EE5139/EE6139
Solution 1

Semester I, AY2021/2022
Marco Tomamichel

Exercise 1.1 Expectation value and variance [EE5139]

Let V and W be discrete random variables defined on some probability space with a joint pmf $P_{VW}(v, w)$. We do not assume independence.

a.) Prove that $\mathbb{E}[V + W] = \mathbb{E}[V] + \mathbb{E}[W]$.

Solution:

$$\begin{aligned}\mathbb{E}[V + W] &= \sum_{v,w}(v + w)p_{V,W}(v, w) = \sum_{v,w} vp_{V,W}(v, w) + \sum_{w} wp_{V,W}(v, w)\\ &= \sum_{v} vp_V(v) + \sum_{w} wp_W(w) = \mathbb{E}[V] + \mathbb{E}[W]\end{aligned}$$

b.) Prove that if V and W are independent, then $\mathbb{E}[VW] = \mathbb{E}[V]\mathbb{E}[W]$.

Solution:

$$\begin{aligned}\mathbb{E}[VW] &= \sum_{v,w} vwp_{V,W}(v, w) = \sum_{v,w} vwp_V(v)p_W(w)\\ &= \sum_{v} vp_V(v)\sum_{w} wp_W(w) = \mathbb{E}[V]\mathbb{E}[W]\end{aligned}$$

c.) Let V and W be independent and let σ_V^2 and σ_W^2 be their respective variances. Find the variance of $Z = V + W$.

Solution:

$$\begin{aligned}\sigma_Z^2 &= \mathbb{E}[Z^2] - \mathbb{E}[Z]^2 = \mathbb{E}[(V + W)^2] - \mathbb{E}[V + W]^2\\ &= \mathbb{E}[V^2 + W^2 + 2VW] - (\mathbb{E}[V]^2 + 2\mathbb{E}[V]\mathbb{E}[W] + \mathbb{E}[W]^2)\\ &= \mathbb{E}[V^2] - \mathbb{E}[V]^2 + \mathbb{E}[W^2] - \mathbb{E}[W]^2 = \sigma_V^2 + \sigma_W^2,\end{aligned}$$

where the second to last equality follows from the fact that $\mathbb{E}[VW] = \mathbb{E}[V]\mathbb{E}[W]$.

Exercise 1.2 Coin flips [EE5139]

Flip a fair coin four times. Let X be the number of Heads obtained, and let Y be the position of the first Heads i.e. if the sequence of coin flips is TTHT, then $Y = 3$, if it is THHH, then $Y = 2$. If there are no heads in the four tosses, then we define $Y = 0$.

a.) Model the experiment completely, i.e. define the sample space and the random variables X and Y as functions from that sample space.

Solution: The underlying sample space is the set

$$\Omega = \{TTTT, TTTH, \dots, HHHH\}.$$

Each outcome $\omega \in \Omega$ can be mapped to $X(\omega)$ and $Y(\omega)$, e.g.,

$$\begin{array}{rl} X(TTTT) = 0 & Y(TTTT) = 0 \\ X(THHT) = 2 & Y(THHT) = 2 \\ X(TTTH) = 1 & Y(TTTH) = 4 \\ \text{etc.} & \end{array}$$

b.) Find the joint pmf of X and Y.

Solution: By listing all 16 elements of Ω, and computing X and Y for each we can see that, e.g.,

$$\{X = 2, Y = 2\} = \{THHT, THTH\}$$

and thus $P_{XY}(2,2) = \frac{1}{16} + \frac{1}{16} = \frac{1}{8}$. This can be repeated for all feasible values of X and Y as in the table below: We can now read off

	x				
y	0	1	2	3	4
0	$\frac{1}{16}$				
1		$\frac{1}{16}$	$\frac{3}{16}$	$\frac{3}{16}$	$\frac{1}{16}$
2		$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{16}$	
3		$\frac{1}{16}$	$\frac{1}{16}$		
4		$\frac{1}{16}$			

c.) Using the joint pmf, find the marginal pmf of X. What is $\Pr[Y = 0|X = 1]$ and $\Pr[Y = 1|X = 3]$?

