operators is an equivalent extensional description, but the pushforward-measure quotient is the concrete definition to keep in mind.

The strength gives

$$X \times \mathbf{P}Y \longrightarrow \mathbf{P}(X \times Y),$$

and commutativity identifies the two canonical ways of combining independent probabilistic choices, giving the Fubini map

$$\mathbf{P}X \times \mathbf{P}Y \longrightarrow \mathbf{P}(X \times Y).$$

These are exactly the structural operations used when separate transition and emission kernels are assembled into a joint HMM step.

Provided $\mathbf{P}$ lies in the chosen accessible enriched endofunctor category, it is a monoid object for composition and hence a residual theory in the candidate QBS Sweedler universe.

8 Hidden Markov models in QBS

Let $Y, O \in \mathbf{Qbs}$. The continuous/higher-order analogue of the finite edge-emitting HMM dynamics is

$$\boxed{\kappa : Y \longrightarrow \mathbf{P}(O \times Y)}.$$

When O is not discrete, however, the theory behind the notation $T_{\text{Mealy}}(1, O)$ must itself be rebuilt as a $\mathbf{Qbs}$ -enriched theory with arity object O . The formal runner equation then has the same appearance as in $\mathbf{Set}$, but its justification is enriched Yoneda rather than ordinary $\mathbf{Set}$ -representability. Subject to the ambient Sweedler existence checks of the previous section, the kernel is therefore read as

$$Y : T_{\text{Mealy}, \mathbf{Qbs}}(1, O) \rightsquigarrow \mathbf{P}.$$

For an input object A , the controlled version is

$$\kappa : A \times Y \longrightarrow \mathbf{P}(O \times Y),$$

corresponding schematically to

$$Y : T_{\text{Mealy}, \mathbf{Qbs}}(A, O) \rightsquigarrow \mathbf{P}.$$

The gain from QBS is not that the formula for an HMM changes. The gain is that continuous state, continuous observations, parameterised kernels and higher-order function spaces live in the same cartesian closed semantic category.

As in the finite case, a particular stochastic process still requires an initial law

$$\pi_0 \in \mathbf{P}Y.$$

The runner is unpointed dynamics; the initial law chooses a generated process.

9 Generative semantics is not yet filtering

A kernel

$$\kappa : Y \longrightarrow \mathbf{P}(O \times Y)$$

generates hidden-state/observation dynamics. Filtering asks a different question. Given a prior law and an observed value, how should the observer update its law on hidden state?

The finite case shows that it is misleading to insert “conditioning” as one undifferentiated operation between these two levels. There are really two stages. First comes a *linear, total* determinisation into unnormalised measures. Only afterwards comes *normalisation*, which is partial at zero evidence and nonlinear. Genuine disintegration enters when the observation object is non-discrete and the first reshaping itself is no longer available by elementary currying.

This distinction will let us see much more of the filtering construction inside ordinary algebra before asking QBS to solve the analytic part.

10 Finite filtering by determinisation

Let Y and O be finite and write

$$\kappa(y)(o, y')$$

for the joint probability of emitting o and moving to y' . Let $\mathcal{M}Y$ denote the commutative monoid of finite nonnegative measures on Y , rather than only probability measures.

For each observation $o \in O$, define a positive linear operator

$$L_o : \mathcal{M}Y \longrightarrow \mathcal{M}Y$$

by

$$(L_o \sigma)(y') = \sum_{y \in Y} \sigma(y) \kappa(y)(o, y').$$

Equivalently, the elementary currying

$$\text{Dist}(O \times Y) \longrightarrow (\mathcal{M}Y)^O, \quad \nu \longmapsto (o \mapsto \nu(o, -)),$$

turns the stochastic runner into a coalgebra

$$Y \longrightarrow (\mathcal{M}Y)^O.$$

The generalised powerset construction then extends it linearly from points of Y to measures on Y , producing

$$\mathcal{M}Y \longrightarrow (\mathcal{M}Y)^O, \quad \sigma \longmapsto (o \mapsto L_o \sigma).$$

This is the unnormalised finite filter [5, 6].

