Computer Science 351

Tail Bounds

Instructor: Wayne Eberly

Department of Computer Science
University of Calgary

Lecture #21

Learning Goals

- Learn about results that can be used to bound the probability that the value of a random variable is larger than a given threshold... or smaller than a given threshold... or far from its expected value.

Restricting Attention to Finite Sample Spaces

- Results introduced, in this lecture, will concern the special case that the sample space, Ω, is ***finite***.
- Similar results can be established for infinite sample spaces — but these only hold when additional technical conditions (which always hold when the sample space is finite) are satisfied.
- Additional information will be provided about the case that the sample space is infinite, either in supplemental material or during the lecture presentation.

A Useful Result about Expectation

Lemma #1: Let Ω be a ***finite*** simple space with probability distribution $\mathrm{P} : \Omega \to \mathbb{R}$, and let $X : \Omega \to V$ be a random variable (so that $V \subseteq \mathbb{R}$). If $g : \mathbb{R} \to \mathbb{R}$ be a total function, and

$$g(V) = \{g(v) \mid v \in V\}$$

— so that "$g(X)$" is a random variable such that $g(X) : \Omega \to g(V) \subseteq \mathbb{R}$ — then

$$\begin{aligned} \mathrm{E}[g(X)] &= \sum_{\mu \in \Omega} g(X(\mu)) \times \mathrm{P}(\mu) \\ &= \sum_{v \in V} g(v) \times \mathrm{P}(X = v) \\ &= \sum_{w \in g(V)} w \times \mathrm{P}(g(X) = w). \end{aligned}$$

A Useful Result about Expectation

How This is Proved:

- The first equation follows by the definition of the expected value of a random variables.
- If the sample space, Ω, is finite, then the sums shown at each of these lines is the sum of a finite number of nonzero terms. The expression on the second and third lines are both obtained by re-ordering terms.

A Useful Result about Expectation

Lemma #2: Let Ω be a ***finite*** sample space with probability distribution $\mathrm{P} : \Omega \to \mathbb{R}$, Let $X : \Omega \to V$ be a random variable (so that $V \subseteq \mathbb{R}$), and let $a, b \in V$.

(i) If $\mathrm{P}(X = b) = 1$ then $\mathrm{E}[X] = b$.

(ii) If $\mathrm{P}(a < X \leq b) = 1$ then $a < \mathrm{E}[X] \leq b$.

(iii) If $g, h : \mathbb{R} \to \mathbb{R}$ (and these are total functions) then

$$\mathrm{E}[g(X) + h(X)] = \mathrm{E}[g(X)] + \mathrm{E}[h(X)].$$

Proof: This is left as an exercise.

A Useful Result about Expectation

Theorem #3 (Basic Inequality): Let Ω be a ***finite*** sample space with probability distribution $\mathrm{P} : \Omega \to \mathbb{R}$, let $X : \Omega \to \mathbb{R}$ be a random variable, and let $h : \mathbb{R} \to \mathbb{R}$ be a total function such that

$$h(x) \geq 0 \quad \text{for all } x \in \mathbb{R}.$$

Then, for every real number a such that $a > 0$,

$$\mathrm{P}(h(X) \geq a) \leq \frac{\mathrm{E}[h(X)]}{a}.$$

A proof of this is given in a supplemental document for this lecture.

Markov's Inequality

Corollary #4 (Markov's Inequality): Let Ω be a ***finite*** sample space with probability distribution $\mathrm{P} : \Omega \to \mathbb{R}$, and let $X : \Omega \to \mathbb{R}$ be a random variable. Then, for every ***positive*** real number a,

$$\mathrm{P}(|X| \geq a) \leq \frac{\mathrm{E}[|X|]}{a}.$$

Proof: This is a straightforward consequence of the "Basic Inequality". □

Applying Markov's Inequality in an Example

Example: Once again, consider the experiment in which n coins are tossed.

