1 Problem 1

Proof. The lecture note 4 has shown that \( \{ \theta > 0 : M(\theta) < \exp(C\theta) \} \) is nonempty. Let

\[
\theta^* := \sup\{ \theta > 0 : M(\theta) < \exp(C\theta) \}
\]

If \( \theta^* = \infty \), which implies that for all \( \theta > 0 \), \( M(\theta) < \exp(C\theta) \) holds, we have

\[
\inf_{t>0} tI(C + \frac{1}{t}) = \inf_{t>0} \{ t(\theta - \log M(\theta)) + \theta \} = \infty = \theta^*
\]

Consider the case in which \( \theta^* \) is finite. According to the definition of \( I(C + \frac{1}{t}) \), we have

\[
I(C + \frac{1}{t}) \geq \theta^*(C + \frac{1}{t}) - \log M(\theta^*)
\]

\[
\Rightarrow \inf_{t>0} tI(C + \frac{1}{t}) \geq \inf_{t>0} t(\theta^*(C + \frac{1}{t}) - \log M(\theta^*))
\]

\[
= \inf_{t>0} t(\theta^*C - \log M(\theta^*)) + \theta^*
\]

\[
\geq \theta^*
\]

(1)

Next, we will establish the convexity of \( \log M(\theta) \) on \( \{ \theta \in \mathbb{R} : M(\theta) < \infty \} \). For two \( \theta_1, \theta_2 \in \{ \theta \in \mathbb{R} : M(\theta) < \infty \} \) and \( 0 < \alpha < 1 \), Hölder’s inequality gives

\[
\mathbb{E}[\exp((\alpha \theta_1 + (1-\alpha) \theta_2)X)] \leq \mathbb{E}[\exp(\alpha \theta_1 X)]^{\frac{1}{\alpha}} \mathbb{E}[\exp((1-\alpha) \theta_1 X)]^{1-\alpha}
\]

Taking the log operations on both sides gives

\[
\log M(\alpha \theta_1 + (1 - \alpha) \theta_2) \leq \alpha \log M(\theta_1) + (1 - \alpha)M(\theta_2)
\]

By the convexity of \( \log M(\theta) \), we have

\[
(C + \frac{1}{t})\theta - \log M(\theta) \leq (C + \frac{1}{t})\theta - \theta^*C - \frac{\dot{M}(\theta^*)}{M(\theta^*)}(\theta - \theta^*)
\]

\[
= (C - \frac{\dot{M}(\theta^*)}{M(\theta^*)} + \frac{1}{t})(\theta - \theta^*) + \frac{\theta^*}{t}
\]

1
Thus, we have
\[
\inf_{t > 0} \sup_{\theta \in \mathbb{R}} \left[ (C + \frac{1}{t})\theta - \log M(\theta) \right] \leq \inf_{t > 0} \sup_{\theta \in \mathbb{R}} \left[ (C - \frac{\dot{M}(\theta^*)}{M(\theta^*)} + \frac{1}{t})(\theta - \theta^*) \right] + \theta^* 
\]
(2)

Then we will establish the fact that \(\frac{\dot{M}(\theta^*)}{M(\theta^*)} \geq C\). If not, then there exists a sufficiently small \(h > 0\) such that
\[
\frac{\log M(\theta^* - h) - \log M(\theta^*)}{-h} < C
\]
which implies that
\[
\log M(\theta^* - h) > \log M(\theta^*) - Ch \\
\Rightarrow \log M(\theta^* - h) > C(\theta^* - h) \Rightarrow M(\theta^* - h) \geq \exp(C(\theta^* - h))
\]
which contradicts the definition of \(\theta^*\). By the facts that
\[
\inf_{t > 0} \sup_{\theta \in \mathbb{R}} \left[ (C - \frac{\dot{M}(\theta^*)}{M(\theta^*)} + \frac{1}{t})(\theta - \theta^*) \right] \geq 0, \text{ (when } \theta = \theta^*)
\]
and \(\frac{\dot{M}(\theta^*)}{M(\theta^*)} \geq C\), we have that
\[
\inf_{t > 0} \sup_{\theta \in \mathbb{R}} \left[ (C - \frac{\dot{M}(\theta^*)}{M(\theta^*)} + \frac{1}{t})(\theta - \theta^*) \right] = 0
\]
and the infimum is obtained at \(t^* > 0\) such that \(C + \frac{1}{t^*} - \frac{\dot{M}(\theta^*)}{M(\theta^*)} = 0\). From (2), we have
\[
\inf_{t > 0} \sup_{\theta \in \mathbb{R}} \left[ (C + \frac{1}{t})\theta - \log M(\theta) \right] \leq \theta^* \\
\Rightarrow \inf_{t > 0} tI(C + \frac{1}{t}) \leq \theta^* 
\]
(3)

