

Data Science Theory and Practices

Chernoff and Hoeffding Bounds

Joseph Chuang-Chieh Lin
Dept. CSIE, Tamkang University

24th May 2021

Moment Generating Functions

- Definition. The moment generating function of a random variable X is

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

Moment Generating Functions

- Definition. The moment generating function of a random variable X is

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

$$\mathbf{E}[X^n] = M_X^{(n)}(0) \quad \text{The } n\text{th derivative of } M_X(t) \text{ at } t = 0.$$

Example

- Consider a geometric random variable X with parameter p .
- For $t < -\ln(1-p)$,

$$\begin{aligned} M_X(t) &= \mathbf{E}[e^{tX}] \\ &= \sum_{k=1}^{\infty} (1-p)^{k-1} p e^{tk} \\ &= \frac{p}{1-p} \sum_{k=1}^{\infty} (1-p)^k e^{tk} \\ &= \frac{p}{1-p} ((1-(1-p)e^t)^{-1} - 1). \end{aligned}$$

Example

- Consider a geometric random variable X with parameter p .
- For $t < -\ln(1-p)$,

$$\begin{aligned} M_X(t) &= \mathbf{E}[e^{tX}] & \therefore M_X^{(1)}(t) &= p(1 - (1-p)e^t)^{-2}e^t, \\ &= \sum_{k=1}^{\infty} (1-p)^{k-1} p e^{tk} & M_X^{(2)}(t) &= 2p(1-p)(1 - (1-p)e^t)^{-3}e^{2t} \\ &= \frac{p}{1-p} \sum_{k=1}^{\infty} (1-p)^k e^{tk} & &+ p(1 - (1-p)e^t)^{-2}e^t. \\ &= \frac{p}{1-p} ((1 - (1-p)e^t)^{-1} - 1). \end{aligned}$$

Example

- Consider a geometric random variable X with parameter p .
- For $t < -\ln(1-p)$,

$$\begin{aligned} M_X(t) &= \mathbf{E}[e^{tX}] && \therefore M_X^{(1)}(t) = p(1 - (1-p)e^t)^{-2}e^t, \\ &= \sum_{k=1}^{\infty} (1-p)^{k-1} p e^{tk} && M_X^{(2)}(t) = 2p(1-p)(1 - (1-p)e^t)^{-3}e^{2t} \\ & & & + p(1 - (1-p)e^t)^{-2}e^t. \\ &= \frac{p}{1-p} \sum_{k=1}^{\infty} (1-p)^k e^{tk} && \mathbf{E}[X] = M_X^{(1)}(0) = \frac{1}{p}, \quad \mathbf{E}[X^2] = M_X^{(2)}(0) = \frac{(2-p)}{p^2}. \\ &= \frac{p}{1-p} ((1 - (1-p)e^t)^{-1} - 1). \end{aligned}$$

MGF for sum of independent r.v.'s

- Theorem. If X and Y are independent random variables, then

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

MGF for sum of independent r.v.'s

- Theorem. If X and Y are independent random variables, then

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

- *Proof.*

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

MGF for sum of independent r.v.'s

- Theorem. If X and Y are independent random variables, then

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

- *Generalization:*

$$M_{X_1+X_2+\dots+X_k}(t) = M_{X_1}(t) \cdot M_{X_2}(t) \cdots M_{X_k}(t).$$

Chernoff bounds: Applying Markov's inequality to e^{tX}

- From Markov's inequality,

For any $t > 0$,

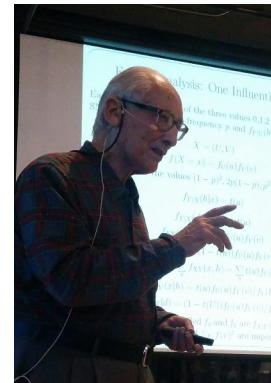
$$\Pr[X \geq a] = \Pr[e^{tX} \geq e^{ta}] \leq \frac{\mathbf{E}[e^{tX}]}{e^{ta}}.$$

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

For any $t < 0$,

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

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



Herman Chernoff

https://en.wikipedia.org/wiki/Herman_Chernoff

Chernoff bounds: Applying Markov's inequality to e^{tX}

- From Markov's inequality,

For any $t > 0$,

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

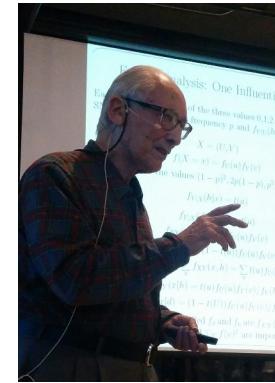
$$\Pr[X \geq a] \leq \min_{t>0} \frac{\mathbf{E}[e^{tX}]}{e^{ta}}.$$

For any $t < 0$,

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

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

- Choose appropriate values for t for specific distributions.



Herman Chernoff

https://en.wikipedia.org/wiki/Herman_Chernoff

Chernoff bounds for sum of Poisson trials

- Poisson trials:
 - ≈ Bernoulli trials
 - while the trials are **not necessarily identical.**
- X_1, \dots, X_n : independent Poisson trials with $\Pr[X_i = 1] = p_i$.
- Let $X = \sum_{i=1}^n X_i$

Chernoff bounds for sum of Poisson trials

- Poisson trials:
 - ≈ Bernoulli trials
 - while the trials are **not necessarily identical.**
- X_1, \dots, X_n : independent Poisson trials with $\Pr[X_i = 1] = p_i$.

