In this notebook, we discuss two standard techniques for constructing joint distributions from simpler building blocks: (1) marginalizing out an unobserved random variable and (2) imposing conditional independence relations. Combining them produces a large class of models known as probabilistic graphical models. As before, we make our derivations in the finite support case, but these can be adapted to continuous cases.

Our starting point for both techniques will be the product rule: if $X = (X_1, \ldots, X_r)$ is a discrete random vector, then for any $x = (x_1, \ldots, x_r) \in S_X$

$$\mathbb{P}[X = x] = \mathbb{P}\left[\bigcap_{i=1}^r \{X_i = x_i\}\right] = \prod_{i=1}^r \mathbb{P}\left[X_i = x_i \big| \bigcap_{j=1}^{i-1} \{X_j = x_j\}\right].$$

Hence, we can specify a joint distribution through a chain marginal and conditional distributions. Note that we can re-order the variables $X_1, \ldots, X_r$ arbitrarily, giving rise to different representations.

### 2.1 Marginalizing out an unobserved random variable
Mixtures are a natural way to define probability distributions. The basic idea is to consider a pair of random vectors \((\mathbf{X}, \mathbf{Y})\) and assume that \(\mathbf{Y}\) is unobserved. The effect on the observed vector \(\mathbf{X}\) is that \(\mathbf{Y}\) is marginalized out. Indeed, by the law of total probability, for any \(\mathbf{x} \in S_X\)
\[
p_X(\mathbf{x}) = \mathbb{P}[\mathbf{X} = \mathbf{x}] = \sum_{y \in S_Y} \mathbb{P}[\mathbf{X} = \mathbf{x}|\mathbf{Y} = \mathbf{y}] \mathbb{P}[\mathbf{Y} = \mathbf{y}] = \sum_{y \in S_Y} p_{X|Y}(\mathbf{x}|\mathbf{y}) p_Y(\mathbf{y})
\]
where we used that the events \(\{\mathbf{Y} = \mathbf{y}\}, \mathbf{y} \in S_Y\), form a partition of the probability space. We interpret this equation as defining \(p_X(\mathbf{x})\) as a convex combination - a mixture - of the distributions \(p_{X|Y}(\mathbf{x}|\mathbf{y}), \mathbf{y} \in S_Y\), with mixing weights \(p_Y(\mathbf{y})\). In general, we need to specify the full conditional probability distribution (CPD):
\[
p_{X|Y}(\mathbf{x}|\mathbf{y}), \forall \mathbf{x} \in S_X, \mathbf{y} \in S_Y.
\]
But assuming that the mixing weights and/or CPD come from parametric families can help reduce the complexity of the model.

In particular, in the parametric context, this gives rise to a powerful approach to expanding distribution families. Suppose \(\{p_\theta : \theta \in \Theta\}\) is a parametric family of distributions. Let \(K \geq 2, \theta_1, \ldots, \theta_K \in \Theta\) and \(\pi = (\pi_1, \ldots, \pi_K) \in \Delta_K\). Suppose \(Y \sim \text{Cat}(\pi)\) and that the conditional distributions satisfy
\[
p_{X|Y}(\mathbf{x}|i) = p_{\theta_i}(\mathbf{x}).
\]
We write this as \(X|\{Y = i\} \sim p_{\theta_i}\). Then we obtain the mixture model
\[
p_X(\mathbf{x}) = \sum_{i=1}^{K} p_{X|Y}(\mathbf{x}|i) p_Y(i) = \sum_{i=1}^{K} \pi_i p_{\theta_i}(\mathbf{x}).
\]

**Example (Mixture of Multinomials):** Let \(n, m, K \geq 1, \pi \in \Delta_K\) and, for \(i = 1, \ldots, K\), \(p_i = (p_{i1}, \ldots, p_{im}) \in \Delta_m\). Suppose that \(Y \sim \text{Cat}(\pi)\) and that the conditional distributions are \(X|\{Y = i\} \sim \text{Mult}(n, p_i)\).

