Tail and Concentration Inequalities

From here on, we use $1_A$ to denote the indicator variable for event $A$, i.e. $1_A = 1$ if $A$ holds and $1_A = 0$ otherwise. Our presentation follows closely the first chapter of [1].

1 Markov Inequality

**Theorem 1.1.** If $X$ is a r.v. taking only non-negative values, $\mu = \mathbb{E}[X]$, then $\forall a > 0$

$$\text{Prob}[X \geq a] \leq \frac{\mu}{a}. \tag{1}$$

**Proof.** From the simple fact that $a1_{\{X \geq a\}} \leq X$, taking expectation on both sides we get $a\mathbb{E}[1_{\{X \geq a\}}] \leq \mu$, which implies (1). $\square$

**Problem 1.** Use Markov inequality to prove the following. Let $c \geq 1$ be an arbitrary constant. If $n$ people have a total of $d$ dollars, then there are at least $(1 - 1/c)n$ of them each of whom has less than $cd/n$ dollars.

(You can easily prove the above statement from first principle. However, please set up a probability space, a random variable, and use Markov inequality to prove it. It is instructive!)

2 Chebyshev Inequality

**Theorem 2.1** (Two-sided Chebyshev’s Inequality). If $X$ is a r.v. with mean $\mu$ and variance $\sigma^2$, then $\forall a > 0$,

$$\text{Prob}[|X - \mu| \geq a] \leq \frac{\sigma^2}{a^2}.$$

**Proof.** Let $Y = (X - \mu)^2$, then $\mathbb{E}[Y] = \sigma^2$ and $Y$ is a non-negative r.v.. From Markov inequality (1) we have

$$\text{Prob}[|X - \mu| \geq a] = \text{Prob}[Y \geq a^2] \leq \frac{\sigma^2}{a^2}. \square$$

The one-sided versions of Chebyshev inequality are sometimes called Cantelli inequality.

**Theorem 2.2** (One-sided Chebyshev’s Inequality). Let $X$ be a r.v. with $\mathbb{E}[X] = \mu$ and $\text{Var}[X] = \sigma^2$, then for all $a > 0$,

$$\text{Prob}[X \geq \mu + a] \leq \frac{\sigma^2}{\sigma^2 + a^2} \tag{2}$$

$$\text{Prob}[X \leq \mu - a] \leq \frac{\sigma^2}{\sigma^2 + a^2}. \tag{3}$$
Proof. Let $Y = X - \mu$, then $E[Y] = 0$ and $\text{Var}[Y] = \text{Var}[X] = \sigma^2$. (Why?) Thus, for any $t$ such that $t + a > 0$ we have

$$\begin{align*}
\text{Prob}[Y \geq a] &= \text{Prob}[Y + t \geq a + t] \\
&= \text{Prob}\left[\frac{Y + t}{a + t} \geq 1\right] \\
&\leq \text{Prob}\left[\left(\frac{Y + t}{a + t}\right)^2 \geq 1\right] \\
&\leq \mathbb{E}\left[\left(\frac{Y + t}{a + t}\right)^2\right] \\
&= \frac{\sigma^2 + t^2}{(a + t)^2}.
\end{align*}$$

The second inequality follows from Markov inequality. The above analysis holds for any $t$ such that $t + a > 0$. We pick $t$ to minimize the right hand side, which is $t = \sigma^2/a > 0$. That proves (2). \qed

**Problem 2.** Prove (3).

### 3 Bernstein, Chernoff, Hoeffding

#### 3.1 The basic bound using Bernstein’s trick

Let us consider the simplest case, and then relax assumptions one by one. For $i \in [n]$, let $X_i$ be i.i.d. random variables which are all Bernoulli with parameter $p$. Let $X = \sum_{i=1}^{n} X_i$. Then, $E[X] = np$. We will prove that, as $n$ gets large $X$ is “far” from $E[X]$ with exponentially low probability.

Let $m$ be such that $np < m < n$, we want to bound $\text{Prob}[X \geq m]$. For notational convenience, let $q = 1 - p$. Bernstein taught us the following trick. For any $t > 0$ the following holds.

$$\begin{align*}
\text{Prob}[X \geq m] &= \text{Prob}[tX \geq tm] \\
&= \text{Prob}\left[e^{tX} \geq e^{tm}\right] \\
&\leq \mathbb{E}\left[e^{tX}\right] \\
&= \mathbb{E}\left[\prod_{i=1}^{n} e^{tX_i}\right] \\
&= \prod_{i=1}^{n} \mathbb{E}\left[e^{tX_i}\right] \\
&= \prod_{i=1}^{n} (pe^t + q) \\
&= \left(pe^t + q\right)^{n}.
\end{align*}$$

