Data-Driven Profit Estimation Error in the Newsvendor Model

Andrew F. Siegel
Professor Emeritus of Information Systems and Operations Management, Michael G. Foster School of Business, University of Washington, asiegel@uw.edu.

Michael R. Wagner
Information Systems & Operations Management, Michael G. Foster School of Business, University of Washington, mrwagner@uw.edu.

In this paper we identify a statistically significant error in naively estimating the expected profit in a data-driven Newsvendor Model, and we show how to correct the error. In particular, we analyze a Newsvendor Model where the continuous demand distribution is not known, and only a sample of demand data is available. In this context, an empirical demand distribution, that is induced by the sample of data, is used in place of the (unknown) true distribution. The quantity at the critical percentile $1 - c/p$ of the empirical distribution is known as the Sample Average Approximation (SAA) order quantity, where $p$ is the unit revenue and $c$ the unit cost. We prove that, if the empirical distribution is used to estimate the expected profit, this estimate exhibits a positive, statistically significant bias. We derive a closed-form expression for this bias, that only depends on $p$ and $c$, and the sample of data. We argue that the dominant source of the bias is underage, and overage plays a minimal role. Furthermore, we provide an in-depth interpretation of the bias, connecting it mathematically with a covariance term as well as visualizing the bias using areas under curves. The bias expression can then be used to design an adjusted expected profit estimate, that we prove is asymptotically unbiased. Numerical hypothesis testing experiments confirm that the unadjusted estimation error is statistically significant, whereas the adjusted estimation error is not significantly different from zero. The bias is not negligible in our numerical experiments: for lognormally and normally distributed demand, the unadjusted error is 2.4% and 3.0% of the true expected profit, respectively. A more detailed exploration with exact finite-sample results, for exponentially distributed demand, demonstrates that the estimation error percentage can be much larger. We also provide an extension for discrete demand distributions that, in numerical experiments, exhibits encouraging bias reduction results.

Key words: Newsvendor, estimation error, statistics, data-driven

1. Introduction

In this paper we identify a statistically significant error in naively estimating the expected profit in a data-driven Newsvendor Model, and we show how to correct the error. In particular, we study a Newsvendor Model that applies the Sample Average Approximation (SAA) order quantity, an approach that replaces the unknown true demand distribution with a data-driven empirical
distribution in the classic Newsvendor formulas. We demonstrate that estimating the maximized expected profit using the empirical distribution systematically overestimates the true expected profit, leading a decision maker to believe that the expected profit is larger than it actually is. Practically speaking, the expected profit serves as a forecast for the future realized profit, and we effectively show that this forecast is biased. We demonstrate how this bias can be estimated from data, which allows us to adjust the expected profit estimate, and obtain an asymptotically unbiased estimate.

Our paper is related to, but considerably distinct from Siegel and Wagner (2021), though further motivations and examples may be found therein. Although Siegel and Wagner (2021) also identified and corrected a profit bias in the Newsvendor Model, the authors do so under the very strong assumption of a parametric continuous demand distribution, which simplifies the analysis. In contrast, in our paper we do not assume that the demand distribution belongs to any particular parametric family of distributions; we only assume that the distribution is smooth, and that we have a single sample of demand data from the distribution. Therefore, our paper does not require a decision maker to assume a specific parametric form for the demand distribution, as Siegel and Wagner (2021) require, which makes the adoption of our results in practice easier than those of Siegel and Wagner (2021). Our analysis techniques are different too; we utilize conditional distribution theory along with both exact and asymptotic properties of order statistics, whereas Siegel and Wagner (2021) relied on the theory of maximum likelihood estimation. Furthermore, and in stark contrast to Siegel and Wagner (2021), we provide the first results for a discrete distribution of demand.

We rigorously prove that using the empirical distribution in place of the true distribution will lead to a positive bias in estimating expected profit, on average. We derive the bias in closed-form, which is a function of the sample size, the selling price and purchasing cost, as well as an estimate of the demand density at the (unknown, but estimated) true optimal ordering quantity. Our result is quite different than that of Siegel and Wagner (2021), whose bias expression depends on the parametric form of the demand distribution, and various partial derivatives of the distribution with respect to its parameters. Furthermore, our bias expression only depends on data, and does not require an additional layer of parametric estimation, as in Siegel and Wagner (2021). Our bias formula can be used directly to adjust the empirical-distribution-based SAA expected profit estimate, to obtain a new estimate that is provably asymptotically unbiased. We also perform numerical experiments that show that 1) using the empirical distribution to estimate expected profit leads to an expected estimation error that is statistically significant and 2) our adjusted expected profit estimate has no statistically significant expected estimation error. A more detailed numerical exploration, for exponentially distributed demand and exact finite-sample results, demonstrate that the bias can be a large percentage of the true expected profit.
1.1. Contributions

The primary contributions of our paper are as follows:

- For a generic smooth demand distribution, from which we have only a sample of data, we derive, using basic probability theory along with conditional distribution theory, the asymptotic properties of order statistics, and Taylor series expansions, a closed-form expression for the (positive) expected bias that results from estimating the expected profit using the empirical distribution in place of the true (unknown) distribution. This bias expression can be estimated in a way that only depends on the selling price, the purchase cost, and the sample data. Using the estimated bias formula, we adjust the empirical-distribution expected profit estimate, and we prove that the resulting adjusted estimate is asymptotically unbiased. We also argue that the dominant source of the bias is underestimation, and overestimation plays a minimal role. Furthermore, we provide an in-depth interpretation of the bias, connecting it mathematically with a covariance term as well as visualizing the bias using areas under curves.

- Hypothesis testing Monte Carlo simulation experiments confirm our theoretical results. We demonstrate that the unadjusted estimation error is significantly positive, whereas the adjusted estimation error is not significantly different from zero. Notably, the unadjusted estimation errors are 2.4-3.0% of the true expected profit values in these experiments, though we also demonstrate that the percentage error can be much higher, depending on the economic parameters, sample size, and underlying true demand distribution. In similar numerical experiments, we also find that our non-parametric environment results in a much larger profit bias than that encountered in the parametric environment studied in Siegel and Wagner (2021); interestingly, our adjusted expected profit expressions eliminate the larger bias just as well as the parametric adjustments derived in Siegel and Wagner (2021) do for the smaller bias. We also, for normal and lognormal distributions of demand, find that our analytical adjustments eliminate bias better than a cross-validated estimation of expected profit.

- While our theory is heavily dependent on the assumption of a continuous demand distribution, our analysis allows us to conjecture the form of the bias for a discrete distribution. We use this conjectured bias to form an adjusted expected profit expression, which, via Monte Carlo simulation experiments, substantially reduces bias by 49.0-94.1%.

1.2. Literature Review

While there is a vast literature related to the Newsvendor Model, we focus on the most relevant data-driven Newsvendor papers. Kleywegt et al. (2002) analyzed the sample average approximation (SAA) method that uses a data-driven empirical distribution in place of the unknown true distribution for general stochastic discrete optimization problems, and identified a bias in estimating
the optimal objective value, but did not provide an adjustment term to correct the bias. Levi et al. (2007) showed that, in the Newsvendor Model, the SAA method provides a solution that is provably near optimal, with high probability, a result that was improved by Levi et al. (2015) by introducing an additive bias into the order quantity. Ban and Rudin (2018) further extended the approach of Levi et al. (2015, 2007) to include explanatory variables that influence the demand distribution using machine learning algorithms. He et al. (2012) similarly studied a features-based Newsvendor Model for staffing hospital operating rooms, and Ban et al. (2018) studied features of demand in a multi-period inventory management setting. In these references, the optimal expected profit depends on the true distribution, which is unknown, which implies that the profit is not calculable. Of course, the empirical distribution could be used to estimate the profit in these references; we show that this estimated profit is biased in the Newsvendor Model, and we also demonstrate how to correct for the bias to obtain an asymptotically unbiased profit estimate.

Other papers have implicitly or explicitly considered bias in a Newsvendor context. Liyanage and Shanthikumar (2005) considered the Newsvendor Model under an exponential demand distribution, where estimation (of the exponential distribution’s mean) and optimization (finding the optimal order quantity) are performed simultaneously, which results in the order quantity being intentionally biased to obtain higher expected profit in a data-driven context. Chu et al. (2008) consider parametric demand distributions characterized by location and scale parameters, and extend the results in Liyanage and Shanthikumar (2005) using Bayesian analysis. Siegel and Wagner (2021) is the paper most related to ours, in that a bias in estimating profit in the data-driven Newsvendor Model is identified and corrected. However, the methods and results presented here are considerably different from those of Siegel and Wagner (2021) for several reasons. First of all, Siegel and Wagner (2021) worked within a finite-dimensional parametric family of distributions, while our work here is nonparametric in the sense that we are working with only a single unknown smooth distribution with no known parametric form. Therefore our work here is much more general. Next, Siegel and Wagner (2021) made extensive use of the theory of maximum likelihood estimation to derive their results (along with the many required assumptions), whereas there is no corresponding theory available in the present nonparametric context; instead, we are able to (surprisingly, perhaps) derive our results for a single unknown distribution using primarily smoothness of the unknown density function together with conditional distribution theory and the asymptotic properties of order statistics.

**Paper Outline:** In Section 2 we introduce the basic Newsvendor model, its data-driven analogue, and our notion of bias. In Section 3 we derive a closed-form asymptotic expression for the bias, use it to create an adjusted expected profit expression that is asymptotically unbiased, and interpret the source of the bias. In Section 4 we provide an exact (non-asymptotic) expression for
the bias when demand is exponentially distributed. In Section 5 we describe numerical experiments that complement our theoretical analysis. In Section 6, we discuss and evaluate an extension of our results for a discrete distribution of demand. In Section 7 we provide concluding thoughts and directions for future research. All proofs appear in the appendix.

2. Preliminaries

The Newsvendor Model determines an order quantity \( y \) that maximizes expected profit in the face of random demand \( X \), where the unit sales price is \( p \) and the unit procurement cost is \( c \), with \( p > c > 0 \). We assume that \( X \) is a continuous random variable, with continuous density \( f \) and cumulative distribution function \( F \), with support on a non-degenerate interval within (or equal to) the non-negative real numbers. We assume that all relevant moments of \( X \) exist and are finite. We also assume that the derivative \( f' \) of the density is continuous and bounded. Note that we make no other assumption about the distribution. In particular, we do not assume that we know the value of any summary statistics (mean, variance, etc.) and we do not assume any parametric form for the distribution. The only information about \( F \) will be via a data sample from the distribution.

If the distribution \( F \) were known, the classic Newsvendor Model is

\[
\max_{y \geq 0} E_F(p \min\{X, y\} - cy),
\]

which has the well-known solution \( y^* = F^{-1}(1 - c/p) \), where \( 1 - c/p \) is known as the critical ratio. Unfortunately, in many situations the demand distribution \( F \) is not known. Instead, data are typically available. In the next subsection, we describe how a data-driven Newsvendor Model can be formulated.

A note on notation: we denote random variables as upper case letters and their realizations (or non-random quantities) in lower case letters.

2.1. A Data Perspective

In practice, one typically has access to data, which, in the Newsvendor context, is demand data \( x_1, \ldots, x_n \). Since the distribution \( F \) of demand \( X \) is not known, one may instead approximate \( F \) using the empirical distribution \( \hat{F} \), which is a discrete distribution with each of the observations \( x_i \) occurring with probability \( 1/n \). In other words, the Newsvendor Model is approximated by

\[
\max_y \frac{1}{n} \sum_{i=1}^{n} (p \min\{x_i, y\} - cy).
\]

This approach is known as the sample average approximation (SAA). Following Levi et al. (2007), on page 825, we define the optimal SAA order quantity \( y_{\text{saa}} \) to be

\[
y_{\text{saa}} \triangleq \min_{j=1,\ldots,n} \left\{ x_j : \frac{1}{n} \sum_{i=1}^{n} I_{x_i \leq x_j} \geq 1 - c/p \right\},
\]

where $I_A$ is the indicator function, which equals 1 if the event $A$ is true, and 0 otherwise. In particular, the expression $\frac{1}{n} \sum_{i=1}^{n} I_{x_i \leq y_{saa}}$ equals $\hat{F}(y_{saa})$ and approximates $F(y_{saa})$; recall that, in the Newsvendor Model, $F(y^*) = 1 - c/p$.

**Lemma 1.** Definition (2) of the optimal SAA order quantity is equivalent to $y_{saa} = x(k)$ with $k = [n(1 - c/p)]$, where $x(k)$ is the $k$-th order statistic and $[\cdot]$ denotes the ceiling function.

Note that, strictly speaking, $k$ is a function of $n$; however, we will use the notation $k$ instead of $k(n)$ for simplicity.

Our paper is primarily concerned with estimating expected profit for the $y_{saa}$ order quantity. A natural data-driven estimate for the expected profit is to simply plug this order quantity into the objective of Formulation (1):

$$\frac{1}{n} \sum_{i=1}^{n} (p \min\{x_i, y_{saa}\} - cY_{saa}).$$

(3)

Our first main result is to show that this approach results in a biased estimate of the true expected profit; to demonstrate this, we first view the data-driven problem from the perspective of random variables in the next section. Our second main result is to show how to correct this bias to provide an asymptotically unbiased data-driven estimate of expected profit.

### 2.2. A Random Variable Perspective

The SAA order quantity $y_{saa}$ depends on the sample of data that was used to calculate it in Equation (2). We therefore wish to study its dependence on the sampling distribution. In particular, we let $X_1, \ldots, X_n$ denote an independent identically distributed (i.i.d.) sample of size $n$ from the (unknown) demand distribution $F$, representing a sample of past demand data. We next introduce a probabilistic analogue of Equation (2):

$$y_{saa} \triangleq \min_{j=1, \ldots, n} \left\{ X_j : \frac{1}{n} \sum_{i=1}^{n} I_{X_i \leq X_j} \geq 1 - c/p \right\}.$$  

(4)

We let $g$ and $G$ denote the density and cumulative distribution function of $Y_{saa}$, respectively (where the functional forms of $g$ and $G$ will be given in Lemma 7). Note that, in Definition (4), $Y_{saa}$ is a random variable that explicitly depends on the random sample $X_1, \ldots, X_n$. Letting $X_{(1)} < X_{(2)} < \ldots < X_{(n)}$ denote the order statistics of the sample, we also provide a probabilistic analogue to the alternative characterization of $y_{saa}$ in Lemma 1:

$$y_{saa} = X_{(k)},$$

where $k = [n(1 - c/p)]$. Let $X$ denote future demand, obtained by drawing a new observation from $F$, independent of the past demand data $(X_1, \ldots, X_n)$. The true expected profit of the SAA order quantity can be written as

$$\bar{\pi}_{true} \triangleq E_F(p \min\{X, Y_{saa}\} - cY_{saa}),$$

(5)
which is the quantity we are focused on estimating unbiasedly. Examining Equation (3), it turns out that the natural estimate of expected profit associated with the SAA order quantity is a realization of the random variable $\frac{1}{n} \sum_{i=1}^{n} (p \min \{X_i, Y_{\text{saq}}\} - cY_{\text{saq}})$, whose expectation is

$$\pi_{\text{naive}} \triangleq E_F \left( \frac{1}{n} \sum_{i=1}^{n} (p \min \{X_i, Y_{\text{saq}}\} - cY_{\text{saq}}) \right). \tag{6}$$

We argue that this approach leads to erroneous, overly optimistic estimates of the true expected profit. Indeed, Theorem 1 implies that $\pi_{\text{naive}} = \text{Bias}_{\text{profit}} + \pi_{\text{true}} + o(1/n)$, where $\text{Bias}_{\text{profit}} > 0$, which demonstrates that the naive estimate in Equation (3) is biased upwards. The expansion of this equation,

$$E_F \left( \frac{1}{n} \sum_{i=1}^{n} (p \min \{X_i, Y_{\text{saq}}\} - cY_{\text{saq}}) \right) = \text{Bias}_{\text{profit}} + E_F (p \min \{X, Y_{\text{saq}}\} - cY_{\text{saq}}) + o \left( \frac{1}{n} \right), \tag{7}$$

lends to a machine learning interpretation, in terms of in-sample versus out-of-sample performance.