Solution: Summing over the columns of the table below, we get the pmf of X as

$$P_X(x) = \begin{cases} \frac{1}{16} & x = 0 \\ \frac{1}{4} & x = 1 \\ \frac{3}{8} & x = 2 \\ \frac{1}{4} & x = 3 \\ \frac{1}{16} & x = 4 \end{cases}$$

We can now use the Bayes' rule to compute

$$\Pr[Y = 0|X = 1] = \frac{P_{XY}(1,0)}{P_X(1)} = 0,$$

$$\Pr[Y = 1|X = 3] = \frac{P_{XY}(3,1)}{P_X(3)} = \frac{3}{16} \cdot \frac{4}{1} = \frac{3}{4}.$$

Exercise 1.3 Property of convex functions [EE5139]

Let f be convex on $[a, b]$. Using only the defining property of convex functions, show that for any $a \leq x_1 < x_2 \leq x_3 < x_4 \leq b$, we have

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1} \leq \frac{f(x_4) - f(x_3)}{x_4 - x_3}.$$

Solution: Recall that a function $f : (a,b) \to \mathbb{R}$ is convex on (a,b) if for all $x, y \in (a,b)$ with $x < y$ and $0 \le \lambda \le 1$,

$$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda)f(y). \tag{1}$$

For brevity in notation, we define $g(x,y) = (f(y)-f(x))/(y-x)$, so we must show that $g(x_1,x_2) \le g(x_3,x_4)$. We first show that for $x_1, x_2, x_3 \in (a,b)$ such that $x_1 < x_2 < x_3$, we have $g(x_1,x_2) \le g(x_2,x_3)$, i.e., that

$$\frac{f(x_2)-f(x_1)}{x_2-x_1} \le \frac{f(x_3)-f(x_2)}{x_3-x_2}. \tag{2}$$

In (1), we let $x \leftarrow x_1$, $y \leftarrow x_3$ and $\lambda = (x_3-x_2)/(x_3-x_1) \in [0,1]$. Then we check that

$$\lambda x + (1-\lambda)y = \lambda x_1 + (1-\lambda)x_3 = \frac{x_3-x_2}{x_3-x_1}\cdot x_1 + \frac{x_2-x_1}{x_3-x_1}\cdot x_3 = x_2. \tag{3}$$

So we have

$$f(\lambda x + (1-\lambda)y) = f(x_2) \le \frac{x_3-x_2}{x_3-x_1}f(x_1) + \frac{x_2-x_1}{x_3-x_1}f(x_3) = \lambda f(x_1) + (1-\lambda)f(x_3). \tag{4}$$

This is equivalent to

$$\lambda f(x_2) + (1-\lambda)f(x_2) = \lambda f(x_1) + (1-\lambda)f(x_3), \tag{5}$$

or

$$\lambda(f(x_2)-f(x_1)) \le (1-\lambda)(f(x_3)-f(x_2)) \tag{6}$$

Recalling the definition of λ shows (2), i.e., that $g(x_1,x_2) \le g(x_2,x_3)$. By the same logic, for three points $x_2 < x_3 < x_4$, we also have $g(x_2,x_3) \le g(x_3,x_4)$. Putting these two inequalities yields $g(x_1,x_2) \le g(x_3,x_4)$ as desired.