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

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

Chernoff bounds for sum of Poisson trials

- For $\delta > 0$, we want to analyze the tail probabilities

$$\Pr[X \geq (1 + \delta)\mu] \text{ and } \Pr[X \leq (1 - \delta)\mu]$$

Chernoff bounds for sum of Poisson trials

- For $\delta > 0$, we want to analyze the tail probabilities

$$\Pr[X \geq (1 + \delta)\mu] \text{ and } \Pr[X \leq (1 - \delta)\mu]$$

$$\begin{aligned} M_{X_i}(t) &= \mathbf{E}[e^{tX_i}] \\ &= p_i e^t + (1 - p_i)e^0 \\ &= 1 + p_i(e^t - 1) \\ &\leq e^{p_i(e^t - 1)}. \end{aligned} \quad \begin{aligned} \therefore M_X(t) &= \prod_{i=1}^n M_{X_i}(t) \\ &\leq \prod_{i=1}^n e^{p_i(e^t - 1)} \\ &= \exp \left\{ \sum_{i=1}^n p_i(e^t - 1) \right\} \\ &= e^{(e^t - 1)\mu}. \end{aligned}$$

Chernoff bounds for sum of Poisson trials

- Theorem. Let X_1, \dots, X_n be independent Poisson trials with $\Pr[X_i = 1] = p_i$.

Let $X = \sum_{i=1}^n X_i$ and $\mu = \mathbf{E}[X]$. Then the following Chernoff bounds hold:

1. For $\delta > 0$,

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

2. For $0 < \delta \leq 1$,

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

3. For $R \geq 6\mu$,

$$\Pr[X \geq R] \leq 2^{-R}.$$

Chernoff bounds for sum of Poisson trials

- **Theorem.** Let X_1, \dots, X_n be independent Poisson trials with $\Pr[X_i = 1] = p_i$.

Let $X = \sum_{i=1}^n X_i$ and $\mu = \mathbf{E}[X]$. Then the following Chernoff bounds hold:

1. For $\delta > 0$,
$$\Pr[X \geq (1 + \delta)\mu] \leq \left(\frac{e^\delta}{(1 + \delta)^{(1+\delta)}} \right)^\mu;$$

2. For $0 < \delta \leq 1$,
$$\Pr[X \geq (1 + \delta)\mu] \leq e^{-\mu\delta^2/3};$$

3. For $R \geq 6\mu$,
$$\Pr[X \geq R] \leq 2^{-R}.$$

Chernoff bounds for sum of Poisson trials

- Applying Markov's inequality for $t > 0$,

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

Chernoff bounds for sum of Poisson trials

- Applying Markov's inequality for $t > 0$,

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

- For any $\delta > 0$, set $t = \ln(1+\delta) > 0$:

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

Chernoff bounds for sum of Poisson trials

- Applying Markov's inequality for $t > 0$,
- For $0 < \delta \leq 1$, $\frac{e^\delta}{(1 + \delta)^{(1+\delta)}} \leq e^{-\delta^2/3}?$

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

- Taking logarithm of both sides:
 $f(\delta) := \delta - (1 + \delta) \ln(1 + \delta) + \frac{\delta^2}{3} \leq 0$

- For any $\delta > 0$, set $t = \ln(1+\delta) > 0$:

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

Chernoff bounds for sum of Poisson trials

- Applying Markov's inequality for $t > 0$,
- For $0 < \delta \leq 1$, $\frac{e^\delta}{(1 + \delta)^{(1+\delta)}} \leq e^{-\delta^2/3}?$

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

- Taking logarithm of both sides:

$$\begin{aligned}f(\delta) &:= \delta - (1 + \delta) \ln(1 + \delta) + \frac{\delta^2}{3} \leq 0 \\ f'(\delta) &= -\ln(1 + \delta) + \frac{2}{3}\delta, \\ f''(\delta) &= -\frac{1}{1 + \delta} + \frac{2}{3}.\end{aligned}$$

- For any $\delta > 0$, set $t = \ln(1+\delta) > 0$:

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

Proof sketch

- Applying Markov's inequality for $t > 0$,
- For $0 < \delta \leq 1$, $\frac{e^\delta}{(1 + \delta)^{(1+\delta)}} \leq e^{-\delta^2/3}?$

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

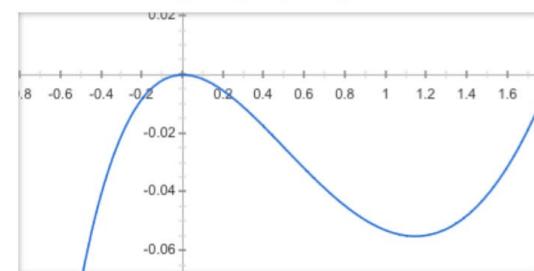
- Taking logarithm of both sides:

$$\begin{aligned}f(\delta) &:= \delta - (1 + \delta) \ln(1 + \delta) + \frac{\delta^2}{3} \leq 0 \\ f'(\delta) &= -\ln(1 + \delta) + \frac{2}{3}\delta, \\ f''(\delta) &= -\frac{1}{1 + \delta} + \frac{2}{3}.\end{aligned}$$

- For any $\delta > 0$, set $t = \ln(1+\delta) > 0$:

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

Graph for $x - (1+x) \ln(1+x) + x^{2/3}$



More info

Chernoff bounds for sum of Poisson trials

- Theorem. Let X_1, \dots, X_n be independent Poisson trials with $\Pr[X_i = 1] = p_i$.

