Stopping problems

Prepared by Mark on June 9, 2026

Instructor's Handout

This file contains solutions and notes.
Compile with the "nosolutions" flag before distributing.
Click [here] for the latest version of this handout.

Part 1: Probability

Definition 1:
A *sample space* is a finite set Ω.
The elements of this set are called *outcomes*.
An *event* is a set of outcomes (i.e, a subset of of Ω).

Definition 2:
A *probability function* over a sample space Ω is a function $\mathcal{P} : P(\Omega) \to [0, 1]$ that maps events to real numbers between 0 and 1.
Any probability function has the following properties:

- $\mathcal{P}(\varnothing) = 0$
- $\mathcal{P}(\Omega) = 1$
- For events A and B where $A \cap B = \varnothing$, $\mathcal{P}(A \cup B) = \mathcal{P}(A) + \mathcal{P}(B)$

Problem 3:
Say we flip a fair coin three times.
List all elements of the sample space Ω this experiment generates.

Problem 4:
Using the same setup as Problem 3, find the following:

- $\mathcal{P}(\ \{\omega \in \Omega \mid \omega \text{ has at least two "heads"}\}\ )$
- $\mathcal{P}(\ \{\omega \in \Omega \mid \omega \text{ has an odd number of "heads"}\}\ )$
- $\mathcal{P}(\ \{\omega \in \Omega \mid \omega \text{ has at least one "tails"}\}\ )$

Definition 5:
Given a sample space Ω and a probability function $\mathcal{P}$,
a *random variable* is a function from Ω to a specified output set.

For example, given the three-coin-toss sample space $\Omega = \{\texttt{TTT}, \texttt{TTH}, \texttt{THT}, \texttt{THH}, \texttt{HTT}, \texttt{HTH}, \texttt{HHT}, \texttt{HHH}\}$,
We can define a random variable $\mathcal{H}$ as "the number of heads in a throw of three coins".
As a function, $\mathcal{H}$ maps values in Ω to values in $\mathbb{Z}_0^+$ and is defined as:

- $\mathcal{H}(\texttt{TTT}) = 0$
- $\mathcal{H}(\texttt{TTH}) = 1$
- $\mathcal{H}(\texttt{THT}) = 1$
- $\mathcal{H}(\texttt{THH}) = 2$
- ...and so on.

Intuitively, a random variable assigns a "value" in $\mathbb{R}$ to every possible outcome.

Definition 6:
We can compute the probability that a random variable takes a certain value by computing the probability of the set of outcomes that produce that value.

For example, if we wanted to compute $\mathcal{P}(\mathcal{H} = 2)$, we would find $\mathcal{P}\big(\{\texttt{THH}, \texttt{HTH}, \texttt{HHT}\}\big)$.

Problem 7:
Say we flip a coin with $\mathcal{P}(\texttt{H}) = 1/3$ three times.
What is $\mathcal{P}(\mathcal{H} = 1)$, with $\mathcal{H}$ defined as above?
What is $\mathcal{P}(\mathcal{H} = 5)$?

Problem 8:
Say we roll a fair six-sided die twice.
Let $\mathcal{X}$ be a random variable measuring the sum of the two results.
Find $\mathcal{P}(\mathcal{X} = x)$ for all x in $\mathbb{Z}$.

Definition 9:
Say we have a random variable $\mathcal{X}$ that produces outputs in $\mathbb{R}$.
The *expected value* of $\mathcal{X}$ is then defined as

$$\mathcal{E}(\mathcal{X}) \coloneqq \sum_{x \in \mathbb{R}} \Big(x \times \mathcal{P}(\mathcal{X} = x) \Big) = \sum_{\omega \in \Omega} \Big(\mathcal{X}(\omega) \times \mathcal{P}(\omega) \Big)$$

That is, $\mathcal{E}(\mathcal{X})$ is the average of all possible outputs of $\mathcal{X}$ weighted by their probability.

Problem 10:
Say we flip a coin with $\mathcal{P}(\texttt{H}) = 1/3$ two times.
Define $\mathcal{H}$ as the number of heads we see.
Find $\mathcal{E}(\mathcal{H})$.

Problem 11:
Let $\mathcal{A}$ and $\mathcal{B}$ be two random variables.
Show that $\mathcal{E}(\mathcal{A} + \mathcal{B}) = \mathcal{E}(\mathcal{A}) + \mathcal{E}(\mathcal{B})$.

Solution

Use the second definition of $\mathcal{E}$, $\sum_{\omega \in \Omega} \Big(\mathcal{X}(\omega) \times \mathcal{P}(\omega) \Big)$.

Make sure students understand all parts of Definition 9, and are comfortable with the fact that a random variable "assigns values" to outcomes.

