

# Probability and Counting, Appendix C

Geoffrey Matthews

April 24, 2018

# Binomial Coefficients

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \binom{n}{n-k}$$

$$(x + y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k}$$

$$2^n = \sum_{k=0}^n \binom{n}{k}$$

$$(n + a)^b = O(n^b)$$

$$(n + a)^b = O(n^b)$$

$$\begin{aligned}(n + a)^b &= \sum_{k=0}^b \binom{b}{k} n^k a^{b-k} \\ &\leq \sum_{k=0}^b \binom{b}{k} n^b a^{b-k} \\ &= n^b \sum_{k=0}^b \binom{b}{k} a^{b-k} \\ &= n^b \sum_{k=0}^b \binom{b}{k} 1^k a^{b-k} \\ &= n^b (1 + a)^b \\ &= O(n^b)\end{aligned}$$

# Binomial Bounds

$$\binom{n}{k} \geq \left(\frac{n}{k}\right)^k$$

$$\binom{n}{k} \leq \left(\frac{en}{k}\right)^k$$

$$\binom{n}{k} \leq \frac{n^n}{k^k (n-k)^{n-k}}$$

$$\binom{n}{\lambda n} \leq 2^{nH(\lambda)}$$

Stirling:  $k! \geq (k/e)^k$

by induction

$$H(\lambda) = -\lambda \lg \lambda - (1 - \lambda) \lg(1 - \lambda)$$

# Sample Space

- ▶ Set of all possible things that can happen.
- ▶ Each thing that can happen is an **elementary event**.
- ▶ Examples:
  - ▶ Flip a coin twice and observe which side is up:  
 $\{HH, HT, TH, TT\}$
  - ▶ Flip a coin twice and count the heads:  $\{0, 1, 2\}$
  - ▶ Throw a coin down the stairs and see what step it lands on:  
 $\{1, 2, 3, \dots, n\}$ , where  $n$  is the number of steps.
  - ▶ Deal two cards:  $\{\{A\spadesuit, 5\clubsuit\}, \{10\heartsuit, K\clubsuit\}, \{A\spadesuit, 3\heartsuit\}, \dots\}$
  - ▶ See who wins the election:  $\{Clinton, Sanders, Trump, Cruz \dots\}$

# Events

- ▶ A subset of the sample space,  $S$ .
- ▶ Examples:
  - ▶  $\{HT, TH\} \subseteq \{HH, HT, TH, TT\}$
  - ▶  $\{\{A\spadesuit, A\clubsuit\}, \{A\heartsuit, A\diamondsuit\}\} \subseteq$   
 $\{\{A\diamondsuit, 5\clubsuit\}, \{10\heartsuit, K\diamondsuit\}, \{A\spadesuit, 3\heartsuit\}, \dots\}$
  - ▶ The **certain event**:  $S$ .
  - ▶ The **null event**:  $\emptyset$

# A probability distribution on a sample space $S$

$\Pr\{\}$  is a mapping from events to real numbers such that:

1.  $\Pr\{A\} \geq 0$
2.  $\Pr\{S\} = 1$
3.  $\Pr\{A \cup B\} = \Pr\{A\} + \Pr\{B\}$  whenever  $A \cap B = \emptyset$

► Theorem:

$$\begin{aligned}\Pr\{A \cup B\} &= \Pr\{A\} + \Pr\{B\} - \Pr\{A \cap B\} \\ &\leq \Pr\{A\} + \Pr\{B\}\end{aligned}$$



# Discrete probability distribution

- ▶ If  $S$  is finite or countably infinite.

$$\Pr\{A\} = \sum_{s \in A} \Pr\{s\}$$

- ▶ If  $S$  is finite and each elementary event has the same probability, we have **uniform probability distribution**.

$$\Pr\{s\} = 1/|S|$$

# Continuous uniform distribution

- ▶ Each real number between  $a$  and  $b$  is equally likely.
- ▶ Not all subsets have probabilities.
- ▶ Just use intervals, and countable unions of intervals.

$$\Pr\{[c, d]\} = \frac{d - c}{b - a}$$

# Conditional probability

$$\Pr\{A \mid B\} = \frac{\Pr\{A \cap B\}}{\Pr\{B\}}$$

- ▶ Probability as if  $B$  were the sample space.
- ▶ Can condition a variable on events:

$$\begin{aligned}\Pr\{B\} &= \Pr\{B \cap A\} + \Pr\{B \cap \bar{A}\} \\ &= \Pr\{A\} \Pr\{B \mid A\} + \Pr\{\bar{A}\} \Pr\{B \mid \bar{A}\}\end{aligned}$$

# Conditioning example

- ▶ Suppose we flip a coin.
- ▶ If it's heads, we roll a 6-sided die; if it's tails an 8-sided die.
- ▶ What's the probability of getting a 6?

