

# Learning Goals

- Define variance and standard deviation
- State Chebyshev inequality and Chernoff inequality
- Compare the conditions and strengths of Markov, Chebyshev and Chernoff inequalities
- Understand the main idea and steps in the proofs of these bounds
- Intuition for the bounds given by the simplified forms of the Chernoff bound

# Chebyshev Inequality

## Definition

The *variance* of a random variable  $X$  is

$\text{Var}[X] := \mathbf{E}[(X - \mathbf{E}[X])^2] = \mathbf{E}[X^2] - (\mathbf{E}[X])^2$ . Its square root,  $\sqrt{\text{Var}[X]}$ , is the *standard deviation* of  $X$ , and is often denoted as  $\sigma$ .

# Chebyshev Inequality

## Definition

The *variance* of a random variable  $X$  is

$\text{Var}[X] := \mathbf{E}[(X - \mathbf{E}[X])^2] = \mathbf{E}[X^2] - (\mathbf{E}[X])^2$ . Its square root,  $\sqrt{\text{Var}[X]}$ , is the *standard deviation* of  $X$ , and is often denoted as  $\sigma$ .

## Theorem (Chebyshev Inequality)

For any  $\alpha > 0$ ,  $\Pr[|X - \mathbf{E}[X]| > \alpha\sigma] \leq \frac{1}{\alpha^2}$ .

# Chebyshev Inequality

## Definition

The *variance* of a random variable  $X$  is

$\text{Var}[X] := \mathbf{E}[(X - \mathbf{E}[X])^2] = \mathbf{E}[X^2] - (\mathbf{E}[X])^2$ . Its square root,  $\sqrt{\text{Var}[X]}$ , is the *standard deviation* of  $X$ , and is often denoted as  $\sigma$ .

## Theorem (Chebyshev Inequality)

For any  $\alpha > 0$ ,  $\Pr[|X - \mathbf{E}[X]| > \alpha\sigma] \leq \frac{1}{\alpha^2}$ .

## Proof.

Apply Markov inequality to the random variable  $(X - \mathbf{E}[X])^2$ :

$$\Pr[|X - \mathbf{E}[X]| \geq \alpha\sigma] = \Pr[(X - \mathbf{E}[X])^2 \geq \alpha^2 \text{Var}[X]] \leq \frac{1}{\alpha^2}.$$



## Question

For what distributions is Chebyshev inequality tight?

## Question

For what distributions is Chebyshev inequality tight?

The Markov inequality in the proof is tight.

## Question

For what distributions is Chebyshev inequality tight?

The Markov inequality in the proof is tight.

$\Rightarrow$  A distribution where  $|X - \mathbf{E}[X]|$  takes two values: 0 and  $\alpha\sigma$

## Question

For what distributions is Chebyshev inequality tight?

The Markov inequality in the proof is tight.

⇒ A distribution where  $|X - \mathbf{E}[X]|$  takes two values: 0 and  $\alpha\sigma$

⇒  $X$  takes three values:  $\mathbf{E}[X]$ ,  $\mathbf{E}[X] + \alpha\sigma$  and  $\mathbf{E}[X] - \alpha\sigma$ , with equal probability on the latter two values.

# Useful Facts for Independent Random Variables

## Lemma

*If  $X$  and  $Y$  are independent random variables, then  $\mathbf{E}[XY] = \mathbf{E}[X] \cdot \mathbf{E}[Y]$ , and  $\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y]$ .*

# Useful Facts for Independent Random Variables

## Lemma

If  $X$  and  $Y$  are independent random variables, then  $\mathbf{E}[XY] = \mathbf{E}[X] \cdot \mathbf{E}[Y]$ , and  $\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y]$ .

## Proof.

$$\mathbf{E}[XY] = \sum_{x,y} (xy) \mathbf{Pr}[X = x, Y = y]$$

# Useful Facts for Independent Random Variables

## Lemma

If  $X$  and  $Y$  are independent random variables, then  $\mathbf{E}[XY] = \mathbf{E}[X] \cdot \mathbf{E}[Y]$ , and  $\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y]$ .

