J. Indian Inst. Sci.
A Multidisciplinary Reviews Journal
ISSN: 0970-4140 Coden-JIISAD

REVIEW ARTICLE

Development of Some Selected Stochastic Models of Human Fertility in India: The Untold Stories

Ram Chandra Yadava[*]

Abstract | Modeling of any phenomenon requires the knowledge of the subject and the mathematical skill that is required to express the phenomenon in the form of mathematical relationship(s). In the process of model building, the researcher(s) experience several failures and learnings before getting the final model. While large number of final models are developed and published, learnings of model development are seldom documented. Hence, there is a need to bring such articles which could present the learning of model development. Under this premise, the author which has experience of developing stochastic models of human reproduction aims, in this article, to narrate those learnings which he had encountered in the development of large number of models. Since detailed descriptions of those models are already available in publications, the article would skips detail derivations of those models but to provide learnings of the development of the models.

1 Introduction

Demography as a subject is considered to be an important branch of Social Sciences. However, 'Mathematical Demography' or 'Mathematics of Population' covers a vast range of literature of the field of Population Studies. Stochastic Modeling of Human Fertility is a narrower branch of 'Mathematical Demography', but still it contains many facets of research work.

It is to be admitted that the things are not so easy to describe, the major contributions made in the field by various researchers even in brief. Therefore, we leave the above matter for the readers to learn from various books and available scientific literature. The major contributions in this respect of Louis[8], Louis (1972), Keyfitz[9], Cox and Miller[4], Chiang[3] and Sheps (1967), and her associates are of worth reading.

Major studies on stochastic modeling of human fertility have also been done at various institutions of India. Some are worth mentioning as International Institute for Population Sciences (IIPS), Mumbai, Indian Statistical Institute (ISI), Kolkata, Centre for Development Studies, Jawaharlal Nehru University, Delhi, Department of Demography, Kerala University, Thiruvananthpuram. The researchers at the Centre of Population Studies, Department of Statistics, Banaras Hindu University, Varanasi have also made significant contributions in the field of Stochastic Modeling. Recently, two chapters on Stochastic Modeling have been published in the two volumes of 'Handbook of Statistics', Integrated Population Biology and Modeling, Vol. 39 and Vol. 40 edited by Arni Srinivasa Rao and C. R. Rao published in the years 2018 and 2019, respectively.

The two chapters are entitled as "Stochastic Modeling of Some Natural Phenomena: A Special Reference to Human Fertility", and "Analyzing Variety of Birth Intervals: A Stochastic Approach". These two chapters extensively describe some of the works done at the Department of Statistics, Banaras Hindu University related to various aspects of stochastic modeling of human fertility. It is to be mentioned that these two chapters do not include an exhaustive list of the work done here in the field of stochastic modeling. Furthermore, the present paper includes only those works where the author of the present article is either main author or a co-author.

[1] *Former Professor, Department of Statistics, Banaras Hindu University, Varanasi, India.*
**rcyadava66@yahoo.co.in*

Springer

Normally, most of the studies are extensions of previous works, while some are based on original thoughts. Usually, the researchers face failures frequently before the final result of any research work, but in some cases, the success comes even at the first attempt itself. On the same line, the young researchers also face many failures in their early research career, and become discouraged, nervous, and frustrated.

The present article is aimed to throw some light on this issue. The author of the present paper also faced many hurdles in his research career and he cites many examples of his failures before success. It is hoped that the researchers, especially young researchers, will learn some new lessons which may be helpful in their future research endeavors.

Since, the stories of failures are normally not mentioned in the final publication of the research work, the author of present article decided to bring such stories of his failures before the researchers and academicians of the field. The article also cites many examples that how a simple thought paved the way for a new result. It is further emphasized that since the stories are related to specific papers, hence the readers are advised to be aware about the fundamental ideas related to the topic of concern. Although, they may not be very much clear about the complex mathematical derivations.

The concept of model, various aspects of model building, their uses, and limitations are extensively described in the two chapters mentioned above as well as in many books and other available literature. Therefore, we are not repeating these concepts here again. The readers are advised to consult these two chapters or other available literature elsewhere for the same.

Now, we present the stories behind our various papers given below. It is also pertinent to mentioned that these stories are not presented in a chronological order, although the similar works have been presented at one place; for example, the models related to number of birth in a given interval of time or births at a given time, variety of birth intervals, and estimation of parity progression ratios utilizing data on open and closed birth intervals have been put in the above sequence without considering their year of publication.

The notations and terminologies such as conception rate/fecundabilty, non-susceptible period have been already explained in our papers, Yadava[37] and Yadava and Rai[36]. Thus, we do not provide the same in the present article. It is also important to mention that different symbols for conception rate, non-susceptible period have been used in different original papers published in various Journals. We have used the same symbols as published in those original papers as we have been advised to see the original papers for more details.

In the context of the present article, it is pertinent to mention the following informations:

A "Demographic Research Centre" presently known as "Centre of Population Studies" was established in the year 1969, in the Department of Statistics, Banaras Hindu University (B.H.U.), Varanasi sponsored by Population Council, New York, USA.

The major objectives of the Centre were:

1. To do basic research in Demography, especially building of appropriate stochastic models and their characterizations and related issues of multidimensional facets of population dynamics.
2. To collect relevant data on various aspects of population dynamics to test the adequacy of different models, developed by the researchers working in the Centre.
3. To train researchers in collection, compilation, and analysis of different types of population data.

The first objective was a long-term goal. The second and third objectives were achieved, at least partially, by conducting "A Demographic Survey of Varanasi (Rural) (1969–1970)" in which data on various aspect of population, such as fertility, mortality, migration, and family planning, were collected from about 2500 households from rural areas of Varanasi.

The data collected in the above survey were much helpful in testing the adequacy of various stochastic models developed by the researchers at the Centre. In addition, the students and researchers who were engaged for the collection of data got practical training to obtain data especially from rural areas.

Since a major objective of the center was to develop stochastic models for various population phenomena, the researchers at the center developed different stochastic models for various aspects of population dynamics consistently overtime. Now, the descriptions of those developed models are presented here. The author of this article was also associated with the development of different stochastic models under different situations. In the development of these models, the authors also faced various hurdles before getting the final result. The present article gives a description of various thoughts which crept in mind in the development of a model and frequent failures before getting the final success.

Now, we present the stories related to specific papers by mentioning the title of the papers, names of authors, and names of Journals where these are published.

2 [A] A parity-dependent model for number of births and its applications

By Singh, Bhattacharya, and Yadava in **Sankhya Series B**[26].

Dandekar[5], Brass[2], and Singh[23,24] proposed probability models for number of births in a given period (0,T) based on some simplified assumptions, including assumption of one-to-one correspondence between conception and birth. Singh and Bhattacharya[25], and Singh and Bhattacharya (1971) published two papers incorporating possibility of foetal loss called by them as incomplete conception. Conception leading to birth was called 'complete conception' by them.

All these models were based on a primary assumption that the conception rate/fecundability remains constant (except during non-susceptible period, where it is zero) throughout the period of observation, say (0, T). It was noticed that the assumption of constant conception rate may be reasonably true for a short period say around 5–10 years, but it may not be reasonable for a larger period say 10 years and above.

When fertility data were being analyzed, it was found that conception rate was considerably low for the period of marriage to first birth in comparison to the conception rate between other consecutive births. This compelled us to think that the assumption of constant conception rate in (0, T) may not be reasonable especially for the locality where the survey was conducted. Considering the above situation, Bhattacharya[1] proposed a model for number of births in (0, T) considering conception rate to be time-dependent. However, the probability expressions were so much complicated that there was almost no scope for applying the model to real data, so that useful conclusion may be drawn.

Keeping the above limitation of the proposed model, it compelled us to think for some alternative way. One day, authors of the paper were discussing about the situation, and then, it came in our mind that the assumption of conception rate to be time-dependent was not workable, so, we decided to assume that it may be parity-dependent instead of time-dependent.

However, no ready solution was available in our knowledge at that time. In this context, it was noticed that the results on birth process are somewhat similar to our thought of assumption of parity dependence of the conception rate, because in birth process also, it is assumed that probability of occurrence of an event in an interval (t, $t+\Delta t$) is dependent on the number of events which have already occurred prior to t'.

Finally, the above paper was prepared in accordance with the assumptions of the human reproduction process. The paper was submitted to Sankhya and the editorial committee of Sankhya agreed to publish the above paper.

It is pertinent to mention that the parity-dependent model was helpful in explaining the variation in the number of births in (0, T), but at the same time, it was also very much helpful in evaluating the impact of any hypothetical family planning programme conceived at that time. This is illustrated in the paper considering various hypothetical family planning programmes.

3 [B] An adjustment of a selection bias in post-partum amenorrhea period from follow-up studies

By Singh, Bhattacharya, and Yadava **in Journal of the American Statistical Association, JASA (1979).**

As mentioned earlier, 'A Demographic Survey of Varanasi (Rural)' was conducted in 1969–1970. Here, apart from other information, the data on the duration of Post Partum Amenorrhea Period (PPA) were also collected for all births occurring during last 7 years before the survey date. The reference date of the survey was Dussehra (Durga Pooja) of 1969 which was in the month of October of 1969.

