

The Effects of Natural Selection and Mutation on
Inbreeding Effective Size

Can Ceník

under the direction of
Fangyun Yang
Department of Mathematics
Massachusetts Institute of Technology

Research Science Institute
July 31, 2007

Abstract

In this paper Ryman's effective size model is extended by integrating basic natural selection and mutation models into it. In addition a novel natural selection model is developed for a sequence of generations that has fluctuating numbers of individuals in each generation. A modified Poisson distribution is used to model different productivities of individuals and therefore natural selection. Finally, we review the related literature to propose future work on the subject.

1 Introduction

When genes are randomly assorted from one generation to the next in finite populations, genetic diversity is lost. This phenomenon is known as genetic drift. For selectively neutral genes, the dynamics of this process depend on the effective population size. The effective size for a given genetic parameter is defined as the size of an ideal population that will show the same genetic behaviour (Wright [6]). Depending on the parameter considered, different effective sizes, such as variance effective size, eigenvalue effective size, and inbreeding effective size can be derived. This paper focuses on inbreeding effective population size (Wright [1]).

Genetic polymorphism within a population is a prerequisite for adaptive change, so it is crucial to understand the factors that promote the retention of genetic variability. The importance of preserving levels of genetic variation within populations has long been recognized for influencing short-term population viability as well as for maintaining a population's adaptability to environmental changes (Frankham [14]). In an isolated population, mutation rate and effective population size determine the level of genetic variation (Nunney [11]). Inbreeding causes a loss in the number of heterozygotes and in genetic diversity. For this reason inbreeding effective size is an important issue in conservation biology. A better understanding of the interaction between natural selection, mutation, gene flow (Hoole et al. [19], Tremblay et al. [20]) and effective population size will help us to design better conservation strategies (Lewandowski et al. [21]) for organisms ranging from long-lived trees of rainforests (Shapcott [13]) to rare cattle breeds of commercial interest (Jordana [16]). Even the controversial issues of conservation strategies like consequences of lineage mixing (Berger [18]), effects of habitat alteration and fragmentation (Scribner [17]) can be understood better when we have more knowledge about effective population size.

Effective size estimations in discrete and overlapping generations are very different. The Wright-Fisher model is derived for non-overlapping generations, whereas Pollak [22, 23] and

Wang et al. [24] suggested models for overlapping generations. Ryman et al. [3] derived a model that can be used for organisms having a combination of discrete and overlapping models. This model can be used for a time-series only, but it generalizes the harmonic and arithmetic model under different demographic assumptions. This paper uses the same approach as Ryman's model.

A great deal of work has been done in order to incorporate natural selection into effective size theories(Caballero [5]). Gillespie [8] , Araki et al. [7] and Tachida [9] used effective size models in order to test the neutral theory of molecular evolution, and Burrows(1984) [4] used phenotypic selection and optimum-index selection models to understand the effects of inbreeding on unrelated families. All of this effort was based on the Wright-Fisher allelic model.

In this paper, the definition of effective size suggested by Ryman's [3] model is used as a basis for the incorporation of natural selection and mutation to develop a more complete model. Ryman et al. [3] defined effective size as the inverse of the probability that two homologous genes in two randomly chosen individuals of one generation were derived from the same individual in the previous generation. This model takes into account different productivities in different reproduction periods in a single generation. This paper uses two different approaches to model natural selection. First, a simple single locus natural selection model is used to investigate the effects of mutation-selection equilibrium on Ryman's model, and equations are derived for calculating effective size after these processes. Second, a modified Poisson distribution is used to model different productivities of individuals, and novel equations modeling natural selection are derived.

2 Theory

2.1 Extension of Ryman's Model

First we consider a monoecious haploid organism. We assume that generations are non-overlapping and population size is stable over generations. In this situation let p be the frequency of the dominant form of the allele. There are N groups, and N_i parents in i th group. x_i is the probability of an individual coming from the i th group, and P is the probability that two homologous genes in two randomly chosen individuals of one generation were derived from the same individual in the previous generation. In this case, P may be expressed as

$$P = \sum_{i=1}^N p^3 \frac{x_i^2}{N_i} + \sum_{i=1}^N (1-p)^3 \frac{x_i^2}{N_i} \quad (1)$$

From this equation ¹ it is possible to define the effective population size N_e through

$$\frac{1}{N_e} \equiv (p^3 + (1-p)^3) \sum_{i=1}^N \frac{x_i^2}{N_i} \quad (3)$$

