Some statistical models for ovarian folliculogenesis in infant rats

K. L. Q. Read, J. C. Mariana, Marie-Madeleine de Reviers

Ann. Biol. anim. Bioch. Biophys., 1979, **19** (5), 1419-1433.

Some statistical models for ovarian folliculogenesis in infant rats

par K. L. Q. READ, J. C. MARIANA *, Marie-Madeleine de REVIERS *

University of Exeter, Dept. of Mathematical Statistics and Operational Research. Exeter EX4 4PU (G. B.)

* *Station de Physiologie de la Reproduction, I.N.R.A. Nouzilly 37380 Monnaie, France.*

Summary. In this paper we introduce a range of probabilistic models for the numbers and sizes of follicles found in the ovaries of infant rats before the first ovulation. Our aim is thus to generate hypotheses for possible further development ; we are not yet in a position to compute reliable estimates for well-proven models. Some key concepts in the modelling problem are (i) the distribution governing the times at which individual follicles become observable in the ovaries ; (ii) the ways in which follicles are lost to observation with the passage of time ; (iii) the distribution of initial sizes of follicles and their modes of growth.

Some assumptions are necessary to give a manageable analysis and the most important that we have made is that follicles are generated (and grow) independently of one another. To reduce the considerable computing problems, we also make some more specific assumptions about the distributions of follicular generation times, times to atresia and rates of growth. This report is concerned with the modelling of follicular numbers, but additional data on follicular sizes may be handled coherently within the same overall approach.

We consider three forms of model for folliculogenesis, based on negative exponential and second and third-order Gamma probability densities for follicular generation times with, in all cases, a displaced negative exponential density for the time from generation to atresia. The negative exponential model seems to be the most satisfactory, as well as the simplest. Estimates of parameters are given, but there is a partial indeterminacy in the estimation of the initial total number of follicles and the folliculogenesis rate parameter, and the atretic rate parameter has had to be restricted from taking excessively high values. Both of these difficulties should disappear, and all the estimates may change, when the full analysis of follicular numbers and growth rates is made simultaneously. In this event, a desirable refinement would be to allow the initial total number of follicles to vary between rats as a distributed parameter : a rough argument given in our paper then yields a simple estimate of the coefficient of variation. Clearly, the numerical estimations reported here must be regarded as provisional, pending the outcome of a comprehensive analysis. However, the results so far obtained, together with the flexibility of the available computable models, encourage us to believe that such a goal may not be unattainable eventually.

Introduction.

The aim of this paper is to introduce a range of probabilistic models for the numbers and sizes of follicles found in the ovaries of infant rats of different ages before ovulation. The emphasis is *exploratory*, in that we seek quantitative hypotheses consis-

tent with the data but are not in a position to compute reliable estimates for well-proven models. Some key concepts in the modelling problem are : 1) the distribution governing the times at which follicles come to be observed in the ovaries ; 2) the ways in which follicles are lost to observation with the passage of time ; 3) the distribution of initial sizes and modes of growth.

This problem divides naturally into two parts : first the modelling of follicular *numbers* and secondly the modelling of follicular *growth*. Our present report is primarily concerned with the first (and simpler) part of the work, but a coherent analysis of the whole problem can be set up within our general approach.

1. Materials and methods.

a) *Animals.* The animals used were Wistar rats (strain 03 INRA), bred so as to minimise variations between animals. The light regime was 14 hrs of daylight and 10 hrs of darkness. The mean temperature was 21 °C. The weaning was carried out at 23 ± 0.5 days.

b) *Ovarian histological methods.* The ovaries were removed immediately after death. They were fixed in Bouin Hollande's solution, then embedded in paraffin. The serial sections were cut at a thickness of 10 μ ; they were stained with Masson's trichrom. All the sections were inspected with a projection microscope. All the normal growing follicles that were non atretic and over 20 granulosa cells were considered The nucleolus of the nucleus of the oocyte which lies close to the section of maximal area, was used as a marker. The outline of the basal membrane was drawn on tracing-paper and the area measured by weighing the paper. Thus, we established the frequency curve according to follicle size. One ovary per rat was studied and two rats from different mothers were considered per age group.

A summary of the raw figures is given in table 1. The main apparent descriptive features of the data are as follows :

(i) The *minimum follicle sizes* recorded (which are necessarily not less than 1259 μ^2) show no obvious trend with time of examination (i. e. age). At different times variations between rats range between 1 p. 100 and 20 p. 100 of current average value.

(ii) *Average Values* of follicle sizes (computed as means or medians) increase at least until age 20 days. On the 20th and 24th days the mean values appear appreciably larger than the medians. This is presumably because small numbers of very large follicles are found at these ages : these follicles have a sizable influence on the mean values and also cause the distributions of follicle size to be positively skew. Variations between the means for rats of the same age can be quite large, too : at 8 days the difference ($3\,446 - 2\,703 = 743$) exceeds 9 times the average estimated standard error (75.0) ; at 20 days the ratio is nearly 5 to 1.

(iii) The *standard deviations* of the follicle sizes appear to increase with age. This trend reflects the general increase in spread of these distributions with time.

(iv) The *maximum sizes* observed increase with time faster than linearly throughout the age range 8-24 days. Variations between rats examined at the same date also increase but are not large enough to upset the overall trend.