The PPA was considered to be an important component of fertility behaviour of females at that time. When the data on PPA duration were to be analyzed, it was thought to examine whether there was any memory bias in the reporting of PPA duration by females. For this purpose, all the births during last 7 years were classified according to their year of birth, i.e., 1962, 1963, 1964, 1965, 1966, 1967, 1968, and 1969. The average duration of PPA related to these year-wise births was computed to examine the possibility of memory bias. Of course, the births occurring in years 1968 and 1969 were excluded, because PPA durations for many of these births might have been continuing at the time of survey. When we examined the results, it was seen that average durations of PPA related to the births of different years were almost equal, indicating that there was no evidence of any memory bias. However, when we computed the average PPA duration related to last but one (next to last) birth, it was found that this average duration was somewhat larger than the average durations of PPA related to births in different years. We tried to search for some plausible explanation, but we were unable to give any convincing explanation for this difference. Ultimately, we abandoned to think on this issue. This was sometime during 1971–1972.

Time passed. One day, the author of this article was sitting in his lawn during winter of 1974–1975, and an idea came in his mind.

Suppose, there are two types of coins. The probability of getting head (success) on first type of coin is 0.5, while probability of getting head (success) for second kind of coin is 0.3. Both the types of coins are in equal proportions i.e., 0.5, Thus, the average probability of success, say $\overline{p}$ will be

$$0.5 \times 0.5 + 0.5 \times 0.3 = 0.4.$$

Suppose, we take 100 coins of each type and toss them once, and then, on an average, there will be 50 coins of first type where success will occur, while there will be 30 coins of second type on which success will occur. Thus, if we take these 80 coins on which success has occurred, then average probability of success for these coins will be

$$\frac{50}{80} \times 0.5 + \frac{30}{80} \times 0.3 = \frac{34}{80} = 0.425,$$

which is larger than $\overline{p} = 0.4$.

This gave the clue that if the coins are of heterogeneous type, then the average probability of success on coins which show a success in any toss will be somewhat larger than the average probability of success on all coins of the population. On this thought, we felt that the females giving birth at a particular time may be somewhat more represented by those females who give births more frequently and such females may be those for whom PPA duration is less.

Next day, we discussed this issue among us and prepared a paper giving rigorous mathematical derivations to explain the phenomenon and sent the paper to Journal of American Statistical Association (JASA) in early 1975 for possible publication. Almost within 2 months, we got reply from JASA, that the editorial committee of JASA is unable to publish the paper mainly on following two grounds:

1. Although the results contained in the paper are interesting as it shows that average duration of PPA of females giving births at a particular time will be smaller than the average duration in the population, but the authors here have not given any solution of this bias adjustment (later on, this was called as selection bias).
2. The results mainly relate to the study of PPA duration which is a demographic issue, so the paper may be sent to some Demography journal for possible publication.

(The reader is advised to study our earlier papers of Singh et.al (1979) and Yadava and Rai[36] to understand about the selection bias.)

Let us consider a population of married women heterogeneous with respect to non-susceptible period g and conception rate φ. Let the joint and marginal distributions of φ and g be denoted as $f(\varphi, g)$, $f_1(\varphi)$ and $f_2(g)$, respectively.

And let $\psi_{\varphi,g} = \varphi/(1 + g\varphi)$ be designated as the asymptotic fertility rate for given φ and g. Also, let the asymptotic fertility rate for given g is ψ_g and ψ_φ be one when g is given. Furthermore, let ψ denote the asymptotic fertility rate in the population. In a population where the women giving birth in a small interval of time are included in a study, the distribution of φ and g will be

$$f_3(\varphi) = (\psi_\varphi/\psi)f_1(\varphi) \text{ and } f_4(g) = \frac{\psi_g}{\psi} \cdot f_2(g).$$

In our earlier paper which was sent for publication and was not accepted for publication, contained only the above result showing the nature of selection bias under different conditions. However, no solution was given for finding $f_2(g)$ from $f_4(g)$, because we were not able to compute the value of ψ for a given population.

Due to not availability of ready solution at that time, the idea of publishing the above paper was almost abandoned.

Around 2 years passed. One day, the author of the present paper was sitting in his room in the department and a point clicked in his mind that instead of considering the equation

$$f_4(g) = \frac{\psi_g}{\psi} \cdot f_2(g),$$

we may consider the equation

$$f_2(g) = \frac{\psi}{\psi_g} \cdot f_4(g),$$

which may yield an easier methodology for finding ψ. Since ψ is a constant and $\psi_g = \frac{\Phi}{1+g\phi}$ hence,

we get,

$$f_2(g) = \frac{\psi}{\frac{\phi}{1+g\phi}} \cdot f_4(g)$$

$$= \frac{\Psi}{\phi}(1+g\phi) \cdot f_4(g).$$

Since $f_2(g)$ is a probability distribution, hence $\int f_2(g)dg = 1$, which gives the result

$$1 = \Psi \frac{1+\overline{g'}\phi}{\phi}$$

that is

$$\Psi = \frac{\phi}{1+\overline{g'}\ \phi},$$

where $\overline{g'}$ is the mean of non-susceptible period associated with $f_4(g)$. Since the distribution of $f_4(g)$ is known, hence ψ can be easily computed for given Φ.

This gave the solution for finding $f_2(g)$ from $f_4(g)$. Then, a revised paper was prepared and sent to JASA again. This time also, the editorial committee of JASA appreciated our results, but still they said that the problem relates to Demography, and hence, it may be sent to some Demographic journal. Then, we replied to the editorial committee that although the problem relates to Demography, but essentially, it is a statistical problem which occurs due to heterogeneity in the population. Therefore, the committee agreed on our justification and finally accepted the paper for publication in JASA.

Thus, a problem which was initiated for investigating the memory bias culminated into the study of impact of heterogeneity on the distribution of a duration variable, i.e., PPA period in this case. This paper became a bench mark for further studies on impact of heterogeneity on the distributions of different duration variables.

4 [C] Mean birth interval characteristic of women

By Singh, Yadava, and Pandey in **Canadian Studies in Population**[28]**.**

After publishing the paper in JASA for adjustment of selection bias in PPA period, we decided to apply the same technique for the adjustment in birth interval data. However, we noticed that problem of selection bias in birth intervals has been solved by Wolfers[35], but still we decided to attempt for it by an alternative procedure discussed in our JASA paper.

The paper was prepared following the procedure used in our JASA paper and expressions were also developed similar to expressions in Wolfers[35]**.** When the results by our procedure and those from Wolfers' procedure were compared, results from both the procedures were almost equal. Finally, the paper was sent to the journal of Canadian Studies in Population and it was accepted for publication with minor revision.

Thus, a paper for which we were initially hesitant to send for publication finally got published in a good journal.

5 [D] A generalised probability model for an equilibrium birth process

By Singh and Yadava in **Demography India**[27]**.**

We have mentioned earlier that Dandekar[5] published a paper for the probability distribution of number of births in an interval (0, T). In the same paper, he has also obtained expressions for probability distribution of number of births in abrupt sequence of trials of a given length. Though, the paper was quite difficult for me to understand, however, probability expressions could be understood easily. To check the validity of probability expressions derived by Dandekar for abrupt sequence of trials, we took some hypothetical values of m, h and T (for symbols see our paper Singh and Yadava[27] and obtained the probability values for $X = 0, 1, 2, \ldots$, n.

For the selected values of m, h and T, the probabilities were in the range (0,1) and sum of probabilities was '1' and average number of births was also equal to $\frac{\text{mT}}{1+\text{mh}}$ which was a theory requirement. Consequently, we became quite sure that the probability expressions derived by Dandekar were correct.

Thus, there was almost no scope to derive these expressions further. Almost at the same time, Singh and Bhattacharya[25], extended the model of Singh[24] incorporating the possibility of foetal loss in the time interval (0, T).

Therefore, it was thought that, could we include the possibility of foetal loss in the model of Dandekar for abrupt sequence of trials? We attempted for this, but could not succeed to get any useful result.

Dandekar[5] has obtained the expression for $P[X=r]$, $r=0,1,2\ldots, n$ where X denotes the number of births in (T^*, T^*+T) where T^* is large, such that the system is in steady state (or equilibrium). The probability expressions were

$$P[X = r] = \Phi_1(r-1) - 2\Phi_1(r) + \Phi_1(r+1), \quad r = 0, 1, 2, \ldots, n,$$

where

$$\Phi_1(r-1) = \frac{1}{1+mh}\sum_{x=0}^{r-1}\sum_{k=0}^{x} e^{-m(T-rh)}\frac{m^k\left(T-ih^k\right)}{k!}, \quad r = 1, 2, \ldots, n \text{ and } 0 < T < \infty.$$

One day, author of this article was sitting in his room in the department and a thought came in his mind that what would be the probability expression if $T<h$.

Obviously for this situation, X can take values '0' or '1'. Since the mean number of births in the interval of length T will be $\frac{mT}{1+mh}$ and X takes only two values 0 or 1, hence, $P[X=1]$ should be $\frac{mT}{1+mh}$, while $P[X=0]$ should be $1-\frac{mT}{1+mh}$

When we tried to obtain these values by Dandekar formula, then these did not match with the values mentioned above, i.e., $1-\frac{mT}{1+mh}$ and $\frac{mT}{1+mh}$.

This created a doubt about Dandekar's formula at least for smaller values of T.

Then, our emphasis shifted towards finding appropriate probability expressions for all values of T (i.e., whether small or large). For this, we were to search for an alternative way to derive the probability distribution in equilibrium birth process.

