Proceedings of the 2018 Winter Simulation Conference
M. Rabe, A. A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, eds.

RISK ASSESSMENT IN PHARMACEUTICAL SUPPLY CHAINS UNDER UNKNOWN INPUT-MODEL PARAMETERS

Alp Akcay
Tugce Martagan

School of Industrial Engineering
Eindhoven University of Technology
5612 AZ, Eindhoven, THE NETHERLANDS

Canan G. Corlu

Metropolitan College
Boston University
Boston, MA 02215, USA

ABSTRACT

We consider a pharmaceutical supply chain where the manufacturer sources a customized product with unique attributes from a set of unreliable suppliers. We model the likelihood of a supplier to successfully deliver the product via Bayesian logistic regression and use simulation to obtain the posterior distribution of the unknown parameters of this model. We study the role of so-called input-model uncertainty in estimating the likelihood of the supply failure, which is the probability that none of the suppliers in a given supplier portfolio can successfully deliver the product. We investigate how the input-model uncertainty changes with respect to the characteristics of the historical data on the past realizations of the supplier performances and the product attributes.

1 INTRODUCTION

The research and development of biological products is often outsourced by large pharmaceutical companies to smaller-scale specialized biomanufacturers (Martagan et al. 2017). The objective of the pharmaceutical company is then to decide which biological products are sourced from which biomanufacturer. A biological product has a set of physical and chemical attributes, such as purity, molecular mass, size, shape, hydrophobicity, endotoxicity, etc. (Akcay and Martagan 2016), and these attributes play an important role in determining the likelihood of a successful delivery of the product by the biomanufacturer. To be specific, if the attributes associated with the product are not satisfied within a specified time window, the biomanufacturer is assumed to have failed. Such supply failures typically have significant financial implications and, therefore, the assessment of the supply failure probability is of critical importance.

We consider the problem of estimating the supply failure probability. We assume that the probability of failure from a supplier is represented with a logistic regression model. However, the parameters of the logistic regression model are unknown to the pharmaceutical company. The estimation of the unknown parameters from the limited amounts of historical data leads to an uncertainty called *input-model uncertainty* in the simulation literature. Accounting for this uncertainty in stochastic simulations is critical and, if ignored, it could lead to suboptimal decisions (Barton 2012). The consequences of ignoring input-model uncertainty have been studied in a number of applications including queueing models and inventory management models. We refer the reader to Akcay and Corlu (2017) for recent work in this area. Our goal in this paper is to study the impact of input-model uncertainty in the estimation of the supply failure probability in a pharmaceutical supply chain.

We capture the uncertainty in the unknown logistic-regression parameters in a Bayesian fashion, and use Markov Chain Monte Carlo (MCMC) simulation to generate random samples from their posterior distribution, because this posterior distribution cannot be computed analytically. Since the logistic-regression parameters are inputs in the assessment of supply failure probability, the uncertainty in these parameters needs to be propagated to the supply failure probability, inducing a probability distribution around the supply failure

regression parameters. Section 5 presents our numerical analysis and Section 6 provides some concluding remarks with future research directions.

2 LITERATURE REVIEW

The relevant literature can be classified into three main research streams: (i) sourcing decisions under supply risk, (ii) the effect of risk attitude in sourcing decisions under supply uncertainty, and (iii) the literature on input uncertainty in stochastic simulations.

The first research stream has been studied extensively in the operations management literature. The supply risk includes both disruption risk and yield risk. Disruption risk refers to the so-called rare events such as hurricanes, tornadoes, earthquakes, and labor strikes whereas the yield risk refers to the suppliers not being able to deliver the promised products in the desired quality or quantity. Although disruption risk and yield risk are related to each other, their modeling and management require the use of different approaches. Therefore, they are considered separately in the literature. Because our focus is on yield risk in this paper, we restrict our review to yield risk in this section and refer the reader to Synder et al. (2015) for a review on supply chain disruption risk. A common assumption in this literature is that the yield distribution of the supplier is known with certainty. Tomlin (2009) relaxes this assumption by proposing a Bayesian model that updates the forecasts about the yield distribution as the decision-maker learns more about the suppliers over time. The resulting inventory management and sourcing strategies are further investigated in Pun and Heese (2014), Saghafian and Tomlin (2016), and Silbermayr and Minner (2016). Another common assumption is that the decision maker is risk-neutral.

The second research stream includes alternative risk measures in supply chains. Among the widely used risk measures are utility functions, mean-variance framework, Value-at-Risk (VaR), and Conditional Value at Risk (CVaR). Lau (1980) considers a newsvendor setting and derives optimal order quantities under two different objectives: maximizing the expected utilization and maximizing the probability of achieving a budgeted profit. Other studies that consider a newsvendor setting under a risk measure include Berman and Schnabel (1986), Choi et al. (2008), Wu et al. (2009), and Choi and Chiu (2012). The risk framework has been extended to more complex problems including multi-period inventory models and outsourcing problems. Buzacott et al. (2011) consider a procurement setting with commitment-option supply contracts under the mean-variance framework. Giri (2011) studies the risk-averse decision-maker's optimal procurement problem. The authors use an exponential utility function to measure the risk attitude of the decision maker. Xue et al. (2016) propose a diversification strategy for a risk-sensitive manufacturer with unreliable suppliers, where the risk attitude of the suppliers is captured by a mean-variance risk measure. In this paper, we are not considering the supply selection problem, but rather focus on the assessment of the supply failure probability, which is an important input to the supply selection problem. Furthermore, our Bayesian credible interval characterization corresponds to the VaR risk measure with respect to the risk of not knowing the true parameters of the logistic-regression model.