TABLE 1

Summary statistics of follicle size distributions in individual rats

	Age (Days) 8	12	16	20	24
Rat number	**207**	**191**	**155**	**159**	**165**
Minimum	1 346	1 405	1299	1 329	1 538
Median	2 409	5 170.5	6 750	6 915	6 552
Mean	2 703	5 194	7 017	8 509	10 891
Maximum	5 935	12 175	25 666	76 573	172 364
Std deviation	898	2 385	3 971	7 311	16 198
Std error of mean	64.0	117.2	173.0	249.3	646.9
Skewness	0.964	0.335	0.923	3.421	4.809
Total number	197	414	527	860	627
Rat number	**208**	**194**	**158**	**161**	**166**
Minimum	1 402	1 604	1 261	1 349	1 264
Median	3 255.5	5 396	6 261	7 551	6 249
Mean	3 446	5 386	6 463	9 986	9 781
Maximum	6 880	13 194	18 668	96 749	167 329
Std deviation	1 203	2 352	3 670	10 033	16 175
Std error of mean	85.9	112.0	141.1	349.1	616.2
Skewness	0.618	0.342	0.524	3.050	6.235
Total number	196	441	677	826	689
Average figures					
Minimum	1 374	1 504.5	1 280	1 339	1 401
Median	2 832	5 283	6 505.5	7 233	6 400.5
Mean	3 074.5	5 290	6 740	9 248	10 336
Maximum	6 407.5	12 684.5	22 167	86 661	169 847
Std deviation	1 051	2 368	3 821	8 672	16 186
Std error of mean	75.0	114.6	157.0	299.2	621.6
Skewness	0.791	0.339	0.724	3.236	5.522
Total number	196.5	427.5	602	843	658

Notes :

1. The units of size are $(\mu m)^2$, i.e. $(10^{-6}\ m)^2$ or 10^{-12} of a square metre.
2. Follicles of size less than 1 259 $(\mu m)^2$ have not been counted.
3. The standard error of the mean is computed as Standard deviation/$\sqrt{\text{Total number}}$.
4. In the Average figures part of the table, each entry is the mean of the two corresponding entries above it : thus 1 374 = 1/2 (1 346 + 1 402), and so on.

(v) The *total numbers* of follicles measured increase at a rate of about 50 per day until at least the 20th day. Variations between rats of the same age vary from 1 on the 8 th day (0.5 p. 100 of the average total) to 150 on the 16th day (24.9 p. 100 of the average total). There is an apparent decline in numbers between 20 and 24 days, but the variations between rats make it difficult to be very precise about this.

2. Basic assumptions.

It seems that we have to begin by considering an « average rat ». To the extent that individual rats vary, in regard to initial follicle sizes, rates of growth of follicles,

numbers of oocyte follicles and any other relevant factors, this may not be valid, but the data from different rats can only be usefully combined on the working assumption that they are based on common (i.e. shared) values of at least some underlying parameters. Our assumptions fall into two groups (A) and (B). In the first set (A), we identify the basic fixed and random quantities in our representation and make certain hypotheses of statistical independence concerning the random elements. Set (B) concerns possible laws of distribution of the random elements.

Assumptions A1 to A5, concerning the folliculogenesis process in general, are as follows :

A1. There is a time origin (t_0 say, measuring from the birth of the rat) after which follicles may be observed in the ovaries if examined and before which no follicles would be observed. Mathematically, this time origin may be before the time of birth ($t_0 < 0$) or after birth ($t_0 > 0$).

A2. Individual follicles come to be observed (or, are generated) in the ovaries at random points in time after t_0 and thereafter behave *independently* (in a statistical sense).

A3. Having appeared, the individual follicles grow in size for a random period of time during which they would be observable in the ovaries if examined.

A4. The times spent in the ovaries are terminated *either* by premature mortality or decay (follicles become atretic), *or* by the reaching of « maturity » (in some sense), at which time the mature follicle leaves the ovary. In either event, follicles become « lost to observation ».

A5. The total number of follicles, N say, is in the first analysis assumed to be a constant for all rats. However, some further results are obtained on the more realistic basis that N is a distributed parameter which varies from rat to rat.

On these assumptions two basic random time periods characterize the appearance and disappearance of a follicle in the ovary. On general grounds, these non-negative quantities may be expected to follow unimodal distributions, and for convenience we work with special Erlangian distributions (Gamma distributions of integer order) (see Read *et al.*, 1968 and Ashford *et al.*, 1970). Displaced distributions, in which the random variables cannot be less than some positive value, are also useful. Graphs of four illustrative distributions are shown in figure 1, together with some basic summary information. Hence for each individual rat we have :

B1. The time delay from the start (t_0) of the generation process until a particular follicle is generated is a random variable following a Special (or General) Erlangian distribution. Different follicles are assumed to be independent and the distribution of the delay variable determines how the generation rate varies through time.

B2. Follicles remain observable in an ovary until lost to observation due to one of two or more causes. Associated with each possible cause is a random variable following a Special (or General) Erlangian distribution, these variables being mutually independent. The actual time from generation until loss to observation is the minimum of these random variables. (Since our data involve rats aged at most 24 days, we have

assumed that no losses to ovulation have occurred and hence that losses are due to atresia only).

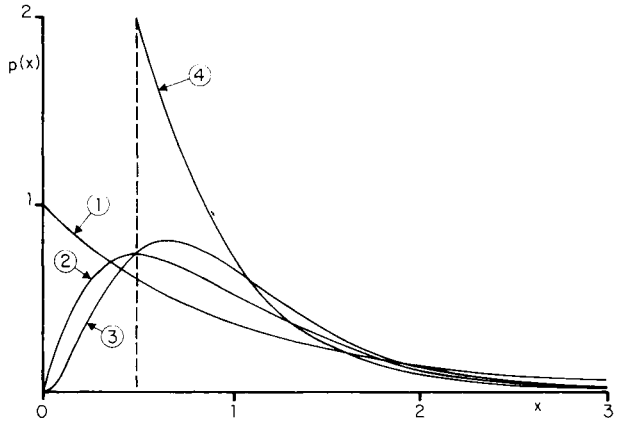

FIG. 1. *Some basic distributions. Probability densities* p(*x*) : (1) Negative exponential (Gamma, order 1) ; (2) 2nd Order Special Erlangian (Gamma, order 2) ; (3) 3rd Order Special Erlangian (Gamma, order 3) ; (4) Displaced Negative Exponential ; all with the same mean 1.

