Massachusetts Institute of Technology
6.042J/18.062J: Mathematics for Computer Science
Professors David Karger and Tom Leighton

Handout 53
November 21, 1995

Notes 21

1 Random Variables

Definition 1.1 *A random variable R on sample space S is a function $R : S \to \mathbb{R}$.*

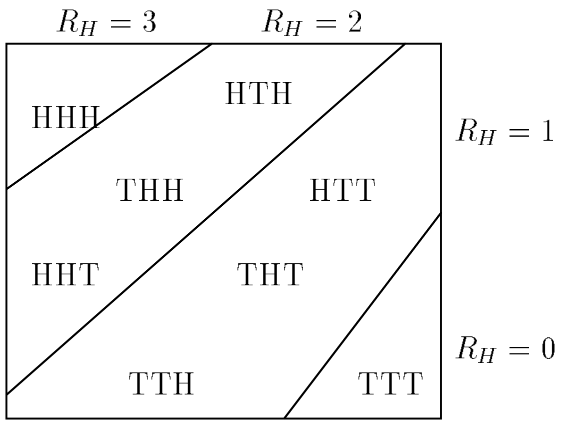
Note defined with respect to a sample space—associates value with each point.

Example 1 *Let S = outcomes of 3 coin tosses = $\{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$.*

- *R_H = number of H in outcome. So $R_H(HHH) = 3, R_H(HHT) = 2$.*
- *$R_M(p)$ = 1 if all coins match, 0 otherwise. So $R_M(HHH) = 1, R_M(HHT) = 0$. R_M is an* indicator *or* characteristic *random variable: 1 if event happens, 0 otherwise.*

Note **functions***:* $R_H : S \to \{0, 1, 2, 3\}$, $R_M : S \to \{0, 1\}$

A random variable partitions S into events (subsets of sample points). "$R = x$" $= \{p \in S | R(p) = x\}$



Definition 1.2 *The* probability distribution function *for a random variable R is*

$$f_R(x) \equiv Pr[R = x] \equiv \sum_{p \in S : R(p) = x} Pr[p]$$

For example,

$$Pr[R_H = 2] = Pr[HHT] + Pr[THH] + Pr[HTH]$$
$$Pr[R_M = 1] = Pr[HHH] + Pr[TTT]$$

Our random variables will be **discrete**–range finite or countably infinite (ie, as here or $R : S \to \mathbb{N}$). There are also *continuous* random variables with values in the reals; for these sums must be replaced by integrals.

2 Independence of Random Variables

Recall, events E_1 and E_2 are independent if $Pr[E_1 \cap E_2] = Pr[E_1]Pr[E_2]$.

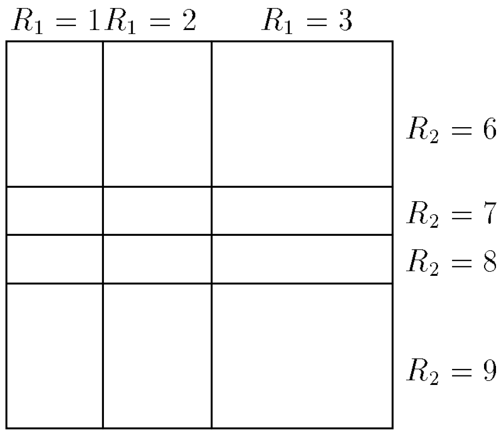
Definition 2.1 *Random variables R_1 and R_2 are independent if for every x, y,*

$$Pr[R_1 = x \cap R_2 = y] = Pr[R_1 = x]Pr[R_2 = y]$$

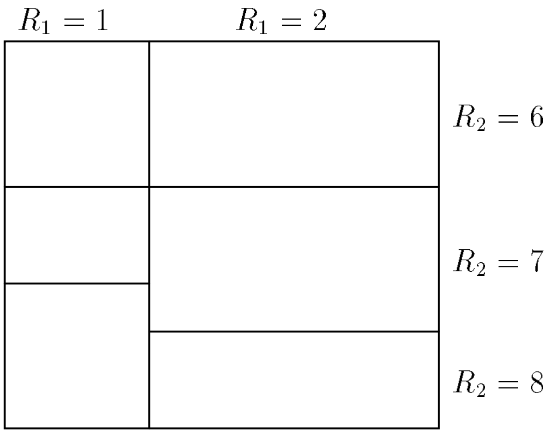
The coins being fair and mutually independent (so each $P \in S$ has probability 1/8) does not mean R_H and R_M are independent, because $Pr[R_H = 2 \cap R_M = 1] = 0 \neq Pr[R_H = 2]Pr[R_M = 1]$.