Definition 12:
Let A and B be events on a sample space Ω.
We say that A and B are *independent* if $\mathcal{P}(A \cap B) = \mathcal{P}(A) \times \mathcal{P}(B)$.
Intuitively, events A and B are independent if the outcome of one does not affect the other.

Definition 13:
Let $\mathcal{A}$ and $\mathcal{B}$ be two random variables over Ω.
We say that $\mathcal{A}$ and $\mathcal{B}$ are independent if the events $\{\omega \in \Omega \mid \mathcal{A}(\omega) = a\}$ and $\{\omega \in \Omega \mid \mathcal{B}(\omega) = b\}$ are independent for all (a, b) that $\mathcal{A}$ and $\mathcal{B}$ can produce.

Part 2: Introduction

Setup:
Suppose we toss a 6-sided die n times.
It is easy to detect the first time we roll a 6.
What should we do if we want to detect the *last*?

Problem 14:
Given $l \leq n$, what is the probability that the last l tosses of this die contain exactly one six?
Hint: Start with small l.

Solution

$\mathcal{P}(\text{last } l \text{ tosses have exactly one } 6) = (1/6)(5/6)^{l-1} \times l$

Problem 15:
For what value of l is the probability in Problem 14 maximal?
The following table may help.
We only care about integer values of l.

l	$(5/6)^l$	$(1/6)(5/6)^l$
0	1.00	0.167
1	0.83	0.139
2	0.69	0.116
3	0.58	0.096
4	0.48	0.080
5	0.40	0.067
6	0.33	0.056
7	0.28	0.047
8	0.23	0.039

Solution

$(1/6)(5/6)^{l-1} \times l$ is maximal at $l = 5.48$, so $l = 5$.
$l = 6$ is close enough.

Problem 16:
Finish your solution:
In n rolls of a six-sided die, what strategy maximizes our chance of detecting the last 6 that is rolled?
What is the probability of our guess being right?

Solution

Whether $l = 5$, 5.4, or 6, the probability of success rounds to 0.40.

Part 3: The Secretary Problem

Definition 17: The secretary problem
Say we need to hire a secretary. We have exactly one position to fill, and we must fill it with one of n applicants. These n applicants, if put together, can be ranked unambiguously from "best" to "worst".

We interview applicants in a random order, one at a time.
At the end of each interview, we either reject the applicant (and move on to the next one), or select the applicant (which fills the position and ends the process).

Each applicant is interviewed at most once—we cannot return to an applicant we've rejected.
In addition, we cannot reject the final applicant, as doing so will leave us without a secretary.

For a given n, we would like to maximize our probability of selecting the best applicant.
This is the only metric we care about—we do not try to maximize the rank of our applicant.
Hiring the second-best applicant is no better than hiring the worst.

Problem 18:
If $n = 1$, what is the best hiring strategy, and what is the probability that we hire the best applicant?

Solution

This is trivial. Hire the first applicant, she's always the best.

Problem 19:
If $n = 2$, what is the best hiring strategy, and what is the probability that we hire the best applicant?
Is this different than the probability of hiring the best applicant at random?

Solution

There are two strategies:
- hire the first
- hire the second

Both are equivalent to the random strategy.

Intuitively, the fact that a strategy can't help us makes sense:
When we're looking at the first applicant, we have no information;
when we're looking at the second, we have no agency (i.e, we *must* hire).

Problem 20:
If $n = 3$, what is the probability of hiring the best applicant at random?
Come up with a strategy that produces better odds.

Solution

Once we have three applicants, we can make progress.

The remark from the previous solution still holds:
When we're looking at the first applicant, we have no information;
when we're looking at the last, we have no choices.

So, let's make our decision at the second candidate.
If we hire only when the second candidate is better than the first,
we end up hiring the best candidate exactly half the time.

This can be verified by checking all six cases.

Problem 21:
Should we ever consider hiring a candidate that *isn't* the best we've seen so far?
Why or why not? *Hint:* Read the problem again.

Solution

No! A candidate that isn't the best yet cannot be the best overall!
Remember—this problem is only interested in hiring the *absolute best* candidate.
Our reward is zero in all other cases.

Remark 22:
Problem 21 implies that we should automatically reject any applicant that isn't the best we've seen.
We can take advantage of this fact to restrict the types of strategies we consider.

Remark 23:
Let B_x be the event "the x^{th} applicant is better than all previous applicants,"
and recall that we only know the *relative* ranks of our applicants:
given two candidates, we know *which* is better, but not *by how much*.

Therefore, the results of past events cannot provide information about future B_x.
All events B_x are independent.

We can therefore ignore any strategy that depends on the outcomes of individual B_x. Given this realization, we are left with only one kind of strategy:
We blindly reject the first $(k-1)$ applicants, then select the next "best-yet" applicant.
All we need to do now is pick the optimal k.