Let $X = \sum_{i=1}^n X_i$ and $\mu = \mathbf{E}[X]$. Then the following Chernoff bounds hold:

For $0 < \delta < 1$,

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

$$\Pr[X \leq (1 - \delta)\mu] \leq e^{-\mu\delta^2/2}.$$

- ✓ Therefore we have:

$$\Pr[|X - \mu| \geq \delta\mu] \leq 2e^{-\mu\delta^2/3}.$$

Example: 75% heads in fair coin flips

- $X_i = \begin{cases} 1 & \text{if the } i\text{th coin flip is head} \\ 0 & \text{otherwise} \end{cases}$

$$\mathbf{E}[X_i] = \Pr[X_i = 1] = \frac{1}{2} \quad \mathbf{E}[X] = \sum_{i=1}^n \mathbf{E}[X_i] = \frac{n}{2}.$$

Example: 75% heads in fair coin flips

- $X_i = \begin{cases} 1 & \text{if the } i\text{th coin flip is head} \\ 0 & \text{otherwise} \end{cases}$

$$\mathbf{E}[X_i] = \Pr[X_i = 1] = \frac{1}{2} \quad \mathbf{E}[X] = \sum_{i=1}^n \mathbf{E}[X_i] = \frac{n}{2}.$$

- Try to use Chernoff bound!

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

Example: 75% heads in fair coin flips

- $X_i = \begin{cases} 1 & \text{if the } i\text{th coin flip is head} \\ 0 & \text{otherwise} \end{cases}$

$$\mathbf{E}[X_i] = \Pr[X_i = 1] = \frac{1}{2} \quad \mathbf{E}[X] = \sum_{i=1}^n \mathbf{E}[X_i] = \frac{n}{2}.$$

- Try to use Chernoff bound!

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

Previous bound using Chebyshev's inequality:

Example: 75% heads in fair coin flips

- $X_i = \begin{cases} 1 & \text{if the } i\text{th coin flip is head} \\ 0 & \text{otherwise} \end{cases}$

$$\mathbf{E}[X_i] = \Pr[X_i = 1] = \frac{1}{2} \quad \mathbf{E}[X] = \sum_{i=1}^n \mathbf{E}[X_i] = \frac{n}{2}.$$

- Actually,

$$\mathbf{E}[X_i^2] = \mathbf{E}[X_i] = \frac{1}{2}.$$

$$\text{Var}[X_i] = \mathbf{E}[X_i^2] - (\mathbf{E}[X_i])^2 = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}.$$

$$\text{Var}[X] = \text{Var} \left[\sum_{i=1}^n X_i \right] = \sum_{i=1}^n \text{Var}[X_i] = \frac{n}{4}.$$

$$\Pr[X \geq 3n/4] \leq \Pr[|X - \mathbf{E}[X]| \geq n/4] \leq \frac{\text{Var}[X]}{(n/4)^2} = \frac{n/4}{(n/4)^2} = \frac{4}{n}.$$

Randomized Algorithms, CSIE, TKU, Taiwan

39

Example: 75% heads in fair coin flips

- $X_i = \begin{cases} 1 & \text{if the } i\text{th coin flip is head} \\ 0 & \text{otherwise} \end{cases}$

$$\mathbf{E}[X_i] = \Pr[X_i = 1] = \frac{1}{2} \quad \mathbf{E}[X] = \sum_{i=1}^n \mathbf{E}[X_i] = \frac{n}{2}.$$

- Try to use Chernoff bound!

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

Previous bound using Chebyshev's inequality:

Example: 75% heads in fair coin flips

- $X_i = \begin{cases} 1 & \text{if the } i\text{th coin flip is head} \\ 0 & \text{otherwise} \end{cases}$

$$\mathbf{E}[X_i] = \Pr[X_i = 1] = \frac{1}{2} \quad \mathbf{E}[X] = \sum_{i=1}^n \mathbf{E}[X_i] = \frac{n}{2}.$$

- Actually,

$$\mathbf{E}[X_i^2] = \mathbf{E}[X_i] = \frac{1}{2}.$$

$$\text{Var}[X_i] = \mathbf{E}[X_i^2] - (\mathbf{E}[X_i])^2 = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}.$$

$$\text{Var}[X] = \text{Var} \left[\sum_{i=1}^n X_i \right] = \sum_{i=1}^n \text{Var}[X_i] = \frac{n}{4}.$$

$$\Pr[X \geq 3n/4] \leq \Pr[|X - \mathbf{E}[X]| \geq n/4] \leq \frac{\text{Var}[X]}{(n/4)^2} = \frac{n/4}{(n/4)^2} = \frac{4}{n}.$$

Randomized Algorithms, CSIE, TKU, Taiwan

39

	$n=50$	$n=100$	$n=200$
$2/n$	0.2	0.02	0.01
$e^{-n/24}$	0.125	0.016	0.00025

Strengthen a weak classifier

- Suppose we have designed a device which can check in a very short time if a diamond is real or fake and the accuracy is around 66% for each examination.
- Such a device is somehow too weak to be used in practical.

Strengthen a weak classifier

- Suppose we have designed a device which can check in a very short time if a diamond is real or fake and the accuracy is around 66% for each examination.
- Such a device is somehow too weak to be used in practical.
- Let say we run the device for $n = 201$ times for each examination and output “True” if more than 101 of the results reveal that the diamond is real and output “False” otherwise.

