A new way of estimating the probability of heads on a coin

Mark Huber
Fletcher Jones Foundation Associate Professor of Mathematics and Statistics and George R. Roberts Fellow
Department of Mathematical Sciences
Claremont McKenna College

31 Jan, 2016

Today

An estimator $\hat{p}$ for the probability of heads p of a coin, where $\hat{p}/p$ does not depend on p.

What is a bit?

A bit is a number that is 0 or 1.

The smallest unit of information in a digital world.

A coin flip

Gives one bit of information

The bit is random

Calling the coin

- England: Heads or Tails
- Ancient Rome: Navia aut Caput (Ship or Head)
- Argentina: *Cara o Cruz* (Face or Cross)

h. The reverse side of a coin; esp. in phr. *head(s) or tail(s)*: see HEAD *n.*[1] 4b.

1684 T. OTWAY *Atheist* II. 17 As the Boys do by their Farthings..go to Heads or Tails for 'em.

1767 T. BRIDGES *Homer Travestie* (ed. 2) I. III. 101 'Tis heads for Greece, and tails for Troy... Two farthings out of three were tails.

1801 J. STRUTT *Sports & Pastimes* IV. ii. 251 The reverse to the head being called the tail without respect to the figure upon it.

1884 *Punch* 16 Feb. 73/1 A sovereign, a half sovereign,..or farthing, so long as it has a 'head' one side, and..a 'tail' the other.

1893 F. W. L. ADAMS *New Egypt* 267 The goddess who sits on the 'tails' side of our bronze currency.

i. The lower, inner, or subordinate end of a long-shaped block or brick; the bottom or visible part of a roofing slate or tile.

1793 J. SMEATON *Narr. Edystone Lighthouse* (ed. 2) §82 The tail of the header was made to..bond with the interior parts.

1856 S. C. BREES *Terms & Rules Archit.* *Tail*,..the lower end of the slate or tile.

The goal

Let p be the probability of heads

Can flip coin as often as I want

Estimate p

Basic estimate

Basic estimate $\hat{p}_n$:

1. Flip coin n times (Draw $X_1, \ldots, X_n \leftarrow \mathrm{Bern}(p)$ iid.)
2. Let $\hat{p}_n$ be fraction of time coin came up heads.

$$\hat{p}_n \leftarrow \frac{X_1 + \cdots + X_n}{n}.$$

Example: Flip coin 5 times

4 out of 5 heads makes $\hat{p}_5 = 4/5 = 0.8000$.

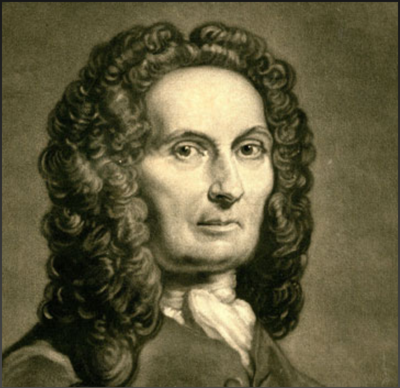
The basic estimate has lead to some great statistics

Jacob Bernoulli took 20 years to prove that Law of Large Numbers holds for $\{0, 1\}$ random variables. (Published posthumously in 1713)

Strong Law of Large Numbers

$\lim_{n \to \infty} \hat{p}_n = p$ with probability 1

How accurate is the basic estimate?



Abraham de Moivre
proved in 1733 an early version of the Central Limit Theorem in order to study how the simple estimate behaves

Relative Error

$$\epsilon_{\mathsf{rel}} = \frac{\hat{\ell}}{\ell} - 1$$

Example

Suppose $p = 20\%$ and $\hat{p} = 22\%$. Relative error is:

$$\frac{22\%}{20\%} - 1 = 1.1 - 1 = 10\%.$$

Relative error using CLT

Use Central Limit Theorem to get rel error at most ϵ with probability at least $1 - \delta$, need

$$2\epsilon^{-2}p^{-1}(1-p)\ln(\delta^{-1})$$

samples.

Problem

- Do not know p
- CLT inaccurate when ϵ, δ small

Relative error for Basic estimate

Relative error depends both on p and n

Example: $n = 5$:

$$\frac{\hat{p}_5}{p} - 1 \in \left\{ \frac{0}{5p} - 1, \frac{1}{5p} - 1, \frac{2}{5p} - 1, \frac{3}{5p} - 1, \frac{4}{5p} - 1, \frac{5}{5p} - 1 \right\}$$

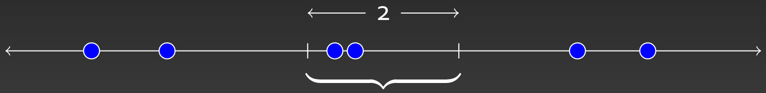
No way is relative error for basic estimate independent of p

The New Algorithm

Point process

Definition

A *point process* is a collection of a random number of points N drawn from a region A, so $\{X_1, \ldots, X_N\} \subseteq A$.