For this purpose, we thought to apply the technique of finding the distribution of $T_1+T_2+\cdots+T_k$ using Laplace transform, where T_1 is the time interval from T^* to occurrence of first birth after T^*, while T_i is the interval between $(i-1)$th and ith births, $i=2, 3, \ldots, n$. Furthermore,

$P\ [T_1+T_2+\cdots+T_k<T]$ implies that at least k births occur in (T^*, T^*+T) from which $p[X=r]$, $r=0, 1, 2, \ldots$, n, can be easily computed.

The details for finding the distribution of $T_1+T_2+\cdots+T_k$ and $P\ [T_1+T_2+\cdots+T_k<T]$ are explained in Yadava[37]. After finding these expressions, it was found that the formula derived by Dandekar was true for $P[X=r]$ for $r=0, 1, 2, \ldots, n$-2 but needed modifications for $r=n-1$ and n. These facts are described in detail in Yadava[37].

Finally, in Singh and Yadava[27], the possibility of foetal loss was also included in the published paper. Thus, an unexpected result was derived by incidentally considering the value of T to be less than h.

6 [E]. On the distribution of births over time in an equilibrium birth process for a female giving specified number of children in a given period

By Yadava and Srivastava in **Demography India**[47].

It is already mentioned that models for number of births in a given time period (0, T) or (T^*, T^*+T) were developed. As a curiosity, it was thought that could we model for number of surviving children out of the births in a given period at the time of end of the interval, because many of the fertility behaviours are more dependent on number of surviving children rather than number of children born.

At first, the problem was thought to be very simple, just a simple application of binomial distribution. As such, if there are n births in an interval, then the probability distribution of number of surviving children out of these n births would be

$$P[X=r] = \binom{n}{r} p^r q^{n-r}, r = 0, 1, 2, \ldots, n,$$

where X denotes the number of surviving children, p denotes the probability of survival of a child; $q=1$-p and $0<p<1$.

However, when a further look on the matter was given, then it was seen that the problem is not so simple, because the survival probabilities for these n children will not be the same, the elder children will face more exposure time for survival, while younger children will have less exposure period for survival. However, binomial distribution assumes constant probability of success (here survival). Thus, the problem needed further investigation. For this, we confined our attention for equilibrium birth process. As we know that in equilibrium birth process, births are uniformly distributed over time and thus, a natural question arose:

Are births uniformly distributed over time for females giving a specified number of births (say n) in the interval?

The issue was quite open. We were not aware of any solution for the problem. We tried to solve it mathematically, but could not succeed, because there were large number of possibilities of occurrence of a given event.

One day, it clicked in our mind that if there is any interval of length $\leq h$, then there are only two possibilities: Either there will be a birth in the interval or there will be no birth in the interval. Furthermore, it was also clear in our mind that if in an interval of length h, there was no birth, then the female will be exposed to the risk of birth at the beginning of the subsequent interval of length h.

This gave a thought for us, i.e., if we divide the interval of length T in segments of length h, then in each segment of length h, there can be either zero birth or one birth and computations of probabilities may be easier.

Then, we took the simplest case. Suppose the length of the interval T is h, i.e., $T = h$, then there are only two possibilities and

$$\mathrm{P[X = 0]} = 1 - \frac{mh}{1 + mh}$$

and $\mathrm{P[X = 1]} = \frac{mh}{1+mh}$ (see Singh and Yadava[27]).

Now, suppose we take $T = 2$ h and divide it into two intervals of length h, then there will be following four possibilities:

1. There is 0 birth in first segment as well as 0 birth in second segment. We may denote it as (0,0).
2. There is 0 birth in first segment and 1 birth in second segment. We denote it as (0,1)
3. There is 1 birth in first segment and 0 birth in second segment. We denote it as (1,0)
4. There is 1 birth in first segment and 1 birth in second segment, which may be denoted as (1,1).

Let the probabilities for these events be denoted as $P_{00}, P_{01}, P_{10}, and\ P_{11}$, respectively.

Now, P_{00} is nothing but probability of exactly '0' birth in the interval of length 2 h, which can be easily computed from Singh and Yadava[27].

Alternatively, P_{00} gives the probability of the event that '0' birth occur in first segment as well as in the second segment.

The probability of '0' birth in first segment is 1-$\frac{mh}{1+mh}$

$$= \frac{1 + mh - mh}{1 + mh}$$

$$= \frac{1}{1 + mh}$$

and the conditional probability of '0' birth in second segment given that there is 0 birth in first segment is e^{-mh}.

Thus, $P_{00} = \frac{1}{1+mh} \times e^{-mh}$.

One can check that the probability of '0' birth in the interval of length 2 h is also $\frac{1}{1+mh} e^{-mh}$.

Similarly, P_{11} is nothing but the probability of exactly 2 births in interval of length 2 h which can be easily computed from Singh and Yadava[27].

Now, P_{01} is easily computed by probability of 0 birth in first segment × conditional probability of one birth in second segment given that there is '0' birth in first segment: thus

$$P_{01} = 1 - \frac{mh}{1 + mh} \times \left(1 - e^{-mh}\right).$$

Now, if $P_{(2h)}(1)$ denotes the probability of one birth in interval of length 2 h, then

$$P_{(2h)}(1) = P_{01} + P_{10}.$$

Since $P_{(2h)}(1)\ and\ P_{01}$ are known, hence P_{10} can be easily computed.

Once the values of $P_{00}, P_{01}, P_{10}, \text{and} P_{11}$ are known, one can easily compute the values of similar probabilities for $T = 3$ h, utilizing the above concept. Proceeding in similar manner, one can easily compute 2^n probabilities of similar type for $T = nh$.

The comprehensive proof of the above result is given in Yadava and Srivastava[47]. (For more details, one can see the above paper.)

Taking certain hypothetical values of m and h, (close to reality), computations of probabilities for $n = 1,2,3,4,5$ (considering $T = nh$) were done. It was seen that the probabilities are symmetrical in nature, for example $P_{011} = P_{110}$ or $P_{0010} = P_{0100}$, etc. However, it was also seen that (for example) $P_{100} = P_{001}$, but these two were not equal to P_{010}. In all the above 3 cases, the probabilities denote that exactly one birth occurs in an interval of length $T = 3$ h. This gives the evidence that births occurring to a female giving specified number of children in a given period are not uniformly distributed over time. This gave an answer to our initial query for which the research was initiated.

It is important to mention that the above-mentioned results are based on assumptions of the model. However, if the assumptions are changed, nothing definite can be said about the situation.

Based on the above results, Srivastava (1992) also computed the probability distribution of number of surviving children.

Although the results contained in Yadava and Srivastava[47] were obtained by utilizing simple laws of probability, but many of the researchers have shown keen interest on the methodology of the paper, i.e., dividing the interval $T = nh$ into n segments of length h.

It is also to be mentioned that observation on symmetry of probabilities was based on certain computed values for assumed values of n and h. However, no proof for this observation was given by Yadava and Srivastava[47].

Later on, Yadava and Tiwari[49] obtained similar probability expressions for the interval (0, T). Obviously, the probability expressions were not similar. For more details, reader is advised to read the above paper.

Incidentally, it may be mentioned that the paper of Yadava and Tiwari[49] was adjudged as the best technical paper published in 'Demography India' in that year.

Later on, Yadava and Tiwari[50] gave a mathematical proof for the symmetry of probabilities under the assumptions of the model. This paper further relaxes the condition that length of each segment should be h'. In this study, they divided the length of interval T into segments of 1 year each (less than h) and obtained similar results. (For more details, the reader is advised to read the above paper in detail.)

7 [F]. A study on the fertility of migrants

By Singh, Yadava, and Yadava **in Health and Population: Perspectives and Issues**[30].

As mentioned earlier, the Demographic Survey of Varanasi (Rural) was conducted in 1969–1970. In this survey, detailed data on fertility, mortality, and migration were collected. First, on priority basis, fertility data were analyzed. Applying appropriate mathematical models, the values of monthly conception rate/fecundability were estimated. The estimates of monthly conception rate/fecundability came in the vicinity of 0.05–0.07 for different situations. Almost at the same time, Sheps and her associates reported the value of fecundability to be around 0.25 for American females, which was too high in comparison to our estimates.

No ready explanation was available at that time for this large difference. However, we reported that the low estimates of conception rate for rural females of the study area may be mainly due to observance of various socio-cultural norms for the frequency of coition. Still, the gap in estimates was too large.

It is important to mention that the data were collected from about 2500 households from the rural areas of Varanasi. Normally, collection of data on migration was not possible if we adopted the usual definition of a household in rural areas, i.e., a group of persons living together and taking food from a common kitchen normally related by birth or marriage. Since all the persons of a household were living in the household itself, hence there was no possibility of any migration date. Therefore, to obtain data on migration, an extended definition of a household was adopted. This definition included such members of the household also who were usually migrated, living elsewhere (normally in big cities) but claim the household to be their own. They normally had all social and economic links with the household and used to visit the household once or twice in a year. Such migrated persons were normally males migrated to elsewhere but leaving their wives in their village. The data on fertility of such migrated couples were also collected.

Therefore, it was thought to study the fertility level of such couples separately. It is to be noticed that in case of such couples, the male and female of the couple used to be separated for a long period. Therefore, it was believed that whenever the male visited the household, the coital frequency would be definitely large. Thus, the coital behaviour will be definitely different from the behaviour of those couples where male and female were living together in the household itself. Therefore, it was thought that, if we can estimate the fecundability level of such couples, then it may represent the fecundability level of females of the area who do not follow any restriction on limiting their coition.

