# Evaluating a Policy under Model and Sampling Uncertainty

We want to evaluate the expected profit of a policy $\pi$ when there are multiple sources of randomness:

- **Model parameters ($S_0$):** e.g. the mean demand $\mu$, which is itself uncertain and follows a distribution.
- **Observed data ($W^1, \ldots, W^N$):** drawn from the demand distribution, conditional on $\mu$.
- **Learned decision ($x^{\pi,N}$):** chosen by policy $\pi$ based on the dataset $W^1, \ldots, W^N$.
- **Future randomness ($\widehat{W}$):** the realized demand after the policy is implemented.

## Dependency chain

1. The model parameter $\mu$ is uncertain:

$$\mu \sim f(\mu),$$

so our knowledge about the distribution of $\mu$ is encoded in the initial state $S_0$.

2. Given $\mu$, we observe a dataset of size $N$:

$$W^1, \ldots, W^N \sim \mathcal{D}(\mu),$$

for example $W^i \sim \text{Poisson}(\mu)$.

3. Based on this dataset, the learning algorithm (policy $\pi$) produces a decision:

$$x^{\pi,N} = \pi(W^1, \ldots, W^N).$$

If $\pi$ is randomized, then $x^{\pi,N}$ is a random variable even given the dataset.

4. Once the policy is fixed, nature reveals new randomness $\widehat{W}$ (future demand):

$$\widehat{W} \sim \mathcal{D}(\mu).$$

5. The realized profit depends on both the decision and the demand:

$$F(x^{\pi,N}, \widehat{W}).$$

## Nested expectations

To evaluate the policy properly, we must average over all sources of uncertainty:

$$F^\pi = \mathbb{E}_{S_0}\left[\mathbb{E}_{W^1,\ldots,W^N|S_0}\left[\mathbb{E}_{x^{\pi,N}|W^1,\ldots,W^N}\left[\mathbb{E}_{\widehat{W}|x^{\pi,N}}\left[F(x^{\pi,N}, \widehat{W})\right]\right]\right]\right].$$

## Interpretation layer by layer

- **Innermost:** $\mathbb{E}_{\widehat{W}|x^{\pi,N}}[F(x^{\pi,N},\widehat{W})]$
  Given a fixed decision $x^{\pi,N}$, compute its average profit over possible future outcomes.

- **Next:** $\mathbb{E}_{x^{\pi,N}|W^1,\ldots,W^N}[\cdot]$
  Average over possible decisions that the algorithm might produce (if randomized).

- **Next:** $\mathbb{E}_{W^1,\ldots,W^N|S_0}[\cdot]$
  Average over possible training datasets drawn from the distribution determined by $S_0$.

- **Outermost:** $\mathbb{E}_{S_0}[\cdot]$
  Average over uncertainty about the initial parameter (e.g. $\mu$).

## Summary

> The profit depends on the realized demand. The demand depends on the policy we chose. The policy depends on the sample we observed. The sample depends on the underlying parameter distribution. Therefore, we must integrate over all these layers of uncertainty.