Then \(X\) is a mixture of multinomials. Its distribution is then
\[
p_X(\mathbf{x}) = \sum_{i=1}^{K} \pi_i \frac{n!}{x_1! \cdots x_m!} \prod_{j=1}^{m} p_{ij}^{x_j}.
\]
Exercise (Moments of a Mixture): For \( i = 1, \ldots, K \), let \( p_i \) be a probability mass function over the set \( S_i \subseteq \mathbb{R} \) with mean \( \mu_i \) and variance \( \sigma_i^2 \). Let \( \pi = (\pi_1, \ldots, \pi_K) \in \Delta_K \). Suppose \( X \) is drawn from the mixture distribution
\[
p_X(x) = \sum_{i=1}^{K} \pi_i p_i(x).
\]

Establish the following formulas:

(a) \( \mathbb{E}[X] = \sum_{i=1}^{K} \pi_i \mu_i \)

(b) \( \text{Var}[X] = \sum_{i=1}^{K} \pi_i (\sigma_i^2 + \mu_i^2 - \mu^2) \).

Next is an important continuous example.

Example (Gaussian mixture model): For \( i = 1, \ldots, K \), let \( \mu_i \) and \( \Sigma_i \) be the mean and covariance matrix of a multivariate normal distribution. Let \( \pi \in \Delta_K \). A Gaussian mixture model is obtained as follows: take \( Y \sim \text{Cat}(\pi) \) and
\[
X|\{Y = i\} \sim \mathcal{N}(\mu_i, \Sigma_i).
\]

It has a probability density function of the form
\[
f_X(x) = \sum_{i=1}^{K} \pi_i \frac{1}{(2\pi)^{d/2} |\Sigma_i|^{1/2}} \exp\left(-\frac{1}{2}(x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i)\right).
\]

This is illustrated below.
We could also consider continuous mixing distributions. Several standard approaches, e.g. probabilistic PCA, factor analysis and independent components analysis, use such mixtures. See [Bis](https://www.microsoft.com/en-us/research/uploads/prod/2006/01/Bishop-Pattern-Recognition-and-Machine-Learning-2006.pdf), Chapter 12].
In Julia, the package Distributions.jl (https://juliastats.org/Distributions.jl/stable/) also provides a way to sample from mixture models (https://juliastats.org/Distributions.jl/stable/mixture/). This is better seen on an example.

```
In [1]: # Julia version: 1.5.1
    using Random, Distributions

In [2]: Random.seed!(123); # Setting the seed

In [3]: # constructs a mixture of three normal distributions, # with prior probabilities [0.2, 0.5, 0.3]
d = MixtureModel(Normal[
    Normal(-2.0, 1.2),
    Normal(0.0, 1.0),
    Normal(3.0, 2.5)],
    [0.2, 0.5, 0.3])

Out[3]: MixtureModel{Normal}(K = 3)
    components[1] (prior = 0.2000): Normal{Float64}(μ=-2.0, σ=1.2)
    components[2] (prior = 0.5000): Normal{Float64}(μ=0.0, σ=1.0)
    components[3] (prior = 0.3000): Normal{Float64}(μ=3.0, σ=2.5)

In [4]: N = 5 # number of samples
    rand(d,N)

Out[4]: 5-element Array{Float64,1}:
    3.0077577268073448
    0.30409716708662865
    0.6175380099263851
    -1.6227093763731233
    -0.8356308611584168
```

### 2.2 Using conditional independence relations

Another powerful approach is the use of conditional independence.

#### 2.2.1 Definitions
We begin with the definition. First an exercise: in words, the conditional probabilities satisfy the product rule.

Exercise: Let $A$, $B$ and $C$ be events. Use the product rule to show that
\[ P[A \cap B | C] = P[A | B \cap C] P[B | C]. \]

Definition (Conditional Independence): Let $A$, $B$, $C$ be events such that $P[C] > 0$. Then $A$ and $B$ are conditionally independent given $C$, denoted $A \perp B | C$, if
\[ P[A \cap B | C] = P[A | C] P[B | C]. \]


$A$ and $B$ are conditionally independent given $C$ if and only if, given knowledge that $C$ occurs, knowledge of whether $A$ occurs provides no information on the likelihood of $B$ occurring, and knowledge of whether $B$ occurs provides no information on the likelihood of $A$ occurring.

In general, conditionally independent events are not (unconditionally) independent.