Strengthen a weak classifier

- Suppose we have designed a device which can check in a very short time if a diamond is real or fake and the accuracy is around 66% for each examination.
- Such a device is somehow too weak to be used in practical.
- Let say we run the device for $n = 201$ times for each examination and output “True” if more than 101 of the results reveal that the diamond is real and output “False” otherwise. **(majority vote)**

$$\Pr[X \leq n/2] = \Pr\left[X - \frac{2n}{3} \leq -\frac{n}{6}\right] \leq e^{-(2n/3) \cdot (1/4)^2 \cdot (1/2)} = e^{-n/48} < 0.016.$$

X_i : 1 if i th test is correct and 0 otherwise

An application: Parameter Estimation

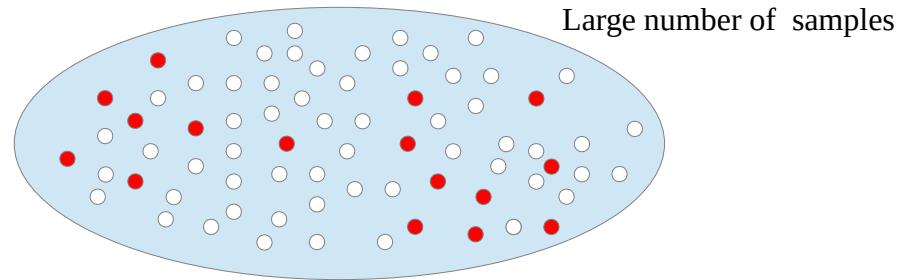
- Goal: evaluating the probability that a particular gene mutation occurs in the population.
- A lab test can determine if a DNA sample carries the mutation.
- However, the test is very **expensive**, so we want to obtain a relatively reliable estimate from a **small** number of samples.



<https://tinyurl.com/8fwxsnab>

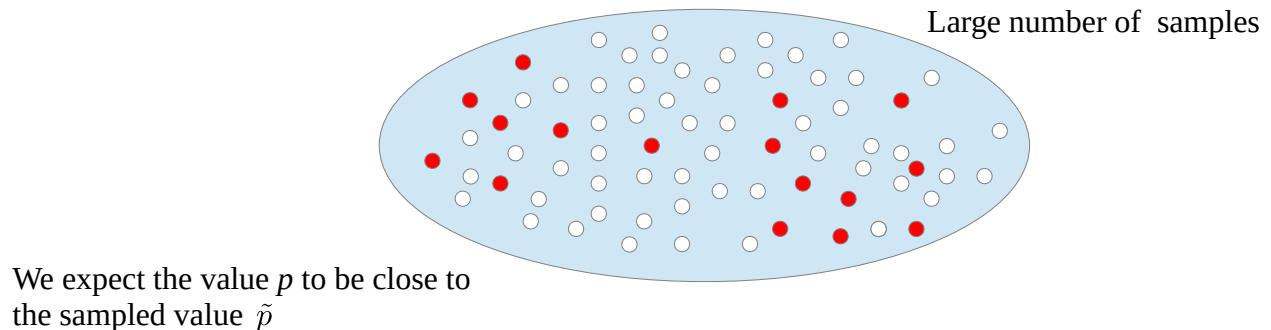
An application: Parameter Estimation

- p : be the unknown value we try to estimate.
- n : the number of samples we have
- $X = \tilde{p}n$: number of samples having the mutation



An application: Parameter Estimation

- p : be the unknown value we try to estimate.
- n : the number of samples we have
- $X = \tilde{p}n$: number of samples having the mutation



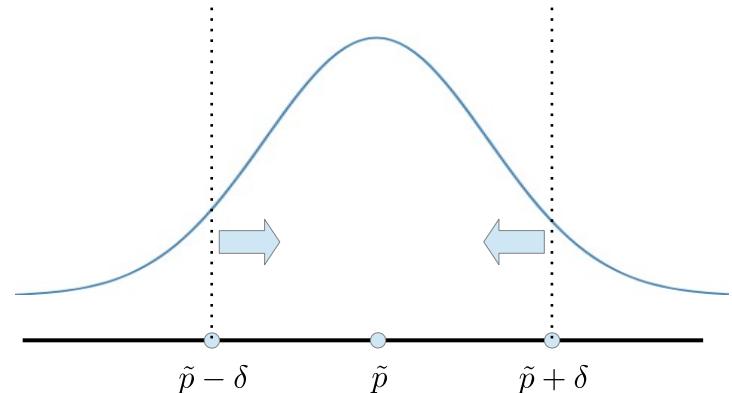
An application: Parameter Estimation

- Definition. A $1-\gamma$ **confidence interval** for a parameter p is an interval

$$[\tilde{p} - \delta, \tilde{p} + \delta]$$

such that

$$\Pr[p \in [\tilde{p} - \delta, \tilde{p} + \delta]] \geq 1 - \gamma.$$



We need to find values of δ and γ such that

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

An application: Parameter Estimation

- Apply the Chernoff bound:

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

An application: Parameter Estimation

- Apply the Chernoff bound:

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

- But we do not know the value of p , so it's not useful...
- Take $p \leq 1$,

$$\Pr[p \notin [\tilde{p} - \delta, \tilde{p} + \delta]] < e^{-n\delta^2/2} + e^{-n\delta^2/3}.$$

An application: Parameter Estimation

- Apply the Chernoff bound:

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

- But we do not know the value of p , so it's not useful...
- Take $p \leq 1$,

$$\Pr[p \notin [\tilde{p} - \delta, \tilde{p} + \delta]] < e^{-n\delta^2/2} + e^{-n\delta^2/3}.$$

Setting $\gamma = e^{-n\delta^2/2} + e^{-n\delta^2/3}$, we obtain a trade-off between δ and n .

