Stat Mech Assignment 2

Debsubhra Chakraborty

April 2022

Contents

1 Problem 1: Conditional Probability

1.1 a

The sample space of the different possible combinations of two children of Mr. Biswas is given by this: $\{BB, BG, GB, GG\}$. Now after stating the fact that, one of his children is a girl, the samle space turns into, $\{BG, GB, GG\}$. Now the event space of the other child to be a boy is given by, $\{BG, GB\}$. So, the probability of his other child to be boy is given by the ratio of the cardinality of the event space and sample space. Thus the probability would be $\frac{2}{3}$.

1.2 b

If Mr. Biswas said that "This is my younger child Aditi.", then the probability of other child to be a boy changes. Because in this case the sample space becomes $\{BG, GG\}$. Here we are assuming that the first letter in each entry stands for the older child and the second letter stand for the younger child. Then the event space becomes $\{BG\}$, So, the probability of the other child to be boy becomes $\frac{1}{2}$.

2 Problem 2: The Three Door Problem

Without any prior knowledge the probability of having heaven behind the chosen door is $\frac{1}{3}$. So, the probability of the heaven being behind any one of the two door that hasn't been chosen is $\frac{2}{3}$. Now, when Saint Peter opened one of the unchosen door and showed that there is hell behind that door. Then the probability of finding heaven behind the door that is unchosen and unopened is increased to $\frac{2}{3}$, because the having the heaven behind the two of the unchosen doors still has to be $\frac{2}{3}$. So, now if Thomas Bayes switch his choice, his chance of getting heaven will be $\frac{2}{3}$. So, he should switch to other closed door.
We can simply support our statement above by simply counting the number of element in the sample space and event space. The elements in the event space without any prior knowledge is $\{H_l H_V H_l, H_l H_l H_v, H_v H_l H_l\}$. Here H_l stand for hell and H_v stands for heaven. The first entry of each element of this set stand for the chosen door by Thomas Bayes and the later 2 stands for the unchosen two. Now the probability of finding the heaven behind any door is equal as $\frac{1}{3}$. But when it is revealed that one of the unchosen door hell behind it, the sample space turns into, $\{H_l H_v, H_l H_v, H_v H_l\}$. Here the the first entry of each element stands for the chosen door and the second entry stands for the rest one. So, we can simply see, the probability of finding the heaven behind the unchosen door is $\frac{2}{3}$. So, to increase the chances of going to heaven, Thomas Bayes should switch his choice.

3 Problem 3: Combinatoric

3.1 a

The probability that the ith card appeared in the same place both the the times is given by, $\frac{1}{52}$. If we call this event E_i, then $P(E_i) = \frac{1}{52}$.

3.2 b

The probability that the ith and the jth card appeared at the same places is equal to the number of instances they appear at the same places /the total number of combinations.
The total number of combination is $52! \times 52!$. The number of instances they appear at the same places are $2! \times \binom{52}{2} \times (50!)^2$. So, the probability would be equal to, $P(E_i \cap E_j) = \frac{1}{52 \cdot 51}$

3.3 c

The probability $P(E_{i_1} \cap E_{i_2} \cap \cdots E_{i_k}) = \frac{(52-k)!}{52!} \quad for\ k \leq n = 52$

3.4 d

we are finally interested to calculate at the probability that at least one of the cards appears at the same place both times. It is simply given by $P(E_1 \cup E_2 \cup \cdots E_n)$.

$$\begin{aligned} P(E_1 \cup E_2 \cup \cdots E_n) &= \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \leq i_1 < i_2 \cdots < i_k \leq n} P(E_{i_1} \cap E_{i_2} \cap \cdots E_{i_k}) \\ &= 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots + (-1)^{n+1} \frac{1}{n!} \\ &= 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots - \frac{1}{52!} \\ &= 0.63212 \end{aligned} \tag{1}$$

3.5 e

For $n \to \infty$, the asymptotic value of the probability is given by, $1 - \frac{1}{e}$.

4 Problem 4: Bayesian Statistics

According to the Bayes Theorem,

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{2}$$

Here $P(A|B)$ is the probability of event A given that event B has happened. For this problem at hand, we can take the event A as being truly positive and the event B as having a positive test result.
Then the probability $P(A) = \frac{1}{20}$ and the probability $P(B) = \frac{1}{20}\frac{72}{100} + \frac{19}{20}\frac{2}{100} = \frac{110}{2000}$. The probability $P(B|A) = \frac{72}{100}$. Then according to the Bayes theorem, we can write,

$$P(A|B) = \frac{72}{110} = \frac{36}{55} \tag{3}$$

So, the probability that someone who got a positive test result is truely infected is given by, $\frac{36}{55}$.

5 Problem 5: Limiting Distribution

Let $\{x_1, \cdots, x_n\}$ be a set of iid random variables taking values $x = \{0, 1\}$ drawn from a Bernoulli distribution,

$$x = \begin{cases} 1, & \textit{with Prob } p \\ 0, & \textit{with Prob } (1-p) \end{cases}$$

From the sequence of x_i's we can construct another set of random variables,

$$Y_i = x_i x_{i+1} \tag{4}$$

5.1 a

The distribution of Y_i is easy to calculate. Y_i will take value 1, when both the x_i and x_{i+1} are 1. The probability of this is simply p^2. So, the probability distribution of Y_i would be as follows,

$$Y_i = \begin{cases} 1, & \textit{with Prob } p^2 \\ 0, & \textit{with Prob } (1-p^2) \end{cases} \tag{5}$$

5.2 b

Now we would like to calculate the variance. From the construction of Y_i variables, it is clear that covariance would be non-zero for $\langle Y_i Y_{i+1}\rangle_c$. Now we would like to calculate it.

$$\begin{aligned} \langle Y_i Y_{i+1}\rangle_c &= \langle Y_i Y_{i+1}\rangle\langle Y_i\rangle\langle Y_{i+1}\rangle \\ &= \langle x_i x_{i+1}^2 x_{i+2}\rangle - \langle x_i x_{i+1}\rangle\langle x_{i+1}x_{i+2}\rangle \\ &= \langle x_i\rangle\langle x_{i+1}^2\rangle\langle x_{i+2}\rangle - \langle x_i\rangle\langle x_{i+1}\rangle\langle x_{i+1}\rangle\langle x_{i+2}\rangle \\ &= p^3 - p^4 \end{aligned} \tag{6}$$