Now, the issue became: how to estimate the fecundability level of such couples? Then, a thought came in our mind that if the male visits in a year and the female becomes pregnant, then if he visits the household next year, then the female is likely to be in her non-susceptible period and she will be susceptible to the risk of conception only when the male visits the household in the subsequent year. Keeping in view the above facts, a probability model was developed and estimate of fecundability of such couples was obtained. The estimate came in the vicinity of 0.2, which was much higher than the estimate for females where male and female were living together in the village itself (for details see Singh et al.[30] and Yadava[37]). This gave the evidence that the fecundability of rural females is low mainly

Springer IISc PRESS J. Indian Inst. Sci.|VOL 102:4|1111–1130 October 2022|journal.iisc.ernet.in

due to observance of many socio-cultural behaviour/norms regarding coital behaviour. However, it is further to be pointed out that the above observations are applicable mainly for the time of 1969–1970 survey. More than 50 years have already passed. With the passage of time, the fertility behaviour now might have changed substantially. Therefore, the present studies must be done in the context of present time.

8 [G] Sex ratio at birth: a model-based approach

By Yadava, Anup Kumar, and Srivastav **in Mathematical Social Sciences**[39].

Son preference has been a common phenomenon not only in India but around the world. Initially, it was conjectured that in the presence of son preference, the sex ratio at birth may tilt towards male births. However, it was mathematically shown that whatever be the sex preference of child, the sex ratio at birth will remain unchanged.

Therefore, people did not show further interest on this issue. However, after the passing of MTP Act (Medical Termination of Pregnancy Act), people started taking shelter of MTP Act to terminate the female foetus resulting in alternation of sex ratio at birth. When the data were analyzed, it was seen that the problem was more severe in the states of Punjab and Haryana, though the phenomenon was present in other states also.

Researchers at that time started studying the phenomenon of "missing girl child" through data available from different sources. Quite lately, it came to our mind that the phenomenon of missing girl child can also be studied through the help of mathematical models. Then we considered various hypothetical stopping rules according to the sex-wise distribution of children accounting the possibility of induced abortion for an unwanted child of undesired sex (say girl). This was achieved making use of simple laws of probability and conditional probability. Consequently, this paper was published by[39,39]

Later on, it was detected that there was a mistake in the probability expressions simply a shift of origin which was rectified by Pandey et al.[16]. (For detailed discussion on the issue, one may read Yadava[37] also.)

9 [H]. Measuring son preference through number of children born

By Shukla, Yadava, and Tiwari **in Demography India (2018).**

Once author of the present article was delivering an invited talk at Women's College (Mahila Maha Vidyalaya) Banaras Hindu University, the topic of talk was on 'son preference' using concepts of statistics.

As an example, suppose couples have a strong son preference and decide to produce children until they get a son. Obviously, the number of children born say X (ignoring mortality) will take values 1, 2, 3,… If p denotes the probability that a born child will be male, the probability distribution of X will be a geometric distribution.

That is, $P[X=x]=pq^{(x-1)}\ x=1,2,3,\ldots q=1\text{-}p$.

That is, q represents the probability that the born child will be female.

The topic was mainly chosen to create a curiosity among audience to study the phenomenon of son preference using concepts of statistics. If the value of 'p' is taken as $\frac{1}{2}$, then

$P\,[X=1]=\frac{1}{2}$

$P\,[X=2]=\frac{1}{4}$

$P\,[X=3]=\frac{1}{8}$,

and so on.

Furthermore, $E(X)=1\times\frac{1}{2}+2\times\frac{1}{4}+3\times\frac{1}{8}+\cdots=2$.

Thus, on an average, a couple will produce 2 children out of which one will be a male. Thus, the sex ratio at birth is unaltered even in the presence of son preference. However, the social implication of the above result would be that larger families will include large number of female children, while smaller families will have more males, making a more disbalanced dowry system.

Therefore, it was thought: whether we can study about the sun preference through data on sex-wise distribution of births? In this context, the data of NFHS-2 (available at that time) were analyzed regarding the sex composition of specified number of children when the couples stopped producing children. From the analysis of data, it was found that if the number of born children is 2, then there will be 4 possible combinations, viz., MM, MF, FM, and FF, where M stands for male child, while F stands for female child. Theoretically, if there is no son preference, then each combination should be of 25% (if $p=\frac{1}{2}$). However, from the analysis, it was found that the percentage for MM was significantly larger than 25%, while for FF, it was much below 25%.

These results were presented before the teachers and research scholars of the department. After listening the presentation of the results, many of the research scholars and teachers pointed out that deviation from 25% may be due to reporting bias (under-reporting of female children). We tried to convince them, but they were adamant about the fact that since the three conditions of Bernoulli trials are satisfied here, hence, if $p = \frac{1}{2}$, then all the 4 percentages should be nearly 25%.

Similar arguments were also given for 3 children. The main aim to report the above story is to demonstrate that many times people hurriedly jump to conclusions without checking the validity of assumptions.

In fact, here, the three assumptions of Bernoulli trials, viz.

1. There are only two possible outcomes of a trial.
2. Probability of success remains constant from trial to trial.
3. Trials are independent are satisfied. However, the fallacy lies with the fact that the decision to perform the *n*th trial depends on the outcomes of previous (*n*-1) trials.

10 [I] Computation of prevalence/ incidence mean

By Shukla in his unpublished Ph.D. Thesis Entitled 'A study on Mathematical Models for Population Dynamics, submitted **in Banaras Hindu University**[22].

While reading NFHS-2 Report, the author of the present article came across the word "Prevalence/Incidence Mean". While describing about the distribution of PPA, the NFHS-2 Report mentioned about Prevalence/Incidence Mean. Since, author was totally unaware of this term, he inquired about it from Dr Arvind Pandey, the then Director of National Institute of Medical Statistics (NIMS), Delhi as he was previously associated with IIPS, Mumbai which was the nodal agency for conducting NFHS-2. Dr. Pandey told that although he has heard about the relationship of incidence, prevalence, and duration of disease in Biostatistics and Epidemiology, but was not aware as to how the mean duration of breastfeeding/PPA is derived from the current status data on the same in NFHS report. When author contacted another person related to analysis of NFHS data, he plainly told that the mean has been mechanically computed and he was not aware of the theory behind it. Author of this article also contacted two professors associated with teaching of Biostatistics/ Epidemiology, but they also did not provide any satisfactory explanation for the same. Almost at the same time, we were doing some work on open birth interval and it came to our mind that the 'Prevalence' seems to be quite similar to 'open birth interval'.

In the PPA case, the prevalence is nothing, but the number of females who have not yet completed their PPA at the time of survey, while open birth interval relates to females who have not yet given their next birth after their last birth. Thus, in both the cases, the concept of [1-$F(x)$] can be applied where $F(x)$ is the distribution function associated with the random variable X; in one case, it relates to PPA duration, while in the other case, it relates to closed birth interval. Thus, we could guess that $\frac{\text{Prevalence}}{\text{Incidence}} = \int [1 - \text{F(x)}]$ dx = mean.

However, we left the issue, because the computed mean by the method was almost true, although the proof was not given in NFHS-2 Report. Author of the present article still does not know whether the proof for it has been given in any Technical Report of NFHS or not, but we can only say that the technical staff engaged in the analysis of data was mostly not aware about the theory behind it. Mostly, they were doing mechanical calculations. Since the result was almost true, we forgot about it for any further study. One day, we saw that in the same NFHS report, Prevalence/Incidence Mean for breastfeeding duration was also reported. We did not care much about it, as we thought that it is almost similar to Prevalence/Incidence Mean for PPA duration.

It is to be mentioned here that for the computation of Prevalence/Incidence Mean, births during last 3 years before the survey were considered. However, we noticed that many of the females were still continuing their breastfeeding at the time of the survey. Thus, if we increase the observational period, the prevalence will increase, while incidence will remain the same. Thus, the mean will increase for the case.

This is mainly because $\int_0^c [1 - F(x)]dx = mean$ only if [1-$F(x)$] = 0 for all $x \geq c$.

In case of breastfeeding, P[(1-$F(x)$) > 36 months] is not zero, and hence, the technique of Prevalence/Incidence cannot be applied unless the observational period is increased to 4–5 years or more. (For more details, one can read the above-mentioned thesis.)

Our main objective to mention about this problem here is to emphasize that before doing any computation mechanically, one should know the theory behind it; otherwise, sometimes, one may obtain incorrect or misleading results.

11 [J]. Estimation of probability of coition on different days of a menstrual cycle near the day of ovulation: an application of theory of Markov chain.

By Yadava, Shruti Verma, and Singh **in Demography India (2015).**

Quite long back (around 1990), author of the present article was reading a book entitled "Society and Fertility" by Potts and Selman (1979) which mainly discussed the interactions among various social issues and fertility.

In the book, at one page, it was shown through a graph that sex of a child is somewhat dependent upon the difference between the time of coition and the time of ovulation. If the day of ovulation and day of coition which results into conception are same, then the probability that the child will be male or female will be almost equal, i.e., 0.5, while if the difference between the day of ovulation and the day of conception is more, then the probability that the child will be male is slightly more than 0.5.