Example: Imagine I have two six-sided dice. Dice $a$ has faces $\{1, 3, 5, 7, 9, 11\}$ and dice $b$ has faces $\{2, 4, 6, 8, 10, 12\}$. Suppose I perform the following experiment: I pick one of the two dice uniformly at random, and then I roll that die twice. Let $X_1$ and $X_2$ be the outcomes of the rolls. Consider the events $A = \{X_1 = 1\}$, $B = \{X_2 = 2\}$, and $C = \{\text{die } a \text{ is picked}\}$. The events $A$ and $B$ are clearly dependent: if $A$ occurs, then I know that die $a$ was picked, and hence $B$ cannot occur. Knowledge of one event provides information about the likelihood of the other event occurring. Formally, by the law of total probability,
\[ P[A] = P[A | C] P[C] + P[A | C^c] P[C^c] = \frac{1}{6} \frac{1}{2} + 0 \frac{1}{2} = \frac{1}{12}. \]

Similarly $P[B] = \frac{1}{12}$. Yet $P[A \cap B] = 0 \neq \frac{1}{12} \frac{1}{12}$.

On the other hand, we claim that $A$ and $B$ are conditionally independent given $C$. Again this is intuitively clear: once I pick a die, the two rolls are independent. For a given die choice, knowledge of one roll provides no information about the likelihood of the other roll. Note that the phrase "for a given die choice" is critical in the last statement. Formally, by our experiment, we have $P[A | C] = 1/6$, $P[B | C] = 0$ and $P[A \cap B | C] = 0$. So indeed
\[ P[A \cap B | C] = P[A | C] P[B | C] \]

as claimed. \(\triangleleft\)
The following exercises are further instances of the important principle that conditional probabilities satisfy the same rules as probabilities.

**Exercise:** Let $A, B, C$ be events such that $\mathbb{P}[C] > 0$ and $A \perp B|C$. Show that $A \perp B^c|C$. \(\triangleleft\)

**Exercise:** Let $A, B, C$ be events such that $\mathbb{P}[B \cap C], \mathbb{P}[A \cap C] > 0$. Show that $A \perp B|C$ if and only if $\mathbb{P}[A|B \cap C] = \mathbb{P}[A|C]$ and $\mathbb{P}[B|A \cap C] = \mathbb{P}[B|C]$. \(\triangleleft\)

**Exercise (Adding Data):** Let $A, B, C$ be events such that $\mathbb{P}[B \cap C] > 0$.

(a) First show that
\[
\mathbb{P}[A|B \cap C] = \frac{\mathbb{P}[C|A \cap B] \mathbb{P}[A|B]}{\mathbb{P}[C|A \cap B] \mathbb{P}[A|B] + \mathbb{P}[C|A^c \cap B] \mathbb{P}[A^c|B]}.
\]

(b) Now suppose $B \perp C|A$. Show that
\[
\mathbb{P}[A|B \cap C] = \frac{\mathbb{P}[C|A] \mathbb{P}[A|B]}{\mathbb{P}[C|A] \mathbb{P}[A|B] + \mathbb{P}[C|A^c] \mathbb{P}[A^c|B]}.
\]
\(\triangleleft\)

Conditional independence is naturally extended to random vectors.

**Definition (Conditional Independence of Random Vectors):** Let $X, Y, W$ be discrete random vectors. Then $X$ and $Y$ are said to be conditionally independent given $W$, denoted $X \perp Y|W$ if for all $x \in S_X, y \in S_Y$ and $w \in S_W$
\[
\mathbb{P}[X = x, Y = y|W = w] = \mathbb{P}[X = x|W = w] \mathbb{P}[Y = y|W = w].
\]
\(\triangleleft\)

An important consequence is that we can drop the conditioning by the independent variable.

**Lemma (Role of Independence):** Let $X, Y, W$ be discrete random vectors such that $X \perp Y|W$. For all $x \in S_X, y \in S_Y$ and $w \in S_W$
\[
\mathbb{P}[X = x|Y = y, W = w] = \mathbb{P}[X = x|W = w].
\]

**Proof:** In a previous exercise, we showed that $A \perp B|C$ implies $\mathbb{P}[A|B \cap C] = \mathbb{P}[A|C]$. That implies the claim. \(\square\)

Here are a couple exercises to get used to the definition.
Exercise: Let \( \mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{W} \) be discrete random vectors. Show that \( \mathbf{X} \perp (\mathbf{Y}, \mathbf{Z}) | \mathbf{W} \) implies that \( \mathbf{X} \perp \mathbf{Y} | \mathbf{W} \) and \( \mathbf{X} \perp \mathbf{Z} | \mathbf{W} \). \(<\)