Intuition: knowing the value of one variable biases the distribution of the other.

The picture, with area proportional to probability:



R_1, R_2 independent



R_1, R_2 not independent

Note that independence of random variables covers *all pairs* of "value events."

Example 2 Suppose we roll two independent fair dice to get values D_1 and D_2. Let $S = D_1 + D_2$ be the sum of values. Let $M = 1$ if $S = 7$ and 0 otherwise.

- S and D_1 are not independent since, for example, $\Pr[D_1 = 6 \mid S = 1] = 0$.
- M and D_1 are independent.

3 The Binomial Distribution

We now study what is perhaps the most common random variable distribution in computer science: the binomial distribution.

3.1 Bernoulli Trials

Coin tossing is a sequence of *Bernoulli Trials*:

- two outcomes (heads/tails, success/failure, good/bad, red/blue)
- trials mutually independent.
- often, Pr[success], denoted p, is the same for each trial (1/2 for a fair coin)

Numerous applications:

- Coding theory: binary messages, faulty channel
- System with n components, each fails with probability p, independent
- false positives in a medical test
- outcome of "are you a democrat" for a random person

With trial i, associate an *indicator random variable* $R_i = 1$ if success, 0 if failure.

- $\Pr[R_i = 1] = p$.
- R_i are mutually independent.

What is the probability distribution for n Bernoulli trials with success probability p?

- Sample space = all $\{H, T\}$ strings of length n.
- independent events $\Rightarrow$ event probabilities multiply
- if a sequence S has k heads, then $\Pr[S] = p^k(1-p)^{n-k}$

3.2 Counting Successes

In many applications, *which* successes/failures happen is less irrelevant—we only care about *how many*.

Let R be the number of heads (a random variable—note $R = R_1 + R_2, \cdots + R_n$). What is the probability distribution for R?

- $\binom{n}{k}$ sequences with k heads
- each has probability $p^k(1-p)^{n-k}$

Definition 3.1 *In n trials with success probability p, the number of successes satisfies the* binomial distribution

$$\Pr[k \text{ } heads] = f_{n,p}(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Sanity check: is this a probability distribution?

$$\begin{aligned} \sum_k f_{n,p}(k) &= \sum \binom{n}{k} p^k (1-p)^{n-k} \\ &= (p + (1-p))^n \quad \text{(binomial theorem)} \\ &= 1 \end{aligned}$$

3.3 Asymptotics

It's hard to see the qualitative behavior of the binomial distribution for large k and n.

Lets consider case $p = 1/2$: uniform distribution. Each outcome has probability

$$f_{n,1/2}(k) = \binom{n}{k} \left(\frac{1}{2}\right)^n$$

Suppose I flip 100 coins

- what is most likely number of heads?
- what probability of exactly 50 heads?
- what probability of exactly 25 heads
- what probability less than 25 heads?
- what probability of exactly 25 heads, given at most 25?

vote: 1, .5, .1, .01, .000001

3.3.1 The central term

Where is $f_{n,p}(k)$ maximized?

$$\begin{aligned} f_{n,p}(k)/f_{n.p}(k-1) &= \frac{n-k+1)p}{k(1-p)} \\ &= 1 + \frac{(n+1)p-k}{k(1-p)} \end{aligned}$$

Therefore, $f_{n,p}(k)$ increases until $k = np$, decreases after.

What is the value at the central term? Let's consider $p = 1/2$, so the central term is at $k = n/2$:

$$\binom{n}{n/2} = \frac{n!}{\frac{n}{2}!\frac{n}{2}!} \sim \frac{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n}{\left(\sqrt{\pi n}\left(\frac{n}{2e}\right)^{n/2}\right)^2} = \sqrt{\frac{2}{\pi n}}2^n$$

(This is the number of subsets of size half the original set)

