Trace Estimation II
– Concentration Inequalities –
Lecture 6

F. M. Faulstich

01/26/2024

Concentration Inequalities

Wikipedia:

> "[...] concentration inequalities provide mathematical bounds on the probability of a random variable deviating from some value"

Concentration Inequalities

Wikipedia:

> "[...] concentration inequalities provide mathematical bounds on the probability of a random variable deviating from some value"

- Markov's inequality:
  Let X be a nonnegative random variable and $a > 0$, then

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}.$$

- Chebychev's inequality:
  Let X be a random variable with finite non-zero variance σ^2.

$$\mathbb{P}(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2} \qquad \forall k > 0.$$

Concentration Inequalities

Wikipedia:

> "[...] concentration inequalities provide mathematical bounds on the probability of a random variable deviating from some value"

- Markov's inequality:
  Let X be a nonnegative random variable and $a > 0$, then
  $$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}.$$
- Chebychev's inequality:
  Let X be a random variable with finite non-zero variance σ^2.
  $$\mathbb{P}(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2} \qquad \forall k > 0.$$

... What do we use them for?

Why concentration inequalities?

- Concentration inequalities can provide quantitative estimates of the likely size of the error when a randomized algorithm is executed.

Why concentration inequalities?

- Concentration inequalities can provide quantitative estimates of the likely size of the error when a randomized algorithm is executed.
- We have seen that for X_i i.i.d Chebychev's inequality yields

$$\mathbb{P}\left(\left|\frac{1}{k}\sum_{i=1}^{k} X_i - \mu\right| \geq t\right) \leq \frac{\sigma^2}{kt^2}$$

where $\mathbb{E}(X) = \mu$ and $\mathbb{V}(X) = \sigma^2$.

Why concentration inequalities?

- Concentration inequalities can provide quantitative estimates of the likely size of the error when a randomized algorithm is executed.
- We have seen that for X_i i.i.d Chebychev's inequality yields

$$\mathbb{P}\left(\left|\frac{1}{k}\sum_{i=1}^{k} X_i - \mu\right| \geq t\right) \leq \frac{\sigma^2}{kt^2}$$

  where $\mathbb{E}(X) = \mu$ and $\mathbb{V}(X) = \sigma^2$.
- Question:
  We want to estimate μ by $\frac{1}{k}\sum_{i=1}^{k} X_i$ up to an error ε with a failure probability δ. How many samples do we need at most?

Why concentration inequalities?

- Concentration inequalities can provide quantitative estimates of the likely size of the error when a randomized algorithm is executed.
- We have seen that for X_i i.i.d Chebychev's inequality yields

$$\mathbb{P}\left(\left|\frac{1}{k}\sum_{i=1}^{k} X_i - \mu\right| \geq t\right) \leq \frac{\sigma^2}{kt^2}$$

  where $\mathbb{E}(X) = \mu$ and $\mathbb{V}(X) = \sigma^2$.
- Question:
  We want to estimate μ by $\frac{1}{k}\sum_{i=1}^{k} X_i$ up to an error ε with a failure probability δ. How many samples do we need at most?
  The accuracy dictates $t = \varepsilon$, and

$$\delta = \frac{\sigma^2}{kt^2} \quad \Leftrightarrow \quad k = \frac{\sigma^2}{\delta\varepsilon^2}$$

How tight are Markov & Chebychev?

What is the ideal scenario?

- Central limit theorem suggests that $\bar{X}_k$ approximates $\mathcal{N}(\mu, \sigma^2/k)$
- So, we would like/expect a result like

$$\mathbb{P}\left(|\bar{X}_k - \mu| \geq t\right) \overset{?}{\lessapprox} \exp\left(-\frac{kt^2}{2\sigma^2}\right)$$

(Probability of being in the tail of a Gaussian)

Hoeffding's Inequality

- Let $X_1, ..., X_k$ be i.i.d and such that $a \leq X_i \leq b$ almost surely. Then

$$\mathbb{P}\left(\bar{X}_k - \mu \geq t\right) \leq \exp\left(-\frac{2kt^2}{(b-a)^2}\right)$$
$$\mathbb{P}\left(|\bar{X}_k - \mu| \geq t\right) \leq 2\exp\left(-\frac{2kt^2}{(b-a)^2}\right)$$

for all $t > 0$.

... We need some tools to prove this!

Tools

- Cramér–Chernoff method:
  Let X be a random variable then

$$\mathbb{P}(X \geq x) \leq \min_{r>0} \exp(-rx)\mathbb{E}(\exp(rX))$$

- Hoeffding's Lemma:
  Let X be a random variable with $a \leq X \leq b$. Then for all $t \in \mathbb{R}$ the momentum generating function of X satisfies:

$$\mathbb{E}(e^{tX}) \leq \exp\left(t\mu + \frac{1}{8}t^2(b-a)^2\right)$$

Proof of Hoeffding's inequality I

- We apply the Cramér-Chernoff method

$$\mathbb{P}\left(\bar{X}_k - \mu \geq t\right) \leq \min_{r>0} \exp(-rt) \prod_{i=1}^{k} \mathbb{E}\left[\exp\left(r\left(\frac{X_i}{k} - \frac{\mu}{k}\right)\right)\right]$$

