These notes are from http://math.cmu.edu/~cocox/teaching/discrete20/rec9.pdf

Let $(\Omega, \mathbf{Pr})$ be a finite probability space. How many mutually independent events can we define in this space? That is to ask, how large can n be (with respect to $|\Omega|$) if $A_1, \ldots, A_n \subseteq \Omega$ are mutually independent? Well, as stated, n could be unbounded since we could take $A_i = \varnothing$ or $A_i = \Omega$ for all i. Therefore, in order to make the question interesting, we must make some assumption on the A_i's.

An event $A \subseteq \Omega$ is called *nontrivial* if $\mathbf{Pr}[A] \notin \{0, 1\}$. This is a reasonable assumption to include on the A_i's since A is independent with A if and only if A is a trivial event. Assuming that each A_i is nontrivial, we can show that m cannot be very large compared to $|\Omega|$; in particular, $n \leq \lg |\Omega|$.

Before proving this, observe that this bound is tight for some probability spaces. Indeed, consider the probability space formed by a sequence of n independent coin-flips; then $|\Omega| = 2^n$ and the events $\{\text{flip } i \text{ is heads}\}$ are mutually independent.

In fact, this is the whole reason that we like to use independent coin-flips to generate probability spaces: we love independence!

Claim 1. *If $A_1, \ldots, A_n \subseteq \Omega$ are mutually independent and nontrivial events, then $|\Omega| \geq 2^n$.*

Proof. For ease of notation, for a set $A \subseteq \Omega$, let $A^1 = A$ and $A^{-1} = \Omega \setminus A$.

For a tuple $x = (x_1, \ldots, x_n) \in \{\pm 1\}^n$, define $f(x) = A_1^{x_1} \cap \cdots \cap A_n^{x_n}$, so $f(x)$ is a subset of Ω formed by intersecting some A_i's and some complements of the A_i's.

Recall that if A, B are independent events, then so are A, B^{-1}, and A^{-1}, B, and A^{-1}, B^{-1}. By a straightforward induction on n, since $A_1, \ldots, A_n$ are mutually independent events, then so are $A_1^{x_1}, \ldots, A_n^{x_n}$ for any $x \in \{\pm 1\}^n$. Using this fact along with the fact that $\mathbf{Pr}[A_i^{x_i}] \neq 0$ since A_i is nontrivial, we see that

$$\mathbf{Pr}[f(x)] = \prod_{i \in [n]} \mathbf{Pr}[A_i^{x_i}] \neq 0.$$

In particular, $f(x) \neq \varnothing$ for every $x \in \{\pm 1\}^n$, i.e. $|f(x)| \geq 1$. Now, consider any $x \neq y \in \{\pm 1\}^n$ and observe that $f(x) \cap f(y) = \varnothing$ since if, say, $x_i = 1$ and $y_i = -1$, then $f(x) \subseteq A$ and $f(y) \subseteq \Omega \setminus A$.

We conclude that

$$|\Omega| \geq \left| \bigcup_{x \in \{\pm 1\}^n} f(x) \right| = \sum_{x \in \{\pm 1\}^n} |f(x)| \geq 2^n. \qquad \square$$

Consider the following random experiment. We begin with r red balls, b blue balls and w white balls in a bucket. We reach in and pick one of the balls uniformly at random:

- If the ball is red, we win.
- If the ball is blue, we lose.
- If the ball is white, we throw it away and redraw.

What is the probability that we win? Well, if $w = 0$, then it's easy to see that $\mathbf{Pr}[\text{win}] = \frac{r}{r+b}$ since we draw either a red or blue ball. What if w is larger?

Claim 2. *For any $w, r, b \in \mathbb{N}$ with $r + b > 0$, we have* $\mathbf{Pr}[win] = \frac{r}{r+b}$.

This makes intuitive sense. Indeed, drawing a white ball simply delays the inevitable event of either winning or losing. So, we can think about the experiment as: wait some random amount of time, and then draw either a red or blue ball uniformly at random.

But how to make this formal? There are a few ways, but let's use the law of total probability. Recall that if events $A_1, \ldots, A_n$ form a partition of our probability space, then for any other event B,

$$\mathbf{Pr}[B] = \sum_{i=1}^{n} \mathbf{Pr}[B \mid A_i] \, \mathbf{Pr}[A_i].$$

Proof of claim. Let W denote the number of white balls that we draw before finally drawing either a red or blue ball. Observe that $W \in \{0, \ldots, w\}$ and the events $\{W = 0\}, \{W = 1\}, \ldots, \{W = w\}$ form a partition of our probability space. We can thus compute,

$$\begin{aligned}
\mathbf{Pr}[\text{win}] &= \sum_{i=0}^{w} \mathbf{Pr}[\text{win} \mid W = i] \, \mathbf{Pr}[W = i] \\
&= \sum_{i=0}^{w} \mathbf{Pr}[\text{win} \mid \text{we draw either a red or blue ball at time } i+1] \, \mathbf{Pr}[W = i] \\
&= \sum_{i=0}^{w} \frac{r}{r+b} \mathbf{Pr}[W = i] = \frac{r}{r+b}.
\end{aligned}$$

where the last equality follows from the fact that $\sum_{i=0}^{w} \mathbf{Pr}[W = i] = 1$. □

