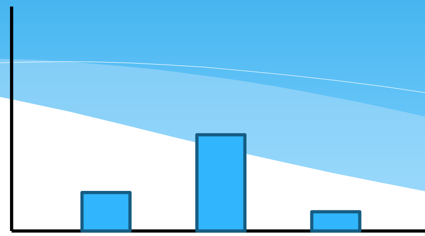
Review: Random Variables

* A ***random variable*** represents a *random quantity* that depends on the *outcome* of *a random experiment*.
* **Example:** Let N be the number of satisfied clauses of a 3CNF formula ϕ when we generate an assignment of ϕ by assigning its variables T/F independently and uniformly at random.

$$\phi = (x \vee x \vee x) \wedge (y \vee y \vee y) \wedge (\neg x \vee \neg y \vee \neg z)$$

Outcome	FFF	FFT	FTF	FTT	TFF	TFT	TTF	TTT
Sat?	NNY	NNY	NYY	NYY	YNY	YNY	YYY	YYN
N	1	1	2	2	2	2	3	2

Review: Distribution of an RV



* The ***probability*** that a random variable is *equal* to a *fixed value* is the *sum of the probabilities* of all outcomes that result in that value occurring.
* **Example:** Let N be the number of satisfied clauses of a 3CNF formula ϕ when we generate an assignment of ϕ by assigning its variables T/F independently and uniformly at random.

$$\phi = (x \vee x \vee x) \wedge (y \vee y \vee y) \wedge (\neg x \vee \neg y \vee \neg z)$$

Outcome	FFF	FFT	FTF	FTT	TFF	TFT	TTF	TTT
Sat?	NNY	NNY	NYY	NYY	YNY	YNY	YYY	YYN
N	1	1	2	2	2	2	3	2

$$\Pr[N = 1] = \frac{2}{8} \qquad \Pr[N = 2] = \frac{5}{8} \qquad \Pr[N = 3] = \frac{1}{8}$$

$$(\Pr[\text{each outcome}] = 1/8)$$

Review: Expected value of an RV

* The ***expected value*** of a random variable is the *weighted average* of its values (weight of value = its probability).
* **Example:** Let N be the number of satisfied clauses of a 3CNF formula ϕ when we generate an assignment of ϕ by assigning its variables T/F independently and uniformly at random.

$$\phi = (x \vee x \vee x) \wedge (y \vee y \vee y) \wedge (\neg x \vee \neg y \vee \neg z)$$

Outcome	FFF	FFT	FTF	FTT	TFF	TFT	TTF	TTT
Sat?	NNY	NNY	NYY	NYY	YNY	YNY	YYY	YYN
N	1	1	2	2	2	2	3	2

$$\Pr[N=1] = \frac{2}{8} \qquad \Pr[N=2] = \frac{5}{8} \qquad \Pr[N=3] = \frac{1}{8}$$

$$\mathbb{E}[N] = 1 \cdot \frac{2}{8} + 2 \cdot \frac{5}{8} + 3 \cdot \frac{1}{8} = \frac{1}{8}(1+1+2+2+2+2+3+2)$$

On average, a random assignment satisfies 15/8 clauses of ϕ.

Review: Linearity of Expectation

Observation: The number of clauses satisfied by the resulting assignment is $N = N_1 + N_2 + N_3 + \cdots + N_m$.

- ***Linearity of expectation:*** If $N = N_1 + N_2 + \cdots + N_m$, then $\mathbb{E}[N] = \mathbb{E}[N_1] + \mathbb{E}[N_2] + \cdots + \mathbb{E}[N_m]$
- **Example:** Fix a 3CNF formula $\phi = C_1 \land C_2 \land \cdots \land C_m$. Suppose we generate an assignment of ϕ by setting its variables to T/F independently, unif. at random.
  - For $1 \leq i \leq m$, let N_i be the indicator random variable for whether clause C_i is satisfied by the assignment.

$$\phi = (x \lor x \lor x) \land (y \lor y \lor y) \land (\neg x \lor \neg y \lor \neg z)$$

Outcome	FFF	FFT	FTF	FTT	TFF	TFT	TTF	TTT
$(\boldsymbol{N_1}, \boldsymbol{N_2}, \boldsymbol{N_3})$	$(0,0,1)$	$(0,0,1)$	$(0,1,1)$	$(0,1,1)$	$(1,0,1)$	$(1,0,1)$	$(1,1,1)$	$(1,1,0)$
$\boldsymbol{N}$	1	1	2	2	2	2	3	2

$\mathbb{E}[N_1] = 1/2$ $\quad$ $\mathbb{E}[N_2] = 1/2$ $\quad$ $\mathbb{E}[N_3] = 7/8$

$$\mathbb{E}[N] = \mathbb{E}[N_1] + \mathbb{E}[N_2] + \mathbb{E}[N_3] = 15/8$$

Review: Independence

- An ***event*** is a set of outcomes.
- **Example:** Let V be an RV for the value when rolling two dice.
  - $V = 5$ is an event: the set of outcomes $\{(1,4), (2,3), (3,2), (4,1)\}$
  - $\Pr[V = 5] = 4/36 = 1/9$
- **Informally:** Two events are ***independent*** if the occurrence of one does not affect the probability of the other occurring.
- **Formally:** Events A and B are ***independent*** if $\Pr[A \cap B] = \Pr[A] \cdot \Pr[B]$.
- **Alternatively:** $\Pr[A|B] \coloneqq \Pr[A \cap B]/\Pr[B]$ = probability of A given (conditioned on) B.
  - A and B are ***independent*** if $\Pr[A \cap B] = \Pr[A]$.
- A *collection of random variables* $Z_1, \dots, Z_n$ is ***independent*** if for *every* $b_1, \dots, b_n : \Pr[Z_1 = b_1, \dots, Z_n = b_n] = \prod_i \Pr[Z_i = b_i]$.

