A Appendix

Proof of Lemma 3.1. We need to show for any stochastic optimization problem, POC $\leq \mathcal{L}(z^*_{\text{two-stage}}; x)/\mathcal{L}(z^*; x)$.

By the definition of the price of correlation, POC $= \mathcal{L}(z^{'}; x)/\mathcal{L}(z^*; x)$, where $z^{'}$ is the optimal solution to a proxy stochastic program over Boolean variables that makes the assumption that all random variables are mutually independent. This is equivalent to reducing all Boolean variables to their marginals.

$z^*_{\text{two-stage}}$ is the optimal solution to a proxy stochastic program where all random variables are reduced to their marginal expectation $E[y|x]$, which is equivalent for mutually independent Boolean variables y. Therefore $z^*_{\text{two-stage}}$ and $z^{'}$ are optimal solutions to the same proxy stochastic program. In case there are multiple optimal solutions, we assume $z^*_{\text{two-stage}}$ and $z^{'}$ are found using the same tie-breaking scheme and therefore have the same loss, $\mathcal{L}(z^{'}; x) = \mathcal{L}(z^*_{\text{two-stage}})$. Therefore, POC $= \mathcal{L}(z^*_{\text{two-stage}}; x)/\mathcal{L}(z^*; x)$. □

Proof of Proposition 3.2. First, we provide some intuition for the proof. We define the end-to-end approach to be restricted to outputting a deterministic prediction vector y for the downstream problem. We construct an example where this y vector represents probabilities of two Boolean events. For end-to-end to be optimal, it must output marginal probabilities y that elicit the optimal solution on the downstream task. The proof shows that there is no such setting of y to elicit the optimal solution. We construct an example where to be optimal, end-to-end would need to output marginal probabilities for events 1 and 2 such that only two scenarios have positive probability (1) when both events happen and (2) when neither event happens. The only way to assign positive probability on both of these scenarios is to also assign positive probability on the two scenarios where only one event happens. This causes the end-to-end approach to find the wrong solution.

Formally, we need to show there exist stochastic optimization problems such that $\mathcal{L}(z^*; x) < \mathcal{L}(z^*_{\text{end-to-end}}; x)$. We construct a setting where $\mathcal{L}(z^*; x)$ is a C factor better for any constant $C > 2$.

Let b_1,b_2 be Boolean events of a set S and $E_{S \in D}[f(z, S)]$ be a stochastic optimization problem. The two-stage approach returns the optimal solution given the marginal probabilities p_1,p_2 of events b_1, b_2 occurring. The end-to-end approach sets p_1, p_2 arbitrarily to produce a corresponding z with best downstream loss:

$$\mathcal{L}(z^*_{\text{end-to-end}}; x) = \min_{p_1, p_2} E_{S \in D}[f(\arg\min_z E_{S' \in p_1, p_2}[f(z, S')], S)]$$

We define the following distribution over b_1, b_2:

$$D = \begin{cases} P(b_1 = T, b_2 = T) = 0.5 \\ P(b_1 = F, b_2 = F) = 0.5 \\ otherwise\ 0. \end{cases}$$

Let $f(z, S)$ have the following cost matrix:

	S_1 $b_1 = T, b_2 = T$	S_2 $b_1 = F, b_2 = F$	S_3 $b_1 = T, b_2 = F$	S_4 $b_1 = F, b_2 = T$
z^*	1	1	∞	∞
z_1	ϵ	C	ϵ	ϵ
z_2	C	ϵ	ϵ	ϵ

z^* is optimal for $C > 2$. z_1 and z_2 both cost an $O(C)$ factor greater than z^*. We prove that end-to-end must select either suboptimal solutions z_1 or z_2 regardless of how p_1 and p_2 are set. There are two cases:

1. If end-to-end sets p_1 and p_2 such that all probability is on S_1 or S_2, z^* will not be chosen since z_1 is best for S_1 and z_2 is best for S_2.
2. If end-to-end sets p_1 and p_2 such that there is non-zero probability on both S_1 and S_2, there must also be non-zero probability on S_3 and S_4 because p_1 and p_2 are marginal probabilities. Since z^* has infinite cost for S_3 and S_4, z_1 and z_2 will always look better to end-to-end than z^*.

