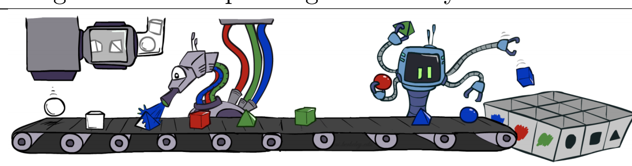
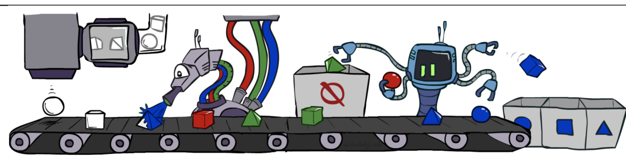
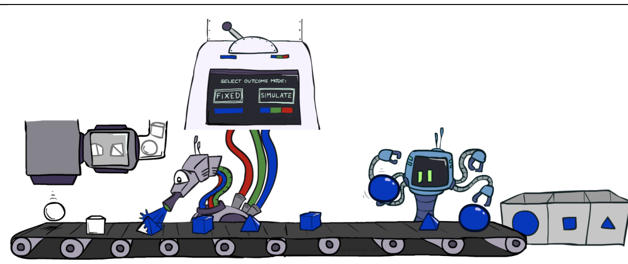
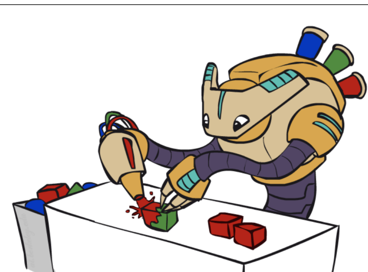
# 1 Candy Sampling

(a) In the year 2020, the Oompa Loompas at Charlie's Chocolate Factory have decided that they want to try a new automated way of sampling their candies for quality assurance. However, they have spilled chocolate sauce on their only copy of the user manual! Help out the Oompa Loompas by filling in the blanks below with the names of the four different types of sampling methods we've discussed in lecture, and then match each one to the corresponding image, probability distribution, and algorithm from the tables below.

| Sampling Method Name | Image and Distribution (A-D) | Algorithm (1-4): |
|---|---|---|
| | | |
| | | |
| | | |
| | | |

| Name: | Images and Corresponding Probability Distributions: |
|---|---|
| (A) |  **P(Q, E)** |
| (B) |  **P(Q \| e)** |
| (C) |  **P(Q, e)** |
| (D) |  **P(Q \| e)** |

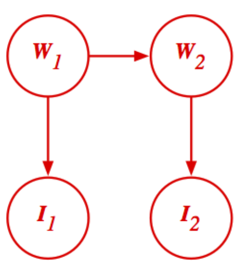
| Name: | Algorithms: |
|---|---|
| (1) | **function** ________________($X$, **e**, $bn$, $N$) **returns** an estimate of **P**($X$\|**e**)<br>**inputs**: $X$, the query variable<br>**e**, observed values for variables **E**<br>$bn$, a Bayesian network specifying joint distribution **P**($X_1, \ldots, X_n$)<br>$N$, the total number of samples to be generated<br>**local variables**: **W**, a vector of weighted counts for each value of $X$, initially zero<br><br>**for** $j$ = 1 to $N$ **do**<br>**x**, $w \leftarrow$ WEIGHTED-SAMPLE($bn$, **e**)<br>**W**[$x$] $\leftarrow$ **W**[$x$] $+ w$ where $x$ is the value of $X$ in **x**<br>**return** NORMALIZE(**W**)<br>---<br>**function** WEIGHTED-SAMPLE($bn$, **e**) **returns** an event and a weight<br><br>$w \leftarrow 1$; **x** $\leftarrow$ an event with $n$ elements initialized from **e**<br>**foreach** variable $X_i$ **in** $X_1, \ldots, X_n$ **do**<br>**if** $X_i$ is an evidence variable with value $x_i$ in **e**<br>**then** $w \leftarrow w \times P(X_i = x_i \mid parents(X_i))$<br>**else** **x**[$i$] $\leftarrow$ a random sample from **P**($X_i \mid parents(X_i)$)<br>**return** **x**, $w$ |
| (2) | **function** __________($bn$) **returns** an event sampled from the prior specified by $bn$<br>**inputs**: $bn$, a Bayesian network specifying joint distribution **P**($X_1, \ldots, X_n$)<br><br>**x** $\leftarrow$ an event with $n$ elements<br>**foreach** variable $X_i$ **in** $X_1, \ldots, X_n$ **do**<br>**x**[$i$] $\leftarrow$ a random sample from **P**($X_i \mid parents(X_i)$)<br>**return x** |
| (3) | **function** ________($X$, **e**, $bn$, $N$) **returns** an estimate of **P**($X$\|**e**)<br>**local variables**: **N**, a vector of counts for each value of $X$, initially zero<br>**Z**, the nonevidence variables in $bn$<br>**x**, the current state of the network, initially copied from **e**<br><br>initialize **x** with random values for the variables in **Z**<br>**for** $j$ = 1 to $N$ **do**<br>**for each** $Z_i$ in **Z do**<br>set the value of $Z_i$ in **x** by sampling from **P**($Z_i \| mb(Z_i)$)<br>**N**[$x$] $\leftarrow$ **N**[$x$] $+ 1$ where $x$ is the value of $X$ in **x**<br>**return** NORMALIZE(**N**) |
| (4) | **function** ________________($X$, **e**, $bn$, $N$) **returns** an estimate of **P**($X$\|**e**)<br>**inputs**: $X$, the query variable<br>**e**, observed values for variables **E**<br>$bn$, a Bayesian network<br>$N$, the total number of samples to be generated<br>**local variables**: **N**, a vector of counts for each value of $X$, initially zero<br><br>**for** $j$ = 1 to $N$ **do**<br>**x** $\leftarrow$ PRIOR-SAMPLE($bn$)<br>**if x** is consistent with **e then**<br>**N**[$x$] $\leftarrow$ **N**[$x$]+1 where $x$ is the value of $X$ in **x**<br>**return** NORMALIZE(**N**) |

