Applied Artificial Intelligence and Computing Systems
https://www.scigate.org
Science Gate Academic
Vol. 5, No. 6, 2026

Article

Learning Censored Demand for One Period Inventory Decisions

Enid Ravel*, Kian Dorsen, Mirel Ashby

Department of Industrial and Systems Engineering, North Carolina State University, Raleigh, United States

*Corresponding author: enidravel914@outlook.com

Abstract: Sales cease to reveal demand when stock runs out. We examine the decision cost of ignoring that fact in an executed twenty-four-item, one-period inventory study. Each item has a fixed generated Poisson demand rate, five hundred historical days with exogenous stock levels, and twenty thousand independent evaluation demands. A sales-mean estimate is compared with a grid maximum-likelihood estimate that treats stock-equal sales as right-censored observations. Both feed the same newsvendor rule with shortage cost nine and overage cost one. Mean evaluation cost is 13.6830 for the sales-mean policy, 5.7428 for censored likelihood, and 5.7136 for a known-rate reference. Rate RMSE falls from 3.7571 to 0.2770. A conserved unit-flow graphic, censored-tail diagram, likelihood ridges, and order staircase connect data recording to economic consequences. Results depend on stationary Poisson demand, correctly recorded stock, and a fixed historical sampling design. They establish neither dynamic-policy regret nor a retail deployment result, but show why an accurate model of the observation process belongs inside a learning-to-decision pipeline.

Keywords: censored demand; inventory learning; newsvendor; Poisson likelihood; decision-focused evaluation

1 The managerial quantity is hidden

A shop that stocks eight units and sells all eight has observed at least eight units of demand. It has not necessarily observed exactly eight. Treating that record as an uncensored demand realization turns a supply constraint into an apparent customer preference. A learning system can then recommend too little inventory, generate further stockouts, and continue receiving sales that appear to confirm its low forecast. This paper studies the first part of that mechanism under a fixed historical design, where the observation rule can be written exactly.

The decision is deliberately simple. A retailer selects a one-period stock quantity for each of twenty-four generated items. Unmet demand costs nine per unit and excess stock costs one per unit. There is no replenishment during the period, no carryover, no lead time, no substitution between items, and no capacity coupling. These restrictions make it possible to identify the consequence of demand estimation without confounding it with a large optimization problem. The resulting objective is an expected cost under a known evaluation generator, not observed revenue from an actual retailer.

Two estimators supply the demand model. The first uses average historical sales as if every day revealed full demand. The second uses the likelihood of the recorded event, distinguishing exact sales below stock from a tail event when sales equal stock. Both choose inventory with the same cost ratio and the same Poisson quantile calculation. A third policy receives the true demand rate as a reference. Holding the decision rule fixed ensures that differences in economic performance arise from the estimated demand distribution rather than from a different optimization procedure.

2 From inventory theory to a learning pipeline

Classical inventory policy [1] and dynamic lot-size planning [2] establish that stock decisions depend on the objective and operational structure. Our one-period problem should not be confused with those broader dynamic settings. Sample-average stochastic optimization [3] provides a general connection between data and decisions, while the big-data newsvendor [4], predictive-to-prescriptive analytics [5], and smart predict-then-optimize [6] emphasize that forecast quality should be assessed in relation to its downstream use. We use a transparent parametric likelihood so that this connection can be inspected directly.

Task-based learning [7] offers another route to incorporating decision consequences into model fitting. Lost-sales theory [8], nonparametric censored-demand planning [9], and analysis of censoring in the newsvendor [10] show that incomplete observations have a substantive effect on what can be learned. Demand estimation using stockout timing [11] highlights the value of richer transaction records. Our data contain only daily stock and sales, not within-day timing. The likelihood therefore uses the information actually generated rather than assuming an observation granularity that the experiment does not possess.

Kaplan–Meier estimation [12], multiproduct data-driven algorithms [13], retail-operations research [14], and adaptive censored-demand control [15] provide alternatives and broader context. General statistical learning [16] and introductory learning methods [17] clarify estimation issues, but no machine-learning benchmark is reproduced. A 2026 exact analysis of the value of censored newsvendor data [18] further emphasizes that information quality matters independently of data volume. Our contribution is an executed parametric case with a visible likelihood and a common downstream rule, not a new worst-case guarantee or nonparametric regret bound.

3 What the historical record contains

For each of twenty-four items, the true Poisson rate is drawn uniformly between three and fifteen. It stays fixed throughout the study. Five hundred historical days per item receive independently drawn integer stock levels from four through twelve. Demand is drawn independently conditional on the item's rate, and observed sales equal the smaller of demand and stock. Random seed 3701 fixes the collection. Historical stock is exogenous: it does not depend on earlier sales or on a fitted policy. This removes adaptive collection effects so that the likelihood comparison has a clear interpretation.

$$D_{it} \sim Poisson(\lambda_i), Y_{it} = min(D_{it}, Q_{it}) \tag{1}$$

The evaluator knows latent demand because the data are generated, but the estimators receive sales and stock only. A censoring flag is defined by sales equal to stock. When demand happens to equal stock exactly, that day is still treated as censored because the observed record does not distinguish exact equality from a larger latent demand. The corresponding event is demand greater than or equal to stock, not strictly greater. This boundary detail matters in a discrete distribution and determines the correct tail probability in the likelihood.