## Proof.

$$\begin{aligned}\mathbf{E}[XY] &= \sum_{x,y} (xy) \Pr[X = x, Y = y] \\ &= \sum_{x,y} (xy) \Pr[X = x] \Pr[Y = y]\end{aligned}$$

# Useful Facts for Independent Random Variables

## Lemma

If  $X$  and  $Y$  are independent random variables, then  $\mathbf{E}[XY] = \mathbf{E}[X] \cdot \mathbf{E}[Y]$ , and  $\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y]$ .

## Proof.

$$\begin{aligned}\mathbf{E}[XY] &= \sum_{x,y} (xy) \Pr[X = x, Y = y] \\ &= \sum_{x,y} (xy) \Pr[X = x] \Pr[Y = y] \\ &= \sum_x x \Pr[X = x] \sum_y y \Pr[Y = y] = \mathbf{E}[X] \cdot \mathbf{E}[Y].\end{aligned}$$

# Useful Facts for Independent Random Variables

## Lemma

If  $X$  and  $Y$  are independent random variables, then  $\mathbf{E}[XY] = \mathbf{E}[X] \cdot \mathbf{E}[Y]$ , and  $\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y]$ .

## Proof.

$$\begin{aligned}\mathbf{E}[XY] &= \sum_{x,y} (xy) \Pr[X = x, Y = y] \\ &= \sum_{x,y} (xy) \Pr[X = x] \Pr[Y = y] \\ &= \sum_x x \Pr[X = x] \sum_y y \Pr[Y = y] = \mathbf{E}[X] \cdot \mathbf{E}[Y].\end{aligned}$$

# Useful Facts for Independent Random Variables

## Lemma

If  $X$  and  $Y$  are independent random variables, then  $\mathbf{E}[XY] = \mathbf{E}[X] \cdot \mathbf{E}[Y]$ , and  $\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y]$ .

## Proof.

$$\begin{aligned}\mathbf{E}[XY] &= \sum_{x,y} (xy) \Pr[X = x, Y = y] \\ &= \sum_{x,y} (xy) \Pr[X = x] \Pr[Y = y] \\ &= \sum_x x \Pr[X = x] \sum_y y \Pr[Y = y] = \mathbf{E}[X] \cdot \mathbf{E}[Y].\end{aligned}$$

Without independence,  $\text{Var}[X + Y]$  in general is not equal to  $\text{Var}[X] + \text{Var}[Y]$ . □

# Application of Chebyshev Inequality

## Theorem (Weak Law of Large Numbers)

Let  $X_1, X_2, \dots$  be *independently, identically distributed* (i.i.d.) random variables, and each has finite variance. For each  $n \geq 1$ , let  $\bar{X}_n$  be  $\frac{1}{n} \sum_{i=1}^n X_i$ . Then for any  $\delta > 0$ ,  $\lim_{n \rightarrow \infty} \Pr[|\bar{X}_n - \mathbf{E}[\bar{X}_n]| > \delta] = 0$ .

# Application of Chebyshev Inequality

## Theorem (Weak Law of Large Numbers)

Let  $X_1, X_2, \dots$  be *independently, identically distributed* (i.i.d.) random variables, and each has finite variance. For each  $n \geq 1$ , let  $\bar{X}_n$  be  $\frac{1}{n} \sum_{i=1}^n X_i$ . Then for any  $\delta > 0$ ,  $\lim_{n \rightarrow \infty} \Pr[|\bar{X}_n - \mathbf{E}[\bar{X}_n]| > \delta] = 0$ .

## Proof.

By independence,  $\text{Var}[\bar{X}_n] = \frac{1}{n^2} \sum_{i=1}^n \text{Var}[X_i] = \frac{1}{n} \text{Var}[X_1]$ .

# Application of Chebyshev Inequality

## Theorem (Weak Law of Large Numbers)

Let  $X_1, X_2, \dots$  be *independently, identically distributed* (i.i.d.) random variables, and each has finite variance. For each  $n \geq 1$ , let  $\bar{X}_n$  be  $\frac{1}{n} \sum_{i=1}^n X_i$ . Then for any  $\delta > 0$ ,  $\lim_{n \rightarrow \infty} \Pr[|\bar{X}_n - \mathbf{E}[\bar{X}_n]| > \delta] = 0$ .

