Learning Goals

- Define variance and standard deviation
- State Chebyshev inequality and Chernoff inequality
- Compare the conditions and strengths of Markov, Chebyshev and Chernoff inequalities
- Understand the main idea and steps in its proof
- Have intuitive understanding of the bounds given by simplified forms of Chernoff inequality

Chebyshev Inequality

Definition

The *variance* of a random variable X is $\mathrm{Var}[X] := \mathbf{E}[(X - \mathbf{E}[X])^2] = \mathbf{E}[X^2] - (\mathbf{E}[X])^2$. Its square root, $\sqrt{\mathrm{Var}[X]}$, is the *standard deviation* of X, and is often denoted as σ.

Chebyshev Inequality

Definition

The *variance* of a random variable X is
$\mathrm{Var}[X] := \mathbf{E}[(X - \mathbf{E}[X])^2] = \mathbf{E}[X^2] - (\mathbf{E}[X])^2$. Its square root, $\sqrt{\mathrm{Var}[X]}$, is the *standard deviation* of X, and is often denoted as σ.

Theorem (Chebyshev Inequality)

For any $\alpha > 0$, $\mathbf{Pr}[|X - \mathbf{E}[X]| > \alpha\sigma] \le \frac{1}{\alpha^2}$.

Chebyshev Inequality

Definition

The *variance* of a random variable X is $\mathrm{Var}[X] := \mathbf{E}[(X - \mathbf{E}[X])^2] = \mathbf{E}[X^2] - (\mathbf{E}[X])^2$. Its square root, $\sqrt{\mathrm{Var}[X]}$, is the *standard deviation* of X, and is often denoted as σ.

Theorem (Chebyshev Inequality)

For any $\alpha > 0$, $\mathbf{Pr}[|X - \mathbf{E}[X]| > \alpha\sigma] \leq \frac{1}{\alpha^2}$.

Proof.

Apply Markov inequality to the random variable $(X - \mathbf{E}[X])^2$:

$$\mathbf{Pr}\left[|X - \mathbf{E}\left[X\right]| \geq \alpha\sigma\right] = \mathbf{Pr}\left[(X - \mathbf{E}\left[X\right])^2 \geq \alpha^2 \,\mathrm{Var}\left[X\right]\right] \leq \frac{1}{\alpha^2}.$$

□

Chebyshev Inequality

Definition

The *variance* of a random variable X is
$\mathrm{Var}[X] := \mathbf{E}[(X - \mathbf{E}[X])^2] = \mathbf{E}[X^2] - (\mathbf{E}[X])^2$. Its square root, $\sqrt{\mathrm{Var}[X]}$, is the *standard deviation* of X, and is often denoted as σ.

Theorem (Chebyshev Inequality)

For any $\alpha > 0$, $\mathbf{Pr}[|X - \mathbf{E}[X]| > \alpha\sigma] \leq \frac{1}{\alpha^2}$.

Proof.

Apply Markov inequality to the random variable $(X - \mathbf{E}[X])^2$:

$$\mathbf{Pr}\left[|X - \mathbf{E}\left[X\right]| \geq \alpha\sigma\right] = \mathbf{Pr}\left[(X - \mathbf{E}\left[X\right])^2 \geq \alpha^2 \,\mathrm{Var}\left[X\right]\right] \leq \frac{1}{\alpha^2}.$$

□

Question

Distribution where Chebyshev inequality is tight?

Question

Distribution where Chebyshev inequality is tight?

It must be distributions for which the Markov inequality in the proof is tight.

Question

Distribution where Chebyshev inequality is tight?

It must be distributions for which the Markov inequality in the proof is tight.

A distribution where $|X - \mathbf{E}[X]|$ takes two values: 0 and $\alpha\sigma$

Question

Distribution where Chebyshev inequality is tight?

It must be distributions for which the Markov inequality in the proof is tight.

A distribution where $|X - \mathbf{E}[X]|$ takes two values: 0 and $\alpha\sigma$

$\Rightarrow X$ takes three values: $\mathbf{E}[X]$, $\mathbf{E}[X] + \alpha\sigma$ and $\mathbf{E}[X] - \alpha\sigma$.

