Notes on Probability

Geoffrey Matthews

April 24, 2018

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: $\{1, 2, 3, \ldots, n\}$, where n is the number of steps.
  - Deal two cards: $\{\{A\diamondsuit, 5\clubsuit\}, \{10\heartsuit, K\clubsuit\}, \{A\spadesuit, 3\heartsuit\}, \ldots\}$
  - See who wins the election: $\{Clinton, Sanders, Trump, Cruz \ldots\}$

Events

- A subset of the sample space, S.
- Examples:
  - $\{HT, TH\} \subseteq \{HH, HT, TH, TT\}$
  - $\{\{A\spadesuit, A\clubsuit\}, \{A\heartsuit, A\diamondsuit\}\} \subseteq \{\{A\diamondsuit, 5\clubsuit\}, \{10\heartsuit, K\diamondsuit\}, \{A\spadesuit, 3\heartsuit\}, \ldots\}$
  - 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:

$$\begin{aligned}\Pr\{B\} &= \Pr\{B \cap A\} + \Pr\left\{B \cap \overline{A}\right\} \\ &= \Pr\{A\} \Pr\{B \mid A\} + \Pr\left\{\overline{A}\right\} \Pr\left\{B \mid \overline{A}\right\}\end{aligned}$$

Independence

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

This implies

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

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

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

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.

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

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**:
$$f(x, y) = \Pr\{X = x \text{ and } Y = x\}$$
- 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 μ_X or μ
- **Linearity of expectation:**

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

  Holds even if X and Y are not independent.

Variance

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

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

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

- The **standard deviation** is the square root of the variance.
  - It is denoted σ_X or σ.
- The variance is denoted σ^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?

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

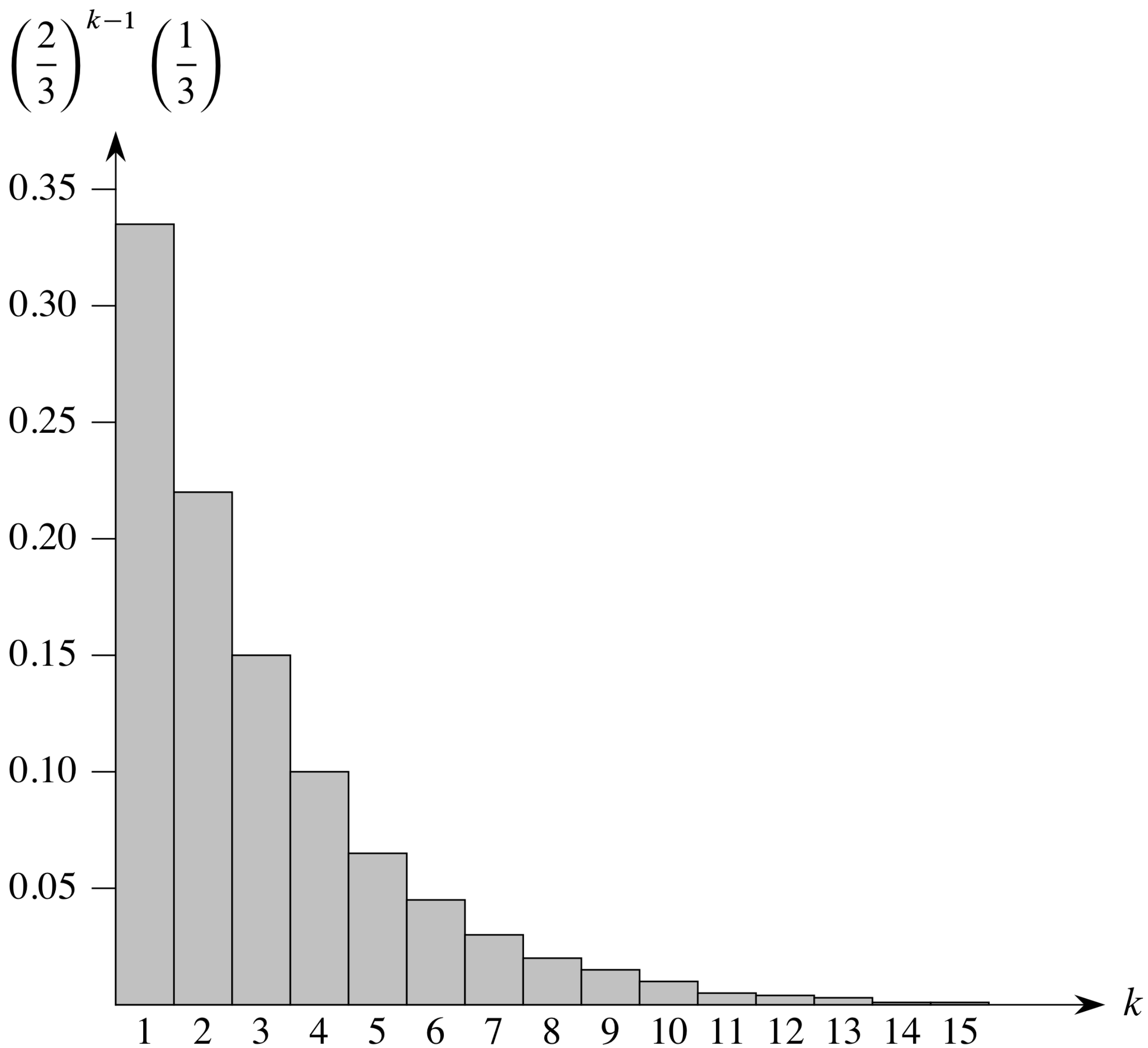
- Expectation:

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

- Variance:

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

Geometric distribution



Binomial distribution

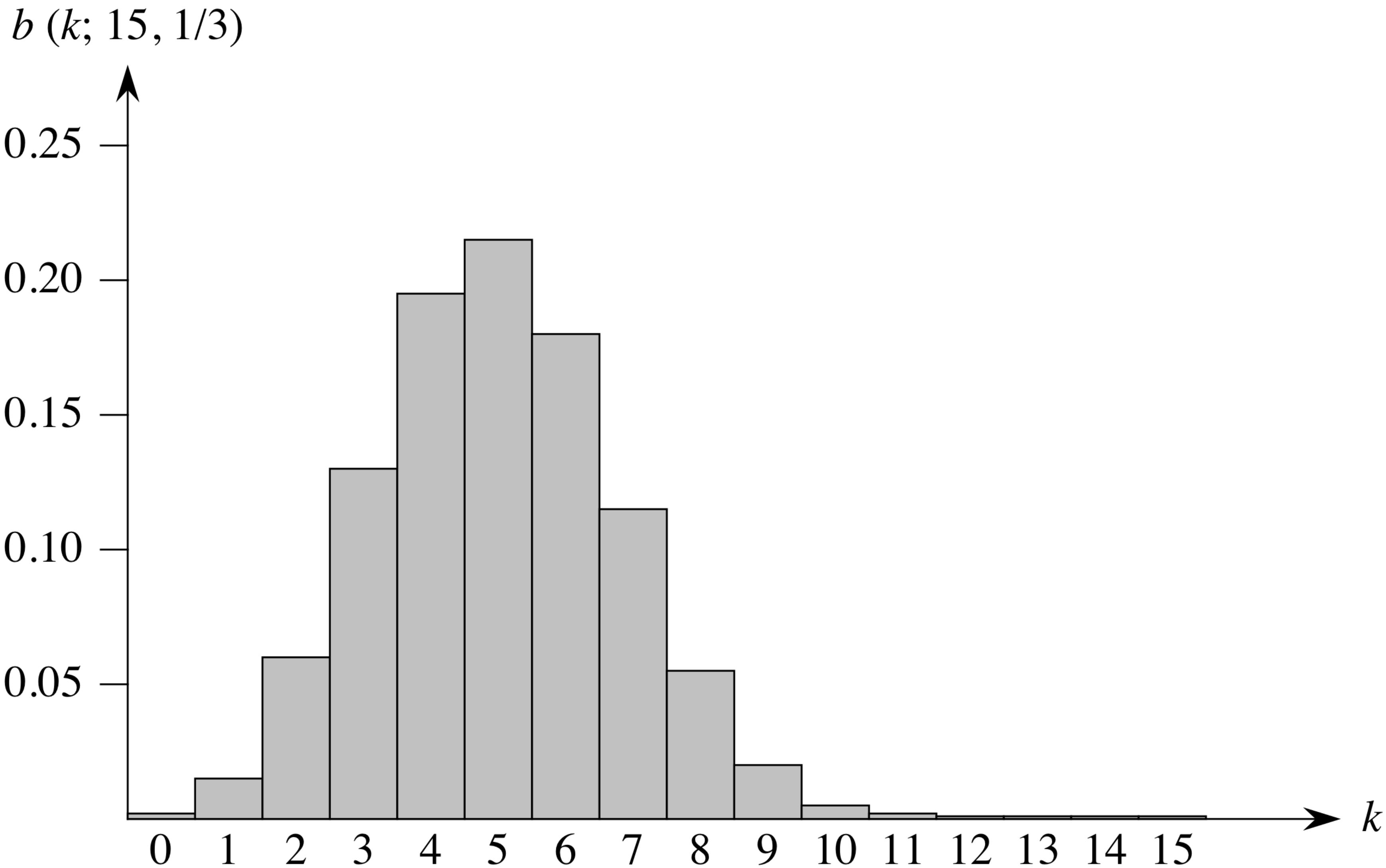
- Probability of k successes occur in n Bernoulli trials with probability p:

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

- 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:

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

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 ith trial.
  - X_i is 1 if the ith 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

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

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

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