6. PKT Problem 3.1 (p. 115) (8 points)

Suppose the synthesis rate for proteins is k_s and the degradation rate is k_d. Then the equation for the total number of proteins N(t) in the cell at time t is

$$\frac{dN}{dt} = k_s - k_d N .$$

Note that previously, we had neglected degradation, so the (old) synthesis rate just needed to satisfy: $N_1 = k_s^0 \tau$.

This differential equation has the solution,

$$N(t) = N_0 e^{-k_d t} + \frac{k_s}{k_d}$$ (verify this!)

We are looking for a solution which starts at the beginning of the cell cycle t=0 with the one-cell required number of protein molecules, say N_1, and ends up at the end of the cell cycle t=τ with twice that number.

Thus, $N_1 = N_0 + \frac{k_s}{k_d}$ and $2N_1 = N_0 e^{-k_d \tau} + \frac{k_s}{k_d}$.

It is easy to solve these two equations for k_s: $k_s = k_d N_1 \left(\frac{2 - e^{-k_d \tau}}{1 - e^{-k_d \tau}} \right)$.

We are told that the destruction rate is such that the protein half-life is 60 minutes. To make things simple, I will assume that this is the same as the cell-cycle life time. This means that $e^{-k_d \tau} = \frac{1}{2}$, so $k_d = \frac{\ln 2}{\tau}$, and

$k_s = 3 k_d N_1 = 3\left(\frac{\ln 2}{\tau}\right) N_1 = (3 \ln 2) k_d^0 = (2.08) \times 10^3$ proteins/sec, where I have used the old calculation (without degradation) of 10^3 protein/sec (see Eq. (3.2) in PKT).

The next question is, how many ribosomes working simultaneously does it take to produce proteins at this rate?

We are told in Sec. 3.2.1 (p. 93) that the *measured* rate of amino-acid production per ribosome is 25/s. It follows that the number of ribosomes needed to produce the required $2 \times 10^3 \cdot 300$ amino-acid units is $\frac{2 \times 10^3 (300)}{25} = 24{,}000$ ribosomes. This compares pretty well with the estimate in PKT Sec. 2.2.1 (p. 32) of 20,000 ribosomes/cell.

Comment: It is tempting to use the value we calculated at Lecture 6.1 of 20 seconds for a ribosome to produce a complete protein. This leads to 40,000 ribosomes needed. However, this reasoning is circular, since we derived this number using 20,000 ribosomes/cell but neglecting the degradation. Since accounting for the degradation required doubling the rate of protein synthesis, we find (of course) that we need twice as many ribosomes as before. Not a big surprise! The logic is better if we just use the *measured* rate of amino-acid incorporation.

How did you all do on this problem?
One person did it completely correctly (congratulations!).

Most others did a rough approximation, to which I gave almost full credit: You were given the protein half-life as essentially the cell doubling time. thus, you argued that over the course of the doubling time half the initial protein inventory degraded. That's $3\text{x}10^6/2$ proteins (or $2.4\text{x}10^6/2$ if you use the figure from Table 2.1). Most people who did this assumed 20,000 ribosomes without degradation. Assuming the same rather of synthesis per ribosome, you will then need 50% more ribosomes, or roughly 30,000.

There are two problems:
a) The overall rate of degradation is proportional to the number of proteins present. Over the course of the cell cycle the number of structural proteins doubles (to provide material for the daughter cells). Thus, there will at most time be more proteins than the initial inventory (up to twice as many), so the overall rate will be significantly higher than you calculated.
b) In using the 20,000 ribosome figure and the15 aa/s/ribosome synthesis rate, you are actually using as input the results of a calculation in which there was no degradation, which is a bit illogical. I prefer to use the measure synthesis rate to CALCULATE (de novo) the number of needed ribosomes.

If you follow this route but use the simplified method, you might take the average of ½ times the initial protein (using the initial protein inventory) and 1 times the initial protein (using the final protein inventory, just before division). Thus you would need to synthesize $\left(1+\frac{3}{4}\right)3\times10^6(300)=1.575\times10^9$ amino-acid residues. At 25 aa/s/ribosome, that would require $\frac{1.575\times10^9}{25(50\cdot60)}=21{,}000$ ribosomes, where I have used the measured per-ribosome synthesis rate referenced above. This approach gives a realistic value for the ribosome number without the circular logic.

