On the Optimistic Bias of Core Damage Frequency,

and why it should be excluded from nuclear regulatory requirements.

Martin Wortman,* Ernie Kee, and Pranav Kannan

HazTechRisk.Org

November 14, 2021

Abstract

Core Damage Frequency (*CDF*) is a risk metric employed by nuclear regulatory bodies worldwide. Numerical values for this metric are required by U.S. regulators, prior to reactor licensing, and reported values can trigger regulatory inspections. *CDF* is reported as a constant, sometimes accompanied by a confidence interval. It is well understood that *CDF* characterizes the arrival rate of a stochastic point process modeling core damage events. However, consequences of the assumptions imposed on this stochastic process as a computational necessity are often overlooked. Herein, we revisit *CDF* in the context of modern counting processes. We will argue that the assumptions required to obtain a numerical value for *CDF* (*e.g.,* with Probabilistic Risk Assessment (*PRA*)) are typically unrealistic and lead to an underestimate bias. We will conclude that:

1. The computation of core damage frequency (CDF) cannot account for the stochastic dependence between reactor protections and arriving initiating events.

---

*corresponding author, wortman@haztechrisk.org

2. In practice, there can be no direct observational data reflecting the propensity for core damage.

3. CDF calculated via PRA or PSA is optimistically biased.

4. The optimistic bias of *CDF* should disqualify its use in regulatory oversight.

Keywords— PRA, PSA, CDF, Core Damage Frequency, counting processes, martingales

1   Overview

CDF has a long history as a risk metric for the U.S. civilian nuclear fleet. It is a data–derived constant intended to provide insight as to the likelihood of core damage events. *CDF* is used to report both historical fleet performance and to predict performance of individual reactor units. As an historical performance metric, *CDF* is simply the quotient of total industry core damage events and total reactor years of operation. For example, see **?** for such history up to 2014. But, risk is concerned with future operational behaviors. Since any practical reactor will experience at most one core damage event, 'frequency' is an unnatural concept that must be understood within the context of modeling used to support the computation of a predictive *CDF*.

A large body of literature is devoted to computing estimates of *CDF*; most of this work extends from **?**. The central theme of this corpus focuses on how to analytically connect possible core damage with the status of reactor protections and the arrival of initiating events that, if unmitigated by protections, might exceed to catastrophe. Examples, among many reported in the literature, that employ or extend Rasmussen's work include **?** who argue that the cost and benefit of safety enhancements can be in part evaluated against predicted *CDF*. **?** describe how, based on *CDF*, some protections are less important than others. **?** describe a method that integrates prediction of conditional probability of core damage events with predicted frequency of core damage. **?** use a combination of *CDF* and core damage probability in a seismic risk study.

These typical analyses rely on the Rasmussen characterization of CDF as the product of a predicted arrival rate of initiating events (denoted by λ) with a predicted proportion of initiating events that exceed to core damage (denoted by p). However, the Rasmussen approach relies on informal mathematics that treats p as the long–run probability reactor protections are unavailable and thereby adopts an implicit assumption that long–run reactor protection unavailability as seen at the epochs of the randomly arriving initiating events is the same as would be observed at an arbitrary (*i.e.,* non–random) time. This assumption leads to computationally tractable estimates of CDF. Unfortunately, however, it holds only in special circumstances that typically defy practical justification. We will show that assuming the proportion of initiating events that exceed to core damage is equal to the limiting probability that protections are unavailable leads to an optimistic bias when estimating CDF. [1] We refer to this assumption as the ? characterization.

We will explore practical conditions under which the ? characterization of CDF is deficient, and then gauge the usefulness of CDF as a risk metric by addressing the following questions:

1. When does CDF exist?

2. Is it feasible to estimate CDF, when it exists?

3. Do estimates of CDF closely approximate its true value?

We approach these questions by first introducing three easily proved propositions and their corollaries in Section 2. We then conclude with Section 3 where we directly address these questions by applying the results of Section 2. Our analyses rely on standard results from the theory of martingales and stochastic counting processes.

---

[1] There is a large literature devoted to bias phenomena that arise with stochastic point and counting process models. Famous well–studied examples include the Inspection Paradox, Feller's Paradox, and Palm Probabilities. A common feature of these bias phenomena is the discrepancy between event probabilities as seen across different classes of stopping times. It should be of no surprise that counting processes used to study CDF would exhibit a bias phenomenon.