## Proof.

By independence,  $\text{Var}[\bar{X}_n] = \frac{1}{n^2} \sum_{i=1}^n \text{Var}[X_i] = \frac{1}{n} \text{Var}[X_1]$ .

By Chebyshev inequality,  $\Pr[|\bar{X}_n - \mathbf{E}[\bar{X}_n]| > \delta] \leq \frac{\text{Var}[X_1]}{n\delta^2}$ .

# Application of Chebyshev Inequality

## Theorem (Weak Law of Large Numbers)

Let  $X_1, X_2, \dots$  be *independently, identically distributed* (i.i.d.) random variables, and each has finite variance. For each  $n \geq 1$ , let  $\bar{X}_n$  be  $\frac{1}{n} \sum_{i=1}^n X_i$ . Then for any  $\delta > 0$ ,  $\lim_{n \rightarrow \infty} \Pr[|\bar{X}_n - \mathbf{E}[\bar{X}_n]| > \delta] = 0$ .

## Proof.

By independence,  $\text{Var}[\bar{X}_n] = \frac{1}{n^2} \sum_{i=1}^n \text{Var}[X_i] = \frac{1}{n} \text{Var}[X_1]$ .

By Chebyshev inequality,  $\Pr[|\bar{X}_n - \mathbf{E}[\bar{X}_n]| > \delta] \leq \frac{\text{Var}[X_1]}{n\delta^2}$ .

The right hand side goes to 0 as  $n$  goes to infinity. □

# Landscape so far

- Markov inequality:  $\Pr[X \geq \alpha \mathbf{E}[X]] \leq \frac{1}{\alpha}$ .
  - $X$  must be nonnegative.

# Landscape so far

- Markov inequality:  $\Pr[X \geq \alpha \mathbf{E}[X]] \leq \frac{1}{\alpha}$ .
  - $X$  must be nonnegative.
- Chebyshev inequality:  $\Pr[|X - \mathbf{E}[X]| > \alpha\sigma] \leq \frac{1}{\alpha^2}$ .
  - In both, the tail bound shrinks polynomially with the distance from  $\mathbf{E}[X]$ .

# Landscape so far

- Markov inequality:  $\Pr[X \geq \alpha \mathbf{E}[X]] \leq \frac{1}{\alpha}$ .
  - $X$  must be nonnegative.
- Chebyshev inequality:  $\Pr[|X - \mathbf{E}[X]| > \alpha\sigma] \leq \frac{1}{\alpha^2}$ .
  - In both, the tail bound shrinks polynomially with the distance from  $\mathbf{E}[X]$ .
  - We often need tighter bounds.

# Landscape so far

- Markov inequality:  $\Pr[X \geq \alpha \mathbf{E}[X]] \leq \frac{1}{\alpha}$ .
  - $X$  must be nonnegative.
- Chebyshev inequality:  $\Pr[|X - \mathbf{E}[X]| > \alpha\sigma] \leq \frac{1}{\alpha^2}$ .
  - In both, the tail bound shrinks polynomially with the distance from  $\mathbf{E}[X]$ .
  - We often need tighter bounds.
- The proof of Chebyshev inequality amplifies the deviation  $|X - \mathbf{E}[X]|$  by taking its square.

# Landscape so far

- Markov inequality:  $\Pr[X \geq \alpha \mathbf{E}[X]] \leq \frac{1}{\alpha}$ .
  - $X$  must be nonnegative.
- Chebyshev inequality:  $\Pr[|X - \mathbf{E}[X]| > \alpha\sigma] \leq \frac{1}{\alpha^2}$ .
  - In both, the tail bound shrinks polynomially with the distance from  $\mathbf{E}[X]$ .
  - We often need tighter bounds.
- The proof of Chebyshev inequality amplifies the deviation  $|X - \mathbf{E}[X]|$  by taking its square.
  - Take a nonnegative, fast growing function  $f(\cdot)$  and apply Markov inequality to  $f(X)$