□

Proof of Theorem 3.3: Optimality of end-to-end for (i) two-stage minimum cost flow. This setting is from Example 1 of Agrawal et al. (2012). There are n Boolean events with marginal probabilities $p_1, ..., p_n$. Each event corresponds to the existence a sink node with unit demand. Each sink node connects to a single hub node. The optimization problem is to buy capacity for the hub node such that all demand is satisfied. Before seeing the realization of sink nodes, in the first stage you can buy capacity z at price $c^I(z)$. Once you have seen the realization of sink nodes, you then must buy enough extra capacity x at higher cost $c^{II}(x)$ to satisfy the demand requests. The goal is to optimize z to minimize expected costs. The problem can be written formally in a single stage as follows:

$$\arg\min_z c^I(z) + \sum_{i=z}^{n} P\left((\sum_j y_j) = i\right) c^{II}(n-i)$$

s.t., $y_1, ..., y_n$ are the realization of Boolean events, and $c^{II}(k) > c^I(k), \forall k$
A two-stage approach wrongly estimates $\sum_{i=z}^{n} P((\sum_j b_j) = i)$ by multiplying marginals together to compute probabilities of sets of events occurring. We now prove $\mathcal{L}(z^*_{\text{end-to-end}}; x) = \mathcal{L}(z^*; x)$ for this setting:

The end-to-end approach can set $p_1...p_n$ arbitrarily. For any given z^*, it is enough to construct $p_1...p_n$ such that $z^*_{\text{end-to-end}} = z^*$. Idea is to set $p_1...p_n$ to make the two-stage costs sufficiently high for $z > z^*$ and low for $z < z^*$ so z^* is optimal.

We can put all probability on a single set S such that $p_1, ..., p_{z^*} = 1$ and $p_{z^*+1}, ..., p_n = 0$. That way, we pay zero second-stage cost for increasing z above z^* since we've satisfied all demand, but add first-stage costs. If we decrease z by k, we must pay $c^{II}(k)$ second-stage cost but only save $c^I(k)$. Since $c^{II}(k) > c^I(k)$, decreasing z also increases costs. Since any change in z increases costs, $z^*_{\text{end-to-end}} = \arg\min E_f(z, S) = z^*$.

□

Proof of Theorem 3.3: Optimality of end-to-end for (i) two-stage stochastic set cover. This setting is from Example 2 of Agrawal et al. (2012). There are n Boolean events with marginal probabilities $p_1, ..., p_n$. Each event corresponds to an item in a set V. There are k disjoint subsets whose union is V. The optimization problem is to buy subsets to cover realized items. Before seeing the realization of events, in the first stage you can buy any subset $z_1, ..., z_k$ at price c^I each. Once you have seen the realization, you then must pay a higher cost c^{II} to buy the uncovered subset with the maximum number of realized items. The goal is to optimize z to minimize expected costs. The problem can be written formally in a single stage as follows:

$$\arg\min_z \sum_i c^I z_i + c^{II} \max_{i=1..k} \left\{ \sum_{S \in S_i} P(S)|S| \cdot (1 - z_i) \right\}$$

s.t. $c^I < c^{II}$
S is subset of items in disjoint subset S_i. i equals ith of k disjoint subsets, whose union is S. We now prove $\mathcal{L}(z^*_{\text{end-to-end}}; x) = \mathcal{L}(z^*; x)$ for this setting:

For any given z^*, we need to construct $p_1...p_n$ such that $z^*_{\text{end-to-end}} = z^*$. For a given z^*, put all probability on a single set S by assigning probability 1 to all events covered by z^* and probability 0 to all events uncovered. z^* can never be improved by adding sets, since all events are already covered and there is only one set that can cover each event (i.e., disjoint subsets). Therefore, z^* will always be optimal as long as c^I (the cost saved by uncovering a set in first stage) is less than $c^{II}|S|$ (the cost of covering in second stage) for smallest S that is covered. Every covered set has a least one event by our construction, therefore uncovering a set would incur a cost of at least c^{II} but only save c^I. Since $c^I < c^{II}$, we can never reduce costs by reducing the covered sets in z. Therefore, $z^*_{\text{end-to-end}} = \arg\min f(z, S) = z^*$. □