# Conditioning example

- ▶ Suppose we flip a coin.
- ▶ If it's heads, we roll a 6-sided die; if it's tails an 8-sided die.
- ▶ What's the probability of getting a 6?
- ▶

$$\begin{aligned}\Pr\{B\} &= \Pr\{B \cap A\} + \Pr\{B \cap \bar{A}\} \\ &= \Pr\{A\} \Pr\{B \mid A\} + \Pr\{\bar{A}\} \Pr\{B \mid \bar{A}\} \\ \Pr\{6\} &= \Pr\{6 \cap H\} + \Pr\{6 \cap \bar{H}\} \\ &= \Pr\{H\} \Pr\{6 \mid H\} + \Pr\{\bar{H}\} \Pr\{6 \mid \bar{H}\} \\ &= (1/2)(1/6) + (1/2)(1/8) \\ &= 7/48\end{aligned}$$

# Independence

$$\Pr\{A \cap B\} = \Pr\{A\} \Pr\{B\}$$

This implies

$$\Pr\{A \mid B\} = \Pr\{A\}$$

## Joint Probability Distribution, Marginal Distributions

		X				$\sum_x f(x, y)$
		2	3	4	5	
Y	A	0.30	0.00	0.10	0.21	0.61
	B	0.03	0.04	0.10	0.08	0.25
	C	0.00	0.03	0.10	0.01	0.14
$\sum_y f(x, y)$		0.33	0.07	0.30	0.30	

# Joint Probability Distribution, Independent Distributions

		X				$\sum_x f(x, y)$
		2	3	4	5	
Y	A	0.18	0.18	0.18	0.06	0.6
	B	0.09	0.09	0.09	0.03	0.3
	C	0.03	0.03	0.03	0.01	0.1
$\sum_y f(x, y)$		0.3	0.3	0.3	0.1	



# Bayes's theorem

$$\Pr\{A \mid B\} = \frac{\Pr\{A\} \Pr\{B \mid A\}}{\Pr\{B\}}$$

- ▶ This follows easily from

$$\Pr\{A \cap B\} = \Pr\{A\} \Pr\{B \mid A\} = \Pr\{B\} \Pr\{A \mid B\}$$

- ▶ We can combine this with a conditioning of  $B$  getting

$$\Pr\{A \mid B\} = \frac{\Pr\{A\} \Pr\{B \mid A\}}{\Pr\{A\} \Pr\{B \mid A\} + \Pr\{\bar{A}\} \Pr\{B \mid \bar{A}\}}$$

## Bayes's theorem example

- ▶ We have a fair coin and a biased coin with  $\Pr\{H\} = 2/3$ . We choose a coin at random and flip it twice. It comes up heads both times. What is the probability we chose the biased coin?
- ▶ Let  $A$  be the event of choosing a biased coin, and let  $B$  be the event of coming up heads twice in a row.

$$\begin{aligned}\Pr\{A \mid B\} &= \frac{\Pr\{A\} \Pr\{B \mid A\}}{\Pr\{A\} \Pr\{B \mid A\} + \Pr\{\bar{A}\} \Pr\{B \mid \bar{A}\}} \\&= \frac{(1/2)(4/9)}{(1/2)(4/9) + (1/2)(1/4)} \\&= \frac{(2/9)}{(2/9) + (1/8)} \\&= \frac{(2/9)}{(25/72)} \\&= 16/25\end{aligned}$$

# Discrete random variables

- ▶ Given finite or countable  $S$ , a **random variable**  $X$  is a function from  $S$  to the real numbers.
- ▶ The event  $X = x$  is

$$\{s \in S : X(s) = x\}$$

- ▶ Therefore

$$\Pr\{X = x\} = \sum_{s \in S: X(s)=x} \Pr\{s\}$$

- ▶ The **probability density function**:

$$f(x) = \Pr\{X = x\}$$

- ▶ With two random variables  $X$  and  $Y$ , the **joint probability density**:

$$f(x, y) = \Pr\{X = x \text{ and } Y = y\}$$

- ▶ What does independence imply about the joint distribution?

# Expected value

- ▶ The **expected value** or **expectation** or **mean**:

$$E[X] = \sum_x x \cdot \Pr\{X = x\}$$

- ▶ Denoted by  $\mu_X$  or  $\mu$
- ▶ **Linearity of expectation:**

$$E[X + Y] = E[X] + E[Y]$$

Holds even if  $X$  and  $Y$  are not independent.

# Variance

$$\begin{aligned}\text{Var}[X] &= E[X - E[X]]^2 \\ &= E[X^2] - E^2[X]\end{aligned}$$

- ▶ The latter can be computed in one pass, but is not as numerically stable as the first.
- ▶ If  $X$  and  $Y$  are independent:

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y]$$

- ▶ The **standard deviation** is the square root of the variance.
  - ▶ It is denoted  $\sigma_X$  or  $\sigma$ .
- ▶ The variance is denoted  $\sigma^2$ .

# Bernoulli trials

- ▶ Repeatedly flip a biased coin with probability  $p$  of heads.
- ▶ Each flip is independent of the others.
- ▶ Instead of heads and tails we say **success** and **failure**.

