SDS 384 11: Theoretical Statistics

Lecture 3: Concentration inequalities

Purnamrita Sarkar

Department of Statistics and Data Science

The University of Texas at Austin

`www.cs.cmu.edu/∼psarkar/teaching`

Remember Markov's inequality?

Theorem

For $X \geq 0$, $E[X] \leq \infty$, $t > 0$, we have:

$$P(X \geq t) \leq \frac{E[X]}{t}$$

Use total expectation theorem.

$$E[X] = E[X|X \geq t]P(X \geq t) + E[X|X < t]P(X < t)$$

$\square$

Remember Markov's inequality?

Theorem

For $X \geq 0$, $E[X] \leq \infty$, $t > 0$, we have:

$$P(X \geq t) \leq \frac{E[X]}{t}$$

Use total expectation theorem.

$$E[X] = E[X|X \geq t]P(X \geq t) + E[X|X < t]P(X < t)$$
$$\geq E[X|X \geq t]P(X \geq t)$$

$\square$

Remember Markov's inequality?

Theorem

For $X \geq 0$, $E[X] \leq \infty$, $t > 0$, we have:

$$P(X \geq t) \leq \frac{E[X]}{t}$$

Use total expectation theorem.

$$
\begin{aligned}
E[X] &= E[X|X \geq t]P(X \geq t) + E[X|X < t]P(X < t) \\
&\geq E[X|X \geq t]P(X \geq t) \\
&\geq tP(X \geq t) \\
P(X \geq t) &\leq \frac{E[X]}{t}
\end{aligned}
$$

$\square$

Higher order moments

Theorem (Chebyshev's)

For $t > 0$

$$P(|X - \mu| \geq t) = P((X - \mu)^2 \geq t^2) \leq \frac{E[(X - \mu)^2]}{t^2} = \frac{var(X)}{t^2}$$

Theorem (Higher order markov)

For $t > 0$

$$P(|X - \mu| \geq t) = P((X - \mu)^k \geq t^k) \leq \frac{E[(X - \mu)^k]}{t^k}$$

Chernoff bound

Theorem (Chernoff bound for Bernoullis)

Let $X_i \in \{0, 1\}$ be independent random variables with $E[X_i] = p_i$. Let $X := \sum_i X_i, \mu := \sum_i p_i$. For $0 < \delta < 1$,

$$P(X \geq \mu(1+\delta)) \leq e^{-\delta^2 \mu/3} \qquad P(X \leq \mu(1-\delta)) \leq e^{-\delta^2 \mu/2}$$

Proof.

$$P(X \geq \mu(1+\delta)) = \inf_{\lambda \geq 0} P(e^{\lambda X} \geq e^{\lambda \mu(1+\delta)}) \leq \inf_{\lambda \geq 0} e^{-\lambda \mu(1+\delta)} \underbrace{E\left[e^{\lambda X}\right]}_{\text{MGF of } X}$$

$\square$

Chernoff continued

$$
\inf_{\lambda \geq 0} e^{-\lambda\mu(1+\delta)} E\left[e^{\lambda X}\right] = \inf_{\lambda \geq 0} e^{-\lambda\mu(1+\delta)} \prod_i E\left[e^{\lambda X_i}\right]
$$

$$
= \inf_{\lambda \geq 0} e^{-\lambda\mu(1+\delta)} \prod_i (e^\lambda p_i + 1 - p_i)
$$

$$
(\text{Since } 1 + x \leq e^x \text{ for } x \geq 0) \leq \inf_{\lambda \geq 0} e^{-\lambda\mu(1+\delta)} \prod_i e^{p_i(e^\lambda - 1)}
$$

$$
= \inf_{\lambda \geq 0} e^{-\lambda\mu(1+\delta) + \mu(e^\lambda - 1)}
$$

$$
(\text{minimized at } \lambda = \log(1+\delta)) = e^{\mu(\delta - (1+\delta)\log(1+\delta))}
$$

$$
((1+\delta)\log(1+\delta) \geq \delta + \delta^2/3, \text{ for } \delta < 1) \leq e^{-\mu\delta^2/3}
$$

Is it tight?

Theorem (Chernoff bound for Gaussians)

Let $X_i \sim N(\mu, \sigma^2)$ be independent random variables. Let $X := \sum_i X_i$.

$$P(X/n - \mu \geq t) \leq e^{-\frac{nt^2}{2\sigma^2}}$$

Proof.

