Lecture 4: Chernoff Bound

- Let $\mathbb{X}$ represent the $\mathrm{Bern}(p)$ random variable
- Let $\mathbb{X}^{(1)}, \dots, \mathbb{X}^{(n)}$ represent n independent and identical copies of the random variable $\mathbb{X}$
- Let $\mathbb{S}_n := \mathbb{X}^{(1)} + \cdots + \mathbb{X}^{(n)}$ represent the sum of these n random variables. That is, $\mathbb{S}_n$ is the Binomial distribution with parameters (n, p).
- In the previous lecture we saw that $\mathbb{E}[\mathbb{S}_n] = np$ by the linearity of expectation
- For example, if $\mathbb{X}$ represents a coin-toss, then $\mathbb{S}_n$ is a random variable representing the number of observed Heads when n coin-tosses are performed
- How does the random variable $\mathbb{S}_n$ concentrate around its mean? What is the probability of $\mathbb{S}_n$ to be "far" from the expected value?

Analysis using Markov Bound

- One can use Markov bound to deduce

$$\mathbb{P}\left[\mathbb{S}_n \geqslant \lambda \cdot (np)\right] \leqslant \frac{1}{\lambda}.$$

- Can we do better?

Analysis using Chebyshev's Inequality

- By Chebyshev's Inequality, we have

$$\mathbb{P}\left[|\mathbb{S}_n - np| \geqslant t\right] \leqslant \frac{\mathrm{Var}\left[\mathbb{S}_n\right]}{t^2}.$$

- In the previous lecture we prove that $\mathrm{Var}\left[\mathbb{S}_n\right] = npq$, where $q = (1 - p)$
- Think: The probability of $\mathbb{S}_n$ being $\Theta\left(\sqrt{npq}\right)$ far from the mean is at most a constant.
- Think: Can we use higher moments to get better bounds?
- Think: Let $(\mathbb{X}_1, \ldots, \mathbb{X}_n)$ be a joint distribution and $\mathbb{S}_n = \sum_{i=1}^{n} \mathbb{X}_i$. Suppose the marginals $\mathbb{X}_i = \mathrm{Bern}\left(p\right)$ and the random variables $\mathbb{X}_i$ and $\mathbb{X}_j$ are *pair-wise independent* when $j \neq i$. Can we still apply this estimation technique?

A Large Deviation Bound

Observe that

$$\mathbb{P}\left[\mathbb{S}_n \geqslant k\right] = \sum_{i=k}^{n} \binom{n}{i} \cdot p^i q^{n-i},$$

where $q = (1 - p)$.

Claim

$$\binom{n}{k} \cdot p^k q^{n-k} \leqslant \mathbb{P}\left[\mathbb{S}_n \geqslant k\right] \leqslant \binom{n}{k} \cdot p^k.$$

- Think: How to prove this claim?
- Think: For what values of p and k is the upper bound meaningful? Hint: Use Stirling's formula.
- Think: When $p = 1/2$, for what values of k is the upper bound < 1?

- Our objective is to study the expression

$$\mathbb{P}\left[\mathbb{S}_n \geqslant k\right] = \sum_{i=k}^{n} \binom{n}{i} \cdot p^i q^{n-i},$$

where $q = (1-p)$ and $k/n > p$. This expression is known as the *upper tail* of the binomial distribution

- By Stirling approximation, we know that

$$\binom{n}{k} \cdot p^k q^{n-k} \sim \frac{1}{\sqrt{2\pi n p' q'}} \exp\left(-n\mathrm{D_{KL}}\left(p', p\right)\right),$$

where $p' = k/n$, $q' = (1-p')$, and

$$\mathrm{D_{KL}}\left(a, b\right) = a \ln\left(\frac{a}{b}\right) + (1-a)\ln\left(\frac{1-a}{1-b}\right)$$

represents the Kullback–Leibler divergence. Recall that $f(n) \sim g(n)$ if (and only if) $f(n) = (1 + \mathrm{o}(1)) \cdot g(n)$.

- Therefore, we have the following lower bound

$$\mathbb{P}\left[\mathbb{S}_n \geqslant k\right] \geqslant \binom{n}{k} p^k q^{n-k} \sim \frac{1}{\sqrt{2\pi n p' q'}} \exp\left(-n \mathrm{D}_{\mathrm{KL}}\left(p', p\right)\right).$$

- For the upper bound, we follow the strategy below.
  1. Consider the sequence

$$\left\{ \binom{n}{i} p^i q^i \right\}_{i \geqslant k}.$$

  2. We show that the following geometric sequence dominates it

$$\left\{ \binom{n}{k} p^k q^{n-k} \cdot \rho^{i-k} \right\}_{i \geqslant k},$$

where $\rho = \frac{q'}{p'} \cdot \frac{p}{q}$.
Think: Why is $\rho < 1$ when $p' = (k/n) > p$?
Think: How to prove this bound?