The fraction of censored item-days is 65.18%. This is a frequency of records whose sales equal stock, not the fraction of demand units lost. Figure 1 uses unit totals and therefore answers a different question. A day with a small stockout and a day with a large stockout each contributes one censored record, while their lost-unit totals can differ greatly. Keeping these quantities separate helps prevent a seemingly intuitive percentage from being assigned the wrong business meaning.

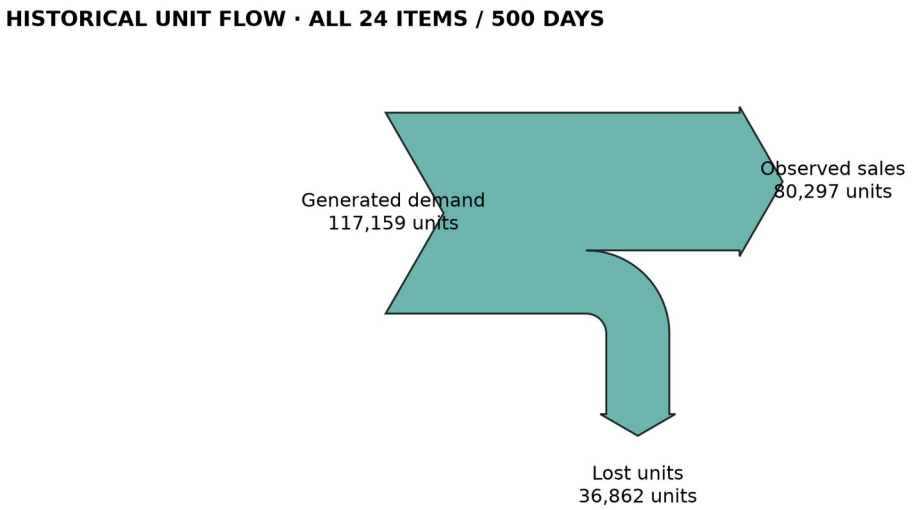


Figure 1. Conserved historical unit flow across all items and days. Generated demand splits into observed sales and unobserved lost units. The evaluator can compute both branches; the fitted estimators receive only sales and historical stock.

4 Estimating a rate from an exact event

For an uncensored day, the contribution to the likelihood is the probability mass at the observed sale. For a censored day, it is the probability that demand reaches or exceeds the known stock level. The product over days becomes a sum of log probabilities. The rate estimate is the maximum over a fixed grid from 0.2 to 25 with 497 points. All generated rates lie inside that range. The grid spacing is 0.05, which adds a small numerical approximation to the maximum-likelihood fit.

$$\ell_i(\lambda) = \sum_{Y<Q} \log p_\lambda(Y) + \sum_{Y=Q} \log Pr_\lambda(D \geq Q) \tag{2}$$

The naive estimate is simply average sales. Its problem is structural: the expectation of a minimum is generally below the expectation of demand. Increasing the number of days can make the sales mean more precise without making it converge to the true rate. It converges to the mean of the censored observation. In contrast, the likelihood encodes how different stock levels reveal different parts of the demand distribution. Under correct specification and adequate stock variation, those events can identify the rate even though individual lost demands remain unobserved.

$$E[min(D,Q)] = E[D] - E[(D-Q)_+] \tag{3}$$

Censoring does not magically supply unlimited information. If all stock levels are far below demand, almost every record is a tail event with probability close to one under many large rates. The likelihood can then be flat across plausible demand levels. Our historical stock range and rate range provide enough variation for useful estimation in the executed sample, but that is an experimental condition. Figure 3 shows normalized likelihood profiles for six items chosen by their rank in true rate, making both the estimated location and the relative concentration of evidence visible.

5 Converting the rate into stock

For a Poisson model with a known or estimated rate, the inventory rule chooses the smallest integer whose cumulative probability reaches 0.9. The value 0.9 is the shortage-cost share of the total shortage-plus-overage cost. Discreteness means the achieved cycle-service probability need not equal exactly ninety percent. The same rule is applied to all three rate inputs. No extra safety-stock adjustment is

tuned on the evaluation demands, and no policy receives a different cost ratio to improve its apparent performance.

$$Q^{*}(\hat{\lambda}) = min\{q \in \mathbb{Z}_{\geq 0} : F_{\hat{\lambda}}(q) \geq 0.9\} \tag{4}$$