The third research stream, which considers the analysis of input uncertainty in stochastic simulations, has been widely studied in the simulation literature. We can broadly categorize the methods used to capture input uncertainty in stochastic simulations as Bayesian methods (e.g., Chick 2001; Biller and Corlu 2011; Xie et al. 2014) or frequentist methods (e.g., Ankenman and Nelson 2012; Barton et al. 2014; Song and Nelson 2015). In contrary to the frequentist methods, the Bayesian approach allows for incorporating expert opinions on unknown model inputs. In this paper, we use a Bayesian model and directly feed the posterior input-distribution parameters into simulation (instead of a using a metamodel as in Xie et al. 2014). The incorporation of risk measures into the decision-making process to hedge against the input uncertainty is a topic that has recently received attention in the simulation literature. Zhou and Xie (2015) are the first to propose a risk-adjusted framework for simulation optimization considering mean-variance, VaR, and CVaR as risk measures. Wu et al. (2018) continue the study of this problem by deriving the consistency and the asymptotical properties of the objective function and the optimal values. Zhu et al. (2015) study risk quantification in stochastic simulations under input uncertainty.

3 MODEL

We consider a pharmaceutical manufacturer who works with $N \geq 1$ suppliers to have a specific pharmaceutical product developed for its downstream manufacturing processes. Each product comes with a set of features of size d denoted by $\mathbf{x} = (x_1, \ldots, x_d) \in \mathscr{X}$ that affect a supplier's ability to successfully develop the product. The examples of these product features include purity, molecular mass, size, shape, hydrophobicity, endotoxicity, etc. (Akcay and Martagan 2016). The outcome of the product development effort by each supplier is either a success or failure. We let the binary random variable Y_n denote the outcome for supplier n; i.e., $Y_n = 1$ and $Y_n = 0$ represent a success and failure, respectively. In practice, it is common to assume that the outcome of the product development effort depends on the product features; i.e., a product with higher purity is typically more difficult to develop than a product with lower purity. Therefore, we model the failure probability of supplier n as

$$\mathbb{P}(Y_n = 0 | \beta_n, \mathbf{x}) = \frac{1}{1 + \exp(\beta_{0n} + \sum_{j=1}^{d} \beta_{jn} x_j)}, \tag{1}$$

where $\beta_n = (\beta_{0n}, \beta_{1n}, \ldots, \beta_{dn})$. Equation (1) is also known as *logistic regression*, a commonly used model in practice for binary classification (Murphy 2012).

The manufacturer can outsource the product to multiple suppliers, while it is sufficient to have only one successful supplier to be able to extract the value of the product. In other words, if none of the suppliers is successful, the manufacturer loses the opportunity to exploit the value of the product. We refer to this situation as the *supply failure*. It is of critical importance to have an accurate estimate of the probability of supply failure. The outcome of the product development effort by a supplier is independent of the other suppliers and, therefore, the probability of supply failure is given by

$$g(\beta; \mathbf{x}) \triangleq \prod_{n=1}^{N} \mathbb{P}(Y_n = 0 | \beta_n, \mathbf{x}), \tag{2}$$

where $\beta = (\beta_1; \ldots; \beta_N)$ is an $N \times (d+1)$ matrix that represents the logistic-regression parameters. The logistic-regression parameters β are estimated from historical data (i.e., the past performance of each supplier under various product attributes). Traditionally, the logistic regression models are usually fit by maximum likelihood, using the conditional likelihood of Y_n given $\mathbf{x}$. We refer the reader to Section 4.4 in Friedman et al. (2001) for further details.

We let $\mathscr{D}_n \triangleq \{(\mathbf{x}_n^t, y_n^t) : t = 1, \ldots, m_n\}$ with $\mathbf{x}_n^t = (x_{n1}^t, \ldots, x_{nd}^t)$ denote the attributes of the tth product assigned by the manufacturer to the supplier n; i.e., x_{ni}^t is the ith attribute of the tth product undertaken by the nth supplier, and y_n^t is a binary variable that takes the value of 1 if the supplier n successfully develops the product t and zero otherwise. Let $\mathscr{D} = \{\mathscr{D}_1, \ldots, \mathscr{D}_n\}$ denote the collection of all the historical data available to the manufacturer to make an inference on the unknown logistic-regression parameters β. In practice, the finiteness of the historical data often leads to an uncertainty in the estimate of the parameters β. Therefore, using only the point estimates of the parameters in Equation (2) can lead to poor decisions if they are based on the estimate of the probability of supply failure.