Proof of Theorem 3.3: Optimality of end-to-end for (i) stochastic optimization with monotone submodular cost function. This setting is from Example 3 of Agrawal et al. (2012). There are n Boolean events with marginal probabilities $p_1, ..., p_n$. Let the realized set of Boolean events be S. Suppose you have a monotone submodular cost function $c(S)$. You want to decide whether to (1) pay the expected cost of the subset $E_S[c(S)]$ or (2) pay a constant cost C.

$$\arg\min_z c(S) z_1 + C z_2$$
$$z_1 + z_2 = 1, z_1, z_2 \in \{0, 1\}$$

We now prove $\mathcal{L}(z^*_{\text{end-to-end}}; x) = \mathcal{L}(z^*; x)$ for this setting:

Given marginal probabilities $p_1, ..., p_n$, the expected subset cost is $\sum_S P(S)c(S)$ where $P(S)$ is the product of probabilities for all events in that set S. For optimality, end-to-end needs to find $p_1, ..., p_n$ such that $\sum_S P(S)c(S) = E_S[c(S)]$. Let $C^* = E_S[c(S)]$. First, there must be some set of events S that costs $\geq C^*$, since $C^* \leq max_S c(S)$. Second, among the sets that satisfy this property, by monotonicity there must be some set S' and event e such that $c(S' - \{e\}) \leq C^*$ since $C^* \geq min_S c(S)$. Setting aside event e, we set probabilities to 1 for events in S' and 0 otherwise. For event e, set $p_e \in [0, 1]$ such that $p_e c(S') + (1 - p_e c(S' - \{e\}) = C^*$. (i.e., linear interpolation between S' and $S' - \{e\}$). There must be a solution for p_e since $c(S' - \{e\}) \leq C^* \leq c(S')$

□

Proof of Theorem 3.4. We first define the optimization problem and distribution that we will use to construct our gap. Our construction only depends on the distribution over Y, so without loss of generality, we let $P(Y|X) = P(Y)$; an analogous construction could be made for every x. We set the optimization problem to be,

$$\min \mathbb{E}_{P(Y)}[C + \sum_{i=1}^{d}(y_{i,1}y_{i,2}z_{i,1}) + (y_{i,3}y_{i,4}z_{i,2})]$$

$$\text{subject to } \sum_{i=1}^{d} z_{i,1} + z_{i,2} \geq d, \quad \forall i \quad 1 \geq z_{i,1}, z_{i,2} \geq 0,$$

where $y \in \mathbb{R}^{d\times 4}$ and $z \in \mathbb{R}^{d\times 2}$. Note that any basic solution to this optimization problem must set at least half of the z variables to be equal to 1.

Now define some large positive $N \in \mathbb{R}$ and some small $\epsilon > 0$. Let $P(Y) = P(Y_1, Y_2)P(Y_3, Y_4)$ be distributed as follows.

$$P(Y_1, Y_2) = \begin{cases} 0.5 & y_{i,1} = 0, y_{i,2} = N \\ 0.5 & y_{i,1} = N, y_{i,2} = 0 \end{cases}$$

$$P(Y_3, Y_4) = \begin{cases} 1. & y_{i,3} = y_{i,4} = \frac{N}{2} - \epsilon \end{cases}$$

The intuition behind this distribution is that $\forall i$, $y_{i,1}y_{i,2} = 0$ and $y_{i,3}y_{i,4} > 0$ despite the fact that $y_{i,1}$ and $y_{i,2}$ have higher expected values than $y_{i,3}$ and $y_{i,4}$.

Notice that this means half of our z variables have coefficients of 0 and we can satisfy our constraints without raising the objective. Therefore, the optimal end-to-end solution is to set z to be 0 anytime the product of targets is nonzero in expectation; i.e., $z^*_{\text{end-to-end}}$ sets $z_{i,1} = 1, z_{i,2} = 0$, for all i.

The loss obtained by choosing $z^*_{\text{end-to-end}}$ is,

$$\mathcal{L}(z^*_{\text{end-to-end}}; x) = \mathbb{E}_{P(Y)}[C + \sum_{i=1}^{d}(y_{i,1}y_{i,2}1.0) + (y_{i,3}y_{i,4}0)] = C.$$

The last equality follows from the fact that $\forall i, y_{i,1}y_{i,2} = 0$.

