Random Walk, Normal Distribution, and Brownian Motion; Linear Regression and the Method of Least Squares

Please attach a cover page (-10%).

Problem 1 (Extra Credit 100%): From Random Walk to Diffusion

From the binomial probability distribution for a "random walking" particle starting from m=0 to migrate to length m after n steps:

$$B(n,k,p) = \frac{n!}{[\frac{1}{2}(n+m)]![\frac{1}{2}(n-m)]!}\left(\frac{1}{2}\right)^n$$

$$k = \tfrac{1}{2}(n+m)$$

$$p = \tfrac{1}{2}$$

derive the normal probability distribution for a "random walking" particle starting from m=0 to migrate to length m after n steps:

$$N(m,m_0,\sigma^2) = \frac{2}{\sqrt{2\pi n}}\exp\left(\frac{-m^2}{2n}\right)$$

$$m_0 = 0$$

$$\sigma^2 = n$$

by using the following approximations for n much larger than m:

a) Stirling's approximation for large n:

$$\ln n! \approx n\ln n - n + \tfrac{1}{2}\ln(2\pi n)$$

b) Taylor's expansion for x much smaller than 1:

$$\ln(1 \pm x) \approx \pm x - \tfrac{1}{2}x^2$$

Hints:

First show that with just the first two terms in Stirling's approximation (a), that is with

c) $\ln n! \approx n\ln n - n$,

and by using Taylor's expansion in (b) in full to write

d) $\ln\left(\frac{n \pm m}{2}\right) = \ln\left(\frac{n}{2}\right) + \ln\left(1 \pm \frac{m}{n}\right) \approx \ln\left(\frac{n}{2}\right) \pm \frac{m}{n} - \frac{1}{2}\left(\frac{m}{n}\right)^2$

the natural logarithm of the binomial probability distribution is approximately

e) $\ln B(n,k,p) \approx \frac{-m^2}{2n}$

Second, show that adding the third term in Stirling's approximation (a), that is by adding $\frac{1}{2}\ln(2\pi n)$ to the approximation of $\ln n!$, and by using only the first term in Taylor's expansion (b), that is by re-writing (d) as follows,

f) $\ln\left(\frac{n \pm m}{2}\right) = \ln\left(\frac{n}{2}\right) + \ln\left(1 \pm \frac{m}{n}\right) \approx \ln\left(\frac{n}{2}\right) \pm \frac{m}{n}$,

the constant $-\frac{1}{2}\ln\left(\frac{2\pi n}{4}\right)$ is added to the natural logarithm $\ln B(n,k,p)$.

```
(* BIOEN 3070/6070: Introduction to Statistics for Bioengineers *)
```



```
(* Assignment 6: Random Walk, Normal Distribution, and Brownian Motion;
Linear Regression and the Method of Least Squares *)
```

```
(* General Commands *)

Clear["Global`*"]
```

```
(* Perrin, Ann Chim Phys (1909), p. 65 *)
```

```
(* Problem 2 (25% Extra Credit):
 Explain the code below that calculates the columns of expected numbers
  for the first and second series of observations. What is the theoretical
  probability distribution that the expected numbers are calculated from? *)

(* Problem 3 (25% Extra Credit): Compare the expected numbers with the observations
  for the first and the second series separately. Use two statistical
  approaches for the comparison. Explain the results of your comparisons. *)

(* First Series *)

observed = {38, 44, 33, 33, 35, 11, 14, 6, 5, 2};
dimension = Dimensions[observed][[1]];
variance = N[Sum[((a - 0.5))^2 * observed[[a]], {a, 1, dimension}]] /
  Sum[observed[[a]], {a, 1, dimension}]
expected = Round[2 * Sum[observed[[a]], {a, 1, dimension}] *
   Table[PDF[NormalDistribution[0, Sqrt[variance]], (a - 0.5)], {a, 1, dimension}]]

14.621

{46, 43, 37, 30, 23, 16, 11, 7, 4, 2}

(* Second Series *)

observed = {48, 38, 36, 29, 16, 15, 8, 7, 4, 4};
dimension = Dimensions[observed][[1]];
variance = N[Sum[((a - 0.5))^2 * observed[[a]], {a, 1, dimension}]] /
  Sum[observed[[a]], {a, 1, dimension}]
expected = Round[2 * Sum[observed[[a]], {a, 1, dimension}] *
   Table[PDF[NormalDistribution[0, Sqrt[variance]], (a - 0.5)], {a, 1, dimension}]]

13.8402

{44, 41, 35, 28, 21, 15, 10, 6, 3, 2}
```

```
(* Kepler's Third Law of Planetary Motion *)

(* Problem 4 (25% Extra Credit):
 Calculate the best linear fit between the square of a planet's period and the cube
  of the planet's mean distance from the sun. Evaluate the fit and explain. *)

(* Problem 5 (25% Extra Credit): Calculate the best linear fit between a planet's
  period and its mean distance from the sun. Explain and compare to Problem 5. *)

stream =
  "http://www.alterlab.org/teaching/BIOEN3070/assignments/Kepler_Third_Law.txt";
data = Import[stream, "Table"];
TableForm[data]
```

Planet	Period_(Years)	Mean_Distance_(Astronomical_Units)
Mercury	0.241	0.39
Venus	0.615	0.72
Earth	1	1
Mars	1.88	1.52
Jupiter	11.8	5.2
Saturn	29.5	9.54
Uranus	84	19.18
Neptune	165	30.06
Pluto	248	39.44