Computer Science 351

Classical Probability Distributions

Instructor: Wayne Eberly

Department of Computer Science
University of Calgary

Lecture #23

Learning Goals

Learning Goals:

- Be aware of — and know the names for — several “classical” probability distributions that have been studied in the literature on probability theory.
- These are studied because they arise in a variety of applications, including applications of computer science. By knowing about them you can avoid “re-inventing the wheel” by trying to carry out analyses that someone else has completed already (and that others can also look at, when needed).

More About This Material

More About This Material:

- It is possible that this should have been included near the *beginning* of the discussion of this topic instead of near the end — because some of it can motivate the study of expectation, and tail bounds, that came before this.
- The material in this presentation is based, heavily, on material from Chapter 4 in David Stirzaker's text, *Probability and Random Variables: A Beginner's Guide*. A link to an electronic copy of this book is available on the course web site.

Waiting: Geometric Distributions

One Version of the Situation:

- You repeatedly try to accomplish something (like getting "Heads" when tossing a fair coin, or getting "6" when you roll a die). Each time you try, you succeed with probability p, where $0 < p < 1$ — so that you fail with probability $1 - p$.
- You try up to K times — giving up and stopping, if you failed K times.
- This is called ***the geometric distribution truncated at K, with parameter p***.

Waiting: Geometric Distributions

- In this case the ***sample space*** can be represented as

$$\Omega = \{0, 1, 2, \ldots, K\}$$

where, for $0 \leq i \leq K - 1$, the outcome "i" represents failing i times and then succeeding after that — and the outcome "K" represents failing K times and then giving up.

- The corresponding ***probability distribution*** is the function

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

such that $\mathrm{P}(i) = p \cdot (1 - p)^i$ for $0 \leq i \leq K - 1$ and such that $\mathrm{P}(K) = (1 - p)^K$.

Waiting: Geometric Distributions

- If we are interested in the number of times we try, before we either succeed or give up, then we are interested in the random variable

  $$X : \Omega \to \mathbb{N}$$

  such that $X(i) = i + 1$ for every integer i such that $0 \leq i \leq K - 1$, and such that $X(K) = K$.

Waiting: Geometric Distributions

Another Version of the Situation:

- A related situation — which might be called "indefinite waiting" — is one where ***you never stop trying, until you succeed***.
- That is, there is no upper bound "K" on the number of attempts you make before you give up — because you never give up, at all.
- This is called the ***geometric distribution with parameter p***.

Waiting: Geometric Distributions

- In this case the ***sample space*** can be represented as

$$\Omega = \{0, 1, 2, \dots\} = \mathbb{N}$$

(assuming that we define the set, $\mathbb{N}$, of natural numbers so that $0 \in \mathbb{N}$) where, for $i \geq 0$, the outcome "*i*" represents failing *i* times and then succeeding after that.

Waiting — Geometric Distributions

- One can also add an outcome "$+\infty$" — so that

$$\Omega = \mathbb{N} \cup \{+\infty\}$$

instead — where the outcome "$+\infty$" represents failing repeatedly, without ever succeeding at all. However, virtually any useful probability distribution $\mathrm{P} : \Omega \to \mathbb{R}$ would set $\mathrm{P}(+\infty)$ to be 0 — and you would get the same answers for interesting questions as you would, if you did not include this outcome, at all.

Waiting — Geometric Distributions

- The corresponding ***probability distribution*** is the function

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

such that $\mathrm{P}(i) = p \cdot (1-p)^i$ for $i \geq 0$ and such $\mathrm{P}(+\infty) = 0$, if an outcome "$+\infty$" is included in Ω.

Waiting: Geometric Distributions

- If we are interested in the number of times we try, before we either succeed or give up, then we are interested in the random variable
$$X : \Omega \to \mathbb{N}$$
such that $X(i) = i + 1$ for every integer i such that $i \geq 0$.
- The lecture presentation will include more information about this.

Counting Successes: The Binomial Distribution

This concerns a situation where we are repeatedly trying to accomplish a task — in particular, where we are trying to do this exactly n times, for a positive integer n.

- In the literature on probability theory, each of these attempts is called a ***Bernoulli trial***.
- On each attempt, we *succeed* with probability p and *fail* with probability $q = 1 - p$, for a real number p such that $0 \leq p \leq 1$.
- We are now interested in counting the number of times we succeed.

