Review of Probability: The Joint Distribution

When there are multiple random variables, the fundamental mathematical object is the joint probability distribution:

	Starts		¬Starts	
	Gas	¬Gas	Gas	¬Gas
GasGauge	.5040	.0006	.2160	.0594
¬GasGauge	.0560	.0014	.0240	.1386

We will use three notations:

- **Single Cell:** $P(\mathsf{Gas}, -\mathsf{GasGauge}, \mathsf{Starts}) = P(\mathsf{Gas} \wedge \neg\mathsf{GasGauge} \wedge \mathsf{Starts}) = .0560.$
- **Whole Table:** $\mathbf{P}(\mathsf{Gas}, \mathsf{GasGauge}, \mathsf{Starts})$ stands for the entire joint distribution.
- **Summation Shorthand:** $\sum_{\mathsf{Gas}} P(\mathsf{Gas}, \mathsf{GasGauge}, \mathsf{Starts}) = P(\mathsf{Gas}, \mathsf{GasGauge}, \mathsf{Starts}) + P(-\mathsf{Gas}, \mathsf{GasGauge}, \mathsf{Starts})$

  For the distribution, we write this as $\sum_{\mathsf{Gas}} \mathbf{P}(\mathsf{Gas}, \mathsf{GasGauge}, \mathsf{Starts}) = \mathbf{P}(\mathsf{GasGauge}, \mathsf{Starts})$

Marginal Probabilities

Suppose we want to know the probability that $\mathsf{GasGauge}$ is true. We obtain this by summing over the entire top row of the table.

$$P(\mathsf{GasGauge}) = .5040 + .0006 + .2160 + .0594 = 0.78$$

Another example: the probability that Gas is false:

$$P(-\mathsf{Gas}) = .0006 + .0014 + .0594 + .1386 = 0.2$$

This is called *marginalizing*. We write it as

$$P(-\mathsf{Gas}) = \sum_{GasGauge, Starts} P(-\mathsf{Gas}, \mathsf{GasGauge}, \mathsf{Starts})$$

$\mathbf{P}(\mathsf{Gas})$ is called the *marginal distribution* of Gas.

Conditional Probability: Belief Updating

Suppose we want to know the probability that the car will start given that we have observed $\mathsf{GasGauge}$ and found it to be true.

$$P(\mathsf{Starts}|\mathsf{GasGauge}) = \frac{P(\mathsf{Starts, GasGauge})}{P(\mathsf{GasGauge})} = \frac{.5040 + .0006}{.5040 + .0006 + .2160 + .0594} = .6469$$

	Starts		–Starts	
	Gas	–Gas	Gas	–Gas
GasGauge	.5040	.0006	.2160	.0594
–GasGauge	.0560	.0014	.0240	.1386

Since $\mathsf{GasGauge}$ is true, one of the four situations from the top row of the table must hold. We *renormalize* the probabilities and then marginalize to compute $\mathbf{P}(\mathsf{Starts})$.

Every time the agent learns the value of some random variable, it can update the joint distribution to include only the remaining relevant variables.

Conditional Probability Distribution

The conditional probability distribution $\mathbf{P}(\mathsf{GasGauge}|\mathsf{Gas})$:

	GasGauge	¬GasGauge
Gas	.9	.1
¬Gas	.3	.7

Each row must sum to 1.

Chain Rule of Probability

$$P(A,B,C) = P(A|B,C) \cdot P(B|C) \cdot P(C)$$

for any number of variables in any order.

Proof:

$$\begin{aligned}
P(A,B,C) &= \frac{P(A,B,C)}{P(B,C)} \cdot P(B,C) \\
&= P(A|B,C) \cdot P(B,C) \\
&= P(A|B,C) \cdot \frac{P(B,C)}{P(C)} \cdot P(C) \\
&= P(A|B,C) \cdot P(B|C) \cdot P(C)
\end{aligned}$$

This holds for distributions as well:

$$\mathbf{P}(A,B,C) = \mathbf{P}(A|B,C) \cdot \mathbf{P}(B|C) \cdot \mathbf{P}(C)$$

Belief Networks (1): Independence

Two random variables X and Y are *independent* if $P(X, Y) = P(X) \cdot P(Y)$ for all values of X and Y. This can be written as $\mathbf{P}(X, Y) = \mathbf{P}(X) \cdot \mathbf{P}(Y)$. Intuitively, there is no causal connection betwen X and Y.