# Landscape so far

- Markov inequality:  $\Pr[X \geq \alpha \mathbf{E}[X]] \leq \frac{1}{\alpha}$ .
  - $X$  must be nonnegative.
- Chebyshev inequality:  $\Pr[|X - \mathbf{E}[X]| > \alpha\sigma] \leq \frac{1}{\alpha^2}$ .
  - In both, the tail bound shrinks polynomially with the distance from  $\mathbf{E}[X]$ .
  - We often need tighter bounds.
- The proof of Chebyshev inequality amplifies the deviation  $|X - \mathbf{E}[X]|$  by taking its square.
  - Take a nonnegative, fast growing function  $f(\cdot)$  and apply Markov inequality to  $f(X)$
  - Historically, Chebyshev  $\rightarrow$  Markov  $\rightarrow$  Kolmogorov

# Landscape so far

- Markov inequality:  $\Pr[X \geq \alpha \mathbf{E}[X]] \leq \frac{1}{\alpha}$ .
  - $X$  must be nonnegative.
- Chebyshev inequality:  $\Pr[|X - \mathbf{E}[X]| > \alpha\sigma] \leq \frac{1}{\alpha^2}$ .
  - In both, the tail bound shrinks polynomially with the distance from  $\mathbf{E}[X]$ .
  - We often need tighter bounds.
- The proof of Chebyshev inequality amplifies the deviation  $|X - \mathbf{E}[X]|$  by taking its square.
  - Take a nonnegative, fast growing function  $f(\cdot)$  and apply Markov inequality to  $f(X)$
  - Historically, Chebyshev  $\rightarrow$  Markov  $\rightarrow$  Kolmogorov
  - Bernstein and Chernoff exploited the idea by looking at  $f(x) = e^{\lambda x}$ .

## Chernoff Bound: I.I.D. Case

Let  $X_1, \dots, X_n$  be i.i.d. Bernoulli variables, such that  $\Pr[X_i = 1] = p$  and  $\Pr[X_i = 0] = q := 1 - p$  for each  $i$ . Define  $X = \sum_{i=1}^n X_i$ .

### Theorem (Chernoff Bound)

For any  $t > 0$ ,

$$\Pr[X > (p + t)n] \leq \exp \left\{ \left( -(p + t) \ln \frac{p + t}{p} - (q - t) \ln \frac{q - t}{q} \right) n \right\}.$$

## Chernoff Bound: I.I.D. Case

Let  $X_1, \dots, X_n$  be i.i.d. Bernoulli variables, such that  $\Pr[X_i = 1] = p$  and  $\Pr[X_i = 0] = q := 1 - p$  for each  $i$ . Define  $X = \sum_{i=1}^n X_i$ .

### Theorem (Chernoff Bound)

For any  $t > 0$ ,

$$\Pr[X > (p + t)n] \leq \exp \left\{ \left( -(p + t) \ln \frac{p + t}{p} - (q - t) \ln \frac{q - t}{q} \right) n \right\}.$$

### Proof.

For any  $\lambda > 0$ , by Markov inequality we have

$$\Pr[X > (p + t)n] = \Pr[e^{\lambda X} \geq e^{\lambda(p+t)n}] \leq \frac{\mathbf{E}[e^{\lambda X}]}{e^{\lambda(p+t)n}}.$$

## Chernoff Bound: I.I.D. Case

Let  $X_1, \dots, X_n$  be i.i.d. Bernoulli variables, such that  $\Pr[X_i = 1] = p$  and  $\Pr[X_i = 0] = q := 1 - p$  for each  $i$ . Define  $X = \sum_{i=1}^n X_i$ .

### Theorem (Chernoff Bound)

For any  $t > 0$ ,

$$\Pr[X > (p + t)n] \leq \exp \left\{ \left( -(p + t) \ln \frac{p + t}{p} - (q - t) \ln \frac{q - t}{q} \right) n \right\}.$$

### Proof.

For any  $\lambda > 0$ , by Markov inequality we have

$$\Pr[X > (p + t)n] = \Pr[e^{\lambda X} \geq e^{\lambda(p+t)n}] \leq \frac{\mathbf{E}[e^{\lambda X}]}{e^{\lambda(p+t)n}}.$$

By independence, we have  $\mathbf{E}[e^{\lambda X}] = \mathbf{E}[e^{\sum_i \lambda X_i}] = \prod_i \mathbf{E}[e^{\lambda X_i}] = (pe^{\lambda} + q)^n$ .