Counting Successes: The Binomial Distribution

This situation is sometimes modelled using a sample space

$$\Omega = \{0, 1, 2, \ldots, n\}$$

where, for each integer i such that $0 \leq i \leq n$, the outcome "i" represents the situation where exactly i of the n attempts are successful.

- Since there are $\binom{n}{i}$ ways to choose the attempts that succeed, the probability distribution corresponding this is the distribution $\text{P} : \Omega \to \mathbb{R}$ such that

  $$\text{P}(i) = \binom{n}{i} \times p^i \times q^{n-i}$$

  for every integer i such that $0 \leq i \leq n$.
- This is called the ***Binomial distribution with parameters n and p.***

Counting How Many Attempts We Need: The Negative Binomial Distribution

Suppose, instead, that we performing a sequence of Bernoulli trials, but we stop when we have succeeded exactly k times — for some positive integer k.

- Since we certainly need at least k trials in order to be successful, the sample space is the set

$$\Omega = \{k, k+1, k+2, \dots\} = \{i \in \mathbb{N} \mid i \geq k\}.$$

  As with the geometric distribution we could also add an outcome "$+\infty$" — but its probability be zero, for any probability distribution of interest, so it can be left out.

Counting How Many Attempts We Need: The Negative Binomial Distribution

- Suppose, once again, that we *succeed* with probability p and *fail* with probability $1 - p$, on every attempt — where, now, $0 < p < 1$ — and that trials are independent.
- For any integer i such that $i \geq k$, we use exactly i trials if and only if the following conditions are satisfied:
  - We succeed on exactly $k - 1$ of the first $i - 1$ trials. Note that there are exactly $\binom{i-1}{k-1}$ choices of the first $i - 1$ trials that can be successful in this case — and, for every such choice, the probability of this set of successes and failures is $p^{k-1}q^{i-k}$.
  - The i^{th} attempt is successful. This happens with probability p.

Counting How Many Attempts We Need: The Negative Binomial Distribution

- This corresponds to a probability distribution $\mathrm{P} : \Omega \to \mathbb{R}$ such that
$$\mathrm{P}(i) = \binom{i-1}{k-1} \times p^k \times q^{i-k}$$
for every integer i such that $i \geq k$.
- This is called the ***negative Binomial distribution*** with parameter p.

Sampling: The Hypergeometric Distribution

Yet another pair of related situations concerning sampling from a large population that consists of two groups. For example, at a huge, huge *science fiction convention* one might understand that the attendees include *Star Trek fans* and *Star Wars fans*.[1]

- There are a $T + W$ conference attendees, in total, for integers $T, W \geq 0$. In particular, the attendees include T "Star **T**rek" fans, and W "Star **W**ars" fans.
- For some positive integer n, we "poll" or *sample* from the collection of attendees, n times. For each integer i such that $0 \leq i \leq n$, we are interested in the probability that i of the attendees whom you poll are "Star Trek" fans.

[1] We are going to pretend, for time being, that no attendees like both, or neither.

Sampling: The Hypergeometric Distribution

One Version of the Situation: *Sampling with Replacement.*

- Suppose, first, that we might poll each attendee more than once. In particular, every time we poll an attendee, we might speak with each conference attendee with probability $\frac{1}{T+W}$.
- This corresponds to a ***Binomial distribution*** with parameters n and $p = \frac{T}{T+W}$ (if "success" corresponds to a "StarTrek" fan — or to the parameter $p = \frac{W}{T+W}$ if "success" corresponds to a "Star Trek" fan, instead).

Sampling: The Hypergeometric Distribution

Another Version of the Situation: *Sampling without Replacement.*

- Suppose, instead, that we *cannot* poll each attendee more than once — so that $n \leq T + W$.
- For the *first* attempt, we see a "Star Trek" fan with probability $\frac{T}{T+W}$ and we see a "Star Wars" fan with probability $\frac{W}{T+W}$, just as before.
- *If* we saw a "Star Trek" fan the first time, then we see a "Star Trek" fan with probability $\frac{T-1}{T+W-1}$ and we see a "Star Wars" fan with probability $\frac{W}{T+W-1}$ the second time.
- If we saw a "Star Wars" fan the first time, then we see a "Star Trek" fan with probability $\frac{T}{T+W-1}$ and we see a "Star Wars" fan with probability $\frac{W-1}{T+W-1}$ the second time.

