Chernoff Bounds

Theme: try to show that it is unlikely a random variable X is far away from its expectation.

The more you know about X, the better the bound you obtain.

Markov's inequality: use $\mathbf{E}[X]$

Chebyshev's inequality: use $Var[X]$

Chernoff bounds: use *moment generating function*

Chernoff Bounds

Let $X_1, ..., X_n$ be independent 0-1 random variables with

$$Pr(X_i = 1) = p_i \qquad Pr(X_i = 0) = 1 - p_i.$$

Let $X = \sum_{i=1}^{n} X_i$,

$$\mu = \mathbf{E}[X] = \sum_{i=1}^{n} \mathbf{E}[X_i] = \sum_{i=1}^{n} p_i$$

We want a bound on

$$Pr(|X - \mu| > \delta\mu).$$

Assume for all i we have $p_i = p; 1 - p_i = q$.

$$\mu = \mathbf{E}[X] = np$$

$$Var[X] = npq$$

If we use Chebyshev's Inequality we get

$$Pr(|X - \mu| > \delta\mu) \leq \frac{npq}{\delta^2\mu^2} = \frac{npq}{\delta^2 n^2 p^2} = \frac{q}{\delta^2 \mu}$$

Chernoff bound will give

$$Pr(|X - \mu| > \delta\mu) \leq 2e^{-\mu\delta^2/3}.$$

The Basic Idea

Using Markov inequality we have:
For any $t > 0$,

$$Pr(X \geq a) = Pr(e^{tX} \geq e^{ta}) \leq \frac{\mathbf{E}[e^{tX}]}{e^{ta}}.$$

Similarly, for any $t < 0$

$$Pr(X \leq a) = Pr(e^{tX} \geq e^{ta}) \leq \frac{\mathbf{E}[e^{tX}]}{e^{ta}}.$$

$$Pr(X \geq a) \leq \min_{t>0} \frac{\mathbf{E}[e^{tX}]}{e^{ta}}.$$

$$Pr(X \leq a) \leq \min_{t<0} \frac{\mathbf{E}[e^{tX}]}{e^{ta}}.$$

Moment Generating Function

Definition

The moment generating function of a random variable X is defined for any real value t as

$$M_X(t) = \mathbf{E}[e^{tX}].$$

Theorem

Let X be a random variable with moment generating function $M_X(t)$. Assuming that exchanging the expectation and differentiation operands is legitimate, then for all $n \geq 1$

$$\mathbf{E}[X^n] = M_X^{(n)}(0),$$

where $M_X^{(n)}(0)$ is the n-th derivative of $M_X(t)$ evaluated at $t = 0$.

Proof.

$$M_X^{(n)}(t) = \mathbf{E}[X^n e^{tX}].$$

Computed at $t = 0$ we get

$$M_X^{(n)}(0) = \mathbf{E}[X^n].$$

□

Theorem

Let X and Y be two random variables. If

$$M_X(t) = M_Y(t)$$

for all $t \in (-\delta, \delta)$ for some $\delta > 0$, then X and Y have the same distribution.

Theorem

If X and Y are independent random variables then

$$M_{X+Y}(t) = M_X(t) M_Y(t).$$

Proof.

$$M_{X+Y}(t) = \mathbf{E}[e^{t(X+Y)}] = \mathbf{E}[e^{tX}]\mathbf{E}[e^{tY}] = M_X(t)M_Y(t).$$

□

Chernoff Bound for Sum of Bernoulli Trials

Let $X_1, \ldots, X_n$ be a sequence of independent Bernoulli trials with $Pr(X_i = 1) = p_i$. Let $X = \sum_{i=1}^n X_i$, and let

$$\mu = \mathbf{E}[X] = \mathbf{E}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathbf{E}[X_i] = \sum_{i=1}^n p_i.$$

$$\begin{aligned} M_{X_i}(t) &= \mathbf{E}[e^{tX_i}] \\ &= p_i e^t + (1 - p_i) \\ &= 1 + p_i(e^t - 1) \\ &\leq e^{p_i(e^t-1)}. \end{aligned}$$

Taking the product of the n generating functions we get

$$\begin{aligned} M_X(t) &= \prod_{i=1}^{n} M_{X_i}(t) \\ &\leq \prod_{i=1}^{n} e^{p_i(e^t-1)} \\ &= e^{\sum_{i=1}^{n} p_i(e^t-1)} \\ &= e^{(e^t-1)\mu} \end{aligned}$$

Theorem

Let $X_1, \ldots, X_n$ *be independent Bernoulli random variables such that* $Pr(X_i = 1) = p_i$.

