```
Universe of n elements [n]
 [ ][ ][ ][ ][ ][ ][ ][ ][ ][ ][ ][ ][ ][ ][ ][ ][ ][ ][ ]
A "trial" draws a random element from [n].
After k trials, what phenomenon occur?

Birthday Paradox:
  after about k = sqrt{n} trials some element appears twice

Coupon Collectors:
  after about k = n log n trials, we see all elements
  each element appears on average log n times
  all "nice" sets w.p. eps  with  (1/eps) log (1/eps) trials

Today:  When is the distribution "even"
         and Tail Bounds

---------------------

What does "even" mean?

S_k := k random samples from [n]
f_i = # trials with i \in [n] in S_k
      E[f_i] = k/n

W_k = max_i |f_i - k/n|
      as k increases, we expect W_k to grow!
      setting k= 0 is silly

Z_k = max_i |~f_i - 1/n|
      ~f_i = f_i/k   |  fraction of elements with value i
      E[~f_i] = 1/n
      as k increases, we expect Z_k to decrease

If Z_k <= eps we say S_k is an "eps-sample".
  (for eps in [0,1], usually eps = {.1, .001})

How large does k need to be for S_k to be an eps-sample?
   k ~ 1/eps^2
   independent of n.
   Setting eps = c/n has Z_k < c/n and W_k < c
      -->  k = n^2/c^2
           c = 0.1  ->  k ~ 100n^2

(a lower bound of (1/16) * 1/eps^2  /  in practice about (1/2) 1/eps^2)

------
Non-uniform distribution:
```

```
  i in [n] w.p.  p_i
then  Z_k = max_i | ~f_i - p_i|
  -->  again k ~ 1/eps^2

------
Continuous domains:
  Consider any distribution D [draw curve]
  Consider any interval I w.p. p_i of having a random sample from D.
  Z_k = max_{I} |~f_i - p_i|
  Still need k~1/eps^2

  Works in higher dimensions R^d as well.  Instead of intervals -> rectangles,
disks, ...
  need about k~d/eps^2

------
Other error measures:
  We consider **worst case bounds**  why?
    - computer scientists like worst case bounds
    - L_1 or L_2? (average error, average squared error)
       similar results (k~1/eps^2) but can be messier
       or show variance decreases at about same rate


-----------------------------------------------
Tail Bounds:
  Markov Inequality
  Chernoff-Hoeffding Inequality
How to combine rare events:
  The Union Bound

--------
Markov Inequality:
  X = random variable, X>0
  Pr[X > a] <= E[X]/a

[draw distribution]
  - plot average
  - if too much "mass" is too large, then average is too low

Consider if not true:  gamma = Pr[X > a]
E[X] = (1-gamma) 0 + gamma a = gamma a > (E[X]/a)/a = E[X]

--------
Chernoff-Hoeffding Inequality:
  r independent random variables {X_1, X_2, ..., X_r}
    -Delta_i <= X_i <= Delta_i
```

M = sum_{i=1}^r X_i   (average of X_i)
E[M] = 0

Pr[M > a] <= 2 exp(-a^2 / 2 sum_i^r Delta_i^2).

idea for proof: (use Markov Inequality)
  Pr[M > a] = Pr[exp(tM) > exp(t a)]
            <= E[exp(tM)]/exp(ta)]
             = prod_{i=1}^r E[exp(t X_i)]/exp(ta)
            <= prod_{i=1}^r exp(t^2 Delta_i^2 /2) /exp(ta)
             = exp((t^2 /2) sum_i Delta_i^2 - ta)
  choose t = a/sum_i Delta_i ...
             = exp (-a^2 / 2 sum_i Delta_i^2)

  magic Lemma is E[exp(t X_i)] <= exp(t^2 Delta_i^2 /2)

[draw cumulative density plot]
  Central Limit Theorem

Use C-H to prove eps-Sample Bound (for one ~f_i).
Each (of k) trials is a random variable Y_i
  Y_j = 1 if it chooses i in [n]
  Y_j = 0 if it does not choose i
New random variable X_j
  X_j = Y_j - 1/n
  E[X_j] = 0
  Delta_i = 1 (-1/n <= X_j <= 1-1/n)

hat{M} = sum_{j=1}^k Y_j | # random trials with i
    M = (1/n) sum_{j=1}^k X_j = (hat{M}-k)/n
         fraction of random trials more than expected to have index i

Set a = eps, and delta in (0,1)
   Pr[hat{M} - E[hat{M}] > eps n] = Pr[M > a]
           < 2 exp(-a^2 / 2 sum_j Delta_i^2)
           = 2 exp(-eps^2 / 2k) < delta

Solve for k -> k >= (2/eps^2) ln (2/delta)

Some probability delta of failure. Trade-off between eps and delta is exponential!
PAC = "probably approximately correct"

[draw eps-(1-delta) trade-off cumulative density plot]

For only one i \in [n]

Union Bound:
Consider t random variables {Z_1, ..., Z_t}
Z_i = 1 wp p_i and Z_1 = 0 w.p. q_i = 1-p_i
All random variables = 1
w.p p >= 1 - sum_{i=1}^ q_i

Z_i do **not** need to be independent

- add probabilities of failure!

Apply to eps-Samples
n indices to consider

Want bound on k, so >= 1-delta probability of failure on **all** indices.
k = (2/eps^2)ln(2/delta') for one index w.p. 1-delta'
Union bound on n indices -> prob. of failure = delta'* n
-> prob. of success = 1-delta'*n = 1-delta
-> delta' = delta/n

-> k = (2/eps^2) ln (2n/delta)
eps-sample (at most eps error on **all** indices)

- factor of n, but inside ln( )
- can remove ln(n) term (not ln(1/delta)), but much more complicated

----
Numbers: Let eps = 1/10 so 1/eps^2 = 100

delta	ln (2/delta)
.1	3
.05	3.7
.01	5.3
.005	6
.001	7.6
.0005	8.3
.0001	9.9
.00001	12.2
.0000001	14.5 (1 in a million)