3. Then, we have the following upper bound

$$\mathbb{P}\left[\mathbb{S}_n > k\right] \leqslant \frac{1}{1-\rho} \cdot \binom{n}{k} p^k q^{n-k}$$
$$\sim \frac{1}{1-\rho} \cdot \frac{1}{\sqrt{2\pi n p' q'}} \exp\left(-n\mathrm{D}_{\mathrm{KL}}\left(p', p\right)\right).$$

- Consequently, we have the following tight bounds

$$1 \lesssim \frac{\mathbb{P}\left[\mathbb{S}_n \geqslant k\right]}{\frac{1}{\sqrt{2\pi n p' q'}} \cdot \exp\left(-n\mathrm{D}_{\mathrm{KL}}\left(p', p\right)\right)} \lesssim (1-\rho)^{-1},$$

where $\rho = \frac{q'}{p'} \cdot \frac{p}{q}$.

- Observe that if p' is a constant $> p$, then
  1. The lower and the upper bounds are within a constant factor of each other!
  2. The probability is exponentially decreasing in n.

- The conclusions are summarized in the next result

Lemma (Conclusions)

Let $\mathbb{S}_n = \mathbb{X}^{(1)} + \cdots + \mathbb{X}^{(n)}$, *where* $\mathbb{X} = \mathrm{Bern}(p)$.

1. $$\mathbb{P}[\mathbb{S}_n \geqslant k] \leqslant \binom{n}{k} p^k.$$

2. $$1 \leqslant \frac{\mathbb{P}[\mathbb{S}_n \geqslant k]}{\binom{n}{k} p^k q^{n-k}} \leqslant \frac{1}{1-\rho},$$

   where $\rho = \frac{q'}{p'} \cdot \frac{p}{q}$, $p' = k/n > p$, $q = (1-p)$, *and* $q' = (1-p')$.

3. $$\binom{n}{k} p^k (1-p)^{n-k} \sim \frac{1}{\sqrt{2\pi n p' q'}} \cdot \exp\left(-n \cdot \mathrm{D_{KL}}(p', p)\right),$$

   where $\mathrm{D_{KL}}(a, b) = a \ln\left(\frac{a}{b}\right) + (1-a)\ln\left(\frac{1-a}{1-b}\right)$ *and* $p' = k/n$.

Chernoff Bound: Proof

- Let us now upper bound the probability $\mathbb{P}\left[\mathbb{S}_{n,p} \geqslant n(p+\varepsilon)\right]$ using the Chernoff bound. Theupper boundd will be slightly better than what we obtained using the naïve Stirling approximation presented above.
- Recall that $\mathbb{X}$ is a r.v. over the sample space $\{0,1\}$. Moreover, we have $\mathbb{P}[\mathbb{X}=1]=p$ and $\mathbb{P}[\mathbb{X}=0]=1-p$. Note that we have $\mathbb{E}[\mathbb{X}]=p$.
- We are studying the r.v.

$$\mathbb{S}_{n,p} = \mathbb{X}^{(1)} + \mathbb{X}^{(2)} + \cdots + \mathbb{X}^{(n)}$$

  Each random variable $\mathbb{X}^{(i)}$ is an independent copy of the random variable $\mathbb{X}$.
- Note that we have $\mathbb{E}\left[\mathbb{S}_{n,p}\right] = n\mathbb{E}[\mathbb{X}] = np$, by the linearity of expectation

Theorem (Chernoff Bound)

$$\mathbb{P}\left[\mathbb{S}_{n,p} \geqslant n(p+\varepsilon)\right] \leqslant \exp\left(-n\mathrm{D}_{\mathrm{KL}}\left(p+\varepsilon, p\right)\right)$$

Before we proceed to proving this result, let us interpret this theorem statement. Suppose $p = 1/2$ and $t = 1/4$. Then, it is exponentially unlikely that $\mathbb{S}_{n,p}$ surpasses $n(1/2 + 1/4) = 3n/4$

Let us begin with the proof.