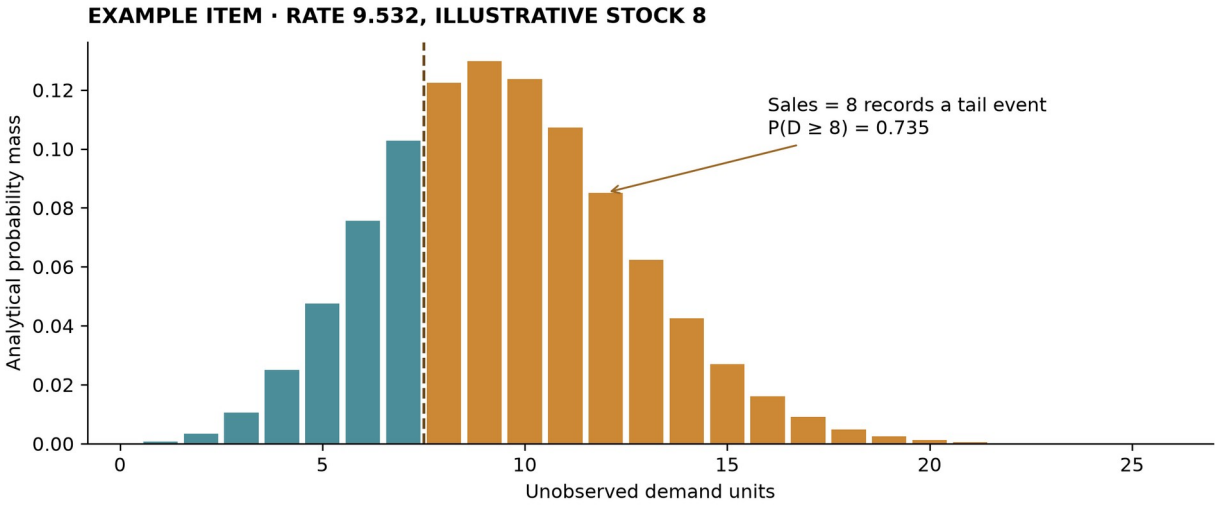


Figure 2. Analytical Poisson mass for one selected item and an illustrative stock level of eight. Ochre bars form the tail event recorded when sales equal stock. This diagram explains the likelihood boundary and is not an empirical histogram of a selected stock day.

The evaluation draws twenty thousand independent demands per item from its fixed generating rate. All policies face the same realized demands. Average cost is shortage units times nine plus excess units times one, averaged over every item and evaluation draw. Cycle service is the fraction of periods whose demand does not exceed stock. Fill rate is one minus total shortage units divided by total demand units. These are related but distinct metrics: a policy can miss some periods while still serving most units.

$$C(Q, D) = 9(D - Q)_{+} + (Q - D)_{+} \tag{5}$$

Table 1. Common-demand evaluation over 24 items and 20,000 test demands per item.

Rate input	Rate RMSE	Mean cost	Cycle service	Fill rate
Sales mean	3.7571	13.6830	62.66%	86.27%
Censored likelihood	0.2770	5.7428	91.96%	98.20%
Known demand rate	0.0000	5.7136	92.29%	98.27%

Censored likelihood reduces rate RMSE from 3.7571 to 0.2770 and brings cost close to the known-rate reference. Mean stock is approximately 10.04 under the sales-mean policy, 13.75 under censored likelihood, and 13.79 under the known-rate policy. The economic improvement is therefore accompanied by a substantial inventory increase. It should not be described as a cost-free accuracy gain. It is the expected consequence of recognizing demand hidden by stockouts under the specified shortage-heavy objective.

6 Why forecast error and decision value differ

An error in a rate estimate matters only through the order quantity it induces and the demand probabilities around that quantity. Two slightly different rates can produce the same integer order and exactly the same realized decision cost. Conversely, a small rate change near a quantile boundary can shift stock by one unit. This discreteness explains why rate RMSE alone cannot fully characterize economic value.

We report it to diagnose estimation, then evaluate cost and service directly to assess the resulting decision.

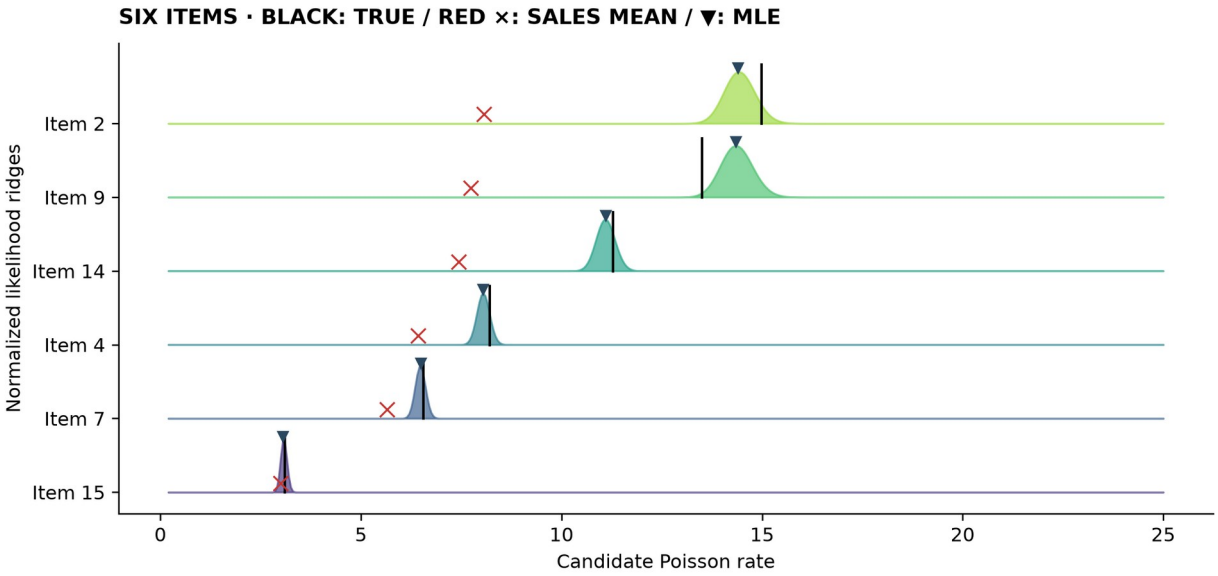


Figure 3. Normalized likelihood ridges for six rate-ranked items. Black marks indicate generated rates, red crosses the sales mean, and downward markers the censored-likelihood estimate. Each ridge is scaled to its own maximum, so ridge height does not compare total evidence across items.

