A New Probability Density Function: (The General Variable-Rate Probability Function)

ABSTRACT

Calculation of probability of events is often based on the assumption of the variable distribution. This is a bold decision that is not always right, which requires adjustments to force conformity to the empirical evidence. The Variable-Rate Probability Density Function (VRPD function) is the result of a proof based on a mental experiment in which the variable was not forced to fit into any preestablished distribution. Instead, the experiment was allowed to guide the proof from its observable behavior.

The result is a density function to calculate the probability of events for any stochastic process by only knowing its rate events function.

I. INTRODUCTION

POISSON distribution is one of the most familiar discrete distribution functions when it comes to calculate the probability of having a certain count of events in a period of time (or in an area or volume). However, a key assumption to successfully applying this distribution is that the random variable is Poisson distributed, which means that:

i) The number of events in an interval is independent of the number of events in any other interval

ii) The probability of a single event in $(t, t+h)$ depends on a constant rate in that interval

iii) The probability of more than one event in $(t, t+h)$ is neglected. i.e.: $o(h)$.

However, in several cases there is need:

i) To know the probability of having an event in a given interval – not the probability of having a certain number of events in that interval

ii) That the probability be in any continuous time interval

iii) That the time interval be as small or large as required by the experiment

The probability of an interruption or failure in a system life cycle or in an organizational process are examples of cases where assuming the variable is Poisson distributed is not appropriate. Here is where the VRPD theorem is the right solution.

II. PROOF OF THE PROBABILITY DENSITY FUNCTION TO ANY VARIABLE-RATE STOCHASTIC PROCESS

A. Theorem

The probability of one or more events (probability of success: P(X=1)) of a stochastic process in any interval $t_a \leq t < t_b$ along its lifecycle with a known interruption rate function λ(t) is given by:

$$P(X=1)_{t_a \leq t < t_b} = \int_{t_a}^{t_b} \frac{\lambda(t)}{K(t)} dt$$

With $K(t)$, the subspace of events of the process along its lifecycle

Proof:

Suppose an experiment with m processes or systems (p/s) of the same kind, starting their lifecycle at the same time $t = 0$. Now let us take a timeframe at out of the p/s lifecycle of size $\Delta T_i = t_{i-1} - t_i$.

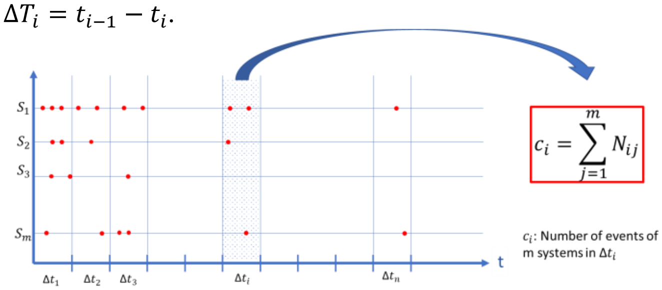

Figure 1.

The number of interruptions of a p/s_j on the interval ΔT will be: Ni_j, and the total number of interruptions (successes count) of the m processes in the interval ΔT is given by:

$$c_i = \sum_{j=1}^{m} N_{ij} \ \ (eq1)$$

The mean interruption count of the m systems on ΔT at the time t_i is given by: $\frac{c_i}{m}$, which is the empirical probability of interruption of the p/s type as it is stablished by the theorem of large numbers for Bernoulli trials:

$$\lim_{m\to\infty} P\left(\left|\frac{c_i}{m} - p_i\right| \geq \epsilon\right) = 0 \ \ (eq2) \ \ \text{(eq2)}$$

With p_i*: Probability of a Bernoulli event* at the time t_i in a timeframe ΔT*;* ϵ*: any positive number.*

An equivalent identity is given by the general low of large numbers which states that:

$$\lim_{m\to\infty} \frac{c_i}{m} = p_i \ \ (eq3)$$

Without losing generality, we can say that the interruption rate of a p/s in an interval $\Delta T = t_0 \leq T < t_n$ is given by:

$$\lim_{m\to\infty} \left(\frac{c_i}{m * \Delta T}\right) = \frac{p_i}{\Delta T} = \lambda_i \ (eq4)$$