The inequality on the third line follows from Markov inequality (1). Naturally, we set $t$ to minimize the right hand side, which is

$$t_0 = \ln \frac{mq}{(n - m)p} > 0.$$
Problem 3. \( np < m < np \) we have

\[
E_{(\text{Bernstein-Chernoff-Hoeffding})} \text{can rewrite (4) simply as}
\]

\[
\text{(0 is below the linear segment connecting the points}
\]

\[
\text{The problem is, we no longer can compute } E.
\]

\[
\text{Thus, (5) holds when the}
\]

\[
\text{a negative real numbers}
\]

\[
\text{The second inequality is due to the geometric-arithmetic means inequality, which states that, for any non-negative real numbers } a_1, \ldots, a_n \text{ we have}
\]

\[
an_1 \cdots a_n \leq \left( \frac{a_1 + \cdots + a_n}{n} \right)^n.
\]

Thus, (5) holds when the \( X_i \) are Bernoulli and they don’t have to be identically distributed.

Finally, consider a fairly general case when the \( X_i \) do not even have to be discrete variables. Suppose the \( X_i \) are independent random variables where \( E[X_i] = p_i \) and \( X_i \in [0, 1] \) for all \( i \). Again, let \( p = \sum_i p_i/n \) and \( q = 1 - p \). Bernstein’s trick leads us to

\[
\text{Prob}[X \geq m] \leq \prod_{i=1}^{n} e^{tm} E\left[ e^{tX_i} \right].
\]

The problem is, we no longer can compute \( E\left[ e^{tX_i} \right] \) because we don’t know the \( X_i \)’s distributions. Hoeffding taught us another trick. For \( t > 0 \), the function \( f(x) = e^{tx} \) is convex. Hence, the curve of \( f(x) \) inside \([0,1]\) is below the linear segment connecting the points \((0, f(0))\) and \((1, f(1))\). The segment’s equation is

\[
y = (f(1) - f(0))x + f(0) = (e^{t} - 1)x + 1 = e^{t}x + (1 - x).
\]

Hence,

\[
E\left[ e^{tX_i} \right] \leq E\left[ e^{tX_i} + (1 - X_i) \right] = p_i e^{t} + q_i.
\]

We thus obtain (4) as before. Overall, we just proved the following theorem.

**Theorem 3.1 (Bernstein-Chernoff-Hoeffding).** Let \( X_i \in [0, 1] \) be independent random variables where \( E[X_i] = p_i, i \in [n] \). Let \( X = \sum_{i=1}^{n} X_i, p = \sum_{i=1}^{n} p_i/n \) and \( q = 1 - p \). Then, for any \( m \) such that \( np < m < n \) we have

\[
\text{Prob}[X \geq m] \leq e^{-n RE(m/n\|p)}.
\]

**Problem 3.** Let \( X_i \in [0,1] \) be independent random variables where \( E[X_i] = p_i, i \in [n] \). Let \( X = \sum_{i=1}^{n} X_i, p = \sum_{i=1}^{n} p_i/n \) and \( q = 1 - p \). Prove that, for any \( m \) such that \( 0 < m < np \) we have

\[
\text{Prob}[X \leq m] \leq e^{-n RE(m/n\|p)}.
\]
3.2 Instantiations

There are a variety of different bounds we can get out of (6) and (7).

**Theorem 3.2** (Hoeffding Bounds). Let $X_i \in [0, 1]$ be independent random variables where $E[X_i] = p_i, i \in [n]$. Let $X = \sum_{i=1}^{n} X_i$. Then, for any $t > 0$ we have

\[
\text{Prob}[X \geq E[X] + t] \leq e^{-2t^2/n}.
\]

and

\[
\text{Prob}[X \leq E[X] - t] \leq e^{-2t^2/n}.
\]

**Proof.** We prove (8), leaving (9) as an exercise. Let $p = \sum_{i=1}^{n} p_i/n$ and $q = 1 - p$. WLOG, we assume $0 < p < 1$. Define $m = (p + x)n$, where $0 < x < q = 1 - p$, so that $np < m < n$. Also, define

\[
f(x) = \text{RE} \left( \frac{m}{n} \parallel p \right) = \text{RE} (p + x \parallel p) = (p + x) \ln \frac{p + x}{p} + (q - x) \ln \frac{q - x}{q}.
\]

Routine manipulations give

\[
f'(x) = \ln \frac{p + x}{p} - \ln \frac{q - x}{q},
\]

\[
f''(x) = \frac{1}{(p + x)(q - x)}
\]

By Taylor’s expansion, for any $x \in [0, 1]$ there is some $\xi \in [0, x]$ such that

\[
f(x) = f(0) + xf'(0) + \frac{1}{2}x^2f''(\xi) = \frac{1}{2}x^2 \frac{1}{(p + \xi)(q - \xi)} \geq 2x^2.
\]