Sampling: The Hypergeometric Distribution

- If we are interested in the number of "Star Trek" fans that we poll then the sample space can be chosen to be

  $$\Omega = \{0, 1, 2, \dots, n\}$$

  where, for $0 \leq i \leq n$, "i" represents the situation where we poll exactly i "Star Trek" fans and exactly $n - i$ "Star Wars" fans.
- There are exactly $\binom{T+W}{n}$ ways to choose attendees to be polled — and each of them is (by assumption) equally likely. (For each of these, there are $n!$ ways to *order* the attendees — but we won't worry about the order in which we see the attendees we poll.)

Sampling: The Hypergeometric Distribution

Let i be an integer such that $0 \leq i \leq \min(n, T)$ and such that $0 \leq n - i \leq \min(n, W)$ (so that it is actually possible to choose i “Star Trek” fans and $n - i$ “Star Wars” fans).

- There are $\binom{T}{i}$ ways to choose the “Star Trek” fans whom we poll. These choices are all equally likely.
- There are $\binom{W}{n-i}$ ways to choose the “Star Wars” fans whom we poll. These choices are equally likely.
- It follows that, for an integer i such that $0 \leq i \leq T$ and $0 \leq n - i \leq W$, the probability that we poll exactly i “Star Trek” fans is

$$\mathrm{P}(i) = \frac{\binom{T}{i} \times \binom{W}{n-i}}{\binom{T+W}{n}}$$

Approximations

- As the Hypergeometric Distribution may suggests, expressions for probabilities can get complicated and difficult to work with.
- In some circumstances, it might be acceptable to work *approximations* instead.
- Consider, “Sampling without Replacement”; When the size of both parts of the population (in the above example, T and W) are both much, much larger than the number n of participants that we poll, and $0 \leq i \leq n$, then the above probability, $P(i)$ has the corresponding probability for sampling *with* replacement, using the parameter $p = \frac{T}{W+T}$.

Approximations

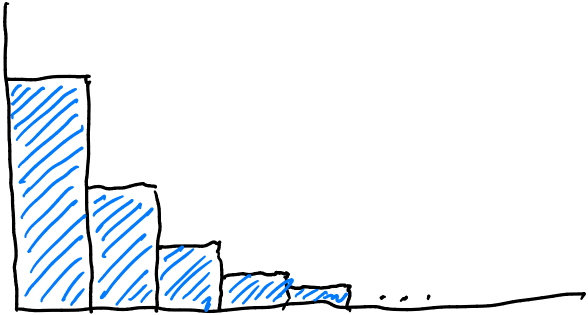
- In particular, $P(i)$ numerically close to the value

$$\binom{n}{i} \times p^i \times q^i$$

— with $p = \frac{T}{W+T}$ and $q = 1 - p = \frac{W}{W+T}$ — that is obtained using the Binomial distribution with parameters n and p.

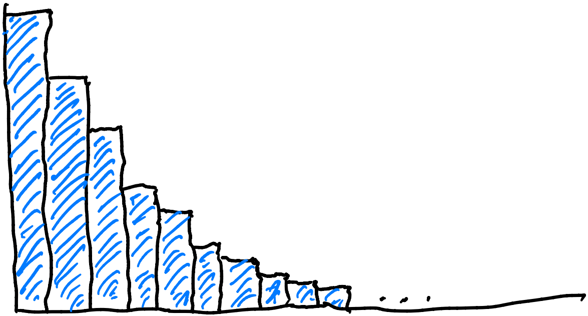
Considering the Geometric Distribution, Again

Once again, consider the geometric distribution — assuming, again, that we succeed with probability p on each attempt and fail with probability $q = 1 - p$ — where $0 < p < 1$. If we were to plot the probability that we need i attempts, for increasing i, then the plot would look like this:



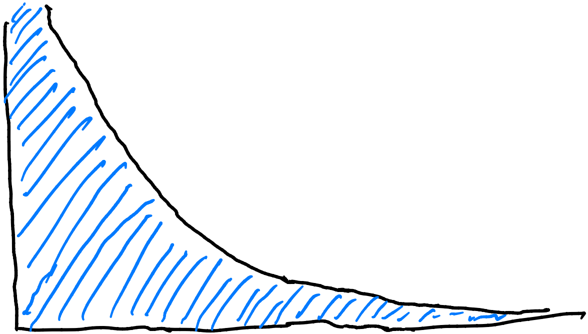
Considering The Geometric Distribution, Again

Suppose we doubled the rate at which we made attempts (so that there were two attempts per time interval instead of one) — with the probability of failure multiplied by $\frac{1}{\sqrt{q}}$ instead of $\frac{1}{q}$ each time we made an attempt. The plot would like this, instead:



Considering the Geometric Distribution, Again

If we continue to do this then, eventually, the plot starts to look more like a continuous curve than a step function:



Considering the Geometric Distribution, Again

- Now — when the number of attempts per unit of time is very large, the probability that we are successful, before a given amount of time, can be ***approximated by the area under a part of this curve***.

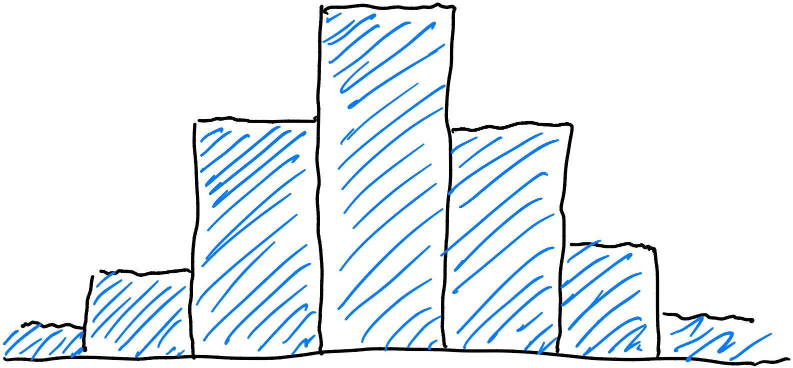
Sampling from Huge Populations

Consider ***sampling from a huge population***.

- If the “science fiction conference”’ gets larger, and larger (taking up more and more of the city) — but the proportion of “Star Trek” fans and “Star Wars” fans does not change — then the effects will resemble what happens if we repeat the above process — starting with a plot for the ***Binomial Distribution*** instead of the ***Geometric Distribution***.

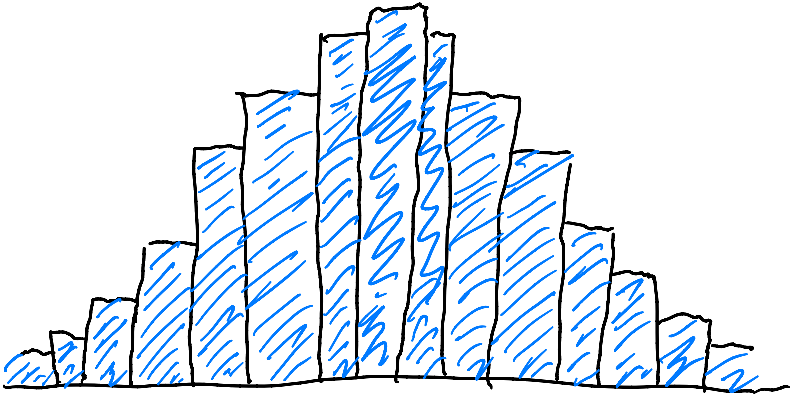
Sampling from Huge Populations

We could go from this...



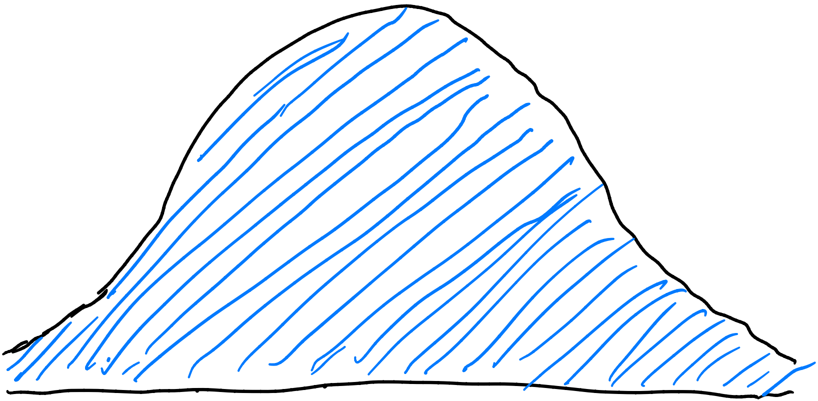
Sampling from Huge Populations

...to this...