Formula	Mean μ	Mode μ^*	p (μ^*)	Variance σ^2	Standard Deviation σ	Coefficient of variation σ/μ
(1) $p(x) = e^{-x}, x > 0$.....	1	0	1	1	1	1
(2) $p(x) = 4x\,e^{-2x}, x > 0$	1	$\frac{1}{2}$	$\frac{2}{e} = 0.736$	$\frac{1}{2}$	$\frac{1}{\sqrt{2}} = 0.707$	$\frac{1}{\sqrt{2}} = 0.707$
(3) $p(x) = \frac{27}{2} x^2 e^{-3x}, x > 0$....	1	$\frac{2}{3}$	$\frac{6}{e^2} = 0.812$	$\frac{1}{3}$	$\frac{1}{\sqrt{3}} = 0.577$	$\frac{1}{\sqrt{3}} = 0.577$
(4) $p(x) = 2\,e^{-2(x-1/2)}, x > \frac{1}{2}$...	1	$\frac{1}{2}$	2	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{2}$

3. Construction of models for the observed numbers of follicles.

We consider three basic forms of model for the observed follicular numbers. These differ in the order (p = 1, 2 or 3) of the Gamma distribution which is assumed to characterise the delay times at which follicles leave the pool of little follicles. Let t_0 be the displacement parameter of effective time origin in all three models, i. e. the time (measuring from the birth of the rat as zero) at which the first follicles leave the pool. Then in Model p, where p = 1, 2 or 3, the probability density function (pdf) of the time at which follicles leave the pool may be written as :

$$f_p(t \mid \lambda, t_0) = \frac{\lambda^p (t - t_0)^{p-1} . \exp(-\lambda(t - t_0))}{(p-1)!}, \qquad t \geqslant t_0 \qquad (3.1)$$

$$= 0, \quad t < t_0.$$

λ is the scale (rate) parameter of the distribution (3.1) and the average follicular delay time is then $t_0 + p/\lambda$.

Once follicles have left the pool they are assumed to be subject to a process of atresia. However, because of the nearly constant rate of increase in observed numbers up to 20 days it seems that few or no follicles become atretic in rats younger than this, and therefore atretic time, i.e. the time from leaving the pool until atresia, is assumed to have a displaced distribution, so that all follicles survive at least a time t* say after leaving the pool (A high order undisplaced Gamma distribution might work but is much more awkward to compute ; also, by our previous assumption, no losses to ovulation occur). If we write t for atretic time, then the fall in numbers between 20 and 24 days suggests that the distribution of t is very heavily concentrated at values just greater than t* : the simplest suitable form is then the displaced exponential

$$\begin{aligned} q(t \mid \mu, t^*) &= \mu \exp(-\mu(t - t^*)) \qquad t \geqslant t^* ; \\ &= 0, \quad t < t^* . \end{aligned}$$

in which the rate parameter μ is relatively large and the average time to atresia beyond t* (i.e. $1/\mu$) is relatively small. The survivor function Q(t) of the time to atresia is then given by

$$\begin{aligned} Q(t \mid \mu, t^*) &= 1, \quad t \leqslant t^* ; \\ &= \exp(-\mu(t - t^*)), \quad t > t^* , \end{aligned}$$

representing the probability that a follicle survives for at least a time t* after leaving the pool before becoming atretic. We now have the basis of a simple probabilstic specification of the numbers of follicles observed at any given time before the onset of ovulation.

The total expected numbers observed at time t aged $\leqslant$ w are given by

$$E(w \mid t) = N \int_0^w f_p(t - u \mid \lambda, t_0)\, Q(u \mid \mu, t^*)\, du \tag{3.2}$$

and the probability that a follicle randomly chosen from the initial effective population is observable at time t is given by $D_p(t)$, say, where

$$D_p(t) = \frac{E(t' \mid t)}{N}$$

$t' = t - t_0$ being the maximum possible age of a follicle observed at time t. The full explicit formula for $E(t' \mid t)$ is derived in the Appendix (A1). We now suppose that the numbers seen in any given rat at age t are binomially distributed with number of trials N = total number in the pool initially and probability of being observed $= D_p(t)$ as given above. We may use this formulation to construct a χ^2-type statistic X^2 given by

$$\left.\begin{aligned} X^2 &= \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left[\frac{(O_{ij} - ND_p(t_i))^2}{ND_p(t_i)} + \frac{[(N - O_{ij}) - N(1 - D_p(t_i))]^2}{N(1 - D_p(t_i))} \right] \\ &= \sum_{i=1}^{k} \sum_{j=1}^{n_i} \frac{(O_{ij} - ND_p(t_i))^2}{ND_p(t_i)\,(1 - D_p(t_i))} , \end{aligned}\right\} \tag{3.3}$$

where O_{ij} = total observed numbers in the jth of the n_i rats examined at the ith of the k age-dates $t_1, \ldots, t_k$; note that to obtain a valid χ^2 statistic it is necessary to sum over an exhaustive set of groups and therefore the contributions to χ^2 from the numbers $N - O_{ij}$ of unobserved follicles and the corresponding probabilities $1 - D_p(t_i)$ must also be included. For $p = 1$, 2 and 3 we seek to estimate N, λ, μ, t' and t^* by minimising the criterion X^2. For this purpose we have used a well-tried minimisation algorithm due to Swann *et al.* (1964), noting that for any given values of λ, μ, t' and t^* the minimising value of N can be found by differentiation as in Appendix (A2). A practical difficulty is the tendency of μ to be estimated very large, leading to inaccurate computations. (It is expected that when the full estimation is carried out, taking into account the distributions of sizes and growth rates of the follicles (including the fact that some very large follicles must have survived for a long time without becoming atretic or else have grown very fast), this difficulty will not arise.) For the present we fix $\mu = 10$, giving a mean time to atresia of $t^* + 0.1$, and then minimise X^2 with respect to the parameters N, λ, t_0 and t^* for $p = 1$ and 2. Unfortunately the minimisation procedure still failed to converge in the case $p = 3$, due possibly to the inter-rat variation between numbers observed at the same age compounded with (in this instance) a poorly fitting model. To overcome this difficulty we modify the X^2 statistic by considering only the average numbers at each age : this gives Y^2, where :