Useful Facts for Independent Random Variables

Lemma

If X and Y are independent random variables, then $\mathbf{E}[XY] = \mathbf{E}[X] \cdot \mathbf{E}[Y]$, *and* $\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$.

Useful Facts for Independent Random Variables

Lemma

If X and Y are independent random variables, then $\mathbf{E}[XY] = \mathbf{E}[X] \cdot \mathbf{E}[Y]$*, and* $\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$.

Proof.

$$\mathbf{E}\left[XY\right] = \sum_{x,y} (xy)\, \mathbf{Pr}\left[X = x, Y = y\right]$$

Useful Facts for Independent Random Variables

Lemma

If X and Y are independent random variables, then $\mathbf{E}[XY] = \mathbf{E}[X] \cdot \mathbf{E}[Y]$, and $\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$.

Proof.

$$\mathbf{E}[XY] = \sum_{x,y} (xy) \, \mathbf{Pr}[X = x, Y = y]$$

$$= \sum_{x,y} (xy) \, \mathbf{Pr}[X = x] \, \mathbf{Pr}[Y = y]$$

Useful Facts for Independent Random Variables

Lemma

If X and Y are independent random variables, then $\mathbf{E}[XY] = \mathbf{E}[X] \cdot \mathbf{E}[Y]$*, and* $\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$.

Proof.

$$\mathbf{E}[XY] = \sum_{x,y} (xy)\,\mathbf{Pr}[X = x, Y = y]$$

$$= \sum_{x,y} (xy)\,\mathbf{Pr}[X = x]\,\mathbf{Pr}[Y = y]$$

$$= \sum_{x} x\,\mathbf{Pr}[X = x] \sum_{y} y\,\mathbf{Pr}[Y = y] = \mathbf{E}[X] \cdot \mathbf{E}[Y].$$

Useful Facts for Independent Random Variables

Lemma

If X and Y are independent random variables, then $\mathbf{E}[XY] = \mathbf{E}[X] \cdot \mathbf{E}[Y]$, *and* $\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$.

Proof.

$$\begin{aligned}\mathbf{E}\left[XY\right] &= \sum_{x,y}(xy)\,\mathbf{Pr}\left[X=x, Y=y\right]\\ &= \sum_{x,y}(xy)\,\mathbf{Pr}\left[X=x\right]\mathbf{Pr}\left[Y=y\right]\\ &= \sum_{x} x\,\mathbf{Pr}\left[X=x\right]\sum_{y} y\,\mathbf{Pr}\left[Y=y\right] = \mathbf{E}\left[X\right]\cdot\mathbf{E}\left[Y\right].\end{aligned}$$

Useful Facts for Independent Random Variables

Lemma

If X and Y are independent random variables, then $\mathbf{E}[XY] = \mathbf{E}[X] \cdot \mathbf{E}[Y]$*, and* $\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$.

Proof.

$$\begin{aligned}
\mathbf{E}[XY] &= \sum_{x,y} (xy)\,\mathbf{Pr}[X = x, Y = y] \\
&= \sum_{x,y} (xy)\,\mathbf{Pr}[X = x]\,\mathbf{Pr}[Y = y] \\
&= \sum_{x} x\,\mathbf{Pr}[X = x] \sum_{y} y\,\mathbf{Pr}[Y = y] = \mathbf{E}[X] \cdot \mathbf{E}[Y].
\end{aligned}$$

The part for variance is left as an easy exercise. □

Application of Chebyshev Inequality: Weak Law of Large Numbers

Theorem

Let $X_1, X_2, \cdots$ be independently, identically distributed (i.i.d.) random variables, and each has finite variance. For each $n \geq 1$, let $\overline{X}_n$ be $\frac{1}{n}\sum_{i=1}^{n} X_i$. Then for any $\delta > 0$, $\lim_{n\to\infty} \mathbf{Pr}[|\overline{X}_n - \mathbf{E}[\overline{X}_n]| > \delta] = 0$.