Next the one-locus natural selection model was incorporated to the haploid model. In order to do this, W_p and $W_{(1-p)}$ are defined as the relative fitnesses of alleles p and $(1-p)$. Relative fitnesses are normalized weight factors of reproductive success associated with each allele. This yields

¹It is also possible to derive a similar equation for randomly mating diploid organisms. Let k be an arbitrary positive integer such that each individual produce gametes proportional to k . We assumed that there are equal number of females and males. Letting N denote the number females and males, because each individual has two copies of the same gene i , it can be stated that each gender produce $2nk$ gametes. All individuals in the next generation will take half of their genes from the mother and half from the father. For simplicity, it was assumed $k = 1$, making the probability of two randomly chosen individual coming from the same parent

$$P = 2 \left[\frac{1}{4} p^2 \left(\frac{1}{2Np} \right)^2 + \frac{1}{4} (1-p)^2 \left(\frac{1}{2N(1-p)} \right)^2 \right] \quad (2)$$

$$\begin{aligned}
P &= \sum_{i=1}^N \left(\frac{pW_p}{pW_p + (1-p)W_{(1-p)}} \right)^2 x_i^2 p^2 \sum_{i=1}^{N_i} \left(\frac{1}{pN_i} \right)^2 \\
&+ \sum_{i=1}^N \left(\frac{(1-p)W_{(1-p)}}{pW_p + (1-p)W_{(1-p)}} \right)^2 x_i^2 (1-p)^2 \sum_{i=1}^{N_i} \left(\frac{1}{(1-p)N_i} \right)^2 \quad (4)
\end{aligned}$$

It is possible to consider the effects of mutation on Ryman's effective size model in a some manner. Letting u be the rate of forward mutation and v be the rate of reverse mutation, we would have

$$\begin{aligned}
P &= \sum_{i=1}^N \left(\frac{(p + v(1-p) - up) - v(1-p)}{(p + v(1-p) - up)} \right)^2 x_i^2 p \sum_{i=1}^{N_i} \left(\frac{1}{pN_i} \right)^2 \\
&+ \sum_{i=1}^N \left(\frac{((1-p) + up - v(1-p)) - up}{(1-p) + up - v(1-p)} \right)^2 x_i^2 (1-p) \sum_{i=1}^{N_i} \left(\frac{1}{(1-p)N_i} \right)^2 \quad (5)
\end{aligned}$$

Equations (4) and (5) can be straightforwardly joined together to obtain a model that takes both mutation-selection equilibrium and inbreeding effective size into account.

All of the above equations are concerned with two consecutive generations. The next step is to generalize these models for an infinite number of generations. We assume that in every generation there are k groups and the progeny of the s th group in the t th generation constitute the s th group in the $(t+1)$ th generation. Each group has either N_1 or N_2 individuals in any given generation, and X_s is the number of individuals in the s th group. Therefore a two-valued stationary Markov chain (See Appendix B) is defined such that its transition functions are:

$$P(X_s(t+1) = X_s(t) | X_s(t) = N_j) = 1 - q_{ij} \quad (6)$$

$$P(X_s(t+1) \neq X_s(t) | X_s(t) = N_j) = q_{ij}, \quad (7)$$

and

$$P(N_s(0) = N_j) = p_{sj}^{(0)} \quad (8)$$

$s \in \{1, 2, \dots, k\}$ and $j \in \{1, 2\}$ where $0 < q_{ij} < 1$

$$(p_{s1}^{(0)}, p_{s2}^{(0)}) = \left(\frac{q_{s2}}{q_{s1} + q_{s2}}, \frac{q_{s1}}{q_{s1} + q_{s2}} \right) \quad (9)$$

is the stationary distribution of the Markov chain. (See Appendix A) E is the expectation calculated over the stationary distribution of the Markov chain. Letting $\pi_2(t)$ be the probability that two randomly chosen individuals from the population in generation t have the same parent in the initial state (Ewens) [12], we obtain

$$1 - \pi_2(t) = E \left[\prod_{g=0}^{t-1} \left(1 - \sum_{s=1}^k \frac{N_s(g)}{\sum_{s=1}^k N_s(g)} \frac{1}{N_s(g-1)} \right) \right] (1 - \pi_2(0)) \quad (10)$$

The inbreeding effective size can now be defined by

$$1 - \frac{1}{N_e} = \lim_{t \rightarrow \infty} \left(E \left[\prod_{g=0}^{t-1} \left(1 - \sum_{s=1}^k \frac{N_s(g)}{\sum_{s=1}^k N_s(g)} \frac{1}{N_s(g-1)} \right) \right] \right)^{1/t} \quad (11)$$

