

Models Based on Conditional Probabilities

Abstract

We know that X has Waiting time distribution if and only if

$$P(X = k \mid X \geq k) = \theta.$$

Here, we propose a two-parameter model by specifying

$$P[X=1] = p \text{ and } P(X = k \mid X \geq k-1) = \theta.$$

The problem of interest is to find the natural parameter space. If $\theta = p(1-p)$, it gives Waiting time distribution.

1. Introduction:

In the literature, models have been proposed based on the failure rate. For example, Cox's proportional hazards model (PHM), Increasing Failure Rate (IFR), DFR, IFRA, DFRA models etc. For details, one may refer to Barlow and Proschan (1975).

If X is a positive integer valued r.v. the failure rate function $r(k)$; $k = 1, 2, 3, \dots$ is defined as

$$r(k) = P\{X = k \mid X \geq k\} = \frac{P_k}{P(X \geq k)} \quad k = 1, 2, 3, \dots$$

If $r(k) = \theta$, for $k = 1, 2, 3, \dots$ and $0 < \theta < 1$, then it corresponds to waiting time distribution. A natural question would be, can one define a probability distribution $\{P_n, n = 1, 2, 3, \dots\}$ by specifying P_1 and $(P_k / 1 - (P_1 + \dots + P_{k-2}))$ for $k \geq 2$? In particular, we consider the case of $1 - (P_1 + \dots + P_{k-2}) = \theta \leq 1 - p$ for $k \geq 2$ with $P_1 = p$. ---- (1)

If $\theta = p(1-p)$ then it corresponds to waiting time distribution. The following is a related question : for $\theta \leq 1 - p$, is every solution $\{P_n\}$ of (1) a proper probability distribution? It is numerically verified that $\{P_n\}$ is a probability distribution for $\theta \leq pq$ and $X_{\theta_2} \stackrel{st}{\leq} X_{\theta_1}$ if $0 < \theta_1 \leq \theta_2 < pq$.

2. A Model Based on Constant Conditional Probabilities

Let $0 \leq p \leq 1$ and $0 \leq \theta \leq 1-p$, be two given numbers and $P_1 = p$, $P_2 = \theta$, further $\{P_n\}_1^\infty$ be a sequence of real numbers such that

$$P_n = \theta \{1 - (P_1 + \dots + P_{n-2})\}, \text{ for } n \geq 2.$$

Equivalently, $\{P_n\}_1^\infty$ is such that $P_1 = p$, $P_2 = \theta$ and it satisfies the difference equation

$$P_n = P_{n-1} - \theta P_{n-2}, \text{ for } n \geq 3 \quad \text{----- (2)}$$

$$P_3 = \theta - \theta p = \theta(1 - p) = \theta(1 - P_1)$$

$$P_4 = \theta - \theta p - \theta \theta$$

$$= \theta \{1 - P_1 - P_2\}$$

$$P_5 = \theta \{1 - P_1 - P_2\} - \theta P_3$$

$$= \theta \{1 - P_1 - P_2 - P_3\}$$

In general, $P_n = \theta \{1 - P_1 - \dots - P_{n-2}\}$

If $\{p_n\}_1^\infty$ is a probability distribution of a r. v. X then,

$$P[X = n] = \theta P[X > n - 2]$$

$$\Rightarrow P[X = n \mid X > n - 2] = \theta \quad \text{----- (3)}$$

Note that if $0 < p < 1$, $\theta = 1 - p$ then $P_1 = p$, $P_2 = 1 - p$, $P_3 = \theta(1 - p) > 0$. Hence, $\{P_n\}$ will not form a probability distribution.

Hence, a problem of interest is for given p , find the possible values of θ (if any exist) for which $\{P_n\}$ is a probability distribution. One way to proceed is to find the generating function and examine for it to be a generating function of a probability distribution. Alternatively, find solution to difference equation (2) and find condition on θ , for which $\{P_n\}$ is a probability distribution.

2.1 Generating function :

Let $G(s) = \sum_1^\infty s^n p_n$ be the generating function corresponding to $P_1 = p$,

$P_2 = \theta$ and $P_n = P_{n-1} - \theta P_{n-2}$, for $n \geq 3$.

Then it is easy to verify that,

$$G(s) = \frac{s\{p + (\theta - p)s\}}{1 - s + \theta s^2} \quad \text{----- (4)}$$

$$\text{If } \theta = pq, G(S) = \frac{sp\{1 + ps\}}{1 - s + pqs^2} = \frac{sp(1 - ps)}{(1 - qs)(1 - ps)} = \frac{sp}{1 - qs},$$

which is the probability generating function of waiting time distribution. From

(4) for any p, θ , we have $G(1) = 1$ and hence

$$\sum_1^{\infty} p_n = 1 \quad \text{----- (5)}$$

For $\{P_n\}$ to be a probability distribution, we must have $P_n \geq 0$. However, if $\{P_n\}$ is a probability distribution of a r.v. X then,

$$\begin{aligned} 1 &= \sum_{n=1}^{\infty} P_n = P_1 + \theta(P_1 + P_2 + \dots) + \theta(P_2 + \dots) \\ &= P + \theta\{P_1 + 2P_2 + 3P_3 + \dots\} \\ &= P + \theta EX \end{aligned}$$

Hence, $EX = \frac{q}{\theta}$. But as $E(X) > 1$ we have $\theta < q$.