7. PKT Problem 3.2 (p.115) (8 points= 2 points for each part)

(a) From PKT image the 0.1 μm scale is about 0.85 cm, while the entire segment is about 11.5 cm, so the total length is about $\frac{11.5}{0.85}\times0.1=1.35$ μm = 1350 nm. There at 0.34 nm/bp, so the number of base pairs is about $\frac{1335}{0.34}\approx4000$.
We are told that the fly genome is$1.8\text{x}10^8$ bp; thus, the fraction of the genome pictured must be about $\frac{4000}{1.8\times10^8}\boxed{=2.2\times10^{-5}.}$

(b) There is a DNA polymerase at each fork. I count 8 forks (2 for each “bubble”). Assuming this segment of DNA is representative (and also, of course, that all the DNA polymerase molecules are attached to the genome), we expect a total of about $8\cdot\frac{1}{2.2\times10^{-5}}=\boxed{3.6\times10^{5}\text{ molecules.}}$

(c) I measure 2.2 cm between forks 4 and 5, so $\frac{2.2}{0.85}\cdot\frac{0.1\times10^{3}}{0.34}\approx\boxed{760\text{ bp.}}$ Replicating at 40x2 bp/s gives $\frac{760}{40\cdot 2}=9.5$ s until forks meet.

(d) There are 8 replication forks over the 4000 bp’s shown, each moving at (we are told 40 bp/s. So, the time to replicate this segment would be approximately $\frac{4000}{8\cdot 40}=\boxed{12.5\text{ s,}}$ if we assume that all the forks start at the same time. If this segment is typical, then that would be an estimate of the time to replicate the entire genome. But, clearly, the forks start at different times. For example the biggest “bubble” is roughly 5 cm larger than the smallest, so they must have started at times different by $\frac{5}{1.2}\cdot\frac{0.1\times10^{3}}{0.34}\cdot\frac{1}{2\times40}\approx15$ s. Any careful calculation would have to include assumptions about whether the bubble initiate at random spatial positions or at fixed ones (both options occur in some organisms). If the positions are random, then the time for completion of replication will correspond to the longest wait-time and the largest interval between. So, I’d estimate $\boxed{25\text{—}30\text{ s.}}$

8. (6 points=3 points for each part)

I will describe a game to you. Your job is to use a probability analysis to calculate the probability of winning using one of two strategies:
You are a contestant on a game show. In this show there are three closed doors A, B, and C. Behind one of them is a prize; behind the other two there is nothing. If you pick the right door, you get the prize. You don’t know which door has the prize behind it but the host does know. The host asks you to pick a door. You choose door A. The host now opens one of the other doors, say, door B and shows you that there is nothing behind it. The host then asks you “Would you like to change your guess to C or would you like to stick with door A?”

(a) What should you do? What is your probability of winning if you switch? What is your probability of winning if you stick? Whichever answer you give, EXPLAIN your reasoning carefully.
Note: You will be tempted to answer that your probability of winning is the same (1/3) with both strategies. This answer is wrong.
(b) What is your probability of winning if there are N(>3) doors?

Note: This is called the Monty Hall problem and you will find long discussions on the web.
(a) Your best strategy is to switch: winning probability 2/3, vs 1/3 if you stick. There are many ways of seeing this. Here's one:
Initially you choose one door. Call that door A and the others B and C.
I will put a circle ◯ around the door behind which is the prize, so the three possible situations are:

(1/3) Ⓐ B C

(1/3) A Ⓑ C

(1/3) A B Ⓒ

At this point the probability of each of these possibilities is 1/3, as indicated at the left. The host always opens a door which does NOT have the prize behind it. Thus, in the first situation, there is a probability ½ for door B or door C. By contrast, in the second and third situations, the host has only a single choice. I will indicate the door opened by the host by putting a square □ around it. The outcomes after the host has chosen are listed below, each with its own probability:

1/3 Ⓐ B C → Ⓐ $\boxed{B}$ C (1/6) or Ⓐ B $\boxed{C}$ (1/6) $\boxed{\text{sticking strategy wins: p=1/3}}$

1/3 A Ⓑ C → A Ⓑ $\boxed{C}$ (1/3) switching strategy wins

1/3 A B C → A $\boxed{B}$ Ⓒ (1/3) switching strategy wins: p=1/3+1/3=2/3

The sticking strategy only wins if the initial guess was correct, which it is 1/3 of the time.
By contrast, the switching strategy is sure to win in the second and third cases, i.e., 2/3 of the time.

(b) Let me do N=4 for an example. I use the same notation for this case: The scenarios are:

(1/4) Ⓐ B C D → Ⓐ $\boxed{B}$ C D or Ⓐ B $\boxed{C}$ D or Ⓐ B C $\boxed{D}$ sticking wins: p=1/4

(1/4) A Ⓑ C D → A Ⓑ $\boxed{C}$ D or A Ⓑ C $\boxed{D}$ switching wins half the time

(1/4) A B Ⓒ D → A $\boxed{B}$ Ⓒ D or A B Ⓒ $\boxed{D}$ switching wins half the time

(1/4) A B C Ⓓ → A $\boxed{B}$ C Ⓓ or A B $\boxed{C}$ Ⓓ switching wins half the time

For N=4, the probability of winning via the stick strategy is p=¼ and the probability of winning by the switch strategy is p=¾ x ½ = 3/8. Note that switching is less of an advantage than for N=3.

For the general case, we have:

Sticking wins with probability $\boxed{p=1/N.}$

Switching wins with probability $\boxed{p=\frac{N-1}{N}\cdot\frac{1}{N-2}}$, which is always better than sticking, although the advantage decreases as N increases.

The first factor is the probability that the prize will be behind a door other than A. The second factor is the probability in each of these cases that the one door you pick out of the remaining N-2 (omitting A and the door opened by the host) is the right one.

9. This is a grab-bag of smaller problems:

(a) (3 points) Testing the Stirling formula. Calculate 10! exactly and compare with the Stirling approximation (without 1/n corrections). What is the absolute error? What is the fractional error? Show that the fractional error is well approximated by the 1/12n correction, i.e., show that the error NOT accounted for by 1/12n is of order $1/n^2$.

(a) I find 10!=3,628,800 while Stirling gives $\sqrt{20\pi}\cdot 10^{10}\cdot e^{-10}=3{,}598{,}695.618\ldots$ Thus, Stirling is too small by 30,104.381…, an absolute error that is numerically LARGE on the scale of unity. On the other hand, the fractional error is this divided by 10!, which gives 0.00829596… Compare this to the Stirling-Formula first correction (absolute) of $\sqrt{20\pi}\cdot 10^{10}\cdot e^{-10}\left(\frac{1}{12\cdot 10}\right)=29{,}989.130$ or 0.00826420 (fractional). The remaining mistake (Stirling is still a little low) is 115.251…(absolute) or $3.176\text{x}10^{-5}$ (fractional). Note that $1/n^2$ here is 0.01, so that the remaining error is comfortably smaller than this.

Note: The next fractional correction to Stirling has the form $\frac{1}{288n^2}=\frac{1}{28{,}800}=3.472\times 10^{-5}$, which captures most of the remaining error but actually over-compensates. The n^{-3} correction is negative. According to my recollection the Stirling series is actually asymptotic rather than convergent (not surprising, since we left out exponentially small tails). This means that successive corrections improve the agreement only up to a finite number of terms (~n). Terms after that make the agreement worse.

(b) (4 points) Consider a continuous probability distribution with the form of a one-sided Gaussian: $p(x)=\begin{cases} Ce^{-\frac{x^2}{2a^2}}, x>0 \\ 0, x<0 \end{cases}$. What is the normalization factor C? Calculate <x>, <x^2>, and σ^2 for this distribution.