From (1) and (3), we have the result \(\inf_{t > 0} tI(C + \frac{1}{t}) = \theta^*\). 
\[\square\]
2 Problem 2 (Based on Tetsuya Kaji’s Solution)

(a). Let $\theta_0$ be the one satisfying $I(a) = \theta_0 a - \log M(\theta_0)$ and $\delta$ be a small positive number. Following the proof of the lower bound of Cramer’s theorem, we have

$$n^{-1} \log \mathbb{P}(n^{-1} S_n \geq a) \geq n^{-1} \log \mathbb{P}(n^{-1} S_n \in [a, a + \delta))$$

$$\geq -I(a) - \theta_0 \delta - n^{-1} \log \mathbb{P}(n^{-1} \tilde{S}_n - a \in [0, \delta))$$

where $\tilde{S}_n = Y_1 + \ldots + Y_n$ and $Y_i (1 \leq i \leq n)$ is i.i.d. random variable following the distribution $\mathbb{P}(Y_i \leq z) = M(\theta_0)^{-1} \int_{-\infty}^{z} \exp(\theta_0 x) dP(x)$. Recall that

$$\mathbb{P}(n^{-1} \tilde{S}_n - a \in [0, \delta)) = \mathbb{P} \left( \frac{\sum_{i=1}^{n} (Y_i - a)}{\sqrt{n}} \in [0, \sqrt{n} \delta) \right)$$

By the CLT, setting $\delta = O(n^{-1/2})$ gives

$$\mathbb{P}(n^{-1} \tilde{S}_n - a \in [0, \delta)) = O(1)$$

Thus, we have

$$n^{-1} \log \mathbb{P}(n^{-1} S_n \geq a) + I(a) \geq -\theta_0 \delta - n^{-1} \log \mathbb{P}(n^{-1} \tilde{S}_n - a \in [0, \delta))$$

$$= -O(n^{-1/2})$$

Combining the result from the upper bound $n^{-1} \log \mathbb{P}(n^{-1} S_n \geq a) \leq -I(a)$, we have

$$|n^{-1} \log \mathbb{P}(n^{-1} S_n \geq a) + I(a)| \leq \frac{C}{\sqrt{n}}$$

(b). Take $a = \mu$. It is obvious, $\mathbb{P}(n^{-1} S_n \geq \mu) \to \frac{1}{2}$ as $n \to \infty$. Recalling that $I(\mu) = 0$, we have

$$|n^{-1} \log \mathbb{P}(n^{-1} S_n \geq \mu) + I(\mu)| \sim \frac{C}{n}$$

Namely, this bound cannot be improved.

3 Problem 3

For any $n \geq 0$, define a point $M_n$ in $\mathbb{R}^2$ by

$$x_{M_n} = \frac{1}{n} \sum_{i \leq n} X_i$$
and
\[ y_{\max} = \frac{1}{n} \sum_{i \leq n} Y_i \]

Let \( B_0(1) \) be the open ball of radius one in \( \mathbb{R}^2 \). From these definitions, we can rewrite
\[
P \left( \left( \frac{1}{n} \sum_{i=1}^{n} X_i \right)^2 + \left( \frac{1}{n} \sum_{i=1}^{n} Y_i \right)^2 \geq 1 \right) = P(M_n \notin B_0(1))
\]