Application of Chebyshev Inequality: Weak Law of Large Numbers

Theorem

Let $X_1, X_2, \cdots$ be independently, identically distributed (i.i.d.) random variables, and each has finite variance. For each $n \geq 1$, let $\overline{X}_n$ be $\frac{1}{n}\sum_{i=1}^{n} X_i$. Then for any $\delta > 0$, $\lim_{n\to\infty} \mathbf{Pr}[|\overline{X}_n - \mathbf{E}[\overline{X}_n]| > \delta] = 0$.

Proof.

By independence, $\mathrm{Var}[\overline{X}_n] = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}[X_i] = \frac{1}{n}\mathrm{Var}[X_1]$.

Application of Chebyshev Inequality: Weak Law of Large Numbers

Theorem

Let $X_1, X_2, \cdots$ be independently, identically distributed (i.i.d.) random variables, and each has finite variance. For each $n \geq 1$, let $\overline{X}_n$ be $\frac{1}{n}\sum_{i=1}^{n} X_i$. Then for any $\delta > 0$, $\lim_{n\to\infty} \mathbf{Pr}[|\overline{X}_n - \mathbf{E}[\overline{X}_n]| > \delta] = 0$.

Proof.

By independence, $\mathrm{Var}[\overline{X}_n] = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}[X_i] = \frac{1}{n}\mathrm{Var}[X_1]$.

By Chebyshev inequality, $\mathbf{Pr}[|\overline{X}_n - \mathbf{E}[\overline{X}_n]| > \delta] \leq \frac{\mathrm{Var}[X_1]}{n\delta^2}$.

Application of Chebyshev Inequality: Weak Law of Large Numbers

Theorem

Let $X_1, X_2, \cdots$ be independently, identically distributed (i.i.d.) random variables, and each has finite variance. For each $n \geq 1$, let $\overline{X}_n$ be $\frac{1}{n}\sum_{i=1}^{n} X_i$. Then for any $\delta > 0$, $\lim_{n\to\infty} \mathbf{Pr}[|\overline{X}_n - \mathbf{E}[\overline{X}_n]| > \delta] = 0$.

Proof.

By independence, $\mathrm{Var}[\overline{X}_n] = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}[X_i] = \frac{1}{n}\mathrm{Var}[X_1]$.
By Chebyshev inequality, $\mathbf{Pr}[|\overline{X}_n - \mathbf{E}[\overline{X}_n]| > \delta] \leq \frac{\mathrm{Var}[X_1]}{n\delta^2}$.
The right hand side goes to 0 as n goes to infinity. □

Overview of the Landscape

- Markov inequality gives a constant bound for deviating from the expectation by a constant factor.
  - Non-negativity is required.

Overview of the Landscape

- Markov inequality gives a constant bound for deviating from the expectation by a constant factor.
  - Non-negativity is required.
- If the variacne is small, Chebyshev can be better. It measures the deviation as a multiple of the standard deviation.
  - Effective when the standard deviation is small.

Overview of the Landscape

- Markov inequality gives a constant bound for deviating from the expectation by a constant factor.
  - Non-negativity is required.
- If the variacne is small, Chebyshev can be better. It measures the deviation as a multiple of the standard deviation.
  - Effective when the standard deviation is small.
  - In the weak law of large numbers, the dependence of the bound on both n and $\frac{1}{\delta}$ is polynomial.

Overview of the Landscape

- Markov inequality gives a constant bound for deviating from the expectation by a constant factor.
  - Non-negativity is required.
- If the variacne is small, Chebyshev can be better. It measures the deviation as a multiple of the standard deviation.
  - Effective when the standard deviation is small.
  - In the weak law of large numbers, the dependence of the bound on both n and $\frac{1}{\delta}$ is polynomial.
- Sometimes we need much tighter bounds.