2 Predictive Model

Consider a single reactor unit equipped with protections having a predictive model constructed on a probability space that is tailored to incorporate the following features.

Predictive Model Features.

- The nuclear reactor unit under investigation is equipped with protections designed for the purpose of mitigating operational anomalies and exogenous influences, called initiating events, that might lead to core damage.

- The sequence of initiating events occur randomly and thus forms a stochastic point process.

- A single reactor can experience multiple core damage events over the course of its deployment.[2]

- The stream of arriving initiating events is bifurcated according to the state of the protections at the time of arrival. Thus, core damage events occur at epochs of those initiating events that breach reactor protections.

In our arguments that follow, we employ the standard analytical framework of filtered probability spaces for characterizing 'historical events' on which stochastic process temporal dynamics depend. Our notation closely follows that of most standard stochastic process textbooks, such as **?** or **?**. All random processes we model are adapted to the filtration $\{\mathcal{F}_t\}_{t \geq 0}$ on the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, P)$.

2.1 Almost Sure Behavior

Let N_t^D, $t \geq 0$ be the number of core damage events the reactor suffers in the interval $[0, t]$. *CDF* is understood to be the limiting number of core damage events per unit time. That is,

Definition 1 (Core Damage Frequency)**.**

$$CDF = \lim_{t \to \infty} \frac{1}{t} N_t^D,$$

1

---

[2]This modeling feature is essential in order to accommodate estimating a core damage 'frequency.'

whenever convergence to a constant occurs almost surely (a.s.)

CDF is reported as a numerical constant. Of course, convergence of eq. 1 is not guaranteed. Further, even when eq. 1 converges, there is no guarantee that its limit is a constant; convergence to a random variable is a completely plausible circumstance. We emphasize that CDF is a numerical constant, estimates of which are used to gauge the risk of suffering a core damage event.

As will be shown, the existence of CDF is predicated in part on the dynamics of initiating event arrivals. Let N_t be the number of initiating events arriving in the interval $[0, t]$. Without loss of generality we require that for almost all $\omega \in \Omega$, the trajectory $N_t(\omega)$ is right continuous and proceeds in jumps of magnitude one. The limiting arrival rate of initiating events λ is defined as follows.

Definition 2 (Initiating Event Frequency)**.**

$$\lambda = \lim_{t \to \infty} \frac{1}{t} N_t, \qquad\qquad 2$$

whenever convergence occurs a.s. In general λ can be a random variable, and convergence to a constant occurs only when $\lambda = E[\lambda]$

Note that $\{N_t^D\}_{t \geq 0}$ inherits right–continuity from $\{N_t\}_{t \geq 0}$, and it now follows directly from Definitions 1 and 2 that when λ is a constant,

$$CDF \stackrel{a.s.}{=} \lim_{t \to \infty} \frac{N_t^D}{N_t} \frac{N_t}{t} = p\lambda \qquad\qquad 3$$

where,

$$p = \lim_{t \to \infty} \frac{N_t^D}{N_t}.$$

Thus, CDF is the product of the initiating event frequency with the proportion of initiating events that exceed to core damage. Typically, estimating λ is straightforward. Estimating p is more challenging. Note that while p takes values in the interval $[0, 1]$, it is defined as the limiting value of a ratio of random variables. However, under certain circumstances p can be interpreted as the probability that an arriving initiating event will exceed to core damage. One such circumstance occurs when initiating events form

an ordinary Poisson process. Then, the well-known Poisson Arrivals See Time Averages (*PASTA*) result applies, and p can be computed as the limiting unavailability of reactor protections (see **?**). Unfortunately, the assumptions needed to justify *PASTA* defy practical justification. Consequently, estimating p is quite difficult in practice.

There are a variety of approaches for crafting estimators for p that incorporate the joint histories of initiating event arrivals, protection maintenance activity, environmental conditions, *etc.* Monte Carlo methods, owing to their adaptability to complex engineering models, have gained acceptance and popularity for estimating p and other statistics. These methods are not, however, a panacea because they require characterizing probability laws on subordinate stochastic processes that must be mapped into the dynamics of protection availability in order to build useful estimators. Characterizing probability laws on stochastic processes is often impractical due to the intensive data support required for all but the most stylized processes.

