Tools from Probability Theory

- Let X be a real-valued random variable with finite support $\{x_1, \ldots, x_n\}$.

- $\mathbb{E}[X] := \sum_{i=1}^{n} x_i \cdot \Pr[X = x_i]$ (sum is replaced with integral if X is continuous)

- Fact 1: Expectation is linear.
  - For all $c \in \mathbb{R}$, $\mathbb{E}[c \cdot X] = c\,\mathbb{E}[X]$
  - For all pairs of real-valued random variables (X, Y), $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$

Variance: $\sigma^2(X) := \mathbb{E}[(X - \mathbb{E}[X])^2]$.

Fact 2: Let $\mu := \mathbb{E}[X]$. Then $\sigma^2(X) = \mathbb{E}[X^2] - \mu^2$.

Proof:
$$\sigma^2(X) \overset{\text{definition}}{=} \mathbb{E}[(X-\mu)^2] = \mathbb{E}[(X^2 - 2X\mu + \mu^2)]$$
$$\overset{\text{linearity}}{=} \mathbb{E}[X^2] - 2\mu \cdot \underbrace{\mathbb{E}[X]}_{=\mu} + \mu^2$$
$$= \mathbb{E}[X^2] - \mu^2.$$

Fact 3: $\sigma^2(cX) = c^2 \sigma^2(X)$

- Two random variables X, Y are <u>independent</u> if for every $x, y \in \mathbb{R}$,
$$\Pr[X = x, Y = y] = \Pr[X = x] \cdot \Pr[Y = y].$$

Variables $X_1, X_2, \ldots, X_n$ are <u>mutually</u> independent if for every $x_1, \ldots x_n \in \mathbb{R}$,
$$\Pr[X_1 = x_1, \ldots, X_n = x_n] = \prod_{i=1}^{n} \Pr[X_i = x_i].$$

Variables $X_1, \ldots X_n$ are pairwise independent if for all pairs i, j, and all $x_i, x_j \in \mathbb{R}$, $\Pr[X_i = x_i, X_j = x_j] = \Pr[X_i = x_i] \cdot \Pr[X_j = x_j]$

Independence and linearity of variance

Fact 4: If $X_1, \ldots X_n$ are pairwise independent, then

$$\sigma^2\left(\sum_i X_i\right) = \sum_i \sigma^2(X_i).$$

Proof sketch: I will just cover the case that $\mathbb{E}[X_i] = 0$ for all i.

Note that this assumption means $\left(\mathbb{E}\left[\sum_i X_i\right]\right)^2 \overset{\text{linearity}}{=} \left(\sum_i \mathbb{E}[X_i]\right)^2 = 0^2 = 0$.

So $\sigma^2\left(\sum_i X_i\right) \overset{\text{Fact 2}}{=} \mathbb{E}\left[\left(\sum_i X_i\right)^2\right] - \left(\mathbb{E}\left[\sum_i X_i\right]\right)^2 = \mathbb{E}\left[\left(\sum_i X_i\right)^2\right]$

$$= \mathbb{E}\left[\sum_{1 \le i, j \le n} X_i \cdot X_j\right] \overset{\text{linearity of expectation}}{=} \sum_{1 \le i, j \le n} \mathbb{E}[X_i X_j]$$

if $i \neq j$, then $= \mathbb{E}[X_i] \cdot \mathbb{E}[X_j] = 0$

$$= \sum_{1 \le i \le n} \mathbb{E}[X_i \cdot X_i] = \sum_i \sigma^2(X_i) \blacksquare$$

- Averaging pairwise-independent random variables reduces variance! ~~Hence~~, If $X_1 \ldots X_n$ are pairwise independent and all have the same mean μ and variance v, then $\frac{1}{n}\sum_i X_i$ also has mean μ (by linearity of expectation) and variance $\frac{1}{n^2}\sigma^2\left(\sum_i X_i\right) \overset{\text{Fact 4}}{=} \left(\frac{1}{n^2}\right) \cdot (n \cdot v) = \frac{v}{n}$.

by Fact 3

We'll see why variance reduction is useful (Chebyshev's inequality)

- Def: An event is a set of possible outcomes of an experiment. ~~[illegible]~~ E.g. if the experiment is "draw to cards from a 52-card deck", example events are "exactly one card is red," "the jack of clubs is one of the cards selected," etc.