Overview of the Landscape

- Markov inequality gives a constant bound for deviating from the expectation by a constant factor.
  - Non-negativity is required.
- If the variacne is small, Chebyshev can be better. It measures the deviation as a multiple of the standard deviation.
  - Effective when the standard deviation is small.
  - In the weak law of large numbers, the dependence of the bound on both n and $\frac{1}{\delta}$ is polynomial.
- Sometimes we need much tighter bounds.
- Methodologically, the proof of Chebyshev inequality amplifies the deviation $|X - \mathbf{E}[X]|$ by taking its square.

Overview of the Landscape

- Markov inequality gives a constant bound for deviating from the expectation by a constant factor.
  - Non-negativity is required.
- If the variacne is small, Chebyshev can be better. It measures the deviation as a multiple of the standard deviation.
  - Effective when the standard deviation is small.
  - In the weak law of large numbers, the dependence of the bound on both n and $\frac{1}{\delta}$ is polynomial.
- Sometimes we need much tighter bounds.
- Methodologically, the proof of Chebyshev inequality amplifies the deviation $|X - \mathbf{E}[X]|$ by taking its square.
  - Take a nonnegative, fast growing function $f(\cdot)$ and apply Markov inequality to $f(X)$.

Overview of the Landscape

- Markov inequality gives a constant bound for deviating from the expectation by a constant factor.
  - Non-negativity is required.
- If the variacne is small, Chebyshev can be better. It measures the deviation as a multiple of the standard deviation.
  - Effective when the standard deviation is small.
  - In the weak law of large numbers, the dependence of the bound on both n and $\frac{1}{\delta}$ is polynomial.
- Sometimes we need much tighter bounds.
- Methodologically, the proof of Chebyshev inequality amplifies the deviation $|X - \mathbf{E}[X]|$ by taking its square.
  - Take a nonnegative, fast growing function $f(\cdot)$ and apply Markov inequality to $f(X)$.
  - Chebyshev $\to$ Markov $\to$ Kolmogorov

Overview of the Landscape

- Markov inequality gives a constant bound for deviating from the expectation by a constant factor.
  - Non-negativity is required.
- If the variacne is small, Chebyshev can be better. It measures the deviation as a multiple of the standard deviation.
  - Effective when the standard deviation is small.
  - In the weak law of large numbers, the dependence of the bound on both n and $\frac{1}{\delta}$ is polynomial.
- Sometimes we need much tighter bounds.
- Methodologically, the proof of Chebyshev inequality amplifies the deviation $|X - \mathbf{E}[X]|$ by taking its square.
  - Take a nonnegative, fast growing function $f(\cdot)$ and apply Markov inequality to $f(X)$.
  - Chebyshev $\rightarrow$ Markov $\rightarrow$ Kolmogorov
  - Bernstein and Chernoff exploited the idea by looking at $f(x) = e^{\lambda x}$.

Chernoff Bound: I.I.D. Case

Let $X_1, \cdots, X_n$ be i.i.d. Bernoulli variables, such that $\mathbf{Pr}[X_i = 1] = p$ and $\mathbf{Pr}[X_i = 0] = q := 1 - p$ for each i. Define $X = \sum_{i=1}^{n} X_i$.

Theorem (Chernoff Bound)

For any $t > 0$,

$$\mathbf{Pr}\left[X > (p+t)n\right] \leq \exp\left\{\left(-(p+t)\ln\frac{p+t}{p} - (q-t)\ln\frac{q-t}{q}\right)n\right\}.$$

Proof.

For any $\lambda > 0$, we have

$$\mathbf{Pr}\left[X > (p+t)n\right] = \mathbf{Pr}\left[e^{\lambda X} \geq e^{\lambda(p+t)n}\right] \leq \frac{\mathbf{E}[e^{\lambda X}]}{e^{\lambda(p+t)n}}.$$

Chernoff Bound: I.I.D. Case

Let $X_1, \cdots, X_n$ be i.i.d. Bernoulli variables, such that $\mathbf{Pr}[X_i = 1] = p$ and $\mathbf{Pr}[X_i = 0] = q := 1 - p$ for each i. Define $X = \sum_{i=1}^n X_i$.

