PRIMAL-DUAL RESAMPLING FOR SOLUTION VALIDATION IN CONVEX STOCHASTIC PROGRAMMING

YI CHU[1], SUSAN R. HUNTER[2*], RAGHU PASUPATHY[1]

[1]*Department of Statistics, Purdue University, West Lafayette, Indiana, USA*

[2]*Edwardson School of Industrial Engineering, Purdue University, West Lafayette, Indiana, USA*

Abstract. Suppose we wish to determine the quality of a candidate solution to a convex stochastic program in which the objective function is a statistical functional parameterized by the decision variable and known deterministic constraints may be present. Inspired by stopping criteria in primal-dual and interior-point methods, we develop cancellation theorems that characterize the convergence of appropriately resampled and standardized primal and dual objective values to a weak limit. The resampled weak limit is distribution free, meaning it does not depend on the data-generating distribution. Furthermore, it is expressed as a functional of the standard Brownian motion, facilitating its implementation and valid use in optimality gap confidence interval construction and KKT point statistical testing. Since our results are general, we anticipate their use in iterative algorithm termination criteria for stochastic linear programs, two-stage convex stochastic programs, and a variety of finite- and infinite-dimensional problems arising in maximum likelihood estimation, nonlinear regression, classification, and portfolio management.

Keywords. statistical inference; statistical testing; confidence interval; bootstrap; batching; resampling; cancellation; self-normalization.

1. Introduction

Suppose a decision-maker or an iterative algorithm suggests a candidate solution to a stochastic program, and we wish to statistically assess its quality. This classic problem of assessing solution quality, or *solution validation*, usually involves constructing a confidence interval on the candidate solution's optimality gap. Such methods can be crucial for algorithm termination and decision-making in a variety of contexts such as stochastic linear programs [30, 42], two-stage convex stochastic programs [33], scalarized two-stage stochastic multi-objective linear programs [11, 17], and maximum likelihood estimation, nonlinear regression, classification, and portfolio management [23, 36, 57].

We consider a version of the solution validation problem in which the stochastic program's objective function is convex but need not be an expectation, and known deterministic constraints may be present. Specifically, we consider the problem of statistically testing the optimality of a given candidate solution to a deterministically constrained convex stochastic program in which

*Corresponding author.

E-mail addresses: yi.chu317@gmail.com, susanhunter@purdue.edu, pasupath@purdue.edu

Funding: Hunter gratefully acknowledges that this material is based upon work supported by the Air Force Office of Scientific Research under award number FA9550-23-1-0488.

Manuscript Revised: Monday December 15, 2025 at 12:11pm.

the objective function is a *statistical functional*, e.g., a nonlinear function of an expectation, a quantile, or a conditional value-at-risk (CVaR), that is "parameterized" by a decision variable. As in classical semiparametric theory [5, 22, 59], we handle such generalization by relying on notions of asymptotic linearity (and functional delta) from empirical processes theory [12, 15, 60]. Empirical process theory tools are especially relevant in the current stochastic programming setting, where the objective function can be seen as a functional of an unknown probability measure that can be sampled to produce a stochastic function estimator. Statistical optimality testing and KKT point validation then become questions of inference in which the estimator is a stochastic function that admits an expansion with a leading term that is a simple sample mean of observations obtained from a dataset or a stochastic oracle. A similar perspective of stochastic programming has been adopted recently in the context of assessing robustness [25].

1.1. **Perspective.** The traditional starting point for solution validation in stochastic programming is a classical central limit theorem (CLT), obtained by appropriately standardizing the objective function estimator. When considering the objective value, the standardization uses a scalar variance parameter, and when considering the solution, the standardization uses a covariance matrix — see, for instance, Shapiro et al. [56, Section 5.1.2], Mak et al. [33, equations (3) and (6)], Bayraksan and Morton [2, Assumption A.5], Morton [34], and Royset [45]. A substantial portion of the literature on M-estimation, e.g., [14], has also proceeded along this route. In all of these cases, when using the established (or assumed) CLT for confidence interval construction, one of two things typically happens. Either the unknown variance parameter is simply assumed to be known (explicitly or implicitly), or a consistent estimator is constructed and "plugged in" for the unknown variance parameter.

In settings where the estimator in question is something different from a simple sample mean, one can estimate the variance parameter by employing resampling techniques such as bootstrap, subsampling, or batching [13, 40, 41, 51, 57]. Resampling, loosely speaking, amounts to constructing multiple estimates of a quantity of interest, e.g., a variance parameter or an optimality gap, which are then appropriately used toward solution validation. In stochastic programming, Higle and Sen [19, 20] incorporated bootstrapping into decomposition methods in the 1990s, with later work by Bayraksan et al. [4], Chen and Woodruff [9], Lam and Qian [26, 27], Love and Bayraksan [32] using resampling and batching methods as a tool for solution validation.

A powerful technique called *cancellation*, introduced by Schruben [49] and analyzed much further in the seminal paper by Glynn and Iglehart [16], sits alongside resampling methods. The essential idea behind cancellation is that if the objective function estimator is standardized using an appropriately chosen functional satisfying specific properties, which we call the *resampling functional* ψ, then there is no need to estimate the variance parameter consistently: an appropriately chosen ψ "cancels out" the variance parameter in a manner reminiscent of the distribution-free t statistic. Cancellation exemplifies the idea that variance estimation and confidence interval construction are different problems, even though good variance estimates help to construct good confidence intervals. Parallel to the original ideas of cancellation introduced in [16, 49], there has been a steady development of similar ideas within mainstream statistics under the term *self-normalization* [24, 31, 52, 53].

1.2. **Primal-dual resampling and cancellation.** Our fundamental premise is that cancellation as an idea can be especially powerful in stochastic programming because of its ability to relieve the need to consistently estimate the variance parameter. Crucially, for computation-heavy optimization contexts, cancellation allows for the use of just a few large (potentially overlapping) batches of data when constructing estimates of the variance parameter in the service of optimality validation. As we show, the resulting limits may be non-normal, but are always distribution-free.

And yet, apart from some general treatment [37] in the context of statistical functionals, we are unaware of any development of cancellation that is tailored to *constrained convex stochastic programming*, especially when the objective function is not necessarily an expectation. It has been commented, for instance in Shapiro [54, Sections 5.6.1, 5.6.2], that in the presence of constraints, one might construct statistical upper and lower bounds on the primal and (Lagrangian) dual objective values at a candidate saddle point, respectively, when constructing a confidence interval on the optimality gap. This idea, when combined with resampling and cancellation, presents an interesting and powerful method for optimality validation.

Cancellation theorems for constrained convex stochastic programming naturally suggest methods for (i) statistically validating candidate KKT points and (ii) stopping an algorithm with probabilistic guarantees. Regarding (i), statistically validating candidate KKT points, or "KKT point testing," is the problem of constructing a test statistic (with a known and computable weak limit) in the service of testing whether a given point is a KKT point. The challenge, however, is that the necessary conditions for a KKT point require it to satisfy a vector equation, which implies the need for appropriate modification of the cancellation ideas, including generalizing the definition of resampling functional ψ. To the best of our knowledge, no work exists on this topic. Regarding (ii), stopping an algorithm in the stochastic programming context refers to terminating an algorithm with some probabilistic guarantee on the quality of the returned solution. The question of stopping has a long history within stochastic programming [3, 18, 20, 28, 34, 35, 38, 39]. However, as noted in Bayraksan and Pierre-Louis [3], their treatment and analysis as stopping times has been somewhat limited, with Patel [38] being an exception. With a view toward implementation, our interest lies in incorporating primal-dual resampling within stopping ideas, especially when constraints are present. This idea is reminiscent of duality gap stopping criterion in primal-dual and interior-point methods [44]. Such incorporation, along with major recent developments on *time-uniform confidence sequences* [21, 46, 58], provides a way to rigorously quantify the complexity of stopping and assessing the quality of stopped solutions.

1.3. **Paper organization.** In the remainder of the paper, we first provide a detailed problem setup with example problem formulations in Section 2 and summarize our results and contributions in Section 3. Section 4 contains mathematical preliminaries including notational conventions and standing assumptions. Our main results appear in Section 5. We provide example resampling functionals in Section 6, and Section 7 contains concluding remarks and a discussion of future research directions.

2. PROBLEM SETTING

To fix ideas, first, we formally define our problem setting. We begin by defining the "true" deterministic primal and dual problems, followed by their sample-path versions. Then, we provide two example formulations for the stochastic programming problems we consider. Finally, we define a linearly interpolated estimator for the objective function which is relevant to the discussion of our main results and contributions.

2.1. **The true primal and dual problems.** First, consider the true primal version of the stochastic program in which the uncertainty has been resolved by a statistical functional,

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & x \in \mathcal{X} := \{x \in \mathcal{D} \colon f_i(x) \leq 0,\, i = 1, \ldots, r\} \end{array} \tag{P}$$

for domain $\mathcal{D} := (\cap_{i=0}^{r} \operatorname{dom} f_i) \subseteq \mathbb{R}^q$, $\mathcal{D} \neq \emptyset$ where the feasible set $\mathcal{X}$ is nonempty, closed, and known. Let a global solution and the optimal value be, respectively,

$$x^* \in \mathcal{X}^* := \operatorname{argmin}\{f_0(x) \colon x \in \mathcal{X}\} \neq \emptyset, \qquad p^* = \min\{f_0(x) \colon x \in \mathcal{X}\} \in (-\infty, \infty). \tag{2.1}$$

The (Lagrangian) dual problem corresponding to (P) is

$$\begin{array}{ll} \text{maximize} & g(y) := \inf_{x \in \mathcal{D}} \left\{ f_0(x) + \sum_{i=1}^{r} y_i f_i(x) \right\} \\ \text{subject to} & y \geq 0, \quad y \in \mathbb{R}^r, \end{array} \tag{D}$$

with solution set $\mathcal{Y}^* := \operatorname{argsup}\{g(y) \colon y \geq 0\}$ and optimal value $d^* = \sup\{g(y) \colon y \geq 0\}$. A given point $\bar{y} \geq 0$ is *dual feasible* if $g(\bar{y}) > -\infty$. Let the set of all dual feasible points be denoted $\mathcal{Y}$,

$$\mathcal{Y} := \{y \in \mathbb{R}^r \colon y \geq 0,\, g(y) > -\infty\}.$$

Duality theory guarantees that $g(y)$ is concave in y even if (P) is not convex, and that if $\bar{x} \in \mathcal{X}$ is primal feasible and $\bar{y} \in \mathcal{Y}$ is dual feasible, then

$$f_0(\bar{x}) \geq p^* \geq d^* \geq g(\bar{y}) \tag{2.2}$$

where the dual feasibility of $\bar{y}$ ensures the right-most bound is meaningful. The inequality in (2.2) allows us to define the duality gap at $(\bar{x}, \bar{y})$ as $f_0(\bar{x}) - g(\bar{y})$. We formalize the standing assumptions on (P) in Assumption 4.1; importantly, we assume (P) is convex and Slater's condition holds so that $p^* = d^*$ in (2.2). The assumption that (P) is convex allows affine equality constraints [8], however, for simplicity, we do not explicitly include them. Finally, the set of allowable statistical functionals in f_0 is determined by the properties of its estimator; these properties are detailed in Assumption 4.2.

