

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.

A Useful Result about Expectation

Let Ω be a sample space, and let $X : \Omega \rightarrow \mathbb{R}$.

- For every total function $g : \mathbb{R} \rightarrow \mathbb{R}$, g and X can be used to another random variable, “ $g(X)$ ” by **composition**: For $\sigma \in \Omega$, if $X(\sigma) = z$ (for a real number z), then the value of the random variable “ $g(X)$ ”, at σ , is $g(z)$.

A Useful Result about Expectation

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

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

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

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 $P : \Omega \rightarrow \mathbb{R}$, Let $X : \Omega \rightarrow V$ be a random variable (so that $V \subseteq \mathbb{R}$), and let $a, b \in V$.

- (i) If $P(X = b) = 1$ then $E[X] = b$.
- (ii) If $P(a < X \leq b) = 1$ then $a < E[X] \leq b$.
- (iii) If $g, h : \mathbb{R} \rightarrow \mathbb{R}$ (and these are total functions) then

$$E[g(X) + h(X)] = E[g(X)] + E[h(X)].$$

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 $P : \Omega \rightarrow \mathbb{R}$, let $X : \Omega \rightarrow \mathbb{R}$ be a random variable, and let $h : \mathbb{R} \rightarrow \mathbb{R}$ 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 $a > 0$,

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

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 $P : \Omega \rightarrow \mathbb{R}$, and let $X : \Omega \rightarrow \mathbb{R}$ be a random variable. Then, for every **positive** real number a ,

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

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

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

— a set with size 2^n .

- Suppose that

$$\mathbf{P} : \Omega_n \rightarrow \mathbb{R}$$

is the uniform distribution, so that $\mathbf{P}(\vec{\alpha}) = 2^{-n}$ for all $\vec{\alpha} \in \Omega_n$.

- Let $X : \Omega_n \rightarrow \mathbb{R}$ 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 $\vec{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_n) \in \Omega_n$,

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

- If

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

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 $E[X_i] = P(A_i) = \frac{|A_i|}{|\Omega|} = \frac{1}{2}$.

Applying Markov's Inequality in an Example

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

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

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

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

Variance and Standard Deviation

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

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

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

Variance and Standard Deviation

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

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

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

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

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

Variance and Standard Deviation

Now

- $X((H, H, H)) = 3$,
- $X((H, H, T)) = X((H, T, H)), X((T, H, H)) = 2$,
- $X((H, T, T)) = X((T, H, T)) = X((T, T, H)) = 1$, and
- $X((T, T, T)) = 0$.

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

Variance and Standard Deviation

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

Variance and Standard Deviation

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

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 $P : \Omega \rightarrow \mathbb{R}$ be a probability distribution for Ω , and let X be a random variable. Then X^2 is also a random variable, and

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

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 $P : \Omega_3 \rightarrow \mathbb{R}$ is the uniform probability distribution. Once again, let $X : \Omega_3 \rightarrow \mathbb{R}$ be the number of heads tossed, so that $X^2 : \Omega_3 \rightarrow \mathbb{R}$ is the random variable with the following values.

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

Alternate Form of Variance

Thus

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

Alternate Form of Variance

It follows, by Theorem #5, that

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

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

Using Pairwise Independence

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

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

Using Pairwise Independence

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

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

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 $P : \Omega \rightarrow \mathbb{R}$ and let $X_1, X_2, \dots, X_n : \Omega \rightarrow \mathbb{R}$ be random variables (for some positive integer n). If $X_1, X_2, \dots, X_n$ are **pairwise independent** then

$$\text{var}(X_1 + X_2 + \dots + X_n) = \text{var}(X_1) + \text{var}(X_2) + \dots + \text{var}(X_n).$$

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

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

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

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

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

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

for $1 \leq i \leq n$.

Using Pairwise Independence

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

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

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

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

Using Pairwise Independence

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

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

and — by Theorem #5 —

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

Using Pairwise Independence

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

- The set

$$\begin{aligned} & \text{"}\alpha_i = \beta_i \text{ and } \alpha_j = \beta_j\text{"} \\ &= \{(\alpha_1, \alpha_2, \dots, \alpha_n) \in \Omega_n \mid \alpha_i = \beta_i \text{ and } \alpha_j = \beta_j\} \end{aligned}$$

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 $\{H, T\}$. Thus

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

Using Pairwise Independence

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

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

as well.

- Thus $P(\alpha_i = \beta_i \text{ and } \alpha_j = \beta_j) = P(\alpha_i = \beta_i) \times P(\alpha_j = \beta_j)$ 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 $1 \leq i, j \leq n$ and $i \neq j$, the random variables

$$X_1, X_2, \dots, X_n$$

are ***pairwise independent***.

Using Pairwise Independence

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

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

Chebyshev's Inequality

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

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

and

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

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

Chebyshev's Inequality

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

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

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

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

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

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

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

Chebyshev's Inequality

Note, as well that, since $E[X] = \frac{n}{2}$, $X \geq \frac{3n}{4}$ *only if* $X - E[X] \geq \frac{n}{4}$ — and that this is true *only if* $|X - E[X]| \geq \frac{n}{4}$.

Thus — applying the *second* inequality given above,

$$\begin{aligned} P(X \geq \frac{3n}{4}) &\leq P(|X - E[X]| \geq \frac{n}{4}) \\ &\leq \frac{\text{var}(X)}{(n/4)^2} \\ &= \frac{(n/4)}{(n/4)^2} \\ &= \frac{4}{n} \end{aligned}$$

— a bound which approaches 0 as n approaches $+\infty$, so that it is a *much* better bound than those obtained, for this example, using either Markov's Inequality or the first form of Cheybshev's Inequality, when n is large.