In the left expression of Equation (7), there are $n$ i.i.d. realizations of demand $X_1, \ldots, X_n$ as well as $Y_{\text{saq}}$, which is a function of these $n$ observations, and the entire expression represents the expected in-sample maximized profit objective. In contrast, in the right expression of Equation (7), there are $n+1$ i.i.d. samples, namely $X_1, \ldots, X_n$ (folded into the definition of $Y_{\text{saq}}$), plus $X$, which represents future demand, and the term $E_F (p \min \{X, Y_{\text{saq}}\} - cY_{\text{saq}})$ represents the expected out-of-sample expected profit exactly as though it had been computed from an infinite virtual holdout sample (of both $X_i$ and $X$); given the special structure of the Newsvendor Model, we can derive this result without needing an actual holdout sample.

In addition to identifying an estimation bias, we also show how to correct for it. In particular, in Theorem 2 we show that a data-driven adjustment results in $\frac{1}{n} \sum_{i=1}^{n} (p \min \{x_i, y_{\text{saq}}\} - cy_{\text{saq}})$ — adjustment being an asymptotically unbiased estimator; in other words, we prove that

$$\pi_{\text{naive}} - E_F (\text{adjustment}) = \pi_{\text{true}} + o(1/n).$$

Finally, we emphasize that the true expected profit $\pi_{\text{true}} = E_F (p \min \{X, Y_{\text{saq}}\} - cY_{\text{saq}})$ is not computable, since $F$ is not known. However, the adjusted estimate $\frac{1}{n} \sum_{i=1}^{n} (p \min \{x_i, y_{\text{saq}}\} - cy_{\text{saq}})$ — adjustment is computable, as it is based entirely on the $n$ data observations $x_1, \ldots, x_n$.

3. The Naive Estimation of Expected Profit Exhibits Statistical Bias

In this section, we first describe the form of the bias in Section 3.1. We then derive the bias in Section 3.2. In Section 3.3 we use the bias expression to adjust the naive expected profit formula, so that it is asymptotically unbiased. Finally, in Section 3.4, we provide an intuitive interpretation of the bias.
3.1. The Form of the Bias

The profit bias can be written as

$$Bias_{profit} \triangleq \pi_{naive} - \pi_{true} = E_F \left( \frac{1}{n} \sum_{i=1}^{n} (p \min\{X_i, Y_{saa}\} - cY_{saa}) \right) - E_F (p \min\{X, Y_{saa}\} - cY_{saa}).$$

For simplicity, we cancel the $E_F (cY_{saa})$ term and $p$ multiplier, and we instead focus on the sales bias

$$Bias_{sales} \triangleq E_F \left( \frac{1}{n} \sum_{i=1}^{n} \min\{X_i, Y_{saa}\} \right) - E_F (\min\{X, Y_{saa}\}),$$

where $Bias_{profit} = pBias_{sales}$. We first define the sales from the naive profit expression in Equation (6) as $\tilde{S} \triangleq \frac{1}{n} \sum_{i=1}^{n} \min\{X_i, Y_{saa}\}$, whose expectation we decompose as follows:

$$E_F(\tilde{S}) = E_F \left( \frac{1}{n} \sum_{i=1}^{n} \min\{X_i, Y_{saa}\} \right) = E_F \left( \frac{1}{n} \sum_{i=1}^{n} X_{(i)} + \left( \frac{n+1-k}{n} \right) Y_{saa} \right) = \left( \frac{k-1}{n} \right) E_F \left( \frac{1}{k-1} \sum_{i=1}^{k-1} X_{(i)} \right) + \left( \frac{n+1-k}{n} \right) E_F (Y_{saa}),$$

(8)

where the second equality is due to $Y_{saa} = X_{(k)}$ and the final expression follows directly; the final expression is convenient for comparing with the sales expression from the exact expected profit.

We next define the sales from the exact profit expression in Equation (5) as $S \triangleq \min\{X, Y_{saa}\}$, whose expectation we decompose as follows:

$$E_F(S) = E_F (\min\{X, Y_{saa}\}) = P(X < Y_{saa}) E(X|X < Y_{saa}) + P(X > Y_{saa}) E(Y_{saa}|X > Y_{saa}) = \left( \frac{k}{n+1} \right) E(X|X < Y_{saa}) + \left( \frac{n+1-k}{n+1} \right) E(Y_{saa}|X > Y_{saa}),$$

(10)

where the second equality is due to the law of total expectation and the final equality is due to properties of order statistics, as shown in Lemma 2 (presented later in Section 3.2.1).

Examining the expressions in Equations (9) and (10), we observe three discrepancies, which may be considered the three potential sources of bias. In particular,

1. The probabilistic multipliers in Equation (9) are $\frac{k-1}{n}$ and $\frac{n+1-k}{n}$, whereas in Equation (10) they are $\frac{k}{n+1}$ and $\frac{n+1-k}{n+1}$, respectively.

2. The first expectation in Equation (9) is the unconditional $E_F \left( \frac{1}{k-1} \sum_{i=1}^{k-1} X_{(i)} \right)$, whereas in Equation (10) it is the conditional $E(X|X < Y_{saa})$. 
3. The second expectation in Equation (9) is the unconditional $E_F(Y_{saa})$, whereas in Equation (10) it is the conditional $E(Y_{saa} \mid X > Y_{saa})$.

The cumulative effect of these discrepancies can be seen by examining the sales bias, which can be written as $Bias_{sales} = E_F(\tilde{S} - S)$. Our first result characterizes this bias asymptotically, with an $o(1/n)$ error term, and we see that it is a positive bias; i.e., the naive expected profit estimate is larger than and overestimates the true expected profit, on average. The following theorem corrects for the above discrepancies, and follows immediately from Propositions 1 and 2 in the next subsection.

**Theorem 1.** The sales bias can be written as

$$Bias_{sales} = \frac{1}{nf(y^*)} \left( \frac{c}{p} \right) \left( 1 - \frac{c}{p} \right) + o \left( \frac{1}{n} \right),$$

where $y^* = F^{-1}(1 - c/p)$, and the profit bias can be written as

$$Bias_{profit} = \frac{c}{nf(y^*)} \left( 1 - \frac{c}{p} \right) + o \left( \frac{1}{n} \right).$$

### 3.2. The Derivation of the Bias

In the following results, and proofs thereof (in the appendix), we suppress the $saa$ subscript for expository clarity, so that $Y = Y_{saa}$. We next decompose the true sales $S$ into the sales that occur under the events of overage and underage, respectively:

$$S = \min\{X, Y\} = XI_{X < Y} + YI_{X > Y} = S_{over} + S_{under},$$

where $S_{over} \triangleq XI_{X < Y}$ and $S_{under} \triangleq YI_{X > Y}$. In particular, $S_{over}$ is the quantity $X$ of items sold if overage occurs, and is zero otherwise, and $S_{under}$ is the quantity $Y$ of items sold if underage occurs, and is zero otherwise. Similarly, we may decompose $\tilde{S} = \frac{1}{n} \sum_{i=1}^{n} \min\{X_i, Y\}$ into $\tilde{S} = \tilde{S}_{over} + \tilde{S}_{under}$, where $\tilde{S}_{over} \triangleq \frac{1}{n} \sum_{i=1}^{n-k} X_i$ and $\tilde{S}_{under} \triangleq (\frac{n+1-k}{n}) Y$; see Equation (8). We also introduce the function $h(y) \triangleq \int_{0}^{y} xf(x)dx = E(XI_{X < y})$, which features prominently in our analyses. In particular, $E(h(Y)) = E(S_{over})$ will be proven in Lemma 3.

Using these decompositions, we divide the evaluation of the sales bias into two subproblems:

$$Bias_{sales} = E_F(\tilde{S} - S)$$

$$= E_F(\tilde{S}_{over} - S_{over}) + E_F(\tilde{S}_{under} - S_{under}),$$

where $E_F(\tilde{S}_{over} - S_{over})$ is the sales bias due to overage and $E_F(\tilde{S}_{under} - S_{under})$ is the sales bias due to underage.
3.2.1. Auxiliary Results to Prove Theorem 1. In this section, we provide a series of auxiliary results that are needed to prove our main propositions, that in turn are used to prove Theorem 1. We make use of the representations $X_i = F^{-1}(U_i)$, $X = F^{-1}(U)$, $X(i) = F^{-1}(U(i))$, $Y = X(k) = F^{-1}(U(k))$, and $F(Y) = U(k)$ where $F^{-1}$ denotes the inverse function of $F$, $U_1, ..., U_n$ is an i.i.d. sample from the uniform distribution on the interval $(0,1)$, and $U$ is independent of $U_1, ..., U_n$ from the same uniform distribution.

Our proof techniques use Taylor series expansions extensively, both for random variables as well as deterministic functions (e.g., expectations). When analyzing deterministic functions, we characterize the finite expansions’ errors precisely using asymptotic notation (e.g., $O(1/n)$ or $o(1/n)$). When expanding random variables, we use $\approx$ to represent a Taylor Series expansion up to a specified degree; however, once we take expectations of these random variable expansions, we again characterize their errors analytically.

The diagram in Figure 1 presents a graph theoretic perspective of the dependencies between the various lemmas (labeled 'L') and propositions (labeled 'P'), that ultimately lead to proving Theorem 1, one of the main contributions of our paper. In particular, a directed arrow from, say, node L2 to L4 signifies that Lemma 2 is used to prove Lemma 4. Propositions 1-2 characterize the sales bias under the cases of overage and underage, respectively. These two propositions then lead naturally to Theorem 1, which characterizes the the overall profit bias.

![Diagram](image-url)

**Figure 1** Diagram describing the dependencies among lemmas, propositions, and Theorem 1.
Lemma 2 applies concepts from order statistics to characterize many useful quantities about the SAA order quantity $Y = X_{(k)}$ and, as is evident from Figure 1, influences many of the subsequent lemmas that are needed to prove Theorem 1.

**Lemma 2.** Here are some basic facts about the order quantity $Y = X_{(k)}$ and the independent future demand observation $X$.

- The probability of underage $P(X > Y) = (n + 1 - k) / (n + 1)$.
- The probability of overage $P(X < Y) = k / (n + 1)$.
- $E[F(Y)] = k / (n + 1)$.
- $Var[F(Y)] = Var(U(k)) = k(n + 1 - k) / [(n + 1)^2(n + 2)] = O\left(\frac{1}{n}\right)$.
- $E(Y) = F^{-1}\left(\frac{k}{n + 1}\right) = \frac{f'\left[F^{-1}\left(k / (n + 1)\right)\right]}{2f^3\left[F^{-1}\left(k / (n + 1)\right)\right]} \frac{k(n + 1 - k)}{(n + 1)^2(n + 2)} + o\left(\frac{1}{n}\right)$.
- $Var(Y) = \frac{k(n + 1 - k)}{\{f[E(Y)]\}^2(n + 1)^2(n + 2)} + o\left(\frac{1}{n}\right) = O\left(\frac{1}{n}\right)$.

Lemma 3 characterizes some expansions of $h(Y) = \int_0^Y xf(x)dx = E(X_{I_{X < Y}})$, where $E(h(Y)) = E(S_{over})$. The first expansion is of a random variable, and we do not characterize the random error term, but the second two expressions (which use the first) are expectations and we characterize the error terms precisely.

**Lemma 3.** Here are some basic facts about $h(y) \triangleq E(X_{I_{X < y}}) = \int_0^y xf(x)dx$.

- To first order, $h(Y) \approx h[E(Y)] + E(Y)\{F(Y) - F[E(Y)]\}$.
- $Cov[h(Y), F(Y)] = E(Y)Var[F(Y)] + o\left(\frac{1}{n}\right)$.
- $E\left(\frac{h(Y)}{F(Y)}\right) = \left(1 + \frac{Var[F(Y)]}{E[F(Y)]}\right) E[h(Y)] - \frac{Var[F(Y)]}{E[F(Y)]} E(Y) + o\left(\frac{1}{n}\right)$.
- $E[h(Y)] = E(S_{over})$.

The next three lemmas use basic algebraic manipulations to simplify some expressions that are functions of $k = \lceil n(1 - c/p) \rceil$.

**Lemma 4.**

\[
1 + \frac{Var[F(Y)]}{\{E[F(Y)]\}^2} - \frac{nE[F(Y)]}{k - 1} = O\left(\frac{1}{n^2}\right)
\]

**Lemma 5.**

\[
\frac{k}{n} = 1 - \frac{c}{p} + O\left(\frac{1}{n}\right), \quad \frac{k}{n + 1} = 1 - \frac{c}{p} + O\left(\frac{1}{n}\right), \quad \text{and} \quad \frac{n + 1 - k}{n} - \frac{n + 1 - k}{n + 1} = \frac{n + 1 - k}{n(n + 1)} = \frac{1}{n} \left(\frac{c}{p}\right) + O\left(\frac{1}{n^2}\right).
\]

**Lemma 6.**

\[
\frac{k - 1}{n} \frac{Var[F(Y)]}{\{E[F(Y)]\}^2} = \frac{1}{n} \left(\frac{c}{p}\right) + O\left(\frac{1}{n^2}\right)
\]
Lemma 7 derives the conditional densities of the SAA order quantity $Y$, given the events of overage and underage, respectively, along with the unconditional density and CDF of $Y$. These results are used to characterize the conditional expected sales when overage occurs, which is used in Lemma 8 which, in turn, is used in the proof of Proposition 1. These conditional densities of the SAA order quantity may also be of independent interest; hence we also include the underage case.

**Lemma 7.** The conditional densities of $Y$ given overage and given underage are:

- The conditional density of $Y$ given overage $X < Y$ is $g_{X,Y}(y) = \frac{F(y) g(y)}{P(X < Y)}$.
- The conditional density of $Y$ given underage $X > Y$ is $g_{X,Y}(y) = \frac{[1 - F(y)] g(y)}{P(X > Y)}$.
- The unconditional density $g$ of $Y$ is $g(y) = \frac{n!}{(k-1)!(n-k)!} [F(y)]^{k-1}[1 - F(y)]^{n-k} f(y)$.
- The unconditional CDF $G(y)$ of $Y$ is given by the incomplete beta function with parameters $k$ and $n+1-k$ evaluated at $F(y)$.

where $g$ denotes the unconditional density of $Y$.

Lemma 8 analyzes expressions that are related to the differences between Equations (9) and (10), the naive and true profit estimations discussed in Section 3.1. Per Figure 1, Lemma 8 is the final auxiliary result needed to prove the overage sales bias result in Proposition 1 (presented in the next section).

**Lemma 8.** The following are relevant to understanding the case of overage:

- $E\left( \frac{1}{k-1} \sum_{i=1}^{k-1} X_{(i)} \mid Y = y \right) = E(X \mid X < y) = \frac{h(y)}{F(y)}$ is an increasing function of $y$.
- $E\left( \frac{1}{k-1} \sum_{i=1}^{k-1} X_{(i)} \right) = E\left( \frac{h(Y)}{F(Y)} \right) < E(X \mid X < Y) = \frac{E[h(Y)]}{E[F(Y)]}$.
- $\frac{k}{n+1} E\left( \frac{1}{k-1} \sum_{i=1}^{k-1} X_{(i)} \right) < E[X I_{X < Y}] = E[h(Y)] = E(S_{over})$.

Lemma 9 builds upon Lemma 2 and provides alternate expressions for certain functions of $Y$ needed for deriving the underage sales bias result in Proposition 2 (presented in the next section).

**Lemma 9.** The following are additional facts about $F(Y)$:

- $E[F(Y)] - \frac{k-1}{n} = \frac{1}{n} \left( \frac{c}{p} \right) + O\left( \frac{1}{n^2} \right)$.
- $\text{Var}[F(Y)] = \frac{1}{n} \left( \frac{c}{p} \right) \left( 1 - \frac{c}{p} \right) + O\left( \frac{1}{n^2} \right)$.
- $E\left( \{|F(Y) - F[E(Y)]|^2\} \right) = \text{Var}[F(Y)] + O\left( \frac{1}{n^2} \right)$.