# Proof of Chernoff Bound (Cont.)

## Proof of Chernoff Bound (Cont.)

So far we have  $\Pr[X \geq (p + t)n] \leq \left( \frac{pe^\lambda + q}{e^{\lambda(p+t)}} \right)^n$ .

# Proof of Chernoff Bound (Cont.)

## Proof of Chernoff Bound (Cont.)

So far we have  $\Pr[X \geq (p+t)n] \leq \left(\frac{pe^\lambda + q}{e^{\lambda(p+t)}}\right)^n$ .

Take a  $\lambda$  that minimizes the RHS: let  $e^\lambda$  be  $\frac{q}{p} \cdot \frac{p+t}{q-t}$ , we get the bound in the theorem statement. □

# Proof of Chernoff Bound (Cont.)

## Proof of Chernoff Bound (Cont.)

So far we have  $\Pr[X \geq (p + t)n] \leq \left(\frac{pe^\lambda + q}{e^{\lambda(p+t)}}\right)^n$ .

Take a  $\lambda$  that minimizes the RHS: let  $e^\lambda$  be  $\frac{q}{p} \cdot \frac{p+t}{q-t}$ , we get the bound in the theorem statement. □

Remarks:

- Note the bound's exponential decay with  $n$ .

# Proof of Chernoff Bound (Cont.)

## Proof of Chernoff Bound (Cont.)

So far we have  $\Pr[X \geq (p+t)n] \leq \left(\frac{pe^\lambda + q}{e^{\lambda(p+t)}}\right)^n$ .

Take a  $\lambda$  that minimizes the RHS: let  $e^\lambda$  be  $\frac{q}{p} \cdot \frac{p+t}{q-t}$ , we get the bound in the theorem statement. □

Remarks:

- Note the bound's exponential decay with  $n$ .
- The negative of the exponent,  $(p+t) \log \frac{p+t}{p} + (q-t) \log \frac{q-t}{q}$ , is the *relative entropy*, a.k.a. *KL-divergence*, from the distribution  $(p, q)$  to the distribution  $(p+t, q-t)$  on the two-point space  $\{1, 0\}$ .

# Proof of Chernoff Bound (Cont.)

## Proof of Chernoff Bound (Cont.)

So far we have  $\Pr[X \geq (p + t)n] \leq \left( \frac{pe^\lambda + q}{e^{\lambda(p+t)}} \right)^n$ .

Take a  $\lambda$  that minimizes the RHS: let  $e^\lambda$  be  $\frac{q}{p} \cdot \frac{p+t}{q-t}$ , we get the bound in the theorem statement. □

Remarks:

- Note the bound's exponential decay with  $n$ .
- The negative of the exponent,  $(p + t) \log \frac{p+t}{p} + (q - t) \log \frac{q-t}{q}$ , is the *relative entropy*, a.k.a. *KL-divergence*, from the distribution  $(p, q)$  to the distribution  $(p + t, q - t)$  on the two-point space  $\{1, 0\}$ .

The same proof yields the same bound for  $\Pr[X \leq (p - t)n]$ .

# Chernoff Bound: Heterogeneous Bernoulli Variables

What if  $X_1, \dots, X_n$  are independent Bernoulli variables but not identically distributed?

# Chernoff Bound: Heterogeneous Bernoulli Variables

What if  $X_1, \dots, X_n$  are independent Bernoulli variables but not identically distributed?

Suppose  $\Pr[X_i = 1] = p_i$ , and  $\Pr[X_i = 0] = q_i$  for each  $i$ , and let  $p = \frac{1}{n} \sum_i p_i$  and  $q = 1 - p$ .

# Chernoff Bound: Heterogeneous Bernoulli Variables

What if  $X_1, \dots, X_n$  are independent Bernoulli variables but not identically distributed?