We have therefore extended Ryman's effective population size model to include the effects of natural selection and mutation. The first equations ((1) to (5)) take into account the allelic frequencies and some basic mutation and selection models for haploid organisms, but they are derived for just two generations. This model is different from the classical Wright-Fisher model because it allows different productivities within a generation so this is very useful in predicting the effective population sizes of semelparous organisms like salmon (Waples [2]) . Since the model for infinitely many generations (equations (6) to (11)) is a two-valued stationary Markov-chain, computer simulations of these equations can be used to estimate the possible consequences of different productivities on the genetic behaviour of different

populations. This will help us develop better conservation strategies when the fluctuations due to different productivities are important.

2.2 A New Natural Selection Model

In this next section we suggest a new natural selection model and integrate it into Ryman's effective size model through creation of diversity in productivities within a generation. This diversity will result in a fluctuation of population size (Iizuka [10]) in time, and we first consider this effect. In this model, population size can be any non-negative integer and the productivity within the generation is uniform and is Poisson distributed with parameter λ . Multiplication of productivity and population size at $(t-1)$ th generation gives the population size in t th generation. $N_1 = cN_0$ and

$$P(N_1 = cN_0) = x_i \tag{12}$$

$$P(N_2 = mN_0) = \sum_{j_1 j_2 = m} P(N_2 = j_1 N_1 | N_1 = j_2 N_0) \tag{13}$$

As these two events are independent it is possible to derive

$$\begin{aligned} \sum_{j_1 j_2 = m} P(N_2 = j_1 N_1) P(N_1 = j_2 N_0) \\ = \sum_{j_1 j_2 = m} x_{j_1} x_{j_2} \end{aligned} \tag{14}$$

It is also possible to obtain the following by induction:

$$\begin{aligned} P(N_t = mN_0) &= P(N_t = j_1 N_{t-1}) \dots P(N_1 = j_k N_0) \\ &= \sum_{j_1 \dots j_k = m} x_{j_1} x_{j_2} \dots x_{j_k} \end{aligned} \tag{15}$$

The next step is considering different productivities within a generation. A new random variable G_i which is a sub-group generated by the i th individual in the 0 th generation is defined. The probability that $G_i = n$ is $\frac{\lambda^n e^{-\lambda}}{n!}$. So the probability that two randomly chosen individuals come from the same parent in the previous generation is

$$P = E \left(\frac{\sum_{i=1}^{N_0} G_i^2}{(G_1 + G_2 + \dots + G_{N_0})^2} \right) \quad (16)$$

After two generations every sub-group in G_1 generates new sub-groups of random variables. Theoretically, after this kind of branching it is possible to sum all probabilities and find the same expectation. To make the model more realistic it is possible to define an upper bound M [15] such that it is not possible for any single individual to have more progeny than that number. With this upper bound set; $\sum_{x=0}^M \frac{\lambda^n e^{-\lambda}}{n!}$ will be smaller than one. To normalize the distribution every single term from 0 to M can be divided by $\sum_{x=0}^M \frac{\lambda^n e^{-\lambda}}{n!}$.

Equations (12) to (14) suggest a novel model of natural selection. The approach is the same as Ryman et al. [3], but it is suggested that every individual can be treated as a sub-group having a specific probability distribution of having different number of progeny, instead of uncertain different productivities within a generation. The model proposed is based on differences in each individual's reproduction rate. This takes every individual as one phenotype that natural selection acts on and does not deal with any allelic frequencies. Our model assumes that every individual has some probability of having 0 to M number of progeny. The probabilities for having different numbers of progeny are distributed as a modified Poisson distribution. Our model is completely governed by a stochastic selection process. It is also possible to find the effective population size (equation (15)) after stochastic selection, but calculations become very complicated when generation numbers increase.

3 Discussion

The genetic structure of populations change according to specific factors such as mutation, migration, natural selection, inbreeding, random sampling of alleles due to finite number of individuals and non-random mating. In order to understand the course and mechanisms of evolution an understanding of the dynamics of population genetics is essential. All existing models consider a relatively small portion of the factors mentioned above.