The known-rate reference is optimal for the specified distribution and cost function in expectation, but its finite evaluation mean is still a Monte Carlo estimate. A fitted policy could occasionally have a lower realized cost on one finite test set without disproving that optimality. Here the censored-likelihood mean is slightly higher. We do not attach a significance claim to the small remaining gap or infer that more historical days would close it at a particular rate. The large contrast with the sales-mean policy is the central practical effect of this construction.

Equal averaging across items also defines a particular business objective. A retailer might weight items by revenue, margin, shelf space, or customer importance. Such weights could change which estimation errors matter most, even when each item's stock rule remains separable. Our equal-item mean is transparent and avoids implying commercial values for generated products. An application study should specify the aggregation rule before comparing algorithms. Otherwise, a convenient weighting can change a headline result without changing any item's actual performance.

7 Information collection is part of the policy

The historical design varies stock independently, which helps distinguish rates from censored observations. A real retailer often chooses stock based on past forecasts, making the observation process adaptive. Correct likelihood conditioning can still be possible under an appropriate policy model, but exploration and coverage become central. A policy that never stocks enough to reveal high demand may learn very slowly. The present offline experiment does not simulate that feedback or establish regret for an adaptive controller. Its fixed data design is a deliberate boundary.

Stock records must also be reliable. If displayed inventory differs from sellable inventory because of shrinkage, delayed replenishment, or recording errors, the censoring threshold in the likelihood is wrong. If customers substitute across items, demand is not independent of other stock decisions. If promotions change demand over time, a fixed Poisson rate may be inappropriate. Each issue changes the observation or demand model. A flexible learner cannot correct these assumptions merely by fitting a larger function to the same sales labels.

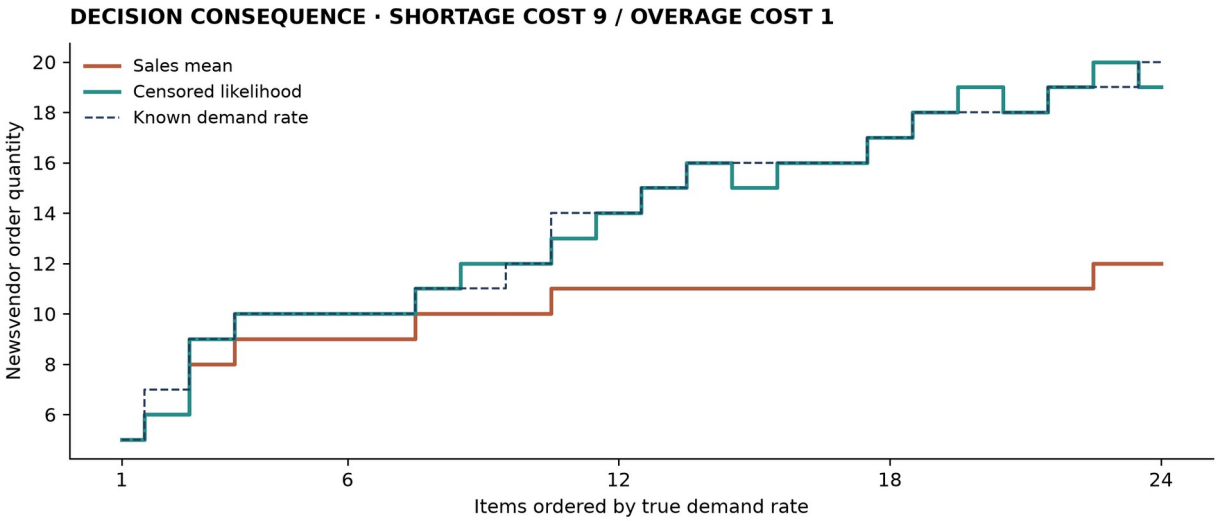


Figure 4. Stock-quantity staircase after sorting items by generated demand rate. All policies use the same 0.9 Poisson quantile rule. The ordering uses evaluator knowledge for presentation and is not supplied to either fitted estimator.

The likelihood method uses no contextual covariates, so it is a statistical learning baseline rather than a deep demand forecaster. Adding prices, calendar variables, and product attributes could improve prediction, but the censoring mechanism would still need to be represented. A neural predictor trained with a censored likelihood would be a natural extension; a network trained to reproduce raw sales would retain the same target problem. The essential distinction is the event being modeled, not the branding of the function approximator.

8 The operational reading

The result supports a specific data requirement: preserve both sales and the stock threshold that limits them. A stockout flag contains useful information even when lost units are not observed. Within-day stockout timing could add more information, but it is absent here. Collecting such records may be more valuable than adding a sophisticated predictor to an incomplete sales table. That statement is conditional on a correctly specified demand model and usable stock variation, rather than a universal return-on-investment claim.