Markov's Inequality

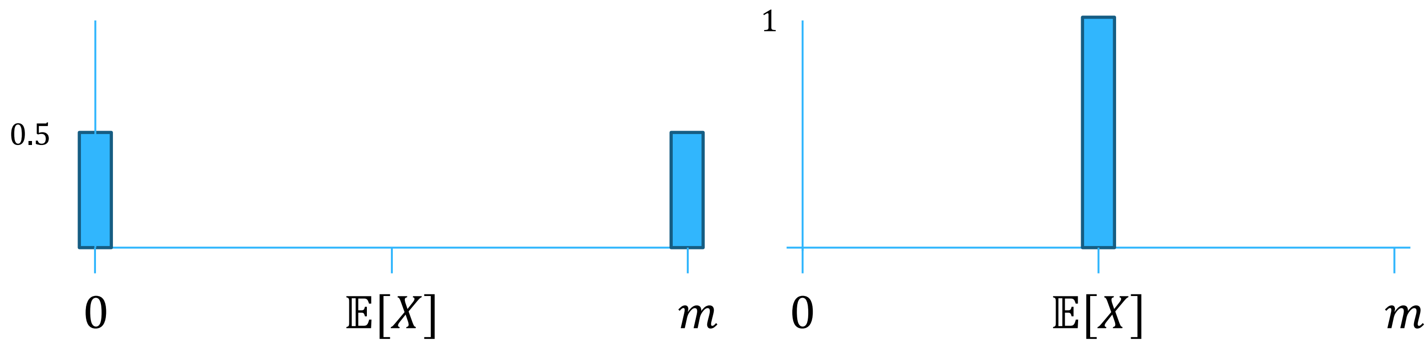
* **Example:** The average score on the midterm was 60. What's the *maximum* fraction of students that got a score of at least 90 (assuming no negative scores)?
  * 1/2? 6/9? 3/4? 99/100?
* ***Markov's Inequality:*** If X is a *non-negative* random variable and $a > 0$, then $\Pr[X \geq a] \leq \mathbb{E}[X]/a$.

Variance

- The ***variance*** of a random variable X is the average *squared-distance* of X from its mean, i.e.,

$$\mathbf{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

- The ***standard deviation*** is $\mathbf{SD}(X) = \sqrt{\mathbf{Var}(X)}$ (it's an upper bound on the average distance of X from $\mathbb{E}[X]$).



Chebyshev's Inequality

$\mathbf{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$

* **(Recall) Markov's Inequality:** For a *non-negative* RV X and $a > 0$:

$$\Pr[X \geq a] \leq \mathbb{E}[X]/a$$

* ***Chebyshev's Inequality:*** For *any* RV X and $a, b > 0$:

$$\Pr[|X - \mathbb{E}[X]| \geq a] \leq \mathbf{Var}(X)/a^2$$
$$\Pr[|X - \mathbb{E}[X]| \geq b \cdot \mathbf{SD}(X)] \leq 1/b^2$$

(proof by applying Markov to $Y = (X - \mathbb{E}[X])^2$)

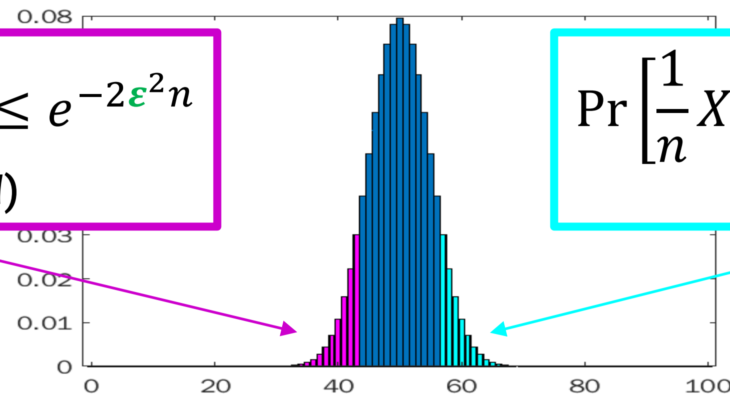
Chernoff-Hoeffding Bounds

(Often tighter than Chebyshev)

- If $X = X_1 + X_2 + \cdots + X_n$ is the sum of n ***i.i.d.*** RVs with **each $\boldsymbol{X_i} \in [\mathbf{0}, \mathbf{1}]$**, then, for any $\boldsymbol{\varepsilon} > 0$:

$$\Pr\left[\frac{1}{n}X \leq \mathbb{E}[X_i] - \boldsymbol{\varepsilon}\right] \leq e^{-2\boldsymbol{\varepsilon}^2 n}$$

(lower tail bound)



$$\Pr\left[\frac{1}{n}X \geq \mathbb{E}[X_i] + \boldsymbol{\varepsilon}\right] \leq e^{-2\boldsymbol{\varepsilon}^2 n}$$

(upper tail bound)

Note: These bounds work *for all* $n \geq 1$!