(b) I use the integrals from Lecture 8:

$\int_0^\infty dx\, e^{-\frac{x^2}{2a}} = I_0\left(\frac{1}{2a^2}\right) = \sqrt{\frac{\pi a^2}{2}}$, so $\boxed{C = \frac{1}{a}\sqrt{\frac{2}{\pi}}.}$

Similarly, $\int_0^\infty dx\, x e^{-\frac{x^2}{2a^2}} = I_1\left(\frac{1}{2a^2}\right) = a^2$, so $\boxed{\langle x\rangle = Ca^2 = a\sqrt{\frac{2}{\pi}}}$.

and $\int_0^\infty dx\, x^2 e^{-\frac{x^2}{2a^2}} = I_2\left(\frac{1}{2a^2}\right) = \frac{\sqrt{\pi}}{4}\left(2a^2\right)^{3/2} = a^3\sqrt{\frac{\pi}{2}}$, so $\boxed{\left\langle x^2\right\rangle = a^2}$, and, finally,

$\boxed{\sigma^2 = \left\langle x^2\right\rangle - \langle x\rangle^2 = a^2\left(1-\frac{2}{\pi}\right)}$.

(c) (5 points) Certain molecular motors walk on cellular filaments. The walking has a fixed step length a but it is not unidirectional. It has at each step a probability p of going forward one step, a probability q of going backwards one step, and a probability r of falling off the "track" (p+q+r=1). Calculate the mean distance <X>=a<M> traveled along the track before falling off.

Hints:

1. The result turns out to be $\langle X\rangle = a\frac{(p-q)}{r}$.
2. In one way to do this problem, the result of problem 10 (a) is useful.

(c) Prob. of staying on track for N steps and falling off at the N+1st: $P_N = (1-r)^N r$

Note that $\sum_{N=0}^{\infty} P_N = r\sum_{N=0}^{\infty}(1-r)^N = \frac{r}{\left[1-(1-r)\right]} = 1$.

Mean distance traveled in N (consecutive) on-track steps with relative probabilities $\frac{p}{p+q}$ of going to the right and $\frac{q}{p+q}$ of going to the right is $\langle M\rangle_N = N\left(\frac{p-q}{p+q}\right)$ (see Problem 10 (a)).

Thus, the mean of M over all walks is

$$\langle M\rangle = \sum_{N=1}^{\infty} P_N \langle M\rangle_N = r\left(\frac{p-q}{p+q}\right)\sum_{N=1}^{\infty} N(1-r)^N = r(p-q)\sum_{N=1}^{\infty} N(1-r)^{N-1},$$

where I have used 1-r=p+q.

Note that the sum is of the form $\sum_{N=0}^{\infty} N x^{N-1} = \frac{d}{dx}\sum_{N=0}^{\infty} x^N = \frac{d}{dx}\left(\frac{1}{1-x}\right) = \frac{1}{(1-x)^2}$.

But, x=1-r, so 1-x=r and $\langle M\rangle = r(p-q)\frac{1}{r^2} = \frac{p-q}{r}$ and $\boxed{\langle X\rangle = \frac{a(p-q)}{r}.}$

Note: You may be tempted to do this problem by averaging (as in the random walk), $\langle M\rangle = \left\langle \sum_{N=1}^{\infty} \sigma_N \right\rangle = \sum_{N=1}^{\infty} \langle \sigma_N \rangle$. The problem can be done this way-not, however, easily. The problem is that $\langle \sigma_M \rangle \neq \langle \sigma_N \rangle$ because we are counting the contribution of $\langle \sigma_N \rangle$ ONLY if the motor has survived on the track for at least N steps. In this sense, the σ's are no longer statistically independent variables.

10. In class, I analysed the random-walk/coin-toss problem for an unbiased coin $p_+ = p_- = 1/2$. In this problem, I want you to do the same analysis for a biased coin $p_+ = p$, $p_- = q$, with $p + q = 1$. Consider a walk of N steps. Proceed as in class:

(a) (2 points) Show by direct calculation that $\langle M\rangle = N(p-q)$,, and $\sigma^2 \equiv 4Npq$.

