Question 1. We are going to prove part of nicer version of the Chernoff Bounds. Prove the following inequalities

i) For $0 < \delta < 1$,
\[
\frac{e^\delta}{(1 + \delta)(1 + \delta)} \leq e^{-\delta^2/3}.
\]

ii) For $0 < \delta < 1$,
\[
\frac{e^{-\delta}}{(1 - \delta)(1 - \delta)} \leq e^{-\delta^2/2}.
\]

iii) Using the Chernoff-Bounds (Slides 8 and 22, Lecture 5) deduce the second part of the Nicer Chernoff Bounds in Slide 23 of Lecture 5.

Solution: i) By taking logarithm the problem is equivalent to prove that
\[
\delta - (1 + \delta) \log(1 + \delta) + \delta^2/3 \leq 0
\]
write $f(\delta) = 2\delta/3 - \log(1 + \delta)$. We will check that $f(\delta) \geq 0$ for $\delta \in [0, 1]$.

A straightforward computation shows that
- $f(0) = 0$,
- $f'(\delta) = 2\delta/3 - \log(1 + \delta)$.
- $f'(0) = 0$ and $f'(1) = 2/3 - \log(2) < 0$

We deduce that $f'(\delta) \leq 0$ for $\delta \in [0, 1]$. This holds because the function $\log(1 + \delta)$ only intersects with the line $2x/3$ in exactly two points, one of them is 0 and the other has to be after 1 (otherwise $f'(1) \geq 0$). Finally,
\[
f(\delta) = \int_0^\delta f'(x)dx - f(0) \leq \int_0^\delta 0dx - 0 = 0
\]
ii) it is done the same way. Alternatively, use the Taylor expansion of $\log(1 - \delta)$,
\[
\log(1 - \delta) = -\delta - \delta^2/2 - \delta^3/3 - \ldots \leq -\delta - \delta^2/2.
\]

Question 2. Chernoff Bounds for other random variables.

i) Let $X$ be a Poisson random variable of mean $\mu$. Prove that
\[
E[e^{\lambda X}] = e^{\lambda(\mu - 1)}
\]
and deduce that for $t \geq \mu$ and for $\delta \in (0, 1)$
\[
P[X \geq t] \leq e^{-\mu} \left( \frac{\mu e^t}{t} \right)^t
\]
and
\[
P[X \geq (1 + \delta)\mu] \leq e^{-\delta^2 \mu/3},
\]
and the corresponding lower tails.
ii) Let \(X\) be a Normal random variable of mean \(\mu\) and variance \(\sigma^2\). Prove that
\[
E[e^{\lambda X}] = e^{\mu \lambda + \sigma^2 \lambda^2 / 2},
\]
and deduce that for \(t > \mu\)
\[
P[X \geq t] \leq \exp \left[ -\frac{(t - \mu)^2}{2\sigma^2} \right].
\]
Find the corresponding lower tail.

Solution: i) Recall that \(P[X = k] = \frac{\mu^k e^{-\mu}}{k!}\). Then
\[
E[e^{\lambda X}] = \sum_{k=0}^{\infty} e^{\lambda k} \frac{\mu^k e^{-\mu}}{k!} = \sum_{k=0}^{\infty} \frac{(e^\lambda \mu)^k e^{-\mu}}{k!} = \frac{e^{-\mu}}{e^{-\mu e^\lambda}} \sum_{k=0}^{\infty} \frac{(e^\lambda \mu)^k}{k!} = (1)
\]
Note that \(\frac{(e^\lambda \mu)^k}{k!}\) is the probability that a Poisson random variable of mean \(\mu e^\lambda\) takes value \(k\), therefore the summation on the RHS of (1) is 1. We conclude
\[
E[e^{\lambda X}] = \frac{e^{-\mu}}{e^{-\mu e^\lambda}} = e^{\mu (e^\lambda - 1)}
\]
The computation of \(P[X \geq t]\) is exactly as in Lecture 5 slide 20.