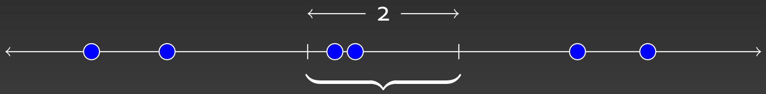


Expected number in interval of length 2 is 2λ

Poisson point process

Definition

A point process is *Poisson* if there is a parameter λ such that for any interval of length a, the average number of points of the process that fall into the interval is λa.



Expected number in interval of length 2 is 2λ

Example: McDonald's

Suppose customers arrive at McDonald's as a Poisson point process

$$\lambda = 90/\text{hour}.$$

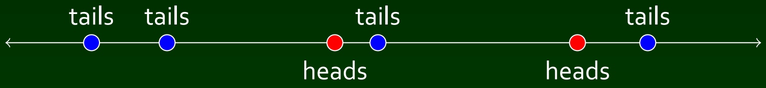
Average number of customers that arrive in the first half-hour is

$$\lambda(1/2)\text{hour} = \frac{90}{\text{hour}} \cdot \frac{1}{2}\,\text{hour} = 45.$$

Changing the rate through thinning

Suppose for each point flip Bern(p)

Only keep points that get heads



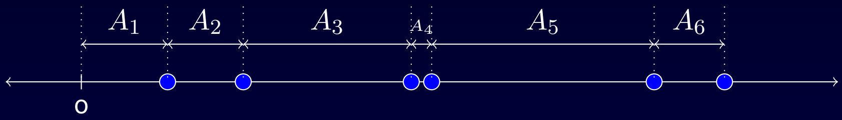
Expected number in interval $[a, b]$ is $\lambda p(b - a)$

New effective rate: λp

Process called *thinning*

Time between points are exponentially distributed

Distances between points are iid exponential r.v.'s of rate λ



$$A_1, A_2, A_3, \ldots \sim \text{Exp}(\lambda) \text{ iid}$$

Drawing $P_1 \sim$ *Exp*(p)

- Generate Poisson process of rate 1 on $[0, 1]$
- Thin it using the p-coin

Time until first head is Exp(p)!

Scaling

Suppose customers arrive McDonald's at rate 90/hour

First customer arrives at time 0.4 hour

Change to minutes:

$$90/\text{hour} \mapsto (90/60) = 1.5/\text{minutes}$$
$$0.4\,\text{hour} \mapsto (0.4)(60) = 24\,\text{minutes}$$

Changing the rate by scaling time

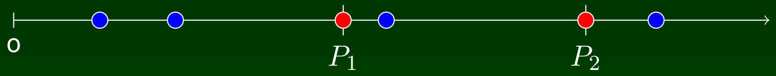
For any constant c:

$$\lambda \mapsto \lambda/c$$
$$P_i \mapsto cP_i$$

Gamma Bernoulli Approximation Scheme

New estimate for p:

- Run Poisson point process of rate 1 forward in time from 0
- Thin the process as it is run forward using p-coin
- Continue until reach k heads
- Let P_k be time of the kth head



[P_k has a gamma distribution with parameters k and p]

The Algorithm

1. Decide what value of k you want to use
2. Flip p-coin until get k heads. Say it takes N flips
3. Generate $A_1, \ldots, A_n$ iid Exp(1) random variables
4. Estimate is $\hat{p} = (k-1)/(A_1 + \cdots + A_N)$.

The Algorithm

1. Decide what value of k you want to use
2. Flip p-coin until get k heads. Say it takes N flips
3. Generate $A_1, \ldots, A_n$ iid Exp(1) random variables
4. Estimate is $\hat{p} = (k-1)/(A_1 + \cdots + A_N)$.

An Example

The Algorithm

1. Decide what value of k you want to use
2. Flip p-coin until get k heads. Say it takes N flips
3. Generate $A_1, \ldots, A_n$ iid Exp(1) random variables
4. Estimate is $\hat{p} = (k-1)/(A_1 + \cdots + A_N)$.

An Example

1. Decide that $k = 4$ is sufficient

The Algorithm

1. Decide what value of k you want to use
2. Flip p-coin until get k heads. Say it takes N flips
3. Generate $A_1, \ldots, A_n$ iid $\text{Exp}(1)$ random variables
4. Estimate is $\hat{p} = (k-1)/(A_1 + \cdots + A_N)$.