Theorem (Chernoff Bound)

For any $t > 0$,

$$\mathbf{Pr}\left[X > (p+t)n\right] \leq \exp\left\{\left(-(p+t)\ln\frac{p+t}{p} - (q-t)\ln\frac{q-t}{q}\right)n\right\}.$$

Proof.

For any $\lambda > 0$, we have

$$\mathbf{Pr}\left[X > (p+t)n\right] = \mathbf{Pr}\left[e^{\lambda X} \geq e^{\lambda(p+t)n}\right] \leq \frac{\mathbf{E}[e^{\lambda X}]}{e^{\lambda(p+t)n}}.$$

By independence, we have $\mathbf{E}[e^{\lambda X}] = \mathbf{E}[e^{\sum_i \lambda X_i}] = \prod_i \mathbf{E}[e^{\lambda X_i}] = (pe^\lambda + q)^n$.

Proof of Chernoff Bound (Cont.)

Proof of Chernoff Bound (Cont.)

So far we have $\mathbf{Pr}[X \geq (p+t)n] \leq \left(\frac{pe^{\lambda}+q}{e^{\lambda(p+t)}}\right)^n$.

Proof of Chernoff Bound (Cont.)

Proof of Chernoff Bound (Cont.)

So far we have $\mathbf{Pr}[X \geq (p+t)n] \leq \left(\frac{pe^{\lambda}+q}{e^{\lambda(p+t)}}\right)^n$.

Take a λ that minimizes the RHS: let e^{λ} be $\frac{q}{p} \cdot \frac{p+t}{q-t}$, we get the bound in the theorem statement. $\square$

Proof of Chernoff Bound (Cont.)

Proof of Chernoff Bound (Cont.)

So far we have $\mathbf{Pr}[X \geq (p+t)n] \leq \left(\frac{pe^{\lambda}+q}{e^{\lambda(p+t)}}\right)^n$.

Take a λ that minimizes the RHS: let e^{λ} be $\frac{q}{p} \cdot \frac{p+t}{q-t}$, we get the bound in the theorem statement. □

Interpretation of the bound:

- Note the exponential decay with n;

Proof of Chernoff Bound (Cont.)

Proof of Chernoff Bound (Cont.)

So far we have $\mathbf{Pr}[X \geq (p+t)n] \leq \left(\frac{pe^{\lambda}+q}{e^{\lambda(p+t)}}\right)^n$.

Take a λ that minimizes the RHS: let e^{λ} be $\frac{q}{p} \cdot \frac{p+t}{q-t}$, we get the bound in the theorem statement. □

Interpretation of the bound:

- Note the exponential decay with n;
- The coefficient $(p+t)\log\frac{p+t}{p} + (q-t)\log\frac{q-t}{q}$ is the *relative entropy distance* of the distribution $p+t, q-t$ from the distribution p, q on the two-point space $\{1, 0\}$.

Proof of Chernoff Bound (Cont.)

Proof of Chernoff Bound (Cont.)

So far we have $\mathbf{Pr}[X \geq (p+t)n] \leq \left(\frac{pe^{\lambda}+q}{e^{\lambda(p+t)}}\right)^n$.

Take a λ that minimizes the RHS: let e^{λ} be $\frac{q}{p} \cdot \frac{p+t}{q-t}$, we get the bound in the theorem statement. □

Interpretation of the bound:

- Note the exponential decay with n;
- The coefficient $(p+t)\log\frac{p+t}{p} + (q-t)\log\frac{q-t}{q}$ is the *relative entropy distance* of the distribution $p+t, q-t$ from the distribution p, q on the two-point space $\{1, 0\}$.

The same proof yields the same bound for $\mathbf{Pr}[X \leq (p-t)n]$.

Chernoff Bound: Heterogeneous Bernoulli Variables

What if $X_1, \cdots, X_n$ are independent Bernoulli variables but not identically distributed?

Chernoff Bound: Heterogeneous Bernoulli Variables

What if $X_1, \cdots, X_n$ are independent Bernoulli variables but not identically distributed?
Suppose $\mathbf{Pr}[X_i = 1] = p_i$, and $\mathbf{Pr}[X_i = 0] = q_i$ for each i, and let $p = \frac{1}{n}\sum_i p_i$ and $q = 1 - p$.

