Measure composition

DENSITY OF A SINGLE READING

$$m(y) = pf(y) + (1-p)g(y)$$

COMPLETE EXAMPLE

Two laws for five readings

```
1  %%%
2  A two-component mixture for five readings.
3  %%%
4  flatppl_compat = "0.1"
5  y_data = [-1.1, -0.7, 0.2, 1.4, 3.0]
6  n = lengthof(y_data)
7  p = elementof(unitinterval)
8  mu = elementof(reals)
A 10 M1 = Normal(mu, 0.7)
11 M2 = Gamma(shape = 2, rate = 1)
12 mix = superpose(weighted(p, M1),
13     weighted(1 - p, M2))
14 readings = iid(mix, n)
B 16 batch = superpose(weighted(p, iid(M1, n)),
17     weighted(1 - p, iid(M2, n)))
C 19 K = functionof(joint(y = readings))
20 K_batch = functionof(joint(y = batch))
21 obs = record(y = y_data)
22 L = likelihoodof(K, obs)
23 L_batch = likelihoodof(K_batch, obs)
D 25 prior = joint(p = Beta(2, 2),
26     mu = Normal(0, 2))
27 posterior = bayesupdate(L, prior)
29 point = record(p = 0.4, mu = -1.0)
30 log_readings = logdensityof(L, point)
31 log_batch = logdensityof(L_batch, point)
```

Keep parameters and predicted readings

`model = jointchain(prior, K)`
A draw contains `p`, `mu` and `y`. All five readings use that draw of the parameters. Under `K`, each reading still has its own component choice.

Keep only the predicted readings

`predictive = kchain(prior, K)`
A draw contains `y`. Its law integrates over `p` and `mu`. Use `K_batch` here to predict a batch with one shared component.

Integrate out the mixing weight

Under the Beta(2, 2) prior, both readings come from the Normal with probability $E[p^2] = \frac{3}{10}$ under `K` and $E[p] = \frac{1}{2}$ under `K_batch`. For either law, each marginal probability is $\frac{1}{2}$.

ANNOTATED LINES

A Choose a component for each reading 10–14

```
M1 = Normal(mu, 0.7)
M2 = Gamma(shape = 2, rate = 1)
mix = superpose(weighted(p, M1),
    weighted(1 - p, M2))
readings = iid(mix, n)
```

`mix` describes one observation. At fixed `p` and `mu`, `iid(mix, n)` repeats that law independently. Two readings in the same batch may come from different components.

B Choose one component for the batch 16–17

```
batch = superpose(weighted(p, iid(M1, n)),
    weighted(1 - p, iid(M2, n)))
```

Each `iid` now describes five observations from a single component. `superpose` mixes the two five-dimensional laws. The readings share the component choice.

C Bind the same five observations 19–23

```
K = functionof(joint(y = readings))
K_batch = functionof(joint(y = batch))
obs = record(y = y_data)
L = likelihoodof(K, obs)
L_batch = likelihoodof(K_batch, obs)
```

Both kernels take `p` and `mu` and return a law for the record field `y`. Binding the same `obs` to each kernel gives two likelihoods on the same parameter space.

D Assign the common prior 25–27

```
prior = joint(p = Beta(2, 2),
    mu = Normal(0, 2))
posterior = bayesupdate(L, prior)
```

`joint` gives independent priors to `p` and `mu`. The field names match both likelihoods. Substitute `L_batch` for `L` to obtain the posterior under a shared component.

Condition the joint model

`given_y = restrict(model, obs)`
Fix the `y` field of `model` at the observed readings. This gives the same unnormalized measure on `p` and `mu` as `bayesupdate(L, prior)`.

CONSEQUENCES OF THE COMPONENT CHOICE

Component pairs for two readings

Rows give the component of the first reading. Columns give the component of the second.

Per reading

	Normal	Gamma
Normal	0.16	0.24
Gamma	0.24	0.36

Per batch

	Normal	Gamma
Normal	0.40	0
Gamma	0	0.60

At $p = 0.4$, either reading has a 0.4 probability of a Normal origin before observation. With a fresh choice per reading, the components differ with probability $0.24 + 0.24 = 0.48$. A shared choice rules out both off-diagonal pairs.

The five-reading likelihoods

Let f_i and g_i be the Normal and Gamma densities at reading y_i. In both models p is the Normal weight.

$$L_{\text{reading}} = \prod_i (pf_i + (1-p)g_i)$$

$$L_{\text{batch}} = p\prod_i f_i + (1-p)\prod_i g_i$$

What one negative reading establishes

The first reading is -1.1. Its Gamma density is zero, so this reading must come from the Normal. At fixed $p = 0.4$, what is the probability that the next reading comes from the Gamma?

Choice	Probability
Per reading	$1 - p = 0.60$
Per batch	0

Score the two laws at the same point

Component choice	Log likelihood
Per reading	-8.93513
Per batch	-27.50312

These scores use `p = 0.4` and `mu = -1`. Set `y_data = [-1.1]` on line 5. With one reading, both log likelihoods become −1.48876.

Jessica Cox · bcox@mpp.mpg.de