This statement was very much exciting for us, because the studies in the field of human reproduction are still not conclusive; that is why, the probability of male birth in the population is more than half, i.e., around 0.514. It was evident that if the above statement by Potts and Selman is true, then it can be a possible explanation for observing the probability of male birth to be around 0.514 rather than 0.5.

However, when we read some other research papers related to the issue, the results were not conclusive. Alternatively, we thought to explore the problem through some mathematical model. However, we could not succeed in our effort even after 1 decade.

It is worthwhile to mention here that the author of the present article taught the course on Stochastic Processes for more than a decade. One day, it clicked in his mind that the *n*th power of a transition probability matrix of a Markov chain becomes almost stable if *n* is large in certain situations. Then the idea came to the mind to solve the issue with the help of theory of Markov chain.

For any study through Markov chain, first, the 'states' of the chain should be clearly defined and the transition probabilities of the chain should also be clearly specified.

Now, the problem became: how to specify the states of the Markov chain? No ready-made solution was available to us. We had to define the states on our own. Since, the matter was to know, whether a coition occurs on the day of ovulation or not, we defined the 'states' as the day of last coition before the day under consideration. For simplicity, it was assumed that if a female has a coition on any particular day, then in the next 5 days, she will definitely have another coition also. Consequently, the following six states $E_0, E_1, E_2, \ldots E_5$ were defined where E_i(i = 0, 1, …, 5) represents the 'state' that last coition occurred *i* days before the considered day.

Now, the problem was to specify the various transition probabilities of the 'chain'. It is evident that the transition probabilities cannot be specified arbitrarily with the condition that the row sums are unity in one step. According to the definition of states, it is evident that the system can move from E_i to either E_0 *or* E_{i+1} only ($i = 0$, 1, 2, 3, 4, 5).

Hence, transition probabilities will be positive only for these two states. Furthermore, if the difference between day of last coition and the considered day increases, then the probability of next coition should normally increase. Keeping in view the above considerations, a few hypothetical transition probability matrices were considered and their *n*th powers were obtained.

For example, one considered matrix is given below:

$$\begin{array}{c} \\ E_0 \\ E_1 \\ E_2 \\ E_3 \\ E_4 \\ E_5 \end{array} \begin{array}{c} \begin{array}{cccccc} E_0 & E_1 & E_2 & E_3 & E_4 & E_5 \end{array} \\ \begin{bmatrix} 0.2 & 0.8 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.3 & 0.0 & 0.7 & 0.0 & 0.0 & 0.0 \\ 0.4 & 0.0 & 0.0 & 0.6 & 0.0 & 0.0 \\ 0.6 & 0.0 & 0.0 & 0.0 & 0.4 & 0.0 \\ 0.8 & 0.0 & 0.0 & 0.0 & 0.0 & 0.2 \\ 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \end{bmatrix} \end{array}$$

We clearly observe here that $p_{00} < p_{10} < p_{20} < p_{30} < p_{40} < p_{50} \leq 1$ where p_{ij} is the transition probability from state E_i *to* E_j in one step. It is also computed that the mean first passage time from E_0 *to* E_0 (i.e.. mean recurrence time forE_0) equals to 2.86 days. This shows a reasonable value for mean recurrence time for E_0. The detailed discussion is given in the above paper.

It was found that for the considered cases, the matrices became stable for $n \geq 10$ and the rows were almost identical. Consequent results are discussed in detail in the above paper.

Thus, a problem which remained unsolved for us for more than 2 decades was finally solved using the theory of Markov chain. It is true that it is just a modest beginning on the issue which can further be extended in near future.

Till now, we described stories related with mainly various models on number of births in the given interval and related topics. Now, we will describe stories related to various probability models on different types of birth intervals.

12 [K]. A probability model for forward birth interval

By S.N. Singh, R. C. Yadava, and Arvind Pandey **in Health and Population: Perspectives and Issues**[29].

Srinivasan[31–34], popularised the concept of open birth interval. Results were also available on straddling birth interval. People were aware of waiting time paradox[7].

In the context of renewal theory, the concepts of backward recurrence time (Left fraction) and Forward Recurrence time (Right fraction) were also known. From renewal theory, it was known that probability density function for backward recurrence time was $\frac{1-F(x)}{\mu}$, where $F(x)$ is the distribution function of X, while μ is the mean of X. Forward Recurrence time was also having the same probability density function, viz.,$\frac{1-F(x)}{\mu}$.

Various uses of open birth interval were demonstrated. However, no attention was paid towards the practical use of Forward Recurrence time. Perhaps, this was mainly due to the reason that to get data for Forward Recurrence time, one has to conduct a prospective study and the result will be similar to open birth interval. Thus, it was thought that there is no use of conducting a prospective study except for theoretical considerations.

One day, it came in our mind that if any family planning programme is launched after the time of a retrospective survey, the fertility parameter, viz., conception rate will be changed and the distribution of Forward Recurrence time will also change.

Keeping in view the above fact, a probability model was worked out for Forward Recurrence time. We called this interval as Forward Birth Interval. The expressions for the probability density function for forward birth interval along with its mean and variance were derived. The detailed derivations of the model are available in the above paper. For more details, one can see Yadava and Rai[36] also. The main advantage of this model is that with the help of this model, one can study the impact of a family planning programme by estimating the change in conception rate after launching of the family planning programme.

13 [L]. On the distribution of straddling birth interval

By R.C. Yadava and Arvind Pandey **in Biometrical Journal**[41].

We were aware about the difference between the usual closed birth interval and straddling birth interval. However, most of the researches were of theoretical nature, because for obtaining data on straddling birth interval, one has to conduct a retrospective survey as well as a prospective survey. Therefore, conducting surveys two times was considered to be a tedious job; that is why, only theoretical results were derived.

We have already mentioned about the conduct of Demographic Survey of Varanasi (Rural), 1969–1970. Fortunately after a gap of about 4–5 years, another revisit survey was also conducted. This gave us an opportunity to find out the data on straddling birth interval. With the availability of such data, we thought to develop a probability model for straddling birth interval and check its adequacy with the help of available data. The model was developed using certain reasonable and simplifying assumptions and applied to the real data and a satisfactory fit to the data was observed. The paper was sent for possible publication to Biometrical journal and was ultimately published in 1989.

14 [M] Closed birth interval versus most recent closed birth interval

By Yadava and Sharma in **Demography India**[45].

Sheps et al.[21] have studied extensively the truncation effect on closed and open birth intervals. Although, the results were extensive, but at that time, it was very difficult for us to understand various derivations of the paper.

Sheps and Menken[19] have also discussed extensively about the effect of sampling frame on birth intervals. Here too, the results were quite difficult to understand by us at that time.

In another paper, Sheps and Menken[20] have remarked that for a given age, the mean of most recent closed birth interval is somewhat larger than the means of other closed birth intervals. This statement was quite exciting for us. However, we could not understand: why the mean of most recent closed birth interval should be larger than the means of other closed birth intervals? We tried to understand the whole logic behind the statement, but could not succeed in finding any solution. Therefore, the problem remained unsolved for us for many years.

With the passage of time, we were also becoming more aware to understand the complex dynamics of birth intervals even including the concept of heterogeneity. The National Family Health Survey (NFHS) was providing extensive raw data on birth histories of all females who were in their reproductive age group at the time of survey. This gave ample opportunities for analyzing the birth interval data under different sampling frames. Therefore, we decided to examine the matter with the help of real data. A Ph.D. student under supervision of the author of this article was engaged in analyzing various types of birth interval data for his Ph.D. work. Author of this article suggested him to analyze the data on closed birth intervals for different orders for different marital duration groups. After obtaining appropriate tables, it was found that for almost all cases, the mean of most recent closed birth interval was larger than the means of other closed birth intervals. This gave empirical evidence that for given marital duration, mean of most recent closed birth interval is somewhat larger than the means of other closed birth intervals. (Note that marital duration and age are highly related.) Thus, we had no option than to accept the statement of Sheps and Menken[20]. Now, the issue became, how to prove it? There was no ready solution.

Then, we tried to examine the issue taking some simplified assumptions. For this, we first assumed that the two main fertility parameters, viz., conception rate and non-susceptible period are same for all females, i.e., the population is homogeneous with respect to these parameters with the further assumption that births are occurring uniformly over time. Under these assumptions, it was found that the distribution of most recent closed birth interval was the same as the distribution of other closed birth intervals. Obviously, their means will also be equal.

Thus, the matter remained unsolved even after using these simplifying assumptions. However, we observed even earlier that many times heterogeneity alters many results. Therefore, we changed the assumption and assumed that the non-susceptible period, h, takes two values say h_1 and h_2 with probabilities α and (1-α). We again obtained the distributions of most recent closed birth interval and other closed birth intervals.

Surprisingly, the two distributions were not same and the mean of most recent closed birth interval was larger than the mean of other intervals. This gave a solution to our long-awaited problem. More elaborate results and discussions are reported in Yadava and Rai[36].

However, after obtaining the above results heuristically, our emphasis now shifted to show the above result with a more theoretic approach. Since age at marriage varies from female to female, so it was thought proper to take fixed marital duration rather than age. Under some specifying assumptions, we attempted to derive the probability distributions of most recent and usual closed birth intervals. Finding the probability density functions of the two was quite difficult as it was found that the integral of the p.d.f. for whole range was not equal to one. Then, we were forced to rethink again and search for finding the cause, and several attempts were made and it took several months to get the appropriate expression. The expressions are quite complex and lengthy. However, these are given in Kumar and Yadava[10]. The reader is advised to see the above paper for better understanding.