- *For any* $\delta > 0$,

$$Pr(X \geq (1+\delta)\mu) \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu}. \tag{1}$$

- *For* $0 < \delta \leq 1$,

$$Pr(X \geq (1+\delta)\mu) \leq e^{-\mu\delta^2/3}. \tag{2}$$

- *For* $R \geq 6\mu$,

$$Pr(X \geq R) \leq 2^{-R}. \tag{3}$$

$$M_X(t) = \mathbf{E}[e^{tX}] \leq e^{(e^t - 1)\mu}$$

Applying Markov's inequality we have for any $t > 0$

$$\begin{aligned} Pr(X \geq (1+\delta)\mu) &= Pr(e^{tX} \geq e^{t(1+\delta)\mu}) \\ &\leq \frac{\mathbf{E}[e^{tX}]}{e^{t(1+\delta)\mu}} \\ &\leq \frac{e^{(e^t-1)\mu}}{e^{t(1+\delta)\mu}} \end{aligned}$$

For any $\delta > 0$, we can set $t = \ln(1+\delta) > 0$ to get:

$$Pr(X \geq (1+\delta)\mu) \leq \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}} \right)^{\mu}.$$

This proves (1).

We show that for $0 < \delta < 1$,

$$\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}} \leq e^{-\delta^2/3}$$

or that $f(\delta) = \delta - (1+\delta)\ln(1+\delta) + \delta^2/3 \leq 0$
in that interval. Computing the derivatives of $f(\delta)$ we get

$$\begin{aligned} f'(\delta) &= 1 - \frac{1+\delta}{1+\delta} - \ln(1+\delta) + \frac{2}{3}\delta = -\ln(1+\delta) + \frac{2}{3}\delta, \\ f''(\delta) &= -\frac{1}{1+\delta} + \frac{2}{3}. \end{aligned}$$

$f''(\delta) < 0$ for $0 \leq \delta < 1/2$, and $f''(\delta) > 0$ for $\delta > 1/2$.
$f'(\delta)$ first decreases and then increases over the interval $[0, 1]$.
Since $f'(0) = 0$ and $f'(1) < 0$, $f'(\delta) \leq 0$ in the interval $[0, 1]$.
Since $f(0) = 0$, we have that $f(\delta) \leq 0$ in that interval.
This proves (2).

Write R as $R = (1+\delta)\mu$. Then for $R \geq 6\mu$, $\delta \geq 5$, so we have.

$$\begin{aligned} Pr(X \geq (1+\delta)\mu) &\leq \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu} \\ &\leq \left(\frac{e}{1+\delta}\right)^{(1+\delta)\mu} \\ &\leq \left(\frac{e}{6}\right)^{R} \\ &\leq 2^{-R}, \end{aligned}$$

that proves (3).

Theorem

Let $X_1, \ldots, X_n$ *be independent Bernoulli random variables such that* $Pr(X_i = 1) = p_i$. *Let* $X = \sum_{i=1}^{n} X_i$ *and* $\mu = \mathbf{E}[X]$.
For $0 < \delta < 1$:

- $$Pr(X \le (1-\delta)\mu) \le \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}} \right)^{\mu}. \tag{4}$$

- $$Pr(X \le (1-\delta)\mu) \le e^{-\mu\delta^2/2}. \tag{5}$$

Using Markov's inequality, for any $t < 0$,

$$\begin{aligned} Pr(X \leq (1-\delta)\mu) &= Pr(e^{tX} \geq e^{(1-\delta)t\mu}) \\ &\leq \frac{\mathbf{E}[e^{tX}]}{e^{t(1-\delta)\mu}} \\ &\leq \frac{e^{(e^t-1)\mu}}{e^{t(1-\delta)\mu}} \end{aligned}$$

For $0 < \delta < 1$, we set $t = \ln(1-\delta) < 0$ to get:

$$Pr(X \leq (1-\delta)\mu) \leq \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{\mu}$$

This proves (4).
We need to show:

$$f(\delta) = -\delta - (1-\delta)\ln(1-\delta) + \frac{1}{2}\delta^2 \leq 0.$$

We need to show:

$$f(\delta) = -\delta - (1-\delta)\ln(1-\delta) + \frac{1}{2}\delta^2 \leq 0.$$

Differentiating $f(\delta)$ we get

$$\begin{aligned} f'(\delta) &= \ln(1-\delta) + \delta, \\ f''(\delta) &= -\frac{1}{1-\delta} + 1. \end{aligned}$$

Since $f''(\delta) < 0$ for $\delta \in (0,1)$, $f'(\delta)$ decreasing in that interval. Since $f'(0) = 0$, $f'(\delta) \leq 0$ for $\delta \in (0,1)$. Therefore $f(\delta)$ is non increasing in that interval.

$f(0) = 0$. Since $f(\delta)$ is non increasing for $\delta \in [0,1)$, $f(\delta) \leq 0$ in that interval, and (5) follows.