Lemma 10 is the auxiliary result that identifies the form of the profit bias, which interestingly equals (asymptotically) the covariance between the SAA order quantity $Y$ and its percentile $F(Y)$. Note that, per Figure 1, Lemma 10 is only used to prove the underage sales bias result in Proposition 2 and we learn that the overall profit bias is driven primarily by the underage sales bias; more discussion on this observation is provided in the next section.
Lemma 10. \( \text{Cov}[Y, F(Y)] = \frac{1}{nf[E(Y)]} \left( \frac{c}{p} \right) \left( 1 - \frac{c}{p} \right) + o\left( \frac{1}{n} \right). \)

Lemma 11 is the final auxiliary result needed to prove Proposition 2 and demonstrates that the SAA order quantity \( Y \) is an (asymptotically) unbiased and consistent estimator of the true optimal order quantity \( y^* = F^{-1}(1 - c/p) \).

Lemma 11. Considering the true (unknown) optimal order quantity \( y^* = F^{-1}(1 - c/p) \), we have \( E(Y) \rightarrow y^* \), and that \( Y \) is a consistent estimator of \( y^* \).

3.2.2. Main Propositions. The following two propositions provide the formulas for the asymptotic biases of the two components of the sales bias.

Proposition 1. The sales bias in overage is
\[
E(\hat{S}_{\text{over}}) - E(S_{\text{over}}) = -\frac{1}{n} \left( \frac{c}{p} \right) E(Y) + o\left( \frac{1}{n} \right).
\]

Proposition 2. The sales bias in underage is
\[
E(\hat{S}_{\text{under}}) - E(S_{\text{under}}) = \frac{1}{n} \left( \frac{c}{p} \right) E(Y) + \frac{1}{nf(y^*)} \left( \frac{c}{p} \right) \left( 1 - \frac{c}{p} \right) + o\left( \frac{1}{n} \right).
\]

Together, Propositions 1-2 imply Theorem 1, due to Equation (11). Interestingly, the \( \frac{1}{n} \left( \frac{c}{p} \right) E(Y) \) term is common to the bias expressions in both the overage and underage cases, which cancel, and the overall sales bias is driven solely by the additional term in the underage sales bias. Fortunately, after this cancellation, all terms in the bias expression are known, except \( f(y^*) \), which we address in the next section.

3.3. Correcting for the Bias

The sales bias \( \text{Bias}_{\text{sales}} = \frac{1}{nf(y^*)} \left( \frac{c}{p} \right) \left( 1 - \frac{c}{p} \right) + o\left( \frac{1}{n} \right) \) from Theorem 1 may be asymptotically unbiasedly estimated using
\[
\widehat{\text{Bias}}_{\text{sales}} = \frac{1}{nf} \left( \frac{c}{p} \right) \left( 1 - \frac{c}{p} \right),
\]
where \( \hat{f} \) is a consistent estimate of the density \( f \) at \( E(Y) \). To estimate the density term \( f(y^*) \), we use the fact that \( f = F' \) to allow us to take a numerical derivative of the empirical CDF using two order statistics with one on each side of \( Y = X_{(k)} \). Choosing order statistics symmetrically placed at ranks \( k - m \) and \( k + m \) with integer \( m > 0 \), we note that the empirical CDF increases by \( 2m/n \) as the density’s argument increases from \( x_{(k-m)} \) to \( x_{(k+m)} \). We therefore define our estimate as
\[
\hat{f} \triangleq \frac{2m}{n \left( x_{(k+m)} - x_{(k-m)} \right)}, \tag{12}
\]
To see that $1/\hat{f}$ is a consistent estimate of $1/f(y^*)$ in the bias equation, we rely upon Equations (8) and (9) and Theorem 1 of Tüsády (1974), where Tüsády’s $k$ corresponds to our $2m$. Thus $\hat{f}$ is a consistent estimator of $f(y^*)$ because the density is being estimated at $Y = X(k)$ with (from Lemma 11) $E(Y) \rightarrow y^*$ and (from Lemma 2) $Var(Y) = O(1/n)$. That a continuous function (in this case, the reciprocal) of a consistent estimator is itself consistent follows, e.g., from the Theorem on page 24 of Serfling (1980). Thus we also have that $1/\hat{f}$ is a consistent estimator of $1/f(y^*)$ as required. Tüsády’s work was extended by Barabás (1987), who shows on page 122 that, asymptotically, the best choice for $m$ is $m \sim C n^{2/3}$, which represents a compromise between a larger $m$ with less density-estimation variability and a smaller $m$ with less bias; in particular, with this choice for $m$ we have both $m \rightarrow \infty$ (so that many data points fall within the interval from $X(k-m)$ to $X(k+m)$) and $m/n \rightarrow 0$ (so that the interval tends to a single point). These results show that the bias estimate is asymptotically unbiased, as stated in the following proposition.

**Proposition 3.** $E[\hat{Bias}_{sales}] = Bias_{sales} + o\left(\frac{1}{n}\right)$.

Using these results, we can create an adjusted sales estimate that is asymptotically unbiased. In particular, referring to Equation (3), we define

$$\hat{S}_{adjusted} = \frac{1}{n} \sum_{i=1}^{n} \min\{x_i, y_{saa}\} - \hat{Bias}_{sales},$$

where $x_i, i = 1, \ldots, n$, is the realized sample of demand and $y_{saa}$ is the realized SAA order quantity. Similarly, an adjusted profit estimate is defined as

$$\hat{\pi}_{adjusted} = \frac{1}{n} \sum_{i=1}^{n} (p \min\{x_i, y_{saa}\} - c y_{saa}) - p \hat{Bias}_{sales}. \quad (13)$$

Since we demonstrated in Section 3.1 that the profit bias is driven solely by the sales bias, Proposition 3 implies that the adjusted profit estimate in Equation (13) is also asymptotically unbiased, which we present in the following theorem.

**Theorem 2.**

$$E(\hat{\pi}_{adjusted}) = \pi_{true} + o(1/n).$$

### 3.4. Interpretation of the Bias

In this subsection, we interpret the bias and show that its dominant source is underage. We begin by showing that the canceling terms $\pm \frac{1}{n} \left( \frac{c}{p} \right) E(Y)$ (negative for the bias in overage from Proposition 1, positive in underage from Proposition 2) are artifacts due to an error of estimation of the probabilities of overage and underage as computed within-sample by the naive Newsvendor. After
correcting for this estimation error, we show that the asymptotic bias from Theorem 1 is due entirely to underage.

We then interpret this bias in two ways: First, we understand the bias as a mathematical consequence of the sign of the covariance of the SAA order quantity $Y$ with a monotonic function of itself $F(Y)$. Second, we use visualization to demonstrate the consequences to the Newsvendor of a varying conditional probability of underage (as experienced by the Newsvendor whose judgment is based on the order quantity from their sample) that interacts with the order quantity itself in such a way that larger (random) realizations of $Y$ are associated with smaller conditional probabilities of underage (because underage requires the future demand $X$ to cross a larger hurdle). While this conditional probability (of underage given $Y$) is not observable at the time by the Newsvendor, we have nonetheless corrected for its overall bias.

3.4.1. Bias is Due to Underage Only. The naive sales estimate in underage is $\bar{S}_{\text{under}} = \left( \frac{n+1-k}{n} \right) Y$ (c.f., the beginning of Section 3.2), which may be interpreted as the within-sample estimated probability of underage, $\frac{n+1-k}{n}$, times the sales $Y$ that would occur with this order quantity when underage occurs. However, the true probability of underage is actually the slightly smaller $\frac{n+1-k}{n+1}$ from Lemma 2. We correct this estimation error by defining the probability-adjusted naive sales estimate in underage given $Y$ to be an $O(1/n)$ perturbation of $\bar{S}_{\text{under}}$ formed by replacing the within-sample estimate with the true probability as follows:

$$\bar{S}_{\text{under},\text{ProbAdj}} \triangleq \left( \frac{n+1-k}{n+1} \right) Y = \bar{S}_{\text{under}} - \left( \frac{n+1-k}{n(n+1)} \right) Y.$$  

The effect of this adjustment on the bias in underage may be computed as follows:

$$E \left( \bar{S}_{\text{under},\text{ProbAdj}} \right) - E \left( S_{\text{under}} \right) = E \left( \bar{S}_{\text{under},\text{ProbAdj}} \right) - E \left( \bar{S}_{\text{under}} \right) + E \left( \bar{S}_{\text{under}} \right) - E \left( S_{\text{under}} \right)$$

$$= \left( \frac{n+1-k}{n+1} - \frac{n+1-k}{n} \right) E(Y) + E \left( \bar{S}_{\text{under}} \right) - E \left( S_{\text{under}} \right).$$  

Using the third assertion of Lemma 5 for the first term and Proposition 2 for the second term, we find the bias of the probability-adjusted sales in underage to be

$$E \left( \bar{S}_{\text{under},\text{ProbAdj}} \right) - E \left( S_{\text{under}} \right)$$

$$= - \left[ \frac{1}{n} \left( \frac{c}{p} \right) + O \left( \frac{1}{n^2} \right) \right] E(Y) + \left[ \frac{1}{n} \left( \frac{c}{p} \right) E(Y) + \frac{1}{n f(y^*)} \left( \frac{c}{p} \right) \left( 1 - \frac{c}{p} \right) + o \left( \frac{1}{n} \right) \right]$$

$$= \frac{1}{n f(y^*)} \left( \frac{c}{p} \right) \left( 1 - \frac{c}{p} \right) + o \left( \frac{1}{n} \right),$$

which we recognize as the full sales bias from Theorem 1 (which combines underage and overage biases). The probability adjustment has eliminated the canceling term $\pm \frac{1}{n} \left( \frac{c}{p} \right) E(Y)$ from the bias in underage.
To preserve the naive estimate in totality, having subtracted \( \left( \frac{n+1-k}{n(n+1)} \right) Y \) from \( \tilde{S}_{under} \) to obtain \( \tilde{S}_{under,ProbAdj} \), we must add \( \left( \frac{n+1-k}{n(n+1)} \right) Y \) to \( \tilde{S}_{over} \) to obtain the probability-adjusted naive sales estimate in overage given \( Y \) as

\[
\tilde{S}_{over,ProbAdj} \triangleq \tilde{S}_{over} + \left( \frac{n+1-k}{n(n+1)} \right) Y = \frac{1}{n} \sum_{i=1}^{k-1} X(i) + \frac{n+1-k}{n(n+1)} X(k),
\]

which is not (strictly speaking) a pure probability adjustment but, instead, is the required adjustment implied by the probability adjustment made in underage. The resulting term \( \tilde{S}_{over,ProbAdj} \) may be interpreted as adding a small fraction of \( X(k) \) to the overstock estimate \( \tilde{S}_{over} \). Adding this small fraction of order \( O(1/n) \) greatly simplifies the bias interpretation because we now have asymptotic bias zero, plus \( o(1/n) \), in overage. To see this, we use the third assertion of Lemma 5 for the first term and Proposition 1 for the second term, and find the bias of the probability-adjusted sales in overage to be:

\[
E \left( \tilde{S}_{over,ProbAdj} \right) - E (\tilde{S}_{over}) = \left( \frac{n+1-k}{n(n+1)} \right) E (Y) + \left[ E \left( \tilde{S}_{over} \right) - E (\tilde{S}_{over}) \right] = \frac{1}{n} \left( \frac{c}{p} \right) + O \left( \frac{1}{n^2} \right) E (Y) + \left[ - \frac{1}{n} \left( \frac{c}{p} \right) E (Y) + o \left( \frac{1}{n} \right) \right] = o \left( \frac{1}{n} \right).
\]

Thus the dominant bias is due entirely to underage after the probability adjustment:

\[
E(\tilde{S}) - E(S) = \frac{1}{nf(y^*)} \left( \frac{c}{p} \right) \left( 1 - \frac{c}{p} \right) + o \left( \frac{1}{n} \right) = E \left( \tilde{S}_{under,ProbAdj} \right) - E (\tilde{S}_{under}).
\]

3.4.2. Interpreting the Bias as a Covariance. Our first interpretation of the bias is as a mathematical consequence of the sign of the covariance \( \text{Cov} [Y, F(Y)] > 0 \) of the SAA order quantity \( Y \) with \( F(Y) \), which is a monotonically increasing function of its argument \( Y \). This covariance emerges from using iterated expectations, first conditioning on \( Y \), then using the definition of covariance, to obtain the true expected sales in underage as follows:

\[
E (\tilde{S}_{under}) = E (Y I_{X>Y}) = E [E (Y I_{X>Y} | Y)] = E [Y E (I_{X>Y} | Y)] = E \{ Y [1 - F(Y)] \} = E [1 - F(Y)] E (Y) + \text{Cov} \{ Y, [1 - F(Y)] \}.
\]

Next, using Lemma 2 for \( E [F(Y)] \) and recognizing that additive constants within a covariance are irrelevant we find \( E (\tilde{S}_{under}) = \left( \frac{n+1-k}{n+1} \right) E (Y) - \text{Cov} [Y, F(Y)] = E \left( \tilde{S}_{under,ProbAdj} \right) - \text{Cov} [Y, F(Y)], \)
which implies that the bias \( E(\bar{S}_{\text{under,ProbAdj}}) - E(S_{\text{under}}) = \text{Cov}[Y, F(Y)] \) is positive because \( F(Y) \) is monotonically increasing in \( Y \). While this interpretation is clear, direct, and true, it is less than fully satisfying. Our second interpretation will establish the intuition of why the bias must occur as a consequence of an error made by the Newsvendor.

**3.4.3. Visualizing the Bias.** Our second interpretation of the bias begins by taking the point of view of the Newsvendor in order to understand why \( \bar{S}_{\text{under,ProbAdj}} \) is lacking the subtracted covariance term \( \text{Cov}[Y, F(Y)] \) found in \( S_{\text{under}} \) and is therefore biased upwards by this amount. Having observed \( Y = y \), the (probability-adjusted) Newsvendor mistakenly applies the unconditional probability of underage \( E[1 - F(Y)] = \frac{n+1-k}{n+1} \) to the amount \( y \) sold in underage, obtaining \( \bar{S}_{\text{under,ProbAdj}} = \left( \frac{n+1-k}{n+1} \right) y \). This is the error that leads to the bias because the actual (true) probability of underage faced by the Newsvendor in this situation is the conditional probability of underage \( 1 - F(y) = P(X > Y | Y = y) \), leading to a true forecast of sales in underage (given this order quantity \( y \)) of \( \bar{S}_{\text{under,ProbAdj,Actual}} = [1 - F(y)] y \) which is, unfortunately, unavailable to the Newsvendor, who does not know whether the estimated order quantity \( y \) is particularly high or low within its distribution \( g(y) \). If, by random chance, the Newsvendor observes a large value of \( y \), this conditional probability will be smaller than the unconditional probability used by the Newsvendor. Just when the Newsvendor believes they will sell a large quantity in underage, the (conditional) probability of underage drops and deflates the expected sales in underage (unbeknownst to the Newsvendor). While we cannot tell the Newsvendor that their \( y \) is high or low, we are nonetheless able to correct for the (overall) bias implied by this variation. The reader might counter that if the Newsvendor observes a small value of \( y \), then the conditional probability of underage is greater than its unconditional probability; while this is true, this larger probability goes with a smaller value, leading to dominance of the larger values over the smaller values, and a positive bias. We demonstrate this next.

To visualize the source of the bias we give an example of two samples, one with a low order quantity (and a negative conditional bias) and the other with a high order quantity (and a positive conditional bias). The sales bias is the expected value of these (and many other) conditional biases, and we know that the positive bias wins out as explained theoretically above. This section shows visually where the disconnect between naïve and true estimated profit originates.

We chose our two samples so that the order quantities were close to the 25th and 75th percentiles of the order quantity distribution \( g(y) \) after sorting 1,000 simulated random samples by the order quantity (the true order quantity 25th percentile of 76.67422 is closely matched to the Sample 1 order quantity of \( y_1 = 76.67222 \); for the 75th percentile these numbers are 118.69706 and \( y_2 = 118.69667 \), respectively for Sample 2). The population is exponential with mean 200, the sample
size is $n = 25$, with $c = 3$ and $p = 5$ so that $k = 10$ and $k/n = 1 - c/p = 0.40$. Figure 2 shows the true exponential CDF (with the true optimal order quantity $y^* = 102.17$) along with the two sample CDFs $\hat{F}_1$ and $\hat{F}_2$ (with their corresponding order quantities low $y_1$ and high $y_2$).