We will apply Cramer’s Theorem in \( \mathbb{R}^2 \):
\[
\lim_{n \to \infty} \frac{1}{n} P \left( \left( \frac{1}{n} \sum_{i=1}^{n} X_i \right)^2 + \left( \frac{1}{n} \sum_{i=1}^{n} Y_i \right)^2 \geq 1 \right) = -\inf_{(x,y) \in B_0(1)^c} I(x,y)
\]

where
\[ I(x,y) = \sup_{(\theta_1, \theta_2) \in \mathbb{R}^2} (\theta_1 x + \theta_2 y - \log(M(\theta_1, \theta_2))) \]

with
\[ M(\theta_1, \theta_2) = \mathbb{E}[\exp(\theta_1 X + \theta_2 Y)] \]

Note that since \((X, Y)\) are presumed independent, \(\log(M(\theta_1, \theta_2)) = \log(M_X(\theta_1)) + \log(M_Y(\theta_2))\), with \(M_X(\theta_1) = \mathbb{E}[\exp(\theta_1 X)]\) and \(M_Y(\theta_2) = \mathbb{E}[\exp(\theta_2 Y)]\).

We can easily compute that
\[ M_X(\theta) = \exp\left(\frac{\theta^2}{2}\right) \]

and
\[ M_Y(\theta) = \mathbb{E}[e^{\theta Y}] = \int_{-1}^{1} e^{\theta y} \frac{1}{2} dy = \frac{1}{2} \left[ e^{\theta y} \right]_{-1}^{1} = \frac{1}{2} \left( e^{\theta} - e^{-\theta} \right) \]

Since \((x, y)\) are decoupled in the definition of \((x, y)\), we obtain
\[ I(x, y) = I_X(x) + I_Y(y) \]

with
\[ I_X(x) = \sup_{\theta_1} g_1(x, \theta_1) = \sup_{\theta_1} (\theta_1 x - \frac{\theta_1^2}{2}) = \frac{x^2}{2} \]
\[ I_Y(y) = \sup_{\theta_2} g_2(y, \theta_2) = \sup_{\theta_2} (\theta_2 y - \log(\frac{1}{2\theta} (e^{\theta_2} - e^{-\theta_2}))) \]

Since for all \(y, \theta_2, g_2(y, \theta_2) = g_2(-y, -\theta_2)\), for all \(y\), \(I_Y(y) = I_Y(-y)\).
Since $I_X(x)$ is increasing in $|x|$ and $I_Y(y)$ is increasing in $|y|$, the maximum is attained on the circle $x^2 + y^2 = 1$, which can be reparametrized as a one-dimensional search over an angle $\phi$. Optimizing over $\phi$, we find that the minimum of $I(x, y)$ is obtained at $x = 1, y = 0$, and that the value is equal to $\frac{1}{2}$. We obtain

$$\lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}\left( \frac{1}{n} \sum_{i=1}^{n} X_i^2 + \frac{1}{n} \sum_{i=1}^{n} Y_i^2 \geq 1 \right) = -\frac{1}{2}$$

4 Problem 4

We denote $Y_n$ the set of all length-$n$ sequences which satisfy condition (a). The first step of our method will be to construct a Markov Chain with the following properties:

- For every $n \geq 0$, and any sequence $(X_1, ..., X_n)$ generated by the Markov Chain, $(X_1, X_2, ..., X_n)$ belongs to $Y_n$.

- For every $n \geq 0$, and every $(x_1, ..., x_n) \in Y_n$, $(x_1, ..., x_n)$ has positive probability, and all sequences of $Y_n$ are “almost” equally likely.

Consider a general Markov Chain with two states $(0, 1)$ and general transition probabilities $(P_{00}, P_{01}; P_{10}, P_{11})$. We immediately realize that if $P_{11} > 0$, sequences with two consecutive ones don’t have zero probability (in particular, for $n = 2$, the sequence $(1, 1)$ has probability $\nu(1)P_{11}$. Therefore, we set $P_{11} = 0$ (and thus $P_{10} = 0$), and verify this enforces the first condition.