- We are interested in upper-bounding the probability

$$\mathbb{P}\left[\mathbb{S}_{n,p} \geqslant n(p+\varepsilon)\right]$$

- Note that, for any positive h, we have

$$\mathbb{P}\left[\mathbb{S}_{n,p} \geqslant n(p+\varepsilon)\right] = \mathbb{P}\left[\exp(h\mathbb{S}_{n,p}) \geqslant \exp(hn(p+\varepsilon))\right]$$

  The exact value of h will be determined later. The intuition of using the $\exp(\cdot)$ function is to consider all the moments of $\mathbb{S}_{n,p}$

- Now, we apply Markov inequality to obtain

$$\mathbb{P}\left[\exp(h\mathbb{S}_{n,p}) \geqslant \exp(hn(p+\varepsilon))\right] \leqslant \frac{\mathbb{E}\left[\exp(h\mathbb{S}_{n,p})\right]}{\exp(hn(p+\varepsilon))}$$

- Now, we need an observation. Suppose $\mathbb{A}$ and $\mathbb{B}$ are two independent random variables. Then, we have $\mathbb{E}\left[\exp(\mathbb{A}+\mathbb{B})\right] = \mathbb{E}\left[\exp(\mathbb{A})\right] \cdot \mathbb{E}\left[\exp(\mathbb{B})\right]$. We emphasize that $\mathbb{A}$ and $\mathbb{B}$ have to be independent to apply this result.
- Note that we have $\mathbb{S}_{n,p} = \sum_{i=1}^{n} \mathbb{X}^{(i)}$. So, we can apply the previous observation iteratively to obtain the following result.

$$\frac{\mathbb{E}\left[\exp(h\mathbb{S}_{n,p})\right]}{\exp(hn(p+\varepsilon))} = \frac{\prod_{i=1}^{n} \mathbb{E}\left[\exp(h\mathbb{X}^{(i)})\right]}{\exp(hn(p+\varepsilon))} = \left(\frac{\mathbb{E}\left[\exp(h\mathbb{X})\right]}{\exp(h(p+\varepsilon))}\right)^{n}$$

- Recall that $\mathbb{X}$ is a random variable such that $\mathbb{P}\left[\mathbb{X}=0\right] = 1-p$ and $\mathbb{P}\left[\mathbb{X}=1\right] = p$. So, the random variable $\exp(h\mathbb{X})$ is such that $\mathbb{P}\left[\exp(h\mathbb{X})=1\right] = 1-p$ and $\mathbb{P}\left[\exp(h\mathbb{X})=\exp(h)\right] = p$. Therefore, we can conclude that

$$\mathbb{E}\left[\exp(h\mathbb{X})\right] = (1-p)\cdot 1 + p\cdot \exp(h) = 1-p+p\exp(h)$$

- Substituting this value, we get

$$\left(\frac{\mathbb{E}\left[\exp(h\mathbb{X})\right]}{\exp(h(p+\varepsilon))}\right)^n = \left(\frac{1-p+p\exp(h)}{\exp(h(p+\varepsilon))}\right)^n$$

- So, let us take a pause at this point and recall what we have proven thus far. We have shown that, for all positive h, the following bound holds

$$\mathbb{P}\left[\mathbb{S}_{n,p} \geqslant n(p+\varepsilon)\right] \leqslant \left(\frac{1-p+p\exp(h)}{\exp(h(p+\varepsilon))}\right)^n$$

- To obtain the tightest upper-bound we should use the value of $h = h^*$ that minimizes the right-hand size expression. For simplicity let us make a variable substitution $H = \exp(h)$. Let us define

$$f(H) = \frac{1 - p + pH}{H^{p+\varepsilon}}$$

Our objective is to find $H = H^*$ that minimizes $f(H)$.

- Let us compute $f'(H)$ and solve for $f'(H^*) = 0$. Note that we have

$$f'(H) = \frac{p}{H^{p+\varepsilon}} - \frac{(p+\varepsilon)(1 - p + pH)}{H^{p+\varepsilon+1}}$$

The solution $f'(H^*) = 0$ is given by

$$H^* = \frac{p+\varepsilon}{1 - p - \varepsilon} \cdot \frac{1-p}{p}.$$

We can check that, for $\varepsilon > 0$, we have $H^* > 1$, that is, $h > 0$.
We can consider the second derivative $f''(H)$ to prove that this extremum is a minima.
Instead of computing $f''(H)$, we can use a shortcut technique.
We know that at H^*, the function $f(H)$ either has a maximum or a minimum. Moreover, there is only one extremum of the function $f(H)$. Note that $\lim_{H\to\infty} f(H) = \infty$, so $f(H^*)$ must be a minimum.