With X: the random variable. The successful event $(X = 1)$ is the occurrence of one or several events in $\Delta T = t_b - t_a$.
λ_i: the rate of events in ΔT.
Now, if we divide the time interval ΔT in two intervals, we will have that the probability of an event in ΔT resulting from any two independent and non-exclusive events, is given by:

$$P(X = 1)_{t_0 \leq t < t_n} = p_1 + p_2 - p_1 p_2 = \lambda_1 \frac{\Delta T}{2} + \lambda_2 \frac{\Delta T}{2} - \left(\frac{\Delta T}{2}\right)^2 \lambda_1 \lambda_2$$

Whereby, from eq4, if we divide ΔT in n intervals, with $n \geq 2$, the probability of one or more interruptions, i.e., $(P(X = 1))$, taking place along $t_0 \leq t < t_n$, can be expressed as follows:

$$P(X = 1)_{t_0 \leq t < t_n} = \sum_{i=1}^{n} p_i - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} p_i p_j + \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{k=j+1}^{n} p_i p_j p_k - \cdots + (-1)^n \sum_{i=1}^{2} \sum_{j=i+1}^{3} \sum_{k=j+1}^{4} \dots \sum_{r=n-1}^{n} p_i p_j p_k \dots p_r + (-1)^{n-1} \prod_{i=1}^{n} p_i \ \ (eq5)$$

$With$:
$[t_0, t_n), any\ interval\ in\ [0, \infty)$.
$The\ time\ interval\ given\ by\ \frac{t_n - t_0}{n} = \frac{\Delta T}{n}$
$X = 1$: $successful\ event$: $one\ or\ more\ events$.
$i \neq j \neq k \neq \cdots r\ \forall i, j, k, \dots, r. With\ i, j, k, \dots, r = 2,3, \dots, n.$
$p_i, p_j, p_k, \dots, p_r$:
$the\ probability\ of\ interruption\ in\ the\ time\ interval\ i, j, k, \dots, r$
n: a possitive integer.

The first sum of eq5 has $\binom{n}{1}$ terms, i.e., n terms; the second, $\binom{n}{2}$ terms; the third, $\binom{n}{3}$ terms, and son on. The sum before the last one has $\binom{n}{n-1}$ terms, i.e., n terms.
For instance, if we divide the subspace of events in ten intervals, i.e., $\frac{\Delta t}{n}$, the sums that compound the eq5 will have the following number of terms:

$$\binom{10}{1} = 10;\ \binom{10}{2} = 45;\ \binom{10}{3} = 120;\ \binom{10}{4} 210;\ \binom{10}{5} = 252;\ \binom{10}{6} 210;\ \binom{10}{7} = 120;\ \binom{10}{8} = 45;\ \binom{10}{9} = 10$$

Plus, the last term, corresponding to $(-1)^{n-1} \prod_{i=1}^{10} p_i$.

When n is even, it appears a mid-term (252 in the example above). When n is odd, the pattern is a perfect mirror. The total number of terms of the eq5 is given by:

$$2\left[n + \binom{n}{2} + \binom{n}{3} + \cdots + \binom{n}{\frac{n-2}{2}}\right] + \binom{n}{\frac{n}{2}}, if\ n\ is\ even$$
$$2\left[n + \binom{n}{2} + \binom{n}{3} + \cdots + \binom{n}{\frac{n-1}{2}}\right], if\ n\ is\ odd$$

Now, to calculate the probability of an event (e.g., one or more interruptions or failures) in any variable rate process using the eq5, we must consider that, if n increases, $\frac{\Delta T}{n}$ decreases in the same proportion. Therefore, if ΔT is divided in n intervals, with n very large, eq5 converges on the second sum and can be written as follows:

$$P(X = 1)_{t_0 \leq t < t_n} = \sum_{i=1}^{n} p_i - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} p_i p_j = \frac{\Delta T}{n} \sum_{i=1}^{n} \lambda_i - \left(\frac{\Delta T}{n}\right)^2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \lambda_i \lambda_j \ \ (eq6)$$

Making $\Delta t = \frac{\Delta T}{n}$, from eq4 we have:

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} p_i p_j = (\Delta t)^2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \lambda_i \lambda_j \ \ (eq7)$$

But,

$$\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\lambda_i\lambda_j = \frac{1}{2}\left[\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j - \sum_{i=1}^{n}{\lambda_i}^2\right] \ (eq8)$$