An Example

1. Decide that $k = 4$ is sufficient
2. Suppose it takes 22 flips to get 4 heads

The Algorithm

1. Decide what value of k you want to use
2. Flip p-coin until get k heads. Say it takes N flips
3. Generate $A_1, \ldots, A_n$ iid Exp(1) random variables
4. Estimate is $\hat{p} = (k-1)/(A_1 + \cdots + A_N)$.

An Example

1. Decide that $k = 4$ is sufficient
2. Suppose it takes 22 flips to get 4 heads
3. Generate $A_1, \ldots, A_{22}$ iid Exp(1)

The Algorithm

1. Decide what value of k you want to use
2. Flip p-coin until get k heads. Say it takes N flips
3. Generate $A_1, \ldots, A_n$ iid Exp(1) random variables
4. Estimate is $\hat{p} = (k-1)/(A_1 + \cdots + A_N)$.

An Example

1. Decide that $k = 4$ is sufficient
2. Suppose it takes 22 flips to get 4 heads
3. Generate $A_1, \ldots, A_{22}$ iid Exp(1)
4. Final estimate $(4-1)/(A_1 + \cdots + A_{22}) = 0.1823\ldots$

Easy to implement

Six lines of pseudocode

GBAS *Input:* $k \geq 2$	
1)	$R \leftarrow 0$, $S \leftarrow 0$
2)	Repeat
3)	$\quad X \leftarrow \text{Bern}(p)$, $A \leftarrow \text{Exp}(1)$
4)	$\quad S \leftarrow S + X$, $R \leftarrow R + A$
5)	Until $S = k$
6)	$\hat{p} \leftarrow (k-1)/R$

The cool part

Consider the relative error

$$\frac{\hat{p}}{p} - 1 = \frac{k-1}{P_k p} - 1$$

But $P_k p$ is the equivalent of scaling time by a factor of p

- Started with rate 1 process
- Thinned to get rate p process
- Scaling time by p gives rate $p/p = 1$ process again!

Relative error does not depend on p!

Distribution of relative error known exactly

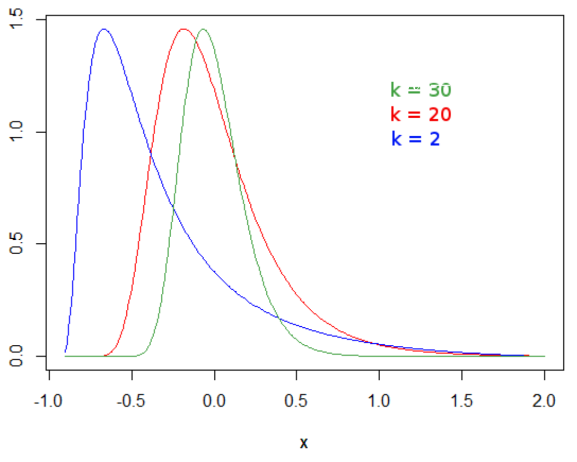
Adding exponentials

- When $T_1, \ldots, T_k \sim$ Exp(λ)...
- ...$T_1 + \cdots + T_k \sim$ Gamma(k, λ)
- So $pP_k \sim$ Gamma$(k, 1)$
- If $X \sim$ Gamma, $1/X \sim$ InvGamma

$$\frac{\hat{p}}{p} \sim \text{InvGamma}(k, 1/(k-1))$$

- $\mathbb{E}(\hat{p}/p) = 1/[(k-1)/(k-1)] = 1)$

As k increases, relative error concentrates about zero



Benefits

Since we know distribution of $\hat{p}/p$ exactly

- Get exact confidence intervals for p easily
- Yields faster randomized approximation schemes
- Theory gives first order same as CLT

Does it work well in practice? YES!

If we knew p exactly

- Exactly find probabilities of tails of binomial distribution
- Use this to find the exact n needed for the basic estimate to be an (ϵ, δ) approximation

The results

- Suppose I want an estimate with absolute relative error at most 10% with probability at least 95%
-

$\epsilon = 0.1$, $\delta = 0.05$			
p	Exact n	$\mathbb{E}[T_p]$	$\mathbb{E}[T_p]/n$
1/20	7 219	7 700	1.067
1/100	37 545	38 500	1.025

The more you know!

Estimators

Math 152 Statistics (every year)

Poisson Processes

Math 156 Stochastic Processes (Fall 2017)

Applications of coin flipping

Math 160 Monte Carlo methods (Spring 2017)

References

Huber, M.,"An unbiased estimate for the mean of a {0,1} random variable with relative error distribution independent of the mean", arXiv:1309.5413, 2013

Huber, M., "A Bernoulli mean estimate with known relative error distribution", Random Structures & Algorithms, to appear