- Now, let us substitute the value of h^* to obtain

$$
\begin{aligned}
\mathbb{P}\left[\mathbb{S}_{n,p} \geqslant n(p+\varepsilon)\right] &\leqslant \left(\frac{1-p+\frac{(1-p)(p+\varepsilon)}{1-p-\varepsilon}}{\left(\frac{(1-p)(p+\varepsilon)}{p(1-p-\varepsilon)}\right)^{p+\varepsilon}}\right)^n \\
&= \left(\frac{\frac{1-p}{1-p-\varepsilon}}{\left(\frac{(1-p)(p+\varepsilon)}{p(1-p-\varepsilon)}\right)^{p+\varepsilon}}\right)^n \\
&= \left(\left(\frac{p}{p+\varepsilon}\right)^{p+\varepsilon}\left(\frac{1-p}{1-p-\varepsilon}\right)^{1-p-\varepsilon}\right)^n \\
&= \exp(-n\mathrm{D}_{\mathrm{KL}}\left(p+\varepsilon, p\right))
\end{aligned}
$$

Our objective is to generalize the Chernoff Bound that we proved above. Let us first recall the Chernoff bound result that we proved.

- Let $\mathbb{X}$ be $\mathrm{Bern}(p)$
- Let $\mathbb{S}_{n,p} = \mathbb{X}^{(1)} + \mathbb{X}^{(2)} + \cdots + \mathbb{X}^{(n)}$
- Chernoff bound states that

$$\mathbb{P}\left[\mathbb{S}_{n,p} \geqslant n(p+\varepsilon)\right] \leqslant \exp(-n\mathrm{D}_{\mathrm{KL}}(p+\varepsilon, p))$$

We shall generalize this result in two ways

1. For $1 \leqslant i \leqslant n$, let $\mathbb{X}_i$ be an independent $\mathrm{Bern}(p_i)$ random variable. That is, $\mathbb{X}_i$ be a r.v. over $\{0,1\}$ such that $\mathbb{P}[\mathbb{X}_i = 0] = 1 - p_i$ and $\mathbb{P}[\mathbb{X}_i = 1] = p_i$. Each $\mathbb{X}_i$ is independent of the other $\mathbb{X}_j$s. Let $\mathbb{S}_{n,p} = \mathbb{X}_1 + \mathbb{X}_2 + \cdots + \mathbb{X}_n$, where $p = (p_1 + \cdots + p_n)/n$.
2. For $1 \leqslant i \leqslant n$, let $\mathbb{X}_i$ be a r.v. over $[0,1]$ such that $\mathbb{E}[\mathbb{X}_i] = p_i$.

Despite these two generalizations, the following bound continues to hold true.

$$\mathbb{P}\left[\mathbb{S}_{n,p} \geqslant n(p+\varepsilon)\right] \leqslant \exp(-n\mathrm{D}_{\mathrm{KL}}(p+\varepsilon, p))$$

- Let $X_1, X_2, \ldots \mathbb{X}_n$ be independent random variables such that $\mathbb{X}_i = \mathrm{Bern}\,(p_i)$, for $1 \leqslant i \leqslant n$
- Let $p := (p_1 + p_2 + \cdots + p_n)/n$
- Define $\mathbb{S}_{n,p} = \mathbb{X}_1 + \mathbb{X}_2 + \cdots + \mathbb{X}_n$
- We bound the following probability. For any $H > 1$, we have

$$\mathbb{P}\left[\mathbb{S}_{n,p} \geqslant n(p+\varepsilon)\right] = \mathbb{P}\left[H^{\mathbb{S}_{n,p}} \geqslant H^{n(p+\varepsilon)}\right]$$

- Now, we apply the Markov inequality

$$\mathbb{P}\left[H^{\mathbb{S}_{n,p}} \geqslant H^{n(p+\varepsilon)}\right] \leqslant \frac{\mathbb{E}\left[H^{\mathbb{S}_{n,p}}\right]}{H^{n(p+\varepsilon)}} = \frac{\mathbb{E}\left[H^{\sum_{i=1}^{n} \mathbb{X}_i}\right]}{H^{n(p+\varepsilon)}} = \frac{\mathbb{E}\left[\prod_{i=1}^{n} H^{\mathbb{X}_i}\right]}{H^{n(p+\varepsilon)}}$$