Then, eq6 becomes:

$$P(X=1)_{t_0\le t<t_n} = \sum_{i=1}^{n}p_i - \frac{(\Delta t)^2}{2}\left[\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j - \sum_{i=1}^{n}{\lambda_i}^2\right] \ (eq9)$$

i. With a Rate Function Continuous by parts

Let us suppose that for $t_0 \le t < t_n$, there is a continuous function by parts $\zeta_i(t)$ that represents the rate of events in every Δt such that:

$$\zeta_i(t) = \begin{cases}\zeta_i(t) > 0, & t_{i-1} \le t < t_i \\ 0, & otherwise\end{cases}$$

And $\zeta_i(t = v) = \lambda_i,\ t_{i-1} \le v < t_i, with\ i = 1,2,3,\dots,n$

According to the Mean Value Theorem we can find a value λ_i such that:

$$\lambda_i = \frac{\int_{t_{i-1}}^{t_i}\zeta_i(u)du}{\Delta t} \ (eq10)$$

$$or, what\ is\ the\ same{:}\ \Delta t * \lambda_i = p_i = \int_{t_{i-1}}^{t_i}\zeta_i(u)du \ (eq11)$$

Now, with:

$$S = \sum_{i=1}^{n}p_i = \int_{t_0}^{t_1}\zeta_1(u)du + \int_{t_1}^{t_2}\zeta_2(u)du + \cdots + \int_{t_{n-1}}^{t_n}\zeta_n(u)du \ (eq12)$$

Eq9 becomes:

$$P(x=1)_{t_0\le t<t_n} = S - \frac{1}{2}\left[S^2 - \sum_{i=1}^{n}\left(\int_{t_{i-1}}^{t_i}\zeta_i(u)du\right)^2\right] \ (eq13)$$

ii. With a Continuous Rate Function

Until now we have $\zeta_i(t)$ as a continuous function by parts, i.e., continuous in every interval $t_{i-1} \le t < t_i$, such that

$$\frac{d\Lambda_i(t)}{dt} = \zeta_i(t)$$

Now, let us suppose that instead of having a continuous function by parts, there exists a continuous function $\lambda(t)$ in $t_0 \le t < t_n$, and:

$$\lambda(t) = \sum_{i=1}^{n}\zeta_i \ (eq14)$$

Besides:

$$\frac{d\Lambda(t)}{dt} = \lambda(t)\ \forall t, t_0 \le t < t_n$$

Thus:

$$\sum_{i=1}^{n}\left(\int_{t_{i-1}}^{t_i}\zeta_i(u)du\right)^2 = \sum_{i=1}^{n}\left(\int_{t_{i-1}}^{t_i}\lambda(u)du\right)^2$$

Therefore:

$$\sum_{i=1}^{n}\left(\int_{t_{i-1}}^{t_i}\lambda(u)du\right)^2 = [\Lambda(t_1)-\Lambda(t_0)]^2 + [\Lambda(t_2)-\Lambda(t_1)]^2 + [\Lambda(t_3)-\Lambda(t_2)]^2 + \cdots + [\Lambda(t_n)-\Lambda(t_{n-1})]^2$$

$$= \Lambda^2(t_0) + \Lambda^2(t_n) + 2[\Lambda^2(t_1) + \Lambda^2(t_2) + \Lambda^2(t_3) + \cdots + \Lambda^2(t_{n-1})] - 2[\Lambda(t_1)\Lambda(t_0) + \Lambda(t_2)\Lambda(t_1) + \cdots + \Lambda(t_{n-1})\Lambda(t_{n-2}) + \Lambda(t_n)\Lambda(t_{n-1})] \ (eq15)$$

As Δt is very small ($\Delta t \ll 1$), we have that: $\Lambda(t_{i+1})\Lambda(t_i) \approx \Lambda^2(t_i)$, from eq 10 and with eq15:

$$(\Delta t)^2\sum_{i=1}^{n}{\lambda_i}^2 = \sum_{i=1}^{n}\left(\int_{t_{i-1}}^{t_i}\lambda(u)du\right)^2 = \Lambda^2(t_n) - \Lambda^2(t_0) \ (eq16)$$