The four visuals trace the business chain without inventing a retail case study. Unit flow shows what the historical system hides. The Poisson tail explains how a censored record enters estimation. Likelihood ridges show uncertainty and bias at the item level. The staircase reveals how estimates become integer stock decisions. Tables then report costs and service under common demands. This sequence avoids treating an abstract forecast score as if it were already a measured business outcome.

9 The observation rule inside the economic model

9.1 A sale can be a value or a set

An uncensored record identifies one demand value, while a stock-equal sale identifies a set of possible values. This distinction is the foundation of the likelihood. The data do not become unusable when a stockout occurs; they become less specific. Treating the set as its smallest member discards the information that larger demands remain possible and replaces uncertainty with a downward point assumption. The censored likelihood preserves that uncertainty through a tail probability. It estimates a rate without pretending to reconstruct the exact lost units on each historical day.

The discrete boundary is important. If stock is eight and sales are eight, demand could be eight, nine, or more. The event therefore includes equality. A likelihood using demand strictly greater than eight

would incorrectly exclude one possible latent outcome. This error can matter when the probability mass near stock is substantial. Our implementation evaluates one minus the cumulative probability through stock minus one. The formula and the Poisson-tail figure state this convention explicitly so that a reproduction does not silently fit a different observation model.

The censoring flag is derived from sales and stock rather than supplied as an independent oracle. That is appropriate for the generated rule because every unit of stock is sellable and recorded correctly. In a real system, sales equal to nominal stock might reflect a recording artifact, while sales below nominal stock could still involve unavailable shelf inventory. Such discrepancies would change the event represented by the record. A correct likelihood is only as credible as the observation process it models, which is why stock-data quality belongs in the method rather than in a generic limitations sentence.

9.2 The naive target has an exact bias identity

The sales mean estimates expected observed sales. Its downward difference from mean demand equals expected lost units under the historical stock process. This identity holds more generally than the Poisson family because it follows from the minimum operation. It explains why adding historical days can make the naive estimator more confidently wrong for demand. The estimator is consistent for a different target, not necessarily poorly optimized. The decision problem requires a demand distribution, so using the observed-sales target without correction creates a structural mismatch.

Historical stock variation changes that target. A retailer with higher past stock may have a sales mean closer to mean demand even if underlying demand is identical. Comparing sales across retailers or periods without accounting for stock can therefore confuse supply exposure with customer demand. Our generated stock levels are independent and known, which makes the correction especially clean. The result should not be interpreted as evidence that every sales forecast is biased in the same way; the issue arises when a demand quantity is inferred from constrained sales without modeling the constraint.

The bias identity also distinguishes record censoring from unit loss. Many lightly censored days can yield a high censoring frequency but modest lost units, while a few severe stockouts can hide many units. Neither quantity alone determines the rate-estimation problem. The thresholds and their variation matter because they shape the likelihood. Figure 1 reports conserved unit totals, while the main text reports censored item-day frequency. Keeping both prevents a single percentage from carrying several incompatible interpretations of information loss and business impact.

9.3 What variation in stock reveals

Each stock threshold provides information about a demand tail. If thresholds vary, the collection of stockout events samples different points of that tail. Under a one-parameter Poisson family, these events can jointly identify the rate even though many individual demands remain hidden. This is a strong benefit of a correctly specified parametric model. It should not be confused with recovering an arbitrary demand distribution from a finite set of low thresholds. The latter problem may leave high-demand behavior unidentified without additional assumptions or exploration.

Suppose every historical day stocks far below the true mean. Nearly all records then indicate a tail event whose probability is close to one over a broad range of large rates. The likelihood can become weakly informative and the maximum may depend on a grid boundary. Our rate grid extends above all generated true rates, and the varied thresholds provide informative uncensored observations for many items. A reproduction should still inspect whether a fitted maximum hits a boundary. A boundary estimate can be a diagnostic of insufficient information or an inadequate search range, not merely a number to report.

At the opposite extreme, if stock almost always exceeds demand, censoring is rare and the censored likelihood approaches ordinary Poisson fitting. In that setting the naive mean and likelihood estimate should be much closer. The current experiment intentionally contains substantial censoring so that the

mechanism is visible. It does not claim that the likelihood always produces a large practical improvement. The expected benefit depends on how much demand the historical policy hides and on whether the recorded thresholds help distinguish plausible rates.

9.4 Likelihood concentration is conditional on the model

The normalized ridges in Figure 3 show which rates fit each selected item's observed events. A narrow ridge indicates concentrated support within the Poisson family and the fixed stock design. It is not a guarantee that the Poisson family is correct. Overdispersion, zero inflation, promotions, or changing demand can produce a narrow but misspecified fit. A practical analysis should examine predictive checks and alternative distributions rather than treating likelihood curvature as a universal measure of confidence in the real demand process.

The ridges are normalized separately to maximum height one. Their heights therefore cannot compare total evidence across items. Their widths and locations remain visually informative, but a formal interval would require a stated construction and appropriate asymptotic or finite-sample justification. We do not label the ridges as posterior densities because no prior is used. They are relative likelihood profiles. This distinction keeps the plot from implying Bayesian uncertainty or calibrated coverage that the experiment has not evaluated.