![CDFs: True F and Two Sampled Empirical $\hat{F}$](image)

**Figure 2** The true exponential CDF, $F$, (with mean 200 and true optimal order quantity $y^* = 102.17$) along with the two sample CDFs: $\hat{F}_1$ (with low order quantity $y_1 = 76.67$) and $\hat{F}_2$ (with high order quantity $y_2 = 118.70$), chosen as the 25th and 75th percentiles of the distribution $g$ of $Y$. Note that $k/n$ is not the true (unconditional) probability of underage, which instead is $k/(n + 1)$ from Lemma 2.

Subsequent figures use the probability on the vertical axis along with the demand amount on the horizontal axis to create rectangles whose area represents the (conditional) expected sales in underage as probability times sales $y$, from which we immediately see that it is the different probabilities (unconditional for the naive Newsvendor, conditional for the truth) that lead to their difference, as the multiplication by $y$ is the same in both cases.

Sample 1, with its low order quantity $y_1$, exhibits negative conditional bias as shown visually in Figures 3 and 4. The shaded area $y_1 (n + 1 - k) / (n + 1)$ in Figure 3 (representing the probability-adjusted naive sales estimate in underage given $y_1$) is smaller than the shaded area $y_1 [1 - F(y_1)]$ in Figure 4 (representing the true expected sales in underage given $y_1$). This indicates a negative conditional sales bias when low order quantities occur.

Sample 2, with its high order quantity $y_2$, exhibits positive conditional bias as shown visually in Figures 5 and 6. The shaded area $y_2 (n + 1 - k) / (n + 1)$ in Figure 5 (representing probability-adjusted naive sales estimate in underage given $y_2$) is larger than the shaded area $y_2 [1 - F(y_2)]$ in Figure 6 (representing the true expected sales in underage given $y_2$). This indicates a positive conditional sales bias when high order quantities occur.

To visualize that the bias ends up positive, with large order quantities having a positive conditional bias and small order quantities having a negative conditional bias, we begin by noting that
the conditional bias in underage is \( y \left[ F(y) - k / (n + 1) \right] \) as may be seen by subtracting the formula for the shaded area of Figure 6 from that of Figure 5. Moreover, with the probability adjustment we know that the bias in overage is negligible. Using iterated expectations, the overall bias must
be the expectation over $Y$ of its conditional bias in underage $Y \left[ F(Y) - k/(n+1) \right]$, so we have

$$
Bias_{sales} = E \left[ Y \left( F(Y) - k/(n+1) \right) \right] = \int_{0}^{\infty} y \left[ F(y) - k/(n+1) \right] g(y) \, dy,
$$

where we use the density $g(y)$ of $Y$ from Lemma 7 to evaluate the expectation. We recognize this integral form of the bias as the area under the curve of the integrand, and this curve is the weighted conditional bias in underage $y \left[ F(y) - k/(n+1) \right] g(y)$. For the current example with $F$
being exponential with mean 200, this density is

\[
g(y) = \frac{n!}{(k-1)!(n-k)!}[F(y)]^{k-1}[1-F(y)]^{n-k}f(y) = 163,438(1-e^{-y/200})^9(e^{-y/200})^{16}.
\]

Figure 7 shows the conditional bias in underage \(y[F(y) - k/(n+1)]\) (dark curve with scale at left) and we note that this bias is positive for large \(y\), specifically those greater than \(F^{-1}[1-k/(n+1)] = 97.10\), and negative for those less. We note that this dividing point is not the same as the true optimal order quantity \(y^* = F^{-1}(1-k/n) = 102.17\) due to the probability adjustment. Also shown is the density \(g(y)\) of \(Y\) (dashed curve, scale not shown). The product of the conditional bias with the density gives us \(y[F(y) - k/(n+1)]g(y)\), which is the weighted conditional bias (scale not shown) whose net area (positive net of negative) is the bias in underage which we know is the bias itself. It is clear in Figure 7 that the positive area dominates, which we know it must. Also shown in Figure 7 are the order quantities \(y_1\) and \(y_2\) at the 25th and 75th percentiles of the distribution \(g\) of \(Y\), from the two random samples of Figures 2 - 6.

![Figure 7](image-url)

**Figure 7** The bias is the net shaded area under the weighted conditional bias curve, and it is clear that the positive area dominates, as we know it must. The population \(F\) is exponential with mean 200. The weighted conditional bias is formed as the product of the conditional bias with the density \(g\) of \(Y\), so that the expected conditional bias is the ordinary integral (and hence the area) of the weighted conditional bias.

4. **Exact Results for an Exponential Distribution of Demand**

In this section we derive exact formulas for the expected true profit, the expected naive profit, and the bias, when the SAA order quantity is used under the assumption that \(F\) is an exponential
distribution with mean $\mu$. The following proposition provides the closed-form expressions for these quantities.

**Proposition 4.** Exact expressions for the expected true profit $\pi_{true}$, the expected naive profit $\pi_{naive}$, and the bias $\text{Bias}_{\text{profit}}$, when $F$ is an exponential distribution with mean $\mu$, are as follows:

$$
\pi_{true} = \mu \left[ \frac{pk}{n+1} - c \sum_{i=n+1-k}^{n} \left( \frac{1}{i} \right) \right],
$$

$$
\pi_{naive} = \mu \left[ \frac{pk}{n} - c \sum_{i=n+1-k}^{n} \left( \frac{1}{i} \right) \right],
$$

$$
\text{Bias}_{\text{profit}} = \frac{pk\mu}{n(n+1)}.
$$

To prove Proposition 4, we first need the following lemma.

**Lemma 12.** When $F$ is an exponential distribution with mean $\mu$, the following expectations hold exactly:

- $E[\min(X,Y)] = \mu \left( \frac{k}{n+1} \right)$.
- $E(Y) = \mu \sum_{i=n+1-k}^{n} \left( \frac{1}{i} \right)$.
- $E \left( \sum_{i=1}^{k-1} X(i) \right) = \mu \left( k - (n+1-k) \sum_{i=n+1-k}^{n} \left( \frac{1}{i} \right) \right)$.

To compare the exact result $\text{Bias}_{\text{profit}} = \frac{pk\mu}{n(n+1)}$ for the exponential distribution to our general asymptotic result $\text{Bias}_{\text{profit}} = \frac{p}{nf(y^*)} \left( \frac{c}{p} \right) \left( 1 - \frac{c}{p} \right) + o \left( \frac{1}{n} \right)$, we now show that they are equal in the exponential case, although technically we have already proven this in the general case. Nonetheless, it is of interest to see how the reciprocal of the density produces the scale factor $\mu$ for the exact bias in the exponential case.

**Proposition 5.** The exact profit bias for the exponential distribution is equal to the general asymptotic profit bias formula in the sense that

$$
\text{Bias}_{\text{profit}} = \frac{pk\mu}{n(n+1)} = \frac{p}{nf(y^*)} \left( \frac{c}{p} \right) \left( 1 - \frac{c}{p} \right) + o \left( \frac{1}{n} \right).
$$

We conclude this section by evaluating the size of the bias, with respect to the true expected profit. In the left plot of Figure 8, we present $\text{Bias}_{\text{profit}}$ as a percentage of $\pi_{true}$, as a function of the sample size $n$, for $p = 5$ and $c = 3$; we obtain qualitatively similar results for different values of $p$ and $c$. We see that, especially for small values of $n$, the percentage is large, which then generally decreases as the sample size increases. In the right plot of Figure 8, we present the bias percentage
as a function of $c/p$ for $n = 25$; again, qualitatively similar plots are obtained for different values of $n$. Here, we see that the bias percentage increases with $c/p$. In both plots, the jagged nature of the lines is due to the value of $k$ changing (c.f., Lemma 1).

5. Numerical Experiments to Evaluate the Adjusted Profit’s Bias Reduction

Simulations were performed for the normal and lognormal demand cases, and considerable expected estimation error reduction was observed using our proposed adjustment in Equation (13), which empirically verifies Theorem 2 for these distributions.

Our experimental primitive is as follows: we generate a sample of $n = 25$ observations $x_i$ from the true distribution, from which we obtain the order size $y = x_{(k)}$. We then compute both the naive and the adjusted expected profit, as perceived by the Newsvendor for this sample. To improve the efficiency of the Monte Carlo simulations, we decrease the noise involved in computing the true profit as follows: Instead of generating a new observation $x$ from the true distribution and using $p \min(x, y) - cy$, we use analytical expressions for the conditional expectation of the true profit given $y$, namely $p E_F \left[ \min(X, y) \right] - cy,$ where $X$ is drawn from the true distribution and $y$ is held fixed at its value for this sample. For the normal distribution $X \sim N(\mu, \sigma^2)$, this conditional expectation is $p \left[ y - \sigma \varphi \left( \frac{y-\mu}{\sigma} \right) - (y-\mu) \Phi \left( \frac{y-\mu}{\sigma} \right) \right] - cy$, which may be derived using basic probability theory. Similarly, for the lognormal distribution $X = e^{\mu + \sigma Z}$, with standard normal $Z$, the conditional expectation is $p \left\{ e^{\mu+\sigma^2/2} \Phi \left( \frac{\ln y - \mu}{\sigma} - \sigma \right) + y \left[ 1 - \Phi \left( \frac{\ln y - \mu}{\sigma} \right) \right] \right\} - cy$.

The estimation error before adjustment is formed by subtracting the conditional expectation (of the true profit) from the naive estimate, where both were formed from this sample. Similarly, the estimation error after adjustment is formed by subtracting the conditional expectation (of the true profit) from the adjusted estimate, again, where both were formed from this sample. We compute
the t-statistic (testing against zero estimation error) for each measure (naive and adjusted) by repeating this procedure for 10,000 independent samples, each of size $n = 25$. The result is a pair of t-statistics for the expected estimation error: one before and one after adjustment. We then repeat this procedure 100 times, to obtain 100 pairs of t-statistics, each based on 10,000 simulations.

In Figure 9, we plot histograms of these 100 paired t-statistics for the normal (left) and lognormal (right) distributions. For both distributions we set the mean at 200 and the standard deviation at 65; in the normal case, we set $\mu = 200$ and $\sigma = 65$, while the lognormal case uses $\mu = 5.248112$ and $\sigma = 0.316877$ so that its mean will be $E(X) = e^{\mu + \sigma^2/2} = 200$ and its standard deviation will be $StDev(X) = e^{\mu + \sigma^2/2} \sqrt{e^{\sigma^2} - 1} = 65$. For estimating the density in Equation (12), we use $m = 2$, which is consistent with the $m \sim C n^{2/3}$ recommendation from Barabási (1987), for $n = 25$ and $C = 1/4$. In both cases, we set $p = 5$ and $c = 3$.

![Figure 9](image_url)

**Figure 9** Adjustment eliminates bias for normal and lognormal demand: normal distribution (left) and lognormal distribution (right) both have mean 200 and standard deviation 65, with $p = 5$, $c = 3$, and sample size $n = 25$.

We observe strong evidence that our asymptotic adjustment eliminates the statistically significant estimation error very effectively even in these finite samples. As shown in Figure 9, for both the normal and the lognormal distributions, the unadjusted estimation error shows high statistical significance (i.e., the t-values in the histograms before adjustment are considerably higher than the standard 1.96 critical value), which is successfully eliminated by the asymptotic correction (i.e., the histogram after adjustment is centered close to zero). For normal demand, the expected profit is 270.4, and the expected estimation error is 8.0, which is a nonnegligible 3.0% of the true expected profit. For lognormal demand, the expected profit is 280.9 and the expected estimation error is 6.8, which is 2.4% of the true expected profit.

We repeated these experiments with smaller values of $c$, which results in larger Newsvendor quantiles $1 - c/p$; these larger quantiles might have caused difficulties in estimating $f(y^*)$, but the experimental results (omitted for brevity) were qualitatively identical to those in Figure 9.
5.1. Comparison with Exact Parametric Results of Siegel and Wagner (2021)

In this section we compare the exact bias result we derived in Section 4 for the exponential distribution with that of Siegel and Wagner (2021). If demand is exponentially distributed with mean \( \mu \), then \( y^* = \mu \ln(p/c) \) maximizes the expected profit, which equals \( p\mu(1 - e^{-y^*/\mu}) - cy^* \). Siegel and Wagner (2021) considered a scenario where it is known that demand is exponentially distributed, but the mean \( \mu \) is not known. It is natural to estimate the mean (unbiasedly) as \( \hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} X_i \), where \( X_1, \ldots, X_n \) is the sample of data. Siegel and Wagner (2021) demonstrate that replacing \( \mu \) with its estimate \( \hat{\mu} \) in the order quantity and profit expressions, namely \( \hat{y} = \hat{\mu} \ln(p/c) \) and

\[
p\hat{\mu}(1 - e^{-\hat{y}/\hat{\mu}}) - c\hat{y},
\]

respectively, results in an exact profit bias of

\[
p\mu \left[ \left( \frac{n}{n + \ln(p/c)} \right)^n - c/p \right],
\]

which is estimated by replacing \( \mu \) with \( \hat{\mu} \); c.f., Equation (3) in Siegel and Wagner (2021). Recall that Proposition 4 derives the non-parametric profit bias as

\[
\frac{pk\mu}{n(n+1)},
\]

which is again estimated by replacing \( \mu \) with \( \hat{\mu} \). Subtracting these estimated biases from their respective expected profit expressions provides adjusted profit estimates.

Unfortunately, comparing Equations (15) and (16) does not lead to an easy analytical conclusion. Consequently, we demonstrate numerically that the profit estimation bias is smaller when there is parametric information available (i.e., the fact that demand is exponentially distributed). Intuitively, more information (the parametric form of the distribution) leads to better decision making, and hence less bias. However, despite the non-parametric environment (the focus of our paper) exhibiting more bias, our data-driven profit adjustment performs as well as that of Siegel and Wagner (2021). We demonstrate this next.

For an exponential distribution with a mean of 200, the naive expected profits for the SAA order quantity and that of Siegel and Wagner (2021) are 101.49 (0.35) and 93.48 (0.19), whereas the true expected profits are 86.21 (0.10) and 90.42 (0.04), respectively, which are estimated over 10,000 Monte Carlo trials; the values in parentheses are standard errors. The biases are 17.7% and 3.4%, respectively. The adjusted profits are 86.02 (0.35) and 90.35 (0.18), representing errors of -0.22% and -0.07%, respectively. Therefore, even though our non-parametric data-driven environment results in a much larger bias (17.7% versus 3.4%), our analytical adjustment is able to effectively eliminate the bias as well as in the parametric environment of Siegel and Wagner (2021).
We also mirror the hypothesis-testing experimental design described earlier in this section. In Figure 10 we present histograms of t-statistics for four expected profit expressions: 1) un-adjusted non-parametric profit, 2) un-adjusted parametric profit, 3) adjusted non-parametric profit, and 4) adjusted parametric profit. Notably, the non-adjusted non-parametric expected profit, from Equation (6), exhibits larger bias than the non-adjusted parametric expected profit from Equation (14); this is due to, in Figure 10, the red histogram being farther to the right than the orange histogram. However, both parametric and non-parametric adjusted profits exhibit almost no bias, evident from their histograms (blue and green) both being centered around zero. Therefore, despite the lack of parametric knowledge causing a larger bias to materialize than in the parametric case studied in Siegel and Wagner (2021), our data-driven adjustment can still neutralize the bias effectively, resulting in an asymptotically unbiased estimate of expected profit in the non-parametric domain.

Figure 10 Non-parametric and parametric (Siegel and Wagner 2021) adjustments eliminate bias for exponential demand with mean 200, $p = 5$, $c = 3$, and sample size $n = 25$. Note that the non-parametric bias is substantially larger than the parametric bias (i.e., the red histogram is farther to the right than the orange histogram).

5.2. Comparison with Cross Validation

In the language of machine learning, we use our sample of data $(X_1, \ldots, X_n)$ as a training data set to calculate $Y_{saa}$. However, evaluating the performance (i.e., expected profit) of an algorithm on the
same data used to train it is a mistake. Typically, in machine learning, data is split into training and testing data sets. The training data set would be used to calculate $Y_{\text{true}}$ and then the testing data set would be used to estimate the expected profit. Of course, this expected profit estimation would depend on the random split of the data into training and testing data sets. A better approach is to use cross validation to estimate the expected profit. In, say, 5-fold cross validation, the data is split into 5 equally sized sets, or folds, of data. Each of the folds serves as a testing data set, with the remaining 4 folds serving collectively as the training data. Averaging the 5 test data set expected profit estimations results in the cross-validated estimate of expected profit. Cross validation is one of the standard approaches in machine learning to estimate the performance of an algorithm on new data. In this section, we examine how cross validation performs with respect to our bias adjustment.