If $\pi_0 \in \text{Dist}Y$ is the initial law and $\mathbf{1} : Y \rightarrow \mathbb{R}_{\geq 0}$ is the constant-one function, then

$$\pi_0 L_{o_1} \cdots L_{o_n} \mathbf{1}$$

is the probability of the observation word $o_1 \cdots o_n$. Thus the pointed HMM gives a weighted automaton over $\mathbb{R}_{\geq 0}$: the initial vector is π_0 , the transition matrices are the L_o , and the final vector is $\mathbf{1}$.

For finite observations, no disintegration is needed to reach the unnormalised filter. The construction is

$$\text{stochastic runner} \longrightarrow \text{linear determinisation on } \mathcal{M}Y.$$

The probabilistic nonlinearity enters only when we insist that the state remain normalised.

11 Normalisation and the ordinary Bayes filter

Write the total mass of $\sigma \in \mathcal{M}Y$ as

$$|\sigma| = \sum_y \sigma(y).$$

Given a prior $\pi \in \text{Dist}Y$, the unnormalised posterior after observing o is $L_o \pi$. Its evidence is

$$Z_\pi(o) = |L_o \pi|.$$

Whenever $Z_\pi(o) > 0$, the usual posterior is

$$\text{Filt}(\pi, o) = \frac{L_o \pi}{|L_o \pi|}.$$

Hence

$$\text{Filt} : \text{Dist}Y \times O \dashrightarrow \text{Dist}Y$$

is the composite of a total linear map into finite measures and a partial nonlinear normalisation.

The correct finite decomposition is therefore

$$\text{finite filter} = \text{determinisation into } \mathcal{M} \text{ followed by normalisation.}$$

The only partiality is zero evidence. This is more informative than saying merely that “the filter requires conditioning.”

There is a familiar continuous-time analogy. In classical nonlinear filtering the Zakai equation evolves an unnormalised conditional law linearly after a suitable change of measure, while the Kushner–Stratonovich filter is obtained by normalisation. We shall treat this as a guiding analogy rather than as a categorical theorem: making it precise in the present framework requires stochastic integration and hypotheses far beyond the finite calculation.

12 Where genuine conditioning enters

For a non-discrete observation object O , the finite currying

$$\nu \longmapsto (o \mapsto \nu(o, -))$$

has no literal analogue for an arbitrary joint probability measure: a singleton observation typically has measure zero. One needs additional analytic structure.

There are at least two related but distinct categorical routes. Markov categories with conditionals axiomatise disintegration/Bayesian inversion of joint laws and support abstract HMM filtering and smoothing [10, 8, 9]. Probabilistic-programming semantics often works instead with unnormalised or s-finite measures, likelihood weighting and partial normalisation; QBS supports synthetic measure-theoretic versions of this idea [11, 12]. These should not be identified merely under the word “conditioning.”

In a dominated model, a kernel can be represented relative to a reference measure on O by likelihoods, and the finite operators L_o are replaced by likelihood-weighted measure transformers. In a non-dominated model, true disintegration or an equivalent conditional structure is unavoidable.

So the QBS question is now sharper:

$$\text{how do measure-valued determinisation,} \\ \text{normalisation and disintegration interact with runners?}$$

This is a narrower and more testable problem than asking vaguely for “Bayesian inversion inside Sweedler theory.”

13 Filtering reverses the interface

The determinised map

$$\mathcal{M}Y \longrightarrow (\mathcal{M}Y)^O$$

changes the polarity of the observation object. In the generative HMM,

$$Y \longrightarrow \text{Dist}(O \times Y),$$

O is an *output* of a Mealy-style server. In the filter,

$$O \times \mathcal{M}Y \longrightarrow \mathcal{M}Y,$$

O is an *input* supplied to an observer. The total mass

$$|L_o \sigma|$$

is then the predictive weight of that input.

This interface reversal is part of the structural content of filtering. It also resolves the apparent tension with the Moore warning of Section 2.1: noisy emission cannot be represented as a pure Moore output on the server side, but the realised observation is perfectly naturally an input on the observer side.

In the four-corner language used elsewhere in the tutorial, this suggests that the filter belongs on the client/observer side rather than being another presentation of the original server. That placement remains a structural proposal until the corresponding double-categorical construction is written down, but the variance is already visible in the finite equations.