- Def: For two events E_1 and E_2:

$$Pr[E_2 | E_1] := \frac{Pr[E_1 \text{ and } E_2]}{Pr[E_1]}$$ (undefined if $Pr[E_1]=0$)

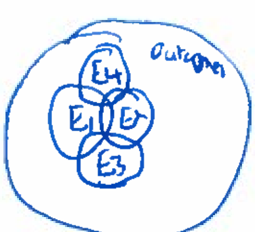
Fact 5: If two random variables X_1, X_2 are independent, then for all $x_1, x_2 \in \mathbb{R}$,

$$Pr[X_2 = x_2 \mid X_1 = x_1] = Pr[X_2 = x_2].$$

Techniques for upper bounding the probability a ~~random~~ "bad event" occurs (typically our randomized algorithms will return the right answer in expectation, and we want to bound the probability the answer is far from its expectation).

Union-Bound: For any events $E_1, E_2, \ldots, E_k$,

$$Pr[E_1 \text{ or } E_2 \ldots \text{ or } E_k] \leq Pr[E_1] + \ldots + Pr[E_k]$$



Inclusion-Exclusion principle:

$$\geq$$

$$- \sum_{i \neq j} Pr[E_i \cap E_j]$$

Markov's Inequality, Chebyshev's Inequality, Chernoff Bounds

Can apply if you have a bound on $\mathbb{E}[X]$

Can apply if you have a bound on $\mathbb{E}[X]$ and $\sigma^2(X)$

Can apply if you have a bound on <u>all</u> moments of X (such as when X is an average of independent variables)

Markov's inequality: Let $\mu = \mathbb{E}X$. If X is a non-negative random variable, then for every $t > 0$, $\Pr[X \geq t\mu] \leq \frac{1}{t}$.

Proof: $\overset{\mu =}{\mathbb{E}[X]} \overset{\text{definition}}{=} \sum_i \Pr[X=i] \cdot i$

$$\overset{X \text{ non-negative}}{\geq} \sum_{i \geq t \cdot \mu} \Pr[X=i] \cdot i$$

$$\geq t\mu \cdot \sum_{i \geq t\mu} \Pr[X=i]$$

$$= t\mu \cdot \Pr[X \geq t\mu]$$

So cancel the μ's and divide both sides by t. ∎

(Bound can be weak and doesn't work for bounding probability X is far <u>below</u> its expectation)

Chebyshev's inequality: Recall $\mu = E[X]$ and σ^2 is variance. $\sigma = \sqrt{\sigma^2}$ is the standard deviation of X.

For every $t > 0$, $\Pr[\,|X-\mu| \geq t\cdot\sigma\,] \leq \frac{1}{t^2}$

Proof: Apply Markov to $|X-\mu|$. That is:

$$\Pr[\,|X-\mu| \geq t\cdot\sigma\,] \overset{\text{square both sides of the inequality}}{=} \Pr[\,|X-\mu|^2 \geq t^2\sigma^2\,]$$

definition of variance $\rightarrow$

$$= \Pr[\,|X-\mu|^2 \geq t^2 \cdot E[|X-\mu|^2]\,]$$

Markov $\rightarrow$

$$\leq \frac{1}{t^2} \quad \blacksquare$$

Chernoff Bound: Let $X_1, \dots X_n$ be i.i.d. random variables with range $[0,1]$ and $E[X_i] = p$.

Then if $X = \sum_{i=1}^{n} X_i$ and $0 < \delta \leq 1$, and $\mu := EX = np$,

$$(*) \quad \Pr[\,|X-\mu| \geq \delta\mu\,] \leq 2\exp\left(-\frac{\mu\cdot\delta^2}{3}\right)$$

Example: Say you toss a fair coin n times, what is the probability more than $\frac{n}{2} + \sqrt{n\log n}$ of them are heads?

Let $X_i = \begin{cases} 0 & \text{if } i^{th} \text{ coin is Tails} \\ 1 & \text{if heads} \end{cases}$. Then $p = \frac{1}{2}$, $\mu = \frac{n}{2}$ and $X = \sum_i X_i$ is #of heads

$$\Pr[\,X \geq \tfrac{n}{2} + \sqrt{n\log n}\,]$$

$$\leq \Pr[\,|X-\mu| \geq \underbrace{\sqrt{n\log n}}_{=\,\delta\mu}\,] \overset{(*)}{\leq} 2\exp\left(-\frac{n}{2}\cdot\frac{\frac{4\log n}{n}}{3}\right) = 2\exp\left(-\tfrac{2}{3}\log n\right) = \frac{2}{n^{2/3}}.$$

for $\delta = 2\cdot\sqrt{\frac{\log n}{n}}$