Now we can calculate the variance $\langle Y_i^2\rangle_c = \langle Y_i^2\rangle - \langle Y_i\rangle^2 = \langle x_i^2 x_{i+1}^2\rangle - \langle x_i x_{i+1}\rangle^2 = p^2(1-p^2)$.So, we can write,

$$\langle Y_i Y_{i+j}\rangle_c = p^2(1-p^2)\delta_{j,0} + p^3(1-p)(\delta_{j,1} + \delta_{j,-1}) \tag{7}$$

5.3 c

Now we would like to find out the asymptotic distribution for S_n for large n, where $S_n = \frac{1}{2}\sum\limits_{i=1}^{n} Y_i$. As we can see the variable S_n is scaled as 1 over n, the using CLT we cannot get to it, because central limit theorem tells about the fluctuation of the order of $\sqrt{n}$ and here we are interested in analysing the fluctuations of the scale N. So, for that reason we will proceed in the following manner. We will first separate the $\{Y_1, \cdots, Y_n\}$ into two groups. The first group contains all the element of the odd positions like, $\{Y_1, Y_3, Y_5, \cdots\}$ and the second group contains all the element of the even positions $\{Y_2, Y_4, Y_6, \cdots\}$. Now we are interested in the asympotic distribution of two quantities, $m1 = \frac{Y_1+Y_3+Y_5+\cdots}{N_1}$ and $m2 = \frac{Y_2+Y_4+Y_6+\cdots}{N_2}$. Now for large n, we take safely take $N_1 = N_2 = \frac{n}{2}$ irrespective of whether n is even or odd. Now, as we can clearly see all the elements of each group are uncorrelated among themselves because we have removed the immediate neighbours of each elements. Now focusing on any one group, we can take the variables to be iid Bernoulli Distribution. Now, for that case, the asymptotic form is well known and derived in the class room. We will simply quote the result here,

$$P(m_1) = \frac{1}{\sqrt{\pi n(1-m_1^2)}} e^{-n\phi(m_1)/2} \tag{8}$$

$$P(m_2) = \frac{1}{\sqrt{\pi n(1-m_2^2)}} e^{-n\phi(m_2)/2} \tag{9}$$

Here $\phi(m_1) = \frac{1+m_1}{2} log \left[\frac{1+m}{2p^2}\right] + \frac{1-m_1}{2} log \left[\frac{1-m_1}{2}\frac{1}{1-p^2}\right]$. Now what we need is to find the probability distribution of $m = \frac{m_1+m_2}{2}$. This is simply given by,

$$\tilde{p}(m) = \iint dm_1 dm_2 P(m_1, m_2)\delta\left(m - \frac{m_1+m_2}{2}\right) \tag{10}$$

Now the thing is as m_1 and m_2 are correlated, we cannot write, $P(m_1, m_2) = P(m_1)P(m_2)$. But we can show that in the large n limit the correlation between m_1 and m_2 vanishes.

$$\begin{aligned} \langle m_1 m_2 \rangle_c &= \frac{4}{n^2}\langle (Y_1 + Y_3 + Y_5 + \cdots)(Y_2 + Y_4 + Y_6 + \cdots)\rangle_c \\ &\sim \frac{4}{n^2} \times 2(n/2)(p^3 - p^4) \\ &= \frac{4}{n}(p^3 - p^4) \end{aligned} \tag{11}$$

So, we can see that the covariance vanishes in the limit large n. This happens because the correlation is in the near neighbour. So, it always scales like n. and we are diving it with n^2. Similarly it is easy to see that all the cumulants of the form $\langle m_1^i m_2^j \rangle_c \to 0$ for large n for $i, j \geq 1$.So, in the large n limit we can take m_1 and m_2 to be indepedent and take $P(m_1, m_2) = P(m_1)P(m_2)$. So, we have,

$$\begin{aligned} \tilde{p}(m) &= \iint dm_1 dm_2 P(m_1)p(m_2)\delta\left(m - \frac{m_1+m_2}{2}\right) \\ &\sim \int dm_1 e^{-\frac{n}{2}(\phi(m_1)+\phi(2m-m_1))} \quad \textit{(Ignoring the factor sitting in front)} \end{aligned} \tag{12}$$

Now in the limit of large n, the lowest value of the exponent can be taken. For this we differentiate the exponent and find the m_1 value for which it vanishes.

$$\begin{aligned} \phi'(m_1) &= \phi'(2m - m_1) \\ m_1 &= m \end{aligned} \tag{13}$$

Putting this in the the integrand we find,

$$\tilde{p}(m) \sim e^{-n\phi(m)} \quad \textit{(Upto a Constant Factor)} \tag{14}$$

6 Problem 6: Ratio of Random Variables

Let $\{x_1, x_2, \cdots, x_{2n}\}$ be iid random variables with

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \tag{15}$$

We now construct a set of random variables $\{Y_1, Y_2, \cdots, Y_n\}$ by

$$Y_i = \frac{x_{2i-1}}{x_{2i}} \tag{16}$$

6.1 a

Let us assume that we have two random number x_1 and x_2 with the pdf $P(x)$ as in equation (6). Now we will like to find the probability distribution of the ratio $y = \frac{x_1}{x_2}$. Then it can be written in the following form,

$$\tilde{P}(y) = \int\int dx_1 dx_2 P(x_1)P(x_2)\delta\left(y - \frac{x_1}{x_2}\right) \tag{17}$$

The delta function only filter out the contributions from the probabilities when $y = \frac{x_1}{x_2}$ condition is satisfied. Now proceeding we find that,

$$\begin{aligned}
\tilde{P}(y) &= \int dx_2 |x_2| P(yx_2) P(x_2) \\
&= \frac{1}{2\pi\sigma^2} \int_{-\infty}^{\infty} dx_2 |x_2| \exp\left(-\frac{x_2^2(y^2+1)}{2\sigma^2}\right) \\
&= \frac{1}{\pi\sigma^2} \int_{0}^{\infty} dx_2 x_2 \exp\left(-\frac{x_2^2(y^2+1)}{2\sigma^2}\right) \\
&= -\frac{1}{\pi\sigma^2} \frac{\sigma^2}{y^2+1} \left(\exp\left(-\frac{x_2^2(y^2+1)}{2\sigma^2}\right)\right)\Bigg|_0^\infty \\
&= \frac{1}{\pi} \frac{1}{y^2+1} \qquad (18)
\end{aligned}$$

So, as we can see the probability distribution function of $y_i = \frac{x_{2i-1}}{x_{2i}}$ is Cauchy distribution.