Grid maximization introduces another finite resolution. The rate estimate can only take values separated by 0.05. This is small relative to the large naive bias in the executed sample, but can matter when comparing a fitted rate with the exact generating value at several decimals. The order rule is discrete as well, so a small grid error may have no decision consequence until it crosses a quantile boundary. The paper reports rate RMSE and economic cost precisely because these two approximation layers should not be collapsed into one score.

9.5 Deriving the critical fractile

The newsvendor condition can be understood by comparing the expected cost of stock q with stock q plus one. The extra unit incurs overage cost when demand is at most q and avoids shortage cost when demand is at least q plus one. The expected difference is overage cost times the cumulative probability through q minus shortage cost times its complement. The first q where this difference is nonnegative is an optimal integer stock. With costs one and nine, that condition produces the cumulative-probability threshold of 0.9 used throughout the experiment.

$$E[C(q+1,D)-C(q,D)]=F(q)-9[1-F(q)] \tag{6}$$

This derivation clarifies why ninety-percent cycle service is not guaranteed exactly. A discrete cumulative distribution can jump past the threshold, and several adjacent quantities can sometimes share the same expected cost at a boundary. We choose the smallest integer reaching the threshold, which fixes a reproducible tie convention. The known-rate policy's realized service above ninety percent is therefore compatible with the optimality condition. It should not be described as exceeding a promised service target through an additional algorithmic improvement.

Changing the cost ratio changes the required quantile and can change which estimation errors matter most. A shortage-heavy objective places greater value on upper-tail demand information, which censoring can hide. A symmetric or overage-heavy objective could reduce the cost gap for the same rate estimates. The current experiment fixes costs before evaluation and does not claim robustness across business preferences. A decision-focused study should report a sensitivity analysis if it intends to recommend a policy under uncertain cost parameters.

9.6 The same demand draws support a paired decision comparison

All policies face the same evaluation demand matrix. This makes their realized cost difference a paired quantity: a high-demand draw affects each policy under the same item and period. Independent test

demands for each policy would add unnecessary Monte Carlo variation and could obscure a small difference near the known-rate reference. The common-demand design does not make fitted policies identical or eliminate sampling uncertainty, but it provides a cleaner estimate of the consequence of their different stock choices under the declared generator.

The mean is averaged equally over items and draws. Because every item has the same number of evaluation demands, this is also an equal-item average of per-item means. It is not a revenue-weighted or unit-weighted cost. The fill-rate calculation, in contrast, uses total shortage units divided by total demand units and therefore weights high-demand items more heavily in its denominator. These aggregation choices are explicit because apparently similar service statistics can answer different questions. An applied retailer should select them according to its decision priorities before comparing models.

The known-rate reference is not a clairvoyant policy that sees each future demand before stocking. It knows the distribution parameter but still chooses one stock quantity before demand is realized. This distinction matters when interpreting its nonzero cost and occasional stockouts. Even perfect rate knowledge cannot remove all uncertainty under a stochastic demand distribution. Calling the reference an oracle is convenient only if the privilege is named precisely: it is a parameter oracle, not a future-demand oracle.

9.7 Near-reference cost does not establish perfect estimation

The censored-likelihood policy has a small mean cost gap to the known-rate reference in this sample, but its estimated rates still have nonzero RMSE. Many estimation errors do not change the chosen integer stock, while others change it by one unit with a modest expected cost effect. Decision quality can therefore be close to optimal even when parameter recovery is imperfect. This is one reason to evaluate the decision directly rather than requiring a forecast metric to stand in for economic performance.

The reverse can also occur. A model with lower average rate error might make an important error near a critical quantile for a high-cost item. Under a different cost structure or demand family, it could then have worse decision value. Our experiment avoids comparing several flexible forecasting models, so it does not establish a general ranking by decision loss. It illustrates the more basic point that modeling censoring changes the estimated distribution in a way that matters under one specified objective.

A small realized gap should not be overinterpreted statistically. The evaluation set is large, but all results remain conditional on one generated collection of twenty-four rates and one historical sample per item. New items or another history would change fitted rates and orders. A broader uncertainty study would repeat the complete historical-generation and fitting process over many item collections. The current experiment provides a transparent case, while its exact likelihood and cost definitions make such an extension straightforward to specify without inventing its results.

9.8 Adaptive inventory would change the evidence process

The historical stock levels in this paper are exogenous. In an adaptive system, a low forecast can cause low stock, which hides demand and reinforces the forecast. That feedback is a natural concern, but it is not simulated by the fixed historical design. The paper's results show the estimation and one-period decision consequences of censoring under known thresholds. They do not measure how quickly a sequential policy learns, how much exploration costs, or whether it recovers from an initially pessimistic forecast.

An adaptive policy might deliberately raise stock on selected days to learn the upper tail. Such exploration can create short-term overage cost while improving later decisions. Its value depends on horizon length, demand stability, and the information provided by each threshold. A likelihood fit alone does not choose an exploration policy. The 2026 censored-data literature motivates these questions, but

the current parametric case should not borrow a regret guarantee or exploration result from a different model. Any dynamic extension would need a new protocol and a clearly stated benchmark.

Lead times and carryover would further change the state of the decision problem. Inventory would become a dynamic variable rather than an independent one-period quantity, and lost sales could affect future customer behavior. A model that performs well under our separable rule might not remain adequate. This is why the title refers to the cost of an inventory forecast rather than claiming a complete inventory-control system. The narrow setting allows a clean information argument while leaving operational complexity to subsequent work.

