Oliver Song 6.041 Exam 1 sheet

Basic shit

Sample space (generally denoted with Ω): the set of all possible outcomes of an experiment
Probability law: every possible outcome of an experiment has a probability of occurring

Axioms

All probabilities must be non-negative: $P(A) \geq 0$ for all A
For disjoint events sum is union: $P(A \cup B) = P(A) + P(B)$, if events are disjoint. Works for more than 2 events too.
Total probability is 1: $P(\Omega) = 1$

Basic laws

- If $A \subset B$, then $P(A) \leq P(B)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $P(A \cup B) \leq P(A) + P(B)$
- $P(A \cup B \cup C) = P(A) + P(A^c \cap B) + P(A^c \cap B^c \cap C)$

Discrete shit

Discrete Probability Law: If Ω is finite, then each event A in Ω can be expressed as $A = \{s_1, s_2, ..., s_n\}$ for s_i in Ω.
Pretty much, the probability of any discrete event can be given by THE SUM OF THE PROB OF ALL ITS SMALLER EVENTS
Discrete Uniform Probability Law: If all outcomes are equally likely, then the probability of the event is the number of events it encompasses over all possible events. $P(A) = |A| / |\Omega|$

Conditional probability

$$\mathbf{P}(A|B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)}$$

s.t. $P(B) > 0$

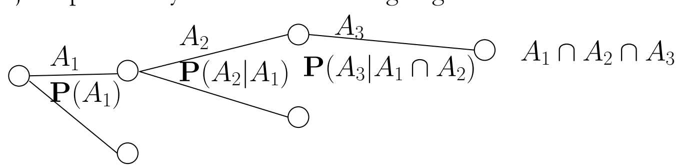
Satisfies
1. $P(A \mid B) \geq 0$
2. $P(\Omega \mid B) = 1$
3. $P(A_1 \cup A_2 \cup ... \mid B) = P(A_1|B) + P(A_2|B) + ...$ s.t. events are disjoint

Multiplication rule

Let $A_1, ... A_n$ be a set of events s.t. their joint probability > 0

$$\mathbf{P}\left(\bigcap_{i=1}^{n} A_i\right) = \mathbf{P}(A_1)\mathbf{P}(A_2|A_1)\mathbf{P}(A_3|A_1 \cap A_2) \cdots \mathbf{P}(A_n|\bigcap_{i=1}^{n-1} A_i)$$

^ joint probability ^ conditionals going down tree all cancel:



Independence

Events A and B are **independent** if $P(A \cap B) = P(A) * P(B)$ or $P(A|B) = P(B)$
Events are **conditionally independent** given an event C if
$P(A \cap B \mid C) = P(A|C) P(B|C)$ or
$P(A|B \cap C) = P(A|C)$ s.t. $P(B \cap C) > 0$

Independence of sets

Pairwise independence is when every pair is independent
$P(A_i \cap A_j) = P(A_i) * P(A_j)$
The entire set is independent if

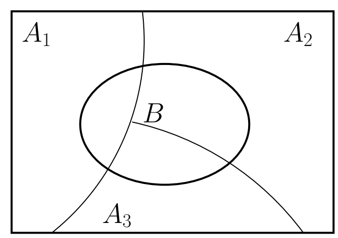
$$\mathbf{P}\left(\bigcap_{i \in S} A_i\right) = \prod_{i \in S} \mathbf{P}(A_i) \quad \forall\ S \subseteq \{1, 2, \ldots, n\}$$

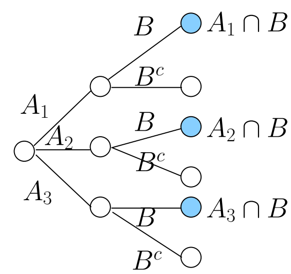
...which is supposed to mean something to me. I think it is the powerset of multiplication

Total Probability Theorem

Let $A_1, ... ,A_n$ be disjoint events that partition Ω. If $P(A_i) > 0$ for each i, then for any event B, this crap:

$$\mathbf{P}(B) = \sum_{i=1}^{n} \mathbf{P}(B \cap A_i) = \sum_{i=1}^{n} \mathbf{P}(B|A_i)\mathbf{P}(A_i)$$





Bayes Rule

$$\mathbf{P}(A_i|B) = \frac{\mathbf{P}(B|A_i)\mathbf{P}(A_i)}{\mathbf{P}(B)} = \frac{\mathbf{P}(B|A_i)\mathbf{P}(A_i)}{\sum_{i=1}^{n} \mathbf{P}(B|A_i)\mathbf{P}(A_i)}$$

Counting rules

Basic counting: For a m-stage process with n_i choices at stage i, the number of total choices is

$$\# \text{ choices} = n_1 * n_2 * ... * n_m$$

Sampling with Replacement: k-length sequences drawn from n distinct items with replacement:

$$\# \text{ sequences} = n^k$$

Permutations: how many ways can you order n things in k buckets?

$$\# \text{ sequences} = \frac{n!}{(n-k)!}$$

Combinations: how many different combinations of k things can you pick from n things?

$$\# \text{ sets} = \frac{n!}{k!(n-k)!}$$

Discrete Random Variables

A **random variable** is a real valued function defined on the sample space:

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

The **probability mass function (PMF)** for the RV, X, assigns a probability to each event $\{X = x\}$

$$p_X(x) = \mathbf{P}(\{X = x\}) = \mathbf{P}\big(\{\omega \in \Omega | X(\omega) = x\}\big)$$

That's all bullshit though. To really find a PMF, you have to go through some base cases for the event you want to find (say, after

2 iterations, after 3 iterations, after...) then find an equation describing the pattern for the probability of the event at the nth iteration. Then you can describe it with the fucking brackets

PMF(X=x) = {
    $x^2 + 2x$ | x > 0
    0 | e/w
}
or whatever shit like that.