Chernoff Bound: Heterogeneous Bernoulli Variables

What if $X_1, \cdots, X_n$ are independent Bernoulli variables but not identically distributed?
Suppose $\mathbf{Pr}[X_i = 1] = p_i$, and $\mathbf{Pr}[X_i = 0] = q_i$ for each i, and let $p = \frac{1}{n} \sum_i p_i$ and $q = 1 - p$.
In the proof where we calculated $\mathbf{E}[e^{\lambda X}]$, we have that it is equal to $\prod_i (p_i e^\lambda + q_i) \leq \left(pe^\lambda + q\right)^n$ by the AM-GM inequality.

Chernoff Bound: Heterogeneous Bernoulli Variables

What if $X_1, \cdots, X_n$ are independent Bernoulli variables but not identically distributed?
Suppose $\mathbf{Pr}[X_i = 1] = p_i$, and $\mathbf{Pr}[X_i = 0] = q_i$ for each i, and let $p = \frac{1}{n}\sum_i p_i$ and $q = 1 - p$.
In the proof where we calculated $\mathbf{E}[e^{\lambda X}]$, we have that it is equal to $\prod_i (p_i e^{\lambda} + q_i) \leq \left(pe^{\lambda} + q\right)^n$ by the AM-GM inequality.
The rest of the proof follows exactly as before.

Chernoff Bound: Heterogeneous Bernoulli Variables

What if $X_1, \cdots, X_n$ are independent Bernoulli variables but not identically distributed?
Suppose $\mathbf{Pr}[X_i = 1] = p_i$, and $\mathbf{Pr}[X_i = 0] = q_i$ for each i, and let $p = \frac{1}{n}\sum_i p_i$ and $q = 1 - p$.
In the proof where we calculated $\mathbf{E}[e^{\lambda X}]$, we have that it is equal to $\prod_i (p_i e^{\lambda} + q_i) \leq \left(pe^{\lambda} + q\right)^n$ by the AM-GM inequality.
The rest of the proof follows exactly as before.

Theorem

For any $t > 0$,

$$\mathbf{Pr}\left[X > (p+t)n\right] \leq \exp\left\{\left(-(p+t)\ln\frac{p+t}{p} - (q-t)\ln\frac{q-t}{q}\right)n\right\}.$$

Chernoff Bound: The Hoeffding Extension

What if $X_1, \cdots, X_n$ always take values from $[0, 1]$ but not necessarily $\{0, 1\}$?
Suppose $\mathbf{E}[X_i] = p_i$, $q_i = 1 - p_i$ for each i, and let $p = \frac{1}{n}\sum_i p_i$ and $q = 1 - p$.

Chernoff Bound: The Hoeffding Extension

What if $X_1, \cdots, X_n$ always take values from $[0, 1]$ but not necessarily $\{0, 1\}$?
Suppose $\mathbf{E}[X_i] = p_i$, $q_i = 1 - p_i$ for each i, and let $p = \frac{1}{n}\sum_i p_i$ and $q = 1 - p$.
Observation: on $[0, 1]$, $e^{\lambda x} \le \alpha x + \beta$ for $\alpha = e^{\lambda} - 1$ and $\beta = 1$.

Chernoff Bound: The Hoeffding Extension

What if $X_1, \cdots, X_n$ always take values from $[0, 1]$ but not necessarily $\{0, 1\}$?
Suppose $\mathbf{E}[X_i] = p_i$, $q_i = 1 - p_i$ for each i, and let $p = \frac{1}{n}\sum_i p_i$ and $q = 1 - p$.
Observation: on $[0, 1]$, $e^{\lambda x} \le \alpha x + \beta$ for $\alpha = e^{\lambda} - 1$ and $\beta = 1$.
So $\mathbf{E}[e^{\lambda X_i}] \le \mathbf{E}[(e^{\lambda} - 1)X_i + 1] = p_i e^{\lambda} + q_i$.