Taking certain assumed values of m, h, and T, the authors have computed means for two types. It was found that the mean of most recent closed birth interval was larger than the means of other intervals, although the differences were varying for different situations.

15 [N]. On the distribution of menstruating interval

By Yadava, Pandey, and Tiwari in **Biodemography Social Biology[43].**

A research scholar under supervision of the author of this present article collected data on menstruating interval from a survey conducted in Lucknow, U.P. While analyzing the fertility data, we tried to examine whether the data on menstruating interval were following an exponential distribution? It was found that the fit was not good.

We were in search of an alternative way to analyze the data. In this context, we thought that many of the females try to delay their next pregnancy either using some contraceptive method or reducing coital frequency under various socio-cultural norms.

Therefore, we decided to classify the females into two categories, viz.

1. Who used some contraceptive method after their last birth.
2. Who did not use any contraceptive method after their last birth.

It was thought that perhaps the menstruating interval of first type of females may follow an exponential distribution, because after discontinuation of contraceptive method, their conception rates may remain constant.

The relevant data were analyzed and it was found that for such females exponential law was providing a satisfactory fit.

However, when the data for second type of females were analyzed, the exponential law (i.e., $f(x) = \lambda e^{-\lambda x}$) did not give a satisfactory fit. Therefore, there was no option other than to change the assumption. Ultimately, we assumed that for such females, the conception rate initially may be low which may increase slowly over time and remain constant after achieving a certain level. Ultimately, we assumed that the conception rate increases linearly up to a certain time and then remains constant. In such situation, finding the expression for p.d.f. of this random variable along with its mean was quite difficult. Ultimately, the expression was obtained, and when this model was applied, we got a satisfactory fit to the data.

For more details, one can see the above paper or Yadava and Rai[36].

16 [O] The distribution of consecutive closed birth intervals in females of Uttar Pradesh

By Yadava and Sharma **in Journal of Bioscience**[46]**.**

Large number of studies were being conducted for analyzing closed birth interval data of different orders. However, mostly data were analyzed considering each birth interval separately. It was thought that why not we make a study on joint distribution of two birth intervals. Perhaps, such studies were rare.

We were aware that while making study on truncation effect on closed and open birth intervals, Sheps et al.[21] obtained some very useful and interesting results. In their research, they assumed that consecutive birth intervals are independent, although may not be identical. In other studies too, for deriving models for number of births in an interval, consecutive intervals have been assumed to be independent.

Then, a problem came in our mind: Are in reality, the consecutive closed birth intervals independent? Of course, if the population is homogeneous with respect to conception rate and non-susceptible period, then consecutive closed birth intervals may be easily assumed to be independent.

However, a natural question arose. If the population is heterogeneous, will the consecutive intervals be independent?

For this, we considered a very simple case. Let all females have same conception rate but suppose that non-susceptible period is not same for all females and suppose that α proportion have non-susceptible period h_1, while (1-α) proportion have non-susceptible period h_2 ($h_1 < h_2$). Can, for this situation, consecutive birth intervals will be independent? We felt that the consecutive intervals should be positively correlated. This was mainly based on the following intuition.

“The females with smaller non-susceptible period are likely to have on an average smaller closed birth intervals, while the females with larger non-susceptible period will have on an average larger closed birth intervals”. Thus, their birth intervals should be positively correlated. This was totally based on our intuition.

One day, we consulted about the issue with a senior Professor of the department. He was also not very much clear about the matter. However, he said that the intervals should be independent. This created a confusion in our mind.

Therefore, we decided to examine the idea empirically. For this, we computed the correlation coefficients between consecutive closed birth intervals for NFHS-2 data and found a positive correlation. This increased our confidence for our thought. Then, a theoretical justification was needed. We obtained the theoretical expression for correlation coefficient making certain reasonable assumptions.

The derived expression is available in Yadava and Sharma[45]. More details are also available in Yadava and Rai[36].

Incidentally, it was also noticed that the PPA duration and menstruating interval were negatively correlated. The referee of the paper appreciated this result, because it was obtained with an indirect way without utilizing the direct data on the two variables.

In the above paper, based on certain assumptions, it has been shown that consecutive closed birth intervals are positively correlated based on certain heuristic assumptions. Later on, it was thought that the relationship between consecutive birth intervals should be studied in a more vigorous way. In the light of the above fact, Kumar and Yadava[37] while considering the impact of heterogeneity on closed and open birth intervals attempted to derive the theoretical expression for the correlation coefficient between the two consecutive closed birth intervals. For this, they assumed that if X and Y represent two consecutive birth intervals having pdf $\lambda e^{-\lambda(x-h)}$, $x > h, x > 0, \lambda > 0$, and $\lambda e^{-\lambda(y-h)}, y > h,\ \lambda > 0$ (here λ is interpreted as conception rate and h is the non-susceptible period associated with a birth). Under this condition, the two random variables X and Y will be independent; however, if the population is considered to be heterogeneous with respect to conception rate λ, then the situation of independence or dependence is to be investigated. Kumar and Yadava [11] assumed that if λ follows gamma distribution $g(\lambda) = \frac{b^a}{\Gamma(a)} \lambda^{(a-1)} e^{-\lambda b}$, $a > 0, b > 0, \lambda > 0$, then the expression for correlation coefficient between X and Y will be $\rho = \frac{1}{a}$, $a > 2$. This implies that the correlation coefficient between X and Y will be always positive as a is positive. This shows that the positive correlation comes only due to the impact of heterogeneity in the population.

17 [P] Estimation of parity progression ratios and instantaneous parity progression ratios from open and closed birth interval data

Under the above title, the works of various authors have been described. Therefore, we are not presenting the names of authors and Journals separately. Alternatively, we are describing each paper separately as given below:

Srinivasan[31,32] developed a methodology to estimate Instantaneous Parity Progression Ratios (IPPR) for different parities utilizing data on open and closed birth intervals. The basic equation for estimating ith order IPPR was

$$E(U_i) = \alpha_i \frac{E(T_i^2)}{2E(T_i)} + (1-\alpha_i)\frac{E(V_i^2)}{2E(V_i)},$$ where U_i is the open birth interval after ith birth, and T_i is the closed birth interval between ith and $(i+1)$ th births. V_i is the interval between last birth and end of reproductive period (say 45 years) for the females who have crossed their reproductive age for whom the ith birth happens to be their last birth.

The basic assumption of this procedure is that ith order births are uniformly distributed over time for females of age group (15–49) years in a retrospective survey.

Almost at the same time, while deriving results for "A parity-dependent model for number of births and its applications"[26], we considered the concept of Parity Progression Ratio (PPR) by defining it to be the conditional probability that a female after giving her ith birth will ever proceed to her next birth (i.e., $(i+1)$ th birth). Although, we did not give any name to this conditional probability, but we were sure that this is nothing but the PPR for ith birth.

What we are going to describe in the following lines is the confusion about the difference between IPPR and PPR. Of course, the confusion was on our part, not on the part of Srinivasan[31,32] applied his procedure to data collected in Gandhigram survey. He obtained the estimates of IPPR for different parities. He also applied his procedure to Fizzi data and obtained relevant estimates. The estimates were a bit lower than our expectation. At that time, we thought that the estimates of PPR (in our understanding) should be in the vicinity of 0.98–0.99 at least for the parities 1 and 2. This created doubt among us about the Srinivasan's procedure itself. Let us clearly mention here that in our thought, IPPR and PPR were the same. We guessed that since Srinivasan has applied entirely a new procedure hence, he has given the name IPPR for PPR (of course we were totally wrong). Another confusion also crept in our mind. We thought that mean open birth interval for fertile females should be half of the mean of closed birth interval. Initially Srinivasan was also confused about it. When he reported this statement in 1969 in Population Studies, almost just after it, Leridon[12] objected about this statement and suggested a rectification for it. Srinivasan readily accepted his modification.

Author of the present article got an opportunity to visit East West Population Institute, Honolulu, USA to attend a workshop for analyzing fertility data from sample surveys in 1984. Feeney was a faculty member at the East West Population Institute and he was also working on finding PPR from birth interval data (See Feeney[6]). When we discussed about Srinivasan's procedure for estimating IPPR, Feeney told that he was also not very clear about the Srinivasan's procedure, and hence, he cannot make any comment on this.

It is pertinent to mention here that the concept of Straddling birth interval was not clear to many of the researchers at that time. Even the problem of waiting time paradox was also not clear to many.

It seems desirable to specify here that although the difference between PPR and IPPR has not been explained in detail in the paper (at least for our understanding), Srinivasan has explained about it in his thesis.

At this stage, it is desirable to explain the difference between PPR and IPPR. To understand IPPR, let us assume that there are B_i females of parity i at the time of survey, i.e., they have given their ith birth sometime before the survey date and have not given their next birth till the survey date. Out of these B_i females, a proportion will give their next birth, i.e., $(i+1)$th birth sometime after the survey date. This proportion indicates the value of IPPR.

On the other hand, PPR can be understood as follows:

Suppose there are B_i^* females who give their ith birth at a given time, then the proportion of these B_i^*females who will ever proceed to give their next birth represents PPR. Thus, there is a clear-cut difference between the two concepts. It is difficult to say that which one is better; both have their advantages and limitations. We are excited to mention here that after a long time we could establish a relationship between the two which we will explain later.