Provided that $\lambda(t) = \zeta_i(t)\ \forall t, with\ t_{i-1} \le t < t_i$, eq12 becomes:

$$S = \int_{t_0}^{t_n}\lambda(u)du = \Lambda(t_n) - \Lambda(t_0) \ (eq17)$$

That with eq16 into eq13 becomes:

$$P(x=1)_{t_0 \le t < t_n} = S - \frac{1}{2}[S^2 - (\Lambda^2(t_n) - \Lambda^2(t_0))]$$

Which is the same as:

$$P(x=1)_{t_0 \le t < t_n} = S - \frac{1}{2}[S^2 - S(S + 2\Lambda(t_0))]$$

Now, factorizing we have:

$$P(x=1)_{t_0 \le t < t_n} = S(1 + \Lambda(t_0))\ (eq18)$$

As we stated, if there is certainty that the event will occur at least once in $0 \le t < t_n$, e.g., the probability of failure or interruption of a process or system along its lifecycle, we may say, from eq18:

$$\lim_{t_n \to \infty} P(x=1)_{0 \le t < t_n} = K(1 + \Lambda(0)) = 1 \iff$$

$$(1 + \Lambda(0)) = \frac{1}{K}\ (eq19)$$

Note: *From eq12 we know that K is the subspace of all the events.* K will depend on the process under analysis, for instance, if we take the subspace of all events along the process' lifecycle, i.e., $t_0 = 0, and\ t_n \to \infty$, and if the limit exists, we have:

$$K = \lim_{t_n \to \infty} S = \lim_{t_n \to \infty} \int_0^{t_n} \lambda(t)dt$$

Or, factorizing, the restriction we have:

$$\Lambda(t_n) - \Lambda(0) = K$$

Then, with eq19 into eq18, without losing generality, the probability of interruption of an EP in any time interval $t: [t_a, t_b)$ is given by:

$$P(x=1)_{t_a \le t < t_b} = \frac{S}{K}$$

Thus, in the most general case:

$$P(X=1)_{t_a \le t < t_b} = \frac{1}{K(t)} \int_{t_a}^{t_b} \lambda(t)dt\ \ (eq20)$$

Conclusion:

The probability of interruption of an EP in a given interval is the ratio between the area under the rate function in that interval and the associated subspace of events, where the subspace of events, K(t), may be a constant or a function of time, as it happens with conditional probability of failure.

The probability density function to any variable-rate processes with a continuous interruption rate function $\lambda(t)$, is given by:

$$f(t) = \frac{\lambda(t)}{K(t)}$$

REFERENCES

[1] J. A. Aristizábal, "Annex-1. Probability of Interruption and Business Impact, 1ed. Menlo Park, CA, USA: IBUKKU, 2020, pp. 107–115.

[2] R. Murray., J. Schiller., R. Srinivasan "Special Probability Distributions" in Probability and Statistics, 4th, USA: Mc Graw Hill, 2013, pp. 108–109.

[3] J. A. Aristizábal, "Annex-4. Variable-Rate processes Probability Distribution (VRPD) applied to Conditional Probability of Failure (CPF)" in Probability of Interruption and Business Impact, 1ed. Menlo Park, CA, USA: IBUKKU, 2020, pp. 133–138.

[4] L. Leithold, "Integral Definida e Integración" in El Cálculo, 7th ed. MX: Oxford University Press, 1998.

[5] A. K. Verma., S. Ajit., D.R. Karanki. "Basic Reliability Mathematics" in Safety Engineering, England: SpringerLink, 2016, pp. 28–36.

Alejandro Aristizábal was born (1972) in Medellín, Colombia, Graduated as Electronics Engineer and with a MSc in statistics.

Developed mathematical models to industrial applications in risk and Business Continuity Management.

Modeled the Business Impact functions, introducing the concepts of Apparent, Inherent and Residual Impacts in Business Continuity.

Reformulated some fundamentals in Reliability Theory through applying the VRPD theorem.

With more than fifteen years of experience in telecommunications networks and business continuity in more than six industries, Alejandro Aristizabal is the author of "Probability of Interruption and Business Impact", a book that approaches business continuity from a mathematical perspective, including fundamentals on risk level calculations and his personal perspective on business continuity, plus the mathematics underlying the Evolutionary Processes' interruption, including some interesting discussions about the effects of the interruption of organizational processes.