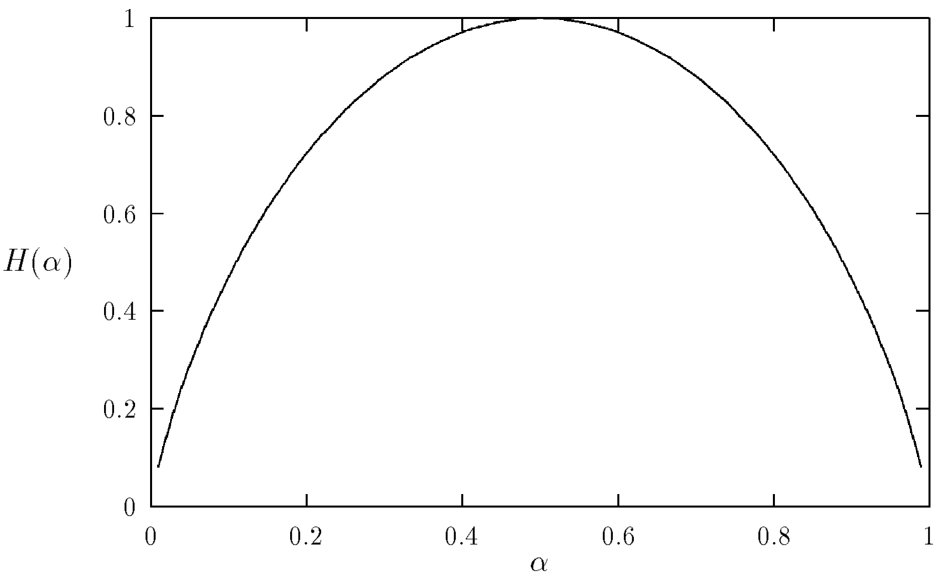
So $f_{n,1/2}(n/2) \sim \sqrt{\frac{2}{\pi n}}$. Note this is an asymptotic bound. For $n = 100$ (our question about coins) is is about 0.080. How close is this to the real answer? To answer this we need to use the form of Stirling's formula that gives upper and lower bounds. If we plug in to that, we find that we are correct to within 1% at $n = 100$.

3.3.2 The tails

Again, let's focus on the case $p = 1/2$, and find $f_{n,p}(k)$ when $k = \alpha n$ for some $\alpha < 1/2$:

$$\begin{aligned} \binom{n}{k} &= \frac{n!}{(\alpha n)!(1-\alpha n)!} \\ &= \sqrt{\frac{1}{2\pi\alpha(1-\alpha)n}}\left(\frac{1}{\alpha^\alpha(1-\alpha)^{(1-\alpha)}}\right)^n \\ &= \sqrt{\frac{1}{2\pi\alpha(1-\alpha)n}}\left(\frac{1}{2^{\alpha\log\alpha+(1-\alpha)\log(1-\alpha)}}\right)^n \\ &= \sqrt{\frac{1}{2\pi\alpha(1-\alpha)n}}2^{nH(\alpha)} \end{aligned}$$

where $H(\alpha) = \alpha\log 1/\alpha + (1-\alpha)\log 1/(1-\alpha)$ is the *entropy function.* It arises in the study of how much information is carried in a binary string with a fraction α of the bits set to one, and thus in the study of error correction over faulty communication links.



Lemma 3.2 *For α a constant $\neq 1/2$, $f_{n,1/2}(\alpha n)$ is exponentially small in n.*

In other words, as n increases, the likely values become concentrated more and more tightly around $\alpha = 1/2$.

Example 3

$$\binom{n}{n/4} \sim \sqrt{\frac{8}{3\pi n}} 2^{0.811n}$$

$$f_{n,1/2}(n/4) \sim \sqrt{\frac{8}{3\pi n}} 2^{-0.189n} \text{ exponetially small}$$

Returning to our beginning questions,

$$Pr[25 \text{ heads}] = f_{100,1/2}(25) \approx 1.88 \times 10^{-7}$$

3.3.3 Cumulative Distribution

For a random variable R, the *cumulative distribution function* $F_R(k) = \Pr[R \leq k]$.

Since $f_{n,1/2}(k)$ is increasing with $k < n/2$, we can bound the probability $F_R(25) \leq 25 \cdot f_{n,1/2}(25)$. But we can do better.