9.9 Recording policy is a learning choice

Preserving a stock threshold can be more informative than preserving sales alone. A stockout flag tells the estimator that demand lies in a tail, while a sale count without threshold context can be interpreted incorrectly as an exact outcome. Within-day timing could add still more information about how quickly stock was exhausted. These observations suggest that data engineering and transaction recording can improve a learning system before its function approximator changes. The experiment quantifies one favorable instance, not a universal financial return from a recording upgrade.

Data quality checks should ensure that stock refers to available sellable units over the same period as sales. Returns, replenishments, reservations, and substitutions can complicate that definition. If the threshold changes within a day, one daily quantity may not describe the censoring event correctly. The current generator deliberately avoids those complications. A real implementation should model the event represented by its records rather than importing a simple minimum formula because it is convenient. Otherwise, a sophisticated likelihood can precisely estimate the wrong observation process.

The unit-flow graphic helps communicate this issue without pretending that lost units are observed in practice. The evaluator's generated demand total splits into sales and losses, but only one branch reaches the fitted estimators. Displaying both is legitimate because the data are synthetic and labeled as such. In a real retailer, the lost branch would need an estimate or an external measurement and should not be drawn with exact counts unless supported. Visual fidelity to the evidence is part of the statistical argument.

9.10 Repeating the calculation and interpreting its boundary

A reproduction must retain the stock range, demand-rate range, number of historical days, and equality convention for censoring. It should evaluate Poisson masses and tails stably, especially when candidate rates make an event unlikely. The grid includes all generated true rates within its interior, and the same estimated rates feed the same quantile routine. These details connect likelihood fitting to the economic result. Changing one without updating the description would create a different experiment even if the overall qualitative lesson remained similar.

The figures have distinct evidential roles. The tail diagram explains a single record, likelihood ridges show parameter support for selected items, and the stock staircase reveals the discrete decision mapping. The cost table then aggregates common-demand outcomes. No one plot is expected to prove the entire pipeline. This separation makes it possible to identify whether a disagreement arises from observation modeling, parameter estimation, decision conversion, or evaluation aggregation, rather than treating the final cost as an unexplained property of an algorithm.

The practical boundary is therefore precise. Censored likelihood helps when stock and sales are correctly recorded, the demand family is adequate, historical thresholds provide information, and the future decision resembles the modeled one-period task. Violating any of these conditions can reduce its value. Stating them does not weaken the large executed contrast with sales-as-demand; it explains why the contrast occurs and what must be checked before applying the method to a new setting. That explanation is the transferable contribution of the study.

9.11 Service metrics imply different customer experiences

Cycle service counts a period as successful only when every demanded unit is supplied. Fill rate counts the fraction of units supplied across periods. A high-demand period with a small shortage can count as a cycle-service failure while contributing little to lost-unit share. Conversely, repeated small-demand stockouts can create many failed periods. The two metrics therefore describe different aspects of service. Our table reports both so that the inventory increase under censored likelihood can be interpreted beyond its average cost.

The shortage-heavy cost ratio encourages high fill rate, but it does not impose a hard service constraint. A different optimization problem could minimize stock subject to a required cycle-service probability or a per-item guarantee. That would change the decision rule and possibly the relative value of estimation errors. The current ninety-percent fractile is derived from costs, not from a contract. A practical recommendation should avoid translating it into a service-level promise without checking the relevant business definition and uncertainty in the demand model.

Average service can also hide item heterogeneity. Some items may be well estimated because their historical stock often exceeds demand, while high-rate items remain heavily censored. The selected likelihood ridges and sorted order staircase expose part of that variation, but the headline table is still an aggregate. A deployment study should preserve item-level outcomes and identify which items drive cost or shortage. The generated collection has no product categories or commercial priority labels, so we do not invent such a segmentation.

9.12 A rate family is a useful simplification with a cost

The Poisson family ties its variance to its mean. That makes one rate sufficient for the full demand distribution and allows a clean quantile rule. Real demand can be more variable, clustered, seasonal, or affected by promotions. Under overdispersion, a rate estimate close to the true mean may still produce an inadequate upper quantile. The experiment therefore establishes the value of modeling censoring within a correctly specified family, not the adequacy of Poisson demand for every retail application.

A richer family could use separate dispersion parameters or contextual covariates, but censored data may identify those less strongly than a single rate. More flexibility can improve realism while increasing the information needed from historical thresholds. A nonparametric approach would face its own coverage limitations above the largest informative stock levels. These tradeoffs explain why modeling the observation process is necessary but not sufficient. It must be paired with a demand model whose tail is credible for the decision being made.

Model checking should therefore examine the distribution of observed sales conditional on stock, not compare an uncensored demand histogram that is unavailable to the estimator. In our generator, latent demands are known only for evaluation and illustration. A real analyst could compare predicted stockout frequencies and uncensored sales masses with held-out records, while recognizing that some tail behavior remains weakly observed. The current study does not execute a misspecification test, but its likelihood formulation identifies the events such a check should target.