Following in the same lines:

$$P(X/n - \mu \geq t) \inf_{\lambda \geq 0} e^{-n\lambda t} E\left[e^{\lambda(X-n\mu)}\right] = \inf_{\lambda \geq 0} e^{-n\lambda t} \prod_i E\left[e^{\lambda(X_i - \mu)}\right]$$

$$\text{(Since } E[e^{\lambda X}] = e^{\lambda\mu + \sigma^2\lambda^2/2}) \qquad = \inf_{\lambda \geq 0} e^{-n\lambda t + n\sigma^2\lambda^2/2}$$

$$\text{(Since } \lambda = t/\sigma^2 \text{ minimizes this)} \qquad = e^{-\frac{nt^2}{2\sigma^2}}$$

$\square$

5

Is it tight?

- Let $Z \sim N(0, 1)$. We can show that for $z > 0$,

$$\phi(z) \left(\frac{1}{z} - \frac{1}{z^3} \right) \leq P(Z \geq z) \leq \phi(z) \left(\frac{1}{z} - \frac{1}{z^3} + \frac{3}{z^5} \right),$$

where $\phi(z)$ is the density of a standard normal.

- Since $\bar{X}_n \sim N(\mu, \sigma^2/n)$, $\lim\limits_{n \to \infty} P(\bar{X}_n - \mu \geq t)/n = -\dfrac{t^2}{2\sigma^2}$

- So the Chernoff bound is asymptotically tight.

Hoeffding's lemma

Theorem

For a random variable $X \in [a, b]$ with $E[X] = \mu$ and $\lambda \in \mathbb{R}$,

$$M_{X-\mu}(\lambda) \leq e^{\frac{\lambda^2(b-a)^2}{8}}$$

- In comparison, for a Gaussian random variable $X \sim N(\mu, \sigma^2)$,

$$M_{X-\mu}(\lambda) \leq e^{\frac{\lambda^2 \sigma^2}{2}}$$

- For a bounded random variable $X \in [a, b]$, $\text{var}(X) \leq (b - a)^2/4$ from Popoviciu's inequality.

Hoeffding's lemma

Theorem

For a random variable $X \in [a, b]$ with $E[X] = \mu$ and $\lambda \in \mathbb{R}$,

$$M_{X-\mu}(\lambda) \leq e^{\frac{\lambda^2(b-a)^2}{8}}$$

- In comparison, for a Gaussian random variable $X \sim N(\mu, \sigma^2)$,

$$M_{X-\mu}(\lambda) \leq e^{\frac{\lambda^2 \sigma^2}{2}}$$

- For a bounded random variable $X \in [a, b]$, $\text{var}(X) \leq (b-a)^2/4$ from Popoviciu's inequality.
  - Recall that $E[(X-t)^2]$ is minimized at $t = E[X]$.

Hoeffding's lemma

Theorem

For a random variable $X \in [a, b]$ with $E[X] = \mu$ and $\lambda \in \mathbb{R}$,

$$M_{X-\mu}(\lambda) \leq e^{\frac{\lambda^2(b-a)^2}{8}}$$

- In comparison, for a Gaussian random variable $X \sim N(\mu, \sigma^2)$,

$$M_{X-\mu}(\lambda) \leq e^{\frac{\lambda^2 \sigma^2}{2}}$$

- For a bounded random variable $X \in [a, b]$, $\text{var}(X) \leq (b-a)^2/4$ from Popoviciu's inequality.
    - Recall that $E[(X - t)^2]$ is minimized at $t = E[X]$.
    - So $\text{var}(X) \leq E[(X - (a+b)/2)^2] \leq \dfrac{(b-a)^2}{4}$

MGF of Rademacher variables

A Rademacher random variable ϵ takes values in $\{-1, 1\}$ equiprobable.

$$E[e^{\lambda\epsilon}] = \frac{e^{\lambda} + e^{-\lambda}}{2}$$
$$= \sum_i \frac{\lambda^{2i}}{(2i)!}$$
$$\leq \sum_i \frac{\lambda^{2i}}{2^i i!}$$
$$= e^{\lambda^2/2}$$

Hoeffding's Lemma: weaker version

Theorem

For a random variable $X \in [a, b]$ with $E[X] = \mu$ and $\lambda \in \mathbb{R}$,

$$M_{X-\mu}(\lambda) \leq e^{\frac{\lambda^2 (b-a)^2}{2}}$$

Hoeffding's Lemma: weaker version

Theorem

For a random variable $X \in [a, b]$ with $E[X] = \mu$ and $\lambda \in \mathbb{R}$,

$$M_{X-\mu}(\lambda) \le e^{\frac{\lambda^2(b-a)^2}{2}}$$

- Consider an iid copy X' of X. Also consider a Radamacher random variable ϵ.