Exercise 1.4 Finite fields [EE5139]

Derive the addition and multiplication tables for F_8 and F_9. You should use the construction described in the lecture notes and the irreducible polynomials x^3+x+1 for F_8 and x^2+1 for F_9.
Hint: You may want to use Matlab to solve this problem. However, you will need to compute some elements by hand to verify the computer-generated output.
Solution:

$+_{F_8}$	0	1	2	3	4	5	6	7
0	0	1	2	3	4	5	6	7
1	1	0	3	2	5	4	7	6
2	2	3	0	1	6	7	4	5
3	3	2	1	0	7	6	5	4
4	4	5	6	7	0	1	2	3
5	5	4	7	6	1	0	3	2
6	6	7	4	5	2	3	0	1
7	7	6	5	4	3	2	1	0

$+_{F_9}$	0	1	2	3	4	5	6	7	8
0	0	1	2	3	4	5	6	7	8
1	1	2	0	4	5	3	7	8	6
2	2	0	1	5	3	4	8	6	7
3	3	4	5	6	7	8	0	1	2
4	4	5	3	7	8	6	1	2	0
5	5	3	4	8	6	7	2	0	1
6	6	7	8	0	1	2	3	4	5
7	7	8	6	1	2	0	4	5	3
8	8	6	7	2	0	1	5	3	4

$\times_{F_8}$	0	1	2	3	4	5	6	7
0	0	0	0	0	0	0	0	0
1	0	1	2	3	4	5	6	7
2	0	2	4	6	3	1	7	5
3	0	3	6	5	7	4	1	2
4	0	4	3	7	6	2	5	1
5	0	5	1	4	2	7	3	6
6	0	6	7	1	5	3	2	4
7	0	7	5	2	1	6	4	3

$\times_{F_9}$	0	1	2	3	4	5	6	7	8
0	0	0	0	0	0	0	0	0	0
1	0	1	2	3	4	5	6	7	8
2	0	2	1	6	8	7	3	5	4
3	0	3	6	2	5	8	1	4	7
4	0	4	8	5	6	1	7	2	3
5	0	5	7	8	1	3	4	6	2
6	0	6	3	1	7	4	2	8	5
7	0	7	5	4	2	6	8	3	1
8	0	8	4	7	3	2	5	1	6

The MatLab Program generating above tables:

Listing 1: addition_table.m

```
p = 3;
degree = 1; % Thus the size of the field is p^(degree+1)
h = [1,0,1]; % The irridicible polynomial of degree 'degree+1'
% List all polynomials
F = zeros(p^(degree+1),degree+1);
for i = 1:p^(degree+1)
    m = i-1;
    for j = 1:degree+1
        F(i,j) = mod(m,p);
        m = (m-mod(m,p))/p;
    end
end
a_table = zeros(p^(degree+1));
for i = 1:p^(degree+1)
    for j = 1:p^(degree+1)
        f = mod(F(i,:)+F(j,:),p);
        [~,k] = ismember(f,F,'rows');
        a_table(i,j) = k-1;
    end
end
disp(a_table);
```

Listing 2: multiplication_table.m

```
p = 3;
degree = 1; % Thus the size of the field is p^(degree+1)
h = [1,0,1]; % The irridicible polynomial of degree 'degree+1'
% List all polynomials
F = zeros(p^(degree+1),degree+1);
for i = 1:p^(degree+1)
    m = i-1;
    for j = 1:degree+1
        F(i,j) = mod(m,p);
        m = (m-mod(m,p))/p;
    end
end
m_table = zeros(p^(degree+1));
```

```
for i = 1:p^(degree+1)
    for j = 1:p^(degree+1)
        f = polyMod(polyMultiply(F(i,:),F(j,:),p),h,p);
        [~,k] = ismember(f,F,'rows');
        m_table(i,j) = k-1;
    end
end
disp(m_table);
```

Listing 3: polyMultiply.m

```
function [ f ] = polyMultiply (g, h, p)
    % Multiply two polynomials on base field Fp where p is a prime number.
    g_degree = length(g)-1;
    h_degree = length(h)-1;
    f = zeros(1, g_degree + h_degree+1);
    for k = 0:(length(f)-1) % k-th order
        for a = max(0,k-h_degree):min(k,g_degree)
            b = k-a;
            f(k+1) = f(k+1) + g(a+1)*h(b+1);
        end
    end
    f = mod(f, p);
end
```