6.2 b

Now we would like to calculate the covariance $\langle Y_i Y_{i+j} \rangle_c$ for $j > 0$,

$$\begin{aligned}
\langle Y_i Y_{i+j} \rangle_c &= \langle Y_i Y_{i+j} \rangle - \langle Y_i \rangle \langle Y_{i+j} \rangle \\
&= \left\langle \frac{X_{2i-1}}{X_{2i}} \frac{X_{2(i+j)-1}}{X_{2(i+j)}} \right\rangle - \left\langle \frac{X_{2i-1}}{X_{2i}} \right\rangle \left\langle \frac{X_{2(i+j)-1}}{X_{2(i+j)}} \right\rangle \\
&= \frac{\langle X_{2i-1} \rangle}{\langle X_{2i} \rangle} \frac{\langle X_{2(i+j)-1} \rangle}{\langle X_{2(i+j)} \rangle} - \frac{\langle X_{2i-1} \rangle}{\langle X_{2i} \rangle} \frac{\langle X_{2(i+j)-1} \rangle}{\langle X_{2(i+j)} \rangle} \quad (\textit{As } X_i \textit{'s are random variable with no correlation}) \\
&= 0 \qquad (19)
\end{aligned}$$

Thus we have show for $j > 0$, $\langle Y_i Y_{i+j} \rangle_c$ is zero. So, the Y_i's are iid random variables.

6.3 c

Now we will like to calculate the characteristic function of $P(Y)$. By the definition of the characteristic function we have proceed as follows,

$$\tilde{p}(k) = \langle \exp(-iky) \rangle = \int dy P(y) \exp(-iky) = \int_{-\infty}^{\infty} dy \frac{1}{\pi} \frac{1}{y^2+1} \exp(-iky) \qquad (20)$$

To do the integration we will use contour integration. So,

$$\oint_C dz \frac{1}{\pi} \frac{1}{z^2+1} \exp(-ikz) \qquad (21)$$

For $k > 0$, we chose the contour to be as shown in the figure (1). As we we know the integrand has a pole at $z = -i$ in the region of integration, we will calculate the residue,

$$\begin{aligned}
&\lim_{z \to -i} (z+i) \frac{1}{\pi} \frac{1}{z^2+1} \exp(-ikz) \\
&= \lim_{z \to -i} \frac{1}{\pi} \frac{1}{z-i} \exp(-ikz) \\
&= -\frac{1}{2\pi i} e^{-k} \qquad (22)
\end{aligned}$$

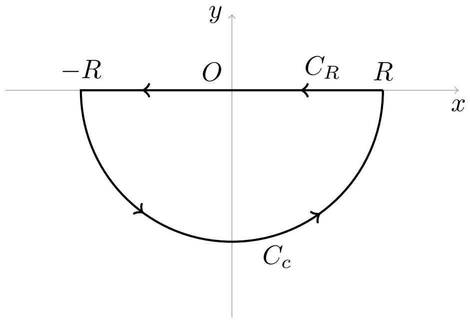


Figure 1: The Contour of Integration for $k > 0$

Now we simply use the residue theorem of the complex integration and find the result that,

$$\oint\limits_C dz \frac{1}{\pi} \frac{1}{z^2+1} \exp(-ikz) = -e^{-k} = \int_{+R}^{-R} dy \frac{1}{\pi} \frac{1}{y^2+1} \exp(-iky) + \int\limits_{C_c} dz \frac{1}{\pi} \frac{1}{z^2+1} \exp(-ikz) \quad (23)$$

Now, in the limit $R \to \infty$, the integral over the semi circular arc C_c vanishes for the exponentially decaying integrand. So, we are left with,

$$\int_{-\infty}^{\infty} dy \frac{1}{\pi} \frac{1}{y^2+1} \exp(-iky) = e^{-k} \quad \textit{for } k > 0 \quad (24)$$

Proceeding in the same way for $K < 0$, we can show that the result in that case is e^k. The contour used in this case is as shown in the figure (2). The residue of the pole at $Z = i$ is given by,

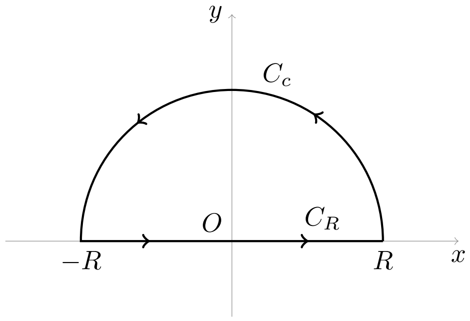


Figure 2: The Contour of Integration for $k < 0$

$$\begin{aligned} &\lim_{z \to i} (z-i) \frac{1}{\pi} \frac{1}{z^2+1} \exp(-ikz) \\ &= \lim_{z \to -i} \frac{1}{\pi} \frac{1}{z+i} \exp(-ikz) \\ &= \frac{1}{2\pi i} e^k \end{aligned} \quad (25)$$

So, again using the residue theorem and taking limit $R \to \infty$ we find that,

$$\int_{-\infty}^{\infty} dy \frac{1}{\pi} \frac{1}{y^2+1} \exp(-iky) = e^k \quad \textit{for } k < 0 \quad (26)$$

So, we can write the characteristic equation as , $\tilde{p}(k) = e^{-|k|}$.