One approach to capture the uncertainty around the unknown logistic-regression parameters β is to adopt a Bayesian point of view and to propagate the uncertainty in the posterior distribution of β into the performance-measure estimate. In the stochastic simulation literature, where the performance measures are estimated via simulation (e.g., discrete-event simulation), Xie et al. (2014), Akcay and Biller (2017) and Akcay and Martagan (2017) are some of the recent examples. We note that the evaluation of the supply failure probability $g(\beta; \mathbf{x})$ in (2) does not itself require simulation because the failure probability of each supplier is available in closed form by Equation (1). In this paper, we use simulation to generate random samples from the posterior distribution of β and build a probability distribution for the supply failure $g(\beta; \mathbf{x})$, which we refer as the *input-model uncertainty*. In other words, we focus only on the input-model uncertainty and do not consider any intrinsic simulation uncertainty. If the intrinsic simulation uncertainty was present (i.e., when the random samples of Y_n were used to estimate the failure probability in Equation 1), then a nested Monte Carlo simulation approach is needed to quantify the role of input-model uncertainty and the intrinsic simulation uncertainty on estimating the mean simulation response.

In order to obtain the posterior distribution of the logistic-regression parameters β_n associated with the nth supplier having the historical data $\mathscr{D}_n$, we first pick a prior $\pi_0(\beta_n)$ that represents the initial belief of the manufacturer about β_n. Since the likelihood of the data $\mathscr{D}_n$ is given by

$$\prod_{t=1}^{m_n} \mathbb{P}(Y_n^t = 1|\boldsymbol{\beta}_n, \mathbf{x}_n^t)^{y_n^t}\, \mathbb{P}(Y_n^t = 0|\boldsymbol{\beta}_n, \mathbf{x}_n^t)^{1-y_n^t},$$

by Bayesian updating, we obtain the posterior distribution of β_n as

$$\pi(\beta_n|\mathscr{D}_n) \propto \pi_0(\beta_n) \prod_{t=1}^{m_n} \left(\frac{\exp\left(\beta_{0n} + \sum_{j=1}^{d} \beta_{jn} x_{ni}^t\right)}{1+\exp\left(\beta_{0n} + \sum_{j=1}^{d} \beta_{jn} x_{ni}^t\right)} \right)^{y_n^t} \left(\frac{1}{1+\exp\left(\beta_{0n} + \sum_{j=1}^{d} \beta_{jn} x_{ni}^t\right)} \right)^{1-y_n^t}, \tag{3}$$

where the notation $\propto$ denotes equivalence up to a normalization constant. It is well known that the normalization constant for the posterior in this model is analytically intractable (Murphy 2012). Furthermore, unlike the linear regression, there is no convenient conjugate prior for β_n in the logistic regression. Therefore, we use a Markov Chain Monte Carlo (MCMC) approach to approximate the posterior distribution $\pi(\beta_n|\mathscr{D}_n)$ via sampling random realizations from this distribution (Section 4).

4 QUANTIFYING THE INPUT UNCERTAINTY IN THE SUPPLY FAILURE PROBABILITY

In this section, we address how the uncertainties around the logistic-regression parameters $\beta = (\beta_1, \ldots, \beta_N)$ translate into the assessment of the probability of supply failure $g(\beta; \mathbf{x})$. Since we take a Bayesian approach, $\boldsymbol{\beta}_n$ is treated as a vector of random variables whose distribution represents the current belief on the likely values of $\boldsymbol{\beta}_n$ for the nth supplier. The MCMC approach allows us for generating random samples from this posterior distribution even though we cannot compute this posterior distribution analytically (Section 4.1). We then use these random samples to quantify the input-model uncertainty in the supply failure probability (Section 4.2).

4.1 Posterior Distribution of the Logistic-Regression Parameters

In Figure 2, we present how a random sample from the posterior distribution of the logistic-regression parameters β_n is generated. The idea behind the MCMC approach is to generate a sequence of realizations of $\boldsymbol{\beta}_n$ such that the stationary distribution of this sequence is the posterior distribution $\pi(\boldsymbol{\beta}_n|\mathscr{D}_n)$; we refer the reader to Andrieu et al. (2003) for a survey on the MCMC algorithms. Following Akcay and Martagan (2016), we adapt the slice sampling algorithm in Neal (2003), which only requires the unnormalized posterior in (3) without specifying a proposal or marginal distribution. The intuition behind slice sampling is based on the observation that to sample a random variable one can sample uniformly from the region (or more specifically, a hyperrectangle slice) under the graph of its density function.
The width parameters in the algorithm presented in Figure 2 represent the length of the interval around the current sample. The algorithm begins with this interval and searches for an appropriate region containing the points of posterior density that evaluate to a large enough value. We refer the reader to Neal (2003) for the details on how to select the width parameters. In our implementation of the algorithm in Figure 2, we choose the initial point β_n^0 and the width parameters randomly from a specified region (see Section 5).