Let $P_1 = p, 0 < p < 1, P_2 = \theta, 0 < \theta < 1 - p$ and

$$P_n = \{1 - P_1 - \dots - P_{n-2}\}, \text{ for } n \geq 3$$

From (5), we have $\lim F(n) = 1$ as n tends to infinity. For $\{P_n\}$ to be a probability distribution it is enough to show that

$$S_n = P_1 + \dots + P_n < 1 \quad \forall n \geq 3$$

For $\theta \leq p(1 - p)$, numerically we have verified that $S_n < 1$ and hence $P_n > 0$ and $S_n \uparrow 1$. However, we are not able to prove it analytically. In the following, some numerical computations are given in the form of a table.

TABLE

X	p = 0.4, pq = 0.24				p = 0.1, pq = 0.09				p = 0.7, pq = 0.21			
	Theta=0.25		Theta=0.26		Theta=0.25		Theta=0.26		Theta=0.210001		Theta = 0.2101	
	P(X)	F(X)	P(X)	F(X)	P(X)	F(X)	P(X)	F(X)	P(X)	F(X)	P(X)	F(X)
0	0.4	0.4	0.4	0.4	0.1	0.1	0.1	0.1	0.7	0.7	0.7	0.7
1	0.25	0.65	0.26	0.66	0.25	0.35	0.26	0.36	0.21	0.91	0.21	0.9101
2	0.15	0.8	0.156	0.816	0.225	0.575	0.234	0.594	0.063	0.973	0.063	0.97313
3	0.088	0.8875	0.088	0.9044	0.163	0.7375	0.1664	0.7604	0.0189	0.9919	0.019	0.992018
4	0.05	0.9375	0.048	0.95224	0.106	0.8438	0.1056	0.86596	0.0057	0.9976	0.006	0.997663
5	0.028	0.9656	0.025	0.977096	0.066	0.9094	0.0623	0.928256	0.0017	0.9993	0.002	0.99934
6	0.016	0.9813	0.012	0.989514	0.039	0.9484	0.0349	0.963106	0.0005	0.9998	5E-04	0.999831
7	0.009	0.9898	0.006	0.995469	0.023	0.9711	0.0187	0.98176	0.0002	0.9999	1E-04	0.99997
8	0.005	0.9945	0.003	0.998195	0.013	0.984	0.0096	0.991352	5E-05	1	4E-05	1.000005
9	0.003	0.9971	0.001	0.999373	0.007	0.9912	0.0047	0.996095	1E-05	1	6E-06	1.000012
10	0.001	0.9984	5E-04	0.999843	0.004	0.9952	0.0022	0.998343	4E-06	1	-1E-06	1.000011
11	7E-04	0.9992	2E-04	1.000005	0.002	0.9974	0.001	0.999358	1E-06	1	-2E-06	1.000008
12	4E-04	0.9996	4E-05	1.000046	0.001	0.9986	0.0004	0.999789	0	1	-2E-06	1.000006
13	2E-04	0.9998	-1E-06	1.000045	6E-04	0.9993	0.0002	0.999956	0	1	-2E-06	1.000004
14	1E-04	0.9999	-1E-05	1.000033	3E-04	0.9996	6E-05	1.000011	0	1	-1E-06	1.000003
15	6E-05	0.9999	-1E-05	1.000021	2E-04	0.9998	1E-05	1.000022	0	1	-1E-06	1.000002
16	3E-05	1	-9E-06	1.000013	1E-04	0.9999	-3E-06	1.000019	0	1	-1E-06	1.000002
17	2E-05	1	-6E-06	1.000007	5E-05	0.9999	-6E-06	1.000014	0	1	0	1.000001
18	8E-06	1	-3E-06	1.000004	3E-05	1	-5E-06	1.000009	0	1	0	1.000001
19	4E-06	1	-2E-06	1.000002	2E-05	1	-4E-06	1.000005	0	1	0	1
20	2E-06	1	-1E-06	1.000001	8E-06	1	-2E-06	1.000003	0	1	0	1
21	1E-06	1	0	1	4E-06	1	-1E-06	1.000001	0	1	0	1
22	1E-06	1	0	1	2E-06	1	-1E-06	1.000001	0	1	0	1
23	0	1	0	1	1E-06	1	0	1	0	1	0	1
24	0	1	0	1	1E-06	1	0	1	0	1	0	1
25	0	1	0	1	0	1	0	1	0	1	0	1

As such we have the following conjectures

1. $\{P_n\}$ is a probability distribution for $\theta \in (0, pq]$.
2. $X_{\theta_2} \stackrel{st}{\leq} X_{\theta_1}$ if $\theta_1 \leq \theta_2 < pq$.
3. For $\theta (= 0.2101) > pq (= 0.21)$ $\{P_n\}$ is not a probability distribution. We conjecture that for $0 < p < 1$ and $0 < \theta < pq$, $\{P_n\}$ is a probability distribution. The analytical work is in progress.