Let $k = \alpha n$ where $\alpha \leq 1/2$. Then

$$\begin{aligned} Pr[R \leq \alpha n] &= \sum_{i=0}^{\alpha n} Pr[R = i] \\ &= \frac{1}{2^n}\left[\binom{n}{\alpha n} + \binom{n}{\alpha n - 1} \cdots \binom{n}{0}\right] \end{aligned}$$

The term in brackets we can bound by a geometric series by looking at the ratio of consecutive terms:

$$\begin{aligned} \frac{\binom{n}{r-1}}{\binom{n}{r}} &= \frac{\frac{n!}{(r-1)!(n-r+1)!}}{\frac{n!}{r!(n-r)!}} \\ &= \frac{r}{n-r+1} \\ &\leq \frac{\alpha n}{n - \alpha n + 1} \\ &\leq \frac{\alpha}{1-\alpha} \end{aligned}$$

(Note that we used $r \leq \alpha n$.) Therefore,

$$\binom{n}{\alpha n - r} \leq \left(\frac{\alpha}{1-\alpha}\right)^r \binom{n}{\alpha n}$$

Plugging in, we find

$$\begin{aligned} Pr[R \leq \alpha n] &\leq 2^{-n}\left[\binom{n}{\alpha n} + \frac{\alpha}{1-\alpha}\binom{n}{\alpha n} + \left(\frac{\alpha}{1-\alpha}\right)^2\binom{n}{\alpha n} + \cdots + \left(\frac{\alpha}{1-\alpha}\right)^{\alpha n}\binom{n}{\alpha n}\right] \\ &\leq 2^{-n}\binom{n}{\alpha n}\frac{1}{1-\frac{\alpha}{1-\alpha}} \quad \left(\alpha < 1/2 \Rightarrow \frac{\alpha}{1-\alpha} < 1\right) \\ &= 2^{-n}\binom{n}{\alpha n}\frac{1-\alpha}{1-2\alpha} \end{aligned}$$

Notice that $\frac{1-\alpha}{1-2\alpha}$ is constant!.

$$Pr[\leq 25 \text{ heads}]\ (n = 100, \alpha = 1/4)\ = (1.88 \times 10^{-7})\frac{3/4}{1/2}$$

For $\beta = o(\sqrt{n})$, $f(n/2 - \beta\sqrt{n}) \sim \sqrt{\frac{2}{\pi n}}e^{-2\beta^2}$.

Furthermore,

$$\Pr[\leq 24 \text{ heads}] \leq \frac{\alpha}{1-\alpha}\Pr[\leq 25 \text{ heads}]$$

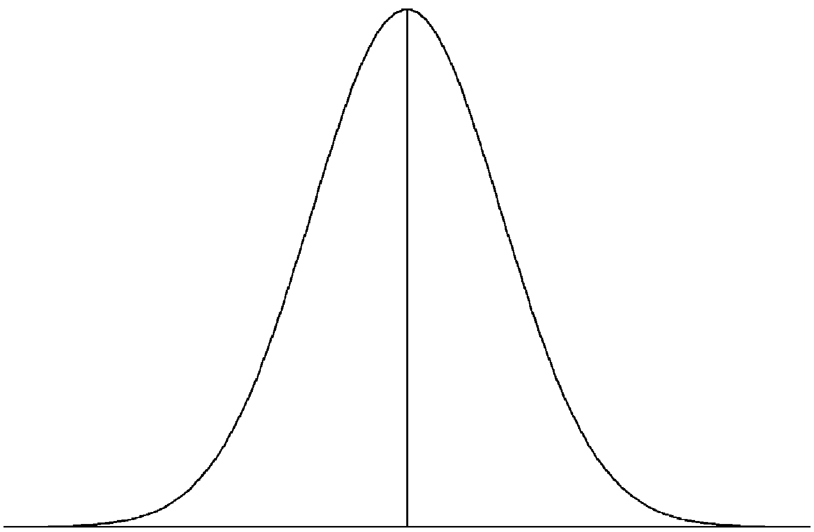
so that

$$\Pr[\text{exactly 25 heads } \mid \text{at most 25 heads}] > 1/2$$

3.3.4 Limiting Distribution

As $n \to \infty$, $f_{n,p}$ approaches "Bell curve" or "normal distribution" (continuous). (Main difference is in the tails, which are exponentially small anyway.)

Here is a plot of the "Bell curve":



4 Applications

4.1 Polling

We want to estimate the percentage of Republicans in the country by conducting a poll. What is the minimum number of people we must sample to be confident of getting an estimate with error at most 5% with a probability of greater than 99%. Assume that every person in the population is chosen uniformly and randomly.