Let us come back again on our original problem. We were observing relatively lower estimates of PPR (in our understanding, although the estimates were of IPPR) in the context of Gandhigram data as well as Fizzi data. We were suspicious about data on V_i, because these related to events occurring quite long back resulting in possibility of some memory bias.

We already were having data of Demographic Survey of Varanasi (Rural) 1969–70. With the passage of time, another survey of about 3500 rural households was also conducted in 1978 under a research project entitled "Evaluation of impact of development activities and fertility regulation programs on population growth rate in rural areas" sponsored by University Grants Commission (UGC), New Delhi.

As mentioned above, we were suspicious about data on V_i. Alternatively, we thought to apply the Srinivasan procedure on our data too. To our surprise, we got similar results on our data also. Therefore, the issue remained unsolved for more than a decade. We were in search of some alternative way. One day, it came to our mind that we are using the relation $\int_0^\infty [1 - F_i(t)]dt = E(T_i)$, where $F_i(t)$ is the distribution function of T_i. It is also true that the value of $[1 - F_i(t)]$ will become zero for larger values of t (say 10 years or more). Therefore, instead of considering the relation

$$\int_0^\infty [1 - F_i(t)]dt = E(T_i),$$

we decided to use the relation

$$\int_0^\infty [1 - F_i(t)]dt = \int_0^C [1 - F_i(t)]dt,$$

where C is such that $P[T_i > C]$ is almost zero, i.e., $[1 - F_i(t)] = 0$ for all $t > C$.

By doing such alternation, we may get rid of V_i as well as of $\frac{E[V_i^2]}{2E[V_i]}$.

Therefore, with this alteration, we decided to modify the Srinivasan procedure. In this context, instead of including all females of parity i at the time of survey, we considered only those females whose open birth interval was less than C. Obviously, in this case also, the females will be of two types, viz., fertile and sterile. For fertile females, mean open birth interval will be

$$\frac{\int_0^C t[1 - F_i(t)]dt}{E(T_i)} = \frac{E(T_i^2)}{2E(T_i)},$$

which is the same as considered by Srinivasan. However, the mean open birth interval for sterile female will be $\frac{C}{2}$, because it has been assumed that births are uniformly distributed overtime. We thought that we have found a solution to the problem by considering the equation

$$E(U_i^*) = \alpha_i \frac{E(T_i^2)}{2E(T_i)} + (1 - \alpha_i)\frac{c}{2},$$

where α_i is the PPR for ith parity, and U_i^* is the open birth interval of females included in the study. Obviously, from the above equation, α_i can be easily computed. We were almost to do relevant computations, but it was suddenly noticed that the estimate of α_i will change if we increase the value of C, say from 10 to 12 years.

In this case, $\frac{E(T_i^2)}{2E(T_i)}$ will remain the same while value of $\frac{C}{2}$ will increase by 1 year. Thus, estimate of α_i will change. However, it is a theory requirement that the value of α_i should not change whatever be the value of C. Thus, we faced a problem again. What to do so that α_i does not change with changing value of C?

Time passed. After a few months, we noticed that by increasing the value of C, number of sterile females in the study will increase, while the number of fertile females will remain the same. This gave a clue that the relationship

$$E(U_i^*) = \alpha_i \frac{E(T_i^2)}{2E(T_i)} + (1 - \alpha_i)\frac{C}{2}$$

must be changed as

$$E(U_i^*) = \alpha_i^* \frac{E(T_i^2)}{2E(T_i)} + \left(1 - \alpha_i^*\right)\frac{C}{2},$$

where α_i^* is the proportion of fertile females included in our study, while $(1-\alpha_i^*)$ is the proportion of sterile female included in the study.

Obviously from the above relation, α_i can be easily computed using the relationship

$$\alpha_i^* = \frac{\alpha_i E(T_i)}{\alpha_i E(T_i) + (1-\alpha_i)C}.$$

Thus, a problem for which search was being made from around 1975 was ultimately solved in 1985 (after a gap of around 1 decade). For more details, see Yadava and Bhattacharya[38] in their paper "Estimation of Parity Progression Ratios from Closed and Open Birth Interval Data" mimeograph, Centre of Population Studies, Banaras Hindu University. For more details, see also Bhattacharya (1984) and Yadava[37].

We became relaxed after solving the above problem. Incidentally, after a gap of about 5 years, it suddenly came to our mind that α_i^* is very much similar to IPPR. It gives the proportion of fertile females out of females under study. It prompted us to believe that we have obtained the expression for inter relationship between IPPR and PPR and vice versa. A paper was prepared and published by Yadava and Saxena[44] entitled "On the Estimation of Parity Progression and Instantaneous Parity Progression Ratios" in the book Population Transition in India edited by Singh, Premi, Bhatia, and Ashish Bose.

After obtaining the above results, a thought came in our mind: Can we estimate α_i by taking even smaller value of C say 4 or 5 years? We tried to solve this problem.

However, for solving this problem, we were required to find the value of $\int_0^c [1 - F_i(t)]dt$ for $C=5$ years (say). This was an incomplete integral. We were not able to find the value of above integral as no explicit form of $\{1 - F_i(t)\}$ was available to us. Therefore, the problem remained unsolved for many years.

One day, author of the present article was taking practical class of B.Sc. students in the department in which a problem related to quadrature formulae was given to the students. The given problem was to find the value of $\int_0^3 e^{-x}dx$ using

1. Trapezoidal rule
2. Simpson's (1/3)rd rule
3. Simpson's (3/8)th rule and
4. Weddle's rule, and compare these values with the actual value. This was a routine exercise in the class. However, when I was examining the answer books of some students, it just clicked in mind that value of integral can be easily found by finding the value of integrand e^{-x} at $x=0$, 0.5, 1.0, 1.5, 2.0, 2.5 and 3.0. Therefore, $\int_0^c [1 - F_i(t)]dt$ can also be easily computed by finding the values of {1-Fi(t)} at some selected points of *t*' with the use of suitable quadrature formula.

Thus, the problem was almost solved. We computed these values and found the estimates of α_i for different parties. As a check, the values of α_i were calculated under the consideration of different values of C say 5,6,7,8, 9, 10 years. For all the cases, the estimate of α_i remained almost constant which was a theory requirement. Finally, a paper was prepared and published in the form of Yadava et al.[42] entitled "Estimation of Parity Progression Ratios from the Truncated Distribution of Closed and Open Birth Intervals". Thus, a problem, which could have been easily solved using some quadrature formula, remained unsolved for about 2–3 years as our mind was never attracted towards this simple solution.

This methodology was providing reasonably reliable estimates of PPR. One of research scholar of the department utilized the data of NFHS-2 to find the values of PPR for different parities. He presented his paper at an annual conference of Indian Association for the Study of Population (IASP) under poster presentation. His paper was adjudged as the best paper under poster presentation. However, when he applied this procedure for NFHS-3 data, for some states, the estimates of PPR were quite unreasonable. This again created a curiosity among us: Why our procedure is not working well for NFHS-3 data especially for southern states like Andhra Pradesh, Tamil Nadu, etc.? No ready solution was available before us. One day, the idea clicked that our procedure is essentially based on the assumption that *i*th order births are uniformly distributed over time. However, fertility of southern states, especially Andhra Pradesh and Tamil Nadu, was declining very fast at that time. Therefore, we thought that, perhaps, the assumption may not be true for some of the southern states. When the data were analyzed according to year of birth, it was found that the assumption of uniform births over time was not true. Births in recent years were less than the births in the previous years. This gave a clue that perhaps we are not getting reasonable estimates of PPR simply because of violation of the assumption. This puts again a new challenge before us. However, due to available data on births year-wise, it was possible for us to compute the value of $[1 - F_i(t)]$ for different values of t. Consequently, it became easy for us to compute α_i for different parities. A paper was prepared and published in the form of[39,39] entitled "Estimation of Parity Progression Ratios from Open and Closed Birth Interval Data".

18 [Q] Extent of infecundity derived from open birth interval data

Yadava and Srivastva **in Demography India**[48].

We have mentioned earlier that α_i^* gives the proportion of fertile females among the females with open interval (0, C). Thus, one can easily compute the proportion of fertile (or sterile) females for whom open birth interval is in (0, C_1) or (0, C_2) or any other group say (0, C^*). Once we know these proportions, the proportion of fertile (or sterile) females, i.e., α_{i}^* can be easily computed for females having open birth interval between C_1 and C_2. Keeping in view the above facts, a paper was prepared and published as Yadava and Srivastava[48] entitled "Extent Of Infecundity Derived from Open Birth Interval Data". The other researchers too have made significant contributions in the field. However, a special mention of Dr. B.N. Bhattacharya seems desirable.

It is worthwhile to mention here that the discussion on the papers listed here does not present an exhaustive list of papers related to stochastic modeling of human fertility at the Department of Statistics, Banaras Hindu University. In fact, Dr. B. N. Bhattacharya who was earlier associated with department of statistics; BHU left the department in 1975 but again came back in 1979. He remained here for about 8 years. During this period, he developed his new team of researchers and published many valuable papers on the topic. During the above period, D.C. Nath (Former Vice-Chancellor, Asam Central University, Silchar), C. M. Pandey (Former head of the department of Biostatistics and Health Informatics, S.G.P.G.I.M.S., Lucknow), and K. K. Singh (Senior Professor, Dept. of Statistics, and Dean Students Welfare) obtained their Ph.D. degrees under the supervision of Dr. B. N. Bhattacharya.