Both models proposed in this paper are extensions of Ryman's effective size model. After the incorporation of single locus natural selection, and basic mutation models, a new stochastic natural selection model based on individuals' difference in reproductivity was suggested, and equation (15) is derived for calculating effective size after this process. This model makes it possible to develop better conservation strategies for organisms that have the characteristics of both overlapping and discrete generation models. In a broader sense, this model is a more complete attempt to understand the genetic structure of populations and therefore the course of evolution.

Effective population size is a useful way to study different factors acting simultaneously on a population, because it considers the divergence times from the ancestral genes and allows us to study the effects of various factors on this process. The approach used in this paper is fixing a starting generation and figuring out the probabilities of divergence from the same individual in this generation. It is also possible to investigate the distributions of times to coalescence as in Polanski et al. [25] without fixing a parental generation. However Polanski et al.'s paper does not take natural selection and mutation into account.

The models in equations 1 to 5 are for autosomal characters that are determined by single genes. The model can be extended for polygenic characters and sex-linked loci by similar approaches used in Burger [26] and Nomura [27] respectively.

Our model is limited in the sense that it does not take some factors into account. It is

a future problem to investigate the properties of our effective size model (equations (12) to (15)) with the effects of stable mutation rate (Ishii et al. [28]), the effects of age-structure (Pollak [29]), spatial subdivision (Wakeley [31]), and/or non-genetic inheritance of fertility (Sibert et al. [30]).

4 Acknowledgments

I would like to thank my mentor Fangyun Yang for her help throughout this program, and Prof. Hartley Rogers for making this mentorship possible. I'm especially grateful to Jeremy England for his patience, and excellent suggestions, and Matt Paschke for his ongoing assistance. I would also like to thank all the other staff members, especially my tutor, Sapan Shah, for helping me organize my ideas and for reading over many drafts of my paper. I am grateful for the support of the Center for Excellence in Education and the Research Science Institute.

References

- [1] Wright, S. : Evolution and the Genetics of Populations. Vol.2 The Theory of Gene Frequencies. *The University of Chicago Press*, 1969: 11(C).
- [2] Waples, R. S.:Effective Size of Fluctuating Salmon Populations *Genetics*,2002: 161, 783-791 .
- [3] Ryman, N. & Laikre, L.: Effects of Supportive Breeding on the Genetically Effective Population Size *Conservation Biology*, 1991: 5, 325-329.
- [4] Burrows, P. M.: Inbreeding under Selection from Unrelated Families *Biometrics*, 1984: 40, 357-366.
- [5] Santiago, E. & Caballero, A.: Effective Size of Populations Under Selection *Genetics*, 1995: 139, 1013-1030.
- [6] Wright, S. :Evolution in mendelian populations. *Genetics*, 1931: 16, 97-159.
- [7] Araki, A. & Tachida, H.: Bottleneck effect on evolutionary rate in the nearly neutral mutation model *Genetics*, 1997: 147, 907-914.
- [8] Gillespie, J. H.: The Neutral Theory in an infinite population *Gene*, 2000: 261, 11-18
- [9] Tachida, H.: DNA evolution under weak selection *Gene*, 2000: 261, 3-9
- [10] Iizuka, M. :The effective size of fluctuating populations *Theoretical Population Biology*, 2001: 59, 281-286
- [11] Nunney, L. :The effective size of annual plant populations: The interaction of aseed bank with fluctuating population size in maintaining genetic variation *The American Naturalist*, 2002: 160, 195-204
- [12] Ewens, W. :”Mathematical Population Genetics” it Springer-Verlag, New York 1979
- [13] Shapcott, A. : Population genetics of the long-lived Huon pine *Lagarostrobos franklinii*: An endemic Tasmanian temperate rainforest tree *Biological Conservation*, 1997: 80, 169-179
- [14] Frankham, R. : Inbreeding and extinction: a treshold effect *Conservation Biology*, 1995: 9, 792-799
- [15] Watson, H. W. & Galton, F. : On the probability of the extinction of families *The Journal of the Anthropological Institute of Great Britain and Ireland*, 1875: 4, 138-144
- [16] Jordana, J. & Piedrafita, J. : The “*Bruna dels Pirineus*” (Pyrenean brown breed) : a genetic study of a rare cattle breed in Catalonia (Spain)