The Hoeffding Bound

Wassily Hoeffding (1914–1991)

refer to <https://tinyurl.com/mzz7x8pb>



- Extending the Chernoff bound to general random variables with a **bounded range**.

Hoeffding's Lemma: Let X be a random such that $\Pr[X \in [a, b]] = 1$ and $\mathbf{E}[X] = 0$. Then for every $\lambda > 0$,

$$\mathbf{E}[e^{\lambda X}] \leq e^{\lambda^2(b-a)^2/8}.$$

The Hoeffding Bound

Wassily Hoeffding (1914–1991)

refer to <https://tinyurl.com/mzz7x8pb>



- Extending the Chernoff bound to general random variables with a **bounded range**.

Hoeffding's Lemma: Let X be a random such that $\Pr[X \in [a, b]] = 1$ and $\mathbf{E}[X] = 0$. Then for every $\lambda > 0$,

$$\mathbf{E}[e^{\lambda X}] \leq e^{\lambda^2(b-a)^2/8}.$$

Hoeffding's Bound: Let X_1, \dots, X_n be independent random variables such that for all $1 \leq i \leq n$, Then for every $\lambda > 0$, $\mathbf{E}[X_i] = \mu$ and $\Pr[a \leq X_i \leq b] = 1$. Then

$$\Pr \left[\left| \frac{1}{n} \sum_{i=1}^n X_i - \mu \right| \geq \epsilon \right] \leq 2e^{-2n\epsilon^2/(b-a)^2}.$$

The Hoeffding Bound

Wassily Hoeffding (1914–1991)

refer to <https://tinyurl.com/mzz7x8pb>



- Extending the Chernoff bound to general random variables with a **bounded range**.

Theorem: Let X_1, \dots, X_n be independent random variables such that

$\mathbf{E}[X_i] = \mu_i$ and $\Pr[a_i \leq X_i \leq b_i] = 1$ for constant a_i and b_i . Then,

$$\Pr \left[\left| \sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i \right| \geq \epsilon \right] \leq 2e^{-2\epsilon^2 / \sum_{i=1}^n (b_i - a_i)^2}.$$

Proofs

Proof of Hoeffding's Lemma

Hoeffding's Lemma: Let X be a random such that $\Pr[X \in [a, b]] = 1$ and $\mathbf{E}[X] = 0$.

Then for every $\lambda > 0$,

$$\mathbf{E}[e^{\lambda X}] \leq e^{\lambda^2(b-a)^2/8}.$$

- We assume $a < 0$ and $b > 0$. (Why?)
- $f(x) = e^{\lambda x}$ is a convex function.

Proof of Hoeffding's Lemma

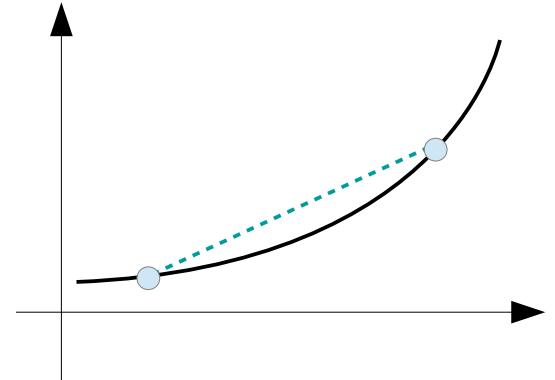
Hoeffding's Lemma: Let X be a random such that $\Pr[X \in [a, b]] = 1$ and $\mathbf{E}[X] = 0$.

Then for every $\lambda > 0$,

$$\mathbf{E}[e^{\lambda X}] \leq e^{\lambda^2(b-a)^2/8}.$$

- We assume $a < 0$ and $b > 0$. (Why?)
- $f(x) = e^{\lambda x}$ is a convex function.
 - For any $\alpha \in (0, 1)$,

$$f(\alpha a + (1 - \alpha)b) \leq \alpha e^{\lambda a} + (1 - \alpha)e^{\lambda b}.$$



Proof of Hoeffding's Lemma

Hoeffding's Lemma: Let X be a random such that $\Pr[X \in [a, b]] = 1$ and $\mathbf{E}[X] = 0$.

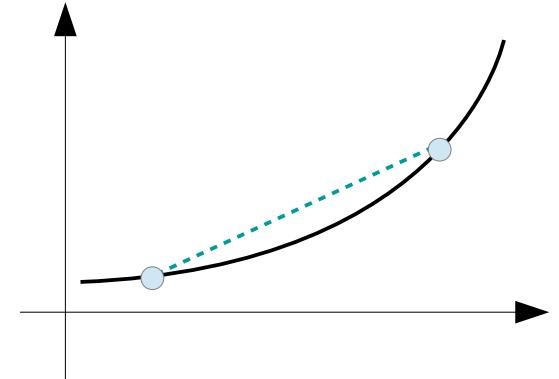
Then for every $\lambda > 0$,

$$\mathbf{E}[e^{\lambda X}] \leq e^{\lambda^2(b-a)^2/8}.$$

- We assume $a < 0$ and $b > 0$. (Why?)
- $f(x) = e^{\lambda x}$ is a convex function.
 - For any $\alpha \in (0, 1)$,

$$f(\alpha a + (1 - \alpha)b) \leq \alpha e^{\lambda a} + (1 - \alpha)e^{\lambda b}.$$

For $x \in [a, b]$, let $\alpha = \frac{b - x}{b - a}$, then $x = \alpha a + (1 - \alpha)b$



Proof of Hoeffding's Lemma

Hoeffding's Lemma: Let X be a random such that $\Pr[X \in [a, b]] = 1$ and $\mathbf{E}[X] = 0$.