Let now $P_{00} = p, P_{01} = 1 - p$, and let’s find $p$ such that all sequences are almost equiprobable. What is the probability of a sequence $(X_1, ..., X_n)$?

Every 1 in the sequence $(X_1, ..., X_n)$ necessarily transited from a 0, with probability $(1 - p)$.

Zeroes in the sequence $(X_1, ..., X_n)$ can come either from another 0, in which case they contribute a $p$ to the joint probability $(X_1, ..., X_n)$, or from a 1, in which case they contribute a 1. Denote $N_0$ and $N_1$ the numbers of 0 and 1 in the sequence $(X_1, ..., X_n)$. Since each 1 of the sequence transits to a 0 of the sequence, there are $N_1$ zeroes which contribute a probability of 1, and thus $N_0 - N_1$ zeroes contribute a probability of $p$. This is only ‘almost’ correct, though, since we have to account for the initial state $X_1$, and the final state $X_n$. By choosing for initial distribution $\nu(0) = p$ and $\nu(1) = (1 - p)$, the above reasoning applies correctly to $X_1$. 

5
Our last problem is when the last state is 1, in which case that 1 does not give a 1 to 0 transition, and the probabilities of zero-zero transitions is therefore $N_0 - N_1 + 1$. In summary, under the assumptions given above, we have:

$$P(X_1, ..., X_n) = \begin{cases} (1 - p)^{N_1} p^{N_0 - N_1}, & \text{when } X_n = 0 \\ (1 - p)^{N_1} p^{N_0 - N_1 + 1}, & \text{when } X_n = 1 \end{cases}$$

Since $N_0 + N_1 = n$, we can rewrite $(1 - p)^{N_1} p^{N_0 - N_1}$ as $(1 - p)^{N_1} p^{2N_1 - 2N_1}$, or equivalently as $(\frac{1 - p}{p})^{N_1} p^n$. We conclude

$$P(X_1, ..., X_n) = \begin{cases} \left(\frac{1 - p}{p}\right)^{N_1} p^n, & \text{when } X_n = 0 \\ \left(\frac{1 - p}{p}\right)^{N_1} p^{n+1}, & \text{when } X_n = 1 \end{cases}$$

We conclude that if $\frac{1 - p}{p} = 1$, sequences will be almost equally likely. This equation has positive solution $p = \frac{\sqrt{5} - 1}{2} \approx 0.6180$, which we take in the rest of the problem (trivia: $1/p = \phi$, the golden ratio). The steady state distribution of the resulting Markov Chain can be easily computed to be $\pi = (\pi_0, \pi_1) = (\frac{1}{2 - p}, \frac{1}{2 - p}) \sim (0.7236, 0.2764)$. We also obtain the “almost” equiprobable condition:

$$P(X_1, ..., X_n) = \begin{cases} p^n, & \text{when } X_n = 0 \\ p^{n+1}, & \text{when } X_n = 1 \end{cases}$$

We now relate this Markov Chain at hand. Note the following: $\log(|Z_n|) = \log(|Y_n|) + \log(|Y_n|)$, and therefore,

$$\lim_{n \to \infty} \frac{1}{n} \log(Z_n) = \lim_{n \to \infty} \frac{1}{n} \log(|Y_n|) + \lim_{n \to \infty} \frac{1}{n} \log\left(\frac{|Z_n|}{|Y_n|}\right)$$

Let us compute first $\lim_{n \to \infty} \frac{1}{n} \log(|Y_n|)$. This is easily done using our Markov Chain. Fix $n \geq 0$, and observe that since our Markov Chain only generates sequences which belong to $Y_n$, we have

$$1 = \frac{P(X_1, ..., X_n)}{(X_1, ..., X_n) \in Y_n}$$

Note that for any $(X_1, ..., X_n) \in Y_n$, we have $p^{n+1} \leq P(X_1, ..., X_n) \leq p^n$, and so we obtain

$$p^{n+1}|Y_n| \leq 1 \leq p^n|Y_n|$$
\[ \phi^n \leq |Y_n| \leq \phi^{n+1}, \quad n \log \phi \leq \log |Y_n| \leq (n + 1) \log \phi \]

which gives \( \lim_{n \to \infty} \frac{1}{n} \log(|Y_n|) = \log \phi \).