14 A second object: the belief-state HMM

There is nevertheless also a genuine stochastic system on belief states with the *same* generative interface as the original HMM. Define

$$\beta : \text{Dist}Y \longrightarrow \text{Dist}(O \times \text{Dist}Y)$$

by

$$\boxed{\beta(\pi) = \sum_{o \in O} Z_\pi(o) \delta_{(o, \text{Filt}(\pi, o))},}$$

ignoring zero-evidence observations. Started at π_0 , this belief-state HMM generates the same observation-word probabilities as the hidden system started at π_0 .

This object should not be conflated with the deterministic filter. The belief HMM is server-side and stochastic,

$$\text{Dist}Y \rightarrow \text{Dist}(O \times \text{Dist}Y),$$

whereas the filter is client-side and observation-driven,

$$O \times \text{Dist}Y \dashrightarrow \text{Dist}Y.$$

Nor is the Dirac map $y \mapsto \delta_y$ a strict runner morphism from the hidden HMM to the belief HMM: pushing $\kappa(y)$ forward along Dirac states is generally supported on Dirac beliefs, whereas $\beta(\delta_y)$ is supported on posterior mixtures. The relation appears only after barycentric flattening.

Behavioural equivalence, or a Kleisli/lax notion of comparison, is therefore the more plausible categorical target.

Finite test problem. A satisfactory categorical account of filtering should distinguish three constructions:

$$\begin{aligned} Y &\rightarrow \text{Dist}(O \times Y) \quad (\text{hidden generative runner}), \\ \mathcal{M}Y &\rightarrow (\mathcal{M}Y)^O \quad (\text{linear observer/determinisation}), \\ \text{Dist}Y &\rightarrow \text{Dist}(O \times \text{Dist}Y) \quad (\text{belief-state generative system}). \end{aligned}$$

It should explain their common trace semantics without forcing them into the same morphism type.

15 Probabilistic handlers and the adjoint operators

The runner also acts on the handler side. Take the expectation algebra on bounded observables. For finite Y , let

$$V : Y \longrightarrow \mathbb{R}$$

be an observable. For each $o \in O$, define the backward operator

$$L_o^*V : Y \longrightarrow \mathbb{R}$$

by

$$(L_o^*V)(y) = \sum_{y'} \kappa(y)(o, y')V(y').$$

More invariantly, delegation of expectation along the runner gives the $T_{\text{Mealy}}(1, O)$ -algebra

$$\begin{aligned} (\mathbb{R}^Y)^O &\longrightarrow \mathbb{R}^Y, \\ (V_o)_{o \in O} &\longmapsto \left(y \longmapsto \sum_{o, y'} \kappa(y)(o, y')V_o(y') \right). \end{aligned}$$

The component selected by o is exactly L_o^* .

Pair finite measures and observables by

$$\langle \sigma, V \rangle = \sum_y \sigma(y)V(y).$$

Then

$$\langle L_o\sigma, V \rangle = \langle \sigma, L_o^*V \rangle.$$

This is the concrete forward/backward duality attached to the very same probabilistic runner. The forward operator belongs to the determinised observer and needs a basepoint to produce a particular trace law; the backward handler action is defined without one.

This pairing is an excellent candidate small model for the measuring pairings of the Sweedler programme. We should not yet identify the two constructions by theorem, but it gives a precise calculation which any proposed abstract pairing should recover.

16 Reachability, observability and weighted minimisation

The determinised finite HMM is a linear weighted automaton. Its reachable forward space is generated from π_0 by the operators L_o ; dually, backward observables are generated from $\mathbf{1}$ by the L_o^* . The Hankel matrix of word probabilities packages the resulting pairing, and its rank gives the usual finite-dimensional minimal realisation. Coalgebraic treatments of linear weighted automata make this reachability/observability structure explicit [7].

Thus Part IX’s abstract asymmetry acquires a probabilistic linear shadow:

forward reachable laws depend on π_0 ,
backward observable action is basepoint-free.

For HMMs, stochastic observability asks whether distinct initial laws can be distinguished by their observation laws. Van Handel’s work formulates this injectivity idea in nonlinear filtering and relates it to filter stability [14]. Kim and Mehta then dualise stochastic observability to controllability of a backward stochastic control system. The finite weighted-automaton pairing is therefore not the Kim–Mehta theorem, but it is a much closer discrete test case than the bare Markov-operator identity.