- The sample space is

  $$\Omega_n = \{(\alpha_1, \alpha_2, \ldots, \alpha_n) \mid \alpha_1, \alpha_2, \ldots, \alpha_n \in \{\mathtt{H}, \mathtt{T}\}\}$$

  — a set with size 2^n.
- Suppose that

  $$\mathrm{P} : \Omega_n \to \mathbb{R}$$

  is the uniform distribution, so that $\mathrm{P}(\vec{\alpha}) = 2^{-n}$ for all $\vec{\alpha} \in \Omega_n$.
- Let $X : \Omega_n \to \mathbb{R}$ be the random variable representing the number of heads tossed.

Applying Markov's Inequality in an Example

Then

$$X = X_1 + X_2 + \cdots + X_n$$

where, for $1 \leq i \leq n$, and $\vec{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n) \in \Omega_n$,

$$X_i(\vec{\alpha}) = \begin{cases} 1 & \text{if } \alpha_i = \mathtt{H}, \\ 0 & \text{if } \alpha_i = \mathtt{T}. \end{cases}$$

- If

  $$A_i = \{\vec{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n) \in \Omega_n \mid \alpha_i = \mathtt{H}\}$$

  then $|A_i| = 2^{n-1}$ and, for $\vec{\alpha} \in \Omega_n$,

  $$X_i(\vec{\alpha}) = \begin{cases} 1 & \text{if } \vec{\alpha} \in A_i, \\ 0 & \text{if } \vec{\alpha} \notin A_i. \end{cases}$$

- This can be used to prove that $\mathrm{E}[X_i] = \mathrm{P}(A_i) = \frac{|A_i|}{|\Omega|} = \frac{1}{2}$.

Applying Markov's Inequality in an Example

- Since $X = X_1 + X_2 + \cdots + X_n$ it now follows by ***Linearity of Expectation*** that

$$\begin{aligned} \mathrm{E}[X] &= \mathrm{E}[X_1 + X_2 + \cdots + X_n] \\ &= \sum_{i=1}^{n} \mathrm{E}[X_i] \\ &= \sum_{i=1}^{n} \frac{1}{2} \\ &= \frac{n}{2}. \end{aligned}$$

Applying Markov's Inequality in an Example

Markov's Inequality can now be used to bound that least $\frac{3n}{4}$ heads are tossed, because this is

$$\begin{aligned} \mathrm{P}(X \geq \tfrac{3n}{4}) &\leq \frac{\mathrm{E}[X]}{(3n/4)} \\ &= \frac{n/2}{3n/4} \\ &= \frac{1}{2} \times \frac{4}{3} \\ &= \frac{2}{3}. \end{aligned}$$

Variance and Standard Deviation

Definition: Let Ω be a sample space with probability distribution $\mathsf{P} : \Omega \to \mathbb{R}$, and let $X : \Omega \to \mathbb{R}$. Then the ***variance*** of X, with respect to P, is

$$\mathrm{var}(X) = \sum_{\mu \in \Omega} (X(\mu) - \mathsf{E}[X])^2 \times \mathsf{P}(\mu)$$

and the ***standard deviation*** of X, denoted $\sigma(X)$, is $\sqrt{\mathrm{var}(X)}$.

Variance and Standard Deviation

Example Suppose that we are tossing *three* coins — so that the sample space is

$$\Omega_3 = \{(\mathtt{H},\mathtt{H},\mathtt{H}),(\mathtt{H},\mathtt{H},\mathtt{T}),(\mathtt{H},\mathtt{T},\mathtt{H}),(\mathtt{H},\mathtt{T},\mathtt{T}) \\ (\mathtt{T},\mathtt{H},\mathtt{H}),(\mathtt{T},\mathtt{H},\mathtt{T}),(\mathtt{T},\mathtt{T},\mathtt{H}),(\mathtt{T},\mathtt{T},\mathtt{T})\}.$$

Suppose, again, that we we are tossing a fair coin — so that the probability distribution used is the ***probability distribution*** used is the ***uniform probability distribution*** $\mathsf{P} : \Omega_3 \to \mathbb{R}$ — that is,

$$\mathsf{P}(\vec{\alpha}) = \frac{1}{|\Omega_3|} = \frac{1}{8}$$

for every outcome $\vec{\alpha} \in \Omega_3$.