However, as mentioned above $y_{i,1}$ and $y_{i,2}$ have higher expected values than $y_{i,3}$ and $y_{i,4}$. Therefore, the optimal two-stage solution is to set z to be 1 for each pair with the lower individual expected values i.e. $z^*_{\text{two-stage}}$ sets $z_{i,1} = 0, z_{i,2} = 1$ for all i.

The loss obtained by choosing $z^*_{\text{two-stage}}$,

$$\begin{aligned}\mathcal{L}(z^*_{\text{two-stage}}; x) &= \mathbb{E}_{P(Y)}[C + \sum_{i=1}^{d}(y_{i,1}y_{i,2}0.0) + (y_{i,3}y_{i,4}1.0)] \\ &= C + \sum_{i=1}^{d}(\frac{N}{2} - \epsilon)^2 1.0 \approx \frac{dN^2}{8}.\end{aligned}$$

Therefore, the multiplicative gap $\left(\frac{\mathcal{L}(z^*_{\text{two-stage}};x)}{\mathcal{L}(z^*_{\text{end-to-end}};x)}\right)$ is dN^2 up to constant factors.

$\square$

Proof of Lemma 3.5. We set end-to-end to output $y'|x$ such that $y'_{i,1} = \mathbb{E}_{y\sim P(Y|x)}[\gamma(y_{i,1}, y_{i,2})]$ for $i = 1, \ldots, d$ and $y'_{i,2} = 1$ for $i = 1, \ldots, d$.

$\mathcal{L}(z^*_{\text{end-to-end}}; x) = \min_z \sum_{i=1}^{d} \gamma(y'_{i,1}, y'_{i,2}) \cdot f_i(z)$

We know that

$$\begin{aligned}\mathcal{L}(z^*; x) &= \min_z \mathbb{E}_{y\sim P(Y|x)}\left[\sum_{i=1}^{d} \gamma(y_{i,1}, y_{i,2}) \cdot f_i(z)\right] \\ &= \min_z \sum_{i=1}^{d} \mathbb{E}_{y\sim P(Y|x)}[\gamma(y_{i,1}, y_{i,2}) \cdot f_i(z)] \\ &= \min_z \sum_{i=1}^{d} f_i(z) \cdot \mathbb{E}_{y\sim P(Y|x)}[\gamma(y_{i,1}, y_{i,2})]\end{aligned}$$

By the definition of γ

$$\begin{aligned}\mathcal{L}(z^*;x) &= \min \sum_{i=1}^{d/2} f_i(z)\cdot\gamma(\mathbb{E}_{y\sim P(Y|x)}[y_{i,1},y_{i,2}],1)\\ &= \min \sum_{i=1}^{d/2} f_i(z)\cdot\gamma(y'_{i,1},y'_{i,2})\\ &= \mathcal{L}(z^*_{\text{end-to-end}};x)\end{aligned}$$

□

Proof of Lemma 3.6. A function $\gamma(y,y')$ is linear if and only if $\forall P(Y|x), \mathbb{E}_{\mathcal{D}}[\gamma(y,y')] = \gamma(\mathbb{E}_{\mathcal{D}}[y],\mathbb{E}_{\mathcal{D}}[y'])$.

The reverse direction follows by linearity of expectation so we must show that if a function $\gamma(y)$ is nonlinear then

$$\exists\mathcal{D}, \mathbb{E}_{\mathcal{D}}[\gamma(y)] \neq \gamma(\mathbb{E}_{\mathcal{D}}[y]).$$

A function γ is linear if and only if $\gamma(\alpha y_1 + (1-\alpha)y_2) = \alpha\gamma(y_1) + (1-\alpha)\gamma(y_2)\ \forall y_1, y_2, \alpha$. If a function is nonlinear then by definition there must exist two points y_1, y_2 between which the function is nonlinear. This means there exists a point $y' = \alpha y_1 + (1-\alpha)y_2$ such that $\gamma(y') \neq \gamma(\alpha y_1 + (1-\alpha)y_2)$ therefore if we choose our distribution such that $Pr[y=y_1]=\alpha$ and $Pr[y=y_2]=1-\alpha$, it is easy to see that $\mathbb{E}_{\mathcal{D}}[\gamma(y)] \neq \gamma(\mathbb{E}_{\mathcal{D}}[y])$. □