9.13 Information and inventory form one pipeline

The economic result follows a chain: historical stock determines which demand events are observed, the likelihood maps those events into a rate, and the rate maps into an integer stock under fixed costs. Changing a neural architecture at the middle of that chain would not remove the need to define the two ends. The transparent baseline makes this dependence visible. It also shows why better transaction recording can matter even when the final predictor is simple.

This is the appropriate scope of the recommendation. Preserve threshold information, distinguish exact and censored observations, and evaluate the resulting decisions under the intended objective. Then test family misspecification, adaptive collection, and operational constraints. The executed study gives a

large and reproducible example of the first three steps. It leaves the later steps open rather than using near-reference synthetic cost to imply a ready retail-control system.

10 Conclusion

In the generated stationary Poisson setting, treating sales as demand yields a persistent downward rate bias and costly understocking. A likelihood that recognizes stock-equal sales as a tail event recovers much of the hidden demand information and produces near-reference one-period decisions. The experiment's lesson is to model the observation process before increasing forecast complexity. Dynamic inventory, unknown demand families, contextual effects, stock-record errors, and adaptive exploration remain separate problems that require additional evidence.

References

[1] K. J. Arrow, T. Harris, J. Marschak, “Optimal Inventory Policy,” Econometrica, vol. 19, pp. 250, 1951. doi: 10.2307/1906813.

[2] H. M. Wagner, T. M. Whitin, “Dynamic Version of the Economic Lot Size Model,” Management Science, vol. 5, pp. 89-96, 1958. doi: 10.1287/mnsc.5.1.89.

[3] A. J. Kleywegt, A. Shapiro, T. Homem-de-Mello, “The Sample Average Approximation Method for Stochastic Discrete Optimization,” SIAM Journal on Optimization, vol. 12, pp. 479-502, 2002. doi: 10.1137/s1052623499363220.

[4] G. Y. Ban, C. Rudin, “The Big Data Newsvendor: Practical Insights from Machine Learning,” Operations Research, vol. 67, pp. 90-108, 2019. doi: 10.1287/opre.2018.1757.

[5] D. Bertsimas, N. Kallus, “From Predictive to Prescriptive Analytics,” Management Science, vol. 66, pp. 1025-1044, 2020. doi: 10.1287/mnsc.2018.3253.

[6] A. N. Elmachtoub, P. Grigas, “Smart “Predict, then Optimize”,” Management Science, vol. 68, pp. 9-26, 2022. doi: 10.1287/mnsc.2020.3922.

[7] P. L. Donti, B. Amos, J. Z. Kolter, “Task-based End-to-end Model Learning in Stochastic Optimization,” arXiv preprint arXiv:1703.04529, 2017. [Online]. Available: https://arxiv.org/abs/1703.04529.

[8] M. Bijvank, I. F. A. Vis, “Lost-sales inventory theory: A review,” European Journal of Operational Research, vol. 215, pp. 1-13, 2011. doi: 10.1016/j.ejor.2011.02.004.

[9] W. T. Huh, P. Rusmevichientong, “A Nonparametric Asymptotic Analysis of Inventory Planning with Censored Demand,” Mathematics of Operations Research, vol. 34, pp. 103-123, 2009. doi: 10.1287/moor.1080.0355.

[10] O. Besbes, A. Muharremoglu, “On Implications of Demand Censoring in the Newsvendor Problem,” Management Science, vol. 59, pp. 1407-1424, 2013. doi: 10.1287/mnsc.1120.1654.

[11] A. Jain, N. Rudi, T. Wang, “Demand Estimation and Ordering Under Censoring: Stock-Out Timing Is (Almost) All You Need,” Operations Research, vol. 63, pp. 134-150, 2015. doi: 10.1287/opre.2014.1326.

[12] E. L. Kaplan, P. Meier, “Nonparametric Estimation from Incomplete Observations,” Journal of the American Statistical Association, vol. 53, pp. 457-481, 1958. doi: 10.1080/01621459.1958.10501452.

[13] C. Shi, W. Chen, I. Duenyas, “Technical Note—Nonparametric Data-Driven Algorithms for Multiproduct Inventory Systems with Censored Demand,” Operations Research, vol. 64, pp. 362-370, 2016. doi: 10.1287/opre.2015.1474.

[14] M. Qi, H. Y. Mak, Z. J. M. Shen, “Data-driven research in retail operations—A review,” Naval Research Logistics (NRL), vol. 67, pp. 595-616, 2020. doi: 10.1002/nav.21949.

[15] W. T. Huh, R. Levi, P. Rusmevichientong et al., “Adaptive Data-Driven Inventory Control with Censored Demand Based on Kaplan-Meier Estimator,” Operations Research, vol. 59, pp. 929-941, 2011. doi: 10.1287/opre.1100.0906.

[16] T. Hastie, R. Tibshirani, J. Friedman, “The Elements of Statistical Learning,” Springer Series in Statistics, 2009. doi: 10.1007/978-0-387-84858-7.

[17] G. James, D. Witten, T. Hastie et al., “An Introduction to Statistical Learning,” Springer Texts in Statistics, 2013. doi: 10.1007/978-1-4614-7138-7.

[18] R. Kumar, O. Mouchtaki, “What is the Value of Censored Data? An Exact Analysis for the Data-driven Newsvendor,” arXiv preprint arXiv:2602.16842, 2026. [Online]. Available: https://arxiv.org/abs/2602.16842.