Variance and Standard Deviation

Now

- $X((\mathtt{H}, \mathtt{H}, \mathtt{H})) = 3$,
- $X((\mathtt{H}, \mathtt{H}, \mathtt{T})) = X((\mathtt{H}, \mathtt{T}, \mathtt{H})), X((\mathtt{T}, \mathtt{H}, \mathtt{H})) = 2$,
- $X((\mathtt{H}, \mathtt{T}, \mathtt{T})) = X((\mathtt{T}, \mathtt{H}, \mathtt{T})) = X((\mathtt{T}, \mathtt{T}, \mathtt{H})) = 1$, and
- $X((\mathtt{T}, \mathtt{T}, \mathtt{T})) = 0$.

As noted above (since $n = 3$ here), $\mathrm{E}[X] = \frac{3}{2}$.

Variance and Standard Deviation

$$\begin{aligned}
\mathrm{var}(X) &= \sum_{\vec{\alpha}\in\Omega_3} (X(\vec{\alpha}) - \mathrm{E}[X])^2 \times \mathrm{P}(\vec{\alpha}) \\
&= \left(X(\mathtt{H},\mathtt{H},\mathtt{H}) = \tfrac{3}{2}\right)^2 \times \mathrm{P}((\mathtt{H},\mathtt{H},\mathtt{H})) \\
&+ \left(X(\mathtt{H},\mathtt{H},\mathtt{T}) = \tfrac{3}{2}\right)^2 \times \mathrm{P}((\mathtt{H},\mathtt{H},\mathtt{T})) \\
&+ \left(X(\mathtt{H},\mathtt{T},\mathtt{H}) = \tfrac{3}{2}\right)^2 \times \mathrm{P}((\mathtt{H},\mathtt{T},\mathtt{H})) \\
&+ \left(X(\mathtt{H},\mathtt{T},\mathtt{T}) = \tfrac{3}{2}\right)^2 \times \mathrm{P}((\mathtt{H},\mathtt{T},\mathtt{T})) \\
&+ \left(X(\mathtt{T},\mathtt{H},\mathtt{H}) = \tfrac{3}{2}\right)^2 \times \mathrm{P}((\mathtt{T},\mathtt{H},\mathtt{H})) \\
&+ \left(X(\mathtt{T},\mathtt{H},\mathtt{T}) = \tfrac{3}{2}\right)^2 \times \mathrm{P}((\mathtt{T},\mathtt{H},\mathtt{T})) \\
&+ \left(X(\mathtt{T},\mathtt{T},\mathtt{H}) = \tfrac{3}{2}\right)^2 \times \mathrm{P}((\mathtt{T},\mathtt{T},\mathtt{H})) \\
&+ \left(X(\mathtt{T},\mathtt{T},\mathtt{T}) = \tfrac{3}{2}\right)^2 \times \mathrm{P}((\mathtt{T},\mathtt{T},\mathtt{T}))
\end{aligned}$$

Variance and Standard Deviation

$$
\begin{aligned}
&= \left(3 - \tfrac{3}{2}\right)^2 \times \tfrac{1}{8} + \left(2 - \tfrac{3}{2}\right)^2 \times \tfrac{1}{8} \\
&\quad + \left(2 - \tfrac{3}{2}\right)^2 \times \tfrac{1}{8} + \left(1 - \tfrac{3}{2}\right)^2 \times \tfrac{1}{8} \\
&\quad + \left(2 - \tfrac{3}{2}\right)^2 \times \tfrac{1}{8} + \left(1 - \tfrac{3}{2}\right)^2 \times \tfrac{1}{8} \\
&\quad + \left(1 - \tfrac{3}{2}\right)^2 \times \tfrac{1}{8} + \left(0 - \tfrac{3}{2}\right)^2 \times \tfrac{1}{8} \\
&= \tfrac{9}{4} \times \tfrac{1}{8} + \tfrac{1}{4} \times \tfrac{1}{8} + \tfrac{1}{4} \times \tfrac{1}{8} + \tfrac{1}{4} \times \tfrac{1}{8} \\
&\quad + \tfrac{1}{4} \times \tfrac{1}{8} + \tfrac{1}{4} \times \tfrac{1}{8} + \tfrac{1}{4} \times \tfrac{1}{8} + \tfrac{9}{4} \times \tfrac{1}{8} \\
&= \tfrac{24}{32} = \tfrac{3}{4}.
\end{aligned}
$$