$$
\begin{aligned}
E[e^{\lambda(X-E[X])}] = E[e^{\lambda(X-E_{X'}[X'])}] &= E_X[e^{\lambda E_{X'}(X-X')}] \\
&\le E_{X,X'} e^{\lambda(X-X')} = E_{X,X'} E_\epsilon e^{\epsilon\lambda(X-X')} \\
&\le E_{X,X'} e^{\frac{\lambda^2(X-X')^2}{2}} \le e^{\frac{\lambda^2(b-a)^2}{2}}
\end{aligned}
$$

Hoeffding's inequality

Theorem

Consider i.i.d $X_i \in [a_i, b_i]$. Let $X = \sum_i X_i$.

$$P(X - E[X] \geq t) \leq e^{-\frac{2t^2}{\sum_i (b_i - a_i)^2}}$$

Proof.

$$
\begin{aligned}
P(X - E[X] \geq t) &\leq \inf_{\lambda \geq 0} e^{-\lambda t} E[e^{\lambda(X - E[X])}] \\
&\leq \inf_{\lambda \geq 0} e^{-\lambda t} \prod_i E\left[e^{\lambda(X_i - E[X_i])}\right] \\
&\leq \inf_{\lambda \geq 0} e^{-\lambda t + \frac{\lambda^2 \sum_i (b_i - a_i)^2}{8}} = e^{-\frac{2t^2}{\sum_i (b_i - a_i)^2}}
\end{aligned}
$$

How do we use this?

Consider n fair coins $X_i \in \{0, 1\}$. The Hoeffding inequality gives us

$$P(|\sum_i X_i - n/2| \geq t) \leq 2e^{-2t^2/n}$$

- How to pick t?
- Set the failure probability at δ.
- So $t = \sqrt{\frac{n}{2} \log(1/\delta)}$, i.e. we can also write the bound as

$$P\left(\left|\sum_i X_i - n/2\right| \geq \sqrt{\frac{n}{2} \log(1/\delta)}\right) \leq \delta$$

Sub Gaussian random variables

Definition

X is sub-gaussian with parameter σ^2 if, for all $\lambda \in \mathbb{R}$,

$$\log M_{X-\mu}(\lambda) \leq \frac{\lambda^2 \sigma^2}{2}$$

Sub Gaussian random variables

Definition

X is sub-gaussian with parameter σ^2 if, for all $\lambda \in \mathbb{R}$,

$$\log M_{X-\mu}(\lambda) \leq \frac{\lambda^2 \sigma^2}{2}$$

- Gaussian random variables are also sub-gaussian.

Sub Gaussian random variables

Definition

X is sub-gaussian with parameter σ^2 if, for all $\lambda \in \mathbb{R}$,

$$\log M_{X-\mu}(\lambda) \leq \frac{\lambda^2 \sigma^2}{2}$$

- Gaussian random variables are also sub-gaussian.
- X is sub-gaussian iff $-X$ is also sub-gaussian

Sub-Gaussian random variables

- X is sub-gaussian implies that
  - $P(X - \mu \geq t) \leq e^{-t^2/2\sigma^2}$
  - $P(X - \mu \leq -t) \leq e^{-t^2/2\sigma^2}$
  - $P(|X - \mu| \geq t) \leq 2e^{-t^2/2\sigma^2}$

Sub-Gaussian random variables

- Let X_1, X_2 be independent sub-gaussian random variables with parameters σ_1 and σ_2. Then $aX_1 + bX_2$ is sub-gaussian with parameter $a^2\sigma_1^2 + b^2\sigma_2^2$.

Sub-Gaussian random variables

- Let X_1, X_2 be independent sub-gaussian random variables with parameters σ_1 and σ_2. Then $aX_1 + bX_2$ is sub-gaussian with parameter $a^2\sigma_1^2 + b^2\sigma_2^2$.

$$\begin{aligned} M_{a(X_1-\mu_1)+b(X_2-\mu_2)}(\lambda) &= E[e^{\lambda(a(X_1-\mu_1)+b(X_2-\mu_2))}] \\ &= E[e^{\lambda a(X_1-\mu_1)}]E[e^{\lambda b(X_2-\mu_2)}] \\ &\leq e^{\frac{\lambda^2(a^2\sigma_1^2+b^2\sigma_2^2)}{2}} \end{aligned}$$