2.2. **The sample-path primal and dual problems.** The stochastic programming context involves an objective function that can only be estimated with stochastic error; that is, f_0 appearing in (P) is unknown but can be estimated using n observations from a stochastic oracle which takes the form of a dataset or a Monte Carlo simulation model. The estimation of f_0 leads to the following *sample-path approximations* of (P) and (D), constructed using the sample-path approximating function $F_{0,n}$:

$$\text{minimize} \quad F_{0,n}(x) \quad \text{subject to} \quad x \in \mathcal{X} \tag{P$_n$}$$

where we assume the domain $\mathcal{D}$ in (P_n) is the same as that in (P). Since the objective function f_0 in (P) does not necessarily involve an expectation, $F_{0,n}(x)$ is not necessarily expressed as a simple sample mean for given $x \in \mathcal{X}$ and may be biased. We denote a global solution and optimal value for (P_n) by the respective estimators

$$X_n^* \in \mathcal{X}_n^* := \operatorname{arginf}\{F_{0,n}(x) \colon x \in \mathcal{X}\} \subseteq \mathcal{X}, \qquad p_n^* = \inf\{F_{0,n}(x) \colon x \in \mathcal{X}\}. \tag{2.3}$$

Likewise, the sample-path dual problem is formed by replacing f_0 in (D) with the corresponding sample-path approximating function $F_{0,n}$,

$$\begin{array}{ll} \text{maximize} & G_n(y) := \inf_{x\in\mathcal{D}} \left\{F_{0,n}(x) + \sum_{i=1}^r y_i f_i(x)\right\} \\ \text{subject to} & y \geq 0, \quad y \in \mathbb{R}^r \end{array} \tag{D_n}$$

having solution set $\mathcal{Y}_n^* = \operatorname{argsup}\{G_n(y) \colon y \geq 0\}$ and optimal value $d_n^* = \sup\{G_n(y) \colon y \geq 0\}$. Let the set of all sample-path dual feasible points be denoted $\mathcal{Y}_n$,

$$\mathcal{Y}_n := \{y \in \mathbb{R}^r \colon y \geq 0,\ G_n(y) > -\infty\}.$$

As in (2.2), duality theory also guarantees that if $\bar{x} \in \mathcal{X}$ is primal feasible and $\bar{y} \in \mathcal{Y}_n$ is sample-path dual feasible, then

$$F_{0,n}(\bar{x}) \geq p_n^* \geq d_n^* \geq G_n(\bar{y}) \tag{2.4}$$

with a meaningful right-side bound. We formalize the standing assumptions on (P_n) in Assumption 4.2; in particular, we assume $F_{0,n}$ is almost surely convex. Therefore, (P_n) is almost surely convex and it inherits Slater's condition from (P), which implies $p_n^* = d_n^*$ in (2.4). Further, in what follows, we shall require $\bar{y} \in \mathcal{Y} \cap \mathcal{Y}_n$ to be both dual feasible and sample-path dual feasible with probability one (w.p.1) for all n. For given n, a sufficient condition to ensure $\mathcal{Y} = \mathcal{Y}_n$ w.p.1 is that $\sup_{x\in\mathcal{D}}|F_{0,n}(x) - f_0(x)| < \infty$ w.p.1.

2.3. **Example formulations.** Two common example problems that fall into our framework are risk-neutral optimization and optimization of risk measures:

Example 2.1 (Risk-neutral optimization)**.** In risk-neutral optimization, the objective function f_0 appearing in (P) takes the form of an expectation, $f_0(x) = \mathrm{E}[F_0(x,\xi)] = \int_\Xi F_0(x,u)d\mathrm{P}(u)$ for all $x \in \mathcal{X}$, where $\xi \colon \Omega \to \Xi$ is a random variable, $F_0 \colon \mathbb{R}^q \times \Xi \to \overline{\mathbb{R}}$ is a random function defined with respect to the probability space induced by ξ, $(\Xi, \mathfrak{A}, \mathrm{P})$, and $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$. To solve (P) in a *sample average approximation (SAA)* framework [55], the random functions appearing in (P_n) are formed using sample means; that is, for $\xi_1, \ldots, \xi_n$ identically distributed to ξ and all $x \in \mathcal{X}$, we set $F_{0,n}(x) = n^{-1}\sum_{j=1}^n F_0(x,\xi_j)$. □

Example 2.2 (Optimization of risk measures)**.** The objective function f_0 may take the form of a risk measure [47], such as the CVaR [48], also called a superquantile. To define the CVaR, using the notation from Example 2.1, for all $x \in \mathcal{X}$, let the probability that the random variable $F_0(x,\xi)$ does not exceed the threshold τ be $\mathrm{P}(F_0(x,\xi) \leq \tau) = \int_\Xi F_0(x,u)\mathbb{I}\{F_0(x,u) \leq \tau\}d\mathrm{P}(u)$, where $\mathbb{I}$ is the indicator function and for simplicity, we assume P is nonatomic. Then for all $x \in \mathcal{X}$, the value-at-risk (VaR) is the quantile $\mathrm{VaR}_\alpha(x) = \inf\left\{\tau \in \mathbb{R} \colon \mathrm{P}\big(F_0(x,\xi) \leq \tau\big) \geq \alpha\right\}$. If $f_0(x)$

takes the form of $\mathrm{CVaR}_\alpha(x) = \mathrm{E}\big[F_0(x,\xi)\,|\,F_0(x,\xi) \ge \mathrm{VaR}_\alpha(x)\big]$ for all $x \in \mathcal{X}$, then following Rockafellar and Uryasev [43], we have

$$f_0(x) = \mathrm{CVaR}_\alpha(x) = \min_{\tau\in\mathbb{R}} \big\{\tau + (1-\alpha)^{-1} \textstyle\int_{u\in\Xi} [F_0(x,u) - \tau]^+ d\mathrm{P}(u)\big\}$$

where $[x]^+ = x$ when $x > 0$ and $[x]^+ = 0$ when $x \le 0$. Then for $\xi_1, \dots, \xi_n$ identically distributed to ξ, one estimator for $f_0(x)$ is

$$F_{0,n}(x) = \min_{\tau\in\mathbb{R}} \big\{\tau + \big(n(1-\alpha)\big)^{-1} \textstyle\sum_{j=1}^n [F_0(x,\xi_j) - \tau]^+\big\}.$$

Other estimators may be possible; see [10]. □

2.4. **Linearly interpolated estimators.** In what follows, we shall require the linearly interpolated versions of estimators which follow the classic construction in [7]. Specifically, let

$$\Delta := \big\{(s,t) \in [0,1]^2 \colon 0 \le s \le t\big\} \tag{2.5}$$

and for the objective function estimator $F_{0,n}$, define

$$F_{0,n}(x,0,t) := \begin{cases} 0 & t = 0 \\ F_{0,nt}(x) & t \in (0,1],\, nt \in \mathbb{Z} \\ F_{0,\lfloor nt \rfloor}(x) + (nt - \lfloor nt \rfloor) F_{0,\lceil nt \rceil}(x) & t \in (0,1],\, nt \notin \mathbb{Z}, \end{cases}$$
$$F_{0,n}(x,s,t) := F_{0,n}(x,0,t) - F_{0,n}(x,0,s), \quad (s,t) \in \Delta,$$

and define the corresponding analogues for f_0,

$$f_0(x,0,t) := t f_0(x),$$
$$f_0(x,s,t) := f_0(x,0,t) - f_0(x,0,s) = (t-s) f_0(x), \quad (s,t) \in \Delta.$$

We also require the linearly interpolated version of G_n the corresponding analogue for g, which are defined similarly. Recalling from (2.1) and (2.3) that p^* and p_n^* are the infima of the true and sample-path problems, respectively, the corresponding *sample-path minimum process* is

$$p_n^*(s,t) := \inf\big\{F_{0,n}(x,s,t) \colon x \in \mathcal{X}\big\}, \quad (s,t) \in \Delta.$$

Conceptually, given appropriate values of (s,t), the quantity $p_n^*(s,t)$ represents the infimum of the sample-path problem analogous to (P_n) but constructed using only the observations $\lfloor ns \rfloor$, $ns+1$, $ns+2$, ..., $\lceil nt \rceil$.

3. Summary of main results and contributions

Given the problem setting from Section 2, we are interested in statistically validating a candidate solution's optimality for (P). Toward this end, we wish to construct the following desired quantities:

(1) Given primal and dual feasible points $\bar{x} \in \mathcal{X}$ and $\bar{y} \in \mathcal{Y}$, respectively, we seek a valid *confidence interval on the optimality gap* at $\bar{x}$, that is, a random quantity U_n^{gap} such that $\mathrm{P}(f_0(\bar{x}) - p^* \le U_n^{\mathrm{gap}}) \to 1 - \alpha$ as $n \to \infty$.
(2) Given a primal feasible point $\bar{x} \in \mathcal{X}$, we seek a *test statistic* that can be used to conduct a valid hypothesis test to check whether $\bar{x}$ satisfies the first-order (KKT) conditions.

(3) Given a primal-feasible (and possibly sample-path optimal) sequence $\{\bar{x}_n \in \mathcal{X}, n \geq 1\}$ revealed in a "streaming" context, we seek to characterize the classical stopping time N_ε described as the time $n \geq 1$ when the width of a confidence interval on the true optimal value p^* first falls below a specified threshold ε. In particular, we seek a guarantee on the stopped solution $\bar{x}_{N_\varepsilon}$, and the magnitude of N_ε as a function of ε.

To construct the desired quantities, we propose a *primal-dual resampling framework* as a way to systematically use solutions to sub-sampled problems having the structure of (P_n) and (D_n), alongside a chosen *resampling functional* $\psi\colon C^q(\Delta) \to \mathbb{S}^q_{++}$ defined on the space $C^q(\Delta)$ of continuous $\mathbb{R}^q$-valued functions on Δ, defined in (2.5), and mapping into the set of all q-by-q symmetric positive definite matrices, $\mathbb{S}^q_{++}$. The proposed primal-dual resampling framework is made possible by our main contribution (Theorem 5.1), which is a general cancellation theorem of the type introduced by Glynn and Iglehart [16]. Theorem 5.1 characterizes the weak convergence of primal and dual objective sequences, standardized using a chosen resampling functional ψ, to a distribution-free limit that depends only on the standard Brownian motion. In particular, for given primal feasible $\bar{x} \in \mathcal{X}$ and dual feasible $\bar{y} \in \mathcal{Y}$, the cancellation theorem demonstrates the weak convergence limits

$$\frac{F_{0,n}(\bar{x}) - f_0(\bar{x})}{\psi\big(F_{0,n}(\bar{x},\cdot,\cdot)\big)} \xrightarrow{\mathrm{d}} \frac{\delta W(0,1)}{\psi\big(\delta W(\cdot,\cdot)\big)} \quad \text{and} \quad \frac{G_n(\bar{y}) - g(\bar{y})}{\psi\big(G_n(\bar{y},\cdot,\cdot)\big)} \xrightarrow{\mathrm{d}} \frac{\delta W(0,1)}{\psi\big(\delta W(\cdot,\cdot)\big)}, \tag{3.1}$$

where the linearly interpolated estimators appearing in the denominators are defined in Subsection 2.4, $\delta W(s,t) := W(t) - W(s)$, $(s,t) \in \Delta$, and $W(t), t \in [0,1]$ is the 1-dimensional standard Brownian motion.

The weak limits in (3.1) suggest asymptotically exact statistical upper and lower bounds U_n, L_n on $f_0(\bar{x})$ and $g(\bar{y})$, respectively, which we use to construct the desired valid confidence interval on the optimality gap at $\bar{x} \in \mathcal{X}$ (Corollary 5.1). Letting T_ψ be the distribution of the right-side weak limit in (3.1) and defining $\gamma_{\psi,\alpha} := T_\psi^{-1}(\alpha) := \inf\{\gamma\colon T_\psi(\gamma) \geq \alpha\}, \alpha \in (0,1)$ as its corresponding critical value, the upper and lower bounds take the form

$$U_n := F_{0,n}(\bar{x}) + \gamma_{\psi,1-\alpha_p}\psi\big(F_{0,n}(\bar{x},\cdot,\cdot)\big) \quad \text{and} \quad L_n := G_n(\bar{y}) - \gamma_{\psi,\alpha_d}\psi\big(G_n(\bar{y},\cdot,\cdot)\big)$$

for given $\alpha_p, \alpha_d \in (0, 0.5)$. Depending on the choice of ψ, the distribution T_ψ may be non-normal but is always *distribution free* and can be calculated "offline" without the need to estimate any problem-specific nuisance parameters. For any given $\alpha \in (0,1)$, choosing $\alpha_p + \alpha_d = \alpha$ suggests $U_n^{\mathrm{gap}} = U_n - L_n$ and $[0, U_n^{\mathrm{gap}}]$ as a candidate $(1-\alpha)$ confidence interval on the optimality gap $f_0(\bar{x}) - p^*$. Figure 1 shows U_n^{gap} in relation to other relevant quantities used to construct the confidence interval.