It follows that the standard deviation of X, $\sigma(X)$, is $\sqrt{(3/4)} = \frac{\sqrt{3}}{2}$.

Alternative Form of Variance

Theorem #5: Let Ω be a ***finite*** sample space, let $\mathrm{P} : \Omega \to \mathbb{R}$ be a probability distribution for Ω, and let X be a random variable. Then X^2 is also a random variable, and

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

Once again, a proof of this claim is included in the supplemental document for this lecture.

Alternate Form of Variance

Continuation of Example: Suppose, again, that the sample space is Ω_3 and we are tossing a fair coin, so that $\mathrm{P} : \Omega_3 \to \mathbb{R}$ is the uniform probability distribution. Once again, let $X : \Omega_3 \to \mathbb{R}$ be the number of heads tossed, so that $X^2 : \Omega_3 \to \mathbb{R}$ is the random variable with the following values.

- $X^2((\mathtt{H},\mathtt{H},\mathtt{H})) = (X(\mathtt{H},\mathtt{H},\mathtt{H}))^2 = 3^2 = 9.$
- $X^2((\mathtt{H},\mathtt{H},\mathtt{T})) = (X(\mathtt{H},\mathtt{H},\mathtt{T}))^2 = 2^2 = 4.$
- $X^2((\mathtt{H},\mathtt{T},\mathtt{H})) = (X(\mathtt{H},\mathtt{T},\mathtt{H}))^2 = 2^2 = 4.$
- $X^2((\mathtt{H},\mathtt{T},\mathtt{T})) = (X(\mathtt{H},\mathtt{T},\mathtt{T}))^2 = 1^2 = 1.$
- $X^2((\mathtt{T},\mathtt{H},\mathtt{H})) = (X(\mathtt{T},\mathtt{H},\mathtt{H}))^2 = 2^2 = 4.$
- $X^2((\mathtt{T},\mathtt{H},\mathtt{T})) = (X(\mathtt{T},\mathtt{H},\mathtt{T}))^2 = 1^2 = 1.$
- $X^2((\mathtt{T},\mathtt{T},\mathtt{H})) = (X(\mathtt{T},\mathtt{T},\mathtt{H}))^2 = 1^2 = 1.$
- $X^2((\mathtt{T},\mathtt{T},\mathtt{T})) = (X(\mathtt{T},\mathtt{T},\mathtt{T}))^2 = 0^2 = 0.$