Listing 4: polyMod.m

```
function [ f ] = polyMod (g, h, p)
    % Find the modulo of polynomial g with respect to irriducible polynomial h on base field Fp where p is a prime number.
    g_length = length(g);
    h_length = length(h);
    while (max([g ~= 0].*[1:g_length]) >= h_length)
        d = max([g ~= 0].*[1:g_length]);
        C = g(d);
        h_shifted = zeros(size(g));
        h_shifted(d-h_length+1:d) = h;
        g = mod(g-C*h_shifted, p);
    end
    f = g(1: h_length-1);
end
```

Exercise 1.5 Continuous and discrete random variables [all]

Consider the following random experiment. A ball is thrown and lands after X meters, where X is distributed uniformly in the interval $[1, 2]$. It either stays there or bounces off and jumps again an additional distance of $\frac{1}{2}X$. The binary random variable $Y \in \{0, 1\}$ takes the value 0 (with probability 50%), indicating that the ball stays put, and 1 (with probability 50%), indicating that the ball jumps again. After this additional bounce the ball rests.

a.) Express the total distance Z that the ball travels in terms of X and Y. Compute and plot the pdf for Z.

Solution:

$$Z = X + \frac{1}{2}XY.$$

We may first compute the pmf for Z ending up in an interval around z, e.g. $[z \pm \epsilon] = [z-\epsilon, z+\epsilon]$.

$$\begin{aligned}
\Pr[Z \in [z \pm \epsilon]] &= \Pr[Z \in [z \pm \epsilon] | Y = 0] \Pr[Y = 0] + \Pr[Z \in [z \pm \epsilon] | Y = 1] \Pr[Y = 1] \\
&= \frac{1}{2} \Pr[X \in [z \pm \epsilon] | Y = 0] + \frac{1}{2} \Pr\left[\frac{3}{2}X \in [z \pm \epsilon] | Y = 1\right] \\
&= \frac{1}{2} \Pr[X \in [z \pm \epsilon]] + \frac{1}{2} \Pr\left[\frac{3}{2}X \in [z \pm \epsilon]\right].
\end{aligned}$$

For $[z \pm \epsilon] \subset [1, 3/2)$,

$$\Pr[Z \in [z \pm \epsilon]] = \frac{1}{2} \times 2\epsilon + 0 = \epsilon.$$

For $[z \pm \epsilon] \subset [3/2, 2]$,

$$\Pr[Z \in [z \pm \epsilon]] = \frac{1}{2} \times 2\epsilon + \frac{1}{2} \times \frac{4}{3}\epsilon = \frac{5}{3}\epsilon.$$

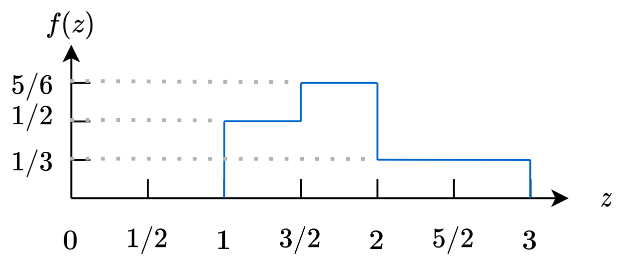
For $[z \pm \epsilon] \subset (2, 3]$,

$$\Pr[Z \in [z \pm \epsilon]] = \frac{1}{2} \times \frac{4}{3}\epsilon = \frac{2}{3}\epsilon.$$

The pdf can then be derived by taking the derivative (note that since we grow the interval on both sides a factor $\frac{1}{2}$ needs to be introduced).