$$Y^2 = \sum_{i=1}^{k} \frac{n_i(\bar{O}_i - ND_p(t_i))^2}{ND_p(t_i)\,(1 - D_p(t_i))}, \tag{3.4}$$

where $\bar{O}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} O_{ij}$ is the mean over the n_i rats which are examined at the ith age-date, $i = 1, \ldots, k$. It is in fact easily shown algebraically that

$$Y^2 = X^2 - \sum_{i=1}^{k} \sum_{j=1}^{n_i} \frac{(O_{ij} - \bar{O}_i)^2}{ND_p(t_i)\,(1 - D_p(t_i))}, \tag{3.5}$$

In the case of our data $k = 5$ and $n_1 = \ldots = n_5 = 2$, so that

$$X^2 = \sum_{i=1}^{5} \sum_{j=1}^{2} \frac{(O_{ij} - ND_p(t_i))^2}{ND_p(t_i)\,(1 - D_p(t_i))},$$

$$= Y^2 + \frac{1}{2} \sum_{j=1}^{5} \frac{(O_{i1} - O_{i2})^2}{ND_p(t_i)\,(1 - D_p(t_i))},$$

Minimisation of Y^2 for $p = 3$ proved successful, and for comparison we also present the results of minimising Y^2 for $p = 1$ and $p = 2$.

The difference $X^2 - Y^2$ can be used to investigate the possible variation between rats of the initial total number of follicles, which up to now we have taken as a constant (N). Suppose now that for all i and j the initial total number of follicles for the jth rat on the ith age-date is N_{ij} instead of N, where the N_{ij}'s are themselves distributed with mean N and variance σ^2. Then the difference $X^2 - Y^2$, which would correspond in the conventional analysis of variance for regression to a sum of squares for pure error, has expectation (see Appendix, A3) given by :

$$E\left\{\sum_{i=1}^{k}\sum_{j=1}^{u_i}\frac{(O_{ij}-\bar{O}_i)^2}{ND_p(t_i)\,(1-D_p(t_i))}\right\}=\sum_{i=1}^{k}(n_i-1)+\frac{\sigma^2}{N}\sum_{i=1}^{k}\frac{(n_i-1)\,D_p(t_i)}{1-D_p(t_i)}\,, \quad (3.6)$$

By calculating X^2 and Y^2 for a set of estimates of N, λ, t_0 and t* (μ being taken as 10) we may obtain an estimate of σ^2 by equating $X^2 - Y^2$ to its expectation as given by (3.6) and solving for σ^2 and hence for the standard deviation σ. The estimates of σ corresponding to p = 1, 2 and 3 and the estimated values of N, λ, t_0 and t* are shown in table 3. It is also of interest to note that $E(Y^2)$ (see Appendix, A3) is given by

$$E\left\{\sum_{i=1}^{k}\frac{n_i(\bar{O}_i-ND_p(t_i))^2}{ND_p(t_i)\,(1-D_p(t_i))}\right\}=k+\frac{\sigma^2}{N}\sum_{i=1}^{k}\frac{D_p(t_i)}{1-D_p(t_i)}\,, \quad (3.7)$$

$$=\frac{1}{2}E(X^2) \text{ when each } n_i=2,\ i=1,\ \ldots,\ k\,,$$

as is the case for our data.

Results.

We consider here the use of first, second and third order (p = 1, 2 and 3) delay time distributions in models as described in section 3, Estimates for (a) analyses of the unaveraged observed numbers for the cases p = 1 and p = 2 and (b) of the averaged figures for p = 1, 2 and 3 are presented in table 2. Judging by the criterion values Y^2 for the runs b (p = 1, 2, 3), we see that the model based on negative exponential delay times gives by far the best (smallest) value of the criterion, the result for p = 3 being worse than that for p = 2. Similarly, run a with p = 1 achieves a much better X^2 value than does run a with p = 2. Comparison of observed and expected numbers of follicles also shows clearly that the trend of the observed numbers (which is best shown in the row of averaged numbers just below the data for the individual rats) is most faithfully followed by the values fitted by the negative exponential model (p = 1, either of runs a and b). All three models predict well for age 12 days but the runs with p = 1 get closer to the peak at 20 days and also fit better in showing bigger reductions in numbers from 20 to 24 days. At 8 days and 16 days the runs with p = 1 are again closer to the averages of the observed numbers. It should be noted that χ^2-type significance tests of goodness of fit using the X^2 and Y^2 values would be incorrect. However, it is encouraging that for the p = 1 runs (and for these only) the discrepancies of the fitted from the average observed figures are in general no greater than the typical variations seen between rats of the same age.