Example:

- **X**: coin with probability $X = true$ of .4, and $X = false$ of .6.
- **Y**: coin with probability $Y = true$ of .8, and $Y = false$ of .2.

Joint distribution:

	$X = true$	$X = false$
$Y = true$	.32	.48
$Y = false$	.08	.12

Belief Networks (2): Conditional Independence

Two random variables X and Y are *conditionally independent given* Z if $P(X, Y|Z) = P(X|Z) \cdot P(Y|Z)$ for all values of X, Y, and Z.

Example:

$$\mathbf{P}(\mathsf{Starts}, \mathsf{GasGauge}|\mathsf{Gas}) = \mathbf{P}(\mathsf{Starts}|\mathsf{Gas}) \cdot \mathbf{P}(\mathsf{GasGauge}|\mathsf{Gas})$$

	Starts		–Starts	
	GasGauge	–GasGauge	GasGauge	–GasGauge
Gas	.630	.070	.270	.030
–Gas	.003	.007	.297	.693

	GasGauge	–GasGauge
Gas	.9	.1
–Gas	.3	.7

	Starts	–Starts
Gas	.7	.3
–Gas	.01	.99

Simplifying the Joint Distribution by Conditional Independence

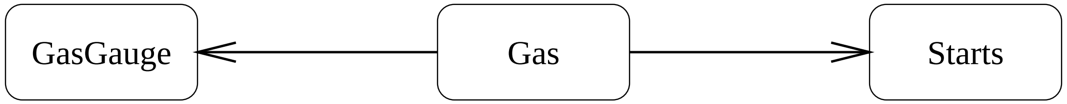
By the definition of conditional probability, we have

$$\mathbf{P}(\mathsf{Gas}, \mathsf{GasGauge}, \mathsf{Starts}) = \mathbf{P}(\mathsf{GasGauge}, \mathsf{Starts} | \mathsf{Gas})\mathbf{P}(\mathsf{Gas})$$

But by conditional independence, the first factor can be rewritten as

$$\mathbf{P}(\mathsf{Gas}, \mathsf{GasGauge}, \mathsf{Starts}) = \mathbf{P}(\mathsf{GasGauge} | \mathsf{Gas})\mathbf{P}(\mathsf{Starts} | \mathsf{Gas})\mathbf{P}(\mathsf{Gas})$$

Bayes Network Diagram



- **One node for each random variable**
- **Nodes with only outgoing arrows store marginal distributions**
  The Gas node stores $\mathbf{P}(\mathsf{Gas})$.
- **Nodes with incoming arrows store conditional probability tables**
  The Starts node stores $\mathbf{P}(\mathsf{Starts}|\mathsf{Gas})$.
- **Only direct dependencies are shown.** If two nodes are not directly connected, then they are conditionally independent on any set of nodes that disconnects them.
  Starts and $\mathsf{GasGauge}$ are conditionally independent given Gas.
- **Joint distribution is product of the distributions of the nodes.**

$$\mathbf{P}(\mathsf{Gas}, \mathsf{GasGauge}, \mathsf{Starts}) = \mathbf{P}(\mathsf{GasGauge}|\mathsf{Gas})\mathbf{P}(\mathsf{Starts}|\mathsf{Gas})\mathbf{P}(\mathsf{Gas})$$

Conformal Products

Given two probability tables, their **conformal product** is constructed by allocating space for the resulting table and then filling in each cell with the product of the corresponding cells.

	Starts		–Starts	
	GasGauge	–GasGauge	GasGauge	–GasGauge
Gas	$a11 \cdot b11$	$a10 \cdot b11$	$a11 \cdot b10$	$a10 \cdot b10$
–Gas	$a01 \cdot b01$	$a00 \cdot b01$	$a01 \cdot b00$	$a00 \cdot b00$

	GasGauge	–GasGauge
Gas	$a11$	$a10$
–Gas	$a01$	$a00$

	Starts	–Starts
Gas	$b11$	$b10$
–Gas	$b01$	$b00$

Commutativity and Associativity of Conformal Products

We can perform conformal products in any order. The only essential thing is that in each cell $x[i, j, k]$ in the final joint distribution, we have the product of the corresponding cells in each of the conditional probability tables.

$$P(\neg\mathsf{Gas}, \mathsf{GasGauge}, \neg\mathsf{Starts}) = P(\neg\mathsf{Starts}|\neg\mathsf{Gas}) \cdot P(\mathsf{GasGauge}|\neg\mathsf{Gas}) \cdot P(\mathsf{Gas})$$