# 2 Sampling Practice

(a) Compared to other sampling methods (rejection, likelihood weighting, Gibbs), what kind of information can prior sampling not use (that other methods can)?

(b) How does rejection sampling work on a high level, and what is its biggest/immediate weakness?

The diagram below describes a person's ice-cream eating habits based on the weather. The nodes $W_i$ stand for the weather on a day $i$, which can either be **s** (sunny) or **r** (rainy). The nodes $I_i$ represent whether the person ate ice-cream on day $i$, which can either be **t** (true) or **f** (false).




| $W_1$ | $P(W_1)$ |
|---|---|
| s | 0.6 |
| r | 0.4 |

| $I$ | $W$ | $P(I\|W)$ |
|---|---|---|
| t | s | 0.9 |
| f | s | 0.1 |
| t | r | 0.2 |
| f | r | 0.8 |

| $W_2$ | $W_1$ | $P(W_2\|W_1)$ |
|---|---|---|
| s | s | 0.7 |
| r | s | 0.3 |
| s | r | 0.5 |
| r | r | 0.5 |

Assume we generate the following six samples given the evidence $I_1 = t$ and $I_2 = f$ using **Likelihood Weighted Sampling**:

$$(W_1, I_1, W_2, I_2) = < s, t, r, f >, < r, t, r, f >, < s, t, r, f >, < s, t, s, f >, < s, t, s, f >, < r, t, s, f >$$

Using these samples, we will complete the following table:

| $(W_1, I_1, W_2, I_2)$ | Count/N | $w$ | Joint |
|---|---|---|---|
| s, t, s, f | 2/6 | 0.09 | 0.03 |
| s, t, r, f | /6 | | |
| r, t, s, f | /6 | | |
| r, t, r, f | /6 | | |

(c) What is the weight of the sample (s, t, r, f) above? Recall that the weight given to a sample in likelihood weighting is:

$$w = \prod_{\text{Evidence variables } e} P(e|\text{Parents}(e)).$$

(d) What is the estimate of P(s, t, r, f) given the samples?

(e) Compute the rest of the entries in the table. Use the estimated joint probabilities to estimate $P(W_2 = r|I_1 = t, I_2 = f)$.

(f) What is a weakness of likelihood weighing sampling? How does Gibbs sampling work, and how does it address this limitation?

# 3 HMMs: Tracking a Jabberwock

You have been put in charge of a Jabberwock for your friend Lewis. The Jabberwock is kept in a large tugley wood which is conveniently divided into an $N \times N$ grid. It wanders freely around the $N^2$ possible cells. At each time step $t = 1, 2, 3, \ldots$, the Jabberwock is in some cell $X_t \in \{1, \ldots, N\}^2$, and it moves to cell $X_{t+1}$ randomly as follows: with probability $1 - \epsilon$, it chooses one of the (up to 4) valid neighboring cells uniformly at random; with probability $\epsilon$, it uses its magical powers to teleport to a random cell uniformly at random among the $N^2$ possibilities (it might teleport to the same cell). Suppose $\epsilon = \frac{1}{2}$, $N = 10$ and that the Jabberwock always starts in $X_1 = (1, 1)$.

(a) Compute the probability that the Jabberwock will be in $X_2 = (2, 1)$ at time step 2. What about $P(X_2 = (4, 4))$?