Exercise: Let \( \mathbf{X}, \mathbf{Y}, \mathbf{Z} \) be discrete random vectors. Suppose that \( \mathbf{X} \perp \mathbf{Y} | \mathbf{Z} \) and \( \mathbf{X} \perp \mathbf{Z} \). Show that \( \mathbf{X} \perp (\mathbf{Y}, \mathbf{Z}) \). \(<\)

2.2.2 The basic configurations

The case of three random variables exemplifies key probabilistic relationships. By the product rule, we can write

\[
P[X = x, Y = y, Z = z] = P[X = x] P[Y = y | X = x] P[Z = z | X = x, Y = y].
\]

This is conveniently represented through a digraph where the vertices are the variables. Recall that an arrow \((i, j)\), from \(i\) to \(j\), indicates that \(i\) is a parent of \(j\) and that \(j\) is a child of \(i\). Let \(\text{pa}(i)\) be the set of parents of \(i\).

The digraph \(G = (V, E)\) below encodes the following sampling scheme, referred as ancestral sampling:

1. First we pick \(X\) according to its marginal \(P[X = x]\). Note that \(X\) has no parent in \(G\).
2. Second we pick \(Y\) according to the CPD \(P[Y = y | X = x]\). Note that \(X\) is the only parent of \(Y\).
3. Finally we pick \(Z\) according to the CPD \(P[Z = z | X = x, Y = y]\). Note that the parents of \(Z\) are \(X\) and \(Y\).

Note that the graph above is acyclic, that is, it has no directed cycle. In particular, the ordering of variables \(X, Y, Z\) is a topological ordering (https://en.wikipedia.org/wiki/Topological_sorting), that is, an ordering of the vertices where all edges \((i, j)\) are such that \(i\) comes before \(j\).

The same joint distribution can be represented by a different digraph if the product rule is used in a different order. For instance,

\[
P[X = x, Y = y, Z = z] = P[Z = z] P[Y = y | Z = z] P[X = x | Z = z, Y = y]
\]

is represented by the following digraph.
The fork: Removing edges in the first graph above encodes conditional independence relations. For instance, removing the edge from $Y$ to $Z$ gives the following graph, known as a fork. We denote this configuration as $Y \leftarrow X \rightarrow Z$.

![Fork Graph](image)

The joint distribution simplifies as follows:

$$P[X = x, Y = y, Z = z] = P[X = x] P[Y = y|X = x] P[Z = z|X = x].$$

So, in this case, what has changed is that the CPD of $Z$ does not depend on the value of $Y$. That is, $Z \perp Y|X$. Indeed, we can check that claim directly

$$P[Y = y, Z = z|X = x] = \frac{P[X = x, Y = y, Z = z]}{P[X = x]} = \frac{P[X = x] P[Y = y|X = x] P[Z = z|X = x]}{P[X = x]} = P[Y = y|X = x] P[Z = z|X = x]$$

as claimed.

The chain: Removing the edge from $X$ to $Z$ gives the following graph, known as a chain (or pipe). We denote this configuration as $X \rightarrow Y \rightarrow Z$.

![Chain Graph](image)
The joint distribution simplifies as follows:
\[ \mathbb{P}[X = x, Y = y, Z = z] = \mathbb{P}[X = x] \mathbb{P}[Y = y|X = x] \mathbb{P}[Z = z|Y = y]. \]

In this case, what has changed is that the CPD of \( Z \) does not depend on the value of \( X \). Compare that to the fork. The corresponding conditional independence relation is \( Z \perp X|Y \). Indeed, we can check that claim directly
\[
\mathbb{P}[X = x, Z = z|Y = y] = \frac{\mathbb{P}[X = x, Y = y, Z = z]}{\mathbb{P}[Y = y]}
= \frac{\mathbb{P}[X = x] \mathbb{P}[Y = y|X = x] \mathbb{P}[Z = z|Y = y]}{\mathbb{P}[Y = y]}
\]

Now we have to use Bayes' rule to get
\[
= \frac{\mathbb{P}[X = x] \mathbb{P}[Y = y|X = x] \mathbb{P}[Z = z|Y = y]}{\mathbb{P}[Y = y]}
= \frac{\mathbb{P}[Y = y|X = x] \mathbb{P}[X = x]}{\mathbb{P}[Y = y]} \mathbb{P}[Z = z|Y = y]
= \mathbb{P}[X = x|Y = y] \mathbb{P}[Z = z|Y = y]
\]
as claimed.