6.4 d

The characteristic equation of $M_n = \sum_{i=1}^{n} Y_i$ can be written as follows,

$$\begin{aligned} \tilde{p}_n(k) &= \langle \exp(-ikM_n) \rangle \\ &= \langle \exp(-ikY) \rangle^n \\ &= e^{-n|k|} \end{aligned} \tag{27}$$

Having the characteristic function for M_n at hand we can easily calculate the pdf by calculating the Fourier transform. But in this case, it is obvious that the Fourier transform of this would yield the Cauchy distribution. Thus we have,

$$P(M_n) = \frac{n}{\pi} \frac{1}{M_n^2 + n^2} \tag{28}$$

The probability distribution function of $m = \frac{M_n}{n}$ is given by,

$$P(m) = \frac{1}{\pi} \frac{1}{m^2 + 1} \tag{29}$$

7 Problem 7: Random Variables on Computer

In different computer languages there are in built Pseudo random number generators that generates random variables in range $x \in [1, 0)$. From such variables we can generate different probability distributions, using proper transformations.

1. We want to make a Bernoulli Probability distribution, for that we can simply make a functional dependence of y on x which is a random variable with uniform distribution. Like, $y = 1 \;\; for \;\; x \leq p$ and $y = 0 \;\; for \;\; x > p$.

2. For the exponential pdf of a random variable y, we can easily work out the functional dependence of y on x.

$$\begin{aligned} p(x)dx &= \tilde{p}(y)dy \\ dx &= \lambda e^{-\lambda y} dy \\ x &= -e^{-\lambda y} + C \end{aligned} \tag{30}$$

   At $y = 0$, $x = 0$. So, we can that, $C = 1$. So from this we can work out,

$$y(x) = -\frac{1}{\lambda} log(1 - x) \tag{31}$$

   y will have the desired exponential distribution.

3. We want to make Gaussian distribution from the uniform distribution. For that we would like to go to two dimension. In 2 dimension the statement of the probability conservation looks like,

$$\begin{aligned} p(x_1, x_2)dx_1 dx_2 &= \tilde{p}(y_1, y_2)dy_1 dy_2 \\ \tilde{p}(y1, y2) &= \left| \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} \right| \end{aligned} \tag{32}$$

   Now, we consider $2\pi x_1$ to be the angle obtained from a uniform distribution x_1 and y_1 and y_2 as Cartesian coordinates that will have a Gaussian Distribution. The relation between them are,

$$\begin{aligned} y_1 &= \sqrt{-2lnx_1} cos 2\pi x_2 \\ y_2 &= \sqrt{-2lnx_1} sin 2\pi x_2 \end{aligned}$$

Then inverting them we can easily find,

$$x_1 = e^{-(y_1^2+y_2^2)/2} \quad (33)$$
$$x_2 = \frac{1}{2\pi} tan^{-1}\frac{y_2}{y_1} \quad (34)$$

Then from equation (32), we can write,

$$\tilde{p}(y1, y2) = \frac{1}{2\pi} e^{-(y_1^2+y_2^2)/2} \quad (35)$$

Now to get a Gaussian distribution with a variance of σ^2, we just have to multiply y_i with σ.

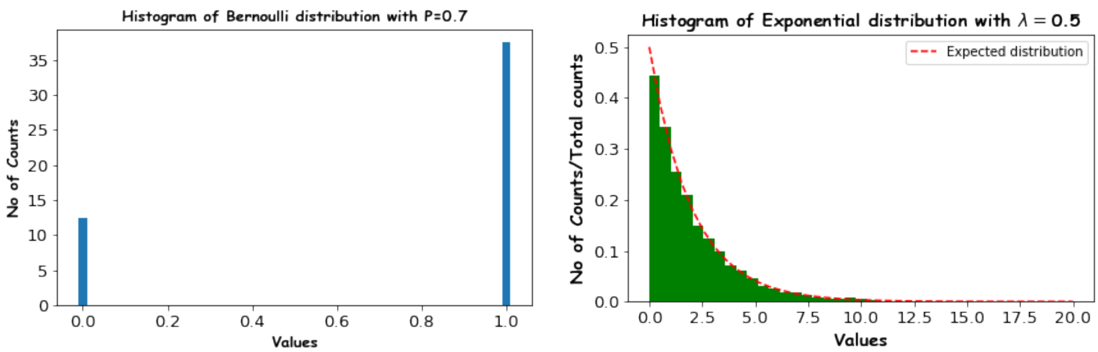
4. Now we would like to make a Cauchy distribution from the uniform distribution. For that, we can simply follow the same procedure for the case of exponential distribution.

$$\begin{aligned} P(x)dx &= \tilde{P}(y)dy \\ dx &= \frac{a}{\pi}\frac{1}{a^2+\pi^2}dy \\ x &= \frac{1}{\pi}tan^{-1}(y/a) + C \end{aligned}$$

Now, for $y = \infty$, $x = 1$. From that we can see that, $C = \frac{1}{2}$. So, the dependence of y on x would be,

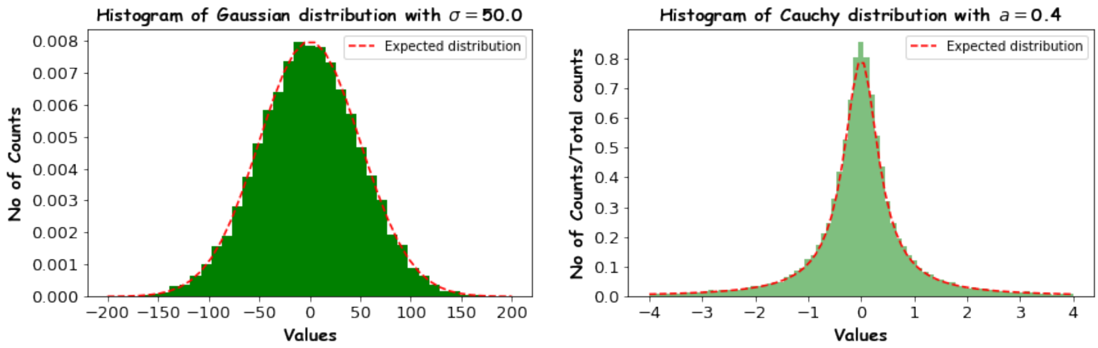
$$y(x) = atan(\pi(x - 1/2)) \quad (36)$$

All the codes are attached with the answer script. Please check them.