Toward obtaining a valid hypothesis test for checking whether a given primal feasible point $\bar{x} \in \mathcal{X}$ is a KKT point, under suitable regularity conditions, Theorem 5.2 demonstrates that a certain computable to-be-defined test statistic $\chi^2_{\psi,n}(\bar{x})$ converges weakly to

$$\chi^2_\psi := \delta W_q(0,1)^\intercal\, \psi^2\big(\delta W_q(\cdot,\cdot)\big)\, \delta W_q(0,1).$$

Here, $\chi^2_{\psi,n}(\bar{x})$ is a function of the sample-path gradient at $\bar{x}$, denoted $\nabla F_{0,n}(\bar{x})$. Again, the weak limit χ^2_ψ is distribution-free and computable offline, allowing for robust implementation of the

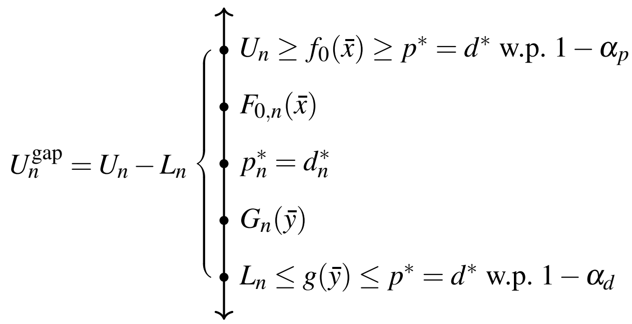

FIGURE 1. Illustration of $U_n^{\text{gap}} = U_n - L_n$ in relation to the sample-path optimal value $p_n^* = d_n^*$ and the true optimal value $p^* = d^*$ for given sample-path primal and dual feasible $\bar{x} \in \mathcal{X}$, $\bar{y} \in \mathcal{Y} \cap \mathcal{Y}_n$, respectively. All probabilistic statements hold as $n \to \infty$.

hypothesis test with null hypothesis that $\bar{x}$ satisfies the KKT conditions versus the alternative that it does not.

Finally, suppose that we terminate a sample-path optimal primal sequence $\{\bar{x}_n \in \mathcal{X}_n^*, n \geq 1\}$ when the confidence interval on p^* falls below a given tolerance ε, that is,

$$N_{\psi,\varepsilon} := \inf\{n \geq 1 \colon U_n^{\text{gap}} \leq \varepsilon\}.$$

Theorem 5.3 demonstrates that under certain regularity conditions, $\mathrm{P}\big(f_0(\bar{x}_{N_{\psi,\varepsilon}(n)}) - p^* > \varepsilon\big) < \alpha_p + \alpha_d$, and that the stopping time $N_{\psi,\varepsilon} = O(\varepsilon^{-2})$ as $\varepsilon \to 0$.

As noted in Section 1, with the aim of subsuming a large swathe of useful stochastic optimization problems, the main results appearing in Theorems 5.1–5.3 are stated for objective functions f_0 that are not necessarily expectations. However, in Subsection 4.2, we assume $F_{0,n}$ is asymptotically linear [50, Section 6.1.2], meaning that $F_{0,n}$ allows a functional expansion involving a sample-mean as the first term. We provide two concrete examples of resampling functionals ψ in Section 6.

4. PRELIMINARIES

Before stating the main results, we provide notational conventions and formally state and discuss standing assumptions that hold throughout the remainder of the paper.

4.1. **Notation and terminology.** $\mathbb{Z}$ is the set of all integers. $\mathbb{S}^d_{++}$ is the set of all $d \times d$ symmetric positive definite matrices, and $\mathbb{S}^d_{+}$ is the set of all $d \times d$ symmetric positive semi-definite matrices. $\mathbb{I}_d$ is the $d \times d$ identity matrix. The L_p norm of $x \in \mathbb{R}^q$ is $\|x\|_p := (\sum_{i=1}^q |x_i|^p)^{1/p}$, and $\|x\|$ refers to the Euclidean norm ($p = 2$). For $A \in \mathbb{S}^d_{++}$, $\|x\|_A = \sqrt{x^\intercal A x}$ is the A-norm or the Mahalanobis norm. For a random sequence $\{X_n, n \geq 1\}$, we write $X_n \xrightarrow{\text{wp1}} X$, $X_n \xrightarrow{\text{p}} X$, and $X_n \xrightarrow{\text{d}} X$ to refer to convergence with probability one (also called almost sure convergence), convergence in probability, and convergence in distribution (also called weak convergence), respectively. For positive-valued sequences $\{a_n, n \geq 1\}$ and $\{b_n, n \geq 1\}$, we say: (i) $a_n \sim b_n$ to mean $a_n/b_n \to 1$ as $n \to \infty$; (ii) $a_n \lesssim b_n$ to mean $a_n/b_n = O(1)$; and (iii) $a_n \gtrsim b_n$ to mean $b_n \lesssim a_n$. For a sequence of positive-valued random variables $\{A_n, n \geq 1\}$, we say $A_n = o_p(1)$ if $A_n \xrightarrow{\text{p}} 0$ as $n \to 0$. The space $C^d(\Delta)$ refers to the set of $\mathbb{R}^d$-valued continuous functions on Δ, where Δ is defined in (2.5); $C(\Delta)$ refers to $C^1(\Delta)$. A differentiable function $h\colon \mathcal{D} \subseteq \mathbb{R}^q \to \mathbb{R}$ is called L-smooth, if its gradient $\nabla h\colon \mathcal{D} \to \mathbb{R}^q$ is L-Lipschitz, that is, $\forall x, y \in \mathcal{D}, \|\nabla h(x) - \nabla h(y)\| \leq L\|x-y\|$.

An $\mathbb{R}^d$-valued stochastic process $\{Y(t), t \geq 0\}$ defined on some probability space $(\Omega, \mathcal{F}, P)$ is called a *Gaussian* process if for any $0 \leq t_0 < t_1 < \cdots < t_n < +\infty$, $(Y(t_1), Y(t_2), \ldots, Y(t_n))$ has a multivariate Gaussian [50, p. 4] distribution. The d-dimensional *Wiener process*, or Brownian motion, $\{W_d(t), t \geq 0\} \subset \mathbb{R}^d$ is a special type of Gaussian process satisfying the following four properties:

(1) $W_d(0) = 0$;
(2) with probability one, the map $t \mapsto W_d(t)$ is continuous in t;
(3) for $0 \leq t_0 < t_1 < \cdots < t_n < +\infty$, the increments

$$W_d(t_1) - W_d(t_0), W_d(t_2) - W_d(t_1), \ldots, W_d(t_n) - W_d(t_{n-1})$$

are independent; and
(4) for all $0 \leq s < t$, $W_d(t) - W_d(s)$ is multivariate Gaussian having mean 0 and covariance matrix $\Sigma \in \mathbb{S}^d_+$.

A d-dimensional Wiener process is called a *standard Wiener process* if $\Sigma = \mathbb{I}_d$. For simplicity, throughout the paper, the one-dimensional case $\{W_1(t), t \geq 0\}$ is denoted by $\{W(t), t \geq 0\}$.

4.2. **Standing Assumptions.** We now formalize our standing assumptions on the functions appearing in (P) and (P$_n$), as well as the required regularity conditions on the estimator and error field structures. We begin with regularity conditions associated with the true problem, coinciding with Shapiro [54, Theorem 5.11, conditions (i)–(iii)].

Assumption 4.1 (True problem regularity)**.** The following hold for the true problem (P):

(a) the problem is convex, that is, the feasible region $\mathcal{X}$ is a convex set, and the objective function $f_0\colon \mathcal{D} \to \mathbb{R}$ is convex;
(b) the solution set $\mathcal{X}^* = \operatorname{arginf}\{f_0(x)\colon x \in \mathcal{X}\}$ is nonempty and bounded;
(c) the functions $f_j, j = 0, 1, \ldots, r$ are finite-valued on a neighborhood of $\mathcal{X}^*$; and
(d) Slater's condition holds: There exists $x \in \mathcal{X}$ such that $f_j(x) < 0$ for all $j = 1, 2, \ldots, r$.

Recall that the objective function f_0 in (P) need not be an expectation, and accordingly the sample-path objective estimator $F_{0,n}$ need not be a sample mean constructed from unbiased estimators of f_0. However, we do require the estimator $F_{0,n}$ to exhibit certain regularity conditions including *asymptotic linearity* [50, 60], as detailed in the following Assumption 4.2.

Assumption 4.2 (Sample-path function regularity)**.** Let $\xi_1, \ldots, \xi_n$ be independent and identically distributed (iid) copies of ξ used to construct the sample-path function $F_{0,n}\colon \mathcal{D} \to \mathbb{R}$.

(a) $F_{0,n}$ is *asymptotically linear* [50, Section 6.1.2]. That is, for all $x \in \mathcal{D}$, we can write

$$F_{0,n}(x) = f_0(x) + n^{-1} \textstyle\sum_{j=1}^n \varepsilon(x, \xi_j) + b_n(x) \tag{4.1}$$

where $\mathrm{E}[\varepsilon(x,\xi)] = 0$, $\mathrm{Var}(\varepsilon(x,\xi)) < \infty$, and $\sup_{x \in \mathcal{D}} |b_n(x)| = o_p(n^{-1/2})$.
(b) There exists measurable $\kappa(\xi)$ with $\mathrm{E}[\kappa(\xi)] < \infty$ such that $\varepsilon(\cdot, \xi)$ is a.s. $\kappa(\xi)$-Lipschitz, that is, for all $x, x' \in \mathcal{D}$ we have that $|\varepsilon(x,\xi) - \varepsilon(x',\xi)| \leq \kappa(\xi)\|x - x'\|$.

(c) $F_{0,n}$ is convex w.p.1.

Assumption 4.2 is satisfied in a wide variety of stochastic optimization contexts. For instance, Assumption 4.2 is trivially satisfied in the classical case where the objective f_0 is an expectation that can be estimated as a sample mean which is constructed from unbiased observations, as in Example 2.1. Assumption 4.2 is also satisfied in other biased contexts such as when f_0 is a quantile, a CVaR as in Example 2.2, or indeed any smooth functional of the probability measure P; see [23, 57] for further discussion.

Assumption 4.2 serves to ensure that the field $F_{0,n}(\cdot) - f_0(\cdot)$ is well-behaved. Specifically, suppose $\mathcal{X}$ is compact. Then f_0 is κ_0-Lipschitz for some $\kappa_0 < \infty$ and Assumption 4.2 implies the following two key observations: (i) there exists $x_0 \in \mathcal{X}$ such that $\mathrm{E}[(f_0(x_0) + \varepsilon(x_0, \xi))^2] < \infty$; and (ii) $f_0(\cdot) + \varepsilon(\cdot, \xi)$ is $(\kappa(\xi) + \kappa_0)$-Lipschitz, where $\mathrm{E}[\kappa(\xi)] < \infty$. Due to (i) and (ii), the functional central limit theorem [1, Corollary 7.17] guarantees that $\sqrt{n}(F_{0,n} - f_0)$ converges weakly to a random element of $C(\mathcal{X})$, and in particular to a Gaussian process in the current context due to the assumed asymptotic linear structure (Assumption 4.2) of $F_{0,n}$. This observation will play an important role in proving the structure of the limits in the cancellation theorem. (Also see [56, Assumptions A.1 and A.2, pp. 164].)

Finally, while we have assumed that $\xi_1, \xi_2, \ldots$ in Assumption 4.2 is an iid sequence, all three theorems that we prove in this paper extend to stationary sequences [6, Chapter 20] with mild additional assumptions.