ii) The density function of a normal random variable with mean \(\mu\) and variance \(\sigma\) is given by
\[
\frac{1}{\sqrt{2\pi\sigma^2}} \int_{\|} \exp \left( -\frac{(x-\mu)^2}{2\sigma^2}\right) dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\|} \exp \left( -\frac{(x-\mu)^2 - 2\sigma^2 \lambda x}{2\sigma^2}\right) dx
\]
We analyse the expression inside the exponential
\[
(x-\mu)^2 - 2\sigma^2 \lambda x = x^2 - 2(\mu + \sigma^2 x) + \mu^2 = (x-(\mu + \sigma^2))^2 + \mu^2 - (\mu + \sigma^2)^2
\]
Therefore
\[
E[e^{\lambda X}] = \exp \left(-\frac{\mu^2 - (\mu + \sigma^2)^2}{2\sigma^2}\right) \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\|} \exp \left(-\frac{(x-(\mu + \sigma^2))^2}{2\sigma^2}\right) dx
\]
Note that \(\exp \left(-\frac{(x-(\mu + \sigma^2))^2}{2\sigma^2}\right)\) is the density function of a normal random variable with mean \(\mu + \sigma^2\) and variance \(\sigma^2\). Therefore we have
\[
E[e^{\lambda X}] = \exp \left(-\frac{\mu^2 - (\mu + \sigma^2)^2}{2\sigma^2}\right) = \exp \left(\mu \lambda + \frac{\sigma^2 \lambda^2}{2}\right)
\]
To compute \(P[X \geq t]\) use the recipe (Lecture 5, Slide 21).

**Question 3.** Show properties 2-6 of slide 19 of Lecture 6.

**Solution:** Property 4: Let \(a\) be a possible value of \(X\). Since \(X\) is independent of \(Y\) we have \(P[Y = y|X = a] = P[Y = y]\) then
\[
E[Y|X = a] = \sum_y y P[Y = y|X = a] \overset{\text{indep}}{=} \sum_y y P[Y = y] = E[Y],
\]
deducing that \(E[Y|X] = E[Y]\).
Property 5: Let \( a \) be a possible value of \( X \), then

\[
\mathbb{E}[YZ|X = a] = \mathbb{E}[F(X)Z|X = a] = \sum_x \sum_z F(x)z \mathbb{P}[X = x, Z = z|X = a]
\]

\[
= F(a) \sum_z z \mathbb{P}[Z = z|X = a] = F(a) \mathbb{E}[Z|X = a],
\]

deducing that \( \mathbb{E}[YZ|X] = Y \mathbb{E}[Z|X] \).

---

**Question 4.** Let \( X_1, \ldots, X_n \) be independent discrete random variables and let \( Z = f(X_1, \ldots, X_n) \) for some function \( f \). Prove that

\[
\mathbb{E}[Z|X_1, \ldots, X_i] = \sum_{x_1, x_2, \ldots, x_n} f(x_1, \ldots, x_i, \ldots, x_n) \mathbb{P}[X_1 = x_1, \ldots, X_i = x_i | X_i = x_i, \ldots, X_n = x_n]
\]

**Solution:** Let \((a_1, \ldots, a_n)\) be a possible value of \((X_1, \ldots, X_n)\), then

\[
\mathbb{E}[Z|X_1 = a_1, \ldots, X_n = a_n] = \sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} f(x_1, \ldots, x_n) \mathbb{P}[X_1 = x_1, \ldots, X_i = x_i | X_i = a_i, \ldots, X_n = a_n]
\]

\[
= \sum_{x_1, x_2, \ldots, x_n} f(a_1, \ldots, a_i, x_i+1, \ldots, x_n) \mathbb{P}[X_i+1 = x_i+1, \ldots, X_n = x_n].
\]

---

**Question 5.** Conditional Variance. Define the conditional variance of \( Y \) given \( X \) as

\[
\text{Var}[Y|X] = \mathbb{E}[ (Y - \mathbb{E}[Y|X])^2 | X].
\]

1. Prove that \( \text{Var}[Y] = \mathbb{E}[\text{Var}[Y|X]] + \text{Var}[\mathbb{E}[Y|X]] \)

2. Consider \( n \) bins and a random number \( M \) of balls, where \( \mathbb{E}[M] = \mu \) and \( \text{Var}[M] = \sigma^2 \). Compute the variance of the number of balls that are assigned to the first bin.

**Solution:** Remember that \( \text{Var}[Y] = \mathbb{E}[ (Y - \mathbb{E}[Y])^2 ] \), and, equivalently, \( \text{Var}[Y] = \mathbb{E}[Y^2] - \mathbb{E}[Y]^2 \)

By definition we have that

\[
\mathbb{E}[\text{Var}[Y|X]] = \mathbb{E}[ (Y - \mathbb{E}[Y|X])^2 | X] \]

\[
\overset{p1}{=} \mathbb{E}[ (Y - \mathbb{E}[Y|X])^2 ] = \mathbb{E}[Y^2 - 2Y\mathbb{E}[Y|X] + \mathbb{E}[Y|X]^2]
\]

(\(p1\) refers to the properties of Lecture 6, slide 19). By \(p1\) we get

\[
\mathbb{E}[Y\mathbb{E}[Y|X]] \overset{p2}{=} \mathbb{E}[\mathbb{E}[Y|X] |X]] \overset{p3}{=} \mathbb{E}[\mathbb{E}[Y|X]^2]
\]

by linearity of conditional expectation (\(p3\)) we get

\[
\mathbb{E}[[\text{Var}[Y|X]] = \mathbb{E}[Y^2] - \mathbb{E}[\mathbb{E}[Y|X]^2]
\]

(2)

Also, note that \( \mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Y] \) then

\[\text{Var}[\mathbb{E}[Y|X]] = \mathbb{E}[\mathbb{E}[Y|X]^2] - \mathbb{E}[Y]^2 \]

(3)

By adding equations (2) and (3) we get the result.