Alternate Form of Variance

Thus

$$
\begin{aligned}
\mathrm{E}[X^2] &= \sum_{\vec{\alpha}\in\Omega_3} X^2(\vec{\alpha}) \times \mathrm{P}(\vec{\alpha}) \\
&= X^2((\mathtt{H},\mathtt{H},\mathtt{H})) \times \mathrm{P}((\mathtt{H},\mathtt{H},\mathtt{H})) + X^2((\mathtt{H},\mathtt{H},\mathtt{T})) \times \mathrm{P}((\mathtt{H},\mathtt{H},\mathtt{T})) \\
&+ X^2((\mathtt{H},\mathtt{T},\mathtt{H})) \times \mathrm{P}((\mathtt{H},\mathtt{T},\mathtt{H})) + X^2((\mathtt{H},\mathtt{T},\mathtt{T})) \times \mathrm{P}((\mathtt{H},\mathtt{T},\mathtt{T})) \\
&+ X^2((\mathtt{T},\mathtt{H},\mathtt{H})) \times \mathrm{P}((\mathtt{T},\mathtt{H},\mathtt{H})) + X^2((\mathtt{T},\mathtt{H},\mathtt{T})) \times \mathrm{P}((\mathtt{T},\mathtt{H},\mathtt{T})) \\
&+ X^2((\mathtt{T},\mathtt{T},\mathtt{H})) \times \mathrm{P}((\mathtt{T},\mathtt{T},\mathtt{H})) + X^2((\mathtt{T},\mathtt{T},\mathtt{T})) \times \mathrm{P}((\mathtt{T},\mathtt{T},\mathtt{T})) \\
&= 9 \times \tfrac{1}{8} + 4 \times \tfrac{1}{8} + 4 \times \tfrac{1}{8} + 1 \times \tfrac{1}{8} \\
&\qquad + 4 \times \tfrac{1}{8} + 1 \times \tfrac{1}{8} + 1 \times \tfrac{1}{8} + 0 \times \tfrac{1}{8} \\
&= \tfrac{24}{8} = 3.
\end{aligned}
$$

Alternate Form of Variance

It follows, by Theorem #5, that

$$\text{var}(X) = \text{E}[X^2] - \text{E}[X]^2 = 3 - \left(\tfrac{3}{2}\right)^2 = 3 - \tfrac{9}{4} = \tfrac{3}{4}$$

— as also shown using definition of the variance of X, above.

Using Pairwise Independence

Suppose Ω is a sample space with probability distribution $\mathrm{P} : \Omega \to \mathbb{R}$, and let $X, Y : \Omega \to \mathbb{R}$ be random variables over Ω.

- ***It is not generally the case*** that $\mathrm{var}(X + Y)$ is equal to $\mathrm{var}(X) + \mathrm{var}(Y)$.

Using Pairwise Independence

Example: Consider the previous example — so that the sample space is Ω_3, $\mathrm{P} : \Omega_3 \to \mathbb{R}$ is the uniform probability distribution, and $X : \Omega_3 \to \mathbb{R}$ is the random variable representing the number of heads tossed. Let $Y = X$.

- $\mathrm{E}[(X+Y)^2] = \mathrm{E}[(2X)^2] = \mathrm{E}[4X^2] = 4\mathrm{E}[X^2] = 4 \times 3 = 12.$
- $\mathrm{E}[X+Y]^2 = \mathrm{E}[2X]^2 = (2\mathrm{E}[X])^2$
  $= 4\mathrm{E}[X]^2 = 4 \times \left(\frac{3}{2}\right)^2 = 9.$
- Thus $\mathrm{var}(X+Y) = \mathrm{E}[(X+Y)^2] - \mathrm{E}[X+Y]^2$
  $= 12 - 9 = 3.$
- $\mathrm{var}(X) + \mathrm{var}(Y) = \mathrm{var}(X) + \mathrm{var}(X) = \frac{3}{4} + \frac{3}{4} = \frac{3}{2}.$
- Thus $\mathrm{var}(X+Y) \neq \mathrm{var}(X) + \mathrm{var}(Y)$.

Using Pairwise Independence

However, something like this can be shown for a useful special case.

Theorem #6: Let Ω be a ***finite*** sample space with probability distribution $\mathrm{P} : \Omega \to \mathbb{R}$ and let $X_1, X_2, \ldots, X_n : \Omega \to \mathbb{R}$ be random variables (for some positive integer n). If $X_1, X_2, \ldots, X_n$ are ***pairwise independent*** then

$$\mathrm{var}(X_1 + X_2 + \cdots + X_n) = \mathrm{var}(X_1) + \mathrm{var}(X_2) + \cdots + \mathrm{var}(X_n).$$

Once again, the supplemental document for this lecture contains a proof of this result.