5. Main results

In this section, we develop *cancellation* and *stopping* results that form the linchpins for constructing confidence bounds and early algorithm termination in stochastic programming. While cancellation as an idea exists in the literature [16, 23, 57], our stochastic programming context requires the development of new results that, to the best of our knowledge, do not currently exist.

5.1. General Cancellation Theorem.

In this section, we present the main theorem which forms the basis for confidence interval construction and for KKT statistical testing, as detailed further in Subsection 5.2. Akin to the seminal idea in Glynn and Iglehart [16], carrying out the cancellation procedure implied by the following Theorem 5.1 involves choosing a resampling functional $\psi\colon C(\Delta) \to \mathbb{R}$ that resides in a class $\mathcal{M}_1$ of measurable real-valued functions, which is a special case of the general class $\mathcal{M}_d$ defined in Definition 5.1.

Definition 5.1. The *resampling class* $\mathcal{M}_d$ is the set of all functionals $\psi\colon C^d(\Delta) \to \mathbb{S}^d_{++}$ that are Borel-measurable, continuous under the sup norm, and satisfying the following four properties.

(a) (positively homogeneous) $\psi(c_1 y) = c_1 \psi(y)$ for $c_1 \in \mathbb{S}^d_{++}$, $y \in C^d(\Delta)$;
(b) (shift invariant) $\psi(y + c_0 k) = \psi(y)$ for $c_0 \in \mathbb{R}^{d\times d}$, $y \in C^d(\Delta)$, and $k(s,t) = (t-s)1_d$ where 1_d refers to the d-dimensional (column) vector of ones.
(c) (positive valued) $\mathrm{P}(\psi(W_d) \in \mathbb{S}^d_{++}) = 1$, where W_d is the d-dimensional standard Brownian motion (Wiener process) on $[0,1]$ (see Section 4).

The following cancellation theorem, Theorem 5.1, holds under Assumptions 4.1 and 4.2, assuming a functional $\psi \in \mathcal{M}_1$ has been chosen. We make several observations about the following Theorem 5.1. First, all of the limits in Theorem 5.1 involve *distribution-free* statistics, in the sense that the right-side limits do not depend on the problem-specific data-generating distributions appearing on the respective left-hand sides. As we make explicit in the proof of Theorem 5.1, the distribution-free nature of these limits is the result of standardization using ψ, leading to a cancellation of the problem-specific variance parameters. For the statistician, these limits are reminiscent of the classical Student's t distribution obtained as the standardized weak limit of sample means. Second, Parts 2 and 3 of Theorem 5.1 play direct roles in constructing upper and lower confidence bounds on the optimality gap at a candidate, using distribution-free limits. Finally, many choices of the function $\psi \in \mathcal{M}_1$ are possible. While all feasible choices result in asymptotically valid confidence intervals, the quality of the confidence intervals and the amount of computation required to construct them depend on the nature of ψ, where we measure the quality of a confidence interval by its rate of convergence to the nominal coverage probability $1-\alpha$ and its expected half-width. We provide a specific recommendation for ψ in Section 6.

Theorem 5.1 (Cancellation)**.** *Suppose Assumptions 4.1 and 4.2 hold, that $\psi \in \mathcal{M}_1$, and $\mathcal{Y} = \mathcal{Y}_n$ for all n. Define*

$$\delta W(s,t) := W(t) - W(s), \quad (s,t) \in \Delta,$$

and $\delta W := \{W(t) - W(s),\, (s,t) \in \Delta\}$.

(1) (Resampled limit) If $x \in \mathcal{D}$, then as $n \to \infty$,

$$\psi\big(\sqrt{n}\, F_{0,n}(x,\cdot,\cdot)\big) \xrightarrow{\text{d}} \sigma(x)\psi\big(\delta W(0,1)\big). \tag{5.1}$$

(2) (Primal candidate) If $\bar{x}$ is primal feasible, that is, $\bar{x} \in \mathcal{X}$, then as $n \to \infty$,

$$\frac{F_{0,n}(\bar{x},0,1) - f_0(\bar{x})}{\psi\big(F_{0,n}(\bar{x},\cdot,\cdot)\big)} \xrightarrow{\text{d}} \frac{\delta W(0,1)}{\psi\big(\delta W(\cdot,\cdot)\big)}. \tag{5.2}$$

(3) (Dual candidate) If $\bar{y}$ is dual feasible, then as $n \to \infty$,

$$\frac{G_n(\bar{y},0,1) - g(\bar{y})}{\psi\big(G_n(\bar{y},\cdot,\cdot)\big)} \xrightarrow{\text{d}} \frac{\delta W(0,1)}{\psi\big(\delta W(\cdot,\cdot)\big)}; \tag{5.3}$$

(4) (Sample optimal value) As $n \to \infty$,

$$\frac{p_n^*(0,1) - p^*}{\psi\big(p_n^*(\cdot,\cdot)\big)} \xrightarrow{\text{d}} \frac{\delta W(0,1)}{\psi\big(\delta W(\cdot,\cdot)\big)}. \tag{5.4}$$

Proof. We start by establishing (5.2), after which the proof of (5.1) will be evident. To prove (5.2), first, recall that solution set $\mathcal{X}^*$ is nonempty and bounded by Assumption 4.1. Since the problem in (P) is convex and Slater's condition holds, we have strong duality and the set $\mathcal{Y}^*$ of optimal solutions to the dual problem (D) is also nonempty and compact. Throughout what follows, we can thus assume (without loss of generality) that the feasible primal and dual sets $\mathcal{X}$ and $\mathcal{Y}$ are compact. Then, the asymptotic linearity and finite variance assumption in Assumption 4.2(a) and Assumption 4.2(b), respectively, imply that the process convergence limit

$$\big\{\sqrt{n}(F_{0,n}(\bar{x},0,t) - t f_0(\bar{x})),\, t \in [0,1]\big\} \xrightarrow{\text{d}} \sigma(\bar{x})W$$

holds, where $\sigma(\bar{x}) \in (0,\infty)$. It can be shown that this process convergence limit implies

$$\begin{aligned}\left\{\sqrt{n}(F_{0,n}(\bar{x},0,t)-tf_0(\bar{x})),\sqrt{n}(F_{0,n}(\bar{x},0,s)-sf_0(\bar{x})),(s,t)\in\Delta\right\}\\ \xrightarrow{\mathrm{d}}\left\{(\sigma(\bar{x})W(t),\sigma(\bar{x})W(s)),(s,t)\in\Delta\right\},\end{aligned}$$

and therefore that

$$\left\{\sqrt{n}(F_{0,n}(\bar{x},s,t)-(t-s)f_0(\bar{x})),\,(s,t)\in\Delta\right\}\xrightarrow{\mathrm{d}}\sigma(\bar{x})\delta W. \tag{5.5}$$

Now define the functional

$$h(y) := \frac{y(0,1)}{\psi(y)}, \quad y \in C(\Delta), \tag{5.6}$$

and write

$$\frac{F_{0,n}(\bar{x},0,1)-f_0(\bar{x})}{\psi\big(F_{0,n}(\bar{x},\cdot,\cdot)\big)} = \frac{\sqrt{n}\big(F_{0,n}(\bar{x},0,1)-f_0(\bar{x},0,1)\big)}{\psi\big(\sqrt{n}\big(F_{0,n}(\bar{x},\cdot,\cdot)-f_0(\bar{x},\cdot,\cdot)\big)\big)} \tag{5.7}$$

$$=: h\big(\sqrt{n}(F_{0,n}(\bar{x},\cdot,\cdot)-f_0(\bar{x},\cdot,\cdot))\big)\xrightarrow{\mathrm{d}} h(\sigma(\bar{x})\delta W)$$

$$= \frac{\sigma(\bar{x})\delta W(0,1)}{\psi(\sigma(\bar{x})\delta W(\cdot,\cdot))} = \frac{\cancel{\sigma(\bar{x})}\,\delta W(0,1)}{\cancel{\sigma(\bar{x})}\,\psi(\delta W(\cdot,\cdot))}, \tag{5.8}$$

where the equality in (5.7) follows from the shift and scale invariance properties of ψ, and the weak limit follows from the application of the mapping theorem [7, p. 20] which is allowed because $\sqrt{n}\,(F_{0,n}(\bar{x},\cdot,\cdot)-f_0(\bar{x},\cdot,\cdot)) \in C(\Delta)$ by Assumption 4.1, ψ is positive valued and a.s. continuous, and (5.5) holds. The last equality in (5.8) follows by (again) applying the assumed scale invariance of ψ. The weak limit implied by (5.8) proves the assertion in (5.2).

Next, we prove the limit in (5.4). First, start by defining

$$\tilde{p}(\phi) := \inf_{x\in\mathcal{X}}\sup_{y\in\mathcal{Y}}\phi(x,y), \quad \phi \in C(\mathcal{X}\times\mathcal{Y}).$$

By Assumptions 4.1 and 4.2, and since $\mathcal{Y} = \mathcal{Y}_n$ the true and sample-path Lagrangian functions

$$\begin{aligned}\tilde{L}(x,y) &:= f_0(x) + \textstyle\sum_{i=1}^r y_i f_i(x),\\ \tilde{L}_n(x,y) &:= F_{0,n}(x) + \textstyle\sum_{i=1}^r y_i f_i(x)\end{aligned}$$

reside in $\mathcal{K} \subset C(\mathcal{X}\times\mathcal{Y})$ formed by convex-concave on $\mathcal{X}\times\mathcal{Y}$ functions. Furthermore, Shapiro [54, Theorem 7.24] assures us that $\tilde{p}$ is Hadamard directionally differentiable at any $h_0 \in \mathcal{K}$ tangentially to the set $\mathcal{K}$ along any $u \in \mathcal{T}_{\mathcal{K}}(\tilde{L})$, where the contingent (Bouligand) cone is

$$\mathcal{T}_{\mathcal{K}}(\tilde{L}) := \left\{h \in C(\mathcal{X}\times\mathcal{Y})\colon \exists t_k \downarrow 0 \text{ with } \tilde{L}+t_k h \in \mathcal{K}\right\},$$

and the derivative at h_0 along u is given by

$$D_1\tilde{p}(h_0;u) = \inf_{x\in\mathcal{X}^*}\sup_{y\in\mathcal{Y}^*} u(x,y).$$

Now notice that $\mathcal{K}$ is convex and apply the Delta Theorem [56, Theorem 7.61] to write an expansion of $\tilde{p}(\tilde{L}_n(\cdot,s,t))$ around $(t-s)\tilde{L}$, where the linearly interpolated estimator for $\tilde{L}_n$ is defined as in Subsection 2.4:

$$\tilde{p}\big(\tilde{L}_n(\cdot,s,t)\big) = \tilde{p}\big((t-s)\tilde{L}\big) + D_1\tilde{p}\big((t-s)\tilde{L};\tilde{L}_n(\cdot,s,t)-(t-s)\tilde{L}\big) + o_p\big(1/\sqrt{n}\big)$$
$$= \tilde{p}\big((t-s)\tilde{L}\big) + \inf_{x\in\mathcal{X}^*}\sup_{y\in\mathcal{Y}^*}\big\{\tilde{L}_n((x,y),s,t)-(t-s)\tilde{L}(x,y)\big\} + o_p\big(1/\sqrt{n}\big). \quad (5.9)$$

Then,

$$\sqrt{n}\,(p_n^*(s,t)-(t-s)p^*) = \sqrt{n}\,\big(\tilde{p}\big(\tilde{L}_n(\cdot,s,t)\big)-\tilde{p}\big((t-s)\tilde{L}\big)\big)$$
$$= \inf_{x\in\mathcal{X}^*}\sup_{y\in\mathcal{Y}^*}\big\{\sqrt{n}\,(F_{0,n}(x,s,t)-(t-s)f_0(x))\big\} + o_p(1)$$
$$\xrightarrow{\text{d}} \inf_{x\in\mathcal{X}^*} Y(x),$$

where the first equality holds because of the strong duality of the sample-path and true problems, the second equality holds from (5.9), and distributional limit to the random element $Y\in C(\mathcal{X})$ holds from Assumptions 4.1 and 4.2 and since we have assumed $\mathcal{X}$ is compact. (See the functional CLT [1, Corollary 7.17].) If $\mathcal{X}^*$ is a singleton, then $\inf_{x\in\mathcal{X}^*} Y(x) = Y(x^*)$ has a Gaussian distribution with mean zero and finite variance. (See the comments that follow Assumption 4.2.) Now use the functional h, defined in (5.6), and retrace steps leading to (5.8) to get

$$\frac{p_n^*(0,1)-p^*}{\psi(p_n^*(\cdot,\cdot))} = \frac{\cancel{\sigma(x^*)}\delta W(0,1)}{\cancel{\sigma(x^*)}\psi(\delta W(\cdot,\cdot))},$$

thus proving the primal weak limit in (5.4).