We reuse the hypothesis-testing experimental design of this section, where we add the t-statistic histograms for 5-fold cross-validated estimated expected profits. We report our findings in Figure 11 for both normally and lognormally distributed demand. We see that cross validation (in green) significantly reduces bias compared to the unadjusted expected profit (in red), as expected. However, our analytical bias adjustment reduces bias better than cross validation: observe that the histograms for our bias-adjusted expected profits (in blue) are centered more closely around zero, whereas the cross-validated histograms are centered to the left of zero, indicating that cross validation actually overcompensates and results in a negative bias. Furthermore, a majority of the cross-validated t-statistics would result in the rejection of the null hypothesis of zero bias at the standard critical value of -1.96. We conclude that our bias adjustment outperforms cross validation in this example.

![Histograms for Profit Estimation Error](image)

**Figure 11** Our analytical adjustment eliminates bias better than cross validation for normal and lognormal demand: normal distribution (left) and lognormal distribution (right) both have mean 200 and standard deviation 65, with $p = 5$, $c = 3$, and sample size $n = 25$. 
6. An Extension for Discrete Demand Distributions

There are many obstacles to extending our work here with continuous distributions to the discrete distribution case. Perhaps the most formidable of these is the density term \( \hat{f} \) in the denominator of the bias estimate \( \frac{\hat{f}}{n f(Y)} \left( 1 - \frac{\hat{f}}{f(Y)} \right) + o \left( \frac{1}{n} \right) \), because discrete distributions do not have a density in the usual sense (i.e., with respect to uniform measure).

One way to overcome this problem, and to extend our bias-reduction method from continuous to discrete distributions, would be to replace this density term with a covariance term that can be estimated for a discrete distribution. The motivation for this modification is the observation that Lemma 10, which demonstrates that \( \text{Cov} [Y, F(Y)] = \frac{1}{n f(Y)} \left( \frac{\hat{f}}{f(Y)} \right) \left( 1 - \frac{\hat{f}}{f(Y)} \right) + o \left( \frac{1}{n} \right) \), is the source of the bias’ form in which the density appears; see also Section 3.4.2 for more information about the covariance term. Therefore, we proceed by assuming that the term \( \text{Cov} [Y, F(Y)] \) is the sales bias for a discrete distribution, namely \( \text{Bias}_{sales} = \text{Cov} [Y, F(Y)] \), which implies that the profit bias \( \text{Bias}_{profit} = p \text{Cov} [Y, F(Y)] \). We then estimate this covariance directly, and use the estimate to similarly define an adjusted profit. While our theory seems to depend heavily on the continuous nature of the distribution, we nonetheless find, through simulation, significant bias reduction with this approach, where we make use of the multinomial distribution and an exact bootstrap calculation for each Monte Carlo sample, using Section 5.4 ‘Bootstrap methods for more general problems’ on page 31 of Efron (1982), and replacing bootstrap simulations with an exact calculation. We use simulations to demonstrate the effectiveness of this method for reducing bias because a full theoretical treatment of the discrete case is beyond the scope of the current paper.

Our simulation design involves three levels. At the top level is the original population from which we choose repeated Monte Carlo random samples of data. Each of these Monte Carlo samples represents the second level. At the third level, we perform an exact Bootstrap calculation on each Monte Carlo sample, effectively reconstructing the exact result as if we had constructed (and averaged results from) an infinite number of Bootstrap samples from that Monte Carlo sample (where each such virtual Bootstrap sample represents the third level of this hierarchy). Note that this third Bootstrap level is allowing the current (random) Monte Carlo sample to act as the (current, temporary) population of interest for this purpose, namely to quantify the effect of random sampling by using sampling variability of Bootstrap samples to estimate the high-level variability inherent in Monte Carlo samples from the original population.

The Bootstrap method motivates our estimation of the covariance bias term \( \text{Cov} [Y, F(Y)] \) for each Monte Carlo sample of data \( x_1, \ldots, x_n \); we outline the approach in this paragraph and provide details in the next paragraphs. First, as in Section 2.1, we estimate \( F(x) \) using the empirical cumulative distribution function \( \hat{F}(x) = \frac{1}{n} \sum_{i=1}^{n} I_{x_i \leq x} \) for this Monte Carlo sample, using the Bootstrap methodology on each Monte Carlo sample (which represents the population from which the
Bootstrap samples will be obtained). Second, we estimate the domain of the $Y$ random variable (where $Y$ is the SAA order quantity of a Bootstrap sample chosen independently with replacement from the Monte Carlo sample $x_1, \ldots, x_n$) as the support of the data $x_1, \ldots, x_n$. Third, we directly compute the probability that $x_i$ is realized as the $k$-th order statistic in a Bootstrap sample (i.e., a sample of size $n$ chosen with replacement from the current Monte Carlo sample); i.e., we estimate the probability $P(Y = x_i)$ as $\hat{P}(Y = x_i)$, $i = 1, \ldots, n$, which provides an estimate of the probability mass function of $Y$ for this Monte Carlo sample. This provides a list of tuples $(y, \hat{F}(y), \hat{P}(Y = y))$, $y \in \{x_1, \ldots, x_n\}$, from which we can compute the probability-weighted covariance, with further details as follows.

The exact Bootstrap calculation begins by observing that, due to the discrete nature of the demand distribution, the pair $(y, \hat{F}(y))$, to be used to find their covariance, may occur across many potential Bootstrap samples from a given Monte Carlo sample, and we need only compute the probability of the pair $(y, \hat{F}(y))$ for each possible value of $y \in \{x_1, \ldots, x_n\}$. To begin, for a given $y$, the probability that a Bootstrap element is strictly less than $y$ is estimated as $\hat{p}_{\text{less}} = \frac{1}{n} \sum_{i=1}^{n} I_{x_i < y}$ and the probability that an element is strictly greater than $y$ is estimated as $\hat{p}_{\text{greater}} = \frac{1}{n} \sum_{i=1}^{n} I_{x_i > y}$; consequently, the probability that a sample equals $y$ is estimated as $\hat{p}_{\text{equal}} = 1 - \hat{p}_{\text{less}} - \hat{p}_{\text{greater}}$.

The set of all Bootstrap samples with $Y = y$ divides neatly into three equivalence classes defined by the following three numbers that sum to $n$: $n_{\text{less}}$ is the number of items in a given Bootstrap sample that are strictly less than $y$, $n_{\text{equal}}$ for the number equal to $y$, and $n_{\text{greater}}$ for the number strictly greater than $y$. The only constraints on these three counts are imposed by the requirement that the $k$-th order statistic of the Bootstrap sample be equal to $y$. Thus we require only that $0 \leq n_{\text{less}} \leq k - 1$ (so that the $k$-th order statistic of the Bootstrap sample is not less than $y$) and $0 \leq n_{\text{greater}} \leq n - k$ (so that the $k$-th order statistic is not greater than $y$) with (by necessity) $n_{\text{equal}} = n - n_{\text{less}} - n_{\text{greater}}$. Note that with all of these possibilities, there is always at least one in the 'equal' category because $n_{\text{less}} + n_{\text{greater}} \leq n - 1$, guaranteeing that $y$ is the $k$-th order statistic of any such Bootstrap sample.

The probability of obtaining a Bootstrap sample with $k$-th order statistic $y$ and permissible counts $(n_{\text{less}}, n_{\text{equal}}, n_{\text{greater}})$ is the multinomial probability

$$\frac{n!}{n_{\text{less}}! n_{\text{equal}}! n_{\text{greater}}!} (\hat{p}_{\text{less}})^{n_{\text{less}}} (\hat{p}_{\text{equal}})^{n_{\text{equal}}} (\hat{p}_{\text{greater}})^{n_{\text{greater}}},$$

which collapses the many equivalent Bootstrap samples we might have chosen and enables us to consider only unique sets of counts for each $y$. We accumulate these probabilities for each $y$ by summing over the permissible values of $(n_{\text{less}}, n_{\text{equal}}, n_{\text{greater}})$, with $n_{\text{equal}} = n - n_{\text{less}} - n_{\text{greater}}$, to find the estimated probability $\hat{P}(Y = y)$:

$$\hat{P}(Y = y) = \sum_{n_{\text{less}}=0}^{k-1} \sum_{n_{\text{greater}}=0}^{n-k} \frac{n!}{n_{\text{less}}! n_{\text{equal}}! n_{\text{greater}}!} (\hat{p}_{\text{less}})^{n_{\text{less}}} (\hat{p}_{\text{equal}})^{n_{\text{equal}}} (\hat{p}_{\text{greater}})^{n_{\text{greater}}}.$$
This gives us the tuple \((y, \hat{F}(y), \hat{P}(Y = y))\). Having these tuples, one for each distinct \(y\) (i.e., the distinct values in this Monte Carlo sample from original data set), we compute the estimated (weighted) covariance for this Monte Carlo sample as

\[
\hat{\text{Cov}}[Y, F(Y)] = \hat{E}[YF(Y)] - \hat{E}[Y] \hat{E}[F(Y)] \\
= \sum_{\text{distinct } y} \hat{P}(Y = y)y\hat{F}(y) - \left( \sum_{\text{distinct } y} \hat{P}(Y = y)y \right) \left( \sum_{\text{distinct } y} \hat{P}(Y = y)\hat{F}(y) \right),
\]

where \(\hat{E}\) denotes the exact bootstrap estimated expectation for this Monte Carlo sample. This estimated covariance is then used to estimate the bias in sales, while \(p\hat{\text{Cov}}[Y, F(Y)]\) estimates the bias in profit, for the original sample of data. Finally, we calculate an adjusted expected profit as in Equation (13) for this Monte Carlo sample with its SAA order quantity \(Y\), using the covariance term as estimated using the Bootstrap:

\[
\hat{\pi}_{\text{adjusted}} = \frac{1}{n} \sum_{i=1}^{n} (p \min\{x_i, Y\} - cY) - p\hat{\text{Cov}}[Y, F(Y)].
\]

To calculate the true expected profit for this Monte Carlo sample with its SAA order quantity \(Y\), we proceed differently, since this calculation utilizes the true distribution of demand \(P(X = x)\) and therefore sums over all distinct (discrete) \(x\) in the original population (from which the Monte Carlo samples were obtained) while using their true population probabilities:

\[
\pi_{\text{true}} = p \sum_{\text{distinct } x} P(X = x) \min\{x, Y\} - cY.
\]

### 6.1. Evaluation of the Proposed Adjustment

To evaluate our proposed adjustment for a discrete demand distribution, we consider 1) a uniform distribution with support on \(\{0, 1, \ldots, 10\}\) and 2) a binomial distribution with 10 trials and probability of success 0.5. We conduct an experiment with 10,000 Monte Carlo trials.

For the uniform distribution, the true average profit is estimated as 3.191 with a standard error of 0.032, the naive average profit is estimated as 3.702 with a standard error of 0.037, and the adjusted average profit is estimated as 3.161 with a standard error of 0.032. The average bias of the naive profit is 0.511 with a standard error of 0.014 and the average bias of the adjusted profit is -0.030 with a standard error of 0.014. The unadjusted bias 0.511 is 16.0% of the true expected profit 3.191, whereas the adjusted bias -0.030 is 0.9% of the true expected profit. Our adjustment reduces the magnitude of the estimated error by \(1 - |-0.030|/0.511 = 94.1\%\).

For the binomial distribution, the true average profit is estimated as 6.855 with a standard error of 0.069, the naive average profit is estimated as 7.059 with a standard error of 0.071, and the adjusted average profit is estimated as 6.751 with a standard error of 0.068. The average bias of
the naive profit is 0.204 with a standard error of 0.008 and the average bias of the adjusted profit is -0.104 with a standard error of 0.008. The unadjusted bias 0.204 is 3.0% of the true expected profit 6.855, whereas the adjusted bias -0.104 is 1.5% of the true expected profit. Our adjustment reduces the magnitude of the estimated error by $1 - | -0.104|/0.204 = 49.0\%$.

We observe that our proposed adjustments reduce the magnitude of the bias, but also converts a positive bias into a negative bias. We next explore this further. We replicate the hypothesis-testing experimental design described in Section 5. In the left panel of Figure 12 we present histograms of $t$-statistics for adjusted (left histogram in blue) and unadjusted (right histogram in red) for the discrete uniform distribution described above; in the right panel we do the same for the binomial distribution described above. Unfortunately, the blue histograms for the adjusted profits are not centered at zero, and our proposed adjustment does not work as well for discrete distributions as it does for continuous distributions. However, there is still significant bias reduction: in the left panel, for uniformly distributed demand, we see that the blue histogram for adjusted profit is centered at approximately -2.5 whereas the red histogram for unadjusted profit is centered as approximately 37.5. Similarly, in the right panel for binomially distributed demand, the blue histogram is centered at approximately -12.5 and the red histogram is centered at approximately 27.5. Thus, while some bias remains after our adjustment, it is reduced considerably compared to the unadjusted profits. In addition, our adjustment results in an underestimation of profit, which is, in our opinion, more palatable in practice than an overestimation of profit. Finally, please recall that this is an initial attempt at bias reduction for discrete distributions, and we hope that the general theory can be worked out eventually, although it is beyond the scope of the current paper.

![Histograms for discrete uniform and binomial distributions](image)

**Figure 12** Proposed discrete distribution adjustment reduces, but does not eliminate, bias: discrete uniform distribution (left) and binomial distribution (right) both have support on $\{0, \ldots, 10\}$ and mean 5, with $p = 5$, $c = 3$, and sample size $n = 25$. 
7. Conclusion

In this paper we identify a statistically significant error in naively estimating the expected profit in a data-driven Newsvendor Model, and we show how to correct the error while assuming only a smooth density, without parametric assumptions. In particular, we analyze a Newsvendor model, where the SAA order quantity is calculated by substituting the empirical demand distribution, generated by a sample of demand data, for the unknown true distribution. We prove that using the same empirical distribution to estimate the expected profit of the SAA order quantity results in a positive asymptotic bias, which we derive in closed-form. We argue that the dominant source of the bias is underage, and overage plays a minimal role. Furthermore, we provide an in-depth interpretation of the bias, connecting it mathematically with a covariance term as well as visualizing the bias using areas under curves. The bias expression allows us to adjust the expected profit estimate, to obtain an asymptotically unbiased expected profit estimate using only information from the sample, while the true distribution remains unknown. The bias is non-negligible: in our numerical experiments, which demonstrate that the estimation error is statistically significant and the adjusted estimation error is generally not significantly different from zero, the bias is approximately 2.4-3.0% of the true expected profit, when demand is generated by a lognormal and normal distribution, respectively. Exact numerical experiments, for exponentially distributed demand, demonstrate that the bias can be a much larger percentage of the true expected profit. We also interpret our results intuitively: while overage contributes negative sales bias, underage contributes a larger magnitude of positive sales bias, which dominates. This implies that the positive profit bias is due to underage. Finally, we provide an initial extension of our results for a discrete demand distribution that exhibits encouraging bias reduction results in numerical experiments.

We conclude by briefly discussing future research directions. First, our results could potentially be extended to a contextual Newsvendor model, where a demand observation $X$ is paired with a contextual vector $Z = (Z_1, \ldots, Z_m)$. A machine learning model could be used to predict the distribution of $X$ from $Z$. If such a prediction model provides an empirical demand distribution, conditional on $Z$, then the Newsvendor could set $Y_{saa}$ to the $1 - c/p$ quantile of this conditional distribution. Ban and Rudin (2018) and Bertsimas and Kallus (2020) study similar models from a variety of perspectives, but do not correct for any biases. Since this approach uses machine learning, our bias characterization and correction could potentially lead to more accurate estimation of performance metrics on hold-out test data sets. Second, since our approach is purely data-driven, it could be worthwhile to extend our analysis to more general optimization problems. Indeed, Kleywegt et al. (2002) demonstrated that the SAA approach, for a general stochastic discrete optimization problem, exhibits an objective function bias, but did not correct it; the techniques in our paper hold strong potential for adjusting such objective function estimations.
Acknowledgments
The first author gratefully acknowledges the support of the Grant I. Butterbaugh Professorship. The second author gratefully acknowledges the support of a Neal and Jan Dempsey Faculty Fellowship.