For each of the analyses tabulated we also indicate, in the lines labelled R and *F*, how the rates at which follicles leave the pool, and the proportions of follicles left in the pool, vary with time as estimated in the different models. For the pth order Gamma distribution of the time (measuring from t_0) at which follicles leave the pool, the instantaneous rate, R say, of leaving the pool at time t is given by $Nf_p(t\,|\lambda,\,t_0)$ where $f_p(t\,|\lambda,\,t_0)$ is given in equation (3.1). Correspondingly, the expected proportion, *F* say, of the total initial numbers still left in the pool at time t is given by integration of (3.1) as

$$F_p(t\,|\lambda,\,t_0)=\exp\left[-\lambda(t-t_0)\right]\sum_{s=0}^{p-1}\frac{\left[\lambda(t-t_0)\right]^s}{s\,!}\,, \text{ for } p=1,\,2,\,3,\,\ldots,\text{ and } t>t_0\,.$$

TABLE 2

Comparison of numbers of follicles observed and expected on the basis of 3 Models

Model		Age (Days)	8	12	16	20	24	λ	μ	t_0	t*	N	σ	Goodness of fit
		Observed Numbers	197 196	414 441	527 677	860 826	627 689							
		Observed Nos. (Averaged)	196.5	427.5	602	843	658							
p = 1 Negative Exponential	a	Fitted nos.	192.0	428.6	630.8	801.2	687.1	0.039 3	10	5.160	14.75	1 820	150	$X^2 = 50.22$
		R	63.9	54.6	46.7	39.9	34.1							
		F	0.895	0.765	0.653	0.559	0.477							
	b	Fitted nos.	190.9	432.1	634.1	802.5	675.8	0.044 4	10	5.271	14.68	1 674	139	$Y^2 = 15.17$
		R	65.8	55.1	46.1	38.6	32.4							
		F	0.886	0.742	0.621	0.520	0.436							
p = 2 Second Order	a	Fitted nos.	185.0	427.2	654.7	777.7	702.1	0.129 3	10	2.875	14.34	1 295	107	$X^2 = 102.32$
		R	57.2	60.7	52.0	40.5	29.8							
		F	0.857	0.670	0.494	0.351	0.243							
	b	Fitted nos.	183.3	432.8	653.0	786.5	690.8	0.1504	10	3.263	14.90	1 144	95	$Y^2 = 54.91$
		R	60.1	60.8	48.5	34.9	23.7							
		F	0.840	0.622	0.429	0.284	0.182							
p = 3 Third Order	b	Fitted nos.	181.7	427.7	660.7	779.2	696.0	0.206 2	10	1.342	14.51	1 136	94	$Y^2 = 71.84$
		R	55.9	62.8	52.1	37.0	23.9							
		F	0.840	0.623	0.418	0.261	0.155							

Notes

R : estimated current daily rates at which follicles leave pool.
F : estimated current proportion of follicles left in pool.
a : obtained by minimising X^2 as defined in eq. (3.2).
b : obtained by minimising Y^2 as defined in eq. (3.3).

Comparing first the analyses 1*a* and 1*b*, we see that the estimated rates are very similar, differing by less than two follicles per day at the ages observed. However, the proportion of follicles left in the pool is consistently higher for Ia, due to the lower estimate of λ and higher estimate of N than in 1*b*. A similar relationship exists between the results in 2*a* and those in 2*b*. This tendency for N and λ to compensate each other in this way presumably arises because the observed numbers effectively determine the rates (which essentially relate to $N\lambda$) at which follicles leave the pool more closely than they do the proportions of the initial population which are left in the pool at 24 days. These estimated proportions thus vary by a factor of more than three to one, from 0.477 in run 1*a* to 0.155 in 3*b*. Put otherwise, the mean delay time in terms of age, i.e. from time $t = 0$, is estimated as 27-30 days for $p = 1$, 16-19 days for $p = 2$ and about 16 days for $p = 3$. In a full analysis in which follicular sizes are taken into account along with some assumptions about growth, we should have better information about the distribution of times at which follicles leave the pool. Even with negative exponentially distributed delay times (i.e. times at which follicles leave the pool), the full model may produce mutually compensating values for N and λ which are noticeably different from those given here.

The estimates of t_0 and t^* (about 5.2 and 14.7 days respectively) are similar for the negative exponential ($p = 1$) analyses and imply that atresia sets in just before age 20 days. This is consistent with the steady rise in numbers up to this time, and a very high mortality or atretic rate thereafter is necessary to fit the subsequent decline. The displaced negative exponential distribution used here is the simplest way of providing for a relatively sudden onset of atresia if this is indeed what happens.