In order to better appreciate the manner in which *CDF* jointly depends on the arrival of initiating events and the efficacy of reactor protections, we will appeal to the martingale characterization of stochastic point processes (see **?**). Let N_t be (as stated previously) the number of initiating events arriving in the interval $[0, t]$. It is reasonable in practice to require that initiating events occur one at a time *a.s.* It follows that the trajectories of $\{N_t\}_{t \geq 0}$ are nonnegative non-decreasing and proceed in jumps of size one *a.s.* We define $\{X_t\}_{t \geq 0}$ as the state of reactor protections at time t taking values in the set $\{0, 1\}$. When an initiating event arrives to find $X_t = 1$, protection holds. Otherwise, with $X_t = 0$, protection fails and core damage will ensue.

Both $\{N_t\}_{t \geq 0}$ and $\{X_t\}_{t \geq 0}$ are adapted to the filtration $\{\mathcal{F}_t\}_{t \geq 0}$. We define X_t to be left continuous and thus $\mathcal{F}_t$-predictable. Clearly, $\{N_t\}_{t \geq 0}$ forms an $\mathcal{F}$-sub–martingale. Appealing to the standard results from the martingale calculus (see **?**), it follows from the Doob–Mayer Decomposition Theorem that

$$N_t = M_t + \Lambda_t, \qquad\qquad 4$$

6

where the process $\{M_t\}_{t\geq 0}$ forms an $\mathcal{F}_t$-martingale with compensator $\{\Lambda_t\}_{t\geq 0}$ (Λ_t is increasing *a.s.* and $\mathcal{F}_t$-predictable), with

$$\Lambda_t = \int_0^t \lambda_s ds. \qquad 5$$

Here, λ_t is well defined when for all nonnegative, $\{\mathcal{F}_t\}$-predictable $\{C_t\}_{t\geq 0}$ such that

$$E\Big[\int_0^\infty C_t dN_t\Big] = E\Big[\int_0^\infty C_t \lambda_t dt\Big]. \qquad 6$$

When well defined, λ_t is a unique Radon–Nikodym derivative defined on the usual equivalence class, with the stochastic intensity process $\{\lambda_t\}_{t\geq 0}$ adapted to $\{\mathcal{F}_t\}_{t\geq 0}$ and predictable. Informally, $\lambda_t = E[dN_t|\mathcal{F}_{t-}]$ and can be understood as the propensity for an initiating event to arrive in the next instant of time given the history of initiating events and reactor protections.

Now, consider the martingale transform M_t^D of protection unavailability $(1 - X_t)$ with respect to M_t of eq. 4, where

$$M_t^D \stackrel{\text{def}}{=} \int_0^t (1 - X_s)dM_s = \int_0^t (1 - X_s)dN_s - \int_0^t (1 - X_s)d\Lambda_s.$$

Proposition 1 (Core Damage Martingale). *$\{M_t^D\}_{t\geq 0}$ is a martingale whenever the stochastic intensity process of arriving initiating events $\{\lambda_t\}_{t\geq 0}$ exists.*

Proof. Since X_t is $\mathcal{F}_t$-predictable and $0 \leq X_t(\omega) \leq 1$ for all $\omega \in \Omega$ and $t \geq 0$, it follows that $\{M_t^D\}_{t\geq 0}$ is also an $\mathcal{F}_t$-martingale (see **?**), and noting that $N_t^D = \int_0^t (1 - X_s)dN_s$ counts the number of core damage events in the interval $[0, t]$, we have that

$$M_t^D = N_t^D - \int_0^t (1 - X_s)d\Lambda_s$$

and, substituting from eq. 5 gives

$$M_t^D = N_t^D - \Lambda_t^D. \qquad 7$$

□

We refer to $\{M_t^D\}_{t \geq 0}$ as the *Core Damage Martingale* and its compensator is given by

$$\Lambda_t^D = \int_0^t \lambda_s^D ds = \int_0^t (1 - X_s)\lambda_s ds.$$

Remark 1. *Equation 7 stands as the most general expression characterizing the relationship among core damage events, the arrival of initiating events, and the efficacy of reactor protections. It is important to keep in mind that, for all $t \geq 0$, M_t^D, N_t^D, λ_t^D, X_t, and (in particular) λ_t are random variables. Hence, eq. 7 is nontrivial as it requires stochastic integration.*