References
Efron B (1982) The jackknife, the bootstrap and other resampling plans (Society for industrial and applied mathematics).
Appendix. Proofs

Proof of Lemma 1. We first note that
\[ \frac{1}{n} \sum_{i=1}^{n} I_{x_i \leq x(k)} = \frac{k}{n} = \frac{n(1 - c/p)}{n} \geq \frac{n(1 - c/p)}{n} = 1 - c/p, \]
which establishes that \( x(k) \in \{ x_j : \frac{1}{n} \sum_{i=1}^{n} I_{x_i \leq x_j} \geq 1 - c/p \} \). To see that \( x(k) \) is the smallest element of this set, note that the next smallest data value is \( x_{(k-1)} \) (and that \( x_{(k-1)} < x(k) \) almost surely, since \( F \) is a continuous distribution) for which we find
\[ \frac{1}{n} \sum_{i=1}^{n} I_{x_i \leq x_{(k-1)}} = \frac{k-1}{n} = \frac{n(1 - c/p)}{n} - 1 < \frac{n(1 - c/p)}{n} = 1 - c/p, \]
where the inequality is due to the fact that \( \lceil n(1 - c/p) \rceil \) is the least integer greater than or equal to \( n(1 - c/p) \), so that \( \lceil n(1 - c/p) \rceil - 1 \), which is a smaller integer, must be strictly less than \( n(1 - c/p) \).

This shows that
\[ x_{(k-1)} \notin \left\{ x_j : \frac{1}{n} \sum_{i=1}^{n} I_{x_i \leq x_j} \geq 1 - c/p \right\}, \]
which completes the proof that \( k = \lceil n(1 - c/p) \rceil \) gives us the optimal order quantity \( y_{\text{soa}} = x(k) \) satisfying (2). □

Proof of Lemma 2. We first prove that \( P(X > Y) = (n+1-k)/(n+1) \): Note that the “extended sample” \( (X, X_1, X_2, \ldots, X_n) \) is itself an i.i.d. sample of size \( n+1 \) from \( f \). It follows from exchangeability that \( X \) is equally likely to be equal to each of the \( n+1 \) order statistics of the extended sample. To satisfy \( X > Y \), where \( Y = X_{(k)} \), the only choices for \( X \) are extended order

statistics \( (k+1), \ldots, (n+1) \), and there are \( n+1-k \) of these. Therefore we have
\( P(X > Y) = \frac{n+1-k}{n+1} \)

and, by the complement rule, \( P(X < Y) = \frac{k}{n+1} \), thereby establishing the first two parts of this lemma.

Next, using iterated expectations, we find
\[ P(X > Y) = E[P(X > Y|Y)] = E[1 - F(Y)], \]
from which we see that \( E[F(Y)] = P(X < Y) = k/(n+1) \), establishing the third part of this lemma.

For the variance of \( F(Y) = F(X_{(k)}) = U(k) \) we use the fact that it is an order statistic from a sample of uniform random variables, and therefore has distribution \( \text{Beta}(k, n+1-k) \) (see page 3 of Arnold et al. (2008)), from which it follows again that \( E[F(Y)] = k/(n+1) \) and also that
\[ \text{Var}[F(Y)] = k(n+1-k) / \left[ (n+1)^2(n+2) \right] = [k/(n+1)]((n+1-k)/(n+1))/(n+2) = O(1)O(1)O(1/n) = O(1/n), \]
establishing the fourth part of this lemma.
We next expand \( Y = F^{-1}(U(k)) \) in a second-order Taylor Series about \( E(U(k)) \), while using standard formulas for derivatives of inverse functions, to find

\[
Y = F^{-1}(U(k)) \\
\approx F^{-1}[E(U(k))] + \left\{ \left( F^{-1}ight)' [E(U(k))] \right\} [U(k) - E(U(k))] + \frac{1}{2} \left\{ \left( F^{-1}ight)'' [E(U(k))] \right\} [U(k) - E(U(k))]^2 \\
= F^{-1}[E(U(k))] + \frac{U(k) - E(U(k))}{f \{ F^{-1}[E(U(k)] \}} - \frac{1}{2} \frac{f' \{ F^{-1}[E(U(k)] \}}{f \{ F^{-1}[E(U(k)] \}}^2 [U(k) - E(U(k))]^2.
\]

Using this representation, it immediately follows (because the expectation of the first-order term is zero) that

\[
E(Y) = F^{-1}[E(U(k))] - \frac{f' \{ F^{-1}[E(U(k)] \}}{2f^3 \{ F^{-1}[E(U(k)] \}} Var(U(k)) + o\left(\frac{1}{n}\right)
\]

\[
= F^{-1}\left(\frac{k}{n+1}\right) - \frac{f' \{ F^{-1}[k/(n+1)] \}}{2f^3 \{ F^{-1}[k/(n+1)] \}} \frac{k(n+1-k)}{(n+1)^2(n+2)} + o\left(\frac{1}{n}\right),
\]

establishing the fifth part of this lemma. Note that the error term \( o(1/n) \) emerges here because, as just established, \( Var(U(k)) = O(1/n) \), and all other terms in the expression are \( O(1) \), thereby fully representing all terms of order \( O(1/n) \) and leaving only terms of order \( o(1/n) \).

Subtracting this expectation from the second-order Taylor Series of \( Y \) itself, and keeping terms only up to first order, we find the first-order series

\[
Y - E(Y) \approx \frac{U(k) - E(U(k))}{f \{ F^{-1}[E(U(k)] \}},
\]

which will be sufficient to compute the variance of \( Y \) to second order:

\[
Var(Y) = E\left\{ [Y - E(Y)]^2 \right\} \\
= \frac{E\left\{ [U(k) - E(U(k))]^2 \right\}}{(f \{ F^{-1}[E(U(k)] \})^2 + o\left(\frac{1}{n}\right)} \\
= \frac{Var(U(k))}{(f \{ F^{-1}[E(U(k)] \})^2 + o\left(\frac{1}{n}\right)}.
\]

Because \( Var(U(k)) = O(1/n) \) in the numerator, we may substitute \( E(Y) \) for \( F^{-1}[E(U(k))] \) in the denominator because \( E(Y) = F^{-1}[E(U(k)] + O(1/n) \), and, due to continuity of \( f \), the denominator becomes \( (f[E(Y)]^2 + O(1/n^2) \), where we recall that \( O(1/n^2) \) may be replaced by \( o(1/n) \). In the numerator, we may also substitute for \( Var(U(k)) \) (from the fourth part of this lemma) to find the main expression for \( Var(Y) \):

\[
Var(Y) = \frac{Var(U(k))}{f[E(Y)]^2 + o\left(\frac{1}{n}\right)} \\
= \frac{k(n+1-k)}{(f[E(Y)]^2(n+1)^2(n+2)} + o\left(\frac{1}{n}\right).
\]
Next, recognizing that $\frac{k}{n+1} = O(1)$ and $\frac{n+1-k}{n+1} = O(1)$, we also have
\[
Var(Y) = \frac{1}{n+2} \left( \frac{k}{n+1} \right) \left( \frac{n+1-k}{n+1} \right) \frac{1}{f^2[E(Y)]} + o\left( \frac{1}{n} \right)
\]
\[
= O\left( \frac{1}{n} \right) + o\left( \frac{1}{n} \right)
\]
\[
= O\left( \frac{1}{n} \right)
\]
completing the proof. □

**Proof of Lemma 3.** We begin by using the fact that $h'(y) = yf(y)$ to establish the first-order expansion for $h(Y)$ around $E[Y]$ to be
\[
h(Y) \approx h[E(Y)] + h'[E(Y)][Y - E(Y)] = h[E(Y)] + E(Y)f[E(Y)][Y - E(Y)].
\]
The first-order expansion of $F(Y)$ around $E[Y]$ is
\[
F(Y) \approx F[E(Y)] + F'[E(Y)][Y - E(Y)] = F[E(Y)] + f[E(Y)][Y - E(Y)],
\]
which is equivalent to
\[
f[E(Y)][Y - E(Y)] \approx F(Y) - F[E(Y)],
\]
which we substitute into the first-order expansion for $h$ to find
\[
h(Y) \approx h[E(Y)] + E(Y)\{F(Y) - F[E(Y)]\},
\]
which proves the first result of this lemma.

Using the definition of covariance, we have
\[
Cov[h(Y), F(Y)] = E(\{h(Y) - E[h(Y)]\}\{F(Y) - E[F(Y)]\})
\]
from which it follows that, for any constant $a$,
\[
Cov[h(Y), F(Y)] = E(\{h(Y) - E[h(Y)] - a\}\{F(Y) - E[F(Y)]\}),
\]
because $E(\{F(Y) - E[F(Y)]\}) = a\{E[F(Y)] - E[F(Y)]\} = 0$. Choosing the constant $a = h[E(Y)] - E[h(Y)] + E(Y)\{E[F(Y)] - F[E(Y)]\}$, and using the first part of this lemma, we find to first order that
\[
h(Y) - E[h(Y)] - a \approx h[E(Y)] + E(Y)\{F(Y) - E[F(Y)]\} - E[h(Y)] - a
\]
\[
= h[E(Y)] - E[h(Y)] + E(Y)\{F(Y) - F[E(Y)]\} - a
\]
\[
= E(Y)\{F(Y) - F[E(Y)]\} - E(Y)\{E[F(Y)] - F[E(Y)]\}
\]
\[
= E(Y)\{F(Y) - E[F(Y)]\},
\]
which allows us to express the covariance as an expected product of first-order terms, from which we find
\[
\text{Cov}[h(Y), F(Y)] = E(Y) E \left( \{F(Y) - E[F(Y)]\}^2 \right) + o(1/n)
= E(Y) \text{Var}[F(Y)] + o(1/n),
\]
where the error term \(o(1/n)\) emerges here because, by multiplying the two first-order terms we find that the expected square term is \(O(1/n)\) from Lemma 2, thereby fully representing all terms of order \(O(1/n)\) and leaving only terms of order \(o(1/n)\), which proves the second result of this lemma.

To simplify notation, we temporarily define first-order quantities
\[
\delta_h \equiv \frac{\{h(Y) - E[h(Y)]\}}{E[h(Y)]} \quad \text{and} \quad \delta_F \equiv \frac{\{F(Y) - E[F(Y)]\}}{E[F(Y)]},
\]
where \(E(\delta_h) = E(\delta_F) = 0\). We expand the term \(1/(1 + \delta_F)\) to second order to find
\[
E\left( \frac{h(Y)}{F(Y)} \right) = E\left( \frac{h(Y)}{F(Y)} \right) E \left( \frac{1 + \delta_h}{1 + \delta_F} \right)
= E\left( \frac{h(Y)}{F(Y)} \right) E \left[ \frac{1 + \delta_h (1 - \delta_F + \delta_F^2)}{1 + \delta_F} \right] + o\left( \frac{1}{n} \right)
= E\left( \frac{h(Y)}{F(Y)} \right) \left( 1 - \delta_F + \delta_F^2 + \delta_h - \delta_h \delta_F + \delta_h \delta_F^2 + o\left( \frac{1}{n} \right) \right)
\]
Keeping terms up to second order while discarding terms with expectation zero, we find
\[
E\left( \frac{h(Y)}{F(Y)} \right) = E\left( \frac{h(Y)}{F(Y)} \right) \left[ 1 + E\left( \delta_F^2 \right) - E\left( \delta_h \delta_F \right) + o\left( \frac{1}{n} \right) \right],
\]
where \(E[\delta_h \delta_F^2] = o(1/n)\). Substituting back for \(\delta_h\) and \(\delta_F\), we find
\[
E\left( \frac{h(Y)}{F(Y)} \right) = E\left( \frac{h(Y)}{F(Y)} \right) \left( 1 + \frac{\text{Var}[F(Y)]}{E[F(Y)]^2} - \frac{\text{Cov}[h(Y), F(Y)]}{E[h(Y)] E[F(Y)]} \right) + o\left( \frac{1}{n} \right)
\]
Using the second result of this lemma, namely \(\text{Cov}[h(Y), F(Y)] = E(Y) \text{Var}[F(Y)] + o(1/n)\), this becomes
\[
E\left( \frac{h(Y)}{F(Y)} \right) = E\left( \frac{h(Y)}{F(Y)} \right) \left( 1 + \frac{\text{Var}[F(Y)]}{E[F(Y)]^2} - \frac{E(Y) \text{Var}[F(Y)]}{E[h(Y)] E[F(Y)]} + o(1/n) \right) + o\left( \frac{1}{n} \right)
= \left( 1 + \frac{\text{Var}[F(Y)]}{E[F(Y)]^2} \right) E\left( \frac{h(Y)}{F(Y)} \right) - \frac{E(Y) \text{Var}[F(Y)]}{E[F(Y)]^2} E(Y) + o\left( \frac{1}{n} \right).
\]
Finally, to show that \(E[h(Y)] = E(S_{\text{over}})\) we use \(S_{\text{over}} = X_{I < Y}\) as specified at the start of Section 3.2. It then follows by iterated expectations while conditioning on \(Y\) that
\[
E(S_{\text{over}}) = E(X_{I < Y}) = E[E(X_{I < Y}|Y)] = E[h(Y)],
\]
completing the proof. \(\square\)
Proof of Lemma 4. We use

\[ E[F(Y)] = \frac{k}{n+1} \quad \text{and} \quad Var[F(Y)] = \frac{k(n+1-k)}{(n+1)^2(n+2)}, \]

both from Lemma 2, to find

\[
1 + \frac{Var[F(Y)]}{E[F(Y)]^2} - \frac{nE[F(Y)]}{k-1} = 1 + \frac{(n+1)^2Var[F(Y)]}{k^2} - \frac{nk}{(k-1)(n+1)}
\]

\[
= 1 + \frac{(n+1)^2k(n+1-k)}{k^2(n+1)^2(n+2)} - \frac{nk}{(k-1)(n+1)}
\]

\[
= \frac{k^2 - (n+1)^2}{k(k-1)(n+1)(n+2)}
\]

\[
= O\left(\frac{n^2}{n^4}\right)
\]

\[
= O\left(\frac{1}{n^2}\right),
\]

completing the proof. \(\square\)

Proof of Lemma 5. For the first assertion, we begin by noting that \(|k-n(1-c/p)|<1\), from which we find

\[
\left|\frac{k}{n} - \left(1 - \frac{c}{p}\right)\right| = \frac{1}{n} \left|k - n\left(1 - \frac{c}{p}\right)\right| < \frac{1}{n},
\]

establishing that \(k/n = 1 - c/p + O(1/n)\) and proving the first assertion of this lemma. We then use this to find

\[
\frac{k}{n+1} = \frac{k}{n} + \left(\frac{k}{n+1} - \frac{k}{n}\right)
\]

\[
= \frac{k}{n} - \frac{k}{n} \frac{1}{n+1}
\]

\[
= \frac{k}{n} + O\left(\frac{1}{n}\right)
\]

\[
= 1 - \frac{c}{p} + O\left(\frac{1}{n}\right) + O\left(\frac{1}{n}\right)
\]

\[
= 1 - \frac{c}{p} + O\left(\frac{1}{n}\right),
\]

which proves the second assertion of this lemma. Finally we compute, using the second assertion of this lemma

\[
\frac{n+1-k}{n} - \frac{n+1-k}{n+1} = \frac{n+1-k}{n(n+1)}
\]

\[
= \frac{1}{n} \left(1 - \frac{k}{n+1}\right)
\]

\[
= \frac{1}{n} \left\{1 - \left[1 - \frac{c}{p} + O\left(\frac{1}{n}\right)\right]\right\}
\]
\[
E[F(Y)] = k/(n + 1) \quad \text{and} \quad \text{Var}[F(Y)] = k(n + 1 - k)/(n + 2)
\]

from Lemma 2 to find
\[
\frac{k - 1}{n} \frac{\text{Var}[F(Y)]}{\{E[F(Y)]\}^2} = \frac{k - 1}{n} \frac{\text{Var}[F(Y)]}{\{E[F(Y)]\}^2} \frac{1}{(n + 1)^2(n + 2)}
\]
\[
= \frac{(k - 1)(n + 1 - k)}{kn(n + 2)}
\]
\[
= \frac{1}{n} \left[ 1 + O\left(\frac{1}{n}\right) \right] \left(1 - \frac{k}{n + 1}\right) \left[1 + O\left(\frac{1}{n}\right)\right]
\]
\[
= \frac{1}{n} \left(1 - \frac{k}{n + 1}\right) + O\left(\frac{1}{n^2}\right).
\]