17 Returning to the Kim–Mehta challenge

Part IX introduced the work of Kim and Mehta as a demanding test of our reachability–observability programme. Their nonlinear filtering theory studies hidden Markov models with continuous-time noisy observations and constructs a backward stochastic differential equation as a dual control system. In their framework, stochastic observability corresponds to controllability of the dual system, while nonlinear filtering itself is paired with an optimal-control problem [16, 17].

The finite analysis above now gives a ladder rather than a leap:

1. **Generative runner.** A stochastic Mealy runner supplies κ .
2. **Forward observer.** Determinisation gives the linear operators L_o on unnormalised laws.
3. **Backward handler.** Delegation gives the adjoint operators L_o^* on observables.
4. **Pointed semantics.** An initial law π_0 turns the forward system into an observation process and generates the reachable law space.
5. **Normalised filter.** Partial normalisation moves from $\mathcal{MY}$ back to $\text{Dist}Y$.
6. **Continuous observations.** Likelihoods or disintegration replace finite currying.
7. **Continuous time.** Stochastic integration and backward equations enter.

A particularly natural intermediate test is the finite-state continuous-time Wonham filter. Kim and Mehta’s stability work treats this setting by a duality method, while their general theory extends the observability/controllability correspondence to nonlinear HMMs [15, 16]. The Wonham case sits much closer to our finite operators than the fully general diffusion problem and should therefore precede any attempt to categorify the full BSDE construction.

The resulting Sweedler question is more precise than before. Can the finite pairing

$$\langle L_o \sigma, V \rangle = \langle \sigma, L_o^* V \rangle$$

be lifted through measure-valued determinisation, QBS enrichment and continuous time until the backward object becomes the dual control system of Kim–Mehta? That remains conjectural, but every stage now has a concrete mathematical target.

18 What QBS buys, and what it does not

We can now answer the question which motivated this Part.

	Residual probability in Set	Probability in QBS
finite stochastic runners	natural	natural
continuous distributions	external/encoded	native
higher-order function spaces	Set-functions only	QBS exponentials
distributions on functions	awkward as measurable semantics	native motivation of QBS
finite HMM determinisation	finite-measure algebra	same algebraic pattern
normalisation	partial at zero evidence	still partial
continuous observations	requires analytic machinery	QBS gives a suitable ambient probability world
disintegration/conditionals	extra structure	extra structure, studied categorically
continuous-time stochastic calculus	not supplied	not supplied by QBS alone

Moving to QBS therefore does not replace the finite residual story. It enlarges the class of state spaces, observations and higher-order parameters on which that story can be attempted. Nor does QBS make conditional probabilities or stochastic differential equations automatic. The advantage is that the relevant continuous probabilistic objects live in a cartesian closed ambient category where enriched algebra and higher-order semantics are available.

19 The relation with Part X: smoothness and probability

Part X supplied smooth families and infinitesimal geometry. Part XA supplies continuous probability and higher-order random structure. Nonlinear filtering ultimately needs both.

The tension can be displayed schematically:

smooth world : tangent vectors, ODEs, Lyapunov structure,
 QBS world : random elements, kernels, higher-order probability.

A continuous-time hidden Markov model with diffusion dynamics asks for stochastic differential geometry, and the Kim–Mehta dual system asks for still more: stochastic integration and backward stochastic differential equations.

There is nevertheless a suggestive structural clue. QBS can be presented via sheaf-like/concrete-sheaf structure built from standard Borel spaces, while diffeological spaces arise from plots on Cartesian spaces. This makes a site carrying both probabilistic probes and smooth plots a natural place to look for a joint ambient world. Work using diffeological semantics for differentiation in

probabilistic programming is adjacent to this idea. At present this is a design heuristic, not a completed Sweedler base.

Thus Part X and Part XA should not be seen as competing choices of base. They isolate two structures which a later analytic universe may need to combine. Before searching for that final universe, we can already test the Sweedler architecture against finite, QBS-valued and finite-state continuous-time probabilistic systems.

20 A staged research programme

The probabilistic branch of the tutorial now has a sequence of concrete tests.

20.1 Stage 1: finite residual probability

Prove explicitly that

$$T_{\text{Mealy}}(A, O) \rightsquigarrow \text{Dist}$$

runners are edge-emitting stochastic Mealy kernels

$$A \times Y \rightarrow \text{Dist}(O \times Y),$$

and record the role of a separate initial law.