$$f(z) = \lim_{\epsilon \to 0} \frac{\Pr\left[Z \in [z-\epsilon, z+\epsilon]\right]}{2\epsilon} = \begin{cases} \frac{1}{2}, & z \in (1, 3/2), \\ \frac{5}{6}, & z \in (3/2, 2), \\ \frac{1}{3}, & z \in (2, 3), \\ 0, & \text{otherwise.} \end{cases}$$

(Obviously, the pdf can also be found more directly.) On the boundary points the values of $f(z)$ are not uniquely defined — they depend on the convention used. If we define them as the limit above then it would simply be the average between the two regions adjoining the point. Plot:



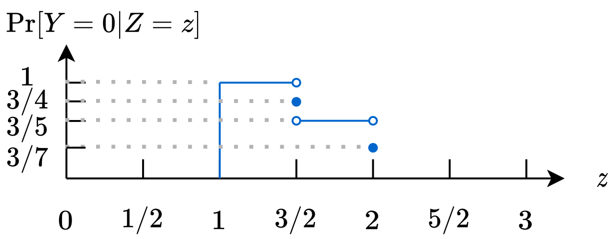
b.) Find the pmf for Y given $Z = z$. Plot $\Pr[Y = 0|Z = z]$ as a function of z.

Hint: We would be inclined to use Bayes' rule here, but the problem is that the $\Pr[Z = z] = 0$ for each z. To avoid this, consider an interval $z \pm \epsilon$ and compute the pmf for Y given $Z \in [z - \epsilon, z + \epsilon]$ and then let $\epsilon \to 0$.

Solution:

$$\begin{aligned}
\Pr[Y = 0|Z = z] &= \lim_{\epsilon \to 0} \Pr[Y = 0|Z \in [z \pm \epsilon]] \\
&= \lim_{\epsilon \to 0} \frac{\Pr[Z \in [z \pm \epsilon]|Y = 0] \Pr[Y = 0]}{\Pr[Z \in [z \pm \epsilon]]} \\
&= \begin{cases} 1, & z \in [1, \frac{3}{2}) \\ \lim_{\epsilon \to 0} (2\epsilon \cdot \frac{1}{2})/\frac{4\epsilon}{3} = \frac{3}{4}, & z = \frac{3}{2} \\ \lim_{\epsilon \to 0} (2\epsilon \cdot \frac{1}{2})/\frac{5\epsilon}{3} = \frac{3}{5}, & z \in (\frac{3}{2}, 2) \\ \lim_{\epsilon \to 0} (\epsilon \cdot \frac{1}{2})/\frac{7\epsilon}{6} = \frac{3}{7}, & z = 2 \\ 0, & \text{otherwise.} \end{cases}
\end{aligned}$$

Plot:



Exercise 1.6 Matrix representation of a communication channel [all]

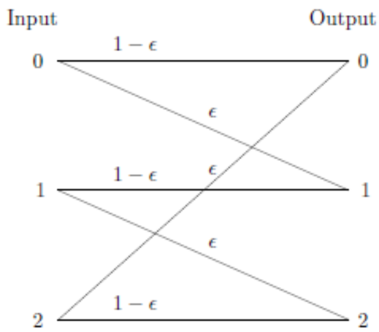


Figure 1: Ternary communication channel

A ternary communication channel is shown in Figure 1.

a.) Represent the channel as a matrix W such that the output distribution of the channel can be written as the matrix product pW, where p is a row vector containing the three input probabilities.

Solution:

$$W = \begin{bmatrix} 1-\epsilon & \epsilon & 0 \\ 0 & 1-\epsilon & \epsilon \\ \epsilon & 0 & 1-\epsilon \end{bmatrix}$$

b.) Suppose that the input probabilities are given by the vector $p = [\frac{1}{2}, \frac{1}{4}, \frac{1}{4}]$. Find the probabilities of the output symbols.

Solution:

$$pW = \begin{bmatrix} \frac{1}{2} - \frac{\epsilon}{4}, & \frac{1}{4} + \frac{\epsilon}{4}, & \frac{1}{4} \end{bmatrix}$$

c.) Suppose that 1 was observed at the output. What's the probability that the input was 0? 1? 2?