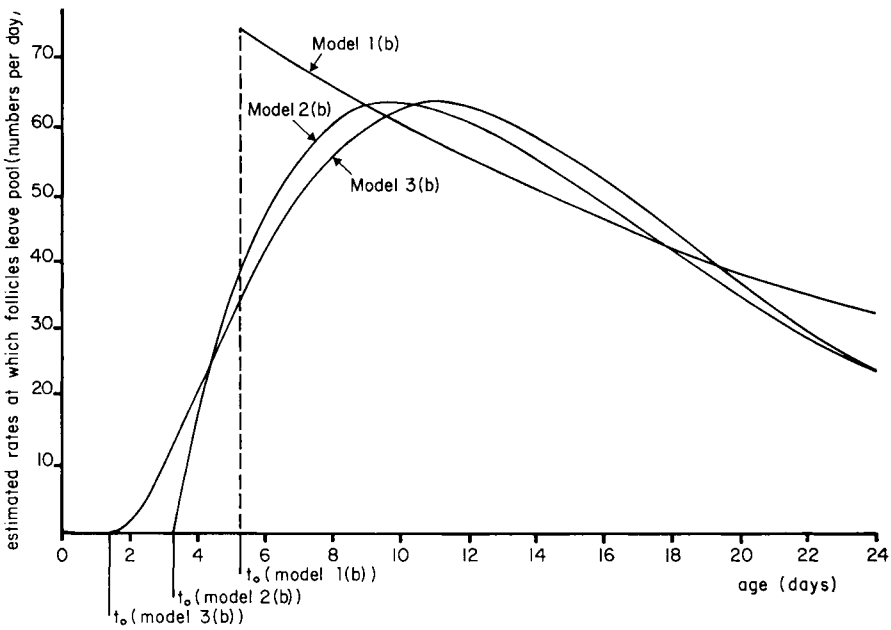

FIG. 2. — *Folliculogenesis rates estimated for 3 Models.*

The higher estimates of λ and lower estimates of t_0 and N for the second and third order models are partly to be expected, as may be seen from figure 2. When $p = 2$ or 3 the delay time density is zero at $t = t_0$ and small for t close to t_0. To allow for this early period during which very few follicles leave the pool it is necessary to estimate t_0 earlier when $p = 2$ and earlier still when $p = 3$. At the same time, to fit the observed trend high values of the rate parameter λ are necessary so that sufficiently fewer follicles leave the pool between 20 and 24 days than the numbers doing so soon after t_0 which become subject to atresia from around 20 days onwards. The difficulty of accomplishing this with the second and third order distributions, which have modes at $t_0 + 1/\lambda$ and $t_0 + 2/\lambda$ respectively, in shown up by the relative success of the negative exponential density which has its mode at t_0. The estimates of N for the second and third order models, about 2/3 of those for $p = 1$, then compensate for the high estimates of λ.

The parameter σ denotes the standard deviation of the initial population of the pool of little follicles in individual rats. Although the estimate σ varies from 94 in run 3*b* to 150 in 1*a*, the ratio σ/N (the coefficient of variation of the N_{ij}'s) is estimated with remarkable consistency, the values in all five runs lying within 1 p. 100 and 2 p. 100 of the average value of about 0.083 or 1/12.

Conclusions.

Even granted the assumptions listed in section 2, because we have concentrated on the observed numbers of follicles rather than their sizes, our conclusions must at this stage be regarded as provisional. It is premature to regard any of the model discussed here as well-proven, and primarily for this reason we have not carried out the (very considerable labour of) calculation of standard errors and confidence limits for our estimates which would conventionally be expected. Our view is that the models formulated must first be accepted as qualitatively realistic, and until this is settled the specification of numerical estimates of uncertainty may be misleading as well as inappropriate. However, a number of conclusions emerge unambiguously from our exercise, and these are as follows :

(i) A displaced negative exponential distribution for the delay times at with follicles leave the pool is definitely preferable to the Gamma distributions which were also tried.

(ii) Losses to atresia appear to be zero or negligible until age 20 days but then they become very high. The distribution of atretic time, i.e. the time from leaving the pool until atretic decay, is for simplicity taken to be of displaced negative exponential form ; a high order Gamma distribution might be a possible though computationally unattractive alternative.

(iii) The model based on two negative exponential distributions gives a fair fit to the observed numbers. There is a partial indeterminacy in the estimation of the initial total number of follicles and the delay time rate parameter, and the atretic rate parameter has had to be restricted from taking excessively high values. Both of these difficulties should disappear in the full analysis when the sizes or ages of individual follicles,

including those large follicles which have survived atresia for some time, are taken into account.

(iv) The variability between rats of the initial total number of follicles is discussed, including the possibility of treating it as a distributed parameter. The estimated coefficient of variation is about 0.083 or 1/12.

As stated in the Introduction, this is an exploratory paper which represents only the first steps in constructing a stochastic representation of ovarian folliculognenesis in infant rats (and, perhaps, more generally). We have considered three forms of model for folliculogenesis, the best of which may well be relevant to a full analysis of folliculogenesis and growth. We feel that this preliminary work is justified because : first, we have gained valuable insights about the form of model which is likely to be satisfactory ; secondly, we have obtained information about the parameters of the distributions of times at which follicles leave the pool or become atretic, and about related population dynamical quantities of interest ; and in the third place computation of the full analysis is likely to be a good deal easier when based on the results of the relatively simple estimations reported here. For these reasons and because of the promise shown by the negative exponential delay time model, we are optimistic that a comprehensive analysis of folliculogenesis and growth will eventually be successful.

EEC Seminar on « Ovarian stimulation and egg quality in mammals », Luynes, France, octobre 1978.

Acknowledgements. — The authors wish to acknowledge the facilities af funding ond the Institut National de la Recherche Agronomique (INRA, France) ; also the help of the University of Exeter (England) Computer Unit in data processing, and the use of the facilities of the UK South Western Universities Computer Network, without which this work could not have been done.

Résumé. Dans cette communication, nous présentons un ensemble de modèles probabilistes pour les nombres des follicules observés dans les ovaires chez la jeune ratte avant la première ovulation. Notre objectif est de développer des hypothèses susceptibles de prolongements ultérieurs ; nous ne sommes cependant pas encore en mesure de calculer des estimations valables à partir de modèles définitifs. Les points-clé de ces modèles sont : 1) la distribution qui régit les instants auxquels les follicules sont observés en phase de croissance dans les ovaires ; 2) la façon dont les follicules deviennent atrétiques au cours du temps ; 3) la distribution des tailles initiales des follicules et leur mode de croissance. Quelques hypothèses sont nécessaires pour aboutir à une analyse mathématique. La plus importante que nous avons faite est que les follicules sont engendrés et croissent indépendamment les uns des autres.

Pour réduire les très importants problèmes de calcul, nous faisons également des hypothèses plus spécifiques concernant les distributions selon lesquelles les follicules sortent de la réserve des primordiaux au cours du temps, les moments d'atrésie et les taux de croissance. Dans la présente communication, on considère seulement une modélisation des nombres de follicules, mais des données plus nombreuses concernant la taille des follicules pourraient être traitées de façon cohérente par la même approche.