- Hoeffding's Lemma yields

$$\mathbb{E}\left[\exp\left(r\left(\frac{X}{k} - \frac{\mu}{k}\right)\right)\right] \leq \exp\left(\frac{1}{8}\frac{r^2}{k^2}(b-a)^2\right)$$

- Hence

$$\mathbb{P}\left(\bar{X}_k - \mu \geq t\right) \leq \min_{r>0} \exp\left(-rt + \frac{1}{8}\frac{r^2}{k}(b-a)^2\right)$$

Proof of Hoeffding's inequality II

- Find

$$\min_{r>0} \exp\left(-rt + \frac{1}{8}\frac{r^2}{k}(b-a)^2\right)$$

Proof of Hoeffding's inequality II

- Find
$$\min_{r>0} \exp\left(-rt + \frac{1}{8}\frac{r^2}{k}(b-a)^2\right) = \exp\left(-\frac{2t^2k}{(b-a)^2}\right)$$

- Putting all together, we obtain
$$\mathbb{P}\left(\bar{X}_k - \mu \geq t\right) \leq \exp\left(-\frac{2t^2k}{(b-a)^2}\right)$$

- Since
$$\mathbb{P}\left(\bar{X}_k - \mu \leq -t\right) = \mathbb{P}\left(\mu - \bar{X}_k \geq t\right) \leq \exp\left(-\frac{2t^2k}{(b-a)^2}\right)$$

  we get the second result
$$\mathbb{P}\left(|\bar{X}_k - \mu| \geq t\right) \leq 2\exp\left(-\frac{2t^2k}{(b-a)^2}\right)$$

Discussion of Hoeffding's Inequality

- Hoeffding's inequality is similar to the anticipated scenario of the central limit theorem

$$\mathbb{P}\left(|\bar{X}_k - \mu| \geq t\right) \overset{?}{\lessapprox} \exp\left(-\frac{kt^2}{2\sigma^2}\right)$$

  but it replaces σ^2 by the larger quantity $(b-a)^2/4$

- Note that

$$\sigma^2 \leq (b-a)^2/4$$

Berstein's Inequality

- Let $X_1, ..., X_k$ be i.i.d and such that $|X - \mathbb{E}(X)| \leq B$ almost surely. Then
$$\mathbb{P}\left(|\bar{X}_k - \mu| \geq t\right) \leq 2\exp\left(-\frac{kt^2/2}{\sigma^2 + Bt/3}\right)$$
- For small values of t σ^2 dominates $Bt/3$ and we get what we anticipated from the CLT
- For large t we get
$$\mathbb{P}\left(|\bar{X}_k - \mu| \geq t\right) \overset{\text{large t}}{\lessapprox} 2\exp\left(-\frac{kt3}{2B}\right)$$
which is exponentially small in t rather than t^2.

$\Rightarrow$ For small t, Berstein's inequality is tighter than Hoeffding's, for large t however Hoeffding's is tighter.

Application to the trace estimator

- Let $X = \boldsymbol{\omega}^* \mathbf{A} \boldsymbol{\omega}$ with $\boldsymbol{\omega}$ Rademacher and $\mathbf{A} \in \mathbb{H}_n$. Then

$$\mathbb{V}(\boldsymbol{\omega}^* \mathbf{A} \boldsymbol{\omega}) = 4 \sum_{i<j} (\mathbf{A})_{i,j}^2 \leq 2 \|\mathbf{A}\|_F^2$$

  and Chebychev yields

$$\mathbb{P}(|\bar{X}_k - \mathrm{Tr}(\mathbf{A})| \geq t) \leq \frac{2\|\mathbf{A}\|_F^2}{kt^2}$$

- Using Bernstein's inequality we can establish

$$\mathbb{P}(|\bar{X}_k - \mathrm{Tr}(\mathbf{A})| \geq t) \leq 2 \exp\left(-\frac{kt^2}{3\|\mathbf{A}\|_F^2 + 4tn\|\mathbf{A}\|/3} \right)$$

More Concentration inequalities

- Vysochanskij–Petunin inequality
- Paley–Zygmund inequality
- Cantelli's inequality
- Azuma's inequality
- McDiarmid's inequality

  ⋮

A prior error estimates for trace estimator

Gratton and Titley-Peloquin (2018):

- Let $\mathbf{A} \in \mathbb{H}_n(\mathbb{R})$ and $0 \preccurlyeq \mathbf{A}$ be non-zero. For $\tau > 1$ and $k \leq n$, the Girard–Hutchinson estimator with $\boldsymbol{\omega} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ then fulfills

$$\mathbb{P}(\bar{X}_k \geq \tau \mathrm{Tr}(\mathbf{A})) \leq \exp\left(-\frac{1}{2} k \ \mathrm{intdim}(\mathbf{A}) (\sqrt{\tau} - 1)^2\right)$$

$$\mathbb{P}(\bar{X}_k \geq \tau^{-1} \mathrm{Tr}(\mathbf{A})) \leq \exp\left(-\frac{1}{4} k \ \mathrm{intdim}(\mathbf{A}) (\tau^{-1} - 1)^2\right)$$