PMF Properties

Let X be a RV and S be a countable subset of the real line
The axioms of probability still hold

$$p_X(x) \geq 0$$
$$\mathbf{P}(X \in S) = \sum_{x \in S} p_X(x)$$
$$\sum_x p_X(x) = 1$$

Remember that a PMF always sums to 1.

Partitions

What is this crap?

The number of ways to partition an n-element set into r disjoint subsets, with n_k elements in the k^{th} subset:

$$\binom{n}{n_1, n_2, \ldots, n_r} = \binom{n}{n_1}\binom{n-n_1}{n_2}\cdots\binom{n-n_1-\cdots-n_r-1}{n_r} = \frac{n!}{n_1!n_2!, \cdots, n_r!}$$

where

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$
$$\sum_{i=1}^{r} n_i = n$$

Expectation

Given a random variable X with PMF $p_X(x)$:

- $\mathbf{E}[X] = \sum_x x p_X(x)$
- Given a derived random variable $Y = g(X)$:

$$\mathbf{E}[g(X)] = \sum_x g(x)p_X(x) = \sum_y y p_Y(y) = E[Y]$$
$$\mathbf{E}[X^n] = \sum_x x^n p_X(x)$$

- **Linearity** of Expectation: $\mathbf{E}[aX + b] = a\mathbf{E}[X] + b$.

Variance

The expected value of a derived random variable $g(X)$ is

$$\mathbf{E}[g(X)] = \sum_x g(x)p_X(x)$$

The variance of X is calculated as

- $var(X) = \mathbf{E}[(X - \mathbf{E}[X])^2] = \sum_x (x - \mathbf{E}[X])^2 p_X(x)$
- $var(X) = \mathbf{E}[X^2] - \mathbf{E}[X]^2$
- $var(aX + b) = a^2 var(X)$

Canonical distributions (discrete)

	X	$p_X(k)$	$\mathbf{E}[X]$	$var(X)$
Bernoulli	$\begin{cases} 1 & \text{success} \\ 0 & \text{failure} \end{cases}$	$\begin{cases} p & k=1 \\ 1-p & k=0 \end{cases}$	p	$p(1-p)$
Binomial	Number of successes in n Bernoulli trials	$\binom{n}{k} p^k (1-p)^{n-k}$ $k = 0, 1, \ldots, n$	np	np(1-p)
Geometric	Number of trials until first success	$(1-p)^{k-1}p$ $k = 1, 2, \ldots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Uniform	An integer in the interval [a,b]	$\begin{cases} \frac{1}{b-a+1} & k = a, \ldots, b \\ 0 & \text{otherwise} \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)(b-a+2)}{12}$
Poisson	Number of rare events	$\frac{e^{-\lambda}\lambda^k}{k!}$ $k = 0, 1, 2, \ldots$	λ	λ

How to solve common problems

Coin flips heads with probability p
K is the # flips up to and including the 2nd head. What is **PMF**?

$$p_K(k) = (k-1)(1-p)^{k-2}p^2$$

Ok, what's the **conditional PMF** of K | first flip is heads?

$$p_{K|H_1}(k) = (1-p)^{k-2}p$$

Ok, what's the **expected value**?

Let T_1 denote the event that the first flip is a tail. Conditional on T_1, $(K-1)$ should have the same distribution as K *without the condition*, so

$$E[K|T_1] = E[K] + 1.$$

We now use the total expectation law, combined with the fact that the geometric random variable of parameter p has mean $1/p$, to write

$$E[K] = E[K|H_1] \cdot p + E[K|T_1] \cdot (1-p) = \left(1 + \frac{1}{p}\right) \cdot p + (E[K] + 1) \cdot (1-p)$$
$$= 2 + (1-p)E[K].$$

This yields

$$E[K] = \frac{2}{p}.$$

Ok, L is the number of identical flips before the first change.
What's the **PMF**?

$$p_L(\ell) = (1-p)^\ell \cdot p + p^\ell \cdot (1-p).$$

What's the **Expected Value**?

$$\mathbf{E}[L] = p \cdot \mathbf{E}[L|H_1] + (1-p) \cdot \mathbf{E}[L|T_1] = \frac{p}{1-p} + \frac{1-p}{p}.$$

What's the **variance**?

We use the same technique as in Part b) to compute $\mathbf{E}[L^2]$, where the expected value of the square of a geometric random variable can be easily computed from its mean and variance, given on Page 2:

$$\mathbf{E}[L^2] = p \cdot \mathbf{E}[L^2|H_1] + (1-p) \cdot \mathbf{E}[L^2|T_1] = \frac{p(1+p)}{(1-p)^2} + \frac{(1-p)(2-p)}{p^2}.$$

Now computing the variance of L is straightforward:

$$\begin{aligned} \mathrm{Var}(L) &= \mathbf{E}[L^2] - (\mathbf{E}[L])^2 \\ &= \frac{p(1+p)}{(1-p)^2} + \frac{(1-p)(2-p)}{p^2} - \left(\frac{p}{1-p} + \frac{1-p}{p}\right)^2 \\ &= \frac{1-p}{p^2} + \frac{p}{(1-p)^2} - 2. \end{aligned}$$