$\langle M\rangle = \left\langle \sum_{n=1}^{N} \sigma_n \right\rangle = \sum_{n=1}^{N} \langle \sigma_n \rangle = N\langle \sigma \rangle$, since the σ's are statistically independent. But, $\langle \sigma \rangle = p(+1) + q(-1) = p - q$, from which the stated result follows.

Similarly, $\left\langle M^2 \right\rangle = N\left\langle \sigma^2 \right\rangle + N(N-1)\langle \sigma \rangle^2 = N + N(N-1)(p-q)^2$ and

$\sigma^2 \equiv \left\langle M^2 \right\rangle - \langle M \rangle^2 = N\left[1-(p-q)^2\right] = \boxed{4pqN\,.}$

(b) (2 points) Show that the discrete probability distribution is

$$P_{N,M} = p^{N_+} q^{N_-} \frac{N!}{N_+!N_-!}$$

The only difference from the previous case (in class) is that each *micro*state with N_+, N_- has probability $p^{N_+} q^{N_-}$. The combinatorial factor simply counts the number of distinct microstates, as before.

(c) (4 points) Show that for N>>1 and $M - \langle M \rangle \sim \sqrt{N}$ the discrete probability distribution is given approximately (but to and excellent approximation for N>>1) by the (discrete) Gaussian distribution, $P_{N,M} \approx \frac{1}{\sqrt{2\pi Npq}} e^{-\frac{(M-\langle M\rangle)^2}{8Npq}}$ with $\langle M \rangle$ as above.

Hints: This is the hard part of the problem. You will need to show that

$$P_{N,M} = \sqrt{\frac{N}{2\pi N_+ N_-}}\, e^{F_N(M)} \text{ with } F_N(M) = N_+ \ln\left(\frac{Np}{N_+}\right) + N_- \ln\left(\frac{Nq}{N_-}\right).$$

Then, carefully find the maximum of $F_N(M)$ and expand around it. The maximum is NOT generally at M=0.

The arithmetic here parallels the discussion from Lecture 9, so I will show the algebra without detailed explanations.
Apply the Stirling formula:

$$P_{N,M} = p^{N_+} q^{N_-} \frac{N!}{N_+!N_-!} = \frac{\sqrt{2\pi N}}{\sqrt{2\pi N_+}\sqrt{2\pi N_-}} \cdot \frac{e^{-N}}{e^{-N_+}e^{-N_-}} \cdot \frac{(pN)^{N_+}(qN)^{N_-}}{N_+^{N_+} N_-^{N_-}} \left(1 + O\left(\frac{1}{N}, \frac{1}{N_+}, \frac{1}{N_-}\right)\right)$$

$$\approx \sqrt{\frac{N}{2\pi N_+ N_-}} \cdot \left(\frac{pN}{N_+}\right)^{N_+} \left(\frac{qN}{N_-}\right)^{N_-} = \sqrt{\frac{N}{2\pi N_+ N_-}} \cdot e^{F_N(M)},$$

with $F_N(M)$ as given. The effect of the prefactor on the maximum of the exponential is negligible for large N (why?), so we look for the maximum of F at fixed N.

$$\frac{\partial F}{\partial M} = \frac{\partial F}{\partial N_+}\frac{\partial N_+}{\partial M} + \frac{\partial F}{\partial N_-}\frac{\partial N_-}{\partial M} = \frac{1}{2}\left(\frac{\partial F}{\partial N_+} - \frac{\partial F}{\partial N_-}\right),$$ since $N_\pm = \frac{1}{2}(N \pm M)$. Thus,

$$\frac{\partial F}{\partial M} = \frac{1}{2}\left(\ln\left(\frac{Np}{N_+}\right) - \ln\left(\frac{Nq}{N_-}\right)\right),$$ which vanishes when $\frac{p}{N_+} = \frac{q}{N_-}$, i.e., for

$N_+ = pN; N_- = qN; M = (p-q)N = \langle M \rangle$, as expected. Note that $F_N(\langle M \rangle) = 0$.