Here are a couple things to think about [not discussed in recitation]:

- If we replace the white ball instead of throwing it away, then we still have $\mathbf{Pr}[\text{win}] = \frac{r}{r+b}$. Here's a brief sketch:

  Using the law of total probability as above, we can show that $\mathbf{Pr}[\text{win}] = \frac{r}{r+b}\big(1 - \mathbf{Pr}[W = \infty]\big)$. Then, for any $n \in \mathbb{N}$, $\mathbf{Pr}[W = \infty] \leq \mathbf{Pr}[W \geq n] = \big(\frac{w}{w+r+b}\big)^n$, so

  $$\mathbf{Pr}[W = \infty] \leq \lim_{n \to \infty} \left(\frac{w}{w+r+b}\right)^n = 0.$$

- If we instead double the number of white balls whenever one is drawn, then $\mathbf{Pr}[\text{win}] < \frac{r}{r+b}$ (assuming $r, w > 0$). Here's a sketch:

  As above, using the law of total probability, we can show that $\mathbf{Pr}[\text{win}] = \frac{r}{r+b}\big(1 - \mathbf{Pr}[W = \infty]\big)$. Now, we know that $\sum_{t \geq 0} 2^{-t} < \infty$, so $\lim_{n \to \infty} \sum_{t \geq n} 2^{-t} = 0$. As such, we can select $m \in \mathbb{N}$ so that

  $$\sum_{t \geq m} 2^{-t} < \frac{w}{r+b}.$$

In order to go further, we'll need to make use of Lemma 3 below with $A_n = \{W \geq n\}$ (noting that $\bigcap_{n\geq 0}\{W \geq n\} = \{W = \infty\}$).

$$\begin{aligned}
\mathbf{Pr}[W=\infty] = \lim_{n\to\infty}\mathbf{Pr}[W\geq n] &= \lim_{n\to\infty}\prod_{t=0}^{n}\frac{2^t w}{2^t w + r + b} = \lim_{n\to\infty}\prod_{t=0}^{n}\left(1 - \frac{r+b}{2^t w + r + b}\right) \\
&\geq \prod_{t=0}^{m-1}\left(1 - \frac{r+b}{2^t w + r + b}\right)\lim_{n\to\infty}\prod_{t=m}^{n}\left(1 - \frac{r+b}{2^t w}\right) \\
&\geq \left(1 - \frac{r+b}{w+r+b}\right)^m \lim_{n\to\infty}\left(1 - \sum_{t=m}^{n}\frac{r+b}{2^t w}\right) \\
&= \left(\frac{w}{w+r+b}\right)^m \left(1 - \frac{r+b}{w}\sum_{t\geq m}2^{-t}\right) > 0,
\end{aligned}$$

by the choice of m.

Lemma 3. *Let $(\Omega, \mathbf{Pr})$ be a finite or countable probability space and let $A_0 \supseteq A_1 \supseteq \cdots$ be events. Then*

$$\mathbf{Pr}\left[\bigcap_{n\geq 0} A_n\right] = \lim_{n\to\infty}\mathbf{Pr}[A_n].$$

Proof. Firstly, let's notice that $\lim_{n\to\infty}\mathbf{Pr}[A_n]$ exists since $A_n \supseteq A_{n+1} \implies \mathbf{Pr}[A_n] \geq \mathbf{Pr}[A_{n+1}]$, so $\big(\mathbf{Pr}[A_n]\big)$ is a sequence of non-increasing numbers in $[0,1]$.

Now, set

$$A = \bigcap_{n\geq 0} A_n, \qquad B = \Omega \setminus A, \qquad B_n = \Omega \setminus A_n,$$

so $B = \bigcup_{n\geq 0} B_n$ and $B_0 \subseteq B_1 \subseteq \cdots$. Now, for $n \geq 0$, let $C_n = B_n \setminus B_{n-1}$ (where $B_{-1} = \varnothing$) and observe that $C_1, C_2, \ldots$ are disjoint and $B = \bigcup_{n\geq 1} C_n$. Thus,

$$\begin{aligned}
\mathbf{Pr}[B] = \sum_{x\in B}\mathbf{Pr}[x] = \sum_{x\in\bigcup_{n\geq 0} C_n}\mathbf{Pr}[x] &= \sum_{n\geq 0}\sum_{x\in C_n}\mathbf{Pr}[x] = \sum_{n\geq 0}\mathbf{Pr}[C_n] \\
&= \sum_{n\geq 0}\mathbf{Pr}[B_n \setminus B_{n-1}] = \lim_{N\to\infty}\sum_{n=0}^{N}\big(\mathbf{Pr}[B_n] - \mathbf{Pr}[B_{n-1}]\big) \\
&= \lim_{N\to\infty}\big(\mathbf{Pr}[B_N] - \mathbf{Pr}[B_{-1}]\big) = \lim_{N\to\infty}\mathbf{Pr}[B_N] - \mathbf{Pr}[\varnothing] = \lim_{N\to\infty}\mathbf{Pr}[B_N].
\end{aligned}$$

Therefore,

$$\mathbf{Pr}[A] = 1 - \mathbf{Pr}[B] = 1 - \lim_{n\to\infty}\mathbf{Pr}[B_n] = \lim_{n\to\infty}\mathbf{Pr}[A_n].$$

□