The dual limit in (5.3) follows along lines similar to what we have adopted for proving the limit in (5.4), but only more simply since $y=\bar{y}$ is fixed. □

5.2. **Confidence interval construction.** Suppose $(\bar{x},\bar{y})$ is a *candidate saddle point*, that is, $\bar{x}\in\mathcal{X}$ is primal feasible, $\bar{y}\in\mathcal{Y}\cap\mathcal{Y}_n$ is dual feasible, and we wish to construct a $(1-\alpha)$ confidence interval on the optimality gap $f_0(\bar{x})-p^*$. Then, given a resampling functional ψ chosen to lie in the resampling class $\mathcal{M}_1$ (see Definition 5.1 and guidance in Section 6), Theorem 5.1 suggests the following upper and lower statistical bounds on the gap $f_0(\bar{x})-p^*$:

$$U_n := F_{0,n}(\bar{x},0,1) + \gamma_{\psi,1-\alpha_p}\,\psi\big(F_{0,n}(\bar{x},\cdot,\cdot)\big); \quad (5.10)$$
$$L_n := G_n(\bar{y},0,1) - \gamma_{\psi,\alpha_d}\,\psi\big(G_n(\bar{y},\cdot,\cdot)\big), \quad (5.11)$$

where

$$\gamma_{\psi,\alpha} := T_\psi^{-1}(\alpha) = \inf\left\{\gamma\colon \mathrm{P}\left(\frac{\delta W(0,1)}{\psi(\delta W(\cdot,\cdot))}\geq\gamma\right)\geq 1-\alpha\right\},\quad \alpha\in[0,1]. \quad (5.12)$$

The following Corollary 5.1 asserts asymptotic validity of the constructed intervals.

Corollary 5.1. *Let all postulates of Theorem 5.1 hold, and let $\alpha_p,\alpha_d\in[0,0.5)$. Then as $n\to\infty$:*

(1) $\mathrm{P}(f_0(\bar{x})\geq U_n)\to\alpha_p$;
(2) $\mathrm{P}(g(\bar{y})\leq L_n)\to\alpha_d$, *and*
(3) $\mathrm{P}(f_0(\bar{x})-p^*\geq U_n-L_n)\lesssim\alpha_p+\alpha_d$, *and*
(4) $\mathrm{P}\left(-\gamma_{\psi,1-\alpha_p}\psi(p_n^*(\cdot,\cdot))\leq p^*-p_n^*\leq\gamma_{\psi,\alpha_d}\psi(p_n^*(\cdot,\cdot))\right)\to 1-(\alpha_p+\alpha_d)$.

Proof. The proofs of Parts 1, 2, and 4 follow from the assertions in (5.2), (5.3), and (5.4), respectively. For proving part (3), notice since

$$\begin{aligned} \mathrm{P}(f_0(\bar{x}) - p^* \geq U_n - L_n) &\leq \mathrm{P}(f_0(\bar{x}) \geq U_n) + \mathrm{P}(p^* \leq L_n) \\ &\leq \mathrm{P}(f_0(\bar{x}) \geq U_n) + \mathrm{P}(g(\bar{y}) \leq L_n) \\ &\to \alpha_p + \alpha_d \text{ as } n \to \infty, \end{aligned} \tag{5.13}$$

where the second inequality above follows because $\bar{y} \in \mathcal{Y} \cap \mathcal{Y}_n$ is dual feasible, and the weak limit is from the first two parts of the theorem. □

5.3. **KKT Testing.** Suppose now that we wish to statistically check whether the first-order optimality conditions associated with (P) hold at a given candidate solution $\bar{x}$. Formally, recall that if $\bar{x}$ is an optimal solution to (P), f_0 is differentiable at $\bar{x}$, and $f_j, j = 1, 2, \ldots, r$ are smooth at $\bar{x}$, then under a constraint qualification, the first-order optimality (KKT) conditions hold at $\bar{x}$, that is, there exist Lagrange multipliers $\lambda_j \geq 0$ such that

$$\nabla f_0(\bar{x}) + \sum_{j \in \mathcal{J}(\bar{x})} \lambda_j \nabla f_j(\bar{x}) = 0,$$

where $\mathcal{J}(\bar{x}) := \{j\colon f_j(\bar{x}) = 0, j = 1, 2, \ldots, r\}$ represents the index set of active constraints at $\bar{x}$. Alternatively, we say $\bar{x}$ satisfies the KKT conditions if

$$\nabla f_0(\bar{x}) \in K(\bar{x}) := \Big\{ z \in \mathbb{R}^q \colon z + \sum_{j \in \mathcal{J}(\bar{x})} \lambda_j \nabla f_j(\bar{x}) = 0 \text{ for some } \lambda_j \geq 0 \Big\}. \tag{5.14}$$

Assuming sample-path differentiability, a natural estimator for the gradient ∇f_0 is simply the gradient $\nabla F_{0,n}$ of $F_{0,n}$. Then, under an analogue of Assumption 4.2 that assures the asymptotic linearity of the gradient sample paths, we can obtain a weak convergence limit that can be used in statistically testing (5.14).

Theorem 5.2 (Cancellation for KKT testing)**.** *Suppose that f_0 is differentiable in $\mathcal{D}$, that each $f_j, j = 1, 2, \ldots, r$ is smooth in $\mathcal{D}$, and that $\psi \in \mathcal{M}_q$. Suppose also that*

(a) $\nabla F_{0,n}$ is asymptotically linear, that is, for each $x \in \mathcal{D}$,

$$\nabla F_{0,n}(x) = \nabla f_0(x) + n^{-1} \textstyle\sum_{j=1}^n \varepsilon_1(x, \xi_j) + b_{1,n}(x),$$

where $\mathrm{E}[\varepsilon_1(x,\xi)] = 0$, $\mathrm{Cov}(\varepsilon(x,\xi))$ exists, and $\sup_{x \in \mathcal{X}} |b_{1,n}(x)| = o_p(n^{-1/2})$; and

(b) there exists measurable $\kappa_1(\xi)$ with $\mathrm{E}[\kappa_1(\xi)] < \infty$ such that $\varepsilon_1(\cdot,\xi)$ is a.s. $\kappa_1(\xi)$-Lipschitz, that is, $\forall x, x' \in \mathcal{X}$ we have that $|\varepsilon_1(x,\xi) - \varepsilon_1(x',\xi)| \leq \kappa_1(\xi)\|x - x'\|$.

For a primal feasible point $\bar{x} \in \mathcal{X}$, denoting

$$T_{\psi,n}(\bar{x}) := \left[\psi(\sqrt{n}\nabla F_{0,n}(\bar{x},\cdot,\cdot))\right]^{-1} \left\{ \sqrt{n} \left(\nabla F_{0,n}(\bar{x}, 0, 1) - \nabla f_0(\bar{x}) \right) \right\},$$

we have as $n \to \infty$,

$$T_{\psi,n}(\bar{x}) \xrightarrow{\mathrm{d}} \left[\psi(\delta W_q(\cdot,\cdot))\right]^{-1} \delta W_q(0,1). \tag{5.15}$$

Suppose further that $\bar{x}$ is a KKT point, that is, $\nabla f_0(\bar{x}) \in K(\bar{x})$. Defining

$$\chi^2_{\psi,n}(\bar{x}) := \inf_{z \in K(\bar{x})} \left\{ n \left\| \nabla F_{0,n}(\bar{x}, 0, 1) - z \right\|^2_{\left[\psi^2(\sqrt{n}\nabla F_{0,n}(\bar{x},\cdot,\cdot))\right]^{-1}} \right\},$$

we have as $n \to \infty$,

$$\chi^2_{\psi,n}(\bar{x}) \xrightarrow{\mathrm{d}} \big\|\delta W_q(0,1)\big\|^2_{[\psi^2(\delta W_q(\cdot,\cdot))]^{-1}} =: \chi^2_{\psi}. \tag{5.16}$$

Proof. The asymptotic linearity assumption of $\nabla F_{0,n}$, and the existence of the covariance $\Sigma_x := \mathrm{cov}(\varepsilon_1(x,\xi(\cdot))$ imply that the process convergence limit

$$\big\{\sqrt{n}(\nabla F_{0,n}(\bar{x},0,t) - t\nabla f_0(\bar{x})), t \in [0,1]\big\} \xrightarrow{\mathrm{d}} \Sigma_{\bar{x}}^{\frac{1}{2}} W_d.$$

It can then be shown that this process convergence limit implies

$$\begin{aligned}\big\{\sqrt{n}(\nabla F_{0,n}(\bar{x},0,t) - t\nabla f_0(\bar{x})), \sqrt{n}(\nabla F_{0,n}(\bar{x},0,s) - s\nabla f_0(\bar{x})), (s,t) \in \Delta\big\}\\ \xrightarrow{\mathrm{d}} \big\{(\Sigma_{\bar{x}}^{\frac{1}{2}} W_d(t), \Sigma_{\bar{x}}^{\frac{1}{2}} W_d(s)), (s,t) \in \Delta\big\},\end{aligned}$$

and therefore that

$$\big\{\sqrt{n}(\nabla F_{0,n}(\bar{x},s,t) - (t-s)\nabla f_0(\bar{x})), (s,t) \in \Delta\big\} \xrightarrow{\mathrm{d}} \Sigma_{\bar{x}}^{\frac{1}{2}} \delta W_d. \tag{5.17}$$

Now generalize the definition of the functional h from (5.6) so that

$$h(y) := \psi(y)^{-1} y(0,1), \quad y \in C^q(\Delta),$$

and write

$$\begin{aligned}&\big[\psi\big(\nabla F_{0,n}(\bar{x},\cdot,\cdot)\big)\big]^{-1} (\nabla F_{0,n}(\bar{x},0,1) - \nabla f_0(\bar{x}))\\ &\quad = \big[\psi\big(\sqrt{n}\big(\nabla F_{0,n}(\bar{x},\cdot,\cdot) - \nabla f_0(\bar{x},\cdot,\cdot)\big)\big)\big]^{-1} \sqrt{n}\big(\nabla F_{0,n}(\bar{x},0,1) - \nabla f_0(\bar{x},0,1)\big) &(5.18)\\ &\quad =: h\big(\sqrt{n}(\nabla F_{0,n}(\bar{x},\cdot,\cdot) - \nabla f_0(\bar{x},\cdot,\cdot))\big)\\ &\quad \xrightarrow{\mathrm{d}} h(\Sigma_{\bar{x}}^{\frac{1}{2}} \delta W_q) := \Big[\psi(\Sigma_{\bar{x}}^{\frac{1}{2}} \delta W_q(\cdot,\cdot))\Big]^{-1} \Sigma_{\bar{x}}^{\frac{1}{2}} \delta W_q(0,1) &(5.19)\\ &\quad = \big[\psi(\delta W_q(\cdot,\cdot))\big]^{-1} \cancel{\Sigma_{\bar{x}}^{-\frac{1}{2}} \Sigma_{\bar{x}}^{\frac{1}{2}}} \delta W_q(0,1), &(5.20)\end{aligned}$$

where the equality in (5.18) follows from the shift and scale invariance properties of ψ, the weak limit in (5.19) follows from the application of the mapping theorem [7, p. 20] which is allowed because $\sqrt{n}(\nabla F_{0,n}(\bar{x},\cdot,\cdot) - \nabla f_0(\bar{x},\cdot,\cdot)) \in C^q(\Delta)$ by assumption, ψ is positive-definite matrix valued and a.s. continuous, and (5.17) holds. The last equality in (5.20) follows by (again) applying the assumed scale invariance of ψ. The weak limit implied by (5.20) proves the assertion in (5.15).