(a) Bernoulli Distribution

(b) Exponential Distribution



(a) Gaussian Distribution

(b) Cauchy Distribution

Figure 4: Histograms of Different Distribution Made from Uniform Distribution

8 Problem 8: Log Normal Distribution and Benford's Law

Let's consider iid random variables $x_i \geq 1$ drawn from an arbitrary distribution with finite mean and variance. We have to find the asymptotic probability distribution of $M_n = \prod_{i=1}^{n} x_i$.

8.1 a

Let us consider the random variable $A_n = ln(M_n) = \sum_{i=1}^{n} ln(x_i)$. It is the sum of the many random variable $y_i = ln(x_i)$, and assuming that the mth order cumulants of this random variables satisfies, $\sum_{i=1}^{n}\langle y_i^m\rangle \ll \mathcal{O}(n^{m/2})$ relation, from the central limit theorem, we can conclude that the the asymptotic probability distribution of A_n would be a Gaussian distribution with mean of $n\langle y_i\rangle$ and variance of $n\langle y_i^2\rangle_c = n\sigma^2$. Thus we can write, in the limit $n \to \infty$,

$$p(A_n) = \frac{1}{\sqrt{2\pi n\sigma^2}} \exp\left[-\frac{(A_n - n\bar{y})^2}{2n\sigma^2}\right] \tag{37}$$

Now we have the probability distribution function for $A_n = ln(M_n)$. So, it is very straight forward to calculate the PDF for the M_n variable.

$$\begin{aligned} p(A_n)dA_n &= \tilde{p}(M_n)dM_n \\ \tilde{p}(M_n) &= p(A_n)\frac{dM_n}{d_An} \\ &= \frac{1}{M_n}\frac{1}{\sqrt{2\pi n\sigma^2}} \exp\left[-\frac{(ln(M_n) - n\bar{y})^2}{2n\sigma^2}\right] \end{aligned} \tag{38}$$

So,equation (38) gives the asymptotic probability distribution of the M_n variable.

8.2 b

Let Z_n be the first digit of M_n when written in the decimal representation. Now lets say we have a digits before the decimal point and the first digit is Z_n. This condition will be satisfied if the value of M_n lies in the range of $Z_n \times 10^a$ to $(Z_n + 1) \times 10^a$. Now this same scenario is possible for all the decimal places. So, we have to sum over of all the a's. Doing this we will find out that,

$$Prob(Z_n) = \sum_a \int_{Z_n \times 10^a}^{(Z_n+1)\times 10^a} \tilde{p}(M_n)dM_n \tag{39}$$

Now, we make a change of variables, $A_n = ln(M_n)$. Thus the integrand now, looks like,

$$Prob(Z_n) = \sum_a \int_{ln(Z_n)+ka}^{ln(Z_n+1)+ka} p(A_n)dA_n \tag{40}$$

Here, $k = ln(10)$. Now, we do another approximation that the integrand is fairly constant in the range of integration. This approximation holds much well if, our $ka \sim n\bar{y}$, because there the distribution is much more slowly varying. Applying this approximation, we find that,

$$Prob(Z_n) \sim \sum_a \exp\left[-\frac{(ka - n\bar{y})^2}{2n\sigma^2}\right] \frac{1}{\sqrt{2\pi n\sigma^2}} (ln(z_n + 1) - ln(z_n)) \propto ln\left(1 + \frac{1}{Z_n}\right) \tag{41}$$

From the normalisation condition of the probability we can find the value of the constant. $\sum_{Z_n=1}^{9} c\, ln\left(1 + \frac{1}{Z_n}\right) = c\, ln 10 = 1$. This implies that,

$$Prob(Z_n) = \frac{1}{ln(10)} ln\left(1 + \frac{1}{Z_n}\right) = log_{10}\left(1 + \frac{1}{Z_n}\right) \tag{42}$$

8.3 c

The Benford's law predicts that the same probability distribution $Prob(Z_n)$ describes the distribution of the first digit of any large data set like stock price and rain fall data. This probability distribution is independent of the unit that has been used to measure the quantities. So, this kind of multiplication by any number can be done many times still getting same distribution of the first digits. In this problem also, we made a random variable M_n by the multiplication of a large number of random variable. But the final result that we obtained is independent of the probability distribution of the x_i's. So, we can see that the central limit theorem is playing a crucial role here. From the result (**??**, we can conclude that the probability of finding $1, 2, 3, \cdots, 9$ as the first digit of a large data set is given by $0.301, 0.176, 0.125, 0.097 \cdots, 0.046$ respectively.

9 Problem 9: Moments and The Probability Distribution Function

A modified log-normal distribution is given by,

$$P)x) = \frac{1}{\sqrt{2\pi}}\frac{1}{x}\exp\left(-\frac{(log\, x)^2}{2}\right)[1 + asin\,(2\pi logx)] \tag{43}$$

where $-1 \leq a \leq 1$.

9.1 a

The distribution is plotted in the figure (5) below for $a = 0$ and $a = 1/2$.

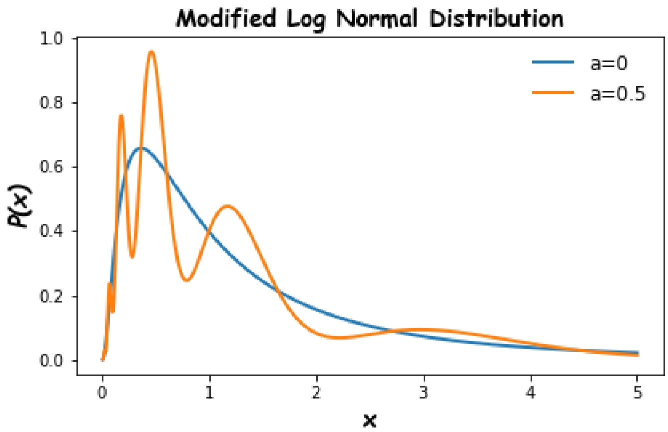


Figure 5: Modified Log Normal Distribution for $a = 0$ and $a = 0.5$

9.2 b,c

Now we want to calculate the moment of this distribution function. The expression for the kth moment would be as following,

$$\begin{aligned}\langle x^k\rangle &= \int_0^\infty x^k p(x)dx \\ &= \int_0^\infty x^k \frac{1}{\sqrt{2\pi}}\frac{1}{x}\exp\left(-\frac{(log\, x)^2}{2}\right)[1 + asin\,(2\pi logx)]\,dx \end{aligned} \tag{44}$$