Chernoff Bound: The Hoeffding Extension

What if $X_1, \cdots, X_n$ always take values from $[0, 1]$ but not necessarily $\{0, 1\}$?
Suppose $\mathbf{E}[X_i] = p_i$, $q_i = 1 - p_i$ for each i, and let $p = \frac{1}{n}\sum_i p_i$ and $q = 1 - p$.
Observation: on $[0, 1]$, $e^{\lambda x} \leq \alpha x + \beta$ for $\alpha = e^{\lambda} - 1$ and $\beta = 1$.
So $\mathbf{E}[e^{\lambda X_i}] \leq \mathbf{E}[(e^{\lambda} - 1)X_i + 1] = p_i e^{\lambda} + q_i$.
Therefore $\mathbf{E}[e^{\lambda X}] = \prod_i \mathbf{E}[e^{\lambda X_i}] \leq \prod_i (p_i e^{\lambda} + q_i) \leq \left(pe^{\lambda} + q\right)^n$.

Chernoff Bound: The Hoeffding Extension

What if $X_1, \cdots, X_n$ always take values from $[0, 1]$ but not necessarily $\{0, 1\}$?
Suppose $\mathbf{E}[X_i] = p_i$, $q_i = 1 - p_i$ for each i, and let $p = \frac{1}{n}\sum_i p_i$ and $q = 1 - p$.
Observation: on $[0, 1]$, $e^{\lambda x} \leq \alpha x + \beta$ for $\alpha = e^{\lambda} - 1$ and $\beta = 1$.
So $\mathbf{E}[e^{\lambda X_i}] \leq \mathbf{E}[(e^{\lambda} - 1)X_i + 1] = p_i e^{\lambda} + q_i$.
Therefore $\mathbf{E}[e^{\lambda X}] = \prod_i \mathbf{E}[e^{\lambda X_i}] \leq \prod_i (p_i e^{\lambda} + q_i) \leq \left(pe^{\lambda} + q\right)^n$.
The rest of the proof remains the same.

Chernoff Bound: The Hoeffding Extension

What if $X_1, \cdots, X_n$ always take values from $[0, 1]$ but not necessarily $\{0, 1\}$?
Suppose $\mathbf{E}[X_i] = p_i$, $q_i = 1 - p_i$ for each i, and let $p = \frac{1}{n}\sum_i p_i$ and $q = 1 - p$.
Observation: on $[0, 1]$, $e^{\lambda x} \leq \alpha x + \beta$ for $\alpha = e^{\lambda} - 1$ and $\beta = 1$.
So $\mathbf{E}[e^{\lambda X_i}] \leq \mathbf{E}[(e^{\lambda} - 1)X_i + 1] = p_i e^{\lambda} + q_i$.
Therefore $\mathbf{E}[e^{\lambda X}] = \prod_i \mathbf{E}[e^{\lambda X_i}] \leq \prod_i (p_i e^{\lambda} + q_i) \leq \left(pe^{\lambda} + q\right)^n$.
The rest of the proof remains the same.

Theorem

For any $t > 0$,

$$\mathbf{Pr}\left[X > (p+t)n\right] \leq \exp\left\{\left(-(p+t)\ln\frac{p+t}{p} - (q-t)\ln\frac{q-t}{q}\right)n\right\}.$$

Useful Forms of Chernoff Bound

Corollary

Let $X_1, \cdots, X_n$ be independently distributed on $[0,1]$ and $X = \sum_i X_i$.

- *For all $t > 0$,*

$$\mathbf{Pr}\left[X > \mathbf{E}\left[X\right] + t\right], \mathbf{Pr}\left[X < \mathbf{E}\left[X\right] - t\right] \le e^{-2t^2/n};$$

- *For any $\epsilon < 1$,*

$$\mathbf{Pr}\left[X > (1+\epsilon)\,\mathbf{E}\left[X\right]\right] \le \left(\frac{e^{\epsilon}}{(1+\epsilon)^{1+\epsilon}}\right)^{\mathbf{E}[X]} \le \exp\left(-\frac{\epsilon^2}{3}\mathbf{E}\left[X\right]\right);$$

$$\mathbf{Pr}\left[X < (1-\epsilon)\,\mathbf{E}\left[X\right]\right] \le \left(\frac{e^{-\epsilon}}{(1-\epsilon)^{1-\epsilon}}\right)^{\mathbf{E}[X]} \le \exp\left(-\frac{\epsilon^2}{2}\mathbf{E}\left[X\right]\right).$$

Useful Forms of Chernoff Bound (Cont.)