Then for every $\lambda > 0$,

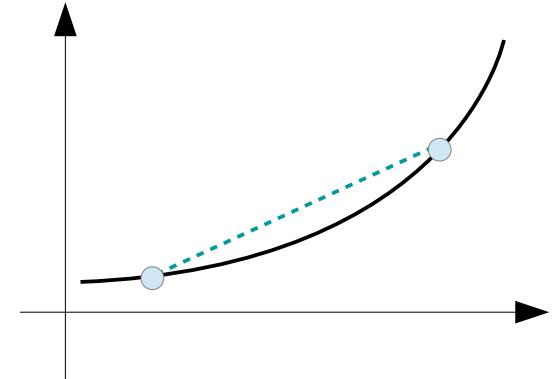
$$\mathbf{E}[e^{\lambda X}] \leq e^{\lambda^2(b-a)^2/8}.$$

- We assume $a < 0$ and $b > 0$. (Why?)
- $f(x) = e^{\lambda x}$ is a convex function.
 - For any $\alpha \in (0, 1)$,

$$f(\alpha a + (1 - \alpha)b) \leq \alpha e^{\lambda a} + (1 - \alpha)e^{\lambda b}.$$

For $x \in [a, b]$, let $\alpha = \frac{b - x}{b - a}$, then $x = \alpha a + (1 - \alpha)b$

$$\therefore e^{\lambda x} \leq \frac{b - x}{b - a} e^{\lambda a} + \frac{x - a}{b - a} e^{\lambda b}.$$



Proof of Hoeffding's Lemma

Hoeffding's Lemma: Let X be a random such that $\Pr[X \in [a, b]] = 1$ and $\mathbf{E}[X] = 0$.

Then for every $\lambda > 0$,

$$\mathbf{E}[e^{\lambda X}] \leq e^{\lambda^2(b-a)^2/8}.$$

$$\begin{aligned}\mathbf{E}[e^{\lambda X}] &\leq \mathbf{E}\left[\frac{b-X}{b-a}e^{\lambda a}\right] + \mathbf{E}\left[\frac{X-a}{b-a}e^{\lambda b}\right] \\ &= \frac{b}{b-a}e^{\lambda a} - \frac{\mathbf{E}[X]}{b-a}e^{\lambda a} - \frac{a}{b-a}e^{\lambda b} + \frac{\mathbf{E}[X]}{b-a}e^{\lambda b} \\ &= \frac{b}{b-a}e^{\lambda a} + \frac{a}{b-a}e^{\lambda b}.\end{aligned}$$

Proof of Hoeffding's Lemma

Hoeffding's Lemma: Let X be a random such that $\Pr[X \in [a, b]] = 1$ and $\mathbf{E}[X] = 0$.

Then for every $\lambda > 0$,

$$\mathbf{E}[e^{\lambda X}] \leq e^{\lambda^2(b-a)^2/8}.$$

$$\begin{aligned}\mathbf{E}[e^{\lambda X}] &\leq \mathbf{E}\left[\frac{b-X}{b-a}e^{\lambda a}\right] + \mathbf{E}\left[\frac{X-a}{b-a}e^{\lambda b}\right] \\ &= \frac{b}{b-a}e^{\lambda a} - \frac{\mathbf{E}[X]}{b-a}e^{\lambda a} - \frac{a}{b-a}e^{\lambda b} + \frac{\mathbf{E}[X]}{b-a}e^{\lambda b} \\ &= \frac{b}{b-a}e^{\lambda a} + \frac{a}{b-a}e^{\lambda b} \\ &= e^{\lambda a} \left(\frac{b}{b-a} - \frac{a}{b-a}e^{\lambda(b-a)} \right).\end{aligned}$$

Proof of Hoeffding's Lemma

Hoeffding's Lemma: Let X be a random such that $\Pr[X \in [a, b]] = 1$ and $\mathbf{E}[X] = 0$.

Then for every $\lambda > 0$,

$$\mathbf{E}[e^{\lambda X}] \leq e^{\lambda^2(b-a)^2/8}.$$

$$\begin{aligned}\mathbf{E}[e^{\lambda X}] &\leq \mathbf{E}\left[\frac{b-X}{b-a}e^{\lambda a}\right] + \mathbf{E}\left[\frac{X-a}{b-a}e^{\lambda b}\right] \\ &= \frac{b}{b-a}e^{\lambda a} - \frac{\mathbf{E}[X]}{b-a}e^{\lambda a} - \frac{a}{b-a}e^{\lambda b} + \frac{\mathbf{E}[X]}{b-a}e^{\lambda b} \\ &= \frac{b}{b-a}e^{\lambda a} + \frac{a}{b-a}e^{\lambda b} \\ &= e^{\lambda a} \left(\frac{b}{b-a} - \frac{a}{b-a}e^{\lambda(b-a)} \right) \quad \text{let } \theta = \frac{-a}{b-a} \\ &= e^{\lambda a}(1 - \theta + \theta e^{\lambda(b-a)}) \\ &= e^{-\theta\lambda(b-a)}(1 - \theta + \theta e^{\lambda(b-a)}).\end{aligned}$$

Proof of Hoeffding's Lemma

Hoeffding's Lemma: Let X be a random such that $\Pr[X \in [a, b]] = 1$ and $\mathbf{E}[X] = 0$.