We next use Lemma 5 to see that
\[
\frac{k - 1}{n} \frac{\text{Var}[F(Y)]}{\{E[F(Y)]\}^2} = \frac{1}{n} \left[ \frac{c}{p} + O\left(\frac{1}{n}\right) \right] + O\left(\frac{1}{n^2}\right) = \frac{1}{n} \left[ \frac{c}{p} \right] + O\left(\frac{1}{n^2}\right),
\]
completing the proof. □

**Proof of Lemma 6.** We use
\[
E[F(Y)] = k/(n + 1) \quad \text{and} \quad \text{Var}[F(Y)] = k(n + 1 - k)/(n + 2)
\]

from Lemma 2 to find
\[
\frac{k - 1}{n} \frac{\text{Var}[F(Y)]}{\{E[F(Y)]\}^2} = \frac{k - 1}{n} \frac{\text{Var}[F(Y)]}{\{E[F(Y)]\}^2} \frac{1}{(n + 1)^2(n + 2)}
\]
\[
= \frac{(k - 1)(n + 1 - k)}{kn(n + 2)}
\]
\[
= \frac{1}{n} \left[ 1 + O\left(\frac{1}{n}\right) \right] \left(1 - \frac{k}{n + 1}\right) \left[1 + O\left(\frac{1}{n}\right)\right]
\]
\[
= \frac{1}{n} \left(1 - \frac{k}{n + 1}\right) + O\left(\frac{1}{n^2}\right).
\]

We next use Lemma 5 to see that
\[
\frac{k - 1}{n} \frac{\text{Var}[F(Y)]}{\{E[F(Y)]\}^2} = \frac{1}{n} \left[ \frac{c}{p} + O\left(\frac{1}{n}\right) \right] + O\left(\frac{1}{n^2}\right) = \frac{1}{n} \left[ \frac{c}{p} \right] + O\left(\frac{1}{n^2}\right),
\]
completing the proof. □

**Proof of Lemma 7.** To find the conditional density \(g_{X\mid Y}(y)\) of \(Y\) given that \(X < Y\), we begin with its CDF \(G_{X\mid Y}(y)\), where we use \(t\) as a variable of integration because \(y\) appears explicitly as an argument,

\[
G_{X\mid Y}(y) = P(Y < y \mid X < Y)
\]
\[
= \frac{P(Y < y \text{ and } X < Y)}{P(X < Y)}
\]
\[
= \int_0^y \int_0^t f(x)g(t)\,dx\,dt
\]
\[
= \int_0^y \left[\int_0^t f(x)\,dx\right] g(t)\,dt
\]
\[
= \int_0^y F(t)g(t)\,dt
\]
\[
= \frac{F(t)g(t)\,dt}{P(X < Y)},
\]
from which it follows that the conditional density is

\[ g_{X \mid Y}(y) = \frac{dG_{X \mid Y}(y)}{dy} = \frac{F(y)g(y)}{P(X \mid Y)}, \]

establishing the first part of this lemma.

We next proceed similarly to find the conditional density \( g_{X \mid Y}(y) \) of \( Y \) given that \( X > Y \), beginning with the relevant CDF:

\[
G_{X \mid Y}(y) = P(Y < y \mid X > Y) = \frac{P(Y < y \text{ and } X > Y)}{P(X > Y)} = \frac{\int_0^y \int_t^\infty f(x)g(t)\,dx\,dt}{P(X > Y)} = \frac{\int_0^y [1 - F(t)]g(t)\,dt}{P(X > Y)},
\]

from which it follows that the conditional density is

\[
g_{X \mid Y}(y) = \frac{dG_{X \mid Y}(y)}{dy} = \frac{(1 - F(y))g(y)}{P(X > Y)},
\]

establishing the second part of this lemma. The third part follows immediately from (2.2.2) on page 10 of Arnold et al. (2008), while the fourth part follows immediately from (2.2.15) on page 13 of the same reference. This completes the proof. \( \square \)

**Proof of Lemma 8.** Given a particular observed order quantity \( Y = y \), we may view the partial sample of order statistics \( (X_{(1)}, \ldots, X_{(k-1)}) \) as though they came from the conditional distribution of \( X \) given that \( X < y \) because we could have created our original sample \( (X_1, \ldots, X_n) \) by first choosing a value \( Y = y \) from its distribution \( g \), then choosing \( k - 1 \) values from the distribution \( (X \mid X < y) \), then choosing the remaining \( n - k \) values from the distribution \( (X \mid X > y) \), and finally placing these values (including \( Y \)) in random order. Because the ordering of the partial sample \( (X_{(1)}, \ldots, X_{(k-1)}) \) is irrelevant to its sum, it follows that this partial sample behaves, conditionally on \( Y = y \), as though it arose as a random sample of size \( k - 1 \) from the conditional distribution \( (X \mid X < y) \). We conclude that, given \( Y = y \), the conditional expected mean of this partial sample is the conditional expected value \( E(X \mid X < y) \) so that

\[
E\left(\frac{1}{k-1} \sum_{i=1}^{k-1} X_{(i)} \mid Y = y \right) = E(X \mid X < y) = \frac{\int_0^y x f(x)\,dx}{\int_0^y f(x)\,dx} = \frac{h(y)}{F(y)}.
\]
To see that this is an increasing function of $y$, we take its derivative to show that it is positive, as follows:

$$
\frac{d}{dy} E(X|X < y) = \frac{d}{dy} \left( \frac{h(y)}{F(y)} \right) = \frac{d}{dy} \int_0^y x f(x) \, dx \int_0^y f(x) \, dx = \left( \frac{d}{dy} \int_0^y x f(x) \, dx \right) \left( \int_0^y f(x) \, dx \right) - \left( \int_0^y x f(x) \, dx \right) \left( \frac{d}{dy} \int_0^y f(x) \, dx \right) \left( \int_0^y f(x) \, dx \right)^2
$$

$$= y f(y) \left( \int_0^y f(x) \, dx \right) - \left( \int_0^y x f(x) \, dx \right) f(y) \left( \int_0^y f(x) \, dx \right)^2 = \frac{f(y) \int_0^y (y-x) f(x) \, dx}{\left( \int_0^y f(x) \, dx \right)^2} > 0,
$$

where the final inequality follows because $x < y$ within the integral, completing the proof of the first result of this lemma.

Next, we use iterated expectations to find

$$E \left( \frac{1}{k-1} \sum_{i=1}^{k-1} X_i \right) = E \left[ E \left( \frac{1}{k-1} \sum_{i=1}^{k-1} X_i \mid Y \right) \right] = E \left( \frac{h(Y)}{F(Y)} \right).$$

To see that $E(X|X < Y) = \frac{E[\min(Y,X)\mid Y]}{E[Y]}$, using iterated expectations, we find

$$E(X|X < Y) = \frac{E(X1_{X < Y})}{P(X < Y)} = \frac{E[E(X1_{X < Y})]}{E[P(X < Y)\mid Y]} = \frac{E[h(Y)]}{E[F(Y)]}.$$

To compare $E[h(Y)/F(Y)]$ to $E[h(Y)]/E[F(Y)]$, we write each in integral form:

$$E \left( \frac{h(Y)}{F(Y)} \right) = \int_0^\infty \int_0^y x f(x) \, dx \int_0^y f(x) \, dx g(y) \, dy = \int_0^\infty E(X|X < y) g(y) \, dy$$

and

$$\frac{E[h(Y)]}{E[F(Y)]} = \frac{1}{P(X < Y)} \int_0^\infty h(y) g(y) \, dy = \int_0^\infty g(y) \left( \frac{F(y) g(y)}{P(X < Y)} \right) \, dy = \int_0^\infty E(X|X < y) g_{X < Y}(y) \, dy,$$

where $g_{X < Y}$ is the conditional density of $y$ given that $X < Y$ as derived in Lemma 7. To establish

that $E \left( \frac{h(Y)}{F(Y)} \right) < \frac{E[h(Y)]}{E[F(Y)]}$, it remains only to show that

$$\int_0^\infty E(X|X < y) g(y) \, dy < \int_0^\infty E(X|X < y) g_{X < Y}(y) \, dy,$$
for which it will be sufficient to establish that \( g_{X < Y} \) first-order stochastically dominates \( g \) (because we have already established that \( E(X \mid X < y) \) is an increasing function of \( y \)). Stochastic dominance follows because these two densities have the monotone likelihood ratio property, which is clear from the fact that

\[
\frac{g_{X < Y}(y)}{g(y)} = \frac{F(y)}{P(X < Y)}
\]

is an increasing function of \( y \) (please note that the denominator on the right is a constant). It now follows that \( g_{X < Y} \) first-order stochastically dominates \( g \), completing the proof of this second assertion of this lemma.

The final assertion, that \( \frac{k}{n+1} E\left(\frac{1}{k-1} \sum_{i=1}^{k-1} X_{(i)}\right) < E[XI_{X < Y}] = E[h(Y)] \), follows by multiplying the previous assertion by \( E[F(Y)] = P(X < Y) = k/(n+1) \), completing the proof. \( \square \)

**Proof of Lemma 9.** Using Lemma 2, we find

\[
E[F(Y)] - \frac{k-1}{n} = \frac{k}{n+1} - \frac{k-1}{n} = \frac{n+1-k}{n(n+1)} = \frac{1}{n} \left( 1 - \frac{k}{n+1} \right) = \frac{1}{n} \left( 1 - \frac{c}{p} + O\left( \frac{1}{n} \right) \right) = \frac{1}{n} \left( \frac{c}{p} \right) + O\left( \frac{1}{n^2} \right),
\]

where the fourth equality is due to Lemma 5, which proves the first result of this lemma.

Using Lemma 2, we find

\[
Var[F(Y)] = \frac{k(n+1-k)}{(n+1)^2(n+2)} = \frac{1}{n} \left( \frac{n}{n+2} \right) \left( \frac{k}{n+1} \right) \left( \frac{n+1-k}{n+1} \right) = \frac{1}{n} \left( \frac{n}{n+2} \right) \left( \frac{k}{n+1} \right) \left( 1 - \frac{k}{n+1} \right) = \frac{1}{n} \left( \frac{n}{n+2} \right) \left( 1 - \frac{c}{p} + O\left( \frac{1}{n} \right) \right) \left( \frac{c}{p} + O\left( \frac{1}{n} \right) \right) = \frac{1}{n} \left( 1 - \frac{2}{n+2} \right) \left( 1 - \frac{c}{p} + O\left( \frac{1}{n} \right) \right) \left( \frac{c}{p} + O\left( \frac{1}{n} \right) \right) = \frac{1}{n} \left( 1 + O\left( \frac{1}{n} \right) \right) \left( 1 - \frac{c}{p} + O\left( \frac{1}{n} \right) \right) \left( \frac{c}{p} + O\left( \frac{1}{n} \right) \right)
\]

\( \square \)

---

1 A random variable \( A \) first-order stochastically dominates random variable \( B \) if and only if \( P(A > x) \geq P(B > x) \) for all \( x \). An alternate and equivalent definition is \( E[\phi(A)] \geq E[\phi(B)] \) for all increasing functions \( \phi \). See Shaked and Shanthikumar (2010) for further details.
\[
\frac{1}{n} \left[ \left( \frac{c}{p} \right) \left( 1 - \frac{c}{p} \right) + O \left( \frac{1}{n} \right) \right] \\
= \frac{1}{n} \left( \frac{c}{p} \right) \left( 1 - \frac{c}{p} \right) + O \left( \frac{1}{n^2} \right),
\]

where the fourth equality is due to Lemma 5, which proves the second result of this lemma.

We next show that \( E[F(Y)] - F[E(Y)] = O(1/n) \) because, expanding \( F(Y) \) to second order, we find

\[
F(Y) \approx F[E(Y)] + [Y - E(Y)] f[E(Y)] + \frac{1}{2} [Y - E(Y)]^2 f'[E(Y)],
\]

and therefore

\[
E[F(Y)] = F[E(Y)] + E[Y - E(Y)] f[E(Y)] + \frac{1}{2} f'[E(Y)] Var(Y) + o(1/n)
\]

\[
= F[E(Y)] + O \left( \frac{1}{n} \right),
\]

since \( Var(Y) = O(1/n) \) from Lemma 2. Subtracting and adding \( E[F(Y)] \) we may write

\[
F(Y) - F[E(Y)] = \{ F(Y) - E[F(Y)] \} + \{ E[F(Y)] - F[E(Y)] \},
\]

and thus

\[
E \left( [F(Y) - F[E(Y)]]^2 \right) = E \left( \{ F(Y) - E[F(Y)] \} + O(1/n) \right)^2
\]

\[
= E \left( \{ F(Y) - E[F(Y)] \}^2 \right) + 2E(\{ F(Y) - E[F(Y)] \}) O(1/n) + E \left\{ \{ O(1/n) \}^2 \right\}
\]

\[
= Var[F(Y)] + O(1/n^2),
\]

completing the proof. □

**Proof of Lemma 10.** Applying the definition of covariance,

\[
Cov[Y, F(Y)] = E \left( \{ Y - E(Y) \} \{ F(Y) - E[F(Y)] \} \right)
\]

\[
= E \left( \{ Y - E(Y) \} \{ F(Y) - F[E(Y)] \} \right),
\]

where the second equality is due to the following: since the first term in the product has expectation zero, the second term may be changed by adding the constant \( E[F(Y)] - F[E(Y)] \) with no effect. The first order Taylor series of \( F(Y) \) around \( E(Y) \) is

\[
F(Y) \approx F[E(Y)] + F'[E(Y)][Y - E(Y)] = F[E(Y)] + f[E(Y)] [Y - E(Y)],
\]

from which it follows that, also to first order,

\[
f[E(Y)] [Y - E(Y)] \approx F(Y) - F[E(Y)],
\]
which can be rearranged to
$$Y - E(Y) \approx \frac{F(Y) - F[E(Y)]}{f[E(Y)]}.$$

Making this substitution into the formula for the covariance, and noting that the error is lower order (because we are working with the expected product of two first-order terms) we find

$$\text{Cov}[Y, F(Y)] = \frac{1}{f[E(Y)]} E(\{F(Y) - F[E(Y)]\} \{F(Y) - F[E(Y)]\}) + o\left(\frac{1}{n}\right)$$

$$= \frac{1}{f[E(Y)]} \left( \text{Var}[F(Y)] + O\left(\frac{1}{n^2}\right) \right) + o\left(\frac{1}{n}\right)$$

$$= \frac{\text{Var}[F(Y)]}{f[E(Y)]} + o\left(\frac{1}{n}\right),$$

where the second equality is due to the third result of Lemma 9.

Finally, using the second result of Lemma 9 we have

$$\text{Cov}[Y, F(Y)] = \frac{1}{nf[E(Y)]} \left( \frac{1}{n} \left( 1 - \frac{c}{p} \right) + O\left(\frac{1}{n^2}\right) \right) + o\left(\frac{1}{n}\right)$$

$$= \frac{1}{nf[E(Y)]} \left( \frac{c}{p} \left( 1 - \frac{c}{p} \right) + O\left(\frac{1}{n^2}\right) + o\left(\frac{1}{n}\right) \right),$$

completing the proof. □

Proof of Lemma 11. From Lemma 2 we know that

$$E(Y) = F^{-1}\left(\frac{k}{n+1}\right) - \frac{f'\{F^{-1}\left[\frac{k}{(n+1)}\right]\}}{2f^3\{F^{-1}\left[\frac{k}{(n+1)}\right]\}} \frac{k(n+1-k)}{(n+1)^2(n+2)} + o\left(\frac{1}{n}\right).$$

Because $f'$ is bounded by assumption, noting that $\frac{k(n+1-k)}{(n+1)^2(n+2)} = \frac{k}{n+1} \left(1 - \frac{k}{n+1}\right) \frac{1}{(n+2)}$, and using $k/(n+1) = 1 - c/p + O(1/n)$ from Lemma 5, it follows that

$$E(Y) = F^{-1}\left(1 - \frac{c}{p} + O\left(\frac{1}{n}\right)\right) - \frac{f'\{F^{-1}\left[\frac{k}{(n+1)}\right]\}}{2f^3\{F^{-1}\left[\frac{k}{(n+1)}\right]\}} \left[1 - \frac{c}{p} + O\left(\frac{1}{n}\right)\right] \left[\frac{c}{p} + O\left(\frac{1}{n}\right)\right] \frac{1}{n+2} + o\left(\frac{1}{n}\right)$$

$$= y^* + O\left(\frac{1}{n}\right),$$

which establishes that $E(Y) \rightarrow y^*$. 