Now, we will do a change of variables, $z = log\, x$. Doing this we obtain,

$$\langle x^k\rangle = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\exp kz \exp\,(-\frac{z^2}{2})\,[1 + asin\,(2\pi z)]\,dz \tag{45}$$

Let's integrate the first term first,

$$\begin{aligned}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\exp kz \exp\,(-\frac{z^2}{2})dz &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(z-k)^2}{2}\right)\exp\left(\frac{k^2}{2}\right)dz \\ &= \exp\left(\frac{k^2}{2}\right)\end{aligned} \tag{46}$$

Now we will come to the second term,

$$\begin{aligned}
&\int_{-\infty}^{\infty} \frac{a}{\sqrt{2\pi}} \exp{(kz)} \exp{(-\frac{z^2}{2})} sin\,(2\pi z)\,dz \\
= &\int_{-\infty}^{\infty} \frac{a}{\sqrt{2\pi}} \exp\left(\frac{k^2}{2}\right) \exp\left(-\frac{(z-k)^2}{2}\right) sin\,(2\pi z)\,dz
\end{aligned} \tag{47}$$

Now we will do a change of variable $z - k = y$. Then the integral will look like,

$$\int_{-\infty}^{\infty} \frac{a}{\sqrt{2\pi}} \exp\left(\frac{k^2}{2}\right) \exp\left(-\frac{y^2}{2}\right) sin\,(2\pi(y+k))\,dy$$

$$= \frac{a}{\sqrt{2\pi}} \exp\left(\frac{k^2}{2}\right) \int_{-\infty}^{\infty} \exp\left(-\frac{y^2}{2}\right) (sin(2\pi y)cos(2\pi k) + cos(2\pi y)sin(2\pi k))\,dy \tag{48}$$

$$= \frac{a}{\sqrt{2\pi}} \exp\left(\frac{k^2}{2}\right) \int_{-\infty}^{\infty} \exp\left(-\frac{y^2}{2}\right) cos(2\pi y)sin(2\pi k)dy \tag{49}$$

To reach equation (49) from equation (48), we have used the fact that the integrand is odd for the first term, so it gets dropped out. Now to proceed we define,

$$I(a) = \int_{-\infty}^{\infty} \exp\left(-\frac{y^2}{2}\right) cos(ay)dy \tag{50}$$

Now, differentiating the above expression under integral sign and using by part integration we find,

$$\begin{aligned}
I'(a) &= -\int_{-\infty}^{\infty} y \exp\left(-\frac{y^2}{2}\right) sin(ay)dy \\
&= \exp\left(-\frac{y^2}{2}\right) sin(ay)\Big|_{-\infty}^{\infty} - a\int_{-\infty}^{\infty} \exp\left(-\frac{y^2}{2}\right) cos(ay)dy \\
&= -aI(a)
\end{aligned} \tag{51}$$

. Solving this differential equation and using the fact that, $I(0) = \sqrt{2\pi}$, we find that,

$$I(a) = \sqrt{2\pi} \exp\left(-\frac{a^2}{2}\right) \tag{52}$$

So, the value of the kth moment of this distribution function becomes,

$$\langle x^k \rangle = \exp\left(\frac{k^2}{2}\right) \left(1 + asin(2\pi k) \exp\left(-2\pi^2\right)\right) \tag{53}$$

Now, as we are calculating the moments for integer k's, the second term vanishes. Thus the final answer is,

$$\langle x^k \rangle = \exp\left(\frac{k^2}{2}\right) \tag{54}$$

So, the first 3 moments are as follows,

$$\langle x \rangle = \sqrt{e} \tag{55}$$
$$\langle x^2 \rangle = e^2 \tag{56}$$
$$\langle x^3 \rangle = e^{\frac{9}{2}} \tag{57}$$

9.3 d

So, as we can see the moments are independent of the parameter a. So, knowing all the moments are not sufficient to construct the probability distribution function. It is because of the fact that the characteristic function of this probability distribution function does not depend on the second term. That is the fourier transform of the second term of the pdf vanishes.

$$\int_0^{infty} \exp{(ikx)} \frac{1}{\sqrt{2\pi}} \frac{1}{x} \exp\left(-\frac{(log\, x)^2}{2}\right) asin\,(2\pi logx)\, dx = 0 \tag{58}$$

So, that is why, this information gets lost in the characteristic function. So, inverse Fourier transform, cannot bring back the 2nd term of the pdf.

10 Problem 10: Random Matrices

Let $\mathcal{M} = (m_{ij})_{n\times n}$ is a real symmetric matrix with element chosen randomly from the distribution,

$$P(m_{ij}) = \frac{1}{\sqrt{\pi}} \exp{(-m_{ij}^2)} \quad for\ i < j \tag{59}$$

$$P(m_{ii}) = \frac{1}{\sqrt{2\pi}} \exp{(-\frac{m_{ii}^2}{2})} \quad for\ i = j \tag{60}$$

10.1 a

The joint probability of all the element can be written as follows,

$$P(\mathcal{M}) \equiv P[\{m_{ij}\}] = A\exp\left[-\sum_{i=1}^{N}\left(\frac{m_{ii}^2}{2} + \sum_{i<j} m_{ij}^2\right)\right] \tag{61}$$

$$= A\exp\left[-\frac{1}{2}\sum_{i=1}^{N}\left(m_{ii}^2 + \sum_{i\neq j} m_{ij}^2\right)\right] \tag{62}$$

$$= A\exp\left[-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left({m_{ij}}^2\right)\right] \tag{63}$$