- Since, each $\mathbb{X}_i$ are independent of other $\mathbb{X}_j$s, we have

$$\frac{\mathbb{E}\left[\prod_{i=1}^{n} H^{\mathbb{X}_i}\right]}{H^{n(p+\varepsilon)}} = \frac{\prod_{i=1}^{n} \mathbb{E}\left[H^{\mathbb{X}_i}\right]}{H^{n(p+\varepsilon)}} = \frac{\prod_{i=1}^{n} 1 - p_i + p_i H}{H^{n(p+\varepsilon)}}$$

- We apply the AM-GM inequality to conclude that

$$\prod_{i=1}^{n} 1 - p_i + p_i H \leqslant \left(\frac{\sum_{i=1}^{n} 1 - p_i + p_i H}{n}\right)^n$$

Equality holds if and only if all $p_i = p$. This bound can now be substituted to conclude

$$\frac{\mathbb{E}\left[\prod_{i=1}^{n} H^{\mathbb{X}_i}\right]}{H^{n(p+\varepsilon)}} \leqslant \left(\frac{1 - p + pH}{H^{p+\varepsilon}}\right)^n$$

- This is identical to the bound that we had in the Chernoff bound proof. We can use the following choice of H in the bound above to obtain the tightest possible bound

$$H^* = \frac{(p+\varepsilon)(1-p)}{p(1-p-\varepsilon)}$$

So, we get the bound

$$\mathbb{P}\left[\mathbb{S}_{n,p} \geqslant n(p+\varepsilon)\right] \leqslant \exp(-n\mathrm{D}_{\mathrm{KL}}\left(p+\varepsilon, p\right))$$

- Let $1 \leqslant \mathbb{X}_i \leqslant 1$ be a r.v. such that $\mathbb{E}[\mathbb{X}_i] = p_i$ and each $\mathbb{X}_i$ is independent of other $\mathbb{X}_j$s
- Just like the previous setting, we have $\mathbb{S}_{n,p} = \mathbb{X}_1 + \mathbb{X}_2 + \cdots + \mathbb{X}_n$, where $p = (p_1 + p_2 + \cdots + p_n)/n$
- Note that if we prove the following bound, then we shall be done

$$\mathbb{E}\left[H^{\mathbb{X}_i}\right] \leqslant 1 - p_i + p_i H$$

We can use this bound in the previous proof and arrive at the identical upper-bound.

The proof follows from the following

$$
\begin{aligned}
\mathbb{E}\left[H^{\mathbb{X}_i}\right] &= \sum_{x\in[0,1]} \mathbb{P}\left[\mathbb{X}_i = x\right] \cdot H^x \\
&= \sum_{x\in[0,1]} \mathbb{P}\left[\mathbb{X}_i = x\right] \cdot H^{(1-x)\cdot 0 + x\cdot 1} \\
&\leqslant \sum_{x\in[0,1]} \mathbb{P}\left[\mathbb{X}_i = x\right] \cdot \left((1-x)\cdot H^0 + x \cdot H^1\right), && \text{(By Jensen's)} \\
&= \sum_{x\in[0,1]} \mathbb{P}\left[\mathbb{X}_i = x\right] \cdot (1 - x + xH) \\
&= \sum_{x\in[0,1]} \mathbb{P}\left[\mathbb{X}_i = x\right] - \sum_{x\in[0,1]} \mathbb{P}\left[\mathbb{X}_i = x\right] \cdot x + H \sum_{x\in[0,1]} \mathbb{P}\left[\mathbb{X}_i = x\right] \cdot x \\
&= 1 - p_i + p_i H, && \text{(Because } \mathbb{E}\left[\mathbb{X}_i\right] = p_i\text{)}
\end{aligned}
$$

The appendix provides additional intuition for this analysis.

Conclusion

- Let $0 \leqslant \mathbb{X}_i \leqslant 1$ are independent random variables, for $1 \leqslant i \leqslant n$. Let $p_i = \mathbb{E}[\mathbb{X}_i]$, for $1 \leqslant i \leqslant n$. Define $\mathbb{S}_{n,p} := \mathbb{X}_1 + \mathbb{X}_2 + \cdots + \mathbb{X}_n$, where $p := (p_1 + \cdots + p_n)/n$.