For the second part, let \( X \) be the number of balls that are assigned to the first bin. We compute \( \text{Var}[X|X] = \mathbb{E}[\text{Var}[X|M]] + \text{Var}[\mathbb{E}[X|M]] \). In lecture 6 slide 24 we computed \( \mathbb{E}[X|M] = \sum_{i=1}^{\infty} 1_{(i \leq M)} = M/n \). Moreover, by definition of conditional variance, we get

\[
\text{Var}[X|M] = \mathbb{E}[ (X - M/n)^2 | M] \overset{p3+p5}{=} \mathbb{E}[X^2|M] - 2(M/n)\mathbb{E}[X|M] + (M/n)^2
\]

\[
= \mathbb{E}[X^2|M] - (M/n)^2
\]

3
We just need to compute $E[X^2|M]$. Let $X_i$ be 1 if ball number $i$ is assigned to the first bin, otherwise $X_i$ is 0. Then $X = \sum_{i=1}^{\infty} X_i 1_{\{i \leq M\}}$ and therefore

$$X^2 = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} X_i X_j 1_{\{i,j \leq M\}}$$

Note that $X_i^2 = X_i$, then

$$E[X^2|M] = \sum_{i=1}^{\infty} E[X_i|M] + 2 \sum_{1<i<j<\infty} E[X_i X_j|M] 1_{\{i,j \leq M\}}$$

Finally, use that the location of a ball is independent of how many balls we assigned in total. Therefore $E[X_i|M] = 1/n$ and $E[X_i X_j|M] = 1/n^2$ for $i \neq j$. We conclude that

$$\text{Var}[X|M] = M/n + M(M-1)/n^2 - (M/n)^2 = M/n - M/n^2$$

and $E[\text{Var}[X|M]] = (\mu/n) (1 - \frac{1}{n})$

On the other hand, remember that $E[X|M] = M/n$. Then

$$\text{Var}[E[X|M]] = \text{Var}[M/n] = \frac{1}{n^2} \text{Var}[M] = \sigma^2/n^2.$$

By adding $\text{Var}[E[X|M]]$ and $E[\text{Var}[X|M]]$ we get the result.

**Question 6.** Consider a coin that shows head with probability $p$. What is the expected number of flips required to observe a run of $n$ consecutive heads?

**Solution:** Done in exercise class.

**Question 7.** Let $X_1, \ldots, X_n$ i.i.d. samples from a distribution of interest. We know that $E[X_i] = \mu$ and $\text{Var}[X_i] = \sigma^2$ for all $i$, but we do not know the exact values of $\mu$ nor $\sigma^2$. We are given the mission to find an estimate $\hat{\mu}$ of the actual mean $\mu$. We want the estimate $\hat{\mu}$ to satisfy the $(\delta, \varepsilon)$ condition: given $\varepsilon$, we want that $\hat{\mu} \in [\mu - \varepsilon \sigma, \mu + \varepsilon \sigma]$ with probability at least $1 - \delta$. How many data points $X_i$ do we need to build an estimator satisfying the $(\delta, \varepsilon)$ condition?

- In a first attempt we can just deliver $\hat{\mu} = \frac{\sum_{i=1}^{n} X_i}{n}$, nevertheless, we cannot guarantee a good behaviour of such estimator, as we do not have enough information to compute a Chernoff Bound for it.

1. Prove that with $m = \lceil \frac{10}{\varepsilon^2} \rceil$ data points, we have that $\hat{\mu}_m = (\sum_{i=1}^{m} X_i) / m$ satisfies the $(1/10, \varepsilon)$ condition.

2. Write an algorithm that uses at most $O\left(\frac{\log(\delta^{-1})}{\varepsilon^2}\right)$ data points to build an estimate of $\mu$ satisfying the $(\delta, \varepsilon)$ condition.

**Hint.**

Q6: Recall how we deduce the expectation of a geometric in class.

Q7: For 2, consider batches of size $m = \lceil \frac{10}{\varepsilon^2} \rceil$. What can you say about more than half of them?

**Solution:** Done in exercise class.