At each time step $t$, you don't see $X_t$ but see $E_t$, which is the row that the Jabberwock is in; that is, if $X_t = (r, c)$, then $E_t = r$. You still know that $X_1 = (1, 1)$.

(b) Suppose we see that $E_1 = 1$, $E_2 = 2$. Fill in the following table with the distribution over $X_t$ after each time step, taking into consideration the evidence. Your answer should be concise. Hint: you should not need to do any heavy calculations.

| $t$ | $P(X_t \mid e_{1:t-1}, X_1 = (1, 1))$ | $P(X_t \mid e_{1:t}, X_1 = (1, 1))$ |
|---|---|---|
| 1 | see tables below | see tables below |
| 2 | see tables below | see tables below |

$t = 1$, $P(X_t \mid e_{1:t-1}, X_1 = (1, 1))$:

| $X_1$ | $P(X_1 \mid X_1 = (1,1))$ |
|---|---|
| (1, 1) | |
| all other values | |

$t = 1$, $P(X_t \mid e_{1:t}, X_1 = (1, 1))$:

| $X_1$ | $P(X_1 \mid e_1, X_1 = (1,1))$ |
|---|---|
| (1, 1) | |
| all other values | |

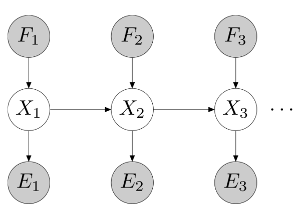
$t = 2$, $P(X_t \mid e_{1:t-1}, X_1 = (1, 1))$:

| $X_2$ | $P(X_2 \mid e_1, X_1 = (1,1))$ |
|---|---|
| (1, 2) | |
| (2, 1) | |
| all other values | |

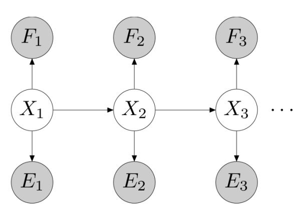
$t = 2$, $P(X_t \mid e_{1:t}, X_1 = (1, 1))$:

| $X_2$ | $P(X_2 \mid e_{1:2}, X_1 = (1,1))$ |
|---|---|
| (2, 1) | |
| $(2, a)$ $(\forall a, a > 1)$ | |
| all other values | |

You are a bit unsatisfied that you can't pinpoint the Jabberwock exactly. But then you remembered Lewis told you that the Jabberwock teleports only because it is frumious on that time step, and it becomes frumious independently of anything else. Let us introduce a variable $F_t \in \{0, 1\}$ to denote whether it will teleport at time $t$. We want to to add these frumious variables to the HMM. Consider the two candidates:




(A)




(B)

| (A) | (B) |
|---|---|
| $X_1 \perp X_3 \mid X_2$ | $X_1 \perp X_3 \mid X_2$ |
| $X_1 \perp E_2 \mid X_2$ | $X_1 \perp E_2 \mid X_2$ |
| $X_1 \perp F_2 \mid X_2$ | $X_1 \perp F_2 \mid X_2$ |
| $X_1 \perp E_4 \mid X_2$ | $X_1 \perp E_4 \mid X_2$ |
| $X_1 \perp F_4 \mid X_2$ | $X_1 \perp F_4 \mid X_2$ |
| $E_3 \perp F_3 \mid X_3$ | $E_3 \perp F_3 \mid X_3$ |
| $E_1 \perp F_2 \mid X_2$ | $E_1 \perp F_2 \mid X_2$ |
| $E_1 \perp F_2 \mid E_2$ | $E_1 \perp F_2 \mid E_2$ |

(c) For each model, circle the conditional independence assumptions above which are true in that model.

(d) Which Bayes net is more appropriate for the problem domain here, (A) or (B)? Justify your answer.

For the following questions, your answers should be fully general for models of the structure shown above, not specific to the teleporting Jabberwock.

(e) For (A), express $P(X_{t+1}, e_{1:t+1}, f_{1:t+1})$ in terms of $P(X_t, e_{1:t}, f_{1:t})$ and the conditional probability tables used to define the network. Assume the $E$ and $F$ nodes are all observed.

(f) For (B), express $P(X_{t+1}, e_{1:t+1}, f_{1:t+1})$ in terms of $P(X_t, e_{1:t}, f_{1:t})$ and the CPTs used to define the network. Assume the $E$ and $F$ nodes are all observed.

Suppose that we don't actually observe the $F_t$s.

(g) For (A), express $P(X_{t+1}, e_{1:t+1})$ in terms of $P(X_t, e_{1:t})$ and the CPTs used to define the network.

(h) For (B), express $P(X_{t+1}, e_{1:t+1})$ in terms of $P(X_t, e_{1:t})$ and the CPTs used to define the network.