We wrote the equation (61) using the fact that the element of matrix are independent of each other. So, we can just write the joint probability as the product of the individual probabilities of each element. Now we will write what $tr(\mathcal{M}^2)$ looks like,

$$tr(\mathcal{M}^2) = \sum_{i=1}^{N}\sum_{j=1}^{N} m_{ij}m_{ji}$$

$$= \sum_{i=1}^{N}\sum_{j=1}^{N} m_{ij}m_{ij} \tag{64}$$

From this we can clearly conclude that,

$$P(\mathcal{M}) = A\exp\left(-\frac{1}{2}tr(\mathcal{M}^2)\right) \tag{65}$$

10.2 b

Now we have to remark on the distribution of $tr(\mathcal{M})$ for large N. For doing that, we will first calculate the first 2 moments of $tr\mathcal{M}$.

$$\begin{aligned}\langle tr\mathcal{M}\rangle &= \langle\sum_{i=1}^{N} m_{ii}\rangle \\ &= \sum_{i=1}^{N}\langle m_{ii}\rangle = 0\end{aligned} \tag{66}$$

The last line follows from the fact that the distribution of each m_{ii}'s is Gaussian with mean 0 and standard deviation of 1.

$$\begin{aligned}\langle (tr\mathcal{M})^2\rangle &= \langle\sum_{i=1}^{N}\sum_{j=1}^{N} m_{ii}m_{jj}\rangle \\ &= \sum_{i=1}^{N}\langle (m_{ii})^2\rangle = N\end{aligned} \tag{67}$$

Now in this case the mean is zero and the second cumulant scales as N and all the higher cumulant also scales as N (as $\langle (tr\mathcal{M})^n\rangle_c = \sum_{i=1}^{N}\langle m_{ii}^n\rangle_c = N\langle m_{ii}^n\rangle_c)$. So, from the central limit theorem, we can conclude that the asymptotic distribution of $tr\mathcal{M}$ for large N, is a Gaussian distribution, with mean of 0 and standard deviation of $\sqrt{N}$,

$$Prob(tr\mathcal{M}) = \frac{1}{\sqrt{2\pi N}}\exp\left(-\frac{(tr\mathcal{M})^2}{2N}\right) \tag{68}$$

10.3 c

The Wigner semi-circle law is given by,

$$\rho(\lambda) = \sqrt{\frac{2}{N\pi^2}}\sqrt{1 = \frac{\lambda^2}{2N}} \quad \textit{for large } N \tag{69}$$

Now assuming that $\lambda_0^2 = 2N$, we can rewrite this as,

$$\rho(\lambda) = \frac{2}{\pi\lambda_0}\sqrt{1-\frac{\lambda^2}{\lambda_0^2}} \tag{70}$$

Now we would like to calculate $\langle\sum_{i=1}^{N}\lambda_i^2\rangle$. Now assuming that the λ_i's are iid. We can write, $\langle\sum_{i=1}^{N}\lambda_i^2\rangle = N\langle\lambda^2\rangle$. So, we would have,

$$\langle\lambda^2\rangle = \frac{2}{\pi}\int_{-\lambda_0}^{\lambda_0}\frac{\lambda^2}{\lambda_0}\sqrt{1-\frac{\lambda^2}{\lambda_0^2}}d\lambda \tag{71}$$

Now, we will perform a change of variable, $\frac{\lambda}{\lambda_0} = sin\theta$. Then the integral transform into the following,

$$\begin{aligned}\langle\lambda^2\rangle &= \frac{2\lambda_0^2}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} sin^2\theta cos^2\theta d\theta \\ &= \frac{4\lambda_0^2}{\pi}\int_{0}^{\frac{\pi}{2}} sin^2\theta cos^2\theta d\theta \\ &= \frac{\lambda_0^2}{\pi}\int_{0}^{\frac{\pi}{2}} sin^2 2\theta d\theta \\ &= \frac{\lambda_0^2}{2\pi}\int_{0}^{\frac{\pi}{2}} (1-cos4\theta)\, d\theta \\ &= \frac{\lambda_0^2}{4}\end{aligned} \tag{72}$$

So, we have , $\langle \sum_{i=1}^{N} \lambda_i^2 \rangle = N\frac{\lambda_0^2}{4} = \frac{N^2}{2}$.

10.4 d

As we have calculated in section(b), $\langle (tr\mathcal{M})^2 \rangle_c = N$. We can express this in the terms of the eigenvalues using the fact that, $tr\mathcal{M} = \sum_{i=1}^{N} \lambda_i$.

$$\begin{aligned} \langle (tr\mathcal{M})^2 \rangle_c &= \sum_{i=1}^{N} \sum_{j\neq i}^{N} \langle \lambda_i \lambda_j \rangle_c + \sum_{i=1}^{N} \langle \lambda_i^2 \rangle_c \\ \sum_{i=1}^{N} \sum_{j\neq i}^{N} \langle \lambda_i \lambda_j \rangle_c &= \langle (tr\mathcal{M})^2 \rangle_c - \sum_{i=1}^{N} \langle \lambda_i^2 \rangle_c = N - \frac{N^2}{2} \end{aligned} \tag{73}$$

Thus we have proved that the $\sum_{i=1}^{N} \sum_{j\neq i}^{N} \langle \lambda_i \lambda_j \rangle_c \neq 0$.

11 Problem 11: Theory of Records

11.1 a

The probability p_n that in a sequence $\{x_1, x_2, \cdots, x_N\}$ a record happened on the nth day is given by the event that,

$$max\{x_1, x_2, \cdots, x_n\} = x_n \quad for\ 1 \leq n \leq N \tag{74}$$

Now, let's assume that the random variables x_i's are drawn from iid with probability distribution of $P(x)$. Now the probability of the nth random variable x_n to be higher than all the past entries can be calculated as following.
We assume that the nth random variable have a value in the range y to $y + dy$ and all the former ones have to be lower than y. The probabilty of this is $P(y)F(y)^{n-1}dy$, where $F(y) = \int_0^y P(x)dx$. Now the y that we have chosen can be any number possible in the applicable range. So, integrating over y we will find the probability of having the maximum value for the nth random variable,

$$p_n = \int_0^\infty P(y)F(y)^{(n-1)}dy \tag{75}$$

11.2 b

Now we will use the fact that, $F(y) = \int_0^y P(x)dx$ implies that $\frac{dF}{dy} = P(y)$. So, using this we can easily do the integration of the last section.

$$p_n = \int_0^\infty \frac{dF}{dy} F(y)^{(n-1)}dy = \int_0^1 F(y)^{(n-1)}dF = \frac{1}{n} \tag{76}$$

As, we can see that the probability p_n is independent of $P(x)$.