It is also critical to verify that the Markov Chain $\{\boldsymbol{\beta}_n^s : s = 0, 1, \ldots\}$ converges to its stationary distribution. We observe in the marginal trace plots that the stationarity is always achieved for the values of s greater than 1000. Thus, we assume that the length of the burn-in period is 1000; i.e., the samples during the burn-in period are discarded. To reduce the serial autocorrelation in the samples, we set the thinning parameter equal to 10; i.e., we collect the samples at every 10 iterations of the algorithm in Figure 2.

Inputs: (i) Unnormalized posterior density in Equation (3); let $\hat{\pi}(\beta)$ denote this density. (ii) The current point of β_n denoted by $\beta_n^s = (\beta_{0n}^s, \beta_{1n}^s \dots, \beta_{dn}^s)$. (iii) The width parameters $\mathbf{w} = (w_0, w_1 \dots, w_d)$.
Output: The new point $\beta_n^{s+1} = (\beta_{0n}^{s+1}, \beta_{1n}^{s+1} \dots, \beta_{dn}^{s+1})$.
Step 1: Generate a random variate $u \sim \text{Uniform}(0, \hat{\pi}(\beta_n^s)$.
Step 2: Randomly position the hyperrectangle $H = (L_0, R_0) \times (L_1, R_1) \times \dots \times (L_d, R_d)$:
    $U_i \leftarrow \text{Uniform}(0,1)$; $L_i \leftarrow \beta_{in}^s - w_i U_i$; and $R_i \leftarrow L_i + w_i$ for $i = 0, \dots, d$.
Step 3: Sample from H and change its size when a new sample is rejected:
    Repeat:
    $U_i \leftarrow \text{Uniform}(0,1)$; and $\beta_{in}^{s+1} \leftarrow L_i + U_i(R_i - L_i)$ for $i = 0, \dots, d$.
    if $u < \hat{\pi}(\beta_n^{s+1})$ **then** Exit loop
    end if
    for $i = 0$ to d **do**
        if $\beta_{in}^{s+1} < \beta_{in}^s$ **then** $L_i \leftarrow \beta_{in}^{s+1}$
        else $R_i \leftarrow \beta_{in}^{s+1}$
        end if
    end for

Figure 2: Sampling from the posterior distribution of the logistic-regression parameters β_n.

4.2 Quantification of the Input-Model Uncertainty

We quantify the uncertainty in the supply failure probability $g(\beta;\mathbf{x})$ by propagating the uncertainty in the logistic-regression parameters β into $g(\beta;\mathbf{x})$. This induces a complete probability distribution for $g(\beta;\mathbf{x})$; we capture this distribution through a Bayesian credible interval (Xie et al. 2014). More specifically, we construct a $(1-\alpha)100\%$ credible interval $[q_L, q_U]$ for $g(\beta;\mathbf{x})$ such that the *probability content*, which is defined as $\mathbb{P}(g(\beta;\mathbf{x}) \leq q_U|\mathscr{D}) - \mathbb{P}(g(\beta;\mathbf{x}) \leq q_L|\mathscr{D})$, is equal to $1-\alpha$. We notice that there is not a unique credible interval meeting this requirement. Therefore, we focus on two-sided and equal-tail probability $(1-\alpha)100\%$ credible intervals in the remainder of the paper. More specifically, we let $q_L \triangleq q_{\alpha/2} = \min\{q : \mathbb{P}(g(\beta;\mathbf{x}) \leq q) \geq \alpha/2\}$ and $q_U \triangleq q_{1-\alpha/2} = \min\{q : \mathbb{P}(g(\beta;\mathbf{x}) \leq q) \geq 1-\alpha/2\}$. As we cannot directly evaluate the posterior distribution of $g(\beta;\mathbf{x})$, we obtain a Monte-Carlo estimate of this credible interval as outlined in Figure 3.

Inputs: Risk level α; confidence level on the estimated values of q_L and q_U; the feature vector of the current product $\mathbf{x}$; the number of samples R to generate to form the distribution of $g(\beta;\mathbf{x})$.
Output: The credible interval $[q_L, q_U]$.
for $r = 1$ to R **do**
    Generate the rth sample β_n^r from the posterior distribution of β_n^r for supplier $n \in \{1, \dots, N\}$.
    $\tau_r \leftarrow g(\beta^r;\mathbf{x})$ where $\beta^r = (\beta_1^r, \dots, \beta_N^r)$.
end for
Estimate the density of $g(\beta;\mathbf{x})$; let $\hat{f}(\cdot)$ denote this density.
Let ω denote $\sqrt{\alpha(1-\alpha)}/\sqrt{R}\hat{f}(\tau_{(\lceil R\alpha/2 \rceil)})$ where $\tau_{(r)}$ denotes the rth smallest value in the set $\{\tau_r : r = 1, 2, \dots, R\}$.
$q_L \leftarrow \tau_{(\lceil R\alpha/2 \rceil)} - z_{\frac{\gamma}{2}}\omega$ and $q_U \leftarrow \tau_{(\lceil R(1-\alpha/2) \rceil)} + z_{1-\frac{\gamma}{2}}\omega$ where $z_{\alpha/2} = \Phi^{-1}(\alpha/2)$ and $\Phi(\cdot)$ is the cdf of a standard normal random variable.