Problem 24:
Consider the secretary problem with a given n.
What are the probabilities of each B_x?

Problem 25:
What is the probability that the n^{th} applicant is the overall best applicant?

Solution

All positions are equally likely. $1/n$.

Problem 26:
Given that the x^{th} applicant is the overall best, what is the probability of hiring this applicant if we use the "look-then-leap" strategy detailed above?
Hint: Under what conditions would we *not* hire this applicant?
This probability depends on k and x.

Solution

Say that the x^{th} applicant is the best overall. If we do not hire this applicant, we must have hired a candidate that came before them.

What is the probability of this? We saw $x-1$ applicants before the x^{th}.
If we hired one of them, the best of those initial $x-1$ candidates did *not* fall into the initial $k-1$ applicants we rejected. (This is again verified by contradiction: if the best of the first $x-1$ applicants *was* within the first $k-1$, we would hire the x^{th})

There are $x-1$ positions to place the best of the first $x-1$ candidates,
and $k-1$ of these positions are initially rejected.
Thus, the probability of the best of the first $x-1$ applicants being rejected is $\frac{k-1}{x-1}$.

Unraveling our previous logic, we find that the probability we are interested in is also $\frac{k-1}{x-1}$.
Assuming that $x \geq k$. Of course, this probability is 0 otherwise.

Problem 27:
Consider the secretary problem with n applicants.
If we reject the first k applicants and hire the first "best-yet" applicant we encounter,
what is the probability that we select the best candidate?
Call this probability $\phi_n(k)$.

Solution

Using Problem 25 and Problem 26, this is straightforward:

$$\phi_n(k) = \sum_{x=k}^{n} \left(\frac{1}{n} \times \frac{k-1}{x-1} \right)$$

Problem 28:
Find the k that maximizes $\phi_n(k)$ for n in $\{1, 2, 3, 4, 5\}$.

Solution

Brute force. We already know that $\phi_1(1) = 1.0$ and $\phi_2(1) = \phi_3(2) = 0.5$.
The maximal value of ϕ_4 is $\phi_4(2) = 0.46$, and of ϕ_5 is $\phi_5(3) = 0.43$.

Problem 29:
Let $r = \frac{k-1}{n}$, the fraction of applicants we reject. Show that

$$\phi_n(k) = r \sum_{x=k}^{n} \left(\frac{1}{x-1} \right)$$

Solution

This is easy.

Problem 30:
With a bit of fairly unpleasant calculus, we can show that the following is true for large n:

$$\sum_{x=k}^{n} \frac{1}{x-1} \approx \ln\left(\frac{n}{k}\right)$$

Use this fact to find an approximation of $\phi_n(k)$ at large n in terms of r.
Hint: If n is big, $\frac{k-1}{n} \approx \frac{k}{n}$.

Solution

$$\phi_n(k) = r \sum_{x=k}^{n} \left(\frac{1}{x-1} \right) \approx r \times \ln\left(\frac{n}{k}\right) = -r \times \ln\left(\frac{k}{n}\right) \approx -r \times \ln(r)$$

Problem 31:
Find the r that maximizes $\lim_{n\to\infty} \phi_n$.
Also, find the value of ϕ_n at this point.
If you aren't familiar with calculus, ask an instructor for help.

Solution

Use the usual calculus tricks:

$$\frac{d}{dr}\big(-r \times \ln(r)\big) = -1 - \ln(r)$$

Which is zero at $r = e^{-1}$. The value of $-r \times \ln(r)$ at this point is also $\frac{1}{e}$.

Thus, the "look-then-leap" strategy with $r = e^{-1}$ should select the best candidate about $e^{-1} = 37\%$ of the time, *regardless of* n. Our probability of success does not change as n gets larger!
Recall that the random strategy succeeds with probability $1/n$.
That is, it quickly becomes small as n gets large.

Part 4: Another Secretary Problem

As you may have already noticed, the secretary problem we discussed in the previous section is somewhat disconnected from reality. Under what circumstances would one only be satisfied with the *absolute best* candidate? It may make more sense to maximize the average rank of the candidate we hire, rather than the probability of selecting the best. This is the problem we'll attempt to solve next.

Definition 32:
The problem we're solving is summarized below. Note that this is nearly identical to the classical secretary problem in the previous section—the only thing that has changed is the goal.

- We have exactly one position to fill, and we must fill it with one of n applicants.
- These n applicants, if put together, can be ranked unambiguously from "best" to "worst".
- We interview applicants in a random order, one at a time.
- After each interview, we either reject or select the applicant.
- We cannot return to an applicant we've rejected.
- The process ends once we select an applicant.

- Our goal is to maximize the rank of the applicant we hire.