Suppose  $\Pr[X_i = 1] = p_i$ , and  $\Pr[X_i = 0] = q_i$  for each  $i$ , and let  $p = \frac{1}{n} \sum_i p_i$  and  $q = 1 - p$ .

In the proof where we calculated  $\mathbf{E}[e^{\lambda X}]$ , we now have

$\mathbf{E}[e^{\lambda X}] = \prod_i (p_i e^{\lambda} + q_i) \leq (pe^{\lambda} + q)^n$  by the AM-GM inequality.

# Chernoff Bound: Heterogeneous Bernoulli Variables

What if  $X_1, \dots, X_n$  are independent Bernoulli variables but not identically distributed?

Suppose  $\Pr[X_i = 1] = p_i$ , and  $\Pr[X_i = 0] = q_i$  for each  $i$ , and let  $p = \frac{1}{n} \sum_i p_i$  and  $q = 1 - p$ .

In the proof where we calculated  $\mathbf{E}[e^{\lambda X}]$ , we now have

$\mathbf{E}[e^{\lambda X}] = \prod_i (p_i e^{\lambda} + q_i) \leq (pe^{\lambda} + q)^n$  by the AM-GM inequality.

The rest of the proof follows exactly as before.

# Chernoff Bound: Heterogeneous Bernoulli Variables

What if  $X_1, \dots, X_n$  are independent Bernoulli variables but not identically distributed?

Suppose  $\Pr[X_i = 1] = p_i$ , and  $\Pr[X_i = 0] = q_i$  for each  $i$ , and let  $p = \frac{1}{n} \sum_i p_i$  and  $q = 1 - p$ .

In the proof where we calculated  $\mathbf{E}[e^{\lambda X}]$ , we now have

$\mathbf{E}[e^{\lambda X}] = \prod_i (p_i e^{\lambda} + q_i) \leq (p e^{\lambda} + q)^n$  by the AM-GM inequality.

The rest of the proof follows exactly as before.

## Theorem

For any  $t > 0$ ,

$$\Pr[X > (p + t)n] \leq \exp \left\{ \left( -(p + t) \ln \frac{p + t}{p} - (q - t) \ln \frac{q - t}{q} \right) n \right\}.$$

# Chernoff Bound: The Hoeffding Extension

What if  $X_1, \dots, X_n$  always take values from  $[0, 1]$  but not necessarily  $\{0, 1\}$ ?  
Suppose  $\mathbf{E}[X_i] = p_i$ ,  $q_i = 1 - p_i$  for each  $i$ , and let  $p = \frac{1}{n} \sum_i p_i$  and  $q = 1 - p$ .

# Chernoff Bound: The Hoeffding Extension

What if  $X_1, \dots, X_n$  always take values from  $[0, 1]$  but not necessarily  $\{0, 1\}$ ?  
Suppose  $\mathbf{E}[X_i] = p_i$ ,  $q_i = 1 - p_i$  for each  $i$ , and let  $p = \frac{1}{n} \sum_i p_i$  and  $q = 1 - p$ .  
Observation: on  $[0, 1]$ ,  $e^{\lambda x} \leq \alpha x + \beta$  for  $\alpha = e^\lambda - 1$  and  $\beta = 1$ .

# Chernoff Bound: The Hoeffding Extension

What if  $X_1, \dots, X_n$  always take values from  $[0, 1]$  but not necessarily  $\{0, 1\}$ ?

Suppose  $\mathbf{E}[X_i] = p_i$ ,  $q_i = 1 - p_i$  for each  $i$ , and let  $p = \frac{1}{n} \sum_i p_i$  and  $q = 1 - p$ .

Observation: on  $[0, 1]$ ,  $e^{\lambda x} \leq \alpha x + \beta$  for  $\alpha = e^\lambda - 1$  and  $\beta = 1$ .

So  $\mathbf{E}[e^{\lambda X_i}] \leq \mathbf{E}[(e^\lambda - 1)X_i + 1] = p_i e^\lambda + q_i$ .