- [17] Scribner, K.T. & Arntzen, J.W. & Cruddace, N. & Oldham, R.S. & Burke, T. : Environmental correlates of toad abundance and population genetic diversity *Biological Conservation*, 2001: 98, 201-210
- [18] Berger, J. & Cunningham, C. : Multiple bottlenecks, allopatric lineages and badlands bison *Bos bison* : Consequences of lineage mixing *Biological Conservation*, 1995: 71, 13-23
- [19] Hoole, J.C. & Joyce, D.A. & Pullin, A.S : Estimates of gene flow between populations of the swallowtail butterfly, *Papilio machaon* in Broadland, UK and implications for conservation *Biological Conservation*, 1999: 89, 293-299
- [20] Tremblay, R.L. & Ackerman, J.D. : Gene flow and effective population size in *Lepanthes* (Orchidaceae): a case for genetic drift *biological Journal of the Linnean Society*, 2001: 72, 47-62
- [21] Lewandowski, A. & Burczyk, J. & Mejnartowicz, L. : Genetic structure of English yew (*Taxus baccata* L.) in the Wierzchlas Reserve: implications for genetic conservation *Forest Ecology and Management*, 1995: 73, 221-227
- [22] Pollak, E.: Effective population numbers and mean time to extinction in dioecious populations with overlapping generations *Mathematical Biosciences*, 1980: 52, 1-25
- [23] Pollak, E.: Effective population numbers and mean time to extinction in monoecious populations with overlapping generations *Mathematical Biosciences*, 1979: 46, 87-106
- [24] Wang, Y. & Pollak, E. : The effective number of a population that varies cyclically in size. II. Overlapping generations *Mathematical Biosciences*, 2002: 179, 161-181
- [25] Polanski, A. & Bobrowski, A. & Kimmel, M. : A note on distributions of times to coalescence, under time-dependent population size *Theoretical Population Biology*, 2003: 63, 33-40
- [26] Burger, R. : Predictions of the Dynamics of a polygenic Character under directional selection *Journal of Theoretical Biology*, 1993: 162, 487-513
- [27] Nomura, T. : Effective population size for a sex-linked locus in populations under selection *Mathematical Biosciences*, 1997: 142, 79-89
- [28] Ishii, K. & Matsuda, H. & Iwasa, Y. & Sasaki, A. : Evolutionary stable mutation rate in a periodically changing environment *Genetics*, 1989: 121, 163-174
- [29] Pollak, E. : The effective population size of some age-structured populations *Mathematical Biosciences*, 2000: 168, 39-56
- [30] Sibert, A. & Austerlitz, F. & Heyer, E. : Wright-Fisher revisited: the case of fertility correlation *Theoretical Population Biology*, 2002: 62, 181-197

- [31] Wakeley, J. : The coalescent in an island model of population subdivision with variation among demes *Theoretical Population Biology*, 2001: 59, 133-144

A Appendix

Let X_s be the random variable denoting the number of population at sth generation, and the state space $\varphi = \{N_1, N_2\}$. And let

$$P(X_s(t+1) = X_s(t) | X_s(t) = N_j) = 1 - q_{ij} \quad (17)$$

$$P(X_s(t+1) \neq X_s(t) | X_s(t) = N_j) = q_{ij} \quad (18)$$

According to the above description

$$P(X_s(t+1) = N_2 | X_s(t) = N_1) = q_{i1} \quad (19)$$

$$P(X_s(t+1) = N_1 | X_s(t) = N_2) = q_{i2} \quad (20)$$

and

$$P(X_s(0) = N_1) = p_{s1}^{(0)} \quad (21)$$

It follows immediately that

$$P(X_s(t+1) = N_1 | X_s(t) = N_1) = 1 - q_{i1} \quad (22)$$

$$P(X_s(t+1) = N_2 | X_s(t) = N_2) = 1 - q_{i2} \quad (23)$$

and that the probability $p_{s2}^{(0)}$ of initially having N_2 individuals is given by

$$p_{s2}^{(0)} = P(X_0 = N_2) = 1 - p_{s1}^{(0)} \quad (24)$$

From this information , we can compute $P(X_s = N_1)$ and $P(X_s = N_2)$ We observe that

$$\begin{aligned}
P(X_s(t+1) = N_1) &= P(X_s = N_1 \wedge X_s(t+1) = N_1) \\
&+ P(X_s = N_2 \wedge X_s(t+1) = N_1) \\
&= P(X_s = N_1)P(X_s(t+1) = N_1 | X_s = N_1) \\
&+ P(X_s = N_2)P(X_s(t+1) = N_1 | X_s = N_2) \\
&= (1 - q_{i1})P(X_s = N_1) + q_{i2}P(X_s = N_2) \\
&= (1 - q_{i1})P(X_s = N_1) + q_{i2}(1 - P(X_s = N_1)) \\
&= (1 - q_{i1} - q_{i2})P(X_s = N_1) + q_{i2}
\end{aligned} \tag{25}$$