To prove (5.16), we write

$$
\begin{aligned}
\chi^2_{\psi,n}(\bar{x}) &:= \inf_{z\in K(\bar{x})} \left\{ n \left\| \nabla F_{0,n}(\bar{x},0,1) - z \right\|^2_{\left[\psi^2(\sqrt{n}\nabla F_{0,n}(\bar{x},\cdot,\cdot))\right]^{-1}} \right\} \\
&= \inf_{z\in K(\bar{x})} \left\{ n \left(\nabla F_{0,n}(\bar{x},0,1) - z\right)^\top \left[\psi^2(\sqrt{n}\nabla F_{0,n}(\bar{x},\cdot,\cdot))\right]^{-1} \left(\nabla F_{0,n}(\bar{x},0,1) - z\right) \right\} \\
&= n \left\| \nabla F_{0,n}(\bar{x},0,1) - \nabla f_0(\bar{x}) \right\|^2_{\left[\psi^2(\sqrt{n}\nabla F_{0,n}(\bar{x},\cdot,\cdot))\right]^{-1}} \\
&\qquad + n \inf_{z\in K(\bar{x})} \left\{ 2\left(\nabla F_{0,n}(\bar{x},0,1) - \nabla f_0(\bar{x})\right)^\top \left[\psi^2(\sqrt{n}\nabla F_{0,n}(\bar{x},\cdot,\cdot))\right]^{-1} \left(\nabla f_0(\bar{x}) - z\right) \right. \\
&\qquad\qquad \left. + \left\| \nabla f_0(\bar{x}) - z \right\|^2_{\left[\psi^2(\sqrt{n}\nabla F_{0,n}(\bar{x},\cdot,\cdot))\right]^{-1}} \right\} \qquad (5.21) \\
&\xrightarrow{\text{d}} \left\| \delta W_q(0,1) \right\|^2_{\left[\psi^2(\delta W_q(\cdot,\cdot))\right]^{-1}} \text{ as } n\to\infty, \qquad (5.22)
\end{aligned}
$$

where the weak limit in (5.22) follows since the second term on the right-hand side of (5.21) goes to zero a. s. as $n\to\infty$, and using (5.15) with the mapping theorem [7, p. 20]. □

Notice that χ^2_ψ in (5.16) is a quadratic form in a Gaussian limit of the gradient residual; hence it is a form of generalized χ^2 that can be used to test KKT optimality. To set up this test formally, consider conducting the following hypothesis test associated with checking whether $\bar{x}\in\mathcal{X}$ is a KKT point:

$$
H_0\colon \nabla f_0(\bar{x}) \in K(\bar{x}) \quad \text{against the alternative} \quad H_1\colon \nabla f_0(\bar{x}) \notin K(\bar{x}), \tag{5.23}
$$

using the statistic

$$
\chi^2_{\psi,n}(\bar{x}) := \inf_{z\in K(\bar{x})} \left\{ n \left\| \nabla F_{0,n}(\bar{x},0,1) - z \right\|^2_{\left[\psi^2(\sqrt{n}\nabla F_{0,n}(\bar{x},\cdot,\cdot))\right]^{-1}} \right\}.
$$

Theorem 5.2 is especially useful two reasons. First, as in the more classical context that does not use resampling [54, p. 213], the test statistic $\chi^2_{\psi,n}(\bar{x})$ can be calculated efficiently since it can be seen as the optimal value associated with the problem of optimizing a convex quadratic objective (since $\psi^2(\cdot)^{-1}$ is positive definite) over the polyhedral cone $K(\bar{x})$. Second, as can be seen in (5.16), $\chi^2_{\psi,n}(\bar{x})$ has a distribution-free weak limit χ^2_ψ whose critical values can be calculated, in principle — Section 6 provides two example ψ functionals for which such critical values have been tabulated extensively. A final useful observation is that the $(1-\alpha)$ confidence region on $\nabla f_0(\bar{x})$ given by

$$
\left\{ z\in\mathbb{R}^q \colon n \left\| \nabla F_{0,n}(\bar{x},0,1) - z \right\|^2_{\left[\psi^2(\sqrt{n}\nabla F_{0,n}(\bar{x},\cdot,\cdot))\right]^{-1}} \le (\chi^2_\psi)^{-1}(1-\alpha) \right\}, \tag{5.24}
$$

is exactly the complement of the rejection region associated with the hypothesis test in (5.23). In (5.24), $(\chi^2_\psi)^{-1}(1-\alpha)$ refers to the $(1-\alpha)$ quantile associated with the weak limit χ^2_ψ.

5.4. **Stopping Theorem.** Consider now a "streaming" context where a fixed sample-path optimal sequence $\{\bar{x}_n \in \mathcal{X}^*_n, n\ge 1\}$ is revealed sequentially. This setting is hypothetical in that we assume the sample-path problems are solved to optimality, whereas this may not be possible in reality. Nevertheless, the analysis that follows sheds much light on the solution quality and effort implied by some natural stopping heuristics.

Suppose we wish to "stop" the sequence $\{\bar{x}_n, n \geq 1\}$ at some (random) time $N_{\psi,\varepsilon}$ with an appropriate guarantee on the solution $\bar{x}_{N_{\psi,\varepsilon}}$. Precisely, suppose we continue to observe the sample-path optimal sequence $\{\bar{x}_n, n \geq 1\}$ until the $(1-\alpha)$ confidence interval on the true optimal value p^* drops below a fixed threshold ε, that is,

$$N_{\psi,\varepsilon} = \inf\{n \geq \max(1, -\log\varepsilon) \colon U_n^* - L_n^* \leq \varepsilon\},$$

where

$$U_n^* = p_n^* + \gamma_{\psi,1-\alpha_p}\psi(p_n^*(\cdot,\cdot)) \quad \text{and} \quad L_n^* = p_n^* - \gamma_{\psi,\alpha_d}\psi(p_n^*(\cdot,\cdot)), \tag{5.25}$$

$\gamma_{\psi,\alpha} := T_\psi^{-1}(\alpha), \alpha \in [0,1]$, T_ψ is the distribution of $\delta W(0,1)/\psi(\delta W(\cdot,\cdot))$, and α_p, α_d are chosen so that $\alpha_p, \alpha_d < 0.5, \alpha_p + \alpha_d = \alpha$. We know from Corollary 5.1 that

$$\mathrm{P}\big(p_n^* - \gamma_{\psi,\alpha_d}\psi(p_n^*(\cdot,\cdot)) \leq p^* \leq p_n^* + \gamma_{\psi,1-\alpha_p}\psi(p_n^*(\cdot,\cdot))\big) \gtrsim 1-\alpha.$$

The following result characterizes the nature of $N_{\psi,\varepsilon}$ by asserting that the sequential procedure will stop almost surely and that the (asymptotic) complexity of $N_{\psi,\varepsilon} = O(\varepsilon^{-2})$.

Theorem 5.3 (Stopping Theorem)**.** *Suppose Assumptions 4.1 and 4.2 hold, and that $\psi \in \mathcal{M}_1$. Suppose further that the following two assumptions hold.*

(a) The sequence $\{\bar{x}_n, n \geq 1\}$ satisfies $\bar{x}_n \in \mathcal{X}_n^*$ for all n.

(b) ψ satisfies $\psi\big(\sqrt{n}\, p_n^*(\cdot,\cdot)\big) \xrightarrow{\text{wp1}} \psi\big(p^*(\cdot,\cdot)\big)$.

Then, as $\varepsilon \to 0$,

$$N_{\psi,\varepsilon} \to \infty \quad \textit{and} \quad N_{\psi,\varepsilon}\varepsilon^2 = O(1).$$

Proof. Notice that from (5.25),

$$\begin{aligned} U_n^* - L_n^* &= (\gamma_{\psi,1-\alpha_p} - \gamma_{\psi,\alpha_d})\,\psi(p_n^*(\cdot,\cdot)) \\ &= (\gamma_{\psi,1-\alpha_p} - \gamma_{\psi,\alpha_d})\,\frac{1}{\sqrt{n}}\,\psi(\sqrt{n}\,(p_n^*(\cdot,\cdot) - p^*)), \end{aligned} \tag{5.26}$$

where the second equality in (5.26) follows from the shift and scale invariance of ψ. Since the expression for $N_{\psi,\varepsilon}$ implies that $N_{\psi,\varepsilon} \geq -\log\varepsilon$, $N_{\psi,\varepsilon} \to \infty$ (trivially) as $\varepsilon \to 0$.

From (5.26) and the assumption in (b), we get that as $n \to \infty$,

$$\sqrt{n}\,(U_n^* - L_n^*) \xrightarrow{\text{wp1}} (\gamma_{\psi,1-\alpha_p} - \gamma_{\psi,\alpha_d})\psi\big(p^*(\cdot,\cdot)\big). \tag{5.27}$$

In particular, we see that (5.27) implies that for large enough n (independent of ε),

$$\frac{1}{4n}(\gamma_{\psi,1-\alpha_p} - \gamma_{\psi,\alpha_d})^2\psi^2(p^*(\cdot,\cdot)) \leq (U_n^* - L_n^*)^2 \leq \frac{4}{n}(\gamma_{\psi,1-\alpha_p} - \gamma_{\psi,\alpha_d})^2\psi^2(p^*(\cdot,\cdot)). \tag{5.28}$$

From (5.28) and since we have proved $N_{\psi,\varepsilon} \to \infty$ as $\varepsilon \to 0$, we conclude that the assertion $\lim_{\varepsilon\to 0} N_{\psi,\varepsilon}\varepsilon^2 = O(1)$ holds. □

Two observations pertaining to Theorem 5.3 are interesting. First, as is common in adaptive stopping, the lower bound $-\log \varepsilon$ in the expression for $N_{\psi,\varepsilon}$ has been introduced to coerce $N_{\psi,\varepsilon}$ to diverge without interfering with the rate at which such divergence occurs. Second, and more important, notice that nothing has been said about the quality of the solution $\bar{x}_{N_{\psi,\varepsilon}}$ at stopping. For example, since the interval $[L_n^*, U_n^*]$ is a $(1-\alpha_p-\alpha_d)$ confidence interval on p^*, and $U^*_{N_{\psi,\varepsilon}} - L^*_{N_{\psi,\varepsilon}} \leq \varepsilon$, it seems logical to wonder if $\mathrm{P}(|p^*_{N_{\psi,\varepsilon}} - p^*| \geq \varepsilon) \leq \alpha$. However, such a guarantee does not hold in general — whereas $[L_n^*, U_n^*]$ is a confidence interval on p^*, the key ingredient to enable a guarantee such as $\mathrm{P}(|p^*_{N_{\psi,\varepsilon}} - p^*| \geq \varepsilon) \leq \alpha$ is that $[U_n^*, L_n^*]$ be a *time-uniform confidence sequence* which covers p^* with probability exceeding $1-\alpha_p-\alpha_d$ *for all* $n \geq k_0$, where k_0 is a constant independent of ε. In the current context, a time-uniform confidence sequence follows after imposing the additional assumption that $\sup_{x\in\mathcal{X}} |F_{0,n}(x) - f_0(x)| \leq \phi(n,\alpha,d)$, where the underlying empirical process class has finite complexity (VC or pseudo-dimension d). See [21, 46, 58] for further detail.