Figure 3: Estimation of the credible interval for the supply failure probability.

In Figure 3, the risk level α represents the sensitiveness to the input-model uncertainty in the assessment of the supply failure probability. For instance, if $\alpha = 0.05$, the risk assessment becomes more conservative as the length of the credible interval (i.e., the likely values of the true supply failure probability) increases. On the other hand, if $\alpha = 0.5$, the risk assessment becomes less conservative as the length of the credible interval decreases (i.e., less importance is given to the tails of the probability distribution of $g(\beta;\mathbf{x})$). Steps $7-9$ in Figure 3 follow from the asymptotic normality of the estimator of the qth quantile of $g(\beta;\mathbf{x})$; i.e., if $\hat{v}_q^R$ is the $(\lceil Rq \rceil)$th largest observation of R independent samples from $g(\beta;\mathbf{x})$ and $\mathbb{P}(g(\beta;\mathbf{x}) \leq v_q) = q$, then $\sqrt{R}(\hat{v}_q^R - v_q)$ converges in distribution to $(\sqrt{\alpha(1-\alpha)}/f(v_q))Z$ as $R \to \infty$ where Z represents a standard normal random variable and $f(\cdot)$ is the density of $g(\beta;\mathbf{x})$ (Hong et al. 2014). A standard method to approximate $f(\cdot)$ is to use the kernel density estimation (Wasserman 2006).

It is of practical importance to know how the input-model uncertainty is affected from the characteristics of the products in the historical data. To capture the similarity of the current product and the old products, we define *resemblance* as $\|\mathbf{x} - \bar{x}_n\|$ where $\|\cdot\|$ is the Euclidian norm, $\bar{\mathbf{x}}_n = (\bar{x}_{n1}, \ldots, \bar{x}_{nd})$ and $\bar{x}_{nd} = \sum_{t=1}^{m_n} x_{ni}^t / m_n$. We choose the Euclidian norm as it gives the intuitive notion of the length of the vector $\mathbf{x} - \bar{x}_n$. Another important characteristic of the historical product data is the dispersion of the product features; i.e., the extent of the variety of the product features developed by the supplier. To capture this characteristic, we define *dispersion* as $\sum_{i=1}^d w_i s_i$ where $s_i = \sqrt{\sum_{t=1}^{m_n} (x_{ni}^t - \bar{x}_{ni})^2 / (m_n - 1)}$ is the standard deviation of the ith feature of the products developed by the nth supplier and w_i is the user-specified weight factor to capture the relative importance of the ith feature. Notice that other representations of the notions of resemblance and dispersion are possible. We leave it as a future research to investigate such possible representations and their impact on the input-model uncertainty in the assessment of the supply failure probability.

5 NUMERICAL RESULTS

In our numerical experiments, we let $N = 1$ and $d = 1$. The feature variable is assumed to take continuous values in the domain (-1,1). We divide the feature domain by $\Delta \in \{2^i : i = 0, 1, 2, \ldots, 8\}$ and randomly generate the feature of the m historical products from one of the Δ slots. Having Δ too small (large) assures that the feature variable of the historical products are generated from a larger (smaller) interval, implying that the dispersion of the feature variable of the historical products are high (low). We generate the feature variable of the current product uniformly from its domain and calculate the resemblance of the current product feature to the features of the products in the historical data. In total, we generate 50,000 random instances and, in each instance, the true values of the logistic-regression parameters are generated from a uniform distribution between -2 and 2. The prior information on the distribution of each parameter is assumed to be normal with mean 0 and standard deviation σ_0.

Figure 4 presents how the credible interval (CrI) width (i.e., $q_U - q_L$) changes as a function of the resemblance and dispersion values generated at these instances. We fit a third-order polynomial surface to better reflect the impact of resemblance and dispersion on the CrI width. Figure 4 shows that (i) the input-model uncertainty (i.e., measured by the CrI width) becomes smaller as the current product characteristic gets closer to the characteristics of the historical product and (ii) for a large number of historical products, the input-model uncertainty becomes larger as the dispersion decreases at low resemblance (i.e., when it is closer to -1 or 1). On the other hand, the input-model uncertainty does not necessarily increase as the dispersion decreases at low resemblance when the number of historical products is small. This observation provides an answer to our first research question listed in Section 1 on whether and when the quantification of input uncertainty is necessary. Essentially, Figure 4 shows that input uncertainty is especially a relevant problem in the estimation of the supply-failure probability when the product features of the current product deviate considerably from the features of the products in the historical data and an attribute varies a lot from one product to another in the historical data.