Next, to show that $Y$ is a consistent estimator of $y^*$, note that $Y = X_{(k)}$ is a sequence (indexed by the implicit $n$) of estimators of $y^*$ and that

$$|Y - y^*| \leq |Y - E(Y)| + |E(Y) - y^*|$$

$$\leq |Y - E(Y)| + O\left(\frac{1}{n}\right),$$
from which it follows that
\[
E \left\{ [Y - y^*]^2 \right\} \leq E \left\{ [Y - E(Y)]^2 \right\} + O \left( \frac{1}{n} \right)
\]
\[
= Var(Y) + O \left( \frac{1}{n} \right)
\]
\[
= O \left( \frac{1}{n} \right)
\]
\[
\rightarrow 0,
\]
where the final equality follows from Lemma 2. It now follows that \( Y \) is a consistent estimator of \( y^* \), by using Lemma 1.1 on page 332 of Lehmann (1983), completing the proof. □

Proof of Proposition 1. From the third and fourth results of Lemma 3, along with
\[
E \left( \frac{1}{k-1} \sum_{i=1}^{k-1} X(i) \right) = E \left( \frac{h(Y)}{F(Y)} \right)
\] from Lemma 8, we have
\[
E \left( \tilde{S}_{\text{over}} \right) - E \left( S_{\text{over}} \right)
\]
\[
= E \left( \frac{1}{n} \sum_{i=1}^{k-1} X(i) \right) - E \left[ h(Y) \right]
\]
\[
= \frac{k-1}{n} E \left( \frac{h(Y)}{F(Y)} \right) - E \left[ h(Y) \right]
\]
\[
= \frac{k-1}{n} \left( 1 + \frac{Var[F(Y)]}{E[F(Y)]^2} \right) \frac{E[h(Y)]}{E[F(Y)]} - \frac{k-1}{n} \frac{Var[F(Y)]}{E[F(Y)]^2} E(Y) - E[h(Y)] + o \left( \frac{1}{n} \right)
\]
\[
= \frac{k-1}{n} \left( 1 + \frac{Var[F(Y)]}{E[F(Y)]^2} \right) \frac{E[h(Y)]}{E[F(Y)]} - \frac{Var[F(Y)]}{E[F(Y)]} \frac{E[h(Y)]}{E[F(Y)]} - \frac{k-1}{n} \frac{Var[F(Y)]}{E[F(Y)]^2} E(Y) + o \left( \frac{1}{n} \right)
\]
\[
= \frac{k-1}{n} \left( 1 + \frac{Var[F(Y)]}{E[F(Y)]^2} \right) \frac{E[h(Y)]}{E[F(Y)]} - \frac{k-1}{n} \frac{Var[F(Y)]}{E[F(Y)]^2} E(Y) + o \left( \frac{1}{n} \right).
\]

Next we use Lemma 4 to find
\[
E \left( \tilde{S}_{\text{over}} \right) - E \left( S_{\text{over}} \right) = \frac{k-1}{n} \frac{O \left( \frac{1}{n^2} \right) E[h(Y)]}{E[F(Y)]} - \frac{k-1}{n} \frac{Var[F(Y)]}{E[F(Y)]^2} E(Y) + o \left( \frac{1}{n} \right)
\]
\[
= - \frac{k-1}{n} \frac{Var[F(Y)]}{E[F(Y)]^2} E(Y) + O \left( \frac{1}{n^2} \right) + o \left( \frac{1}{n} \right)
\]
\[
= - \frac{k-1}{n} \frac{Var[F(Y)]}{E[F(Y)]^2} E(Y) + o \left( \frac{1}{n} \right)
\]
\[
= - \left( \frac{1}{n} \left( \frac{c}{p} \right) + O \left( \frac{1}{n^2} \right) \right) E(Y) + o \left( \frac{1}{n} \right)
\]
\[
= - \frac{1}{n} \left( \frac{c}{p} \right) E(Y) + o \left( \frac{1}{n} \right),
\]
where the second equality is due to \( \frac{k-1}{n} = O(1) \) and \( \frac{E[h(Y)]}{E[F(Y)]} = O(1) \), and the fourth equality is due to Lemma 6, completing the proof. □
Proof of Proposition 2. We begin by using iterated expectations, first conditioning on $Y$, to obtain the expected sales in underage:

$$E(S_{\text{under}}) = E(Y I_{X > Y}) = E[E(Y I_{X > Y}|Y)] = E[Y E(I_{X > Y}|Y)] = E\{Y [1 - F(Y)]\}.$$

We next apply the definition of covariance to obtain

$$E(S_{\text{under}}) = E(Y) E[1 - F(Y)] + \text{Cov}[Y, 1 - F(Y)].$$

Next, recalling that $\tilde{S}_{\text{under}} = \frac{n + 1 - k}{n} Y/n$, the sales bias in underage may be written as

$$E(\tilde{S}_{\text{under}}) - E(S_{\text{under}}) = E\left(\left(\frac{n + 1 - k}{n}\right) Y\right) - E(Y) E[1 - F(Y)] - \text{Cov}[Y, 1 - F(Y)]$$

$$= E\left[E(F(Y)) - \frac{k - 1}{n}\right] E(Y) + \text{Cov}[Y, E(F(Y))].$$

Substituting from the first result of Lemma 9 and Lemma 10 we find

$$E(\tilde{S}_{\text{under}}) - E(S_{\text{under}}) = \frac{1}{n} \left(\frac{c}{p}\right) E(Y) + \frac{1}{nf[E(Y)]} \left(\frac{c}{p}\right) \left(1 - \frac{c}{p}\right) + o\left(\frac{1}{n}\right)$$

$$= \frac{1}{n} \left(\frac{c}{p}\right) E(Y) + \frac{1}{nf[E(Y)]} \left(\frac{c}{p}\right) \left(1 - \frac{c}{p}\right) + o\left(\frac{1}{n}\right).$$

From Lemma 11, we have $E(Y) \to y^*$, and from continuity and positivity of the function $1/f$ we also have $\frac{1}{f[E(Y)]} \to \frac{1}{f(y^*)}$, from which it follows that $\frac{1}{f[E(Y)]} = \frac{1}{f(y^*)} + o(1)$. Using this, we find

$$E(\tilde{S}_{\text{under}}) - E(S_{\text{under}}) = \frac{1}{n} \left(\frac{c}{p}\right) E(Y) + \frac{1}{n} \left(\frac{1}{f(y^*)} + o(1)\right) \left(\frac{c}{p}\right) \left(1 - \frac{c}{p}\right) + o\left(\frac{1}{n}\right)$$

$$= \frac{1}{n} \left(\frac{c}{p}\right) E(Y) + \frac{1}{nf(y^*)} \left(\frac{c}{p}\right) \left(1 - \frac{c}{p}\right) + o\left(\frac{1}{n}\right).$$

completing the proof. □

Proof of Lemma 12. These three results rely upon a representation for order statistics from a sample from the standard (mean 1) exponential distribution that may be found, e.g., in Theorem 4.6.1 on page 72 and Equation 4.6.5 on page 73 of Arnold et al. (2008). In particular, because $Y/\mu$ is the $k$-th order statistic from a sample of $n$ from a standard exponential distribution, it follows that $Y/\mu \overset{d}{=} \sum_{i=1}^{k} \frac{Z_i}{n + 1 - i}$, where $Z_i$ are independent standard exponential random variables.

For the first expectation, noting that $X$ (being independent of $Y$) is exponential with mean $\mu$, we first evaluate $E[\min(X, y)]$ for any fixed value of $Y = y$, and we find

$$E[\min(X, y)] = \int_0^\infty \min(x, y) \frac{1}{\mu} e^{-x/\mu} dx$$
\[
\begin{align*}
&= \frac{1}{\mu} \int_{0}^{y} x e^{-x/\mu} dx + \frac{y}{\mu} \int_{y}^{\infty} e^{-x/\mu} dx \\
&= \mu \int_{0}^{y} \frac{x}{\mu} e^{-x/\mu} d\left(\frac{x}{\mu}\right) + y \int_{y}^{\infty} e^{-x/\mu} d\left(\frac{x}{\mu}\right) \\
&= \mu \int_{0}^{y/\mu} w e^{-w} dw + y \int_{y/\mu}^{\infty} e^{-w} dw \\
&= -\mu \left( we^{-w}_{0} - \int_{0}^{y/\mu} e^{-w} dw \right) + ye^{-y/\mu} \\
&= -\mu \left( \frac{y}{\mu} e^{-y/\mu} - (1 - e^{-y/\mu}) \right) + ye^{-y/\mu} \\
&= -ye^{-y/\mu} + \mu (1 - e^{-y/\mu}) + ye^{-y/\mu} \\
&= \mu (1 - e^{-y/\mu}),
\end{align*}
\]

which implies, by iterated expectations, that \(E[\min(X,Y)] = \mu - \mu E\left( e^{-Y/\mu} \right)\). Next, to evaluate the expectation \(E\left( e^{-Y/\mu} \right)\), we again apply the identity from Arnold et al. (2008) to find \(E\left( e^{-Y/\mu} \right) = E\left[ \exp\left( - \sum_{i=1}^{k} \frac{Z_i}{n+1-i} \right) \right]\). Using independence of the \(Z_i\) to exchange product and expectation, we find

\[
E\left( e^{-Y/\mu} \right) = \prod_{i=1}^{k} E\left[ \exp\left( - \frac{Z_i}{n+1-i} \right) \right] \\
= \prod_{i=1}^{k} \int_{0}^{\infty} e^{-z/(n+1-i)} e^{-z} dz \\
= \prod_{i=1}^{k} \int_{0}^{\infty} e^{-z[1+1/(n+1-i)]} dz \\
= \prod_{i=1}^{k} \int_{0}^{\infty} \exp\left( - \frac{n+2-i}{n+1-i} z \right) dz \\
= \prod_{i=1}^{k} \frac{n+1-i}{n+1-i} \\
= \frac{n+1-k}{n+1},
\]

where the final equality follows because this is a telescoping product. It now follows that \(E[\min(X,Y)] = \mu - \mu_{n+1-k} = \mu_{n+1},\) completing the proof of the first expectation in this lemma.

Next, we again apply the representation from Arnold et al. (2008), this time to find

\[
E(Y) = \mu E\left( \frac{Y}{\mu} \right) \\
= \mu E\left( \sum_{i=1}^{k} \frac{Z_i}{n+1-i} \right) \\
= \mu \sum_{i=1}^{k} \frac{1}{n+1-i} \\
= \mu \sum_{i=n+1-k}^{n} \left( \frac{1}{i} \right),
\]
completing the proof of the second expectation in this lemma.

For the third expectation in this lemma, we yet again use the representation from Arnold et al. (2008), this time for various order statistics \( X_{(i)} \), while noting that the same \( Z_i \) values may be used for each of these order statistics, to find

\[
E\left( \sum_{i=1}^{k-1} X_{(i)} \right) = \mu \sum_{i=1}^{k-1} E \left( \frac{X_{(i)}}{\mu} \right) \\
= \mu \sum_{i=1}^{k-1} E \left( \sum_{j=1}^{i} \frac{Z_j}{n + 1 - j} \right) \\
= \mu \sum_{i=1}^{k-1} \sum_{j=1}^{i} \frac{1}{n + 1 - j}.
\]

We next change the order of summation, noting that the resulting inner summation then involves a constant (i.e., does not contain the inner index of summation) to find

\[
E\left( \sum_{i=1}^{k-1} X_{(i)} \right) = \mu \sum_{j=1}^{k-1} \sum_{i=j}^{k-1} \frac{1}{n + 1 - j} = \mu \sum_{j=1}^{k-1} \frac{k - j}{n + 1 - j}.
\]

We may then extend the upper summation limit from \( k - 1 \) to \( k \), because the summand is zero when \( j = k \), to find

\[
E\left( \sum_{i=1}^{k-1} X_{(i)} \right) = \mu \sum_{j=1}^{k} \frac{k - j}{n + 1 - j} \\
= \mu \sum_{j=1}^{k} \frac{n + 1 - j - (n + 1 - k)}{n + 1 - j} \\
= \mu \sum_{j=1}^{k} \left( 1 - \frac{n + 1 - k}{n + 1 - j} \right) \\
= \mu \left( k - (n + 1 - k) \sum_{i=n+1-k}^{n} \left( \frac{1}{i} \right) \right),
\]

completing the proof. \[\square\]

**Proof of Proposition 4.** For the true profit, it follows immediately from Lemma 12 that

\[
\pi_{true} = E[p \min(X, Y) - cY] = p \mu \frac{k}{n + 1} - c \mu \sum_{i=n+1-k}^{n} \left( \frac{1}{i} \right),
\]

completing the proof for true profit. For the naive profit, using Lemma 12, we find

\[
\pi_{naive} = pE\left( \frac{1}{n} \sum_{i=1}^{k-1} X_{(i)} + \frac{n + 1 - k}{n} Y \right) - cE(Y) \\
= \frac{p}{n} E\left( \sum_{i=1}^{k-1} X_{(i)} \right) + \left( \frac{p n + 1 - k}{n} - c \right) E[Y] \\
= \frac{\mu p}{n} \left[ k - (n + 1 - k) \sum_{i=n+1-k}^{n} \left( \frac{1}{i} \right) \right] + \mu \left( \frac{p n + 1 - k}{n} - c \right) \sum_{i=n+1-k}^{n} \left( \frac{1}{i} \right) \\
= \mu \left[ \frac{pk}{n} - c \sum_{i=n+1-k}^{n} \left( \frac{1}{i} \right) \right],
\]
completing the proof for naive profit. Taking the difference, and recognizing that the $c$ terms cancel, we find

$$
\text{Bias}_{\text{profit}} = \pi_{\text{naive}} - \pi_{\text{true}} = \mu \left[ \frac{pk}{n} - c \sum_{i=n+1-k}^{n} \left( \frac{1}{i} \right) \right] - \mu \left[ \frac{pk}{n+1} - c \sum_{i=n+1-k}^{n} \left( \frac{1}{i} \right) \right] = \frac{pk\mu}{n(n+1)},
$$

completing the proof. □

**Proof of Proposition 5.** We need the density $f(y^*) = \exp(-y^*/\mu)/\mu$ at the true (unknown) optimal order quantity $y^* = F^{-1}(1-c/p)$ for which $1-c/p = F(y^*) = 1 - \exp(-y^*/\mu)$, from which we deduce that $\exp(-y^*/\mu) = c/p$ and therefore the density we need is $f(y^*) = c/(\mu p)$. Using this density in the general profit bias formula from Theorem 1, we find

$$
\frac{p}{nf(y^*)} \left( \frac{c}{p} \right) \left( 1 - \frac{c}{p} \right) + o \left( \frac{1}{n} \right) = \frac{p}{n [c/(\mu p)]} \left( \frac{c}{p} \right) \left( 1 - \frac{c}{p} \right) + o \left( \frac{1}{n} \right)
$$

$$
= \frac{\mu p}{n} \left( 1 - \frac{c}{p} \right) + o \left( \frac{1}{n} \right).
$$

We next use $1-c/p = k/(n+1) + O(1/n)$, from Lemma 5, to find

$$
\frac{p}{nf(y^*)} \left( \frac{c}{p} \right) \left( 1 - \frac{c}{p} \right) + o \left( \frac{1}{n} \right) = \frac{\mu p}{n} \left( 1 - \frac{c}{p} \right) + o \left( \frac{1}{n} \right)
$$

$$
= \frac{\mu p}{n} \left[ \frac{k}{n+1} + O \left( \frac{1}{n} \right) \right] + o \left( \frac{1}{n} \right)
$$

$$
= \frac{pk\mu}{n(n+1)} + O \left( \frac{1}{n^2} \right) + o \left( \frac{1}{n} \right)
$$

$$
= \frac{pk\mu}{n(n+1)} + o \left( \frac{1}{n} \right),
$$

completing the proof. □