6. Example Resampling Functionals

It should be clear from Subsection 5.1 that the resampling functional ψ plays a crucial role in our construction. In effect, the resampling functional ψ is the analogue of the chi random variable with ν degrees of freedom, denoted χ_ν, that appears in the formulation of the Student's t distribution in classical statistics [29, pp. 75]. In our context, ψ features prominently when constructing U_n, L_n in (5.10) and (5.11), respectively, and in particular when computing the critical value $\gamma_{\psi,q}$ in (5.12). The potential choices of ψ are vast since it is stipulated (only) by the four conditions appearing in Definition 5.1. In what follows, we describe two specific choices that have recently demonstrated good performance in a variety of settings. We provide concrete expressions in each case, and briefly discuss properties.

6.1. **OB-I resampling functional.**

The OB-I resampling functional $\psi_{\text{OB-I}}\colon C(\Delta) \to \mathbb{R}$ is a two-parameter ($\beta \in [0,1); b_\infty \in \{\mathbb{N}\cup\infty\}\setminus\{1\}$) family of functions given by:

$$\psi^2_{\text{OB-I}}(y;\beta,b_\infty) := \begin{cases} \frac{1}{1-\beta}\int_0^{1-\beta} \left(y(s,s+\beta)-y(0,1)\right)^2 ds & \beta=0, b_\infty=\infty; \\ \frac{1}{\kappa_1(\beta,b_\infty)}\frac{1}{\beta(1-\beta)}\int_0^{1-\beta} \left(y(s,s+\beta)-y(0,1)\right)^2 ds & \beta\in(0,1), b_\infty=\infty; \\ \frac{1}{\kappa_1(\beta,b_\infty)}\frac{1}{\beta}\sum_{j=1}^{b_\infty} \left(y(c_j,c_j+\beta)-y(0,1)\right)^2 & b_\infty\in\mathbb{N}\setminus\{1\}, \end{cases} \tag{OB-I}$$

where

$$\kappa_1(\beta,b_\infty) := 1-\beta; \quad c_j := (j-1)\frac{1-\beta}{b_\infty-1}.$$

The quantity $\psi_{\text{OB-I}}(y;\beta,b_\infty)$ should be interpreted as the positive square-root of $\psi^2_{\text{OB-I}}(y;\beta,b_\infty)$ appearing in (OB-I). Also, it can be shown that $\psi_{\text{OB-I}}(y;\beta,b_\infty)$ indeed satisfies the stipulations (1)–(4) of Subsection 5.1.

The expression for $\psi^2_{\text{OB-I}}(y;\beta,b_\infty)$ in (OB-I) can be understood as the (well-defined) continuous time limit of the bias-corrected squared standard deviation of overlapping batch estimates obtained from a function $y\colon \Delta \to \mathbb{R}$. To see this more concretely, consider computing $\psi_{\text{OB-I}}(F_{0,n}(\bar{x},\cdot,\cdot);\beta,b_\infty)$ and $\psi_{\text{OB-I}}(p_n^*(\cdot,\cdot);\beta,b_\infty)$ for "plugging in" the expressions of U_n and

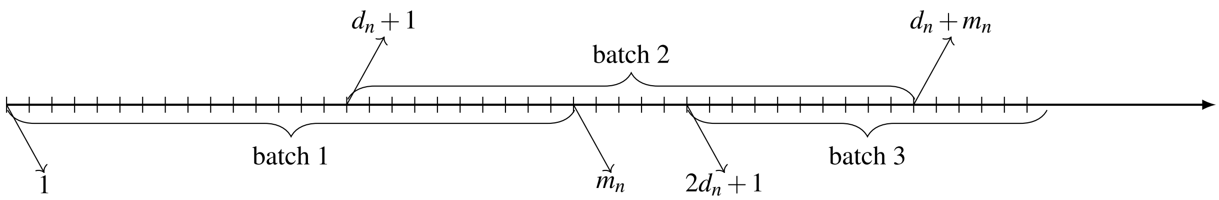

FIGURE 2. The figure, adapted from [57], depicts partially overlapping batches. Batch 1 consists of observations $X_j, j = 1,2,\ldots,m_n$; batch 2 consists of observations $X_j, j = d_n+1, d_n+2,\ldots,d_n+m_n$, and so on, where batch i consists of $X_j, j = (i-1)d_n+1,(i-1)d_n+2,\ldots,(i-1)d_n+m_n$. There are $b_n := d_n^{-1}(n-m_n)+1$ batches in total, where n is the size of the dataset.

L_n appearing in (5.10). Figure 2 depicts the process implicit in such calculation — partition the available observations $\xi_1,\xi_2,\ldots,\xi_n$ into b_n potentially overlapping batches of size m_n, with the i-th batch consisting of observations $(i-1)d_n+1,(i-1)d_n+2,\ldots,(i-1)d_n+m_n$. Choose β, b_∞ to be the asymptotic batch size and number of batches, respectively, that is:

$$\beta := \lim_{n\to\infty} \frac{m_n}{n}; \quad b_\infty := \lim_{n\to\infty} b_n.$$

For instance, we might fix $\beta \in (0,1)$ and choose $b_\infty = \beta^{-1}$. If we assume for ease of exposition that β^{-1} and $n\beta$ are integers, then this choice corresponds to partitioning the available data into $b_n = \beta^{-1}$ non-overlapping batches each having size $m_n = n\beta$, and thus (OB-I) gives

$$\psi^2_{\text{OB-I}}(F_{0,n}(\bar{x},\cdot,\cdot);\beta,b_\infty) := \frac{1}{\kappa_1(\beta,b_\infty)}\frac{1}{\beta(1-\beta)}\int_0^{1-\beta}\left(F_{0,n}(\bar{x},s,s+\beta)-F_{0,n}(\bar{x},0,1)\right)^2 ds \tag{6.1}$$

$$= \frac{1}{\kappa_1(\beta,b_\infty)}\frac{m_n}{b_n}\sum_{i=1}^{b_n}\left(F_{0,n}\left(\bar{x},(i-1)\frac{m_n}{n},i\frac{m_n}{n}\right)-F_{0,n}(\bar{x},0,1)\right)^2, \tag{6.2}$$

and

$$\psi^2_{\text{OB-I}}(p_n^*(\cdot,\cdot);\beta,b_\infty) = \frac{1}{\kappa_1(\beta,b_\infty)}\frac{m_n}{b_n}\sum_{i=1}^{b_n}\left(p_n^*\left((i-1)\frac{m_n}{n},i\frac{m_n}{n}\right)-p_n^*(0,1)\right)^2. \tag{6.3}$$

Expressions for $\psi_{\text{OB-I}}(F_{0,n}(\bar{x},\cdot,\cdot);\beta,b_\infty)$ and $\psi_{\text{OB-I}}(p_n^*(\cdot,\cdot);\beta,b_\infty)$ for other choices of $\beta \in [0,\infty), b_\infty \in \{\mathbb{N}\cup\infty\}\setminus\{1\}$ are similarly calculated with ease, modulo the effort due to the nuisance of $n\beta$ or β^{-1} not being integers. Tables and code for the α-critical value (see 5.12) specific to the OB-I resampling functional are available — see for instance [57].

6.2. **OB-II resampling functional.** The OB-I resampling functional $\psi_{\text{OB-I}}$ from Subsection 6.1 can be computationally intensive. To see this precisely, suppose $c_0(n(t-s))$ is the computational cost of calculating $\inf_{x\in\mathcal{X}} F_{0,n}(s,t)$, that is, solving a sample-path optimization problem with $n(t-s)$ observations. Then since calculating $\psi^2_{\text{OB-I}}(p_n^*(\cdot,\cdot);\beta,b_\infty)$ in (6.3) involves solving b_n optimization problems of size m_n, and one optimization problem of size n, the computation complexity becomes $b_nc_0(m_n)+c_0(n)$.

Toward mitigating the computational burden of OB-I, the OB-II resampling functional, denoted $\psi_{\text{OB-II}} \colon C(\Delta) \to \mathbb{R}$, is a two-parameter ($\beta \in [0,1); b_\infty \in \{\mathbb{N} \cup \infty\} \setminus \{1\}$) family:

$$\psi^2_{\text{OB-II}}(y;\beta,b_\infty) := \begin{cases} \frac{1}{\kappa_2(\beta,\infty)} \frac{\beta^{-1}}{1-\beta} \int_0^{1-\beta} \left(y_u(\beta) - \frac{1}{1-\beta} \int_0^{1-\beta} y_s(\beta)\, ds \right)^2 du & b_\infty = \infty; \\ \frac{1}{\kappa_2(\beta,b_\infty)} \frac{1}{\beta} \frac{1}{b_\infty} \sum_{j=1}^{b_\infty} \left(y_{c_j}(\beta) - \frac{1}{b_\infty} \sum_{i=1}^{b_\infty} y_{c_i}(\beta) \right)^2 & b_\infty \in \mathbb{N} \setminus 1, \end{cases}$$

where $y_x(\beta) := y(x+\beta) - y(x), x \in [0,1-\beta]$, $c_i := (i-1)\frac{1-\beta}{b_\infty - 1}, i = 1,2,\ldots,b_\infty$, and $\kappa_2(\beta,b_\infty)$ is the "bias-correction" factor given by

$$\kappa_2(\beta,b_\infty) := \begin{cases} 1 & \beta = 0; \\ 1 - 2\left(\frac{\beta}{1-\beta} \wedge 1\right) + \frac{1}{\beta}\left(\frac{\beta}{1-\beta} \wedge 1\right)^2 - \frac{2}{3}\frac{1-\beta}{\beta}\left(\frac{\beta}{1-\beta} \wedge 1\right)^3 & \beta > 0, b_\infty = \infty; \\ 1 - \frac{1}{b_\infty} - \frac{2}{b_\infty} \sum_{h=1}^{b_\infty} \left(1 - \frac{h}{b_\infty - 1}\frac{1-\beta}{\beta}\right)^+ \left(1 - \frac{h}{b_\infty}\right) & \beta > 0, b_\infty \in \mathbb{N} \setminus 1. \end{cases}$$

$\psi_{\text{OB-II}}(y;\beta,b_\infty)$ differs from $\psi_{\text{OB-I}}(y;\beta,b_\infty)$ mainly in its choice of the "centering" variable, the quantity about which the deviations of y or y_s are integrated. Specifically, notice that the centering variable for $\psi_{\text{OB-II}}(y;\beta,b_\infty)$ is $(1-\beta)^{-1}\int_0^{1-\beta} y_s(\beta)\, ds$ (or the analogous average when b_∞ is finite), whereas for $\psi_{\text{OB-I}}(y;\beta,b_\infty)$ the centering is simply $y(0,1)$. See [57] for more details on this calculation and an analysis of how these resampling functionals compare in terms of computation. Like $\psi_{\text{OB-I}}(y;\beta,b_\infty)$, the resampling functional $\psi_{\text{OB-I}}(y;\beta,b_\infty)$ can also be shown to satisfy the stipulations (1)–(4) of Section 5.1.

7. Concluding remarks and future research

This paper outlines methods to systematically incorporate primal-dual resampling when assessing solution quality within deterministically constrained convex stochastic programs. The key enabling mathematical machinery is a primal-dual cancellation theorem for ensuring that the resampled weak limit (whose critical values are used for confidence interval construction and for hypothesis testing) is distribution free, that is, no parameters associated with the data-generating distribution need be estimated explicitly or consistently. The primal-dual cancellation theorems characterized in this paper might provide a principled path for analogous resampling based solution quality assessment within useful but even more challenging settings like multistage stochastic programs [54]. Indeed, they may also provide a principled basis, e.g., through a hypothesis test, for checking if strong duality holds for the true problem. We have assumed throughout the paper that the constraints underlying the problem are deterministic and known. Such assumption avoids the well-recognized challenge that comes with having stochastic constraints — dealing with solutions that are sample-path dual feasible but not truly dual feasible.