These include huge work on stochastic modeling of human fertility.

We are satisfied to mention that the seed which was sown about 52 years back (1969) has grown consistently over time and has now taken the shape of big tree with its strong branches. It is hope that in future also new relay team of researchers will further strengthen it with more vigour and enthusiasm.

Finally, it is to be said that I have given the relevant stories based on my personal experience. For some readers, these may be very interesting, for some these may be making them curious, while for others, these may be worthless or boring. Whatever be the situation, I can only say, especially to young researchers, that continuous effort is needed in the research career. We should not get discouraged from our failures. Continuous effort may yield some fruitful result on some day. This all depends on chance, the essence of life.

Publisher's Note

Acknowledgements

The author is extremely thankful to Prof. Piyush Kant Rai, Professor, Department of Statistics for his valuable help and suggestions in the preparation of this article.

Declarations

Conflict of interest

There is no any financial or personal relationship with a third party whose interests could be positively or negatively influenced by this present article's contents. And this article is written by me as a single author; therefore, I (R.C. Yadava) declare that there is no any conflict of interest in writing this article.

Received: 24 April 2022 Accepted: 5 September 2022
Published online: 26 October 2022

References

1. Bhattacharya BN (1970) On some stochastic models for couple fertility. An Unpublished Thesis Submitted in Banaras Hindu University, Varanasi
2. Brass W (1958) The distribution of births in human population. Popul Stud 12:51–72
3. Chiang CL (1968) Introduction to stochastic processes in biostatistics. John Wiley and Sons
4. Cox DR, Miller HD (1965) The theory of stochastic processes. Methuen and Co.Ltd., London
5. Dandekar VM (1955) Certain modified forms of binomial and poisson distributions Sankhya. Indian J Stat 15(3):237–250

6. Feeney G, Ross JA (1984) Analyzing open birth interval distributions. Popul Stud 38:473–478
7. Feller W (1966) An introduction to probability theory and its applications. Oliver and Boyd Ltd
8. Henry L (1953) Fondements theoriques des mesures de la fecondite naturelle. Rev Inst Int Stat 2:135–151
9. Keyfitz N (1969) Introduction to mathematics of population. Addison-Wesley, Reading
10. Kumar A, Yadava RC (2015) Usual closed birth interval versus most recent closed birth interval. J Data Sci 13(1):73–94
11. Kumar A, Yadava RC (2018) Impact of heterogeneity on closed and open birth intervals. Sankhya B Indian J Stat 2:1
12. Leridon H (1969) Some comments on an article by K Srinivasan. Popul Stud 23:102
13. NFHS-2 (1998) National family health survey-2 Report, India
14. NFHS-3 (2005) National family health survey-2 Report, India
15. Pandey A, Dwivedi SN, Mishra RN (1990) Distribution of closed birth interval with some biosocial components: a stochastic model and its applications. J Math Sociol 16:89–106
16. Pandey A, Singh KK, Singh A (2016) Modelling the proportion of male births under sex preferred stopping rules. Janasamkhya 34:1–8
17. Pathak KB (1971) A stochastic model for the study of open birth interval—a cohort approach. Sankhya B 33:305–314
18. Sharma SS (2004) A study on some mathematical models for birth intervals. A Ph.D. Thesis, Banaras Hindu University
19. Sheps MC, Menken JA (1972) Distribution of birth intervals according to the sampling frame. Theor Popul Biol 3:1–26
20. Sheps MC, Menken JA (1973) Mathematical models of conception and birth. University Of Chicago Press, Chicago
21. Sheps MC, Menken JA, Ridley JC, Linger JW (1970) Truncation effect in closed and open birth interval data. J Am Stat Assoc 65:678–693
22. Shukla AK (2006) A study on mathematical models for population dynamics. An Unpublished Thesis Submitted in Banaras Hindu University, Varanasi
23. Singh SN (1963) Probability models for the variation in the number of births per couple. J Am Stat Assoc 58(303):721–727
24. Singh SN (1968) A chance mechanism of variation in the number of births per couple. J Am Stat Assoc 63(321):209–213
25. Singh SN, Bhattacharya BN (1970) A generalized probability distribution for couple fertility. Biometrics 26:33–40
26. Singh SN, Bhattacharya BN, Yadava RC (1974) A parity dependent model for number of births and its applications. Sankhya B Indian J Stat 36(1):93–102
27. Singh SN, Yadava RC (1977) A generalized probability model for an equilibrium birth process. Demogr India 6(1–2):163–173
28. Singh SN, Yadava RC, Pandey A (1987) On estimating the mean birth interval characteristic of women. Can Stud Popul 14(1):1–8
29. Singh SN, Yadava RC, Pandey A (1978) A probability model for forward birth interval. Health Popul Perspect Issues 1(4):295–301
30. Singh SN, Yadava RC, Yadava KNS (1981) A study on the fertility of migrants. Health Popul Perspect Issues 4:959–965
31. Srinivasan K (1967) Analytical models for two types of birth intervals with applications toindian population. A Ph.D. Thesis, University of Kerala, Trivandrum, India
32. Srinivasan K (1967) A probability model applicable to the study of inter-live birth intervals and random segments of the same. Popul Stud 21:63–70
33. Srinivasan K (1968) A set of analytical models for the study of open birth intervals. Demography 5(1):34–44
34. Srinivasan K (1980) Birth interval analysis in fertility surveys: an illustrative application to datafrom fiji fertility survey. Scientific Report No. 7. International Statistical Institute, Netherlands
35. Wolfers D (1968) Determinants of birth intervals and their means. Popul Stud 22:253–262
36. Yadava RC, Rai PK (2019) Analyzing variety of birth intervals: a stochastic approach. In: Srinivasa Rao AS, Rao CR (eds) Integrated population biology and modeling Part—B In the handbook of statistics, vol 40. Elsevier, Amsterdam
37. Yadava, R.C., 2018. Stochastic Modelling of Some Natural Phenomena: A Special Reference to Human Fertility. In: Srinivasa Rao, A.S.R., Rao, C.R. (Eds.), Handbook of Statistics:Integrated Population Biology and Modelling Part A, Vol. 39, Pp. 187–274.
38. Yadava RC, Bhattacharya M (1985) Estimation of parity progression ratios from closed andopen birth interval data. Centre of Population Studies, Banaras Hindu University, Varanasi, India, Mimeo
39. Yadava RC, Kumar A, Pratap M (2013) Estimation of parity progression ratios from open and closed birth interval data. J Data Sci 11(3):607–621
40. Yadava RC, Kumar A, Srivastava U (2013) Sex ratio at birth: a model based approach. Math Soc Sci 65(1):36–39
41. Yadava RC, Pandey A (1989) On the distribution of straddling birth intervals. Biom J 31(7):855–863
42. Yadava RC, Pandey A, Saxena NC (1992) Estimation of parity progression ratios from thetruncated distribution of closed and open birth intervals. Math Biosci 110(2):181–190

43. Yadava RC, Pandey R, Tiwari AK (2009) On the distribution of menstruating interval. Biodemogr Soc Biol 55(1):1–11
44. Yadava RC, Saxena NC (1989) On the estimation of parity progression and instantaneous parity progression ratios. In: Singh SN, Premi MK, Bhatia PS, Ashish B (eds) Population transition in India, vol 1. BR Publishing, Delhi, pp 357–362
45. Yadava RC, Sharma SS (2004) Closed birth intervals versus most recent closed birth interval. Demogr India 33(2):249–263
46. Yadava RC, Sharma SS (2007) The distribution of consecutive closed birth intervals in females in Uttar Pradesh. J Biosci 39:185–199
47. Yadava RC, Srivastava M (1993) On the distribution of births over time in an equilibrium birth process for a female giving specified number of children in a given period. Demogr India 22(2):241–246
48. Yadava RC, Srivastava M (1998) Extent of infecundity derived from open birth interval data. Demogr India 27(1):205–211
49. Yadava RC, Tiwari AK (2007) On the pattern of births over time. Demogr India J Indian Assoc Study Popul 36(2):287–302
50. Yadava RC, Tiwari AK (2010) Symmetry in equilibrium birth process: a mathematical approach. Demogr India 39(2):287–296

Prof. R.C. Yadava born on July 6, 1946, in Azamgarh District of Uttar Pradesh, India, obtained his B.Sc., M.Sc. and Ph.D. degrees from Banaras Hindu University. Prof. Yadava started his teaching and research career from his parent institution and retired as Professor of Statistics. He has more than 50 years of teaching experience at Under-graduate and Post-graduate levels. Nearly 20 students have obtained their Ph.D. degrees under his supervision. During his academic career, he has published more than 100 research articles in National and International Journals of repute. Some of them are Journal of American Statistical Association (JASA), Canadian Studies in Population, Journal of Biosocial Science, Journal of Data Science, Sankhya, and Demography India. During his tenure at B.H.U., he was head of the Department of Statistics, Dean Faculty of Science, Chief Proctor and many other administrative posts in B.H.U. He has also conducted many research projects as Principal Investigator/Co-investigator sponsored by Indian Council of Medical Research, New Delhi, University Grants Commission, Indian Council of Social Science Research, New Delhi. The major areas of research of Prof. Yadava are Stochastic Modeling of Human Fertility, Effects of Socio-cultural factors on Fertility and conduct of Large-Scale Sample Surveys.