Nous considérons trois types de modèles pour la folliculogenèse fondés sur des lois de probabilité exponentielles négatives et gamma jusqu'au troisième ordre pour les temps de génération des follicules. Dans tous les cas, la fonction de densité des intervalles de temps séparant le moment du démarrage de la croissance du moment de l'atrésie est une loi

exponentielle négative déplacée. Le modèle exponentiel semble le plus satisfaisant ainsi que le plus simple. Chacun des paramètres est estimé, mais il y a une indétermination partielle dans l'estimation du nombre initial des follicules et du paramètre caractérisant le taux de sortie des follicules. Il a été nécessaire de limiter la valeur que pouvait prendre le taux d'atrésie. L'ensemble de ces difficultés devrait disparaître dans une analyse complète portant sur les nombres et les taux de croissance des follicules. Dans ce cas, il serait nécessaire de tenir compte du fait que le nombre total des follicules varie d'un animal à l'autre, selon une certaine distribution. Nous donnons dans cet article les éléments permettant d'estimer le coefficient de variation. Cependant, les estimations numériques que nous fournissons doivent être considérées comme provisoires, mais les résultats obtenus et la flexibilité des modèles utilisés nous encouragent à penser qu'une analyse plus globale du phénomène peut être réalisée.

References

ASHFORD J. R., READ K. L. Q., VICKERS G. G., 1970. A system of stochastic models applicable to studies of animal population dynamics. *J. Anim. Ecol.*, **39**, 29-50.

READ K. L. Q., ASHFORD J. R., 1968. A system of models for the life cycle of a biological organism. *Biometrika*, **55**, 211-221.

SWANN W. H., DAVIES D., CAMPEY I. G., 1964. Report on the development of a new direct search method of optimisation. ICI (UK), Centr. Instrum. Lab. Res. Note 64/3.

Appendix

A.1. *The expected total number of follicles observable at time t*

In this paragraph we derive the expectation, denoted by $E(t' \mid t) = ND_p(t)$ in section 3, of the number of follicles observed at time t. We begin by substituting the formula (3.1) for $f_p(t - u \mid \lambda, t_0)$ into equation (3.2), and for brevity will write t' for $t - t_0$. We suppose first that w in (3.2) does not exceed t* : then

$$E(w \mid t) = N \int_0^w \frac{\lambda^p (t' - u)^{p-1} \exp\big(-\lambda(t' - u)\big)}{(p-1)!}\, du$$

$$= N \left\{ \exp\big(-\lambda(t' - w)\big) \sum_{s=0}^{p-1} \frac{[\lambda(t' - w)]^s}{s!} - e^{-\lambda t'} \sum_{s=0}^{p-1} \frac{(\lambda t')^s}{s!} \right\}.$$

The maximum possible age of a follicle at time t is clearly t' ; hence, if $t' \leqslant t^*$ the total expected numbers at time t are given by

$$E(t' \mid t) = N \left\{ 1 - \exp(-\lambda t') \sum_{s=0}^{p-1} \frac{(\lambda t')^s}{s!} \right\},$$

$$= ND_p(t), \text{ say, if } t' \leqslant t^* \text{ or } t \leqslant t_0 + t^*.$$

If $t' > t^*$ and $w > t^*$ then

$$E(w \mid t) = N \left\{ \int_0^{t^*} \frac{\lambda^p (t' - u)^{p-1} . \exp\big(-\lambda(t' - u)\big)}{(p-1)!}\, du + \int_{t^*}^w \frac{\lambda^p (t' - u)^{p-1} . \exp\big(-\lambda(t' - u) - \mu(u - t^*)\big)}{(p-1)!}\, du \right\}$$

$$= N \left\{ \exp(-\lambda(t'-t^*)) \sum_{s=0}^{p-1} \frac{[\lambda(t'-t^*)]^s}{s!} - \exp(-\lambda t') \sum_{s=0}^{p-1} \frac{(\lambda t')^s}{s!} , \right.$$

$$\left. + \exp(-\mu(t'-t^*)) \int_{t'-w}^{t'-t^*} \frac{\lambda^p v^{p-1} \exp((\mu-\lambda)v)}{(p-1)!} \, dv \right\},$$

$$= N \left\{ \exp(-\lambda(t'-t^*)) \sum_{s=0}^{p-1} \frac{[\lambda(t'-t^*)]^s}{s!} - \exp(-\lambda t') \sum_{s=0}^{p-1} \frac{(\lambda t')^s}{s!} \right.$$

$$\left. + \frac{\exp(-\mu(t'-t^*))\lambda^p}{(-1)^{p-1}\nu^p} \left[\exp(\nu(t'-t^*)) \sum_{s=0}^{p-1} \frac{[-\nu(t'-t^*)]^s}{s!} - \exp(\nu(t'-w)) \sum_{s=0}^{p-1} \frac{[-\nu(t'-w)]^s}{s!} \right] \right\},$$

where $\nu = \mu - \lambda$ and we assume $\nu \neq 0$. If we put $w = t'$ we obtain the total expected numbers as

$$E(t' \mid t) = N \left\{ \exp(-\lambda(t'-t^*)) \sum_{s=0}^{p-1} \frac{[\lambda(t'-t^*)]^s}{s!} - \exp(-\lambda t') \sum_{s=0}^{p-1} \frac{(\lambda t')^s}{s!} \right.$$

$$\left. + \frac{\lambda^p . \exp(-\mu(t'-t^*))}{(-1)^p \nu^p} \left[1 - \exp(\nu(t'-t^*)) \sum_{s=0}^{p-1} \frac{[-\nu(t'-t^*)]^s}{s!} \right] \right\},$$

$$= N D_p(t), \text{ say, if } t' > t^*, \text{ ie if } t > t_0 + t^* .$$

A.2. *The estimation of* N

For brevity we write p_i for $D_p(t_i)$, q_i for $1 - D_p(t_i)$, $i = 1, \ldots, k$. In the formula

$$X^2 = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \frac{(O_{ij} - Np_i)^2}{Np_i q_i} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left\{ \frac{O_{ij}^2}{Np_i q_i} - \frac{2 O_{ij}}{q_i} + \frac{Np_i}{q_i} \right\},$$

we require

$$\frac{\partial X^2}{\partial N} = - \sum_{i=1}^{k} \sum_{j=1}^{n_i} \frac{O_{ij}^2}{N^2 p_i q_i} + \sum_{i=1}^{k} \frac{n_i p_i}{q_i}$$

to be zero, which occurs when

$$N^2 = \left[\sum_{i=1}^{k} \frac{1}{p_i q_i} \sum_{j=1}^{n_i} O_{ij}^2 \right] \Big/ \left[\sum_{i=1}^{k} \frac{n_i p_i}{q_i} \right] = \left[\sum_{i=1}^{k} \frac{n_i \overline{O_i^2}}{p_i q_i} \right] \Big/ \left[\sum_{i=1}^{k} \frac{n_i p_i}{q_i} \right],$$

showing that the minimising value of N is a root (weighted) mean square. We note that this differs from the condition that the total observed and expected numbers be equal, which may be written

$$\sum_{i=1}^{k} \sum_{j=1}^{n_i} O_{ij} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} Np_i ,$$

i. e.