**The collider:** Removing the edge from \( X \) to \( Y \) gives the following graph, known as a collider. We denote this configuration as \( X \rightarrow Z \leftarrow Y \).
The joint distribution simplifies as follows:
\[
P[X = x, Y = y, Z = z] = P[X = x] P[Y = y] P[Z = z | X = x, Y = y].
\]
In this case, what has changed is that the CPD of \( Y \) does not depend on the value of \( X \). Compare that to the fork and the chain. This time we have \( X \perp \perp Y \). Indeed, we can check that claim directly
\[
P[X = x, Y = y] = \sum_{z \in S_z} P[X = x, Y = y, Z = z]
\]
\[
= \sum_{z \in S_z} P[X = x] P[Y = y] P[Z = z | X = x, Y = y]
\]
\[
= P[X = x] P[Y = y]
\]
as claimed.

Perhaps counter-intuitively, conditioning on \( Z \) makes \( X \) and \( Y \) dependent in general. This is known as explaining away or Berkson’s Paradox. Quoting my colleague Jordan Ellenberg (https://en.wikipedia.org/wiki/Berkson%27s_paradox#Ellenberg_example):

Suppose Alex will only date a man if his niceness plus his handsomeness exceeds some threshold. Then nicer men do not have to be as handsome to qualify for Alex’s dating pool. So, among the men that Alex dates, Alex may observe that the nicer ones are less handsome on average (and vice versa), even if these traits are uncorrelated in the general population. Note that this does not mean that men in the dating pool compare unfavorably with men in the population. On the contrary, Alex’s selection criterion means that Alex has high standards. The average nice man that Alex dates is actually more handsome than the average man in the population (since even among nice men, the ugliest portion of the population is skipped). Berkson’s negative correlation is an effect that arises within the dating pool: the rude men that Alex dates must have been even more handsome to qualify.

### 2.2.3 Some key examples

We end this subsection with two key examples.

**Example (Naive Bayes):** This model is useful for document classification. We assume that a document has a single topic \( C \) from a list \( C = \{c_1, \ldots, c_K\} \) with probability distribution \( \pi_k = P[C = c_k] \). There is a vocabulary of size \( M \) and we record the presence or absence of a word \( m \) in the document with a Bernoulli variable \( X_m \), where \( p_{km} = P[X_m = 1 | C = c_k] \). We denote by \( \mathbf{X} = (X_1, \ldots, X_M) \) the corresponding vector.
The conditional independence assumption comes next: we assume that, given a topic $C$, the word occurrences are independent. That is,

$$
P[X = x | C = c_k] = \prod_{m=1}^{M} P[X_m = x_m | C = c_k]
$$

$$
= \prod_{m=1}^{M} \pi_m^x (1 - p_{km})^{1-x_m}.
$$

Finally, the joint distribution is

$$
P[C = c_k, X = x] = P[X = x | C = c_k] P[C = c_k]
$$

$$
= \pi_k \prod_{m=1}^{M} \pi_m^x (1 - p_{km})^{1-x_m}.
$$

Example (Markov Chain): Markov chains (https://en.wikipedia.org/wiki/Markov_chain) are widely used to model sequential data. We restrict ourselves here to discrete-time, time-homogeneous, finite-space Markov chains. But the conditional independence is at the heart of Markov chains more generally. Let $X_0, X_1, \ldots, X_T$ be a sequence of random vectors taking values in a common space $S$, where the index represents time. In words the basic assumption is the following: we assume that the past and future are conditionally independent given the present. Formally, for all $1 \leq t \leq T - 1$,

$$(X_0, \ldots, X_{t-1}) \perp (X_{t+1}, \ldots, X_T) | X_t.$$  

The implication of these assumptions is that the joint distribution (also known referred to as the distribution of a sample path) is

$$
P[X_0 = x_0, X_1 = x_1, \ldots, X_T = x_T]
$$

$$
= P[X_0 = x_0] \prod_{t=1}^{T} P[X_t = x_t | X_{t-1} = x_{t-1}, \ldots, X_0 = x_0]
$$