Using Pairwise Independence

Example: Once again, considering the experiment of tossing a sequence of n fair coins, where $n \geq 2$, so that the sample space is the set

$$\Omega_n = \{(\alpha_1, \alpha_2, \ldots, \alpha_n) \mid \alpha_i \in \{\mathtt{H}, \mathtt{T}\} \text{ for } 1 \leq i \leq n\}$$

with size 2^n, and $\mathrm{P} : \Omega_n \to \mathbb{R}$ is the uniform probability distribution. Once again, let X be the random variable whose value is the number of heads tossed, so that

$$X = X_1 + X_2 + \cdots + X_n$$

where $X_i : \Omega_n \to \mathbb{R}$ is the random variable such that, for $\vec{\alpha} = (\alpha_1, \alpha_2, \ldots \alpha_n) \in \Omega_n$,

$$X_i(\vec{\alpha}) = \begin{cases} 1 & \text{if } \alpha_i = \mathtt{H}, \\ 0 & \text{if } \alpha_i = \mathtt{T}. \end{cases}$$

for $1 \leq i \leq n$.

Using Pairwise Independence

Since X_i is an ***indicator random variable***, it can be shown that

$$\mathrm{E}[X_i] = \mathrm{P}(X_i = 1) = \tfrac{1}{2}$$

for $1 \leq i \leq n$, and it follows by ***Linearity of Expectation*** that

$$\begin{aligned}
\mathrm{E}[X] &= \mathrm{E}[X_1 + X_2 + \cdots + X_n] \\
&= \mathrm{E}[X_1] + \mathrm{E}[X_2] + \cdots + \mathrm{E}[X_n] \\
&= \sum_{i=1}^{n} \mathrm{E}[X_i] \\
&= \sum_{i=1}^{n} \tfrac{1}{2} \\
&= \tfrac{n}{2}.
\end{aligned}$$

Using Pairwise Independence

Since X_i is an indicator random variable, X_i^2 is the same random variable as X, so that

$$\mathrm{E}[X_i^2] = \mathrm{E}[X_i] = \tfrac{1}{2}$$

and — by Theorem #5 —

$$\mathrm{var}(X_i) = \mathrm{E}[X_i^2] - \mathrm{E}[X_i]^2 = \tfrac{1}{2} - \left(\tfrac{1}{2}\right)^2 = \tfrac{1}{4}.$$

Using Pairwise Independence

Let i and j be integers such that $1 \leq i, j \leq n$ and $i \neq j$, and let $\beta_i, \beta_j \in \{\mathtt{H}, \mathtt{T}\}$.

- The set

$$\text{“}\alpha_i = \beta_i \text{ and } \alpha_j = \beta_j\text{”} \\ = \{(\alpha_1, \alpha_2, \ldots, \alpha_n) \in \Omega_n \mid \alpha_i = \beta_i \text{ and } \alpha_j = \beta_j\}$$

  has size 2^{n-2}, since each of the values α_h such that $1 \leq h \leq n$ and $h \notin \{i, j\}$ can be chosen freely from $\{\mathtt{H}, \mathtt{T}\}$. Thus

$$\mathrm{P}(\alpha_i = \beta_i \text{ and } \alpha_j = \beta_j) = \frac{2^{n-2}}{2^n} = \frac{1}{4}.$$

Using Pairwise Independence

- Since $\text{P}(\alpha_h = \texttt{H}) = \text{P}(\alpha_h = \texttt{T}) = \frac{1}{2}$ for every integer h such that $1 \le h \le n$,

$$\text{P}(\alpha_i = \beta_i) \times \text{P}(\alpha_j = \beta_j) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

  as well.
- Thus $\text{P}(\alpha_i = \beta_i \text{ and } \alpha_j = \beta_j) = \text{P}(\alpha_i = \beta_i) \times \text{P}(\alpha_j \times \beta_j)$ for all values $\beta_i, \beta_j \in \{0, 1\}$ and (since X_i and X_j are indicator random variables) this establishes that the random variables X_i and X_j are ***independent***.
- Since this is true for all choices of integers i and j such that $1 \le i, j \le n$ and $i \ne j$, the random variables

$$X_1, X_2, \ldots, X_n$$

  are ***pairwise independent***.