Solution: We use the formula

$$P_{X|Y}(x|y) = \frac{P_{Y|X}(y|x)P_X(x)}{P_Y(y)}$$

with $y = 1$. Now $P_Y(1) = \frac{1}{4} + \frac{1}{4}\epsilon$ and so

$$\begin{aligned} P_{X|Y}(0|1) &= \frac{P_{Y|X}(1|0)P_X(0)}{\frac{1}{4} + \frac{1}{4}\epsilon} = \frac{\frac{1}{2}\epsilon}{\frac{1}{4} + \frac{1}{4}\epsilon} \\ P_{X|Y}(1|1) &= \frac{P_{Y|X}(1|1)P_X(1)}{\frac{1}{4} + \frac{1}{4}\epsilon} = \frac{\frac{1}{4}(1-\epsilon)}{\frac{1}{4} + \frac{1}{4}\epsilon} \\ P_{X|Y}(2|1) &= \frac{P_{Y|X}(1|2)P_X(2)}{\frac{1}{4} + \frac{1}{4}\epsilon} = 0 \end{aligned}$$

We may check that $\sum_{x=0}^{2} P_{X|Y}(x|1) = 1$.

Exercise 1.7 Further tail bounds [EE6139]

a.) For a nonnegative integer-valued random variable N, show that $\mathbb{E}[N] = \sum_{n>0} \Pr(N \geq n)$.

Solution: Solution:

$$\begin{aligned} \mathbb{E}[N] &= \sum_{n=0}^{\infty} n p_N(n) = 0 \cdot p_N(0) + 1 \cdot p_N(1) + 2 \cdot p_N(2) + 3 \cdot p_N(3) + \ldots && (7) \\ &= [p_N(1) + p_N(2) + p_N(3) + \ldots] \\ &\quad + [p_N(2) + p_N(3) + p_N(4) + \ldots] \\ &\quad + [p_N(3) + p_N(4) + p_N(5) + \ldots] + \ldots && (8) \\ &= \Pr(N \geq 1) + \Pr(N \geq 2) + \Pr(N \geq 3) + \ldots = \sum_{n>0} \Pr(N \geq n) && (9) \end{aligned}$$

b.) Derive the Cauchy-Schwarz inequality, which says that $\mathbb{E}[AB] \leq \sqrt{\mathbb{E}[A^2]\mathbb{E}[B^2]}$.

Hint: Consider the non-negative random variable $(X - \alpha Y)^2$ and compute its expectation, then choose α appropriately.

Solution:

We have

$$0 \leq \mathbb{E}\left[(X-\alpha Y)^2\right] = \mathbb{E}[X^2] - 2\alpha\mathbb{E}[XY] + \alpha^2\mathbb{E}[Y^2]\,. \tag{10}$$

Choosing $\alpha = \mathbb{E}[XY]/\mathbb{E}[Y^2]$ yields

$$0 \leq \mathbb{E}[X^2] - 2\frac{\mathbb{E}[XY]^2}{\mathbb{E}[Y^2]} + \frac{\mathbb{E}[XY]^2}{\mathbb{E}[Y^2]}, \tag{11}$$

which equals the desired statement after multiplication with $\mathbb{E}[Y^2]$ on both sides.

c.) Derive the one-sided Cheybyshev inequality, which says that $\Pr(Y \geq a) \leq \sigma_Y^2/(\sigma_Y^2 + a^2)$ if $\mathbb{E}[Y] = 0$ and $a > 0$.