20.2 Stage 2: finite determinisation and duality

Construct

$$\mathcal{M}Y \rightarrow (\mathcal{M}Y)^O$$

by the generalised powerset construction, prove the word-probability semantics, derive the partial normalised filter, and identify the adjoint operators L_o/L_o^* . Compare reachable and observable subspaces with Hankel-rank minimisation.

20.3 Stage 3: the QBS ambient theorem

Fix κ and verify the enriched duoidal and accessibility hypotheses for

$$\mathcal{D}_{\text{Qbs}, \kappa} = \text{Qbs}[\text{Qbs}, \text{Qbs}]_{\kappa}.$$

In particular, locate $\mathbf{P}$ inside a workable accessibility rank and establish the measuring objects needed for runner classification.

20.4 Stage 4: continuous observations

Replace finite currying by dominated likelihood transformers where possible, and by Markov-category/QBS disintegration where necessary. Determine how unnormalised measure evolution and partial normalisation are expressed in the runner/handler architecture.

20.5 Stage 5: probabilistic behaviour

Ask what the Part IX behaviour construction becomes for probabilistic residual runners. Does the runner-to-profunctor construction interact with weighted determinisation, posterior belief, or behavioural equivalence of the hidden and belief-state systems? Continuous probabilistic bisimulation will be relevant here.

20.6 Stage 6: the Wonham bridge

Before the full nonlinear theory, treat the finite-state continuous-time Wonham filter and its dual stability analysis. This is the smallest setting in which stochastic calculus enters while the hidden state remains algebraically finite.

20.7 Stage 7: continuous-time duality

Only then attempt the full Kim–Mehta system: nonlinear continuous-time HMMs, BSDEs, and observability–controllability/filtering–control duality.

21 What is established and what remains conjectural

It is useful to keep the status of the chapter explicit.

Status	Claim
established	QBS is cartesian closed, locally $\mathfrak{c}^+$ -presentable, and supports a strong affine commutative probability monad.
established	A stochastic Mealy kernel $A \times Y \rightarrow \text{Dist}(O \times Y)$ has the generator-level form of a residual Dist -runner; a pointed HMM additionally requires an initial law.
established	Finite filtering factors as total linear determinisation on finite measures followed by partial normalisation.
established	The finite forward operators L_o and delegated backward operators L_o^* satisfy the expectation pairing.
established	Generalised powerset/determinisation and weighted-automaton minimisation provide the finite coalgebraic comparison.
established	Markov categories with conditionals support abstract HMM filtering and smoothing.
structural check	the precise fixed- κ duoidal/Sweedler theorem for QBS-enriched accessible endofunctors, including accessibility of P .
open	a canonical Sweedler account of likelihood weighting, normalisation and disintegration for continuous observations.
open	the exact double-categorical status of the interface reversal and of hidden-versus-belief behavioural equivalence.
open	recovery of the Wonham and Kim–Mehta dualities from the Sweedler architecture.

This division is a feature rather than an embarrassment. It tells us where ordinary coalgebra, linear algebra and categorical probability already solve the small test cases, and where genuinely new Sweedler work begins.

22 Where Part XA leaves us

The probabilistic story is now substantially closer to the main line of the tutorial than it first appeared.

First, probabilistic systems do not require a new ambient category merely in order to enter **Run**. Residual probability already gives stochastic implementations in the Set-based theory. More precisely,

$$T_{\text{Mealy}}(1, O) \rightsquigarrow \text{Dist}$$

is the dynamics of an edge-emitting finite HMM; an initial law points the runner and selects an observable process. Moore observation is genuinely different because the reader equations force finite probabilistic reads to remain deterministic.

Second, finite filtering itself is much less mysterious than a generic appeal to conditioning suggests. The runner determines positive linear operators

$$L_o : MY \rightarrow MY,$$

whose generalised-powerset determinisation is total. Bayes filtering is obtained by normalising $L_o\pi$, and the backward handler action supplies the adjoints L_o^* . Thus the chapter already contains a concrete forward/backward reachability–observability pairing before QBS enters.

Third, QBS offers something additional rather than merely more of the same. It allows continuous probability, higher-order function spaces, and probability distributions on those spaces to coexist in the ambient category. This makes

$$\mathbf{Qbs}\text{-}[\mathbf{Qbs}, \mathbf{Qbs}]_\kappa$$

a serious candidate for a probabilistic Sweedler universe, subject to the fixed-rank enriched duoidal checks recorded above.