Proof of Theorem 3.7. We consider the case of $d=2$; the extension to arbitrarily many dimensions is trivial. We first define our function f,

$$f(y,z) = \gamma(y_{1,1},y_{2,1})f_1(z) + \gamma(y_{2,1},y_{2,2})f_2(z).$$

We construct $P(Y)$ and small $\epsilon > 0$ such that $\mathbb{E}_{P(Y)}[\gamma(y_{1,1},y_{1,2})] \neq \gamma(\mathbb{E}_{P(Y)}[y_{1,1}],\mathbb{E}_{P(Y)}[y_{1,2}])$. Such a distribution is guaranteed to exist by Lemma 3.6.

Now there are two cases,

1. $\mathbb{E}_{P(Y)}[\gamma(y_{1,1},y_{1,2})] < \gamma(\mathbb{E}_{P(Y)}[y_{1,1}],\mathbb{E}_{P(Y)}[y_{1,2}])$ where we choose point masses for our remaining two values such that $p(y_{2,1} = \mathbb{E}_{P(Y)}[y_{1,1}] - \epsilon) = 1.0$ and $p(y_{2,2} = \mathbb{E}_{P(Y)}[y_{1,2}] - \epsilon) = 1.0$;
2. $\mathbb{E}_{P(Y)}[\gamma(y_{1,1},y_{1,2})] > \gamma(\mathbb{E}_{P(Y)}[y_{1,1}],\mathbb{E}_{P(Y)}[y_{1,2}])$, where we choose point masses for our remaining two values such that $p(y_{2,1} = \mathbb{E}_{P(Y)}[y_{1,1}] + \epsilon) = 1.0$, and $p(y_{2,2} = \mathbb{E}_{P(Y)}[y_{1,2}] + \epsilon) = 1.0$.

Without loss of generality, we assume that we are in the first case and our constructed optimization problem will minimize f. The example works symmetrically in the second case where we can construct the optimization problem to maximize f.

We show that two-stage is suboptimal for the following optimization problem

$$\min_z\ \ \mathbb{E}_{y\sim P(Y)}[\gamma(y_{1,1},y_{1,2})z_1 + \gamma(y_{2,1},y_{2,2})z_2] \quad \text{subject to } z_1 + z_2 \geq 1.$$

It is easy to see that the optimal choice of z is $z^*_{OPT} = \{z_1 = 1.0, z_2 = 0.0\}$, which gives a loss of $\mathcal{L}(z^*_{OPT}) = \mathbb{E}_{P(Y)}[\gamma(y_{1,1},y_{1,2})]$. However, since two-stage makes its choices with respect to $\gamma(\mathbb{E}_{P(Y)}[y_{1,1}],\mathbb{E}_{P(Y)}[y_{1,2}])$ and $\gamma(\mathbb{E}_{P(Y)}[y_{2,1}],\mathbb{E}_{P(Y)}[y_{2,2}])$, it chooses the solution $z^*_{\text{two-stage}} = \{z_1 = 0.0, z_2 = 1.0\}$, giving it a loss of $\mathcal{L}(z^*_{\text{two-stage}}) = \mathbb{E}_{P(Y)}[\gamma(y_{2,1},y_{2,2})]$.

If we choose a small enough ϵ,$\mathbb{E}_{P(Y)}[\gamma(y_{1,1},y_{1,2})] < \mathbb{E}_{P(Y)}[\gamma(y_{2,1},y_{2,2})]$, and hence $\mathcal{L}(z^*_{OPT}) < \mathcal{L}(z^*_{\text{two-stage}})$. This construction trivially extends to $d > 2$ by making $f_i(z) = 0, \forall i > 2$. □

Proof of Theorem 3.8, Condition 1. We know that $\mathcal{L}_{OPT} = \min \mathbb{E}_{y\sim P(Y|x)}[f(y,z)]$ and $\mathcal{L}_{\text{two stage}} = \min_z f(\mathbb{E}_{y\sim P(Y|x)}[y],z)]$