Using Pairwise Independence

It now follows, by Theorem #6, above, that

$$\begin{aligned}
\text{var}(X) &= \text{var}(X_1 + X_2 + \cdots + X_n) \\
&\qquad\qquad (\text{since } X = X_1 + X_2 + \cdots + X_n) \\
&= \text{var}(X_1) + \text{var}(X_2) + \cdots + \text{var}(X_n) \\
&\qquad\qquad (\text{since } X_1, X_2, \ldots, X_n \\
&\qquad\qquad \text{are pairwise independent}) \\
&= \sum_{i=1}^{n} \text{var}(X_i) \\
&= \sum_{i=1}^{n} \tfrac{1}{4} \\
&= \tfrac{n}{4}.
\end{aligned}$$

Chebyshev's Inequality

Theorem #7: Let Ω be a ***finite*** sample space with probability distribution $P : \Omega \to \mathbb{R}$, let X be a random variable, and let $a \in \mathbb{R}$ such that $a > 0$. Then

$$P(|X| \geq a) \leq \frac{E[X^2]}{a^2}.$$

The lecture presentation will include a proof of Chebyshev's Inequality.

Chebyshev's Inequality

Example: Consider the sample space Ω_n, probability distribution $\mathrm{P} : \Omega_n \to \mathbb{R}$ and random variables $X, X_1, X_2, \ldots, X_n$ from the previous example — so that

$$X = X_1 + X_2 + \cdots + X_n.$$

- As noted above, $\mathrm{E}[X] = \frac{n}{2}$ and $\mathrm{var}(X) = \frac{n}{4}$.
- Since $\mathrm{var}(X) = \mathrm{E}[X^2] - \mathrm{E}[X]^2$, by Theorem #5, it follows that

$$\mathrm{E}[X^2] = \mathrm{var}(X) + \mathrm{E}[X]^2 = \frac{n}{4} + \left(\frac{n}{2}\right)^2 = \frac{n^2+n}{4}.$$

Chebyshev's Inequality

Once again, let us consider the probability that $X \geq \frac{3n}{4}$. Since X is a *non-negative* random variable, $X = |X|$, and it follows by Chebyshev's Inequality (with $a = \frac{3n}{4}$) that

$$\begin{aligned} \mathrm{P}(X \geq \tfrac{3n}{4}) &= \mathrm{P}(|X| \geq \tfrac{3n}{4}) \\ &\leq \frac{\mathrm{E}[X^2]}{(3n/4)^2} \\ &= \frac{(n^2+n)/4}{9n^2/16} \\ &= \tfrac{4}{9} \times \left(1 + \tfrac{1}{n}\right) \end{aligned}$$

— a considerably smaller bound than the bound that was obtained above, with Markov's Inequality, when n is large.

Cantelli's Inequality

Theorem #8: Let Σ be a ***finite*** sample space with probability distribution $\text{P} : \Omega \to \mathbb{R}$, let $X : \Omega \to \mathbb{R}$ be a random variable, and let $a \in \mathbb{R}$ such that $a > 0$. Then

$$\text{P}(X - \text{E}[X] \geq a) \leq \frac{\text{var}(X)}{a^2 + \text{var}(X)}.$$

- This result is sometimes called the "One-Sided Chebyshev's Inequality".
- A proof of this will be considered in the tutorial exercise for this topic.

Cantelli's Inequality

Once again, consider the sample space Ω_n, probability distribution $\mathrm{P} : \Omega_n \to \mathbb{R}$, and the above random variable $X : \Omega_n \to \mathbb{R}$ — so that $\mathrm{E}[X] = \frac{n}{2}$ and $\mathrm{var}(X) = \frac{n}{4}$.

$$
\begin{aligned}
\mathrm{P}(X \geq \tfrac{3n}{4}) &= \mathrm{P}(X - \mathrm{E}[X] \geq \tfrac{n}{4}) && (\text{since } \mathrm{E}[X] = \tfrac{n}{2}) \\
&\leq \frac{\mathrm{var}(X)}{(n/4)^2 + \mathrm{var}(X)} && (\text{by Canelli's Inequality}) \\
&= \frac{n/4}{(n/4)^2 + (n/4)} && (\text{since } \mathrm{var}(X) = \tfrac{n}{4}) \\
&= \frac{4}{n+4}
\end{aligned}
$$

— a bound with approaches 0 as n approaches $+\infty$, and which is a *much* better bound than can be obtained using either Markov's Inequality or Chebyshev's Inequality, when n is large.