Figure 4 only considers the CrI width as the measure of input-model uncertainty. In Tables 1 and 2, we further report the probability content (PC) as a measure of validity in the input-model uncertainty quantification. To be specific, in principle, as the number of historical data increases, PC must converge

(a) $m = 10$

(b) $m = 40$

(c) $m = 80$

(d) $m = 160$

Figure 4: $\sigma_0 = 1$ and $\alpha = 0.1$.

to $1-\alpha$. A higher (lower) PC is an indication that the CrI width is larger (smaller) than it is supposed to be. In Tables 1 and 2, we investigate how the deviation from $1-\alpha$ is affected by the dispersion and resemblance values. For ease in presentation, we group the dispersion values in the 50,000 instances into low, medium, and high, captured by DI, DII, and DIII, respectively. More specifically, DI corresponds to the instances with $\Delta \in \{1,2,4\}$, DII corresponds to the instances with $\Delta \in \{8,16,32\}$, and DIII corresponds to the instances with $\Delta \in \{64,128,256\}$. We group the resemblance values in the 50,000 instance into three groups as well: RI represents the instances where the resemblance is less than the 20% quantile of all the resemblance values, RII represents the instances where the resemblance is within the 40%–60% quantile of all the resemblance values, and, finally, RIII represents the instances where the resemblance is greater than the 80% quantile of all the resemblance values.

Table 1 provides an answer to our second research question listed in Section 1 on how the historical product data characteristics affect the input-model uncertainty quantified in terms of Bayesian credible intervals. Specifically, Table 1 shows that, for a given value of *m*, the CrI width is much smaller for RII compared to RI and RIII while having a PC closer to 0.95. This means that if the current product is similar to the products in the historical data, the input-model uncertainty is lower. Table 1 also shows that, as *m* increases, the CrI width typically becomes much larger in DIII compared to DI. This indicates that as the feature variable of the historical product becomes less dispersed (i.e., the products developed by the supplier are more similar), the input-model uncertainty continues to be significant even for large values of *m*. On the other hand, if the products developed by the supplier become more dispersed, the input-model uncertainty decreases much faster with the increasing number of products.

Table 1: The effect of the number of historical products (n) on the credible interval (CrI) width and the probability content (PC) for $\sigma_0 = 1$ and $\alpha = 0.025$.

(a) $m = 10$

		RI		RII		RIII	
DI	CrI width	0.717	± 0.014	0.479	± 0.007	0.707	± 0.012
	PC	0.950	± 0.031	0.956	± 0.026	0.932	± 0.036
DII	CrI width	0.741	± 0.010	0.453	± 0.009	0.748	± 0.012
	PC	0.979	± 0.019	0.995	± 0.010	0.956	± 0.027
DIII	CrI width	0.733	± 0.012	0.468	± 0.008	0.733	± 0.013
	PC	0.904	± 0.038	0.978	± 0.020	0.948	± 0.028

(b) $m = 40$

		RI		RII		RIII	
DI	CrI width	0.615	± 0.020	0.272	± 0.006	0.568	± 0.022
	PC	0.873	± 0.046	0.952	± 0.027	0.978	± 0.021
DII	CrI width	0.715	± 0.012	0.275	± 0.007	0.686	± 0.016
	PC	0.938	± 0.027	0.971	± 0.023	0.947	± 0.027
DIII	CrI width	0.705	± 0.016	0.268	± 0.007	0.719	± 0.013
	PC	0.922	± 0.037	0.946	± 0.028	0.929	± 0.031

(c) $m = 80$

		RI		RII		RIII	
DI	CrI width	0.494	± 0.026	0.206	± 0.005	0.562	± 0.0225
	PC	0.897	± 0.044	0.930	± 0.033	0.926	± 0.0361
DII	CrI width	0.689	± 0.015	0.208	± 0.005	0.717	± 0.0144
	PC	0.935	± 0.029	0.974	± 0.021	0.895	± 0.037
DIII	CrI width	0.684	± 0.018	0.199	± 0.005	0.715	± 0.0146
	PC	0.907	± 0.038	0.930	± 0.033	0.922	± 0.0357

(d) $m = 160$

		RI		RII		RIII	
DI	CrI width	0.441	± 0.025	0.151	± 0.003	0.458	± 0.0231
	PC	0.928	± 0.033	0.991	± 0.013	0.924	± 0.0358
DII	CrI width	0.659	± 0.016	0.154	± 0.005	0.699	± 0.0164
	PC	0.931	± 0.034	0.946	± 0.030	0.947	± 0.0271
DIII	CrI width	0.666	± 0.018	0.154	± 0.005	0.714	± 0.0146
	PC	0.940	± 0.031	0.996	± 0.008	0.920	± 0.0365

Table 2: The effect of the risk level (α) on the credible interval (CrI) width and the probability content (PC) for $\sigma_0 = 1$ and $m = 10$.

