Concentration Inequalities

- Is there a way to measure how close the expectation $E[X]$ is to the actual values of X?

- We'll prove a series of increasingly stronger interpretation of expectation:

Markov inequality. Thm (Markov 1890s): For a positive random var. X & $a > 0$, $P(X \geq a) \leq E[X]/a$;
(it is of interest if $a > E[X]$)

Pf: • Idea: Larger the value is, smaller the prob.

• $E[X] = \sum_{x \geq 0} P(X=x) \cdot x = \sum_{0 \leq x < a} P(X=x) \cdot x + \sum_{x \geq a} \cdots$

$\geq 0 + \sum_{x \geq a} P(X=x) \cdot a = P(X \geq a) \cdot a$

$\Rightarrow P(X \geq a) \leq E[X]/a$. $\square$

Variance. Let's now make "discrepancy" $X - E[X]$ estimate more formal.

$\triangleright$ $E[X - E[X]] = E[X] - E[E[X]] = 0$.

+ve & -ve discrepancies cancel out!

- Let's suppress the sign to avoid cancellation.

Defn: • Variance of X is $\mathrm{var}(X) := E[(X-E[X])^2]$

• Standard-deviation of X is $\sigma(X) := \sqrt{\mathrm{var}(X)}$.

▷ $\mathrm{var}(X) = E[X^2 - 2\cdot E[X]\cdot X + E[X]^2] = E[X^2] - E[X]^2$.

▷ $\forall a \in \mathbb{R}$, $\mathrm{var}(a\cdot X) = a^2 \cdot \mathrm{var}(X)$.

Standard deviation of Bernoulli (with $P(H) =: p$).

▷ $\mathrm{var}(X) = E[X^2] - E[X]^2 = p - p^2$.

▷ $\sigma(X) = \sqrt{p(1-p)} \leq 1/2$.

▷ Unbiased coin ($p = 1/2$): $\sigma(X) = E[X] = E[X^2] = \frac{1}{2}$.

$\Rightarrow$ Unbiased coin maximizes the deviation!

- But, what about $|X - E[X]| = ?$

Chebyshev inequality. Thm (Chebyshev 1867): For a random variable X & $a > 0$,

$$P(|X - E[X]| \geq a) \leq Var(X)/a^2 \leq (\sigma(X)/a)^2.$$

makes sense if $a > \sigma(X)$

Pf: • Idea: Use Markov's on $(X - E[X])^2$.

• $LHS = P((X - E[X])^2 \geq a^2) \leq \dfrac{E[(X - E[X])^2]}{a^2} = \dfrac{Var(X)}{a^2} = \left(\dfrac{\sigma(X)}{a}\right)^2.$ $\square$

$\Rightarrow \sigma(X)$ tightly controls <u>deviation from the mean</u>!

$\triangleright\ P(X \geq E[X] + 2\cdot\sigma \text{ OR } X \leq E[X] - 2\cdot\sigma) \leq 1/4.$

Weak Linearity of Variance

Lemma: Let $\{X_i \mid i \in [n]\}$ be 2-wise (pairwise) independent random variables. Then, $var(\sum_i X_i) = \sum_i var(X_i)$.

Pf: • $var(\sum X_i) = E[(\sum_i X_i)^2] - (E[\sum_i X_i])^2$

$= E[\sum_{i,j} X_i X_j] - \sum_{i,j} E[X_i] \cdot E[X_j]$

$= \sum_{i,j} \left(E[X_i X_j] - E[X_i] \cdot E[X_j]\right)$

• Note: $E[X_1 X_2] = \sum_{k_1, k_2 \in \mathbb{R}} P(X_1 = k_1 \wedge X_2 = k_2) \cdot k_1 k_2$

$= \sum_{k_1, k_2} P(X_1 = k_1) \cdot k_1 \cdot P(X_2 = k_2) \cdot k_2$ [by 2-wise indep.]

$$= \left(\sum_{k_1} P(X_1 = k_1) \cdot k_1 \right) \cdot \left(\sum_{k_2} P(X_2 = k_2) \cdot k_2 \right)$$

$$= E[X_1] \cdot E[X_2]$$

↳ Expectation is <u>multiplicative</u> on indep. rnd. var.!

• So, $\mathrm{var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=j \in [n]} \left(E[X_i \cdot X_j] - E[X_i] \cdot E[X_j] \right)$

other terms $i \neq j$ cancel by <u>2-wise</u> indep.

$$= \sum_{i=1}^{n} \mathrm{var}(X_i).$$

$\square$

Weak Law of large numbers

Corollary: Define $\bar{X} := (\sum_{i=1}^{n} X_i)/n$ as the average of 2-wise indep. rnd. variables X_i's (each identical to rnd. variable X). Then, $\forall a > 0$:

$$P(|\bar{X} - E[X]| \geq a) \leq var(X)/na^2.$$

Pf: • Apply Chebyshev's; linearity of variance & $E[\cdot]$.

▷ $E[\bar{X}] = \sum_i E[X_i]/n = E[X]$.

▷ $var(\bar{X}) = var(\sum X_i/n) = \frac{var(\sum X_i)}{n^2} = \frac{n \cdot var(X)}{n^2}$

• Now, Chebyshev gives $P(\cdots) \leq var(X)/na^2$. □

↳ As $n \to \infty$, $\bar{X} \to E[X]$ with prob. $\to 1$!
↳ Thus, repeating an experiment really takes you close to the expectation.

- Next, we present the strongest concentration inequality around!