Consider now, the following proposition:

Proposition 2 (Existence of *CDF*)**.** *If*

$$\lim_{t \to \infty} \frac{M_t^D}{t} \overset{a.s.}{=} 0,$$
8

then λ^D exists (and is possibly a random variable) and for almost all (a.a.) $\omega \in \Omega$

$$\lambda^D(\omega) = \lim_{t \to \infty} \frac{N_t^D(\omega)}{t} = \lim_{t \to \infty} \frac{1}{t} \int_0^t (1 - X_s(\omega))\lambda_s(\omega)ds$$

where, $0 < \lambda^D(\omega) < \infty$.

That is,

$$\lim_{t \to \infty} \frac{M_t^D}{t} \overset{a.s.}{=} 0 \text{ if and only if}$$
$$\lim_{t \to \infty} \frac{N_t^D}{t} \overset{a.s.}{=} \lambda^D \text{ and}$$
$$\lim_{t \to \infty} \frac{1}{t} \int_0^t (1 - X_s)\lambda_s ds \overset{a.s.}{=} \lambda^D.$$
9

Proof. Proposition 2 is an obvious consequence of Definition 1 and eq. 7. □

Remark 2. *Clearly, CDF exists only if $\frac{M_t^D}{t} \overset{a.s.}{\to} 0$ and $\lambda^D \overset{a.s.}{=} E[\lambda^D] < \infty$. Proposition 2 reveals the challenge in estimating CDF. In the absence of observed core damage events, predictive estimates of CDF must*

be formulated in terms of phenomena that can be observed. To this end, analysts must rely on observations of initiating event arrival times, and reactor protection performance (principally in the form of maintenance records and failure data). These observations are, of course, insufficient to capture the joint dynamics of $\{(X_t, \lambda_t)\}_{t \geq 0}$ needed to directly employ the strong law relationship of Proposition 2, where,

$$\lambda^D \overset{a.s.}{=} \lim_{t \to \infty} \frac{\Lambda_t^D}{t} = \lim_{t \to \infty} \frac{1}{t} \int_0^t (1 - X_s)\lambda_s ds.$$

Monte Carlo methods do not escape the difficulty of computing λ^D. $\{X_t\}_{t \geq 0}$ and $\{\lambda_t\}_{t \geq 0}$ are not mutually independent (even when $\{N_t\}_{t \geq 0}$ is Poisson with rate λ) and λ_t is not directly observable. Since this dependence requires a Monte Carlo model to rely on an accurate characterization of the probability law on the joint stochastic process $\{(X_t, \lambda_t)\}_{t \geq 0}$, it is clear that the data requirements to support accurate estimation of this probability law are beyond the practical reality of reactor unit operations records.

2.2 Behavior in Expectation

Important insights regarding *CDF* are revealed by exploring the expectation of $\frac{M_t^D}{t}$. In particular, we are interested in the consequences of the stochastic dependence the state of system protections $\{X_t\}_{t \geq 0}$ and the arrival of initiating events $\{N_t\}_{t \geq 0}$ and consequently $\{\lambda_t\}_{t \geq 0}$. Consider now,

Proposition 3 (Moment Convergence). *Suppose that $\frac{M_t^D}{t} \overset{a.s.}{=} 0$, then*

$$\lim_{t \to \infty} E\Big[\frac{M_t^D}{t}\Big] = 0 \ \textit{if and only if}$$
$$\lim_{t \to \infty} E\Big[\frac{N_t^D]}{t}\Big] = E[\lambda^D] \ \textit{and} \qquad\qquad 10$$
$$\lim_{t \to \infty} E\Big[\frac{1}{t} \int_0^t (1 - X_s)\lambda_s ds\Big] = E[\lambda^D].$$

And, with the additional condition that $\lambda^D = E[\lambda^D] < \infty$,

$$CDF = \lim_{t \to \infty} \frac{1}{t} \int_0^t E[(1 - X_s)\lambda_s] ds. \qquad\qquad 11$$

Proof. Recall that almost sure convergence implies convergence in expectation (see **?**). Hence, eq. 10 follows directly from Proposition 2. Nonnegativity of the integrand in eq. 10 allows a routine application of Tonelli's Theorem to exchange the order of expectation and integration to show eq. 11 (see **?**). Finally, $\lambda^D = E[\lambda^D] < \infty$ implies that $\frac{N_t^D}{t} \overset{a.s.}{\to} E[\lambda^D]$, a finite constant, thus ensuring the existence of *CDF*. $\qquad\square$