Definition 33:
Just like before, we need to restate this problem in the language of probability.
To do this, we'll say that each candidate has a *quality* rating in $[0, 1]$.

Our series of applicants then becomes a series of random variables $\mathcal{X}_1, \mathcal{X}_2, ..., \mathcal{X}_n$,
where each $\mathcal{X}_i$ is drawn uniformly from $[0, 1]$.

Problem 34:
The modification in Definition 33 doesn't fully satisfy the constraints of the secretary problem.
Why not?

Solution

If we observe $\mathcal{X}_i$ directly, we obtain *absolute* scores.
This is more information than the secretary problem allows us to have—we can know which of two candidates is better, but *not by how much*.

Ignore this issue for now. We'll return to it later.

Problem 35:
Let $\mathcal{X}$ be a random variable uniformly distributed over $[0, 1]$.
Given a real number x, what is the probability that $\mathcal{P}(\mathcal{X} \leq x)$?

Solution

$$\mathcal{P}(\mathcal{X} \leq x) = \begin{cases} 0 & x \leq 0 \\ x & 0 < x < 1 \\ 1 & \text{otherwise} \end{cases}$$

Problem 36:
Say we have five random variables $\mathcal{X}_1, \mathcal{X}_2, ..., \mathcal{X}_5$.
Given some y, what is the probability that all five $\mathcal{X}_i$ are smaller than y?

Solution

Naturally, this is $\mathcal{P}(\mathcal{X} \leq y)^5$, which is y^5.

Definition 37:
Say we have a random variable $\mathcal{X}$ which we observe n times. (for example, we repeatedly roll a die)
We'll arrange these observations in increasing order, labeled $x_1 < x_2 < ... < x_n$.
Under this definition, x_i is called the i^{th} *order statistic*—the i^{th} smallest sample of $\mathcal{X}$.

Problem 38:
Say we have a random variable $\mathcal{X}$ uniformly distributed on $[0, 1]$, of which we take 5 observations.
Given some y, what is the probability that $x_5 < y$? How about $x_4 < y$?

Solution

$x_5 < y$: This is a restatement of the previous problem.

$x_4 < y$: We need 4 measurements to be smaller, and one to be larger. Accounting for permutations, we get $5\mathcal{P}(\mathcal{X} \leq y)^4 \mathcal{P}(\mathcal{X} > y) + \mathcal{P}(\mathcal{X} \leq y)^5$, which is $5y^4(1 - y) + y^5$.

Problem 39:
Consider the same setup as Problem 38, but with n measurements.
What is the probability that $x_i < y$ for a given y?

Solution

$$\mathcal{P}(x_i < y) \ = \ \sum_{j=i}^{n} \binom{n}{j} \times y^j (1 - y)^{n-j}$$

Remark 40:
The expected value of the i^{th} order statistic on n samples of the uniform distribution is below.

$$\mathcal{E}(x_i) = \frac{i}{n + 1}$$

We do not have the tools to derive this yet.

Definition 41:
Recall Problem 34. We need one more modification.
In order to preserve the constraints of the problem, we will not be allowed to observe $\mathcal{X}_i$ directly.
Instead, we'll be given an "indicator" $\mathcal{I}_i$ for each $\mathcal{X}_i$, which produces values in $\{0, 1\}$.
If the value we observe when interviewing $\mathcal{X}_i$ is the best we've seen so far, $\mathcal{I}_i$ will produce 1.
If it isn't, $\mathcal{I}_i$ produces 0.

Problem 42:
Given a secretary problem with n applicants, what is $\mathcal{E}(\mathcal{I}_i)$?

Solution

$$\mathcal{E}(\mathcal{I}_i) = \frac{1}{i}$$

Problem 43:
What is $\mathcal{E}(\mathcal{X}_i \mid \mathcal{I}_i = 1)$?
In other words, what is the expected value of $\mathcal{X}_i$ given that we know this candidate is the best we've seen so far?

Solution

This is simply the expected value of the i^{th} order statistic on i samples:

$$\mathcal{E}(\mathcal{X}_i \mid \mathcal{I}_i = 1) = \frac{i}{i+1}$$

Problem 44:
In the previous section, we found that the optimal strategy for the classical secretary problem is to reject the first $e^{-1} \times n$ candidates, and select the next "best-yet" candidate we see.

How effective is this strategy for the ranked secretary problem?
Find the expected rank of the applicant we select using this strategy.

Problem 45:
Assuming we use the same kind of strategy as before (reject k, select the next "best-yet" candidate), show that $k = \sqrt{n}$ optimizes the expected rank of the candidate we select.

Solution

This is a difficult bonus problem. see `Neil Bearden, J. (2006). A new secretary problem with rank-based selection and cardinal payoffs.`