References

[1] A. Araújo and E. Giné. *The Central Limit Theorem for Real and Banach Valued Random Variables*. Wiley Series in Probability and Mathematical Statistics. Wiley, New York, 1980.

[2] G. Bayraksan and D. P. Morton. A sequential sampling procedure for stochastic programming. *Operations Research*, 59(4):898–913, 2011. doi:10.1287/opre.1110.0926.

[3] G. Bayraksan and P. Pierre-Louis. Fixed-width sequential stopping rules for a class of stochastic programs. *SIAM Journal on Optimization*, 22(4):1518–1548, 2012. doi:10.1137/090773143.

[4] G. Bayraksan, D. P. Morton, and A. Partani. Simulation-based optimality tests for stochastic programs. In *Stochastic Programming: The State of the Art In Honor of George B. Dantzig*, pages 37–55. Springer, 2010. doi:10.1007/978-1-4419-1642-6_3.

[5] P. Bickel, C. A. J. Klaassen, Y. Ritov, and J. A. Wellner. *Efficient and Adaptive Estimation for Semiparametric Models*. Springer, New York, 1998.

[6] P. Billingsley. *Probability and Measure*. John Wiley and Sons, New York, 3 edition, 1995.

[7] P. Billingsley. *Convergence of Probability Measures*. Wiley Series in Probability and Statistics. Wiley-Interscience, New York, second edition, 1999.

[8] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, 2004.

[9] X. Chen and D. L. Woodruff. Distributions and bootstrap for data-based stochastic programming. *Computational Management Science*, 21(33):1–21, 2024. doi:10.1007/s10287-024-00512-3.

[10] S. Y. Chun, A. Shapiro, and S. Uryasev. Conditional value-at-risk and average value-at-risk: Estimation and asymptotics. *Operations Research*, 60(4):739–756, 2012. doi:10.1287/opre.1120.1072.

[11] O. Dowson, D. P. Morton, and A. Downward. Bi-objective multistage stochastic linear programming. *Mathematical Programming*, 196:907–933, 2022. doi:10.1007/s10107-022-01872-x.

[12] R. M. Dudley. *Uniform Central Limit Theorems*. Cambridge University Press, 1999.

[13] B. Efron and R. Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall, Inc., New York, NY, 1993.

[14] C. J. Geyer. On the asymptotics of constrained M-estimation. *Annals of Statistics*, 22(4):1993–2010, 1994. URL `https://www.jstor.org/stable/2242495`.

[15] E. Giné and R. Nickl. *Mathematical Foundations of Infinite-Dimensional Statistical Models*. Cambridge University Press, 2016.

[16] P. W. Glynn and D. I. Iglehart. Simulation output analysis using standardized time series. *Mathematics of Operations Research*, 15(1):1–16, 1990. doi:10.1287/moor.15.1.1.

[17] A. Gupta and S. R. Hunter. Properties of two-stage stochastic multi-objective linear programs. *Mathematical Programming*, 2025. doi:10.1007/s10107-025-02280-7.

[18] J. Higle and S. Sen. Stochastic decomposition: An algorithm for two-stage stochastic linear programs with recourse. *Mathematics of Operations Research*, 16:650–669, 1991. doi:10.1287/moor.16.3.650.

[19] J. L. Higle and S. Sen. Statistical verification of optimality conditions for stochastic programs with recourse. *Annals of Operations Research*, 30(1):215–239, 1991. doi:10.1007/BF02204818.

[20] J. L. Higle and S. Sen. Duality and statistical tests of optimality for two stage stochastic programs. *Mathematical Programming*, 75(2):257–275, 1996. doi:10.1007/BF02592155.

[21] S. R. Howard, A. Ramdas, J. McAuliffe, and J. Sekhon. Time-uniform, nonparametric, nonasymptotic confidence sequences. *The Annals of Statistics*, 49(2), 2021. doi:10.1214/20-aos1991.

[22] P. J. Huber and E. Ronchetti. *Robust Statistics*. Wiley Series in Probability and Statistics. Wiley, Hoboken, N.J., 2nd edition, 2009.

[23] Y. Jeon, Y. Chu, R. Pasupathy, and S. Shashaani. Uncertainty quantification using simulation output: Batching as an inferential device. *Journal of Simulation*, 2024. doi:10.1080/17477778.2024.2425311.

[24] N. M. Kiefer, T. J. Vogelsang, and H. Bunzel. Simple robust testing of regression hypotheses. *Econometrica*, 68(3):695–714, 2000. URL `https://www.jstor.org/stable/2999605`.

[25] H. Lam. On the impossibility of statistically improving empirical optimization: A second-order stochastic dominance perspective, 2021. URL https://arxiv.org/abs/2105.13419.

[26] H. Lam and H. Qian. Assessing solution quality in stochastic optimization via bootstrap aggregating. In *Proceedings of the 2018 Winter Simulation Conference*, pages 2061–2071, 2018. doi:10.1109/WSC.2018.8632334.

[27] H. Lam and H. Qian. Bounding optimality gap in stochastic optimization via bagging: Statistical efficiency and stability. *arXiv*, 2022. doi:10.48550/arXiv.1810.02905.

[28] G. Lan, A. Nemirovski, and A. Shapiro. Validation analysis of mirror descent stochastic approximation method. *Mathematical Programming*, 134(2):425–458, 2012. doi:10.1007/s10107-011-0442-6.

[29] E. L. Lehmann. *Elements of Large-Sample Theory*. Springer Texts in Statistics. Springer, New York, 1999.

[30] S. Liu, F. Bunea, and J. Niles-Weed. Asymptotic confidence sets for random linear programs. In G. Neu and L. Rosasco, editors, *Proceedings of Thirty Sixth Conference on Learning Theory*, volume 195 of *Proceedings of Machine Learning Research*, pages 3919–3940. PMLR, 12–15 Jul 2023. URL https://proceedings.mlr.press/v195/liu23d.html.

[31] I. N. Lobato. Testing that a dependent process is uncorrelated. *Journal of the American Statistical Association*, 96(455):1066–1076, 2001. doi:10.1198/016214501753208726.

[32] D. Love and G. Bayraksan. Overlapping batches for the assessment of solution quality in stochastic programs. *ACM Transactions on Modeling and Computer Simulation*, 25(3):1–20, 2015. doi:10.1145/2701421.

[33] W.-K. Mak, D. P. Morton, and R. K. Wood. Monte Carlo bounding techniques for determining solution quality in stochastic programs. *Operations Research Letters*, 24:47–56, 1999. doi:10.1016/S0167-6377(98)00054-6.

[34] D. P. Morton. Stopping rules for a class of sampling-based stochastic programming algorithms. *Operations Research*, 46(5):609–748, 1998. doi:10.1287/opre.46.5.710.

[35] V. I. Norkin, G. C. Pflug, and A. Ruszczyński. A branch and bound method for stochastic global optimization. *Mathematical Programming*, 83(1):425–450, 1998. doi:10.1007/BF02680569.

[36] R. Pasupathy. Bootstrapping and batching for output analysis. In C. G. Corlu, S. R. Hunter, H. Lam, B. S. Onggo, J. Shortle, and B. Biller, editors, *Proceedings of the 2023 Winter Simulation Conference*, pages 61–75, Piscataway, NJ, 2023. IEEE. doi:10.1109/WSC60868.2023.10407686.

[37] R. Pasupathy, P. W. Glynn, S. Ghosh, and F. Hashemi. On sampling rates in simulation-based recursions. *SIAM Journal on Optimization*, 28(1):45–73, 2018. doi:10.1137/140951679.

[38] V. Patel. Stopping criteria for, and strong convergence of, stochastic gradient descent on bottou-curtis-nocedal functions. *Mathematical Programming*, 195:693–734, 2022. doi:10.1007/s10107-021-01710-6.

[39] G. C. Pflug. Stepsize rules, stopping times and their implementation in stochastic quasigradient algorithms. In Y. Ermoliev and R. J.-B. Wets, editors, *Numerical Techniques for Stochastic Optimization*, volume 10 of *Springer Series in Computational Mathematics*, chapter 17, pages 353–372. Springer-Verlag, New York, 1988.

[40] D. N. Politis and J. P. Romano. Large sample confidence regions based on subsamples under minimal assumptions. *The Annals of Statistics*, 22(4):2031–2050, 1994. URL https://www.jstor.org/stable/2242497.

[41] D. N. Politis, J. P. Romano, and M. Wolf. *Subsampling*. Springer Series in Statistics. Springer, New York, 1999.

[42] A. Prékopa. On the probability distribution of the optimum of a random linear program. *SIAM Journal on Control*, 4(1):211–222, 1966. doi:10.1137/0304020.

[43] R. T. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2(3): 21–41, 2000. doi:10.21314/JOR.2000.038.

[44] C. Roos, T. Terlaky, and J.-P. Vial. *Interior Point Methods for Linear Optimization*. Springer, New York, NY, 2nd edition, 2005.

[45] J. O. Royset. Optimality functions in stochastic programming. *Mathematical Programming*, 135(1): 293–321, 2012. doi:10.1007/s10107-011-0453-3.

[46] D. Russo and J. Zou. How much does your data exploration overfit? Controlling bias via information usage. *IEEE Transactions on Information Theory*, 66(1):302–323, 2020. doi:10.1109/TIT.2019.2945779.

[47] A. Ruszczyński and A. Shapiro. Optimization of risk measures. In G. Calafiore and F. Dabbene, editors, *Probabilistic and Randomized Methods for Design under Uncertainty*, chapter 4, pages 119–157. Springer, London, 2006. doi:10.1007/1-84628-095-8_4.

[48] S. Sarykalin, G. Serriano, and S. Uryasev. Value-at-risk vs. conditional value-at-risk in risk management and optimization. In *TutORials in Operations Research*, pages 270–294. INFORMS, Catonsville, MD, 2008. doi:10.1287/educ.1080.0052.

[49] L. Schruben. Confidence interval estimation using standardized time series. *Operations Research*, 31(6):1090–1108, 1983. doi:10.1287/opre.31.6.1090.

[50] R. J. Serfling. *Approximation Theorems of Mathematical Statistics*. Wiley Series in Probability and Mathematical Statistics. Wiley & Sons, New York, 1980.

[51] J. Shao and D. Tu. *The Jackknife and Bootstrap*. Springer Series in Statistics. Springer, New York, NY, 1995.

[52] X. Shao. A self-normalized approach to confidence interval construction in time series. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 72(3):343–366, 2010. doi:10.1111/j.1467-9868.2010.00754.x.

[53] X. Shao. Self-normalization for time series: A review of recent developments. *Journal of the American Statistical Association*, 110(512):1797–1817, 2015. doi:10.1080/01621459.2015.1050493.

[54] A. Shapiro. Stochastic programming by Monte Carlo simulation methods. In J. L. Higle, W. Römisch, and S. Sen, editors, *Stochastic Programming E-Print Series*. Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät II, Institut für Mathematik, 2000. doi:10.18452/8226.

[55] A. Shapiro. Monte Carlo sampling methods. In A. Ruszczyński and A. Shaprio, editors, *Stochastic Programming*, Handbooks in Operations Research and Management Science, Volume 10, pages 353–425. Elsevier, Amsterdam, The Netherlands, 2003.

[56] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory*. MPS-SIAM Series on Optimization. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2009.

[57] Z. Su, R. Pasupathy, Y. Yeh, and P. W. Glynn. Overlapping batch confidence intervals on statistical functions constructed from time series: Application to quantiles, optimization, and estimation. *ACM Transactions on Modeling and Computer Simulation*, 34(2):10:1–10:43, 2024. doi:10.1145/3649437.

[58] S. van de Geer. *Empirical Processes in M-Estimation*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, UK, 2000.

[59] A. W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998.

[60] A. W. van der Vaart and J. A. Wellner. *Weak Convergence and Empirical Processes with Applications to Statistics*. Springer Series in Statistics. Springer, New York, NY, 1996.