Example: Coin flips

Let X be the number of heads in a sequence of n independent fair coin flips.

$$Pr\left(\left|X - \frac{n}{2}\right| \geq \frac{1}{2}\sqrt{6n\ln n}\right)$$

$$= Pr\left(X \geq \frac{n}{2}\left(1 + \sqrt{\frac{6\ln n}{n}}\right)\right)$$

$$+ Pr\left(X \leq \frac{n}{2}\left(1 - \sqrt{\frac{6\ln n}{n}}\right)\right)$$

$$\leq e^{-\frac{1}{3}\frac{n}{2}\frac{6\ln n}{n}} + e^{-\frac{1}{2}\frac{n}{2}\frac{6\ln n}{n}} \leq \frac{2}{n}.$$

Using the Chebyshev's bound we had:

$$Pr\left(\left|X-\frac{n}{2}\right|\geq\frac{n}{4}\right)\leq\frac{4}{n}.$$

Using the Chernoff bound in this case, we obtain

$$\begin{aligned}Pr\left(\left|X-\frac{n}{2}\right|\geq\frac{n}{4}\right) &= Pr\left(X\geq\frac{n}{2}\left(1+\frac{1}{2}\right)\right)\\ &+ Pr\left(X\leq\frac{n}{2}\left(1-\frac{1}{2}\right)\right)\\ &\leq e^{-\frac{1}{3}\frac{n}{2}\frac{1}{4}}+e^{-\frac{1}{2}\frac{n}{2}\frac{1}{4}}\\ &\leq 2e^{-\frac{n}{24}}.\end{aligned}$$

Example: Estimating a Parameter

- Evaluating the probability that a particular DNA mutation occurs in the population.
- Given a DNA sample, a lab test can determine if it carries the mutation.
- The test is expensive and we would like to obtain a relatively reliable estimate from a minimum number of samples.
- $p =$ the unknown value;
- $n =$ number of samples, $\tilde{p}n$ had the mutation.
- Given sufficient number of samples we expect the value p to be in the neighborhood of sampled value $\tilde{p}$, but we cannot predict any single value with high confidence.

Confidence Interval

Instead of predicting a single value for the parameter we give an *interval* that is *likely* to contain the parameter.

Definition

A $1-q$ **confidence interval** *for a parameter* T *is an interval* $[\tilde{p}-\delta, \tilde{p}+\delta]$ *such that*

$$Pr(T \in [\tilde{p}-\delta, \tilde{p}+\delta]) \geq 1-q.$$

We want to minimize 2δ and q, with minimum n.
Using $\tilde{p}n$ as our estimate for pn, we need to compute δ and q such that

$$Pr(p \in [\tilde{p}-\delta, \tilde{p}+\delta]) = Pr(np \in [n(\tilde{p}-\delta), n(\tilde{p}+\delta)]) \geq 1-q.$$

- The random variable here is the interval $[\tilde{p} - \delta, \tilde{p} + \delta]$ (or the value $\tilde{p}$), while p is a fixed (unknown) value.
- $n\tilde{p}$ has a binomial distribution with parameters n and p, and $\mathbf{E}[\tilde{p}] = p$. If $p \notin [\tilde{p} - \delta, \tilde{p} + \delta]$ then we have one of the following two events:
  1. If $p < \tilde{p} - \delta$, then $n\tilde{p} \geq n(p + \delta) = np\left(1 + \frac{\delta}{p}\right)$, or $n\tilde{p}$ is larger than its expectation by a $\frac{\delta}{p}$ factor.
  2. If $p > \tilde{p} + \delta$, then $n\tilde{p} \leq n(p - \delta) = np\left(1 - \frac{\delta}{p}\right)$, and $n\tilde{p}$ is smaller than its expectation by a $\frac{\delta}{p}$ factor.

$$
\begin{aligned}
& \Pr(p \notin [\tilde{p} - \delta, \tilde{p} + \delta]) \\
= \; & \Pr\left(n\tilde{p} \leq np\left(1 - \frac{\delta}{p}\right)\right) + \Pr\left(n\tilde{p} \geq np\left(1 + \frac{\delta}{p}\right)\right) \\
\leq \; & e^{-\frac{1}{2}np\left(\frac{\delta}{p}\right)^2} + e^{-\frac{1}{3}np\left(\frac{\delta}{p}\right)^2} \\
= \; & e^{-\frac{n\delta^2}{2p}} + e^{-\frac{n\delta^2}{3p}}.
\end{aligned}
$$

But the value of p is unknown, A simple solution is to use the fact that $p \leq 1$ to prove

$$\Pr(p \notin [\tilde{p} - \delta, \tilde{p} + \delta]) \leq e^{-\frac{n\delta^2}{2}} + e^{-\frac{n\delta^2}{3}}.$$

Setting $q = e^{-\frac{n\delta^2}{2}} + e^{-\frac{n\delta^2}{3}}$, we obtain a tradeoff between δ, n, and the error probability q.

$$q = e^{-\frac{n\delta^2}{2}} + e^{-\frac{n\delta^2}{3}}$$

If we want to obtain a $1 - q$ confidence interval $[\tilde{p} - \delta, \tilde{p} + \delta]$,

$$n \geq \frac{3}{\delta^2} \ln \frac{2}{q}$$

samples are enough.