# Chernoff Bound: The Hoeffding Extension

What if  $X_1, \dots, X_n$  always take values from  $[0, 1]$  but not necessarily  $\{0, 1\}$ ?

Suppose  $\mathbf{E}[X_i] = p_i$ ,  $q_i = 1 - p_i$  for each  $i$ , and let  $p = \frac{1}{n} \sum_i p_i$  and  $q = 1 - p$ .

Observation: on  $[0, 1]$ ,  $e^{\lambda x} \leq \alpha x + \beta$  for  $\alpha = e^\lambda - 1$  and  $\beta = 1$ .

So  $\mathbf{E}[e^{\lambda X_i}] \leq \mathbf{E}[(e^\lambda - 1)X_i + 1] = p_i e^\lambda + q_i$ .

Therefore  $\mathbf{E}[e^{\lambda X}] = \prod_i \mathbf{E}[e^{\lambda X_i}] \leq \prod_i (p_i e^\lambda + q_i) \leq (p e^\lambda + q)^n$ .

# Chernoff Bound: The Hoeffding Extension

What if  $X_1, \dots, X_n$  always take values from  $[0, 1]$  but not necessarily  $\{0, 1\}$ ?

Suppose  $\mathbf{E}[X_i] = p_i$ ,  $q_i = 1 - p_i$  for each  $i$ , and let  $p = \frac{1}{n} \sum_i p_i$  and  $q = 1 - p$ .

Observation: on  $[0, 1]$ ,  $e^{\lambda x} \leq \alpha x + \beta$  for  $\alpha = e^\lambda - 1$  and  $\beta = 1$ .

So  $\mathbf{E}[e^{\lambda X_i}] \leq \mathbf{E}[(e^\lambda - 1)X_i + 1] = p_i e^\lambda + q_i$ .

Therefore  $\mathbf{E}[e^{\lambda X}] = \prod_i \mathbf{E}[e^{\lambda X_i}] \leq \prod_i (p_i e^\lambda + q_i) \leq (p e^\lambda + q)^n$ .

The rest of the proof remains the same.

# Chernoff Bound: The Hoeffding Extension

What if  $X_1, \dots, X_n$  always take values from  $[0, 1]$  but not necessarily  $\{0, 1\}$ ?

Suppose  $\mathbf{E}[X_i] = p_i$ ,  $q_i = 1 - p_i$  for each  $i$ , and let  $p = \frac{1}{n} \sum_i p_i$  and  $q = 1 - p$ .

Observation: on  $[0, 1]$ ,  $e^{\lambda x} \leq \alpha x + \beta$  for  $\alpha = e^\lambda - 1$  and  $\beta = 1$ .

So  $\mathbf{E}[e^{\lambda X_i}] \leq \mathbf{E}[(e^\lambda - 1)X_i + 1] = p_i e^\lambda + q_i$ .

Therefore  $\mathbf{E}[e^{\lambda X}] = \prod_i \mathbf{E}[e^{\lambda X_i}] \leq \prod_i (p_i e^\lambda + q_i) \leq (p e^\lambda + q)^n$ .

The rest of the proof remains the same.

## Theorem

For any  $t > 0$ ,

$$\Pr [X > (p + t)n] \leq \exp \left\{ \left( -(p + t) \ln \frac{p + t}{p} - (q - t) \ln \frac{q - t}{q} \right) n \right\}.$$

# Useful Forms of Chernoff Bound

## Corollary

Let  $X_1, \dots, X_n$  be independently distributed on  $[0, 1]$  and  $X = \sum_i X_i$ .