To expand around the maximum, we need further derivatives of F:

$$\frac{\partial^2 F}{\partial M^2} = \frac{\partial}{\partial N_+}\left(\frac{\partial F}{\partial M}\right)\frac{\partial N_+}{\partial M} + \frac{\partial}{\partial N_-}\left(\frac{\partial F}{\partial M}\right)\frac{\partial N_-}{\partial M} = -\frac{1}{4}\left(\frac{1}{N_+} + \frac{1}{N_-}\right) \xrightarrow[M=\langle M \rangle]{} -\frac{1}{4Npq}$$

$$\frac{\partial^3 F}{\partial M^3} = \frac{\partial}{\partial N_+}\left(\frac{\partial^2 F}{\partial M^2}\right)\frac{\partial N_+}{\partial M} + \frac{\partial}{\partial N_-}\left(\frac{\partial^2 F}{\partial M_2}\right)\frac{\partial N_-}{\partial M} = -\frac{1}{8}\left(\frac{1}{N_+^2} - \frac{1}{N_-^2}\right) \xrightarrow[M=\langle M \rangle]{} -\frac{1}{8N^2}\left(\frac{1}{p^2} - \frac{1}{q^2}\right)$$

and so forth. You can see that $\left.\frac{\partial^n F}{\partial M^n}\right|_{M=\langle M \rangle} \sim \frac{1}{N^{n+1}}$.

Thus, $e^{F_N(M)} = e^{-\frac{(M-\langle M \rangle)^2}{8Npq}} \cdot e^{-\frac{(M-\langle M \rangle)^3}{8\cdot 3! N^2\left(p^{-2}-q^{-2}\right)}} \cdot \ldots$

As for the case done in class, the first factor limits the difference $|M - \langle M \rangle|$ to $O(\sqrt{N})$, so the exponents of subsequent factors decrease with powers of $1/\sqrt{N}$, giving corrections which are negligibly small for large N.

Finally, because the exponential factor is sharply peaked, the factor $N_+(M)N_-(M)$ in the prefactor can be expanded around M=<M>, giving

$$N_+(M)N_-(M) = \left[N_+(\langle M \rangle) + \frac{1}{2}(M - \langle M \rangle)\right]\left[N_-(\langle M \rangle) - \frac{1}{2}(M - \langle M \rangle)\right]$$

$$= pqN^2\left[1 + \frac{1}{2pN}(M - \langle M \rangle)\right]\left[1 - \frac{1}{2qN}(M - \langle M \rangle)\right] = pqN^2\left(1 + O\left(\frac{1}{N}\right)\right)$$

The 1/N corrections are negligible, and we arrive at the result quoted.

(d) (4 points) Now, reinterpret the problem as a random walk with step-length a and time-step τ. Without taking any limits, show that the continuum probability distribution can be written as $P(X,t)=\frac{1}{\sqrt{4\pi Dt}}e^{-\frac{(X-vt)^2}{4Dt}}$, where $v=\frac{a(p-q)}{\tau}$ is a drift velocity and $D=\frac{2pqa^2}{\tau}$ is a diffusion coefficient.
This time-dependent probability distribution represents a particle starting at the origin at time t=0 and subject to a combination of drift and diffusion.

In this interpretation, $X=aM$ and $t=N\tau$, so $M\to\frac{X}{a};N\to\frac{t}{\tau}$,

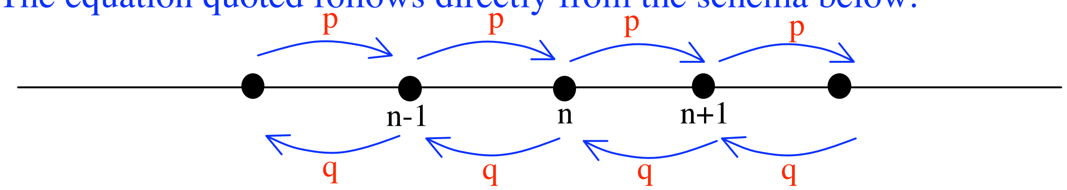
making the exponent $-\frac{(M-\langle M\rangle)^2}{8Npq}\to-\frac{\left(X-a(p-q)\frac{t}{\tau}\right)^2}{8a^2pq\frac{t}{\tau}}=-\frac{(X-vt)^2}{4Dt}$

and the prefactor $\frac{1}{\sqrt{2\pi Npq}}\to\frac{1}{\sqrt{2\pi pq\frac{t}{\tau}\cdot\frac{a^2}{a^2}}}=\frac{a}{\sqrt{\pi Dt}}$.