$$
= P[X_0 = x_0] \prod_{t=1}^{T} P[X_t = x_t | X_{t-1} = x_{t-1}]
$$

where we used that, by a previous exercise, $(X_0, \ldots, X_{t-2}) \perp (X_t, \ldots, X_T) | X_{t-1}$ implies $(X_0, \ldots, X_{t-2}) \perp X_t | X_{t-1}$. That in turn implies the above simplification by the Role of Independence.
Now we assume that the common state space \( S \) is finite, say \( \{1, \ldots, m\} \) for concreteness. We also assume the process is time-homogeneous, which means that

\[
\mathbb{P}[X_t = x' \mid X_{t-1} = x'] = \mathbb{P}[X_1 = x' \mid X_0 = x'] =: p_{x',x}, \quad \forall t = 1, \ldots, T.
\]

We can think of the transition probabilities \( P = (p_{x',x})_{x,x' \in S} \) as an \( m \times m \) nonnegative matrix. In fact, \( P \) is a stochastic matrix, that is, its rows necessarily sum to one. Indeed,

\[
\sum_{x \in S} p_{x,x'} = \sum_{x \in S} \mathbb{P}[X_1 = x \mid X_0 = x'] = \mathbb{P}[X_1 \in S \mid X_0 = x'] = 1
\]

by the properties of the conditional probability. We also let \( \mu_x = \mathbb{P}[X_0 = x] \) and we think of \( \mu = (\mu_x)_{x \in S} \) as a row vector.

With this notation, the joint distribution becomes simply

\[
\mathbb{P}[X_0 = x_0, X_1 = x_1, \ldots, X_T = x_T] = \mu_{x_0} \prod_{t=1}^{T} p_{x_{t-1}, x_t}.
\]

This formula has a remarkable consequence. The marginal distribution of \( x_s \) is a matrix power of \( P \). Indeed, for any \( 1 \leq s \leq T \),

\[
\mathbb{P}[X_s = x_s] = \sum_{x_0, \ldots, x_{s-1} \in S} \mu_{x_0} \prod_{t=1}^{T} p_{x_{t-1}, x_t}
= \sum_{x_0, \ldots, x_{s-1} \in S} \mu_{x_0} \prod_{t=1}^{s} p_{x_{t-1}, x_t} \sum_{x_{s+1}, \ldots, x_T \in S} \prod_{t=s+1}^{T} p_{x_{t-1}, x_t}.
\]

The most interior sum simplifies to

\[
\sum_{x_{s+1}, \ldots, x_T \in S} \prod_{t=s+1}^{T} p_{x_{t-1}, x_t} = \sum_{x_{s+1}, \ldots, x_T \in S} \prod_{t=s+1}^{T} p_{x_{t-1}, x_t} \sum_{x_T \in S} p_{x_{T-1}, x_T}
= \sum_{x_{s+1}, \ldots, x_{T-1} \in S} \prod_{t=s+1}^{T-1} p_{x_{t-1}, x_t}
= \cdots
= 1
\]

where we repeatedly used the fact that the rows of \( P \) sum to 1.
The remaining sum can be simplified in a similar fashion

\[
\sum_{x_0, \ldots, x_{s-1} \in S} \mu_{x_0} \prod_{t=1}^{s} p_{x_{t-1}, x_t} = \sum_{x_0 \in S} \mu_{x_0} \sum_{x_1 \in S} p_{x_0, x_1} \cdots \sum_{x_{t-2} \in S} p_{x_{t-3}, x_{t-2}} \sum_{x_{t-1} \in S} p_{x_{t-2}, x_{t-1}} p_{x_{t-1}, x_t} = \sum_{x_0 \in S} \mu_{x_0} \sum_{x_1 \in S} p_{x_0, x_1} \cdots \sum_{x_{t-2} \in S} p_{x_{t-3}, x_{t-2}} (P^2)_{x_{t-2}, x_t} = \cdots = (\mu P^s)_{x_s}.
\]

Thus we have established that, for any \(1 \leq s \leq T\),

\[
P[X_s = x_s] = (\mu P^s)_{x_s}.
\]

This type of recursive simplification will play a major role in the next section. \(\triangleq\)

The time homogeneity in the model above is an example of parameter sharing, a standard way to reduce the complexity of a model.