The last inequality follows from the fact that $(p + \xi)(q - \xi) \leq ((p + q)/2)^2 = 1/4$. Finally, set $x = t/n$. Then, $m = np + t = E[X] + t$. From (6) we get

\[
\text{Prob}[X \geq E[X] + t] \leq e^{-nf(x)} \leq e^{-2x^2n} = e^{-2t^2/n}.
\]

\[
\square
\]

**Problem 4.** Prove (9).

**Theorem 3.3** (Chernoff Bounds). Let $X_i \in [0, 1]$ be independent random variables where $E[X_i] = p_i, i \in [n]$. Let $X = \sum_{i=1}^{n} X_i$. Then,

(i) For any $0 < \delta \leq 1$,

\[
\text{Prob}[X \geq (1 + \delta)E[X]] \leq e^{-E[X]\delta^2/3}.
\]

(ii) For any $0 < \delta < 1$,

\[
\text{Prob}[X \leq (1 - \delta)E[X]] \leq e^{-E[X]\delta^2/2}.
\]

(iii) If $t > 2E[X]$, then

\[
\text{Prob}[X \geq t] \leq 2^{-t}.
\]
Proof. To bound the upper tail, we apply (6) with \( m = (p+\delta)p n \). Without loss of generality, we can assume \( m < n \), or equivalently \( \delta < q/p \). In particular, we will analyze the function

\[
g(x) = RE(p + xp|p) = (1 + x)p \ln(1 + x) + (q - px) \ln \frac{q - px}{q},
\]

for \( 0 < x \leq \min\{q/p, 1\} \). First, observe that

\[
\ln \frac{q - px}{q} = \ln \left(1 + \frac{px}{q - px}\right) \leq \frac{px}{q - px}.
\]

Hence, \( (q - px) \ln \frac{q - px}{q} \geq -px \), from which we can infer that

\[
g(x) \geq (1 + x)p \ln(1 + x) - px = p[(1 + x) \ln(1 + x) - x].
\]

Now, define

\[
h(x) = (1 + x) \ln(1 + x) - x - x^2/3.
\]

Then,

\[
h'(x) = \ln(1 + x) - 2x/3
\]

\[
h''(x) = \frac{1}{1 + x} - 2/3.
\]

Thus, \( 1/2 \) is a local extremum of \( h'(x) \). Note that \( h'(0) = 0, h'(1/2) \approx 0.07 > 0, \) and \( h'(1) \approx 0.026 > 0 \). Hence, \( h'(x) \geq 0 \) for all \( x \in (0, 1) \]. The function \( h(x) \) is thus non-decreasing. Hence, \( h(x) \geq h(0) = 0 \) for all \( x \in [0, 1] \). Consequently,

\[
g(x) \geq p[(1 + x) \ln(1 + x) - x] \geq px^2/3
\]

for all \( x \in [0, 1] \). Thus, from (6) we have

\[
\text{Prob}[X \geq (1 + \delta)E[X]] = \text{Prob}[X \geq (1 + \delta)pn] \leq e^{-n \cdot g(\delta)} \leq e^{-\delta^2 E[X]}/3.
\]

Problem 5. Prove (12).

Problem 6. Let \( X_i \in [0, 1] \) be independent random variables where \( E[X_i] = p_i, i \in [n] \). Let \( X = \sum_{i=1}^{n} X_i \), and \( \mu = E[X] \). Prove the following

(i) For any \( \delta, t > 0 \) we have

\[
\text{Prob}[X \geq (1 + \delta)\mu] \leq \left(\frac{e^{e^t-1}}{e^t(1+\delta)}\right)^\mu
\]

(Hint: repeat the basic structure of the proof using Bernstein’s trick. Then, because \( 1 + x \leq e^x \) we can apply \( 1 + p_i e^t + 1 - p_i \leq e^{p_i e^t - p_i} \).)

(ii) Show that, for any \( \delta > 0 \) we have

\[
\text{Prob}[X \geq (1 + \delta)\mu] \leq \left(\frac{e^\delta}{(1 + \delta)^{1+\delta}}\right)^\mu
\]
(iii) Prove that, for any $t > 2eE[X],$

$$\text{Prob}[X \geq t] \leq 2^{-t}.$$  

**Problem 7.** Let $X_i \in [a_i, b_i]$ be independent random variables where $a_i, b_i$ are real numbers. Let $X = \sum_{i=1}^{n} X_i.$ Repeat the basic proof structure to show a slightly more general Hoeffding bounds:

$$\text{Prob}[X - E[X] \geq t] \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^{n} (a_i - b_i)^2}\right)$$

$$\text{Prob}[X - E[X] \leq -t] \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^{n} (a_i - b_i)^2}\right)$$

**Problem 8.** Prove that, for any $0 \leq \alpha \leq n,$

$$\sum_{0 \leq k \leq \alpha n} \binom{n}{k} \leq 2^{H(\alpha)n},$$

where $H(\alpha) = -\alpha \log_2 \alpha - (1 - \alpha) \log_2 (1 - \alpha)$ is the binary entropy function.

**References**