We now consider the term \( \frac{|Z_n|}{|Y_n|} \). The above reasoning shows that intuitively, \( |Y_n| \) is the probability of the equally likely sequences of \((X_1, \ldots, X_n)\), and that \( |Z_n| \) is the number of such sequences with more than 70% zeroes. Basic probability reasoning gives that the ratio is therefore the probability that a random sequence \((X_1, \ldots, X_n)\) has more than 70% zeroes. Let us first prove this formally, and then compute the said probability. Denote \( G(X_1, \ldots, X_n) \) the percent of zeroes of the sequence \((X_1, \ldots, X_n)\). Then, for any \( k \in [0, 1] \)

\[
\mathbb{P}(G(X_1, \ldots, X_n) \geq k) = P(X_1, \ldots, X_n)_{(X_1, \ldots, X_n) \in Z_n}
\]

Reusing the same idea as previously,

\[
\mathbb{P}(G(X_1, \ldots, X_n) \geq k) \leq \frac{1}{n} \log |Z_n| p^n \leq |Z_n| p^n \leq |Z_n| = |Z_n| \left( \frac{1}{1-p} \right)^n \leq \left( \frac{1}{p^n} \right)^n \frac{|Z_n|}{|Y_n|}
\]

Similarly,

\[
\mathbb{P}(G(X_1, \ldots, X_n) \geq k) \leq p \left( \frac{|Z_n|}{|Y_n|} \right)
\]

Taking logs, we obtain

\[
\lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}(G(X_1, \ldots, X_n) \geq k) = \lim_{n \to \infty} \frac{1}{n} \log \left( \frac{|Z_n|}{|Y_n|} \right)
\]

We will use large deviations for Markov Chain to compute that probability. First note that \( G(X_1, \ldots, X_n) \) is the same as \( F(X_i) \), when \( F(0) = 1 \) and \( F(1) = 0 \). By Miller’s Theorem, obtain that for any \( x \),

\[
\lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}(F(X_i) \geq nk) = - \inf_{x \geq k} I(x)
\]

with

\[
I(x) = \sup_{\theta} (\theta x - \log \lambda(\theta))
\]

where \( \lambda(\theta) \) is the largest eigenvalue of the matrix

\[
M(\theta) = \begin{pmatrix} p \exp(\theta) & 1 - p \\ \exp(\theta) & 0 \end{pmatrix}
\]
The characteristic equation is \( \lambda^2 - (p \exp(\theta)) \lambda - (1 - p) \exp(\theta) = 0 \), whose largest solution is \( \lambda(\theta) = \frac{p \exp(\theta) + \sqrt{p^2 \exp(2\theta) + 4(1 - p) \exp(\theta)}}{2} \). The rate function of the MC is

\[
I(x) = \sup_{\theta} (\theta x - \log(\lambda(\theta)))
\]

Since the mean of \( F \) under the steady state distribution \( \pi \) is above 0.7, the minimum \( \min_{x \geq 0.7} I(x) = I(\mu) = 0 \). Thus, \( \lim_{n \to \infty} \frac{1}{n} \log |Z_n| = 0 \), and we conclude

\[
\lim_{n \to \infty} \frac{1}{n} \log |Z_n| = \log \phi = 0.4812
\]

In general, for \( k \leq \mu \), we will have

\[
\lim_{n \to \infty} \frac{1}{n} \log |Z_n(k)| = \log \phi = 0.4812
\]

and for \( k > \mu \),

\[
\lim_{n \to \infty} \frac{1}{n} \log |Z_n(k)| = \log \phi - \sup_{\theta} (\theta k - \log(\lambda(\theta)))
\]

5 Problem 5

5.1 1(i)

Consider a standard Brownian motion \( B \), and let \( U \) be a uniform random variable over \([1/2, 1]\). Let

\[
W(t) = \begin{cases} B(t), & \text{when } t \neq U \\ B(U) = 0, & \text{otherwise} \end{cases}
\]