# Geometric distribution

- ▶ With Bernoulli trials with probability of success  $p$ , what is the probability we try  $k$  times to get the first success?

$$\Pr\{X = k\} = q^{k-1}p$$

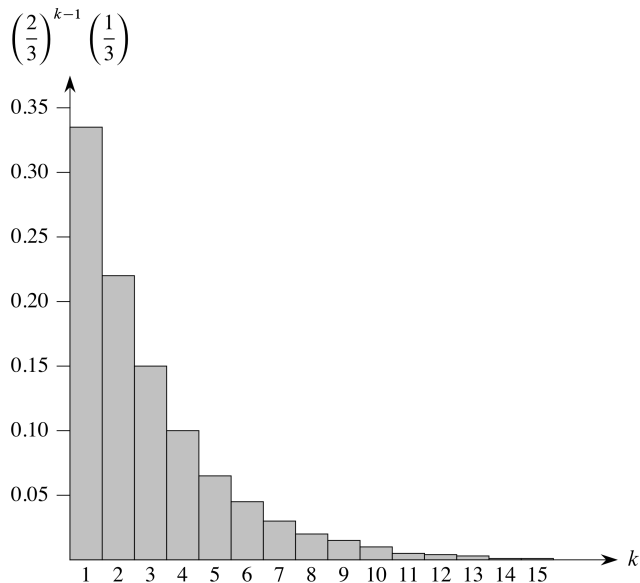
- ▶ Expectation:

$$\begin{aligned} E[X] &= \sum_{k=1}^{\infty} kq^{k-1}p \\ &= 1/p \end{aligned}$$

- ▶ Variance:

$$\text{Var}[X] = q/p^3$$

# Geometric distribution





# Binomial distribution

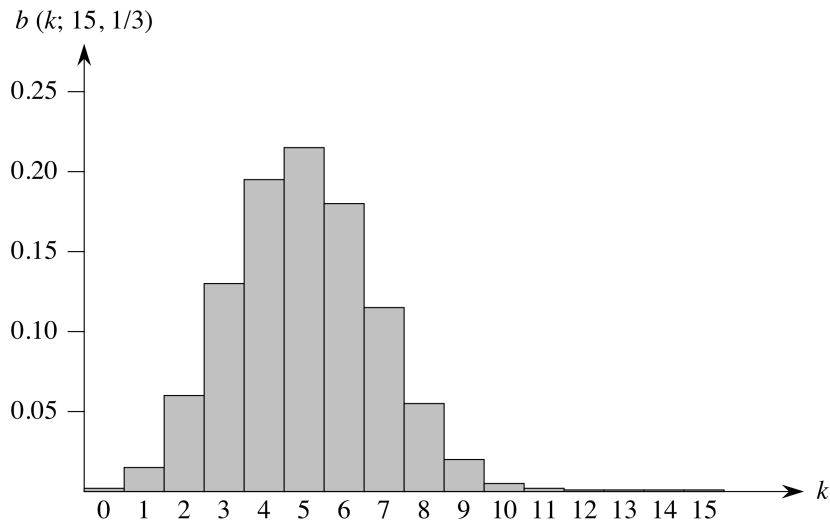
- ▶ Probability of  $k$  successes occur in  $n$  Bernoulli trials with probability  $p$ :

$$\Pr\{X = k\} = \binom{n}{k} p^k (1 - p)^{n-k}$$

- ▶ We define a family of distributions:

$$b(k : n, p) = \binom{n}{k} p^k (1 - p)^{n-k}$$

# Binomial distribution



# Binomial distribution expectation

$$\Pr\{X = k\} = \binom{n}{k} p^k (1 - p)^{n-k}$$

► Expectation:

$$\begin{aligned} E[X] &= \sum_{k=0}^n k \cdot \Pr\{X = k\} \\ &= \sum_{k=0}^n k \binom{n}{k} p^k q^{n-k} \\ &= ? \end{aligned}$$

# Binomial distribution expectation

$$\Pr\{X = k\} = \binom{n}{k} p^k (1 - p)^{n-k}$$

- ▶ Let  $X_i$  be an **indicator random variable** for the  $i$ th trial.
  - ▶  $X_i$  is 1 if the  $i$ th trial is a success, 0 otherwise.
- ▶ Easier math:

$$\begin{aligned} E[X] &= E\left[\sum_{i=1}^n X_i\right] \\ &= \sum_{i=1}^n E[X_i] && \text{linearity of expectation} \\ &= \sum_{i=1}^n p = np \end{aligned}$$

## Binomial distribution variance

$$\Pr\{X = k\} = \binom{n}{k} p^k (1-p)^{n-k}$$

$$\begin{aligned}\text{Var}[X_i] &= E[X_i^2] - E^2[X_i] \\ &= E[X_i] - E^2[X_i] \\ &= p - p^2 = pq\end{aligned}$$

$$\begin{aligned}\text{Var}[X] &= \text{Var}\left[\sum_{i=1}^n X_i\right] \\ &= \sum_{i=1}^n \text{Var}[X_i] \\ &= \sum_{i=1}^n pq = npq\end{aligned}$$

# Binomial tails

$$\Pr\{X \geq k\} = \sum_{i=k}^n b(i : n, p)$$

$$\leq \binom{n}{k} p^k$$

$$\Pr\{X \leq k\} = \sum_{i=0}^k b(i : n, p)$$

$$\leq \binom{n}{k} (1-p)^{n-k}$$