Sampling from Huge Populations

...to this.



Sampling from Huge Populations

- Under *these* circumstances, things might also have turned around: We might now think of *the area of the continuous curve* as the “probability” that we wish to discover, while *the area under the step function* is the ***approximation of that probability*** that we know how to calculate.
- Now we need a way to define and study “probabilities” where the sample space is ***uncountably infinite*** — either the set $\mathbb{R}$ of real numbers, or a subset of the set real numbers inside some interval.

Continuous Probability Theory

Definition: Let $\Omega = \mathbb{R}$ — or let $\Omega = \{x \in \mathbb{R} \mid a \leq x \leq b\}$ for a pair of real number a and b such that $a < b$. An integrable function $f : \Omega \to \mathbb{R}$ is a ***probability density function*** for Ω if it satisfies the following properties.

(a) $f(x) \geq 0$ for every real number x such that $x \in \Omega$.

(b) $\int_{t \in \Omega} f(t) \mathrm{d}t = 1$.

Continuous Probability Theory

- In this definition, "$\int_{t\in\Omega} f(t)\mathrm{d}t$" is

$$\int_{-\infty}^{+\infty} f(t)\mathrm{d}t$$

if $\Omega = \mathbb{R}$, and "$\int_{t\in\Omega} f(t)\mathrm{d}t$" is

$$\int_{a}^{b} f(t)\mathrm{d}t$$

if $\omega = \{x \in \mathbb{R} \mid a \leq x \leq b$.

Continuous Probability Theory

Definition: A random variable $X : \Omega \to \mathbb{R}$ is said to be ***continuous***, with density $f : \Omega \to \mathbb{R}$, if

$$\mathrm{P}(\alpha \le X \le \beta) = \int_{\alpha}^{\beta} f(t)\mathrm{d}t$$

for all real numbers α and β such that $\alpha, \beta \in \Omega$ and $\alpha \le \beta$.

Continuous Probability Theory

Similarly,

$$\mathsf{P}(X \leq \beta) = \int_{\substack{t \in \Omega \\ t \leq \beta}} f(t) \mathrm{d}t$$

and

$$\mathsf{P}(X \geq \alpha) = \int_{\substack{t \in \Omega \\ t \geq \alpha}} f(t) \mathrm{d}t.$$

With continuous random variables, we only consider probabilities of the above form (that is, probabilities that the values of X lies in some interval).

Continuous Probability Theory

Definition: If a random variable $X : \Omega \to \mathbb{R}$ is continuous, with density $f : \Omega \to \mathbb{R}$, then the ***expected value*** of X is

$$\mathsf{E}[X] = \int_{t \in \Omega} t \cdot f(t) \mathrm{d}t.$$

Now the ***variance*** of X can be defined as before:

$$\begin{aligned} \mathrm{var}(X) &= \mathsf{E}[(X - \mathsf{E}[X])^2] \\ &= \int_{t \in \Omega} (t - \mu)^2 \mathrm{d}t \end{aligned}$$

where $\mu = \mathsf{E}[X]$.

Results similar to those established for *discrete random variables* can (sometimes) be established for continuous random variables too. This is outside the scope of this course.

Continuous Probability Theory

One can also consider "classical" *continuous* probability distributions

- ***Exponential distributions*** are distributions resembling the ones that we get by starting with *geometric distributions* and applying the process that has now been described.
- ***Gaussian distributions*** — also called "normal distributions" — are distributions resembling the ones that we get by starting with *Binomial distributions* and applying the process that has now been described.

Continuous Probability Theory

- There is not time, in this course, to discuss continuous probability theory. However, this has significant applications in computer science, just as discrete probability theory does.
- It can be helpful to be able to recognize “classical” continuous probability distributions, just like it can be helpful to recognize “classical” discrete probability distributions, because they might arise in the applications you are interested in.