Then for every $\lambda > 0$,

$$\mathbf{E}[e^{\lambda X}] \leq e^{\lambda^2(b-a)^2/8}.$$

$$\begin{aligned}\mathbf{E}[e^{\lambda X}] &\leq \mathbf{E}\left[\frac{b-X}{b-a}e^{\lambda a}\right] + \mathbf{E}\left[\frac{X-a}{b-a}e^{\lambda b}\right] \\&= \frac{b}{b-a}e^{\lambda a} - \frac{\mathbf{E}[X]}{b-a}e^{\lambda a} - \frac{a}{b-a}e^{\lambda b} + \frac{\mathbf{E}[X]}{b-a}e^{\lambda b} \\&= \frac{b}{b-a}e^{\lambda a} + \frac{a}{b-a}e^{\lambda b} \\&= e^{\lambda a}\left(\frac{b}{b-a} - \frac{a}{b-a}e^{\lambda(b-a)}\right) \\&= e^{\lambda a}(1 - \theta + \theta e^{\lambda(b-a)}) \\&= e^{-\theta\lambda(b-a)}(1 - \theta + \theta e^{\lambda(b-a)}) \\&= e^{\phi(\lambda(b-a))}.\end{aligned}$$

Let $\phi(t) = -\theta t + \ln(1 - \theta + \theta e^t)$.

Proof of Hoeffding's Lemma

Hoeffding's Lemma: Let X be a random such that $\Pr[X \in [a, b]] = 1$ and $\mathbf{E}[X] = 0$. Then for every $\lambda > 0$,

$$\mathbf{E}[e^{\lambda X}] \leq e^{\lambda^2(b-a)^2/8}.$$

$\mathbf{E}[e^{\lambda X}] \leq e^{\phi(\lambda(b-a))}$. Let $\phi(t) = -\theta t + \ln(1 - \theta + \theta e^t)$.

Proof of Hoeffding's Lemma

Hoeffding's Lemma: Let X be a random such that $\Pr[X \in [a, b]] = 1$ and $\mathbf{E}[X] = 0$.

Then for every $\lambda > 0$,

$$\mathbf{E}[e^{\lambda X}] \leq e^{\lambda^2(b-a)^2/8}.$$

$\mathbf{E}[e^{\lambda X}] \leq e^{\phi(\lambda(b-a))}$. Let $\phi(t) = -\theta t + \ln(1 - \theta + \theta e^t)$.

$\phi(0) = 0$, $\phi'(0) = 0$, and $\phi''(t) \leq 1/4$ for all t .

Proof of Hoeffding's Lemma

Hoeffding's Lemma: Let X be a random such that $\Pr[X \in [a, b]] = 1$ and $\mathbf{E}[X] = 0$.

Then for every $\lambda > 0$,

$$\mathbf{E}[e^{\lambda X}] \leq e^{\lambda^2(b-a)^2/8}.$$

$\mathbf{E}[e^{\lambda X}] \leq e^{\phi(\lambda(b-a))}$. Let $\phi(t) = -\theta t + \ln(1 - \theta + \theta e^t)$.

$\phi(0) = 0$, $\phi'(0) = 0$, and $\phi''(t) \leq 1/4$ for all t .

$$\phi(t) = \phi(0) + t\phi'(0) + \frac{1}{2}t^2\phi''(t') \leq \frac{1}{8}t^2. \leftarrow \text{Taylor's theorem, } \forall t > 0, \exists t' \in [0, t]$$

Proof of Hoeffding's Lemma

Hoeffding's Lemma: Let X be a random such that $\Pr[X \in [a, b]] = 1$ and $\mathbf{E}[X] = 0$.

Then for every $\lambda > 0$,

$$\mathbf{E}[e^{\lambda X}] \leq e^{\lambda^2(b-a)^2/8}.$$

$\mathbf{E}[e^{\lambda X}] \leq e^{\phi(\lambda(b-a))}$. Let $\phi(t) = -\theta t + \ln(1 - \theta + \theta e^t)$.

$\phi(0) = 0$, $\phi'(0) = 0$, and $\phi''(t) \leq 1/4$ for all t .

$$\phi(t) = \phi(0) + t\phi'(0) + \frac{1}{2}t^2\phi''(t') \leq \frac{1}{8}t^2.$$

For $t = \lambda(b - a)$,

$$\phi(\lambda(b - a)) \leq \frac{\lambda^2(b - a)^2}{8}.$$

Thus, $\mathbf{E}[e^{\lambda X}] \leq e^{\phi(\lambda(b-a))} \leq e^{\lambda^2(b-a)^2/8}$.