Recalling the definition of covariance[3], it immediately follows that

Corollary 1. *When $\lambda^D = E[\lambda^D] < \infty$, then*

$$CDF \overset{a.s.}{=} \lim_{t \to \infty} \frac{1}{t} \int_0^t cov((1 - X_s), \lambda_s) ds + \lim_{t \to \infty} \frac{1}{t} \int_0^t E[(1 - X_s)] E[\lambda_s] ds \qquad 12$$

Proof. Simply apply the definitions of *CDF* and covariance. $\qquad\square$

Corollary 2. *When $\lambda^D = E[\lambda^D] < \infty$,*

$$CDF \overset{a.s.}{=} \lim_{t \to \infty} \frac{1}{t} \int_0^t E[(1 - X_s)] E[\lambda_s] ds \qquad 13$$

if and only if $E[\lambda_t | \mathcal{F}_t] = E[\lambda_t]$ for all $t \geq 0$.

Proof. We need only show that $cov((1 - X_t), \lambda_t) = 0$ if and only if $E[\lambda_t | \mathcal{F}_t] = E[\lambda_t]$.

First, assume that $E[\lambda_t | \mathcal{F}_t] = E[\lambda_t]$ for all $t \geq 0$ for all $t \geq 0$. Note that

$$E[(1 - X_t)\lambda_t] = E[E[(1 - X_t)\lambda_t | \mathcal{F}_t]$$

$$= E[E[(1 - X_t)E[\lambda_t] | \mathcal{F}_t]$$

$$= E[\lambda_t] E[E[(1 - X_t) | \mathcal{F}_t]$$

$$= E[\lambda_t] E[(1 - X_t)].$$

Thus, it follows from the definition of covariance that $cov((1 - X_t), \lambda_t) = 0$.

---

[3]For two random variables Y and Z, $cov(Y, Z) \overset{\text{def}}{=} E[ZY] - E[Z]E[Y]$.

Now, assume that $cov((1 - X_t), \lambda_t) = 0$. It follows trivially that

$$E[(1 - X_t)\lambda_t] = E[\lambda_t]E[(1 - X_t)].$$

$\square$

Corollary 3 (Rasmussen Characterization of CDF). *When,* $E[\lambda_t|\mathcal{F}_t] = E[\lambda_t] = \lambda < \infty$ *and* $X_t \overset{a.s.}{\to} X$,

then

$$CDF \overset{a.s.}{=} \lambda \lim_{t \to \infty} \frac{1}{t} \int_0^t (1 - X_s)ds. \qquad 14$$

Proof. It follows from eq. 13 that when $E[\lambda_t|\mathcal{F}_t] = E[\lambda_t] = \lambda < \infty$,

$$CDF \overset{a.s.}{=} \lambda \lim_{t \to \infty} \frac{1}{t} \int_0^t E[(1 - X_s)]ds,$$

and since $X_t \overset{a.s.}{\to} X$,

$$\lambda \lim_{t \to \infty} \frac{1}{t} \int_0^t E[(1 - X_s)]ds = \lambda \lim_{t \to \infty} \frac{1}{t} \int_0^t (1 - X_s)ds,$$

eq. 14 follows. $\square$

Remark 3. *It is important to appreciate that Corollary 3 does not imply that the state of reactor protections* $\{X_t\}_{t \geq 0}$ *is independent of initiating event arrivals* $\{N_t\}_{t \geq 0}$. *The condition* $E[\lambda_t|\mathcal{F}_t] = \lambda$ *implies that* X_t, *the state of system protections at time t, does not influence the arrival times of future initiating events (a circumstance that we discuss more fully in Section 3). The conditions establishing Corollary 3 allow for the possibility that initiating events can cause a failure of system protections (in addition to the possibility protections were already failed immediately prior to arrival). Equation 14 has, as a special case, the condition that arriving initiating events* $\{N_t\}_{t \geq 0}$ *form an ordinary Poisson process of rate* λ. *We also could have shown eq. 14 as a special case of eq. 9 where, the constant* λ *can be moved outside the integral (see ?) for a full development of when Poisson arrivals see time averages. We point out, however, that the condition* $E[\lambda_t|\mathcal{F}_t] = \lambda$ *can hold for initiating event streams that are not Poisson (see ?). Hence, eq. 14 is a bit more*