Fourth, the true analytic boundary is now visible. For non-discrete observations, the finite currying step fails: one needs likelihoods, unnormalised measure transformers, disintegration, or related conditional structure. These are the constructions which must be integrated with runners, handlers and behaviour.

Finally, the Kim–Mehta duality from Part IX now has intermediate targets: weighted determinisation, the L_o/L_o^* pairing, Hankel reachability/observability, the Wonham filter, and only then the full BSDE dual control system. The gap remains substantial, but it is now a staircase of testable constructions rather than a jump from stochastic runners directly to nonlinear filtering.

Working conclusion. Probability enters the Sweedler programme twice:

residual probability explains stochastic implementation,

while

a probabilistic ambient category makes continuous stochastic state
and higher-order probability native.

Between generation and Bayesian filtering lies an algebraically visible intermediate object: the unnormalised measure-valued observer. Its forward operators and the runner’s backward handler action already form the smallest probabilistic duality that the later Sweedler theory should explain.

References and orientation

References

- [1] C. Heunen, O. Kammar, S. Staton, and H. Yang, *A Convenient Category for Higher-Order Probability Theory*, LICS 2017.
- [2] C. Heunen, O. Kammar, S. Moss, A. Ścibior, S. Staton, M. Vákár, and H. Yang, *The Semantic Structure of Quasi-Borel Spaces: Algebra, Logic, and Recursion*, PPS 2018.
- [3] T. Uustalu, *Stateful Runners of Effectful Computations*, Electronic Notes in Theoretical Computer Science 319 (2015), 403–421.
- [4] D. Ahman and A. Bauer, *Runners in Action*, ESOP 2020.
- [5] A. Silva, F. Bonchi, M. M. Bonsangue, and J. J. M. M. Rutten, *Generalizing Determinization from Automata to Coalgebras*, Logical Methods in Computer Science 9(1), 2013.
- [6] B. Jacobs, A. Silva, and A. Sokolova, *Trace Semantics via Determinization*, Journal of Computer and System Sciences 81 (2015), 859–879.
- [7] F. Bonchi, M. M. Bonsangue, M. Boreale, J. J. M. M. Rutten, and A. Silva, *A Coalgebraic Perspective on Linear Weighted Automata*, Information and Computation 211 (2012), 77–105.
- [8] T. Fritz, *A Synthetic Approach to Markov Kernels, Conditional Independence and Theorems on Sufficient Statistics*, Advances in Mathematics 370 (2020), 107239.
- [9] K. Cho and B. Jacobs, *Disintegration and Bayesian Inversion via String Diagrams*, Mathematical Structures in Computer Science 29 (2019), 938–971.
- [10] T. Fritz, A. Klingler, D. McNeely, A. Shah-Mohammed, and Y. Wang, *Hidden Markov Models and the Bayes Filter in Categorical Probability*, arXiv:2401.14669, 2024.
- [11] A. Ścibior, O. Kammar, M. Vákár, S. Staton, H. Yang, Y. Cai, K. Ostermann, S. K. Moss, C. Heunen, and Z. Ghahramani, *Denotational Validation of Higher-Order Bayesian Inference*, Proceedings of the ACM on Programming Languages 2 (POPL), 2018.
- [12] D. Ahman, O. Kammar, and R. E. Møgelberg, *A Convenient Fibration for Dependently-Typed Probability Theory*, LICS 2026.
- [13] P. Panangaden, *Labelled Markov Processes*, Imperial College Press, 2009.
- [14] R. van Handel, *Observability and Nonlinear Filtering*, Probability Theory and Related Fields 145 (2009), 35–74.
- [15] J. W. Kim and P. G. Mehta, *A Dual Characterization of the Stability of the Wonham Filter*, Proceedings of the 60th IEEE Conference on Decision and Control, 2021, 1621–1628.
- [16] J. W. Kim and P. G. Mehta, *Duality for Nonlinear Filtering I: Observability*, IEEE Transactions on Automatic Control 69(2) (2024), 699–711.
- [17] J. W. Kim and P. G. Mehta, *Duality for Nonlinear Filtering II: Optimal Control*, IEEE Transactions on Automatic Control 69(2) (2024), 712–725.