(a) $\alpha = 0.05$

		RI		RII		RIII	
DI	CrI width	0.639	± 0.014	0.409	± 0.006	0.627	± 0.012
	PC	0.864	± 0.048	0.895	± 0.038	0.854	± 0.050
DII	CrI width	0.664	± 0.011	0.387	± 0.008	0.672	± 0.012
	PC	0.945	± 0.029	0.975	± 0.021	0.948	± 0.029
DIII	CrI width	0.658	± 0.012	0.401	± 0.008	0.656	± 0.013
	PC	0.862	± 0.044	0.914	± 0.037	0.873	± 0.041

(b) $\alpha = 0.10$

		RI		RII		RIII	
DI	CrI width	0.530	± 0.014	0.325	± 0.005	0.520	± 0.012
	PC	0.798	± 0.056	0.778	± 0.052	0.724	± 0.063
DII	CrI width	0.554	± 0.011	0.307	± 0.007	0.562	± 0.012
	PC	0.817	± 0.050	0.892	± 0.043	0.838	± 0.048
DIII	CrI width	0.548	± 0.012	0.320	± 0.006	0.548	± 0.013
	PC	0.778	± 0.053	0.833	± 0.049	0.765	± 0.053

(c) $\alpha = 0.25$

		RI		RII		RIII	
DI	CrI width	0.306	± 0.010	0.180	± 0.003	0.301	± 0.008
	PC	0.475	± 0.070	0.500	± 0.062	0.432	± 0.070
DII	CrI width	0.324	± 0.008	0.168	± 0.004	0.330	± 0.009
	PC	0.528	± 0.064	0.483	± 0.069	0.467	± 0.065
DIII	CrI width	0.322	± 0.009	0.175	± 0.004	0.323	± 0.009
	PC	0.460	± 0.063	0.550	± 0.066	0.406	± 0.061

In Table 2, we investigate the effect of the risk level α on the CrI width and the PC. Since the risk level $1-\alpha$ can be interpreted as the sensitiveness to the input-model uncertainty, we observe that, as α decreases, the CrI width increases in order to ensure that the true supply failure probability is included in the CrI. We note that the aforementioned effect of the resemblance on the input-model uncertainty continues to hold for the values of the risk level $\alpha \in \{0.05, 0.1, 0.25\}$.

6 CONCLUSION

We consider a pharmaceutical supply chain where the manufacturer sources a customized product with unique attributes. Different from a classical random yield setting, the attributes of the products determine the likelihood of the successful development of the products by the suppliers. We model the supply failure probability as Bayesian logistic regression and use simulation to obtain the posterior distribution of the unknown parameters of this model. We define two characteristics of the historical products, namely, the dispersion and resemblance, and investigate their impact on the input-model uncertainty due to not knowing the true values of the logistic-regression parameters.

ACKNOWLEDGMENTS

The second author is grateful to the generous support of Marie Sklodowska-Curie Fellowship by the European Commission under the Horizon 2020 program, and the Innovational Research Incentives Scheme VENI grant by The Netherlands Organisation for Scientific Research (NWO).

REFERENCES

Akcay, A., and B. Biller. 2017. "Input Uncertainty in Stochastic Simulations in the Presence of Dependent Discrete Input Variables". *Journal of Simulation* https://doi.org/10.1057/s41273-017-0051-3.

Akcay, A., and C. G. Corlu. 2017. "Simulation of Inventory Systems with Unknown Input Models: A Data-Driven Approach". *International Journal of Production Research* 55(19):5826–5840.

Akcay, A., and T. Martagan. 2016. "Stochastic Simulation under Input Uncertainty for Contract-Manufacturer Selection in Pharmaceutical Industry". In *Proceedings of the 2016 Winter Simulation Conference*, edited by T. M. K. Roeder et al., 2292–2303. Piscataway, New Jersey: IEEE.

Akcay, A., and T. Martagan. 2017. "Simulation-based Production Planning for Engineer-to-Order Systems with Random Yield". In *Proceedings of the 2017 Winter Simulation Conference*, edited by W. K. V. Chan et al., 3275–3286. Piscataway, New Jersey: IEEE.

Andrieu, C., N. De Freitas, A. Doucet, and M. I. Jordan. 2003. "An Introduction to MCMC for Machine Learning". *Machine Learning* 50(1-2):5–43.

Ankenman, B., and B. L. Nelson. 2012. "A Quick Assessment of Input Uncertainty". In *Proceedings of the 2012 Winter Simulation Conference*, edited by C. Laroque et al., 241–250. Piscataway, New Jersey: IEEE.

Barton, R. 2012. "Tutorial: Input Uncertainty in Output Analysis". In *Proceedings of the 2012 Winter Simulation Conference*, edited by C. Laroque et al., 67–78. Piscataway, New Jersey: IEEE.

Barton, R., B. Nelson, and W. Xie. 2014. "Quantifying Input Uncertainty via Simulation Confidence Intervals". *INFORMS Journal on Computing* 26(1):74–87.

Berman, O., and J. A. Schnabel. 1986. "Mean-Variance Analysis and the Single-Period Inventory Problem". *International Journal of Systems Science* 17(8):1145–1151.

Biller, B., and C. Corlu. 2011. "Accounting for Parameter Uncertainty in Large-Scale Stochastic Simulations with Correlated Inputs". *Operations Research* 59(3):661–673.

Buzacott, J., H. Yan, and H. Zhang. 2011. "Risk Analysis of Commitment-Option Contracts with Forecast Updates". *IIE Transactions* 43(6):415–431.

Chick, S. 2001. "Input Distribution Selection for Simulation Experiments: Accounting for Input Uncertainty". *Operations Research* 49(5):744–758.