11

*general than the **?** result. Of course, Corollary 3 also holds under the more restrictive condition that $\{X_t\}_{t \geq 0}$ and $\{N_t\}_{t \geq 0}$ are independent processes.*

3 Summary and Conclusions

The appeal of *CDF* as a risk metric is that it reports as a single numerical value that is computed using historical data captured for a given reactor unit using physics and operations-based predictive modeling. The magnitude of the reported value is intended to quantify expectations regarding time between core damage events. That is, the reciprocal of *CDF* estimates the "expected" time between core damage events. And, since in practice, an operating reactor unit has never suffered a core damage event, *CDF* is intended to offer insight as to expectation of the remaining time until the first core damage event might occur.

Estimates of *CDF* are almost always developed through *PRA* and Probabilistic Safety Analysis (*PSA*) studies. These studies use combinational logic models (fault trees) of reactor protections subordinated to a Monte Carlo simulation experiment. Monte Carlo relies on estimates of the probability law for the various fault tree elements and initiating events. Fault tree element probability laws (*i.e.,* reliability functions) are estimated from historical data associated with reactor unit operations and maintenance. Initiating event laws are developed from histories of both exogenous influences (*e.g.,* weather events, seismic disturbances, *etc.*) and endogenous influences (*e.g.,* equipment failures, human operator errors, *etc.*). Modeling reliance on historical data reveals the necessity of strong law characterizations of event probabilities. But, it is important to appreciate that Level-1 *PRA* studies rely on the Rasmussen characterization of *CDF*. That is, either through explicit or implicit assumption, future arriving initiating events are taken to be independent of the present state of reactor protections. Consequently, Monte Carlo estimates of *CDF* developed through *PRA* converge to the form given in eq. 13.

Whether or not the appeal of *CDF* as a risk metric is justified is the subject of much literature that we will not review. It is nonetheless fact that *CDF* is a quantity of interest to regulators, design engineers, utility operators, and the public, and is a focus of Level-1 *PRA* studies. Our focus is limited to the existence

of *CDF* and the fidelity of its estimates. To this end, we revisit the three questions posed in Section 1.

3.1 When does *CDF* exist?

We establish the conditions required for the existence of *CDF* in Proposition 2. We observe that existence relies on the strong convergence to zero of the core damage martingale. As Proposition 2 shows, convergence cannot be achieved when either reactor protections or initiating events do not achieve a steady state regime. For instance, if climate change is deemed to be in play during the operational life time of a reactor unit, then exogenously arriving weather–response initiating events are unlikely to be in steady state. Similarly, our present understanding of seismic activity relies on self–exciting stochastic point processes. This suggests that exogenous initiating events arising from seismic activity are unlikely to achieve steady state during a reactor unit's operating lifetime. Finally, and perhaps most importantly, much engineering effort is devoted to discovering and eliminating reactor protection failure modes not included in preliminary Failure Mode and Effects Analysis (*FMEA*) studies. These discoveries and corresponding design corrections dramatically influence the probability law on $\{X_t\}_{t \geq 0}$ and whether or not reactor protections converge to a steady state regime. It is important recognize that the discovery of any new failure mode after time t cannot appear in the prior history captured in the filtration $\{\mathcal{F}_t\}_{t \geq 0}$; this practical engineering reality suggests that "good engineering practice" and innovation contribute to the lack of stationarity for reactor protections.

3.2 Is it feasible to estimate *CDF*, when it exists?

Accepting the existence of *CDF*, it is a straightforward matter to compute estimates under the Rasmussen characterization; this is the function of Level-1 *PRA* studies. If, however, the Rasmussen characterization is rejected, *CDF* is not easily estimated. In order to better understand this, let T_n, $n \geq 1$ be the time of the nth arriving initiating event.

$$T_n = \inf\{t > 0 : N_t \geq n\}, n \geq 1.$$

When *CDF* exists,

$$X_{T_n} \overset{a.s.}{\to} \tilde{X} \text{ and } X_t \overset{a.s.}{\to} X. \qquad\qquad 15$$

Under the Rasmussen characterization,

$$\tilde{X} \overset{a.s.}{=} X \implies E[\tilde{X}] = E[X] \overset{a.s.}{=} \lim_{t\to\infty} \frac{1}{t} \int_0^t X_s ds. \qquad\qquad 16$$

That is, the Rasmussen characterization requires that the state of reactor protections as seen at epochs of initiating event arrivals is the same as would be seen at an arbitrary time.