$$N = \left[\sum_{i=1}^{k} \sum_{j=1}^{n_i} O_{ij} \right] \Big/ \left[\sum_{i=1}^{k} n_i p_i \right] \quad \text{or} \quad \left[\sum_{i=1}^{k} n_i \overline{O}_i \right] \Big/ \left[\sum_{i=1}^{k} n_i p_i \right].$$

The argument for Y^2 is similar. We have

$$Y^2 = \sum_{i=1}^{k} \frac{n_i(\overline{O}_i - Np_i)^2}{Np_i q_i} = \sum_{i=1}^{k} \left\{ \frac{n_i \overline{O_i}^2}{Np_i q_i} - \frac{2 n_i \overline{O}_i}{q_i} + \frac{N n_i p_i}{q_i} \right\},$$

so that

$$\frac{\partial Y^2}{\partial N} = - \sum_{i=1}^{k} \frac{n_i \overline{O_i}^2}{N^2 p_i q_i} + \sum_{i-1}^{k} \frac{n_i p_i}{q_i} ,$$

which is zero when

$$N^2 = \left[\sum_{i=1}^{k} \frac{n_i \overline{O_i}^2}{p_i q_i}\right] \Big/ \left[\sum_{i=1}^{k} \frac{n_i p_i}{q_i}\right],$$

A.3. *The expected values of the statistics* X^2 *and* Y^2

For brevity we write p_i for $D_p(t_i)$ and q_i for $1 - D_p(t_i)$, $i = 1, \ldots, k$. Our model for variation between rats is that each O_{ij} is binomially distributed with number of trials N_{ij} and probability parameter p_i, whilst each N_{ij} follows (or is drawn from) a distribution with mean N and variance σ^2 which apart from this is unspecified. It follows that the first two unconditional moments of O_{ij} are :

$$E_{N_{ij}}\left[E(O_{ij} \mid N_{ij})\right] = E_{N_{ij}}\left[N_{ij} p_i\right] = Np_i, \; i = 1, \ldots, k \; ;$$

$$E_{N_{ij}}\left[E(O_{ij}^2 \mid N_{ij})\right] = E_{N_{ij}}\left[N_{ij} p_i q_i + N_{ij}^2 p_i^2\right]$$

by the standard results for the binomial distribution,

$$= Np_i q_i + p_i^2(\sigma^2 + N^2), \; i = 1, \ldots, k \, .$$

From (3.5) in terms of our shortened notation,

$$E(X^2 - Y^2) = E\left\{\sum_{i=1}^{k}\sum_{j=1}^{n_i} \frac{(O_{ij} - \overline{O}_i)^2}{Np_i q_i}\right\}$$

$$= \sum_{i=1}^{k}\left\{\frac{n_i}{Np_i q_i} E(O_{ij}^2) - \frac{2}{Np_i q_i} E\left[O_{ij}\sum_{m=1}^{n_i} O_{im}\right] + \frac{1}{Nn_i p_i q_i} E\left[\left(\sum_{m=1}^{n_i} O_{im}\right)^2\right]\right\},$$

$$= \sum_{i=1}^{k}\left\{\frac{n_i}{Np_i q_i} E(O_{ij}^2) - \frac{1}{Nn_i p_i q_i}\left[\sum_{j=1}^{n_i} E(O_{ij}^2) + 2\sum_{1 \leqslant j < m \leqslant n_i} E(O_{ij} O_{im})\right]\right\},$$

$$= \sum_{i=1}^{k}\left\{\frac{(n_i - 1)}{Np_i q_i}\left[Np_i q_i + p_i^2(\sigma^2 + N^2)\right] - \frac{n_i(n_i - 1)}{Nn_i p_i q_i} \cdot N^2 p_i^2\right\},$$

$$= \sum_{i=1}^{k}(n_i - 1) + \frac{\sigma^2}{N}\sum_{i=1}^{k}\frac{(n_i - 1)\, p_i}{q_i} \, ,$$

which is equivalent to equation (3.6).

Similarly, starting from (3.4) we have

$$E(Y^2) = E\left\{\sum_{i=1}^{k}\frac{n_i(\overline{O}_i - Np_i)^2}{Np_i q_i}\right\},$$

$$= \sum_{i=1}^{k}\left\{\frac{E\left[\left(\sum_{j=1}^{n_i} O_{ij}\right)^2\right]}{Nn_i p_i q_i} - \frac{2}{q_i} E\left[\sum_{j=1}^{n_i} O_{ij}\right] + \frac{Nn_i p_i}{q_i}\right\},$$

$$= \sum_{i=1}^{k}\left\{\frac{n_i\left[Np_i q_i + p_i^2(\sigma^2 + N^2)\right] + n_i(n_i - 1) N^2 p_i^2}{Nn_i p_i q_i} - \frac{Nn_i p_i}{q_i}\right\},$$

$$= k + \frac{\sigma^2}{N}\sum_{i=1}^{k}\frac{p_i}{q_i} \, ,$$

which is equivalent to (3.7).