Theorem (Chernoff Bound)

$$\mathbb{P}\left[\mathbb{S}_{n,p} \geqslant n(p+\varepsilon)\right] \leqslant \exp(-n\mathrm{D}_{\mathrm{KL}}(p+\varepsilon, p))$$

- **Objective of the next lecture.** We shall obtain easier to compute, albeit weaker, upper bounds on this probability. These bounds shall rely on the following inequalities
  1. $\mathrm{D}_{\mathrm{KL}}(p+\varepsilon, p) \geqslant 2\varepsilon^2$,
  2. $\mathrm{D}_{\mathrm{KL}}\left(p(1+\varepsilon), p\right) \geqslant \frac{p\varepsilon^2}{2\left(1+\varepsilon/3\right)}$, and
  3. $\mathrm{D}_{\mathrm{KL}}\left(1-p(1-\varepsilon), 1-p\right) \geqslant p\varepsilon^2/2$.

  Check them out at:

  `https://www.desmos.com/calculator/pyessio3v2`

- Let $\mathbb{X}$ be an r.v. over $[a, b]$ such that $\mathbb{E}[\mathbb{X}] = \mu$
- Let $f\colon \mathbb{R} \to \mathbb{R}$ be a concave upwards function (that is, it looks like $f(x) = x^2$)
- Jensen's inequality states that $f(\mathbb{E}[\mathbb{X}]) \leqslant \mathbb{E}\left[f(\mathbb{X})\right]$, and equality holds if and only if $\mathbb{X}$ has its entire probability mass at μ. Therefore, we can conclude that $f(\mu) \leqslant \mathbb{E}\left[f(\mathbb{X})\right]$
- So, we have a lower-bound on $\mathbb{E}\left[f(\mathbb{X})\right]$. Now, we are interested in obtaining an upper-bound on $\mathbb{E}\left[f(\mathbb{X})\right]$
- For the upper-bound note that is $\mathbb{X}$ deposits more probability mass away from μ, then $\mathbb{E}\left[f(\mathbb{X})\right]$ increases. In fact, increasing the mass further away increases $\mathbb{E}\left[f(\mathbb{X})\right]$ more. So, the maximum value of $\mathbb{E}\left[f(\mathbb{X})\right]$ is achieved when $\mathbb{X}$ deposits the entire probability mass either at a or b only. Let us find such a probability distribution under the constraint that $\mathbb{E}[\mathbb{X}] = \mu$

- Suppose $\mathbb{P}[\mathbb{X}^* = a] = p$. Then, we have $\mathbb{P}[\mathbb{X}^* = b] = 1 - p$. Further, the constraint $\mathbb{E}[\mathbb{X}^*] = \mu$ becomes $pa + (1-p)b = \mu$. Solving, we get

$$p = \frac{b - \mu}{b - a}$$

Therefore, we get $1 - p = \frac{\mu - a}{b - a}$. For this probability, we get

$$\mathbb{E}\left[f(\mathbb{X}^*)\right] = \frac{b - \mu}{b - a} f(a) + \frac{\mu - a}{b - a} f(b)$$

So, we expect the following bound to hold for a general r.v. $\mathbb{X}$

$$\mathbb{E}\left[f(\mathbb{X})\right] \leqslant \mathbb{E}\left[f(\mathbb{X}^*)\right] = \frac{b - \mu}{b - a} f(a) + \frac{\mu - a}{b - a} f(b)$$

This is not a formal proof. Let us prove this intuition formally.

- Let $\mathbb{X}$ be an r.v. over $[a, b]$ with $\mathbb{E}\left[\mathbb{X}\right] = \mu$. Note that by Jensen's inequality, we have

$$f(x) = f\left(\frac{b-x}{b-a}a + \frac{x-a}{b-a}b\right) \leqslant \frac{b-x}{b-a}f(a) + \frac{x-a}{b-a}f(b)$$

  Now, we take expectation on both sides to conclude that

$$\begin{aligned}\mathbb{E}\left[f(\mathbb{X})\right] &\leqslant \mathbb{E}\left[\frac{b-\mathbb{X}}{b-a}f(a) + \frac{\mathbb{X}-a}{b-a}f(b)\right] \\ &= \frac{b-\mathbb{E}\left[\mathbb{X}\right]}{b-a}f(a) + \frac{\mathbb{E}\left[\mathbb{X}\right]-a}{b-a}f(b) \\ &= \frac{b-\mu}{b-a}f(a) + \frac{\mu-a}{b-a}f(b)\end{aligned}$$

- To conclude, we have the following bound.

$$f(\mu) \leqslant \mathbb{E}\left[f(\mathbb{X})\right] \leqslant \frac{b-\mu}{b-a}f(a) + \frac{\mu-a}{b-a}f(b)$$