12 Bonus Question: The Weierstrass Walk

Let's consider a one dimensional random walker that jumps at every timestep either to the right or to the left but jump in lengths of $1, b, b^2, b^3, \cdots$ so on. The probability of different jump length are different,

$$Prob(\Delta x = \pm b^m) = \frac{a-1}{2a} a^{-m} \tag{77}$$

for $m = 0, 1, 2, 3, \cdots$ and $a > 1$ and $b > 1$.

12.1 a

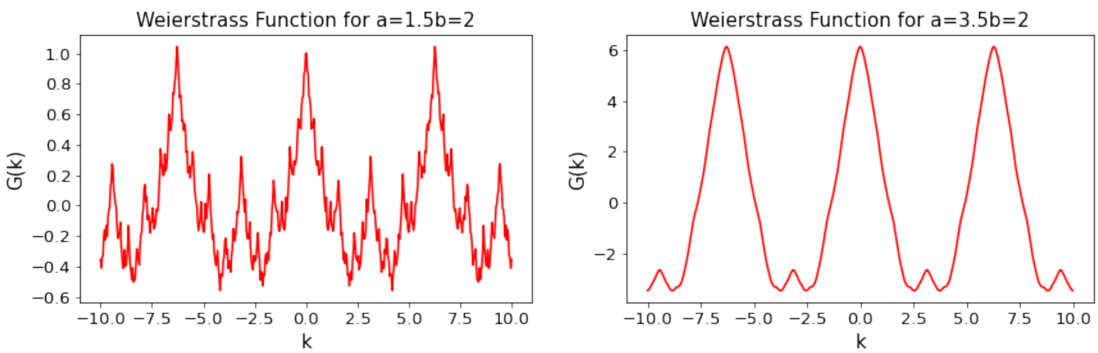
The characteristic function of $Prob(\Delta x)$ is given by,

$$
\begin{aligned}
G(k) &= \langle e^{-ikx} \rangle \\
&= \sum_i e^{-ikx_i} P(x_i) \\
&= \frac{a-1}{2a} \left[(e^{ik} + e^{-ik}) + a^{-1}(e^{ikb} + e^{-ikb}) + a^{-2}(e^{ikb^2} + e^{-ikb^2}) + \cdots \right] \\
&= \frac{a-1}{a} \left[cosk + a^{-1} coskb + a^{-2} coskb^2 + \cdots \right] \\
&= \frac{a-1}{a} \sum_{m \geq 0}^{\infty} a^{-m} coskb^m
\end{aligned}
\tag{78}
$$

Thus, we can conclude $A_m = a^{-m}$ and $B_m = b^m$.

12.2 b

Below we plot the Weierstrass function by taking the sum up to $m = 10$, for two sets of parameters ($b = 2$ and $a = 1.5$) and ($b = 2$ and $a = 3.5$).



(a) For the case $a = 1.5$ and $b = 2$

(b) For the case $a = 3.5$ and $b = 2$

Figure 6: Plots of the Weierstrass Function

12.3 c

Now we would like to calculate the mean and the variance of the probability distribution.

$$
\begin{aligned}
\langle \Delta x \rangle &= \sum_i \Delta x_i P(\Delta x_i) \\
&= \frac{a-1}{2a} \left[1 - 1 + \frac{b}{a} - \frac{b}{a} + \frac{b^2}{a^2} - \frac{b^2}{a^2} + \cdots \right] \\
&= 0
\end{aligned}
\tag{79}
$$

So, we can see that this probability distribution have a zero mean what we should expect of a even distribution. Now we will move on to calculate the variance.

$$\begin{aligned}
\langle(\Delta x)^2\rangle &= \sum_i \Delta x_i^2 P(\Delta x_i) \\
&= \frac{a-1}{2a}\left[1+1+\frac{b^2}{a}+\frac{b^2}{a}+\frac{b^4}{a^2}+\frac{b^4}{a^2}+\cdots\right] \\
&= \frac{a-1}{a}\left[1+\alpha+\alpha^2+\alpha^3+\cdots\right] \quad \textit{(here } \alpha = \frac{b^2}{a}\textit{)}
\end{aligned} \tag{80}$$

Now, for $\frac{b^2}{a} < 1$, we the variance would be finite and will take the form,

$$\begin{aligned}
\langle(\Delta x)^2\rangle &= \frac{a-1}{a}\frac{1}{1-\frac{b^2}{a}} \\
&= \frac{a-1}{a-b^2}
\end{aligned} \tag{81}$$

For the case $\frac{b^2}{a} > 1$, the variance will diverge and would not be finite anymore.

12.4 d

Now we would like to see behaviour of the probability distribution for large value of x, and try to infer what should be the asymptotic form of the probability distribution. For large value of Δx, the m can be written as $m = log_b \Delta x$. Substituting this value of m in the pdf we find,

$$\begin{aligned}
Prob(\Delta x) &= \frac{a-1}{2a}\frac{1}{a^{log_b \Delta x}} \\
&= \frac{a-1}{2a}\frac{1}{\Delta x^{1/log_a b}}
\end{aligned} \tag{82}$$

Now if $P(\Delta x) \propto \frac{1}{\Delta x^{1+\alpha}}$ for large Δx, then if $\alpha > 1$, then the CLT holds (Generally when we talk about probability distribution function then for $\alpha > 2$, the CLT holds. But in this case we are talking about probability itself.). Otherwise in the asymptotic limit it will give Levy stable distribution. Now if $b^2/a < 1$, then $log_a b < \frac{1}{2}$. So, in this case $\alpha > 1$, and we have a finite variance. So, for the case $\frac{b^2}{a} < 1$, the asymptotic distribution of the position of the position of the random walk after, N number of steps is Gaussian with 0 mean and $N\frac{a-1}{a-b^2}$ variance.
Now, for the case $\frac{b^2}{a} > 1$, the variance is not finite and the asymptotic distribution of the position after N steps will be Levy stable distribution, the particular form will depend on $\frac{b^2}{a}$.