Proof of The Hoeffding's Bound

Hoeffding's Bound: Let X_1, \dots, X_n be independent random variables such that for all $1 \leq i \leq n$, Then for every $\lambda > 0$, $\mathbf{E}[X_i] = \mu$ and $\Pr[a \leq X_i \leq b] = 1$. Then

$$\Pr \left[\left| \frac{1}{n} \sum_{i=1}^n X_i - \mu \right| \geq \epsilon \right] \leq 2e^{-2n\epsilon^2/(b-a)^2}.$$

Let $Z_i = X_i - \mathbf{E}[X_i]$ and $Z = \frac{1}{n} \sum_{i=1}^n Z_i$.

For any $\lambda > 0$, by Markov's inequality:

Proof of The Hoeffding's Bound

Hoeffding's Bound: Let X_1, \dots, X_n be independent random variables such that for all $1 \leq i \leq n$, Then for every $\lambda > 0$, $\mathbf{E}[X_i] = \mu$ and $\Pr[a \leq X_i \leq b] = 1$. Then

$$\Pr\left[\left|\frac{1}{n} \sum_{i=1}^n X_i - \mu\right| \geq \epsilon\right] \leq 2e^{-2n\epsilon^2/(b-a)^2}.$$

Let $Z_i = X_i - \mathbf{E}[X_i]$ and $Z = \frac{1}{n} \sum_{i=1}^n Z_i$.

For any $\lambda > 0$, by Markov's inequality:

$$\begin{aligned}\Pr[Z \geq \epsilon] &= \Pr[e^{\lambda Z} \geq e^{\lambda \epsilon}] \leq \frac{\mathbf{E}[e^{\lambda Z}]}{e^{\lambda \epsilon}} &= \frac{\prod_{i=1}^n \mathbf{E}[e^{\lambda Z_i/n}]}{e^{\lambda \epsilon}} \\ &\leq \frac{\prod_{i=1}^n e^{\lambda^2(b-a)^2/(8n^2)}}{e^{\lambda \epsilon}} \\ &\leq \frac{e^{\lambda^2(b-a)^2/8n}}{e^{\lambda \epsilon}}.\end{aligned}$$

Proof of The Hoeffding's Bound

Hoeffding's Bound: Let X_1, \dots, X_n be independent random variables such that for all $1 \leq i \leq n$, Then for every $\lambda > 0$, $\mathbf{E}[X_i] = \mu$ and $\Pr[a \leq X_i \leq b] = 1$. Then

$$\Pr\left[\left|\frac{1}{n} \sum_{i=1}^n X_i - \mu\right| \geq \epsilon\right] \leq 2e^{-2n\epsilon^2/(b-a)^2}.$$

Let $Z_i = X_i - \mathbf{E}[X_i]$ and $Z = \frac{1}{n} \sum_{i=1}^n Z_i$.

For any $\lambda > 0$, by Markov's inequality:

$$\Pr[Z \geq \epsilon] \leq \frac{e^{\lambda^2(b-a)^2/8n}}{e^{\lambda\epsilon}}$$

Note: $Z_i/n \in [(a - \mu)/n, (b - \mu)/n]$.

Setting $\lambda = \frac{4n\epsilon}{(b-a)^2}$,

Proof of The Hoeffding's Bound

Hoeffding's Bound: Let X_1, \dots, X_n be independent random variables such that for all $1 \leq i \leq n$, Then for every $\lambda > 0$, $\mathbf{E}[X_i] = \mu$ and $\Pr[a \leq X_i \leq b] = 1$. Then

$$\Pr\left[\left|\frac{1}{n} \sum_{i=1}^n X_i - \mu\right| \geq \epsilon\right] \leq 2e^{-2n\epsilon^2/(b-a)^2}.$$

Let $Z_i = X_i - \mathbf{E}[X_i]$ and $Z = \frac{1}{n} \sum_{i=1}^n Z_i$.

For any $\lambda > 0$, by Markov's inequality:

$$\Pr[Z \geq \epsilon] \leq \frac{e^{\lambda^2(b-a)^2/8n}}{e^{\lambda\epsilon}}$$

$$\Pr\left[\left|\frac{1}{n} \sum_{i=1}^n X_i - \mu\right| \geq \epsilon\right] = \Pr[Z \geq \epsilon] \leq e^{-2n\epsilon^2/(b-a)^2}.$$

Proof of The Hoeffding's Bound

Hoeffding's Bound: Let X_1, \dots, X_n be independent random variables such that for all $1 \leq i \leq n$, Then for every $\lambda > 0$, $\mathbf{E}[X_i] = \mu$ and $\Pr[a \leq X_i \leq b] = 1$. Then

$$\Pr\left[\left|\frac{1}{n} \sum_{i=1}^n X_i - \mu\right| \geq \epsilon\right] \leq 2e^{-2n\epsilon^2/(b-a)^2}.$$

Let $Z_i = X_i - \mathbf{E}[X_i]$ and $Z = \frac{1}{n} \sum_{i=1}^n Z_i$.

$$\Pr\left[\left|\frac{1}{n} \sum_{i=1}^n X_i - \mu\right| \geq \epsilon\right] = \Pr[Z \geq \epsilon] \leq e^{-2n\epsilon^2/(b-a)^2}.$$

For $\Pr[Z \leq -\epsilon]$ with $\lambda = -\frac{4n\epsilon}{(b-a)^2}$

$$\Pr\left[\left|\frac{1}{n} \sum_{i=1}^n X_i - \mu\right| \geq -\epsilon\right] = \Pr[Z \leq -\epsilon] \leq e^{-2n\epsilon^2/(b-a)^2}.$$