With probability 1, \( B(U) \) is not zero, and therefore \( \lim_{t \to U} W(t) = \lim_{t \to U} B(t) = B(U) \neq 0 = W(U) \), and \( W \) is not continuous in \( U \). For any finite collection of times \( t = (t_1, ..., t_n) \) and real numbers \( x = (x_1, ..., x_n) \), denote \( W(t) = (W(t_1), ..., W(t_n)), x = (x_1, ..., x_n) \)

\[
\mathbb{P}(W(t) \leq x) = \mathbb{P}(U \notin \{t_1, 1 \leq i \leq n\}) \mathbb{P}(W(t) \leq x | U \notin \{t_i, 1 \leq i \leq n\}) + \mathbb{P}(U \in \{t_i, 1 \leq i \leq n\}) \mathbb{P}(W(t) \leq x | U \in \{t_i, 1 \leq i \leq n\})
\]

Note that \( \mathbb{P}(U \in \{t_i, 1 \leq i \leq n\}) = 0 \), and \( \mathbb{P}(W(t) \leq x | U \notin \{t_i, i \leq n\}) = \mathbb{P}(B(t \leq x)) \), and thus the Process \( W \) has exactly the same distribution properties as \( B \) (gaussian process, independent and stationary increments with zero mean and variance proportional to the size of the interval).
5.2  1(ii)

Let $X$ be a Gaussian random variable (mean 0, standard deviation 1), and denote $\mathbb{Q}_X$ the set $\{q + x, q \in \mathbb{Q}\} \cup \mathbb{R}_+$ where $\mathbb{Q}$ is the set of rational numbers.

$$W(t) = \begin{cases} B(t), & \text{when } t \notin \mathbb{Q}_X \setminus \{0\} \\ B(t) + 1, & \text{when } t \in \mathbb{Q}_X \setminus \{0\} \end{cases}$$

Through the exact same argument as 1(i), $W$ has the same distribution properties as $B$ (this is because $\mathbb{Q}_X$, just like $\{t_i, 1 \leq i \leq n\}$, has measure zero for a random variable with density).

However, note that for any $t > 0$, $|t - x - \frac{[(t-x)10^n]}{10^n}| \leq 10^{-n}$, proving that $\lim_n (x + \frac{[(t-x)10^n]}{10^n}) = t$. However, for any $n$, $x + \frac{[(t-x)10^n]}{10^n} \in \mathbb{Q}_X$, and so $\lim_n W(x + \frac{[(t-x)10^n]}{10^n}) = B(t) + 1 = B(t)$. This proves $W(t)$ is surely discontinuous everywhere.

5.3  2

Let $t \geq 0$, and consider the event $E_n = \{|B(t + \frac{1}{n}) - B(t)| > \epsilon\}$. Then, since $B(t + \frac{1}{n}) - B(t)$ is equal in distribution to $\frac{1}{\sqrt{n}}N$, where $N$ is a standard normal, by Chebychev’s inequality, we have

$$\mathbb{P}(E_n) = \mathbb{P}(n^{-1/2}|N| > \epsilon) = \mathbb{P}(|N| > \epsilon n^{-1/2}) = \mathbb{P}(N^4 > \epsilon^4 n^{-2}) \leq \frac{3}{\epsilon^4 n^2}$$

Since $\sum_n \mathbb{P}(E_n) = \sum_n \frac{1}{n} < \infty$, by Borel-Cantelli lemma, we have that there almost surely exists $N$ such that for all $n \geq N$, $|B(t + 1/n) - B(t)| \leq \epsilon$, proving $\lim_{n \to \infty} B(t + 1/n) = B(t)$ almost surely.

6  Problem 6

The event $B \in A_R$ is included in the event $B(2) - B(1) = B(1) - B(0)$, and thus

$$P(B \in A_R) \leq P(B(2) - B(1) = B(1) - B(0)) = 0$$

Since the probability that two atomless, independent random variables are equal is zero (easy to prove using conditional probabilities).