Of course, relaxing the requirements of the Rasmussen characterization implies that $E[\lambda_t | \mathcal{F}_t] \neq \lambda$ which has the consequence that eq. 16 does not hold. Corollary 1 reveals the bias between true *CDF* and the Ramussen characterization as

$$CDF - \lambda(1 - E[X]) \overset{a.s.}{=} \lim_{t\to\infty} \frac{1}{t} \int_0^t cov((1 - X_s), \lambda_s) \neq 0, \qquad\qquad 17$$

and while the bias is known to exist, stochastic intensity λ_t is a Radon–Nikodym derivative that is not directly observable; hence, there are no data to support a Bayesian estimator either parametric or non–parametric.

3.3   Do estimates of *CDF* closely approximate its true value?

Any special properties associated with the bias, given eq. 17, are of considerable practical interest. To this end, recall the definition of the Pearson correlation coefficient $\rho(Y, Z)$ between any two random variables Y and Z defined on a common probability space where

$$\rho(Y, Z) \overset{\text{def}}{=} \frac{cov(Y, Z)}{\sqrt{var(Y)}\sqrt{var(Z)}}. \qquad\qquad 18$$

14

$\rho(Y, Z)$ is positive if and only if Y and Z tend to be simultaneously greater than, or simultaneously less than, their respective expected values.[4]

It immediately follows from the definition of the correlation coefficient (eq. 18) that $(1 - X_t)$ and λ_t are either positively correlated or uncorrelated if and only if $cov((1 - X_t), \lambda_t) \geq 0$. Observe that for each $t, s \geq 0$, and $\omega \in \Omega$, $X_{t+s}(\omega) - X_t(\omega)$ takes values only the set $\{-1, 0, 1\}$.

Straightforward engineering reasoning reveals that for all $t, s > 0$,

$$\lim_{s \downarrow 0} P(X_{t+s} - X_t = 1, N_t - N_{t-s} = 1) = 0.$$

This is to say, *almost surely an arriving initiating event will not cause non-functioning reactor protections to suddenly function*, and it follows that

$$\lim_{s \downarrow 0} P(X_{t+s} - X_t \leq 0 | N_t - N_{t-s} = 1) = 1.$$

This is equivalent to saying that X_t almost surely will not increase given an increase in N_t. Since N_t is nondecreasing in t, we observe that X_t and N_t must be either negatively correlated or uncorrelated. As an immediate consequence, $(1 - X_t)$ and N_t must be either positively correlated or uncorrelated. Noting the informal relationship that $E[dN_t | \mathcal{F}_{t-}] = \lambda_t dt$, it is easily reasoned that λ_t and $(1 - X_t)$ must be either positively correlated or uncorrelated. This circumstance agrees with engineering experience.[5] Thus, when λ_t and $(1 - X_t)$ are positively correlated or uncorrelated, increasing the propensity for an initiating event to arise in the next instant of time increases the propensity for core damage. Similarly, decreasing the propensity of an arriving initiating event in the next instant of time decreases the propensity for core damage. More precisely,

$$\lim_{s \downarrow 0} P(X_{t+s} - X_t \leq 0 | \tilde{\lambda}_t) \overset{a.s.}{>} \lim_{s \downarrow 0} P(X_{t+s} - X_t \leq 0 | \hat{\lambda}_t), \text{ for any } \tilde{\lambda}_t \overset{a.s.}{>} \hat{\lambda}_t$$

---

[4]The converse is true for a negative value of $\rho(Y, Z)$.