- For all  $t > 0$ ,

$$\Pr [X > \mathbf{E} [X] + t], \Pr [X < \mathbf{E} [X] - t] \leq e^{-2t^2/n};$$

- For any  $\epsilon < 1$ ,

$$\Pr [X > (1 + \epsilon) \mathbf{E} [X]] \leq \left( \frac{e^\epsilon}{(1 + \epsilon)^{1+\epsilon}} \right)^{\mathbf{E} [X]} \leq \exp \left( -\frac{\epsilon^2}{3} \mathbf{E} [X] \right);$$

$$\Pr [X < (1 - \epsilon) \mathbf{E} [X]] \leq \left( \frac{e^{-\epsilon}}{(1 - \epsilon)^{1-\epsilon}} \right)^{\mathbf{E} [X]} \leq \exp \left( -\frac{\epsilon^2}{2} \mathbf{E} [X] \right).$$

## Useful Forms of Chernoff Bound (Cont.)

## Corollary ((Cont.))

- For any  $\epsilon > 1$ ,

$$\Pr [X > (1 + \epsilon) \mathbf{E} [X]] \leq \left( \frac{e^\epsilon}{(1 + \epsilon)^{1+\epsilon}} \right)^{\mathbf{E}[X]} \leq \exp \left( -\frac{\epsilon}{3} \mathbf{E} [X] \right);$$

- If  $t > 2e \mathbf{E}[X]$ , then

$$\Pr [X > t] \leq 2^{-t}.$$

# Proof Sketch

## Proof Sketch.

- Let  $f(t)$  be  $(p+t) \ln \frac{p+t}{p} + (q-t) \ln \frac{q-t}{q}$ . Show  $f(t) \geq 2t^2$  by showing  $f(0) = f'(0) = 0$  and  $f''(t) \geq 4$  for all  $0 \leq t \leq q$  followed by Taylor's theorem with remainder.

# Proof Sketch

## Proof Sketch.

- Let  $f(t)$  be  $(p+t) \ln \frac{p+t}{p} + (q-t) \ln \frac{q-t}{q}$ . Show  $f(t) \geq 2t^2$  by showing  $f(0) = f'(0) = 0$  and  $f''(t) \geq 4$  for all  $0 \leq t \leq q$  followed by Taylor's theorem with remainder.
- Let  $g(x)$  be  $f(px)$ , then  $g'(0) = pf'(px)$ , and so  $g(0) = g'(0) = 0$ . Show  $g'(1) > p \ln 2 > \frac{2}{3}p$ . Deduce that for  $x \in (0, 1)$ ,  $g(x) \geq px^2/3$ .

# Proof Sketch

## Proof Sketch.

- Let  $f(t)$  be  $(p+t) \ln \frac{p+t}{p} + (q-t) \ln \frac{q-t}{q}$ . Show  $f(t) \geq 2t^2$  by showing  $f(0) = f'(0) = 0$  and  $f''(t) \geq 4$  for all  $0 \leq t \leq q$  followed by Taylor's theorem with remainder.
- Let  $g(x)$  be  $f(px)$ , then  $g'(0) = pf'(px)$ , and so  $g(0) = g'(0) = 0$ . Show  $g'(1) > p \ln 2 > \frac{2}{3}p$ . Deduce that for  $x \in (0, 1)$ ,  $g(x) \geq px^2/3$ .
- Set  $h(x) := g(-x)$ . Then  $h'(x) = -g'(-x)$ , and  $h(0) = h'(0) = 0$ . Show then  $h''(x) \leq p$  for  $x \in (0, 1)$ . Deduce that  $h(x) \geq px^2/2$ .

# Proof Sketch

## Proof Sketch.

- Let  $f(t)$  be  $(p+t) \ln \frac{p+t}{p} + (q-t) \ln \frac{q-t}{q}$ . Show  $f(t) \geq 2t^2$  by showing  $f(0) = f'(0) = 0$  and  $f''(t) \geq 4$  for all  $0 \leq t \leq q$  followed by Taylor's theorem with remainder.
- Let  $g(x)$  be  $f(px)$ , then  $g'(0) = pf'(px)$ , and so  $g(0) = g'(0) = 0$ . Show  $g'(1) > p \ln 2 > \frac{2}{3}p$ . Deduce that for  $x \in (0, 1)$ ,  $g(x) \geq px^2/3$ .
- Set  $h(x) := g(-x)$ . Then  $h'(x) = -g'(-x)$ , and  $h(0) = h'(0) = 0$ . Show then  $h''(x) \leq p$  for  $x \in (0, 1)$ . Deduce that  $h(x) \geq px^2/2$ .

See assigned reading for more details. Or take them as exercises. □