Corollary ((Cont.))

- *For any* $\epsilon > 1$,

$$\mathbf{Pr}\left[X > (1+\epsilon)\,\mathbf{E}\left[X\right]\right] \leq \left(\frac{e^{\epsilon}}{(1+\epsilon)^{1+\epsilon}}\right)^{\mathbf{E}[X]} \leq \exp\left(-\frac{\epsilon}{3}\,\mathbf{E}\left[X\right]\right);$$

- *If* $t > 2e\,\mathbf{E}[X]$, *then*

$$\mathbf{Pr}\left[X > t\right] \leq 2^{-t}.$$

Proof Sketch

Proof Sketch.

- Let $f(t)$ be $(p+t)\ln\frac{p+t}{p} + (q-t)\ln\frac{q-t}{q}$. Show $f(t) \geq 2t^2$ by showing $f(0) = f'(0) = 0$ and $f''(t) \geq 4$ for all $0 \leq t \leq q$ followed by Taylor's theorem with remainder.

Proof Sketch

Proof Sketch.

- Let $f(t)$ be $(p+t)\ln\frac{p+t}{p}+(q-t)\ln\frac{q-t}{q}$. Show $f(t)\geq 2t^2$ by showing $f(0)=f'(0)=0$ and $f''(t)\geq 4$ for all $0\leq t\leq q$ followed by Taylor's theorem with remainder.
- Let $g(x)$ be $f(px)$, then $g'(0)=pf'(px)$, and so $g(0)=g'(0)=0$. Show $g'(1)>p\ln 2>\frac{2}{3}p$. Deduce that for $x\in(0,1)$, $g(x)\geq px^2/3$.

Proof Sketch

Proof Sketch.

- Let $f(t)$ be $(p+t)\ln\frac{p+t}{p}+(q-t)\ln\frac{q-t}{q}$. Show $f(t)\geq 2t^2$ by showing $f(0)=f'(0)=0$ and $f''(t)\geq 4$ for all $0\leq t\leq q$ followed by Taylor's theorem with remainder.
- Let $g(x)$ be $f(px)$, then $g'(0)=pf'(px)$, and so $g(0)=g'(0)=0$. Show $g'(1)>p\ln 2>\frac{2}{3}p$. Deduce that for $x\in(0,1)$, $g(x)\geq px^2/3$.
- Set $h(x):=g(-x)$. Then $h'(x)=-g'(-x)$, and $h(0)=h'(0)=0$. Show then $h''(x)\leq p$ for $x\in(0,1)$. Deduce that $h(x)\geq px^2/2$.

Proof Sketch

Proof Sketch.

- Let $f(t)$ be $(p+t)\ln\frac{p+t}{p} + (q-t)\ln\frac{q-t}{q}$. Show $f(t) \geq 2t^2$ by showing $f(0) = f'(0) = 0$ and $f''(t) \geq 4$ for all $0 \leq t \leq q$ followed by Taylor's theorem with remainder.
- Let $g(x)$ be $f(px)$, then $g'(0) = pf'(px)$, and so $g(0) = g'(0) = 0$. Show $g'(1) > p\ln 2 > \frac{2}{3}p$. Deduce that for $x \in (0,1)$, $g(x) \geq px^2/3$.
- Set $h(x) := g(-x)$. Then $h'(x) = -g'(-x)$, and $h(0) = h'(0) = 0$. Show then $h''(x) \leq p$ for $x \in (0,1)$. Deduce that $h(x) \geq px^2/2$.

See reading for more details. Or take them as exercises.

□