[5]For example, when 12 of 13 emergency backup diesel generators were washed away at Fukushima Daiichi, loss of onsite power and thus the ability to sustain cooling lead to core meltdown in each of the reactor units.

and

$$\lim_{s\downarrow 0} P(X_{t+s} - X_t \le 0|\tilde\lambda_t) \stackrel{a.s.}{<} \lim_{s\downarrow 0} P(X_{t+s} - X_t \le 0|\hat\lambda_t), \text{ for any } \tilde\lambda_t \stackrel{a.s.}{<} \hat\lambda_t.$$

We, thus, conclude that since for all $t \ge 0$, $(1 - X_t)$ and λ_t cannot be negatively correlated, the bias of eq. 17 must be nonnegative. The nonnegative bias appearing in Corollary 1 ensures that when CDF exists, the Rasmussen characterization always underestimates the true value of CDF.

In practice, numerical values for CDF are typically computed within a PRA or PSA study that relies on the Rasmussen characterization for Level-1 studies. The inadequacy of the Rasmussen characterization, of course, derives from how (typically very sophisticated) protection (un-)availability analysis is incorporated into PRA. Rasmussen insightfully seeks to extended traditional protection availability analysis, by noting that the risk of a core damage event should be understood as a consequence of an unmitigated initiating event. Hence, it is only at arrival times of initiating events that protection unavailability would lead to a core damage event. Treating $\lim_{t\to\infty}(1 - X_t)$ as the top event of a fault tree analysis of the unit protective system, the Rasmussen characterization reasons that $p = \lim_{t\to\infty} E[(1-X_t)]$ is the likelihood of core damage at the time of an initiating event. However, a closer examination reveals that unavailability at instants of initiating events should be written as $\lim_{n\to\infty} E[(1 - X_{T_n})]$. Since the index on unavailability at the time of initiating event arrivals is the sequence of random variables $\{T_n; n \ge 1\}$, it is logical that, except under very special conditions, $\lim_{n\to\infty} E[(1 - X_{T_n})] \ne \lim_{t\to\infty} E[(1 - X_t)]$.

An arriving initiating event will find system protections either working or failed. If the protections are working, they will ensure that the initiating event does not lead to catastrophe. On the other hand, if protections are failed upon arrival, catastrophe will ensue. Catastrophe ensues in two distinct scenarios:

1. System protections are failed at time T_n for reasons unrelated to the arrival of the nth event.

2. System protections were working, but the arrival of the nth event caused protection failure.

Clearly, the covariance bias term in Corollary 1 is associated with item 2. Because the Rasmussen characterization relies on protective system reliability modeling which only describes $(1 - X_t)$, $t \ge 0$, uncertainty associated with the initiating event arrival time point process $\{T_n; n \ge 0\}$ is not represented. Upon rec-

ognizing this modeling deficiency, it might seem reasonable to extend the Rasmussen characterization so as to overcome its optimistic bias. But, the very purpose of exploring CDF is to approximate it with a numerical value that is supported by operations and maintenance data. Yet, observational data will not exist to support estimating protection unavailability $(1 - X_{T_n})$, for any $n \geq 1$ because in practice we have no historical core damage events. In practice, an unbiased estimate of CDF, in the absence of any observed core damage events should (except under very special circumstances) seem too good to be true, whenever initiating events are believed to affect protection reliability. Rasmussen and successor PRA researchers are ensnared by a common paradox (the random observer problem) that is untangled in the large and well–known Palm Calculus literature. Unfortunately, understanding the Rasmussen bias is not the same as escaping it. There is no "fix" to correct this bias, and this deficiency has important implications for PRA and PSA and risk-informed regulatory oversight.

3.4   Final Observations

We have shown that the Rasmussen characterization of CDF will at best yield values that underestimate true CDF and at worst provide values for a metric that does not exist. We observe from corollary 1 of proposition 3 that the optimistic bias of CDF occurs under the Rasmussen characterization because it generally does not account for the possibility that an arriving initiating event exceeds to catastrophe by damaging the functionality of system protections. Clearly, it is infeasible to quantify such catastrophes since historical data do not exist. Further, reliance on *ad hoc* characterizations of the Rasmussen characterization bias lack the rigor necessary to support the objectives of high–consequence risk analyses like PRA and PSA. Thus, philosophical objections regarding the usefulness of CDF aside, reported values of CDF are not particularly informative as a risk metric quantifying reactor safety because their optimistic bias cannot be analytically quantified. This observation calls into question the use of CDF in regulatory language. The intent of regulators' oversight is to ensure adequate safety for the public. To this end, the public is better served by well–developed prescriptive regulations than by optimistically biased risk–informed regulations

built upon *PRA* or *PSA* analyses.