Solution: Since Y has zero mean, we have

$$a = \mathbb{E}[a - Y]$$

Consider the expectation above: We have

$$\mathbb{E}[a-Y] = \sum_y P_Y(y)(a-y) = \sum_{y:y<a} P_Y(y)(a-y) + \sum_{y:y\geq a} P_Y(y)(a-y) \leq \sum_{y:y<a} P_Y(y)(a-y) \tag{12}$$

because the second sum is non-positive. This can be written as

$$a \leq \sum_y P_Y(y)(a-y)\mathbf{1}\{y < a\} \tag{13}$$

where $\mathbf{1}\{\text{statement}\}$ returns 1 if the statement is true and 0 otherwise. Thus, we have

$$a \leq \mathbb{E}[(a-Y)\mathbf{1}\{Y < a\}]\mathbf{1} \tag{14}$$

Now square both sides,

$$a^2 \leq (\mathbb{E}[(a-Y)\mathbf{1}\{Y < a\}])^2 \tag{15}$$

Apply Cauchy-Schwarz inequality to the expectation,

$$a^2 \leq \mathbb{E}[(a-Y)^2]\mathbb{E}[\mathbf{1}\{Y<a\}^2] = \mathbb{E}[(a-Y)^2]\mathbb{E}[\mathbf{1}\{Y<a\}] = (a^2 + \mathbb{E}[Y^2])\Pr(Y < a) \tag{16}$$

Since $\mathbb{E}[Y^2] = \sigma_Y^2$, rearrangement of the above inequality yields the one-sided Chebyshev inequality as desired.

d.) Derive the reverse Markov inequality: Let X be a random variable such that $\Pr(X \leq a) = 1$ for some constant a. Then for $d < \mathbb{E}[X]$, we have

$$\Pr(X > d) \geq \frac{\mathbb{E}[X] - d}{a - d}.$$

Solution: Apply Markov's inequality to the non-negative random variable $\tilde{X} := a - X$. Then one has

$$\Pr(X \leq d) = \Pr(a - \tilde{X} \leq d) = \Pr(\tilde{X} \geq a - d) \leq \frac{\mathbb{E}[\tilde{X}]}{a-d} = \frac{\mathbb{E}[a-X]}{a-d}$$

Hence,

$$\Pr(X < d) \geq 1 - = \frac{\mathbb{E}[a-X]}{a-d} = \frac{\mathbb{E}[X] - d}{a-d}.$$

e.) **Chernoff Bound**: Let $X_1, \ldots, X_n$ be a sequence of i.i.d. rvs with zero-mean and moment generating function $M_X(s) := \mathbb{E}[e^{sX}]$. Show that for any $\epsilon > 0$,

$$\Pr\left(\frac{1}{n}(X_1 + \ldots + X_n) > \epsilon\right) \leq \exp\left[-n \max_{s \geq 0}(\epsilon s - \log M_X(s))\right].$$

Hint: Note that the event $\{\frac{1}{n}(X_1 + \ldots + X_n) > \epsilon\}$ occurs if and only if $\{\exp(s(X_1 + \ldots + X_n)) > \exp(ns\epsilon)\}$ occurs for any fixed $s \geq 0$. Now apply Markov's inequality.

Solution: Using the hint, we know that

$$\Pr\left(\frac{1}{n}(X_1 + \ldots + X_n) > \epsilon\right) = \Pr\left(\exp(s(X_1 + \ldots + X_n)) > \exp(ns\epsilon)\right) \tag{17}$$

$$\leq \frac{\mathbb{E}[\exp(s(X_1 + \ldots + X_n))]}{\exp(n\epsilon s)} \tag{18}$$

where the inequality is from Markov's inequality. Since the X_i's are independent,

$$\Pr\left(\frac{1}{n}(X_1 + \ldots + X_n) > \epsilon\right) \leq \frac{\prod_{i=1}^n \mathbb{E}[\exp(sX_i)]}{\exp(n\epsilon s)} = \exp[-n(\epsilon s - \log M_X(s))]$$

where the final equality is due to the fact that the X_i are identically distributed. Since $s \geq 0$ is arbitrary, we can minimize over this parameter.