Plugging in our definition of $f(y,z)$ from Equation 5 we get

$$\begin{aligned}\mathcal{L}_{OPT} &= \min \mathbb{E}_{y\sim P(Y|x)}[\sum_{i=1}^{d}\gamma(y_{i,1},y_{i,2})\cdot f_i(z)]\\ &= \min\sum_{i=1}^{d}\mathbb{E}_{y\sim P(Y|x)}[\gamma(y_{i,1},y_{i,2})\cdot f_i(z)]\\ &= \min\sum_{i=1}^{d} f_i(z)\cdot\mathbb{E}_{y\sim P(Y|x)}[\gamma(y_{i,1},y_{i,2})].\end{aligned}$$

Since we know γ is a linear function, by linearity of expectation

$$\begin{aligned}\mathcal{L}_{OPT} &= \min_z \sum_{i=1}^{d} f_i(z) \cdot \gamma(\mathbb{E}_{y\sim P(Y|x)}[y_{i,1}], \mathbb{E}_{y\sim P(Y|x)}[y_{i,2}]) \\ &= \mathcal{L}_{\text{two stage}}.\end{aligned}$$

$\square$

Proof of Theorem 3.8, Condition 2. We know that $\mathcal{L}_{OPT} = \min \mathbb{E}_{y\sim P(Y|x)}[f(y,z)]$ and $\mathcal{L}_{\text{two stage}} = \min_z f(\mathbb{E}_{y\sim P(Y|x)}[y], z)]$

Plugging in our definition of $f(y,z)$ from Equation 5 we get

$$\begin{aligned}\mathcal{L}_{OPT} &= \min_z \mathbb{E}_{y\sim P(Y|x)}[\sum_{i=1}^{d} \gamma(y_{i,1}, y_{i,2}) \cdot f_i(z)] \\ &= \min_z \sum_{i=1}^{d} \mathbb{E}_{y\sim P(Y|x)}[\gamma(y_{i,1}, y_{i,2}) \cdot f_i(z)] \\ &= \min_z \sum_{i=1}^{d} f_i(z) \cdot \mathbb{E}_{y\sim P(Y|x)}[\gamma(y_{i,1}, y_{i,2})].\end{aligned}$$

Since $\gamma(y, y') = y \cdot y'$:

$$\begin{aligned}\mathcal{L}_{OPT} &= \min_z \sum_{i=1}^{d} f_i(z) \cdot \mathbb{E}_{y\sim P(Y|x)}[y_{i,1} \cdot y_{i,2}] \\ &= \min_z \sum_{i=1}^{d} f_i(z) \cdot \int_{y\sim P(Y|x)} y_i P(y_{i,1}) \cdot y_{i,2} P(y_{i,2}|y_{i,1})).\end{aligned}$$

Since $P(y_{i,2}) = P(y_{i,2}|y_{i,1})$, by independence:

$$\begin{aligned}&\min_z \sum_{i=1}^{d} f_i(z) \cdot \int_{y\sim P(Y|x)} y_{i,1} P(y_{i,1}) \cdot \int_{y\sim P(Y|x)} y_{i,2} P(y_{i,2}) \\ &\min_z \sum_{i=1}^{d} f_i(z) \cdot \gamma(\mathbb{E}_{y\sim P(Y|x)}[y_{i,1}], \mathbb{E}_{y\sim P(Y|x)}[y_{i,2}]) \\ &= \mathcal{L}_{\text{two stage}}.\end{aligned}$$

$\square$

Generalization of Theorem 3.8 for objective functions that are linear in y. We know that $\mathcal{L}_{OPT} = \min \mathbb{E}_{y\sim P(Y|x)}[f(y,z)]$ and $\mathcal{L}_{\text{two stage}} = \min_z f(\mathbb{E}_{y\sim P(Y|x)}[y], z)]$. Since $f(y,z)$ is linear in y, by linearity of expectation we can move expectation inside f:

$$\begin{aligned}\mathcal{L}_{OPT} &= \min_z f(\mathbb{E}_{y\sim P(Y|x)}[y, z]) \\ &= \min_z f(\mathbb{E}_{y\sim P(Y|x)}[y], z) && \text{pulling out } z \text{ since expectation over } y \\ &= \mathcal{L}_{\text{two stage}}.\end{aligned}$$

$\square$