Now $P(X_0 = N_1) = p_{s1}^{(0)}$, so

$$P(X_1 = N_1) = (1 - q_{i1} - q_{i2})p_{s1}^{(0)} + q_{i2} \tag{26}$$

and

$$P(X_2 = N_1) = (1 - q_{i1} - q_{i2})P(X_1 = N_1) + q_{i2} \tag{27}$$

$$= (1 - q_{i1} - q_{i2})^2 p_{s1}^{(0)} + q_{i2}[1 + (1 - q_{i1} - q_{i2})] \tag{28}$$

It is easily seen by repeating this procedure s times that

$$P(X_s = N_1) = (1 - q_{i1} - q_{i2})^s p_{s1}^{(0)} + q_{i2} \sum_{j=0}^{s-1} (1 - q_{i1} - q_{i2})^j \tag{29}$$

As indicated in the paper let $0 < q_{ij} < 1$. Then by the formula for the sum of a finite geometric progression,

$$\sum_{j=0}^{s-1} (1 - q_{i1} - q_{i2})^j = \frac{1 - (1 - q_{i1} - q_{i2})^s}{q_{i1} + q_{i2}} \quad (30)$$

Then we conclude from (30) that

$$P(X_n = N_1) = \frac{q_{i2}}{q_{i1} + q_{i2}} + (1 - q_{i1} - q_{i2})^s (p_{s1}^{(0)} - \frac{q_{i2}}{q_{i1} + q_{i2}}) \quad (31)$$

and consequently that

$$P(X_n = N_2) = \frac{q_{i1}}{q_{i1} + q_{i2}} + (1 - q_{i1} - q_{i2})^s (p_{s2}^{(0)} - \frac{q_{i1}}{q_{i1} + q_{i2}}) \quad (32)$$

As $0 < q_{i1} + q_{i2} < 2$, which implies that $|1 - q_{i1} - q_{i2}| < 1$, we can let $s \rightarrow \infty$ in 32 and 33 and conclude that

$$\lim_{s \rightarrow \infty} P(X_s = N_1) = \frac{q_{i2}}{q_{i1} + q_{i2}} \text{ and } \lim_{s \rightarrow \infty} P(X_s = N_2) = \frac{q_{i1}}{q_{i1} + q_{i2}} \quad (33)$$

It is clear from the previous three equations that we should have

$$(p_{s1}^{(0)}, p_{s2}^{(0)}) = (\frac{q_{s2}}{q_{s1} + q_{s2}}, \frac{q_{s1}}{q_{s1} + q_{s2}}) \quad (34)$$

as starting distributions for $P(X_s = N_1)$ and $P(X_s = N_2)$ to be independent of s .

B Appendix

A stochastic process can be defined as any collection of random variables $X_t, t \in T$, defined on a common probability space, where T is a subset of $(-\infty, \infty)$ and is thought of as the

time parameter set. The process is called *continuous parameter process* if T is an interval having positive length and a *discrete parameter process* if T is a subset of the integers. If the random variable X_t all take on values from the fixed set φ , then φ is called the *state space* of the process.

Many systems of theoretical and applied interest have the property that given the present state, the past states have no influence on the future. This property is called the *Markov property*, and systems having this property are called *Markov chains*. The Markov property is defined precisely by the requirement that

$$P(X_{n+1} = x_{n+1} | X_0 = x_0, \dots, X_n = x_n) = P(X_{n+1} = x_{n+1} | X_n = x_n) \quad (35)$$

for every choice of nonnegative integer n and the numbers x_0, \dots, x_{n+1} , each in φ . The conditional probabilities $P(X_{n+1} = y | X_n = x)$ are called the *transition probabilities* of the chain. If transition probabilities are independent of n , they are called *stationary transition probabilities*.

$P(x, y), x$ and $y \in \varphi$, is a *transition function* if it satisfies

$$P(x, y) \geq 0 \quad (36)$$

and

$$\sum_y P(x, y) = 1 \quad (37)$$

$p_s^{(0)}, s \in \varphi$, is an *initial distribution* if it satisfies

$$p_s^{(0)} = P(X_0 = s) \quad (38)$$

$$p_s^{(0)} \geq 0, s \in \varphi \quad (39)$$

and

$$\sum_s p_s^{(0)} = 1 \quad (40)$$