But, this is still a *discrete* distribution with all-even or all-odd values of M. To transform to a continuous distribution, we want the discrete weight associated with $M\to M+2$ to be associated with the continuous weight of $X\to X+2a$. Thus, $P_{N,M}=P(X)dX$ with $dX=2a$; hence, $P(X)=\frac{1}{2a}P_{N,M}$. The final form of the continuous distribution is $P(X,t)=\frac{1}{2a}\cdot\frac{a}{\sqrt{\pi Dt}}\cdot e^{-\frac{(X-vt)}{4Dt}}$, which gives the result quoted.
Note that it is properly normalized!

(e) (2 points) Consider a lattice model as done in class for the case without drift. Show that with the *biased* walk the discrete equation becomes
$P_n(t+\tau)-P_n(t)=p\left(P_{n-1}(t)-P_n(t)\right)+q\left(P_{n+1}(t)-P_n(t)\right)$.
Now, take an appropriate continuous limit and show that
$\frac{\partial P}{\partial t}=-v\frac{\partial P}{\partial X}+D\frac{\partial^2 P}{\partial X^2}$, which is sometimes called the Smoluchowski equation (for a linear potental).

The equation quoted follows directly from the schema below:



As in class, we take the continuum limit with:

$P_n(t) \to P(x = na, t)$, so

$$P((n \pm 1)a, t) = P(na, t) + a\left.\frac{\partial P}{\partial x}\right|_{x=na} + \frac{a^2}{2}\left.\frac{\partial^2 P}{\partial x^2}\right|_{x=na} + O\left(a^3\right)$$

$$P(na, t+\tau) = P(na, t) + \tau\left.\frac{\partial P}{\partial t}\right|_{t} + O\left(\tau^2\right).$$

Substituting these:

$$\tau\frac{\partial P}{\partial t} + O\left(\tau^2\right) = -pa\frac{\partial P}{\partial x} + \frac{pa^2}{2}\frac{\partial^2 P}{\partial x^2} + qa\frac{\partial P}{\partial x} + \frac{qa^2}{2}\frac{\partial^2 P}{\partial x^2} + O\left(a^3\right).$$ Divide by τ, so

$$\frac{\partial P}{\partial t} + O(\tau) = -\frac{a(p-q)}{\tau}\frac{\partial P}{\partial x} + (p+q)\frac{a^2}{2\tau}\frac{\partial^2 P}{\partial x^2} + O\left(\frac{a^3}{\tau}\right).$$

Note that p+q=1.

Now take $a \to 0, \tau \to 0$ but with $\frac{a^2}{2\tau} \to D$. This makes the correction terms vanish.

The only difficulty is the drift term: If $p \neq q$, then $\frac{a}{\tau} = \frac{2a}{2a}\cdot\frac{a}{\tau} = \frac{2D}{a} \to \infty$, so the drift velocity diverges in the limit. The solution is to take $p, q \to \frac{1}{2}, so\ (p-q) \to 0$ with $\frac{a(p-q)}{\tau} \to v$.

(f) (2 points) Show that the diffusive drift distribution derived in (d) satisfies the Smoluchowski equation.

Calculate:

$$\frac{\partial P}{\partial t} = -\frac{1}{2t}P(X,t) + \left[\frac{v(X-vt)}{2Dt} + \frac{(X-vt)^2}{4Dt^2}\right]P(X.t)$$

$$\frac{\partial P}{\partial X} = -\left(\frac{X-vt}{2Dt}\right)P(X,t),$$

$$\frac{\partial^2 P}{\partial X^2} = \left[-\frac{1}{2Dt} + \frac{(X-vt)^2}{4D^2t^2}\right]P(X,t),$$

and the algebra shows that Smoluchowski is satisfied.