Let us assume that the correct estimate of the percentage of Republicans is 50%. Let our sample size be n. Let X be the random variable that denotes the number of people in our sample who are Republicans. Then we want to find the minimum n such that

$$Pr[|\frac{X}{n} - .5| > 0.05] < 0.01$$

We have that

$$\begin{aligned} Pr[|\frac{X}{n} - .5| > 0.05] &= Pr[X > 0.55n] + Pr[X < 0.45n] \\ &= 2 * Pr[X < 0.45n] && \text{by symmetry, since } p = 1/2. \\ &\leq 2 * (\tfrac{1-0.45}{1-2*0.45}) 2^{-n} \tbinom{n}{0.45n} && \text{geometric series} \\ &\leq 11 * 2^{-n} 2^{H(0.45)n} && \text{Stirling} \\ &= 11 * 2^{-(1-H(0.45))n} \end{aligned}$$

Setting the above to be less than 0.01 we find that

$$11 * 2^{-(1-H(0.45))n} < 0.01 \Rightarrow n > \frac{\log_2 1100}{1 - H(0.45)}.$$

i.e. $n > 1398$.

Now, note that, for all ϵ, $Pr[|\frac{X}{n} - \mu| > \epsilon]$ is maximized when $\mu = \frac{1}{2}$. (Take this fact for granted.)

From this we conclude that a sample size of 1399 suffices even when we drop the assumption that Republicans form exactly 50% of the population.

4.2 Counting Fish

There is a lake with 1 million fish in it. We conduct an experiment on the fish in two phases. In phase 1 we catch A fish, tag them and release them back into the lake. In phase 2 we repeatedly catch and release the fish - we do this B times. What is the minimum total number $(A+B)$ of fish we should catch to make sure that we catch a tagged fish (during phase 2) with probability greater than 99%.

The probability we don't catch even one tagged fish during phase 2 is $(1 - \frac{A}{10^6})^B$. We want this probability to be less than 0.01. Now,

$$(1 - \frac{A}{10^6})^B < e^{-(\frac{AB}{10^6})}$$

. Hence if we assure that

$$e^{-(\frac{AB}{10^6})} < 0.01$$

we are done. This implies that $AB > 4.6 * 10^6$. $A + B$ is minimized when $A = B$ for a given AB. Hence the minimum total number of fish we should catch is $2 * \sqrt{4.6 * 10^6} \sim 4292$

4.3 Leader Election

There are n persons in a room. They wish to pick one of themselves as their leader. They wish to do this in a fair and democratic way, so that each and everyone has the same chance to be the leader. The scheme they employ is for everyone to toss a coin. If exactly one person tosses a head that person is elected the leader. If no persons or more than one person tosses heads then they repeat the entire process. If the coins they use have probability p of coming up heads then what should p be to maximize the probability of selecting a leader in a given round.

If n coins are tossed then the probability of having exactly one head is $\binom{n}{1}p(1-p)^{n-1}$. Differentiating w.r.t. p and equating to 0 we find the maximum occurs when $p = 1/n$. Hence, they should use coins so that the probability of coming up heads is $1/n$. When they use such coins then the probability of selecting a leader in a given round is $\binom{n}{1}(1/n)(1-(1/n))^{n-1} \sim 1/e$.

5 Summary

Unbiased ($p = 1/2$) Binomial Distribution:

$$f(\frac{n}{2}) \sim \sqrt{\frac{2}{\pi n}} = \Theta(\frac{1}{\sqrt{n}})$$

$$f(\alpha n) \sim \sqrt{\frac{2}{\pi n}}\sqrt{\frac{1}{4\alpha(1-\alpha)}}\frac{1}{2^{n(1-H(\alpha))}} \quad (\alpha < 1/2)$$

General ($p \leq 1/2$) Binomial Distribution:

$$f(pn) \sim \sqrt{\frac{1}{2\pi p(1-p)n}} = \Theta(\frac{1}{\sqrt{pn}})$$

$$f(\alpha n) \sim \sqrt{\frac{1}{2\pi\alpha(1-\alpha)n}}2^{n(\alpha\log\frac{p}{\alpha}+(1-\alpha)\log\frac{p}{1-\alpha})}$$