Choi, T.-M., and C.-H. Chiu. 2012. "Mean-Downside-Risk and Mean-Variance Newsvendor Models: Implications for Sustainable Fashion Retailing". *International Journal of Production Economics* 135(2):552–560.

Choi, T.-M., D. Li, and H. Yan. 2008. "Mean–Variance Analysis for the Newsvendor Problem". *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 38(5):1169–1180.

Friedman, J., T. Hastie, and R. Tibshirani. 2001. *The Elements of Statistical Learning*, Volume 1. New York: Springer.

Giri, B. 2011. "Managing Inventory with Two Suppliers Under Yield Uncertainty and Risk Aversion". *International Journal of Production Economics* 133(1):80–85.

Hong, L. J., Z. Hu, and G. Liu. 2014. "Monte Carlo Methods for Value-at-Risk and Conditional Value-at-Risk: A Review". *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 24(4):22.

Lau, H.-S. 1980. "The Newsboy Problem Under Alternative Optimization Objectives". *Journal of the Operational Research Society* 31(6):525–535.

Martagan, T., A. Krishnamurthy, P. A. Leland, and C. T. Maravelias. 2017. "Performance Guarantees and Optimal Purification Decisions for Engineered Proteins". *Operations Research* 66(1):18–41.

Murphy, K. P. 2012. *Machine Learning: A Probabilistic Perspective*. Cambridge: MIT press.

Neal, R. M. 2003. "Slice Sampling". *Annals of Statistics* 31(3):705–767.

Pun, H., and H. S. Heese. 2014. "Outsourcing to Suppliers with Unknown Capabilities". *European Journal of Operational Research* 234(1):108–118.

Saghafian, S., and B. Tomlin. 2016. "The Newsvendor under Demand Ambiguity: Combining Data with Moment and Tail Information". *Operations Research* 64(1):167–185.

Silbermayr, L., and S. Minner. 2016. "Dual Sourcing under Disruption Risk and Cost Improvement Through Learning". *European Journal of Operational Research* 250(1):226–238.

Song, E., and B. L. Nelson. 2015. "Quickly Assessing Contributions to Input Uncertainty". *IIE Transactions* 47(9):893–909.

Synder, L., Z. Atan, P. Peng, Y. Rong, A. J. Schmitt, and B. Sinsoysal. 2015. "OR/MS Models for Supply Chain Disruptions: A Review". *IIE Transactions* 48(2):89–109.

Tomlin, B. 2009. "Impact of Supply Learning When Suppliers are Unreliable". *Manufacturing & Service Operations Management* 11(2):192–209.

Wasserman, L. A. 2006. *All of Nonparametric Statistics*. New York: Springer.

Wu, D., H. Zhu, and E. Zhou. 2018. "A Bayesian Risk Approach to Data-driven Stochastic Optimization: Formulations and Asymptotics". *SIAM Journal on Optimization* 28(2):1588–1612.

Wu, J., J. Li, S. Wang, and T. Cheng. 2009. "Mean–Variance Analysis of the Newsvendor Model with Stockout Cost". *Omega* 37(3):724–730.

Xie, W., B. L. Nelson, and R. R. Barton. 2014. "A Bayesian Framework for Quantifying Uncertainty in Stochastic Simulation". *Operations Research* 62(6):1439–1452.

Xue, W., T.-M. Choi, and L. Ma. 2016. "Diversification Strategy with Random Yield Suppliers for a Mean–Variance Risk-Sensitive Manufacturer". *Transportation Research Part E: Logistics and Transportation Review* 90:90–107.

Zhou, E., and W. Xie. 2015. "Simulation Optimization When Facing Input Uncertainty". In *Proceedings of the 2015 Winter Simulation Conference*, edited by L. Yilmaz et al., 3714–3724. IEEE.

Zhu, H., T. Liu, and E. Zhou. 2015. "Risk Quantification in Stochastic Simulation Under Input Uncertainty". *arXiv preprint arXiv:1507.06015*.

AUTHOR BIOGRAPHIES

ALP AKCAY is an Assistant Professor in the School of Industrial Engineering at Eindhoven University of Technology. He holds a Ph.D. in Operations Management from the Tepper School of Business at Carnegie Mellon University. His research interests include statistical decision making under uncertainty, simulation design and analysis, and approximate dynamic programming with applications in manufacturing, maintenance, and supply chain management. His email address is a.e.akcay@tue.nl.

TUGCE MARTAGAN is an Assistant Professor in the School of Industrial Engineering at Eindhoven University of Technology. She holds a Ph.D. in Industrial Engineering from the University of Wisconsin-Madison. Her research interests include stochastic modeling and optimization with applications in engineer-to-order systems. She is the recipient of the first prize in 2016 POMS Applied Research Challenge, and the first prize in 2017 INFORMS TIMES Dissertation Award. Her email address is t.g.martagan@tue.nl.

CANAN G. CORLU is an Assistant Professor at Administrative Sciences Department of the Metropolitan College at Boston University. Her research interests are in the areas of design and analysis of stochastic simulations with applications to operations management. Her email address is canan@bu.edu.