The Chernoff Bound

The ***Chernoff Bound*** cannot be used in all the cases when the above can — but can provide significantly better results, when it is applicable. A sketch of a proof is given in the supplemental document

Theorem #9: Let Ω be a finite sample space with probability distribution $\mathrm{P} : \Omega \to \mathbb{R}$. Suppose that $X_1, X_2, \dots, X_n$ are mutually independent random variables such that $X_i : \Omega \to \{0, 1\}$ for $1 \le i \le n$, and suppose that $\mathrm{P}(X_i = 1) = p$ for every integer i such that $1 \le i \le n$, for a real number p such that $0 \le p \le 1$. Let $X = X_1 + X_2 + \cdots + X_n$. Then, for every real number θ such that $0 \le \theta \le 1$,

$$\mathrm{P}(X \ge (1 + \theta)pn) \le e^{-\frac{\theta^2}{3}pn}.$$

The Chernoff Bound

Example: Once again, consider the experiment of tossing a sequence of n fair coins, where $n \geq 2$, so that the sample space is the set

$$\Omega_n = \{(\alpha_1, \alpha_2, \ldots, \alpha_n) \mid \alpha_i \in \{\mathtt{H}, \mathtt{T}\} \text{ for } 1 \leq i \leq n\}$$

with size 2^n, and $\mathrm{P} : \Omega_n \to \mathbb{R}$ is the uniform probability distribution. Once again, let X be the random variable whose value is the number of heads tossed, so that

$$X = X_1 + X_2 + \cdots + X_n$$

where $X_i : \Omega_n \to \mathbb{R}$ is that random variable such that, for $\vec{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n) \in \Omega_n$,

$$X_i(\vec{\alpha}) = \begin{cases} 1 & \text{if } \alpha_i = \mathtt{H}, \\ 0 & \text{if } \alpha_i = \mathtt{T}. \end{cases}$$

The Chernoff Bound

- Since the ***uniform probability distribution*** is being used, it can be shown that the random variables $X_1, X_2, \ldots, X_n$ are ***mutually independent***.
- $X_i : \Omega_n \to \{0, 1\}$ for every integer i such that $1 \leq i \leq n$.
- $\text{P}(X_i = 1) = p$ for every integer i such that $1 \leq i \leq n$, when $p = \frac{1}{2}$.
- $X = X_1 + X_2 + \cdots + X_n$.
- Thus the conditions, included in the statement of the ***Chernoff Bound***, are satisfied.

The Chernoff Bound

- Let $\theta = \frac{1}{2}$. Then

$$(1+\theta)pn = \frac{3}{2} \times \frac{1}{2} \times n = \frac{3n}{4},$$

so that it now follows, by the Chernoff Bound, that

$$\begin{aligned} P(X \geq \tfrac{3n}{4}) &= P(X \geq (1+\theta)pn) \qquad \text{(for } \theta = p = \tfrac{1}{2}\text{)} \\ &\leq e^{-\frac{\theta^2}{3}pn} \\ &= e^{-\frac{n}{12}}. \end{aligned}$$

The Chernoff Bound

- Like the bound obtained using Cantelli's Inequality, this bound approaches 0 as n approaches $+\infty$.
- While the results given before this might give smaller bounds when n is quite small, this result gives smaller (and, therefore, better) bounds when n is large — because the bound, given here, approaches 0 *much* more quickly than the bound obtained using Cantelli's Inequality.

What About Countably Infinite Sample Spaces?

- Results like the ones given in these notes can also be established for countably infinite sample spaces ***when additional technical conditions are satisfied***.
- This will be discussed in *another* supplemental document for this topic.