Cantelli's Inequality

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

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

- 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 $P : \Omega_n \rightarrow \mathbb{R}$, and the above random variable $X : \Omega_n \rightarrow \mathbb{R}$ — so that $E[X] = \frac{n}{2}$ and $\text{var}(X) = \frac{n}{4}$.

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

— a bound which also approaches 0 as n approaches $+\infty$, and which is *slightly* better than the bound obtained using the second form of Chebyshev's Inequality.

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.

Let e be **Euler's number** — so that e is approximately (but not exactly) 2.71828.

Theorem #9: Let Ω be a finite sample space with probability distribution $P : \Omega \rightarrow \mathbb{R}$. Suppose that $X_1, X_2, \dots, X_n$ are **mutually independent random variables** such that $X_i : \Omega \rightarrow \{0, 1\}$ for $1 \leq i \leq n$, and suppose that $P(X_i = 1) = p_i$ for every integer i such that $1 \leq i \leq n$ — so that $p_1, p_2, \dots, p_n$ are real numbers such that $0 \leq p_i \leq 1$ for $1 \leq i \leq n$. Let $X = X_1 + X_2 + \dots + X_n$, and let $\mu = E[X] = p_1 + p_2 + \dots + p_n > 0$.

The Chernoff Bound

(a) For every real number δ such that $\delta > 0$,

$$\mathbf{P}(X \geq (1 + \delta)\mu) \leq \left(\frac{e^\delta}{(1 + \delta)^{1+\delta}} \right)^\mu.$$

(b) For every real number δ such that $0 < \delta < 1$,

$$\mathbf{P}(X \leq (1 - \delta)\mu) \leq \left(\frac{e^{-\delta}}{((1 - \delta)^{1-\delta})} \right)^\mu,$$

The Chernoff Bound

Corollary #10: Let Ω be a finite sample space with probability distribution $P : \Omega \rightarrow \mathbb{R}$. Suppose that $X_1, X_2, \dots, X_n$ are mutually independent random variables such that $X_i : \Omega \rightarrow \{0, 1\}$ for $1 \leq i \leq n$, and suppose that $P(X_i = 1) = p_i$, for $1 \leq i \leq n$ — so that $p_1, p_2, \dots, p_n$ are real numbers such that $0 \leq p_i \leq 1$ for $1 \leq i \leq n$. Let $X = X_1 + X_2 + \dots + X_n$ and let $\mu = E[X] = p_1 + p_2 + \dots + p_n > 0$.

- (a) If $0 < \delta \leq 1$ then $P(X \geq (1 + \delta)\mu) \leq e^{-\frac{\delta^2 \mu}{3}}$.
- (b) For all positive real numbers γ and δ , if $\delta \geq \gamma$ then $P(X \geq (1 + \delta)\mu) \leq (e^{\gamma - (1 + \gamma) \ln(1 + \gamma)})^\mu$.
- (c) If $0 < \delta < 1$ then $P(X \leq (1 - \delta)\mu) \leq e^{-\frac{\delta^2 \mu}{2}}$.

The Chernoff Bound

Corollary #11: Let Ω be a finite sample space with probability distribution $P : \Omega \rightarrow \mathbb{R}$. Suppose that $X_1, X_2, \dots, X_n$ are mutually independent random variables such that $X_i : \Omega \rightarrow \{0, 1\}$ for $1 \leq i \leq n$, and suppose that $P(X_i = 1) = p$, for $1 \leq i \leq n$ — for a real number p such that $0 < p < 1$. Let $X = X_1 + X_2 + \dots + X_n$.

(a) If $\delta > 0$ then

$$P(X \geq (1 + \delta)pn) \leq \left(\frac{e^\delta}{(1 + \delta)^{1+\delta}} \right)^{pn}.$$

(b) If $0 < \delta < 1$ then

$$P(X \leq (1 - \delta)pn) \leq \left(\frac{e^{-\delta}}{(1 - \delta)^{1-\delta}} \right)^{pn}.$$

The Chernoff Bound

- (c) If $0 \leq \delta \leq 1$ then $P(X \geq (1 + \delta)pn) \leq e^{-\frac{\delta^2 pn}{3}}$.
- (d) For all positive real numbers γ and δ , if $\delta \geq \gamma$ then $P(X \geq (1 + \delta)pn) \leq (e^{\gamma - (1 + \gamma) \ln(1 + \gamma)})^{pn}$.
- (e) If $0 < \delta < 1$ then $P(X \leq (1 - \delta)pn) \leq e^{-\frac{\delta^2 pn}{2}}$.

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

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

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

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

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

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

The Chernoff Bound

- Since the ***uniform probability distribution*** is being used, it can be shown that the random variables $X_1, X_2, \dots, X_n$ are ***mutually independent***.
- $X_i : \Omega_n \rightarrow \{0, 1\}$ for every integer i such that $1 \leq i \leq n$.
- $P(X_i = 1) = p$ for every integer i such that $1 \leq i \leq n$, when $p = \frac{1}{2}$.
- $X = X_1 + X_2 + \dots + X_n$.
- Thus the conditions, included in the version of the Chernoff Bound given as “Corollary #11” are satisfied.

The Chernoff Bound

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

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

so that it now follows, by part (a) of Corollary #11, that

$$P(X \geq \frac{3n}{4}) \leq e^{(\frac{1}{2} - \frac{3}{2} \ln(\frac{3}{2}))n/2} < e^{-0.054n}.$$

It follows by part (c) of Corollary #11 that

$$P(X \geq \frac{3n}{4}) \leq e^{-\frac{n}{24}} < e^{-0.04n}.$$

The Chernoff Bound

- Like the bounds obtained using the second form for Chebyshev's Inequality, and using Cantelli's Inequality, these bounds approach 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 bounds, given here, approach 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.